

類神經網路映射模型之人工濕地水質生化需氧量 去除率分析

陳鈞華⁽¹⁾、宋建明⁽²⁾、鄭富尚⁽³⁾、徐貴新⁽⁴⁾、甘俊二⁽⁵⁾

⁽¹⁾屏東科技大學土木工程系副教授、水利技師

⁽²⁾屏東科技大學土木所博士班

⁽³⁾屏東科技大學土木所碩士班

⁽⁴⁾東南科技大學環境管理系副教授兼研發處處長、水利技師

⁽⁵⁾七星農田水利研究發展基金會董事長、水利技師

摘 要

本研究目的為預測花蓮縣玉里鎮大禹里地表面流人工濕地(surface-flow constructed treatment wetlands)之污染質BOD₅去除率，研究期間(2006年9~1月,2007年3~6月)研究方法乃應用Kohonen網路映射模型(Kohonen network model) ”self-organizing map” 結合主成分分析(principal components analysis,PCA)和倒傳遞類神經網路(back-propagation neural network,BPN network)。其過程係經由Kohonen網路模型之自組織映射圖網路(self-organizing map, SOM)技術，將本研究中複雜之水質資料由7維向量，自動映射降維至2維向量群集，將此2維向量群集水質資料，以PCA找出群聚的水質參數的累積之最大變異量，將其構成BOD₅主成分之參數應用於BPN之輸入層中。簡言之，經由SOM-PCA-BPN步驟後，可快速收斂和避免局部最小值，準確預測出流水BOD₅去除率。

關鍵字：人工濕地，生化需氧量，SOM，主成份分析，倒傳遞類神經網路。

一、研究動機目的

自我組織特徵映射類神經網路(self-organizing map,SOM)理論模型，為一種模擬生物大腦之非監督學習式類神經網路運算法，SOM網路具有將資料維度”縮減”或稱”映射” Teuvo Kohonen(1982)，由高維度(輸入)樣本空間以非線性映射運算到低維度二維(輸出)特徵空間，換句話說，SOM運算模型，是以簡單的特徵性質描述原本複雜原始資料型態。Kohonen (1995)以拓撲鄰域(Topological neighborhood)和定義一個核

(Kernel)函數，提出改良式類神經網路學習法，修正只有唯一能獲勝之神經元擁有學習之權力的Winner-Take-All運算模型，Kohonen推導出來之運算模型，主要特色為神經元在競爭後不僅只有獲勝的一個神經元獲得權重學習，周圍其他的神經元也獲得不同權重可以學習。Kohonen et al.(1996a)，SOM理論配合Matlab軟體寫出軟體套件，達到高維度可視化統計數據圖形。(Kohonen et al., 1996; Lee and Scholz, 2006; Kalteh et al., 2007).SOM 基

於一個非監督的類神經網路演算法，可運用於分析大型數據資料庫，有效率的劃分成不同集群。其中重點為在原始資料映射成特徵過程中，其轉換信號之拓撲關係維持不變。由此特性，SOM被廣泛運用至多種複雜數據分析領域。其中，運用於濕地水質分類，Liang Zhanga et al.(2008)運用SOM，預測一個農業生態系統人工濕地(ICWs)之逕流排放的氨氮和可溶性活性磷的濃度，在中，以SOM模式分析得知，鹽分濃度低和較高的溫度時，可有效去除氨氮。可溶性活性磷(Soluble reactive phosphorus,SRP)去除率主要受到鹽份和溶氧量(dissolved oxygen,DO)的影響，溫度和pH值與SRP去除率呈現弱相關。延續此研究結論繼續討論其他營養質去除率主要受到哪些因素影響。Carroll et al.,(2005)ICWs為自由表面流式人工濕地，設

計理念基於有效率使用土地，包括設計適宜的景觀造物 and 維持生物多樣性以及水體和營養質流入和流出的適當控制。Liang Zhanga et al.(2008)，以SOM評估綜合型人工濕地(ICWs)之脫氮、除磷功能 Lee, B.-H., Scholz, M.,(2006)，運用SOM分析評估重金屬去除表現在人工濕地。由以上SOM運用於表面流人工濕地之水質分析(包括氨氮、可溶性活性磷)研究文獻，延伸出本研究所要探討之重點為運用SOM探討表面流人工濕地之水質分析(包括入流量、SS、VSS、TCOD、BOD₅、NH₃-N、PO₄³⁻)。接續主成分分析得知，對於BOD₅的主要因子為何，接著以徑路分析得知，入流量、SS、VSS、TCOD、BOD₅、NH₃-N、PO₄³⁻彼此的直接和間接相關係數，得到一個表面流人工濕地各污染質之間的定量關係模式。

