
* 論 著 *

類神經網路於區域地下水砷濃度的推估

Artificial Neural Networks for Estimating Arsenic Variation in the Regional Ground Water

高 力 山

LI-SHAN KAO

國立台灣大學

生物環境系統工程學系

博士班研究生

吳 子 青

TZU-CHING WU

國立台灣大學

生物環境系統工程學系

碩士班研究生

劉 振 宇

CHEN-WUING LIU

國立台灣大學

生物環境系統工程學系

教授

張 斐 章*

FI-JOHN CHANG

國立台灣大學

生物環境系統工程學系

教授

摘 要

致癌毒物—砷對環境與人類健康之影響係為社會所關切的議題，本研究架構砷濃度類神經網路推估模式，對區域內砷濃度進行推估工作，提供區域內砷濃度變化，以維護當地民眾使用地下水之安全。台灣西南沿海地區有嚴重砷污染地下水情形，水利署於民國81年至94年間於雲林沿海地區，共設置28口水質監測井，蒐集此區域水質資料作為架構模式之用，在過程中發現本區域雖有長期水質監測資料，惟資料時常發生零星遺漏情形，為解決此問題，本研究以雲林沿海地區為研究範例，架構類神經網路模式推估各水質站內遺漏砷濃度資料，並對沿海地區進行區域砷濃度推估工作。另一方面，由於本區域內水質站之資料不足，造成模式之穩定性不佳，無法建構可靠補遺推估模式，為改善此問題，首先採用交叉驗證以確認模式架構後，並採修正型目標函數搜尋模式中參數，其結果證明此兩種方法可改善模式不穩定性及過度訓練情況，擴大類神經網路之使用範圍及應用領域，提供可靠砷濃度空間推估結果，對於了解此區域內砷濃度在時間與空間變化有很大助益，同時依據此結果可減少居民誤飲用含高濃度砷地下水之危險，達地下水有效管理及利用之目的。

關鍵詞：砷，地下水，人工智慧，類神經網路，修正型目標函數。

ABSTRACT

With the great concern for the potential effects of arsenic on human health and the environment, there is a growing need for efficiently determining and modeling the presence and amount of arsenic in the ecohydrogeological systems. In this study, we construct the ANNs model to complete the lost arsenic data according to the relationship of the arsenic concentration of the monitoring wells in the region. The results offer the realization for the arsenic variation and keep the safe from the resident ingesting and usage of the groundwater. Blackfoot disease was once common on the southwestern coast of Taiwan, especially in the alluvial fan of Chou-Shui River. In order to monitoring the arsenic concentration, the Water

* 通訊作者，國立台灣大學生物環境系統工程學系教授，106臺北市大安區羅斯福路4段1號，changfj@ntu.edu.tw

Resource Agency has setup twenty-eight water quality wells which distribute in coastal area in Yun-Lin county from 1992 to 2005. However, some water quality data were lost or not record because of the man-made and money factors. The lost data affect the development of the arsenic associate research. Therefore, we choose this region as the example to construct ANNs model to estimate the lost arsenic data.

Due to sparse and lost data, there are not steady and over-fitting problem in the process of constructing the ANNs model. To solve the problems, we first apply cross-validation to assure the architecture of the model and adopt the modify-objective function to search the optimal weights. It is proven that these two methods can reduce the unsteady and over-fitting problems. The results have been apparently improved the realization of the spatial-temporal distribution of arsenic. Based on the results, the risk of ingesting the high arsenic groundwater can be decreased to reach the goal of efficiently controlling and usage of the groundwater.

Keywords: Arsenic, Groundwater, Artificial Intelligence (A.I.), Artificial Neural Networks (ANNs), Modify-objective function.

一、前言

致癌毒物砷對環境與人類健康之影響長久以來一直為社會所關切的議題，而砷分佈、生成與運移涉及複雜生物化學環境改變、水文條件變遷及岩心物質砷濃度等十分難以掌握，為有效監控地下水中水質變化情形，架設可靠水質網為最直接方法，由於地下水水質站網設置涉及經費及取樣次數限制，水質資料常有缺漏情形，需藉由補遺模式將資料缺漏填補，過去補遺工作多利用線性模式藉由已知站資料補遺未知站之資料；若空間資料未呈現線性關係，常會發生推估及補遺誤差過大情形，且線性加權方式通常僅能解決簡單空間推估問題，對於高度不均勻且複雜空間關係，採用線性模式將產生很大誤差，因此，本研究以類神經網路，架構砷污染補遺推估模式，填補水質站網中砷濃度資料缺漏，藉由此架構有效掌握砷在地下水層中空間變化，並有效管理地下水抽取量，以利控管雲林縣沿海地區有限之地下水資源，其雲林縣沿海集水區分佈如圖1。

經濟部水利署鑒於雲林沿海地區之地下水污染具擴散趨勢，已嚴重影響沿海居民之健康，於民國81起執行台灣地區地下水觀測站網整體計畫(1992~2008)，在雲林離島式基礎工業區，設置地下水質監測網(1992~2005)，設置監測井28口，其中21口淺水井，7口深水井，每年約四次水質取樣分析，然因人為或經費考慮，本資料於81年至88年間，砷濃度採樣有零星缺漏情形，至



圖1 雲林沖積扇位置圖

88年後部份水質站，更因經費縮減而停止採樣，這些資料缺漏對研究砷污染之擴散及後續對於砷污染與地下水面或降雨量空間分布之相關性研究有重大影響。

類神經網路的應用相當廣泛，包括圖形辨識、智慧型控制、污染量預測、降雨及逕流預測模式均有良好表現(張和張，2005)，本研究採用倒傳遞類神經網路，對雲林縣沿海地區砷濃度進行空間補遺推估工作，運用過去資料作訓練與驗證，架構類神經網路推估模式，期使補遺推估結果與過去資料產生資料結構一致性，由於水質資料取得較困難，資料筆數較少，應用倒類神經網路常發生過度訓練情形，影響補遺及推估準確性甚大，為有效減少過度訓練發生，本研究採用不同目標函數及交叉驗證方法，改善網路穩定性，準確推估砷在地下水中變化。

