

科技部補助  
大專學生研究計畫研究成果報告

\* \*\*\*\*\* \*  
\* 計 畫 \*  
\* : 空間異質性對全域及區域空間相依性的影響 \*  
\* 名 稱 \*  
\* \*\*\*\*\* \*

執行計畫學生： 杜承軒  
學生計畫編號： MOST 105-2815-C-002-159-M  
研 究 期 間： 105 年 07 月 01 日至 106 年 02 月 28 日止，計 8 個月  
指 導 教 授： 溫在弘

處理方式： 本計畫可公開查詢

執 行 單 位： 國立臺灣大學地理環境資源學系暨研究所

中華民國 106 年 03 月 27 日

## 空間異質性對全域及區域空間相依性的影響

### 一、摘要

空間自相關的檢定，能看出事件分布整體的集中趨勢及熱區的發生位置，在生態學、流行病學、犯罪地理學、經濟地理學、人口研究等領域都有應用。全域空間自相關能看出整體空間分布型態，區域空間自相關能找出熱區，然而，在全域空間自相關不顯著時，代表事件在整體是隨機分布，但區域端仍然會有可能出現熱區。

本研究以模擬實驗的方式，觀察在研究區整體呈現完全空間隨機分布時，區域端熱區辨識的情況。首先確認經過多重檢定的校正後，是否還會出現顯著熱區，並探討熱區的出現是否隨著事件密度影響。結果顯示，經校正後仍會出現熱區，且隨密度越小，辨識出熱區的機率與出現熱區的網格數都增加。即在空間隨機情形下，仍會產生熱區，且熱區的辨識與事件密度呈現負相關。

關鍵字：空間相依、空間自相關、熱區分析、Getis-Ord General G、Local G-statistic

### 二、研究動機與研究問題

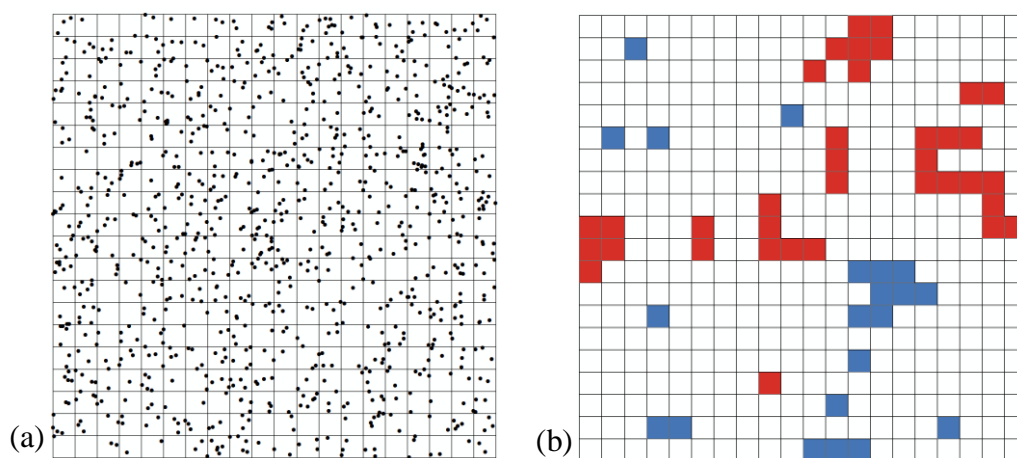
Tobler (1970) 提出地理學第一定律，指出「每個事物都彼此相關，而越接近者相較於距離越遙遠者的關聯性會越大」，說明資料存在空間相依的特性，因此討論事件在空間之間的關聯性，對於空間資料的分析中是常常被討論的。空間自相關 (spatial autocorrelation) 是空間相依程度的指標，代表空間單元與鄰近單元是否相似，空間單元與鄰近空間的同一變量之相似性，空間自相關為正相關且程度越高，則表示該變量之空間分布特性越群聚，反之，空間自相關為負相關，則空間分布越分散，若沒有空間自相關，則為隨機的空間分布形態。空間自相關即可以量測事件的發生或某屬性變量，是否受到空間鄰近的因素影響。而空間異質性 (spatial heterogeneity) 通常被定義為在空間中，一系統屬性複雜性與多變性 (Li and Reynolds 1994)，Li and Reynolds (1994) 提出空間異質性與空間中區域區塊的特性有關，如區塊種類、每個種類的比例、不同區塊的空間排列、區塊的形狀、鄰近區塊的對比，並由這些組成因素來量化空間異質性。

