

結合溫室氣體排放、綠蔽度衛星影像與土地利用資料的環境溫度機器學習預測模型開發

張皓庭¹ 陳映融² 柳婉郁^{3,4} 吳治達^{5,6,7*}

摘要

本研究綜合考量了溫室氣體、環境和人為活動相關變數，以利用大數據與五種機器學習演算，包含：隨機森林 (RF)、梯度提升 (GBR)、輕量梯度提升 (LGBMR)、類別提升 (CBR) 和極限梯度提升 (XGBoost) 來建立兩種溫室氣體 CO₂ 和 CH₄ 推估環境溫度的模型，其中 LGBMR 模型在 CO₂ 方面表現最佳，而 XGBR 模型在 CH₄ 方面效果較好。CO₂ 和 CH₄ 推估環境溫度模型的表現，R²值分別為 0.993 和 0.999。SHAP 值的分析確認了溫室氣體濃度、濕度、風速等因素對預測的關鍵影響。本研究成果為溫室氣體減排策略提供了新的評估方法，並為全球氣候變化研究與政策制定提供了重要參考，凸顯了跨學科合作的重要性。

關鍵詞：環境溫度、二氧化碳、甲烷、機器學習預測模型、機器學習演算法

1. 前言

面對全球暖化日益嚴峻的挑戰，深入理解並有效控制溫室氣體排放已成為一項急迫且至關重要的任務。自工業革命以來，我們見證了煤炭和石化燃料的大量使用不僅大幅增加了二氧化碳 (Carbon Dioxide, CO₂) 的排放量，還導致其他溫室氣體，如甲烷 (Methane, CH₄) 和氧化亞氮 (Nitrous oxide, N₂O) 等，大量進入大氣層。這些溫室氣體的累積不僅加劇了溫室效應，而且顯著推動了全球暖化現象，進而引發了一連串的環境、社會經濟，以及公共健康問題 (IPCC, 2022、World Health Organization, 2021)。

在全球性挑戰下，國際社會已經積極行動起來。從 1992 年聯合國氣候變遷綱要公約 (United Nations Framework Convention for Climate Change, UNFCCC) 的通過開始，全球各國一直在共同努力

減少溫室氣體排放。特別是在 2015 年的巴黎協定 (Paris Agreement) 確立之後，各國紛紛制定了更加明確和具體的減排目標 (UNFCCC, 2015)。臺灣在這全球共同應對氣候變化的大背景下，也積極參與其中，不僅制定了「溫室氣體減量及管理法」，而且進一步修訂為「氣候變遷因應法」，銘記著到 2050 年達到淨零排放的宏偉目標。

然而，目前的減排努力主要集中在控制排放源上，而尚未直接針對溫室氣體在環境中的實際濃度進行深入研究。考慮到溫室氣體在大氣中的長期存在及其對環境的深遠影響，迫切需要一種能直接評估減排策略對大氣中溫室氣體濃度影響的方法。因此，本研究旨在運用先進的大數據分析和機器學習技術，建立一個機器學習預測模型，以深入探究溫室氣體濃度與環境溫度之間的相互關係。相對於過去的研究，本研究更進一步考慮了溫室氣體、環境，以及人為活動相關的變數，以建立更全面的模型，

¹ 國立成功大學測量及空間資訊學系 博士後研究員

² 國立成功大學測量及空間資訊學系 研究助理

³ 國立中興大學森林學系 特聘教授

⁴ 國立中興大學永續農業創新發展中心 特聘教授

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

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

收到日期：民國 113 年 02 月 26 日

修改日期：民國 113 年 05 月 09 日

接受日期：民國 113 年 05 月 31 日

用以推估環境溫度的變化。我們致力於捕捉模型中的溫室氣體監測資料，以及與環境和人為活動相關的多項資料，進一步揭示它們與環境溫度之間的潛在趨勢。研究的目標在於實現對臺灣環境中溫室氣體濃度對環境溫度變化的準確預測。在建模過程中，將運用 SHAP 值深入研究溫室氣體與環境溫度之間的相關性。這項研究的重要性不僅體現在對臺灣應對氣候變化的理論和實踐層面的貢獻，同時也為全球氣候變化策略提供了有價值的參考。透過深入了解溫室氣體與環境溫度之間的複雜關係，有望為制定更有效的環境政策和氣候變遷適應策略提供具體的科學基礎。本研究的成果將有助於推動全球範圍內更可持續的發展和氣候變化防控工作。

2. 材料及方法

本研究計畫旨在結合過去在空氣污染模擬方面的豐富經驗，融合土地利用迴歸、機械學習和集成學習等多元方法學，旨在開發一套創新的溫室氣體濃度估計環境溫度的機器學習預測模型。圖 1 展示了本研究的整體流程，其中涵蓋了以下關鍵階段：(1) 資料庫準備 (Database preparation)；(2) 運用自動機器學習 (AutoML) 技術構建預測模型 (Leveraging SHAP values to choose variables and AutoML to streamline algorithm enhancement and model development)；(3) 進行模型的驗證與評估 (Assessment of algorithm effectiveness and outcome confirmation in machine learning models)。接下來，將詳細介紹每個階段的具體操作流程。

2.1 資料庫準備

本研究運用了 ArcMap 10.8 軟體的核心功能，即環域 (buffer) 空間分析，以進行溫室氣體排放源與溫度的深入探究。具體而言，研究團隊計算了測站周圍方圓 50 m 至 5000 m 的 14 個不同環域範圍內，溫室氣體排放與溫度相關的空間排放源分佈密度。此外，本研究還涵蓋了測站與這些排放源之間的最短距離、測站所處的海拔和地形等重要的地理環境特徵。研究總共收集了約四百多個與空間排放源相關的變數，為後續的分析工作提供了豐富的數

據基礎。關於這些變數和資料庫的詳細圖示和介紹，將在接下來的研究材料部分展示。此外，為了建立完整的模型並進行測試和驗證，我們對收集到的研究資料進行了時間段分割，涵蓋 2000 年至 2019 年。將 2000 年至 2018 年的資料分為訓練集 (Training dataset) 和測試集 (Testing dataset)，其中 80% 作為訓練集用於模型開發，20% 作為測試集用於模型初步測試。2019 年的測量數據則作為獨立的外部數據 (External dataset) 來驗證模型的性能，不會納入模型構建過程中。

