

國立臺灣師範大學文學院地理學系

碩士論文

Department of Geography

College of Liberal Arts

National Taiwan Normal University

Master's Thesis

應用自組織映射神經網路於環境品質模型之建構：

以都市熱島因子資料為例

Application of Self-Organizing Map in Constructing
Environmental Quality Model:

A Case Study of Urban Heat Island Factors



劉宗妮

Liu, Tsung-Ni

指導教授：張國楨 博士

Advisor : Chang, Kuo-Chen, Ph.D.

共同指導教授：郭乃文 博士

Co-Advisor : Kuo, Nae-Wen, Ph.D.

中華民國 114 年 7 月

July 2025

摘要

本研究探討自組織映射神經網路 (Self-Organizing Map, SOM) 於環境品質評估的應用潛力，以都市熱島環境因子作為變數進行分群。過去環境品質評估以專家法為主，但當環境資料涉及多個面向的議題時較難以專家法訂定指數，故本研究以非監督式方法對環境資料進行分群，藉由資料本身的特徵區分不同環境樣態。相較於傳統統計方法在處理非線性與高維資料時的侷限性，以類神經網路分群方法 SOM 能有效映射高維資料至二維空間並保留拓撲關係，在視覺化與探索隱藏資料結構上具有更大的潛力。相似環境樣態的區塊本研究稱之為環境均質區，相較於綜合的指數，環境均質區可更直觀的應用於民眾理解及政策決策。

研究區選定縣市合併前的原台中市，整合以下變數：NDVI、NDBI、NDWI、LST、建蔽率、容積率、地表粗糙度、道路密度、與工業區及大型開放空間距離等 10 項變數，建置 SOM 模型。為評估模型有效性，以 Z-score 大於 0.5 作為都市熱區範圍與分群結果進行比對驗證，並透過 Random Forest (RF) 進行後驗的解釋與分群合併依據。研究結果顯示 SOM 能有效捕捉都市相對高溫區域，以多年分 LST 作為變數捕捉長期熱區範圍的正確性達 73%，以不同年份環境變數之環境熱區範圍正確性則介於 59-66% 間。

本研究以 2014 年 SOM 模型進行後處理，透過 RF 模型進行分群合併後以新分群作為訓練標籤，模型之 Accuracy 達 90.5%，Macro 平均 F1-score 為 90.4%，不同分群特徵重要性不同，顯示新分群在解釋邏輯上具有差異。綜合上述，以 SOM 作為分群方法並結合 RF 於探索都市環境熱區中展現良好適用性，可作為後續綜合環境品質評估模型建構之新方向與方法論基礎。

關鍵字：自組織映射神經網路、都市熱島、環境均質區

Abstract

This study discusses the potential of Self-Organizing Maps (SOM) for assessing the environmental quality, and using urban heat island (UHI) related factors as a case study. Traditional assessment methods are difficult to process complex, multidimensional data. SOM is an unsupervised method that maps high-dimensional data onto a two-dimensional space. It offers advantages in visualizing and identifying hidden environmental patterns. In this study, similar environmental patterns are called environmentally homogeneous zones. Compared to traditional composite indices, representing environmental quality through homogeneous zones provides a more intuitive way for the public to understand and for policymakers to make decisions.

The study area is located in Taichung, Taiwan. The study integrated ten variables: NDVI, NDBI, NDWI, building coverage ratio, floor area ratio, surface roughness, road density, and distances to industrial areas and large open spaces. Clusters were validated using a 0.5 standard deviation threshold of land surface temperature (LST). SOM successfully identified heat-prone areas, with long-term LST-based accuracy of 73%. The annual models, using the ten integrated variables, achieved accuracies ranging from 59% to 66%.

A post-analysis using the 2014 SOM model showed that reclassifying clusters with a Random Forest model yielded 90.5% accuracy and a macro-average F1-score of 90.4%. Variations in feature importance across clusters suggest distinct explanatory logic. The integration of SOM and RF is effective in identifying urban heat zones and offers a promising methodological foundation for future work.

Keywords: SOM, Urban Heat Island, Environmentally Homogeneous zones

目次

摘要	i
Abstract.....	ii
目次	iii
圖目錄	iv
表目錄	v
公式目錄	vi
第一章 緒論	1
第一節 研究背景與動機	1
第二節 研究目的	4
第三節 研究限制	5
第四節 研究範圍	6
第二章 文獻回顧	7
第一節 環境品質	7
第二節 傳統集群分析	9
第三節 深度學習	11
第四節 自組織映射神經網路	13
第五節 都市熱島因子	18
第三章 研究方法	22
第一節 研究流程	22
第二節 研究資料與資料前處理	24
第三節 模型建置	28
第四節 模型驗證	30
第四章 結果與討論	34
第一節 GWR 模型成果	34
第二節 SOM 模型成果	36
第三節 SCMC 模型成果	47
第四節 SOM 分群結果後處理	49
第五章 結論與建議	59
第一節 研究結論	59
第二節 研究建議	61
參考文獻	62
附錄一 SOM 驗證數結果-分群數於 300 至 350 群間	69
附錄二 One-vs-Rest RF 模型各分群 Feature Importance	70

圖目錄

圖 1 梧棲、台中測站溫度比較	6
圖 2 研究區範圍	6
圖 3 Self-Organizing Map 模型示意圖	13
圖 4 研究架構	23
圖 5 Random Forest 後處理流程	33
圖 6 2014 年 GWR 模型推估值與地真資料對照	34
圖 7 GWR 模型殘差分布與 Local R-squared	35
圖 8 SOM 分群結果（長期 LST）	36
圖 9 SOM 分群（長期 LST）二元化驗證	36
圖 10 長期都市相對高溫區範圍	36
圖 11 500 分群結果	43
圖 12 500 分群驗證結果	43
圖 13 SCMC 18 分群結果	47
圖 14 300 分群與驗證結果	47
圖 15 原始分群結果 Feature Importance	55
圖 16 合併新分群結果 Feature Importance	57
圖 17 原始分群結果與合併新分群結果空間分布對照	57

表目錄

表 1 分群模型示意圖	9
表 2 SOM 於各領域應用	16
表 3 環境模型選定影像日期	22
表 4 Landsat-8 衛星影像波段	24
表 5 變數資料及定義	27
表 6 參數設定	28
表 7 Random Forest 模型設定	31
表 8 溫度 SOM 模型驗證 (Pixel 數)	37
表 9 歷年選定影像之都市熱區界定範圍	38
表 10 SOM 環境熱區分群分布-分群數於 15 至 20 群間	40
表 11 SOM 環境熱區分群驗證-分群數於 15 至 20 群間	41
表 12 以 2014 年不同分群數驗證	42
表 13 SOM 環境熱區分群結果-分群數於 300 至 350 群間	44
表 14 SOM 環境熱區分群驗證-分群數於 300 至 350 群間	45
表 15 SCMC 與 SOM 分群結果驗證比較 (Pixel 數)	48
表 16 SOM 分群變數組內變異	50
表 17 分群合併列表與組內相似度	55
表 18 RF 測試集評估	56
表 19 新分群 Feature Importance 組成	58

公式目錄

公式 1	14
公式 2	14
公式 3	14
公式 4	14
公式 5	25
公式 6	25
公式 7	25
公式 8	25
公式 9	25
公式 10	26
公式 11	30



第一章 緒論

社會經濟、環境因子等與人類行為息息相關的空間資料時常會有很高的複雜度，往往呈現非線性分布且涉及不同結構的資料，在傳統統計研究上往往會造成困難。隨著人為開發、科技進步所帶來的環境改變，在不同的領域及政策上環境議題近年來逐漸受到重視，如何找尋適當的環境指標且有效量化環境品質的方法成為十分值得討論的議題。

第一節 研究背景與動機

隨著科技日新月異，人類加速的消耗資源、技術迭代以追求更加便捷的生活，但與之同時面對的是地球環境破壞造成負面影響，大自然的反撲帶來極端氣候、環境污染等問題，對人們的生活品質與健康而言是一大考驗。有鑑於資源消耗所帶來的環境變遷愈發顯著，環境議題成為重要的探討議題之一，如聯合國早在 2004 年提出之 ESG（Environmental, Social, and Governance）概念在國際間廣受重視，其中 Environmental 的層面注重環境永續的概念，致力於降低業界造成的環境汙染、碳排放等方面。聯合國永續發展目標（Sustainable Development Goals, SDGs）亦將環境議題列為全球發展核心，其中 9、11、13 與 15 項目標在基礎建設、永續城市、氣候行動與生態保育等議題中，顯示環境品質監測與永續治理的重要性。隨著環境品質越發被重視，廣泛建置監測空氣微粒、噪音分貝、水質等感測器也成為趨勢，使得環境資料的越發完整。另外也有許多指標的建立如空氣品質指標（Air Quality index, AQI），河川汙染指標（River Pollution Index）等用於量化分析各方向環境品質，環境的重要性可見一斑。

都市環境的生活機能較佳，卻需付出環境品質被犧牲的代價，除了噪音、空污等生活中常見的外部成本，有許多研究發現都市溫度高於周遭鄉村的情況越發明顯，Howard（1818）認為此現象來自都市建物易聚集熱能且散熱緩慢、缺乏綠地調節溫度，加上交通工具、工廠機具等人造熱能的大量增加，導致都市地表溫度快速

升高的狀況被稱為都市熱島（Urban Heat Island）。

由此可見環境因子十分複雜，有許多面向的資料需要考量，若要將其量化建構指標十分不易，過去在環境資料指標的建立提供民眾及決策參考時，常見的方式是使用層級分析法（AHP）等專家法建立指數，但當涉及多面向的環境資料時，專家法訂定出的指數較難以顧及其可解釋性，使用者較不易直觀了解其背後邏輯。另外環境資料往往十分複雜，可能會有非線性、高維度等問題，若是使用傳統的統計方法進行建模會造成準確性降低、結果難以解釋等情況發生。尋求一個適當的方法，將不同來源且屬性、特性也不同的環境資料整合，找出環境特性分布的樣態是個十分值得探討的議題。

為解決上述傳統統計常遇到的問題，近年來機器學習與延伸之深度學習技術相當盛行，分別有監督式與非監督式學習等兩大範疇。監督式學習是透過學習已知資料的特徵標籤分析資料，由於建構綜合環境指標其資料的複雜性及地域差異都可能造成影響，因此較不適用預先設置標籤的方法。而非監督式學習則根據資料本身特徵和結構進行聚合，對於處理較高資料維度、不具有資料標籤且複雜的環境資料而言可能是十分適合的方法。這些聚合而成具相似環境特性的空間區塊，本研究稱其為環境均質區。相較於綜合的指數，環境均質區能更直觀在空間上呈現不同環境條件區塊的分布樣態，有助於民眾理解環境樣態的區域差異，亦可為決策者提供明確的空間規劃參考。

深度學習（Deep Learning）其核心概念是透過神經網路的非線性映射來學習資料中特徵，相較於傳統的機器學習方法，深度學習在處理複雜、高維度且非線性資料時表現出更好的結果。自組織映射神經網路（Self-Organizing Map, SOM）該方法結合非監督式學習及神經網路的特色能有效處理高維資料，本研究欲應用 SOM 模型在資料視覺化與保留資料間距與關聯性的優勢，適合探索性分析找出環境均質區，做為未來探討更多面向環境品質模型建立的參考。

相較於其他環境資料，都市熱島因子具有較完整的研究基礎，過去都市熱島相關研究的環境因子包括探討土地利用及土地覆蓋與溫度間的關係，或是都市建物

樓高、容積率、距離工業區、道路、公園及水體等都市環境因素對溫度影響之研究。有關量化都市熱島範圍的研究已相當成熟，可利用其範圍作為地真資料（Ground Truth）來驗證模型的有效性，因而更適合作為 SOM 在環境品質指標建構上的案例，對複雜的都市熱島現象進行多變量分析，來識別都市區域中不同的熱環境模式，進一步為環境品質指標的建構提供新的方法論。



第二節 研究目的

由於環境品質資料的複雜性及高維度，傳統統計方法可能較難處理這些資料，本研究欲探討非監督式學習方法於建立綜合環境品質評估相關研究的可行性，其中過往研究認為 SOM 方法對高維資料探索與分群具有相當的優勢。因此本研究欲探討 SOM 方法於建構綜合性環境品質指標之可行性，並以都市熱島資料實作環境品質模型，作為評估環境品質模型可行性之驗證。分群成果中相似環境樣態的區塊稱之為環境均質區，相較於綜合的指數，環境均質區可於空間視覺化上提供民眾更直觀的解讀，並在政策治理上提供空間樣態的參考。

綜合上述，本研究欲達成之研究目的如下：

- 一、整合不同資料來源之都市熱島環境因子資料建立環境熱區 SOM 模型。
- 二、探討 SOM 模型於區分環境均質區之可行性與潛在優勢，並與 SCMC 模型進行實作比較。
- 三、建立尋找環境均質區的流程與驗證方法，最終研究結果可用於提升民眾對於環境空間狀態之理解，並在政策治理上空間參考的依據。

第三節 研究限制

本研究受限於取得資料內容、研究方法等因素，有以下研究限制：

- 一、 本研究所使用之環境因子資料有向量及網格資料，但 SOM 模型訓練需將其處理為相同型態、相同解析度的資料，本研究以 Landsat-8 影像資料解析度作為標準，將資料解析度統一為 30 公尺網格資料，因此經轉換後可能產生空間誤差。
- 二、 本研究之溫度資料為透過衛星影像資料取得地表溫度，與氣象觀測取得之大氣溫度可能會有不同，此外衛星影像可能會受雲層遮蔽而使地表溫度與地表真實狀態有誤差。
- 三、 本研究著重於分群結果與特徵映射，無法對模型中各環境因子的數值組成進行深入分析解釋，且研究結果之環境均質區是根據資料的特性聚合而成，其空間分布與體主觀認知的環境分區無法完全一致。
- 四、 原台中市少數地區時間變化較大且快速，部分資料可能無法提供連續性的資料，可能導致模型結果與現實出現誤差。

第四節 研究範圍

原臺中市位於臺中盆地內，由於夏季西南風吹拂，氣溫及濕度高，高溫約在六月至九月間，以臺中、梧棲測站為代表市區與沿海地區之長期測站進行比較，觀察 7 月均溫、當月最高溫、最低溫變化，可發現兩側站皆呈現溫度上升之趨勢，而台中測站之變化幅度明顯較大（圖 1），推測來自原臺中市開發較早，建物密度高、大量人工鋪面等因素，共同加深了溫度升高的幅度。

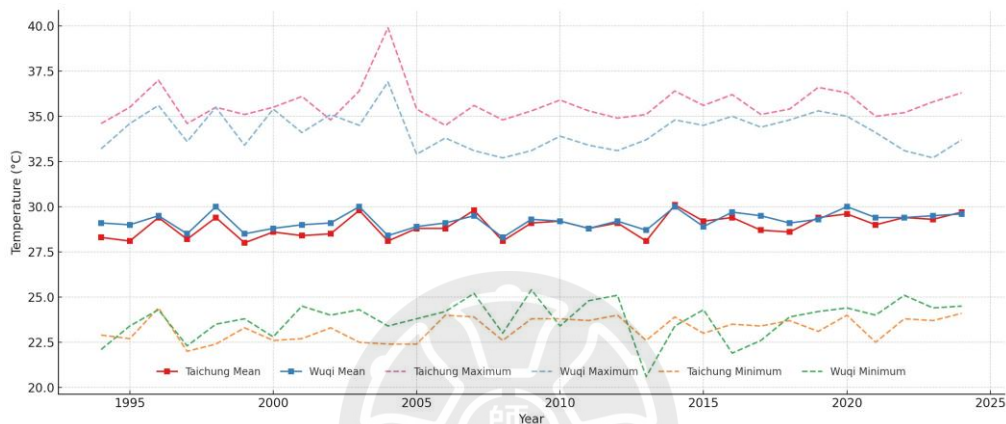


圖 1 梧棲、台中測站溫度比較

因此本研究將研究範圍選定於縣市合併前原臺中市範圍：中區、東區、西區、南區、北區、西屯區、南屯區及北屯區（以下簡稱原臺中市）。為維持研究區內都市區域地貌的均質性，排除於北屯區大坑風景特定計畫區，及北屯、西屯、南屯於 74 號快速道路以西，農業區、工業區為主的區塊。（圖 2）



圖 2 研究區範圍

第二章 文獻回顧

本章節首先回顧環境品質因子於現今研究的重要性與過去建立環境指標之模式，其次回顧深度學習非監督式學習與自組織映射神經網路方法的特色，並探討自組織映射神經網路於不同領域的應用及該方法於環境資料的選用優勢。其次回顧過去研究於都市熱島影響因子及量化應用的方法，初步了解可選用之因子資料。

第一節 環境品質

一、環境品質指標的發展

環境議題包括許多不同層面，包括碳排放、污染管理、資源永續利用、生物多樣性等。經濟合作暨發展組織（OECD）在許多報告書中提及有關環境問題的影響性，包括業界的永續性與環境保護呈現高度相關、人類污染暴露與健康影響及自然資源保護與生態等層面的環境議題（OECD environmental performance reviews），顯示了環境品質在經濟發展與生活層面中的重要性。

隨著工業化的發展各種環境惡化所帶來的影響也接踵而至，如空氣污染物可能會導致呼吸道疾病、心血管疾病、癌症等疾病，世界衛生組織（WHO）指出每年約 700 萬人死於空污影響（WHO, 2018）。交通運輸、工業設施、建築施工等活動產生的過量噪音，對健康也造成極大的影響，生活在機場或高速公路周邊的居民，長期暴露於高噪音環境導致中風與高血壓的風險顯著提升（Munzel et al., 2021）。

一直以來綠地在維持氣候穩定、生物多樣性與水資源循環中、人類生活品質等方面扮演重要角色，但隨著土地不斷開發，都市化與農業擴張導致綠地的銳減，影響生態系統的韌性、降低都市環境生活品質。

由此可見環境的劣化與經濟發展活動有著密切的關連，隨著全球人口增長與經濟活動的擴張，資源的消耗速度加快，使得環境負荷日益加重，這些對環境造成的傷害是不可逆的，考量到維護環境的永續性，如何在環境品質的維護與經濟發展之間取得平衡，可謂是現今最迫切的議題之一。

二、環境品質評估指標

為了有效改善環境品質，各國政府與國際組織採取政策治理、環境監測與技術等來達成環境永續目標，如頒布環境保護相關法令、建置感測器偵測相關資料、建立量化環境指標等。

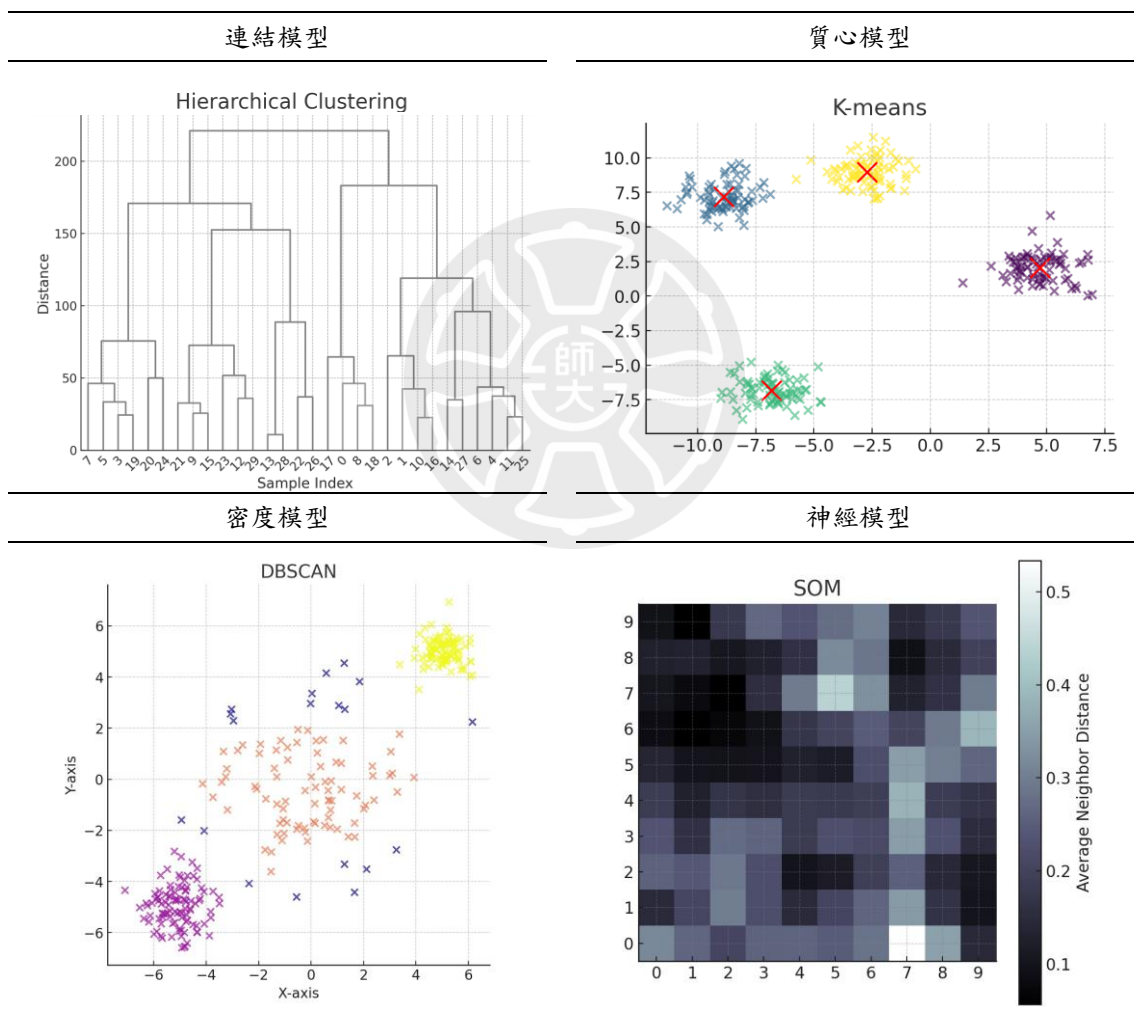
其中建立量化環境指標，有助於對環境狀態的追蹤與理解，過去研究大多以層次分析法（AHP）對環境品質進行量化分析，包括使用於建立空氣品質指標（Banerjee et al., 2018; Sahin et al., 2020）、都市環境品質指標（Nguyen et al., 2021; Lee & Xue, 2021）等。

