農業工程學報｜第 69 卷 – 第 3 期｜
中華民國 112 年 09 月出版

應用非監督學習與時間融合變換器優化水庫集水區中展期降雨預報

APPLYING UNSUPERVISED LEARNING AND TEMPORAL FUSION TRANSFORMER MODELS TO OPTIMIZE MEDIUM AND EXTENDED-RANGE RAINFALL FORECAST IN RESERVOIR CATCHMENT AREAS

國立臺灣大學
生物環境系統工程學系
博士班研究生
經濟部水利署
水利規劃試驗所
助理研究員

許 家 銓*

Chia Chuan Hsu

摘　要

近年來，臺灣地區乾旱頻傳，水資源管理上對於有效降雨預報的需求日益增加；而隨著數值天氣預報技術精進，降雨預報的準確性和可靠性已有長足進展。本研究提出一種結合非監督式學習 (Unsupervised learning) 及時間融合變換器模型 (Temporal fusion transformer model, TFT) 的方法，用於優化中展期系集降雨預報之準確性。非監督式學習是一種機器學習的方法，無需事先進行資料標記，這大幅增加了資料處理效率及使用彈性，且可自動對輸入的資料進行分群。時間融合變換器模型是一種基於時間序列轉譯概念所發展的機器學習模式，能夠有效處理時序列資料，並捕捉其內在的關聯性和變化性，更能有效學習長範圍的時間依賴性，對於時間序列預測問題具有較高的效能。

本研究以臺灣地區石門水庫、德基水庫及曾文水庫等 3 座水庫集水區為研究案例，以本研究提出之方法推估未來 30 日的逐旬累積降雨之預報，並與常見的機率擬合法平均 (Probability-matched ensemble mean, PM) 進行比較，在 3 處研究案例水庫集水區中，可降低約 39 %至 61 %的旬降雨預報誤差。藉由本研究所發展之優化架構，可提升降雨預報於水庫集水區的應用效能，為水資源管理單位理解雨量預報產品在水庫水源管理上的應用提供實用的工具。

關鍵詞：非監督式學習、時間融合變換器、中展期降雨預報。

* 國立臺灣大學生物環境系統工程學系博士班研究生、經濟部水利署水利規劃試驗所助理研究員
台中市南區忠明南路 932 號 5 樓之 1，d08622003@ntu.edu.tw

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 69, NO. 3, SEPTEMBER 2023
DOI - 10.29974/JTAE.202309_69(3).0005

APPLYING UNSUPERVISED LEARNING AND TEMPORAL FUSION TRANSFORMER MODELS TO OPTIMIZE MEDIUM AND EXTENDED-RANGE RAINFALL FORECAST IN RESERVOIR CATCHMENT AREAS

Chia Chuan Hsu*
National Taiwan University
Department of Bioenvironmental Systems Engineering;
Water Resources Planning Institute, Water Resources Agency,
Ministry of Economic Affairs

ABSTRACT

In recent years, Taiwan has frequently experienced droughts, leading to an increasing demand for effective rainfall forecasts in water resource management. Alongside advancements in numerical weather forecasting technology, significant progress has been made in the accuracy and reliability of rainfall forecasts. This study proposes a method that combines unsupervised learning and the Temporal Fusion Transformer model (TFT) to enhance the accuracy of medium and extended-range ensemble rainfall forecasts. Unsupervised learning, a method of machine learning that does not require prior data labeling, significantly improves data processing efficiency and flexibility, and it can automatically cluster the input data. The Temporal Fusion Transformer model is a machine learning model developed based on the concept of time-series translation, which effectively handles time-series data, captures its inherent correlations and variations, and efficiently learns long-range temporal dependencies, providing higher performance for time-series prediction problems.

This study uses the cstchment areas of the Shimen, Techi, and Zengwen Reservoirs in Taiwan as case studies. The proposed method can estimate forecasts for accumulated rainfall over each ten-day period for the next 30 days, and the results were compared with the commonly used Probability-Matched Ensemble Mean (PM). In the three case study catchment areas, the method reduces ten-day rainfall forecast errors by approximately 39% to 61%. Through the optimization framework developed in this study, the application performance of rainfall forecasts in reservoir catchment areas can be enhanced. This study provides a practical tool for water resource management institutes to understand the application of rainfall forecast products in reservoir water management.

Keywords: Unsupervised learning, Temporal fusion transformer model, Medium and extended-range rainfall forecast.

Hsu, C.C.*. (2023). “Applying Unsupervised Learning and Temporal Fusion Transformer Models to Optimize Medium and Extended-range Rainfall Forecast in Reservoir Catchment Areas.” *Journal of Taiwan Agricultural Engineering,* 69(3), 57-71. https://doi.org/10.29974/JTAE.202309_69(3).0005

一、前言

臺灣近年多次發生嚴重的乾旱事件，尤其是 2021 年初的「百年大旱」，造成西部地區的水庫蓄水量降至歷史低點，影響民生、農業和工業的用水。2020 年的颱風季節中，臺灣 56 年來首次未有颱風登陸，這導致了史上最嚴重的乾旱之一。由於水庫蓄水量降至 5 %以下，超過百萬的家庭和企業需進行分區供水。在 2020 年和 2021 年上半年，臺灣的降雨量遠低於平均水平。根據中央氣象局 2020 年台灣氣候分析，2020 年夏季的降雨量比平均水平少約 25 %，秋季的降雨量比平均水平少約 50 %。冬季也異常乾燥，而 2021 年春季的情況更糟，降雨量比平均水平少了 75 %。因此，到 2021 年夏季時，水庫的蓄水量已降至 50 年來的最低水平。乾旱在臺灣並不罕見，但這種情況的嚴重性是前所未有的。

為能提早掌握未來可能的降雨情況並提升水資源利用率，數值天氣預報 (Numerical Weather Prediction, NWP) 模型的發展對於提高預測精確性至關重要，相關研究中已多有運用季節性預報輔助水文預測及水資源管理的案例 (Chang and Guo, 2020; Peñuela *et al*., 2020)，惟受限於天氣預報系統的複雜性，長期數值天氣預報仍具較高不確定性，但中期天氣預報 (3 至 10 日) 及展期天氣預報 (10 至 30 日) (WMO and WHO, 2016)，已有不錯的精度 (Bauer *et al*., 2015; Frnda *et al*., 2015)。

數值天氣預報是以物理機制為基礎，使用數學模型來預測未來天氣的方法。傳統的數值天氣預報通常只產生一種可能的未來天氣情況，然而，由於氣候系統的複雜性和混沌性，這種單一預測的方法可能無法準確反映所有可能的未來天氣情況。因此，多數現代的數值天氣預報方法已經轉向使用系集預報 (Ensemble Forecasting)，優點在於提供更全面的資訊，這有助於更好理解和量化預測的不確定性。Du 與 Chen 於 2010 年的研究表明，系集平均預測通常比單一預測更準確，而缺點在於計算資源需求大及解釋困難且可能包含雜訊，因此，為能進一步提升系集預報的效能，在 Lou 等人 (2021)、Modali 等人 (2022) 及 Lamberson 等人 (2023) 的研究中均提出將系集預報進行集群分析，能增強模式效能及適應性，且有助於理解不同氣象或降雨事件間的共通性和差異性，運用於後續預報模式的優化。

綜合上述觀點，在本研究中採取了結合機器學習方法中的非監督學習和時間序列分析的方法，旨在分析並優化系集預報中逐次預報的成員關聯性。通過探索預報資料的結構進行資料分群，並根據不同的分群原則選用了四種常見的分群方法：K 平均法 (K-means)、高斯混合模型 (Gaussian Mixture Models, GMM)、層次密度基礎空間集群 (Hierarchical Density-Based Spatial Clustering of Applications with Noise, HDBSCAN) 和排序點識別集群結構法 (Ordering points to identify the clustering structure, OPTICS)。這種非監督式分群技術可以評估系集預報資料的一致性及各分群成果的效能，並且可以節省標記分群所需的時間。並進一步運用時間融合變換器模型 (Temporal Fusion Transformer, TFT) 的時間序列轉譯特性，有效地處理時間序列資料並捕捉其內在的關聯性和變化性。相較於傳統的遞歸神經網路 (RNN) 或長短期記憶網路 (Long short-term memory, LSTM)，TFT 模型更能有效地學習長範疇的時間依賴性，因此在處理時間序列預測問題時具有較高的效能。透過整合這些分群結果，優化整體的預報效能，產生出更符合水資源管理需求的降雨預報結果。這種方法不僅提高系集預報的效能，而且對於每次預報中的系集成員的關聯性，以及這些成員在臺灣地區北中南水庫集水區的運用上的差異，有更深入的理解。本研究成果為水資源管理單位在水庫水源管理上，理解雨量預報產品的應用，提供了一個實用的工具。

