

# 2021 年臺灣國際科學展覽會 優勝作品專輯

作品編號 190021

參展科別 電腦科學與資訊工程

作品名稱 臺南市違停的離散小波與 ARIMA 分析

得獎獎項 大會獎 三等獎

耶魯科學與工程學會獎

就讀學校 國立臺南女子高級中學

指導教師 洪士薰、謝淑莉

作者姓名 施珽希、林亮萱、楊雁筑

關鍵詞 MODWT、ARIMA、違停

## 作者簡介



我是林亮萱(右)，目前就讀台南女中高二，喜歡觀星、看電影還有喝水。這次的研究過程中，認識了一點跟電腦對話的方法，但學到更多的是團隊合作及一份研究的產出是多麼的困難，期待我們的研究能落實在生活中！

我是施玗希(左)，就讀台南女中二年級，喜歡旅遊及閱讀。我於高一曾參加過一些資訊相關自學課程及講座，所以對資訊科技結合生活的應用相當有興趣。這次選定的題目是台南市所有用路人最頭痛的違停問題，希望能藉著這一次的研究，對自己家鄉的交通問題帶來些幫助。

我叫楊雁筑(中)，是台南女中的學生，喜歡跟天文社上山看星星，或是賞月，喜歡天文故事或是認識星座位置，但是討厭天文物理。由於爸爸常因為違停被罰錢，所以萌生對違停的興趣，個人藉由這次機會跟電腦變熟，但是還只是泛泛之交的程度，希望未來能更加進步，也希望我們的研究真的帶給社會一些小小的改變。

## 中 文 摘 要

現今逐漸有許多科學家研究出一些能夠進行預測事件的方法，諸如天氣預報、價格變化的幅度等等。在眾多生活常見的問題中，我們選定違規停車為例，作為研究主題，而我們推測違規停車是一種具有週期性的問題，有多樣的潛在變因存在，於是我們跳脫以往多以心理層面、法規制度與地理熱點等的視角去看待違規停車，以資訊科學及數學的方法，建立一套分析的模式加以進行研究。

本文以連續小波變換 **CWT**、最大重複離散小波轉換 **MODWT**、差分整合移動平均自迴歸模型 **ARIMA**、混合 **MODWT+ARIMA** 模型及混合 **Harr** 離散小波變換 **DWT+ARIMA** 模型，對臺南市三個不同特性的行政里的共五個聚類點，進行預測。原始的資料以週為單位，分成全日 (**allDay, 0~24** 時) 及白天 (**Day, 08~16** 時) 兩種型態，並以連續 **46** 週的資料預測接下來 **4** 週。結果採用方均根誤差 **RMSE** 相對平均值、標準差、最大值的比值，進行預測品質的比較。

## 英 文 摘 要

The research on parking violations has always been an important issue in Tainan. However, it is difficult to effectively control the variables due to the complexity and the various factors involved. Therefore, researches in the past are generally based on the psychological cause of "violation", geographical location of business districts, and the formulation of laws and regulations. However, in our research, we want to predict the amount of parking violation using mathematics and information sciences .

This article aims to establish analytical models by using Continuous Wavelet Transform (CWT), Harr Discrete Wavelet Transform (DWT), Maximal Overlap Discrete Wavelet Transform (MODWT), Autoregressive Integrated Moving Average Model (ARIMA) and the combination of ARIMA and wavelet transform to predict a total of five clustering points in three administrative districts with three different characteristics in Tainan City.

We analyzed the data by two types: all-day-mode (00-24 o'clock) and day-mode ( 08-16 o'clock) by week, and took 46 consecutive weeks to train our model and the following 4 weeks to predict. Finally, we used the ratio of the root-mean-square error (RMSE) to the average value, standard deviation, and maximum value respectively to measure the quality of each method.

Comparing the performances of these methods, we found that the combination of MODWT and ARIMA actually improves the prediction of ARIMA in general. As for stability, the combination of MODWT and ARIMA acts better than mere ARIMA and combination of DWT and

ARIMA, which acts case by case. We want to prove that our research can be used to predict data with potential variables, and it can be widely applied to situations such as short-term electricity prices, wind power, peak season for specific diseases, stocks, and commodity prices.

## 壹、前言動機及簡介

違規停車是目前台灣尚未解決的一項問題，108年台南市的違停數量高達15萬，而違停不僅造成用路人不便，也耗費龐大的警力資源，其中，根據不同的地區及時段違規停車的數量也不同，於是本研究從時間、地點等方面切入研究台南市不同地區各時段違停彼此間的關係，而我們採用的方式依序為(1)以K-means將不同里的違停依照地點做分類(2)以時間序列分析不同時段違停的關係(3)以小波分析預測違停趨勢。

本文討論及研究了連續小波變換CWT、Harr離散小波變換DWT、最大重複離散小波轉換MODWT (Maximal Overlap Discrete Wavelet Transform)、差分整合移動平均自迴歸模型ARIMA以及混合MODWT+ARIMA模型。對臺南市三個不同特性的行政里的共五個聚類點，進行預測。以週為單位，分成全日(allDay)及白天(Day, 08~16時)兩種型態交叉比較分析，並以連續46週的資料預測接下來4週。結果採用方均根誤差RMSE相對平均值、標準差、最大值的比值，做預測品質的比較。

## 貳、文獻討論

關於違規停車的研究，一直是重要的問題但因為牽涉層面太多太廣，實難以有效控制變因的方式進行，所以一般多從「違規行為」的心理成因或者商圈地理位置，法規的制定著手。但因為難以有效控制變因來進行研究。

[1] 摘要：國內外對於「違規行為」和「停車管理」之研究甚多，且均提出許多具體的結論或建議。然而對於兩者之關聯卻著墨甚少，也就是有關「違規停車」之研究實際上相當有限，鮮少有人探討違規停車之成因及其間之因果關係。因此本研究乃從「外在的環境構面」與「駕駛人內在的心理構面」兩方面著手探討駕駛人違規停車之成因，分析駕駛人對違規停車之態度與行為，期能藉此提出改善之建議，供政府相關單位及決策者參考使用。

### [2] 摘要

都市因機動車輛數量多且受限於停車空間不足，常使違規停車成為大都市的交通安全議題…。近年來政府機關為了改善違規停車的問題，除規劃增設停車空間、鼓勵善用大眾運輸

外，並強化執法取締及拖吊，將併排停車違規項目提高罰鍰至 2400 元。依據內政部警政署資料統計顯示，全國取締違規停車件數仍逐年增加。…鑑此，臺南市政府警察局自 104 年 4 月 9 日起展開「違規停車零容忍」專案執法，並在 105 年 4 月 15 日起實施委外拖吊四區執法政策。本研究考量在原有交通違規執法強度未予改變，僅強化拖吊執法能量，探討臺南市在執行違規停車取締政策從公營拖吊改以委託民營拖吊政策，對於違規行為發生的時段、道路型態及態樣等執法成效影響做事前事後分析，提供執法改善空間之建議。本研究根據臺南市交通執法管理系統之資料，分析 104 年至 106 年四個行政區(臺南市東區、仁德區、中西區、北區)取締違規停車之成效，以敘述性統計分析方法、交叉分析、卡方檢定，探討違規停車取締政策與違規停車行為關聯性，以研擬更適合的違規停車取締方式，減少違規停車行為的發生，以提供政府機關後續參考。

商圈地理環境，經濟活動周期，法規制度，「違規行為」的心理都是一、二年內不會有重大改變，所以若假設人民經濟活動的一年周期，本文從時間序列的觀點進行分析，依[3]的想法，以離散小波與 ARIMA 來分析違規停車數量，並預測，良好的預測模型有助於了解違規停車數量的模式，進一步可成為政策及法規制度規劃之基礎。

### 參、研究方法及過程：

原始違規停車的資料為 2016 年 8 月至 2017 年 8 月，資料欄位包含有違規停車時間、車種代碼及名稱、違規停車地點緯度、經度、行政里。

表 1. 原始違規停車示例

B	C	D	E	F	L	M	O
違規時間	簡車種代碼	車種名稱	條款1	緯度	經度	行政里	
2017-02-28 11:40:00	112	自用小客車	5500101	23.019934	120.135128	四草里	
2016-09-01 08:13:00	H	普通重型	5500105	23.003848	120.218783	中樓里	
2016-09-01 08:27:00	H	普通重型	5500105	23.005415	120.218438	中樓里	

初步統計 2016 年 8 月至 2017 年 8 月，各行政里的違規停車數量如下表以 A、B、C…里代稱：

表 2. 各行政里的違規停車數量示例

代碼	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	2365	5546	3534	4566	5920	1694	6187	5377	2273	1868	2146	5059	3117	1601	1720	7186	6546	9398	1967	4031	2798	2950	3219	2540	2257	1741

有些行政里數量太小，所以我們選定一年時間內至少有 1500 件違規停車的里進行分析。因此由 323 個有資料的行政里中過濾剩 26 個行政里。

#### 一、違規停車聚點的關聯性分析

網路上購買商品的點擊率行為有些差異，例如某人在購買某一商品 A 後也會同時購買

另一商品 **B**，我們可以借由條件機率的想法去計算其關聯性，雖然違規停車的行為與此有些不同，但在以區域的停車需求及經濟活動與空間的配置來看在，既成的環境與需求，會造成 **A** 地與 **B** 地之間，違規停車的行為的關聯性。因此我們首先，以每日同一小時為單位，進行違規停車資料的關聯性規則之計算。

關聯規則學習(**Association rule**)可以在大量資料庫中發現變數之間的關係。例如由超市的 **POS** 系統記錄的大批交易資料中產品之間的規律性。資料中發現{小蛋糕, 甜點}→{牛奶} 會表明如果顧客一起買小蛋糕, 甜點，他們也很有可能買牛奶的規則。但反過來則不一定。

假設  $S = \{A, B, \dots, Z\}$  是每一里的集合。每一天從 8:00~21:00 以兩小時為單位的違規停車資料在每一里的違規停車數量統計，其中若 **A** 里的違規停車資料，其集合記為  $\mathcal{A}$ ，另有一 **B** 里的違規停車資料，其集合記為  $\mathcal{B}$ ，則定義關聯規則  $\mathcal{A} \Rightarrow \mathcal{B}$  的支持度 (support) 是  $\mathcal{A} \cup \mathcal{B}$  在所有違規停車資料中，所占的比例，或者機率；信度 (confidence) 則是  $\mathcal{A}$  中同時也在  $\mathcal{B}$  中的違規停車資料即條件機率  $P(\mathcal{B}|\mathcal{A}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{A})}$ 。我們可以依資料的情況設定如果同時滿足的最小支持度閾值和最小信度閾值，以建立關聯規則。

表 3. 每一天兩小時為單位的違規停車資料

Month	Weekday	HOUR	up	A里	B里	C里	D里	E里	郡王里	F里	G里	H里	I里	J里	K里
8	2	16	1	22	0	0	0	10	0	0	2	0	0	0	1
8	2	18	2	0	2	0	1	0	0	0	4	1	0	15	0
8	2	20	0	11	1	0	12	0	0	5	0	0	0	1	0
8	2	24	0	1	0	1	0	1	0	1	0	0	0	0	0
8	3	10	7	8	0	2	5	0	1	0	0	0	0	0	0
8	3	12	4	0	2	1	6	0	0	0	0	0	0	0	6
8	3	14	1	8	1	4	5	1	1	0	1	4	0	0	0
8	3	16	3	18	2	4	13	0	2	2	0	0	0	0	1
8	3	18	1	0	0	1	0	0	0	0	0	0	0	10	1
8	3	20	1	0	0	0	8	0	0	3	2	2	0	0	1
8	3	24	0	0	7	4	6	0	0	3	0	1	0	0	1
8	4	10	2	0	0	1	2	1	0	3	0	0	4	0	0

