

系集定量降水群集分析技術之測試與應用

王潔如¹ 吳明璋¹ 徐理寰¹ 蕭玲鳳¹ 洪景山² 李清勝¹

台灣颱風洪水研究中心¹ 中央氣象局氣象資訊中心²

摘要

日趨成熟的系集預報技術已成為近年數值天氣預報領域的重要發展趨勢，系集預報產品實際應用的關鍵價值也日益提升。如何於大量系集預報產品當中獲取有用的訊息，並提供作業單位於預報應變時參考運用，係為未來增進系集預報實務效益的重要關鍵。

本研究將針對已開發之系集定量降水群集分析技術，透過2012年至2017年20場劇烈降雨個案資料進行敏感度測試與系統性分析。結果顯示，群集分析技術可正確分辨不同特徵之系集定量降水預報，明確且合理地將不同降雨事件的系集定量降水預報資料分至不同類別。此外，使用群集分析加值後某一群集之系集平均，相較於未分類前之所有系集成員平均，有八成機率得到較佳之定量降雨預報結果。

關鍵字：群集分析(Cluster Analysis)、系集預報(Ensemble Forecasting)

一、前言

有鑑於系集預報系統產出之預報產品資料量急速成長，隨著電腦計算與儲存設備技術之進步，已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。

在應用系集預報產品時，若假設系集預報之降水頻率為高斯分布，則可以系集平均值代表發生頻率最高的事件，是系集預報中最常使用的分析方法。但從實際觀測降雨中，常發現雨量在時間、空間上的分布多呈現不連續的特徵，且隨著地形有獨特對流尺度的降雨特性。因此使用系集平均雖可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而低估極端量值。為了改善系集平均低估降雨極端值的缺點，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

偏差時，上述之降雨預報結果即不適用。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，本研究應用機器學習(machine learning)理論中之自組織映射圖(Self Organizing Map, SOM)技術為基礎，開發系集定量降水預報群集分析技術(吳等, 2016)。透過敏感度試驗與多組實際案例測試，結果顯示該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，並將具有不同降雨特徵的系集定量降水預報資料聚類至不同類別。有助於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辨識系集定量降水預報特性之掌握度，亦有助於能更深入地了解系集預報系統特性，進而更有效、合理地使用系集預報產品。

二、研究方法

自組織映射圖係由Kohonen於1982年提出之一種特殊類型之人工類神經網路(artificial neural network)技術(Kohonen, 2001)。

許多研究均已指出，自組織映射圖具有資料特徵萃取與樣式辨識技術的能力，且可提供比傳統分析技術更為客觀的分析結果(Mangiameli et al., 1996; Michaelides et al., 2001)。近年來，自組織映射圖亦已應用至數值天氣模式之預報資料的分析研究上，並已獲得十分良好的應用成效(Borah et al., 2013; Yin et al., 2011)。

自組織映射圖透過反覆自我組織(self-organizing)的迭代學習過程，使內含的神經元權重(neuron weights)不斷地依據資料的特徵而調整，最終將可描述整體資料在拓樸(topology)空間中分佈情況，並以可視化(visualization)的方式展現全體資料間的拓樸分佈關係。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。上述之自組織映射圖的學習過程，可如圖1所示。圖中藍色區塊為資料的實際分布情況，方形網格則表示自組織映射圖的神經元權重，每個格子均代表一個神經元權重。整體而言，自組織映射圖的學習策略中共包含三個重要的學習階段，依序分別為競爭、合作、調適三個階段。首先，在競爭過程中，可針對某一輸入資料(圖中白點)，找出與該輸入資料最為近似的神經元，並視之為優勝神經元(圖中黃點)。接著，在合作過程中，則根據前述過程中找出之優勝神經元，定義其鄰近拓樸區域中神經元間相互的鄰近關係。最後，在調適過程中，則是依據合作過程中決定之拓樸鄰域，調整整個拓樸鄰域中所有神經元的權重值，使優勝神經元權重與拓樸鄰域神經元權重均將朝該輸入資料的方向接近。

