國立臺灣師範大學理學院

資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

College of Science

National Taiwan Normal University

Master's Thesis

使用排隊理論分析室外空氣品質與人口流動對於 COVID-19 的爆發風險

Predicting Risk of COVID-19 Outbreak by using Outdoor Air Pollution Indicators and Population Flow with Queueing Theory

蔣宜芳

Yi-Fang Chiang

指導教授： 賀耀華 博士

Advisor: Yao-Hua Ho, Ph,D

中華民國 111 年 6 月

June 2022

摘要

自2019年12月首次被發現以來 COVID-19 已經席捲各個國家，因其高度傳染性使得各國難以預防，當相關機構要執行防疫措施時，其執行的時間點需要更多資訊進行評估。COVID-19 主要透過飛沫傳染，因此本研究將人口流動視為其中的關鍵影響，但是增加傳染的風險並不局限於人口流動一種。空氣汙染物也被認為與呼吸道相關疾病有著高度的相關性，先前研究也使用空氣品質以預測流感病毒爆發時間，此外研究也觀察到人口流動與空氣汙染物間存在一定的相關性，像是因為開車而造成 NO_2 以及 $PM_{2.5}$ 的增加等。

本研究所選擇的影響因素可以觀察到相關機構的防疫措施對於這些因素的影響性，例如是否強制規定戴口罩、居家令的執行以及重新開放等相關防疫措施。戴口罩可以減少飛沫傳染的風險也較能減少空氣汙染物的影響，而居家令則是會減少人口的流動也減少人與外面空氣的接觸，重新開放則是使得人口流動恢復也增加空氣污染物。因此本研究希望能透過所提出的影響因素來協助評估目前疫情爆發的風險並做為防疫措施的一參考數值。

本研究使用美國加利福尼亞州、佛羅里達州、紐澤西州以及紐約州的人口流動、室外空氣汙染物及疫苗接種預測COVID-19 確診人數及活動病例，並以排隊理論預測其平均所需康復時間，將預測結果帶入的風險分析。最終以斜率來進行風險分析，獲取爆發閥值，預測可能存在的高風險時段，將這些時段與防疫措施結合，提供相關機構一個評估防疫措施的參考數據，期望達到對疫情有更好的控制。

Abstract

COVID-19 has swept all countries since it was first discovered in December, 2019. The high infectivity of COVID-19 makes it more difficult to prevent it, and air quality has been considered to be highly correlated with respiratory diseases. Previous study shown that air quality has been used for influenza viruses outbreak time prediction, but air is not the only one that increases the risk of infection. Population movement is also one of the key influences. We have observed that population movement is also correlated with air pollutions. In addition, the study also observed a certain correlation between population movement and air pollutants, such as the increase in NO_2 and $PM_{2.5}$ caused by driving.

In addition, the government's epidemic prevention strategies are also one of the factors that affect the risk of infection, such as whether it is mandatory to wear masks or household orders, etc. Wearing masks can reduce the risk of droplet infection and the impact of air pollution, while household orders can reduce the mobility of the population and the contact between people and the outside air.

In this study, the number of COVID-19 cases and active cases will be predicted according to population flow, outdoor air pollutions and vaccination in California, Florida, New Jersey and New York, and the average recovery time will be predicted by queuing theory, and the predicted results will be brought into risk analysis. Finally, the slope is used to analyze the risk, predict the possible high-risk time periods, and combine these time periods with the epidemic prevention measures to provide a reference data for relevant institutions to evaluate the epidemic prevention measures, hoping to achieve better control of the epidemic situation.

目錄

第一章 緒論 1
第二章 文獻探討 3
第一節 人口流動與 COVID-19 的相關性 3
第二節 空氣汙染物與 COVID-19 的相關性 4
第三節 人口流動與空氣品質的相關性 6
第四節 呼吸道疾病預測 7
2.4.1 機器學習及人工神經網路演算法 8
2.4.2 實際研究 12
第五節 風險評估 13
2.5.1 排隊理論 15
第三章 研究方法 17
第一節 資料來源、相關性分析及前處理 18
3.1.1 資料來源 18
3.1.2 資料前處理 19
3.1.3 各州資料分析 20
3.1.4 資料相關性分析 24
3.1.5 滯後影響 28
3.1.6 資料合併及切割 30
第二節 資料預測 32
3.2.1 機器學習模型 33
3.2.2 LSTM 模型 34
3.2.3 模型評估 35
第三節 排隊理論及風險分析 36
3.3.1 排隊理論 36
3.3.2 風險分析 38
第四章 實驗結果分析 42
第一節 案例探討 42
4.1.1 加利福尼亞州 45
第五章 結論與未來展望 54

第六章 參考文獻 56

附圖目錄

圖 1. LSTM 示意圖 11
圖 2. 排隊理論示意圖 15
圖 3. 研究架構圖 17
圖 4. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州2019 年 NO_2 盒狀圖 21
圖 5. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州2019 年 SO_2 盒狀圖 21
圖 6. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州 2019 年 Ozone 盒狀圖 21
圖 7. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州 2019 年 $PM_{2.5}$ 盒狀圖 21
圖 8. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州 2019 年 CO 盒狀圖 22
圖 9. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 SO_2 盒狀圖 23
圖 10. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 NO_2 盒狀圖 23
圖 11. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 $PM_{2.5}$ 盒狀圖 23
圖 12. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 Ozone 盒狀圖 23
圖 13. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 CO 盒狀圖 24
圖 14. Pearson Correlation 結果圖 27
圖 15. 加利福尼亞州不同滯後天數相關係數圖 29
圖 16. 佛羅里達州不同滯後天數相關係數圖 29
圖 17. 紐澤西州不同滯後天數相關係數圖 29
圖 18. 紐約州不同滯後天數相關係數圖 29
圖 19. 回歸預測流程圖 32
圖 20. 機器學習模型示意圖 33
圖 21. LSTM 模型圖 34
圖 22. COVID-19 狀態轉換圖 36
圖 23. 排隊理論模型示意圖 37
圖 24. 風險分析流程圖 38
圖 25. 英國2020/01/31~ 2021/11/30 每日確診人數圖 38
圖 26. 英國2020/01/31~ 2021/11/30 每日確診人數斜率圖 39
圖 27. 英國2020/01/31~ 2021/11/30 確診人數斜率持續增長天數圖 40
圖 28. 英國2020/01/31~ 2021/11/30 持續增長天數斜率圖 40
圖 29. 英國確診人數風險分析結果圖 41

圖 30 PR 預測每日確診人數分析結果圖 42
圖 31 PR 預測每日活動病例分析結果圖 42
圖 32 SVR預測每日確診人數分析結果圖 43
圖 33 SVR 預測每日活動病例分析結果圖 43
圖 34. 加利福尼亞州防疫措施時間線圖 45
圖 35. 加利福尼亞州 2020/ 03 PR 風險分析圖 46
圖 36.加利福尼亞州 2020/ 03 SVR 風險分析圖 46
圖 37. 加利福尼亞州 2020/ 11 PR 風險分析圖 47
圖 38. 加利福尼亞州 2020/ 11 SVR 風險分析圖 48
圖 39. 加利福尼亞州 2020/ 05 PR 風險分析圖 48
圖 40. 加利福尼亞州 2020/ 05 SVR 風險分析圖 49
圖 41.加利福尼亞州 2021/ 01 PR 風險分析圖 50
圖 42.加利福尼亞州 2021/ 01 SVR 風險分析圖 50
圖 43.加利福尼亞州 2021/ 06 PR 風險分析圖 51
圖 44. 加利福尼亞州 2021/ 06 SVR 風險分析圖 52
圖 45. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州的 PR 風險分析圖 52

表目錄

表 1.州人口及土地面積表 20
表 2. 不同時期的起迄時間表 25
表 3. 根據不同時期，空氣汙染物、人口流動及疫苗施打率與確診人數相關性分析表 26
表 4. 模型評估結果表 36

第一章 緒論

自2019年12月起首次在武漢被發現以來嚴重特殊傳染性肺炎(Coronavirus disease 2019 , COVID-19)已經席捲各個國家， COVID-19的高度傳播性及其較高的死亡率讓各國政府不得不重視它所帶來的問題，COVID-19 的這些特性使它更加難以去預測及預防，各個國家在疫情爆發初期經歷了一段艱難的時期後了解如何去應對疫情爆發，通過限制人口流動、減少人與人的接觸、戴口罩等等防疫措施去控制疫情，但我們接下來所要面對的是何時可以解除這些防疫措施，已經有許多國家因為過早的放寬相關防疫措施而導致疫情再次爆發，進而造成更嚴重的情形發生。而如何知曉目前疫情爆發的風險，以及有哪些因素影響疫情爆發，因此這些資訊對於目前政府因應相關防疫措施，變得極為重要。

據 WHO 所公布的研究報告中指出[1]， COVID-19 是藉由人與人之間的飛沫進行傳染。但因為飛沫顆粒大小不一，較大直徑的飛沫會快速降落至地面或物體表面，而小直徑的飛沫則會在空氣中懸浮著，時長可從數分鐘至數小時。而這些飛沫會經由感染者的行為，如呼吸、說話或咳嗽等等行為而釋放出來，導致周遭人群有機會透過直接或間接途徑又或是直接吸入而遭到感染。所以當人群與感染者足跡有重疊時會大大地增加感染風險，因此人口流動可以視為影響感染風險的一大因素。

但是增加傳染的風險並不只有人口流動一種，近幾年受空氣污染的影響，也讓我們對於空氣品質與疾病之間的相關有更進一步的了解。已有需多研究證實空氣汙染會增加呼吸道疾病的風險，像是懸浮微粒(PM)、氣態污染物（臭氧、二氧化氮和二氧化硫）對於哮喘有較明顯影響[2]，而長期暴露空氣污染物較高的環境會影響兒童呼吸系統發育[3]。由此可見空氣品質及汙染物對於疾病存在一定的影響因素。對於呼吸道疾病而言，汙染物中的懸浮微粒 (PM) 已被證實容易引發呼吸道

相關疾病[4]。因為其粒徑較小可經由呼吸進入人體，進而穿透肺泡在人體內循環。除了容易引發呼吸道疾病以外，細懸浮微粒也可能攜帶依附在顆粒上的汙染物一同進入體內。此細懸浮微粒的特性也同樣可以讓 COVID-19 的病毒依附在顆粒上面進入體內而導致感染[5]。而目前多數國家中空氣污染較嚴重的地區，同樣也會有較嚴重的影響[6]。因此空氣品質及空氣汙染物同樣可視為 COVID-19 傳染的途徑。

而上述的人口流動與空氣品質除了可單獨視為 COVID-19 的傳染途徑外，經研究證實人口流動與空氣品質之間有一定的相關性，例如因為開車等交通行為而造成二氧化氮 (NO_2) 以及細懸浮微粒 ($PM_{2.5}$) 的增加 [7]。而感染者的交通行為會使得疫情更進一步的傳播，病毒會隨著這些感染者的途徑進行擴散。以最先感染的武漢市為例，有研究發現其他城市的確診者增加，與武漢市的人口流入有著密切的相關[8]。

這兩者之間的相關更進一步表示人口流動及空氣汙染物對於 COVID-19 有著明確的影響。此外在 2020 年 12月時針對 COVID-19 的首支疫苗研發完成並開始施打，因為疫苗的出現有效降低感染後的重症率，進而減少了死亡風險，同時也因為重症的減少，大多確診者轉變成無症狀患者，確診數也跟著下降，所以我們也將疫苗施打比例是為影響疫情的重要因素之一。

本研究將透過上述所提到的影響疫情的三個因素，人口流動、空氣品質及疫苗施打比例來進行未來確診人數及活動病例預測。本研究將針對預測出的人數進行排隊理論分析，獲取平均康復時間，進而評估未來的疫情高風險時段，提供相關機構對於防疫措施的參考數據，以達到對疫情更好的控制。

第二章 文獻探討

第一節 人口流動與 COVID-19的相關性

鑒於 COVID-19 是透過飛沫傳染，因此在其傳染途徑中，人與人的之間的足跡重疊扮演著重要的角色。現今的防疫措施大多也是限制人們的行動，減少重疊避免互相感染。但在疫情初期還不太了解 COVID-19 極強的傳播性，可以觀察到確診案例與最開始爆發疫情地區的人口流動有著高度的正相關[8]。2019年末做為首例 COVID-19 出現地的中國就有研究人員利用 COVID-19 病例的時間和位置信息及武漢市的人口流動進行研究。中國一些病例數較少的城市，當其病例數突然急增，其原因來自於武漢市的人口外流，因此確定來自武漢市人口流動與鄰近地區確診案例有高度的正相關0。

而同樣位於中國的研究將研究時間延長至政府進行防疫措施，防疫措施將所有可以進行人口流動的交通方式都進行暫停，進而發現限制人口流動對於 COVID-19 而言是有效的防疫措施[9]。而之後許多國家在疫情爆發後也開始限制市民的活動量來達到疫情控制，但限制的幅度可能不盡相同，時長也漸漸拉長。以瑞典為研究背景的研究就指出，雖然瑞典採取較溫和的限制措施，並沒有強制禁止人口流動，但依舊是有降低確診率的效果[10]。以及在英國背景下的研究，以英國六個常見的移動空間根據時間變化進行分析，發現零售超市及交通設施的人口流動減少幅度最大。但公園的使用起初是有減少但也漸漸回復到限制之前的流動量，以及其他空間在後期隨時間推移流動量也有所增加[11]。同樣的情形在南韓的研究中也有提到，南韓 2020 年 1 月至 3 月的全國的日均交通量與疫情尚未爆發的 2019 年相同月份進行分析，雖然 2020 年 1 月至 3 月與 2019 年同期相比有明顯下降，但將 2020 年 1 月至 3 月進行比較，發現交通量正在慢慢回升，研究指出這種情形可能導致疫情傳播風險的增加[12]。

從上面的研究可以發現，在進行限制市民活動量的措施之後，市民可能會因為長期的行動受限制導致在限制解除後有報復性出遊的行為。而在另一個南韓研究中證實了這項可能，研究中有提到在第一波疫情結束後，市民因為不再被限制而擁入其他地區造成第二波的疫情[13]。可以看出人口流動對於 COVID-19 的影響性是一直存在，並不只停留於首次爆發，而是持續性的影響，但也並非所有區域都是相同的。在美國的研究中還有提到城市與郊區對於限制人口流動的反應可能不太一樣，城市可以較明顯的看出限制效果，但因為郊區本身人口流動就沒有城市頻繁，所以限制前後的差別並不顯著。因此較難看出郊區人口流動與新增確診之間的相關性[14]。這點同樣可以在南韓研究中驗證，雖然交通量確實與確診數有著正相關，但並不是所有區域都一樣[13]。根據上述所提到研究我們可以發現人口流動確實與 COVID-19 有著一定的相關性，但對於人口流動本就比較稀少的地區而言，兩者之間的相關性就稍嫌不足，需要有其他特徵來補足其相關性。

第二節 空氣汙染物與 COVID-19的相關性

空氣汙染對於呼吸道疾病的影響已經是經過多年的驗證，不論是對於兒童、成年人[15]，又或是對於老年人[16]，對於短期暴露在空氣污染下而造成的急性影響[16]，以及長期暴露的慢性影響[15]，都可以證實空氣汙染與呼吸道疾病的相關性極高。而 COVID-19 也不例外，已經有研究指出長期暴露在空氣污染下會增加 COVID-19 的感染和 COVID-19 疾病嚴重程度的風險[17]。其中特指汙染物細懸浮微粒，在相關研究結果指出細懸浮微粒可能增加肺部纖維化的嚴重程度[18]，而短期暴露在空氣污染的影響則是與病毒在人群中的傳播有相關性[8]，除了會引起肺部炎症以外，細懸浮微粒可能會協助病毒傳播更遠，這個觀點在意大利環境醫學學會 (SIMA) 提出的研究獲得證實，同樣的[19]研究認為懸浮微粒是除了飛沫之外促成 COVID-19 傳播的重要載體，而懸浮微粒能使傳播距離更遠，空氣污染不只本身對於呼

吸系統就有危害性，更可以增加病毒的傳播性。

義大利研究學者認為其中二氧化氮 (NO_2) 以及細懸浮微粒 ($PM_{2.5}$) 是促進的 COVID-19 傳播的原因，研究中對義大利全國各地二氧化氮 (NO_2) 以及細懸浮微粒 ($PM_{2.5}$) 與 COVID-19 發病率進行線性回歸分析，並將老年人口及人口密度作為潛在因素。研究最後證實二氧化氮 (NO_2) 以及細懸浮微粒 ($PM_{2.5}$) 濃度每增加 1 $\mu g/m^3$ 就會使 COVID-19 發病率增加1.24 和 1.56 $*10^4$人 [20]。同樣的情形在祕魯的研究也經過證實，祕魯的研究證實二氧化氮 (NO_2) 對於 COVID-19 感染的影響性 [21]。除了汙染物之外氣候因素也可能是影響 COVID-19 傳播的因素，學者針對紐約市的空氣和氣候數據包括溫度、濕度、風速、空氣品質及降雨量與 COVID-19 病例使用 Kendall 和 Spearman 進行相關性分析，結果顯示空氣品質與溫度對 COVID-19 有著顯著相關 [22]。這個發現與針對米蘭的空氣與氣候進行 Person 相關性分析的結果相同，研究發現對於米蘭而言 COVID-19 與相對溼度呈負相關。乾燥的空氣更適合病毒的傳播，並與氣溫呈正相關。此外也證實每日新增病例與空氣汙染中的懸浮微粒($PM_{2.5}$、PM_{10})之間的正相關性 [23]。從上述中可以證明細懸浮微粒與新增病例間的相關。

而對於死亡病例，二氧化氮 (NO_2)的影響則是更加明顯，研究指出因為暴露在二氧化氮 (NO_2) 環境下會導致肺部發炎，進而影響 COVID-19 後的致死率，研究中在選定的研究國家中，有78%的死亡案例發生在二氧化氮 (NO_2) 高濃度地區 [24]。另外研究學者對於法國三個城市的二氧化氮 (NO_2) 與 COVID-19 死亡病例使用Artificial Neural Networks (人工神經網路) 估算其閥值並使用Causal Direction from Dependency (D2C) algorithm 驗證其因果性，結果表明二氧化氮 (NO_2) 與 COVID-19 死亡病例存在因果關係並且不同城市有的不一樣的二氧化氮 (NO_2) 閥值 [25]。這個結果與研究背景為紐約的研究也同樣被證實，研究驗證其因果性，並且提出細懸浮微粒

($PM_{2.5}$) 對於死亡病例同樣存在著因果關係 [26]。此外有研究發現在防疫政策下可以更明顯的看出限制人口流動與空氣汙染的因果關係 [27]。

第三節 人口流動與空氣品質的相關性

對於空氣品質而言，存在許多影響因素。許多研究 [28][29]指出在人口較密集的城市地區，其空氣質量會隨著人口增加而下降。在德國研究 [28] 中使用當地的人口密度與空氣汙染進行分析，進而發現到較高的人口密度會惡化當地的空氣品質。在中國的研究 [29] 也得到相同的結論，研究結果顯示城市面積、城市人口和人口密度與細懸浮微粒 ($PM_{2.5}$) 呈正相關。但主要影響空氣汙染的因素不僅只有人口密度，例如在台灣主要的汙染源除了氣候帶來的境外汙染源之外，中南部的境內汙染源多為發電廠、重工業及焚化爐等等的廢氣排放，而人口較密集的北部則是來自都會區的汽機車廢氣排放。從上述可得知對於汽機車等交通方式的廢氣排放是造成空氣汙染物大增的重要因素。而多個研究 [30][31] 也已證實此觀點，在羅馬尼亞的研究 [30] 中指出交通壅塞與一般交通產生的汙染物與空氣品質之間存在著相關性。同樣在馬來西亞的研究 [31]也證實，人口流動所造成的交通現象會影響其空氣品質。從上述的結果可以看出人口流動對於空氣品質的影響性，同樣的結果在 COVID-19 期間也可以發現到，在孟加拉的研究 [32] 中，以 COVID-19 最嚴重的地區的空氣品質為資料進行分析，觀察限制人口流動前後的空氣品質變化。結果發現在限制人口流動後的空氣汙染物大幅減少，特別是交通量較大的城市，其空氣汙染物減少幅度最為顯著。而在巴西也做了相同的研究 [33]，結果與孟加拉類似與封鎖前相比大部分的汙染物都有下降的趨勢。由此可知人口流動與空氣汙染間有著顯著的相關，這能協助我們更進一步地去預測 COVID-19 。

第四節 呼吸道疾病預測

呼吸道疾病包含流行性感冒、哮喘、慢性阻塞性肺病、肺部纖維化、百日咳、肺結核以及肺癌等等。其中症狀因為感染的位置而有所不同，上呼吸道感染包括一般感冒，常見症狀為咳嗽、喉嚨癢、喉嚨痛、流鼻水，而下呼吸道感染包括肺炎，其常見症狀為反覆咳嗽、胸痛、咳出濃痰、喘、高燒。此外在 2003 年在台灣肆虐的嚴重急性呼吸道症候群(Severe Acute Respiratory Syndrome, SARS) 也是屬於呼吸道疾病的一種，而近兩年造成人心惶惶的疫情 COVID-19 ，其病原新型冠狀病毒 (SARS-CoV-2)，在受病毒感染後感染者會出現輕度至中度呼吸道症狀，症狀為上呼吸道感染症狀與一般感冒類似，所以通常無需特殊治療即可康復。但部分個案特別是老年人和患有心血管疾病、糖尿病、慢性呼吸道疾病或癌症等疾病的族群，可能出現嚴重的肺炎與呼吸衰竭等症狀。因此任何人都可能感染 COVID-19 並都會有病重或死亡[34][35]。