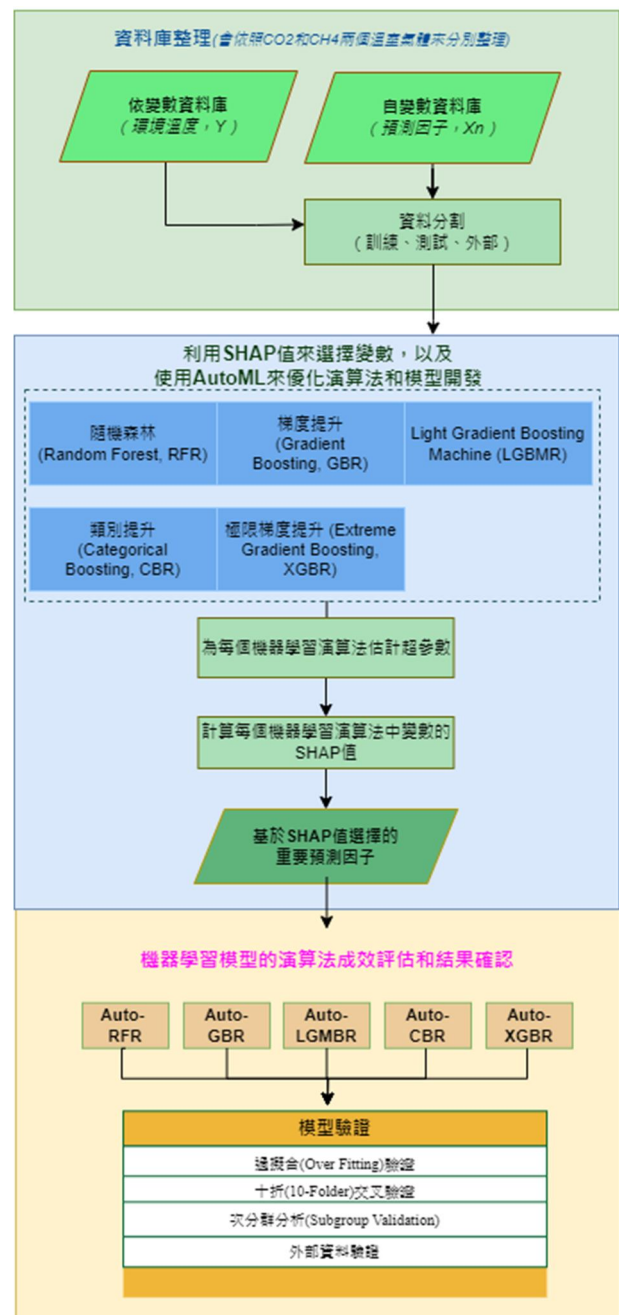


圖 1 研究流程圖

2.1.1 溫室氣體與氣象監測資料庫

環境部於臺灣本島設立了約 76 個空氣品質監測站，這些站點不僅專注於空氣品質的監控，同時也追蹤溫室氣體濃度的變化。具體來說，目前有 6 個和 37 個監測站分別對 CO₂ 和 CH₄ 濃度進行持續的監測和記錄（詳見圖 2）。除 CH₄ 外，這些監測站還對各種空氣污染物（例如：細懸浮微粒、懸浮微粒、氮氧化物等）和氣象因素（如：溫度、降雨量、相對濕度等）進行追蹤。本研究搜集公開溫度數據（涵蓋 2000 年至 2019 年），以及相同時段的 CO₂（2000 年至 2019 年）和 CH₄（2002 年至 2019 年）日平均監測數據，作為研究建模的基礎。

另外，根據既往研究，溫度、其他氣象因子、空氣污染與溫室氣體濃度分布之間有顯著關係 (Ramanathan & Feng, 2009)。因此，本研究將利用研究團隊以往使用土地利用迴歸和 XGBoost 演算法估算出的臺灣 PM_{2.5}、NO₂ 和 O₃ 濃度分布 (Hsu *et al.*, 2019、Wong *et al.*, 2021a、Wong *et al.*, 2021b) 作為解釋變數。這些方法在估算臺灣全島 PM_{2.5}、NO₂ 和 O₃ 的時空變異時，準確度分別達到了 94%、91% 和 92%，經過多次驗證後，模型的穩定性和可靠性得到了證實。至於氣象資料，包括相對濕度、風速、風向、降雨量和紫外線等，則是來自中央氣象局的資料庫，這些數據經過普通克利金法處理後，研究使用溫室氣體監測站的點位萃取出的氣象數值，這些數據能夠反映全臺灣氣象狀況的分布，為研究接下來溫室氣體推估模型的分析提供了有力的數據支持。此外，本研究根據臺灣的地理座標和研究時間，額外計算和整理了太陽天頂角和太陽赤緯角等變數。

2.1.2 MODIS 衛星觀測之環境綠蔽度資料庫

鑑於全球環境變遷的重要性，美國國家航空暨太空總署 (National Aeronautics and Space Administration, NASA) 運用先進的遙測技術和衛星資源，建立了地球觀測系統 (Earth Observing System, EOS)。該系統致力於長期監測地球的大氣

層、水圈、生物圈以及植被等關鍵地表參數的變化。這些珍貴的資料由在美國地質調查所 (The U.S. Geological Survey, USGS) 整理後，為全球學者和專家提供了重要的學術研究資源。本研究將利用 EOS 系統中的 MOD13Q1 第 6 版全球植被監測資料庫，來估算區域環境的綠蔽度。這些資料基於 Moderate Resolution Imaging Spectroradiometer (MODIS) 衛星的紅光 (645 納米) 和近紅外光 (858 納米) 波段，進行常態化植生差異指標 (Normalized Difference Vegetation Index, NDVI) 的計算，空間解析度為 250 m×250 m。自 2000 年 2 月起，每 16 天生成一次的全球綠蔽度監測資料，已被 NASA 和 USGS 透過多時期的地表觀測驗證，確保了第 6 版 NDVI 產品的準確性，可靠地支持學術發表 (USGS, 2020)。考慮到植物和森林在減少溫室氣體排放、碳固定與受溫度影響方面的關鍵作用，本研究根據兩種溫室氣體監測數據庫的時間軸，來準備相對應時間長度的環境綠蔽度資料庫。此外，本研究利用 ArcGIS Pro 提供的“最近鄰取樣”函數，將 250 m 空間分辨率的 NDVI 放大至 50 m。這種重新取樣技術是在影像處理中使用的，其目的是識別原始資料的最近值 (Brandsma & Können, 2006)。該技術的優點在於保持原始值，同時避免在資料之間進行插值，確保了原始值在重新取樣後的資料中被保留。通過在放大後保持準確性，該函數在重新取樣的資料中盡可能地複製了原始值。

2.1.3 全國土地利用調查資料

本研究利用由內政部國土測繪中心進行的第三次全國土地利用調查數據，目的在於收集與溫度、溫室氣體排放等相關的土地利用資料。該調查的土地分類系統採用層次化樹狀結構，分為三個層級。例如，在第三次調查中，第一級分類涵蓋了農業、森林、交通、建設、公共設施、休閒娛樂、採礦、鹽業以及其他土地使用等 9 大類；第二級對這些大類進一步細化為 61 小類；而第三級則基於第二級，進行更細分的劃分，如稻田等 101 個細分類別。本研究選取其中具有代表性的類別，如住宅區、國際機場、畜牧業、水稻農田、垃圾掩埋場、鐵路、道

路、製造業、污水處理設施、水體等，作為預測變數，並整合實驗的數據庫。此外，與環境綠覆率密切相關的土地利用類型，如植被、森林、公園綠地等，也被納入考量，作為估計環境綠蔽度的第二種指標。

