

應用 Geo-AI 模型推估苗栗地區 NO_2 濃度及影響因子貢獻解析

陳潔瑩¹ 賴忻宜¹ 曾于庭² 吳治達^{3,4,5,6,7*}

摘要

本研究以苗栗縣為示範區，整合氣象、地面監測、衛星遙測、土地利用與通霄電廠排放等多源資料，建構逐時、 $50\text{ m} \times 50\text{ m}$ 解析度之 NO_2 濃度推估模型。採用 XGBoost 演算法 SHAP 遞增篩選機制，選取一小時滯後，模型 R^2 為 0.80，RMSE 為 1.78 ppb。SHAP 分析顯示 NO_x 與道路密度為最主要驅動因子，台電固定源亦具區域影響力。多層次驗證證實模型對時間、空間與高污染情境皆具穩健表現。研究展現 Geo-AI 於中小型縣市空品推估之應用潛力，亦為風險預警與污染治理提供量化依據。

關鍵詞：二氧化氮、地理人工智慧、機器學習、空間推估圖

1. 前言

空氣污染已被世界衛生組織列為全球最嚴峻的環境健康挑戰之一，而 NO_2 更是同時兼具高社會關注度與高度健康風險的指標性污染物。 NO_2 主要產生於高溫燃燒過程，與化石燃料使用密切相關 (Vardoulakis *et al.*, 2020、Achakulwisut *et al.*, 2019)。在戶外環境中，道路交通可貢獻高達 80 % 的都市 NO_2 濃度，燃煤或燃氣電力設施則構成重要背景來源；室內方面，燃氣爐灶、熱水器與暖爐等燃燒器具同樣能顯著升高 NO_2 水平 (Tiotiu *et al.*, 2020)，而戶外濃度也會透過自然通風滲入室內。

NO_2 已被確認為呼吸道刺激物，可引發急性咳嗽、喘鳴與支氣管痙攣，長期暴露則與全因死亡率與兒童氣喘顯著相關 (Qian *et al.*, 2021、Kashtan *et al.*, 2024)。全球疾病負擔研究估計，每年約有 400 萬例兒童新發氣喘可歸因於 NO_2 ，佔總病例數的 13 %，其中 92 % 發生於年均 NO_2 低於世界衛生組織 (World Health Organization, WHO) 指南值 21 ppb 的

地區，突顯「低濃度亦具健康風險」的警訊 (Achakulwisut *et al.*, 2019)。長期暴露於較高 NO_2 濃度還會導致呼氣一氧化氮 (FeNO) 上升，並使第一秒用力呼氣量 (forced expiratory volume in one second, FEV1) 和用力肺活量 (forced vital capacity, FVC) 下降。此外，孕期高暴露亦與嬰幼兒肺功能不良相關 (Tiotiu *et al.*, 2020、Vardoulakis *et al.*, 2020)。因此，降低 NO_2 暴露已成為交通與都市環境管理的優先議題。

面對都市化、機動車輛與多元燃燒源交織的排放格局，如何在不同時空尺度上精準推估 NO_2 濃度並量化各影響因子貢獻，已是公共衛生與環境決策的關鍵課題。近年來，地理人工智慧 (Geo-AI) 結合遙測影像、地面監測、氣象與土地利用資料，為高解析度空氣品質推估提供強大工具；然而，現有研究大多侷限於靜態特徵，較少探討排放與氣象交互作用的時間延遲 (lag effect)，亦缺乏系統化方法來解析各變數對預測結果的相對影響。

國際上，結合衛星觀測與機器學習以推估近地

收到日期：民國 114 年 07 月 11 日
修改日期：民國 114 年 09 月 03 日
接受日期：民國 114 年 10 月 15 日

¹ 國立成功大學測量及空間資訊學系 碩士生

² 國立成功大學測量及空間資訊學系 助理研究員

³ 國立成功大學測量及空間資訊學系 教授

⁴ 國家衛生研究院國家環境醫學研究所 兼任研究員

⁵ 國立中興大學永續農業創新發展中心 兼任研究員

⁶ 高雄醫學大學精準環境醫學研究中心 兼任研究員

⁷ 長庚科技大學慢性疾病暨健康促進研究中心 合聘研究員

* 通訊作者, E-mail: chidawu@mail.ncku.edu.tw

表 NO_2 已成主流。近年大型研究多以集成學習整合多源資料 (如衛星柱量、化學傳輸模型 (CTM)、地理與氣象)，在國家尺度以約 1 km、逐日產出地表 NO_2 ；例如美國全境研究以神經網路、隨機森林與梯度提升等模型集成，達到整體 $R^2=0.788$ (Di *et al.*, 2020)。在中國，全國尺度的三階段集成模型整合 OMI/TROPOMI、AOD、CTM 與高解析地理變數，於 $1 \times 1 \text{ km}$ 、逐日期間取得整體 $R^2=0.72$ 、空間 $R^2=0.85$ (Huang *et al.*, 2022)。亦有區域尺度研究以 TROPOMI 為核心，結合「氣象滯後項」與「時空項」的集成模型，10 折 $R^2=0.89$ 、時間與空間 R^2 分別為 0.71 與 0.81，顯示其在擷取短時變化與非線性耦合關係方面具體效益 (He *et al.*, 2022)。

在單一模型方面，多項研究證實，僅以單一樹系或增強式學習亦能達到穩健的 NO_2 推估表現。日本大都市區以隨機森林為核心的時空 LURF 架構，在留一站交叉驗證下整體優於傳統線性模型，顯示單一演算法能有效擷取土地利用、排放源距離遞減與氣象要素之非線性關係 (Araki *et al.*, 2018)。於東亞尺度、以 TROPOMI 為主的研究系統比較 SVR、RF、XGBoost、LightGBM 等單一模型建立近地表 NO_2 與 O_3 的高空間解析度估測，結果顯示梯度提升法在多種驗證方案下具良好泛化表現，並可藉由特徵重要度解析衛星柱量與地表濃度的關聯 (Kang *et al.*, 2021)。在都會區交通主導情境中，採用隨機森林之「分段 (partition)」單一模型更能處理高、低濃度區間的結構性差異，提升 NO_2 預測精度，較傳統單一 RF 具有顯著增益 (Kamińska, 2019)。此外，在感測器校正應用上，單一隨機森林可顯著提升低成本 NO_2 感測器相對監測站的準確度與穩定性，支持其於實務部署的可行性 (Zimmerman *et al.*, 2018)。另有泛歐洲尺度之比較研究指出，多種單一機器學習法 (含樹系與增強式演算法)在年均 NO_2/PM 的空間建模上優於基礎線性模型 (Chen *et al.*, 2019)，但具有地理加權的線性模型 (GWR/GTWR)則展現出更佳或並列最佳的性能 (Shen *et al.*, 2022)，進一步凸顯單一非線性模型於廣域暴露評估的效能基準價值。

在單一學習的比較中，XGBoost 多次呈現 NO_2

的最佳或並列最佳預測力。以東亞 TROPOMI 資料同時比較 SVR、RF、XGBR、LGBMR 的研究顯示：在隨機/空間/時間三種交叉驗證下，樹系方法整體優於 SVR，而樹系中以「提升式」的 XGBR/LGBMR 表現最佳；其中針對 NO_2 ，XGBR 略勝並被選為後續最佳單一模型 ($R^2 \approx 0.63\text{--}0.70$) (Kang *et al.*, 2021)。在短時序列預測 (如 1、8、12、24 小時同步預測) 與低成本感測器應用的情境下，研究指出 M5P 模型樹憑藉其樹狀結構效率與較強的泛化能力，在多變量建模中整體表現最佳，優於 SVM 與 ANN；相對地，ANN 在資料量相對較小且特徵多樣時較易過度擬合、表現最差。同時，納入氣象條件 (如濕度、風速) 與共存污染物 (如 SO_2) 的多變量建模可顯著提升 NO_2 的預測精度 (Shaban *et al.*, 2016)。在空間建模領域，XGBR 的優勢仍然穩固。在中國東南沿海的區域尺度研究中，於相同資料與特徵配置下，三種單一樹系模型之表現排序一致為 XGBR > EXT (極端隨機樹 > RF (10 折交叉驗證 R^2 最高達 0.88))，再次支持 XGBR 為 NO_2 單一演算法的領先者 (He *et al.*, 2022)。此外，在全中國 $1 \times 1 \text{ km}$ 的長期集成研究中，雖三個單一模型 (RF/ERT/XGB) 整體 R^2 相近 (≈ 0.71)，但變數重要度分析顯示 XGBR 對 TROPOMI NO_2 訊號的依賴度最高 (相對重要度約 43.9%)，突顯 XGBR 對衛星柱量訊號的敏感度與表徵力 (Huang *et al.*, 2022)。綜合而言，在特徵與驗證設計同質條件下，XGBR 通常能在 NO_2 單一模型中取得最佳或並列最佳的準確度，且可配合 SHapley Additive exPlanations (SHAP) 等方法提供穩健的變數貢獻解讀 (Kang *et al.*, 2021、Hsu *et al.*, 2024)，適合作為本研究的單模基準與集成關鍵子模型。

跨國與中尺度研究普遍以「多源資料 + 多演算法」的集成框架顯著優於單一模型。美國全境日尺度 1 km 的研究以 NN、RF、GB 並結合 CTM/遙測等特徵建構集成，達成整體 10 折交叉驗證 $R^2=0.79$ 、空間 $R^2=0.84$ 、時間 $R^2=0.73$ ，並提供位址層級不確定度估計，適用於長短期健康研究 (Di *et al.*, 2020)。中國東南沿海的 TROPOMI 主導日尺度集成模型，納入「氣象滯後項」與「時空項」，10 折交叉驗證 $R^2=0.89$ 、時間/空間 $R^2=0.71/0.81$ ，顯示延