為了預防 COVID-19 大肆傳染，我們期望使用機器學習 (Machine Learning) 與人工神經網路 (Artificial Neural Network) 技術去對疾病進行預測。機器學習是從資料中自主分析進而獲得資料中存在的規律，並且使用這些規律對未知資料進行預測的演算法[36]，人工神經網路則是通過一個基於數學統計學類型的學習方法得以最佳化[37]。

機器學習演算法基本上分成四類，監督式學習 (Supervised Learning) 、半監督式學習 (Semi-supervised Learning) 、非監督式學習 (Unsupervised Learning)、強化式學習 (Reinforcement learning)。監督式學習是使用已知或是已標記完成的資料集進行學習，半監督式學習則是使用少部分資料進行標記來學習，無監督式學習利用未知資料集進行學習，而強化式學習則是剛開始不訓練，從錯誤的過程裡學習。對於使用機器學習來進行疾病預測，已經有許多研究可以證實其準確性，不論是用於預測心血管疾病 [38]

又或是呼吸道疾病中的慢性阻塞性肺病 [39]、肺結核相關治療 [40] 以及 COVID-19[41]。

但隨著資料的複雜度增加，機器學習已經無法滿足我們的需求，因此我們會使用人工神經網路來進行更複雜的學習。人工神經網路是由一個輸入層、一個或多個隱藏層，以及一個輸出層所組成，稱為一個節點。每一個節點或人工神經元會連接到另一個，而且具有相關加權和臨界值。如果任何個別節點的輸出高於指定臨界值，隨即啟動該節點，並將資料傳送至網路的下一層。 否則，不會將任何資料傳遞至網路的下一層。人工神經網路對於疾病的預測與機器學習相同也已經被多個研究證實其有效性，例如預測糖尿病[42]、心血管疾病 [43]、呼吸道疾病[44][45]以及流感[46]，還有近期的 COVID-19[47]。

2.4.1 機器學習及人工神經網路演算法

本段將介紹常用於預測疾病的機器學習演算法包括多元線性回歸 (Multiple Linear Regression, MLR)、多項式回歸 (Polynomial Regression, PR)、支持向量回歸 (Support Vector Regression, SVR) 以及人工神經網路演算法中最常被使用來預測疾病的長短期記憶 (Long Short-Term Memory, LSTM)。

- 多元線性迴歸 (Multiple Linear Regression, MLR)

  屬於線性迴歸的一種，線性迴歸是利用線性迴歸方程式的最小平方函數對自變數 (independent variable) 和應變數 (dependent variable) 之間關係進行建模的一種迴歸分析。多元線性迴歸自變數為多個，因此稱為多元線性迴歸，其公式與簡單迴歸類似但有更多項，公式為:

  假設資料集輸入 (x,z) 以及輸出 y, 資料長度為n, 資料集$T=\{(x_1, z_1, y_1), (x_2, z_2, y_2), \ldots (x_n, z_n, y_n)\}$，將資料集 T 帶入多元線性

回歸

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{1}$$

β_0為截距，β_1 、β_2為斜率可視為x_1、x_2變動一個單位導致y變動的量，ε為誤差。

線性回歸其計算量較小，所以訓練速度快，但也容易受離群值的影響而產生誤差。此外實際資料通常不會有如此明顯的線性相關，而非線性的關聯則會被線性回歸給忽略掉。

- 多項式回歸 (Polynomial Regression, PR)

多項式回歸同樣為回歸分析的一種，與線性回歸不同，其可以模擬非線性相關性。當自變數和應變數之間的關係為非線性時就可以使用，其會將關係建模為自變數的 n 次多項式。其公式為:

假設輸入為 x, 輸出 y

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \varepsilon \tag{2}$$

β_0為截距，β_1, $\beta_2, \cdots \beta_n$為斜率。

多項式回歸雖然可以模擬出資料間的非線性關係，但使用前需要先了解資料特性，因為需要選擇適合的指數進行擬合，但在選擇指數過程中容易會選擇太貼合訓練資料的指數而導致模型過擬合(overfitting)。

- 支持向量回歸 (Support Vector Regression, SVR)

支持向量回歸為支持向量機 (Support Vector Machine, SVM)的

延伸，SVM 適合用於分類。SVR 在高維或無限維空間中構造超平面或超平面集合，其將資料投射為空間中的一點，進而找到一個超平面。接著將新資料投射到同一空間，並基於它們的落點來對資料進行預測。其公式為:

假設資料包含輸入 x 以及輸出 y, 資料長度為n, 資料集T= $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$，其線性模型:

$$\hat{y}(x) = \langle w, x \rangle + b \tag{3}$$

$\hat{y}$ 為模型的估計輸出，w為權重向量，b為截距，但大多數據不為線性，因此有非線性的模型:

$$\hat{y}(x) = \langle w, \phi(x) \rangle + b \tag{4}$$

ϕ 為非線性核函數，模型需要選擇正確非線性映射$\phi(x)$並你核出最佳的w，其公式表現為$\mathrm{i} = \{1, 2, 3, \cdots, n\}$, ε 為最大誤差:

$$\begin{aligned} &minimize \quad \frac{1}{2}\|w\|^2 \\ &subject\ to\ y_i - \langle w, \phi(x_i) \rangle - b \leq \varepsilon \end{aligned} \tag{5}$$

SVR 與 SVM的區別在於， SVM是最近的樣本點與超平面的距離最大，SVR則是最遠的樣本點到超平面的距離最小。

- 長短期記憶 (Long Short-Term Memory, LSTM)

LSTM為時間循環神經網路 (Recurrent Neural Network, RNN) 中的一種模型，其適合處理時間序列資料，表現通常會較時間循環神經網路來的更好，因為 LSTM 補足 RNN 的梯度消失，而

LSTM 透過記憶功能來解決這個問題 。

LSTM利用三個控制閥來決定記憶的儲存與使用，分別為輸入閥(Input Gate)、輸出閥 (Output Gate) 以及遺忘閥 (Forget Gate)。

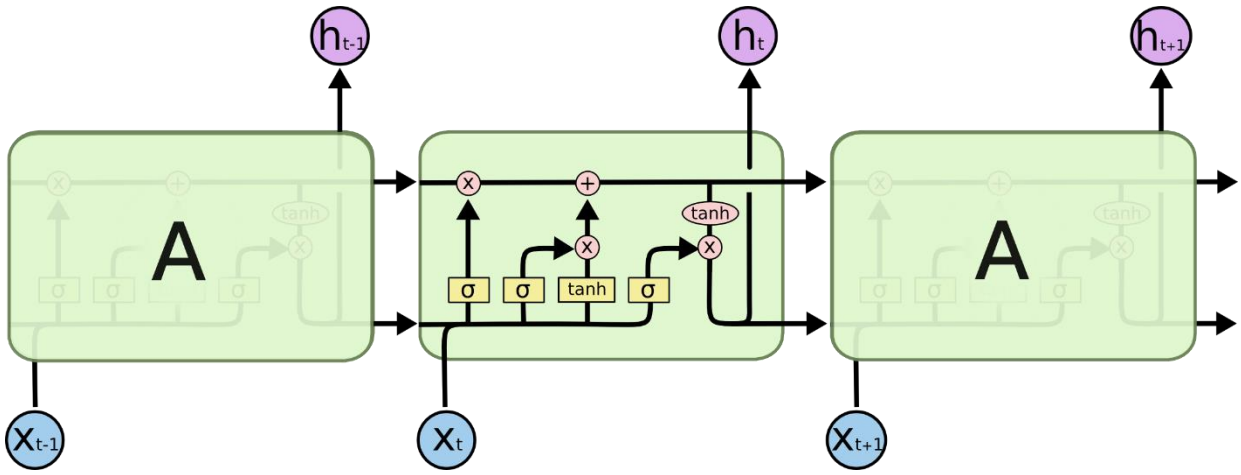

圖 1. LSTM 示意圖

從可以看到圖 1中有一條由左至右貫穿的水平線，其代表單元狀態而黃色的為閥。單元狀態會透過閥做調控。接下來會依照圖 1 由左至右介紹各個閥:

公式()~()中，LSTM 輸入為 $x_t \in \mathbb{R}^d$，遺忘閥為 $f_t \in \mathbb{R}^h$，輸入閥為 $i_t \in \mathbb{R}^h$，輸出閥為 $o_t \in \mathbb{R}^h$，隱藏狀態為 $h_t \in \mathbb{R}^h$，單元狀態為 $C_t \in \mathbb{R}^h$，$W \in \mathbb{R}^{h \times d}$ 、$b \in \mathbb{R}^h$為訓練中的矩陣及網路學習計算元值。

1. 遺忘閥:

   遺忘閥可以決定哪些資料需要被記住或是遺忘，它透過邏輯函數 (σ)來決定在上一層輸出的資料 (h_{t-1})以及這一層新增加的輸入 (x_t)中多少比例的資料需要被遺忘，邏輯函數通常使用乙狀函數 (Sigmoid)，若其數值趨近於零則表示資料要被遺忘，不會帶進下一層。其公式為:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_t) \tag{6}$$

2. 輸入閥:

輸入閥的功能則是決定是否將新輸入的資料記錄到單元狀態中，包含資料是否會被記錄以及對主要單元進行更新。輸入閥與遺忘閥一樣透過 σ 來決定有多少新輸入資料 (i_t) 會被記錄，同時 tanh 激勵函數會計算出一個向量 ($\widetilde{C_t}$)，決定出更新到主要單元 (C_t)的資料有多少。其公式為:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{8}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \tag{9}$$

輸出閥:

輸出閥使用 σ 來決定有多少資訊會被輸出(o_t)， C_t 透過 tanh 激勵函數來決定主要單元中要被輸出進入下一層的資料 (h_t) 是什麼。其公式為:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

$$h_t = o_t * \tanh(C_t) \tag{11}$$

2.4.2 實際研究

在預測 COVID-19 方面，有研究證實使用最簡單的線性回歸也能獲得一定的效果，在印度尼西亞為研究背景的研究中使用羅吉斯回歸 (Logistic Regression) 來針對 COVID-19 的爆發進行預測。預測所使用的訓練資料為 2020/03/02 至 2020/11/12 的印度尼西亞 COVID-19 每日確診人數以及康復人數，並使用平均絕對誤差 (Mean Absolute Error , MAE) 以及均方誤差 (Mean Square Error , MSE) 來衡量模型的準確性。研究結果顯示羅吉斯回歸確實可以針對 COVID-19 的爆發進行具有準確性的預測[48]。

另外在印度的研究使用多元線性回歸 (Multiple Linear Regression) 與線性回歸 (Linear Regression) 進行比較，結果顯示多元線性回歸的預測結果較線性回歸來的好[49]，因此選擇較多的特徵進行預測可能會帶來更好的結果。

另外也有研究使用其它預測模型，分別為線性回歸[49]、隨機森林回歸（Random Forests Regression , RFR）[50]、支持向量回歸 (Support Vector Regression , SVR) [51]、以及長短期記憶（Long Short-Term Memory，LSTM）[52]。研究[53]以各國的確診病例總數、死亡總數、新確診病例總數、新死亡總數、康復患者總數、活動患者總數和傳播分類為特徵進行預測，並計算其平均絕對百分比誤差(Mean absolute percentage error , MAPE) 以及均方根誤差 (Root Mean Square Error , RMSE) 來進行準確性比較。除了上述方法以外有研究也使用支援向量機 (Support Vector Machine , SVM) [54] 、高斯過程回歸(Gaussian Process Regression , GPR) [55] 、差分整合移動平均自我迴歸模型 (Autoregressive Integrated Moving Average model , ARIMA) [56] 以及使用流行病學的房室模型 (SEIR/SIR) [57][58] 來進行預測。

第五節 風險評估

疾病的傳播風險在流行病學上被稱為基本傳染數 (Basic Reproduction Number , R_0) ， R_0 是指在沒有任何防疫措施的介入，以及所有人都沒有免疫力的情況下，平均一個感染到某種疾病的確診者能把疾病傳染給幾個人。 而當 R_0 的數字愈大，表示能傳染的人數越多，疾病的控制愈難，病毒傳染力強，繁殖速度快。 R_0 小於一時表示其傳染力減弱，疾病會逐漸消失。R_0 等於一時，其感染力維持原狀，但並不會成為流行病 [59]。像是我們所熟知的麻疹、天花以及風疹等等，其 R_0 值都大於一，可以稱為流行病。而近兩年來肆虐全球的 COVID-19 其 R_0 值同樣也大於

一，且不同的變異株其 R_0 值也不盡相同，有研究進行計算，Delta 變異株 R_0 為 5.1 [60]，Alpha 變異株為 4 ~ 5 [61] 而近期的 Omicron變異株 R_0 值則高達 7 [62]。此外 R_0 值在不同地區其數值也不盡相同，有研究就以西歐地區與中國進行比較，研究得出的 R_0 值與中國相比較低 。因此研究指出，如果想要阻止疫情爆發所需的歐洲人口中免疫人群的比例可能會比中國所需的人口比例再來的少 [63]。但 R_0 值只用於計算流行病初期，當公共衛生干預措施開始及人群裡出現有抵抗力的人口時 R_0 值就失去其有效性。我們則需要改算有效再生數 (Effective Reproductive Number , Rt ）估算在一定時間內傳染數量的變化，去看其趨勢，得知其傳播能力。若 Rt 值小於 1 則表示疫情有好轉的趨向。再面對 COVID-19 同樣有研究使用 Rt 值估算疫情的風險。在加拿大及印度的研究[64][65]中使用不同方法去估算 Rt 值，在研究的結果中可以看到研究人員有說到估計的結果只是近似並不為準確值，但有其參考價值。因此我們可以知道 Rt 值的計算並不容易。

從上述可知因為 R_0 值以及 Rt 值的計算較複雜，所以衍生出其他評估方式。例如流行病學常使用的房室模型 (Compartmental models)[66] ，以及排隊理論 (Queueing Theory)[67]。房室模型試圖預測諸如疾病如何傳播、感染總數或流行病持續時間等事物，並估計各種流行病學參數，例如繁殖數。這樣的模型可以顯示不同的公共衛生干預措施如何影響流行病的結果，例如，在特定人群中 發布有限數量的疫苗最有效的技術是什麼。排隊理論則是處理到達具有固定數量服務器的設施的服務客戶的分析，雖然將排隊論與流行病學聯繫起來的工作相對較少，但也有研究[68]將排隊理論應用到房室模型上。另外排隊理論對於評估醫療量能的部分也會較其他方法來的直觀，有許多研究[69][70]使用排隊理論來評估醫院的病床使用，或是醫療資源分配。但對於預測流行病的流行時長，在葡萄牙的研究[71]使用排隊理論來預估流感的流行時長。

同樣在面對 COVID-19 時，我們可以使用排隊理論來預估醫療量能

是否充足。以荷蘭的研究[69]為例，研究COVID-19 的入院消耗了荷蘭ICU 和診所的相當大的資源，研究還發現當第二波疫情發生時每日入院人數還沒有達到第一波的峰值水平，但第二波的上升和下降速度要慢得多，導致患者人數和住院天數更高。這不可避免地會導致更多的醫療資源的消耗。同樣的在澳大利亞也有關於面對 COVID-19 時需要有多少床位才能負荷的疫情進行研究[72]。研究同樣使用排隊理論來進行分析，結果表明對ICU 床位的迫在眉睫的需求可能會壓倒附近最大的澳大利亞醫院的容量 。同樣在美國對於在 COVID-19 期間的呼吸機資源分配進行研究[73]，先預測 COVID-19 的患者生存概率以及呼吸機使用時間，再使用排隊理論進行分析，並得到一個合理的分配方案。從上述可以看到對於分析目前疫情風險，除了可以預防未來的爆發也可以評估政府對於疫情的控制能力。在評估政府的控制能力方面可以從醫療體系來看。當醫療量能還足夠負荷當前感染人數時表示疫情還在政府可控制的範圍，一旦不足時表示政府失去對疫情的控制。這個結果可以在台灣的研究 [74]中發現，研究利用各國確診病例、死亡病例和康復病例的數量來評估不同國家實時控制 COVID-19 流行病的能力 。上述這些研究可以證實排隊理論可使用於評估疫情的風險。

2.5.1 排隊理論

排隊理論是一種簡單的數學研究，用於預測隊列或等待區域的長度以及等待時間而最簡單的排隊理論為圖 2所示。

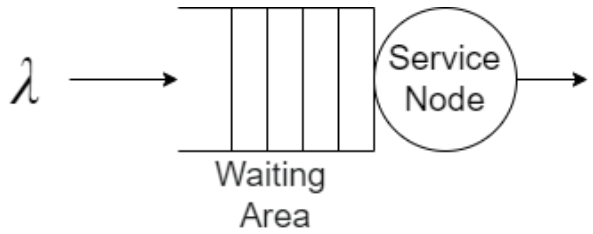

圖 2. 排隊理論示意圖

可以從圖 2看到 λ 也就是輸入，在排隊理論可以視為顧客，排隊區

(Waiting area)，在顧客進入等候區之後要進行排隊，排隊區是採用先進先出的模式，但也有模型可以設定有特權要插隊為更進階的模式。服務(Service node)，可以視為服務顧客的地方，其為正在處理工作的單位。圖()其假設只有一個服務櫃台。而基於排隊理論而提出的公式為利特爾法則 (Little's law) [75]其公式為:

$$L = \lambda W \tag{12}$$

L為平均等候的顧客人數，λ 為長期的顧客有效抵達率，W 為顧客平均等待時間。

第三章 研究方法

本研究目標為預測分析 COVID-19 的爆發風險。研究會透過美國國家環境保護局[76]架設之室外空氣品質感測器，獲取各州感測器之 NO_2 、SO_2 、 Ozone、 $PM_{2.5}$ 、PM_{10} 及 CO 數據。以及 Apple Maps [77]所提供移動趨勢報告，其報告包含各州開車、走路以及搭乘大眾運輸工具數據。還有完整接種疫苗的人數比，透過上述的資料針對各州的感染風險進行研究。圖 3為研究架構，此章節會介紹資料收集、相關性分析以及資料前處理，而在第二節研究則會介紹如何將處理好的資料放入模型內進行訓練，最終獲得預測值。在第三節則會介紹研究如何將預測結果帶入排隊理論中進行分析以及研究如何使用排隊理論的結果進行風險分析。研究會依照圖 3 的研究架構進行。

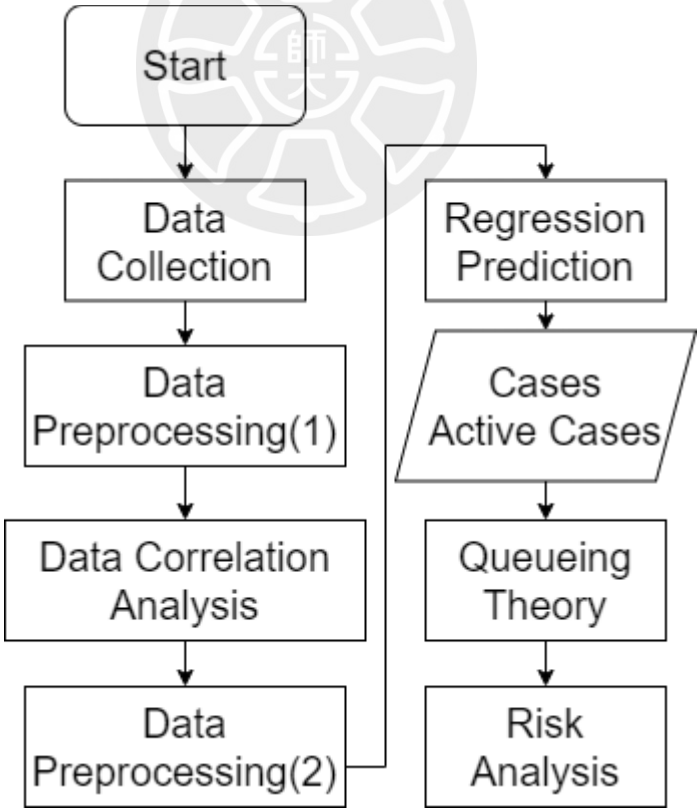

圖 3. 研究架構圖

第一節 資料來源、相關性分析及前處理

3.1.1 資料來源

本研究選定美國為實驗對象，因為其人口數較充足，自2020年至今經歷過不只一次的疫情爆發，更適合用來分析每次疫情爆發間的差別。透過 Worldometer[78] 及 USAFacts [79] 獲取美國各州從 2020/01/13 起每日確診人數及活動病例。空氣汙染物的資料則是來自美國國家環境保護局官方網站[76]獲取自 2020/01/13 起每日各州的資料，分別為每日最大 1 小時 NO_2 濃度、每日最大 1 小時 SO_2 濃度、每日最高 8 小時 Ozone 濃度、每日平均 $PM_{2.5}$ 濃度、每日平均 PM_{10} 濃度以及每日最高 8 小時 CO 濃度。而人口流動資料則是來自 Apple Maps [77]所提供自 2020/01/13 起的移動趨勢報告，以各州 2020/01/13 的交通量為基準值每日的開車、走路以及搭乘大眾運輸工具的變化量，Apple Maps 的交通量為統計使用者向Apple Maps 提出導航要求的數量。研究選擇Apple Maps 而非 Google Mobility 的原因為根據 Statista 數據庫所提供美國手機系統統計[80]，iOS 系統在美國使用人口中佔超過半數的比例，因此研究會選擇使用較多人使用的Apple Maps 交通量數據做為研究。兩劑疫苗接種的人數比例則是來自 Our World in Data [81]所提供美國各州資料。由於美國州數眾多本研究將挑選其中四個州作為主要研究目標，其分別為加利福尼亞州、佛羅里達州、紐澤西州以及紐約州，資料收集的時間為2020/01/13 ~ 2021/12/31 共 719 天。

