

# 銘傳大學

應用統計與資料科學學系  
專題研究

指導教授：陳明輝博士

高雄地區 PM2.5 分析及預測-以高斯過程迴歸為  
工具

Analysis and prediction of PM2.5 in  
Kaohsiung area by using Gaussian process  
regression

05170463 蔡宜誼、05170561 張芳瑜

中 華 民 國 1 0 8 年 1 2 月

## 摘要

空氣污染是目前世界上熱門的議題之一，其中直徑小於 2.5 微米的細懸浮微粒(PM2.5)，由於其直徑細小容易影響身體各種器官，更成為受關注的議題。傳統對於細懸浮微粒的研究，會以氣象變數(meteorological variables)或其他污染物為自變項，先假設其影響方式的模型，再利用各種方法估計模型中的參數，如複迴歸法、時間數列法等。但是相關變數對於 PM2.5 的影響方式複雜，事先假設的模型難以考慮周全。機器學習(machine learning)中的高斯過程迴歸(Gaussian process regression)是一種無母數(nonparametric)的分析方法，可以避開事先定義明確模型，我們的主要方法是利用空間模型的觀念，藉由已知地區的 PM2.5，來預測無測站地區的 PM2.5。在此論文中，我們會先拿可能與 PM2.5 最為相關之變數，再利用 11 個測站去預測剩下的 1 個測站再和我們原先有的觀測值做比較，藉此我們可以計算得出以高斯過程預測的誤差值，並比較用不同的變數建立的模型。高雄地區的懸浮微粒濃度一直以來都是相當嚴重，本研究彙整高雄地區 2017、2018 年的 12 個測站資料，先分析和整理 PM2.5，再比較兩年的測站、月分和時段三個項目，之後再試圖用較嚴重的年份來建立高斯過程迴歸為研究工具，考慮時間及空間趨勢，預測高雄地區細懸浮微粒濃度。

關鍵詞：細懸浮微粒、機器學習、高斯過程迴歸、無母數

# 目錄

	頁次
摘要.....	I
目錄.....	II
表次.....	III
圖次.....	VI
第一章 緒論.....	1
第一節 研究背景 .....	1
第二節 研究動機 .....	2
第三節 研究目的 .....	4
第四節 名詞釋義 .....	4
第五節 研究架構 .....	7
第二章 文獻探討.....	9
第一節 空氣品質監測站以及空氣的簡介 .....	9
第二節 PM2.5 相關介紹 .....	10
第三節 空氣汙染物的相關研究 .....	12
第四節 高斯迴歸過程相關應用 .....	13
第五節 高斯過程之套件 .....	14
第三章 研究方法及套件介紹.....	17
第一節 介紹高斯過程回歸 .....	17
第二節 共變異函數 .....	19
第三節 套件介紹 .....	22
第四章 資料分析.....	24
第一節 資料整理 .....	24
4.1.1 介紹監測站、位置及監測項目 .....	24
4.1.2 資料前處理.....	28
第二節 高雄 PM2.5 分析 .....	31
4.2.1 測站.....	32
4.2.2 月份.....	34
4.2.3 時段.....	37
第三節 預測 PM2.5. ....	40
4.3.1 單變量-PM10、風速和風向 .....	40
4.3.2 雙變量-經緯度 .....	43

4.3.3 多變量-經緯度和 PM10 .....	49
4.3.4 多變量-經緯度和風速風向 .....	54
4.3.5 多變量總整理.....	59
第五章 結論與建議.....	63
第一節 結論 .....	63
第二節 建議 .....	65
參考文獻.....	66



## 表次

	頁次
表 1-2-1 全台 PM2.5 年均值最差的 10 個測站 .....	3
表 1-4-1 細懸浮微粒(PM2.5)指標 .....	6
表 2-1-1 空氣品質指標(AQI) .....	10
表 2-1-2 PM2.5 和 PM10 的比較 .....	10
表 2-5-1 高斯過程回歸套件 .....	16
表 3-5-1 GPFIT 主要參數介紹 .....	22
表 4-1-1 汙染物測英文簡稱、單位及測項名稱 .....	27
表 4-2-1 2017、2018 PM2.5 年平均值 .....	31
表 4-2-2 2017、2018 年高雄測站年平均比較 .....	32
表 4-2-3 2017 年 PM2.5 月平均 .....	34
表 4-2-4 2018 年 PM2.5 月平均 .....	35
表 4-2-5 2017 年 24 小時平均 .....	37
表 4-2-6 2018 年 24 小時平均 .....	38
表 4-3-1 PM10 單變量 .....	40
表 4-3-2 風速單變量 .....	41
表 4-3-3 風向單變量 .....	42
表 4-3-4 經緯度轉換 .....	43
表 4-3-5 24 小時平均誤差值(A) .....	44
表 4-3-6 24 小時平均誤差值(B) .....	45
表 4-3-7 24 小時誤差平均 .....	46
表 4-3-8 季節平均誤差(上) .....	47
表 4-3-9 季節平均誤差(下) .....	47
表 4-3-10 四季誤差平均 .....	48
表 4-3-11 24 小時平均誤差值(A) .....	49
表 4-3-12 24 小時平均誤差值(B) .....	50
表 4-3-13 24 小時誤差平均 .....	51
表 4-3-14 季節平均誤差(上) .....	52
表 4-3-15 季節平均誤差(下) .....	52
表 4-3-16 四季誤差平均 .....	53
表 4-3-17 24 小時平均誤差值(A) .....	54
表 4-3-18 24 小時平均誤差值(B) .....	55
表 4-3-19 24 小時平均誤差 .....	56
表 4-3-20 季節平均誤差(上) .....	57
表 4-3-21 季節平均誤差(下) .....	57
表 4-3-22 四季誤差平均 .....	58

表 4-3-23	以測站比較各模型 .....	59
表 4-3-24	以時段比較各模型 .....	60
表 4-3-25	以四季比較各模型 .....	61
表 4-3-26	多變量總比較 .....	62



## 圖次

	頁次
圖 1-1-1 全台測站.....	2
圖 1-4-1 機器學習的過程.....	5
圖 1-4-2 人工智慧、機器學習和深度學習的關係.....	5
圖 1-5-1 研究步驟流程圖.....	8
圖 2-2-1 臺灣 PM2.5 主要來源.....	11
圖 3-5-1 GPFIT 例子.....	23
圖 3-5-2 GPFIT 例子的結果圖.....	23
圖 4-1-1 高雄測站位置圖.....	26
圖 4-1-2 原始檔案(EXCEL).....	29
圖 4-1-3 轉置合併後的檔案(已匯入 R).....	30
圖 4-2-1 2017、2018 PM2.5 年平均値比較圖.....	31
圖 4-2-2 2017&2018 高雄測站 PM2.5 比較圖.....	33
圖 4-2-3 高雄測站 2017 年 PM2.5 月平均.....	34
圖 4-2-4 高雄測站 2018 年 PM2.5 月平均.....	35
圖 4-2-5 2017 年每小時平均.....	39
圖 4-2-6 2018 年每小時平均.....	39
圖 4-3-1 24 小時平均誤差値比較.....	46
圖 4-3-2 24 小時誤差平均圖.....	47
圖 4-3-3 季節平均誤差比較.....	48
圖 4-3-4 四季誤差平均圖.....	48
圖 4-3-5 24 小時平均誤差値比較.....	51
圖 4-3-6 24 小時誤差平均圖.....	51
圖 4-3-7 季節平均誤差比較.....	52
圖 4-3-8 四季誤差平均圖.....	53
圖 4-3-9 24 小時平均誤差値比較.....	56
圖 4-3-10 24 小時誤差平均圖.....	57
圖 4-3-11 季節平均誤差比較.....	58
圖 4-3-12 四季誤差平均圖.....	58
圖 4-3-13 以測站比較各模型.....	59
圖 4-3-14 以時段比較各模型.....	61
圖 4-3-15 以四季比較各模型.....	62

# 第一章 緒論

## 第一節 研究背景

Nery(2018, May, 2)在世界健康組織(World Health Organization ,WHO)的網站上提到，光是呼吸，全球每年就造成 700 萬人死亡，數十億人受到傷害；Taylor(2018, October, 29)也在英國衛報(The Guardian)的文章中寫道，世界上 90 %以上，總計 18 億的兒童正呼吸著有毒空氣。空氣汙染中直徑小於 2.5 微米的細懸浮微粒(PM2.5)，由於其直徑細小容易影響呼吸器官，也造成癌症和心血管等疾病，更成為受關注的議題。

為使民眾事先瞭解周遭環境空氣品質狀況，作為戶外活動參考，環保署於民國 79 年起推動建置臺灣地區空氣品質監測網(Taiwan Air Quality Monitoring, 簡稱 TAQM)。網址為 <http://taqm.epa.gov.tw/taqm/tw/default.aspx>)。如今全台已設立了 77 個空氣品質測站(如圖 1-1-1)，採取自動化每天 24 小時的偵測。目前，環保署(2018)的預報作業是根據空氣品質狀況及以空氣品質防制區為依據，參考氣象分區並考量臺灣地區空氣品質監測站分布與污染源分布特性，利用相關氣象參數及預報輔助系統中的氣候法、持續法、迴歸法及類比法得到客觀預報數值，再由預報人員主觀研判修正；每天提供北部、竹苗、中部、雲嘉南、高屏、宜蘭、花東七大空品區的 3 日空氣品質預報，以及 3 個離島地區(澎湖、馬祖及金門)隔日的空氣品質預報。



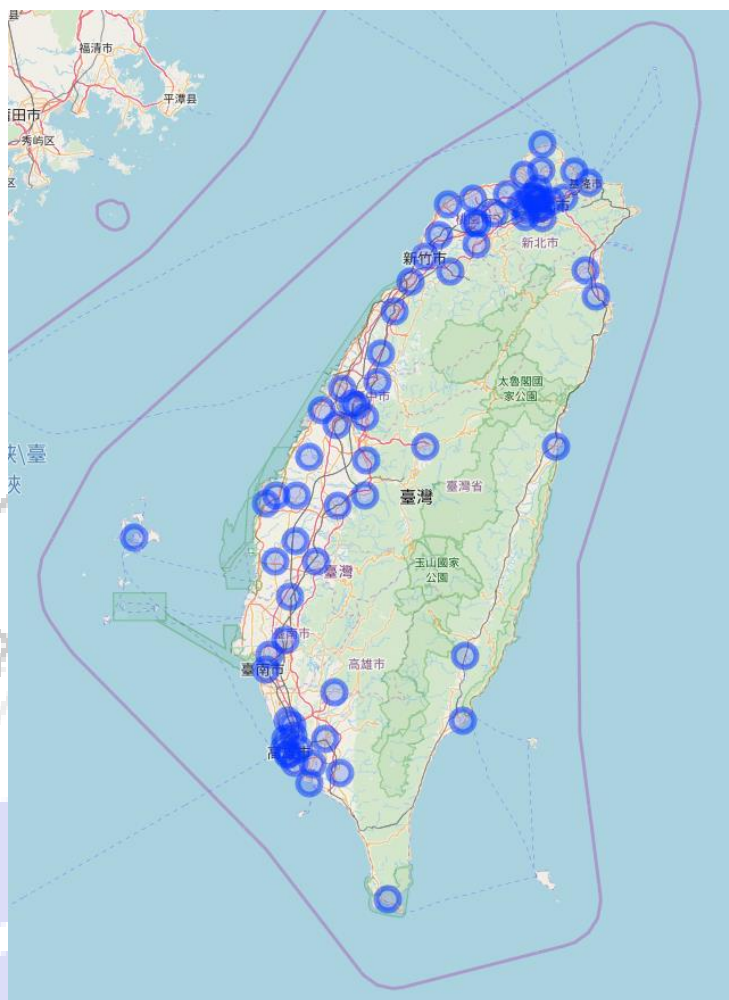


圖 1-1-1 全台測站

## 第二節 研究動機

近年來國人對空氣汙染逐漸重視，不外乎是因為它影響我們的生活品質，更侵略了我們的健康，然而在空氣汙染的主要汙染物中，更讓我們束手無策的莫非是直徑小於 2.5 微米之懸浮微粒，它幾乎等於整體空氣汙染的代名詞，整治困難度居於各種空氣汙染之首。Cifuentes, et. al. (2000) 的研究指出，PM<sub>2.5</sub> 相較於大顆粒可能具有更大的毒性，並認為 PM<sub>2.5</sub> 會被吸入到肺部深處且具更大的沉降表面積，每單位質量對有毒汙染物具更大吸附及濃縮能力。從上述兩點能判斷 PM<sub>2.5</sub> 比 PM<sub>10</sub> 對健康的危害，因此此次研究以 PM<sub>2.5</sub> 為主要研究之空氣汙染物。

根據環境資訊中心(2019)的報導，日前環保署宣布「PM2.5 紅害減半大作戰」的計劃，原定 2020 年要達成的目標，許多縣市在 2018 年就提前達標，但唯獨高雄地區的 PM2.5 年平均值大多不降反升，而年平均值最高的前四測站皆在高雄市 (表 1-2-1)。高雄為臺灣的重工業之都，長期發展工業拚經濟，使高雄成為臺灣重工業龍頭，但在繁榮工業發展下，卻也帶來相對嚴重的空氣汙染。

表 1-2-1 全台 PM2.5 年均值最差的 10 個測站

取自環境資訊中心 TEIA (<https://e-info.org.tw/node/215820>)

	PM 2.5 監測站	PM2.5 自動原始 2018 年均值
1	高雄市 左營站	37.2
2	高雄市 前金站	33.4
3	高雄市 小港站	33.2
4	高雄市 橋頭站	32.3
5	彰化縣 二頭站	31.4
6	南投縣 南投站	30.6
7	台南縣 新營站	30.5
8	高雄市 鳳山站	30.5
9	高雄市 仁武站	30.4
10	高雄市 楠梓站	30.2

近年許多關於預測空氣汙染的相關研究，線性迴歸(Linear regression)或時間序列(time series)等方法都有前人使用，像是簡單的線性迴歸只能用在包含一項獨立變數的線性關係且所有觀測值之權重皆相等，而時間序列中要注意未來發展變化規律和發展水平，不一定與歷史和現在的發展完全一致。這些缺點讓本研究決定使用近年來被廣泛應用在資料的分析與預測的機器學習，而高斯過程迴歸(Gaussian process regression,GPR)是機器學習領域的一種無母數貝氏迴歸(nonparametric Bayesian regression)分析方法，不須事先假設迴歸模型樣態，藉由訓練資料建立無母數模型，以預測新資料的輸出，而且可以求得該輸出的不確定性，對於本研究的資料此方法可以應用。因此，在這次研究中，我們的主要方法是利用空間模型的觀念，藉由已知地區的 PM2.5，來預測無測站地區的

PM2.5。在此論文中，我們會先拿可能與 PM2.5 最為相關之變數，再利用 11 個測站去預測剩下的 1 個測站再和我們原先有的觀測值做比較，藉此我們可以計算得出以高斯過程預測的誤差值，並比較用不同的變數建立的模型。

### 第三節 研究目的

本研究目的如下所示：

- (一)探討高雄地區 PM2.5 相關因素。
- (二)分析比較測站、月份和時段的 PM2.5 值。
- (三)運用高斯過程建立高雄地區 PM2.5 之模型並加以預測。
- (四)運用估計值與實際觀察值計算出誤差值，並比較不同變數建立之模型。

### 第四節 名詞解釋

#### (一) 細懸浮微粒(PM2.5)

細懸浮微粒(PM2.5)係指微粒氣動粒徑小於 2.5 微米，它的直徑還不到人的頭髮絲粗細的 1/28，比 PM10 更容易深入人體肺部，對健康影響更大，若細微粒又附著其他污染物，將更加深呼吸系統之危害。而 PM2.5 來源可分為自然源、原生性和衍生性，自然源為經由火山爆發、地殼變動或風化等作用形成，原生性為人為產生如火力發電廠、石化工廠、一般工廠燃燒不完全所產生的等，衍生性來源為工廠、汽機車、石化業所排的硫氧化物或氮氧化物、有機碳化合物，受到日照產生硫酸鹽、硝酸鹽及有機碳等有毒的細懸浮微粒，形成了二次污染物。而臺灣 PM2.5 主要來源有境內和境外兩種不同因素，汽機車排放、工廠廢氣排放或稻草垃圾露天燃燒等為境內因素，而中國大陸帶來的沙塵暴為境外因素。

## (二) 機器學習(Machine Learning)

「機器學習是用數據或以往的經驗，以此優化電腦程式的性能標準。」，這是 Alpaydin (2004) 對機器學習的定義。機器學習就是要讓電腦像人類一樣具有學習的能力，機器學習能夠幫助電腦從現有的資料中學習，利用從過往的資料和經驗中學習並找到其運行規則，以便預測未來的趨勢和結果(圖 1-4-1)。機器學習的厲害之處就在於它可以自主學習，例如當年 AlphaGo 挑戰世界冠軍韓國職業棋士李世石時，造成了不只圍棋界的關注，乃至於全球人類皆震驚不已。這項技術在近來非常熱門，就成為各家科技巨頭，包括微軟、Google、Facebook 等公司的研發重心，而蘋果手機的智慧助理 Siri 更是我們最熟知的例子。



圖 1-4-1 機器學習的過程

機器學習是從人工智慧(Artificial Intelligence, AI)衍生出來的，定義是製造出來的機器所表現出來的智慧，而深度學習(deep learning)是機器學習的分支，一種以人工神經網路為架構，對資料進行表徵學習的演算法。他們三者之間的關係為圖 1-4-2 所顯示。

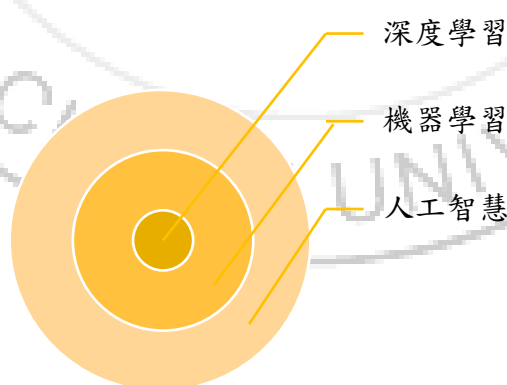


圖 1-4-2 人工智慧、機器學習和深度學習的關係

取自 GCP 專門家部落格(<https://blog.gcp.expert/ml-1-ai-ml-deep-learning-intro/>)

### (三) 紅害減半大作戰

我國環保署建立「細懸浮微粒(PM2.5)指標」，提供民眾日常生活上之行動建議，與空氣品質指標(AQI)併行，提升對民眾的健康保障的情況，指標分為綠、黃、紅、紫四種警示顏色等級，另外再細分為 10 等級(表 1-4-1)，並且在第 4 級開始，有敏感性體質的民眾就必須注意戶外活動及身體狀況。而「紫爆」就是屬於第 10 級的狀況，且空污紅害就是以前形容的紫爆，行政院環保署在 2017 年 12 月 21 日提出「紅害減半大作戰」，此計畫的目標就是更積極地對空汙防制有一定的計畫及改善目標，也提出更具體的行動方案，將確實的解決國內空氣污染的問題，且這些目標需要各界各部門及民眾的參與支持才能夠達成。

此次行動方案提出的政策目標，如下：

- 一、2019 年空污紅害減半。
- 二、2030 年新購公務車輛及公共運輸大巴士全面電動化。
- 三、2035 年新售機車全面電動化。
- 四、2040 年新售汽車全面電動化。

表 1-4-1 細懸浮微粒(PM2.5)指標

取自南投縣政府環境保護局

(<https://www.ntepb.gov.tw/sub/content/index.aspx?Parser=1,17,511,494,496>)

指標等級	1	2	3	4	5	6	7	8	9	10
分類	低	低	低	中	中	中	高	高	高	高
PM2.5 濃度 ( $\mu\text{g}/\text{m}^3$ )	0-11	12-23	24-35	36-41	42-47	48-53	54-58	59-64	65-70	$\geq 71$

## 第五節 研究架構

本論文共計五章，第一章為緒論，首先說明研究背景、動機與目的，再來解釋重要名詞並說明本研究架構與流程圖，見圖 1-5-1。第二章為文獻探討，說明相關測站和汙染物基本資料，將過去空氣汙染之相關研究做整理與討論，以及高斯迴歸的相關文獻也一併一起做介紹。第三章為研究方法，說明高斯過程的架構及共變異數的應用，以及實作高斯過程所用的 R 套件 GPFit。第四章為資料分析，蒐集並整理資料、使用問題的解決方法，先對 PM2.5 作分析再利用高斯過程迴歸模型建立模型及預測結果，分單變量、雙變量和多變量一一去研究，並整理出實際和預測之間的誤差值，第五章為研究結果，闡述本研究過程與結果，並提出本研究的結論與建議。