遲效應與相對位置編碼可顯著強化時空推論 (He *et al.*, 2022)。在中國全國 1×1 km、2013–2019 年的集成模型中，整合 OMI/TROPOMI、AOD、CTM 與高解析地理變數與三種樹系演算法，總體 $R^2=0.72$ 、空間 $R^2=0.85$ ，並於都會區進一步降維至 100 m，以捕捉道路鄰近的細尺度暴露梯度 (Huang *et al.*, 2022)。混合統計與機器學習的技術亦日益成熟：南加州以「一般化加性混合模型 + 集成學習 + 約束優化」建立 22 年雙週 NO₂/NO_x 序列，NO₂ 的集成 $R^2=0.85$ ，顯著優於初始單模型，且時間序列更平滑穩定 (Li *et al.*, 2017)。在台灣，本土 Geo-AI 已將集成混合空間模型 (EMSM) 推展至 50 m 超高解析度與分時段建模；針對 NO₂ 的日/夜兩時段模型，調整後 R^2 分別達 0.92/0.93，並驗證夜間與冬季濃度偏高之時段與季節型態，彰顯在地化集成模型對健康風險評估與政策應用的即時價值 (Babaan *et al.*, 2024)。相同 EMSM 策略也用於台灣通勤時段 O₃ 評估，在多種內外部驗證 (含極值、降雨/非降雨、平假日) 下仍維持高穩健性 (調整後 $R^2 \approx 0.91$)，提供高時空解析度的暴露資訊以支援決策 (Hsu *et al.*, 2024)。

綜合而言，上述研究多在國家或大區域尺度產製日尺度網格；常用方法包含隨機森林、XGBoost、集成學習與卷積式神經網路等，證實多源資料融合對地表 NO₂ 估測之關鍵性 (He *et al.*, 2022、Huang *et al.*, 2022)。在地面監測導向的路徑中，土地利用與氣象特徵結合機器學習可有效解釋 NO₂ 的空間異質性；其中集成與混合建模能在高度非線性、多來源特徵的情境下提升穩健性，並支援後續的變數貢獻解析 (Li *et al.*, 2017、Di *et al.*, 2020)。台灣近年亦有 Geo-AI 於 50 m 生成 NO₂ 細網格並提供日/夜成果：如 Hybrid Kriging-LUR 結合 XGBoost 於 50 m 網格估計逐日 NO₂，解釋力可達 $R^2 \approx 0.91$ ，並經外部、季節與城市分層等多重驗證證實模型穩健性 (Wong *et al.*, 2021)；進一步以 EMSM 建置日間/夜間 NO₂，於 50 m 解析度下調整後 $R^2=0.92/0.93$ ，呈現夜間與冬季濃度偏高之型態，顯示在本土資料條件下可達實務所需的時空解析度與準確度 (Babaan *et al.*, 2024)。

相較於多數聚焦「日尺度、 ≥ 1 km 空間解析度」

的研究，本研究於苗栗地區建置「逐時、50 m × 50 m」之高解析度 NO₂ 模型，並引入 1–12 小時滯後資料庫以辨識短時延遲效應。流行病學證據顯示，污染物對健康影響常呈現「當日與滯後」的累積效應：以野火相關 PM_{2.5} 的全球時間序列研究為例，作者同時評估當日與單日滯後效應 (至 7 天)，並以 0–2 天移動平均為主要指標，且於敏感度分析中將最大滯後延長至 10 天以檢驗穩健性；結果顯示短期暴露對死亡風險的影響主要集中於前 0–2 天，而延長滯後窗可更全面地捕捉短期健康衝擊之輪廓 (Chen *et al.*, 2021)。基於此一證據鏈，本研究在 NO₂ 的短時暴露推估中納入多小時滯後資料庫，以提升風險評估之精確度與完整性。同時，結合 SHAP 量化氣象、土地利用與遙測等變數的相對影響 (Babaan *et al.*, 2024、He *et al.*, 2022、Huang *et al.*, 2022、Li *et al.*, 2017、Di *et al.*, 2020)，以兼顧模型之可解釋性與準確度。綜上，本研究旨在提升 NO₂ 推估的解釋力與精度，並為健康風險評估、污染管制與即時預警提供科學依據。

2. 研究材料

2.1 研究區域

本研究以臺灣中部西側的苗栗縣為對象，如圖 1 所示。苗栗縣幅員約 1820 km²，下轄 18 個鄉鎮縣市 (內政部戶政司，2023、內政部國土測繪中心，2023)，地形由西側濱海平原向東漸次過渡至丘陵與中低海拔山地，形塑出複雜的氣流場與污染傳輸路徑。縣內設有 11 座空氣品質監測站，其中 3 座為環境部常規測站，8 座為台電背景監測站；監測點位涵蓋沿海、平原與內陸谷地等多種地貌，有助於解析污染物在不同地形條件下的時空分布。

苗栗縣坐擁竹南、頭份、銅鑼等 6 座工業區，並鄰接大型天然氣複循環機組的通霄發電廠。相較於既有文獻多聚焦六都，苗栗縣雖屬非直轄市，仍受工業活動與能源設施影響，具備潛在空氣污染風險，卻常被忽略。本研究選擇苗栗作為樣區，旨在補足中小型縣市空品研究的缺口，驗證 Geo-AI 模型在複雜地形與多重排放源情境下的適用性。

本研究的資料期間設定為 2022 – 2023 年，以避開 COVID-19 防疫措施對交通流量與排放行為所造成非常態擾動，確保基準年的排放與氣象條件具有代表性。許多研究證實，疫情期間的封城與行動限制使都會區 NO_2 在短期內下降約 10 – 60%，並常伴隨 O_3 上升；因此，本研究排除 COVID-19 衝擊年份，以維持模型在常態排放與氣象條件下的代表性與穩健性 (Cole *et al.*, 2020、Wang *et al.*, 2020、Petetin *et al.*, 2020、Keller *et al.*, 2021、Shi *et al.*, 2021、Yang *et al.*, 2021)。

2.2 資料庫介紹

本研究所使用之變數彙整於表 1，類別包含空氣汙染物、氣象因子、土地利用變數、衛星資料、台電因子與其他虛擬變數，以下逐一說明各類別資料。

2.2.1 空氣汙染物與氣象因子

本研究的空氣污染物資料取自環境部開放平臺，涵蓋 $\text{PM}_{2.5}$ 、 NO_2 、 O_3 、 NO_x 、 PM_{10} 與 SO_2 六項指標。所有測站逐時觀測值先依 24 小時分段整理，再分別以空間克利金法 (Kriging Interpolation) 內插至 50 m × 50 m 格網，產製高解析度濃度圖層。研究的依變數為 NO_2 ，其餘五項污染物作為自變數；

清理後共取得 164759 筆有效紀錄，整體缺漏率 14.5 %。

氣象資料則來自大氣科學研究與應用資料庫，處理流程與污染物相同：逐時觀測值經克利金法內插至 50 m × 50 m，生成與污染物圖層對應的高解析度網格。無降雨時雨量設定為 0，風向以箭矢標示方向與格點位置。

2.2.2 土地利用變數

本研究的土地利用資料同時涵蓋面狀、線狀與點狀三類空間要素。面狀部分包括農業、森林、交通、水利、建築、公共、遊憩與礦鹽等八大土地利用類型，並額外整合「工業區」圖層，將整合加工出口區、科學園區、報編工業區、農業生技園區與一般編定工業區統一編碼，以反映區域產業活動強度。線狀資料則採用路網數值圖，區分主要道路、當地道路與其他道路，作為交通排放與通風走廊的重要指標。點狀資料涵蓋各類商業與服務業據點，以及寺廟位置；後者因焚香與燒金紙為傳統文化排放來源，已被證實與 NO_2 濃度呈正相關 (Babaan *et al.*, 2024)，故一併納入以完善空間解釋變數。圖 2 彙整上述要素：工業區以紅色面填色，道路以灰色線條表示，通霄發電廠則以符號點標示其位置。

