
* 論 著 *

應用類神經網路推估溪流之生物多樣性

A Study of Artificial Neural Networks for Estimating Riverine Biodiversity

蔡 文 柄

WEN-PING TSAI

國立台灣大學生物環境系統工程學系
研究助理

張 斐 章*

FI-JOHN CHANG

國立台灣大學生物環境系統工程學系 教授
農業工程研究中心 主任

摘 要

河川流量管理為一兼顧人類使用需求及河川生態系統需求之理念，將生態觀念融入河川流量經營管理之中，以達到人類與河川生態系統共存之理想。本研究建立自組特徵輻狀基底類神經網路架構，透過此模式對流量資料進行型態判別分類並利用台灣生態水文指標系統推估溪流生態中魚類的生物多樣性。共選取全台灣河川中未受人為控制、流量資料長度大於二十年且其上下游10公里內有魚類調查資料之流量站作為研究測站，模式於訓練階段的科技異度實際值與網路推估值之相關係數高達0.82，而測試階段的實際值與推估值表現上，相關係數仍然有0.66，顯示模式除有對流量資料進行型態判別分類的能力外，也能準確推估生物多樣性。

關鍵詞：自組特徵輻狀基底類神經網路，台灣生態水文指標系統，生物多樣性。

ABSTRACT

Stream flow management is the idea that combines the concept of ecology and provides the demand for both human and river ecosystem. This study were built Self-Organizing Radial Basis Neural Networks. By this model, it can categorize the stream flow data and be estimated the diversities of fish families in river ecosystem by using the index of Taiwan Ecohydrology Index System. In this research, the stream flow data which are only collected with records more than 20 years and without anthropogenic control would be tested. The output of the model showed a high authenticity of the prediction ($r > 0.82$ and 0.66 for training and testing process, respectively) for the diversities. The result shows that this model not only can categorize the stream flow data but also can estimate the bio-diversity efficiently and precisely.

Keywords: Self-Organizing Radial Basis Neural Networks, Taiwan Ecohydrology Index System, Bio-diversity.

* 通訊作者，國立台灣大學生物環境系統工程學系教授，10617台北市大安區羅斯福路4段1號，changfj@ntu.edu.tw

一、前言

台灣的自然環境水文、氣候條件都極具有獨特性，坡陡流急及降雨型態分佈不均，使得大部份雨水直接逕流入海，大部分的水都無法被利用，因此若要對河川流量做有效管理及永續經營以達到人類與生態系統共存的理想，彙整相關資料評估分析生態與水文氣候條件，是一個必要的過程與方法，藉由資料分析進一步推估生物多樣性，以作為日後生態環境復育決策的一個重要參考指標；一般來說，生物多樣性可分為三個層面討論，即物種多樣性(species diversity)、遺傳多樣性(genetic diversity)及生態系多樣性(ecosystem diversity)；其中物種多樣性被普遍認為最接近生物多樣性之基本涵義(廖，2000)，因此選定針對物種多樣性探討；本研究選取全台灣河川未受人為控制、流量資料長度大於二十年且其上下游10公里內有魚類調查資料之20個流量站作為研究測站，架構一個具有對河溪型態作分類，並可以推估魚類生物多樣性指標之類神經網路；在研究中對於溪流生態的討論係以河川流量流態為主，暫不考慮水質、水溫等因子的影響。

二、文獻回顧

Indicators of Hydrologic Alteration (IHA)為Richter *et al.* (1996)所提出，是現今使用最廣泛且最完整的水文指標，這些指標能夠具體呈現並有效描述集水區生態永續經營中，扮演關鍵角色的自然河川流態之特性，如河川流量的量值、頻率、延時、極端值發生的時間及改變速率等。「台灣生態水文指標系統(Taiwan Ecohydrology Index System, TEIS)」即依據Olden和Poff (2003)的研究成果為基礎，並參考Richter等人在1996年所發展之IHA架構，將河川流態的水文變數分為一般流量變數、高/低流量變數、頻率變數及時間變數指標等四大類別，此指標不僅考慮台灣獨特的水文氣候特性，其中更同時考量人類及河溪生態系統的需求為一兼顧生態與水文之指標系統(Suen *et al.*, 2004, Suen & Herricks, 2006)。

類神經網路應用於生態領域則有Parka等人(2003)應用反傳遞類神經網路對於水中巨型無脊椎動物進行生活型態的判別與預測，以物種豐富度(Species richness, SR)與歧異度指數(Shannon index, SH)為預測標的。在研究中所使用的資料為荷蘭地區23個不同之水域型態，如：江、河、溪流、湖…等地點所收集的資料，每筆資料包含34種環境因子變數，透過自組映射圖類神經網路的分類將23個不同水域型態之輸入資料分為五類，藉由分類結果之關聯性，即可以得知種的豐富度與歧異度指數；結果顯示，不管是在訓練階段或測試階段的物種豐富度與歧異度指數其觀測值與推估值的相關性有不錯的表現，顯示反傳遞類神經網路具有對生態資料進行型態判別與預測之能力。

三、理論概述

本研究以類神經網路推估溪流之生物多樣性，以下各節所用之理論依序介紹：(1)台灣生態水文指標系統；(2)生物多樣性指標；(3)集群分析概述；(4)類神經網路。

3.1 台灣生態水文指標系統(TEIS)

台灣生態水文指標系統(TEIS)係基於台灣特殊的水文及地文環境，並結合台灣魚類個體生態矩陣所提供的生態環境需求之資訊所發展，為一兼顧生態與水文之指標系統。TEIS為國內相關研究針對台灣河川生態(魚類)系統所提出的指標，已有多篇國內外論文說明其合宜性(Suen and Herricks, 2009; Chang *et al.*, 2008; Suen and Eheart, 2006)，本研究中選用TEIS有兩點原因，第一點，本研究對於溪流生態以河川流態為主，目的是希望透過既有的流量測站資料推估生物多樣性，達到方便快速推估的效果，而不必擔心現地調查資料受到氣候季節等限制；第二點，台灣生態水文指標系統另一重要的考量為了解魚類族群所偏好的生態水文環境，以魚類個體生態矩陣與水文因子之間的關係為基礎，並透過台灣魚類個體生態矩陣中的生物性因子(食物偏好、產卵時間(月份)與卵的型式等)及非生物性/物理環境因子(棲地型態偏好、流速偏好、遷徙習

性、產卵所需的流速等)將魚類族群透過這些因子所呈現的特性反應至水文指標中，找出符合該地河川生態環境之河川流態，因此TEIS與溪流生態是息息相關。台灣生態水文指標系統以季節性、颱風效應、用水之管理考量及魚類個體生態矩陣與水文因子之間的關係為基礎，將水文指標分為一般流量變數指標、高/低流量變數指標、頻率變數指標及時間變數指標四大類，TEIS中各指標如表3.1。

