JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 4, DECEMBER 2021
DOI - 10.29974/JTAE.202112_67(4).0005

結合非線性主成分分析之神經網路降尺度模式評估臺中與花蓮未來降雨

INTEGRATING NONLINEAR PRINCIPAL COMPONENT ANALYSIS AND NEURAL NETWORKS TO DEVELOP A DOWNSCALING MODEL EVALUATING FUTURE RAINFALL OF TAICHUNG AND HUALIEN

永聯物流開發股份有限公司
產品開發處
副主任
胡 鈞 甯
Jun-Ning Hu

中原大學
土木工程學系
副教授
林 旭 信*
Shiu-Shin Lin

中原大學
土木工程學系
碩士生
朱 楷 洋
Kai-Yang Zhu

摘　要

本研究提出一結合主成份分析(Linear Principal Components Analysis，PCA)、非線性主成分分析(Non-Linear Principal Components Analysis，NLPCA)及輻狀基底函數類神經網路(Radial Basis Function Neural Network，RBFNN) 之非線性主成分分析統計降尺度模式(Nonlinear Principal Component for Genetic Algorithm based- Radial Basis Function Neural Network，NGRB)。首先使用 PCA 及 NLPCA 分析氣象站與 GCM (General Circulation Model)模式資料，其次透過基因演算法(Genetic Algorithm，GA)優化 NGRB 模式參數，最後再以 4 種 GCM 模式的未來情境資料預測臺中與花蓮測站之中期(2050/01 ~ 2069/12)及長期(2080/01 ~ 2099/12)之降雨趨勢。

研究顯示以 NLPCA 較 PCA 能解釋原資料之特徵，並且能有效減少 NGRB 模式運算時間，GA 世代數為 25、族群數為 25 時 NGRB 能合理地收斂；預測 2 個測站中及長期未來降雨趨勢，花蓮測站長期降雨變異性較臺中測站大，兩測站之長期降雨相比中期降雨有減少之趨勢。臺中測站於測試結果顯示中及長期之豐/枯水期約提前 1 ~ 2 個月，然而未來模擬結果並無明顯趨勢。

關鍵詞：非線性主成份分析、輻狀基底函數類神經網路、降雨變異性、氣候變遷。

* 通訊作者，中原大學土木工程學系副教授
32023 桃園市中壢區中北路 200 號，linxx@cycu.edu.tw

INTEGRATING NONLINEAR PRINCIPAL COMPONENT ANALYSIS AND NEURAL NETWORKS TO DEVELOP A DOWNSCALING MODEL EVALUATING FUTURE RAINFALL OF TAICHUNG AND HUALIEN

Jun-Ning Hu
Ally Logistic Property Co., Ltd.

Shiu-Shin Lin*
Department of Civil Engineering,
Chung Yuan Christian University

Kai-Yang Zhu
Department of Civil Engineering,
Chung Yuan Christian University

ABSTRACT

The study combines principal component analysis (PCA), nonlinear principal component analysis (NLPCA), and genetic algorithm (GA) based radial basis function neural network (RBFNN) to develop a statistical downscaling model (NGRB). Firstly, the PCA and NLPCA are used to analyze data from meteorological stations and GCM model. GA is then employed to optimize parameters of NGRB models. Finally, four GCM models outputs from A1B scenario are applied to predict near (2050/01 to 2069/12) and far future (2080/01 to 2099/12) monthly rainfall of Taichung and Hualien stations. The simulated results show that NLPCA can extract features of data better than PCA and can reduce much computation time of NGRB; GA can converge effectively with 25 generations and populations. Predicting near and far future rainfall trends at two stations. Long-term rainfall variability at the Hualien station is greater than that at the Taichung station. The long-term rainfall at the two stations tends to decrease compared to the medium-term rainfall. It reveals that the wet and dry season may happen one to two months ahead of schedule for the near and far future rainfall in Taichung station. The annual average monthly rainfall of Hualien station may small increase for the near future. Rainfall trends for the far future in Hualien station has a gentle bell curve.

Keywords: Nonlinear principal component analysis, Radial basis neural networks, Rainfall variability, Climate change.

Hu, J.N., Lin, S.S.*, & Zhu, K.Y. (2021). "Integrating Nonlinear Principal Component Analysis and Neural Networks to Develop a Downscaling Model Evaluating Future Rainfall of Taichung and Hualien." *Journal of Taiwan Agricultural Engineering,* 67(4), 78-90. https://doi.org/10.29974/JTAE.202112_67(4).0005

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 4, DECEMBER 2021
DOI - 10.29974/JTAE.202112_67(4).0005

一、緒論

根據 2007 年發表的第四次評估報告(AR4)指出暖化現象影響未來降水、大氣、海洋、冰圈等。隨著全球暖化造成天氣型態的改變，降水的型態與溫度的變化也隨之改變，如未來極端降雨事件的增加、降雨強度增加、降雨延時縮短及不降雨的天數增加等現象(陳，2007)。目前未來氣候狀況使用的方法為全球環流模式(General Circulation Model，GCM)，配合預設之情境(Scenarios)進行未來中、長期之氣候推估(Rashid *et al*., 2016；Li *et al*., 2020；Gupta *et al*., 2020)，由於 GCM 所制定的資料網格大小約為 150 ~ 300 公里，且同一網格內以平均值表現氣候資料，對於一區域性的氣候而言，其空間解析度過於粗糙(Wigley and Raper, 1990)。為了使 GCM 模式輸出資料能應用在尺度較小之區域進行模擬推估，發展許多方法將 GCM 所提供的資料進行降尺度，表現出區域氣候特徵。

統計降尺度(Statistical Downscaling)是利用長期的觀測資料(氣象站或雨量站)建立 GCM 模式和區域氣候條件之間的統計關係，應用此統計關係與 GCM 提供之未來氣候模擬資料，輸出未來區域氣候資料。藉由 GCM 模式資料及區域氣象站資料建立模式，透過此模式掌握歷史氣候變化趨勢，再藉由未來 GCM 情境資料輸入此模式之中，模擬未來區域氣候資料，藉以掌握未來氣候變化的趨勢。

轉換函數法為目前較常使用的統計降尺度方法，其中又以分類於非線性轉換法且最具代表性的方法為仿人體神經訊息傳遞的類神經網路(Artificial Neural Networks, ANN)理論，常見的神經網路架構有倒傳遞類神經網路(Back-Propagation Neural Network, BPNN)、輻狀基底函數類神經網路(Radial Basis Function Neural Network, RBFNN)、自聯想式類神經網路(Auto-Associative Neural Networks, AANN)等。若資料具有非線性的特徵，則可考慮使用非線性的分析方法，如非線性主成分分析(Non-Linear Principal Components Analysis, NLPCA)，廣泛被應用在大氣水文領域，以分析資料非線性特徵 (Kramer, 1991；甘等，2017；Canchala *et al*., 2020)。