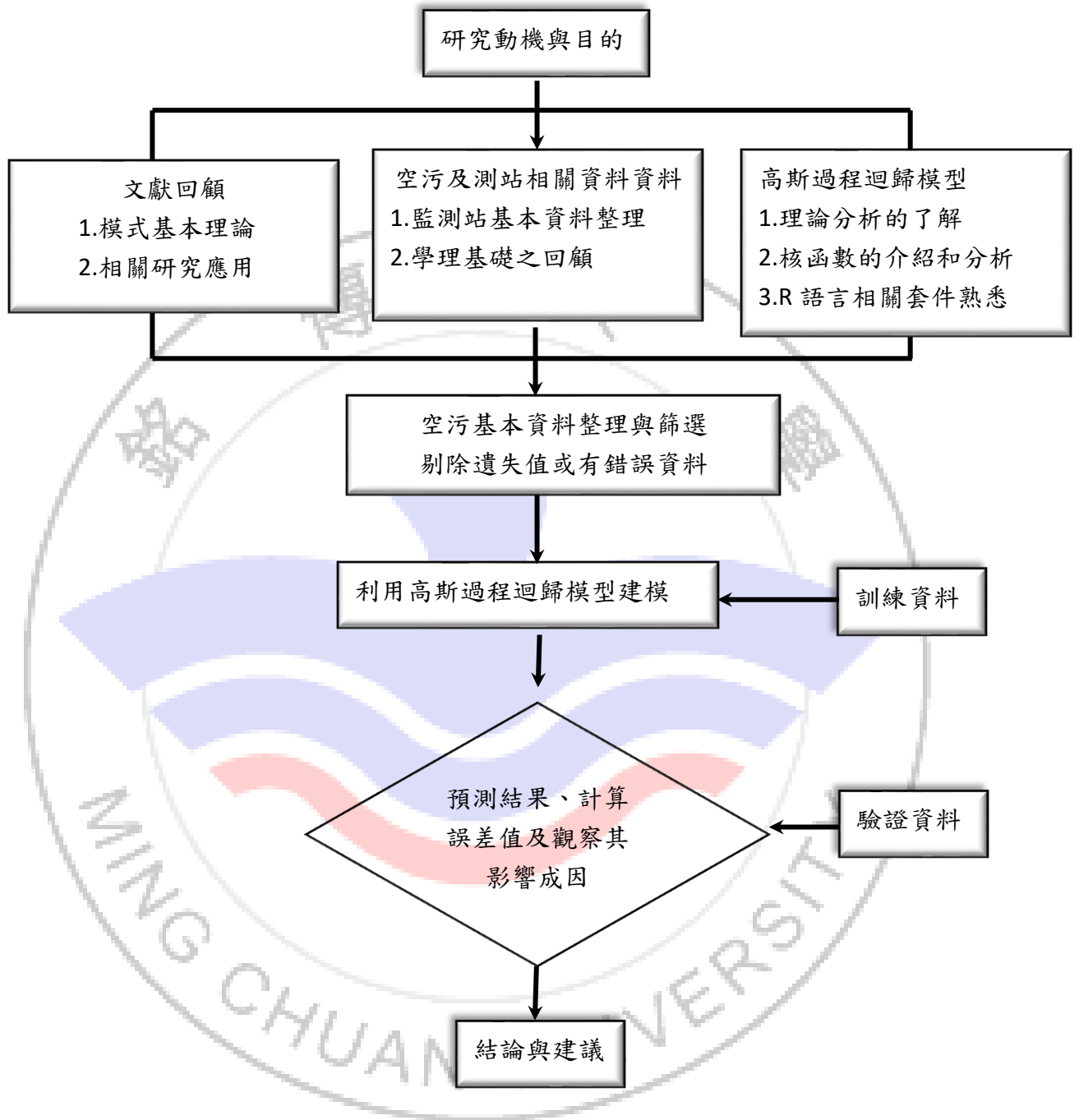


圖 1-5-1 研究步驟流程圖



## 第二章 參考文獻

在這一章中，我們將先介紹國內關於空氣品質監測的相關資訊，再來針對 PM2.5 做詳細的介紹，接著回顧懸浮微粒的相關研究文獻和高斯迴歸過程的相關主題研究，最後介紹 R 語言中適用套件以及關於高斯過程迴歸的相關套件。

### 第一節 空氣品質監測站以及空氣的簡介

為推動空氣品質保護及防制空氣污染，環保署已在全台設立了 77 個空氣品質測站，其中高雄地區有 12 個測站，採取自動化每天 24 小時的監測。監測項目包含直徑小於 10 微米的懸浮微粒(PM10)、細懸浮微粒(PM2.5)，二氧化硫(SO<sub>2</sub>)、一氧化碳(CO)、二氧化氮(NO<sub>2</sub>)、氮氧化物(NO<sub>x</sub>)、臭氧(O<sub>3</sub>)、碳氫化合物(HC)；及相關的氣象變數，如溫度、濕度、酸雨等。並將當日空氣中 PM10、PM2.5、SO<sub>2</sub>、NO<sub>2</sub>、CO 及 O<sub>3</sub> 六大污染物的濃度，以其對人體健康的影響程度，分別換算出副指標值，再以當日各副指標之最大值作為該測站當日的空氣品質指標值(Air Quality Index, AQI)。AQI 依據各空氣污染物對人體健康影響濃度大小，採用 6 等級、6 顏色方式劃分，可以提供民眾清楚表示指標及顏色，可經由環保署空氣品質監測網或環境即時通 APP 查詢即時空氣品質指標訊息，作為日常活動參考指南(表 2-1-1)。PM2.5 以及 PM10 一直都是空氣污染很嚴重的原因，它們都是非常小的懸浮微粒，然而愈小粒徑的懸浮微粒對於人體健康影響愈顯著，亦是民眾所關切，如何減少細懸浮微粒成為當前空氣污染管制政策的重點。由表 2-1-2 可以看得出來細懸浮微粒 PM2.5 直徑較小、生命週期較長及傳送距離較遠。



表 2-1-1 空氣品質指標(AQI)

取自環保署空氣品質監測網(<https://taqm.epa.gov.tw/taqm/tw/b0201.aspx>)

空氣品質指標(AQI)	0~50	51~100	101~150	151~200	201~300	301~500
對健康影響與活動建議	良好	普通	對敏感族群不健康	對所有族群不健康	非常不健康	危害
狀態色塊	綠	黃	橘	紅	紫	褐紅

表 2-1-2 PM2.5 和 PM10 的比較

取自 ScienceDirect

(<https://www.sciencedirect.com/science/article/pii/S0160412014002992>)

PM2.5 和 PM10 的比較	細懸浮微粒 PM2.5	懸浮微粒 PM10
直 徑	小於及等於 2.5 微米	小於及等於 10 微米
生命週期	數天至數週	數分鐘至數小時
傳送距離(公里)	100 至 1000	1 至 10

## 第二節 PM2.5 相關介紹

PM2.5 以及 PM10 一直都是空氣污染很嚴重的原因，它們都是非常小的懸浮微粒。其中 PM2.5 的大小約為頭髮直徑的 1/28，體積小、質量輕這樣的特性使它容易在大氣中的停留時間長，能輸送距離較遠，這樣小的粒子會穿透肺泡，跟著血液循環全身。依照生成性質分類，PM2.5 能夠分成原生性及衍生性兩大類，原生性係指被直接排放到大氣的粒狀物，例如火山爆發、海水飛沫、沙塵暴、營建工地粉層及汽機車排放不完全燃燒之氣體等原因；衍生性是排放至大氣中的化學物質經過物理反應（如凝結）或化學反應（如光化學反應）等二次形成約微米大小的粒狀物，例如硫酸鹽，主要來源為化石燃料的燃燒產生硫氧化物（SO<sub>2</sub> 與 SO<sub>3</sub>，合稱為 SOX），經由氣相或液相反應而成，而這些污染物不僅僅只有在戶外，在室內，塵蟎及二手菸是懸浮微粒最主要的來源。根

據張艮輝(2017)在強化空氣品質模式制度建立計畫中提到，臺灣 2010 年 PM2.5 主要來源為，自身排放佔 66%，境外影響佔了 34%，其中中國大陸就佔 30%，其餘為其他周邊國家及海鹽影響(如圖 2-2-1)，而國內排放 PM2.5 前三高的產業分別為鋼鐵基本工業、電力業及化學材料製造業。高雄是臺灣石化發展的火車頭，1946 年中油接收日本海軍第六燃料廠，重新整頓為高雄煉油廠，石化相關產業陸續進駐，從此高雄被重工業盤據，也註定了空氣污染的命運。在高雄市，PM2.5 前三大來源，分別為鋼鐵工業、車輛揚塵和柴油車排放，其中又以工業排放的 PM2.5，含有比較多有害物質。現今三座化石燃料發電廠，三個大型石化專區，兩百萬輛的機車，眾多的煙囪在高雄繼續排放污染物，而空氣汙染對人體的影響不僅僅在於與呼吸系統相關的肺炎及氣喘，更有研究發現更多相關的疾病都受到空氣汙染的影響，像是環境中空氣汙染物濃度與心血管疾病住院人數存在相關性(Dominici et al.,2006)、空氣汙染將引起凝血功能改變(Seaton et al., 1995)、自主神經系統對心臟控制的改變(Devlin et al., 2003)以及(Pope et al., 2002) 研究指出 PM2.5 會導致動脈斑塊沉積，引發血管炎症，最終導致心臟病或其他心血管問題。從上述研究可得知空氣汙染還會造成心血管疾病死亡率與罹病率的增加。

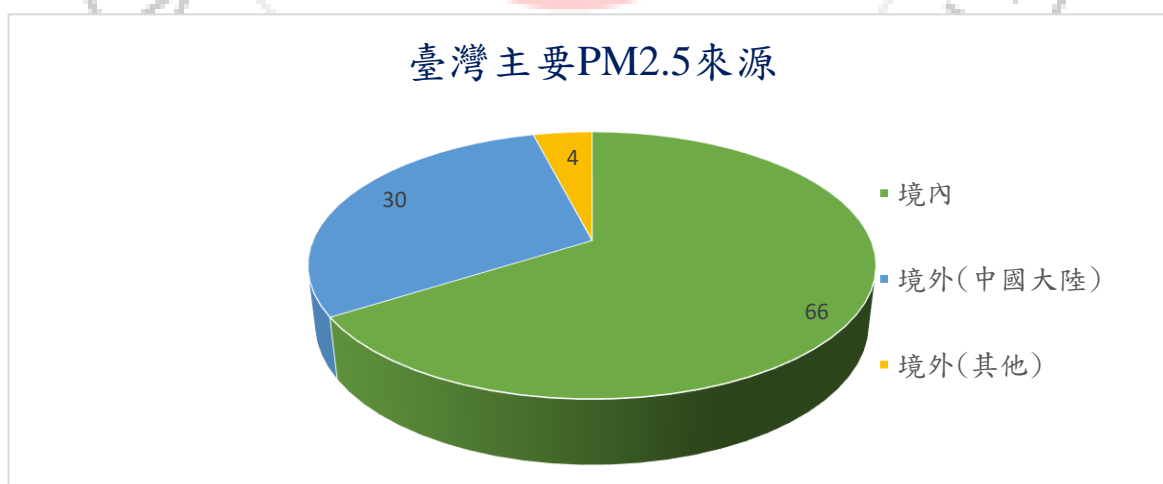


圖 2-2-1 臺灣 PM2.5 主要來源

### 第三節 空氣污染物的相關研究

對於預測空氣污染的研究，近年有許多研究者使用各種方式、各種範圍及各種污染物做研究，而最熱門的話題不外乎就是懸浮微粒，傳統使用複線性迴歸研究相關變數(如氣象變數及相關污染物濃度)對於懸浮微粒濃度的影響。陳奕愷(2012)、黃翊涵(2014)都是以複線性迴歸分析空氣品質資料。陳奕愷(2012)使用鄰近濁水溪之七個空氣品質監測站的資料，以多種方法分析，再分別建立預測模式，並探討污染物與各種因子之關係；黃翊涵(2014)研究 2009 年至 2013 年中部 11 座空氣品質自動監測站的各污染物，討論污染物與氣象因子時空變化趨勢以及相關性。研究結果顯示出 Pearson 檢定發生懸浮微粒(PM10)不良日時，主要顯著相關因子為 PM2.5 與 SO<sub>2</sub>。

傅文閣(2018)利用高雄市一般空氣品質監測站資料，分析 PM2.5 濃度分布及其變化，並建立複線性迴歸模型，用以預測次日 PM2.5 濃度日均值。利用 2016 年資料作為測試資料，驗證模型性能，結果顯示春季模型預測性能最佳；就預測濃度區間而言，夏季模型的預測準確率最高，達 98%以上，最低為冬季，準確率僅達 50.0~64.4%。

俞淑惠(2014)考慮時間季節的因素，以廣義自迴歸條件異方差模型 (generalized autoregressive conditional heteroskedastic model, GARCH)，研究相關變數對於細懸浮微粒濃度的影響；在其分析結果中，不同測站的顯著影響變數及模型均不相同。陳明輝、周錦慧、周佳穎(2016)認為相關變數對於懸浮微粒的影響複雜，或許不只是線性影響方式，以桃園地區的測站為研究對象，應用廣義加法模型研究懸浮微粒的相關因素，以了解其非線性影響方式。

陳騰瑞(2013)針對 2006 年至 2010 年雲嘉南空品區的主要污染物作分析，使用 Matlab 倒傳遞類神經網路預測空氣品質，再為 PM10 及 O<sub>3</sub> 做預測，研究結果用 PSI 指標看來將懸浮微粒分為三級，作為預報之等級。測試結果而言，PSI 指標預報平均正確率為 88.25%，其中以嘉義測站冬季網路預報最為準確，

PSI 指標預報正確率高達 92.82%。

機器學習近年來已廣泛應用在資料的分析與預測。簡家宏(2004)以台中沙鹿站的 PM10 濃度為預測對象，利用基因演算法(genetic algorithm)加強類神經網路分析法(neural network analysis)，使得在求解過程中能全域搜尋避免落入局部最佳解。林冠名(2018)開發三種機器學習方法：決策樹回歸(decision tree regression)、梯度增強樹回歸(gradient boosting regression tree)和支持向量回歸(support vector regression)，建立一個 8 小時的空氣品質預測模型，即時預測 PM2.5 的濃度。Rasmussen and Williams(2006)一書中提到，類神經網路分析法著重在預測，對於整個輸入輸出之間的關係較無著墨；而高斯過程迴歸是機器學習領域的一種無母數貝氏迴歸分析方法，可以不須事先假設明確的迴歸模型，藉由訓練資料建立無母數模型，以預測新資料的輸出，而且可以求得該輸出的不確定性。

#### 第四節 高斯迴歸過程相關應用

陳亭蓉(2006)採用貝氏分析配合高斯過程迴歸模型來估算降雨量，並使用 Leave One Out(LOO)法求出單一颱風模型之估算誤差，以檢核該模型之適用性，並建立一個當颱風來受到地形強迫產生的降雨結構，可以推估降雨空間及區域降雨量。安靈雅(2014)採用高斯過程迴歸的預測模型，給定傳感器讀數數據，找到一個合適的核函數的過程。提出了一個各向異性核函數的高斯過程，對傳感器讀數造型空間和時間之間的關係平衡，以天氣數據為例進行，實驗結果證明，高斯過程可以給我們很好的歸集。蔡効耕(2015)針對美國加州高速公路資料，利用高斯過程為主要的方法來預測未來交通狀況，並且探討當前車況與歷史車況中的時間關係以及不同路段交通的關係來改善預測結果。

Guizilini and Ramos (2015)使用高斯過程迴歸預測澳洲地區不同時間及空間的 PM2.5 濃度；在此研究中，作者使用了能夠兼顧時間及空間效應的共變異數

矩陣，並提出一套系統訓練及更新模型。

## 第五節 高斯過程之套件

高斯過程迴歸需要大量計算，我們運用高斯過程專業網站 [www.gaussianprocess.org](http://www.gaussianprocess.org) 列出了目前常用的數個軟體及套件(表 2-5-1)，主要是用 Python、matlab 或 R 寫成，這裡面的套件都可以成為我們往後的工具。

而高斯過程其中列出使用 R 語言寫成的套件有 Gramacy and Taddy (2007) 的 tgp 套件、Kalaitzis, Honkela, Gao and Lawrence (2014) 的 gptk 套件，以上兩套件都是高斯過程的工具。另外，MacDonald, Ranjan and Chipman (2015) 提出以基因演算法為基礎的高斯過程套件 GPfit，宣稱計算結果較為穩健並且速度較快。

Tgp 套件是由貝式無母數和非平穩迴歸的高斯過程轉換到限制線性模型 (LLM)，特別的例子有貝葉斯線性迴歸(Bayesian linear models)、決策樹(linear CART)、可分離(stationary separable)以及高斯過程迴歸。gptk 套件使用高斯徑向基函數核(Radial basis function kernel)可當作高斯過程迴歸的工具。GPfit 將高斯過程模型合到確定性模擬器輸出的穩定方法，假設輸入變量是連續的，輸出是從標量質確定性計算模擬器中獲得的。

套件	標題	作者	適用語言
bcm	The Bayesian Committee Machine	Anton Schwaighofer	matlab and NETLAB
fbm	Software for Flexible Bayesian Modeling	Radford M. Neal	C for linux/unix
gp-lvm and fgp-lvm	A (fast) implementation of Gaussian Process Latent Variable Models	Neil D. Lawrence	matlab and C
gpml	Code from the Rasmussen and Williams: Gaussian Processes for Machine Learning book.	Carl Edward Rasmussen and Hannes Nickisch	matlab and octave
c++-ivm	Sparse approximations based on the Informative Vector Machine	Neil D. Lawrence	C++
BFD	Bayesian Fisher's Discriminant software	Tonatiuh Peña Centeno	matlab
gpor	Gaussian Processes for Ordinal Regression	Wei Chu	C for linux/unix
MCMCstuff	MCMC Methods for MLP and GP and Stuff	Aki Vehtari	matlab and C
spgp .tgz or .zip	Sparse Pseudo-input Gaussian Processes	Ed Snelson	matlab
tgp	Treed Gaussian Processes	Robert B. Gramacy	C/C++ for R

Tpros	Gaussian Process Regression	David MacKay and Mark Gibbs	C
GP Demo	Octave demonstration of Gaussian process interpolation	David MacKay	octave
GPClass	Matlab code for Gaussian Process Classification	David Barber and C. K. I. Williams	matlab
VBGP	Variational Bayesian Multinomial Probit Regression with Gaussian Process Priors	Mark Girolami and Simon Rogers	matlab
pyGPs	Gaussian Processes for Regression and Classification	Marion Neumann	Python
gaussian-process	Gaussian process regression	Anand Patil	Python
gptk	Gaussian Process Tool-Kit	Alfredo Kalaitzis	R

表 2-5-1 高斯過程回歸套件 取自 The Gaussian Processes Web Site ([www.gaussianprocess.org](http://www.gaussianprocess.org))



### 第三章 研究方法及套件介紹

本章為研究方法，共分為三小節。第一節為介紹高斯回歸過程，第二節介紹其中常用的共變異數函數，第三節介紹我們執行高斯回歸過程所使用的 R 套件。

#### 第一節 介紹高斯回歸過程

Guizilini and Ramos (2015) 指出氣象變數及相關污染物濃度對於懸浮微粒濃度的影響方式是相當複雜，且可能存在化學交互作用，並不容易事先假設模型的樣態。以往的研究通常針對某些測站資料建模，較少將空間因素(經度、緯度)考慮在內，若該地區沒有測站資料，將無法預估其懸浮微粒濃度。作者使用高斯過程迴歸預測澳洲地區不同時間及空間的 PM10 及 PM2.5 時，在類似的考慮下，結合週期性核函數作為其最後的共變數矩陣算法。在此研究中，使用了能夠兼顧時間及空間效應的共變異數矩陣，並提出一套系統訓練及更新模型。

本研究目的為用過去的資料來預測未來的變化，在空汙資料方面，自變項  $x$  可以是經度、緯度、時間、或是其他的氣象變數或污染物，而我們的因變項  $y$  就是 PM2.5 濃度。假設我們有一組訓練資料  $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$ ，其中  $y_i$  表示第  $i$  個觀察值的因變項(或稱輸出變項)， $x_i = (1, x_{i1}, \dots, x_{ip})^T$  表示第  $i$  個觀察值的自變項(或稱輸入變項)向量。 $X_n$  為  $d \times n$  的矩陣，稱為設計矩陣。 $y_n$  為向量，是因變項資料的集合，整理起來可以寫成  $D = (X_n, y_n)$ ，然後我們利用手上現有的資料，來推估未知的值，並且推估資料的變異程度。以符號來講就是給定觀察到的資料  $X_{n1}$ 、 $y_{n1}$ ，當自變項為  $X_{n2}$  時的預測值  $y_{n2}$  的條件分布，也就是  $p(y_{n2} | y_{n1}, X_{n1})$ 。

$$y = \begin{bmatrix} y_{n1} \\ y_{n2} \end{bmatrix}, \mu = \begin{bmatrix} \mu_{n1} \\ \mu_{n2} \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$