二、文獻回顧

根據前人研究過去嘉南沿海地區發生嚴重的烏腳病事件，其主因可能為飲用水中含有較高砷濃度引起(曾文賓，1976)；王榮德等人(1983)應用統計分析出嘉南沿海地區的烏腳病是由於飲用當地深井地下水引起，並在地下水抽出時檢測出高濃度的砷；而雲林地區沿海地區歷年地下水水質檢驗成果(成功大學水工試驗所，1991~1998)發現，雲林四湖鄉沿海區域砷濃度皆有偏高現象。在水質因子相關研究方法方面，Ruiz *et al.* (1990)利用因子分析研究沿海地下含水層的水文地質化學並分析離子主要來源；廖少威(1991)利用主成份分析法探討台灣灌溉水質受工業、都市、畜牧廢水污染之變異性。

在空間推估上，近年來較多研究採用克利金方法(Kriging method)，但由於該法建構在線性假設及一階與二階定常性，限制了使用範圍，惟其具較強解釋空間變異之能力，因此，仍有許多相關研究採用此一方法，如Liu *et al.* (2004)利用指標克利金方法在不同的門檻值下推估雲林地區砷的空間變化及風險，結果顯示很高機率(92%)在淺層含水層(<180 m)的砷污染超過水質安全標準。Liu *et al.* (2006)藉由當地地層鑽井取出的樣本進行分析發現在地下含水層及限制層同樣發現高濃度的砷，而且岩心樣本的砷和地層所屬的地質年代有很高的相關性。Lin *et al.* (2006)利用multivariate factorial kriging方法分析雲林地區砷的空間變化，並指出砷主要來源為地層中的鐵氧化而被地下水沖洗出來，而砷的分佈主要受鋁濃度的影響，並受沿海海水入侵的影響；另也發現在自然界中砷主要以五價砷型式，但無機砷佔總砷比率約為3~8.3%，比較吳郭魚體內、水體及沉積物中以沉積物中砷濃度最高，甚至可高至1,000倍或100,000倍之水體砷濃度。Nolan *et al.* (2003)應用達西公式及地質狀態估計地下水補注量，並應用克利金方法，推估未設站處地下水補注量，研究中並採用克利金方法評估不同地區之地下水補注量差異，研究中依據補注量與污染濃度關係，發現在農業區內隨地下水補注量增加，硫酸根離子濃度明顯下降，但於舊市區及未開發區內，硫酸根離子濃度卻維持相同水準。

類神經網路應用水文量及水質推估相當廣泛，在國內張斐章等人(1998)即採用反傳遞模糊類神經網路應用於河川流量推估。張和孫(1997)應用倒傳遞類神經網路於降雨及逕流過程之研究。近年來，Chang *et al.* (2001, 2002, 2006)更採用不同類神經網路架構於不同水文量之推估與研究，均獲致可靠水文量預測與推估成果，如2001年應用模糊倒傳遞類神經網路對河川流量推估，2002年應用即時回饋類神經網路預測河川流量，研究中並對各種網路架構之優劣進行探討，提供後續研究之參考依據。郭益銘(1999)應用多變量統計分析與類神經網路分析雲林沿海地區地下水質變化。盧和駱(1998)採用類神經網路於污染物之空間推估之研究。在國外研究方面，Gumrah *et al.* (2000)考量石油抽取後大量廢鹽水注入地下水層造成地下水污染，採用氯離子與地下水水位之過去資料作為訓練，架構類神經網路，有效預測鹽水擴散情形，並可預測地下水水位變化情形，可有效控制地下水污染與使用；Martos *et al.* (2002)採用SOM推估地下水質推估變化；Coppola *et al.* (2003)考慮狀態變數、氣候及抽水條件，架構類神經網路預測多層地下水中地下水位變化，並將類神經模式與物理模式之數值方法相比較，證實本法可精確預測複雜之地下水變化。

三、理論說明

3.1 類神經網路

類神經網路主要的基本概念為模仿人類腦神經系統，因此在邏輯的組成單元與架構上皆與生物神經網路類似，其中最重要的三項功能為學習、回想及歸納推演。而類神經網路經過了長時間的發展，已有許多不同的網路模型被提出，常見的網路架構有倒傳遞類神經網路、輻狀基底函數網路及自組織特徵圖網路等，因其具有高度訊息處理與非線性映射的能力，可解決物理模式難以描述與建構之非線性關係，因此被大量地應用於水文領域上。

倒傳遞類神經網路為發展最早且應用最廣之網路，其架構為一多層前饋式網路，可分為輸入、隱藏及輸出三層(如圖2所示)。其使用監督

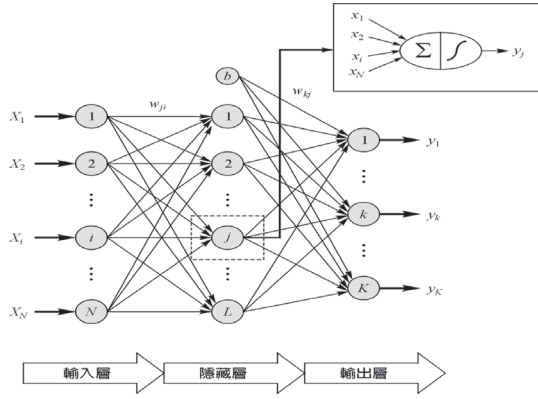


圖2 倒傳遞類神經網路架構圖

式學習法，藉由網路輸出值與目標輸出值間之誤差來修正模式，因此網路學習的過程即為使誤差函數 E 達到最小值。

在倒傳遞類神經網路中，第 n 層第 j 個神經元的輸入值為第 $n-1$ 層神經元輸出值的非線性函數：

$$y_j^n = f(\text{net}_j^n) \quad (1)$$

式中 y_j^n 為第 n 層的輸出值，對第一層來說為輸入值， f 為活化函數， net_j^n 為第 $n-1$ 層輸出值的權重累加值，可表示為(2)式：

$$\text{net}_j^n = \sum_i w_{ji}^n y_i^{n-1} + b_j^n \quad (2)$$

式中 w_{ji}^n 為第 n 層第 j 個神經元與第 $n-1$ 層第 i 個神經元的連結權重， b_j^n 為第 n 層第 j 個神經元的偏權值。由於倒傳遞類神經網路屬監督式學習，其學習的目的在降低網路輸出值與目標輸出值間的差距，誤差函數 E 定義為

$$E = (1/2) \sum_k (d_k - y_k)^2 \quad (3)$$

式中 d_k 為第 k 個神經元的目標輸出值， y_k 為輸出層第 k 個神經元的網路輸出值。因此，網路的學習過程，即為使誤差函數 E 最小化的過程，在此方面前人研究提供不同演算法搜尋 E 的最佳解。

