

時間序列模型之比較：以新北市空氣污染 PM2.5 指數預測為例

王建智、簡均樺

摘要

由於氣候劇烈變化與工業迅速發展，空氣污染所造成對於人民生活品質及健康的影響，已成為無法忽視的重大議題。根據衛生福利部指出，空氣污染的項目中以細懸浮微粒(Particulate Matter 2.5：PM2.5)對於人體健康影響最大。若可以即時預測預測未來數小時 PM2.5 指數，則能夠減少空污對民眾健康損害並降低社會成本的消耗。本研究透過新北市新莊空氣品質監測站的 PM2.5 指數資料為例，比較傳統時間序列 ARIMA 模式及深度學習類神經網路 LSTM 於 PM2.5 指數的預測成效。研究結果顯示 LSTM 估計的 RMSE 低於 ARIMA 模式，結論將藉由模式比較說明兩者演算法差異以及使用時機，以提供後續建模參考。

關鍵詞：空氣污染、預測、整合移動平均自回歸模型、長短期記憶模式

王建智(通訊作者)，明志科技大學工業工程與管理系教授。E-mail: ieccwang@mail.mcut.edu.tw

簡均樺，明志科技大學工業工程與管理系碩士生。E-mail:a0952261106@gmail.com

The study of comparison of time series models: New taipei city air pollution PM2.5 index forecast as an example

Chien-Chih Wang & Chun-Hua Chien

Abstract

Due to drastic climate change and industrial development, the impact of air pollution on people's quality of life and health has become a major issue that cannot be ignored. According to the Ministry of Health and Welfare, the Particulate Matter 2.5 (PM2.5) is the most serious form of air pollution affecting human health. If the PM2.5 index can be predicted in real time in the next few hours, it can reduce the damage to people's health caused by air pollution and reduce the consumption of social costs. This study uses the PM2.5 index data from the Xin-Zhuang Air Quality Monitoring Station in New Taipei City to compare the prediction performance of the traditional time series ARIMA model and LSTM for PM2.5 index. The results show that the RMSE estimated by LSTM is lower than that of the ARIMA model. The conclusion will be based on the model to compare the differences in algorithms and the motivations used, and to provide a reference for subsequent model building.

Keywords: air pollution, prediction model, integrated moving average regression model, short- and long-term memory model

Chien-Chih Wang (Corresponding Author), Professor, Industrial Engineering and Management, Ming Chi University of Technology, E-mail: ieccwang@mail.mcut.edu.tw

Chun-Hua Chien, Master student, Industrial Engineering and Management, Ming Chi University of Technology, E-mail: a0952261106@gmail.com

壹、緒論

近年來即時監控空氣品質變化一直都是社會關注的焦點，世界衛生組織(World Health Organization, WHO)國際癌症研究中心(International for Research on Cancer, IARC)發布一份報告，說明室外空氣污染是造成癌症主要環境因素，並宣布將空氣污染對人類致癌的第一級致癌因。懸浮微粒(Particulate Matter, PM)為室外空氣污染主要成分，也被列為對人類第一級致癌物，其中 IARC 負責人 Christopher Wild 表示，當人類暴露在空氣污染的環境中越長，罹患肺癌的風險越大，2010 年全球已有 223,000 件死於肺癌的肇因於空氣污染(IARC,2013)。PM_{2.5} 泛指空氣中直徑小於 2.5 微米(μm)的 PM 粒子，PM_{2.5} 成分包含鉛、砷等重金屬元素及戴奧辛等有毒物質，對健康影響高於其他污染物。由於 PM_{2.5} 可以穿透肺部氣泡，直接進入血管中，隨著血液循環影響至人類各個器官。許多研究證實 PM_{2.5} 會導致氣喘、肺部等呼吸系統傷害以及心臟病、動脈粥樣硬化等心血管疾病(高雄市空氣污染防制，2020)。台灣肺癌醫學會指出，台灣肺癌女性比例中，在既定印象中廚房油煙或二手菸是主要引發肺癌之主因；現今，已研究顯示 PM_{2.5} 已小到比毛細孔還要細，長期居住在 PM_{2.5} 濃度 35 微克的環境下，肺癌患者死亡率會提升至 15%，「WHO 研究顯示，PM_{2.5} 濃度每增加 10 微克，死亡率會再增加 8%」，養成查詢空氣品質標準(Air Quality Indicator；AQI)，一旦超過警戒值，就要盡量減少待在戶外的時間(行政環境保護署，2020)。鑒於上述 PM_{2.5} 對於人類健康威脅，WHO 在 2015 年世界衛生大會(World Health Assembly, WHA)中，通過決議並督促所有會員國，發展空氣品質監測系統，以監控並降低空氣污染對人體健康造成傷害(衛生福利部，2016)。

行政院環境保護署在全台各地均設有空氣品質監測站，依據各地區污染源排放資料、氣象及空氣品質濃度分布資料等，分別為一般空氣品質監測站(60 站)、交通空氣品質監測站(鳳山、三重、中壢、永和、復興、大同，共 6 站)、工業空氣品質監測站(頭份、線西、麥寮、臺西、前鎮，共 5 站)、國家公園空氣品質監測站(恆春、陽明共 2 站，其中恆春站兼一般測站)、背景空氣品質監測站(萬里、觀音、三義、橋頭、富貴角，共 5 站；其中萬里、三義 2 站兼一般測站)及其他空氣品質監測站(埔里、關山)。環保署為使民眾方便查詢該區域空氣品質指標監測空氣品質項目有二氧化硫(SO₂)、一氧化碳(CO)、二氧化碳(CO₂)、臭氧(O₃)、細懸浮微粒(PM_{2.5})及懸浮微粒(PM₁₀)等。為使民眾方便了解地區空氣品質，一般是以 O₃ 八小時值來計算，並且將空氣品質指標(AQI)

依據對人體危害來做顏色做區分，來了解空氣品質狀態已能夠選擇適當活動空間，來做好保護措施。行政院主計於 2016 年發布國民幸福指數年報(行政院,2016)指出，台灣空氣污染指標相較經濟合作暨發展組織(Organization for Economic Co-operation and Development; OECD)中國家排名高居第三，顯示台灣在空氣污染防治上還有改進空間。

行政院環保署於 2012 年修正「空氣品質標準」，考量 PM_{2.5} 對國人健康影響，專家學者建議應將其納入空氣污染管制為優先考量，並訂定 PM_{2.5} 空氣品質標準為二十四小時平均值 35 $\mu\text{g}/\text{m}^3$ 、年平均 15 $\mu\text{g}/\text{m}^3$ ，為目前國際間納入法規規範眾較為嚴格的標準(環境工程會刊，2013)。此外，行政院環保署於 2017 年修正「空氣品質嚴重惡化緊急防制辦法」，規定中央主管機關應按日發布空氣品質狀況及預測資料，若二十四小時平均值達 35.5 $\mu\text{g}/\text{m}^3$ 應發布預警警告；二十四小時平均值達 150.5 $\mu\text{g}/\text{m}^3$ 則應發布嚴重惡化警告(全國法規資料庫，2017)。