則

$$y_{n1} \sim N(\mu_{n1}, \Sigma_{11})$$

且

$$y_{n2}|y_{n1}, X_{n1}, X_{n2} \sim N(\mu_{n2} + \Sigma_{21} \Sigma_{11}^{-1}(y_{n1} - \mu_{n1}), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})$$

如果我們假設此高斯過程的期望值為 0，則為

$$y_{n2}|y_{n1}, X_{n1}, X_{n2} \sim N(\Sigma_{21} \Sigma_{11}^{-1} y_{n1}, \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})$$

接下來還有共變異數和超參數為未知，而 GPfit 內的公式共變異數的寫法為

$$R_{ij} = \prod_{k=1}^d \exp(-10^{\beta k} |x_{ik} - x_{jk}|^{power})$$

在(式)當中，有  $\beta$  和  $power$  為超參數， $power$  內定值為 1.95，但它是可以自己設定的，只要介於 0 到 1 之間即可； $\beta$  我們還尚未知道它的功用。

假設觀察資料  $D_{n1} = (X_{n1}, y_{n1})$ ，測試資料  $D_{n2} = (X_{n2}, y_{n2})$ ，在高斯過程中，我們假設隨機變數的聯合分配均為聯合高斯分配。

$$y = \begin{bmatrix} y_{n1} \\ y_{n2} \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_{n1} \\ \mu_{n2} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

透過公式(3-3-3)的觀念，我們一樣可以把公式帶入

$$y_{n2}|y_{n1}, X_{n1}, X_{n2} \sim N(\mu_{n2} + \Sigma_{21} \Sigma_{11}^{-1}(y_{n1} - \mu_{n1}), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})$$

假設此期望值為 0，則測試資料的事後分配為

$$y_{n2}|y_{n1}, X_{n1}, X_{n2} \sim N(\Sigma_{21} \Sigma_{11}^{-1} y_{n1}, \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})$$

總結前幾個小節，任意數個隨機變數的聯合分配均為聯合高斯分配。若  $y_{n1}$  為觀察資料， $y_{n2}$  為測試資料，SE 函數  $(K = \{k(x_i, x_j)\})$  來為其共變異數，則其聯合機率分配為

$$y = \begin{bmatrix} y_{n1} \\ y_{n2} \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_{n1} \\ \mu_{n2} \end{bmatrix}, \begin{bmatrix} K(X_{n1}) & K(X_{n1}, X_{n2}) \\ K(X_{n2}, X_{n1}) & K(X_{n2}) \end{bmatrix} \right)$$

考慮到觀察資料的量測有誤差，則其聯合機率分配為

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_{n1} \\ \mathbf{y}_{n2} \end{bmatrix} \sim N \left( \begin{bmatrix} \boldsymbol{\mu}_{n1} \\ \boldsymbol{\mu}_{n2} \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X}_{n1}) + \sigma_n^2 \mathbf{I}_{n1} & \mathbf{K}(\mathbf{X}_{n1}, \mathbf{X}_{n2}) \\ \mathbf{K}(\mathbf{X}_{n2}, \mathbf{X}_{n1}) & \mathbf{K}(\mathbf{X}_{n2}) \end{bmatrix} \right)$$

考慮到觀察資料的量測有誤差，則測試資料 $\mathbf{y}_{n2}$ 的條件分配為底下式

$$\mathbf{y}_{n2} | \mathbf{X}_{n1}, \mathbf{X}_{n2}, \mathbf{y}_{n1} \sim N(E(\mathbf{y}_{n2} | \mathbf{X}_{n1}, \mathbf{X}_{n2}, \mathbf{y}_{n1}), \text{cov}(\mathbf{y}_{n2} | \mathbf{X}_{n1}, \mathbf{X}_{n2}, \mathbf{y}_{n1}))$$

$$(E(\mathbf{y}_{n2} | \mathbf{X}_{n1}, \mathbf{X}_{n2}, \mathbf{y}_{n1}) = \boldsymbol{\mu}_{n2} + \mathbf{K}(\mathbf{X}_{n1}, \mathbf{X}_{n2}) [\mathbf{K}(\mathbf{X}_{n1}) + \sigma_n^2 \mathbf{I}_{n1}]^{-1} (\mathbf{y}_{n1} - \boldsymbol{\mu}_{n1})$$

$$(\text{cov}(\mathbf{y}_{n2} | \mathbf{X}_{n1}, \mathbf{X}_{n2}, \mathbf{y}_{n1})))$$

$$= \mathbf{K}(\mathbf{X}_{n2} - \mathbf{K}(\mathbf{X}_{n1}, \mathbf{X}_{n2}) [\mathbf{K}(\mathbf{X}_{n1}) + \sigma_n^2 \mathbf{I}_{n1}]^{-1} \mathbf{K}(\mathbf{X}_{n1}, \mathbf{X}_{n2}))$$

如果假設此過程為一平穩過程，即平均數為 0 的情況下，則 $\mathbf{y}_{n2}$ 的事後機率分配的期望值為

$$E(\mathbf{y}_{n2} | \mathbf{X}_{n1}, \mathbf{X}_{n2}, \mathbf{y}_{n1}) = \mathbf{K}(\mathbf{X}_{n2}, \mathbf{X}_{n1}) [\mathbf{K}(\mathbf{X}_{n1}) + \sigma_n^2 \mathbf{I}_{n1}]^{-1} \mathbf{y}_{n1}$$

但我們還是不知道超參數的值，只能實際代入才可能知道結果。

## 第二節 共變異函數

再來共變異函數，一般的共變異數可以由觀察值估計出來，但高斯過程的假設並不同，首先，我們必須明白一個觀念，當越相近的距離，它們的關係一定也比較強，接下來介紹幾個有這樣性質的函數。

### (一)、歐式函數(Euclidean Distance)

歐氏距離是一種距離計算方法，源自歐氏空間中兩點間的距離公式。在歐幾里得空間中

(1) 二維平面上兩點  $a(\mathbf{X}_1, \mathbf{y}_1)$  與  $b(\mathbf{X}_1, \mathbf{y}_2)$  間的歐氏距離：

$$d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

(2) 三維空間兩點  $a(\mathbf{X}_1, \mathbf{y}_1, \mathbf{Z}_1)$  與  $b(\mathbf{X}_2, \mathbf{y}_2, \mathbf{Z}_2)$  間的歐氏距離：

$$d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

(3) 兩個  $n$  為向量， $a(x_1, x_2 \dots, x_n)$  與  $b(y_1, y_2 \dots y_n)$  間的歐氏距離：

$$d_{12} = \sum_{k=1}^n |x_{1k} - x_{2k}|$$

向量  $\vec{x}$  的自然長度，為一個純數值，就是該點到原點地距為

$$\|\vec{x}\|^2 = \sqrt{|x_1|^2 + \dots + |x_n|^2}$$

針對歐式函數的缺點再改進成標準化歐氏距離 (Standardized Euclidean distance)，因為各分量分布不一樣，標準化歐氏距離先將各分量都「標準化」到均值、方差相等。假設  $X$  的均值為  $m$ ，標準差為  $s$ ， $X$  的「標準化變量」表示為：

$$X^* = \frac{X - m}{s}$$

標準化歐式距離公式為

$$d_{12} = \sqrt[p]{\sum_{k=1}^n |x_{1k} - x_{2k}|^p}$$

取自 MBA 智庫 (<https://wiki.mbalib.com/zh-tw/%E6%AC%A7%E6%B0%8F%E8%B7%9D%E7%A6%BB>)

(二)、平方指數函數(squared exponential function, SE)

SE 函數如下：

$$k(x_p, x_q^T) = \exp\left(-\frac{1}{2}|x_p - x_q|^2\right)$$

在使用時，會引進參數，而改寫成

$$k(x_p, x_q^T) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}|x_p - x_q|^2\right)$$

考慮到資料的測量誤差，那再加上誤差變異，共變異數矩陣如下：

$$\begin{aligned} cov(x_p, x_q) &= k(x_p, x_q) + \sigma_n^2 \delta_{pq} \\ &= \sigma_f^2 \exp\left(-\frac{1}{2l^2}(x_p - x_q)^2\right) + \sigma_n^2 \delta_{pq} \end{aligned}$$

$$\sigma_n^2 \delta_{pq} = \begin{cases} 1 & p = q \\ 0 & p \neq q \end{cases}$$

而其中有三個參數，length-scale( $l$ )、signal variance( $\sigma_f^2$ )和 noise variance( $\sigma_n^2$ )，我們稱它們為超參數(hyperparameters)，但我們只知道 $\sigma_n^2$ 為誤差變異，其他還要再測試才會知道功用。

(三)、有理二次核函數(Rational Quadratic Kernel, RQ kernel)

$$k_{RQ}(\mathbf{r}) = \left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha}, \alpha, l > 0$$

上式 $\alpha, l$ 為超參數。

(四)、週期性核函數(Periodic Kernel)

平穩核函數可以用於構建週期核函數  $K = k\left[\sin\left(\frac{\pi}{p}r\right)\right]$ ，其中  $p$  表示該核函數具有週期，而後整理出：

$$k_{PR}(\mathbf{a}, \mathbf{b}) = \sigma_f^2 \exp\left(-\frac{2\sin^2(\|\mathbf{a}-\mathbf{b}\|)}{l^2}\right)$$

上述 $\sigma_f^2, l^2$ 為我們的超參數。

以上對於(三)、(四) GP 模型來說，kernel 的形式都是事先選擇好的，kernel 中的超參數都是不會事先給定的，而是需要通過具體的學習方法自動學習而得。

取自知乎(<https://zhuanlan.zhihu.com/p/35396322>)

### 第三節 套件介紹

此研究主要利用 R 套件中的 GPfit 來做為執行高斯回歸過程的工具。GPfit 能運用輸入的訓練資料建立一模型，在幫你用觀測資料去估計出相對應的數值，也可以用圖呈現。先介紹我們 GPfit 主要參數(表 3-5-1)

表 3-5-1 GPfit 主要參數介紹

X	為一個( $n \times d$ )的矩陣
Y	為輸出( $n \times 1$ )的矩陣
corr	為共變異數 $Cov(Z(x_i), Z(x_j))$

GP model:  $y(x) = \mu + Z(x), x \in [0,1]^d$ ，當 $Z(x)$ 是 GP，平均數為 0，而  $Var(Z(x_i)) = \sigma^2$ 、 $Cov(Z(x_i), Z(x_j)) = \sigma^2 R_{ij}$ ，此外還有  $R_{\delta_{lb}} = R + \delta_{lb}$ ，其中  $\delta_{lb}$  是代表加一個數可以讓模型更穩定，是代表加一個數可以讓模型更穩定，而 0 是最好的。統整後我們可以得知共變異數  $\sigma^2 = \frac{Cov(Z(x_i), Z(x_j))}{R}$

為了能夠更加清楚知道函數的成果，我們先使用 R 裡面附的例子(圖 3-5-1) 來做說明，GPfit 能夠幫我把模型建好，下面的例子我們先給定五組數據去建立模型，之後利用 plot 可以幫我們把圖畫出來。圖 3-5-2 中黑色的點就是我們給定的資料，實線是它幫我預測的樣子，虛線是誤差範圍，而當虛線離黑色點距離越近，表示誤差值越小。

```

1  ## 1D Example 1
2  n = 5
3  d = 1
4  computer_simulator <- function(x){
5    x = 2 * x + 0.5
6    y = sin(10 * pi * x) / (2 * x) + (x - 1)^4
7    return(y)}
8  }
9  set.seed(3)
10 library(lhs)
11 x = maximinLHS(n, d)
12 y = computer_simulator(x)
13 GPmodel = GPfit(x, y)
14 print(GPmodel)

```

圖 3-5-1 GPfit 例子

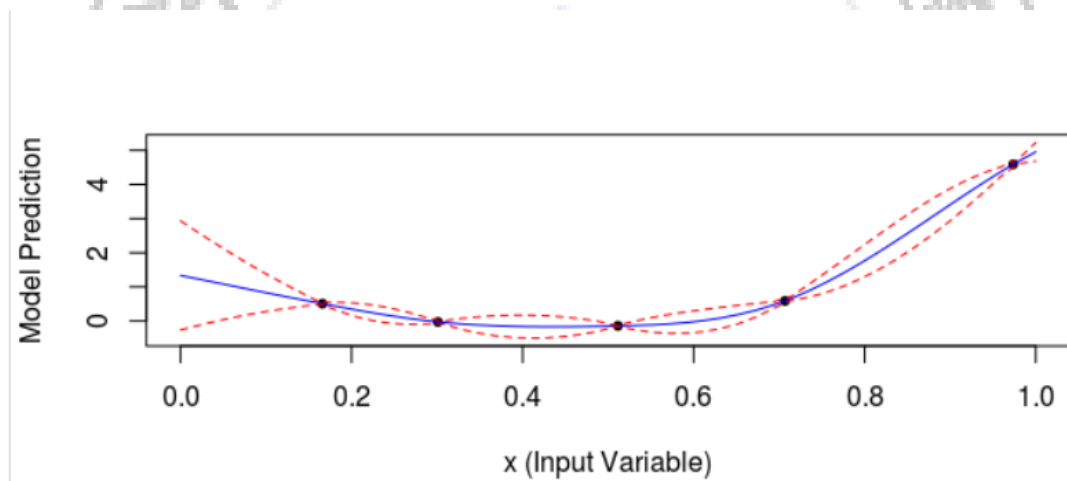


圖 3-5-2 GPfit 例子的結果圖

## 第四章 資料分析

第四章為資料分析，第一小節是整理與合併資料，第二小節以測站、月分和時段分類來統整分析資料，並透過圖表呈現結果，第三小節使用高斯迴歸過程，針對 2017 年的資料來建模預測；過程中，我們依照單變量、雙變量及多變量分別建模。

### 第一節 資料整理

這一章節為資料分析，4.1.1 介紹我們研究的測站及監測項目，運用 R 繪圖、點出高雄地區所有測站的相對位置，並說明各測站的性質和功能。再來 4.1.2 透過環保署所提供的空污資料附件，我們適當地做整理、轉換，以方便我們使用需要。

#### 4.1.1 介紹監測站、位置及監測項目

根據環保署空氣品質監測網的解釋，監測空氣品質資料是為了推動空氣品質保護及防制空氣污染工作。監測站的運作，從空間配置、儀器種類、維護及品質保護都要做好，才能期待獲得良好的監測數據。

行政院環境保護署建立監測站的目的，為以下取自環保署空氣品質監測網 (<https://taqm.epa.gov.tw/taqm/tw/b0106.aspx>)：

- [1] 法規的符合性。
- [2] 監測空氣污染物的長期趨勢。
- [3] 評量管制策略的有效性。
- [4] 研判及評估對民眾健康的威脅及損害。
- [5] 了解及評估對自然環境的威脅及損害。
- [6] 發展及驗證具預測能力的空品擴散模式。

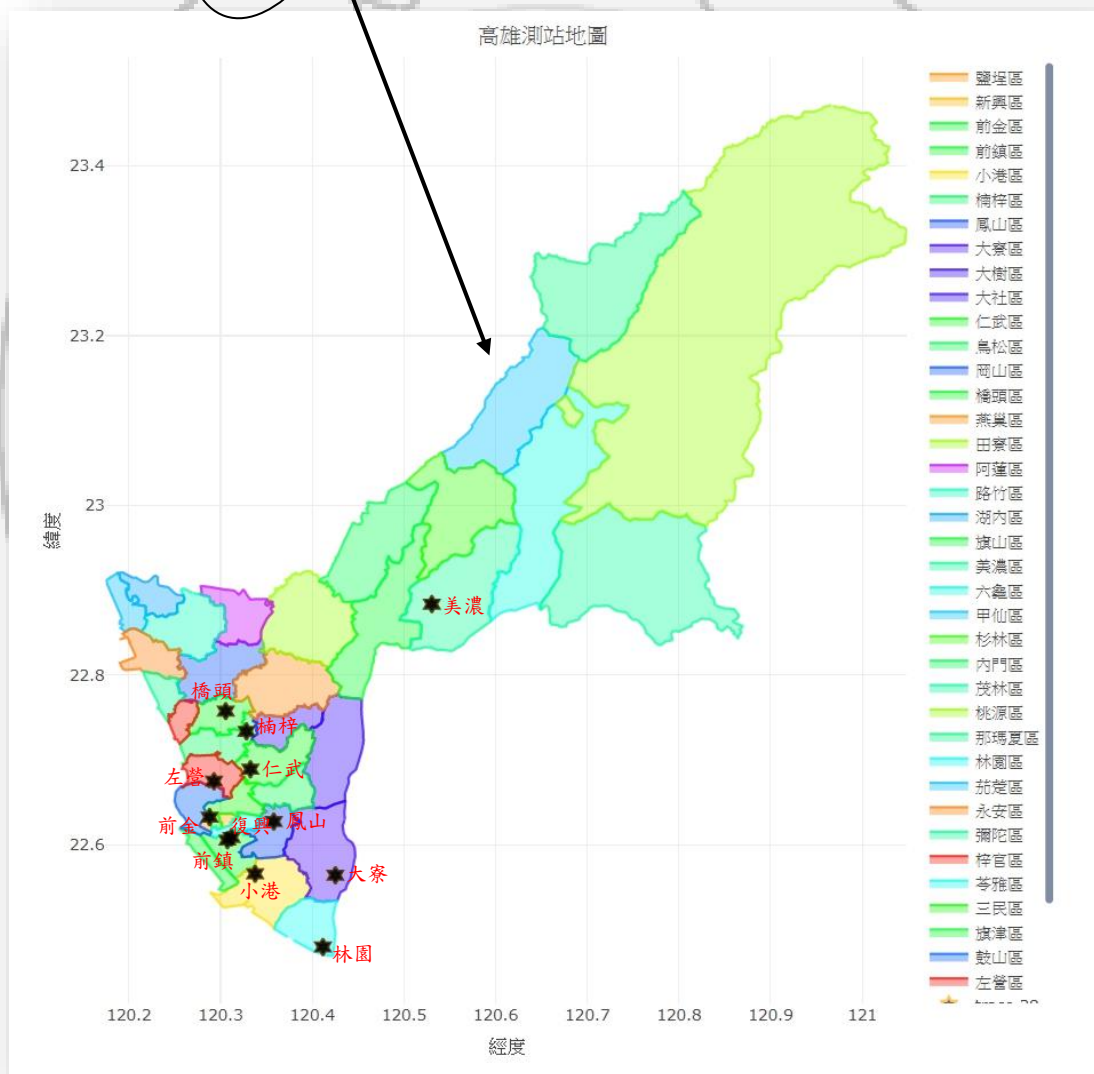
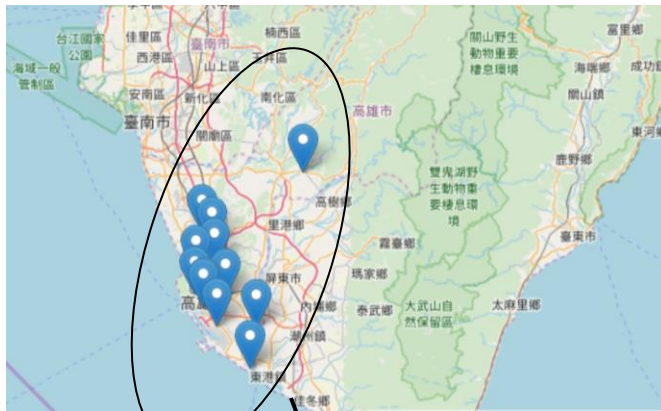
考量臺灣地形分布與污染源分布特性，將臺灣分為 7 個空氣品質預報區，包括北部、竹苗、中部、雲嘉南、高屏、宜蘭、花東，總共有 77 個監測站。而

監測站也分為不同的的類型，一般測站通常設置在人口集或可能發生高污染的地區；交通測站設置於交通量頻繁的地區，可評估車輛管制及廢氣污染狀況；工業測站設置於工業區之盛行風下風區，以了解工業污染之影響；背景測站設置於人口較少的地方或總量管制區盛行風上風區，遠離其他污染源可以判斷有無臺灣以外傳來的污染物。

本研究的高雄市屬於高屏空品區，目前高雄地區共有 12 個測站（圖 4-1-1），其中有八個一般測站（美濃、楠梓、仁武、左營、前金、小港、林園、大寮），兩個交通測站（鳳山、復興），一個工業測站（前鎮），一個背景測站（橋頭）。