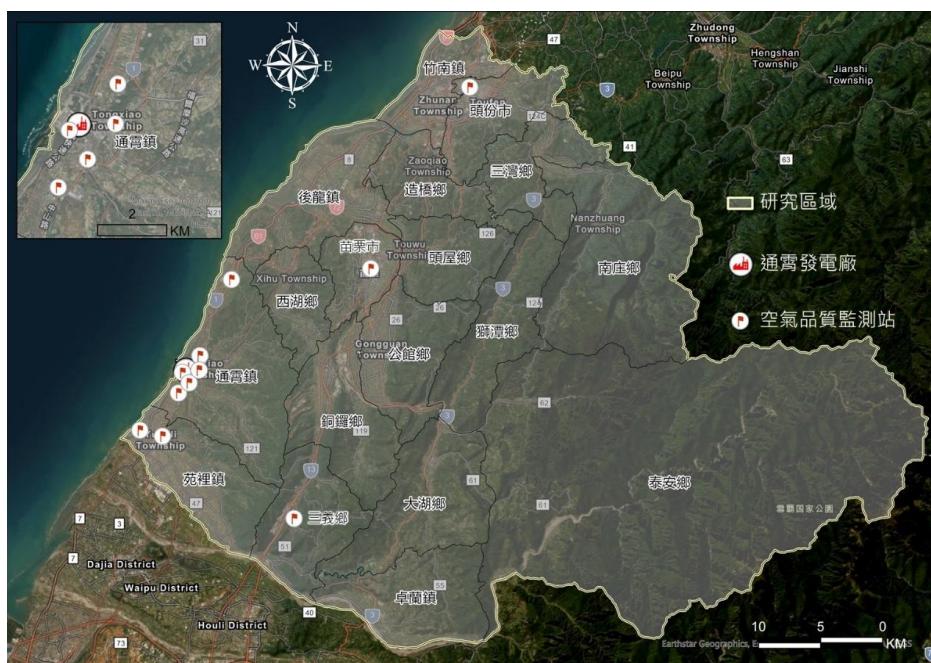


圖 1 研究試區

表 1 本研究所用之變數清單

資料分類	資料來源	內容	時間解析度	空間解析度
空氣汙染物	環境部環境資料開放平臺	PM _{2.5} ($\mu\text{g}/\text{m}^3$)、SO ₂ (ppb)、NO _x (ppb)、PM ₁₀ ($\mu\text{g}/\text{m}^3$)、O ₃ (ppb)、NO ₂ (ppb)	每小時	50 m × 50 m
氣象因子	大氣科學研究與應用資料庫	平均溫度(°C)、平均相對溼度(%)、平均大氣壓力(hPa)、平均風速(m/s)、總降雨量(mm)	每小時	50 m × 50 m
土地利用變數	內政部國土測繪中心國土利用調查資料庫	森林、農地、水體、住宅區、商業區、工業區、工業園區、寺廟、火化場、焚化爐、道路等		50 m × 50 m
	MODIS 全球常態化 NDVI 資料庫 (MOD13Q1)第 6.1 版	NDVI	每月	250 m × 250 m
衛星資料	歐洲太空局(ESA)的哨兵 -5P 衛星 (Sentinel-5 Precursor)	對流層二氧化氮垂直柱濃度	每日	5.5 km × 3.5 km
	歐洲中期天氣預報中心(ECMWF)的第五代再分析資料	ERA5 行星邊界層 Boundary Layer Height (BLH)	每小時	0.25° × 0.25°
台電因子	台灣電力公司資料開放平台	連續排放監測系統(CEMS)、發電量、污染物排放量	每小時	50 m × 50 m
其他	虛擬變數(節慶因子)	新年、元宵節、清明節、端午節、中元節、中秋節		

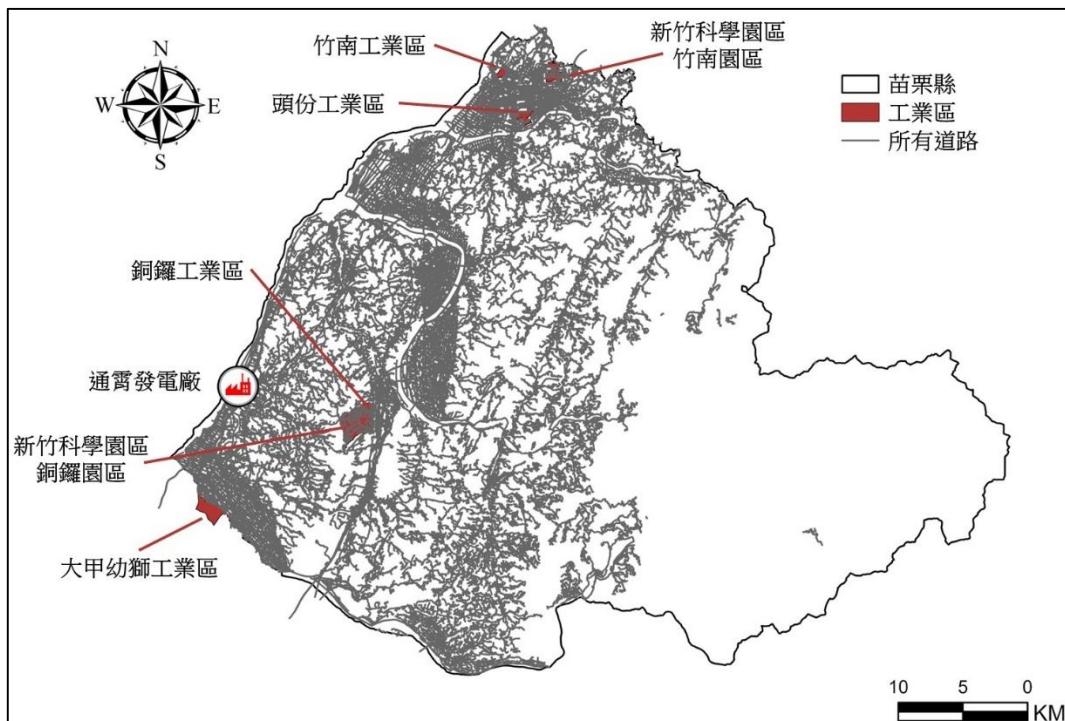


圖 2 研究區域內「點、線、面」示範圖層

2.2.3 衛星資料

本研究選用的衛星遙測變數可分為三類。第一類為綠覆度指標，採用 NASA MODIS 250 m × 250

m 解析度的常態化差值植生指數 (Normalized Difference Vegetation Index, NDVI)；該產品每月提供兩景完整覆蓋臺灣的影像，透過紅光與近紅外光反射率差異評估植被生長程度。第二類為大氣污染

指標，取用歐洲太空總署 Sentinel-5P (TROPOMI) 每日資料，擷取對流層二氧化氮柱濃度 (mol/m^2)，空間解析度為 $5.5 \text{ km} \times 3.5 \text{ km}$ ，以描繪境外輸送與整體氣膠負荷。第三類為邊界層結構參數，使用 ECMWF ERA5 再分析的行星邊界層(m)，解析度 $0.25^\circ \times 0.25^\circ$ ，時間尺度為逐時，可捕捉大氣混合作用隨時間之變化。

2.2.4 台電因子

電力相關資料由臺灣電力公司提供，內容包括：通霄電廠廠址與六座機組煙囪的精確坐標、研究期間各機組逐時的發電量與 NO_x 排放量，以及連續排放監測系統 (CEMS) 追蹤的 NO_x 排放紀錄。

2.2.5 其他

為反映傳統節慶活動帶來的短期排放衝擊，本研究參考 Hsu *et al.* (2024) 的作法，將農曆新年、元宵、清明、端午、中元與中秋六大節日設計為二元變數：節日當日標記為 1，其餘時間為 0，並納入模型。相關研究指出，焚香、燒金紙與燃放爆竹等習俗會在節慶期間造成顆粒物與氣體污染物瞬間增加，進而影響區域空氣品質。表 2 詳列六大節慶的具體日期，涵蓋節日當天及其相應的國定連假時段。

表 2 虛擬變數之日期

節慶	2022 年	2023 年
春節	1/29~2/6	1/20~1/29
元宵節	2/15	2/5
清明節	4/2~4/4	4/1~4/5
端午節	6/3~6/5	6/22~6/25
中元節	8/12~8/14	8/30
中秋節	9/9~9/11	9/29~10/1

3. 研究方法

本研究方法流程如圖 3 所示，可分為五大步驟：(1)資料庫建置；(2)時間延遲效應評估；(3)模型建構與變數篩選；(4)分層驗證；(5)變數影響度分析與高解析度 NO_2 空間推估。以下分述各環節內容。



圖 3 研究流程圖

3.1 資料庫建置

首先彙整六大資料來源：衛星、土地利用、氣象、空氣污染物、節慶活動與台電因子。所有資料經投影統一至 TWD97 / TM2 座標系，時間戳調整為臺灣標準時間並依逐時尺度對齊。各變數重採樣至 $50 \text{ m} \times 50 \text{ m}$ 格網後，剔除缺漏值與異常值，最終形成涵蓋 2022–2023 年 164759 筆時空樣本、逾 500 個候選變數的整合資料庫。

為確保不同資料源在同一空間與時間尺度下運作，所有原始變數先完成座標轉換、單位統一與標準化。土地利用相關變數進一步計算 14 種環境範圍 (50、150、250、500、750、1000、1250、1500、1750、2000、2500、3000、4000、5000 m)：

- (1) 對面狀與線狀資料 (如工業區、路網) 計算「面積/長度密度」，反映排放源集中度；
- (2) 對點狀資料 (如寺廟、商服據點) 計算「點密度」；
- (3) 同時求取每一格網到最近同類排放源的直線距離。

衛星資料 (NDVI、對流層二氧化氮柱濃度、行星邊界層) 則先重新取樣至 50 m × 50 m，確保解析度與格網一致。完成上述處理後，各類變數即可在統一的時空框架下供 XGBoost 模型訓練使用。

3.2 時間延遲效應評估

為捕捉 NO₂ 濃度對短期排放與氣象波動的「短尾」滯後反應，本研究先為所有隨時間劇烈變動的自變數 (例如氣象參數、機組排放量等) 建立 1–12 小時的滯後版本。對每一滯後期 (lag 1–lag 12)，計算這些變數與 NO₂ 濃度的皮爾森相關係數，並取其絕對值後的平均值，作為該滯後期的整體關聯強度指標。

皮爾森相關係數適用於衡量變數間的線性關聯，雖然其嚴格假設變數近似常態分布，但在大樣本下對偏態分布具有相當穩健性。本研究僅將相關係數用作篩選延遲效應的相對比較依據，而非進行顯著性檢定，因此即使部分變數分布未完全符合常態，仍可合理應用。最終選取平均相關性最高的單一滯後期，並僅納入該視窗下的變數進入後續 XGBoost 建模，以保留具代表性的延遲訊號，同時降低資料冗餘與模型複雜度。先前研究亦指出，在逐時或短時尺度之污染預測中納入短期滯後特徵，可更有效表徵排放與氣象的交互延遲影響並提升模型表現 (Wei *et al.*, 2021; He *et al.*, 2022)。