由此可見過去研究環境品質指標的建立大多以專家法單一面向之環境資料為主，然而考慮到周遭不同面向的環境因子並將其量化，可更完整的分析環境品質。但若要將專家法應用於建立綜合性環境指標，由於涉及多層面環境資料時，專家法的答案是否適當往往是難以驗證，透過專家的標準作為答案進行學習可能會造成誤差傳遞。且這些綜合的指標在解釋性往往較為模糊，無法較直觀的了解環境樣態的組成與不同區塊間的異同，因此本研究欲使用集群分析方法讓資料特性自行群聚。

第二節 傳統集群分析

集群分析的主要目標為找出資料間的同質性（Homogeneity）與異質性（Heterogeneity），Cormack（1971）指出在良好的分群結果中，劃分在同分群的資料模式應相似、相異分群的資料應盡可能相異，根據不同的演算法邏輯可以將集群方法歸類為連結模型、質心模型、密度模型及神經模型等類別（表 1），適用於不同的資料型態與應用場景。

表 1 分群模型示意圖



連結模型基於距離的相似性來建立集群結構，其主要方法是層次集群（Hierarchical Clustering），此方法的優勢在於無需事先指定群數，透過計算樣本之間的距離逐步合併或分割集群，形成樹狀結構（Dendrogram）提供集群間的層級關係，再根據結果決定最後的分割的分群數。

質心模型則是以代表向量來描述每個集群的特徵，K-means 是最典型的方法，K-means 透過迭代方式計算每個群組的中心 (Centroid)，並將資料指派至距離最近的中心。該方法計算簡單且執行效率高，但需要預先指定群數 K ，且由於其演算法是將資料指派至距離最近的中心點，因此較適合用於球形分佈的資料，且結果易受到初始中心點選擇的影響，且在非線性資料上的表現較差。

在處理非線性資料的分群，則可考慮使用密度模型或神經模型。DBSCAN (Density-based spatial clustering of applications with noise) 透過鄰近點的密集程度來劃分集群，適合處理非線性或不規則形狀的資料結構，但 DBSCAN 在面對不同密度的群集時可能會出現判斷困難，並且會將密度較低的區域視為「異常點」或「雜訊」，並非適用於所有非線性資料的分群目標。神經模型則是利用神經網路來執行集群分析，自組織映射神經網路 (Self-Organizing Map, SOM) 是其中最經典的方法，其透過競爭學習法 (Competitive Learning) 來建立拓撲結構，使得高維資料映射至低維空間低維空間時保留拓撲關係，視覺化高維資料並保持局部相似性。

集群分析可針對資料的特性選擇適合的演算法進行運算，不必有資料標籤的訓練前置作業，所以可以避免人為的誤判，因此本研究欲選用集群分析作為環境資料建立綜合環境因子之方法。但由於缺乏標籤資料，也導致了分群結果的不穩定性，分群結果受到參數設定的影響極大，如事先設定的集群數、分群初始點的位置等，都會影響分群的結果，且即使參數的設定皆相同，仍可能每次分群的結果會有不同。此外在分群結果的解釋困難，各集群的意義以及其與研究目標的關聯性，需要專業知識輔助分析賦予分群結果意義。

第三節 深度學習

傳統的集群分析方法雖然在許多領域得到了廣泛應用，但在高維或非線性分佈資料上，這些方法的表現往往欠佳。隨著電腦運算功能的增加，使用深度學習的技術來解決問題成為近年來十分流行的解決方案。

一、深度學習的定義

深度學習是透過神經網絡 (Neural Networks) 模擬人類大腦的學習方式設計不同層數的神經元，使其能從大量資料中自動學習特徵，並反覆調整權重 (iteration) 來優化模型更有效地捕捉資料特性，經常被應用於在影像識別、自然語言處理及空間資料分析等大量非線性資料的領域 (LeCun, Bengio & Hinton, 2015)。深度學習的方法又可分為監督式學習 (Supervised Learning) 及非監督式學習 (Unsupervised Learning)。

監督式學習透過建立資料標籤 (Label) 來訓練模型，透過梯度下降 (Gradient Descent) 與反向傳播 (Backpropagation) 的方式不斷調整權重，獲取模型輸入特徵 (Features) 與標籤之間的最佳映射 (mapping) 關係 (Rumelhart, Hinton & Williams, 1986)，使模型能夠在最終測試資料中獲得良好的預測，可應用於進行分類 (Classification) 和回歸 (Regression)。

非監督式學習則不需要資料標籤，根據演算法的不同由輸入資料關聯性、距離、密度等進行運算找出資料間潛在的關係或模式。常見的應用包括透過分群 (clustering) 找出資料間隱藏的群體模式，對於高維度、具複雜性的資料進行降維 (Dimensionality Reduction) 在提升計算效率的同時保留最大的資料特徵等。非監督式學習亦作為探索性分析 (Exploratory Data Analysis, EDA) 的工具 (Abukmeil et al., 2021)，初步了解資料的結構並為後續研究提供更多的方向。

監督式學習與非監督式學習各有其適用的場景與領域，並在不同類型的問題中發揮重要作用。監督式學習被應用於影像分類、回歸預測等領域，資料的輸入與輸出間關係往往較為明確，能透過資料標籤來指引模型學習而獲取較高的正確率。

而非監督式學習則更適合用於資料探索與潛在結構的發掘，尤其適用於無法事先提供標籤或標籤取得困難的領域，但由於缺乏標籤資料，非監督式學習的結果可能較難解釋，並且在準確性往往取決於資料的結構以及演算法選用與設定的適當性。

二、深度學習方法適用性

本研究欲使用類神經網路捕捉特徵的優勢進行環境指標之建模，尋找環境同質區，因此監督式學習中的分類，與非監督式學習中的分群都是可能可使用的方法，但考量到環境品質指標難以建立資料標籤，因此非監督式學習中的分群方法較為適用。

SOM 是最經典 Neural-Based 的分群方法，其特色在於可將高維資料輸入資料映射至低維空間中保留拓撲關係，視覺化高維資料並在平面上保持高維空間時之局部相似性。在此背景下，本研究最終選擇非監督式學習中的自組織映射神經網路，其透過競爭學習演算法，能夠將高維度數據映射至低維度的同時保留資料間的相對空間關係，應用於探討環境同質區建立的可行性。

第四節 自組織映射神經網路

一、自組織映射神經網路

Kohonen 在 1982 年提出自組織映射神經網路 (Self-Organizing Map, SOM)，該方法基於類神經網路，透過模擬大腦神經元的演算法，將高維度的資料映射至二維的拓樸空間以達成分群及將特徵降至二維，根據 Kohonen (2013) 的整理，自組織映射神經網路在資料分析上可用於大型資料集的資料擷取與組織化、視覺化等功能。

自組織映射神經網路模型 (圖 3) 可分為輸入層與輸出層，變數 X 透過權重連結至輸出層的神經元，在訓練模型的過程中不斷調整權重，模型收斂可獲得其拓樸關係及分類結果。

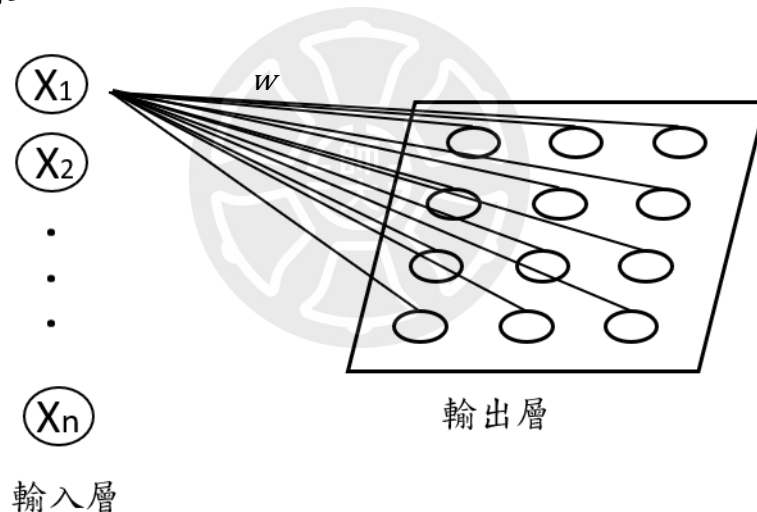


圖 3 Self-Organizing Map 模型示意圖

如同所有的深度學習方法，SOM 亦受到參數設定影響，由於 SOM 模型的結構相對簡單，因此對於太複雜的資料結構，如果沒有適當的參數設定，可能會造成收斂困難或是無法良好的劃分資料使得結果不佳，在 SOM 演算法中設定 SOM 網格的大小決定訓練神經元數量、鄰域函數公式及優勝神經元的影響範圍是其中最重要的參數設定。

在神經元進行權重調整前，需要根據競爭學習演算法，決定出優勝神經元 (Best Matching Unit, BMU)，只有優勝神經元及其鄰近神經元會進行權重調整，每個神

經元權重調整受到距離優勝神經元的遠近、使用鄰近函數的不同、學習率等因素所影響。

優勝神經元是透過歐基里德距離公式計算各輸入向量 X 與每個拓樸上神經元 i 的連結權重 (w) 的距離 (公式 1)，其中最小值即為優勝神經元。

$$D_c(x) = \operatorname{argmin} \|x - w_{cj}\| \quad \text{if } c \in i \quad \text{公式 1}$$

鄰近神經元調整權重之幅度與範圍由鄰近函數 η 所決定，以最常使用的高斯函數為例， r_q 為優勝神經元在拓樸上的位置， r_i 為其鄰近神經元的拓樸位置 (公式 2)，位於鄰近半徑 R 範圍內的神經元距離優勝神經元越近可獲得的權重調整幅度越大，隨著模型訓練次數增加，逐漸縮小 R 之範圍，每次影響的神經元數量則逐漸縮小。

$$\eta_{qi}(X) = \exp(-\|r_i - r_q\|^2 / R^2) \quad \text{公式 2}$$

Δw_{ij} (公式 3) 為最後權重調整變化值， μ 為學習率，學習率應隨訓練時間遞減，通常範圍介於 $0 < \mu < 1$ 之間。

$$\Delta w_{ij} = \mu \times (x - w_{ij}) \times \eta_{qi} \quad \text{公式 3}$$

綜合上述步驟，即可獲得每次訓練模型 (t) 調整之權重公式與權重變化 (公式 4)：

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij} \quad \text{公式 4}$$

隨者每次競爭學習法訓練模型調整神經元的權重，形成各神經元間的拓樸結構，逐漸獲得分群的資料結果。

二、應用領域

由於自組織映射神經網路方法在大型資料分析、視覺化的良好結果，醫學分析、金融資料、市場消費、異常檢測皆是常使用的領域。於醫學領域可應用於分析影像資料協助疾病的判斷，如對腦部磁共振影像資料的分群，用於腫瘤影像鑑別，分離出病灶如壞死和水腫區域 (Mei et al., 2015)、協助識別基因表達的模式和細胞分化過程中的重要基因和途徑 (Tamayo et al., 1999)。語音識別亦是一個經常使用的領域，用於語音訊號的特徵提取聲譜圖進行語音辨識，有優於傳統聲譜圖的穩定性、

直觀性及適用性 (Jia et al., 2021)。於金融資料可應用於時間序列的資料分析，如利用迴歸樹結合 SOM 提取歷史股價資料的隱藏模式，用於分析股票趨勢與投資推薦 (Nair et al., 2017)。異常檢測監控也是一個經常使用的領域，用以提高設備的可靠性，Concetti 等人 (2023) 應用 SOM 於油氣工業設備中感測器進行油氣提取生命週期分析，建立 6 個監控狀態分區，整體正確率達到 90%，由於其結果視覺化的易讀性，對於非專家操作人員可容易解讀成果。

地理資料亦有應用 SOM 的研究，Andrienko 等人 (2010) 以 SOM 方法結合互動式視覺化工具建立專家分析和探索資料的框架，透過城市交通資料及犯罪率資料進行實作驗證，適合用以探索複雜的空間與時間資料模式及找出隱藏模式。Arribas-Bel 等人 (2011) 以 SOM 進行歐洲城市蔓延的分析，使用連通性、密度、分散程度、開放空間可用性、土地使用六個指標作為訓練資料，其分群結果映射至地圖可成功反映出城市蔓延的模式分布。在氣候研究上，Hua 等人 (2022) 使用龍捲風相關環境參數用於分析其時空模式，透過第一次 SOM 將龍捲風機率分布分群來分類地理區，第二次 SOM 則透過第一次的分群結果針對每個環境因子異常機率進行分群。此套流程可辨識出具有相似龍捲風機率和環境參數異常之區域，提高對不同地區龍捲風風險的理解和預報能力。

整體而言，SOM 在多個領域有應用的價值 (表 2)，其在特徵提取、影像分析、尋找隱藏模式及視覺化等任務目標上達到十分良好的結果。在地理領域的應用上，可以發現 SOM 方法於資料探索的優勢受到研究者們的青睞，特別是在複雜的時空維度資料的應用，除了可以透過該 SOM 的分群結果對分群模式進行分析解讀，更可以結合其他方法達到互動式視覺化分析、提升預測準確度等目標。

表 2 SOM 於各領域應用

應用 領域	文獻	達成目標	應用 價值
醫學 分析	腦部磁共振影像資料判釋 (Mei et al., 2015)、識別基因表達模式 (Tamayo et al., 1999)	協助病灶判斷	特徵提取、 影像判讀
語音 分析	提取特徵聲譜圖 (Jia et al., 2021)	語音辨識特徵提取	特徵提取
金融 資料	迴歸樹結合 SOM 提取歷史股價資料隱藏模式 (Nair et al., 2017)	分析趨勢與投資目標	尋找隱藏模式、 特徵提取
異常 檢測	油氣提取生命週期分析 (Concetti et al., 2023)	找出異常因子	視覺化、 特徵提取
地理 資料	建立專家分析和探索資料的框架 (Andrienko et al., 2010)、城市蔓延模式分布 (Arribas-Bel et al., 2011)、龍捲風相關環境參數分析其時空模式與預測 (Hua et al., 2022)	高維資料探索	尋找隱藏模式、 視覺化、 特徵提取

三、過去研究 SOM 與其他方法之比較

當資料的規模和維度不斷增大，導致多元迴歸的結果變得難以解釋，機器學習方法則會有過度擬合使模型泛化能力不佳、分類結果準確性降低等問題 (Hastie et al., 2009; Domingos, 2012)，且為了提升模型效能、減少計算成本等，資料降維技術的重要性也隨之提升。此外，若分析含有非線性的資料往往會導致結果不佳缺乏解釋力，需要將資料經過額外的數學轉換才能使結果得到改善，然而選擇適當的數學方法轉換對於研究者往往是一大難題，因此能處理非線性資料的方法研究逐漸成為資料探勘的熱門議題 (Van et al., 2009; Jia et al., 2021)。此外，環境資料通常與周遭資料性質相似且呈擴散分布特性，SOM 的演算機制可將距離相近的地理資料分配到相同分群，因此相較於其他分群或非監督式學習方法，SOM 更適合用於劃分環境均質區。

與其他分群法相比，自組織映射神經網路其中一個優勢在於不需要事先給予分群數目，舉例而言，K-means 分群方法結果受到起始分群數目影響 (Pham et al., 2005)，然而由研究者決定良好的分群數目是不易處理的問題，涉及到經驗與足夠的專業知識，而自組織映射神經網路能透過模型訓練自行組成最終集群解決這個

難題。在適用資料上 SOM 在線性資料中的分群結果較 K-means 差，但在捕捉非線性資料的趨勢上能達到更好的效果，適合用於高維度資料探索性資料分析 (Mingoti & Lima, 2006)。K-means 有收斂在局部最優解、對初始條件 (centroids) 較敏感會影響分群結果的問題，SOM 可藉由其訓練過程中鄰域半徑和學習率的調整使得分群結果減少這些問題，達成更好、更穩定的分群 (Bação, Lobo & Painho, 2005)。亦有研究結合 K-means 及 SOM 的方法應用於提升高維資料分群的準確性，Mishra & Behera (2012) 首先透過 SOM 將高維資料降維，用於確認數據的拓撲結構和分群的數量，接著使用遺傳算法 (GA) 找出最佳初始質心再使用 K-means 分群，比起僅用單一方法，KSOM 可降低 K-means 對初始條件的敏感性，從而提高集群的準確性，此外亦結合 SOM 的視覺化優勢幫助研究者直觀的觀察資料模式與結構以決定分群數。

與常見降維方法之一的主成分分析 (Principal Components Analysis, PCA) 相比，SOM 亦作為降維方法之一，透過類神經網路的訓練，則能更好的處理非線性資料的降維。此外 SOM 在處理環境研究的常見缺失資料方面具有良好的穩定性，相比於 PCA 能更好的區分出環境變數間的關係 (Licen et al., 2023)。

Aguado 等人 (2008) 在對廢水去磷監測資料進行 PCA 與 SOM 兩種方法的降維研究中，發現兩者能達到相似的成果，但 SOM 在降維後結果的直觀性讓高維度資料間趨勢更容易被解讀分析。

四、小節

整理上述 SOM 之演算法、應用領域及其他方法比較之文獻後，可發現 SOM 對高維度資料的可達到更好、更穩定的分群結果，且更容易解讀降維後資料間趨勢。雖然過去研究認為 SOM 處理缺失資料、非線性映射的特性適合用於分析環境資料，但鮮少應用其於綜合性環境因子的評估並探討其於空間分布趨勢模型建置的可行性，因此以 SOM 模型建立環境均質區是個十分值得探討的議題。

第五節 都市熱島因子

一、都市熱島效應與量化

最早的都市熱島現象發覺於1818年英國氣象學家Howard在觀測倫敦氣溫時，發現倫敦地區的都市溫度高於鄉村溫度，此現象可能來自都市密集的建築、人造物導致熱量發散較慢，而真正使用都市熱島一詞於學術是1958年由Gordon Manley於英國皇家氣象學會學報中所使用。在都市化中人口集中、人造物及土地利用改變等因素會造成都市地區溫度的升高，其溫度立體剖面曲線高於周遭地區呈現鐘形分布，如同一座島嶼而被稱之為熱島（Chang, 2016）。

根據Rizwan等人（2008）的統整，都市熱島效應的成因可能包含交通工具及工廠運作等人為活動排放的熱能，以及太陽輻射的反射率由於都市建物的材質、密集地分布及綠地的減少而降低等原因。造成的影響除了高溫導致居民生活條件的惡化，亦會提高因高溫空調使用量帶來的能源消耗，高溫導致的熱中暑、熱衰竭等健康問題更是造成醫療資源的消耗，對身心皆可能造成傷害。

近年來隨著都市高速發展，人為開發對環境的破壞是無法避免的結果，都市環境劣化成為都市繁華背後的外部成本，探討都市熱島效應與都市內部溫度分布的關係成為人們關注的問題，使得都市熱島的量化與其影響因子及其量化的方法亦為研究重要的課題。都市熱島影響範圍的量化主要可分為空間自相關、熱點分析或律定比較標準值方式探討都市熱島效應的影響分布。

