

基於土地利用迴歸之機器學型模型分析新冠肺炎 三級警戒政策對臺灣工業城市細懸浮微粒之影響

蘇均珺¹ 翁佩詒² 曾于庭³ 李佳禾⁴ 吳治達^{5,6,7,8*}

摘要

COVID-19 疫情對全球帶來巨大衝擊，臺灣政府於 2021 年 5 月 19 日宣布三級警戒，限制民眾活動。本研究旨在評估警戒期間臺灣工業城市 PM_{2.5} 濃度之變化。本研究以代表性的工業城市高雄市為研究區域，蒐集 1994 至 2020 年的空污觀測數據和地理變量，利用土地利用迴歸和逐步變量選擇建立模型、選取重要變數，再使用不同機器學習演算法建立模型，其中結果以 Random Forest (RF) 演算法的模型表現最佳，R² 達 0.95；推估成果顯示封鎖期間空氣品質改善，高雄市平均 PM_{2.5} 濃度為 18.1 μg/m³，低於警戒前 19.9 μg/m³。Paired t-test 結果顯示差異達到統計顯著水準 (p 值<0.001)，各土地利用區域 (居住區、工業區、街道和綠地) 亦呈現一致結果。

關鍵詞：空氣污染、細懸浮微粒物 (PM_{2.5})、基於土地利用的機器學習模型、COVID-19、三級警戒

1. 前言

新冠肺炎 (COVID-19) 疫情對全球人類的生活帶來了前所未有的巨大影響，讓各個國家和地區都感到措手不及。病毒的傳播速度之快以及其高度傳染性使得政府、醫療機構和社會各個層面都面臨了重大挑戰。人們的日常生活、工作、旅行和社交習慣都發生了巨大的變化，許多國家實施了緊急應對措施，如封鎖、社交距離和口罩規定，以遏制病毒的傳播。而為了應對這種情況，臺灣政府於 2021 年 5 月 19 日發布了三級警戒，限制民眾外出和工作，迫使多的公司和工廠停止生產運作。有鑑於空氣汙染對於人體健康會造成相當顯著之危害，其中直徑小於 2.5 微米 (micrometer, μm) 的懸浮微粒

(particulate matter, PM) PM_{2.5}，是環境中對人體健康影響最大的物質之一，不僅會造成及加劇心血管與呼吸道疾病，更被國際癌症研究署 (International Agency for Research on Cancer, IARC) 列為第一級致癌物。

從上世紀至今日，不斷有各類研究在探討懸浮微粒對於健康的影響。美國一項為期 7 年的研究表明，PM_{2.5} 每減少 10 μg/m³，平均壽命就會延長 0.35 年 (Correia *et al.*, 2012)；在歐盟國家，PM_{2.5} 使平均壽命縮短了 8.6 個月 (Orru *et al.*, 2011)；在中國的追蹤研究中，PM_{2.5} 每增加 10 mg/m³，與心肌缺血症、出血性中風和缺血性中風的死亡風險分別增加了 9.7%、4.4%和 13.5% (Yin *et al.*, 2015)。因此評估 PM_{2.5} 在空間中的濃度與分布至關重要。而由於空氣

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 國立成功大學測量及空間資訊學系 助理研究員

⁴ 嘉義市政府環境保護局 局長

⁵ 國立成功大學測量及空間資訊學系 教授

⁶ 國家衛生研究院國家環境醫學研究所 兼任研究員

⁷ 國立中興大學永續農業創新發展中心 兼任教授

⁸ 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 04 月 12 日

接受日期：民國 113 年 06 月 24 日

汙染主要源自於人類的各種活動，我們預期在封鎖期間的空氣品質會較於先前有所改善。

本研究旨在利用基於土地利用迴歸之機器學習模型，評估封鎖政策對臺灣工業城市 PM_{2.5} 濃度的影響。目前關於 COVID-19 與空氣汙染之間關係的研究主要集中在醫學方面，探討空氣汙染與 COVID-19 感染風險或死亡率之間的關聯。對於政策是否會影響 COVID-19 期間空氣中污染物的濃度和分佈，國際間已有部分研究，例如中國的研究發現，封鎖城市的空氣品質指標 (Air quality index, AQI) 相較於對照組下降了 19.84 $\mu\text{g}/\text{m}^3$ (PM_{2.5} 下降了 14.07 $\mu\text{g}/\text{m}^3$) (He *et al.*, 2020)；美國的研究顯示，NO₂ 顯著下降，PM_{2.5} 在 COVID-19 期間也有所降低，並且在城市縣和早期實施非必要業務關閉的州的縣中的減少是統計顯著的 (Berman & Ebisu, 2020)；印度的研究表明，PM_{2.5}、NO₂ 和 SO₂ 的平均質量濃度顯著下降，尤其是在封鎖的第一階段 (主要是 NO₂)，空氣污染物的減少更為顯著 (Pratap *et al.*, 2021)，然而國內相關研究仍然不足。本研究旨在填補這一空白，探討三級警戒對地區空氣汙染的影響，為未來的醫學研究提供有價值的洞察。

過往已發展出許多推估二維空間大範圍空氣汙染的方法，例如土地利用迴歸 (Land Use Regression, LUR) 模型，在環境研究中被廣泛應用於預測空氣污染物的濃度，是一種基於多變量線性回歸的實證建模方法，它將污染監測資料與由地理資訊系統 (Geographic Information System, GIS) 中所獲得的空間變量相結合，最終產製出一個線性推估模型 (Bitta *et al.*, 2018)。過去有許多應用 LUR 模型的研究，透過地理資訊系統以及遙感探測等空間資訊技術獲取測站周邊的土地利用資訊，進而建立 LUR 模型，並得到穩定的模型結果來推估國內高屏空品區 PM_{2.5} 的時空分布 (曾芷琳等, 2019) 以及北部空品區 PM_{2.5} 的時空分布 (曾于庭等, 2018)。此外，還有研究結合克里金空間內插 (Kriging Interpolation) 發展出混合模型來推估臨海石化工業區 PM₁₀ 的濃度 (吳昭儀等, 2020)。總結來說，LUR 模型結合 GIS 和遙感探測技術，已在臺