3.3 模型建構與變數篩選

本研究以 NO₂ 濃度作為目標變數，將其餘時空因子納為解釋變數，建立梯度提升決策樹 (XGBoost) 模型。之所以採用 SHAP，是因其 (1) 基於合作賽局理論，具一致性與可加總性，能提供穩健、可比較的特徵貢獻；(2) 同時給出「全域」與「個別預測」(局部)的解釋，有利於釐清不同時段、氣象條件與空間位置下之機制；(3) 對樹系模型可用

TreeSHAP 快速計算，適合高維、多滯後與大量資料；(4) 可輔助偏差診斷與政策溝通，提升模型的可稽核性與可用性 (García & Aznarte, 2020)。

變數挑選採用 SHAP 值遞增策略，流程如下：首先以完整模型計算各變數之 SHAP 全域重要度並依數值大小排序；接著按此順序將變數逐一加入模型並重新訓練；若新增變數使模型解釋力 (R²) 提升至少 0.01，則視為關鍵變數並予以保留。透過此遞增篩選，可在保留關鍵資訊的同時排除冗餘特徵，降低過度擬合風險並優化模型之可解釋性與穩健性。

3.4 分層驗證

為評估模型的穩健度與泛化能力，本研究規劃四套互補的驗證機制。首先，將全部資料隨機分割出 20% 作為獨立測試集，用以檢視基本預測誤差；其次，從年、月、日、時四個時間尺度比對預測與觀測值，以驗證時間一致性；第三，針對 11 座空品測站評估空間殘差分布，並探討不同區域下的敏感度；最後，結合十折交叉驗證以降低抽樣偏差，並選取濃度前 10%、20%、25% 的樣本進行極端濃度測試，檢視模型在高污染情境中的表現。

3.5 變數影響度分析與高解析度 NO₂ 空間推估

在確定模型穩健可靠後，本研究以 SHAP 值評估各變數對 NO₂ 預測的相對貢獻，並將數值轉換為百分比，以凸顯主要驅動因子。隨後擇定研究期間 NO₂ 觀測值最高的日期，將最終模型套用至完整的 50 m × 50 m 網格，逐時推估當日 NO₂ 濃度分布，生成高解析度時空圖，以輔助後續熱點檢測與風險評估。

4. 結果

4.1 NO₂ 敘述統計

表 3 彙整研究期間 NO₂ 的敘述統計結果。整體而言，NO₂ 平均濃度為 7.04 ± 3.95 ppb，範圍介於 0.0–49.6 ppb，顯示本研究區域 NO₂ 濃度在不同時段

具有相當變動性。進一步比較年度差異，2022 年平均濃度為 $7.37 \pm 3.97 \text{ ppb}$ ，整體高於 2023 年的 $6.72 \pm 3.90 \text{ ppb}$ 。四分位數亦呈現一致趨勢：2022 年的第一、第二及第三四分位數 ($4.4, 6.8, 9.7 \text{ ppb}$) 均略高於 2023 年 ($3.8, 6.0, 8.9 \text{ ppb}$)。這反映 2022 年在整體與分布中值上皆普遍高於 2023 年，可能與當年度氣象條件、排放量變化或區域背景污染有關。最大值方面，兩年度分別落在 47.2 與 49.6 ppb ，顯示偶發性高濃度事件在不同年份皆存在。

4.2 時間延遲選擇

圖 4 顯示 NO_2 與各自變數在 1–12 小時滯後 (lag1-lag12) 下之絕對平均皮爾森相關係數。可見相關性在 lag1 達到最高峰 (≈ 0.12 ，紅點標示)，隨滯後時間遞增而快速衰減，至 lag7 降至最低 (≈ 0.075)，其後略有回升但始終未超過 0.085 。此趨勢印證 NO_2 對短期排放與氣象變動的「短尾型」延遲特性：暴露後 1 小時即展現最大聯動，6–8 小時後影響顯著減弱。

根據絕對平均相關係數最高原則，本研究選擇 lag1 作為所有高變動變數的代表滯後期，並將該時間窗下的資料納入後續 XGBoost 模型，以捕捉最具解釋力的延遲訊號，同時避免引入冗餘滯後變數造成模型複雜度上升。

4.3 模型表現

表 4 彙整 XGBoost 模型在三種評估情境下的績效指標：訓練集 (80%)、獨立測試集 (20%)、以及十折交叉驗證 (10-Fold Cross Validation)。訓練階段模型可解釋 80% 的 NO_2 變異 ($R^2 = 0.80$)，均方根誤差 (Root Mean Square Error, RMSE) 為 1.78 ppb ，顯示在已見過的資料上具備良好擬合度。當模型應用到未參與建模的外部測試集時， R^2 下降至 0.69，RMSE 上升至 2.23 ppb ，仍維持相當水準的預測能力，表示模型泛化性尚佳但略有高估趨勢。十折交叉驗證所得 R^2 (0.68) 與測試集幾乎一致，RMSE 為 2.22 ppb ，進一步證實抽樣方式不同時模型的穩定性。整體而言，XGBoost 能在多時空變數環境中有效捕捉 NO_2 濃度分布，且訓練、測試與交叉驗證結果差異不大 (R^2 僅差 0.12)，僅出現適度的擬合落差，反映所採 SHAP 遞增篩選有助抑制過度擬合並保留關鍵訊息。

圖 5 進一步以散佈圖呈現 NO_2 的觀測值與模型預測值之間的對應關係。點位大致分布於 45° 參考線附近，顯示模型預測結果與實際觀測值整體趨勢一致。圖中亦標示 $R^2 = 0.80$ 與 $\text{RMSE} = 1.78 \text{ ppb}$ ，與表 3 中訓練集的數值相符，直觀展現模型在資料擬合上的表現。此結果不僅呼應統計指標，也提供視覺化證據支持模型具備合理的解釋力與預測力。

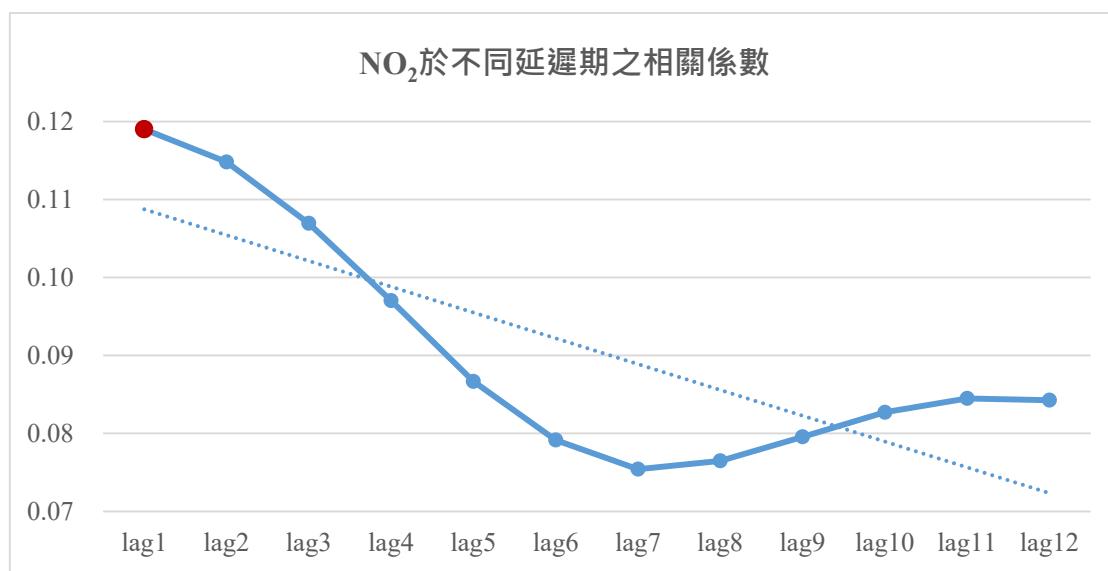


圖 4 lag1 至 lag12 之絕對平均相關係數折線圖

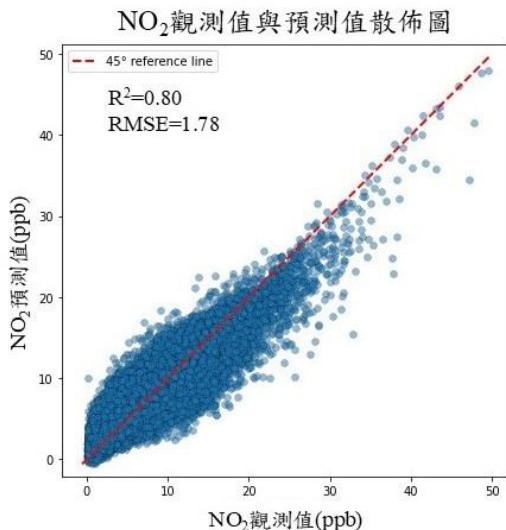
表 3 研究期間 NO₂ 敘述統計之結果

	平均值±標準差	最小值	第一四分位數	第二四分位數	第三四分位數	最大值
整體	7.04±3.95	0.0	4.1	6.4	9.3	49.6
2022 年	7.37±3.97	0.0	4.4	6.8	9.7	47.2
2023 年	6.72±3.90	0.1	3.8	6.0	8.9	49.6