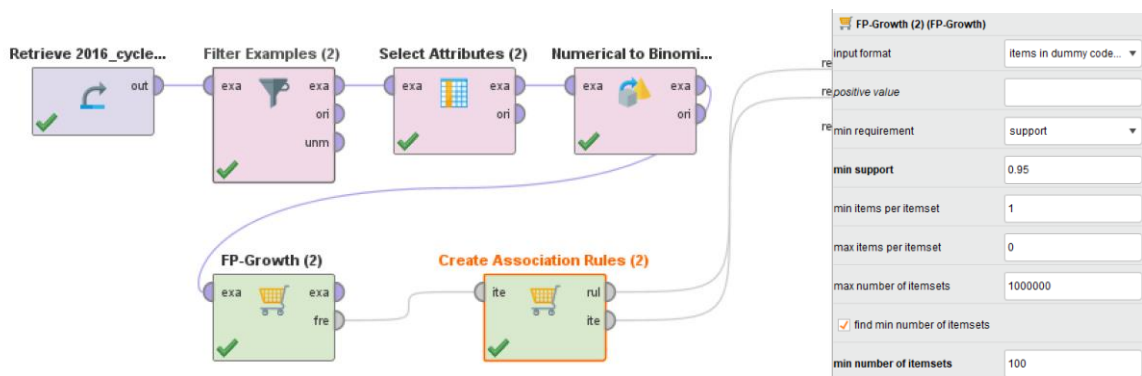


圖 1. RapidMinerStudio9.8 版本的計算流程及參數

在以 RapidMinerStudio9.8 版本的計算結果如下：

(一)同一時段像 $G, L, K, W, \dots \Rightarrow R$ 都有很大的信度，而 $R, L, X, W, \dots \Rightarrow G$ 也許是因為  $R$ 、 $G$ 所占的違規停車資料比例高，但是像 $D \Rightarrow L, W \Rightarrow L$ 的相對較低的支持度，但仍有很高的信度，這應該與區域的停車需求及經濟活動與空間的配置等潛在的環境與需求有關。

表 4. RapidMinerStudio9.8 計算的支持度及信度

No.	Premises	Conclusion	Support	Confiden... ↓
409	D里, M里	L里	0.782	0.974
410	L里, Q里	R里, G里	0.782	0.974
403	R里, L里, B里	G里	0.777	0.974
404	R里, D里, E里	G里	0.777	0.974
405	R里, D里, E里	L里	0.777	0.974
401	W里, M里	R里, L里	0.773	0.974

## 二、K-means 聚點分類取代地理熱點

傳統上，帶有地理位置的資料多可以在 TGOS 地理資訊系統平台上標記，所以違規停車的資料也可以如此處理，但因為使用免費的 TGOS 平台上傳資料的運算時間很長，所以我們試著將資料以 1000 筆分批上傳，得到如下的熱點圖。

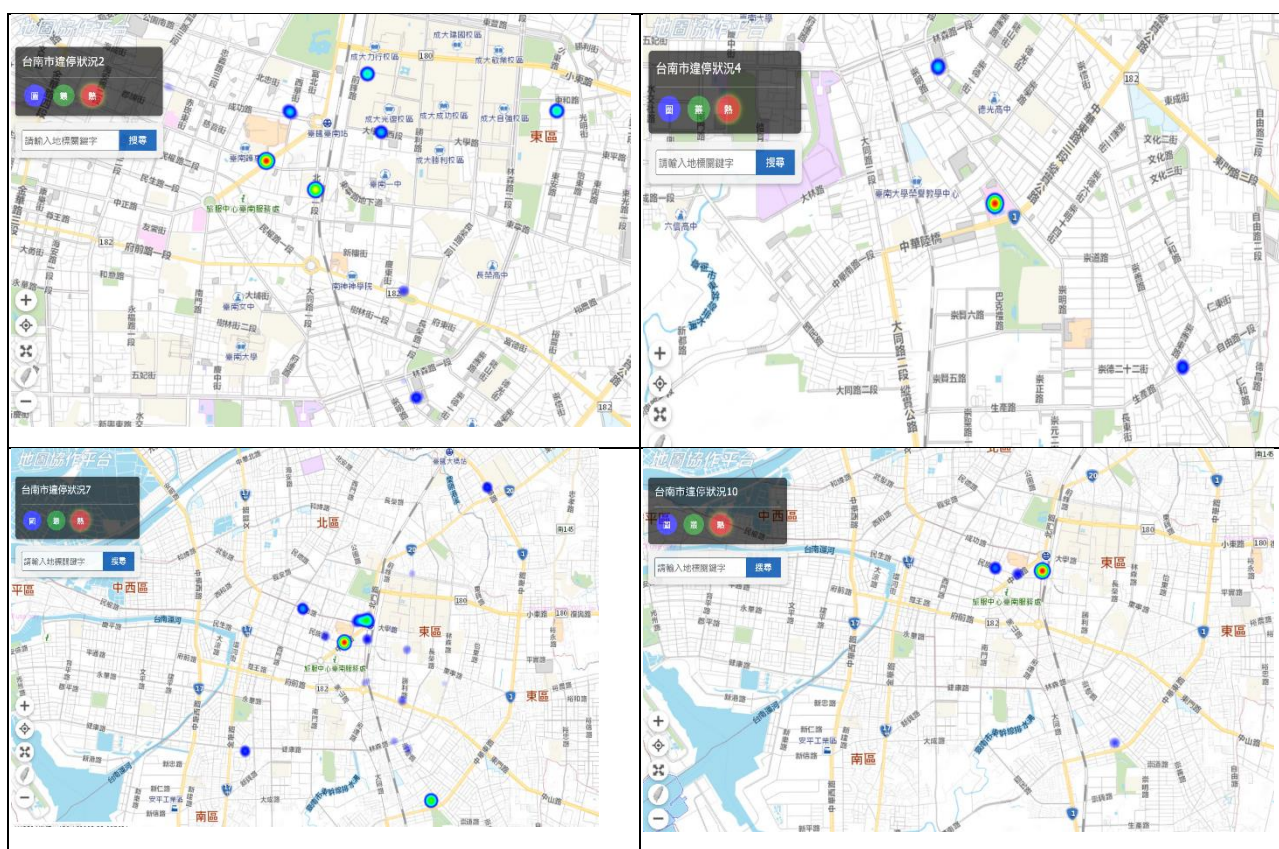


圖 2. TGOS 平台違規停車熱點圖示例

但傳統的熱點分析，不容易看出統計的關聯性，所以我們引入非監督式的機器學習方法 **K-means**，來進行運算。針對每一里的違規停車資料，以經緯度的坐標距離，進行 **K-means** 分類。

**K-means** 是一種聚類 (**Cluster**) 的方式，原始每一筆資料的經緯度分別為  $x, y$  坐標，記為  $x_j$ ，一開始選定  $k$  個初始值資料  $\mu_1, \mu_2, \dots, \mu_k$ ，計算  $\|x_j - \mu_i\|^2$ ， $i=1, 2, \dots, k$  若  $\|x_j - \mu_i\|^2$  最小，則  $x_j \in S_i$ ，如此原始每一筆資料可分成  $S_1, S_2, \dots, S_k$  類，再計算

$$J = \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

的值，對初始值資料  $\mu_1, \mu_2, \dots, \mu_k$  分別做變動並進行疊代求  $J$  的最小值，此時所得的  $\mu_1^*, \mu_2^*, \dots, \mu_k^*$ ，即是  $k$  個聚類中心。再將原來資料依與聚類中心的距離歸類，即可將資料分類。

平均中心距離 =  $\sqrt{\frac{1}{n} \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i^*\|^2}$ 。

當然  $k$  越大，**K-means** 的平均中心距離當然會越小，但當  $k$  達一個值後，平均中心距離下降的情形會趨緩。以行政里 **G** 里為例方案的流程為

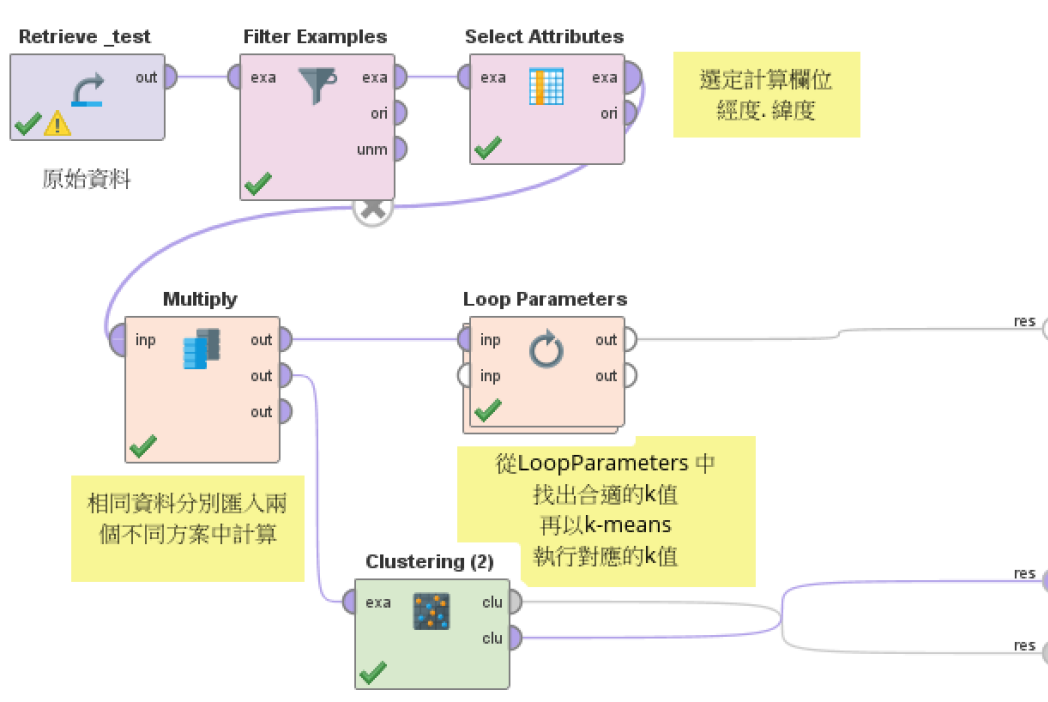


圖 3. 以 RapidMiner 進行的 K-means 流程

在 **Loop Parameters** 中，考慮  $k$  值由 2 至 9 進行計算，並輸出  $k-d$  圖，其中的  $d$  為對應 **K-means** 的平均中心距離。**Loop Parameters** 內的流程如下圖：



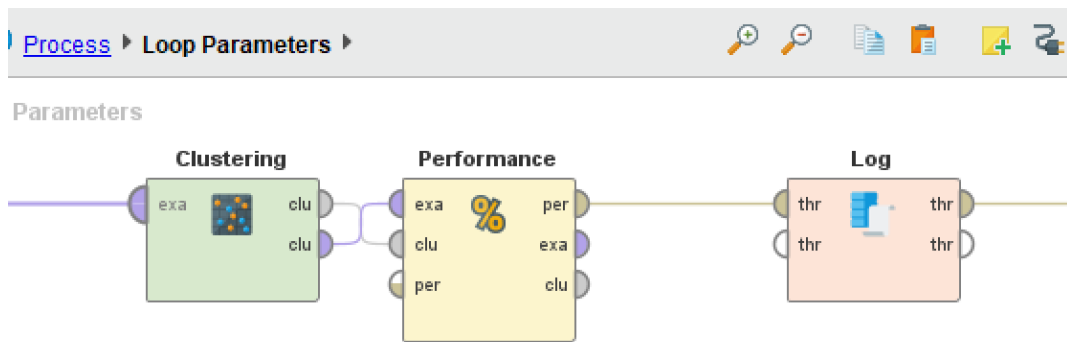


圖 4. Loop Parameters 內的流程

### 三、ARIMA

ARIMA 模型 (Autoregressive Integrated Moving Average model) , 差分整合移動平均自迴歸模型, 又稱整合移動平均自回歸模型, 時間序列預測分析方法之一。