灣的多個地區和不同空氣污染物的時空分布推估中展現了其穩定性和可靠性。

然而，基於 LUR 的機器學習 (Machine Learning, ML) 模型近年來嶄露頭角，成為一種具有潛力的替代方法，因為它們具有更低的誤差和更強的解釋能力。與 LUR 模型相比，基於 LUR 的機器學習模型可以融入更廣泛的預測變量和非線性關係，從而能夠提供更準確和細緻的預測。此外，基於 LUR 的機器學習模型在處理缺失資料和異常值方面更為有效 (Basu *et al.*, 2019)。過去便有結合 LUR 與極限梯度提升演算法建立推估模型以推估高雄都會區二氧化氮，且模型具有高度預測與解釋力 (翁佩詒等, 2021)。這些優勢使得基於 LUR 的機器學習模型成為環境研究領域的一項有前景的進展。

本研究收集了 1994 年至 2020 年的每日 PM_{2.5} 觀測數據和地理變量，用於模型的建置。地理變量資料主要基於航遙測技術產製，包括土地利用資料、NDVI、DEM 和 DTM 等。其中，土地利用資料的產製其除了國土利用調查的現場勘查外，亦利用航空測量攝影影像、遙感探測影像、地籍圖、地形圖及電子地圖等圖資，並使用影像判釋進行土地覆蓋調查，從而掌握現況地景環境的分布及規模，包括植被、水體、建物及裸露地等，全面了解土地的使用樣態。這些豐富的地理變量資料經過逐步變量選擇程序 (Stepwise) 結合土地利用迴歸 (LUR) 模型，確定了影響 PM_{2.5} 變異性的重要因素。本研究所建置之模型基於不同的機器學習演算法，包括梯度提升機 (Gradient Boosting Machine, GBM)、隨機森林 (Random Forest, RF) 和 CatBoost (Categorical Boosting) 等等。最終的模型推估了高雄市所有鄉鎮的 PM_{2.5} 濃度，藉此研究分析了三級警戒前後 PM_{2.5} 濃度的變化，以及不同土地利用類型之間的差異。透過本研究的成果，不僅實現了對 PM_{2.5} 濃度的準確預測，還展示了航遙測技術在環境監測和公共衛生研究中的巨大潛力。這一貢獻不僅擴展了航遙測資料在公共衛生等領域的應用廣度，還為航遙測技術在不同領域中的廣泛應用提供了新的視角和方法，促進了其在公共衛生、環境保護和城市規劃等方面的發展。

2. 材料及方法

研究流程基本上可以分為五個部分：資料庫建置、篩選重要變數、建立推估模型、模型驗證以及最後的 PM_{2.5} 濃度分布推估與分析 (圖 1)。

2.1 研究地區

高雄市位於臺灣的西南邊，是一個擁有大量石化產業的城市，在其都市地區內具有相當稠密的人口與可觀的交通流量，工商業活動蓬勃發展；此外，高雄市的沿海地區坐擁著臺灣最大且最繁忙的貨運港口。此港口不僅是臺灣主要的貿易門戶，還扮演著國際貨物運輸和物流的關鍵角色，它承載了大量的進出口貨物，促進了該地區的經濟活動。在這一帶，工業區亦繁忙發展，擁有各種規模的製造和生產設施。這些工業區域涵蓋了多個產業領域，包括石化、製造、化學和重工業等，為當地提供了大量的就業機會，但也伴隨著相應的環境挑戰，如廢氣排放和廢水處理等。再者，這個地區還擁有一座重要的火力發電廠，為高雄市及其周邊地區提供電力供應。然而，這種發電方式通常伴隨著空氣污染風險，因為它釋放出二氧化硫、氮氧化物和微粒物質等污染物質，對當地空氣質量造成了一定的不利影響。綜合以上各項區域因素，對高雄市的空氣品質造成了複雜的影響，也成為本研究的重要研究對象。

2.2 研究資料

本研究從以下四個資料庫蒐集 1994 年至 2020 年之資料：

(1) 環保署空氣污染監測數據庫

行政院環境保護署在全臺灣，包括離島地區，共設有 77 個空氣品質監測站，這些監測站的功能包括數據收集，資料涵蓋了多種空氣污染物，如細懸浮微粒、氣體污染物，以及周圍的氣象因素，這些數據的時間尺度可達每小時的平均值。本研究蒐集研究試區測站之 PM_{2.5} 監測值，作為本研究之依變量。

(2) 土地利用數據庫

本研究使用之土地利用資料庫來自網路及本系實驗室提供之資料，包含：

- A.內政部國土測繪中心之國土利用調查資料
- B.經濟部工業局工業區範圍資訊 (含工業區、加工出口區、科學園區等)
- C.勤崴國際科技股份有限公司之地標資料庫 (如寺廟、中式餐飲等)
- D.Google Earth/Google Map 地標資料庫 (如全臺火力發電廠位置資訊)
- E.行政院環境保護署全臺焚化爐煙囪分布資訊
- F.內政部與交通部共同發行之臺灣地區交通路網數值圖

