

# 1999 臺北市民當家熱線受理違規 停車處理之空間資料探勘： 以核密度分析為工具\*

廖興中\*\*、廖洲棚\*\*\*、陳敦源\*\*\*\*

## 摘要

本研究利用空間資料探勘的方法，分析臺北市政府 1999 市民熱線在 2010 年 4 月、5 月與 6 月的違規停車紀錄，並試著瞭解其發生位置與整體趨勢。平均最近鄰分析與核密度估計的結果顯示，違規停車紀錄確實呈現顯著的空間群聚現象，最主要的高密度區域在中正區的北部、中正區的西部、松山區的中部、信義區的西北部與大安區的中南部。總之，本研究呈現空間資料探勘方法如何協助政府發掘資料背後所隱藏的民意知識，以作為未來擬定或支援決策的參考。

關鍵詞：空間資料探勘、地理資訊系統、平均最近鄰分析、核密度估計、1999 市民熱線

---

\* 本文為行政院國家發展委員會（前研考會）委託電子治理中心研究成果（研究報告代碼為 RDEC-RES-101002），感謝行政院國家發展委員會與臺北市政府協助提供 1999 相關資料。初稿曾發表於 2013 年臺灣公共行政與公共事務系所聯合會年會暨國際學術研討會「續階：對臺灣公共行政之檢討與展望」，會中感謝中國文化大學推廣教育部國際企業學系郭耀昌副教授給予的寶貴意見。同時也由衷感謝兩位匿名審查人細心的指正與評論，使得本文更加完善。

\*\* 本文通訊作者，世新大學行政管理學系副教授，電子郵件：lhc@cc.shu.edu.tw。

\*\*\* 空中大學公共行政學系助理教授。

\*\*\*\* 政治大學公共行政學系教授。

## 壹、前言

資料探勘 (Data Mining) 的概念首次出現在 1989 年，第 11 屆人工智慧國際聯合會議 (International Joint Conference on Artificial Intelligence, 簡稱 IJCAI) (Wu, 2004: 23)。雖然各家的定義並不完全一致 (Han and Kamber, 2006)，但總體而言，資料探勘主要是指從資料庫中發掘潛在資訊或知識，例如：潛藏於資料中的知識規則、模式、規律與其他的趨勢。現今由於空間資料迅速的大量增加，以及空間資料庫廣泛的使用；如何針對空間資料的特性，充分利用空間分析的方法，挖掘出重要的管理或決策知識，並根據這些知識制定相關策略或支援決策，已經成為重要的課題 (Chen et al., 1996; Han and Kamber, 2006; Witten and Frank, 2005; Tan et al., 2005)。

而空間資料探勘 (Spatial Data Mining) 意指一種資料分析的過程，並在該過程自空間資料庫中探索出有趣且過往未知的有用資訊或知識 (Roddick and Spiliopoulou, 1999; Shekhar and Chawla, 2003)。然而，由於空間資料的複雜程度較高，包括這些資料主要為長時間記錄的連續性資料、資料的收集經常以地區為單位、而且資料通常具有距離或空間座標的訊息 (石計生, 2001)。上述這些空間資料的特質，往往使得傳統資料探勘技術在試圖針對空間資料進行知識萃取時，顯得能力十分有限，且容易產生錯誤的分析結果。因此，如何使用有效的工具，針對大量空間資料進行與行政管理相關的知識萃取，對於公共組織進行決策而言，顯為越發的關鍵且重要。

空間資料探勘可以提供公部門利用空間資料分析，開拓優化公共政策制定的機會。藉由空間資料探勘過程所獲取的知識，可以協助公共組織改善其知識累積的策略與方法。這些經由探勘所產生的知識，一方面可以作為政府改善治理績效的工具；另一方面，則可作為政府長期管理知識建立的方法。像美國太空總署 (National Aeronautics and Space Administration,

簡稱 NASA)、美國國家地理空間情報局(National Geospatial-Intelligence Agency, 簡稱 NGA)、美國癌症研究中心(National Cancer Institute, 簡稱 NCI)與美國交通部(U.S. Department of Transportation, 簡稱 USDOT)等機構,便在許多不同的政策領域中,利用空間資料探勘的技術挖掘管理知識;這些領域包含生態與環境管理、公共安全、交通運輸、地球科學、人口學與生態學等。此外,在國內像林祥偉(2011)利用地理資訊系統的空間資料處理能力,以花蓮 119 急難救護的空間資料為操作案例,分析案例發生區位、人口密度、道路密度、消防分隊責任範圍,利用地理加權迴歸,檢視整個花蓮縣急難救護的醫療資源,並推估村里間醫療資源利用的空間分佈,藉以發掘出發現先前關心卻未曾有效擷取的重要知識。由上述的例子來看,空間資料探勘的方法,不論在國外與國內都漸漸受到重視並開始廣泛應用到許多公共事務的領域之中。

而本研究主要是根據臺北市政府「1999 臺北市民當家熱線」(以下簡稱 1999 市民熱線)於 2010 年 4 月到 6 月間,所受理之違規停車處理資料,利用地理資訊系統來處理平均中心點分析、平均最鄰近分析、與點密度分析,進行空間資料探勘研究。主要是分析三個月內民眾向 1999 市民熱線通報違規停車派工紀錄分布之空間特性,探索臺北市這些可能現象的高密度顯著區域。換言之,本研究的主要目的為:分析臺北市違規停車之空間分佈特性,探討臺北市違規停車之空間群聚型態,並歸納出臺北市違規停車派工案件密度較高且可能的需要注意的區域。

## 貳、地理資訊系統與空間資料探勘

早在 1960 年代,加拿大、美國等地已開始將地理資訊系統運用於國土利用、自然資源及地質調查、人口普查等工作,但直到 1990 年代個人電腦以及網際網路的快速發展與普及後,地理資訊系統的應用才開始日趨個人化、生活化與普及化。隨著地理資訊系統技術的蓬勃發展,越來越多學術與實務領域都開始將地理資訊系統作為研究與分析的工具。

地理資訊系統主要是為著建立地理或空間的相關資料所設計的資訊系統，該系統具有空間資料描述與分析的能力（Star and Estes, 1990）。簡而言之，地理資訊系統視為一套結合軟體、硬體與空間資料的系統，並用以進行資料記錄、儲存、上傳、操作、分析與展示各種地圖資料。根據 Maguire（1991: 13-14）對地理資訊系統的界定，地理資訊系統可以包含以下三種主要的功能：

1. 地圖：主要強調地理資訊系統製作地圖的功能，並將地理資訊系統視為地圖處理與展示的系統。
2. 資料庫：主要強調在地理資訊系統中，完善的資料庫與良好資料結構的重要性。
3. 空間分析：主要強調空間分析功能的重要性，地理資訊系統已不再被視為只是資訊處理的工具，而是空間資訊科學。

此外，Goodchild（1992）所提出的觀點與 Maguire 類似，認為地理資訊系統可以分為兩個面向。首先，強調空間資料的建置與管理，扮演空間資料管理與存取的角色。其次，則強調豐富多樣之空間分析模型的建立。由此看來，地理資訊系統與一般資訊系統最大的不同之處，主要是在於獨特的空間資料分析能力。因此，綜合過去學者的觀點來看，地理資訊系統主要是結合了資訊科技與空間資料的整合性系統，運用資訊科技進行空間資料之獲取、儲存、管理、分析與展現等相關功能。

長久以來，無論學術或實務領域對於地理資訊系統，都有著相當程度的期待。無論公私部門都期望能利用地理資訊系統的空間分析能力來提升生產力，並且改善空間資料的管理能力。甚至，由更具策略性的觀點來運用空間資料，以作為決策訂定時的支援或參考依據（Huxhold, 1991）。空間分析主要是利用地理資訊系統，將空間資料轉化成為另一種形式，並且從資料中萃取出額外的意涵。從廣泛的地理資訊系統空間資料分析的角度而言，空間資料探勘可以歸屬於進階的空間資料分析。透過空間資料探勘的過程，利用地理資訊系統複雜的空間分析功能，進行相關空間資料的探勘，進而找出其中所隱含的管理知識。Ester 等人（1997）便將空間資料探勘（Spatial Data Mining）定義為，針對空間資料所進行的資料分析。換言之，使用地理資訊系統進行資料探勘的目的，主要是針對資料庫中空