計算空間自相關的方法有許多，主要分成「全域型空間自相關」與「區域型空間自相關」。全域型空間自相關能描述整個空間整體的分布情形，計算出一個統計值來說明，此變量在研究區內有無空間自相關的特性，用來分析地理現象有無群聚的現象，如 Moran (1950) 提出 Moran's I、Geary (1954) 提出 Geary's C 等等，而 Cliff and Ord (1973) 提出 General G，更針對群聚現象，能進一步檢測是強群聚或弱群聚。區域型空間自相關則可以進一步了解每一個空間單元與其鄰近單元的相關趨勢，經由統計後判斷聚集區域的範圍，觀察出空間聚集的地區，如 Getis and Ord (1992) 提出的 Local G-statistic ( $G_i^*$ ) 與 Anselin (1995) 提出的 Local Moran's I (LISA)。

透過計算空間自相關，能看出事件是否有群聚的現象，以及群聚的熱區分布。在確認是否有空間相依存在時，進行全域型空間自相關分析計算確認是否整體有聚集的趨勢後，並進行區域型空間自相關分析，觀察聚集區域的分布位置，進而對該事件熱區做出解釋，如對於疾病的擴散現象與犯罪的聚集，先進行研究區整體的自相關計算，確認有聚集趨勢後，進而對於疫區或熱區加以治療、防範，能更有效的解決問題。

然而，並非所有研究區內整體空間分布情形都有顯著的群聚趨勢，若是在全域分析中沒有群聚的趨勢，有研究就不再進行區域型空間自相關的分析，例如，薛益忠與郭士鳳 (2010) 對於 1999 年~2005 年腸病毒病例進行空間自相關的分析，發現有些年份的全域空間自相關不顯著而呈現隨機分布，如 1999、2001 與 2005 的一般病例，1999、2004 年的重症病例，便沒有進行這些年份的區域空間自相關的分析，跳過這些年份的資料，只針對有全域空間自相關的年度進行區域指標的分析，在觀察隨時間改變的變遷現象。

然而，在全域空間自相關中為隨機分布的事件，在熱區的分析中，仍然會有群聚區域的出現，如圖 1。圖 1(a)為20×20的網格資料與隨機產生的 1000 筆



▲圖 1

在20×20的網格中，隨機產生 1000 個資料點，加總到網格後進行空間自相關計算。圖 1(a)為原始資料，圖 1(b)為判釋出的熱區(紅)及冷區(藍)。其全域空間自相關不顯著，但仍產生 36 個熱區與 21 個冷區，表示區域空間有相依的現象。

資料點，用 Moran's I 與 General G 計算全域空間自相關，Moran's I 指標為 0.005，Z 分數為 0.28，General G 指標為 0.019，Z 分數為 0.37，結果皆判定為隨機分布，而使用 Local G-statistics ( $G_i^*$ ) 分析的結果如圖 1(b)，在 5% 的顯著水準下，將  $G_i^*$  中 Z 值超過  $\pm 1.96$  的視為有群聚產生，定義出熱區及冷區，出現熱區的範圍如紅色區塊，冷區為藍色區塊，因此，表示在隨機狀況下，區域端仍然會有熱區產生。

區域端產生的熱區，受到全域空間的自相關不顯著，是否為真正事件發生的聚集處，或疫情需要注意的重點，或犯罪頻率高、需要更多警力關注的發生地區。以上述案例為例，在薛益忠與郭士鳳 (2010) 的腸病毒病例分析中，因全域空間自相關不顯著，則不討論區域之間的聚集狀態，但如果全域分析不顯著的年份之資料，也做出區域的自相關計算，所出現的熱區是否也能被視為疫情嚴重的地點，進而與一般群聚年份的熱區進行比較，能更完整看出腸病毒隨著時間在空間的變遷樣態。

因此，確定全域空間自相關不顯著時所產生的熱區之意義，能不受到全域相依性低的限制，從區域尺度看出真正的空間分布樣態，更進一步，若能確定區域空間自相關會受到全域空間自相關或其他參數所影響，能釐清區域與全域相依性的關係，即使全域相依性低，也能找出區域型熱區出現率的關係。

首先，需要檢測在全域空間自相關不顯著時，產生的熱區是否是真正顯著的熱區，抑或是受到多重檢定因素影響，而使熱區判識的標準降低，而多判識出許多假熱區。透過校正方法，得到更嚴謹也更準確的熱區，能得知在全域空間是隨機的狀態，是否能檢測到顯著的熱區，以及熱區被判識的比例與數量。若經過校正後，仍能判識到更嚴格的熱區，便關注熱區被判識是否隨著空間資料屬性的因素而改變。本研究假設在完全空間隨機的狀態，在研究區內事件發生的密度，會影響熱區的判識。

