




環境部環境管理署

112 年度土壤及地下水污染整治基金補助研究與模場試驗專案

以人工智慧結合遠端監控優化場址之 現地整治成效

期末報告（定稿）

主辦單位： 環境部環境管理署
專案執行單位：國立臺灣大學／環境工程學研究所
專案主持人：駱尚廉 教授
專案執行期間：112 年 12 月 15 日起至
113 年 11 月 29 日止

中 華 民 國 113 年 11 月 印製



環境部環境管理署「土壤及地下水污染整治基金補助研究與模場試驗專案」

☐構想書 ☐申請計畫書 ☐期中報告
☐修正計畫書 ☒成果報告

審查意見回覆對照表

| 計畫年度 | 112 年度 | 計畫類型 | <input type="checkbox"/> 先導型 <input type="checkbox"/> 研究型 <input checked="" type="checkbox"/> 模場型 |
|---|--|---|---|
| 計畫類別 | <input type="checkbox"/> 調查 <input checked="" type="checkbox"/> 整治 <input type="checkbox"/> 其他 | 主持人：駱尚廉 | NO：B7 |
| 計畫名稱 | 以人工智慧結合遠端監控優化場址之現地整治成效 | | |
| 審查意見 | | 執行單位回覆 | |
| 委員一 | | 1. 謝謝委員指教，已修正自評表 | |
| 1. 請更新專案成果績效自評表的內容（目前僅有申請預估數）。 2. 本年度著重於建立模型建立與學習訓練，並建立遠端監控設備與數據連線收集；對預測結果品質、目標值測試（如污染物移除效率等）、整治工法的生命週期評估等實際應用面仍待進行。第二年應有預測結果驗證與應用面向與限制說明。 | | 2. 根據本年度執行過程中發現特徵工程是提高學習效能重要的步驟，並耗費較多時間蒐集特徵參數的數據，加諸遠端監控與數據收集工作上，場址內變異條件較多，設備穩定性有待因應調整，故測試耗時較多，於本年度中尚未能將模型應用到實場中進行驗證，但基礎工作已大致完成，後續的應用應可於第二年第一季開展預測與驗證作業，並分析相關影響因素酌予修正模型或探索其他特徵參數以提高效能。 | |
| 委員二 | | 1. 謝謝委員指教 | |
| 1. 本研發計畫「以人工智慧結合遠端監控優化場址之現地整治成效」為現地污染整治管理及成效提升技術開發計畫，目前已完成以人工智慧之機器學習方法，對於蒐集數據進行訓練、建立對於場址概念模式、整治成效的預測模型，並從整治過程數據修定場址概念模式、整治系統操作參數調整與整治成效評估指標，並已提出第一年研發成果報告，報告內容、已完成之工作項目及相關成果符合預定進度。 2. 針對計畫後續工作，亦已於本成果報告中說明，應可順利推展後續技術開發作業。 3. 研發團隊選擇模場整治系統中重要參數的遠端監控設備，將包括井內水壓、氣壓、整治設備流量、真空度、尾氣濃度等傳感器取得設備運轉過程中的各項即時參數，並搭配現場取樣與場址其他定期監測數據，導 | | 2. 謝謝委員指教 3. 謝謝委員指教，目前現場已安裝的遠端設備，都可以正常遠端監看運轉情況並針對 PID 傳感器定期進行查核與校正作業 4. 目前訓練的模型初步將不同場址中地下水水位、污染源深度、抽除設備能力等場址特徵納入參數進行學習，後續蒐集彙整的更多場址數據如發現其他可用於提高預測效能的特徵參數，亦將修正模型用於預測與驗證。 | |



| | |
|---|---|
| <p>入建立模型，同時也作為機器學習數據的一部分，驗證模型;相關參數之數據品質對整體模型具關鍵影響，建議研發團隊確保場址中設置之各項遠端監控設備運作、控管需有良好品保、品管與定期校正程序。</p> <p>4. 本研發計畫主要以雙相抽除工法為研究對象，有關人工智慧訓練學習及模型建立優化等項目，建議多蒐集並採用不同地理場域、地質特性之場址資料以提升人工智慧訓練學習資訊之多樣性與完整性。</p> | |
| <p>委員三</p> <p>1. 本年度成果符合計畫書執行內容與進度。</p> <p>2. 成果績效自評表申請預估項目與數量，截至目前都尚未達成(或是未填)，請補充，至今未達成也請填結案後半年是否會投稿或說明原因。</p> <p>3. 本期成果顯示以尾氣濃度為指標，已達成3種演算法預測值與實際值的良好比對。遠端監控系統也完成設置，兩者後續如何鏈結來達到91頁”提供現場操作人員更即時的操作調整建議；尾氣與真空度的預測結果可以提供有用的資訊輔助人員進行異常檢查與排除，並透過實場中蒐集即時的異常發生資訊。最終達到第7頁LNAPL型污染場址整治效率進而減低整治成本20-30%。提高能源使用效率20-30%。建議以實施例說明。</p> | <p>1. 謝謝委員指教</p> <p>2. 本年度預定114/1月底前將會提出兩篇論文投稿，目前專利申請書正在撰製中。</p> <p>3. 已補充透過調整建議後對於整治成本(時間)、能源使用效率的效益說明如表4-10。</p> |
| <p>委員四</p> <p>1. 本報告應為期末成果報告，然而報告中章節均說明為期中報告，請確認是否如期完成，並請修正。另外第四章開頭說明「資料建置預定於6月中前完成」筆誤?</p> <p>2. 4.2.3以不同場址模擬抽提井污染分層推估，進而獲得KRR-5效能最佳，以較表4-8列出各類模型輸入值，其中KRR-5的未來應用性，是否受限與資料完整性?是否有替代方案。</p> | <p>1. 謝謝委員指教，誤植內容以進行修訂</p> <p>2. 由於目前已採用的特徵參數約在20個以內，以此作為機器學習的輸入值，在數百筆到數千筆的訓練數據都是足夠的，而KRR模型對於萬筆以內的數據，特徵值100個以內的應用都是適用的。本研究另外比對的MLP跟RBFN對於更大型的數據集(10萬或更大)皆可適用，目前受限數據集屬於中小型的數據量，其效能反而較差，後續倘若數據量或特徵量增加，仍可仍作為模型評估的選項之一。</p> |



環境部環境管理署「土壤及地下水污染整治基金補助研究與模場試驗專案」

☐構想書 ☐申請計畫書 ☒期中報告
☐修正計畫書 ☐成果報告

審查意見回覆對照表

| 計畫年度 | 112 年度 | 計畫類型 | <input type="checkbox"/> 先導型 <input type="checkbox"/> 研究型 <input checked="" type="checkbox"/> 模場型 |
|---|--|--|---|
| 計畫類別 | <input type="checkbox"/> 調查 <input checked="" type="checkbox"/> 整治 <input type="checkbox"/> 其他 | 主持人：駱尚廉 NO：B7 | |
| 計畫名稱 | 以人工智慧結合遠端監控優化場址之現地整治成效 | | |
| 審查意見 | | 執行單位回覆 | |
| 委員一 3. 期中進度較預定查核點慢，應加速資料庫建置進度以開始進行數據訓練，於成果達到模型訓練、模型選擇及效能評估。 4. 4.3 節文中圖名缺漏應補正。 | | 1. 謝謝委員指教。現場基礎數據的彙總已經如期完成，但是數據品質的檢核會影響預測效能，因此，本研究中先建立了 CSMcheck 程序來檢核數據品質並回溯檢討數據是否有誤，此工作隨著後續訓練過程中仍會持續進行。 2. 遵照辦理，已補正。 | |
| 委員二 5. 本研發計畫「以人工智慧結合遠端監控優化場址之現地整治成效」為現地污染整治管理及成效提升技術開發計畫，目前已完成(已解除列管)場址操作參數及監測數據之資料庫建置以及人工智慧模型建立等研究工作，並提出期中成果報告，報告內容、已完成之工作項目及相關成果符合預定進度。 6. 針對計畫後續工作，亦已於期中報告中說明，應可順利推展後續技術開發作業。 7. 本研發計畫將 SVE、AS、DPE 等常用技術納入本案整治管理及優化之技術工項，本項專案計畫成果應有助於提升國內現地污染整治之管理及成效，進一步達到節能減排及縮短整治期程之目標。 8. 國內目前仍有許多運作中之污染場址應用本計畫目標技術辦理整治工作，建議於計畫後期挑選合適場址將研發成果進行實場測試，以即時修訂並提升其應用成效。 | | 謝謝委員指教 | |



| | |
|--|---|
| <p>委員三</p> <p>4. 本計畫擬建立一套智能判斷的整治系統輔助現場人員操作，進而達到優化整治系統成效。目前收集 16 場址資料，但此 16 場址整治歷程、TPH 污染量/濃度範圍、土壤質地；地下水位等關鍵項目又大多不同，有限資料是否會 AI 產出的智能判讀結果？</p> <p>5. 預定以智能判斷的整治系統輔助現場人員操作。目前正建置遠端監控設備，但 AI 學習與判讀也剛起步進行中，宜規劃與說明兩個工項之結合與搭配。</p> <p>6. 中文摘要有敘及：「在考慮生命週期碳排影響下，讓整治過程朝向兼具永續及韌性提升」建議補充工項中預期成果達到或產生此效益之具體說明。</p> | <p>1. 根據目前研究成果，在統一化每個場址從水位深度上下進行分層同時加入場址特徵參數後就學習成效來看並無明顯影響效能，但在學習過程中會發現不同特徵參數的添加都會必須進行超參數的調整以得到最佳的模型，因此，在以既有場址訓練之後，實場大量數據進入，還需要繼續評估既有模型是否最優化。</p> <p>2. 已補充說明在第 4.3 節</p> <p>3. 已補充說明於第 4.4 節</p> |
| <p>委員四</p> <p>3. AI 應用與整治工法的文獻收集及分析整理內容豐富，建議在成果報告中補充其應用優劣勢比較，後續應用可行性的分析。</p> <p>4. 第三章前期訓練場址彙整表 3-2，最後欄位整治系統的內容意義說明。</p> <p>5. 場址遠端監控系統架構清楚呈現在圖 3-8，建議補充模塊一、二、三監測系統設置相對位置的場址配置圖，可與圖 3-7 既有整治系統配置圖作為比較。</p> <p>6. 第四章初步結果與討論 4.2 節分別比較不同測試參數、項目製作說明預測圖及影響，建議本節增加摘要說明-整合初步成果，初步的發現及重點參數控制的篩選。</p> <p>7. 後續工作計劃，除了條列工作項目，建議有具體的工作規劃說明。</p> | <p>1. 針對</p> <p>2. 已補充說明之</p> <p>3. 已補充說明之</p> <p>4. 已增加相關說明於 4.4 節</p> <p>5. 本計畫為兩年期計畫，後續工作規畫說明將進一步細化於第二年度工作計畫書中</p> |
| <p>委員五</p> <p>1. 申請計畫書審查意見回覆表請於定稿補上。</p> | <p>遵照辦理</p> |



專案基本資料表

| | | | | | | | | |
|--|--|--|----------|----------------------------|-----------|---|------------------|---------------|
| 專案性質 | | <input checked="" type="checkbox"/> 實驗性質 <input type="checkbox"/> 非實驗性質 | | 專案技術編碼 | | LAB - S - R -R1 -M - (如非實驗性質含氣溶劑之生物整治模場為 NLAB-S-R-C1-B，填寫請參考附件 1-4) | | |
| 專案類別(單選) | | <input type="checkbox"/> 先導型 <input type="checkbox"/> 研究型 <input checked="" type="checkbox"/> 模場型 | | 研究主題 | | <input checked="" type="checkbox"/> 整治 <input checked="" type="checkbox"/> 調查 <input type="checkbox"/> 政策類或其他 | | |
| 申請機構系所 | | 國立臺灣大學環境工程學研究所 | | | | | | |
| 機構地址 | | 臺北市大安區羅斯福路四段 1 號 | | | | | | |
| 專案主持人 | | 駱尚廉 | | 職等／職稱 | | 特聘教授 | | |
| 協同主持人 | | 關蓓德 | | 職等／職稱 | | 教授 | | |
| 協同主持人 | | 楊裕堅 | | 職等／職稱 | | 新野科技股份有限公司副總經理 | | |
| 專案名稱 | | 中文 | | | | | | |
| | | 英文 | | | | | | |
| | | 關鍵字 | | | | | | |
| | | 以人工智慧結合遠端監控優化場址之現地整治成效 | | | | | | |
| | | Optimization of remediation performance with artificial Intelligence and remote sensing Technology | | | | | | |
| | | 人工智慧, 整治成效優化, 遠端監控, 機器學習 | | | | | | |
| 執行期程 | | 自 民國 112 年 12 月 15 日 起 至 民國 113 年 11 月 29 日 止 | | | | | | |
| 專案主持人 | | 姓名：駱尚廉 | | E-mail：sllo@ntu.edu.tw | | 專線：02-23625373 手機：0915617061 | | |
| 專/兼任人員 | | 姓名：翁瑞蓮 | | E-mail：p96541205@gmial.com | | 專線：02-33664377 手機：0910-282849 | | |
| 經費分析總表 (僅模場試驗專案需填寫兩年度金額) | | 專案預估 總經費 | | 第一年 申請經費 | | 第二年 申請經費 | | |
| | | | | | | 編列說明 | | |
| | | 1. | 人事費用 | | 1,080,887 | 1,093,700 | (1~5 項相加之 50%為限) | |
| | | 2. | 設備使用含維護費 | | 78,650 | 79,000 | (與計畫實驗相關) | |
| | | 3. | 耗材與主要費用 | | 1,208,538 | 698,000 | (與計畫主體相關) | |
| | | 4. | 其它研究相關費用 | | 875 | 18,000 | (含差旅與租賃費用) | |
| | | 5. | 行政管理費 | | 418,050 | 333,300 | (1~5 項相加之 15%為限) | |
| | | 6. | 自籌款 | | 1,018,600 | 752,400 | (申請單位自行籌備款項) | |
| | | | | 申請補助經費(1~5 項) | | 2,787,000 | 2,222,000 | 總金額：5,009,000 |
| | | 計畫總經費(1~6 項) | | 3,805,600 | 2,974,400 | 總金額：6,780,000 | | |

說明：

1. 本署以部份補助為原則，得指定不補助項目，不補助項目應為自籌款項，應自行核銷不得申請
2. 非實驗性質之專案，若第一項費用編列金額超出比率，須於申請公文內向本署說明，俾利審查
3. 補助專案經費編列自申請起經專案核准後，經費編列中各項費用單價與需求只可減少與下修。

專案主持人：_____（簽名或蓋章） 日期：_____

112 年度專案成果績效自評表

填表日期：113 年 11 月 29 日

一、 專案基本資料

| | | | |
|--------|--|-------|---|
| 專案性質 | <input checked="" type="checkbox"/> 實驗性質 <input type="checkbox"/> 非實驗性質 | 專案類別 | <input type="checkbox"/> 先導型 <input type="checkbox"/> 研究型 <input checked="" type="checkbox"/> 模場型 |
| 研究主題 | <input checked="" type="checkbox"/> 整治 <input type="checkbox"/> 調查 <input type="checkbox"/> 其他 | | |
| 執行機構 | 國立臺灣大學/環境工程學研究所 | 專案主持人 | 駱尚廉特聘教授 |
| 專案名稱 | 以人工智慧結合遠端監控優化場址之現地整治成效 | | |
| 專案執行期程 | <input type="checkbox"/> 申請階段 <input type="checkbox"/> 期中 <input checked="" type="checkbox"/> 成果 | | |

二、 成果績效自評

「計畫總預估數」應與計畫審查核定值相符，請執行單位依實際達成之量化成果填寫於欄位中。

(一) 學術面

| 目標達成程度 | | | 申請 預估數 | 期中 達成數 | 成果 達成數 | 結案後半 年達成率 | 備註 (說明未達成 原因或學術產 出發表名稱) |
|--------------------------------------|------------------|--------------|-----------|-----------|-----------|--------------|----------------------------------|
| 項目 | | | | | | | |
| A 學 術 產 出 及 活 動 | 1.國內投稿 (篇數) | (1)論文 | 0 | 0 | 0 | | |
| | | (2)研討會論文 | 2 | 0 | 0 | | |
| | 2.國外投稿 (篇數) | (1)期刊論文 | 2 | 0 | 2 | | 註一 |
| | | (2)國際研討會論文發表 | 1 | 0 | 0 | | |
| | 3.報告(篇數) | (1)技術報告 | 0 | 0 | 0 | | |
| | | (2)研究報告 | 1 | 1 | 1 | | |
| | 4.專著 (本數) | | | | | | |
| | 5.辦理學術 會議(場數) | (1)研討/說明會 | | | | | |
| | | (2)成果發表會 | | | | | |
| | | (3)論壇 | | | | | |
| | 6.研發改良 技術(項數) | (1)已開發技術 | | | | | |
| | | (2)技術平台 | | | | | |
| | 7.技術獎項(項數) | | | | | | |
| B 人 才 培 育 | 8.研發人員 (人數) | (1)碩士 | 2 | 2 | 2 | | |
| | | (2)博士 | 2 | 2 | 2 | | |
| | 9.研究團隊 (個數) | (1)跨領域團隊 | | | | | |
| | | (2)跨機構團隊 | 1 | 1 | 1 | | |
| | | (3)形成研究中心 | | | | | |
| | | (4)形成實驗室 | | | | | |
| 10.其他指標 (請自行命名) | | (請自填) | | | | | |



(二) 產業面

| 項目 \ 目標達成程度 | | | | 申請 預估數 | 期中 達成數 | 成果 達成數 | 結案後半年 達成率 | 備註 (說明未達成 原因或專利、 技術轉移相關 詳細資料) |
|---------------------------------|------------------|---------------|-------|-----------|-----------|-----------|--------------|---|
| A 智慧 財產 權 | 1.專利 (件數) | 已 核 准 | 發明 | | | | | |
| | | | 新型/設計 | | | | | |
| | | | 合計 | | | | | |
| | | 申 請 中 | 發明 | | | | | |
| | | | 新型/設計 | 1 | 0 | 1 | | |
| | | | 合計 | 1 | 0 | 1 | | |
| B 研 發 技 術 轉 移 | 2.先期技術 成果移轉 | 件數 | | | | | | |
| | | 授權金(仟元) | | | | | | |
| | | 衍生利益金(仟元) | | | | | | |
| | 3.技術移轉 (專利) | 件數 | | | | | | |
| | | 授權金(仟元) | | | | | | |
| | | 衍生利益金 (仟元) | | | | | | |
| | 4.技術移轉 (應用技術) | 件數 | | 1 | 0 | 0 | | |
| | | 授權金(仟元) | | 300 | 0 | 0 | | |
| | | 衍生利益金(仟元) | | | | | | |
| | 5.可移轉 產業技術 | (1)技術(件數) | | | | | | |
| (2)品種/系 (件數) | | | | | | | | |
| C 產 學 研 合 作 | 6.促成合作研 究 | 件數 | | | | | | |
| | | 金額(仟元) | | | | | | |
| | 7.促成投資 | 件數 | | | | | | |
| | | 投資金額 (仟元) | | | | | | |
| | 8.促成取得業 界科專 | 件數 | | | | | | |
| 業界投資金額 (仟元) | | | | | | | | |
| 9.其他指標 (請自行命名) | | (請自填) | | | | | | |



(三) 政策面

| 項目 | | 目標達成程度 | 申請預估數 | 期中達成數 | 成果達成數 | 結案後半年達成率 | 備註 (說明未達成原因或其他詳細資料) |
|-------------------|------------------|--------|-------|-------|-------|----------|------------------------|
| | | | | | | | |
| A 服務 便民 | 1.技術服務 | 次數 | | | | | |
| | | 收入(仟元) | | | | | |
| | 2.諮詢服務 | 次數 | | | | | |
| | | 收入(仟元) | | | | | |
| B 政策 推動 | 3.協助政府推動 (件數) | (1)政策 | | | | | |
| | | (2)法規 | | | | | |
| C 技術 效益 | 4.整治技術提升(%) | | | 25% | 25% | | |
| | 5.整治成本降低(%) | | 25% | 0 | 0 | | |
| | 6.提升能源效率(%) | | 25% | 0 | 0 | | |
| 7.其他指標 (請自行命名) | | (請自填) | | | | | |

三、請依前述學術面、產業面、政策面等預期量化成果，具體敘明研究成果對本署政策推動之助益。(200 字為限)

| |
|--|
| 學術面(如國內外研討會、期刊投稿件數、或人才培育碩博士生說明) 1. 投稿國外論文期刊 2 篇、參與國內研討會發表論文 2 篇、參與國外研討會發表論文 1 篇，預定於 114 年 1 月底前投出以下兩篇論文： <ul style="list-style-type: none"> ◆ Payal Maharathi, Kiran Kumar Eripogu, Shang-Lien Lo, and Hsiao-Fen Wu, "A Synergistic Approach: For Promoting Soil Vitality and Crop Diversity With Eco-Friendly Biochar And Advanced Nutrient Absorption Technologies" Submitted to <i>J. Environmental Management</i>. ◆ Hsiao-Fen Wu, Shang-Lien Lo., Pei-Te Chiueh, "Impact of Features on Machine Learning Predictions of Extracted Gas Concentrations in the In Situ Dual-Phase Extraction System." Submitted to <i>J of Hydrology</i> ◆ 預定於 114 年後半年參加國外研討會一場 2. 培育 2 名碩士生、2 名博士生。 |
| 產業面(如合作研發產業、申請專利、洽談技術移轉廠商件數說明) 1. 申請監控設備新型專利 1 件，目前著手申請以下一件專利： 「一種應用於現地雙相抽除系統中的多管道有機氣體光離子偵測系統」 2. 技術移轉 1 件，授權金 300(仟元)。 |
| 政策面(如整治費用降低、特定污染物整治效益提升、或提供政府作為監測/管制標準、污場址管理等政策及法規研訂之參考) |



1. 整治技術提高的評估方法，以專案團隊反饋用於系統操作、判斷的投入時間為基礎，判斷其團隊對於應用本技術後的反饋時間之提升程度。客觀的處理方式可以用 AI 判斷須排查系統的次數作為技術提升的評估指標，例如 AI 建議當月排查次數有 5 件，按照專案若原先每月進度報告分析 1 次 3 件來看，增加 2 件，技術提高比例為 $(5-3)/3$ 。按照本年度執行情形，在遠端監控系統安裝後，搭配場址 CSM 建立下，專案團隊反饋減少數據分析的時間與周期，該部分提高 25% 的技術能力。
2. 因為本年度安裝與測試完成後，尚未針對操作部分執行到提供現場操作建議，故上無法評估具體數據，原則上，暫定以操作時數的調整作為整治成本降低的評估指標，例如，從原先操作條件例如每日抽 5 小時休 1 小時，透過 PID 實時監測及 AI 預測調整操作條件為抽 4 小時休 2 小時，移除量一樣，則每日操作時數從 20 小時降為 16 小時，操作成本減少比例為 $(20-16)/20$ 。
3. 能源效率提升：同上，原先 5 小時的移除量，透過監控與 AI 預測於 4 小時達到相當的移除量，則能源效率提升比例為 $(5-4)/5$



四、主要研究人力

| 類別 | 姓名 | 服務機構/系所 | 職稱 | 在本研究專案內擔任之具體工作性質、項目及範圍 | *每月平均投入工作日數比率(%) |
|-------|-----|--------------------|------|----------------------------|------------------|
| 計畫主持人 | 駱尚廉 | 國立臺灣大學 環境工程學研究所 | 特聘教授 | 整體計畫執行規劃、人工智慧與機器學習方法建立 | 25% |
| 協同主持人 | 闕蓓德 | 國立臺灣大學 環境工程學研究所 | 教授 | 指導機器學習、生命週期評估等研究方法之應用與結果闡釋 | 20% |
| 協同主持人 | 楊裕堅 | 新野科技股份有限公司 | 副總經理 | 協助模場場址之相關協調聯繫工作與現場安全衛生管理事宜 | 20% |
| 專任人員 | 翁瑞蓮 | 國立臺灣大學 環境工程學研究所 | 專任助理 | 協助報告撰寫、樣品採樣分析、數據整理及行政等事務 | 100% |
| 兼任人員 | 吳曉芬 | 國立臺灣大學 環境工程學研究所 | 博士生 | 場址相關數據蒐集、分析、程序訂立與模式運算 | 50% |
| 兼任人員 | 施承恩 | 國立臺灣大學 環境工程學研究所 | 博士生 | 協助報告撰寫、樣品採樣分析、數據整理 | 50% |
| 兼任人員 | 黃詩雯 | 國立臺灣大學 環境工程學研究所 | 碩士生 | 協助報告撰寫、樣品採樣分析、數據整理 | 50% |
| 兼任人員 | 吳立芹 | 國立臺灣大學 環境工程學研究所 | 碩士生 | 協助報告撰寫、樣品採樣分析、數據整理 | 50% |

五、產業界資源投入表

專案執行若有與產業界合作，請執行單位依實際量化成果填寫於欄位中。

| 投入資源類別 | 數量 | 說明 |
|-------------------|--------------|------------|
| 1.自籌款 | 1,018,600 元 | |
| 2.人力 | 8 人月 | |
| 3.設備 | 9410VTR 直推鑽機 | 提供現場採樣機具使用 |
| 4.其他資源 (請自行增列) | (請自填) | |



中文摘要

污染場址採取現地整治工法進行整治作業時，場址各類數據的取得與分析為優化整治成效中關鍵性的因素，而這些數據的變異性將影響到場址概念模式的修正、整治系統的調整與列車型整治方案的決策，因此，本研究將以 LNAPL 污染類型的場址採取現地地下水抽提、土壤氣體抽提、雙相抽提、地下水注氣為主的工法者為研究對象，蒐集 10-20 處於同類型工法的污染場址整治成功的案例，以人工智慧中機器學習的方法，對於蒐集數據進行訓練並建立對於場址概念模式、整治成效的預測模型，進而達到從整治過程的數據決策出修定場址概念模式、整治系統操作參數調整與整治成效評估指標；其中，指標之決定將納入生命週期評估的參數。為了驗證前述訓練的模型，本研究於挑選的模場中安裝整治系統中重要參數的遠端監控設備，其中，遠端監控設備將包括井內水壓、氣壓、整治設備流量、真空度、尾氣濃度等傳感器取得設備運轉過程中的各項即時參數，並搭配必要的現場取樣比對與蒐集場址其他定期監測數據，導入建立的模型內評估模型的準確性，同時也繼續作為機器學習數據的一部分，持續驗證模型，以最終確認模型的可行性。

本研究整合場址內土壤氣體、土壤及地下水頂空濃度等現場快篩數據利用 RBF 方法建立場址分層污染分布模型，並對應抽提井所在網格分層污染分布作為輸入參數，加入地下水水位、整治設備操作參數對應抽除系統尾氣濃度，操作 MLP、RBFN 及 KRR 三種演算法建立預測模型，初步得到 KRR 的模型效能最佳，以回歸模型 R^2 值來看，訓練集、驗證集與測試集的效能達到 0.87、0.75 及 0.76；利用 SVM 分類機針對尾氣濃度是否達到尾氣處理設備採用觸媒處理($\geq 750\text{ppmV}$)以及是否濃度偏低應採取替代方案的分類上($<150\text{ppmV}$)，最佳模型準確度與 AUC 分別達到 0.95, 0.91 及 0.88、0.95，顯示在分類上已經具有一定預測能力；遠端監控系統已於現場安裝完成，後續將持續收集場址資料並導入邊緣計算器以利預測模型於現場應用，進而達到提高整治操作效率並降低整治成本，提高節能減碳，進而達到永續韌性的整治目標。。



英文摘要

When implementing on-site remediation using in situ methods for contaminated sites, the acquisition and analysis of various types of data on the site are crucial factors for optimizing the effectiveness of the remediation. The variability of these data will affect the adjustment of the site conceptual model, the tuning of the remediation system, and the decision-making for the train-type remediation scheme. Therefore, this study will use machine learning techniques to train and establish predictive models for the conceptual model and remediation performance with 10-20 successfully remediated sites, which are LNAPL (Light Non-Aqueous Phase Liquid) contaminated sites with groundwater extraction, soil gas extraction, dual phase extraction, and air sparging remediation technology implemented. Furtherly, this approach aims to achieve data-driven decision-making in the remediation process, including the revision of site conceptual models, adjustment of operating parameters for the remediation system, and evaluation of indicators for remediation effectiveness. The determination of these indicators will incorporate parameters from lifecycle assessments. To validate the trained models, the study will install remote sensing devices for important parameters of the remediation system at selected pilot sites. The remote sensing devices will include sensors for in-well water pressure, air pressure, flow rates of remediation equipment, vacuum levels, exhaust gas concentrations, and other real-time parameters during the operation of the equipment. These parameters will be compared and correlated with necessary on-site sampling and other regularly collected monitoring data to assess the accuracy of the models. Additionally, these data will continue to be part of the machine learning process to continuously validate the models, ultimately confirming their feasibility.

This study integrates on-site rapid screening data, including soil gas, soil, and groundwater headspace concentrations, to establish a layered contamination distribution model of the site using the Radial Basis Function (RBF) method. The model uses the layered contamination distribution within the grid corresponding to extraction wells as input parameters, along with groundwater level, operational parameters of remediation equipment, and exhaust gas concentrations from the extraction system as the output. It employs three algorithms—MLP, RBFN, and KRR—to build predictive models, with preliminary results showing that the KRR model performs best. In terms of regression



model R^2 values, the training, validation, and testing sets achieve scores of 0.87, 0.75, and 0.76, respectively. An SVM classifier is used to determine whether the exhaust gas concentration meets the threshold for catalytic treatment (≥ 750 ppmV) or whether alternative measures are required due to low concentrations (< 150 ppmV). The best classification model achieves accuracy and AUC of 0.95, 0.91 for the 750ppmV classification, and 0.88, and 0.95 for 150ppmV, indicating substantial predictive capability. A remote monitoring system has been installed on-site, and data collection will continue to feed into an edge computer, enabling on-site application of the predictive model. This approach aims to enhance remediation operation efficiency, reduce remediation costs, improve energy conservation and carbon reduction, and ultimately achieve the goals of sustainable and resilient remediation.



目錄

| | |
|---|------------|
| 第一章 前言 | 19 |
| 1.1 計畫緣起 | 19 |
| 1.2 研究目的 | 21 |
| 第二章 文獻回顧 | 23 |
| 2.1 雙相抽除處理技術原理與應用 | 23 |
| 2.2 地下水注氣技術原理與應用 | 30 |
| 2.2 土壤及地下水調查與整治應用機器學習方法 | 33 |
| 2.3 人工神經網絡演算法 | 44 |
| 第三章 研究方法及步驟 | 50 |
| 3.1 整體研究流程 | 50 |
| 3.2 前期訓練場址資料收集 | 51 |
| 3.3 數據預處理 | 55 |
| 3.4 模型選擇與訓練 | 57 |
| 3.5 模場試驗場址資料 | 58 |
| 3.6 場址遠端監控系統架構規劃與建置 | 65 |
| 3.7 整治成效評估指標之建立 | 67 |
| 3.8 本年度計畫執行進度規劃 | 68 |
| 第四章 結果與討論 | 72 |
| 4.1 資料前處理執行情形 | 72 |
| 3.2 模型建立 | 72 |
| 3.2.1 以抽提井為核心的分層汙染特性做為輸入值的方法 | 72 |
| 3.2.2 建立場址概念模式提取至抽提井的分層汙染特性做為輸入值的方法 | 81 |
| 3.2.3 不同場址特徵參數對於模擬結果的影響 | 84 |
| 3.2.4 回歸模型的選擇 | 87 |
| 3.2.5 尾氣濃度分類的學習成果 | 90 |
| 4.3 遠端監控設備設置情形 | 95 |
| 8.4 初步工作成果與後續工作計畫 | 103 |
| 第五章 參考文獻 | 106 |



表目錄

| | |
|---|-----|
| 表 1-1、台灣 LNAPL 解列耗時與未解列已耗時統計數據比較 | 21 |
| 表 2-1、不同土壤質地對應水力傳導係數範圍 | 24 |
| 表 2-2、不同土壤質地抽氣影響半徑 | 27 |
| 表 2-3、利用 AI 建立場址特徵評估模型的相關研究列表 | 37 |
| 表 2-4、利用 AI 建立場址整治優化模型的相關研究列表 | 41 |
| 表 2-5、機器學習預測模型驗證誤差指標類型 | 48 |
| 表 3-1、本計畫預定取得數據與參數類型 | 53 |
| 表 3-2、本研究前期訓練場址彙整 | 54 |
| 表 3-3、本計畫偵測器查核、校正與允收標準 | 66 |
| 表 3-4、工作進度甘特圖 | 71 |
| 表 4-1、資料庫建置情形說明 | 72 |
| 表 4-2、抽提井核心法建立輸入參數 | 73 |
| 表 4-3、抽提井為核心的輸入參數共變異數(Covariance) | 73 |
| 表 4-4、激活函數列表 | 78 |
| 表 4-5、輸入、目標變數相關性分析表 | 85 |
| 表 4-6、各特徵參數數據統計表 | 85 |
| 表 4-7、在不考慮超參數下的輸入變數以 MLP 演算的效能比較 | 85 |
| 表 4-8、本計畫選擇模型輸入值列表 | 88 |
| 表 4-9、不同模型對於尾氣濃度之預測成效 | 89 |
| 表 4-10、KRR-5 模型於 DPE 整治系統調整之應用與效益 | 101 |



圖目錄

| | |
|--------------------------------------|----|
| 圖 1-1、台灣 LNAPL 污染類型解列與未解列歷時統計 | 20 |
| 圖 1-2、不同思維的整治行動與其成效示意 | 21 |
| 圖 2-1、雙相抽除處理系統(DPE)示意圖 | 23 |
| 圖 2-2、不同土層質地在給定毛細壓力與飽和度下的空氣滲透性 | 24 |
| 圖 2-3、不同質地含水度與毛細壓力對照圖 | 25 |
| 圖 2-4、不同物質在相同流量條件的抽氣有效影響半徑變化 | 27 |
| 圖 2-5、雙泵浦式雙相抽除處理井設井構造圖 | 28 |
| 圖 2-6、不同型式抽取泵的馬達用電、轉速與功能曲線 | 29 |
| 圖 2-7、空氣注入系統配置示意圖 | 31 |
| 圖 2-8、曝氣井設置規範..... | 32 |
| 圖 2-9、人工智慧與機器學習涵蓋分析方法的體系圖 | 35 |
| 圖 2-10、機器學習與深度學習涵蓋方法的體系圖 | 36 |
| 圖 2-11、涉及 AI 分析地下水水質類型之研究數量分析 | 36 |
| 圖 2-12、多層感知器架構..... | 45 |
| 圖 2-13、RBF 訓練架構..... | 47 |
| 圖 3-1、本計畫整體研究步驟..... | 50 |
| 圖 3-2、模型建立與測試流程..... | 56 |
| 圖 3-3、本計畫模場試驗預定位置 | 59 |
| 圖 3-4、本計畫預定場址之地質特性 | 60 |
| 圖 3-5、本場址地下水流向推估圖 | 61 |
| 圖 3-6、本場址土壤污染範圍推估圖 | 62 |
| 圖 3-7、本場址地下水中 TPH 污染範圍推估圖 | 62 |
| 圖 3-8、整治工作整體流程圖..... | 63 |
| 圖 3-9、模場內污染土壤開挖與生物復育場區域圖 | 64 |
| 圖 3-10、模廠內既有地下水整治系統配置圖 | 64 |
| 圖 3-11、模場監控架構規劃圖..... | 69 |
| 圖 3-12、現地壓力與土壤氣體偵測井設計規劃圖 | 70 |
| 圖 4-1、期中報告中輸入變數的數據分布直方圖 | 74 |
| 圖 4-2、不同隱藏層數量訓練誤差箱線圖 | 75 |
| 圖 4-3、不同隱藏層數量訓練曲線與準確度對比 | 76 |
| 圖 4-4、變更激活函數後的模型變化 | 77 |



| | |
|---|-----|
| 圖 4-5、不同 RBF 數量對於準確度的影響 | 79 |
| 圖 4-6、不同 RBF 數量訓練集的實際值與預測值比較圖 | 79 |
| 圖 4-7、不同 RBF 數量測試集的實際值與預測值比較圖 | 80 |
| 圖 4-8、以 MLP 訓練尾氣濃度在不同隱藏層之效能..... | 81 |
| 圖 4-9、場址污染分布的演化..... | 84 |
| 圖 4-10、不同特徵參數數據分布型式 | 86 |
| 圖 4-11、輸入特徵對於 MLP 預測模型的影響..... | 87 |
| 圖 4-12、不同演算法最佳模型以預測值與實際值對比圖 | 90 |
| 圖 4-13、以 SVM 預測 750ppmv 成效對照圖..... | 92 |
| 圖 4-14、SVM 針對 750ppm 分類模型 ROC 曲線評估選型 | 93 |
| 圖 4-15、以 SVM 預測 150ppmv 成效對照圖 | 94 |
| 圖 4-16、SVM 針對 150ppmv 分類模型 ROC 曲線評估選型 | 95 |
| 圖 4-17、本研究中增設試驗裝置之現場設施配置圖 | 97 |
| 圖 4-18、遠端監控系統現場安裝完成照片 | 97 |
| 圖 4-19、已完成 SVE 遠端數據監控單元與遠端下載平台界面..... | 100 |
| 圖 4-20、PID 尾氣監測系統架構圖..... | 103 |



第一章 前言

1.1 計畫緣起

自國內於 2002 年實施土壤及地下水污染整治法以來，無論經政府查證發現或是自行發生洩漏採取改善作業的污染場址，常因場址持續營運需要、經費與技術可行等因素而採取現地整治的技術進行污染控制與整治作業，其中，針對土壤及地下水同時存在有機類型污染物的情況，採取地下水抽出處理(pump and Treat,P&T)、土壤氣體抽除處理(Soil vapor extraction, SVE)、雙相抽除(Dual Phase extraction, DPE)、地下水注氣(Air Sparging, AS/Biosparging, BS)、注藥氧化(In situ chemical oxidation, ISCO)或生物處理(Bioremediation)為常見的現地整治技術，此些技術主要利用建井或鑽孔的方式，結合可以實現提供物理、化學、生物作用之設備，將土壤及地下水中污染物移除或分解後達到污染物濃度下降至低於標準之程度，進而達到整治完成之目的，因此，現地整治技術常被認為具有對場址環境衝擊低(無大幅開挖動作)、初設投入成本(建井與設備安裝)相對較低等優點而被選用，然而，亦往往容易因場址概念模式未即時修正、整治系統參數取得與應變不及時、現場專業人力不足等原因導致整治失敗或成效不彰，尤其現地整治系統需要一段操作期間，根據所取得的各類數據綜合研判場址概念模式是否正確並據以修正操作參數或模式，一旦數據取得缺漏或專業人力不足未即時完成數據分析與評估，便容易錯失修正時機，進而導致成效不彰並拖延數月甚或數年消耗無效之成本。