由上述規定可知，如何精準預測 PM_{2.5} 濃度，是政府是否能夠發布正確空氣品質警告之關鍵因素。若預報低估 PM_{2.5} 濃度可能會使民眾健康面臨風險，因此如何有效且正確預測各地區空氣中 PM_{2.5} 濃度，便成為有關當局重要議題。然而，目前鮮少人針對新北地區之 PM_{2.5} 月平均濃度進行預測研究。新北市新莊區在全國 368 個鄉鎮市區人口高居全國第三 (新北市政府民政局，2016)。本研究為了解新北市新莊區 PM_{2.5} 濃度趨勢，使用行政院環保署「環境資源資料庫」中，PM_{2.5} 及相關變數之歷史數據，建立預測模型。

本研究分別比較傳統時間序列 ARIMA 模式及深度學習類神經網路 LSTM 於 PM_{2.5} 指數的預測成效。藉由模式比較說明兩者演算法差異以及使用時機，並提供後續建模參考說明。

貳、文獻探討

一、細懸浮微粒的定義

美國環保署將細懸浮微粒(PM_{2.5})係指空氣中粒徑(aerodynamic diameter)等於或小於 2.5 微米(μm)的粒狀污染物，亦稱「細懸浮微粒」。由於粒徑小 PM_{2.5} 具有更強大的穿透力且易吸附有毒物質，容易經由人體呼吸進入至肺部，或是由血管中循環至全身，可能都會導致人體器官不同之危害影響身體健康。PM_{2.5} 來源可分為原生性(primary)和衍生性(secondary)；原生性細懸浮微粒是直接從自然與人為活動所排放，未經過化學反應生成；衍生性細懸浮微粒是由自然與人為活動排放到大氣環境中 SO₂、NO_x、

VOCs 與 NH₃ 等氣態前驅物，在大氣中經過複雜化學反應形成，其反應過程包括氣相光化學反應、液相反應與氣固相反應與不同反應物間之競合作用等。因細懸浮微粒濃度與前驅物排放量之間有顯著非線性關係，都可能由自然或人為而產生，如表 1 所示(楊宏隆，1998)。

表 1 細懸浮微粒的來源

來源 區分	自然界產生	地殼、火山爆發、岩石等。
	人類行為產生	移動源、工業排放和石化燃料等廢棄燃燒行為。
性質 區分	原生性 PM _{2.5}	從自然及人為活動直接排放，在大氣環境之未經化學反應之微粒，如營建工地粉塵、天然之海鹽飛沫、車行揚塵和工廠直接排放。
	衍生性 PM _{2.5}	自然及人為活動排放到大氣環境中之化學物質，經過太陽光照或其他化學反應後生成，如燃油、燃氣及燃煤電廠、石化相關產業工廠、煉鋼廠、禽畜排泄、農業施肥、生活污水、機動車輛、建物塗料及船舶等。

原生性及衍生性 PM_{2.5} 來源比例約為 50%，顯示原生性與衍生性來源的管制與境外影響傳輸影響的掌握與改善，然而衍生性 PM_{2.5} 來源，由上述所說化學物質經過光化學反應後而生成，一般所見之粒狀物質種類如表 2 所示。化學成分組成包含金屬元素、有機碳類、硫、梨子物質、硝酸鹽類等(陳維新、江金龍，2009)。

表 2 粒狀物質種類

名詞	形成機制與說明	尺寸 (μm)
飛灰	由不可燃燒的微粒所組成，其長時間存在煤廢棄物焚化或燃燒後的氣體中。	1~300
灰塵	主要來自機械操作或物質的生成處理過程。	1~1000
煙	由碳氫化合物如木材、煙草或煤等在不完全燃燒下所產生的固體微粒。	0.05~1
燻煙	由於化學反應或蒸發，昇華而後凝結成固體顆粒，通常為金屬氧化物。	0.03~0.3
霧及霧氣	因化學反應或蒸氣凝結所形成的液滴稱為霧，而如果霧的濃度高達阻礙視線，則稱為霧氣。	<10

二、細懸浮微粒危害

空氣微粒會依據不同粒徑大小，可經由呼吸進入身體不同的部位。落塵粒徑約在 100 微米(μm)以上，無法長期漂浮於空氣中會逐漸沉降於顆粒。對於所有粒徑的粒狀污染中，1~5 μm 粒狀污染是值得重視。當人體呼吸時，較大粒徑顆粒 (>5 μm) 非常容易藉由氣管纖毛運動排出或被鼻毛截留，因此對於人體危害性較小；當小顆粒

($<1\mu\text{m}$) 進入呼吸器官時，較易隨呼出的氣體排出。但對於 $1\sim 5\mu\text{m}$ 粒徑的粒狀污染物則非常容易停留於肺泡 ($1\sim 3\mu\text{m}$) 或呼吸道及支氣管 ($3\sim 5\mu\text{m}$) 中。圖 1 說明細懸浮微粒 ($\text{PM}_{2.5}$) 對人體健康之嚴重性(高雄市空氣污染防制, 2020)。

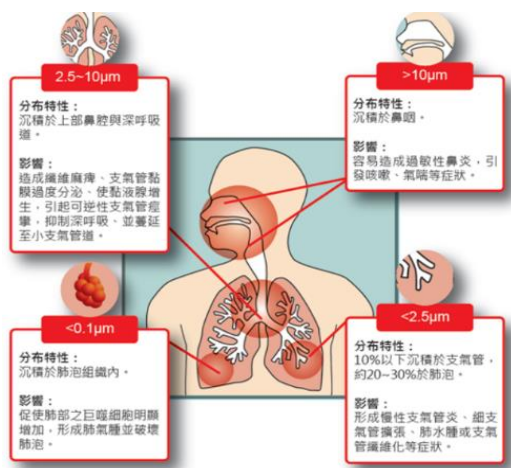


圖 1 吸入懸浮微粒沉積位置圖

(圖片來源:行政環境保護署「空氣品質改善維護資訊網」)

三、台灣近年死因統計

依據世界衛生組織(WHO)第十版，國際疾病分類標準(ICD-10)及死因選碼準則，統計台灣 107 年死因。如表 3 所示 107 年十大死因(死亡率為每十萬人口)，依死亡率排序，107 年十大死因依序為(1)惡性腫瘤(癌症)(2)心臟疾病(3)肺炎(4)腦血管疾病(5)糖尿病(6)事故傷害(7)慢性下呼吸道疾病(8)高血壓性疾病(9)腎炎、腎病症候群及腎病變(10)慢性肝病及肝硬化，排名順位自 105 年以來維持相同。