3.2 生物多樣性概述

生物多樣性指標是由族群中的種類數與個體數所構成，可以反應群聚的特性及功能，生物種組成的多樣性與群聚的穩定程度有密切關係。一般而言，均勻的分布表示其生物多樣性較好，而不均勻的分布則相反之(Stiling, 2002)。由於生物群落對環境因子變化較具敏感性，故可使用生物群落來了解集水區的環境狀況，本研究配合台灣生態水文指標系統，採用魚類作為生物多樣性推估之物種，且魚類位於集水區河川生態系統食物網的上層包括了多種營養階層，成為最廣泛應用的指標生物。Fisher *et al.* (1943)開啟了研究物種多樣性的先河，爾後生態學家常用兩種

指標來代表物種的多樣性，即Shannon (1948)提出的Shannon index，及Simpson (1949)所提出的Simpson index。其中，Simpson index是指在一個群聚之中，任意取得兩種生物個體的樣本，都為同種類的機率，而其指數對優勢種的變化較具敏感度。而Shannon index是指在一群聚中隨機選取一個體時，預測屬於某一特定種的不確定程度，此種不確定性會隨著種類數增加而增加。Simpson index與Shannon index最大的差異在於Simpson index在計算上忽略稀有種的存在，僅考慮到優勢種的變化，而Shannon index則對稀有種的變化較具敏感度，較重視社群中族群的相互比例。本研究範圍是以台灣地區為主，其中有許多台灣稀有種，最著名的稀有種為櫻花鉤吻鮭，在研究中所收集的魚類資料中，有收集到此魚種的資料，故在此採用Shannon index (*SI*)的概念，其公式如(3.1)所示：

$$SI = -\sum_{i=1}^N P_i \log_2 P_i \tag{3.1}$$

P_i 代表第 i 種生物出現的機率， N 為該群落的科別數。

表3.1 台灣生態水文指標系統

水文指標	
一般流量變數(量值，改變率，趨勢)	
各旬平均流量(36旬)	
枯水期平均流量上升速率(DR)、豐水期平均流量上升速率(WR)	
枯水期平均流量下降速率(DD)、豐水期平均流量下降速率(WD)	
高/低流量變數(量值，延時)	
枯水期單日最小值(D1MIN)、枯水期十日最小值(D10MIN)、枯水期三十日最小值(D30MIN)、枯水期九十日最小值(D90MIN)、枯水期單日最大值(D1MAX)、枯水期十日最大值(D10MAX)、枯水期三十日最大值(D30MAX)	
豐水期單日最小值(W1MIN)、豐水期十日最小值(W10MIN)、豐水期三十日最小值(W30MIN)、豐水期單日最大值(W1MAX)、豐水期三日最大值(W3MAX)、豐水期十日最大值(W10MAX)、豐水期三十日最大值(W30MAX)	
頻率變數(頻率)	
枯水期低流量事件(DLE)、豐水期低流量事件(WLE)	
枯水期高流量事件(DHE)、豐水期高流量事件(WHE)	
連續三年低流量事件(3LE)、連續三年高流量事件(3HE)	
枯水期水文歷線反轉數(DHyR)、豐水期水文歷線反轉數(WHyR)	
時間變數(時間，延時)	
枯水期單日最小值發生日、枯水期單日最大值發生日、豐水期單日最大值發生日	
枯水期低流量事件平均延時(DLD)、枯水期高流量事件平均延時(DHD)豐水期低流量事件平均延時(WLD)、豐水期高流量事件平均延時(WHD)	
豐水期：五月至十月	枯水期：十一月至隔年四月
低流量：平均日流量之0.25倍	高流量：平均日流量之2倍

3.3 集群分析概述

集群分析主要在對觀察體進行分類，期望分類後同一集群內觀察體的相似性愈高愈好，而不同集群間觀察體的相異性則要愈高愈好。在分類方法中除要達到上述之目標外，最重要的即為集群個數的決定，本研究利用三個常用於判斷群集個數的指標為依據(陳正昌，2001)，綜合推算群集的個數，其方法如下所述：

(1) R^2 (R-squared)：

R^2 代表集群間的異質性(Sharma, 1996)，若趨近於0，表示群體間幾乎沒有差異性，分群效果差；若趨近於1則表示分群效果佳。在變異數分析中， SS (sum of square)為聯合組內變異量，以離均差平方和表示，如公式(3.2)：

$$SS = \sum_{i=1}^p \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2 \quad (3.2)$$

其中 i 代表變項， j 代表觀察體。

SS 分為組間 SS_B (sum of square between)、組內 SS_W (sum of square within)及全體 SS_T (total sum of square)，其關係如(3.3)、(3.4)式所示：

$$SS_B + SS_W = SS_T \quad (3.3)$$

$$R^2 = SS_B / SS_T \quad (3.4)$$

由於集群分析主要目的在使集群內的變異較小、集群間的變異較大，因此若觀察體的合併，使得 R^2 突然減小，則表示應停止合併。

(2) 半淨 R^2 (semi-partial R-squared)：

代表由於加入新的觀察體到集群中而比上一個階段增加的異質性。即為每個集結階段比上一階段增加的組內 SS 與全體 SS 的比，如(3.5)式，即稱為半淨 R^2 。由於集群分析的目的在於將不同的觀察體合併而又不使集群內的異質性太大，因此若半淨 R^2 突然增加，就表示集結過程應該停止。

$$\text{半淨}R^2 = (|SS_W - SS_W'|) / SS_T \quad (3.5)$$

其中 $|SS_W - SS_W'|$ 為每個集結階段比上一階段增加的組內 SS 。

(3) 均方根標準差(root-mean-square standard

deviation, $RMSSTD$)：

是集群中所有觀察體的聯合標準差，代表新集群的異質性， $RMSSTD$ 可用來衡量經過分群變數分群後，群體內個體間相似程度，越小則表示群體內相似性越高，其公式如(3.6)式：

$$RMSSTD = \sqrt{SS / P} \quad (3.6)$$

P 為集群內所有變項的聯合自由度($\sum_{i=1..m} \sum_{j=1..n} (n_{ij} - 1)$)。

(4) K-Means

本研究對資料選擇分類個數時並無訓練目標，僅只單純的對資料作型態上的分類，故採用非監督式之K-Means聚類法，搭配上上述三種分類指標作為選定分類個數之理論依據；其K-Means理論概述如下：