ARIMA (p, d, q) 中, AR 是"自回歸", p 為自回歸項數; MA 為"移動平均", q 為滑動平均項數, d 為使之成為平穩序列所做的差分階數。

ARIMA (p, d, q) 模型是 ARMA (p, q) 模型的擴展。ARIMA (p, d, q) 模型可以表示為：

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

其中 L 是遲滯算子。即  $L^k x_t = x_{t-k}$

或者也可以整理成下列形式

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

$\mu$  是  $X_t$  平均數,  $\phi_i$  是自回歸參數,  $\theta_i$  是移動平均參數,  $\varepsilon_t$  是誤差量。從文獻[4]一般至少要有 50 個歷史資料才能有好的預測。

### 四. 連續小波轉換 CWT

違規停車雖有各式各樣的成因, 法規、心理、經濟區域等。但其統計數字也呈現了類似週期的行為, 所以像風力場的風力、金融市場的債券交易價格、歐洲各地區每週的豬隻平均交易價格, 等等, 都有以小波進行的相關研究。

把資料  $X_t$ , 想像成函數  $f(t), 0 \leq t \leq L$  的取樣, 連續小波的想法是  $f(t)$  表示成一族正交週期函數的組合:  $\langle \phi_k(t) \rangle_{k \in N}$ ,  $f(t) = \sum_{k=1}^{\infty} \alpha_k \phi_k(t)$ , 因為  $\int_0^L f(t) \phi_k(t) dt = \alpha_k$ 。

由  $X_1, X_2, X_3, \dots, L$  求係數  $\alpha_1, \alpha_2, \dots$ , 再由  $\alpha_1, \alpha_2, \dots$  反向求預測值。

對應 Fourier Transform

$$f(t) \mapsto F(\xi), F(\xi) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i \xi t} dt$$

與反向 Fourier Transform 的過程

$$F(\xi) \mapsto f(t), \quad = f(t) \int_{-\infty}^{\infty} F(\xi) e^{2\pi i \xi t} dt$$

本文中 **CWT** 選擇以  $\phi_k(t)$ ,  $k = 1, 2, 3, \dots$  為基底函數

$\phi_k(t) = \frac{\sqrt{k\alpha L}}{2\sqrt{\pi}} \sin(\alpha kt)$ ,  $L$  是  $\frac{2\pi}{\alpha}$  的倍數, 對於原始數據  $X_0, X_1, X_2, \dots, X_L$ , 我們假設

$f(t) = X_t$ , 且是一折線函數。因此

$$f(t) = \frac{(t-k+1)X_k + (k-t)X_{k-1}}{1} = a_k t + b_k, \quad k-1 \leq t \leq k$$

因為  $\int_0^L (\sin(\alpha kt))^2 dt = \frac{L}{2}$ , 及

$$\int_0^L f(t) \sin(\alpha nt) dt = \int_0^L \sum_{k=1}^{\infty} a_k \sin(\alpha kt) \sin(\alpha nt) dt = \alpha_k \int_0^L (\sin(\alpha nt))^2 dt = \frac{L}{2} \alpha_n$$

另一方面

$$\sum_{k=1}^L \int_{k-1}^k (a_k t + b_k) \sin(\alpha nt) dt = \sum_{k=1}^L \left( \int_{k-1}^k a_k t \sin(\alpha nt) dt + \int_{k-1}^k b_k \sin(\alpha nt) dt \right)$$

$$= \sum_{k=1}^L \left( -\frac{a_k}{n} (k \cos n\alpha k) + \frac{a_k}{n} ((k-1) \cos n\alpha(k-1)) + \frac{a_k}{n^2} (\sin n\alpha k - \sin n\alpha(k-1)) \right)$$

$$+ \sum_{k=1}^L \left( -\frac{b_k}{n^2} (\cos n\alpha k - \cos n\alpha(k-1)) \right)$$

因此,

$$\alpha_n = \frac{2}{L} \sum_{k=1}^L \left( -\frac{a_k}{n} (k \cos n\alpha k) + \frac{a_k}{n} ((k-1) \cos n\alpha(k-1)) + \frac{a_k}{n^2} (\sin n\alpha k - \sin n\alpha(k-1)) \right) + \frac{2}{L} \sum_{k=1}^L \left( -\frac{b_k}{n^2} (\cos n\alpha k - \cos n\alpha(k-1)) \right)$$

我們利用 **Excel** 即可進行順向及反向的變換。

以 **A** 里之資料選定  $\alpha = 10, L = 56$ ,  $f(t) \sim \sum_{k=1}^{15} a_k \phi_k(t)$  為例。

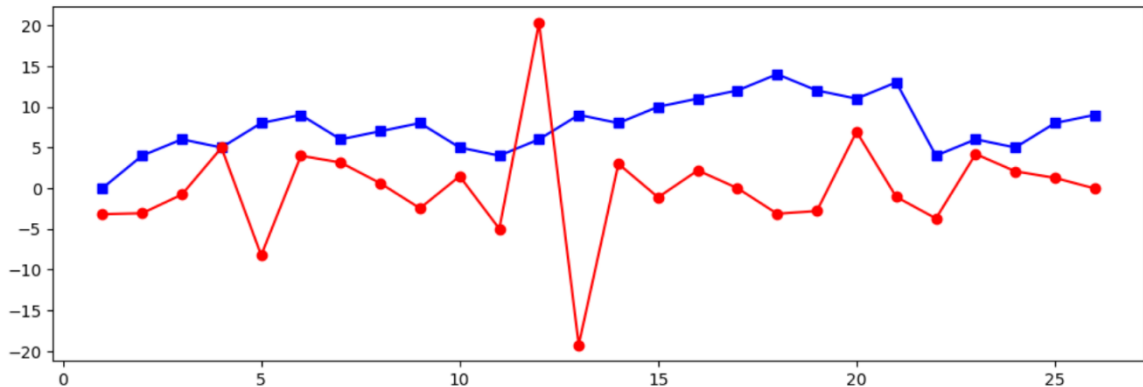


圖 5. 藍色為實際違停數, 紅色為 **CWT** 計算結果

## 五、MODWT(Maximal Overlap Discrete Wavelet Transform)

### (一) Harr 離散小波(DWT)

選擇不連續的週期函數為基底函數，進行類似於 **Fourier Transform** 與反向

**Fourier Transform**，進行演算。首先考慮 **Harr** 函數

$$\psi(x) = \begin{cases} 1 & 0 \leq x \leq 1/2 \\ -1 & 1/2 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

我們以  $\psi(x)$  為小波演算的母函數 (mother wavelet function)， $\phi(x)$  為趨勢函數 (scaling function)，

$$\begin{aligned} \psi(x) &= \phi(2x) - \phi(2x - 1) \\ \phi(x) &= \phi(2x) + \phi(2x - 1) \end{aligned}$$

定義

$$\psi_{n,k}(x) = \sqrt{2}^n \psi(2^n x - k), \quad \phi_{n,k}(x) = \sqrt{2}^n \phi(2^n x - k)$$

$$\int_{-\infty}^{\infty} \psi_{n,k}(x) dx = 0, \quad \int_{-\infty}^{\infty} \psi_{n,k}(x)^2 dx = 1, \quad \int_{-\infty}^{\infty} \psi_{n,k}(x) \psi_{m,l}(x) dx \text{ 僅當 } n = m, \text{ 且 } k = l \text{ 為 } 1, \text{ 否則為 } 0。$$

因此  $\psi_{n,k}(x)$ ，及  $\phi_{n,k}(x)$  形成一組小波演算的基底函數。

資料  $x_1, x_2, x_3, \dots, x_L$ ，視為  $x(t)$  在時間  $t = 1, 2, 3, \dots, L$  的取值

$$\begin{aligned} \alpha_{n,k} &= \int_{-\infty}^{\infty} x(t) \phi_{n,k}(t) dt \\ \beta_{n,k} &= \int_{-\infty}^{\infty} x(t) \psi_{n,k}(t) dt \end{aligned}$$

由  $\phi(x), \psi(x)$  的性質，有

$$\begin{aligned} \alpha_{n,k} &= \frac{1}{\sqrt{2}} (\alpha_{n+1,2k} + \alpha_{n+1,2k+1}) \\ \beta_{n,k} &= \frac{1}{\sqrt{2}} (\alpha_{n+1,2k} - \alpha_{n+1,2k+1}) \end{aligned}$$

若在時間  $k - 1 \leq t \leq k$ ， $x(t) = x_k$  並進行  $\alpha_{n,k}, \beta_{n,k}$  演算

$$\frac{1}{\sqrt{2}} [1 \ 1 \ 1 \ -1] [x_1 \ x_2] = [s_0 \ d_0], \text{ 依此類推得到}$$

$$(x_0, x_2, x_3, \dots, x_{2n+1}) \mapsto ((s_0, d_0), (s_1, d_1), \dots, (s_n, d_n))$$

$(d_0, d_1, \dots, d_n)$  為第一層高通濾波(D1 係數)。 $(s_0, s_1, \dots, s_n)$  為第一層低通濾波(S1 係數)。

再將此方法套用到  $(s_0, s_1, \dots, s_n)$  即可得到第二層的高通濾波(D2)及低通濾波(S2)，依此可接連不斷地操作下去。

我們以底下資料來示範

表 5. DWT 高通及低通濾波演算

time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
no of illStop	4	6	5	8	9	6	7	8	5	4	6	9	8	10	11	12	14	12	11
D1	1.41		2.12		-2.12		0.71		-0.71		2.12		1.41		0.71		-1.41		1.41
S1		7.07		9.19		10.6		10.6		6.36		10.6		12.7		16.3		18.4	
D2		1.5				0				3				2.5					-1
S2				11.5				15				12				20.5			
D3				2.47								6.01							
S3								18.7								23			
D4								3											
S4																29.5			

若原始資料為  $a_1, a_2, a_3, \dots$

S1 數列為  $s1(2) = \frac{1}{\sqrt{2}}(a_1 + a_2)$  ,  $s1(4) = \frac{1}{\sqrt{2}}(a_3 + a_4)$ , ...

D1 數列為  $D1(1) = \frac{1}{\sqrt{2}}(-a_1 + a_2)$  ,  $D1(3) = \frac{1}{\sqrt{2}}(-a_3 + a_4)$ , ...

S2 數列為  $s2(4) = \frac{1}{\sqrt{2}}(s1(2) + s1(4))$  ,  $s2(8) = \frac{1}{\sqrt{2}}(s1(6) + s1(8))$ , ...

D2 數列為  $D2(2) = \frac{1}{\sqrt{2}}(-s1(2) + s1(4))$  ,  $D2(6) = \frac{1}{\sqrt{2}}(-s1(6) + s1(8))$ , ...

依此方式

$$x_t \rightarrow \{S1 \rightarrow \{S2 \rightarrow \dots D2 \quad D1$$

可將資料分類成一系列的高通及低通濾波，高通符合數據的高頻特徵，低通濾波則符合數據的趨勢特徵。

(二) MODWT 最大重複離散小波轉換

[3]CHAPTER3 p.45 中說：

The MODWT addresses some of the issues with the DWT but is a non-orthogonal redundant transform [48]. The MODWT is a non-decimated transform meaning that the number of coefficients produced at each level has the same cardinality as the original time series, providing for more robust ARIMA modelling. The DWT coefficients contain enough information to reconstruct a series, while the MODWT contains this information but it is repeated as each neighbouring coefficient is constructed from overlapping time series values as we will see in our next section.