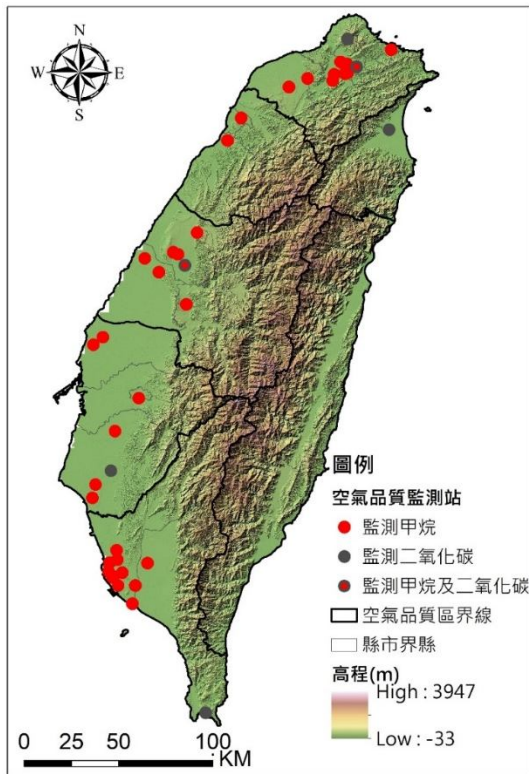


圖 2 環境部甲烷及二氧化碳測站空間分佈圖

2.1.4 地理標記資料庫

本研究所採用的地標數據資源，主要來自於勤崙國際科技股份有限公司建立的 Point of Interest (POI) 地標資料庫，以及環境部所提供的全臺灣焚化爐煙囪位置資料。POI 資料庫，作為 Google Earth 和 Google Map 中的重要地標資訊來源，不僅豐富了地圖的內容，也提供了實時的地理參考資訊。這個資料庫每季更新一次，並且按照三個層次組織：主要分類、次要分類和具體細項。為了與本研究的時間範圍相符，研究特別選用了 2004 年、2006 年、2008 年及 2010 年的 POI 數據。在主分類中，包括政府機構、公共事業、金融證券等 12 至 14 個大類，進一步細分為 119 至 160 的次分類，以及 920 至 1,350 的具體細項。本研究將從 POI 資料庫

中提取寺廟、中式餐廳、火力發電廠等特定地標的座標信息。另外，關於焚化爐煙囪的資料，則是從環境部在政府開放資料平臺 (MOI Open Data) 上所發布的官方開放數據中獲得。

2.1.5 路網數值圖

鑒於能源相關活動，如煤炭燃燒、石油使用和交通運輸，是溫室氣體排放和改變溫度的主要來源之一，本研究計劃利用交通部運輸研究所提供的 2010 年和 2015 年全臺灣 1/5000 比例尺的 GIS 路網數值圖，以收集交通排放相關的關鍵數據，用於建立模型。這些路網數據根據其特性，將臺灣的道路分類為國道、省道、縣道及鄉道等八大類別。本研究使用這些路網數值圖，將道路進行更精細的分類：主要道路（包括國道、省道、縣道及市區快速道路）、地方道路（包括市區道路、鄉道、產業道路及無名道路）以及所有類型的道路，從而提供各類道路的密度資訊，以支持我們的研究分析。

2.1.6 數值地形模型資料

本研究中使用的數值地形模型 (Digital Terrain Model, DTM) 是將臺灣地區細致劃分成一系列規則的網格，並在這些網格的每一點記錄相應的海拔高度。這些數據被儲存於 GIS 系統中，以便進行深入的地形分析。考慮到地形特徵，如：海拔高度，可能與溫室氣體排放源或是與溫度改變之間存在關聯，並可能對溫室氣體濃度或溫度的變化產生影響，本研究選擇採用由中華民國農業部航測的最新數據，該數據提供了 20 m×20 m 解析度的全臺灣數值地形模型。通過這些數據，本研究將收集測站周圍地區的海拔、坡度、坡向等關鍵地形資訊，以支持使用溫室氣體來估算溫度的研究分析。

2.1.7 工業區綜合資料庫

自 1970 年起，經濟部工業局致力於引領與促進臺灣工業的全面發展，其職責範圍涵蓋了制定工業發展策略、管理工業區開發，以及處理工業環境污染問題等重要任務。為了深入研究工業活動對環境的影響，本研究引入了經濟部工業局所提供的

2010 年與 2015 年臺灣各主要加工出口區與科學園區的地理資訊系統 (Geographic Information System, GIS) 數據。這些數據將協助研究精確地分析工業區分布對環境因素的相互作用及其可能的影響。

2.1.8 溫室氣體排放清冊資料庫

本研究依據環境部依據溫室氣體減量及管理法 (溫管法) 第 13 條之規定，以及聯合國政府間氣候變化專門委員會 (Intergovernmental Panel on Climate Change, IPCC) 在 2006 年發布的《國家溫室氣體清冊指南》(2006 IPCC Guidelines for National Greenhouse Gas Inventories) 所採納的統計方法，進行了詳實的資料整合。環境部每年更新並發布《中華民國國家溫室氣體清冊報告》，詳細記錄自 1990 年起我國溫室氣體排放及吸收的趨勢。

本研究主要分析我國最新版的《中華民國國家溫室氣體排放清冊報告 (2021 年版)》，該報告綜合了全臺灣各年度的溫室氣體排放數據。此外，亦從各縣市政府的溫室氣體管制執行方案中搜集資料，整合各地區的溫室氣體排放量，以建立一個全面且精確的溫室氣體排放清冊資料庫。這些資料將為研究建立溫室氣體濃度推估溫度模型提供關鍵支持，助研究更深入了解和應對氣候變化的挑戰。所有上述的空間資料及相關資訊已整合於表 1 中。表格內容包括：資料來源 (例如資料庫)、變數、單位、地理統計格式 (例如網格大小)、以及時間解析度或時間點。

2.2 運用自動機器學習技術構建預測模型

本研究以 CO₂ 和 CH₄ 兩種溫室氣體為基準，分別建立轉換環境溫度的模型。模型的依變數為環境溫度，自變數包括溫室氣體 (CO₂、CH₄)、環境綠蔭度資料、國土地利用調查資料、地理標記資料、路網數據、數值地形模型資料以及工業區綜合資料。為了建立模型，採用了五種不同的機器學習演算法：隨機森林 (Random Forest, RF)、梯度提升 (Gradient Boosting Regressor, GBR)、輕量梯度提升 (Light Gradient Boosting Machine Regressor, LGBMR)、類

別提升 (CatBoost Regressor, CBR) 和極限梯度提升 (eXtreme Gradient Boosting, XGBoost)。這些演算法各具特色：隨機森林通過建立多個迴歸樹來提升計算效能；梯度提升通過迭代弱分類器來優化模型；輕量梯度提升在維持高準確度的同時，擁有更快的訓練速度和較低的記憶體需求；類別提升強化了對類別特徵的支持；極限梯度提升結合了並行與串行建樹技術，在減少過擬合的同時實現更優的預測效果。這五個演算法的公式如下：

