

國立臺灣大學理學院地理環境資源學系



碩士論文

Department of Geography
College of Science
National Taiwan University
Master Thesis

建立共識自組織對映於預測登革熱擴散的時空範圍

Creating a consensus self-organizing map
for predicting dengue diffusion in time and space

徐品翰

Pin-Han Hsu

指導教授：溫在弘 博士

Advisor: Tzai-Hung Wen, Ph.D.

中華民國 112 年 1 月

January 2023



國立臺灣大學碩（博）士學位論文 口試委員會審定書

建立共識自組織對映於預測登革熱擴散的時空範圍
Creating a consensus self-organizing map for predicting dengue diffusion in time and space

本論文係徐品翰君（r07228003）在國立臺灣大學地理環境資源學系、所完成之碩（博）士學位論文，於民國 110 年 8 月 25 日承下列考試委員審查通過及口試及格，特此證明。

口試委員： 溫在弘

(簽名)

甘泉源 (指導教授)

徐品翰

郭巧玲

謝辭

總共經歷四年半的碩士生生涯，建立共識自組織對映於預測登革熱擴散的時空範圍主題的論文總共寫了三個版本，第一個版本是研究計畫書，當時滿懷興致的從事研究過程，然而真正進入到研究方法與研究成果，經歷了多次的挫敗，由於欠缺思慮，導致受到其他博士生的質疑，幸好老師都能夠及時給予建議，並且安慰學生，給予學生一些空間去思索該如何完成一個論述，並且達成學術論文的要求。

然而到了口試期間又發展了另一個版本的論文，已經完成整個論文的架構與結果，可惜的是論文寫作的的能力有待加強，無法有效地說明研究貢獻，以及清楚的呈現成果，儘管老師們給予再努力的評價，然而這又是另一個開始，如何清楚的表達研究過程與成果。

口試之後是漫漫長日，儘管進度緩慢，但總在充滿靈感的時候完成一部份的論文補充，並且重新檢核模式與成果圖表，並且健全模式與加強論文的主題性。

口試過後經歷的一個學期獲得公費留學，前往俄羅斯聯邦莫斯科國立大學，在課業上面臨很大的挑戰，並且當時已通過俄文檢定的國家考試為目標，但僅有一個學期的公費留學，在當學期並沒有考取俄文檢定，幸好沒有繳交論文因此又延長學籍半年，持續在莫斯科國立大學學習，當時暑假有些空閒時間，便拿出論文寫了一半，直到下學期開始論文又呈現擱置的情況，但俄檢在九月順利通過，同時當年度的社會氛圍趨於正常化，開始不必膽戰心驚的生活，在下半學期過了一個豐富而且充滿喜悅的日子。

然而回到臺灣，已經是元旦了，依然念念不忘要完成論文，在研究生最後一段日子裡奮鬥，不管何時都是沉浸在論文裡，希望能奮力一搏，在最後時刻完成論文並繳交，取得研究所的學位。

然而當下回想起研究所的論文寫作的過程，不管是每個禮拜的會議，可以跟老師介紹我的研究發現，或者寫到天亮的研究專題討論，或者是漫長的口試與預口試，或者是開車到台北跟老師討論，開車到臺南參加研討會，這些點點滴滴在我人生中留下美好的印記。

國立臺灣大學地理環境資源學系

徐品翰 2023.2.1

中文摘要

發展登革熱每日滾動預測擴散模式目的，掌握登革熱疫情傳染的動向，採用時空預測的框架，達到精準匡列登革熱高風險區，縮小打擊範圍的進程，以 2010 年至 2015 年臺灣臺南市和高雄市的登革熱疫情作為研究案例。

我們將疫情分成兩個階段，疫情前期發病的登革熱病例資料作為訓練資料集，疫情中後期發病的登革熱病例資料作為測試資料集，為了發展預測模式，而建構時空擴散區的概念，並提出四種時空擴散每日滾動預測模式，模式一預設固定時空邊界、模式二歸納時空邊界、模式三利用空間異質性與模式四利用氣候、環境與空間異質性，模式三與模式四時空擴散區滾動預測模式都採用 MST-DBSCAN (Modified space time DBSCAN algorithm)與自組織對映類神經網路採用共識決演算法，我們定義擴散預測的綜合指標，挑選最佳的登革熱滾動擴散預測模式。

根據研究模式預測成果，在登革熱疫情非爆發年份之 2010 年至 2012 年，模式四利用氣候、環境與空間異質性的綜合指標表現最佳，該模式在疫情中後期，每日預測高達 80% 的登革熱病例，在登革熱疫情爆發年份之 2014 年與 2015 年，模式三利用空間異質性的綜合指標表現最佳，該模式在疫情中後期，每日預測高達 90% 的登革熱病例，彙整上述模式預測細節，發現登革熱的傳染半徑約 800 公尺，在疫情非爆發年份發病間距是從 6 至 11 日，在疫情爆發年份發病間距是從 6 至 9 日。

最後我們開發動態地圖，視覺化呈現登革熱每日滾動預測的擴散特徵，根據最佳模式生成的時空擴散區細節，協助我們歸納登革熱的傳染半徑與發病間距。

關鍵字：登革熱、自組織對映、DBSCAN、時空模式、滾動預測、共識決

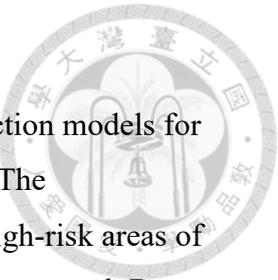
Abstract

The objectives of the study are to develop the daily rolling prediction models for the diffusion of dengue fever and to track the spread of the epidemic. The spatiotemporal prediction framework would accurately estimate the high-risk areas of dengue fever in near future for effective spatially targeting of epidemic control. Dengue fever epidemics in Tainan and Kaohsiung Cities of Taiwan from 2010 to 2015 were used for the case study.

Our approach divided the epidemic into two stages: an early stage for data training and a middle-late stage for data testing. To develop our predictions, we incorporated the concept of spatial-time diffusion zone. Four models for daily rolling predictions of dengue fever diffusion are proposed. Model I is based on a fixed spatial-time boundary, while Model II summarizes the spatial-time boundary. Model III utilizes the concept of spatial heterogeneity, and Model IV integrates the concepts of climate, environment, and spatial heterogeneity. Both Models III and IV utilized MST-DBSCAN (Modified Space Time Density-Based Spatial Clustering of Applications with Noise) and a self-organizing map modified by consensus decision-making algorithm. A comprehensive indicator also defined for determining the best daily rolling prediction model for dengue fever diffusion.

Our results show that, for the non-outbreak years of dengue fever (2010-2012), Model IV, incorporating climate, environment, and spatial heterogeneity, performed the best with the comprehensive indicators. This model accurately predicted up to 80% of daily dengue cases during the middle-late stage of the epidemic. Conversely, for the outbreak years of dengue fever (2014-2015), Model III, considering spatial heterogeneity, performed the best with the comprehensive indicators. This model achieved daily prediction accuracy of up to 90% of dengue cases during the middle-late stage of the epidemic. Our model analysis also found that the infection radius of dengue fever was approximately 800 meters. In the non-outbreak years, the serial intervals ranged from 6 to 11 days, while in the outbreak years, the serial intervals ranged from 6 to 9 days.

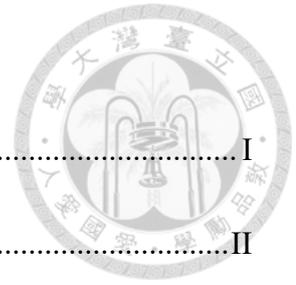
Finally, we created dynamic mapping animations for visualizing the diffusion patterns generated by the daily rolling prediction models of dengue fever. It provided detailed spatial-time diffusion zones, enabling us to summarize the infection radius and serial intervals of dengue fever.



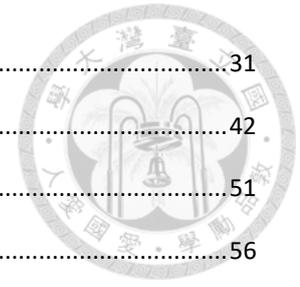
**Keywords: Dengue fever, Self-organizing map, DBSCAN, Spatial-time model,
Rolling prediction, Consensus**



目錄



中文摘要	I
Abstract.....	II
目錄	IV
圖目錄	VI
表目錄	VIII
第一章 緒論	1
第一節 介紹.....	1
第二節 登革熱的傳播.....	2
第三節 研究動機.....	3
第四節 研究目的	4
第二章 文獻回顧	6
第一節 登革熱的擴散模式	6
第二節 登革熱的預測模式	7
第三節 登革熱擴散與預測模式結合	10
第四節 小結.....	13
第三章 研究區與資料	15
第一節 研究區概述	15
第二節 資料來源	21
第四章 研究方法	23
第一節 研究架構.....	23
第二節 預測模式的資料集	24
第三節 模式共同項目	25
第四節 模式一	30



第五節 模式二.....	31
第六節 模式三.....	42
第七節 模式四.....	51
第八節 驗證研究成果.....	56
第五章 研究成果.....	62
第一節 綜合指標評量.....	63
第二節 時空擴散區滾動預測.....	70
第三節 正規化均方根誤差.....	77
第四節 時空擴散區細節.....	81
第五節 成果統整.....	93
第六章 討論.....	95
第一節 擴散預測模式.....	95
第二節 時空擴散區的特色.....	97
第三節 研究限制.....	99
第七章 結論.....	100
參考文獻.....	101
附錄.....	110
附錄一 研究成果補充.....	110
附錄二 平台.....	116
附錄三 演算法.....	117
附錄四 證明最佳化密度方程式.....	126
附錄五 門檻.....	127
附錄六 機器學習系統效能評估.....	129

圖目錄

圖 1、登革熱傳播週期的氣象和其他決定因素與臨床疾病的相互作用的方法論 (WHO, 2012).....	1
圖 2、時空擴散區概念示意圖	4
圖 3、登革熱病例發病位置最小統計區中心 2010 至 2015 年在研究區內的空間分布	17
圖 4、登革熱病例發病日按月份累積病例數	18
圖 5、研究區範圍包含高雄市與臺南市	21
圖 6、2010 年度登革熱每日病例數分布圖	24
圖 7、登革熱病例 x 屬性示意圖	26
圖 8、登革熱病例傳染關係	26
圖 9、登革熱時空擴散區	27
圖 10、模式一流程圖	30
圖 11、模式二流程圖	31
圖 12、疫情樹	32
圖 13、最短潛伏期估計	34
圖 14、最長潛伏期估計	34
圖 15、模式三流程圖	42
圖 16、輸入層示意圖	44
圖 17、自組織對映運作示意圖	47
圖 18、時空擴散區模式示意圖	49
圖 19、模式四輸入層示意圖	54
圖 20、測試資料集輸入滾動預測的輸入層向量示意圖	56
圖 21、單日滾動預測時空擴散區	57
圖 22、預測資料集的每日涵蓋率分布盒狀圖	68
圖 23、預測資料集的每日合適度分布盒狀圖	70
圖 24、2011 年測試資料集滾動預測時空擴散區與即時運算綜合指標的過程	73
圖 25、2015 年測試資料集滾動預測時空擴散區與即時運算綜合指標的過程	76
圖 26、各研究年份模式三的正規化方均根誤差衰減過程	80

圖 27、各研究年份模式四的正規化方均根誤差衰減過程.....	81
圖 28、測試資料集時空擴散區的病例密度.....	86
圖 29、疫情中後期時空擴散區的傳染半徑與發病間距上界組圖.....	91



表目錄

表 1、登革熱空間預測成果地圖分類與文獻說明	9
表 2、統整登革熱位於研究區各年度的指標病例發病日與嚴重程度	20
表 3、定義訓練資料集與測試資料集時間區間	25
表 4、模式變數代號統整表	29
表 5、測試資料集綜合指標各模式各研究年份預測成果表	64
表 6、訓練資料集綜合指標各模式各研究年份預測成果表	65
表 7、測試資料集的模式正規化方均根誤差	78
表 8、訓練資料集的模式正規化方均根誤差	78
表 9、測試資料集的時空擴散區的時空擴散範圍每日平均值	83
表 10、訓練資料集的時空擴散區的時空擴散範圍每日平均值	83
表 11、測試資料集時空擴散區的傳染半徑與發病間距細節	88
表 12、模式特性與適用情境統整	93
表 13、最佳模式的時空擴散區性質統整	94
表 14、清除半徑與相關文獻論點	98

第一章 緒論

第一節 介紹

登革熱透過蚊子進行人與人的傳染，疫情迅速擴張至全球是公共衛生挑戰 (Simmons et al., 2012)，氣候變遷導致登革熱現今比起過去五十年發生率增加率增三十倍，是目前傳播最快速的病媒傳染(WHO, 2015)，尤其是熱帶與亞熱帶地區面臨高疫情風險(Bhatt et al., 2013)，常見的登革熱傳染形式，是疫情流行地區透過境外移入病例造成疫情在當地擴散，導致當地疫情爆發(Semenza et al., 2014)。

臺灣位處熱帶與亞熱帶地區，鄰近於西太平洋登革熱流行區，臺灣與西太平洋人群的區域互動頻繁，導致境外移入病例增加臺灣本土登革熱的流行風險。臺灣地區的境外移入病例，幾乎每年都產生大小規模不一的本土病例，在本土登革熱疫情初期，局部地區會出現零星本土病例，但在 2014 年臺灣爆發數千例的本土登革熱病例，2015 年分別在臺南市與高雄市爆發嚴重的登革熱本土疫情，當年度全臺灣統計出四點三萬登革熱病例數(楊芝青, 2018)，臺灣衛生福利部提供登革熱病例詳細的開放資料，提供登革熱的研究堅實的基礎。