以下對於本研究所選擇之空氣汙染物其主要來源進行介紹，NO_2 來自於石化原料的不完全燃燒以及汽車排放之廢氣。SO_2 工業過程中多產生以及燃燒煤或石油時也會排放。Ozone 為二次污染物，其是因為工業過程及汽車排放之廢氣在經過化學反應後，而產生。也就是空氣中的有機揮發氣體 (VOC) 和來自工廠和汽機車排氣管所排出的廢氣 (氮氧化

物 , NOx)，被強烈陽光照射後，和空氣中的氧結合形成臭氧。$PM_{2.5}$ 來源同樣多來自於工業燃燒及汽機車的廢氣排放為主。CO 來自於汽車所排放之廢氣，因燃燒不完全而產生。可以發現這些汙染大多都與汽機車也就是交通行為有相關。

3.1.2 資料前處理

資料收集完成後，首先我們針對空氣汙染物資料進行缺失值處裡處理方法分為兩種，第一種為監測站的空缺是一天長度的缺失值，這種空值的部分，研究會利用前後天的汙染物數值平均數去進行補值。而第二種的缺失值為缺失長期資料，這種資料若是使用平均值處裡可能會導致資料特徵的消失，因此研究會將缺失天數大於一的監測站在後面進行監測站是值平均時暫時剔除。處理好缺失值後由於空氣汙染物資料為每個州內各個監測站的空氣汙染物資料，因此研究需要將該州有數值的監測站的數據進行算術平均後將結果視為是該州的空氣汙染物資料。最後完成前處理後，研究發現各州多數監測站 PM_{10} 的數值有長期缺漏的情形，進行平均數補值可能會造成數據不真實，所以將 PM_{10} 從特徵中剔除。

第二個進行前處理的資料為 COVID-19 中的活動病例資料(Active Cases) ，因為網站所提供的活動病歷資料其數據的公式為:

$$Active\ Cases = Total\ Cases - Total\ Deaths - Rcovered$$

其為累積資料，而研究想要的是每日活動病例，因此研究針對這個資料進行前處理，會使用公式:

$$AC_t{}' = AC_t - AC_{t-1} \tag{13}$$

AC為活動病例資料集，而 AC_t 則是在 t 時間點的 AC 資料。將 t 時間點的 AC 資料減去$t-1$時間點的資料後我們就可以獲得 t 時間點的每日活動病例($AC_t{}'$)。

最後會針對疫苗接種資料進行前處裡，由於美國疫苗施打的日期

是由 2021 年年初開始，所以在2020 年時的資料將會進行補零，零值有其存在的意義，確實表明 2020 年時疫苗還為研發完成，市民們也未接種過疫苗。

3.1.3 各州資料分析

進行完第一次資料前處理後我們將會針對處理好的資料進行一次相關性分析。首先會針對各州的空氣汙染物資料進行分析了解資料的特性。

	加利福尼亞州	**佛羅里達州**	**紐澤西州**	**紐約州**
人口	39,538,223	21,538,187	9,288,994	20,201,249
土地面積	423,970 平方公里	170,304 平方公里	22,588 平方公里	141,299 平方公里

人口數為美國人口調查局 2020年6月調查結果

表 1.州人口及土地面積表

表 1.州人口及土地面積為加利福尼亞州、佛羅里達州、紐澤西州以及紐約州的人口總數、土地面積及是否有公共運輸。從表格數據中可得知，人口密度由高至低分別為紐澤西州、紐約州、佛羅里達州，最後則為加利福尼亞州。其中紐澤西州的人口密度為其他三州的四倍。在交通方面四個州都具有鐵路及公路系統，而在公共運輸方面，巴士交通則是只有紐澤西州以及紐約州具備。

資料搜集完成後，對各州的資料進行分析，研究會先使用2019 年四個州的空氣汙染平均數進行觀察了解各州特性。再使用各州 2019/01/01 ~ 2021/12/31 的空氣汙染物資料依照疫情出現前後進行分段比較，繪製盒狀圖將資料視覺化，進而了解各州在疫情發生前後的空氣汙染物變化。

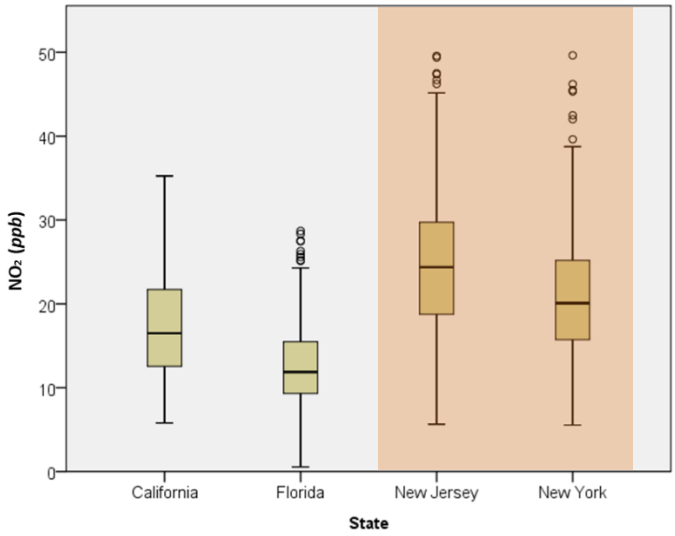

圖 4. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州2019 年 NO_2 盒狀圖

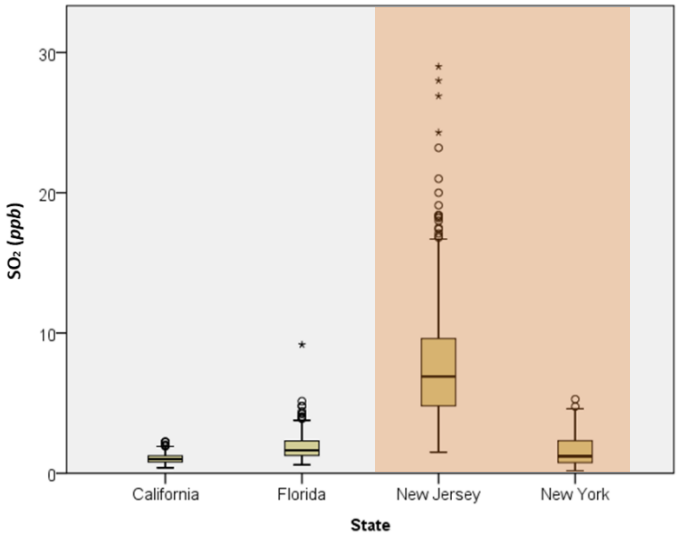

圖 5. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州2019 年 SO_2 盒狀圖

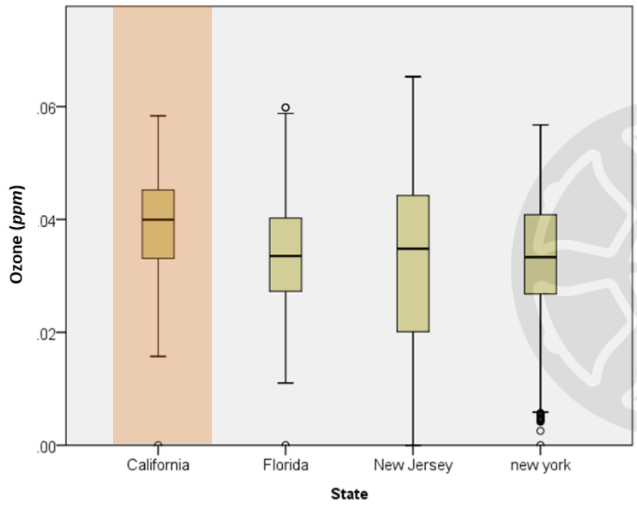

圖 6. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州 2019 年 Ozone 盒狀圖

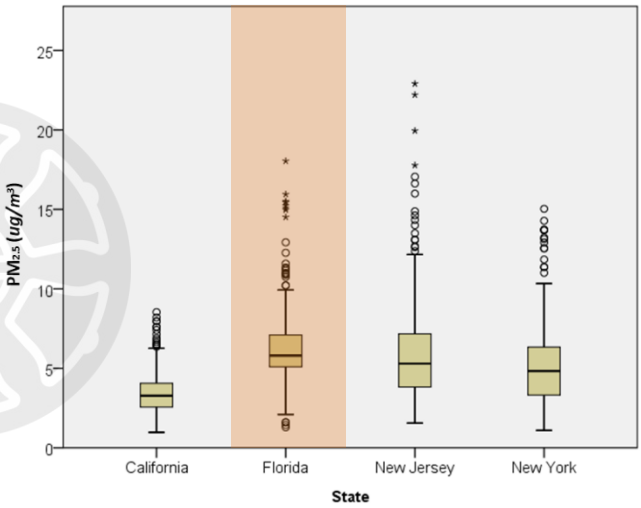

圖 7. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州 2019 年 $PM_{2.5}$ 盒狀圖

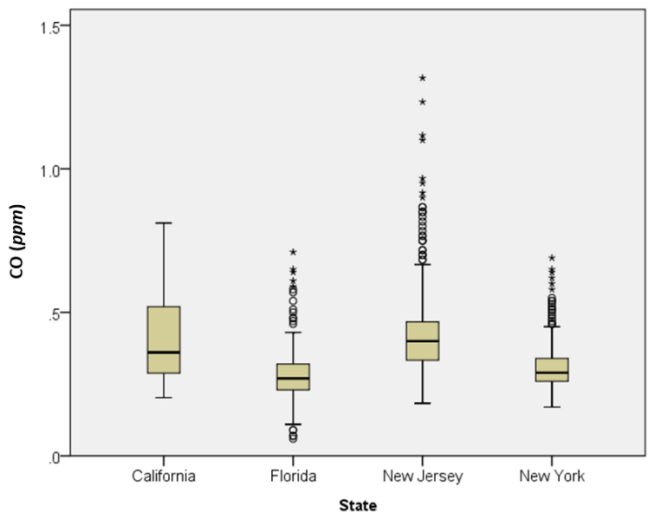

圖 8. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州 2019 年 CO 盒狀圖

美國肺臟協會 (ALA) 20 年空氣狀況報告中[82]有提到加利福尼亞州較糟糕的空氣品質，其為美國空氣汙染嚴重的地區。加利福尼亞州的空氣污染源來自汽車交通以及工業污染。該州的當地地形，特別是在農業地區，加上溫暖的天氣，有助於將污染捕獲在谷壁內，從而增加臭氧水平。從圖 6中也可以看出加利福尼亞州的臭氧水平高於其他州。而從圖 4及圖 5可以看到紐澤西州及紐約州有較高的趨勢，紐澤西州人口較為密集其主要的汙染源來自汽車排放以及燃燒化石燃料的發電廠。而紐約州也有相同情形，大部分污染來自車輛排放、用於發電的化石燃料燃燒和住宅建築供暖及製冷。佛羅里達州的空氣汙染情形可以從圖 7看到，以$PM_{2.5}$ 為大宗，根據佛羅里達州環境保護部的報告中可以看到，佛羅里達州的 $PM_{2.5}$ 來源可分為內部及外部，外部為撒哈拉沙漠吹過大西洋的沙塵，內部則為野火的影響。由上述可以看出各州的汙染物生成原因較為不同，紐約州及紐澤西州主要受交通及發電廠影響、佛羅里達州則受野火影響而加利福尼亞州是因為地形及其氣候。

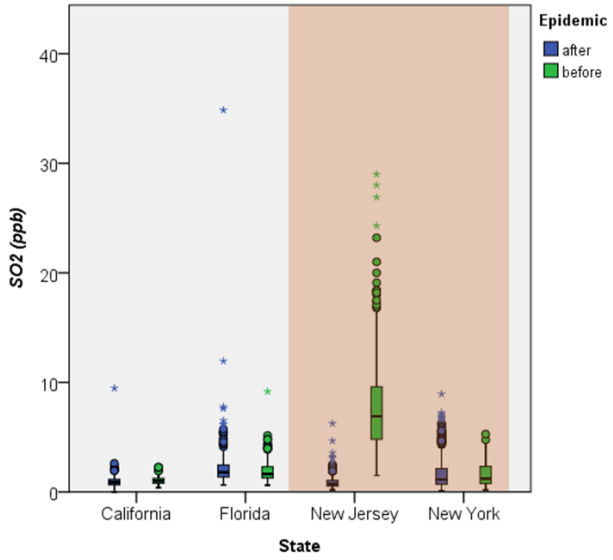

圖 9. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 SO_2 盒狀圖

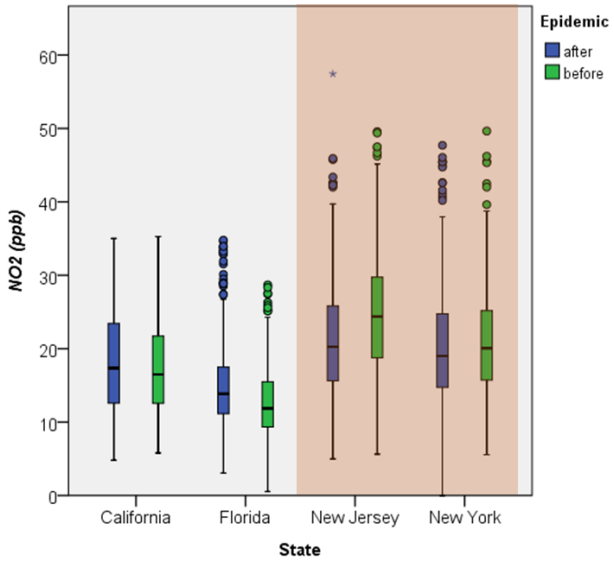

圖 10. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 NO_2 盒狀圖

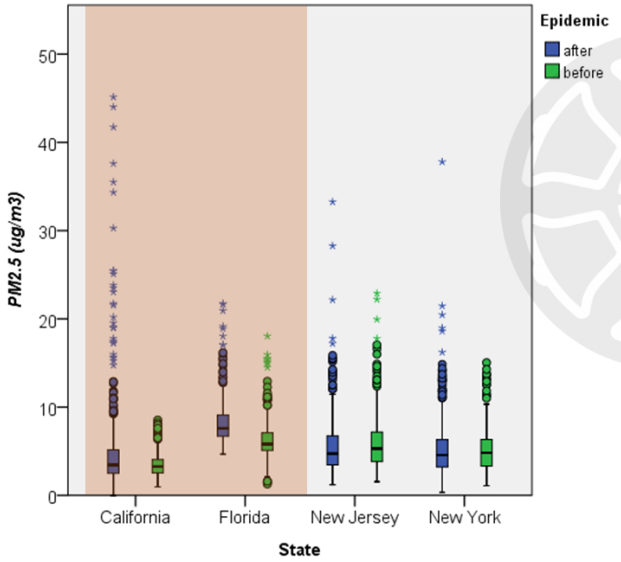

圖 11. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 $PM_{2.5}$ 盒狀圖

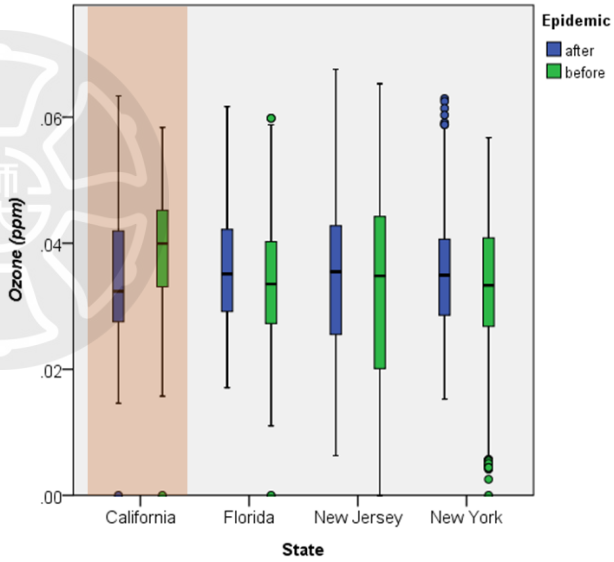

圖 12. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 Ozone 盒狀圖

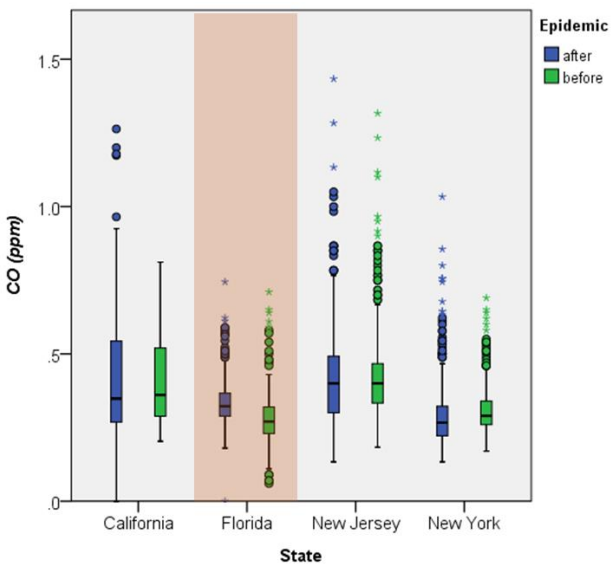

圖 13. 疫情前後加利福尼亞州,佛羅里達州,紐澤西州以及紐約州的 CO 盒狀圖

從圖 9~ 圖 13中我們可以看到在上面由提到的汙染物來源多為車輛排放的地區也就是紐約州及紐澤西州在疫情前後 NO_2 及 SO_2 汙染物水平有著較明顯的下降。但 $PM_{2.5}$ 於佛羅里達州及加利福尼亞州而言則是有些汙染物平均水平有上升的情形，推測原因可能與這兩周的野火情形相關，野火的燃燒產生$PM_{2.5}$。而在 Ozone 的情形則可以發現原本 Ozone 濃度最高的加利福尼亞州有著明顯的下降。CO 則是只有佛羅里達州有最為明顯的上升。五種汙染物在疫情發生前後都能看到存在著起伏，這是否意味著它們與 COVID-19 存在著一定的相關性，又或是在 COVID-19 中扮演傳播的角色，後面研究會針對這一點做更進一步的分析。

3.1.4 資料相關性分析

為了解研究所選擇的影響因素是否與COVID-19有相關性，研究針對這些資料進行了相關性分析，我們使用皮爾森基差相關係數[83]以確保其相關性。皮爾森基差相關系數適用於測量兩個變量之間的線性關係程度。皮爾森基差相關系數 (Pearson Correlation) 以測量兩個變量之間的

線性關係程度。它被稱為測量感興趣變量之間關係的最佳技術，因為它基於協方差方法。它提供了有關變量之間聯繫幅度的信息，並給出了這種關聯的方向。計算出的相關值應該在 -1 和 1 之間，分別表示強負相關或強正相關，以下為其公式:

r為相關係數，n為資料集大小，x_t為(空氣汙染物/ 人口流動/ 疫苗接種率) 在 t 時間的資料，y_t 則是確診人數在 t 時間的資料，$\bar{x}$, $\bar{y}$則是資料的算術平均數

$$r = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}} \tag{14}$$

本研究會先根據不同防疫政策將資料做切割，主要以防疫條件最為嚴格的居家令作為劃分點，分為還未執行防疫措施時期、居家令執行時期以及解除居家令時期。研究所選用的居家令時期為全州的居家令，只有特定區域的居家令並不列入考慮。而在解除居家令之後，州政府都有發布戴口罩等相關較不嚴苛的防疫措施。另外疫苗開始施打為 2020 年底開始，所以在前兩個時期還未有數據。研究選擇將資料劃分不同時期的主要原因為不同的防疫政策對於空氣汙染物及人口流動有一定的影響，在這些影響下研究所選取的特徵對於確診人數及死亡人數也有著不一樣的相關性。所以我們會先將不同時期的空氣汙染物及人口流動資料依照表 2所列出的時間段對資料進行相關性的觀察。

Period	**加利福尼亞州**	**佛羅里達州**	**紐澤西州**	**紐約州**
還未執行防疫措施 **Period 1**	2020/01/13~ 2020/03/18	2020/01/13~ 2020/03/31	2020/01/13~ 2020/03/20	2020/01/13~ 2020/03/20
居家令 **Period 2**	2020/03/19~ 2020/05/07	2020/04/01~ 2020/04/30	2020/03/21~ 2020/06/09	2020/03/21~ 2020/05/14
解除居家令 **Period 3**	2020/05/07~ 2021/11/04	2020/05/01~ 2021/11/04	2020/06/10~ 2021/11/04	2020/05/15~ 2021/11/04

表 2. 不同時期的起迄時間表

State	Period	NO_2	Ozone	$PM_{2.5}$	SO_2	CO	Driving	Walking	Transit	Vaccinated
California	Period 1	-0.564**	0.130	-0.533**	-0.321**	-0.578**	-0.463**	-0.483**	-0.802**	0
	Period 2	-0.148	0.280*	0.302*	-0.032	-0.151	0.211	0.283*	-0.355*	0
	Period 3	0.356**	-0.369**	-0.020	0.224**	0.479**	-0.522**	-0.430**	-0.468**	-0.387**
Florida	Period 1	-0.055	0.034	0.424**	0.462**	-0.022	-0.603**	-0.589**	-0.693**	0
	Period 2	0.260	0.028	-0.086	-0.064	0.038	-0.363*	-0.357	-0.305	0
	Period 3	0.281**	-0.071	0.014	0.049	0.141**	-0.151**	-0.084	-0.011	-0.388**
New Jersey	Period 1	-0.066	0.152	-0.073	-0.074	-0.020	-0.492**	-0.363**	-0.609**	0
	Period 2	0.186	0.167	-0.070	-0.193	-0.103	-0.161	-0.183	-0.270*	0
	Period 3	0.107*	-0.098*	-0.016	0.139**	-0.081	-0.127**	-0.128**	-0.203**	-0.173**
New York	Period 1	-0.186	0.232	-0.230	0.005	-0.182	-0.365**	-0.362**	-0.592**	0
	Period 2	-0.124	-0.060	0.023	-0.043	-0.147	-0.084	-0.167	-0.188	0
	Period 3	0.235**	-0.108*	0.036	0.269**	0.073	-0.479**	-0.397**	-0.350**	-0.257**