二、研究方法

2.1 非監督式學習資料分群方法

為探討降雨預報資料中潛在的資料結構，並將資料進行有效的分群，藉以探討對於特定時段及區域，各系集成員是否具有群聚情況，透過此集群結果，找出提升預報效能之可能性。本研究根據不同的分群原則，選擇了四種廣泛應用的資料分群方法：K 平均法 (K-means)、高斯混合模型 (Gaussian Mixture Models, GMM)、層次密度基礎空間集群 (Hierarchical Density-Based Spatial Clustering of Applications with Noise, HDBSCAN) 和排序點識別集群結構法 (Ordering Points to Identify the Clustering Structure, OPTICS)。以下簡述 4 種方法之集群概念：

K 平均法 (K-means) 是由 MacQueen 於 1967 年提出，主要是通過最小化每個群組中的資料點與其群組中心的距離之和來進行分群，如式 (1) 所示。

主要優點是算法複雜度低和演算速度快，因此對大型資料集特別有效。此外，其集群結果也相對易於理解，每個資料點都能明確劃分到一個群集中。然而，其局限在於，需要預先確定群集數量，對於非凸形狀、密度不均、差異性大的數據集分群效果較有限，且對初始中心點的選擇敏感。

$$\arg\min_{\mu}\sum_{c=1}^{K}\sum_{i=1}^{n_c}\left\|x_i-\mu_c\right\|^2 \quad \text{.....(1)}$$

其中，x_i 為資料點，K 為群心數，n_c 為在第 c 群中的資料數，μ_c 為群心，$\|x-\mu\|$為歐基李德距離 (Euclidean distance)。

高斯混合模型(GMM)是由 Duda 及 Hart 於 1973 年提出，假設所有的資料點都是由單一混合高斯分佈中生成的，每個高斯分佈對應一個分群，而每個資料點對應概率 (π_k)，如公式 (2) 及公式 (3)，據以決定該點所屬的分群。參數的擬合多通過最大化似然函數實現 (例如：期望最大化法 (Expectation-Maximization))。優點是其靈活性，能夠模擬具有多個峰值的分佈。此外，GMM 可使每個資料點都有屬於每個群集的概率，而不僅是被劃分到特定群集；這使得 GMM 可以提供更多的集群資訊，使用上更為彈性。然而，GMM 需要對每個群集的數量進行預設，且對於資料結構和變異性較為敏感。

$$p(x)=\sum_{k=1}^{K}\pi_k\,N(x|\mu_k,\Sigma_k) \quad \text{.....(2)}$$

$$\sum_{k=1}^{K}\pi_k=1 \quad \text{.....(3)}$$

其中，各高斯分布出現的機率為 π k，μ_k 及 Σ_k 為對應之平均值及標準差。

排序點識別集群結構法 (OPTICS) 是由 Ankerst 等人於 99 年提出的基於密度 (Density-based) 的集群分析算法，是一種用於探索基於密度的空間群集結構的演算法。OPTICS 的主要概念是創建一種群集順序或者排序，這種排序對資料集的密度和空間結構有良好的表示，原理是基於基於密度的帶噪聲應用空間聚類法 (Density-based spatial clustering of applications with noise, DBSCAN) (Ester *et al*., 1996)，再加入核心距離 (core-distance) 及可達距離 (reachablity-distance) 的概念，從而改善處理密度變化較劇烈的時所遇到的問題。OPTICS 不需預先設定群集的數量，並且可以處理任何形狀的群集和不同密度的群集。

層次密度基礎空間集群 (HDBSCAN) 是由 Campello 等人於 2013 年所開發的集群演算法，同樣是基於密度的分群方法，可識別出具有各種形狀和大小的群集並處理噪聲點。該方法首先構建一個距離矩陣，然後進行層次集群以找出密度峰值。接著，將這些密度峰值進行修剪來形成群集。最後，透過評比群集的穩定性，來決定最終的群集劃分。這種方法的優點是能夠自動決定群集的數量，無需預設。此外，HDBSCAN 還具有處理不同密度群集的能力，使其在處理具有不同密度區域的資料集時表現良好。

2.2 時間融合變換器

在機器學習於時間序列分析領域的應用中，相對於傳統的循環神經網路(RNN) 和長短期記憶 (LSTM) 模型，近年由 Lim 等人於 2021 年所提出的時間融合變換器 (Temporal Fusion Transformer, TFT)，能靈活的進行多時態的預測，更有效的處理多變量時間序列。TFT 便於將靜態和時間變化特徵的結合，且可透過注意力機制 (Attention mechanism)，提供對模型預測行為的解釋。在 Lim 等人 (2021) 及 Koya 與 Roy (2023) 的研究中均顯示，相較於 RNN 和 LSTM，TFT 能夠透過注意力機制和變量選擇網路 (Variable selection network) 增強其可解釋性。這有助於觀察模型在特徵資料時間序列中所關注的時間點，也能透過變量選擇網路評估各項特徵對預測結果的重要性 (通常介於 0 到 1 之間)，進而提升模型的可解釋性 (interpretability) 及預報效能。

時間融合變換器 (TFT) 是一種專為多時態預測設計的深度學習模型，其結構如圖 1 所示。TFT 接收靜態元數據 (Static Metadata)、過去輸入資料 (Past Inputs) 以及已知的未來輸入資料 (Known Future Inputs)；變量選擇 (Variable Selection) 用於選擇最顯著的特徵資料；閘控殘差網絡 (GRN) 通過殘差連接 (Residual Connection) 和閘控層 (Gate) 實現有效的信息傳遞；時間依賴性以 LSTM 進行處理，並使用多頭注意力機制 (Multi-head Attention) 來整合任何時間步長的資訊；通過將多種輸入融合，並學習這些輸入與目標變量的交互關係，以達到高效的預測。TFT 的主要優勢在於其對多時態預測的高效能，以及對時間動態的可解釋性。該模型可以幫助使用者識別對預測問題的重要全域變量、持續性的時間關聯性，以及顯著的事件。

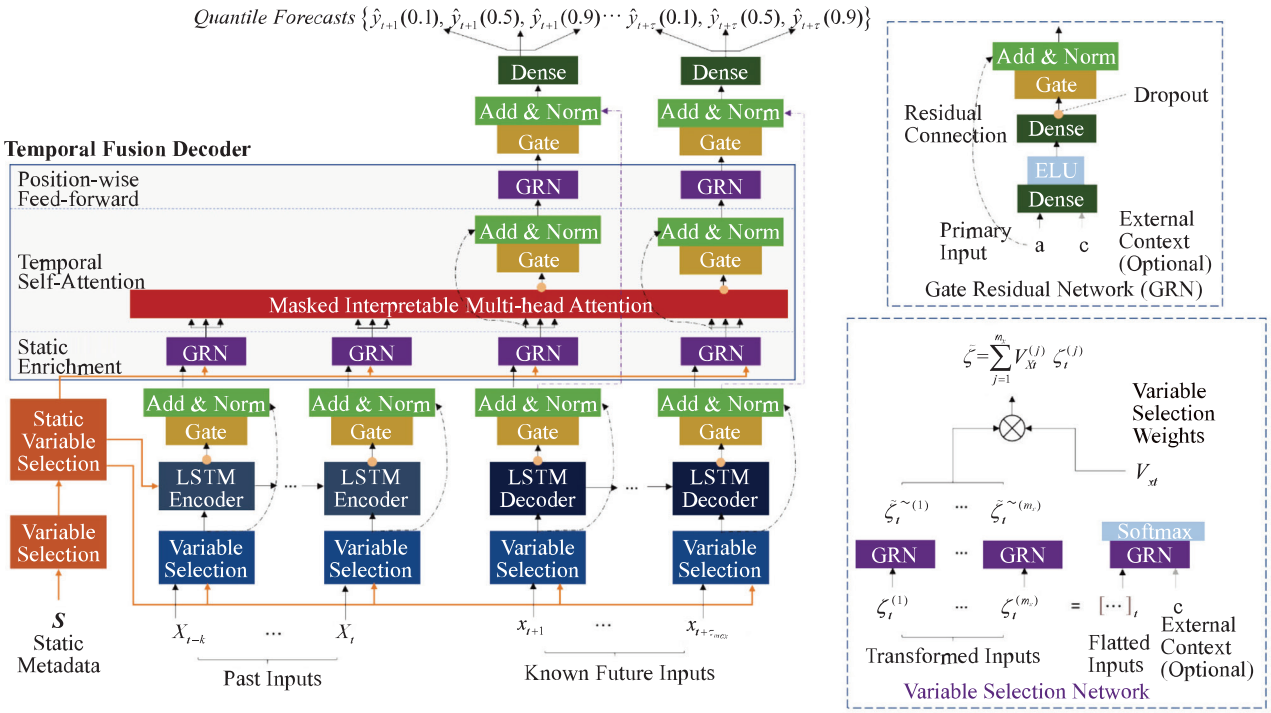


圖 1　時間融合變換器 (Temporal Fusion Transformer, TFT) 架構圖 (重繪自 Lim et al., 2021)