4-1-1 高雄測站位置圖

空氣品質監測項目有分為氣象測項及污染物，氣象測項包含溫度、降雨、風速及風向等，空氣污染物有相當多的測項(表 4-1-1)，共有 21 個不同的測項名稱，雖然有這麼多測項，但卻不是所有監測站的測項都完全相同。

表 4-1-1 污染物測英文簡稱、單位及測項名稱)

測項簡稱	單位	測項名稱
SO <sub>2</sub>	ppb	二氧化硫
CO	ppm	一氧化碳
O <sub>3</sub>	ppb	臭氧
PM <sub>10</sub>	µg/m <sup>3</sup>	懸浮微粒
PM <sub>2.5</sub>	µg/m <sup>3</sup>	細懸浮微粒
NO <sub>x</sub>	ppb	氮氧化物
NO	ppb	一氧化氮
NO <sub>2</sub>	ppb	二氧化氮
THC	ppm	總碳氫化合物
NMHC	ppm	非甲烷碳氫化合物
CH <sub>4</sub>	ppm	甲烷
UVB	UVI	紫外線指數
AMB_TEMP	°C	大氣溫度
RAINFALL	mm	雨量
RH	%	相對溼度
WIND_SPEED	m/sec	風速(以每小時最後 10 分鐘算術平均)
WIND_DIREC	degress	風向(以每小時最後 10 分鐘向量平均)
WS_HR	m/sec	風速小時值(以整個小時算術平均)
WD_HR	degress	風向小時值(以整個小時向量平均)
PH_RAIN	pH	酸鹼值(酸雨)
RAIN_COND	µS/cm	導電度(酸雨)

### 4.1.2 資料前處理

本研究資料於行政院環保署空氣品質監測網所下載的歷年資料(<https://taqm.epa.gov.tw/taqm/tw/YearlyDataDownload.aspx>)，資料原始檔案如(圖 4-1-2)所示，以 2017 年大寮站為例，第一欄日期(2017/01/01)，第二欄測站(大寮)，第三欄為測項(AMB\_TEMP、CH4、CO...)，從第四欄開始為 0-23 監測時間。另外，資料內有許多不同的符號，可能是有儀器異常或是人為等問題，其中包括「#」表示儀器檢視為無效值、「\*」表示程式檢視為無效值、「x」表示人工檢視為無效值、「NR」表示無降雨、空白表示缺值，分為「"」和「"」。

當把資料匯入到 R 的時候，發現到一些遺失值或是一些標點符號(#、\\\*、-、x)，針對這些特殊符號我們都將它轉換成「NA」，也把「NR」轉換成「0」，整理好資料後就可以做資料轉置的工作，第一筆資料就為仁武站第一小時的各項測項，第二筆為仁武站第二小時，以此類推，而第一欄為日期，第二欄為測站，第三欄為小時(1-24)，之後就是各項污染物，這樣可以讓資料更清楚的呈現，最後再所有測站的資料合併起來。在合併的時後卻發現並不是每個測站的污染物都相同，而且順序也不一樣，舉例來說，正常應該要有 21 個測項，但仁武站的只有 18 個，缺少了 PH\_RAIN、RAIN\_COND 和 UVB 這三項，美濃站缺少 CH4、NMHC、PH\_RAIN、RAIN\_COND、THC 和 UVB 足足缺了 6 項，而這 12 個測站就分別有 12 個不一樣的形式，這讓我們在合併的時候就會有些困難，必須將這問題考慮進去，因此我們把測站中缺少的項目補上，並把裡面的數值都設定為 NA 值，由此一來我們即可方便的合併。還有我們也把 year、mon、wkday、long 和 lat 也放上去，分別就是年分、月份、星期(0-6)、緯度和經度，所以我們透過 rbind 這個函數，它可以幫我們解決這個問題，最後可以合併出我們整理過全部測站的完整資料(圖 4-1-3)。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	日期	測站	測項	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
2	2017/01/01	大寮	AMB_TEM	20	20	19	18	18	18	18	18	20	22	24	25	26	26	25	25	24	
3	2017/01/01	大寮	CH4	2.3	2.3	2.6	2.3	2.3	2.2	2.3	2.4	2.2	2.1	2	2	2	2	1.9	2	2	
4	2017/01/01	大寮	CO	0.58	0.76	0.75	0.65	0.67	0.7	0.8	0.97	0.7	0.59	0.57	0.71	0.65	0.43	0.32	0.34	0.38	
5	2017/01/01	大寮	NMHC	0.27	0.29	0.64	0.19	0.18	0.23	0.32	0.42	0.21	0.12	0.09	0.09	0.17	0.07	0.06	0.07	0.09	
6	2017/01/01	大寮	NO	3.6	5.3	11	1.1	0.7	0.4	1.6	13	5.6	2.8	2.1	2.2	1.9	2	1.3	1.3	1.1	
7	2017/01/01	大寮	NO2	26	28	29	19	17	18	24	29	19	14	14	18	22	18	15	19	22	
8	2017/01/01	大寮	NOx	30	33	40	20	17	19	26	42	25	17	17	20	24	20	17	20	23	
9	2017/01/01	大寮	O3	4.8	1.4	1.3	8.6	11	14	7.4	4.2	16	35	54	67	71	61	66	64	58	
10	2017/01/01	大寮	PM10	96	107	102	98	85	106	111	120	95	94	81	74	69	67	51	55	65	
11	2017/01/01	大寮	PM2.5	43	48	53	56	53	51	58	64	60	53	45	44	42	33	25	18	21	
12	2017/01/01	大寮	RAINFAL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	
13	2017/01/01	大寮	RH	89	90	91	93	93	92	92	92	88	77	68	62	62	61	65	67	71	
14	2017/01/01	大寮	SO2	8.4	3.1	5.3	2.6	2.5	3.3	3.3	4.1	3.9	5.8	7.7	9.1	12	8.2	4.7	4.6	4.6	
15	2017/01/01	大寮	THC	2.6	2.6	3.2	2.5	2.5	2.5	2.7	2.8	2.4	2.2	2.1	2.1	2.2	2	2	2	2	
16	2017/01/01	大寮	WD_HR	71	326	28	303	309	330	322	312	319	322	329	230	205	277	283	276	280	
17	2017/01/01	大寮	WIND_DIR	66	319	302	325	309	357	2.4	327	322	320	305	206	255	277	292	269	283	
18	2017/01/01	大寮	WIND_SPEED	0.6	0.5	0.6	1.2	0.8	0.7	0.5	1.1	1.3	2	1.2	1.8	2.4	3.5	3.3	2.6	1.3	
19	2017/01/01	大寮	WS_HR	0.7	0.6	0.5	1	0.9	0.8	0.5	0.8	0.9	1.1	0.8	0.6	2	2.9	3.2	2.7	1.9	
20	2017/01/02	大寮	AMB_TEM	20	20	19	19	19	18	18	18	19	20	23	25	26	27	27	26	25	
21	2017/01/02	大寮	CH4	2.1	2.2	2.2	2.2	2.2	2.2	2.2	2.1	2.1	2.1	2	2	2	2	2	2	2	
22	2017/01/02	大寮	CO	0.5	0.5	0.49	0.44	0.39	0.4	0.45	0.53	0.56	0.5	0.47	0.4	0.49	0.49	0.46	0.46	0.59	
23	2017/01/02	大寮	NMHC	0.16	0.18	0.17	0.32	0.23	0.18	0.18	0.28	0.19	0.11	0.11	0.08	0.11	0.1	0.09	0.08	0.13	
24	2017/01/02	大寮	NO	1.1	1.1	0.9	1	0.7	0.5	1.1	6.5	5.8	4.2	3.2	1.2	1.4	1	0.7	0.6	0.8	
25	2017/01/02	大寮	NO2	20	20	19	21	18	15	17	22	21	15	14	11	15	14	15	15	21	
26	2017/01/02	大寮	NOx	21	21	20	22	18	16	18	29	27	19	18	13	16	15	16	16	22	
27	2017/01/02	大寮	O3	19	13	12	5.6	8.7	7.5	5.9	5.4	13	22	38	61	77	84	79	75	68	
28	2017/01/02	大寮	PM10	98	104	99	102	105	92	93	74	111	85	86	67	66	67	70	63	63	
29	2017/01/02	大寮	PM2.5	46	48	49	47	52	48	43	39	42	51	54	45	42	38	32	36	46	
30	2017/01/02	大寮	RAINFAL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	

圖 4-1-2 原始檔案(Excel)

Filter																									🔍
dt	stateName	hr	AMB_TEMP	CH4	CO	NMHC	NO	NO2	NOx	O3	PH_RAIN	PM10	PM2.5	RAINFALL	RAIN_COND	RH	SO2	THC	WD_HR	UVB	WIND_DIREC	WIND_SP			
1	2017-01-01	仁武	1	20	2.2	0.71	0.30	6.8	21.0	28.0	3.3	NA	128	27	0	NA	87	3.3	2.5	86.0	NA	34.0			
2	2017-01-01	仁武	2	20	2.2	0.77	0.29	5.2	22.0	27.0	3.1	NA	139	25	0	NA	87	3.0	2.5	16.0	NA	342.0			
3	2017-01-01	仁武	3	20	2.4	0.91	0.54	20.0	24.0	45.0	2.4	NA	120	25	0	NA	87	3.2	2.9	14.0	NA	64.0			
4	2017-01-01	仁武	4	19	2.4	0.91	0.47	13.0	22.0	35.0	2.7	NA	155	45	0	NA	88	2.3	2.8	338.0	NA	305.0			
5	2017-01-01	仁武	5	19	2.3	0.74	0.40	8.5	23.0	31.0	3.5	NA	152	55	0	NA	88	3.1	2.7	338.0	NA	61.0			
6	2017-01-01	仁武	6	19	2.4	0.69	0.45	6.4	21.0	27.0	3.1	NA	147	54	0	NA	88	2.6	2.8	67.0	NA	76.0			
7	2017-01-01	仁武	7	18	2.3	0.66	0.32	3.5	18.0	21.0	4.8	NA	123	37	0	NA	89	2.4	2.6	63.0	NA	7.0			
8	2017-01-01	仁武	8	19	2.4	0.95	0.50	28.0	23.0	51.0	4.5	NA	115	33	0	NA	89	3.5	2.9	360.0	NA	2.5			
9	2017-01-01	仁武	9	21	2.4	0.96	0.64	17.0	31.0	48.0	17.0	NA	157	48	0	NA	86	6.4	3.0	68.0	NA	65.0			
10	2017-01-01	仁武	10	22	2.0	0.57	0.24	4.0	15.0	19.0	43.0	NA	153	65	0	NA	75	10.0	2.3	11.0	NA	308.0			
11	2017-01-01	仁武	11	24	2.0	0.39	0.17	2.5	6.0	8.6	60.0	NA	154	64	0	NA	68	5.4	2.1	316.0	NA	313.0			
12	2017-01-01	仁武	12	24	2.0	0.37	0.08	1.6	2.6	4.2	66.0	NA	125	54	0	NA	64	3.9	2.0	295.0	NA	272.0			
13	2017-01-01	仁武	13	25	1.9	0.26	0.05	1.5	0.5	2.1	61.0	NA	97	38	0	NA	62	2.8	1.9	276.0	NA	284.0			
14	2017-01-01	仁武	14	25	1.9	0.21	0.03	NA	NA	NA	62.0	NA	89	29	0	NA	64	2.2	1.9	271.0	NA	268.0			
15	2017-01-01	仁武	15	24	1.9	0.23	0.05	NA	NA	NA	68.0	NA	57	14	0	NA	70	2.5	1.9	285.0	NA	290.0			
16	2017-01-01	仁武	16	23	1.9	0.29	0.06	1.4	2.0	3.4	72.0	NA	79	15	0	NA	74	4.5	2.0	291.0	NA	284.0			
17	2017-01-01	仁武	17	23	2.0	0.33	0.08	1.1	2.8	3.8	72.0	NA	80	18	0	NA	76	2.9	2.1	291.0	NA	302.0			
18	2017-01-01	仁武	18	23	2.0	0.37	0.07	0.9	4.2	5.2	77.0	NA	89	41	0	NA	78	2.4	2.0	289.0	NA	297.0			
19	2017-01-01	仁武	19	22	1.9	0.38	0.07	1.1	6.2	7.4	68.0	NA	78	35	0	NA	79	2.5	2.0	281.0	NA	245.0			
20	2017-01-01	仁武	20	23	1.9	0.42	0.09	0.9	9.7	11.0	63.0	NA	72	43	0	NA	78	2.7	2.0	264.0	NA	282.0			
21	2017-01-01	仁武	21	22	1.9	0.37	0.08	1.2	9.0	10.0	54.0	NA	69	29	0	NA	79	3.0	2.0	308.0	NA	315.0			
22	2017-01-01	仁武	22	22	2.1	0.52	0.18	1.1	15.0	17.0	35.0	NA	80	48	0	NA	80	2.8	2.2	320.0	NA	325.0			
23	2017-01-01	仁武	23	21	2.1	0.50	0.17	1.2	17.0	18.0	22.0	NA	88	40	0	NA	83	2.6	2.3	320.0	NA	321.0			
24	2017-01-01	仁武	24	20	2.1	0.47	0.20	1.3	15.0	16.0	17.0	NA	106	45	0	NA	86	2.0	2.3	318.0	NA	317.0			

圖 4-1-3 轉置合併後的檔案(已匯入 R)

## 第二節 高雄 PM2.5 分析

這一節我們將使用高雄 2017 和 2018 年的歷史資料，分別對測站、月分和時段來整理和統計結果，統計出各項 PM2.5 最嚴重的情況，之後取較嚴重的之年份來做第三節的預測。由於我們都知道高雄的空氣污染很嚴重，所以我們先算出各自的年均值，因為環保署公告的 PM2.5 管制值年均值為  $15(\mu\text{g}/\text{m}^3)$ ，一旦超過這個標準，表示為不良狀況。所以我們針對高雄所有測站，來計算年均值是否超標，由表 4-2-1 可以看出結果 2017 年和 2018 兩年的年均值都大於  $15(\mu\text{g}/\text{m}^3)$ 。

表 4-2-1 2017、2018 PM2.5 年平均値

測站	2017 年	2018 年	管制值
大寮	25.80271241	27.58866909	年平均値( $\mu\text{g}/\text{m}^3$ ) > 15
小港	28.48045279	23.28848597	
仁武	27.90337711	24.24938359	
左營	28.7254902	23.30254335	
林園	21.52402402	22.73652482	
前金	26.96862049	24.93722339	
前鎮	29.51866018	23.32459701	
美濃	26.32403433	17.65087781	
復興	27.3097769	23.95824125	
楠梓	29.42601418	27.63176225	
鳳山	27.73727441	24.94673237	
橋頭	31.0198491	32.12670565	結論：2017 年和 2018 年的年均值都大於 $15(\mu\text{g}/\text{m}^3)$ 。
總計	27.68075606	24.50955136	

2017、2018 PM2.5 平均値

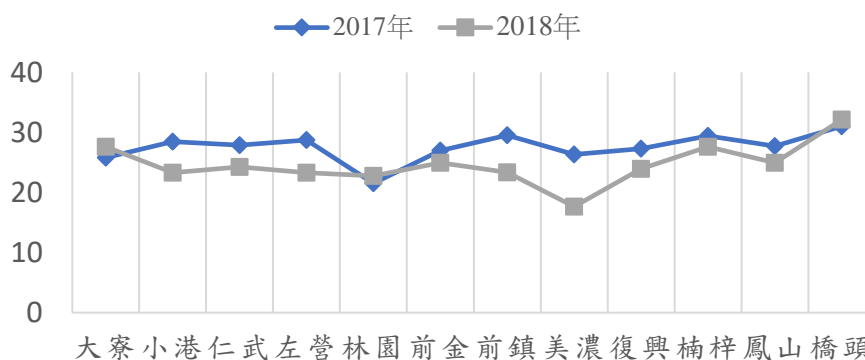


圖 4-2-1 2017、2018 PM2.5 年平均値比較圖

### 4.2.1 測站

整理完資料後，依據研究目的我們對高雄空污資料進行分析，我們要找出在這 12 個測站當中，哪些監測站污染較年平均嚴重，先分別看這兩年 PM2.5 年平均(四捨五入取到小數點後一位)的資料。

表 4-2-2 2017、2018 年高雄測站年平均比較

測站	2017 年	2018 年	平均
仁武	28	24	26
美濃	26	18	22
楠梓	29	28	28.5
左營	29	23	26
前金	27	25	26
小港	28	23	25.5
林園	22	23	22.5
大寮	26	28	27
鳳山	28	25	26.5
復興	27	24	25.5
前鎮	30	23	26.5
橋頭	31	32	31.5

由上表 4-2-2 得知，橋頭、楠梓和大寮為平均最高的前三名，美濃站為年平均最低的測站，單看數字可能感覺不出來差異，所以接下來以圖顯示，將會更清楚的顯示各站之間的差別。



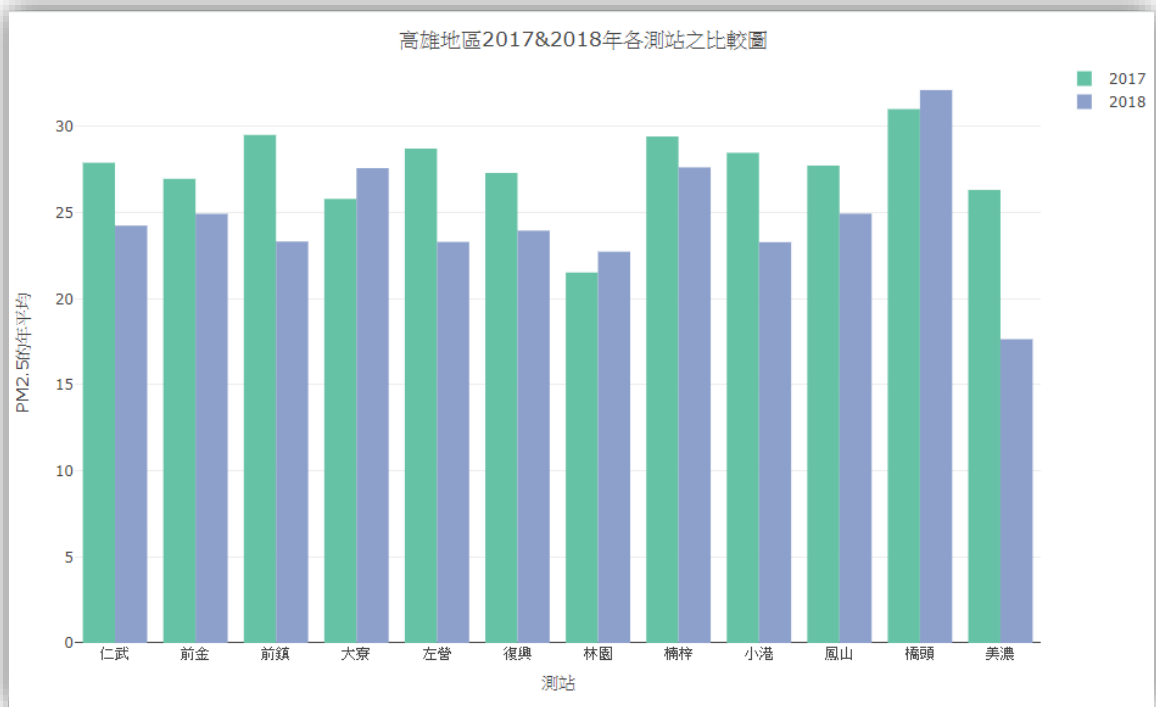


圖 4-2-2 2017、2018 高雄測站 PM2.5 比較圖

由圖 4-2-2 得知，2017 年為橋頭站、前鎮站、楠梓站為前三高的測站，2018 年為橋頭站、大寮站、楠梓站。很明顯的橋頭和楠梓的嚴重程度都有在前三名。橋頭站為背景測站，當南風速偏弱，可能就會導致擴散條件不佳，且 PM2.5 就容易累積，再加上尤其雲嘉南及高屏空品區午後因光化反應易使臭氧濃度升高，空氣汙染就相當嚴重。楠梓站為一般測站，由於附近都是工業區，與火力發電廠、煉鋼廠、煉油廠做鄰居，讓楠梓站的 PM2.5 一直都是居高不下的狀況，嚴重的空氣汙染也大大降低高雄地區的能見度。



