

國立臺灣大學生農學院生物環境系統工程學系

碩士論文

Department of Bioenvironmental Systems Engineering

College of Bioresources and Agriculture

National Taiwan University

Master Thesis

以支持向量機界定農地重金屬高污染潛勢區

Delineation of Heavy Metal Pollution Potential Areas in

Agricultural Land Using Support Vector Machine

Analysis

劉修銘

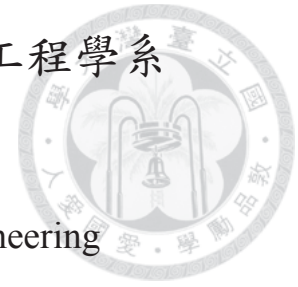
Xiu-Ming Liu

指導教授：張尊國 博士


Advisor: Tsun-Kuo Chang, Ph.D.

中華民國 104 年 6 月

June, 2015



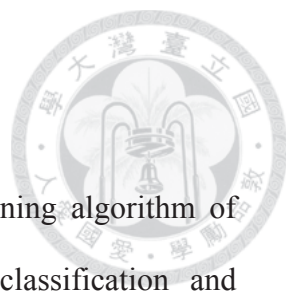
摘要



支持向量機 (Support Vector Machine, SVM) 屬於資訊科學領域中的機器學習 (Machine learning)，為一種監督式學習 (Supervised Learning) 演算法，其分類、回歸的功能亦可應用於地質、環境科學等相關領域。本研究將農地重金屬含量調查資料透過內梅羅指標 (Nemerow index, P_N) 轉換後，以 SVM 搭配地理資訊系統 (geographic information system, GIS) 劃分土壤農地重金屬高污染潛勢區，過程中透過 10 次交叉驗證優選訓練集標籤組成比例、樣本數量。結果顯示，彰化縣以 7,353 筆點位在陽性 ($P_N \geq 1.0$)、陰性 ($P_N < 1.0$) 標籤比 1:2 下建立之模型進行土壤重金屬污染潛勢預測，結果準確度 (Accuracy) 為 85.37%、F1-measure 為 0.692；桃園市在標籤比為 1:1 下，共 3,288 筆資料模型，污染潛勢預測之結果準確度為 71.58%、F1-measure 為 0.506。並將結果套疊河川流域、工廠、工業區等空間分布資訊，評析以 SVM 劃分農地重金屬污染潛勢區域之肇因及關連性，證實 SVM 演算法能有效地應用於土壤重金屬污染潛勢劃分，且在低訓練集樣本數即可達良好的分類效能。

關鍵詞：支持向量機、重金屬、土壤污染、污染潛勢圖

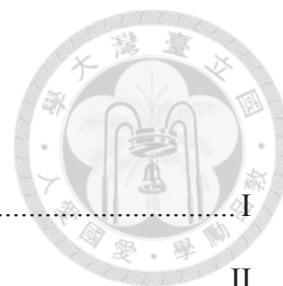
Abstract



Support Vector Machine (SVM) is a kind of supervised learning algorithm of machine learning in computer science, its function such as classification and regression could also be applied to related field e.g. geoscience and environmental science. In this research, the data of heavy metal pollution areas in agricultural land converted by Nemerow index (P_N) combined with SVM and geographic information system (GIS) classifies the highly potential heavy metal pollution areas in agricultural land. For modeling, the samples were optimized into an ideal proportion for training data set by using 10-fold cross validation. In Changhua County, at 7,353 points with the sample labeled ratio of positive ($P_N \geq 1.0$) and negative ($P_N < 1.0$) set to 0.5, results show the potential heavy metal pollution area with an accuracy of 85.37% and F1-measure of 0.692; In Taoyuan city, at 3,288 points with sample labeled ratio set to 1, results show the potential heavy metal pollution area with accuracy of 71.58% and F1-measure of 0.506. By interpreting the mapping of results with the information of surrounding geological features such as the distribution of river basins, factories and industrial zones, it allowed us to divide the causes and relationships of potential heavy metal polluted area with the use of SVM. Thus, the algorithm had proved that it could be validly applied to classify the potential heavy metal pollution areas in agricultural land even with low training data set.

Keywords: Support Vector Machine, heavy metals, soil pollution, pollution potential map

目錄



摘要.....	I
Abstract.....	II
目錄.....	III
圖目錄.....	V
表目錄.....	VI
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 研究目的.....	2
1.3 研究架構.....	2
第二章 文獻回顧.....	4
2.1 支持向量機 (Support Vector Machine, SVM) 發展及應用.....	4
2.1.1 核心函數 (kernel function) 及參數選定.....	6
2.1.2 網格搜尋 (grid-search) 及交叉驗證 (Cross Validation).....	7
2.2 混淆矩陣 (confusion matrix).....	8
2.3 內梅羅指標 (Nemerow Index, P_N) 與分級.....	9
第三章 研究材料、方法及流程.....	11
3.1 研究材料.....	11
3.1.1 研究區域.....	11
3.1.2 農地調查資料.....	14
3.2 研究方法.....	15
3.2.1 內梅羅指標 (P_N).....	15
3.2.2 支持向量機 (SVM).....	17
3.3 研究流程.....	20
3.3.1 支持向量機 (SVM) 分類.....	20

3.3.2	網格搜尋	20
第四章	結果與討論	21
4.1	內梅羅指標 (P_N) 評析結果	21
4.1.1	訓練集樣本資料標籤	24
4.2	訓練集標籤組成比例探討	26
4.3	訓練集樣本數量結果探討	27
4.4	農地重金屬污染關聯性評析	42
第五章	結論	49
	參考文獻	50



圖目錄



圖 1-1 研究架構圖	3
圖 3-1 彰化縣行政區域	12
圖 3-2 桃園市行政區域	13
圖 3-3 支持向量機概念示意	17
圖 3-4 最大間隔超平面	18
圖 3-5 支持向量機 (SVM) 分類流程	20
圖 3-6 網格搜尋流程	20
圖 4-1 彰化縣各採樣點 P_N 分布圖	22
圖 4-2 桃園市各採樣點 P_N 分布圖	23
圖 4-3 彰化縣閾值為 1.0 之各採樣點 P_N	24
圖 4-4 桃園市閾值為 1.0 之各採樣點 P_N	25
圖 4-5 彰化縣不同訓練集數量與網格搜尋結果	29
圖 4-6 彰化縣不同訓練集數量與高污染潛勢分布界定	32
圖 4-7 桃園市不同訓練集數量與網格搜尋結果	36
圖 4-8 桃園市不同訓練集數量與高污染潛勢分布界定	39
圖 4-9 彰化縣高污染潛勢區綜合研析圖	43
圖 4-10 桃園市高污染潛勢區綜合研析圖	44
圖 4-11 彰化縣水利小組評析污染及危害等級套疊 SVM 的預測結果	45
圖 4-12 桃園市水利小組評析污染及危害等級套疊 SVM 的預測結果	46
圖 4-13 彰化縣第 III 級圳路具污染潛勢區域套疊 SVM 的預測結果	47
圖 4-14 桃園市第 III 級圳路具污染潛勢區域套疊 SVM 的預測結果	48

表目錄



表 2-1 混淆矩陣 (Confusion matrix)	8
表 2-2 P_N 地表水之分級標準	10
表 2-3 P_N 地下水之分級標準	10
表 2-4 P_N 土壤農地重金屬分級標準	10
表 3-1 研究區概述	11
表 3-2 彰化縣農地表土重金屬濃度調查統計	14
表 3-3 桃園市農地表土重金屬濃度調查統計	14
表 3-4 P_N 的評析基準值 (C_{si})	15
表 3-5 農地重金屬分級標準	16
表 4-1 彰化縣 P_N 分級分布表	22
表 4-2 桃園市 P_N 分級分布表	23
表 4-3 彰化縣 P_N 分級 (閾值設定為 1.0)	24
表 4-4 桃園市 P_N 分級 (閾值設定為 1.0)	25
表 4-5 彰化縣樣本比例實驗結果	26
表 4-6 桃園市樣本比例實驗結果	26
表 4-7 彰化縣不同訓練集數量對應預測結果	28
表 4-8 桃園市不同訓練集數量對應預測結果	35
表 4-9 行業類別及其重金屬污染種類	42



第一章 緒論

1.1 研究動機

土壤是人類賴以生存的基本自然資源，亦是農業發展的重要基礎。過去數十年間，臺灣工商業高速發展，土壤接受來自灌溉、掩埋、棄置等途徑的累積，成為各種污染物之最終承受體。

農地土壤污染的來源，主要由水污染隨著灌溉渠道或灌排渠道引入農地，造成重金屬在農地中殘留累積。而種植高含量重金屬之農地的水稻，其重金屬含量亦具有高污染之潛勢，其中以農地鎘、鉛污染為主，公害事件則造成鎘米事件為最多（張，2002）。1950年於日本富山縣發生全球第一起鎘米污染事件，調查其原因是當地礦業排放出重金屬鎘至河川中，農民引灌到農田以致種出鎘米，並造成該地居民遭受痛痛病危害。1982年於桃園縣觀音鄉大潭村發生第一宗臺灣鎘米事件，因當地化工廠生產含鎘和鉛的安定劑，隨渠道排放至農地所致；隨後彰化縣、台中縣、雲林縣等縣市亦接連發生鎘米污染事件，引發國民關注。根據農委會統計，近十年來臺灣每人每年白米攝取量平均約49公斤，鎘米問題須慎重處理（Chen & Lee, 1995; Hsu et al., 2010; Yao et al., 2014）。

早期環保單位是採系統性採樣方式，首先將臺灣以網格劃分後進行農地重金屬污染調查，再依據調查結果將重金屬污染較嚴重之區域進行網格細分並調查，以逐步發覺污染範圍。由於土壤採樣及監測成本昂貴，且欠缺有效率反映高污染潛勢區域之方法。在現代科技進步下，搜尋污染熱點之演算法持續發展，透過將過往農地土壤巨量調查資料，透過篩選機制、地理資訊系統（Geographic Information System, GIS）、監督式學習等整合分析，以有效方法確認高污染潛勢之農地區域，找出重金屬高污染潛勢之土壤區域，提高土壤污染監測、防治等效率，解決改善臺灣地區土壤污染問題。



1.2 研究目的

本研究之目的為應用機器學習，以支持向量機(Support Vector Machine, SVM)區分農地重金屬高污染潛勢區。條列目的如下：

1. 應用支持向量機將農試所的採樣資料，訂定閾值、標籤及訓練模型，並透過預測分類、驗證，篩選出高污染潛勢區域。
2. 透過調整訓練資料集(training data set)的不同標籤比例及數量，探討其對 SVM 分類結果之影響程度。
3. 將 SVM 分類結果套疊整合環境地表特徵，評析土壤重金屬高污染潛勢區域之污染肇因。

1.3 研究架構

本研究之架構如圖 1-1 所示。

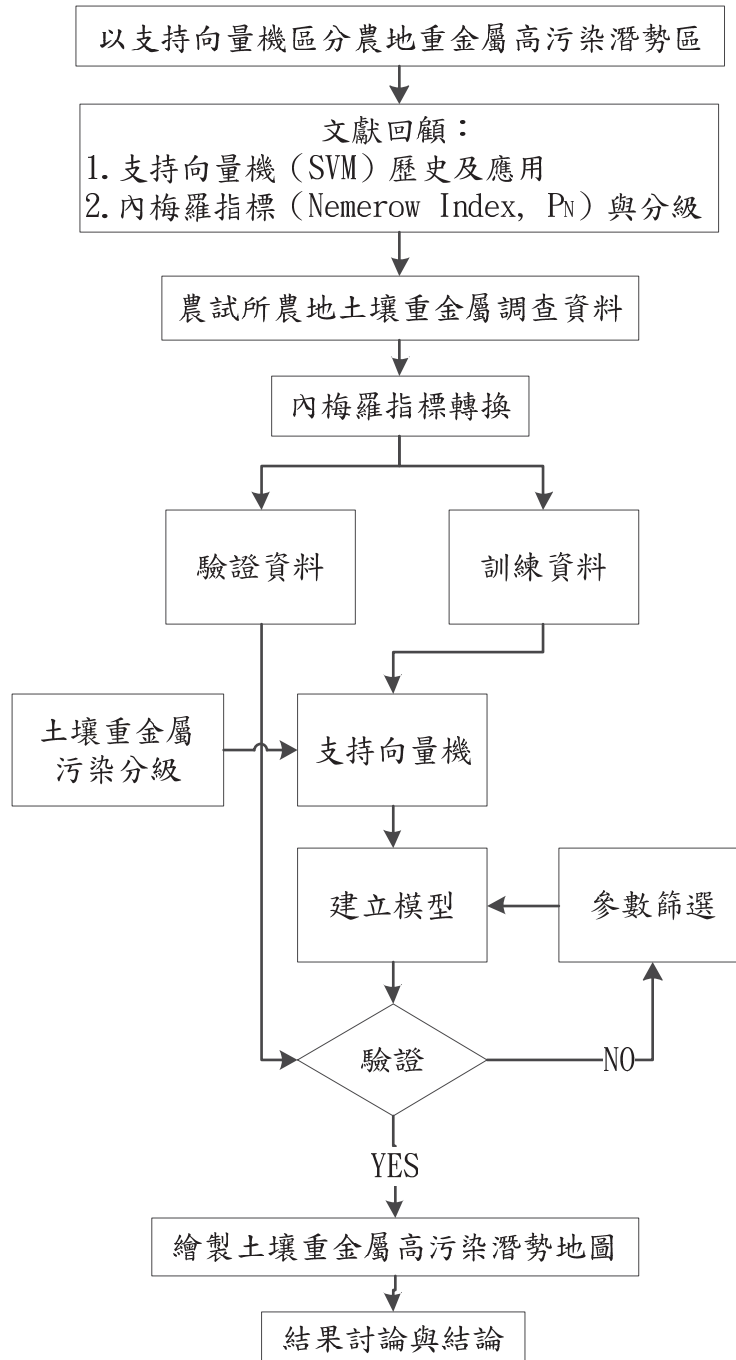


圖 1-1 研究架構圖

第二章 文獻回顧

2.1 支持向量機 (Support Vector Machine, SVM) 發展及應用

支持向量機 (Support Vector Machine, SVM) 為一種根據統計學習理論 (Statistical Learning Theory) 而發展出的機器學習方法，最早由 Bernhard E. Boser 等在 1992 年的計算機學習理論 (Computational Learning Theory, COLT) 的研討會提出概念。Corinna Cortes 和 Vladimir N. Vapnik 在 1995 年《Machine Learning》期刊中發表 (Boser et al., 1992; Cortes & Vapnik, 1995)。因其優秀且具高準確性的分類能力，且能處理高維度的資料，SVM 現今已被廣泛地運用在處理分類的問題上。

SVM 於地球科學與環境領域之應用有，1999 年於日內瓦湖 (Lake Geneva)，透過湖底沉積物得到鎘 (Cd) 之濃度分布，並以閾值 (threshold value) 為 $0.8 \mu\text{g}\cdot\text{g}^{-1}$ 及 $1.0 \mu\text{g}\cdot\text{g}^{-1}$ 將訓練資料區分，以空間座標、鎘濃度資料透過 SVM 建立模型。另外使用測試集作準確度驗證，並進一步預測整個湖底的重金屬鎘濃度分布 (Mikhail et al., 1999)。

2008 年於中國四川省之汶川大地震，以湍江流域地震引發滑坡、高程、坡角、坡度等資料，作為 SVM 訓練資料之類別。結果得到滑坡敏感性映射之區域分布，研究討論中以不同滑坡數量作為訓練資料，探討對結果準確度之影響 (Chong et al., 2012)。

2010 年於塞爾維亞，透過土壤調查的定量技術來預測和分類土壤種類。其中透過採樣點內含之 pH、N、 K_2O 、 P_2O_5 、CEC、clay sand (粘土砂) 等作為訓練之類別，用以建立模型並預測結果，將土壤區分為黑鈣土、淋溶土等種類 (Miloš et al., 2010)。

2012 年於伊朗克爾曼省 (Kerman province) 的 Now Chun 礦床區，Abedi Maysam 等將地球物理及地球化學等不同的數據層集成分析。透過 SVM 訓練、預測礦產勘探有前景之區域，以評估銅礦區，並找尋一個前景更加好的地區進行

礦產勘查。此研究結果顯示，支持向量機可作為礦物前景的預測，並有效提高勘探前景潛勢區域的分辨率和降低鑽井風險 (Maysam et al., 2012)。

2012 年於西班牙的戈梅拉島 (Gomera Island)，該地理區位裡，分別以傳統農業、有機農業法種植芒果。透過土壤中有些金屬含量，如鈣、鈷、銅、鐵、鉀、鎂、錳、鈉、鎳和鋅的濃度含量，作為關係類別並以 SVM 訓練分類。結果顯示高達 93 % 預測能力，準確度相當高 (Hernández-Sánchez et al., 2012)。

由上述文獻得知，SVM 所涉及的环境相關層面非常廣泛，從地理環境、土壤性質、探勘調查、底泥重金屬濃度分布等領域皆有涵蓋，且結果顯示在區分類別上具有良好效能。



2.1.1 核心函數 (kernel function) 及參數選定

SVM 可透過選擇核心函數 (kernel function) 處理應用非線性問題 (Zuo et al., 2011), 以函數 ϕ 映射到高維度空間, 求訓練向量集 x_i 之線性最佳超平面 (Optimal Hyperplane)。定義 $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ 稱為核心函數, 常見有以下四種:

1. linear : $K(x_i, x_j) = x_i^T x_j$
2. polynomial kernel : $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma \in \mathfrak{R}^+$
3. radial basis function (RBF) : $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma \in \mathfrak{R}^+$
4. sigmoid : $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

