

# 系集定量降水預報群集分析技術之發展與評估

吳明璋<sup>1</sup> 王潔如<sup>1</sup> 徐理寰<sup>1</sup> 張龍耀<sup>1</sup> 陳熾竹<sup>1</sup> 蕭玲鳳<sup>1</sup> 洪景山<sup>2</sup> 李清勝<sup>1</sup>

台灣颱風洪水研究中心<sup>1</sup> 中央氣象局氣象資訊中心<sup>2</sup>

## 摘要

隨著數值預報和高效能計算機技術的發展，系集預報不僅成為數值預報技術的一個重要發展趨勢，而且日趨成熟。系集預報的應用也越來越受到重視，顯示出它在日常天氣預報與短期氣候預報方面不可取代的作用與優勢。一個系集預報系統所能提供的產品信息量非常的大，是單一數值預報的幾十倍，要在如此龐大的資訊當中獲取有用的訊息，同樣是數值預報技術改進的重要關鍵。目前系集定量降水預報之分析以系集平均之降水預報分布為主要應用，若透過對降雨型態特性之分析，可更深入瞭解系集預報系統特性，幫助更有效、合理地使用系集預報產品。

本研究針對現有群集分析技術之特性與實用性進行評估，依據評析結果初步完成以自組織映射圖為基礎之系集定量降水預報群集分析技術建置。並以颱風中心定量降水系集實驗(TAPEX)資料進行之敏感性測試，結果顯示已建置之技術確可正確分辨不同特徵之系集定量降水預報結果，未來研究方向可考慮群集結果之加值應用。

關鍵字：群集分析(Cluster Analysis)、系集預報

## 一、前言

系集預報技術之發展日趨成熟，系集預報系統產出之預報產品資料量急速成長，透過電腦計算與儲存設備技術之進步，我們可獲得比過去龐大的氣象預報資料庫，如何從這些資料中提煉出有價值的資訊，才是能獲得預報技術改善及科學進步的重要關鍵。系集預報系統的發展是為了處理單一數值天氣預報中預報不確定性的問題，不確定性來源可包含初始資料誤差，模式參數化方法的假設，以及關注的天氣系統本身預報度等等。透過多個不同系集成員的設計，系集預報系統希望能將不確定性量化，改善單一決定性預報的不足，而以機率預報的概念呈現預報結果。目前國內外各大數值天氣預報中心皆以系集預報的方式進行作業，並致力於改善系集預報的表現。

在系集預報中，若假設降水頻率分布為高斯分布，系集平均則通常代表發生頻率最高的事件，也是系集預報中常使用的分析方法，但雨量的分布通常並不連續，且隨著地形有獨特

對流尺度的降雨特性，因此系集平均雖然可以大致上反應出正確的降雨空間分布，但對於極端降雨之區域往往會因為系集平均後低估。為了改善系集平均的缺點，Ebert (2001)以機率擬合方法(Probability-Matched Mean, PM)，保持系集平均的空間分布，重新分配降雨頻率的分布，使得極端降雨極值能夠維持。過去將 PM 方法應用在台灣之系集預報系統中，也證實在台灣這種地形複雜的環境下也能有效提高被系集平均低估的降雨極值，雖然有時會在極值處有過度預報的情形發生，但透過不同降雨頻率分布的調整或挑選，可有效降低 PM 過度預報的情況(Fang and Kuo, 2013; 李與洪, 2014; 葉世瑄, 2014; 江等, 2015; 蘇等, 2015)。雖然上述 PM 方法可改善系集平均低估的極端值預報，但當系集平均及 PM 方法之降雨空間分布發生誤差時，降雨預報之結果即不適用，因此，分析系集系統中各系集成員在每次預報中可能造成的不同降雨型態亦為一重要課題。

我們發現在台灣的定量降水系集預報中，若能根據不同降雨型態特性，對系集預報結果

進行客觀分析，透過適當挑選有相似特性的系集成員，可得到更為合理的預報結果。若再配合歸納不同特性之系集預報結果的機率分布，還可進一步提供不同情境下風險管理的有效資訊。因此，若能發展與建置有效且適用之群集分析技術並應用到系集定量降水預報中，則可透過客觀的統計分類方法，能更深入了解系集預報系統特性，幫助我們更有效、合理地使用系集預報產品。