因此，由圖1即可瞭解自組織映射圖的神經元權重經過不斷自我組織之學習過程。學習過程中，神經元權重將依序回應輸入資料特徵而逐次調整。經不斷反覆迭代後，最後即可描述全體資料間的拓樸分佈關係的能力(方形網格疊合資料的真實分布情況)。

三、 資料來源

本研究之系集資料來自台灣颱風洪水研究中心「台灣定量降雨系集預報實驗」(Taiwan Cooperative Precipitation Ensemble Forecast Experiment, TAPEX)，該系集實驗係起始於2010年，結合國內學研界與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。現階段，該系集實驗中使用4種數

值天氣預報模式，配合不同邊界與初始條件、不同參數化技術、以及不同執行策略，共設計有26組系集成員，分別為20組WRF模式、2組MM5模式、2組CRSS模式，以及2組HWRP模式成員。這些系集成員每天均執行4次實驗，每次實驗均進行未來3天之數值模擬與雨量預估。

四、 群集分析技術分析結果

案例測試以蘇拉颱風7月31日18Z的初始場所產出之24小時累積降雨(2012-08-01 00Z~2012-08-02 00Z)系集資訊。多數成員預估之主要降雨區域集中在北部，東北部及中部山區，雨量極大值出現在東北部地區。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的預估結果十分相似，也有些成員的預估結果與其他成員的預估結果十分相異。雖說如此多變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。透過群集分析後(結果如圖2所示)，則可找出定量降水系集預估資訊中隱含的群集資訊。顯示有5個系集成員預估全台灣均將有較大的降雨潛勢，有4個系集成員預估台灣的中北部區域將有較大的降雨潛勢，有8個系集成員預估台灣的北部與中部山區將有較大的降雨潛勢，有5個系集成員預估僅台灣北部區域將有較大的降雨潛勢。

更進一步，我們亦針對蘇拉颱風的系集資料，探討使用不同拓樸維度設定下的群集分析結果。前述之分析結果係為使用 1×4 之一維Kohonen拓樸層的設定進行分析。使用此設定表示我們係藉由使用4個神經元權重來描述蘇拉颱風22筆系集定量降水預估資料的分布情況，最後可獲取(至多)4類的群集分析結果。後續則嘗試使用 5×5 之二維Kohonen拓樸層的設定進行分析，亦即表示我們將使用25個神經元權重來描述22筆資料的分布情況(結果如圖3所示)。

在圖3約略可看出22筆系集資料的群聚情況，這些情況也反應資料間彼此的相異與相同程度。在特徵圖上越相近的兩筆資料表示兩者間有較高的相似程度(例如M12與M19)，反之則表示兩者的相異程度越高(例如M19與M21)。對比圖2與圖3的結果可以發現，使用不同維度大小的SOM所獲取的分析結果實際上仍為一致，圖3中由右上至左下的區塊，分別對應圖2中由上至下的1~4類分類結果。

因此可知，使用不同維度大小的SOM均可獲取正確分析結果，不同的是使用較大維度的SOM可以得到更多細節的分類分析結果。然而，實際作業上則可能因為較多的細節(例如子群集的存在)，導致群集數目的決定上有較多的不確定性。因此，對於拓樸維度大小的設定，建議仍應依據實際需求而決定。若需求為快速獲取分類分析結果，則建議使用一維的拓樸結構(1 × n)；但若需求為欲獲取更為細節的分類分析結果，則建議可使用二維的拓樸結構(n × n或n × m)。

此處亦進一步將真實觀測降雨資料輸入到此學習完成的SOM網路中，並且標記該筆觀測資料的優勝神經元位置。結果顯示，該筆觀測資料被映射至(3,5)神經元，值得一提的是該神經元亦為M10與M14之優勝神經元。如圖4所示，真實觀測降雨、(3,5)神經元、及M10與M14預估降雨間有高度的相似，均於台灣東北部之宜蘭區域有較為極端的降雨。