其中, γ, r, d 皆是核心參數。

由於線性核心 (linear kernel) 是 RBF 中的特例 (Keerthi & Lin, 2003); sigmoid 核心在適當訓練參數下, 分類行為像 RBF (Lin & Lin, 2003)。此外, 與 RBF 相比, 多項式核心 (polynomial kernel) 需要調整較多參數, 導致建立模型、選擇參數時較於複雜, 且在 Huang 等 (2004); Tay 和 Cao (2001) 研究指出, 多項式核心在訓練階段需要較長的時間, 其結果不優於 RBF, 故本研究選擇 RBF 為核心函數。

2.1.2 網格搜尋 (grid-search) 及交叉驗證 (Cross Validation)

在 RBF 核心函數中有兩個參數，分別懲罰值 (penalty value, C) 及參數 γ ，必須選擇一組好的參數來解決問題。本研究使用網格搜尋 (grid-search)，透過改變間距測試每對 (C, γ) 之精確度，為增加網格搜尋效率，本研究以指數型 (exponential) 增加的序列做測試，如： $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ ； $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ 。

本研究並透過交叉驗證篩選出最高精準度時之參數對 (C, γ) ，稱為最佳參數對 (best pair of parameters)，其中用於訓練資料建立模型的樣本子集稱為訓練集 (Training Set)，而剩下用以驗證模型之樣本子集則稱為驗證集 (Validation Set)。

保持交叉檢定法 (Holdout Cross Validation)，即將原始資料集隨機劃分成兩個獨立的子集合，亦為訓練集與驗證集，並只使用訓練集建立模型，再使用驗證集預測輸出值並判斷其正確性。

K 次交叉驗證 (K-fold Cross Validation) 的作法是將訓練集均分成 K 份，每次取其中一份做為驗證集，剩下 K-1 份做為訓練集，重複步驟驗證 K 次。



2.2 混淆矩陣 (confusion matrix)

透過混淆矩陣 (confusion matrix) 將演算法的性能視覺化，是一種統計模型評估的標準工具。其中每行 (column) 代表一個實際分類；而列 (row) 則表示預測分類，結果會透過預測值與實際值之關係，將模型中的所有案例分類到不同的類別目錄。如表 2-1 所示，並透過準確度 (Accuracy)、精確度 (Precision)、召回率 (Recall) 及 F1-measure 來評估測量 (Evaluation Measurement) (Kohavi & Provost, 1992; Powers et al., 2011)。

儘管準確度同時考量真陽性 (True positive)、真陰性 (True negative) 預測分類結果，於環境污染領域應用時，以精確度、召回率、F1-measure 等指標計算表達，可使污染分類結果更清晰 (如：Pandey et al., 2013)。

表 2-1 混淆矩陣 (Confusion matrix)

		實際分類	
		陽性 (Condition Positive)	陰性 (Condition Negative)
預測分類	Test Positive	真陽性 (True positive)	偽陽性 (False positive)
	Test Negative	偽陰性 (False negative)	真陰性 (True negative)
		總量 (Total population)	

$$\text{➤ 準確度 (Accuracy)} = \frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$$

$$\text{➤ 精確度 (Precision)} = \frac{\sum \text{True positive}}{\sum \text{Test Positive}}$$

$$\text{➤ 召回率 (Recall)} = \frac{\sum \text{True positive}}{\sum \text{Condition Positive}}$$

$$\text{➤ F1 - measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$



2.3 內梅羅指標 (Nemerow Index, P_N) 與分級

內梅羅指標 (Nemerow Index, P_N) 用於評價環境品質時，不僅考慮評估之監測項平均值，亦加重計算超標程度最大項，可突顯超標程度最大之項目影響。(Nemerow, 1974；張，2012)。

目前臺灣環保署已建置多項環境指標監測系統，例如：空氣污染綜合指標 (PSI)、水質指數 (WQI) 及河川污染指數 (RPI) 等。但 RPI 判定河川水質污染程度之判斷基準僅為四分法，並不考量原本不同水體分類之意涵。翁 (2010) 指出水源保護區之河段監測，少數水質參數超過標準值時，在水質指標上並無顯著反應，實失去評估並採取適當管制措施之先機；然而， P_N 可以判斷環境質量與評價標準之間的關係，透過選擇適當的評析基準值 (C_{si}) 調適之。有關土壤污染文獻中指出，若選用該國土壤背景值為 C_{si} ，評價後之 P_N 數值通常小於 3；而 P_N 大於 3 的地區，幾乎確定為污染區域 (劉等，2009；張等，2012)。

P_N 屬於綜合指標，較單一指標易了解整體土壤重金屬污染。目前臺灣環保署訂定各項重金屬 (砷、鎘、鉻、銅、汞、鎳、鉛、鋅) 之土壤管制標準及監測標準，以此固定門檻作為評估基準屬於單一指標方法，較難知其污染全貌。許多研究提到 P_N 會因地區、背景不同及評價對象的改變而有差異，如評估陸域水體及土壤污染，則有不同的定義，需依照研究對象提出適當的分類，並針對差異等條件選定等級分類標準 (如洪等，2012；洪等，2013；劉等，2014)。

以下所列為各定地透過內梅羅綜合指標評價其環境質量制定之分類標準，如表 2-2，位於中國雲南省通海縣的杞麓湖評價地表水，因其地表水的質量已為中國顯著環境規劃和管理的問題之一 (Tang et al., 2011)；及表 2-3，為在中國太行山東麓所監測地下水的分類標準 (李等，2009)；另外，國內外應用在土壤農地重金屬之分級，如表 2-4 所示 (張等，2012)。

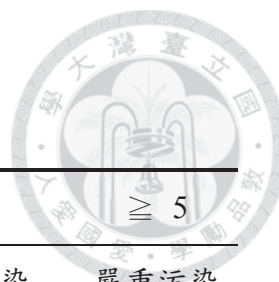


表 2-2 P_N地表水之分級標準

P _N 分級	≤ 1	1~2	2~3	3~5	≥ 5
地表水等級	清潔	輕度污染	污染	重度污染	嚴重污染

(Tang et al., 2011)

表 2-3 P_N地下水之分級標準

P _N 分級	≤ 0.8	0.8~2.5	2.5~4.25	4.25~7.2	≥ 7.2
地下水等級	優良	良好	較好	較差	極差

(李等, 2009)

表 2-4 P_N土壤農地重金屬分級標準

P _N 分級	< 0.7	0.7~1.0	1.0~2.0	2.0~3.0	≥ 3.0
土壤污染等級	優良	安全	警戒	污染	危害

(張等, 2012)

第三章 研究材料、方法及流程



3.1 研究材料

3.1.1 研究區域

截至 102 年末，依環保署公告資料顯示，累計各縣市農地被調查且公告列管之農地共 4,402 筆（面積 746 公頃），其中以桃園市公告之 1,701 筆（面積 211 公頃）為最多，其次為彰化縣 328 筆（面積 61 公頃），故本研究選擇彰化縣、桃園市作為研究區域。

表 3-1 研究區概述

	彰化縣	桃園市
土地面積（公頃）	107,440	122,100
耕地面積（公頃）	69,111	37,544
主要產業	染整、畜牧、電鍍、 金屬表面處理等	金屬製品製造業、 機械設備製造修配 業，及電子機械器 材製造、修配業等 為主

資料整理自環保署計畫，2011



(一) 彰化縣

彰化縣位於臺灣省中部，如圖 3-1 彰化縣行政區域所示，共分 26 鄉鎮市，北起大肚溪與台中縣為界，南至濁水溪與雲林縣毗鄰，東倚八卦山與南投縣相隔，西臨臺灣海峽。土地面積 107,440 公頃，耕地面積 69,111 公頃(佔全縣面積 65.5%)，為臺灣重要稻米產區，也是農牧重點縣市。然而，彰化縣內小型工廠零星散佈於田間之工農混合型態，工廠登記多以金屬製品、紡織業及塑膠製品業，區內潛在污染來源有染整、畜牧、電鍍、金屬表面處理等。另外，彰化水利會彰化工作站轄區內之灌溉渠道東西二、三圳幹、支線受重金屬污染嚴重，已成為重點重金屬改善復育地區(張，2010)。



圖 3-1 彰化縣行政區域

(二) 桃園市

桃園市位於臺灣省西北，如圖 3-2 桃園市行政區域所示，全市共劃分為 13 區，西臨臺灣海峽，東南以達觀山、蹦蹦山與台北及宜蘭線分界，為臺北盆地與桃園台地的天然劃區。全縣土地總面積約 122,100 公頃，耕地面積 37,544 公頃（佔全縣面積 30.7 %）。市內工廠以金屬製品製造業、機械設備製造修配業，及電子機械器材製造、修配業等為主。其盜採砂石並回填有害事業廢棄物，導致該區域之土壤污染更為複雜（張，2011；環保署，2011；桃園縣政府，1998）。

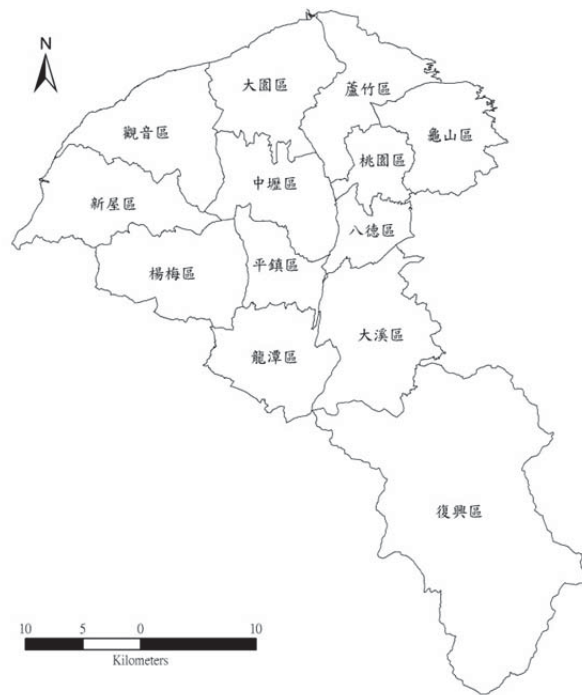


圖 3-2 桃園市行政區域



3.1.2 農地調查資料

自 1992 年起，行政院農業委員會農業試驗所（簡稱農試所），為農地品質現況進行具體調查並分析目前耕作方式，使國人了解農地土壤肥力、土壤污染等疑慮，並期待能防範污染區域、改善農田地力品質，故其進行全臺農地面積約 78 萬公頃之土壤品質及生產力調查。調查方式以間隔 250 公尺為單位之網格式採集土壤樣本，以 0.1 M HCl 萃取方法檢測鎘 (Cd)、鉻 (Cr)、銅 (Cu)、鎳 (Ni)、鉛 (Pb)、鋅 (Zn)，共六項重金屬濃度。為研析農地重金屬高污染潛勢區域，本研究採用彰化縣 10,371 筆、桃園市 9,005 筆之表土數據，表 3-2、表 3-3 分別為彰化縣、桃園市農地表層土壤的重金屬調查濃度基本統計分析。

表 3-2 彰化縣農地表土重金屬濃度調查統計

項目	Cd	Cr	Cu	Ni	Pb	Zn
平均值	0.28	0.71	10.88	3.56	8.07	37.54
中位數	0.18	0.31	7.46	2.17	6.44	8.98
第 1 四分位數	0.10	0.19	5.01	1.47	4.36	5.98
第 3 四分位數	0.37	0.52	10.40	3.34	9.79	14.04
最大值	89.14	363.07	5512.46	911.32	2674.47	188590.86
標準差	1.07	5.10	59.03	12.13	28.22	1871.12

資料總數 10,371 筆，單位：mg·kg⁻¹，檢測單位：農試所

表 3-3 桃園市農地表土重金屬濃度調查統計

項目	Cd	Cr	Cu	Ni	Pb	Zn
平均值	0.17	0.39	18.93	1.33	8.29	18.60
中位數	0.13	0.24	4.38	0.77	5.89	7.78
第 1 四分位數	0.07	0.14	2.65	0.48	4.12	3.93
第 3 四分位數	0.21	0.40	8.57	1.30	8.47	15.94
最大值	24.37	40.49	6411.70	158.09	2200.25	3206.07
標準差	0.45	0.90	129.25	3.47	29.39	59.65

資料總數 9,005 筆，單位：mg·kg⁻¹，檢測單位：農試所



3.2 研究方法

3.2.1 內梅羅指標 (P_N)

本研究選用內梅羅指標 (Nemerow index, P_N) 進行土壤重金屬污染之評價，
計算公式如：

$$P_N = \sqrt{(P_{\text{avg}}^2 + P_{\text{max}}^2)/2} = \sqrt{\frac{1}{2} \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{C_i}{C_{si}} \right)^2 + \left(\frac{C_m}{C_s} \right)^2 \right]}$$

式中， $P_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{C_i}{C_{si}} \right) ;$

$$P_{\text{max}} = \max \left(\frac{C_1}{C_{s1}}, \frac{C_2}{C_{s2}}, \dots, \frac{C_i}{C_{si}} \right) = \left(\frac{C_m}{C_s} \right) ;$$

C_i : i 監測項目的實測值；

C_{si} : i 監測項目的評價基準值；

n : 受評估監測項目的總數；

農試所以 0.1 M HCl 萃取土壤重金屬資料，與「臺灣地區土壤重金屬含量標準與等級」之檢測方法相同，故以「臺灣地區土壤重金屬含量標準與等級」之第三級背景值上限，做為監測項目的評價基準值 (C_{si}) 進行評價，如表 3-4 所列。

表 3-4 P_N 的評析基準值 (C_{si})

重金屬	鎘	鉻	銅	鎳	鉛	鋅
C _{si} (mg·kg ⁻¹)	0.39	10.0	20.0	10.0	15.0	25.0

採「臺灣地區土壤重金屬含量標準與等級」之第三級背景值上限

$P_N < 1.0$ ，說明環境質量達評析標準的要求； $P_N \geq 1.0$ ，說明環境質量已不能滿足評析標準的要求，本研究定義 $P_N < 1.0$ 為未污染農地以陰性 (negative) 表示，而 $P_N \geq 1.0$ 為污染農地以陽性 (positive) 表示，如表 3-5。



表 3-5 農地重金屬分級標準

P_N 分級	< 1.0	≥ 1.0
等級	未污染	污染



3.2.2 支持向量機 (SVM)

支持向量機 (Support Vector Machine, SVM) 為一種監督式學習 (Supervised Learning)。當有資料要分成兩類或更多類別，如圖 3-3 支持向量機概念示意，藉由 SVM 方法找出超平面 (Hyperplane) 將兩種不同類別的資料分群，另外將超平面與最接近的訓練資料點間的距離定義為間隔 (Margin)。支持向量機的目標即尋找具有最大間隔的超平面作為分類邊界，此時稱為最佳超平面 (Optimal Hyperplane)，能將訓練資料盡量明確地區隔開來。空間中最靠近最佳超平面的資料點稱為支持向量 (Support Vector)，可利用此超平面判定新進資料之所屬類別。(Boser et al., 1992; Cortes & Vapnik, 1995)

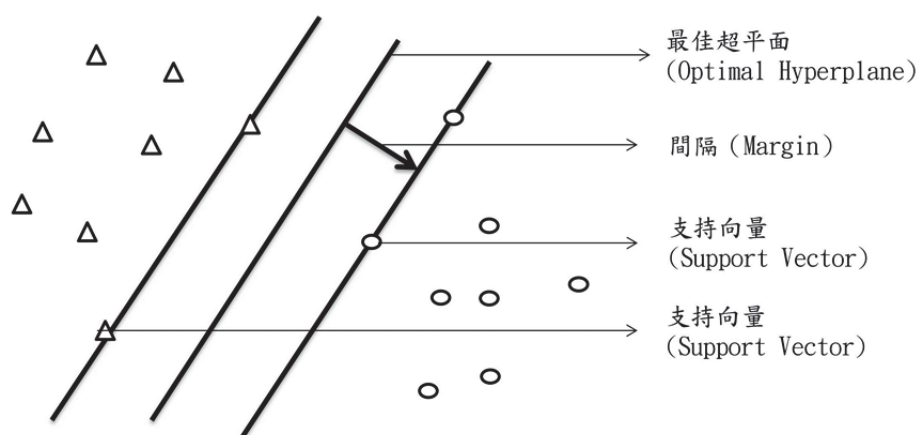


圖 3-3 支持向量機概念示意

設訓練資料表示為 $D = \{x_i, y_i\}_{i=1}^n \mid \forall x_i = (x_i^{(1)}, \dots, x_i^{(n)}) \in R^n \mid y_i \in \{+1, -1\}^n$ ，超平面的數學形式可以表示為：

$$w \cdot x - b = 0$$

其中 x 是超平面上的點； w 是垂直於超平面的向量； b 為一常數。



如圖 3-4 最大間隔超平面，平行的並且離支持向量最近的超平面由以下方程族表示：

$$\begin{cases} w \cdot x - b = +1 \\ w \cdot x - b = -1 \end{cases}$$

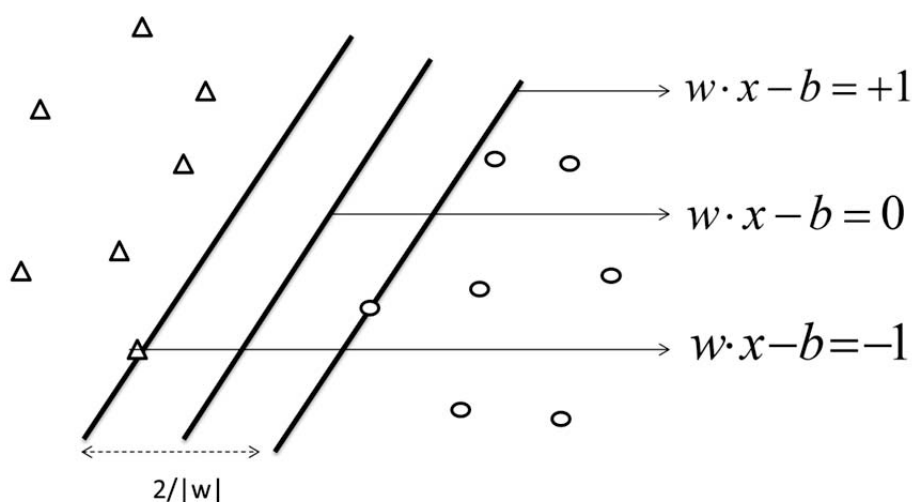


圖 3-4 最大間隔超平面

兩個超平面之間的距離是 $2/|w|$ ，目標求最小化 $|w|$ 。軟性間隔 (Soft margin) 中，透過加入鬆弛參數 ξ_i (Slack Variable) 以衡量對數據 x_i 的誤分類度，可以處理標記錯誤的訓練資料，區分正負例的超平面不存在時，則將選擇一個超平面儘可能清晰的區分資料，同時使其與分界最清晰的樣本的距離最大化，且同時對於所有的 i 滿足，目標函數可以表示為：

$$y_i(w \cdot x + b) \geq 1 - \xi_i \quad 1 \leq i \leq n$$

隨後，在增大間距和縮小懲罰值 (penalty value, C) 兩大目標之間進行權衡優化，其中 $C > 0$ ，如果懲罰值為線性函數，則表示為

$$\text{Min}_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i$$

$$\text{s.t. } y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, n$$

$$\xi \geq 0$$

為二次規劃最佳化問題 (Quadratic Programming Optimization Problem)，通過 LIBSVM (Chang & Lin, 2011) 求解，得超平面的分類判別函數為：

$$f(x) = \text{sgn} \left(\left[\sum_{i=1}^N \alpha_i y_i (x_i, x) + b \right] \right)$$