表 3. 根據不同時期，空氣汙染物、人口流動及疫苗施打率與確診人數相關性分析表
(**為在0.01層級上有顯著性，*為在0.05層級上有顯著性)

可以從表 3與確診人數的相關性看到與上一小節分析的結果一致，紐澤西及紐約兩個空氣汙染情形較相似的州，其在三個時期下空氣汙染物、人口流動與確診人數間的相關性也較為類似。在疫情初期還未進行防疫措施時，人口流動會是疫情傳播的一大因素 。而進到居家令時期時，因為被限制人口流動，所以相關性有下降的趨勢，而這時空氣汙染物中的 SO_2 其相關性逐漸上升。在上一小節有提到這兩州的 SO_2 平均濃度本身就較其他兩州來的高，是因為本身交通污染的程度較高。推測是因為居家令限制了大眾運輸的乘載量，進而導致有需求的人外出改為汽車，進而有較高的 SO_2 汙染。當居家令解除後，人們重新回歸生活，其餘汙染物水平的影響也逐漸上升，而人口流動的影響則較剛開始有下降的情形。而對於汙染情形本身就比較嚴重的加利福尼亞州而言，在第一個時期與人口流動相比，空氣汙染物與確診人數有較高的相關性。但隨著防疫政策的影響，在居家令時期可以看到空氣汙染的相關性降低，反而是汽車與步行的相關性增加。而結束居家令後空氣汙染物的相關性又再度升高，人口流動的相關性下降。由此可以看出加利福尼亞的社會型態與紐澤西及紐約都不相同，其受空氣汙染的狀況也較為嚴重。佛羅里

達州則是最為特殊的一個州，可以從上一小節的分析中看到，佛羅里達州本身的$PM_{2.5}$有著較高的水平。因此我們可以看到$PM_{2.5}$在第一個時期就有與確診人數有著正相關，且同時也能看到人口流動的相關性。這些相關性在居家令執行後快速的消失，但居家令結束後，空氣汙染的相關性重新恢復。而佛羅里達州則是四個州裡唯一一個顯示人口流動相關性不足的州。這與第二章的相關研究[13]中所提到人口較不密集的地區對於疫情造成的人口流動變化會比較不明顯示，有相同之處。空氣汙染物、人口流動及疫苗施打率與確診人數間存在些相關性。可以看到有些州是因為居家令而導致人口流動的相關性降低，以及空氣汙染物在疫情初期對於傳播的貢獻，而在進行防疫措施後汙染物的貢獻有降低但也並未完全消失。

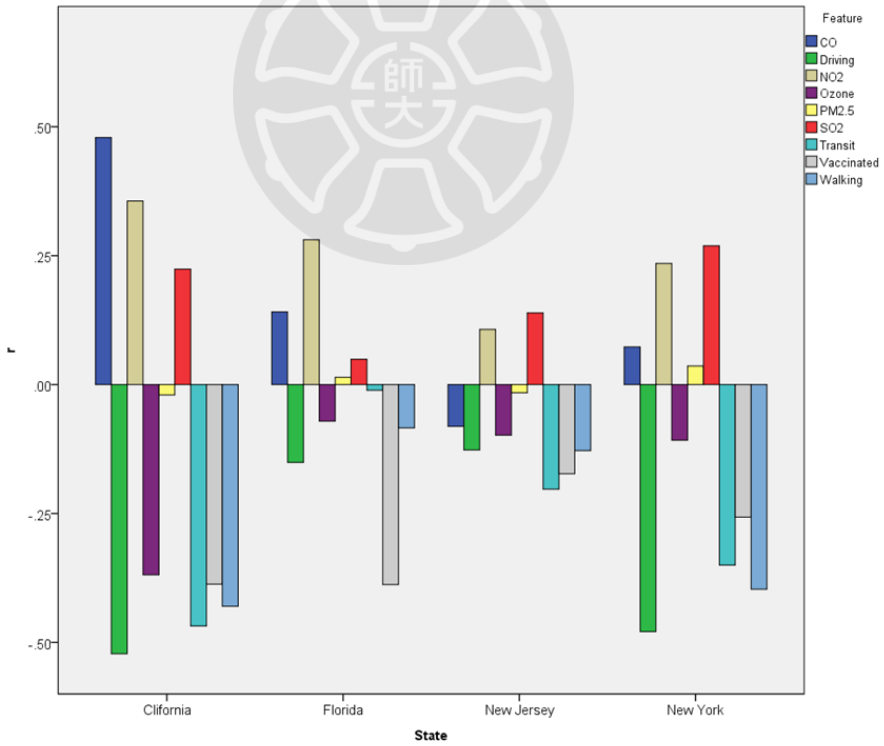

圖 14. Pearson Correlation 結果圖

結果我們可以從圖 14 看到，其中可以看到 NO_2 對於確診人數都有著較高的相關性。Ozone 及 SO_2 與確診人數有一定的相關性，而 CO 對於確診人數有相關性則較低。但最出乎意料的是 $PM_{2.5}$ 的影響性並沒有

想像中的高，主要原因有二，第一點為研究採用的資料為平均值，有些城市可能確實有高相關，但因為研究需要州資料將數值平均後造成相關性降低。第二點，$PM_{2.5}$ 需要在人長期在一定濃度的環境下才能有其影響如有著平均較高濃度的佛羅里達州以及$PM_{2.5}$的影響可能不是及時的，所以下一小節我們將會分析滯後影響對於資料相關性會有什麼影響。

3.1.5 滯後影響

由於空氣污染對於人體的影響性可能不是及時的，需要短期待在這個環境才會有較明顯的影響或是經過一段時間後才能看到其影響性，這個過程就會被稱為滯後效應 (Lagged Effect)。所以本研究將計算空氣汙染物與確診人數間的滯後天數。而最佳的滯後天數則是會透過 Pearson correlation 來確定，有較強的相關性的天數即為本研究所選擇的滯後天數。

為了計算出最佳的滯後天數，本研究將空氣汙染數及確診人數根據所假設滯後天數 (1, 3, 7, 12, 14) 進行算術平均數的運算，其公式如下:

t為時間，d為研究所選擇的滯後天數 (1, 3, 7, 12, 14)，AP_t 則是空氣汙染物在 t 時間的資料，$\overline{AP_t}'$ 為研究所需要的根據 d 天數滯後的 t 時間空氣汙染物資料。

$$\overline{AP_t}' = \frac{\sum_{t=d}^{t-d-1} AP_t}{d} = \frac{AP_t + APt_{t-1} + \cdots + AP_{t-d-1}}{d} \tag{15}$$

並將得出的資料透過 SPSS 計算它的皮爾森積差相關系數 (Pearson Correlation)，並以圖像化的方式呈現圖 15~ 圖 18，而有較高相關性的天數將會成為本研究的之後研究分析的滯後天數。

California

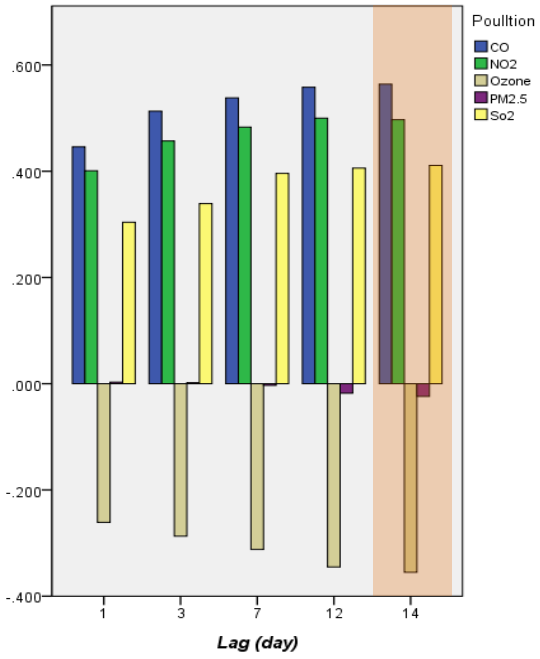

圖 15. 加利福尼亞州不同滯後天數相關係數圖

Florida

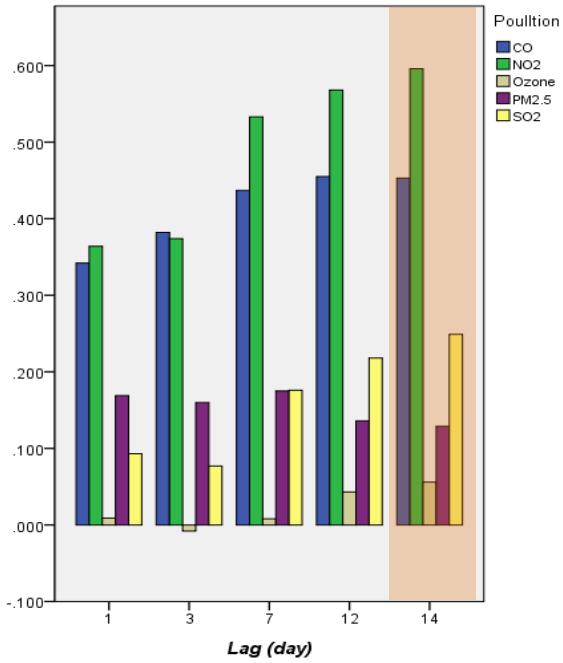

圖 16. 佛羅里達州不同滯後天數相關係數圖

New Jersey

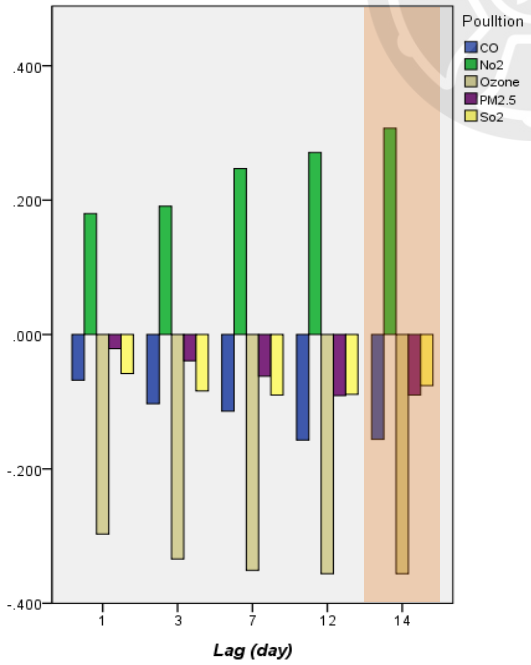

圖 17. 紐澤西州不同滯後天數相關係數圖

New York

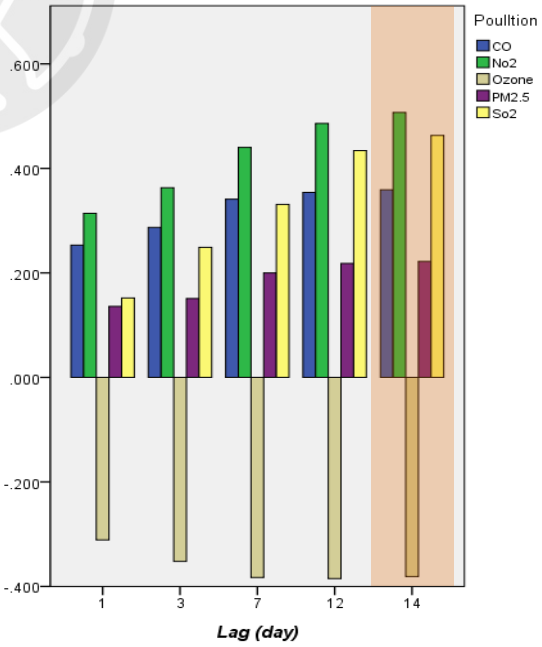

圖 18. 紐約州不同滯後天數相關係數圖

從圖 15~ 圖 18中可以看到與預期相同隨著滯後天數的增長相關係數有增加的趨勢，證實空氣汙染物的影響會需要短期的存在而非立即就能發現影響。根據圖 15~ 圖 18 可以發現大部分的相關係數在滯後 7 天都有較明顯的增加，而在滯後 12 天及滯後 14 天時相關係數來到高峰。其中可以看到NO_2 、CO 以及 Ozone 的貢獻是較高的，雖然有幾個汙染物的相關係數在 滯後 12 天時會比滯後 14 天再高一點，但縱觀整體滯後 14 天會是四個州的空氣汙染物相關係數都是較高的天數。且原先相關性較低且顯著性不足的 $PM_{2.5}$ 在經過滯後計算後，相關性上升且顯著性也有明顯增加，因此最後將列入計算。本研究會選擇滯後天數為 14 天的資料進行後面的研究。另外也可以觀察到每個州的係數以及相關性都有所不同這點在第二章時有相關研究[14]就有提出，不同地方所受到的影響不盡相同，所以相關性也會有所不同。這個問題本研究會採取每個州分別訓練一個模型來解決，這樣不同相關性就不會產生干擾反而會有較好的結果。

3.1.6 資料合併及切割

在完成相關性分析後，研究發現計算出來的相關係數結果有存在線性相關。但因為資料有可能存在我們為發現的非線性關係，所以模型會分為兩種，首先會使用線性的Multiple Linear Regression，非線性Polynomial Regression、SVR，上述這些回歸方式不會考慮到資料的時序性，所以還會加入LSTM 來補足時間序列的特徵。不同的回歸預測模型會有不同的切割方式下面會將資料切割方式分成兩種進行介紹:

- 機器學習資料合併及切割

機器學習的模型並不會抓取時間序列特徵，所以並不會對於時間方面做其他調整。我們先將每日空氣汙染物 (NO_2, $PM_{2.5}$, SO_2, Ozone, CO) 以公式(15)求出滯後 14 天的空氣汙染物資料，再將這個資料集與每日人

口流動(Driving, Walking, Transit) 以及每日疫苗施打率 (Vaccinated) 資料進行合併每州獲得一個資料集(D) 其大小為 X ∈ R^{n*9} 的矩陣 ，n 表示資料所包含的天數。COVID-19 每日確診 (Cases, C) 以及每日活動病例 (Active Cases, AC) 將前 13 天資料刪除，因為前 13 天無法進行滯後 14 天的計算，最後兩個資料集各自形成一個 y ∈ N^{n*1}矩陣，n 與資料集 D 的長度相同。

在訓練模型時因為機器學習模型資料不需要按照時序輸入，所以會將資料集 D, C, AC 以隨機選取的方式挑選出 0.8 n的資料及成為訓練集，而剩餘的0.2 n則成為驗證集，n 為資料長度，最後我們可以獲得訓練集 (D_{R_Train} ， C_{R_Train} ， AC_{R_Train}) 與驗證集 ($D_{R_Validate}$ ， $C_{R_Validate}$ ， $AC_{R_Validate}$)。

- LSTM資料合併及切割

LSTM 模型我們使用多對一 (many to one) 的模型架構，資料使用的是還未經過滯後天數計算的資料及為最原始的資料。首先將原始的每日空氣汙染物 (NO_2, $PM_{2.5}$, SO_2, Ozone, CO)、每日人口流動(Driving, Walking, Transit) 以及每日疫苗施打率 (Vaccinated) 進行合併。再將以14天為一組進行資料重新排列。原始資料其中一列為例。$D = [NO_{2_t}, SO_{2_t}, \cdots Vaccinated_t]$，$t$ 為預測當天，而經過重新排列過的資料集其樣式為$D = [NO_{2_t-13}, SO_{2_t-13}, \cdots Vaccinated_t]$。可以看到特徵從原本的只有 9 種當天數據變為 9 種共 14 天的數據 ，研究將利用一組 14 天的來進行預測。最後將 D 進行資料切割，轉變為資料集 D_{14}其大小為X ∈ R^{n*14*9}的三維矩陣，n 表示資料所包含的天數。COVID-19 每日確診 (Cases, C) 以及每日活動病例 (Active Cases, AC) 將前 13 天資料刪除，因為前 13 天無法進行滯後 14 天的計算，最後兩個資料集各自形成一個 y ∈ N^{n*1}矩陣，n 與資料集 D 的長度相同。

在進行模型訓練時一樣會將D_{14}, C 及 AC 按照時間序列以 8:2 的比例分為訓練集 (D_{14_Train}, C_{14_Train}, AC_{14_Train}) 以及驗證集 ($D_{14_Validate}$, $C_{14_Validate}$, $AC_{14_Validate}$)。

第二節 資料預測

在預測 COVID-19 相關人數方面，會根據圖 19的流程進行訓練及驗證，因為預測目標有兩個分別為預測每日確診人數 (P_C) 以及每日活動病例 (P_{AC})，所以每種模型會各自有兩個不同預測目標的模型產出。將上一節所完成的前處理資料帶入，先使用訓練集資料，將模性訓練好再使用驗證集資料對模型進行驗證。最後預測出的值會進行評估。

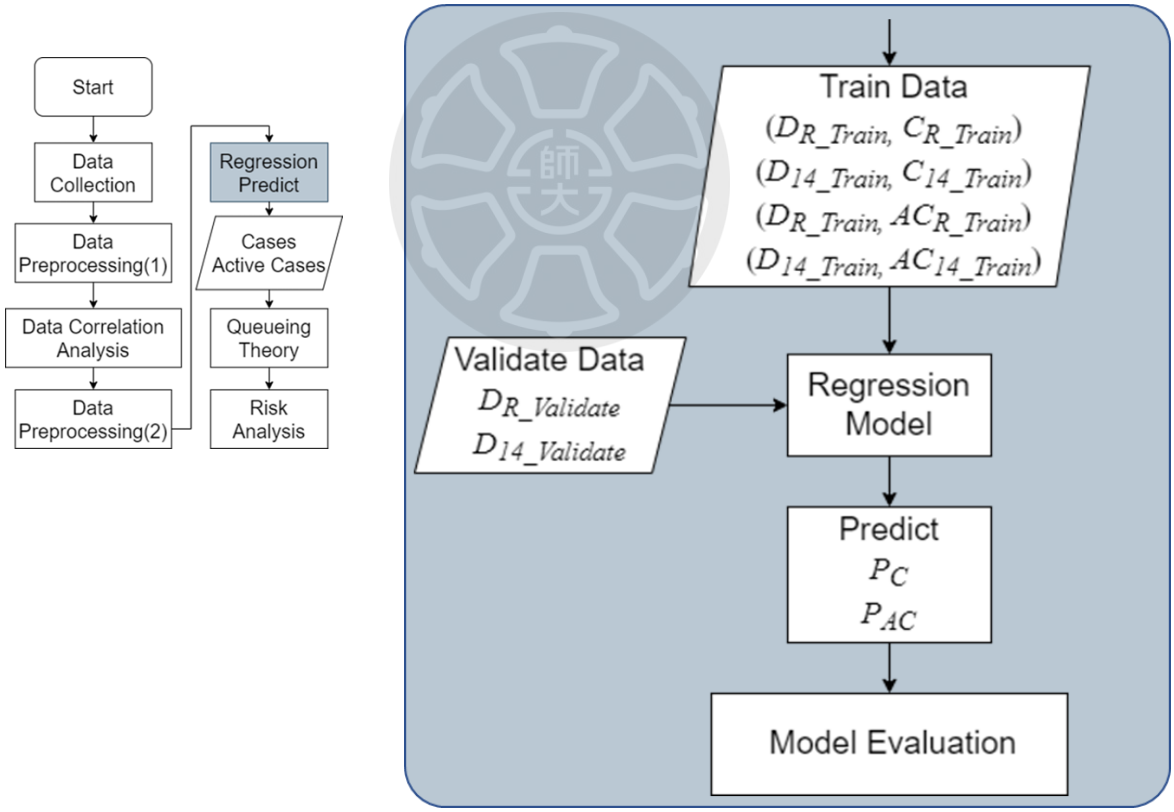

圖 19. 回歸預測流程圖

3.2.1 機器學習模型

在使用機器學習的回歸模型方面可以分為線性回歸或非線性回歸，在線性回歸方面使用到最簡單的多元線性回歸。將特徵輸入後找到一條包含最多特徵的直線，其訓練速度快，但容易被資料中的離群值影響以及如果資料存在非線性相關性會被忽略掉。為補足線性回歸模型的缺點使用非線性回歸模型進行訓練也就是我們的多項式回歸。可以使用多次方的方式來找到一條與特徵資料最貼近的多項式，但訓練時要小心容易發生過擬合 (overfitting) 的情形。本研究使用到的特徵較多，因此額外選擇了有利於高維特徵的SVR， SVR 是將所有的輸入特徵投射到空間中進而找到一個超平面與最遠特徵點的距離最，小平面的搜尋方式可分為線性(linear)以及非線性(poly, rbf)，會有利於我們的預測。但以上三種模型都無法兼顧時間特徵。