三、研究區域概述與資料蒐集

本研究以臺灣地區石門水庫 (北部)、德基水庫 (中部)、曾文水庫 (南部)集水區為研究區域。彙整各水庫集水區內近 20 年 (2001 年迄今) 雨量站資料，以及經氣象局統計後處理 (Statistic post-processing, SPP) 之歐洲中期天氣預報中心 (European Centre for Medium-Range Weather Forecasts, ECMWF) 之中展期 (1 至 30 日) 雨量預報資料。資料年限為 2001 年至 2019 年及 2022 年，約 250 萬張全臺降雨系集成員之預報網格，作為主要分析資料，探討本研究所提出之降雨預報優化方案於臺灣北部、中部及南部地區之預報效能。

3.1 資料前處理

本研究係基於中央氣象局取得歐洲中期天氣預報中心 (ECMWF) 之中展期雨量預報，進行統計後處理 (SPP) 之降雨預報資料。其中統計後處理主要包含：類比後處理 (Analog Post- Processing)、偏差修正 (Bias correction) 與降尺度 (Downscaling) 等流程。透過氣象局作業化程序，可產製臺灣含離島地區高解析格點 (1 km × 1 km) 即時展期雨量預報，共有 42,158 個格點，進行後續分析。

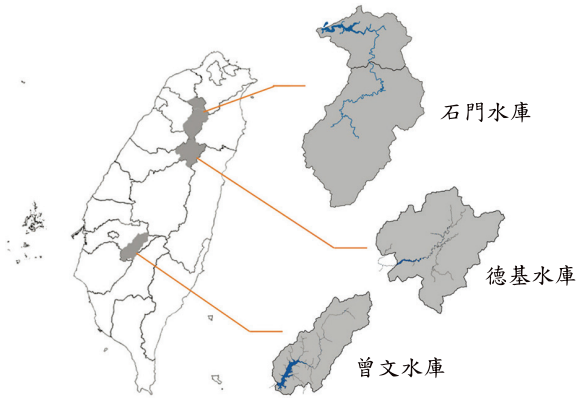


圖 2　研究範圍圖

首先，為獲得各水庫集水區內的高解析度網格位置 (1 km × 1 km)，在空間上，將高解析度網格與各水庫集水區圖層進行套疊，以確定每個水庫集水區內的所有網格點位，如圖 2 所示。分別擷取石門水庫、德基水庫及曾文水庫集水區所在區位，從而計算水庫集水區內所有網格的預報降雨量平均值，作為該水庫集水區的代表降雨量預報值。

在時間上，因統計後處理的降雨預報資料時間格式為未來 30 日的逐日 7 日累計降雨預報 (每次預報 24 組)，約每 1～2 日進行 1 次預報。輸出結果包含 (1) 20 組系集成員降雨預報、(2) 1 組決定性降雨預報；決定性降雨預報係由機率擬合平均法

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 69, NO. 3, SEPTEMBER 2023
DOI - 10.29974/JTAE.202309_69(3).0005

預報＼時間	1	2	3	4	5	6	7	8	9
$F_{1\text{-}4}$	X_1	X_1	X_1	X_1					
$F_{2\text{-}5}$		X_2	X_2	X_2	X_2				
$F_{3\text{-}6}$			X_3	X_3	X_3	X_3			
$F_{4\text{-}7}$				X_4	X_4	X_4	X_4		
$F_{5\text{-}8}$					X_5	X_5	X_5	X_5	
$F_{6\text{-}9}$						X_6	X_6	X_6	X_6
	S_1			S_2			S_3		

註：F：預報降雨量，例：$F_{1\text{-}4}$ 為第 1 至 4 日預報累積降雨量。
X：預報日降雨量，例：因採均化分配，X_1 為 $F_{1\text{-}4}$ 累積降雨除以 4。
S：目標天數之累積降雨，例：S1 為為第 1 至 3 日預報累積降雨量。
計算方式如下

$S_1 = X_1 + (X_1+X_2)/2 + (X_1+X_2+X_3)/3$

$S_2 = (X_1+X_2+X_3+X_4)/4 + (X_2+X_3+X_4+X_5)/4 + (X_3+X_4+X_5+X_6)/4$

$S_3 = (X_4+X_5+X_6)/3 + (X_5+X_6)/2 + X_6$

圖 3　均化分配法示意圖 (以日更新之 4 日累積預報轉換為 3 日累積預報為例)

(Probability-matched ensemble mean, PM) 方式推估。機率擬合平均法係由 Ebert 於 2001 年提出，概念是基於系集平均的空間降雨強度分布，將系集平均的降雨頻率以整體系集預報系統的降雨頻率取代，因其演算概念簡單且具備一定效能，故常用於空間降雨分布之分析。

考量預報資料不確定性及水資源管理應用之便利性，本研究透過均化分配方式進行預報時間維度轉換，如圖 3 所示 (以日更新之 4 日累積預報轉換為 3 日累積預報為例)：將統計後處理過後之 20 組 7 日累計降雨預報轉換為未來 30 日的第 1 旬、第 2 旬及第 3 旬之旬累計降雨預報。

四、模式建置及效能評估

4.1 模式建置

本研究結合非監督式集群及深度學習的方法來預測未來 30 日的石門水庫、德基水庫及曾文水庫集水區的旬降雨量。首先，通過資料前處理步驟，生成 20 組旬降雨預報資料。為瞭解這些預報的資料結構及一致性，使用 4 種非監督式集群方法 (K-means, GMM, HDBSCAN, OPTICS) 進行分群。因分群方式的差異，K-means 與 GMM，須明確指定群集數量，而 HDBSCAN 及 OPTICS 是基於密度的集群演算法，並且可以自動決定分群群數。為利分析評比，本研究優先進行 HDBSCAN 及 OPTICS 之分群，藉以得出此降雨預報資料集之分群情況，而 K-means 與 GMM 等 2 種方法，則參考前面 2 類之結果進行群集數量設定。

初步分群結果顯示，這 20 組預報結果多數能呈現出群聚特性，考量此等分群情況及 TFT 具備能分析時序特性之能力，將未來第 1 旬、第 2 旬、第 3 旬分群結果之旬雨量及日雨量與過去第 1 旬、第 2 旬的歷史觀測降雨資料 (旬雨量及日雨量) 作為時間融合變換器 (TFT) 的輸入項。通過此種方式，來優化並產製未來第 1 旬、第 2 旬、第 3 旬的水庫集水區旬降雨預報。TFT 模式輸入項及輸出項詳如表 1。

表 1　優化模式預設輸入項與輸出項

項目	因子
輸入項	未來第 1 旬、第 2 旬、第 3 旬雨量分群結果之旬雨量及日雨量。 過去第 1 旬、第 2 旬歷史觀測旬雨量及日雨量
輸出項	未來第 1 旬、第 2 旬、第 3 旬雨量預報

4.2 集群分析結果

蒐集 2001 至 2022 年的預報降雨資料，經資料前處理過程後共轉換為 4554 次預報資料。為利於後

續分析，將資料分為 2 類，第一類為單群 (single cluster) 表示資料中的資料點都被認為有足夠的相似性，可被視為一個群，或因為集群演算法的設定，使得所有的資料點都被歸入一個群；第二類為多群 (multiple cluster) 表示資料有明顯分群，且可分成 2 群以上，代表著資料中的點被分成了多個不同的群，每個群內的數據點有足夠的相似性，而不同群組之間的數據點則有顯著差異。

