

具離群值資料之水文頻率分析

Hydrological Frequency Analysis With Presence of Outliers

國立臺灣大學生物環境系統工程學系
教授

鄭克聲*
Ke-Sheng Cheng

國立臺灣大學生物環境系統工程學系
研究助理

連琮勛
Tsung-Hsun Lien

摘 要

異常極端降雨會使水文頻率分析在判斷重現期雨量上有所失準，進而影響水文治理、水文建設等相關規範。為了能合理推估特定重現期之設計雨量，本文係就異常極端降雨(雨量資料離群值)對水文頻率分析的影響加以探討。

本文之內容，首先探討台灣南部五個縣市其 28 個雨量測站的區域頻率分析，判斷測站的分區方式與各分區極端降雨之機率分布型態。接著採用兩種區域調整變數進行頻率分析，形式一：測站雨量區域調整變數為各測站經過標準化的原始降雨量；形式二：測站雨量區域調整變數為各測站原始雨量值除以該站雨量平均值。估算出各測站在延時 1、2、6、12、18、24、48 與 72 小時下重現期 5 年、20 年、50 年、100 年及 200 年的設計雨量。最後，針對兩種區域調整變數作 1000 次出象之區域頻率分析和單站頻率分析序率模擬，結果顯示以形式一之區域調整變數進行區域頻率分析模擬，結果有近 8 成的機率比單站頻率分析佳。

關鍵詞：異常極端降雨，水文頻率分析，重現期雨量。

ABSTRACT

Rainfalls of extraordinary amounts are outliers in the observed data series. Presence of outliers, particularly in short data series, often causes significant problems in goodness-of-fit tests and distribution parameter estimation which in turn can affect accuracies in estimation of return periods for hydrological observations in hydrological frequency analysis. A regional frequency analysis approach is proposed in this study to cope with problems induced by outliers. Annual maximum series (AMS) of rainfalls of 1, 2, 6, 12, 18, 24, 48 and 72-hour durations from 28 rainfall stations in southern Taiwan

*通訊作者，國立臺灣大學生物環境系統工程學系教授，10617 台北市大安區羅斯福路 4 段 1 號，rslab@ntu.edu.tw

were collected. These rainfall stations were grouped into two homogeneous regions by a k-means cluster analysis which took into account the mean, standard deviation and coefficient of skewness of annual maximum rainfalls of individual durations. Two types of regional rescaled variables were considered for regional frequency: (1) regional variables rescaled with respect to the mean and standard deviation of AMS of individual stations and (2) regional variables rescaled with respect to mean AMS only. Parameters of the regional probability distributions of AMS were estimated independently for individual homogeneous and were then used for regional frequency analysis. A stochastic simulation process which involved 1000 simulation runs was also implemented to provide a quantitative and comparative evaluation of the results of regional frequency analysis and site-specific frequency analysis. It was found that there is an approximately 80% chance that the regional frequency analysis using the first-type rescaled regional variable outperforms the site-specific frequency analysis. Such results provide a theoretical basis for adopting regional frequency analysis when outliers are present in observed data series.

Keywords: Rainfalls of Extraordinary Amounts, Hydrological Frequency Analysis, Return Period Rainfalls.

一、前言

台灣位於西北太平洋颱風經常行進的路線範圍，在西北太平洋生成的颱風平均有 13% 侵襲台灣，而台灣極端氣候的影響，大都來自颱風(許晃雄等，2011)。由國家災害防救科技中心 1970-2009 年的颱風降雨統計資料顯示，極端強降雨颱風(排名前 10% 的颱風降雨)，常造成台灣重大災害，如 2009 年之莫拉克颱風為台灣帶來近 50 年來災情最慘重之水患-八八水災，中南部地區多處更降下刷新歷史紀錄的大雨，此類長延時的異常極端降雨事件在統計上被視為離群值。

異常極端降雨會使得水文頻率分析在判斷重現期洪水量上有所失準，其特定重現期洪水量之準確推估即為水文頻率分析最重要之目的，因此異常極端降雨等離群值資料對水文頻率分析的影響之探討極為重要。由於單一測站之異常極端值紀錄甚少，因此本研究使用台灣南部五個縣市 28 個雨量測站之時雨量資料(最長紀錄年份為 1951-2010 年)，建立歷年 1、2、6、12、18、24、

48、72 小時累積最大降雨量序列。並以群集分析(cluster analysis)方式將雨量資料分群(雨量測站分區)以進行區域頻率分析，再利用適合度檢定選取分區最適分布並計算重現期 5、20、50、100、200 年設計雨量，佐以序率模擬的方式驗證區域頻率分析之結果。

二、研究地區與資料

研究地區為南部五縣市(嘉義市、台南市、高雄市、屏東縣和台東縣)，雨量資料採用水利署 23 個雨量站及中央氣象局的 5 個局屬測站之歷年逐時雨量資料，各站雨量資料長度為 41 年以上，最長紀錄年份為 1951-2010 年。歷年逐時雨量資料經過 1 至 72 小時(1、2、6、12、18、24、48、72 小時)降雨延時之累積雨量運算，得到歷年 1 至 72 小時延時之年最大降雨量序列。表 1 為各測站之 24 小時延時之年最大降雨量序列之資訊與敘述統計結果，圖 1 為各測站分布情形。圖 2 顯示 24 小時延時之年最大降雨量平均值與海拔高度呈現正相關的現象。

