

使用空間混合模型分析 PM_{2.5} 的長期變化趨勢 (1994 年至 2020 年) – 以臺中市為例

林祐如¹ 許家瑋² 李佳禾³ 曾于庭⁴ 翁佩詒⁵ 陳保中^{6,7,8,9}
陳裕政^{10,11,12} 吳治達^{13,14,15,16*}

摘要

研究以臺中市為例，利用空間模型推估 1994 至 2020 年 PM_{2.5} 濃度趨勢，並評估城市開發對空氣品質的影響。研究使用 PM_{2.5} 相關汙染物、氣象資料、土地利用、地標、路網、地形、植生指數等作為預測變數。結合土地利用迴歸和機器學習方法，使用隨機森林、梯度提升機、極限梯度提升、輕量梯度提升機和基於梯度提升的決策樹模型擬合預測模型。通過數據拆分、十折交叉和外部驗證確認模型穩健性，結果顯示模型穩定且可信，Adj-R² 為 0.93。結果表明多數地點的「年份」變數係數為負，顯示過去 25 年空氣污染顯著改善。研究強調在城市開發規劃中管理和控制空氣污染的重要性。

關鍵詞： PM_{2.5}、都市開發、空間混合模型、趨勢分析

1. 前言

隨著都市人口的快速成長與都市數量的增加，都市開發所延伸的環境問題在 21 世紀成為一項重要議題。都市範圍的擴大和和城市基礎設施的不斷發展，使人民生活水平的提高、也使經濟增長以及就業機會增加。然而，這些好處卻是以環境為代價，因為都市化也帶來了環境惡化、生態系統破壞、空氣污染等負面影響。在這些負面影響中，空氣污染對公眾健康和環境構已然成嚴重威脅，因此已然成

為全球許多城市必須要面對的重要議題。

暴露於高濃度的 PM_{2.5} 與呼吸和心血管疾病 (Ren *et al.*, 2006)、肺癌 (Lepeule *et al.*, 2012) 等疾病相關，且可能加重現有的心肺疾病導致早逝 (Anderson *et al.*, 2012)，因此減少推估 PM_{2.5} 濃度暴露的誤差對流行病學研究至關重要。基於此，空氣污染濃度的空間建模在空氣污染流行病學研究中越來越普遍，其中與傳統空間內插法相比，土地利用迴歸 (Land-use Regression, LUR) 更可以考慮監測站間特定排放源造成的影響，因此已被證明在表

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 博士生

³ 嘉義市政府環境保護局 局長

⁴ 國立成功大學測量及空間資訊學系 研究助理

⁵ 國立成功大學工業衛生學科暨環境醫學研究所 博士生

⁶ 國家衛生研究院國家環境醫學研究所 特聘研究員兼所長

⁷ 國立臺灣大學公共衛生學院環境與職業健康科學研究所 特聘教授

⁸ 國立臺灣大學公共衛生學院公共衛生學系 特聘教授

⁹ 國立臺灣大學醫學院附設醫院環境及職業醫學部 主治醫師

¹⁰ 國家衛生研究院國家環境醫學研究所 研究員

¹¹ 國立聯合大學環境與安全衛生工程學系 兼任教授

¹² 中國醫藥大學職業安全與衛生學系 兼任教授

¹³ 國立成功大學測量及空間資訊學系 教授

¹⁴ 國家衛生研究院國家環境醫學研究所 兼任副研究員

¹⁵ 國立中興大學永續農業創新發展中心 教授

¹⁶ 高雄醫學大學精準環境醫學研究中心 兼任研究員

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

收到日期：民國 113 年 02 月 27 日

修改日期：民國 113 年 05 月 09 日

接受日期：民國 113 年 05 月 22 日

現局部排放源和城市內污染變化之間的空間關係方面具有優勢 (Michanowicz *et al.*, 2016)。因此本研究進一步使用了混合克利金土地利用迴歸模型 (hybrid Kriging-LUR)，該方法與傳統的 LUR 模型不同，將克利金插值的預測濃度作為 LUR 建模中的一個變量，以提高對 PM_{2.5} 變數預測的準確性 (Wu *et al.*, 2018)。

過去已有許多研究旨在探討都市開發與空氣污染之關聯，且前人研究已表明都市開發與空氣污染有直接的影響。都市化透過交通、工業和民生來源的污染物排放直接或間接地影響空氣品質，同時城市的高人口密度也加劇了空氣污染程度。儘管前人研究已表明都市發展與空氣污染相關，卻鮮少有研究深入探討都市開發與空氣品質的長期影響，基於此，本研究將深入探討與分析以全面了解都市開發與空氣品質之間的複雜關聯。釐清這項議題可以更加有效的訂定相關政策以及指引人們實現永續發展的方向，以減輕都市化對空氣品質的負面影響。

2. 材料及方法

2.1 研究試區

臺中市位於中部地區，地理位置優越，為主要交通幹線的交會處，是臺灣的重要交通樞紐，也是六個直轄市之一。臺中市土地主要用途為住宅、商業及工業用途。該市擁有發達的交通系統，包括公車系統、鐵路、捷運和機場。與臺灣許多其他大都市相同，由於工業活動和交通排放，臺中市近年來也出現了空氣污染問題。地方政府針對這項問題採取了各種措施來改善空氣品質，例如推廣大眾運輸的使用、鼓勵採用綠色能源和加強環境管理等。雖然有了一些措施改善空氣品質，臺中市僅有五座國家空氣品質監測站 (圖 1)，亦指每 50 萬人共享一座監測站，且對於那些離監測站的地區，監測站測量的數值即無法正確反映該地區實際的空氣品質狀態。故本研究欲利用地理資訊系統空間分析 (Geographic Information System, GIS)，結合土地利用迴歸和機械學習 (Machine Learning) 減少因測站數量限制所造成的空氣污染濃度誤差。