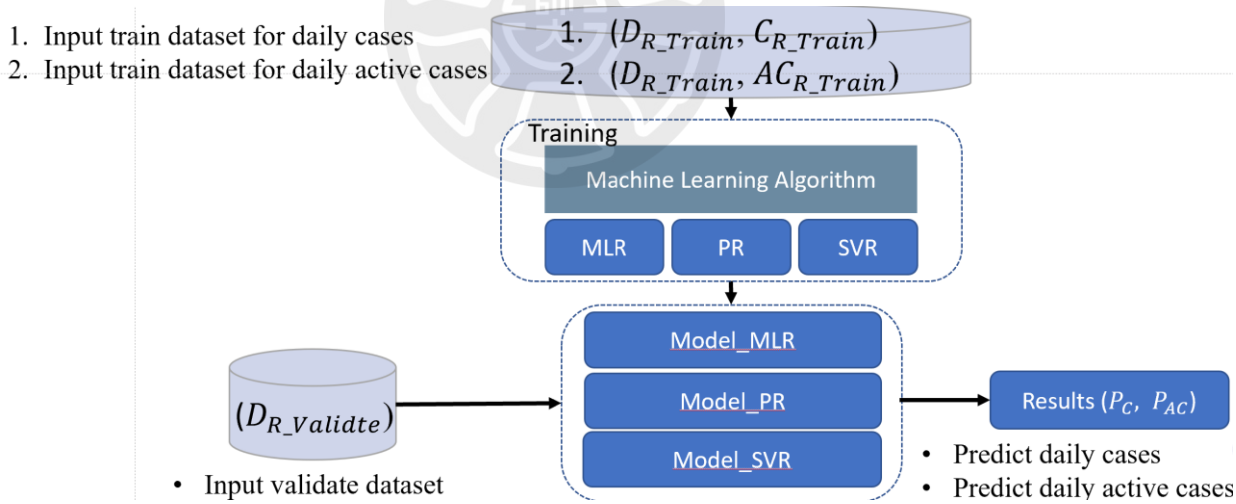

圖 20. 機器學習模型示意圖

以上的模型訓練方式為，將上章節所切割好的測試集輸入，透過三種回歸方式訓練出模型。並將驗證集$D_{R_Validate}$ 輸入進行驗證，因為會與用三種方式因此兩個預測目標 (P_C, P_{AC}) 各會有三個不同模型的預測結果。

3.2.2 LSTM 模型

由於資料間可能存在著時間相關性，所以使用最能利用時間序列特性的 LSTM 來補足這方面的特徵。為了使資料的滯後影響維持所以選擇 many to one 的 LSTM 模型，從上一小節可以看到 LSTM 所使用的訓練資料為三維矩陣的資料集。輸入從一個改為多個但輸出維持一個不變。而在本研究中每次輸入大小為$X \in R^{14*9}$的資料，表示其有 14 個時間單位，每個時間單位有 9 個特徵的進行訓練，輸出值會是每次以 14 天特徵值所預測出的目標人數，全部完成後預測結果會與訓練集的長度相同。

從圖 21我們可以看到輸入為 D，n 為資料長度，第一次輸入的資料為訓練集資料$(D_{14_Tarin}, C_{14_Train})$, $(D_{14_Train}, AC_{14_Train})$ ，進入到LSTM Cell 裡計算，將結果輸入至h_n 隱藏狀態最終，訓練出兩個不同預測目標的模型。模型建好後放入 $D_{14_Validate}$ 進行驗證，輸出我們預測目標(P_C, P_{AC})。

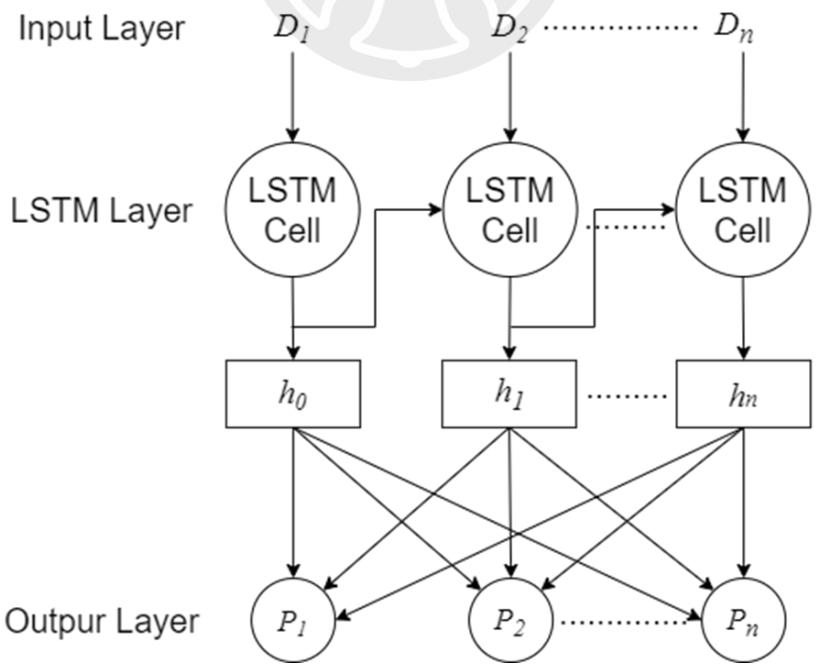

圖 21. LSTM 模型圖

3.2.3 模型評估

獲取到預測結果後我們要針對所使用的四種預測方式共八個模型進行評估，研究所使用的評估方法為均方誤差 (Mean Square Error, MSE) 以及決定係數(R-square)，兩個方式的公式如下:

n為資料長度，$C_{Validate_i}$ 為第 i 個驗證資料，C_{Pred_i} 則是第 i 個預測資料 (P_C, P_{AC})，$\overline{C_{Validate}}$ 為驗證集平均值

$$Mean\ Square\ Error = \frac{1}{n}\sum_{i=1}^{n}(C_{Validate_i} - C_{Pred_i})^2 \tag{16}$$

$$R - square = 1 - \frac{SS_{res}}{SS_{tot}} \tag{17}$$

$$SS_{tot} = \sum_i (C_{Validate_i} - \overline{C_{Validate}})^2 \tag{18}$$

$$SS_{res} = \sum_i \left(C_{Validate_i} - C_{Pred_i}\right)^2 \tag{19}$$

$$\overline{C_{Validate}} = \frac{1}{n}\sum_{i=1}^{n} C_{Validate_i} \tag{20}$$

MSE 會計算實際值與預測值的差異並進行平均，差異所使用的是計算距離所以會進行平方，但也因為是使用平方所以會受離群值的影響。計算出的結果就為誤差的平均。可以獲得模型的精準度。而 R-square 分母會先計算實際值的平均變化，而分子一樣是誤差平方，但因為是用百分比顯示所以離群值的影響會較 MSE 小。最後可以獲得實際值與預測誤差平方之間的變化量，也就是誤差平均的變化量。可以用於評估模性是否具有解釋性。

將實際的四個州預測結果帶入後我們可以獲得表 4

State	PR				MLR				SVR				LSTM			
	MSE		R square		MSE		R square		MSE		R square		MSE		R square	
	P_C	P_{AC}	P_C	P_{AC}	P_C	P_{AC}	P_C	P_{AC}	P_C	P_{AC}	P_C	P_{AC}	P_C	P_{AC}	P_C	P_{AC}
California	0.00324	0.0047	0.886	0.65451	0.01378	0.01239	0.516	0.08955	0.00483	0.00583	0.83037	0.49952	0.00676	0.00597	0.73563	0.57083
Florida	0.01457	0.0058	0.65139	0.33889	0.01838	0.00713	0.56019	0.18664	0.0101	0.0101	0.75838	0.75838	0.01736	0.00271	0.53151	0.24857
New Jersey	0.00393	0.00502	0.62787	0.46549	0.00679	0.0071	0.35721	0.24467	0.00621	0.00554	0.41206	0.4109	0.00715	0.0073	0.00982	-0.09045
New York	0.00238	0.00355	0.75274	0.4912	0.00555	0.00576	0.42343	0.17532	0.0042	0.00466	0.56441	0.33204	0.01094	0.00814	0.57357	0.42024

表 4. 模型評估結果表

我們的評估結果 MSE 要盡量小而 R square 則是要越接近 1 表示效果越好，結果可以從表 4 發現，對四個州而言 Polynomial Regression (PR) 以及SVR 的效果會是最好的，因此研究會將這兩個模型的預測結果帶入後面的排隊理論中，進行更進一步的分析。

第三節 排隊理論及風險分析

3.3.1 排隊理論

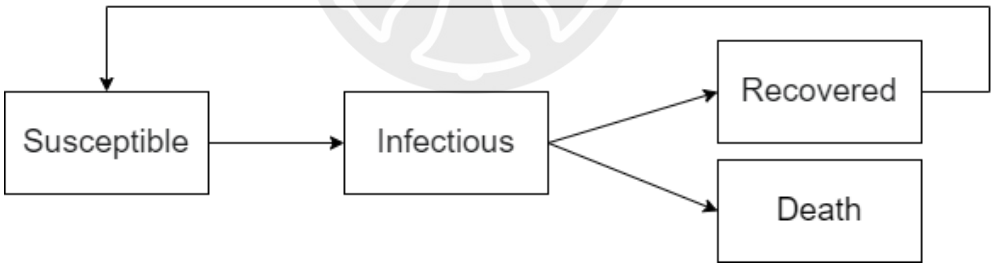

圖 22. COVID-19 狀態轉換圖

使用排隊理論前我們需要先了解 COVID-19 的狀態轉換，依照圖22可以看到未受感染的大眾會處於易感染狀態 (Susceptible)，而受到感染後會進入到感染狀態 (Infectious)，再受到感染狀態後會有兩種狀態變化，一個為康復 (Recovered) 另一個為死亡 (Death)，而康復狀態最後會重回易感染狀態 (Susceptible)，死亡狀態 (Death) 則會離開人群。

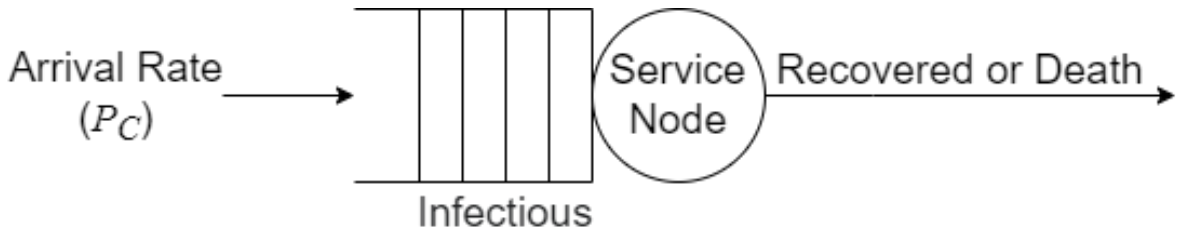

圖 23. 排隊理論模型示意圖

研究會將上述的狀態轉換帶入排隊理論模型，研究所使用分析爆發風險的排隊理論選擇最簡單的M/ M/ 1 模型，只有單一服務台 (server)、等候隊伍長度的無限制以及可加入隊伍的人數以同樣無限制。根據圖 23 可以看到，到達率 (arrive rate) 可以視為從易感染狀態轉變為受感染的狀態，進而進入到 Infectious 隊列中，也就是上一節透過空氣汙染物、人口流動以及疫苗施打率所預測出的每日確診人數 (P_C)。研究的預測目標 (P_C)，而 Service Node 裡則是在等待從受感染狀態轉換成康復狀態或死亡狀態，一旦狀態轉換完成則會離開隊列。

$$L = \lambda W \tag{12}$$

我們將上述的排隊理論套用進 Little's law 公式(12)的話，L為在一段時間內的活動病例數(P_{AC})，研究將一段時間定義為 60 天也就是兩個月，λ 為每日的預測確診人數 (P_C)，W為研究想算出的從確診狀態至死亡或康復狀態所需要花費的時間。

我們所擁有 P_C 以及 P_{AC} 這兩個數值，進而能計算出 W (Service Time) 也就是確診者平均所花費的康復時間 (Average Recovery Time)。康復時間在研究中可以將其視為一個州的醫療量能使用程度，當一位確診病患要花費較長天數才能恢復時，我們將會視為州的醫療量能較吃緊或正在面對疫情爆發的時期。但多少天數可以視為量能吃緊，這點還並未有確切的數據可以告訴我們，因此研究會使用計算出的 Average Recovery Time 進行更進一步的分析。

3.3.2 風險分析

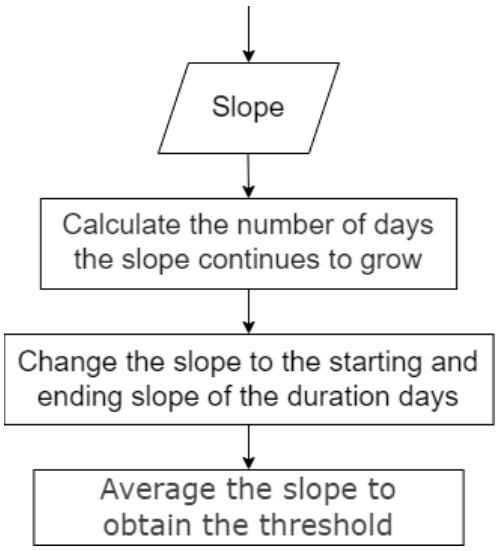

圖 24. 風險分析流程圖

在風險分析的部分本研究會利用從 Our Wolrd in Data[84]獲取世界各國的每日確診人數。透過圖 24的方法以驗證方法的可行性。

研究會使用英國的每日確診人數數據作為示範:

研究將挑選兩個上升與下降都較明顯的時間段來觀察風險分析方法其有效性。在圖 25 中可以看到紅底的時間點 (2020/11/26 ~2021/01/25) 為人數增加較快速的時間點，而黃底的時間點 (2021/05/25 ~ 2021/07/24) 為人數增加較緩慢的時間點。研究將會根據這兩個時間點做更進一步的觀察。

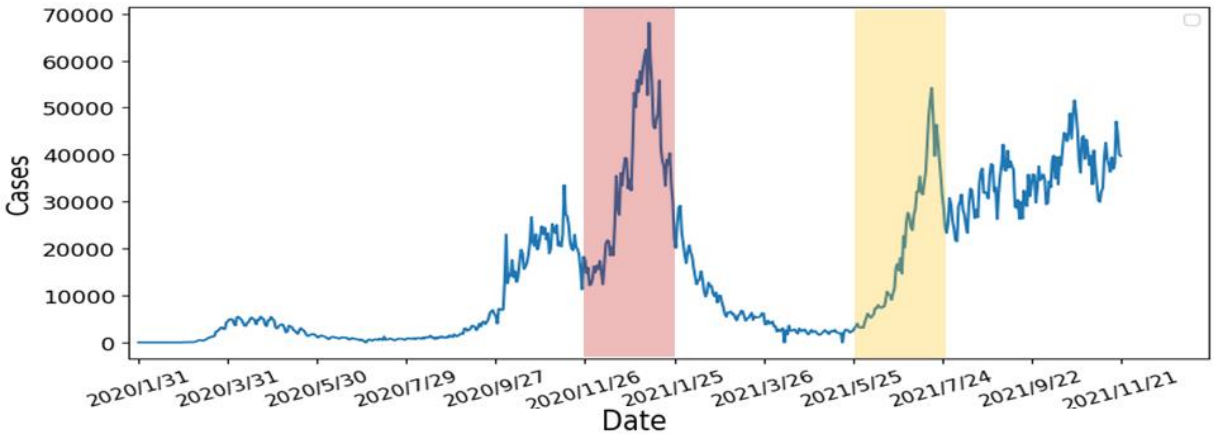

圖 25. 英國2020/01/31~ 2021/11/30 每日確診人數圖

- 風險分析第一步 (Slope)

  將這些確診人數資料以 7 天為單位進行斜率計算。根據其他研究[85][86]指出有時候民眾會因為處於節假日而無法進行篩檢，所以會有急速下降後又急速上升的情形，以及美國的篩檢政策也有所不同，後期開放快篩後確診人數會大幅上升，但經歷一段時間後民眾的篩檢意願又會逐漸降低。因此研究將資料進行 7 天的平均，來避免掉較高的起伏來獲取更平穩的斜率。但本研究目標為提供參考數據，可能無法完全避免掉這種情形。計算後的斜率結果如圖 26

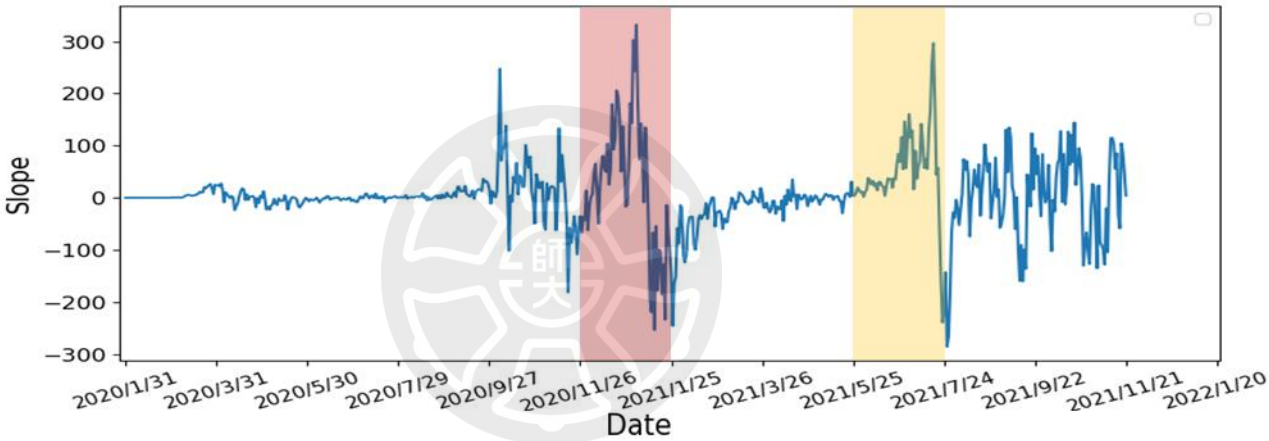

圖 26. 英國2020/01/31~ 2021/11/30 每日確診人數斜率圖

可以透過圖 26 發現紅底時間點的斜率確實能看到快速上升以及下降的斜率，而與黃底時間點相比增長的快慢也能觀察得出來。根據圖 26 的結果可以看到斜率確實能表現出確診人數的升降。

- 風險分析第二步 (Calculate the number of days the slope continues to grow)

  為計算其斜率持續增長的連續天數，研究所使用計算方式為當日斜率大於前一日的斜率時即視為斜率有持續增長的跡象，會計算這種情形的持續天數。當連續天數較長時，可視為確診人數持續多天非線性增長的情形。 結果如圖 27。

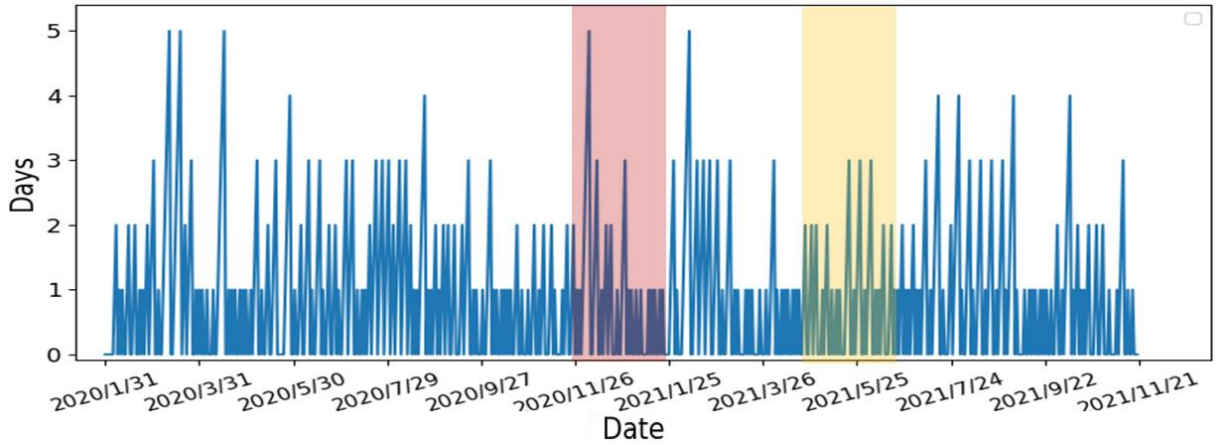

圖 27. 英國2020/01/31~ 2021/11/30 確診人數斜率持續增長天數圖

透過圖 27 可以發現增長較快速紅底時間點的持續天數有達到持續 5 天的情形，而黃底時間點因為增長速度較慢所以持續天數只有 3 天，能確實底叫出兩個時間的差別。與圖 25 結合來看這持續 5 天的斜率增長確實與確診人數增長相關。

- 風險分析第三步 (Change the slope to the starting and ending slope of the duration days)

  重新計算出持續天數起迄點的斜率，根據圖 27 計算的持續天數，研究將斜率更換為持續天數起訖點的斜率，以獲取持續增長天數時間段的斜率，這個斜率會是研究所認為高風險的特徵斜率。結果如圖 28 所示。

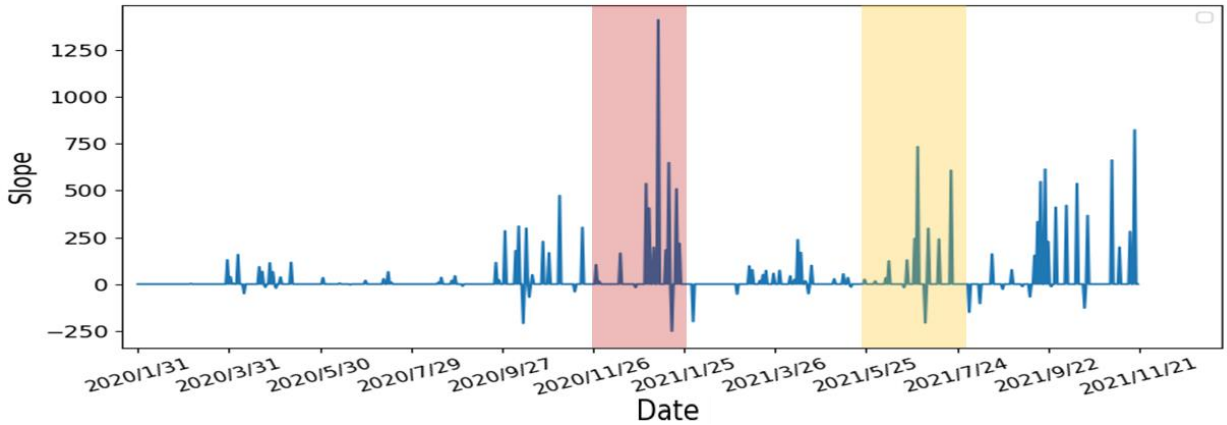

圖 28. 英國2020/01/31~ 2021/11/30 持續增長天數斜率圖

透過圖 28 可以發現與原始的 7 天斜率圖 26 相比，圖 28 的斜率圖

能將風險分析所需要的持續增長斜率特徵抓取出來，以紅色時間點與黃色時間點進行比較能看出紅色時間點的增長確實比黃色時間點來的更為快速且更為密集。除此之外其他的增長高峰也都能抓取出。