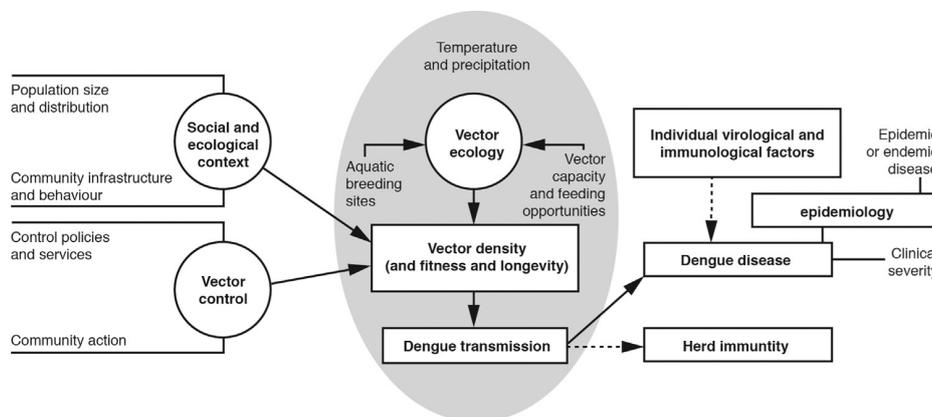


圖 1、登革熱傳播週期的氣象和其他決定因素與臨床疾病的相互作用的方法論 (WHO, 2012)。

登革熱高度貢獻的研究領域，他們根據當地環境氣候條件，預警各地區的登革熱發生率，提供疫情預防與管控的參考底本(Bhatt et al., 2013; Campbell et al., 2015)，世界衛生組織提供登革熱傳播週期的氣象和其他決定因素與臨床疾病的相互作用的方法論，期望研究人員能解決登革熱對人類的威脅(WHO, 2012)。世界衛生組織強調病媒蚊生態學是登革熱疫情研究的核心，病媒蚊生存條件構成病媒

蚊族群密度，病媒蚊族群密度可能受到社經條件與病媒蚊控制(Vector control)影響，病媒蚊族群密度直截影響登革熱的傳染，登革熱的傳染衍生流行病學主題(圖1)。

環境與氣候、病媒蚊與登革熱病毒存在複雜的回饋機制(Morin, Comrie, and Ernst, 2013)，預防登革熱疫情的傳播有效方法是病媒蚊控制，防止登革熱病毒地理邊緣擴張(Ebi and Nealon, 2016)，跨越多重領域了解登革熱傳染的機制，改進對未來疾病的風險預測模式，從而改善防疫策略阻斷登革熱的傳播(Morin, Comrie, and Ernst, 2013)，使得人類脫離登革熱疫情的威脅。

第二節 登革熱的傳播

阻斷登革熱的傳播是當前的研究問題，登革熱的病例來源可分成境外移入病例與本土病例，境外移入病例是區域邊界之間的人口移動(Liebig et al., 2019)，夾帶零星的病例在進入當地，在當地發生登革熱傳染事件，造成當地本土病例的登革熱疫情爆發的現象(Semenza et al., 2014)，本土病例是病媒蚊媒介造成的傳染事件，病媒蚊種類有埃及斑蚊(Thavara et al., 2001)與白線斑蚊(Semenza et al., 2014)，病媒蚊的密度嚴重影響登革熱疫情的傳染規模，但病媒蚊的族群分布與數量規模受制於氣候與當地人造環境與自然環境所影響。

由於登革熱的傳染受病媒蚊族群影響，因此當地的氣候環境條件可作為登革熱疫情的預測模式因子，透過預測因子進行各種預測模式建構，產生有效的風險預測模式，進行登革熱疫情評估，在登革熱疫情高風險地區採取噴藥消毒是有效阻絕登革熱傳播的手段(Othman et al., 2012; Calderón-Arguedas et al., 2009)，但登革熱的疫情分布具有空間的局限性(Adams and Kapan, 2009)，受制於病媒蚊、人造空間、環境、氣候、與登革熱病毒存在複雜的回饋機制。

另一方面在登革熱的疫情主要在熱帶地區，全年度可能有登革熱傳染的現象，但在亞熱帶地區由於病媒蚊族群受到季節氣候的影響，因此登革熱傳染的規模受氣候條件產生急劇的變異，其中高溫多雨是促使病媒蚊族群增長(Thavara et al., 2001; Watts et al., 1987)，促使病媒蚊的叮咬事件，構成本土病例急劇增長的條件(Platt et al., 1997)。

由於登革熱透過病媒蚊叮咬傳染的特性，因此發病病例在空間分布的位置呈現聚集的現象，疫情必須透過病媒蚊傳播特性，病媒蚊的移動距離相當的狹小，

大多數的登革熱傳染事件都有空間群聚的現象，局部區域因病媒蚊族群在三維空間的移動分布，導致在區域容易發生短時間登革熱傳播，病例急遽暴增的現象。

臺灣地區登革熱本土疫情面臨登革熱境外移入病例輸入，本土病例數成長發生春、夏、秋季且明顯分布在局部地區，分別在不同年度局部地區發生小規模的群聚傳染與大規模的疫情爆發現象，適合做為登革熱疫情傳染與風險預測模式的研究區。

第三節 研究動機

過去的登革熱的風險預測研究，是預測研究區的登革熱的發生率、病例數或分類的危險程度，但受限於研究區的空間單元，發生實際疫情控制區域與研究空間單元產生落差，且幾乎難以捕捉登革熱聚傳染的特性，導致登革熱預測模式應用在疫情管控的挑戰。

世界衛生組織針對疫情管控實務，避免過去的登革熱的預測高風險地區空間單元劃分問題，認為登革熱發病病例的傳染半徑 400 公尺，且 4 至 10 日以內從事孳生源清除(WHO, 2009a)，在病媒蚊的移動範圍有效清除登革熱的傳播。但參考疫情管控研究，我們發現各地區年份的研究，建議的傳染半徑具有明顯差異。另一方面疫情控制的時間點選擇更加複雜，必須考慮登革熱的潛伏期之後的發病日，來自不同的環境或模式設計，各種對於登革熱潛伏期具有不同論點。

本研究認為在多樣化的環境的疫情管控上，傳染空間半徑與潛伏期多樣性是必然的結果，由於研究採用的氣候、環境與病媒蚊條件不同，導致估計登革熱的管控空間範圍與潛伏期面臨挑戰，本研究推論可能是研究區當下的環境與氣候差異，可能多樣化傳染空間半徑與發病日。

預測傳染半徑與發病日模式是有效的時空模式，應用於匡列登革熱的高風險地區，模式沒有上述嚴重的空間單元劃分的問題，但多樣性的傳染半徑與潛伏期造成的挑戰，盡可能藉由氣候與環境因子影響病媒蚊族群概念，發展出適合預測當地登革熱疫情的時空模式。

將氣候與環境因子預測登革熱的疫情風險的模式，改造成預測登革熱的傳染半徑與發病日的模式，病例的發病位置的傳染半徑構成的環域是空間環域，潛伏期間產生最短發病日與最長發病日是時間間距，空間環域與時間間距構成時空模式，將傳統的風險預測模式，轉換成針對登革熱病例的時空範圍預測，有效解決

空間單元的問題，進而提升登革熱的時空預測精度，提升預測模式應用在疫情管控的實用性，有效縮小登革熱疫情打擊範圍。



第四節 研究目的

針對每日登革熱病例進行滾動預測，在登革熱疫情期間匡列疫情高風險區，縮小疫情的打擊範圍，預測登革熱病例周圍可能傳染半徑¹構成的空間環域，與病例發病日未來最短與最長可能發病日構成發病間距²，共同構成的時空擴散區，時空擴散區描述登革熱各別病例未來可能被傳染病例之間的傳染半徑與發病日區間(圖 2)，滾動預測的過程描繪出登革熱病例構成疫情傳染時空區，每日病例發病的預測時空生成的擴散動態。

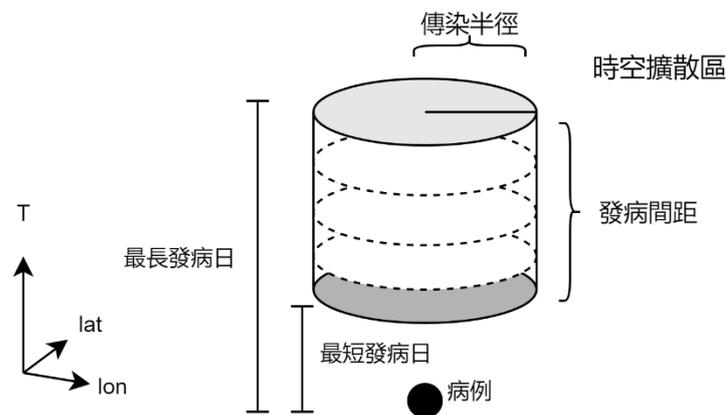


圖 2、時空擴散區概念示意圖。病例是登革熱病例，發病間距由最長發病日與最短發病日構成的時間區間，傳染半徑是未來被傳染病例可能的發病位置分布的環域，時空擴散區由傳染半徑與發病間距的概念構成。lat 緯度軸、lon 經度軸、T 時間軸。

在登革熱疫情期間，每日登革熱病例時空擴散區的滾動預測，是新的登革熱預測研究議題，本研究根據登革熱風險預測與登革熱傳染與擴散過程文獻進行模式改造，發展出具有代表性的登革熱病例時空擴散區的滾動預測模式，嘗試預測登革熱病例未來發展的時空範圍。

但我們無法得知根據過去的哪些登革熱的預測模式進行改造，對於登革熱疫

¹ 傳染半徑: 登革熱病例發病位置，與發生傳染事件生成新的病例位置之間的直線距離。

² 發病間距: 登革熱病例發病日，與發生傳染事件生成新的病例發病日之間的時間差距。

情期間的擴散預測，提供最佳的預測成果，因此本研究嘗試設計綜合指標，評估不同時空範圍預測滾動模式產生的長期時空擴散預測成果，評估登革熱病例預測的時空區是否足以包含未來的病例，並且評估時空擴散區是否合乎未來疫情的空間分布的範圍與發病日落在預測的發病間隔內，透過綜合指標揀選出最佳的登革熱時空擴散區滾動預測模式。

我們希望挑選出最佳的模式，可以模擬登革熱的疫情擴散過程，例如繪製每日滾動預測地圖，觀察每日的疫情發展與檢核模式生成時空擴散區的相關指標。我們還希望根據每日滾動預測過程中的時空擴散區細節，與模式的設計細節，可以提供未來發展登革熱擴散預測研究的經驗與基礎，同時可以掌握疫情登革熱疫情的趨勢，達到精確匡列登革熱高風險區，縮小打擊範圍的進程。

第二章 文獻回顧



第一節 登革熱的擴散模式

歸納疾病由源頭向外的時空效應，是釐清傳染的擴散途徑的關鍵(Sabel, Pringle, and Schærström, 2009)，空間視覺化平臺展示疫情的傳染過程(Carroll et al., 2014)，掌控疫情與防堵疫情擴散的有效方法。

一、群聚特徵

登革熱疫情的空間分布，具有明顯群聚傳染現象，早期的文獻中試圖採用非監督空間分析模式，登革熱群聚分析透過核密度基礎的演算法(Carvalho, Magalhães, and Medronho, 2017)，透過發病病例分布位置的空間群聚特性捕捉傳染源病例。

登革熱群聚分析透過聚類分析演算法(Kan et al., 2008; Shaukat et al., 2015)，比起核密度基礎的演算法，更加著重於多樣化的摸索模式、門檻值與參數的設定，根據演算法設置，模擬出合理的登革熱疫情空間分布群聚，同樣以演算法獲得相同的群聚作為相應的感染源。

登革熱群聚特徵藉由演算法模擬個別傳染源的空間分布，皆採用非監督空間聚類演算法，此類研究挑戰是缺乏時間軸、因此難以尋找疫情源頭的輸入病例，也難以解釋登革熱疫情隨時間的病程。

二、擴散特徵

必需在空間群聚加入時間軸的概念，釐清登革熱疫情在空間中隨時間的演進過程，專指登革熱病例視為單一群聚空間分布特徵變化過程，初探疫情移動的範圍(Wen and Tsai, 2015)，但往往登革熱的疫情是多重的感染源，因此會有多地在相近時間發生群聚傳染的現象，因此以個別時間點產生的子群聚，通過子群聚的時空相似性，反向推論登革熱群聚移動的過程(Chin et al., 2017)，這個過程是疫情在空間中擴散的現象。

登革熱疫情群聚移動過程無法滿足探索傳染事件，在近年的研究針對傳染的空間侷限性與時間序列掌握得更加全面，由於登革熱的傳染半徑與時間序列，具有大量的文獻提出佐證，可以更進一步建構疫情樹(Epidemic tree)架構，重建登革熱本土病例傳染與新的病例的傳染關係。採用 MST-DBSCAN 演算法(Modified Space Time DBSCAN Algorithm)演算法，重建多點疫情群聚形變與移動的疫情擴

散現象(Kuo, Wen, and Sabel, 2018)。結合蒙地卡羅法評估傳染半徑與發病間距，生成疫情傳遞鏈，重建疫情擴散特徵(Guzzetta et al., 2018; Morelli et al., 2012)。

擴散模式的生成促使學者證明，地區氣候環境條件對於登革熱疫情擴散具備高相關性(Li et al., 2019)，疾病的擴散在空間中的有效再生率(Effective reproduction number)具備嚴重的空間異質性(Haydon et al., 2003; Napp et al., 2016)，此推論區域環境對於疫情的擴散扮演重要的角色。