表 1 台灣南部五縣市 28 個觀測站 24 小時延時年最大降雨量序列之敘述統計

站號	站名	記錄年數	海拔高度 (m)	平均值 (mm)	標準差 (mm)	最小值 (mm)	中位數 (mm)	最大值 (mm)	最大最小值差異 (mm)	縣市
00J810	北港(2)	45	9	199.33	103.08	38.0	201.4	500.0	462.0	嘉義市
01J960	大埔	49	205	216.18	155.01	2.3	187.0	720.0	717.7	嘉義市
01L390	大湖山	46	725	414.16	318.20	42.0	325.1	1329.0	1287.0	嘉義市
01L480	樟腦寮(2)	45	545	358.75	226.60	50.0	253.9	907.0	857.0	嘉義市
01L490	沙坑	43	78	228.02	151.81	35.0	197.0	632.0	597.0	嘉義市
01L910	中坑(3)	45	95	201.29	135.19	10.5	174.3	522.0	511.5	嘉義市
01M010	溪口(3)	41	17	159.80	95.78	9.9	164.0	352.0	342.1	嘉義市
01O070	關子嶺(2)	47	350	311.47	229.09	27.4	246.0	1098.0	1070.6	台南市
01O080	六溪	51	86	239.66	190.98	31.8	179.0	793.0	761.2	台南市
01P190	木柵	44	78	293.82	133.02	58.1	280.6	778.0	719.9	高雄市
01P500	阿蓮(2)	41	21	256.38	126.60	37.9	237.0	630.0	592.1	高雄市
00P470	旗山(4)	47	64	271.11	116.82	85.0	272.0	558.0	473.0	高雄市
01P660	甲仙(2)	50	355	366.11	175.72	149.0	329.1	1040.0	891.0	高雄市
01P770	美濃(2)	47	61	295.26	105.76	114.7	295.5	551.0	436.3	高雄市
01Q160	新豐	48	166	330.58	158.10	123.7	291.0	908.0	784.3	高雄市
00Q070	屏東(5)	47	25	258.74	159.74	47.6	209.9	799.0	751.4	屏東縣
01Q250	泰武(1)	48	950	643.39	333.78	180.0	584.0	1747.0	1567.0	屏東縣
01Q360	南和	41	140	321.04	167.09	84.0	303.0	773.5	689.5	屏東縣
01Q910	阿禮	42	1320	551.80	304.60	138.0	494.0	1473.0	1335.0	屏東縣
01Q930	三地門	42	150	374.41	207.54	28.5	363.9	867.0	838.5	屏東縣
00S120	鹿鳴橋	43	195	286.15	141.19	36.5	267.6	720.0	683.5	台東縣
01S360	紹家	53	480	269.73	158.58	34.6	232.9	838.0	803.4	台東縣
01S430	霧鹿	47	925	274.80	129.63	44.8	286.0	671.5	626.7	台東縣
467480	嘉義	43	27	244.32	146.07	64.4	212.6	871.0	806.6	嘉義市
467410	台南	60	14	237.03	103.80	63.1	218.2	557.6	494.5	台南市
467440	高雄	51	2	260.30	130.53	65.9	240.4	630.3	564.4	高雄市
467590	恆春	51	22	261.23	98.84	83.5	247.7	528.0	444.5	屏東縣
467660	台東	60	9	257.22	88.35	91.8	246.8	538.3	446.5	台東縣

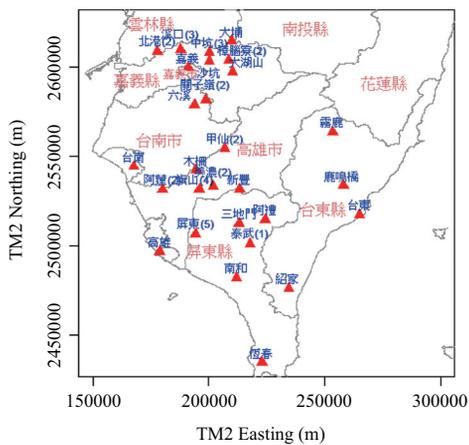


圖 1 台灣南部五縣市 28 個觀測站之分布情形

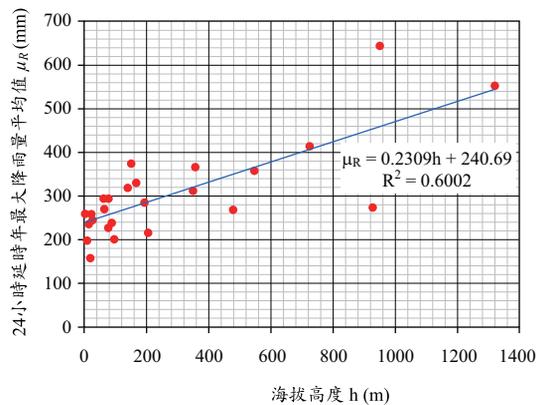


圖 2 台灣南部五縣市 28 個觀測站之 24 小時延時年最大降雨量平均值與海拔高度之關係

三、區域頻率分析理論介紹

本文以區域頻率分析(regional analysis)方式進行水文頻率分析，此章節說明區域頻率分析之理論與計算流程，並敘述研究如何利用線性動差法進行參數推估。

水文頻率分析(hydrological frequency analysis)是利用水文觀測記錄，以統計方法分析水文變量設計值與重現期間的關係，其分析結果常作為水工結構物設計及規劃之依據。

Chow (1951)提出以下之頻率分析通式：

$$x_T = \mu + \sigma K_T \quad \dots\dots\dots(1)$$

式中 x_T 為重現期為 T 之水文變量值， K_T 稱為頻率因子(frequency factor)， μ 及 σ 為水文變數之平均值及標準差。

一般水文頻率分析作法為：(1)擷取水文變數之隨機樣本；(2)進行適合度檢定，決定機率分布；(3)參數推估；(4)決定水文變量設計值。

若水文資料之記錄年期不長且有離群值出現，將影響機率分布之參數推估並造成較高之參數不確定性，最後影響單站頻率分析之結果。因此本文採用區域頻率分析來降低參數不確定性對單站水文頻率分析結果的影響。

區域頻率分析在水文研究及應用方面已行之有年，最早是由 Dalrymple (1960)提出指數洪水法(index-flood method)，現今成為區域頻率分析最常用的方法。

區域頻率分析主要的好處是克服單一測站水文資料紀錄短缺的問題。其透過測站分群的方式將水文特性相似的測站劃分在同一均勻區(homogenous region)，藉此獲得比單一測站還多的水文資訊。區域頻率分析最主要的步驟為擷取水文變數隨機樣本、均勻區劃分、適合度檢定及參數推估，以下說明均勻區劃分、適合度檢定及參數推估之作法。

