

國立臺灣大學生物資源暨農學院生物環境系統工程學研究所

碩士論文

Graduate Institute of Bioenvironmental Systems Engineering

College of Bioresources and Agriculture

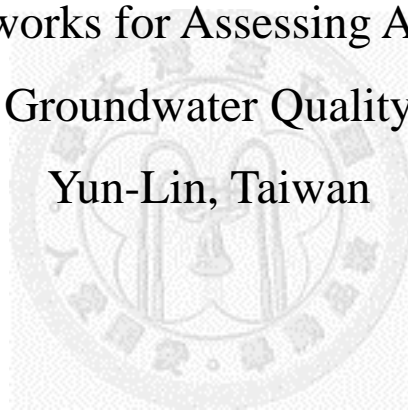
National Taiwan University

Master Thesis

以類神經網路探討雲林沿海地區地下水砷濃度與水質特徵

Artificial Neural Networks for Assessing Arsenic Concentrations
and Characteristics of Groundwater Quality in the Coastal Area of

Yun-Lin, Taiwan



林政華

Cheng-Hua Lin

指導教授：張斐章

Advisor: Fi-John Chang

中華民國一百零一年七月

July 2012

誌謝

很幸運能夠在臺大就讀研究所，進入水資源資訊系統研究室並順利完成碩士論文，因為有各位的幫助，我才能夠在這短短兩年的研究之路，讓本篇論文能夠順利產出，我要感謝非常多人。最由衷感謝的是恩師 張斐章老師，有您的悉心指導與栽培，無論是研究上適切的提點或是生涯的規劃與建議，都使我獲益匪淺，讓我們擁有一個充滿研究氣息、完善資源與良好讀書風氣的研究室。在這樣的環境下，從學長姐身上學習，得到了研究方面的協助與鼓勵，與學弟妹互相切磋、解決問題，這段過程中學習到很多待人處世的道理，我也與研究室成員們建立深厚的感情，這些都是在人生的道路上，很難得的緣分與機會。

在此也由衷的感謝學位論文口試委員臺灣大學生物環境系統工程學系劉振宇教授、淡江大學水資源及環境工程學系張麗秋教授、交通大學土木工程學系張良正教授、新北市政府水利局高力山博士針對論文內容的不吝指正，並於論文口試給予諸多寶貴建議與指教，使得本篇研究更臻完整。

在研究所的這段時間，也感謝臺大與系上各位老師的幫助以及教導，在此也特別感謝劉振宇老師研究室的高雨瑄學姊，提供本篇研究中許多化學領域的專業知識與建議，給予我相當大的協助。

感謝水資源資訊系統研究室各位夥伴的照顧與陪伴，熱心指導學弟妹的學長姐—衍銘、國威、孟蓉、文柄、昌翰、品安、逸鴻、孫維，論文研究以及生活上的經驗分享都受到你們很大的照顧與關心；感謝

辛苦的助理們—郁垂在生活上的照顧與叮嚀、恒玥對於論文的修改與建議，以及惠茵在許多細節及資料上的協助；也感謝學弟妹英秦、宇軒、琬渝、昱中、英睿的陪伴與幫忙；感謝一起為論文奮鬥的同學承賢與桂宏，這段時間的共事，有你們的幫助與照應，讓我順利完成碩士學位。感謝研究室的夥伴們營造這麼美好的環境，總是充滿著活力與朝氣。

最後要感謝我的家人，感謝爸爸、媽媽無私的支持與照顧，讓我在求學路上無後顧之憂，以及妹妹、弟弟的鼓勵，我才能順利完成碩士學位；感謝和我一起攝影、出遊的文藝青年們，紓解了很多在做研究時累積的壓力；感謝女友心好的陪伴，你的支持讓我更有力量繼續學習且對自己有信心。在此謹與你們分享研究成果與畢業的喜悅心情，由衷的感謝你們一路的鼓勵與相伴。

政華 2012.7

摘要

臺灣西南部沿海地區過去為烏腳病流行地區，經研究發現此情況與當地居民長期飲用含砷量高之地下水有密切關係，雖然當地居民已不再直接飲用地下水，但仍大量抽取地下水供給灌溉、養殖、公共及民生用水等多項用水標的，此舉可能造成砷透過生物累積過程進而對人體健康造成危害，實有必要建立可靠之推估模式，掌握該區域地下水砷污染情形，並分析該區域地下水之水質特性。

本研究目的為探討臺灣西南部沿海地區的水質特徵、影響地下水砷濃度消長之因子與可能影響砷釋出於地下水中的機制，並推估此區域地下水砷濃度之空間分布。本研究採用水利署於雲林縣沿海地區設置 28 座監測井之水質資料作為分析對象，首先將所有監測井水質資料輸入類神經網路-自組特徵映射網路(Self-Organizing Feature Map, SOM)進行聚類分析，經由視覺化結果呈現有效率地檢視水質因子間暨水質因子與砷濃度之關係，並由聚類的結果對測站進行分類以探討砷濃度在空間上的分布關係。進一步將自組特徵映射網路(SOM)的聚類結果移除砷的資訊作為決定輻狀基底函數類神經網路(Radial Basis Function Neural Network, RBFNN)之隱藏層神經元數目及中心點，再以 RBFNN 計算隱藏層至輸出層間神經元之權重值，以有限之水質因子推估砷濃度。上述推估結果亦與倒傳遞類神經網路(BPNN)模式的推估結果作比較，結論是 SOM 與 RBFNN 建立之模式精確度優於 BPNN，且 SOM 於聚類時加入砷濃度資訊可使分類的結果更為顯著，並增加推估精確度；聚類分析的結果指出沿海地區測站地下水層鹽化與高砷污染情況，導致部分測站存在著高濃度砷，推論可能原因為此區域地下水環境趨

於還原環境(如環境中存在高 pH 值、高鹼度、低溶氧量與低硫酸鹽濃度等)，有利於砷釋出，此結論與國內外相關研究結果一致。最後，應用地理資訊系統(GIS)將本模式推估之地下水砷濃度繪製地下水砷污染區域潛勢圖，呈現研究區域內砷濃度推估結果在 1998 年與 1999 年在空間上的變化，經由輸入水質因子便可了解此區域地下水中砷之釋出與遷移情形。

本模式特色在於可藉由 SOM 拓樸圖建構與解釋砷濃度與水質因子間關係，同時聚類結果有助於 RBFNN 網路之砷濃度推估能力，並以 GIS 呈現區域地下水砷濃度空間分布情形，展現以水質因子推估區域砷濃度分布圖，可供決策者透過視覺化了解區域時間與空間上的砷污染情形。

關鍵字：砷、地下水水質、類神經網路、自組特徵映射網路、輻狀基底函數類神經網路



ABSTRACT

In the past, Blackfoot disease commonly occurred along the southwestern coast of Taiwan. A number of investigations revealed this epidemic disease was highly related to arsenic (As) concentration in groundwater, which is the main source of drinking water to local residents. Although local residents do not directly drink groundwater any more in the Yun-Lin County of Taiwan, groundwater is still a main water source in this area because surface water suffers from limited sources. A large quantity of groundwater has been extracted from the aquifer for supplying water to the public, fish ponds and crop lands, which has resulted in the accumulation of arsenic in crops and fish. Products highly-contaminated with arsenic has threatened the health of residents. Therefore, it's essential to construct a reliable model for estimating As concentration in groundwater.

The aims of this study are to assess the characteristics of groundwater quality, extract the factors affecting As concentration, investigate the sources releasing As, and estimate As concentration in groundwater. The Water Resources Agency (WRA) have set up 28 monitoring wells for investigating groundwater pollution, and water quality data collected by the WRA were used in this study. The first subject of this study is to import all the data collected form 28 wells to the Self-Organizing Feature Map (SOM) network. The SOM was applied to classifying all the water quality data into

a topology map for finding the hidden relations among data and the spatial patterns between water quality variables and As concentration. Then the clustering results were adopted as the centroids of the Radial Basis Function Neural Networks (RBFNN) for accurately estimating As concentration based on water quality variables. In addition, the Back Propagation Neural Network (BPNN) was built to compare with the proposed model that integrates the SOM and RBF. The results demonstrate that the performance of the proposed model is better than the BPNN. When comparing the clustering results, adding As concentration to the SOM could make the clustering results more obvious and therefore achieves much accurate estimation. Moreover, the results demonstrate the characteristics of groundwater quality in coastal areas correlate with salinization and arsenic pollution factors. According to the clustering results, we surmise that the occurrence of high arsenic concentration in parts of wells is mainly because groundwater is in the reduction phase, especially at higher pH and Alk values and lower dissolved oxygen levels and SO_4^{2-} concentration. Finally, the Geographic Information System (GIS) is applied to the results of groundwater quality models for displaying the spatial distribution map of As pollution so that we can realize the temporal and spatial variation in arsenic concentration in the study area.

Keywords: Arsenic (As), Groundwater quality, Artificial neural network

(ANN), Self-Organizing Feature Map (SOM), Radial Basis Function Neural Networks (RBFNN).



目錄

誌謝.....	i
摘要.....	iii
ABSTRACT.....	v
目錄.....	viii
圖目錄.....	x
表格目錄.....	xiii
第一章 緒論.....	1
1.1 研究緣起.....	1
1.2 研究目的.....	2
1.3 研究架構.....	4
第二章 文獻回顧.....	6
2.1 地下水中砷污染相關研究.....	6
2.2 分析地下水水質方法.....	9
2.3 自組特徵映射網路之應用.....	10
第三章 理論概述.....	12
3.1 類神經網路.....	12
3.1.1 自組特徵映射網路(SOM).....	16
3.1.2 輻狀基底函數類神經網路(RBFNN).....	22
3.1.3 倒傳遞類神經網路(BPNN).....	27
3.2 交叉驗證法優選模式.....	33
3.3 評估指標.....	35

3.3.1	均方根誤差(RMSE)	35
3.3.2	平均絕對誤差(MAE)	36
3.3.3	正規均方誤差(NMSE)	36
3.3.4	相關係數(CC)	36
第四章	研究案例	37
4.1	研究區域概述	37
4.2	資料蒐集與處理	37
4.2.1	資料蒐集	37
4.2.2	資料處理	45
4.3	模式架構	45
第五章	結果與討論	54
5.1	SOM 拓樸架構	54
5.1.1	聚類內測站之位置分布	58
5.1.2	聚類內水質因子探討	60
5.2	區域地下水模式推估結果	68
5.2.1	輻狀基底函數類神經網路(RBFNN)	68
5.2.2	倒傳遞類神經網路(BPNN)	71
5.2.3	空間推估	71
第六章	結論與建議	79
6.1	結論	79
6.2	建議	81
第七章	參考文獻	82

圖目錄

圖 1-1	研究架構圖	5
圖 2-1	全球地下水含水層受砷區域分布圖	9
圖 3-1	前饋式類神經網路架構圖	13
圖 3-2	回饋式類神經網路架構圖	14
圖 3-3	類神經網路學習方法	15
圖 3-4	二維矩陣的 SOM 架構圖	17
圖 3-5	SOM 網路神經元間拓樸座標	17
圖 3-6	拓樸層網路圖	19
圖 3-7	演算代數與鄰近半徑之關係	20
圖 3-8	SOM 演算流程	21
圖 3-9	輻狀基底函數類神經網路架構圖	23
圖 3-10	各形式之輻狀基底函數	25
圖 3-11	架構自組特徵映射網路輸出層	26
圖 3-12	倒傳遞類神經網路架構圖	28
圖 4-1	雲林縣沿海地區監測井分佈圖	40
圖 4-2	雲林地區水文地質概念圖	41
圖 4-3	SOM 不同拓樸層之模式表現	49
圖 4-4	SOM 聚類砷濃度分布圖(無加入砷)	50
圖 4-5	SOM 聚類砷濃度分布圖(加入砷)	51
圖 4-6	地下水砷濃度推估模式流程圖	52

圖 4-7	交叉驗證流程圖	53
圖 5-1	SOM 聚類中心點.....	55
圖 5-2	模式輸入與輸出架構圖	56
圖 5-3	SOM 聚類資料筆數分布圖.....	56
圖 5-4	SOM 聚類測站分布圖.....	57
圖 5-5	SOM 聚類對應測站位置分布圖.....	59
圖 5-6	SOM 聚類電導度中心點.....	61
圖 5-7	SOM 聚類總溶解固體濃度中心點.....	61
圖 5-8	SOM 聚類鈉濃度中心點.....	61
圖 5-9	SOM 聚類鉀濃度中心點.....	61
圖 5-10	SOM 聚類鎂濃度中心點.....	61
圖 5-11	SOM 聚類鈣濃度中心點.....	61
圖 5-12	SOM 聚類砷濃度中心點.....	62
圖 5-13	SOM 聚類溶氧量中心點.....	62
圖 5-14	SOM 聚類氯鹽濃度中心點.....	62
圖 5-15	SOM 聚類硫酸鹽濃度中心點.....	62
圖 5-16	SOM 聚類鹼度中心點.....	62
圖 5-17	SOM 聚類 pH 值中心點.....	62
圖 5-18	1998 年 1 月模式推估地下水砷濃度分布圖	73
圖 5-19	1998 年 3 月模式推估地下水砷濃度分布圖	74
圖 5-20	1998 年 6 月模式推估地下水砷濃度分布圖	75
圖 5-21	1998 年 9 月模式推估地下水砷濃度分布圖	76

圖 5-22 1999 年 1 月模式推估地下水砷濃度分布圖77

圖 5-23 1999 年 3 月模式推估地下水砷濃度分布圖78



表格目錄

表 4-1	1999 年至 2005 年雲林沿海地區觀測井地下水砷濃度資料表	41
表 4-2	雲林沿海地區 28 座監測井座標資料表	42
表 4-3	雲林沿海地區 21 口監測井及水利處地下水井基本資料	43
表 4-4	雲林沿海地區地下水水質相關性	44
表 4-5	訓練與測試資料	46
表 5-1	SOM 聚類中心點資料表 1.....	63
表 5-2	SOM 聚類中心點資料表 2.....	63
表 5-3	RBFNN 地下水砷濃度推估結果.....	70
表 5-4	BPNN 地下水砷濃度推估結果.....	70



第一章 緒論

本研究以類神經網路探討區域地下水水質特徵與推估地下水砷濃度，並針對雲林沿海地區砷污染嚴重區域作一探討，在此章節對此研究的動機、目的與研究架構作一簡介，分述如下：

1.1 研究緣起

臺灣雲林地區因濁水溪流量豐枯懸殊，枯水期之水量經上游灌溉取水後，已幾無地表餘水可用，故雲林縣目前自來水每日用水 26 萬噸，長期以來均使用地下水。地下水年抽用量推估約為 5 至 7 億立方公尺，年超抽量介於 5000 萬至 2 億立方公尺之間(水利署中區水資源局)，長期的地下水超抽，導致雲林沿海地區地層下陷、海水倒灌、海水入侵、土質鹽化與排水不良等問題。本區域地表水嚴重不足卻有大量農業、養殖與工業用水需求，此區域之地下水主要特性為海水導致之鹽化與高砷濃度污染，大量超抽地下水情形普遍不僅可能造成地下水鹽化，亦可能使得地下水氧化還原狀態改變，進而使原本地層中含砷之鐵氧化物的還原與溶解，高含量砷溶解釋放入地下水中，增加地下水之砷濃度(Liu et al, 2003)。

砷(Arsenic, As)是一種毒性很強且具有致癌性的重金屬元素，過去發現無機砷具有急性與慢性毒性，與人類癌症的發生機率呈現高度相關。而地下水中砷污染問題為世界環境問題之一，這些情形發生於阿根廷、智利、墨西哥、中國與孟加拉，近二十年來許多研究嘗試探討地層中砷之來源與釋出，砷對環境與人類健康之影響係為社會所關切

之議題。1960 年代出現大量烏腳病之疫情，主要原因可能與當地居民長期飲用高砷含量之地下水有關(曾文賓等人，1961)，雖然目前居民已無直接飲用，但當地養殖漁業與民生、公共用水皆因缺水而抽取地下水使用，在生物累積與食物鏈的因素下，仍存在影響人類的風險；而臺灣含高砷濃度之地下水區域，分佈於蘭陽平原、濁水溪沖積扇與嘉南平原。其中雲林麥寮地區，地下水砷濃度高於其他地區，雖然自來水公司有實施計畫改善，但依環保署所作之全臺飲用水質抽驗報告顯示，麥寮水質砷含量依然在超標臨界點。

根據臺灣飲用水水質標準(環保署，2000)、美國飲用水標準(USEPA, 2006)、世界衛生組織的飲用水標準(NRC, 2001; WHO, 2006)，含砷量濃度標準之 $10\mu\text{g/L}$ 與臺灣飲用水水源保護區內之管制標準 $50\mu\text{g/L}$ ，整體而言，臺灣西南部沿海地區地下水一般有砷濃度過高之問題，而濁水溪沖積扇原為臺灣之重要農業地區，受到灌溉水源不足限制與養殖漁業興盛造成地下水之抽取量大幅增加，抽水量約為地下水天然補注量之兩倍，本區之水源供應無法避免偏重地下水之開發，因此對地下水之依賴甚般，故砷污染情形影響甚鉅。

1.2 研究目的

建立可靠的模式最有效且直接之方法了解地下水中砷濃度變化，但受限於地下水模式建立過程，需要相當完整之地層參數、地下水水位及初始污染物質濃度等，其中自然狀態下之土壤與地下水化學環境呈現高度複雜性，大部份現地採樣調查，受限於經費僅可針對有限之觀測井進行水質分析及容易測量參數分析。因此地下水觀測資料往往

相當稀少且珍貴，在統計模式需要有相當的資料、在物理模式需要假設與繁瑣之建立過程，且具有高度不確定性(高力山，2011)。地下水砷濃度之趨勢尚未有明確定論。類神經網路，其在圖形辨識、智慧型控制、污染量預測、降雨與逕流預測等模式上均有良好預測及推估結果(張斐章與張麗秋，2010)，本研究嘗試利用類神經網路，其建立模式的方法簡明、結構單純，模擬物理方程式及高度複雜性之非線性關係有較佳表現的特點，以水質資料為切入點，蒐集地下水水質資料，探討此區域地下水中的水質特徵、影響砷濃度消長之原因與砷污染可能成因。本研究使用自組特徵映射網路(Self-Organizing Feature Map, SOM)對所有測站之水質進行聚類分析，由視覺化的方法有效率地檢視水質因子間或水質因子與砷濃度之潛在關係，由聚類的結果對應測站，可視為對測站分類，其在空間上的分布關係亦可作為探討的對象；另將SOM的聚類結果移除砷的資訊作為決定RBFNN之隱藏層神經元數目及中心點，再以RBFNN之原理計算隱藏層至輸出層間神經元之權重值，以有限之水質因子推估砷濃度。特點在藉由拓樸圖建構與解釋砷與水質因子間關係，同時亦保有網路推估能力。藉由應用地下水水質資料，推估之地下水砷濃度結果繪出地下水砷污染潛勢圖，呈現研究區域內1998與1999年的推估結果在空間的變化，有助於以觀察其他水質因子便可了解此區域砷在地下水中之釋出與遷移情形。以分析此區域地下水特性及建立模式推估地下水中砷污染情形，作為地下水使用管制與污染防治之參考。

1.3 研究架構

本文共分為六章，總研究架構如圖 1-1：第一章為緒論，包括研究動機之說明、研究方法之概述及章節架構之編排；第二章為文獻回顧，針對與本研究相關之研究進行回顧，包括類神經網路、砷污染相關研究、研究區域內以其他角度切入砷污染之發展與應用；第三章為理論概述，介紹本研究使用到之理論：類神經網路理論、交叉驗證法及數種檢視模式優劣之評估指標公式；第四章為研究區域概述、資料蒐集及資料輸入前之處理方式，另外討論在模式的輸入部分加入砷濃度對聚類過程的幫助；第五章將研究內容分為兩大部分：第一部分為以自組特徵映射網路(SOM)對水質資料進行聚類結果，著重拓撲圖的呈現及砷濃度與水質因子間的關係的關係，探討研究區域內之地下水特性；第二部分為採用輻狀基底函數類神經網路(RBFNN)建立地下水中砷濃度推估模式，以砷濃度以外水質因子濃度作為推估砷濃度之主要依據，其中探討 SOM 聚類決定 RBFNN 隱藏層神經元個數與權重，並與倒傳遞(Backpropagation, BP)類神經網路推估之結果進行比較，並繪出測試資料 1998 與 1999 年雲林縣沿海地區地下水中砷濃度分布情形。

第六章為結論與建議，針對此研究結果做出總結，並就未來改進方向與研究方向提出建議。

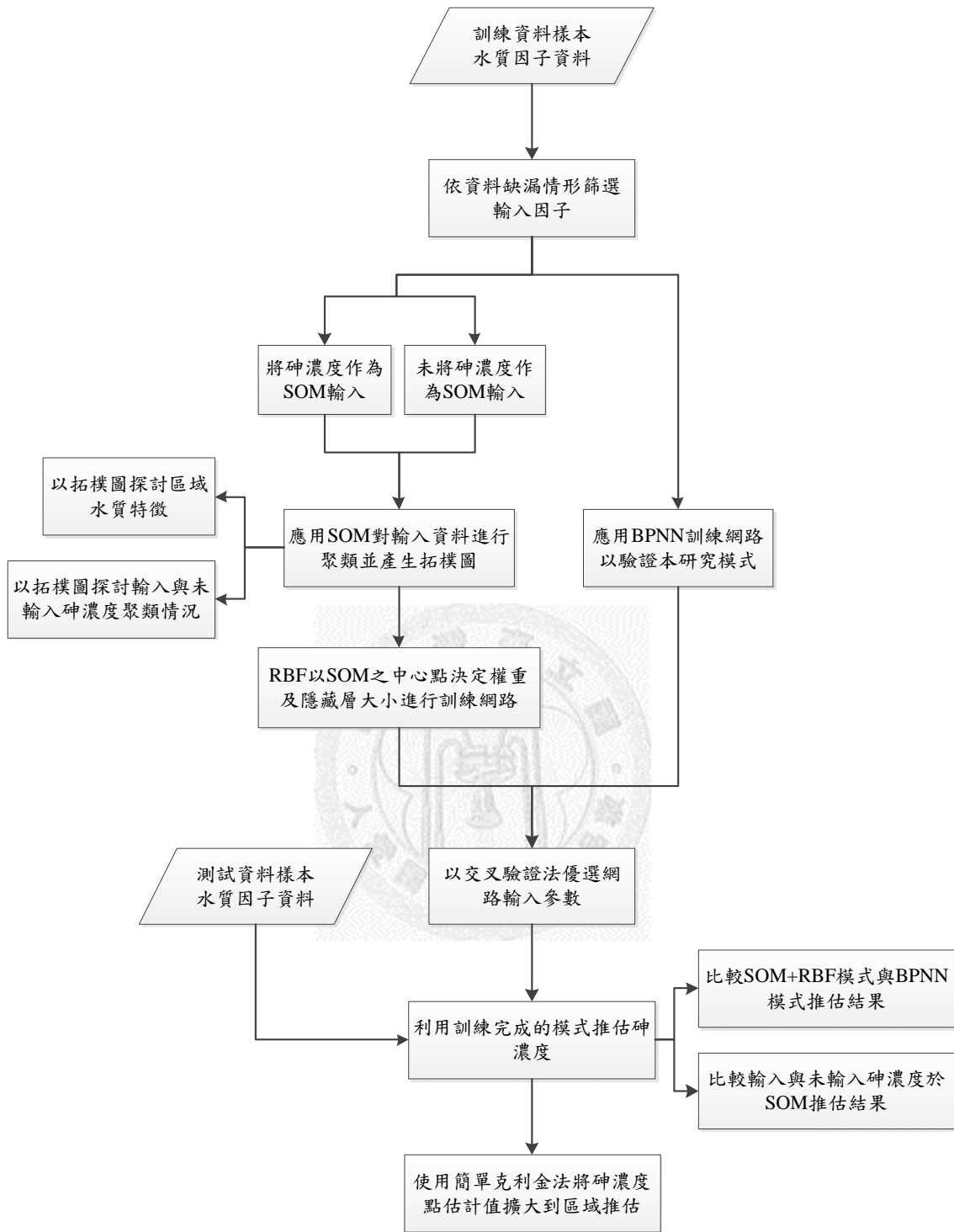


圖 1-1 研究架構圖

第二章 文獻回顧

一般對於地下水水質分析以建立地下水監測井網最為直接，並現地採樣水質資料或進研究室作進一步分析，應用統計方法而得到此區域之地下水特性。地下水水質具有高度複雜及異質性，本研究應用類神經網路具有強大非線性及複雜問題之處理能力，將水質因子高維度資料運算至低維度特徵空間，以簡單的特徵性質描述原本複雜的資料型態。由以下水質分析之研究文獻，包含應用其他統計分析方法分析水質、以地質年代觀點探討砷濃度來源及離子之化學性質探討水質特徵之文獻，延伸出本研究所探討之重點，應用自組特徵映射網路(SOM)探討地下水之水質分析，並推估地下水砷濃度。以下文獻回顧中將探討砷污染相關研究，一般地下水水質統計分析方法與應用類神經網路於分析水質之研究，另外對自組特徵映射網路的應用也將做一簡介，分述如下：