註：單位均為 ppb，四捨五入至小數第一位。

表 4 XGBoost 模型之訓練集、測試集與十折交叉驗證表現

	R²	Adjusted R²	RMSE	MSE	MAE
Training	0.80	0.80	1.78	3.15	1.01
Testing	0.69	0.69	2.23	4.99	1.22
10-Fold CV	0.68	0.68	2.22	4.92	1.24

圖 5 NO₂ 原始觀測值與最終模型之預測值散佈圖

4.4 變數篩選結果與重要性比較

圖 6 以 SHAP 分析呈現最終模型的關鍵變數：在圖 6(a)中，NO_x 的 SHAP 分布最寬且均位於正值區域，顯示其濃度越高，對 NO₂ 預測的推升效應越大；所有道路環域 1000 m 位居其次，驗證交通排放為區域背景 NO₂ 的主要來源。氣溫、相對濕度與風速等氣象因子，則透過抑制或促進大氣混合作用調節 NO₂；其中高溫與低濕條件傾向升高濃度，而高風速則顯著降低濃度。

臭氧 (O₃) 則反映了 NO-NO₂-O₃ 光化學平衡，其與 NO₂ 既可能呈現正相關 (透過 O₃ 氧化 NO 生成 NO₂)，也可能因交通排放下的 O₃ 消耗作用而呈現負相關 (Crutzen, 1970)，因此在模型中成為重要的解釋變數。

植被指標 NDVI 最大值環域 5000 m 與 NO₂ 呈負相關，支持綠覆度具有稀釋與沉降作用。柱狀圖

圖 6 (b) 將前十大變數的平均 SHAP 值量化，可見 NO_x (+1.52) 遠高於其他因子，凸顯同源污染在短尺度內的直接影響，而交通、氣象與植被變數分別扮演次要但不可或缺的調節角色。

進一步將所有變數依屬性歸納 (表 5)，可發現空氣汙染物類別占總貢獻的約 42%，其中 NO_x 與 O₃ 為最具代表性的化學驅動因子；氣象因子與土地利用變數分別貢獻約 24% 與 22%，顯示大氣擴散條件及交通/綠地格局共同主宰區域 NO₂ 分布；台電因子 (9%) 則反映固定源排放的區域影響力；最後，境外污染指標僅佔 3%，顯示在 2022–2023 年的苗栗背景條件下，對流層 NO₂ 柱濃度對地面濃度的直接解釋力有限。整體而言，交通相關排放 (NO_x、道路) 加上氣象穩定度 (氣溫、風速、相對濕度、行星邊界層) 構成影響苗栗地區 NO₂ 的核心組合，而植被覆蓋與固定源距離則提供額外的空間異質性訊息，使模型能夠細緻捕捉高低濃度帶。

4.5 分層驗證

圖 7 與表 6 整合模型在多層次驗證中的指標。整體而言，模型在不同時空層級與極端濃度情境皆維持良好且一致的預測能力，細部結果如下。

- (1) 時間驗證：逐日驗證的 R² 介於 0.76–0.82，對應 RMSE 1.62–1.88 ppb；逐月 R² 落在 0.71–0.88，呈現春、冬兩季較高，夏、秋稍低的季節律。兩年平均 R² 均為 0.80，RMSE 分別為 1.80 ppb (2022 年) 與 1.75 ppb (2023 年)，顯示模型能穩定捕捉長週期變化。逐時尺度 R² 為 0.74–0.83，RMSE 1.44–2.01 ppb，說明模型對短時波動亦具備充足敏感度。

- (2) 極端值驗證：當樣本限制在 NO_2 濃度前 25%、20% 與 10% 時， R^2 分別為 0.61、0.61、0.65；雖較全樣本略低，但各組 RMSE 皆小於 2 ppb，證實即便在高污染事件期間模型仍具可接受的預測精度。
- (3) 空間驗證（表 6）：11 座監測站 R^2 分佈於 0.66–0.85。其中三義（0.66）與苑裡（0.71）因位於丘陵轉折與背風側，預測較為保守；內島里與海岸里均為 0.74，其餘測站皆 ≥ 0.75 。站點 RMSE 介於 1.54–2.13 ppb，與全區平均 RMSE 相近，顯示空間誤差無明顯集聚。

綜合上述，模型在年、月、日、時四層時間尺度的 R^2 差距不超過 0.2，RMSE 差異小於 0.6 ppb；空間與極端情境驗證亦呈現穩定表現，證明 SHAP 遞增篩選與滯後優選策略有效抑制過度擬合、保留主要物理與排放訊號，使模型具備良好的預測能力與實務應用潛力。

4.6 NO_2 時空推估圖

圖 8 至圖 11 連續呈現 2023 年 2 月 1 日

0–23 時的 NO_2 推估圖，是研究期間原始觀測值最高的高污染事件日。當日 NO_2 的高濃度始終鎖定在竹南與頭份一帶，即便午間海風與高邊界層將全區大部分濃度壓至 10–20 ppb，該區仍維持 25–40 ppb 的局部熱點；傍晚邊界層收縮後，熱點更迅速堆升至 40 ppb 以上。若與圖 2 的土地利用示意相互參照，高值中心正好覆蓋竹南科學園區、竹南工業區與頭份工業區等複合產業聚落，同時也是全縣道路最密集的交通走廊。圖中灰色線網標示的主要與次要道路在竹南–頭份交會成輻射狀，連接國道一號、省道台 1 線及多條縣鄉道，使車流量遠高於其他鄉鎮。工業排放與高密度車流疊加，再加上地形開闊、海風午後轉內陸的條件，形成固定源與移動源共同推升的區域性污染。因此，高值區的空間位置並非孤立的氣象現象，而是工業區與道路網絡共同主導的結果；這也印證 SHAP 分析中「 NO_x 濃度」與「道路環域 1000m」分別排名第一、第二的重要性，兩者在竹南–頭份的高度耦合，解釋了該區日內濃度始終居高不下的原因。

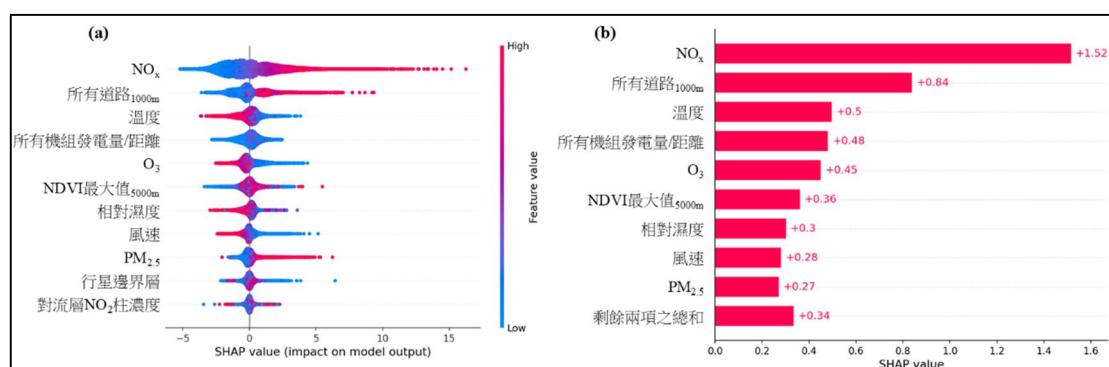


圖 6 變數 SHAP 值及作用方向：(a) 逐筆 SHAP 分布；(b) 全域平均 SHAP 值柱狀圖

表 5 變數重要性百分比

類別	百分比	變數	百分比
空氣汙染物	41.9%	NO_x	28.4%
		O_3	8.4%
		$\text{PM}_{2.5}$	5.1%
氣象因子	23.9%	溫度	9.3%
		相對濕度	5.7%
		風速	5.3%
土地利用變數	22.5%	行星邊界層	3.6%
		所有道路 1000m	15.7%
		NDVI 最大值 5000m	6.8%
境外汙染物	2.6%	對流層 NO_2 柱濃度	2.6%
台電因子	9.0%	所有機組發電量/距離	9.0%