根據擴散模式的演進過程，近代研究證明登革熱的擴散與區域氣候環境具有高度相關性，可惜尚無研究可以透過區域的環境氣候建構登革熱的擴散現象。如果透過登革熱的傳遞鏈建構疫情樹，可能缺乏考慮地區氣候環境對於登革熱擴散的影響。另一方面地區氣候環境條件，對於登革熱疫情擴散具備高相關性(Li et al., 2019)，此論點是建構在未考慮氣候環境的擴散模式，進行再生率與氣候環境之間的相關性分析，此分析法無法應用地區氣候環境的特性，更加精確的建構疫情樹，作為重建疫情擴散的基礎，此研究僅止於推論環境氣候的相關性。

第二節 登革熱的預測模式

登革熱的預測模式是根據當地的環境氣候特性，預測該地當時的登革熱風險程度，此類研究幾乎無法考慮登革熱群聚傳染導致的空間異質性，另一方面未考慮疾病本身的空間侷限性，登革熱疾病的發病間距導致疫情的時序變化過程。

一、時序預測

登革熱主題最常見的預測模式，時序預測登革熱病例數、爆發或發生率，常見研究類型是監督式預測模式，關注季節性與氣候對於長期預測與實際的登革熱病例數或發生率的擬合現象，非線性迴歸系列模式的預測，通常採用季節性或自相關的特性預測登革熱的嚴重程度(Hii, Zhu, et al., 2012; Hii, Rocklöv, et al., 2012; Wu et al., 2007; Wu et al., 2008; Ho and Yee, 2015; Althouse, Ng, and Cummings, 2011)，操作的因子需要經過前處理階段，預測方法採用混合模式進行交岔驗證，每個研究區都有獨特的模式擬合病例數或發生率，例如馬來西亞的登革熱預測研究採用差分整合移動平均自迴歸模型(Autoregressive Integrated Moving Average model)，應用季節性特徵預測登革熱病例數(Ho and Yee, 2015)，新加坡與曼谷研究區使用聲量趨勢因子，分別複合選擇線性迴歸、負二項迴歸、梯度提升迴歸

(Generalized boosted regression)預測進行比較，加上自動向量機預測病例數(Althouse, Ng, and Cummings, 2011)，利用氣候因子透過複迴歸加上波松迴歸預測病例數(Hii, Zhu, et al., 2012)，然而這些當地獨特的預測模式可能無法通用於其他研究區。

類神經模式通常會採用前一期的病例數作為輸入層，採用監督式的類神經網路預測登革熱的病例數(Baquero, Santana, and Chiaravalloti-Neto, 2018; Laureano-Rosario et al., 2018; Aburas, Cetiner, and Sari, 2010)，類神經模式是相對迴歸模式容易操作，整體的擬和度非常高，而且相同多層式網路的模式可以適用在多個區域，但面對疫情嚴重爆發期間，預測效果普遍不如預期，無法反映大量病例數與高風險的現象，類神經模式適用性高但會面臨低估當地疫情風險的隱憂。

時序預測登革熱登風險或病例數，根據研究區內氣候環境或社會經濟發展作為因子，致力於擬和真實值與預測值之間的差異，模式只能針對於該單一特定研究區域，對於其他區域或其他空間單元，可能存在不適用的隱憂。

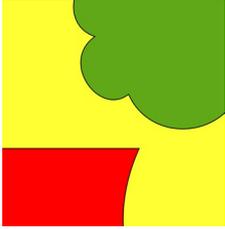
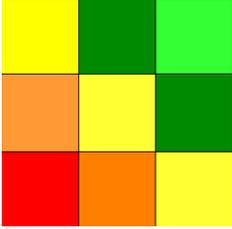
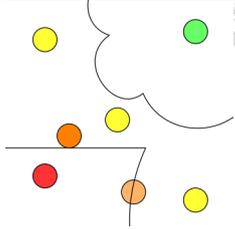
二、時空預測

針對時序預測解決特定區域模式問題，預測研究中加入考量研究區中，不同的空間單元的預測成果，可能提升模式的通用性作法。

時空預測作法類似時序預測，只是在時序預測中，考量不同空間單元預測成果，但空間單元卻面臨設定空間單元，與採用空間解析度的困難。本處根據空間預測的主題與呈現的主題地圖進行分類，說明過去研究處理空間單元與空間解析度的問題，繪製的地圖種類包含面量圖、網格圖、點位圖(表 1)，該類型主題圖以顏色鑑別該空間單元的預測風險程度。

面量圖根據原始資料行政區空間單元繪製風險圖，相關文獻凸顯行政區的氣候環境條件導致空間單元中的發生率的差異(Guo et al., 2017)，行政區中心法取得內插資料的前處理(Chien and Yu, 2014)，行政區中心的預測值對應輸出在該行政區，原始行政區的空間單元也許不能代表登革熱疫情的空間分布。

表 1、登革熱空間預測成果地圖分類與文獻說明。以下地圖分類紅色代表高風險，黃色代表低風險，綠色代表低風險的抽象概念。

地圖種類	面量圖	網格圖	點位圖
示意圖			
文獻	Guo et al., 2017 Semenza et al., 2014 Chien and Yu, 2014 Lowe et al., 2011 Hales et al., 2002	Ong et al., 2018 Bhatt et al., 2013 Buczak et al., 2012 Lowe et al., 2011	Liebig et al., 2019 Gomide et al., 2011
時空解析度	每周一個市行政區 (例如廣東省中山市)	每年 1x1 平方公里網格	每月份的機場點位
空間單元	行政區空間	網格	地標

網格圖是空間預測最為常見的總類，避免受制於行政區空間單元影響預測成果，網格會受到空間解析度的選取影響網格圖的功能性，例如空間解析度高達 2.5 乘 2.5 經緯度子區域，預測巴西全境的年登革熱風險(Lowe et al., 2011)。或者全球作為空間尺度，根據各年度資料預測登革熱的發生率、潛在受感人口，以 5 乘 5 平方公里每年的空間解析度繪製全球的疫情風險圖(Bhatt et al., 2013)。例如新加坡的登革熱疫情年度風險，繪製空間解析度 1 乘 1 平方公里網格(Ong et al., 2018)。

網格中心加值資料預測，以新加坡的登革熱風險程度分析為例，採用年度一平方公里的網格(Ong et al., 2018)，每個網格代表當年度的風險，在空間解析度似乎可以滿足登革熱的主題圖，但在時間解析度上在非新加坡地區可能不適用。原因是新加坡全年皆有登革熱病例，但以臺灣為例，登革熱疫情期間通常僅有半年(Wang et al., 2016)，將年度作為時間解析度對於臺灣登革熱得預測，可能發生不合適的情形。

點位圖具有地標特性，例如機場透過航空網路導致登革熱境外移入的病例數(Liebig et al., 2019)，都市族群網路聲量與情感反映登革熱疫情的發生率(Gomide et al., 2011)，這些登革熱的空間預測針對特殊地標進行論述，出自於地標的空間特性，不適用於預測登革熱的時空擴散。

根據登革熱空間預測的文獻，針對登革熱的預測主題，網格圖可能是最適合呈現登革熱的風險，但採用網格圖的預測模式，會落入時空解析度取決於主題圖的功能性，與當地的登革熱疫情的時空分布，對於登革熱預測的通用性，甚至在臺灣的登革熱預預測上面臨挑戰。

第三節 登革熱擴散與預測模式結合

擴散與預測模式的結合目的，根據登革熱的傳播特性，挑選過去的登革熱對於發展登革熱擴散與預測的模式，進行微幅度的模式修整後，發展出固定時空範圍的擴散預測模式，經過參數估計的不固定時空範圍擴散預測模式，結合地理位置的時空範圍擴散預測模式，與結合地理位置與環境氣候的擴散預測模式，繪製疫情預測高風險區，比較各種觀點的適用性。

一、擴散模式

由於時空預測的採用網格圖的形式，可能會落入時空間解析度的爭論，本研究欲探討登革熱的擴散預測，目的是在登革熱疫情期間匡列疫情高風險區，縮小疫情的打擊範圍，根據登革熱疫情擴散的性質，挑選出合適預測登革熱擴散的時空單元。

登革熱的病例可能發生超過一次的傳染現象，尤其是血性登革熱(Dengue hemorrhagic fever)與休克型登革熱(Dengue shock syndrome)，可能被傳染者帶源兩種以上的血清型的登革熱病毒(Kalayanaroj, 2011; Changal et al., 2016; WHO, 2009b)，病毒血清型的差異會造成不同的受感性 (Amâncio et al., 2014)，單一登革熱病例可能又藉由蚊蟲叮咬的過程傳染給多個病例(Thavara et al., 2001; Platt et al., 1997)，登革熱病例扮演傳染者與被傳染者的關係都是多對多的關係。

疫情樹架構透過演算法模擬登革熱傳染的過程，疫情樹參數設定是參考過去登革熱對於病媒蚊的飛行距離，與潛伏期推測發病間距。研究登革熱擴散採用MST-DBSCAN 演算法(Modified Space Time DBSCAN Algorithm)演算法(Kuo,

Wen, and Sabel, 2018)是很好的例子，採用病例 300 公尺的傳染半徑，與 6 至 23 日的發病間距，作為重建群聚傳染的演算模式。擴散預測的時空單元，病例周圍的傳染半徑公尺，與發病間距日數，共同構成的時空範圍，可能符合登革熱疫情擴散與縮小打擊範圍的研究目的。

此研究提出的時空解析度分別是公尺與日數，明確對應世界衛生組織建議，發病病例的傳染半徑 400 公尺 4 至 10 日以內從事孳生源清除(WHO, 2009a)，潛伏期日數作為單位是登革熱估計發病日的文獻(Chan and Johansson, 2012)。

DBSCAN 演算法特性，不需預設中心點為前提，只要預設中心性門檻繪製有邊界的分群(Parimala, Lopez, and N C, 2011; Birant and Kut, 2007; Ester et al., 1996)，不需預設中心點性質符合登革熱傳染特徵，只能重建疫情擴散的過程推論可能的傳染源，本研究根據 MST-DBSCAN 演算法設計固定時空範圍的擴散預測模式，此基本模式與其他模式比較的差異進行探討。

疫情樹的實作不僅能透過 MST-DBSCAN 固定時空範圍的演算法，例如參數估計不固定時空範圍可能，比起固定時空範圍的演算法更加有參考價值，例如透過馬可夫蒙地卡羅法估計傳染半徑與發病間距(Guzzetta et al., 2018)，因此改良 MST-DBSCAN 演算法中固定時空範圍，利用參數估計產生不固定時空範圍，經過改良後是否比起基本模式提升預測能力，也是本研究探討的議題。

二、預測模式

建構在 MST-DBSCAN 固定時空範圍的演算法，除了透過參數估計產生不固定時空範圍，透過地理位置與環境氣候資料是否能提升擴散預測的能力。本研究根據地區氣候環境條件對於登革熱疫情擴散具備高相關性(Li et al., 2019)，但疾病傳染的空間分布可能有嚴重的空間異質性(Haydon et al., 2003; Napp et al., 2016)，空間異質性可能也是造成登革熱的擴散特徵差異，可能來自於不同病媒蚊族群的移動距離(Verdonschot and Besse-Lototskaya, 2014)，另外氣候條件可能會造成病媒蚊的族群消長，可能造成登革熱的傳播距離(Shannon and Davis, 1930)與發病間距(Chan and Johansson, 2012)波動的现象。

本研究應該發展考量空間異質性的不固定時空範圍預測模式，其中一種模式僅採用病例的發病位置預測登革熱的擴散，另一種模式除考慮病例的發病位置，並考慮環境與氣候的不固定時空範圍預測模式，這兩種模式分別採用不同的因子



進行訓練，用以比較考量氣候、環境與空間異質性，比起考量空間異質性的模式在預測能力獲得提升。

三、預測模式實作

由於建構在 MST-DBSCAN 固定時空範圍的演算法，產生不固定的固定時空範圍，必須採用預測模式進行實作，過去登革熱的預測研究通常採用監督式學習，但這方面的研究通常針對預測風險程度與病例數，但面對預測登革熱擴散的時空範圍，監督式學習無法符合病媒蚊叮咬的過程傳染給多個病例(Thavara et al., 2001; Platt et al., 1997)，另一方面我們無法得知真實的傳染關係，只能根據重件傳染關係進行推測，非線性的困難複雜度問題(Non-deterministic polynomial hard, NP-hard)(Sipser, 2006)，非線性面臨嚴重的模糊化與雜訊的挑戰，自組織對映分群演算法降低熵提供這類困難問題的解法(Vesanto and Alhoniemi, 2000; Mu-Chun, DeClaris, and Ta-Kang, 1997)。

自組織對映是成功的案例是良好的處理雜訊能力(Lippmann, 1987; Kohonen, 2001)，例如辨識文字能力超越傳統多層感知器(Multi layers perceptron)的監督式學習(Sung-Bae, 1997)，僅定義各種形式的拓樸圖，與結合最佳匹配單元(Best match unit) (方程式 1)與鄰近修正單元進行迭代(方程式 2)(Kohonen, 1990)，便能有效的分群進行且標定進行預測。

$$c = \underset{i}{\operatorname{argmin}} \{\|x - m_i\|\} \quad (\text{方程式 1})$$

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x - m_i(t)]$$
$$h_{ci}(t) = \alpha(t) \exp\left(-\frac{\|y_c - y_i\|}{2\sigma^2(t)}\right) \quad (\text{方程式 2})$$

m 神經元權重、 argmin 運算最小(距離)、 i 選定的神經元、 c 活化神經元、 x 輸入層一筆資料的向量、 t 迭代數、 h_{ci} 鄰近性函數又名高斯方程式、 $\alpha(t)$ 學習率、 $\sigma(t)$ 核心寬度。