空間自相關（Spatial Autocorrelation）又可分為Global和Local尺度，Global Spatial Autocorrelation常見的方法為Moran's I和Geary's C，可用於評估整體空間的聚集或分散狀況，但無法分析個別區域的分布情形，因此前人研究多採用Local Spatial Autocorrelation，找出特定區域與其鄰近區域的高低值分布。張晏菁等人（2019）運用Local Moran's I分析臺中市地表溫度，指出高溫與低溫聚集區，並探討其與綠地和PM2.5濃度的關聯。

熱點分析則透過計算空間單元溫度與鄰近單元溫度是否為有高溫或低溫聚集

現象，常使用 Getis-Ord G_i^* 方法來辨別溫度聚集現象，如 Grigoros 和 Uritescu (2018) 利用 MODIS 衛星數據分析羅馬尼亞布加勒斯特都市熱島效應，根據 G_i^* 值和 Z 分數將溫度聚集分為「極冷區」至「極熱區」等類別。

律定比較標準值則透過設定門檻值界定異常高溫區域，Shih 等人 (2020) 以溫度高於或低於研究區平均溫度一個標準差界定高低溫聚集區，探討其與局部氣候區域 (LCZ) 的關聯性。李育寬 (2020) 則利用 Landsat 影像分析大台北都市熱島效應，探討 0.5 與 1 標準差下的熱島範圍差異，指出一個標準差最適合該區域影響範圍界定。

本研究的欲以都市熱島資料探討 SOM 建立環境同質區之效益，因此本研究欲量化都市熱島效應的範圍作為地真資料，驗證分群結果，本研究選擇以律定比較標準值界定都市熱島影響範圍，因研究目標為找出都市相對環境熱區而非極端熱島區塊之討論，本研究所應用之標準差範圍較一般熱島研究定義寬鬆，以 0.5 標準差為界定熱區之標準。

二、都市熱島因子量化與應用

過去多項研究找出使都市溫度升降之變數，主要影響的因素為氣候、地物覆蓋、土地利用及都市空間結構 (Zhi et al., 2020)，其中土地覆蓋、土地利用及都市空間結構造成溫度升高皆與都市發展過程中人為改變地貌有關，也是造成都市熱島效應的主因，本節將探討這些因素之因子量化與應用。

1. 地物覆蓋與土地利用

不同類型的地物覆蓋、土地利用會影響地表的吸收及輻散能力，Bokaie 等人 (2016) 針對土地利用與土地覆蓋及都市地表溫度間的關係進行探究，其研究結果發現土地利用及土地覆蓋類別中，水體、綠地、森林公園為地表溫度最低處，而裸露地、人工鋪面、工業區、商業區、住宅區及交通用地通常擁有較高的地表溫度。

量化土地利用及土地覆蓋的方式，部分研究採用常態化差異植生指數 (Normalized Difference Vegetation Index, NDVI) 及常態化差異建物指數

(Normalized Difference Built-Up Index, NDBI) 作為變數，分別代表地表植被覆蓋程度及建物覆蓋程度 (Shih et al., 2020; Zhi et al., 2020)，常態化差異水體指數 (Normalized Difference Water Index, NDWI) 則可用來量化水體分布情形，作為影響地表溫度的重要變數 (Xin et al., 2022)。

2. 立體環境因子

Rizwan 等人 (2007) 認為建物排放的輻射熱能是影響熱島效應強度的主因，建築物高度及分布密度連帶影響周遭散熱及風場，因此都市溫度的環境因子除了地物覆蓋、土地利用等平面地表特徵，亦須考慮都市立體空間的影響，如建築物的高度、密度、空間分布等 (Park, Ha & Lee, 2017)。

建蔽率與容積率可作為衡量建物在都市土地使用規範的兩個重要指標，建蔽率的定義為該土地可作為建物基地面積之占比，容積率的定義則為規定建蔽率內，建物樓地板面積的比率。潘豐家 (2015) 以建蔽率、容積率定義都市中心、都市及郊區後，探討熱島效應與建蔽率間的關係，其研究結果發現在郊區及開發中區域，建蔽率與熱島效應強度相關性高，而容積率則與高度開發地區的相關性最為顯著。Yu (2021) 等人發現容積率越高能代表都市建物密度的增加，在不同建物型態之都市容積率增高皆會加劇都市熱島效應。

地表粗糙度用以分析建物高度與開放空間的空間配置，其主要影響都市空間中風速及風廊的形成進而影響都市溫度。Park 等人 (2017) 以建物高度的標準差作為地表粗糙度變數進行都市內部三維空間與都市氣溫關聯性分析，指出地表粗糙度在都市區域中較高，且較高的地表粗糙度與較高的氣溫具有關聯性。建物高度標準差計算較簡易且直觀，可直接反映建物高度對於地表溫度的影響。而粗糙長度除考慮建物高度外，亦將空間中之空地、植栽的面積及高度納入考量 (林子平, 2021)，但植栽高度資料難以取得，因此本研究僅以建物高度標準差作為地表粗糙度參數。

3. 周遭環境因子

空間本身的土地利用與土地覆蓋外，周遭環境亦可調節或加劇地表溫度，研究指出綠地與水體均有助於降低周遭環境溫度並調節微氣候，常見的都市綠地包括公園、廣場等，綠地面積越大，降溫效果越明顯（石婉瑜，2018；羅子雯，2018；吳宜郡，2020）；水體的降溫效果則隨距離的減少而增強（Bokaie et al., 2016；Gao et al., 2022），過去研究指出 NDWI 與地表溫度呈現負相關，說明了水體有助於降低城市熱島效應的強度（Xin et al., 2022）。距離人造熱源越近則會導致周遭溫度的升高，若土地作為道路使用，車流的廢氣及排熱等會導致周圍溫度上升（Gao et al., 2022），工業區內密集分布的工廠於機具運作所排放出的熱能亦為影響都市溫度升高的因素（何佳薇等人，2019）。

三、小節

綜合以上前人研究的成果，本研究將以「NDVI」、「NDBI」、「NDWI」作為土地覆蓋類型影響因子，並以「建蔽率」、「容積率」及「地表粗糙度」作為都市三維空間結構影響因子。最後以「與大型開放空間距離」、「道路密度」及「與工業區距離」作為平面空間影響因子，建立環境因子模型訓練所需資料集。

第三章 研究方法

透過前面兩章的內容可得知過去在 SOM 與環境品質研究的趨勢及本研究使用的資料與處理方法，本章節闡述本研究欲進行的研究流程與應用方法、使用之資料及處理方式，以建立都市熱島環境品質模型。

第一節 研究流程

本研究首先透過文獻回顧獲取適合之都市熱島環境因子，將使用「溫度」、「NDVI」、「NDBI」、「NDWI」、「建蔽率」、「容積率」、「地表粗糙度」、「與大型開放空間距離」、「道路密度」及「與工業區距離」10 個變數作為都市熱島環境因子。透過為衛星影像可獲取地表溫度 (LST) 及 NDVI、NDBI、NDWI，其餘環境因子資料則來自政府開放資料。以 Landsat-8 之解析度作為統一標準，將資料前處理後所有資料之最終型態皆為 30 公尺解析度網格資料，並對所有資料進行 z-score 標準化。地真資料的產製則使用衛星影像資料進行輻射溫度反演獲取 LST，並以律定比較法門檻為溫度 0.5 標準差建置熱區門檻資料。

為了確認 SOM 環境模型在不同時間點環境狀態下的通用性，本研究將模型建置分為兩步，首先以 2013 至 2024 所有可用影像之 LST 資料建立 SOM 長期 LST 模型，第二步建置各年分之環境資料進行模型建置。考量政府開放資料之更新頻率約為一年，且更新時間不一致，本研究每年僅建立一組環境模型，並搭配該年度一張晴朗無雲之影像資料，最終納入之研究年度及對應影像時間如下表 3 所示：

表 3 環境模型選定影像日期

選定年分	2014	2015	2016	2018	2019	2020	2021	2022
Date	0901	0920	0922	1030	0729	0816	0828	0822

本研究欲建置三個模型進行分析，分別為 SOM 模型、Geographically Weighted Regression (GWR) 模型、Spatially Constrained Multivariate Clustering (SCMC) 模型，使用都市熱島因子資料各自訓練模型，並與地真資料驗證環境因子模型分群結

果，若模型表現未達預期，將進行模型超參數的調整重新訓練，研究流程如圖 4 所示。

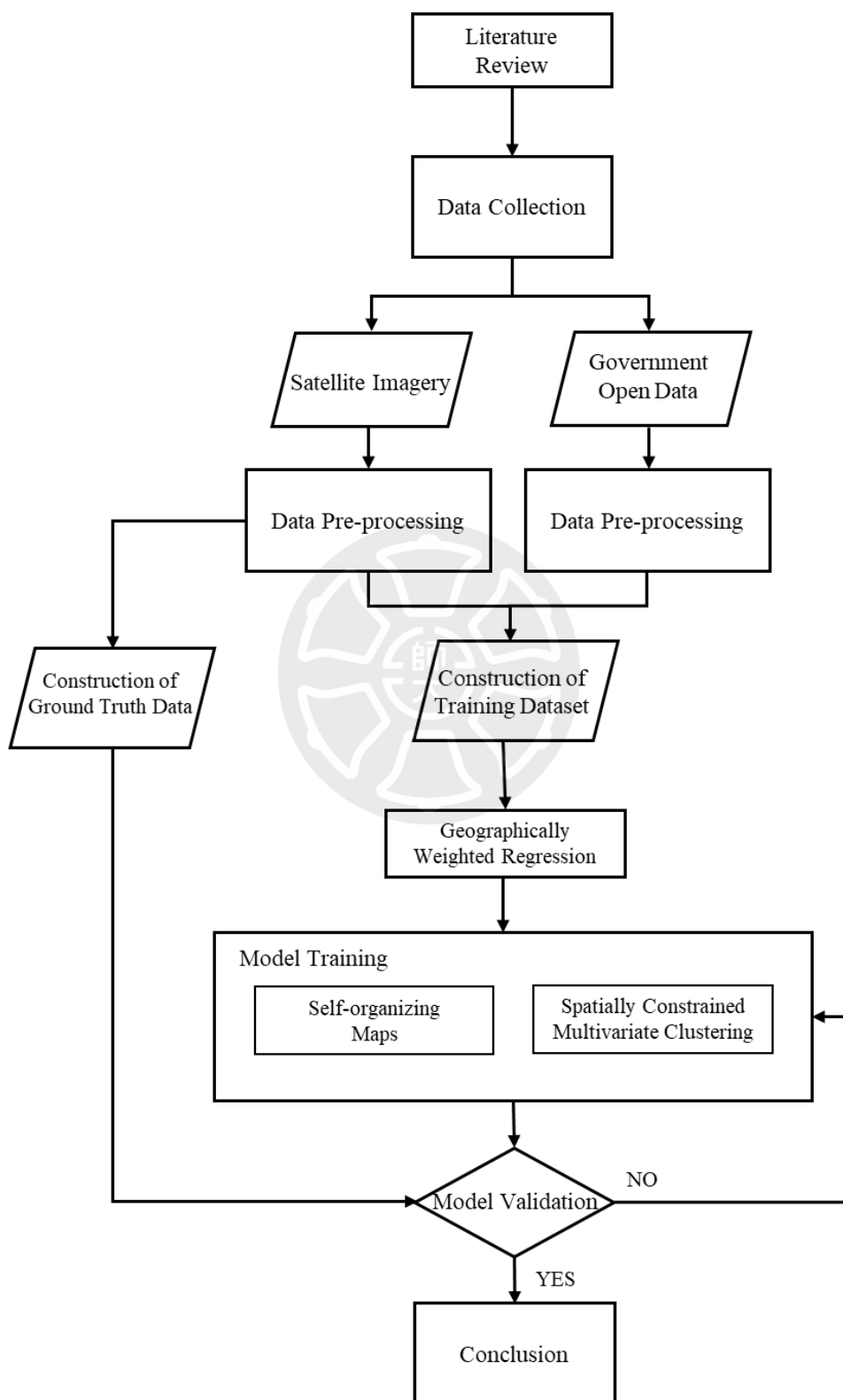


圖 4 研究架構

第二節 研究資料與資料前處理

本研究欲使用 2014 至 2022 年的 Landsat-8 影像及政府開放資料提取所需資料，為避免影像資料受大氣條件影響衛星影像將選擇雲覆率較低之影像，進行資料前處理並將資料型態皆轉為 30 公尺解析度的網格資料以進行 SOM 模型訓練。

一、衛星影像資料

本研究欲使用 Landsat 衛星拍攝之衛星影像波段作為研究資料，Landsat 系列衛星是由美國太空總署（National Aeronautics and Space Administration, NASA）於 1972 年起陸續發射的資源衛星，提供珍貴的影像資源可用於監測災害、農業、森林、海洋環境資料蒐集等方面。Landsat-8 為現今仍在運作之衛星，可見光與近紅外波段的空間解析度為 30 公尺，熱紅外波段解析度為 100 公尺，全色態波段解析度則是 15 公尺（表 4），時間解析度為 16 天。

表 4 Landsat-8 衛星影像波段

波段 (Band)	名稱	波長範圍 (μm)	解析度 (m)
Band 1	Coastal/Aerosol	0.43 - 0.45	30
Band 2	Blue	0.45 - 0.51	30
Band 3	Green	0.53 - 0.59	30
Band 4	Red	0.64 - 0.67	30
Band 5	Near Infrared	0.85 - 0.88	30
Band 6	SWIR-1	1.57 - 1.65	30
Band 7	SWIR-2	2.11 - 2.29	30
Band 8	Panchromatic	0.50 - 0.68	15
Band 9	Cirrus	1.36 - 1.38	30
Band 10	TIRS-1	10.6 - 11.19	100
Band 11	TIRS-2	11.5 - 12.51	100

資料來源：美國地質調查局 U.S. Geological Survey (2023) . Landsat 8 | U.S. Geological Survey.

1. 衛星影像指標計算

本研究之土地利用與土地覆蓋資料以常態化差異植生指數 (NDVI)、常態化差異建物指數 (NDBI)、常態化差異水指數 (NDWI) 作為代表，透過 Landsat-8 波段計算取得資料，其值範圍接介於 -1 至 1 間。

NDVI (公式 5) 計算中 NIR 為近紅外光波段 (Band 5)， R 為紅外光波段 (Band 4)。

$$NDVI = \frac{NIR - R}{NIR + R} \quad \text{公式 5}$$

NDBI (公式 6) 計算中 NIR 為近紅外光波段 (Band 5)， $SWIR$ 為短波紅外光波段 (Band 6)。

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad \text{公式 6}$$

NDWI (公式 7) 值計算中為 NIR 為近紅外光波段 (Band 5)， $GREEN$ 綠光波段 (Band 3)。

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad \text{公式 7}$$

2. 溫度反演

地表溫度資料使用衛星影像資料進行溫度反演計算，考量 Landsat-8 波段 11 之校正作業尚未完成，本研究選用僅需波段 10 之單窗演算法。

首先須計算輻射強度值 (L_λ) (公式 8)，其中輻射度乘數 ML 及輻射度加數 AL 為轉換常數，可透過 Landsat-8 衛星影像詮釋資料獲取， ML 值為 0.0003342， AL 值為 0.1， DN 值則為影像灰階值。

$$L_\lambda = ML \times DN + AL \quad \text{公式 8}$$

再將輻射強度值 (L_λ) 轉換為亮度溫值 BT (公式 9)， K_1 、 K_2 為波段溫度轉換常數，可由 Landsat-8 影像詮釋資料中取得，其中 Thermal Band 的 K_1 常數為 774.89、 K_2 常數為 1321.08。

$$BT = \frac{K_2}{\ln \left(\frac{K_1}{L_\lambda} + 1 \right)} \quad \text{公式 9}$$

將上述公式轉換為 LST (公式 10) 將單位由克氏溫度 (K) 轉換為攝氏溫度 (°C)，再加上絕對零度-273.1 進行溫度單位轉換。

$$LST = BT - 273.15 \quad \text{公式 10}$$

二、政府開放資料

1. 立體環境變數

本研究透過臺中市政府開放資料平台取得立體環境變數資料，透過都市計畫圖擷取建蔽率、容積率，若資料內容為「空值」或「不予規定」大多為規範用地如綠地、停車場等，則將值輸入為 0，並以平均計算將其轉換為 30 公尺網格資料。

地表粗糙度則以建物圖資計算建物高度之標準差作為依據，但建物圖資僅提供樓層資料因此以此推估高度資料，假設每層樓高為 3 公尺，將樓層數轉換為樓層高度，以反距離加權插值法 (Inverse Distance Weighting, IDW) 計算研究區域中每 30 公尺網格中之平均樓層高度，以 9×9 的滑動窗格進行目標網格與周圍網格之建物高度的標準差計算。

2. 周遭環境變數

每筆資料與工業區、大型開放空間的最近距離，將研究區範圍資料由向量資料轉為 30 公尺網格資料後，均以歐基里德直線距離計算，資料皆來自都市計畫圖，與大型開放空間距離選定臺中市都市計畫範圍中計畫使用分區為「公園」、「綠地」、「廣場」、「園道」及「兒童遊樂場」且面積大於 1.5 公頃之用地。道路密度本研究使用臺中市道路圖資作為依據，計算 30 公尺網格內道路長度總和與面積之比值計算。

綜合上述，本研究所使用變數與定義如下表 5，共 10 個變數，經處理後皆為 30 公尺解析度資料，後續再將其進行資料標準化等步驟以利建置訓練資料集。

表 5 變數資料及定義

因子名稱	資料來源	定義
LST	Landsat-8 衛星影像	$\frac{K_2}{\ln\left(\frac{K_1}{L_\lambda} + 1\right)} - 273.15$
NDVI	Landsat-8 衛星影像	$\frac{NIR - R}{NIR + R}$
NDBI	Landsat-8 衛星影像	$\frac{SWIR - NIR}{SWIR + NIR}$
NDWI	Landsat-8 衛星影像	$\frac{GREEN - NIR}{GREEN + NIR}$
建蔽率	臺中市都市計畫圖	建蔽率平均值
容積率	臺中市都市計畫圖	容積率平均值 (或取最高上限容積率)
地表粗糙度	臺中市建物圖資	建物高度標準差之平均
與工業區距離	臺中市都市計畫圖	至工業區最小直線距離
道路密度	臺中市道路圖資	網格內面積道路長度佔比
與大型開放空間距離	臺中市都市計畫圖	至大型開放空間最小直線距離

三、資料集建置

本研究的資料蒐集可分為使用衛星影像資料與開放資料，衛星影像資料之 NDVI 由於研究區皆有被衛星軌道覆蓋，因此全區皆有資料。但與建物分布位置相關的建蔽率、容積率及地表粗糙度為向量資料，僅於建物所在空間有資料，無建築區域則為大量空值。道路密度資料則有大量值為 0 可能影響分群結果。為改善這些問題，本研究採用核密度估計 (Kernel Density Estimation, KDE) 對建蔽率、容積率與道路密度進行平滑化處理，使原本在空間上呈離散分布的資料能透過密度估算方式展現較為連續的空間分布特徵。

此外考量 SOM 演算法需計算輸入資料與神經元間的距離進行分群，導致在 SOM 模型訓練時數值較大的變數影響較大。本研究各資料範圍存在較大的數值差異，導致影響訓練結果。因此本研究對所有資料進行 Z-score 標準化，將各變數轉換為平均值為 0、標準差為 1 的標準化數值，以去除變數尺度不一致的影響，確保各變數在訓練過程中具同等影響力。

第三節 模型建置

本研究除了使用 SOM 模型外，並建置 GWR 模型及 SCMC 模型，評估三種模型在熱島分析目的之適用性。

一、GWR 模型

GWR 模型是區域迴歸模型，相較於傳統的普通最小平方法 (Ordinary Least Squares, OLS)，GWR 的局部偏迴歸係數可根據區域有不同的權重，描述出不同空間上變數的變異性，比起傳統迴歸，GWR 能增加預測局部估計值的準確度與解釋性。本研究首先將使用環境變數作為自變數，LST 作為應變數，迴歸的結果用於探討整體環境變數在溫度估計具有解釋力再進行後續分群模型的探討。

二、SOM 模型

SOM 模型參數設定如下，SOM Grid 決定了 SOM 訓練神經元的數量，越多的 SOM Grid 可以學習更多不同的特徵映射，但同時會提高時間成本降低訓練效率，若已知資料複雜度較高可選擇較多的 SOM Grid 開始訓練，相反則可選擇較少的 SOM Grid 提高訓練效率，因此本研究設置兩組參數進行模型訓練（表 6）。本研究設定 Neighborhood 初始權重影響範圍為 25，隨訓練過程遞減至 1。