(1) 隨機森林公式

A. 建樹：

- 對於每棵樹，從訓練集隨機抽取一個有放回的樣本子集 (bootstrap sampling)。
- 對於每個節點，隨機選擇特徵子集，基於這些特徵構建最佳分裂。

B. 集成：

- 對於分類問題，取所有樹的投票結果的多數作為最終預測。
- 對於迴歸問題，取所有樹的預測平均值作為最終預測。

(2) 梯度提升公式

A. 初始化模型 $F_0(x)$ 為常數值 (如訓練集目標值的平均)。

B. 對於每一步 $m = 1, 2, \dots, M$ ：

- 計算當前模型的殘差 $ri^{(m)} = yi - F_{m-1}(xi)$ 。
- 基於殘差訓練一個新的樹 $h_m(x)$ 。
- 更新模型 $F_m(x) = F_{m-1}(x) + \eta h_m(x)$ ，其中 η 是學習率。

(3) 輕量梯度提升公式

- ##### A. 使用直方圖算法來加速尋找最佳分裂點。
- ##### B. 使用基於葉子的增量 (Leaf-wise Growth) 來構建樹，而不是基於層的增量 (Level-wise Growth)，從而提高效率和準確性。

(4) 類別提升公式

- ##### A. 使用 Ordered Boosting 技術，在構建每個新的樹時，避免信息洩漏。
- ##### B. 使用目標編碼 (Target Encoding) 來處理類別特徵。

表 1 在模型建構中使用的可能預測變項及其數據來源

資料來源 (資料庫)	變數	單位	地理統計格式 (網格大小)	時間解析度/時間點
環境部 (溫室氣體、空污和環境溫度氣象監測資料庫)	CO ₂	ppm	網格 (50 m)	逐時資料/2000 年至 2019 年
	CH ₄	ppm		
	O ₃	ppb		
	PM _{2.5}	µg/m ³		
	NO ₂	ppm		
	環境溫度	°C		
中央氣象署 (氣象監測資料庫)	溫度	°C	網格 (50 m)	逐時資料/2000 年至 2019 年
	降雨量	mm		
	風速	m/s		
	相對濕度	%		
	風向	360 度		
	紫外線	UVI		
自行計算 (太陽天頂角赤緯角)	太陽天頂角	°	網格 (50 m)	逐時資料/2000 年至 2019 年
	太陽赤緯角	°		
美國國家航空暨太空總署 (MODIS 衛星觀測之環境綠蔽度資料庫)	常態化植生差異指標	-	網格 (50 m)	16 天/2000 年至 2019 年
內政部國土測繪中心 (全國土地利用調查資料)	住宅區	-	密度/最近距離 (50 m)	2015 年
	國際機場			
	畜牧業			
	水稻農田			
	垃圾掩埋場			
	鐵路			
	道路			
	製造業			
	污水處理設施			
	水體			
	植被			
	森林			
勤崙國際科技股份有限公司 (地理標記資料庫)	寺廟	-	密度/最近距離 (50 m)	2004、2006、2008 和 2010 年
	中式餐廳			
	火力發電廠			
交通部 (路網數值圖)	主要道路	-	密度/最近距離 (50 m)	2010 年和 2015 年
	地方道路			
	所有類型的道路			
中華民國農業部 (數值地形模型資料)	海拔高度	m	網格 (20 m)	2016 年
	坡度	°		
	坡向	360 度		
經濟部工業局 (工業區綜合資料庫)	加工出口區	-	密度 / 最近距離 (50 m)	2010 年和 2015 年
	科學園區			
環境部 (溫室氣體排放清冊資料庫)	溫室氣體排放量	千公噸 二氧化碳 碳當量	-	2000 年至 2019 年

(5) 極限梯度提升公式：

A. 初始化模型 $F_0(x)$ 為常數值。

B. 對於每一步 $m = 1, 2, \dots, M$ ：

定義目標函數： $L(m) = \sum_{i=1}^n l(y_i, F_{m-1}(x_i) + fm(x_i)) + \Omega(fm)$

其中 l 是損失函數， Ω 示正則化項。

訓練新的數 $fm(x)$ 以最小化目標函數。

更新模型 $F_m(x) = F_{m-1}(x) + \eta fm(x)$

在建模過程中，引入了自動機器學習(AutoML)技術來自動化演算法的超參數優化。接著，對每個演算法中的所有變數進行 Shapley Additive exPlanations (SHAP) 值計算，這些 SHAP 值代表了變數對模型的貢獻程度。因此，SHAP 值越大的變數在模型中越重要。在選擇模型中的變數時，首先考慮變數值的大小順序及其對模型解釋量的貢獻度；如果連續兩次新增變數的解釋增加量小於 0.01，則視為變數篩選的終點。

為了評估模型的效能，將使用相關係數 (r)、判定係數 (R^2) 以及校正後判定係數 (Adjusted R^2) 來衡量模型的準確度。同時，也將採用平均絕對誤差 (MAE)、均方誤差 (MSE) 和均方根誤差 (RMSE) 來衡量模型誤差。根據這些指標的表現，將對各演算法進行排序，並以準確度高、誤差小為原則，最終確定最合適的演算法。

2.3 進行模型的驗證與評估

在本研究中，從先前提及的五種機器學習演算法中挑選出解釋能力最佳的一種作為最終模型。為了確保所選演算法的穩定性和有效性，採用了四種不同的方法來進行模型的穩定性 (Robustness) 測試與驗證，包括：過擬合驗證 (Overfitting validation)、十折交叉驗證 (10-fold cross-validation, 10-fold CV)、次分群分析 (Subgroup validation) 以及外部資料驗證 (External validation)。這些方法的具體說明如下：

(1) 過擬合驗證：通過比較訓練集和測試集計算出的 R^2 值之間的差異來檢驗模型。如果這兩者之間的差異不超過 0.1，則可以認為模型沒有發生過擬合現象。

(2) 十折交叉驗證：這一方法旨在測試模型效能的

穩定性。將整體樣本分成十等份，每次選取其中九份作為訓練集，剩餘一份作為測試集。這個過程重複十次，以確保每一份樣本都有機會作為測試集。通過這種方式，能夠更準確地評估模型的預測準確度和穩定性。

(3) 次分群分析：這一方法評估模型在不同時間單位下的預測能力差異，包括春、夏、秋、冬四季及研究期間 2000 年至 2019 年各年的數據。將資料按不同季節和年份進行分層，然後分別計算每個分層的模型解釋力，進行比較分析。