將 n 筆輸入資料依照其資料特性聚類為 K 類，其中 K 為一正整數，其演算法的目的在於求取各個輸入資料與其相對應聚類中心點距離平方和之最小值，故我們可將目標函數定義為下式，以尋求目標函數最小值而找出每一聚類族群之聚類中心點。

$$J = \sum_{i=1}^k J_i = \sum_{i=1}^k \sum_{j=1}^n w_{ji} \|X_j - C_i\|^2 \quad (3.7)$$

其中， k 為聚類個數， X_j 為第 j 筆輸入向量， C_i 為第 i 個聚類中心(向量)， J_i 為第 i 類聚類之目標函數。 $w = [w_{ji}]$ 為 $n \times k$ 的二位元矩陣，如第 j 個輸入向量屬於第 k 個群集，則 $w_{ji} = 1$ 表示與輸入向量相對應的群集，其餘群集的矩陣單元 w_{ji} 為0。藉由中心點的調整與輸入向量資料的相對應之二位元權重矩陣 w ，逐步降低目標函數值以求得最小值或收斂值，即可獲得最佳中心點位置(張&張，2005)。其演算步驟如下：

(a) 隨機選取 k 個資料點 C_i , $i = 1, \dots, k$ ，並將之分別視為各群聚的初始聚類中心。

(b) 以3.8式決定各資料點所屬之群聚類別，式中若資料點 X_j 判定屬於 i 類別，則權重值 $w_{ji} = 1$ ，否則為0。

$$w_{ji} = \begin{cases} 1, & \text{if } \|X_j - C_i\| \leq \|X_j - C_m\|, \quad \forall m \neq j \\ 0, & \text{其他} \end{cases} \quad (3.8)$$

且滿足

$$\sum_{i=1}^k w_{ji} = 1, \forall j = 1, \dots, n, \quad \sum_{i=1}^k \sum_{j=1}^n w_{ji} = n \quad (3.9)$$

(c)由3.7式計算目標函數J，如果J保持不變，代表分群結果已經穩定不變，則可結束此迭代方法，否則進入步驟d。

(d)以3.10式更新群集的中心點。回到步驟b。

$$C_i = \frac{\sum_{j=1}^n w_{ji} X_j}{\sum_{j=1}^n w_{ji}} \quad (3.10)$$

本研究中對分類個數的選擇，是針對SAS軟體所得三種分類指標的結果與使用非監督式之K-Means聚類法結果綜合評估決定。

3.4 類神經網路

類神經網路源自於模擬人類腦神經系統運作的方式，擁有許多與人類大腦相同功能的特性，一般類神經網路分為三層：輸入層、隱藏層及輸出層。(張&張，2005)。

3.4.1 自組特徵映射網路(SOM)

自組特徵映射網路(Self-Organizing Map, SOM)於1982年首先由Kohonen提出。自組特徵映射網路演算是以特徵映射的方式，將任意維度的輸入向量，映射至較低維(度)的特徵映射圖上，如一維向量方式或二維矩陣方式排列形成拓樸層架構的映射圖(圖3.1為二維矩陣映射圖)，並且依據目前的輸入向量在神經元間彼此相互競爭，優勝的神經元可獲得調整連結權重向量的機會；而最後輸出層的神經元會依據輸入向量的「特徵」以有意義的「拓樸結構」(topological structure)(圖3.2表示神經元間的拓樸關係與其空間位置)展現在輸出空間中。從另一個角度來看，自組特徵映射的過程也是一個聚類的過程，其可將一群未經標示的樣本，透過演算，從中尋找某些相似的特性，然後再將這些具有相似特性的樣本聚集成一類。

3.4.2 輻狀基底函數類神經網路(RBFNN)

輻狀基底函數(Radial Basis Function, RBF)類神經網路，其特質主要在於模擬大腦皮質層軸

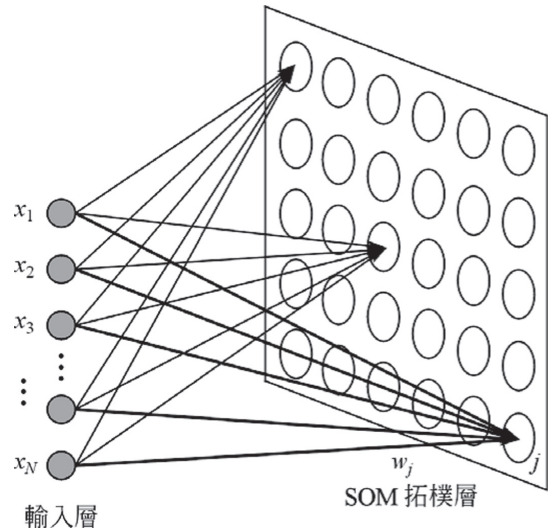


圖3.1 二維矩陣的SOM架構圖

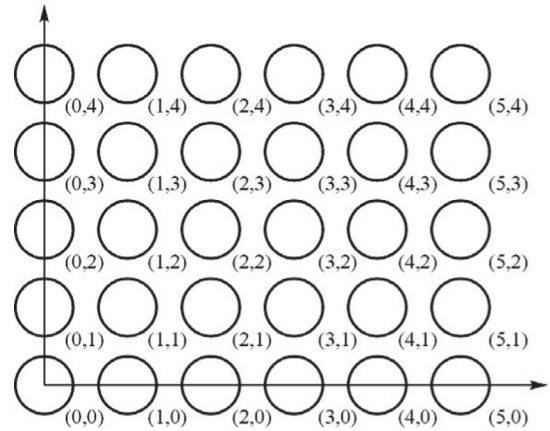


圖3.2 SOM網路神經元間的拓樸座標

突的局部調整功能，具備相當良好的映射能力；其架構與多層感知器相同，具有輸入層、一層隱藏層與輸出層，其優點在於可大量減少學習時間(圖3.3)；但不同於其他的類神經網路型態，RBF類神經網路是以函數逼近(curve fitting)的方式來建構網路，最早衍生於解決多變數的內插(interpolation)問題，並由Powell (1987)針對這個問題做了一些探討，以不規則位置的資料點，求得相對應的輻狀基底函數，之後陸續有許多學者也提出類似的架構，解決內插問題與近似的理論。RBFNN主要概念是建構許多輻狀基底函數 $\phi(\|x - c\|)$ ，以函數逼近法(curve fitting)找出輸入/輸出間的映射關係，也就是在隱藏層的