表 6 參數設定

參數名稱	設置參數	
SOM Grid	10×10	50×50
Epoch	8000	
Initial Neighborhood	25	
Final Neighborhood	1	
Initial Learning Rate	0.1	
Final Learning Rate	0.01	

三、SCMC 模型

Assunção 等人 (2006) 提出了可用於將社會經濟地理單元劃分為具備空間連續性與內部同質性的演算法 SKATER (Spatial 'K'luster Analysis by Tree Edge Removal)，SCMC 模型即是應用該演算法而來，該方法不僅考慮變數間的相似性，亦納入空間鄰近性約束，以避免傳統分群方法可能導致的空間碎片化 (Spatial Fragmentation) 的問題，使其分群呈現空間上連續變化。其目標與本研究欲將環境分為具備空間連續性與內部同質性的概念相似，因此本研究欲使用該方法作為比較模型。



第四節 模型驗證

根據前述文獻回顧，本研究選定率定標準值作為都市熱島範圍的界定，由於研究目標為區分出都市相對環境熱區而非詳盡分析都市熱島位置，因此在驗證範圍的界定較為寬鬆，以大於 0.5 個標準差作為都市相對高溫區範圍的標準。為確保熱區範圍不受到單一影像之天候影響熱區分布，本研究選用 2013 至 2024 年間所有可用影像反演之 LST，找出長期高於 0.5 標準差之區域作為第一步驗證 SOM 模型之結果。各年份之模型則以該模型使用之 LST 資料大於 0.5 個標準差作為該模型都市相對高溫區範圍之驗證標準。

一、分群結果驗證

由於本研究分群結果較多難以用肉眼判斷成果，為了視覺化分群結果，本研究將分群合併為二元分類，在後續分析以環境熱區分群為探討之重點。

環境熱區的定義為若該分群超過 50%面積落在地真資料之熱區則被歸類於環境分群熱區（以下簡稱環境熱區）反之則為非環境分群熱區（以下簡稱非環境熱區），將分群分為誤判區塊(Commission)、失判區塊(Omission)、正確區塊(Correct)數量作為衡量模型結果與地真資料間的評估，衡量三者間的占比做為評估分結果的依據（公式 11），僅關注環境熱區的 Correct 區塊與 Commission 區塊、Omission 區塊之間的相對關係。

$$Correct\ cluster\ rate = \frac{Correct}{Omission + Commission + Correct} \quad \text{公式 11}$$

二、分群結果評估

1. 敘述統計

在分析分群的有效性前，應先確認分群內部的一致性（internal cohesion），再進一步檢驗各分群間是否在各變數上呈現顯著差異。群內資料分布若高度分散代表其內部異質性高，不適合作為環境樣態的代表。本研究首先計算各分群在變數上的組內標準差（within-cluster standard deviation）評估分群結果的內部

一致性 (internal cohesion)，數值越低代表該群樣本在該變數上的分布越集中，亦即群內成員於該變數上的差異程度較小。由於研究資料已事先進行標準化處理，因此組內變異可直接作為跨變數之比較基礎。

確認分群內部的一致性後，再進一步檢驗各分群間是否呈現顯著差異，由於本研究分群數量較多，因此擬在分群合併後再進行組間差異的探討。

2. 後處理評估

由於都市熱島現象是一個概念，其判定並無統一標準，本研究所定義之 0.5 標準差門檻僅為驗證 SOM 模型是否能捕捉熱區空間趨勢的暫訂標準。為進一步分析分群間的因子差異以及對分群結果進行合併與評估，本研究採用 Random Forest 模型邏輯作為後驗分析，了解熱區分群組間差異。

本研究以符合 0.5 標準差門檻之 SOM 環境熱區分群結果作為訓練樣本區塊，建立熱區分類 RF 模型，並透過 GridSearchCV 方法調整參數，以尋找最佳模型設定組合（表 7）。

表 7 Random Forest 模型設定

Input X: SOM 環境熱區變數（環境因子）		
Output Y: SOM 環境熱區分群		
Hyperparameter	Name	value
	Max depth	15
	Max features	sqrt
	Min samples split	2
	N estimators	300
	Class weight	balanced
	Random state	42

本研究 RF 研究流程分成三個部分，首先建置環境熱區 RF 模型，並以其 feature importance 了解，其次為 One-vs-Rest RF 模型所產生的 feature importance 進行相似性指標分析，並將相似分群進行合併，再針對合併後分群進行評估。

相似性指標分析上，本研究參考 Wang 等人（2016）對 RF 模型中「特徵重要性向量穩定性」的探討。該研究透過多次訓練隨機森林模型，並以 Spearman 與 Pearson 等相關係數，量化各次訓練所產生的特徵重要性向量

(feature importance vectors) 間的相似程度，以評估模型在不同條件下是否維持一致的特徵判斷邏輯，亦即模型對同一資料是否給出穩定的解釋。

然而本研究關注的並非模型本身的穩定性，而是不同分群之間是否展現出相近的解釋結構。即為在都市熱島的不同分群中，是否存在類似的環境因子組合，對熱島現象扮演相同的解釋角色。為此將多個 One-vs-Rest RF 模型所產生的 feature importance，視為各分群的解釋特徵結構，並進一步比較其間的結構相似性。

為將相似特徵結構之分群進行合併，本研究選擇使用 cosine similarity 作為相似性指標。近年來可解釋人工智慧 (Explainable AI, XAI) 領域廣泛應用 cosine similarity 作為評估模型解釋邏輯差異的工具。特別是在 SHAP (SHapley Additive exPlanations) 等模型解釋方法中，研究者常透過計算 SHAP 向量間 cosine similarity，來評估不同樣本或模型對於輸出結果是否具備一致的解釋邏輯 (Lundberg et al., 2020; Covert et al., 2021)。相較於 Pearson correlation 著重於數值的線性變化，cosine similarity 著重於向量的方向一致性代表各特徵在解釋邏輯中相對排序是否一致，因此更適合應用於本研究欲合併之解釋邏輯相似分群之目標。若發現某些分群的解釋邏輯相近，則可進一步考慮合併標籤，根據過往研究，cosine similarity 值高於 0.7 可作為判定高相似的門檻 (López Zapata et al., 2018; Dey et al., 2020)，本研究亦以此門檻值作為分群合併的依據，將相似分群進行標籤合併後重新訓練 RF 模型，依 6:4 比例將新標籤資料劃分為訓練集與測試集，並以測試集結果之 Accuracy、評估合併後分群的解釋力與預測表現，並透過 feature importance 了解每個分群的組成邏輯差異。

綜合前面所述，本研究的後處理流程如圖 5，若新分群的驗證結果不如預期，則會再重新進行合併門檻分析並再次重複研究流程。

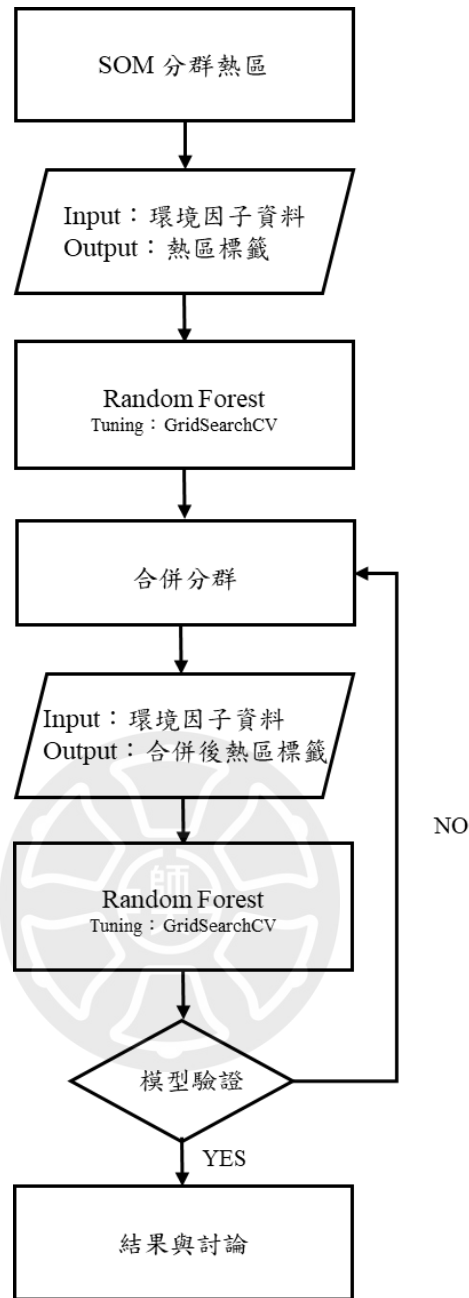


圖 5 Random Forest 後處理流程

第四章 結果與討論

第一節 GWR 模型成果

GWR 是都市熱島推估常使用的研究方法，以 2014 年資料為例（圖 6），R-squared 及調整後 R-squared 代表 GWR 模型整體可解釋約 85% 的變異量，並把模型推估的溫度值 Z-score 大於 0.5 之區塊與 2014 年地真資料進行對照，在空間分布上的趨勢差異不大，本研究所使用的變數在整體推估熱區的趨勢上具有一定的解釋力。

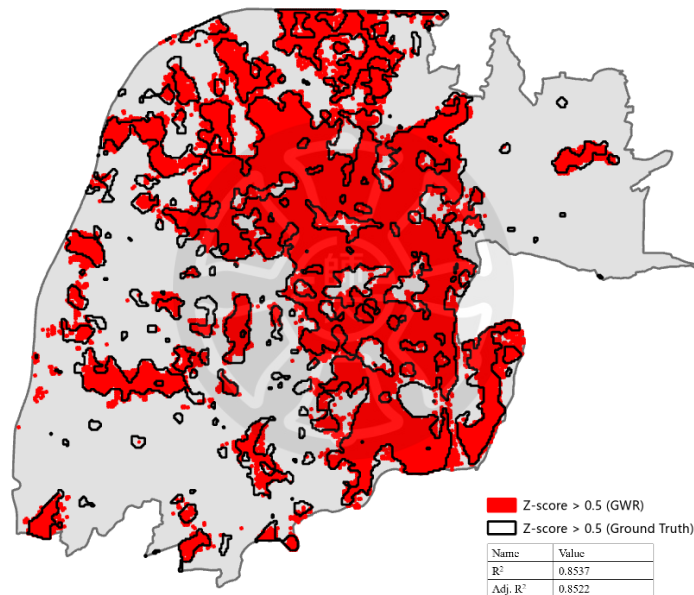


圖 6 2014 年 GWR 模型推估值與地真資料對照

GWR 模型在空間統計中廣泛使用，其本身仍具有一定的模型限制，其模型屬於線性推估，若資料為非線性分布或資料間共線性會影響模型推估的結果，由圖 7 可見殘差分布與 Local R-squared 在部分區域的解釋力表現不佳。GWR 模型可產出每個位置的偏迴歸係數，進而針對熱島區塊不同解釋變數組成的區域，但其局部的估計結果為數值之預測及異質性區塊的變數組成與解釋，難以直接區分具空間連續性的環境同質區，本研究認為該方法可與 SOM 模型的用途互補，SOM 模型無法直接看出每一個區塊的變數組成與解釋，但可藉由劃分連續的環境熱區幫助研究者探索空間趨勢。

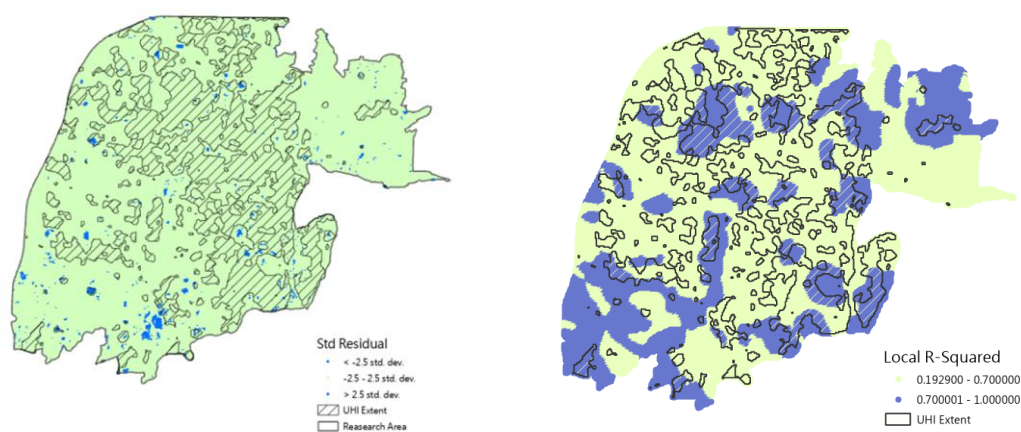


圖 7 GWR 模型殘差分布與 Local R-squared



第二節 SOM 模型成果

一、溫度 SOM 模型

挑選 2013 至 2024 所有可用衛星影像 B10 進行溫度反演計算為 LST，並進行 SOM 模型訓練後，最終結果分為 18 分群（圖 8），將其合併為二元分類（圖 9），與長年平均標準差高於 0.5（圖 10）之都市相對高溫區進行比對驗證。

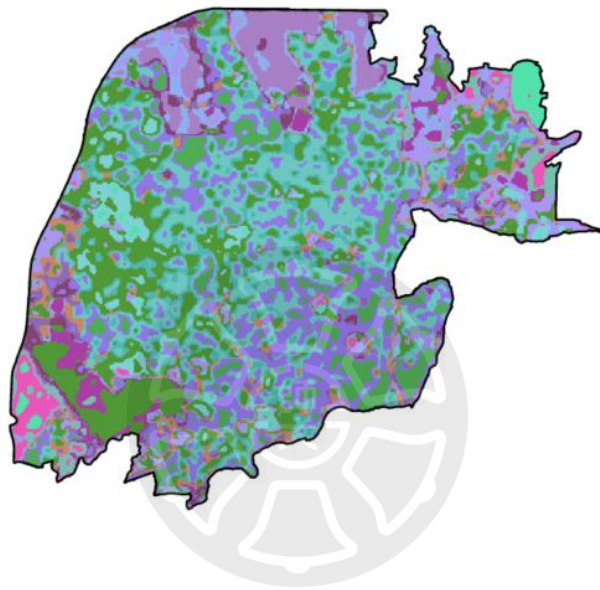


圖 8 SOM 分群結果（長期 LST）

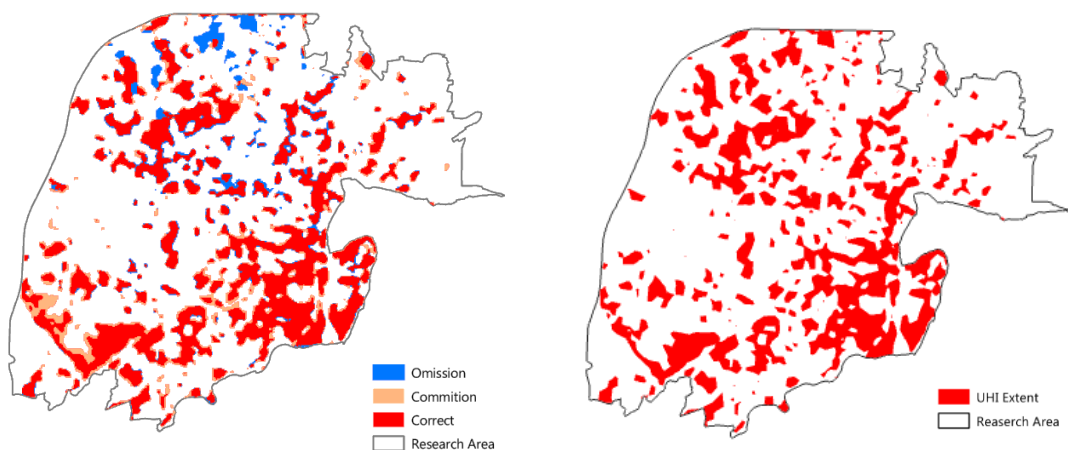


圖 9 SOM 分群（長期 LST）二元化驗證

圖 10 長期都市相對高溫區範圍

與長期都市相對高溫區範對照後，可計算其分群之二元分布（表 8），熱區之

Omission 為 0.1764，Comission 為 0.086，Correct 為 0.7376，由此可見 SOM 可良好的區分出都市相對高溫區。

表 8 溫度 SOM 模型驗證 (Pixel 數)

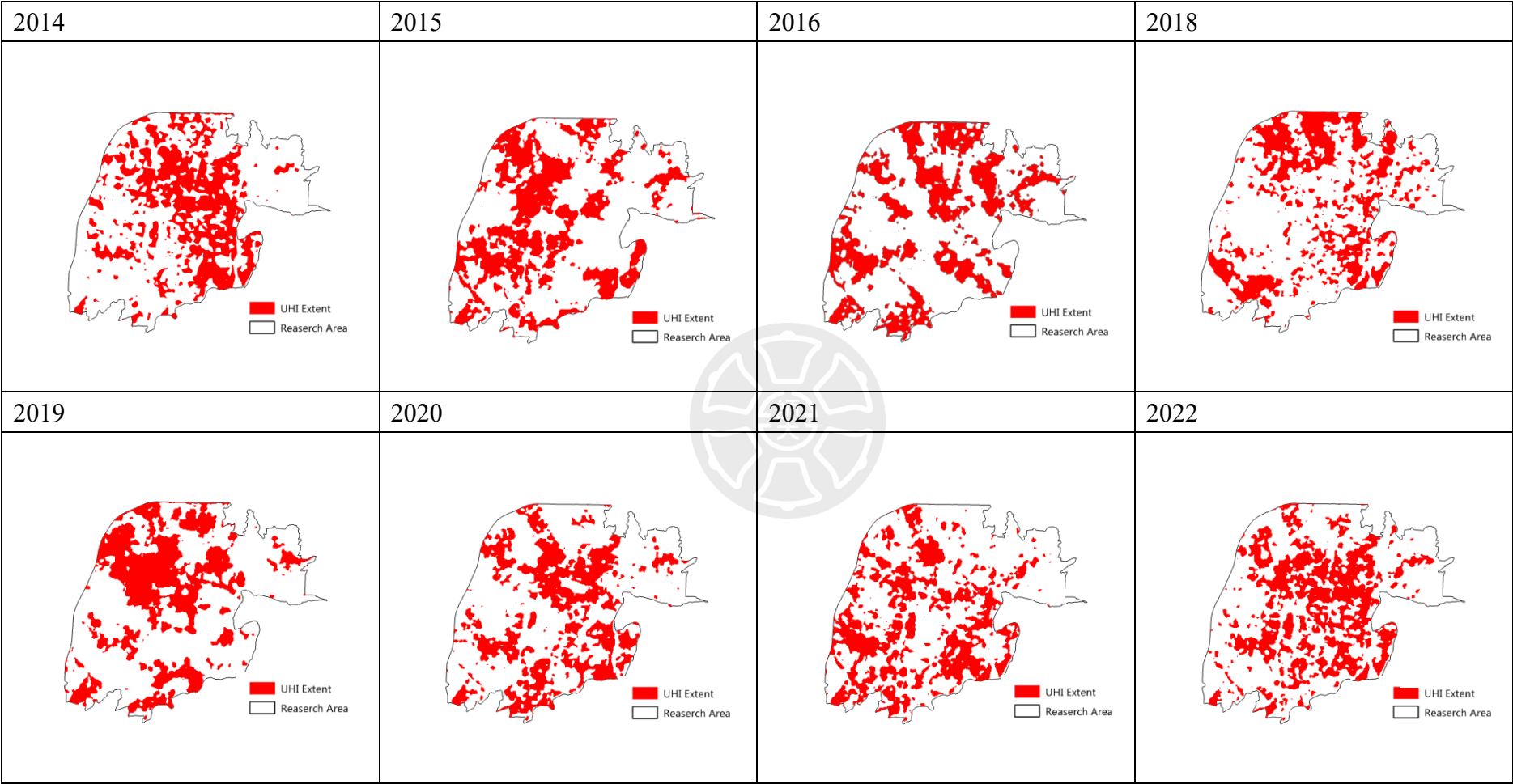
ALL Years LST	環境熱區分群	非環境熱區分群
地真資料熱區	20461	4902
地真資料非熱區	2412	70397

Correct = 0.7376; Omission = 0.086; Commission = 0.1764

確認 SOM 可應用多年分之溫度特徵良好區分出長期都市相對高溫區後，本研究將進一步探討 SOM 應用於環境因子之通用性，因此本研究先根據所使用之衛星影像建置之地真資料（表 9），以便後續驗證驗證環境因子模型。