表 3 十大癌症死亡人數及死亡率

	死亡人數 (人)		死亡率 (每十萬人口)			
	107年	較上年增減%	106年 順位	107年 順位	106年	較上年增減%
所有死亡原因	48,784	1.6			203.9	1.5
氣管、支氣管和肺癌	9,388	1.7	1	1	39.2	1.5
肝和肝內膽管癌	8,222	-2.1	2	2	35.7	-2.2
腸、直腸和肛門癌	5,823	0.2	3	3	24.7	-
女性乳癌	2,418	1.7	4	4	20.1	1.5
口腔癌	3,027	6.5	5	5	12.1	5.8
前列腺(攝護腺)癌	1,377	-1.1	6	6	11.9	-0.8
胃癌	2,299	-0.2	7	7	9.8	-1.0

表 3 十大癌症死亡人數及死亡率(續)

	死亡人數 (人)		死亡率 (每十萬人口)			
	107年	較上年增減%	106年順位	107年順位	106年	較上年增減%
胰臟癌	2,292	10.1	8	8	8.8	10.2
食道癌	1,929	7.3	9	9	7.6	7.9
子宮頸及部位未明示子宮癌	653	0.3	10	10	5.5	-

附註:

1. 107年年中人口數統計 23,580,080 人，男性 11,716,247，女性 11,863,833 人。
2. 106年年中人口數統計 23,555,522 人，男性 11,719,425，女性 11,836,097 人。
3. 標準化死亡率係以 2000 年 WHO 之世界標準人口數為準。
4. (1)每十萬女性人口死亡率；(2)每十萬男性人口死亡率。

107 年癌症死亡人數為 4 萬 8,784 人，占有所有死亡人數 28.2%，死亡率每十萬人口 206.9 人，較上年上升 1.5%，癌症多集中於 55 歲以上之族群，107 年占 8 成 5，並依照十大癌症死亡率依序為(1)氣管、支氣管和肺癌(2)肝和肝內膽管癌(3)結腸、直腸和肛門癌(4)女性乳癌(5)口腔癌(6)前列腺(攝護腺)癌(7)胃癌(8)胰臟癌(9)食道癌(10)子宮頸及部位未明示子宮癌，順位和 106 年相同(衛生福利部，2018)。在 2014 年，世界衛生組織對於空氣汙染發布警訊，顯示空汙已造成全球 7 百萬人死亡，當中低收入到中等收入的東南亞國家以及西太平洋地區人民，是空氣汙染最大的健康受害者，估計有 330 萬人死於室內空氣汙染、260 萬人死於戶外空氣汙染，這個數字遠高於 2004 年的統計資料。因戶外空氣汙染所致的相關死因，40%為缺血性心臟病(Ischemic heart disease)、40%為中風 (stroke)、11%為慢性阻塞性肺病(COPD)、6%為肺癌(lung cancer)，另 3%為兒童急性下呼吸道感染。2014 年發表於 Environmental Health Perspectives，蒐集 18 個相關研究進行戶外空氣汙染與肺癌的綜論分析，研究顯示 PM_{2.5} 暴露導致罹患肺癌的風險為 1.09，PM_{2.5} 產生肺腺癌的風險是 1.40 (WHO，2014)。

參、研究方法

本研究建立 ARIMA 與 LSTM 模型來預測新北市 PM_{2.5} 濃度值，於環保署空氣品質監測資料庫(圖 2)，下載新北市新莊空氣監測站 PM_{2.5} 濃度值資料，資料期間於 2019 年 1 月 1 日至 2020 年 3 月 31 日，共 410 筆資料，並填補缺失值。建立預測模型透過預測評估指標找出最佳預測方法。



圖 2 環保署空氣品質監測資料庫

時間序列自變數為時間，應變數為各時間點所應對的數量或數值，利用時間序列對未來的數值能經由過去的數值來估計 Donald S. Remer and Craig Jorgens (1978)，時間序列是依事件或資料先後次序，並將統計的數據排序，反映出發展過程、方向和趨勢，進行預測。廣泛存在於金融預測、天氣預報、災害預測及其他時間實用工程領域和科學研究。

一、ARIMA 時間序列模型

自我回歸整合移動平均模式 (Auto Regression Integrated Moving Average, ARIMA) 由 Box 和 Jenkins(1976) 所提出時間序列預測方法，又稱 Box-Jenkins 模型。主要是對過去及現在資料分析檢視其自我相關與偏自我相關函數等特性進行認定、估計、診斷三階段模式建構過程，配適最佳模式並進行資料分析預測 (楊浩彥、郭迺峰、林政勳 2013)。

ARIMA(p,d,q)模式經由 p、d、q 三個引數用整數值來代替，藉著 p、d、q 的變化而產生無數多模式，必須找出最佳模型來描述此時間序列。AR 是自迴歸 p 為自迴歸項；MA 為移動平均，q 為移動平均項數；d 為時間序列成為平穩時所做的差分次數。時間序列是從分解法、平均法、指數平滑法及 AR 系列模式，運用機器、深度學習歸類在時間序列分析數據工具上，再執行需求預測。預測工具方式眾多，準確度亦不相同，藉著對於研究之事物觀察，分析及比較各種方法優異性，提供相關建議，研究成果才算發揮功效。本研究將 ARIMA(p,d,q)預測模型調整步驟如下：

- Step1：檢視資料是否為定態序列：

在進行時間序列分析前，需先判定數列是否為定態(Stationary)，當序列資料呈現定態時，其序列會在長期平均值附近波動，若序列不穩定則 d 次差分(difference)後才能成為定態序列，以 I(d)表示。本研究利用自我相關函數圖 (Autocorrelation Function ,ACF)、偏自我相關函數(Partial Autocorrelation Function ,PACF)與單根檢定法

(Dickey and Fuller)來判斷序列是否為定態。

● Step2：單根檢定決定差分(d)次數

單根檢定(Unit Root Test)主要目的在於確定變數之間時間序列的整合級次，藉以判定時間序列是否達到定態性質。由於大多時間序列具有自我相關與異質變異之特性，因此本研究採用 Dickey(1979)提出 ADF(Augmented Dickey Fuller Test)及檢定 KPSS (Kwiatkowski-Phillips-Schmidt-Shin, 1992)檢定，來判斷序列是否為定態。