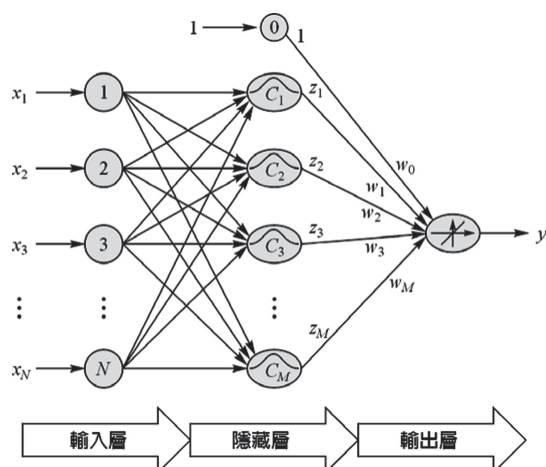


圖3.3 輻狀基底函數類神經網路架構圖

各神經元中建立相對應的輻狀基底函數。

3.4.3 自組特徵輻狀基底類神經網路(SORBNN)

複合式學習法即將網路分別以兩種不同的學習法來學習、調整連結權重；以輻狀基底函數類神經網路(RBFNN)為例，建構RBFNN一般採用複合式學習法，前階段以非監督式學習選取中心點，而後階段採監督式學習調整連結權重向量。在前階段非監督式學習中，採用中心點選取法，亦即在選取中心點的過程中，同時決定了隱藏層神經元的個數與位置。本研究即在RBFNN的前階段部份，使用自組特徵映射網路(SOM)取代原先的中心點選取法，透過SOM對資料特性做分類，並以拓模結構的中心點作為RBFNN之輻狀基底函數中心點，架構一自組特徵輻狀基底類神經網路(Self Organizing Radial Basis Neural Networks, SORBNN) (Moradkhani *et al.*, 2004)，此網路不但具有將TEIS指標依特性分類的能力，也具備了對生物多樣性做推估的能力。

四、研究案例

4.1 資料收集與處理

本研究以台灣本島河川未受人為控制、紀錄流量資料長度大於二十年且其上下游10公里內具有魚類調查資料之流量站作為研究測站。其中收集之資料分為三個部分，如下所述：

(1) 水文資料：

從經濟部水利署「水文水資源資料管理供應系統」中，收集符合本研究設定篩選條件之流量站，其面積及高程等地文資料。

(2) 流量資料：

經濟部水利署之「台灣地區河川流量資料庫」，計有430個流量測站，然而全台只有20個流量站符合篩選條件。並以經濟部水利署水利規劃試驗所「水資源最佳化管理於河川生態復育之研究(1/3)」計畫中所撰寫的程式將收集的流量資料轉換為TEIS。

(3) 魚類資料：

在魚類資料收集方面，主要為經濟部水利署水利規劃試驗所建制之「河川生態情勢調查資料庫」，目前所收集到之資料為2000~2004年，共有18條河川約160筆資料；魚類資料收集原則為周圍10公里的魚類調查資料收集整理成一筆資料，此筆資料即為流量測站的魚類資料。本研究以魚類之科技異度為代表探討生物多樣性，故採用Shannon's歧異度指數(Shannon and Weaver, 1949)，將這些魚類資料轉換，其計算公式如前述(3.1)式。

4.2 模式架構

在本研究中資料筆數只有20筆，因此僅將資料分為訓練資料與測試資料兩部分，並應用交叉驗證法分10組進行訓練與測試，以K-fold Cross-Validation之理論選用K = 10，即將資料分為10個子集合，每個子集合中的資料採用隨機取樣之方法，任選兩筆資料作為一個子集合，並對此網路架構訓練10次，使每個子集合都輪值過測試資料，每筆資料均會被當作測試資料1次、訓練資料9次。由於僅將資料區分為訓練資料與測試資料，缺少選定網路參數及決定網路架構大小(即決定隱藏層神經元個數)的驗證資料，為解決此問題，本研究應用三個常用於判斷群集個數的指標及K-Means聚類法，來決定網路架構大小(隱藏層神經元個數)，選擇分類個數大致上分為兩部分，先計算出各分類數的均方根標準差(root-mean-square standard deviation, RMSSTD)、 R^2 (R-squared)、半淨 R^2 (semi-partial R-squared)，從中選定較佳的

分類個數範圍，之後與K-Means聚類法比較，以確定分類個數。依這三個指標在群集分析時所代表的意義得知：若RMSSTD越小越好，表示群體內相似性高；半淨 R^2 也是越小越好，若突然增加，代表集結應該停止；而 R^2 值則是越大代表分類越佳。但分類個數之選取時仍要依據樣本數的多寡來決定，在小樣本時分類個數不宜太多，若以較少的分類數就能達到將樣本分類則屬較佳分類。

本研究的樣本屬於小樣本，由表4.1可知在10到19的分類範圍內，指標的半淨 R^2 都無突然增加之情況，因此僅討論2到9的分類個數，由圖4.1可以看出，2到4的分類個數為較佳的分類範圍；分類個數為4類時， R^2 與半淨 R^2 表現都更佳，RMSSTD值也很小，各指標表現都很好，似乎是很好的分類個數，但是RMSSTD由4類變3類時所增加的量(集群內個體之間的異質性)並沒有比由3類變2類時所增加的量來的大，因此若要以最少的分類個數來達到較佳的分類效果，也許分3類時就已足夠將資料做較佳的分類，故本研究將針對3及4類之分類數，再透過K-Means聚類方法來確定。

一般而言判斷分類結果之優劣，會隨著應用範圍不同而有所不同，在本研究中利用K-Means聚類方法來確定分類個數原則為依聚類後每個類別是否有明顯之區分表現作為判斷標準，如圖4.2最佳分類示意圖，分成三類時各類別指標之間並無明顯重疊之情況，而分成四類時第三類與第四類並無法明顯區分開來，故藉由此判斷標準，將分類後各類別之台灣生態水文指標分別繪出，以判斷最佳分類數；結果顯示除了少數

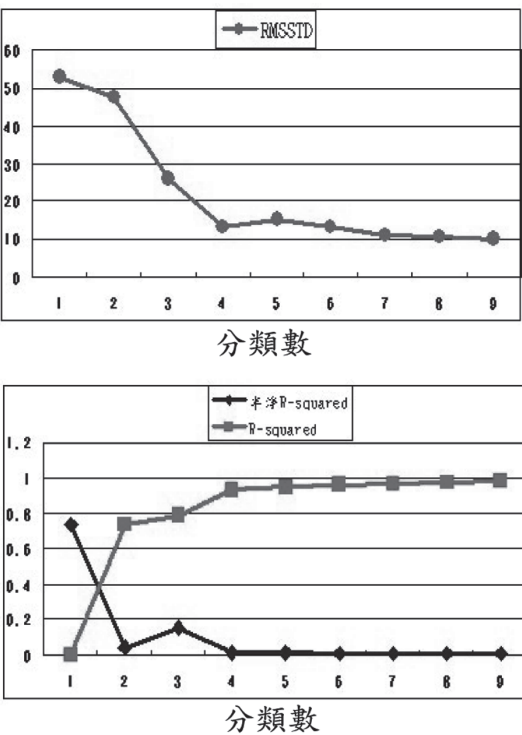


圖4.1 分類評比指標

幾個生態水文指標無法被明顯的區分外，使用K-Means聚類方法將資料區分為三類時為最佳如圖4.3所示，因此在本研究中決定將資料分為三類，即網路架構的隱藏層神經元個數決定為三個；其對模式的具體幫助為：(1)決定了網路大小，確定適當的中心點個數有效過濾雜訊(2)模式依輸入資料的相關性將其分類，故每個中心點的初始位置不再是隨機選擇，而是依資料特性產生，對網路收斂速度及穩定性有很大的幫助(3)選取最佳分類數為三類，代表將這20筆資料分成