自組織對映類型分群特性，處理地理區域性與環境問題領域獲得廣大的應用，原因來自於拓樸圖地圖可視化，容易凸顯區域特性(Chang, Shen, and Chang, 2014; Chang et al., 2010; Aggarwal and Yonghua, 1998)，拓樸圖模擬環境空間非線性的數值分布給予最合理的解釋(Vesanto and Alhoniemi, 2000; C. LourenÇO, Lobo, and Bação, 2018; Smith, 2002)，空間科學領域甚至利用自組織拓樸圖與空間異質性改良自組織對映，限制拓樸空間修正神經源權重(Bação, Lobo, and Painho, 2005)。處理病媒蚊叮咬的過程傳染給多個病例(Thavara et al., 2001; Platt et al., 1997)，這些多的病例的傳染關係是根據時空範圍推論，因此雜訊處理能力的特性對於估計時空擴散的過程是重要的利器。

另外一個技術層面的自組織對映網路的優化改善，由於登革熱的傳染與被傳染關係是多對多的傳染關係，為了提升擴散預測的穩定性應該採取共識決，赫賓學習法認為人腦神經有集體觸發作用(Shaw, 1986)，自組織對映的特色是臨近性修正權重類似集體觸發的概念，過去的研究集體觸發作用後產生共識(Consensus)決定輸出，此類型模式的類神經網路可以提升預測的穩定度(Hopfield, 1982; Ackley, Hinton, and Sejnowski, 1988)，而本研究期望能採用活化神經元產生進行共識決，非傳統類神經的純量輸出，由於過去文獻常見登革熱擴散研究，透過傳染半徑與發病間距的最佳化(Wen, Lin, and Fang, 2012; Guzzetta et al., 2018)，登革熱的傳染與被傳染關係是多對多的傳染關係，因此共識決的作法可能更適合登革熱疫情擴散的研究主題，然而共識的決議結果進行最佳化，此做法是統計上專家模式(Expert modeling)的共識決議，面對每個分布可能具有共同性與相異性，尋求共同性與相異性之間的平衡(Winkler, 1981; Genest and Zidek, 1986)，透過共識決議登革熱傳染的傳染過程，在利用最佳化的概念更進一步推論擴散得時空趨勢。

第四節 小結

研究發現登革熱有群聚傳染現象，然而群聚傳染隨時間的移動，是疫情擴散的過程，然而近代學者為了捕捉更細緻的疫情擴散特徵，因此認為登革熱傳染有如疫情樹，傳染者與被傳染者間具有傳遞鏈關係，因此疫情擴散的觀點不在只是單純的群聚傳染隨時間移動的過程，是從重建傳遞鏈的觀點切入疫情擴散的過程，重建傳遞鏈的研究，發現傳染特徵可能具有空間異質性與環境氣候的異質性。

然而空間異質性與環境氣候的異質性的觀點，已經產生大量的預測登革熱病例與風險的研究，但登革熱的時序預測與時空預測，往往面臨時空解析度的挑戰，尤其是針對臺灣登革熱疫情期短暫而多變，很難產生常態性的時序或空間預測模式。

因此我們透過疫情擴散的觀點來預測登革熱，利用傳遞鏈重建登革熱疫情擴散，傳遞鏈給予我們的時空解析度，是傳染半徑單位是公尺；是發病間距單位是日，傳染關係可能是多個病例對多個病例的傳染，病媒蚊受到氣候、環境與區位條件影響族群。

利用傳遞鏈重建登革熱擴散，在缺乏實際傳染關係的資料，也可以透過 MST-DBSCAN 進行重建傳遞鏈，然而我們無法確認與實際傳染關係相輔，因此我們應該透過自組織對映與共識決的演算法設計，盡可能找到最符合疫情擴散的傳遞鏈形式。

我們認為應該要發展時空擴散區概念的預測模式，同時採用傳遞鏈給予時空解析度，另一方面根據過去文獻對於擴散的觀點，發展四個擴散預測模式，預設固定時空邊界、歸納時空邊界、空間異質性和氣候、環境與空間異質性的模式，比較四種模式的擴散預測成果，由於缺乏實際傳染關係的資料，過去的文獻也沒有一致的檢核模式優劣標準，因此我們必須發展擴散預測的指標，評估各種概念建立模式是否合乎登革熱的擴散過程。

第三章 研究區與資料



第一節 研究區概述

本研究區地處於臺灣西南部，由北往南涵蓋高雄市與臺南市兩大行政區，兩地互相接壤，臺南市行政區位於北緯 23°24' 到 22°53'；東經 120°01'36" 到 120°38'53" 之間共 37 個鄉鎮市區；高雄市行政區位於北緯 22°28' 到 23°28'，東經 120°10'32" 到 121°01'15" 之間共 38 個鄉鎮市區，西半部地區與海洋接壤，東半部地區為丘陵與山地，人口分布由西部濱海地區往東部山區遞減。氣候屬於熱帶季風氣候區，以 2015 年為例臺南市年平均溫度 25 攝氏度、高雄市為 26 攝氏度 (Central Weather Bureau, 2015)。

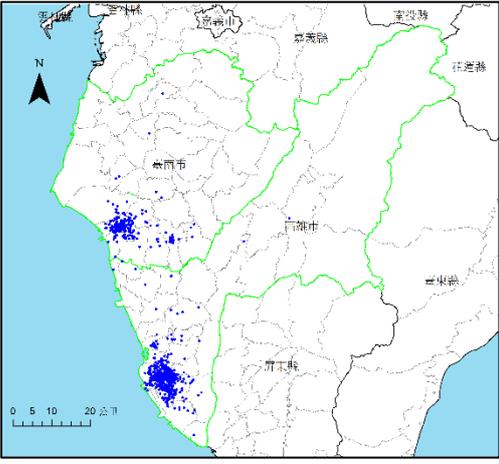
一、登革熱病例的空間分布

臺灣地區登革熱病例主要分布於高雄市與臺南市兩大行政區範圍，於 2010 至 2015 年六年度該研究區共計 61828 例登革熱發病病例，接近全國登革熱病例數量 76874 例，並且在該地好發大小規模的登革熱群聚傳染現象，因此兩大行政區作為登革熱疫情擴散的研究區。

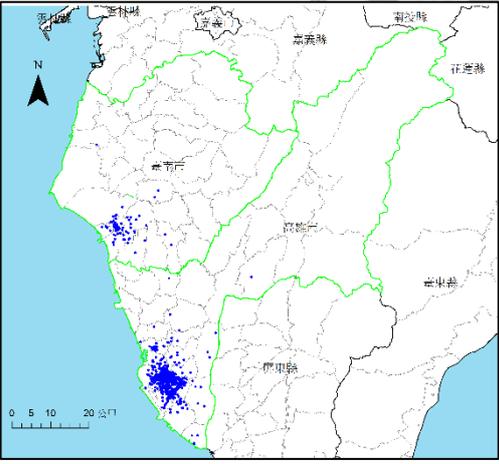
就研究區各年度統計累積登革熱病例數於 2010 年共 1555 例、2011 年共 1297 例、2012 年共 1286 例、2013 年共 152 例、2014 年共 15198 例與 2015 年共 42340 例，以上各年度的病例發病位置分布如登革熱病例發病位置最小統計區中心 2010 至 2015 年在研究區內的空間分布組圖 3a-f，其中病例的發病位置由於個人資料機敏性，因此疾管署僅提供病例發病位置的最小統計區中心點經緯度座標。



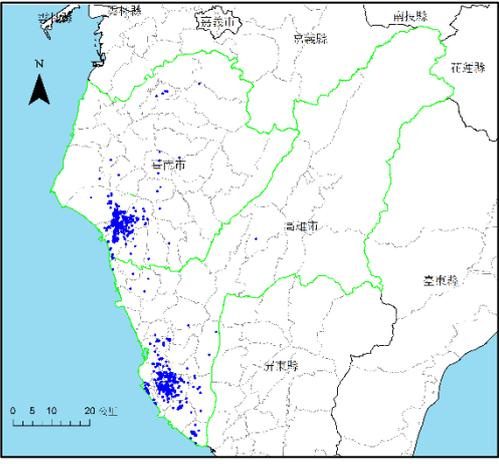
a.



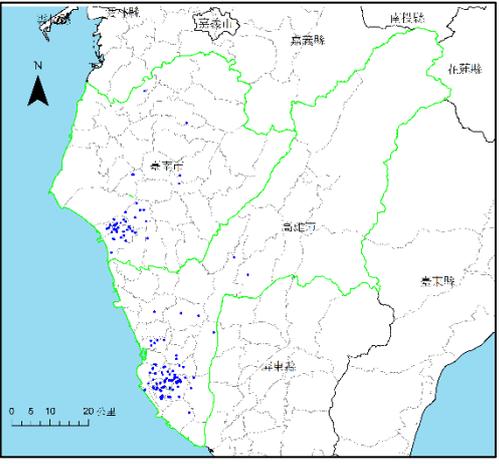
b.



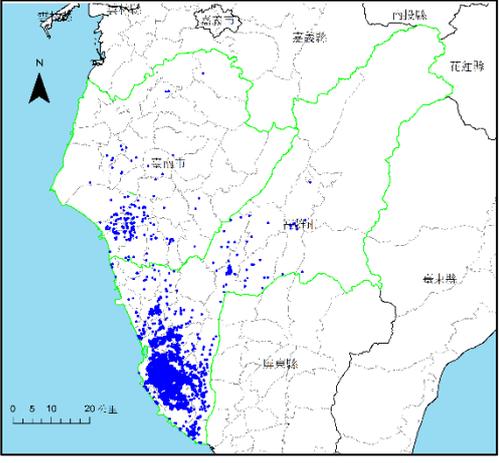
c.



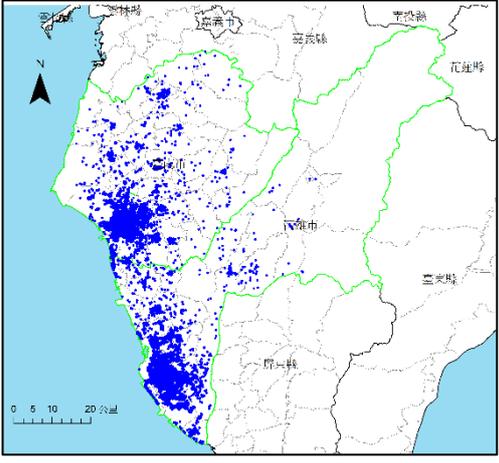
d.



e.



f.



續下頁

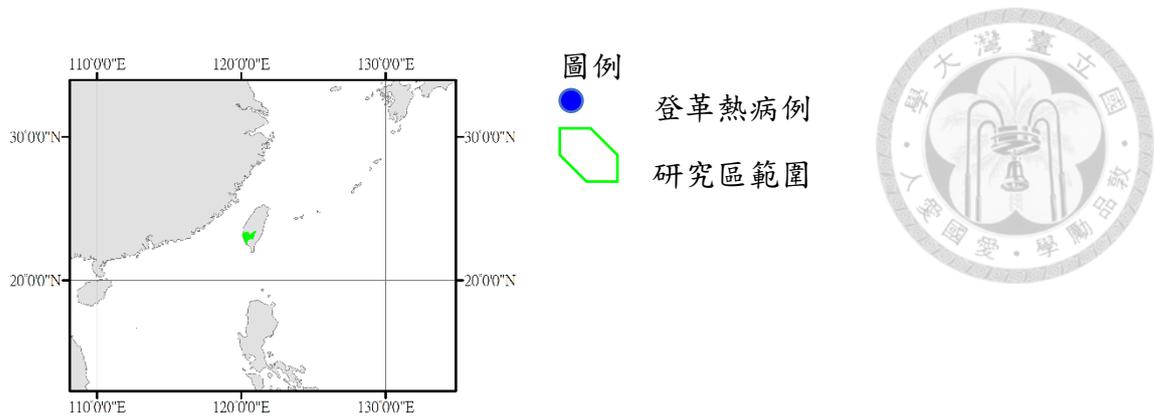


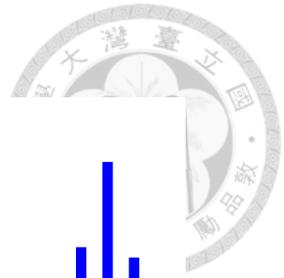
圖 3、登革熱病例發病位置最小統計區中心 2010 至 2015 年在研究區內的空間分布。a-f 分別是 2010 至 2015 年度的登革熱病例發病位置最小統計區中心的分布情形。

研究區就 2010 至 2012 年期間登革熱病例發病位置分布，集中於臺南市與高雄市人口密集區域，其餘位置病例分布稀疏。而 2013 年度病例發病位置空間分布最為稀疏，病例數僅數百例，群聚傳染現象並不明顯。然而 2014 至 2015 年期間病例發病位置遍布研究區西半部臨海地區，且集中於人口密集地區現象，群聚傳染現象明顯，該兩年度視為疫情爆發期，超過一萬例的病例數量。

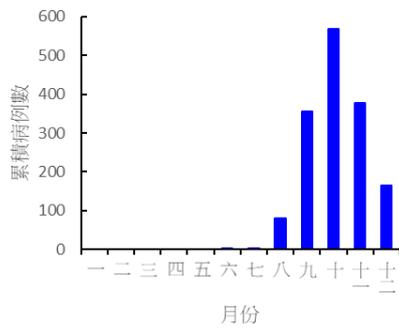
研究區於兩行政區就整體而言登革熱的病例數由西向東遞減，登革熱病例好發於西部平原地區，西部平原地區病例特別集中於人口密集且靠近濱海地區，以臺南市的西南端與高雄市的西南端之都會區登革熱疫情特別嚴重。其他零星病例分布於由臺南市與高雄市都會區向外延伸的郊區與鄉村地區，再向東方延伸至高雄市與臺南市的山區地帶，登革熱的病例數量在六年間的累積低於百例，在兩行政區的東半部登革熱的病例稀少。

二、登革熱病例的發病日程

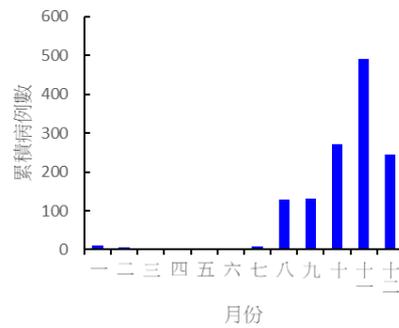
登革熱由 2010 至 2015 年位於臺南市與高雄市兩行政區的病例數案月份累積的分布，位於該研究區分別藉由個年度統計，當月份登革熱發病日所累積的病例數，如下組圖 4a-f 長條圖所示，觀察登革熱病例發病日，針對各月份與季節觀察長期趨勢。



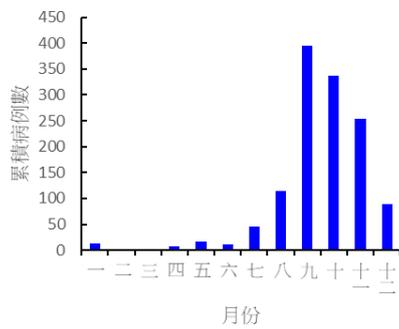
a.