3.1 均勻區劃分

均勻區劃分的方式有很多種，除了數學計算外也需要一些主觀判斷。用來分區的依據主要有

兩種：(1)單站樣本統計參數；(2)單站地理特徵，包括地理位置、海拔高度、降雨量等。Hosking and Wallis(1997)建議應以群集分析方式，將觀測資料的特性值作分類依據，則同一均勻區內的測站並不一定具有地理上的鄰近關係。

群集分析通常以地文因子(測站座標、高程、面積等)或水文因子(觀測資料之平均值、標準差、偏態係數等)作群集分析之特徵變數。本文採用 k-means 群集分析將研究區域內各雨量站依長短延時年最大降雨量之統計特性(包括期望值、標準差與偏態係數)做為群集分析之特徵變數來進行分群。

3.2 適合度檢定

區域頻率分析在劃分完均勻區之後，各均勻區內之水文變數需通過適合度檢定(goodness-of-fit test)，以決定該均勻區水文變數之最適機率分布。傳統上以無母數統計法之卡方檢定及科司檢定(Haan, 2002)判斷一組隨機變數是否為假設之機率分布。

在分析資料的過程中，我們常有多個機率分布模型可以採用，且適合度檢定之結果有多個機率分布模型未被拒絕，此時可利用模型選取資訊準則(information-criteria-based model selection)選定最適機率分布(best-fit distribution)。

常用的模型選取資訊準則如(2~4)式所示，Akaike information criterion,

$$AIC = \frac{p - l(\theta)}{n} \quad \dots\dots\dots(2)$$

Schwarz's Bayesian information criterion,

$$BIC = \frac{p \log(n) - 2l(\theta)}{2n} \quad \dots\dots\dots(3)$$

Hannan - Quinn information criterion,

$$HQIC = \frac{p \log(\log(n)) - l(\theta)}{n} \quad \dots\dots\dots(4)$$

其中 $l(\theta)$ 為模型之對數概似函數(log-likelihood function)， θ 為模型之參數， n 為樣本量(sample size)， p 為參數空間維度。如該機率分布計算出之 AIC、BIC 與 HQIC 值為最小值，則此機率分

表 2 常用機率分布之機率密度函數 $f(x;\theta)$ 與對數概似函數 $l(\theta)$

Distribution	Density function and log-likelihood function
Gamma distribution	$f_X(x) = \frac{1}{\alpha \Gamma(\beta)} \left(\frac{x-\varepsilon}{\alpha} \right)^{\beta-1} e^{-[(x-\varepsilon)/\alpha]}, \quad \varepsilon \leq x < +\infty$ $\alpha = \sigma / \sqrt{\beta}, \quad \beta = (2/\gamma)^2, \quad \varepsilon = \mu - \sigma \sqrt{\beta}$ $l(\alpha, \beta, \varepsilon) = -n \ln(\Gamma(\beta)) - n\beta(\ln(\alpha)) + (\beta-1) \sum_{i=1}^n \ln(x_i - \varepsilon) - \frac{1}{\alpha} \sum_{i=1}^n (x_i - \varepsilon)$
Log Normal distribution	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_y x} \exp \left[-\frac{1}{2} \left(\frac{y - \mu_y}{\sigma_y} \right)^2 \right], \quad x > 0$ $\mu = e^{\mu_y + \sigma_y^2/2}, \quad \sigma^2 = (e^{\sigma_y^2} - 1)\mu^2$ $l(\mu, \sigma^2) = \frac{n}{2} \{ \ln(2\pi) + \ln(\sigma_y^2) \} - \sum_{i=1}^n \ln x_i - \frac{1}{2\sigma_y^2} \sum_{i=1}^n (\ln x_i - \mu_y)^2, \quad y = \ln x$ <p style="text-align: center;">μ_y 及 σ_y 為 y 之平均值與標準偏差</p>
EV1 distribution	$f_X(x) = \alpha \exp[-\alpha(x-\beta) - e^{-\alpha(x-\beta)}], \quad -\infty < x < +\infty$ $\alpha = \pi / \sqrt{6\sigma}, \quad \beta = \mu - (0.5772/\alpha)$ $l(\alpha, \beta) = n \ln \alpha - \alpha \sum_{i=1}^n (x_i - \beta) - \sum_{i=1}^n e^{-\alpha(x_i - \beta)}$

布為最適機率分布。常用機率分布之機率密度函數與對數概似函數如表 2 所示。

3.3 線性動差法參數推估

本文以線性動差法(L-moment method)進行參數推估，而線性動差法之參數推估方式由 Hosking (1990)提出，為機率權重動差(probability-weighted moment)之線性組合。

Hosking and Wallis(1997)定義隨機變數 X 之 r 階線性動差公式如下：

$$\lambda_r = \int_0^1 x(u) P_{r-1}^*(u) du, \quad r = 1, 2, \dots \quad (5)$$

其中

$$P_r^*(u) = \sum_{k=0}^r P_{r,k}^* u^k, \quad r = 0, 1, 2, \dots \quad (6)$$

$$P_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k} (r+k)!}{(k!)^2 (r-k)!} \quad \dots (7)$$

$x(u)$ 為隨機變數 X 之分位數函數 (quantile function)。

而隨機變數 X 之各階線性動差可由機率權重動差推演而來，其關係式如下：

$$\lambda_{r+1} = \sum_{k=0}^r P_{r,k}^* \beta_k, \quad r = 0, 1, 2, \dots \quad (8)$$

推算可得前四階線性動差與機率權重動差之互換關係如下：

$$\lambda_1 = \beta_0 \quad \dots \dots \dots (9a)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \quad \dots \dots \dots (9b)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad \dots \dots \dots (9c)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad \dots \dots \dots (9d)$$

其中 $\beta_r (r = 0, 1, 2, \dots)$ 為機率權重動差量。第 r 階機率權重動差如(10)式所示：

$$\beta_r = \int_0^1 x(u) u^r du, \quad r = 0, 1, 2, \dots \quad (10)$$

線性動差亦可表示成隨機變數 X 之順序統計量(order statistics)的線性組合：

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}), \quad r = 1, 2, \dots \quad (11)$$