近年來基因演算法(Genetic Algorithm，GA)廣泛的被使用在調整參數的方面，Li *et al*. (2018)以輻狀基底函數類神經網路結合 GA 針對臺灣探討不同天氣狀況下各流域之流量差別，其結果顯示有最高而穩定的預報技術。為使用 GCM 資料推估區域降雨量之特性，考量機器學習之特性，本研究結合 PCA、NLPCA 與 RBFNN 以建立非線性主成分分析之 GA-RBFNN 統計降尺度模式(Nonlinear Principal Components for GA-Radial Basis Function Neural Network，簡稱為 NGRB 模式)，以同一緯度且位於中央山脈東西側之臺中及花蓮氣象站為例，透過 NGRB 模式推估未來情境的單月總降雨量，分析處於背風面及迎風面之二測站之推估雨量有何趨勢及異同之處。

二、相關理論介紹

2.1 全球環流模式

GCM 模式為三維的數值模型，具有 150 到 300 公里的水平解析度與 10 到 30 層的垂直圖層，其主要概念為將地球系統建構成立體格網，再輸入網格的邊界，以求解動量方程式，如此便能模擬出未來該網格內的氣候資料。GCM 屬於全球尺度(> 104 km)，相較於臺灣集水區水文尺度而言，其空間解析度相對粗糙，而臺灣不同地區的地形各異，降雨型態也十分極端，若以 GCM 結果直接預估臺灣地區未來氣候型態，則無法掌握較細緻的變異。

2.2 非線性主成份分析

NLPCA 利用類神經網路 NN-model 實作以求解非線性主成份，基本架構如圖 1 所示(Hsieh, 2001)。NN-model 屬於基本的前饋式類神經網路，為三層隱藏層之神經網路，介於輸入層 x_0 及輸出層 $x_0^{'}$ 之間的隱藏層依序為編碼層(Encoding Layer)、瓶頸層(Bottleneck Layer)及解碼層(Decoding Layer)，編碼層及解碼層的神經元數目皆以 m 表示，而 m 須小於輸入變數個數。

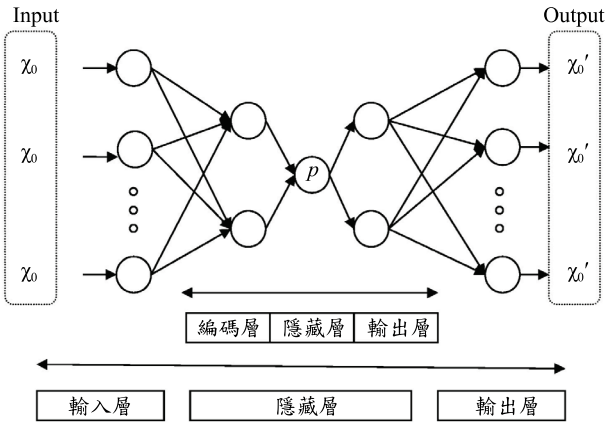

圖 1　非線性主成份分析網路架構圖

圖中 x_0、$x_0^{'}$ 皆為以向量表示。x_0 透過輸入層輸入 NN-model 網路中，經由提取函數(Extraction Function) Φ將資料從高維度的輸入層空間轉換到低維度的瓶頸層空間，如式(1)所示:

$$p = \Phi(x_0) \quad \text{.....(1)}$$

再藉由重建函數(Reconstruction Function)ψ 將資料映射回與輸入層空間同維度的輸出層空間中，如式(2)所示：

$$x_0 = \psi(p) \quad \text{.....(2)}$$

非線性主成份須使用參數收斂才可求得，梯度坡降法(Gradient Decent Method)中的最小均方根誤差演算法(Least Mean Square Error，LMS)廣泛被使用在求解參數最佳化問題上。而 NN-model 透過 LMS 使輸出值與輸入值兩者之間的目標函數(Objection Function) J，亦即均方根誤差(Mean Squared Error，MSE)隨著訓練次數增加而逐漸減少，當 MSE 幾乎不再變化時，則表示模式收斂，目標函數公式如式(3)所示。

$$Min\ J = (\left\| x_0^{'} - x_0 \right\|^2) \quad \text{.....(3)}$$

式中，$\left\| x_0^{'} - x_0 \right\|$ 表示 $x_0^{'}$ 與 x_0 的歐式距離。

對於一般的非線性轉換函數法而言，若輸入資料中含有一部份 noisy，可透過大量的樣本以降低 noisy 所造成的影響，但對於 NN-model 而言，含有 noisy 的資料會使活化函數產生區域極小值，從而導致 overfitting，因此 Hsieh(2004)於活化函數中加入了懲罰項(Penalty)，透過調整 Penalty 可避免 overfitting。Hsieh (2007)將 MSE 及相關係數(Coefficient of Correlation, CC)整合成一個新的指標函數，並稱之為全訊息準則(Holistic Information Criteria, HIC)，如式(4)所示，HIC 越小越表示此網路參數設定越佳。

$$HIC = MSE - CC \times MSE \quad \text{.....(4)}$$

式中，MSE 為均方根誤差、CC 為相關係數。

模式最佳的 m 與 Penalty 之組合為第一非線性主成分，代表著輸入變數中最大的非線性特徵。將原 NN-model 的輸入與輸出之間的殘差做新的 NN-model 的輸入值，重新訓練即可得到對應的第二非線性主成份，以此類推。NN-model 除了可計算 NLPCA 之外，尚可計算 PCA。將活化函數換成線性函數，且隱藏神經元數(m)設定為 1、懲罰項 Penalty 設定為為 0，在目標函數收斂後即可從瓶頸層單一神經元中求得主成份。

2.3 輻狀基底函數類神經網路

RBFNN 優點在於可大量減少學習所需之時間，並可應用於分類及非線性模式之建立，RBFNN 屬於基本的前饋式類神經網路，其架構如圖 2 所示(Elanayar V.T. & Shin, 1994)。其主要的概念是建構許多的輻狀基底函數 $\Phi(\left\| x - c \right\|)$，即隱藏神經元，以函數逼近的方式來建立輸入與輸出之間的映射關係。

當將訓練資料輸入後，直接由輸入層將輸入向量傳給隱藏層中的每個輻狀基底函數，亦即計算輸入向量與隱藏層各神經元中心點的距離後，經函數轉換獲得隱藏層各神經元的輸出如式(5)：

$$z_j(x) = \Phi(\left\| x - c_j \right\|) \qquad j=1, 2, \ldots, m \quad \text{.....(5)}$$

式中，$\Phi(\bullet)$ 表輻狀基底函數、c_j 表隱藏層第 j 個神經元中心點、$\left\| x - c_j \right\|$ 表示 x 與 c_j 間之歐氏距離。

將隱藏層的輸出值經加權傳至輸出層，即可求得網路輸出值如式(6)：

$$y = \sum_{j=0}^{M} w_j \bullet z_j(x) \quad \text{.....(6)}$$

式中，y 為輸出層的輸出值、w_j 為隱藏層第 j 個神經元至輸出層的權重值、z_j 為隱藏層第 j 個神經元的輸出值。

隱藏層中的輻狀基底函數 $\Phi(x-c)$ 有許多種型式。本研究採用一般較廣泛使用的高斯函數當作輻狀基底函數，其型式如式(7)：

$$\Phi(\left\| x - c_j \right\|) = exp\,(-\frac{\left\| x - c_j \right\|^2}{2\sigma^2}) \quad \text{.....(7)}$$