因此本研究欲討論在無全域空間相依性下，是否能會出現顯著的區域空間相依性，即是否會出現熱區，以及事件在研究區內發生的密度，是否影響熱區在全域空間隨機下的判識，探討密度對於全域及區域型空間自相關的影響。

### 三、文獻回顧與探討

空間自相關探討鄰近空間的群聚關係，觀察空間相依的現象，最早的概念是從全域的空間自相關來觀察出整體的聚集情形。Moran (1950) 提出的方法，基於統計學相關係數的共變異數關係推算得來，觀察兩個鄰近的空間單元  $i$  與  $j$  之間的變量與整體變量平均的趨勢，若兩者同時皆高或低於平均，表示兩者相較於平均數的變化是呈現相同方向的，則  $(x_i - \bar{x})(x_j - \bar{x})$  數值為正；反之，若

兩者對於平均數為一正一負，則 $(x_i - \bar{x})(x_j - \bar{x})$ 數值為負值，由此概念經統計的結果後，得到 Moran's I，因此 Moran's I 正相關越大也代表越群聚。Geary (1954) 提出另一個方法來量測空間自相關，相較於 Moran 使用相對於平均的概念，Geary 將 $(x_i - \bar{x})(x_j - \bar{x})$ 替換成 $(x_i - x_j)^2$ ，看出兩空間單元之間變數實際值的差異程度，稱為 Geary's C Ratio。Geary's C 的值越小，表示兩者越相似，也代表空間自相關程度越大。前兩者的方法都是判定空間自相關，因此，無論聚集處數值是高或低，都會被判定成高度空間自相關，Getis and Ord (1992) 提出的 Getis-Ord General G，在空間自相關之上更使用空間群聚度的概念，用兩者相乘的概念 $x_i x_j$ ，可以判斷檢定出的聚集是否為熱區或冷區，及高數值或的數值的聚集，能看出 Moran's I 或 Geary's C 聚集是受到高數值或低數值聚集的結果，使用熱區更能捕捉事件在空間中發生的聚集強度。

若只進行全域空間自相關的計算，只能透露出整體是否有群聚現象，而不能看出產生群聚的現象在何處發生，因此出現區域型的空間自相關，如 Getis and Ord (1992) 提出的 $G_i^*$ 統計值，針對每一個空間單元都做檢驗，觀察此空間單元所有鄰近單元的值之和，與所有研究區內的值總合的比例，並用統計方法計算出 z 分數，檢定判斷是否顯著性，定義出高數值聚集的熱區及低數值聚集的冷區的空間。Anselin (1995) 提出空間關聯性的區域指標 (Local Indicators of Spatial Association，簡稱 LISA)，並用 Moran's I 的概念設計出在區域端計算的空間自相關公式，LISA (Local Moran's I) 的值越高，表示該空間單元與其鄰近單元有空間自相關現象，因此只要檢視該空間單元的資料，判斷原始值是高或低於平均，就能判斷熱區和冷區的分布。

但在衡量區域空間自相關時，也因為全域尺度下空間相依存在，會造成區域空間自相關定義的熱區受到影響 (Ord and Geits 2001)，因此為了在考慮到全域自相關的狀況下，更精確的定義區域尺度下分布的情況，Ord and Geits (2001) 因此在有出現全域型空間自相關的情形下，修正區域空間自相關的數值方法，來檢測熱區的存在，找出真正顯著的熱區之分布狀況。

至今有許多研究也都會應用空間自相關來觀察空間關係，其應用跨足許多領域，國內研究如紀凱婷 (2008) 分析台北市建商推案的單價與平均主力坪數是否有聚集的現象；薛益忠與郭士鳳 (2010) 探討腸病毒在時間變化下的群聚現象與熱區範圍的變遷；溫在弘、劉擇昌、林民浩(2010)繪製台北市住宅竊盜犯罪地圖，也運用 Global Moran's I 和 LISA、 $G_i^*$ 來觀察聚集和分析熱區。國外研究如 Sokal and Thomson (2006) 以一個印地安部落的人口為例，用區域空間自關推論人口結構；Wulder et al.(2007) 亦用區域空間自相關分析森林成長模型中，葉面積指數和林分蓄積的結果分布。

然而，上述的研究都是先進行全域型空間自相關的分析，確認事件資料有群聚後，才接著進行區域型熱區的辨識 (Sokal and Thomson 2006；Wulder et al.