表 9 歷年選定影像之都市熱區界定範圍



二、環境 SOM 模型

經過 LST 模型可確認 SOM 可良好的捕捉特徵，分群之空間分布吻合都市相對高溫區。於是本研究加入環境因子依年度訓練 SOM 模型（表 10），先以與 LST 相同之參數設定模型。由於加上環境因子的變化較大較不穩定，因此各年度的分群結果分為 15 至 20 類時，驗證結果不甚理想（表 11），多數年份之 Correct 落在 0.4 至 0.5 間，但仍可看出環境分群之空間分布可根據當下的狀態有所變動，可看出不同年分區塊性的差異。但可發現在 2021 年之分群結果較差，經過分析發現與其衛星影像時間點位於颱風過境時間，因此在整體大氣環境會有更高的複雜度，而導致分群結果之正確率特別低。

由 LST 模型至環境因子模型之差異可知推測，模型僅能捕捉大致空間變化而無法捕捉到細緻的環境因子變化，與設定較少之 SOM Grid 或分群數無法區分出有效空間有關，本研究調整 SOM Grid 增加權重特徵的數量後分群結果無明顯改變，因此本研究改以增加權重特徵數量外亦增加分群數來改善分群正確率。

表 10 SOM 環境熱區分群分布-分群數於 15 至 20 群間

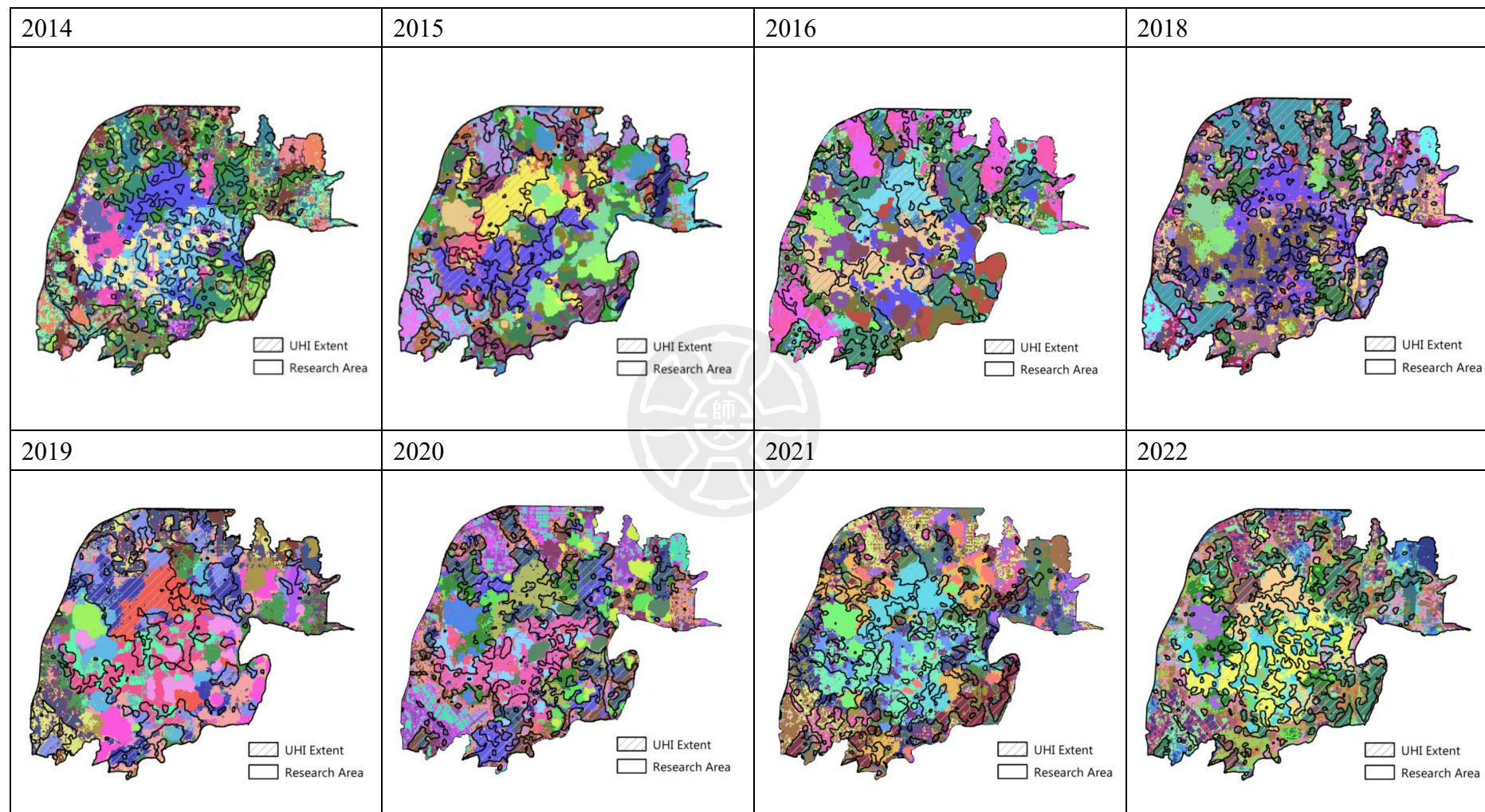
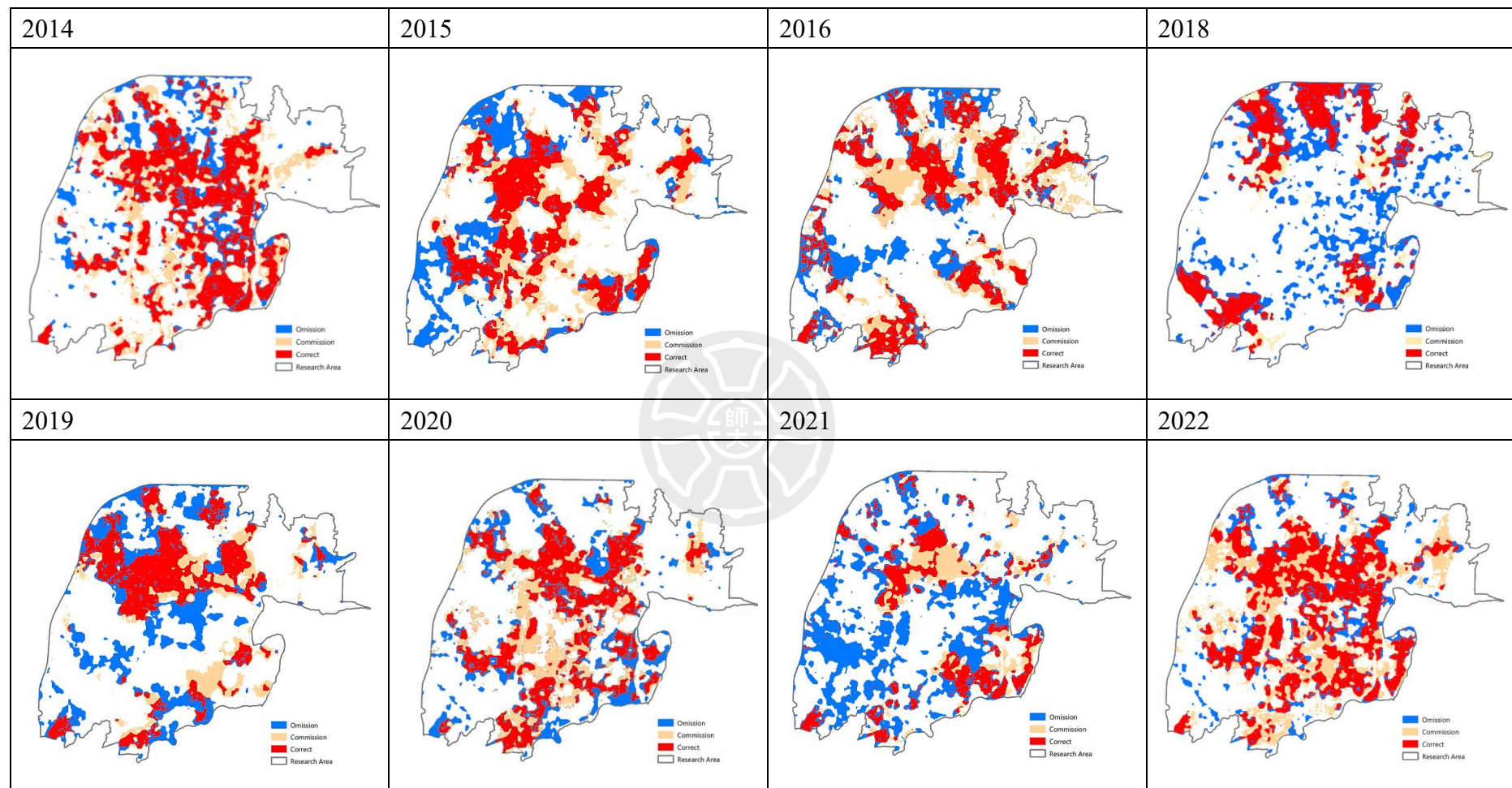
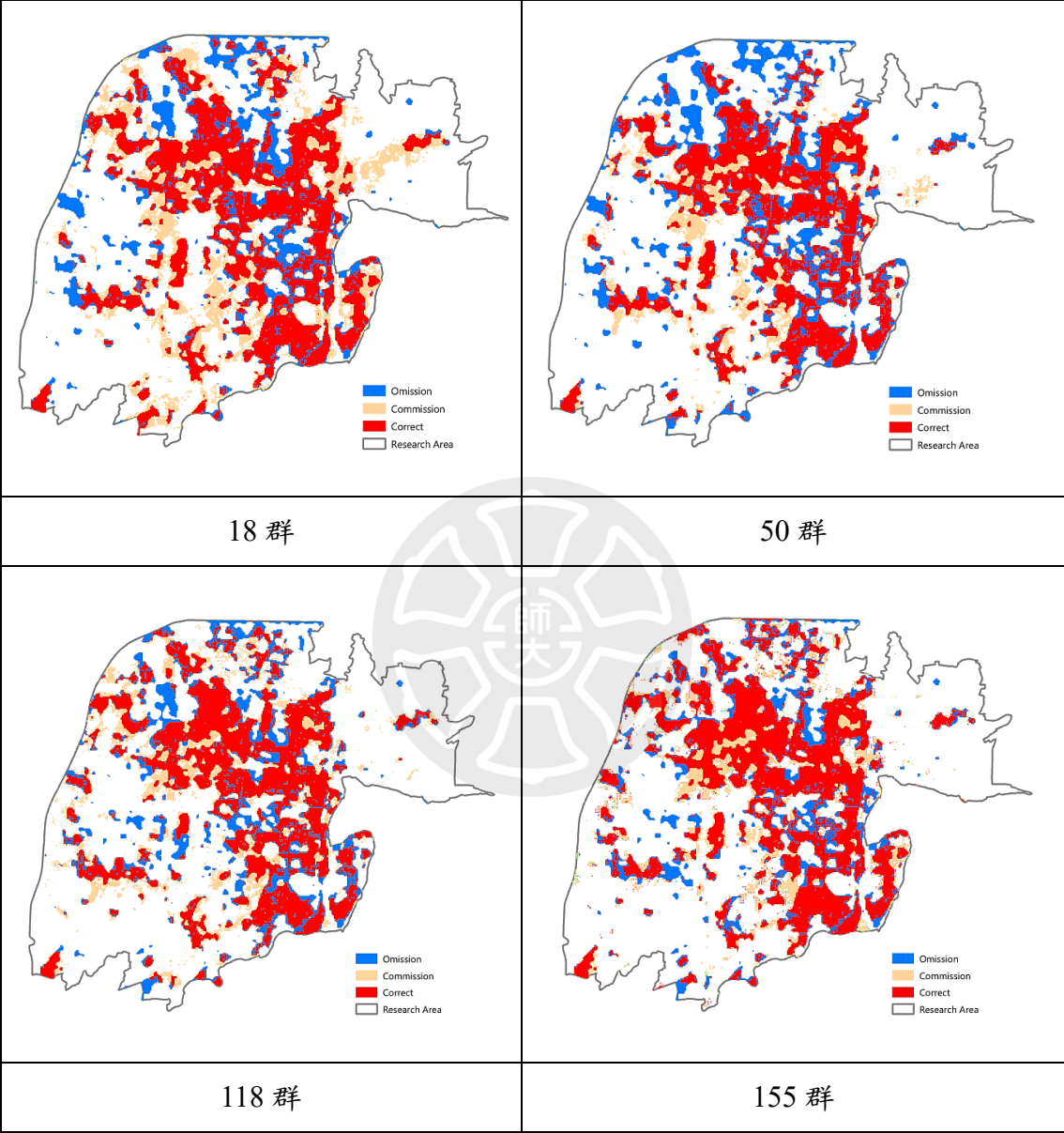


表 11 SOM 環境熱區分群驗證-分群數於 15 至 20 群間



由於增加分群數量可以區分出更多的細微特徵變化，但在 50、118、155 等分群數量找出更細緻的分區特徵，但在分群結果之正確率並沒有顯著改善（表 12）。

表 12 以 2014 年不同分群數驗證



但當分群數量至 300 類以上時(表 13、表 14)。分群之正確率達到顯著提升可達到約 0.59 至 0.66 間(詳見附錄一)，但當 2014 年資料分群數提升至 500 以上(圖 11、圖 12)，分群正確率僅提升約 0.06，考量到訓練成本及使用目標為找出環境因子分布之空間趨勢並盡量提升準確性，其成本與效益明顯不符，本研究評估較無提升正確率之急迫性，因此本研究僅止於 300 群之驗證與其他模型進行其結果空間趨勢之比較。

