

專 論

結合自組織映射圖網路與支撐向量機於颱風  
期間水庫入流量預測之研究

**Reservoir Inflow Forecasting During Typhoon Periods  
by Combining Self-Organizing Map with Support  
Vector Regression**

國立臺灣大學  
土木工程學系  
博士班研究生

王宗惇\*

Tsung-Chun Wang

台灣首府大學  
休閒管理學系  
副教授

陳儒賢

Lu-Hsien Chen

摘 要

臺灣由於地狹人稠，水資源一直都仰賴水庫供給，而水庫每年主要水量均是來自梅雨、颱風等降雨集中事件，尤其颱風降雨為在短延時中動輒降下數百釐米雨量，對於水庫儲水雖有其利，亦可能因庫容難以負荷，反而造成災害，所以在即使是短時間的入流預測，對於水庫入流及放流的掌握也是極其重要。本研究以翡翠水庫集水區入流量為例，進行 25 場颱風降雨事件短延時水庫入流量預測。為能加強預測能力，運用非監督式的自組織映射圖網路(Self-Organizing Map, SOM)先行針對已降下的前期雨量與水庫入流量，以不同水庫入流歷程進行分類後，再利用支撐向量機(Support Vector Machines, SVM)於迴歸能力上的優勢，進行未來 1 到 3 小時水庫入流量的預測。結果顯示本研究所建立，先行經過分類後再行預測的 SOM-SVM 模式，不僅在短延時確實造成 25 場事件整體均方根誤差均下降，預測效率則使 25 場事件有 24 場都達到 0.9 以上。而在 1 到 3 小時平均效率係數則分別提昇 17.5%、22.7% 及 23.2%。研究成果有助於未來颱風期間水庫入流量之預測。

**關鍵詞：**水庫入流量，颱風降雨，自組織映射圖網路，支撐向量機。

ABSTRACT

The reservoir inflow forecasting during typhoon periods is always an important task

\*通訊作者，國立臺灣大學土木工程學系博士班研究生，26050 宜蘭市民權新路四號，I600060@wra01.gov.tw

for water resources management and disaster mitigation. For improving the short lead-time flood forecasting performance, a reservoir inflow forecasting model is proposed in this paper. The observed inflows of Feitsui reservoir for 25 typhoon events are adopted. Firstly, the Self-Organizing Map (SOM) is used to classify the data into different regions relative for different inflow process. After the inputs are arranged into different regions, the Support Vector Regression (SVR) is implemented for determining the 1- to 3-h ahead inflow forecasts. The results show that the performance of the proposed model can provide improved forecasts of hourly inflows. In conclusion, the proposed model, which considers higher related inputs instead of all inputs, can generate better forecasts in different clusters. This proposed model is recommended as an alternative to the original SVR model because of its accuracy and efficiency.

**Keywords:** Reservoir inflows, Typhoon rainfall, Self-organizing map, Support vector machine.

## 一、緒 論

同為熱帶氣旋的旋風、颱風及颶風，雖是同樣的氣候現象，隨著在地球上不同地區才有著不同的稱呼。它們均是環繞著低壓氣旋，並挾帶著強風的天氣系統，在北半球為逆時針旋轉，南半球則為順時針。臺灣位在北太平洋西側，為颱風的主要路徑之一，平均每年會受到 3 至 5 個颱風的侵襲。只是，由於隨著臺灣地形的複雜性，導致了颱風降雨在時間與空間上的不平均，猛烈的降雨頻繁地產生了像是洪水、山崩及土石流等災害。但隨著颱風遠離臺灣，因颱風降雨所挾帶的大量雨水反而成了必須被儲存的重要水資源。因此水庫實為面對臺灣地區不均勻降雨條件下，最重要，也是最有效的儲水設施。這使得水庫入流量的預測工作在水庫的調節與容量管理，扮演了一個相當重要的角色。因為合理、正確的入流量預測，對於水庫容量來說是個重要的參考依據。因此，強化入流量預測的正確性一直是一個重要的課題。

在從颱風降雨到匯流形成水庫入流量過程，由於其極端的複雜度及高度非線性，對於以物理性水文模式進行建構其過程，一直是個相當不易的課題。作為傳統物理性水文模式的取代，

以人工智慧為理論基礎，好比類神經網路就是個受到矚目且合適的數值學習模式，可以被應用到水庫入流的預測上(Coulibaly *et al.*, 2000, Lin and Chen, 2004; Chen and Chang, 2009; Lin *et al.*, 2009a; Lin *et al.*, 2010; Tserng *et al.*, 2011; Lin *et al.*, 2013a, 2013b, 2013c; Wu *et al.* 2014)。

支撐向量機(SVM)則是一個較新式、以人工智慧為概念為基礎的數值方法，是由 Vapnik 和其工作團隊所以統計理論基礎所發展出一種針對資料分類、再進而發展到迴歸與圖形辨識的學習工具(1998, 1999)。支撐向量機是以「結構風險最小化(structural risk minimization)」為機器學習的主軸，理論性地將機器學習預期誤差最小化，並降低了學習過程中過度擬合的問題。此外，其 SRM 的結構主軸由於是以規劃求解，可說是保證可獲取全域最佳化的惟一解，其詳細的求解內容則可從各類教導 SVM 的書籍進行參考(Vapnik, 1995, Cristianini and Shaw-Taylor, 2000)

支撐向量機作為一個數值學習工具，因為有許多出色的表現，被拿來應用在許多的領域。(Bovolo *et al.*, 2008; Hiew *et al.*, 2009; Song and Wyrwicz, 2009; Baylar and Batan, 2010; Munoz and Gonzalez, 2010) 而水文問題的領域裡，則以近幾年應用在時間序列的表現上為最多(Liong

and Sivapragasam, 2002; Yu *et al.*, 2004; Sivapragasam and Liong, 2005; Yu *et al.*, 2006; Yu and Liong, 2007; Pan and Wang, 2004; Chen and Yu, 2007; Lin *et al.*, 2009b, c; Lin *et al.*, 2010; Chen, Yang *et al.*, 2011)。

雖然支撐向量機在迴歸問題上，是種傑出的數值學習工具，只是當在迴歸問題上碰到太多無用或是低相關性的數值向量，它的迴歸能力也會跟著降低。是以在本研究的水文問題中，它也可能出現使用過多低物理值去迴歸產生高物理值的情形。再者，對於長延時預測結果，當支撐向量機只能在取得數量有限且均含高度雜訊的資料組，它也就必然無法有效地得出本研究的水庫入流量預測成果。

為了解決上述的低迴歸能力問題，自組織映射圖(Self-Organizing Map, SOM)在本研究中被用來預先針對資料作出有意義的群組分類。在各個分類群組中，輸入項被分類到所涉及颱風兩事件中的不同入流過程中。也因此，由支撐向量機所得出的預測結果可以因為只考量同一入流情形群的群組，得到較正確的值。

SOM 是由 Kohonen (1990, 1995)所提出，它是個類神經網路較特別的範疇，能夠將高度非線性的輸入值投射到較低維的拓撲圖像上，令高維度資料群組的情形可以被觀察出來。此一能力在近幾年的使用(Hsu *et al.*, 2002; Moradkhani *et al.*, 2004; Lin and Chen, 2005; Yang *et al.*, 2005; Samsonova *et al.*, 2006; Srinivasulu and Jain, 2006; Lin and Wang, 2006; Lin and Wu, 2007; Alvarez-Guerra *et al.*, 2008; Muller, 2009; Chang *et al.*, 2010)，SOM 的分類表現較都優於常見的分類方法，使得人們可以從複雜的資料中發現其中的關聯性。

本論文為增加水庫入流量預測能力，先由 SOM 進行分類後，再由 SVM 進行迴歸預測。首先，先由過去 2 小時的雨量及流量作為資訊，對不同事件及延時以 SOM 來進行分類。再用 SVM 對被 SOM 分類後的資料進行預測。最後，本文所建議的方法被應用在翡翠水庫來預測 1 到 3 小時的入流量，本研究所建立模式圖詳如圖 1。