#### 4.2.2 月份

接下來我們看 2017、2018 每個月的月平均(四捨五入取整數)，透過圖表來顯示一年之中是哪些月份較為嚴重，再說明其中的可能性。

表 4-2-3 2017 年 PM2.5 月平均

2017 測站	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
仁武	41	43	40	24	21	7	15	16	22	29	39	37
美濃	38	35	35	24	23	13	17	18	25	25	31	33
楠梓	44	47	43	29	24	9	13	14	21	26	40	43
左營	45	43	41	28	23	9	13	13	20	27	40	41
前金	42	40	37	24	21	7	11	12	20	28	40	40
小港	49	44	40	23	18	10	12	16	21	29	40	40
林園	NA	NA	29	21	18	9	11	10	21	28	32	33
大寮	38	35	33	22	18	8	13	16	23	29	36	38
鳳山	43	40	37	26	22	9	12	14	20	26	43	40
復興	45	42	38	24	20	8	12	14	19	26	38	40
前鎮	45	42	43	26	23	7	12	14	21	32	44	45
橋頭	48	46	46	29	24	11	16	17	23	31	38	44
平均	43	42	39	25	21	9	13	15	21	28	38	40



圖 4-2-3 2017 年 PM2.5 月平均圖

表 4-2-4 2018 年 PM2.5 月平均

2018 測站	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
仁武	36	42	35	26	15	13	11	8	16	28	31	33
美濃	28	30	23	20	12	7	5	3	11	24	24	23
楠梓	43	41	34	28	12	12	14	12	22	38	40	37
左營	39	39	32	22	13	12	11	9	16	30	30	28
前金	41	41	33	23	13	12	11	9	19	32	35	32
小港	40	39	32	21	11	11	9	8	15	29	32	33
林園	30	26	25	21	13	13	12	10	16	29	38	40
大寮	40	39	35	27	19	14	15	12	20	35	38	37
鳳山	39	45	33	24	15	12	10	7	16	31	35	35
復興	42	44	35	22	12	9	8	8	15	29	31	33
前鎮	42	41	35	24	8	8	8	6	14	29	34	32
橋頭	39	41	33	24	12	17	NA	NA	22	40	46	38
平均	38	39	32	24	13	12	10	8	17	31	35	33

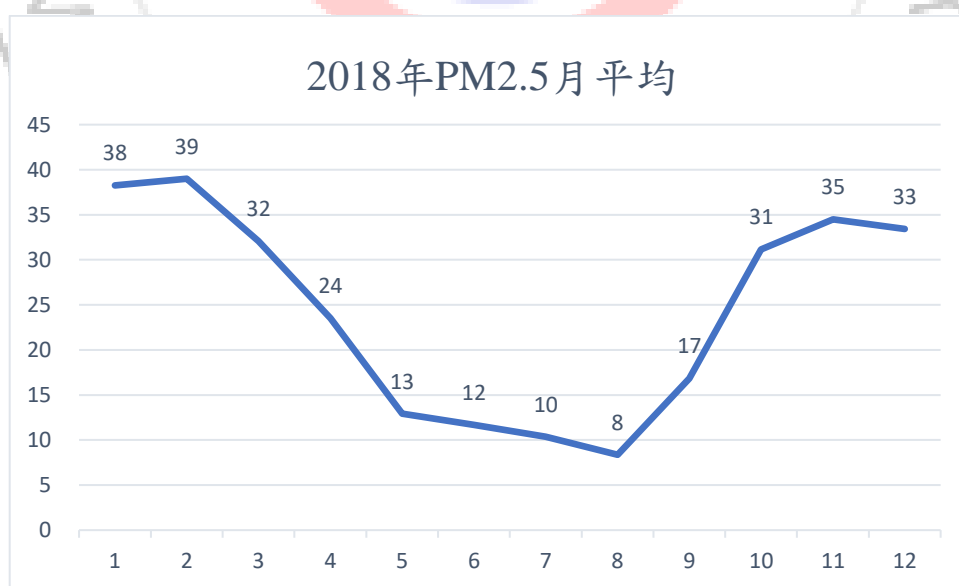


圖 4-2-4 2018 年 PM2.5 月平均圖

由圖 4-2-3 和圖 4-2-4 我們可以看出在 2017 年時，1 月和 2 月為最高而 6 月時為最低，在 2018 年時仍然為 1、2 月為最高而在 8 月最低，雖然兩年中最低濃度的 PM2.5 不同一個月，但可以看出來高雄的 PM2.5 一年之中，高峰集中於 10 月至翌年 3 月，我們認為可能是因為受到大陸冷高壓系統影響，導致部分擴散不良所致；反之，6 月至 8 月因受到西南季風和颱風的影響，易有較強風速與降雨，該段期間為 PM2.5 濃度為較低的月份，所以我們可以整理出冬季的污染程度應該是大於夏季。



### 4.2.3 時段

我們分別討論一天平均是哪個時段汙染比較嚴重，一樣我們分兩年來看各自的結果，再把所有測站的 24 小時平均(四捨五入取整數)算出。

表 4-2-5 2017 年 24 小時平均

測站 hr	仁武	美濃	楠梓	左營	前金	小港	林園	大寮	鳳山	復興	前鎮	橋頭	平均
1	27	26	28	28	27	32	21	26	27	27	27	31	27
2	27	25	28	28	27	29	21	26	27	28	30	32	27
3	28	25	28	28	26	29	21	27	27	28	31	32	28
4	27	24	28	29	26	31	22	28	27	28	31	31	28
5	28	24	28	29	27	31	22	28	27	28	30	31	28
6	27	24	27	30	28	31	22	28	27	29	30	31	28
7	27	25	28	29	29	31	23	28	27	28	32	31	28
8	29	26	29	30	30	33	24	29	28	29	33	33	29
9	31	27	31	31	32	33	25	28	30	31	34	34	31
10	32	26	33	33	34	32	27	27	32	32	35	35	32
11	33	26	34	35	34	31	27	25	33	32	36	34	32
12	32	27	35	35	33	30	26	25	32	32	33	34	31
13	32	26	34	33	30	28	20	25	31	30	30	32	29
14	29	27	33	31	27	26	18	24	30	29	29	30	28
15	27	27	30	28	23	25	17	23	28	26	27	29	26
16	25	27	29	26	22	25	17	23	26	25	27	27	25
17	25	27	27	25	22	25	17	23	25	24	26	27	24
18	24	28	27	25	22	25	19	24	25	24	25	27	25
19	25	28	27	25	22	25	19	25	25	23	26	28	25
20	26	28	28	26	23	26	20	25	25	23	26	29	25
21	26	28	29	26	24	26	20	25	26	23	27	30	26
22	27	27	29	27	26	27	22	25	27	24	28	31	27
23	27	27	29	27	26	28	22	25	27	25	29	31	27
24	28	26	29	28	27	28	22	26	27	26	29	32	27

表 4-2-6 2018 年 24 小時平均

測 站 hr	仁 武	美 濃	楠 梓	左 營	前 金	小 港	林 園	大 寮	鳳 山	復 興	前 鎮	橋 頭	平 均
1	25	17	28	24	26	27	22	27	25	21	22	33	25
2	24	17	27	24	25	24	22	26	25	22	23	32	24
3	24	16	27	23	25	24	22	28	25	27	24	33	25
4	23	15	27	23	24	24	23	28	24	26	24	33	25
5	23	14	27	23	26	24	23	28	24	25	24	32	24
6	23	14	26	23	26	24	23	29	23	25	24	32	24
7	23	15	27	23	25	24	23	29	24	23	24	33	24
8	25	17	28	24	27	25	24	30	25	24	26	35	26
9	27	18	30	25	29	25	25	30	27	27	27	36	27
10	28	19	30	27	30	26	27	29	28	28	27	36	28
11	28	19	30	27	30	26	28	29	28	29	28	36	28
12	27	19	30	27	29	26	28	29	28	28	26	36	28
13	27	19	30	27	28	24	26	28	28	28	25	34	27
14	26	19	30	25	25	23	23	25	27	28	24	32	26
15	24	19	28	23	24	22	21	24	26	26	22	30	24
16	23	18	27	21	22	20	19	25	24	23	21	28	23
17	22	18	26	20	21	21	19	26	23	21	20	27	22
18	22	18	26	20	21	20	19	27	22	21	20	27	22
19	22	20	25	20	21	20	20	28	22	18	19	28	22
20	22	19	26	21	22	21	20	28	23	18	20	29	22
21	23	19	27	21	22	21	22	28	23	21	21	31	23
22	23	19	27	22	24	23	22	27	24	22	22	33	24
23	24	18	28	23	24	23	22	27	25	23	23	33	24
24	24	17	28	23	25	24	22	27	25	24	23	33	25

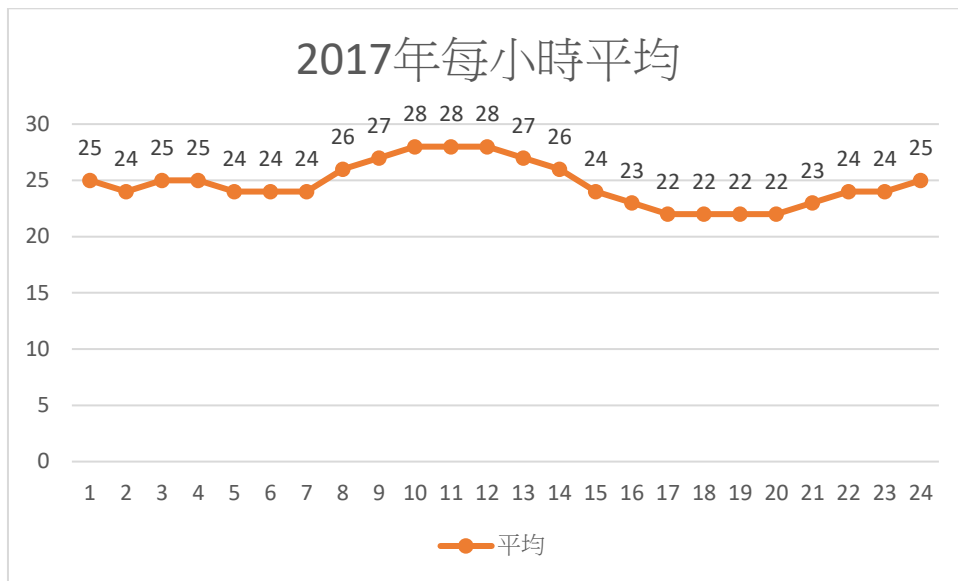


圖 4-2-5 2017 年每小時平均

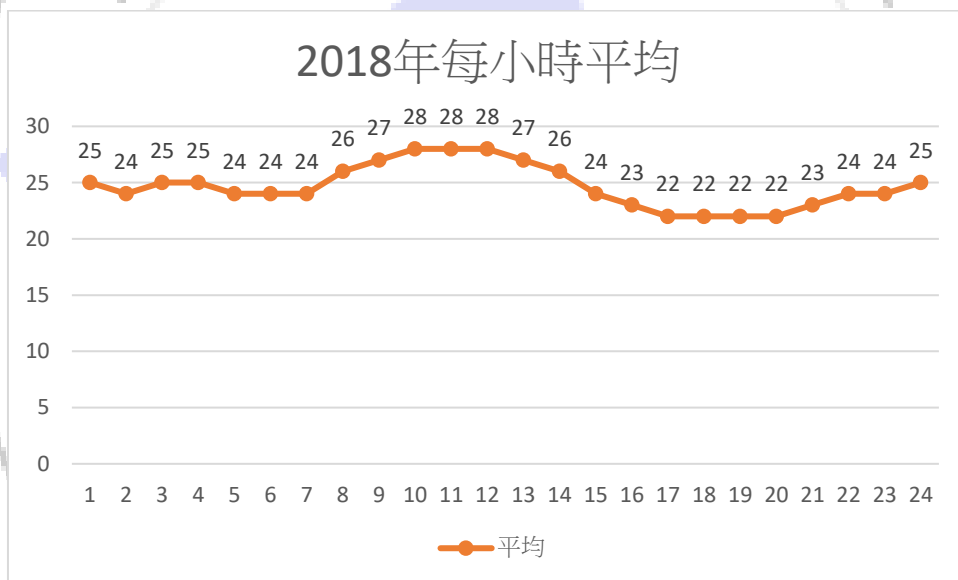


圖 4-2-6 2018 年每小時平均

圖 4-2-5 和圖 4-2-6 是運用每小時平均畫出來的圖，圖中我們可以發現平均早上的 PM2.5 濃度比較高，而晚上是相對比較低。我們認為通常白天是交通巔峰的時刻，汽機車等交通運輸排放量較大，所以可能也是導致這個時段空氣中 PM2.5 較高的原因之一，而且白天通常都比晚上氣溫還較高，當氣溫高 PM2.5 的濃度就會升高，且少了風的流動，就會導致汙染散不去，使 PM2.5 在都市和郊區間反覆循環。

### 第三節 預測 PM2.5

對於 PM2.5 的預測，我們使用了三種不同的方式來建模。第一種方式，使用一天的其他污染物測量值或氣象變數測量值為自變項，當天的 PM2.5 測量值為因變項，使用高斯過程回歸建模。目的是想要了解各變數對 PM2.5 的影響程度。第二種方式，利用當天其他測站的經緯度為自變項，這些測站某特定一小時 PM2.5 的測量值為因變項，利用高斯回歸過程建模，來預測剩下那個測站的特定一小時的 PM2.5。第三種方式與第二種方式類似，但再加入其他變項的資料，以求更精準的預測 PM2.5。

#### 4.3.1 單變量-PM10、風速和風向

這一小節我們將分別利用 PM10、風速和風向，單獨的預測 PM2.5，以 2017 年 1 月 1 日為例。由於 PM10 和 PM2.5 極為相似我們認為影響程度是很接近的，而風向和風速是因為 PM10 會受到風的影響，所以也將這兩項列入，其中林園站的 PM2.5 人工檢核為無效值，所以我們在編碼時對 PM2.5 就設為

NA。表列的  $\hat{\beta}$  為超參數、 $\hat{\sigma}$  為共變異數矩陣、誤差值為  $\frac{\sum(|\text{估計值}-\text{觀察值}|)}{\text{個數}}$ 、標

準差為  $\sqrt{\frac{\sum(\text{估計值}-\text{觀察值})^2}{\text{個數}}}$  計算出來並平均比較。

表 4-3-1 PM10 單變量

PM10/測站	$\hat{\beta}$	$\hat{\sigma}$	誤差平均值	誤差標準差
仁武	-5.246114	213373287	18.31537	21.92588
美濃	-5.359575	273932763	8.968928	9.505494
楠梓	-4.895108	264858820	8.354554	11.05321
左營	-5.144176	251290236	11.66531	14.01956
前金	-5.320054	279239748	5.212859	6.789013
小港	-5.395527	233133810	14.1886	17.18884
林園	-5.28501	239744658	NA	NA
大寮	-5.367112	279516865	6.070258	7.418749
鳳山	-5.303893	270018402	8.412718	10.0278
復興	-5.413132	282392099	3.854698	6.619918
前鎮	-5.335969	260381632	9.392878	12.23292
橋頭	-5.40647	268143144	7.454134	10.79107
平均	-5.289345	259668788.7	9.262755	11.5975

由上表 4-3-1 可看出仁武、小港和左營的誤差平均值以及誤差標準差較大，估計的應該比較不準確，而復興、前金和大寮的誤差平均值以及誤差標準差較小，可能估計的相對準確。2017/01/01 這天剛好 PM2.5 整天都為無效值，因此我們沒有驗證資料，所以誤差平均值以及誤差標準差都為 NA。

表 4-3-2 風速單變量

風速/測站	$\hat{\beta}$	$\hat{\sigma}$	誤差平均值	誤差標準差
仁武	-5.051439	447802297	12.22719	14.62292
美濃	-4.870315	464787962	8.738282	11.14659
楠梓	-5.036073	443360434	10.69685	15.32116
左營	-4.740865	427134011	13.86099	17.08932
前金	-4.992754	452409937	10.79036	13.26897
小港	-5.01406	425894351	13.61834	17.62962
林園	-4.904939	402651619	NA	NA
大寮	-4.84829	459591308	9.496696	11.81522
鳳山	-4.881679	438400430	13.28331	15.41484
復興	-4.855488	413277240	15.90782	17.74584
前鎮	-5.018205	440564841	13.71486	15.53925
橋頭	-5.030434	446835468	10.14868	14.7273
平均	-4.937045083	438559158.2	12.04394345	14.93827545

由表 4-3-2 可看出誤差值以仁武和復興的誤差值以及誤差值標準差較大，估計的可能比較不準確，而美濃和大寮之誤差值以及誤差值標準差較小，就有可能估計的相對準確。2017/01/01 這天剛好 PM2.5 整天都為無效值，因此我們沒有驗證資料，所以誤差平均值以及誤差標準差都為 NA。



#### 4-3-3 風向單變量

風向/測站	$\hat{\beta}$	$\hat{\sigma}$	誤差平均值	誤差標準差
仁武	-5.539662	44802437	12.96952	16.32154
美濃	-5.675422	47514325	9.975157	12.19857
楠梓	-5.444415	43186500	14.04598	18.57487
左營	-5.261989	45041078	14.06145	15.72984
前金	-5.89628	46355530	11.44916	14.05685
小港	-7.035215	45403904	14.36835	17.51541
林園	-5.654185	41550961	NA	NA
大寮	-6.65099	48124575	10.29279	12.34591
鳳山	-5.85815	46262791	12.25559	14.51117
復興	-6.69775	44760986	13.04447	14.23211
前鎮	-6.735092	46525112	13.1418	15.19709
橋頭	-5.3784	44545238	11.34938	16.37601
平均	-5.985629167	45339453.08	12.45033155	15.18721545

由表 4-3-3 可觀察誤差值以楠梓、小港和左營的誤差值以及誤差值標準差較大，估計的應該比較不準確，而美濃、前金和大寮之誤差值以及誤差值標準差較小，可能估計的相對準確。2017/01/01 這天剛好 PM2.5 整天都為無效值，因此我們沒有驗證資料，所以誤差平均值以及誤差標準差都為 NA。

總結本小節的表可以看出，PM10 建模的誤差值介在  $9.262755 \pm 11.5975$ 、風速建模的平均誤差值為  $12.04394345 \pm 14.93827545$  和風向建模的為  $12.45033155 \pm 15.18721545$ 。由此可知以 PM10 建立模型之誤差值及標準差比較小，可能較準確；相較之下也可看出以風向估計之誤差值以及標準差都較大，可能相較之下估計得比不準確。其中仁武、小港以及左營之誤差值和誤差標準差較大，應是估計的比較不準確之測站，美濃、前金以及大寮之誤差值和誤差標準差較小，應是估計的比較準確。

### 4.3.2 雙變量－經緯度

前一節的單變量，只是用一個自變項來預測我們的 PM2.5，而我們最想研究的是透過經緯度來預測，這時候就必須要有兩個自變項，也就是我們這一小節要探討的雙變量預測。首先由於自變項必須介於 0 到 1 之間，所以我們都要必須把我們 12 個測站的經緯度轉換成表 4-3-4 的形式，轉換後我們才能進行接下來的動作。

表 4-3-4 經緯度轉換

	經度(x1)	緯度(x2)
仁武	0.183724057148519	0.518596426971685
美濃	1	1
楠梓	0.165815653149408	0.628997012989903
左營	0.019925264790296	0.483467505438235
前金	0	0.378800889916181
小港	0.204779423895477	0.213651650774717
林園	0.510047183818903	0
大寮	0.565978981753358	0.209452018520947
鳳山	0.288699805325459	0.365994115070416
復興	0.098702444979672	0.319763513931539
前鎮	0.0803362259543717	0.311535006421942
橋頭	0.0726028640247912	0.687992318409826

接下來我們要利用當天其他測站的經緯度為自變項，這些測站某特定一小時 PM2.5 的測量值為因變項，利用高斯回歸過程建模，來預測剩下那個測站的特定一小時的 PM2.5。以預測仁武站的 PM2.5 為例，我們就用除了仁武的其他 11 個測站的經緯度及其當天某特定小時 PM2.5 的測量值建模，來預測出當天某特定小時，仁武測站 PM2.5 的值。而為了瞭解預測的準確度，拿原本仁武站當天特定小時的實際測量值減去我們的預測值，得到誤差值。這樣的過程，每天每一小時都做。2017 年共 365 天，我們就建了  $365 \times 24 = 8760$  個模型，得到 8760 個誤差值。12 個測站都如此去做，最後將誤差值整理分析於後，以了解在此方式下，高斯回歸過程的預測能力。

表 4-3-5 24 小時平均誤差值(a)