式中 c_j 表示群聚之中心點；σ 代表兩個最近之群聚中心的距離。

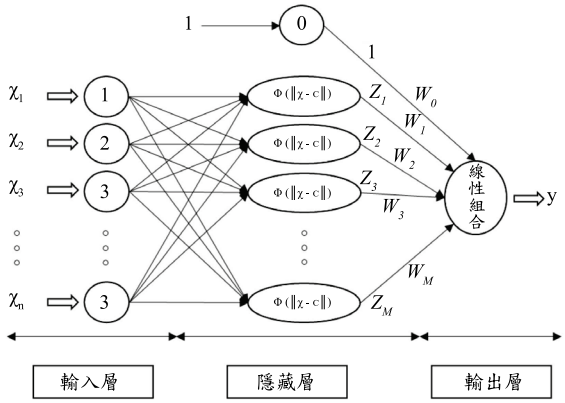

圖 2　輻狀基底函數類神經網路架構圖

RBFNN 架構最重要的部分分別為(1)隱藏層神經元的個數(2)中心點的選取。因隱藏神經元的個數代表了整個網路架構之複雜度，決定中心點的個數代表決定了整個網路架構，且中心點的初始位置對於網路訓練的收斂速度與穩定性有相當大之影響，選擇適當的中心點位置能有效的減少中心點個數，降低網路的複雜度同時可提升網路整體訓練收斂速度。本研究使用的監督式選取法為垂直最小平方法，此方法選取中心點之概念為逐次由輸入資料中，一次挑選一個能讓輸出誤差最小者作為新加入的隱藏層中心點，當達到本研究所設定誤差值時，便停止增加隱藏層中心點個數(Chen *et al*., 1991; Chng *et al*., 1996; Li *et al*., 2020)。當決定好隱藏層中心點個數之後，即可推求隱藏層與輸出層間之最佳權重，並進一步求得網路之輸出值。

2.4 基因演算法

GA 是強調基因型的轉變，將問題的可能解答經過編碼成為基因型式，利用遺傳運算子演化找到最佳解，主要的三步驟為：複製(Reproduction)、交配(Crossover)、突變(Mutation)；複製是指染色體依照適應函數值高低來決定被複製到下一代之機率，適應程度高的染色體有較高的機率被大量複製成為下一代之新染色體，而適應程度低之染色體有較高的機率被淘汰。交配則為基因演算法中最重要的一環，以隨機方式選取兩個不同的染色體進行配對，藉由交換彼此基因位元而產生兩個新的染色體。突變則是為避免近似的兩個染色體經由交配過程產生之下一代過於相似而侷限於某一範圍，利用隨機選取一物種的字串，隨機選取突變點來改變物種字串裡的位元資訊，突變過程發生之機率則依「突變率」設定而改變(陳與廖，2005；Cheng *et al*., 2006)。

三、資料讀取與降尺度模式建立

3.1 資料簡介與分析

本研究以臺中及花蓮二緯度接近且分別位於中央山脈背風面及迎風面之測站做為研究區域，探討此二測站未來逐月降雨之趨勢。使用資料由颱洪中心與 DBAR 大氣資料庫所提供及由 IPCC 網站下載之 4 個 GCM 模式資料。表 1 為臺中與花蓮測站地理位置與所選資料年份，臺中與花蓮二測站之緯度分別為北緯 24°08’50.98”與北緯 23°58’ 37.10”，甚為接近。降雨資料所使用的資料時間及資料長度，由 1950 年 1 月起至 2000 年 12 月迄共 51 年之降雨資料。

AR4 列出了多個全球不同科學機構利用 GCM 所發展的未來氣候推估，於 IPCC 網站(http://www.ipcc-data.org/sres/gcm_data.html)提供下載以供研究者利用。本研究選擇其中的 4 個模式，其模式名稱、縮寫、發展單位(所屬國家)與水平網格解析度分別為 CCCma_CGCM3(T47 resolution)、CGMR、Canadian Center for Climate Modelling and Analysis (Canada)、96(km) × 48(km)；CSIRO_ Mk3.0、CSMK3、Australia's Commonwealth Scientific and Industrial Research Organization (Australia)、192(km) × 96(km)；GFDL_CM2.1、GFCM2.1、Geophysical Fluid Dynamics Laboratory (USA)、144(km) × 90(km)；CGCM2.3.2、MRCGCM、Meteorological Research Institute (Japan)、128(km) × 64(km)。為方便後續資料整理與使用，以 GrADS 將各 GCM 鄰近臺灣之網格點繪輸出，以了解各測站落於哪幾個網格內。以 CGMR 為例，如圖 3 所示，鄰近臺灣的網格數共 6 格，臺中與花蓮測站皆

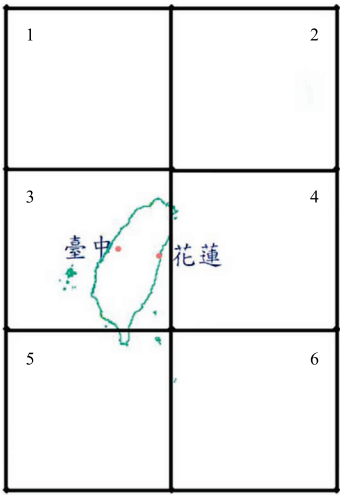

圖 3 CGMR 模式位於臺灣網格及臺中與花蓮測站位置示意

表 1 選用測站資料簡介

測站名稱	測站編號	經度	緯度	海拔高度(m)	資料時間	資料長度(年)
臺中	467490	120°40’ 33.31” E	24°08’50.98” N	16.0	1950/01 ~ 2000/12	51
花蓮	466990	121°36’17.98” E	23°58’ 37.10” N	34.0	1950/01 ~ 2000/12	51

選用第 3 網格資料，其餘模式皆以此方式讀取位於臺灣附近的鄰近網格資料。研究中使用 GCM 模式 1950/1 ~ 2000/12 之月資料，其採用時間與各氣象測站選用時間相同，及 GCM 模式的未來情境 A1B 中期(2050/01 ~ 2069/12)、長期(2080/01 ~ 2099/12)之月平均資料。待選用測站及 GCM 模式資料時間範圍後，將 GCM 模式及測站資料進行正規化，其目的在於消除單位對權重造成的影響，本研究資料正規化範圍為 0 ~ 1。

3.2 NGRB 統計降尺度模式建立

NGRB 模式核心架構為 RBFNN，RBFNN 之建立主要分為輸入變數及參數之選擇。資料處理的方式共分為三種，分別為透過非線性主成份分析處理後之資料，以 NLS 表示；經一般主成份分析處理後之資料，以 LS 表示；未經過任何分析或處理之資料，以 OS 表示。研究地點於模式的名稱中皆以縮寫表示，如臺中以 C 表示、花蓮以 H 表示。本研究依據 3 種輸入 NGRB 模式資料處理的方式、4 種選用的 GCM 模式及 2 種選擇之研究地點的不同而建立模式，以資料處理方式_GCM 模式_測站命名。