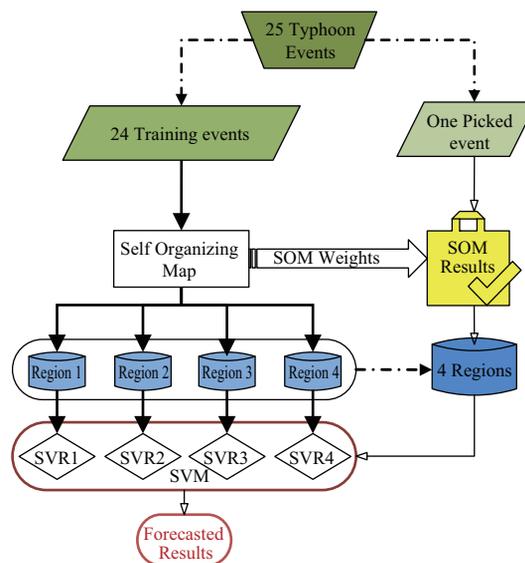


圖 1 本研究 SOM-SVM 模式

## 二、研究方法

### 2.1 自組織映射圖網路(SOM)

SOM 是一種前饋、非監督式與競爭學習式的網路模型，由 Kohonen (1982)所提出，其網路架構詳圖 2 所示。在 SOM 網路架構中，當使用者將變數向量經由輸入層輸入後，每個變數透過鍵結權重連結至輸出層的各個神經元。這些輸出層神經元則代表上述變數向量輸入後每筆資料在不同維度空間上的拓撲映射結果，輸出層的神經元會依輸入向量的特徵或趨勢，以有意義的分佈在輸出空間中進行拓撲結構 (topological structure) 的展現。而前述拓撲映射的排列方式，在不同維度空間中可以一維的線狀 (linear) 排列、二維的網格式 (lattice) 進行排列，此外亦可以更高維度的進行排列，只是一般以二維的排列方式較為常用。

SOM 網路架構建立之學習過程，包含有三個重要的程序，分別是 ① 競爭程序 (competitive process)、② 合作程序 (cooperative process) 以及 ③ 調整程序 (adaptive process)，演算過程則簡述如下：

假設每個輸入的變數如向量式(1)，為具有  $M$  維度的向量群  $X$ 。在最初的競爭程序中，基於存

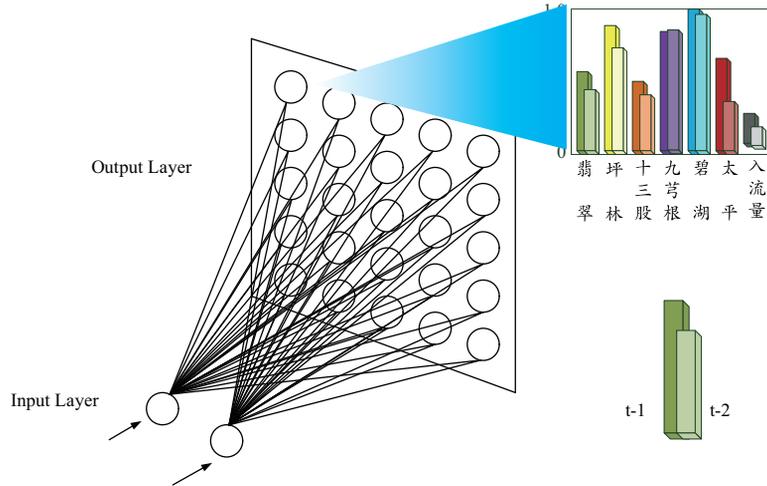


圖 2 自組織映射圖網路架構圖與分類後輸入項分佈示意圖

在輸入層及輸出層中為各神經元間的鍵結權重向量，因此均具有與輸入向量相同的  $M$  維度，而介於輸入層與輸出層中的神經元鍵結權重初始化，則以隨機變數產生如式(2)。

$$\mathbf{X} = [x_1, x_2, \dots, x_M]^T \dots\dots\dots (1)$$

$$\mathbf{W}_j = [w_{j1}, w_{j2}, \dots, w_{jM}]^T \quad j = 1, 2, \dots, l \dots\dots\dots (2)$$

經過最初競爭過程後的優勝神經元，其權重則如下式

$$i(\mathbf{X}) = \arg \min_j \|\mathbf{X} - \mathbf{W}_j\| \dots\dots\dots (3)$$

其中， $i(\mathbf{X})$ 表示與輸入之變數向量  $\mathbf{X}$  最為相似的神經元； $\|\bullet\|$  為歐氏距離 (Euclidean distance)。

換言之，於競爭過程中的每筆輸入樣本經輸入層資料輸入網路後，緊接著便開始評估此輸入樣本資料與網路中各神經元之相似程度，並進而挑出與該輸入樣本資料相似度最高之神經元，令其受到激發並進入後續程序。相似度的評估則依每一輸入樣本資料與各神經元鍵結權重之內積 (inner product)，再依所得出歐氏距離進行評估，距離越小者代表相似度越高，也就是當  $\|\mathbf{X} - \mathbf{W}_j\|$  的數值最小時，輸入向量  $\mathbf{X}$  與鍵結權重  $\mathbf{W}_j$  便具有最大相似度，而此一內積值最小者之神經元便是競爭過程中的優勝神經元。

輸入網路後的樣本資料歷經競爭程序並得出優勝神經元後，該神經元則成為拓樸鄰域之中心，並進入第二階段的合作程序。在合作程序中，由於不同神經元間存在側向間交互作用之影響，且每個神經元與不同神經元間之距離成反比，距離愈遠則影響相對愈小，此一拓樸鄰域之概念可以高斯函數表示如式(4)。

$$h_{j,i(x)} = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right) \dots\dots\dots (4)$$

上式中鄰近區域函數  $h_{j,i(x)}$  為優勝神經元  $i$  及鄰近神經元  $j$  之鄰近關係值。 $d_{j,i}^2$  則是神經元  $j$  與優勝神經元  $i$  之歐式距離。其中當歐式距離值隨著距離增加至趨近無限大時，拓樸鄰域將趨近於零，此一機制為合作程序中 SOM 網路收斂之必要條件。

最後再進入第三階段的調整程序，於神經元鍵結權重之調整過程中，鍵結權重值會隨著每一筆輸入項，往靠近輸入樣本資料之方向進行調整，調整方法以原始 Hebbian (1949) 假設為本並引入一失憶項 (forgetting term) 即調整前後之神經元鍵結權重差值，可由下式得知。

$$\Delta \mathbf{W}_j = \eta h_j (\mathbf{X} - \mathbf{W}_j) \dots\dots\dots (5)$$

$$\mathbf{W}_j(t+1) = \mathbf{W}_j(t) + \eta(t) h_j(t) (\mathbf{X} - \mathbf{W}_j(t)) \dots\dots (6)$$