在完成倒傳遞演算法流程後，尚有許多網路架構問題須決定，例如如何決定輸入層個數，選用何種因子作為輸入層、選用何種活化函數、隱藏層個數及選擇何種演算法等均會影響網路輸出誤差。

在活化函數選用上，過去研究較常使用對數雙彎曲活化函數、正切彎曲活化函數、線性轉移活化函數，使用者可根據問題須要選用較適當活化函數。至於隱藏層神經元個數的常用方法有兩種：其一是使用網路修剪法(pruning algorithm)由Abrahart *et al.* (1998)提出，另一種方法則是網路增長法(constructive algorithm)是在1997年由Kwok和Yeung提出。另一重要問題即是演算法選用，在過去較常使用為梯度下降法與動量的梯度下降法，但是這兩種演算法對於實際問題求解太過耗時，近年來提出一些較高效率之演算法，第一類為啟發式演算法，第二類為標準的數值最佳化，其中啟發式演算法又可分為1.可變學習速率(Variable Learning Rate)及2.有彈性的(Resilient)倒傳遞演算法；而數值最佳化可分為共軛梯度、擬牛頓及Levenberg-Marquardt演算法，在此要決定何種演算法需考慮問題的複雜度、訓練數據多寡、網路中權重值及誤差目標值大小，因此，各種演算法須配合整體網路架構，以找尋最佳化演算法。

3.2 交叉驗證法

交叉驗證法(cross validation)常應用資料過少時校驗網路推估值之良莠，其在類神經網路應用上，為每筆資料交替使用在訓練時期及驗證時期，由於本區之砷濃度資料過少，若由人為選擇訓練與驗證資料，在架構類神經網路時，其人為選擇結果將顯著影響模式表現優劣，為消除人為效應，本研究採用交叉驗證法，其計算步驟說明如下。

- (1) 假設 N 個(本研究 N 為22)已知地下水水質測站，每個測站 T 筆砷濃度資料(本研究 T 為28筆)，砷濃度有共計 $N \times T$ (22×28)筆資料。
- (2) 將各水質站 T 筆資料切割為 W 組(本研究 W 為7)，每組內具相等資料筆數(本研究中每組內資料為4筆)，並將每組資料分別編號(本研究中編號為 $W1, W2 \dots W7$)，每次由 $W1 \sim W7$ ，中選擇一組為驗證組，另外六組為訓練資料。
- (3) 在訓練階段，採用 i 號站除外之其餘 $N-1$ 站(本研究 $N-1$ 為21)之資料路輸入項(共21水質站為輸入神經元，每站共24筆)，第 i 號水質站

為網路輸出項(共24筆)，訓練網路中各參數值，也就是網路具有21個輸入神經元，1個輸出神經元，共有24筆資料作訓練，在完成訓練後，藉由演算法得到網路中最佳連結參數，提供驗證模式使用，且在完成網路訓練後，即再將原輸入層資料投入訓練模式中，則可得到網路輸出值，並計算與實際值之差，即為訓練誤差(rmse)。

- (4) 驗證階段，首先假設i號水質站其一驗證段內4筆砷濃度為未知，以其餘N-1站水質站同一時期編號之砷濃度為已知，分別推估i號水質站之4筆砷濃度，網路具有21個輸入神經元，及1個輸出神經元，但僅有4筆資料作驗證，將網路輸出值與i號站實際值作比較，即可得到驗證誤差(rmse)。
- (5) 在完成i號站作為W1驗證模式之資料後，重複(2)~(4)項之工作，直到每一組都曾被選取為驗證組，此時即完成第i號井之驗證工作。
- (6) 接續進行選取各站之砷濃度為網路輸出神經元之工作，重複步驟(3)至步驟(5)，直至22個水質站井均完成交叉驗證工作，則停止整個交叉驗證工作。

整體交叉驗證過程，資料充份混合及重新排序，可得到整體網路較穩定之參數值，比較交叉驗證時，訓練與驗證之誤差，即可了解模式之穩定性及可靠性。

3.3 修正型目標函數

為減少過度訓練情形，過去相關研究提出改善過度訓練方法(羅華強，2005)，主要可分為三類，分別為降低隱藏層數目、修改目標函數及提早停止模式，本研究嘗試採用修改目標函數改善過度訓練情形，在修改目標函數上，一般倒傳遞類神經網路採用式(4)作為目標函數及評估模式之表現，本研究考量採用修正型目標函數式(5)，而式(5)為結合式(6)及式(7)，其中 t_i 為目標輸出值， a_i 為網路輸出值， w_j 為網路中連結權重， γ 為性能比，使用此修正型目標函數將造成網路有較小的權重值，並且強制網路之輸入項與輸出項之影響性平滑化，有效減少過度訓練之機會及網路不穩定情形，如果設定較大，對改善過度訓練效果較差，模式穩定度不佳，但對於極

端值預測卻較準確，相反地，如果性能比較小，較能減少過度訓練情形，但卻需犧牲極端值之推估精確度，因此，在採用修正型目標函數時，須兼顧精確度與減少過度訓練發生，採用合適之性能比。

$$rmse = \sqrt{\frac{1}{N} \sum_{i=1}^n (e_i)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^n (t_i - a_i)^2} \quad (4)$$

$$msereg = \gamma mse + (1 - \gamma) msw \quad (5)$$

$$mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (6)$$

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (7)$$

四、研究案例

本研究針對砷在水文地質系統中的空間時間變化，架構適宜的砷濃度類神經網路補遺推估模式並探討砷濃度變化情形。

1. 研究區域現況

本研究以臺灣西南沿海之濁水溪下游沖積扇(雲林縣)為例。雲林縣位在台灣的中南部，在嘉南平原最北端，北邊沿著濁水溪和彰化縣接壤，東西最寬的地方有50公里，南北最長的地方有38公里，全縣面積總計1,290平方公里，其中十分之九為平原，十分之一為山地，人口約73萬5千人。

