

## 運用時序地理加權迴歸建構多時序PM<sub>2.5</sub>推估模型

張家浚<sup>1</sup>、陳嘉惠<sup>2</sup>、張國楨<sup>3</sup>

### 摘要

隨著物聯網科技的發展，環境感測資料之精確性與穩定性已日漸成熟，其中如空氣品質感測器能測得多項空品數值，並建構多時序空品資料集。空氣污染是目前全世界都關注的環境議題之一，其中又以細懸浮微粒（PM<sub>2.5</sub>）對人體的危害有重大威脅而被重視，如何透過各項環境因子去解釋 PM<sub>2.5</sub> 之濃度並達成多時序 PM<sub>2.5</sub> 濃度之預測是本研究欲探討的議題。

有鑑於空間現象是隨時間維度而進行的，以往許多空間統計模型只考量到單一時間點的空間現象，於是本研究試圖使用時空地理加權迴歸(Geographical and Temporal Weighted Regression, GTWR)對多時序的 PM<sub>2.5</sub> 濃度推估進行建模，相較於傳統的地理加權迴歸(Geographically Weighted Regression, GWR)，GTWR 同時將空間與時間納入了權重計算，將過去多時期的資料點考量進迴歸模型的建構，並且透過 GTWR 中提出的 Spatial&Temporal bandwidth 優化方法決定各時期最佳的 bandwidth，用以計算時空權重矩陣。本研究所使用的感測資料之時間解析度為一小時，並將 PM<sub>2.5</sub> 濃度設定為依變數，於自變數中加入溫度、濕度，並選擇台中市(不包含和平區)作為研究區，研究時間區間則限制為 24 小時，由 2021/01/04 08:00AM 持續到 2021/01/05 07:00AM。

研究結果顯示 2021/01/04 17:00PM 至 2021/01/05 07:00AM 為模型適配性最佳時段，於此時段 R 平方平均為 0.7 至 0.85 之間，其中包含交通繁忙時段，根據殘差分布可看出模型於該時段將交通要道周遭的推估值低估，或許是因模型本身只有溫、濕度兩個自變數，未來可將交通、風向與風速資料一併作為模型自變數，期望能透過更多自變數與其在時間維度上的變化，建構出解釋力更佳之時序推估模型。

**關鍵字：**時序地理加權迴歸、多時序空品模式推估、空氣品質

### 壹、緣起

公共衛生及醫學領域研究證實，細懸浮微粒對人體健康會造成不良影響，世界衛生組織旗下的國際癌症研究機構已於 2013 年將空氣污染列為一級環境致癌物，我國亦於 2017 年修正災害防救法令，將懸浮微粒物質認列為法定災害類型之一，空氣污染已然成為國內外備受關注的環境與健康議題，故此，空氣品質數據的需求及重要性便日益彰顯。

<sup>1</sup> 國立臺灣師範大學地理學系，學生

<sup>2</sup> 國立臺灣師範大學地理學系，博士候選人

<sup>3</sup> 國立臺灣師範大學地理學系，教授

## (一)物聯網空品監測技術與發展

隨著近年無線通訊建設日趨完善，加上硬體研發技術提升，物聯網於各領域的應用蔚為盛行，環保署已結合當前新興科技，建構「臺灣空氣品質感測物聯網」，截至 2020 年底，微型感測器布建範圍已擴及全台 6 都與 11 縣市，累計數量高達 9,250 台，為當今全球空氣品質感測網絡密度最高的國家，資料時間解析度可細緻到分鐘等級的即時監測，相較國家標準精密儀器，微型感測器價格低廉，有廣泛布建的優勢，其感測數據不論在時間序列上的污染樣態分析，或是空間上的污染熱區分布研究，都具備了良好的分析應用價值。

## (二)空品推估模式與環境影響因子探討

依據我國《空氣污染防制法》訂定，空氣污染係為空氣中足以直接或間接妨害國民健康或生活環境之物質。其污染來源可分為固定式及移動式，固定式污染多指稱工廠煙囪排放，移動式污染為汽機車等移動載具排放，但不論來源為何，污染物排放至大氣後，會不斷地在其中進行傳輸、擴散、轉化與沉降，運動過程受氣象因素影響甚多，Monks et al. (2009) 指出在所有時空尺度上，氣象條件均會影響大氣污染濃度變化，或於跨距傳輸及交互作用過程中，改變周遭環境的空氣品質，增加污染趨勢分析的困難性 (Sorbian, 2003; Elminir, 2005)。