因此，對於上述群集分析結果在實際預報作業上的加值應用，可如圖5所示。由圖中4個群集各別之系集降雨平均與系集路徑平均結果可發現，群集IV之系集降雨平均與真實觀測降雨有極高的相似程度，對於極端降雨之區域與量值的預估表現，明顯優於未分類前的系集降雨平均(觀測：1024mm；群集IV之系集平均：861mm；未分類之系集平均：589mm)。對於路徑的預估上也可發現相同的情況，群集IV之系集路徑平均(黑色線)與真實颱風路徑(紅色線)有極高的相似程度，預估表現亦優於未分類前的系集路徑平均。故可知，採用群集分析加值後之各群集系集平均，將有機會可獲取預估表現更優之降雨預報結果。

最後，透過2012年~2017年共20場實際降雨案例，廣泛評析現有技術對於24小時累積之系集定量降水預報資料的特徵萃取能力。這20場降雨事件共有11場颱風事件、6場梅雨降雨事件、2場西南氣流降雨事件與1場冷鋒過境降雨事件(表1)。群集分析結果顯示，M06、M08與M15有較高的比率被歸至同一群集中，M13、M14與M15亦有較高的比率被歸至同一群集中。後續亦可透過這些大量的系統系分類結果，進行系統的改進研究，以期可使系集系統產出更具效能的系集資料。

另外，20場降雨事件系集定量降水預報資訊之群集校驗結果如圖6所示。使用某一群集之系集平均(黑線)，相較於使用未分類前的整體系集平均(紅點)，有8成5的機率得到較佳的相關係數CORR(如圖6(a)所示)，並且有8成

的機率得到較小的均方根誤差RMSE(如圖6(b)所示)，亦有8成的機會找到某一群集之系集平均在豪雨門檻(200mm/24hr)有較高的預兆得分TS(如圖7所示)，亦即，使用某一群集之系集平均，相較於使用未分類前的整體系集平均，有機會可獲取預估表現更優良之定量降水預報結果的潛力。

五、 結論

本研究以2012年蘇拉颱風之實際案例，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的測試，評析每次系集預報之群集分析結果作為下一次預報參考依據的助益。結果顯示，使用不同設定下之自組織映射圖所獲取之群集分析結果，均有助於預報參考依據。亦即，透過群集分析技術對資料特徵的萃取能力，可合理地篩選部分系集成員，進而將有機會可產出表現更佳之定量降水預報結果。

由20場實際降雨案例的校驗結果亦可發現，若使用某一群集之系集平均，相較於使用未分類前的整體系集平均，有機會可獲取預估表現更優良(較高的相關係數、較低的均方根誤差、以及較高的預兆得分)之定量降水預報結果的潛力。

然而，雖已知若使用某一群集之系集平均，相較於使用未分類前的整體系集平均，有機會可獲取預估表現更優良之定量降水預報結果。但該如何事先得知該使用哪一個群集之系集平均才能獲取較佳之定量降水預報結果，則仍需做進一步的探討。

六、 參考文獻

1. 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。
2. 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。
3. 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。
4. 江宙君、陳嫻竹、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。
5. 吳明璋、王潔如、徐理寰、張龍耀、陳嫻竹、蕭玲鳳、洪景山、李清勝, 2016, 系集定量降水預報群集分析技術之發展與評估, 105年天氣分析與預報研討會, A2-11。
 6. 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.
 7. 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.
 8. 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.
 9. Kohonen, T., 2001: Self-Organizing Maps, 3rd edn. Springer-Verlag: Berlin.
 10. 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.
 11. Michaelides, S.C., C.S., Pattichis, and G. Kleovoulou, 2001: Classification of rainfall variability by using artificial neural networks. *Int. J. Climatol.*, **21**, 1401–1414.
 12. Borah N., A.K. Sahai, R. Chattopadhyay, S. Joseph, S. Abhilash, B. N. Goswami, 2013: A self-organizing map-based ensemble forecast system for extended range prediction of active/break cycles of Indian summer monsoon. *Journal of Geophysical Research: Atmospheres*, **118(16)**, 9022–9034.
 13. Yin, C.H., Y.P. Li, W. Ye, J. F. Bornman, X.D. Yan, 2011: Statistical downscaling of regional daily precipitation over southeast Australia based on self-organizing maps, *Theoretical and Applied Climatology*, **105(1-2)**, 11–26.