(一)ADF 的檢定模型如下：

單根檢定法由 Dickey and Fuller(1979)所提出的 DF 單根檢定法，主要針對 AR(1)模式檢定序列的定態性，但未考慮到殘差效果可能存在自我相關，因此發展出 ADF 單根檢定法，共有以下三種模式如下：

模式一：沒有截距項、沒有趨勢

$$\Delta Y_t = \beta Y_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta Y_{t-i} + \varepsilon_t \quad (3.1)$$

模式二：有截距項、沒有趨勢

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta Y_{t-i} + \varepsilon_t \quad (3.2)$$

模式三：有截距項、有趨勢

$$\Delta Y_t = \alpha + \gamma t + \beta Y_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta Y_{t-i} + \varepsilon_t \quad (3.3)$$

其中 Δ 為一階差分， α 為截距項， β 為自我迴歸係數， t 為時間趨勢項， p 表示自我迴歸的落後期數， ε_t 為殘差項並且服從白噪音過程。其假設檢定如下：

虛無假設 H_0 ： $\beta=0$ (序列為非定態序列，具有單根)

對立假設 H_1 ： $\beta < 0$ (序列為定態序列，不具有單根)

當單根檢定結果不拒絕虛無假設 H_0 ，則表示此序列為非定態序列且具有單根；則當檢定結果為拒絕虛無假設 H_0 ，表示序列定態序列且不具有單根。雖然 ADF 單根檢定沒有考慮到殘差是否存在著自我相關及異質變異的問題，但仍然有低檢定力的問題。因此除了 ADF 較常見的單根檢定之外，本研究加入在不同虛無假設的 KPSS 檢定。KPSS 檢定與 ADF 檢定最大的差異為虛無假設不同，KPSS 檢定的虛無假設 H_0 為「變數為定態」說明 H_0 不具單根。進行單根檢定時，利用不同的檢定方法可以更精準的判定非定態序列轉換為定態序列的差分次數。

(二)KPSS 的檢定模型如下：

Kwiatkowski et al., (1992)提出以變數服從定態程序作為虛無假設，由於大部分的單根檢定其虛無假設都設定序列具有單根，則對立假設為序列不具有單根，但這些單根

檢定力並不高，KPSS 檢定則提供一個反向的檢定方法。

KPSS 單根檢定假設變數是由一個定項趨勢、隨機漫步程序(random walk)以及一個定態白噪音(white noise)所組成

$$y_t = \xi_t + \mu_t + \varepsilon_t \quad (3.4)$$

其中， ε_t 是一個定態程序， ξ_t 是一個 random walk， $\mu_t = \mu_{t-1} + u_t, u_t \xrightarrow{iid} N(0, \sigma_u^2)$ 。此時虛無假設為 $H_0: \sigma_u^2 = 0$ ，對立假設 $H_1: \sigma_u^2 > 0$ ，在虛無假設可推導出 KPSS 檢定統計量為

$$LM = \sum_{t=1}^T \frac{S_t^2}{\hat{\sigma}_\varepsilon^2} \quad (3.5)$$

S_t^2 是指殘差的累積總合， $\hat{\sigma}_\varepsilon^2$ 為殘差變異數的估計值。KPSS 檢定的虛無假設為 $H_0: \sigma_u^2 = 0$ ，假設變數為定態，因此無法拒絕虛無假設，表示序列為定態，亦指資料沒有單根；反之，拒絕虛無假設，表示資料為非定態，須將資料做差分直到呈現定態為止。

● Step3: 判斷 ARIMA(p,d,q) 的落後期 p 與 q

在實際應用上，時間序列會呈現不同形態，並且需搭配參數以配適出較合適的模型。1970 年 Box-Jenkins Methodology 提出定態時間序列 ARMA 線性模型，當原始序列具單根為非平穩序列，利用 ARMA 模型精簡配適模型參數。本研究使用 ACF 與 PACF，作為最基本方法與圖形初步判斷自我迴歸模型(Autoregressive model, AR)、移動平均模型(Moving average model, MA)或自我迴歸混合移動平均模式(Auto Regression Mixed Moving Average Model, ARMA)。為避免人為判斷產生誤差，在實務上常用試誤 (Trial and Error) 的方法，藉由利用參數估計選取最小之 AIC 與 BIC 值，配適出最佳模型。在試誤的過程中，將 p 與 q 值階次皆限制在 3 以下(蔡卓翰，2010)。

AR(p)的一般化模型為：

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t$$

(3.6)

α_0 為常數截距項，p 為落後期數， α_i 為 y_{t-i} 的係數， ε_t 為白噪音。AR(p)模型描述當期 y_t 變數和過去部分 p 期的 y 變數有關係。

MA(q)一般化模型為：

$$y_t = \alpha_0 + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

(3.7)

α_0 為常數截距項，q 為落後期數， b_j 為 ε_{t-j} 的係數， ε_t 為白噪音。MA(q)模型描述現在

的 y_t 變數和過去部分 q 期的誤差項影響之關係。

ACF 圖衡量序列指數遞減或正負之間遞減的趨勢，即一個序列與其序列差距 k 步之相關係數稱 ACF(k) 來判別移動平均模型階次，PACF 來判別自我迴歸模型階次，以下表為 ACF 與 PACF 圖形判斷參考依據：

表 4 判斷 ARMA(p,q) 參考依據

Model	典型 ACF 型態	典型 PACF 型態
AR(p)	漸漸消失	落後 p 期後截斷
MA(q)	落後 q 期後截斷	漸漸消失

● Step4：ARIMA(p,d,q) 模型選擇

當有多組 ARIMA(p,d,q) 都符合檢定時，決定落後期 p 與 q 顯得十分重要。若選擇太長落後期會使得參數過度化，導致預估的結果缺乏效率；若選擇太短落後期，會使得參數過度簡化，則造成估計結果產生誤差。所以需要從中選擇最適當落後期數作為預測模型，來衡量模型之配適度，並且判斷配適度以及找出最佳落後期數。AIC 與 BIC 是較常使用準則：

Akaike 訊息準則(Akaike's Information Criterion, AIC)

$$AIC = 2k + \ln(SSE)$$

(3.8)

Bayesian 訊息準則(Bayesian Information Criterion, BIC)

$$BIC = N \ln(SSE) + k \ln(N)$$

(3.9)

其中 k 表示模式中要估計的參數數目， N 為總樣本數， SSE 為殘差平方和，選擇最佳模式會以 AIC 及 BIC 越小越好為標準。一般而言，小樣本資料會使用 AIC 來評估模型配適度，而大樣本資料會使用 BIC 來評估模型之配適度。

● Step5: 檢查殘差是否為白噪音(White noise)