2007；紀凱婷 2008；溫在弘、劉擇昌、林民浩 2010；薛益忠、郭士鳳 2010)，即全域的指標顯示有聚集後，才會進行區域指標的分析。但會造成全域型空間自相關低時，無法與一般群聚狀況一併討論，如薛益忠與郭士鳳 (2010) 論文中，在腸病毒病例全域自相關不顯著的年份時，即不進行區域自相關檢測，無法得知該年度是否有熱區及其發生位置，因此無法和其它年度的資料做比較。

在 Ord and Geits (2001) 論文中，主要強調於在出現空間相依的特性時，如何尋找區域尺度的熱區，且試驗的流程中必須依賴於區域端熱區與全域空間的自相關之間的尺度差異，資料研究區、事件數必須大夠才能進行分割，觀察在從區域漸漸擴張到全域之間的尺度影響。在研究流程也會為了避免過度強調固有空間自相關的存在，而在區域檢驗時有所忽略全域自相關，並沒有提出完整區域與全域空間相依的關係，也沒有說明，在全域空間自相關低的情形，區域端產生的熱區如何修正，以及全域空間自相關為隨機狀態時，熱區出現的意涵與一般群聚現象的差別。

綜合上述研究，對於全域空間相依度低時，對於區域端產生熱區的討論較缺乏，以及尚未有全域空間隨機下，區域空間自相關變化的關係之文獻，因此本研究想釐清在全域空間相依度低時，全域與區域空間自相關是否會受到事件發生的密度而產生變化。

#### 四、 研究方法及步驟

##### (一)研究流程

本研究使用程式模擬，建立隨機空間分布型態，改變屬性資料的數值，進而分析在全域為隨機分布下，區域端顯著熱區的判識，觀察事件發生的密度對全域與區域空間自相關的影響。

本研究的研究區為 $20 \times 20$ 的網格，每個網格都有一個數值屬性資料，代表事件在此網格中發生的次數，而為了產生全域空間分布是隨機的型態，每個網格的數值是由 Poisson 分配隨機生成，因此需決定 Poisson 分配的參數 $\lambda$ ，代表事件發生的密度，即事件在一個網格中平均發生幾次。固定一密度後，經過 Poisson 分配隨機生成每個網格的數值，進行 Getis-Ord General G 全域檢定與 Local G-statistic ( $G_i^*$ ) 區域檢定，顯著水準設定 $\alpha = 0.05$ ，觀察 General G 的 Z 分數，判斷產生的數值是否真的在全域空間中是隨機分布，以及 $G_i^*$ 經過 FDR 校正後，所判定出來為熱區的網格數。

針對同一密度，做 1000 次模擬，記錄每一次 General G 的 Z 分數與熱區判定的網格數。並改變密度值，從平均發生次數少，每一網格平均發生 0.01 次，逐漸增加到平均發生 1000 次的情形，共模擬 46 種密度情況，並將各種情況皆做 1000 次模擬。觀察在不同密度下，是否影響熱區的判識。

## (二)方法介紹

### Getis-Ord General G

Getis-Ord General G 為全域型空間自相關與空間群聚度的指標之一，用於判斷研究區整體是否有群聚的現象。相較 Moran's I 只能判斷是否有群聚現象，General G 能進而判斷是否有高數值或低數值的群聚。以式(1)表示(Getis and Ord 1992)：

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}; j \neq i \quad \dots\dots\dots \text{式(1)}$$

其中，n 為資料數量； $x_i$  為空間單元 i 中變數的數值； $w_{ij}$  為空間單元 i 與 j 之間的鄰近權重，用於衡量空間單元 i 與 j 的鄰近性，在本研究中，鄰近定義為 Queen，即空間單元有相同的共邊或共點，就算是鄰近單元。General G 中規定計算鄰近性時不將自身空間單元納入考慮。

此方法的計算概念使用相鄰空間單元的乘積作比較，固定一空間單元 i，則比較鄰近的空間單元 j 之間的變量，若  $x_i$  與  $x_j$  的數值皆與平均沒有顯著差異，或一高一低，則其乘積與平均值不會有太大差異；若兩者數值同時都很高或很低，其乘積也會明顯高於或低於平均值，進而可以判斷是否有高數值或低數值的群聚。

計算出來的 General G 無法直接觀察，需和平均值做比較，理論上 General G 的期望值如式(2)表示(Getis and Ord 1992)：

$$E(G) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}{n(n-1)} \quad \dots\dots\dots \text{式(2)}$$

若 G 接近  $E(G)$ ，則表示沒有顯著的群聚發生。若 G 大於  $E(G)$ ，則表示空間中有高數值的群聚發生；反之，若 G 小於  $E(G)$ ，則表示空間中有低數值的群聚發生。因此，計算出 General G 的指標後，透過蒙地卡羅顯著性檢定或統計數值方法 (Cliff and Ord 1972)，將 General G 轉換成 Z 值，以判斷群聚的顯著性，如式(3)：

$$Z(G) = \frac{G - E(G)}{\sqrt{Var(G)}} \quad \dots\dots\dots \text{式(3)}$$

將 General G 轉換成 Z 分數後，在 5% 顯著水準下，若 Z 值介於 -1.96 與 1.96 之間，則整體的群聚現象較不顯著。Z 值若大於 1.96 或小於 -1.96，都代表空間分布為群聚的型態，有顯著的正向空間自相關，而前者代表整體的空間形態中有出現高強度數值的群聚現象，後者代表整體的空間形態中有出現低強度數值的群聚現象。