## 二、 研究方法

現有的群集分析技術，一般而言可概分成階層式(hierarchical)群集分析法與非階層式(nonhierarchical)群集分析法兩大類型(Johnson and Wichern, 2002)。使用階層式的群集方法，最終都將以階層形式的樹狀圖(dendrogram)呈現群集分析結果。一開始將每一筆資料均視為一個群集，接著找出所有群集間最相似的兩個群集進行合併而成為一個新的群集。重複上述步驟，直到最後所有資料均歸為同一群集中。整體來說，階層式群集分析法的優點為概念簡單，只需要資料間的距離就可以建構分群結果，而且最後可用樹狀結構圖來表現整個群集分析的聚類過程。而缺點即為此種分析技術通常較適用於分析少量資料，而不易處理大量資料並展示分析結果。

在非階層式群集方法中，於文獻中最常見的代表即為  $K$  均值法 ( $K$ -means method) (MacQueen, 1967)。使用這種類型的群集分析方法，必須預先訂定群集數目與相對應的初始群集中心位置，再藉著反覆疊代運算與調整，而得到群集分析的最後結果。 $K$  均值法具有概念簡單以及可以快速獲得分析結果的優點，然而實際使用上可能面臨無法處理群集重疊、群集分界為曲線和折線的問題，以及遭遇到離群值時容易導致群集中心偏移、難以預先決定群集數量與群集中心等缺點。

此外，近年來亦有許多以機器學習

(machine learning)理論為基礎所發展之新式的群集分析方法，例如自組織映射圖(self-organizing map)法。自組織映射圖係於 1982 年由 Kohonen 提出之一種特殊類型之人工類神經網路(artificial neural network)技術(Kohonen, 2001)，許多研究均指出自組織映射圖具有可提供較傳統資料特徵萃取與樣式辨識技術更為客觀的分析結果(Mangiameli et al., 1996; Michaelides et al., 2001)。透過自我組織的學習過程，自組織映射圖內含的神經元權重(neuron weights)終將可描述整體輸入資料在拓樸(topology)空間中分佈情況，並以可視化的方式展現全體資料間的拓樸分佈。因此，依據這些資料間的拓樸分佈關係，即可對資料進行客觀的聚集或分類分析。

由於此研究之主要研究目標為針對現有系集預報系統中系集定量降水資訊研發群集分析技術，因此較適合本研究目標的群集分析技術，必須擁有可處理大量且高維度資料特性的能力，亦即應為非階層式及機器學習式群集分析技術。而且，考慮主要影響台灣降雨之颱風或梅雨系統等系統裡，降雨分布會受到颱風路徑、鋒面位置不同，而有複雜多變的降雨分布與特性。因此，依據評析結果，我們認為利用機器學習式的群集分析技術，應會比需要預先決定群集數量與群集中心的非階層式方法更加適合實際應用需求。

## 三、 資料來源

台灣颱風洪水研究中心自 2010 年起於颱風影響台灣期間進行「台灣定量降雨系集預報實驗」(Taiwan Cooperative Precipitation Ensemble Forecast Experiment, TAPEX)，其目的在研發測試定量降雨預報技術，並提供災防單位應變作業時決策參考之資訊，協助提升災害防救效能。本實驗結合學研界(台灣大學、中央大學、台灣師範大學及文化大學)與作業單位(中央氣象局)之研發能量，並結合國家

高速網路與計算中心的計算資源及國家災害防救科技中心的實務經驗。

目前該實驗中已超過 20 組系集成員，各成員為考慮初始場、積雲參數法、雲微物理參數法、邊界層參數化等差異及採用不同模式而設計。透過不同物理參數法之組合，針對歷史個案進行定量降水預報校驗，找出適合台灣地區及華南地區降水模擬的最佳物理組合。系集預報所針對的區域範圍約為東經 80 至 160 度、北緯 0 至 40 度，三層巢狀網格之水平解析度分別為 45、15、5 公里，台灣本島及周遭區域範圍均採用 5 公里之最高解析度，每天皆可產出 4 組系集定量降水預報資訊，預報長度達到 72 小時的雨量資料。