表4.1 分類評比指標報表

分類個數	RMSSTD	半淨 R^2	R^2	分類個數	RMSSTD	半淨 R^2	R^2
19	3.051	0.0002	1.000	9	10.090	0.0071	0.983
18	4.064	0.0004	0.999	8	10.759	0.0037	0.980
17	4.738	0.0006	0.999	7	11.136	0.0088	0.971
16	4.008	0.0003	0.998	6	13.380	0.0096	0.961
15	4.233	0.0003	0.998	5	15.165	0.0110	0.950
14	6.001	0.0011	0.997	4	13.424	0.0118	0.938
13	7.567	0.0026	0.994	3	25.943	0.1531	0.785
12	6.569	0.0020	0.992	2	47.624	0.0426	0.743
11	6.871	0.0009	0.992	1	52.962	0.7427	0.000
10	7.898	0.0012	0.990				

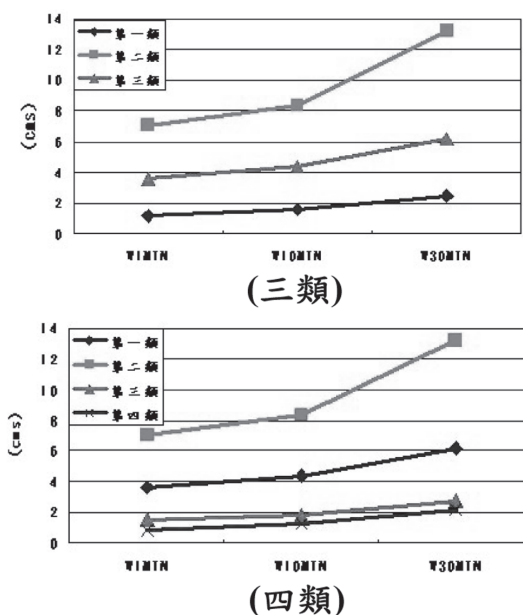


圖4.2 選擇最佳分類示意圖

三類是最理想的分類狀況，有最佳的聚類中心點，就類神經網路來說，正確的中心點數目及有依據的中心點初始位置，其推估結果也會較佳。

4.2.1 自組特徵輻狀基底類神經網路架構(SORBNN)

在本研究中所採用的類神經網路為結合自組特徵映射網路(SOM)與輻狀基底函數類神經網路(RBFNN)如圖4.4所示，其架構由輸入層、一層隱藏層(拓樸層)及輸出層結合而成，並採用複合式學習法建構此網路，稱為自組特徵輻狀基底類神經網路(SORBNN)，整個網路分為兩階段(圖4.5)，第一階段利用非監督式學習法之SOM將輸入資料依特性分類，資料依其特性分類後，會以有意義的拓樸結構展現在拓樸層中，在本研究中選擇將資料分為三類，故SOM拓樸層中的神經

元個數為三個，每個SOM拓樸層的神經元代表一個聚類，而每一聚類都會有一個聚類中心點；而第二階段則是架構RBFNN，採用監督式學習法來推求網路輸出值，將第一階段拓樸層之聚類中心點作為第二階段RBFNN中輻狀基底函數的中心點，本研究以高斯函數為輻狀基底函數，透過所建構的輻狀基底函數，以函數逼近方法找出輸入/輸出間的映射關係，即計算輸入向量與隱藏層各神經元中心點的距離後，經由函數轉換獲

得隱藏層各神經元的輸出，再經加權傳至輸出層，即可求得網路輸出值。

4.2.2 模式評比指標

評比指標可對網路模式判斷其優劣，因此本研就選用兩個常見的評比指標來評估SORBNN推估能力，其定義如下：

(1) Correlation Coefficient (CC)

$$CC = \frac{\sum_{i=1}^N [H_{pre}(i) - \bar{H}_{pre}(i)][H_{obs}(i) - \bar{H}_{obs}(i)]}{\sqrt{\sum_{i=1}^N [H_{pre}(i) - \bar{H}_{pre}(i)]^2} \sqrt{\sum_{i=1}^N [H_{obs}(i) - \bar{H}_{obs}(i)]^2}} \quad (4.2)$$

(4.2)式中 N 為資料筆數， $H_{pre}(i)$ 為推估之科技異度， $H_{obs}(i)$ 為實際之科技異度， $\bar{H}_{pre}(i)$ 為推估科技異度之平均值， $\bar{H}_{obs}(i)$ 為實際科技異度之平均值，當 CC 越接近1時，表示模式推估能力精確度越高。

(2) Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^N [H_{pre}(i) - \bar{H}_{obs}(i)]^2}{N}} \quad (4.3)$$

$RMSE$ 可以評估推估模式之優劣，當其值越小則代表其準確性越高。

五、結果討論

將TEIS指標分別對流量站之面積及高程，計算其相關係數，結果如表5.1所示。從表5.1中可發現，面積與一般流量變數及高/低流量變數共18項水文指標，其相關係數大部份皆高於0.75以上，而頻率變數及時間變數等12項變數，其指標與面積、高程之間都沒有很高的相關性。

在SORBNN第一階段所架構的SOM網路具有依資料特性進行聚類的的能力，因此本研究將符合篩選條件的20個流量測站之TEIS指標值輸入SOM網路中，對資料分為三類，依各分類結果將流量測站之科技異度、集水區面積及高程整理如表5.2所示。