間性及非空間性的資料，分析出彼此間的關係。而這些關係可能包含了空間上的關係與非空間上的關係；甚至於推敲出具有規則的模式出來。

空間資料探勘是一個正在蓬勃成長的研究領域，而且目前的成熟度越來越高。在過去的十幾年當中，由於全球定位系統（Global Positioning System，簡稱 GPS）的發展、以及網路為基礎之空間資料的可分享與地圖化，使得許多的研究領域開始建立並使用空間資料來進行分析與研究。例如 Spielman 與 Thill（2008）便將資料探勘的概念與地理資訊系統的功能進行結合，利用紐約市的人口社會經濟資料，嘗試找出社會相似性與空間鄰近性之間的關連性。除了學術應用的興起外，實務界的應用也漸漸受到重視，私人企業與民眾也開始產生大量的空間資料，並且為了相關的目的開始利用這些資料（Guo and Mennis, 2009）。因此，空間資料探勘的研究，在未來勢必會產生更新更多的應用與分析方法。

到目前為止，依據分析任務或目的的不同，空間資料探勘有許多的方法可應用，包括運算、視覺化分析、統計，甚至是前述方法的綜合。而依據不同的分析任務，空間資料探勘可分為以下幾類（Ester et al., 1997）：

1. 空間特徵分析（Spatial Characterization）：針對所關注的議題從資料庫中選取符合議題的相關資料，進行空間或是非空間上的描述。
2. 空間區分（Spatial Discriminate）或極端值分析（Outlier Analysis）：是指將「標的」物件的一般特徵與一個或多個「對比」物件的一般特徵進行比較。進而以空間視覺化為基礎，界定出空間資料異常的位置與成分，往往極端值就是變化最明顯的空間成分。
3. 空間分類分析（Spatial Classification Analysis）：是針對空間物件根據某些空間或非空間特質所建立的分類架構進行歸類。譬如分析「ATM 增設合理區域」，空間分類分析就可以根據 ATM 設置相關標準，針對空間資料進行分類分析，歸納出哪些區域符合增設 ATM 的標準。
4. 空間群聚分析（Spatial Clustering）：先針對與探勘工作有關點和面的空間資料，運用有效的空間群聚分析法，將資料分成若干不同的群聚類別，之後再針對各個群聚類別的屬性，歸納出能夠描述空間群聚一般特性的規則。
5. 空間關聯分析（Spatial Association Analysis）：針對大量的空間以及非

空間資料，找出彼此之間的關連性，並且歸納成規則方式來做展現。而空間關聯有許多的不同的空間描述方法可用來表達包括資料的相交、重疊、分離、方位與距離。此外，空間關聯程度也可應用空間自相關（*spatial autocorrelation*）來量測，用來表達空間數值變數本身與多個同時發生數值變數之間的關係。

綜合而言，由於資料庫建構的目的不應僅僅是一個大型整合的資料庫，其重要的使命在於能透過各種分析方法產生初步的決策資訊，並發掘出可用的資訊來促進組織知識的升級。換言之，透過針對民眾回應紀錄的資料探勘，以及結合地理資訊系統空間分析的強大功能，政府便能從資料庫中發現隱藏且有價值的資訊，並將其轉化成人們容易瞭解且具有應用價值的知識，作為支援政府進行空間策略決策及解決相關政策問題時的參考。空間資料探勘技術應用於公共行政專業領域，有待行政專業人員的專業背景知識與資訊科學領域提供先進技術，共同攜手合作才能在理論與應用實務兼顧下，得到實質發展。以下本研究將以行政機關中常見的點資料為例，以臺北市政府 1999 市民熱線受理違規停車派工紀錄之點資料為個案，依據其資料的特質，進行空間資料探勘的說明與分析。

## 參、研究方法

在向量式的空間資料探索研究之中，有許多的地理資訊系統軟體經常扮演了十分重要的空間分析工具（Eck et al., 2005；林祥偉，2011），例如：GeoDa 與 ArcView。GeoDa( Anselin, 2004)可以運用空間自相關(*spatial autocorrelation*)，呈現面資料的群聚是否顯著，以及不同群聚型態在空間中的分佈狀態。Environmental Sciences Research Institute（簡稱 ESRI）的 ArcView 中的空間統計工具（*Spatial Statistics*），亦可進行類似的空間分析研究。在向量式 GIS 資料模式的空間資料探勘研究中，以「點」資料為探勘研究的樣本最為常見，透過地理資訊系統將這個資料以空間分布的方式展現，便能夠解讀出研究標的在時空中的趨勢（林祥偉，2011）。

點資料的分析在地理學、生態學與病理學則有著較長的傳統（Clark and Evans, 1954; Cliff and Ord, 1981; Harvey, 1966; Knox, 1964; Mantel, 1967; Snow, 1855）。由於地理運算方法的更新，以及由全域轉為著重區域分析的趨勢，使得研究者對於點資料分析產生了不同以往的興趣。地理資訊系統的出現，提升了研究者對於空間資料的運算與分析能力，而且該工具近來更是結合許多不同的方法，例如空間統計（spatial statistics）或模糊運算（fuzzy computation）等，使其地理運算的功能再進階（Brimicombe, 2003）。再加上目前空間計量學開始重視區域的差異性，而非全域的一致性，使得點資料空間群聚性的探索與分析成為關鍵。

目前處理點資料群聚現象分析的方法主要以熱點分析（hot spot analysis）最常用，其焦點在於偵測空間中是否有事件或活動產生不尋常的群聚，並且大量運用在疾病、犯罪或交通事故的分析。例如在空間流行病學（spatial epidemiology）當中，這樣的分析方法便是主流（Lawson, 2001），主要在界定疾病風險人口群聚的區域，並偵測是否有傳染擴散的問題。然而，在眾多群聚分析的方法中，又以核密度分析的應用最廣泛。本研究便是嘗試利用熱點分析中常用的核密度分析（Kernel Density Analysis），以臺北市政府 1999 市民熱線受理違規停車派工紀錄之點資料進行空間資料探勘。以下為本研究分析的步驟說明，主要分為空間資料定位、平均最近鄰分析、以及核密度分析等三個部分。

## 一、空間資料定位

本研究以臺北市政府 1999 市民熱線受理違規停車派工紀錄之點資料為基礎，繪製地圖並進行該資料空間分佈型態之分析。藉由空間資料的分析，挖掘出未來相關議題的行政管理知識。本研究所使用之分析資料，是由臺北市政府研考會所提供之 1999 市民熱線於 2010 年 4 月至 6 月受理之違規停車派工紀錄<sup>1</sup>。

對於 1999 派工服務的資訊而言，資料屬性的錯誤便是最大的雜訊來

---

<sup>1</sup> 本研究所使用之 1999 市民熱線派工案件發生地點資料，為研究團隊在執行 2011-2012 年電子治理研究中心研究案時，向臺北市政府申請取得，但所提供之資料時間僅受限於 2010 年之 4 月至 6 月。

源。不論從民眾陳情的語音檔案自動轉換為地址、從來電顯示顯示電信公司登錄的地址擷取資訊、或是直接由 1999 市民熱線話務人員鍵入地址資料<sup>2</sup>，都有可能產生許多種錯誤。這些錯誤類型主要包括兩類：地址紀錄不完整、以及語音轉換時的口語化地名（可參考表 1）。其次，亦可能出現同樣的派工案件，於相近的時間內由不同的民眾通報。因此，本研究團隊一方面自臺北市政府研考會取得內含地址資訊的之違規停車派工紀錄之外；另一方面則利用人工過濾的方式，將資訊不完整的紀錄剔除，亦即將所有地址資料不全或相關訊息不足者，排除在本分析的範圍之外。在經過資料雜訊剔除的處理後，總計 1999 市民熱線於 2010 年 4 月受理之違規停車派工紀錄有 8,346 筆記錄（排除 143 筆無法分析資料，占原資料 1.68%），5 月及 6 月則分別有 8,717（排除 158 筆無法分析資料，占 1.78%）與 8,551（排除 167 筆無法分析資料，占 1.92%）筆記錄，被納入本研究空間資料探勘的分析之中。而各月份因為地址紀錄不完整、以及口語式位置資料等問題而被排除的資料，在各月份資料中所占比率皆低於 2%，因此在分析上所產生的誤差不大。最後，利用地理資訊系統之座標定位功能，依據各紀錄所在位置之經緯度座標定位於 TWD97 二度分帶座標系統，並將其轉換成違規停車派工紀錄之點地圖資料檔案。