表1、20場降雨事件資料表

降雨事件	發生時間 (UTC)	降雨成因	降雨事件	發生時間 (UTC)	降雨成因
1	2012年07月31日 18Z	蘇拉颱風	11	2015年04月20日 12Z	冷鋒過境
2	2013年05月18日 12Z	梅雨	12	2015年07月18日 00Z	西南氣流過境
3	2013年07月12日 12Z	蘇力颱風	13	2015年08月07日 06Z	蘇迪勒颱風
4	2013年08月28日 12Z	康芮颱風	14	2015年09月27日 12Z	杜鵑颱風
5	2013年10月06日 12Z	菲特颱風	15	2016年07月07日 12Z	尼伯特颱風
6	2014年05月05日 12Z	梅雨(1)	16	2016年09月13日 12Z	莫蘭蒂颱風
7	2014年05月20日 12Z	梅雨(2)	17	2016年09月26日 12Z	梅姬颱風
8	2014年07月22日 12Z	麥德姆颱風	18	2017年06月01日 00Z	梅雨(1)
9	2014年08月11日 12Z	西南氣流過境	19	2017年06月03日 06Z	梅雨(2)
10	2014年09月20日 18Z	鳳凰颱風	20	2017年06月16日 00Z	梅雨(3)

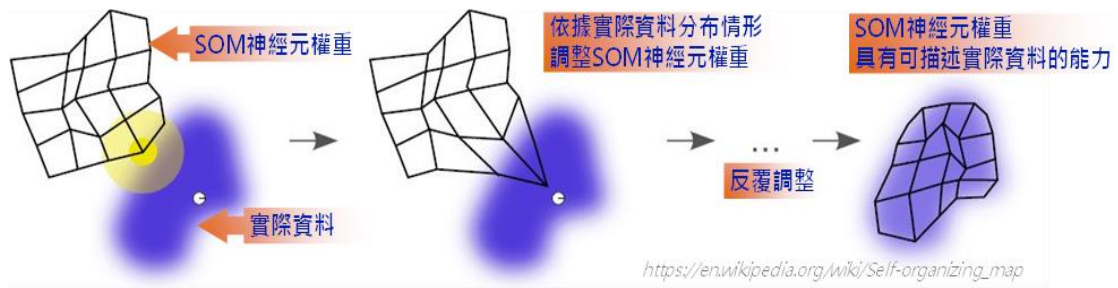


圖1、自組織映射圖法之分類過程示意圖(來源：Wikipedia及後製)。

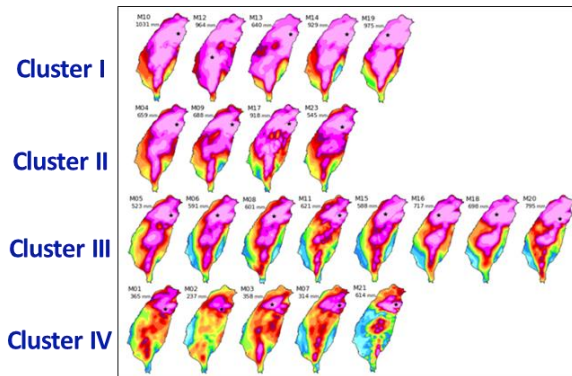


圖2、蘇拉颱風事件的群集分析結果

4,9		13		12,19
23				
5				10,14
	11			17
1,2,3,7,21		6,8,15	16	18,20

圖3、蘇拉颱風事件系集資料的二維拓樸群集分析結果

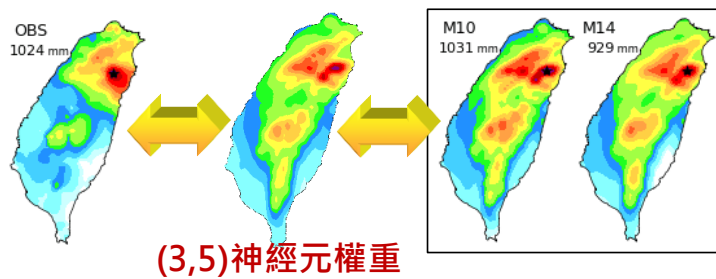


圖4、蘇拉颱風真實觀測降雨、(3,5)神經元、以及M10與M14預估降雨

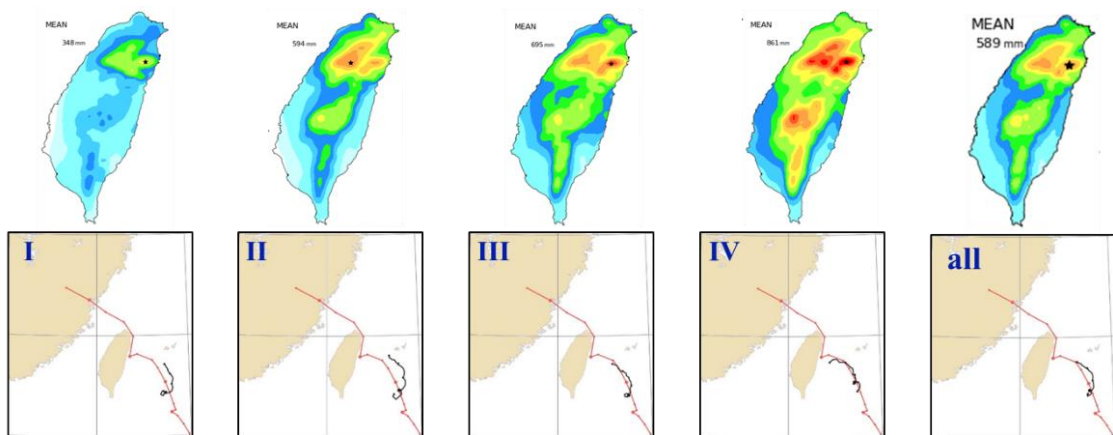


圖5、蘇拉颱風之4個群集個別之系集降雨平均與系集路徑平均

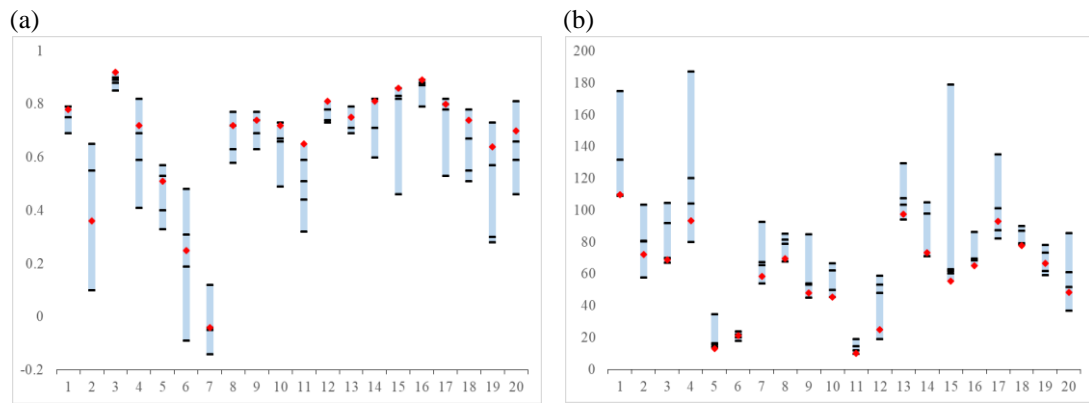


圖6、20場降雨事件群集系集平均(黑線)與系集平均(紅點)的校驗結果(a)相關係數(b)均方根誤差

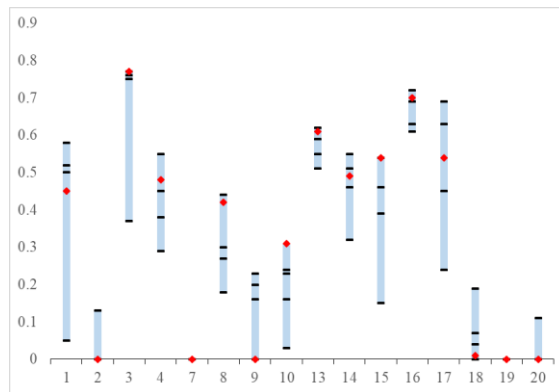


圖7、16場降雨事件群集系集平均(黑線)與系集平均(紅點)在豪雨門檻(200mm/24hr)的預兆得分