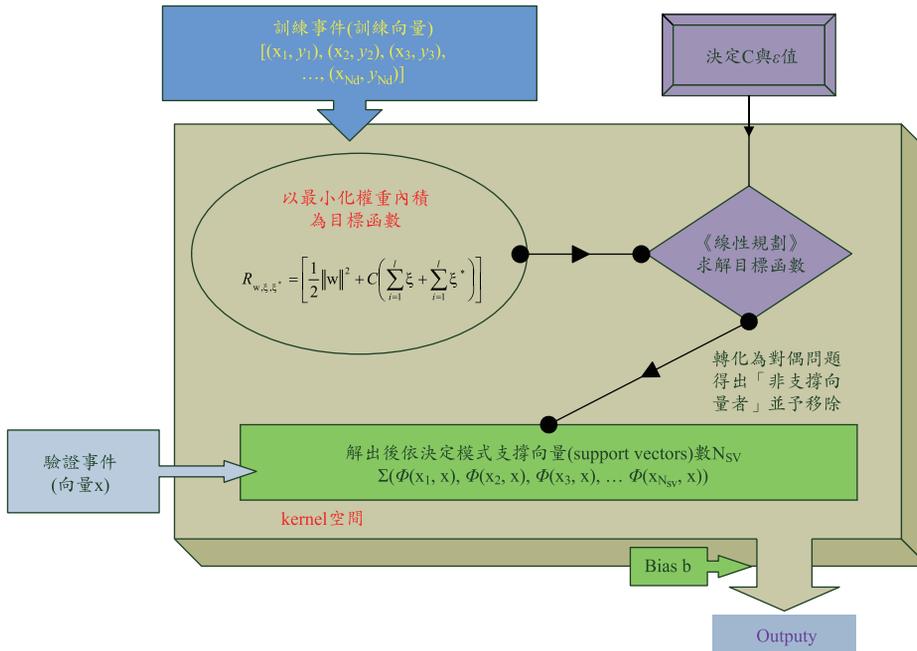


圖 3 支撐向量機演算架構

經調整程序後各神經元之新鍵結權重值將可由(6)式推求。此時，受調整程序作用的僅有拓樸鄰域內之神經元鍵結權重。這也意味著於網路收斂後，相鄰神經元將具有相近鍵結權重，而相距較遠者之鍵結權重差異也就相對較大。此處的鍵結權重調整過程包含排序階段及收斂階段，在排序階段時各神經元鍵結權重將進行其拓樸結構排序，其學習率建議需維持接近 0.1，且並不宜低於 0.01，另外疊代次數則大概需 1000 次以上(葉怡成，1995)。而在收斂階段中，初始時網路架構為提供更精確描述有關輸入樣本資料之統計量，此階段中的學習率微調在 0.01 左右則需經數千次疊代。

經由上述競爭、合作及調整等三程序的反覆進行，直到網路收斂為止，並將輸入樣本與其所對應激發之神經元在以二維空間上進行網路排列，其中並將編號或名稱標示於排列後的網格中，便可得出特徵圖(feature map)。受標示之網格元素代表網路中神經元受某特定之輸入樣本資料激發，此受標示之網格元素則稱為該特定輸入

樣本資料之映像(image)。此外，於上述得出特徵圖中，各映像所對應之輸入樣本數目累計並填入該網格即得密度圖(density map)，亦即可顯示出輸入樣本之分佈情形。

## 2.2 支撐向量機

SVM 是由 Vapnik 和其工作團隊於 1990 年初所發展出一種最初用於資料分類，後被延伸至迴歸(1995)與圖形辨識的學習工具。SVM 是建立在 VC 維度(dimension)理論和結構風險最小化原理(structural risk minimization, SRM)的歸納基礎上，取代常見的倒傳遞式類神經網路(BPN)理論中使誤差值之經驗風險最小化(empirical risk minimization)的機器演算學習工具。其原理是根據上述 SRM 基礎，利用一個二次規劃(quadratic programming)演算法保證所得到的解為全域最佳解，來同時達成經驗風險和模式複雜度最小化。此外，不同於倒傳遞式類神經網路需一再經由試誤求解的耗時方式，SVM 使用的二次規劃求解可迅速的得出最佳惟一解，其架構圖詳圖 3。

本節將 SVR 的原理簡短整理如下：

首先，訓練資料集合以數學式表示如  $[(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_{N_d}, y_{N_d})]$ ，其中  $y$  是  $\mathbf{x}$  向量對應的輸出值， $N_d$  則是訓練資料集合的數目。SVR 迴歸的目標便是在認定一個可接受誤差值  $\varepsilon$  的門檻下，產生一組需求的非線性  $y$  估算值。其步驟如下：

首先，輸入項  $\mathbf{x}$  先由非線性方程式  $\phi(\mathbf{x})$  映射到高維度的空間，在此空間中，輸入項  $\mathbf{x}$  與輸出項的關係式為：

$$\hat{y} = f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b \quad \dots\dots\dots(7)$$

其中的  $\mathbf{w}$  和  $b$  分別代表迴歸式中的權重及偏估值，本式依據前述的 SRM 基礎，對於  $\mathbf{w}$  與  $b$  值以最小化結構風險進行規劃求解。其風險式如下：

$$R = \min \left( \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{N_d} L_\varepsilon(\hat{y}_i) \right) \quad \dots\dots\dots(8)$$

式中前項所計算的是模式的複雜度，後項則是經驗誤差。其中後項的  $C$  值是一個可由使用者調控定義的參數，用以給予預測值與訓練誤差間的估算進行不利的估算條件。 $\varepsilon$  則是一被稱為 Vapnik's  $\varepsilon$ -insensitive (1995) 的可容誤差區間，其損失函數為  $L_\varepsilon(\hat{y}_i)$ ，這個函數的定義為：

$$L_\varepsilon(\hat{y}_i) = |y - \mathbf{w} \cdot \phi(\mathbf{x})|_\varepsilon = \begin{cases} 0 & \text{if } |y - \mathbf{w} \cdot \phi(\mathbf{x})| < \varepsilon \\ |y - \mathbf{w} \cdot \phi(\mathbf{x})| - \varepsilon & \text{otherwise} \end{cases} \quad \dots\dots\dots(9)$$

如此，目標函數式便可導為下式：

$$\text{Minimize } R(\mathbf{w}, \xi, \xi') = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N_d} (\xi_i + \xi'_i) \quad \dots\dots\dots(10)$$

subject to

$$\begin{aligned} y_i - (\mathbf{w} \cdot \phi(\mathbf{x}) + b) &\leq \varepsilon + \xi_i \\ (\mathbf{w} \cdot \phi(\mathbf{x}) + b) - y_i &\leq \varepsilon + \xi'_i \\ \xi_i, \xi'_i &\geq 0, \quad i = 1, 2, \dots, l \end{aligned}$$

其中  $\xi$  與  $\xi'$  是針對容許誤差上、下界受限於  $\varepsilon$  管界中的鬆弛係數。也就是說，模式複雜度及可容許誤差的交易取決於  $C$  與  $\varepsilon$  的值，當  $C$  值增大時，能夠給予誤差較大的懲罰乘積，不利於規劃求解之目標函數的最小化。如此便可控制使容許誤差空間變小，增加估計值的準確性，只是相對地會因有效反應太多非離群支撐向量的個數，應加深模式的複雜度，對於未經學習的輸入向量可能因過度適用而無法正確反應；反過來說，若是容許誤差過大，也許可以對尚未學習的輸入向量給予適度的容錯值，進而令模式有足夠的空間捕捉離群的向量，得到較大的解空間支撐向量。只是，在同樣不利於最小化的目標函數，一旦容許誤差過大，相對無用的離群向量過多，亦可能反而使估計值準確度不足。

本目標函數式則常常以 Lagrangian multipliers 將二次規劃問題轉化為 Lagrangian 對偶問題求解：

$$\begin{aligned} &\frac{1}{2} \sum_{i,j=1}^{N_d} (\alpha_i - \alpha'_i)(\alpha_j - \alpha'_j) K(\mathbf{x}_i, \mathbf{x}_j) \\ &+ \varepsilon \sum_{i=1}^{N_d} (\alpha_i - \alpha'_i) - \sum_{i=1}^{N_d} y_i (\alpha_i - \alpha'_i) \end{aligned} \quad \dots\dots\dots(11)$$

subject to

$$\begin{aligned} &\sum_{i=1}^{N_d} (\alpha_i - \alpha'_i) = 0 \\ &C \geq \alpha_i \geq 0 \\ &C \geq \alpha'_i \geq 0 \\ &i = 1, \dots, N_d \end{aligned}$$

這裡的  $\alpha_i$  及  $\alpha'_i$  為 dual Lagrange multipliers

$$f(\mathbf{x}) = \sum_{i=1}^{N_d} \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b \quad \dots\dots\dots(12)$$

此處的  $K(\mathbf{x}, \mathbf{x}_i)$  函式定義為：

$$K(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x})^T \phi(\mathbf{x}_i) \quad \dots\dots\dots(13)$$