表1 臺北市1999違規停車派工資料雜訊例子

雜訊類型	紀錄地址
地址資料不完整	臺北市忠孝東路
	臺北市萬大路與雙和街交叉
	經貿二路62巷巷內
口語式位置資料	東湖捷運站2號出口附近
	捷運新北投站出口與七虎公園游泳池之間單行道

資料來源：本研究自行整理。

<sup>2</sup> 目前 1999 市民熱線派工案件發生地點資料的取得，主要是仰賴話務人員的輸入，亦即話務人員先記錄民眾來電陳述後再轉化為發生地點或地址資訊輸入派工系統。



## 二、檢視臺北市政府違規停車派工案件之聚集或離散程度－平均最近鄰分析

接著本研究利用第一步驟所完成的點資料，繪製 2010 年違規停車派工之地點分布圖，透過視覺分析的方式初步瞭解該資料的空間分布型態。利用 ArcView 10.0 軟體中的「平均最近鄰分析」（Average Nearest Neighbor Analysis）工具，可進一步獲得較確切的評估結果。此處的平均最近鄰分析指數是由植物生態學者 Clark 與 Evans (1945) 所建立的指標，主要用來檢測空間點分佈的型態。平均最近鄰分析主要用於計算空間中點資料之分佈狀態，檢測資料分佈的型態屬於隨機、聚集或分散。公式如下：

$$ANN = \frac{\sum_{i=1}^n di}{n} \bigg/ \frac{0.5}{\sqrt{\frac{n}{A}}}$$

$ANN$ ：平均最近鄰指數

$di$ ：資料點  $i$  與其最近資料點的距離

$n$ ：資料點總數

$A$ ：研究區域總面積

$$Z_{ANN} = \frac{\frac{\sum_{i=1}^n di}{n} - \frac{0.5}{\sqrt{\frac{n}{A}}}}{\frac{0.26136}{\sqrt{\frac{n^2}{A}}}}$$

如同前述所言，平均最近鄰分析是用來計算點資料分佈狀況的指數。假若  $ANN$  值大於 1，則表示臺北市違規停車的派工資料是屬於分散的分佈；假若  $ANN$  值小於 1，則表示違規停車派工資料屬於具有群集分佈的現象。但是若計算分析結果數值接近於 1，則此資料比較接近隨機的空間分佈狀態。此外，當  $Z_{ANN}$  值小於 -2.58 時，則是表示資料分佈有顯著的空間群聚狀態。

### 三、界定出臺北市違規停車派工案件較密集之區域—核密度分析

在確定資料的空間分布有顯著群聚的現象之後，本研究以違規停車派工資料進行核密度分析。該方法主要是在研究範圍中建立均勻的網格（cell），以每個網格中心點透過搜尋半徑建立其搜尋的範圍，並藉此推估違規停車派工地點在空間上出現的密度，可用來表示這些事件發生群聚的強度。通常採用建立核密度分配的方式推估，其數學關係式如下（Burt and Barber, 1996）：

$$\lambda(s) = \frac{1}{b^2} \sum_{i=1}^n K\left(\frac{di}{b}\right)$$

$\lambda(s)$ ：估算出之核密度值

$K$ ：雙變量的概率密度函數，稱為核心 Kernel

$b$ ：搜尋半徑

$di$ ：網格中心點至點  $i$  的距離

根據上述的公式，本研究以違規停車派工資料地點作為定點核心，搜尋的半徑為 500 公尺，網格尺寸為 10 平方公尺，並計算半徑範圍內的派工案件數，產生臺北市違規停車派工案件核密度趨勢圖。利用核密度分析的功能，分別界定出 2010 年 4 月到 6 月臺北市違規停車派工集中的區域。同時進一步分析這三個月臺北市違規停車派工之點密度深色區塊（點密度值高）的位置，找出三個月中共同出現點密度值較高的區域位置。

### 肆、研究結果

在敘述統計的結果部分，其分析結果顯示臺北市 2010 年 4 月共計有 8,346 筆違規停車的派工紀錄，5 月及 6 月則分別有 8,717 與 8,551 筆紀錄。若將紀錄依時間加以區隔，本研究發現 1999 市民熱線在接獲民眾通報違規停車派工數量最多的時段為晚間 19 點 00 分至 19 點 59 分之間（如圖

1)。表 2 顯示在 2010 年 4 月，19 點 00 分至 19 點 59 分之間所發生之違規停車派工件數為 653 件，約占當月總件數之 7.8%；5 月同時段的件數則為 670（占 7.7%），5 月則為 613（占 7.2%）。至於其他各時段，除了晚上 23 點 00 分至 23 點 59 分之間所發生的件數較少，約占總數量 2%外；其他各時段之違規停車派工件數，大約占總數量之 4%到 7%之間。而在凌晨 00 點 00 分至早上 06 點 59 分之間民眾通報違規停車派工的數量最少；在 4 月間該時段的派工件數為 337，僅占當月總件數的 4%。5 月也僅有 357 件違規停車派工（占 4.1%），6 月則更只有 285 件（占 3.3%）。