對於非線性問題，若核函數 $K(x_i, x_j)$ 滿足 Mercer 條件 $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ ，訓練向量 x_i 透過函數 ϕ 映射到高維度空間，在高維度空間找尋線性最佳超平面，即得高維空間的分類函數 (Yang et al., 2008) 為：

$$f(x) = \text{sgn} \left(\left[\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b \right] \right)$$

上式，將高維空間的內積計算轉換為低維度函數計算，此分類函數學習機稱為支持向量機，其中 $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ 為核心函數 (kernel function)，本研究選擇 RBF 為核心函數，表示為：

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad \gamma \in \mathfrak{R}^+$$



3.3 研究流程

3.3.1 支持向量機 (SVM) 分類

支持向量機將已標籤 (label) 成陽性及陰性的訓練資料集 (training data set) 做訓練建模, 其後, 以驗證集資料 (Validation data set) 或新資料集 (New data set) 依該模型透過 SVM 預測分類結果並分析驗證, 流程如圖 3-5。

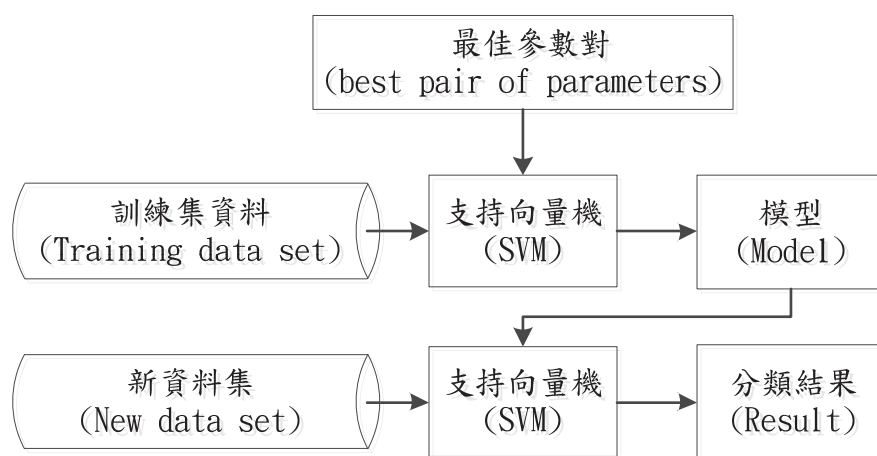


圖 3-5 支持向量機 (SVM) 分類流程

3.3.2 網格搜尋

由圖 3-5 所示, 以 SVM 訓練、建立模型須給予參數對 (C, γ), 為篩選出最佳參數對 (best pair of parameters), 本研究透過 10-交叉驗證 (10-fold cross validation) 進行網格搜尋 (grid-search) 進行測試參數對篩選, 操作流程如圖 3-6。

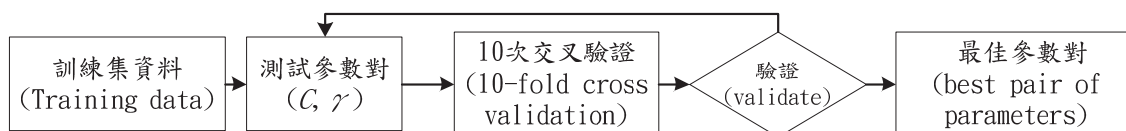


圖 3-6 網格搜尋流程

第四章 結果與討論



4.1 內梅羅指標 (P_N) 評析結果

本研究使用農試所檢測之農地重金屬資料採樣時間為民國 81 年至 97 年，以彰化縣、桃園市為研究區域，透過內梅羅指標 (P_N) 綜合計算六項重金屬 (Cd、Cr、Cu、Ni、Pb、Zn) 濃度資料，結合調查點位的空間座標 (X_i, Y_i)，藉助地理資訊系統 (geographic information system, GIS) 展示空間分布並依張等 (2012) 建議之五級分類呈現，所繪研究區域之 P_N 值空間分布，如圖 4-1、圖 4-2。

(一) 彰化縣

彰化縣評析結果如表 4-1 所示，25.22% (6,461 筆) 屬於「優良 ($P_N < 0.7$)」；21.25% (1,459 筆) 屬於「安全 ($0.7 \leq P_N < 1.0$)」；23.82% (1,954 筆) 屬於「警戒 ($1.0 \leq P_N < 2.0$)」；23.82% (248 筆) 屬於「污染 ($2.0 \leq P_N < 3.0$)」；最後 29.71% (249 筆) 屬於「危害 ($P_N \geq 3.0$)」。其中， $P_N \geq 1.0$ 佔該區筆數 23.63%，表示土壤污染物質超過背景值上限，作物生產可能受到重金屬污染。

(二) 桃園市

桃園市評析結果如表 4-2 所示，72.40% (6,520 筆) 屬於「優良 ($P_N < 0.7$)」；9.34% (841 筆) 屬於「安全 ($0.7 \leq P_N < 1.0$)」；9.46% (852 筆) 屬於「警戒 ($1.0 \leq P_N < 2.0$)」；2.89% (260 筆) 屬於「污染 ($2.0 \leq P_N < 3.0$)」；最後 5.91% (532 筆) 屬於「危害 ($P_N \geq 3.0$)」。其中土壤污染物質超過背景值上限 ($P_N \geq 1.0$) 佔該區 18.26%。



表 4-1 彰化縣 P_N 分級分布表

污染等級	P_N 分級	筆數	百分比
優良	$P_N < 0.7$	6,461	62.30
安全	$0.7 \leq P_N < 1.0$	1,459	14.07
警戒	$1.0 \leq P_N < 2.0$	1,954	18.84
污染	$2.0 \leq P_N < 3.0$	248	2.39
危害	$P_N \geq 3.0$	249	2.40
總數		10,371	100

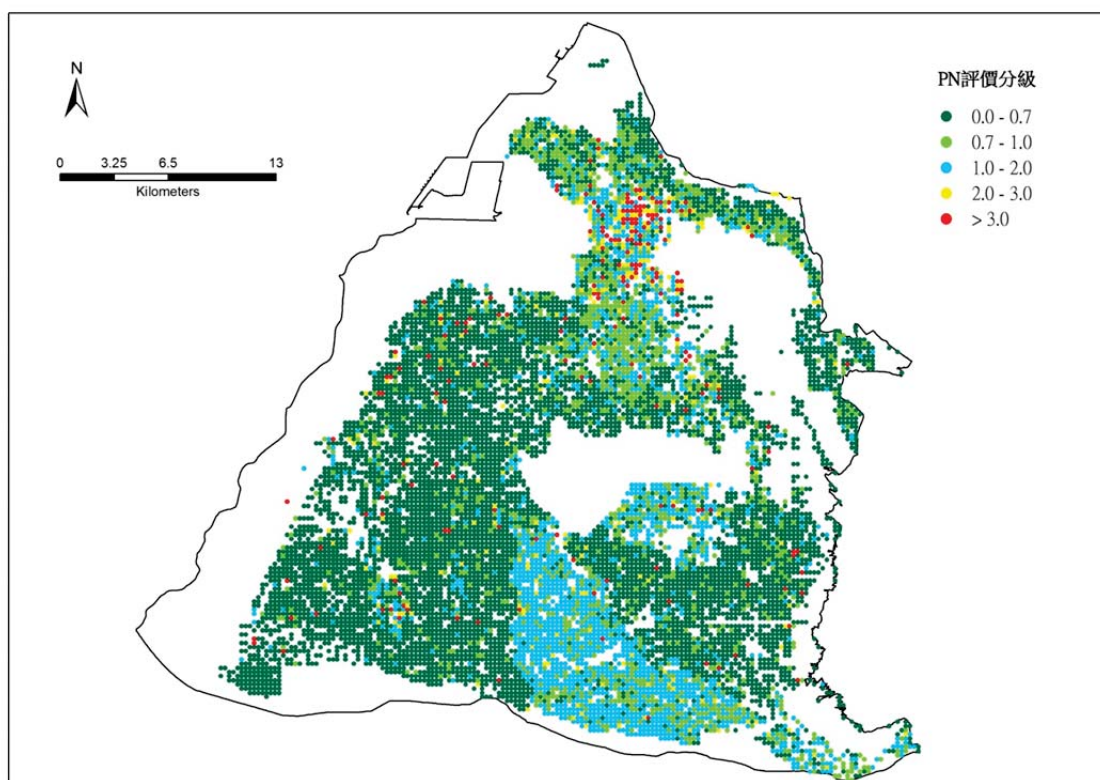


圖 4-1 彰化縣各採樣點 P_N 分布圖



表 4-2 桃園市 P_N 分級分布表

污染等級	P_N 分級	筆數	百分比
優良	$P_N < 0.7$	6,520	72.40
安全	$0.7 \leq P_N < 1.0$	841	9.34
警戒	$1.0 \leq P_N < 2.0$	852	9.46
污染	$2.0 \leq P_N < 3.0$	260	2.89
危害	$P_N \geq 3.0$	532	5.91
總數		9,005	100

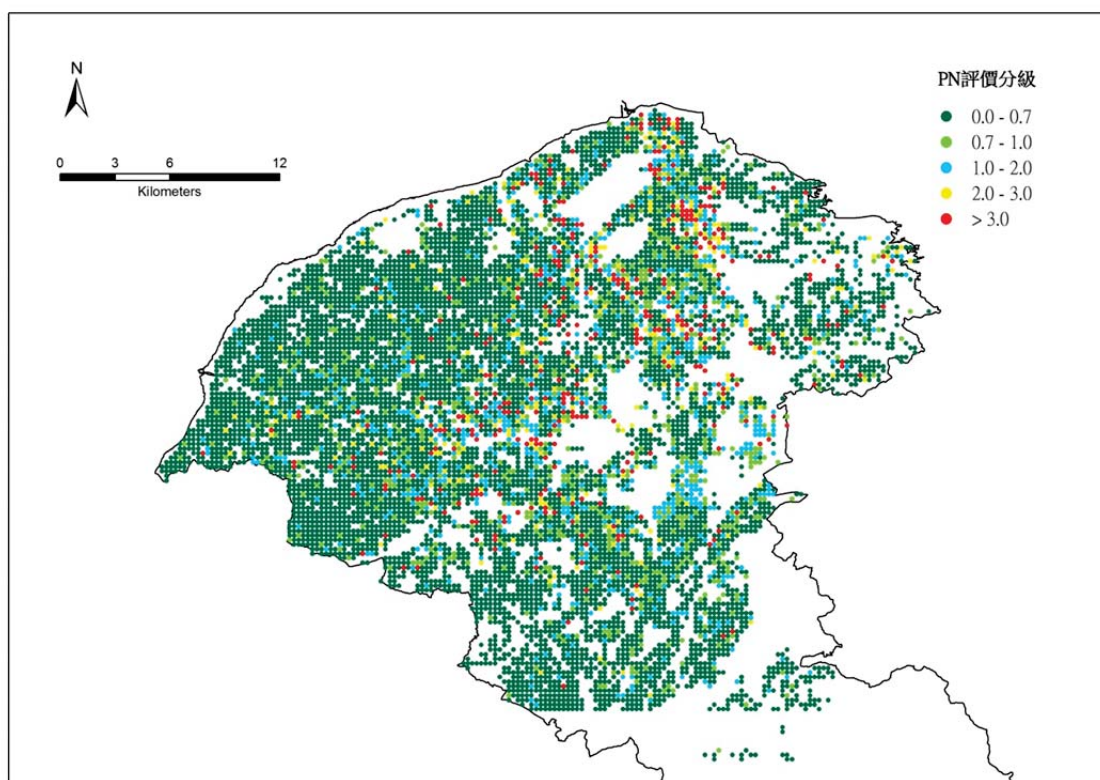


圖 4-2 桃園市各採樣點 P_N 分布圖



4.1.1 訓練集樣本資料標籤

(一) 彰化縣

支持向量機 (SVM) 需將訓練資料標籤後進行分類，在此將彰化縣所有採樣點 P_N ，以 1.0 作為閾值 (threshold value) 區分為兩類。透過 GIS 繪出，如圖 4-3。其中，2,451 筆屬於 $P_N \geq 1.0$ (佔 23.63%)，標籤為陽性 (positive)；以及 7,920 筆屬於 $P_N < 1.0$ (佔 76.37%)，標籤為陰性 (negative)，如表 4-3。

表 4-3 彰化縣 P_N 分級 (閾值設定為 1.0)

P_N 分級	標籤	筆數	百分比 (%)
$P_N \geq 1.0$	陽性	2,451	23.63
$P_N < 1.0$	陰性	7,920	76.37
總數		10,371	100

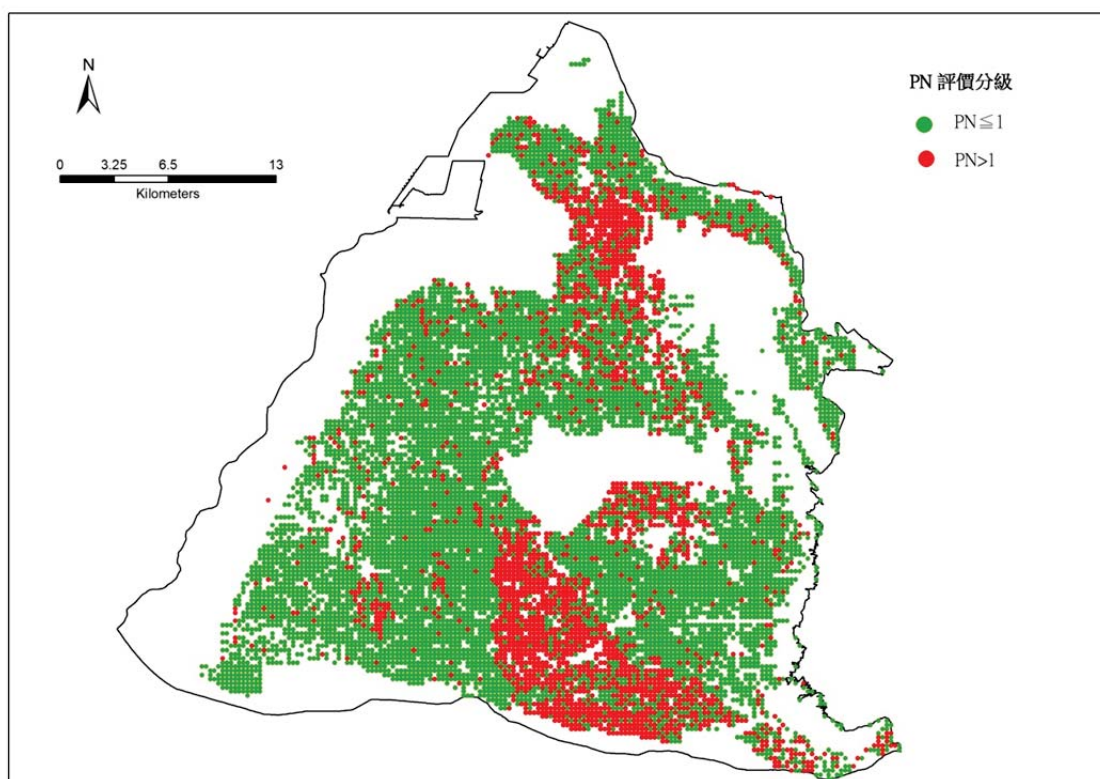
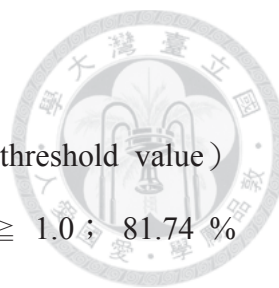


圖 4-3 彰化縣閾值為 1.0 之各採樣點 P_N



(二) 桃園市

首先，同樣地將桃園市所有採樣點以 P_N 為 1.0 作為閾值 (threshold value) 區分為兩類，如圖 4-4。其中，18.26 % (1,644 筆) 屬於 $P_N \geq 1.0$ ；81.74 % (7,361 筆) 屬於 $P_N < 1.0$ ，如表 4-4。

表 4-4 桃園市 P_N 分級 (閾值設定為 1.0)

P_N 分級	標籤	筆數	百分比 (%)
$P_N \geq 1.0$	陽性	1,644	18.26
$P_N < 1.0$	陰性	7,361	81.74
總數		9,005	100