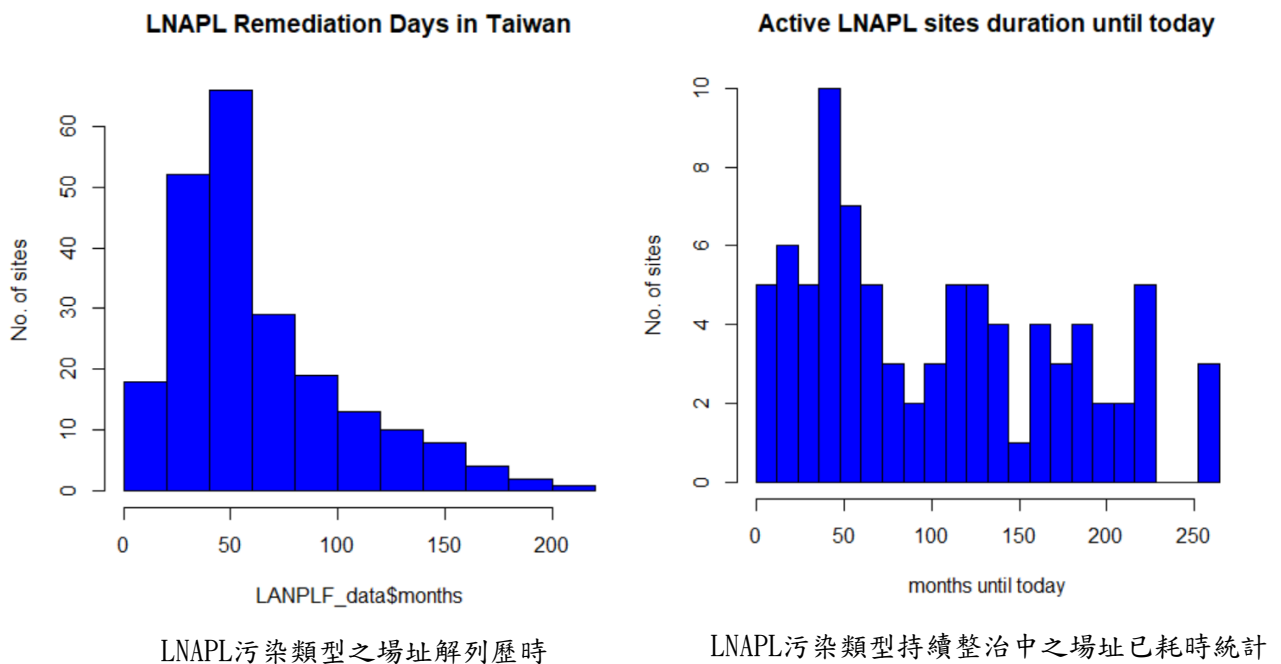
本工作團隊從環管署土污基管會列管場址查詢網站(<https://sgw.moev.gov.tw/SgwSiteInfo/SituationMap/>)，篩選有機類比水輕的污染物類型(LNAPL)進行統計分析，可以發現從已解列的污染場址取得之場址資料(篩選 228 處)中，平均從列管到解列歷時 64.09 個月，超過 25%的場址需要 81 個月以上完成解列；而從尚未解列迄今仍在控制或整治的場址(84 處)進行統計分析則發現，其平均之控制或整治消耗月份已達 103.9 個月，且因尚未解列，還需要持續進行整治，平均月份會持續增加(請參考圖 1-1 及表 1-1)。對於同類型污染物之場址存在不同的整治管理成效，影響的因素可能包括場址持續運營影響整治作業、場址污染量體大、場址整治團隊專業度差異、主管機關管制態度差異、場址所有人決策差異等等，其中，以平均歷時 60 個月能達成解列之場址或是平均 100 個月都未能達成解列之場址，其於控制或整治過程中所獲得的各類監測數據已相當龐大，無論是整治公司、業主、環保機關對於龐大持續增加的數據，與這些數據如何去體現仍屬於有效管理



以人工智慧結合遠端監控優化場址之現地整治成效

的評估方式都沒有一定的方法，即使有專家審查，也無法在短期數月間立刻取得整治成效提升的關鍵因素給與環保局或顧問公司相關建議，尤其不同場址仍存在不同變異特性，這些變異特性可能成為整治成效提高的關鍵因素或期間，往往因為數據欠缺評估分析而錯過黃金整治期間，因此猝然放棄一個有效工法或是決策錯誤的方向都可能導致該階段整治不成功而消耗時間與金錢，尤其以採取了現地整治(In situ)作為的場址而言，設備的運轉與操作是否有效，建立評估因子對於優化場址現地整治成效將有助益於業主、整治公司與環保單位進行決策

而美國 ITRC 於 2000 年初期即意識到整治工作中初始的場址特性與整治方案都存在很大的不確定性，因此需要在過程中採取更靈活且反覆運算的方法來管理不確定性，尤其隨著場址整治的時間愈久，可發現大部分資源都用於監控和運營和維護(如圖 1-2(a))，透過 RPO(Remediation Process Optimization)審閱，希望能降低成本及縮短完成時間(如圖 1-2(b))。因此，基於 RPO 的思維與目的，本計畫以 LNAPL 類型的場址為主，選擇過去曾應用土壤氣體抽除、雙相(氣液)抽除與土壤或地下水注氣技術進行土壤及地下水整治的場址，利用其歷史操作數據進行數據分析並建立成效預測模式，搭配具有傳感器並可利用物聯網(The Internet of Things, Iot)技術上傳並回控操作參數之整治設備，可實現蒐集更多即時數據應用不同演算方法進行分析，透過學習建立預測模式而讓設備操作人工智慧(Artificial Intelligence, AI)化，即實現人工智慧物聯網(AIot)的運作方式，進而達到提高整治成效，降低操作成本及達成節能減碳的目的。



數據來源：<https://sgw.moenv.gov.tw/SgwSiteInfo/SituationMap/>於本計畫以 R 分析

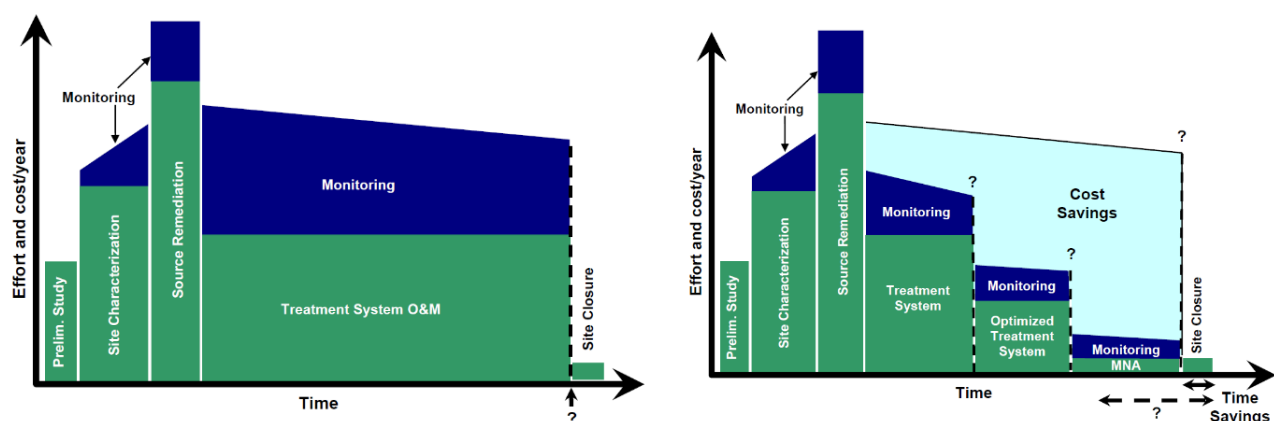
圖 1-1、台灣 LNAPL 污染類型解列與未解列歷時統計



表 1-1、台灣 LNAPL 解列耗時與未解列已耗時統計數據比較

| 場址類型 | 場址從公告列管日期到解除列管日期歷時(月數) | | | | | |
|------------------------|------------------------|------|------|-------|-------|-----|
| | 最小值 | 25%值 | 中位數 | 平均值 | 75%值 | 最大值 |
| 解除列管之 LNAPL 場址(228 筆) | 6 | 37 | 53.5 | 64.09 | 81 | 211 |
| 尚未解除列管之 LNAPL 場址(84 筆) | 1 | 44 | 92.5 | 103.9 | 159.2 | 263 |

數據來源：<https://sgw.moe.gov.tw/SgwSiteInfo/SituationMap/>於本計畫以 R 分析



(a) 傳統整治行動的時間與成效示意

(b) 在 RPO 下的整治行動時間與成效示意

資料來源：Remediation Process Optimization: Identifying Opportunities for Enhanced and More Efficient Site Remediation, 2004

圖 1-2、不同思維的整治行動與其成效示意

1.2 研究目的

本計畫主要便在以不同時間完成整治成功的場址數據，以機器學習的方法進行數據分析與建立預測模型，預期達到以下研究成果：

- 目前通行的場址概念模式(Conceptual Site Model, CSM)的建立，透過 2D 或 3D 成果進行展現的方式，雖然圖像化已更容易呈現污染的外貌，但是對於整治工作而言，概念模式與監測數據間的數據相符性則仍需要憑藉個人的專業、記憶與基礎的數據比較方法去判斷是否採取對應的動作，而針對取得數據的變異性，往往相隔一段時間才能發現變異的存在，其中，包括了數據本身的變異原因與場址概念模式間的變異原因，透過多次試錯的操作型態進而產生無法彌補的時間成本或是重要線索的丟失。因此，本計畫根據同類型污染物(LNAPL)之不同場址提供的場址概念模式(水文地質與污染特性)，以機器學習的方式去建立與現地整治系統操作參數相關的預測模型，並將操作參數與



以人工智慧結合遠端監控優化場址之現地整治成效

定期監測的各類數據回導修正預測模型，亦即以人工智慧輔助場址概念模式的修訂過程以優化整治成效。

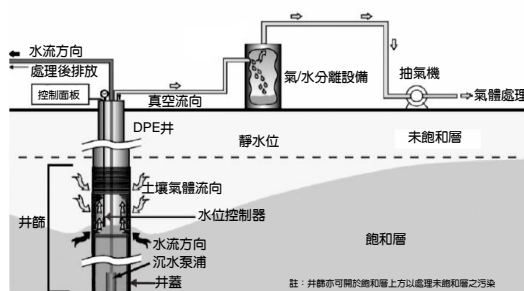
- 一般對於現地整治的場址成效評估常利用污染移除率、污染團削減率等之類的指標去評估整治成效，本計畫則將利用整治過程中取得的設備操作參數、定期監測數據或其他有效測試的數據，利用機器學習的方法去預測整治成效變化的趨勢與不確定性的量化，提供業主、環保單位與整治公司一個更為具體的判斷依據。
- 本計畫將透過遠端監控設備對於現場設備操作參數進行數據遠端擷取、分析與建立變異的預測模型，該項分析將最直接的影響移除量的污染因子導入到設備端的自動控制或是參數調整判斷提供現場操作人員的即時參考，以優化整治系統的操作性能。



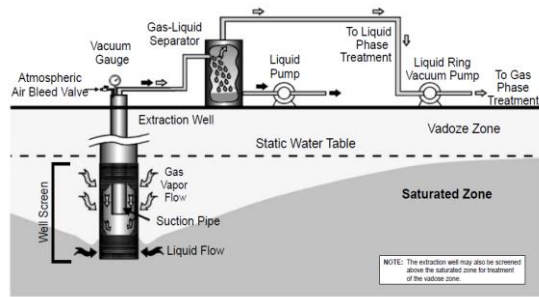
第二章 文獻回顧

2.1 雙相抽除處理技術原理與應用

參考美國陸軍 1999 年對於多相抽除處理技術(Multiple Phase Extraction, MPE)的定義為現地抽出超過一種以上的流體便稱之為多相抽除，其中定義了兩種抽除的方式，第一種為氣相與液相經由兩道管路從井中抽出，稱為 Dual Phase Extraction(DPE)(圖 2-1(a))，另一種為氣相與液相經由單一管路從井中抽出，稱為 Two Phase Extraction(TPE)(圖 2-1(b))；美國環保署 Solid Waste and Emergency Response 單位則對於 MPE 又給予不同的名稱定義，單泵 MPE 為單管伸入到水位面下進行抽水，雙泵 MPE 與前述 DPE 技術一致，新增了生物漱洗法(Bioslurping)的定義為類似單泵操作，但是伸入單管後放在水氣交界面；美國環保署 1995 年關於地下儲槽系統洩漏整治的技術指引(How to Evaluate Alternative Cleanup Technologies for Underground Storage Tank Sites)則將前述各類名詞都定義為 DPE 技術，並分成雙泵(Double Pump)DPE 技術與單泵(Single Pump)DPE 技術，本研究所稱的 DPE 技術即依據最後一個定義的內涵。



(a) Double Pump DPE



(b) Single Pump DPE

資料來源：U.S. Army Corps of Engineers, Engineering and Design Multi-phase Extraction, EM 1110-1-4010, 1999

圖 2-1、雙相抽除處理系統(DPE)示意圖

DPE 技術為一有效之土壤及地下水整治技術，主要用於抽除、處理含有 VOCs 及石油類碳氫化合物污染之土壤及地下水。其原理為利用單泵浦或雙泵浦將地下水及浮油抽出處理，同時造成雙相井旁地下水水位洩降，增加抽氣深度，再以機械設備如真空幫浦或鼓風機所提供之抽除能力，將含有揮發性有機物之土壤氣體藉由雙相井收集，經管線抽送至末端處理設備，經處理後排放至大氣中。影響整治成效的因子包括：

以人工智慧結合遠端監控優化場址之現地整治成效

1. 場址特性

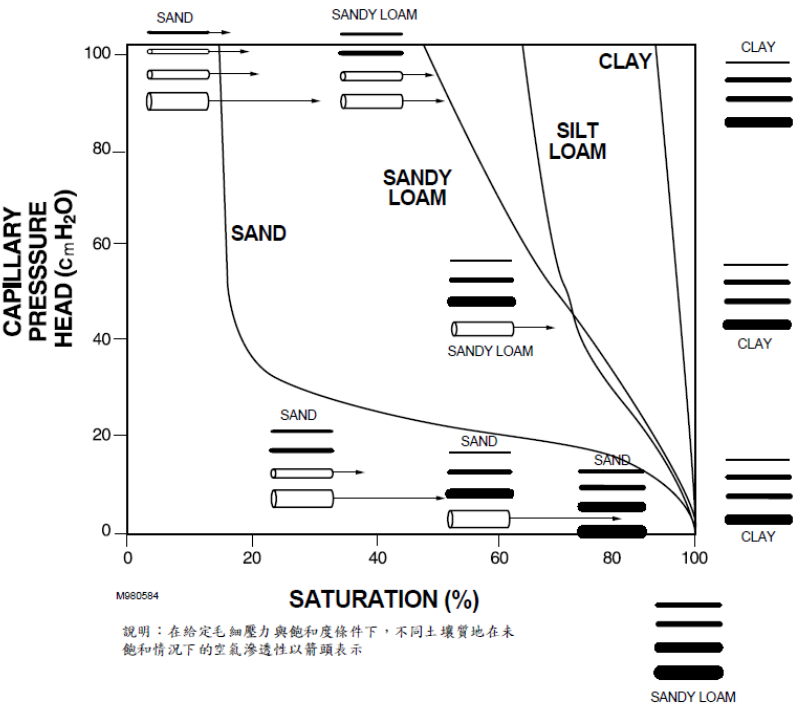
■ 滲透性

單泵浦 DPE 系統內在滲透性的 k 值最佳範圍為 10^{-9} ~ 10^{-11} cm²，大於 10^{-9} cm² 則不適用；雙泵浦系統則無上限限制； k 值可由水力傳導係數計算，參考 Rohit et al. 2016 的研究對於不同土壤質地的水力傳導係數分類如表 2-1 所示；而在空氣滲透率(Air Permeability)上，US Army 1999 年則對於各種土壤在飽和程度不同的情況下，空氣滲透性的範圍提供如圖 2-2。

表 2-1、不同土壤質地對應水力傳導係數範圍

| Type of Soil | Hydraulic conductivity K(cm/sec) | Remarks |
|------------------------|-------------------------------------|----------------------|
| Gravel | 1.00E-01 | Very High |
| Coarse Sand, Fine Sand | 1.00E-03<K<1.00E-01 | High to Medium |
| Silty Sand | 1.00E-05<K<1.00E-03 | Low |
| Silt, Silty Clay | 1.00E-07<K<1.00E-05 | Very Low |
| Clay | <1.00E-07 | Tends to Impermeable |

資料來源：Rohit Maurya, Umesh Kumar, Comparative study of various soils upon addition of different materials on the basis of hydraulic conductivity parameter, International Journal of Engineering Research & Technology (IJERT),2016



資料來源：U.S. Army, Engineering and Design of Multiphase Extraction,1999

圖 2-2、不同土層質地在給定毛細壓力與飽和度下的空氣滲透性



以人工智慧結合遠端監控優化場址之現地整治成效

DPE 在地下水位過淺(<90cm)的場址應用上較為困難，主要因為地下水位過淺的時候，抽取井容易造成垂直的短流，而減低水平方向的回收處理能力；但在 SVE 系統中因水位過淺於抽氣中所造成的湧升現象，在 DPE 系統則可獲得較佳的控制。

2. 化學特性

■ 有效揮發度

以下污染物條件適用於 DPE 抽氣系統移除：

蒸氣壓>0.5mm Hg

沸點< 250~300°C

亨利常數> 100 atm

■ 化學吸附能力

土壤的吸附能力可以用土壤-水分配係數來表示，愈高的吸附能力，污染物欲難抽除，相對地污染傳輸的風險亦較低。

3. 雙相系統單元

■ 抽取井水平佈設

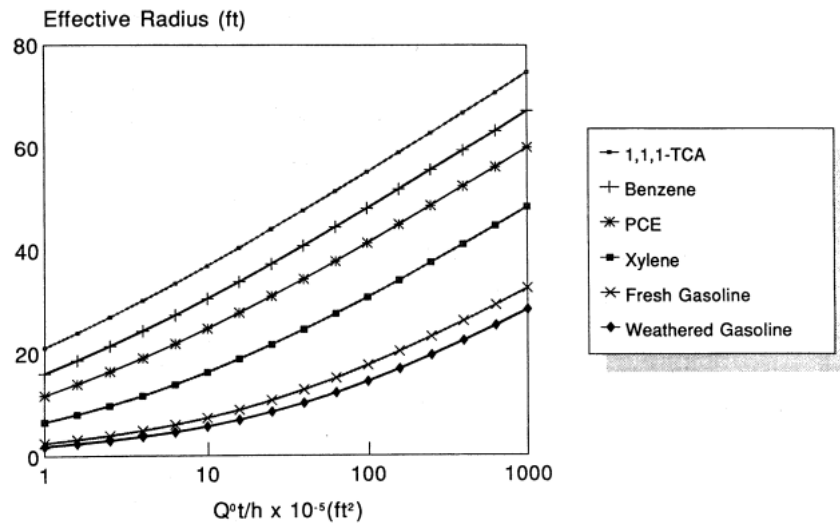
影響半徑的決定經常為現地整治系統中探討的重點，而基於土壤氣體抽除的影響半徑推估方式也相當成熟，如 1993 年 David H. Bass 提出影響半徑與真空度的關係如下式，同時提出當 $\frac{Q_0 \times t}{h}$ (Q_0 : 抽氣量， t : 抽氣時間， h : 有效開篩範圍) 為常數時，影響半徑將會一致，影響半徑受到待抽除的污染物影響較大(如圖 2-4)

$$P_r^2 - P_w^2 = (P_{RI}^2 - P_w^2) \frac{\ln\left(\frac{r}{R_w}\right)}{\ln\left(\frac{R_I}{R_w}\right)} \quad (2.1)$$

Jimmy H.C. Wang 於 1997 年發行的「Design of remediation systems」著作中建議影響半徑透過現場模場試驗可以取得，但是從經驗中觀察到不同的質地在搭配適當的系統條件下，可以達到的影響半徑如表 2-2 所示，而現場模場試驗中，建議以 0.01"H₂O 或抽取井的 1~10% 的真空度範圍作為影響半徑；而在相關的整治操作參數上，也有一些研究試圖能夠對於操作參數透過模擬提出相關的優化建議，其中包括如 Yan 等人於 2023 年對於 SVE 系統的設備與現場參數進行了實場的研



究並建立了數值模型，其提出各抽取井的流量與抽除設備的頻率呈線性關係；Michelle Matos de Souza 等人(2016)則簡化不同井剖面複雜的變化，以 peaceman 模型結合 STOMP 模擬 bioslurping 井的性能與操作預測情況，藉此提供整治操作中的評估參考；Shengqi Qi 等人研究利用數值模式假設抽除井的柱狀結構對雙泵 DPE 技術進行浮油回收的影響因子分析，得到在不同油厚情況，泵浦放置的最佳深度，及發現浮油回收泵的流量可以達到一個最佳化的程度下，浮油回收量不會再增加。



資料來源：David H. Bass, Sc.D., CHMM, Estimation of Effective Cleanup Radius for Soil-Vapor Extraction Systems, 1993

圖 2-4、不同物質在相同流量條件的抽氣有效影響半徑變化

表 2-2、不同土壤質地抽氣影響半徑

| Soil Type | Radius of Influence (ft) |
|-------------|--------------------------|
| Coarse Sand | >100 |
| Fine Sand | 60-100 |
| Silt | 20-40 |
| Clay | <20 |

資料來源：Jimmy H.C. Wong, Chin Hong Lim, Greg L. Nolen, Design of remediation systems, Lewis Publishers, 1997

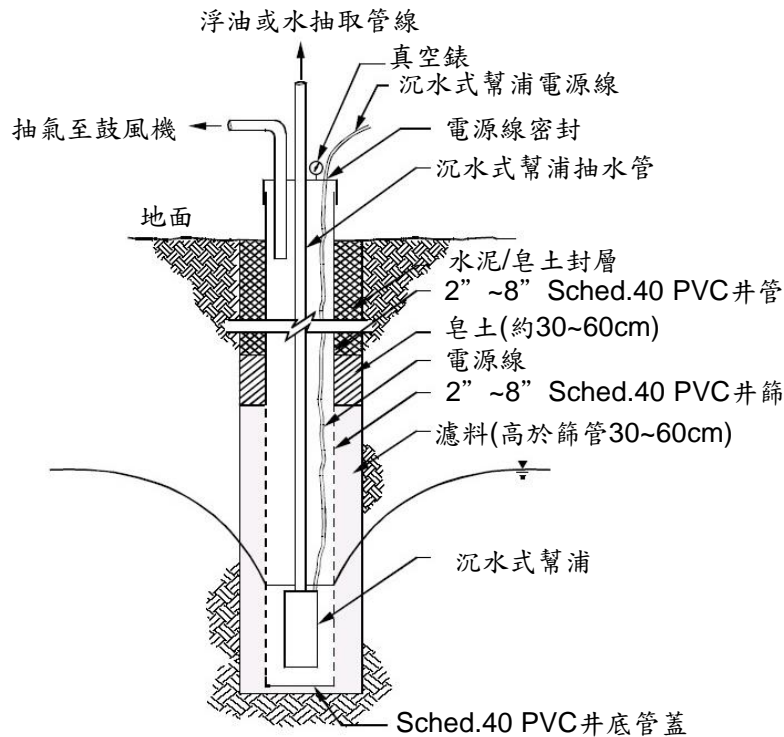
■ 設井規格

抽取井的管徑通常設計為 2~8 吋，抽取井的構造示意如圖 2-5 所示，通常抽取井開篩底部需足夠涵蓋枯水期低水位深度，未飽和層開篩長度是否涵蓋污染團



以人工智慧結合遠端監控優化場址之現地整治成效

厚度將影響到抽出氣的濃度、真空度與流量，進而導致整治成效好否。



資料來源：How to evaluate alternative cleanup technologies for underground storage tank sites-Aguid for corrective action plan reviewers, USEPA, EPA 510-R04-002, 2004

圖 2-5、雙泵浦式雙相抽除處理井設井構造圖

■ 管線系統

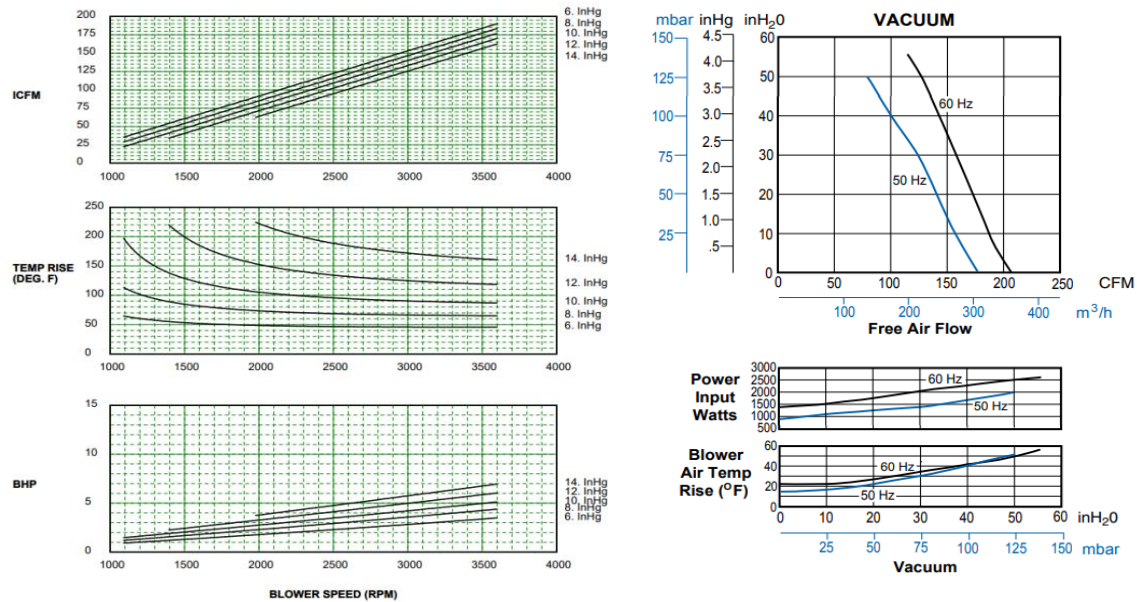
配管系統係指自抽取井與地面上整治設備如抽氣機組的連接系統，需注意配管時須自末端至抽取井傾斜以使地下水濃縮後可以回流至井內，管線系統設計會影響整治設備的動力輸出壓力損失，造成井頭與設備端的真空度的差異，為了減少管線壓力損造成不同抽取井的效能差異，一般建議不同抽取井與設備距離差異不要超過 30m。

■ 鼓風機選擇

在雙相抽除系統中的設備選擇上，一般雙泵系統的真真空泵選擇，氣相回收的部分與 SVE 設備相近，Jimmy H.C. Wong 提出鼓風機型式(Regenerative Blower)的真真空泵適用真空度小於 120" H₂O，魯式泵(Rotary lobe)則可以產生較高的真空度(10-20"Hg)；而參考美國陸軍的 MPE 操作指引，單泵系統的真真空泵通常選擇高真空度的液環式設備，真空度可以到達 29"Hg，以滿足較深深度的氣液相回收。而在實務上的應用，不同真空度能力的抽除設備主要取決於土壤質地、濕度，愈緊緻的土層特性，需要真空度能力更大的設備；而在真真空泵的動力選擇上，要選擇

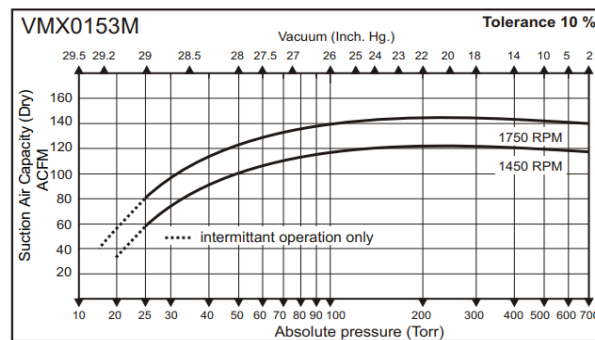


距離與數量適合的風量，原則上動力愈大，可以提供更大的抽除風量，可能在真空能力有些微差距，但真空能力仍與泵浦型式關係較為密切，圖 2-6 為本研究中場址內選用的泵浦型式及其馬達頻率與功能曲線圖。



(a) Regenerative Blower

(b) Roots Vacuum Pump



(c) Liquid Ring Vacuum Pump

資料來源：Gast、Howden、Dekker 廠商型錄，選擇風量配接 6 口井規格為參考型號之功能曲線

圖 2-6、不同型式抽取泵的馬達用電、轉速與功能曲線

■ 儀控系統

在控制系統上，DPE 多會採取變頻的控制調整輸出動力，以因應過程測試輸出不同的動力來測試抽取井的性能，也可以透過降頻操作減少抽出過多的液相。在 DPE 系統主要監控參數包括壓力(真空度)、流量、地下水抽取率、VOCs 濃度(PID)、CO₂/O₂ 濃度(生物降解反應)、污染量移除率、溫度等，土壤氣體抽出氣體理論流量如 2.2 式，在現場監控數據中可以做為對比修正的參考。



以人工智慧結合遠端監控優化場址之現地整治成效

$$Q_w = 2\pi R_w u_w H = H \left(\frac{\pi k}{\mu} \right) \left[\frac{P_w}{\ln \left(\frac{R_w}{R_I} \right)} \right] \left[1 - \left(\frac{P_{RI}}{P_w} \right)^2 \right] \quad (2.2)$$

而污染物移除量(kg)=抽氣濃度(C_{offgas} , kg/l)×平均流量(F , L/hr)×操作時數(t , hr)，公式如下：

$$C_{offgas} \left(\frac{kg}{l} \right) = \frac{C_{offgas}(ppmv) \times 10^{-6} \left(\frac{l}{l} \right) \times 56 \left(\frac{g}{mole(iC_4H_8)} \right)}{0.082 \left(atm \cdot l / ^\circ K \cdot mole \right) \times T(^{\circ}K)} \times 10^{-3} \left(\frac{kg}{g} \right) \dots (2.3)$$

$$F \left(\frac{l}{hr} \right) = F(scfm) \times 60 \left(\frac{min}{hr} \times \left(\frac{l}{ft^3} \right) \right)$$

2.2 地下水注氣技術原理與應用

空氣注入法為現地地下水污染整治方法，主要乃將空氣或氧氣注入飽和層中，促使地下水中高揮發性污染物揮發後降低地下水中污染物濃度，此外，亦會提高地下水中溶氧進而促進原生細菌的活性來分解有機物質，由於在飽和層中注氣，通常空氣注入法會併同土壤氣體抽除處理方法以控制氣相污染團的移動，一般空氣注入法搭配土壤氣體抽除處理方法之系統配置示意如圖 2-7 所示。影響整治成效因素說明如後：

1. 場址特性

■ 內在滲透率(Intrinsic Permeability, k)

空氣注入系統 k 值在大於 10^{-9} cm^2 時效果最好，但如小於 10^{-10} cm^2 則不適用。

■ 土壤分層組成

土壤中孔隙分佈會影響注氣壓力及於飽和層中空氣散佈情形，土壤分層不均質情形會造成空氣短流會部份區域無法到達。

■ 地下水溶解鐵濃度

地下水中亞鐵離子容易因曝氣導致被氧化生成鐵氧化物造成井管阻塞，一般當 $\text{Fe}^{2+} < 10 \text{ mg/L}$ 效果最好，但如 $\text{Fe}^{2+} > 20 \text{ mg/L}$ 時，則不建議採用空氣注入氣法。

2. 污染物特性

■ 化學結構



低分子量(含碳數<9)的脂肪族及單環芳香化合物較易被微生物分解。

■ 蒸氣壓

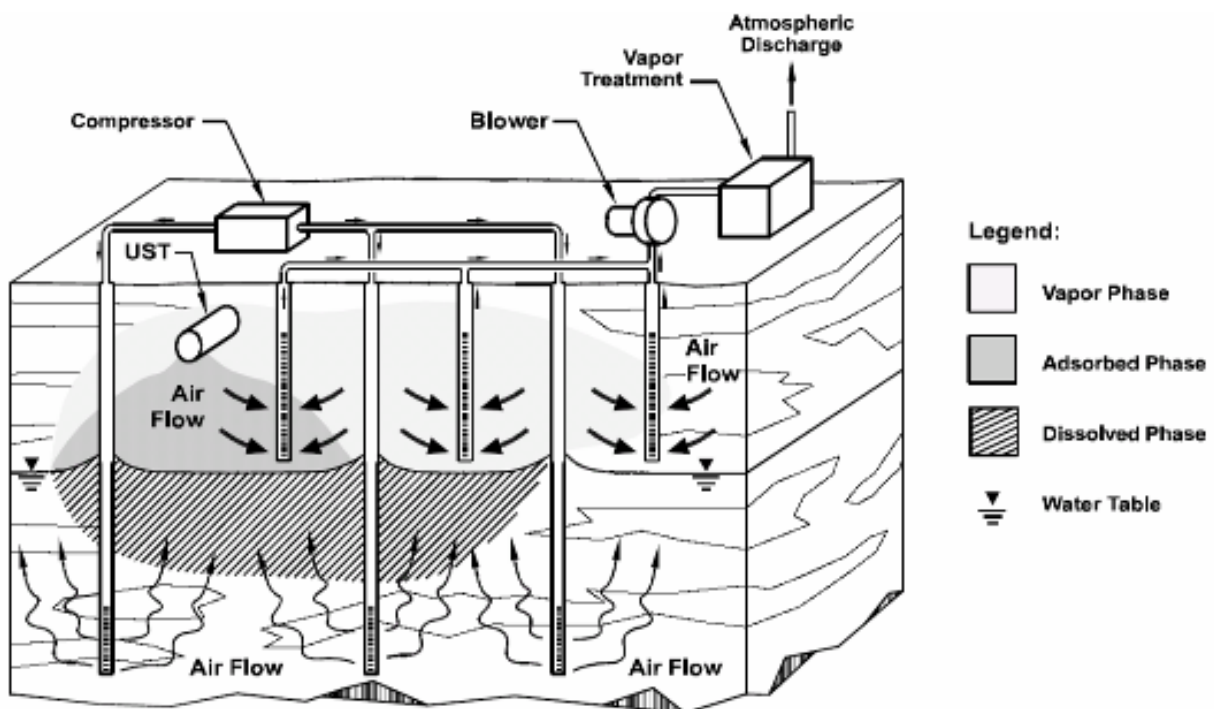
污染物之蒸氣壓大於 0.5 mmHg 容易因注入氣體而揮發，反之則以生物分解作用為主。

■ 污染物組成及沸點

一般沸點低於 250~300°C 表示容易揮發，此範圍包括所有汽油及部份柴油或煤油之成分。

■ 亨利定律常數

亨利常數愈高，污染物較易由水溶液態揮發至氣態，通常以 100 atm 為一參考值。



資料來源：How to evaluate alternative cleanup technologies for underground storage tank sites-Aguid for corrective action plan reviewers, USEPA, EPA 510-R04-002, 2004

圖 2-7、空氣注入系統配置示意圖

3. 系統單元說明

■ 曝氣井

圖 2-8 為一般曝氣井設置規範，一般尺寸可以為 1"~5" PVC 材質井管，其中

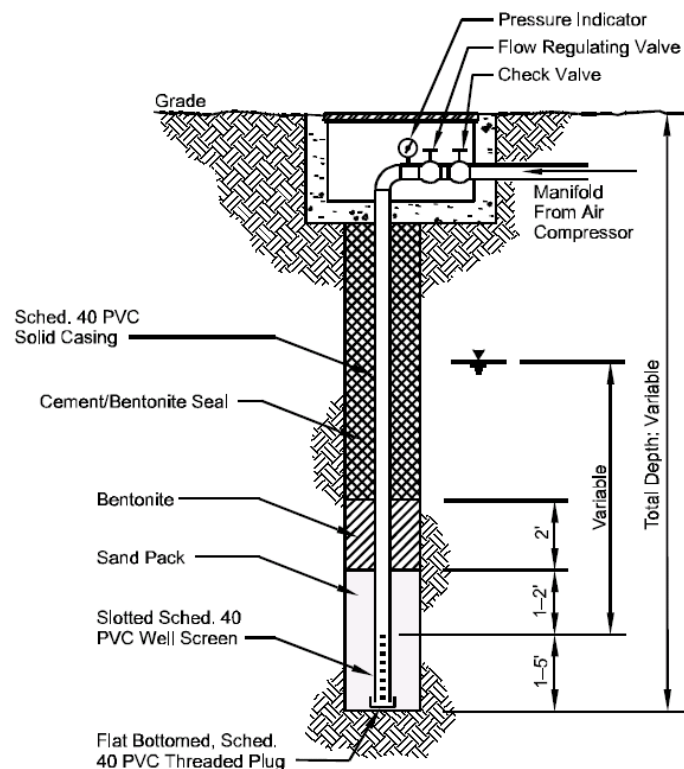


以人工智慧結合遠端監控優化場址之現地整治成效

開篩範圍建議在距離含水層污染團下 5~15' 開始開篩，篩管長度約為 1'~3'，注氣壓力一般在 10~15psig，為避免造成曝氣井損壞，注氣壓力避免不超過以下上限 (Battelle, Final air sparging guidance document, NFESC Technical Report TR-2193-ENV, 2001)：

$$P_{fracture}(psig) = 0.73 \times D_{soil}[ft] \quad (2.4)$$

$D_{soil}[ft]$ ：地表到開篩篩段頂部的距離



資料來源：How to evaluate alternative cleanup technologies for underground storage tank sites-Aguid for corrective action plan reviewers, USEPA, EPA 510-R04-002, 2004

圖 2-8、曝氣井設置規範

■ 空壓機

空壓機一般選用無油式較好，但如選擇一般空壓機需加裝去油過濾裝置以確保不會有二次污染物注入地下水中；一般各單井曝氣量約為 3~25 scfm，其搭配抽氣量需為注氣量的 1.25~5 倍範圍都算合理，進行曝氣操作時，SVE 抽氣流量與地下真空度皆會發生變化，此時，應找尋最佳化條件操作 SVE/AS 系統。

■ 配管等其他設施



一般建議配管方式與前述 SVE 系統相近，但由於接空壓機端壓力較大，建議選用金屬管材；為控制注氣量，於配管中須設置壓力表、流量表、控制閥等，為有效進行後續系統控制，在控制系統上可設計計時器以啟動注氣系統。搭配 SVE 的抽除量與抽除氣體濃度可以評估 AS/SVE 系統的整治成效，而在沒有 SVE 的操作下，單純 AS 的操作只具備生物降解的影響因子，可以採用以下方式評估降解效果：

$$R_{gw} = V_{soil} \cdot n \cdot O \cdot \frac{10^3 L}{m^3} \cdot \frac{10^{-6} kg}{mg} \cdot \frac{0.33 kg - HC}{kg - O_2} \quad (2.5)$$