第一類其地文因子特色為集水區平均面積是三個分類中最大且平均高程最高，科技異度之

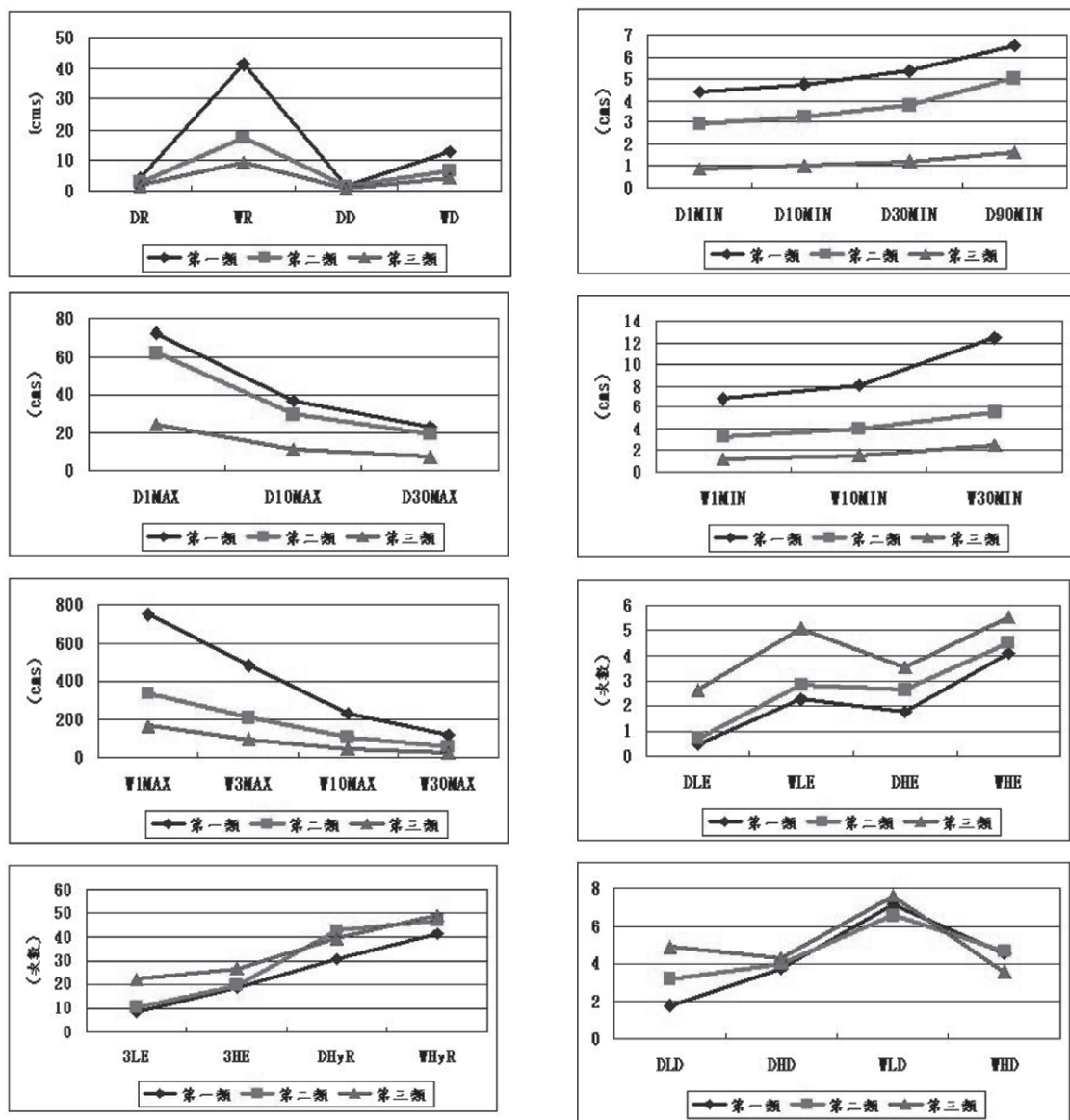


圖4.3 K-Means分為三類之台灣生態水文指標

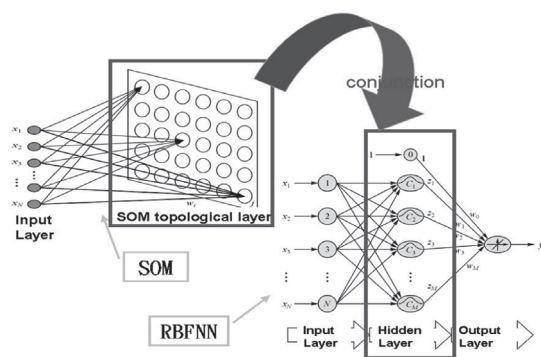


圖4.4 SOM結合RBFNN示意圖

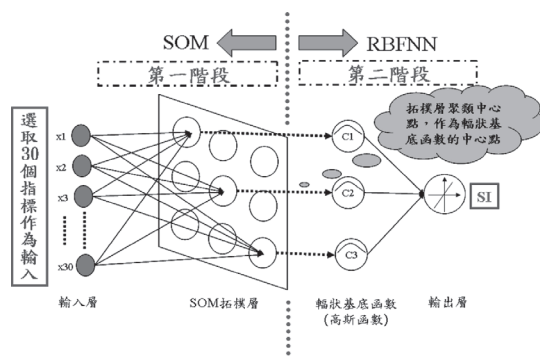


圖4.5 自組特徵輻狀基底類神經網路(SORBFNN)架構圖

表5.1 面積及高程與TEIS指標之相關係數

一般流量變數	面積	高程	高／低流量變數	面積	高程
DR	0.712	0.191	D1MIN	0.653	0.06
WR	0.825	0.083	D10MIN	0.649	0.05
DD	0.637	-0.028	D30MIN	0.649	0.046
WD	0.825	-0.121	D90MIN	0.675	0.058
頻率變數			D1MAX	0.838	-0.047
DLE	-0.231	-0.125	D10MAX	0.878	0.047
WLE	-0.237	-0.283	D30MAX	0.832	0.073
DHE	-0.302	-0.008	W1MIN	0.764	0.208
WHE	-0.239	-0.33	W10MIN	0.788	0.196
3LE	-0.2	-0.2	W30MIN	0.811	0.211
3HE	-0.295	-0.219	W1MAX	0.83	-0.032
DHyR	-0.119	-0.298	W3MAX	0.877	0.018
WHyR	-0.213	-0.401	W10MAX	0.914	0.088
時間變數			W30MAX	0.929	0.116
DLD	0.055	-0.274	*面積：km2 高程：m		
DHD	0.119	0.381			
WLD	0.016	-0.143			
WHD	0.379	0.368			