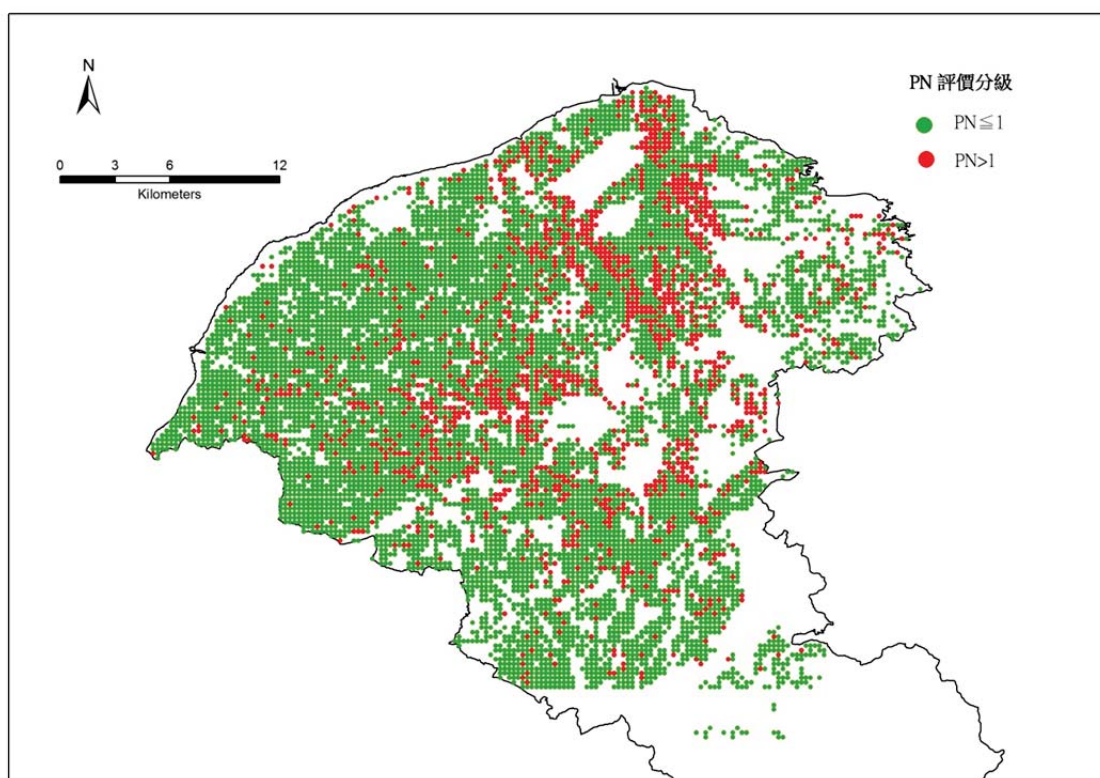


圖 4-4 桃園市閾值為 1.0 之各採樣點 P_N



4.2 訓練集標籤組成比例探討

本節固定樣本中標籤為陽性(positive)的資料筆數，改變標籤為陰性(negative)的資料選取量，以不同比例隨機抽取數量組成訓練集(training data set)，並以該區域總資料點測試模型(model)效能，探討訓練集樣本組成比例對測試結果的影響。彰化縣、桃園市的核心函數參數對 (C, γ) 分別設為定值 $(0.5, 256.0)$ 及 $(0.125, 256.0)$ 。測試結果分別如表 4-5、表 4-6 所示。檢測結果顯示，在當訓練集樣本標籤陽性、陰性比，較理想之結果發生於，彰化縣陽性、陰性比為 1:2 時 F1-measure 為 0.692；桃園市陽性、陰性比為 1:1 時，F1-measure 為 0.497。

表 4-5 彰化縣樣本比例實驗結果

	1 : 1	1 : 2	1 : 3
P : Positive	2451	2451	2451
N : Negative	2451	4902	7353
True positive	1924	1706	1444
False negative	527	745	1007
True negative	6524	7148	7421
False positive	1396	772	499
準確度 (Accuracy)	81.46%	85.37%	85.48%
召回率 (Recall)	78.50%	69.60%	58.91%
精確度 (Precision)	57.95%	68.85%	74.32%
F1-measure	0.667	0.692	0.657

表中灰底為 F1-measure 最高項

表 4-6 桃園市樣本比例實驗結果

	1 : 1	1 : 2	1 : 3	1 : 4
P : Positive	1644	1644	1644	1644
N : Negative	1644	3288	4932	6576
True positive	1310	486	292	0
False negative	334	1158	1352	1644
True negative	5136	6961	7172	7361
False positive	2225	400	189	0
準確度 (Accuracy)	71.58%	82.70%	82.89%	81.74%
召回率 (Recall)	79.68%	29.56%	17.76%	0.00%
精確度 (Precision)	37.06%	54.85%	60.71%	0.00%
F1-measure	0.506	0.384	0.275	0.000

表中灰底為 F1-measure 最高項



4.3 訓練集樣本數量結果探討

本節探討訓練集樣本數量變化對於結果的影響。根據 4.2 結果，訓練集樣本數量陽性、陰性比例於彰化縣固定為 1：2，而桃園市固定為 1：1。

以各採樣點座標 (X_i, Y_i) 作為訓練之索引類別，選定 RBF 為核心函數，依序取不同比例隨機抽取數量的訓練資料集樣本 (10%, 20%, ..., 100% 訓練集)，透過網格搜尋 (grid-search) 搭配 10-交叉驗證 (10-fold cross validation)，篩選最佳參數對 (best pair of parameters)，以總資料集 (Total data set) 預測，探討其準確度 (Accuracy)、召回率 (Recall)、精確度 (Precision)、F1-measure 及污染潛勢範圍等結果。

彰化縣、桃園市訓練集數量與預測結果分別如表 4-7 及表 4-8；訓練集數量與網格搜尋最佳參數對落在圖中綠色區間，驗證及預測結果如表分別如圖 4-5、圖 4-7；訓練集數量與高污染潛勢圖分別如圖 4-6、圖 4-8 所示。結果顯示，當訓練集樣本數量越高，模型預測結果 F1-measure 越高，代表 SVM 的分類效能隨學習樣本數增多而提高。

彰化縣以 7,353 筆點位在陽性、陰性標籤比 1：2 下訓練建立之模型進行土壤重金屬污染潛勢預測，並以總資料測試結果彰化縣高污染潛勢點位數為 2,478 筆 (佔 23.89%)，準確度 (Accuracy) 為 85.37%、F1-measure 為 0.692，主要聚集分布於北彰化和美鄉、南舊濁水溪流域上游。

桃園市在標籤比為 1：1 下，以 3,288 筆資料訓練建立模型，並以總資料測試結果高污染潛勢點位數為 3,535 筆 (佔 39.25%)，準確度為 71.58%、F1-measure 為 0.506，聚塊分布散於桃園區域內。

且在小樣本數 10% 訓練時彰化縣、桃園市 F1-measure 分別達 0.611 及 0.421，與區域各別總訓練資料量的 F1-measure 相比，表示此研究案例的 SVM 在低訓練集樣本數即可達良好的分類效能。

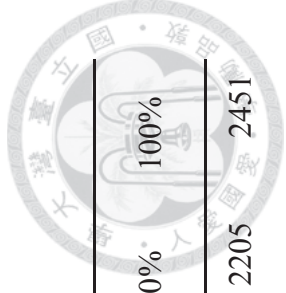
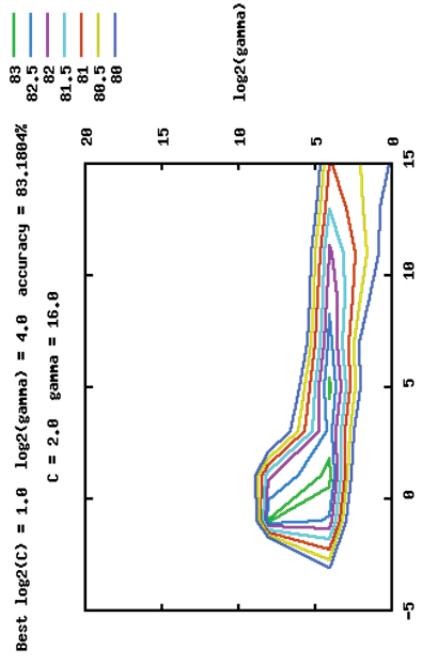


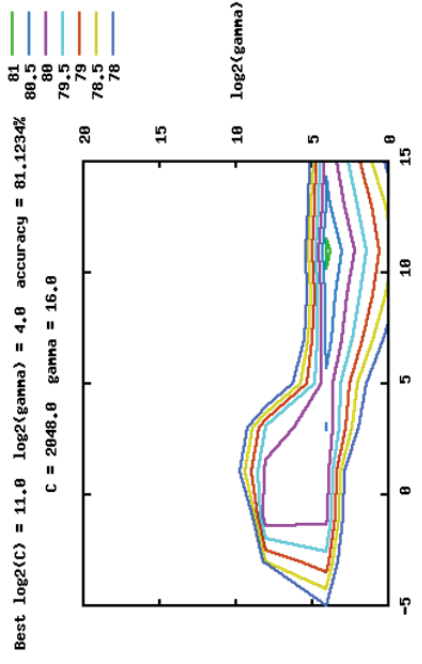
表 4-7 彰化縣不同訓練集數量對應預測結果

樣本數量	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
	P : Positive ($P_N \geq 1.0$)	245	490	735	980	1225	1470	1715	1960	2205
N : Negative ($P_N < 1.0$)	490	980	1470	1961	2451	2941	3432	3922	4412	4902
最佳參數對	C	0.5	2.0	8.0	2048.0	8192.0	0.5	8192.0	2.0	2.0
	γ	16.0	16.0	16.0	16.0	16.0	16.0	256.0	256.0	256.0
True positive		1519	1441	1498	1636	1614	1585	1583	1666	1718
False negative		932	1010	953	815	837	866	868	785	745
True negative		6921	7215	7118	6731	7098	7235	6769	7170	7086
False positive		999	705	802	1189	822	685	1151	750	834
準確度 (Accuracy)		81.38%	83.46%	83.08%	80.68%	84.00%	85.04%	80.53%	85.20%	84.89%
召回率 (Recall)		61.97%	58.79%	61.12%	66.75%	65.85%	64.67%	64.59%	67.97%	70.09%
精確度 (Precision)		60.33%	67.15%	65.13%	57.91%	66.26%	69.82%	57.90%	68.96%	67.32%
F1-measure		0.611	0.627	0.631	0.620	0.661	0.671	0.611	0.685	0.687

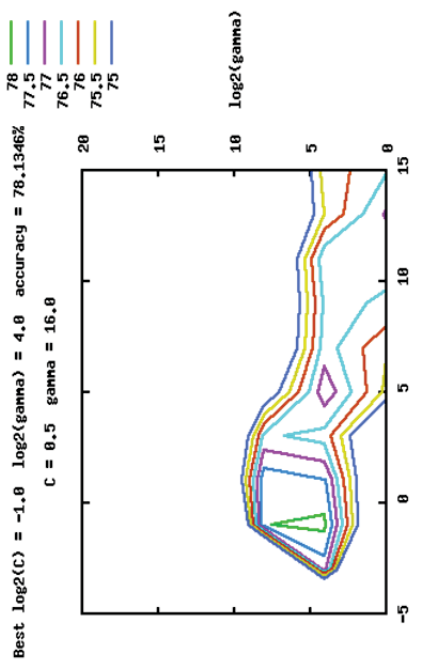
訓練集：7,353 筆；預測集：10,371 筆



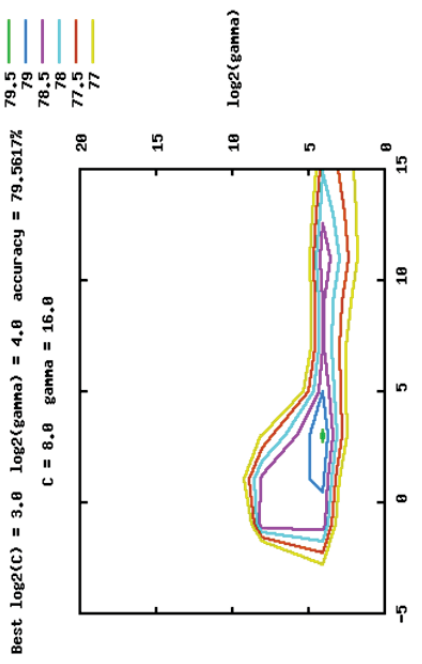
(b) 20% 訓練集



(d) 40% 訓練集

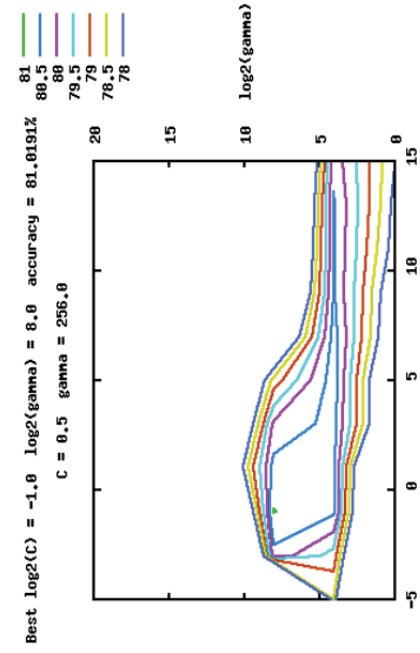


(a) 10% 訓練集

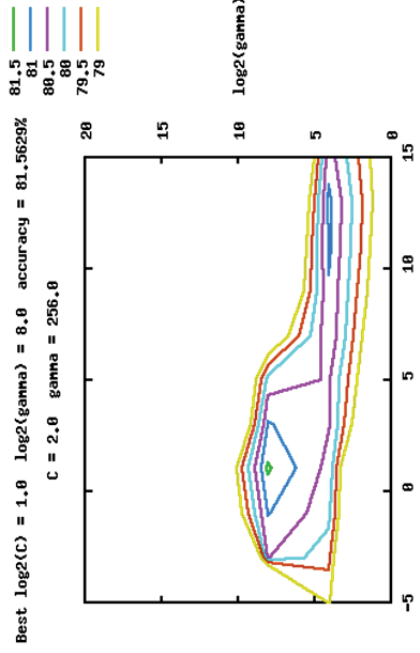


(c) 30% 訓練集

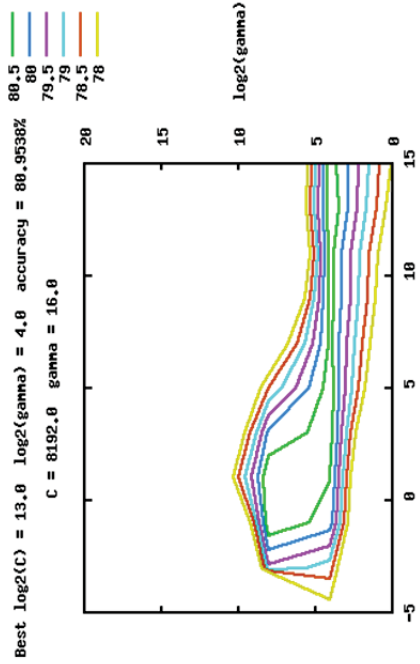
圖 4-5 彰化縣不同訓練集數量與網格搜尋結果



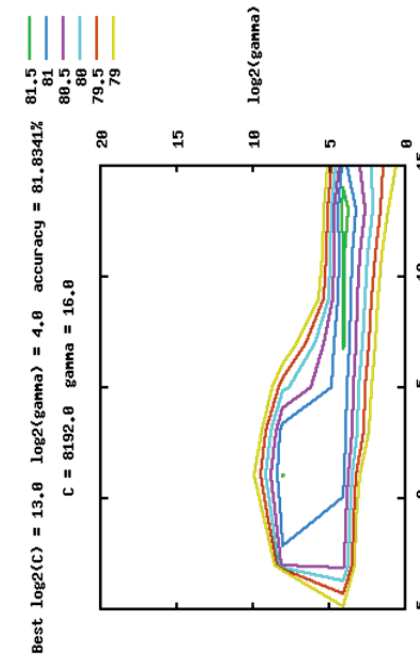
(f) 60% 訓練集



(h) 80% 訓練集

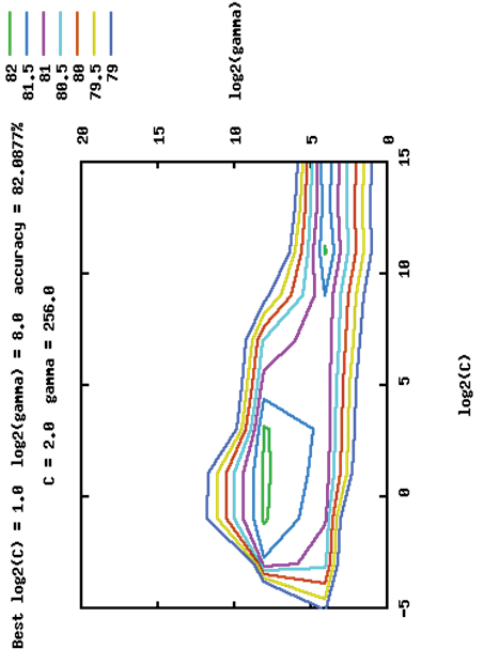


(e) 50% 訓練集

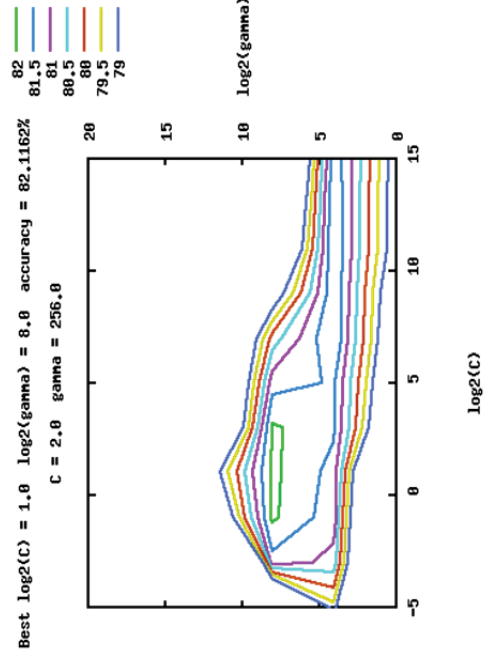


(g) 70% 訓練集

圖 4-5 彰化縣不同訓練集數量與網格搜尋結果 (續)