R_{gw} ：在氧氣傳輸至地下水下污染物的好氧生物降解速率(kg/day)

V_{soil} ：處理區域體積(m^3)

n ：孔隙率(L-pores/L-soil)

O ：氧氣輸送到地下水的速率($mg - O_2/L - water/d$)

2.2 土壤及地下水調查與整治應用機器學習方法

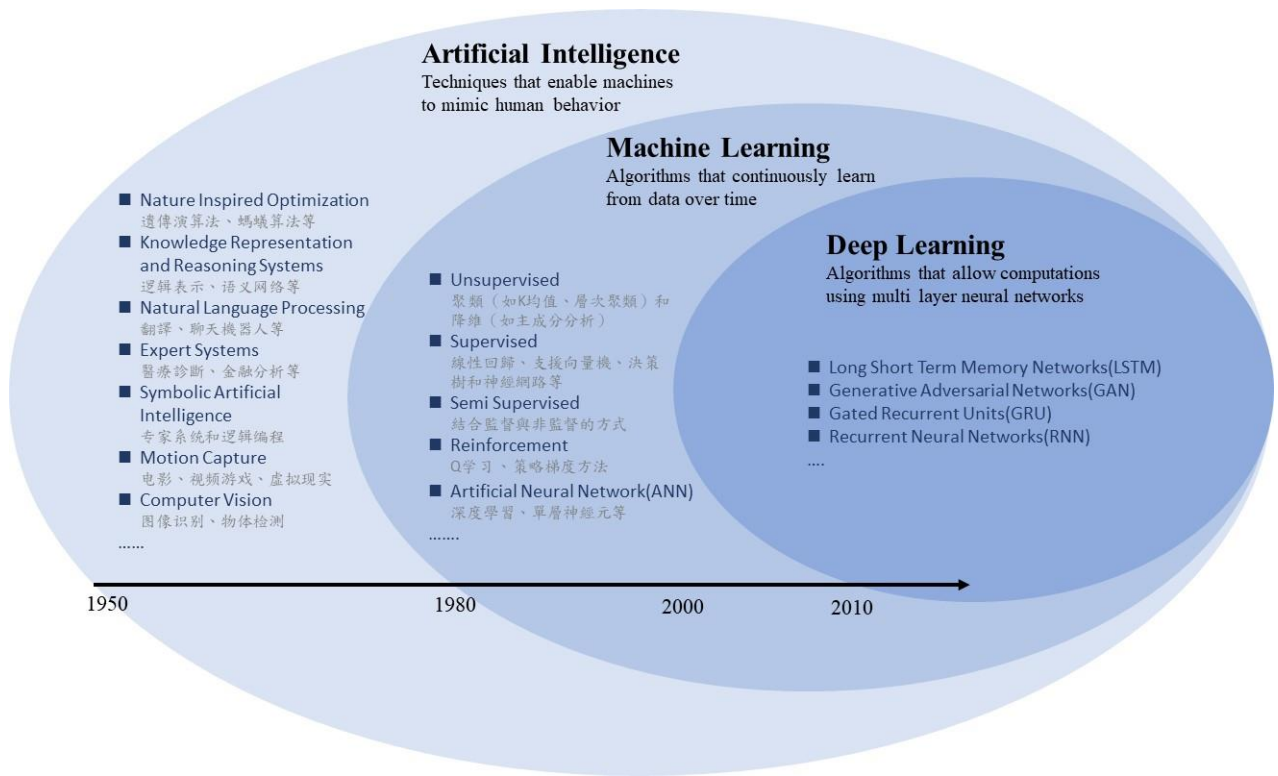
人工智慧技術(Artificial Intelligence, AI)自 1950 年發展以來，迄今已經 70 幾年，隨著電腦技術發展，從 1980 年代開始機器學習的技術發展到 2010 更多深度學習技術的研究，讓人工智慧的發展與應用層面演變得更加寬廣，Jagadeesh 等人(2023)將前述演變的技術內容製作如圖 2-9 的說明，並將機器學習與深度學習涵蓋的技術內涵也進行了彙整(如圖 2-10(a))，其中，從人工神經網絡(Artificial Neuron Network, ANN)方法衍生的更深層的學習方法被稱為深度學習(如圖 2-10(b))，也是目前影像辨識、語言分類等 AI 技術採用的學習技術，他可以處理複雜非線性的多維度參數的問題，獲得良好的預測結果。

Ryan Haggerty 等人(2023)利用 AI 預測地下水污染物的模型中，主要以基礎水質參數之研究較多(如圖 2-11)，其彙整了在建立場址特徵與整治技術優化等方面機器學習的相關研究如表 2-3 及表 2-4；從彙整的表 2-4 中可以看到主要的技術領域包括了 P&T、界面活性劑沖排(SEAR)、現地生物整治(In situ Bioremediation)、電動力(Electric Kinetic)、植物復育(Phytoremediation)等，對於 LNAPL 應用的 DPE 技術則著墨甚少，而從表 2-4 可以看到對於不同應用的目的，選擇適當的模型是關鍵的步驟，其中包括評估模型是否適合數值或分類數據，並做出相應的選擇，不同的模型有其優點和缺點，必須選擇適合特定問題的模型，此外，還應考慮模型的複雜性，因為過於複雜的模型會導致過擬合，而過於簡單的模型可能無法捕捉到問題的複雜性。通過選擇正確的模型，可以確保模型能夠準確有效地解決問



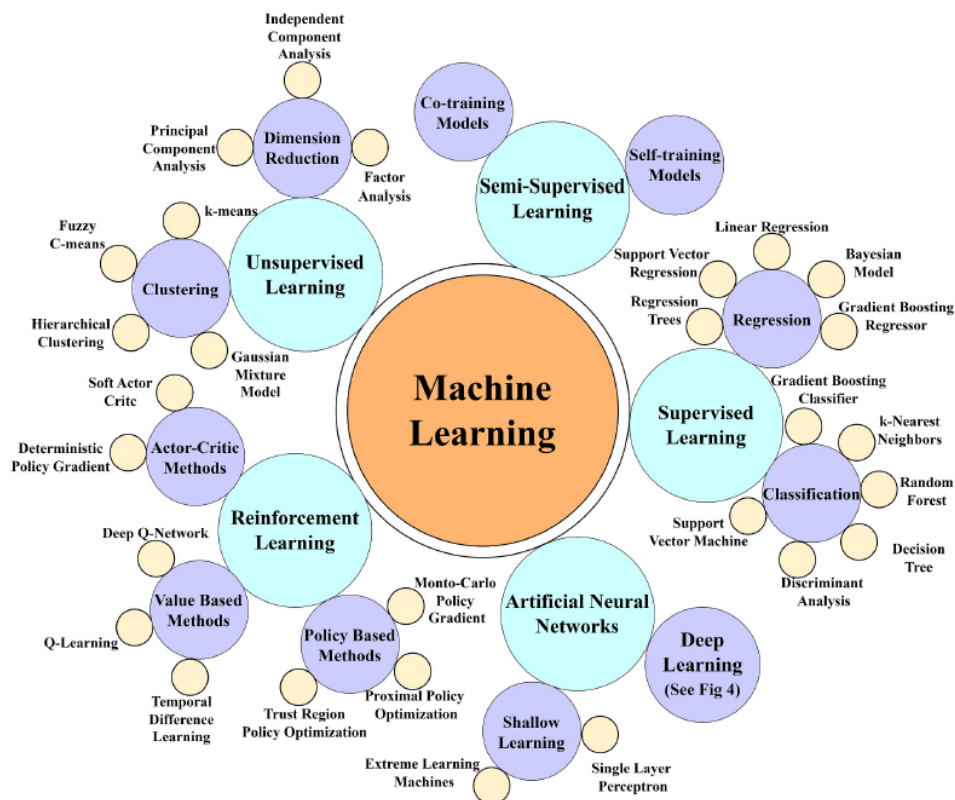
以人工智慧結合遠端監控優化場址之現地整治成效

題，從而獲得更可靠和準確的結果。可以使用各種模型精度參數來評估模型精度。其中，對於從土壤剖面、水文特性轉換為整治井功能上可以參考相關研究進行建立，例如 Shengquan 等人(2006)在地下水整治中井位的優化上利用 Adaptive Neural Network Genetic Algorithm (ANGA)預測優化後的井位得到與傳統數值模型 85-90%準確的結果，Michelle Matos de Souza 等人(2016) 利用 STOMP 模擬器評估生物漱洗井(Bioslurping)的性能進而預測移除浮油的成效，Shengqi Qi 等人(2020)利用數值模式預測 MPE 系統中浮油回收的優化、Junqi Huang 等人(2017)利用解析解推估 SVE 系統中的氣相傳輸模式、Christian Wolfsteiner 等人(2003)建立了井效指標用以評估井對於不同相的回收模型、M. TH. VAN GENUCHTEN(1980)利用簡單的土壤中含水壓力曲線建立了未飽和層水力傳導係數的預測模型、Hung-Kl Yen 等人(2003)針對 bioslurping 的場址浮油回收成效進行了模擬、Samanta 等人(2019)針對 MPE 單泵的抽取管(Bioslurping)深度利用 STOMP 模擬回收量變化；而在成效指標的模型建立上，X. S. Qin 等人(2008) 利用 SCA(stepwise cluster analysis)為基礎的方法預測浮油回收量再以 GA(Genetic Analysis)演算對多重目標進行成效計算進而提出成效優化的方法、Yan Shi 等人針對燒煤電場在數據上先利用 CFD(Computational Fluid Dynamics)模型產出計算數據做為輸入數據的補充後，再利用 ANN(Artificial Neural Networks)預測熱效率與 NOX 排放量，再利用 GA 的演算法對於操作參數的優化等，對於數據量不足或有缺漏情況下的模擬方式，可以提供本研究之參考；Pyo 等人(2020)利用了深度學習 Convolutional autoencoder 方法降低維度去預測重金屬污染情形、Cozad 等人(2014)利用替代模型(Surrogate model)的方法處理複雜、計算昂貴或耗時的模擬或實驗不可行時，提供更快、更高效的替代方法。本計畫將優先以人工神經網絡(ANN, Artificial Neuron Network)中多層感知器(Multilayer Perceptron)及徑向基函數(Radial Basis Function Network, RBFN)的方式作為深度學習的方法，如果有缺漏數據的情況下，將考慮以線性回歸、數學模型等方式推演後作為輸入數據，以提高深度學習的效能。



資料來源：Jagadeesh Kumar Janga , Krishna R. Reddy , K.V.N.S. Raviteja , Integrating artificial intelligence, machine learning, and deep learning approaches into remediation of contaminated sites: A review, Chemosphere 345 (2023) 140476

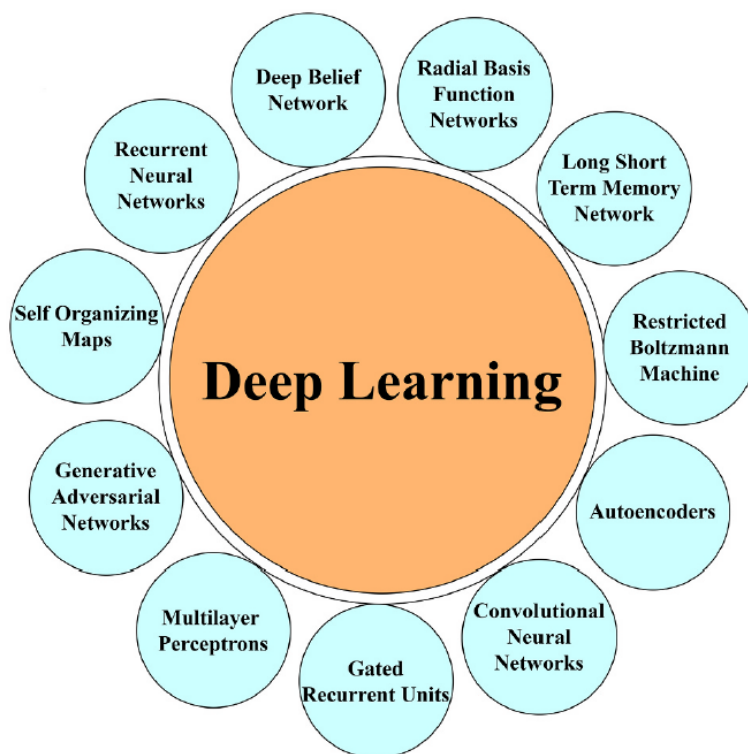
圖 2-9、人工智慧與機器學習涵蓋分析方法的體系圖



(a)機器學習涵蓋方法體系圖



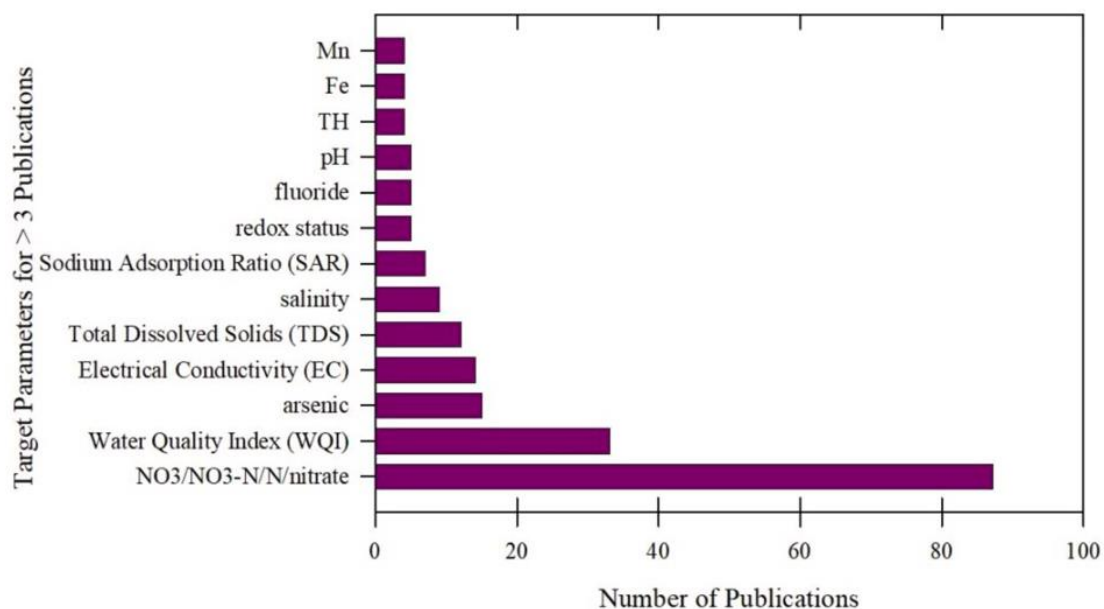
以人工智慧結合遠端監控優化場址之現地整治成效



(b)深度學習涵蓋方法體系圖

資料來源：Jagadeesh Kumar Janga , Krishna R. Reddy , K.V.N.S. Raviteja , Integrating artificial intelligence, machine learning, and deep learning approaches into remediation of contaminated sites: A review, Chemosphere 345 (2023) 140476

圖 2-10、機器學習與深度學習涵蓋方法的體系圖



資料來源：Ryan Haggerty , Jianxin Sun , Hongfeng Yu , , Yusong Li , Application of machine learning in groundwater quality modeling - A comprehensive review, Chemosphere 345 (2023) 140476

圖 2-11、涉及 AI 分析地下水水質類型之研究數量分析

表 2-3、利用 AI 建立場址特徵評估模型的相關研究列表

| 研究污 染 物 | 使用模 型 | 模型效能 評估因子 | 研究目的 | 研究成果 | 文獻來 源 |
|------------------------------------|---|---|--|--|-------------------------------|
| 1,2,3 trichloropropane (TCP) | CART, RF, and BRT | R ² , RMSE, and MAE | 在沒有地下水監測數據的情 況下，利用如歷史土地利 用、溶解氧和共污染物-硝酸 鹽數據，預測 TCP 的空間分 佈。比較各種模式的效能 | BRT, and RF models 相較 CART 更好，R ² 為 0.44, and 0.41 ， CART 為 R ² = 0.020. | Hauptma n et al. (2023) |
| DEHP | Bi-LSTM, kNN, and RF | Confusion matrix | 開發一種基於深度學習的新 型模型，以 預測研究區 DEHP 的空間變化 | The LSTM 為基礎的 模式顯示有較好的效 能 | Zheng et al. (2022) |
| PFAS | RF, and LR | Spearman’s correlation coefficient, and AUC | 為預測 PFAS 的汙染情形建立 ML 模型，採用共污染物指紋 圖譜、靠近機場和軍事設施 以及其他地表和地下特徵等 特徵做預測，進一步評估地 下水調查優先順序 | 透過機器學習建立模 式減少了 70%的監測 井數量 | George and Dixit (2021) |
| PAHs | SVR GA for optimizatio n | RMSE, MAE, and MAPE | 建立 SVR 模型，用 TPH 來預 測 PAHs 濃度 | 與研究中的其他優化 方法相比，使用 GA 方法和高斯核函數的 SVR 模型產生了最準 確的預測。研究表 明，TPH 濃度可以作 為預測 PAH 濃度的可 靠手段，而不會影響 預測精度。 | Akinpelu et al. (2020) |
| As | RF, ERF, SVM, MLP | Accuracy, Precision, Recall, F1 score, and Cohen’s Kappa coefficient | 使用高解析度航空成像 (HRAI) 和 ML 繪製土壤 As 污染的風險水平圖 | 與其他三種 ML 演算 法相比，ERF 給出了 更好的預測，平均分 類準確率為 0.87。 | Jia et al. (2021a) |
| Cr | GRNN, and MLP with and without kriging | MAE, RMSE, Spearman’s rank correlation coefficient, | 用 ML 與殘差克利金法混合 模型預測鉻濃度分布 | 使用殘差克里金法估 計 ML 模型殘差有助 於消除異常的高預測 和低預測，並且與純 基於 ML 的模型相比 可能更勝一籌。 | Tarasov et al. (2018) |



以人工智慧結合遠端監控優化場址之現地整治成效

表 2-3、利用 AI 建立場址特徵評估模型的相關研究列表(續)

| 研究污染 物 | 使用模 型 | 模型效能 評估因子 | 研究目的 | 研究成果 | 文獻來 源 |
|--------------|--|---|---|--|------------------------------|
| Cr(VI) | XGBoost kNN for imputation of missing data | R^2 , MSE, and RMSE | 使用貝葉斯搜索交叉驗證方 法優化的 XGBoost 模型預 測 Cr (VI) 等污染物對地 下水的長期污染 | Optimized XGBoost 預測模型在訓練時達 到 $R^2 = 0.99$ ，測試 時達到 $R^2 = 0.85$ | Mazumdar et al. (2022) |
| Cu,Mn,Ni | GRNN, and MLP with and without kriging | R^2 , RMSE, Spearman's correlation coefficient, Willmott's index of agreement, IQ | 利用 ML 模型和地統計克里 金法構建的混合模型預測重 金屬污染的空間分佈 | 與基本 ML 模型相 比，採用的混合方法 確實提高了預測準確 性 | Sergeev et al. (2019) |
| As,Cu,Pb | CACNN, CNN, ANN, RFR PCA for dimensiona lity reduction | R^2 , and RMSE | 採用深度學習方法估計重金 屬污染 | CACNN 提供較好的 估計結果，而 ANN 和 RFR 則較差 | Pyo et al. (2020) |
| Fe, Mn, Zn | MLP | Sum of Squares Error, RMSE, and Relative Error | 在 Ida 國家公園使用宏觀元 素和海拔水平數據估計和預 測重金屬水準 | ANN 可以從土壤的 參數中有效預測污染 物濃度 | Sari et al. (2022) |
| Heavy metals | DL with nearest neighbor neural network | RMSE, α , β , and TErate | 用 DL(ResNet)為基礎的空間 內插模型比較傳統克里金法 | 深度學習演算法通過 為污染物濃度的空間 插值提供更高的精 度，為基於克里金法 的插值提供了一種強 大的替代方案 | Man et al. (2021) |
| Heavy metals | RF | R^2 , ME, MSE, and RMSE | 決定沿海開墾土壤中重金 屬濃度分布的重要因子 | RF 模型成功建立 了不同因子(土壤礦 物組成、有機物組 成、化學特性等)對 重金屬分布的重要 性。 | Zhang et al. (2021) |



表 2-3、利用 AI 建立場址特徵評估模型的相關研究列表(續)

| 研究污染 物 | 使用模 型 | 模型效能 評估因子 | 研究目的 | 研究成果 | 文獻來 源 |
|--------------------------------|--|----------------------------------|---|--|--|
| Heavy metals | ANN, BP- FFNN | R^2 , MSE, MAE, and RMSE | 確定最佳污染指數方法， 通過基於 ML 和 DL 的方 法評估地下水污染 | DL (BP-FFNN)相較 ML (ANN) 更好的 評估地下水污染 | Singha et al. (2020) |
| Heavy metals | MLR1, RF with fuzzy c- means | R^2 , and RMSE | 預測重金屬污染的空間變 化，識別關鍵影響因素， 確定風險等級，劃定風險 區 | RF 模型在評估污 染空間分佈和識別 關鍵影響因素方面 的有效性獲得證 實。成功應用模糊 c-means 來檢測和 勾勒風險區域已被 證明在可視化和精 確定位關注區域方 面很有價值。 | Chen et al. (2023) |
| Soil Microplastics (MPs) | SVR-RBF, BPNN, RF, RBFN, LSTM, XGBoost, CART, RR, and LASSO regression | R^2 , RMSE, and MAE | 評估和比較不同 ML 模型 在預測土壤 MPs 豐度方 面的準確性和適用性 | SVR-RBF 模型被 發現在預測 MPs 的 準確度最好，而 RF 為基礎的集成 模型能最好解釋出 影響 MPs 分布的環 境影響因子 | Qiu et al. (2023) |
| Boron | BPNN, SVM, and LR | MAE, and RMSE | 利用傳統的 ML 模型- SVM、線性回歸和基於 DL 的模型預測研究區地 熱源硼的濃度 | DNN 的預測效能最 好 | Tut Haklidir and Haklidir (2020) |
| Nitrate | BRT, SVM, and MDA | AUC, Kappa, and MSE | 利用 ML 預測模型產生污 染物風險地圖 | 所有三個模型和三 個模型的集合在預 測地下水污染 (AUC>80%) 和 風險水準方面表現 良好 | Sajedi- Hosseini et al. (2018) |



以人工智慧結合遠端監控優化場址之現地整治成效

表 2-3、利用 AI 建立場址特徵評估模型的相關研究列表(續)

| 研究污 染 物 | 使用模 型 | 模型效能 評估因子 | 研究目的 | 研究成果 | 文獻來 源 |
|-------------|-------------------------|--|--|--|-------------------------------|
| Fluoride | ELM, MLP, and SVM | R ² , RMSE, and MAE | 觀察 ELMs 模型在預測地 下水中氟濃度上與 MLP 及 SVE 效能比較 | ELM 模型在預測地 下水氟化物污染方 面優於 MLP 和 SVM 模型與線性、 多項式和基於 Sigmoid 的核函數 相比，基於 RBF 的 ELM 給出了最 好的結果。ELM 模型的計算效率也 更高 | Barzega r et al. (2017) |
| Pb,Mn, etc. | MLR2, and RF | Confusion matrix, F1 score, Accuracy, Precision, and Recall | 利用氮指數等相關變數， 預測地下水污染情況，劃 定風險區空間分佈。 | RF 模型相較 MLR 更好，準確度達 93%. | Madani et al. (2022) |
| Radium | ANN, SVM | RMSE ^b | 開發基於各種探測器演算 法組合的優化檢測系統， 以提高檢測「熱」顆粒的 準確性 | 與傳統的計數率演 算法相比，ML 演 算法可以通過專注 於光譜形狀的變化 來提高檢測限。與 SVM 相比，ANN 尤其給出了更好的 結果。 | Varley et al. (2015) |

資料來源：Ryan Haggerty , Jianxin Sun , Hongfeng Yu , , Yusong Li , Application of machine learning in groundwater quality modeling - A comprehensive review, Chemosphere 345 (2023) 140476

表 2-4、利用 AI 建立場址整治優化模型的相關研究列表

| 要優化的 整治技術 | 整治對 象 | 採用模 型 | 模型效能 指標 | 研究目的 | 數據 | 成果 | 參考文 獻 |
|--|--|--|--|--|---|---|-----------------------------------|
| Pump, Treat, and Inject (PTI) | TDS in groundw ater | SFL and ANN models as surrogate s and GA for optimizat ion | R ² and RMSE | 優化泵壽命以 降低TDS濃度 與成本 | MODFLOW - 2000, and MT3DMS | SFL模型並未明顯 影響到泵壽命但是 明顯降低模式的運 算時間，以ML為 基礎的替代演算法 可以有效優化整治 決策 | Sadegh fam et al. (2019) |
| Pump and Treat | Groundw ater contamin ated with CCl ₄ | 3-D CNN | Precision, Accuracy, Sensitivity, and Specificity | 利用過去效能 數據去預測井 未來的效能 | Field data and simulation data from custom made fate and transport model | DL模型可以有效 預測汙染團分布而 協助決策過程。 3D-CNN演算法架 構很靈活，易於擴 展更多變數 | Song et al. (2023) |
| SEAR | DNAPL contamin ated aquifer | PRS, RBFN, SVR, GP, and Kriging | R ² and RMSE | 借助五種不同 的代理模型創 建高效的整合 代理模型並驗 證其準確性。 驗證自適應順 序抽樣(ASS) 在整治策略優 化中的應用。 | Simulations using UTCHEM model | 集成替代模型 (surrogate)比單一 替代模型可以提高 準確性。ASS取樣 技術可以改善預測 成果可信度。 | Ouyang et al. (2017c) |
| SEAR | | RBFN, SVR and GA to solve optimizat ion problem | R ² , and Absolute errors, and Relative errors (mean and maximum) | 利用集合對分 析(set pair analysis)來構 建集成代理模 型以優化整治 策略 | Simulations using UTCHEM model | 集合對分析是提高 集成代理模型選 擇、構建和精度的 有效方法 | Hou et al. (2017) |



以人工智慧結合遠端監控優化場址之現地整治成效

表 2-4、利用 AI 建立場址整治優化模型的相關研究列表(續)

| 要優化的 整治技術 | 整治對 象 | 採用模型 | 模型效 能指標 | 研究目的 | 數據 | 成果 | 參考文 獻 |
|-------------------------------|--|------------------------------------|---------------------|---|---|--|--------------------------|
| SEAR | | 3-D CNN | R^2 , and SSIM | 採用CNN模 型，通過考慮 含水層的非均 質性來準確預 測DNAPL的 去除率，並識 別整治后的風 險區，以幫助 決策 | Simulations using UTCHEM model | 與傳統的模擬優化 相比，該優化策略 利用k值的多次實 現和源區架構，能 夠準確識別最優 解，速度提高了 99.8%。這種方法 還可以根據整治后 含水層中留下的 NAPL來劃定風險 區。 | Du et al. (2022) |
| In-situ bioremedia tion | Petroleu m contamin ated site | Fuzzy based optimizatio n | - | 開發基於模糊 規則的預測控 制系統，以優 化原位生物整 治過程，並使 用案例研究進 行演示 | Simulations using UTCHEM model | 開發基於模糊的預 測控制系統允許在 整個淨化過程中對 原位生物整治過程 進行在線、即時、 經濟高效和優化的 控制。它的主要優 點是可以即時處理 模擬模型中的不確 定性。 | Hu and Chan (2015) |
| In-situ bioremedia tion | Chlorinat ed ethenes | CART | AUC | 開發可以確定 最佳生物整治 策略的數據挖 掘方法並確定 現地還原脫氯 的影響因子 | Groundwater monitoring wells data | CART模型能夠有 效地對訓練集和測 試集的3個月脫氯 電位進行分類，真 陽性率分別為 75.8%和69.5%。 研究表明，使用數 據挖掘來確定影響 原位脫氯潛力的因 素。 | Lee et al. (2016) |

表 2-4、利用 AI 建立場址整治優化模型的相關研究列表(續)

| 要優化的 整治技術 | 整治對 象 | 採用模型 | 模型效 能指標 | 研究目的 | 數據 | 成果 | 參考文 獻 |
|---|--|---------|--|--|---|---|---------------------------|
| Electro-kinetic enhanced bioremediation | Chlorinated solvents in low permeable porous media | ANN | R ² | 建立代理模型到基於過程的數值模型，類比電動增強生物整治，並執行靈敏度和不確定性分析。 | Simulation results from process-based numerical model | 模型在預測相對分布面積上R ² >0.99。 | Sprocati and Rolle (2021) |
| Phytoremediation | PAH contaminated soil | BP-FFNN | R ² and RMSE | 預測Melilutusalba最大吸收多環芳烴的理想條件 | Experimental dataset obtained by performing pot experiments | ANN模型準確利用土壤特性準確預測了植物根部PAH的濃度，並依據ANN預測結果建議不同土壤pH調節方式 | Olawoyin (2016) |
| Phytoremediation | Cadmium contaminated soil | BP-FFNN | R ² , MSE, MAE, and Correlation coefficient | 預測Sinapis alba L用下水道污泥調整後對於Cd的吸收變化 | Experimental dataset by performing pot experiments | ANN的模型對於預測Cd去除率方面的準確度較佳，以模型確定污水添加比例對整治的最佳化操作。 | Jaskulski et al (2020) |
| Phytoremediation | Soil contaminated with Heavy metals | XGBoost | R ² , RMSE, F1 score, Precision, Recall, and accuracy | 用ML基礎模型預測影響植物萃取重金屬的影響因子 | Data from literature | 輸入土壤重金屬、植物特性後可以準確預測重金屬濃度、枝條產量、生物濃度因子、金屬萃取率及整治時間 | Shi et al. (2023) |

表 2-4、利用 AI 建立場址整治優化模型的相關研究列表(續)

| 要優化的 整治技術 | 整治對 象 | 採用模型 | 模型效 能指標 | 研究目的 | 數據 | 成果 | 參考文 獻 |
|--|---|----------------|------------------------------|----------------------------------|-------------------------------------|--|---------------------------------|
| Monitored natural attenuation | PAH contamin ated soil | RF, and LDA | Accuracy | 預測PAH在土 壤中的降解率 及影響因子 | Experimenta l mesocosm trials | RF模型比LDA模 型更能準確預測 PAHs的降解程 度，從不同變數的 相關性可以了解影 響降解的因子 | Sicariel lo et al. (2022) |
| ZVI based permeable reactive barriers | Chlorinat ed organic compoun ds | XGBoost | RMSE, MAE, and MAPE | 用ML模型作 為含鐵反應牆 的最佳化填充 成分 | Data from literature | 利用ZVI基材料的 粒徑、表面積、孔 徑等描述符和反應 條件，建立了 XGBoost模型，以 預測ZVI基材料的 動力學反應常數。 結果表明，比面積 是決定反應速率的 最重要因素之一 | Ren et al.(202 3) |

資料來源：Ryan Haggerty , Jianxin Sun , Hongfeng Yu , , Yusong Li , Application of machine learning in groundwater quality modeling - A comprehensive review, Chemosphere 345 (2023) 140476

2.3 人工神經網絡演算法

要採用何種模型達到解決研究目的主題是關鍵的步驟，其中包括評估模型是否適合數值或分類數據，並做出相應的選擇，不同的模型有其優點和缺點，必須選擇適合特定問題的模型，此外，還應考慮模型的複雜性，因為過於複雜的模型會導致過擬合，而過於簡單的模型可能無法捕捉到問題的複雜性。通過選擇正確的模型，可以確保模型能夠準確有效地解決問題，從而獲得更可靠和準確的結果。可以使用各種模型精度參數來評估模型精度。考慮到研究的新穎性以及相關文獻回顧中深度學習的方法對於非線性多維度的情境有比較好的表現，本計畫將採用監督式的人工神經網絡(ANN, Artificial Neuron Network)的方式作為機器學習的演算法，，初步選擇多層感知器學習方法(Multilayer Perceptron, MLP)與 Radial Basis Function Network（RBFN）兩種訓練模型應用於回歸性質的輸出數據，摘要說明兩種方法的應用：



1. 多層感知器學習方法(Multilayer Perceptron, MLP)

多層感知器 (Multilayer Perceptron, MLP) 是一種人工神經網絡 (Artificial Neural Network, ANN) 的基本類型。它由多個神經元層組成，每一層與下一層之間都有全連接的神經元。MLP 通常由三部分組成：輸入層、隱藏層和輸出層。

- 輸入層 (Input Layer)：接受來自數據源的輸入數據，每個輸入特徵都與輸入層的一個神經元相關聯。
- 隱藏層 (Hidden Layer)：位於輸入層和輸出層之間的一個或多個層。隱藏層的神經元對輸入進行非線性轉換，從而使得 MLP 可以學習和捕捉輸入之間的複雜關係。
- 輸出層 (Output Layer)：生成 MLP 的最終輸出結果。每個輸出神經元代表了一個預測值或類別概率。

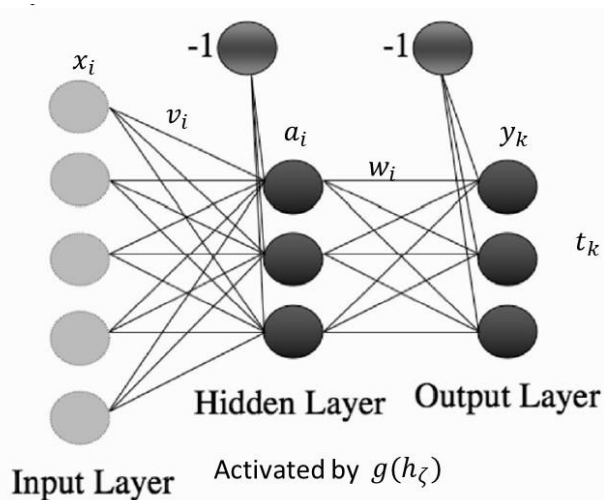


圖 2-12、多層感知器架構

MLP 的訓練過程通常使用反向傳播算法 (Backpropagation) 進行，該算法通過調整網絡中的權重和偏置來最小化預測輸出與實際輸出之間的誤差。MLP 是一種通用的機器學習模型，可以應用於各種任務，如分類、回歸、時間序列預測等。參考 Marsland 著作「Machine Learning: An Algorithmic Perspective」(2015)關於 MLP 的演算法如下：

- 初始階段：先給予每個 input 一個初始化的權重，這個值是包括正負的隨機值
- 訓練階段：

前向傳播演算(Forward Phase)：對於每個神經元先估計隱藏層(j)是否要被活化，



以人工智慧結合遠端監控優化場址之現地整治成效

導入活化函數，輸入層(i)、隱藏層(j)，輸出層(k)的權重與運算公式如下：

$$h_{\zeta} = \sum_{i=0}^L x_i v_{1\zeta} \quad (2.6)$$

$$a_{\zeta} = g(h_{\zeta}) = \frac{1}{a + \exp(-\beta h_{\zeta})} \quad (2.7)$$

同樣地，從隱藏層到輸出層也要生成銓重去運算最終出來的 y 值，

$$h_k = \sum_j a_j w_{jk} \quad (2.8)$$

$$y_k = g(h_k) = \frac{1}{a + \exp(-\beta h_k)} \quad (2.9)$$

回推演算狀態(Backward Phase)：

計算輸出值的誤差：

$$\delta_o(\mathbf{k}) = (y_k - t_k) y_k (1 - y_k) \quad (2.8)$$

計算隱藏層的誤差：

$$\delta_h(\zeta) = a_{\zeta} (1 - a_{\zeta}) \sum_{k=1}^N w_{\zeta} \zeta_0(k) \quad (2.9)$$

更新輸出層的權重：

$$w_{\zeta k} \leftarrow w_{\zeta k} - \eta \delta_o(\mathbf{k}) a_{\zeta}^{hidden} \quad (2.10)$$

更新隱藏層的權重：

$$v_i \leftarrow v_i - \eta \delta_h(\mathbf{k}) x_i \quad (2.11)$$

前述步驟演算至訓練完成(η 為學習速率，ML 的參數)

■ 調用(Recall)：使用模型的前向演算法演算進行預測

2. 徑向基函數(Radial Basis Function Network ,RBFN)

Radial Basis Function Network (RBFN) 是一種人工神經網絡 (Artificial Neural Network, ANN) 的特殊類型。與傳統的多層感知器 (Multilayer Perceptron, MLP) 不同，RBFN 具有一層徑向基函數 (Radial Basis Function, RBF) 作為隱藏層，以及一個線性輸出層。RBFN 的學習架構如圖 2-13，主要組成部分包括：



- 輸入層 (Input Layer)：接受來自數據源的輸入特徵向量。
- 徑向基函數層 (Radial Basis Function Layer)：這是 RBFN 的隱藏層，每個神經元代表一個徑向基函數。這些函數對輸入進行非線性映射，通常使用高斯函數來計算輸入和中心之間的距離。
- 線性輸出層 (Linear Output Layer)：這是 RBFN 的輸出層，它將徑向基函數的輸出加權線性組合，產生最終的輸出。

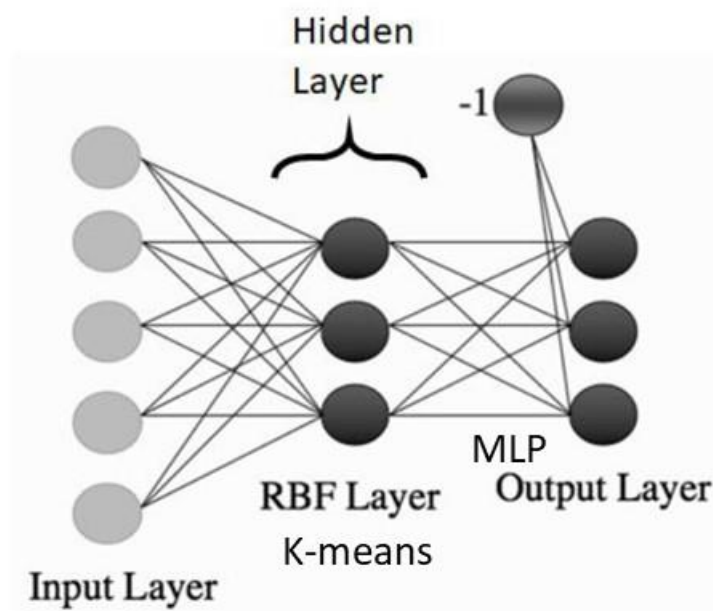


圖 2-13、RBF 訓練架構

RBFN 的訓練通常分為兩個階段：

- 中心初始化：首先，需要初始化徑向基函數的中心，通常是通過聚類算法（如 k-means）來實現。這些中心點表示了輸入空間中的重要特徵，以下列高斯分布型態計算出 RBF 元：

$$g(\mathbf{x}, \mathbf{w}, \sigma) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{w}\|^2}{2\sigma^2}\right) \quad (2.12)$$

- 權重學習：在初始化中心之後，通過最小化輸出與實際目標之間的誤差，輸出層以 MLP 輸出層的方式加上 bias 學習權重，這部分從輸出層得到的權重再返回 RBF 層。

RBFN 的優勢在於其良好的擬合能力和較快的收斂速度。它們通常用於函數逼近、模式識別、時間序列預測等應用中。然而，RBFN 的設計和訓練也具有一

以人工智慧結合遠端監控優化場址之現地整治成效

些挑戰，例如中心的初始化和選擇，以及訓練算法的穩定性。

在訓練過程中，還需要取得影響學習成效的參數，例如學習速度(Learning Rate)、迭代次數(Iteration,Epoch)、活化函數(Logistic、Tanh 等)，RBF 中則可以選擇除了高斯以外的分布函數特性去生成 RBF 隱藏層。從驗證集與測試集的效能可以了解預測結果的準確度，參考 Jagadeesh et al(2023)彙整用以評估學習效能的指標如表 2-5。

表 2-5、機器學習預測模型驗證誤差指標類型

| Various model-performance evaluation metrics used to evaluate AI/ML/DL based models. | | | |
|--|---|---|--|
| Metric | Formula | Metric | Formula |
| R^2 | $1 - \frac{\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2}$ | Root mean square error (RMSE) | $\sqrt{\frac{\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2}{n}}$ |
| Mean absolute error (MAE) | $\frac{\sum_{i=1}^{i=n} y_i - \hat{y}_i }{n}$ | Mean absolute percentage error (MAPE) | $\frac{1}{n} \sum_{i=1}^{i=n} \left \frac{y_i - \hat{y}_i}{y_i} \right \times 100\%$ |
| Mean square error (MSE) | $\frac{1}{n} \sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2$ | Willmott's index of agreement, d | $1 - \frac{\sum_{i=1}^{i=n} y_i - \hat{y}_i }{2 \times \sum_{i=1}^{i=n} (y_i - \bar{y})}$ |
| Relative absolute error (RAE) | $\frac{\sum_{i=1}^{i=n} y_i - \hat{y}_i }{\sum_{i=1}^{i=n} y_i - \bar{y} }$ | Relative error (RE) | $\sum_{i=1}^{i=n} \left \frac{y_i - \hat{y}_i}{y_i} \right \times 100\%$ |
| Precision | $\frac{TP}{TP + FP}$ | Recall | $\frac{TP}{TP + FN}$ |
| Accuracy | $\frac{TP + TN}{TP + FP + TN + FN}$ | Cohen's Kappa coefficient | $\frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$ |
| F1 score | $\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$ | AUC | Area under ROC curve |
| IQ | $Q_3 - Q_1$ | RPIQ | $\frac{IQ}{RMSE}$ |
| α | Probability of FP | β | Probability of FN |
| TErate | $\alpha + \beta$ | Spearman's rank correlation coefficient | $1 - \frac{6 \times \sum d_i^2}{n(n^2 - 1)}$ |

Notes: \hat{y}_i - predicted value, y_i - actual observed value, \bar{y} - mean of observed values, n-total no. Of test values, TP- True Positives, TN - True Negatives, FP - False Positives, FN - False Negatives, ROC - Receiver Operating Characteristic Curve, Q_1 - value below which 25% samples can be found, Q_3 - Value below which 75% samples can be found, d_i -difference in the ranks given to the two variable values for each item of the data.