2.1 地下水中砷污染相關研究

砷(Arsenic, As)是一種毒性很強且具有致癌性的重金屬元素，對於生物體具有強烈之毒性，因此，砷對環境與人類健康之影響係為社會所關切之議題，含砷之飲用水被國際癌症研究署(IARC)認定為第一類致癌物質，而在 1993 年世界衛生組織(World Health Organization, WHO)已經將飲用水中之砷濃度最大允許標準由 50 μ g/L 降低為 10 μ g/L(WHO, 1993)，我國於 2000 年 12 月實行飲用水水質標準為 10 μ g/L(環保署，1998)，至今臺灣飲用水水質標準(環保署，2000)、美國飲用水標準

(USEPA, 2006)、世界衛生組織的飲用水標準(NRC, 2001; WHO, 2006)皆以 10 μ g/L 作為標準，顯示砷之毒性與對人體健康之危害，受到國際間的重視。

近幾年全世界有不少國家的地下水受到高濃度砷之影響，這個問題受到各地關注且提出相關研究，在孟加拉、越南、智利、中國、墨西哥等地皆有飲用地下水導致砷中毒的案例(Polya et al., 2005; Berg et al., 2001)。其中以孟加拉超過四千萬人飲用含砷之地下水，引起了大規模之癌症、色素沉著症與類似烏腳病之四肢壞疽最受到關切。Smedley and Kinniburgh (2002)彙集相關文獻，分析全球地下水含水層受到砷影響之區域，將造成高砷濃度的原因分成三類：(一)還原環境，如臺灣、孟加拉、中國等；(二)氧化環境且在 pH 值偏高的環境下，如智利、墨西哥等；(三)含水層氧化、還原變動大所引起，如美國西南部。其中孟加拉為受到地下水砷污染最廣的國家，含水層屬於全新紀沉積物(Holocene Alluvial)，地下水之酸鹼度偏中性，有高鹼度存在，含水層為還原環境，且流速慢；臺灣地下水處於強還原環境，部分受砷污染之地下水有腐植酸存在；中國內蒙古地下水之特性處於強還原環境，且具有高鹼度與部分地下水存在腐植酸；智利與墨西哥之含水層屬於火成岩，地下水處於氧化狀態；另外，美國達科塔南部含水層特性與湖泊的沈積物有關。地下水受到高砷濃度影響之在全球分布如圖 2-1 所示。

目前地下水中高濃度的砷釋出，因地下水中地質環境不同與存在複雜的化學反應，目前相關研究有很多不同方面的證據，得到不同的結論。本研究歸納相關文獻，篩選出與本文中相關的因子，在下面列

出過去研究探討適宜砷釋出於地下水中，造成砷濃度上升的環境與水質特徵：

1. 氧化還原環境：在氧化地層中，伴隨著鐵及硫酸根離子濃度上升且 pH 值下降的環境，利於砷釋出而存在高砷濃度地下水 (Schreiber et al., 2000; Raessler et al., 2000)。另外有研究指出在還原地層中，也有利於砷釋出的條件，地下水中砷的化合物會被膠體、鐵錳氧化物或氫氧化物所吸附，當氧化還原電位到達還原電位時，膠體變得不穩定，或是對於砷有高吸附能力的鐵錳氧化物或氫氧化物被還原，生成溶解度較高的亞鐵離子，則吸附在表面的砷隨著釋放到地下水中，因此在此類地下水中，處於高鐵、高錳、低溶氧且硫酸根濃度較低的環境下，可能使得砷釋出，在地下水中存在高濃度砷 (Smedley and Kinniburgh, 2002)。
2. 陰離子競爭吸附：Anawar et al. (2004)、Kim et al. (2000) 與 Goh and Lim (2005) 指出當環境的陰離子濃度高時，會使吸附在污染土壤上的砷脫附釋出。硫酸根離子與碳酸根亦會與砷離子競爭吸附在土壤顆粒上 (Kim et al., 2002)。
3. pH 值條件：pH 值也會影響砷在沉積物表面的吸附與去吸附作用。當環境處在 pH 低值時，砷離子較易被吸附在金屬氧化物或氫氧化物表面；一旦 pH 值增加時，吸附在氧化物或氫氧化物表面的砷離子將會減少 (Pierce and Moore, 1982; Xu et al., 1988)。Smedley and Kinniburgh (2002) 也指出，高 pH 值環境的地下水造成五價砷離子不容易被吸附在鐵或錳的氧化物上。

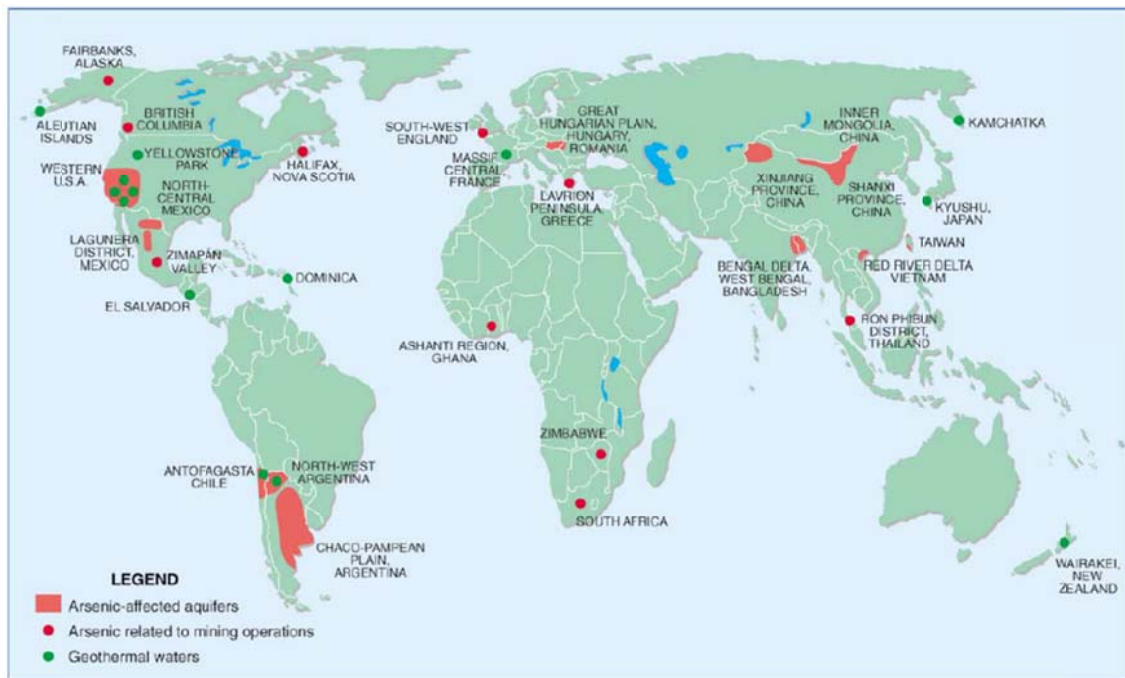


圖 2-1 全球地下水含水層受砷區域分布圖

(Smedley and Kinniburgh, 2002)

2.2 分析地下水水質方法

地下水水質變化呈現高度複雜與異質性，在本研究區域與臺灣其他區域之地下水分析中，應用統計多變數理論為一種相當普遍的方法，Lee et al. (2008)應用判別分析(Discriminant analysis)及化學地質模式分析臺灣宜蘭平原地層狀況，將此區域分類為氧化、過度與還原狀態，並研究在還原狀態下，地層中砷釋出與移動的機制。Kuo and Chang (2010)應用動態因子分析(Dynamic Factor Analysis, DFA)探討影響地下水中砷濃度的關鍵因子，並以臺灣雲林沿海為例，結果顯示沿海地下水中的砷濃度受到鄰近地區地下水中砷濃度、鹼度與總有機碳的濃度

影響最為明顯。Lu et al. (2010)以礦物主成分及表面化學分析與連續萃取方法，針對濁水溪沖積扇岩心中砷釋出的機制作探討，於濁水溪沖積扇扇央與扇尾監測出高砷濃度地下水，而地下水砷濃度與碳酸根(HCO_3^-)、氨根(NH_4^+)之相關性佳，但與總有機碳(TOC)及鈣離子(Ca^{2+})之關係較差，顯示有機質參與反應，主要作用為增加地下水之碳酸根及氨離子濃度，使地下水環境趨於還原狀態。本區地層中砷可能釋出機制，應與無晶型之鐵氧化物或鐵氫氧化物還原溶解有關，且碳酸根之取代也是主要反應之一。Chang et al. (2010)應用倒傳遞類神經網路，建立空間模式補遺地下水中砷濃度資料，並應用主成分分析、交叉驗證法及修正型目標函數，解決建立空間模式時遭遇到資料過少、模式過度訓練等問題。Liu et al. (2003)利用因子分析法將雲林沿海地區地下水水質歸納為「鹽化因子」與砷污染因子，其中鹽化因子包含電導度(EC)、總溶解固體(TDS)、氯鹽(Cl^-)、硫酸鹽(SO_4^{2-})、鈉(Na^+)、鉀(K^+)與鎂(Mg^{2+})；而砷污染因子為鹼度(Alk)、總有機碳(TOC)與砷(arsenic)。

2.3 自組特徵映射網路之應用

類神經網路之應用相當廣泛，在國內外已有於不同領域之應用，如流量預測(Chang et al., 2001, 2002, 2006)、水資源管理(Chang and Chang, 2001)、蒸發散量推估(Chang et al., 2010；高慧珊，2007；李品輝，2009)、生態多樣性研究(蔡文柄，2007)。自組特徵映射網路(SOM)適合將高維度且有非線性關係的資料進行聚類，廣泛地應用在影像、語音及訊號辨識，也應用於生態與環境學各方面。Grieu et al. (2006)應用 SOM 在推估污水處理廠化學物質的排放量。將污水處理過程相關的

變數進行聚類，並將結果中變數的權重，依各變數權重相關性較高的變數，作為倒傳遞類神經網路的輸入，進行化學物質入流量與出流量的推估。Martos et al. (2002)應用 SOM 分析多種地下水水質離子濃度變化，歸納在卡氏座標網格，以簡化西班牙南部複雜的水質狀況，呈現地下水水質在卡氏座標中的位置即為該水質污染的情形與特性。Lee and Miklas (2006)應用 SOM 之拓樸圖解釋實驗溼地中，不同離子濃度影響濕地去除重金屬的關係，並挑選出達成推估有效去除重金屬濃度關鍵的輸入因子。



第三章 理論概述

本研究以類神經網路為主要研究方法，使用自組特徵映射網路(SOM)、輻狀基底函數類神經網路(RBFNN)與作為驗證本研究所建立之模式的倒傳遞類神經網路(BPNN)。另外因地下水水質因子變異性大，避免人為決定資料分配造成誤差，故應用交叉驗證法優選模式。而模式推估的誤差以本研究所應用評估指標作為評估的參考。以上方法與理論分述如下：

3.1 類神經網路

類神經網路(Artificial Neural Networks, ANN)係以電腦來模擬人類大腦運作的科學，具有從輸入的資訊中擷取並累積經驗、儲存知識，進而加以利用的智慧型演算程序。其模仿生物神經網路架構的平行計算模式，透過許多高度連結的神經元(或處理單元)的組成，以模仿生物神經系統的功能與運作，其主要功能在於學習(Learning)、回想(Recall)及歸納推演(Generalization)。我們能夠藉由調整單元間的連結值(即權重值)來訓練類神經網路以執行某個特定的功能，將輸入與輸出變數間複雜的對應關係充份呈現，為解決非線性動態問題的強健工具。

一般類神經網路結構包含：輸入層、隱藏層、輸出層。輸入層是由外界接受訊息並將之傳入類神經網路中，以方便進行訊息的處理，輸入層單元數目由影響問題之輸入變數個數加以決定；隱藏層可能有數層，可將輸入與輸出層各處理單元間的相互關係透過連結權重值充份地表現出來，其單元數目需以試驗的方式決定其最佳數目。原則上

隱藏層處理單元太多除了效率不好外，也會造成過度訓練(over fitting)的問題，然而太少又無法處理較複雜的事件(張斐章與張麗秋，2010)；輸出層則接受網路處理後的訊息，並將結果傳送到外界，其神經元個數與待預測之目標數相同。

類神經網路的連結架構，可分為前饋式類神經網路與回饋式類神經網路兩類。前饋式類神經網路其連結方式為單一方向的向前傳遞連結，且網路的所有神經元皆無後向或是側向的傳遞連結，如圖 3-1 所示，本研究後續所使用的即為前饋式類神經網路；回饋式類神經網路或稱為遞迴式網路與前饋式類神經網路最大不同在於網路的回饋項傳到前一層產生了遞迴的演算，如圖 3-2 所示。

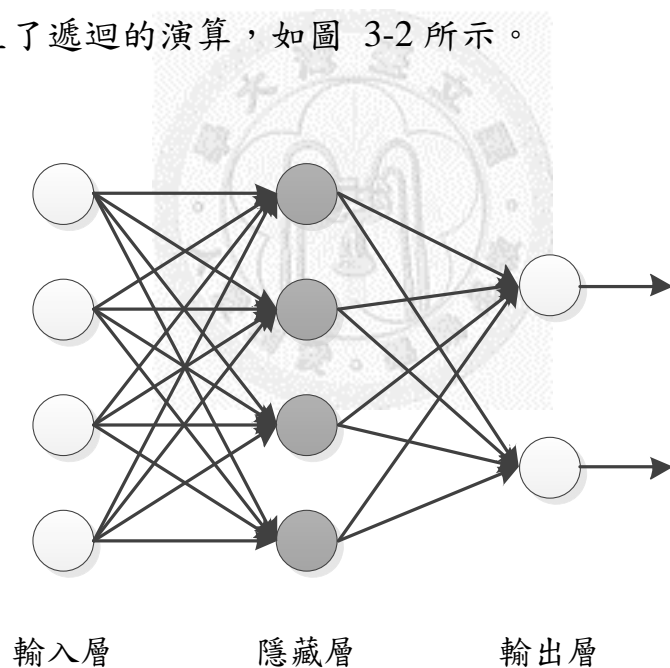


圖 3-1 前饋式類神經網路架構圖

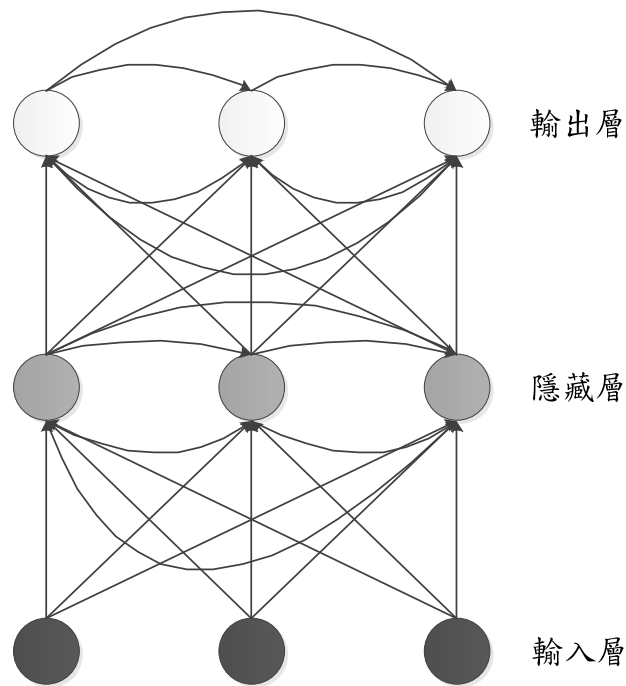


圖 3-2 回饋式類神經網路架構圖

類神經網路能由過去之經驗加以學習訓練，以不斷的訓練調整網路中各神經元間之連結權重值。當訓練完成時，訓練資料中所有的經驗知識即儲存於這些連結權重值上。類神經網路的整體運作模式包含(I)學習過程(Learning)：網路依學習演算法，從範例中學習，以調整網路權重值的過程；(II)回想過程(Recalling)：網路依回想演算法，以輸入資料決定網路輸出資料的過程。

針對不同的問題，類神經網路訓練時所應用之學習法則也有所差異，而配合學習方式的不同亦發展出許多不同型式之類神經網路。一般而言，學習過程的學習演算法可分為兩類：(1)監督式(Supervised)學習：從問題領域中取得訓練範例，並從中學習輸入變數與輸出變數的內在對應規則，以應用於新的案例。然而學習結果會受限於學習樣本資料的影響，在預測值大於學習樣本時往往會產生比較大的誤差，此

類學習法多應用於求解型態判別(pattern classification)及曲線配適(curve fitting)問題。(2)非監督式(Unsupervised)學習：從問題領域中取得訓練範例，並從中學習範例的內在分類規則，以應用於新的案例，就輸入變數值，推論它與那些訓練範例屬同一類別，多應用於求解群落分析問題。圖 3-3 為監督式學習與非監督式學習示意圖。

類神經網路之所以能廣為眾人重視及投入的研究，主要原因在其具有處理訊息的強健能力，與優越的非線性映射能力。過去應用各種不同架構之類神經網路於河川水位及流量的預測(Chang and Chen, 2003； Chang et al., 2005； Chang et al., 2007)，顯示依河川特性及資料結構等考量，不同類神經網路模式皆有良好的預報能力。

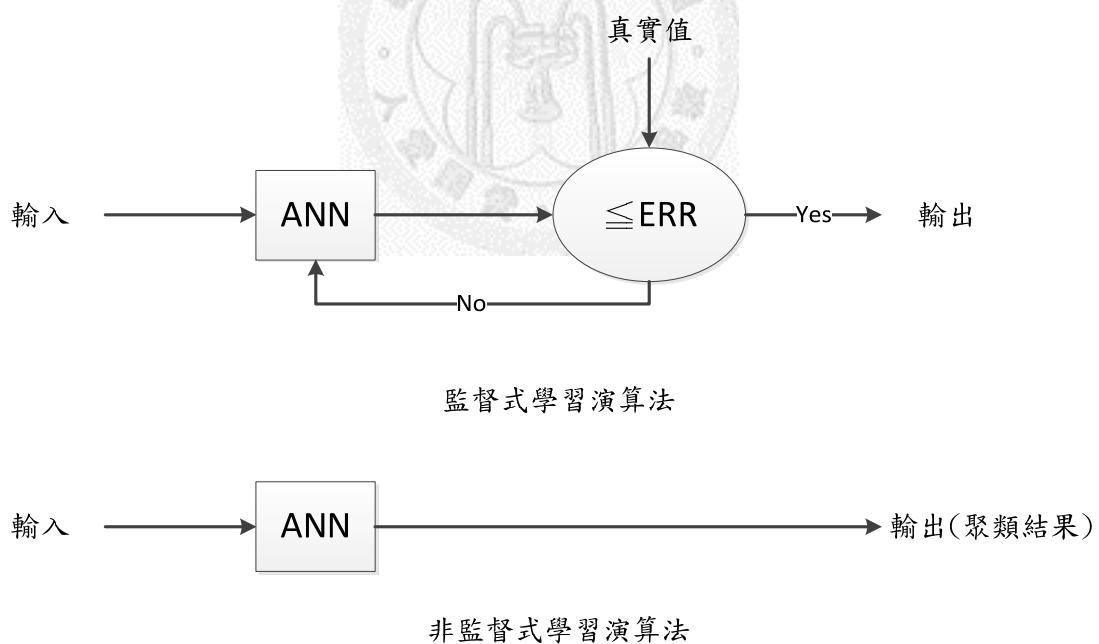


圖 3-3 類神經網路學習方法

3.1.1 自組特徵映射網路(SOM)

自組特徵映射網路(Self-Organizing Map, SOM)於 1992 年首先由 Kohonen 提出，屬於前饋式、非監督式神經網路，同時也是競爭式神經網路。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如圖 3-4 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向量向量的機會；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)如圖 3-5 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，由於所產生的拓樸結構圖可以反映所有輸入值間的分布關係，因此將此網路稱作為自組特徵映射網路，而該映射圖也可稱為拓樸圖(topology)。

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，透過此演算法，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。而 SOM 拓樸層也可稱之為聚類中心層，其中的每一個神經元皆可代表輸入向量的中心點。由圖 3-4 所示，輸入層代表的是輸入變數，每個輸入變數透過權重來連結拓樸層中的神經元，這些神經元代表著拓樸映射的結果。拓樸輸出層上的每個神經元都隱含著拓樸座標，如圖 3-4 所示，輸入層代表的是輸入變數，每個輸入變數透過權重來連結拓樸層中的神經元，這些神經元代表著拓樸映射的結果。拓樸輸出層上的每個神經元都隱含著拓樸座標，如圖 3-4 中的 25 個神經元間的拓樸座標關係可根據其鄰近的相對座標位置，對應圖 3-5 的拓樸座標圖。

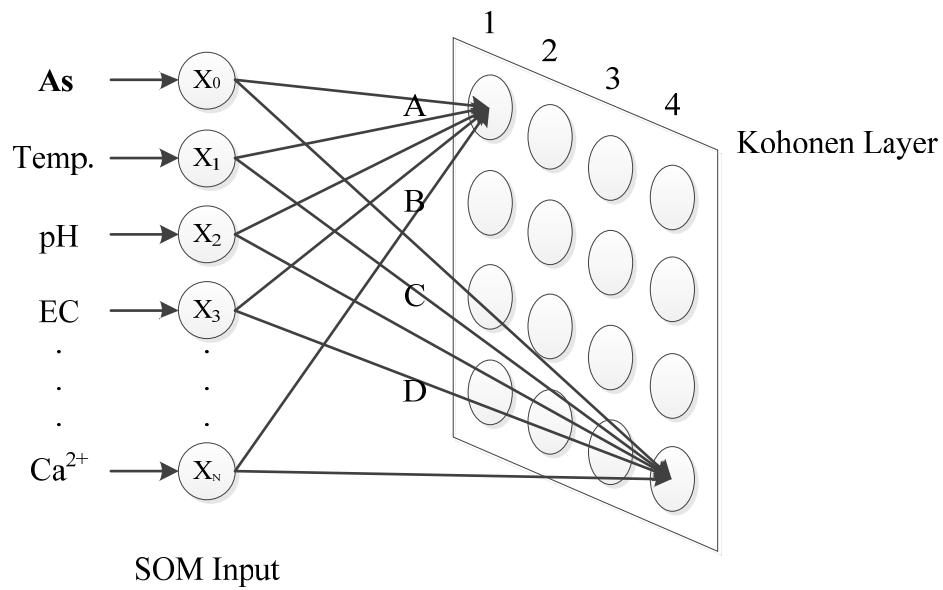


圖 3-4 二維矩陣的 SOM 架構圖

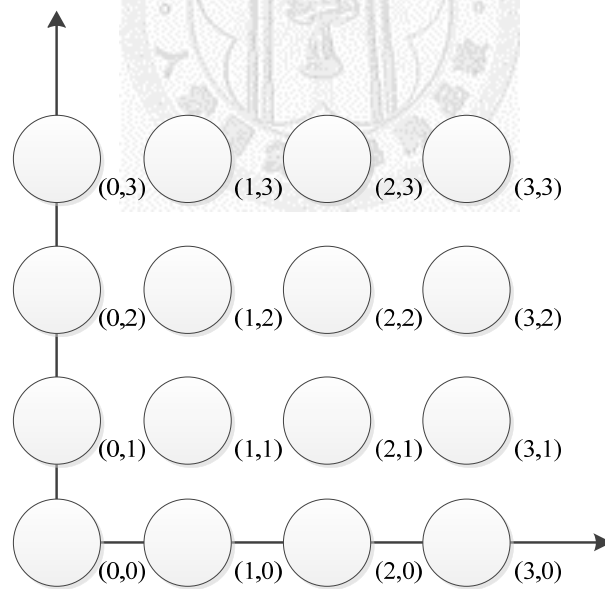


圖 3-5 SOM 網路神經元間拓樸座標

在 SOM 網路中，每一筆輸入向量透過連結權重與網路神經元相連接，各神經元間以競爭式學習法則來決定被調整的神經元，即為說，每筆輸入向量都必須尋找其對應的優勝神經元，其為該輸入向量最近似的神經元，進而調整該神經元的連結權重。