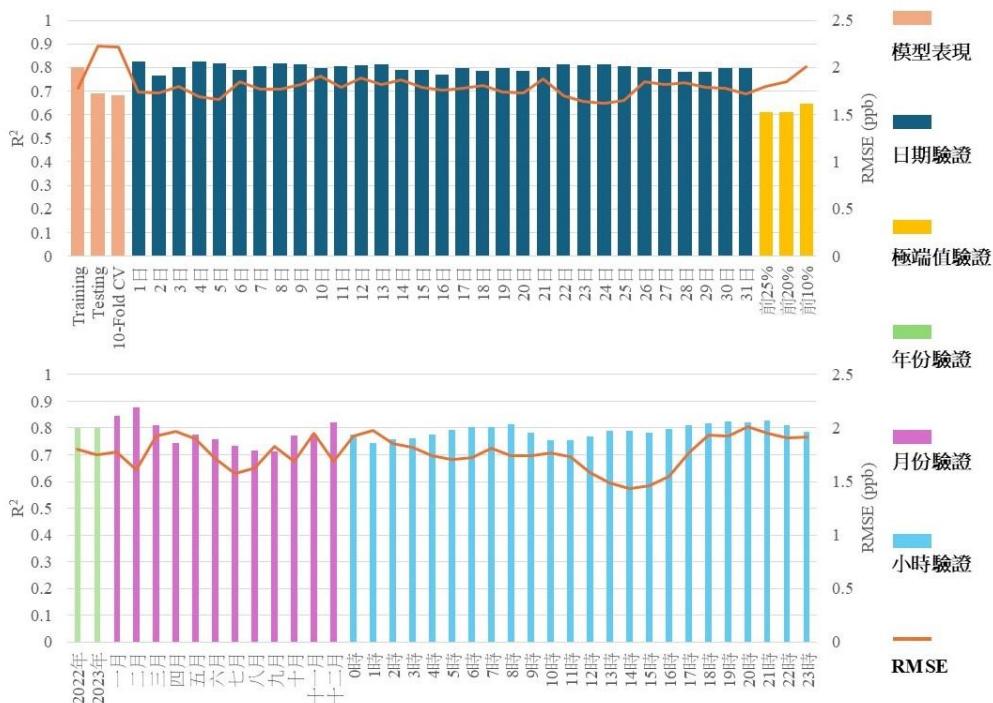
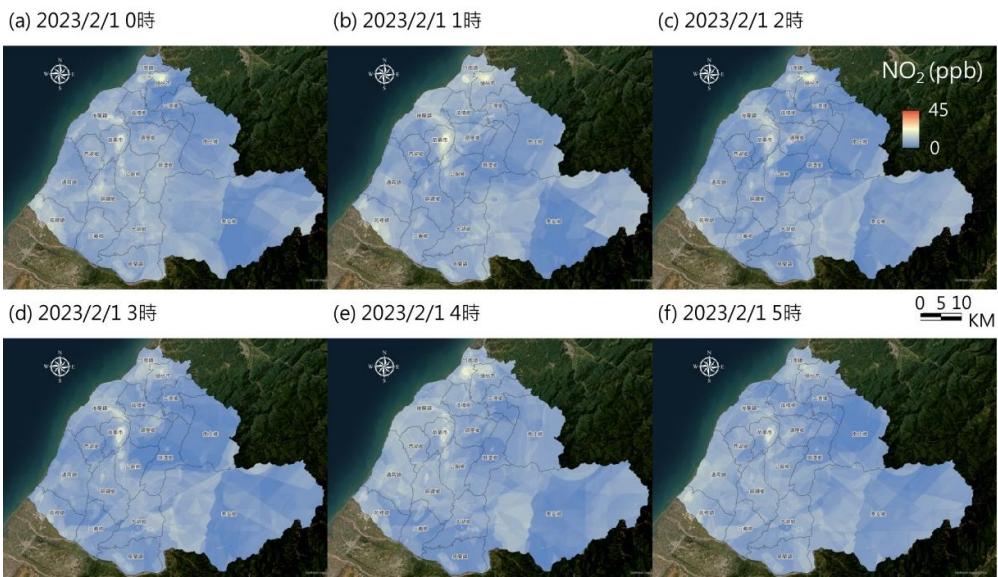
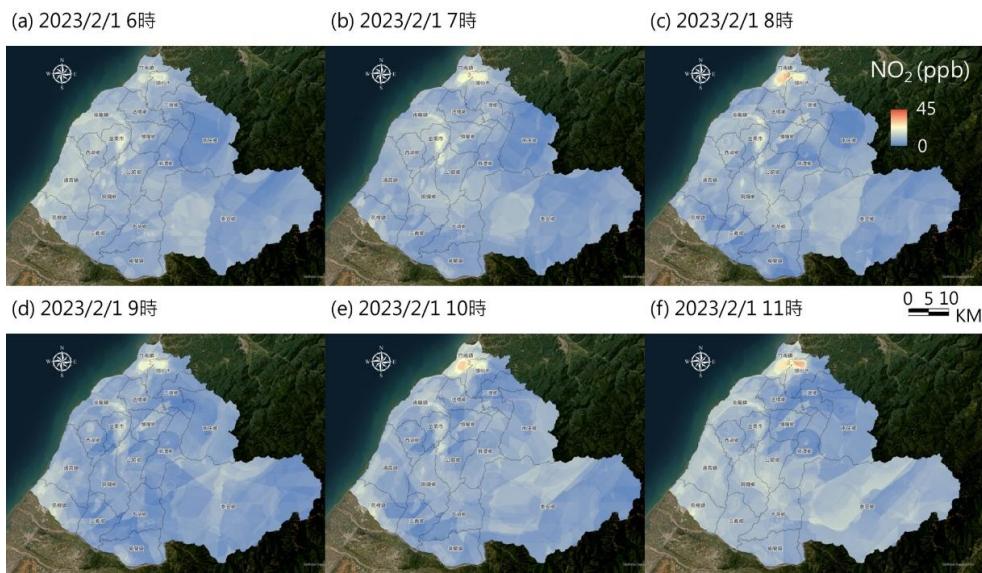
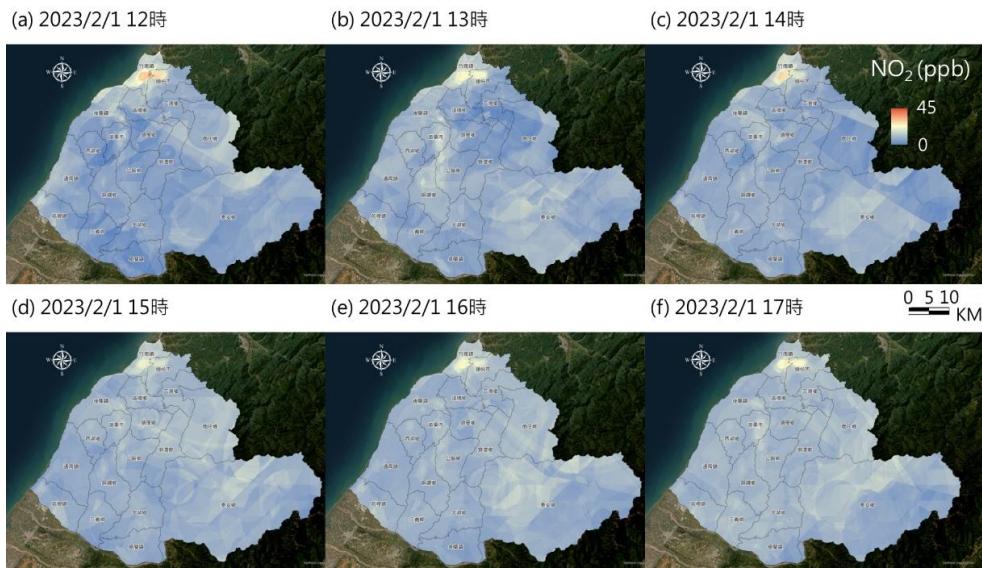
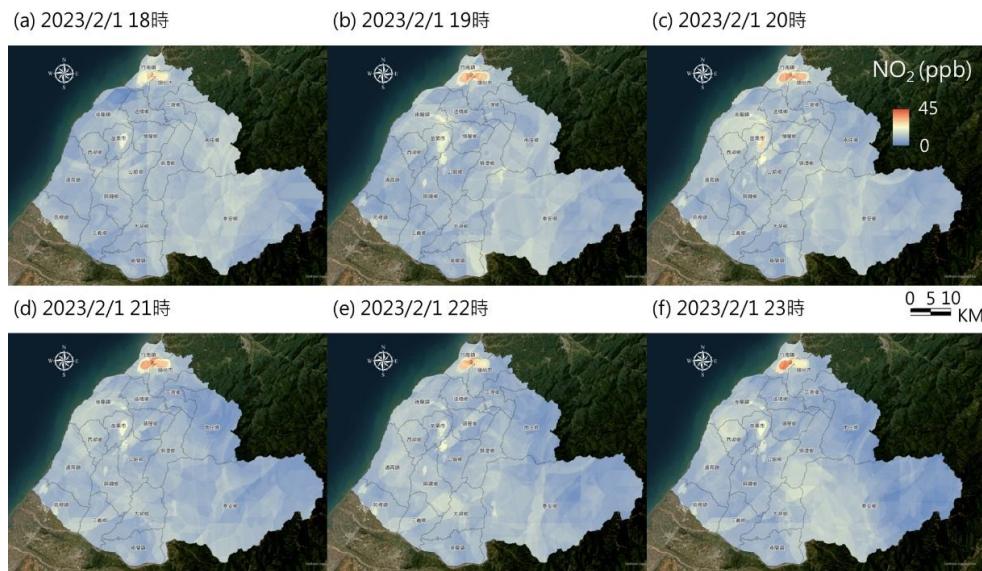


圖 7 模型多層次驗證直方圖

表 6 各測站之分層驗證

測站名	筆數	R^2	Adjusted R^2	R	MSE	RMSE	MAE
海岸里	11010	0.74	0.74	0.86	3.78	1.94	1.48
苗栗	12669	0.83	0.83	0.91	4.02	2.00	1.47
內島里	11445	0.74	0.74	0.86	2.47	1.57	1.18
三義	12864	0.66	0.66	0.81	3.32	1.82	1.32
通南	12573	0.75	0.75	0.87	2.81	1.68	1.29
通灣	12185	0.79	0.79	0.89	2.67	1.63	1.24
通霄	9387	0.80	0.80	0.90	2.47	1.57	1.18
通霄所	12606	0.80	0.80	0.89	2.36	1.54	1.17
頭份	12858	0.85	0.85	0.92	4.53	2.13	1.54
五北	12279	0.75	0.75	0.87	3.16	1.78	1.35
苑裡	11905	0.71	0.71	0.84	2.87	1.70	1.31

圖 8 NO₂ 2023 年 2 月 1 日 0~5 時推估圖

圖 9 NO₂ 2023 年 2 月 1 日 6~11 時推估圖圖 10 NO₂ 2023 年 2 月 1 日 12~17 時推估圖圖 11 NO₂ 2023 年 2 月 1 日 18~23 時推估圖