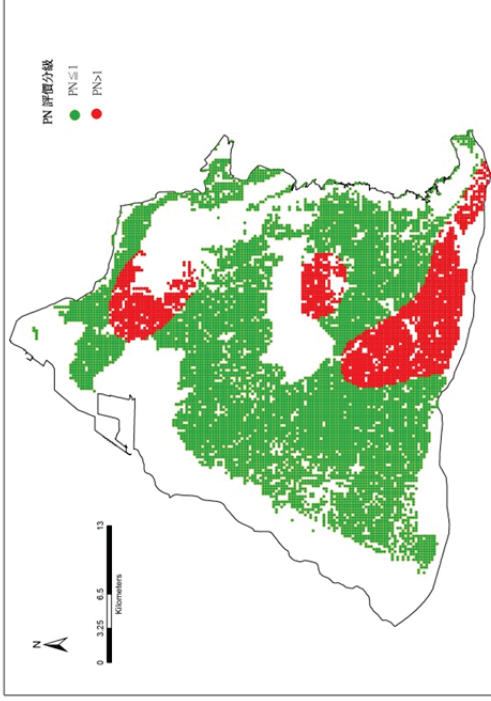


(j) 100 % 訓練集

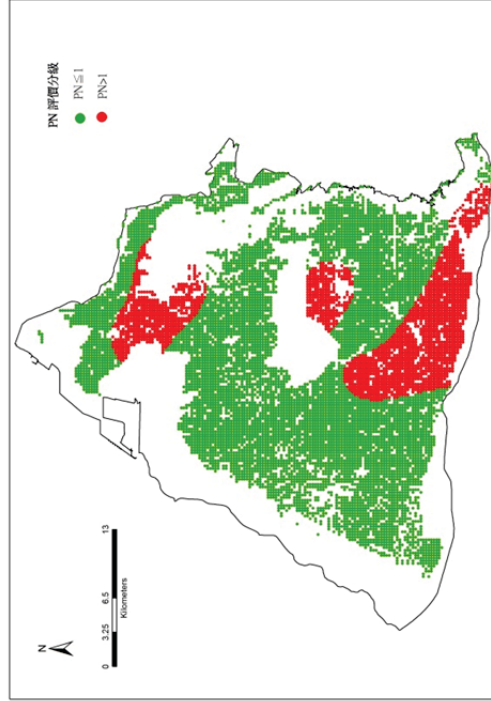


(i) 90 % 訓練集

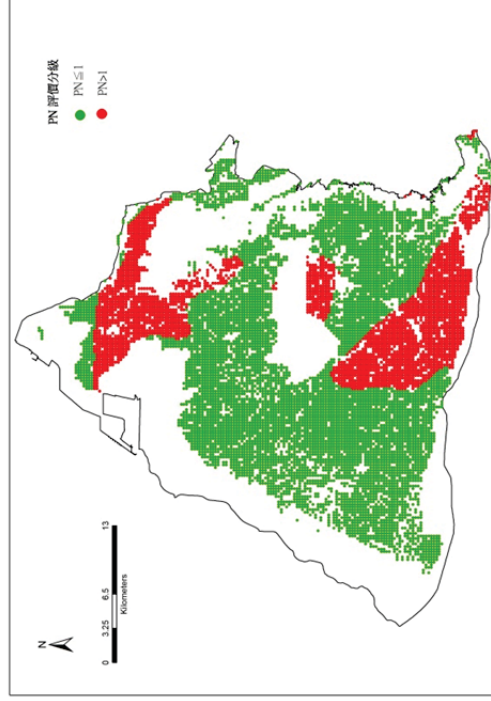
圖 4-5 彰化縣不同訓練集數量與網格搜尋結果 (續)



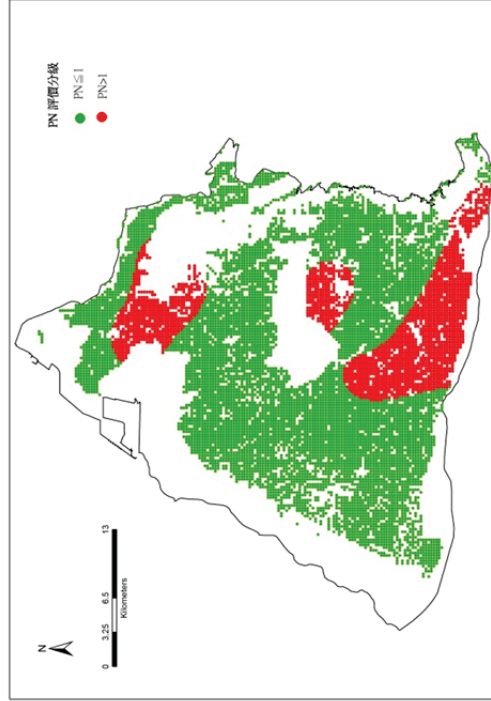
(a) 10% 訓練集



(b) 20% 訓練集

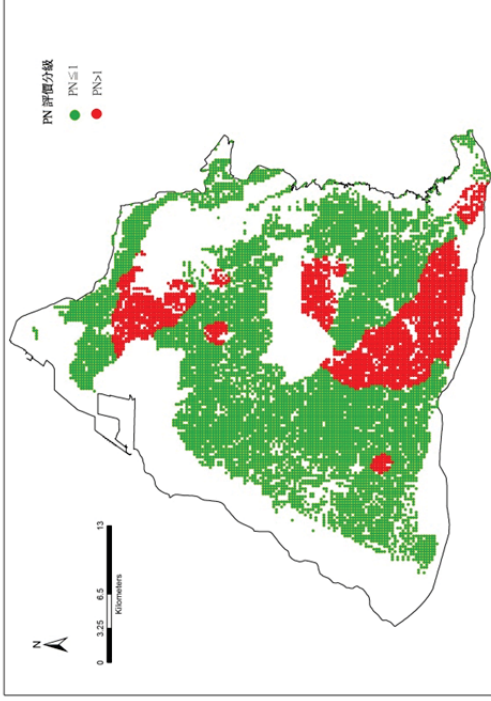


(c) 30% 訓練集

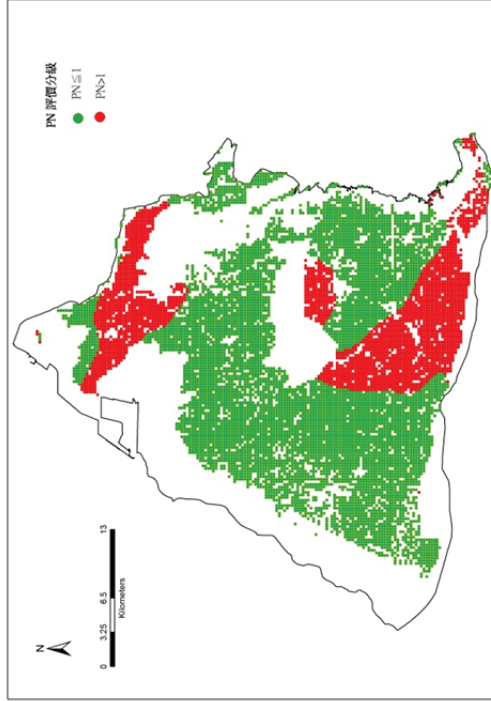


(d) 40% 訓練集

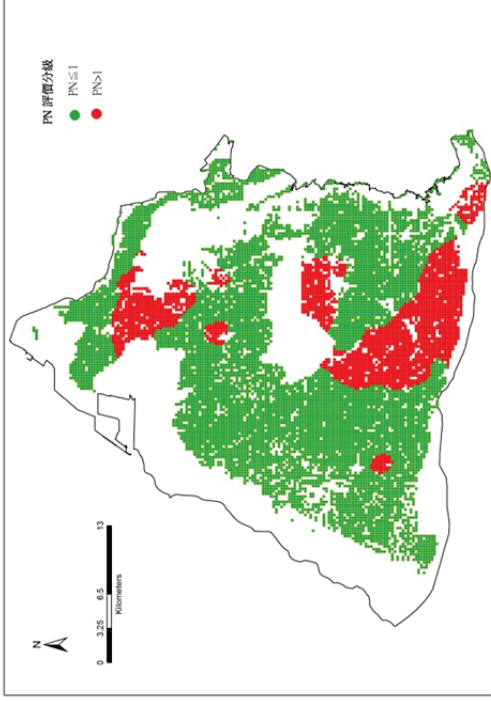
圖 4-6 彰化縣不同訓練集數量與高污染潛勢分布界定



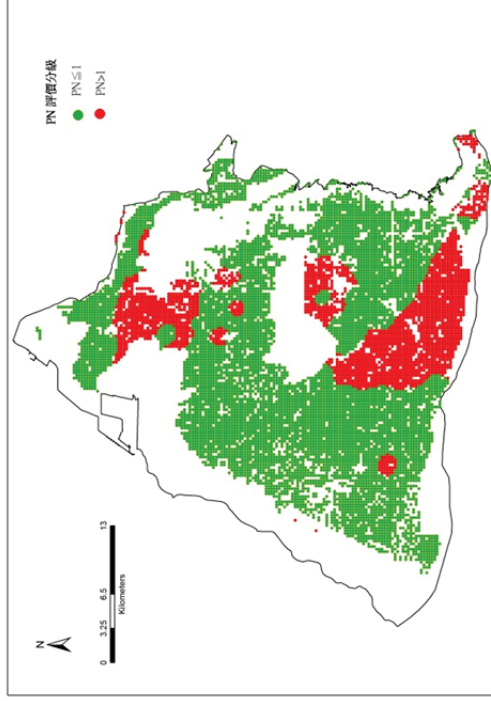
(e) 50% 訓練集



(g) 70% 訓練集

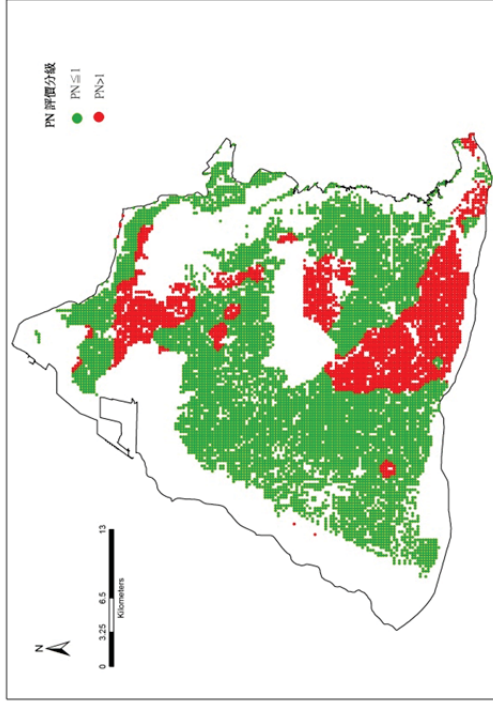


(f) 60% 訓練集

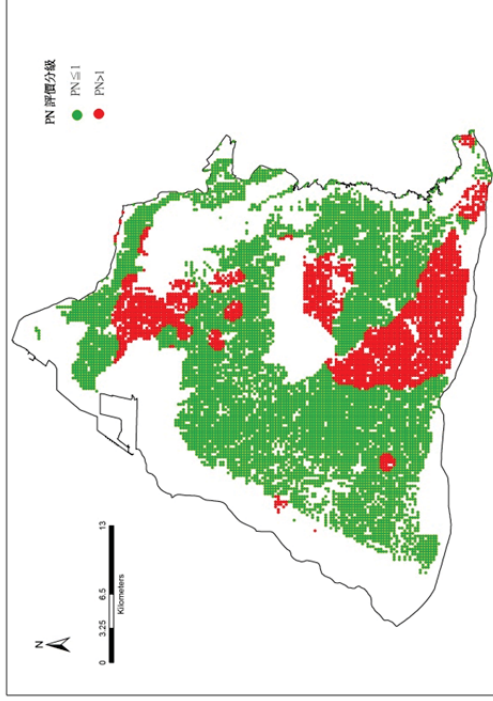


(h) 80% 訓練集

圖 4-6 彰化縣不同訓練集數量與高污染潛勢分布界定 (續)



(i) 90 % 訓練集



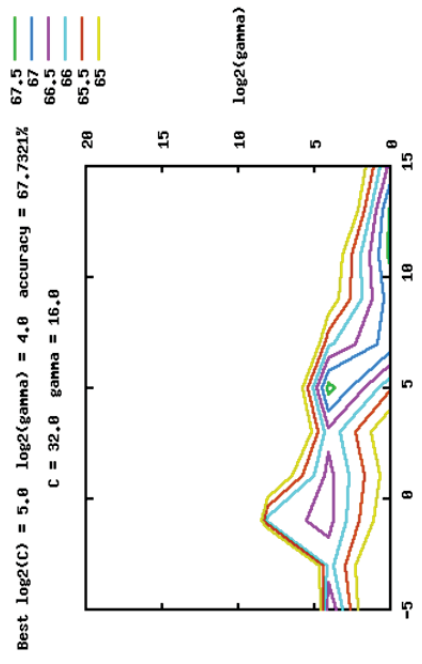
(j) 100 % 訓練集

圖 4-6 彰化縣不同訓練集數量與高污染潛勢分布界定 (續)

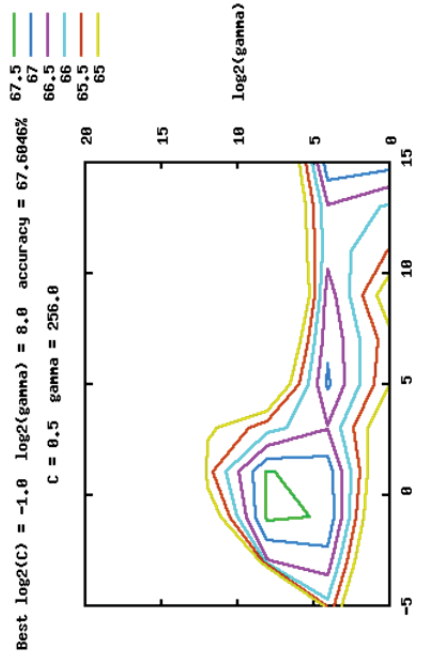
表 4-8 桃園市不同訓練集數量對應預測結果

樣本數量	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
	P : Positive ($P_N \geq 1.0$)	164	328	493	657	822	986	1150	1315	1479
N : Negative ($P_N < 1.0$)	164	329	493	658	822	986	1151	1315	1480	1644
最佳參數對	C	0.5	32.0	32.0	0.5	2.0	2.0	8.0	2.0	2.0
	γ	16.0	16.0	1.0	256.0	16.0	16.0	256.0	256.0	256.0
True positive	1191	1099	1253	1081	1185	1249	1210	1181	1212	1310
False negative	453	545	391	563	459	395	434	463	432	334
True negative	4535	4520	4196	5315	4762	4985	4606	4962	5226	5136
False positive	2826	2841	3165	2046	2599	2376	2755	2399	2135	2225
準確度 (Accuracy)	63.59%	62.40%	60.51%	71.03%	66.04%	69.23%	64.59%	68.22%	71.49%	71.58%
召回率 (Recall)	72.45%	66.85%	76.22%	65.75%	72.08%	75.97%	73.60%	71.84%	73.72%	79.68%
精確度 (Precision)	29.65%	27.89%	28.36%	34.57%	31.32%	34.46%	30.52%	32.99%	36.21%	37.06%
F1-measure	0.421	0.394	0.413	0.453	0.437	0.474	0.431	0.452	0.486	0.506