其中國土利用調查資料之土地使用分類系統為三層式之樹狀結構類。本研究參考其中之工商 (人工構造物、工業、商業建築)、水利 (水體、水利設施等)、森林 (含農業)、交通 (交通道路、鐵路、機場等) 等四大類型之土地利用位置資訊；幾項地標資料庫則被用以獲取各類地標與特徵土地利用點位之分布及位置資訊，包含透過工業區資料庫中之工業區、加工出口區、科學園區；勤崴資料庫中之寺廟、中式餐飲、Google Earth/Google Map 資料庫之火力發電廠，以及環保署全臺焚化爐煙囪分布資訊。路網數值資料則係使用 2020 年交通部運輸研究所全島 1/5000 比例尺之 GIS 路網數值圖。該資料將全臺之道路分為國道、省道、縣道以及鄉道等 8 種類型。本研究透過該資料庫再將其合併為三大類，包含主要道路 (國道、省道、縣道 以及市區快速道路)、地方道路 (市區道路、鄉道、產業道路以及無路名) 和所有道路。上述土地利用類型在計算後，將做為後續空污建模之解釋變數。本研究使用之土地利用空間變數如表 1 所示。

(3) MODIS NDVI 圖像數據庫

本研究使用 NASA (National Aeronautics and Space Administration) 所拍攝，並由美國地質調查所 (The U.S. Geological Survey, USGS) 校驗計算後提供之常態化植生差異指標 (Normalized Difference Vegetation Index, NDVI)，以代表植生綠度分布情形，並自其 Earth Data 網站取得監測日期下研究試區 NDVI 數值，搭配國土利用調查資訊中的森林、綠

地等土地利用，代表研究試區之植生狀態及指標，作為模擬懸浮微粒濃度變異之解釋因子之一。

(4) 氣象數據庫

交通部中央氣象署 (Central Weather Administration, CWA) 是臺灣的地球科學 (含氣象、氣候、海象、地震、天文) 專責機構，負責上述項目之觀測及預報業務。除了向公眾發布氣象預報、

海象測報與地震報告之外，也是災害性天氣預報 (颱風警報、豪雨特報、強風特報、低溫特報等) 發布、以及中華民國《氣象法》法定氣象事業督導之機關。本研究配合研究時期，將自大氣水文資料庫平台下載對應時間之平均降雨量、溫度、濕度、風向以及風速等氣象資料，供後續模式之用。

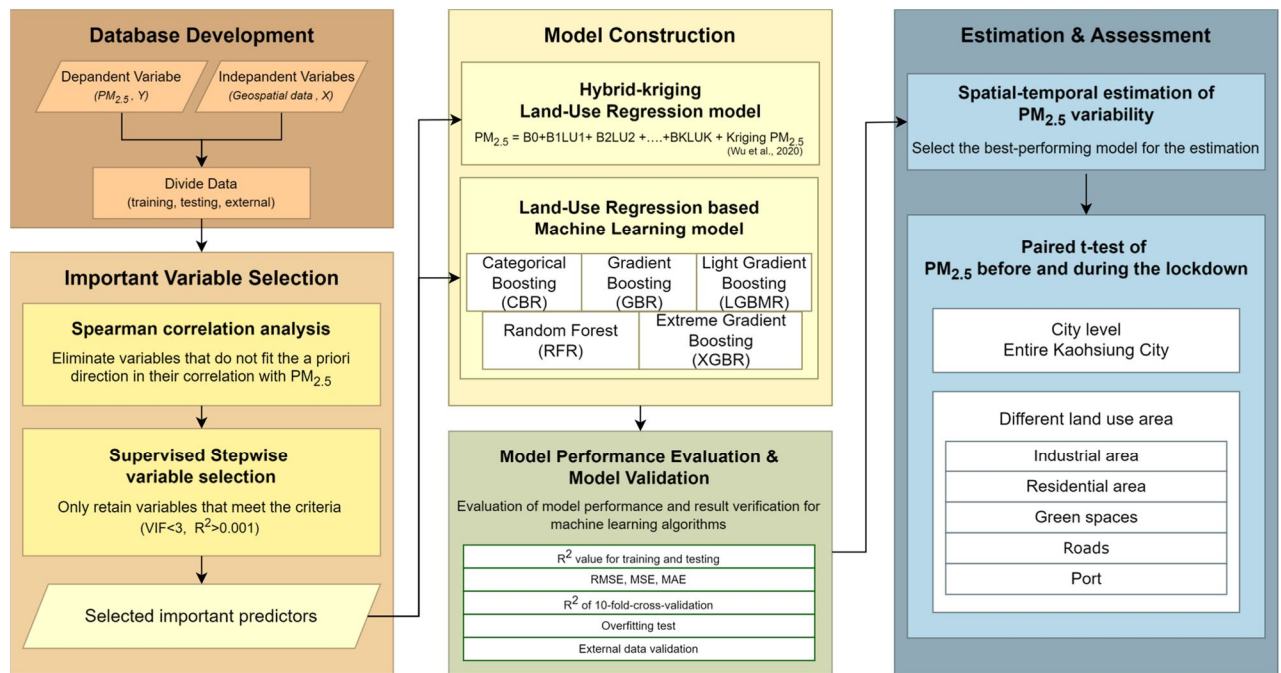


圖 1 研究流程圖

表 1 土地利用空間變數

資料來源	變數	資料描述	圖資型態	單位
交通部運輸研究所 路網數值圖	當地道路	鄉村、城市、工業和無路名道路	體積、比例	$m^3, \%$
	主要道路	國道、省道、縣道、城市高速公路	體積、比例	$m^3, \%$
	全部道路	當地道路+主要道路	體積、比例	$m^3, \%$
經濟部工業局 工業區資料庫	工業園區	科學園區、加工出口區及工業園區	體積、比例	$m^3, \%$
	純住宅	-	體積、比例	$m^3, \%$
	混合住宅	商業住宅+工業住宅	體積、比例	$m^3, \%$
內政部國土測繪中心 國土利用調查資料	所有住宅	純住宅地+混合住宅地	體積、比例	$m^3, \%$
	稻田	-	體積、比例	$m^3, \%$
	旱田	-	體積、比例	$m^3, \%$
	果樹	-	體積、比例	$m^3, \%$
	水體	-	體積、比例	$m^3, \%$
	公園綠地廣場	-	體積、比例	$m^3, \%$
	森林用地	-	體積、比例	$m^3, \%$
	工業	所有製造業	體積、比例	$m^3, \%$
	鐵路	-	體積、比例	$m^3, \%$
	地標資料庫	寺廟	-	點資料
中式餐廳		中式餐廳和夜市	點資料	count