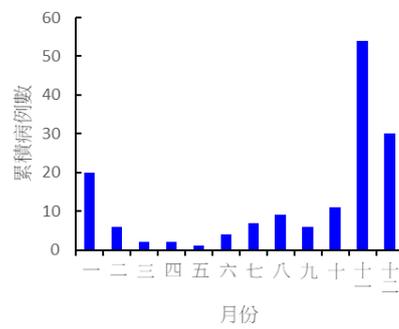
b.



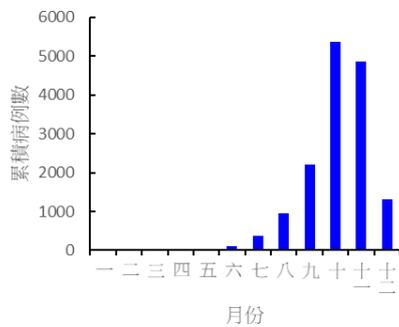
c.



d.



e.



f.

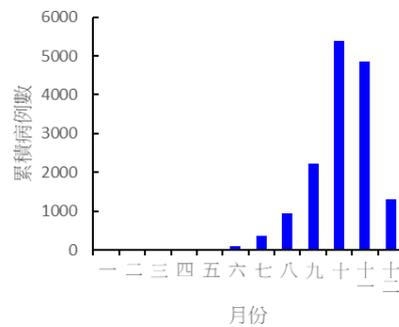


圖 4、登革熱病例發病日按月份累積病例數。a-f 分別是 2010 至 2015 年度的當月累積的病例數，經由登革熱病例發病日統整後的成果。

在 2010 年六月開始出現 4 例登革熱病例，至當年度九至十一月每月累積病例數超過 300 例以上，至十二月疫情緩和當月累積病例數 164 例，至隔年二月疫情完全結束，隔年一月與二月合併僅累積 17 病例數。

2011 年五月出現極少數病例，到八月、九月與十月期間，每月累積病例數超過一百筆以上，當年最嚴重為十一月當月累積病例數 491 例，往後月份疫情逐漸緩和至隔年一月消失。



2012 年二月至四月發生極少數病例，直到當年五月當月累積病例數 16 例，直到當年八至十一月每月累積病例數皆超過百例小於千例以下，直到十二月緩和當月累積病例數低於百例以下共 89 例。

隔年 2013 年全年度每月累積病例數皆低於百例以下。

2014 年一月與二月當月累積病例數小於 10，當年三月無任何病例發生，但四月至六月疫情趨向百例當月累積病例數，至七月與八月疫情趨向千例當月累積病例數，九至十二月疫情相當嚴重；當月累積病例數分別是 2211、5379、4858、1305 例，至 2015 年一月與二月當月累積病例數下降至百例以下。

2015 年從四月當月累積病例超過十例以上開始，至當年度八月當月累積病例數高達 4390 例，九月更超過 14909 例，十月與十一月當月累積接近一萬例，直到十二月疫情明顯趨緩當月累積病例數 3040 例，至隔年 2016 年一月疫情結束。

統整登革熱由 2010 至 2015 年位於臺南市與高雄市兩行政區的病例數與月份分布，總共經歷 5 波大幅度的病例數升降，總共經歷 5 波疫情，除 2013 年各月份每月累積病例數低於百例以下，其中 2014 與 2015 年病例數隨月份升降的幅度最為明顯，在疫情爆發的月份高達千例至萬例的當月累積病例數，其他年份與月份僅千例以下的每月累積病例數。

統整各年度疫情開始是四月份，往後五月至八月疫情逐漸升溫，到九至十一月達疫情高峰，十二月後疫情逐漸趨緩與消失。

三、各年度的登革熱疫情概述

登革熱主要發生在臺灣南部的高雄市與臺南市，臺南市與高雄市是白線斑蚊、埃及斑蚊與人類宿主混合區位特性，導致該區成為登革熱疫情流行地區 (Yang et al., 2014)，首先在 2010 至 2015 年度當地疫情概況進行介紹，臺灣地區已經發生四種血清型的登革熱，2010 年於臺南市發生第四血清型登革熱，高雄市發生第二與第三血清型登革熱，2011 與 2012 年高雄市以第一血清型登革熱大宗，2013 年非常少數且分散於屏東縣以第二血清型登革熱，2014 年高雄市嚴重爆發第一與第二血清型登革熱，隔年 2015 年高雄市與臺南市爆發第二血清型登革熱為主，高雄市出現少數的第一血清型登革熱病例，2016 年往後的病例數極少，在 2015 年以前幾乎每年都有一波疫情 (Hsieh, 2018)，因此登革熱擴散模式的研究年份僅至 2015 年。

2010 至 2013 年屬於常態性少數地區發生幾波零星病例，2014 年僅高雄市發生登革熱大流行，大流行的原因推測當年七月高雄市發生氣爆後導致產生許多積水與登革熱孳生源(Hsieh, 2018; Wang et al., 2016)，隔年 2015 年登革熱疫情分別於高雄市與臺南市爆發一波嚴重的大流行總數超過四萬例，但無法提出 2015 年爆發具體的解釋(Hsieh, 2018)，唯一根據長期氣候推測 2014 與 2015 年的每月降雨與溫度明顯略高於十年平均溫度與雨量，些微的氣候變遷導致登革熱疫情大流行(Wang et al., 2016)，統整結果 2014 與 2015 年是屬於登革熱疫情大流行的年份(Wang et al., 2016; Hsu, Hsieh, and Lu, 2017; Hsieh, 2018)，本研究定義 2014 與 2015 年為登革熱疫情爆發年，其他年份 2010 至 2013 年為登革熱疫情非爆發年(表 2)。

表 2、統整登革熱位於研究區各年度的指標病例發病日與嚴重程度。

年分	2010	2011	2012	2013	2014	2015
指標病例發病日	6/15	7/3	6/17	6/16	5/20	5/16
疫情嚴重程度	非爆發	非爆發	非爆發	非爆發	爆發	爆發

我們觀察圖 4a-f 的疫情，我們歸納大致疫情走向在 2010 至 2013 年間之登革熱非爆發年分，各個年度在六月中旬爆發零星病例，在社區中爆發零星的群聚傳染案例，在秋季進入疫情的高峰，直到冬季疫情進入尾聲，直至隔年一月登革熱疫情消失的現象，我們透過登革熱疫情在社區的分布情形觀察指標病例，在 2010 至 2013 年間發生首起社區群聚傳染的月份集中於六月或七月。然而登革熱爆發年份之 2014 年與 2015 年疫情首起指標病例發生於五月中旬，延續疫情至隔年一月後消失。

四、研究區人口分布

研究區範圍包含高雄市與臺南市，觀察圖 5 人口主要分布於西部沿海地區，人口呈現東高西低的現象，於高雄市與臺南市西南邊構成人口密集的都會區，都會區的分布位置，似乎與登革熱病例發病位置有空間重疊的現象(圖 4a-f)，我們可能觀察人口密集地區，可能病例數量較高的現象，然而這個推論必須透過模式的建立說明這個推論的可靠度。

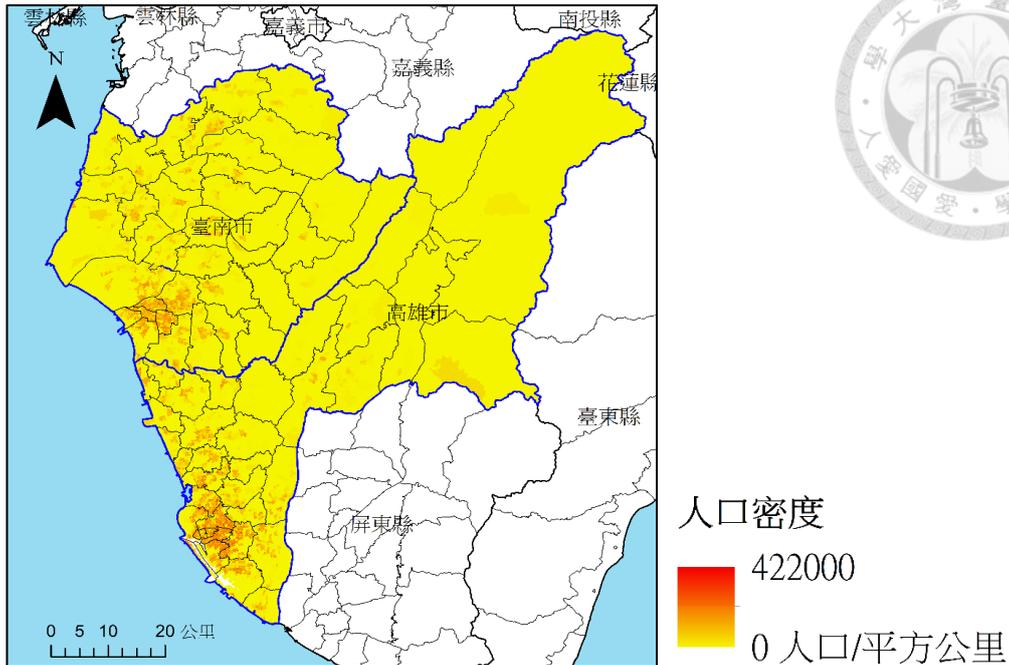


圖 5、研究區範圍包含高雄市與臺南市。研究區範圍是藍色線內的兩個行政區臺南市與高雄市，著色黃紅漸層是研究區的人口密度分布以 2010 年 12 月人口密度為例，白色部分是研究區鄰近縣市，淺藍色部分是海洋。

第二節 資料來源

一、登革熱病例資料

登革熱病例資料於中華民國衛生福利部疾病管制署公布登革熱 1998 年公布每日確定病例統計資料政府開放資料平臺，資料更新頻率是每日更新，本研究擷取 2010 年至 2015 年共六年度，擷取位於臺南市與高雄市的登革熱病例發病日、經緯度座標、境外移入屬性資料。

二、氣候資料來源

每日天氣資料來源為中華民國交通部中央氣象局提供觀測資料查詢平臺，於 2010 至 2015 年共六年間每日臺灣全域 595 有效測站資料擷取每日最低溫度、每日雨量，進行反距離加權(Inverse distance weighting)內插平行運算結果(Shepard, 1968)，產生五年之間每日臺灣本島所有地區的最高溫度與日累積雨量資料的面量圖，提供登革熱病例發病日與發病位置的氣候背景資料。

以下列出反距離加權的演算法，反距離權重次方採 p 等於 3 的形式，反距離權重採測站點與臺灣全島各地的幾何距離單位公尺作為距離(方程式 3)。

$$u(x) = \frac{\sum_{j=1}^n w_j(x) u_j}{\sum_{j=1}^n w_j(x)} \quad \text{if } d(x, x_j) \neq 0$$

$$u(x) = u_j \quad \text{if } d(x, x_j) = 0$$

$$w_j(x) = \frac{1}{d(x, x_j)^p}$$



(方程式 3)

$u(x)$ 是 u 內插結果當位於 x 位置，本處是臺灣本島各地的溫度或雨量的內插值、 x 是內插結果所處在的位置、 j 是任意實際存在測量值的點位，本處是臺灣本島各地的氣象站點、 n 是實際存在測量值的總數量，本處是臺灣本島各地的氣象站點數共 595 點、 w_j 是任意實際存在測量值的點位與 x 內插結果所處在的位置之間生成的權重、 u_j 是任意實際存在測量值的點位 j 的測量值、 x_j 是任意實際存在測量值的點位 j 的位置、 d 函數代表任意兩位置的距離、 p 是兩反距離權重次方項，決定各點內插權重的影響性。

三、人口資料

人口密度採用中華民國開放統計資料全國人口統計資料村里界之最小統計區各年度全國人口統計資料，選取 2010 至 2015 年各年度 12 月結算人口數除以當地最小統計區面積之單位平方公里，獲得臺灣共六年度之各年度每一平方公里之人口數作為人口密度圖。

第四章 研究方法



第一節 研究架構

針對每日登革熱病例進行滾動預測，預測登革熱病例周圍可能傳染半徑構成的空間環域，與病例發病日未來最短與最長可能發病日構成發病間距，透過視覺化時空共同構成的時空範圍，時空範圍描述登革熱各別病例未來可能被傳染病例的傳染半徑範圍與發病日間隔。