在統計時間序列中，殘差分析是用來檢視是否滿足線性迴歸假設，檢查誤差是否服從常態分配。檢查白噪音視為一種時間序列的隨機變數，亦可以說明白噪音為平均數等於 0 與變異數為常數的假設，且具有獨立性(independently and identically distribute, 簡稱 iid)，檢定假設如下：

$$H_0: \text{殘差服從白色雜訊}$$

H_1 : 殘差不服從白色雜訊

本研究在診斷殘差時，利用標準化殘差圖、常態機率圖、殘差值方圖來檢測模型，若殘差圖落在以 0 為中心線，則表示此模式的殘差為白色的隨機過程；反之，檢定結果為否，則表示此模型不合適必須重新選擇參數。

(一)標準化殘差圖(Standardized residual plot):

說明殘差與數據順序圖中各數值隨機分佈，當序列不會顯示趨勢、週期性等模式，可以驗證殘差獨立。

(二)常態機率圖 (Normal Quantile-Quantile plot):

較常使用檢定方法有單一樣本檢定(Kolmogorov-Smirnov)及卡方適度檢定(Goodness of fit test, χ^2)。在圖形判斷方面有直方圖(Histogram)，缺點為當樣本少時不適用；盒子圖(Box plot)與常態機率圖等(Normal Quantile-Quantile plot)，簡稱 Q-Q plot，樣本的累積頻率分佈與理論正態分佈的累積機率分佈之間的關係，若圖形上的點接近於一條 45° 的直線時，表示其滿足常態分配的假設。在本研究中，使用常態機率圖來判斷殘差是否有滿足常態分配假設。

(三)殘差直方圖

檢查殘差是否服從正態分佈，由於直方圖的外觀會根據對數據進行分組所用的區間數值間變化，因此使用正態機率圖與擬合配適度檢驗來評估殘差是否為正態。

二、長短期記憶(LSTM)預測模型

Sepp and Jurgen(1997)提出長短期記憶模型(Long Short-Term Memory, LSTM)是回饋式神經網路(RNN)的一種變化形式，可以解決梯度下降法(Gradient descent)問題。依據錯誤修改每次的權重，演算法式子經過不斷改進訓練後，從輸出值影響每一階段輸入值，直到誤差值被過濾找出最適合的權重比例，將參數輸出使循環神經網路具有更好的記憶性能內部神經元的運算結構變得完整。LSTM 在處理與預測時間序列相關數據時，其預測出來的無論是誤差還是成果皆優於一般的遞歸類神經網路模型。

時間序列為一組按時間排序之序列，但為可以將時間序列預測問題套用在機器學習中，監督式學習模型構建在一個輸入(X)即對應一個輸出(Y)配對之訓練，訓練資料通常需由人工建立，並對資料進行預測。因此在建構 LSTM 模型之前需先將資料進行預處理，會將序列轉換為監督式學習問題，透過正規化方式(Normalization)將資料特徵進行縮放，始資料介於[0,1]之間，主要目的是防止某一個特徵值過大影響結果。

由於長短期記憶類神經網路具有學習能力高、能夠訓練誤差，以及將錯誤儲存並

且能夠反覆思考的功能，能夠處理非線性函數的問題。由於本研究所蒐集到資料有不規則之變動，並且是以過去時間的資料分析各成分之變化形態，再據以預測未來之變化，符合時間序列之特性，因此數據分為訓練集與測試集兩類，以建立需求預測模型。

LSTM 的基本架構圖 3.4 所示， X 為縮放信息、 $+$ 為添加訊息、 σ 為 Sigmoid 層、 h_{t-1} 為上一個 LSTM 單元輸出、 C_{t-1} 為上一個 LSTM 單元記憶、 X_t 為輸入、 C_t 為最新的記憶以及 h_t 為輸出，本研究將 LSTM 預測模型調整分為以下三個步驟：

- Step1：遺忘門(忘記不必要的訊息)

每個輸入都有自身的權重 LSTM 單元，Sigmoid 激活函數介於 0 至 1 之間做非線性轉換，其運算後之資訊會再由儲存記憶單元內的資訊(C_{t-1})進行點積運算，控制著儲存記憶單元內上一期的資訊傳遞，會取得當期的輸入(X_t)及前一期時間點的輸出(h_{t-1})兩個向量，並決定從過去輸出中刪除哪些部分。例如當輸入是「他有一位女性朋友叫靜香」時，「大雄」的姓名可以被遺忘，因為主題已經變成了靜香。這個門被稱之為遺忘門(f_t)，門輸出為 $(f_t) * (C_{t-1})$ 。

- Step2：決定並儲存記憶單元新輸入之 X_t 訊息

在 Sigmoid 層決定需更新或忽略那些新訊息，並且可以根據 tanh 層輸入新的數值建立所有可能值的向量之輸入已更新這個新的記憶單元添加到舊記憶(C_{t-1})中，得出 C_t 。在上述例子當中，對於新的輸入，他有一位女性朋友叫靜香，靜香的性別已被更新；當輸入訊息為「靜香在新北市一家著名的科大擔任教授，最近他們參加同學會上碰到面」，此時像「著名」、「同學會」這樣的詞可以被忽略，而「教授」、「科大」與「新北市」這樣的詞將被更新。

- Step3:決定輸出內容

記憶單元透過 tanh 生成所有可能的值，矩陣相乘以 Sigmoid 門之輸出，更新後的值將會透過雙曲正切函數進行非線性轉換後與 O_t 進行點積運算，代表著 O_t 控制計算後的值 h_t 是否能夠順利輸出。在例子，若要預測空白的單詞，模型知道在記憶中的「老師」相關名詞，它可以很容易回答「教書」，並沒有直接給模型這個答案，而是讓它長期學習中得出結果。

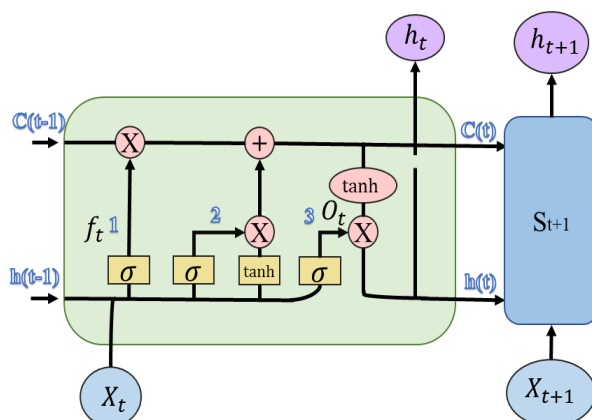


圖 3 LSTM 架構圖

(一) 梯度下降法

在類神經網路中，最佳化理論透過梯度下降法變形而成，較常見的梯度下降法有隨機梯度下降法(Stochastic gradient descent, SGD)、批量梯度下降法(Batch Gradient Descent, BGD)、(Adaptive Gradient, AdaGrad)、(Root Mean Square Prop, RMSProp)、自適應時刻估計法(Adaptive Moment Estimation, Adam)等，最一開始為 BGD，每次訓練皆使用所有的訓練樣本進行訓練。因此當訓練樣本變多時需要使用全部訓練樣本來更新模型參數，易造成訓練時間過長，而 SGD 即是用來解決 BGD 之弊端，對「權重與 bias 的組合」進行最佳化，其沿著目標函數參數梯度的反方向將各個參數的梯度值來最小化損失函數。