其中 $X_{r-k:r}$ 指樣本量(sample size)為 r 之樣本中，

序級 r - k 之樣本。

前四階線性動差則表示如下：

$$\lambda_1 = E[X_{1:1}] \dots\dots\dots (12a)$$

$$\lambda_2 = \frac{1}{2}E[X_{2:2} - X_{1:2}] \dots\dots\dots (12b)$$

$$\lambda_3 = \frac{1}{3}E[X_{3:3} - 2X_{2:3} + X_{1:3}] \dots\dots\dots (12c)$$

$$\lambda_4 = \frac{1}{4}E[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}] \dots\dots\dots (12d)$$

為了便於把高階線性動差 $\lambda_r (r \geq 3)$ 標準化，可依動差定義變異數(CV)、偏態係數(CS)和峰度係數(CK)為動差量之比值，同樣定義線性動差比值(L-moment ratio)，定義如下：

$$\tau = \frac{\lambda_2}{\lambda_1} \dots\dots\dots (13a)$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \dots\dots\dots (13b)$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \dots\dots\dots (13c)$$

其中 τ 、 τ_3 、 τ_4 分別為線性動差變異數(LCV)、線性動差偏態係數(LCS)與線性動差峰度係數(LCK)。

而 Landwehr *et al.* (1979) 提出機率權重動差值之無偏估計方式。對於 β_r 的樣本推估值 b_r 可由機率權重動差之參數不偏推估式計算，其適用於任何機率分布，計算式如(14)式所示：

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \dots\dots\dots (14)$$

其中 $x_{j:n}$ 指樣本量為 n 的樣本中，序級為 j 之樣本。則任何分布的樣本之前四階線性動差與線性動差比可由下式計算：

$$l_1 = b_0 \dots\dots\dots (15a)$$

$$l_2 = 2b_1 - b_0 \dots\dots\dots (15b)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \dots\dots\dots (15c)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \dots\dots\dots (15d)$$

$$t_r = l_r / l_2 \dots\dots\dots (16)$$

$$t = l_2 / l_1 \dots\dots\dots (17)$$

其中樣本線性動差 l_r 為線性動差 λ_r 之不偏推估值 (unbiased estimator)；樣本線性動差比 t_r 則為線性動差比 τ_r 之推估值，具有推估偏差(bias)存在，而大部分機率分布在樣本量大於 20 時可忽略推估偏差值。 t 、 t_3 、 t_4 分別為樣本線性動差變異數 (sample LCV)、樣本線性動差偏態係數 (sample LCS) 與樣本線性動差峰度係數 (sample LCK)。

本文均勻區配適之機率分布主要以皮爾遜第三型分布(Pearson Type III, PT3)和極端值第一型分布(Extreme Value Type I, EV1)為主，其線性動差法參數推估說明如下。

皮爾遜第三型分布之機率密度函數 (probability density function) 和主要參數如下：

$$f_x(x) = \frac{1}{\alpha \Gamma(\beta)} \left(\frac{x-\varepsilon}{\alpha} \right)^{\beta-1} e^{-[(x-\varepsilon)/\alpha]}, \dots\dots (18)$$

$$\varepsilon \leq x < +\infty$$

$$\alpha = \sigma / \sqrt{\beta}, \quad \beta = (2/\gamma)^2, \quad \varepsilon = \mu - \sigma \sqrt{\beta}$$

其中 μ 、 σ 和 γ 為平均值、標準偏差和偏態係數， α 、 β 和 ε 為尺度(scale)、形狀(shape)和位置(location)參數。

當 α 值大於 0 時 ($0 < \alpha < \infty$)，PT3 分布之線性動差與機率分布參數之關係如下式：

$$\lambda_1 = \varepsilon + \alpha\beta \dots\dots\dots (19)$$

$$\lambda_2 = \pi^{-1/2} + \alpha \Gamma\left(\beta + \frac{1}{2}\right) / \Gamma(\beta) \dots\dots\dots (20)$$

$$\tau_3 = 6I_{1/2}(\beta, 2\beta) - 3 \dots\dots\dots (21)$$

其中 $I_x(p, q)$ 表示不完全 Beta 函數； $\Gamma(\cdot)$ 表示 Gamma 函數。

平均值(μ)、標準差(σ)、偏態係數(γ)與線性動差之關係如下：

$$\mu = \lambda_1 \dots\dots\dots (22)$$

$$\sigma = \lambda_2 \pi^{1/2} \beta^{1/2} \frac{\Gamma(\beta)}{\Gamma\left(\beta + \frac{1}{2}\right)} \quad (23)$$

$$\gamma = 2\beta^{-1/2} \text{sign}(\tau_3) \quad (24)$$

其中 $\text{sign}(x) = \frac{x}{|x|}$ 。

PT3 分布之線性動差法參數推估可令：

$$l_1 = \hat{\varepsilon} + \hat{\alpha}\hat{\beta} \quad (25)$$

$$l_2 = \pi^{-1/2} + \hat{\alpha}\Gamma\left(\hat{\beta} + \frac{1}{2}\right)/\Gamma(\hat{\beta}) \quad (26)$$

$$t_3 = 6I_{\gamma}(\hat{\beta}, 2\hat{\beta}) - 3 \quad (27)$$

其中 $\hat{\alpha}$ 、 $\hat{\beta}$ 和 $\hat{\varepsilon}$ 為尺度、形狀和位置參數之推估值，可由(25~27)式解聯立求得。