在雲林離島基礎工業區及工業區附近，其水質監測井分佈於台17、雲129及雲3公路；於口湖、四湖、台西與麥寮等四個濱海鄉鎮，本研究共收集21口淺層水質監測井(編號#1~#21)資料，以及收集水利處(現改名為水利署)在許厝寮、新興、飛沙及口湖四處，共7口深水質監測井(編號#22~#28)資料，其28口觀測井座標如表1，其相對於雲林縣位置如圖3，本研究所蒐集資料時間長度為民國81至94年間調查的地下水質資料，每站約具10~50次調查資料。其中地下水質調查項目包括砷、溫度、pH、電導度、溶氧、鹼度、TDS、氯鹽、硫酸鹽、游離氨氮、硝酸鹽氮、亞硝酸鹽、COD、總有機碳、BOD、鎘、鉻、

表1 雲林沿海地區28口觀測井座標資料表

站號	站名	X	Y	站號	站名	X	Y
#1	(口湖鄉成龍)	163913	2606249	#15	(台西鄉)	168938	2625517
#2	(口湖鄉成龍)	163124	2605124	#16	(台西鄉三姓寮)	171097	2626709
#3	(口湖鄉臺子)	162785	2607417	#17	(台西鄉蚊港)	169435	2628182
#4	(口湖鄉厝子)	163875	2609319	#18	(麥寮鄉)	171634	2629150
#5	(口湖鄉金湖)	162364	2608682	#19	(麥寮鄉後安寮)	170709	2631869
#6	(口湖鄉下崙)	163550	2611887	#20	(麥寮鄉中山)	172359	2632279
#7	(口湖鄉下崙)	163814	2614327	#21	(麥寮鄉許厝寮)	171855	2633618
#8	(口湖鄉箔子寮)	162095	2614628	#22	(許厝寮井)	171324	2634277
#9	(四湖鄉廣溝厝)	162831	2616185	#23	(新興淺井)	169728	2627209
#10	(四湖鄉林厝寮)	163455	2618581	#24	(新興深井)	169808	2627209
#11	(四湖鄉三條崙)	164462	2617465	#25	(飛砂淺井)	164841	2616491
#12	(四湖鄉林厝寮)	165379	2619664	#26	(飛砂深井)	164921	2616491
#13	(台西鄉五條港)	168511	2624556	#27	(口湖淺井)	164560	2609882
#14	(台西鄉瓦厝)	169456	2623115	#28	(口湖深井)	164640	2609882



圖3 雲林沿海地區水質站分布圖

銅、鋅、鎳、鈉、鉀、鐵、鉛、鎂、鈣、硬度、鈉吸著率，共28項水質指標。

根據水利署及中央地質調查所於雲林縣設置地下水位觀測站址之水文地質鑽探資料，顯示本地區自地表至300公尺深度間，共包含4個含水層和4個阻水層，由上而下命名為含水層一、阻

水層一、含水層二、阻水層二、含水層三、阻水層三、含水層四及阻水層四。本研究所採用28口水質井，其中21口淺水井位於含水層一，7口深水井位於阻水層一及阻水層二。

本區域內砷濃度升高與超抽地下水有高度相關，雲林縣位於濁水溪下游南岸，因濁水溪含砂量甚大，且豐枯水量懸殊，故雲林縣自來水每日用水24萬噸，大部份來自地下水，且沿海地區之農業及養殖用水亦大部分抽用地下水，造成嚴重地層下陷，局部地區下陷量甚至超過2公尺，且地層下陷中心已由沿海移往內陸，政府注意到此地區地下水超抽問題之嚴重性，於民國64年11月管制四湖、口湖、水林及北港四鄉鎮地下水之開發。近年來，雲林縣沿海地區下陷趨勢已較為緩和穩定，其年平均下陷速率約3~5公分/年(工業技術研究院能源與資源研究所，2002，2004)，但台灣高鐵雲林站周邊之土庫鎮及元長鄉，地層下陷速率卻有增加現象，目前量測下陷速率約每年9.5公分，嚴重威脅高速鐵路行車安全。

為了解本區域砷污染之嚴重性，及評估本區砷污染程度，根據行政院環保署90年地下水污染管制標準，地下水飲用水含砷量濃度(As)標準為10 ug/l以下，一般地下水使用含砷量濃度為500 ug/l以下，為符合安全之標準，比對雲林縣沿海地區21處淺水井及7處深水井，81年至94年地下水水質平均值資料(如表2)顯示，28口井中除7號井水質平均值不符合500 ug/l外，其他各

表2 81年至94年雲林沿海地區22口水質站砷污染之平均值與標準偏差

單位：(ug/l)

站 號	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
平均值	28.3	21.5	109.9	77.6	9.5	161.8	593.8	11.1	477.5	93.0	83.3	49.5	55.6	31.4
標準偏差	41.1	15.9	67.4	47.6	10.5	123.9	349.4	6.7	208.2	101.3	26.8	32.4	90.9	16.9
站 號	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27	#28
平均值	111.2	6.7	66.0	19.5	32.4	20.0	7.5	226.1	10.5	19.1	122.2	38.0	22.4	68.5
標準偏差	171.1	8.9	49.7	14.8	26.2	12.6	3.8	79.7	2.9	7.9	24.1	14.4	24.5	65.4

站砷濃度平均值均低於地下水使用標準，惟28口水井僅3口水井砷濃度平均值符合飲用水標準，其餘大部分水井均不符合飲用水標準，顯見本區仍為砷污染較嚴重地區，需有效掌握砷在本區內之空間與時間變化，藉此管理地下水使用量與使用安全。

2. 應用交叉驗證法於模式之穩定性分析及參數檢定

本研究中採用倒傳遞類神經網路作為砷濃度推估補遺方法，惟應用於本區砷濃度補遺前仍有困難先須克服，由於砷濃度資料取得不易，本區內大部份水質站僅具10~50筆資料，在資料過少情形下，要架構穩定模式有其困難度，為求模式可獲穩定推估補遺結果，我們先採用交叉驗證測試模式中參數，確保模式穩定之後，再將本模式應用於實際補遺推估工作。