為探討集群分析成果是否具優化預報之可能性，先進行定性分析，比較集群結果與機率擬合平均法 (PM) 結果之優劣。經過資料前處理，每次預報資料可轉換為未來第 1 旬、第 2 旬及第 3 旬降雨預報等 3 個維度，在資料空間中可繪製為 3 維圖形，預報誤差則可以預報資料與觀測資料在資料空間中的距離表示，詳如式 (4) 及圖 4 所示。以各集群結果之質心至觀測資料 (Obs) 之距離與機率擬合平均法 (PM) 至觀測資料 (Obs) 之距離進行比較，並將集群結果為多群者，依群內數量進行排序，如 C-1 表示群內數量最多的群，C-2 表示群內數量次多的群，依此類推，用以評估優化可能性。

$$D_{Error} = \sum_{i=1\sim3} \sqrt{(f_i - o_i)^2} \quad \text{.....(4)}$$

D_{Error} ：誤差距離

f_i ：預報數值

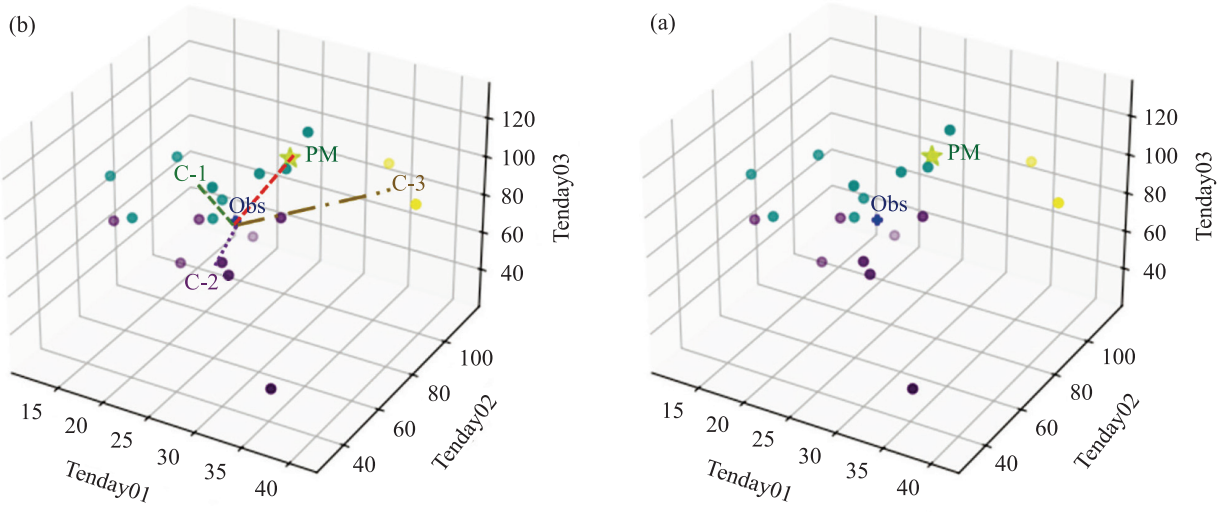
o_i ：觀測數值

由 HDBSCAN 集群分析結果顯示，單群佔約 54 %，多群佔約 46 % (分成 2 群的約佔 3 %、分成 3 群的約佔 42 %、分成 4 群的約佔 1 %)，以分成 1 群的最大宗。此分群結果於石門水庫、德基水庫及曾文水庫分布趨勢大致相同，如表 2 所示。依定性分析結果而言，分成 1 群有約 56 %報能被優化、分成 2 群有約 75 %的預報能被優化、分成 3 群有約 79 %的預報能被優化、分成 4 群有約 84 %的預報能被優化，如表 3 所示，此處優化是指能由集群結果之質心數值中，直接找出一組優於 PM 之預報結果。比較 3 處資料可知，集群分析結果佔優依序為石門水庫 (69 %)、曾文水庫 (67 %)、德基水庫 (63 %)。

由 OPTICS 集群分析結果顯示，單群佔約 4 %，多群佔 96 % (分成 2 群的約佔 27 %、分成 3 群的約佔 47 %、分成 4 群的約佔 20 %、分成 5 群的約佔 2 %)，其中分成 2 群及 3 群為最大宗，共佔約 74 %。

表 2　HDBSCAN 集群分析結果-集群數

分群數	石門水庫	德基水庫	曾文水庫
1 群	2434	2480	2487
2 群	136	124	132
3 群	1918	1891	1868
4 群	66	59	67



註：三軸 Tenday01、Tenday02 及 Tenday03 分別代表第 1 旬、第 2 旬及第 3 旬的旬降雨量；線段代表各集群質心 (C-1、C-2、C-3) 及機率擬合平均法 (PM) 至觀測降雨 (Obs) 之誤差距離。

圖 4　資料空間示意圖，(a) 系集預報、觀測資料 (Obs) 及機率擬合平均法 (PM) 空間分布；(b) 集群質心 (C-1、C-2、C-3)、機率擬合平均法 (PM) 與觀測資料 (Obs) 誤差距離圖。

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 69, NO. 3, SEPTEMBER 2023
DOI - 10.29974/JTAE.202309_69(3).0005

表 3　HDBSCAN 集群分析結果-定性分析

分群數	1 群	2 群	3 群	4 群	總計
石門水庫					
集群佔優	1375	103	1618	62	3158
PM 佔優	1059	33	300	4	1396
小計	2434	136	1918	66	4554
德基水庫					
集群佔優	1269	90	1471	48	2878
PM 佔優	1211	34	420	11	1676
小計	2480	124	1891	59	4554
曾文水庫					
集群佔優	1507	100	1383	52	3042
PM 佔優	980	32	485	15	1512
小計	2487	132	1868	67	4554

表 4　OPTICS 集群分析結果-集群數

分群數	石門水庫	德基水庫	曾文水庫
1 群	193	146	188
2 群	1199	1218	1211
3 群	2193	2138	2097
4 群	868	944	939
5 群	101	108	119

表 5　OPTICS 集群分析結果-定性分析

分群數	1 群	2 群	3 群	4 群	5 群	總計
石門水庫						
集群佔優	95	926	1813	763	96	3693
PM 佔優	98	273	380	105	5	861
小計	193	1199	2193	868	101	4554
德基水庫						
集群佔優	76	803	1606	748	93	3326
PM 佔優	70	415	532	196	15	1228
小計	146	1218	2138	944	108	4554
曾文水庫						
集群佔優	108	804	1551	764	96	3323
PM 佔優	80	407	546	175	23	1231
小計	188	1211	2097	939	119	4554

此分群結果於石門水庫、德基水庫及曾文水庫均呈現相同的趨勢，如表 4 所示。依定性分析結果所示，分成 1 群有約 53 %的預報能被優化、分成 2 群有約 70 %的預報能被優化、分成 3 群有約 77 %的預報能被優化、分成 4 群有約 83 %的預報能被優化、分成 5 群有約 87 %的預報能被優化，如表 5 所示。比較 3 處資料可知，集群分析結果佔優依序為石門水庫 (81 %)、曾文水庫 (73 %)、德基水庫 (73 %)。

依上述 2 種集群分析結果，考量 TFT 對特徵資料筆數及優化可能性之需求，分為 3 群之資料筆數多 (HDBSCAN：5,677 筆、OPTICS：6,428 筆)，且優化情況佳 (HDBSCAN：79 %、OPTICS：77 %)，故以 3 群作為 K-means 及 GMM 分群數設定依據，以利在相同基準分析。

表 6　K-means 及 GMM 集群分析結果-定性分析

方法	K-means	GMM
石門水庫		
集群佔優	3840	3819
PM 佔優	714	735
德基水庫		
集群佔優	3426	3409
PM 佔優	1128	1145
曾文水庫		
集群佔優	3398	3378
PM 佔優	1156	1176

K-means 及 GMM 在分群數設定為 3 的條件下，3 處案例整體定性分析結果相似，集群分析結果佔優依序為石門水庫 (K-means：84 %、GMM：84 %)、德基水庫 (K-means：75 %、GMM：75 %)、曾文水庫 (K-means：75 %、GMM：74 %)，詳如表 6。

4.3 特徵相關性分析

為瞭解不同特徵變數與觀測降雨量之間的關係，作為後續 TFT 模式輸入項調整之參考，本研究採用 Pearson 相關性分析，此為一種衡量兩個變數線性關係強度和正負相關的方法。首先，基於特徵變數對觀測降雨量可能影響性，將降雨預報集群分析後的結果及歷史觀測降雨量納入，旨在初步檢驗此等特徵變數與觀測降雨量 (第 1、2、3 旬) 之間是否存在顯著相關性。其次，分別將日預報雨量 (未來 1 ~ 30 日)、旬預報雨量 (未來 1、2、3 旬) 和觀測雨量 (前 1 ~ 20 日，日及旬雨量)與實際雨量 (未來 1、2、3 旬) 進行 Pearson 相關性分析。