	仁武	前金	前鎮	大寮	小港	左營
1	6.23	7.72	5.50	6.28	6.97	5.26
2	6.15	7.83	5.75	6.01	5.81	5.51
3	6.30	8.45	5.45	5.28	5.29	5.66
4	6.54	7.20	5.82	5.26	5.57	5.64
5	6.57	7.38	5.31	5.45	5.99	5.68
6	6.48	7.63	5.11	5.01	6.19	5.96
7	6.61	7.04	6.22	4.74	5.99	5.93
8	6.85	7.39	6.55	4.77	6.13	5.97
9	6.65	7.70	6.39	5.38	6.00	5.52
10	6.86	9.01	7.06	7.07	7.00	6.18
11	7.08	10.31	7.80	8.51	6.95	6.57
12	7.57	10.05	7.01	8.64	7.10	7.63
13	7.62	10.43	6.97	8.51	6.59	8.13
14	7.13	10.42	6.95	7.80	6.37	7.45
15	7.15	10.88	6.39	6.98	5.65	6.97
16	6.92	8.71	6.25	6.41	4.89	6.50
17	6.54	8.25	5.77	5.77	4.75	5.78
18	5.65	7.41	5.53	4.97	4.36	5.55
19	5.40	7.79	5.86	4.59	4.01	5.19
20	5.37	6.78	6.26	4.73	4.23	5.34
21	5.77	7.49	6.27	4.94	4.27	5.47
22	5.80	7.61	6.33	5.42	4.94	5.38
23	6.25	7.68	5.92	5.57	5.29	5.93
24	6.19	8.13	5.97	6.10	5.04	6.07
平	6.49	8.30	6.19	6.01	5.64	6.05

表 4-3-6 24 小時平均誤差值(b)

	復興	林園	楠梓	橋頭	美濃	鳳山
1	5.23	7.72	6.32	5.60	7.01	5.60
2	5.31	7.83	5.83	5.57	7.23	5.57
3	5.03	8.45	6.60	5.90	6.98	5.90
4	5.17	7.20	6.64	5.63	7.35	5.63
5	4.79	7.38	6.63	5.25	7.48	5.25
6	4.68	7.63	6.48	5.61	7.50	5.61
7	5.62	7.04	6.32	5.16	7.52	5.16
8	6.07	7.39	6.87	5.84	7.47	5.84
9	5.86	7.70	6.38	6.15	7.55	6.15
10	6.45	9.01	6.86	6.31	9.08	6.31
11	7.36	10.31	7.30	6.20	9.60	6.20
12	6.37	10.05	7.75	5.81	9.59	5.81
13	6.71	10.43	8.12	5.54	9.53	5.54
14	6.52	10.42	7.47	5.03	9.45	5.03
15	5.94	10.88	7.99	5.21	8.74	5.21
16	5.70	8.71	7.77	5.15	8.41	5.15
17	5.34	8.25	7.33	4.81	7.95	4.81
18	4.93	7.41	6.76	4.88	7.50	4.88
19	5.48	7.79	6.98	5.03	6.96	5.03
20	6.04	6.78	6.54	5.49	6.93	5.49
21	6.17	7.49	6.94	5.44	7.27	5.44
22	5.96	7.61	6.54	5.72	6.82	5.72
23	5.46	7.68	6.37	6.00	6.93	6.00
24	5.47	8.13	6.49	6.01	6.46	6.01
平	5.74	8.30	6.89	5.56	7.80	5.56

透過表 4-3-5 和表 4-3-6 的資料算出平均後，再利用折線圖來觀察變化。

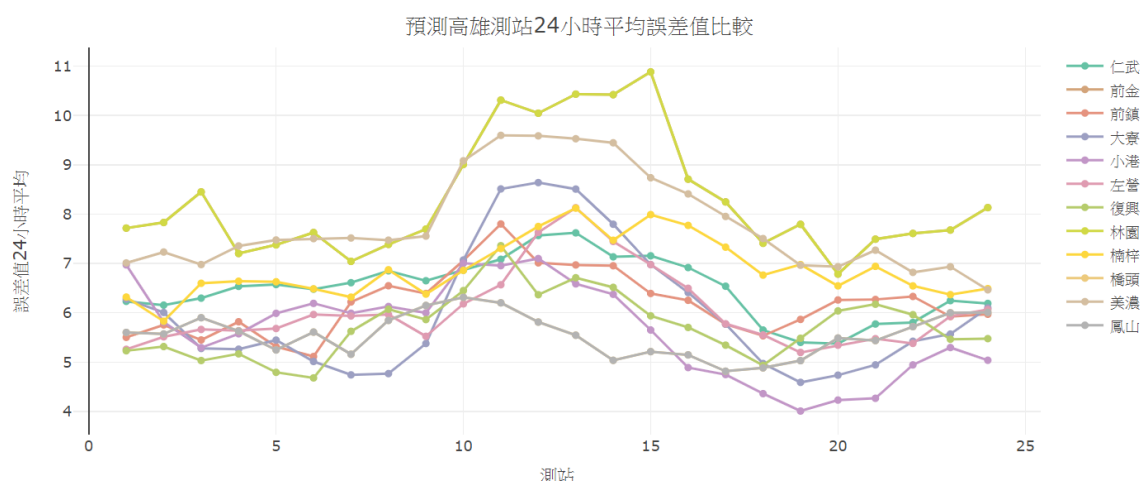


圖 4-3-1 24 小時平均誤差值比較

表 4-3-5(a) 與 4-3-5(b)是以小時來看，以解高斯回歸過程在個小時的預測能力。表中算出個小時預測誤差的平均值。而從圖 4-3-1 可看出林園(8.3)以及前金(8.3)之平均的誤差值相較之下比較大，可能估計的相對不準確，而橋頭(5.56)以及鳳山(5.56)之平均誤差值相較之下比較小，可能估計出來的數值相對準確。從這個結論我們可推斷，因此次的預測是經由經緯度做預測的，所以在高雄市中心的小港以及鳳山的誤差值是較小的，因四周圍有很多的測站幫助建模，相較之下從地理位址可明顯看到的較偏遠的林園和美濃，估計的不太準確，所以我們可以知道在越多母體的建模的狀況下，估測的值相對的是更加的準確的。之後我們再把表 4-3-5 和表 4-3-6 的第 1 小時到第 24 小時平均，可以看出一天之中哪個小時的誤差值最大，我們將資料整理到下表。

表 4-3-7 24 小時誤差平均

1	6.29	6.20	6.27	6.14	6.10	6.16
7	6.11	6.43	6.45	7.27	7.85	7.78
13	7.84	7.50	7.33	6.71	6.28	5.82
19	5.84	5.83	6.08	6.15	6.26	6.34

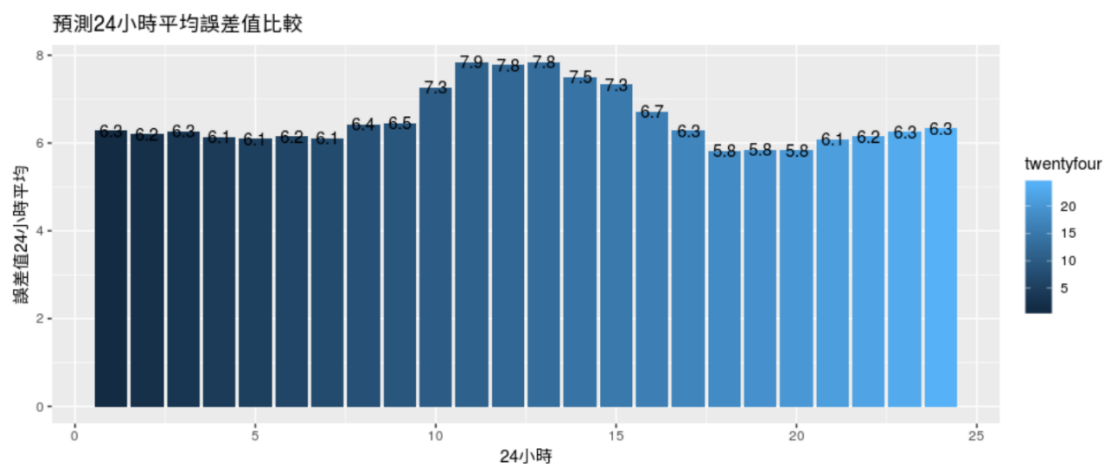


圖 4-3-2 24 小時誤差平均圖

上圖 4-3-2 可以看出時段 10~15 點平均的誤差值相較之下比較大，可能估計的相對不準確，18~22 點之間平均誤差值相較之下比較小，可能估計出來的數值相對準確。接下來我們看每個四季平均誤差，這樣我們可以清楚看出哪些測站四季之間誤差較大。我們透過圖表來呈現。

表 4-3-8 季節平均誤差(上)

	仁武	前金	前鎮	大寮	小港	左營
春	7.07	8.36	5.59	7.12	6.26	5.39
夏	4.80	6.04	4.56	5.04	6.20	4.43
秋	5.76	8.54	5.70	5.31	6.62	5.40
冬	7.26	7.94	6.16	7.66	8.80	5.79

表 4-3-9 季節平均誤差(下)

	復興	林園	楠梓	橋頭	美濃	鳳山
春	5.02	8.36	7.60	7.77	7.10	7.77
夏	4.04	6.04	4.35	4.53	4.60	4.53
秋	5.41	8.54	5.72	4.70	7.08	4.70
冬	6.46	7.94	7.62	5.36	9.48	5.36

表 4-3-9 季節平均誤差(下)

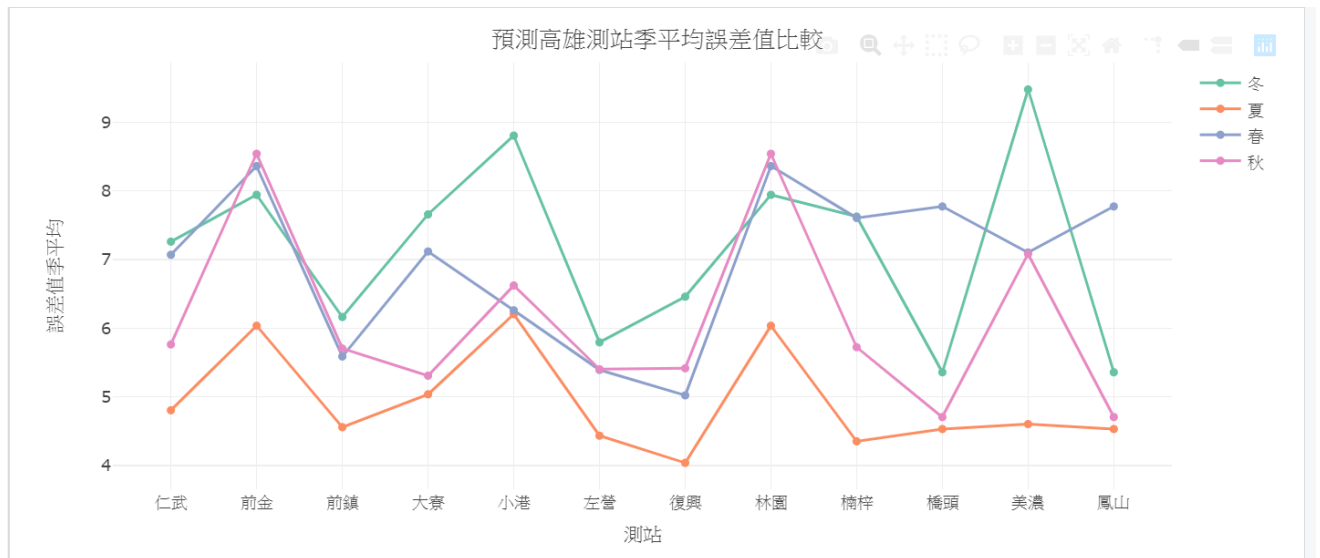


圖 4-3-3 季節平均誤差比較

由圖 4-3-3 可看出林園以及美濃之四季的平均誤差值相差比較大，可能四季估計的準確度相差較大，而前鎮以及復興之四季的平均誤差值相較之下相差比較小，可能四季估計的準確度相差較小。接下來我們一樣將春、夏、秋和冬加總誤差平均做比較。

表 4-3-10 四季誤差平均

春	夏	秋	冬
6.95	4.93	6.12	7.15

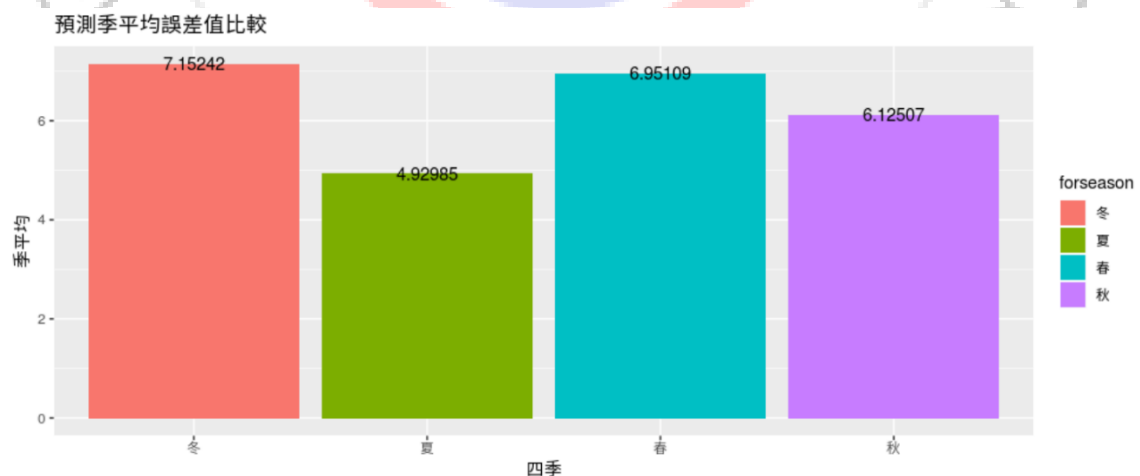


圖 4-3-4 四季誤差平均圖

由圖 4-3-4 看出，發現冬天時平均的誤差值相較之下比較大，表示可能估計的相對不準確，而夏天平均的誤差值相較之下比較小，表示可能估計出來的數值相對準確。

### 4.3.3 多變量-經緯度和 PM10

前一小節的雙變量，只單用經緯度來建模，但我們認為模型應該可以再更好一點，所以除了經緯度之外，這次的多變量再加上 PM10 去估計 PM2.5。如同前一小節，這一小節也分成以 24 小時以及以季節區分的 PM2.5 之估計值去比較以及探討。所以接下來先呈現的是 24 小時各測站的平均誤差值。

表 4-3-11 24 小時平均誤差值(a)

	仁武	前金	前鎮	大寮	小港	左營
1	6.43	6.16	5.53	6.16	7.29	5.51
2	6.51	6.04	5.71	6.11	6.09	5.98
3	6.32	6.31	5.46	5.68	5.67	6.03
4	6.55	6.24	5.79	5.44	5.92	5.99
5	6.5	6.75	5.59	5.52	6.07	5.8
6	6.95	6.33	5.24	5.34	6.53	6.27
7	6.96	5.83	6.11	5.12	6.28	6.39
8	7.28	6.15	6.68	4.88	6.6	6.19
9	7.19	6.21	5.88	5.3	6.45	5.71
10	6.87	6.47	7.15	7.41	7.25	5.97
11	7.53	6.94	8.18	8.86	7.61	7.2
12	7.7	6.9	6.81	8.57	7.44	7.87
13	7.48	6.98	7.21	8.58	6.71	7.97
14	7.5	6.68	6.81	8.12	6.51	7.36
15	7.6	7.16	6.25	7	6.36	7.42
16	7.15	6.64	6.52	6.52	5.56	6.69
17	6.44	6.37	6.15	5.86	5.03	5.9
18	5.93	6.16	5.54	5.14	4.78	5.74
19	5.81	6.05	6.07	4.77	4.53	5.41
20	5.71	5.88	6.02	4.66	4.44	5.57
21	6.1	5.92	6.15	4.73	4.65	5.96
22	6.08	5.83	6.44	5.4	5.13	5.75
23	6.73	6.3	5.87	5.63	5.51	6.07
24	6.28	6.1	6.04	6.33	5.39	6.31
平	6.73	6.35	6.22	6.13	5.99	6.29



表 4-3-12 24 小時平均誤差值(b)

	復興	林園	楠梓	橋頭	美濃	鳳山
1	5.27	7.74	6.66	6.04	6.99	6.04
2	5.35	7.53	6.33	6.29	7.22	6.29
3	5.28	8.26	6.68	6.15	6.99	6.15
4	5.12	7.09	6.69	5.71	7.19	5.71
5	4.71	7.24	6.61	5.59	7.48	5.59
6	4.58	7.35	7.16	5.78	7.47	5.78
7	5.4	6.83	6.81	5.76	7.35	5.76
8	5.98	7.12	6.99	6.43	7.4	6.43
9	5.69	7.48	6.73	6.5	7.49	6.5
10	6.41	9.07	7.06	6.15	8.88	6.15
11	7.3	10.56	7.53	6.24	9.45	6.24
12	6.54	10.06	7.86	5.89	9.43	5.89
13	6.72	10.44	7.83	5.65	9.08	5.65
14	6.35	9.91	7.82	5.39	9.38	5.39
15	5.81	10.8	7.93	5.54	8.87	5.54
16	5.67	8.67	7.73	5.55	8.54	5.55
17	5.37	8.06	7.71	5.11	8.02	5.11
18	4.83	7.55	6.79	5.37	7.53	5.37
19	5.89	7.79	7.19	5.54	6.85	5.54
20	6.07	6.64	6.53	5.4	6.89	5.4
21	5.95	7.45	6.91	5.34	7.32	5.34
22	5.65	7.53	6.64	5.74	6.96	5.74
23	5.27	7.66	6.5	5.97	6.86	5.97
24	5.34	8.06	6.67	6.11	6.49	6.11
平	5.69	8.2	7.06	5.8	7.76	5.8

透過表 4-3-11 和 4-3-12 的資料，再利用折線圖來觀察變化。

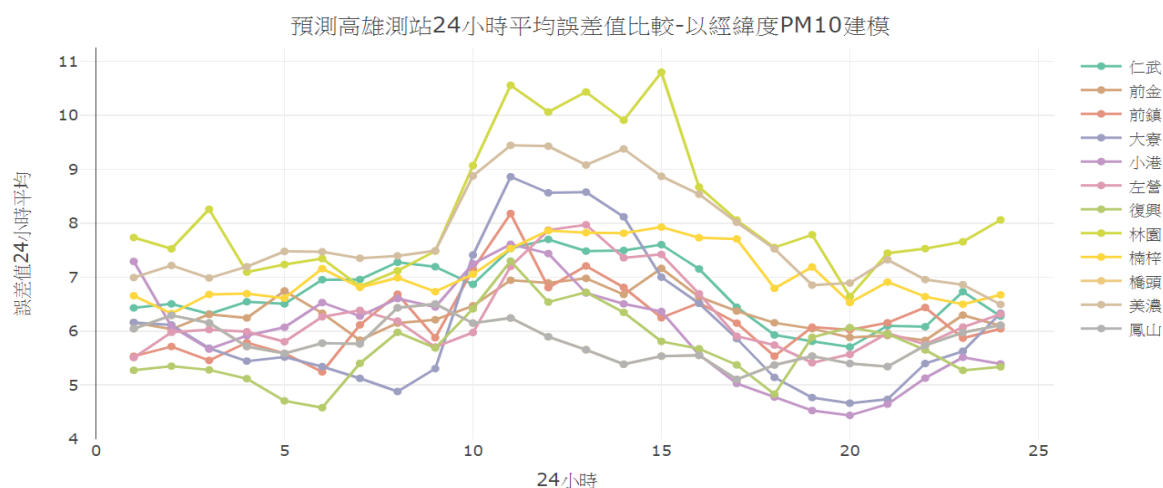


圖 4-3-5 24 小時平均誤差值比較

由上圖以及表可看出楠梓(7.06)以及美濃(7.76)之平均的誤差值相較之下比較大，可能估計的相對不準確，而復興(5.69)、橋頭(5.8)以及鳳山(5.8)之誤差平均值相較之下比較小，可能估計出來的數值相對準確。加上 PM10 的多變量模型估計出來的誤差值最大的楠梓(7.06)以及美濃(7.76)相對於雙變量誤差值最大的林園(8.3)以及前金(8.3)降低了一些，可以看得出來 PM10 對於這個模型是有幫助的。接著我們再把 24 小時平均誤差值的第 1 小時到第 24 小時平均，可以很明顯的看出一天之中哪個小時的誤差值最大，我們將資料整理到以下。

表 4-3-13 24 小時誤差平均

1	6.32	6.29	6.25	6.12	6.12	6.23
7	6.22	6.51	6.43	7.07	7.8	7.58
13	7.53	7.27	7.19	6.73	6.26	5.89
19	5.95	5.77	5.99	6.07	6.2	6.27