#### 四、 群集分析技術測試結果

研究個案選取 2015 年台灣地區所發生的降雨事件。嘗試以不同發生時間之不同類型降水事件下之 TAPEX 系集定量降水預報資訊，評析已建置完成的系集定量降水資訊群集分析技術是否具有可正確將不同降雨事件資料分辨(分類)的能力。本研究採用之降雨事件如圖 1 所示，分別為 2015 年 4 月 20 日之冷鋒降雨(圖 1a)、2015 年 7 月 18 日之西南氣流降雨(圖 1b)、以及 2015 年 9 月 27 日之颱風降雨(圖 1c)。此 3 個降雨事件均為明顯不同的成因，因此導致之降雨分布特性在不同降雨事件彼此間亦為不同。

此處同時以 3 個降雨事件進行測試。此測試之明確群集數目為 3，因此 SOM Kohonen 層則設定為 1x3。測試結果顯示，自組織映射圖群集分析技術亦可正確將 3 場降雨事件之系集定量降水預報資訊聚集成 3 類，但正確率為 97%(圖 2)，即 66 筆測試資料中，有 2 筆資料未能被正確歸至應屬的類別(此處意指降雨事件)中。

進一步，對於此一測試中分類未能獲取 100% 正確分辨率之原因，此處將進行自組織

映射圖群集分析技術之參數敏感度試驗。於該群集分析技術中，主要共有三個參數需由使用者事先訂定，分別為 Kohonen 層之拓樸維度(n)、初始學習速率( $\eta(0)$ )、以及拓樸鄰域參數中之鄰近半徑( $\alpha$ )。由於此測試中已明確群集數目為 3，因此此處將不變動 Kohonen 層拓樸維度 n 值(n=3)。至於剩下之初始學習速率與鄰近半徑兩參數，則分別於 0.5 至 1.0 之範圍間，以 0.1 的間隔進行變動測試，因此總共將有 36 組試驗結果。在 36 組敏感度試驗之正確率範圍為 97%~100%間(表 1)，亦即表示在 66 筆資料中僅有 1~2 筆資料未能被正確歸類至其應屬的群集(此處意指降雨事件)中。然而，這樣的結果對於群集分析的後續加值應用上(例如計算各群集間的系集平均)將造成何種程度的影響，需進行更詳細之分析與探討。

若以單一事件進行群集分析技術測試，選取 2012 年之蘇拉(Saola)颱風，颱風中心系集實驗之系集定量降水預估結果如圖 3(a)所示，為以 2012-07-31-18Z 的初始資料場進行數值天氣實驗所產出之 24 小時累積降雨(2012-08-01-00Z~2012-08-02-00Z)系集預估資訊。圖中可約略發現，系集實驗中各成員產出之 24 小時累積降雨預估結果似乎有部分成員(例如 M05、M06 與 M08)的預估結果十分相似，也有些成員的預估結果(例如 M02)與其他成員的預估結果十分相異。雖說如此的預估情況係反應大氣系統的混沌與不確定性，但卻往往造成後續使用者資訊判讀或使用上的困惑。若進一步利用自組織映射圖可建置多維度複雜資料間拓樸分佈關係並萃取資料特性之能力，即可對 3(a)中之定量降水系集資訊建置群集分析技術並進行分析，4 個群集及其歸屬成員之初步分析結果如下圖 3(b)所示。可發現，透過群集分析後確可找出颱風中心系集實驗之定量降水系集預估資訊中隱含的資訊，22 組系集成員中有 5 個系集成員(M10、M12、M13、M14、M19)預估全台灣均將有較大的降雨潛勢，

有 4 個系集成員(M04、M09、M17、M23)預估台灣的中北部區域將有較大的降雨潛勢，有 8 個系集成員(M05、M06、M08、M11、M15、M16、M18、M20)預估台灣的北部與中部山區將有較大的降雨潛勢，而有 5 個系集成員(M01、M02、M03、M07、M21)則預估僅台灣北部區域將有較大的降雨潛勢。因此，透過群集分析，確可幫助後續使用者快速獲取系集定量降水預報資訊間的相似與相異性，並得知不同降雨預估情境下的系集成員個數。這樣的資訊將可增進使用者判識系集定量降水預報特性之掌握度，進一步亦有助於依據群集分析成果而進行後續系集機率預報研究。

## 五、 結論

使用 TAPEX 之定量降水系集資料進行該群集分析技術之實際案例測試與參數敏感度分析。測試結果顯示，對於具有明顯差異之不同降雨事件下的系集定量降水預報資料，使用本案已初步建置之群集分析技術確可正確分辨不同特徵之系集定量降水預報結果而明確地將這些不同降雨事件下的系集定量降水預報資料聚類至不同類別。另外，2012 年蘇拉颱風降雨事件之實際案例測試結果可知，透過該群集分析確可幫助後續使用者快速獲取系集定量降水預報資訊間的相似與相異性，並得知不同降雨預估情境下的系集成員個數。這樣的資訊將可增進使用者判識系集定量降水預報特性之掌握度，亦將有助於進行後續系集機率預報等加值應用研究。