但我們無法得知根據過去的哪些登革熱的預測模式進行改造，對於登革熱疫情期間的時空範圍的預測提供最佳的預測成果，我們發展四個時空範圍滾動預測模式，模式一透過預設固定的傳染半徑與發病間距，滾動預測登革熱病例周遭可能潛在的病例。模式二透過推論並建構登革熱疫情前期的傳播過程，運算潛在的傳染時空範圍參數，設置未來固定的傳染半徑與發病間距，滾動預測登革熱病例周遭可能潛在的病例。模式三透過推論並建構登革熱疫情前期的傳播過程，結合登革熱病例的傳播的時空範圍空間分布，運算潛在的傳染時空範圍參數，設置未來隨空間變化的傳染半徑與發病間距，滾動預測登革熱病例周遭可能潛在的病例。模式四透過推論並建構登革熱疫情前期的傳播過程，結合登革熱病例的傳播的時空範圍空間分布、氣候與人口背景資料，運算潛在的傳染時空範圍參數，設置未來隨空間、氣候與人口變化的傳染半徑與發病間距，滾動預測登革熱病例周遭可能潛在的病例。

本研究嘗試設計綜合指標，以每日滾動預測形式，長期的預測成果評估四種模式的表現。綜合指標包含涵蓋率(Coverage)與合適度(Fit)兩種概念。涵蓋率是評估不同時空範圍滾動預測模式，評估登革熱病例預測的時空擴散區是否足以包含未來的病例，另外一方面時空範圍的合適度評估時空範圍是否合乎未來疫情分布的範圍。

根據綜合指標成果挑選最佳的模式，作為長期滾動預測登革熱擴散的時空範圍基礎，發展空間視覺化模式，透過每日時序動態地圖，匡列登革熱每日未來擴散的潛在高風險區域，作為有效防堵登革熱疫情擴散的參考底本。



第二節 預測模式的資料集

本研究挑選登革熱擴散滾動預測模式，必須訓練模式並且測試模式的預測能力，因此我們根據防疫需求，並且考量登革熱疫情的特性，將每年度登革熱疫情期間分成登革熱疫情前期與疫情中後期。

疫情前期的概念來自於每一波完整疫情的第一個登革熱病例，需要觀察長達四個月約略 20 周期間的疫情季節週期性氣候發展，才可能斷定當年度疫情的嚴重性(Hii, Zhu, et al., 2012; Laureano-Rosario et al., 2018; Méndez-Lázaro et al., 2014)，完整疫情的第一個登革熱病例又名指標病例。

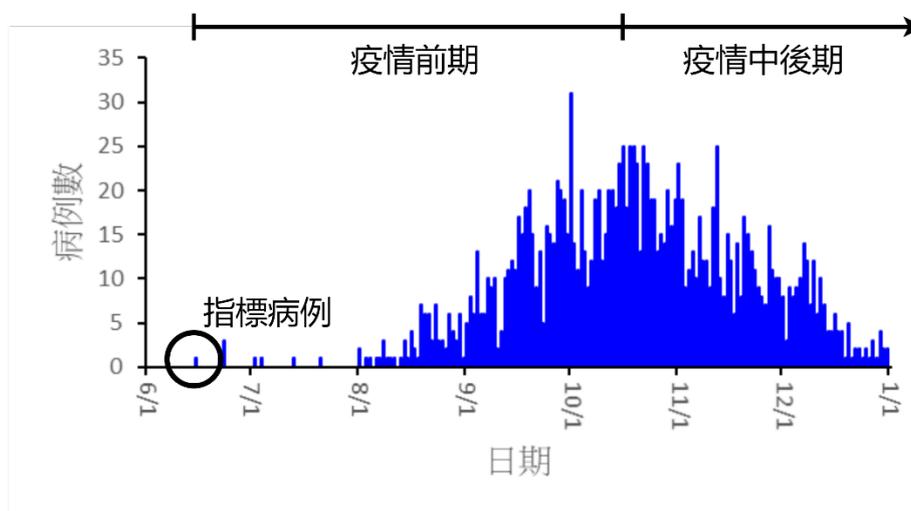


圖 6、2010 年度登革熱每日病例數分布圖。疫情前期與疫情中後期與指標病例示意圖。橫軸是日期，縱軸是每日累積病例數。

我們以圖 6 的 2010 年度登革熱每日病例數分布圖為例，當年度疫情首例指標病例發生於 6 月 15 日，往後計算四個月份 10 月 15 日，這段期間有明顯疫情升溫的跡象，由零星的每日累積病例數，至十月份每日登革熱疫情達到巔峰的現象，由此跡象表明四個月的疫情季節週期性氣候發展規律，因此我們定義由 6 月 15 日至 10 月 15 日為疫情前期，在當年度 10 月 15 日後每日累積病例數開始呈現長期下降的趨勢，因此定義為疫情中後期，由於季節性的影響與人為防疫的行為導致疫情趨緩或消失的情形。

延續先前的概念我們定義疫情前期的每日病例是訓練資料集，然而疫情高峰前後的疫情中後期為測試資料集，由於疫情前期我們需要觀察每日疫情的嚴重程

度與空間分布，在疫情高峰期後積極的指明未來登革熱病例發生的位置與發病日，未來登革熱的時空擴散範圍與趨勢，提供防堵登革熱疫情擴散的基礎。

本研究根據登革熱疫情非爆發年與爆發年，觀察指標病例的發病日作為基礎，以 2010 年為例，指標病例發生於該研究 6 月 15 日，為求登革熱非爆發年份之時間區段一致性。因此我們在非爆發年份之 2010 年至 2013 年採用 6 月 15 日至 10 月 15 日的病例資料作為訓練資料集，往後兩個月 10 月 16 日至 12 月 15 日作為測試資料集。另外 2014 年與 2015 年是登革熱疫情爆發年分，由於外在的氣候環境條件，導致登革熱的指標病例發病日提早至五月中旬的現象，因此我們在爆發年份之 2014 年至 2015 年採用 5 月 15 日至 9 月 15 日的病例資料作為訓練資料集，往後兩個月 9 月 16 日至 11 月 15 日作為測試資料集(表 3)。

表 3、定義訓練資料集與測試資料集時間區間。統整登革熱位於研究區各年度的指標病例發病日與嚴重程度，與資料集相對應的病例數。

年分	2010	2011	2012	2013	2014	2015
指標病例發病日	6/15	7/3	6/17	6/16	5/20	5/16
疫情嚴重程度	非爆發	非爆發	非爆發	非爆發	爆發	爆發
訓練資料集區間	6/15-10/15	6/15-10/15	6/15-10/15	6/15-10/15	5/15-9/15	5/15-9/15
訓練資料集病例數	700	360	722	28	2334	1196
測試資料集區間	10/16-12/15	10/16-12/15	10/16-12/15	10/16-12/15	9/16-11/15	9/16-11/15
測試資料集病例數	813	858	479	88	9593	22581

第三節 模式共同項目

一、資料集變數

此處本研究整理模式常見的共同函數與變數，方便敘述模組化的運算細節。在模組當中常見的是資料流，資料流提供模組登革熱病例與登革熱環境氣候相關的屬性資料，*TrD* 是訓練資料集；代表登革熱疫情前期的病例，*TD* 是測試資料集，代表登革熱疫情中後期的病例，之對應的病例與屬性背景資料。我們每日滾動預測每個病例的時空擴散區，然而在模式二至三需要透過訓練階段獲得預測時空擴散範圍的參數，因此我們採用疫情前期的病例重建模式並取得參數，這登革

熱病例資料通稱訓練資料集，在疫情中後期的部分我們稱為疫情中後期，評估模式的預測時空擴散範圍表現。

x 任意登革熱病例，定義 x 是討論單一登革熱病例對於各模組運算過程， x 任意登革熱病例存在各種屬性 $x.pos$ 是登革熱病例的發病位置、 $x.onset$ 是登革熱的發病日、 $x.imported$ 是否登革熱屬於境外移入的病例、 $x.env$ 時間氣候與環境因子屬性資料(圖 7)。

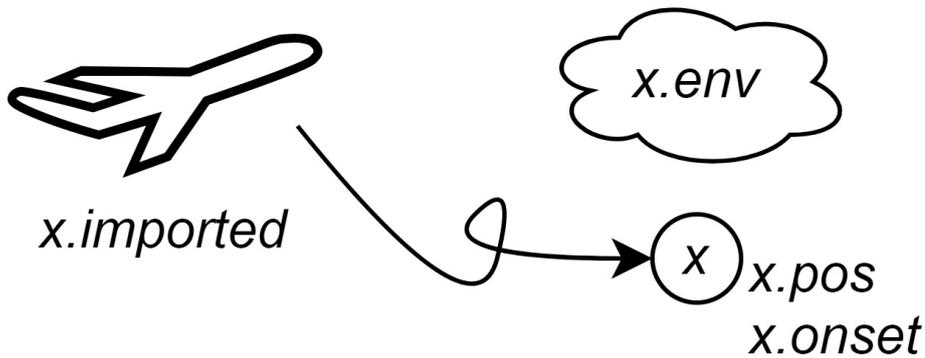


圖 7、登革熱病例 x 屬性示意圖。 x 任意登革熱病例、 $x.pos$ 是登革熱病例 x 的發病位置、 $x.onset$ 是登革熱病例 x 的發病日、 $x.imported$ 是否登革熱病例 x 屬於境外移入的病例、 $x.env$ 登革熱病例 x 時間氣候與環境因子屬性資料。

二、傳染關係項變數

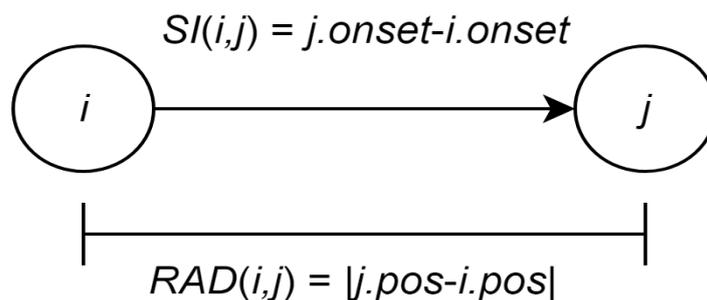


圖 8、登革熱病例傳染關係。 i 是傳染病例、 j 是被傳染病例、 $onset$ 發病日、 pos 發病位置、 RAD 傳染半徑函數、 SI 發病間距函數。

根據登革熱病例在疫情樹模組建構的傳遞鏈推估的傳染關係 i 是傳染病例 (Infector) 與 j 是被傳染病例 (Infectee)，傳染病例與被傳染病例構成的時空範圍，

空間半徑的運算採用傳染半徑 RAD 函數(方程式 4)，發病間距運算是發病間距 SI 函數(方程式 5)，兩個函數分別輸入 i 與 j 病例作為參數(圖 8)。

傳染半徑的運算模式 $RAD(i, j)$ ，登革熱病例 i 的發病位置經緯度座標與病例 j 的發病位置經緯度座標求取歐幾里得距離，採用的單位為公尺(方程式 4)。發病間距的運算方式 $SI(i, j)$ ，將登革熱病例 j 的發病日與登革熱病例 i 發病日之間的時間差，以日為單位表示(方程式 5)。

傳染半徑:

$$RAD(i, j) = |j.pos - i.pos| \text{ (單位:公尺)} \quad \text{(方程式 4)}$$

發病間距:

$$SI(i, j) = j.onset - i.onset \text{ (單位:日)} \quad \text{(方程式 5)}$$

三、時空擴散區

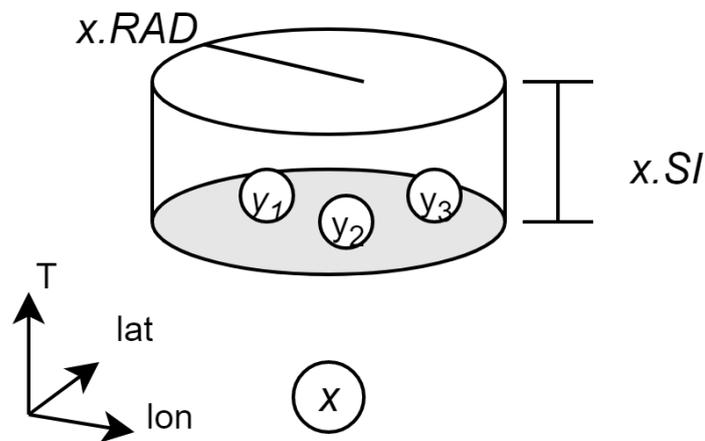


圖 9、登革熱時空擴散區。 x 是登革熱病例、 $x.RAD$ 傳染半徑、 $x.SI$ 發病間距、 y_1 到 y_3 時空擴散區涵蓋未來的病例、 lat 緯度軸、 lon 經度軸、 T 時間軸。

表示預測時空範圍的成果的變數項目 $x.RAD$ 是任意登革熱病例 x 的預測傳染半徑， $x.SI$ 是任意登革熱病例 x 的預測發病間距(圖 9)，所有登革熱病例的時空擴散範圍構成的時空擴散區的時空擴散範圍。

c 時空擴散區內的病例數，指的是我們預測的時空擴散區，在未來發生的病例，它的發病日與發病位置剛好落在我們時空擴散區內，因此我們計數多少病例



落在時空擴散區內，例如圖 9 為例， y_1 到 y_3 是時空擴散區涵蓋未來的病例，因此我們的 c 等於 3 個病例。

而時空擴散範圍的定義在研究區某日存在 n 個病例總數，任意 x 病例有傳染半徑 $x.RAD$ 與發病間距 $x.SI$ 構成 at_x ，運算 $x.RAD$ 與 $x.SI$ 都需要加絕對值，原因是傳染半徑與發病間距是區間的概念，因此絕對值代表區間上界減去區間下界， $x.RAD$ 傳染半徑從病例 x 周遭 0 公尺至最遠的傳染半徑， $x.SI$ 發病間距是新病例最短發病日與最長發病日的區間(方程式 6)。