資料來源：Jagadeesh Kumar Janga , Krishna R. Reddy ' , K.V.N.S. Raviteja , Integrating artificial intelligence, machine learning, and deep learning approaches into remediation of contaminated sites: A review, Chemosphere 345 (2023) 140476

3.支持向量機(Support Vector Machine，SVM)

SVM 的核心概念為尋找一個決策邊界（超平面）能夠把資料樣本分開，在二維空間中，超平面是直線；在三維空間中，它是平面；在高維空間中，它是一個更高維的結構。而運作上則是找到一個能最大化兩類樣本之間間隔（即邊界）的超平面，稱為最大間隔超平面，此處所稱的間隔為從超平面到最近樣本點（稱為支援向量）的距離，定義離決策邊界最近的樣本點為支持向量，決定了超平面的位置和方向，而這些支援向量對模型的最終決策有直接影響，對於線性可分數據，SVM 可以直接找到一個線性超平面將資料分開；對於線性不可分數據，SVM 使用核技巧（Kernel Trick）將資料映射到高維空間，在該空間中資料可能是線性可分的，常見的核函數包括：

- 線性核(linear)：適用於線性可分數據。
- 多項式核(polynomial)：通過多項式映射捕捉複雜的模式。



- 徑向基函數核 (RBF)：最常用的核函數之一，適合大多數非線性資料。
- Sigmoid 核：類似於神經網路中的啟動函數。

以下為 SVM 的演算法：

SVM Algorithm

The Support Vector Machine Algorithm

• Initialisation

- for the specified kernel, and kernel parameters, compute the kernel of distances between the datapoints
 - * the main work here is the computation $\mathbf{K} = \mathbf{X}\mathbf{X}^T$
 - * for the linear kernel, return \mathbf{K} , for the polynomial of degree d return $\frac{1}{\sigma} \mathbf{K}^d$
 - * for the RBF kernel, compute $\mathbf{K} = \exp(-(\mathbf{x} - \mathbf{x}')^2 / 2\sigma^2)$

• Training

- assemble the constraint set as matrices to solve:

$$\min_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T \mathbf{t}_i \mathbf{t}_j \mathbf{K} \mathbf{x} + \mathbf{q}^T \mathbf{x} \text{ subject to } \mathbf{G} \mathbf{x} \leq \mathbf{h}, \mathbf{A} \mathbf{x} = \mathbf{b}$$

- pass these matrices to the solver
- identify the support vectors as those that are within some specified distance of the closest point and dispose of the rest of the training data
- compute b^* using equation (8.10)

• Classification

- for the given test data \mathbf{z} , use the support vectors to classify the data for the relevant kernel using:
 - * compute the inner product of the test data and the support vectors
 - * perform the classification as $\sum_{i=1}^n \lambda_i t_i \mathbf{K}(\mathbf{x}_i, \mathbf{z}) + b^*$, returning either the label (hard classification) or the value (soft classification)

基於 SVM 對於非線性數據的處理方式，核嶺回歸法(Kernel Ridge Regression, KRR) 保留了線性回歸的方法及核技巧投射高維空間的作法，與 SVM 所不同的地方在於採取最小平方差做為損失參數，並採取了閉式解方式運算，求取損失函數的最小化解，不需到迭代過程即可找到參數，對於數據量較小的數據建立預測模型較為適用，過大的數據在求解過程會消耗太多時間，則不適用。

本研究終將利用 SVM 進行分類以評估整治效能與終端處理設備的選擇。並由於場址數據量仍在百到千的數據量，將嘗試 KRR 的方式進行回歸分析。



第三章 研究方法及步驟

3.1 整體研究流程

本計畫整體研究步驟如圖 3-1，首先先從已整治成功解列的案例進行數據彙整分析與訓練建立預測模型，再搭配本計畫組建的遠端監控系統擷取更多的數據繼續訓練模型，於第一年度計畫中如發現增加現地監測因子可以提高預測準確度，將於第二年計畫增設可遠端監控之監測因子，其中，第一年度可以得到初步的預測模型與初步訓練成效，第二年如增加監測因子後繼續優化模型進而最終確認達到本計畫預設目的的模型建立與應用。茲說明各項研究方法及步驟如後。

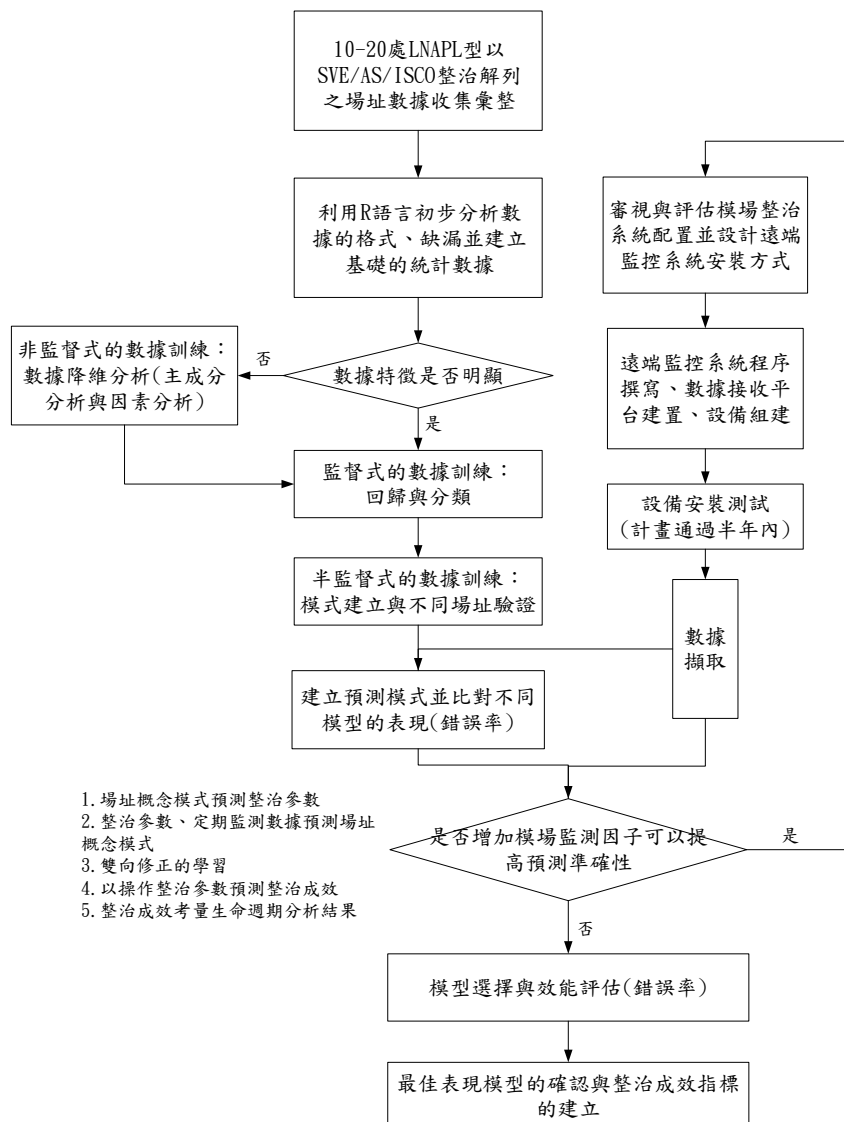


圖 3-1、本計畫整體研究步驟



3.2 前期訓練場址資料收集

本計畫將由合作單位新野科技股份有限公司提供其經由土壤氣體抽除處理、雙相抽除處理、地下水抽除處理、地下水注氣、土壤通氣等工法整治完成的場址資料進行前期的數據訓練，主要資料來源包括各場址之控制(整治)計畫、調查評估計畫、調查評估成果報告、控制(整治)計畫執行進度報告報告、控制(整治)計畫執行成果報告及整治過程現場紀錄、工作日誌等，涵蓋的資料類型如表 3-1 所示，前期已蒐集與模場試驗相似場址之資訊彙整如表 3-2，其中整治系統中的整治井數量可以了解該場址可能涵蓋的數據量規模；各場址雖都已經經過整治解除列管，但具體歷程不太相同，有的經過多次控制計畫變更，或增設整治井，或變更工法或變更操作條件後才在整治成效上產生明顯影響，因此，針對場址性質不同，本研究會因應時間階段不同拆分為訓練集、驗證集與測試集，以更好的擬合與選擇模型。本項工作將於第一年完成。不分場址以整治井抽除污染物對照單井尾氣數據約 1000 到 1500 筆數據，不同場址土壤質地與地下水頂空濃度數據約有數千筆，其中單一場址最多數據有數百筆，由於數據分散且格式不同，需要於分析工作展開前統整數據格式建入資料庫，部分以場址為座標系統的點位要轉換為二度分帶座標，以利後續數據計算與應用。針對有些特徵性質通用的部分，則需要給予共同可以在後續計算中標準化的參數，例如座標系統的統一、深度表示方式的統一、數值標準化，標準化的方式也會影響模擬結果，後續會在執行中將標準化的型式也納入研究中。參考表 3-1，本計畫預定建立的數據類型與項目包括：

1. 土壤現場篩測紀錄資料庫：監測日期、採樣點位座標(TM_X、TM_Y)、採樣點編號、採樣深度、土壤質地、土壤含水特性、氣味、是否有自由相，分段快篩 PID 篩測值、分段 FID 篩測值、送樣與否、若有送樣則有實驗室 BTEX、TPH 分析值。
2. 地下水水位量測記錄：監測日期、監測井位座標(TM_X、TM_Y)、警號井頂高程、監測井口徑、開篩範圍、井深、水位、PID 頂空濃度，此部分將彙整水位紀錄、頂空濃度量測記錄及地下水採樣紀錄，若有委託實驗室檢測之標準採樣，則彙整現場水質參數、實驗室檢測數據 BTEX、TPH。
3. 土壤氣體量測記錄：監測日期、監測井位座標(TM_X、TM_Y)、水位、井深、真空度、是否有油、抽出氣體之 PID、FID 檢測值。
4. 單井尾氣紀錄：整治井編號、座標(TM_X、TM_Y)、開篩範圍、井深、口徑、設井日期、馬達最大轉速、變頻器頻率、井內抽水泵控制深度、流量、真空度、溫度、尾氣 PID 或 FID 測值。



以人工智慧結合遠端監控優化場址之現地整治成效

5. 場址尾氣紀錄：量測日期、抽氣設備編號、型式、馬達最大轉速、操作頻率、累計時數、流量、真空度、溫度、變頻器頻率、井內抽水泵控制深度、抽出氣體之 PID 或 FID 值、放流水水量
6. 影響半徑試驗數據：量測日期、抽氣井號、真空度量測點位、真空度

其中，如果要建立場址概念模式與抽氣整治操作的關聯，必須將土壤與地下水性質與對於整治作為的影響與已關聯，初步規劃建立以下輸入項：

- 從土壤特性與含水程度建立影響半徑 ROI(參考表 2-2)，邏輯如下：

第一層：黏土 6m、砂(或礫石)10m、其他 8m

第二層：如紀錄中黏土帶有砂或石字樣者，+2m；砂(或礫石)無其他字樣，+5m

第三層：如含水紀錄為含水者，-2m；為濕潤者，-1m

- 原先紀錄中每一段 j (120cm)有多點的篩測值(i)，將依照權重進行平均，公式如下，可將原先同一段深度的篩測濃度平均為每一段固定深度(0~120cm、120~240cm，餘類推)

$$C_j = \sum C_i w_i, w_i = \frac{r_i}{\sum r_i} \quad (3.1)$$

同理，以上質地特性可以利用圖 2-3 與表 2-1 新增透水係數與毛細層壓力欄位作為該深度的水文地質特性。而單井尾氣測試井對應的土壤剖面與水位變化將採取以下方式進行建立：

- 初始水位(WT_{ini})：用以評估整治系統初始未飽和層抽氣範圍(開篩深度($h_{w,upper}$)到水位深度)，計畫採用抽除井於設井前後三個月內最近日期的水位，如果沒有抽除井水位數據的，則選擇最近的監測井水位數據。
- 整治前未飽和層油氣濃度(C_{ini}^{Unsat})：用以評估抽除井於啟動後可能影響的初始尾氣濃度，初始值將採用該抽除井於設井前指定距離(D_1)內土壤採樣剖面快篩數據的反距離權重平均值(通常為控制計畫調查數據)，所以將與初始水位進行如下計算：



$$C_{ini \text{ from soil}} = \frac{C_i R_i^{h_d}}{\sum C_i R_i^{h_d}}, \left\{ \begin{array}{l} R_i^{h_d} = 0, \text{ when } h_i^{bhb} < h_{w,upper} \text{ or } h_i^{bht} > WT_{ini} \\ R_i^{h_d} = 1, \text{ when } h_i^{bht} \geq h_{w,upper} \text{ and } h_i^{bhb} \leq WT_{ini} \\ R_i^{h_d} = \frac{h_i^{bhb} - h_{w,upper}}{h_i^{bhb} - h_i^{bht}}, \text{ when } h_i^{bht} < h_{w,upper} \text{ and } h_i^{bhb} \leq WT_{ini} \\ R_i^{h_d} = \frac{WT_{ini} - h_i^{bht}}{h_i^{bhb} - h_i^{bht}}, \text{ when } h_i^{bhb} > WT_{ini} \text{ and } h_i^{bht} < WT_{ini} \end{array} \right\} \quad (3.2)$$

$$C_{unsat \text{ of well } j} = \frac{\sum C_{ini \text{ from soil } i} \times \frac{1}{d_{soil \text{ to extraction well } j, i}}}{\sum \frac{1}{d_{soil \text{ to extraction well } j, i}}} \quad (3.3)$$

而已經進行整治中的井，則未飽和層油氣濃度可以從前述最新最近的土壤篩測數據或土壤氣體監測數據得到，若採用土壤篩測數據，則水位將依據該整治井最近距離監測井的最新水位數據為代表值進行計算；若採用土壤氣體監測數據，則選擇指定距離(D₂)內土壤氣體監測數據的反距離權重平均值。

- 飽和層油氣濃度(C_{sat,i})：以尾氣量測日期前最近一口監測井最新日期的地下水頂空濃度，亦選擇指定距離(D₂)內地下水頂空監測數據的反距離權重平均值。

表 3-1、本計畫預定取得數據與參數類型

| 控制計畫階段關於場址概念模式建立可已取得的數據 | 場址整治期間定期監測可取得的數據 | 整治設備操作可以取得的數據 |
|---|---|--|
| <ul style="list-style-type: none"> ➤場址基本資料 場址配置圖、製程與運作物質、洩漏源、洩漏量、洩漏歷史等 ➤水文地質調查數據 土壤質地 地層(土壤)剖面 水位流向資訊 地表高程 水力試驗數據 地下水監測井設置紀錄 ➤污染特性數據 土壤氣體檢測數據 土壤地下水檢測數據 土壤現場快篩數據 地下水現場量測水質參數數據 ➤模擬數據 土壤及地下水污染範圍 推估污染量、污染範圍 | <ul style="list-style-type: none"> 土壤、地下水定期監測數據 土壤質地 水位數據 土壤採樣現場快篩紀錄 浮油厚度紀錄 地下水採樣水質參數 土壤氣體監測數據 二次污染防治監測數據 回收浮油數量 推估污染移除量 整治井設置紀錄 | <ul style="list-style-type: none"> 設備馬力、種類 整治系統運轉型態(每日操作時數、時段、開口井號) 抽氣流量、真空度監測記錄(馬達變頻器頻率) 抽氣尾氣紀錄 單井尾氣紀錄 影響半徑試驗紀錄 其他監測參數紀錄 注氣流量、壓力紀錄 化學藥劑灌注紀錄 整治系統故障紀錄 設備運轉時數 抽水量記錄 其他系統異常紀錄 |



以人工智慧結合遠端監控優化場址之現地整治成效

表 3-2、本研究前期訓練場址彙整

| 編號 | 場址代號 (縣市) | 整治歷 時(月) | TPH 污 染量 kg | 地下水污 染面積 m ² | 土壤/含水層 質地 | 地下水位 (距地表) mbgs | 土壤污染物 (mg/kg) | 地下水污染物 (mg/l) | 整治系統 |
|----|----------------|-------------|----------------|----------------------------|------------------------|-----------------------|--|--|-------------------------------------|
| 1 | H10048(桃園) | 30 | NA | 1519 | 粉土夾礫石 | 1.4-3.2 | 苯：6.05 | 苯：25.4/甲苯： 35/總酚：3.11/ 萘：0.457 | SVE：6 DPE：4 AS：9 |
| 2 | F00015(新北) | 27 | 6700 | 465 | 黏土/粉質細 砂 | 1.3-3.2 | 苯：31.4 甲苯：789 二甲苯：1615 TPH：9323 | 苯：6.34/甲苯： 14.9/萘：0.404/ TPH：35.7 | SVE：6 DPE：14 AS：7 |
| 3 | J10038 (新竹) | 49 | 6061 | 337 | 粉土夾礫石/ 砂夾礫石 | 3.2-3.9 | 苯：102 甲苯：561 二甲苯：826 TPH：7920 | 苯：43/甲苯： 72.2/萘：0.484 TPH：75.5 總酚：0.309 | SVE：11 DPE：3 AS：15 |
| 4 | J10008 (新竹) | 38 | 3331 | 360 | 粉土/砂夾礫 石 | 7.8-8 | 苯：47.9 甲苯：648 TPH：5764 | 苯：1.47 萘：0.404 | P&T：2 SVE：11 AS：4 BV：2 |
| 5 | F10013(新北) | 51 | 10300 | 590 | 黏土/粉質細 砂 | 3-4.9 | 苯：16.7/乙苯： 262/二甲苯 1310/TPH： 9688 | 苯：14.6 | SVE：12 DPE：8 AS：2 |
| 6 | H10096(桃園) | 21 | NA | 1410 | 粉黏土/砂夾 礫石 | 3.2-4.1 | NA | 苯：5.02 甲苯：15 | DPE：16 BS：8 |
| 7 | F10066(新北) | 90 | 3166.3 | 730 | 粉土/砂 | 2.8-3.9 | 苯：56.4/甲苯： 1840/乙苯：599 二甲苯：2724 TPH：14200 | 苯：0.775 | SVE：5 DPE：25 AS：18 |
| 8 | P10037(雲林) | 30 | 6610kg | 478 | 砂 | 1.5~3m | 苯：29.2 二甲苯：883 TPH：8560 | 苯 0.564 | SVE:4 口 DPE:4 口 AS:8 口 |
| 9 | A00604(台北) | 45 | NA | 233 | 回填粉土/黏 土 | 0.4~1.3 | 苯：32.2 TPH：7090 | 苯：1.03 MTBE：1.29 | SVE：16 DPE：3 DPE/BV：34 |
| 10 | H10161 (桃園) | 24 | 1010 | 1028 | 粉黏土夾礫 石/礫石夾砂 | 1.5-2.5 | TPH：2240 | 苯：0.486 TPH：74.2 | SVE：5 DPE：8 AS：18 |
| 11 | A00615 (台北) | 18 | 9226 | 338.2 | 粉土/砂夾粉 質砂 | 2.4-2.8 | 苯：156/甲苯： 1120/乙苯：536 二甲苯：2467 TPH：13950 | 苯：0.421 MTBE：3.44 | SVE：8 BV/SVE：7 DPE：4 AS：12 |
| 12 | H10295 (桃園) | 62 | 1609 | 697 | 砂、粉土、 黏土夾層/砂 夾礫石 | 2~3m | 苯：8.99 二甲苯：618 TPH： 4280 | 苯：1.40 TPH：37.6 MTBE：2.37 | SVE:19 口 DPE:7 口 AS:22 口 |
| 13 | A00636 (台北) | 24 | 109 | NA | 砂夾礫石/砂 | 3.5-4.4 | 甲苯：678 乙苯：291 二甲苯：1470 TPH：13100 | NA | SVE：8 P&T：3 DPE：1 AS：8 |

| | | | | | | | | | |
|----|---------|--------|----|------|-----------------|---------|-------------------------------|---|--------------------------|
| 14 | PR1(新竹) | 12 | NA | 1900 | 細砂夾礫石/ 粗砂夾礫石 | 2-3.2 | 苯：120 二甲苯：521 TPH：15130 | 苯：11.6 萘：0.685 | SVE：2 DPE：4 AS：15 |
| 15 | PR2(台中) | 100/12 | NA | 797 | 粉土/卵礫石 | 2.7-5.1 | NA | 苯：0.984 萘：1.02 TPH：17.4 | SVE：18 AS：12 |
| 16 | PR3(桃園) | 24 | NA | 4539 | 黏土質細砂/ 砂夾礫石 | 1.4-4.7 | 苯：22.1 TPH：6240 | 苯：14.3/萘： 0.722MTBE： 186/總酚：0.155 | SVE：4 DPE：14 AS：56 |

NA：無資料

3.3 數據預處理

這是開發準確可靠的模型的關鍵步驟，隨機化數據可確保數據均勻分佈，這對於防止模型中的偏差很重要。數據清理對於刪除不需要的數據（如缺失值、行和列）以及重複項以及在必要時轉換數據類型也是必不可少的。利用研究課題的理論，提取對於目標有關的特徵參數後，數據進一步分為實際訓練算法的訓練集(Training set)、用於跟蹤學習效果的驗證集(Validation set)及用於產生最終結果的測試集(Test set)，訓練集用於訓練模型，驗證集用於調整參數，完全獨立的測試集則最終用於評估模型的性能，這樣的劃分可以幫助確保模型在新數據上的泛化能力，從而提高模型的實用價值。在數據充足的條件下，訓練集、驗證集與測試集的比例為 50、25、25，沒那麼充足的情況下，可以增大訓練集的比例到 60、20、20 或 70、15、15，在數據非常不足的情況下，可以採取交叉驗證（Cross-Validation）、留出法（Hold-Out Method）或信息標籤法（Information Criterion）等方式進行數據集的分配，其中，數據量與模擬中特徵的維度有關，如果維度多但是數據量不足，可能導致維度災難、過度擬合(無法泛化)的問題。本研究目標中涉及到雙相抽除系統的操作過程都與場址內水文地質特性有密切的關聯，本研究中會將場址內土壤調查剖面、地下水位等數據利用統計或數學模式的方式轉換為整治井的特徵後再進行後續的機器學習訓練。本研究中關於機器學習數據集的訓練流程如圖 3-2，原則上，數據量應至少滿足總權重數的 10 倍，因此，以目前收集到的數據量約可以有 100 到 200 個權重數，初步在學習的時候，還是先優先找到較佳的表現的參數，再從後續的測試或模場中修正，針對輸入層變數如過多，將採取降維方式減少變數以利機器學習上的合理性。本研究中將會先將所有參數進行變異數分析，了解各個參數的相關性，並研判數據的合理性，進行數據的檢查與回看是否需要修正或拋棄，原則上會先從物理上的意義去設定初始變數，如果從效能中發現變數不足(預測值明顯不發生變化)，將逐步增加變數或重新檢討變數的正規化方法，再重新學習檢視效能，此步驟係將物理意義與實驗方式結合透過 try and error 的過程直到獲得滿意的效能。

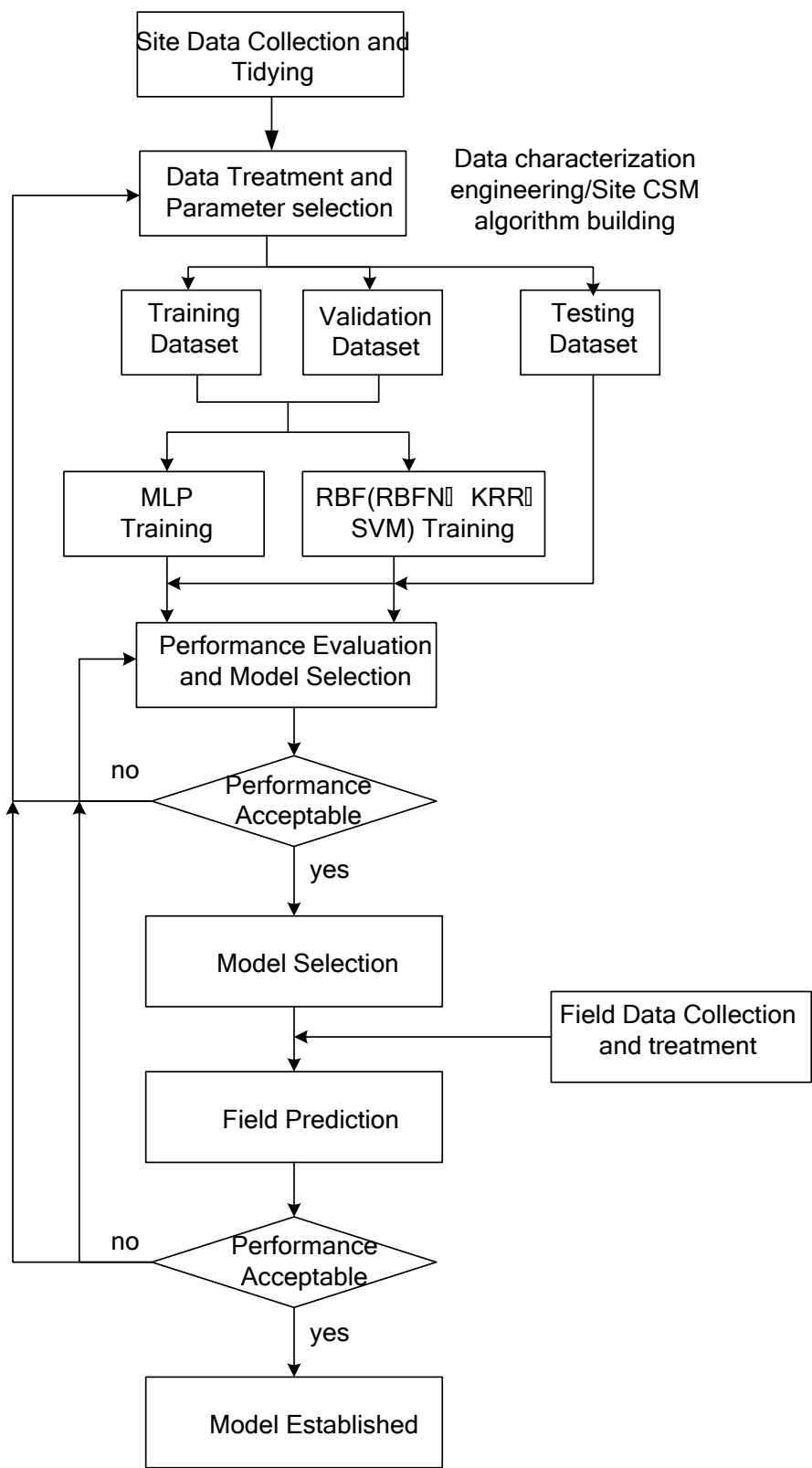


圖 3-2、模型建立與測試流程



3.4 模型選擇與訓練

要採用何種模型達到解決研究目的主題是關鍵的步驟，其中包括評估模型是否適合數值或分類數據，並做出相應的選擇，不同的模型有其優點和缺點，必須選擇適合特定問題的模型，此外，還應考慮模型的複雜性，因為過於複雜的模型會導致過擬合，而過於簡單的模型可能無法捕捉到問題的複雜性。通過選擇正確的模型，可以確保模型能夠準確有效地解決問題，從而獲得更可靠和準確的結果。可以使用各種模型精度參數來評估模型精度。其中，對於從土壤剖面、水文特性轉換為整治井功能上可以參考第二章相關參考文獻的參數值進行轉換與設定，針對不同時序的數據，除了目標值以外，輸入值在不同時序的紀錄初步規劃採取以下方式進行內插或選擇：

1. 基於場址內水文地質與污染分布特性有其時空上的關聯性，因此，對於在整治井尚欠缺調查數據時，將根據周邊土壤及地下水相關數據進行推估，原則上將採取反距離權重法推估水位、水文地質參數；至於在不同時間點下，哪些點位數據適合作為權重的組成，這部分將設定一定時間範圍內的點才可作為權重組成，以同時將時間與空間的變異考慮進去。而除了權重法，也會比對最近點的特性直接帶入，而在機器學習過程中，不需要取捨這些數據，可以直接納入輸入層進行訓練，再視訓練模型的效能進行取捨或調整
2. 輸入值的正規化操作也是機器學習重要的數據處理工作，正規化的方式很多，包括如下幾種，本研究中會變更正規化的方式比對模型效能：
 - 依照物理意義上的方式進行正規化，例如 $\log x$, $1/x$, \sqrt{x} 等
 - 利用統計學上的標準偏差範圍，以平均值或最大值-最小值等取得比例
 - 以相關標準值為正規化的參考數
 - 以該參數可對比的全範圍值，例如 1atm，馬達性能轉數、電力頻率規格等
3. 這是典型機器學習工作流程的最後一步，使用看不見的數據對模型進行訓練和測試后，可以使用它對新數據進行預測。新的數據亦須要確保數據與用於訓練和測試的數據具有相同的品質。本研究中採行遠端監控的數據進行評估，期間傳感器的性能需經過定期的檢驗與校正確認數據可信度。



以人工智慧結合遠端監控優化場址之現地整治成效

3.5 模場試驗場址資料

基於本計畫將先針對已成功解列完成且相關整治歷程資料完整的場址進行機器學習訓練，再應用到場址上進行模場試驗，因此，選擇的模場需同時為LNAPL類型場址及適合採用現地土壤氣體抽除、雙相抽除及空氣注入進行整治的場址，因此，本計畫原先選擇位於台中港區的控制場址進行後續模場試驗，惟因該場址經歷本計劃期間的尾氣濃度已經明顯下降，後續持續監測恐對於研究成果沒有助益，故於計劃期間變更場址至桃園縣為 H10047 場址（新屋鄉台 15 線 53 公里 500 公尺處）崁頭厝段下庄子小段 710、710-10、710-12、711、711-1、712、713 地號，茲就該場址的相關資訊說明如後。

1. 場址基本資料

H10047 場址位於新屋區省道台 15 線 53.5 K 兩側之空地，接近永安漁港，鄰近用地多為埤塘或草生空地，過去為農地及畜牧用地。污染事件發生於民國 96 年 8 月 1 日，中油公司埋設於台 15 線 53 公里 500 公尺處之 12 吋汽、柴油輸油管線遭竊油導致油料外洩污染，應變人員巡查水路，同時於漏油處附近設置數處集油坑及攔截溝，以攔截外洩油料，阻止油料隨排水溝渠擴散。民國 96 年 8 月 9 日~10 月 5 日間，以回收油灌車抽除油水則估計約有 7,950 公秉，其中約有 350 公秉為純油品、7,600 公秉為受污染之地下水。桃園市政府環境保護局（下稱桃園環保局）於民國 96 年 8 月 6 日以及、8 月 27、28 日，分次派員前往現場進行土壤及地下水採樣檢測，發現地下水中苯與甲苯濃度超過地下水污染管制標準，土壤中苯與總石油碳氫化合物濃度超過土壤污染管制標準，隨後桃園環保局於民國 96 年 9 月 27 日行文要求採取適當措施並提送污染改善計畫書，中油公司於民國 97 年 2 月 27 日提送本場址污染改善計畫書至桃園環保局，並經多次審查後於民國 98 年 4 月 2 日核定通過，改善期限為 30 個月，以浮油回收、DPE 法、AS 法及抽出處理法等改善。改善期限屆滿，桃園環保局於民國 101 年 10 月進場查證發現場址土壤中污染物苯及總石油碳氫化合物（TPH）等 2 項污染物，濃度達土壤污染管制標準（苯：5 mg/kg，TPH：1,000 mg/kg）；地下水中苯及柴油總碳氫化合物（TPH-d）等 2 項污染物，濃度達第二類地下水污染管制標準（苯：0.05 mg/L，TPH-d：10 mg/L），故依土污法第 12 條第 2 項及同法施行細則第 10 條公告本場址為土壤及地下水污染控制場址，中油公司於 102 年 10 月 21 日提送「新屋區台 15 線 53.5 K 土壤及地下水污染控制計畫書（第一版）」，並民國 103 年 6 月 11 日核定污染控制計畫至民國 107 年 12 月 10 日執行完成，但因場址合約因素，並未能如期完成，再於 110 年 1 月 4 日核定控制計畫書（第二次變更），其中受到國營單位招投標程序、疫情影響等因素未能順利發包廠商執行工作，再於 110 年 11 月



提出場址控制計畫(第三次變更)申請，並於年月核定，目前場址正處於執行第三次變更期間。

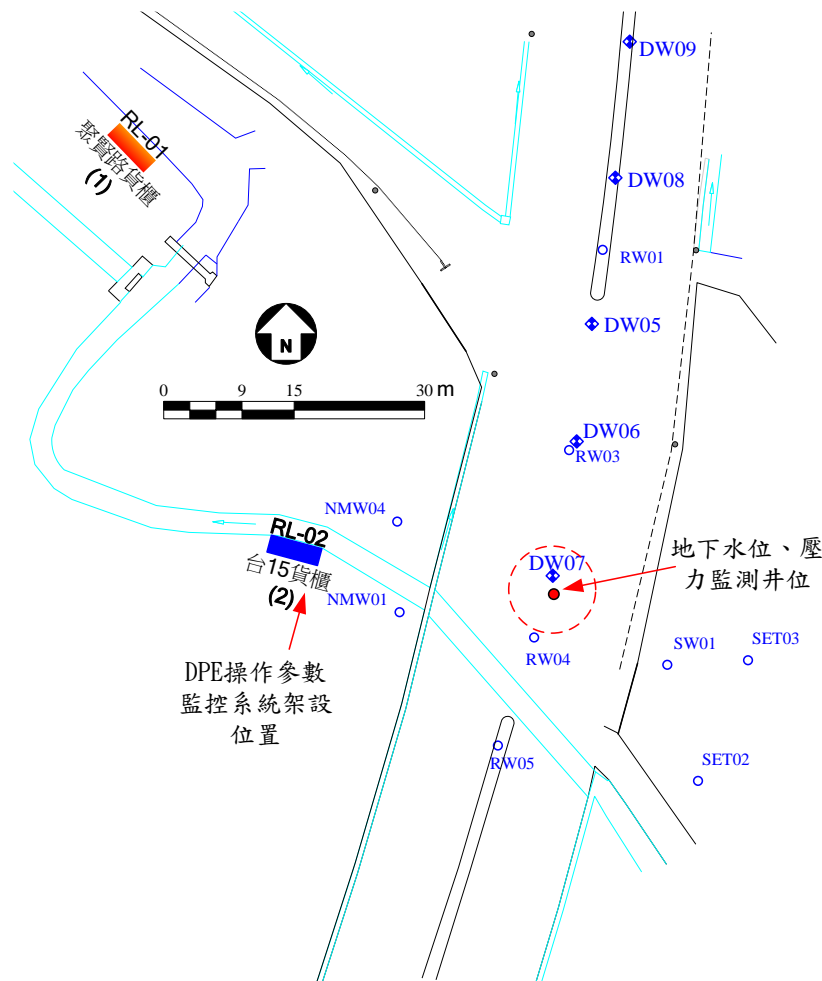


圖 3-3、本計畫模場試驗預定位置

2. 場址水文地質特性

本場址於民國 104 年 12 月依調查佈點規劃進行 5 點地質鑽探，調查結果如圖 3-4 所示，範圍內 G03、G04 及 G05 井孔區域地質發現地表下約 0~1 公尺為回填之粉土，夾砂夾礫及磚塊回填層，地表下約 1~2.4 公尺則為透水性不佳之黏土層；場內多數地層為桃園中壢地區之紅土礫石層，地表下約 1.45~6.85 公尺範圍為主要油品污染範圍，污染物較易隨裂隙流布；地表下約 6.85 公尺上下有一層厚約 40 公分上下之不透水黏土層存在，且 G03 與 G04 範圍內該黏土層連續性佳，因此污染往更深層之可能性較低。