依 4 類 (HDSCAN、OPTICS、Kmeans、GMM) 集群特徵變數對觀測降雨量之相關係數顯示，此等特徵變數與觀測降雨第 1 旬相關性最高、第 2 旬次之、第 3 旬居末，可進一步說明，此線性關係隨著預報時間愈遠而愈降低。整體而言，各集群分析結果 (C-1、C-2、C-3) 對應各旬觀測雨量 (T01、T02、T03)，均有高相關性。此外，在該旬內的部分日雨量，通常也存在較高相關性，因總體變數較多。選用 4 類集群方法分析後，相關性排名前 20 者進行展示，詳如表 7 及表 8 所示，而過去 20 日觀測雨量與觀測降雨

量關聯性低 (約 0.04 至 0.06)，可解釋降雨多為獨立事件。綜上而論，選用集群分析結果 (旬雨量及日雨量) 為主要資料，進行優化旬降雨預報之基礎，應屬合理。

表 7　HDSCAN、OPTICS 特徵變數相關性分析結果

HDBSCAN						OPTICS					
OBS_T01		OBS_T02		OBS_T03		OBS_T01		OBS_T02		OBS_T03	
C-1_D01	0.65	C-1_T02	0.48	C-1_D21	0.44	C-1_T01	0.64	C-1_T02	0.48	C-2_D20	0.41
C-1_D02	0.64	C-1_D13	0.48	C-1_D20	0.43	C-1_D01	0.64	C-1_D12	0.47	C-2_D19	0.40
C-1_D03	0.64	C-1_D14	0.48	C-1_D23	0.43	C-1_D02	0.63	C-1_D14	0.47	C-2_D21	0.40
C-1_T01	0.64	C-1_D12	0.47	C-1_D24	0.43	C-1_D03	0.63	C-1_D13	0.47	C-2_D22	0.40
C-1_D04	0.63	C-1_D15	0.47	C-1_D22	0.43	C-1_D04	0.62	C-1_D11	0.47	C-2_D18	0.40
C-1_D05	0.62	C-2_D12	0.47	C-2_D20	0.42	C-1_D05	0.61	C-1_D15	0.47	C-2_D23	0.40
C-2_T01	0.62	C-2_T02	0.47	C-1_D19	0.42	C-1_D06	0.61	C-1_D16	0.46	C-2_D24	0.39
C-2_D03	0.61	C-2_D13	0.47	C-1_T02	0.42	C-1_D07	0.61	C-1_D17	0.46	C-2_T02	0.39
C-1_D06	0.61	C-2_D14	0.46	C-2_D21	0.42	C-2_T01	0.60	C-1_D18	0.46	C-1_D21	0.39
C-2_D02	0.61	C-1_D11	0.46	C-1_D25	0.42	C-2_D03	0.60	C-1_D10	0.45	C-1_D24	0.39
C-2_D01	0.61	C-1_D16	0.46	C-2_D19	0.42	C-3_T01	0.59	C-1_D19	0.45	C-1_D20	0.38
C-1_D07	0.60	C-2_D11	0.46	C-2_D22	0.42	C-2_D04	0.59	C-3_T02	0.44	C-1_T02	0.38
C-3_T01	0.60	C-1_D17	0.45	C-1_D26	0.41	C-2_D05	0.59	C-2_D13	0.44	C-1_D23	0.38
C-2_D04	0.60	C-2_D15	0.45	C-1_D18	0.41	C-2_D02	0.59	C-1_D20	0.44	C-3_D19	0.38
C-2_D05	0.59	C-3_T02	0.45	C-2_D18	0.41	C-2_D06	0.59	C-3_D13	0.44	C-3_D18	0.38
C-2_D06	0.59	C-1_D10	0.45	C-2_D23	0.41	C-2_D07	0.58	C-2_D12	0.44	C-1_D25	0.38
C-2_D07	0.59	C-1_D18	0.45	C-2_D17	0.40	C-3_D03	0.58	C-3_D12	0.43	C-1_D22	0.38
C-3_D04	0.59	C-2_D10	0.44	C-2_T02	0.40	C-2_D01	0.57	C-2_T02	0.43	C-3_D17	0.38

註：OBS：觀測雨量；C：集群分析結果；D：日雨量，D01 代表第 1 日之日雨量，依此類推；T：旬雨量，T01 代表第 1 旬之旬雨量，依此類推。

表 8　K-Means、GMM 特徵變數相關性分析結果

K-Means						GMM					
OBS_T01		OBS_T02		OBS_T03		OBS_T01		OBS_T02		OBS_T03	
C-1_D02	0.59	C-1_T02	0.32	C-1_D21	0.23	C-1_D01	0.63	C-1_D13	0.49	C-1_D21	0.44
C-1_D03	0.59	C-1_D14	0.31	C-1_D22	0.23	C-1_D02	0.63	C-1_D12	0.49	C-1_D20	0.43
C-1_D01	0.58	C-1_D15	0.31	C-1_D23	0.22	C-1_D03	0.63	C-1_T02	0.49	C-1_D19	0.43
C-1_T01	0.58	C-1_D16	0.31	C-1_D20	0.22	C-1_T01	0.62	C-1_D14	0.48	C-1_D22	0.43
C-1_D04	0.58	C-1_D17	0.30	C-1_D19	0.21	C-1_D04	0.62	C-1_D11	0.47	C-2_D20	0.43
C-1_D05	0.57	C-1_D13	0.30	C-1_D24	0.21	C-1_D05	0.61	C-1_D15	0.47	C-1_D23	0.43
C-1_D06	0.56	C-1_D18	0.29	C-1_T02	0.21	C-1_D06	0.60	C-2_T02	0.46	C-1_D18	0.43
C-1_D07	0.54	C-1_D12	0.28	C-1_D25	0.20	C-1_D07	0.59	C-1_D16	0.46	C-2_D21	0.43
C-2_D02	0.54	C-1_D19	0.28	C-1_D18	0.20	C-2_T01	0.59	C-2_D12	0.46	C-1_T02	0.42
C-2_D03	0.53	C-1_D20	0.27	C-1_D26	0.20	C-2_D02	0.59	C-2_D13	0.46	C-1_D24	0.42
C-2_T01	0.53	C-1_D11	0.26	C-2_D21	0.20	C-2_D01	0.58	C-2_D11	0.46	C-2_D19	0.42
C-2_D01	0.53	C-2_T02	0.25	C-2_D20	0.19	C-2_D03	0.58	C-1_D17	0.46	C-1_D25	0.41
C-2_D04	0.52	C-1_D21	0.25	C-2_D22	0.19	C-2_D05	0.57	C-1_D10	0.45	C-2_D22	0.41
C-2_D05	0.51	C-2_D13	0.24	C-3_D21	0.19	C-3_T01	0.57	C-2_D14	0.45	C-1_D17	0.41
C-2_D06	0.50	C-2_D12	0.24	C-2_D19	0.19	C-2_D04	0.57	C-1_D18	0.45	C-1_D26	0.41
C-2_D07	0.49	C-2_D14	0.24	C-1_D17	0.19	C-2_D06	0.57	C-2_D10	0.44	C-1_T02	0.41
C-1_D08	0.49	C-2_D15	0.24	C-1_D27	0.19	C-2_D07	0.56	C-2_D15	0.44	C-2_D18	0.41
C-3_T01	0.49	C-2_D11	0.24	C-3_D22	0.19	C-3_D03	0.56	C-1_D19	0.44	C-2_D23	0.40

註：OBS：觀測雨量；C：集群分析結果；D：日雨量，D01 代表第 1 日之日雨量，依此類推；T：旬雨量，T01 代表第 1 旬之旬雨量，依此類推。

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 69, NO. 3, SEPTEMBER 2023
DOI - 10.29974/JTAE.202309_69(3).0005

4.4 TFT 模式效能分析

以集群分析 (分群數為 3) 之成果作為特徵資料，輸入時間融合變換器 (TFT)，進行未來第 1 旬、第 2 旬及第 3 旬之旬降雨量預測。因所蒐集的資料近 20 年，訓練資料與驗證資料比採用 8:2，16 年訓練資料 (2001 年至 2016)，4 年驗證資料 (2017 年至 2019 年，以及 2022 年)，採用平均誤差距離作為模式效能評估指標，如公式 (5) 所示，數值愈小表示效能愈佳。將 TFT 與機率擬合平均 (PM) 及集群結果之質心進行比較，優化模式可分為： (1) HDBSCAN-TFT、(2) OPTICS-TFT、(3) Kmeans-TFT、(4) GMM-TFT 等 4 種。

$$MD_{Error} = \frac{1}{n}\sum_{i=1\sim3}\sqrt{(f_i - o_i)^2} \quad \text{.....(5)}$$