本研究目前針對系集預報系統，分類結果可透過計算各類別中成員數目來計算不同降雨型態之機率，提供未來可能進行的後續加值應用，例如：透過分類大量歷史天氣資料，建立特殊天氣狀況下對應的降雨型態資料庫，當未來雨量預報屬於其中一種特殊天氣狀況時，可對未來降雨進行機率推論。

## 六、 參考文獻

- 李志昕、洪景山，2014：區域系集定量降水預報之應用與分析研究。103年天氣分析與預報研討會，A2-19。
- 葉世瑄，2014：系集定量降水預報方法之研究。國立中央大學，大氣物理研究所，碩士論文，共 82 頁。
- 蘇奕叡、李志昕、洪景山，2015：系集機率擬合定量降水預報產品之特性分析。104 年天氣分析與預報研討會，A2-66。
- 江宙君、陳熾竹、徐理寰、吳明璋、黃麗蓉、林忠義，2015：定量降雨系集預報加值雨量測試分析。104年天氣分析與預報研討會，A2-65。
- Ebert, E. E., 2001：Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480.
- Fang, X., and Y.-H. Kuo, 2013: Improving ensemble-based quantitative precipitation forecasts for topography-enhanced typhoon heavy rainfall over Taiwan with a modified probability-matching technique. *Mon. Wea. Rev.*, **141**, 3908–3932.
- Johnson, R.A., and D.W. Wichern, 2002: Applied Multivariate Statistical Analysis, 5th edn. Prentice Hall: Upper Saddle River.
- MacQueen, J.B., 1967: Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, Berkeley, CA: University of California Press, 281–297.
- Kohonen, T., 2001: Self-Organizing Maps, 3rd edn. Springer-Verlag: Berlin.
- Mangiameli, P., S.K., Chen, and D. West, 1996: A comparison of SOM neural network and hierarchical clustering methods. *Eur. J. Oper. Res.*, **93**, 402–417.

Kleovoulou, 2001: Classification of rainfall variability by using artificial neural networks.

表 1、參數敏感度測試之分類結果正確率彙整表。

	$\eta(0) = 0.5$	$\eta(0) = 0.6$	$\eta(0) = 0.7$	$\eta(0) = 0.8$	$\eta(0) = 0.9$	$\eta(0) = 1.0$
$\alpha = 0.5$	98.5	100	100	97	97	97
$\alpha = 0.6$	98.5	100	100	97	97	97
$\alpha = 0.7$	98.5	100	100	97	97	97
$\alpha = 0.8$	98.5	100	100	97	97	97
$\alpha = 1.0$	98.5	100	100	97	97	97
$\alpha = 1.0$	98.5	100	100	97	97	97

(a) 事件1 (22個成員)

(b) 事件2 (21個成員)

(c) 事件3 (23個成員)

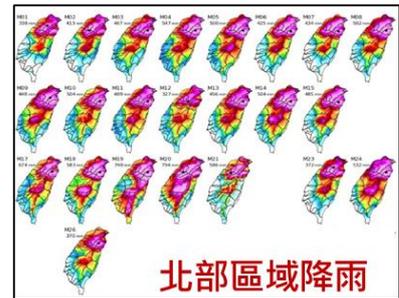
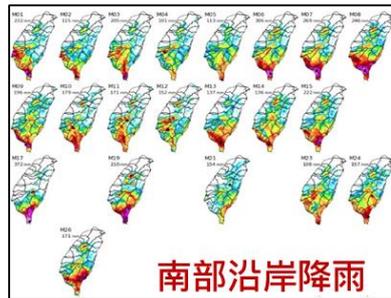
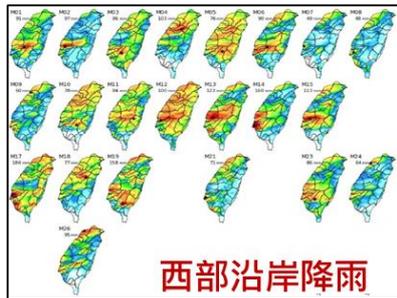
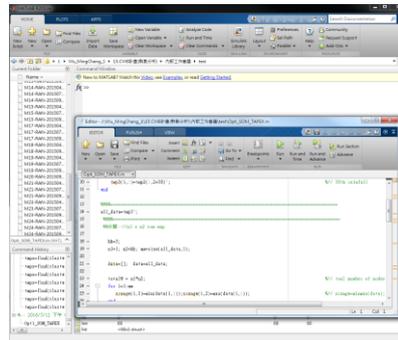