二、材料及方法

2.1 實驗場地

本試驗田區(玉里鎮禹東段36號地目田面積一·0一七三公頃)位於花蓮縣玉里鎮大禹里，與玉里鎮中心相鄰。秀姑巒溪位於本區東側，本區東西寬八00公尺，南北長四000公尺，對外交通非常便利。台灣省水利局陸續興建完成秀姑巒溪大禹堤防計一三五0公尺，東部土地開發處配合延建完成二六五0公尺，開發土地二四0公頃。

2.1.1 計畫需水量

全區開發可得耕地淨面積一九二公頃，其中五四公頃水稻田，成田耕地已由水利會配水灌溉，餘一三八公頃配合承購戶多角化經營，以最大水田需水量推估灌

溉用水，即整田田間需水一八公分，日數二十日，本田灌溉率二00ha/cms，混凝土給水渠道輸水損失採二0%，需水量為一·0四二CMS。經洽水利會同意自該會所屬大禹峻末端引用0·八CMS，以埋設涵管1.8M長度九00公尺，橫跨卓溪引入本區供水外，不足數0·二四二CMS引至秀姑巒溪水源補注。

2.1.2 地下流人工濕地設計單元

地下流人工濕地設計單元係利用加強污染質入滲率，來作為小面積地區去除污染質之淨化設施，其構造為以大顆粒砂石(3-6公分)填塞之溝渠，在砂石與土壤間置有過濾纖維或濾布(filter fabric)，作為減少對地下水之污染及土

壤細顆進入砂石間空隙，造成砂石間淤塞(clogging)。其剖面構造可參考(圖2.1)地下流人工濕地設計單元(Schueler, 1987)^[16]，其對污染質去除之原理包括重力沉澱、吸附等，當污染質藉雨水進入土壤中，因物理(沉澱、吸附)、化學(反應)及生物(根部攝取、轉換)等作用而被去除。其設計單元為長條狀梯形斷面，以抽水機截取溪水分流後分別流進，人工濕地底層先鋪設不透水塑膠布，每道長50 m，石層上底淨寬1.98m，下底寬0.45 m，總高1.0m。最底層放置3支4" ϕ PVC排泥管，礫石層高0.5 m(排水管上方鋪設20cm厚5" ϕ 卵石後；上方再鋪設15cm厚3" ϕ 碎石層；最上層鋪設15cm厚1" ϕ ~2" ϕ 礫石)，每道總體積38.4 m³，平均孔隙率0.457，有效體積17.54m³，有效面積49.5m²。石床上方種植空心菜。實驗自1998年11月至1999年5月，以及1999年11月至2000年5月兩個枯水期操作，進流量1.7~127.6 m³/day，水力停留時間HRT為120~1.6 hrs，水力負荷從0.034~2.57 m³/m²-d。

2.2 水體採樣分析

定期採樣分析入流量(Q)、生化需氧量(Biochemical Oxygen Demand, BOD₅)、總化學需氧量(Total Chemical Oxygen Demand, TCOD)、懸浮固體物(Suspended Solids, SS)、VSS(揮發性懸浮固體)、氨氮(NH₃-N)、正磷酸鹽(PO₄³⁻)。分析方法，依照環境保護署環境檢驗所公告的標準方法[2]

2.3 統計分析工具

本研究圖表使用Matlab R2007b, R, SPSS13軟體。統計資料型態表達方式，其顯著性(p<0.05)。

2.4 自我組織神經網路映射模型

SOM是一個神經網路模型和算法，其實現了非線性特徵向量的預測，可從高維空間的資料型態或其他輸入資料型態降維到一個低維數空間，並已廣泛運用於三維可視化系統和數據採礦^{[3][4][5]}。SOM屬於一種競爭學習神經網路和無監督學習的模式。^[7]

2.4.1 參數設定及建立網絡

首先建立SOM網絡，網絡數量依照(Vesanto et al., 2000)^[12]的建議， $M \approx \sqrt{n}$ ，M=神經元數目，n=樣本數目。本研究樣本數目為350將建立，一個二維分隔數為[4x5]，為六角網格，輸入變化範圍分別為[0 1]、[0 2]。接續根據Kohonen學習規則，要同時修正獲勝神經元附近區域 $N_{i^*}(d)$ 內所有神經元，其中， $N_{i^*}(d)$ 包括獲勝神經元以d為半徑區域內的所有神經元。如下為調整公式，依此加以訓練SOM網路：

$$\begin{aligned} i^{w(q)} &= i^{w(q-1)} + \alpha(p(q) - i^{w(q-1)}) \\ i^{w(q)} &= (1 - \alpha)i^{w(q-1)} + \alpha p(q) \end{aligned} \quad (1)$$