以細懸浮微粒而言，Tai et al. (2010) 利用 11 年在美國的連續觀測紀錄，建構多元線性回歸模型，探究細懸浮微粒濃度、組成成分與氣象因子間的相關性，並對數據進行去季節化及去趨勢化處理後，著重觀察天氣尺度下氣象因子的影響，結果發現溫度、濕度與降水的日變化，可在模型中達到 50% 的解釋力，表示三者為模型的重要變數。Luo et al. (2017) 使用普通最小二乘法 (OLS) 觀察 PM<sub>2.5</sub> 濃度與潛在地理環境因子之間的關係，結果如下圖顯示，道路、農業、人口、工業、經濟、城市和 PM<sub>2.5</sub> 濃度之間存在顯著的正相關。

Tian & Chen (2010) 利用 MODIS 衛星影像的氣膠光學厚度 (Aerosol Optical Depth, AOD) 數據，與地面氣象觀測值進行資料同化，建立半經驗模型，獲取每小時細懸浮微粒濃度值，該模型均方根誤差為  $6.1\mu\text{g}/\text{m}^3$ ，可解釋地表 65% 的細懸浮微粒濃度變化，並且發現結合地面溫度和相對濕度，對模型的預測效能提升具有重要意義。

## 貳、課題

長期以來空間推估模型一直是地理資訊科學的關注重點，其中回歸方法常被用來探討自變數與依變數之間的線性關係，或用以推論預測依變數，地理加權回歸 (Geographically Weighted Regression, GWR) 分析考量了變數的空間相依性特徵，分別針對預測目標周圍的統計單元建構空間的權重矩陣，將非平穩的空間變化關係整合至回歸模型中，有別於以往全域型的最小二平方法，地理加權回歸分析是一種局部估計方法，可提升整體模型推估的準確性 (Brunsdon et al., 1996; Fotheringham et al., 2002)。

除了空間外，時間同樣也是型塑自然環境過程與社會活動的基本要素，分析帶有時間維度的空間資料，能挖掘到隱含於空間過程的動態變化，例如：疾病傳播、空氣或水污染的擴散，以及都市擴張...等研究 (Crespo et al. 2007; Cressie & Wilkie, 2011)。Crespo et al. (2007)提出時空帶寬的概念，建立時空權重核函數，將時間維度的非平穩(nonstationary)特徵，整合至地理加權回歸模型中，以衰減的時空帶寬作為隨時間變化的局部空間效應。Huang et al. (2010)正式將時間效應納入 GWR 模型，開發了時序地理加權回歸(Geographical and Temporal Weighted Regression, GTWR)方法，並用於建構房價的時空變化模型，藉以解決房地產市場資料中空間、時間的非平穩特性。

## 參、研究方法

### (一)資料集準備

本研究資料集主要以臺中市(不含和平區)多時序空品感測器資料為主，研究區內共有 1145 台感測器，依變數為 PM<sub>2.5</sub> 濃度，自變數則使用溫、濕度，研究區間為 2021/01/04 08:00AM 至 2021/01/05 07:00AM，時間解析度為一小時，並由 2021/01/04 08:00AM 向前擴展 6 小時的資料，用於多時序 OLS(Ordinary Least Squares)模型建構。

### (二)OLS 建構與 Koenker(BP)檢驗

給予時間點  $t$  至  $t-q$  之多時期資料，其中  $q \in \mathbb{Z}^+$ ，對此區間之所有資料進行 OLS 間模，並根據 Koenker(BP)之顯著性診斷此全域模型是否有非穩態情形發生，若存在非穩態情形，則使用 GTWR 進行模式改進。

### (三)時空帶寬優化演算法

本研究使用 GTWR 對多時序空品感測資料進行建模，並透過 Golden section 提出候選空間帶寬，接著再對候選帶寬使用 GTWR Spatial & Temporal bandwidth optimization 演算法進行優化。