對於輸入向量 X ，可以用(3-1)式表示，

$$X = [x_1, x_2, x_3, \dots, x_n]^T \quad (3-1)$$

其中 n 表示輸入維度。第 i 個神經元的連結權重可用 (3.2) 式表示，

$$W_i = [w_{i1}, w_{i2}, w_{i3}, \dots, w_{im}]^T \quad (3-2)$$

其中 $i=1,2,\dots,m$ ， m 代表的是神經元數目；其初始權重值可以隨機亂數產生，建議在輸入範圍的中心位置設定一均勻分佈的隨機亂數，以讓網路在形成階段中有較快的收斂速度(張斐章與張麗秋，2010)。每一筆輸入向量都可透過連結權重與網路中的每個神經元連結，各神經元間以競爭式學習法來決定被調整的神經元。經由這層關係，可求得每筆輸入向量所對應出的優勝神經元，也就是與該輸入向量最近似的神經元。決定優勝神經元的方式是比較所有神經元的連結權重與輸入向量間的距離，並選取距離最近的神經元為優勝神經元。在此採用歐氏距離(Euclidean norm)評估輸入向量與神經元所連結的權重向量是否近似，

$$q(X) = \min_{\forall i} \|X - W_i\|^2 \quad (3-3)$$

在(3-3)式中的 $q(X)$ 為神經元權重 W 與輸入向量 X 的最短距離，造成此最短距離的神經元即為優勝神經元。

由於每筆輸入向量所對應的優勝神經元可能相同也可能不同，因此每個神經元被調整的次數皆不盡相同，這樣的初步分類是一種聚類分析的過程；然而這樣的過程是無法將神經元間的拓樸關係表現出來，即聚類間無連續的關係存在，因此設定神經元間的鄰近關係，如鄰近半徑、鄰近區域、鄰近函數及鄰近距離並適切地隨著優勝神經元進行調整，有助於網路神經元間的拓樸映射關係。鄰近半徑、鄰近區域、中心的意義，如圖 3-6 所示。

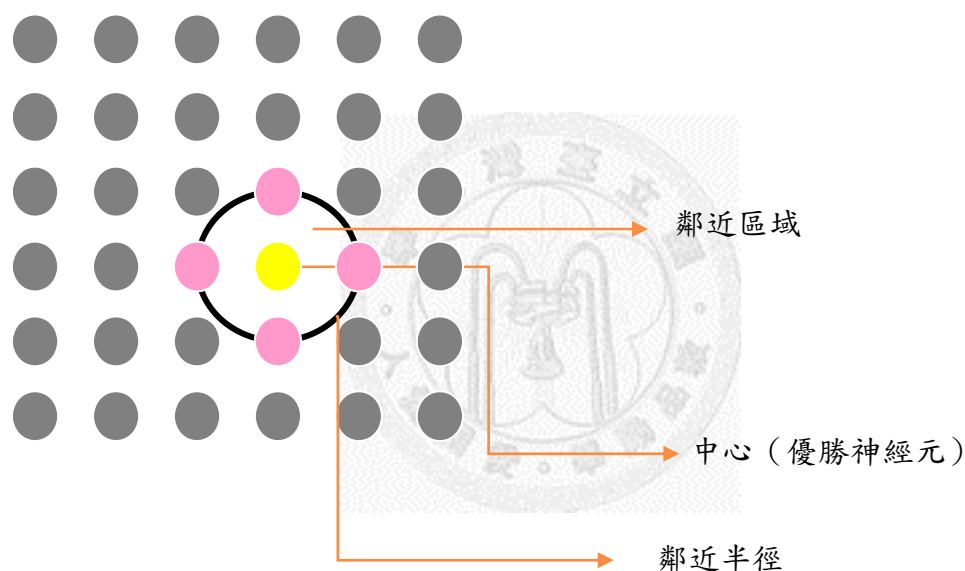


圖 3-6 拓樸層網路圖

(高慧珊，2007)

在找出優勝神經元之後，自組特徵映射網路演算法會依據優勝神經元於拓樸結構上的相對應位置，尋找出其鄰近區域相關的「鄰近神經元」，並對於優勝神經元以及其鄰近的神經元進行權重的調整工作。一般而言，初期所設定的鄰近半徑約略包含整個拓樸結構的範圍，隨著演算代數的增加，鄰近半徑會逐漸縮減，如圖 3-7 所示；這是因為

演算初期的鄰近半徑較大，整個網路能夠快速達到收斂的效果，而在演算後期則由於網路已達相當程度的穩定性，因此只需針對少數的神經元進行調整即可。

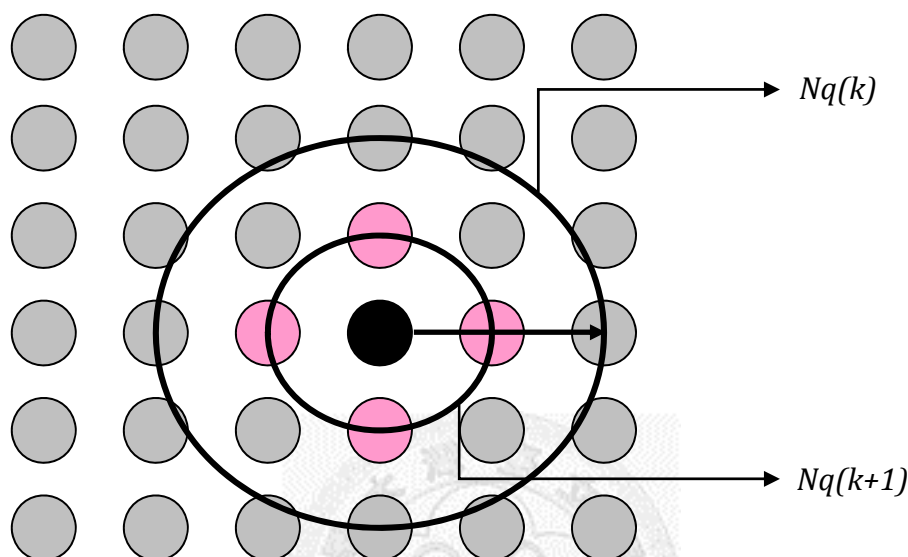


圖 3-7 演算代數與鄰近半徑之關係

(高慧珊，2007)

對每一筆輸入向量而言，自組特徵映射網路演算法只會更新其優勝神經元與其相對應的鄰近神經元，這也就是自組特徵映射網路為競爭式學習最具代表性的原因之一；其更新權重的公式則如(3-4)式所示。

$$\eta_{qj} = \exp\left(\frac{-\|r_j - r_q\|^2}{R^2}\right) \quad (3-4)$$

$$W_j(k+1) = W_j(k) + \Delta W_j$$

$$\Delta W_j = \mu(k)\eta_{qj}(k)[X(k) - W_j(k)]$$

其中 W 表聚類中心層神經元權重、 X 為輸入向量、 μ 表學習速率、

η 表示鄰近函數、 η_{qj} 為第 j 個神經元與優勝神經元 q 之鄰近關係值、 r_j 為鄰近區域中其他神經元於拓樸圖上的位置座標、 r_q 為優勝神經元於拓樸層上的位置、 R 為鄰近半徑、 k 為迭代次數。一般而言學習速率 μ 的值介於 0 至 1 之間，而其與鄰近半徑相同地也會隨著演算代數的增加而逐漸減少。圖 3-8 為 SOM 的演算流程圖。

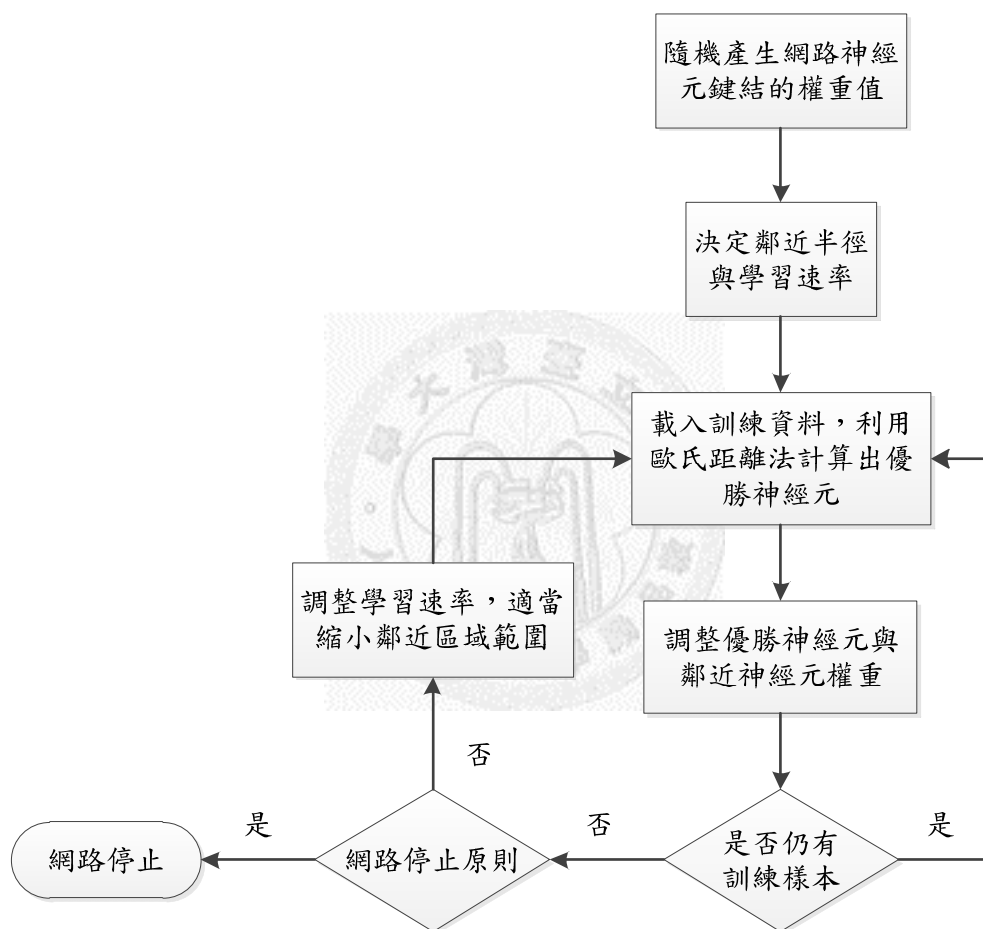


圖 3-8 SOM 演算流程

3.1.2 輻狀基底函數類神經網路(RBFNN)

輻狀基底函數類神經網路(Radial Basis Function Neural Network, RBFNN)，或稱半徑式類神經網路，其特質主要在於模擬大腦皮質層軸突的局部調整功能，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，是屬於基本前饋式類神經網路的組合，其優點在於可大量減少學習時間；但不同於其他的類神經網路型態，RBFNN 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)針對這個問題作了一些探討，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(Broomhead and Lowe, 1988)，解決內插與近似的理論。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，也就是在隱藏層的各神經元中建立相對應的輻狀基底函數。在此以 N 個維度的輸入值、 M 個神經元的隱藏層與一個輸出值作為 RBFNN 架構與演算法說明，其架構如圖 3-9 所示。

在 RBFNN 中每一層所扮演的角色各有不同，輸入層是輸入資料與網路連結的介面層，再將輸入資料經過非線性活化函數轉換到隱藏層，也就是將輸入空間 $X(X \in \mathfrak{R}^{N \times 1})$ 進行非線性映射到隱藏空間 $Z(Z \in \mathfrak{R}^{M \times 1})$ ，輸出層則是將隱藏層的輸出進行線性組合（或線性映射）獲得輸出值。

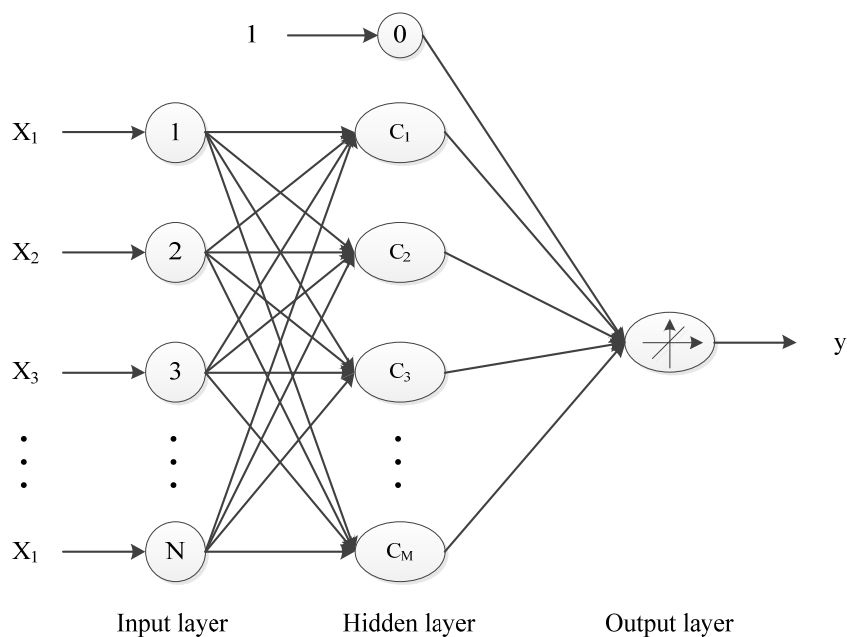


圖 3-9 輻狀基底函數類神經網路架構圖

RBFNN 演算方法：

當訓練範例資料輸入網路後，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，也就是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(3-5)式：

$$z_j(x) = \phi(\|x - c_j\|) \quad j = 1, 2, \dots, M \quad (3-5)$$

式中 $\phi(\cdot)$ 表輻狀基底函數， c_j 表隱藏層第 j 個神經元中心點， $\|x - c_j\|$ 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(3-6)式：

$$y = \sum_{j=0}^M w_j \cdot z_j(x) \quad (3-6)$$

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。將(3-5)式代入(3-6)式可寫成(3-7)式：

$$y = \sum_{j=1}^M W_j \cdot \phi(\|x - c_j\|) + w_0 \quad (3-7)$$

一般來說，隱藏層中的輻狀基底函數，其型式有：線性函數、三次函數、薄平面曲線函數、高斯函數等，如圖 3-10 所示，在本研究選擇以高斯函數為輻狀基底函數，如(3-8)式所示。

$$\phi(\|x - c\|) = \exp\left(-\frac{\|x - c\|^2}{2\sigma^2}\right) \quad (3-8)$$

式中 σ_j 為控制輻狀基底函數平緩度之參數。

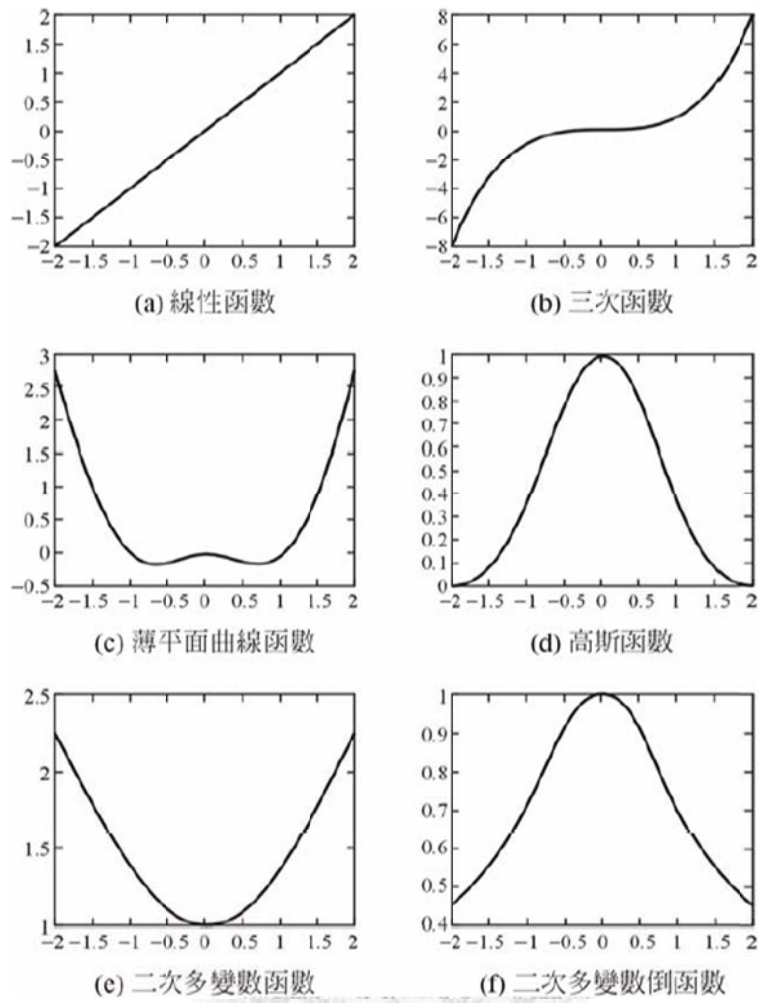


圖 3-10 各形式之輻狀基底函數

(摘自張斐章、張麗秋，2010，類神經網路導論)

由於自組特徵映射網路(SOM)演算法為非監督式的學習演算法，因此在其訓練階段完成後，任一輸入向量皆可獲得其相對應的聚類中心層之神經元，但整個網路架構沒有輸出向量，因此針對聚類中心層所產生的結果再加以架構一層網路輸出層如圖 3-11 使得自組特徵映射網路得以擷取輸入向量的統計聚類特性而獲得輸出值 (Chang and Chang, 2010)；本研究以接輻狀基底函數類神經網路來架構網路輸出層。

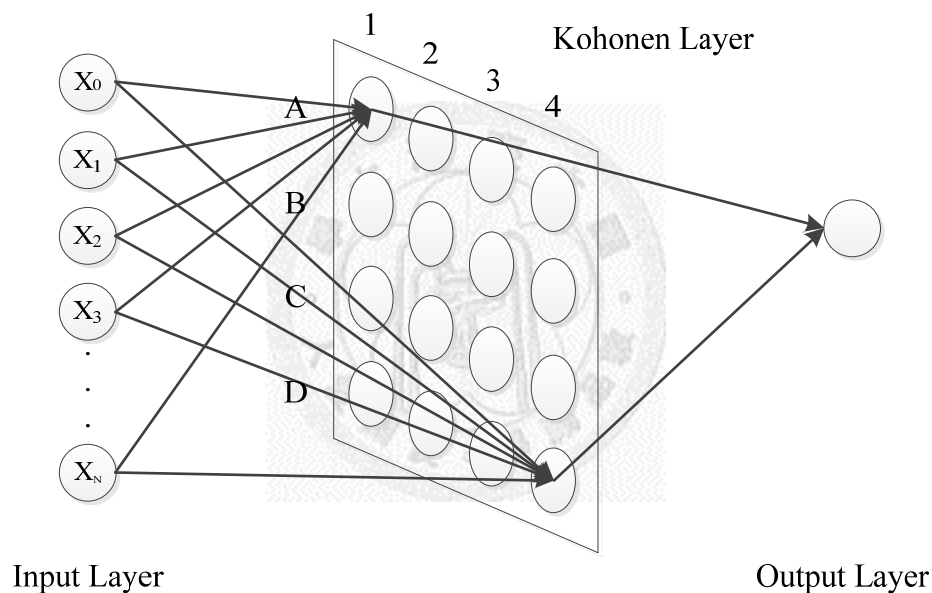


圖 3-11 架構自組特徵映射網路輸出層

3.1.3 倒傳遞類神經網路(BPNN)

倒傳遞類神經網路的架構為多層感知器 (Multilayer Perceptron, MLP)，如圖 3-12 所示，其架構為多層前饋式類神經網路，採用監督式學習法(supervised learning)，一般使用的學習演算法為誤差倒傳遞演算法(Error Back Propagation, EBP)，簡稱為 BP 演算法。結合 MLP 與 EBP 的類神經網路即稱之為倒傳遞類神經網路(BPNN)。一般常用的倒傳遞類神經網路架構由三層所構成，包含輸入層、隱藏層(hidden layer)及輸出層，隱藏層是在輸入層與輸出層之間，與外界資訊隔離，實際上輸入層如同單層感知器的輸入層，並不對輸入資訊作處理，只是當外界輸入訊息的傳遞介面，將輸入訊息直接往後傳遞，因此實際有作用的神經元只有隱藏層及輸出層兩層。隱藏層的層數可依問題複雜度由一層增加到數層，但一般來說不需超過兩層以上。至於神經元數目則依問題的型式而定，隱藏層神經元數目若過少則無法有效描述問題的型態，過多則可能造成過度描述(Over-fitting)而無法有效應用，且會使得網路歸納推演的能力降低並徒增訓練的時間。至於選取多少神經元數目並無一定標準，一般皆依問題的形式而定，本研究則以多數專家建議之試誤法(trial-and-error)來決定網路神經元個數。

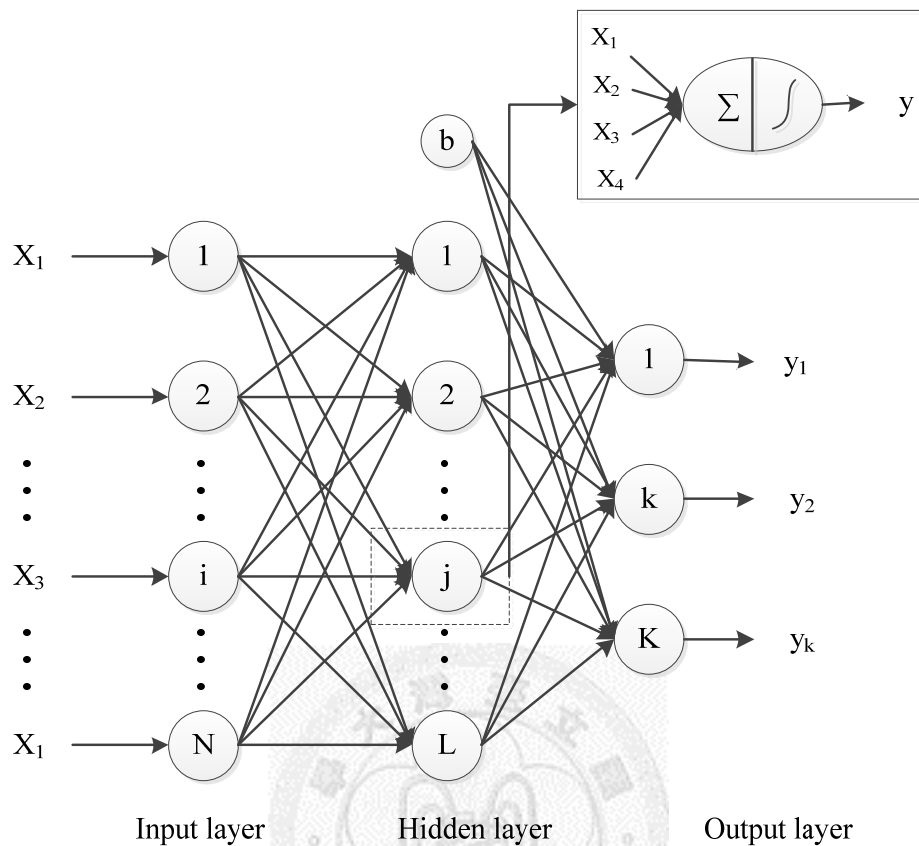


圖 3-12 倒傳遞類神經網路架構圖

網路中靠相關權重連結各層間之神經元，輸入值由輸入層直接傳入隱藏層，經加權累加後再透過活化函數轉換可得一輸出值，同理再傳入輸出層。常使用的活化函數型式相當多，包括線性轉移函數、對數雙彎曲轉移函數、正切雙彎曲轉移函數及線性轉移函數等。

倒傳遞演算法的學習過程，其學習過程是由正向傳遞與負向傳遞所組成，在正向傳遞過程中，輸入訊息從輸入層經由隱藏層加權運算、透過轉換函數處理後，再傳向輸出層運算後輸出，每一層神經元的狀態只影響下一層神經元的狀態；若網路輸出值與目標輸出值誤差範圍不在容忍範圍內，則轉為反向傳播，將誤差訊號沿原來的連接通路回

傳，透過修改各層神經元的權重與偏權值，期能使誤差值達到容忍誤差範圍之內而停止。

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：

$$y_j^n = f(\text{net}_j^n) \quad (3-9)$$

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為

$$\text{net}_j^n = \sum_i w_{ji}^n y_i^{n-1} - b_j^n \quad (3-10)$$

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，目標函數 E 定義為

$$E = (1/2) \sum_k (d_k - y_k)^2 \quad (3-11)$$

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，即為使目標函數 E 最小化的過程，而求解最佳化問題之數值方法有很多種，可分為一階梯度法與二階梯度法，一階梯度法包括梯度下降法 (gradient method)、最陡坡降法 (steepest descent method) 及共軛梯度法 (conjugate gradient method)；二階梯度法包括牛頓法 (Newton method)、高斯-牛頓法、擬牛頓法及 Levenberg-Marquardt 法，若此兩種演算法求解過於耗時可採用啟發式

的技巧加以改善，如可變學習速率(Variable Learning Rate)或有彈性的(Resilient)演算法改善傳統演算法過於耗時之缺點。一般而言，標準的倒傳遞類神經網路是指使用最陡坡降法來搜尋 E 的最小值；最陡坡降法是一個相當古老且廣為用於求解最佳化問題的數值方法，搜尋向量方向則是沿著目標函數負梯度的最陡方向搜尋目標函數的最小值。

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} \quad (3-12)$$

式中 η 為學習速率，其大小決定了最陡坡降法修正的步幅，而 $\frac{\partial E}{\partial w_{ji}}$ 可用微積分學的連鎖律 (chain rule) 得