3.2.1 預測因子之變數優選

本研究將測站雨量資料定義為預報變量， GCM 資料定義為預測因子，以 PCA 及 NLPCA 兩種方法挑選適合之預測因子做為 NGRB 模式之輸入。各 GCM 模式候選預測因子列於表 2 中，CSMK3、GFCM2.1 及 MRCGCM 模式擁有較多變數，CSMK3 模式無變數 uas、vas 與 mrso；GFCM2.1 及 MRCGCM 無變數 tasmax、taxmin。

表 2　各 GCM 模式大氣水文因子

CGMR			
因子縮寫	氣象因子名稱		單位
huss	surface specific humidity	表面比濕度	(kg/kg)
pr	total precipitation	總降雨量	$(kg/m^2 \cdot s)$
psl	sea level pressure	海平面氣壓	$(pascal)$
rsds	surface downwelling shortwave radiation	表面輻射向下短波	$(watt/m^2)$
tas	near surface air temperature	高度 2 m 平均溫度	$(kelvin)$
uas	zonal surface wind speed	表面緯向風速	(m/s)
vas	meridional surface wind speed	表面經向風速	(m/s)
CSMK3、GFCM2.1、MRCGCM			
因子縮寫	氣象因子名稱		
hur200	relative humidity at 200 hPa	200 hPa 相對溼度	$\left(\frac{g/cm^3}{g/cm^3}\right)$
hur500	relative humidity at 500 hPa	50 hPa 相對溼度	$\left(\frac{g/cm^3}{g/cm^3}\right)$
hur850	relative humidity at 850 hPa	850 hPa 相對溼度	$\left(\frac{g/cm^3}{g/cm^3}\right)$
prc	convective precipitation	對流降雨量	$(kg/m^2 \cdot s)$
rsds	surface downwelling shortwave radiation	表面輻射	$(watt/m^2)$
ta200	air temperature at 200 hPa	200 hPa 平均溫度	$(kelvin)$
ta500	air temperature at 500 hPa	500 hPa 平均溫度	$(kelvin)$
ta8500	air temperature at 850 hPa	850 hPa 平均溫度	$(kelvin)$
ts	surface skin temperature/SST	海表面溫度	$(kelvin)$
tas	near surface air temperature	高度 2 m 平均溫度	$(kelvin)$
ua200	zonal wind component at 200 hPa	200 hPa 緯向風速	(m/s)

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 4, DECEMBER 2021
DOI - 10.29974/JTAE.202112_67(4).0005

表 2　各 GCM 模式大氣水文因子

因子縮寫	氣象因子名稱	單位	
ua500	zonal wind component at 500 hPa	500 hPa 緯向風速	(m/s)
ua850	zonal wind component at 850 hPa	850 hPa 緯向風速	(m/s)
va200	meridional wind component at 200 hPa	200 hPa 經向風速	(m/s)
va500	meridional wind component at 200 hPa	500 hPa 經向風速	(m/s)
va850	meridional wind component at 200 hPa	850 hPa 經向風速	(m/s)
pr	total precipitation	總降雨量	$(kg/m^2 \cdot s)$
psl	sea level pressure	海平面氣壓	$(pascal)$
zg200	geopotential height at 200 hPa	200 hPa 重力位高度	(m)
zg500	geopotential height at 500 hPa	500 hPa 重力位高度	(m)
zg850	geopotential height at 850 hPa	850 hPa 重力位高度	(m)
tasmax	2m mean max. air temperature	高度 2m 最高溫度	$(kelvin)$
tasmin	2m mean min. air temperature	高度 2m 最低溫度	$(kelvin)$
uas	zonal surface wind speed	表面緯向風速	(m/s)
vas	meridional surface wind speed	表面經向風速	(m/s)
mrso	total soil moisture content	土壤含水量	(g/g)

表 3　LS 與 NLS 的 MRCGCM 模式氣候因子篩選結果

GCM 模式	主成分	Penalty	m	HIC	Var %	累計 Var%
LS_MRCGCM	主成分 1	0	1	0.1221	51.931	51.931
	主成分 2	0	1	0.52866	16.681	70.627
	主成分 3	0	1	1.6058	10.240	80.867
NLS_MRCGCM	主成分	Penalty	m	HIC	Var %	累計 Var%
	主成分 1	0.1	4	0.09308	51.931	51.931
	主成分 2	0.1	4	0.47047	21.681	73.612
	主成分 3	0.0001	3	1.5673	13.240	86.852

以 MRCGCM 模式為例，表 3 為 MRCGCM 模式氣候因子篩選結果。NN-model 挑選氣候因子以模式穩定為最優先，因此挑選評估指標 HIC 最小之參數組合及其對應之非線性主成分。利用原 NN-model 之輸入值與輸出值之間的殘差輸入至新的 NN-model 中，並且使網路重新訓練，再藉由評估指標 HIC 求得輸入變數相同個數之主成份及非線性主成份。從表 3 可看出相同主成份個數之 Var%，非線性主成份之 Var%略大於主成份之 Var%，表示 MRCGCM 模式為具有非線性特徵之資料，並且表示 NLPCA 比 PCA 更能解釋原資料之變異性。

3.2.2 NGRB 統計降尺度模式參數優選

NGRB 模式主要的核心為類神經網路 RBFNN，本研究使用 Matlab 中的 newrb 函式來建構，所需調整之參數有二，為 GOAL 及 SRPEAD。隱藏層神經元間之歐式距離對於 SRPEAD 參數調整扮演重要的角色，應盡量能涵蓋左右隱藏神經元且含蓋範圍不超過 2 個隱藏神經元為佳。

本研究使用基因演算法優化網路參數，由於使用 GA 優化參數時必須給定參數範圍，故須先決定上述兩個參數之範圍，最後以基因演算法優化參數。參數範圍選擇之概念為透過 GOAL 與 SPREAD 的組合找到其對應的隱藏層神經元所生成的個數，其個數最小不得低於輸入變數的個數，最大不得超過輸入變數的資料筆數。根據臺中測站及花蓮測站之參數訓練結果顯示，GOAL 上限分別約為 2.0 及 0.008。不同資料處理方式以 OS 之 SPREAD 較高，輸入模式之變數個數較多會影響 SPREAD 之上限，二測站之 OS 模式，其 SPREAD 在 4.0 ~ 5.0 之間，而 NLS 與 LS 則在 0.5 ~ 3.0 之間。至於不同輸入資料由於經 PCA 與 NLPCA 處理後其變數在 2 ~ 3 之間，則無明顯之特性。以下簡述 NGRB 模式選取參數範圍之步驟：