給予一組時間點  $t$  之 GTWR 時空帶寬，定義如下：

$$S_t = \{b_{st}, b_{s(t-1)}, \dots, b_{s(t-q)}, b_{Tt}\}$$

每組時空帶寬會有一個時間帶寬  $b_{Tt}$ ，其中  $b_{st}$  代表時間點  $t$  之空間帶寬， $b_{s(t-1)}$  則為時間點  $t-1$  之空間帶寬，以此類推。每組時空帶寬均由 cross-validation(CV)計算並選擇最小 CV score 的那組時空帶寬，優化函數如下：

$$\operatorname{argmin}_{S_t} CV(S_t)$$

### (四) GTWR 模型建構

本研究根據 GTWR 中提出之方法進行演算法實作，並根據其提出之時空權重核函數，計算單一時間資料點之時空權重矩陣，時空權重核函數如下：

$$\omega_{ijS,T}^t = \exp\left(-\frac{d_{Sij}^2}{b_{St}^2}\right) * \exp\left(-\frac{d_{tij}^2}{b_{Tt}^2}\right)$$

$d_{Sij}$  為點之間之歐基里德距離， $b_{St}$  為時間點  $t$  之空間帶寬， $d_{tij}$  則為時間點  $t$  之資料點相較於其他資料點的時間距離， $b_{Tt}$  為時間點  $t$  之時間帶寬。接著根據上述方法產製出之時空權重矩陣  $W$ ，進行局部參數推估，推估式如下：

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y$$

其中  $i$  代表時間點  $t$  的資料點， $(u_i, v_i)$  為其空間座標， $X$  為自變數加上常數所組成之矩陣， $y$  則是由實際值所組成之向量。

## 肆、主要成果

### (一) 多時序 OLS 結果

表一為多時序 OLS 報表，經過 F 檢定後可看出各時間點的模型具有顯著性，Wald 檢定於各時間點也具有顯著性，唯獨 Koenker(BP)於 2021/01/04 07:00AM 至 2021/01/04 16:00AM 有部分呈現不顯著，意味著這部分時間點所構成之全域回歸式不具有顯著非穩態情形發生，本研究後續仍會對這部分時間點進行 GTWR 模式推估，並從中比較模型適配性。

表一、多時序 OLS 報表

date	time	f-value	f-pvalue	r-squared	adj-r-squared	K(BP)	K(BP)-Prob	Wald	Wald-Prob
2021/1/4	08:00	363.45	0	0.09	0.09	1.59	0.66	363.45	0
2021/1/4	09:00	346.34	0	0.08	0.08	0.4	0.94	346.34	0
2021/1/4	10:00	247.43	0	0.05	0.05	0.55	0.91	247.43	0
2021/1/4	11:00	135.68	0	0.03	0.03	2.03	0.57	135.68	0
2021/1/4	12:00	114.96	0	0.02	0.02	2.99	0.39	114.96	0
2021/1/4	13:00	134.1	0	0.02	0.02	1.69	0.64	134.1	0
2021/1/4	14:00	176.11	0	0.03	0.03	2	0.57	176.11	0
2021/1/4	15:00	232.07	0	0.03	0.03	2.74	0.43	232.07	0
2021/1/4	16:00	297.28	0	0.04	0.04	5.24	0.15	297.28	0
2021/1/4	17:00	284.47	0	0.04	0.04	30.04	0	284.47	0
2021/1/4	18:00	305.28	0	0.04	0.04	64.91	0	305.28	0
2021/1/4	19:00	376.76	0	0.04	0.04	163.28	0	376.76	0
2021/1/4	20:00	485.53	0	0.05	0.05	250.13	0	485.53	0
2021/1/4	21:00	550.19	0	0.06	0.06	266.8	0	550.19	0
2021/1/4	22:00	565.46	0	0.06	0.06	284.46	0	565.46	0
2021/1/4	23:00	466.68	0	0.05	0.05	378.76	0	466.68	0
2021/1/5	00:00	389.47	0	0.04	0.04	407.33	0	389.47	0
2021/1/5	01:00	325.49	0	0.03	0.03	489.56	0	325.49	0
2021/1/5	02:00	268.43	0	0.02	0.02	569.86	0	268.43	0
2021/1/5	03:00	232.63	0	0.02	0.02	592.83	0	232.63	0
2021/1/5	04:00	365.3	0	0.03	0.03	519.81	0	365.3	0
2021/1/5	05:00	526.56	0	0.04	0.04	505.59	0	526.56	0
2021/1/5	06:00	675.3	0	0.05	0.05	532	0	675.3	0
2021/1/5	07:00	794.3	0	0.06	0.06	566.19	0	794.3	0