2.3 使用機器學習和土地利用數據處理開發和驗證 PM_{2.5} 推估模型

2.3.1 土地利用資料前處理

為獲取研究區內土地利用空間資訊，本研究參考了先前研究的估算和模擬案例 (Wong *et al.*, 2021)，使用 ArcGIS Pro 2.6 軟體將研究區域內的各種土地利用排放源和空間變數轉換成 50 m×50 m 的網格數據；並計算了兩種空間分布資料：每個網格中心點從 25 到 5000 m 半徑的緩衝區，以及每個網格中心點到研究區域內每個土地利用空間變數的歐幾里德距離，以此獲得研究區域內土地利用排放變數的空間密度分布。最後在每個監測站點處提取了點位所在之土地利用密度值，並將其添加到資料庫中，以供後續變數選擇和機器學習推估使用。

2.3.2 土地利用迴歸模型建置

本研究使用傳統的土地利用迴歸和克利金/土地利用迴歸混合模型 (Hybrid Kriging/Land-use Regression) 來建立初步模型，並選擇重要的空間解釋變數來預測 PM_{2.5} 濃度。土地利用迴歸的本質為多元線性迴歸，在此步驟基於測站之濃度資訊為依變數、各項土地利用空間變數為解釋變數進行線性模型建置，並且搭配逐步變數篩選 (Stepwise Variable Selection) 方法，以 p 值小於 0.1、大於 0.3 作為變數納入以及移除之標準，同時於模型建置過程中判斷選入之變數與預測對象之相關性方向是否正確 (如人工覆蓋物比例與空氣污染微粒應呈正相關；環境綠蔽度、植被涵蓋比則應與空氣污染呈現負相關)。此外，為避免共線性造成模型參數估計上之誤差，故同時利用變異數膨脹係數 (Variance Inflation Factor, VIF) 為標準，排除 $VIF > 3$ 之變數，以完成重要土地利用變數篩選以及土地利用迴歸模型建立。

其中克利金/土地利用迴歸混合模型的原理在於結合 Ordinary Kriging 空間內插法和土地利用迴歸的優勢，以提高模型對濃度的準確估算。而混合

模型的核心思想是，即使在沒有真正濃度值的情況下，如果在相同的時間點附近有監測資料的點位，那麼這些附近點位的監測濃度值與欲評估地點的濃度值之間必定存在某種關聯性。當兩地之間的距離越近，這種關聯性就越能代表未監測地區空氣污染程度，這比周圍的土地利用和污染源更具代表性。因此，將周邊點位的濃度信息納入土地利用迴歸模型中，可以提高模型的解釋能力 (Lung *et al.*, 2004)。基於這一理念，本研究通過 Leave-one-out Kriging 方法估算了每個未知濃度點位的 Kriging PM_{2.5} 污染估算值，並將其納入統計模型中作為解釋變數，其餘步驟皆與土地利用迴歸模型相似，並於最終建立混合式推估模型。

2.3.3 整合機器學習之推估模型

為了降低資料的維度並確保所選的變數在統計上具有意義，本研究以模型篩選出的變數為基礎，運用 Python 中的 AutoML 套件，精心挑選了五種機器學習算法，包括隨機森林、梯度提升、極限梯度提升 (Extreme Gradient Boosting, XGBoost)、Light Gradient Boosting Machine (Light GBM) 和 CatBoost 等，進行模型超參數的自動調整，並打造了不同機器學習算法為基礎的混合式土地利用集成學習模型。

每個機器學習演算法都擁有其獨特的優勢，隨機森林以同時建立多個迴歸樹來提升計算效能，確保模型的穩定性和可靠性。XGBoost 進一步精煉了模型，降低了過度擬和的風險，同時兼具並行建樹和串行建樹的特性，使得預測結果更加優化。而 LightGBM 則在 XGBoost 的基礎上進行了進一步的優化，特別在快取速度方面取得了巨大的改進，有助於縮短模型訓練的時間。這些演算法的選用，使研究能夠更好地滿足各種不同的建模需求，並提高了模型的預測性能。

至此，包含以統計為基礎之：(1) 土地利用迴歸、(2) 克利金/土地利用迴歸混合模型、機械學習為基礎之(3) 隨機森林、(4) 梯度提升、(5) 極限梯度提升、(6)Light Gradient Boosting Machine 及 (7) CatBoost，共使用七個模型來推估研究區之空氣污

染分佈。本研究以利金/土地利用迴歸混合模型為基礎，結合隨機森林、梯度提升、極限梯度提升、Light Gradient Boosting Machine 及 CatBoost 五種機器學習演算法，比較五種方法之預測成果。

2.3.4 模型驗證

在模型表現能力之驗證方面，本研究透過數種方法以驗證模型穩定度 (Robustness) 以及模型是否具有過擬合 (Overfitting) 之問題，包含：

(1) Data Split 驗證法

將總樣本資料的 80% 作為訓練樣本進行模型建立，並以其餘 20% 樣本作為測試樣本，並將測試樣本帶入模型當中與模擬之結果進行比較。

(2) 十折交叉驗證法 (10-Fold Cross Validation)

每次以總樣本數中 90% 資料做為訓練樣本，扣除訓練樣本後剩餘之 10% 資料做為測試樣本，整個過程重覆十次，以訓練樣本建立模型後，據以進行預測試樣本之預測，並計算其與測試樣本真實值之相關性。

(3) 外部資料驗證 (External Data Validation)