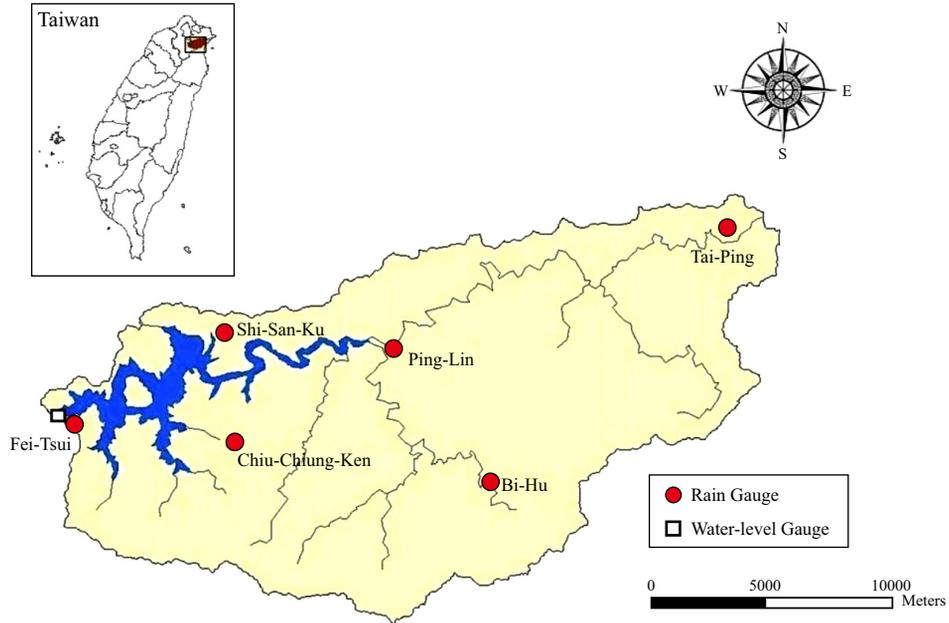


圖 4 翡翠水庫集水區

即為建構出最佳超平面核心函式，在本研究上，本函式使用輻射基底方程式。

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma |\mathbf{x} - \mathbf{x}_i|^2) \dots\dots\dots(14)$$

部份的 Lagrange multipliers 的對偶項會因為  $(\alpha - \alpha')$  的值为 0，使得在迴歸過程中被視為無影響性而被清除。如此，在迴歸方程式中，訓練資料 Lagrange multipliers 對偶過程未被移除的向量被視為有效的輸入項，這些向量即所謂「支撐向量群」，最後，原迴歸式便可寫成：

$$f(\mathbf{x}) = \sum_{k=1}^{N_{\text{sv}}} \alpha_k K(\mathbf{x}, \mathbf{x}_k) + b \dots\dots\dots(15)$$

### 三、應用實例

在本研究中，所推導的模式被應用在台灣北部翡翠水庫集水區中，翡翠水庫位於三條主要支流下游(金瓜寮溪、魚逮魚堀溪及北勢溪)，總表面積為 10 平方公里，平均深度 40 公尺，最大深度 120 公尺，總庫容則為 406 百萬噸，總流域面積則有 303 平方公里(如圖 4)。正因翡翠水庫供應

作為首都台北市的用水，它等同是北台灣最重要的水庫。

降雨資料及水庫入流量資料收集自 1998 至 2013 年水利署災害緊急應變系統，最大降雨及年平均降雨量分別為 5,736.6 mm 和 3,808.6 mm。在本文中，以其中 25 個颱風事件的時雨量及時入流量資料應用在本研究，各颱風事件發生時間、延時長度詳如表 1 所示，這 25 個颱風事件被交叉用來預測彼此。亦即使用其中 24 場颱風事件前 2 期不同雨量站的時雨量及水庫時入流量資料經過 SOM 分類後，依 24 場在 SOM 圖像所得的分類權重結果對餘下 1 場的時雨量進行分類。由於輸入項分別有時雨量與時入流量，基於尺度上的不同，各項輸入項均需先由各輸入項的最大及最小值範圍，經過正規化後方能作為輸入 SOM 及其後 SVM 模式的訓練項。

建構得出的分類成果，再以這 24 場的分類資料為訓練值，針對分類成果在 SOM 上的分佈，分別作為 SVR 模式建構的訓練資料，所建構得出模式再對餘下 1 場分類成果進行 SVR 迴歸，得出預測結果後再重組。同樣情形對餘下一場颱風進

表 1 本研究所用 25 場颱風事件及事件長度

颱風名稱	日期 (年/月/日)	延時 (h)	尖峰流量 (m <sup>3</sup> /s)
寶莉	1992/08/27	91	970
泰德	1992/09/20	71	922
提姆	1994/07/09	51	674
弗雷特	1994/08/19	76	718
葛拉絲	1994/08/31	34	1,449
席斯	1994/10/08	77	1,457
賀伯	1996/07/30	72	2,586
瑞伯	1998/10/15	52	2,642
象神	2000/10/31	34	4,075
納莉	2001/09/16	85	4,022
海燕	2001/10/15	30	560
雷馬遜	2002/07/03	34	1,096
艾利	2004/08/23	70	2,416
納坦	2004/10/24	46	1,719
南瑪都	2004/12/03	33	1,690
海棠	2005/07/17	68	1,476
馬莎	2005/08/04	50	1,226
泰利	2005/08/31	45	1,798
柯羅莎	2007/10/04	63	2,356
鳳凰	2008/07/26	43	969
辛樂克	2008/09/11	107	3,598
薔蜜	2008/09/26	52	2,513
南瑪都	2011/08/27	100	521
蘇拉	2012/07/30	131	2,882
潭美	2013/08/20	68	1,351

行替換，直至 25 場不同颱風都被預測以進行模式驗證。

延時長度決定：

在模式建構中，首先決定預測項長度是必要的。每個颱風事件的模式的項測項為：

$$F_{t+\Delta t} = f(R_{G,t}, R_{G,t-1}, \dots, R_{G,t-(L_R-1)}, F_t, F_{t-1}, \dots, F_{t-(L_F-1)}) \quad G = 1, 2, \dots, 7 \quad (16)$$

t 代表當前時間；Δt 代表延時(1 到 3 小時)； $R_{G,t}$  是指第 G 個雨量站在第 t 小時的雨量； $F_t$  為在第 t 小時的水庫入流量；f 則是依據  $L_F$  和  $L_R$  導出的項。

對於水庫入流預測模式，適合的稽延時間輸入對模式建構是非常重要的工作。在本文中，

RPE 標準被應用來計算輸入項的延時長度。RPE 定義為：

$$RPE = \frac{E(L) - E(L+1)}{E(L)} \times 100 \quad (17)$$

此處 E(L)和 E(L+1)分別為模式在 L 延時及 L+1 延時的均方根誤差，均方根誤差可以從下式得到。

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{F}_t - F_t)^2} \quad (18)$$

在此  $F_t$  是時間 t 的入流量， $\hat{F}_t$  是被預測時間 t 的入流量，n 是時間階數。一般來說，RMSE 會隨著時間增加而減少，當 RPE 值少於 5%時，延時長度的增加便予以停止。

經由此一過程，對於一個可靠的需求時間來說，最適當的颱風降雨延時長度  $L_R$  及入流量延時長度  $L_F$  可由此求得。這個適當的 2 小時降雨及入流延時長度使用來預測未來 1 到 3 小時的入流量。其對於 SVM 為工具表現的一般項可寫為：

$$F_{t+\Delta t} = f(R_{G,t-1}, R_{G,t-2}, F_{t-1}, F_{t-2}) \quad G = 1, 2, \dots, 7 \quad (19)$$

最後，為了令 2 個以 SVM 進行預測的模式能夠合理的被比較，SVM 中的 C 及 ε 分別固定為 1 及 0.1。

評估指標：

為評估模式的效能，本研究以效率係數 (coefficient of efficiency, CE)及均方根誤差(relative root-mean-square-error, RRMSE)作為評估指標。

$$RRMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n \left( \frac{\hat{F}_t - F_t}{F_t} \right)^2} \quad (20)$$

$$CE = 1 - \frac{\sum_{t=1}^n (F_t - \hat{F}_t)^2}{\sum_{t=1}^n (F_t - \bar{F}_t)^2} \quad (21)$$