在架構交叉驗證模式時，考量本區域雖設有28口水質站，但23號至28號水質站之資料有大量缺漏，若使用這些資料進行模式之訓練及驗證將因部分水質站資料缺乏而需刪除整筆資料之情況，在資料已極度缺乏下，又需刪除可用資料，相當不利於架構穩定可靠模式，故僅採用資料較完整1號至22號水質站之砷濃度作為交叉驗證補遺推估模式之訓練與驗證資料，其驗證過程如理論說明先採用1號水質站以外之21處水質站作為網路輸入值，以1號水質站之歷年砷濃度作為網路輸出值，也就是推估對象，其網路架構為輸入層為21個神經元(21處已知站之砷濃度)，輸出層為1個神經元架構(未知站之砷濃度)，如圖4，分別選用7段資料中6段作訓練，7段中1段作驗證，完成1號站後，接續2號水質站，最後完成22處水質站并交叉驗證工作為止，得到各水質站訓

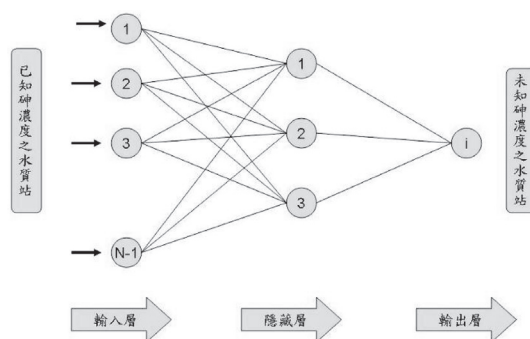


圖4 類神經網路補遺推估模式架構圖

練期及驗證期之平均誤差量，如表3，表中並顯示各水質站在訓練及驗證過程之rmse值，且標註是否大於該站之平均值或標準偏差。

參考表1現有22水質站之砷濃度平均值與標準偏差，另表3為各站砷濃度交叉驗證之rmse，比較兩表得知大部份水質站之驗證期平均誤差(rmse)在飲用水質標準以下，且小於該水質站之砷濃度平均值與標準偏差(僅#1、#5及#19大於該站平均值與標準偏差)，顯示本模式可達粗略之補遺推估效果，但在6號、7號、9號及15號站經交叉驗證結果其驗證期之平均誤差較高，主要因為這些站原始資料平均值與標準偏差均較大。

另比對表中各水質站之訓練與驗證誤差比相當大(約1,000倍)，顯示在資料過少時，應用倒傳遞類神經網路，易發生過度訓練情形且表示模式之推估結果較不穩定，推測可能原因為模式中優選參數過大或過小，使得網路推估不準確，需提出進一步改善方法，確保本模式可適用於任何狀況。

為解決過度訓練問題，我們利用修正型目標函數以有效解決過度訓練問題，但須先決定性能比，首先採用22處水質站之砷濃度測試模式合

表3 雲林沿海地區22口水質站採類神經網路交叉驗證誤差值(rmse)值

單位(ug/l)

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11
訓練期之 平均誤差值	0.034	0.001	0.002	0.003	0.112	3.928	0.021	0.732	0.010	0.003	1.538
驗證期之 平均誤差值	45c	14	66	70b	13c	148b	281	7b	121	68	35b
	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22
訓練期之 平均誤差值	0.0025	0.0045	1.7047	0.0264	0.0005	0.0040	0.0003	0.0021	0.0005	0.0004	11.4709
驗證期之 平均誤差值	36b	74a	11	108	8a	44	11	36c	11	3	61

a 表示該值大於該站之平均值，b 表示該值大於該站之標準偏差，c 表示該值大於該站之平均值與標準偏差。

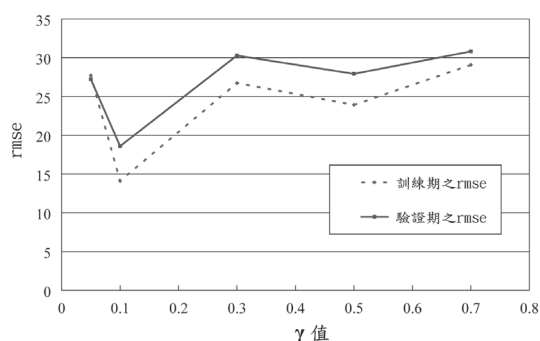


圖5 應用倒傳遞類神經網路之性能比測試圖

理之性能比，經採用不同性能比測試22處水質站平均誤差如圖5，發現當性能比設為0.1時，最能消除模式中發生過度訓練情形且能兼顧模式精確度。

另外在降低隱藏層數目上，依過去選用隱藏層個數之經驗，即隱藏層數需在輸入層個數及輸出層個數平均值11個 $((21+1)/2=11)$ 以下，可減少過度訓練情形。

依上述經驗採11以下之隱藏層神經元個數，並使用網路修剪法，逐步減少網路中隱藏層之神經元個數，測試網路訓練期及驗證期誤差，發現當隱藏層之神經元，設定為3或4個神經元具較佳穩定度，故本模式選擇在隱藏層採用3個神經元作為本模式架構。

3. 雲林縣沿海地區砷濃度推估及區域砷濃度變化

本研究應用倒傳遞類神經網路結合修正型目標函數建構類神經網路推估模式，在採用修正型

目標函數時，過去研究建議需配合Levenberg-Marquardt演算法可得到良好參數最佳化結果，且Levenberg-Marquardt演算法為較近期發展標準數值最佳化演算法，雖本研究中無大量參數，但是在有限資料下找尋最適合參數，為需優先克服之問題，因此，本研究採用Levenberg-Marquardt演算法結合修正型目標函數結，架構類神經網路補遺推估模式。

考量本區域採樣水質資料有限，13年觀測期間，全區28站(#1~#28)水質測站中較完整水質站共7口水質井(#3, #6, #7, #12, #13, #17, #19)具50筆資料，其餘各站砷濃度資料皆有零星缺漏情形，在此情形下，若僅選擇7口水質測站中砷濃度作為輸入層之神經元，建立推估補遺模式，補遺81年至94年間其它21口水質站之零星缺漏砷濃度資料，由於網路輸入層之神經元過少，將擴大推估補遺之誤差，因此，將模式分割為兩部份，首先，採用17口水質站(#1, #2, #3, #4, #6, #7, #8, #9, #11, #12, #14, #16, #17, #18, #19, #20, #21)補遺81年5月至88年3月間零星缺漏水質站之砷濃度(共11站，如表4)，其網路架構為輸入層為17個神經元，隱藏層為3個神經元，輸出層為1個神經元，目標函數採用修正型目標函數，其中訓練資料、驗證資料及補遺資料之筆數，及其訓練與驗證平均誤差，如表4。