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial net_j^n} \frac{\partial net_j^n}{\partial w_{ji}} = \frac{\partial E}{\partial y_j^n} \frac{\partial y_j^n}{\partial net_j^n} \frac{\partial net_j^n}{\partial w_{ji}} \quad (3-13)$$

將(3-9)式代入 $\frac{\partial y_j^n}{\partial net_j^n}$ 得

$$\frac{\partial y_j^n}{\partial net_j^n} = f'(net_j^n) \quad (3-14)$$

將(3-10)式代入 $\frac{\partial net_j^n}{\partial w_{ji}}$ 可得

$$\frac{\partial net_j^n}{\partial w_{ji}} = \frac{\partial}{\partial w_{ji}} (\sum_i w_{ji} y_i^{n-1} - b_j) = y_i^{n-1} \quad (3-15)$$

(3-13)式的 $\frac{\partial E}{\partial y_j^n}$ 可分成兩種情況：

(1) 若第 n 層為最終層，即網路的輸出層：

此時， $j=k$ ；將(3-12)式與 $j=k$ 代入 $\frac{\partial E}{\partial y_j^n}$ 得

$$\frac{\partial E}{\partial y_j^n} = \frac{\partial}{\partial y_j^n} [(1/2) \sum_k (d_k - y_k^n)^2] = -(d_j - y_j^n) \quad (3-16)$$

(2) 第 n 層不是最終層，也就是網路的隱藏層之一，可連鎖律得

$$\frac{\partial E}{\partial y_j^n} = \sum_k \left(\frac{\partial E}{\partial net_k^{n+1}} \frac{\partial net_k^{n+1}}{\partial y_j^n} \right) \quad (3-17)$$

將(3-10)式代入 $\frac{\partial net_k^{n+1}}{\partial y_j^n}$ 可得

$$\frac{\partial net_k^{n+1}}{\partial y_j^n} = \frac{\partial}{\partial y_j^n} (\sum_i w_{ki} y_i^n - b_k) = w_{kj} \quad (3-18)$$

為簡明定義，令

$$\frac{\partial E}{\partial net_k^n} = -\delta_k^n \quad (3-19)$$

將(3-18)、(3-19)式代入(3-17)式：可得下式(3-20)

$$\frac{\partial E}{\partial y_j^n} = -\sum_k (\delta_k^{n+1} w_{kj}) \quad (3-20)$$

總結上述 $\frac{\partial E}{\partial w_{ji}}$ 可分為兩種情形：

(1) w_{ji} 處於輸出層與隱藏層之間

$$\frac{\partial E}{\partial w_{ji}} = -(d_j - y_j^n) f'(net_j^n) y_i^{n-1} \quad (3-21)$$

將(3-15)、(3-19)式代入(3-13)式：

$$\frac{\partial E}{\partial w_{ji}} = -\delta_j^n y_i^{n-1} \quad (3-22)$$

比較(3-21)式與(3-22)式得

$$\delta_j^n = (d_j - y_j^n) f'(net_j^n) \quad (3-23)$$

(2) w_{ji} 處於隱藏層與隱藏層之間

$$\frac{\partial E}{\partial w_{ji}} = \left[-\sum_k \delta_k^{n+1} w_{kj} \right] f'(net_j^n) y_i^{n-1} \quad (3-24)$$

將(3-15)、(3-19)式代入(3-13)式：

$$\frac{\partial E}{\partial w_{ji}} = -\delta_j^n y_i^{n-1} \quad (3-25)$$

比較(3-24)式與(3-25)式得

$$\delta_j^n = \left[\sum_k \delta_k^{n+1} w_{kj} \right] f'(net_j^n) \quad (3-26)$$

因此不論 w_{ji} 處於隱藏層與隱藏層之間，或 w_{ji} 處於輸出層與隱藏層之間，均可寫成通式：

$$\frac{\partial E}{\partial w_{ji}} = -\delta_j^n y_i^{n-1} \quad (3-27)$$

式中 δ_j^n 為將(3-27)式代入(3-12)式可得

$$\Delta w_{ji} = \eta \delta_j^n y_i^{n-1} \quad (3-28)$$

此式即倒傳遞演算法之關鍵公式，每輸入一組訓練範例(也就是有輸入項與目標輸出值 input-output pair)，各連結的權重即可依此方式，加以調整。

$$w_{ji}(p) = w_{ji}(p-1) + \Delta w_{ji} \quad (3-29)$$

式中 p 為第 p 組訓練範例。由於本研究採用地下水水質資料作為網路之訓練範例，水質因子的變異性大，避免因人為選用訓練樣本資料造成誤差，本研究另採用交叉驗證法對於網路架構與參數進行優選模式，於後續章節中詳細說明。

3.2 交叉驗證法優選模式

交叉驗證法(cross validation)應用於資料過少或資料變異性大時，避免人為決定訓練與驗證資料所造成的誤差，一般決定訓練與驗證資料是參考資料的統計特性作分配，但是若資料變異性很大時，以人為選擇訓練與驗證資料，部分極端值容易因為被平均掉，造成資料之統

計特性無法正確表現出資料的真實情況，在架構類神經網路時，將顯著影響模式學習，為消除人為效應，本研究採用交叉驗證法，使訓練與驗證資料充分交換，使建立之類神經網路達到充分訓練。其計算步驟說明如下。

1. 假設每組資料集合具有相同 N 筆(假設 50 筆)水質資料，包括砷濃度及其他水質因子。而不同資料集合差別在模式開始的輸入值不一樣，例如學習速率、迭代次數與神經元個數等。
2. 將每組資料集合之 N 筆水質，切割為 M 段(假設 5 段)，每段中資料筆數為 $N/M(50/5=10$ 筆)。
3. 在訓練階段，選用 M 段中 $M-1$ 段作為訓練段，在本研究中為選擇 4 段(編號 11~50 資料，共 40 筆)作為訓練資料，在訓練過程中，藉由目標函數調整網路中權重，最後計算網路之估計值與實際值之差，計算訓練誤差。
4. 驗證過程中，選用訓練不使用之 1 段(1~10 筆資料)作為驗證之用，並根據訓練過程中得到網路權重，計算驗證過程中，網路輸出之估計值與實際值之差，即為驗證誤差，此為完成第 1 段交叉訓練，接著回到步驟(3)，重新選擇第 2 段(編號 11~20 資料)作為驗證資料，其餘資料第 1 段(編號 1~10 資料)及第 3~5 段(編號 21~50 資料)作為訓練資料，直至 1~5 段資料均作為驗證資料，並且其他段資料均作為訓練資料後，即可稱完成本資料集合之交叉驗證工作，可得到 M 組(5 段)之誤差。
5. 在本模式中將資料分為 M 段(假設 5 段)，故在分段訓練及驗證過程中可得到 M 組訓練及驗證之誤差值，將 M 組誤差值平均後，可

得到平均訓練及驗證之誤差值，本誤差可客觀表示本網路架構之誤差值。

6. 完成第 1 組資料集合之交叉訓練工作後，接續第 2 組資料集合之交叉驗證工作，直至所有水質模式均完成交叉驗證工作，即可優選出一組最佳的網路參數。

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，即可瞭解模式之穩定性及可靠性。

3.3 評估指標

本研究採用兩種評估指標，評估水質推估模式之準確性與適用性，針對此兩種評估指標之定義與公式，說明如下：

3.3.1 均方根誤差(RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (o_i - d_i)^2} \quad (3-30)$$

其中 d_i 為目標輸出值， o_i 為網路輸出值，RMSE 可避免誤差值因正負號而抵銷，較接近數據真實的情況，其值愈小表示預測值與實際值越接近，表示準確性愈高。

3.3.2 平均絕對誤差(MAE)

$$MAE = \frac{1}{N} \sum_{t=1}^N |d_i - y_i| \quad (3-31)$$

MAE 越小表示模式的準確性越高，尤其是再評估具有較多中低數值時，MAE 較能表現出模式之優劣。

3.3.3 正規均方誤差(NMSE)

$$NMSE = \frac{\sum(o_i - d_i)^2}{\sum(d_i - \text{average}(d_i))^2} \quad (3-32)$$

當 NMSE 值越小時，表示模式的精確度愈高，且本評估指標不受預測值大小之影響，能客觀表示不同資料的誤差。

3.3.4 相關係數(CC)

$$R = \frac{\sum_{t=1}^N (d_i - \bar{d})(y_i - \bar{y})}{\sqrt{\sum_{t=1}^N (d_i - \bar{d})^2 \sum_{t=1}^N (y_i - \bar{y})^2}} \quad (3-33)$$

CC 用於測量兩個變數之間的相關(線性相依)，其值介於-1 與 1 之間，值越接近 1 意味著變數間可以很好的由直線方程來描述，越接近 0 則表示兩個變數之間沒有線性關係。

第四章 研究案例

本章對研究區域之背景作一簡單描述，另外探討所蒐集到此研究區域相關資料，並闡述本研究在建立模式前對資料進行基本統計分析與資料前處理過程，分述如下：

4.1 研究區域概述

本研究之研究區域為雲林沿海地區，雲林縣位於臺灣的中南部，在嘉南平原最北端，北邊沿著濁水溪與彰化縣接壤，東西最寬有 50 公里，南北最長有 38 公里，全縣面積共 1,290 平方公里，其中九成為平原，一成為山地，人口約 73 萬人。

濁水溪流量豐枯懸殊，枯水期之水量在經上游灌溉取水後，已幾無地表餘水可用，故雲林縣目前自來水每日用水 26 萬噸，長期以來均使用地下水。地下水年抽用量推估約為 5 至 7 億立方公尺，年超抽量介於 5000 萬至 2 億立方公尺之間，長期的地下水超抽，導致雲林沿海地區地層下陷、海水倒灌、海水入侵、土質鹽化與排水不良等問題，靠海地區之自來水地下水井已檢驗出受砷污染之情況。

4.2 資料蒐集與處理

4.2.1 資料蒐集

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 與雲 3 公路；與口湖、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，以及蒐集水利署在許厝寮、新興、飛沙及口湖四處，共 7 口深水質監測井(編號#22 ~ #28)資料，其相對位置如圖 4-1，而各 28 口監測井座標如表 4-2，根據中央地質調查所於雲林此區域設置之地下水位觀測站的水文地質鑽探資料，顯示此區域自地表至 300 公尺間，共有 4 個含水層與 4 個阻水層，由上而下分別為含水層一、阻水層一、含水層二、阻水層二、含水層三、阻水層三、含水層四與阻水層四，其概念圖如圖 4-2 所示，阻水層透水性差而含水層透水性佳，其中第一阻水層除透水性差且含高濃度砷，因此需要作為分類特性的參考。本研究所使用之監測井資料，其中有 21 口淺水井位於含水層一，其井深介於 8.9 至 22.8 公尺之間，另外 7 口深淺水井之井深，分布於地表下 26 至 110 公尺間，監測井深與用途如表 4-3。

本研究蒐集地下水水質資料時間長度為 1992 至 2005 年間，每站約具 10~50 次調查資料。其中地下水質調查項目包括砷(As)、溫度(Temperature)、酸鹼度(pH)、電導度(Electrical Conductivity, EC)、溶氧(Dissolved Oxygen, DO)、鹼度(Alkalinity, Alk)、總溶解固體(Total Dissolved Solids, TDS)、氯鹽(Cl⁻)、硫酸鹽(SO₄²⁻)、遊離氨氮(NH₃-N)、硝酸鹽氮(NO₃⁻-N)、亞硝酸鹽(NO₂)、生化需氧量(Chemical Oxygen Demand, COD)、總有機碳(Total Organic Carbon, TOC)、生化需氧量(Biochemical Oxygen Demand, BOD)、鎘(Cd)、鉻(Cr)、銅(Cu)、鋅(Zn)、鎳(Ni)、鈉(Na)、鉀(K)、鐵(Fe)、鉛(Pb)、鎂(Mg)、鈣(Ca)、硬度、鈉吸著率，共 28 項水質指標。另外，本研究蒐集雲林縣總區域與沿海區域兩部分地下水水位資料，其中，統計雲林縣總區域內約有 7 個井鄰近沿海地區，因地下水位井與水質監測井位置有一段距離，無法確定

地下水位井與各水質監測井之間的關聯；而沿海區域之 21 個地下水位井雖鄰近各水質監測井，但因資料缺漏過多，有部分測站未有量測資料，因此未作為輸入。

由過去文獻可得其中有部分因子對於了解砷濃度釋出環境為重要的因子，如總有機碳、鐵等，因本研究目的在分析水質因子在聚類中的關係與砷濃度的推估，需考慮用來建立模式的資料筆數是否足夠，因此篩選掉缺漏過多的水質因子，以溫度(Temperature)、酸鹼度(pH)、電導度(EC)、溶氧(DO)、鹼度(Alk)、總溶解固體(TDS)、氯鹽(Cl⁻)、硫酸鹽(SO₄²⁻)、鈉(Na⁺)、鉀(K⁺)、鎂(Mg²⁺)、鈣(Ca²⁺)作為模式的輸入因子。水質因子間的相關性如表 4-4，可以看出後面章節探討之鹽化因子彼此間相關性高，而與砷濃度的相關性普遍都很低，相關性最高的水質因子為鹼度的 0.32。

為瞭解本區域砷污染之嚴重性，及評估本區砷污染程度，根據臺灣飲用水水質標準(環保署，2000)、美國飲用水標準(USEPA, 2006)、世界衛生組織之飲用水標準(NRC, 2001; WHO, 2006)，地下水含砷量濃度(As)標準為 10μg/L 以下及臺灣飲用水水源保護區內管制標準 50μg/L，為符合安全之標準，表 4-1 顯示雲林縣沿海地區 21 口淺水井及 7 口深淺水井，除#7 井水質平均值大於 500μg/L 外，且有超高砷濃度存在，有部分資料甚至超過 1000μg/L，其他各站僅 3 口水井砷濃度平均值符合飲用水標準 10μg/L，大部分水井之砷濃度皆大於飲用水標準，顯示本研究區域為砷污染較嚴重地區，需掌握砷在本區內之變化，藉此管理地下水使用安全。

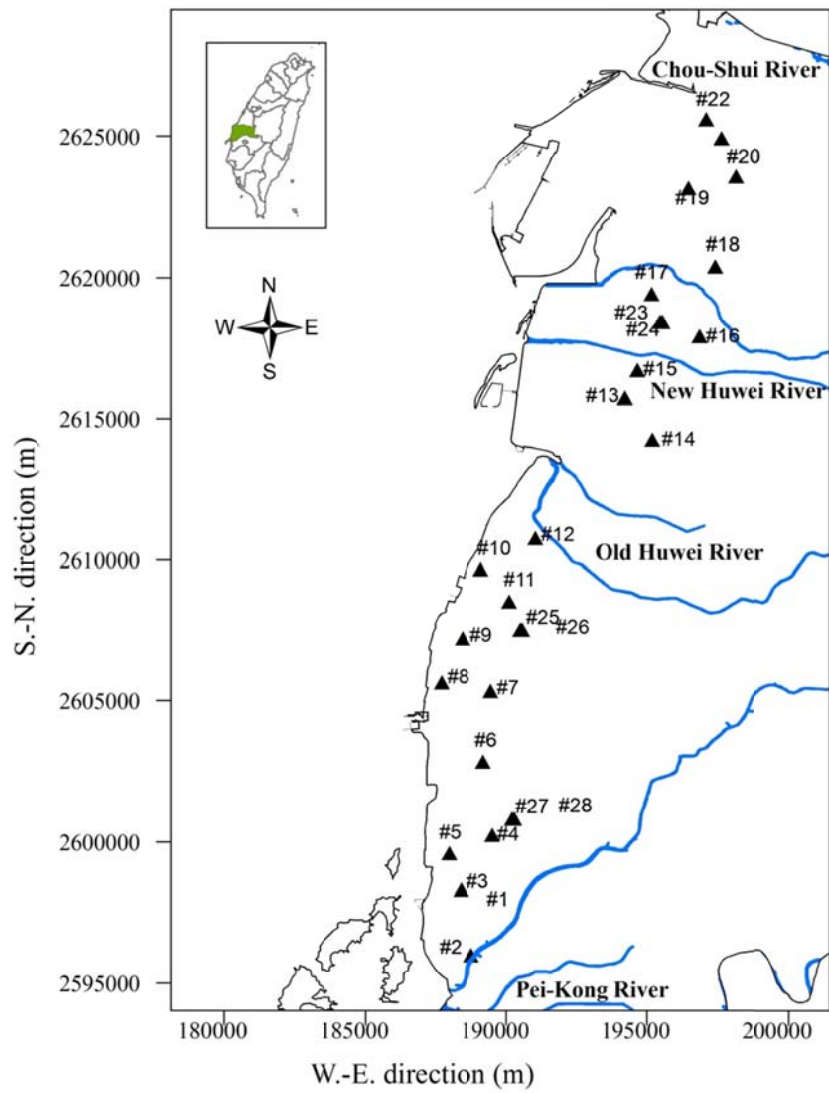


圖 4-1 雲林縣沿海地區監測井分佈圖

表 4-1 1999 年至 2005 年雲林沿海地區觀測井地下水砷濃度資料表

站號	1992-1999 筆數	1999-2005 筆數	砷濃度 ($\mu\text{g/L}$)		站號	1992-1999 筆數	1999-2005 筆數	砷濃度 ($\mu\text{g/L}$)	
			平均	標準 偏差				平均	標準 偏差
#1	28	0	28.3	41.1	#15	19	0	111.0	171.1
#2	28	0	21.5	15.9	#16	28	0	6.7	8.9
#3	28	22	110.0	67.4	#17	27	18	66.0	49.7
#4	28	0	77.6	47.6	#18	28	0	19.5	14.8
#5	24	0	9.5	10.5	#19	28	22	32.4	26.2
#6	27	22	162.0	124.0	#20	28	0	20.0	12.6
#7	28	22	594.0	349.0	#21	28	0	7.5	3.8
#8	28	0	11.1	6.7	#22	24	0	226.0	79.7
#9	28	0	478.0	208.0	#23	24	0	10.5	2.9
#10	26	0	93.0	101.0	#24	24	0	19.1	7.9
#11	28	0	83.3	26.8	#25	23	0	122.0	24.1
#12	28	22	49.5	32.4	#26	24	0	38.0	14.4
#13	12	21	55.6	90.9	#27	24	0	22.4	24.5
#14	28	0	31.4	16.9	#28	24	0	68.5	65.4

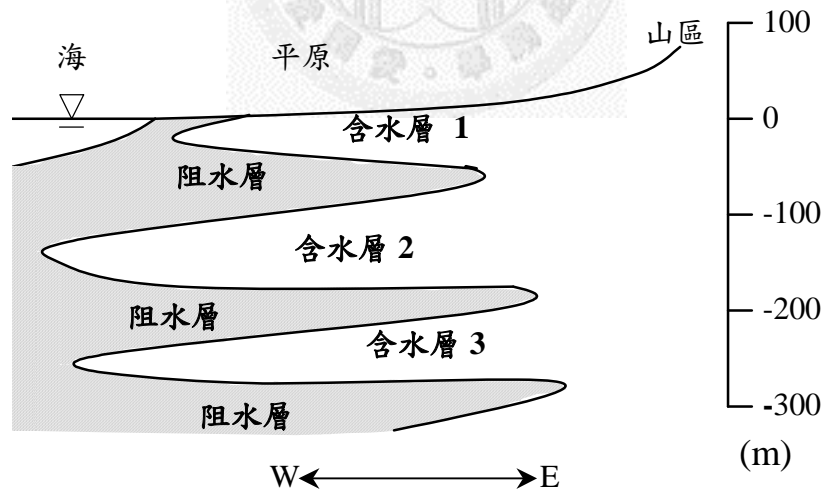


圖 4-2 雲林地區水文地質概念圖

(劉振宇等人，2006)

表 4-2 雲林沿海地區 28 座監測井座標資料表

(高力山, 2011)

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

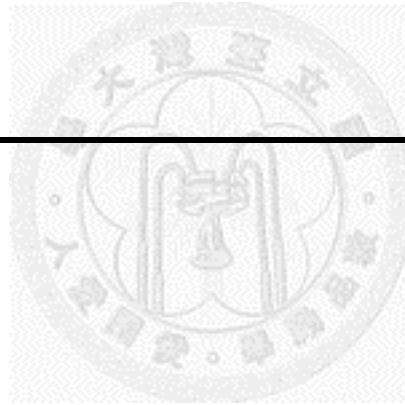
表 4-3 雲林沿海地區 21 口監測井及水利處地下水井基本資料

(成功大學水工試驗所，1998)

井號	地理位置	井管深度(公尺)	影響地下水變動因素及可能抽水原因
#1	口湖鄉成龍	13.8	氣候
#2	口湖鄉成龍	15.2	氣候、潮汐
#3	口湖鄉臺子	22.8	氣候、養殖抽水
#4	口湖鄉厝子	19.1	氣候、建築用抽水
#5	口湖鄉金湖	13.0	氣候、潮汐
#6	口湖鄉下崙	17.0	氣候、潮汐
#7	口湖鄉下崙	19.0	氣候、農田抽水
#8	口湖鄉箔子寮	19.9	氣候、水渠水位
#9	四湖鄉廣溝厝	19.2	氣候
#10	四湖鄉林厝寮	13.0	氣候、水渠水位
#11	四湖鄉三條崙	24.4	氣候
#12	四湖鄉林厝寮	19.6	氣候
#13	台西鄉五條港	13.0	氣候
#14	台西鄉瓦厝	8.9	氣候、畜牧抽水
#15	台西鄉	12.0	氣候、養殖抽水
#16	台西鄉三姓寮	8.9	氣候、農田抽水
#17	台西鄉蚊港	8.4	氣候、養殖抽水
#18	麥寮鄉	15.2	氣候、畜牧抽水
#19	麥寮鄉後安寮	14.9	氣候、水渠水位
#20	麥寮鄉中山	9.4	氣候、養殖抽水
#21	麥寮鄉許厝寮	12.4	氣候、養殖抽水
#22	水利處麥寮鄉許厝寮井	36.0	氣候、養殖抽水
#23	水利處台西鄉新興淺井	36.0	氣候
#24	水利處台西鄉新興深井	97.0	氣候
#25	水利處四湖鄉飛砂淺井	35.0	氣候、農田抽水
#26	水利處四湖鄉飛砂深井	110.0	氣候
#27	水利處口湖鄉口湖淺井	36.0	氣候、農田抽水
#28	水利處口湖鄉口湖深井	78.0	氣候

表 4-4 雲林沿海地區地下水水質相關性

	As	Temp	pH	EC	DO	Alk	TDS	Cl ⁻	SO ₄ ²⁻	Na ⁺	K ⁺	Mg ²⁺	Ca ²⁺
As	1.00	-0.02	0.25	-0.15	0.03	0.32	-0.15	-0.16	-0.17	-0.14	-0.12	-0.18	-0.27
Temp		1.00	0.00	-0.05	0.02	0.08	-0.05	-0.05	-0.04	-0.08	-0.05	-0.03	-0.06
pH			1.00	-0.40	0.21	0.23	-0.36	-0.39	-0.43	-0.35	-0.32	-0.44	-0.59
EC				1.00	-0.13	-0.09	0.95	0.97	0.88	0.96	0.88	0.90	0.76
DO					1.00	-0.05	-0.11	-0.11	-0.07	-0.12	-0.06	-0.15	-0.15
Alk						1.00	-0.06	-0.09	-0.12	-0.06	-0.01	-0.10	-0.18
TDS							1.00	0.95	0.85	0.95	0.87	0.86	0.72
Cl ⁻								1.00	0.88	0.98	0.91	0.91	0.77
SO ₄ ²⁻									1.00	0.87	0.76	0.84	0.80
Na ⁺										1.00	0.90	0.89	0.75
K ⁺											1.00	0.80	0.66
Mg ²⁺												1.00	0.81
Ca ²⁺													1.00



4.2.2 資料處理

本研究選用全區水井中砷與其他 13 項較資料筆數較完整水質因子，包括溫度、pH、電導度、溶氧、鹼度、TDS、氯鹽、硫酸鹽、鈉、鉀、鎂、鈣及硬度為本研究主要探討對象。考慮到 SOM 類神經網路以特徵映射方式，可以降低高維度系統的複雜性，因此，本研究採取不在輸入模式之前作篩選因子的動作，以上述所有因子作為模式之輸入項，並於輸入時對所有因子之原始輸入值加以規格化(Scaling)或稱為正規化(Normalizing)，將輸入資料之範圍正規化至 0 至 1 之間，因為計算優勝神經元是根據歐氏距離公式計算，若未正規化則會受到各個變數的範圍、濃度值的大小與變異程度不同，造成計算歐氏距離時會對訓練過程造成不同程度之影響(Kohonen, 1995)。本文採用的正規化方法如下：

$$N_{normal} = \frac{N_i - N_{min}}{|N_{max} - N_{min}|} \quad (4-1)$$

N 為水質因子之濃度值，下標 max 與 min 分別表示各種水質因子本身的最大值與最小值，而所謂「反正規化」處理，即是將類神經網路的預測結果利用上式反求得其對應的原始值。