極端值第一型分布之機率密度函數和主要參數如下：

$$f_x(x) = \alpha \exp[-\alpha(x - \beta) - e^{-\alpha(x - \beta)}], \quad -\infty < x < +\infty \quad (28)$$

$$\alpha = \pi / \sqrt{6}\sigma, \quad \beta = \mu - (0.5772 / \alpha)$$

其中 μ 和 σ 為平均值與標準偏差， α 、 β 為尺度(scale)和位置(location)參數。

Hosking and Wallis(1997)針對極端值第一型分布之線性動差定義如下：

$$\lambda_1 = \beta + 0.5772 / \alpha \quad (29)$$

$$\lambda_2 = \log 2 / \alpha \quad (30)$$

$$\tau_3 = 0.1699 \quad (31)$$

$$\tau_4 = 0.1504 \quad (32)$$

EV1 分布之線性動差法參數推估可令：

$$l_1 = \hat{\beta} + 0.5772 / \hat{\alpha} \quad (33)$$

$$l_2 = \log 2 / \hat{\alpha} \quad (34)$$

則 EV1 分布之尺度與位置參數推估值 $\hat{\alpha}$ 、 $\hat{\beta}$ 如下所示：

$$\hat{\alpha} = \log 2 / l_2 \quad (35)$$

$$\hat{\beta} = l_1 - 0.5772 / \hat{\alpha} \quad (36)$$

四、降雨均勻區劃分

首先將研究區域內各雨量站延時 1、2、6、12、18、24、48 及 72 小時年最大降雨量之期望值、標準差與偏態係數先標準化後，再做為群集分析之特徵變數來進行降雨均勻區劃分，結果如表 3 所示。

將降雨均勻區劃分為兩個分區，則大湖山、樟腦寮(2)、關子嶺(2)、甲仙(2)、泰武(1)、阿禮及紹家測站為第二分區(subregion II)，因這 7 個測站皆有 4 個延時以上，且為水文頻率分析重點之長延時(48、72 小時)年最大降雨量序列被分為第二分區，而其餘 21 個測站則是有 4 個延時以上及長延時部分列為第一分區而被定為第一分區(subregion I)。

上述分區結果呼應了于與李(2011)之研究，如圖 3 所示，即 1992 至 2010 年歷史颱風侵台期間，南部最容易發生極端強降雨的位置分別是阿里山區以及高屏溪流域山區(本研究之第二分區)，其中又以高屏溪流域山區(阿禮、泰武(1)測站，海拔及降雨量最高之測站)為最。

五、結果討論

5.1 決定分區之機率分布與參數推估

研究採用卡方檢定和科司檢定並配合模型選取資訊準則 AIC、BIC、HQIC 來判斷各測站長短延時年最大降雨量序列之機率分布型態。而區域機率分布之判斷則以各分區測站隸屬機率分布最多之類型為主，結果如表 4 所示。以第一分區來說，卡方檢定判斷極端值第一型分布為最適分布，所佔總比例為 49.69%，而科司檢定則是對數皮爾遜第三型分布佔 47.62%超過極端值第一型分布的 41.67%。考量兩種檢定的統計結果做加權平均，以極端值第一型分布所佔 45.59%比例最高，因此第一分區為極端值第一型分布(EV1)；第二分區在卡方及科司檢定都是對數皮爾遜第

表 3 各測站 1 至 72 小時延時年最大降雨量序列之 K-means clustering 分區結果

測站編號	1	2	3	4	5	6	7	8	9	10	11	12	13	14
站名	北港(2)	大埔	大湖山	樟腦寮(2)	沙坑	中坑(3)	溪口(3)	關子嶺(2)	六溪	木柵	阿蓮(2)	旗山(4)	甲仙(2)	美濃(2)
1hr	2	1	1	2	2	1	2	1	2	1	1	1	2	1
2hr	1	1	2	2	2	1	2	2	2	1	1	1	2	1
6hr	1	1	2	2	1	1	1	2	1	1	1	1	2	1
12hr	1	1	2	2	1	1	1	1	1	1	1	1	1	1
18hr	1	1	2	1	1	1	1	1	1	1	1	1	1	1
24hr	1	1	2	1	1	1	1	1	1	1	1	1	1	1
48hr	1	1	2	2	1	1	1	2	1	1	1	1	2	1
72hr	1	1	2	2	1	1	1	2	1	1	1	1	2	1
測站編號	15	16	17	18	19	20	21	22	23	24	25	26	27	28
站名	新豐	屏東(5)	泰武(1)	南和	阿禮	三地門	鹿鳴橋	紹家	霧鹿	嘉義	台南	高雄	恆春	台東
1hr	1	2	1	2	2	2	1	2	1	2	1	2	1	1
2hr	1	2	2	2	2	2	1	2	1	2	1	2	1	1
6hr	1	1	2	1	2	2	1	1	1	1	1	2	1	1
12hr	1	1	2	1	2	2	1	1	1	1	1	1	1	1
18hr	1	1	2	1	2	1	1	1	1	1	1	1	1	1
24hr	1	1	2	1	2	1	1	1	1	1	1	1	1	1
48hr	1	1	2	1	2	1	1	2	1	1	1	1	1	1
72hr	1	1	2	1	2	1	1	2	1	1	1	1	1	1

註：數字 1 表示分區結果為第一分區，數字 2 為第二分區。

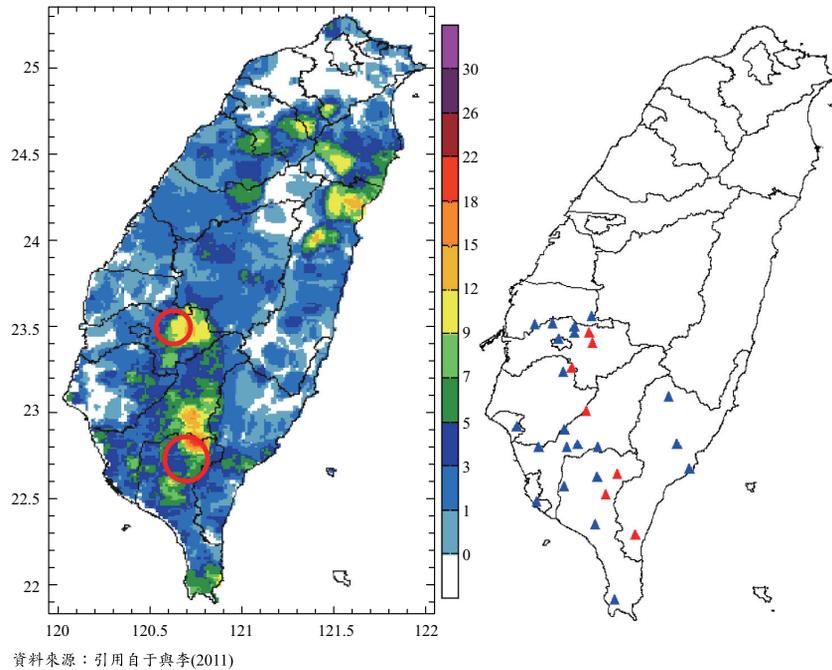


圖 3 台灣 1992 至 2010 年颱風期間強降雨頻率分析與對應之分區位置，左圖紅圈處為本研究之第二分區（右圖之紅色標記點為第二分區測站），且左圖右方之顏色標示代表強降雨之發生次數