表5.2 分類後測站之相關資料

第一類-3站					
流域	魚類資料測站	流量測站	科技異度	集水區面積(km2)	高程(m)
卑南溪	初來橋	新武呂(4)	0.00	638.78	316
清水溪	瑞草橋	桶頭(2)	0.00	259.2	214
濁水溪	龍神橋(中)	內茅埔	2.28	367.4	475
平均值			0.76	421.79	335
第二類-8站					
流域	魚類資料測站	流量測站	科技異度	集水區面積(km2)	高程(m)
頭前溪	南昌大橋	上坪	2.00	221.73	211
蘭陽溪	牛鬥橋	牛鬥	2.2	446.74	216
和平北溪	北溪第1樣點	和平北	0.81	272.35	157
和平南溪	南溪第1樣點	和平南	1.00	190.2	189
花蓮溪	馬太鞍溪上游	光復	0.92	136.36	171
花蓮溪	馬太鞍溪上游	馬鞍溪橋	0.92	136.46	166
坪林溪	包尾	南崗大橋	1.84	253.58	75
濁水溪	春陽(上)	櫻社	2.16	188	1006
平均值			1.48	230.68	273.88
第三類-9站					
流域	魚類資料測站	流量測站	科技異度	集水區面積(km2)	高程(m)
大漢溪	興漢橋	稜角	2.16	107.76	525
頭前溪	比麟大橋	內灣	1.46	139.07	249
後龍溪	汶水橋	彼岸橋	2.06	109.95	226
大安溪	鯉魚潭1橋	鯉魚潭	1.92	53.45	2
北港溪	崙子大橋(下)	溪口	2.13	129.38	12
八掌溪	行圓橋(上)	觸口	1.42	83.15	221
秀姑巒溪	馬蘭鉤溪橋	富源	1.45	83.12	194
大安溪	雪山坑橋	雪山坑	0.92	45.32	598
水里溪	13哩	水里	1.92	55.46	302
平均值			1.72	89.63	258.78

值為三個分類中最小，表示此分類裡的流量站其生物多樣性並不高，並沒有很好的環境提供多種不同科的魚類生存，因此可知地文因子對生物多樣性是有很重要的影響。在此區由於高程較高，屬於較高海拔之溪流段，其生存環境較為嚴苛，可能只適合少數喜愛高海拔環境的魚種生存，故此區的科技異度為三類中最小。

第二類中，地文因子平均值則介於第一類與第三類之間，而科技異度亦介於三類中間且比第一類還高，推論原因可能為此區的高程比第一類低，其海拔高度較低，生存環境較第一類佳。

第三類的集水區平均面積及平均高程皆為所有分類中最小，推論其一般流量變數與高/低流量變數之值亦為最小，故在此分類的流量測站不但不會有太劇烈的流量變化，其高流量值亦不會太大，並且此分類流量測站位處於較低海拔地區，生存環境應屬於三類之中較適合多數魚種生存的，因此在第三類之科技異度為所有分類中最高，有較豐富的生物多樣性。

從圖5.1可清楚的驗證TEIS中一般流量變數與高/低流量變數指標與生物多樣性的相關性，在第一分類中，集水區平均面積是三個分類中最大，其一般流量變數與高/低流量變數皆為最高，流量變化急速且高流量值都很大，急速的流量變化可能會把水中的生物給沖走而造成大量死亡及生物遷徙情形的變動，對魚類生存有較大的威脅，故科技異度最小，生物多樣性並不豐富；而第三類中一般流量變數與高/低流量變數皆為最小，因此流量變化較平緩，高流量值亦不會太大，故不會對魚類生存產生嚴重威脅，所以科技

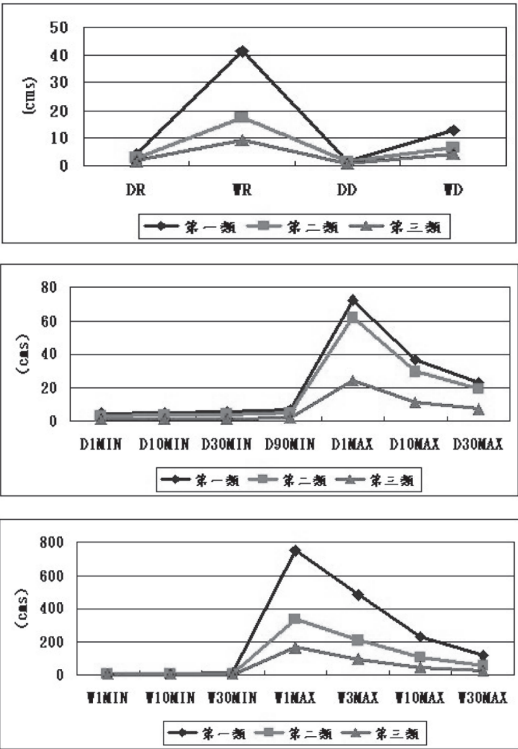


圖5.1 各分類之一般流量與高/低流量變數指標

異度在第三類最大，生物多樣性最豐富。

本研究之SORBNN架構是利用20個流量測站資料計算出台灣生態水文指標系統中的30個指標為網路輸入項推估測站之科技異度，其結果如表5.3。

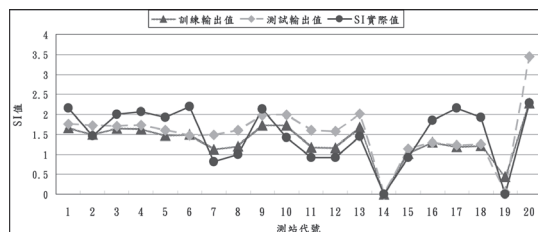
如表5.4所示，可得知SORBNN在訓練部份RMSE為0.419，相關係數高達0.82，而在測試部分雖然RMSE增加為0.538，但其相關係數仍有

表5.3 訓練、測試輸出及實際值

流量測站	棱角	內灣	上坪	彼岸橋	鯉魚潭	桶頭(2)	內茅埔
訓練推估值	1.66	1.49	1.65	1.63	1.47	0.43	2.29
測試推估值	1.75	1.72	1.70	1.73	1.59	0.05	3.44
科技異度實際值	2.16	1.46	2.00	2.06	1.92	0.00	2.28
流量測站	光復	馬鞍溪橋	富源	新武呂(4)	雪山坑	櫻社	水里
訓練推估值	1.17	1.16	1.68	0.00	1.01	1.19	1.21
測試推估值	1.60	1.57	2.00	0.00	1.13	1.23	1.26
科技異度實際值	0.92	0.92	1.45	0.00	0.92	2.16	1.92
流量測站	牛鬥	和平北	和平南	溪口	觸口	南崗大橋	
訓練推估值	1.50	1.13	1.21	1.73	1.73	1.31	
測試推估值	1.47	1.48	1.61	1.97	1.98	1.29	
科技異度實際值	2.20	0.81	1.00	2.13	1.42	1.84	

表5.4 模式評比指標結果比較

Training		Testing	
RMSE	CC	RMSE	CC
0.419	0.82	0.538	0.66



測站名稱依序為：1.棱角 2.內灣 3.上坪 4.彼岸橋 5.鯉魚潭 6.牛鬥 7.和平北 8.和平南 9.溪口 10.觸口 11.光復 12.馬鞍溪橋 13.富源 14.新武呂(4) 15.雪山坑 16.南崗大橋 17.櫻社 18.水里 19.桶頭(2) 20.內茅埔