其中， $N_i(d) = \{j, d_{ij} \leq d\}$

其中，神經元間距離為歐幾里德距離，如下：

$$D = \sqrt{\sum_{i=1}^n (x_i - y_j)^2}, X = (x_1, x_2, \dots, x_n), Y = (y_1, y_2, \dots, y_n) \quad (2)$$

2.4.2 SOM 模型訓練

SOM為單層神經網路，分為輸入層與輸出層(自組織映射層)，主要訓練階段分為兩階段，一為排列順序階段，其此階段以指定的步數，將排序階段學習率降至調整階段之學習速率。二為調整權值階段，此階段是對神經元的權值進行調整，調整中，網路的鄰域距離將調成排列順序將不

會再改變，則其網路學習率將會慢慢收斂。

本研究排列順序學習率為0.9，排列順序步數為1000，調整學習率為0.02，調整鄰域距離為：1。依照輸入矩陣P、激發矩陣 A_2 、學習率lr，可得到權值變化矩陣值，其公式如下：

$$\Delta W_{(i,j)} = lr * A_2 * (P_{(i)} - W_{(i,j)}) \quad (3)$$

其中，激發矩陣 A_2 經由輸入矩陣A、目前鄰域距離nd和神經元距離D而定。

$$a_2(i, q) \begin{cases} = 1, a(i, q) = 1 \\ = 0.5, a(i, q) = 1, D(i, j) \leq nd \\ = 0, other \end{cases} \quad (4)$$

2.5 主成分分析 (principal components analysis, PCA)

主成分分析也稱主分量分析，旨在利用降維的思想，把多指標轉化為少數幾個綜合指標。本研究目的在使個指標，轉化為每一個成分能夠代表最大的觀察變異量，第一個主成分為觀察變項的線性整合，能夠反應最大的變異量，依序發展各主成分可以得到最大的解釋變異量，其以下為主成分分析之步驟：

1、原始指標數據的標準化採集 p 維隨機向量 $x = (x_1, x_2, \dots, x_p)^T$ n 個樣品 $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ ， $i=1, 2, \dots, n$ ， $n > p$ ，構造樣本陣，對樣本陣元進行如下標準化變換：

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, i=1, 2, \dots, n, j=1, 2, \dots, p \quad (5)$$

$$\text{其中 } \bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}, s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}, \text{得標準化陣 } Z。$$

2. 對標準化陣 Z 求相關係數矩陣

$$R = [r_{ij}]_p \quad xp = \frac{Z^T Z}{n-1} \quad (6)$$

$$\text{其中, } r_{ij} = \frac{\sum Z_{kj} \cdot Z_{ki}}{n-1}, i, j=1, 2, \dots, p。$$

3. 解樣本相關矩陣 R 的特征方程

$|R - \lambda I_p| = 0$ 得 p 個特征根，確定主成分

$$\text{按 } \frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.85 \text{ 確定 } m \text{ 值，使信息的利用}$$

率達 85% 以上，對每個 $\lambda_j, j=1, 2, \dots, m$ 解方程組 $Rb = \lambda_j b$ 得單位特征向量 b_j^o 。

4. 將標準化後的指標變數轉換為主成分

$$U_{ij} = Z_i^T b_j^o, j=1, 2, \dots, m$$

U_1 稱為第一主成分， U_2 稱為第二主成分， \dots, U_p 稱為第 p 主成分。

5. 對 m 個主成分進行綜合評價對 m 個主成分進行加權求和，即得最終評價值，權數為每個主成分的方差貢獻率。

在統計學中，主成分分析 (principal components analysis, PCA) 是一種簡化數據集的技術。它是一個線性變換。這個變換把數據變換到一個新的坐標系統中，使得任何數據投影的第一大方差在第一個坐標(稱為第一主成分)上，第二大方差在第二個坐標(第二主成分)上，依次類推。主成分分析經常用減少數據集的維數，同時保持數據集的對方差貢獻最大的特征。這是通過保留低階主成分，忽略高階主成分做到的。這樣低階成分往往能夠保留住數據的最重要方面的方法。