表 4 各分區測站所有延時之最適機率分布統計結果

檢定方法	Chi-Square test				Kolmogorov-Smirnov test			
	PT3	LPT3	EV1	LN	PT3	LPT3	EV1	LN
機率分布種類	PT3	LPT3	EV1	LN	PT3	LPT3	EV1	LN
第一分區 count	20	50	80	11	18	80	70	0
第一分區 percent %	12.42	31.06	49.69	6.83	10.71	47.62	41.67	0.00
第二分區 count	6	32	14	3	3	48	5	0
第二分區 percent %	10.91	58.18	25.46	5.45	5.36	85.71	8.93	0.00

註：機率分布計數方式(count)是將隸屬該分區內所有測站及延時之分布種類一起加總

表 5 各分區 1 至 72 延時年最大降雨量頻率因子序列機率分布參數

Duration time	Parameters	α (Type 1)	β (Type 1)	ε (Type 1)	α (Type 2)	β (Type 2)	ε (Type 2)
1hr	Subregion I	1.271	-0.454	-	3.213	0.820	-
1hr	Subregion II	0.151	43.727	-6.613	0.014	43.727	0.374
2hr	Subregion I	1.274	-0.453	-	3.031	0.810	-
2hr	Subregion II	0.066	229.116	-15.134	0.006	229.116	-0.379
6hr	Subregion I	1.260	-0.458	-	2.758	0.791	-
6hr	Subregion II	0.129	59.985	-7.733	0.013	59.985	0.243
12hr	Subregion I	1.260	-0.458	-	2.570	0.775	-
12hr	Subregion II	0.172	34.176	-5.871	0.018	34.176	0.390
18hr	Subregion I	1.262	-0.458	-	2.440	0.763	-
18hr	Subregion II	0.155	42.636	-6.589	0.017	42.636	0.291
24hr	Subregion I	1.260	-0.458	-	2.420	0.761	-
24hr	Subregion II	0.130	60.041	-7.806	0.014	60.041	0.150
48hr	Subregion I	1.270	-0.455	-	2.359	0.755	-
48hr	Subregion II	0.093	115.871	-10.741	0.010	115.871	-0.190
72hr	Subregion I	1.278	-0.452	-	2.353	0.755	-
72hr	Subregion II	0.105	89.924	-9.461	0.012	89.924	-0.058

三型分布比例最高，分別為 58.18%與 85.71%，則第二分區為對數皮爾遜第三型分布(LPT3)。

接著利用線性動差法(L-moment method)推估兩分區之隸屬測站年最大降雨量區域調整變數序列之機率分布參數，並以兩種形式之區域調整變數(regional rescaled variable)進行區域頻率分析，茲介紹如下：

形式一：測站雨量值區域調整變數為各測站經過標準化的原始降雨量(即頻率因子)。

$$\text{Type 1 : } q_1 = (Q - \mu_Q) / \sigma_Q \dots\dots\dots (37a)$$

形式二：測站雨量值區域調整變數為各測站原始雨量值除以該站雨量平均值。

$$\text{Type 2 : } q_2 = Q / \mu_Q \dots\dots\dots (37b)$$

其中 q_1 、 q_2 為區域調整變數， Q 為原始降雨量值，

μ_Q 為測站雨量平均值， σ_Q 為測站雨量標準差。

依分區各測站年份加權得到區域之機率分布參數，計算方式如(38)式。

$$\alpha_i = \frac{\sum_{l=1}^{N_i} \alpha_l L_l}{\sum_{l=1}^{N_i} L_l} \dots\dots\dots (38)$$

其中 N_i 為第 I 分區測站總數， α_i 為第 I 分區之區域機率分布參數， α_l 為第 l 測站之機率分布參數， L_l 為第 l 測站之紀錄年限。

分區之機率分布參數結果如表 5 所示，分別以形式一及形式二之區域調整變數計算區域機率分布參數，結果第一分區各延時之間的 EV1 參數的變動不大，第二分區各延時之間的 LPT3 參數較有變動。

表 6 甲仙站 24 小時降雨延時，重現期 100 年與 200 年之設計雨量值

區域頻率分析方式	$q_1=(Q-\mu_Q)/\sigma_Q$ (形式一)	$q_2=Q/\mu_Q$ (形式二)
$T=100\text{yr}$, $t_r=24\text{hr}$	1,018 mm	1,648 mm
$T=200\text{yr}$, $t_r=24\text{hr}$	1,165 mm	2,000 mm

5.2 兩種區域頻率分析特定重現期之設計雨量值比較

使用延時 1、2、6、12、18、24、48 及 72 小時年最大降雨量序列進行區域頻率分析，以前述之兩種區域調整變數形式來計算重現期 5、20、50、100 及 200 年之設計降雨量，結果顯示以兩種方法計算特定重現期之設計雨量在各分區各有高低。以單一測站甲仙站為例，如表 6 所示。

2009 年莫拉克颱風造成甲仙測站當年度最大日雨量值為 1,040 mm，用形式一推估大約 100 年會發生一次這樣的降雨量，形式二推估則約 20 多年會發生一次。以甲仙站的紀錄年份 54 年看來，只有 1 次年最大值雨量超過 900 mm，採用形式二推估之異常極端降雨發生頻率顯然較高，則以形式一之區域調整變數進行區域頻率分析可能為較合理之推估方式。