MD_{Error} ：平均誤差距離

f_i ：預報數值

o_i ：觀測數值

4.4.1 HDBSCAN-TFT

以 PM 法為比較基準，HDBSCAN-TFT 可有效降低誤差，降低比例依序為曾文水庫 50 %、德基水庫 39 %、石門水庫 30 %。以誤差距離來看，依序為德基水庫 60.91 表現最佳、石門水庫 72.97 次之、曾文水庫 98.7 居末。C-1、C-2、C-3 分別表示集群分析後群內數量最多、次多及第三的集群質心與觀測值之平均誤差距離，若納入集群質心進行比較，C-1 結果在曾文水庫及石門水庫，會略優於 PM 結果，但兩者均不如 HDBSCAN-TFT，詳如表 9。

4.4.2 OPTICS-TFT

以 PM 法為比較基準，OPTICS-TFT 可降低誤差，但效果有限，降低比例依序為曾文水庫 18 %、德基水庫 7 %、石門水庫 5 %。以誤差距離來看，依序為德基水庫 91.5 表現最佳、石門水庫 103.65 次之、曾文水庫 169.54 居末。若納入集群質心進行比較，於石門及曾文水庫案例中，C-1 結果尚可，甚至略優於 OPTICS -TFT，詳如表 10。

4.4.3 Kmeans-TFT

以 PM 法為比較基準，Kmeans-TFT 可有效降低誤差，降低比例依序為石門水庫 61 %、曾文水庫 55 %、德基水庫 39 %。以誤差距離來看，依序為石門水庫 39.64 表現最佳、德基水庫 60.62 次之、曾文水庫 88.5 居末。若納入集群質心進行比較 C-1 結果會略優於 PM 結果，但兩者均不如 Kmeans-TFT，詳如表 11 。

表 9　HDBSCAN-TFT 平均誤差距離

項目	PM	HDBSCAN-TFT	C-1	C-2	C-3
石門水庫					
訓練	116.87(-)	81.04(31%)	112.84(3%)	117.97(-1%)	128.01(-10%)
驗證	103.97(-)	72.97(30%)	94.61(9%)	102.26(2%)	111.93(-8%)
德基水庫					
訓練	125.2(-)	87.51(30%)	124.82(<1%)	131.09(-5%)	142.25(-14%)
驗證	99.33(-)	60.91(39%)	101.22(-2%)	104.91(-6%)	124.8(-26%)
曾文水庫					
訓練	188.2(-)	86.6(54%)	136.3(28%)	148.69(21%)	166.96(11%)
驗證	196.68(-)	98.7(50%)	139.02(29%)	145.65(26%)	168.8(14%)

表 10　OPTICS -TFT 平均誤差距離

項目	PM	OPTICS-TFT	C-1	C-2	C-3
石門水庫					
訓練	117.12(-)	117.31(0%)	113.71(3%)	114.82(2%)	119.02(-2%)
驗證	108.87(-)	103.65(5%)	98.79(9%)	103.27(5%)	106.78(2%)
德基水庫					
訓練	122.24(-)	130.42(-7%)	122.31(0%)	124.63(-2%)	128.89(-5%)
驗證	98.4(-)	91.5(7%)	105.08(-7%)	100.88(-3%)	106.07(-8%)
曾文水庫					
訓練	192.65(-)	159.38(17%)	146.57(24%)	148.98(23%)	158.85(18%)
驗證	207.63(-)	169.54(18%)	141.63(32%)	147.11(29%)	157.65(24%)

表 11　K-means -TFT 平均誤差距離

項目	PM	Kmeans-TFT	C-1	C-2	C-3
石門水庫					
訓練	117.42(-)	44.9(62%)	111.42(5%)	123.34(-5%)	150.86(-28%)
驗證	101.48(-)	39.64(61%)	91.12(10%)	105.3(-4%)	146.18(-44%)
德基水庫					
訓練	125.2(-)	85.74(32%)	123.49(1%)	133.24(-6%)	167.46(-34%)
驗證	99.33(-)	60.62(39%)	96.37(3%)	110.4(-11%)	156.93(-58%)
曾文水庫					
訓練	188.31(-)	84.25(55%)	132.61(30%)	155.97(17%)	210.01(-12%)
驗證	196.68(-)	88.5(55%)	135.44(31%)	157.45(20%)	211(-7%)

表 12　GMM-TFT 平均誤差距離

項目	PM	GMM-TFT	C-1	C-2	C-3
石門水庫					
訓練	116.87(-)	99.67(15%)	113.16(3%)	118.71(-2%)	145.38(-24%)
驗證	103.97(-)	83.7(19%)	92.48(11%)	107.91(-4%)	139.41(-34%)
德基水庫					
訓練	125.2(-)	107.8(14%)	123.6(1%)	133.45(-7%)	164.07(-31%)
驗證	99.33(-)	71.57(28%)	96.43(3%)	111.35(-12%)	153.21(-54%)
曾文水庫					
訓練	188.31(-)	142.3(24%)	134.57(29%)	152.41(19%)	203.51(-8%)
驗證	196.68(-)	156.97(20%)	136.07(31%)	155.64(21%)	204.89(-4%)

4.4.4 GMM-TFT

以 PM 法為比較基準，GMM-TFT 可有效降低誤差，降低比例依序為德基水庫 28 %、曾文水庫 20 %、石門水庫 19 %。以誤差距離來看，依序為德基水庫 71.57 表現最佳、石門水庫 83.7 次之、曾文水庫 156.97 居末。若納入集群質心進行比較，於曾文水庫案例中，C-1 結果尚可，甚至略優於 GMM -TFT，詳如表 12。

總體而言，Kmeans-TFT 在預測效能上表現較佳，未來第 1 旬、第 2 旬及第 3 旬之觀測值、PM 及 TFT 於測試組 (2017、2018、2019、2022) 之表現如圖 5 所示。依據此結果，進行逐月誤差討論，如表 13、表 14、表 15 所示，在 3 處分析案例中，除德基水庫預報未來第 1 旬、第 2 旬及第 3 旬效能相近外，在石門水庫及曾文水庫，預報第 2 旬的效能明顯較佳，推測原因為在資料前處理階段，預報第 2 旬資料相對多，如圖 3 所示，故預報效能相對穩定。此外，在每年的 3 月至 8 月，無論是絕對誤差 (Absolute Error, AE) 或相對誤差 (Relative Error, RE)，三處案例的預報效能均良好。相比於 PM，Kmeans-TFT 在豐枯轉換期的月份 (4 至 7 月) 都顯示出一定程度的優化效果，推測於此時段，各系集成員預報分歧較明顯，而產生多群集之結果，因此 Kmeans-TFT 透過分群能夠有效的優化旬降雨預報結果。

然而，於颱風事件中，Kmeans-TFT 會有低估降雨量的情況，推測原因為：(1) 模式架構：TFT 是一種相對複雜的時間序列模型，能夠處理多個來源和不同時間尺度的特徵，並且學習其間的複雜互動。然而，這種複雜性也意味著模型需要大量的資料來有效地學習，由於颱風事件的資料相對少，可能導致 TFT 對颱風時期之降雨預報效能不佳；(2) 資料來源：因所使用的是中展期降雨預報資料，對颱風降雨這類短期事件掌握度較低；上述因素可能導致在颱風期間預報效能下降。