4.3 模式架構

本研究將 28 個測站透過聚類方式依測站的特性分區，找出具有相似特性的測站，由自組特徵映射網路(SOM)產生拓撲圖以探討不同聚類之特性，並期望能有助於類神經網路推估的精準度，進而將聚類的結果作為輸入輻狀基底函數類神經網路(RBFNN)的資訊，建構區域性砷

濃度推估模式，最後以克利金(Kriging)方法作砷濃度空間分布之呈現，其模式架構之流程圖如圖 4-6 所示。

聚類的概念乃是將輸入的資料加以分成不同的類別，將同質性高的資料聚集為一類，進而分離出同質性低的資料，換句話說同一群同一聚類內的資料特性相似度高，而不同聚類間的資料特性差異較大。本研究將 28 個測站，將包含砷之 13 個水質因子濃度作為分類所輸入的變數。本研究考慮在建立好砷濃度推估模式後，未將砷濃度作為推估模式之輸入，單輸入其他水質因子至模式中推估砷濃度，故初步將蒐集的資料分成訓練資料(1992-1997 年與 2000-2005 年)、測試資料(1998-1999 年)如表 4-5，其中 2000 年後因研究區域內測站因經費縮減，僅剩下 6 站繼續監測水質，故這部份不適合代表全區域測站作為測試資料。接著將訓練部份包含砷之 13 個水質因子濃度輸入自組特徵映射網路(SOM)，進一步找出具有相同特性的測站，探討各測站間空間分布的關聯性及各水質因子與砷濃度的關係。

表 4-5 訓練與測試資料

	年份	筆數
Training	1992-1997 年、2000-2005 年	共 654 筆
Testing	1998-1999 年	共 179 筆

本模式因研究區域內水質監測值變異性大，各監測站之水文、地質結構造成特性不同。若由人工選取訓練及驗證資料，容易因選取資

料差異造成訓練及驗證結果產生人為誤差，為解決此問題以達到良好訓練，採用 K 疊交叉驗證法(K-fold cross-validation)進行訓練與驗證。K-fold cross-validation 主要將資料分為 K 個子集合，每次選取一個子集合作為驗證資料，剩餘的 K-1 個子集合作為訓練資料；本研究選用 K=6，即將資料分為 6 個子集合，每個子集合中的資料以約兩年為單位作為一個子集合，並對此網路架構訓練 6 次，使每個子集合都輪值過驗證資料，直到訓練完成之後，再將訓練與驗證資料分別計算平均誤差；另外，將獨立出來的測試資料輸入至模式中推估砷濃度，並計算誤差作為優選模式之參考。本模式的網路訓練與測試流程如圖 4-7 所示。

由於 SOM 拓樸層的分類個數較難決定，本研究利用訓練、驗證資料之評估指標來決定拓樸層大小、類神經網路的初始輸入參數。圖 4-3 為不同拓樸層大小之模式表現，類神經網路訓練結果表示網路的學習情況，而驗證結果則決定網路架構，並搭配交叉驗證方法嘗試多次不同初始輸入參數時，所產生之結果作為優選模式之用。依驗證資料的評估指標所示，以 4×4 具有最小的誤差，而往後誤差逐漸增加，有過度描述的情形，因此選用 4×4 拓樸架構作為 SOM 分類個數的標準。

本研究提出與以往自組特徵映射網路(SOM)結合網路輸出層推估模式相異的方法，在 SOM 聚類過程加入砷濃度資訊於輸入資料，作為其中一聚類的資訊，以期望有助於聚類的結果，如此也使得輸出拓樸圖，包含砷之中心點資料可以與其他輸入之水質因子可直接比較對應關係。本研究針對不同的砷輸入情況：無加入砷濃度與加入砷濃度於 SOM 聚類作探討。依照聚類內的砷濃度，參考國際及臺灣飲用水水質標準與臺灣飲用水水源水質標準分別設定 10($\mu\text{g/L}$)、50($\mu\text{g/L}$)與極大砷

濃度 1000($\mu\text{g/L}$)來探討聚類情況，圖 4-4 為僅輸入 12 個水質因子而未加入砷濃度進入 SOM 進行聚類；圖 4-5 為 13 個水質因子，包含砷濃度皆加入 SOM 進行聚類，每一個圓餅圖相當於一個聚類，對應圖 5-1、圖 5-3 與聚類的位置，每一個圓餅圖表示不同砷濃度區間在此聚類所佔的比例，由此可得知不同區間的砷濃度分布在各聚類的情形。

結果顯示未加入砷濃度於 SOM 進行聚類，無法將高濃度與低濃度砷分開，區隔出來做單一聚類特性之探討，如圖 4-4 超過 1000($\mu\text{g/L}$)的極高砷濃度落於聚類 A-2，而中高濃度受到嚴重污染的 50-1000($\mu\text{g/L}$)分布在上方的聚類，而低於 50($\mu\text{g/L}$)分布於下方聚類，每個聚類皆有中、低濃度分布，未能分離。而圖 4-5 為有加入砷於 SOM 輸入進行聚類，結果顯示 1000($\mu\text{g/L}$)極高濃度分布在聚類 A-4，而砷濃度超過 50($\mu\text{g/L}$)的聚類落於最上方與右方，擁有低濃度的小於 50($\mu\text{g/L}$)聚類則落於其他聚類。其高濃度 1000($\mu\text{g/L}$)、中濃度 50-1000($\mu\text{g/L}$)皆有顯著分離，如此可以對於不同區間的砷濃度作不同高低濃度的特性探討。另外，對於地下水砷濃度推估亦有精度上的改善，在第五章詳細探討，對照表 5-3 推估結果表。

當聚類中心層的神經元完成 SOM 演算法過程後，每個 SOM 拓樸層的神經元代表一聚類，而每一聚類都將有一個固定的聚類中心點。本研究從前階段拓樸層之聚類中心點移除砷濃度資訊，作為下一階段架構 RBFNN 中輻狀基底函數的中心點，以 RBFNN 監督式學習法，輸入砷以外水質因子至模式中，透過所建構的輻狀基底函數，以函數逼近方法找出輸入—輸出間的關係，即計算輸入向量與隱藏層各神經元中心點距離後，經由函數轉換獲得隱藏層各神經元的輸出，再加權傳

至輸出層，得到網路輸出之砷濃度推估值。

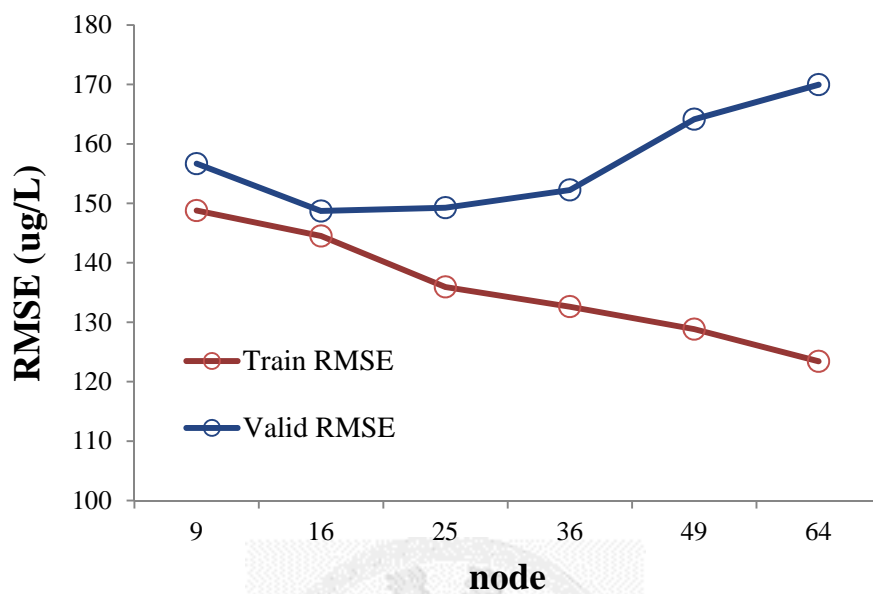


圖 4-3 SOM 不同拓樸層之模式表現

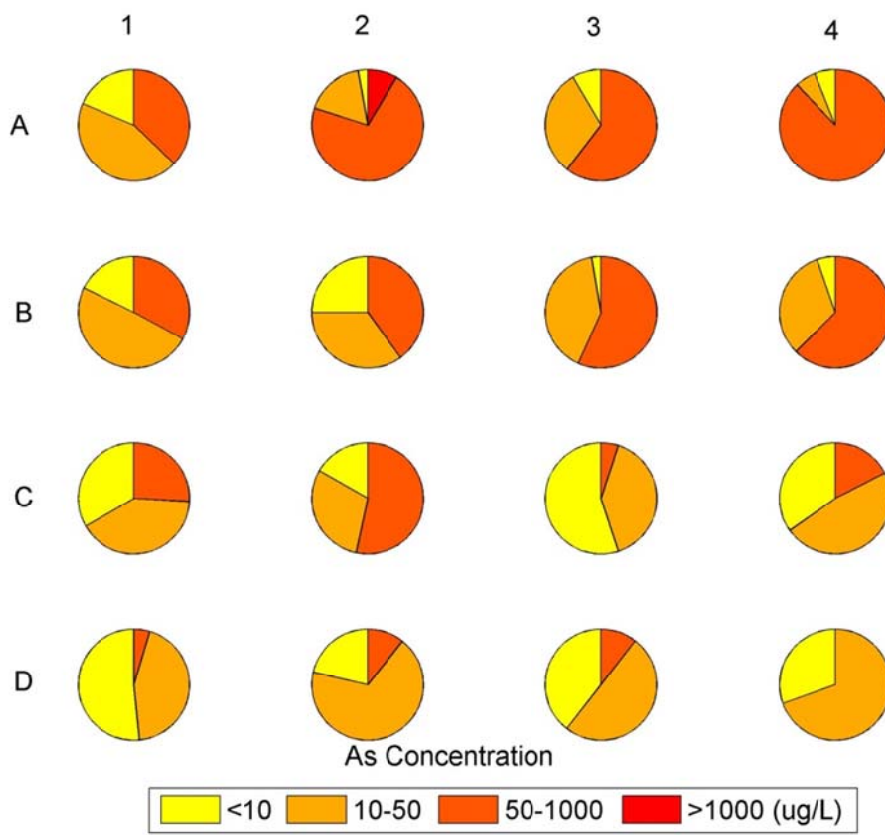


圖 4-4 SOM 聚類砷濃度分布圖(無加入砷)

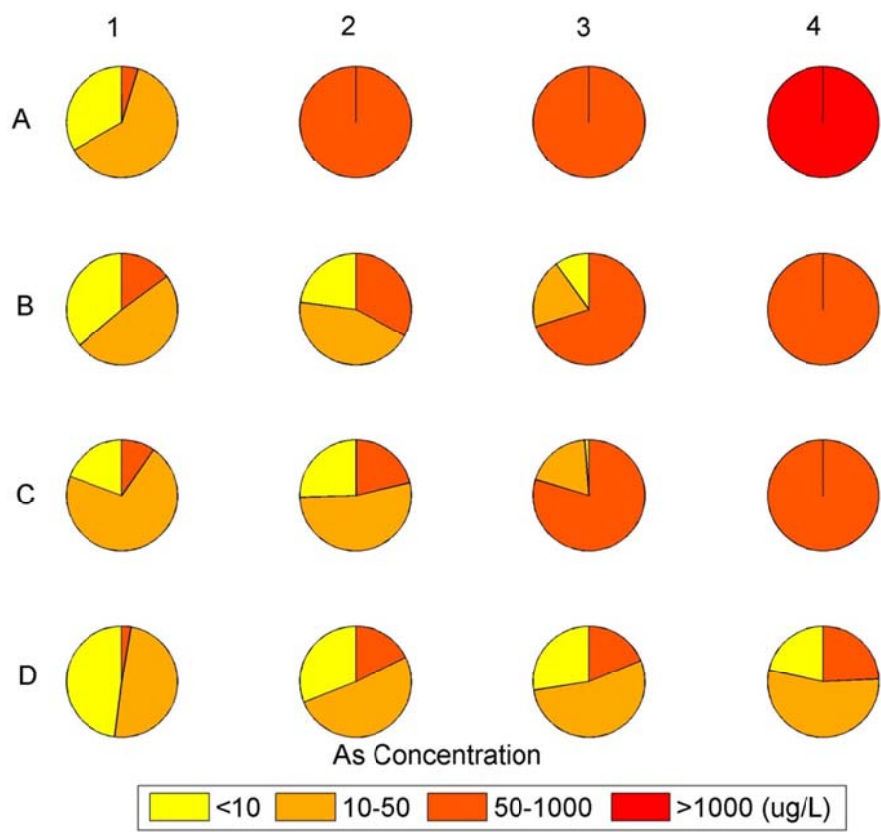


圖 4-5 SOM 聚類砷濃度分布圖(加入砷)

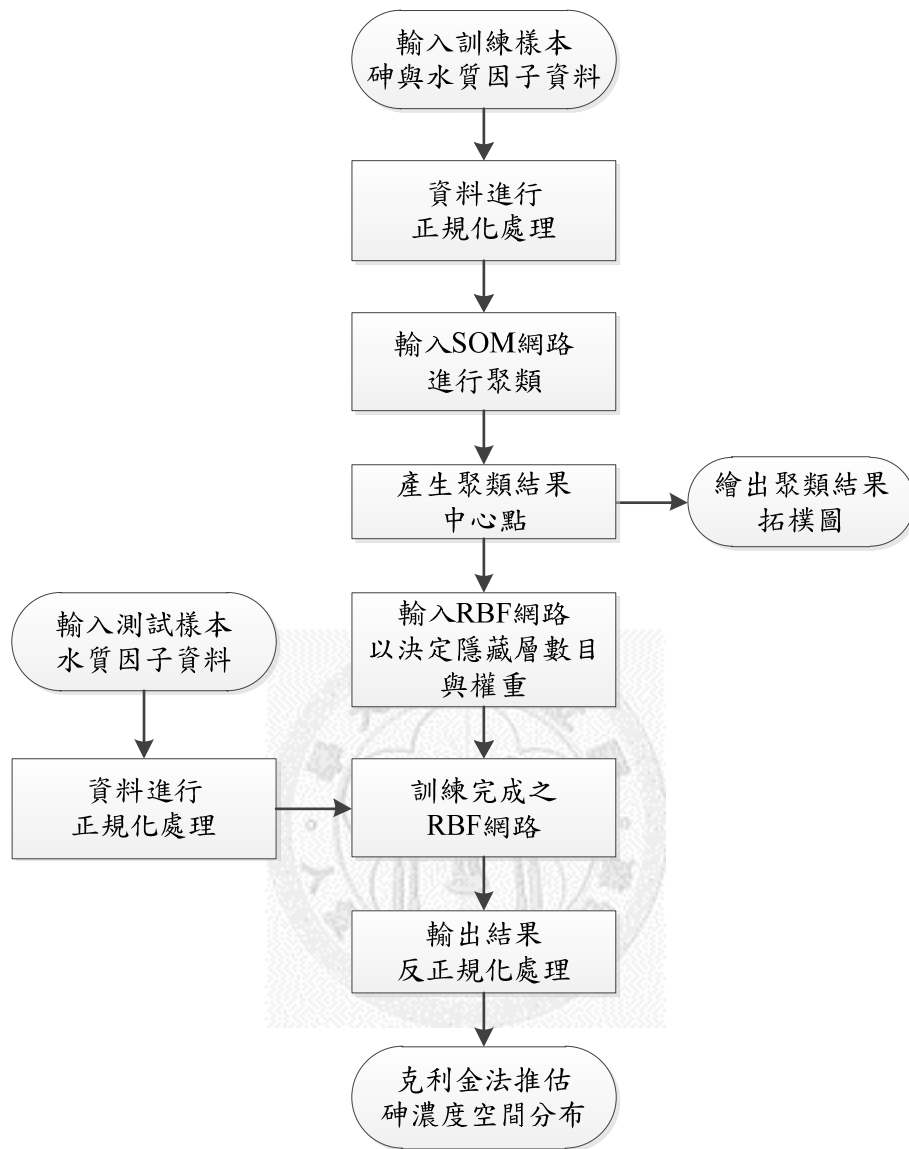


圖 4-6 地下水砷濃度推估模式流程圖

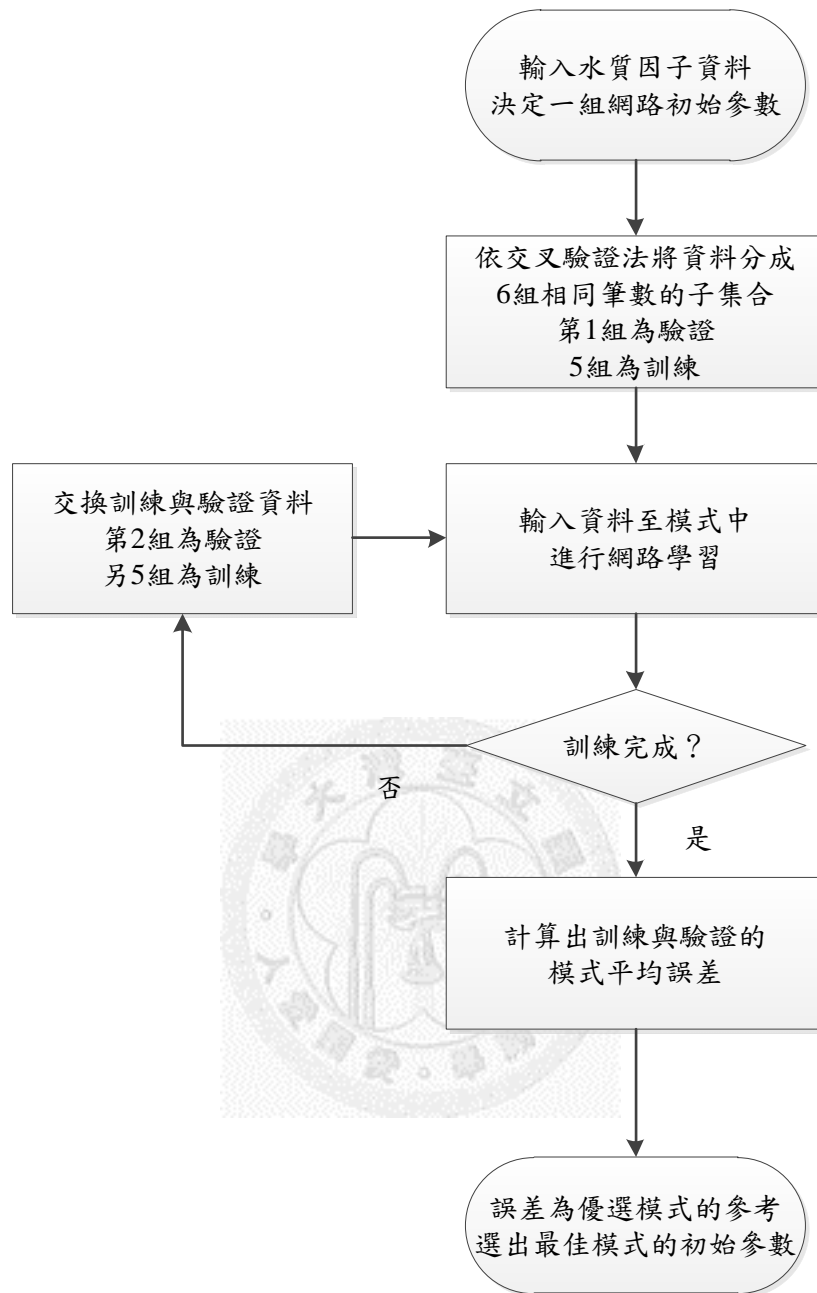


圖 4-7 交叉驗證流程圖

第五章 結果與討論

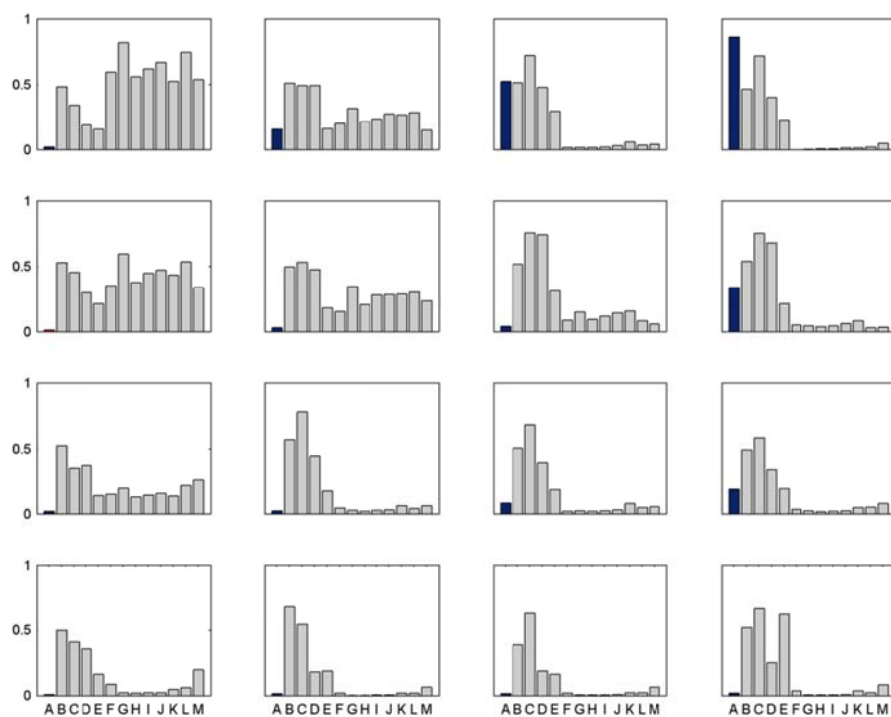
本章節為研究的呈現，主要分成兩部分，第一部分首先解釋拓樸圖的架構與如何對應不同水質因子的圖，並探討每個聚類中所表現的地下水水質特徵，與砷濃度的關係；第二部分為建立地下水砷濃度推估模式之誤差結果與呈現區域內的砷濃度空間推估圖，並比較本研究所應用的模式(SOM 與 RBFNN)與倒傳遞類神經網路(BPNN)的結果，分述如下：

5.1 SOM 拓樸架構

本研究應用自組特徵映射網路(SOM)具備的分類特性如圖 5-2 的左半部分，並依模式所得最後的權重作為聚類中心點以代表該聚類內資料之特性，以圖 5-1 表示，圖中每一方格代表一個聚類，聚類中的長條圖代表著輸入的資料落於該聚類的中心點大小，所呈現的是正規化後的數值，橫坐標由左到右分別為輸入因子：砷(As)、溫度(Temperature)、酸鹼度(pH)、電導度(EC)、溶氧量(DO)、鹼度(Alk)、總溶解固體(TDS)、氯鹽(Cl⁻)、硫酸鹽(SO₄²⁻)、鈉(Na⁺)、鉀(K⁺)、鎂(Mg²⁺)與鈣(Ca²⁺)。

自組特徵映射網路(SOM)將 28 個水質監測井分為 16 類之後，由圖 5-3 可以得知每一個聚類內所代表的資料筆數，由每個聚類內的資料可以找到相對應的水質監測井如圖 5-4 所示，其在拓樸圖內的聚類位置可以對照著看，例如圖 5-3 中最左上角我們稱作聚類 A-1，可以得知此聚類中有 42 筆資料，對應圖 5-4 內可知這 42 筆資料是來自水質監

測井#2、#3、#10、#12 與#13；圖 5-3 最右下角稱作聚類 D-4，其中有 37 筆資料，均來自水質監測井#14、#27 與#28。因圖 5-4 內必須呈現監測井資料落於聚類中的情況，而分類的結果顯示每一監測井均有若干筆資料且依據水質因子的資訊無法完全分離至其中一個聚類，因此為簡易統計測站分類後所對應的聚類，本研究訂定以超過每聚類總資料筆數之 10% 作為門檻值，超越此值的測站則歸屬於此聚類。



橫軸：水質因子(As, Temp., pH, EC, DO, Alk., TDS, Cl, SO₄, Na⁺, K⁺, Mg²⁺, Ca²⁺)

縱軸：變數中心點大小(正規化)