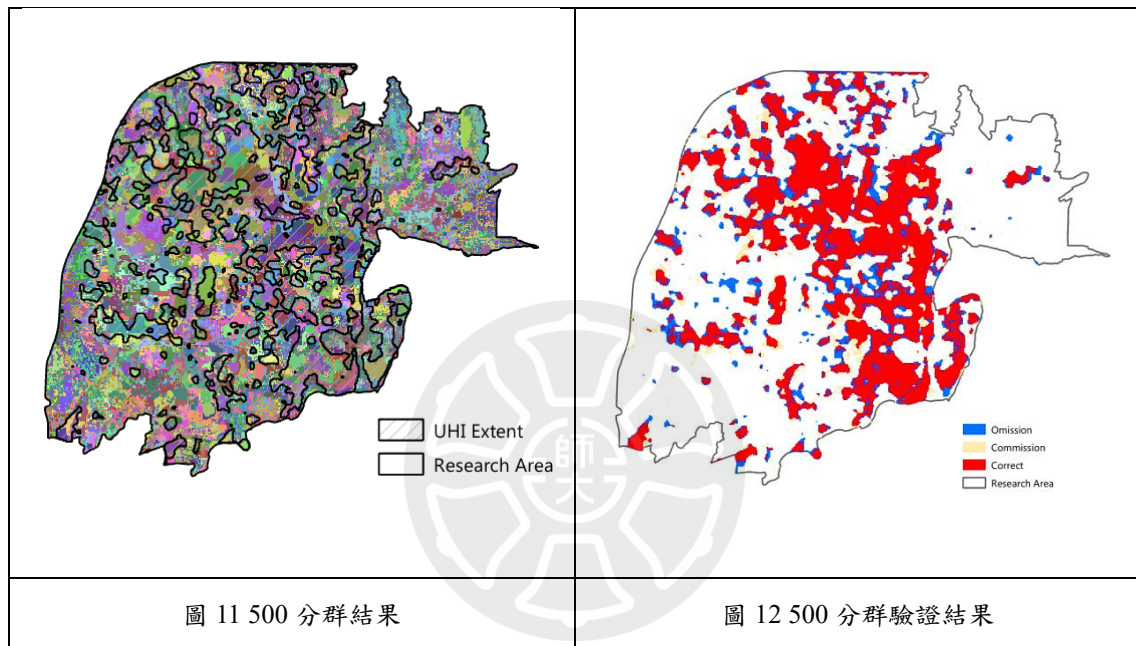


表 13 SOM 環境熱區分群結果-分群數於 300 至 350 群間

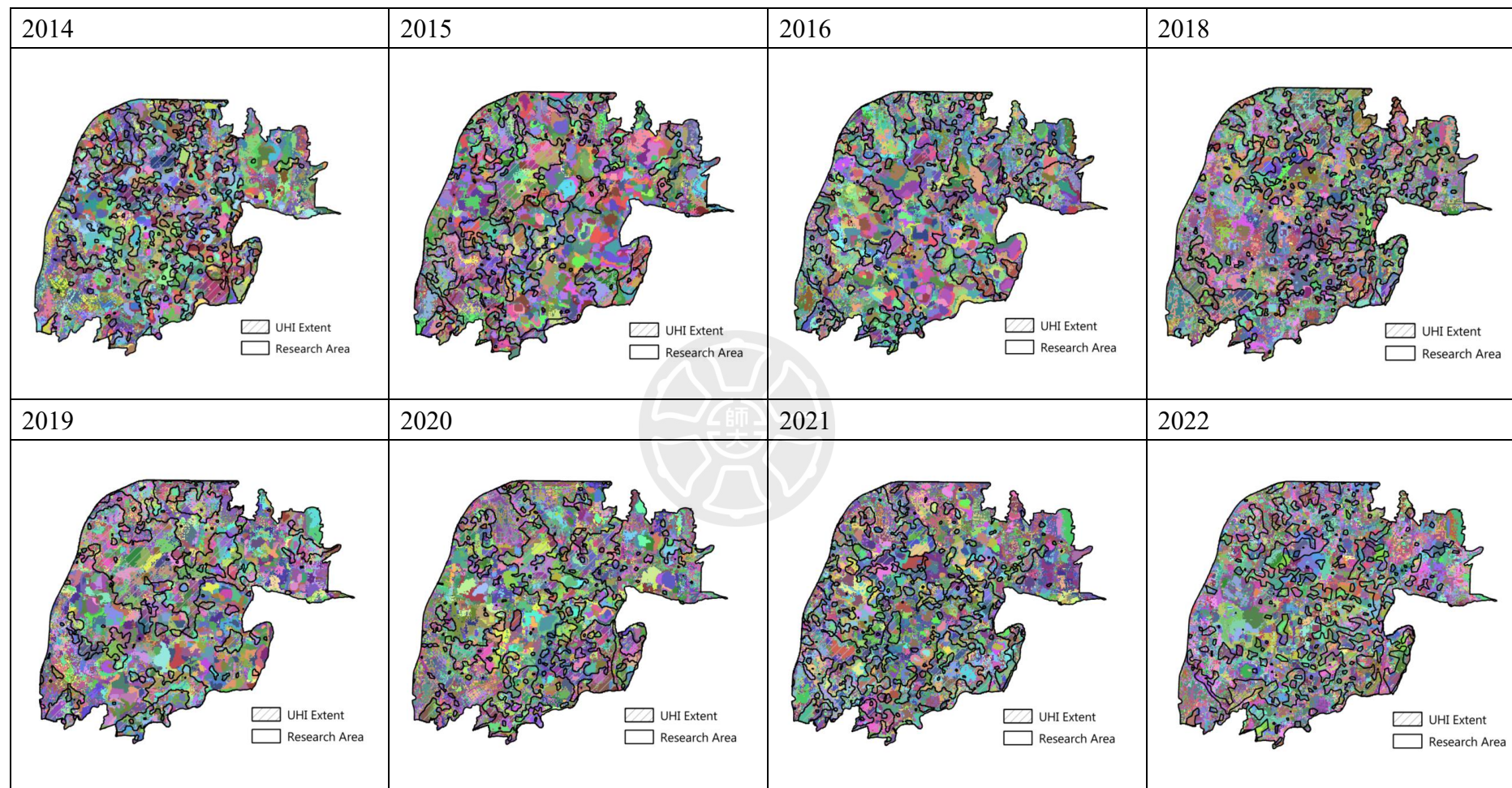


表 14 SOM 環境熱區分群驗證-分群數於 300 至 350 群間

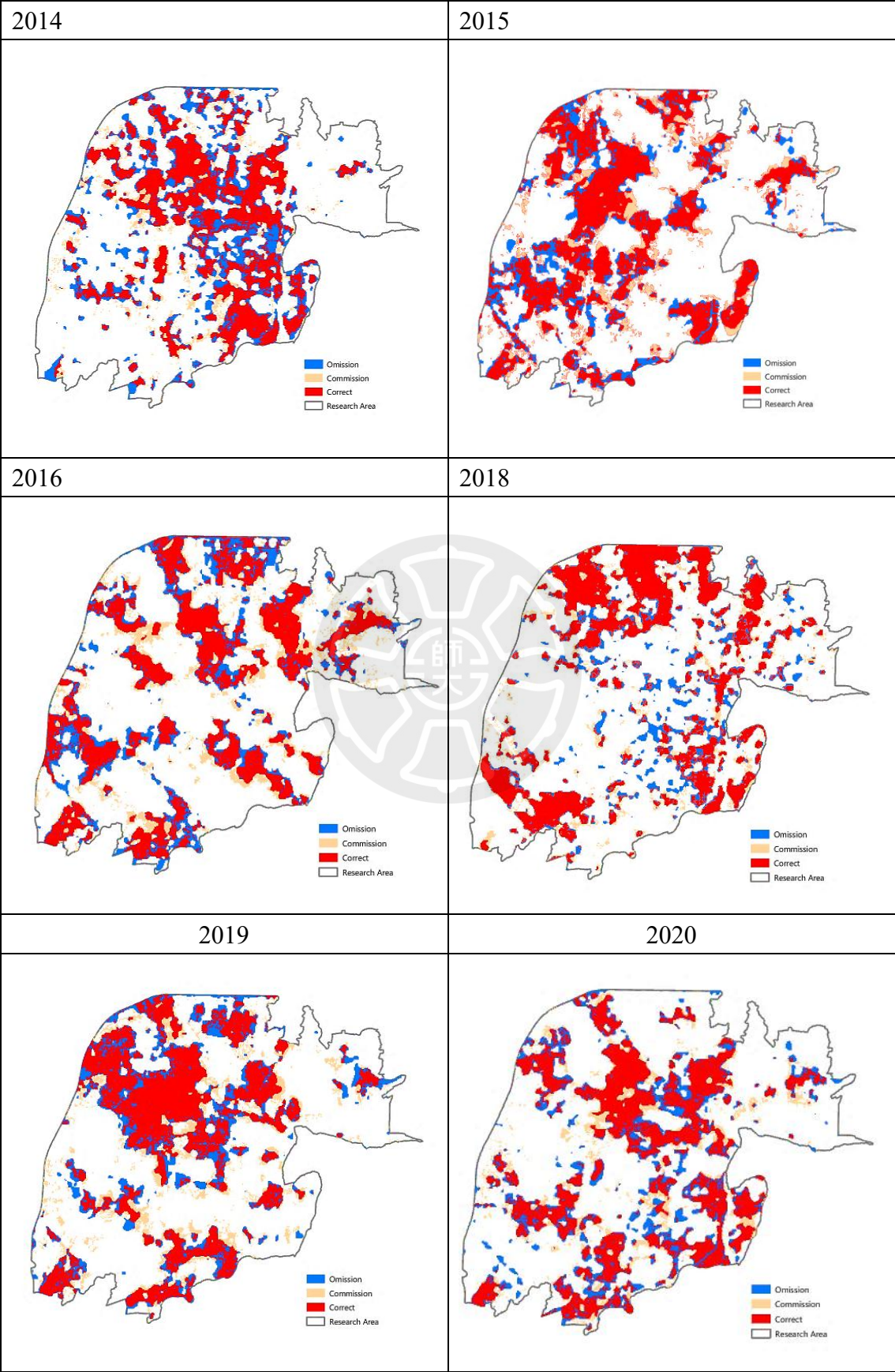
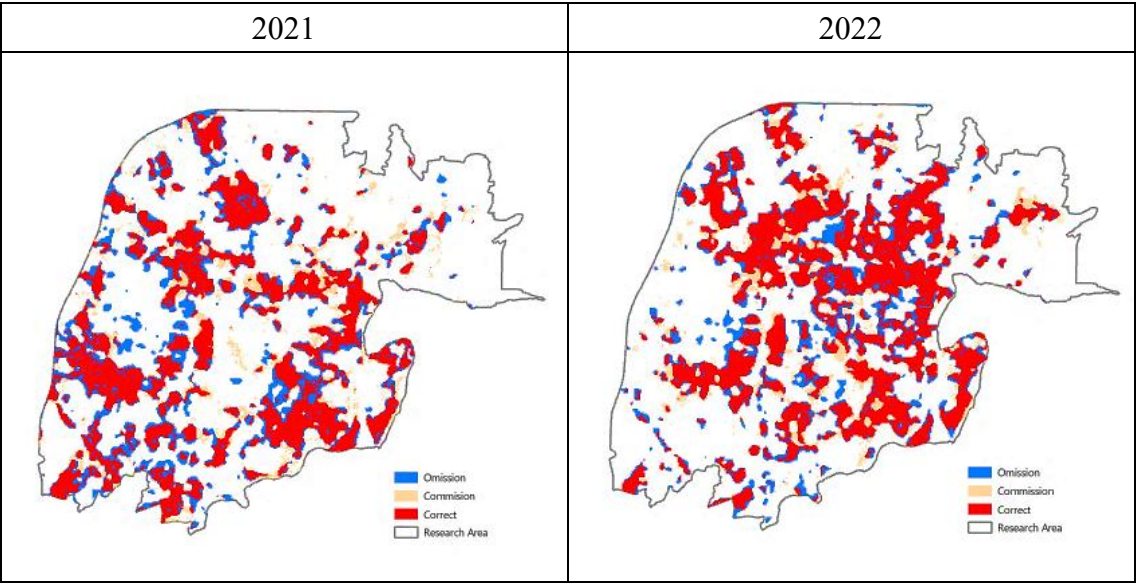


表 14（續前頁）SOM 環境熱區分群驗證-分群數於 300 至 350 群間



第三節 SCMC 模型成果

以 2014 年資料為例，雖然 SCMC 之分群（圖 13）具有空間上的連續性，但若以一樣分群數 15 至 20 時，其分群空間分布與都市熱區不甚吻合，無法看見明確的都市熱區空間分布趨勢。

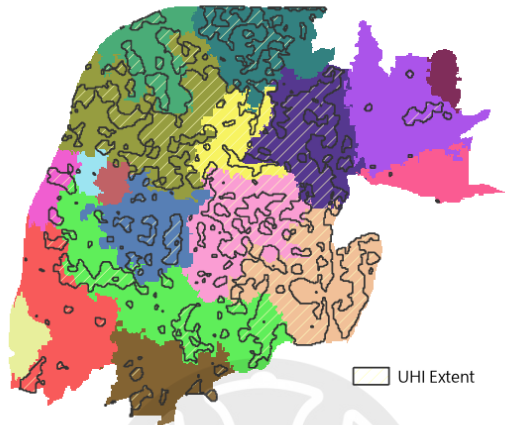


圖 13 SCMC 18 分群結果

為了確定其模型在捕捉更多分群特徵上的變化，本研究進一步將 SCMC 之分群數調整至 300 類，探索其分群分布隨分群數之趨勢，結果（圖 14）顯示 SCMC 在呈現都市熱區的空間分布上達到顯著的效果，其空間碎片化之問題較 SOM 模型少，能更好的保留空間連續性。

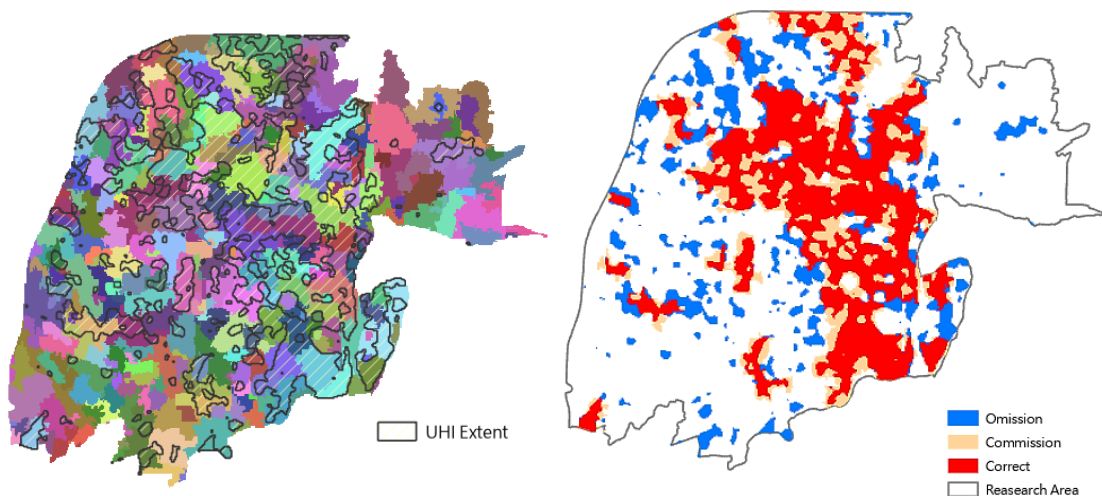


圖 15 300 分群與驗證結果

但在 SCMC 分群結果驗證上(表 15)，其驗證之準確率約為 0.5489，與 SOM 模型 2014 年之準確率相比結果較不理想。此外，SCMC 模型在 commission 與 omission 亦相對偏高，分別為 0.2069 與 0.2441，顯示模型較容易將實際非熱區錯誤判斷為熱區及漏判實際的熱區範圍，相較之下，SOM 模型在整體指標仍表現更為均衡。

綜合比較上，SOM 模型在參數設置與模型調整上的探索效率較高，能較快速地形成具備辨識性的空間分布模式，對於初期探索環境因子的空間特性具有相對優勢。整體而言，SOM 模型在準確率、誤判率與實際熱區辨識能力等各項評估指標皆優於 SCMC 模型，顯示其在都市熱島熱區的空間區辨能力上更佳，亦為後續進行分群後驗分析的較佳依據。

表 15 SCMC 與 SOM 分群結果驗證比較 (Pixel 數)

2014 SCMC Model	環境熱區分群	非環境熱區分群
地真資料熱區	20819	9262
地真資料非熱區	7847	52629

Correct = 0.5489; Omission = 0.2441; Commission = 0.2069

2014 SOM Model	環境熱區分群	非環境熱區分群
地真資料熱區	21424	8657
地真資料非熱區	4542	55934

Correct = 0.6187; Omission = 0.2500; Commission = 0.1311

第四節 SOM 分群結果後處理

本研究發現在分群結果準確度上在使用多年份 LST 資料進行模型建置時，僅需少量的 SOM Grid 及分群數即可捕捉良好的環境特徵建立環境均質區。但當資料加入環境因子後，可發現以同樣參數設置 SOM 分群時，僅可捕捉粗略的環境分區的樣態，推測若欲捕捉更多的細部分區樣態及空間分群，需要大量的分群數來更細緻的區分些微環境特徵的變化。

透過多次試驗與視覺化結果評估調整後，本研究將分群數由 15-20 類提高至 300 類，達到整體空間分布更接近 0.5 標準差都市熱區分布之環境均質區結果。由於本研究將以 2014 年分群結果進行分群後續分析探討對象。

2014 年環境分群中共有 94 群屬於環境熱區，在環境熱區之組內變異中（表 16），由 within-cluster standard deviation 可得知組內差異高於門檻值僅有 4 筆資料，多數群體在組內變異控制範圍內，顯示其分群內部組成具備後續分析之代表性與穩定性。

表 16 SOM 分群變數組內變異

cluster	Industrial Dist	OpenSpace Dist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
2	0.361	0.428	0.255	0.344	0.328	0.615	0.281	0.323	0.331	0.243
4	0.182	0.324	0.292	0.354	0.482	0.431	0.303	0.33	0.216	0.359
6	0.248	0.434	0.316	0.327	0.351	0.475	0.379	0.404	0.287	0.178
7	0.285	0.551	0.147	0.129	0.311	0.535	0.277	0.348	0.365	0.258
8	0.31	0.376	0.111	0.145	0.251	0.41	0.222	0.259	0.193	0.19
9	0.029	0.232	0.262	0.439	0.385	0.517	0.323	0.307	0.394	0.253
10	0.32	0.349	0.142	0.135	0.286	0.378	0.138	0.196	0.23	0.198
14	0.383	0.301	0.176	0.557	0.279	0.439	0.308	0.382	0.488	0.289
20	0.139	0.531	0.231	0.237	0.263	0.404	0.245	0.306	0.229	0.311
21	0.348	0.458	0.283	0.281	0.382	0.509	0.436	0.411	0.24	0.321
24	0.38	0.347	0.357	0.285	0.423	0.536	0.31	0.408	0.308	0.41
26	0.182	0.351	0.265	0.233	0.276	0.369	0.202	0.257	0.256	0.392
38	0.352	0.398	0.238	0.227	0.375	0.496	0.359	0.361	0.488	0.293
39	0.129	0.188	0.192	0.133	0.213	0.402	0.184	0.242	0.218	0.25
46	0.401	0.282	0.247	0.317	0.425	0.432	0.291	0.32	0.223	0.368
47	0.205	0.376	0.229	0.254	0.394	0.476	0.276	0.29	0.482	0.407
56	0.363	0.24	0.242	0.149	0.287	0.265	0.191	0.244	0.257	0.304
57	0.402	0.284	0.334	0.174	0.32	0.359	0.208	0.289	0.367	0.371
58	0.324	0.441	0.339	0.395	0.372	0.429	0.356	0.368	0.455	0.345
60	0.134	0.297	0.234	0.192	0.231	0.35	0.204	0.251	0.339	0.268

cluster	Industrial Dist	OpenSpace Dist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
62	0.253	0.338	0.428	0.278	0.26	0.437	0.198	0.254	0.276	0.324
63	0.334	0.578	0.586	0.565	0.571	0.552	0.477	0.49	0.398	0.469
66	0.361	0.335	0.251	0.21	0.336	0.481	0.233	0.344	0.375	0.376
69	0.412	0.377	0.268	0.294	0.526	0.457	0.267	0.303	0.319	0.429
72	0.37	0.329	0.306	0.302	0.331	0.467	0.253	0.314	0.408	0.249
73	0.359	0.297	0.226	0.211	0.258	0.424	0.242	0.319	0.317	0.315
74	0.408	0.42	0.318	0.258	0.395	0.444	0.313	0.375	0.312	0.46
75	0.114	0.251	0.235	0.253	0.267	0.409	0.291	0.291	0.18	0.292
84	0.222	0.435	0.238	0.256	0.374	0.407	0.212	0.268	0.383	0.715
92	0.217	0.492	0.226	0.243	0.371	0.401	0.203	0.244	0.332	0.356
96	0.179	0.204	0.235	0.169	0.36	0.475	0.198	0.253	0.294	0.128
98	0.368	0.311	0.479	0.328	0.482	0.452	0.189	0.234	0.437	0.367
106	0.325	0.318	0.358	0.266	0.371	0.465	0.205	0.273	0.354	0.297
108	0.281	0.517	0.431	0.388	0.386	0.435	0.239	0.319	0.326	0.495
115	0.2	0.516	0.125	0.138	0.291	0.366	0.222	0.332	0.176	0.344
116	0.262	0.371	0.153	0.157	0.322	0.359	0.164	0.221	0.261	0.208
118	0.29	0.443	0.299	0.274	0.322	0.337	0.251	0.283	0.278	0.261
121	0.312	0.449	0.295	0.282	0.359	0.425	0.188	0.279	0.314	0.29
128	0.191	0.371	0.239	0.208	0.352	0.387	0.236	0.333	0.272	0.533
139	0.287	0.402	0.355	0.309	0.369	0.448	0.347	0.441	0.242	0.209

cluster	Industrial Dist	OpenSpace Dist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
140	0.178	0.408	0.182	0.187	0.321	0.46	0.233	0.254	0.269	0.158
141	0.323	0.331	0.253	0.219	0.28	0.393	0.163	0.212	0.353	0.38
144	0.277	0.244	0.252	0.275	0.389	0.383	0.29	0.338	0.293	0.21
149	0.211	0.365	0.307	0.254	0.411	0.35	0.2	0.263	0.295	0.305
151	0.318	0.36	0.155	0.321	0.457	0.51	0.417	0.489	0.228	0.907
154	0.343	0.418	0.19	0.255	0.302	0.329	0.172	0.214	0.297	0.499
156	0.104	0.291	0.142	0.255	0.282	0.462	0.324	0.383	0.236	0.239
174	0.211	0.411	0.385	0.197	0.313	0.383	0.28	0.304	0.296	0.496
175	0.224	0.286	0.159	0.207	0.33	0.394	0.233	0.267	0.278	0.231
179	0.384	0.442	0.323	0.287	0.383	0.489	0.323	0.339	0.477	0.377
183	0.443	0.31	0.284	0.283	0.272	0.377	0.196	0.241	0.293	0.402
185	0.365	0.478	0.338	0.223	0.351	0.416	0.183	0.248	0.284	0.369
186	0.225	0.377	0.356	0.284	0.417	0.428	0.461	0.522	0.326	0.309
189	0.323	0.342	0.177	0.209	0.296	0.417	0.191	0.224	0.378	0.274
193	0.191	0.287	0.159	0.377	0.34	0.53	0.209	0.228	0.3	0.166
207	0.271	0.279	0.244	0.25	0.236	0.327	0.294	0.295	0.343	0.299
213	0.271	0.328	0.338	0.412	0.341	0.37	0.279	0.263	0.295	0.508
216	0.276	0.313	0.302	0.258	0.285	0.389	0.194	0.236	0.319	0.376
219	0.332	0.378	0.131	0.217	0.238	0.424	0.242	0.255	0.253	0.22
224	0.113	0.267	0.377	0.185	0.343	0.42	0.368	0.399	0.226	0.314

cluster	Industrial Dist	OpenSpace Dist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
225	0.226	0.321	0.248	0.211	0.492	0.402	0.255	0.294	0.3	0.278
228	0.331	0.43	0.347	0.234	0.356	0.419	0.226	0.268	0.436	0.204
229	0.24	0.407	0.206	0.308	0.264	0.385	0.226	0.317	0.247	0.235
231	0.172	0.305	0.188	0.193	0.276	0.374	0.223	0.175	0.362	0.323
238	0.173	0.342	0.097	0.161	0.258	0.498	0.266	0.301	0.255	0.073
241	0.307	0.384	0.195	0.171	0.561	0.317	0.36	0.497	0.233	0.337
242	0.348	0.336	0.172	0.323	0.365	0.32	0.256	0.278	0.407	0.183
244	0.405	0.377	0.326	0.271	0.367	0.422	0.267	0.289	0.285	0.171
245	0.274	0.258	0.254	0.229	0.257	0.333	0.287	0.33	0.284	0.391
246	0.554	0.374	0.361	0.311	0.372	0.543	0.349	0.357	0.407	0.292
254	0.003	0.317	0.186	0.257	0.243	0.439	0.215	0.279	0.159	0.261
256	0.315	0.146	0.134	0.159	0.338	0.343	0.175	0.297	0.404	0.183
260	0.163	0.296	0.23	0.191	0.263	0.402	0.249	0.274	0.294	0.283
268	0.227	0.319	0.236	0.22	0.282	0.432	0.21	0.256	0.297	0.172
269	0.251	0.347	0.492	0.232	0.267	0.385	0.214	0.271	0.178	0.387
270	0.323	0.499	0.277	0.169	0.499	0.391	0.243	0.25	0.462	0.191
271	0.367	0.326	0.268	0.209	0.352	0.404	0.242	0.253	0.332	0.327
283	0.381	0.466	0.271	0.245	0.358	0.418	0.335	0.365	0.434	0.4
284	0.264	0.278	0.184	0.246	0.276	0.378	0.19	0.244	0.3	0.15
286	0.315	0.371	0.244	0.164	0.324	0.366	0.183	0.232	0.282	0.373

cluster	Industrial Dist	OpenSpace Dist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
287	0.404	0.368	0.233	0.284	0.241	0.394	0.233	0.273	0.343	0.195
298	0.29	0.316	0.167	0.182	0.286	0.393	0.256	0.285	0.183	0.235
299	0.342	0.318	0.245	0.209	0.381	0.442	0.194	0.238	0.323	0.298
310	0.269	0.269	0.174	0.217	0.271	0.354	0.254	0.307	0.24	0.234
311	0.315	0.305	0.205	0.196	0.32	0.352	0.198	0.271	0.622	0.436
314	0.401	0.442	0.324	0.346	0.381	0.437	0.38	0.406	0.409	0.278
319	0.326	0.361	0.168	0.153	0.263	0.34	0.152	0.224	0.249	0.275
325	0.216	0.355	0.194	0.119	0.202	0.336	0.177	0.26	0.152	0.331
329	0.366	0.313	0.419	0.387	0.396	0.448	0.308	0.359	0.277	0.307
334	0.274	0.351	0.296	0.26	0.283	0.283	0.211	0.253	0.346	0.265
340	0.239	0.427	0.275	0.348	0.276	0.566	0.367	0.408	0.381	0.506
344	0.303	0.353	0.137	0.059	0.561	0.365	0.208	0.259	0.451	0.113
346	0.351	0.326	0.12	0.194	0.441	0.399	0.226	0.253	0.423	0.244
348	0.215	0.381	0.212	0.26	0.321	0.352	0.285	0.324	0.32	0.262