### Local G-statistic ( $G_i^*$ )

Local G-statistic 能檢測空間中熱區或冷區的方法，能判識出區域端空間群聚的現象，計算每一個空間單元之鄰近單元的數值，是否高於或低於研究區的平均值。分成計算鄰近時不考慮自身單元的  $G_i$  統計量，以及包含自身單元的  $G_i^*$  統計量，前者多用於討論擴散的情況，後者較多用於討論找出實際群聚的位置，即熱區或冷區出現的地點。 $G_i^*$  的計算公式如式(4) (Getis and Ord 1992)：

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j} \dots\dots\dots \text{式(4)}$$

其中， $x_j$  為空間單元  $j$  中變數的數值， $w_{ij}$  為空間單元  $i$  與  $j$  之間的鄰近權重，用於衡量空間單元  $i$  與  $j$  的鄰近性，在本研究中，鄰近定義為 Queen，即空間單元有相同的共邊或共點，就算是鄰近單元。

此方法的計算概念為計算空間單元  $i$  的  $G_i^*$  時，討論自身值與鄰近值得總和相對於整體數值和的大小。若自身值與鄰近值很大，表示是熱區發生的地點，則佔整體數值有很大的貢獻， $G_i^*$  值也會高於平均；反之，若自身值與鄰近值很小，表示是冷區發生的地點，則相對整體數值有很小， $G_i^*$  值也會低於平均。

計算後  $G_i^*$  值即為空間單元  $i$  的統計量，和 General G 一樣無法直接觀察比較，因此需透過蒙地卡羅法來進行顯著性檢定，將每一個  $G_i^*$  值轉換成 Z 分數，如式(5)：

$$Z(G_i^*) = \frac{G_i^* - E(G_i^*)}{\sqrt{Var(G_i^*)}} \dots\dots\dots \text{式(5)}$$

然而，因為每個網格都需要進行顯著性檢定，為了避免多重檢定使型一錯誤提高，本研究使用錯誤發現率(False Discovery Rate，簡稱 FDR)方法來校正，降低偽陽性的比例，找出更嚴格也更精確的熱區出現範圍(Caldas de Castro and Singer 2006)。

經過 FDR 校正過後，計算出每一空間單元的 Z 分數。在 5% 顯著水準下，若 Z 值介於 -1.96 與 1.96 之間，則此空間單元與鄰近的群聚現象較不顯著。Z 值若大於 1.96 或小於 -1.96，都代表區域端空間有群聚的型態，而 Z 值若大於 1.96 代表該空間單元被判識成熱區，而 Z 值小於 -1.96 則代表該空間單元被判識成冷區。



## 五、結果與討論

利用 Poisson 分配生成完全隨機空間分布，並改變密度從每網格平均發生 0.01 次到 1000 次，各情形都模擬 1000 次，觀察密度對全域群聚度及區域熱區辨識的影響。統計各個密度在模擬 1000 次中，全域指標 General G 轉換成 Z 分數的平均值，以及  $G_i^*$  呈現有顯著熱區的平均網格數量與次數，前者代表平均一次試驗中，有多少網格會被判識成熱區，後者為 1000 次中有幾次試驗是有判識出熱區，即平均一次試驗中有多少機率會有熱區被判識。將上述兩者相除，得到熱區判識次數除以熱區網格數量的比例，能得知在確定有熱區的情況下，每一次有多少網格被判定成熱區。結果如表 1 所呈現。