2.2 研究材料

本研究使用之資料庫包含 PM_{2.5} 與相關空氣污染物監測數據、土地利用清單、數位路網地圖、興趣點 (Point of Interest, POI) 地標資料庫、中級解析度成像分光輻射度計 (Moderate-resolution Imaging Spectroradiometer, MODIS) 常態化差異植生指標 (Normalized Difference Vegetation Index, NDVI) 資料庫、高精數值地形模型 (Digital Terrain Model, DTM) 和氣象資料集，作為建構土地利用迴歸模型的變數。

其中 PM_{2.5} 相關空氣污染物包含 1994 至 2020 年之懸浮微粒 (Particulate matter, PM₁₀)、二氧化氮 (Nitrogen dioxide, NO₂)、氮氧化物 (Nitrogen oxides, NO_x)、臭氧 (Ozone, O₃) 及二氧化硫 (Sulfur dioxide, SO₂)，各項空污濃度觀測值取自於環保署 (今環境部) 環境資料開放平臺，在本研究區內有豐原、忠明、沙鹿、西屯與大里站，共 5 個測站。2007 年和 2015 年的土地利用清單取自內政部國土測繪中心，用於獲取土地利用分類 (Land-use/Land-cover, LULC) 資訊，包括住宅區、農田、綠地、水體和機場。道路模式數據則是採用交通部製作的 2001、2006、2013、2017 及 2020 年數位道路網路地圖。中式餐廳、寺廟等特定文化排放源的分布是取得自 2006、2008、2010 及 2012 年的 POI 地標資料庫。POI 地標資料庫經過 Google Earth 改編，其代表了全臺灣地標的分布 (Chen *et al.*, 2020)。MODIS NDVI 資料庫係使用美國國家航空暨太空總署 (National Aeronautics and Space Administration, NASA) 獲得的 250m × 250m 網格解析度的 NDVI 資料，資料期程為 2000 至 2020 年，用為說明研究期間研究試區植被的分布情況。我們也基於 2016 年內政部 20 m × 20 m 解析度數值地形模型資料 (DEM) 計算監測點的高程、坡度等地形特徵。而研究期間的氣象變數，包括溫度、相對濕度、風速、風向、降水量和紫外線，紫外線資料庫取自於環保署環境資料開放平臺及中央氣象局 CODIS 觀測資料，資料期程為 2000 至 2020 年，其餘氣象變數來自大氣水文資料庫，資料期程為 1994 至 2020 年。