一、分群整體評估與相似性指標

大量分群使空間分群碎片化在解釋分析上變得十分困難，儘管可取得每一分群的變數組成資訊，但仍難以直接分析環境分群的組間差異，為補足結果的解釋性，本模型以 RF 中之 feature importance 對分群因子組成的重要性進行後驗分析。

本研究以 2014 年分群結果作為分群結果探討的案例，進行整體環境熱區的 feature importance 分析。從整體的 feature importance 結果來看（圖 15），模型重要性最高的三項變數分別為工業區距離、開放空間距離與道路密度，代表在鄰近工業設施、開放空間分布與道路密度等空間環境特性，在區分不同類型的都市熱區中可能扮演重要角色。

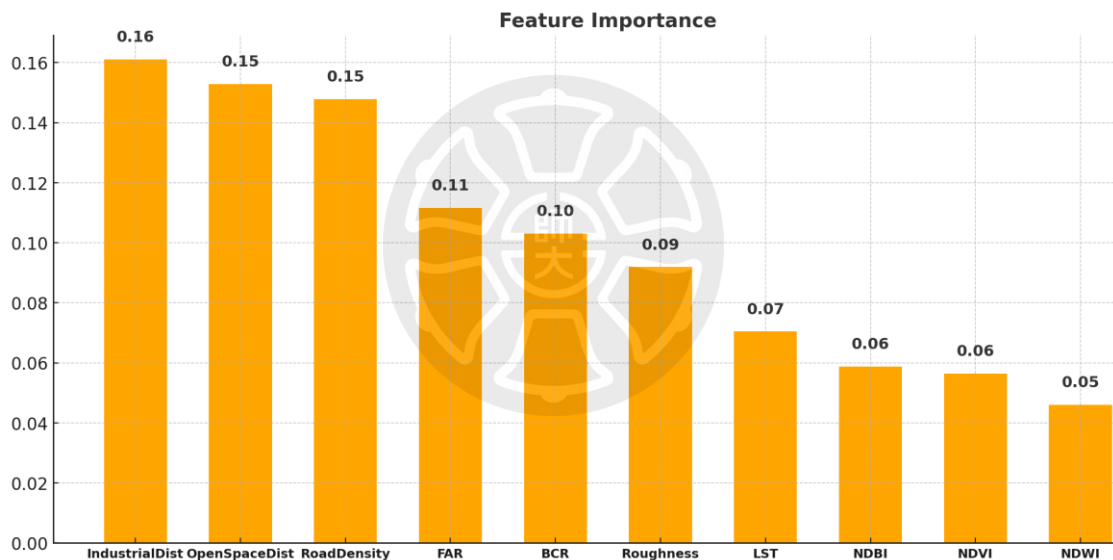


圖 16 原始分群結果 Feature Importance

One-vs-Rest RF 模型所產生的 feature importance(附錄二)進行相似性指標分析，根據 cosine similarity 的門檻結果，部分分群間存在高度相似的特徵權重分佈，其對熱區分布判斷的依據具有極高的相似性，本研究經過敏感性分析後根據相似性之結果將分群重新合併為 6 群（表 17）。

表 17 分群合併列表與組內相似度

New Cluster Label	合併之分群	群內平均相似度
1	139、231、213、7、21、72、74、246、310、63、38、47、2、271、69、186、179、314	0.818
2	156、98、174、185、286、24、73	0.927

New Cluster Label	合併之分群	群內平均相似度
3	346、242、219、39、269、57、118、241、344、75、46、144、4、238、260	0.85
4	115、128、151、141、66、228、298、8、10、270、325、14、183、56、207、225	0.848
5	58、229、116、256、245、340、319、287、84、329、92、154、62、106、108、283	0.796
6	193、96、216、311、334、189、348、175、26、149、9、284、121、244、20、6、254、224、140、60、268、299	0.875

二、合併分群評估

為了評估分群合併後之成果，本研究將模型再次以 RF 進行新標籤的分類作為驗證，由測試集成果（表 18）作為評估，其中 F1-score 均值（Macro Average）為 90.4%、整體準確率（accuracy）達 90.5%。各群別之 F1-score 皆高於 88%，合併後的分群標籤具備良好的區別性，可有效區分不同環境特徵之樣本。

表 18 RF 測試集評估

Cluster	Precision	Recall	F1-score	Support
1	0.889731	0.938909	0.913659	1375
2	0.844138	0.927273	0.883755	660
3	0.888728	0.932525	0.9101	1319
4	0.928571	0.902186	0.915188	1830
5	0.935667	0.862249	0.897459	2294
6	0.900034	0.90809	0.904044	2905
Accuracy				0.905326
Macro avg	0.897812	0.911872	0.904034	10383
Weighted avg	0.906583	0.905326	0.905306	10383

當分群數進行合併並重新劃分為六大類型後，其所對應的 RF 模型中的 feature importance 亦隨之產生變化，原先在細分類群中最具區辨識力的變數為「工業區距離」、「開放空間距離」與「道路密度」，但在六大類型的分群中，變數重要性排序則轉變為「地表粗糙度」、「開放空間距離」與「道路密度」為前三大解釋變數。當分群標籤由細緻的熱區類型合併為較大範疇的環境群體後，模型的分類判斷邏輯亦隨之調整，合併後的 feature importance（圖 16）排序中「地表粗糙度」成為更具

重要性的分辨因子。

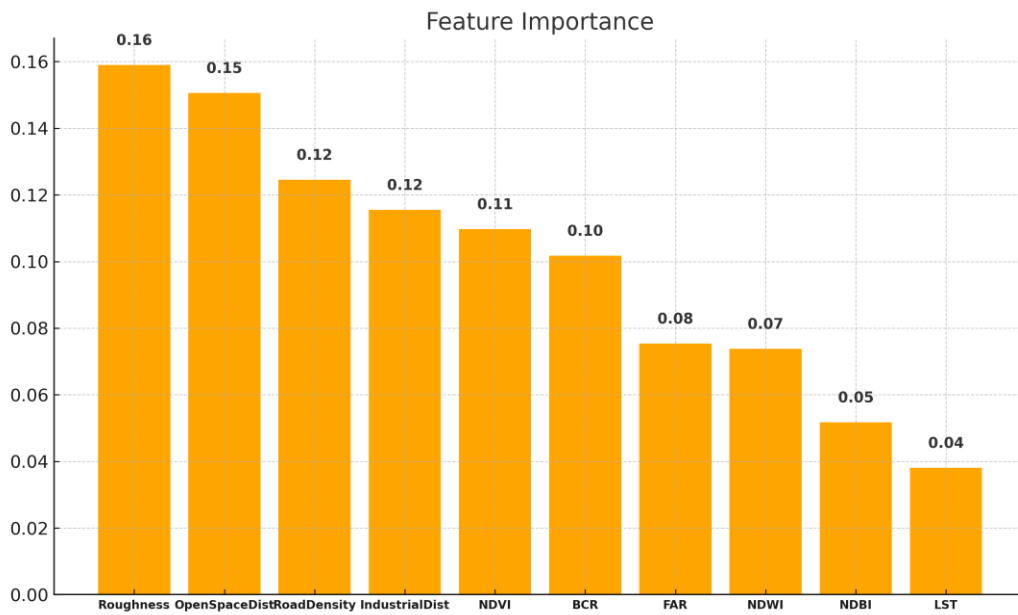


圖 17 合併新分群結果 Feature Importance

經過整併的分群在空間上的連續性更佳，圖 17 中左側為原熱區 94 分群分布，空間分布呈現碎片化及複雜化，後續分析及應用上解釋困難；而右側為合併後分群結果，可見整體空間分布更具連續性，分析及應用上解釋具備更高的應用潛力。

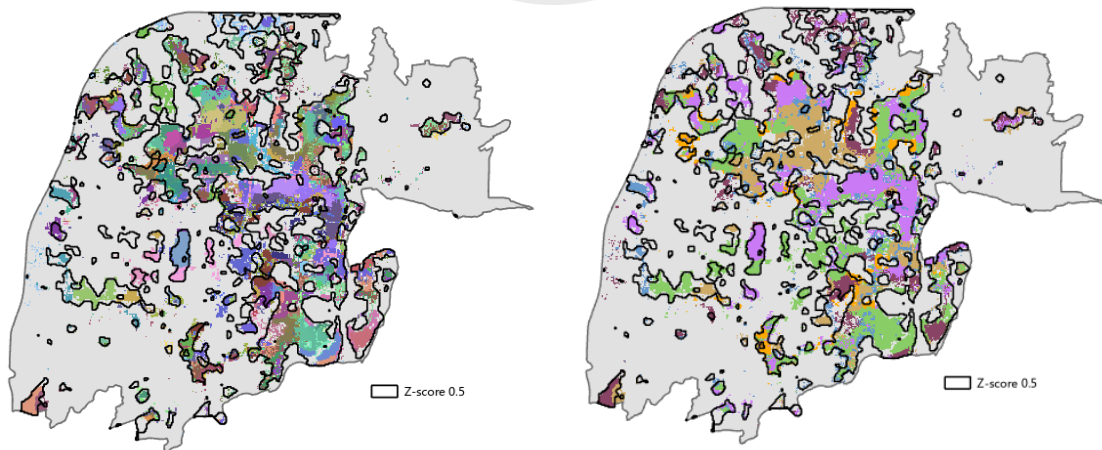
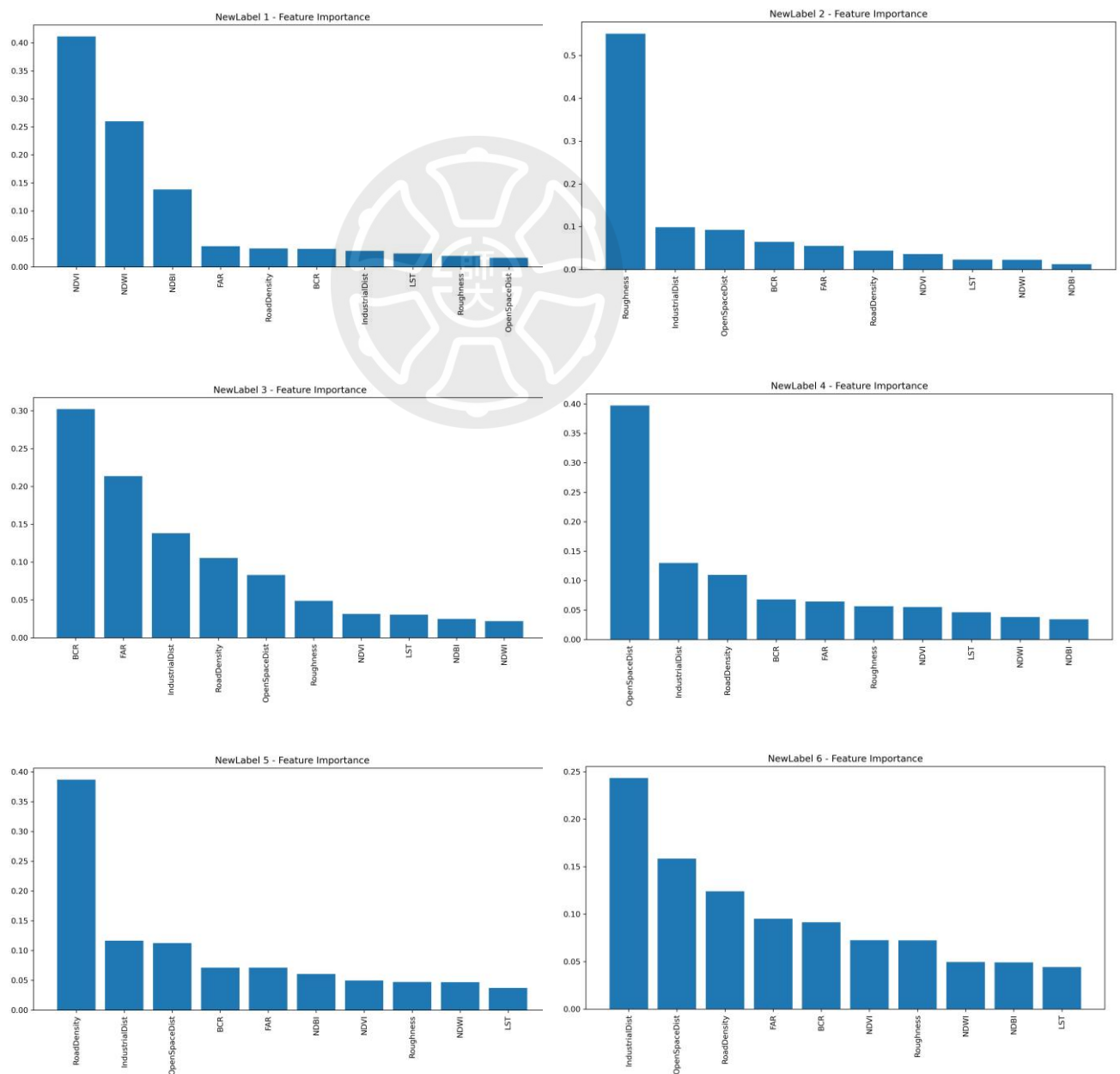


圖 18 原始分群結果與合併新分群結果空間分布對照

合併後分群的特徵重要性分析（表 19）顯示，各分群在模型判斷過程中所依賴的變數組成存在明顯差異。NewLabel 1 的模型區分的重要變數為 NDVI 與 NDBI，NewLabel 2 則以地表粗糙度（Roughness）為主要判斷依據，其餘變數的重要性相對較低。NewLabel 3 中，建蔽率（BCR）與容積率（FAR）為主要特徵；NewLabel 4 則以大型開放空間距離（OpenSpaceDist）為最具重要性的變數。NewLabel 5 的判斷主要仰賴道路密度（RoadDensity），而 NewLabel 6 則以工業區距離（IndustrialDist）與大型開放空間距離為重要性最高之變數。

表 19 新分群 Feature Importance 組成



第五章 結論與建議

第一節 研究結論

本研究欲探討 SOM 於建立環境品質均質區之研究，透過都市熱島相關之環境因子資料建立綜合年分與單一年份之 SOM 模型，並與 SCMC 模型進行比較，可得出以下結論：

比較 SOM、SCMC 模型之結果後，本研究認為 SOM 之結果較適合用於找尋環境均質區。SCMC 模型雖然具有空間上的連續性，在偵測小範圍區域間的變化之能力不如 SOM，無法區分較為細緻的環境樣態。

本研究首先以多年分 LST 作為因子，透過 SOM 將其分為 15 至 20 類之分群，與長期溫度高於 0.5 標準差之地真資料比對，結果顯示其可良好的區分出環境熱區分布，正確率達到 73.86%。但當 SOM 模型引入各年分之遙測因子、立體環境變數、周圍環境變數等環境因子建立相同參數之模型時，可發現各年分之 SOM 模型 Correct 皆顯著將低，推測主因在於環境因子特徵的多樣性使資料複雜度提高，若仍以少量的分群數欲找出空間上的同質區，難以精確劃分出空間同質區，會導致較高的誤判及失判區。但 SOM 模型仍可呈現在不同時間點下環境熱區的變異樣態，可見其在辨識環境模式上的潛力，但如何提升低分群數之模型精準度需再研擬更適當的環境因子或其他優化模型的方式。若將 SOM 分群之數量提升至 300 分群以上，可將正確率提升至 59-66%間，但在可解釋性上難以兼顧，較難直接作為實際政策或應用的依據，因此本研究以 RF 模型對分群結果進行後處理分析及研擬分群整併。

在本研究以 RF 模型的分類邏輯作為分析複雜分群背後的环境因子組成的依據，並作為後續檢驗合併分群合理性，增加本研究之應用潛力與解釋力。RF 模型整體熱區分類的 feature importance 分析中，環境因子中以「工業區距離」、「開放空間距離」與「道路密度」三者對分群結果最具解釋力，顯示這些因素在劃分不同空間熱區上扮演關鍵角色，亦反映都市熱島空間分布與人為開發強度之間的密切

關聯。

本研究其次將各分群視為獨立樣本分別建立 One-vs-Rest RF 模型，並以 cosine similarity 衡量各分群 feature importance 解釋結構的相似性，使用門檻值 0.7 進行合併，以此簡化原始 94 個環境熱區至 6 個環境熱區。以合併後的 6 個分群作為新訓練標籤，重新建立 RF 模型，並以測試集作為分類效能驗證。整體模型之 Accuracy 達 90.5%，Macro 平均 F1-score 為 90.4%，顯示新劃分之環境熱區在 RF 模型邏輯上可被良好的區分出來，且在空間上具有較佳的連續性。

合併後的模型 feature importance 排序亦隨之產生變化，相較於初始分群結果中整體模型以「工業區距離」、「開放空間距離」與「道路密度」為最高的三個特徵重要性因子，在合併後模型中，「地表粗糙度」重要性取代「工業區距離」，成為重要性最高的因子，與現實情況較為相似。此外合併後的分群標籤 feature importance 具備良好的區別性，可有效區分不同環境特徵之樣本。以上驗證了本研究使用 SOM 進行探索性分析並使用 RF 作為後處理的研究流程合理性，可提升環境因子分群空間分布的連續性與可解釋性。

相比於一般的環境指標使用分級或是數值表示，SOM 分群能將空間單元依環境特徵劃分為同質區，有助於深入理解各區域的環境特性，經過整併的分群在視覺化及解釋上都更加容易。但針對細部的環境均質區分析仍需要專業的輔助判決，如瞭解該區的既有條件、當時環境條件、天候問題等對均質區進一步應用。

第二節 研究建議

本研究探討 SOM 用於建立環境品質分區已有初步研究成果，但在模型準確性的提升及後續如何應用分群結果於實務上仍待後續研究探討，本研究的研究建議如下：

將不同面相的環境因子進行分群時會十分複雜，雖然在少量的分群時能看出環境因子大致空間分布的趨勢，若要同時顧及精確率就需大量的分群來捕捉更細緻的特徵，但大量的分群數也導致在解讀上的不易，難以進行有效的視覺解釋與後續應用。本研究已實作以 RF 模型邏輯作為分群解釋及合併機制的參考，未來可接續探討優化結合 SOM 模型與 RF 模型的研究流程及更深入的進行後續探討。

本研究已獲得一定正確率之結果，但模型部分區塊結果較不理想推測可能與資料尺度皆配合 Landsat-8 影像轉為 30 公尺解析度使得部分資料特徵損失，本研究有兩個方向的研究建議。首先可考慮將資料解析度提高至 5 公尺，使環境因子具有更多細節特徵可供模型學習。其次考量到環境多屬於大尺度範圍的變化，未來可考慮將資料解析度降低至 200 公尺以上進行研究，縮減細小範圍之空間雜訊，以更貼近大環境的上空間的特性。亦可考慮測試不同都市規模是否會影響 SOM 的熱島範圍界定能力，以及在不同氣候區或加上更多的環境因子來驗證 SOM 分群的適用性。

最後本研究僅以都市熱島相關環境因子作為探討其作為以 SOM 綜合環境品質建立環境品質之可行性，可作為 ESG 架構中環境面向評估之探索性應用。未來可考量加入建築資訊模型 (Building Information Modeling, BIM)、智慧電表等變數，以分析更多面向的環境關係。

參考文獻

一、中文文獻

何佳薇、周天穎、楊龍士（2011）。臺中地區土地利用變化於熱島效應之研究。航測及遙測學刊，16（2），139-149。

吳宜郡（2020）。都市公園綠地微氣候調節服務之能力與流動（碩士論文）。國立臺灣大學園藝暨景觀學系。

李育寬（2020）。應用 Landsat 衛星資料探討大台北都會區都市熱島效應之時空分析（碩士論文）。國立中央大學大氣科學學系大氣物理碩士班。

林子平（2021）。都市的夏天為什麼愈來愈熱？：圖解都市熱島現象與退燒策略。臺北市：商周出版。

林子平、蔡沛淇、歐星妤與張洲滄（2023）。熱島效應緩解策略之風廊系統的指認與應用。土木水利，50（1），24-29。

潘豐家（2015）。結合衛星資料與建物資訊解析台北市空間發展與都市熱島效應之鏈結（碩士論文）。國立中央大學遙測科技碩士學位學程。

羅子雯（2018）。結合局部氣候分區及景觀生態指標之都市氣候地圖建置及應用（碩士論文）。國立成功大學建築研究所。

石婉瑜. (2018). 臺北盆地的熱環境特徵與都市綠色基盤的影響. *都市與計劃*, 45(4), 283-300.