基於離散小波演算的基礎，我們以實際資料來說明示範 MODWT 的演算流程

表 6. MODWT 高通及低通濾波演算

time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
no of illStop	1	13	3	5	0	4	4	3	5	6	1	0	1	3	2	16	7
D1	0	6	-5	1	-2.5	2	0	-0.5	1	0.5	-2.5	-0.5	0.5	1	-0.5	7	-4.5
S1	1	7	8	4	2.5	2	4	3.5	4	5.5	3.5	0.5	0.5	2	2.5	9	11.5
D2	-2	2.75	3.5	-1.5	-2.75	-1	0.75	0.75	0	1	-0.25	-2.5	-1.5	0.75	1	3.5	4.5
S2	4.75	6.75	4.5	5.5	5.25	3	3.25	2.75	4	4.5	3.75	3	2	1.25	1.5	5.5	7

離散小波中每一層的高通及低通濾波會進行縮減取樣 (Downsampling)

所以由原始資料為  $a_1, a_2, a_3, \dots$

S1 數列為  $s1(1) = \frac{1}{2}(a_L + a_1)$  ,  $s1(2) = \frac{1}{2}(a_1 + a_2)$ , ...

D1 數列為  $D1(1) = \frac{1}{2}(-a_L + a_1)$  ,  $D1(2) = \frac{1}{2}(-a_1 + a_2)$ , ...

S2 數列為  $s2(1) = \frac{1}{2}(s1(L-1) + s1(1))$  ,  $s2(2) = \frac{1}{2}(s1(L) + s1(2))$ , ...

D2 數列為  $D2(1) = \frac{1}{2}(-s1(L-1) + s1(1))$  ,  $D2(6) = \frac{1}{2}(-s1(L) + s1(2))$ , ...

依此方式將原始資料，經一系列的高通及低通濾波運算得

$$\{D1, D2, \dots, Dj, Sj\}$$

#### 肆、研究結果

在各行政里的資料中，我們挑選了相對數量較多及有不同特性的行政里作為分析的範本，如下表

表 7. 違停數量

行政里	代碼	資料數量
三民里	C	3534
成大里	P	7182
赤崁里	Q	6546

三民里位於台南永康區，在高速公路交流道北側延伸至國立歷史博物館與鹽水溪兩岸，鄰永康工業區及台南科學園區行政里。



圖 6. 三民里 google 衛星地圖(google map 截圖)

赤崁里是從台南火車站延伸至赤崁樓及成功路民族路之間的市中心地區，鄰近台南市中西區的歷史地區。

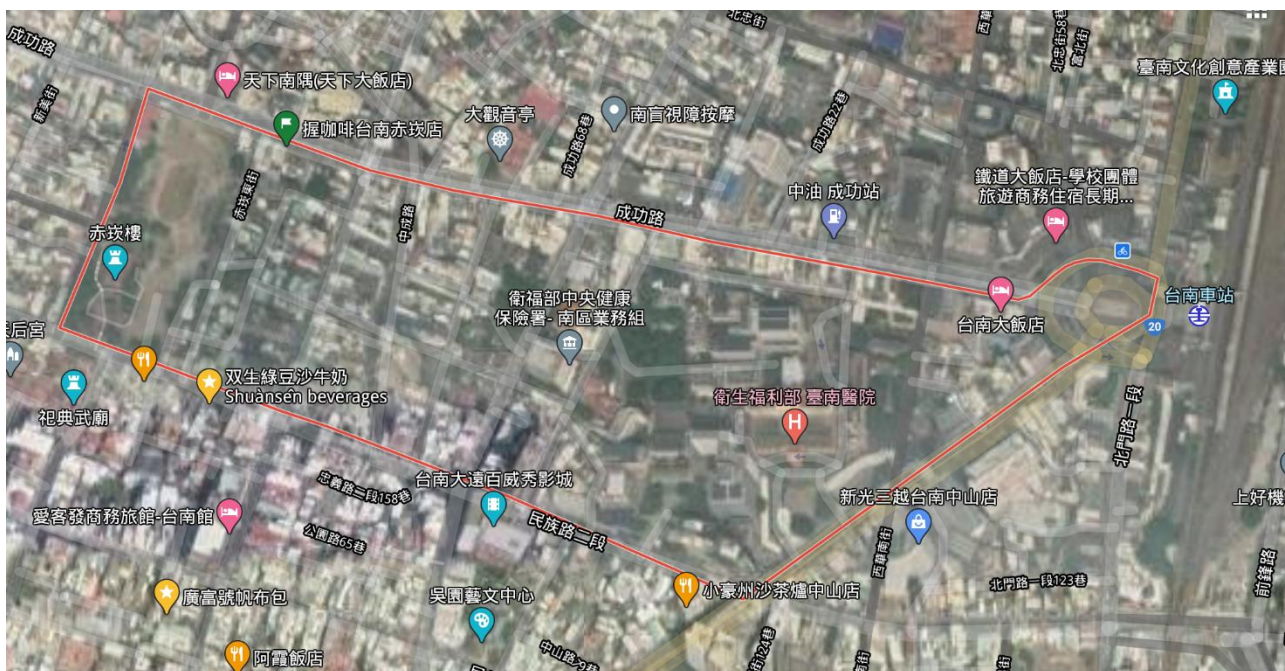


圖 7. 赤崁里 google 衛星地圖(google map 截圖)

成大里從台南火車站另一側延伸至勝利路，是受成功大學影響的典型大學周邊大學城行政里，生活作息往往會受大學的學期生活所影響。

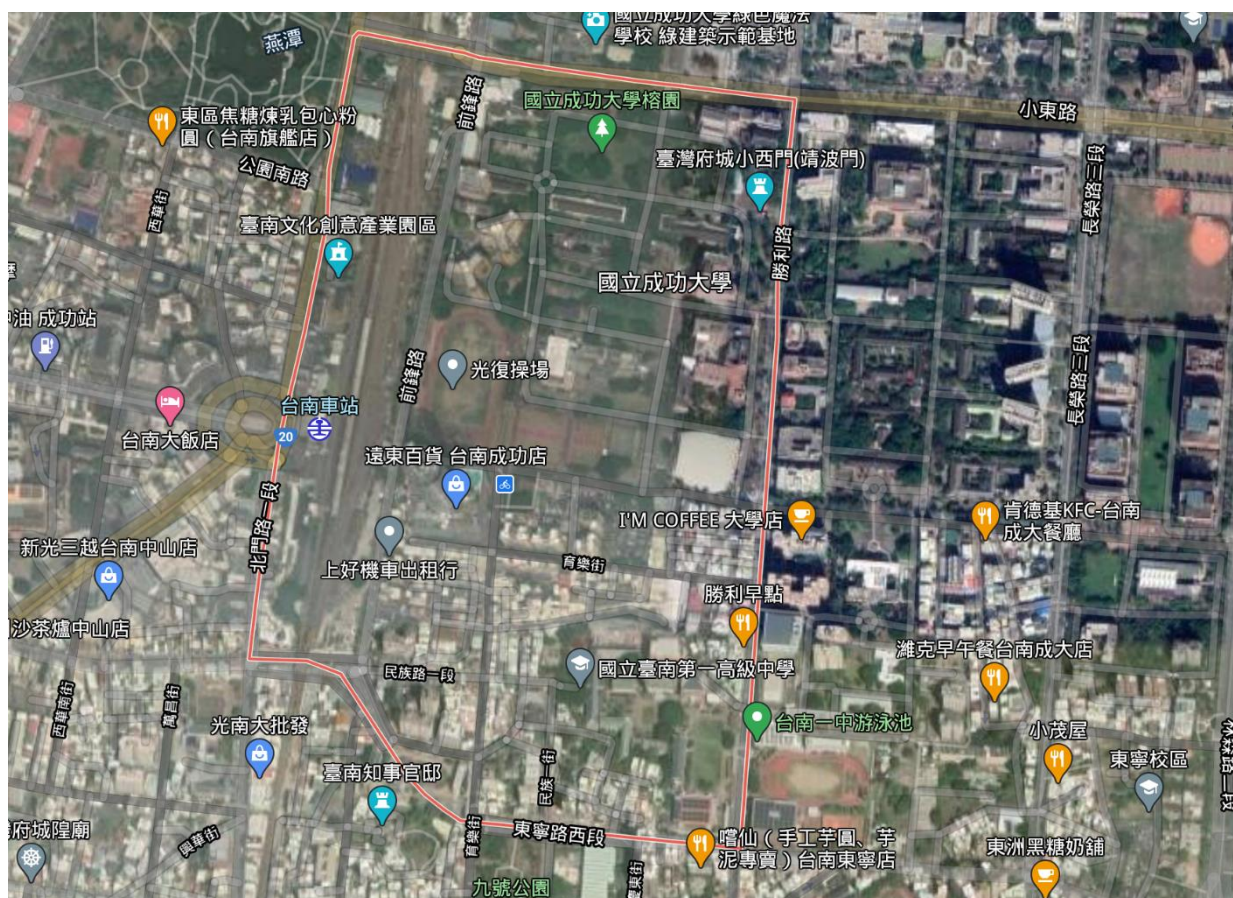


圖 8. 成大里 google 衛星地圖(google map 截圖)

選定此三個行政區是因為經濟活動、地理面貌、商業活動性質、房屋配置分佈都有很大差異，對「違停行為」的心理、經濟、地理等等成因各不相同，而根據我們的分析方法有哪些會相似？哪些會不同？

### 一、K-means 聚點分類取代地理熱點

K-means 聚點法的最合適  $k$  值選定如 三-2 K-means 聚點分類取代地理熱點 所介紹例如

下圖 17. 為例，當  $k=4$  之前；K-means 的平均中心距離隨著  $k$  增加而有大幅下降，但  $k>4$  開始，隨著  $k$  增加平均中心距離下降幅度有限，因此可以選定， $k=4$ ，進行分析。