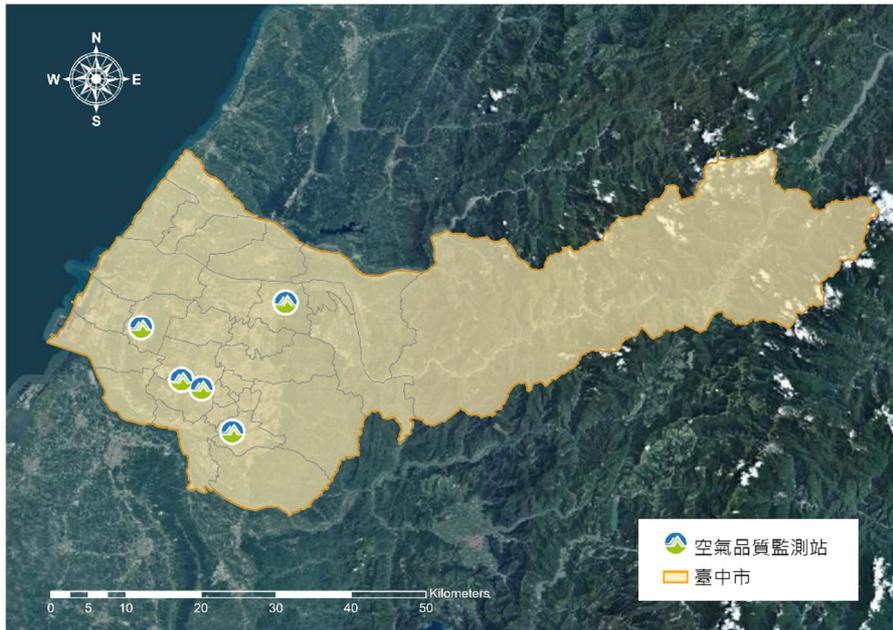


圖 1 臺中市空氣品質監測站分布圖

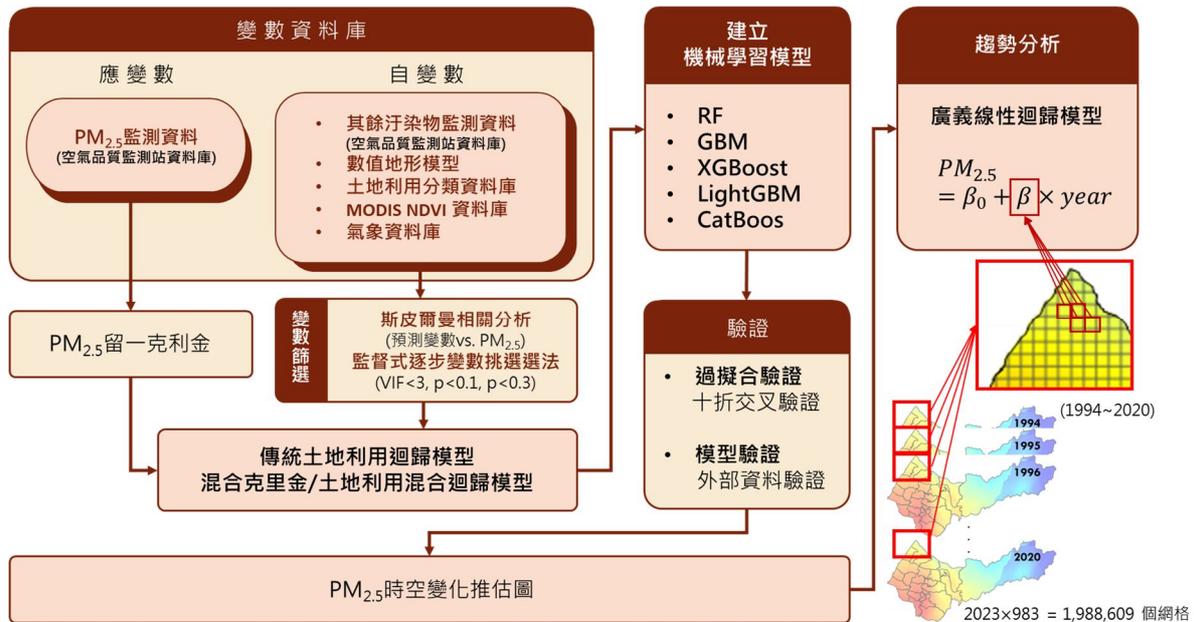


圖 2 研究流程圖

2.3 研究方法

本研究流程大致分為五個部分，依序為土地利用資料庫前處理、變數篩選、建立土地利用迴歸模型之架構、建立機械學習推估模型及趨勢分析。以下將對各個步驟逐一說明，研究流程圖如圖 2。

2.3.1 土地利用資料庫前處理

基於電腦效能能夠負荷之最細解析度 50 m × 50 m 的網格，本研究利用 ArcGIS Pro 2.6 軟體將研

究區域內各項土地利用排放源和空間變數的數據轉換為 50 m×50 m 的網格數據，然後計算兩種空間分布資訊，分別為 (1) 計算半徑為 25、50、150、250、500、750、1000、1250、1500、1750、2000、2500、3000、4000 和 5000 m 的各個網格中心點的長度 and 面積等資訊，以獲得研究區域範圍內這些土地利用排放變數的空間密度分布；(2) 網格中心點到每個土地利用空間變數的歐幾里得距離 (Euclidean Distance)。預計總共將考慮約 520 個空間排放源變數。

2.3.2 變數篩選

本研究結合斯皮爾曼相關分析 (Spearman Correlation Analysis) 及監督式逐步變數挑選選法 (Supervised Stepwise Variable Selection), 共兩種方法用於選擇土地利用迴歸模型之變數。其中斯皮爾曼相關分析用於檢查各項預測變數與 PM_{2.5} 間的關聯, 僅保留符合預定義效應方向的預測變數, 然後將其納入監督逐步變數選擇過程中。保留或刪除變數的統計標準分別為 p 值 < 0.1 和 < 0.3 。最後, 採用變異數膨脹因子 (Variance inflation factor, VIF) 來檢驗所發展模型中變數的相關性。在建立最終的常規土地利用迴歸模型時保留了 $VIF < 3$ 以及與預測方向一致的預測變數。

2.3.3 土地利用迴歸模型之架構

在空氣汙染推估研究領域, 前人研究多數僅基於單一測站的 PM_{2.5} 資料進行分析, 而利用土地利用迴歸將這些測站資料轉換為面狀資料的目的在於提高空間解析度和涵蓋範圍。通過將測站資料轉換為面狀資料, 可以更全面地捕捉空間變化和分布模式, 從而更好地理解 and 預測 PM_{2.5} 的空間分佈情況。這樣的分析可以為制定區域性的空氣汙染控制策略提供更具體和準確的信息。本研究採用傳統的土地利用迴歸 (Land-use Regression) 和混合克利金/土地利用混合迴歸 (Hybrid Kriging/Land-use Mixed Regression) 進行初步模型建構和重要空間解釋變數的篩選。克利金/土地利用迴歸混合模型的原理為, 在同一測量時間有相鄰點的監測數據, 而這些相鄰點的監測濃度值必須與待推估地點的濃度有一定的關係 (Wu *et al.*, 2018)。我們再進一步透過留一克利金法 (Leave-one-out Kriging) 推估出各未知濃度點的克利金污染物的推估值, 並將其作為解釋變數納入統計模型中, 其餘過程與土地利用迴歸相同, 即可完成混合估計模型的建構。

2.3.4 建立機械學習推估模型

基於傳統的土地利用迴歸和混合克利金/土地利用混合迴歸, 挑選使用上述選取的變數並以兩個統計模型為基礎, 利用 Python 中的 AutoML 套

件, 選擇了隨機森林 (Random Forest, RF)、梯度提升機 (Gradient Boosting Machine, GBM)、極限梯度提升 (eXtreme Gradient Boosting, XGboost)、輕量梯度提升機 (Light Gradient Boosting Machine, Light GBM) 和基於梯度提升的決策樹模型 (CatBoost) 進行模型構建, 其中 RF 是由 Breiman (2001) 提出的一種集成方法。這種方法結合了多個決策樹, 其結果比單個決策樹具有更高的準確性。在 RF 中包含的預測變量是通過自助法 (bootstrap method) 隨機分割的, 並且被輸入到不同的決策樹中。每棵樹中的數據被用來訓練預測模型, 且樹生長的程度影響了模型的性能。GBM 是一種將多個學習器 (決策樹) 迭代結合以提高預測準確性的方法 (Friedman, 2001)。這種方法使用梯度下降函數來最小化誤差。使用 GBM 的動機在於該方法允許用戶選擇與其目的相對應的特定損失函數。XGBoost 是一種梯度提升方法, 它使用多個梯度提升的決策樹 (GBDTs) (Chen & Guestrin, 2016), 也利用損失的泰勒展開近似 (Taylor expansion approximation) 來獲得對模型誤差的精確評估並找到最合適的目標函數。XGBoost 的優點包含更少的時間消耗和更好的預測能力。LightGBM 是梯度提升決策樹方法的一個分支, 它具有快速、分散式、高性能和較低的內存使用率的優點 (Ke *et al.*, 2017)。該算法引入了基於葉子的分割與有限樹生長深度方法, 而不是基於層級的分割方法來訓練複雜的決策樹。通過使用具有最高訊息增益的葉節點進行分割, 模型可以比其他提升算法具有更好的準確性, 並提高訓練效率。CatBoost 是一種基於梯度提升機器學習的決策樹方法, 用於處理分類特徵 (Prokhorenkova *et al.*, 2018)。這種方法採用了提升方法的優勢, 並進一步提出了一種無偏差的程序來處理分類特徵。該方法允許使用整個數據集來訓練模型, 並使用目標統計訊息來處理具有最小訊息損失的分類特徵。上述演算法過去常應用於空氣汙染推估相關研究, 故選之應用於本研究 (Hsu *et al.*, 2022、Wong *et al.*, 2021)。最後選擇其中表現最好的機械學習方法, 再結合兩個統計模型, 作為推估臺中市 PM_{2.5} 長期趨勢的模型架構。