(4) 外部資料驗證：使用 2019 年的資料來驗證基於 2000 至 2018 年資料所建立的模型。其目的是評估模型的外推能力是否良好。

通過這四種驗證方法的綜合應用，能夠全面評估所選演算法的效能，保證模型的可靠性和準確性。

3. 結果

3.1 描述性統計分析：建構溫室氣體轉換對環境溫度影響的模型變數

本研究分別針對 CO_2 和 CH_4 建立了溫室氣體轉換環境溫度的模型。模型建立時所挑選到的變數呈現於表 2。從表中可觀察到，建立 CO_2 和 CH_4 轉換環境溫度模型所使用到的樣本數依序為 23,223 與 294,475 筆。 CO_2 轉換環境溫度的模型中選到的變數 (研究期間的平均值) 除了 CO_2 (415.67 ppm) 本身，還有溫度 (22.92 °C)、濕度 (79.02%)、風速 (1.66 m/s)、雨量 (6.48 mm)、數值高程模型 (16.55 m)、NDVI (0.25)、150m 方圓範圍內佔地比建築物 (0.41)、太陽天頂角 (24.32°) 和太陽赤緯角 (0.12°)。 CH_4 轉換環境溫度的模型中選到的變數除了 CH_4 (2.01 ppm) 本身，還有溫度 (23.22 °C)、濕度 (77.7%)、風速 (1.76 m/s)、雨量 (5.58 mm)、數值高程模型 (34.42 m)、NDVI (0.24)、2500 m 方圓範圍內佔地比建築物 (0.32)、太陽天頂角 (23.91) 和太陽赤緯角 (0.13)。

表 2 建立溫室氣體轉換環境溫度模型的變數描述性統計

Variable	CO ₂ 模型 (N=23,223)				CH ₄ 模型 (N=294,475)				變數中文名稱
	平均值	標準差	最小值	最大值	平均值	標準差	最小值	最大值	
AMB_TEMP (°C)	23.87	5.06	5.5	33.8	24.07	4.97	0.02	37.14	環境溫度
CO ₂ (ppm)	415.67	19.24	234.52	533.25	-	-	-	-	二氧化碳
CH ₄ (ppm)	-	-	-	-	2.01	0.27	0	8.78	甲烷
TH01 (°C)	22.92	4.95	4.24	32.57	23.22	4.91	3.79	32.79	溫度 (由氣象測站資料進行空間內插)
RH01 (%)	79.02	8.02	34.6	99.78	77.7	7.58	28.45	99.84	濕度 (由氣象測站資料進行空間內插)
WD01 (m/s)	1.66	0.81	0.06	12.24	1.76	0.95	0.02	13.32	風速 (由氣象測站資料進行空間內插)
PP01 (mm)	6.48	20.48	0	569.85	5.58	19.03	0	642	雨量 (由氣象測站資料進行空間內插)
Dem (m)	16.55	16.09	5	44	34.42	67.58	1	448	數值高程模型
NDVI	0.25	0.12	0.01	0.74	0.24	0.12	-0.01	0.94	常態化植生差異指標
LUIT_B1_B150 (Cover ratio)	0.41	0.44	0	1	-	-	-	-	建築物 (150m 方圓範圍內佔地比)
LUIT_B1_B2500 (Cover ratio)	-	-	-	-	0.32	0.27	0	0.93	建築物 (2500m 方圓範圍內佔地比)
Solar_Zenith (°)	24.32	16.41	0.26	48.53	23.91	16.4	0.01	48.62	太陽天頂角
Solar_Declination (°)	0.12	16.43	-23.44	23.44	0.13	16.45	-23.44	23.44	太陽赤緯角

表 3 五個機器學習演算法建立溫室氣體轉換環境溫度模型之結果

Outcome Algorithm	Temperature (CO ₂)					Temperature (CH ₄)					
	GBR	XGBR	CBR	LGBMR	RFR	GBR	XGBR	CBR	LGBMR	RFR	
Training dataset (80%)	R square	0.997	0.989	0.992	0.993	0.988	0.985	0.999	0.979	0.983	0.992
	Adjusted R square	0.997	0.989	0.992	0.993	0.988	0.985	0.999	0.979	0.983	0.992
	RMSE	0.3	0.533	0.456	0.416	0.566	0.611	0.056	0.729	0.648	0.449
	MSE	0.09	0.284	0.208	0.173	0.32	0.374	0.003	0.531	0.42	0.202
	MAE	0.146	0.301	0.259	0.258	0.292	0.3	0.017	0.383	0.359	0.213
Testing dataset (20%)	R square	0.982	0.98	0.982	0.98	0.973	0.978	0.981	0.977	0.978	0.975
	Adjusted R square	0.981	0.98	0.982	0.98	0.973	0.978	0.981	0.977	0.978	0.975
	RMSE	0.687	0.717	0.681	0.716	0.836	0.742	0.685	0.761	0.739	0.783
	MSE	0.472	0.513	0.463	0.512	0.698	0.551	0.469	0.579	0.547	0.613
	MAE	0.366	0.409	0.383	0.392	0.467	0.388	0.327	0.399	0.399	0.414
10-fold cross-validation	R square	0.981	0.979	0.981	0.98	0.973	0.978	0.98	0.976	0.978	0.975
	Adjusted R square	0.981	0.979	0.981	0.98	0.973	0.978	0.98	0.976	0.978	0.975
	RMSE	0.71	0.735	0.703	0.733	0.848	0.746	0.701	0.769	0.747	0.797
	MSE	0.504	0.54	0.495	0.537	0.719	0.556	0.492	0.591	0.559	0.635
	MAE	0.381	0.41	0.386	0.399	0.469	0.38	0.334	0.397	0.399	0.419
External validation	R square	0.977	0.974	0.975	0.979	0.98	0.981	0.982	0.983	0.981	0.979
	Adjusted R square	0.977	0.974	0.975	0.978	0.98	0.981	0.982	0.983	0.981	0.979
	RMSE	0.775	0.801	0.803	0.757	0.743	0.595	0.58	0.565	0.584	0.622
	MSE	0.601	0.642	0.645	0.573	0.551	0.354	0.337	0.32	0.341	0.386
	MAE	0.408	0.408	0.448	0.379	0.396	0.338	0.32	0.31	0.335	0.368
Averaged validation MAE*	0.394	0.409	0.417	0.389	0.433	0.359	0.327	0.354	0.367	0.393	
Averaged validation R ²	0.981	0.979	0.981	0.98	0.973	0.978	0.98	0.976	0.978	0.975	

3.2 五個機器學習演算法的模型建立結果

本研究採用五種不同的演算法來構建溫室氣體轉換環境溫度模型，其結果詳細列於表 3 中。對於 CO₂ 轉換環境溫度模型，從數據中明顯看出，GBR 和 LGBMR 在多個性能指標上展示出了類似且較高的準確度，特別是在訓練數據集上的 R² 值，分別高達 0.997 和 0.993。然而，RFR 在某些測試條件下表現略遜一籌，特別是在測試數據集和 10 折交叉驗證的均方誤差 (MSE) 及均方根誤差 (RMSE) 方面。綜合考量所有性能指標，LGBMR 的整體表現最為優異，特別是在平均驗證的均絕對誤差 (MAE) 為 0.389 和平均驗證 R² 值為 0.980。因此，選擇 LGBMR 演算法作為 CO₂ 轉換環境溫度的最終模型。