二、英文文獻

Abukmeil, M., Ferrari, S., Genovese, A., Piuri, V., & Scotti, F. (2021). A survey of unsupervised generative models for exploratory data analysis and representation learning. *Acm computing surveys (csur)*, 54 (5), 1-40.

- Aguado, D., Montoya, T., Borrás, L., Seco, A., & Ferrer, J. (2008) . Using SOM and PCA for analysing and interpreting data from a P-removal SBR. *Engineering Applications of Artificial Intelligence*, 21 (6) , 919-930.
- Andrienko, G., Andrienko, N., Bremm, S., Schreck, T., Von Landesberger, T., Bak, P., & Keim, D. (2010, June) . Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. In *Computer Graphics Forum* (Vol. 29, No. 3, pp. 913-922) . Oxford, UK: Blackwell Publishing Ltd.
- Arribas-Bel, D., Nijkamp, P., & Scholten, H. (2011) . Multidimensional urban sprawl in Europe: A self-organizing map approach. *Computers, environment and urban systems*, 35 (4) , 263-275.
- Ashok, A., Natarajan, G., Elmasri, R., & Smith-Stvan, L. (2020, July) . SimsterQ: A similarity based clustering approach to opinion question answering. In *Proceedings of the 3rd Workshop on e-Commerce and NLP* (pp. 69-76) .
- Assunção, R. M., Neves, M. C., Câmara, G., & da Costa Freitas, C. (2006) . Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20 (7) , 797-811.
- Bação, F., Lobo, V., & Painho, M. (2005) . Self-organizing maps as substitutes for k-means clustering. In *Computational Science–ICCS 2005: 5th International Conference, Atlanta, GA, USA, May 22-25, 2005, Proceedings, Part III 5* (pp. 476-483) . Springer Berlin Heidelberg.
- Banerjee, P., Ghose, M. K., & Pradhan, R. (2018) . AHP-based Spatial Air Quality Impact Assessment Model of vehicular traffic change due to highway broadening in Sikkim Himalaya. *Annals of GIS*, 24 (4) , 287-302.
- Bokaie, M., Zarkesh, M. K., Arasteh, P. D., & Hosseini, A. (2016) . Assessment of urban heat island based on the relationship between land surface temperature and land use/land cover in Tehran. *Sustainable Cities and Society*, 23, 94-104.

- Chang, H. T. (2016) . A temporal and spatial analysis of urban heat island in basin city utilizing remote sensing techniques. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41, 165-170.
- Comberti, L., Baldissoni, G., & Demichela, M. (2015) . Workplace accidents analysis with a coupled clustering methods: SOM and K-means algorithms. *Chemical Engineering Transactions*, 43, 1261-1266.
- Concetti, L., Mazzuto, G., Ciarapica, F. E., & Bevilacqua, M. (2023) . An Unsupervised Anomaly Detection Based on Self-Organizing Map for the Oil and Gas Sector. *Applied Sciences*, 13 (6) , 3725
- Cormack, R. M. (1971) . A review of classification. *Journal of the Royal Statistical Society, Series A*, 134, 321-367.
- Covert, I., Lundberg, S. M., & Lee, S.-I. (2021) . Explaining by removing: A unified framework for model explanation. In *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)* (pp. 2136–2146) . PMLR.
- Domingos, P. (2012) . A few useful things to know about machine learning. *Communications of the ACM*, 55 (10) , 78-87.
- Gao, Y., Zhao, J., & Han, L. (2022). Exploring the spatial heterogeneity of urban heat island effect and its relationship to block morphology with the geographically weighted regression model. *Sustainable Cities and Society*, 76, 103431.
- Ghahramani, Z. (2003) . Unsupervised learning. In *Summer school on machine learning* (pp. 72-112) . Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009) . The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758) . New York: springer.
- Hua, Z., & Anderson-Frey, A. K. (2022) . Self-organizing maps for the classification of spatial and temporal variability of tornado-favorable parameters. *Monthly Weather Review*, 150 (2) , 393-407.

- Jia, W., Sun, M., Lian, J., & Hou, S. (2022) . Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, 8 (3) , 2663-2693.
- Jia, Y., Chen, X., Yu, J., Wang, L., Xu, Y., Liu, S., & Wang, Y. (2021) . Speaker recognition based on characteristic spectrograms and an improved self-organizing feature map neural network. *Complex & Intelligent Systems*, 7, 1749-1757.
- Kohonen, T. (2013) . Essentials of the self-organizing map. *Neural networks*, 37, 52-65.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015) . Deep learning. *nature*, 521 (7553) , 436-444.
- Lee, S. W., & Xue, K. (2021) . An integrated importance-performance analysis and modified analytic hierarchy process approach to sustainable city assessment. *Environmental Science and Pollution Research*, 28, 63346-63358.
- Licen, S., Astel, A., & Tsakovski, S. (2023) . Self-organizing map algorithm for assessing spatial and temporal patterns of pollutants in environmental compartments: A review. *Science of the Total Environment*, 878, 163084.
- Liu, P., Ranka, S., & Kahveci, T. (2017) . Margin-based permutation of random forest for feature importance. *IEEE Transactions on Knowledge and Data Engineering*, 31 (11) , 2124-2135.
- Lundberg, S. M., & Lee, S.-I. (2017) . A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 30 (NeurIPS 2017) .
- Münzel, T., Sørensen, M., & Daiber, A. (2021) . Transportation noise pollution and cardiovascular disease. *Nature Reviews Cardiology*, 18 (9) , 619-636.
- Mei, P. A., de Carvalho Carneiro, C., Fraser, S. J., Min, L. L., & Reis, F. (2015) . Analysis of neoplastic lesions in magnetic resonance imaging using self-organizing maps. *Journal of the neurological sciences*, 359 (1-2) , 78-83.

- Mingoti, S. A., & Lima, J. O. (2006) . Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European journal of operational research*, 174 (3) , 1742-1759.
- Mishra, M., & Behera, H. S. (2012) . Kohonen self organizing map with modified k-means clustering for high dimensional data set. *International Journal of Applied Information Systems*, 2 (3) , 34-39.
- Münzel, T., Sørensen, M., & Daiber, A. (2021). Transportation noise pollution and cardiovascular disease. *Nature Reviews Cardiology*, 18(9), 619-636.
- Nair, B. B., Kumar, P. S., Sakthivel, N. R., & Vipin, U. (2017) . Clustering stock price time series data to generate stock trading recommendations: An empirical study. *Expert Systems with Applications*, 70, 20-36.
- Nguyen, T. T. M., Nguyen, H. T. T., Doan, T., & Tri, D. Q. (2021) . Application Analytic Hierarchical Process (AHP) in Setting up Local Community Urban Environmental Quality of Life Index in a Developed Metropolitan Area in Ho Chi Minh City, Vietnam. *Current Urban Studies*, 9 (3) , 376-391.
- OECD. (n.d.) . OECD environmental performance reviews. OECD.
https://www.oecd.org/en/publications/oecd-environmental-performance-reviews_199000090.html
- Park, C., Ha, J., & Lee, S. (2017). Association between three-dimensional built environment and urban air temperature: Seasonal and temporal differences. *Sustainability*, 9(8), 1338.
- Pham, D. T., Dimov, S. S., & Nguyen, C. D.(2005) . Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219 (1) , 103-119.
- Rizwan, A. M., Dennis, L. Y., & Chunho, L. I. U. (2008) . A review on the generation, determination and mitigation of Urban Heat Island. *Journal of environmental sciences*, 20 (1) , 120-128.

- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.
- Sahin, F., Kara, M. K., Koc, A., & Sahin, G. (2020) . Multi-criteria decision-making using GIS-AHP for air pollution problem in Igdir Province/Turkey. *Environmental Science and Pollution Research*, 27, 36215-36230.
- Shih, W. Y., Ahmad, S., Chen, Y. C., Lin, T. P., & Mabon, L. (2020). Spatial relationship between land development pattern and intra-urban thermal variations in Taipei. *Sustain Cities Soc*, 62, 102415.
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., ... & Golub, T. R. (1999) . Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences*, 96 (6) , 2907-2912.
- Tian, B., Shaikh, M. A., Azimi-Sadjadi, M. R., Haar, T. H. V., & Reinke, D. L. (1999) . A study of cloud classification with neural networks using spectral and textural features. *IEEE transactions on neural networks*, 10 (1) , 138-151.
- Van Der Maaten, L., Postma, E. O., & Van Den Herik, H. J. (2009) . Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10 (66-71) , 13.
- Wang, H., Sun, Y., Li, Y., & Zhang, X. (2016) .An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics*, 17 (1) , 60.
- World Health Organization. (2018, May 2) . 9 out of 10 people worldwide breathe polluted air, but more countries are taking action. <https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action>
- Yang, C., Flanagan, B., Akcapinar, G., & Ogata, H. (2018) . Maintaining reading experience continuity across e-book revisions. *Research and practice in technology enhanced learning*, 13, 1-16.

- Yu, X., Liu, Y., Zhang, Z., & Xiao, R. (2021) . Influences of buildings on urban heat island based on 3D landscape metrics: an investigation of China's 30 megacities at micro grid-cell scale and macro city scale. *Landscape Ecology*, 36 (9) , 2743-2762.
- Xin, J., Yang, J., Wang, L. E., Jin, C., Xiao, X., & Xia, J. (2022) . Seasonal differences in the dominant factors of surface urban heat islands along the urban-rural gradient. *Frontiers in Environmental Science*, 10, 974811.
- Zhi, Y., Shan, L., Ke, L., & Yang, R. (2020) . Analysis of land surface temperature driving factors and spatial heterogeneity research based on geographically weighted regression model. *Complexity*, 2020, 1-9.



附錄一 SOM 驗證數結果-分群數於 300 至 350 群間

2014 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	21424	8657
地真資料非熱區	4542	55934

Correct = 0.6187; Omission = 0.2500; Commission = 0.1311

2015 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	23219	8128
地真資料非熱區	7232	51978

Correct = 0.6018; Omission = 0.2106; Commission = 0.1874

2016 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	21472	7839
地真資料非熱區	6531	54715

Correct = 0.5990; Omission = 0.2187; Commission = 0.1822

2018 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	20136	6331
地真資料非熱區	4503	59587

Correct = 0.6501; Omission = 0.2044; Commission = 0.1453

2019 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	22264	8357
地真資料非熱區	6002	53934

Correct = 0.6079; Omission = 0.2281; Commission = 0.1638

2020 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	20802	7713
地真資料非熱區	5403	56639

Correct = 0.6133; Omission = 0.2274; Commission = 0.1592

2021 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	20411	8595
地真資料非熱區	4631	56920

Correct = 0.6068; Omission = 0.2555; Commission = 0.1376

2022 環境分群	環境熱區分群	非環境熱區分群
地真資料熱區	23615	7044
地真資料非熱區	4817	55081

Correct = 0.6656; Omission = 0.1985; Commission = 0.1357

附錄二 One-vs-Rest RF 模型各分群 Feature Importance

cluster	Industrial Dist	OpenSpaceDist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
2	0.047	0.022	0.147	0.2	0.005	0.017	0.205	0.247	0.091	0.019
4	0.078	0.094	0.301	0.364	0.004	0.004	0.003	0.003	0.133	0.016
6	0.398	0.221	0.058	0.036	0.038	0.072	0.037	0.012	0.064	0.064
7	0.057	0.217	0.076	0.073	0.014	0.005	0.269	0.246	0.042	0.002
8	0.199	0.377	0.13	0.116	0.033	0.008	0.007	0.006	0.062	0.061
9	0.496	0.196	0.035	0.037	0.051	0.002	0.004	0.007	0.138	0.034
10	0.088	0.43	0.057	0.083	0.105	0.01	0.027	0.015	0.154	0.03
14	0.155	0.161	0.303	0.221	0.074	0.01	0.013	0.009	0.028	0.026
20	0.414	0.21	0.055	0.014	0.028	0.001	0.003	0.002	0.134	0.14
21	0.047	0.117	0.043	0.054	0.004	0.04	0.298	0.356	0.023	0.019
24	0.128	0.134	0.053	0.089	0.054	0.043	0.015	0.01	0.079	0.394
26	0.473	0.079	0.194	0.12	0.017	0.002	0.006	0.007	0.047	0.055
38	0.032	0.188	0.061	0.045	0.004	0.154	0.283	0.171	0.059	0.005
39	0.063	0.109	0.363	0.265	0.021	0.002	0.01	0.021	0.023	0.123
46	0.053	0.162	0.221	0.3	0.012	0.013	0.075	0.107	0.04	0.018
47	0.023	0.162	0.067	0.028	0.001	0.142	0.355	0.211	0.01	0.001
56	0.086	0.322	0.149	0.192	0.087	0.029	0.018	0.023	0.079	0.015
57	0.073	0.089	0.412	0.266	0.085	0.003	0.008	0.007	0.043	0.015
58	0.101	0.037	0.089	0.156	0.017	0.016	0.114	0.054	0.4	0.017

cluster	Industrial Dist	OpenSpaceDist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
60	0.406	0.048	0.183	0.247	0.019	0.005	0.006	0.01	0.062	0.016
62	0.327	0.033	0.028	0.13	0.044	0.007	0.008	0.004	0.332	0.087
63	0.052	0.156	0.004	0.003	0.006	0.272	0.303	0.187	0.017	0.001
66	0.108	0.458	0.197	0.113	0.009	0.028	0.007	0.003	0.032	0.046
69	0.016	0.073	0.08	0.122	0.003	0.063	0.31	0.196	0.132	0.006
72	0.053	0.077	0.034	0.017	0.008	0.068	0.416	0.284	0.033	0.009
73	0.108	0.162	0.101	0.088	0.018	0.003	0.005	0.007	0.071	0.438
74	0.065	0.049	0.02	0.023	0.014	0.052	0.335	0.321	0.06	0.061
75	0.036	0.116	0.241	0.353	0.004	0.007	0.002	0.001	0.08	0.159
84	0.257	0.095	0.053	0.035	0.007	0.005	0.013	0.015	0.403	0.119
92	0.269	0.094	0.06	0.033	0.046	0.006	0.01	0.009	0.463	0.011
96	0.276	0.201	0.065	0.097	0.284	0.003	0.005	0.002	0.028	0.04
98	0.078	0.183	0.124	0.159	0.006	0.004	0.017	0.026	0.092	0.31
106	0.22	0.183	0.089	0.09	0.038	0.205	0.025	0.025	0.088	0.037
108	0.33	0.027	0.056	0.091	0.011	0.221	0.01	0.01	0.194	0.05
115	0.172	0.236	0.068	0.027	0.264	0.003	0.003	0.001	0.216	0.01
116	0.081	0.067	0.186	0.182	0.033	0.005	0.027	0.021	0.345	0.053
118	0.192	0.08	0.192	0.251	0.022	0.008	0.021	0.032	0.176	0.026
121	0.393	0.131	0.085	0.052	0.012	0.041	0.038	0.01	0.204	0.032
128	0.111	0.388	0.132	0.033	0.001	0.001	0.007	0.005	0.023	0.3

cluster	Industrial Dist	OpenSpaceDist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
139	0.105	0.39	0.011	0.018	0.002	0.062	0.206	0.146	0.049	0.01
140	0.386	0.045	0.173	0.206	0.023	0.007	0.006	0.006	0.1	0.047
141	0.132	0.528	0.151	0.051	0.044	0.001	0.028	0.04	0.017	0.01
144	0.046	0.21	0.231	0.295	0.007	0.006	0.015	0.016	0.121	0.054
149	0.497	0.187	0.04	0.033	0.027	0.016	0.008	0.005	0.181	0.005
151	0.098	0.398	0.203	0.088	0.002	0.036	0.102	0.042	0.01	0.021
154	0.359	0.047	0.139	0.125	0.007	0.005	0.011	0.013	0.288	0.006
156	0.305	0.037	0.067	0.154	0.006	0.016	0.024	0.002	0.019	0.371
174	0.137	0.06	0.036	0.125	0.009	0.006	0.007	0.006	0.096	0.518
175	0.55	0.08	0.133	0.08	0.027	0.004	0.005	0.003	0.094	0.025
179	0.232	0.078	0.023	0.014	0.006	0.098	0.287	0.181	0.074	0.007
183	0.176	0.226	0.272	0.157	0.014	0.008	0.028	0.021	0.083	0.016
185	0.133	0.048	0.048	0.116	0.008	0.005	0.01	0.016	0.052	0.564
186	0.16	0.025	0.032	0.033	0.003	0.07	0.252	0.123	0.296	0.006
189	0.374	0.132	0.156	0.094	0.05	0.015	0.02	0.016	0.101	0.042
193	0.439	0.132	0.076	0.043	0.188	0.005	0.006	0.009	0.037	0.066
207	0.148	0.26	0.18	0.099	0.018	0.037	0.021	0.011	0.15	0.076
213	0.043	0.214	0.094	0.028	0.004	0.097	0.206	0.207	0.102	0.005
216	0.282	0.227	0.113	0.085	0.025	0.007	0.028	0.022	0.062	0.149
219	0.121	0.07	0.349	0.185	0.021	0.025	0.014	0.014	0.16	0.041

cluster	Industrial Dist	OpenSpaceDist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
224	0.342	0.034	0.132	0.311	0.007	0.008	0.002	0.003	0.106	0.056
225	0.121	0.265	0.237	0.136	0.035	0.004	0.01	0.018	0.134	0.041
228	0.255	0.417	0.029	0.082	0.01	0.011	0.036	0.034	0.054	0.072
229	0.05	0.042	0.238	0.271	0.017	0.026	0.01	0.005	0.285	0.055
231	0.073	0.307	0.079	0.02	0.002	0.082	0.201	0.187	0.047	0.001
238	0.242	0.065	0.291	0.118	0.188	0.001	0.001	0.001	0.03	0.062
241	0.118	0.029	0.244	0.295	0.044	0.127	0.013	0.006	0.117	0.006
242	0.028	0.021	0.329	0.239	0.005	0.013	0.082	0.1	0.142	0.041
244	0.375	0.206	0.037	0.083	0.009	0.003	0.01	0.019	0.193	0.064
245	0.18	0.3	0.067	0.066	0.024	0.01	0.015	0.015	0.311	0.01
246	0.078	0.071	0.029	0.036	0.006	0.142	0.359	0.209	0.065	0.006
254	0.401	0.024	0.03	0.247	0.175	0.001	0.004	0.001	0.111	0.006
256	0.022	0.129	0.175	0.235	0	0	0.015	0.006	0.364	0.054
260	0.096	0.082	0.186	0.224	0.328	0.003	0.005	0.006	0.056	0.014
268	0.269	0.127	0.172	0.203	0.043	0.008	0.018	0.009	0.07	0.082
269	0.116	0.074	0.359	0.285	0.035	0.003	0.014	0.007	0.085	0.02
270	0.135	0.471	0.064	0.107	0.131	0.007	0.004	0.006	0.037	0.039
271	0.062	0.15	0.082	0.096	0.011	0.057	0.221	0.124	0.173	0.024
283	0.321	0.092	0.032	0.036	0.028	0.158	0.074	0.05	0.195	0.015
284	0.395	0.131	0.062	0.125	0.062	0.003	0.015	0.006	0.144	0.057

cluster	Industrial Dist	OpenSpaceDist	BCR	FAR	LST	NDBI	NDVI	NDWI	Road Density	Roughness
286	0.114	0.133	0.044	0.102	0.009	0.007	0.028	0.02	0.174	0.368
287	0.192	0.112	0.068	0.033	0.027	0.01	0.021	0.015	0.43	0.092
298	0.211	0.467	0.106	0.119	0.01	0.004	0.007	0.008	0.044	0.024
299	0.261	0.216	0.16	0.141	0.025	0.035	0.033	0.022	0.078	0.028
310	0.06	0.026	0.057	0.024	0.018	0.193	0.317	0.182	0.12	0.003
311	0.357	0.181	0.068	0.057	0.023	0.043	0.018	0.013	0.02	0.22
314	0.295	0.073	0.062	0.11	0.01	0.098	0.182	0.124	0.026	0.021
319	0.227	0.154	0.048	0.098	0.022	0.051	0.019	0.006	0.33	0.045
325	0.09	0.476	0.127	0.148	0.057	0.011	0.005	0.003	0.034	0.049
329	0.201	0.122	0.026	0.059	0.005	0.003	0.023	0.026	0.511	0.024
334	0.257	0.168	0.109	0.086	0.016	0.029	0.041	0.066	0.193	0.036
340	0.107	0.276	0.183	0.095	0.013	0.004	0.009	0.014	0.291	0.008
344	0.135	0.023	0.262	0.414	0.007	0	0.006	0.003	0.034	0.116
346	0.048	0.018	0.308	0.23	0.186	0.002	0.019	0.017	0.139	0.035
348	0.357	0.125	0.15	0.073	0.012	0.035	0.04	0.035	0.114	0.059