2.3.5 模型驗證和模型表現評估

本研究為驗證模型是否有過擬合之情形，以及確認模型的穩定性，進行了數據拆分、十折交叉驗證以及外部驗證等方法。過度擬合測試同時採用了數據拆分法和十折交叉驗證。數據拆分法係將 80% 的數據用於模型訓練，其中五個測站皆約有 7600 筆數據 (圖 3 (a))，另外的 20% 用於模型測試，其中五個測站皆約有 1900 筆數據 (圖 3 (b))，資料皆均勻分布在研究試區中。十折交叉驗證法則係將 90% 的數據隨機選擇用於模型訓練，剩下的 10% 用於模型測試。此過程重複十次，以便每組觀測資料至少在測試集中出現一次。為進一步確認模型的擬合程度和穩健性本研究也進行了外部驗證。本研究使用 1994 至 2019 年的每日 PM_{2.5} 觀測數據用於模型訓練，並將 2020 年的每日 PM_{2.5} 觀測數據視為外部數據，並與現地觀測值進行比較。

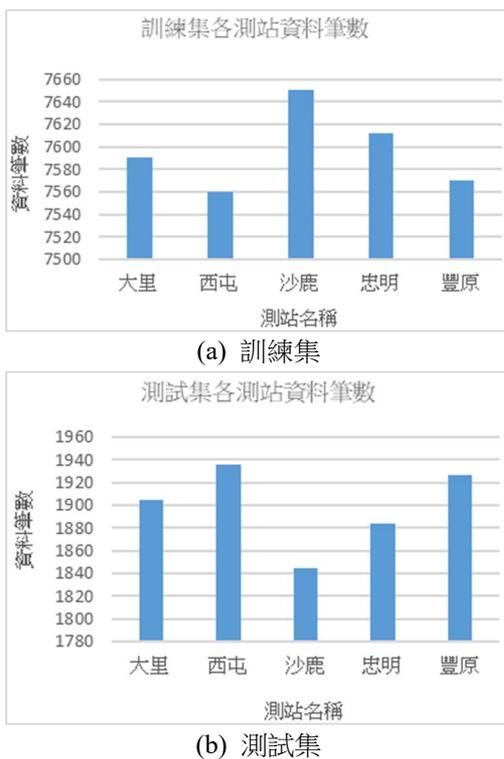


圖 3 訓練集與測試集使用之各測站資料筆數

為評估模型的效能，在本研究中使用五種統計數據做為指標，其中判定係數 (Coefficient of Determination, R²) 及調整後判定係數 (Adjusted R²) 用於評估模型的預測能力。而 3 種誤差指標均方誤

差 (Mean Squared Error, MSE)、均方根誤差 (Root Mean Square Error, RMSE) 與平均絕對誤差 (Mean Absolute Error, MAE)，則是用於來確認預測值與觀測值之間的殘差。首先，判定係數 (R²) 和調整後判定係數 (Adjusted R²) 用於評估模型的預測能力。R² 衡量了模型對因變量變異性的解釋程度，其取值範圍在 0 到 1 之間，越接近 1 表示模型的預測能力越強。調整後判定係數則修正了 R² 的偏誤，特別是當模型包含多個自變量時，使得不同模型間更具可比性。其次，三種誤差指標均方誤差 (MSE)、均方根誤差 (RMSE) 和平均絕對誤差 (MAE) 用於確認模型預測值與觀測值之間的殘差。MSE 量化了預測值與觀測值之間的平方差的平均值，RMSE 提供了對應的標準化評估，而 MAE 則是平均絕對誤差的絕對值的平均值，對實際預測誤差更加直觀。綜合這些指標可以全面評估模型的預測性能和誤差水平。

2.3.6 趨勢分析

廣義線性回歸模型 (Generalized Linear Regression) 是一種用於探索因變量與一組預測因子之間關係的統計方法。與傳統的線性回歸模型不同，廣義線性回歸模型允許因變量不需要服從常態分佈，並且可以使用不同的響應分佈和連結函數來捕捉不同數據模式中的變化。為了探討臺中市過去 25 年的空氣污染趨勢變化，我們使用廣義線性回歸模型對每個網格位置的 PM_{2.5} 濃度與年份之間的關係進行分析。該模型的形式如下：

$$PM_{2.5} = \beta_0 + \beta \times year \dots\dots\dots(1)$$

其中 PM_{2.5} 表示 PM_{2.5} 濃度，year 表示年分，β₀ 為廣義線性模型的截距項，β 是年份變數的係數。通過評估每個網格的 β 係數，我們可以了解 PM_{2.5} 濃度隨時間的變化趨勢，若 β 為正值，代表濃度隨時間增加，反之，若 β 為負值，代表濃度隨時間減少，根據這些係數的正負可以得知該網格的 PM_{2.5} 濃度是隨時間增加還是減少。基於此，可以更深入地了解臺中市過去 25 年 PM_{2.5} 濃度的變化趨勢。

3. 研究結果

在此部分中，我們將簡單地透過研究期間 $PM_{2.5}$ 的觀測值了解 $PM_{2.5}$ 長期濃度的變化，再進一步比對五種機械學習演算法，並分析最佳演算法的模型效能，最後我們將進一步分析臺中市 $PM_{2.5}$ 濃度在空間與長時間中的變化趨勢。

3.1 平均 $PM_{2.5}$ 濃度

圖 4 為臺中市 1994 年至 2020 年平均 $PM_{2.5}$ 濃度觀測資料折線圖。從圖 4 初步觀察可以發現，自 1994 年以來，臺中市的空氣污染逐年加重隨著城市的發展。2005 年達到頂峰，但隨後逐漸下降。