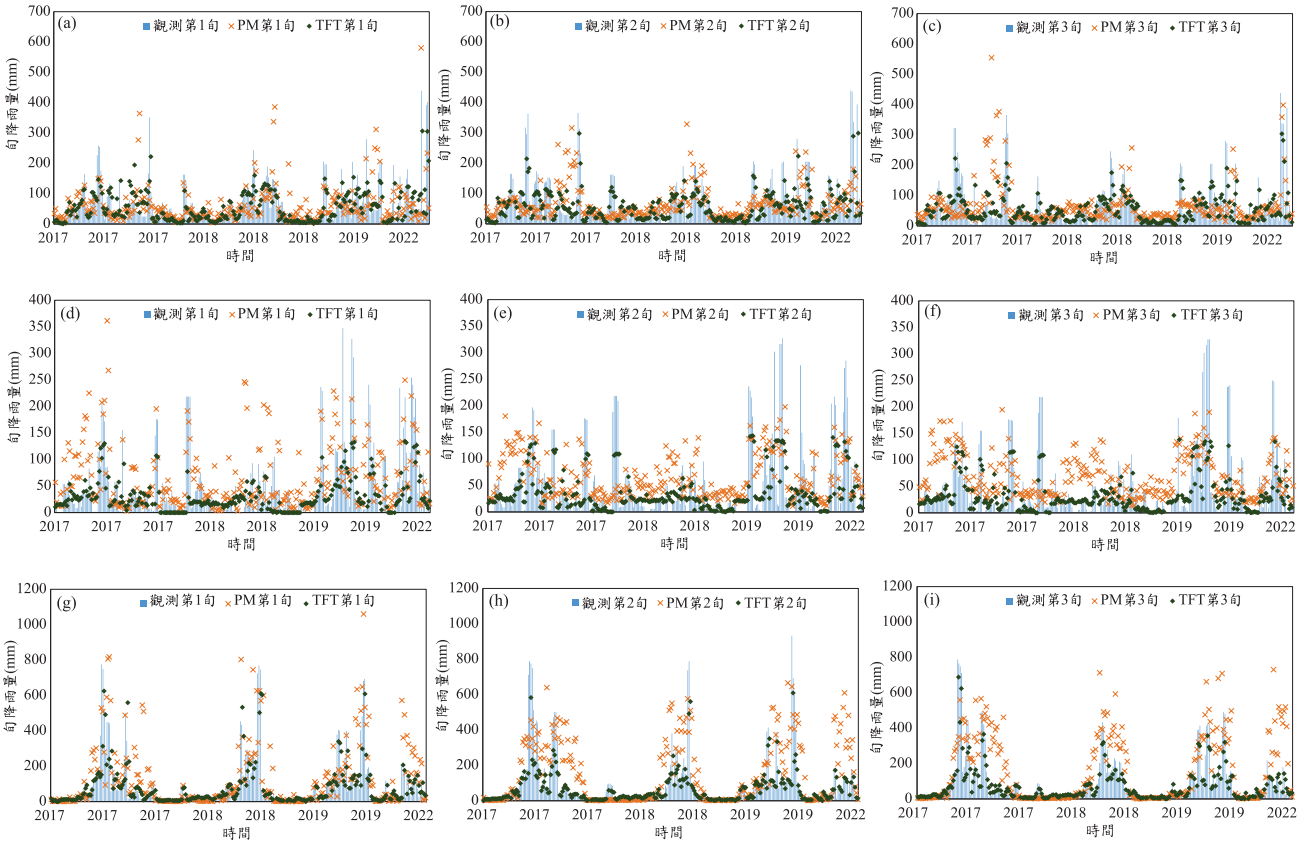


圖 5　Kmeans-TFT 預報雨量-驗證組 (2017、2018、2019、2022)，(a) 石門水庫-第 1 旬、(b) 石門水庫-第 2 旬、(c) 石門水庫-第 3 旬、(d) 德基水庫-第 1 旬、(e) 德基水庫-第 2 旬、(f) 德基水庫-第 3 旬、(g) 曾文水庫-第 1 旬、(h) 曾文水庫-第 2 旬、(i) 曾文水庫-第 3 旬。

表 13　石門水庫旬預報誤差分析

月份	預報第 1 旬						預報第 2 旬						第 3 旬					
	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化
1	11.04	23.49	-12.46	25%	54%	-29%	2.48	3.46	-0.99	10%	14%	-4%	4.93	6.41	-1.47	19%	25%	-6%
2	4.00	19.52	-15.52	10%	48%	-38%	6.64	8.49	-1.85	12%	15%	-3%	4.83	8.22	-3.39	9%	16%	-7%
3	29.46	8.27	21.19	66%	18%	47%	18.83	2.97	15.86	50%	8%	42%	21.12	3.64	17.48	51%	9%	42%
4	63.37	2.36	61.01	176%	7%	169%	10.00	9.26	0.74	13%	12%	1%	4.57	5.49	-0.92	8%	10%	-2%
5	13.45	31.06	-17.61	13%	30%	-17%	41.39	21.48	19.91	43%	22%	21%	46.98	26.12	20.86	55%	30%	24%
6	46.96	11.87	35.10	71%	18%	53%	35.70	11.99	23.71	39%	13%	26%	20.61	8.59	12.03	29%	12%	17%
7	14.71	0.01	14.71	47%	0%	47%	10.77	9.80	0.97	13%	12%	1%	3.35	5.61	-2.26	6%	10%	-4%
8	2.20	12.80	-10.61	4%	25%	-21%	23.65	13.39	10.25	23%	13%	10%	111.14	4.96	106.18	124%	6%	119%
9	33.31	3.94	29.36	105%	12%	93%	49.19	29.98	19.21	52%	32%	20%	18.21	18.59	-0.38	25%	26%	-1%
10	8.76	29.08	-20.32	20%	65%	-46%	10.79	3.05	7.73	21%	6%	15%	8.84	1.93	6.92	35%	8%	27%
11	11.24	17.74	-6.50	59%	93%	-34%	0.86	3.67	-2.80	3%	14%	-11%	12.00	3.30	8.71	61%	17%	45%
12	10.34	19.42	-9.08	53%	100%	-47%	28.82	20.49	8.33	51%	36%	15%	4.55	10.82	-6.27	14%	34%	-20%
平均	20.74	14.96	—	54%	39%	—	19.93	11.50	—	28%	16%	—	21.76	8.64	—	36%	17%	—

註：絕對誤差 (AE)、相對誤差 (RE)

表 14　德基水庫旬預報誤差分析

月份	預報第 1 旬						預報第 2 旬						預報第 3 旬					
	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化
1	3.52	18.84	-15.32	9%	51%	-41%	12.95	2.59	10.36	57%	11%	46%	19.90	0.29	19.61	94%	1%	93%
2	5.55	7.50	-1.96	17%	23%	-6%	5.06	25.52	-20.45	7%	34%	-27%	55.84	4.71	51.13	191%	16%	175%
3	5.17	10.92	-5.74	10%	22%	-11%	57.23	7.24	49.99	196%	25%	171%	73.06	1.27	71.79	249%	4%	244%
4	9.60	0.17	9.43	19%	0%	19%	49.38	11.48	37.91	88%	20%	67%	40.95	25.96	14.98	53%	33%	19%
5	14.30	3.84	10.47	22%	6%	16%	13.56	39.20	-25.64	11%	32%	-21%	8.68	18.34	-9.66	9%	20%	-10%
6	49.56	22.52	27.03	43%	20%	24%	19.07	2.80	16.28	45%	7%	38%	27.32	8.42	18.91	150%	46%	104%
7	14.30	11.01	3.29	23%	18%	5%	8.72	11.49	-2.77	15%	20%	-5%	41.54	40.26	1.28	47%	46%	1%
8	4.91	6.86	-1.94	6%	8%	-2%	8.59	14.84	-6.25	20%	35%	-15%	34.48	1.27	33.21	137%	5%	132%
9	78.73	27.26	51.47	113%	39%	74%	30.67	11.78	18.89	80%	31%	50%	2.48	12.56	-10.08	5%	24%	-19%
10	20.94	24.14	-3.19	25%	29%	-4%	10.26	6.15	4.11	41%	25%	17%	16.87	8.00	8.87	116%	55%	61%
11	1.69	14.57	-12.88	6%	50%	-45%	14.14	13.90	0.25	76%	75%	1%	28.63	9.52	19.11	225%	75%	150%
12	9.39	7.26	2.13	62%	48%	14%	28.28	29.82	-1.54	47%	50%	-3%	23.63	30.35	-6.72	39%	51%	-11%
平均	18.14	12.91	—	30%	26%	—	21.49	14.73	—	57%	30%	—	31.11	13.41	—	110%	31%	—

註：絕對誤差 (AE)、相對誤差 (RE)