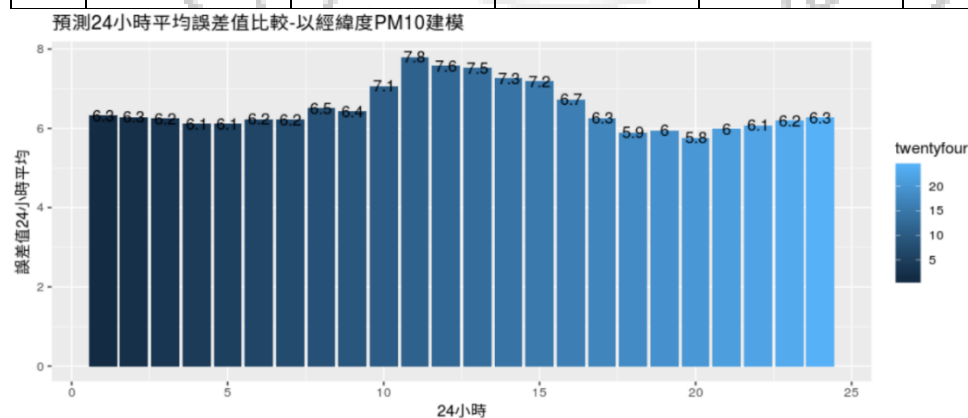


圖 4-3-6 24 小時誤差平均圖

由表 4-3-13 和圖 4-3-6 可以看出時段 10~15 點平均的誤差值相較之下比較大，可能估計的相對不準確，19~21 點之間平均誤差值相較之下比較小，可能估計出來的數值相對準確。接下來我們一樣看每個四季平均誤差值，這樣我們可以清楚看出哪些測站四季之間誤差較大。我們一樣透過圖表來呈現資料。

表 4-3-14 季節平均誤差(上)

	仁武	前金	前鎮	大寮	小港	左營
春	7.44	8.36	5.57	7.13	7.22	6.36
夏	4.91	6.04	4.32	4.95	5.95	4.22
秋	6.13	8.54	5.47	5.01	7.02	5.92
冬	7.03	7.94	6.76	7.57	9.02	5.49

表 4-3-15 季節平均誤差(下)

	復興	林園	楠梓	橋頭	美濃	鳳山
春	4.93	8.13	8.12	7.63	6.98	7.63
夏	3.96	6.08	4.79	5.02	4.88	5.02
秋	6.01	8.76	6.27	5.79	6.96	5.79
冬	6.21	7.95	7.43	5.7	9.33	5.7

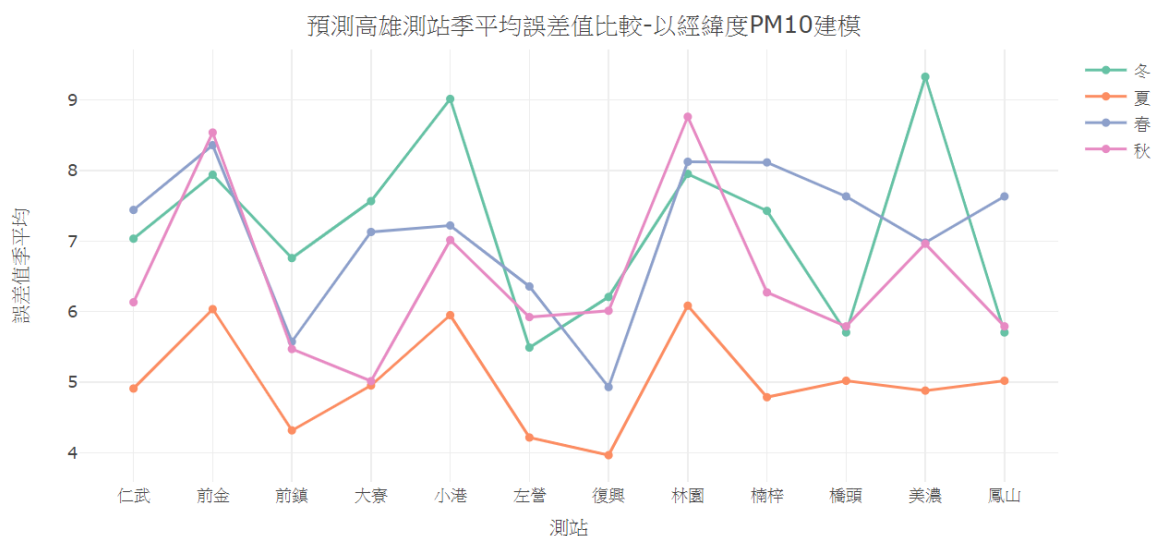


圖 4-3-7 季節平均誤差比較

由圖 4-3-7 可看出小港以及美濃之四季的平均誤差值相差比較大，四季估計的準確度相差較大，而前鎮以及復興之四季的平均誤差值相較之下相差比較小，可能四季估計的準確度相差較小，接下來我們一樣將春、夏、秋和冬做比較。

表 4-3-16 四季誤差平均

春	夏	秋	冬
7.13	5.01	6.47	7.18

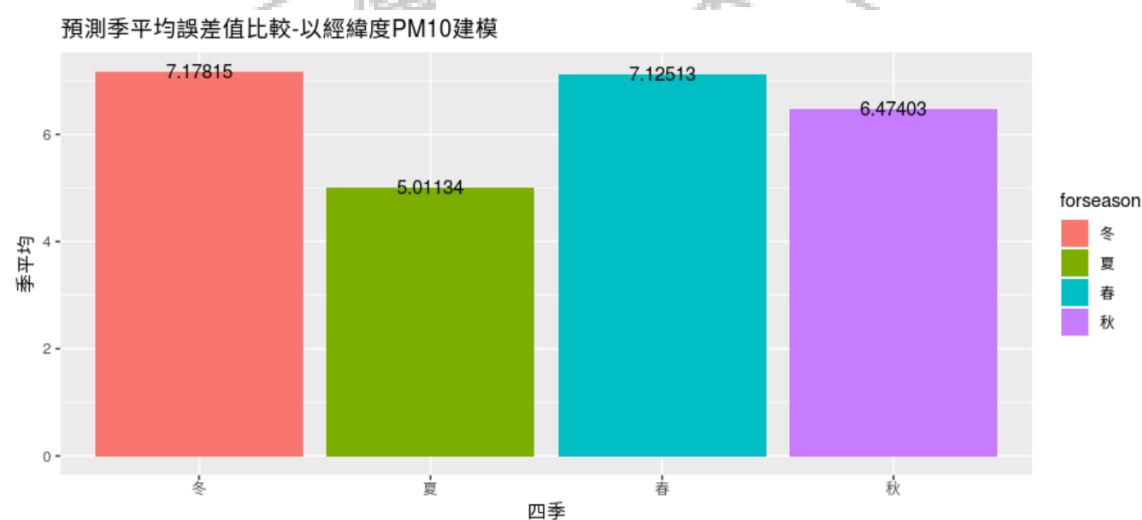


圖 4-3-8 四季誤差平均圖

由以上表圖整理出後，可以發現冬天和春天時，平均的誤差值相較之下比較大，表示可能估計的相對不準確，而夏天平均誤差值相較之下比較小，表示可能估計出來的數值相對準確些。

#### 4.3.4 多變量-經緯度和風向風速

前一小節運用經緯度和 PM10 來建模，發現加上 PM10 後對於結果是有顯著的改善，所以在此章節我們決定加入風速和風向來建立模型，因為台灣有一部份的 PM2.5 是來自於境外，因此我們猜測風速和風向對於 PM2.5 的建模應該也是有關聯的。而這一小節我們也分成 24 小時和季節區分的 PM2.5 之估計值去比較以及探討，接下來先呈現的是 24 小時各測站的平均誤差值。

表 4-3-17 24 小時平均誤差值(a)

	仁武	前金	前鎮	大寮	小港	左營
1	6.93	6.48	5.98	7.06	7.77	6.08
2	6.81	6.4	6.26	6.56	5.86	6.43
3	6.56	6.21	5.81	5.51	5.86	6.11
4	7.15	7.17	6.36	5.78	6.48	6.27
5	6.98	6.96	5.93	5.97	6.26	5.93
6	7.02	6.54	5.51	5.85	6.48	6.53
7	7.31	6.82	6.28	5.57	6.2	6.35
8	7.8	6.62	6.57	5.66	6.75	6.67
9	7.22	6.66	6.56	5.87	6.62	6.73
10	7.08	6.8	7.51	7.26	7.3	6.79
11	7.66	7.19	8.22	8.73	7.63	7.15
12	8.16	7.6	7.78	8.97	7.76	8.07
13	8.52	7.86	7.37	8.91	7.4	8.89
14	7.76	7.2	6.9	8.48	7.06	7.91
15	7.34	7.15	6.25	7.21	6.09	7.61
16	7.32	7.3	6.91	7.21	5.58	6.88
17	6.96	6.64	6.32	6.02	5.59	6.35
18	6.17	6.82	5.76	5.49	4.88	6.42
19	6.11	6.49	5.73	5.15	5.05	5.86
20	6.28	6.13	6.13	4.96	4.99	6
21	6.39	5.9	6.14	5.36	5.13	6.21
22	6.51	5.95	6.04	5.75	5.5	5.98
23	6.6	6.5	6.01	6.08	5.82	6.63
24	6.88	6.52	6.27	6.73	5.68	6.72
平	7.06	6.75	6.44	6.51	6.24	6.69

表 4-3-18 24 小時平均誤差值(b)

	復興	林園	楠梓	橋頭	美濃	鳳山
1	5.2	7.78	6.57	6.28	7.19	6.28
2	5.73	7.58	6.52	6.17	7.33	6.17
3	5.33	8.17	7.17	6.6	7.15	6.6
4	5.25	7.28	7.2	6.07	7.47	6.07
5	5.14	7.72	7.01	6.05	7.67	6.05
6	4.84	7.79	6.96	5.94	7.41	5.94
7	5.11	7.46	6.73	6.04	7.35	6.04
8	5.74	7.76	7.28	6.67	7.33	6.67
9	5.67	7.75	7.09	7.06	7.58	7.06
10	6.16	8.35	7.22	6.89	9.06	6.89
11	7.01	10.45	7.41	6.74	9.55	6.74
12	6.04	10.18	8.05	6.68	9.75	6.68
13	6.94	10.4	8.68	6.27	9.51	6.27
14	6.51	10.35	7.9	6.32	9.68	6.32
15	5.54	10.66	7.86	6.12	8.77	6.12
16	5.37	8.97	7.71	6.53	8.72	6.53
17	4.77	8.22	7.65	5.56	8.03	5.56
18	4.79	7.45	7.25	5.66	7.44	5.66
19	5.32	7.77	7.09	5.65	7.23	5.65
20	5.91	6.64	6.64	5.51	7.07	5.51
21	6.09	7.34	7.35	5.86	7.13	5.86
22	5.86	7.53	7.02	5.87	6.59	5.87
23	5.51	7.45	6.7	6.52	6.93	6.52
24	5.33	8.28	7	6.41	6.54	6.41
平	5.63	8.31	7.25	6.23	7.85	6.23

透過表 4-3-17 和、4-3-18 的資料，再利用折線圖來觀察變化。

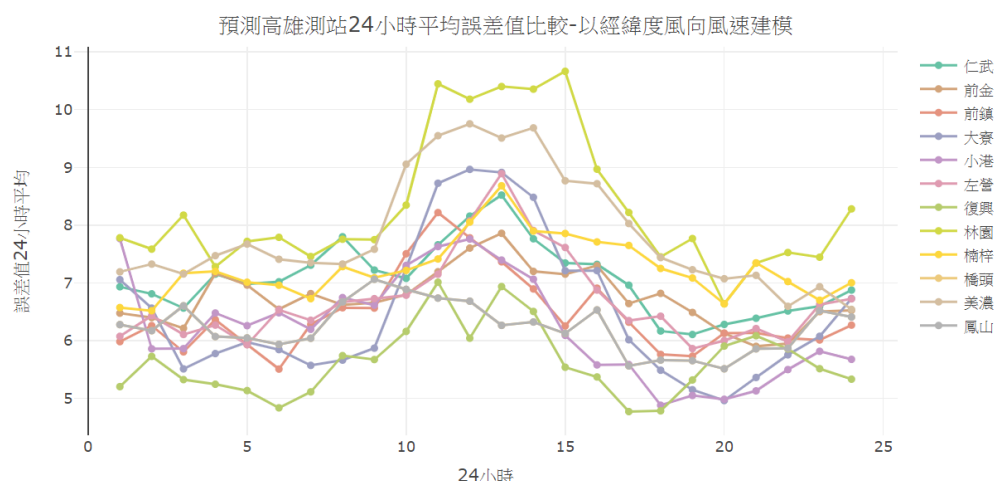


圖 4-3-9 24 小時平均誤差值比較

由上圖以及表可看出林園(8.31)以及美濃(7.85)之平均的誤差值相較之下比較大，可能估計的相對不準確，而復興(5.63)、橋頭(6.23)以及鳳山(6.23)之誤差平均值相較之下比較小，可能估計出來的數值相對準確。加上風速以及風向的多變量模型估計出來的誤差值最大的林園(8.31)以及美濃(7.85)相對於雙變量誤差值最大的林園(8.3)以及前金(8.3)降低了一些，可以看得出來風向以及風速對於這個模型是有幫助的，但對於林園測站仍估計的較不準確，可能是因為風速風向對於距離遠的測站仍無法讓模型更加完整。

接著我們再把 24 小時平均誤差值的第 1 小時到第 24 小時平均，可以很明顯的看出一天之中哪個小時的誤差值最大，我們講資料整理到以下。

表 4-3-19 24 小時誤差平均

1	6.63	6.49	6.42	6.55	6.47	6.4
7	6.44	6.79	6.82	7.28	7.87	7.98
13	8.09	7.7	7.23	7.09	6.47	6.15
19	6.09	5.98	6.23	6.21	6.44	6.56

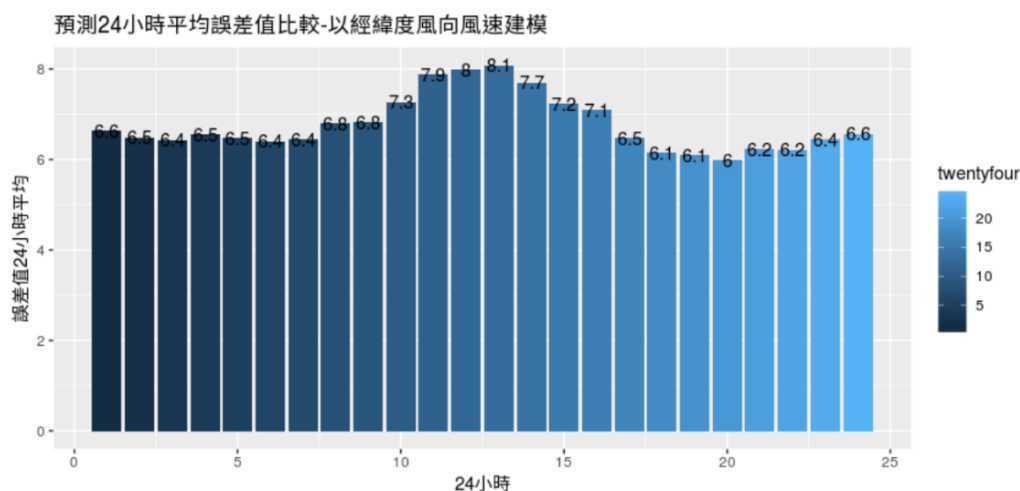


圖 4-3-10 24 小時誤差平均圖

由表 4-3-19 和圖 4-3-10 可以看出時段 11~13 點平均的誤差值相較之下比較大，可能估計的相對不準確，19~21 點之間平均誤差值相較之下比較小，可能估計出來的數值相對準確。接下來我們一樣看每個四季平均誤差值，這樣我們可以清楚看出那些測站四季之間誤差較大。我們一樣透過圖表來呈現資料。

表 4-3-20 季節平均誤差(上)

	仁武	前金	前鎮	大寮	小港	左營
春	7.9	6.58	5.9	6.97	6.91	6.79
夏	5.42	4.81	5.62	5.62	7.1	4.61
秋	7.08	6.68	5.88	6.08	7.34	6.61
冬	7.3	7.84	6.55	9.58	9.72	6.26

表 4-3-21 季節平均誤差(下)

	復興	林園	楠梓	橋頭	美濃	鳳山
春	5.42	8.28	7.41	7.97	7.31	7.97
夏	4.16	6.54	4.91	5.3	4.86	5.3
秋	5.44	8.18	6.02	5.61	7.61	5.61
冬	5.8	8.5	7.99	6.21	9.17	6.21

表 4-3-21 季節平均誤差(下)



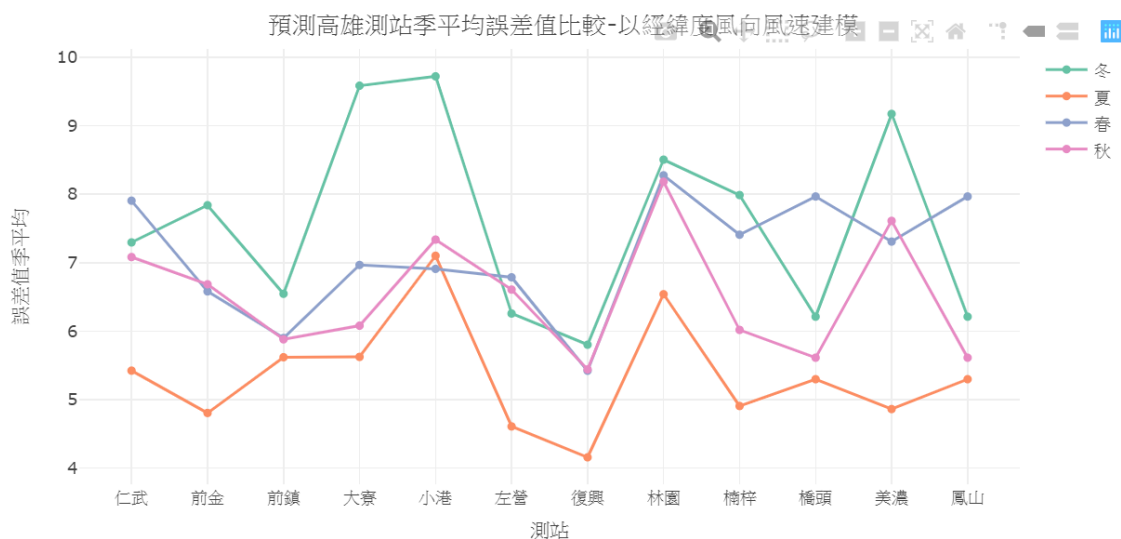


圖 4-3-11 季節平均誤差比較

由圖 4-3-11 可看出大寮以及美濃之平均的誤差值相較之下比較大，可能四季估計的準確度相差較大，而前鎮以及復興之平均誤差值相較之下比較小，可能四季估計的準確度相差較小，接下來我們一樣將春、夏、秋和冬做比較。

表 4-3-22 四季誤差平均

春	夏	秋	冬
7.12	5.36	6.52	7.60

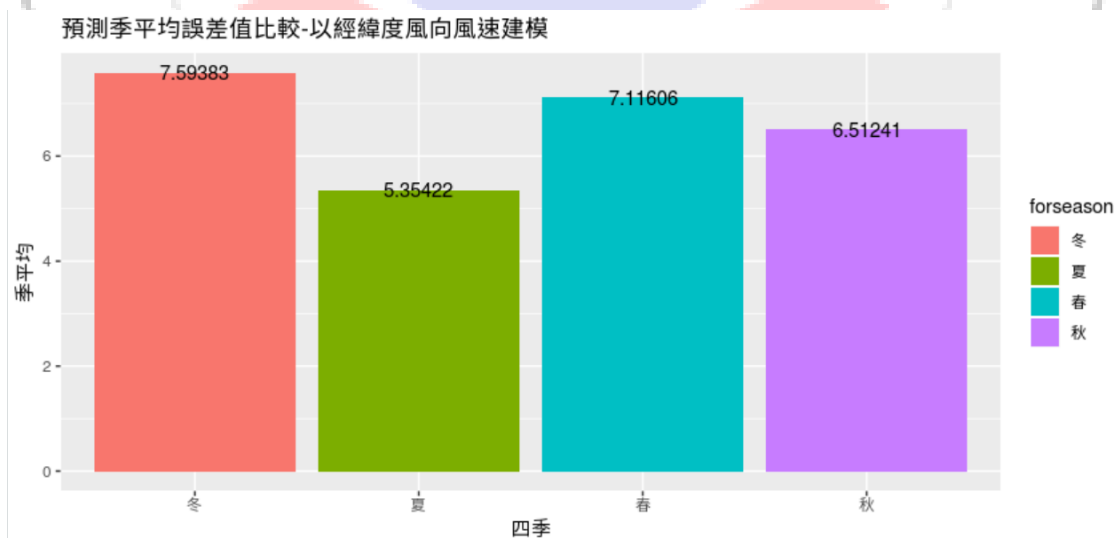


圖 4-3-12 四季誤差平均圖

由此以上表圖整理出後，可以發現冬天和春天時，平均的誤差值相較之下比較大，表示可能估計的相對不準確，而夏天平均誤差值相較之下比較小，表示可能估計出來的數值相對準確些。