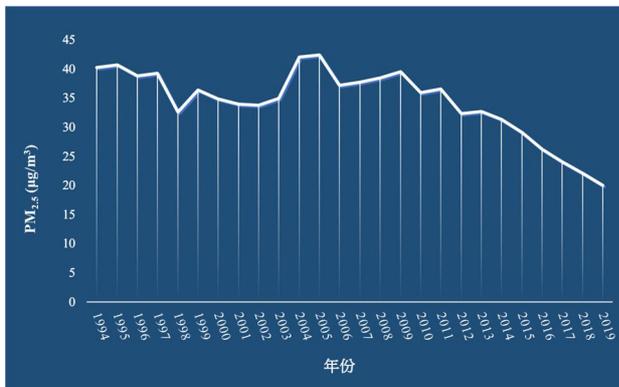


圖 4 1994 至 2020 年臺中市 $PM_{2.5}$ 觀測濃度折線圖

3.2 機械學習模型效能

表 1 為從兩種基於土地利用模型方法篩選出的預測變數，其中， $PM_{2.5}$ 克利金濃度表示由留一克利金法推估得到的 $PM_{2.5}$ 濃度。其他則是土地利用排放源與相關污染物自變量，在這些變數中農作用地、日 SO_2 濃度、寺廟排放源、溫度與工業用住宅對 $PM_{2.5}$ 濃度的增加為正相關，海拔和 NDVI 則是負相關。

承在研究方法中提到，本研究使用了五種機械學習演算法用於建立模型。表 2 顯示了每個模型的結果，針對觀察訓練資料集的結果，可以看出 RF、GBM 和 XGBoost 在 R^2 值方面表現相似，都在 0.92 至 0.93 之間，而 LightGBM 和 CatBoost 則稍微低一些。然而，LightGBM 和 CatBoost 在訓練資料上的 MSE、RMSE 和 MAE 都比其他算法要高，這表

示上述方法在預測方面的準確度稍低。在測試資料集上，各算法的表現相對一致，但是 RF 和 XGBoost 在 RMSE 和 MAE 方面稍微優於其他算法，這表明 RF 和 XGBoost 在避免過度擬合方面的表現較好。十折驗證結果顯示，各算法的性能基本保持一致，但隨機森林和 XGBoost 在 MSE、RMSE 和 MAE 方面的表現仍然相對較好。在外部驗證方面，隨機森林和 XGBoost 在所有指標上都表現最佳，具有最低 MSE、RMSE 和 MAE 值，並且在 R^2 和調整後的 R^2 值方面也較高。總體而言，RF 和 XGBoost 在這個案例中展現了較好的性能，不僅在訓練和測試資料上，而且在十折驗證和外部驗證方面也都表現出色。綜合 Adjusted R^2 與 RMSE 來看，在本研究中 RF 為最佳的機械學習演算法，對 $PM_{2.5}$ 濃度的推估具有最高的解釋力且誤差最小，其 Adjusted R^2 為 0.93，RMSE 為 $4.86 \mu\text{g}/\text{m}^3$ ，表 3 以訓練集資料為例進行模型效能排序，除訓練集成果，在測試集資料與外部驗證中 RF 模型效能皆為最佳，僅十折驗證些微低於 GBM 與 XGBoost，故綜合所有驗證結果 RF 擁有最佳的模型效能。我們也針對該演算法做了過擬合驗證與模型驗證，在十折交叉驗證和外部驗證中也證實了該模型的可靠性，其 Adjusted R^2 分別為 0.86 與 0.89，RMSE 分別為 $6.77 \mu\text{g}/\text{m}^3$ 與 $3.16 \mu\text{g}/\text{m}^3$ 。

表 1 基於 LUR 篩選的預測變數

變數	係數	p 值
常數	-2.94	
$PM_{2.5}$ 濃度 _{克利金}	0.96	
農作用地	0.01	9.10×10^{-3}
高程	-0.01	1.21×10^{-3}
SO_2 濃度 _{克利金}	0.51	8.51×10^{-4}
寺廟 _{1000m}	2357257.71	6.78×10^{-4}
NDVI _{150m}	-4.11	7.33×10^{-4}
日溫度	0.11	6.99×10^{-4}
工業用住宅 _{500m}	0.01	2.70×10^{-4}
LUR 模型 Adjusted R^2 : 0.88		

3.3 臺中市空氣污染趨勢分析

使用模型效能最佳的 RF 演算法完成空間混合

模型後，我們亦使用廣線性迴歸對研究區各網格進行 25 年趨勢分析。在 LUR 篩選的預測變數中，包含隨空間變化之變數，即 NDVI、高程、農作用地、寺廟、工業用住宅；也包含隨時間同時隨空間變化之變數，即日溫度、PM_{2.5} 濃度、SO₂ 濃度，將上述變數帶入廣義線性回歸模型中，即可透過不同時期的網格資料的 PM_{2.5} 濃度高低，觀察 PM_{2.5} 濃度在時間與空間中的變化。圖 5 與圖 6 為空間趨勢分析的

最終結果，就時間上來說，所有網格的迴歸係數值均為負值，故我們可以得知 PM_{2.5} 濃度與時間呈負相關，且都達到顯著呈度 ($p < 0.05$)，這代表隨著年份增加，臺中各地的 PM_{2.5} 濃度都呈現顯著的下降趨勢。從空間上看，則可以發現臺中市西南區的迴歸係數遠低於其他地區，這代表該地區的 PM_{2.5} 濃度改善最多。