2.6 徑路分析

徑路分析中有 m 個預測變項，則可以找到 m 個迴歸方程式及 $[m(m+1)]/2$ 條徑路途徑，而其標準化迴歸係數即為徑路係

數。徑路係數其為預測變項影響效標變項的大小。其統計模式共有 m 個徑路模式（迴歸方程式）如下表示：

$$\begin{aligned} X_2 &= \beta_{12}X_1 + e_2 \\ X_3 &= \beta_{13}X_1 + \beta_{23}X_2 + e_3 \\ &\vdots \\ X_m &= \beta_{1m}X_1 + \beta_{2m}X_2 + \dots + \beta_{(m-1)m}X_{m-1} + e_m \\ Y &= \beta_{1Y}X_1 + \beta_{2Y}X_2 + \dots + \beta_{mY}X_m + e_Y \end{aligned} \quad (7)$$

2.7 倒傳遞分析

本研究採用倒傳遞網路作為模擬診斷污染水質評估的模式，原因是主要應用在樣本識別分類問題函數合成適應控制雜訊過濾資料壓縮專家系統，最大特點是可解XOR問題、應用最普遍、學習精度高、學習速度慢、回想速度快。倒傳遞類神經網路具有一層的輸入層、一層或一層以上的隱藏層、以及一層輸出層；輸入信號以前饋方式輸入層傳向輸出層，其中以「多層感知機 (multilayer perceptrons)」或稱為「倒傳遞類神經網路 (backpropagation networks)」最為著稱。

「倒傳遞演算法」的網路訓練方式包含兩個階段：前饋階段以及倒傳遞階段。在前饋階段時，輸入向量由輸入層引入，前饋方式經由隱藏層傳導至輸出層，並計算出網路輸出值，此時，網路的鍵結值是固定的；而在倒傳遞階段時，網路的鍵結值則根據錯誤更正法則來進行修正，藉由鍵結值的修正，以使網路的輸出值趨向於期望輸出值，更明確地說，我們以期望（正確）輸出值減去網路輸出值以得到誤差信號，然後將此誤差信號倒傳遞回網路中，因此我們將此演算法稱之為「倒傳遞演算法」。倒傳遞演算法，為一種有系統

的多層感知 (Multi-Layer Perceptron, MLP) 訓練方法，被廣泛運用在醫療預測作業。圖5為代表一個有隱藏層的MLP，此結構顯示在輸入層有 i 個節點，隱藏層有 h 個節點，輸出層有 t 個節點，實際上輸出值 Y 在倒傳遞網路的輸出節點上完成，如公式 $Y = f(X*W)$ ， Y 代表輸出向量， X 為輸入 (row) 向量， W 為權重向量 (包括偏權值)， $f(\cdot)$ 為一個活化函數，這個活化函數 $f(\cdot)$ ，將輸入值的總和轉換成輸出值節點，選擇典型的活化函數是由邏輯、正切、訊號及線性所組成，而邏輯函數如公式， $Y = f(X*W) = 1 / (1 + e^{-(X*W)})$ 。

實際輸出值 Y 與目標值 (欲達成的值) 可比較如本研究中圖4，誤差值在目標值與由一些錯誤的預測所產生的網路實際輸出值之間，例如預測的總和平方。倒傳遞網路主要的目的在於調整權重向量直到降低錯誤到一個令人滿意的程度。研究者使用倒傳遞網路相關的參數前必須找到一個理想的結構。此作業牽涉到倒傳遞網路中定義的數值 (例如：活化函數、誤差預測函數、多層感知結構及特殊的訓練方法)。顯示出在倒傳遞網路中要決定一個適當的結構和參數是不容易的。有很多學者提出不同的學習方式稱作學習法則 (Learning Rule)，在此說明常用的基本學習法則 (Basic Learning Rule)。它主要是用梯度降階法 (Gradient Descent Method) 及連鎖律 (Chain Rule)，求出各結點的導函數，由此導函數來決定結點內的參數以及網路中的加權值。基本學習法則是 P. Werbos 在 1974 年提出的，又稱作 Delta 學習法則或梯度降階學習法則 (Gradient Descent Learning Rule)。

三、結果與討論

3.1 自我組織神經網路映射模型與徑路分析比較

本研究中經由SOM方法得圖2(Final quantization error: 0.846; Final topographic error: 0.000)由圖中視覺化判斷可得知。依7種取樣污染質，進行相關分析，取得有95%信賴區間下，達顯著水準之連續變項，排除線性關係及交互作用，得本研究人工濕地之污染物標準化之相關徑路係，由圖3表示之。