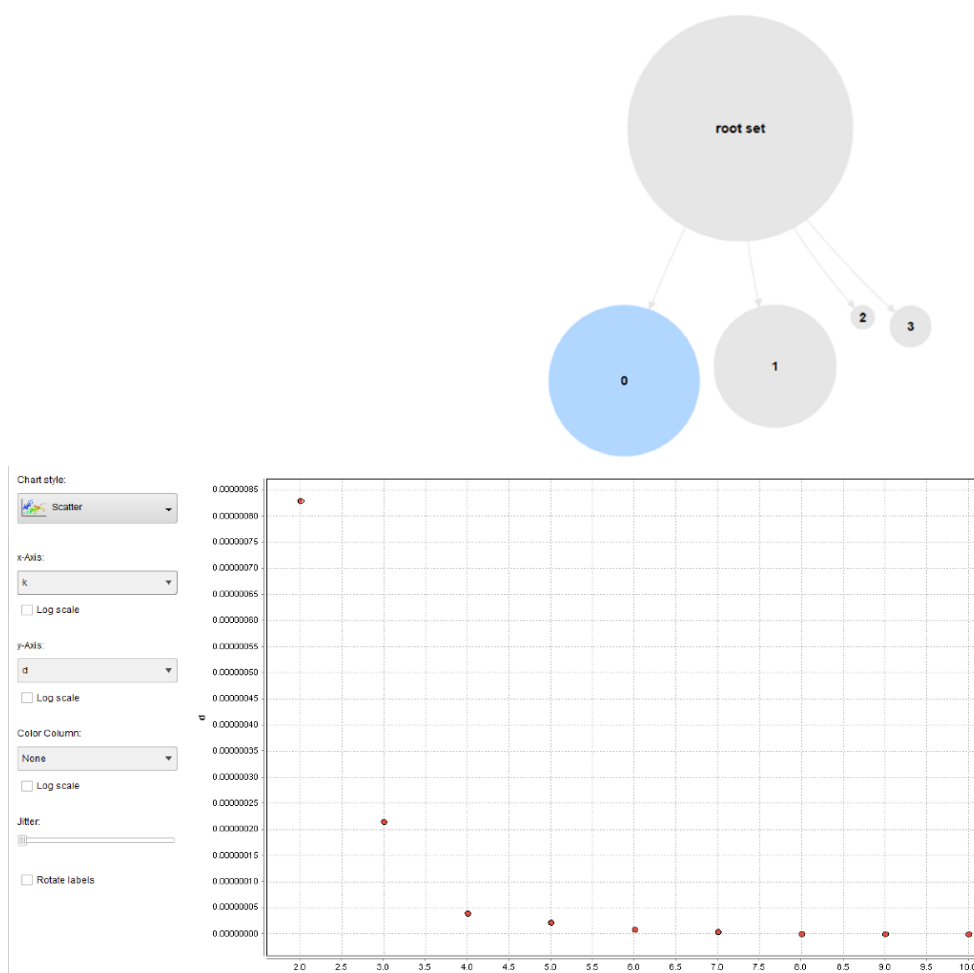


圖 9.不同  $k$  值， K-means 結果的平均中心距離示例

表 8.  $k=4$ , K-means 的分類情形

類別	資料數
Cluster 0	730 items

Cluster 1	565 items
Cluster 2	100 items
Cluster 3	174 items
Total number of items: 1569	

圖 10.  $k=4$ ,  $K$ -means 的分類情形

為重新將資料地點分類成  $G_0, G_1, G_2, G_3$  取代傳統  $TGOS$  的熱點來進行分析。

原則上就是根據  $Loop$  Parameters 中，選定適當的  $k$  值，再依此  $k$  值進行  $K$ -means 分類。

依前面  $3-2$   $K$ -means 聚點分類我們分別得到三個行政里的最合適  $k$  值與聚點分類

表 9.  $K$ -means 聚點分類

行政里	k 值	類別 0	類別 1	類別 2	類別 3	類別 4	類別 5	類別 6
三民里	2	3532	2					
成大里	7	2986	795	653	52	285	1559	851
赤崁里	6	5425	190	213	60	151	506	

考量到資料數量，我們選擇三民里(0)、成大里(0),(5),(6)及赤崁里(0)作為繼續分析的資料樣本。

## 二、資料設定及整理

違規停車的資料時間範圍是 2016 年 8 月 22 至 2017 年 8 月 27 日。考量國人生活作息以一星期為基礎的週期，我們設定「違停行為」會有以 7 日為週期的因子，但若單以一日為計數週期違停的數量太小，所以我們把同一聚類點一週的累計做為計算的基礎。另外白天與夜晚的「違停行為」也會有不同，因此我們把資料分成兩個類型，即一週間每天 8 時至 16 時的違停累計數量，與一週間全天的違停累計數量。前者資料標記為  $Day$ ，後者標記為  $allDay$ 。因為統計後有記錄的週數約略為 58 週，我們取第 8 週至 57 週，共 50 筆數字，其中最後 4 筆用來做比並較預測結果，而前 46 筆進行各種數學模型的運算，這樣所有的模型就有共同的比較基礎。

所以在三個里中共 5 個聚類( $k$ -means)，每一個聚類都以一週間全天及白天的違停累計數量，得到  $allDay$  及  $Day$  的兩個數列，每一個數列各有 50 項。前 46 項作為預測，並以最後 4 項來判斷預測的結果。因此

違停數列  $(X_t)_{t=1}^{t=50}$ ，以  $X_1, X_2, X_3, \dots, X_{46}$  進行模型訓練，得到模型結果  $(y_t)_{t=1}^{t=50}$ ，再比較



$y_{47}, y_{48}, y_{49}, y_{50}$  與  $X_{47}, X_{48}, X_{49}, X_{50}$ ，比較時我們採用方均根誤差(Root Mean Square Error),

$$RMSE = \sqrt{\frac{1}{4} \sum_{i=47}^{50} (y_i - X_i)^2}$$

做為同一個聚類的比較。而不同聚類間我們則以

$$\frac{RMSE}{\text{Mean of } X_t} \text{ 或 } \frac{RMSE}{\text{Standard deviation of } X_t} \text{ 或 } \frac{RMSE}{\text{Maximal of } X_t}$$

做為比較基準。

### 三、CWT

違規停車的數量呈現上下波動的特性，很自然聯想到小波分析，因此我們分別考慮了連續及離散小波的方法，試圖去找出三個里中共 5 個聚類(k-means)的模式。如同三-4 介紹的方法：

$$\alpha_n = \frac{2}{L} \sum_{k=1}^L \left( -\frac{a_k}{n} (k \cos nak) + \frac{a_k}{n} ((k-1) \cos n\alpha(k-1)) + \frac{a_k}{n^2} (\sin nak - \sin n(k-1)\alpha) \right) + \frac{2}{L} \sum_{k=1}^L \left( -\frac{b_k}{n^2} (\cos nak - \cos n(k-1)\alpha) \right)$$

$$L = 46, f(t) \sim \sum_{k=1}^{15} a_k \sin(kat)。$$

選擇不同的 $\alpha$ 值，每次變動 0.05 並比較  $\sum_{t=1}^{50} |f(t) - X(t)|$  值的大小，去找相對較佳的模型。利用 Excel 即可進行順向及反向的變換。

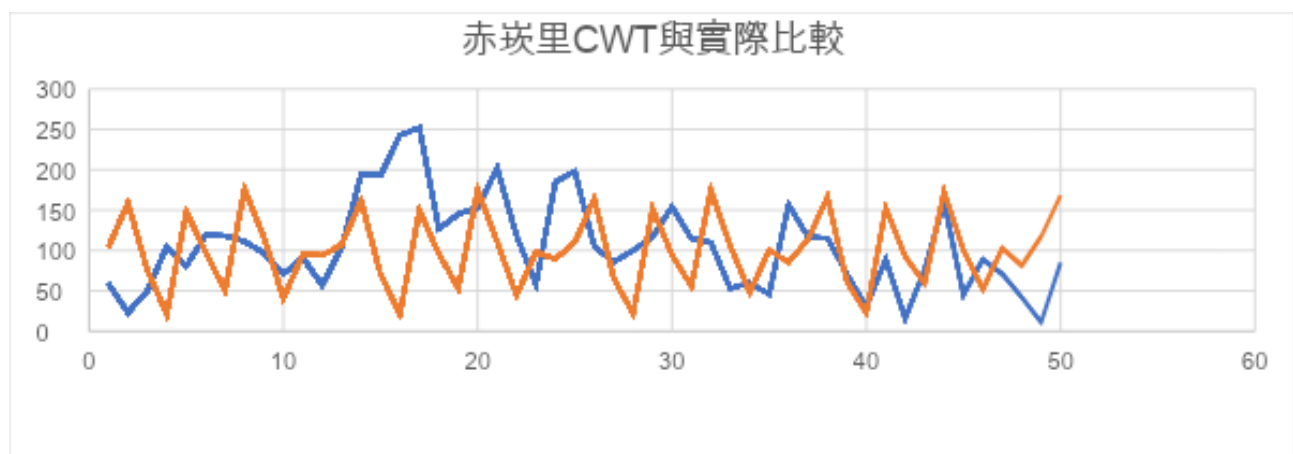


圖 11(1). 赤崁里 CWT( $\alpha = 10$ )實際與預測值比較

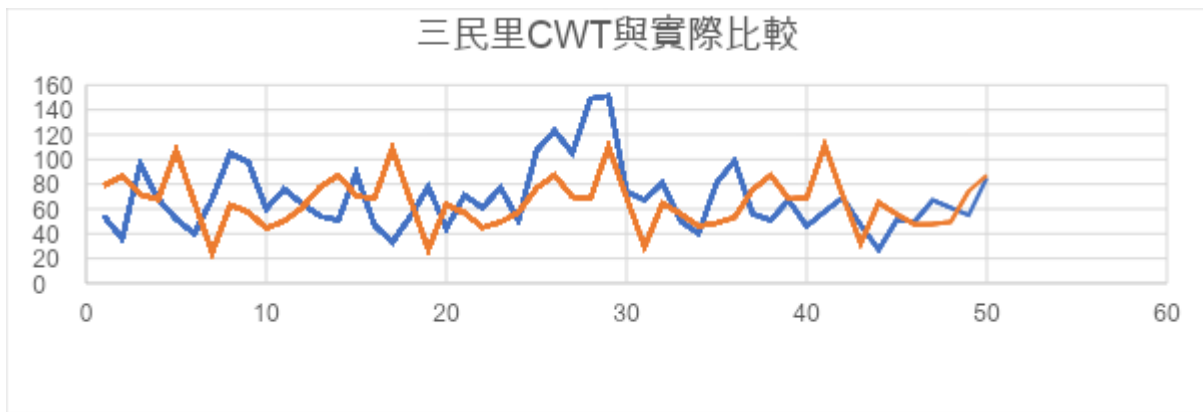


圖 11(2). 三民里 CWT( $\alpha = 10$ )實際與預測值比較

因為成大里的違規停車較特殊，以成大里聚類(0)作示例

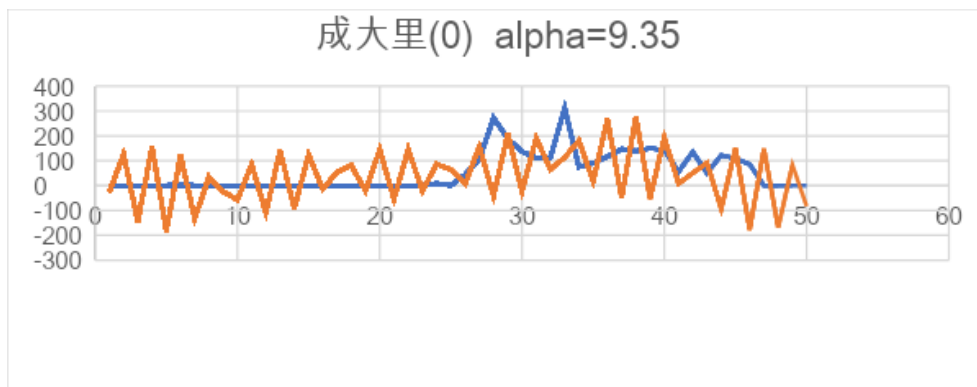


圖 12(1). 成大里 CWT( $\alpha = 9.35$ )實際與預測值比較

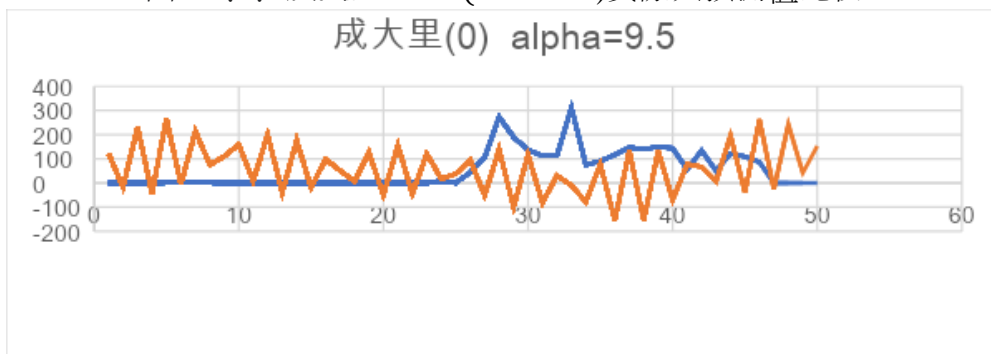


圖 12(2). 成大里 CWT( $\alpha = 9.5$ )實際與預測值比較



圖 13(1). 成大里 CWT( $\alpha = 9.75$ )實際與預測值比較

表 10. 成大里(0)  $\alpha = 9.35, 9.5, 9.75$  CWT 比較

週數	47	48	49	50	RMSE	總離差
違停數量	0	0	0	0		
$\alpha=9.35$	91.08099	62.60904	92.76574	81.20162	82.78711	1997.72
$\alpha=9.5$	9.738833	34.94883	7.393885	16.1155	20.19062	2050.708
$\alpha=9.75$	89.77793	55.52759	-18.3442	24.12839	54.91375	3518.934