- 風險分析第四步(Average the slope to obtain the threshold)

  統計出圖 28上的斜率平均數，將這個斜率視為英國確診人數的高風險時段的閥值(threshold)。並將 threshold 帶回圖 25及圖 26。其結果如圖 29。

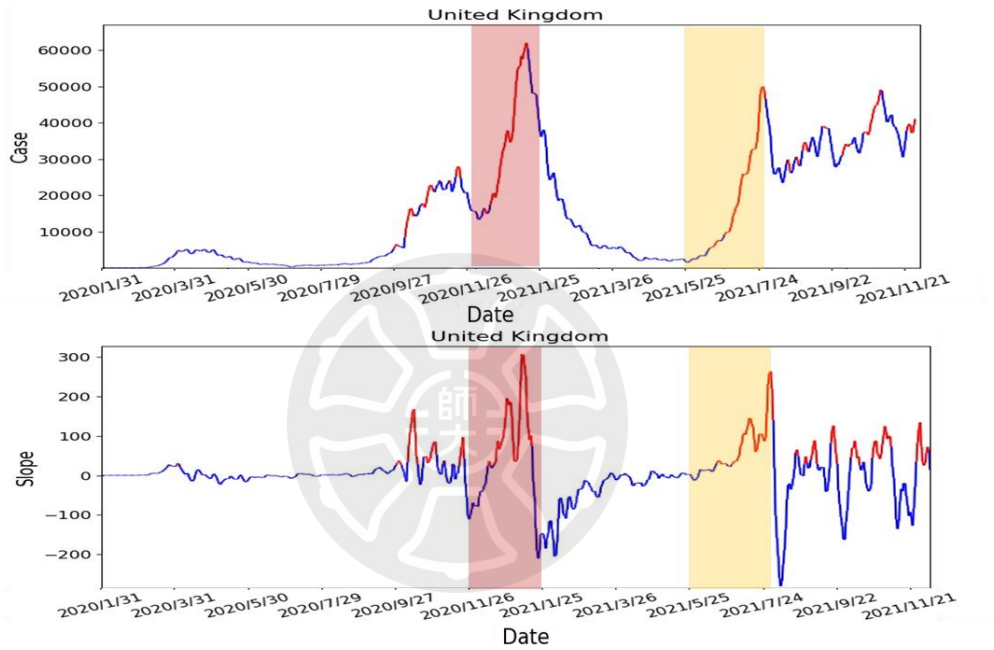

圖 29. 英國確診人數風險分析結果圖

風險分析會將超出 threshold 的斜率標記進行高風險警告，標記方式使用紅色線段表示為高風險時段，而藍色線段則表示低風險時段。透過圖 29圖 29，可以看到風險分析結果確實有將高風險時段標記出來。這個結果表示研究所使用的風險分析方法是有效的，因此研究會將上小節使用排隊理論所預測出的平均康復時間 (W) 帶入風險分析方法進行風險預測，並使用預測結果來評估各州的防疫措施是否妥當。

第四章 實驗結果分析

第一節　模型分析

根據 3.2.3 節模型評估的結果將針對 Polynomial Regression (PR)以及 SVR 這兩個模型進行更進一步的分析。研究希望能夠提供防疫措施評估的參考，但所提供的參考數據是事前提醒又或是事後檢討這點會根據模型能預測的準確度而有所影響。因此研究將會針對預測當日(Day 0)，提前 1, 3, 5, 7 天(Day 1, 3, 5, 7) 的結果進行比較，結果如圖 30 ~ 圖 33。

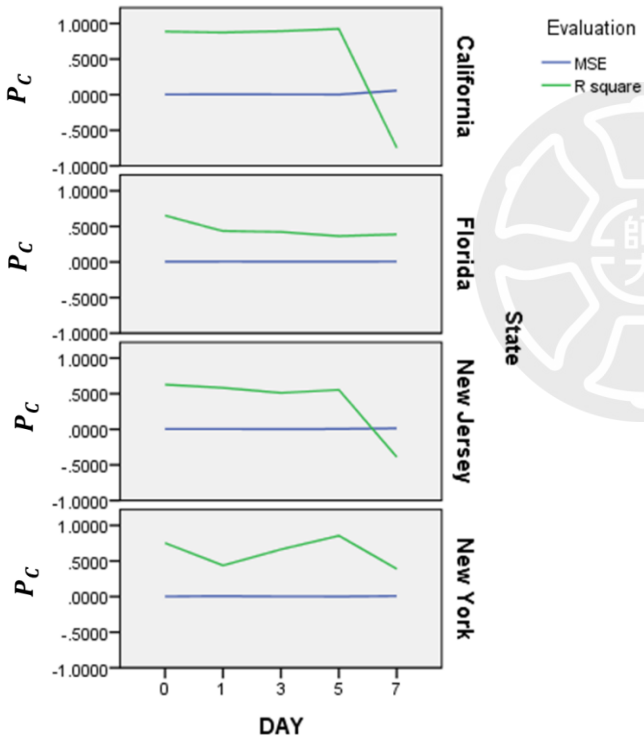

圖 30 PR 預測每日確診人數分析結果圖

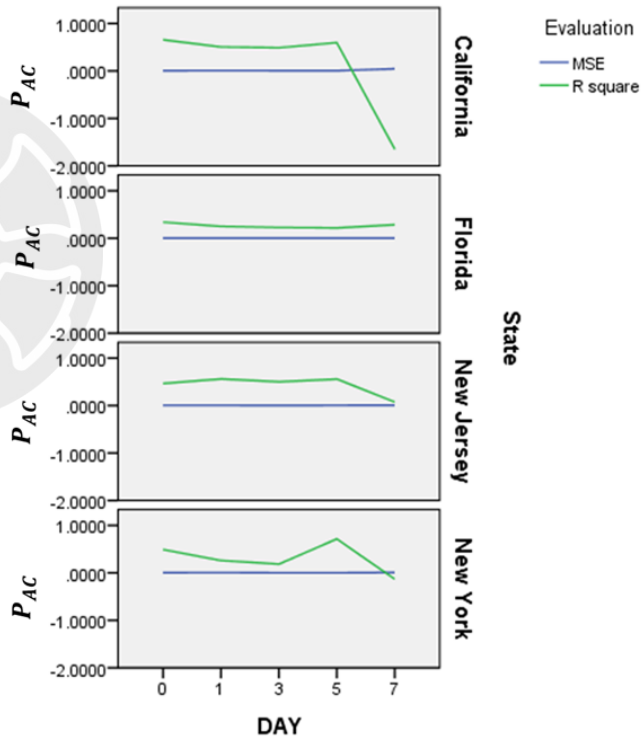

圖 31 PR 預測每日活動病例分析結果圖

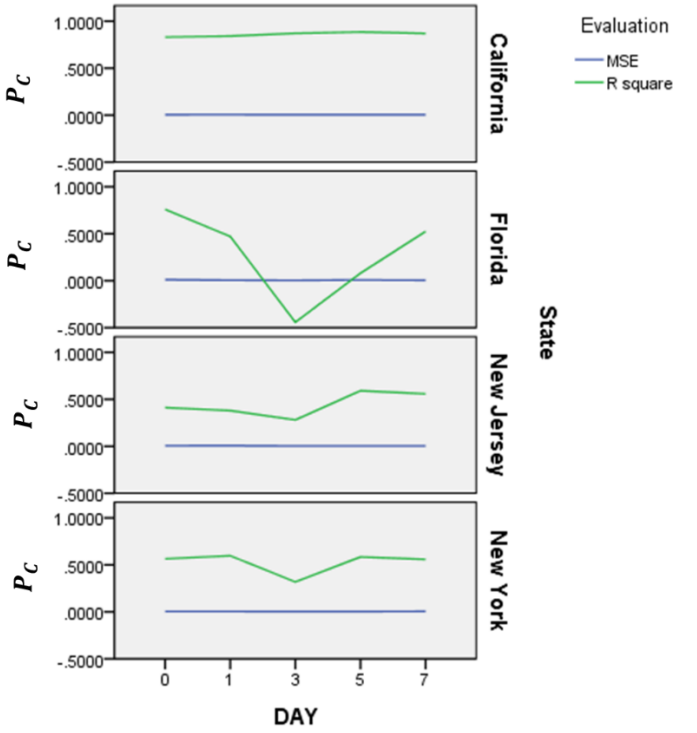

圖 32 SVR預測每日確診人數分析結果圖

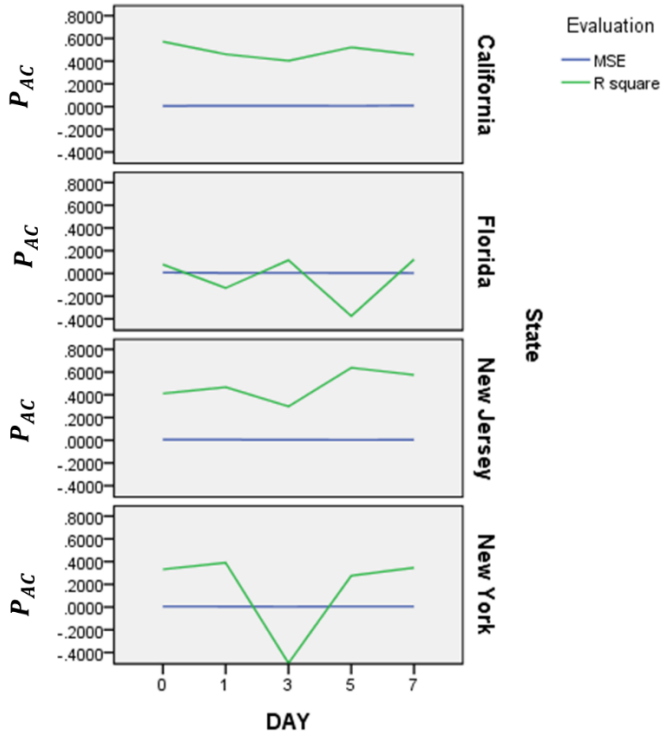

圖 33 SVR 預測每日活動病例分析結果圖

可以從圖 30 以及圖 31 的 PR 預測結果中看到，MSE 的結果相差並不明顯，但 R Square在提前 7 天 (Day 7) 預測時其數值相較其他有大幅下降的趨勢，顯示其預測結果並不理想。而綜觀各天數可以發現預測當日 (Day 0) 的結果在四個州有的相較較好的結果。而圖 32 以及圖 33 的 SVR 預測結果與 PR 相同在 MSE 方面其結果相差並不明顯，但 R Square 在提前 3 天 (Day 3) 有較明顯的下降情形，並在提前 5, 7 天 (Day 5, 7) 有重新上升的趨勢，雖然有上升但大多數都並未超過當日 (Day 0)。從 PR 以及 SVR 的結果來看可以發現當日 (Day 0) 的結果還是較佳的，所以研究還是會已當日 (Day 0) 為主要的分析目標。

第二節　案例探討

如上節所示研究將以 Polynomial Regression (PR)以及 SVR 這兩個模型所預測當日 (Day 0) 的結果帶入排隊理論，並將排隊理論所預測的平均康復時間帶入風險分析進行高風險時段的預測。預測結果會與約翰·霍普金斯大學[87]以及Wikipedia [88] 所整理的加利福尼亞州的防疫措施進行比較，並以每日活動病例作為輔助條件來觀察預測的高風險時段是否有其準確性。

研究首先會針對限制措施進行 PR、SVR 預測結果的評估，限制措施研究評估方式為若措施發布時間點為風險分析所預測的高風險時段，並使用施行後每日活動病例進行驗證。

首先會針對美國的防疫政策以及名詞進行介紹:

- 居家令: 該命令要求居民除了工作之外不得離開家中。適用於抑制或減輕流行病的發展
- 四階段解封:
  - 第一階段: 居家令時期
  - 第二階段: 逐步開放一些低風險工作場所與適應，零售(路邊取貨)、製造業、辦公室(無法遠程辦公時)
  - 第三階段: 開放更高風險的環境和限制聚會，如個人護理(美髮沙龍、健身房)，娛樂場所(電影院)，需要面對面的宗教儀式(教堂、婚禮)
  - 第四階段: 重新開放風險最高的工作場所(音樂廳、會議中心以及有現場觀眾的運動賽事)
- 宵禁: 非必要工作、活動和集會在晚上 10 點至凌晨 5 點之間停止
- 重新開放: 取消了大多數 COVID-19 限制，但企業仍然可以要

求員工配戴口罩

- 限制措施: 如居家令、宵禁，是限制市民的行為的措施
- 放寬措施: 將限制解開，對市民的行為鬆綁

4.1.1 加利福尼亞州

如圖 34所示，研究會選擇幾個較有代表性的政策進行與我們的風險分析結果進行評估。

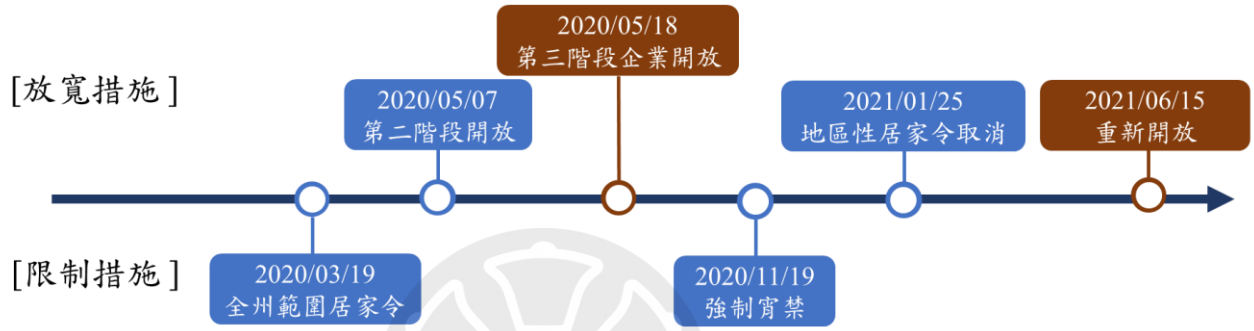

圖 34. 加利福尼亞州防疫措施時間線圖

評估圖會以紅色點線表示所預測的高風險時段並以紅色方框所框出此時段，藍色點線表示為低風險的時段，以而藍色方框所框出的則為錯誤預測的高風險時段。綠色點線則表示每日活動病例。

接下來會先針對放寬措施來進行評估，評估方式為，放寬措施發布時是否為風險分析所預測的高風險時段以及措施前後期是否有高風險時段出現，措施是否應提前或延後，並使用每日活動病例進行驗證。

- 2020/ 03/ 19 全州範圍居家令

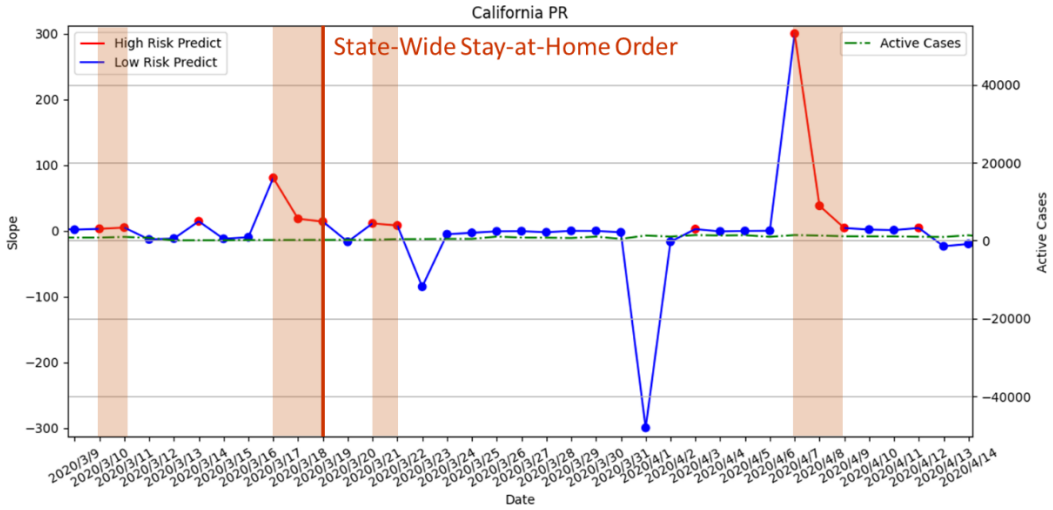

圖 35. 加利福尼亞州 2020/ 03 PR 風險分析圖

透過圖 35可以看到有使用紅色方框圈出來的部分為 PR 經過排隊理論以及風險分析所得出的高風險時段，而措施發布時間使用紅色直線表示，可以發現第居家令措施所發布的時間點有 PR 所預測的高風險時段存在，這表示說這個時段確實為疫情的高風險時段，所以 PR 的風險分析會認為說這個措施在這個時間點發布是合適的。並且可以看到綠色的活動病例線段，發現在措施實施後有逐漸升高的趨勢，且在措施發布後還有一個高風險時段，因此 PR 的風險分析會認為說措施可能需要更強的約束力。

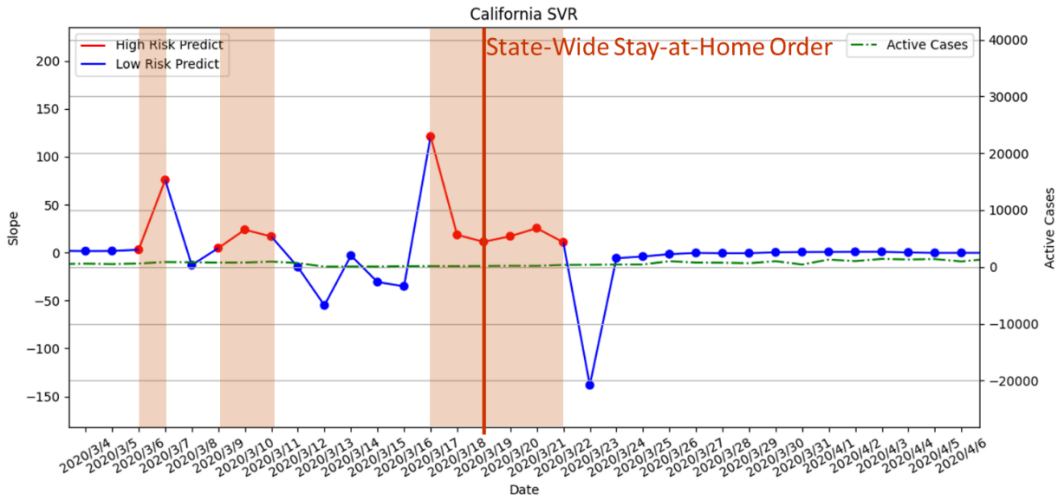

圖 36.加利福尼亞州 2020/ 03 SVR 風險分析圖

透過圖 36 可以看到與PR結果相同都有在這個時段確實有預測到為高風險時段，且綠色的活動病例線段發現其有逐漸升高的傾向。所以使用的兩種風險分析皆認為說這個措施在這個時間點發布是合適的。但後面可能為高風險時段這個部分 SVR 並未預測到。因此PR 的風險分析效果可能較 SVR 來的好。

- 2020/ 11/ 19 強制宵禁

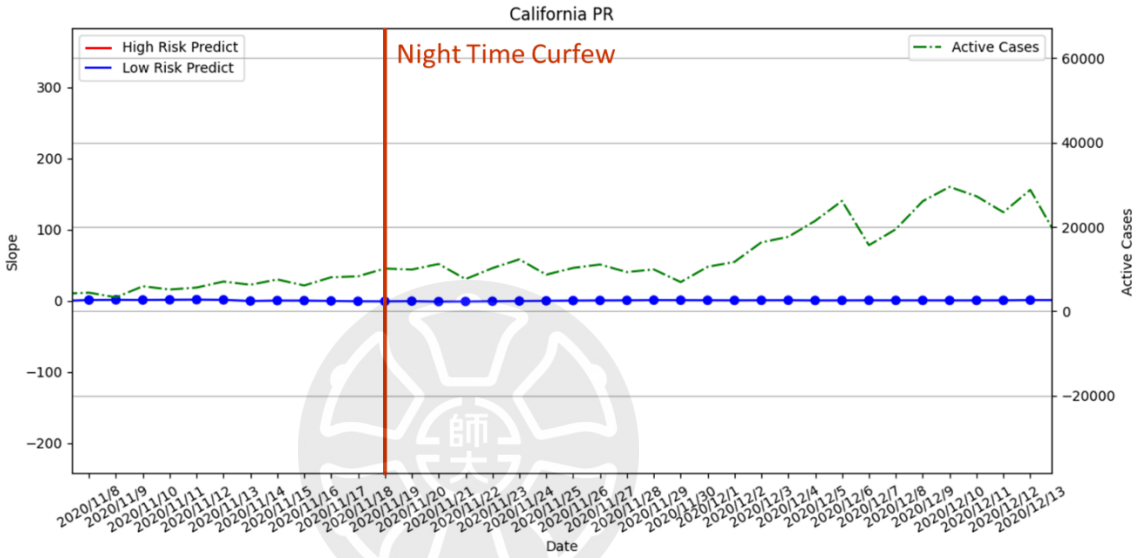

圖 37. 加利福尼亞州 2020/ 11 PR 風險分析圖

對於強制宵禁措施，透過圖 37 PR的風險分析並未預測到高風險時段，但觀察綠色的每日活動病例可以發現 11 月的活動病例是處於逐漸升高的趨勢，表示這個時段可能存在高風險，但 PR 的風險分析並未預測到。