預先進行一次採樣，其結果不納入建模，利用其餘資料建立模型進行預測後，與未納入模型之資料 (真實值) 比較之。

至於在模型表現能力之評估指標方面，本研究將採用 R^2 與 Adjusted R^2 代表模型的準確度；平均絕對誤差 (Mean Absolute Error, MAE)、均方誤差 (Mean Square Error, MSE) 及均方根誤差 (Root Mean Square Error, RMSE) 用以代表模型誤差。

2.3.5 $PM_{2.5}$ 濃度分布推估

這項研究的主要目標是通過製作並比較兩張高雄市 $PM_{2.5}$ 濃度地圖，深入瞭解高雄市政府宣布實施三級警戒前後 (分別為 2021 年 5 月 1 日至 19 日和 2021 年 5 月 20 日至 31 日) 的 $PM_{2.5}$ 之濃度及其於空間中分佈狀況。研究中特別關注不同土地利用區域 (如圖 3 所示) 的濃度分佈情形，並使用配對 t 檢定以確定在不同類型的土地利用地區是否存在明顯的濃度變化趨勢。研究成果不僅有助於更了解環境改善和城市政策實施的效果，還具有重要的

政策參考價值，可以為改善城市空氣質量、制定相關政策和採取措施提供實際指導，特別是在應對環境變遷挑戰並提高城市居民的生活品質方面。

3. 結果

3.1 重要變數

表 2 顯示了從傳統土地利用迴歸與逐步迴歸分析中所獲得的 8 個模型預測變數。在此過程中，本研究對每項加入模型之變數係數和 R^2 改變量進行了檢查，以確定它們與 $PM_{2.5}$ 濃度之間的關係是否顯著並合理。結果顯示，水體 1500 m (緩衝區) 和天然/人工森林 750 m (緩衝區) 的變數與 $PM_{2.5}$ 濃度呈負相關，這表示當周圍有水體或天然/人工森林的情況越多時， $PM_{2.5}$ 濃度通常越低；而其他變數則與 $PM_{2.5}$ 濃度呈正相關，即當這些變數的值越高時，則通常伴隨著較高的 $PM_{2.5}$ 濃度。這可能是由於水體和森林有助於 $PM_{2.5}$ 濃度的分散和減少；另一方面，其他變數，如鐵路、港口和其他土地利用等區域，對 $PM_{2.5}$ 濃度的正向影響較為顯著，這意味著這些區域可能更容易產生或累積污染物，導致較高的 $PM_{2.5}$ 濃度。

表 2 顯示所列所有變數之 p 值皆小於 0.05，表明這些變數在推估模型中具有統計學上的顯著性；此外，每個變數的 R^2 變化值皆大於 0.001，顯示它們在解釋 $PM_{2.5}$ 濃度變異性方面都有顯著的貢獻。這些結果的統計顯著性進一步強調了所選變數的重要性，並確保了我們建立的模型的可信度。

表 2 土地利用逐步迴歸變數篩選

變數	coefficient	p-value	R^2 change
截距	41.331	2.63E-12	
$PM_{2.5}$ Kriging	0.967	0	
鐵路 150m	0.016	8.53E-142	0.007
商港與漁港 2000m	0.007	1.20E-67	0.009
水體 1500m	-0.004	1.34E-141	0.005
寺廟 1250m	4276490.417	0	0.003
兼工業使用住宅 1250m	0.015	0	0.004
天然/人工林 750m	-0.031	2.15E-111	0.002
果樹 1500m	0.009	2.29E-62	0.001

3.2 土地利用迴歸模型表現

表 3 展示了土地利用迴歸模型結合克利金空間內插後在預測 PM_{2.5} 濃度方面的表現。該模型呈現了相當不錯的預測能力，其 R² 值達到 0.91，顯示其在估算 PM_{2.5} 濃度方面的有效性。此外，我們也可以看到該模型的表現令人滿意，這可以從 MAE、MSE 和 RMSE 的數值中看出，分別為 3.10、33.30 和 5.77。測試數據的 R² 值與訓練數據相近，這說明了該模型並未出現過度擬合的情況。而模型的外部驗證結果則獲得了 0.84 的 R² 值，進一步確保了模型的穩健性。綜合來看，這些結果強烈表明了該模型在預測 PM_{2.5} 濃度方面的可靠性。

表 3 土地利用迴歸模型表現

Statistic/Data	train	test	external
r square	0.91	0.92	0.84
adjusted r square	0.91	0.92	0.84
MAE	3.10	3.15	3.12
MSE	33.30	31.66	21.47
RMSE	5.77	5.63	4.63

3.3 基於土地利用迴歸之機器學習模型推估表現

進一步探討機器學習對於提升模型表現的關鍵作用，本研究運用了五種不同的機器學習模型，並深入評估它們的性能。其中，隨機森林(RF) 模型表現出眾，擁有最高的 R² 值，達到 0.95，這意味著它在預測 PM_{2.5} 濃度方面具有卓越的能力。此外，該模型的 MAE、MSE 和 RMSE 值也均優於其他模

型，分別為 3.14、19.78 和 4.45，顯示其預測準確性和誤差控制方面的卓越表現。值得一提的是，XGBoost 模型在計算速度方面表現出色，僅需 26.81 秒即可完成模型訓練，這對於處理大規模數據集至關重要。研究進一步對測試數據的 R² 值和 10 折交叉驗證結果進行比較，結果表明，這五個基於 LUR 的機器學習模型均未受到過度擬合的影響，確保了模型的穩健性和可靠性。在外部驗證方面，所有五個機器模型均表現出色，這進一步證實了它們對於預測外部數據的能力。