很明顯從週期現象來看  $\alpha=9.75$  似乎比較符合真實的波動行為，但以 RMSE 及整體絕對離差來看預測能力卻是最差的。而 RMSE 及整體絕對離差在  $\alpha=9.5$  最佳但在週期現象來看卻不符合。這樣的結果在赤崁里及三民里也有類似的結果。

表 11. 赤崁里及三民里 CWT 預測的 RMSE 值

CWT 預測											RMSE
週數	1	2	3	4	5	6	47	48	49	50	
赤崁里違停數	59	23	50	105	81	119	71	42	11	85	
赤崁里 CWT 值	104.33	158.96	75.621	20.942	148.15	97.565	102.84	81.408	117.29	168.58	72.197
三民里違停數	54	36	96	68	52	40	67	61	55	85	
三民里 CWT 值	78.886	86.55	71.679	67.722	107.05	65.905	47.889	49.545	74.449	86.994	14.821

#### 四、DWT+ARIMA 與 MODWT+ARIMA

ARIMA 演算在三-3 節中已有介紹，ARIMA (p, d, q) 模型是

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

$\mu$  是  $X_t$  平均數， $\phi_i$  是自回歸參數， $\theta_i$  是移動平均參數， $\varepsilon_t$  是誤差量。

實際執行時 (p, d, q) 的參數則要依經驗去調整，因為每一數列中取前 46 項作為預測，並

以最後 4 項來判斷預測的結果。所以選定在 windows Size=46, step Size=4, horizon Size=4

我們試了一些情形 p=2,3,4,5,6,7,8,9,10,12, d=0,1, q=1,2,3,4,5,6,8 後發現選定 (p, d, q) = (6,0,1) 可以得到各方案較好的預測結果。

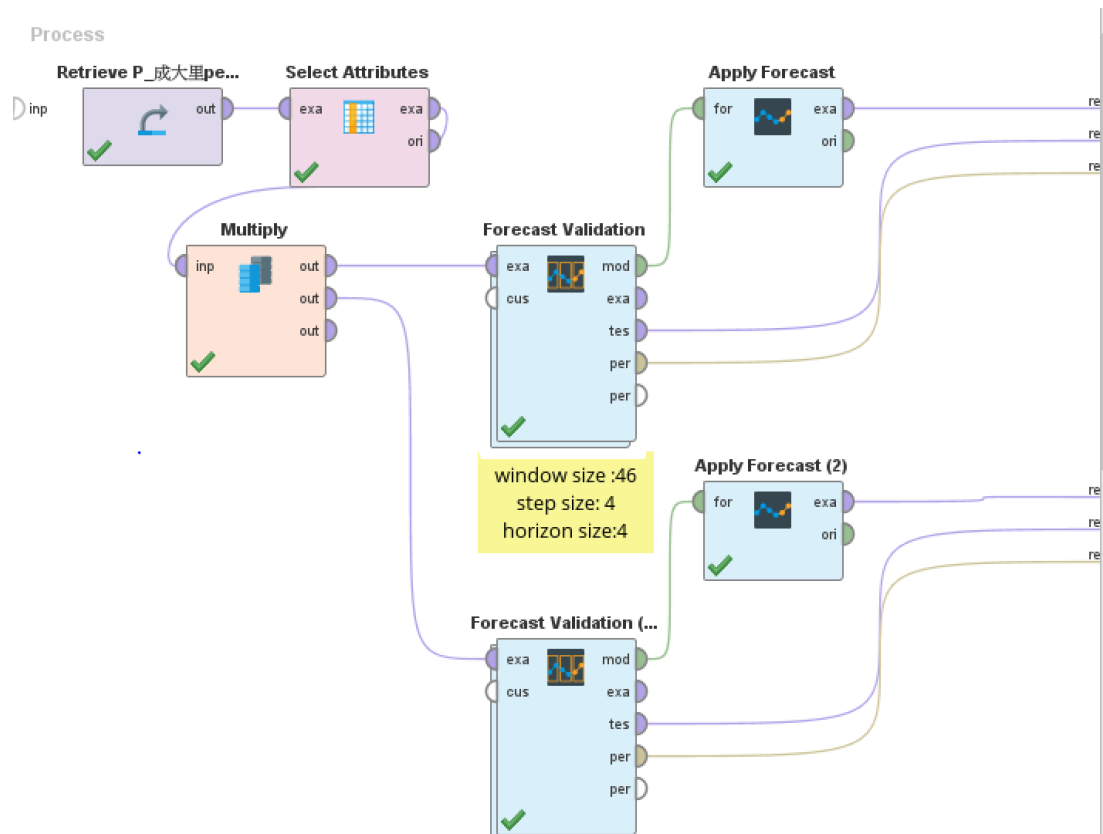


圖 14(1). 各聚類點資料進行 ARIMA 方案

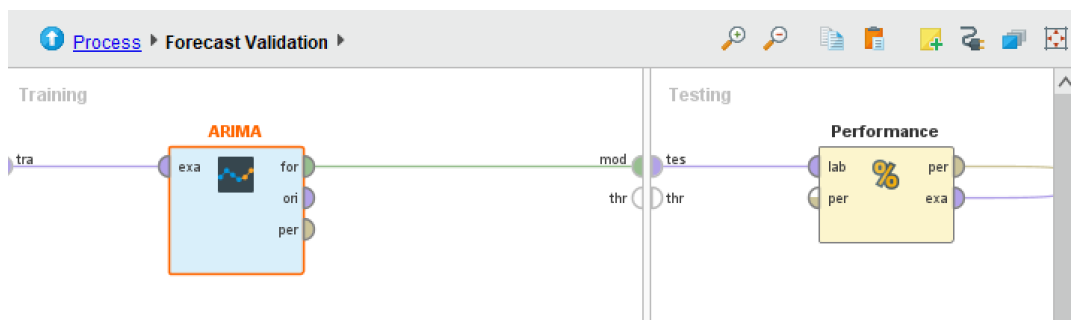


圖 14(2). 各聚類點 ARIMA 方案 Forecast Validation 內 ARIMA 流程

ARIMA 演算程序：

- (1) 選定各聚類點中分別以 allDay、Day 的資料
- (2) ARIMA 演算
- (3) 以最後 4 項預測值與實際值做比較，計算 RMSE 值

ARIMA+DWT 演算程序：

- (1) 選定各聚類點中分別以 allDay、Day 的資料
- (2) 每一數列做 DWT 演算，選擇 Level 1、Level 2 或 Level 4
- (3) ARIMA 演算 DWT 結果的前 46 項，以最後 4 項預測(第 47 至 50 項)
- (4) 以最後 4 項預測值與實際值做比較，計算 RMSE 值

ARIMA+MODWT 演算程序：

- (1) 選定各聚類點中分別以 allDay、Day 的資料
- (2) 選定 Level 後將高通濾波(D)全部相加，將相加後的高通濾波進行 ARIMA 演算
- (3) 將低通濾波(S)單獨進行 ARIMA 演算
- (4) 將其與進行 ARIMA 演算後的高通濾波相加
- (5) 以最後 4 項預測值與實際值做比較，計算 RMSE 值

表 12. 各聚類點資料的最大最小範圍、平均值及標準差

	allDay			Day		
	範圍	平均	標準差	範圍	平均	標準差
赤崁里	[16, 252]	110.78	54.923	[6,162]	50.52	36.744
三民里	[27,151]	69.87	27.318	[20,140]	51.85	25.840
成大(0)	[0,312]	64.76	76.845	[0,267]	47.87	63.644
成大(5)	[0,101]	28.76	23.121	[0,70]	23.28	19.818
成大(5)	[0,67]	16.72	13.551	[0,61]	14.02	12.398

