

訊息熵在雨量站網設計之應用

Application of Information Entropy to Raingauge Network Design

國立臺灣大學農業工程學系
博士班研究生

江介倫

Jie-Lun Chiang

國立臺灣大學農業工程學系副教授
兼水工試驗所副研究員

鄭克聲

Ke-Sheng Cheng

摘要

本研究主要引進降雨資料的熵值(Entropy)概念，經群集(Cluster)分析後，將之應用在雨量站網設計。即在探討如何利用熵理論，分析既有的雨量資料在時間及空間上分佈的特性，來度量集水區中已有雨量資料的訊息量及其中資訊的轉移；以利掌握整個集水區系統的降雨資訊，進而能設計出適合的雨量站網，以便使雨量站所記錄的降雨資料能充分表現出該區的降雨分佈特性。

分析淡水河流域，40個雨量站的時雨量資料，由群集分析可以將雨量站分為三群，第一群（偏西）有16站、第二群（居中）有15站、第三群（偏東）有9站；分析三群資料在空間上之分佈，顯示此降雨資料之結構為非等向性，在東西向變化較快。

評估結果三群中以第三群之降雨變化最大，第一群最小。三群之代表測站由東至西分別為巴陵、木柵、桶後等三站。而在獲取相同資訊量的要求下，所需雨量測站數目亦以第三群最多，第一群最少；本研究並探討欲獲得不同資訊量，如60%、70%、80%、90%、95%等所需測站數目。就所需測站的數目與分佈和淡水河流域的降雨特性來看，以訊息熵的概念用在雨量站網設計應有相當的可行性。

關鍵詞：訊息熵，雨量站網。

ABSTRACT

In this study a rain gauge network design approach using cluster analysis and the theory of entropy is proposed and applied to Tansui River Basin. Both representative information and redundant information are considered through a repeated ranking procedure.

Total information of rainfall of the watershed was determined and used to evaluate the number of stations that are needed to obtain 60%, 70%, 80%, 90%, and 95% of total information. The results exemplify the potential use of entropy theory to monitoring network design.

Keywords: Information entropy, Rain gauge network.

一、前言

熵 (Entropy) 原屬熱力學之領域，十九世紀中，克勞修斯 (Clausius) 最先將熵的概念引進熱力學，熵在熱力學中的定義為從一狀態到另一狀態所經歷之最大亂度。熵也是一個衡量秩序的度量，熵值愈大，代表系統中的混亂程度愈大。

而在早期，熵就被以能量的觀點應用在氣象的分析與預報上，即氣象領域裡的等熵分析 (Isoentropic Analysis) 所謂的等熵分析就是在等位溫面上分析氣象變數。在另一方面 Boltzmann 將之用於統計力學，而給了熵機率的意義；近來引入統計學中，以熵來判斷重要分佈模式之不確定性 (Uncertainty) (張學文、馬力，1992)。

於 1940 年代末期，由於訊息理論 (Information Theory) 的需要而有 Shannon 熵的出現。1950 年代末為因應遍歷理論 (Ergodic Theory) 問題而有 Kolmogorov 熵；到了 1960 年代中期發展了拓樸熵 (Topological Entropy) 的概念，以研究拓樸動力系統 (Topological System) (李天岩，1989)，這些都是有關不確定性的度量方法。Burg (1975) 之最大熵法說明有關消息之不確定性或不連續之隨機變數的時間序列。且將之應用在各種自不足資料 (Incomplete data) 中，決定未知參數的問題。

雨量站網評估與設計的研究為水利工程中之一項重要課題，早期就開始這方面的研究，而其中所用的方法、原理有變異數 (Variance) 分析、空間相關函數 (Spatial-correlation Function)、克利金法 (Kriging) 還有本研究所採用的訊息熵 (Information Entropy) 等等。在此僅針對以訊息熵的方法為出發點的研究作簡單的介紹。

Casleton and Husain (1980) 將 Shannon 的

Entropy 用在水文站網的設計評估，他求出各站之熵值，再內插得較多點位之熵值，以此值畫出等值線，用以設計站網，但其中只考慮到各測站以及內插加密以後各點的熵值，並未考慮研究區域整體資訊的重覆或相關性。

在 Husain, Ukayli and Khan (1986) 氣象站網的設計研究中，直接以各站之熵值來做設計之依據，亦未曾考慮到各站與整個集水區系統間的整體的資訊轉移之相關性。此外 Harmancioglu and Yevjevich (1987) 曾以熵度量同一流域中，河川的各流量站水文資料的最大可轉移資訊。

Kristanovic and Singh (1992a,b) 應用熵理論來做雨量站網的評估。他們認為不確定性愈大的地方，表示其降雨的變化愈大，愈是難以掌握，而若欲掌握降雨的分佈則需在不確定性較大的位置設站，以取得該處之降雨資料，故選擇不確定性最高的站為最重要的站，然而此站對系統的代表性尚值得探討；但其中另需考慮各站間資訊轉移的情況；並畫出等資訊線，即由非轉移資訊畫出等值線 (依所需站數畫出該站數之等值線)，則可以瞭解在整個集水區系統中，資訊轉移的狀況。以考慮整個集水區系統內之降雨分佈，在該研究中考慮到了各站與整個系統之資訊相關性 (取相關性較小的站加入)，故能達到重覆資訊最少的要求。但此分析無法客觀判斷實際需要測站數目，且測站對整體系統之代表性也值得商榷，故本文以此為出發點進行探討與研究。

二、研究方法與理論

本研究採用訊息熵的理論為主，並先以群集 (Cluster) 分析初步分類，再分區做雨量測站重要性的排序，最後以資訊量達到飽和的程度評估所需要的測站數目，以下將對本研究的概念以及方