表 4 為各種不同的機器學習土地利用迴歸模型在訓練集上的表現。明顯地，傳統土地利用迴歸模型相較於加入機器學習演算法的模型，在模型解釋力方面較差 (R² 值為 0.91，是所有模型中最低的)，並伴隨較高的誤差。而以隨機森林演算法為基礎的模型表現最為優越，其模型解釋力最高 (R² 值達 0.95)，同時具有較低的 MAE (0.95)、MSE (19.78) 以及 RMSE (4.45)，誤差表現優於所有其他模型。這些結果強調了加入機器學習演算法對土地利用迴歸模型的改進效果，尤其是基於隨機森林的模型在預測方面表現卓越。

表 5 呈現了隨機森林演算法模型在測試和驗證上的結果，可以觀察到測試集和驗證集之間的 R² 值差異小於 0.05，這表明該模型無過擬合的問題。因此，本研究決定在後續的 PM_{2.5} 濃度地圖製作中，採用基於隨機森林機器學習演算法的土地利用迴歸模型。

表 4 傳統 LUR 模型以及機器學習模型表現

	Hybrid LUR	LUR based ML Model				
		RandomForest	GBM	XGBoost	LightGBM	CatBoost
R ²	0.91	0.95	0.92	0.93	0.92	0.92
adjusted R ²	0.91	0.95	0.92	0.93	0.92	0.92
MAE	3.10	3.14	3.83	3.75	3.81	3.83
MSE	33.30	19.78	29.30	27.60	29.16	29.30
RMSE	5.77	4.45	5.41	5.25	5.40	5.41

表 5 隨機森林演算法模型成果

Statistic/Data	10-fold	External	
		20% Data	External (2020)
R ²	0.92	0.92	0.91
adjusted R ²	0.92	0.92	0.82
MAE	3.87	3.84	3.68
MSE	30.00	28.73	22.57
RMSE	5.48	5.36	4.75

3.4 隨機森林模型推估圖

本研究針對兩個時間段的平均 $PM_{2.5}$ 濃度進行了推估：2021 年 5 月 1 日至 19 日（三級警戒實施前），以及 2021 年 5 月 20 日至 31 日（三級警戒期間）。如圖 2 所示，清晰呈現了警戒期間 (b) 的 $PM_{2.5}$ 濃度低於警戒實施前 (a)，與原先的預期一致。通過將警戒前後的濃度進行相減比較 (c)，可以看出於西南方整體呈現藍色（負值），尤其以南方顏色最深，代表 $PM_{2.5}$ 濃度下降最為顯著。圖 2 亦展示了 $PM_{2.5}$ 濃度的分佈，特別是在高雄市的西南部，包括人口稠密的都市地區和沿海地區，顯示出較高的濃度。都市地區的濃度較高，由於人口密度和交通流量大，而沿海地區的工業區域則是該地區濃度較高的主要原因。

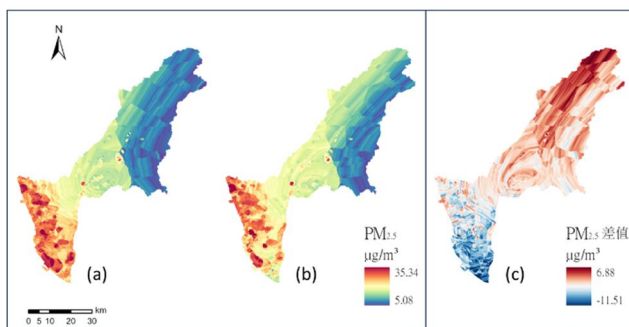


圖 2 高雄市三級警戒前 (a)與三級警戒期間；(b) $PM_{2.5}$ 濃度推估圖；(c)警戒後與警戒前相減

3.5 配對 t 檢定

根據模型估算結果，高雄市在實施三級警戒期間的平均 $PM_{2.5}$ 濃度為 $18.1 \mu g/m^3$ ，低於警戒前 ($19.9 \mu g/m^3$)。研究選取了特定土地利用分類地區來進行配對 t 檢定以確定是否存在明顯的濃度變化趨勢，土地利用類型分別為：工業區、純住宅地區、森林綠地、街道以及港口 (如圖 3 所示)。

如表 6 所示，配對 t 檢定的結果明確驗證了在警戒期間，不同土地利用區域的 $PM_{2.5}$ 濃度皆有顯著差異 (p 值 < 0.001)，尤其以沿海的工業區以及港口最為顯著。具體而言，警戒期間 $PM_{2.5}$ 濃度在所有土地利用中一致呈現下降的趨勢。尤其以高雄市西南沿海地區的港口和工業區最為顯著。

在警戒期間，由於 COVID-19 疫情的嚴重性，各國的海上運輸活動受到了巨大的制約，臺灣亦同；此外，許多工業區域也實施了工作限制和防疫措施，導致工業生產活動的減少或暫停。所有這些因素的共同作用下減少了空氣汙染排放的活動，有助於解釋為什麼港口和工業區域的 $PM_{2.5}$ 濃度下降得更加明顯，並在配對 t 檢定中呈現顯著的差異。

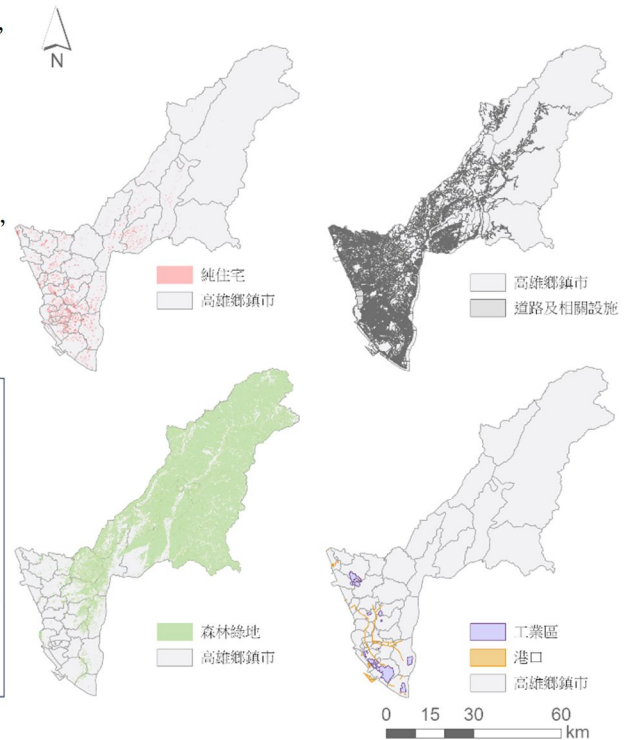


圖 3 土地利用變數之空間分布示意圖

表 6 依據不同土地利用進行三級警戒前後 $PM_{2.5}$ 濃度的配對 t 檢定 ($\mu g/m^3$)