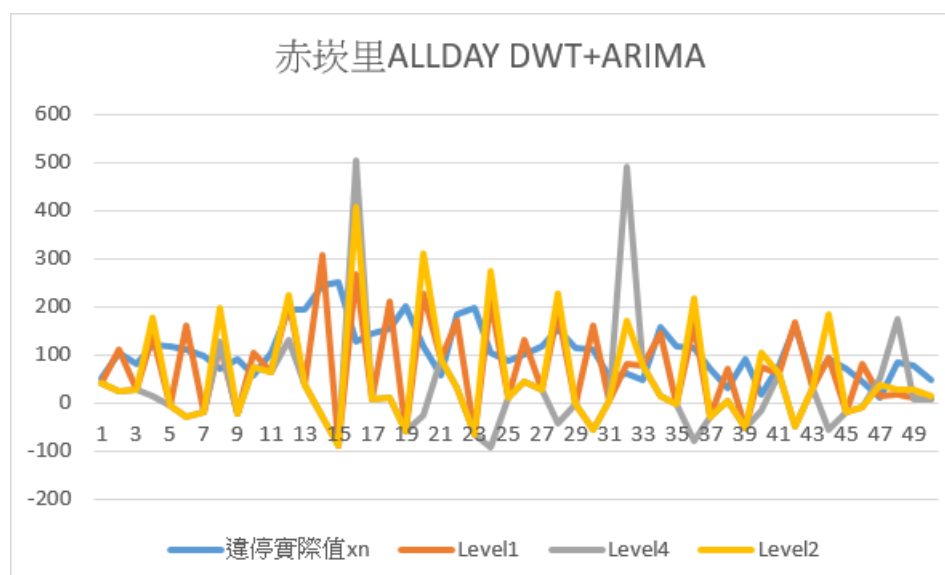


圖 15. 赤崁里 allDay DWT+ARIMA

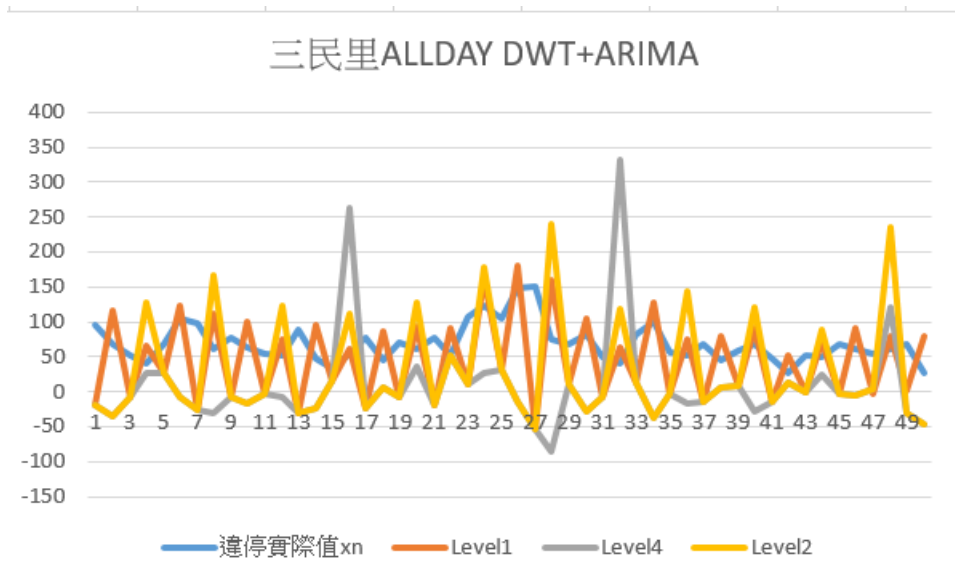


圖 16.三民里 allDay DWT+ARIMA

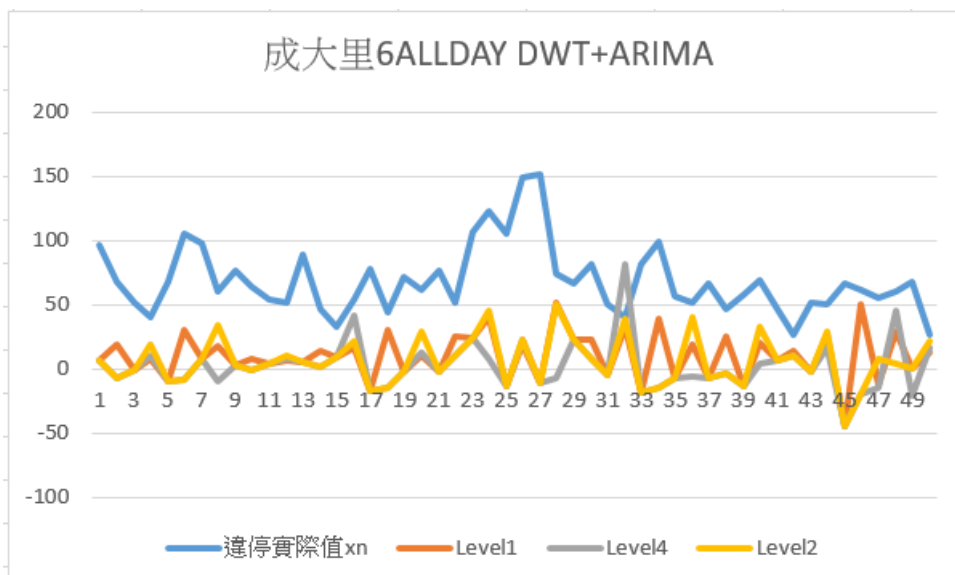


圖 17.成大里(6) allDay DWT+ARIMA

根據圖 15、16、17，可以發現 DWT+ARIMA 在各里中表現皆不太穩定。

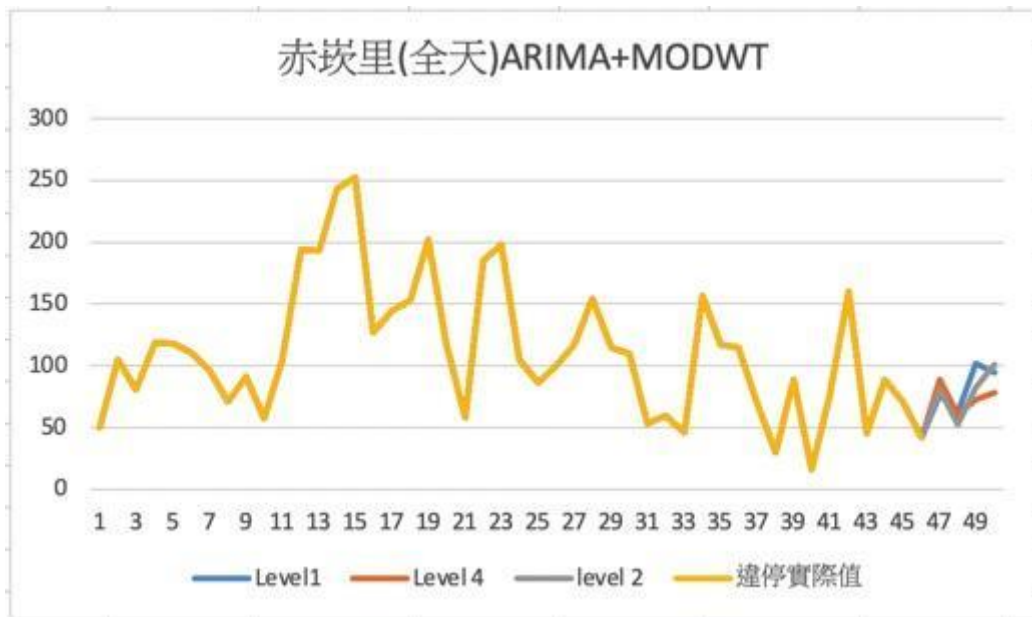


圖 18.赤崁里 allDay ARIMA+MODWT

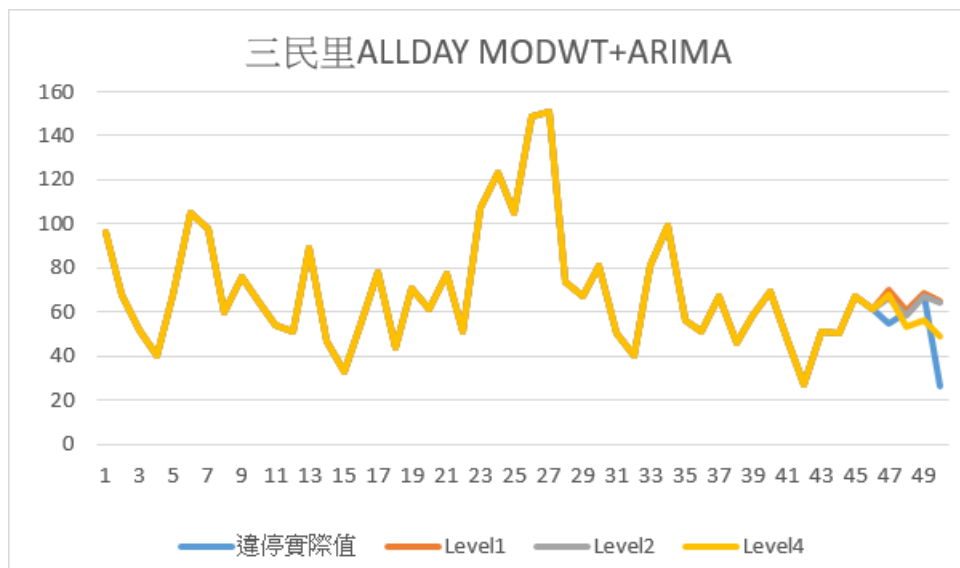


圖 19.三民里 allDay ARIMA+MODWT

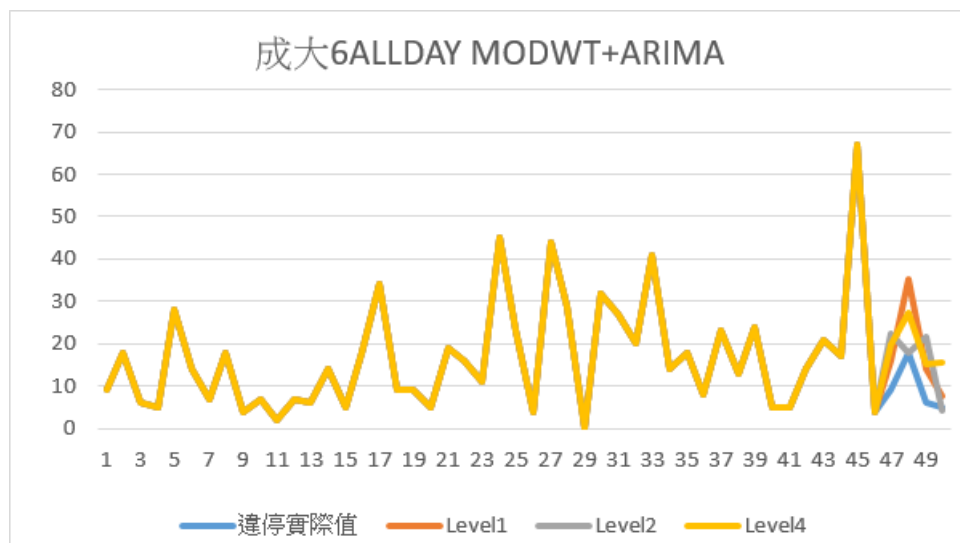


圖 20.赤崁里 allDay ARIMA+MODWT

表 13. 三民里、赤崁里 ARIMA 及 ARIMA+MODWT 預測結果

		三民里				RMSE	赤崁里				RMSE
allDay	週數	47	48	49	50		47	48	49	50	
	違停數量	55	60	68	26		11	85	77	47	
	Level 1	70.295	60.876	68.798	64.952	20.932	76.337	61.942	101.31	94.567	43.744
	Level 2	66.691	58.207	66.736	64.272	20.039	81.050	52.400	82.599	101.31	47.303
	Level 4	67.947	53.299	56.260	48.555	14.654	0	60.276	72.904	78.328	20.800
	ARIMA	72.553	63.096	67.302	61.616	19.917	103.53	82.661	95.608	104.56	55.289
Day	違停數量	23	40	31	23		4	55	42	26	
	Level 1	44.399	41.500	41.854	38.195	14.220	27.428	44.403	53.088	37.172	15.074
	Level 2	39.325	39.197	41.980	39.934	12.986	19.312	23.455	41.802	46.590	20.331
	Level 4	39.838	33.167	33.095	28.734	9.585	26.960	29.889	32.526	32.918	17.995
	ARIMA	43.751	42.148	42.312	37.991	14.035	38.114	42.708	44.656	42.866	20.060

據圖 18、19、20 及表 13，我們發現：

- 1.三民里與赤崁里中不管 allDay 還是 Day 皆是 ARIMA+MODWT Level 4 最優
- 2.整體來說，預測結果 ARIMA+MODWT>ARIMA，ARIMA+MODWT 預測能改善單獨使用 ARIMA 結果。

因此，我們初步發現在三民里與赤崁里中 ARIMA+MODWT 的預測可以改善 ARIMA 的預測結果。

表 14. 成大里(0)、(5)、(6) ARIMA 及 ARIMA+MODWT 預測結果

全天	類 0				RMSE	類 5				RMSE	類 6				RMSE
週數	47	48	49	50		47	48	49	50		47	48	49	50	
違停數	4	3	0	0		1	39	80	16		6	25	10	19	
Level 1	74.78	77.81	95.37	68.80	78.16	12.48	82.07	25.92	33.51	29.51	15.92	35.24	13.97	7.81	9.28
Level 2	33.84	76.46	77.08	82.35	68.94	15.97	25.29	27.00	31.74	29.45	22.48	18.00	21.63	4.36	12.94
Level 4	-69.15	-7.76	-1.94	-2.24	37.00	16.34	16.26	13.74	14.63	35.86	19.75	27.30	15.10	15.77	7.59
ARIMA	57.17	56.88	67.45	61.06	59.18	59.18	22.16	22.13	25.19	31.92	16.75	25.99	9.78	12.36	6.34
白天	類 0				RMSE	類 5				RMSE	類 6				RMSE
違停數	4	1	0	0		1	25	79	16		6	23	8	17	
Level 1	55.47	62.12	53.91	58.05	56.26	7.86	27.47	19.78	30.25	30.67	16.05	19.69	12.03	7.24	7.51
Level 2	22.91	29.56	47.09	57.98	41.08	12.32	21.94	20.16	27.39	30.53	16.44	12.77	15.78	4.18	10.47
Level 4	41.01	34.64	23.16	16.31	28.57	14.62	19.76	18.24	22.40	31.41	17.59	20.59	14.15	11.18	7.28
ARIMA	50.12	51.37	46.93	46.56	47.52	15.79	21.13	17.61	22.57	31.80	12.61	15.50	9.34	6.85	7.16