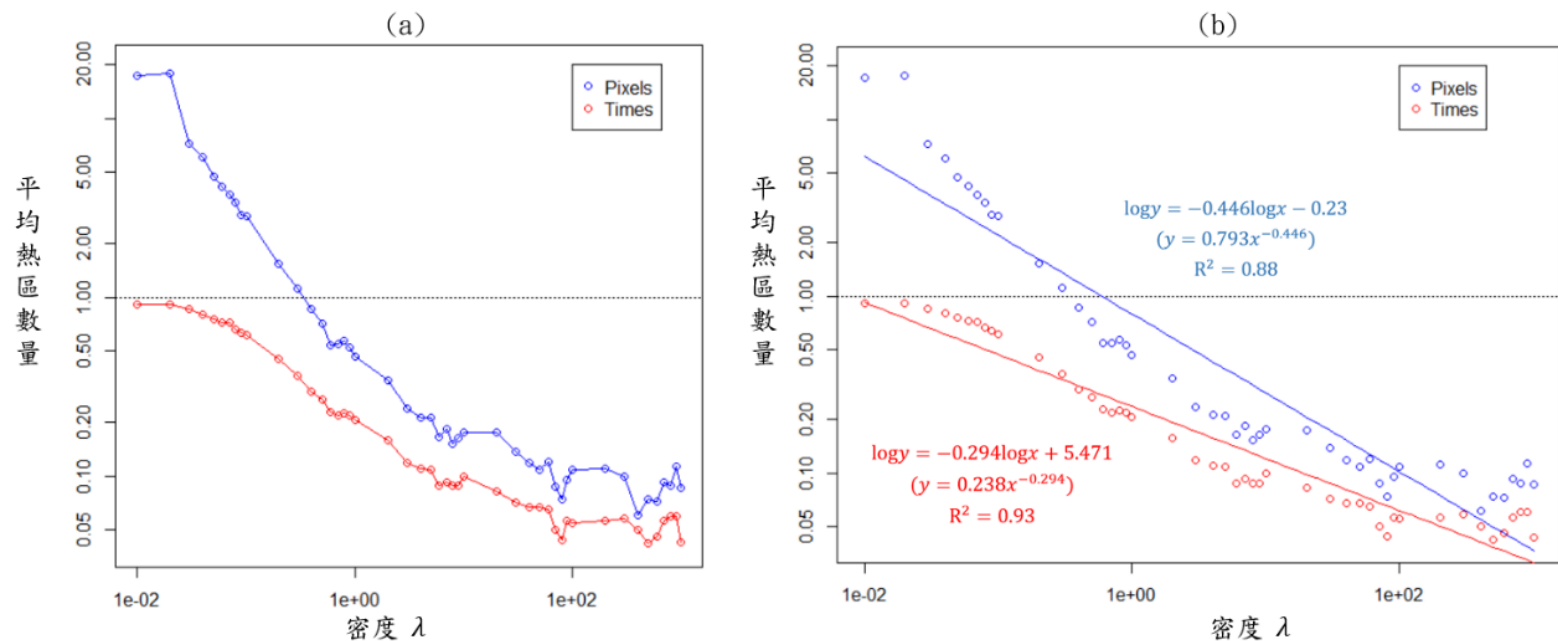
▼表 1、密度與全域群聚度及區域熱區辨識關係表。

其中  $\lambda$  為密度，表示平均一個網格中事件發生的次數； $Z(G)$  為 General G 統計量經過換算後的 Z 分數，1000 次模擬的平均，由表得知在密度改變的情況下， $Z(G)$  都不顯著，表示全域空間皆為隨機分布；熱區網格為 1000 次模擬中，平均一次會判識出幾格熱區的數量；熱區次數為 1000 次模擬中，平均有幾次會判識出熱區；比例為熱區網格除以熱區次數，表示每次判識出熱區時，平均會有幾個網格被判識。

$\lambda$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.3
Z(G)	-0.04	0.01	0	-0.05	-0.02	0.01	0.02	0.04	-0.03	-0.02	0.01	-0.02
熱區網格	17.306	17.773	7.256	6.056	4.718	4.182	3.764	3.403	2.903	2.833	1.543	1.12
熱區次數	0.915	0.916	0.855	0.801	0.76	0.725	0.719	0.667	0.635	0.612	0.451	0.364
比例	18.9	19.4	8.5	7.6	6.2	5.8	5.2	5.1	4.6	4.6	3.4	3.1
$\lambda$	0.4	0.5	0.6	0.7	0.8	0.9	1	2	3	4	5	6
Z(G)	-0.02	0.01	0	-0.02	-0.01	0	-0.06	-0.05	-0.06	0.04	0	-0.01
熱區網格	0.867	0.713	0.54	0.545	0.569	0.525	0.466	0.345	0.237	0.213	0.211	0.165
熱區次數	0.296	0.27	0.229	0.22	0.225	0.218	0.207	0.158	0.118	0.11	0.109	0.088
比例	2.9	2.6	2.4	2.5	2.5	2.4	2.3	2.2	2	1.9	1.9	1.9
$\lambda$	7	8	9	10	20	30	40	50	60	70	80	90
Z(G)	0	0.03	0	0.02	0.04	0	-0.02	0	0	0	-0.06	0.05
熱區網格	0.184	0.153	0.164	0.176	0.175	0.138	0.118	0.108	0.12	0.087	0.074	0.095
熱區次數	0.093	0.088	0.088	0.1	0.082	0.071	0.067	0.067	0.065	0.05	0.044	0.056
比例	2	1.7	1.9	1.8	2.1	1.9	1.8	1.6	1.8	1.7	1.7	1.7
$\lambda$	100	200	300	400	500	600	700	800	900	1000		
Z(G)	-0.02	-0.02	-0.03	0.05	0.02	-0.05	-0.03	-0.02	-0.04	0		
熱區網格	0.108	0.111	0.1	0.061	0.074	0.072	0.092	0.088	0.114	0.086		
熱區次數	0.055	0.056	0.058	0.05	0.042	0.046	0.056	0.06	0.06	0.043		
比例	2	2	1.7	1.2	1.8	1.6	1.6	1.5	1.9	2		