- (1)圖2得知，當Q高時，SSr中，VSSr低，TCODr低，BOD₅低，NH₃Nr中，PO₄³⁻r低。經圖2辨別後可得定性的基礎判斷。再由徑路分析中得知Q與NH₃Nr關係係數-0.31，與上述定性符合。
- (2)圖2得知，當Q低時，SSr高，VSSr中高，TCODr高，BOD₅中高，NH₃Nr高，PO₄³⁻r高。

四、結果與建議

1. 由SOM視覺化圖型中可以有效率地，分類出七種污染質之間的定性關係。
2. 表面流型人工濕地中入流量(Q)、NH₃N去除率、PO₄P去除率對於將污染質BOD₅去除率，具有高相關性，可經由路徑分析得知直接和間接相關係數，得到定量

當NH₃Nr中高，PO₄³⁻r部分中高。

再經由徑路分析中得到相關係數+0.33，與上述定性符合。

- (3)當BOD₅r中高與PO₄-Pr部分中高。經由徑路分析中得到相關係數+0.55，與上述定性符合。

3.2 倒傳遞類神經模擬 BOD₅

本研究輸入層為入流量(Q)、SSr、TCODr、NH₃-Nr、PO₄³⁻r(r代表去除率)，輸出層為BOD₅去除率，重點為，先正規化各參數(mean=0.0, stddev=1.0)，以Random方法將各參數分成兩等分(N=72，以5:5比例分配，N=36為訓練樣本，N=36為測試樣本)，即可得到本研究模擬出高方法。以倒傳遞類神經模擬分析BOD₅，其模擬後之準確度與原始BOD₅比較後得，R²=0.9987, (Hidenu11=21, Slope=3.2000, Radius=0.1000, DistMean=1.7729)。

關係。

3. 經由倒傳遞類神經網路，利用主成分分析出之入流量(Q)、NH₃-N去除率、PO₄³⁻視為輸入層，可以有效率地預測出BOD₅去除率，

五. 圖表範例

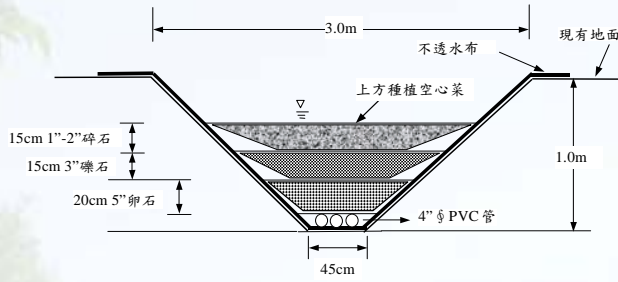


圖 1 人工濕地設置(資料來源：郭文健研究室)

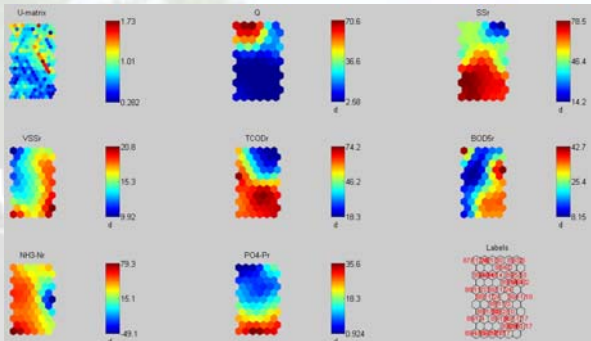


圖 2 使用 SOM 於入流量 (Q)、生化需氧量 (Biochemical Oxygen Demand, BOD₅)、總化學需氧量 (Total Chemical Oxygen Demand, TCOD)、懸浮固體物 (Suspended Solids, SS)、VSS (揮發性懸浮固體)、氨氮 (NH₃-N)、正磷酸鹽 (PO₄-P) 的視覺化圖