1. 將 GCM 模式及測站資料以 4:1 的方式切割為建模資料及測試資料，再將建模資料以 3:1 的比例切割成訓練資料與驗證資料，並且只使用訓練資料進行

NGRB 參數範圍測試，其中 Test_1 與 Test_2 分別為 LS 與 NLS 模式的第一組和第二組測試資料。

2. Test_1 假設 GOAL 最小為 0，最大為 1，SRPEAD 範圍則給定 0.1～5.0，測試過程中的 GOAL 範圍由 1～0 遞減。當 GOAL 每減少 0.001，則 SRPEAD 由 0.1 增加到 5.0，而每一個 GOAL 及 SPREAD 之組合皆會對應一個神經元數目；Test_2 則依據 GOAL 與其對應神經元數，選擇神經元數有變動之範圍，將 GOAL 範圍縮小，而 SPREAD 變動範圍一樣由 0.1 增加到 5.0。
3. 決定 GOAL 上限值：GOAL 由 1 遞減至 0，假設有一數值 L，只要 GOAL 小於數值 L，其對應之神經元個數會大於輸入變數之個數。而 GOAL 在 1 至數值 L 之間，其對應之神經元數會等於輸入變數之個數，則此數值 L 為 GOAL 之上限值。
4. 決定 GOAL 下限值值：GOAL 由 1 遞減至 0，假設有一數值 D，GOAL 在數值 D 至 0 之間其對應之神經元個數會等於輸入變數之個數，則此數值 D 為 GOAL 之下限值。
5. OS 模式之變數個數皆大於 NLS 模式及 LS 模式，輸入變數增加則導致模式複雜度上升，為了避免 OS 因模式複雜度而導致模式在設定較低的世代數與族群數之值無收斂，因此以 OS_MRCGCM_C 模式進行測試。待所有 NGRB 模式之 GOAL 與 SPREAD 範圍測試完成後，使用 GA 優化參數，本研究將測試 GA 之資料組命名為 Test_3。以 Test_3 測試世代數(NG)、族群數(PG)之值，分別測試 { NG,PG }= { {3,3}、{25,25}、{50,50} } 之三種組合之模式運算時間與均方根誤差(Root Mean Squared Error，RMSE)。式(8)為 RMSE 計算式，式中 x_i 與 y_i 分別為預測值與觀測值。

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)^2} \quad \text{.....(8)}$$

GA 挑選世代數及族群數之測試，NG 為 3、PG 為 3 之組合其 RMSE 範圍於 36.23 ~ 174.64 (mm)、運算之平均時間為 35(s)；NG 為 25、PG 為 25 之組合其 RMSE 範圍為 10.10～21.83 (mm)、運算之平均時間為 850(s)；NG 為 50、PG 為 50 之組合其 RMSE 範圍 9.7～11.67 (mm)、運算之平均時間為 5647(s)，其誤差雖然為三組世代數、族群數中最低，而其模式運算時間遠大於另外兩組參數。而 NG 為 25、PG 為 25 之組合其誤差已控制在小範圍中， 所需要的模式運算時間雖然比 NG 為 3、PG 為 3 之組合高，但尚在可接受的範圍之中，因此以世代數、族群數為 NG 為 25、PG 為 25 之組合，做為 NGRB 模式的 GA 參數之設定。

3.3 模式效能評估

待預測因子、NGRB 網路參數範圍及 GA 世代數、族群數確定後，即可建立 NGRB 模式。如圖 4 所示，NGRB 模式主要分為訓練及測試兩大部分，透過訓練流程及測試流程建立模式後，再以 RMSE 評估測試資料之效能。

以 MRCGCM 模式為例，表 4 為臺中 NGRB 模式建立之基本資料，模式輸入項依 LS、NLS、OS 分別為主成分第 1 及第 2 項、非線性主成分第 1 及第 2 項、7 個原始變數。而三個模式之模式輸出項、雨量／建模／測試資料時間長度與單位皆一致，分別為臺中測站單月總降雨、1950/01～2000/12 (共 51 年)、1950/01 ～ 1990/10 (共 490 筆)、1990/11 ~ 2000/12 (共 122 筆)、以月為單位(共 612 筆資料)。RBFNN 參數設定時，三個模式之 Goal 參數範圍分別為 0.005 ~ 0.020、0.005 ~

表 4　臺中 NGRB 模式建立資料及參數簡介

	LS_MRCGCM_C	NLS_MRCGCM_C	OS_MRCGCM_C
Input	主成分 1 主成份 2 主成份 3	非線性主成分 1 非線性主成份 2 非線性主成分 3	25 個原始變數
Output	臺中測站單月總降雨		
雨量資料時間長度	1950/01~2000/12 (共 51 年)		
建模資料時間長度	1950/01~1990/10 (共 490 筆)		
測試資料時間長度	1990/11~2000/12 (共 122 筆)		
單位	以月為單位(共 612 筆資料)		
	RBFNN 參數設定		
GOAL 參數範圍	0.001~0.025	0.003~0.022	0.0001~0.020
SPREAD 參數範圍	0.1~1.7	0.1~0.5	0.1~5.0
世代數	25	25	25
族群數	25	25	25

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 4, DECEMBER 2021
DOI - 10.29974/JTAE.202112_67(4).0005

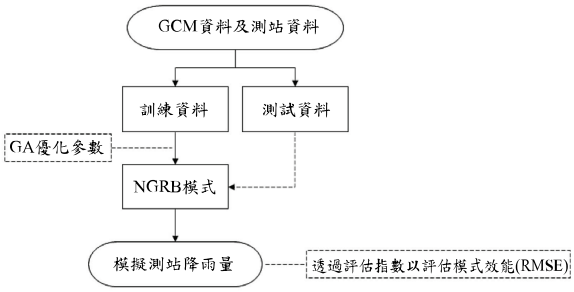

圖 4　NGRB 模式建立步驟

0.020 與 0.005 ~ 0.020；Spread 參數範圍分別為 0.1 ~ 0.9、0.1 ~ 0.9、0.1 ~ 4.0；世代數與族群數一致皆為 25。對於 GOAL 與 SPREAD 而言，因為以相同的測站作為研究地點差異不大，三個模式最大的差異點在於以變數個數，OS_MRCGCM_C 模式使用 25 個變數，但 LS_ MRCGCM _C 及 NLS_ MRCGCM _C 模式只使用 3 個輸入變數，或許可對於輸入變數之個數進行調整。

表 5 為臺中測站與花蓮測站 LS、NLS、OS 之 MRCGCM_C 模式測試評估指標模擬結果。臺中測站與花蓮測站三個模式分別的 RMSE 相近。然而 OS_MRCGCM_C 模式之運算時間大於 NLS_MRCGCM_C 及 LS_MRCGCM_C 模式。由於輸入變數個數影響模式複雜度導致計算時間增加，OS_CGMR_C 模式之輸入變數共有 25 個，而 NLS_MRCGCM_C 及 LS_MRCGCM_C 模式之輸入變數只有 3 個。

圖 5 與圖 6 分別為臺中測站與花蓮測站於不同情

表 5　臺中測站與花蓮測站 LS、NLS、OS 之 MRCGCM 模式測試評估指標

臺中測站_單月累積降雨			
模式	LS_MRCGCM_C	NLS_ MRCGCM _C	OS_ MRCGCM _C
RMSE(mm)	148.03	166.18	136.84
時間(s)	579.76	459.04	803.33
花蓮測站_單月累積降雨			
模式	LS_MRCGCM_H	NLS_ MRCGCM _H	OS_ MRCGCM _H
RMSE(mm)	188.46	208.46	181.47
時間(s)	478.73	636.63	1223.11