法、理論做簡單的介紹。

(一) 研究概念

在以往雨量站網設計、規劃的研究中，大多分析站網的密度及測站與測站之間的距離，但是台灣地屬海島型氣候而且地形複雜、高程差距相當大，高山常將氣流抬升造成地形雨，使迎風面多雨，背風面少雨，如此雖然距離不大但是降雨型態的差異卻相當大，故若直接分析降雨的資料而非取決於其單一的因子（例如測站間的距離）應當較為合理。

設計雨量站網時，需要讓測站所記錄的雨量資料能夠充分表現該區的降雨分佈型態，而且測站間的訊息重覆性應盡量減低，以符合經濟效益。故雨量站網設計的兩大訴求為：

1. 測站具代表性，所測得的降雨記錄不只是區域中幾個點的降雨狀況而且要能真實記錄該區的降雨特性，能夠充分代表該地區的狀況。
2. 減少不必要的測站，即重覆資訊夠多的測站就無須設立，換言之，存在的各個測站間的重覆情況應當盡可能減少。

以上兩者要兼顧，在實際設計上實有困難，選擇重覆資訊較少的測站即會選出資料結構比較不同的測站，但因此又會使其代表性產生問題；以 Kristanovic and Singh (1992a,b)的雨量站網評估之研究為例，選擇不確定性較大的測站，做為中心站（最早加入的測站），雖然它的訊息量較大，但是其資料特性可能大異於區域內的其他測站，即其代表性值得懷疑，若是選用區域中心站為代表站則會因為區域內的地形、氣候、高程等各項條件為不均勻時而有所偏差；而在依序選擇重覆資訊少的第二、第三、第四、...等站加入站網系統時，所得到的測站大多是距離較遠的站。亦即避免了重複資訊但卻使代表性出現問題；但可由其中發現降雨在空間上的變異性可由直接由降雨記錄得到。

若直接於全區域之所有測站中，選出重覆資訊最大的測站為代表站，因為其涵蓋範圍太廣可能有多種不同型態的降雨結構，非常難以單一測

站來完全代表全區域之降雨特性，故先將各雨量測站作群集分析，把性質較相似的測站歸類之後，再於各群中選取與其他測站重覆資訊最多的測站為代表站，藉此使全區中的主要特性能由各分群之代表站來代表；如此較能兼顧代表性以及重覆資訊較少的兩個要求。

而群集分析時之所以選用矯值的相似程度來作聚類，是因為降雨受地形、高程、緯度、距海遠近等綜合的影響，若只取其中幾個要素則不夠客觀、完整的分析其特性。故在本研究中，直接由記錄資料的資訊量來作分析，比較能夠客觀的看出各因素影響之下的綜合結果。

而在做測站重要性排序時需考慮增加測站與原有系統間資料的重覆性，當測站增加至一定數目以後資訊量的增加會漸趨緩和，終而達到飽和的狀態。故可以藉此來評估所需要的測站數目。

(二) 研究方法

本研究以矯來度量降雨資訊的不確定性及其資料之時間變化特性，再以群集分析，將矯值較相近（即資料結構較類似）的測站歸納為同一類；如此可以分析出整個大區域中降雨的時間與空間分佈型態，將區域中的雨量站作合理的歸類。

於各群中選出重覆資訊最多的測站，作為該群的代表站。最後在各群中依序選擇重覆資訊最少的測站加入，以便使重覆資訊達到最小的情形。最後繪出測站數目與資訊量遞增的關係圖，以評估欲獲得的資訊量與其所需要的測站數目。

(三) 研究理論

在本研究中先以矯值來度量降雨資料的資料結構型態，再以群集分析將各個既有的雨量測站加以分門別類，再利用整個系統的條件矯來度量單站資料對於整個系統訊息的重覆量，以此對各測站的重要性做排序，再評估所需的測站數目。其中群集分析顧名思義群集分析就是一種把資料分類歸納成群的一種分析方法，它根據事物

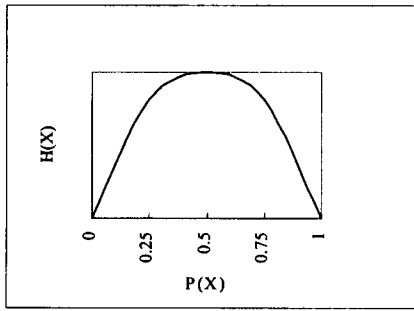


圖 1 兩種訊息時熵與機率關係圖

的異同等性質，將相似的事物歸為同一類，在一般統計的方法中常可以見到。以下將針對熵的理論作簡單的介紹。

如果只有一種可能訊息，即事件百分之百會發生，則熵值為零，不確定性為零。若只有兩種可能訊息時，例如擲硬幣只有正面或反面兩種結果；或是例如降雨與否，只有發生降雨與不發生降雨兩種選擇時，則降雨的機率 P 與其對應的熵值 H 的關係如圖 1 所示，由於事件的總合為 1 當發生之機率為 0.5 時，不發生之機率亦為 0.5，此時有最高的不確定性，故此時熵值為最大。當發生的機率等於 0.25 時，不發生的機率等於 0.75，反之若是發生的機率等於 0.75 時，不發生的機率就等於 0.25，所以二者的不確定性是一樣的。

Hartly (1928)將事件發生結局種類的對數值 ($\log n$) 稱為訊息量，例如投擲骰子時，可能出現 1~6 等 6 種狀況，即 $n=6$ ，其訊息量為 $\log 6$ ；而在投擲硬幣時可能出現的結果只有正面與反面兩種情況即 $n=2$ ，故訊息量為 $\log 2$ ；而訊息量愈大，則表示可能發生的變化較多，也就是說不確定性也愈大，故由以上擲骰子的訊息量及擲硬幣的訊息量可以知道，擲骰子的不確定性比擲硬幣要高。而 Shannon 以 $H = C \log n$ 表示某實驗結果之不確定性，即稱訊息熵。