5. 討論

5.1 模型表現與篩選變數分析

本研究以 lag 1 為代表滯後期，結合 XGBoost 與 SHAP 遞增篩選建立 NO₂ 預測模型。訓練集、獨立測試集與十折交叉驗證的 R² 分別為 0.80、0.69、0.68，RMSE 介於 1.78–2.23 ppb；三組結果差距僅 0.12，顯示過度擬合已被抑制，模型具有良好穩定性。

在時間分層驗證中，日、月、年、時四層尺度的 R² 介於 0.71–0.88，RMSE 變化僅 0.6 ppb；即使限制至濃度前 10 % 的極端事件，RMSE 仍 < 2 ppb，證明模型能穩定捕捉短-中期波動並對高污染情境保持可接受精度。空間驗證顯示 11 座測站 R² 範圍 0.66 – 0.85，除丘陵背風側的三義 (0.66) 與苑裡 (0.71) 略低外，其餘站點皆 ≥ 0.74；RMSE 1.54 – 2.13 ppb，與全區平均誤差相近，顯示空間殘差無明顯集聚。

SHAP 分析指出，NO_x (28.4 %) 為首要推升因子，反映同源污染對 NO₂ 的直接影響；道路環域 1000 m (15.7 %) 排名第二，驗證交通排放的重要性。整體貢獻依序為空氣污染物 41.9 %、氣象因子 23.9 %、土地利用 22.5 %、台電固定源 9 %、境外輸送 2.6 %—突顯交通與擴散條件的共同主導力。

5.2 研究優勢與限制

本研究最大的優勢，在於首度於中小型縣市層級，同步整合氣象、地面監測、衛星遙測、土地利用與通宵電廠資料等五百餘項變數至 50 m × 50 m、逐時解析度的網格，並以 SHAP 遞增篩選機制在每一次特徵增刪時設定 $\Delta R^2 \geq 0.01$ 的門檻，兼顧模型效能與解釋力，同時有效抑制冗餘資訊導致的過擬合。分層驗證進一步證實模型在各種情境下皆維持穩健精度，顯示其作為政策評估與動態預警工具的可行性。然而，本研究亦存在若干限制：首先，監測站點主要分布於沿海與平原，丘陵及背風地形樣本相對不足，導致三義等站點誤差稍高，暗示地形複雜區的模型不確定性仍待下降；其次，固定源排

放係以推估瞬時排放，可能低估尖峰排放對濃度場的瞬時影響；第三，本研究時間跨度僅涵蓋 2022–2023 兩年，尚不足以解析長期的排放趨勢與氣候變化效應；最後，雖使用節慶虛擬變數扣除短期行為干擾，仍缺乏移動監測與個體暴露資料，無法直接量化健康效應或交通動態對模型的影響。

5.3 未來方向

未來研究可從資料與方法兩方面同步精進：在資料層面，建議於山區與背風側增設固定站或微型感測器，以補齊地形複雜區的觀測缺口，並持續擴充時間序列至五年以上，以利評估能源轉型與氣候變動的長期效應；同時，可引入高時空解析衛星產品，提升夜間與雲遮條件下的填補能力。在方法層面，未來可以採用「整合單一機器學習至集成混和空間模型 (Ensemble Mixed Spatial Model, EMSM)」的概念，將不同演算法的優勢匯聚於一體，以提高濃度預測的精度與穩健性。應用方面，未來可將逐時濃度與健保資料鏈結，探討不同暴露水平對急診就診、氣喘發生等健康指標的短期效應，並搭配在地政府排放管制清單，建立高頻空品預報暨早期警報系統；同時，透過將本方法移植至其他縣市，可比較不同能源結構下的排放特徵與污染熱點，進一步驗證模型並支持區域合作的污染治理策略。

6. 結論

本研究整合五百餘項多源資料，於 50 m × 50 m、逐時尺度上建構具解釋性的 NO₂ 機器學習模型，驗證顯示在時空層級皆維持 $R^2 \geq 0.70$ 且 RMSE < 2.3 ppb 的穩健表現。SHAP 解析揭示 NO_x 與交通密度為最關鍵驅動因子，與以往文獻結果一致。方法上，SHAP 遞增篩選有效平衡效能與可解釋力，為中小型縣市空品評估提供可複製的框架；應用上，逐時推估圖與特徵貢獻度可直接支援風險預警與排放管制情境模擬。

本研究亦存在限制：監測佈點與觀測期程仍有限，且瞬時固定源排放可能被低估。未來可透過增設監測據點、延長時序，並導入 EMSM 以提升外推能力與政策決策價值，為能源轉型與區域空品改善

奠定更扎實的科學基礎。此外，建議結合在地微型感測與移動監測以補足複雜地形區的觀測缺口，並在集成模型框架下強化夜間與極端天氣情境；國內既有研究亦顯示，Geo-AI 與集成式方法可在本土資料條件下產製高品質細網格成果，支援實務端之預警與管制 (Hsu *et al.*, 2022、Hsu *et al.*, 2024、Babaan *et al.*, 2024)。

參考文獻

- 內政部戶政司，2023。人口統計資料，
<https://www.ris.gov.tw/>，引用 2024/04/18。
 [Department of Household Registration, M.O.I., 2023. Population statistics by administrative division, Available at: <https://www.ris.gov.tw/> , Accessed April 18, 2024. (in Chinese)]
- 內政部國土測繪中心，2023。行政區域統計，
<https://www.nlsc.gov.tw/>，引用 2024/04/18。
 [National Land Surveying and Mapping Center, 2023. Administrative area statistics, Available at: <https://www.nlsc.gov.tw/>, Accessed April 18, 2024. (in Chinese)]
- Achakulwisut, P., Brauer, M., Hystad, P., and Anenberg, S.C., 2019. Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO₂ pollution: estimates from global datasets, *Lancet Planetary Health*, 3(4): e166-e178, DOI: 10.1016/S2542-5196(19)30046-4.
- Araki, S., Shima, M., and Yamamoto, K., 2018. Spatiotemporal land use random forest model for estimating metropolitan NO₂ exposure in Japan, *Science of The Total Environment*, 634: 1269-1277, DOI: 10.1016/j.scitotenv.2018.03.324.
- Babaan, J., Wong, P.-Y., Chen, P.-C., Chen, H.-L., Lung, S.-C.C., Chen, Y.-C., and Wu, C.-D., 2024. Geospatial artificial intelligence for estimating daytime and nighttime nitrogen dioxide concentration variations in Taiwan: A spatial prediction model, *Journal of Environmental Management*, 360: 121198, DOI: 10.1016/j.jenvman.2024.121198.

- Chen, G., Guo, Y., Yue, X., Tong, S., Gasparrini, A., Bell, M.L., Armstrong, B., Schwartz, J., Jaakkola, J.J.K., Zanobetti, A., Lavigne, E., Saldiva, P.H.N., Kan, H., Roye, D., Milojevic, A., Overcenco, A., Urban, A., Schneider, A., Entezari, A., Vicedo-Cabrera, A.M., Zeka, A., Tobias, A., Nunes, B., Alahmad, B., Forsberg, B., Pan, S.C., Iniguez, C., Ameling, C., Valencia, C.D.L.C., Astrom, C., Houthuijs, D., Dung, D.V., Samoli, E., Mayvaneh, F., Sera, F., Carrasco-Escobar, G., Lei, Y., Orru, H., Kim, H., Holobaca, I.H., Kysely, J., Teixeira, J.P., Madureira, J., Katsouyanni, K., Hurtado-Diaz, M., Maasikmets, M., Ragettli, M.S., Hashizume, M., Stafoggia, M., Pascal, M., Scortichini, M., Coelho, M.D.S.Z.S., Ortega, N.V., Ryti, N.R.I., Scovronick, N., Matus, P., Goodman, P., Garland, R.M., Abrutzky, R., Garcia, S.O., Rao, S., Fratianni, S., Dang, T.N., Colistro, V., Huber, V., Lee, W., Seposo, X., Honda, Y., Guo, Y.L., Ye, T., Yu, W., Abramson, M.J., Samet, J.M., and Li, S., 2021. Mortality risk attributable to wildfire-related PM_{2.5} pollution: A global time series study in 749 locations, *Lancet Planetary Health*, 5(9): e579–e587, DOI: 10.1016/S2542-5196(21)00200-X.
- Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Janssen, N.A.H., Martin, R.V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Vermeulen, R., Brunekreef, B., and Hoek, G., 2019. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide, *Environment International*, 130: 104934, DOI: 10.1016/j.envint.2019.104934.
- Cole, M.A., Elliott, R.J.R., and Liu, B., 2020. The