以人工智慧結合遠端監控優化場址之現地整治成效



圖 3-4、本計畫預定場址之地質特性

地下水文以場內現有監測井（共計 17 口）以水位計進行地下水量測，並將結果以 Surfer 軟體繪製地下水流向，地下水流向如圖 3-5 所示，本場址地下水位大約介於地表下 1~2 公尺，地下水流場大致向東及向北流動，地下水受場址旁埤塘補助之影響明顯，水力傳導係數約為 $1.1\times10^{-4}\sim3.2\times10^{-4}$ cm/sec。



資料來源:本場址污染控制計畫書(變更)

圖 3-5、本場址地下水流向推估圖

3. 場址污染特性

依據本場址 105~107 年控制計畫定期監測及 108 年、109 年土壤污染補充調查結果，本場址主要為 TPH 之污染物，然因本場址於過去進行整治作業時，於場內設置相關鋼板樁及連續壁(CCP)等地下阻隔措施，且場址內多處已進行開挖作業，故以克利金(Surfer 軟體)方式推估污染範圍已不符現況情形，因此土壤污染範圍係以現地鋼板樁或連續壁(CCP)設置區域進行污染區域劃分，其推估範圍如圖 3-6 所示，污染範圍包括溝渠北側區域、工務所區域及北側窪地區域，整體土壤污染面積推估約 1,935m²，地下水污染範圍部分，本公司於 109 年 10~12 月發現台 15 線道路上地下水中 TPH 污染物超過第二類地下水污染管制標準，污染範圍台 15 線道路區域，地下水污染面積推估約 1,935m²，污染範圍推估如圖 3-7 所示。

以人工智慧結合遠端監控優化場址之現地整治成效

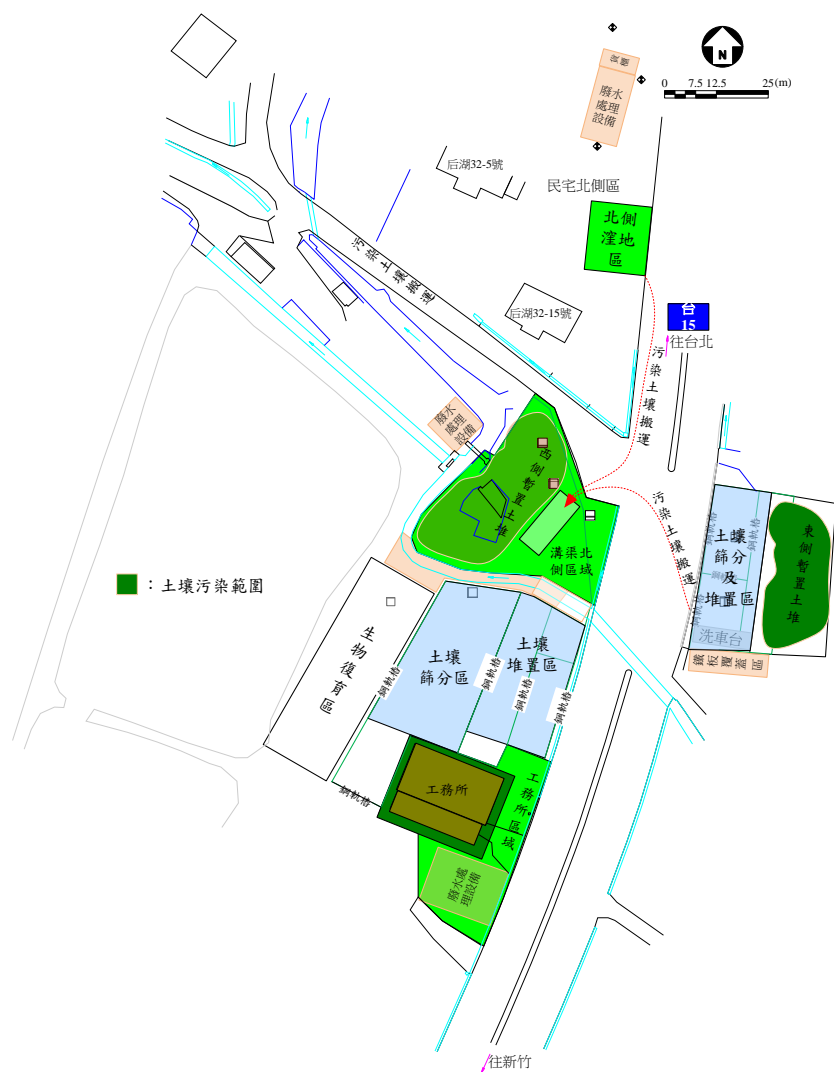


圖 3-6、本場址土壤污染範圍推估圖



圖 3-7、本場址地下水中 TPH 污染範圍推估圖



4. 場址內現有整治系統

本場址目前執行控制計畫中，分區採取不同整治工法，分區作業流程規劃如圖 3-8，主要為 TPH 污染土壤採取開挖與現場復育作業(分區如圖 3-9)，地下水污染區域採取現地雙相抽除與生物曝氣作業(系統配置如圖 3-10)，部分無法開挖區域之污染土壤整治採取現地化學氧化作業。本計畫預定於現地雙相整治作業區域進行模場試驗。主要位置如圖 3-3 所示，現場設備傳感器將安裝於 RL-2 機組。

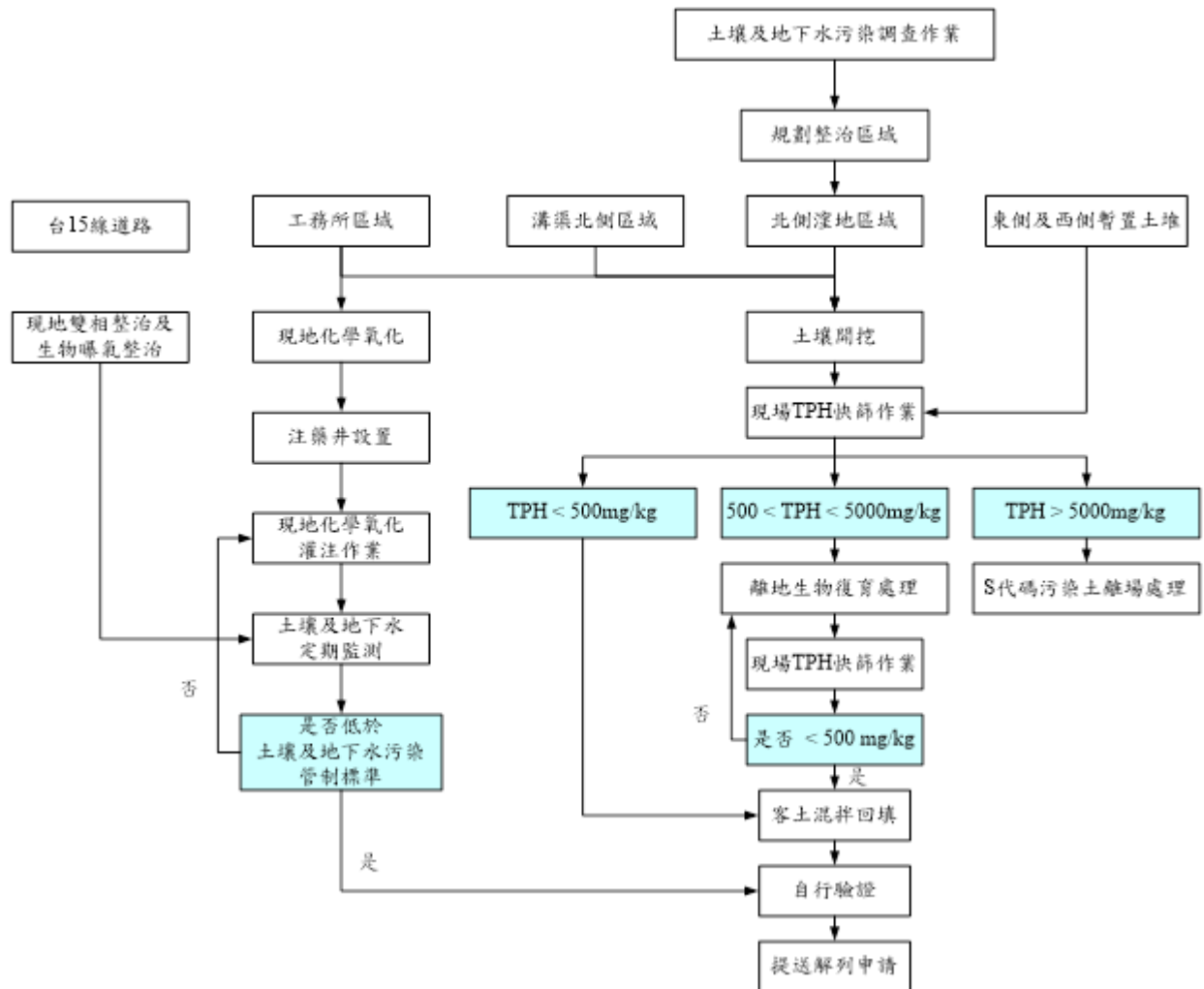
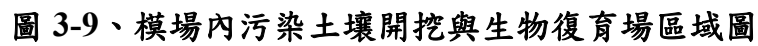


圖 3-8、整治工作整體流程圖

64





3.6 場址遠端監控系統架構規劃與建置

本研究中預定安裝完成的遠端監控系統架構如圖 3-11，主要包括三大部分：

1. 模塊一：土壤氣體抽除設備相關傳感器

參考圖 3-11，設備上需安裝以下傳感器：

- 壓力(真空)傳感器：壓力感測器原理為當外力(壓力)使膜片變形發生壓阻效應，進而使電阻變化時，感測器會將電阻的變化轉換為電子訊號(如 4-20mA)，而可進一步轉換為壓力讀值，本計畫中將選擇 0.1 級以上、溫度飄移 0.01%FS/°C 等級之市售壓力計。
- 流量傳感器：本研究選擇熱感式流量傳感器，熱流傳感技術依賴於流動液體的冷卻效果。流速越高，冷卻效果越好，可以可靠地測量。感測器尖端包含兩個鉑 RTD 元件和一個熱源。通電時，感測器的尖端被加熱。通過改變施加在加熱器上的功率，兩個 RTD 之間的溫差保持恆定 ($\Delta 4^{\circ}\text{C}$)。當這些 RTD 對流動介質的物理效應做出反應時，流速就會從感測器尖端傳導熱能。此類型的流量傳感器體積較小，以現地整治設備常具有移動性需要的情況，比較好進行設計組裝，選擇的型號主要需考慮目標量測的流量範圍及管徑，以 5HP 的土壤氣體抽除設備，通常排放管道為 2"，最大流量量測範圍可以達 4000lpm；排水管道為 1"，流量量測範圍 50-100lpm。
- 溫度傳感器：溫度傳感器分為兩種主要類型：接觸式和非接觸式。接觸式感測器包括熱電偶和熱敏電阻，在量測時會接觸受測物體；非接觸式感測器則會量測熱源所釋放的溫度輻射，藉此判定其溫度，本計畫選擇熱電偶型式溫度傳感器，量測範圍 0-100°C。
- 變頻器與定時器：直接導出變頻器頻率數據與馬達運轉時數表數據，以評估馬達運轉動力輸出情形。

2. 模塊二：土壤氣體濃度監測系統

土壤氣體濃度監測系統為本研究中重要需要組裝的監控系統，監測參數以抽除氣體的尾氣濃度為主，主要包括揮發性有機氣體濃度(VOCs)評估移除低碳數污染物的濃度計算移除量、CO₂/CH₄/O₂ 量測地下環境好(厭)氧情況間接評估生物降解的現象變化；除了尾氣濃度外，本系統將架設地下水水流元吹捕裝置將抽出的地下水經曝氣取得的氣體濃度採集後進行監測，相關數據做為輔助評估使用(替代人工的 PID 頂空濃度快篩)。第一年的工作成果將進一步於第二年繼續優化監測



以人工智慧結合遠端監控優化場址之現地整治成效

系統，其中，涉及傳感器的部分，將視第一年工作成果做必要的修正與改善。本研究採用的氣體偵測器工作原理與規格簡述如下，於本計畫中亦會進行傳感器的可靠度評估，並定期以標準品進行查核，各儀器之查核、校正與允收標準如表 3-3，以便評估尾氣監控設備產品的穩定性、影響因素與耗材成本等，儀器取樣邏輯中都會先經過乾淨空氣歸零後再取樣：

- VOCs 傳感器：VOCs 選擇以光離子偵測器(Photoionization Detector, PID)作為偵測器，以本場址主要污染物為二甲苯而言，應具有代表性，根據場址內過去的 PID 量測數據顯示，選擇量程範圍為 0-2000ppmv 應可以適合本研究之使用。本研究採用的 PID 偵測器為英國品牌 Alphasense 公司製造，精度在 3% 量程範圍內。
- CH₄ 偵測器：甲烷採用熱導偵測器（Thermal Conductivity Detector，縮寫為 TCD）進行偵測，是常見的氣體偵測器，對於乙烷、丙烷等烷類的選擇性不佳，但以本研究中現地土壤氣體的偵測以評估現場好氧與厭氧的現象且場地內並無運作其他烷類化合物來講，應可作為偵測器。本研究採用日本 FIGRO 廠牌的甲烷偵測器，偵測範圍為 0-10000ppmv，精度在 3% 量程範圍內。
- CO₂ 偵測器：採用紅外線吸收原理的偵測器，主要用來偵測場址在好氧(曝氣)下，是否因生物降解增加 CO₂ 濃度變化而可作為推判場址內整治行為的參考因素，本計畫採用 Dynament 的偵測器，偵測範圍為 0-10000ppmv，精度在 3% 量程範圍內。
- O₂ 偵測器：採用電化學原理進行氧氣的偵測，主要可用以評估曝氣後的下環境的影響以及生物降解的現象，本研究採用英國 Alphasense 製造的偵測器，偵測範圍為 0-30%，誤差同樣為量程的 3%。

表 3-3、本計畫偵測器查核、校正與允收標準

| 儀器名稱 | 查核標準品 | 查核頻率 | 允收標準 | 校正頻率 |
|-----------------|-----------------|------|------|-----------|
| PID | n-異丁烯 | 每周一次 | ±10% | 查核不過或每季一次 |
| CH ₄ | CH ₄ | 每周一次 | ±15% | 查核不過或每季一次 |
| CO ₂ | CO ₂ | 每月一次 | ±20% | 查核不過或每年一次 |
| O ₂ | O ₂ | 每月一次 | ±20% | 查核不過或每年一次 |

3. 模塊三：地下環境壓力監控系統

為了評估模型推估數據的相關性及測試是否需要增加評估參數，本計畫於第



一年度將於整治井與新設監測井共兩口井內安裝水壓傳感器與氣壓傳感器，規劃示意如圖 3-12，其中，監測井將同時設置為土壤氣體濃度監測井，將參考「地下儲槽系統土壤氣體監測井中油氣檢測方法」(NIEA M203.11C)施作。於本模塊中安裝的水壓與氣壓傳感器說明如下：

- 投入式水位計/氣壓計：兩者都是基於所測液體靜壓與該液體高度成正比的原理，採用擴散矽或陶瓷敏感元件的壓阻效應，將靜壓轉成電信號。經過溫度補償和線性校正，轉換成 4-20mA DC 標準電流信號輸出。本研究中預定採用液位計量程在 5m，外殼 316 不銹鋼，以陶瓷材質為感應膜片的機種，氣壓計則量測範圍為 0-1000mmHg，用以補償液位計因抽氣或注氣影響造成的水位值。電線都將採用鐵弗龍材質電線，以減少可能因為地下環境油氣影響所導致的損壞。

4. 遠端數據傳輸

- 本研究中的遠端數據傳輸的架構，第一層為在需要作動的元件與讀取數據的傳感器以 Modbus 協議或直接電流(4-20mA)的型式傳輸數據到 PLC，而第二層則為 PLC 到路由器，本層則可以用 Modbus-RTU(Remote Terminal Unit)二進制數據形式或 Modbus TCP(Transmission Control Protocol)用乙太網路線型式傳輸數據，Modbus 協定定義了一組通信規則，包括數據的格式、功能碼、地址等，透過通用的協定可以與不同讀取器取得相同的數據組態，進而可以從源頭(傳感器)到遠端伺服器得到相同的數據內容。第三層為路由器至遠程數據庫，本層主要將 Modbus 的數據協定透過 4G/5G/NB/Wifi 等無線通訊方式傳輸到遠端數據庫，而設定傳輸頻率本身決定了數據傳輸的量體，以本研究而言，現場的變動並非瞬間差異極大者，因此，採取每 10-30 分鐘進行一筆數據傳輸，應已足夠。本次收集數據的遠端平台將採用新野科技既有數據平台進行下載至專用的資料庫內做後續應用，於本計畫實施的第二年將導入預測模型的演算結果於新建的平台呈現預測結果。

3.7 整治成效評估指標之建立

透過量化性的評估成效包括兩項指標：

- 整治成效 Remedial Performance：符合整治目標的程度
- 系統成效 System Performance：特定的整治單元符合當初設計期望值的程度

其中，本計畫透過預測模型將進一步建立以下是否獲得優化的指標內涵：



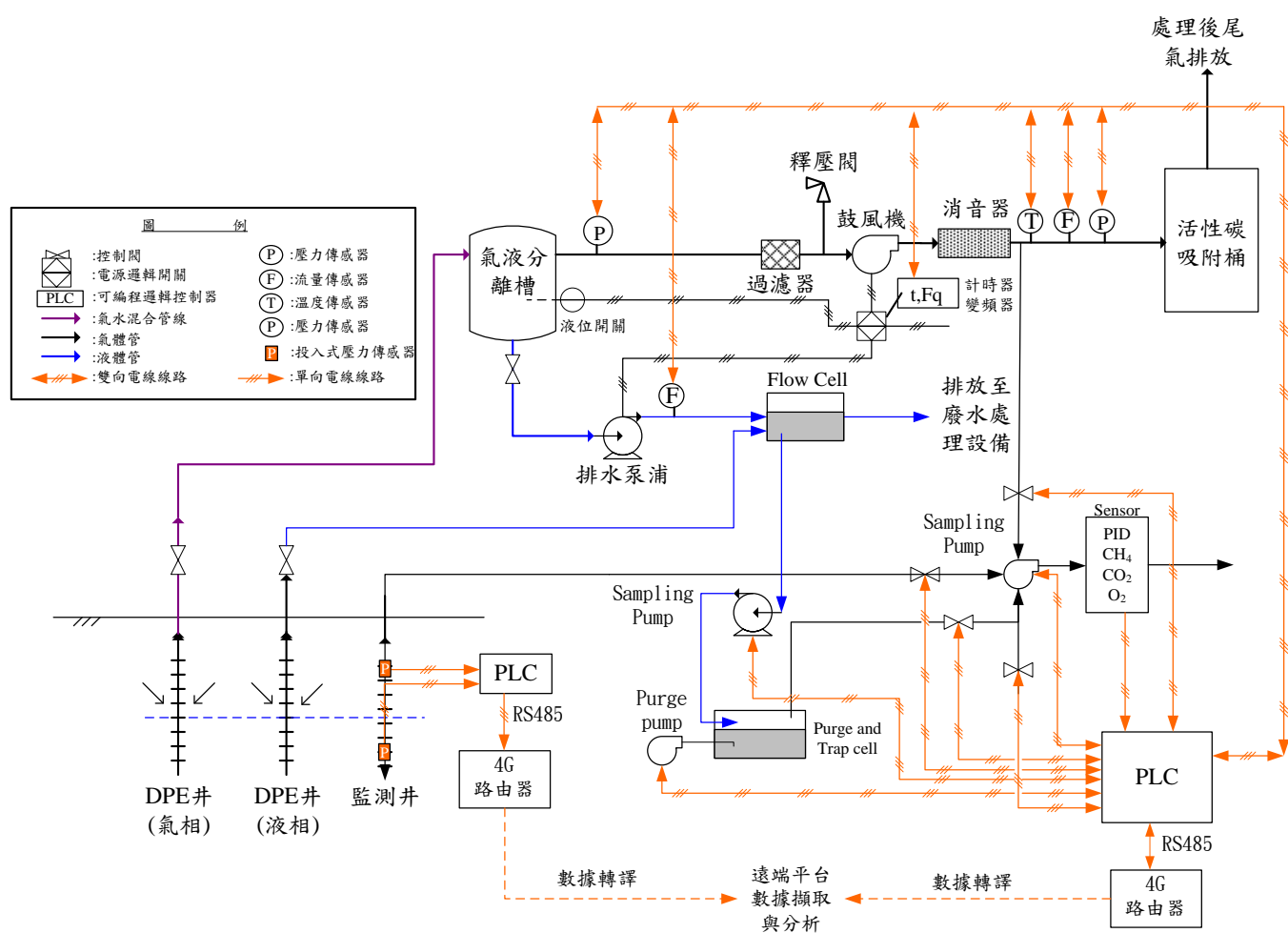
以人工智慧結合遠端監控優化場址之現地整治成效

- 監測頻率是否優化
- 整治設備操作條件是否優化
- CSM 是否及時反饋提供整治決策的調整
- 人力操作成本是否優化
- 整治時程是否縮減?
- 能源損耗是否優化
- 總體成本(涵蓋碳排放成本)是否優化

3.8 本年度計畫執行進度規劃

依照計畫書規劃之工作進度甘特圖如表 3-4，目前資料庫建置因為計畫起步時間較慢，估計要到 6 月中左右完成，但機器學習的程序已同步進行中，預估於遠端系統架設完成後可以獲得更進一步的成果，目前尚在進行參數分析，詳第四章結果與討論。

第三章 研究方法及步驟



說明：橘色線路為本計畫需架構部分

圖 3-11、模場監控架構規劃圖



以人工智慧結合遠端監控優化場址之現地整治成效

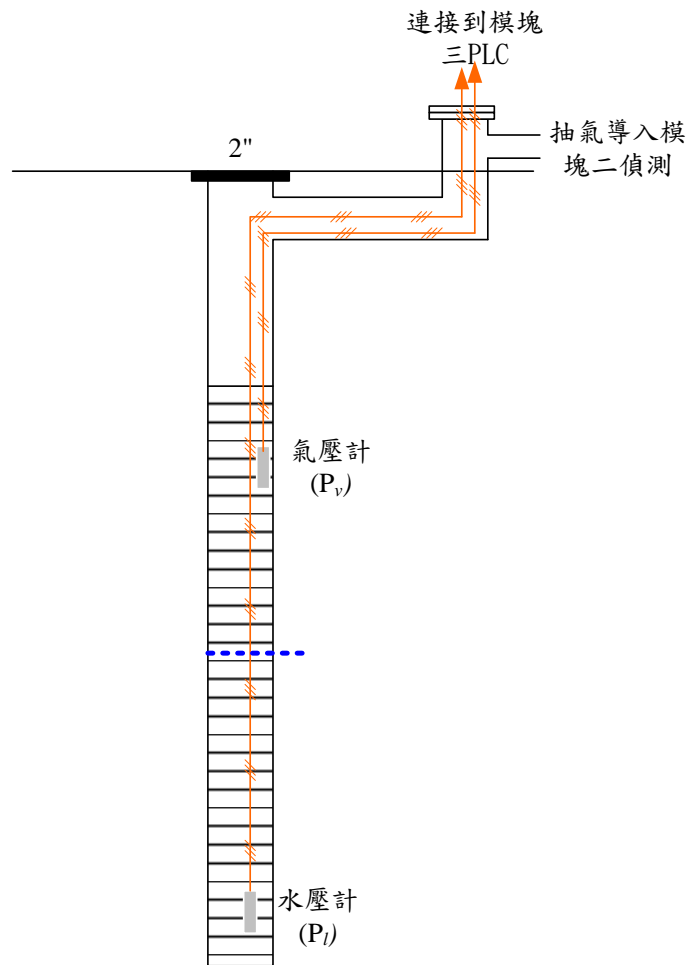


圖 3-12、現地壓力與土壤氣體偵測井設計規劃圖

表 3-4、工作進度甘特圖

| <div>年月</div> <div>工作項目</div> | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 備註 |
|-------------------------------|----|-----|--|-----|-----|-----|-----|-----|-----|-----|-----|------|----|
| 1.已解列場址資料彙整與數據前處理 | | | | | | | | | | | | | |
| 2.數據訓練與測試 | | | | | | | | | | | | | |
| 3.遠端監控設備設計組建 | | | | | | | | | | | | | |
| 4.遠端監控設備於現場安裝 | | | | | | | | | | | | | |
| 5.監控設備測試 | | | | | | | | | | | | | |
| 6.數據採集與設備校正 | | | | | | | | | | | | | |
| 7.模型參數調整與預測 | | | | | | | | | | | | | |
| 8.初步模型選擇 | | | | | | | | | | | | | |
| 9.提送期中報告 | | | | | | | | | | | | | |
| 10.提送期末報告 | | | | | | | | | | | | | |
| 工作進度估計百分比（累積數） | 5% | 10% | 25% | 30% | 40% | 45% | 60% | 65% | 80% | 90% | 95% | 100% | |
| 預定查核點 | 期中 | | ※完成數據訓練集與測試集資料整理並建立開始數據訓練 ※組裝完成遠端監控設備及架構完成數據收集資料庫 | | | | | | | | | | |
| | 成果 | | ※安裝完成遠端監控系統並開始擷取數據到遠端資料庫 ※利用模型進行現場數據模擬與參數調整 ※初步模型選擇效能比較與後續建議 | | | | | | | | | | |



第四章 結果與討論

4.1 資料前處理執行情形

本計畫已建置完成資料如表 4-1 所示，已於本計畫期間內完成，惟由於模型建立過程中需要增補確認的輸入參數，經常需要透過模型訓練成果回顧資料正確性，尤其 CSM 的建立部分，本研究中建立 CSMcheck.py 程序用於產製不同場址不同時間的概念圖，其經場址專案經理協助確認是否存在不合理性時需再回溯燈打值確認取樣方法、數據正確性、代表性後進行修正，為特徵工程中重要的工作，亦影響機器學習效能甚重，故後續在持續優化模型的過程中，仍需回顧 CSM 的性質是否合理並酌予修正，重新訓練。

表 4-1、資料庫建置情形說明

| 資料類型 | 數據來源 | 建置情況說明 |
|-----------|--------------------------------------|------------------------|
| 土壤剖面資料 | 土壤採樣紀錄表 | 建置完成 |
| 土壤垂直濃度分布 | 土壤篩測紀錄表 | 建置完成 |
| 地下水水位 | 地下水水位量測紀錄表 地下水頂空濃度紀錄表 地下水採樣紀錄表 | 建置完成 |
| 地下水頂空濃度 | 地下水頂空濃度紀錄表 | 建置完成 |
| 土壤檢測數據 | 檢測報告 | 建置完成 |
| 地下水檢測數據 | 檢測報告 | 建置完成 |
| 單(多)井尾氣數據 | 整治系統執行效能評估表 整治系統操作日誌 | 建置完成 |
| 整治系統操作參數 | 整治系統操作日誌 | 建置完成 |
| 整治系統操作歷程 | 場址改善完成報告書 | 建置完成 |
| 場址概念模型 | 前述資料綜合 | 部分完成(仍有些彙整資料另需要場址經理確認) |

3.2 模型建立

3.2.1 以抽提井為核心的分層汙染特性做為輸入值的方法

本研究初期採取的模式架構為從抽提井的角度去綜合鄰近土壤及地下水調查資訊，並主要引用土壤調查數據區分為飽和層與未飽和層數據利用反距離權重法去換算抽提井的飽和層與未飽和層的汙染程度，評估相關變數對於抽提井真空度與尾氣濃度的影響，相關結果說明如下：

1. 輸入變數分析：

抽提井為核心的輸入值取用方法採用的數據如表 4-2，預測目標為真空度與尾氣濃度，表 4-2 的數據進行統計彙整如圖 4-1 直條圖(Histogram)，各變數間的變異數分析如表 4-3，其中正相關底色為淺綠色，負相關為粉紅色，數值愈大代表相關性愈高，愈接近 0 表示愈沒有相關性，其中發現未飽和層的濃度值與尾氣值呈負相關，而飽和層的濃度值反而對於尾氣值存在更大的相關性，初步認為將土壤剖面資料搭配選定的固定水位資訊造成了極大的偏誤，如何將場址資料投射到抽提井為模型數據預處理上須優先解決的問題，將在第 4.2.2 節說明。

表 4-2、抽提井核心法建立輸入參數

| 指引代號 | 欄位名稱 | 代號 | 指引代號 | 欄位名稱 | 代號 |
|------|-------------|--------------|------|------------------|----------|
| 1 | 場址名稱 | Site_Name | 10 | 量測日期 | Date |
| 2 | 整治井號 | Well_No | 11 | 設備頻率(Hz) | Freq |
| 3 | 整治井座標(TM_X) | TM_X | 12 | 流量(cfm) | Flow |
| 4 | 整治井座標(TM_Y) | TM_Y | 13 | 溫度() | Temp |
| 5 | 整治井設置日期 | Install_Date | 14 | 真空度(mmAq) | Vacuum |
| 6 | 口徑(“) | Dia_In | 15 | 量測 PID 尾氣(ppmv) | PID_bAC |
| 7 | 井深(cm) | Well_Depth | 16 | 量測日期代表水位(cm) | Ope_wt |
| 8 | 開篩起始深度(cm) | Screen_From | 17 | 未飽和層平均 PID(ppmv) | UnsatPID |
| 9 | 初始水位(cm) | Ini_wt | 18 | 飽和層平均 PID(ppmv) | SatPID |

表 4-3、抽提井為核心的輸入參數共變異數(Covariance)

| | Freq | Vacuum | PID_bAC | Dia_in | Depth | Screen_From | ini_wt | ope_wt | UnsatPID | SatPID |
|-------------|---------|---------|----------|---------|---------|-------------|---------|---------|----------|------------|
| Freq | 16.92 | 687.37 | 251.9 | -0.36 | 57.89 | 35.15 | 10.91 | -37.59 | 37.46 | 2996.64 |
| Vacuum | 687.37 | 101375 | 29062.39 | 32.38 | 17152.9 | 12175.16 | 5071.04 | 7218.14 | 6802.94 | 138460.6 |
| PID_bAC | 251.9 | 29062.4 | 63645.8 | -24.76 | 6507.84 | 3962.59 | 2586.86 | 4672.66 | -3759.95 | 178382.42 |
| Dia_in | -0.36 | 32.38 | -24.76 | 5.92 | 66.38 | -22.93 | -18.82 | -23.95 | 91.98 | -320.75 |
| Depth | 57.89 | 17152.9 | 6507.84 | 66.38 | 45268.9 | 10275.56 | -1743.3 | -1007.8 | 5804.04 | 42552.52 |
| Screen_From | 35.15 | 12175.2 | 3962.59 | -22.93 | 10275.6 | 4891.62 | 1982.98 | 2511.72 | 280.9 | 27691.35 |
| ini_wt | 10.91 | 5071.04 | 2586.86 | -18.82 | -1743.3 | 1982.98 | 4358.67 | 3420.1 | -2608.68 | 39264.63 |
| ope_wt | -37.59 | 7218.14 | 4672.66 | -23.95 | -1007.8 | 2511.72 | 3420.1 | 7592.37 | -3418.57 | 73759.56 |
| UnsatPID | 37.46 | 6802.94 | -3759.95 | 91.98 | 5804.04 | 280.9 | -2608.7 | -3418.6 | 38569.38 | -68292.24 |
| SatPID | 2996.64 | 138461 | 178382.4 | -320.75 | 42552.5 | 27691.35 | 39264.6 | 73759.6 | -68292.2 | 4665286.07 |



以人工智慧結合遠端監控優化場址之現地整治成效

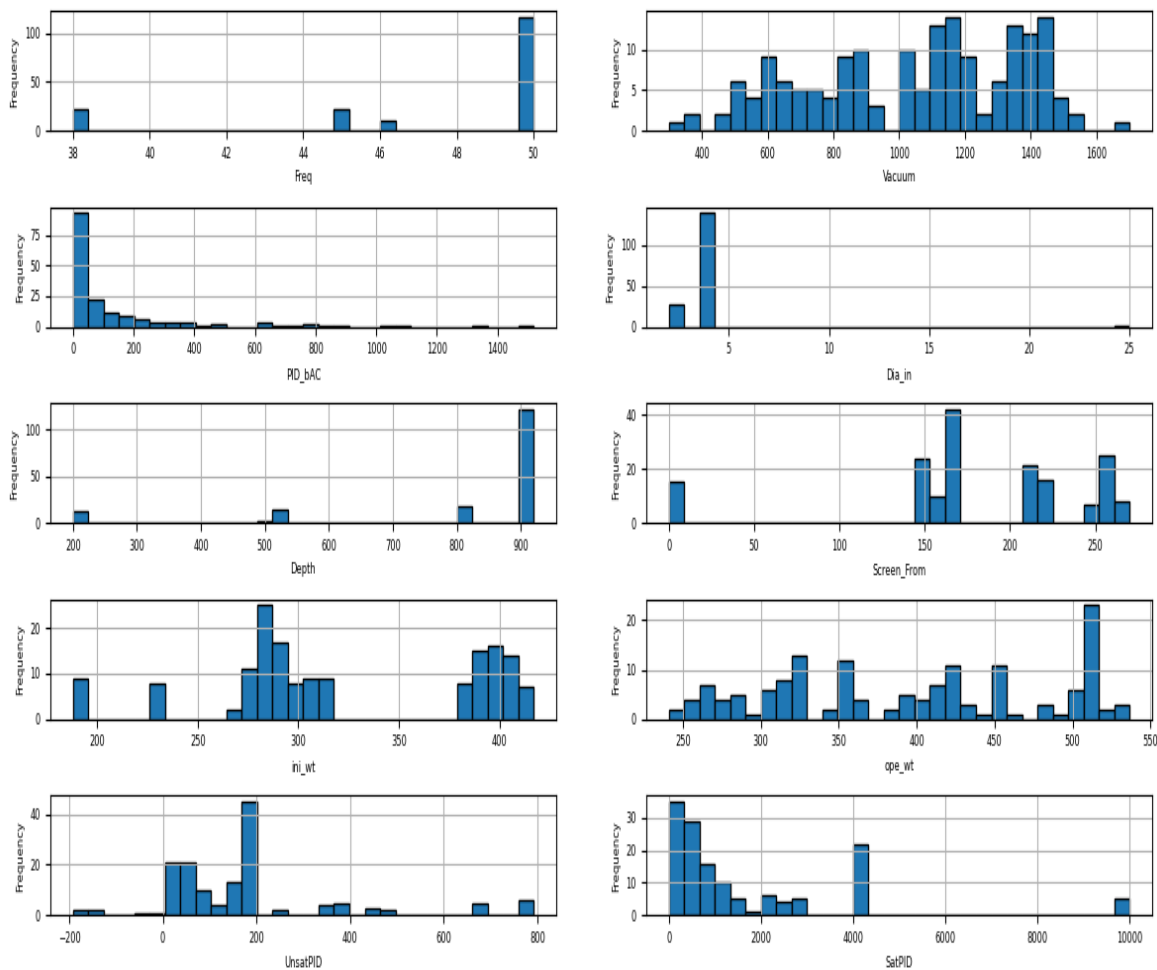


圖 4-1、抽提井為核心的輸入變數的數據分布直方圖

從圖 4-1 的直方圖可以看到不同數據分布的特性，其中可以發現飽和層相較未飽和層與抽出尾氣 PID 之間的分布更為相近。有了數據的基本分析，將先針對兩個應變數進行學習與預測，包括真空度與量測尾氣，由於真空度的測值受到設備性能而有較大的變化，吾人對其進行了標準化，以不同設備可以達到的最大真空度進行正規化，PID 則仍先用原數值來看預測結果。

$$\text{Vac_NM} : \frac{\text{Vaccum}}{\text{設備最大真空度}}, \left\{ \begin{array}{l} \text{鼓風機} : 2000\text{mmAq} \\ \text{魯式泵} : 5000\text{mmAq} \\ \text{液環式} : 10000\text{mmAq} \end{array} \right\} \quad (4.1)$$

2. 以真空度為目標值的參數設定—MLP

在 MLP 中重要的訓練參數包括訓練與測試集比例、隱藏神經元的數量、迭代次數的優化、活化函數的選擇，茲就以 Freq、Screen_From、ini_wt 及 Depth 作為輸入值，輸出為 Vac_NM 進行訓練，單一場址的有效數據量為 158 筆，得到以



下的參數結果：

■ 最佳化隱藏層的評估(# of Hidden Layer)

設定不同隱藏層數量(1,3,5,10,20,70,100,200)，隨機亂數產生訓練集:驗證集:測試集為 80:10:10 的數據評估最佳的神經元數量，每種數量跑 10 次，執行三次後得到如圖 4-2 的箱線圖，因為只有 158 筆數據，原則上應選擇可接受誤差範圍內的最小數量隱藏層較佳，從圖上看到隱藏神經元數量在 20 到 70 間，誤差較低且相對穩定。