為比較以 SVM 預測為基礎，先期颱風降雨量及水庫入流量資料經 SOM 分類過後，再由 SVM 的模式預測能力，原始 SVM 由每場被預測

場次的颱風以外的其他颱風事件建構得出，並直接用以預測該颱風事件。接著，本研究所建構，以 SVM 為基礎的模式具備 2 個部份：SOM 的分類和 SVM 的迴歸。

首先，本模式如同本節最初說明，說需先選擇 1 場颱風事件，作為驗證模式可行性的事件，餘下的其他 24 場次事件則作為建構本模式的所有輸入資料。選擇作為驗證的事件為在模式建構完成後，依序再選擇其餘 24 場的其中 1 個事件作為驗證事件，一直至此所有 25 場颱風事件均由其他 24 場次事件進行建構完畢為止。將所建構輸入資料經由 RPE 指數測試，決定出各颱風事件的前兩期時雨量及前 1 期出流量作為輸入項長度，亦即：

$$F_{t+\Delta t} = f(R_{G,t}, R_{G,t-1}, F_t) \dots\dots\dots(22)$$

決定輸入項長度後，便將本訓練項輸入 SOM 分類工具中，令其以各輸入項的歐式距離對所有輸入項進行自動分類。在分類過程中，以產出由小至大的二維圖像作為分類結果判定，即從 2×2 網格開始，擴張到 3×3 與 4×4，一直到可以有效判定分類在觀察上有一定成果才停止。本研究中經過測試，在觀察網格上所得出各輸入項能夠有效區分出颱風事件的前、中及後的分佈時，所得出最佳分類的成果是 5×5 網格，為進行說明，本研究將所得出 25 個網格從 1 到 25 號，分別編號如圖 5 中編號數值所示。

在得出二維分類網格的成果後，經由觀查其輸入項對於原始颱風事件的分佈情況，發現這些輸入項在二維 SOM 圖像上，共 25 個群組的分佈，大致可以分成 4 區：分別是颱風開始發生的水位上升區、颱風事件後的水位下降區、颱風事件的洪峰流量區，以及颱風影響還不大的基流量區。

在此，為了加強 SVR 模式的預測成果，本研究基於各段颱風過程已經大致可以辨識的條件下，將最初選為驗證事件作為驗證事件的輸入項，依照已得出 SOM 圖像所產生的權重值，將驗證事件輸入項進行再分類。

此時，我們已經分別擁有分類完畢訓練事件

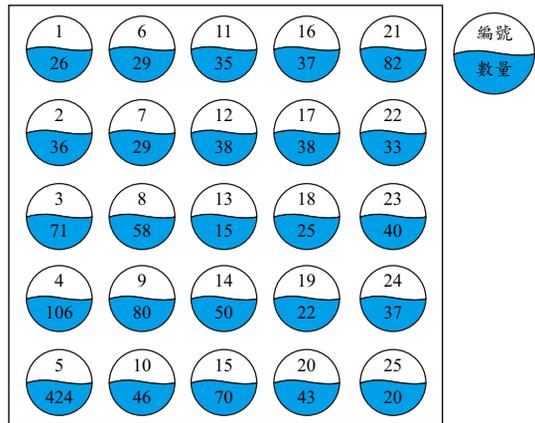


圖 5 在 5 × 5 自組織映射圖上所得出的分類編號及輸入項數量

的輸入群組及分類完畢驗證事件的輸入群組，這些群組便可開始分別由訓練項群組架構 SVR 模式，並經由預測驗證群組的結果與原始事件流量比較。只是，由於各颱風事件的長度不一，最長達 107 小時，最短則只有 33 小時，加上本二維 5 × 5 的 SOM 圖像上共有 25 個分類群組，並非每個事件都能均勻散佈在這 25 個分類群組中，而是與該筆資料中前 1 及 2 小時的各項時雨量與時入流量分佈相關，所產生之分佈圖像。再者，由 SOM 分類過後的密度圖像(圖 5 中各圓下方數量值)來比對，網格中訓練資料密度雖由圖中可以觀察到，資料密度者為 5 號網格中因為在雨量無太大變化，且流量最低時的 424 筆。但訓練資料密度最小的網格上卻只有 15 筆資料，如單純只以這 15 筆資料進行模式的架構所做的預測於同網格中的資料進行評估，即使具有高相關性，卻也容易造成偏估。是以考量上面提到分成 4 區的情形，本研究亦在進行 SVR 模式建構及預測的過程中，以建構 4 個不同颱風降雨出流過程的模式來強化 SVR 的模式。

在一開始，每個事件的預測項都經由受預測事件外的颱風事件所建立 SOM 圖像進行分類。接著，這些預測項再經由不同的策略進行預測。各經過預測後的成果再由時間序列組為颱風事件預測。以寶拉颱風為例，我們將寶拉颱風選為驗證事件，其餘 24 場颱風則均為模式的訓練事

件。將這 24 場颱風的各時段前 2 小時雨量及前 1 小時流量作為輸入項，輸入至 SOM 後，可以得出如圖 5 所示的  $5 \times 5$  二維 SOM 圖像。接著於圖 6 便可觀察這得出的 SOM 圖內各輸入項與原始不同颱風事件的關係，便可發現分佈在左上角 6 個群組內的輸入項，幾乎都是代表著颱風事件流量的上升段，左下 9 個群組則代表的是基流量、右上 4 個群組代表的則為各事件的洪峰流量，最後，右下的 6 個群組代表的是颱風事件入流量的退水段，4 個入流量對照的分類情況詳如圖 7 及圖 8。因此，在此我們便依據颱風事件在這 4 個不同分區上的分佈，建構出 4 個 SVR 的入流量預測模式。然後，就將寶拉颱風先依照 SOM 圖像所得出的權重，將寶拉颱風的輸入項乘上權重後，便可將輸入項分配到 SOM 圖上的 25 個群組中，得到寶拉颱風各時段在 SOM 圖上的分佈情形。接著，再依據 4 個不同颱風降雨出流過程的分區所建構的 SVR 模式，將 25 個群組中的寶拉颱風輸入項輸入到 4 個模式中，得出下一時段入

流量預測值，最後，再將各時段預測入流值依時間序列過程重組，便可得出寶拉颱風入流量的預測值。

如此，上述如寶拉颱風作為驗證的過程，套用在其他 21 場事件中，也就是說每一場颱風事件的入流量值預測，均以其他 21 場事件進行 SOM 分類及 SVR 模式的建構後，再將受驗證事件進行分類並得出入流量，進行交叉驗證。

#### 四、成果討論

在 SOM 的分類過程中，由於 25 次的分類過程資料大致固定，只是依序抽出一場事件，對於分類結果影響並不大，其分佈情形大致就如圖 5 的密度圖。比較不同事件抽出後的分類過程，大致只有 14 筆資料會隨事件不同的情況造成所在網格位置移動 1 格，這些變動位置則大多位在編號 8 號、12 號、13 號、14 號及 18 號的位置。