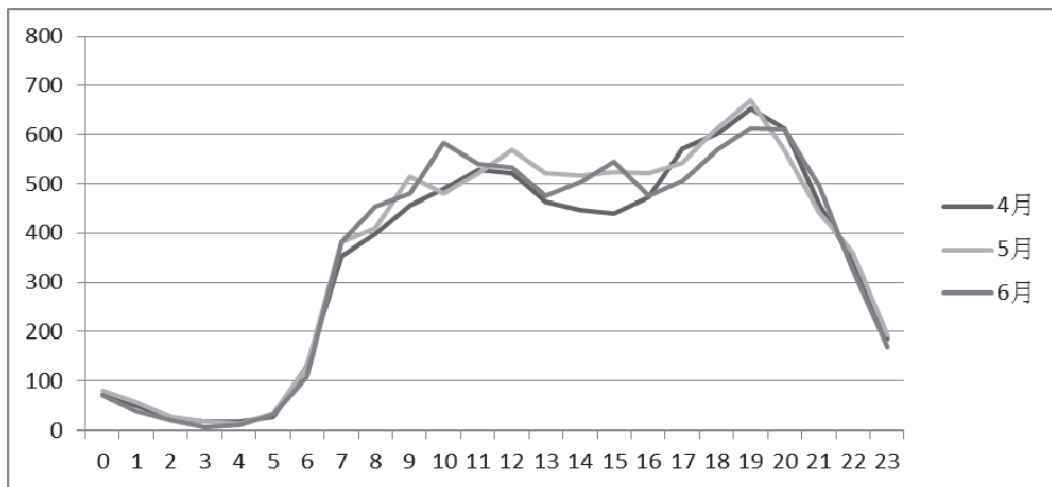


圖1 臺北市政府違規停車派工件數線型圖（每小時）

資料來源：本研究整理自2010年4月至6月臺北市1999違規停車派工資料。

表2 臺北市1999違規停車派工各時段次數表

時間	4月	時段總合	5月	時段總合	6月	時段總合
0000-0059	73 (0.9%)	337 (4.0%)	80 (0.9%)	357 (4.1%)	71 (0.8%)	285 (3.3%)
0100-0159	49 (0.6%)		56 (0.6%)		39 (0.5%)	
0200-0259	25 (0.3%)		27 (0.3%)		20 (0.2%)	
0300-0359	17 (0.2%)		18 (0.2%)		5 (0.1%)	
0400-0459	17 (0.2%)		13 (0.1%)		10 (0.1%)	
0500-0559	27 (0.3%)		33 (0.4%)		31 (0.4%)	
0600-0659	129 (1.5%)		130 (1.5%)		109 (1.3%)	
0700-0759	355 (4.3%)	2231 (26.7%)	383 (4.4%)	2311 (26.5%)	381 (4.5%)	2440 (28.5%)
0800-0859	399 (4.8%)		411 (4.7%)		454 (5.3%)	
0900-0959	457 (5.5%)		515 (5.9%)		481 (5.6%)	
1000-1059	490 (5.9%)		480 (5.5%)		583 (6.8%)	
1100-1159	530 (6.4%)		522 (6.0%)		541 (6.3%)	
1200-1259	523 (6.3%)	986 (11.8%)	569 (6.5%)	1092 (12.5%)	534 (6.2%)	1010 (11.8%)
1300-1359	463 (5.5%)		523 (6.0%)		476 (5.6%)	
1400-1459	447 (5.4%)	2535 (30.4%)	517 (5.9%)	2720 (31.2%)	503 (5.9%)	2599 (30.4%)
1500-1559	440 (5.3%)		524 (6.0%)		546 (6.4%)	
1600-1659	474 (5.7%)		523 (6.0%)		476 (5.6%)	
1700-1759	572 (6.9%)		543 (6.2%)		505 (5.9%)	
1800-1859	602 (7.2%)		613 (7.0%)		569 (6.7%)	
1900-1959	653 (7.8%)	2256 (27.0%)	670 (7.7%)	2237 (25.7%)	613 (7.2%)	2217 (25.9%)
2000-2059	613 (7.3%)		570 (6.5%)		612 (7.2%)	
2100-2159	460 (5.5%)		443 (5.1%)		496 (5.8%)	
2200-2259	345 (4.1%)		362 (4.2%)		327 (3.8%)	
2300-2359	185 (2.2%)		192 (2.2%)		169 (2.0%)	
總數	8346		8717		8551	

資料來源：本研究整理自2010年4月至6月臺北市1999違規停車派工資料。

其次，本研究將三個月間之違規停車派工紀錄，根據違規停車時段（最早 7 點、最晚 23 點）與一般生活作息，分為五個時段加以分類，包括深夜到清晨之 00 點 00 分至 06 點 59 分、涵蓋早上交通尖峰時間之 07 點 00 分至 11 點 59 分、中午午休時間之 12 點 00 分至 13 點 59 分、涵蓋下午交通尖峰時間之 14 點 00 分至 18 點 59 分、以及夜間休閒應酬時間之 19 點 00 分至 23 點 59 分。根據上述分類，本研究針對臺北市違規停車派工紀錄位置點進行平均最近鄰分析。平均最近鄰分析指數是針對點資料的群聚狀態進行統計評估，提供點的群集程度指數。若平均最近鄰分析指數大於 1，則證明該資料屬於分散類型的分佈；計算分析結果的數值小於 1，則證明該資料屬於具有群聚現象的分佈狀態。計算的結果若接近 1，則顯示資料的分佈情形屬於隨機的型態。分析結果顯示，臺北市在三個月之中的五個時段中之派工資料，其平均最近鄰分析指數皆小於 1，而且 Z 值也都小於 -2.58。因此，本研究發現臺北市的違規停車，於空間分布狀態呈現顯著的群聚現象（參見表 3）。

表3 臺北市1999違規停車派工各時段平均最近鄰分析結果

時段	4月ANN (Z)	5月ANN (Z)	6月ANN (Z)
0000-0659	0.388423 (-21.47815) **	0.447363 (-19.975834) **	0.420817 (-18.705505) **
0700-1159	0.400576 (-54.164472) **	0.368895 (-58.040682) **	0.359787 (-60.499245) **
1200-1359	0.455017 (-32.738013) **	0.453531 (-34.546765) **	0.497339 (-30.560949) **
1400-1859	0.360485 (-61.598599) **	0.379234 (-61.936055) **	0.360801 (-62.340433) **
1900-2359	0.313016 (-62.423368) **	0.338491 (-59.854915) **	0.32546 (-60.760558) **

註：\*\*表示  $p\text{-value} \leq 0.01$ 。

資料來源：本研究分析整理。

至於點資料常用的群聚分析方法，主要為核密度估計。圖 3 係分別表示 2010 年 4 月、5 月與 6 月間等三個月，以「核密度估計」所呈現的臺北市違規停車密度群聚差異。本研究利用核密度估計，以 500 公尺為估計違規停車的影響範圍，設定其搜尋半徑，以建立空間影響連續變化的過程。本研究以點資料進行局部群聚分析，並比較三個月內各個時段的時空群聚差異。首先，在 4 月深夜清晨時段，本研究發現違規停車派工紀錄密度較高的區域，主要分布在以下幾個區域：中正區西部（建國里、南門里

與龍福里的交界區域)、中正區北部(梅花里);中山區南部(松江里、中央里、中庄里與中吉里的交界處)、以及信義區中部(松友里與安康里交界處)。至於5月的部分,本研究發現密度較高的區域,分佈狀態與4月的分佈相似,但中正區密度較高的位置移往西南部的聚盛里與聚葉里交界處。6月違規停車派工紀錄密度較高的區域,除了與前兩個月的分佈狀態相類似外,另外信義區中部的興隆里附近與北投區中部的林泉里與文化里,也出現密度較高的情形。總體而言,臺北市深夜清晨時段違規停車的密度分佈狀態差異不大,主要集中於中山區、中正區與信義區。此外,4月的高密度區域分佈狀態較聚集,5月與6月的分佈的範圍有擴大現象(參見圖2)。

其次,在涵蓋上午尖峰時間7點00分至11點59分的部分,本研究發現無論在4月、5月與6月,臺北市違規停車密度最高的區域都位在中正區北部的梅花里附近。而4月與5月的高密度區域分佈狀態較集中,但6月則有其他區域亦出現較高密度的現象,例如:士林區的蘭雅里;中正區建國里、南門里與龍福里的交界區域;松山區的精忠里與東勢里;以及信義區西村里與六合里(參見圖3)。總體而言,臺北市早上至中午時段,違規停車的密度分佈集中於中正區北部。

至於,在中午12點00分至13點59分的休息用餐時間部分,本研究發現4月與5月違規停車派工密度較高的區域主要在大安區西部的龍安里、福住里、永康里與景安里的交界處,以及中正區北部的梅花里。然而,6月的密度分佈狀態範圍較擴大,像中正區便由梅花里向西南延伸至幸福里與東門里。此外,密度較高的區域還有松山區中部的精忠里、信義區西北部的興隆里,以及士林區中部的芝山里與蘭雅里(參見圖4)。總體而言,密度較高的區域集中在中正區與大安區。

針對涵蓋下午交通尖峰時間之14點00分至18點59分的分析發現,連續三個月在該時段密度最高的區域皆為中正區北部的梅花里。而5月與6月的核密度估計圖顯示,信義區的興隆里亦存在較高的違規停車密度(參見圖5)。總體而言,核密度估計值較高的區域位在中正區與信義區。至於涵蓋夜間應酬休閒時間之19點至24點的分析部分,本研究發現4月核密度較高的區域分布在中正區的梅花里與建國里,信義區的興隆里、興仁里,信義區三張里、雙和里、景新里與景勤里交界處,以及松山區的

精忠里等。5 月的核密度估計結果則顯示，除上述區域密度較高外，大安區學府里、文山區興光里、文山區華興里、文山區興得里、士林區百齡里、士林區福華里、內湖區紫雲里、內湖區清白里、北投區溫泉里與北投區林泉里亦呈現密度較高的現象。至於 6 月的分布狀態，則與 4 月的分布型態相類似（參見圖 6）。總體而言，密度較高的區域集中在中正區、中山區、信義區與大安區。至於內湖區、文山區、松山區、士林區、北投區等地，也是有某種程度的密集現象，但未較前面所提及的中正區、中山區、信義區與大安區來得高。

由前述的空間平均近鄰分析結果來看，本研究發現臺北市違規停車派工案件的分佈，確實存在顯著的群聚現象。另外，經由核密度估計，本研究則發現臺北市違規停車派工案件的空間分佈特質。首先，不論時間點的早晚，中正區北部梅花里附近的密度都相當的高。其次，在 19 點 00 分至隔天早晨 06 點 59 分之間，中正區西部建國里、南門里與龍福里的交界區域也是密度較高的區域。除此之外，松山區的中部、信義區的西北部與大安區的中南部，於特定時段亦有違規停車派工案件高密度的現象出現。這些位置都是未來臺北市政府在思考取締或降低違規停車問題時，需要特別留意的區域。