## (二) GTWR 推估結果

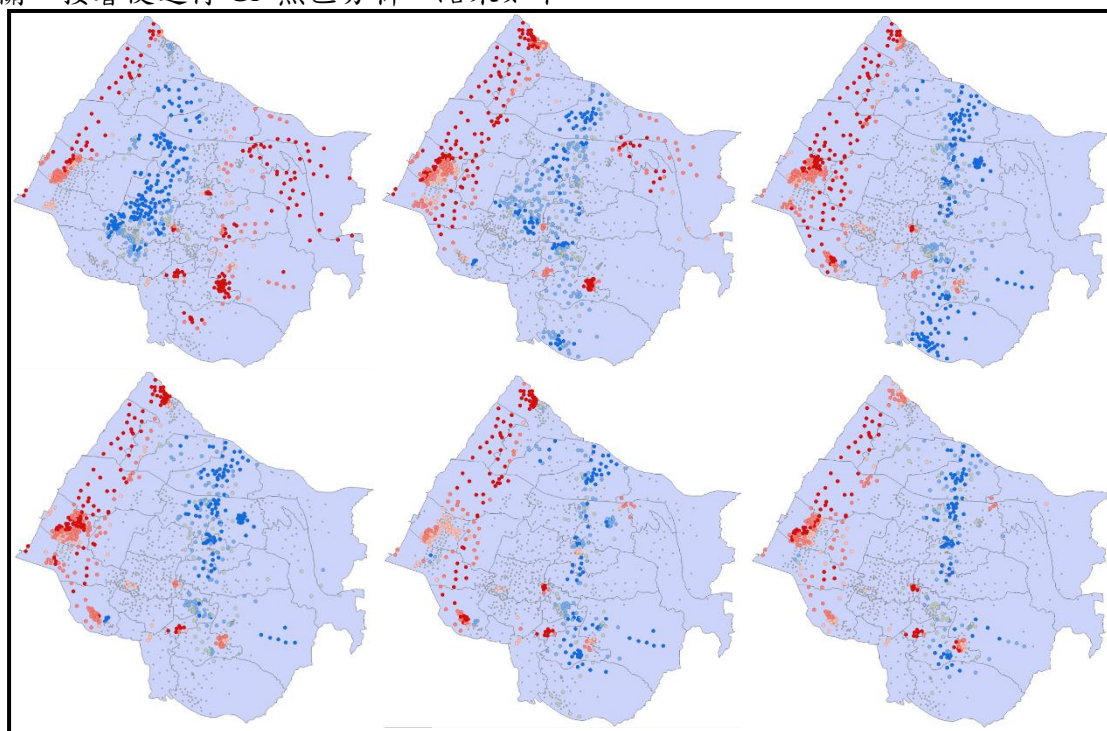
表二為 GTWR 推估結果，主要以 R-squared 作為模型適配性評判指標，可以發現於前部分多時序 OLS 結果中的 Koenker(BP)不顯著時段，同時間段的 R-squared 同樣不是很高，而其餘時段的 R-squared 除少部分擾動外，大部分時段的模型適配性是較可被接受的，均維持在 0.7 至 0.85 之間，本研究接著對殘差的空間分布進行分析。

表二、GTWR 推估結果

date	time	R-squared	date	time	R-squared
2021/1/4	08:00	0.3394724	2021/1/4	20:00	0.80178
2021/1/4	09:00	0.3714363	2021/1/4	21:00	0.748718
2021/1/4	10:00	0.1142509	2021/1/4	22:00	0.801321
2021/1/4	11:00	0.2655785	2021/1/4	23:00	0.749143
2021/1/4	12:00	0.3442073	2021/1/5	00:00	0.696977
2021/1/4	13:00	0.4112827	2021/1/5	01:00	0.742445
2021/1/4	14:00	0.5207502	2021/1/5	02:00	0.750484
2021/1/4	15:00	0.6414667	2021/1/5	03:00	0.672761
2021/1/4	16:00	0.6492375	2021/1/5	04:00	0.72471
2021/1/4	17:00	0.7423688	2021/1/5	05:00	0.657619
2021/1/4	18:00	0.7450555	2021/1/5	06:00	0.711311
2021/1/4	19:00	0.8129325	2021/1/5	07:00	0.840191