圖 5-1 SOM 聚類中心點

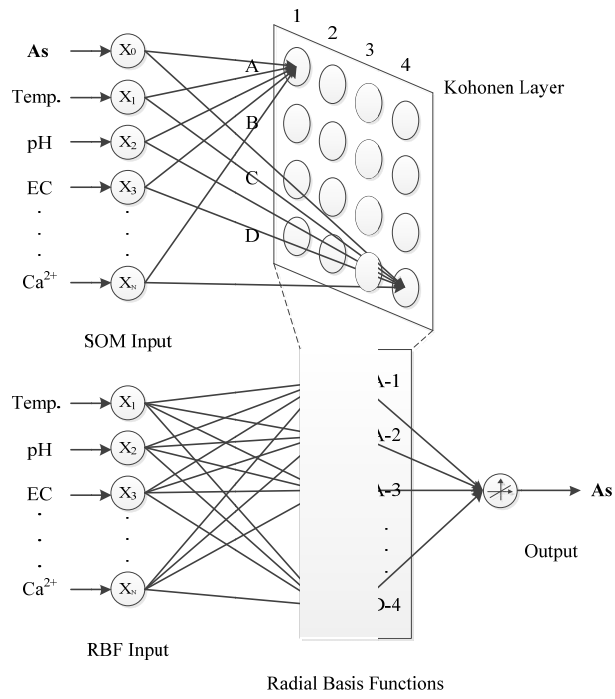


圖 5-2 模式輸入與輸出架構圖

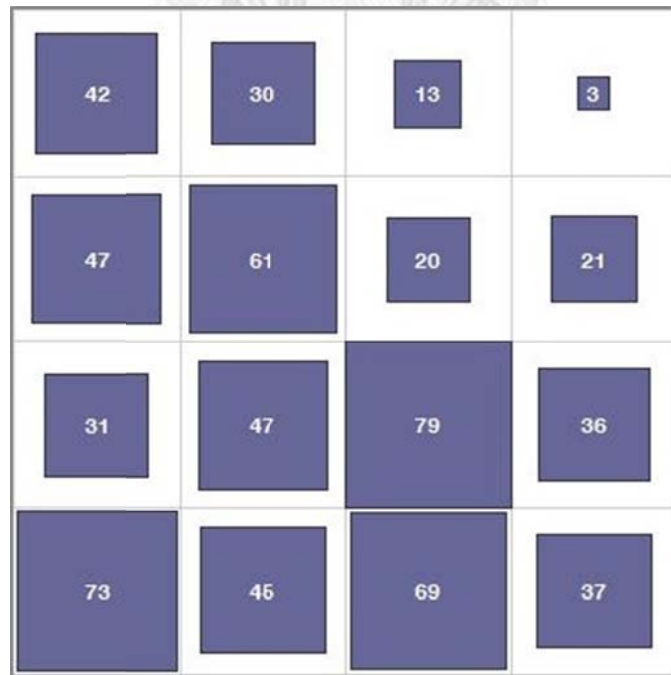


圖 5-3 SOM 聚類資料筆數分布圖

	1	2	3	4
A	#2, #3, #10, #12, #13	#3, #6, #10	#7	#7
B	#2, #5, #8, #12	#3, #5, #6, #8, #12, #19	#6, #12	#7, #9, #15
C	#13, #15, #19	#1, #4	#4, #11, #14, #25	#7, #22
D	#1, #16, #17, #18, #20	#17, #20, #21, #26, #28	#17, #21, #23, #24	#14, #27, #28

圖 5-4 SOM 聚類測站分布圖



5.1.1 聚類內測站之位置分布

圖 5-4 內聚類所代表的顏色對應圖 5-5 左方標記的顏色，聚類的編號順序依據圖 5-4 拓樸圖由上到下進行 A 至 D 四列的聚類編號，顏色為四個區域。由圖 5-4 的聚類顏色可以對應到圖 5-5 的測站分布位置，顏色表示聚類，地圖上的數字表示井號，因地圖上每個測站位置只能表示一種顏色，則以每個測站落於最多筆數的聚類為主，作為代表測站特性的聚類。

依照圖 5-5 可知，在拓樸圖上層的聚類，顯示為 Type I 與 II 的測站較多靠近研究區域西南沿海地區；下層聚類，顯示為 Type III 與 IV 的測站則較靠近北邊，濁水溪沖積扇尾南側、離海岸較遠之測站。另外將結果對照表 4-3 顯示區域內大部分監測井井深約在 20 公尺上下，而水利署之深淺井均被分到聚類下方 D-2~D-4 聚類中，最大井深度約在 78~110 公尺之間。

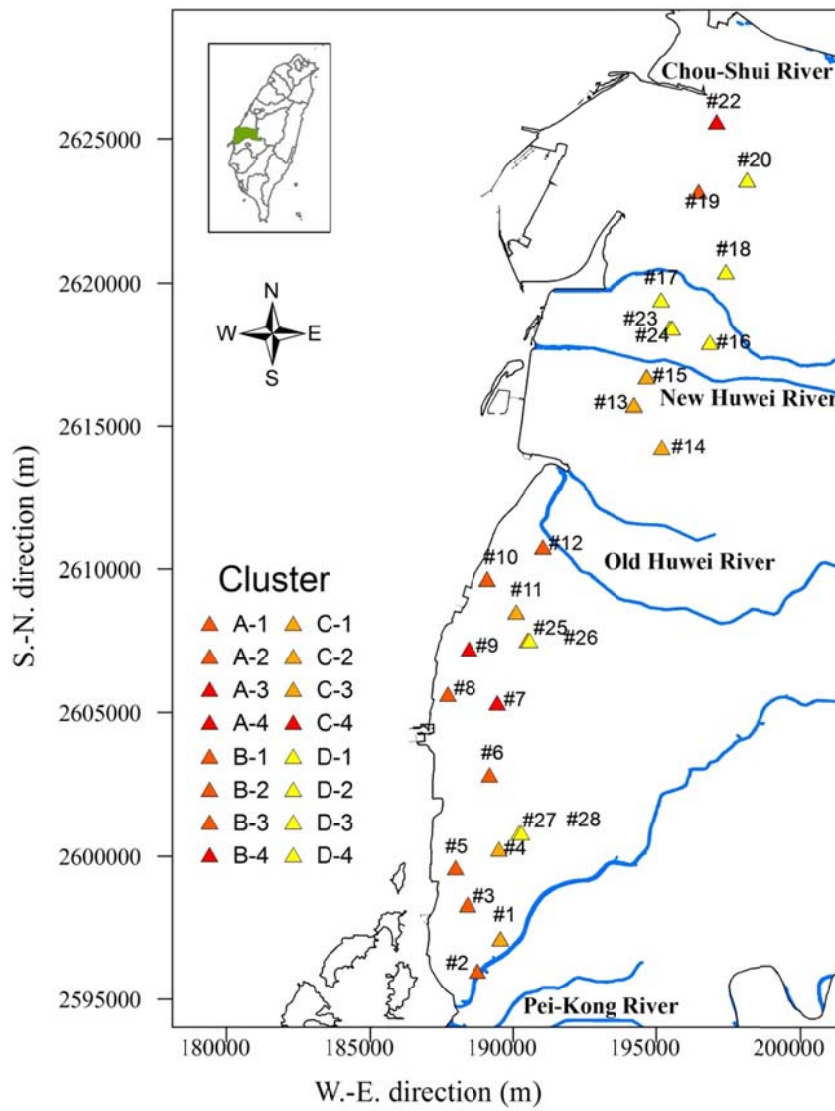


圖 5-5 SOM 聚類對應測站位置分布圖

Type I : A-3, A-4, B-4, C-4

Type II : A-1, A-2, B-1, B-2, B-3

Type III : C-1, C-2, C-3

Type IV : D-1, D-2, D-3, D-4

5.1.2 聚類內水質因子探討

由於聚類間的拓樸結構彼此間有連續的關係存在，圖 5-1 不易表現各個變數在 16 個聚類內變化的趨勢，因此繪出圖 5-12 至圖 5-11 探討水質因子間，在不同聚類的特性與現象，並佐以文獻以了解聚類內測站的特性。圖中 X - Y 軸對應二維聚類別，聚類的編號依據圖 5-1 先往右再往下的方向進行 A-1 至 D-4 聚類編號，Z 軸為中心點大小。前面提到深淺井落於聚類 D-2 至 D-4 中，在圖 5-5 測站位置偏北部，對照圖 5-14，氯鹽係指水中的氯離子，天然淡水中的含量不多，氯鹽濃度低($<100\text{mg/L}$)一般視為淡水(陳文福等人，2010)，可能原因為井管深度很深且受到海水鹽化之影響較輕微，且除了氯鹽以外的部分因子也與上方聚類有不一樣的特性，如圖 5-12 砷、圖 5-15 硫酸鹽與圖 5-6 至圖 5-11 所代表的電導度、TDS、鈉、鉀、鎂與鈣在聚類 D-2 至 D-4 中，未如上方聚類呈現正相關或負相關。

鹽化因子

西南沿海一帶分布拓樸圖左上方的聚類，除了聚類 A-4 之外，其氯鹽如圖 5-14，濃度約在 $100\sim 19000\text{mg/L}$ 間，一般視為淡海水(陳文福等人，2010)，且電導度與水中總離子相關，水中溶解離子之濃度與電導度大致呈正比，見圖 5-6；氯離子、硫酸根離子、鈉、鉀、鎂與鈣離子均為海水之主要成分，其中氯離子與電導度常作為海水鹽化之指標，與其他因子趨勢相似，見圖 5-6 至圖 5-11，顯示此區域海水鹽化情形較嚴重，此結果與 Liu et al. (2003)對濁水溪沖積扇所進行因子分析結果相仿。

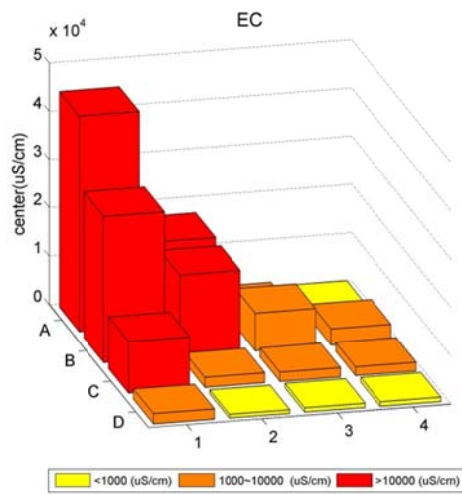


圖 5-6 SOM 聚類電導度中心點

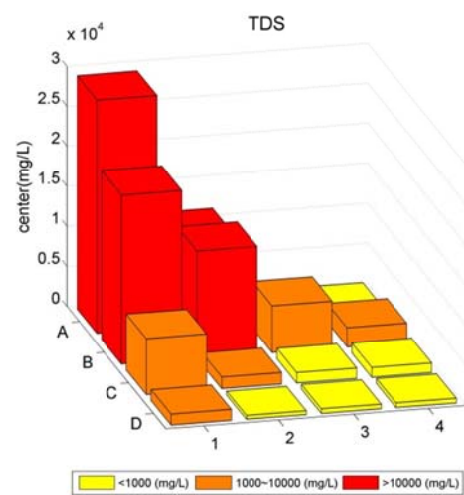


圖 5-7 SOM 聚類總溶解固體濃度中心點

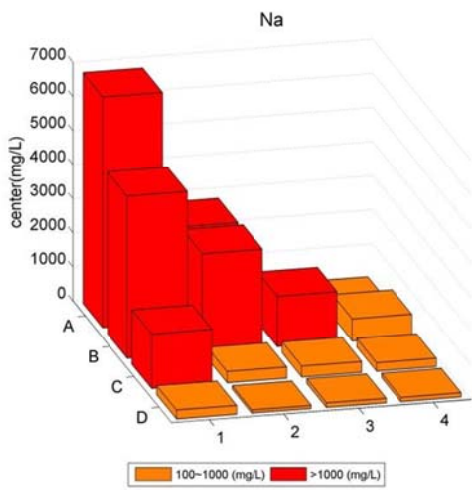


圖 5-8 SOM 聚類鈉濃度中心點

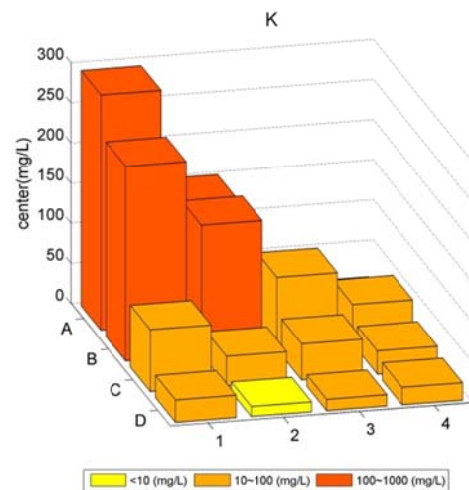


圖 5-9 SOM 聚類鉀濃度中心點

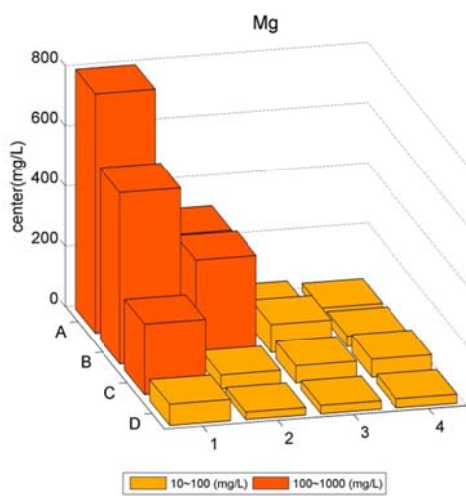


圖 5-10 SOM 聚類鎂濃度中心點

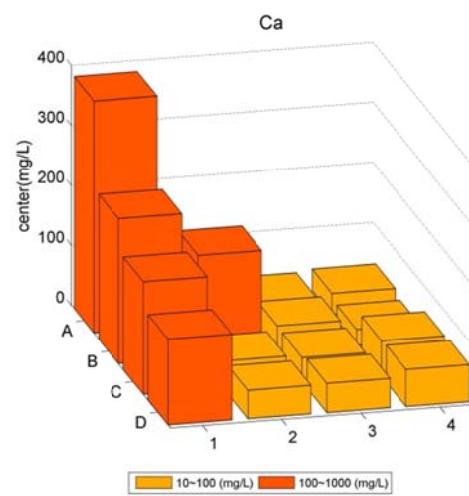


圖 5-11 SOM 聚類鈣濃度中心點

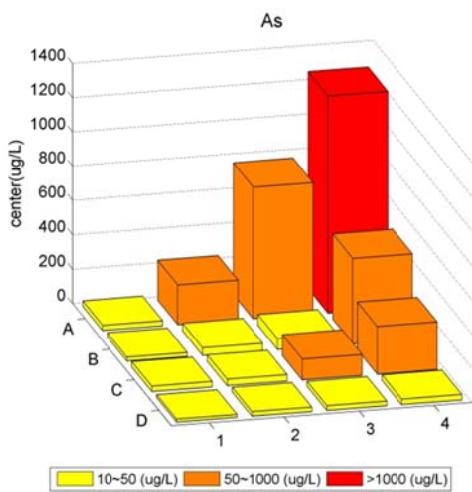


圖 5-12 SOM 聚類砷濃度中心點

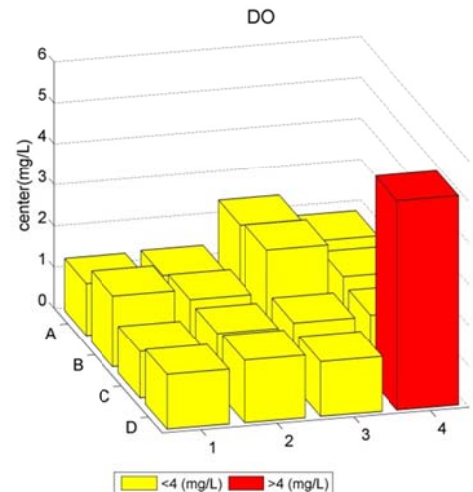


圖 5-13 SOM 聚類溶氧量中心點

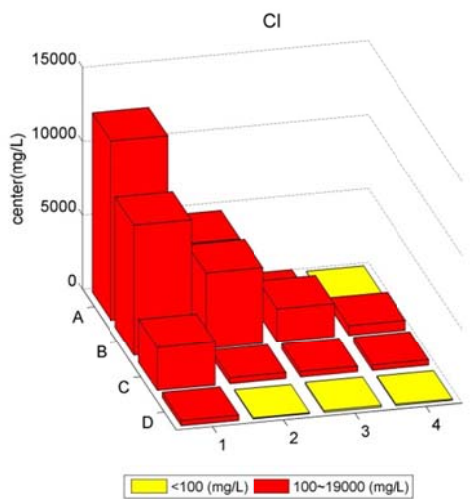


圖 5-14 SOM 聚類氯鹽濃度中心點

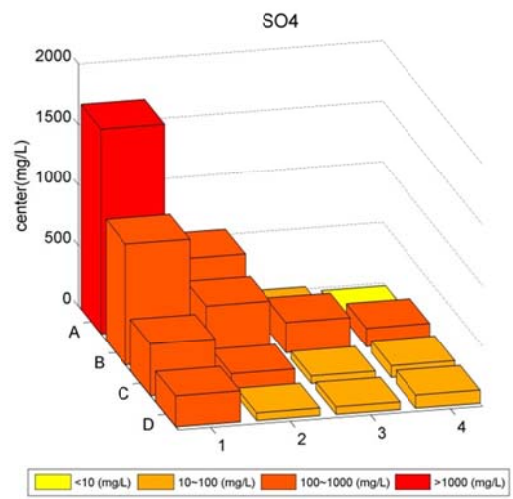


圖 5-15 SOM 聚類硫酸鹽濃度中心點

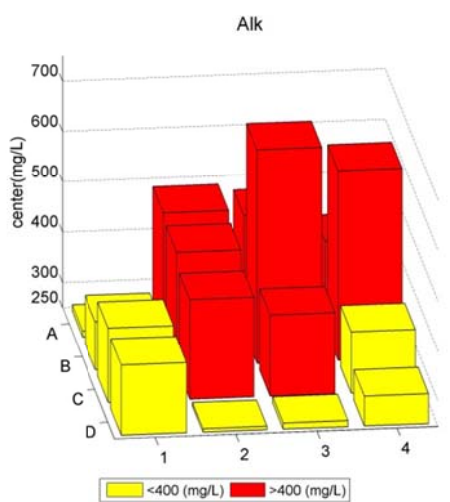


圖 5-16 SOM 聚類鹼度中心點

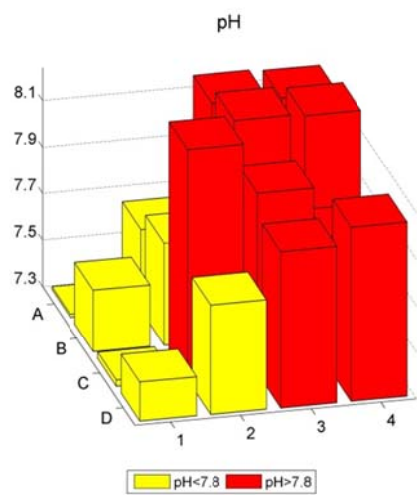


圖 5-17 SOM 聚類 pH 值中心點

表 5-1 SOM 聚類中心點資料表 1

		A-1	A-2	A-3	A-4	B-1	B-2	B-3	B-4
As	($\mu\text{g/L}$)	30.6	232.4	766.1	1265.0	22.6	44.4	62.0	496.7
Temp.	($^{\circ}\text{C}$)	25.7	26.0	26.0	25.5	26.1	25.8	26.0	26.2
pH		7.3	7.7	8.2	8.2	7.6	7.8	8.3	8.3
EC	(uS/cm)	44584.3	17291.3	1734.2	873.3	30143.8	16863.2	7800.0	3331.2
DO	(mg/L)	1.3	1.3	2.4	1.8	1.7	1.5	2.5	1.7
Alk.	(mg/L)	261.9	489.9	477.8	421.7	345.5	476.9	674.8	627.0
TDS	(mg/L)	29155.3	11253.8	1216.7	566.7	21083.3	13376.6	5824.5	2357.5
Cl ⁻	(mg/L)	12123.3	4688.8	257.9	48.0	8787.1	5123.5	2238.5	657.7
SO ₄ ²⁻	(mg/L)	1693.3	569.9	47.3	4.3	998.3	436.6	248.7	150.1
Na ⁺	(mg/L)	6769.1	2780.9	351.3	141.3	4756.7	2895.9	1479.8	647.9
K ⁺	(mg/L)	293.2	150.6	33.5	8.7	242.5	161.9	89.9	48.6
Mg ²⁺	(mg/L)	795.7	307.8	38.8	27.0	569.9	326.0	92.7	35.0
Ca ²⁺	(mg/L)	386.8	108.0	31.4	35.7	241.6	170.4	42.5	26.3

表 5-2 SOM 聚類中心點資料表 2

		C-1	C-2	C-3	C-4	D-1	D-2	D-3	D-4
As	($\mu\text{g/L}$)	32.0	38.1	122.7	278.8	17.7	26.8	28.1	32.5
Temp.	($^{\circ}\text{C}$)	26.1	26.6	25.9	25.8	25.8	27.7	24.7	26.1
pH		7.3	8.3	8.1	7.9	7.5	7.8	8.0	8.1
EC	(uS/cm)	10701.3	2081.6	1916.8	1828.0	2148.5	752.0	908.6	959.2
DO	(mg/L)	1.1	1.4	1.5	1.6	1.3	1.5	1.3	5.1
Alk.	(mg/L)	397.6	448.4	411.5	372.1	389.4	256.8	262.1	309.4
TDS	(mg/L)	6999.4	1474.9	1308.4	1230.0	1381.9	478.1	584.8	603.2
Cl ⁻	(mg/L)	2917.0	401.0	365.7	358.0	347.4	62.9	114.4	90.2
SO ₄ ²⁻	(mg/L)	428.0	131.5	63.1	96.6	254.0	63.1	62.4	109.0
Na ⁺	(mg/L)	1599.9	328.6	321.1	246.9	260.7	94.3	133.7	125.0
K ⁺	(mg/L)	77.2	36.9	45.2	29.0	27.9	12.4	14.1	21.8
Mg ²⁺	(mg/L)	235.5	46.0	57.2	59.6	70.4	24.5	27.5	29.9
Ca ²⁺	(mg/L)	187.7	45.1	40.9	58.3	142.7	47.4	48.7	61.7

電導度

本區域的電導度的差異很大，由圖 5-6 的中心點濃度從最小聚類 D-2 的 $752 \mu\text{S}/\text{cm}$ 到最大聚類 A-1 的 $44584 \mu\text{S}/\text{cm}$ 。依據環保署飲用水水質標準規定總溶解固體的限值為 $250\text{mg}/\text{L}$ ，電導度約在 $350 \mu\text{S}/\text{cm}$ ，全區水樣都超過飲用水水質標準，表示此地區大部分的水質不適合直接飲用。對照圖 5-5 測站位置分布，北部區域的監測井位於拓樸圖下方聚類，電導度較小；往西南區域越往海邊電導度漸漸升高，沿海地下水的電導度最大值甚至高達 $50000 \mu\text{S}/\text{cm}$ ，大約海水電導度的平均值 ($47800 \mu\text{S}/\text{cm}$)。在電導度較小的地下水區域，部分可能位於淺層含水層受到地表水補注，因而沖淡離子濃度；或因為位於深層的關係，而未受到海水的影響。

溶氧量

圖 5-13 為溶氧的聚類中心點濃度，溶氧量指溶解於水中的氧分子，聚類 D-4 的溶氧濃度達到 $5.1\text{mg}/\text{L}$ ，超過欲維持魚類良好之棲息環境，水中溶氧量至少需達到 $5\text{mg}/\text{L}$ 以上的指標。而其他聚類溶氧都在 $1\text{mg}/\text{L}$ 左右，顯示地下水氧氣消耗呈缺氧狀態。Anawar et al. (2003)指出孟加拉處在缺氧的還原環境，其地下水溶氧都小於 $2.1\text{mg}/\text{L}$ ，且大部分水質樣本低於監測極限，與本區域的環境相似。地下水中的溶氧主要受到地面補注影響，地面水的溶氧介於 $0\sim 8\text{mg}/\text{L}$ ，受到溫度與生物影響而有大小值之分，但地下水流動緩慢，受到有機物與其他化學物質反應，而消耗氧氣使地下水溶氧值降低至缺氧狀態。

砷濃度

圖 5-12 為砷濃度的中心點，根據臺灣飲用水水質標準(環保署，

2000)、美國飲用水質標準(USEPA, 2006)與世界衛生組織的飲用水水質標準(NRC, 2001; WHO, 2006),其砷濃度的標準皆為 $10\mu\text{g/L}$;而臺灣飲用水源的水質標準則為砷濃度 $50\mu\text{g/L}$ (環保署, 1997)。用以上標準來評估此研究區域地下水的砷濃度高低,結果顯示幾乎所有測站均高過國際使用的飲用水水質標準,由表 4-1 顯示 7 號井與 9 號井的平均值約在 $500\mu\text{g/L}$ 上下,為本區最嚴重之兩個站,而拓樸聚類上也分布在圖 5-4 右上角的聚類中,其中聚類 A-3 與 A-4 僅有監測井 7 號的資料,且中心點最高濃度值達到 $1265\mu\text{g/L}$,可見其砷污染情況嚴重且變異性大。

pH 值

圖 5-17 所代表的 pH 值中心點,本研究區域 pH 介於 7~8.6 之間,大部分的水樣呈微鹼性,對照圖 5-12 砷濃度,顯示出在右方的聚類除了最下方聚類 D-2~D-4 以外都是高濃度砷的分布聚類,其 pH 值也高於左方的聚類。而隨著 pH 增高,膠體和黏土礦物帶更多負電荷,降低對以陰離子形式存在的離子吸附能力,有利於砷的釋出,因此使地下水砷濃度提高 (Park et al., 2006)。對照圖 5-16 的鹼度拓樸圖,可以得到鹼度與 pH 值在聚類分布上,除了靠研究區域北部之聚類 D-2 至 D-4 之外,皆呈現高度正相關,pH 值在聚類 B-3、B-4 也是呈現相對高值,而偏左方的鹽化因子高的聚類則呈現低值。pH 增加時吸附在氧化物或氫氧化物表面的砷將會減少 (Pierce and Moore, 1982; Xu et al., 1998)。