至於，這些違規停車派工案件密度較高區域的現象，其背後的原因當然也十分關鍵。本研究一面參考過去相關的研究，另一面訪談臺北市研考會相關負責的同仁，大致發現一些可能的相關因素。在廖興中、廖洲棚、陳敦源（2013）研究中，就曾針對 2010 年 4 月至 6 月期間臺北市 1999 市民熱線違規停車案件，利用著重空間異質性分析的地理加權迴歸模型（Geographically Weighted Regression，簡稱 GWR），針對這些熱點出現的地點之重要相關因素進行探索。其研究結果顯示這些熱點與生活活動聚點（包括中央與市府相關機關、學校、圖書館、醫院及工商業聚點）的密度具有較高程度的正相關。換言之，民眾為了生活相關事務至公務或私人機構洽辦，可能因著圖方便而臨時停車成為熱點可能出現的相關因素。其次，臺北市政府研考會相關負責同仁則提到在當時的 2010 年 4 月、5 月、6 月交通會報中，針對 1999 違規停車處理陳情案件數高密度的區域也提出幾個可能的原因，主要包括市中心交通脈衝、住商混和區、常態性巷道內違規停車等，也都可能與違規停車陳情密度較高有關。

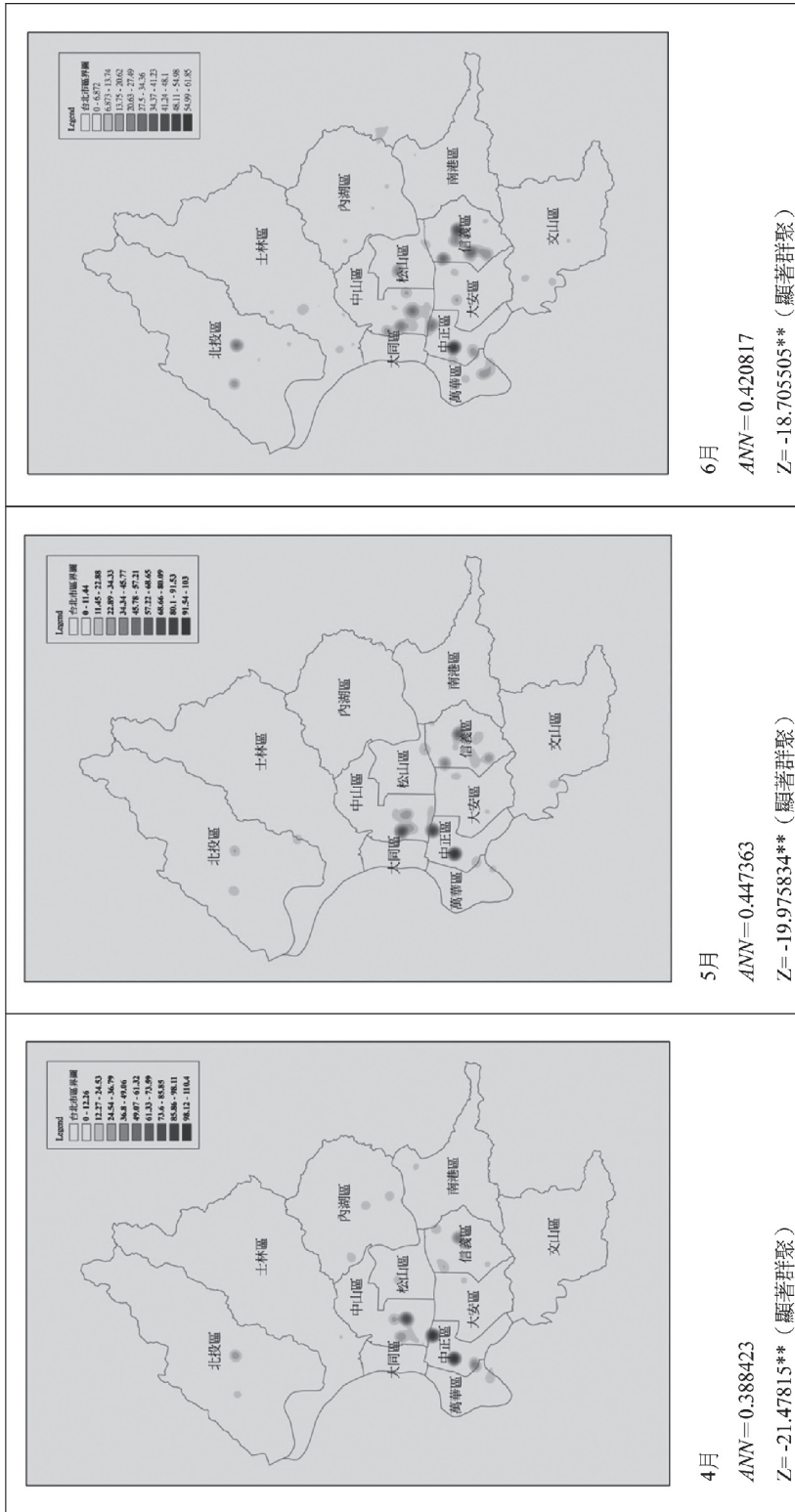


圖2 00:00至06:59違規停車派工次數之核密度估計結果

資料來源：本研究根據2010年4月至6月臺北市1999違規停車派工資料製作。



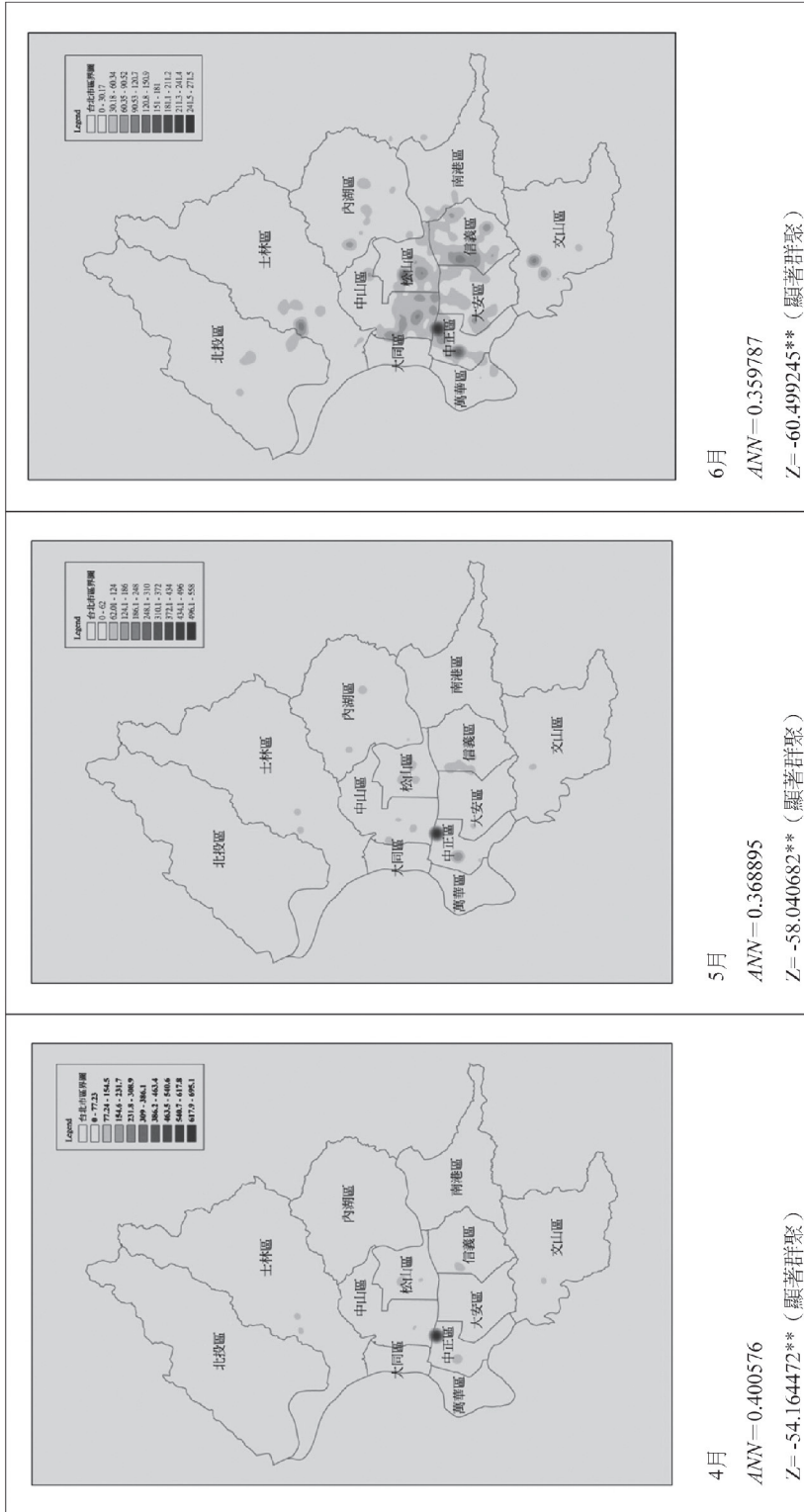


圖3 07:00至11:59違規停車派工次數之核密度估計結果

資料來源：本研究根據2010年4月至6月臺北市1999違規停車派工資料製作。



圖4 12:00至13:59違規停車派工次數之核密度估計結果

資料來源：本研究根據2010年4月至6月臺北市1999違規停車派工資料製作。

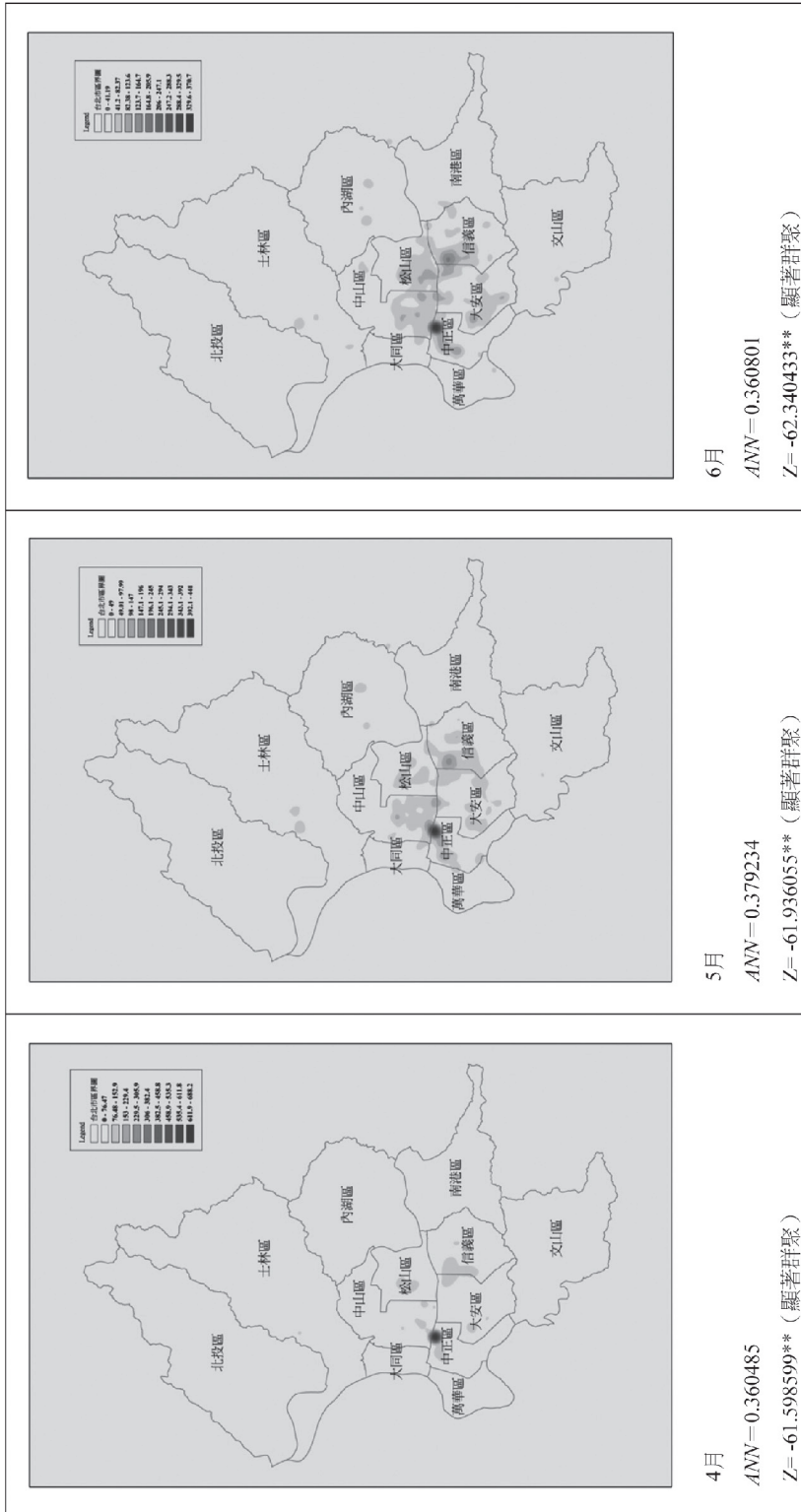


圖5 14:00至18:59違規停車派工次數之核密度估計結果

資料來源：本研究根據2010年4月至6月臺北市1999違規停車派工資料製作。

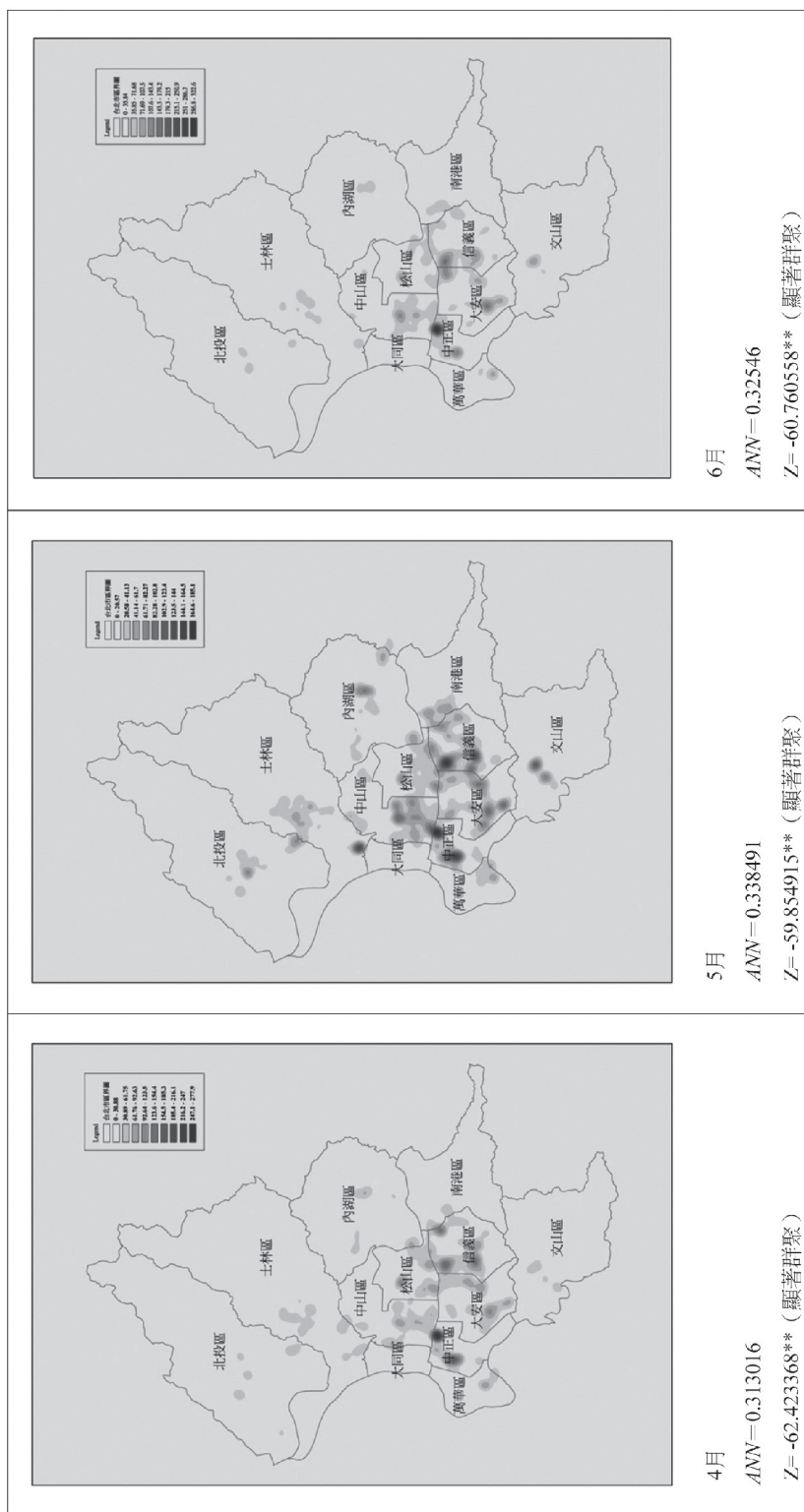


圖6 19:00至23:59違規停車派工次數之核密度估計結果

資料來源：本研究根據2010年4月至6月臺北市1999違規停車派工資料製作。

## 伍、結論

本研究說明如何利用地理資訊系統中空間分析與統計的方法，來進行空間資料探勘；並且以 1999 市民熱線在 2010 年 4 月至 6 月之違規停車派工案件為範例，說明空間資料探勘的分析結果。值得注意的是，當使用者在利用各種空間分析或統計方法進行資料探勘時，務必先行瞭解各種方法的基本假設、分析方法的限制及分析結果正確的解讀方式。以本研究為例，由於分析資料是以 1999 市民熱線派工案件呈現的點資料為主，因此必須因應點資料的型態，選擇合適的空間分析方法。