表 15　曾文水庫旬預報誤差分析

月份	預報第 1 旬						預報第 2 旬						預報第 3 旬					
	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化	AE-PM	AE-TFT	優化	RE-PM	RE-TFT	優化
1	3.39	1.33	2.06	20%	8%	12%	5.26	3.08	2.19	47%	28%	20%	4.65	3.52	1.13	48%	36%	12%
2	4.51	4.42	0.10	35%	34%	1%	6.14	0.92	5.22	25%	4%	22%	5.51	3.18	2.34	29%	17%	12%
3	1.62	0.56	1.05	5%	2%	3%	3.44	6.78	-3.33	17%	34%	-17%	4.13	6.27	-2.14	18%	27%	-9%
4	16.81	2.97	13.84	40%	7%	33%	8.62	0.55	8.08	16%	1%	15%	13.17	27.60	-14.43	12%	24%	-13%
5	23.25	27.65	-4.39	14%	17%	-3%	90.27	89.59	0.68	37%	37%	0%	103.16	67.80	35.36	36%	24%	12%
6	93.50	45.17	48.34	39%	19%	20%	170.57	4.15	166.43	126%	3%	123%	228.88	25.63	203.24	169%	19%	150%
7	130.29	17.31	112.99	91%	12%	79%	175.04	95.68	79.36	79%	43%	36%	204.93	53.03	151.91	100%	26%	74%
8	78.08	47.56	30.52	32%	19%	12%	181.33	38.73	142.60	100%	21%	79%	253.75	5.46	248.30	413%	9%	404%
9	121.06	38.27	82.79	432%	136%	295%	102.80	1.41	101.39	378%	5%	373%	1.37	1.51	-0.14	4%	4%	0%
10	8.67	1.20	7.48	32%	4%	28%	3.64	4.25	-0.61	109%	128%	-18%	2.79	8.32	-5.53	139%	415%	-276%
11	2.40	6.92	-4.52	91%	263%	-172%	4.91	1.34	3.57	59%	16%	43%	2.29	1.65	0.64	30%	22%	8%
12	3.79	2.79	1.00	40%	29%	10%	27.37	10.40	16.98	87%	33%	54%	17.73	6.26	11.47	82%	29%	53%
平均	40.61	16.34	—	73%	46%	—	64.95	21.40	—	90%	29%	—	70.20	17.52	—	90%	54%	—

註：絕對誤差 (AE)、相對誤差 (RE)

五、結論

本研究運用 HDBSCAN、OPTICS、K-means、GMM 等 4 種非監督式學習方法進行集群分析。在多數情況下，僅使用集群分類，已具備優化系集降雨預報之效果，尤其在曾文水庫若直接選用群內數量最多之群集質心 (C-1) 作為降雨預報結果，相比於 PM 法之結果，可有約 30%的提升。在資料前處理上，主要差異在於 PM 法係以全臺範圍進行降雨重新分配，而本優化方法則是聚焦於集水區範圍。在多數情況下，本優化方法預報效果優於 PM 法，惟於颱風事件，當預報不確定性提高時，颱風路徑對降雨量空間分布的影響非常顯著，使得本優化方法預報

效能下降，當然也與所採資料來源為中展期降雨預報有關。

總體來看，Kmeans-TFT 在預測效能上表現優異，相比於 PM 法，能有效降低誤差，對於石門水庫、曾文水庫和德基水庫，分別降低 61 %、55 %和 39 %。因此，雖 K-means 的集群機制不如 HDBSCAN、OPTICS 和 GMM 複雜，但在本研究案例中，每次預測的系集成員相對較少 (20 組) 且資料維度較低 (3 維)，資料結構和規則相對簡單，更適合 K-means 發揮其效用；此外，HDBSCAN 和 OPTICS 得出的分群數也有助於 K-means 分群數的給定。綜上，K-means 對於資料使用的靈活性以及低演算複雜度，都使其在本研究資料集與未來應用情境上具備優勢。此外，結合 TFT 對時間序列資料處理的優勢，也進一步提升了降雨預測效能。

本研究所設計的優化方法，可將氣象單位所產製的全臺降雨系集預報網格，映射至水庫集水區，藉以產製特定水庫集水區未來 30 日的逐旬降雨預報，能得到更佳的預報效能，對於水資源管理應有所助益。

六、建議

本研究結果顯示，特別是對於颱風事件的短期預報，其預測能力仍存在改進的潛力。因此，建議採納如 ETQPF、CWBWRF 以及 WEPSPro 等短期預報工具，進行組合預報，從而提高颱風期間的預報準確性。此外，透過擴大資料維度，例如採用 7 日、3 日及 1 日等不同時間範疇的預報資料，可以生成更多功能性的預報產品。在預報模式的產出方面，建議在現有的決定性預報方法之外，納入 TFT 概率預報，這將擴大預報的覆蓋範疇並提升其對決策支援的參考價值。最後，預報成果可再結合降雨逕流模式，提供逕流預報，且就不同大小的水庫集水區進一步評估模式效能及預報不確定性。上述建議應有助於未來降雨預報在水資源領域產生積極影響。

誌謝

本研究承蒙交通部中央氣象局提供的歐洲中期天氣預報中心 (ECMWF) 中展期雨量預報的統計後處理 (SPP) 資料，以及水利規劃試驗所相關長官及同仁專業技術支援，得以順利完成階段性成果並持續精進，謹致謝忱。

參考文獻

1. Ankerst, M., Breunig, M.M., Kriegel, H.P., Jörg, S., “OPTICS: Ordering Points to Identify the Clustering Structure,” *ACM SIGMOD Record*, 1999-06-01, 28 (2): 49-60, 1999.
2. Bauer, P., Thorpe, A., Brunet, G., “The Quiet Revolution of Numerical Weather Prediction,” *Nature*, 525, 47-55, 2015.
3. Campello, R.J.G.B., Moulavi, D., Sander, J., “Density-based Clustering Based on Hierarchical Density Estimates,” *Advances in Knowledge Discovery and Data Mining*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 160-172, 2013.
4. Chang, F.J. and Guo, S., “Advances in Hydrologic Forecasts and Water Resources Management,” *Water*, 12, 1819, 2020.
5. Du, J. and Chen, J., “The Corner Stone in Facilitating the Transition from Deterministic to Probabilistic Forecasts-Ensemble Forecasting and Its Impact on Numerical Weather Prediction,” *Meteorological Monthly*, 36(11): 1-11, 2010.
6. Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis. Wiley, New York, 1973.
7. Ebert, E.E., “Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation,” *Monthly Weather Review*, 129, 2461-2480, 2001.
8. Ester, M., Kriegel, H. P., Sander, J., Xu, X., “A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD’96. AAAI Press, pp. 226-231, 1996.
9. Frnda, J., Durica, M., Rozhon, J., Vojtekova, M., Nedoma, J., Martinek, R., “ECMWF Short-term Prediction Accuracy Improvement by Deep Learning,” *Scientific Reports*, 12, 7898, 2022.
10. Koya, S. R. and Roy, T., “Temporal Fusion Transformers for Streamflow Prediction: Value of Combining Attention with Recurrence,” Machine Learning, Geophysics, 2023.
11. Lamberson, W. S., Bodner, M. J., Nelson, J. A., and Sienkiewicz, S. A., “The Use of Ensemble Clustering

on a Multimodel Ensemble for Medium-Range Forecasting at the Weather Prediction Center", *Weather Forecasting*, 38, 539-554, 2023.

12. Lim, B., Arık, S.Ö., Loeff, N., Pfister, T., "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting," *International Journal of Forecasting*, Volume 37, Issue 4, Pages 1748-1764, 2021.
13. Lou, D., Yang, M., Shi, D., Wang, G., Ullah, W., Chai, Y., Chen, Y., "K-Means and C4.5 Decision Tree Based Prediction of Long-Term Precipitation Variability in the Poyang Lake Basin, China," *Atmosphere*, 12(7):834, 2021.
14. MacQueen, J. B., "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley and Los Angeles: University of California Press, 1967.
15. Modali, K., Sander, D., Brune, S., Rupp, P., Garny, H., Baehr, J., and Rautenhaus, M., "A Framework for Comparative Cluster Analysis of Ensemble Weatherprediction Data," *EMS Annual Meeting Abstracts*, Vol. 19, EMS2022-531, 2022 (updated on 23 Aug 2023).
16. Peñuela, A., Hutton, C., and Pianosi, F., "Assessing the Value of Seasonal Hydrological Forecasts for Improving Water Resource Management: Insights from a Pilot Application in the UK," *Hydrology and Earth System Sciences*, 24, 6059-6073, 2020.
17. World Meteorological Organization (WMO) and World Health Organization (WHO), Use of Climate Predictions to Manage Risks, 2016.
18. 中央氣象局，2020 臺灣氣候分析，https://www.cwb.gov.tw/Data/climate/Watch/twn/twn-monitor_2020-0.pdf。

收稿日期：民國 112 年 07 月 20 日
修改日期：民國 112 年 08 月 25 日
接受日期：民國 112 年 09 月 07 日