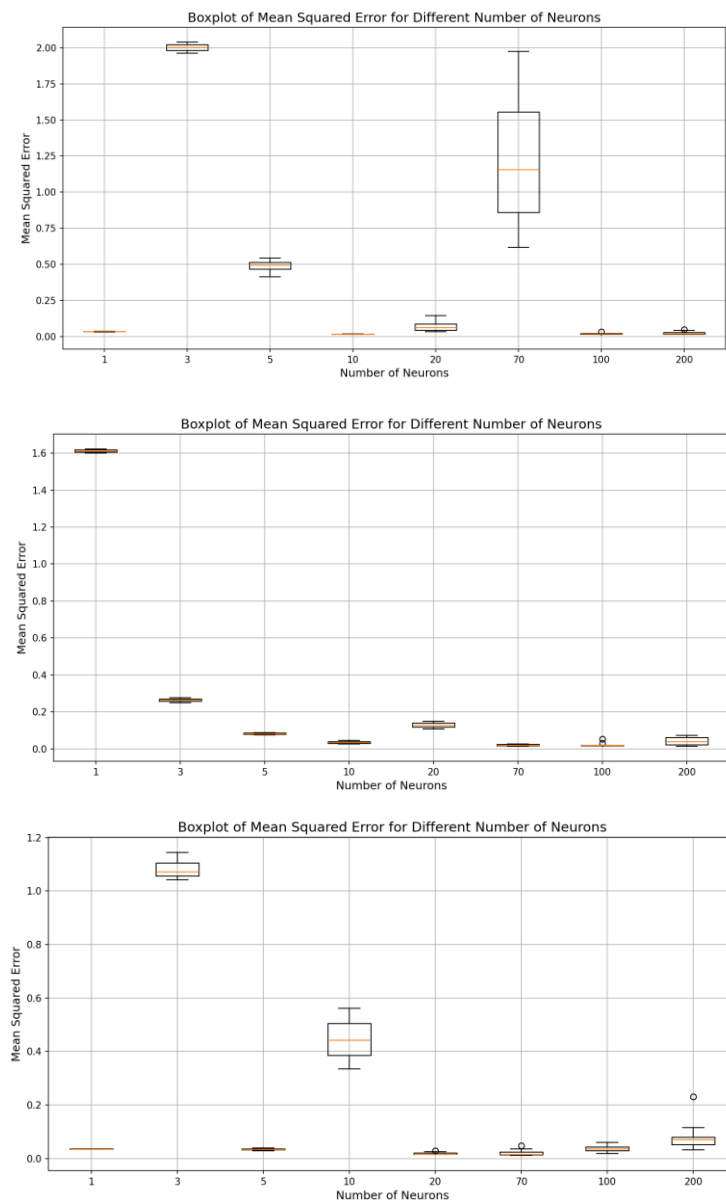


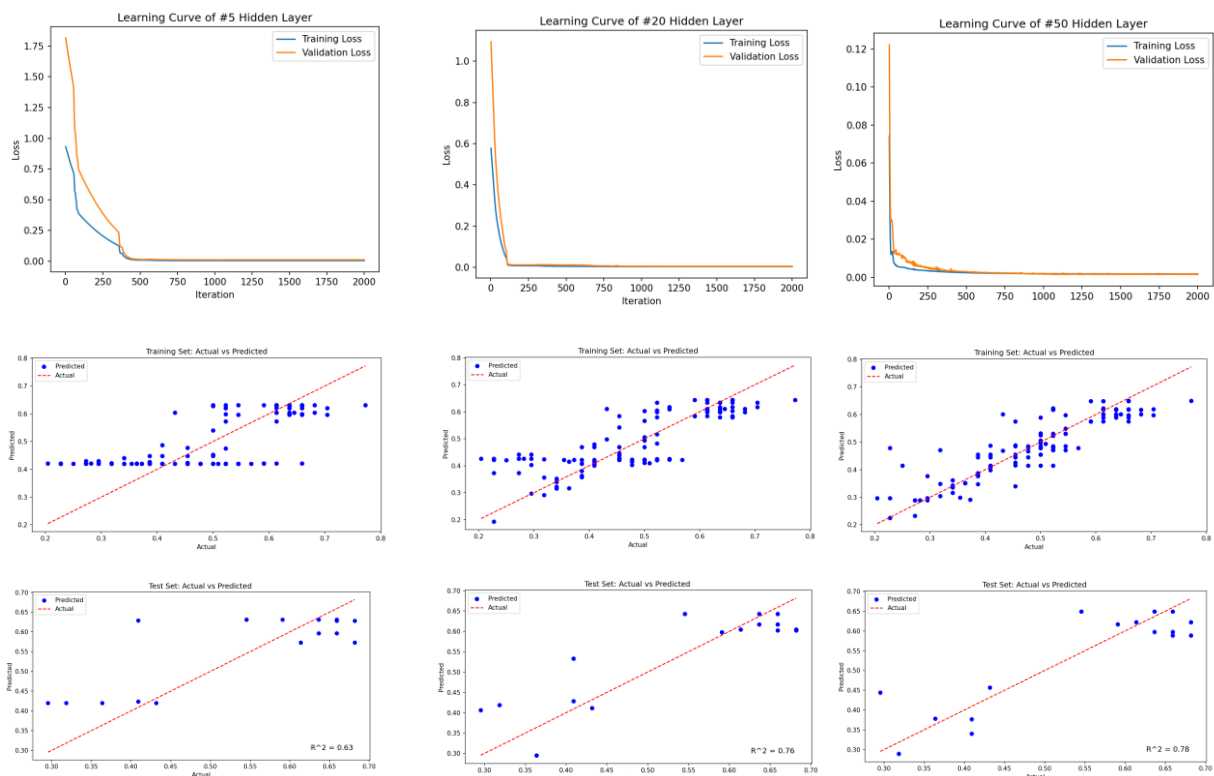
圖 4-2、不同隱藏層數量訓練誤差箱線圖

■ 迭代次數的選擇



以人工智慧結合遠端監控優化場址之現地整治成效

為了評估迭代次數的影響，將最大迭代次數 2000 內評估訓練集與驗證集的誤差變化，於此同時可以檢視不同神經元下的影響，從圖上可以看到當迭代次數 500 時，訓練次數足夠；而從預測值與量測值的散布圖可以看到當隱藏層只有 5 個的時候，有些數據明顯無法區分而有不同的預測值，隱藏神經元愈多，明顯可以提高訓練集的準確度至 0.8~0.9，但隱藏神經元過高，可能造成後續泛化上的問題，這部分則需要更多數據量來訓練。而從測試集的結果來看，因為數量少，無論哪種條件下得到的準確度不會有明顯差異，但是從預測值與量測值的散布圖可以看到當隱藏層較少的情況(#5)，預測值無法區分出差異。



(a)Hidden Layer #=5
 R^2 of train: 0.49
 R^2 of validation: 0.22
 R^2 of test: 0.63
 MSE: 0.007

(b)Hidden Layer #=20
 R^2 of train: 0.62
 R^2 of validation: 0.67
 R^2 of test: 0.67
 MSE: 0.004

(c)Hidden Layer #=50
 R^2 of train: 0.80
 R^2 of validation: 0.90
 R^2 of test: 0.78
 MSE: 0.004

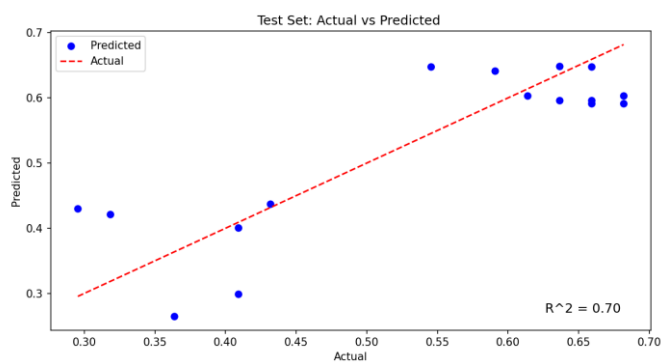
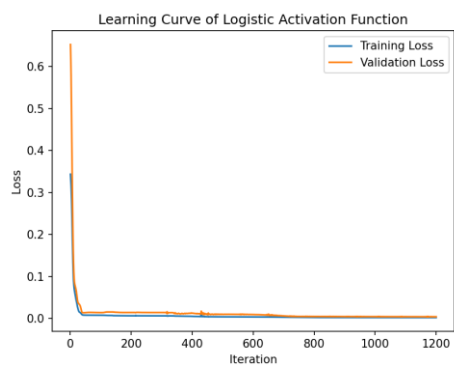
圖 4-3、不同隱藏層數量訓練曲線與準確度對比

■ 激活函數的選擇

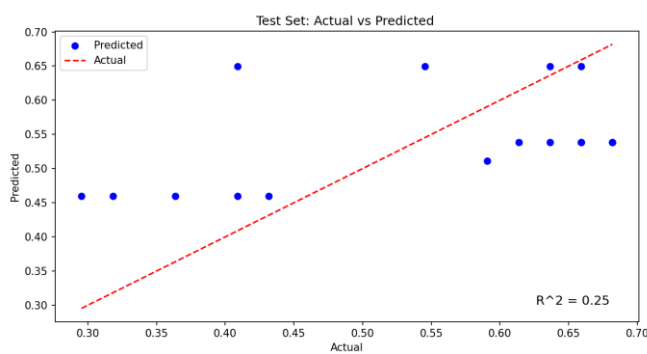
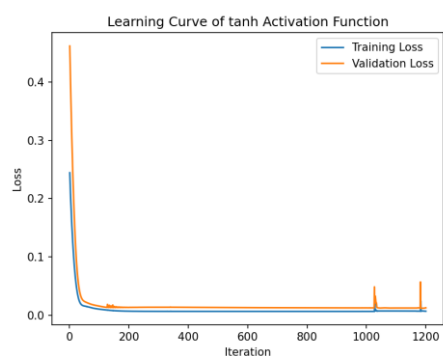
以 MLP 訓練中常見的四種激活函數代入可以得到如圖 4-4 中造成模型變化的影響，其中，前述的迭代次數、隱藏神經元的數量皆以 logistic 函數所獲得，代入其他激活函數可能相對不穩定，仍需要再測試相關條件，以前述條件設定後發現 logistic 的模型較佳。四種模型的公式如表 4-4 所列。



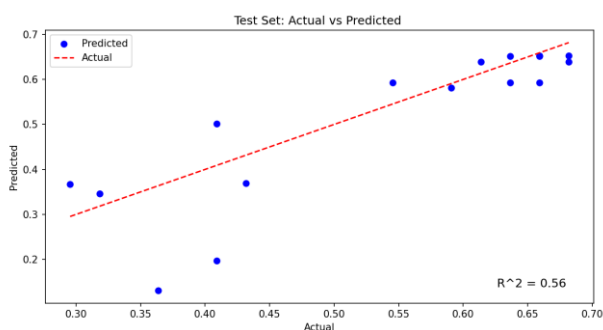
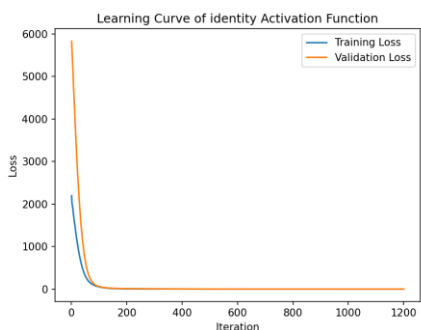
第四章 結果與討論



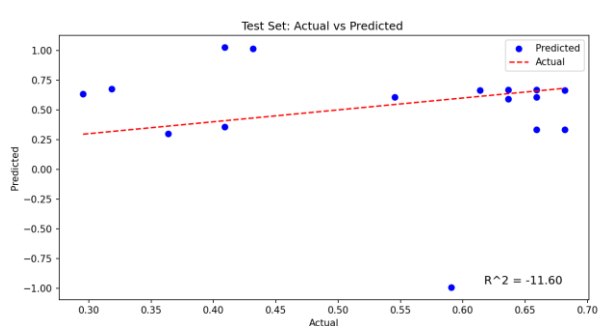
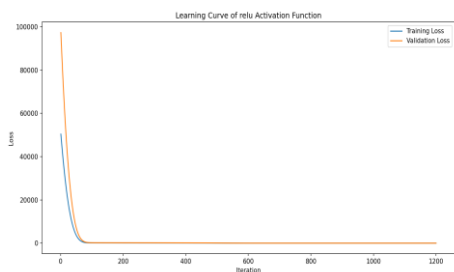
(a) 激活函數=Logistic， R^2 of train/Validation/test: 0.72,0.70,0.70，MSE: 0.006



(b) 激活函數=tanh， R^2 of train/Validation/test: 0.20,0.12,0.25，MSE: 0.014



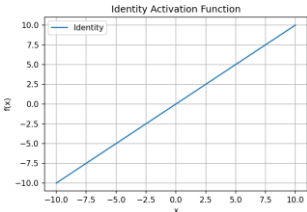
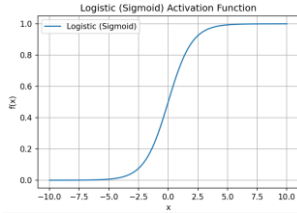
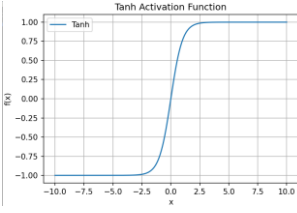
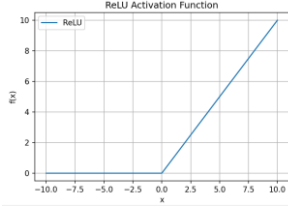
(c) 激活函數=identity， R^2 of train/Validation/test: 0.04,0.12,0.56，MSE: 0.008



(d) 激活函數=Relu， R^2 of train/Validation/test: -41.06,-29.60,-11.60，MSE: 0.23

圖 4-4、變更激活函數後的模型變化

表 4-4、激活函數列表

| 函數名稱 | 函數內容 | 函數圖形 |
|----------|---|---|
| identity | $f(x) = x \quad (4.2)$ |  |
| logistic | $f(x) = \frac{1}{1+e^{-x}} \quad (4.3)$ |  |
| tanh | $f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4.4)$ |  |
| Relu | $f(x) = \max(0, x) \quad (4.5)$ |  |

3. 以真空度為目標值的參數設定－RBNF

以相同的輸入變數與目標值採用 RBNF 的模型進行訓練，嘗試原始的基函數，以集 x 轉換提高維度後的基函數：

$$\text{RBFsampler: } K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (4.6)$$

調整 $\gamma = 38$ ， $\alpha = 0.08$ 下，不同 RBF 數量對於準確度的影響彙整如圖 4-5，從圖上可以發現 RBF 數超過 100 時才能獲得穩定較佳的性能表現。不同 RBF 數的訓練集與測試集，比較預測值與實際值如圖 4-6 及 4-7，可以發現當設定 RBF#=100 時，才可以達到 MLP－logistic 的模型性能。但是 RBF 可以改變核函數以提高預測準確度，後續將嘗試調整核函數看是否能提高預測準確度。

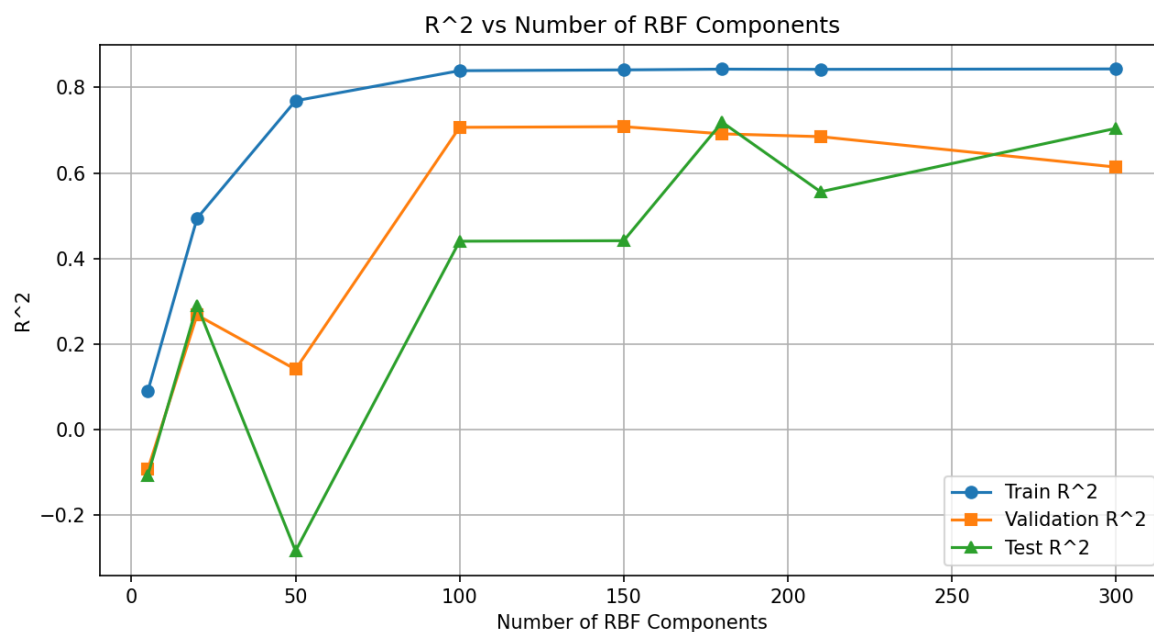


圖 4-5、不同 RBF 數量對於準確度的影響

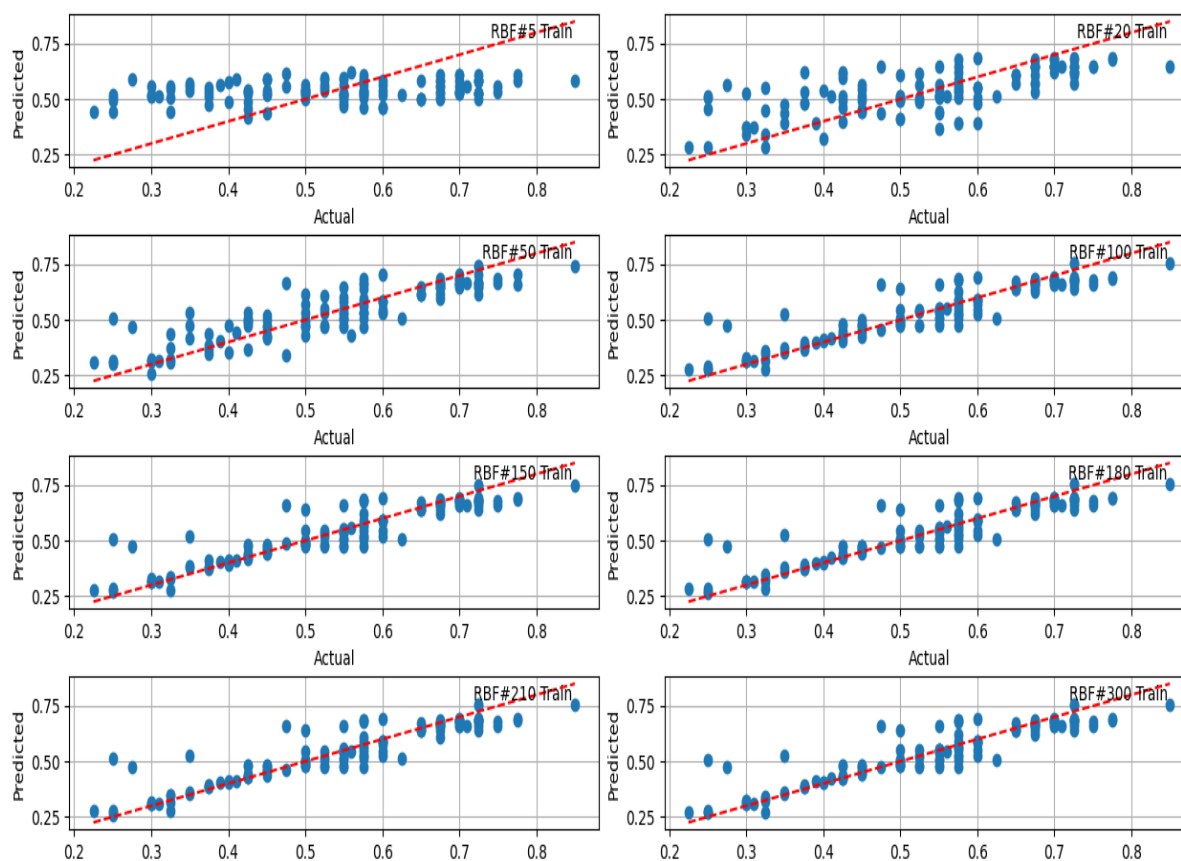


圖 4-6、不同 RBF 數量訓練集的實際值與預測值比較圖



以人工智慧結合遠端監控優化場址之現地整治成效

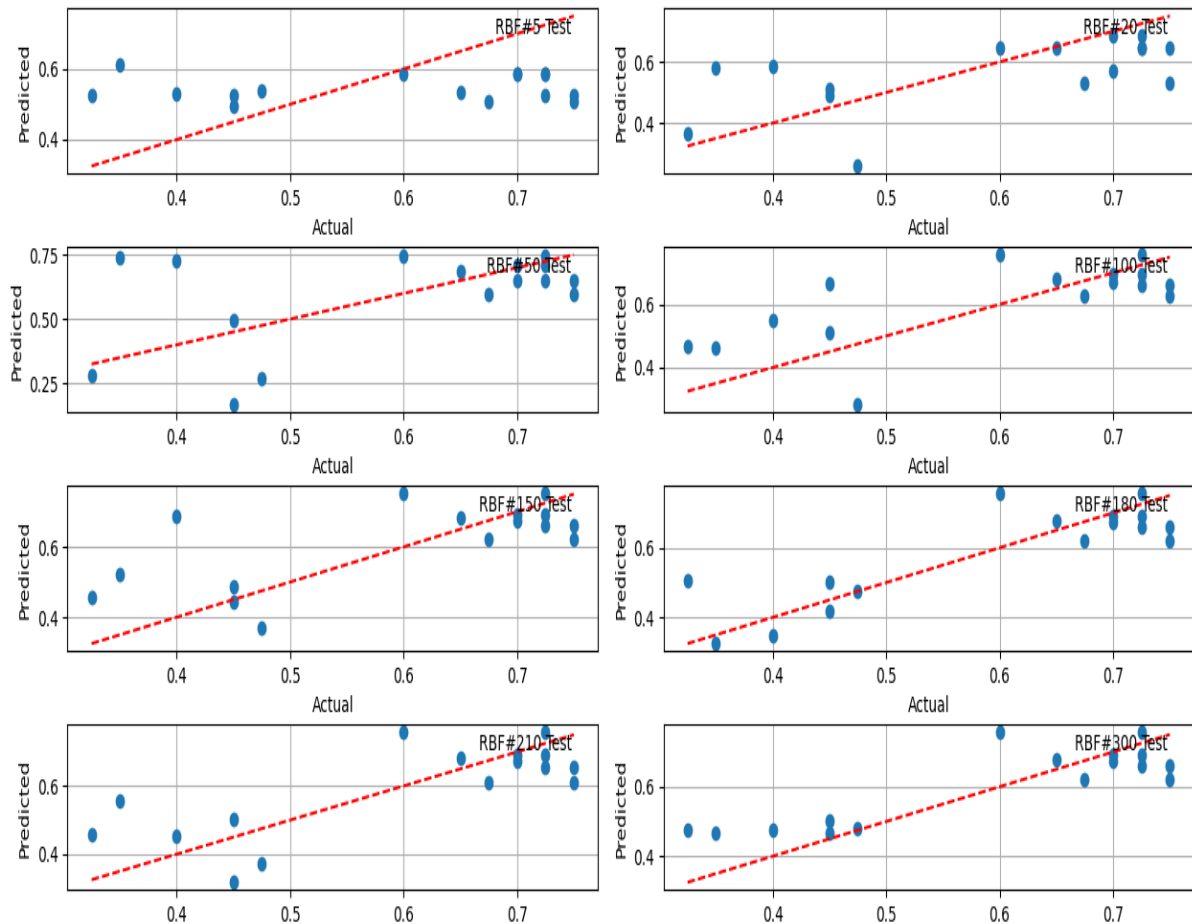


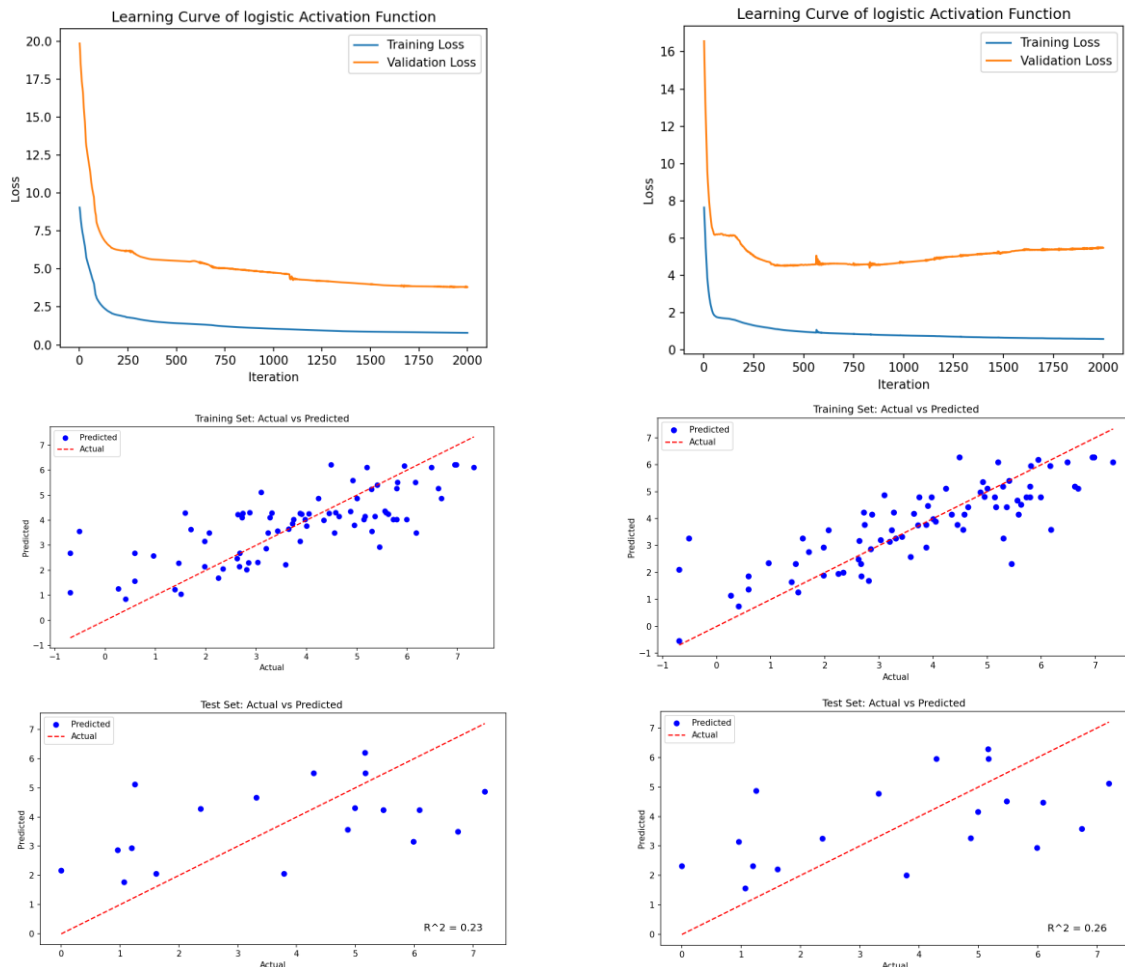
圖 4-7、不同 RBF 數量測試集的實際值與預測值比較圖

4. 以尾氣濃度為目標值的參數設定—MLP

將尾氣值取 \log 值作為正規化輸入變數，選擇 Freq、ini_wt、Depth、Screen_From、UnsatPID_NM 及 SatPID_NM 當作輸入值，將數據集設定為訓練：驗證：測試=70:15:15，經過多次測試得到測試集的最佳效能可在 0.2~0.3 之間，圖 4-8 為上述輸入參數設定 logistic 激活函數下得到隱藏層#30 與#100 下的學習曲線、預測值與實際值散步圖及效能值，從結果來看，訓練效能不佳，因此，將針對抽提井對應分層污染層的形成方式進行檢討，並對應傳統概念模型形成的方法來生成抽提井的分層污染特性，相關修正如 4.2.2 說明。



第四章 結果與討論



Hidden Layer=30

R^2 of train/Validation/test:
0.57,0.39,0.23 , MSE: 3.645

Hidden Layer=100

R^2 of train/Validation/test: 0.67,0.13,0.26 ,
MSE: 3.487

圖 4-8、以 MLP 訓練尾氣濃度在不同隱藏層之效能

3.2.2 建立場址概念模式提取至抽提井的分層污染特性做為輸入值的方法

不同於原先從 DPE(SVE)井周邊一定距離內的採樣點濃度以反距離權重平均法估計 DPE 井的飽和層與未飽和層濃度，其受到水位分類的影響過大，為了減少影響並區別出更多不同深度的污染程度，以提供機器學習上更多的特徵，針對場址內土壤、土壤氣體與地下水的現場調查數據，採取了以下處理方式：

- 1 針對土壤調查數據以地表為基礎點，定義每個場址每 100~150cm 深度為一個分層，因此，可以取得每個場址不同採土點不同污染層的現場快篩數據(代號：lev1)。
2. 參考土壤分層數據，建立每個場址特徵的分層編號(level index)，透過土壤氣



以人工智慧結合遠端監控優化場址之現地整治成效

體監測的數據，其測值，代表了水位以上範圍的污染濃度值；對於地下水頂空濃度，其測值則代表了水位以下範圍的污染濃度值，但因為監測井深度有時過深，對於 LNAPL 而言並不適當認定自水位以下深度都達到了現場監測值，因此，定義了每個場址的 pollution depth，當到達污染深度層，污染濃度降為 1/10 以下。

3. 建立數據處理程序，對於三個資料庫數據分別處理形成場址污染概念中的分層污染特徵，添加座標後作為場址概念模式中的基礎數據(HSbase、Soilgasbase、Soilbase)
4. 建立場址污染概念檢核程序(CSMcheck)，該程序將會根據前項基礎數據進行地理上的模擬，本研究經過篩選後選擇 RBF(徑向基函數)的方式進行污染分布的推估，RBF 是常見的插值推估方法，每個場地皆劃分為 100x100 的網格進行插值並繪製分布圖，常見的核函數有：

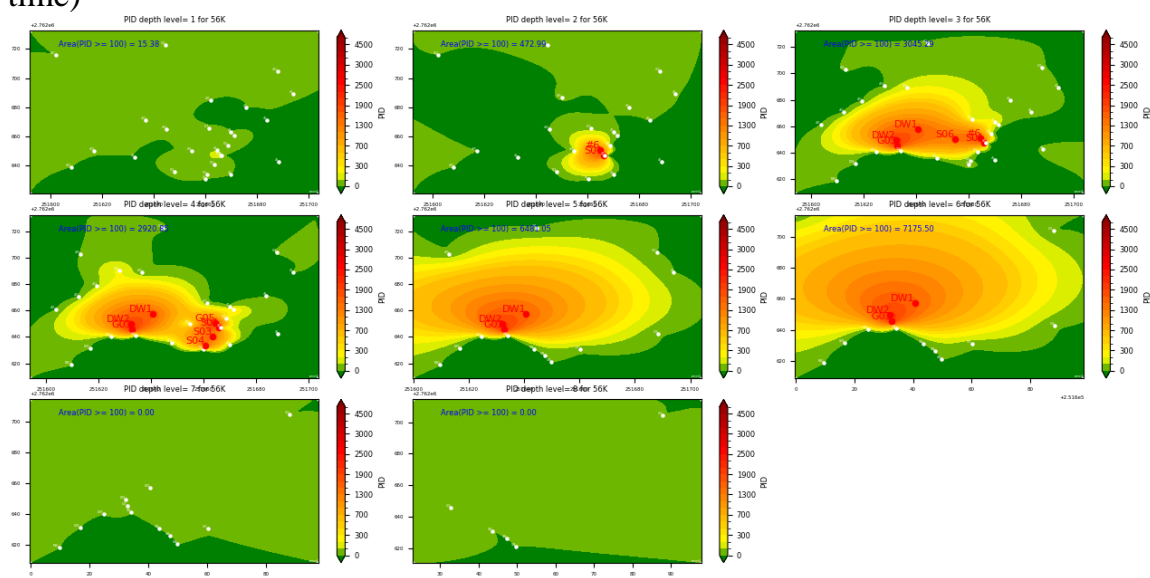
- 高斯核(gaussian)： $\phi(r) = \exp(-\gamma r^2)$ ，利用 γ 控制核的寬度，適合描述局部特徵。
- 多項式核(Polynomial)： $\phi(r) = (1 + r^2)^k$ ， k 為多項式階數，適合全局特徵的描述。
- 反高斯核(Thin_plate)： $\phi(r) = r^2 \log(r)$ ，適合全局的平滑插值。
- 線性核(linear)： $\phi(r) = r$ ，適合簡單平均的預測。

本研究中採取線性核作為推估的基礎，並根據專案經理對於推估後的分布特性進行評估，確認該時間段的污染推估合理；在數據處理上亦雷同專案經理對於場址概念模型的評估方式，採取以下原則處理數據：

- 對於相同的井，後面時間段的數據會更新原先的數據
- 對於土壤採樣點，由場址專案經理協助判讀原先的高值是否已經被取代，並將取代點的座標統整一致，通常因為採土點不可能為同一個點，所以後面的採樣點座標會不一致，為了更新場址污染概念，將會針對後段時間無法對應的點位確認高值是否仍存在。
- 優先以監測井(尚未進行整治作業的井)作為場址概念模型的基礎數據，但若因為不同時間或空間點有數據缺漏，將會採納抽水井抽出水作為缺漏位置的補充數據。



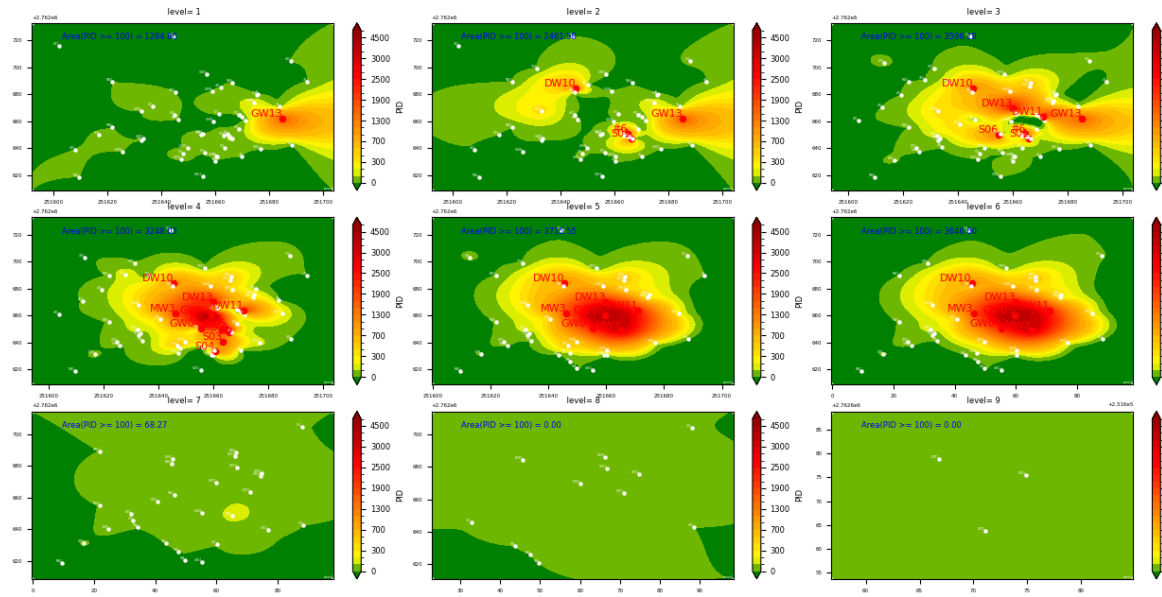
- 每個場址會引用最大與最小周界的座標生成無污染邊界點。
 - 對於每個場址中有更新調查數據的日期即生成類似圖 4-9 的污染分布圖，再經由場址經理確認是否合理並回溯檢驗數據品質。
 - 在確認每一個場址的分層污染分布圖正確性後，套用抽除井的座標對應所在網格的插值，作為該抽取井該分層的濃度值。
 - 考慮不同場址的污染來源、水文地質差異，前述分層的污染層將再進行重新分層，其中初始水位深度將調整為第 0 層，往上為 1、3、5..層，往下為 2、4、6..層，以便模擬的時候可以具有可以比較的表徵，分別命名為 PID_0、PID_1、...餘類推。本研究為便利大量圖層數據的處理，以建立相關演算程序可以快速得到場址污染分布情形，並標示高值位置，以利查對。由於目前污染分布的判定仍然交由人員檢視，後續研究將進一步提高機器學習判讀圖面部分。
5. 除了取得不同分層的數據，也針對不同抽取井選定一處鄰近的監測井作為水位觀測井，命名為 Ind_wt，物理意義上，抽氣能力會受到水位影響，因此，對於尾氣與設備操作都有一定的影響；此外，亦考慮場址洩漏源可能的影響，對於不同場址的洩漏源深度(Source_depth)也加入參數
6. 在系統操作上，也導入對於抽氣效能可能產生影響的參數，包括抽水泵抽水控制深度(Pump_depth)、抽氣泵最大真空度(Max_vac)、設備操作頻率(Freq)、抽取井數(Well_number)、運轉時數(duration)及每日運轉時數(daily operation time)。



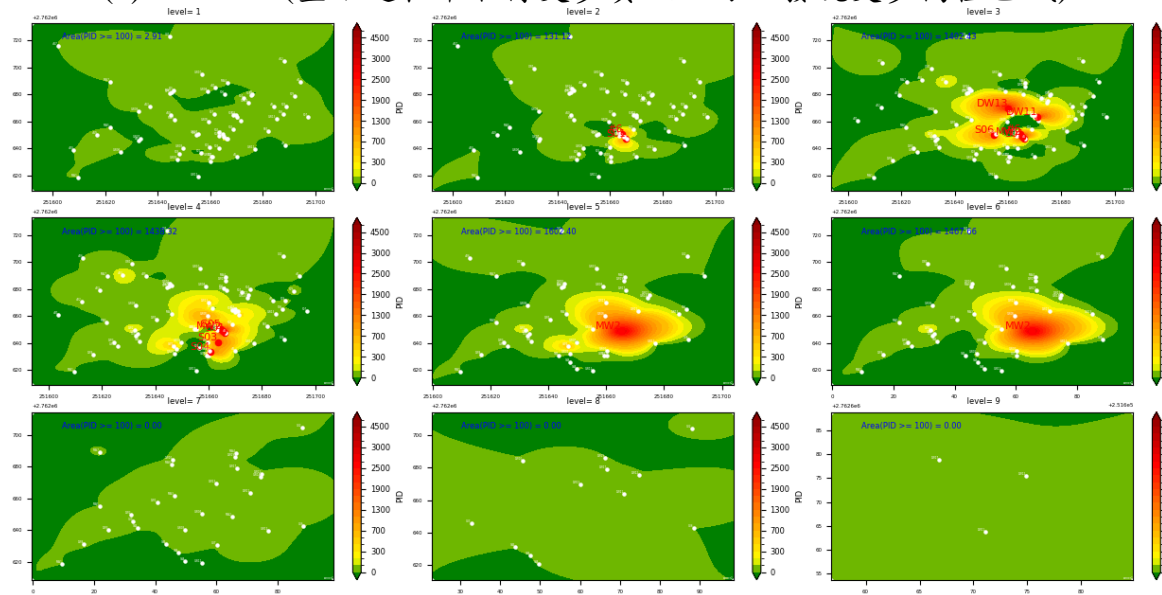
(a)2011/4/19 (初期調查點位資訊較少，推估範圍找不到更多高值)



以人工智慧結合遠端監控優化場址之現地整治成效



(b)2012/1/16(整治過程中取得更多資訊，可以發現更多高值區域)



(c)2012/6/30(隨著整治成效逐漸產生，污染範圍縮減)

圖 4-9、場址污染分布的演化

7. 生成的場址污染分布後，將與不同時間段的尾氣數據整合為一個基礎模擬數據(MLbase_level)，其分別參數相關性與統計數據如表 4-5 及 4-6，各數據直方圖如圖 4-10(a)(b)。

3.2.3 不同場址特徵參數對於模擬結果的影響

在評估採取不同的抽提井污染分層推估方法下，本研究先以 MLP 兩層神經元(1,1)來測試不同類型的輸入值對於訓練集、驗證集與測試集的影響，並逐漸增加參數來看相關變因。參考表 4-6，先選擇抽提井分層推估的 PID 值作為輸入值，