若 n 種結局發生的機率皆為 $P(x) = 1/n$ ，則亦可寫成：

$$H(X) = -C \log P(X_n) \dots\dots\dots(1)$$

$$H(X) = -C \sum_{n=1}^N P(X_n) \log P(X_n) \dots\dots\dots(2)$$

將(2)式用於本研究領域時可將 X 視為雨量測站的資料，其記錄長度為 N ，變量 X 可視為一隨機變數，且其第 i 個事件發生的機率為 P_i 。

除了以熵值來度量各站的雨量資料在時間上的變異結構之外(以(2)式度量單一測站本身的平均不確定性)，再引入聯合熵及條件熵等概念用以考量整個系統資訊的相關性。

聯合熵 (Joint Entropy)：用在本研究內，可代表幾各測站聯合的不確定性。若 X_i 與 X_j 為兩個雨量測站之變量資料，如時雨量資料； X_i 與 X_j 資料長度均為 N ，且其聯合機率為： $P(X_i, X_j) = P_{ij}$ ，則測站 X_i 與 X_j 之聯合熵表示如下：

$$H(X_i, X_j) = -\sum_{i,j} P_{ij} \log(P_{i,j}) \dots\dots\dots(3)$$

由於上式可以表示測站 X_i 與 X_j 間全部的不確定性，故可以得知測站 X_i 與 X_j 個別的邊際熵之和應大於或等於其聯合熵，如(4)式；當兩測站資料為統計獨立時，邊際熵的和等於聯合熵。

$$H(X_i, X_j) \leq H(X_i) + H(X_j) \dots\dots\dots(4)$$

在本研究中我們所關心的是當一測站已知時，第二個測站的不確定性被降低了多少？為了解決此問題我們引進條件熵的概念；其定義如右，條件熵 (Conditional Entropy)：當 X_j 發生已知的條件下， X_i 所剩的不確定性。所以條件熵小於等於邊際熵，如(5)式，當兩測站資料為統計獨立時邊際熵將等於條件熵。

$$H(X_i|X_j) \leq H(X_i) \dots\dots\dots(5)$$

$$H(X_i|X_j) = -\sum_{i,j} P_{ij} \log(P_{i,j}) \dots\dots\dots(6)$$

$$H(X_i|X_j) = H(X_i, X_j) - H(X_j) \dots\dots\dots(7)$$

若要瞭解兩測站間共有的資訊量有多少？可以可轉移資訊來解決此問題。

可轉移資訊 (Transferable Information) $T(X_i, X_j)$ 表示可由 X_j 轉移到 X_i 的訊息，可以看

成兩者重覆的訊息，如(8)式。

$$T(X_i, X_j) = H(X_i) - H(X_i|X_j) \dots\dots\dots(8)$$

(四) 雨量站網評估步驟

應用上述之理論，進行雨量站網評估的步驟簡單說明如下：

1. 選定研究區域。
2. 選出欲研究之兩型

台灣之氣候依季節之不同其降雨類型可粗略分為五類，有東北季風型、春雨型、梅雨型、颱風雨型、秋雨型（李富城、林沛練，1981）；另外在夏季又常有短延時的對流雨等。本研究選出降雨變化大之夏季（6月1日--10月30日）資料來作分析。

3. 以群集分析將區域內測站作合理的分群

$$H(X) = -\sum_{n=1}^N P(X_n) \log P(X_n) \dots\dots\dots(9)$$

由(9)式分別求出各個已知雨量站之熵值 $H(X)$ 。再依熵值的近似程度將測站分類。

4. 測站資訊重覆性排序

分別在各群內，將其所屬的測站依資料重覆性多寡來排序，選出重覆性最多的測站為該群代表站（因為它含有各測站的共同資訊較多，故具代表性）。做重覆性排序時，首先選出熵值最大 $MAX \{H(X_1)\}$ （表示其不確定性最大，即資料中含有之資訊量較大）的測站，此為第一個站。

再選擇與第一站重覆資訊最少的測站為第二站（可使新加入的測站能使整個系統增加較多的資訊），即選

$$MIN\{H(X_1) - H(X_1|X_2)\} \dots\dots\dots(10)$$

其中， $H(X_1|X_2)$ 為 X_2 發生已知的條件下， X_1 所剩的不確定性。則 $H(X_1) - H(X_1|X_2)$ 為 X_1, X_2 二者中重覆的資訊。其中 $H(X_1|X_2)$ 愈大則重覆資訊愈小，亦即選擇 $H(X_1|X_2)$ 最大者。

則第 j 個站即如下式：

$$MIN\{H(X_1, \dots, X_{j-1}) - H[(X_1, \dots, X_{j-1})|X_j]\} \dots\dots\dots(11)$$

在運算時可以選擇 $H[(X_1, \dots, X_{j-1})|X_j]$ 最

大者，依此類推將全區內之測站之重覆性，依序排列，如此最後一站即為重覆資訊最多的測站，也就是代表站。

5. 測站重要性排序

於各群內，以代表站為基礎，依序選擇重覆資訊較少的測站依序加入；此排序亦及其重要性之順序，可以作為刪減測站之順序。

選擇與第一站重覆資訊最少的為第二重要站，以使該站加入後有較多新的資訊，即選

$$MIN\{H(X_1) - H(X_1|X_2)\}$$

則第 j 個重要的站即如下式：

$$MIN\{H(X_1, \dots, X_{j-1}) - H[(X_1, \dots, X_{j-1})|X_j]\} \dots\dots\dots(12)$$