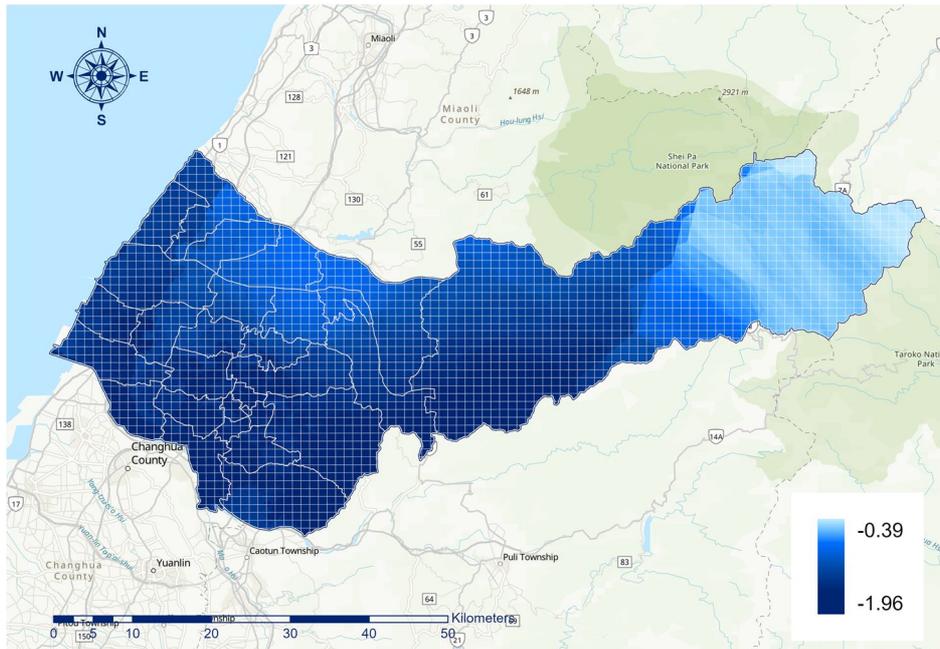


圖 5 臺中市空間模型係數趨勢分析圖。圖中所有網格之係數均小於 1，代表隨著年份增加，臺中各地的 PM_{2.5} 濃度都呈現下降趨勢

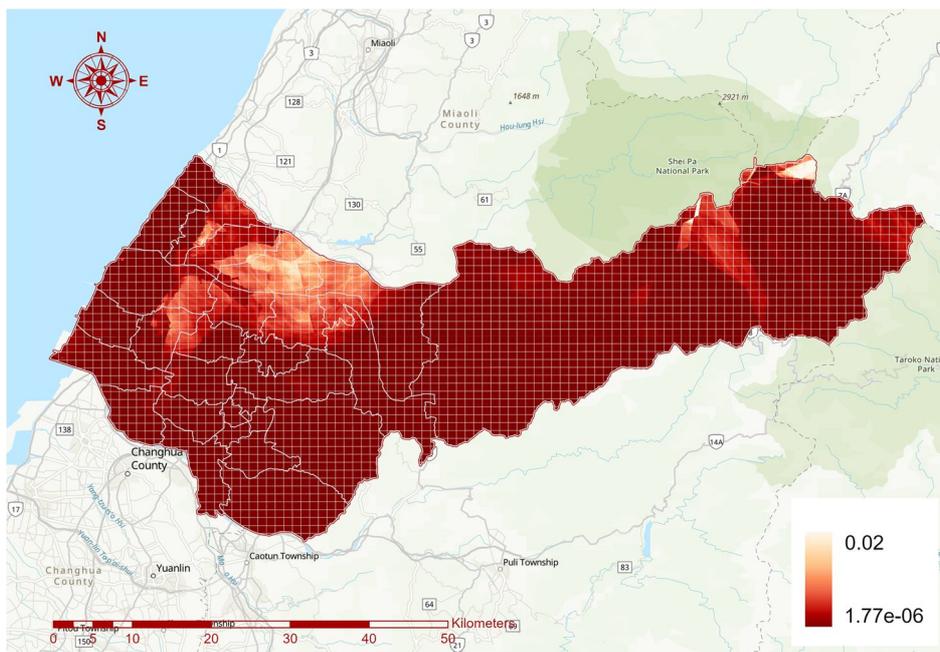


圖 6 臺中市 PM_{2.5} 空間模型 p 值趨勢分析圖。圖中所有網格之 $p < 0.05$ ，代表均達到統計上的顯著程度

表 2 機械學習混合空間模型結果

	RF	GBM	XGBoost	LightGBM	CatBoost	
訓練資料 (80%)	Adjusted R ²	0.93	0.92	0.93	0.86	0.87
	MSE (µg/m ³)	23.60	25.81	23.77	50.36	44.23
	RMSE (µg/m ³)	4.86	5.08	4.88	7.10	6.65
	MAE (µg/m ³)	3.19	3.35	3.45	4.69	4.55
	time(s)	42.84	186.52	27.57	49.00	166.03
	測試資料 (20%)	Adjusted R ²	0.87	0.88	0.88	0.86
MSE (µg/m ³)		43.31	41.13	41.09	49.79	43.95
RMSE (µg/m ³)		6.58	6.41	6.41	7.06	6.63
MAE (µg/m ³)		4.51	4.33	4.35	4.70	4.52
R ²		0.86	0.87	0.87	0.85	0.86
十折驗證		Adjusted R ²	0.86	0.87	0.87	0.85
	MSE (µg/m ³)	45.89	42.99	43.12	52.60	46.00
	RMSE (µg/m ³)	6.77	6.56	6.57	7.25	6.78
	MAE (µg/m ³)	4.63	4.41	4.44	4.80	4.61
	R ²	0.89	0.88	0.89	0.87	0.89
	外部驗證	Adjusted R ²	0.89	0.88	0.89	0.87
MSE (µg/m ³)		9.97	11.02	10.10	13.13	11.00
RMSE (µg/m ³)		3.16	3.32	3.18	3.62	3.32
MAE (µg/m ³)		2.39	2.56	2.41	2.82	2.60

表 3 模型效能排序

	Adjusted R ²	RMSE (µg/m ³)	Ranking
RF	0.93	4.86	1
GBM	0.92	5.08	3
XGBoost	0.93	4.88	2
LightBoost	0.86	7.10	5
CatBoost	0.87	6.65	4