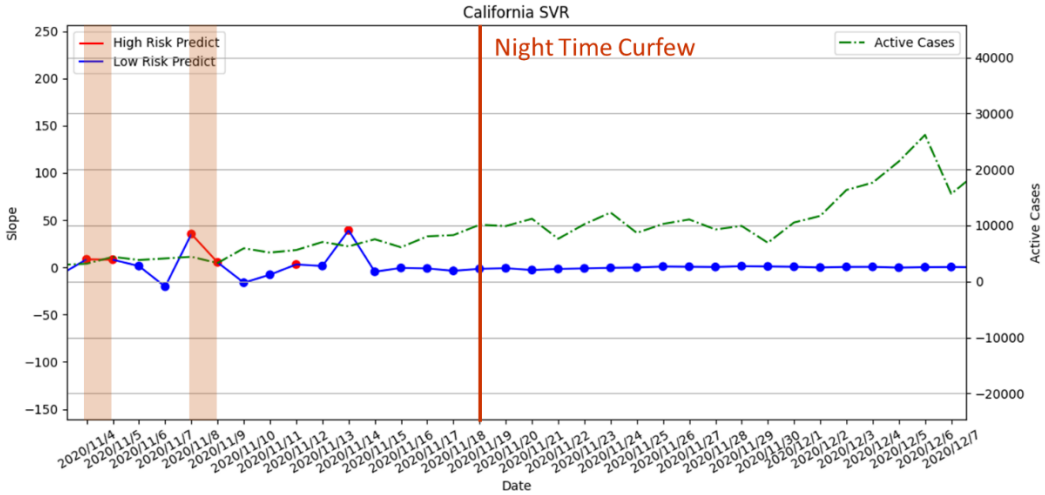

圖 38. 加利福尼亞州 2020/ 11 SVR 風險分析圖

透過圖 38 可以看到，SVR的風險分析雖然並未在措施發布時有預測到高風險時段但在措施發布前就有兩次的高風險時段預測。因此評估結果為應該要將措施時間往前，更早的發布此措施，且從每日活動病例來看是有持續上升的趨勢。

- 2020/ 05/ 07 第二階段開放以及 2020/ 05/ 18 第三階段開放

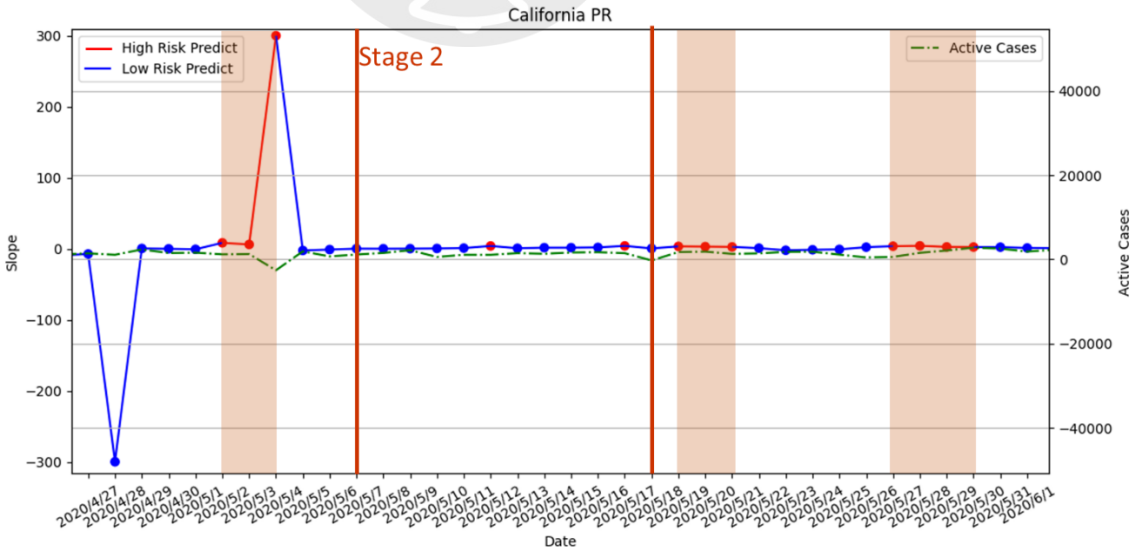

圖 39. 加利福尼亞州 2020/ 05 PR 風險分析圖

針對開放措施由於這兩個措施時間接近因此會一同進行評估，首先是第二階段開放，透過圖 39可以看到這個時間點並未有高風險時段預測，

且至第三階段開放前每日活動病例也未有明顯增長。所以這第二階段開放措施可以認為是在合適時間點發布的措施。但是對於第三階段開放，透過圖 39 可以看到在措施之後有兩次的高風險時段，因此 PR 評估的結果為第三階段開放措施可能是比較不合適的，有過早開放的情形發生，且活動病例也確實開始有上升的趨勢。以此驗證評估結果。

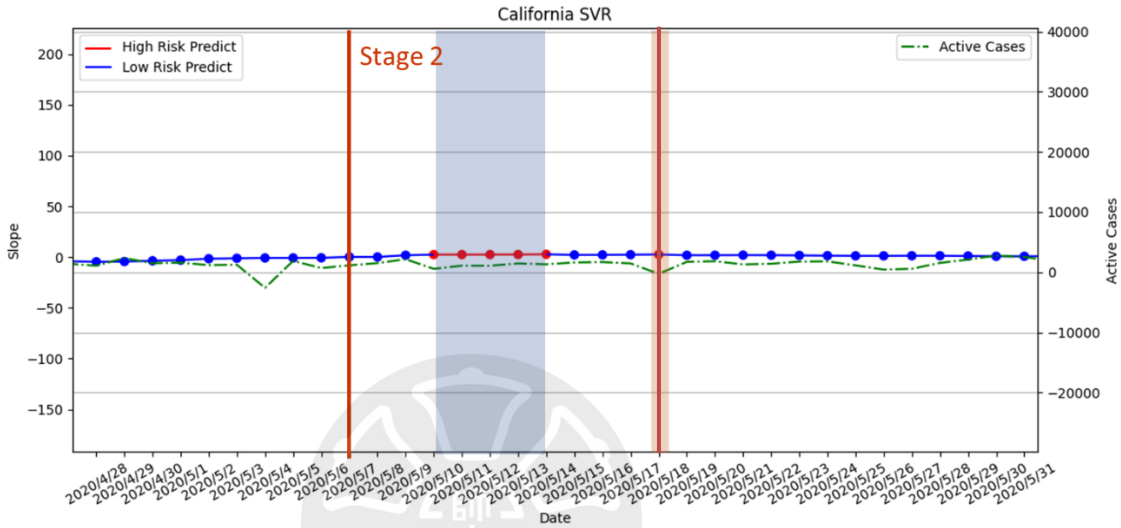

圖 40. 加利福尼亞州 2020/ 05 SVR 風險分析圖

透過圖 40 可以看到在第二階段開放時 SVR 風險分析並未預測到高風險時段，反而是在措施發布後有高風險時段的預測，評估結果為過早開放。但根據每日活動病例的數據是為下降的趨勢可以發現這個高風險預測是錯誤預測。而對於第三階段的 SVR 有給出高風險預測，但只存在於一天，較難看出是否為合適措施。且措施後的每日活動病例是有上升情形，SVR 也並未預測出高風險時段。與 PR 相比 SVR 的準確度較為不佳。

- 2021/ 01/ 25 地區居家令取消

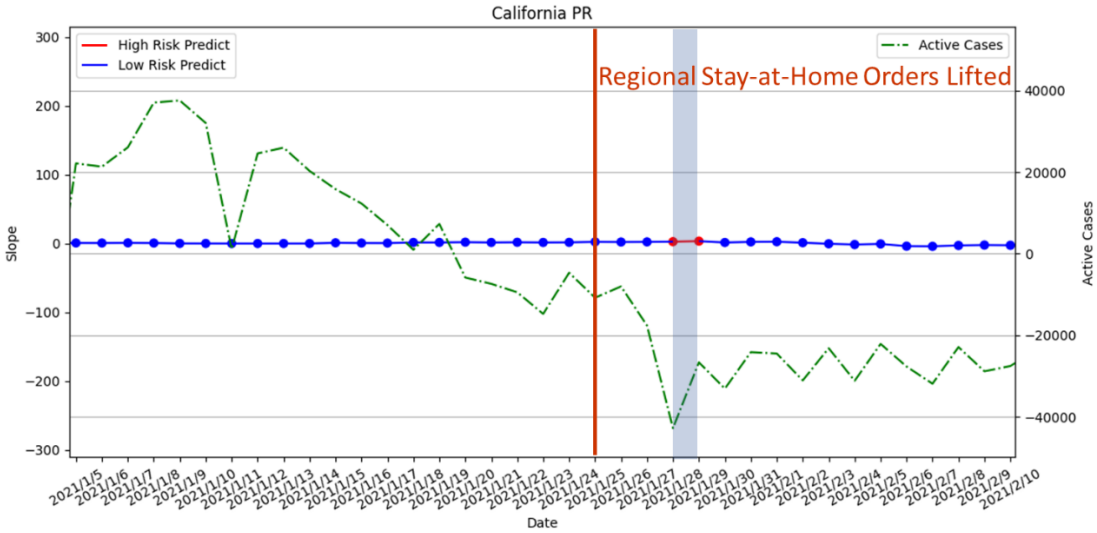

圖 41.加利福尼亞州 2021/ 01 PR 風險分析圖

可以透過圖 41看到在措施發布時並未有高風險時段，但後期存在高風險時段，因此評估為過早開放。每日活動病例卻有明顯的下降趨勢。因此這個高風險時段會被認為為錯誤預測，但在仔細觀察所標記的高風險時段確實是有明顯的活動病例上升，但這時的活動病例與前一周相比是明顯的下降。這是研究使用斜率來進行運算所產生的問題。

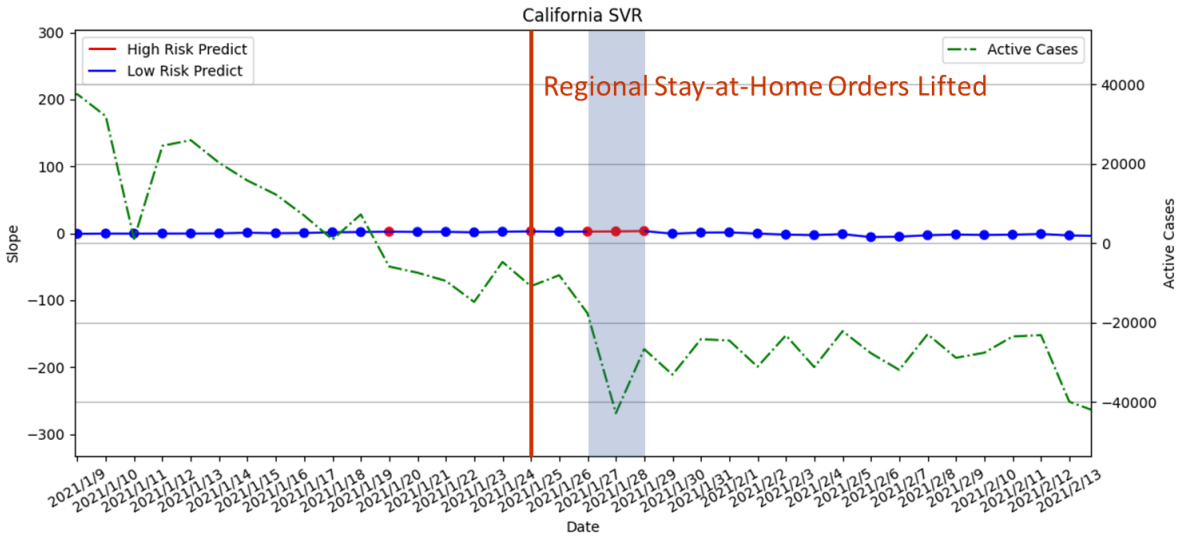

圖 42.加利福尼亞州 2021/ 01 SVR 風險分析圖

透過圖 42可以看到，在措施發布時有一點的預測高風險時段，且後期出現高風險時段，因此評估為過早開放，但根據每日活動病例能看出此時是處於下降的趨勢，因此 SVR 的預測風險為預測錯誤的情形。這與 PR 的預測風險結果類似但與其相比 SVR 結果存在較多錯誤，時間點的表記也較不準確。措施後的高風險時段雖然是有上升但時段多預測一天。

- 2021/ 06/ 15 重新開放

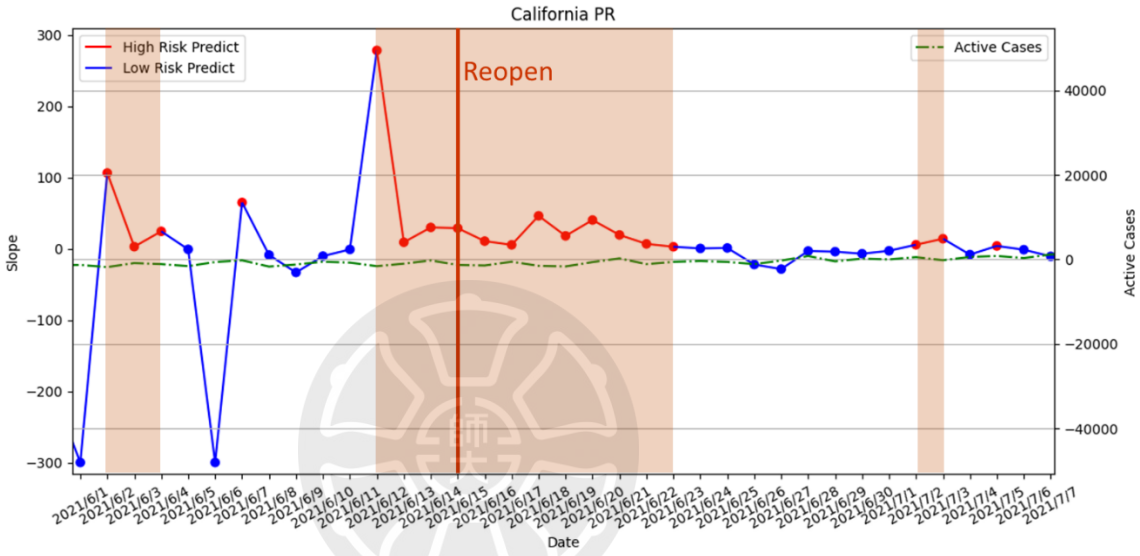

圖 43.加利福尼亞州 2021/ 06 PR 風險分析圖

透過圖 43 可以看到措施發布時間點為高風險時段，且是一個長時間的高風險時段。因此評估為過早開放，從措施發布後的每日活動病例也有上升的趨勢這點驗證。

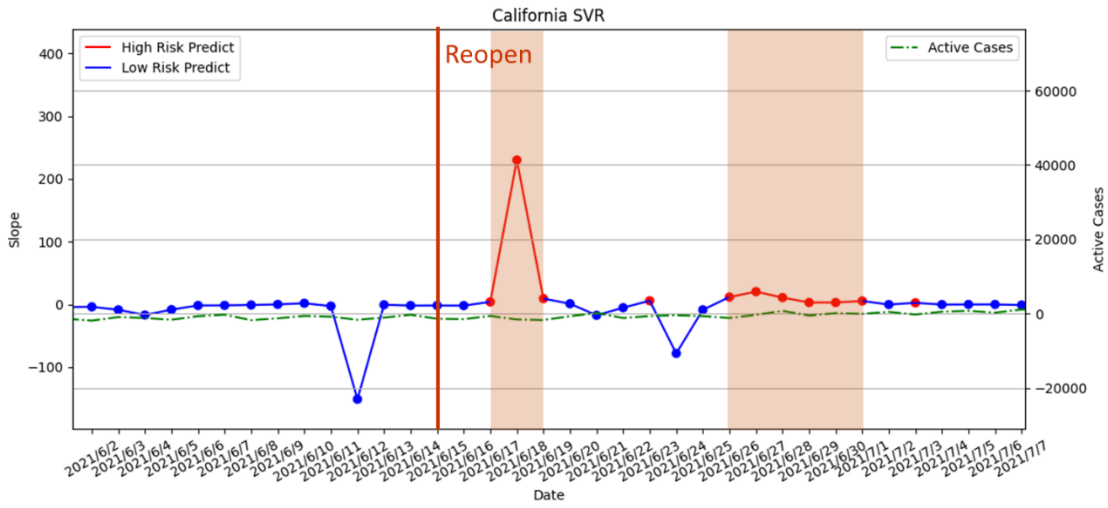

圖 44. 加利福尼亞州 2021/ 06 SVR 風險分析圖

透過圖 44 的 SVR 風險預測結果發現，雖然在措施發布當日並未有高風險時段預測，但後期存在高風險預測，也有長時間的高風險時段預測，因此風險評估結果認為此措施發布時間過早。而每日活動病例顯示有上升的趨勢，以此驗證。

以上將所選擇的措施評估結束之後，可以發現到 PR 所產生的預測錯誤會比 SVR 來的少且時間點的標記也相對較精準，因此研究將四個州的 PR 預測結果放在一起進行分析，希望能找出其時間上的共通性，結果如圖 45 所示。

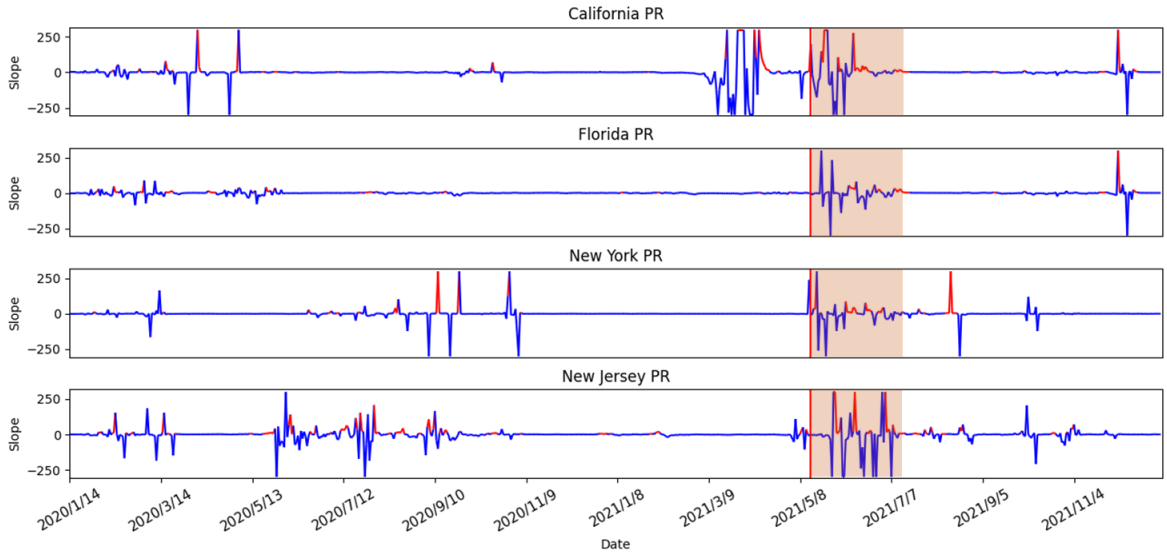

圖 45. 加利福尼亞州、佛羅里達州、紐澤西州以及紐約州的 PR 風險分析圖

可以從圖 45 發現到在 2021/05/08~ 2021/07/07 期間四個州都有較為明顯的起伏且都存在較多的高風險時段，因此針對這期間研究進行更進一步分析，發現到美國在 2021/04/12 時在加利福尼亞發現了首 6 例的 Delta 變異病毒確診案例[89]，但在 4 月中至 5 月中研究發現到所選擇的四個州並未有明顯的相關影響。在 2021/05/13 時美國政府發部一項措施[89]，其措施為完整接種過兩劑疫苗的國民可以不用配戴口罩，而在此之後就為研究在圖 45 所框出的時間區間，推測可能因為此措施而導致後面疫情再度爆發。另外美國CDC 有數據指出在 7 月 7 日以前 Delta 病毒株已經成為美國的主要感染病毒[89]，佔美國新感染 COVID-19 病例的 51.7%。由此可以看到防疫措施對於疫情的影響性是非常明顯的也證實風險分析方法有其參考價值。

第五章 結論與未來展望

本研究先對美國的人口流動、空氣汙染物及疫苗接種率與 COVID-19 進行相關性分析並使用分析結果利用回歸方式預測 COVID-19 的每日確診人數以及每日活動病例。並將預測結果帶入排隊理論來預測平均康復時間，也可視為醫療量能使用度。預測結果再帶入研究所提出的風險分析方法進而找出高風險時段來評估防疫措施。

在相關性分析中，先針對美國四個州疫情前後的空氣汙染物變化進行觀察，發現到在疫情前後空氣汙染物確實存在著起伏情形。而後使用了皮爾森相關係數計算，結果發現人口流動中以自駕的相關性為最高，而在空氣汙染物中也發現有相關性存在，而疫苗接種率則是發現與疫情存在著負相關。而相關性分析完之後，則是進行 COVID-19 的每日確診人數以及每日活動病例的預測，我們發現 (Polynomial Regression, PR) 以及 (Support Vector Regression, SVR) 有著最好的預測結果，並將這個結果帶入至風險分析裡面。並得出最後的高風險時段預測結果。將這個結果與防疫措施進行比較評估，發現我們的高風險時段預測確實能評估出措施是否適合發布於這個時段。其中又以 PR 的效果最佳，能準確地預測出高風險時段並提供措施評估。透過 PR 風險分析，我們能更加瞭解疫情的爆發風險存在於何時，並且也能看出防疫措施對於風險的影響。最後利用預測結果提供相關機構的措施評估的參考數據。期望我們將會對於疫情有更好的控制。

在未來研究希望能將風險進行量化，能更好的評估即將遇到的風險，以及在預測美國四州每日確診人數以及每日活動病例時可以發現各個地區其特徵值的相關性都不同，在未來預測時可以針對這些相關性進行參數的權重調整，但僅限於研究所挑選的美國四州，因不同地區其人口結構或是移動趨勢等組成因素不同。如果未來想要將此風險分析套用至其他地區或國家時需要在進行預測前了解其地區的人口結構特性來進行修改。而如果

想使用相同的風險分析方式套用至其他地區或國家時，可以先找有著相同防疫方針的國家且人口結構類似會有著較好的結果。此模型雖然在美國四州預測上有著良好的效果，但並不表示其適合用於任何地區，COVID-19的預測存在許多變因，想要一個模型就套用至全部地區是個巨大的挑戰。

第六章 參考文獻

[1] “World Health Organization. Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations.” 2020 Available from: https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations. [(accessed on 22 May 2020)]

[2] Guarnieri M., Balmes J.R. “Outdoor air pollution and asthma.” The Lancet 2014, 383(9928), 1581-1592.