如步驟四之方法將全區內之測站的重要性，依序排列。

6. 決定所需測站數目

若以測站數目為橫軸，以 $H[(X_1, \dots, X_{j-1})|X_j]$ 為縱軸，將兩者的關係繪圖之後可以發現，測站數目增加到一定數目以後 $H[(X_1, \dots, X_{j-1})|X_j]$ 將不再增加，則可轉移資訊亦不會改變，即資訊量亦將不再增加，故若以此為資訊量的一個指標，可套配出上述關係圖的模式，找出臨界資訊量與影響站數。若不考慮經濟等因素時，則測站一直增加，直到加入的測站不會提供任何新的資訊時才停止加入。亦可由套配所得的模式估算達欲獲得的資訊量時，所應設置測站數目。

7. 決定設站位置

由步驟 6. 決定所需測站數目之後，若所需站數比目前已有站數少，則依步驟 5. 的重要性排序，選出所需測站，其餘則刪除；若所需站數比目前已有的測站多，則需要再增加測站。

三、研究實例

- (一) 研究地區與資料

本研究針對淡水河流域（包括大漢溪、新店溪及基隆河）分析其降雨之空間分布特性。計收集該流域內 40 個雨量站之小時雨量記錄，加以分析；各雨量站之編號及位置如圖 2 所示。

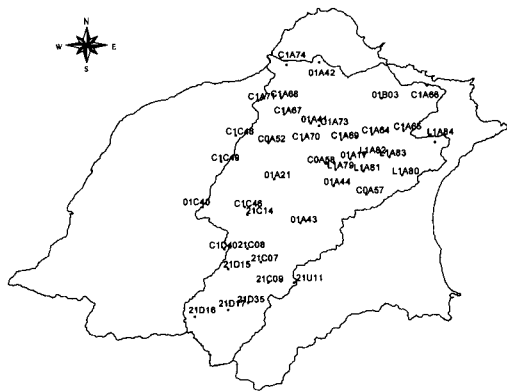


圖 2 淡水河流域雨量測站位置、編號圖

1. 淡水河流域概況

淡水河流域雨量豐沛，全流域年平均降雨量約為 2966 公釐。淡水河位於台灣北部，為台灣的主要河川之一；它是台灣的第三大河川，發源於中央山脈北部末端標高 3529 公尺之品田山，全長 158.7 公里，流域面積 2726 平方公里。流域內南部之大漢溪流域地屬山區地勢較高，而北部屬於台北盆地地勢最低，東南部的新店溪上游地勢居中。

2. 資料選取

由於台灣地區夏季多為對流雨(Convective Storm)，而冬季則多為鋒面雨(Frontal Rainfall)，兩者之降雨強度及空間分布特性均不相同，而且台灣的水資源多集中於夏季，水患也多集中於夏季發生，而本研究欲探討夏季的降雨情況故僅對歷年六月至十月間各雨量站記錄中之降雨事件進行分析。

而資料來源為由中央氣象局取得之玉峰等 40 個雨量站(其所屬單位包括中央氣象局、台灣省水利局、石門水庫管理局、翡翠水庫管理局、台灣電力公司)的時雨量資料，其記錄年限自 1922 年至 1995 年一月；在三年間 6~10 月的雨量記錄中，扣除損壞及遺漏的資料後，40 各測站同時可用的資料為 7296 小時。本研究中將降雨大小分為十級來討論，如表 1 所示。

表 1 雨量資料分級表

分級	時雨量(mm)
0	0
1	0~4
2	4~6
3	6~8
4	8~10
5	10~15
6	15~20
7	20~30
8	30~40
9	大於 40

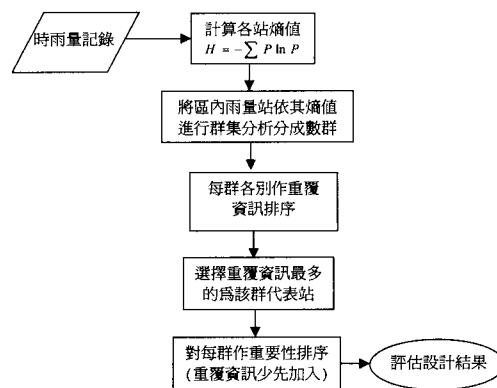


圖 3 雨量站網設計流程圖

(二) 演算流程

整個站網設計的步驟如前所述，而簡單的演算流程示意如圖 3 所示，本研究即以此流程來針對淡水河流域既有之 40 個雨量測站作分析。

四、結果討論

(一) 雨量站分群

依據全區內各個雨量站的矯值大小，進行群集分析後，全區 40 個雨量站可明顯分為三群，且此三群在空間位置上亦有明顯之區分，如圖 4 所示；由左至右分別為第一群有 16 站、第二群有 15 站、第三群有 9 站。其矯值的平均分別為，0.34、0.43、0.53，由此可知第一群的不確定性最低，即其雨量資料變化最少，而第三群的不確定性最高，表示此群之雨量資料變化較大，可能是因為降雨的頻率較高所造成。