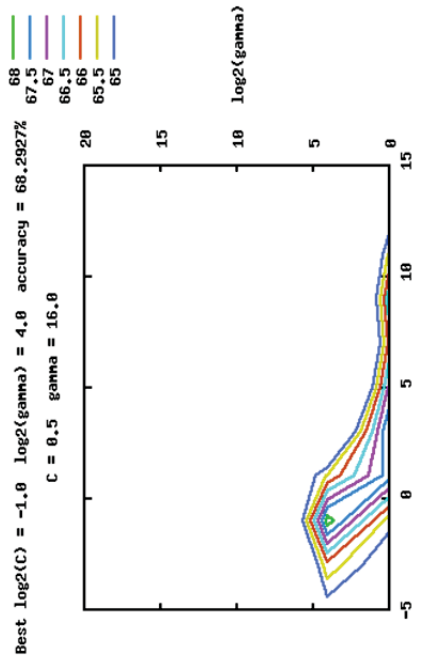
訓練集：3,288 筆；預測集：9,005 筆



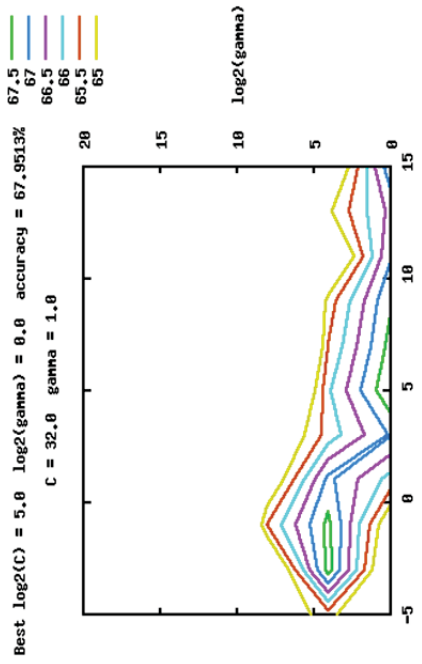
(b) 20% 訓練集



(d) 40% 訓練集

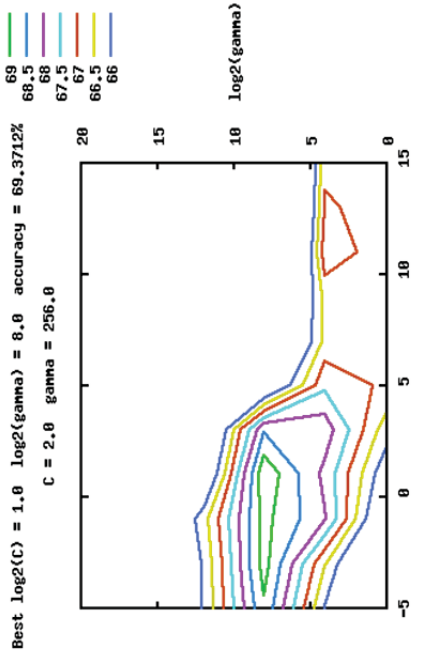


(a) 10% 訓練集

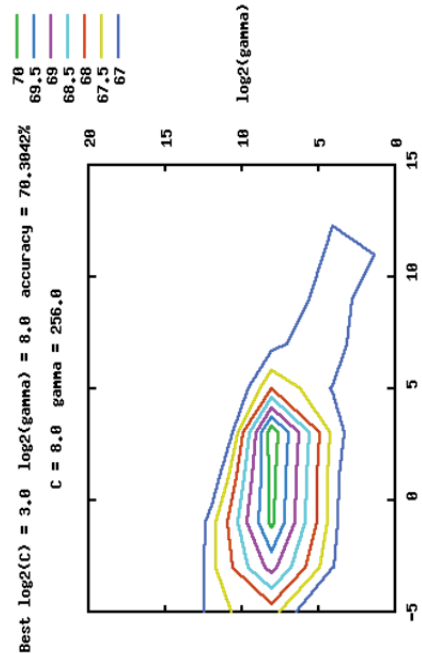


(c) 30% 訓練集

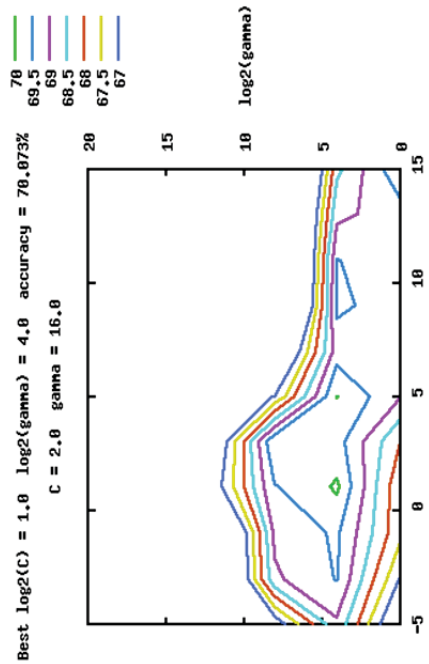
圖 4-7 桃園市不同訓練集數量與網格搜尋結果



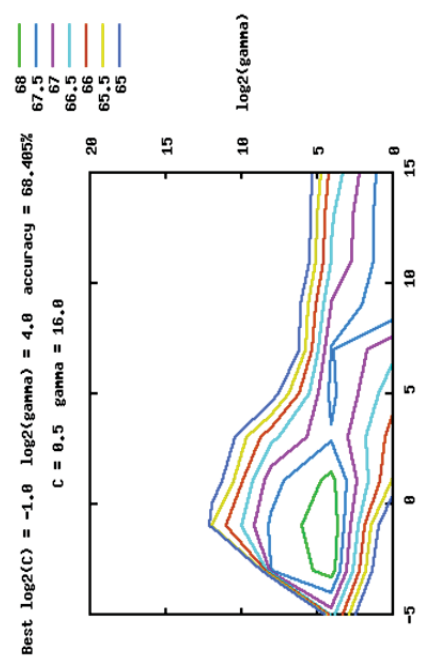
(f) 60% 訓練集



(h) 80% 訓練集

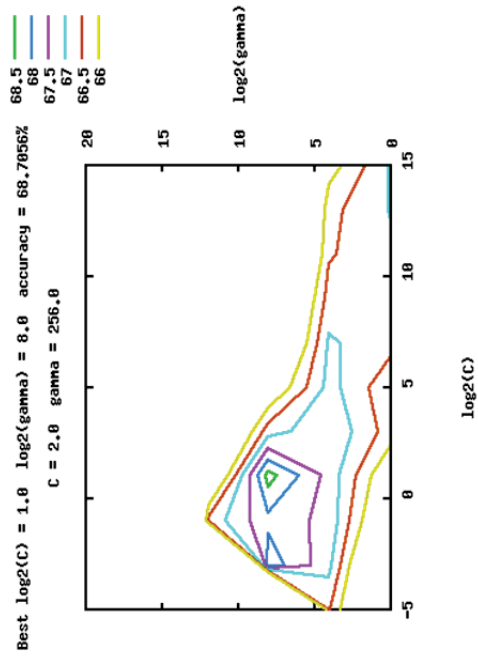


(e) 50% 訓練集

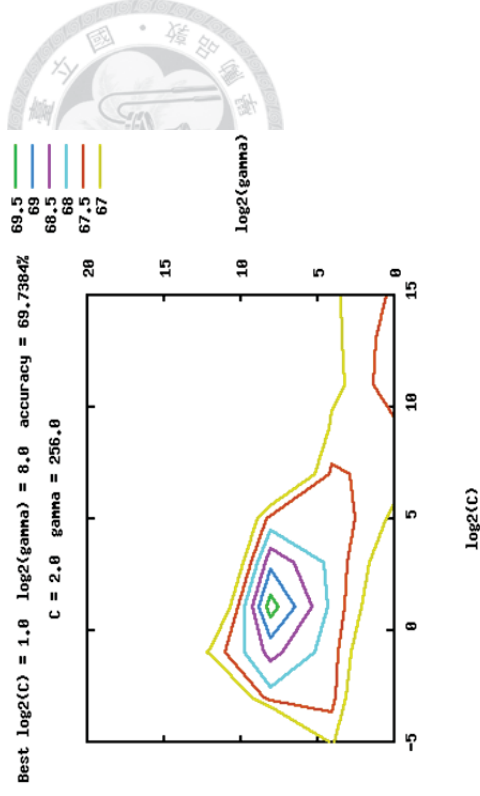


(g) 70% 訓練集

圖 4-7 桃園市不同訓練集數量與網格搜尋結果 (續)

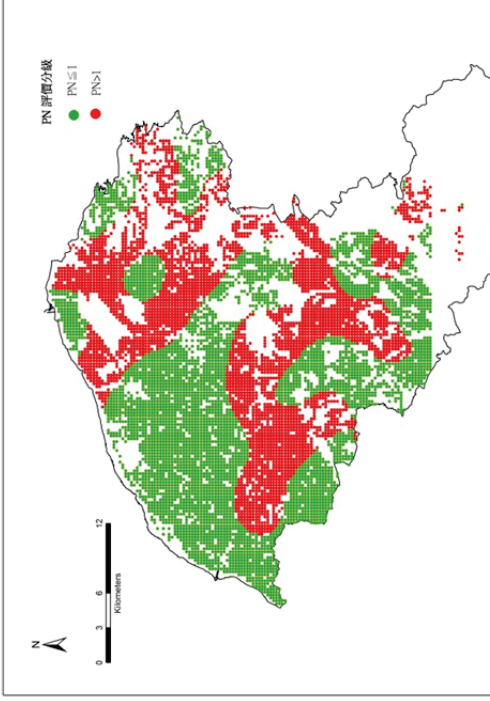


(i) 90 % 訓練集

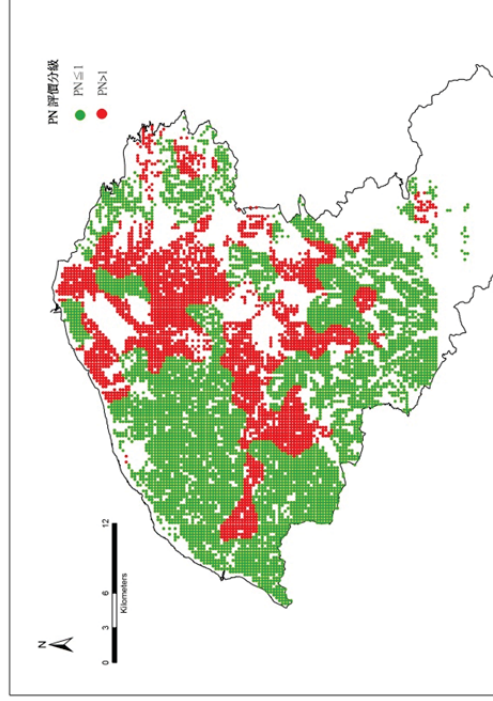


(j) 100 % 訓練集

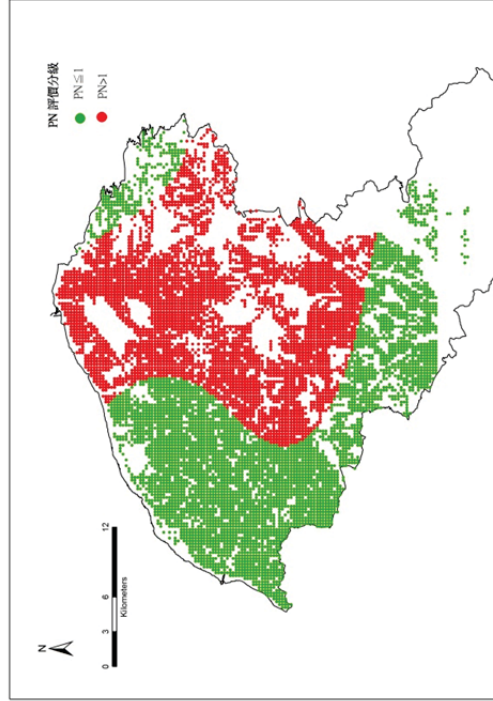
圖 4-7 桃園市不同訓練集數量與網格搜尋結果 (續)



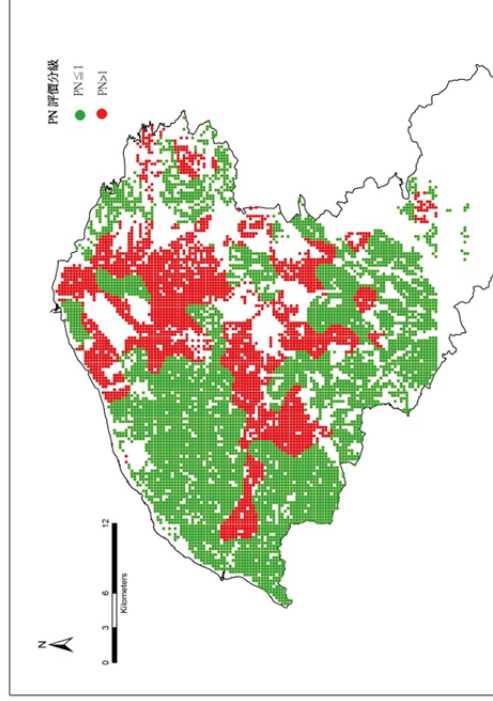
(a) 10% 訓練集



(b) 20% 訓練集

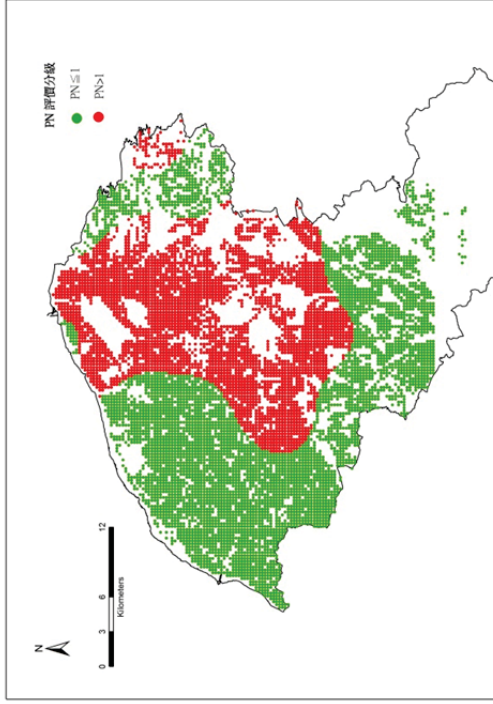


(c) 30% 訓練集

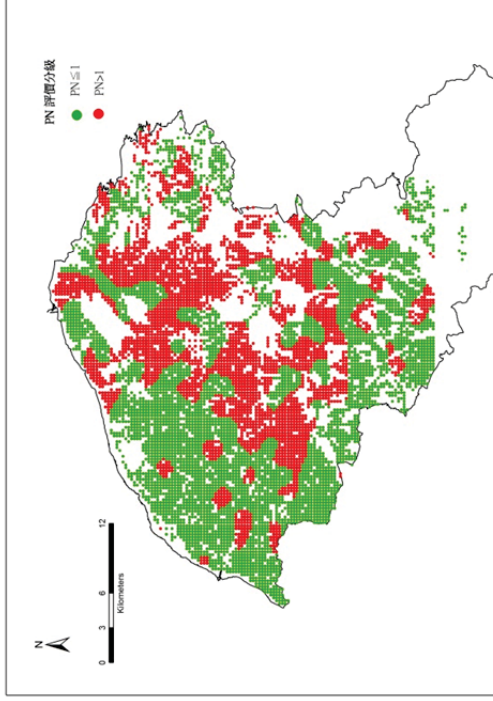


(d) 40% 訓練集

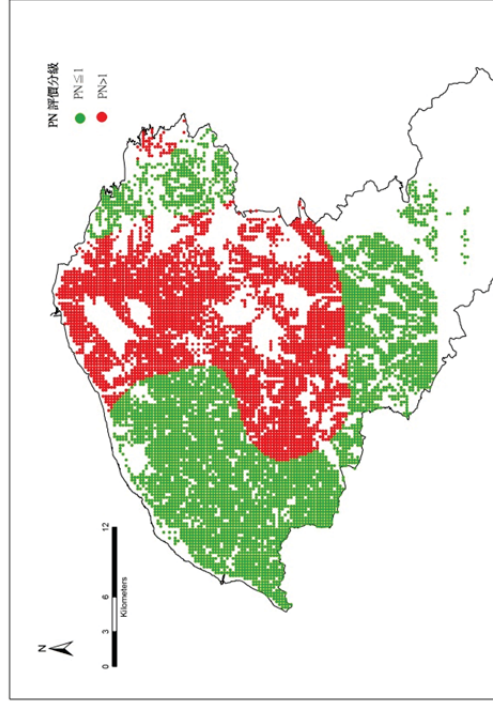
圖 4-8 桃園市不同訓練集數量與高污染潛勢分布界定



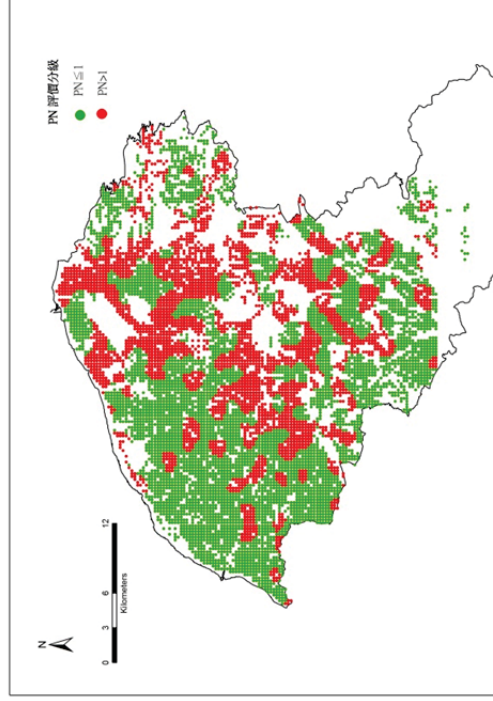
(e) 50% 訓練集



(f) 60% 訓練集

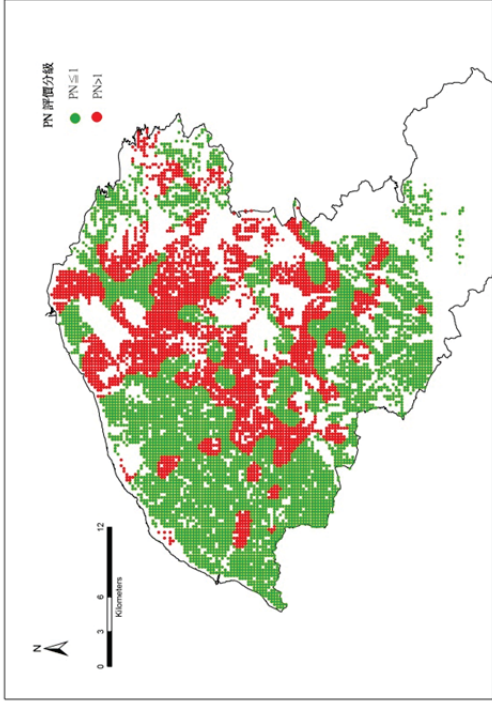


(g) 70% 訓練集

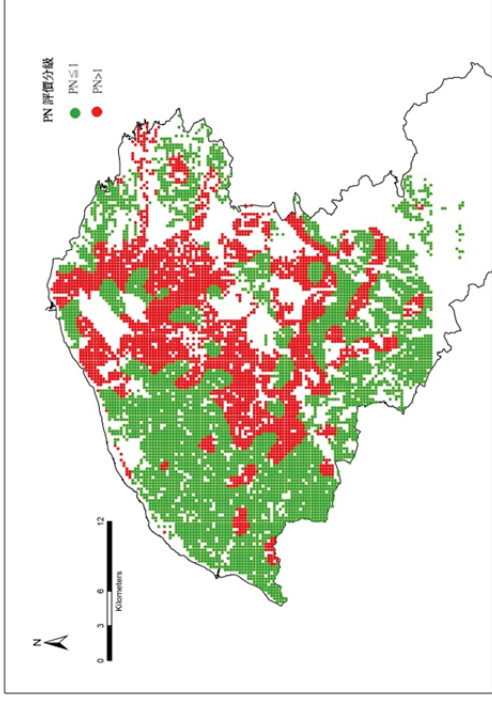


(h) 80% 訓練集

圖 4-8 桃園市不同訓練集數量與高污染潛勢分布界定 (續)