AdaGrad 在 2011 年由 Duchi John、Elad Hazan、Yoram Singer(2011)，提出可以運用不同參數調整適合學習率之梯度下降法，此方法稱之為「學習率衰減(Learning rate decay)」遇到稀疏參數時會大幅更新、當參數較為頻繁則僅有小幅度更新，因此 AdaGrad 適用於處理稀疏數據，但當迭代次數越大，梯度平方和會越大，使得學習率越小，易影響訓練提前結束。

$$X_i^{(t+1)} = X_i^{(t)} - \frac{\gamma}{\sqrt{G_{t,ii}} + \epsilon} g_{t,i}$$

(3.10)

ϵ 為平滑項避免分母為 0 的問題， γ 為學習率、 g_t 為參數梯度平方和，其缺點是在訓練中後段時，可能會因分母累積越來越大導致梯度趨近於 0，若有設定 early stop 會使得訓練提前結束。

RMSProp 與 AdaGrad 一樣自適應調整學習率，但 AdaGrad 的分母是從第一次梯度到第 t 次梯度之和易導致學習率下降過快的問題，而 RMSProp 的分母則是取平均數。

$$X_i^{(t+1)} = X_i^{(t)} - \frac{\gamma}{\sqrt{E[gi^2]_t} + \epsilon} g_{t,i}$$

(3.11)

$E[\cdot]$ 在統計上取期望值，因此取 g_t^2 之期望值、 g_t 為參數梯度平方和。

Adam 在 2014 由 Diederik P.Kingma 與 Jimmy Lei Ba 提出，計算參數更新方向前會考慮前一次參數更新之方向，RMSProp 則在學習率上會依據梯度大小對學習率進行加強或是衰退， m_t 和 v_t 分別對梯度的指數移動平均值(Exponential Moving Averages)和平均梯度(Square Gradient)，各自做偏差修正。Adam 的優點在於所需內存少、適合用於大數據、能對模型參數進行較佳效能、高計算效率以及僅需調整少量參數(Kingma & Ba, 2014)。

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

(二) 超參數

依據類神經網路模型學習方法，調整模型之參數，模型建立後需由學習率(Learn Rate)、丟棄率(Dropout)以及激活函數(Activation Function)等調校模型，其目的主要是為了讓模型辨識能力可以達到較佳的效果。

1. 學習率(Learning Rate)

學習率在梯度下降法中代表著依據不同學習問題而改變，達到最佳學習效果(Ruder, 2017)，當學習率設定值越大時，會使得誤差值無法下降，或是收斂會落在較大誤差範圍內；反之，則可能會讓誤差下降的速度變得緩慢，使模型訓練時間變長，會依據模型使用不同之演算法，學習率的大小會直接影響到優化之效率。

2. 丟棄率(Dropout)

類神經網路中包含非線性之隱藏層，當隱藏層結構越趨近於複雜龐大時，越能夠時找出資料間複雜之關係，但若在訓練資料有限之情況下，部分訓練出來之結果可能無法代表測試資料間之實際情況，進而出現模型過度適配(overfitting)之現象，為了解決此現象，一般會採用設定一個比率，使類神經的節點會依照設定比率將部分之輸出為零，可以明顯減少特徵檢測器(隱藏節點)間之相互作用，增強模型泛化性，並不會太依賴某些局部特徵，以避免過擬合情況發生，如圖4所示(Srivastava, Hinton et al., 2014)。

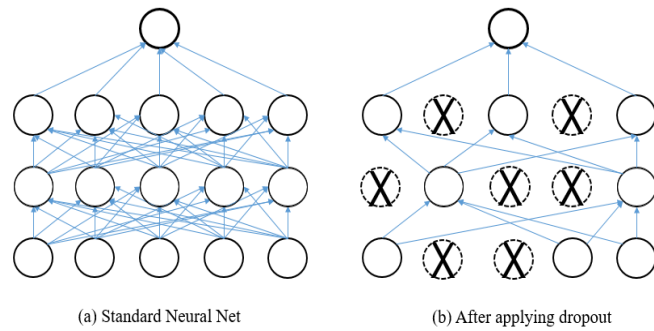


圖 4 使用 Dropout 神經網路模型

若以節點輸入矩陣為 x ，節點間權重為 w ，bias 為 b ，激活函數為 σ ，輸出矩陣為 y ，沒有設定資料丟棄率時，輸出如下所示。

$$y = \sigma(wx + b)$$

(3.14)

一般在複雜模型中設定資料丟棄率會達到較佳效果，其中 r 為資料丟棄向量、 p 為機率值、白努力(Bernoulli)函數會產生一個以 0 至 1 為值之向量，依據輸入之機率 p 值來決定 0 在向量中出現之機率(Srivastava et al., 2014)。

$$r = \text{Bernoulli}(p)$$

$$y = f(w(rx) + b)$$

(3.16)

(三) 激活函數

激活函數又稱為啟動函數或轉換函數，主要為一個可微分的非線性的函數，神經網路結構中最重要的是透過激活函數，來控制神經元的輸出幅度(Karlik & Olgac, 2011)。在類神經網路使用激活函數，主要是利用非線性方程式，解決非線性問題，因為無論是隱藏層或輸出層，每一層都是上一層輸出的線性函數，而激活函數等於是給神經元導入了非線性函數，使得類神經網路可以任意逼近任何非線性函數。

激活函數使用的選擇需要根據模型的需求才能組合出較好的效果，通常是在類神經網路節點處理完權重及bias 後才使用，如前式 (3.14) 所示常見的激活函數分別為 Sigmoid、Tanh 以及 ReLU 等三種。

1. Sigmoid

函數也稱邏輯函數，其圖形像個 S 形，它會把輸出值限定在 0 與 1 之間，使得數據在傳遞過程中不容易發散，再大的值都會被限縮在 0 與 1 之間 (Karlik & Olgac,

2011)。依圖形顯示，輸出值在接近 1 或者是 0 時，其梯度也趨近於 0，所以在進行倒傳遞時會有梯度消失的問題，除此之外，Sigmoid 的輸出值不是以 0 為均值，這對梯度會造成影響，若輸入值均為正數或負數時，權重值會往正方向或負方向更新，進而造成收斂速度緩慢。