5.3 區域頻率分析之序率模擬驗證結果

為驗證區域頻率分析比單站頻率分析更能合理計算特定重現期之設計雨量，研究利用序率模擬的方式，針對兩種形式之降雨量區域調整變數個別作 1,000 次出象之序率模擬，模擬各站 24 小時延時($t_r=24\text{hr}$)年最大降雨量序列，以評估兩分區重現期 5、20、50、100 及 200 年($T=5\sim 200\text{yrs}$)設計雨量之區域頻率分析模擬值、單站頻率分析模擬值何者較貼近理論值。

以兩種形式之降雨量區域調整變數進行序率模擬，模擬步驟敘述如下：

(1) 模擬降雨量序列：

使用測站 24 小時延時年最大降雨量觀測值之區域調整變數序列，以線性動差法(L-moment method)推估該序列之機率分布參數，以此推估參數產生測站之區域調整變數模擬值序列，各測站序列模擬長度為該站之紀錄年限。最後將各站之區域調整變數模擬值序列轉換為降雨量模擬值

序列。

(2) 計算理論之重現期設計雨量 $d_i(t_r=24,T)$ ：

用步驟(1)得到之各測站機率分布參數，計算各站理論重現期 5~200 年之設計雨量區域調整變數，並將理論重現期設計雨量之區域調整變數轉換為理論重現期設計雨量 $d_i(t_r=24,T)$ 。

(3) 分區之機率分布參數推估與計算區域頻率分析模擬之重現期設計雨量 $d_i^{(R)}(t_r=24,T)$ ：

用步驟(1)求得之各站降雨量模擬值序列，計算該模擬值序列之平均值與標準差，接著將各站模擬值序列之降雨量轉換成雨量區域調整變數模擬值。

以線性動差法推估各站雨量區域調整變數模擬值序列之機率分布參數，配合測站紀錄年限，加權計算分區之機率分布推估參數值(計算方式如(38)式)。

最後以分區之機率分布推估參數，來計算各分區隸屬測站，其區域頻率分析模擬之重現期 5~200 年設計雨量區域調整變數，配合前述計算之各站降雨量模擬值序列的平均值與標準差，將該區域調整變數轉換成區域頻率分析模擬之重現期設計雨量 $d_i^{(R)}(t_r=24,T)$ 。

(4) 計算單站頻率分析模擬之重現期設計雨量：

用步驟(3)推估之各站雨量區域調整變數模擬值序列之機率分布參數，計算各站單站頻率分析模擬之重現期 5~200 年設計雨量區域調整變數，將該區域調整變數轉換成單站頻率分析模擬之重現期設計雨量 $d_i^{(S)}(t_r=24,T)$ 。

(5) 以均方根誤差之計算結果驗證區域頻率分析較能合理計算特定重現期之設計雨量：

比較 $d_i(t_r=24,T)$ 與 $d_i^{(R)}(t_r=24,T)$ 、 $d_i(t_r=24,T)$ 與 $d_i^{(S)}(t_r=24,T)$ 於不同分區重現期 5~200 年之均方根誤差(RMSE)，其均方根誤差平均值較小者表示此種頻率分析之重現期設計雨量較貼近理論值。

表7 兩種區域調整變數形式之區域及單站頻率分析模擬重現期雨量值與理論重現期雨量值(mmm)之均方根誤差(RMSE)統計敘述

Type	RMSE 平均值	第一分區之5年 重現期雨量值			第一分區之20年 重現期雨量值			第一分區之50年 重現期雨量值			第一分區之100年 重現期雨量值			第一分區之200年 重現期雨量值			第一分區之5個 重現期雨量值			全區5個 重現期之雨量值														
		區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比												
		1	30.08	29.61	33.70	49.28	48.01	32.50	62.15	60.39	32.00	71.95	69.82	32.00	81.79	79.29	32.10	62.05	60.29	31.90	244.38	362.97	76.80											
2	5.32	5.22		8.87	8.54		11.24	10.78		13.04	12.48		14.84	14.18		11.15	10.68		112.18	247.26														
Type 1	RMSE 平均值	31.79	30.32	34.40	59.69	49.44	12.90	78.81	62.28	7.90	93.38	72.05	6.00	108.00	81.86	5.00	79.34	62.17	6.80	359.59	335.68	27.20												
Type 2	RMSE 標準差	5.15	5.50		7.78	8.70		9.41	10.81		10.61	12.42		11.81	14.03		9.20	10.69		58.73	220.40													
Type	RMSE 平均值	第一分區之20年 重現期雨量值			第二分區之20年 重現期雨量值			第一分區之50年 重現期雨量值			第二分區之50年 重現期雨量值			第一分區之100年 重現期雨量值			第二分區之100年 重現期雨量值			第一分區之200年 重現期雨量值			第二分區之200年 重現期雨量值			第一分區之5個 重現期雨量值			第二分區之5個 重現期雨量值					
		區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比	區域- 理論值 RMSE	各站- 理論值 RMSE	區域 分析 較佳 百分比						
		1	73.29	71.98	46.40	206.81	232.68	71.00	367.43	475.02	75.10	546.61	786.89	76.40	793.85	1273.06	77.50	479.35	723.51	76.90	479.35	723.51	76.90	25.72	24.39		82.91	101.84		231.68	503.62		231.68	503.62
Type 2	RMSE 標準差	95.48	71.12	17.40	314.68	222.20	13.30	561.46	444.67	20.10	823.39	728.25	25.90	1169.66	1167.38	31.00	713.22	666.83	27.90	713.22	666.83	27.90	95.48	71.12		61.52	94.04		120.56	449.92		120.56	449.92	

最後模擬之結果如表 7 所示，以形式一之區域調整變數進行 1,000 次序率模擬中，全區域重現期 5 至 200 年有 768 次模擬顯示區域頻率分析結果比單站頻率分析好(佔 76.8%)；以形式二模擬則只有 272 次(佔 27.2%)，但兩種模擬方式皆顯示區域頻率分析與理論值之 RMSE 標準差較小。