(i) 90 % 訓練集



(j) 100 % 訓練集

圖 4-8 桃園市不同訓練集數量與高污染潛勢分布界定 (續)



4.4 農地重金屬污染關聯性評析

本研究應用 GIS 技術繪製污染空間地圖，比對污染的空間流佈情形，以空間相關性來呈現多方資料的關聯性分析。藉彙整 SVM 區分農地重金屬高污染潛勢區之空間與屬性資料，套疊河川流域、工廠、工業區的空間分布，掌握各行業之污染特徵（表 4-9）、空間分布，綜合分析主要污染之重金屬項目，並解析其污染成因，如圖 4-9、圖 4-10。

分析結果顯示，高污染潛勢之區域多數非偶發場址，多處工廠、工業區周邊具高污染潛勢，污染隨渠道流動傳輸擴散的現象，其之間確實存在關聯性，唯張（2011）提醒，以推估圖套疊工廠，雖可推測污染來源，但農地重金屬污染具有累積性問題，工廠會開業停業，新工廠登機可以精確定位，但卻非污染的直接來源，舊工廠可能停業、轉業，從環境資料庫裡消失，但卻是真正的污染來源，以環境資料庫解釋問題有一定的風險性，需定期更新資料庫與現地調查深入了解問題來源。另外，與過往類似的研究比較，將篩選之危害及污染等級的水利小組範圍(洪等,2013)及農田水利會篩選出第 III 級圳路具污染潛勢區域(農委會,2014)，與本研究結果套疊進行比對。發現本研究挑選範圍為 $P_N \geq 1$ ，故範圍會比水利小組篩選評析較大一些，提供更多未來可能受影響的範圍，如圖 4-11、圖 4-12，而以灌溉系統顯示污染潛勢區域沒有分的很細密，如污染圳路可能為整條圳路或支線，若搭配本研究結果更能清楚的了解受污染影響範圍，如圖 4-13、圖 4-14。

表 4-9 行業類別及其重金屬污染種類

行業類別	主要污染重金屬種類
製革業	Cd, Cr, Cu, Pb, Zn
金屬基本工業	As, Cd, Cr, Cu, Hg, Ni, Pb, Zn
金屬表面處理業、電鍍業	Cd, Cr, Cu, Hg, Ni, Pb, Zn
晶圓製造及半導體製造業	As, Cd, Cr, Cu, Hg, Ni, Pb