2. Tanh (Hyperbolic Tangent)

Tanh 函數又稱為雙曲正切函數，圖形類似於 Sigmoid 函數，但是效果會較 Sigmoid 好一些，取的範圍值是-1 至 1 之間。依圖形顯示 Tanh 函數的輸出以 0 為均值，所以收斂效果會較 Sigmoid 函數好一些，但是在輸出值接近 1 和-1 的地方還是會有梯度消失的問題。

3. ReLU (Rectified Linear Unit)

ReLU 函數若輸入的值為正，則輸出該值大小，若輸的值為負，輸出則為 0，因此 ReLU 函數僅需判斷輸入值是否大於 0，不需要經過複雜的指數運算，且在輸入值大於 0 的區域上，不會因梯度擬合而出現梯度消失的問題，另外在模型的收斂速度上也較 Sigmoid 和 Tanh 函數快許多 (Krizhevsky, Sutskever, & Hinton, 2012)。

三、預測評估指標

常見的預測準確度檢驗方法有 MAE (Mean Absolute Error)、MAPE (Mean Absolute Percent Error)、RMSE (Root-Mean-Square Error) 三種，上列幾總計算結果愈小者表示預測誤差愈小，預測能力愈佳。本研究採用 RMSE 作為判斷預測能力指標。

(一) Mean Absolute Error (MAE)

為衡量每筆資料預測值與實際值之誤差，MAE 法將每筆資料誤差取絕對值後加總，再計算平均誤差，公式如下：

$$MSE = \frac{1}{k} \sum_{t=1}^k (y_t - \bar{f}_t)^2$$

平均絕對誤差能夠獲得一個評價值，但由於不知道這個價值代表模型擬合優劣，因此需要透過對比才能達到評價指標的效果。

(二) Mean Absolute Percent Error (MAPE)

MAPE 係以每筆資料之相對預測誤差衡量，可避免 MAD 法及 MSE 法中，因資料本身數值較大而導致計算結果過大的缺點，公式如下：

$$MAPE = \frac{1}{k} \sum_{t=1}^k \frac{|y_t - \bar{f}_t|}{y_t} \times 100\%$$

表 5 MAPE 預測百分比對照表

MAPE(%)	說明
<10	高準確的預測
10~20	優良預測
20~50	合理預測
>50	不準確預測

(三) 均方根誤差RMSE(Root-Mean-Square Error)

均方根誤差亦稱標準誤差，它是觀測值與實際值偏差的平方與觀測次數比值的平方根。均方根誤差是用來衡量觀測值與實際值之間的偏差。標準誤差對一組測量中的特大或特小誤差反映非常敏感。所以，標準誤差能夠很好地反映出測量的精密度。可用標準誤差作為評定這一測量過程精度的標準，公式如下：

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{k} \sum_{t=1}^k |y_t - \bar{f}_t|^2}$$

(3.19)

其中 y_t 為實際值， \bar{f}_t 為預測值，k 為樣本數，預測模型能力指標為 RMSE 樣本規模越小，均方根誤差越可靠。

肆、模型建立與評估

本研究蒐集 2019 年 1 月 1 日至 2020 年 3 月 31 日新北市新莊空氣監測站 PM_{2.5} 濃度值 410 筆資料，利用滾動式預測未來數值，並以樣本內資料進行時間序列建模，樣本外資料進行預測，建立較佳預測。

一、建立 ARIMA 模型

首先，執行 Python 將原始資料及取對數資料進行 Augmented Dicky Fuller Test，取對數的原因主要是希望序列數值不會因趨勢及季節性所影響，結果如表 6。

表 6 Augmented Dicky Fuller Test 結果

檢定量	ADF		KPSS	
d=0	T-Statistic=-4.107524		T-Statistic=0.305061	
	p-value=0.000941		p-value=0.1	
	1% level	-3.479007	1% level	0.739
	5% level	-2.882878	2.5% level	0.574
	10% level	-2.578149	5% level	0.463
			10% level	0.347

由表 6 可見，序列沒有經過差分(d=0)，無論在 ADF 檢定統計量其檢定結果(T-Statistic) 小於顯著水準(α=0.01,0.05,0.1)，以及 KPSS 檢定統計量故拒絕(T-Statistic) 小於顯著水準

($\alpha=0.01,0.25,0.05,0.1$)，故無法拒絕單根存在的虛無假設，表示該序列資料為不存在單根之定態序列。

在決定差分次數後可以利用自我相關途中之落後 p 階之自我相關函數 (Autocorrelation function ,ACF)與偏自相關函數(Partial autocorrelation function ,PACF)是最直接判斷ARMA中的落後期數，針對模型組合進行樣本模組內評估，選擇ARMA(2,2)作為參數。

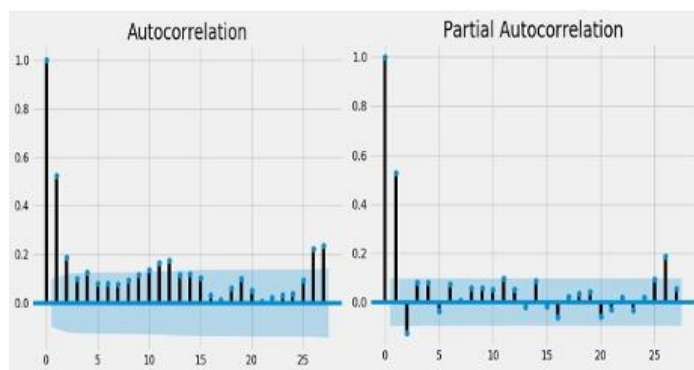


圖 5 ACF 和 PACF 結果

選擇最佳參數，並分析特徵對於序列的影響，當 $P>|z|$ 低於或接近 0.05，表示模型中所保留所有權重為合理，如圖 6 所示。

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.4697	0.044	10.733	0.000	0.384	0.556
ma.L1	-0.9702	0.018	-54.837	0.000	-1.005	-0.936
ar.S.L12	0.0102	0.005	2.042	0.041	0.000	0.020
ma.S.L12	-0.9003	0.033	-27.146	0.000	-0.965	-0.835
sigma2	43.7870	2.311	18.947	0.000	39.258	48.316

圖 6 PM_{2.5} 特徵權重

進行模型診斷，主要是確保模型的殘差為不相關並且平均分布為零如圖 7 示，診斷殘差正常如下，①左上角圖片為標準化殘差序列②右上圖主要是紅色 KED 線與 N(0,1) 為常態分配，平均值 0，標準偏差為 1，表示殘差為正常分布③左下角為 qq 圖表示，殘差(藍點)有序分布遵循 N(0,1)的標準分布線性趨勢，如上述所言為殘差正常分布④右下角為自相關圖表示殘差與本身滯後具有低相關。