在空間群聚性分析的方法論上，點資料的分析主要常使用平均近鄰分析以及核密度推估兩種方法。平均近鄰分析主要用來確認，空間中點資料的分佈是否在統計呈現顯著的群聚現象。本研究結果顯示，臺北市違規停車之平均近鄰分析的檢測值均小於 1 且 Z 值皆小於 -2.58，證實這些案件於不同的時段，在空間中的分佈確實呈現顯著的群聚現象。就核密度分析的方法論而言，其目的在於幫助研究者界定出特定現象發生密度較高的區域，以便據此資訊進一步探索此些現象發生的原因。研究分析結果顯示，臺北市於 2010 年 4 月和 6 月違停密度較高的區域集中在中正區的梅花里與建國里，信義區的興隆里、興仁里、三張里、雙和里、景新里與景勤里交界處，以及松山區的精忠里等；5 月違停包括除上述區域密度較高外，還包括大安區學府里、文山區興光里、華興里、興得里、士林區百齡里、福華里、內湖區紫雲里、內湖區清白里、北投區溫泉里和林泉里等地域也有略高的密度。由此可知，這個分析工具的優勢在於提供可靠性的地理預測資訊，來協助公共管理者進一步探索造成違停群聚的原因，以進一步思考解決特定區域發生高密度違停的方法。

伴隨著資訊科技的進步，政府已有能力大量收集與建構空間資料，因此如何利用空間資料探勘幫助政府自空間資料中挖掘有用的管理知識，以及如何將地理資訊系統和空間資料探勘方法結合，已成為未來重要的

行政研究議題。從本研究分析成果顯示，空間資料探勘分析可用於挖掘出發生於行政區域內的特定現象的分佈狀態，若配合長期的時空分析，更可進一步挖掘特定現象發生密度較高的區域。以本研究為例，若公共管理者能得知違規停車密度較高的區域，將持續集中於中正區北部某些地域，那麼處理違規停車的相關資源配置，就可以進行因地制宜的考量，以達到快速回應民眾需求的目的。為此，建議臺北市政府權責單位可持續針對 1999 市民熱線違規停車派工案件進行更長時間的分析，觀察是否呈現與本研究發現一致的違停空間分佈樣態。若仍然有固定的空間分佈樣態，表示這些群聚區域的發生原因有可能與當地的空間環境系絡有關，可以參考本研究根據過去研究與實務訪談所得的相關因素來分析瞭解。例如：是否為生活活動聚點密度較高的區域？是否為交通脈衝？住商混和的情況？是否為常態性巷道內違規停車的區域？像目前在臺北市政府的交通會報中，便已經開始針對 1999 資料與交通大隊取締數目所顯示的交通尖峰時間違規停車熱點，進行實地的現況查證。然而這項違規停車熱點的現況查證，仍僅就執勤情形、違規現況及硬體設施現況進行紀錄。未來建議臺北市政府，可以在現況查證中納入空間環境系絡的項目進行紀錄。此外，可以針對違規停車行為者的反應與行為，進行深入的瞭解與分析，也是一種可行的研究方向。畢竟，違規停車的行為除了環境系絡外，也涉及到個人行為認知的反應。這些相關資料的分析，都有助於未來臺北市政府對於違規停車問題政策的再思考與再創新。