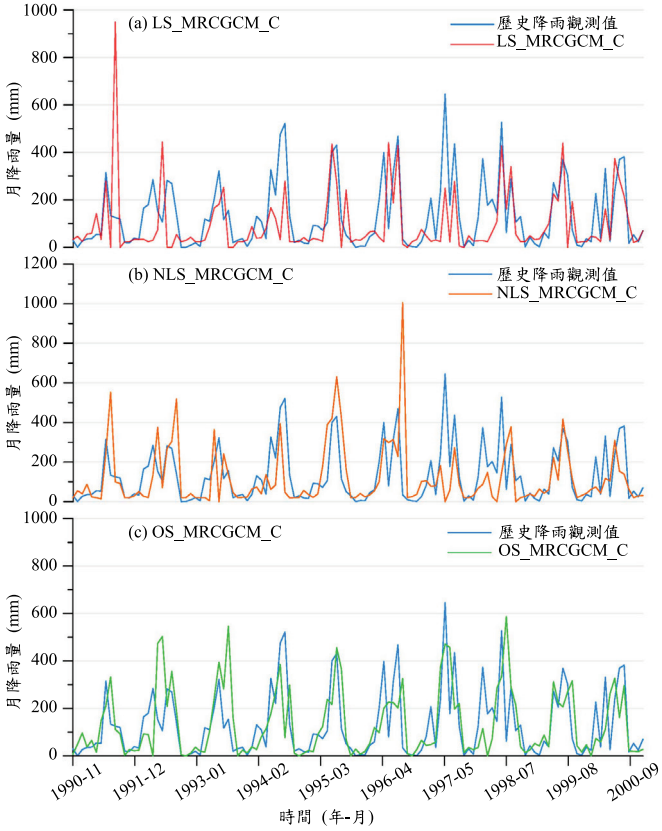

圖 5　MRCGCM_C 於不同資料處理方式：(a) LS；(b)NLS；(c) OS 測試結果

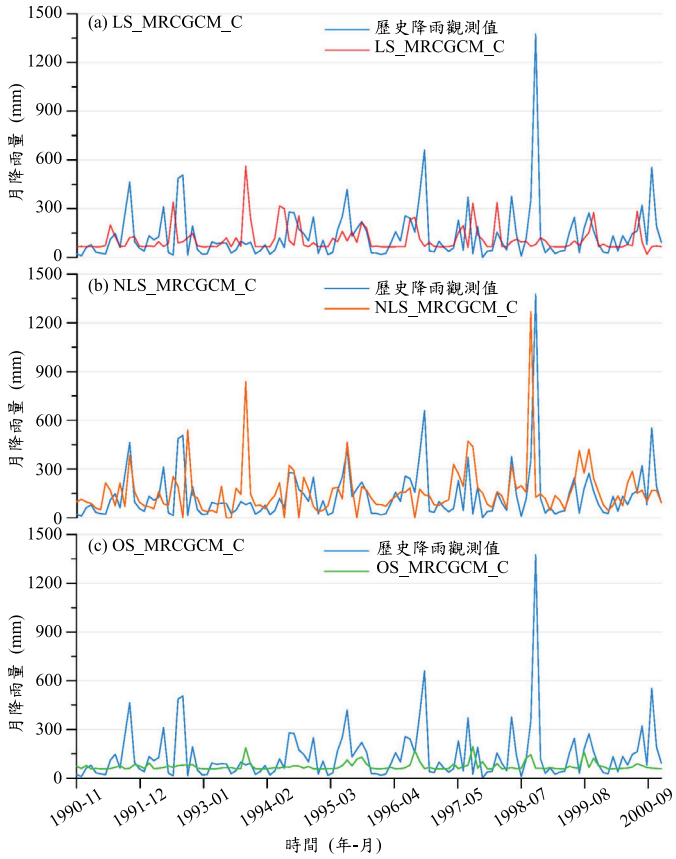

圖 6　MRCGCM_H 於不同資料處理方式：(a) LS；(b) NLS；(c) OS 測試結果

境之未來雨量趨勢預測結果。研究顯示臺中測站以 OS_MRCGCM_C 模式之月份輸出值與歷史降雨觀測值較相近；花蓮測站顯示 NLS_MRCGCM_H 模式除各別高估值外，其餘月份輸出值與歷史降雨觀測值較相近。

四、結果與討論

臺中測站之中期及長期預測只有 OS_MRCGCM_C 較接近歷史觀測值之趨勢，故採用該模式進行臺中未來監督式選取法情境之分析；花蓮測站之中期與長期預測，NLS_CGMR_H 與 NLS_MRCGCM_H 表現較佳，故採用此二模式作為未來情境之分析。圖 7 為臺中與花蓮在中、長期情境的月降雨量統計比較，十字標記為降雨量上限值，方形標記為降雨平均值，三角形標記為降雨下限值。由圖可得知臺中在未來平均月降雨量及最大月降雨量均有減少之趨勢，平均月降雨量與歷史觀測平均值差異不大；花蓮測站在未來月降雨量及最大月降雨量亦有減少之趨勢，其中平均月降雨量低於歷史觀測平均值，最大月降雨量則減少了約 400 毫米的雨量。由圖 7 結果可得知花蓮未來降雨相較於臺中有更大的變異性。

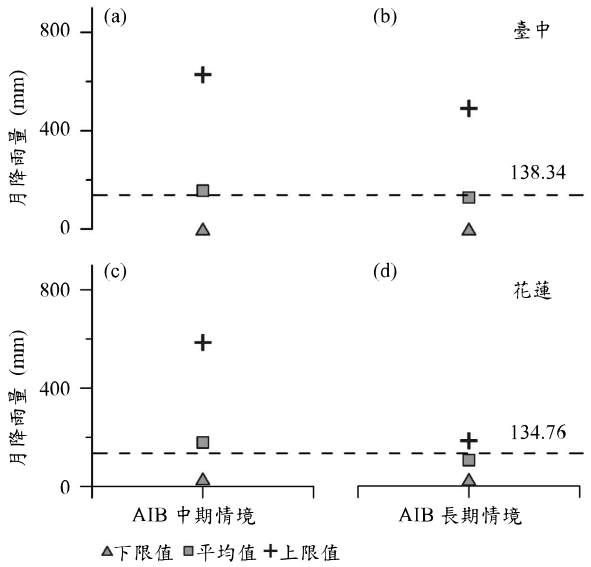

圖 7　臺中與花蓮中期/長期預測值比較

圖 8 與圖 9 分別為臺中與花蓮中期(2050/1 ~ 2069/12)、長期(2080/1 ~ 2099/12)的月平均降雨量距平圖，圓形標記為冬季(12 月 ~ 2 月)，十字標記為夏季(6 月 ~ 8 月)。由圖 8(a)臺中中期夏季、冬季的雨量距平值可得知，夏季雨量自 2050 年起至 2069 年普遍高