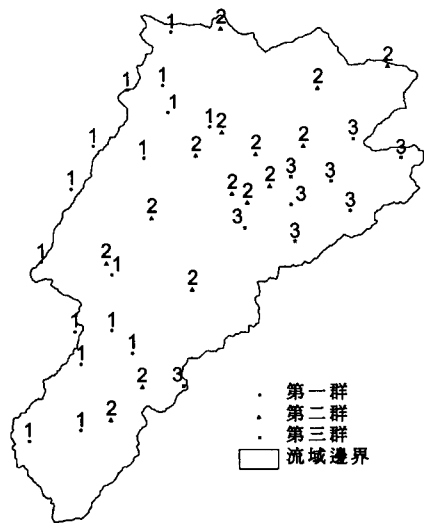


圖 4 分成三群後各群位置圖

表 2 第一群高程、累積雨量、熵值比較表

測站編號	站名	高程(m)	總雨量(mm)	熵值
21D17	鎮西堡	1,630.00	2,698.00	0.376
C0A52	山佳	10.00	2,197.00	0.370
C1A67	新莊	9.00	2,251.00	0.349
C1A74	淡水	40.00	1,863.50	0.323
01A41	中正橋	15.00	2,617.00	0.313
C1A71	林口	238.00	1,878.00	0.317
C1A68	五股	45.00	1,773.00	0.307
C1C48	桃園	105.00	1,901.50	0.320
C1C49	八德	175.00	1,813.00	0.315
01C40	石門	142.00	2,616.00	0.307
C1D40	烏嘴山	839.00	2,455.50	0.338
21C14	霞雲	350.00	3,021.00	0.360
21D15	玉峰	780.00	2,364.00	0.321
21C08	高義	620.00	2,557.00	0.338
21C07	巴陵	1,220.00	2,856.00	0.371
21D16	白石	1,630.00	2,282.00	0.335
累積值		7,848.00	37,143.50	5.361
平均		490.5	2321.5	0.335
標準偏差		571.22	394.15	0.0236
變異係數		1.16	0.17	0.071

由表 2、表 3 及表 4 比較各區的測站高程及本研究所採用的 7296 小時之總雨量與各測站熵

表 3 第二群高程、累積雨量、熵值比較表

測站編號	站名	高程(m)	累積雨量(mm)	熵值
C1A64	石碇	140.00	3,070.00	0.462
01B03	五堵	16.00	3,132.00	0.392
01A42	竹子湖	600.00	3,639.00	0.451
C1A66	瑞芳	101.00	2,933.50	0.436
C1A70	南勢角	197.00	2,700.00	0.395
C1A73	公館	10.00	3,112.00	0.410
C1A69	木柵	260.00	3,294.00	0.456
C0A58	屈尺	90.00	2,825.00	0.400
01A17	坪林	200.00	3,860.00	0.450
L1A79	翡翠	202.00	2,944.00	0.444
01A21	大豹	600.00	3,457.00	0.413
C1C46	大溪	350.00	2,842.50	0.415
21C09	嘎拉賀	1,160.00	3,288.00	0.409
21D35	西丘斯山	2,000.00	3,540.00	0.419
01A43	福山	500.00	4,381.00	0.428
累積值		6,426.00	49,018.00	6.380
平均		428.4	3267.9	0.425
標準偏差		528.52	4450.12	0.023
變異係數		1.23	0.14	0.054

表 4 第三群高程、累積雨量、熵值比較表

測站編號	站名	高程(m)	累積雨量(mm)	熵值
C1A65	火燒寮	380.00	4,397.50	0.563
11A84	太平	403.00	4,268.50	0.559
11A82	十三股	530.00	3,754.50	0.492
11A81	九芎根	378.00	3,636.00	0.496
11A83	坪林	210.00	3,949.00	0.516
11A80	碧湖	386.00	4,232.50	0.526
01A44	大桶山	916.00	4,217.00	0.508
21U11	池瑞	1,070.00	4,156.00	0.525
C0A57	桶後	387.00	3,907.50	0.547
累積值		4,660.00	36,518.50	4.731
平均		517.8	4057.6	0.526
標準偏差		283.85	257.30	0.260
變異係數		0.55	0.06	0.049

值的關係，可以發現地形、累積雨量與熵值之間的關係。就高程來看，第一群的高程平均為 490.5 公尺，而第二群的高程平均為 428.4 公尺、第三群的高程平均為 517.8 公尺，與熵值的大小順序不同，而且三群測站高程的變異係數分別為

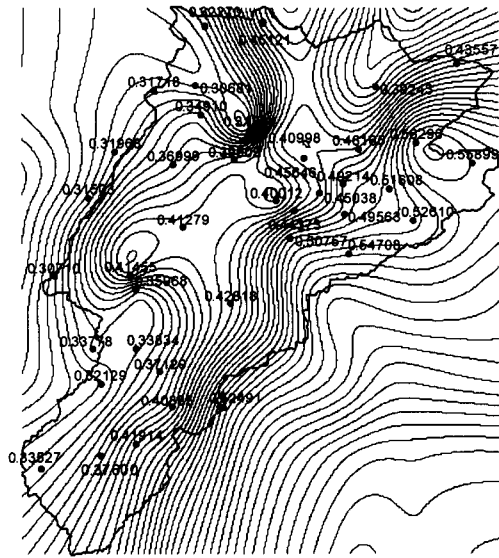


圖 5 淡水河流域等熵值線

1.16、1.23、0.55，故之在同一群內高程的離散程度相當大，即同一群內測站之高程不盡相似；而由地形圖來看第二群及第三群位於山區故降雨機會大於第一群，又第三群靠近東部降雨機會更大於第二群；由此可以發現降雨的分佈雖然會受到地形的影響，但與測站個別的高程並沒有絕對的關係，與所在位置整個地區的整體地形關係較大，也就是說，若測站在山區，則普遍來說比平地的降雨機率高而且強度較大，又例如在迎風面可能高程不需很高就會有豐富的降雨，而在背風面即使有較高的高程，其降雨機率仍然不高。