## (三) GTWR 殘差分析

礙於篇幅限制，本研究選擇了 2021/01/04 15:00PM 至 2021/01/04 20:00PM 等 6 小時區間進行殘差分析，此段時間為交通繁忙時刻，並且模型具有良好適配性。本研究對這六小時區間進行了全域空間自相關，結果均顯示殘差具有顯著正空間自相關，接著便進行 Gi\*熱區分析，結果如下：



可看出冷區不斷向東方移動，最終停留於此並漸漸淡出，初步推斷因此區為交通熱區，於該時段有頻繁極端事件發生，導致 PM<sub>2.5</sub> 濃度上升，而模型低估了此區的 PM<sub>2.5</sub> 濃度，意味著模型缺乏其他環境變數來解釋這部分的極端事件。

## (四) 結論與未來建議

本研究使用 GTWR 進行多時序 PM<sub>2.5</sub> 濃度建模，將時間維度納入迴歸式估計，藉此探討時間維度於空間現象的重要性，並將結論分為兩大方向，分別朝向演算法與實際應用面進行改進。



演算法方面有 1.非穩態問題：過去透過 Koenker(BP)指標能進行非穩態診斷，並且因以往 OLS 與 GWR 應用都是以一個時間點的空間資料為主，因此能透過此指標來推測非穩態情形於空間中聚集，但本研究所使用之 OLS 含括時間維度，因此無法透過單一指標來定位非穩態情形發生於空間或時間維度上；2.核函數適用性：原始 GTWR 之核函數被用於倫敦房價推估，時間單位為年，將其直接應用於感測器資料上實屬強人所難，並且根據研究所選之時間單位不同，會產生可調整時序單元問題(Modifiable Temporal Unit Problem, MTUP)。

實際應用面則有 1.自變數數量增加：本研究將溫、濕度作為自變數，對 PM<sub>2.5</sub> 濃度進行推估，未來可加入更多自變數以增加模型解釋力。2.核函數與時序單元選擇：如何正確的從時間維度上描述空品擴散，並選擇有效的時序單元，這幾個關鍵性的因素為未來研究的重要目標。

## 伍、參考文獻

1. Brunson, C., Fotheringham, A. S., & Charlton, M. E. (1996). "Geographically weighted regression: a method for exploring spatial nonstationarity". *Geographical analysis*, 28(4), 281-298.
2. Crespo, R., Fotheringham, S., & Charlton, M. (2007). Application of geographically weighted regression to a 19-year set of house price data in London to calibrate local hedonic price models. In *Proceedings of the 9th International Conference on Geocomputation*. National University of Ireland Maynooth.
3. Cressie, N. A., & Wikle, C. (2011). *Statistics for Spatio-Temporal Data*.
4. Elminir, H. K. (2005). Dependence of urban air pollutants on meteorology. *Science of the Total Environment*, 350(1-3), 225-237.
5. Fotheringham, Stewart A., Chris Brunson, and Martin Charlton. *Geographically Weighted Regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2002.
6. Huang, Bo, Bo Wu, and Michael Barry. "Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices." *International Journal of Geographical Information Science* 24.3 (2010): 383-401.
7. Luo, J., Du, P., Samat, A., Xia, J., Che, M., & Xue, Z. (2017). Spatiotemporal pattern of PM 2.5 concentrations in mainland China and analysis of its influencing factors using geographically weighted regression. *Scientific reports*, 7(1), 1-14.
8. Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., Akimoto, H., ... & Blake, N. (2009). Atmospheric composition change—global and regional air quality. *Atmospheric environment*, 43(33), 5268-5350.
9. Sorbjan, Z. (2003). *Air pollution meteorology. AIR QUALITY MODELING Theories, Methodologies, Computational Techniques and Available Databases and Software*, 1.
10. Tai, A. P., Mickley, L. J., & Jacob, D. J. (2010). Correlations between fine particulate matter (PM<sub>2.5</sub>) and meteorological variables in the United States: Implications for the sensitivity of PM<sub>2.5</sub> to climate change. *Atmospheric environment*, 44(32), 3976-3984.
11. Tian, J., & Chen, D. (2010). A semi-empirical model for predicting hourly ground-level fine particulate matter (PM<sub>2.5</sub>) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sensing of Environment*, 114(2), 221-229.