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 4, DECEMBER 2021
DOI - 10.29974/JTAE.202112_67(4).0005

於歷史平均，然而自 2057 年起的雨量有減少之趨勢。冬季雨量於中期降雨預測自 2057 年起的雨量亦有減少之趨勢，且多數低於歷史平均。圖 8(b)花蓮中期夏季雨量除 2050 年、2057 年與 2059 年的夏季雨量明顯高於歷史平均外，其餘年份的夏季雨量皆與歷史平均差異不大，冬季雨量有兩個明顯增加趨勢出現在 2050 年至 2057 年與 2058 年至 2067 年，2056 年與 2067 年出現高於歷史平均約 400 毫米之雨量。圖 9(a)臺中長期夏季雨量距平值最高點出現在 2087 年的夏季，有高於歷史平均約 600 毫米的雨量。冬季雨量無明顯趨勢，自 2080 年至 2099 年的雨量普遍低於歷史平均。由圖 9(b)花蓮長期夏季雨量距平值可得知，除 2087 年的雨量明顯大於歷史平均外，其餘年份皆與歷史雨量差異甚小。冬季雨量除 2086 年高於歷史平均外，其餘年份普遍小於歷史平均 0～200 毫米不等。

圖 10 與圖 11 為臺中與花蓮中/長期降雨預測，箱型圖為模式平均值，黑色實線為兩測站歷史降雨觀測值。由圖 10(a)與圖 11(a)可知臺中枯水期(11～4 月)月雨量有減少趨勢，豐水期(5～10 月)降雨不確定性高；圖 10(b)與圖 11(b)為花蓮中期與長期雨量預測，中期月平均雨量大於歷史降雨觀測值，長期冬季雨量低於歷史降雨觀測值且相對於其他月份有較高之不確定性，長期雨量預測結果與 Li *et al*. (2020)研究預測之花蓮長期降雨有相似趨勢。綜合上述，臺中中期與長期豐水期有較高不確定性，枯水期雨量低於歷史平均；花蓮中期豐水期雨量大於歷史降雨觀測值，枯水期相較於豐水期有較高不確定性。長期豐水期雨量大於歷史降雨觀測值且不確定性較小，枯水期與豐水期有相反結果，雨量低於歷史降雨觀測值且不確定相對大。

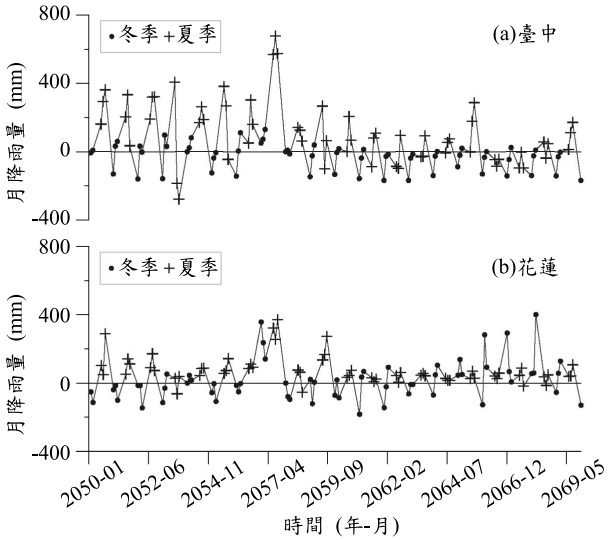

圖 8　臺中與花蓮中期的月降雨量距平圖

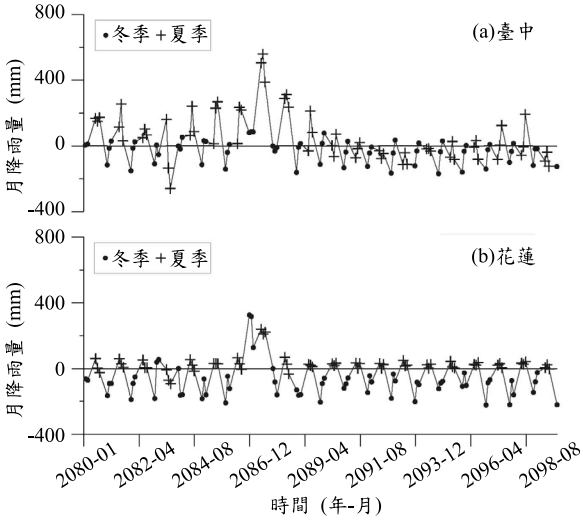

圖 9　臺中與花蓮長期的月降雨量距平圖

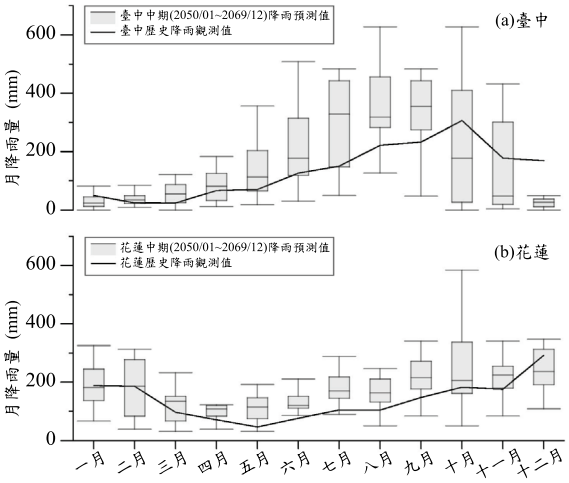

圖 10　臺中與花蓮中期月降雨量預測

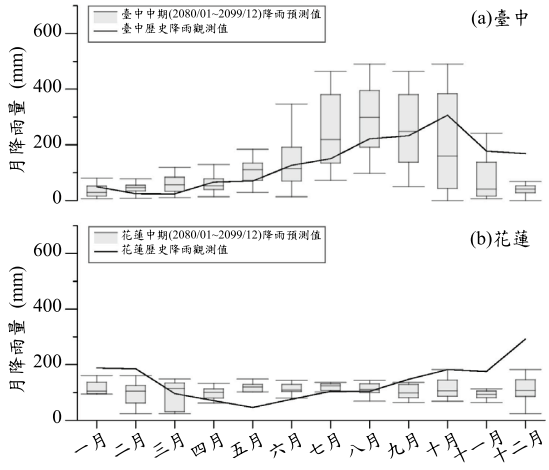

圖 11　臺中與花蓮長期月降雨量預測

五、結論

本研究以 RBFNN 為核心，利用 GA 挑選網路參數，及使用 PCA 及 NLPCA 對 CGMR、CSMK3、GFCM2.1、MRCGCM 等四個 GCM 模式進行資料特性的萃取，以建立 NGRB 統計降尺度模式預測臺中及花蓮測站之單月總降雨。經案例研究分析，得到具體結論如下:

1. 非線性主成分的累計變異數百分比大於線性主成份。對 CGMR 而言，LS 與 NLS 於第二主成份的累積變異數百分比分別為 80.627 及 83.723；對 GFCM2.1 而言，LS 與 NLS 於第三主成份的累積變異數百分比分別為 79.63 及 82.589；對 CSMK3 而言，LS 於第二主成份的累積變異數百分比為 83.564，NLS 於第三主成份的累積變異數百分比為 82.907；以 MRCGCM 而言，LS 模式於第三主成份的累積變異數百分比為 80.867 及 86.852。
2. NGRB 模式參數範圍測試結果顯示：相同研究地點之 GOAL 上限值相近，而輸入變數個數會影響 SPREAD 之下限大小，若輸入變數越多則 SPREAD 下限值越高。臺中測站 GOAL 上限約為 2.0、花蓮測站 GOAL 上限約為 0.008；使用原始資料(OS)做為輸入值之模式其 SPREAD 範圍約在 3.0 ~ 5.0 之間，而使用 LS、NLS 作為輸入變數之模式其 SPREAD 範圍約在 0.5 ~ 2.1 之間。
3. 由 GA 挑選世代數與族群數測試結果可得，世代數與族群數各為 25 的組合為最適合且能合理收斂，其 RMSE 範圍為 10.10 ~ 21.83(mm)，運算之平均時間為 850(s)。
4. LS_MRCGCM_H 模式其輸入變數個數為 3 個主成份、累積變異數達 80.867、RMSE 為 188.46 (mm)；NLS_ MRCGCM _H 模式其輸入變數個數為 3 個非線性主成份、累積變異數達 86.852、RMSE 為 208.46 (mm)。相同主成份個數的情況下，非線性主成份之累積變異數大於主成份，做為輸入值其模式預測結果也優於使用主成份做為輸入值之模式，花蓮測站之結果亦同，於臺中測站之結果則相反，LS_ MRCGCM_H 模式優於 NLS_ MRCGCM_H 模式。
5. 臺中中期與夏季雨量有減少趨勢，冬季雨量普遍低於歷史平均；花蓮中期夏季雨量 2050~2059 年有較大變異性，2060 ~ 2069 與歷史平均差異不大。中期冬季雨量前 10 年與後 10 年各有一增加趨勢。長期夏季雨量除了 2087 年有明顯增加外，其餘年份與歷史平均差異不大。長期冬季雨量普遍低於歷史平均 0 ~ 200 毫米不等。
6. 臺中測站中期與長期豐水期相較於枯水期變異性較高；花蓮測站中期降雨枯水期相較豐水期變異性高。兩測站長期降雨相比中期有減少之趨勢。
7. 臺中測站於測試結果顯示豐/枯水期有平移 1 ~ 2 個月之現象，但未來模擬結果未能看出明顯趨勢。

誌謝

本研究承蒙科技部計畫補助計畫編號(MOST 108-2625-M-033-001-)得以完成，特此致謝。

參考文獻

1. Canchala, T., Alfonso-Morales, W., Cerón, W. L., Carvajal-Escobar, Y., & Caicedo-Bravo, E. (2020). “Teleconnections between Monthly Rainfall Variability and Large-Scale Climate Indices in Southwestern Colombia.” *Water*, *12*(7), 1863-1863. https://doi.org/10.3390/w12071863
2. Chen, S., Cowan, C. f. n., & Grant, P. m. (1991). “Orthogonal least squares learning algorithm for radial basis function networks.” *IEEE Transactions on Neural Networks, 2*(2)*,* 302-309. https://doi.org/10.1109/72.80341
3. Cheng, C. T., Zhao, M. Y., Chau, K. W., & Wu, X. Y. (2006). “Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure.” *Journal of Hydrology*, *316*(1-4), 129-140.
4. Elanayar V. T., S., & Shin, Y. C. (1994). “Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems.” *IEEE Transactions on Neural Networks*, *5*(4), 594-603. https://doi.org/10.1109/72.298229
5. Chng, E. S., Yang, H. h., & Bos, S. (1996). “Orthogonal least-squares learning algorithm with local adaptation process for the radial basis function networks.” *IEEE Signal Processing Letters*, *3*(8), 253-255. https://doi.org/10.1109/97.511811
6. Gupta, V., Singh, V., & Jain, M. K. (2020). “Assessment of precipitation extremes in India during the 21st century under SSP1-1.9 mitigation scenarios of CMIP6 GCMs.” *Journal of Hydrology*, *590*, 125422.

https://doi.org/10.1016/j.jhydrol.2020.125422

7. Hsieh, W. W. (2001). "Nonlinear principal component analysis by neural networks." *Tellus*, *53*, 599-615. https://doi.org/10.1034/j.1600-0870.2001.00251.x
8. Hsieh, W. W. (2007). "Nonlinear principal component analysis of noisy data." *Neural Networks*, 4, 434.
9. Hsieh, W. W. (2004). "Nonlinear multivariate and time series analysis by neural network methods." *Reviews of Geophysics*, *42*(1).
10. Kramer, M. A. (1991). "Nonlinear principal component analysis using autoassociative neural networks." *AIChE Journal*, *37*(2), 233-243.
11. Li, C. Y., Lin, S. S., Chuang, C. M., & Hu, Y. L. (2020). "Assessing future rainfall uncertainties of climate change in Taiwan with a bootstrapped neural network-based downscaling model." *Water and Environment Journal*, *34*(1), 77-92. https://doi.org/10.1111/wej.12443
12. Li, C. Y., Lin, S. S., Lin, Y. F., & Kan, P. S. (2018). "A bootstrap regional model for assessing the long-term impacts of climate change on river discharge." *International Journal of Hydrology Science and Technology*, *9*(1), 84-108. https://doi.org/10.1504/IJHST.2019.096802
13. Rashid, M. M., Beecham, S., & Chowdhury, R. (2016). "Statistical downscaling of rainfall: A non-stationary and multi-resolution approach." *Theoretical & Applied Climatology*, *124*(3-4), 919-933. https://doi.org/10.1007/s00704-015-1465-3
14. Sedki, A., Ouazar, D., & El Mazoudi, E. (2009). "Evolving neural network using real coded genetic algorithm for daily rainfall–runoff forecasting." *Expert Systems with Applications*, *36*(3, Part 1), 4523-4527. https://doi.org/10.1016/j.eswa.2008.05.024
15. Wigley, T. M., & Raper, S. C. (1990). "Natural variability of the climate system and detection of the greenhouse effect." *Nature*, *344*(6264), 324.
16. 甘秉玄、林旭信、胡衍立與宋政樺，2017，結合非線性主成份分析與 GA-FUZZY 之颱風降雨預測模式，*臺灣水利*，*65*(4)，42–53。
17. 陳正達，2007，IPCC 第一工作分組之第四次評估報告:科學基礎 第十章 "全球氣候變遷的推估" 重點說明，*全球變遷通訊雜誌*，*55*，27–31.
18. 陳惠菁與廖宜恩，2005，結合自適應共振理論網路二代與基因演算法之分類器，第十屆人工智慧與應用研討會(TAAI)。

收稿日期：民國 110 年 03 月 23 日
修改日期：民國 110 年 06 月 10 日
接受日期：民國 110 年 08 月 11 日