另一模式架構，則因88年後持續具砷濃度完整資料，僅剩7口井(#3, #6, #7, #12, #13, #17, #19)，本補遺推估模式僅可採用資料長度

表4 81年至88年砷污染補遺之rmse值及訓練、驗證及補遺模式採用資料筆數統計表

	#5	#10	#13	#15	#22	#23	#24	#25	#26	#27	#28
訓練期rmse	5.1	24.0	16.9	15.9	49.2	0.6	2.8	8.6	4.2	4.1	21.0
驗證期rmse	6.1	35.0	14.3	12.4	83.5	3.1	9.2	23.1	22.1	58.6	53.4
訓練期資料筆數	19	22	7	15	20	20	20	19	20	20	20
驗證期資料筆數	4	4	4	4	4	4	4	4	4	4	4
補遺資料筆數	5	2	17	9	4	4	4	5	4	4	4

表5 88年至94年砷污染補遺之rmse值及訓練、驗證及補遺模式採用資料筆數統計表

	#1	#2	#4	#5	#8	#9	#10	#11	#14	#15	#16
訓練期rmse	30.4	11.2	41.2	9.0	3.7	151.3	81.7	22.5	13.4	117.6	6.6
驗證期rmse	16.9	7.7	27.1	9.6	8.0	132.2	39.0	16.6	8.1	87.6	6.8
訓練期資料筆數	24	24	24	24	24	24	24	24	24	24	24
驗證期資料筆數	4	4	4	4	4	4	4	4	4	4	4
補遺資料筆數	24	24	24	24	24	24	24	24	24	24	24
	#18	#20	#21	#22	#23	#24	#25	#26	#27	#28	
訓練期rmse	12.2	9.3	3.8	96.8	3.8	9.1	33.5	12.6	8.4	60.1	
驗證期rmse	5.6	10.6	1.4	132.1	1.5	6.0	17.0	20.5	61.6	36.2	
訓練期資料筆數	24	24	24	24	24	24	24	24	24	24	
驗證期資料筆數	4	4	4	4	4	4	4	4	4	4	
補遺資料筆數	24	24	24	24	24	24	24	24	24	24	

較長之上述7口水質站井之砷濃度補遺88年9月至94年5月撤除水質站之砷濃度(共21站,如表5),其網路架構仿照上述架構修改輸入層為7個神經元,隱藏層為3個神經元,輸出層為1個神經元,目標函數同樣採用修正型目標函數,其中訓練資料、驗證資料及補遺資料之筆數,及其訓練與驗證平均誤差如表5。

採用倒傳遞類神經網路結合修正型目標函數之推估結果(表4與表5)與交叉驗證結果(表3)比較,採用修正型目標函數後,在同一設站處驗證時之平均誤差明顯降低,且訓練時與驗證時平均誤差比為1至2,與交叉驗證過程之訓練時與驗證時平均誤差比為1000,明顯降低很多,說明加入修正型目標函數後,可有效減少過度訓練情形,而本補遺結果以過去資料為基礎,在訓練時可藉由類神經網路保留過去資料各站間砷濃度相關訊息,再藉由驗證時測試各站補遺可能誤差,為一種較可信且非線性補遺方式,在複雜地下水層中由於地層狀況不易獲得,污染傳輸機制很難掌控,採用本補遺模式可藉由簡單網路架構,將缺漏資料補齊,在最少設置站之數目及採樣次數下,達到可接受水質監測結果。

在表4與表5同時出現之水質站(#5, #10, #15, #22, #24, #25, #26, #27, #28),表4採用17處水質站進行砷濃度空間推估,其誤差較表5為小,說明在輸入層採用較多神經元推估準確度較高,但越多網路輸入層之神經元表示須需增加監測經費,在經費有限及可容許誤差範圍內,可依靠模式運算能力推估區域砷空間污染情形,作為地下水控管之依據。

將補遺結果之81年至94年砷濃度,分別繪製81年、85年、89年及93年砷污染情形等值線圖,如圖6(a)~圖6(d),由圖中可見81年後雲林地區地下水中砷濃度逐年增加,圖中代表砷污染區嚴重之深色部份(砷濃度大於300 ug/l)位於雲林縣西南沿海地區,至85年砷污染達到最高峰,尤其在雲林縣西南區沿海地區代表黑色嚴重污染區擴散至台17線東側,至89年後砷濃度緩慢下降且黑色代表嚴重污染區面積縮小至台17線以西地區,至93年雲林縣西南沿海地區部份砷濃度嚴重污染區又有些微擴大傾向,深色區域再次擴大,但93年後砷污染情形已漸趨穩定已獲有效控制。