接下來，針對 CH₄ 轉換環境溫度的模型，XGBR 在大多數的性能指標上表現卓越，尤其是在訓練數

據集上的 R² 值高達 0.999，顯示出對數據的極高適配度。相較之下，RFR 在某些測試條件下的表現不佳，特別是在 RMSE 和 MSE 上。綜合考量，XGBR 成為 CH₄ 轉換環境溫度模型的首選演算法，不僅擁有最低的平均驗證 MAE (0.327)，還有較高的平均驗證 R² 值 (0.980)。

在深入分析溫室氣體轉換環境溫度模型之後，本研究利用 SHAP 值來揭示 CO₂ 和 CH₄ 轉換環境溫度模型中的關鍵影響因素(請參見圖 3)。SHAP 值提供了一種強大的解釋方法，能夠清晰地展示各個變數對模型預測的貢獻程度。

從圖 3 中的 CO₂ 轉換環境溫度的 LGBMR 模型和 CH₄ 轉換環境溫度的 XGBR 模型中，可以觀察到 CO₂ 和 CH₄ 對環境溫度的影響程度。具體而言，它們的平均 SHAP 值分別為第 4 名 (+0.11) 和第 6 名 (+0.52)。這意味著這兩種溫室氣體確實與環境溫度存在關聯。

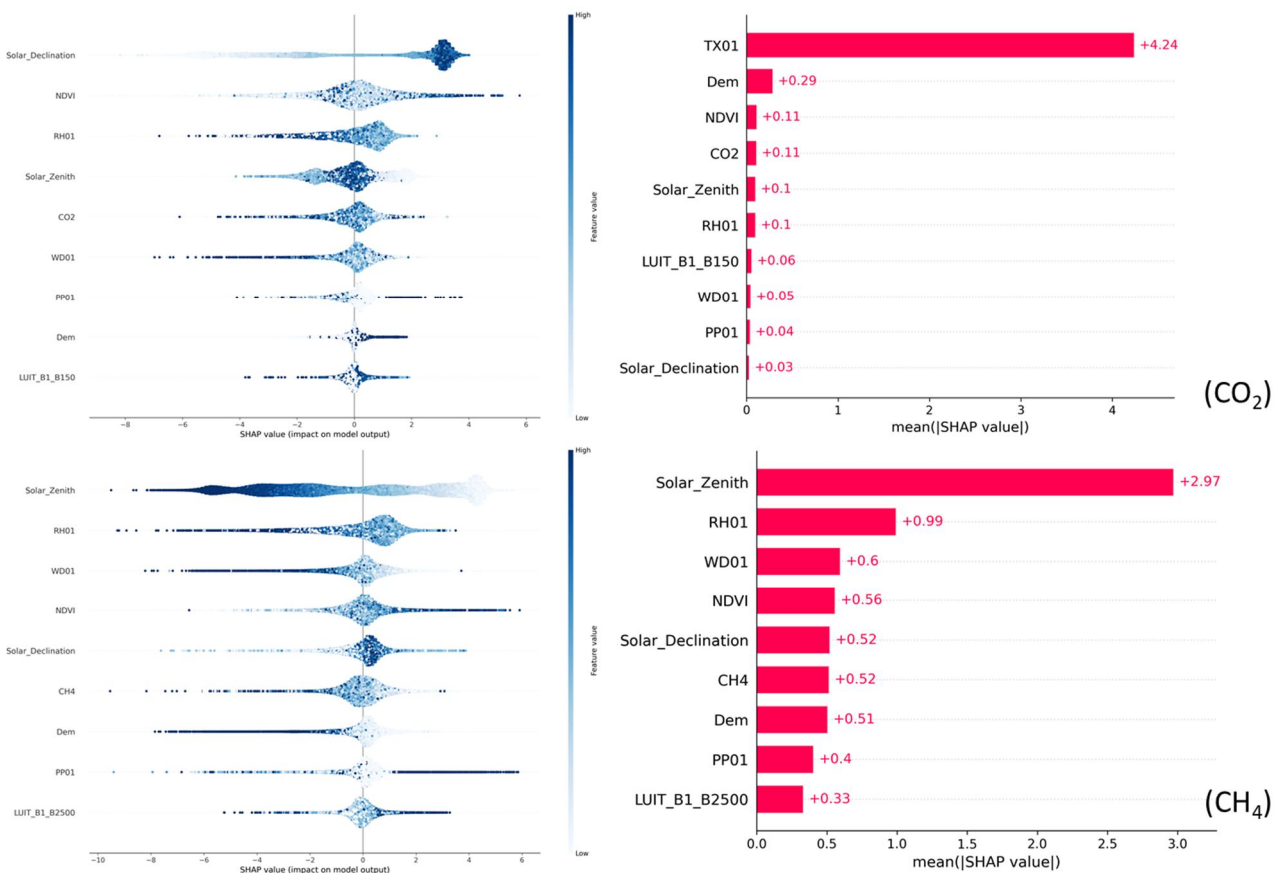


圖 3 SHAP 值解析：探索溫室氣體轉換溫度模型中關鍵變數的影響力 (上圖展示 CO₂ 轉換溫度模型的關鍵因素，下圖揭示 CH₄ 轉換溫度模型的主要影響變數)

對於 CO₂ 轉換環境溫度的 LGBMR 模型，分析結果顯示，影響該模型預測最顯著的五大因素依序為：由氣象測站資料進行空間內插的溫度、數值高程模型 (DEM)、常態化植生差異指標 (NDVI)、二氧化碳濃度、以及太陽天頂角。這些因素的排列反映了它們對預測結果的相對重要性，其中氣象測站資料進行空間內插的溫度影響最大，顯示了地理位置和即時氣溫條件對環境溫度變化有顯著的影響。

在 CH₄ 轉換環境溫度的 XGBR 模型中，最重要的五個影響因子則略有不同，分別為太陽天頂角、相對濕度、風速、NDVI、以及太陽赤緯角。這表明在預測 CH₄ 對環境溫度的影響時，太陽的位置和氣象條件尤為關鍵。特別值得注意的是，太陽天頂角和太陽赤緯角的重要性，這可能與太陽輻射對氣溫和 CH₄ 轉換率的影響有關。

綜合來看，這些分析結果不僅增進了對溫室氣體與環境溫度關係的理解，而且還為未來的氣候變化研究和策略制定提供了有價值的參考。透過深入分析這些關鍵因素，可以更有效地預測和應對氣候變化帶來的挑戰。

3.3 模型的驗證與評估

本研究針對溫室氣體轉換環境溫度模型進行了周全的驗證，其中分析了 CO₂ 與 CH₄ 轉換溫度的兩種模型，分別採用 LGBMR 和 XGBR 演算法。以下是對這兩種模型的驗證過程和結果的整理與分析。