另外有許多其他在都市中可能有的交通問題也可以透過本研究的方法，進行循證（evidence-based）研究，挖掘出熱點區域。例如臺北市政府日益增多的 YouBike，與其相關車禍意外（與行人以及與車）的資料便可以進行空間資料探勘，針對這些熱點區域進行環境系絡相關分析，進而嘗試規畫對應的策略，例如自行車道的規劃便可以參考這類的分析結果。另外像是路面不平的通報資料，也可以利用空間資料探勘進行分析。一面可以嘗試找出道路不平頻率較高的可能區域外，進而深入瞭解其道路經常不平的關鍵因素；另一面則藉著這樣道路不平熱點位置的公開，達到透明甚至課責的效果。這樣的分析有助於道路品質的管理與提升。此外，建議地方政府可以針對有留下資料的市政議題，可以用去識別化之後對外公

開，請民間力量群眾外包（crowd-sourcing）來進行資料探勘分析，結合民間分析資料的能量，讓政府決策更有循證的基礎。

然而，本研究仍然資料與研究方法等兩方面的限制。首先在資料部分，本研究在利用數量龐大與多元性繁雜的資料庫，進行資料探勘分析的同時，除了需要較大的人力之外，最關鍵的問題還是資料雜訊的解決。因此，研究團隊面對到的最大限制即在於地址定位的問題。在 1999 市民熱線的資料中，雖然單位所提供的紀錄已經包含位置的相關資訊。研究者仍然必須利用大量人力，進行資料雜訊的解決，除了費時、費力之外，還必須面對大量資料無法定位的嚴重問題。因為許多的個案可能記錄了不完整的地址（例如：臺北市萬大路與雙和街交叉）、以及口語式的位置資訊（例如：東湖捷運站 2 號出口附近）；而因著這些資料並未記錄派工案件的完整地址，導致系統無法提供少數個案的座標，造成某些資料可能被捨棄於分析之外。若這樣的資料數量過於龐大，將可能造成分析結果的偏誤。因此，本研究建議未來相關單位在進行資料建構的同時，應特別考慮到空間位置資料取得的完整性與精確性，以便分析者在後續進行資料處理時，可避免花費過多的人力與時間成本，並降低因資料不完整而導致的分析結果偏誤。另外，正如先前已經提到的，臺北市政府僅提供 2010 年 4 月至 6 月期間的違規停車派工資料，本研究所獲得的分析結果，並無法類推至其他月分。因此，本研究僅能就臺北市政府所能提供之具時間區間限制的資料，承襲當前大數據研究強調循證研究的觀點，進行空間資料探勘的嘗試，並無法證明其他月份是否與本研究所發現的現象一致。未來若要取得可類推的結果，可能需要時間範圍更長（跨年）的相關資料。

其次在研究方法方面，本文對於資料探勘結果的討論，主要是聚焦於 2010 年 4 月至 6 月違規停車派工事實的詮釋，這主要是受限於本研究的分析方法，僅能提供現象在空間上群聚的偵測，並無法解讀現象背後的相關因素。至於這些群聚現象發生的相關環境系絡因素（例如停車空間、人口密度、商業活動據點密度等），本研究則是引用其他相關研究成果，以及徵詢主管機關同仁的主觀意見後，提出了一些可能的解釋。基於大數據循證研究的觀點，後續的研究可以結合更多其他的客觀數據進行關聯分析，例如違停地點的停車場距離、停車位轉換率、鄰近特殊商家（如便利