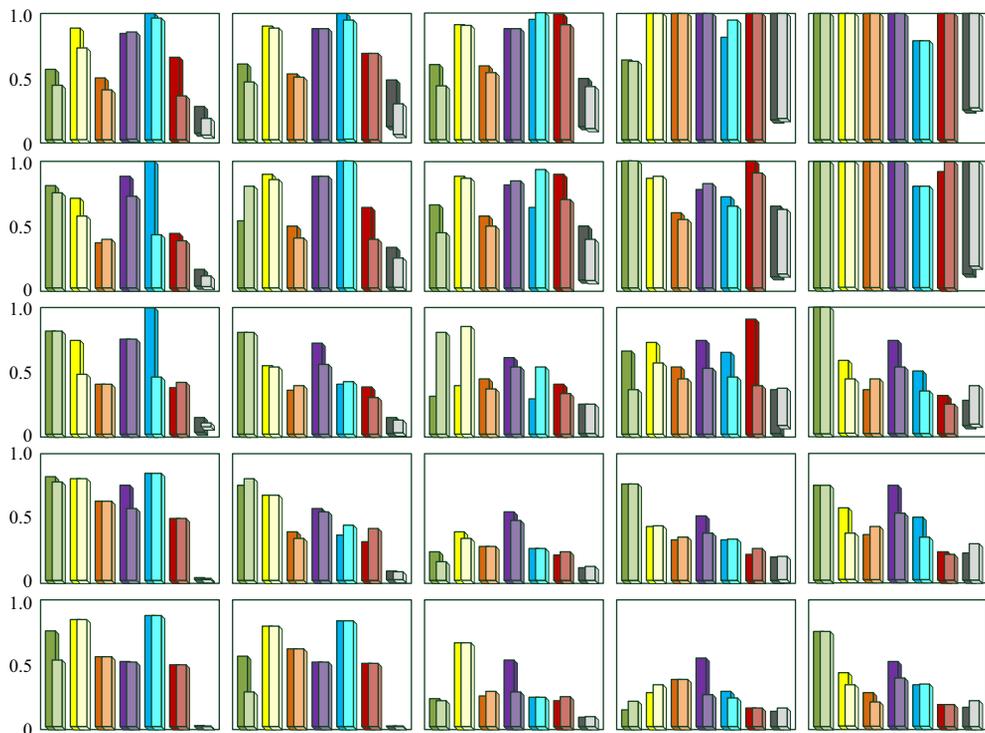


圖 6 在  $5 \times 5$  自組織映射圖上各分類單元內所包含的分類情形

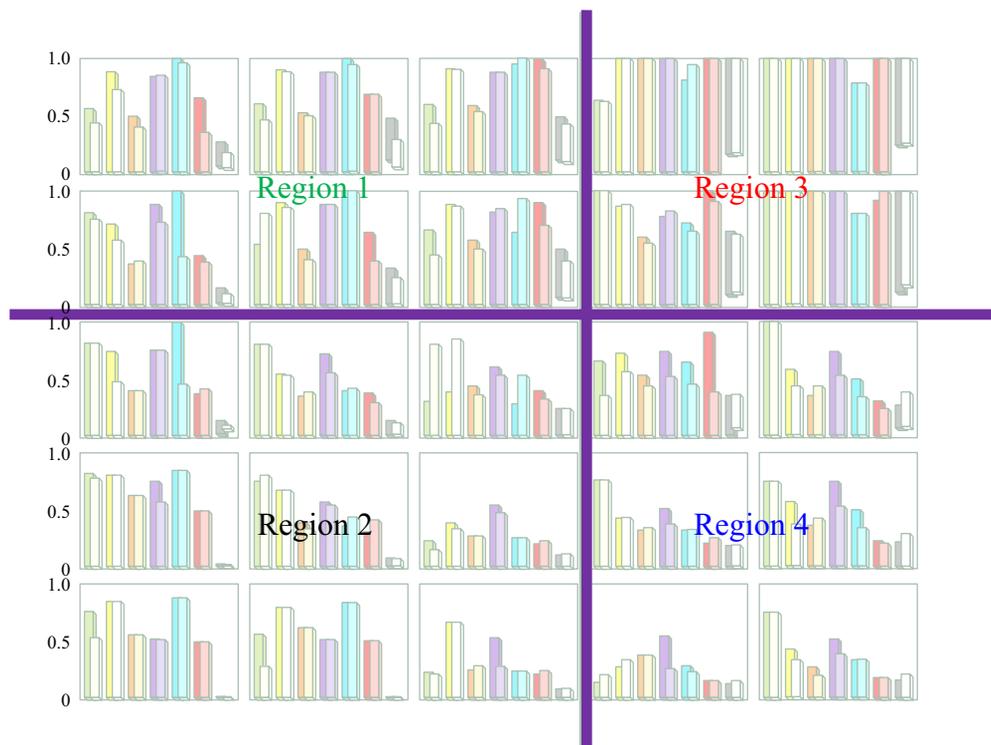


圖 7 5 × 5 自組織映射圖經過水庫入流量分區考量所得出的區間

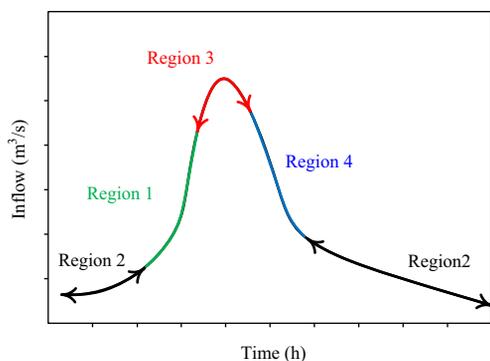


圖 8 對照 SOM 圖像上的區間與水庫入流歷線之相關過程

也因為圖像的分類情形大致穩定，依據 SOM 分類所產生的資料分佈並比對降雨逕流的過程所進行的模式建構分類，大致如圖 7 與圖 8 所示。而其中 Region 2 的分類結果所擁有的資料數量最為龐大，達到 662 筆；最小的 Region 3 則只有 76 筆。對照圖 8 的降雨-逕流過程可以了解，屬於水庫基礎入流量的資料量最大，而尖峰流量

資料量最少，這樣的分類成果其實並不令人感到意外。

圖 9 是實際入流量及經由原始 SVR 的 1 小時入流量預測結果的散佈圖，對照圖 10 為經過 SOM 強化模式預測能力後，得出的散佈圖，可以明顯看出，在經過依照不同降雨-逕流過程分類的模式建構，所產生的資料散佈情況，的確是較為集中。原始模式中，由於模式的建構考量的可用資訊來自所有可用資料，因此會出現低流量可能高估，高流量反而可能低估，較偏中間值的流量平均上也許相差無幾，但散佈圍範仍是較廣；相對經過分類後再行預測的 SOM-SVR 模式則因為經過入流過程的分類在先，所引為訓練的資訊均是較具有高度相關性的資料組，是以無論是入流過程的前、中、後，在散佈上均相當地貼近 45 度線。這表示經過 SOM-SVR 的預測後，模式表現確實優於原始 SVR 的預測結果。