### 4.3.5 多變量總整理

這一節總整理是把以經緯度建模、以經緯度和 PM10 建模及以經緯度、風向風速建模這三個模型的平均誤差值做一個整理和比較，分為測站、時段和季節比較，這三個模型在不同條件之間的差異，並可以找出誤差最小的模型。

一、以測站去做比較如下：

表 4-3-23 以測站比較各模型

測站	以經緯度建模	以經緯度、PM10 建模	以經緯度、風向風速建模
仁武	6.49	6.73	7.06
前金	8.3	6.35	6.75
前鎮	6.19	6.22	6.44
大寮	6.01	6.13	6.51
小港	5.64	5.99	6.24
左營	6.05	6.29	6.69
復興	5.74	5.69	5.63
林園	8.3	8.2	8.31
楠梓	6.89	7.06	7.25
橋頭	5.56	5.8	6.23
美濃	7.8	7.76	7.85
鳳山	5.56	5.8	6.23
平均	6.54	6.50	6.77

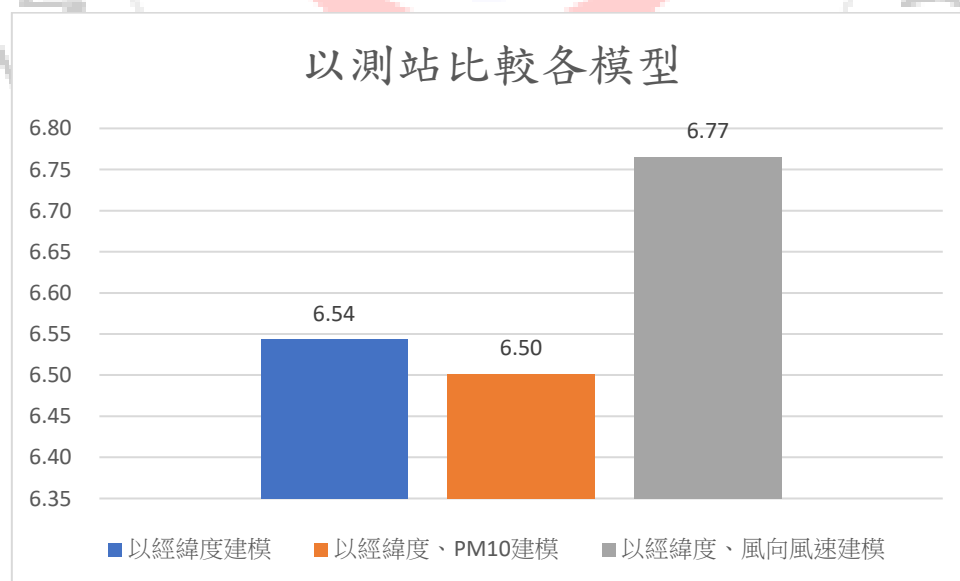


圖 4-3-13 以測站比較各模型

表 4-3-23 中，表中數值表示誤差平均值，紅色表示誤差最大之測站，綠色代表誤差最小之測站。從中我們可以清楚的比較三個模型，然而最大誤差值出現在以經緯度、風向風速建模的模型中，而最小誤差值出現在以經緯度、PM10 建模的模型中。以上我們可以很清楚的觀察出，不管運用哪個方式建立模型，誤差最大值大多都落在林園，也就是位於最南方之測站，我們認為可能是因為四周都無其他測站幫助建模。

二、以時段去做比較如下：

表 4-3-24 以時段比較各模型

	以經緯度建模	以經緯度、PM10 建模	以經緯度、風向風速建模
1	6.29	6.32	6.63
2	6.2	6.29	6.49
3	6.27	6.25	6.42
4	6.14	6.12	6.55
5	6.1	6.12	6.47
6	6.16	6.23	6.4
7	6.11	6.22	6.44
8	6.43	6.51	6.79
9	6.45	6.43	6.82
10	7.27	7.07	7.28
11	7.85	7.8	7.87
12	7.78	7.58	7.98
13	7.84	7.53	8.09
14	7.5	7.27	7.7
15	7.33	7.19	7.23
16	6.71	6.73	7.09
17	6.28	6.26	6.47
18	5.82	5.89	6.15
19	5.84	5.95	6.09
20	5.83	5.77	5.98
21	6.08	5.99	6.23
22	6.15	6.07	6.21
23	6.26	6.2	6.44
24	6.34	6.27	6.56
平均	6.54	6.50	6.77

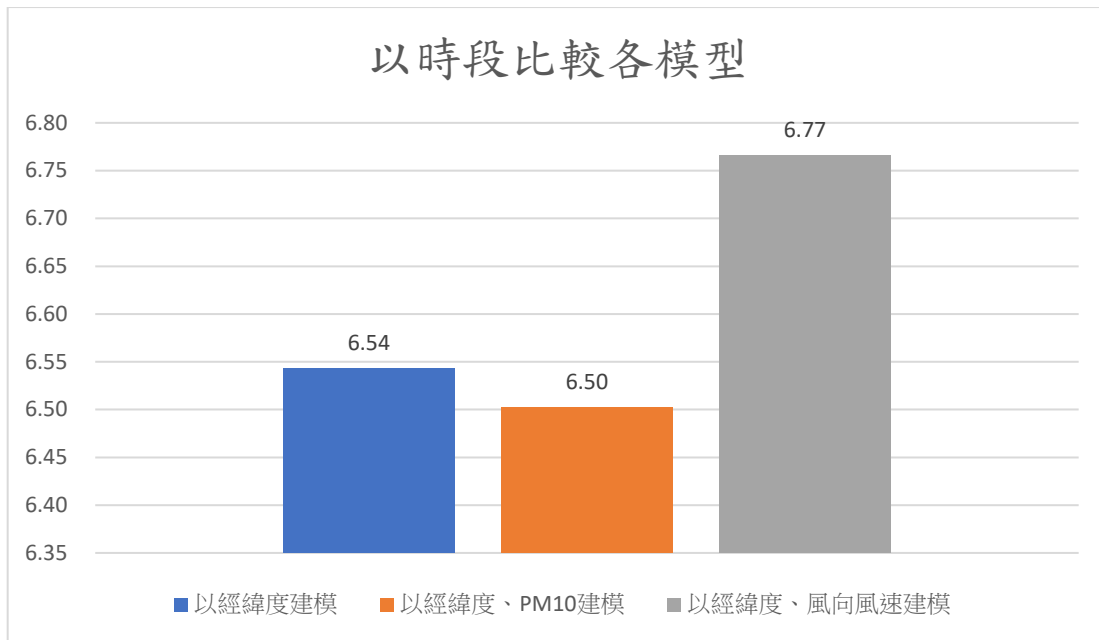


圖 4-3-14 以時段比較各模型

上表中，紅色表示誤差最大之測站，綠色代表誤差最小之測站。從中我們可以清楚的比較三個模型，然而最大誤差值出現在以經緯度、風向風速建模的模型中，而最小誤差值出現在以經緯度建模的模型中。我們發現不管運用哪個方式建立模型，誤差最大值都在 11 點時，表示 11 點時的預測普遍較為不準確，誤差最小值大多落在 18-19 點時，表示這個時段普遍預測得較為準確。

二、以四季去做比較如下：

表 4-3-25 以四季比較各模型

	以經緯度建模	以經緯度、PM10 建模	以經緯度、風向風速建模
春	6.95	7.13	7.12
夏	4.93	5.01	5.36
秋	6.12	6.47	6.52
冬	7.15	7.18	7.6
平均	6.29	6.45	6.65

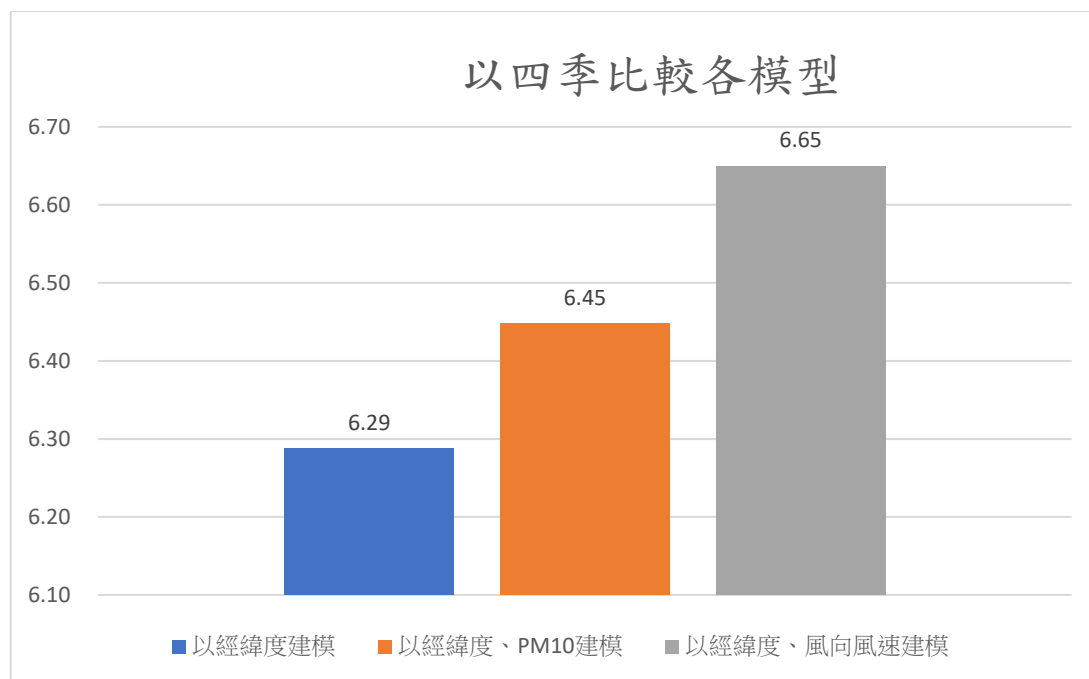


圖 4-3-15 以四季比較各模型

上表 4-3-25 中，紅色表示誤差最大之測站，綠色代表誤差最小之測站。從中我們可以清楚的比較三個模型，然而最大誤差值出現在以經緯度、風向風速建模的模型中，而最小誤差值出現在以經緯度建模的模型中。我們也可以很清楚的看出，不管運用哪個方式建立模型，誤差最大值都為冬天，表示冬天時的預測普遍較為不準確，誤差最小值夏天時，表示這個時段普遍預測得較為準確。

#### 四、總表如下

表 4-3-26 多變量總比較

	以經緯度建模	以經緯度、PM10 建模	以經緯度、風向風速建模
全部平均值	6.54	6.50	6.77

從此表 4-3-26 中我們可以看到以經緯度、PM10 建模之模型估計得較為準確，然而以經緯度、風向風速建模之模型估計得較為不準確。

## 第五章 結論與建議

### 第一節 結論

本研究為高雄地區的 2017 和、2018 年兩年的 PM2.5 來做分析及預測，透過環保署提供的歷年監測資料，我們加以整理及合併，探討這 12 個測站的影響程度，再分析 PM2.5 前我們先區分測站、月分和時段，觀察和彙整他們的相異之處。之後建立模型來預測其他測站的 PM2.5，並算出誤差值，再討論此模型的誤差大小，統整以上所得結果我們將列為以下八點：

一、分析 2017 年和 2018 年高雄地區 12 個測站，我們發現到全部都超出環保署公告的 PM2.5 管制值(年均值為  $15(\mu\text{g}/\text{m}^3)$ )；且橋頭站和楠梓站他測站相比與汙染都較嚴重，由於楠梓站附近都為重工業區，讓 PM2.5 特別嚴重，我們認為這是一個非常關鍵的因素；橋頭站為背景測站，當風速減弱導致擴散條件不佳，就容易累積 PM2.5。

二、關於 PM2.5 在一年之中，統整出以上資料，雖然兩年最高以及最低的月份不太一樣，但我們發現冬季的汙染程度是明顯大於夏季的，位於南部的高雄在冬季時受到大陸冷高壓引響，導致部分擴散不良，在夏季時有西南季風或颱風等因素，有較強的降雨和風速，可以使 PM2.5 不容易累積，相對之下汙染程度就較為低。

三、在一天 24 小時之中，我們也發現到白天的 PM2.5 濃度是大於晚上的 PM2.5，我們認為可能是因為白天大量的汽機車排放量，加上白天氣溫高，當氣溫高 PM2.5 的濃度就會升高，且少了風的流動，就會導致汙染散不去，使 PM2.5 在都市和郊區間反覆循環。反之，晚上的氣溫較低，車流量也不多，所以使得 PM2.5 濃度也較低。

四、由用 PM10、風速以及風向之高斯過程單變量結果中，PM10 建模的誤差值介在  $9.262755 \pm 11.5975$ 、風速建模的平均誤差值為  $12.04394345 \pm 14.93827545$  和風向建模的為  $12.45033155 \pm 15.18721545$ 。由此可知以 PM10 建立模型之誤差值及標準差比較小，可能較準確；相較之下也可看出以風向估計之誤差值以及標準差都較大，可能相較之下估計得比不準確。

五、從高斯過程單變量的結果中可以看得出來，其中仁武、小港以及左營之誤差值和誤差標準差較大，認為應該是估計的比較不準確，而美濃、前金以及大寮之平均誤差值和誤差標準差較小，應該是估計的比較準確。

六、從經緯度的高斯過程雙變量可以看出來，林園以及美濃之平均誤差值相較之下比較大，可能估計的相對不準確，而小港以及鳳山之平均誤差值相較之下比較小，可能估計出來的數值相對準確。我們推斷是因為地理位置的疏密影響。

七、從雙變量只有經緯度到多變量加上了 PM10 之後，發現誤差值變小了，所以我們認為多了一個 PM10 這個變數，是可以讓模型預測得更準確些；而經緯度加上風速和風向共四個 X，好像效果沒有比經緯度加上 PM10 的誤差來的小，但如果只是跟有單獨雙變量的經緯度比較，經緯度加上風速和風向的結果還是有讓誤差變小。

八、我們發現誤差低的都是本身測站 PM2.5 也比較低的，例如我們分析 PM2.5 的時段時，研究出是白天比晚上嚴重，而在預測雙變量或多變亮時，誤差值也是白天大於晚上，季節也是，冬天的誤差值都大於夏季，所以我們猜想污染高低可能會有些影響預測結果。

## 第二節 建議

一、本研究觀察 2017 和 2018 高雄 PM2.5 各測站年平均值，發現基本上都有下降的趨勢，未來可以再統計今年 2019 的情況，是否也有持續下降。且 24 小時平均也有比較低，這些都是可以列入待觀察的目標。

二、目前在高雄市只有 12 個測站，位於較偏遠的地區(例如：最北部的美濃和最南部的林園)利用經緯度估計出來的數值較不準確，但這些地區仍為 PM2.5 之嚴重污染測站，建議高雄市可以在較偏遠的郊區興增測站。而且林園站監測站裡面的數值很多的 NA 值，這對於估計的數據有相對的影響，希望高雄市能更加的重視較偏遠地區的監測站。

三、我們使用經緯度預測 PM2.5，我建議可以擴大範圍，例如台南和屏東這些鄰近的縣市，不單單只有高雄地區，加上風向和風速的因素共同考量，可以尋找是否也有關聯。

四、本研究關於預測 PM2.5 的方法，因為我們是用現有的資料，例如我是用 2017 年 1 月 1 日的 PM10 來預測當時的 PM2.5，但這樣的預測方法只能知道這模型準不準確，還沒到真正有預測到下一小時，所以我建議這套模型和資料可以再多加變化，讓預測更有意義。

五、由於本研究預測出來的誤差質都偏大，建議如果之後還是用這方法研究時，可以改變研究目標，例如適用大範圍的台北市來預測信義區等，不再是小範圍的一個點，可以研究是否誤差會比較小。



## 參考文獻

英文部分：

- Alpaydin. (2004). Introduction to Machine Learning. Retrieved from <https://www.cmpe.boun.edu.tr/~ethem/i2ml/>
- Cifuentes, L. A., Vega, J., Kopfer, K., and Lave, L. B. (2000). Effect of the fine fraction of particulate matter versus the coarse mass and other pollutants on daily mortality in Santiago, Chile. *Journal of the Air & Waste Management Association*, 50(8), 1287-1298. doi: 10.1080/10473289.2000.10464167
- Devlin, R. B., Ghio, A. J., Kehrl, H., Sanders, G., and Cascio, W. (2003). Elderly humans exposed to concentrated air pollution particles have decreased heart rate variability. *European Respiratory Journal*, 21(40), 76-80. doi: 10.1183/09031936.03.00402403
- Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L. and Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*, 295(10), 1127. doi: 10.1001/jama.295.10.1127
- Gramacy, R.B. (2007). tgp: An R Package for Bayesian Nonstationary, Semiparametric Nonlinear Regression and Design by Treed Gaussian Process Models. *Journal of Statistical Software*, 19(9), 1-46. doi: 10.18637/jss.v019.i09
- Guizilini, V., and Ramos, F.A., (2015). A Nonparametric Online Model for Air Quality Prediction, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence.
- Kalaitzis, A., A. Honkela, P. Gao and N. D. Lawrence (2014). gptk: Gaussian Processes Tool-Kit. R package version 1.08. Retrieved from <https://CRAN.R-project.org/package=gptk>.
- MacDonald, B., Ranjan P., and Chipman H. (2015). GPfit: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs. *Journal of Statistical Software*, 64(12), 1-23. doi: 10.18637/jss.v064.i12
- Nery, M.C.D. (2018). One third of global air pollution deaths in Asia Pacific. Retrieved from <https://www.who.int/news-room/detail/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action>.

- Pope C.A. 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. (2002). Lung Cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*, 287(9), 1132. doi: 10.1001/jama.287.9.1132.
- Rasmussen, C.E. and Williams, C.K.I. (2005). *Gaussian Processes for Machine Learning*. doi: 10.7551/mitpress/3206.001.0001
- Seaton, A., MacNee, W., Donaldson, K., and Godden, D. (1995). Particulate air pollution and acute health effects. *The Lancet*, 345(8943), 176-178. doi: 10.1016/S0140-6736(95)90173-6
- Taylor. (2018). 90% of world's children are breathing toxic air, WHO study finds. Retrieved from <https://www.theguardian.com/environment/2018/oct/29/air-pollution-worlds-children-breathing-toxic-air-who-study-finds>.

中文部分：

- 安靈雅(2014)。基於非等向性高斯進程迴歸之運用時空關係內插機制於感測器網路。國立臺灣科技大學，臺北市。
- 林冠名(2018)。在大數據平台使用機器學習方法預測空氣汙染。國立台北大學，臺北市。
- 俞淑惠(2014)。細懸浮微粒觀測數值時間序列分析及測站代表性研究。102 年度「環保署/國科會空污防制科研合作計畫」期末報告。
- 張艮輝(2017)。強化空氣品質模式制度建立計畫(第 2 年)。
- 陳明輝、周錦慧、周佳穎(2016)。應用廣義加法模型於懸浮微粒 PM10 的研究，銘傳大學 2016 追求高等教育卓越國際學術研討會。
- 陳奕愷(2012)。應用迴歸分析與類神經網路於懸浮微粒之研究。國立中興大學，臺中市。
- 陳亭蓉(2006)。以高斯過程迴歸分析臺灣中部山區颱風降雨量之空間分佈。國立臺灣科技大學，臺北市。
- 陳騰瑞(2013)。應用類神經網路預測雲嘉南空品區懸浮微粒與臭氧趨勢之研究。私立環球科技大學，雲林縣。
- 傅文閣(2018)。2012-2016 年高雄市 PM2.5 汙染分布展建構 PM2.5 之預測模型。國立中興大學，臺中市。
- 黃翊涵(2014)。中彰投地區空氣品質影響因子分析。私立大葉大學，彰化縣。

蔡効耕(2015)。基於多模型高斯過程回歸分析之車流速度預測。國立臺灣科技大學，臺北市。

環保署(2018)。空氣品質預報作業。2019 年 8 月 19 日取自行政院環保署  
<https://taqm.epa.gov.tw/taqm/tw/b0204.aspx>。

環境資訊中心 TEIA(2019)。2019 年 1 月 3 日取自環境資訊中心 TEIA。  
<https://e-info.org.tw/node/215820>

簡家宏(2004)。應用基因類神經網路於空氣品質短期預測及監測資料異常值診斷之研究，以台中縣沙鹿測站為例。國立雲林科技大學環境，雲林縣。