(二) 測站間資訊相關性分析

淡水河流域實例研究中，三群各別的重覆資訊之排序結果如圖 6，各群中與其他測站重覆資訊最多者，為該群之代表站(重覆資訊最多者即重覆資訊的排序中最後一個的測站)；排序之結果第一群、第二群、第三群的代表站分別為巴陵(21C07)、木柵(C1A69)以及桶後(C0A57)等三站。

求得每群之代表站之後，再選擇重覆資訊最少的測站逐站加入。而所得到的各群中每個測站之重要性排序，即測站加入系統的順序如圖 7 所

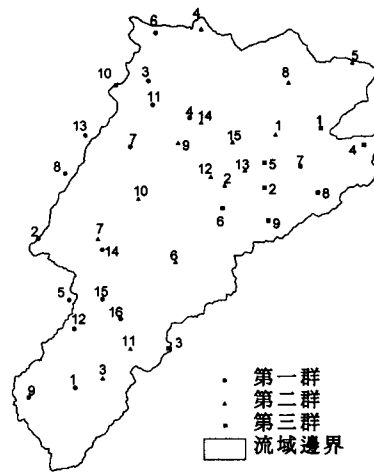


圖 6 淡水河流域三群重覆資訊排序圖

註：各群之順序為獨立

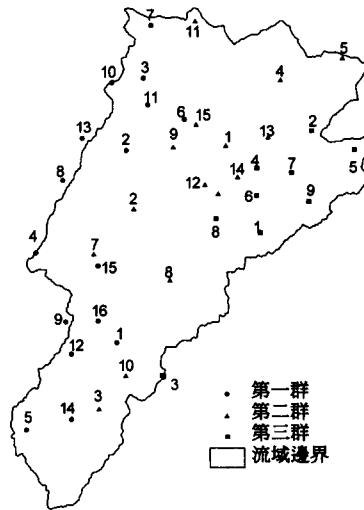


圖 7 淡水河流域三群測站重要性排序圖

註：各群之順序為獨立

示。

(三) 需要測站數目評估

經過重要性的排序之後，我們已知道各測站應加入的順序，但是究竟需要多少測站呢？要解決此問題，故最後需要評估所需要的測站數目。於重要性排序過程中，可以發現持續增加測站時，資訊量的增加幅度愈來愈小，即重覆資訊所

表 5 $H[(X_1, \dots, X_{j-1})|X_j]$ 套配球型模式參數值

球型模式	ω	a
9 站	3.147	13.431
15 站	3.447	18.521
16 站	2.564	17.968

表 6 $H[(X_1, \dots, X_{j-1})|X_j]$ 套配指數模式參數值

指數模式	ω	a
9 站	5.503	14.604
15 站	4.247	12.717
16 站	2.735	9.868

佔的比例愈來愈大，可視為最後的資訊量會趨近於一個定值，但是究竟整個系統的總資訊量是多少我們無法確知；又加入第 j 站時的條件熵可以表示加入第 j 站時，其前一系統整體所剩的不確定性，而此值亦會隨測站增加，漸趨一定值；故以測站數目為橫軸、以 $H[(X_1, \dots, X_{j-1})|X_j]$ 為縱軸則此關係圖可視為測站數目與資訊量的關係圖；再套配適當模式後，可藉以評估該區所需要的測站數目。

本研究在訊息量套配時，嘗試以指數及球型兩種模式（如(13)(14)式）套配：

指數模式 $H(n) = \omega[1 - \exp(-n/a)]$ (13)

球型模式 $H(n) = \omega[\frac{3}{2}(n/a) - \frac{1}{2}(n/a)^3]$, $n < a$;
.....(14a)

$H(n) = \omega, n \geq a$ (14b)

結果如表 5、表 6、圖 8、圖 9 及圖 10 所示。由圖形上可知以指數模式來套配較吻合訊息量的增加趨勢，故選擇指數模式。由套配的結果可知第一區若設置 30 站時資訊量達到飽和，即再增加測站將不會提供任何新資訊，第二區則需要 38 站，第三區需要 44 站；如此全淡水河流域共需 112 個雨量測站即可幾乎掌握所有資訊；而各區所需測站的數目與先前群集分析時所得到的降雨資料變化之大小的推論相當吻合。由於第一區不確定性最低，即降雨的變化性較少故所需的