再由圖 11 的入流歷線圖比較，較之原始 SVR 模式所得出預測值，SOM-SVR 模式的預測成果

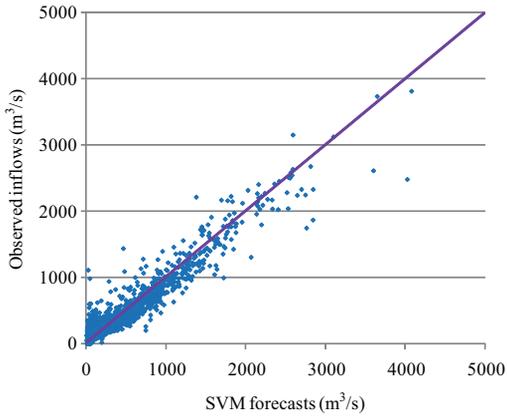


圖 9 原始 SVM 對水庫入流量的預測結果散佈圖

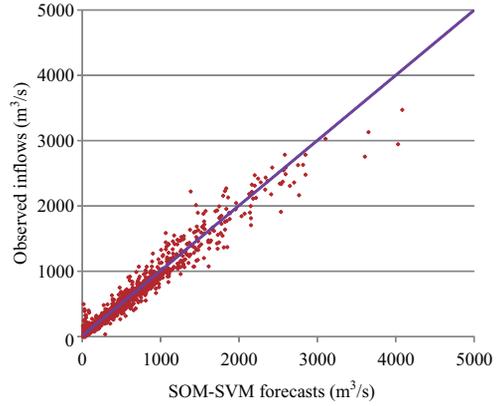


圖 10 由 SOM-SVM 模式所推得水庫入流量的預測結果散佈圖

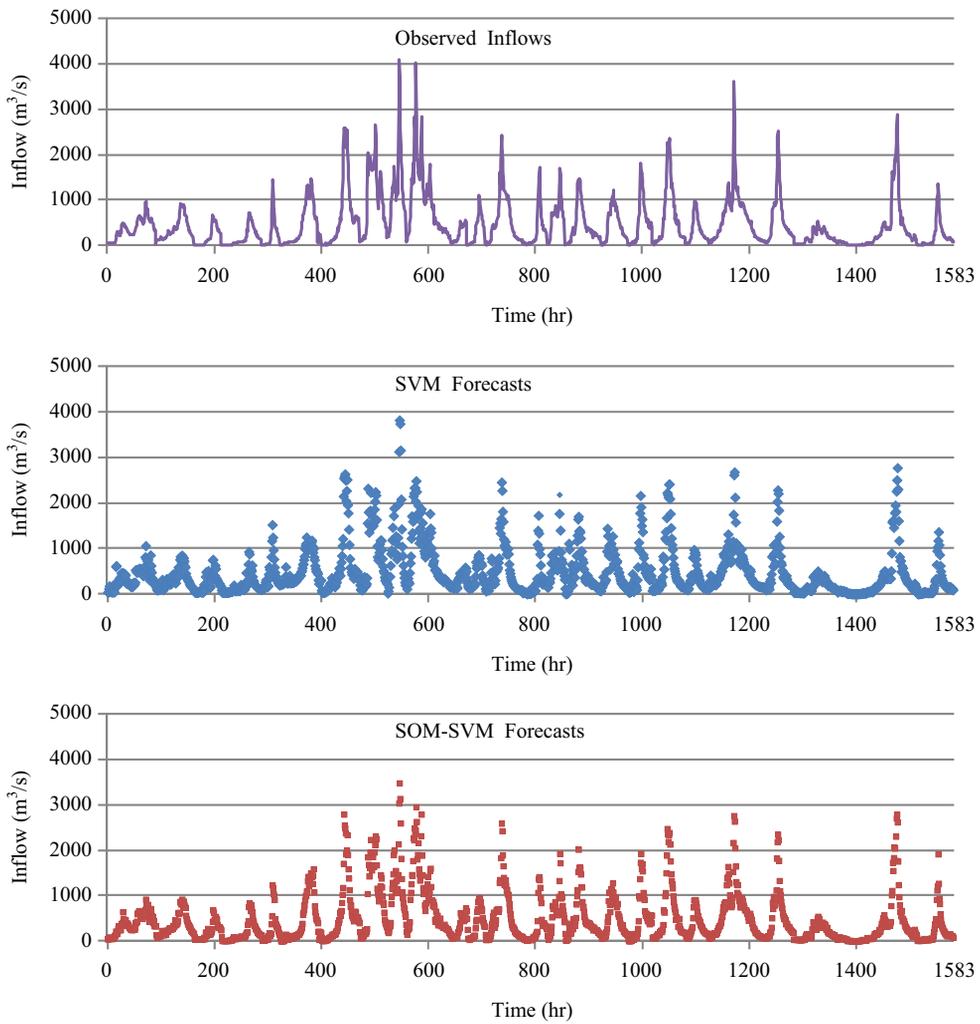


圖 11 兩模式水庫入流量預測比較

表 2 SVM 模式與 SOM-SVM 模式 1 小時預測成果比較

颱風事件	SVM		SOM-SVM	
	RRMSE	CE	RRMSE	CE
寶莉	1.358	0.583	0.447	0.913
泰德	0.316	0.770	0.159	0.940
提姆	3.627	0.551	1.722	0.872
弗雷特	1.460	0.745	0.369	0.959
葛拉絲	7.430	0.644	1.354	0.921
席斯	5.882	0.650	0.817	0.936
賀伯	1.375	0.959	0.368	0.967
瑞伯	0.340	0.905	0.187	0.958
象神	0.280	0.918	0.185	0.915
納莉	0.268	0.808	0.263	0.895
海燕	1.222	0.743	0.473	0.912
雷馬遜	2.975	0.643	0.740	0.932
艾利	2.784	0.935	1.245	0.977
納坦	0.919	0.788	0.769	0.961
南瑪都	1.681	0.806	0.259	0.924
海棠	7.049	0.727	2.329	0.923
馬莎	4.292	0.523	1.474	0.904
泰利	3.623	0.874	0.875	0.883
柯羅莎	0.667	0.960	0.554	0.980
鳳凰	0.876	0.830	0.289	0.979
辛樂克	0.696	0.858	0.212	0.938
薔蜜	0.400	0.937	0.133	0.986
南瑪都	0.885	0.901	0.787	0.937
蘇拉	0.888	0.940	0.653	0.973
潭美	0.466	0.947	0.328	0.953

的確較能貼近原始入流量觀測值。而表 2 中在分析成果的表現上，在 25 場颱風事件在 1 小時預測成果中，原始 SVM 模式僅有 9 場能令效率係數達到 0.9 以上，而經過 SOM-SVR 的強化模式，則只剩下 3 場事件無法達到 0.9 以上。而相較兩模式的成果，SOM-SVR 模式只有在象神颱風事件的效率係數上略低原始 SVM 的預測成果，但其實在 RMSE 的成果上，SOM-SVR 的效果均優於原始 SVM 模式。在象神颱風事件的預測成果中，其實兩個模式的效率係數均已超過 0.9，RMSE 的成果也是由 SOM-SVR 模式表現較優，差別只在於原始 SVM 模式局部流量預測可能稍有優於 SOM-SVR 模式。

表 3 則顯示在不同延時的預測上，比起原始

表 3 SVM 模式與 SOM-SVM 模式 1~3 小時預測平均成果比較

Lead time (h)	MRMSE (m <sup>3</sup> /s)		MCE	
	SVM	SOM-SVM	SVM	SOM-SVM
1	167.078	96.023	0.798	0.938
2	202.715	149.595	0.710	0.871
3	253.993	223.094	0.560	0.690

SVR 模式在未來 1 小時的預測能力，SOM-SVR 從均方根誤差的 167.078 降低到 96.023；未來 2 小時從 202.715 降低至 149.595；未來 3 小時則從 253.993 降低至 223.094。總的來說，SOM-SVR 模式較之原始 SVM 模式在均方根誤差表現上雖然隨著延時長度而有所遞減。但在效率係數上，1 至 3 小時的效率係數卻各自從 0.798、0.710、與 0.560 提昇至 0.938、0.871 及 0.690，分別提昇了 17.5%、22.7%及 23.2%。是以對於水庫在颱風期間入流量的預測上，有著穩定的提昇。此一預測的方式及成果，期望能夠對於臺灣面臨颱風事件侵襲時，在水庫入流量預測及下游地區因水庫容量操作時的防災應變，能夠隨著入流量預測效果的提昇，使水庫的蓄水效果提昇外，也能讓水庫下游地區的災害最小化。

## 五、結 論

1. 經過 SOM 先將資料分類後，再進行 SVR 所做預測結果，確實能夠強化 SVR 對於水庫在颱風事件中對入流量的預測能力。
2. 對於颱風事件的分類預測過程，除了對於分類情形需進行觀察外以進行颱風降雨-逕流過程辨識，作為 SVR 預測組別依據外，SOM 圖像在 5 × 5 的二維網格所做分類效果是較佳的選擇。
3. 經過 SOM 依照降雨-逕流過程分類強化後的水庫入流量預測，不僅在 25 場颱風降雨的水庫入流量事件預測中，已有 22 場事件達到效率係數 0.9 以上，較之原始 SVM 的事件只有 9 場達到 0.9 以上為優外。在 1~3 小時的平均效率係數分別可增強 17.5%、22.7%及 23.2%的預測效果。