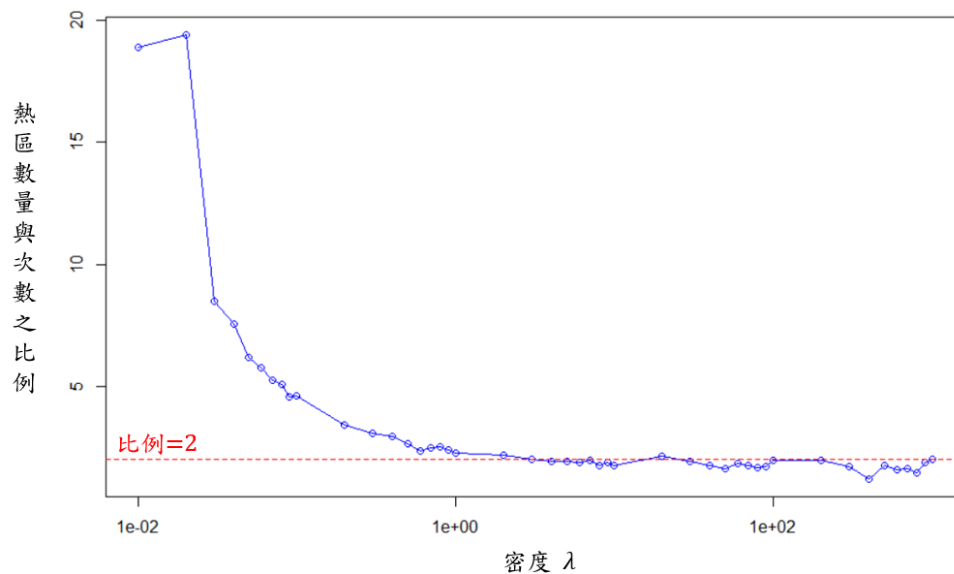
由表 1 的資料可知，我們從 Poisson 生成的數值，在全域空間中，無論密度如何改變，General G 都不顯著，證明所產生的是全域隨機分布。經過 FDR 校正後，仍能辨識出熱區，表示即使是完全空間隨機的分佈，仍可以判釋出顯著且更嚴格的熱區。

而隨著密度的改變，熱區辨識的程度也隨不同密度而影響，如圖 2(a)。能看出密度改變時，平均辨識出熱區的次數，以及每次辨識出熱區的網格數量，都隨著密度降低而升高。顯示在完全空間隨機的情況，較低的密度，有更高的機率會存在熱區，且若確定此次試驗有產生熱區，則出現熱區的網格數也隨密度降低而增多。將密度與顯著熱區的平均網格數量、次數都取對數後進行迴歸，能得到對數迴歸模型，如圖 2(b)。密度與熱區數量、次數的迴歸線之判定係數為 0.88 與 0.93，表示這兩條線對於密度與熱區有良好的解釋力，能由此推測密度與熱區辨識的關係，不僅能說明隨密度降低，辨識出熱區的次數與其網格數量都增加，密度與熱區更可能是隨某一乘冪函數關係而改變。



▲圖 2

針對表 1 中，密度與平均熱區辨識次數以及平均熱區網格數量作圖，其中 x 軸與 y 軸均取對數，藍色(上方線段)為平均熱區網格數量，紅色(下方線段)為平均熱區辨識次數，即單次試驗之熱區辨識機率。由圖 2(a)能觀察到密度下降時，熱區辨識次數與網格數量都隨之升高，表示密度越小越容易出現熱區；圖 2(b)中對兩組資料取對數後取迴歸，得到兩條乘冪形式的迴歸線，判定係數為 0.88 與 0.93，表示迴歸線對於資料有解釋力，推測密度與熱區辨識次數、網格數量可能有乘冪的函數關係。