首先，在 CO₂ 轉換環境溫度的 LGBMR 模型中，(1) 過擬合驗證中 (表 3)，模型的過擬合值僅為 0.013，這顯示了模型具有良好的泛化能力，能夠適應多變的數據條件；(2) 十折交叉驗證 (表 3)，此模型在十折交叉驗證中表現出色，R² 值高達 0.980。此外，均方根誤差 (RMSE) 為 0.733，均方誤差 (MSE) 為 0.537，平均絕對誤差 (MAE) 為 0.399，這些指標共同證明了模型的高準確性；(3) 次分群驗證中 (圖 4)，無論是根據季節還是年份進行的次分群驗證，R² 值均超過 0.97，MAE 均低於 0.05，彰顯了模型在不同條件下的高穩定性；(4) 對 2019 年的外部資料進行驗證時，MAE 稍高為 0.32，但依

然在接受範圍內。從表 3 中的數據顯示，外部資料驗證的 R² 值為 0.979，RMSE 為 0.757，MSE 為 0.573，MAE 為 0.379，進一步確認了模型的穩健性。

接著，CH₄ 轉換環境溫度的 XGBR 模型中，(1) 過擬合驗證 (表 3) 中可以看到，該模型的過擬合值為 0.018，表示在避免過度複雜化方面做得相當不錯；(2) 十折交叉驗證中 (表 3) 的模型結果同樣表現優異，R² 值達到 0.98，RMSE 為 0.701，MSE 為 0.492，MAE 為 0.334，顯示出高準確度；(3) 次分群驗證 (圖 4) 的分析結果，無論是根據季節或年份進行的次分群驗證，R² 值均超過 0.92，MAE 均低於 0.4，證明了模型在不同子集中的良好性能；(4) 外部資料驗證中 (表 3)，結果進一步展現了模型的堅實性，外部資料驗證的 R² 值為 0.982，RMSE 為 0.58，MSE 為 0.337，MAE 為 0.32。

總結來說，無論是 CO₂ 還是 CH₄ 轉換環境溫度模型，在進行多角度的驗證過程中均表現出卓越的性能。這些結果不僅證明了模型的高效性，也顯示了其在面對多樣化數據時的穩健性和可靠性。

4. 討論

在本研究的前言中，闡述了溫室氣體排放對環境溫度的潛在影響，以及這一問題的全球重要性。在結果章節中，本研究進一步展示了這些假設的實證分析。在此基礎上，本討論章節在深入解析這些研究發現，將其與現有的學術文獻和實際情況相結合，並探討其對未來研究方向的可能影響。

首先，本研究利用先進的機器學習技術，深入剖析了 CO₂ 和 CH₄ 與環境溫度之間的關係。首先，從模型中觀察到 CO₂ 和 CH₄ 與環境溫度的關係，背後可能的原因為 CO₂ 和 CH₄ 是兩種主要的溫室氣體，這兩種溫室氣體的作用是吸收地球表面向外輻射的熱能，並將部分熱能重新輻射回地球表面。這個過程類似於一個溫室，使得地球的溫度保持較高 (Al-Ghussain, 2019)。此外，從衡量溫室氣體對氣候變化影響的指標—輻射強迫，CO₂ 的輻射強迫約為每平方米 2 瓦特 (W/m²)，而 CH₄ 的輻射強迫約為 0.65 W/m²。儘管 CH₄ 的輻射強迫較低，但它的影響

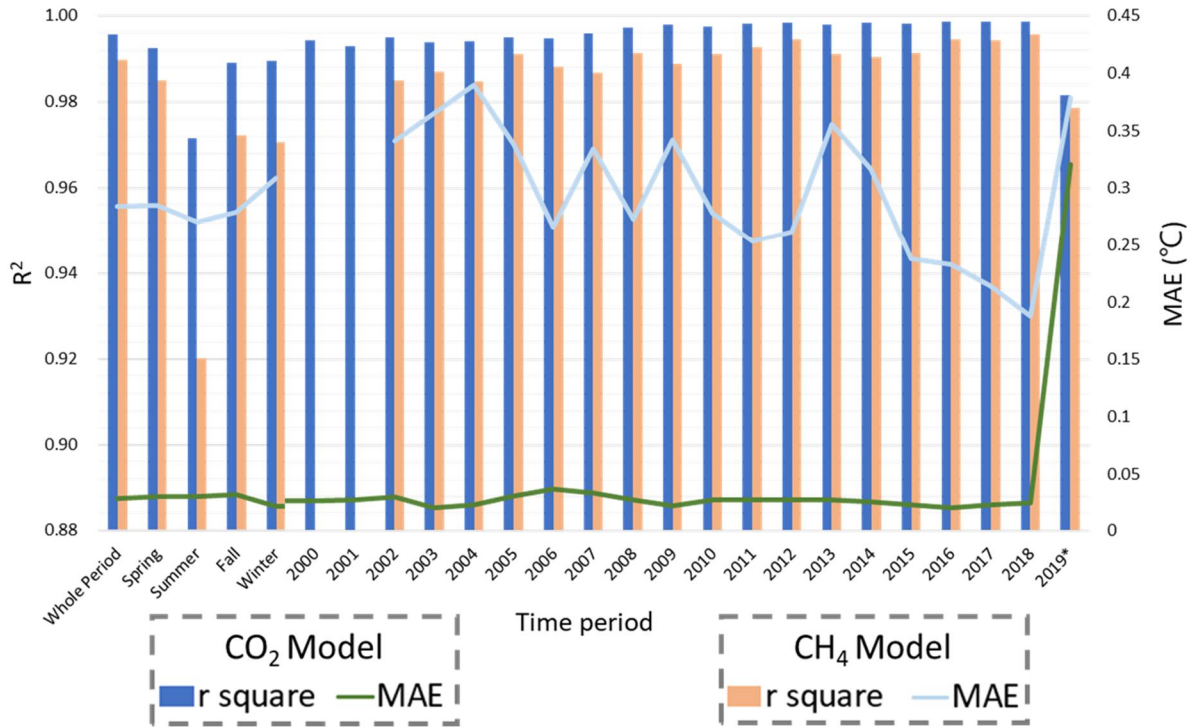


圖 4 溫度轉換模型之次分群驗證

與 CO₂相當 (Etmnan *et al.*, 2016)。根據最新報告，自 19 世紀末以來，CH₄對全球變暖的貢獻約為 0.5°C，而 CO₂的貢獻約為 0.75°C (Etmnan *et al.*, 2016)。總之，CO₂和 CH₄的變化對環境溫度產生影響，需要共同努力減少它們的排放，以應對氣候變化。另外，本研究還觀察到太陽天頂角和氣象因素等關鍵變數對環境溫度具有顯著影響。這一結果與 IPCC (2022) 的研究相契合，凸顯了溫室氣體排放與氣候變化之間錯綜複雜的相互關係。另外，本研究採用了衛星遙測影像的常態化植生差異指標，以反映全台大範圍長時間的綠地生長情形。相較於小範圍的林木調查或是航拍影像，這種方法可以獲取長期且大空間尺度的資料，進一步凸顯了航遙測資料的價值。在模型建立過程中，常態化植生差異指標被選為一個影響環境溫度的重要因子，這一事實表明衛星遙測的資料確實可作為預測因子，有助於準確預測環境溫度的模型建立。此外，憑藉機器學習模型的精準預測能力，可以更深入地理解未來氣候變化的走勢，這對於擬定有效的減排策略至關重要。