4. 本研究所得 SOM-SVR 預測模式，可供未來水庫在颱風期間庫容操作上，除使水庫的蓄水效果提昇外，也能讓水庫下游地區的災害最小化。

## 六、參考文獻

1. Alvarez-Guerra, M., González-Piñuela, C., Andrés, A., Galán, B. & Viguri, J. R., "Assessment of Self-Organizing Map artificial neural networks for the classification of sediment quality." *Environment International* **34(6)**, 782-790, 2008.
2. Baylar, A., Batan & M., "Usage of artificial intelligence methods in free flowing gated closed conduits for estimation of oxygen transfer efficiency." *Advances in Engineering Software* **41(5)**, 729-736, 2010.
3. Bovolo, F., Bruzzone, L. & Marconcini, M., "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure." *IEEE Transactions on Geoscience and Remote Sensing* **46(7)**, 2070-2082, 2008.
4. Chen, S. T. & Yu, P. S. 2007. "Pruning of support vector networks on flood forecasting." *J. Hydrol.* **347**, 67-78.
5. Chen, Y.H. & Chang, F. J. 2009 "Evolutionary artificial neural networks for hydrological systems forecasting." *J. Hydrol.* **367**, 125-137.
6. Coulibaly, P., Anctil, F. & Bobée, B., "Daily reservoir inflow forecasting using artificial neural networks with stopped training approach." *J. Hydrol.* **230**, 244-257, 2000.
7. Cristianini, N. & Shaw-Taylor, J., "An introduction to support vector machines and other kernel-based learning methods." Cambridge university press, New York, 2000.
8. Chang, F. J., Chang, L. C., Kao, H. S., & Wu, G. R., "Assessing the effort of meteorological variables for evaporation estimation by Self-Organizing of Neural Network" *J. Hydrol.* **384**, 118-129, 2010.
9. Hiew, B. Y., Teoh, A. B. J. & Yin, O. S., "A secure digital camera based fingerprint verification system." *Journal of Visual Communication and Image representation* **21(3)**, 219-231, 2009.
10. Hsu, K. L., Gupta, H. V., Gao, X. & Sorooshian, S., "Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrology modeling and analysis." *Wat. Res. Res.* **38(12)**, 1302, doi:10.1029/2001WR000795, 2002.
11. Kohonen, T., "The Self Organizing Map." *Proceedings of the IEEE* **78**, 1464-1480, 1990.
12. Kohonen, T., "Self-Organizing Maps." Springer, Berlin, 1995.
13. Kohonen, T., "Self-Organizing Maps." Springer, New York, 2001.
14. Lin, G. F. & Chen, L. H., "A non-linear rainfall-runoff model using radial basis function network." *J. Hydrol.* **289**, 1-8, 2004.
15. Lin, G. F. & Chen, L. H., "Time series forecasting by combining the radial basis function network and the self-organizing map." *Hydrol. Process.* **19(10)**, 1925-1937, 2005.
16. Lin, G. F. & Chen, L. H., "Identification of homogeneous regions for regional frequency analysis using the self-organizing map." *J. Hydrol.* **324**, 1-9, 2006.
17. Lin, G. F. & Wang, C. M., "Performing cluster analysis and discrimination analysis of hydrological factors in one step." *Adv. Water Res.* **29(11)**, 1573-1585, 2006.
18. Lin, G. F. & Wu, M. C., "A SOM-based approach to estimating design hyetographs of ungauged sites." *J. Hydrol.* **339**, 216-226, 2007.
19. Lin, G. F., Wu, M. C., Chen, G. R. & Tsai, F. Y., "An RBF-based model with an information processor for forecasting hourly reservoir inflow during typhoons." *Hydrol. Process.* **23**, 3598-3609, 2009a.
20. Lin, G. F., Chen, G. R., Wu, M. C. & Chou, Y.

- C., "Effective forecasting of hourly typhoon rainfall using support vector machines." *Wat. Res. Res.* **45**, W08440. doi:10.1029/2009WR007911, 2009b.
21. Lin, G. F., Chen, G. R., Huang, P. Y. & Chou, Y. C., "Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods." *J. Hydrol.* **372(1-4)**, 17-29, 2009c.
  22. Lin, G. F., Chen, G. R. & Huang, P. Y., "Effective typhoon characteristics and their effects on hourly reservoir inflow forecasting." *Adv. Water Res.* **33(8)**, 887-898, 2010.
  23. Lin, G. F., Chou, Y. C. & Wu, M. C., "Typhoon flood forecasting using integrated two-stage support vector machine approach." *J. Hydrol.* **486**, 334-342, 2013a.
  24. Lin, G. F., Huang, P. K. & Lin, H. Y., "Forecasting tropical cyclone intensity change in the western North Pacific." *Journal of Hydroinformatics* **15(3)**, 952-966, 2013b.
  25. Lin, G. F., Jhong, B. C. & Chang, C. C., "Development of an effective data-driven model for hourly typhoon rainfall forecasting." *J. Hydrol.* **495**, 52-63, 2013c.
  26. Liong, S. Y. & Sivapragasam, C., "Flood stage forecasting with support vector machines." *J. AWWRA* **38(1)**, 173-186, 2002.
  27. Mangiameli, P., Chen, S. K. & West, D., "A comparison of SOM neural network and hierarchical clustering methods." *European Journal of Operational Research* **93**, 402-417, 1996.
  28. Michaelides, S. C., Pattichis, C. S., Kleovoulou, G., "Classification of rainfall variability by using artificial neural networks." *International Journal of Climatology* **21**, 1401-1414, 2001.
  29. Moradkhani, H., Hsu, K. L., Gupta, H. V. & Sorooshian, S., "Improved streamflow forecasting using self-organizing radial basis function artificial neural networks." *J. Hydrol.* **295**, 246-262, 2004.
  30. Muller, D., "Self organized mapping of data clusters to neuron groups." *Neural Networks* **22(4)**, 415-424, 2009.
  31. Munoz, A. & Gonzalez, J., "Representing functional data using support vector machines." *Pattern Recognition Letters* **31(6)**, 511-516, 2010.
  32. Pan, T. Y. & Wang, R. Y., "State space neural networks for short term rainfall-runoff forecasting." *J. Hydrol.* **297(1-4)**, 34-50, 2004.
  33. Samsonova, E. V., Kok, J. N. & Ijzerman, A. P., "TreeSOM: Cluster analysis in the self-organizing map." *Neural Networks* **19**, 935-949, 2006.
  34. Sivapragasam, C. & Liong, S. Y., "Flow categorization model for improving forecasting." *Nordic Hydrology* **36(1)**, 37-48, 2005.
  35. Song, X. & Wyrwicz, A. M., "Unsupervised spatiotemporal fMRI data analysis using support vector machines." *NeuroImage* **47(1)**, 204-212, 2009.
  36. Srinivasulu, S. & Jain, A., "A comparative analysis of training methods for artificial neural network rainfall-runoff models." *Applied Soft Computing* **6**, 295-306, 2006.
  37. Tserng, H. P., Lin, G. F., Tsai, L. K. & Chen, P. C., "An enforced support vector machine model for construction contractor default prediction", *Automation in Construction* **20(8)**, 1242-1249, 2011.
  38. Vapnik, V. N., "The Nature of Statistical Learning Theory." Springer Verlag, New York, 1995.
  39. Vapnik, V. N., "Statistical Learning Theory." Wiley, New York, 1998.
  40. Vapnik, V. N., "An overview of statistical learning theory." *IEEE Transactions on Neural Networks* **10(5)**, 988-999, 1999.
  41. Wu, M. C., Lin, G. F. & Lin, H. Y., "Improving

- the forecasts of extreme streamflow by support vector regression with the data extracted by self organizing map.” *Hydrol. Process.* **28(2)**, 386-397, 2014.
42. Yang, H., Hsu, K. L., Sorooshian, S. & Gao, X., “Self-organizing nonlinear output (SONO): a neural network suitable for cloud patch-based rainfall estimation from satellite imagery at small scales.” *Wat. Res. Res.* **41**, W03008. doi:10.1029/2004WR00314, 2005.
43. Yang, T. C., Yu, P. S., Wei C. M. Wei. & Chen, S. T., “Projection of climate change for daily precipitation: A case study in Shih-Men Reservoir catchment in Taiwan.” *Hydrological Processes.* **25(8)**, 1342-1354, 2011.
44. Yu, P. S., Chen, S. T. & Chang, I. F., “Support vector regression for real-time flood stage forecasting.” *J. Hydrol.* **328**, 704-716, 2006.
45. Yu, X. Y. & Liong, S. Y., “Forecasting of hydrologic time series with ridge regression in feature space.” *J. Hydrol.* **332(3-4)**, 290-302, 2007.
46. Yu, X. Y., Liong, S. Y. & Babovic, V., “EC-SVM approach for real-time hydrologic forecasting.” *J. Hydroinf.* **6(3)**, 209-223, 2004.
47. 葉怡成：類神經網路-模式應用與實作，儒林圖書有限公司，4版，1995。

收稿日期：民國 104 年 5 月 27 日

修正日期：民國 104 年 12 月 21 日

接受日期：民國 105 年 1 月 4 日