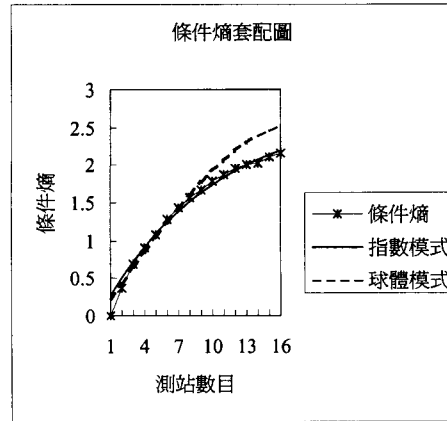


圖 8 第一區 $H[(X_1, \dots, X_{j-1})|X_j]$ 套配圖

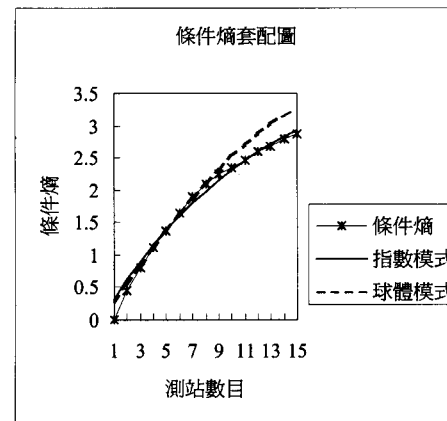


圖 9 第二區 $H[(X_1, \dots, X_{j-1})|X_j]$ 套配圖

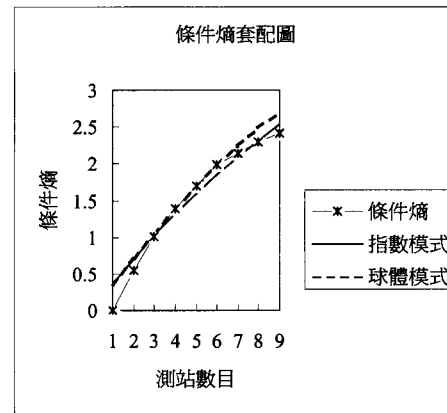


圖 10 第三區 $H[(X_1, \dots, X_{j-1})|X_j]$ 套配圖

測站數目最少；而第三區不確定性最大，即降雨的變化性最大，所以需要最多的測站才足以掌握該區之資訊；又第二區之不確定性居中，故所需的測站數目，亦介於第一區與第三區之間。

又雨量站分佈之代表性根據世界氣象組織(WMO)之規定，在地形複雜的海島上，雨量器的代表面積為 25 平方公里，但在特定目標的研究時，此密度仍嫌太小(曲克恭, 1981)。據此規範淡水河全區總面積共 2726 平方公里，若每個測站代表面積為 25 平方公里，則淡水河流域需要 109 個測站，此與本研究評估所得的結果相當接近。又分析目前三區既有的測站數目，發現目前已設置的測站，在第一區內的 16 個測站，能獲得第一區大約 83% 的資訊量、第二區的 15 個測站能獲得第二區 71.42% 的資訊量，第三區的 9 個測站只可以獲得第三區 46.23% 的資訊量，故若單純針對降雨資訊之取得，要增加測站應該選擇在第三區內優先增加。

本研究並分別計算了獲得 60%、70%、80%、90%、95% 等五種百分比的資訊量時各區所需的測站數目，如表 7 所示。而各區所需要的測站為哪些？若是需要的訊息量要求在目前已有測站可掌握的訊息量之下，則可依照前述的各測站重要性排序，將重要性排序在所需站數以後的測站刪除。若所需訊息量大於目前已有測站可掌握的訊息量，則需要增加測站；而增加測站的位置可選擇資訊變化較快的位置，即等燭圖中坡降較陡處，但詳細方法需要日後進一步研究。若以需要訊息量 70% 為例，由表 7 可知需 70% 訊息量時，第一區需要 11 個測站、第二區需要 14 個測站；再由圖 7 得知，可以分別將重要性在 11(第一區)以下及 14(第二區)以下的測站刪除，即第一區需要巴陵(21C07)~新莊(C1A67)等 11 站、第二區需要木柵(C1A69)~坪林(01A17)等 14 站。而第三區需要 16 站個測，由圖 7 知道該區內目前只有 9 個測站，訊息量僅達 46.23%，故除了原有的桶後(C0A57)~碧湖(L1A80)等 9 個測站以外，另需增加 7 個雨量測站。由此可知淡水河流域之雨量站分佈看似均勻，但是由於地形、氣候等因素的影響，所需的測站分佈不應該是平均分配的方

表 7 需要測站數目評估表

欲獲得資訊量	區域	所需測站數	全流域總站數
60%	第一區	8	31
	第二區	11	
	第三區	12	
70%	第一區	11	41
	第二區	14	
	第三區	16	
80%	第一區	14	53
	第二區	18	
	第三區	21	
90%	第一區	19	73
	第二區	25	
	第三區	28	
95%	第一區	23	87
	第二區	30	
	第三區	34	
100%	第一區	30	112
	第二區	38	
	第三區	44	

式，而需要在第三區內設立較多測站，於第一區可以設立較少的測站，故在一定的經費條件下，針對需求的資訊量來設計雨量站網，使其適切的分佈就相當重要。

(四) 綜合討論

本研究中採用每年 6~10 月的時雨量資料來作分析，因為期間包含的降雨型態仍屬複雜，期間可能有梅雨、對流雨以及颱風暴雨。未來可以僅針對特定某一種降雨來作分析。又本研究中採用的資料為連續的，並非挑選出單一的暴雨事件，加以探討故期間包含了各站皆不降雨的情形，即各站降雨皆為 0 的資料並沒有被剔除，因此發生零雨量的頻率相當大，此可能會降低燭值使不確定性下降，可能會影響評估的結果，若將晴天的記錄扣除則評估結果可能需要更多的測站。

由群集分析分三群後，其在空間上的分佈可以發現同一群內的各測站分佈屬於南北長，東西向狹窄的狹長狀，此即資料結構在空間上的變異

為非等向性的變化在南北向變化較緩慢，在東西向上變化較快速。

評估淡水河流域的雨量站網分佈，結果發現欲獲得相同資訊量時，三群之中，以最東邊的一群（第三群）需要的測站數目最多，偏西的一群（第一群）所需的測站數目最少；此與本島的氣候、降雨狀況頗為吻合，即第三區降雨事件的變化最大，第二區次之，第一區變化最小，此由三群中各站的熵值平均亦可發現相同趨勢。