將砷濃度變化與本區域地下水位做比對,發現地下水位與砷濃度有密切關係,由圖7中本

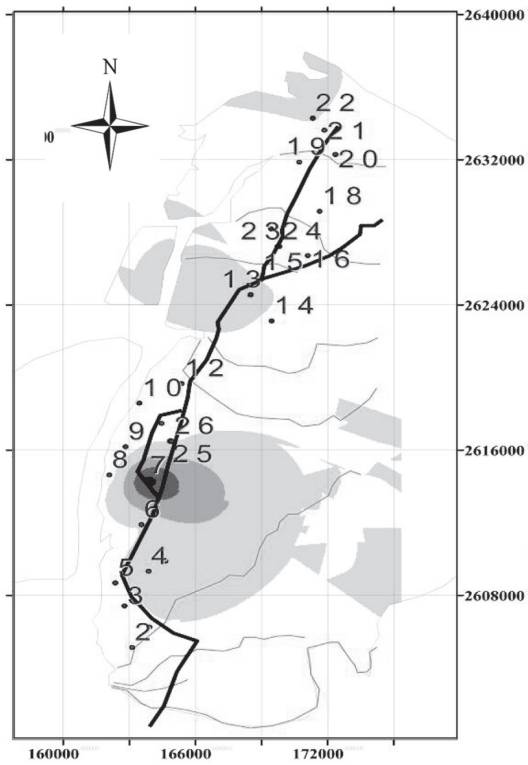


圖6(a) 81年雲林縣沿海地區砷污染等值線

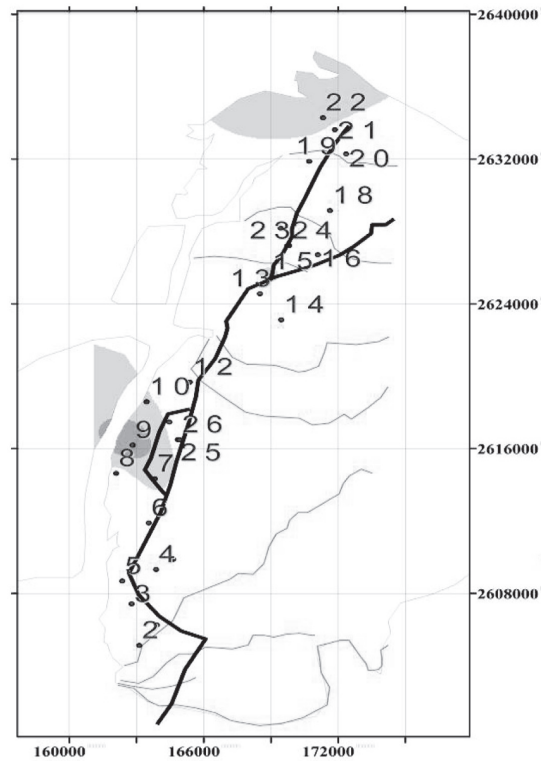


圖6(c) 89年雲林縣沿海地區砷污染等值線

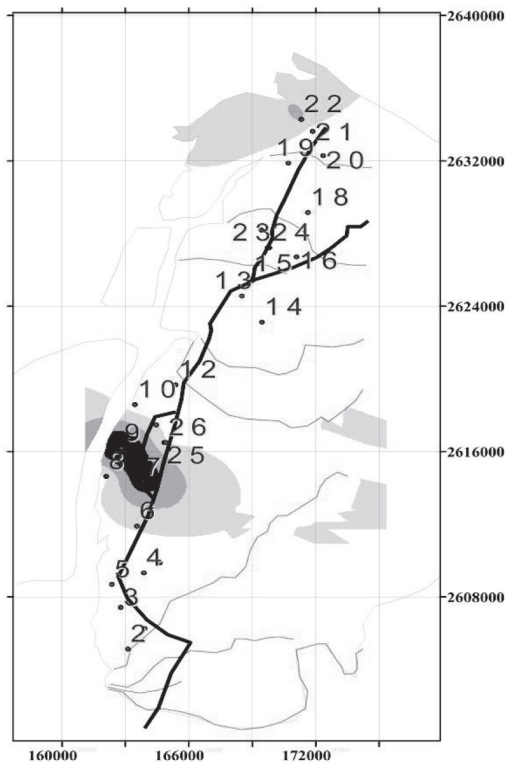


圖6(b) 85年雲林縣沿海地區砷污染等值線

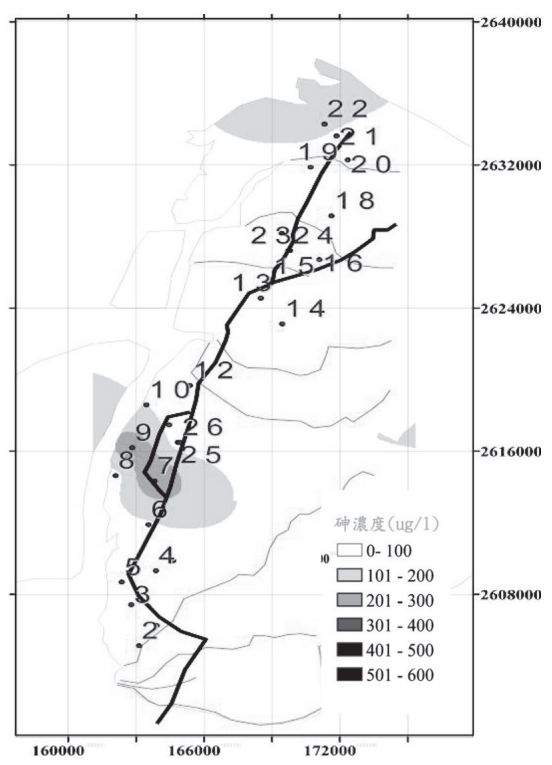


圖6(d) 93年雲林縣沿海地區砷污染等值線

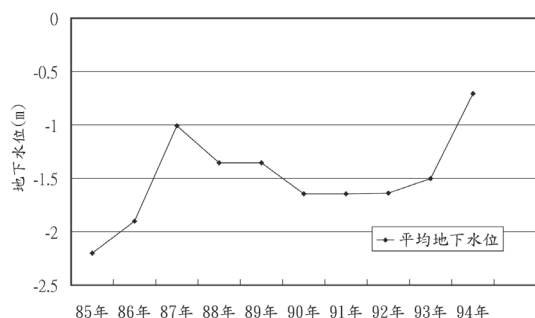


圖7 雲林地區沿海22口地下水井之85年至94年平均地下水位圖

區域歷年地下水位資料，本區域平均地下水位於85年達到最低-2.2 m，同時，砷濃度達嚴重污染之面積(深色部份)達到最大，污染源中心區內，最高砷濃度到達743 ug/l，地下水水質達嚴重砷污染程度，85年後地下水超抽情形得到有效控制，地下水位逐年緩慢提昇，至89年地下水位升至-1.7 m，砷濃度嚴重污染區面積大幅減少，在93年前後，由於地下水位又往下降，砷濃度再次上升，顯見本區域地下水中砷濃度確與本區域內地下水水位高低有密切關係，有效控制地下水用量或加強污染地區地下水補注將有明顯改善地下水砷污染情形，另由歷年砷濃度較高區域除西南沿海地區，西北區同樣有砷污染情形，與雲林縣西南區相較其污染程度雖較輕微且面積較小，但仍有較高砷污染潛勢，必須加強地下水管理，以防止砷污染擴散。

六、結 論

1. 類神經網路可用於補遺水質站中資料零星缺漏情形，惟地下水之水質資料取得耗時且成本過高，無法取得大量水質資料，因此，如何在有限資料下架構穩定網路是本研究首要課題，藉由導入修正型目標函數，將隱藏層至輸出層中權重納入目標函數後，顯示訓練期與驗證期之誤差比從原1000倍減為2倍，明顯改善過度訓練情況，藉此補遺結果可掌握區域砷污染擴散情形，並提供地下水管理者有效控管地下水使用。
2. 本研究利用有限資料架構類神經網路模式推估砷濃度平均誤差值僅29.2 ug/l，在本區砷