▲圖 3

密度與熱區網格數的比例關係，表示在每次辨識出熱區時，平均被判識的網格數量。當密度高於 1，則比例都接近 2。當密度少於 1，比例則隨密度減少而提升。

另外，觀察密度與熱區網格數的比例關係(即熱區網格除以熱區次數，表示每次辨識出熱區時，平均會有幾個網格被判識)，如圖 3。能看出在密度超過 1，即每個網格平均發生次數大於一次時，當確定有熱區出現，每次試驗平均會出現兩格熱區，且不隨著密度改變而有變化。而當密度少於每網格發生一次時，每次有出現熱區之試驗中，平均出現熱區的網格數會隨著密度的遞減而有所增加，表示有更多網格被判定成熱區。

綜合實驗的結果，經過 FDR 校正後，即使全域空間自相關不顯著，即整體空間是呈現隨機分布的情形下，仍然可以找到更嚴格且顯著的熱區。而熱區的辨識會受到事件發生的密度所影響，經過多次模擬發現，密度越小，每一次試驗有辨識出熱區的機率越高，且每次出現熱區的平均網格數量，也隨密度變小而增加。對於密度大於一的狀況，當試驗中已知辨識出熱區，平均出現熱區的網格數，不隨密度影響呈現定值，平均有兩個網格為熱區，而密度小於一時，則會隨密度減少而使網格數比例增加。因此並非在完全空間隨機下就不會有熱區出現，而熱區出現的機率以及被辨識出熱區的網格數量，和事件發生的密度有密切相關，當密度越低，在區域端會更容易出現熱區。

因此，文獻回顧中，許多研究都是先判斷全域空間自相關存在，才進行熱區的辨識 (Sokal and Thomson 2006；Wulder et al. 2007；紀凱婷 2008；溫在弘、劉擇昌、林民浩 2010；薛益忠、郭士鳳 2010)，但透過此研究得知，在全域端完全隨機時，區域端仍然能找到顯著的熱區。例如，在薛益忠與郭士鳳 (2010) 的腸病毒研究中，即使在 1999~2005 年中，其中有幾年是隨機分布，我們仍然能對每一年做區域空間自相關，來監測出熱區發生地區的變化，或許更

能看出腸病毒隨年份變化，熱區移動或分布改變的現象。

透過此研究我們能得知，在全域空間自相關不顯著，區域仍然有機率找到顯著的熱區，因此在做熱區分析時，或許能不考慮事件在空間整體的分布，即全域的空間相依性低時，也針對區域自相關進行檢定。例如更能對長時間追蹤的事件，可能有些年份顯著群聚，而有些隨機分布，對於在不同時間下對於同一事件的時空變遷分析，會有更清楚和完整的結果。

綜合上述所言，在完全空間隨機的情況下，經過多重檢定的校正，仍然能找到顯著熱區存在，且存在熱區的機率與事件密度呈現負相關。尤其在罕見事件中，事件發生的密度低，即使整體事件為隨機分布，也有很高的機率能檢測出熱區。

## 六、引用文獻

紀凱婷 2008。台北市新推個案訂價之時間與空間相依性分析。國立政治大學地政研究所碩士論文。

溫在弘、劉擇昌、林民浩 2010。犯罪地圖繪製與熱區分析方法及其應用：以1998~2007 年台北市住宅竊盜犯罪為例。地理研究 52: 43-64。

薛益忠、郭士鳳 2010。利用「空間自相關」探索 1999 年至 2005 年台灣腸病毒病例之空間擴散。華岡地理學報 25: 37 - 52。

Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27(2): 93-115.

Caldas de Castro, M., and B. H. Singer. 2006. Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. *Geographical Analysis* 38(2), 180-208.

Cliff, A., and K. Ord. 1972. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4(3), 267-284.

Geary, R. C. 1954. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* 5 (3): 115-146.

Getis, A., and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24(3): 189-206.

Griffith, D. A. 1987. *Spatial autocorrelation*. Washington DC: Association of American Geographers.

Li, H., and J. F. Reynolds. 1994. A simulation experiment to quantify spatial heterogeneity in categorical maps. *Ecology* 75(8): 2446-2455.

- Moran, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika*. 37(1): 17–23.
- Ord, J. K., and A. Getis. 2001. Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of Regional Science* 41(3): 411-432.
- Sokal, R. R., and B. A. Thomson. 2006. Population structure inferred by local spatial autocorrelation: an example from an Amerindian tribal population. *American Journal of Physical Anthropology* 129(1): 121-131.
- Tobler W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(2): 234-240.
- Wulder, M. A., J. C. White, N. C. Coops, T. Nelson, and B. Boots. 2007. Using local spatial autocorrelation to compare outputs from a forest growth model. *Ecological Modelling* 209(2): 264-276.