商店、銀行)於違停時間的來客數、電話通報者報案頻率、違停時間車流量等客觀數據，來進一步推測違停事實與其他事件間的關聯性。但是在行為者的動機部分，由於本研究資料所記錄之案件已發生較久，加上原始資料未提供具體違停行為人的個資，確實不容易利用其他調查方法來瞭解其違規停車的動機。當然研究者可以先利用違規停車相關資料先進行熱點偵測，除了針對這些熱點除了進行環境系絡特質的相關性分析外，可以再輔以質性訪談來瞭解違停動機。訪談部分的操作可以透過警政單位提供過去違停熱點附近曾收到違規停車罰單之民眾資料，在取得當事人同意下進行深度訪談。抑或可以直接於熱點區域，在取得違停民眾同意下，進行現地簡略訪談。當然這兩種訪談方式，都需要留意其具有相當程度的侵入性，在爭取訪談違規停車民眾的機會方面也相對較低。當然訪談違停熱點當地的里長、店家、或居民的機會較大，也可以是一個用以瞭解違停現象或行為的另一個調查管道。然而，研究者還是需要留意這也僅能作到某種程度的推論，畢竟受訪者並非行為者本身。不論如何，未來針對違規停車行為背後動機的研究，對於城市交通管理確實有其重要性。

藉由我國電子化政府的政策的推展，各政府機關已陸續將相關行政作業電子化，惟在達成行政電子化以提高行政作業效率目的的同時，各政府機關亦留下了大量的、龐雜的、積累速度極快卻沒有好好利用的數位化資料—政府大數據 (Government Big Data)。隨著愈來愈多大數據分析工具的發明以及有創意的使用，政府大數據分析將可以成為找出公共問題的病因、協助解決公共問題以達到政府回應價值的利器。本研究展示了公共管理者在面對龐雜的空間資料時，可以如何透過適當的空間資料探勘分析與統計方法，來挖掘出隱藏在資料中的空間分佈樣態，相信對於其未來進行深入的調查或相關政策的規劃都有相當的助益。綜合而言，本研究嘗試利用 1999 市民熱線的違規停車派工資料來呈現這個方法於分析地理數位資料的優勢，未來除期望能引起研究者關注並共同思考引入這個分析工具於公共行政研究的可能性外，亦期望能刺激公共管理者進一步思考如何有效運用這個分析工具來分析政府大數據，俾找出更有意義的行政管理知識來提高政府的治理能力。



## 參考文獻

### 一、中文部分

- 石計生，2001，《人文社會 GIS》，臺北：儒林圖書公司。
- 林祥偉，2011，〈119 急難救護資料探勘之研究－以花蓮縣為例〉，《地圖》，21: 1-14。
- 廖興中、廖洲棚、陳敦源，2013，〈1999 臺北市民當家熱線違規停車陳情個案之空間資料探勘：以地理加權迴歸（Geographically Weighted Regression, GWR）為例〉，發表於「2013 年中國政治學會年會暨全球政經局勢的變與不變：國家 vs 社會／環境 vs 發展／分配 vs 正義國際學術研討會」（11 月 16 日），臺中市：中興大學國家政策與公共事務研究所。

### 二、英文部分

- Anselin, Luc. 2004. *GeoDa 0.9.5-i, An Exploration Spatial Data Analysis (ESDA) Software Application*. from <http://geodacenter.asu.edu/>. Retrieved June 26, 2014.
- Brimicombe, Allen. J. 2003. *GIS, Environmental Modelling and Engineering*. London: Taylor & Francis.
- Burt, James, and Gerald Barber. 1996. *Elementary Statistics for Geographers*. London: The Guilford Press.
- Chen, Ming-Syan, Jiawei Han, and Philips S. Yu. 1996. "Data Mining: An Overview from Database Perspective." *IEEE Transaction on Knowledge and Data Engineering*, 8: 866-883.
- Clark, J. Philip, and Francis C. Evans. 1954. "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations." *Ecology*, 35: 445-453.
- Cliff, Andrew D., and John K. Ord. 1981. *Spatial Processes: Models and Applications*. London: Pion.
- Eck, John E., Spencer Chainey, James G. Cameron, Michael Leitner, and Ronald E. Wilson. 2005. *Mapping Crime: Understanding Hot Spots*. from <http://discovery.ucl.ac.uk/11291/1/11291.pdf>. Retrieved June 26, 2014.
- Ester, Martin, Hans-Peter Kriegel, and Jörg Sander. 1997. "Spatial Data Mining: A Database Approach." In *Advances in Spatial Databases*. Berlin Heidelberg: Springer, 47-66.

- GoodChild, Michael F. 1992. "Geographical Data Modeling." *Computers and Geosciences*, 18(4): 401-408.
- Guo, Diansheng, and Jeremy Mennis. 2009. "Spatial Data Mining and Geographic Knowledge Discovery-An Introduction." *Computers, Environment and Urban Systems*, 33(6): 403-408.
- Han, Jiawei, and Micheline Kamber. 2006. *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan Kaufmann.
- Harvey, David W. 1996. "Geographical Processes and the Analysis of Point Patterns: Testing Models of Diffusion by Quadrat Sampling." *Transactions of the Institute of British Geographers*, 40: 81-95.
- Huxhold, William E. 1991. *An Introduction to Urban Geographic Information Systems*. New York: Oxford University Press.
- Knox, E. G. 1964. "The Detection of Space-Time Interactions." *Applied Statistics*, 13: 25-29.
- Lawson, Andrew. B. 2001. *Statistical Methods in Spatial Epidemiology*. Chichester: Wiley.
- Maguire, David J. 1991. "An Overview and Definition of GIS." *Geographical Information Systems: Principles and Applications*, 1: 9-20.
- Mantel, Nathan. 1967. "The Detection of Disease Clustering and a Generalized Regression Approach." *Cancer research*, 27(2): 209-220.
- Roddick, John F., and Myra Spiliopoulou 1999. "A Bibliography of Temporal, Spatial and Spatio-temporal Data Mining Research." *ACM SIGKDD Explorations Newsletter*, 1(1): 34-38.
- Shekhar, Shashi, and Sanjay Chawla. 2003. *Spatial Databases :A Tour*. Englewood Cliffs, NJ: Prentice-Hall.
- Snow, John. 1855. *On the Mode of Communication of Cholera*. London: John Churchill.
- Spielman, Seth E., and Jean-Claude Thill. 2008. "Social Area Analysis, Data Mining and GIS." *Computers Environment and Urban Systems*, 32(2): 110-122.
- Star, Jeffrey, and John Estes. 1990. *Geographic Information Systems: An Introduction*. Englewood Cliffs, NJ: Prentice-Hall.
- Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. 2005. *Introduction to Data Mining*. Reading, MA: Addison and Wiley.
- Witten, Ian H., and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Morgan Kaufmann Publishers Inc.
- Wu, Xindong. 2004. "Data Mining: An AI Perspective." *IEEE Intelligent Informatics Bulletin*, 4: 23-26.

# Spatial Data Mining on Parking Violation Cases in 1999 Citizen Hotline of Taipei City Government: Kernel Density Estimation as the Tool

Hsin-Chung Liao\*    Zhou-Peng Liao\*\*    Don-Yun Chen\*\*\*

## Abstract

Using spatial data mining method, this study analyzed parking violation cases from April to June in 2010, which were provided by 1999 citizen hotline of Taipei City Government, and tried to understand the locations of incidents and the trend. The results of Average Nearest Neighbor and Kernel Density Estimation revealed that parking violation cases did exhibit significant spatial dependence in Taipei City, and the areas with higher density of parking violation are included northern and western Zhongzheng, middle Songshan, northwestern Xinyi, and middle southern Da'an. In general, this study presented how Spatial Data mining can help the government to find the public opinion knowledge hidden in the data in order to further develop strategies and support policies.

**Key Words:** spatial data mining, geographic information system, average nearest neighbor, Kernel Density Estimation, 1999 citizen hotline

---

\* Corresponding author, Associate Professor, Department of Public Policy and Management, Shih Hsin University. E-mail: [lhc@cc.shu.edu.tw](mailto:lhc@cc.shu.edu.tw).

\*\* Assistant Professor, Department of Public Administration, National Open University.

\*\*\* Professor, Department of Public Administration, National Chengchi University.