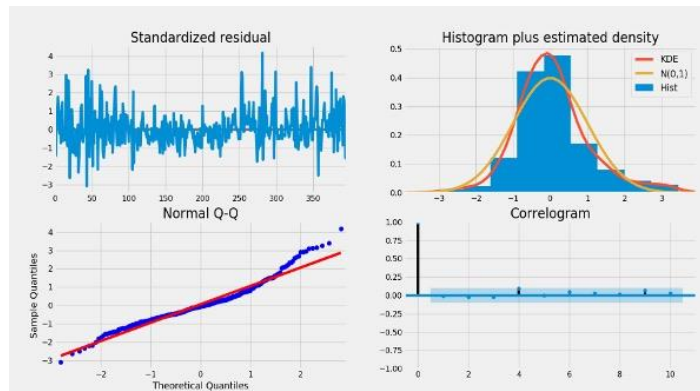


圖 7 PM_{2.5} 殘差圖

由下圖 8 示 ARMA(2,2)之趨勢圖預測結果，易由實際值較高時預測偏低的情形，RMSE 為 7.667。

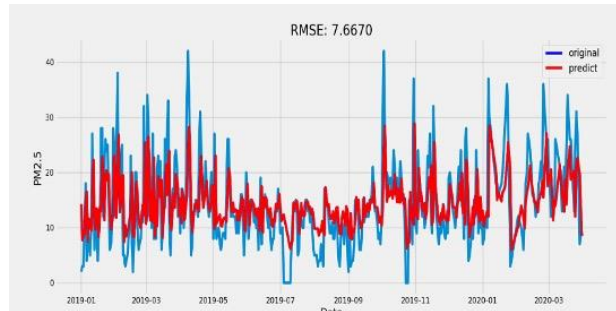


圖 8 ARIMA PM_{2.5} 實際值與預測值趨勢圖

由下圖 9 所示 LSTM 趨勢圖預測結果，預測準確性 RMSE 為 6.25 比 ARIMA 來得精準。

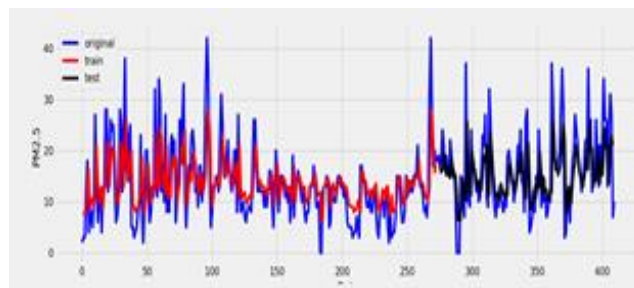


圖 9 LSTM PM_{2.5} 實際值與預測值趨勢圖

伍、結論與建議

近年來空氣品質嚴重惡化一直是社會關心的重要議題，準確的空氣品質預測是民眾從事戶外活動的參考，也可以提供政府管制污染源的決策指標，因此對於空氣污染的預測有許多的理論與方法。本研究以新北市環保署空氣品質監測站新莊站之 PM2.5 濃度進行測試，建立單變量的時間序列模型 ARIMA 及類神經模型 LSTM，經測試 LSTM 模型預測結果確實較優於 ARIMA 模型。預測準確率不理想，可能有以下幾個原因：

(1) 建立多變量模型

加入溫度、雨量、風向及風速等氣象條件，及蒐集鄰近測站的空間關係資料建立的預測模型。

(2) 更多預測模型比較

本研究以統計時間序列與類神經網路方法論建構 ARIMA 與 LSTM 兩種預測模型，並依預測結果進行比較，而現今人工智慧預測方法工具眾多，後續研究者可以嘗試更多不同預測模型的建立，如基因演算法、改良後的遞歸神經網路模型(GRU)...等，藉由更多預測模型的比較，挑選出最佳的預測模型來建構產品需求預測模式。

參考文獻

- 全國資料庫(2020, 4月20日)。空氣品質標準。取自網址: <https://law.moj.gov.tw/>
- 行政院環境保護署(2019, 3月25)。空氣品質標準。取自網址: <https://www.epa.gov.tw/>
- 高雄市空氣污染防制(2020, 4月12日)。空氣污染防制。取自網址: <https://air.ksepb.gov.tw/News/List/1>
- 陳維新、江金龍(2009)。空氣污染與控制。臺北市:高立圖書。
- 新北市政府民政局(2020, 4月18日)。環境資源新北市新莊區 PM2.5 濃度值資料庫。取自網址: <https://www.ca.ntpc.gov.tw/>
- 楊宏隆(1998)。大氣懸浮微粒 PM2.5 及 PM10 之特性及來源分析，國立中興大學環境工程學研究所，碩士論文。
- 楊浩彥、郭迺峰、林政勳(2013)。實用財經計量方法 EViews 之應用。臺北市:雙葉書廊。
- 蔡卓翰(2010)。澎湖國家風景區遊客量預測模式之研究，東海大學工業工程與經營資訊學系，碩士論文。
- 衛生福利部國民健康署(2022年, 4月12)。空氣品質監測系統。取自網址: <https://www.hpa.gov.tw/Home/Index.aspx>
- Box, G. E. P and Jenkins, G.M., (1976). "Time series analysis: „Forecasting and control," Holden-Day, San Francisco. *Journal of Applied Mathematics and Statistics*. 2014, Vol. 2 No.

1, 16-28

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.

Donald S. Remer and Craig Jorgens. "Ethylene economics and production forecasting in a changing environment". 1978, vol. 3, issue 4, 267-278.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).

Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

IARC (2013). "Outdoor Air Pollution a Leading Environmental Cause of Cancer Deaths". Retrieved from: <https://monographs.iarc.fr/agents-classified-by-the-iarc/>

Karlik, B., & Olgac, A. V. (2011). Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. *International Journal of Artificial Intelligence and Expert Systems* 1(4), 111-122.

Kingma, D. P. (2014). and Jimmy Lei Ba, "Adam: A method for stochastic optimization,". *arXiv preprint arXiv,1412.6980*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*.1097-1105.

Kwiatkowski,Phillips,Schmidt,&Shin(1992). *Journal of Econometrics*, 1992, vol. 54, issue 1-3, 159-178.

Lewis, C.D. (1982) *International and Business Forecasting Methods*. Butterworths, London.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv,1706.05098*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014).

Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.

WHO(2014). " 7 million premature deaths annually linked to air pollution. " Retrieved from:<http://www.who.int/mediacentre/news/releases/>.