圖5.2 模式訓練、測試輸出值及實際值

0.66，由圖5.2可更清楚的看出網路推估值與實際值的關係，雖然每個點無法皆準確推估，但整體來說，各測站生物多樣性的推估結果皆相當不錯。在模式中RMSE的表現不佳，但由於本研究以相關係數之表現較為重要，因為環境生物多樣性的推估屬於較大範圍、大尺度，故可以容忍較大誤差，並不像洪水預報模式需要精準的推估其洪水量值，否則將會對生命財產有嚴重的影響，並且計算科技異度實際值需要依靠現地調查來獲取資料，而現地調查的準確性則會因調查方式或天候環境影響等因素而有誤差，因此就CC及RMSE兩評比指標來看，SORBNN對生物多樣性(魚類之科技異度)的推估具有非常不錯的效果。

六、結 論

本研究利用流量資料計算台灣生態水文指標，架構可有效推估溪流生物多樣性(魚類科技異度)之SORBNN，所得結論如下：

- (1)研究結果顯示水文、地文與生態三者是息息相關的；台灣河川流態除明顯受水文氣象影響外，集水區面積、高程亦對河川流態影響甚大，較大的集水區面積會有較大的流量變數，而急速的流量變化可能將水中的生物給沖走而造成大量死亡及生物遷徙情形的變動，導致魚類食物來源減少或魚類本身的死

亡；而集水區面積較小的測站流量較小，相對的不會有急速的流量變化等惡劣生存條件，故生物多樣性較佳。

- (2)TEIS指標分別對流量站之面積及高程計算其相關係數，發現面積與一般流量變數及高/低流量變數共18項水文指標的相關係數大部份皆高於0.75以上，證明流量變量與流域面積的高度相關性。
- (3)SORBNN將20個流量測站依其30個台灣生態水文指標值建立關聯性，藉由此關聯性將20站分成3大類，可以有效推估流量測站的生物多樣性。
- (4)在網路輸出結果部分，訓練階段的科技異度實際值與網路推估值之相關係數高達0.82，而測試階段的實際值與推估值表現上，相關係數仍然有0.66，其結果顯示SORBNN除了具有對流量資料進行型態判別分類的能力外，也具備推估河溪生態中魚類生物多樣性之能力，故在日後於溪流生態系統管理上，可以利用本研究所架構之SORBNN推估生態狀況作為參考依據。

謝 誌

本研究承蒙經濟部水利署水利規劃試驗所計畫(ISBN-978-986-00-7663-9)經費補助及相關資料提供，使得以順利進行，僅此致謝。

文 獻 回 顧

1. 廖大經，2000，「生物物種多樣性指標熵值的估計方法及模擬研究」，國立清華大學統計學研究所碩士論文。
2. 陳正昌，2001，「觀察體的分類」，教育學門提昇研究方法論研討會。
3. 張斐章、張麗秋，2005，「類神經網路」，東華書局。
4. 張斐章，2004，「生態水文於多目標水資源管理之研究與應用(1/2)」，經濟部水利署水利規劃試驗所期末報告 ISBN-957-01-9429-4。
5. 張斐章，2005，「生態水文於多目標水資源管理之研究與應用(2/2)」，經濟部水利署水利規劃試驗所期末報告 ISBN-986-00-3534-2。
6. 張斐章，2006，「水資源最佳化管理於河川生態復育之研究(1/3)」，經濟部水利署水利規劃試驗所期末報告 ISBN-978-986-00-7663-9。

7. Chang Fi-John, Tsai Meng-Jung, Tsai Wen-Ping, Edwin E. Herricks, 2008, "Assessing the ecological hydrology of natural flow conditions in Taiwan", *Journal of Hydrology*, 354, 75-89.
8. Fisher, R. A., Corbet, A. S., and Williams, C. B., 1943, "The relation between the number of species and the number of individuals in a random sample of an animal population", *Journal of Animal Ecology*, 12, 42-58.
9. Kohonen, T., 1982, "Self-Organizing formation of topologically correct feature maps", *Biological Cybernetics* 43, 59-69.
10. Moradkhani H., Hsu Kuo-lin, Gupta Hoshin V., Sorooshian Soroosh, 2004, "Improved streamflow forecasting using self-organizing radial basis function artificial neural networks." *Journal of Hydrology* 295, 246-262.
11. Olden, J. D. and Poff, N. L., 2003, "Redundancy and the choice of hydrologic indices for characterizing streamflow regimes", *River Res. Applic.*, 19: 101-121.
12. Parka Young-Seuk, Verdonshotb Piet F.M., Chonc Tae-Soo, Lek Sovan., 2003, "Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network", *Water Research*. 37: 1749-1758.
13. Powell M. J. D., 1987, "Radial Basis Functions For Multivariable Interpolation: A Review." IMA Conference on Algorithms for Approximation of Function and Data. RMCS, Shrivenham, England. 143-167.
14. Richter, B. D., Baumgartner, J. V., Powell, J., Braun, D. P., 1996, "A method for assessing hydrologic alteration within ecosystems", *Conservation Biology* 10: 1163-1174.
15. Shannon, C. E., and W. Weaver., 1949, "The Mathematical Theory of Communication", Urbana: University of Illinois Press.
16. Shannon, C. E., 1948, "A mathematical theory of communication", *Bell System Technical Journal*, 27, 379-423, 623-656.
17. Sharma, S., 1996, "Applied multivariate techniques", NY: John Wiley & Sons, Inc.
18. Simpson, E. H., 1949, "Measurement of diversity", *Nature*, 163, 688.
19. Stilling, P., 2002, "Ecology Theories and Applications", Prentice Hall, 273-288.
20. Suen, J.P. and Eheart, J.W., 2006, "Reservoir management to balance ecosystem and human needs: incorporating the paradigm of the ecological flow regime", *Water resources research*, 42(3): W03417.
21. Suen, J. P., Herricks, E. E. and Eheart, J. W., 2004, "Ecohydrologic Indicators for Rivers of Northern Taiwan", ASCE/EWRI World Water & Environmental Resources Congress 2004, Salt Lake City, Utah, USA. Paper in CD ROM.
22. Suen, J. P. and Herricks, E. E., 2009, "Developing fish community based ecohydrological indicators for water resources management in Taiwan", *Hydrobiologia*, 625: 223-234.

收稿日期：民國 98 年 01 月 22 日

修正日期：民國 98 年 03 月 24 日

接受日期：民國 98 年 05 月 08 日