以不同分區之模擬結果而言，用形式一之區域調整變數進行 1,000 次序率模擬在第二分區之區域頻率分析表現較佳(佔 76.9%)，而在第一分區則區域頻率分析與單站頻率分析之重現期雨量模擬結果相差不大；形式二之區域調整變數在兩個分區的序率模擬皆是單站頻率分析表現較佳。

總結序率模擬驗證成果，在 1,000 次序率模擬中，用形式一之區域調整變數進行區域頻率分析模擬結果有近 8 成的機率比單站頻率分析好，且形式一之區域調整變數比形式二更適合用來做區域頻率分析，因此建議以形式一之區域調整變數($q_1 = (Q-\mu_Q)/\sigma_Q$)進行水文區域頻率分析。

六、結 論

台灣南部在夏秋之際常有颱風來襲，往往伴隨極端強降雨的發生，因此颱風是造成異常極端降雨的關鍵。本文係探討異常極端降雨(雨量資料離群值)對水文頻率分析的影響，並得出以下結論：

1. 研究使用台灣南部五個縣市 28 個雨量測站之時雨量資料，建立延時 1、2、6、12、18、24、48、72 小時之年最大降雨量序列，以 k-means 群集分析將雨量測站分為兩區，第二分區符合于與李(2011)之研究，即 1992 至 2010 年歷史颱風侵台期間，南部最容易發生極端強降雨的位置分別是阿里山區以及高屏溪流域山區。
2. 採用卡方檢定與科司檢定配合模型選取資訊準則，決定分區極端降雨之機率分布型態，判斷第一分區為極端值第一型分布(EV1)，第二分區為對數皮爾遜第三型分布(LPT3)。利用線性動差法(L-moment method)推估兩分區各測站年最大降雨量區域調整變數 $q_1 = (Q-\mu_Q)/\sigma_Q$

(形式一)和 $q_2 = Q/\mu_Q$ (形式二)之機率分布參數，第一分區各延時之間的 EV1 參數分別在形式一及形式二的變動不大，第二分區之 LPT3 參數較有變動。

3. 為驗證區域頻率分析之結果，分別以形式一與形式二之區域調整變數作 1000 次出象之序率模擬，則形式一之區域頻率分析模擬結果有近 8 成的機率比單站頻率分析佳，且形式一之區域調整變數比形式二更適合進行區域頻率分析，因此建議以形式一之區域調整變數($q_1 = (Q-\mu_Q)/\sigma_Q$)進行水文區域頻率分析。本文之研究結果可作為水文治理、水文建設等參考。

參考文獻

1. Ani, S. and Abdul, A. J., "Estimation of the Extreme Value Type I Distribution by the Method of LQ-Moments," *Journal of Mathematics and Statistics*, 5(4), pp. 298-304, 2009.
2. Chou, C., Tu, J.Y. and Tan, P.H., "Asymmetry of tropical precipitation change under global warming," *Geophysical Research Letters*, 34, 2007.
3. Chow, V. T., "A general formula for hydrologic frequency analysis," *Transaction of American Geophysics Union*, 32, 1951.
4. Coles, S., "An introduction to Statistical Modeling of extremes values," Springer-Verlag: London, 2001.
5. Dalrymple, T., "Flood frequency analyses," U.S. Geological Survey Water Supply Paper, 1543-A, U.S., 1960.
6. Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R., "Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form," *Water Resources Research*, 15(5), pp. 1049-1054, 1979.
7. Haan, C. T., Statistical methods in hydrology, Iowa State University Press, Ames, Iowa, 2002.
8. Held, I.M. and Soden, B.J., "Robust responses

- of the hydrological cycle to global warming,” *Journal of Climate*, 19, pp. 5686-5699, 2006.
9. Hosking, J.R.M. and Wallis, J.R., Regional frequency analysis: an approach based on L-moments, Cambridge University Press, Cambridge, U.K., 1997.
 10. Hosking, J.R.M., “L-moments—Analysis and estimation of distributions using linear combinations of order statistics,” *Journal of the Royal Statistical Society, Series B*, vol. 52, pp. 105-124, 1990.
 11. Landwehr, J. M., Matalas, N. C., and Wallis, J. R., “Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles,” *Water Resources Research*, 15(5), pp. 1055-1064, 1979.
 12. Liou, J.J., Wu, Y.C., and Cheng, K.S., “Establishing acceptance regions for L-moments based goodness-of-fit tests by stochastic simulation,” *Journal of Hydrology*, 355, pp. 49-62, 2008.
 13. Liu, S. C., Fu, C., Shiu, C. J., Chen, J. P. and Wu, F., “Temperature dependence of global precipitation extremes,” *Geophysical Research Letters*, 36, pp. 2-4, 2009.
 14. and Analysis of Multivariate Observations, pp. 281-297, University of California Press, Los Angeles, 1967.
 15. Sun, Y., Solomon, S., Dai, A. and Portmann, R.W., “How often will it rain?” *Journal of Climate*, 20, pp. 4801-4818, 2007.
 16. 于宜強、李宗融，「高屏河流域極端降雨事件研究」，國家災害防救科技中心 2011 年 10 月第 75 期災害防救電子報，2011。
 17. 吳宜珍，「序率模擬應用於氣候變遷衝擊評估與線性動差機率分布適合度檢定」，國立台灣大學生物環境系統工程學研究所博士論文，2011。
 18. 涂建翊、陳俐盈，「從時雨量看台灣降雨強度的頻率變化」，2011 年地球科學系統學術論壇論文集，113-117 頁，2011。
 19. 許介維，「序率模擬應用於機率分佈適合度檢定之評估」，國立台灣大學生物環境系統工程學研究所碩士論文，2004。
 20. 許晃雄、陳正達、盧孟明、陳永明、周佳、吳宜昭等，「臺灣氣候變遷科學報告 2011」，科技部，2011。

收稿日期：民國 104 年 4 月 8 日

修正日期：民國 104 年 6 月 2 日

接受日期：民國 104 年 6 月 10 日