[3] Goldizen F.C., Sly P.D., Knibbs L.D. “Respiratory effects of air pollution on children.” Pediatric pulmonology 2016, 51(1), 94-108.

[4] Jiang X. Q., Mei X. D., Feng D. “Air pollution and chronic airway diseases: what should people know and do?.” Journal of thoracic disease 2016, 8(1), E31–E40. https://doi.org/10.3978/j.issn.2072-1439.2015.11.50

[5] Setti L., Passarini F., de Gennaro G, et al. “Evaluation of the Potential Relationship between Particulate Matter (PM) Pollution and COVID-19 Infection Spread in Italy.” Available online:http://www.simaonlus.it/wpsima/wp-content/uploads/2020/03/COVID_19_position-paper_ENG.pdf. [(accessed on 22 May 2020)]

[6] Gupta A., Bherwani H., Gautam S., et al. “Air pollution aggravating COVID-19 lethality? Exploration in Asian cities using statistical models.” Environ Dev Sustain 23, 6408–6417 (2021). https://doi.org/10.1007/s10668-020-00878-9

[7] Beckerman B.S., Jerrett M., Brook, J.R, et al. “Correlation of nitrogen dioxide with other traffic pollutants near a major expressway.” Atmospheric Environment 2018, 42, 275-290.

[8] Chen ZL, Zhang Q, Lu Y, et al. “Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan, China.” Chin Med J (Engl). 2020;133(9):1044-1050. doi:10.1097/CM9.0000000000000782

[9] Chen H, Chen Y, Lian Z, et al. “Correlation between the migration scale index and the number of new confirmed coronavirus disease 2019 cases in China.” Epidemiology and Infection. 2020;148:e99. doi:10.1017/S0950268820001119

[10] Dahlberg M, Edin P.A., Grönqvist E, et al. “Effects of the COVID-19 Pandemic on Population Mobility under Mild Policies: Causal Evidence from Sweden.” 2020. [J]. arXiv preprint arXiv:2004.09087

[11] Drake TM, Docherty AB, Weiser TG, et al. “The effects of physical distancing on population mobility during the COVID-19 pandemic in the UK.” Lancet Digit Health. 2020 Aug, 2(8):e385-e387. doi: 10.1016/S2589-7500(20)30134-5.

[12]Lee H, Park SJ, Lee GR, et al. “The relationship between trends in COVID-19 prevalence and traffic levels in South Korea.” Int J Infect Dis. 2020 Jul;96:399-407. doi: 10.1016/j.ijid.2020.05.031.

[13]Lee H, Noh E, Jeon H, Nam EW. “Association between traffic inflow and COVID-19 prevalence at the provincial level in South Korea.” Int J Infect Dis. 2021 Jul;108:435-442. doi: 10.1016/j.ijid.2021.05.054

[14]Li X, Rudolph AE, Mennis J. “Association Between Population Mobility Reductions and New COVID-19 Diagnoses in the United States Along the Urban–Rural Gradient.” February–April, 2020. Prev Chronic Dis 2020;17:200241. DOI: http://dx.doi.org/10.5888/pcd17.200241external icon

[15] C A Pope, 3rdBrigham Young University, Provo, UT 84602. “Respiratory disease associated with community air pollution and a steel mill, Utah Valley.” American Journal of Public Health 79, no. 5 (May 1, 1989): pp. 623-628. https://doi.org/10.2105/AJPH.79.5.623

[16] Schwartz J. “Short term fluctuations in air pollution and hospital admissions of the elderly for respiratory disease.” Thorax. 1995 May;50(5):531-8. doi: 10.1136/thx.50.5.531.

[17] Li HH., Liu CC., Hsu TW. et al. “Upregulation of ACE2 and TMPRSS2 by particulate matter and idiopathic pulmonary fibrosis: a potential role in severe COVID-19.” *Part Fibre Toxicol* **18,** 11 (2021). https://doi.org/10.1186/s12989-021-00404-3

[18] Winterbottom CJ, Shah RJ, Patterson KC, et al. “Exposure to Ambient Particulate Matter Is Associated With Accelerated Functional Decline in Idiopathic Pulmonary Fibrosis.” Chest. 2018 May;153(5):1221-1228. doi: 10.1016/j.chest.2017.07.034.

[19] Comunian, S., Dongo, D., Milani, C., et al. “Air Pollution and Covid-19: The Role of Particulate Matter in the Spread and Increase of Covid-19's Morbidity and Mortality.” International journal of environmental research and public health 2020, 17(12), 4487. https://doi.org/10.3390/ijerph17124487

[20] Fiasca F, Minelli M, Maio D, et al. “Associations between COVID-19 Incidence Rates and the Exposure to PM2.5 and NO2: A Nationwide Observational Study in Italy.” International Journal of Environmental Research and Public Health. 2020; 17(24):9318. https://doi.org/10.3390/ijerph17249318

[21] Arias Velásquez RM, Mejía Lara JV. “Gaussian approach for probability and correlation between the number of COVID-19 cases and the air pollution in Lima.” Urban Clim. 2020 Sep;33:100664. doi: 10.1016/j.uclim.2020.100664.

[22] Bashir M. F., Ma B., Bilal, Komal, et al. “Correlation between climate indicators and COVID-19 pandemic in New York, USA.” The Science of the total environment 2020, 728, 138835. https://doi.org/10.1016/j.scitotenv.2020.138835

[23] Zoran MA, Savastru RS, Savastru DM, et al. “Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy.” Sci Total Environ. 2020 Oct 10;738:139825. doi: 10.1016/j.scitotenv.2020.139825. Epub 2020 Jun 2. PMID: 32512362; PMCID: PMC7265857.

[24] Ogen Y. “Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality.” Sci Total Environ. 2020 Jul 15;726:138605. doi: 10.1016/j.scitotenv.2020.138605.

[25] Mele M., Magazzino C., “Schneider N., et al. NO2 levels as a contributing factor to COVID-19 deaths: The first empirical estimate of threshold values.” Environmental research 2021, 194, 110663. https://doi.org/10.1016/j.envres.2020.110663

[26] Magazzino C, Mele M, Sarkodie SA. “The nexus between COVID-19 deaths, air pollution and economic growth in New York state: Evidence from Deep Machine Learning.” Journal of Environmental Management. 2021;286:112241. pmid:33667818

[27] Chen Z, Hao X, Zhang X, et al. “Have traffic restrictions improved air quality? A shock from COVID-19.” J Clean Prod. 2021 Jan 10;279:123622. doi: 10.1016/j.jclepro.2020.123622.

[28]Rainald B. “Philipp Schrauth,Population density and urban air quality,Regional Science and Urban Economics” Volume 86, 2021, 103596, ISSN 0166-0462, https://doi.org/10.1016/j.regsciurbeco.2020.103596.

[29] Yifan F, Xiaotian D, Jian H, et al. “Characteristics of urban air pollution in different regions of China between 2015 and 2019,” Building and Environment, Volume 180, 2020, 107048, ISSN 0360-1323, https://doi.org/10.1016/j.buildenv.2020.107048.

[30] Blagoiev M., Gruicin I., Ionascu M., et al "A Study on Correlation Between Air Pollution and Traffic," 2018 26th Telecommunications Forum (TELFOR), 2018, pp. 420-425, doi: 10.1109/TELFOR.2018.8612084.

[31] Azmi S.Z., Latif M.T., Ismail A.S. et al. “Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia.” Air Qual Atmos Health 3, 53–64 (2010). https://doi.org/10.1007/s11869-009-0051-1.

[32] Islam M.S., Tusher T.R., Roy S. et al. “Impacts of nationwide lockdown due to COVID-19 outbreak on air quality in Bangladesh: a spatiotemporal analysis.” Air Qual Atmos Health 14, 351–363 (2021). https://doi.org/10.1007/s11869-020-00940-5

[33] Liane Y. K. N., Rodrigo C. “Urban,COVID-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil,” Science of The Total Environment,Volume 730, 2020, 139087, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2020.139087.

[34] World Health Organization. (n.d.). Coronavirus. World Health Organization. Available from: https://www.who.int/health-topics/coronavirus#tab=tab_1. [(accessed on 23 May 2020)]

[35] 疾病介紹. 衛生福利部疾病管制署. (n.d.). Available from: https://www.cdc.gov.tw/Category/Page/vleOMKqwuEbIMgqaTeXG8A. [(accessed on 23 Mar 2020)]

[36] Wikimedia Foundation. (2022, January 26). 機器學習. Wikipedia. Available from: https://zh.wikipedia.org/zh-tw/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0. [(accessed on 23 Mar 2020)]

[37] Wikimedia Foundation. (2022, May 31). 人工神經網路. Wikipedia. Available from: https://zh.wikipedia.org/zh-tw/%E4%BA%BA%E5%B7%A5%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9

C. [(accessed on 23 Mar 2020)]

[38] Mohan S., Thirumalai C., Srivastava G. “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques,” in IEEE Access 2019, vol. 7, pp. 81542-81554, doi: 10.1109/ACCESS.2019.2923707.

[39] Moll M, Qiao D, Regan EA, et al. “Machine Learning and Prediction of All-Cause Mortality in COPD.” Chest. 2020 Sep;158(3):952-964. doi: 10.1016/j.chest.2020.02.079.

[40] Samaneh K, Yang Y, Timothy M.W., et al. “Application of machine learning techniques to tuberculosis drug resistance analysis,” Bioinformatics, Volume 35, Issue 13, 1 July 2019, Pages 2276–2282, https://doi.org/10.1093/bioinformatics/bty949

[41] Gergo P, Imre F, Amir M, et al. “COVID-19 Pandemic Prediction for Hungary; a Hybrid Machine Learning Approach.”
medRxiv .2020.05.02.20088427; doi: https://doi.org/10.1101/2020.05.02.20088427

[42] Pradhan, Manaswini, Sahu, et al. “Predict the onset of diabetes disease using Artificial Neural Network (ANN).” International Journal of Computer Science & Emerging Technologies (E-ISSN) 2011. 2. 2044-6004.

[43] Junejo A., Shen Y., Laghari A. A., et al. “Notice of Retraction: Molecular Diagnostic and Using Deep Learning Techniques for Predict Functional Recovery of Patients Treated of Cardiovascular Disease,” in IEEE Access, vol. 7, pp. 120315-120325, 2019, doi: 10.1109/ACCESS.2019.2937290.

[44] Gabriela P., Yara S.T., Hugo V.S., et al. “Assessing the impact of PM2.5 on respiratory disease using artificial neural networks,” Environmental Pollution, Volume 235,2018,Pages 394-403,ISSN 0269-7491,https://doi.org/10.1016/j.envpol.2017.12.111.

[45] K. L. Khatri and L. S. Tamil, "Early Detection of Peak Demand Days of Chronic Respiratory Diseases Emergency Department Visits Using Artificial Neural Networks," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 285-290, Jan. 2018, doi: 10.1109/JBHI.2017.2698418.

[46] Wu, Neo G., Bradley B., et al. “Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case.” 2020, https://doi.org/10.48550/arXiv.2001.08317

[47] Pal R, Sekh AA, Kar S, et al. “Neural Network Based Country Wise Risk Prediction of COVID-19.” Applied Sciences. 2020; 10(18):6448. https://doi.org/10.3390/app10186448

[48] F W Wibowo and Wihayati 2021 J. Phys.: Conf. Ser. 1803 012015

[49] Rath S, Tripathy A, Tripathy AR. “Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model.” Diabetes Metab Syndr. 2020 Sep-Oct;14(5):1467-1474. doi: 10.1016/j.dsx.2020.07.045.

[50] YeŞİlkanat CM. “Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm.” Chaos Solitons Fractals. 2020 Nov;140:110210. doi: 10.1016/j.chaos.2020.110210.

[51] A. U. Mandayam, R. A.C, S. Siddesha, et al, “Prediction of Covid-19 pandemic based on Regression,” 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2020, pp. 1-5, doi: 10.1109/ICRCICN50933.2020.9296175.

[52] Chimmula VKR, Zhang L. “Time series forecasting of COVID-19 transmission in Canada using LSTM networks.” Chaos Solitons Fractals. 2020 Jun;135:109864. doi: 10.1016/j.chaos.2020.109864. Epub 2020 May 8. PMID: 32390691; PMCID: PMC7205623.

[53] M. Gupta, R. Jain, S. Arora, et al., “Ai-enabled covid-19 outbreak analysis and prediction: indian states vs. union territories,” Computers, Materials & Continua, vol. 67, no.1, pp. 933–950, 2021.

[54] Vijander S., Ramesh C.P., Sandeep K., et al. “Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine,” Journal of Discrete Mathematical Sciences and Cryptography 2020, 23:8, 1583-1597, DOI: 10.1080/09720529.2020.1784535

[55] Ricardo Manuel Arias Velásquez, Jennifer Vanessa Mejía Lara. “Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression,” Chaos, Solitons & Fractals,Volume 136,2020,109924,ISSN 0960-0779,https://doi.org/10.1016/j.chaos.2020.109924.

[56] Chaurasia V, Pal S. “COVID-19 Pandemic: ARIMA and Regression Model-Based Worldwide Death Cases Predictions.” SN Comput Sci. 2020;1(5):288. doi: 10.1007/s42979-020-00298-6.

[57] López L, Rodó X. “A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics.” Results Phys. 2021 Feb;21:103746. doi: 10.1016/j.rinp.2020.103746.

[58] Rajan G., Gaurav P., Poonam C., et al. “SEIR and Regression Model based COVID-19 outbreak predictions in India.” medRxiv 2020.04.01.20049825; doi: https://doi.org/10.1101/2020.04.01.20049825.

[59] Heesterbeek JA. “A brief history of R0 and a recipe for its calculation.” Acta Biotheoretica. 2002, 50 (3): 189–204. doi:10.1023/A:1016599411804. hdl:1874/383700.

[60] Liu, Ying, Rocklöv, et al. “The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus.” Journal of Travel Medicine. 2021-10-01, 28 (7) [2022-03-26]. doi:10.1093/jtm/taab124

[61] Gallagher, James. “Covid: Is there a limit to how much worse variants can get?”. BBC News. 12 June 2021

[62] “Omicron is the Dominant COVID Variant for Two Reasons.” Vitals. 2021-12-28

[63] Locatelli I, Trächsel B, Rousson V. “Estimating the basic reproduction number for COVID-19 in Western Europe.” PLoS One. 2021 Mar 17;16(3):e0248731. doi: 10.1371/journal.pone.0248731.

[64] Knight J, Mishra S. “Estimating effective reproduction number using generation time versus serial interval, with application to covid-19 in the Greater Toronto Area, Canada.” Infect Dis Model. 2020;5:889-896. doi: 10.1016/j.idm.2020.10.009.

[65] Marimuthu S, Joy M, Malavika B, et al. “Modelling of reproduction number for COVID-19 in India and high incidence states.” Clin Epidemiol Glob Health. 2021 Jan-Mar;9:57-61. doi: 10.1016/j.cegh.2020.06.012.

[66] Brauer F. “Compartmental Models in Epidemiology.” Mathematical Epidemiology 2008, 1945, 19–79. https://doi.org/10.1007/978-3-540-78911-6_2

[67] Robert B. Cooper. “Queueing theory.” In Proceedings of the ACM '81 conference (ACM '81). Association for Computing Machinery 1981, New York, NY, USA, 119–122. https://doi.org/10.1145/800175.809851

[68] Trapman P, Bootsma MC. “A useful relationship between epidemiology and queueing theory: the distribution of the number of infectives at the moment of the first detection.” Math Biosci. 2009 May;219(1):15-22. doi: 10.1016/j.mbs.2009.02.001.

[69] Bekker R, Broek MUH, Koole G. “Modeling COVID-19 hospital admissions and occupancy in the Netherlands.” Eur J Oper Res. 2022 Jan 5. doi: 10.1016/j.ejor.2021.12.044.

[70] Furman, E., Cressman, A., Shin, S. et al. “Prediction of personal protective equipment use in hospitals during COVID-19.” Health Care Manag Sci 24, 439–453 (2021). https://doi.org/10.1007/s10729-021-09561-5

[71] Ferreira, Manuel. “An Approach with a Queue System to the Pandemic Period Study.” Journal of Modern Accounting and Auditing 2018. 14. 10.17265/1548-6583/2018.05.006.

[72] Meares, H.D., Jones, M.P. “When a system breaks: queueing theory model of intensive care bed needs during the COVID-19 pandemic.” Med. J. Aust. 2020, 212: 470-471. https://doi.org/10.5694/mja2.50605

[73] Raffensperger John F., Marygail K. Brauner, Briggs R.J. “Planning Hospital Needs for Ventilators and Respiratory Therapists in the COVID-19 Crisis.” Santa Monica, CA: RAND Corporation, 2020. https://www.rand.org/pubs/perspectives/PEA228-1.html

[74] Ho YH, Tai YJ, Chen LJ. “COVID-19 Pandemic Analysis for a Country’s Ability to Control the Outbreak Using Little’s Law.” Infodemiology Approach. Sustainability. 2021; 13(10):5628. https://doi.org/10.3390/su13105628

[75] Little, J.D.C. “A Proof for the Queuing Formula: L = λW.” Oper. Res. 1961, 9, 383–387.

[76] Environmental Protection Agency. (n.d.). EPA. Available from: https://www.epa.gov/outdoor-air-quality-data/download-daily-data. [(accessed on 10 Jan 2020)]

[77] Covid-19 - mobility trends reports. Apple. Available from: https://covid19.apple.com/mobility. [(accessed on 10 Jan 2020)]

[78] United States. Worldometer. Available from: https://www.worldometers.info/coronavirus/country/us/. [(accessed on 10 Jan 2020)]

[79] Understanding the COVID-19 pandemic. USAFacts. Available from: https://usafacts.org/issues/coronavirus/. [(accessed on 10 Jan 2020)]

[80] Statista Research Department. "Mobile OS share in North America 2018-2021". Statista. Available from: https://www.statista.com/statistics/1045192/share-of-mobile-operating-systems-in-north-america-by-month/. [(accessed on 1 July 2020)]

[81] Ritchie H., Mathieu E., Rodés-Guirao L., et al. (2020, March 5). Coronavirus (COVID-19) vaccinations. Our World in Data. Available from: https://ourworldindata.org/covid-vaccinations. [(accessed on 22 Feb 2020)]

[82] Most polluted places to live: State of the Air. State of the Air | American Lung Association. (n.d.). Available from: https://www.lung.org/research/sota/key-findings/most-polluted-places\. [(accessed on 5 May 2020)]

[83] Benesty, J., Chen, J., Huang, Y., et al. "Pearson correlation coefficient. In Noise reduction in speech processing." Springer 2009, Berlin, Heidelberg, (pp. 1-4).

[84] Ritchie H., Mathieu E., Rodés-Guirao, L., et al. (2020, March 5). Coronavirus (COVID-19) cases. Our World in Data. Available from: https://ourworldindata.org/covid-cases. [(accessed on 5 May 2020)]

[85] Simpson R.B., Lauren B.N., Schipper K.H., et al. "Critical Periods, Critical Time Points and Day-of-the-Week Effects in COVID-19 Surveillance Data: An Example in Middlesex County," Massachusetts, USA. Int. J. Environ. Res. Public Health 2022, 19, 1321. https://doi.org/10.3390/ijerph19031321

[86] Aragão D.P., dos Santos D.H., Mondini, A., et al. "National Holidays and Social Mobility Behaviors: Alternatives for Forecasting COVID-19 Deaths in Brazil." Int. J. Environ. Res. Public Health 2021, 18, 11595. https://doi.org/10.3390/ijerph182111595

[87] Wikimedia Foundation. Covid-19 pandemic in California. Wikipedia. Available from: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_California. [(accessed on 23 May 2020)]

[88] Impact of opening and closing decisions in California, new cases - johns hopkins. Johns Hopkins Coronavirus Resource Center. (n.d.). Available from: https://coronavirus.jhu.edu/data/state-timeline/new-confirmed-cases/california. [(accessed on 23 May 2020)]

[89] "Timeline of the COVID-19 pandemic in the United States (2021). Wikipedia. Available from: https://en.wikipedia.org/wiki/Timeline_of_the_COVID-19_pandemic_in_the_United_States_(2021) . [(accessed on 23 May 2020)]