圖 1 觀測與台灣定量降雨系集預報實驗各成員 24 小時累積雨量圖(a) 2015 年 4 月 20 日之冷鋒降雨(b) 2015 年 7 月 18 日之西南氣流降雨(c) 2015 年 9 月 27 日之颱風降雨。

1. 將所有系集降雨檔案  
放置於同一資料節中  
(21+23+22=66個檔案)

- M01-RAN-2015042012-grid
- M01-RAN-2015071800-grid
- M01-RAN-2015092712-grid
- M02-RAN-2015042012-grid
- M02-RAN-2015071800-grid
- M02-RAN-2015092712-grid
- M03-RAN-2015042012-grid
- M03-RAN-2015071800-grid
- M03-RAN-2015092712-grid
- M04-RAN-2015042012-grid
- M04-RAN-2015071800-grid
- M04-RAN-2015092712-grid
- M05-RAN-2015042012-grid
- M05-RAN-2015071800-grid
- M05-RAN-2015092712-grid
- M06-RAN-2015042012-grid
- M06-RAN-2015071800-grid
- M06-RAN-2015092712-grid
- M07-RAN-2015042012-grid
- M07-RAN-2015071800-grid
- M07-RAN-2015092712-grid

2. 執行分類程式  
(約28秒)



3. 輸出分類結果

第1群:23  
第2群:24  
第3群:19

分辨率  
97%

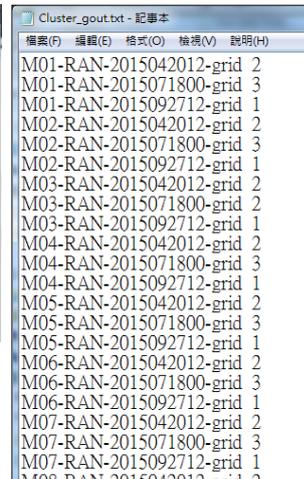


圖 2 冷鋒降雨、西南氣流降雨與颱風降雨之測試結果 (以  $\eta(0) = 0.8$  ,  $\alpha = 0.5$  為例)

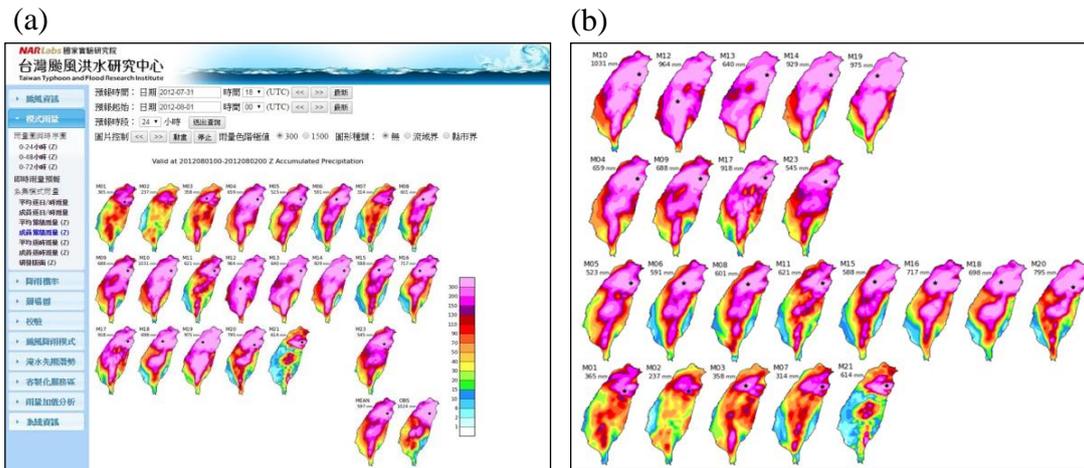


圖 3 2012 年蘇拉颱風事件之(a)系集定量降水預估資訊與(b)群集分析之分類結果。