選擇 PID_0~PID_3 及 PID_5，其中 PID_4 因為生成的方式與 PID_2 的相關性極高，故排除不納入輸入值，得到預測值與實際值的模型如圖 4-11(a)(b)，效能評估指標採取 RMSE、 R^2 、AIC 值彙總如表 4-7。從圖 4-11 跟表 4-6 可以看出在經過先建立概念模型再投射到 DPE 井與排放尾氣進行學習，其效能相較期中報告的做法得到更穩定的預測模型。而增加了場址與整治設備的特徵後，學習成效約提高 10%，顯示增加場址特徵對於學習成效仍有明顯的效果，但是對於特徵的選擇與數據處理部分，可能有必要更細化中間的差異性。本研究尚未將整治井開篩範圍、曝氣井的操作情形、雙相抽除抽水的操作情形納入特徵參數，後續可以繼續進行優化。然於本計畫中，仍已選擇其他演算方式就相同的數據進行學習，相關研究成果如後章節說明。

表 4-5、輸入、目標變數相關性分析表

| Variables | Well_number | offgas_PID | Max_vac | freq | source_elev | duration | p_depth | daily operation time | ind_wt | PID_0 | PID_1 | PID_2 | PID_3 | PID_4 | PID_5 | PID_6 | PID_8 |
|----------------------|-------------|------------|---------|-------|-------------|----------|---------|----------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| Well_number | 1 | -0.15 | 0.32 | 0.14 | -0.36 | 0.69 | 0.42 | 0.24 | 0.06 | -0.22 | -0.62 | -0.4 | -0.34 | -0.4 | -0.41 | -0.01 | -0.58 |
| offgas_PID | -0.15 | 1 | 0.43 | 0.11 | 0.18 | -0.49 | -0.43 | 0.11 | 0.49 | 0.61 | 0.02 | 0.64 | 0 | 0.64 | 0.29 | 0.65 | 0.04 |
| Max_vac | 0.32 | 0.43 | 1 | 0.23 | -0.13 | -0.15 | -0.42 | 0 | 0.45 | 0.54 | -0.38 | 0.27 | -0.27 | 0.28 | -0.19 | 0.66 | -0.36 |
| freq | 0.14 | 0.11 | 0.23 | 1 | -0.28 | -0.2 | -0.31 | 0.36 | 0.05 | 0.43 | 0.07 | 0.31 | 0.33 | 0.32 | -0.05 | 0.36 | 0.27 |
| source_elev | -0.36 | 0.18 | -0.13 | -0.28 | 1 | -0.35 | 0.11 | 0.02 | 0.48 | 0.07 | 0.2 | 0.2 | -0.26 | 0.17 | 0.44 | -0.05 | 0.05 |
| duration | 0.69 | -0.49 | -0.15 | -0.2 | -0.35 | 1 | 0.6 | 0.17 | -0.19 | -0.64 | -0.48 | -0.73 | -0.25 | -0.72 | -0.33 | -0.47 | -0.53 |
| p_depth | 0.42 | -0.43 | -0.42 | -0.31 | 0.11 | 0.6 | 1 | 0.07 | -0.11 | -0.59 | -0.26 | -0.5 | -0.21 | -0.52 | -0.2 | -0.42 | -0.44 |
| daily operation time | 0.24 | 0.11 | 0 | 0.36 | 0.02 | 0.17 | 0.07 | 1 | 0.2 | 0.2 | 0 | 0.14 | 0.01 | 0.14 | 0.15 | 0.2 | -0.02 |
| ind_wt | 0.06 | 0.49 | 0.45 | 0.05 | 0.48 | -0.19 | -0.11 | 0.2 | 1 | 0.26 | -0.25 | 0.19 | -0.32 | 0.17 | 0.38 | 0.35 | -0.19 |
| PID_0 | -0.22 | 0.61 | 0.54 | 0.43 | 0.07 | -0.64 | -0.59 | 0.2 | 0.26 | 1 | 0.3 | 0.89 | 0.15 | 0.9 | 0.14 | 0.86 | 0.27 |
| PID_1 | -0.62 | 0.02 | -0.38 | 0.07 | 0.2 | -0.48 | -0.26 | 0 | -0.25 | 0.3 | 1 | 0.43 | 0.39 | 0.42 | 0.23 | 0.06 | 0.59 |
| PID_2 | -0.4 | 0.64 | 0.27 | 0.31 | 0.2 | -0.73 | -0.5 | 0.14 | 0.19 | 0.89 | 0.43 | 1 | 0.22 | 1 | 0.24 | 0.76 | 0.4 |
| PID_3 | -0.34 | 0 | -0.27 | 0.33 | -0.26 | -0.25 | -0.21 | 0.01 | -0.32 | 0.15 | 0.39 | 0.22 | 1 | 0.24 | 0.18 | 0.03 | 0.59 |
| PID_4 | -0.4 | 0.64 | 0.28 | 0.32 | 0.17 | -0.72 | -0.52 | 0.14 | 0.17 | 0.9 | 0.42 | 1 | 0.24 | 1 | 0.24 | 0.77 | 0.41 |
| PID_5 | -0.41 | 0.29 | -0.19 | -0.05 | 0.44 | -0.33 | -0.2 | 0.15 | 0.38 | 0.14 | 0.23 | 0.24 | 0.18 | 0.24 | 1 | 0.02 | 0.44 |
| PID_6 | -0.01 | 0.65 | 0.66 | 0.36 | -0.05 | -0.47 | -0.42 | 0.2 | 0.35 | 0.86 | 0.06 | 0.76 | 0.03 | 0.77 | 0.02 | 1 | 0.02 |
| PID_8 | -0.58 | 0.04 | -0.36 | 0.27 | 0.05 | -0.53 | -0.44 | -0.02 | -0.19 | 0.27 | 0.59 | 0.4 | 0.59 | 0.41 | 0.44 | 0.02 | 1 |

表 4-6、各特徵參數數據統計表

| Variable s | offgas_PID | Well_number | Max_vac | freq | source_elev | duration | p_depth | daily operation time | ind_wt | PID_0 | PID_1 | PID_2 | PID_3 | PID_4 | PID_5 | PID_6 | PID_8 |
|------------|------------|-------------|----------|-------|-------------|----------|---------|----------------------|--------|--------|---------|---------|--------|---------|-------|--------|-------|
| count | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 | 422 |
| mean | 352.5 | 6.59 | 3048.55 | 53.28 | 171.45 | 2290.65 | 410.95 | 15.95 | 344.09 | 727.82 | 500.07 | 732.63 | 68.48 | 707.79 | 6.12 | 29.08 | 14.71 |
| std | 545.31 | 2.17 | 2493.89 | 5.36 | 69.07 | 1788.55 | 77.38 | 6.25 | 50.85 | 643.42 | 288.08 | 617.19 | 91.56 | 618.53 | 13.68 | 41.49 | 13.63 |
| min | 0.1 | 2 | 2003.55 | 22 | 100 | 1 | 280 | 0 | 238 | 42.31 | 88.43 | 41.41 | 0 | 33.87 | 0 | 1.71 | 0.08 |
| 25% | 51.88 | 5 | 2003.55 | 50 | 100 | 908.44 | 350 | 12.79 | 307.6 | 223.17 | 257.5 | 216.69 | 4.65 | 172.85 | 0.17 | 6.77 | 2.68 |
| 50% | 146.15 | 7 | 2003.55 | 53.5 | 150 | 1930.47 | 360 | 18 | 344 | 592.7 | 522.52 | 622.16 | 28.73 | 601.32 | 1.48 | 18.78 | 8.77 |
| 75% | 416.75 | 8 | 2003.55 | 55.4 | 250 | 3346.9 | 500 | 20 | 379.5 | 970 | 652.6 | 1047.24 | 122.18 | 1048.3 | 3.85 | 33.68 | 24.41 |
| max | 4723 | 12 | 10121.39 | 60 | 250 | 8054 | 500 | 28.6 | 458 | 2772.4 | 1580.52 | 2899.75 | 430.7 | 2900.05 | 71.53 | 291.26 | 51.97 |

表 4-7、在不考慮超參數下的輸入變數以 MLP 演算的效能比較

| Model | Training | | Validation | | Testing | | |
|-------|----------|------|------------|------|---------|------|-----|
| | R^2 | RMSE | R^2 | RMSE | R^2 | RMSE | AIC |
| MLP-1 | 0.52 | 335 | 0.52 | 456 | 0.52 | 267 | 917 |
| MLP-2 | 0.61 | 303 | 0.51 | 459 | 0.48 | 276 | 927 |
| MLP-3 | 0.55 | 325 | 0.62 | 403 | 0.59 | 244 | 911 |
| MLP-4 | 0.63 | 293 | 0.7 | 360 | 0.58 | 250 | 922 |

MLP-1=PID_0,PID_1,PID_2,PID_3,PID_5：分層污染分布
MLP-2=MLP-1+Ind_rw,Source：分層污染分布、周邊井水位變化及場址洩漏源深度
MLP-3=MLP-1+SVE_p,freq,well_number：分層污染分布、設備能力、系統配接



以人工智慧結合遠端監控優化場址之現地整治成效

MLP-4=MLP-1+MLP-2+MLP-3+運轉時數：所有因素添加

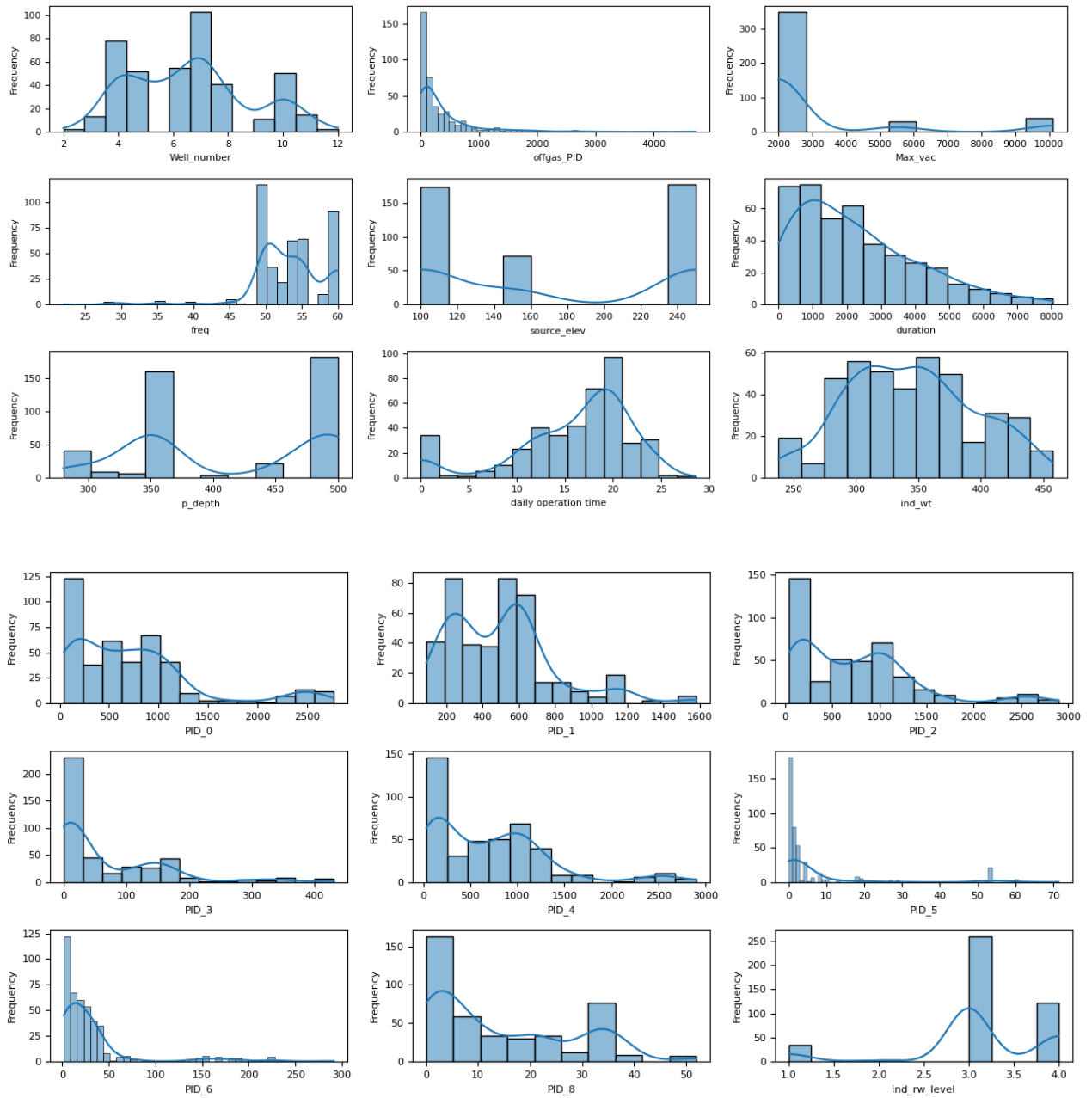
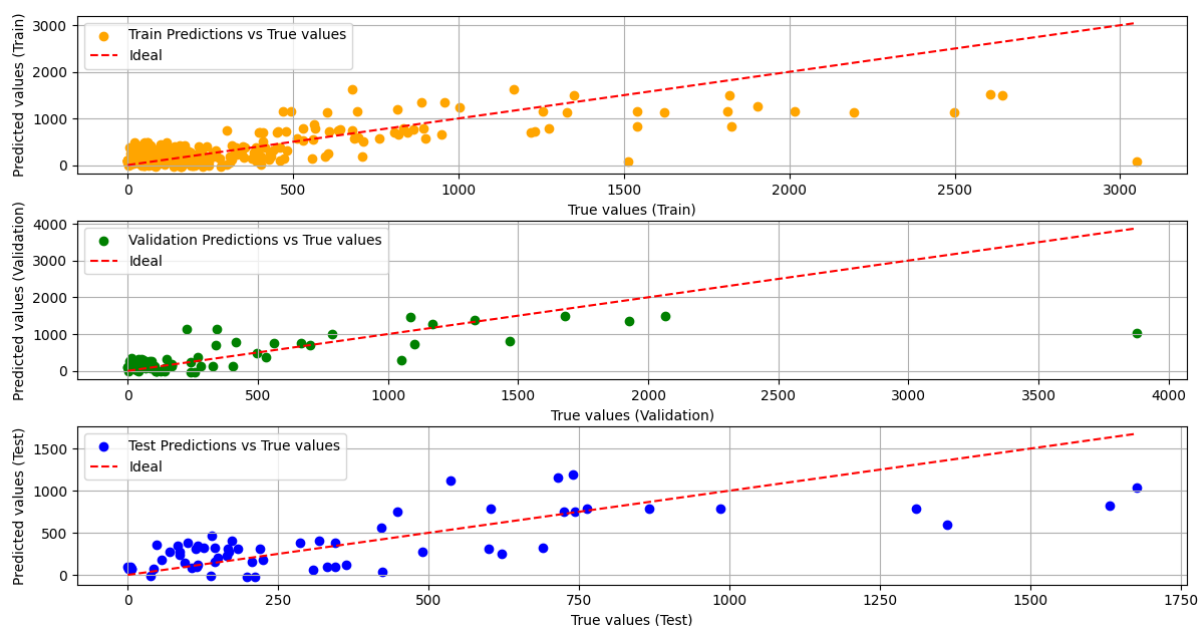
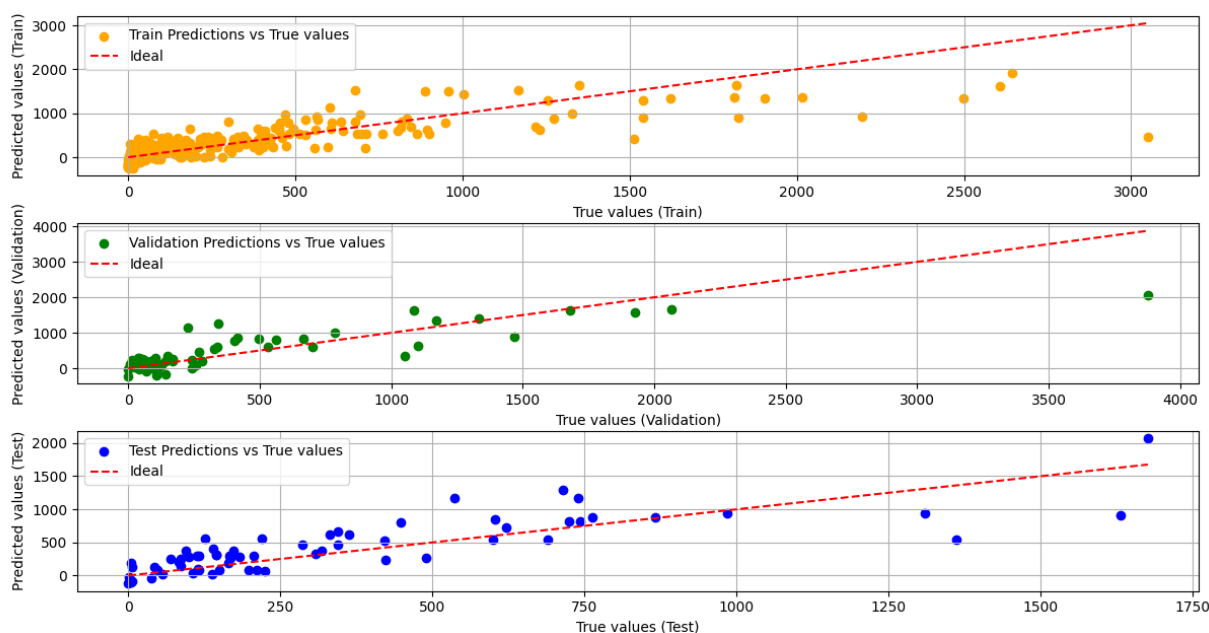


圖 4-10、不同特徵參數數據分布型式



(a)MLP-1 預測值與實際值對照圖



(b) MLP-4 預測值與實際值對照圖

圖 4-11、輸入特徵對於 MLP 預測模型的影響

3.2.4 回歸模型的選擇

本研究中採用了三種演算法，分別為 MLP、RBFN 及 KRR，在找尋最佳效能的模型下，分別對於三種演算法設定不同輸入參數、變更超參數，都從相同的輸入參數開始，再變更超參數後，調整輸入參數類別去取得最好效能的模型。模型選擇考慮了 RMSE、 R^2 及 AIC 值，表 4-8 呈現了三種演算法不同輸入參數與模型代號，經過機器學習訓練得到的效能列於表 4-9，其中，三種演算法中效能最



以人工智慧結合遠端監控優化場址之現地整治成效

佳的模型以預測值對應實際值繪製對照圖如圖 4-12，而透過各項效能指標的比較，KRR-5 的效能最好，其 AIC 值最低(875)、且無論是訓練集、驗證集及測試集的 RMSE、 R^2 都最佳，顯示其具有一定的預測成功率且泛化性也達到一錠接受的程度，但須注意 KRR 的演算方法對於數據規模愈大的情況可能會需要更長的計算時間，隨著數據集數量的增加，屆時需要再評估是否有其他替代的模型，但本研究中所提供的相關參數可作為後續選型的參考。此外，透過 MLP 不同模型的比對也可以發現場址中仍然有一些尚未考量進去的特徵可以再進一步優化模型，尚須考量的特徵包括場址抽水操作的特徵、氣候變化如大降雨的影響、地表鋪面的影響、特定設備異常的特徵摘取等等需要更多的數據才能獲得更多該特徵的變因影響。

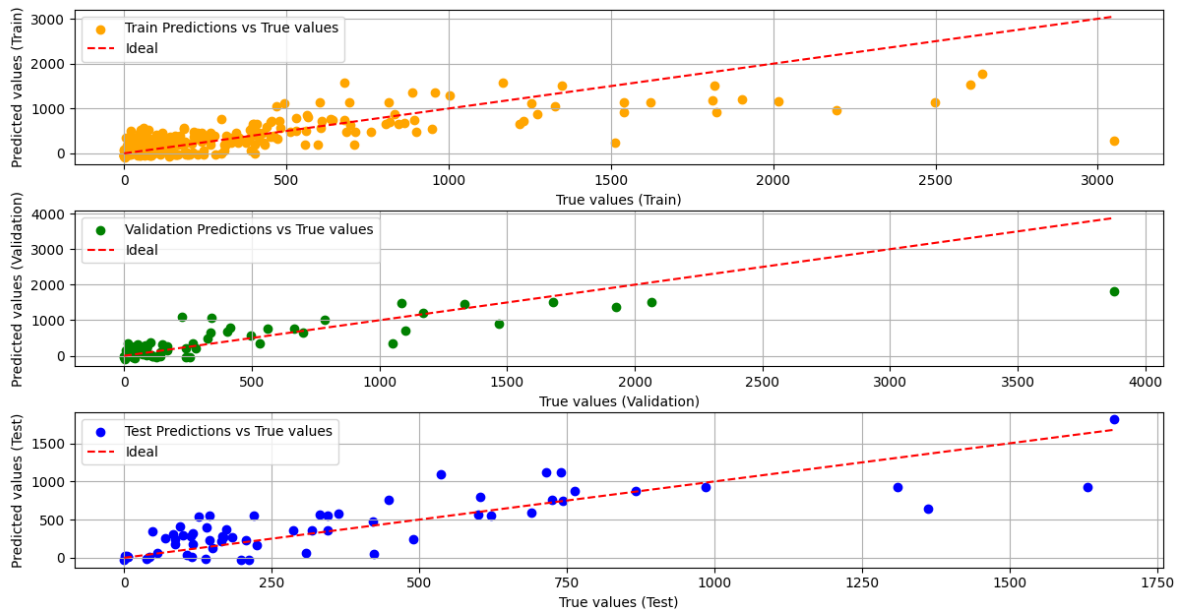
表 4-8、本計畫選擇模型輸入值列表

| Model# | 輸入值(X) |
|--------|--|
| MLP-1 | 'PID_0','PID_1','PID_2','PID_3','PID_5' |
| MLP-2 | 'PID_0','PID_1','PID_2','PID_3','PID_5','ind_rw','source','p_depth' |
| MLP-3 | 'PID_0','PID_1','PID_2','PID_3','PID_5','wn','SVE_p','freq' |
| MLP-4 | 'PID_0','PID_1','PID_2','PID_3','PID_5','ind_rw','SVE_p','dt','wn','freq' |
| MLP-5 | 'PID_0','PID_1','PID_2','PID_3','PID_5','ind_rw','SVE_p','dt','wn','freq','source','p_depth' |
| MLP-6 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','freq','dt','p_depth','ind_rw','wn' |
| RBFN-1 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6' |
| RBFN-2 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','dt' |
| RBFN-3 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','dt' |
| RBFN-4 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','dt','p_depth','SVE_p' |
| RBFN-5 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','dt','p_depth','SVE_p','freq' |
| KRR-1 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6' |
| KRR-2 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','dt' |
| KRR-3 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','dt' |
| KRR-4 | 'PID_0','PID_1','PID_2','PID_3','PID_5','PID_6','dt','p_depth','SVE_p' |
| KRR-5 | 'PID_0','PID_1','PID_2','PID_3','PID_5','p_depth','SVE_p','freq','dt','source','wn' |



表 4-9、不同模型對於尾氣濃度之預測成效

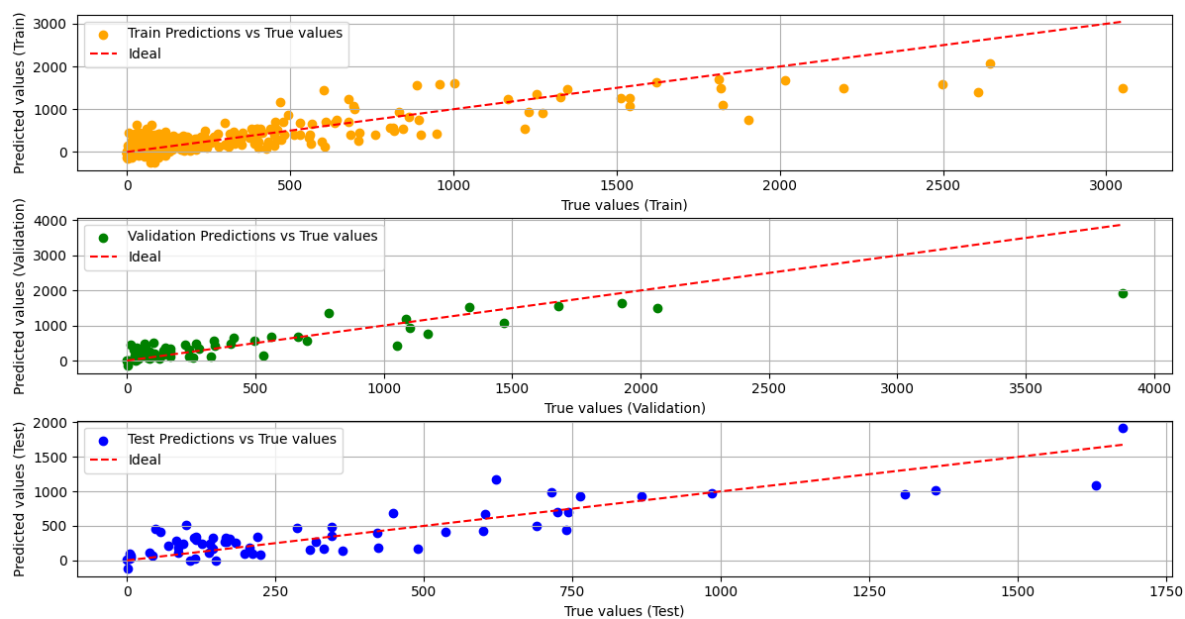
| Model# | Training | | Validation | | Testing | | | hyperparameters | | |
|---------------|----------------|------------|----------------|------------|----------------|------------|------------|-----------------|---------------|---------------|
| | R ² | RMSE | R ² | RMSE | R ² | RMSE | AIC | RBN# | alpha | gamma |
| MLP-1 | 0.52 | 335 | 0.52 | 456 | 0.52 | 267 | 917 | — | — | — |
| MLP-2 | 0.61 | 303 | 0.51 | 459 | 0.48 | 276 | 927 | — | — | — |
| MLP-3 | 0.55 | 325 | 0.62 | 403 | 0.59 | 244 | 911 | — | — | — |
| MLP-4 | 0.6 | 304 | 0.66 | 385 | 0.58 | 248 | 917 | — | — | — |
| MLP-5 | 0.63 | 293 | 0.7 | 360 | 0.58 | 250 | 922 | — | — | — |
| MLP-6 | 0.61 | 302 | 0.67 | 377 | 0.66 | 223 | 906 | — | — | — |
| RBFN-1 | 0.59 | 309 | 0.56 | 436 | 0.57 | 252 | 952 | 30 | 0.01 | 0.01 |
| RBFN-2 | 0.67 | 279 | 0.66 | 381 | 0.68 | 217 | 933 | 30 | 0.01 | 0.01 |
| RBFN-3 | 0.7 | 263 | 0.72 | 347 | 0.65 | 228 | 939 | 30 | 0.003 | 0.01 |
| RBFN-4 | 0.71 | 259 | 0.65 | 390 | 0.68 | 217 | 932 | 30 | 0.015 | 0.05 |
| RBFN-5 | 0.71 | 259 | 0.74 | 334 | 0.73 | 198 | 921 | 30 | 0.05 | 0.08 |
| KRR-1 | 0.67 | 279 | 0.66 | 381 | 0.57 | 251 | 901 | — | 0.00192 | 0.0073 |
| KRR-2 | 0.77 | 230 | 0.68 | 373 | 0.65 | 227 | 890 | — | 0.00192 | 0.0073 |
| KRR-3 | 0.79 | 221 | 0.7 | 358 | 0.65 | 227 | 890 | — | 0.001 | 0.0073 |
| KRR-4 | 0.79 | 220 | 0.75 | 327 | 0.72 | 203 | 880 | — | 0.0048 | 0.0077 |
| KRR-5 | 0.87 | 174 | 0.75 | 327 | 0.76 | 190 | 875 | — | 0.0018 | 0.0075 |



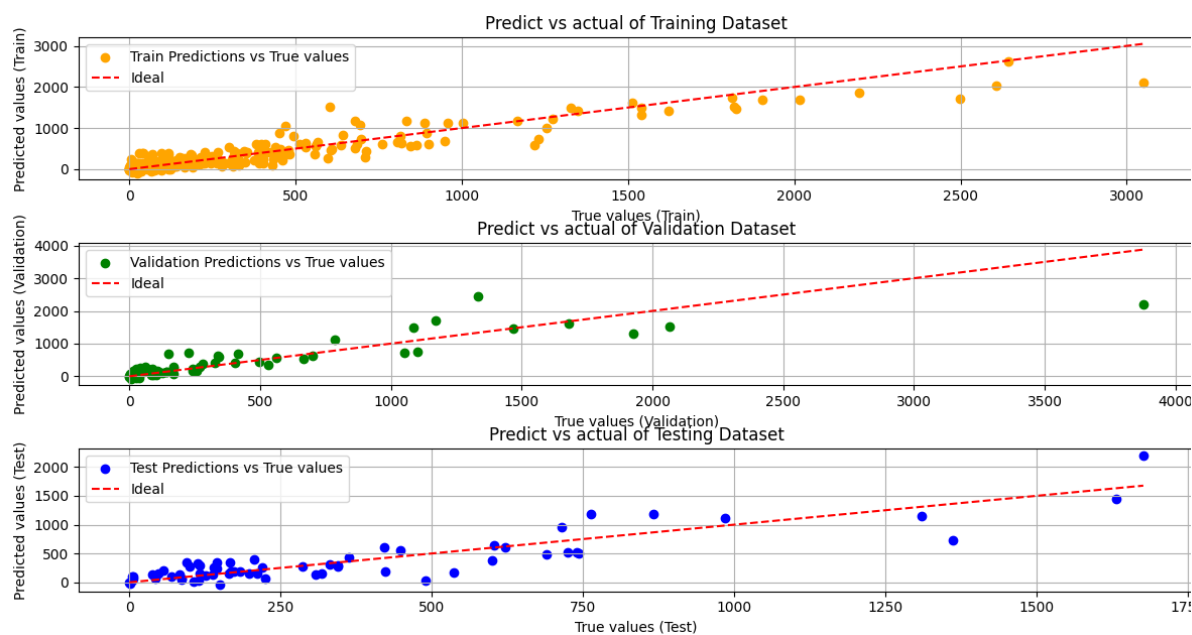
(A)MLP-6 預測值與實際值對照圖



以人工智慧結合遠端監控優化場址之現地整治成效



(B)RBFN-5 預測值與實際值對照圖



(C)KRR-5 預測值與實際值對照圖

圖 4-12、不同演算法最佳模型以預測值與實際值對比圖

3.2.5 尾氣濃度分類的學習成果

本年度研究中將主要預測目標放在抽出氣體尾氣濃度的預測上有以下幾個考量：

- 針對移除成效的評估需利用尾氣濃度、操作時間、溫度、流量等參數進行推算，而其中尾氣濃度變異性較大，透過預測結果可以進而評估移除成效的達成情形



- 在選擇尾氣處理設備時，通常以濃度超過 750ppm 作為需考量觸媒熱處理設備的參考濃度，以維持一定熱值有利處理設備的操作適用性
- 在評估是否仍須採用雙相抽除技術時，尾氣濃度過低將會產生單位成本過高的情形，或是亦會作為替代工法選用時的參考時間點，因此，濃度過低的情況也是需要考慮，通常以濃度小於 100-150ppm 以下，會啟動各類替代工法。

針對第 2 及第 3 點，本研究將採用 SVM 的分類方法進行訓練，透過對於場址污染程度、採用設備的能力與水文地質特性進行分類的預測，成效將以混淆矩陣(confusion matrix)配合 AUC curve 呈現模型效能。

1. 預測尾氣濃度是否超過 750ppm?

在進行 SVM 的分類時，也需要經過類似迴歸時所做的測試不同輸入值及超參數後得到預測上的效能，在報告中摘要列出兩種模型，其一為僅採用分層 PID 濃度(PID_0~3、PID_5)(SVM750-1)，其二為經過類似迴歸分析的效能評估後選擇最優的模型(SVM750-2)進行預測，兩種模型都採用 $C=3$ 及 $\text{Gamma}=0.09$ ，兩者結果說明如下：

■ SVM750-1

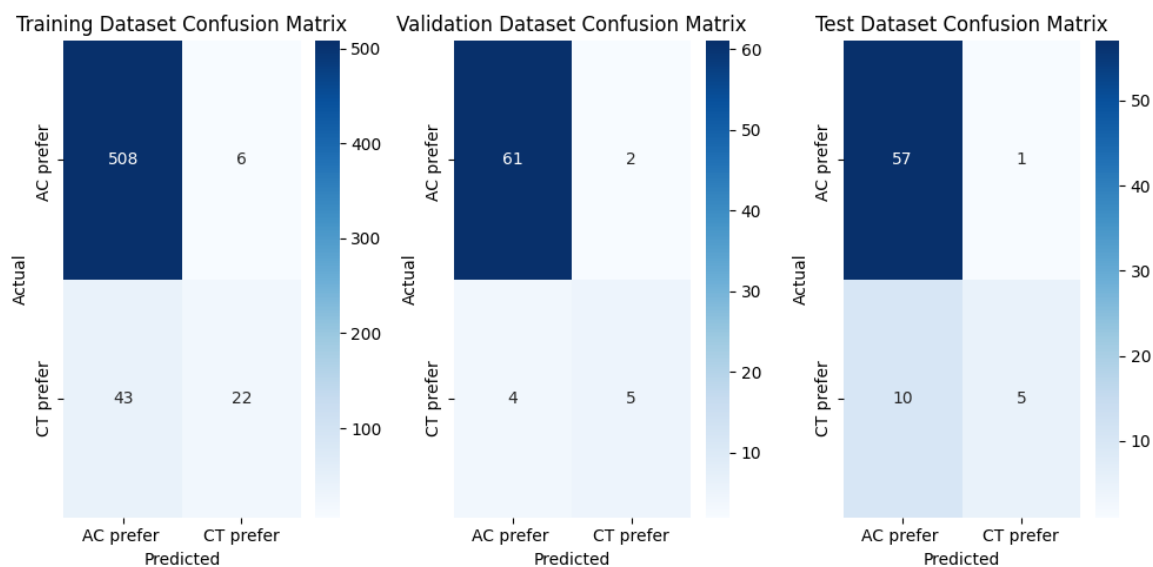
從學習結果可以看到在採用先建立場址分層污染分布圖再內插抽出井分層污染分布的輸入數據，對於預測濃度低於 750ppmv 的成效已經可以達到 80%以上，但是對於實際值為超過 750ppmv 以上的預測能力仍然較差，從測試集的結果來看 recall 值只達到 33%(5/15)(如圖 4-13(a)，成效並不好，整體 accuracy 約達 85%。

■ SVM750-2

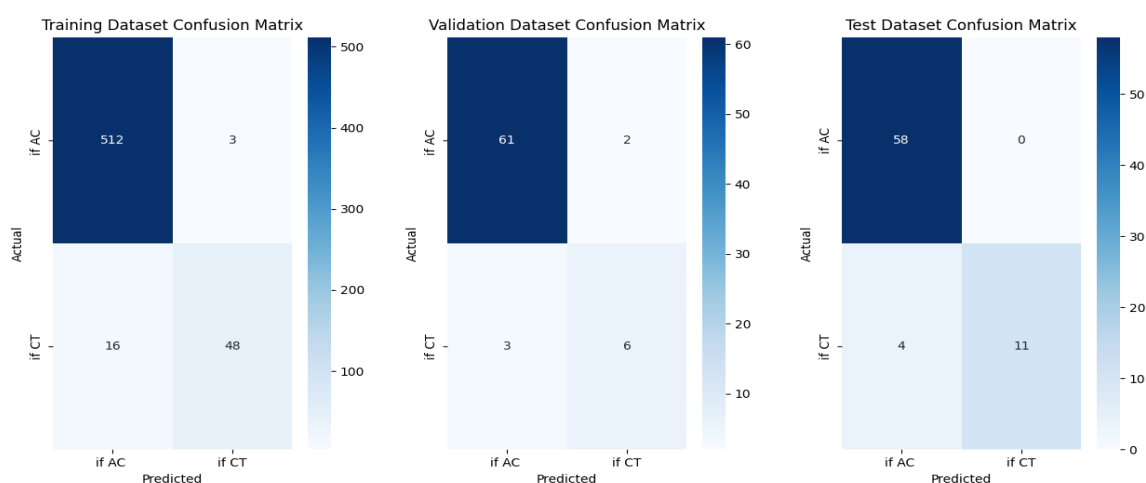
除了分層污染特性外，將設備操作能力、場址水位、污染源、泵控制深度等特徵加入後，SVM750-2 呈現了較佳的效能(如圖 4-13(b)，58 筆小於 750ppm 的尾氣實測值中，模型都預測正確($\text{Recall}=1$)，而預測出 62 筆小於 750ppm 的樣本中，正確的有 58 筆($\text{Precision}=0.94$)；而對於 15 筆超過 750ppm 的數據，則僅預測出 11 筆($\text{Recall}=0.73$)，但預測會大於 750ppm 的 11 筆，則全部正確($\text{Precision}=1$)，顯示模型對於實際大於 750ppm 的預測能力較弱，仍有些樣品未能正確預測出來。整體的準確率仍達到 0.95。而以 ROC curve 來檢視訓練集、驗證集與測試集的效能(如圖 4-14)，預測模型對於三個數據集 AUC 都能達到 90%以上的效能。



以人工智慧結合遠端監控優化場址之現地整治成效



(a)SVM750-1 模型混淆矩陣圖



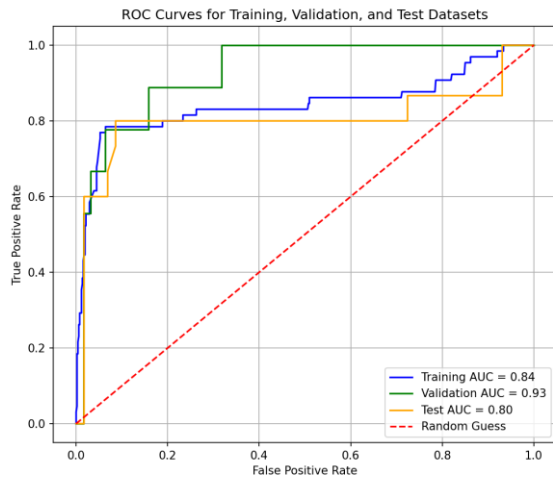
(b)SVM750-2 模型混淆矩陣圖

| model | precision of AC prefer | precision of CT prefer | Recall of AC prefer | Recall of CT prefer | Accuracy | AUC |
|----------|---------------------------|---------------------------|------------------------|------------------------|----------|------|
| SVM750-1 | 0.85 | 0.83 | 0.98 | 0.33 | 0.85 | 0.8 |
| SVM750-2 | 0.94 | 1 | 1 | 0.73 | 0.95 | 0.91 |

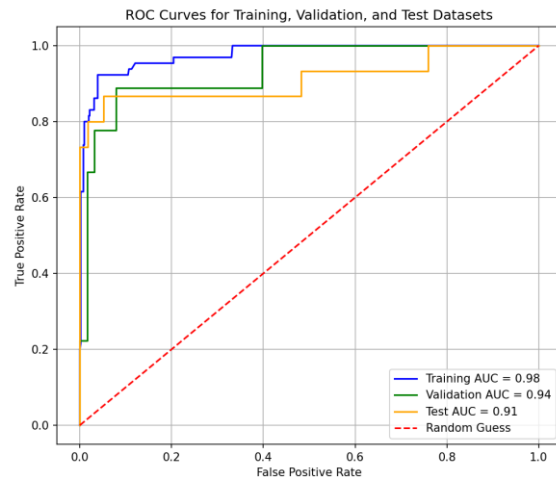
AC prefer：預測尾氣值是否低於 750ppmv 而建議用活性碳(AC)處理廢氣

CT prefer：預測尾氣值是否高於 750ppmv 而建議用觸媒熱處理(CT)廢氣

圖 4-13、以 SVM 預測 750ppmv 成效對照圖



(a)SVM750-1 ROC 曲線圖



(b)SVM750-2 ROC 曲線圖

圖 4-14、SVM 針對 750ppm 分類模型 ROC 曲線評估選型

2. 預測尾氣濃度是否小於 150ppm?