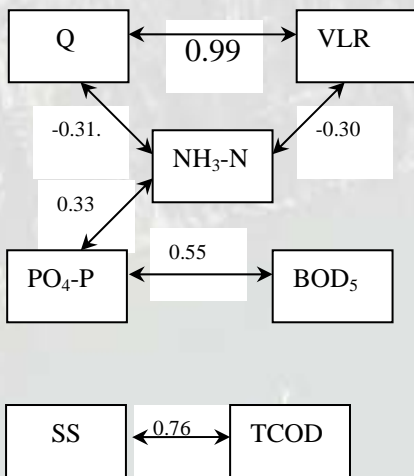


圖 3 玉里試驗區表面流人工濕地標準化之徑路關係係數圖

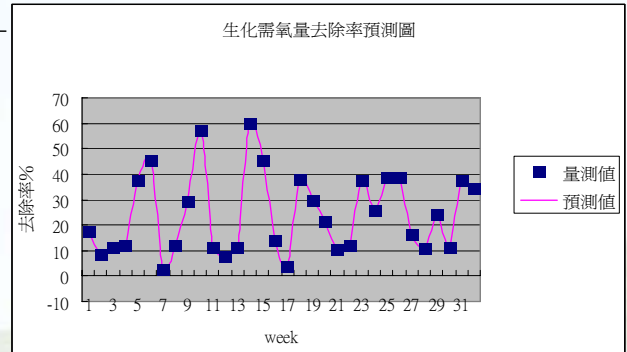


圖 4 玉里試驗區表面流人工濕地生化需氧量去除率估算、預測圖

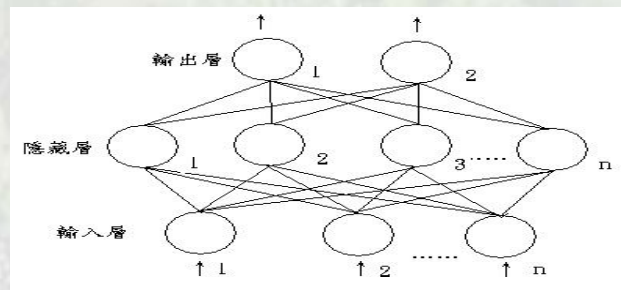


圖 5 一個有隱藏層之倒傳遞類神經網路架構 (MLP)

六. 參考文獻

1. Application of the self-organizing map as a prediction tool for an integrated constructed wetland agroecosystem treating agricultural runoff
2. Assessment of the nutrient removal performance in integrated constructed wetlands with the self-organizing map
3. Carroll, P., Harrington, R., Keohane, J., Ryder, C., 2005. Water treatment performance and environmental impact of integrated constructed wetlands in the Anne Valley watershed, Ireland. In: Dunne, E.J., Reddy, K.R., Carton, O.T. (Eds.), *Nutrient Management in Agricultural Watersheds: A Wetlands Solution*. Wageningen Academic Publishers, Wageningen, The Netherlands.
4. Kalteh, A.M., Hjorth, P., Berndtsson, R., 2007. Review of the selforganizing map (SOM) approach in water resources: analysis, modelling and application. *Environ. Model. Softw.* (in press). Available online since 19 November 2007.
5. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59-69.
6. Kohonen, T. (1995). *Self-Organizing Maps*. Springer, Berlin.
7. Kohonen, T., Hynninen, J., Kangas, J., and Laaksonen, J. (1996a). SOM_PAK: The self-organizing map program package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo.
8. Kohonen, T., Oja, E., Simula, O., Visa, A., Kangas, J., 1996. Engineering applications of the self organizing map. *Proc. IEEE* 84 (10), 1358–1384.
9. LE´ OPOLD SIMAR, (2003), *Detecting Outliers in Frontier Models: A Simple Approach*, Kluwer Academic Publishers. Manufactured in The Netherlands., *Journal of Productivity Analysis*, 20, 391–424.
10. Lee, B.-H., Scholz, M., 2006. Application of the self-organizing map (SOM) to assess the heavy metal removal performance in experimental constructed wetlands. *Water Res.* 40, 3367–3374.
11. Schueler, T. R. (1987). "Controlling Urban Runoff : A Practical Manual for Planning and Design Urban BMPs." Metropolitan Washington Council of Governments Washington
12. *Standard Methods for the Examination of Water and Wastewater*, 1992, 18th Ed., APHA-AWWA-WEF, Washington, D.C
13. Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. SOM Toolbox for Matlab 5 Documentation. Helsinki University of Technology, Helsinki, Finland Available online at [/http://www.cis.hut.fi/projects/somtoolbox/S](http://www.cis.hut.fi/projects/somtoolbox/S).
14. 宋建明、陳鈞華、郭文健(2005)，休耕水田復育成生態濕地之水質淨化分析研究。
15. 黃俊英(1995)，多變量分析，(第五版)，華泰書局。
16. 楊浩二(1995)，多變量統計方法，(第一版)，華泰書局。

收稿：100年5月4日
修改：100年5月7日
接受：100年5月11日