濃度變化劇烈情況下，如本區平均標準偏差為158.87 ug/l，本區砷濃度最高7號站之標準偏差為394.4 ug/l，顯示本補遺推估結果屬可容許誤差，對於中低污染區之水質站補遺推估結果有較高可信度，相對於極端值與高污染區之砷濃度，由於屬外插推估方法其補遺推估結果則較差，但整體而言，類神經網路模式提供一種在資料極度缺乏情況下可行之補遺方法。

3. 本研究應用類神經網路補遺雲林縣沿海地區81年至94年28口水質站之砷濃度，將補遺結果繪製等值線圖，由圖中砷濃度分佈與地下水位及地層下陷數據相比對，可知砷濃度升降與地下水位及地層下陷深度具高度相關，有效控制地下水抽取，對於本區砷污染擴散控制相當有助益。

謝 誌

本研究承蒙國科會專案(NSC 95-2313-B-002-051-MY3)經費之補助，特此致謝。

參 考 文 獻

- 【1】Abrahart, R. J., See, L. and Kneale, P. E., 1998, "New Tools for Neurohydrologists: Using Network Pruning and Model Breeding Algorithms to Discover Optimum Inputs and Architectures", Proceedings of the 3rd International Conference on Geocomputation, University of Bristol, 17-19 September.
- 【2】Chang, F. J. and Chen, Y. C., 2001, "A counterpropagation fuzzy-neural network modeling approach to real-time streamflow prediction", Journal of Hydrology, Vol. 245, pp. 153-164.
- 【3】Chang, F. J., Chang, L. C. and Huang, H. L., 2002, "Real-time recurrent neural network for streamflow forecasting", Hydrological Processes, Vol. 16, No. 13, pp. 2577-2588.
- 【4】Chang, F. J. and Chang, Ya-Ting, 2006, "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir", Advances in Water Resources, Vol. 29, pp.1-10.
- 【5】Chang, F. J., Chang, L. C. and Wang, Y. S., 2006, "Enforced Self-Organizing Map Neural Networks for River Flood Forecasting", Hydrological Processes, Vol. 21, pp.741-749.
- 【6】Coppola, E., Szidarosky F., Poulton, M. and

- Charles, E., 2003, "Artificial Neural Network Approach for Predicting Transient Water Levels in a Multilayered Groundwater System under Variable State, Pumping, and Climate Conditions", *Journal of Hydrologic Engineering*, Vol. 8, No. 6, pp. 348-360.
- 【7】Gumrah, F., Oz, B., Guler, B. and Evin, S., 2000, "The Application of Artificial Neural Networks for the Prediction of Water Quality of Polluted Aquifer", *Water, Air and Soil Pollution* 119, pp. 275-294.
- 【8】Kwok, T. Y. and Yeung, D. Y., 1997, "Constructive algorithms for structure learning in feedforward neural networks for regression problems", *IEEE Transaction on Neural Networks*, Vol. 8, pp. 630-645.
- 【9】Liu, C. W., Jang, C. S., and Liao, C. M., 2004, "Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan)", *Science of the Total Environment* 321(1-3): 173-188.
- 【10】Liu, C.W., Wang, S.W., Jang, C.S. and Lin, K.H., 2006, "Occurrence of Arsenic in Ground Water in the Choushui River Alluvial Fan, Taiwan", *Journal of Environmental Quality* 35, pp. 68-75.
- 【11】Lin, Y. B., Lin, Y. P., Liu, C. W. and Tan, Y. C., 2006, "Mapping of spatial multi-scale sources of arsenic variation in groundwater on ChiaNan floodplain of Taiwan", *Science of the Total Environment*, 370: 168-181.
- 【12】Martos, F. S., Aguilera, P. A., Garrido-Frenich, A., Torres, J. A. and Pulido-Bosch, A., 2002, "Assessment of Groundwater Quality by Means of Self-Organizing Map: Application in a Semiarid Area", *Environment Management*, Vol. 30, No.5, 716-726.
- 【13】Nolan, B. T., Arthur, L. Baehr and Leon, J. Kauffman, 2003, "Spatial Variability of Groundwater Recharge and its Effect on Shallow Groundwater Quality in Southern New Jersey", *Vadose Zone Journal*, Vol. 2, issue 4, pp. 677-691.
- 【14】Ruiz, F., Gomis, V. and Blasco, P., 1990, "Application of factor analysis to the hydrogeochemical study of a coastal aquifer", *Journal of Hydrology* 119, pp. 169-177.
- 【15】王榮德、胡賦強、江錦華、吳新英，1983，「飲用水改善前後，烏腳病發生率之研究」，烏腳病之研究報告，第18輯，台灣省烏腳病防治中心，第21-25頁。
- 【16】郭益銘，1999，「應用多變量統計分析與類神經網路分析雲林沿海地區地下水水質變化」，國立台灣大學農業工程研究所碩士論文。
- 【17】張斐章、胡湘帆、黃源義，1998，「反傳遞模糊類神經網路於流量推估之應用」，農業工程學報，第四十四卷，第二期，第26-38頁。
- 【18】張斐章、孫建平，1997，「類神經網路及其應用於降雨-逕流過程之研究」，中國農業工程學報，第四十三卷，第一期，第34-49頁。
- 【19】張斐章、張麗秋，2005，「類神經網路」，東華書局。
- 【20】曾文賓，1976，「烏腳病之診斷、治療與預防」，台灣省政府烏腳病防治中心，第一輯，第1-25頁。
- 【21】廖少威，1991，「以主成份分析法探討台灣灌溉水質受工業、都市、畜牧廢水污染之變異分析」，國立台灣大學農業工程研究所碩士論文。
- 【22】盧瑞山、駱尚廉，1998，「類神經網路於污染物空間推估之研究」，中國環境工程學刊，第八卷，第三期，第239-245頁。
- 【23】工業技術研究院能源與資源研究所，2002，「臺灣地區地層下陷之監測、調查及分析(2/4)」，經濟部水利署。
- 【24】工業技術研究院能源與資源研究所，2003，「台灣地區地層下陷之監測、調查及分析(3/4)」，經濟部水利署。
- 【25】羅華強，2005，「類神經網路-MATLAB的應用」，高立圖書有限公司。
- 收稿日期：民國 97 年 04 月 29 日
修正日期：民國 97 年 07 月 31 日
接受日期：民國 97 年 08 月 14 日