	警戒前	警戒期間	差異	P-value
鄉鎮市區	19.94	18.15	-1.80	<0.001
工業區	24.99	21.43	-3.56	
住宅區	21.29	18.70	-2.59	
綠地	15.06	14.88	-0.17	
街道	18.41	17.47	-0.93	
港口	23.57	20.45	-3.12	

4. 討論

綜上研究結果可見，傳統的土地利用迴歸模型結合克利金空間內插後，在預測 $PM_{2.5}$ 濃度方面已經相當出色；然而，透過引入機器學習 (ML) 技術

後，研究成功地將模型的性能提到了更高的水準。這些 ML 模型呈現出多重優點，有助於增強研究成果。

首先，這些 ML 模型展現出更低的誤差，意味著它們能夠更準確地推估 PM_{2.5} 濃度。其次，這些模型具有更強的解釋能力，能夠更清晰地揭示不同因素對 PM_{2.5} 濃度的影響，有助於更深入地理解空氣污染的來源和變化。最重要的是，這與先前研究的發現一致，進一步鞏固了它們在土地利用回歸領域的卓越性 (Wong *et al.*, 2021)。

本研究亦觀察到，透過模型推估之成果，以及配對 t 檢定的證實，在三級警戒前與警戒期間前後不同時期的空氣污染物濃度確實存在顯著差異。此外，這些結果在不同的土地利用情況下，包括居住區、工業區、繁忙街區和綠地等，都得到了一致性的成果，進一步說明了在三級警戒期間高雄市的 PM_{2.5} 濃度確實較警戒前有所減少。圖 2 展示了不同時期的 PM_{2.5} 在高雄市中的空間分佈與濃度變化，提供了兩時期差異的視覺呈現。

與其他國際間的研究結果相比 (如表 7 所示)，高雄的 PM_{2.5} 亦為顯著下降，只是下降幅度遠小於其他各國。而根據過往研究大多都以一段期間內 PM_{2.5} 下降或上升 10 µg/m³ 為單位進行研究分析其與人體健康效益的關聯性，顯示高雄市 PM_{2.5} 在三級警戒期間短期內的輕微下降可能無法產生與其他地區相當的顯著健康效益。

表 7 不同國家地區 PM_{2.5} 濃度在封鎖與封鎖前之差異

	臺灣 高雄市	美國	印度 德里	中國
平均下降 濃度(µg/m ³)	1.80	1.44	21.29	14.07
下降百分比	9.05%	18.6%	27.35%	24%

5. 結論

綜上所述，臺灣工業城市實施的三級警戒政策對環境中 PM_{2.5} 濃度產生了顯著的影響，在此期間，空氣品質明顯的改善。本研究凸顯了政府政策、公共健康和環境問題之間的錯綜複雜關係，並強調了

需要採取綜合方法來應對這些相互關聯的挑戰。

這些研究成果在環境政策制定、公共健康保護和城市規劃等方面具有重要的應用價值。研究顯示，透過實施類似的警戒措施，不僅可以有效控制病毒傳播，還能在短期內顯著改善空氣質量，進而減少相關健康風險。此外，本研究引入的機器學習技術展示了其在提高模型預測精度和解釋能力方面的優勢。未來的研究可以進一步探討更多的空間變量和更長期的數據，以持續提升模型的準確性和穩定性。最後，未來的研究應該更多地關注不同環境政策在不同地區和城市的效果比較，以及其對長期健康效益的影響。通過綜合考慮各種因素並採取有力行動，可以更有效地應對公共健康和環境保護的雙重挑戰，實現可持續發展的目標。

參考文獻

吳昭儀、吳治達、陳裕政、許金玉、陳穆貞，2020。

應用克利金／土地利用迴歸混合模式推估林園臨海石化工業區懸浮微粒之時空分布，航測及遙測學刊，25(1)：11-23，DOI：10.6574/JPRS.202003_25(1).0002。 [Wu, J.Y., Wu, C.D., Chen, Y.C., Hsu, C.Y., and Chen, M.J., 2020. Estimate particulate matter concentrations variations in a petrochemical parks area using a hybrid kriging/land-use regression model, Journal of Photogrammetry and Remote Sensing, 25(1):11-23, DOI: 10.6574/JPRS.202003_25(1).0002. (in Chinese)]

曾于庭、吳治達、龍世俊，2018。應用土地利用迴歸模式推估北部空品區細懸浮微粒之時空分布，航測及遙測學刊，23(3)：191-204，DOI：10.6574/JPRS.201809_23(3).0004。 [Yu-Ting Zeng, Y.T., Wu, C.D., and Lung, S.C.C., 2018. Spatial-temporal variability of fine particulate matter in northern air quality zone of Taiwan using land use regression, Journal of Photogrammetry and Remote Sensing, 23(3): 191-204, DOI: 10.6574/JPRS.201809_23(3).0004. (in Chinese)]