本研究存在幾個限制。首先，本研究僅聚焦於溫室氣體對環境溫度變化的推估。為了增強研究成果的豐富性，建議未來研究可以考慮在不同溫室氣

體濃度排放強度下估算環境溫度的變化。此外，未將溫度的衛星影像數據納入評估資料集，因為這些影像數據中存在大量因天氣條件（如雲和雨）引起的缺失值，可能影響模型的準確性。接著，鑑於研究數據主要來自西部地區的連續監測站，可能在東部地區的預測結果上會有疑慮。因此，建議未來的研究努力納入更多監測站的數據，如果可能的話。這不僅會增加數據的樣本量，有助於更好地訓練模型，而且可能擴大數據值的覆蓋範圍，提高濃度預測的可信度。最後，研究僅檢驗了有限的土地利用資料集，且僅適用於少數時期。未來研究應考慮使用更全面和多樣化的資料集，以提高估算模型的性能。同時，分析更長時間段的數據並考慮更多影響因素，將有助於進一步提升模型的預測能力和準確性。

本研究開發的模型不僅提供了一種新的方法來評估溫室氣體減排策略的有效性，還強調了考慮地理和氣象等多重因素的重要性。這一成果為臺灣和其他地區提供了更精確的氣候模擬和政策規劃參考。特別是在臺灣這樣的島嶼地區，深入了解各種環境因素對氣候的綜合影響，對於制定有效的策略以達成 2050 年淨零排放目標至關重要。此外，研

究方法和結果展示了大數據和機器學習技術在設計減排和氣候適應策略方面的潛力，強調了跨學科合作，尤其是環境科學與數據科學結合的重要性，以便更全面地理解和應對全球暖化的挑戰。儘管本研究提供了寶貴的見解和工具，但也認識到其局限性。未來的研究應擴展數據集，涵蓋更多地理位置和不同類型的環境因素，以提升模型的普遍適用性和準確性。此外，探索該模型在不同氣候條件和地理區域的應用，將有助於提高其在全球氣候變化研究中的實用性和有效性。

總之，本研究通過創新的機器學習方法，為理解溫室氣體排放對環境溫度的影響提供了新視角。這不僅對臺灣，也對全球的氣候變化研究和政策制定提供了寶貴的參考，為未來的相關工作指明了方向。

5. 結論

本研究成功考慮了涉及溫室氣體、環境和人為活動的相關變數，以建立一個能夠推估環境溫度變化的模型。研究使用描述性統計分析和五種機器學習演算法來驗證 CO₂ 和 CH₄ 轉換環境溫度的模型。研究結果顯示，LGBMR 演算法在 CO₂ 模型中表現最佳，而 CH₄ 模型則以 XGBR 演算法為優。透過 SHAP 值的分析，研究揭示了溫室氣體、濕度、風速等關鍵因素對模型預測的顯著影響。這些模型在多重驗證過程中展現出高度的準確性和穩定性，提供了對於溫室氣體與環境溫度關係的新見解，同時也為應對氣候變化提供了實用的工具。

參考文獻

- Al-Ghussain, L., 2019. Global warming: Review on driving forces and mitigation, *Environmental Progress and Sustainable Energy*, 38(1): 13-21, DOI: 10.1002/ep.13041.
- Brandsma, T., and Können, G., 2006. Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 26(1): 75-89, DOI: 10.1002/joc.1236.
- Etminan, M., Myhre, G., Highwood, E.J., and Shine, K.P., 2016. Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing, *Geophysical Research Letters*, 43(24): 12614-12623, DOI: 10.1002/2016GL071930.
- Hsu, C.Y., Wu, J.Y., Chen, Y.C., Chen, N.T., Chen, M.J., Pan, W.C., Lung, S.C.C., Guo, Y.L., and Wu, C.D., 2019. Asian culturally specific predictors in a large-scale land use regression model to predict spatial-temporal variability of ozone concentration, *International Journal of Environmental Research and Public Health*, 16(7): 1300, DOI: 10.3390/ijerph16071300.
- Intergovernmental Panel on Climate Change, IPCC, 2022. *Climate Change 2022: Impacts, Adaptation and Vulnerability, Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, DOI: 10.1017/9781009325844.
- Ramanathan, V., and Feng, Y., 2009. Air pollution, greenhouse gases and climate change: Global and regional perspectives, *Atmospheric Environment*, 43(1): 37-50, DOI: 10.1016/j.atmosenv.2008.09.063.
- UNFCCC, 2015. Historic paris agreement on climate change: 195 nations set path to keep temperature rise well below 2 degrees celsius, Available at: <https://unfccc.int/news/finale-cop21>, Accessed March 15, 2023.
- USGS, 2020. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid.
- Wong, P.Y., Lee, H.Y., Chen, Y.C., Zeng, Y.T., Chern, Y.R., Chen, N.T., Lung, S.C.C., Su, H.J., and Wu, C.D., 2021a. Using a land use regression model

with machine learning to estimate ground level PM_{2.5}, *Environmental Pollution*, 277: 116846, DOI: 10.1016/j.envpol.2021.116846.

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., 2021b. 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.

World Health Organization, 2021. Climate change and health, Available at: <https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health>, Accessed March 15, 2023.

Development of an Ambient Temperature Prediction Model Using Machine Learning by Integrating Greenhouse Gas Emissions, Vegetation Index Satellite Images, and Land Use Data

Hao-Ting Chang¹ Yinq-Rong Chern² Wan-Yu Liu^{3,4} Chih-Da Wu^{5,6,7*}

Abstract

This study integrated greenhouse gases, environmental, and anthropogenic variables, utilizing big data and five machine learning algorithms, including Random Forest (RF), Gradient Boosting (GBR), Light Gradient Boosting Machine Regressor (LGBMR), CatBoost Regressor (CBR), and eXtreme Gradient Boosting (XGBoost), to establish models for estimating ambient temperatures based on two greenhouse gases, CO₂ and CH₄. The LGBMR model performed best for CO₂, while the XGBR model showed better performance for CH₄. The R² values for the CO₂ and CH₄ estimation models were 0.993 and 0.999, respectively. Analysis of SHAP values confirmed the significant influence of greenhouse gas concentration, humidity, wind speed, and other factors on predictions. The findings of this study offer new evaluation methods for greenhouse gas emission reduction strategies and provide crucial insights for global climate change research and policy-making, highlighting the importance of interdisciplinary collaboration.

Keywords: Ambient Temperature, Carbon Dioxide, Methane, Machine Learning Predictive Model, Machine Learning Algorithms

¹ Postdoctoral Research Fellow, Department of Geomatics, National Cheng Kung University

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

³ Distinguished Professor, Department of Forestry, National Chung Hsing University

⁴ Distinguished Professor, Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University

⁵ Professor, Department of Geomatics, National Cheng Kung University

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

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

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

Received Date: Feb. 26, 2024

Revised Date: May. 9, 2024

Accepted Date: May. 31, 2024