五、結論與建議

(一) 結論

1. 台灣地區之降雨受地形影響很大，由群集分析與資訊重覆性或重要性的排序過程中，都可以很明顯的看出淡水河流域的降雨記錄有很強的空間分佈特性，而且在空間上之分佈變化屬於非等向性，降雨資料在東西方向上變化較快。
2. 設計雨量站網時，除了需要考慮重覆資訊之外，尚需考慮對於整體系統之代表性問題。兩者要兼顧，在實際設計上實有困難，目前以群集分析分成幾群之後，再於各群中選取與其他測站重覆資訊最多的站為代表站，此後於各群中選擇重覆資訊較少者，逐站加入；較能兼顧代表性以及重覆資訊較少的兩個要求。
3. 由本研究之結果可知，淡水河流域之雨量測站，可以群集分析分為明顯的三群，第一群包括巴陵(21C07)等 16 站，第二群包括木柵(C1A69)等 15 站、第三群包括桶後(C0A57)等 9 站；其中巴陵、木柵、桶後分別為其代表測站。又三群中以第三群（最東邊的一群）之降雨變化最大，第一群(位於最西邊)變化最小，第二群居中。故欲取得相同比例的資訊量時，第三區需要最多測站。又本研究選出所需的雨量站在空間上的分佈應屬合理。故就所需測站的數目與位置分佈和淡水河流域的降雨特性來看，以訊息熵的概念用在雨量站網設計應有相當的可行性。

(二) 建議

1. 若以需要訊息量 70% 為例，第一區需要巴陵(21C07)~新莊(C1A67)等 11 站、第二區需要木柵(C1A69)~坪林(01A17)等 14 站則。而第三區需要 16 站個測，但該區內目前只有桶後(C0A57)~碧湖(L1A80)等 9 個測站，訊息量僅達 46.23%，故應考慮於第三區內增加雨量測站；增加測站的位置大概可以選擇在等熵圖中坡降較陡的區域，至於詳細的位置則需未來針對增加測站的位置做更進一步的研究。
2. 後續研究，可選出單一雨型如梅雨(取每年 5 月中~6 月中之降雨記錄)或颱風雨（選定數場颱風事件）來分析使問題單純化，同時可以解決資料中含大多數 0 可能造成的問題，並比較不同雨型所需測站數，及在不同延時資料時（如 2 小時、6 小時、12 小時、日雨量或月雨量），所需要的測站數目與分佈。
3. 未來此類的研究可以嘗試在不同分區選出不同數目的代表站加以探討(若選兩代表站時即選出重覆資訊最多的兩個測站)。並可以與其他不同的評估資訊量或雨量站網設計的方法做比較。
4. 由於台灣為一海島且地形變化大，故可能在小範圍內就有極相異的降雨分佈，故若站網不夠密集則所測得的降雨記錄有可能有所偏差，又資訊量無法增加不能無中生有，故若經費許可時，能先於研究區內初步設置較密集的均勻網格式站網，用以獲得研究區域內較完整的降雨資訊，再加以探討、評估較能真實的反應該區域之降雨狀態而不致有所偏差。

參考文獻

1. 曲克恭，「台灣地區降雨量分析與預測之討論」，氣象預報與分析，第 88 期，1981。
2. 李富城、林沛練，「石門地區降水特性之分析」，氣象預報與分析，第 86 期，1981。
3. 李天岩，「熵(Entropy)(上)(下)」，數學傳播，第十三卷，第四期，1989。
4. 施孫富，「石門灌區之雨量站設計」，台灣水利，第三十卷，第四期，1982。
5. 黃俊英，多變量分析(Multivariate Analysis An

- Introduction Fourth Edition), 中國經濟企業研究所, 1991。
6. 張學文、馬力, 熵氣象學 (Entropy Meteorology), 氣象出版社, 1992。
 7. 葉惠中, 主成份、群集及判別分析法應用於台灣中部地區降雨空間分佈之研究, 國立台灣大學農業工程研究所碩士論文, 1991。
 8. 劉廣英, 等熵分析, 氣象預報與分析, 第 99 期, 1984。
 9. 鄭克聲、許敏楓、葉惠中, 「雨量站網設計雨評估-區域化變數理論之應用」, 台灣水利, 第 44 卷, 第 1 期, 1996。
 10. 蘇賢錫, 「熵是什麼? 規則性抑或無規性」, 科學教育月刊, 第 43 期, 1981。
 11. Amorocho, J. and B. Espildora, "Entropy in Assessment of Uncertainty in Hydrologic Systems and Models", Water Resources Research, Vol. 9, No. 6, 1973.
 12. Caselton, W. F. and T. Husain, "Hydrologic Networks Information Transmission", Journal of The Water Resources Planning and Management Division, pp.504-519, 1980.
 13. Chapman, T. G., "Entropy as a Measure of Hydrologic Data Uncertainty and Model Performance", J. Hydrol., Vol. 85, pp.111-126, 1986.
 14. Gray, R. M., Entropy and Information Theory, Springer-Verlag New York Inc, 1990.
 15. Husain, T., Ukayli, M. and H. U. Khan, " Meteorological Network Expansion for Saudi Arabia", J. Rech. Atmos., Vol. 16, pp. 281-294, 1982.
 16. Husain, T., Ukayli, M. and H. U. Khan, " Meteorological Network Expansion Using Information Decary Concept", Journal of Atmospheric and Oceanic Technology Vol.3, 1986.
 17. Harmancioglu, N. and V. Yevjevich, " Transfer of Hydrologic Information Among River", Journal of Hydrology, 91, pp.103-118, 1987.
 18. Johnson, R. A. and D. W. Wichern, " Applied Multivariate Statistical Analysis Third Edition", Prentice Hall, 1992.
 19. Krstanovic, P. F. and V. P. Singh, " Evaluation of rainfall network using entropy I", Theoretical development, Water Resour. Management, 6, pp.279-293, 1992a.
 20. Krstanovic, P. F. and V. P. Singh, " Evaluation of rainfall network using entropy II", Theoretical development, Water Resour. Management, 6, pp.295-314, 1992b.
 21. Vijay P. Singh, Elementary Hydrology, Prentice-Hall, Inc., 1992.

收稿日期：民國 89 年 11 月 4 日

接受日期：民國 89 年 11 月 24 日