- 曾芷琳、蕭雅萍、曾于庭、陳穆貞、龍世俊、吳治達，2019。整合空間資訊技術與土地利用迴歸模式推估高屏空品區細懸浮微粒之時空分布，航測及遙測學刊，24(2)：79-87，DOI：10.6574/JPRS.201906_24(2).0001。[Zeng, C.L., Hsio, Y.P., Zeng, Y.T., Mu-Jean Chen, M.J., Lung, S.C.C., and Wu, C.D., 2019. Integrate geospaital information technologies and land-use regression to estimate the spatial-temporal variability of fine particulate matter in kaohsiung-pingtung air quality zone, *Journal of Photogrammetry and Remote Sensing*, 24(2):79-87, DOI: 10.6574/JPRS.201906_24(2).0001. (in Chinese)]
- 翁佩詒、吳治達、蘇慧貞，2021。結合土地利用迴歸與極限梯度提升演算法發展高雄都會區二氧化氮之推估模型，航測及遙測學刊，26(1)：1-12，DOI：10.6574/JPRS.202103_26(1).0001。[Wong, P.Y., Wu, C.D., and Su, H.J., 2021. Development of an integrated model for NO₂ variation prediction in kaohsiung metropolis using land-use regression and XGBoost, *Journal of Photogrammetry and Remote Sensing*, 26(1): 1-12, DOI: 10.6574/JPRS.202103_26(1).0001. (in Chinese)]
- Basu, B., Alam, M.S., Ghosh, B., Gill, L., and McNabola, A., 2019. Augmenting limited background monitoring data for improved performance in land use regression modelling: Using support vector regression and mobile monitoring, *Atmospheric Environment*, 201: 310-322, DOI: 10.1016/j.atmosenv.2018.12.048.
- Berman, J.D., and Ebisu, K., 2020. Changes in U.S. air pollution during the COVID-19 pandemic, *Science of The Total Environment*, 739: 139864, DOI: 10.1016/j.scitotenv.2020.139864.
- Bitta, J., Pavlíková, I., Svozilík, V., and Jancík, P., 2018. Air pollution dispersion modelling using spatial analyses, *ISPRS International Journal of Geo-Information*, 7(12): 489, DOI: 10.3390/ijgi7120489.
- Correia, A., Pope, C., Dockery, D., Wang, Y., Ezzati, M., and Dominici, F., 2012. Effect of air pollution control on life expectancy in the United States: An analysis of 545 US counties for the period from 2000 to 2007, *Epidemiology*, 24(1): 23-31, DOI: 10.1097/EDE.0b013e3182770237.
- He, G., Pan, Y., and Tanaka, T., 2020. The short-term impacts of COVID-19 lockdown on urban air pollution in China, *Nature Sustainability*, 3(12): 1005-1011, DOI: 10.1038/s41893-020-0581-y.
- Lung, S.C.C., Guo, K.J., Chen, P.Y., Tsai, P.F., and Chen, P.C., 2004. Participants' exposure to PM_{2.5} and gaseous/particulate polycyclic aromatic hydrocarbons during the Ma-tsu Goddess parade, *Journal of Exposure Science and Environmental Epidemiology*, 14(7): 536-543, DOI: 10.1038/sj.jea.7500372.
- Orru, H., Maasikmets, M., Lai, T., Tamm, T., Kaasik, M., Kimmel, V., Orru, K., Merisalu, E., and Forsberg, B., 2011. Health impacts of particulate matter in five major Estonian towns: Main sources of exposure and local differences, *Air Quality, Atmosphere & Health*, 4(3-4): 247-258, DOI: 10.1007/s11869-010-0075-6.
- Pratap, V., Tiwari, S., Kumar, A., and Singh, A.K., 2021. COVID-19 lockdown induced air pollution reduction over India: A lesson for future air pollution mitigation strategies, *Journal of Earth System Science*, 130(4): 230, DOI: 10.1007/s12040-021-01722-y.
- 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.

Yin, P., Brauer, M., Cohen, A., Burnett, R.T., Liu, J., Liu, Y., and Zhou, M., 2015. Ambient fine particulate matter exposure and cardiovascular mortality in China: A prospective cohort study, *The Lancet*, 386: S6, DOI: 10.1016/S0140-6736(15)00584-X.

Impacts of the Level 3 Alert Brought by COVID-19 on Fine Particulate Matter of an Industrial City in Taiwan Using a Land-Use Based Machine Learning Model

Jun-Jun Su¹ Pei-Yi Wong² Yu-Ting Zeng³ Chia-Ho Lee⁴
Chih-Da Wu^{5,6,7,8*}

Abstract

The COVID-19 epidemic has brought significant changes to human activities worldwide, including in Taiwan. On May 19, 2021, the government announced a level 3 alert to restrict public movement. This study aims to assess the impact of the lockdown policy on PM_{2.5} concentrations in Taiwan's industrial city, Kaohsiung. Daily PM_{2.5} observations and geographic data from 1994 to 2020 were collected. A land-use regression model, combined with stepwise variable selection, was used to identify important factors affecting PM_{2.5} variability. These predictors were used to develop machine learning models with algorithms such as Random Forest (RF), which showed the best performance with an R² of 0.95. Paired t-tests indicated that PM_{2.5} levels were significantly lower during the alert (18.1 µg/m³) compared to before (19.9 µg/m³), with consistent results across residential, industrial, street, and green areas ($p < 0.001$).

Keywords: Air Pollution, Fine Particulate Matter (PM_{2.5}), Land-Use Based Machine Learning Model, COVID-19, Level 3 Alert

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, Environmental Protection Bureau of Chiayi City

⁵ Professor, Department of Geomatics, National Cheng Kung University

⁶ Adjunct Researcher Fellow, National Institute of Environmental Health Sciences, National Health Research

⁷ Adjunct Professor, Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University

⁸ Adjunct Researcher Fellow, 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: Apr. 12, 2024

Accepted Date: Jun. 24, 2024