鹼度

鹼度一般是用來量測中和酸的能力,天然水中的鹼度大部分由弱酸的鹽類所組成,尤其是碳酸氫根(HCO_3^-),其乃是鹼度的主要形式。

由圖 5-16 鹼度的拓樸圖可以看出最大濃度出現在聚類 B-3、B-4，對應圖 5-4 為測站#6、#7、#9、#12 及#15，接著往左方及往下方聚類減少。對照圖 5-12 顯示鹼度與砷濃度有一些正相關的趨勢，在圖 5-12 的聚類 A-2 至 A-4、B-4 與 C-4 為高濃度砷的聚類，對應的測站#7、#9、#15 與#22，與鹼度存在著相對高值；相對的，左方沿海海水鹽化特徵的聚類與下方北邊分布的聚類則存在低值。Anawar et al. (2004)利用不同的 CO_3^{2-} 和 HCO_3^- 對孟加拉的沉積物進行萃取，其中 NaHCO_3 的濃度增加與 pH 增加，都會加強砷被萃取出來的效率。Kim et al. (2000)利用 HCO_3^- 對硫化砷進行萃取，砷與碳酸氫根形成錯合物進入水中，再進一步形成 H_3AsO_3 或 HAsO_4^{2-} ，結果顯示碳酸氫根最可能是造成砷自沉積物上脫附的媒介。且 pH 值也會影響砷在沉積物表面的吸附與去吸附作用；另外歐東坤(2005)對嘉南平原的砷濃度研究指出，高 HCO_3^- 濃度和高 pH 值可以造成鐵氧化物表面所吸附的砷釋出，而此論點可以由砷濃度與碳酸氫根及 pH 值有正相關性而得到支持。

硫酸鹽

對照圖 5-12 與圖 5-15 顯示，在高濃度砷分布的聚類，如聚類 A-3、A-4 與 B-4 濃度皆在 $500\mu\text{g/L}$ 以上，其對應的硫酸鹽濃度較低，顯示出在第 1~3 列砷濃度在處於高濃度值時，呈現負相關的結果。部分有高濃度硫酸根離子的聚類可能與此區域原是處於海相沉積環境所造成，或由於與圖 5-11 鈣離子濃度有正相關且顯著的關係，也顯示此區域高濃度硫酸根之產生與地質中鹽類溶解有關，與砷硫鐵礦之氧化作用應較無直接關係。由 Lu et al. (2010)的結果得到，濁水溪沖積扇往下游應呈現較還原之狀態。一般而言，還原地層中，則是伴隨著鐵、碳酸氫

根、磷酸根離子濃度與 pH 值上升(Shimada, 1996; McCreadie and Blowers, 2000; McArthur et al., 2001)。氧化還原電位達到還原時，膠體變得不穩定或對砷有高吸附能力的鐵錳氧化物或氫氧化物被還原，生成溶解度較高的亞鐵離子，吸附在表面的砷也隨著被釋放至地下水中。因此，此類型地下水中，高濃度砷伴隨著高濃度鐵、錳且溶氧量與硫酸根濃度較低，與砷濃度呈負相關(Smedley and Kinniburgh, 2002)。

由 SOM 的聚類結果可以得到下列幾個結果，對照，(1)代表拓樸圖下方的聚類 D-2 至 D-4 的測站主要分布舊虎尾溪以北，與西南沿海的測站位置區隔開，因此推論可能為拓樸圖中水質因子性質與其他聚類相異的主要原因，這些測站的鹽化情形相對不明顯且砷濃度也未存在高值，與上方的聚類特性不一致；(2)拓樸圖左上方的聚類中，測站皆分布於研究區域的西南沿海，鹽化因子 SO_4^{2-} 、EC、TDS、 Cl^- 、 Na^+ 、 K^+ 、 Mg^{2+} 、 Ca^{2+} 均呈現高濃度值，顯示此區域受到海水鹽化的情形非常顯著，也與測站近海的位置有關；(3)拓樸圖右上方的聚類雖然測站也分布在西南沿海，但鹽化因子的濃度未達到高值，而這些聚類有著高濃度砷存在，推論此區域高濃度砷的存在可能與此區域還原地層的環境有關，伴隨著 pH 值趨於鹼性、溶氧量低與硫酸根濃度低為利於砷釋出的環境。

5.2 區域地下水模式推估結果

本研究另一部分為建立區域地下水的推估模式。為了解地下水中的水質變化，建立一可靠模式為有效且直接之方法；傳統的地下水模式建立條件，必須先蒐集完整的地層參數、地下水水位與初始污染物濃度，且需要長期且持續地蒐集區域內地質、水文與水質資料，常常受到經費或設井區域的限制。傳統地下水水質推估模式經常使用數值分析法中有限元素法或有限差分法，將地下水區域劃分成規則相同與不同之網格，應用地下水傳輸理論計算每一網格之地下水水位，再依據地下水水位呈現地下水流向，進而推論出地下水污染擴散情形，數值方法之優點在於可應用在模擬抽水、河川、排水、河川水位變化、補注與蒸發散以及邊界條件等對於地下水水位與地下水水質之影響。數值方法雖有嚴謹的理論基礎，但受限於需具備完整地下水相關參數，如出使地下水水位、邊界地下水水位、地下水污染源位置、水平水力傳導係數與垂直水力傳導係數等，且對於地下水其他水質因子與砷的交互影響，較難以數值模是模擬。另外，數值模式在應用時需反覆驗證與校正參數以得到合理結果，屬於較繁複的方法。本研究以類神經網路具有解決高度非線性函數之能力，且具有整合異質性資料及容許高雜訊資料之特性，能彌補傳統數值模式，並提高推估砷濃度之精確度，為一有效且快速之方法。

5.2.1 輻狀基底函數類神經網路(RBFNN)

本模式以 SOM 聚類階段拓樸層之聚類中心點移除砷濃度資訊，作

為下一階段架構 RBFNN 中輻狀基底函數的中心點，以 RBFNN 監督式學習法，輸入砷以外水質因子至模式中，透過所建構的輻狀基底函數，以函數逼近方法找出輸入－輸出間的關係，經由函數轉換獲得隱藏層各神經元的輸出，再加權傳至輸出層，得到網路輸出之砷濃度推估值。

利用評估指標來檢視此地下水砷濃度推估模式之推估表現如表 5-3 所示，以 1998 年與 1999 年包含所有測站的最新資料作為測試資料，結果分為未加入砷濃度於 SOM 聚類(RBFNN)與加入砷濃度於 SOM 聚類(RBFNN_{As})，在第四章討論過加入砷濃度於 SOM 聚類有助於高濃度與低濃度砷的分離如圖 4-4、圖 4-5 所示，較適合來討論不同濃度的聚類內的水質特性，從而找出關係；其對於砷濃度的推估精度也有提升。從兩年的測試資料來看，未加入砷濃度於 SOM 聚類的誤差表現 RMSE 值 1998 年與 1999 年分別為 114.3 $\mu\text{g/L}$ 與 128.2 $\mu\text{g/L}$ ，若以標準差當作資料變異的程度標準，誤差表現雖優於原始資料標準差分別為 127.4 $\mu\text{g/L}$ 與 131.6 $\mu\text{g/L}$ ，但在加入砷濃度於 SOM 聚類時，誤差表現 RMSE 值分別降到 103.1 $\mu\text{g/L}$ 與 116.9 $\mu\text{g/L}$ 。因此本研究以加入砷濃度於 SOM 聚類為最終模式，其中 1999 年的砷濃度較 1998 年高，可能原因為 1999 年水質因子變異較大，造成推估誤差偏高。

表 5-3 RBFNN 地下水砷濃度推估結果

	RBFNN _{As***}			RBFNN		
	Training	Testing 1998	Testing 1999	Training	Testing 1998	Testing 1999
Ave. *($\mu\text{g/L}$)	97.10	75.20	85.00	97.10	75.20	85.00
SD. **($\mu\text{g/L}$)	168.80	127.40	131.60	168.80	127.40	131.60
RMSE($\mu\text{g/L}$)	144.90	103.10	116.90	149.20	114.30	128.20
MAE($\mu\text{g/L}$)	87.90	54.30	66.90	89.00	59.10	70.50
NMSE	0.74	0.66	0.80	0.78	0.81	0.96
CC	0.52	0.76	0.63	0.47	0.67	0.55

*為平均值(Average)

**為標準差(Standard Deviation)

***為加入砷濃度於 SOM 輸入因子

表 5-4 BPNN 地下水砷濃度推估結果

	Training	Testing - 1998	Testing - 1999
Ave.*($\mu\text{g/L}$)	97.1	75.2	85
SD.** ($\mu\text{g/L}$)	168.8	127.4	131.6
RMSE($\mu\text{g/L}$)	144.1	109.1	128.2
MAE($\mu\text{g/L}$)	82.7	60.6	71.0
NMSE	0.73	0.74	0.96
CC	0.52	0.60	0.45

*為平均值(Average)

**為標準差(Standard Deviation)

5.2.2 倒傳遞類神經網路(BPNN)

為探討 RBFNN 模式的適用性，本研究以 BPNN 所建立之作為驗證前面所建立模式與比較，結果如表 5-4 所示，1998 年與 1999 年的測試資料誤差表現 RMSE 值分別為 109.1 $\mu\text{g/L}$ 與 128.2 $\mu\text{g/L}$ ，對照表 5-3，BPNN 的推估結果優於未加入砷濃度於 SOM 聚類的 RBFNN 推估結果 1998 年與 1999 年分別為 114.3 $\mu\text{g/L}$ 與 128.2 $\mu\text{g/L}$ ，而誤差大於 RBFNN_{As} 推估結果分別為 103.1 $\mu\text{g/L}$ 與 116.9 $\mu\text{g/L}$ 。另外在 MAE 值的部分，BPNN 在 1998 年與 1999 年的測試資料分別為 60.6 $\mu\text{g/L}$ 與 71 $\mu\text{g/L}$ ，均大於 RBFNN 兩個情況的 MAE 值。整體來說，以 RMSE 值評估結果為 RBFNN_{As}<BPNN<RBFNN，而以 MAE 值評估結果為 RBFNN_{As}<RBFNN<BPNN，顯示 BPNN 在極值的推估結果優於 RBFNN，但中低值的表現則未優於 RBFNN_{As} 與 RBFNN 的兩個情況。因此可知加入砷濃度於 SOM 聚類之 RBFNN 推估模式效果最好。

5.2.3 空間推估

為呈現區域地下水水質模式推估地下水砷濃度之空間分布情形，本研究應用地理資訊系統(Geographic Information system, GIS)將 RBFNN 類神經網路推估 28 座監測井之地下水砷濃度推估值，藉由 GIS 提供克利金(Kriging)方法之空間推估功能，將本模式推估砷濃度值轉換為雲林沿海地區地下水砷污染空間分布圖，如圖 5-18 至圖 5-23 分別表示 1998 年 1 月至 1999 年 3 月應用模式推估地下水中砷污染分布情形。圖 5-18、圖 5-19 分別為 1998 年 1 月與 3 月，呈現在研究區域西南沿海一帶，地下水砷濃度有較高的趨勢，高濃度砷出現在監測井#9；

往北過了新虎尾溪(#15、#16、#18、#20)，往南至北港溪附近(#1、#2、#3、#27、#28)，砷濃度有漸漸下降的情況。圖 5-20 為 1998 年 6 月的砷濃度推估圖，監測井#7、#9 與北邊#19 都呈現高濃度值，其他區域與 1998 年 1 月、3 月相較起來，砷濃度有減輕的趨勢。到了 1998 年 9 月如圖 5-21 所示，監測井#7、#9 濃度有稍微減輕，但鄰近區域的砷濃度有提高的情況，延伸到新虎尾溪。1999 年 1 月與 3 月分別如圖 5-22、圖 5-23 所示，中部區域砷污染呈現往北港溪，向南方移動的趨勢。