4. 討論

本研究進一步將趨勢分析圖與污染排放源相關地標重疊，以觀察兩者之間的關係，其中污染排放源包含本地污染源，即中式餐廳與寺廟 (圖 7 (a))；工業相關地標，即火力發電廠與製造業工廠 (圖 7 (b))；和交通相關污染源，即道路分布 (圖 7 (c))。各自重疊後可以觀察到這些地標分布密集的地區與係數高的區域相匹配，此發現與模型數據中的高 Adjusted R² 和低 RMSE 結果相同，代表本研究開發模型具有高效能與高可信度。

本研究針對臺中市過去 25 年 PM_{2.5} 污染的空間模式和時間趨勢進行深入分析，根據(圖 4)

1994 至 2020 年臺中市 PM_{2.5} 觀測濃度折線圖，2006 年為 PM_{2.5} 濃度減少最多的年度，這可能代表地方政府在空氣污染整治方面做出了一些努力，根據李立德等 (2007)，自 2003 年起，臺中市針對區域特性，增列執行了許多專案性及研究性計畫，包括餐飲業 (2003、2005、2006 年)、露天燃燒 (2003、2004 年)、紙錢集中燃燒 (2005、2006 年)、臺中工業區及中科周圍空氣調查(2004、2006 年)、空氣中重金屬之污染調查 (2004 年)、因應京都議定書之規劃 (2006 年) 等，對改善臺中市的空氣品質有一定的幫助。上述發現強調了有效的環境法規和管理策略在減輕空氣污染方面的重要性。該研究成果可以讓公眾更好地了解臺中市空氣污染分布，亦能提供一種系統化檢查隨時間變化的環境的方法，進而可以幫助地方政府相關單位了解當地空氣品質的變化情況，並制定針對性的政策作為改善空氣品質的利具。未來可以透過納入有關排放源和氣象因素等更詳細的數據來擴展研究規模，以更好地了解導致該地區空氣污染的潛在機制，盼能更加有效地改善全球空氣品質。

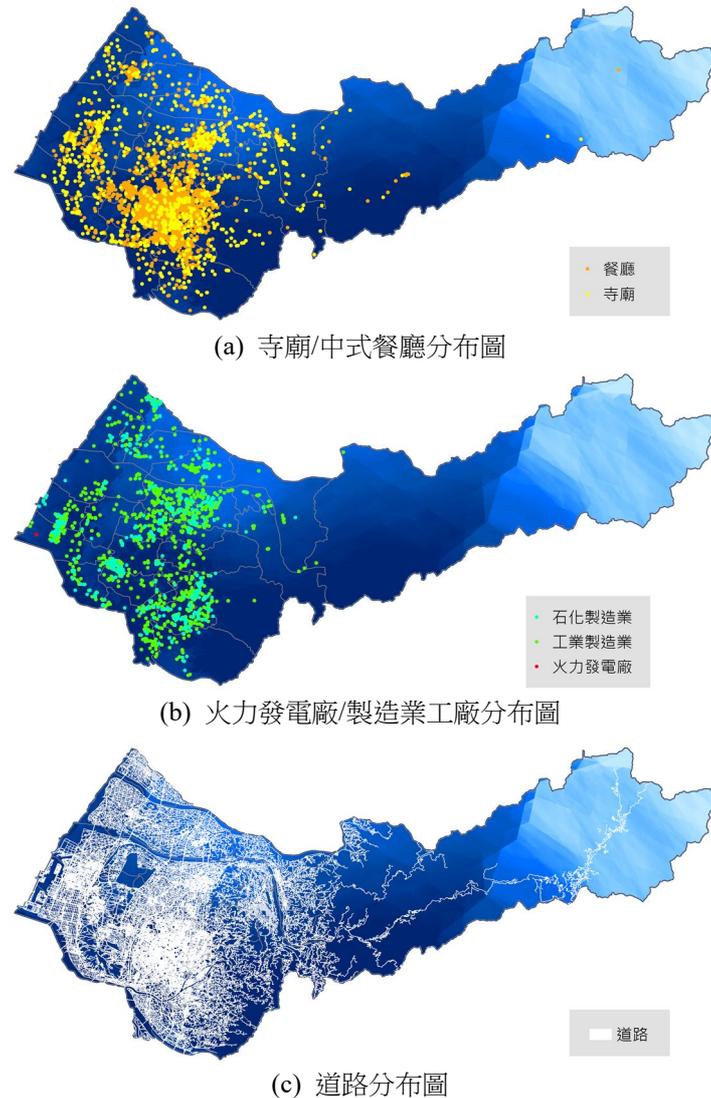


圖 7 臺中市空氣污染排放源分布圖

5. 結論

這項研究的結果凸顯了空氣污染監測和環境整治政策的重要性。雖然自 2006 年以來觀察到的 PM_{2.5} 濃度下降令人鼓舞，這可能歸功於政府控制工業排放和提高燃料品質的政策。但是，目前臺中市的 PM_{2.5} 濃度仍超過世界衛生組織 10 $\mu\text{g}/\text{m}^3$ 的年度建議值，表示我們需要持續努力減少排放和改善空氣品質。本研究產生的空間推估圖可以幫助識別高風險區域並促進有針對性的干預措施以減少空氣污染，但可以進一步討論的是，本研究仍有其局限性，例如使用的監測站數據有限及缺乏排放源時空變化的資訊。針對上述研究限制，未來可以透過整合更多資料來源，如空氣汙染排放清冊，並應用

更先進的建模技術來減少相關研究限制的影響，以便更好地了解空氣污染的潛在機制。

參考文獻

- 李立德、白珏瑛、顏淑琪、苗宛陶、施桂英、郭怡婷，2007。國內外空氣污染管制工作現況與臺中市未來展望，台灣環境資源永續發展研討會，pp.3-336-3-347。[Li, L.T, Pai, C.Y., Yan, S.C., Miao, W.T., Shi, G.Y., and Kuo, Y.T., 2007. Guo nei wai kung chi wu ran guan jr gung tzuo shian kuang yu tai jung shr wei lai jan wang, The Seminar Environmental Resources Sustainable Development, pp.3-336-3-347. (in Chinese)]