■ SVM150-1

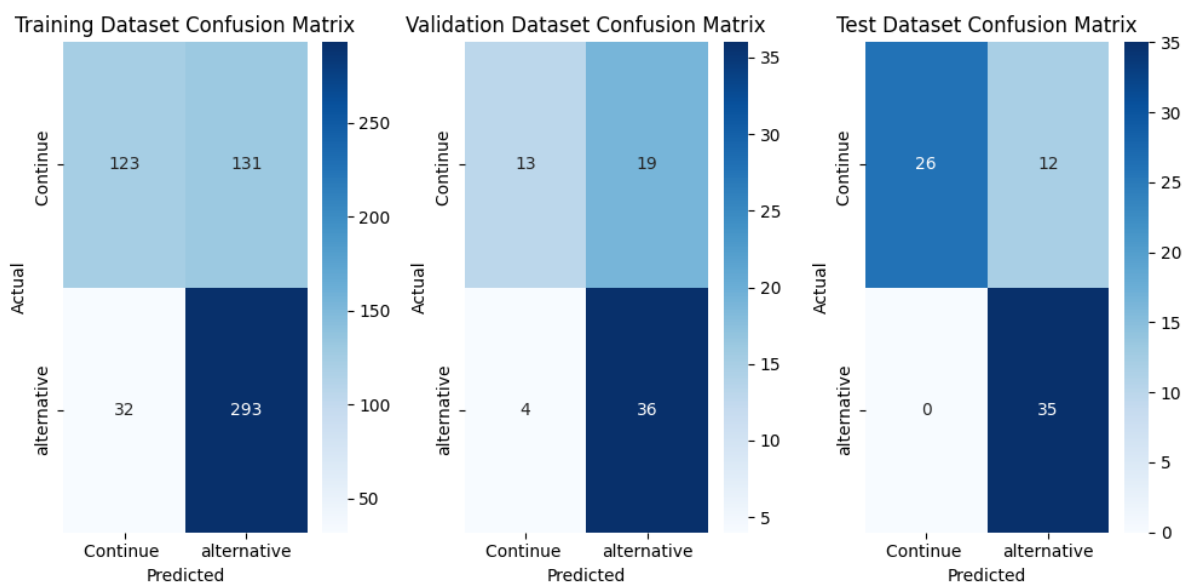
從學習結果可以看到在採用先建立場址分層污染分布圖再內插抽出井分層污染分布的輸入數據，對於濃度低於 150ppmv 的在分類上的正確率是高的，測試集的甚至達到 100%以上(圖 4-15(a))，但在預測精確度上來看，則約僅有 74%的正確性，整體 accuracy 約達 84%，而從 ROC 圖(圖 4-16(a))可以看到驗證集的預測準確度特別差(68%)，顯示在泛化能力上還可以接受，AUC 值也未達 90%。

■ SVM150-2

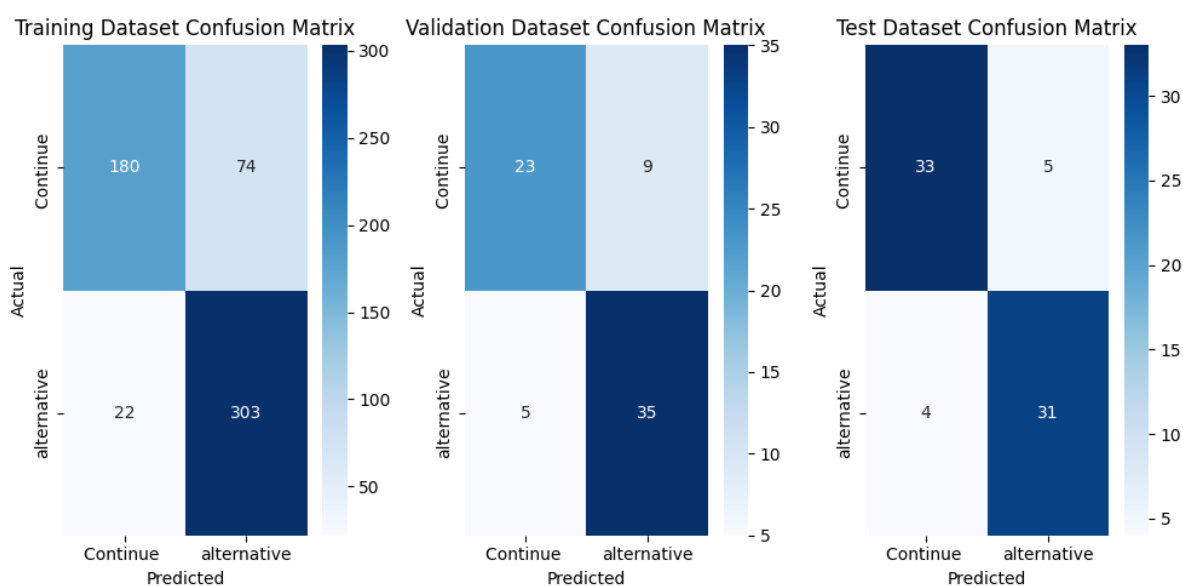
除了分層污染特性外，將設備操作能力、場址水位、污染源、泵控制深度等特徵加入後，SVM750-2 呈現了較佳的效能(如圖 4-13(b))，其中，無論 precision, recall 都呈現了平衡沒有特別偏向的結果，整體準確度達到 88%，AUC 達到 0.95。而以 ROC curve 來檢視訓練集、驗證集與測試集的效能(如圖 4-14)，預測模型中驗證集的效能較差約 0.87，訓練與測試集都能達到 90%以上。由於 150ppmv 相當於目標值分布的中位數，在樣本數相較比較均衡條件下，相對更高值的準確度會相對較差，而對比以 750ppm 做篩選值確實也得到類似的結果。



以人工智慧結合遠端監控優化場址之現地整治成效



(a)SVM150-1 模型混淆矩陣圖



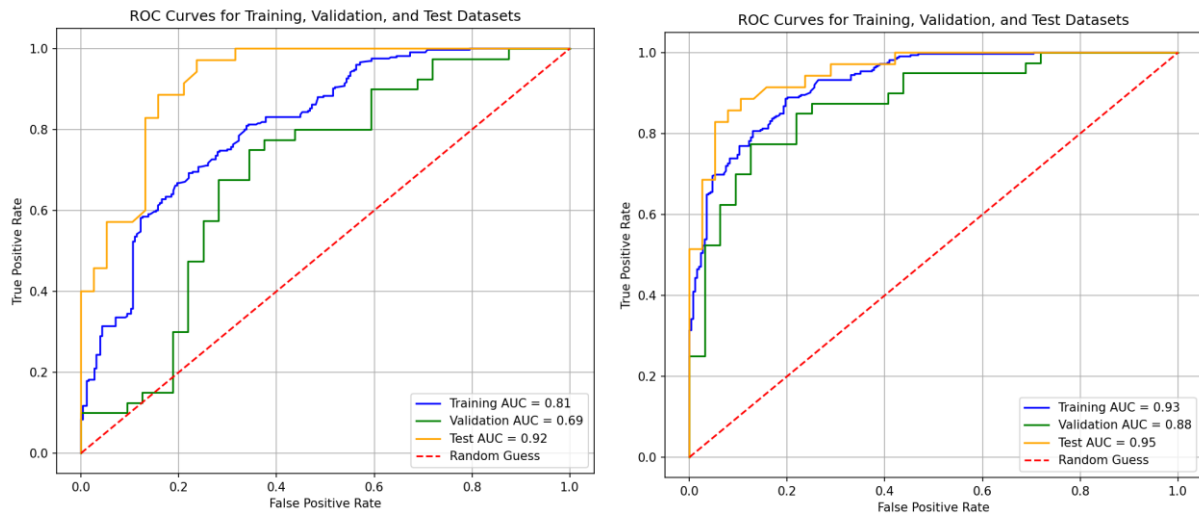
(b)SVM150-2 模型混淆矩陣圖

| model | precision of continue | precision of alternative | Recall of continue | Recall of alternative | Accuracy | AUC |
|----------|--------------------------|-----------------------------|-----------------------|--------------------------|----------|------|
| SVM150-1 | 1 | 0.74 | 0.68 | 1 | 0.84 | 0.92 |
| SVM150-2 | 0.89 | 0.86 | 0.87 | 0.89 | 0.88 | 0.95 |

Continue：預測尾氣值是否高於 150ppmv 而建議持續用雙相抽除法處理

Alternative：預測尾氣值是否低於 150ppmv 而建議評估替代方案

圖 4-15、以 SVM 預測 150ppmv 成效對照圖



(a)SVM150-1 ROC 曲線圖

(b)SVM150-2 ROC 曲線圖

圖 4-16、SVM 針對 150ppmv 分類模型 ROC 曲線評估選型

4.3 遠端監控設備設置情形

本研究中配置遠端監控設備取得的數據，將以可以滿足前述既有場址利用機器學習中所發現重要的特徵參數作為後續實場繼續學習與驗證的基礎，同時在實場中透過超參數調整得到最佳化的模型下，做到邊緣計算後，提供現場操作人員更即時的操作調整建議，其中，參考 KRR-5 的預測模型涉及的特徵參數為例，摘要說明調整上的決策影響到整治過程的成本與效益如表 4-10，後續將會在邊緣設備端利用工業電腦安置於現場提供判斷依據，俾利現場人員參考後決定動作方針。本研究為了解決抽提井的污染分層特徵所建立的 CSMcheck 程序可以輔助專案經理即時的評估場址污染分布特性，尾氣與真空度的預測結果可以提供有用的資訊輔助人員進行異常檢查與排除，並透過實場中蒐集即時的異常發生資訊，在整理既有的場址特徵時，涉及設備故障或場址特定變化的紀錄往往不及時造成應變數據的特徵無法對應，而此也是實場上可以提高觀察並納入特徵參數以利模型優化。

目前現場已於 8 月底安裝完成兩套監控系統及相關的傳感器，安裝完成迄今一直進行系統穩定性測試、數據下載、上傳及軟件因應更新，現場施工完成照片如圖 4-18 所示，相關設備配置位置如圖 4-17，茲就兩套系統目前運行狀況說明如後。

1. SVE 設備監控系統：

本系統組裝架構如圖 4-19，主要包括 PLC、變頻控制器、4G 路由器及 HMI 面板，可以收集到變頻器頻率、抽氣流量、真空度、溫度、抽水流量、運轉時間



以人工智慧結合遠端監控優化場址之現地整治成效

等資訊，並且可以遠端控制頻率與操作時間的調整。本系統的 PLC 採用西門子 S7-200，透過 Modbus 協議傳輸數據以 4G 通訊路由器傳送到遠端平台，遠端平台目前架設於合作公司的獨立伺服器一台，該平台係由合作公司提供做為數據下載使用，本研究並未涉及該平台管理。在安裝測試期間，原先由合作公司提供的平台存放區域為一虛擬空間，容易受到防火牆影響無法登入上傳資料，爾後經過協商，合作公司同意提供單獨一台伺服器電腦作為平台後台使用 (<http://114.32.253.50:20000/>)，可以透過平台進行數據下載，採取獨立伺服器已初步改善原先防火牆阻擋的影響，目前仍在持續測試中。

2. PID 尾氣監測系統：

本系統組裝架構如圖 4-20，主要包括 PLC、三組泵(P1~P3)(氣體取樣、水取樣及曝氣泵)、4 組氣體閥(A0-A3)及 3 組水閥(W1-W3)。透過邏輯控制多管道分時間取樣動作，為確保取樣品質，每批次管道都會用乾淨空氣排空，每批次完成取樣的時間約 25 分鐘，偵測的邏輯將配合 SVE 的運轉時間進行設定。由於 PID 尾氣監測系統為本研究中設計開發出來，尚未有現成的平台可以使用，目前相關數據先採用 HMI 下載功能以記憶卡存儲，亦已經安裝 4G 路由器將數據上傳至遠端伺服器，目前尚在測試中，相關數據暫時以記憶卡存儲並定時取用。

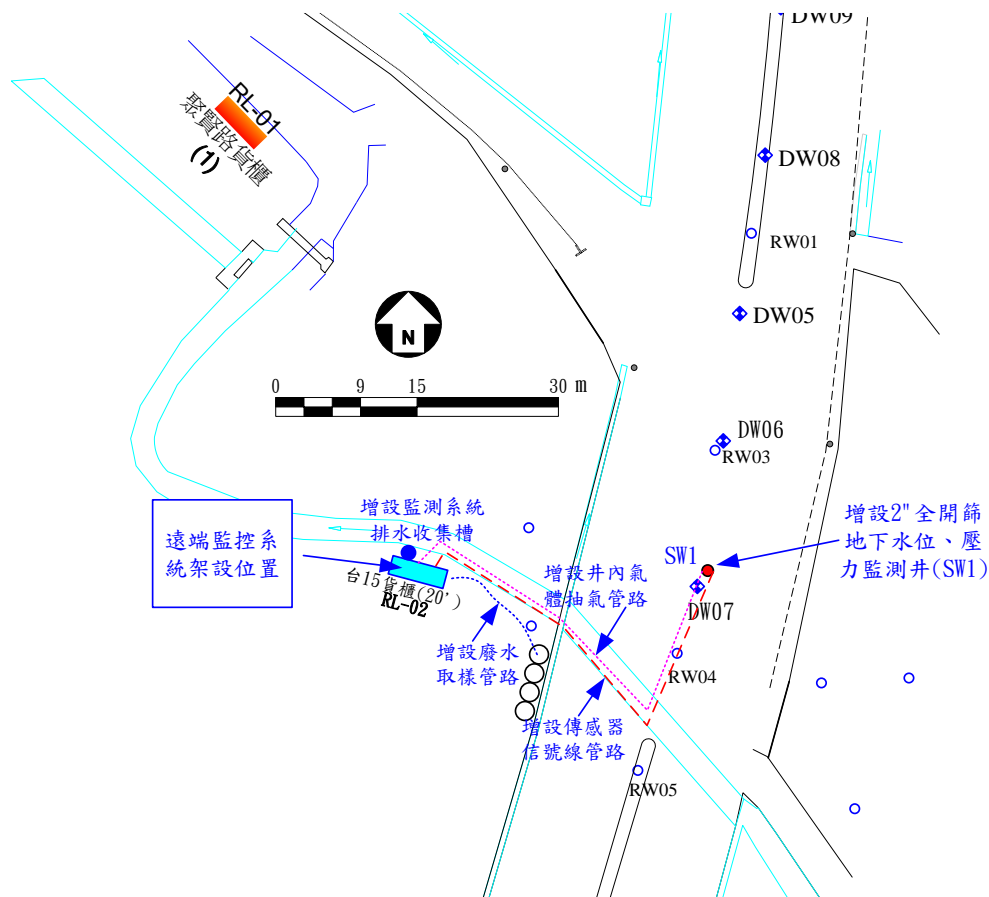




圖 4-17、本研究中增設試驗裝置之現場設施配置圖



抽除設備控制系統安裝及測試



排出氣體流量計



排出氣體壓力計



排出氣體溫度計



氣液分離桶排水流量計

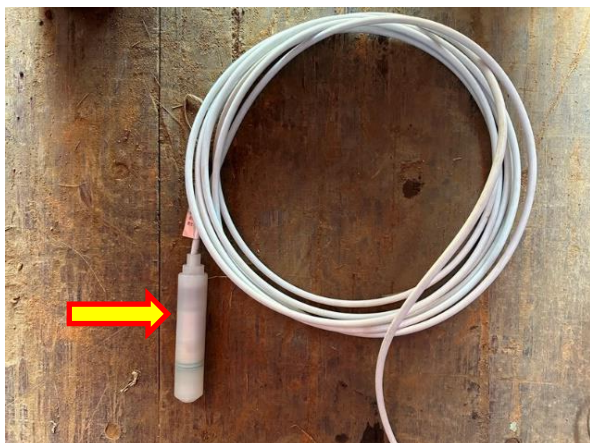


抽除系統真空度偵測器

圖 4-18、遠端監控系統現場安裝完成照片



以人工智慧結合遠端監控優化場址之現地整治成效



監測井內投入式水位計



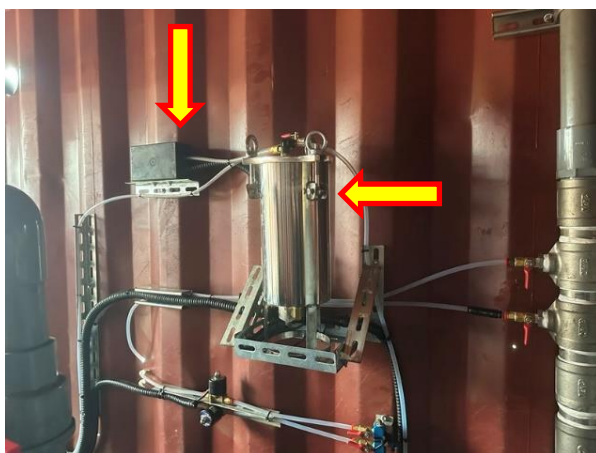
監測井內氣壓計



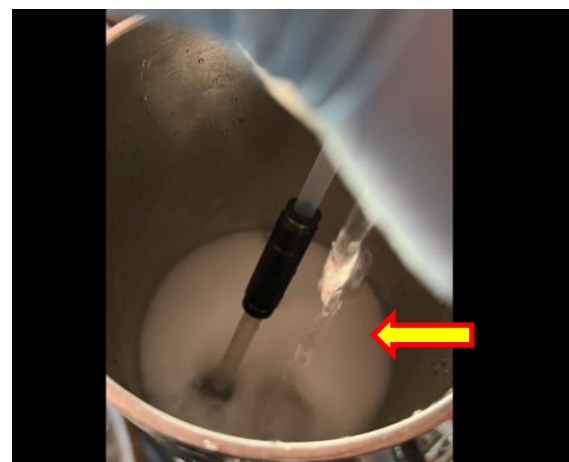
水位(氣壓)監測井設置



監測井水位計、氣壓計及管線安裝



頂空濃度取樣曝氣裝置



取樣裝置內曝氣盤

圖 4-18、遠端監控系統現場安裝完成照片(續)



第四章 結果與討論



取樣暫存槽由現場現有設備利用



廢水系統取樣蠕動泵



尾氣監測與設備管路配置



尾氣監測與設備管路配置



系統測試及運作



系統測試及運作

圖 4-18、遠端監控系統現場安裝完成照片(續)



以人工智慧結合遠端監控優化場址之現地整治成效

The DPE PLC control system (finished)



圖 4-19、已完成 SVE 遠端數據監控單元與遠端下載平台界面



表 4-10、KRR-5 模型於 DPE 整治系統調整之應用與效益

| 情境 | 影響因素說明與因應行動 | 成本效益影響說明 |
|---|---|---|
| KRR-5 預測濃度等於實測濃度 (定義預測濃度與實測濃度的誤差範圍為相等) | 表示 CSM 跟現地 DPE 系統操作參數相符，系統維持原設定條件持續運轉 | 減少人員巡視頻率，若變化週期以 30 天為單位，原先每周一次 1hr(通常控制計畫的頻率)改為發生變化時去修正，則人力上從每月 4hr 變成 1hr，人力、差旅成本減少 75%。 |
| KRR-5 預測濃度大於實測濃度 (定義預測濃度與實測濃度經過一段時間的誤差範圍下都是大於的情況(可以用統計檢驗的方法定義之)) | <p>1.CSM 可能需要進行修正，建議監測點位與項目(產出 csmcheck 的圖像資料標示最高點供現場精準調查)</p> <p>2.涉及 KRR-5 的操作參數如抽水泵浦深度、設備操作頻率、設備連續運轉時間、抽取井口數須於實際系統中調整，現場檢視參數條件是否仍與原先輸入值相符並進行問題排除與參數調整，以減少錯誤資訊的操作狀態持續進行的時間</p> <p>3.模型應進行修正與增加預測錯誤數據重新訓練導入新模型，根據 1,2 檢視後的數據彙整加入訓練集，重新修正模型重新導入使用</p> | <p>1.針對特定點進行及時調查，可以加速成效評估速度，例如原先的三個月一期頻率所得到的判斷，倘若提前一個月做判斷，可以縮減 33%的低效操作時間</p> <p>2. 操作上面的影響可以分成以下三部分描述效益：</p> <p>整治技術提升：從 AI 建議的排查數跟原先專案人員經過數據分析後的排查數比例可以看出提升效果，例如 AI 建議當月排查次數有 5 件，按照專案若原先每月進度報告分析 1 次 3 件來看，增加 2 件，技術提高比例為 $(5-3)/3$</p> <p>整治成本減少：從原先操作條件例如每日抽 5 小時休 1 小時，透過 PID 實時監測及 AI 預測調整操作條件為抽 4 小時休 2 小時，移除量一樣，則每日操作時數從 20 小時降為 16 小時，操作成本減少比例為 $(20-16)/20$</p> <p>能源效率提升：同上，原先 5 小時的移除量，透過監控與 AI 預測於 4 小時達到相當的移除量，則能源效率提升比例為 $(5-4)/5$</p> <p>3.累積預測較差的數據集繼續訓練並更新模型，以提高模型效能。</p> |

以人工智慧結合遠端監控優化場址之現地整治成效

| | | |
|--|--|---|
| KRR-5 預測濃度大於實測濃度 (定義預測濃度與實測濃度經過一段時間的誤差範圍下都是小於的情況) | 1.CSM 可能需要進行修正，建議監測點位與項目(產出 csmcheck 的圖像資料標示最高點供現場精準調查) 2.模型應進行修正與增加預測錯誤數據重新訓練導入新模型 | 1.針對特定點進行及時調查，可以加速成效評估速度，例如原先的三個月一期頻率所得到的判斷，倘若提前一個月做判斷，可以縮減 33%的低效操作時間 2. 累積預測較差的數據集繼續訓練並更新模型，以提高模型效能。 |
|--|--|---|

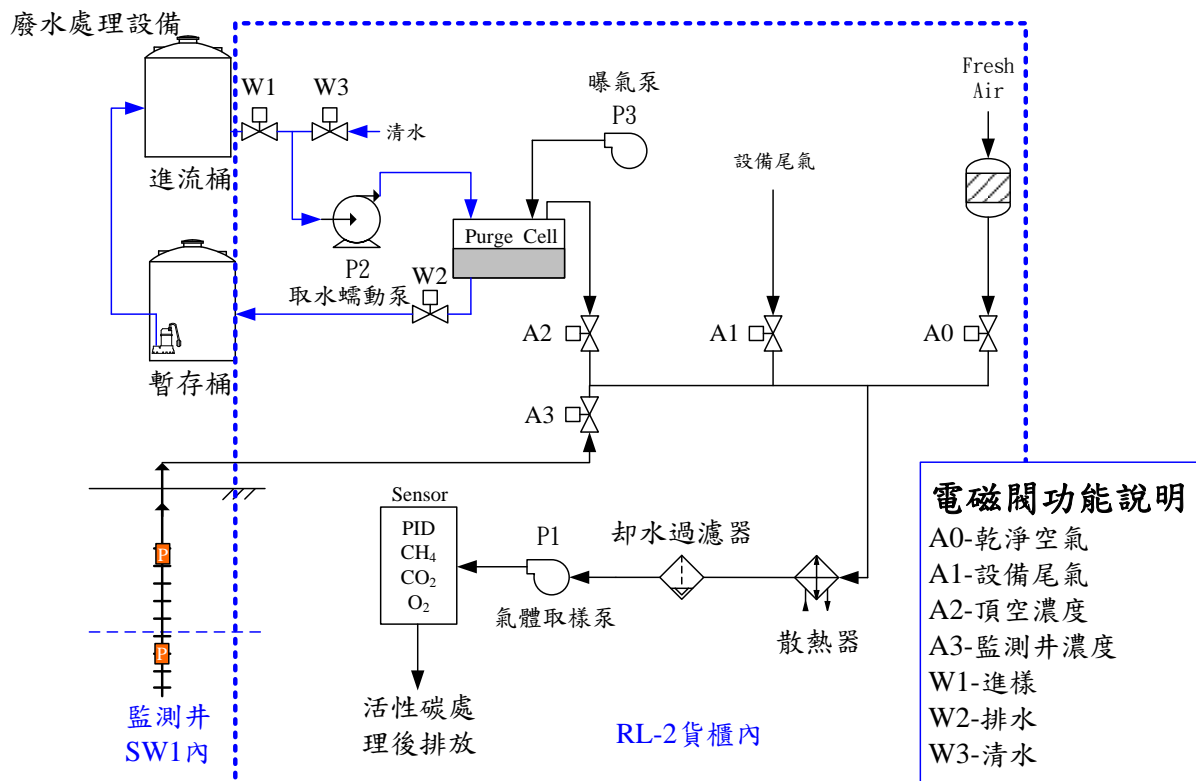


圖 4-20、PID 尾氣監測系統架構圖

4.4 初步工作成果與後續工作計畫

1. 根據目前已完成的研究成果總結如下發現：

- 在進行期中報告的研究階段，採取粗分飽和層與未飽和層的數據特徵不夠明顯，且在推估抽提井代表性的分層污染特性上，不容易進行檢查。因此，本報告中採取先建立場址污染分層分布的圖像(CSMcheck)後，再透過場址經理協助確認分布合理性後，無論採取何種演算法，對於目標值的預測能力都大幅提升，亦即機器學習的數據品質重大的影響了學習效果。而原先擔憂不同場址是否存在差異而無法綜合數據進行學習的部分，在將分層以水位層為 0



層區別污染層並添加場址洩漏源深度及場址整治設備真空度能力的參數後，多場址(本研究報告選擇已經專案經理確認 CSM 的 3 處)的訓練效能並未因此變差，但會發現效能會受到輸入參數與超參數設定影響，後續需要建立可以檢核模型的程序以利繼續學習更多數據時快速修正模型。

- 在進行場址污染分層分布後內插抽提井的周邊污染情形採取了 RBF 的方式，其中，選擇用 linear 的核函數相較多項式(cubic)的核函數對於後續機器學習的效果更佳。
- 本研究中採用了 MLP、RBFN 及 KRR 針對尾氣濃度預測進行了學習訓練，結果顯示輸入值涵蓋採用 $X = ['PID_0', 'PID_1', 'PID_2', 'PID_3', 'PID_5', 'p_depth', 'SVE_p', 'freq', 'dt', 'source', 'wn']$ 的 KRR 的效能最佳($R^2_{train} = 0.87, R^2_{validation} = 0.75, R^2_{testing} = 0.76$)，選用的超參數分別為 $\alpha = \gamma =$ ，而在尋求最佳效能的超參數調整過程，可以發現對於效能提升可以達到 10-20%的 R^2 的提高，而針對不同輸入參數亦必須調整超參數才能得到最佳的效能。
- 本研究採用 SVM 的方法針對尾氣濃度 750ppmV 跟 150ppmV 進行了分類預測，其中在以 $X = ['PID_0', 'PID_1', 'PID_2', 'PID_3', 'PID_5', 'p_depth', 'SVE_p', 'freq', 'source', 'wn', 'ind_rw']$ 作為輸入值，超參數 $C=3, \text{Gamma}=0.09$ 下，對於是否超過 750ppm 的預測效能最佳，準確度可以達到 0.95，惟在實際超過 750ppmV 的預測上仍只有 0.73 的 recall，顯示模型可能還欠缺更多的特徵參數去彌補這部分的不足；而在 150ppmV 的分類預測模型上，則呈現較為均衡的效能(0.86-0.89)，但準確度較低為 0.88，其所選擇的超參數為 $C=2, \text{Gamma}=0.075$ 。
- 在遠端監控數據管理上，為了能得到更多特徵數據，往往希望能有更多數據取得，但數據的存儲空間涉及到成本費用，邊緣計算有其必要性，本計畫後期將會篩選需要邊緣計算的特徵以更符合經濟有效性。

2.根據第一年的執行成果，綜合本計畫目標將持續進行以下工作：

- 本年度建立了數據預處理的方法以及提高數據品質的方法，後續將沿用並發掘更多場址特徵以提高預測效能。
- 本年度尚未針對移除效率(污染物移除量需綜合溫度、流量、操作時數)及整治效能(需考慮單位移除量所需耗能或碳排的檢核)作為機器學習目標來修正操作參數，後續將進一步增加目標值並評估對於場址操作參數是否有不同的決定。



- 後續將針對整治工法的生命週期進行評估，並同前一項工作納入整治效能的評估因子做為目標值繼續學習。
- 本年度已著手針對邊緣計算器進行資料收集並規劃兼容的系統架構，下一年度將落實邊緣計算的作業。
- 本年度蒐集的資料中對於場址發生異常的情況(如設備故障、水文地質變異、地表鋪面或其他整治行為影響等)納入特徵參數進行分析，主因為過去資料並未針對異常情況提供異常的起始點，導致該特徵對應數據無法應用，在實場的作業中，除了持續修正模型、繼續學習外，亦將會納入本年度未考慮的影響整治成效的特徵為因子持續提高模型的效能。



第五章 參考文獻

1. 行政院環保署，2015.8 「土壤及地下水油品類污染整治作業參考指引」，EPA-25-III-06B-2015-002.
2. USEPA, 2004, “How to evaluate alternative cleanup technologies for underground storage tank sites-Aguid for corrective action plan reviewers”, EPA 510-R04-002.
3. USEPA, 1999, “Multi-Phase Extraction: State-of-the-Practice”, EPA 542-R-99-004.
4. US Army Corps of Engineers, 1999, “Engineering and Design : Multi-Phase Extraction”, EM 1110-1-4010.
5. Naval Facilities Engineering Service Center, Battelle, 2001, “Air Sparging Guidance Document”, NFESC Technical Report, TR-2193-ENV.
6. ITRC, 2004, “Remediation Process Optimization: Identifying Opportunities for Enhanced and More Efficient Site Remediation”
7. ITRC, 2006, “Technology Overview: Exit Strategy-Seeing the Forest Beyond the Trees”.
8. ITRC, 2006, “Technology Overview: Life Cycle Cost Analysis”
9. ITRC, 2006, “Technology Overview: Above Ground Treatment Technologies”.
10. Mineraud, J., Mazhelis, O., and Su, X., Tarkom, S., 2016, “A gap analysis of Internet-of-Things platforms”, Computer. Communications Vol. 89, 5-16.
11. Yang, S. and Chang, L., 2012, “Regional Inundation Forecasting Using Machine Learning Techniques with the Internet of Things”, Water Vol. 6, 65-80.
12. Ziaosheng Qin, G.H. Huang, 2008, “Numerical Simulation of Dual-Phase Vacuum Extraction to remove Nonaqueous Phase Liquids in Subsurface”, Water Resources Research 44(4).
13. Wang, X., Ratnaweera, H., Holm, J. A., and Olsbu, V. J. J. o. e. m., 2017, “Statistical monitoring and dynamic simulation of a wastewater treatment plant: a combined approach to achieve model predictive control. ” Journal of environmental



- management ,193, 1-7
14. Sarkar, A., Pandey, P., 2015, “River water quality modelling using artificial neural network technique.” *Aquat. Procedia* 4, 1070–1077
 - 15.. 鄭詠紘、駱尚廉，2021「建置智慧物聯網工業園區以即時監測管理廢污水系統」，*土木期刊* 110 年 10 月第 33 卷第 6 期.
 16. Ryan Haggerty , Jianxin Sun , Hongfeng Yu , Yusong Li, 2023, “Application of machine learning in groundwater quality modeling - A comprehensive review”, *Water Research* 233, 119745
 17. Jagadeesh Kumar Janga , Krishna R. Reddy , K.V.N.S. Raviteja, 2023, “Integrating artificial intelligence, machine learning, and deep learning approaches into remediation of contaminated sites: A review”, *Chemosphere* 345, 140476
 18. Xueyuan Kang , Amalia Kokkinaki , Christopher Power , Peter K. Kitanidis , Xiaoqing Shi ,Limin Duan , Tingxi Liu , Jichun Wu, 2021, “Integrating deep learning-based data assimilation and hydrogeophysical data for improved monitoring of DNAPL source zones during remediation”, *Journal of Hydrology* Vol. 601.
 19. Xueyuan Kang, Amalia Kokkinaki, Peter K. Kitanidis, Xiaoqing Shi1, Jonghyun Lee, Shaoxing Mo, and Jichun Wu, 2021, “Hydrogeophysical Characterization of Nonstationary DNAPL Source Zones by Integrating a Convolutional Variational Autoencoder and Ensemble Smoother”, *Water Resour. Res.* 57.
 20. C.C. Mei, Z. Cheng, C.-O. Ng, 2002, “A model for flow induced by steady air ventingand air sparging”, *Applied Mathematical Modelling* 26.
 21. Rohit Maurya, Umesh Kumar, 2016, “Comparative Study of Various Soils Upon Addition of Different Materials on the Basis of Hydraulic Conductivity Parameter”, *International Journal of Engineering Research* ISSN: 2278-0181 Vol. 5 Issue 5.
 22. Xueyuan Kang, Amalia Kokkinaki, Christopher Power, Peter K. Kitanidis, Xiaoqing Shi, Limin Duan, Tingxi Liu, Jichun Wu, 2021, “Integrating deep learning-based data assimilation and hydrogeophysical data for improved monitoring of DNAPL source zones during remediation”, *Journal of Hydrology* Vol. 601, 126655.
 23. Xuehang Song, Huiying Ren, Zhangshuan Hou, Xinming Lin, Marinko Karanovic, Matt Tonkin, Vicky L. Freedman, Inci Demirkanli, Rob Mackley, 2023, “Predicting future well performance for environmental remediation design using deep learning”,



以人工智慧結合遠端監控優化場址之現地整治成效

- Journal of Hydrology Vol. 617 part C, 129110.
24. Masoud Arshadi¹, M. Clara De Paolis Kaluza, Eric L. Miller, and Linda M. Abriola, 2020, “Subsurface Source Zone Characterization and Uncertainty Quantification Using Discriminative Random Fields”, Water Resources Research Vol. 56 Issue 3.
 25. Xueyuan Kang, Amalia Kokkinaki, Peter K. Kitanidis, Xiaoqing Shi, Jonghyun Lee, Shaoxing Mo, and Jichun Wu, 2021, “Hydrogeophysical Characterization of Nonstationary DNAPL Source Zones by Integrating a Convolutional Variational Autoencoder and Ensemble Smoother”, Water Resources Research Vol. 57 Issue 2.
 26. Liz McConnell, Kayvan Karimi Askarani, Kristen E. Cognac, E. Erin Mack, Craig Bartlett, Michael J. Ronayne and Jens Blotvogel, 2022, “Forecasting Groundwater Contaminant Plume Development Using Statistical and Machine Learning Methods”, Groundwater Monitoring & Remediation Vol. 42 Issue 3, pp. 34-43.
 27. Gordana Kaplan, Hakan Oktay Aydinli, Andrea Pietrelli, Fabien Mieyeville, Vincenzo Ferrara, 2022, “Oil-Contaminated Soil Modeling and Remediation Monitoring in Arid Areas Using Remote Sensing”, Remote Sensing 14(10), 2500.
 28. Khamis Al-Mahallawi, Jacky Mania, Axxedine Hani, Isam Shahrouh, 2012, “Using of neural networks for the prediction of nitrate groundwater contamination in rural and agricultural areas”, Environ Earth Sci Vol. 65, pp. 917-928.
 29. 台灣中油股份有限公司，民國 111 年，「新屋區台 15 線 53 公里 500 公尺處土壤及地下水污染控制計畫書(第三次變更)」，桃園市環境保護局。
 30. Wu, C.I., S.L. Lo, and H.C. Chung, 2022, “Automated Monitoring of a High-Speed Flocculation Flat-Bottomed Sludge Blanket Clarifier Pond during Drought and Flood Conditions”, Water, 14: 1170.
 31. Cheng, Y.Y., and S.L. Lo, 2021, “The Real Time Water Quality and Alerting for Wastewater Treatment Plant by Intelligent System”, *J. of the Chinese Institute of Civil and Hydraulic Engineering*, 33(6): 451-459.,
 32. 洪一丁, 2021, 「應用機械學習預測污水廠放流水之氨氮」, 臺灣大學環境工程學研究所碩士論文,
 33. Michelle Matos de Souza, Mart Oostrom, Mark D. White, Gerson Cardoso da Silva Jr, Maria Claudia Barbosa, 2016, “Simulation of Subsurface Multiphase Contaminant Extraction Using a Bioslurping Well Model”, Transp Porous Med 114:649–673,
 34. Shengqi Qi, Jian Luo, David O'Connor, Yidong Wang, Deyi Hou, 2020, “A



- numerical model to optimize LNAPL remediation by multiphase Extraction”, Science of The Total Environment Vol. 718, 137309.
36. Junqi Huang, Mark N. Goltz, 2017, “Analytical solutions for a soil vapor extraction model that incorporates gas phase dispersion and molecular diffusion”, Journal of Hydrology Vol. 549, pp. 452-460.
 37. Christian Wolfsteiner, Louis J. Durlofsky and Khalid Aziz, 2003, “Calculation of well index for nonconventional wells on arbitrary grids”, Computational Geosciences Vol.7, pp. 61-82.
 38. M.TH. VAN GENUCHTEN, 1980, “A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils”, Soil Science Society of America Journal Vol. 44 Issue 5, pp. 892-898.
 39. Hung-Kl Yen; Ni-Bin Chang, M. ASCE and Tsair-Fuh Lin, 2003, “Bioslurping Model for Assessing Light Hydrocarbon Recovery in Contaminated Unconfined Aquifer.I: Simulation Analysis”, Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management Vol. 7 Issue 2.
 40. X. S. Qin, G. H. Huang, G. M. Zeng, and A. Chakma, 2008, “Simulation-based optimization of dual-phase vacuum extraction to remove nonaqueous phase liquids in subsurface”, Water Resources Research, VOL. 44 Issue 4.
 41. Marsland, S., 2015, “Machine Learning – An Algorithmic Perspective”, CRC Press.
 42. Yan Shi, Wenqi Zhong, Xi Chen, A.B. Yu, Jie Li, 2019, “Combustion optimization of ultrasupercritical boiler based on artificial intelligence”, Energy Vol. 170, pp. 804-817.