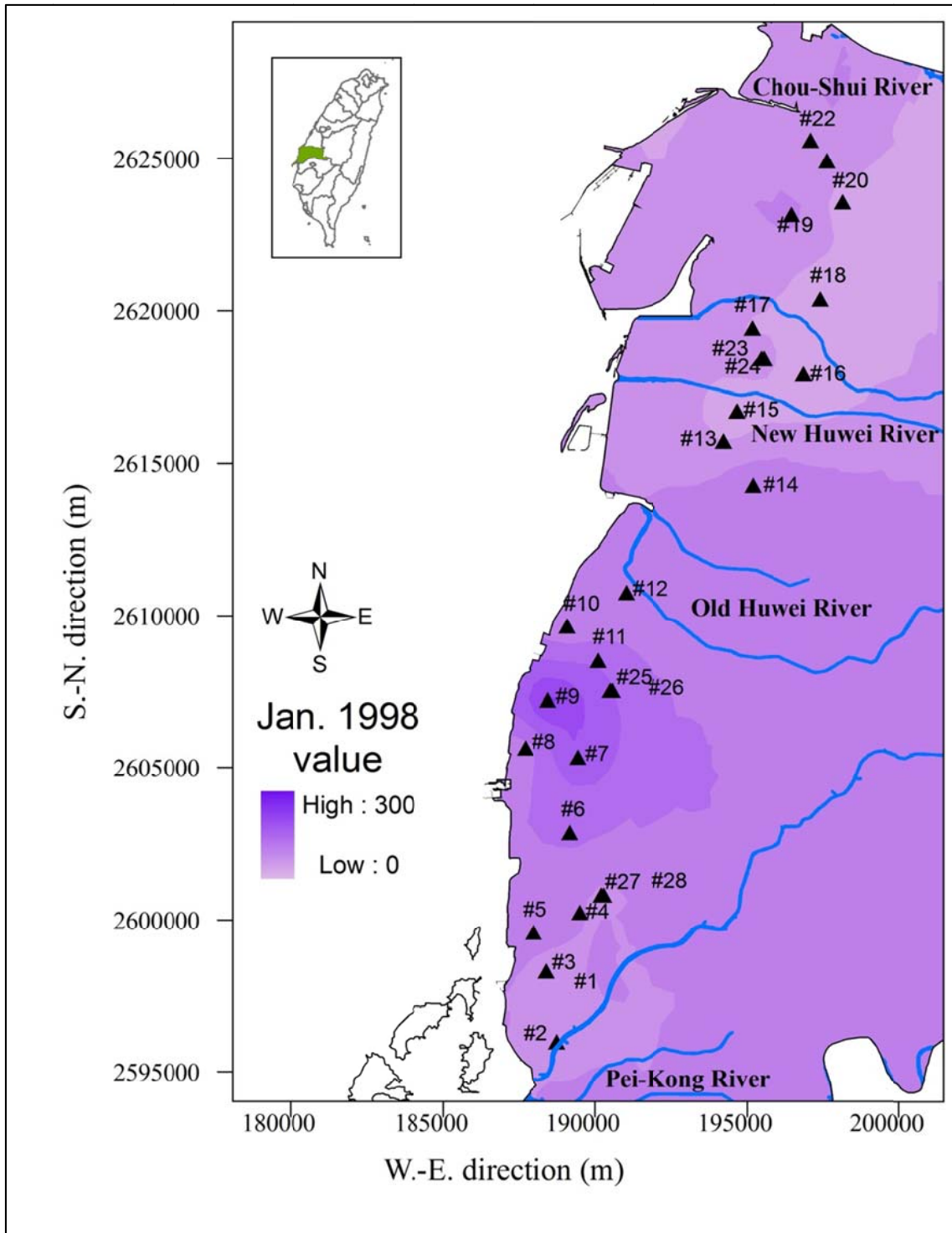


圖 5-18 1998 年 1 月模式推估地下水砷濃度分布圖

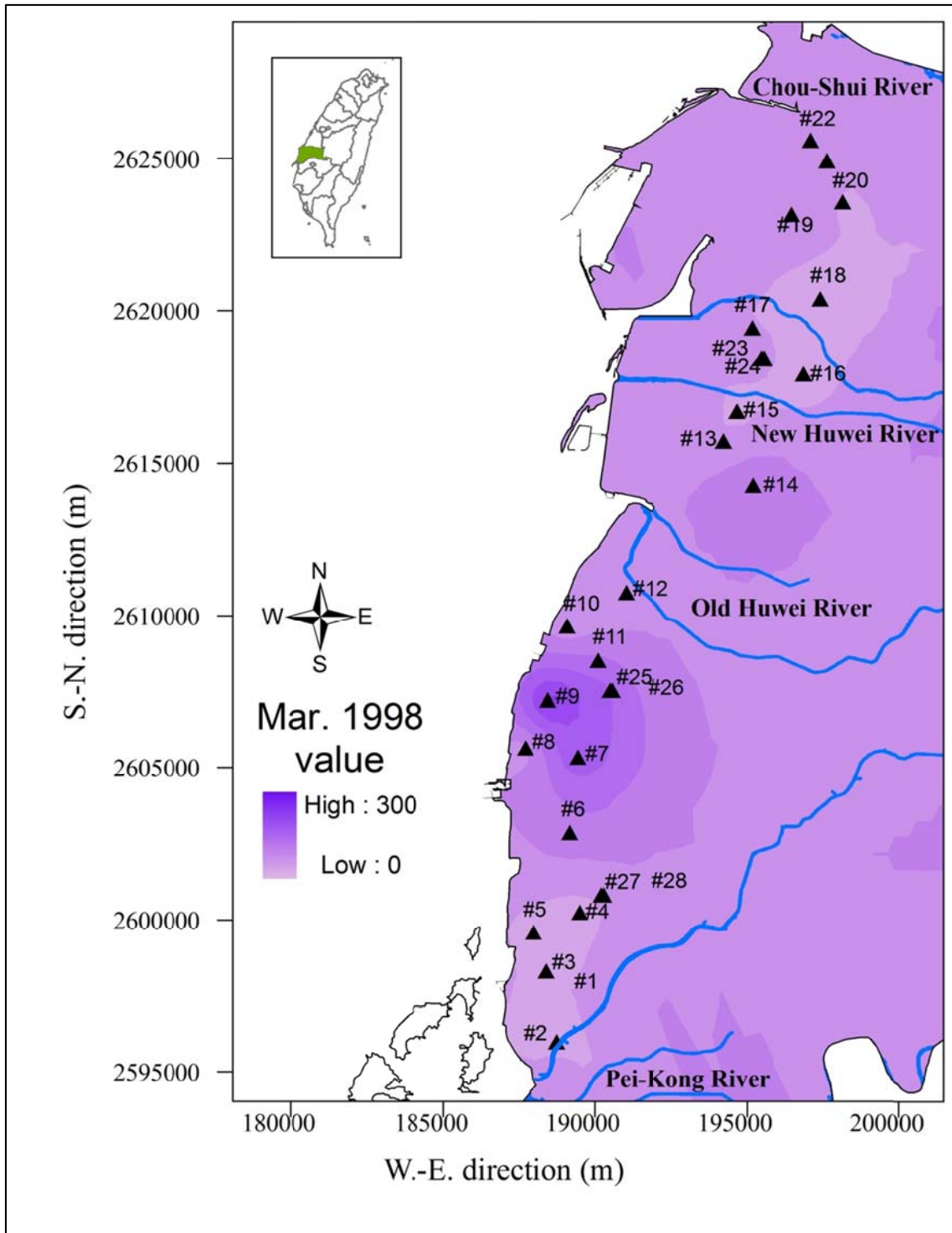


圖 5-19 1998 年 3 月模式推估地下水砷濃度分布圖

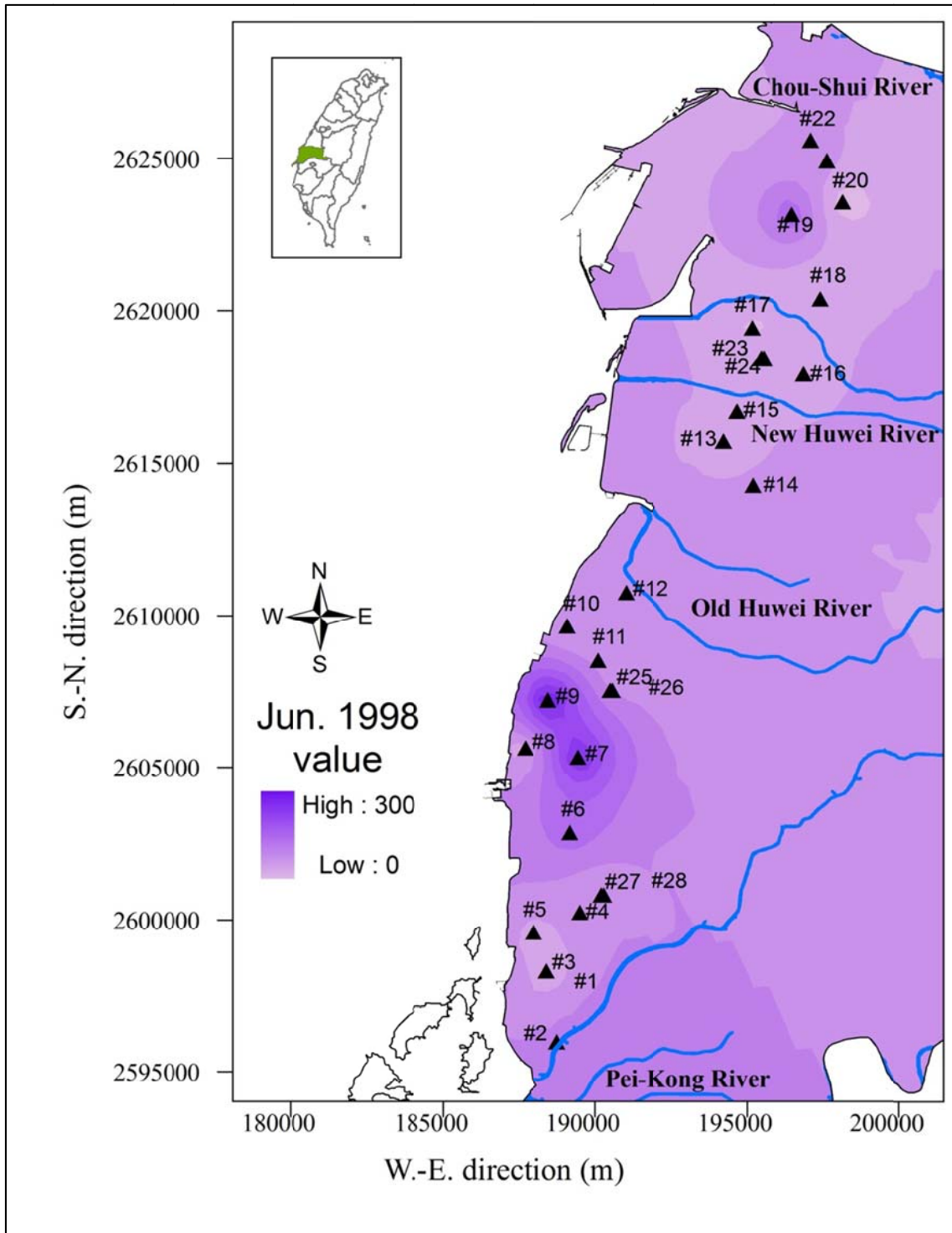


圖 5-20 1998 年 6 月模式推估地下水砷濃度分布圖

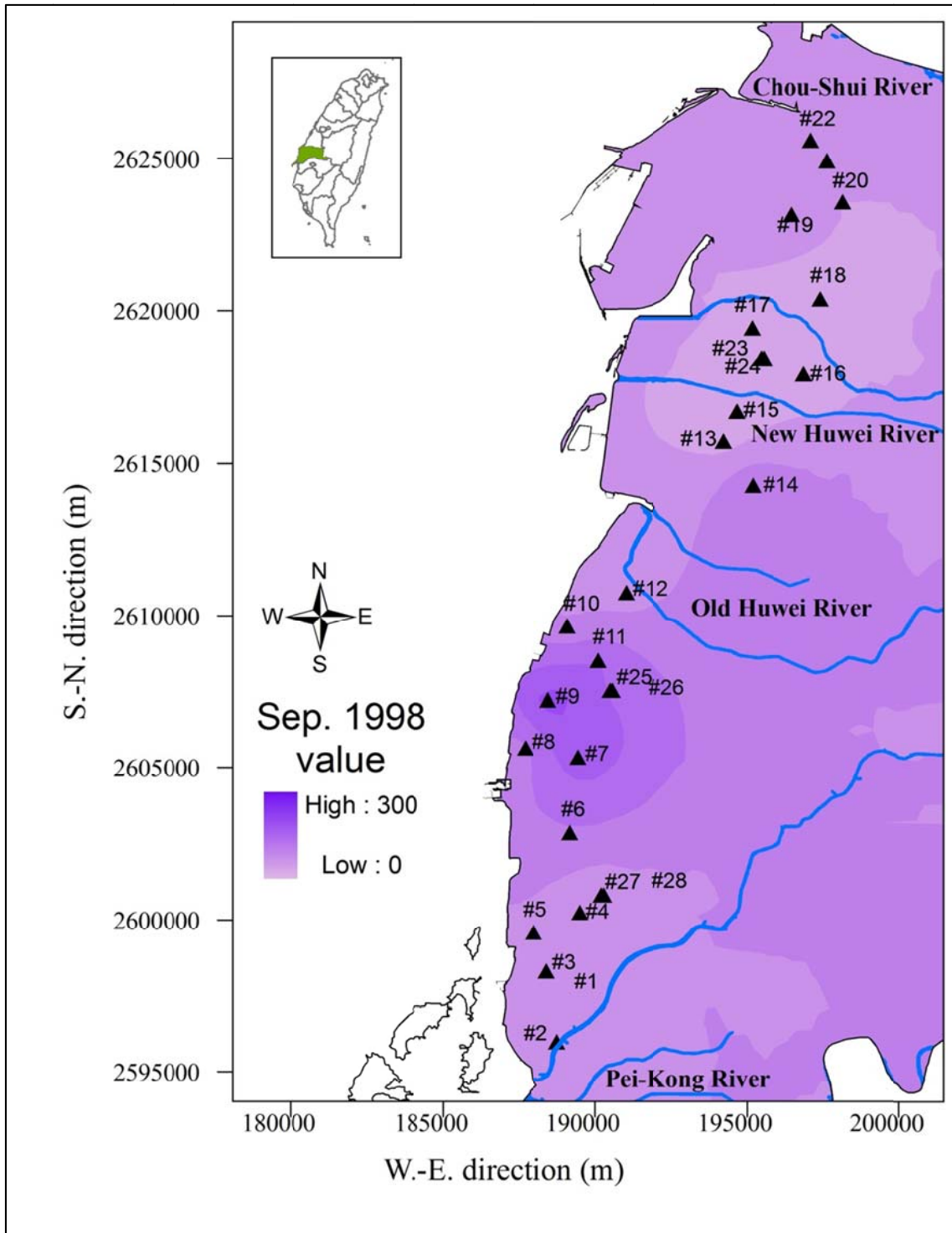


圖 5-21 1998 年 9 月模式推估地下水砷濃度分布圖

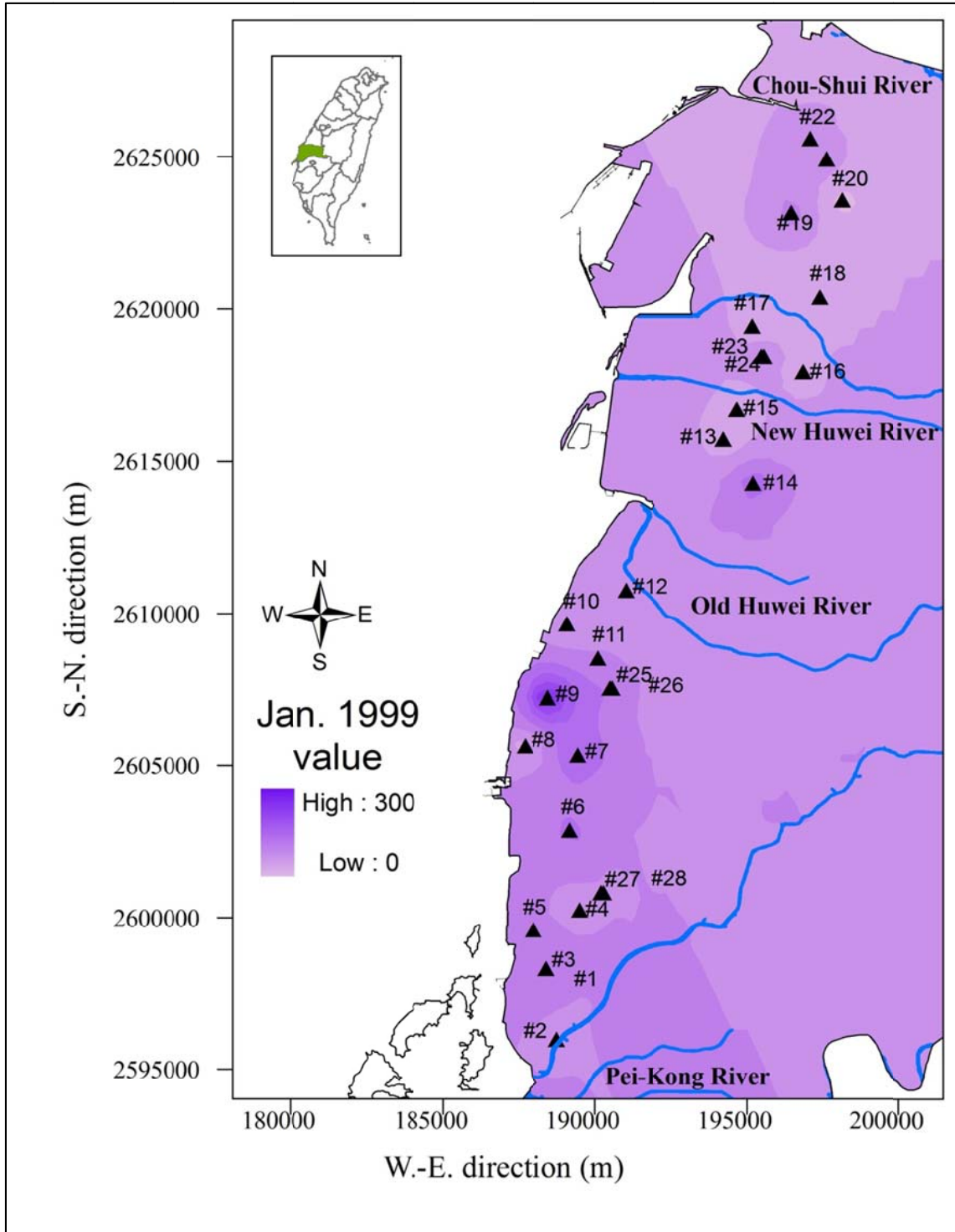


圖 5-22 1999 年 1 月模式推估地下水砷濃度分布圖

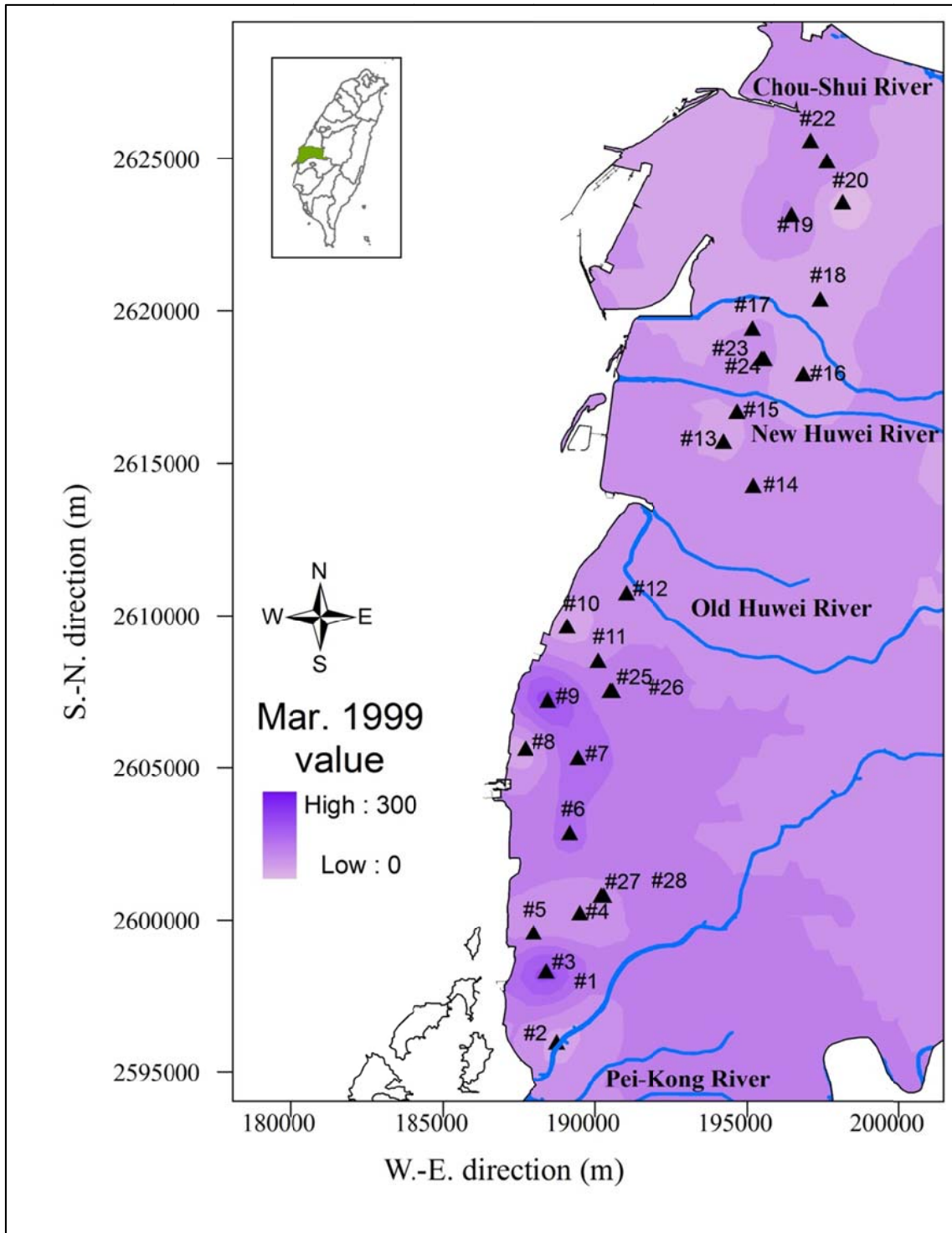


圖 5-23 1999 年 3 月模式推估地下水砷濃度分布圖

第六章 結論與建議

本研究蒐集雲林沿海地區地下水水質資料，以砷濃度與其他易測得的水質因子作為模式輸入，分別為：砷、溫度、pH 值、電導度、溶氧量、鹼度、總溶解固體、氯鹽、硫酸鹽、鈉、鉀、鎂與鈣，以自組特徵映射網路(SOM)方法先進行聚類，聚類所得的聚類中心點繪為拓樸圖，進行區域水質特徵分析；另外，輸入中心點於輻狀基底函數類神經網路(RBFNN)以決定隱藏層數目與權重，透過建立模式的方式推估砷濃度，最後以克利金方法做砷濃度的空間推估。其中探討加入砷濃度於自組特徵映射網路作為聚類資訊，對於聚類結果與推估結果的助益。另外將本模式(SOM 與 RBFNN)與倒傳遞類神經網路推估結果作比較，以驗證本研究之模式之穩定性與適用性。研究之結論與建議如下：

6.1 結論

1. 本研究在自組特徵映射網路(SOM)加入砷濃度資料與水質因子進行聚類，目的為探討砷濃度與水質因子之特徵，聚類結果顯示以砷濃度作為一輸入資訊，使得高、低值砷濃度的分離較顯著，有助於探討與了解高、低砷濃度聚類的特性。另外對於自組特徵映射網路(RBFNN)推估的結果亦有改善。
2. 以 SOM 對水質資料聚類，其特點在於透過拓樸圖的架構顯示輸入因子間的關係，聚類結果顯示，拓樸圖可以分出北部與南部區域的監測井，同時與聚類內的水質資料特性互相呼應，北部區域的砷污染問題較輕微且受到鹽化影響較輕微；而南部區

域部分測站存在高濃度砷。

3. SOM 拓樸圖結果顯示，研究區域受到海水鹽化影響嚴重，其中海水主要成份之因子如氯離子、硫酸根離子、鈉、鉀與鎂離子，與水中總離子相關的電導度之濃度高印證這點。位於拓樸圖左上方聚類的鹽化情形特別嚴重，部分因子甚至超過海水平均值，而這些聚類的測站分布在西南沿海地區。
4. 研究區域內溶氧量約在 1 mg/L 左右，顯示地下水氧氣消耗呈缺氧，與孟加拉受到砷污染的地下水環境相似，是容易形成砷釋出於地下水中之還原環境，此一特性可能為此區域地下水砷濃度皆高於國際間飲用水水質標準的因素，高砷濃度值最大超過 1200($\mu\text{g/L}$)，顯示此區域地下水存在高濃度砷的問題嚴重。
5. pH 與鹼度因子相關性高，與砷濃度在拓樸圖上也顯示出有正相關的關係，故推測 pH 值在偏鹼性的環境下，使得沉積物表面帶電荷，降低離子吸附能力，利於砷的釋出。另外，自然水中的鹼度主要形式為碳酸氫根，亦有研究指出碳酸氫根亦可能影響沉積物表面電荷的變化，為造成砷自沉積物上脫附的媒介。
6. 在硫酸根濃度低且存在高濃度砷的聚類中，溶氧量為低值、pH 值偏鹼性對應前人對於濁水溪沖積扇南翼的研究結果皆顯示此區域為還原地層，故推測此區域為伴隨高濃度鐵、錳且溶氧量與硫酸根濃度低之環境，而此環境亦有利於砷釋放至地下水中。
7. BPNN 建立之模式驗證 RBFNN 模式，結果顯示 RBFNN 之誤

差優於 BPNN 之推估誤差，表示 RBFNN 模式有相當穩定性，此模式特點為同時可對輸入因子進行水質特徵之分析，亦可達到良好的推估結果。

8. 最後以 1998 年與 1999 年的資料進行模式驗證，推估結果以地理資訊系統呈現 1998 年與 1999 年區域地下水砷濃度變化情形，由砷濃度在空間上的分布供決策者了解區域內砷污染情況，作為參考。

6.2 建議

1. 為更深入探討水質因子對地下水中砷濃度的直接影響，建議可以加入更多水質因子作為輸入，如鐵與總有機碳等，能夠更完整分析砷濃度在不同聚類釋出機制，且可能有效改善模式誤差的效果。
2. 區域地下水砷濃度為動態變化，未來可考慮加入地下水位進行探討，並建立即時的地下水砷濃度推估模式，對於砷污染嚴重地區應有很大幫助。
3. 對於了解砷污染的分布情形，可以空間推估方式或機率分布的方式呈現。因地下水水質變異性大，以水質因子推估砷濃度的精確度仍未達到相當準確，呈現分布情形與趨勢是一個比較好的方式。

第七章 參考文獻

1. Anawar, H.M., Akai, J., Komaki, K., Terao, H., Yoshioka, T., Ishizuka, T., Safiullah, S. and Kato, K. (2003) Geochemical occurrence of arsenic in groundwater of Bangladesh: sources and mobilization processes. *Journal of Geochemical Exploration* 77(2-3), 109-131.
2. Anawar, H.M., Akai, J. and Sakugawa, H. (2004) Mobilization of arsenic from subsurface sediments by effect of bicarbonate ions in groundwater. *Chemosphere* 54(6), 753-762.
3. Berg, M., Tran, H.C., Nguyen, T.C., Pham, H.V., Schertenleib, R. and Giger, W. (2001) Arsenic contamination of groundwater and drinking water in Vietnam: A human health threat. *Environmental Science & Technology* 35(13), 2621-2626.
4. Chang, F.J., Chang, L.C. and Huang, H.L. (2002) Real-time recurrent learning neural network for stream-flow forecasting. *Hydrological Processes* 16(13), 2577-2588.
5. Chang, F.J., Chang, L.C., Kao, H.S. and Wu, G.R. (2010) Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network. *Journal of Hydrology* 384(1-2), 118-129.
6. Chang, F.J. and Chang, Y.T. (2006) Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in Water Resources* 29(1), 1-10.

7. Chang, F.J. and Chen, Y.C. (2001) A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *Journal of Hydrology* 245(1-4), 153-164.
8. Chang, F.J. and Chen, Y.C. (2003) Estuary water-stage forecasting by using radial basis function neural network. *Journal of Hydrology* 270(1-2), 158-166.
9. Chang, F.J., Hu, H.F. and Chen, Y.C. (2001) Counterpropagation fuzzy-neural network for streamflow reconstruction. *Hydrological Processes* 15(2), 219-232.
10. Chang, F.J., Kao, L.S., Kuo, Y.M. and Liu, C.W. (2010) Artificial neural networks for estimating regional arsenic concentrations in a blackfoot disease area in Taiwan. *Journal of Hydrology* 388(1-2), 65-76.
11. Chang, L.C. and Chang, F.J. (2001) Intelligent control for modelling of real-time reservoir operation. *Hydrological Processes* 15(9), 1621-1634.
12. Chang, Y.T., Chang, L.C. and Chang, F.J. (2005) Intelligent control for modeling of real-time reservoir operation, part II: artificial neural network with operating rule curves. *Hydrological Processes* 19(7), 1431-1444.
13. D.S. Broomhead and D. Lowe (1998) Multivariate Function Interpolation and Adaptive Networks. *Complex Systems* 2: 321-355.
14. Goh, F., Lim, C., Sih, V.K.T., Ismail, Z. and Chooi, S.Y.M. (2005) Occurrence of arsenic-based defects and techniques for their

- elimination. *Ultra Clean Processing of Silicon Surfaces* VII 103-104, 87-90.
15. Grieu, S., Thiery, F., Traore, A., Nguyen, T.P., Barreau, M. and Polit, M. (2006) KSOM and MLP neural networks for on-line estimating the efficiency of an activated sludge process. *Chemical Engineering Journal* 116(1), 1-11.
 16. Harvey, C.F., Swartz, C.H., Badruzzaman, A.B.M., Keon-Blute, N., Yu, W., Ali, M.A., Jay, J., Beckie, R., Niedan, V., Brabander, D., Oates, P.M., Ashfaque, K.N., Islam, S., Hemond, H.F. and Ahmed, M.F. (2002) Arsenic mobility and groundwater extraction in Bangladesh. *Science* 298(5598), 1602-1606.
 17. Kim, M.J., Nriagu, J. and Haack, S. (2000) Carbonate ions and arsenic dissolution by groundwater. *Environmental Science & Technology* 34(15), 3094-3100.
 18. Kim, M.J., Nriagu, J. and Haack, S. (2002) Arsenic species and chemistry in groundwater of southeast Michigan. *Environmental Pollution* 120(2), 379-390.
 19. Kohonen, T. (1982) Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics* 43(1), 59-69.
 20. Kohonen, T. (1995) *Self-Organizing Maps*. Berlin: Springer-Verlag.
 21. Kuo, Y.M. and Chang, F.J. (2010) Dynamic Factor Analysis for Estimating Ground Water Arsenic Trends. *Journal of Environmental Quality* 39(1), 176-184.

22. Lee, B.H. and Scholz, M. (2006) Application of the self-organizing map (SOM) to assess the heavy metal removal performance in experimental constructed wetlands. *Water Research* 40(18), 3367-3374.
23. Lee, J.J., Jang, C.S., Wang, S.W., Liang, C.P. and Liu, C.W. (2008) Delineation of spatial redox zones using discriminant analysis and geochemical modelling in arsenic-affected alluvial aquifers. *Hydrological Processes* 22(16), 3029-3041.
24. Liu, C.W., Lin, K.H. and Kuo, Y.M. (2003) Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment* 313(1-3), 77-89.
25. Lu, K.L., Liu, C.W., Wang, S.W., Jang, C.S., Lin, K.H., Liao, V.H.C., Liao, C.M. and Chang, F.J. (2010) Primary sink and source of geogenic arsenic in sedimentary aquifers in the southern Choushui River alluvial fan, Taiwan. *Applied Geochemistry* 25(5), 684-695.
26. McArthur, J.M., Ravenscroft, P., Safiulla, S. and Thirlwall, M.F. (2001) Arsenic in groundwater: Testing pollution mechanisms for sedimentary aquifers in Bangladesh. *Water Resources Research* 37(1), 109-117.
27. McCreadie, H. and Blowes, D.W. (2000) Influence of reduction reactions and solid phase composition on porewater concentrations of arsenic. *Environmental Science & Technology* 34(15), 3159-3166.

28. NRC (2001) Arsenic in Drinking Water. Washiongton, DC: National Academy Press.
29. Park, J.M., Lee, J.S., Lee, J.U., Chon, H.T. and Jung, M.C. (2006) Microbial effects on geochemical behavior of arsenic in As-contaminated sediments. *Journal of Geochemical Exploration* 88(1-3), 134-138.
30. Pierce, M.L. and Moore, C.B. (1982) Adsorption of Arsenite and Arsenate on Amorphous Iron Hydroxide. *Water Research* 16(7), 1247-1253.
31. Polya, D.A., Gault, A.G., Diebe, N., Feldman, P., Rosenboom, J.W., Gilligan, E., Fredericks, D., Milton, A.H., Sampson, M., Rowland, H.A.L., Lythgoe, P.R., Jones, J.C., Middleton, C. and Cooke, D.A. (2005) Arsenic hazard in shallow Cambodian groundwaters. *Mineralogical Magazine* 69(5), 807-823.
32. Powell M. J. D. (1987) Radial Basis Function for Multivariable Interpolation: A Review. IMA Conference on Algorithms for Approximation of Function and Data. RMCS, Shrivenham, England. 143-167.
33. Raessler, M., Michalke, B., Schulte-Hostede, S. and Kettrup, A. (2000) Long-term monitoring of arsenic and selenium species in contaminated groundwaters by HPLC and HG-AAS. *Science of the Total Environment* 258(3), 171-181.
34. Sanchez-Martos, F., Aguilera, P.A., Garrido-Frenich, A., Torres, J.A. and Pulido-Bosch, A. (2002) Assessment of groundwater quality by

- means of self-organizing maps: Application in a semiarid area. *Environmental Management* 30(5), 716-726.
35. Schreiber, M.E., Simo, J.A. and Freiberg, P.G. (2000) Stratigraphic and geochemical controls on naturally occurring arsenic in groundwater, eastern Wisconsin, USA. *Hydrogeology Journal* 8(2), 161-176.
 36. Shimada, N. (1996) Geochemical conditions enhancing the solubilization of arsenic into groundwater in Japan. *Applied Organometallic Chemistry* 10(9), 667-674.
 37. Smedley, P.L. and Kinniburgh, D.G. (2002) A review of the source, behaviour and distribution of arsenic in natural waters. *Applied Geochemistry* 17(5), 517-568.
 38. Tseng, W. P., Chen, W. Y., Sung, J. L. and Chen J. S. (1961) A Clinical Study of Blackfoot Disease in Taiwan, An Endemic Peripheral Vascular Disease. *Memoire College Med., National Taiwan University*, 7, 1-18.
 39. USEPA (2006) Drinking Water Standards. <http://www.epa.gov/>
 40. WHO (2006) Guidelines for Drinking-Water Quality. <http://www.who.int/>
 41. Xu, H., Allard, B. and Grimvall, A. (1988) Influence of Ph and Organic-Substance on the Adsorption of as(V) on Geologic Materials. *Water Air and Soil Pollution* 40(3-4), 293-305.
 42. 水利署中區水資源局，2010，「湖山水庫工程計畫」，<http://www3.wracb.gov.tw/index.asp>。

43. 李品輝，2009，「以類神經網路探討全台蒸發量區域性分類與推估之成效」，國立臺灣大學生物環境系統工程學研究所碩士論文。
44. 高力山，2011，「人工智慧應用於區域地下水系統中砷污染推估之研究」，國立臺灣大學生物環境系統工程學研究所博士論文。
45. 高慧珊，2007，「以自組特徵映射網路推估蒸發量」，國立臺灣大學生物環境系統工程學研究所碩士論文。
46. 國立成功大學台南水工試驗所，1997~1998，「雲林離島式基礎工業區整體開發規劃調查研究報告」，第一部份，第七冊。
47. 張斐章、張麗秋，2010，「類神經網路導論－原理與應用」，滄海書局。
48. 陳文福、呂學諭、劉聰桂，2010，「台灣地下水之氧化還原狀態與砷濃度」，農業工程學報，第56卷第2期。
49. 曾文賓，1976，「烏腳病之診斷、治療與預防」，臺灣省政府烏腳病防治中心，第一輯，1-25。
50. 維基百科網站，www.wikipedia.org/。
51. 劉振宇、王聖瑋、盧光亮，2006，「台灣地區濁水溪沖積扇南翼之地下水砷污染可能來源與成因」，台灣土壤及地下水環境保護協會簡訊，第二十一期，第15頁-第19頁。
52. 歐東坤，2005，「嘉南地區地下水砷濃度之研究」，國立臺灣大學生物環境系統工程學研究所碩士論文。
53. 蔡文柄，2007，「應用類神經網路推估溪流之生物多樣性」，國立臺灣大學生物環境系統工程學研究所碩士論文。
54. 環保署，1998，「飲用水水源水質標準」，<http://www.epa.gov.tw/>。
55. 環保署，2000，「飲用水水質標準」，<http://www.epa.gov.tw/>。