據表 14，我們發現：

- 1.成大里(0)的 Day 跟 allDay 中 ARIMA+MODWT Level 4 效果最優，而 ARIMA+MODWT Level 1 最差
- 2.成大(5)的 Day 跟 allDay 中皆是 ARIMA+MODWT Level 2 最優，而 Day 是 ARIMA 效果最差，allDay 則是 ARIMA+MODWT Level 4 最差，但整體來說數值都很差。



3.成大里(6) allDay 跟 Day 皆是 ARIMA 表現最優，但 ARIMA+MODWT Level 4 也很好，整體來看都不錯。

由此可知 ARIMA+MODWT Level 4 都不會差 ARIMA 太多，而 ARIMA+MODWT 幾乎都有某個 Level 可以改善 ARIMA，成大(6)中則是整體都很好。

伍、研究結果與討論：

本文討論及研究了連續小波變換 CWT、Harr 離散小波變換 DWT、最大重複離散小波轉換 MODWT(Maximal Overlap Discrete Wavelet Transform)、差分整合移動平均自迴歸模型 ARIMA 以及混合 MODWT+ARIMA 模型。對臺南市三個不同特性的行政里的共五個聚類點，進行預測。以週為單位，分成全日(allDay)及白天(Day, 08~16 時)兩種型態交叉比較分析，並以連續 46 週的資料預測接下來 4 週。結果採用方均根誤差 RMSE 相對平均值、標準差、最大值的比值，進行預測品質的比較，三個比值分別是：

$$\frac{RMSE}{Mean\ of\ X_t}, \frac{RMSE}{Standard\ divation\ of\ X_t}, \frac{RMSE}{Maximal\ of\ X_t}$$

三個值分別簡記為 R/M, R/Sd, R/Mx。得到結果如下列表 15。

表 15.預測結果

方案	成大里(0)				成大里(5)				成大里(6)				三民里(0)				赤崁里(0)				
	RMSE	R/M	R/Sd	R/Mx	RMSE	R/M	R/Sd	R/Mx	RMSE	R/M	R/Sd	R/Mx	RMSE	R/M	R/Sd	R/Mx	RMSE	R/M	R/Sd	R/Mx	
白天	ARIMA	47.5248	1.0767	0.7626	0.1780	31.8028	1.3340	1.5248	0.4026	7.1554	0.5118	0.5941	0.1173	14.0350	0.2805	0.5482	0.1002	20.0610	0.4092	0.5573	0.1238
	ARIMA+DWT L1	45.8589	1.0389	0.7359	0.1718	31.2979	1.3128	1.5006	0.3962	17.5879	1.2581	1.4604	0.2883	26.9985	0.5395	1.0545	0.1928	33.6690	0.6868	0.9353	0.2078
	ARIMA+DWT L2	61.2586	1.3878	0.9830	0.2294	32.8427	1.3776	1.5747	0.4157	20.8680	1.4927	1.7328	0.3421	24.0468	0.4806	0.9392	0.1718	24.5042	0.4999	0.6807	0.1513
	ARIMA+DWT L4	19.7803	0.4481	0.3174	0.0741	38.1656	1.6009	1.8299	0.4831	10.2791	0.7353	0.8535	0.1685	20.5430	0.4105	0.8023	0.1467	64.1247	1.3081	1.7813	0.3958
	ARIMA+MODWT L1	56.2598	1.2746	0.9028	0.2107	30.6717	1.2866	1.4706	0.3882	7.5122	0.5374	0.6238	0.1232	14.2205	0.2842	0.5554	0.1016	15.0744	0.3075	0.4187	0.0931
	ARIMA+MODWT L2	41.0848	0.9308	0.6593	0.1539	30.5329	1.2807	1.4639	0.3865	10.4711	0.7490	0.8695	0.1717	12.9855	0.2595	0.5072	0.0928	20.3318	0.4148	0.5648	0.1255
ARIMA+MODWT L4	28.5719	0.6473	0.4585	0.1070	31.4068	1.3174	1.5058	0.3976	7.2786	0.5206	0.6044	0.1193	9.5846	0.1915	0.3743	0.0685	17.9952	0.3670	0.4999	0.1110	
全天	ARIMA	59.1788	0.9909	0.7824	0.1897	31.9249	1.0941	1.3425	0.3161	6.3367	0.3822	0.4812	0.0946	19.9165	0.2909	0.7376	0.1319	55.2885	0.5200	0.9974	0.2194
	ARIMA+DWT L1	61.8233	1.0352	0.8173	0.1982	32.4296	1.1114	1.3637	0.3211	11.1511	0.6726	0.8468	0.1664	52.0697	0.7606	1.9283	0.3448	52.4967	0.4938	0.9470	0.2083
	ARIMA+DWT L2	97.9027	1.6394	1.2943	0.3138	33.0793	1.1336	1.3910	0.3275	21.1419	1.2751	1.6056	0.3156	72.9809	1.0660	2.7027	0.4833	63.8422	0.6005	1.1517	0.2533
	ARIMA+DWT L4	25.3340	0.4242	0.3349	0.0812	44.1368	1.5126	1.8560	0.4370	11.7677	0.7098	0.8937	0.1756	10.2119	1.6099	4.0815	0.7299	44.2767	0.4165	0.7987	0.1757
	ARIMA+MODWT L1	78.1602	1.3088	1.0333	0.2505	29.5053	1.0111	1.2407	0.2921	9.2773	0.5595	0.7045	0.1385	20.9320	0.3058	0.7752	0.1386	43.7445	0.4114	0.7891	0.1736
	ARIMA+MODWT L2	68.9383	1.1544	0.9114	0.2210	29.4487	1.0092	1.2383	0.2916	12.9446	0.7807	0.9831	0.1932	20.0392	0.2927	0.7421	0.1327	47.3034	0.4449	0.8533	0.1877
ARIMA+MODWT L4	36.9996	0.6196	0.4891	0.1186	35.8623	1.2290	1.5080	0.3551	7.5940	0.4580	0.5767	0.1133	14.6548	0.2141	0.5427	0.0971	20.7996	0.1956	0.3752	0.0825	

根據表 15.可以得到：

- 1.ARIMA+MODWT L4 除了成大里(5)之外，預測結果皆表現不錯。
- 2.ARIMA+DWT L2 普遍表現不理想。
- 3.成大(5)任何方法皆不理想。
- 4.除了成大里(6)之外，ARIMA+MODWT 及 ARIMA+DWT 皆有其中一個Level 能夠確實改善單獨使用 ARIMA 的預測結果。

5.整體來說，預測結果  $ARIMA+MODWT > ARIMA > ARIMA+DWT$ ， $ARIMA+MODWT$  符合預測能改善單獨使用  $ARIMA$  結果， $ARIMA+DWT$  則較不穩定，甚至有時結果比單獨使用  $ARIMA$  差。

6.赤崁里(0)由於數據較分散，在使用  $ARIMA+MODWT$  時若將  $D1$  數列視為噪聲，則  $ARIMA+MODWT$  預測效果會更精確。

7.各里白天及全天表現最好及最差的方法幾乎皆相同。

從結果來看我們有下列結論：

1.整體來說，預測結果  $ARIMA+MODWT > ARIMA > ARIMA+DWT$ ， $ARIMA+MODWT$  符合預測能改善單獨使用  $ARIMA$  結果， $ARIMA+DWT$  則較不穩定。

2. $ARIMA+MODWT$  整體來看真的能改善單獨使用  $ARIMA$  情況，符合研究假設。

3.成大里(0)中有連續多週資料為 0，以一年的資料長度而言影響太大，所以預測結果參考性較低。

4.本文中所使用的資料為 2016 年 8 月到 2017 年 8 月的資料，受限於資料僅有一年，但隨經濟地理活動及城市規劃的改變，長時間的資料容易受外來因素影響，若能有兩年以上的資料相信會有更好的結果。

5.在確定比較標準前，我們也試過各聚類點其他資料形態：

(1)  $X_1$  = 第一週 8~16 時,  $X_2$  = 第一週 16~24 時,  $X_3$  = 第二週 8~16 時,  $X_4$  = 第二週 16~24 時,...。並以前 92 項預測最後 8 項。

(2)  $X_1$  = 第一週 8~14 時,  $X_2$  = 第一週 14~18 時,  $X_3$  = 第一週 18~24 時,  $X_4$  = 第二週 8~14 時,...,  $X_{144}$  = 第 46 週 18~24 時。並以前 132 項預測最後 12 項。

(3)  $X_1$  = 第一週 8~12 時,  $X_2$  = 第一週 12~16 時,  $X_3$  = 第一週 16~20 時,  $X_4$  = 第一週 20~14 時,...,  $X_{168}$  = 第 42 週 18~24 時。並以前 142 項預測最後 16 項。最後的結果都很差，推測因為以週為單位，前後時段混合了不同時序的資料，所以影響了預測結果。因此本文選定了 **allDay** 及 **Day** 兩種資料型態，而藉由比對兩種資料型態的模擬結果，也可以知道日間及夜間的違停模式。

6.藉由本研究推論結合小波分析及時間序列的此種預測方法可以應用在風力發電、疾病高峰期此種帶有潛在變因的趨勢上，且不需長週期即可預測。

陸、參考文獻

- [1] 張新立, 張起豪, 影響小客車駕駛人違規停車意向成因研究分析, 九十三年 道路交  
通 安全與執法國際研討會, 2004 /9
- [2] 黃詩萍, 違規停車執法政策對違規停車行為影響之分析-以臺南市為例 The  
Influence of Banning Illegal Parking Policy on Parking Behavior-A Case  
Study of Tainan City, 交通管理科學系碩士在職專班, 2018 碩士論文, 成功大學
- [3] Ken Bailey, “A Combined Wavelet and ARIMA Approach to Predicting Financial Time Series”  
, Dublin City University, Master of Science

附錄：

表 1.成大(0),(6),(5) allDay 及 Day 資料

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
(0)allDay	0	0	0	0	1	3	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	9	0
Time	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	44	105	273	190	137	112	114	312	75	91	117	147	140	150	144	56	135	50	122	109	86	0	0	0	0
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
(0)Day	0	0	0	0	1	3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	9	0
Time	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	36	68	267	167	85	83	66	264	56	67	97	111	94	106	103	22	93	24	87	58	70	53	55	54	4
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
(6)allDay	3	10	9	9	18	6	5	28	14	7	18	4	7	2	7	6	14	5	18	34	9	9	5	19	16
Time	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	11	45	23	4	44	29	0	32	27	20	41	14	18	8	23	13	24	5	5	14	21	17	67	4	6
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
(6)Day	2	9	9	7	17	2	1	22	12	6	13	1	6	0	5	2	14	5	17	34	9	8	5	19	16
Time	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	9	43	22	3	21	24	0	22	26	18	39	13	7	4	21	9	24	3	5	13	18	15	61	4	6
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
(5)allDay	5	40	55	51	21	31	15	46	8	15	57	8	3	1	60	6	101	28	28	56	82	67	45	50	42
Time	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	26	51	55	12	37	1	0	28	4	7	27	32	25	15	37	13	37	9	3	11	14	29	3	26	1
Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
(5)Day	4	8	15	31	21	31	15	37	8	14	57	1	3	1	60	4	62	14	16	37	70	63	45	50	36
Time	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	15	26	55	12	25	1	0	22	0	0	27	19	13	13	37	13	36	3	2	11	8	29	2	26	1

## 【評語】 190021

本研究主題清楚且聚焦，且可用科學方法檢驗研究成果。實驗設計完整，資料與分析佳。建議可多思考並針對所提出方法套用在不同場域進行可行性分析。