整理自土污法第八、九條；行政院；張和鄭（2012）

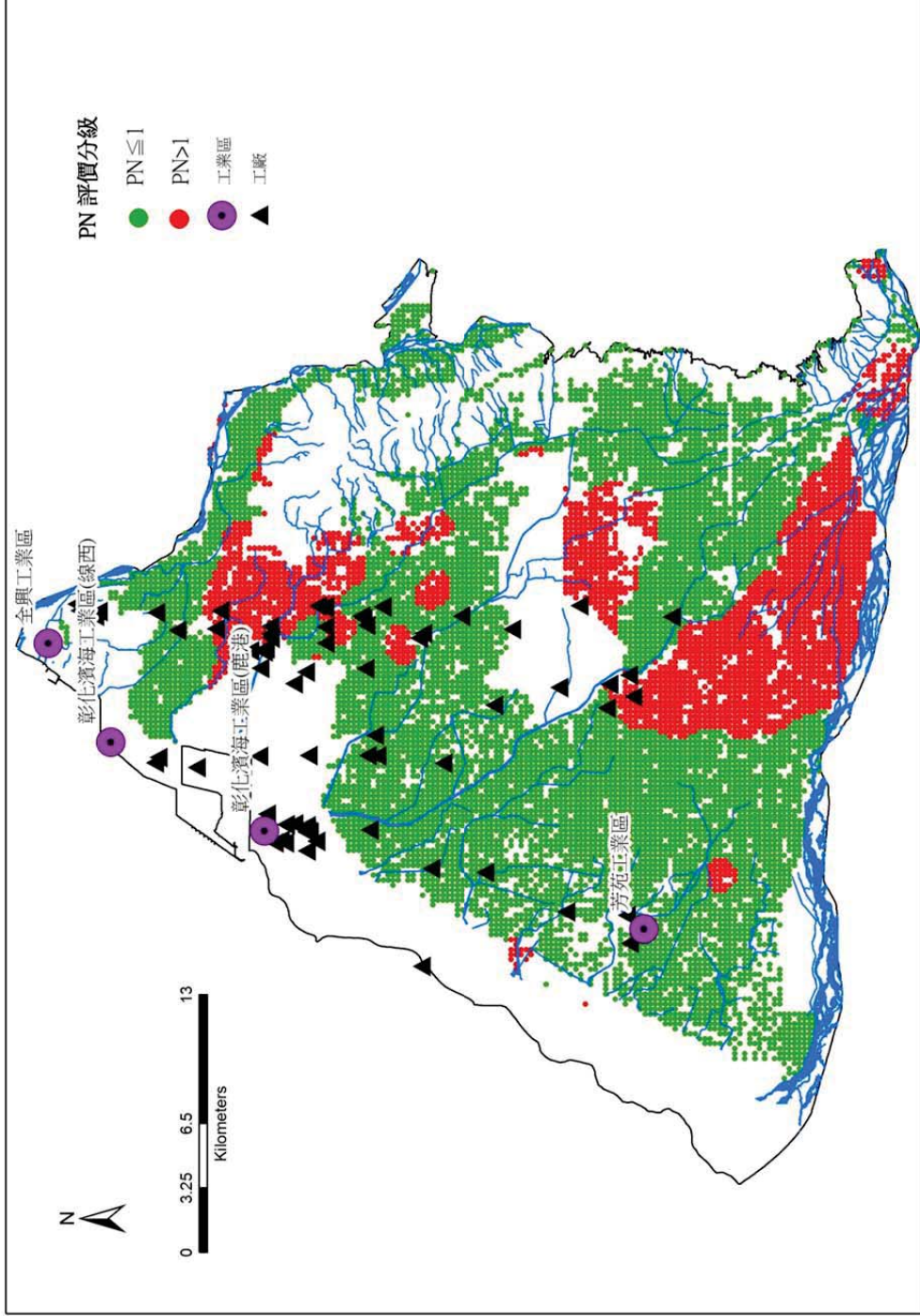


圖 4-9 彰化縣高污染潛勢區綜合研析圖

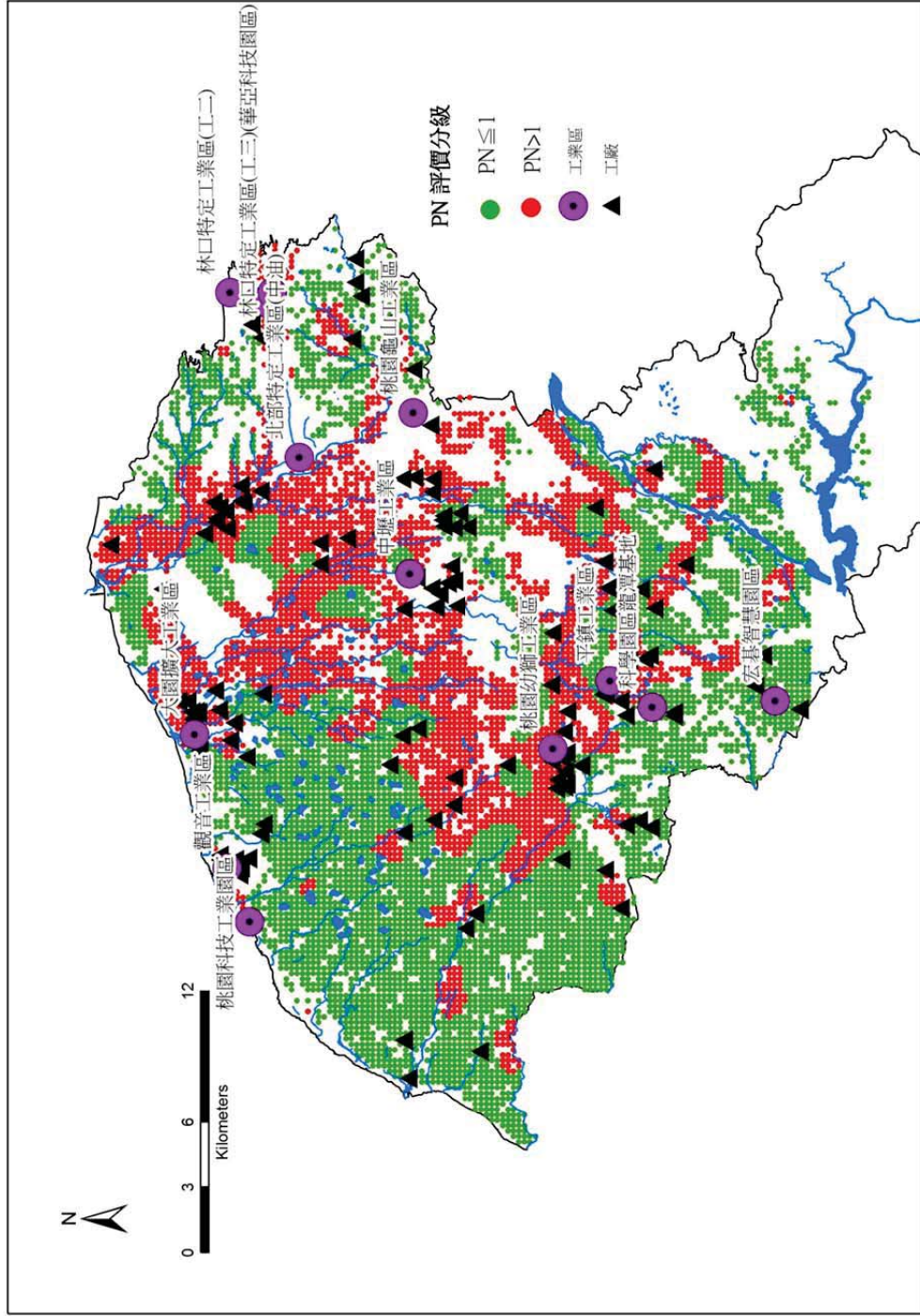


圖 4-10 桃園市高污染潛勢區綜合研析圖

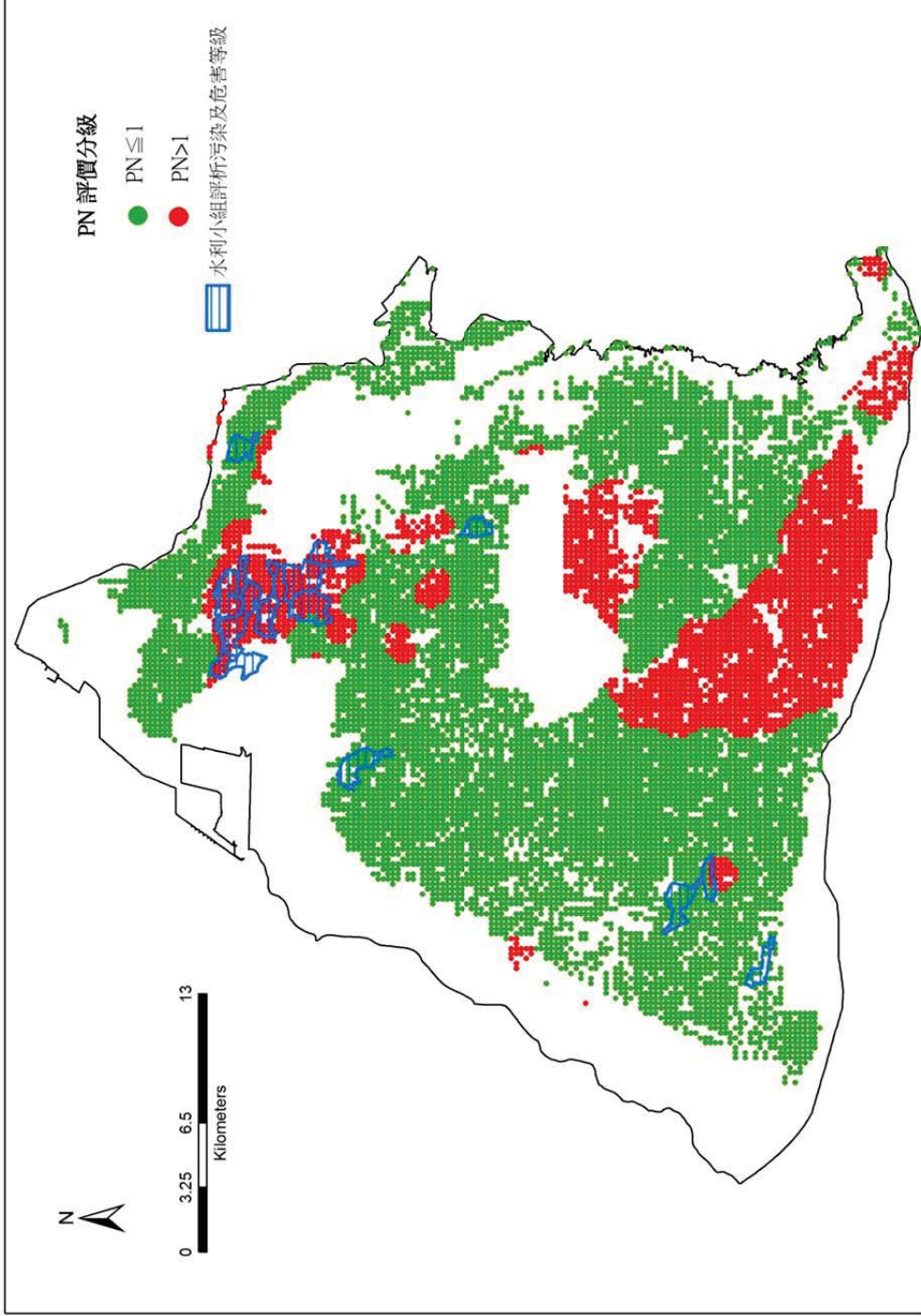


圖 4-11 彰化縣水利小組評析污染及危害等級套疊 SVM 的預測結果

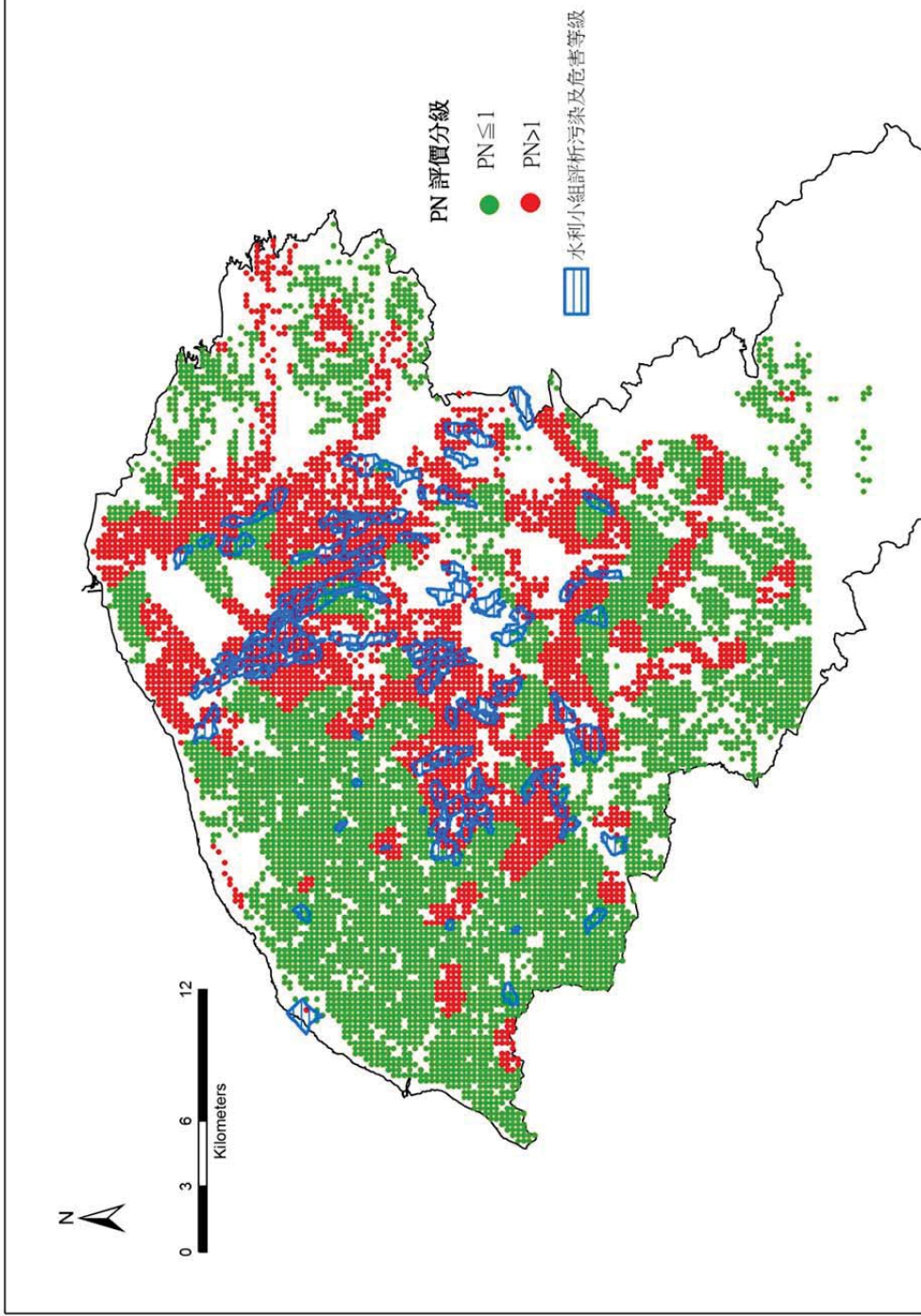


圖 4-12 桃園市水利小組評析污染及危害等級套疊 SVM 的預測結果

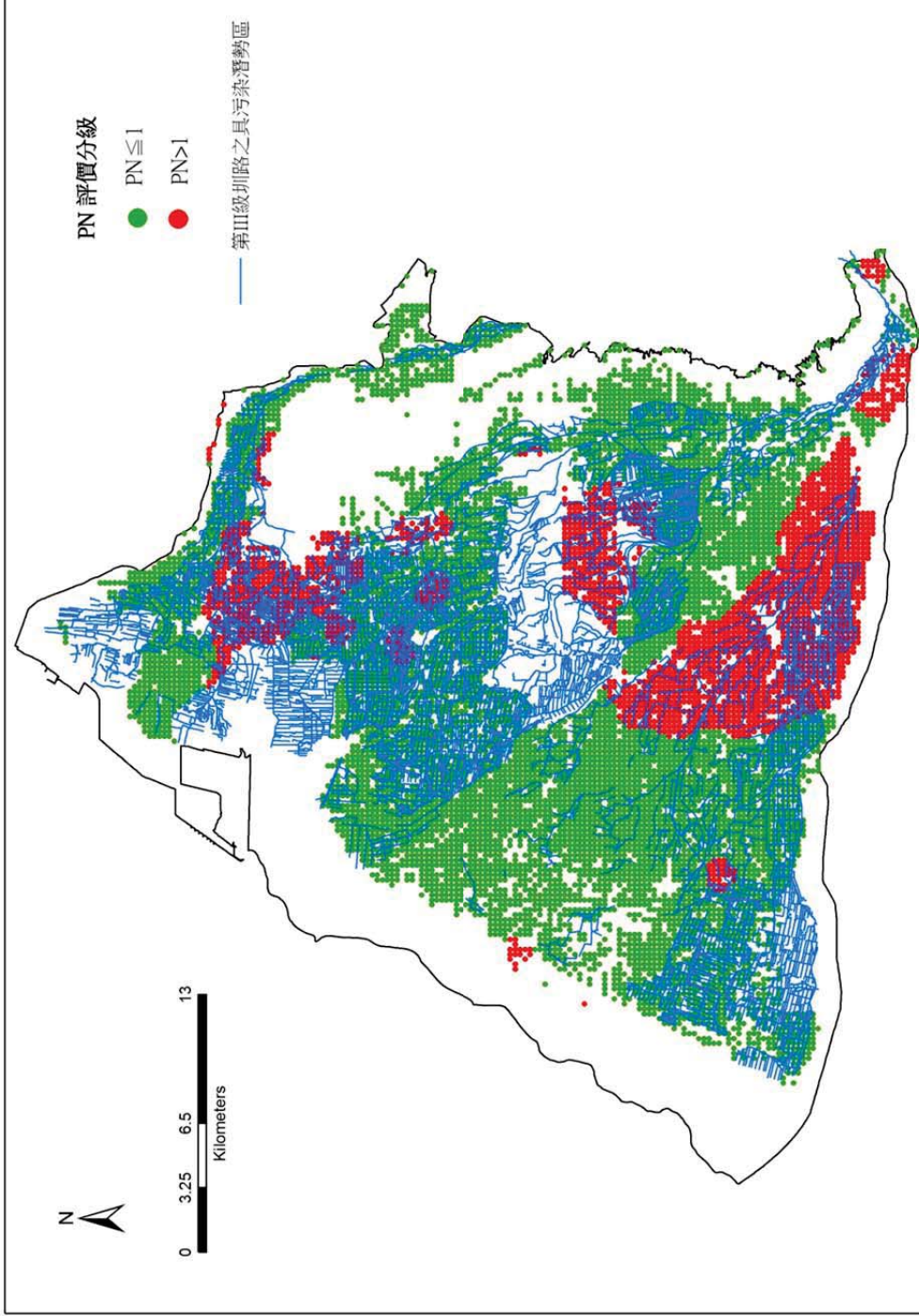


圖 4-13 彰化縣第 III 級圳路具污染潛勢區域套疊 SVM 的預測結果

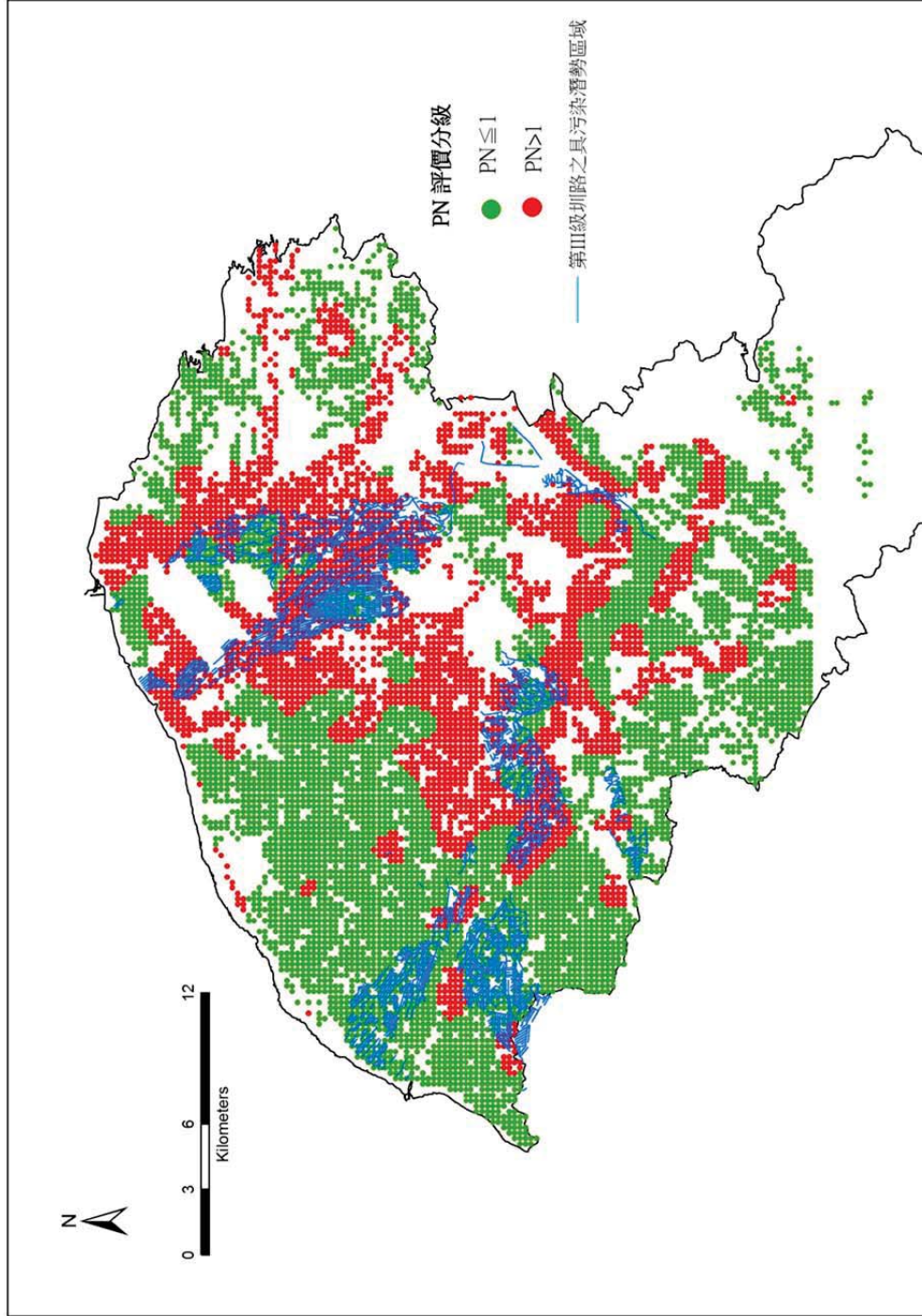


圖 4-14 桃園市第三級圳路具污染潛勢區域套疊 SVM 的預測結果

第五章 結論



1. 經支持向量機 (Support Vector Machine, SVM) 分類，彰化縣高污染潛勢點位數為 2,478 筆 (佔 23.89 %)，主要聚集分布於北彰化和美鄉、南舊濁水溪流域上游；桃園市高污染潛勢點位數為 3,535 筆 (佔 39.25 %)，聚塊分布散於桃園區域內，此結果能作為政府後續污染管理與防治規劃參考之依據。
2. 彰化縣以 7,353 筆點位在陽性、陰性標籤比 1:2 下訓練建立之模型進行土壤重金屬污染潛勢預測，結果準確度 (Accuracy) 為 85.37%、F1-measure 為 0.692；桃園市在標籤比為 1:1 下，共 3,288 筆資料訓練建立模型，污染潛勢預測之結果準確度為 71.58 %、F1-measure 為 0.506。
3. 本研究證實以 SVM 演算法能有效地應用於土壤重金屬污染潛勢劃分，且在低訓練集樣本數即可達良好的分類效能。
4. 本研究顯示將 SVM 區分農地重金屬高污染潛勢區之空間與屬性資料，套疊河川流域、工廠、工業區的空間分布，發覺多處高污染潛勢之區域屬非偶發場址，多位於工廠、工業區周邊，且污染隨渠道流動傳輸擴散，故 SVM 演算法所劃之污染潛勢圖具高合理性，可助有關單位判釋與決策。

參考文獻



- 年素磊，2011。半監督支持向量機綜述，南京大學計算機科學與技術系。
- 行政院，2011。中華民國行業標準分類（第9次修訂），行政院主計處編印。
- 行政院農業委員會，2014。灌溉水質監測調查及技術輔導，103年度農業發展計畫。
- 行政院環境保護署，1998。修正「地面水體分類及水質標準」，環署水字第0039159號令。
- 行政院環境保護署，1999。發佈「土壤及地下水污染整治法公布施行後過渡時期執行要點」，環署廢字第0024062號令。
- 行政院環境保護署，2011。修正「土壤污染整治」，環署土字第0990119003號令。
- 李亞松、張兆吉、費宇紅、王昭，2009。內梅羅指數評價法的修正及其應用，水資源保護，第6期。
- 周建成，1990。臺灣河川水質指數之建立，國立成功大學環境工程研究所碩士論文。
- 洪美秀，2013。臺灣農地重金屬高污染潛勢區域篩選方法之探討，臺灣大學生物環境系統工程學系碩士論文。
- 洪美秀、鄭百佑、徐貴新、張尊國，2012。內梅羅指標法評析淡水河水質，101年度農業工程研討會。
- 桃園縣政府，1998。桃園縣統計要覽，桃園：桃園縣政府出版。
- 翁煥廷，2010。探討「水源保護區河川污染水質指標」之應用問題，環保簡訊，第6期。
- 張尊國，1994。利用地理資訊系統於土壤污染等級區分與潛勢預測，行政院國家科學委員會專題研究計畫成果報告，第四章，17-22。
- 張尊國，2002。臺灣地區土壤污染現況與整治政策分析，財團法人國家政策研究基金會國政分析，永續（析）091021號。



- 張尊國，2010。桃園農地污染之環境資料蒐集與污染關聯性分析計畫報告書，行政院環保署計畫，EPA-100-GA101-03-A209。
- 張尊國，2010。彰化農地污染之環境資料蒐集與污染關聯性分析計畫，行政院環保署計畫，EPA-100-GA103-02-D054。
- 張尊國，2011。全國重金屬高污染潛勢農地之管制及調查計畫報告書，行政院環保署計畫，EPA-99-G101-03-A181。
- 張尊國、林裕彬，2000。地理統計模擬與估計法評估土壤重金屬污染範圍，行政院國家科學委員會專題研究計畫成果報告，NSC89-2621-B-002-004。
- 張尊國、管永愷、鄭百佑等，2012。全國重金屬高污染潛勢農地之管制及調查計畫，行政院環保署計畫，EPA-99-G101-03-A181。
- 張尊國、鄭百佑，2012。101 年度新北、台中、高雄農地污染之環境資料蒐集與污染關聯性分析計畫，行政院環保署計畫，EPA-101-GA11-03-D167。
- 馮秀娟、肖敏志、閻思諾、陳沛雲，2011。贛州不同級公路沿線農田土壤重金屬污染評價研究，*有色金屬科學與工程*，第 2 期第 1 卷，68-73。
- 經濟部水利署，2008。地理資訊倉儲中心。
- 劉衍君、湯慶新、白振華、張秀玲、張保華，2009。基於地質累積與內梅羅指數的耕地重金屬污染研究，*中國農學通報*，第 25 期第 20 卷，174-178。
- 劉修銘、姚佩萱、徐貴新、鄭百佑、張尊國，2014。內梅羅指標評析河川水體水質分類之達標程度，103 年度農業工程研討會。
- Boser, B., Guyon, M., Vapnik, V., 1992. A training algorithm for optimal margin classifiers. 5th Annual ACM Workshop on COLT, pp. 144-152.
- Chang, C., Lin, C., 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2(3), 27:1-27:27.

Chen, Z.S., Lee, D.Y., 1995. Heavy metals contents of representative agricultural soils in Taiwan. *Journal of Chinese Institute of Environmental Engineering* 5(3), 205-211.

Cheng, B.Y., Fang, W.T., Shyu, G.S., Chang, T.K., 2012. Distribution of heavy metals in the sediments of agricultural fields adjacent to urban areas Central Taiwan. *Paddy and Water Environment* 11, 343-351.

Chong, X., Fuchu, D., Xiwei, X., Yuan, H., 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* 145-146, 70-80.

Deng, H.G., Gu, T.F., Li, M.H., Deng, X., 2012. Comprehensive Assessment Model on Heavy Metal Pollution in Soil. *International Journal of Electrochemical Science* 7, 5286-5296.

Ding, Y., Song, X., Zen, Y., 2008. Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications* 34, 3081-3089.

Hernández-Sánchez, C., Luis, G., I.Moreno, Cameán, A., 2012. Differentiation of mangoes (*Mangifera indica* L.) conventional and organically cultivated according to their mineral content by using support vector machines. *Talanta* 97, 325-330.

Hsu, Z.Y., Su, S.W., Lai, H.Y., Guo, H.Y., Chen, T.C., Chen, Z.S., 2010. Remediation techniques and heavy metal uptake by different rice varieties in metal-contaminated soils of Taiwan: New aspects for food safety regulation and sustainable agriculture. *Soil Science and Plant Nutrition* 56(1), 31 - 52.

Huang, S.W., Jin, J.Y., 2008. Status of heavy metals in agricultural soils as affected by different patterns of land use. *Environmental Monitoring and Assessment* 139, 317-327.



Huang, Z., Chen, H., Hsu, C.J., 2004. Credit rating analysis with support vector machine and neural networks: A market comparative study. *Decision Support Systems* 37, 543-558.

Keerthi, S.S., Lin, C.J., 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation* 15(7), 1667-1689.

Kohavi, R., Provost, F., 1998. Glossary of terms - Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. *Machine Learning* 30(2-3), 271-274.

Lin, H.-T., Lin, C.J., 2003. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Technical Report, Department of Computer Science, National Taiwan University.

Maysam, A., Gholam-Hossain N., Abbas, B., 2012. Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences* 46, 272-283.

Mikhail, K., Nicolas, G., 1999. Environmental Spatial Data Classification with Support Vector Machines. IDIAP-RR-99-07.

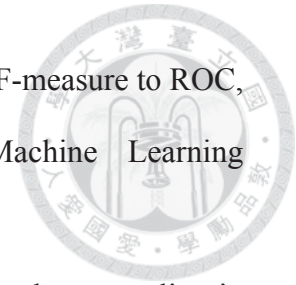
Miloš, K., Branislav, B., Boško, G., 2010. Soil type classification and estimation of soil properties using support vector machines. *Geoderma* 154, 340-347.

Nemerow, N.L., 1974. *Scientific Stream Pollution Analysis*. McGRAW-WILL BOOK COMPANY, New York, pp. 1-358.

Nemerow, N.L., 1985. *Stream, Lake, Estuary and Ocean Pollution*. Van Nostrand Reinhold Publishing Co., New York, pp. 303-309.

Pandey, G., Bin, Z., Le, J., 2013. Predicting submicron air pollution indicators: a machine learning approach. *Environmental Science: Processes & Impacts* 15.5, 996-1005.

Powers, David, M.W., 2011. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies* 2(1), 37-63.



Taner San, B., 2014. An evaluation of SVM using polygon-based random sampling in landslide susceptibility mapping: The Candir catchment area (western Antalya, Turkey). *International Journal of Applied Earth Observation and Geoinformation* 26, 399-412.

Tang, T., Zhai, Y., Huang, K., 2011. Water Quality Analysis and Recommendations through Comprehensive Pollution Index Method. *Management science and engineering* 5(2), 95-100.

Tay, F.E.H., Cao, L., 2001. Application of support vector machines in financial time series forecasting. *Omega* 29, 309-317.

Vapnik, V., 1995. Support-vector networks. *Machine Learning* 20, 273-297.

Yang, Q., Lia X., Shi X., 2008. Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences* 34, 592-602.

Yao, P.-H., Chang, T.-K., Shyu, G.-S., Cheng, B.-Y., Kuan, A., Chen, S.-D., Ho, J.-R., 2014. The course of protecting agricultural land from heavy metal pollution in Taiwan. 2014 International Conference on Remediation and Management of Soil and Groundwater Contaminated Sites, Taipei, Taiwan.

Yao, X., Tham, L.G., Dai, F.C., 2008. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* 101, 572-582.

Zuo, R., Carranza E.J., 2011. Support vector machine: A tool for mapping mineral prospectivity. *Computers & Geosciences* 37, 1967-1975.