- Anderson, G.B., Krall, J.R., Peng, R.D., and Bell, M.L., 2012. Is the relation between ozone and mortality confounded by chemical components of particulate matter? Analysis of 7 components in 57 US communities, *American Journal of Epidemiology*, 176(8): 726-732, DOI: 10.1093/aje/kws188.
- Breiman, L., 2001. Random forests, *Machine Learning*, 45: 5–32, DOI: 10.1023/A:1010933404324.
- Chen, T., and Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp.785-794, DOI: 10.1145/2939672.2939785.
- Chen, T.H., Hsu, Y.C., Zeng, Y.T., Lung, S.C.C., Su, H.J., Chao, H.J., and Wu, C.D., 2020. A hybrid kriging/land-use regression model with Asian culture-specific sources to assess NO₂ spatial-temporal variations, *Environmental Pollution*, 259: 113875, DOI: 10.1016/j.envpol.2019.113875.
- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine, *The Annals of Statistics*, 29(5): 1189–1232, DOI: 10.1214/aos/1013203451.
- 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.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems (NIPS 2017)*, Vol.30: pp.3146–3154, Long Beach, CA, USA.
- Lepeule, J., Laden, F., Dockery, D., and Schwartz, J., 2012. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard six cities study from 1974 to 2009, *Environmental Health Perspectives*, 120(7): 965-970, DOI: 10.1289/ehp.1104660.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., and Gulin, A., 2018. CatBoost: Unbiased boosting with categorical features, *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, Montréal, Canada, DOI: 10.48550/arXiv.1706.09516.
- Michanowicz, D.R., Shmool, J.L., Cambal, L., Tunno, B.J., Gillooly, S., Hunt, M.J.O., Tripathy, S., Shields, K.N., and Clougherty, J.E., 2016. A hybrid land use regression/line–source dispersion model for predicting intra–urban NO₂, *Transportation Research Part D: Transport and Environment*, 43:181-191, DOI: 10.1016/j.trd.2015.12.007.
- Ren, C., Williams, G.M., and Tong, S., 2006. Does particulate matter modify the association between temperature and cardiorespiratory diseases?, *Environmental Health Perspectives*, 114(11): 1690-1696, DOI: 10.1289/ehp.9266.
- 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.
- Wu, C.D., Zeng, Y.T., and Lung, S.C.C., 2018. A hybrid kriging/land-use regression model to assess PM_{2.5} spatial-temporal variability, *Science of The Total Environment*, 645: 1456-1464, DOI: 10.1016/j.scitotenv.2018.07.073.

The Long-term Trend Analysis of PM_{2.5} Variability From 1994 to 2020 Using a Hybrid Spatial Model: A Case Study of Taichung City, Taiwan

Yu-Ju Lin¹ Chia-Wei Hsu² Chia-Ho Lee³ Yu-Ting Zeng⁴ Pei-Yi Wong⁵
Pau-Chung Chen^{6,7,8,9} Yu-Cheng Chen^{10,11,12} Chih-Da Wu^{13,14,15,16*}

Abstract

This study takes Taichung City as an example and aims to investigate the long-term impact of urban development on air pollution. By establishing a spatial model, we estimate the concentration trends of fine particulate matter (Particulate Matter 2.5, PM_{2.5}) over the past 25 years (from 1994 to 2020) and further assess the influence of urban development on air quality. Various databases were utilized as sources of spatial predictor variables, including the Environmental Resources Database, meteorological database, land-use inventory, landmark database, digital road network map, digital terrain model, MODIS Normalized Difference Vegetation Index (NDVI) database, and power plant distribution database. The spatial hybrid model in this study combines Hybrid Kriging/Land-Use Regression and machine learning methods. Initially, important predictor variables were determined using traditional Land-Use Regression (LUR) and Hybrid Kriging-LUR. Subsequently, prediction models based on the selected variables from LUR models were fitted using Random Forest (RF), Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (Light GBM), and CatBoost algorithms. Validation methods such as data splitting, 10-fold cross-validation, and external data verification were employed to confirm the robustness of the developed models. The results indicate that the model is stable and reliable, with an Adj-R² of 0.93. Through linear regression, it was observed that the estimated values of the predictor variable 'year' for most locations in the city are negative, indicating a significant improvement in air pollution over the past 25 years. This study emphasizes the importance of managing and controlling air pollution in urban development planning.

Keywords: PM_{2.5}, Urban Development, Hybrid Spatial Model, Trend Analysis

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

² Ph.D Student, Department of Geomatics, National Cheng Kung University

³ Assistant Researcher, Department of Geomatics, National Cheng Kung University

⁴ Director General, Environment Protection Bureau Chiayi City

⁵ Ph.D. Student, Department of Environmental and Occupational Health, National Cheng Kung University

⁶ Distinguished Investigator and Director, National Institute of Environmental Health Sciences, National Health Research Institutes

⁷ Associate Investigator, National Institute of Environmental Health Sciences, National Health Research Institutes

⁸ Doctor, Department of Environmental and Occupational Medicine, National Taiwan University Hospital

⁹ Distinguished Professor, College of Public Health, National Taiwan University

¹⁰ Distinguished Professor, Department of Public Health, National Taiwan University College of Public Health

¹¹ Adjunct Assistant Professor, Department of Safety, Health and Environment Engineering, National United University

¹² Adjunct Assistant Professor, Department of Occupational Safety and Health, China Medical University

¹³ Professor, Department of Geomatics, National Cheng Kung University

¹⁴ Joint Associate Investigator, National Institute of Environmental Health Sciences, National Health Research Institutes

¹⁵ Professor, Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University

¹⁶ Joint Investigator, Research Center for Precision Environmental Medicine, Kaohsiung Medical University

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

Received Date: Feb. 27, 2024

Revised Date: May. 09, 2024

Accepted Date: May. 22, 2024