- impact of the Wuhan Covid-19 lockdown on air pollution and health: A machine learning and augmented synthetic control approach, *Environmental and Resource Economics*, 76: 553-580, DOI: 10.1007/s10640-020-00483-4.
- Crutzen, P.J., 1970. The influence of nitrogen oxides on the atmospheric ozone content, *Quarterly Journal of the Royal Meteorological Society*, 96: 320-325, DOI: 10.1002/qj.49709640815.
- Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., and Schwartz, J., 2020. Assessing NO₂ concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging, *Environmental Science & Technology*, 54(3): 1372-1384, DOI: 10.1021/acs.est.9b03358.
- García, M.V., and Aznarte, J.L., 2020. Shapley additive explanations for NO₂ forecasting, *Ecological Informatics*, 56: 101039, DOI: 10.1016/j.ecoinf.2019.101039.
- He, S., Dong, H., Zhang, Z., and Yuan, Y., 2022. An ensemble model-based estimation of nitrogen dioxide in a southeastern coastal region of China, *Remote Sensing*, 14(12): 2807, DOI: 10.3390/rs14122807.
- Hsu, C.-Y., Lee, R.-Q., Wong, P.-Y., Lung, S.-C.-C., Chen, Y.-C., Chen, P.-C., Adamkiewicz, G., and Wu, C.-D., 2024. Estimating morning and evening commute period O₃ concentration in Taiwan using a fine spatial-temporal resolution ensemble mixed spatial model with Geo-AI technology, *Journal of Environmental Management*, 351: 119725, DOI: 10.1016/j.jenvman.2023.119725.
- Hsu, C.-Y., Xie, H.-X., Wong, P.-Y., Chen, Y.-C., Chen, P.-C., and Wu, C.-D., 2022. A mixed spatial prediction model in estimating spatiotemporal variations in benzene concentrations in Taiwan, *Chemosphere*, 301: 134758, DOI: 10.1016/j.chemosphere.2022.134758.
- Huang, C., Sun, K., Hu, J., Xue, T., Xu, H., and Wang, M., 2022. Estimating 2013-2019 NO₂ exposure with high spatiotemporal resolution in China using an ensemble model, *Environmental Pollution*, 292: 118285, DOI: 10.1016/j.envpol.2021.118285.
- Kamińska, J.A., 2019. A random forest partition model for predicting NO₂ concentrations from traffic flow and meteorological conditions, *Science of The Total Environment*, 651: 475-483, DOI: 10.1016/j.scitotenv.2018.09.196.
- Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C.-K., and Kim, S., 2021. Estimation of surface-level NO₂ and O₃ concentrations using TROPOMI data and machine learning over East Asia, *Environmental Pollution*, 288: 117711, DOI: 10.1016/j.envpol.2021.117711.
- Kashtan, Y., Nicholson, M., Finnegan, C.J., Ouyang, Z., Garg, A., Lebel, E.D., Rowland, S.T., Michanowicz, D.R., Herrera, J., Nadeau, K.C., and Jackson, R.B., 2024. Nitrogen dioxide exposure, health outcomes, and associated demographic disparities due to gas and propane combustion by U.S. stoves, *Science Advances*, 10(18): eadm8680, DOI: 10.1126/sciadv.adm8680.
- Keller, C.A., Evans, M.J., Knowland, K.E., Hasenkopf, C.A., Modekurty, S., Lucchesi, R.A., Oda, T., Franca, B.B., Mandarino, F.C., Díaz Suárez, M.V., Ryan, R.G., Fakes, L.H., and Pawson, S., 2021. Global impact of COVID-19 restrictions on the surface concentrations of nitrogen dioxide and ozone, *Atmospheric Chemistry and Physics*, 21(5): 3555-3592, DOI: 10.5194/acp-21-3555-2021.
- Li, L., Lurmann, F., Habre, R., Urman, R., Rappaport, E., Ritz, B., Chen, J.C., Gilliland, F.D., and Wu, J., 2017. Constrained mixed-effect models with ensemble learning for prediction of nitrogen

- oxides concentrations at high spatiotemporal resolution, *Environmental Science & Technology*, 51(17): 9920-9929, DOI: 10.1021/acs.est.7b01864.
- Petetin, H., Bowdalo, D., Soret, A., Guevara, M., Jorba, O., Serradell, K., and Pérez García-Pando, C., 2020. Meteorology-normalized impact of the COVID-19 lockdown upon NO₂ pollution in Spain, *Atmospheric Chemistry and Physics*, 20(18): 11119-11141, DOI: 10.5194/acp-20-11119-2020.
- Qian, Y., Li, H., Rosenberg, A., Li, Q., Sarnat, J., Papatheodorou, S., Schwartz, J., Liang, D., Liu, Y., Liu, P., and Shi, L., 2021. Long-term exposure to low-level NO₂ and mortality among the elderly population in the Southeastern United States, *Environmental Health Perspectives*, 129(12): 127009, DOI: 10.1289/EHP9044.
- Shaban, K., Kadri, A., and Rezk, E., 2016. Urban air pollution monitoring system with forecasting models, *IEEE Sensors Journal*, 16(8): 2598-2606, DOI: 10.1109/JSEN.2016.2514378.
- Shen, Y., de Hoogh, K., Schmitz, O., Clinton, N., Tuxen-Bettman, K., Brandt, J., Christensen, J.H., Frohn, L.M., Geels, C., Karssenberg, D., Vermeulen, R., and Hoek, G., 2022. Europe-wide air pollution modeling from 2000 to 2019 using geographically weighted regression, *Environment International*, 168: 107485, DOI: 10.1016/j.envint.2022.107485.
- Shi, Z., Song, C., Liu, B., Lu, G., Xu, J., Van Vu, T., Elliott, R.J.R., Li, W., Bloss, W.J., and Harrison, R.M., 2021. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns, *Science Advances*, 7(3): eabd6696, DOI: 10.1126/sciadv.abd6696.
- Tiotiu, A.I., Novakova, P., Nedeva, D., Chong-Neto, H.J., Novakova, S., Steiropoulos, P., and Kowal, K., 2020. Impact of air pollution on asthma outcomes, *International Journal of Environmental Research and Public Health*, 17(17): 6212, DOI: 10.3390/ijerph17176212.
- Vardoulakis, S., Giagloglou, E., Steinle, S., Davis, A., Sleeuwenhoek, A., Galea, K.S., Dixon, K., and Crawford, J.O., 2020. Indoor exposure to selected air pollutants in the home environment: A systematic review, *International Journal of Environmental Research and Public Health*, 17(23): 8972, DOI: 10.3390/ijerph17238972.
- Wang, Y., Wen, Y., Wang, Y., Zhang, S., Zhang, K.M., Zheng, H., Xing, J., Wu, Y., and Hao, J., 2020. Four-month changes in air quality during and after the COVID-19 lockdown in six megacities in China, *Environmental Science & Technology Letters*, 7(11): 802-808, DOI: 10.1021/acs.estlett.0c00605.
- Wei, J., Li, Z., Pinker, R.T., Wang, J., Sun, L., Xue, W., Li, R., and Cribb, M., 2021. Himawari-8-derived diurnal variations in ground-level PM_{2.5} pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM), *Atmospheric Chemistry and Physics*, 21(10): 7863-7880, DOI: 10.5194/acp-21-7863-2021.
- Wong, P.-Y., Su, H.-J., Lee, H.-Y., Chen, Y.-C., Hsiao, Y.-P., Huang, J.-W., Teo, T.-A., Wu, C.-D., and Spengler, J.D., 2021. Using land-use machine learning models to estimate daily NO₂ concentration variations in Taiwan, *Journal of Cleaner Production*, 317: 128411, DOI: 10.1016/j.jclepro.2021.128411.
- Yang, J., Wen, Y., Wang, Y., Zhang, S., Pinto, J.P., Pennington, E.A., Wang, Z., Wu, Y., Sander, S.P., Jiang, J.H., Hao, J., Yung, Y.L., and Seinfeld, J.H., 2021. From COVID-19 to future electrification: Assessing traffic impacts on air quality by a machine-learning model, *Proceedings of the National Academy of Sciences of the United States of America*, 118(26): e2102705118, DOI:

- 10.1073/pnas.2102705118.
- Zimmerman, N., Presto, A.A., Kumar, S.P.N., Gu, J., Hauryliuk, A., Robinson, E.S., Robinson, A.L., and Subramanian, R., 2018. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques*, 11(1): 291-313, DOI: 10.5194/amt-11-291-2018.

Application of Geo-AI Modeling to Estimate NO₂ Concentrations and Analyze Contributing Factors in Miaoli, Taiwan

Chieh-Ying Chen¹ Sin-Yi Lai¹ Yu-Ting Zeng² Chih-Da Wu^{3,4,5,6,7*}

Abstract

This study takes Miaoli County as a demonstration area to develop a high-resolution NO₂ estimation model by integrating multiple data sources, including meteorological data, ground monitoring, satellite remote sensing, land use, and data from the Tongxiao Power Plant. An hourly 50 m × 50 m resolution model was constructed using the XGBoost algorithm combined with a SHAP-based incremental feature selection mechanism, with lag1 (1-hour lag) identified as the optimal time delay. The model achieved an R² of 0.80 and an RMSE of 1.78 ppb. SHAP analysis revealed that NO_x and road density were the most influential predictors, while emissions from the power plant also exhibited regional impact. Multilevel validation confirmed the model's robustness across temporal, spatial, and high-pollution scenarios. The results demonstrate the potential of Geo-AI in air quality estimation for small and medium-sized counties and provide a quantitative basis for risk warning and pollution control.

Keywords: Nitrogen Dioxide, Geo-AI, Machine Learning, Spatial Estimation Map

¹ Master Student, Department of Geomatics, National Cheng Kung University

Received Date: Jul. 11, 2025

² Assistant Researcher, Department of Geomatics, National Cheng Kung University

Revised Date: Sep. 03, 2025

³ Professor, Department of Geomatics, National Cheng Kung University

Accepted Date: Oct. 15, 2025

⁴ Adjunct Associate Research, National Institute of Environmental Health Sciences, National Health Research Institutes

⁵ Adjunct Professor, Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University

⁶ Adjunct Research, Research Center for Precision Environmental Medicine, Kaohsiung Medical University

⁷ Joint Appointment Research, Chronic Diseases and Health Promotion Research Center, Chang Gung University of Science and Technology

* Corresponding Author, E-mail: chidawu@mail.ncku.edu.tw