時空擴散範圍的概念是所有的 1 至 n 病例數，若採用原始時空擴散範圍加總所有 at_i 第 i 個病例的時空擴散範圍，取得每日滾動預測的原始時空擴散範圍(方程式 7)，我們的時空解析度傳染半徑是公尺，發病間距是日，因此時空擴散範圍原始單位應該是公尺平方日($m^2 \text{ day}$)，然而為了方便觀察可能會轉換單位成公里平方日($km^2 \text{ day}$)。

若採用原始時空擴散範圍，會面臨當日各別病例重複計算時空擴散範圍空間重疊部分，原因出自於登革熱疫情具有群聚傳染的現象，當日各別病例發病位置在空間分布十分相近，因此本研究應該採用聯集時空擴散範圍的定義，發展運算時空擴散範圍的演算法，去除時空擴散範圍重疊部分，在聯集時空擴散範圍採用聯集所有 at_i 第 i 個病例的時空擴散範圍，經過修正得到聯集時空擴散範圍，是本研究時空擴散範圍的完整定義(方程式 8)。

$$\begin{aligned} \text{Let } x &= \forall i, i = 1 \sim n \\ at_x &= |x.RAD|^2 \pi |x.SI| \end{aligned} \tag{方程式 6}$$

原始時空擴散範圍:

$$\widetilde{AT} = \sum_{i=1}^n at_i \tag{方程式 7}$$

聯集時空擴散範圍:

$$AT = \bigcup_{i=1}^n at_i \tag{方程式 8}$$

at_i 第 i 個病例的時空擴散範圍、 $i.RAD$ 是 i 病例的預測傳染半徑、 $i.SI$ 是 i 病例的預測發病間距、 \overline{AT} 原始時空擴散範圍、 AT 聯集時空擴散範圍、 n 病例總數。

以上敘述常見函數與變數整理於表 4，來自所有模式共同的資料集變數、傳染變數與時空擴散範圍定義。

表 4、模式變數代號統整表。

代號	定義
TrD	訓練資料集
TD	測試資料集
x	任意的登革熱病例
$x.pos$	任意的登革熱病例 x 的發病位置經緯度座標
$x.onset$	任意的登革熱病例 x 的發病日
$x.imported$	任意的登革熱病例 x 境外移入病例
$x.env$	任意病例的環境與氣候因子屬性資料向量
i	登革熱傳染病例(Infector)
j	登革熱被傳染病例(Infectee)
$RAD(i,j)$	函數運算登革熱傳染病例 i 與登革熱被傳染病例 j 的傳染半徑 (Transmission radius)，計算公式: $ j.pos - i.pos $ (單位:公尺)
$SI(i,j)$	函數運算登革熱傳染病例 i 與登革熱被傳染病例 j 的發病間距 (Serial interval)，計算公式: $j.onset - i.onset$ (單位:日)
$x.RAD$	任意的登革熱病例 x 的預測傳染半徑， RAD 是傳染半徑的代號
$x.SI$	任意的登革熱病例 x 的預測發病間距， SI 發病間距的代號
AT	聯集時空擴散範圍，簡稱時空擴散範圍，原始單位公尺平方日 ($m^2 \text{ day}$)
c	時空擴散區內(涵蓋)的病例數

第四節 模式一

模式一是不實行任何訓練模式，我們預設登革熱病例 x 的預測傳染半徑是 0 至 300 公尺內，預測發病間距是 6 至 23 日，繪製各別病例的時空擴散範圍，生成每日的時空擴散區(圖 10)，根據文獻推論外生潛伏期與內生潛伏期構成的發病間距，可能介於 6 至 23 日產生新的登革熱病例，若傳染半徑設置 300 公尺以內可以獲得較佳的中心性(Kuo, Wen, and Sabel, 2018)，有助於重建登革熱的擴散過程，以相同研究區為例，預設固定的時空邊界試圖捕捉登革熱的擴散過程。

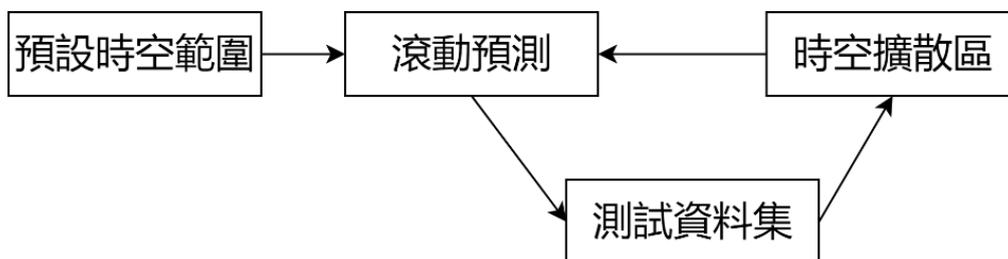


圖 10、模式一流程圖。模式一僅有預設時空範圍階段，與滾動預測生成時空擴散區的流程。

一、預設時空範圍

模式一採用預設時空範圍的方法，任意病例 x 可能屬於訓練資料集或測試資料集，設定任意病例 x 的預測傳染半徑介於離病例的發病位置周圍 0 至 300 公尺範圍，另外一方面發病間距根據文獻設置 6 至 23 日，以此構成時空擴散區，生成時空擴散範圍，預測未來的登革熱病例的發病位置與發病日(方程式 9)。

$$\begin{aligned}x &\in TrD \vee TD \\x.RAD &= [0,300] \\x.SI &= [6,23]\end{aligned}\tag{方程式 9}$$

二、滾動預測循環

(1) 滾動預測

滾動預測過程根據模式，每日的登革熱病例，透過模式預設的時空擴散範圍，繪製各別病例的時空擴散範圍，構成當日整體的時空擴散區，預測未來的登革熱病例的發病位置與發病日。



(2) 測試資料集

提供每日新的登革熱病例，給予模式每日作滾動預測的過程，病例的發病位置發病日與病例相關的背景資料，以利模式進行模擬每日的滾動預測流程。

(3) 時空擴散區

時空擴散區的生成藉由每日滾動預測的過程，當日病例產生的時空擴散範圍，根據聯集時空擴散範圍的概念，劃設登革熱的時空擴散區，嘗試捕捉未來的疫情擴散動向，同時回復至滾動預測過程，等待隔日新的病例資料進入模式中預測時空擴散區。

第五節 模式二

模式二是從疫情前期的訓練資料集重建疫情樹取得傳遞鏈集合，進行時空擴散範圍的彙整，分別包含傳染半徑的直方圖與發病間距的直方圖，藉由設定門檻值的方法，尋找大多數登革熱疫病例的傳染半徑與發病間距，預設固定的時空擴散範圍，包含未來病例的預測傳染半徑與預測發病間距，並進入滾動預測的循環(圖 11)。

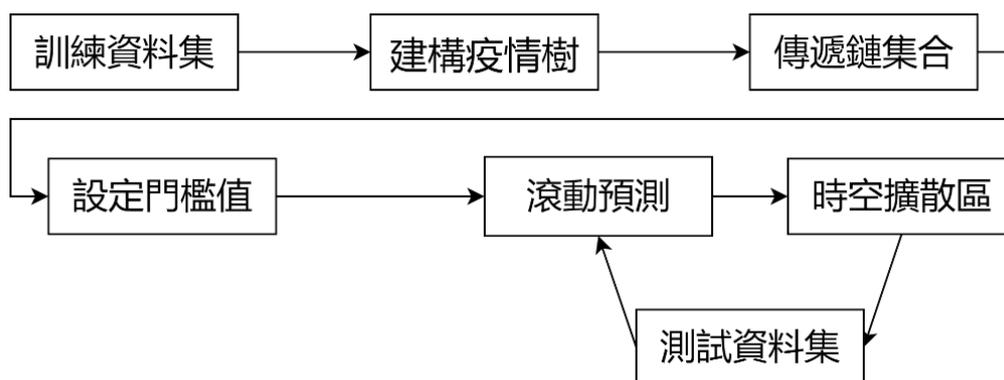


圖 11、模式二流程圖。模式二比起模式一複雜許多，必須設置訓練階段而採用訓練資料集，並建構疫情樹，產生傳遞鏈集合，並且歸納設置門檻值與滾動預測生成時空擴散區的流程。



一、建構疫情樹

從疫情前期的訓練資料集 TrD 從中選出所有的登革熱病例，透過 MST-DBSCAN 嘗試建構可能傳染關係(方程式 10)，並改寫成 i 傳染病例與未來發病點 j 被傳染病例(圖 12)，構成 i 與 j 的傳染關係必須滿足三項條件:

第一項條件 i 傳染病例與未來發病點 j 被傳染病例， i 傳染病例發病日必須早於 j 被傳染病例的發病日，傳染需要透過病媒蚊作為傳染媒介與人類宿主(Adams and Kapan, 2009)，傳染者與被傳染者不會同時發病。

第二項條件 i 與 j 病例是可能的傳染關係， j 必定不是境外移入，由於 j 境外移入病例病毒來源明確，無法透過境內傳染病例 i 推論 j 病例由 i 所傳染，境外移入病例只可能扮演 i 傳染者，由於臺灣的登革熱受到境外移入病例，導致兩日至兩個月內的本土疫情發生變化，境外移入病例扮演病毒帶原者(Pan et al., 2020)，。

第三項條件 i 傳染病例在有限的時空半徑內傳染給 j 被傳染病例， j 超過 i 傳染病例的有限時空半徑邊界，則 i 與 j 無法構成傳染關係。

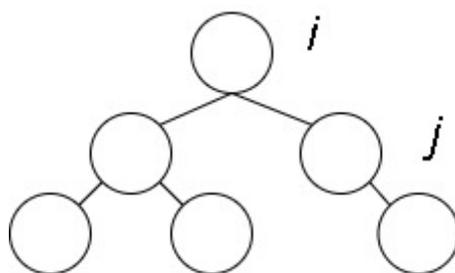


圖 12、疫情樹。病例是空白圓圈，推估傳染用連結表示。 i 與 j 分別代表傳染者與被傳染者。

$$MST - DBSCAN(\forall(i, j) \in TrD)$$

(方程式 10)

MST - DBSCAN該建構疫情樹的演算法。

有限的時空範圍中預設寬鬆的空間範圍，空間範圍即可能的傳染半徑，登革熱的區域擴散範圍受到病媒蚊的移動所影響，本研究預設傳染半徑邊界是 0 至 1000 公尺為登革熱擴散的傳染半徑的邊界，1 公里是文獻建議的傳染半徑上界，

多數被感染病例會落在距離傳染病例的環域空間半徑(Guzzetta et al., 2018; Kan et al., 2008)。

有限的時空範圍預設寬鬆的時間間距，時間間距即可能的發病間距，本研究預設發病間距邊界是 6 至 50 日，發病間距來自於潛伏期(Incubation period)與傳染期(Infectious period)，潛伏期分成內生潛伏期(Extrinsic incubation period)與外生潛伏期(Intrinsic incubation period)，分別用 EIP 與 IIP 作為代號，傳染期的代號是 $Infectious$ 。登革熱最短外生潛伏期需要 2.5 日(Chan and Johansson, 2012)，最短內生潛伏期為 3.5 日(Nishiura and Halstead, 2007)(圖 13)， t_0 是傳染病例 i 發病日是第 0 日發生病媒蚊叮咬事件， t_1 是登革熱病毒在病媒蚊體液感染得外生潛伏期(方程式 11)， t_2 是被傳染病例 j 的發病日需要經過 t_1 的外生潛伏期後被病媒蚊叮咬在被傳染病例 j 經過內生潛伏期發展至具有傳染力後經診斷確診，因此 $t_2 - t_0$ 可以代表最短的發病間距(方程式 12)，經過計算帶入內生潛伏期與外生潛伏期發病後 6 日 i 病例可能會傳染給 j 病例(方程式 13)。最長外生潛伏期 33 日(Chan and Johansson, 2012)與最長內生潛伏期 10 日(Chan and Johansson, 2012; WHO, 2009a)，登革熱在發病日前兩天和發病含當日後五天共 7 日具有傳染性(Matangkasombut et al., 2020; Nishiura and Halstead, 2007)(圖 14)， t_0 是傳染病例 i 發病確診日， t_1 是登革熱病例具有感染力的時距，假設病媒蚊在 i 病例具有傳染力的最後一刻叮咬，則 t_1 等於具有傳染力的時距， t_2 是登革熱病毒在病媒蚊體液感染得外生潛伏期加上 t_1 ， t_3 是被傳染病例 j 的發病日需要經過 t_2 的外生潛伏期後被病媒蚊叮咬在被傳染病例 j 經過內生潛伏期發展至具有傳染力後經診斷確診(方程式 14)，因此 $t_3 - t_0$ 可以代表最長的發病間距(方程式 15)，經過計算帶入傳染期、內生潛伏期與外生潛伏期發病後 50 日內 i 病例可能會傳染給 j 病例(方程式 16)。以上潛伏期統整最短傳染情境是傳染者發病日經過 6 日以後被傳染者發病，最長傳染情境是傳染者發病日經過 50 日以內被傳染者發病(方程式 18)。

最短期傳染情境:

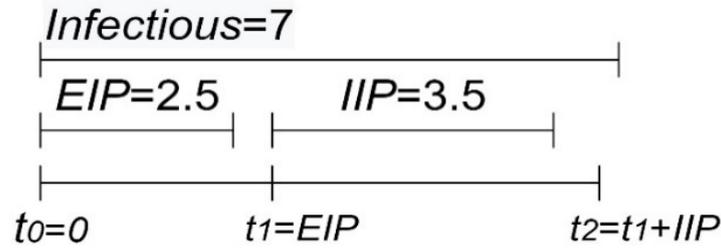


圖 13、最短潛伏期估計。外生潛伏期 (Extrinsic incubation period, EIP) 與內生潛伏期 IIP (Intrinsic incubation period, IIP) 根據文獻估計的最短天數， t_0 至 t_2 分別為登革熱傳染者具傳染性受媒介叮咬時間、被傳染者被媒介叮咬時間與被傳染者發病時間。

$$t_0 = 0, t_1 = EIP, t_2 = t_1 + IIP \quad (\text{方程式 11})$$

$$t_2 - t_0 \geq EIP + IIP - 0 \quad (\text{方程式 12})$$

$$EIP = 2.5, IIP = 3.5 \Rightarrow t_2 - t_0 \geq 6 \quad (\text{方程式 13})$$

t_0 登革熱傳染者具傳染性受媒介叮咬時間、 t_1 被傳染者被媒介叮咬時間、 t_2 被傳染者發病時間、 EIP 外生潛伏期、 IIP 內生潛伏期。

最長期傳染情境:



圖 14、最長潛伏期估計。外生潛伏期 (Extrinsic incubation period; EIP) 與內生潛伏期 IIP (Intrinsic incubation period, IIP) 根據文獻估計的最長天數， t_0 至 t_3 分別為登革熱傳染者免疫前受媒介叮咬、登革熱傳染者發病、被傳染者被叮咬時間與被感染者發病時間。

$$t_0 = 0, t_1 = \text{Infectious}, t_2 = t_1 + \text{EIP}, t_3 = t_2 + \text{IIP}$$

(方程式 14)

$$t_3 - t_0 \leq \text{Infectious} + \text{EIP} + \text{IIP} - 0$$

(方程式 15)

$$\text{Infectious} = 7, \text{EIP} = 33, \text{IIP} = 10 \Rightarrow t_3 - t_0 \leq 50$$

(方程式 16)

t_0 登革熱傳染者免疫前受媒介叮咬、 t_1 登革熱傳染者發病、 t_2 被傳染者被叮咬時間、 t_3 被感染者發病時間、*EIP* 外生潛伏期、*IIP* 內生潛伏期。

$$\text{RadBound} = \{ 0 \text{ meter} \leq |j.\text{pos} - i.\text{pos}| \leq 1000 \text{ meters} \}$$

(方程式 17)

$$\text{SIBound} = \{ 6 \text{ days} \leq j.\text{onset} - i.\text{onset} \leq 50 \text{ days} \}$$

(方程式 18)

RadBound 傳染半徑邊界、*SIBound* 發病間距半徑邊界。

經討論統整登革熱傳染的寬鬆化時空範圍邊界，傳染半徑邊界 *RadBound* 是介於 0 至 1000 公尺(方程式 17)，是 j 發病位置 $j.\text{pos}$ 與 i 發病位置 $i.\text{pos}$ 的幾何距離；發病間距半徑邊界 *SIBound* 是介於 6 至 50 日(方程式 18)，是 j 發病日 $j.\text{onset}$ 與 i 發病日 $i.\text{onset}$ 的發病日差，MST-DBSCAN 的演算法透過 i 病例作為中心觀察時空半徑閉集合內的所有候選病例 j ，傳染半徑屬於 *RadBound* 聯集發病間距屬於 *SIBound*，聯集候選病例 j 不是境外移入條件，以上條件本研究稱為時空範圍閉集合，推論 j 是 i 的鄰近且 j 、 i 因為鄰近關係 j 、 i 是屬於某傳遞鏈因此歸化為未分類標籤，未分類的標籤是 NOTCLASSIFY(方程式 19)，本模組鄰近關係的標籤單數是 Neighbor 或複數是 Neighbors，MST-DBSCAN 是指 i 病例可能傳染登革熱給 j 病例的表示方法， j 病例可能是多個採用複數；若單個 j 病例則是單數。



checkNearPoints:

$$|j.pos - i.pos| \in RadBound \cap j.onset - i.onset \in SIBound \cap j \neq imported$$

$$\Rightarrow j \text{ is the neighbor of } i, i \leftarrow NOTCLASSIFY, j \leftarrow NOTCLASSIFY$$

(方程式 19)

imported 境外移入、*neighbor* 臨近性是借用 MST-DBSCAN 的定義，也可解釋為易達性，在疫情樹代表兩個點(病例)存在傳染關係、*NOTCLASSIFY* 是 DBSCAN 的演算法標籤，在疫情樹代表屬於未知傳遞鏈。

根據病例 *i* 觀察時空範圍閉集合的所有鄰近病例 *j* 的數量，所有鄰近病例 *j* 的數量可以藉由 *CountNeighbors* 函數取得，根據中心性門檻 *minPts* 的定義根據有效再生率大於 1 作為前提(Codeço, Villela, and Coelho, 2018)，所有未分類病例依照 *minPts* 分析，鄰近點大於 *minPts* 門檻視為未分類點；反之鄰近點小於 *minPts* 門檻視為雜訊點，中心性門檻不足無法確認傳染關係，未分類點 *NOTCLASSIFY* 與雜訊點 *NOISE* 在 DBSCAN 演算法代表的是點，但在本研究主題代表的是登革熱病例，未分類點在登革熱病例是屬於某個傳染或被傳染者，雜訊點是無法找到病例傳染與被傳染者(方程式 20)。

向下處理未分類點，已知未分類點在登革熱病例是屬於某個傳染或被傳染者(方程式 21)，根據病例發病日時間先後預選傳染者 *i* 病例，*i* 病例向後找出符合時空半徑閉集合的所有病例 *j*，且病例 *j* 也是未分類點，透過 *connect* 函數重建 *i* 病例與 *j* 病例的傳染與被傳染關係(方程式 22)，這時原本的所有的 *j* 病例 $\forall j$ 病例被傳染者病例變成傳染者病例的角色 *i* 傳染者，候選排隊進行深度搜尋法(Deep first search)的規則，此時所有的 *j* 病例 $\forall j$ 依然是任選符合時空範圍閉集合的所有候選病例進行連結(方程式 22)，不斷遞迴操作此程序直到所有分類點變成分類點或雜訊點，以上所有登革熱病例依造時空半徑閉集合與中心性門檻推估出可能的連結與傳染關係，以上疫情樹虛擬碼參考附錄三之一疫情樹採用 MST-DBSCAN 演算法。

Deep first search rule:

$$minPts = 1$$

$$CountNeighbors(i) > minPts: i \leftarrow CLASSIFY$$

$$CountNeighbors(i) \leq minPts: i \leftarrow NOISE$$

$$i = \text{NOTCLASSIFY} \ \& \ j = \text{NOTCLASSIFY} \Rightarrow \text{connect } i, j \quad (\text{方程式 20})$$

$$i \leftarrow \forall j, \forall j \leftarrow \text{Neighbors of } j \text{ as candidate run } dfs(j) \quad (\text{方程式 21})$$

$$\text{Epidemic tree} \leftarrow \text{connected } dfs(\forall i, j) \quad (\text{方程式 22})$$

CountNeighbors 函數用於說明 *i* 病例有多少鄰近，在疫情樹說明 *i* 病例可以傳給幾個人、*CLASSIFY* 是 DBSCAN 的演算法標籤，在疫情樹代表有傳染力的病例、*NOISE* 是 DBSCAN 的演算法標籤，在疫情樹代表沒傳染力的病例、*Neighbors* 是臨近性是借用 MST-DBSCAN 的定義，也可解釋為易達性，在疫情樹代表多個點(病例)與 *i* 點(病例)存在傳染關係，*Epidemic tree* 疫情樹是 DBSCAN 演算法完成深度搜尋後的承接變數。

二、傳遞鏈集合

傳遞鏈集合流程擷取來自疫情樹的所有傳遞鏈，以代號 $\forall(i, j) \in \text{Epidemic tree}$ 表示，*Hist_{RAD}* 傳染半徑的直方圖由所有傳遞鏈的傳染半徑輸入生成直方圖，每個直方圖記錄所有傳遞鏈的傳染半徑，以 10 公尺作為區間，由 0 至 1000 公尺的值域(方程式 23)，*Hist_{SI}* 發病間距的直方圖是紀錄所有傳遞鏈的發病間距以 1 日作為區間，記錄所有傳遞鏈的發病間距，由 0 日至 50 日值域(方程式 24)，透過訓練資料集之疫情前期，記錄所有疫情樹重建傳遞鏈的時空範圍。

$$\text{Hist}_{RAD} \leftarrow RAD(\forall(i, j) \in \text{Epidemic tree}) \quad (\text{方程式 23})$$

$$\text{Hist}_{SI} = SI(\forall(i, j) \in \text{Epidemic tree}) \quad (\text{方程式 24})$$

Hist_{RAD} 傳染半徑的直方圖、*Hist_{SI}* 發病間距的直方圖、*Epidemic tree* 疫情樹是 DBSCAN 演算法完成深度搜尋後的承接變數、*RAD* 函數是擷取所有疫情樹的傳遞鏈的傳染半徑、*SI* 函數是擷取所有疫情樹的傳遞鏈的發病間距。



三、設定門檻值

(1) 基本性質介紹

門檻的取得是藉由模式時空擴散區的病例數與時空擴散範圍找到適當的預估涵蓋率 ec ，門檻的基本概念是修正模式的缺陷，模式在估計時空擴散區放大時空範圍的過程為了預測更多病例數量，導致高估群聚擴散源的登革熱時空擴散範圍，被時空擴散區的包含的病例不屬於此時空擴散區的群聚傳染源。大多數屬於該時空擴散區的登革熱病例會落在發病源有限的時空擴散範圍內，少數距離傳染源時空距離遙遠的病例，本研究推論不屬於該時空擴散區，如果納入少數距離傳染源時空距離遙遠的病例會導致時空擴散區匡列過大的現象，預測不屬於該群聚傳染源的病例。

因此設計門檻的目的，讓模式在有限的時空擴散範圍預測多數登革熱病例的發病位置與發病日，並且在有限的時空擴散範圍匡列高風險區，避免匡列過大的登革熱時空擴散區，造成高估登革熱的高風險區。

門檻值的設置是根據預估涵蓋率 ec ，預估涵蓋率對應的病例數關係必須透過模式先行訓練的結果，觀察模式預估涵蓋率與時空擴散區的病例數、時空擴散區內的病例數與時空擴散範圍，模式時空擴散區的病例數與密度的關係，然而本處發現模式時空擴散區的病例數與密度數值變化不明顯，且時空擴散區的病例數與密度之間的關係不易運算，因此本研究採取密度倒數的概念，時空擴散範圍對病例數進行一階導數，以系統效能評估模式的等距格子法設定範圍從 0.1 至 1 間距 0.05 的預估涵蓋率，觀察在每個預估涵蓋率時空擴散範圍對病例數一階導數的成長趨勢(方程式 25)。

Let $ec=[0,1]$

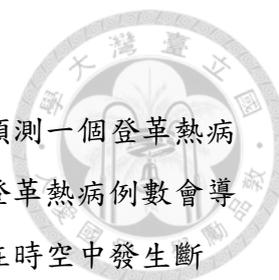
$x \in TrD$

for ec plus 0.05 $x.RAD, x.SI \leftarrow cdf^{-1}(Hist_{RAD}, ec), cdf^{-1}(Hist_{SI}, ec)$

$$AT = \bigcup_{i=1}^n |x.RAD|^2 \pi |x.SI|$$

(方程式 25)

ec 預估涵蓋率、 $Hist_{RAD}$ 傳染半徑的直方圖、 $Hist_{SI}$ 發病間距的直方圖。



(2) 有效與無效預測

時空擴散範圍對病例數一階導數的數值，代表模式每增加預測一個登革熱病例數會導致時空擴散範圍成長的趨勢，萬一發生增加預測一個登革熱病例數會導致時空擴散範圍成長非常嚴重的現象，代表登革熱群聚擴散區在時空中發生斷點，這個斷點代表該群聚傳染的時空擴散區的邊界，本就定義這個邊界以內是時空擴散區的有效預測，代表模式預測該群聚傳染非常有可能發生在該時空擴散區內，反之時空擴散區的邊界或以外的時空擴散範圍，定義時空擴散區的無效預測範圍。

(3) 門檻的數學性質

根據密度定義，密度下界之分母是時空擴散範圍，等距設定預估涵蓋率，觀察模式時空擴散區的病例數與時空擴散範圍，任何完美的模式沒有系統性雜訊，時空擴散範圍因為隨機雜訊導致指數成長的現象， p 的成長必定與 x 同時會呈現指數放大的現象，因此各模式必然存在時空擴散範圍的過度放大現象，為了避免 p 、 x 、 δ 過度放大導致時空擴散範圍高估的現象，本研究採取預估涵蓋率的法則，所以定義 q 等於 0；又 e^p 遠大於 $1/e^q$ 因此不需要考慮 $1/e^q$ (方程式 26)。

$$\begin{aligned} range_i &= [0, ec] = [q, p] \\ \left| \left(e^p - \frac{1}{e^q} \right) e^{x+\delta} \right|, e^p \gg \frac{1}{e^q}, \text{ then } |(e^p)e^{x+\delta}| \end{aligned} \tag{方程式 26}$$

$range_i$ 預估涵蓋率範圍再任意點 i 、 i 在此方程式是可變預估涵蓋率點、 ec 預估涵蓋率，本研究設定預估涵蓋率上界等於預估涵蓋率，預估涵蓋率下界等於 0，因此預估涵蓋率範圍 $range_i$ 是 0 至 ec 的封閉區間、 p 預估涵蓋率上界、 q 預估涵蓋率下界、 x 病例的時空擴散半徑、 δ 隨機性誤差。

定義時空擴散範圍對病例數的一階導數 $\partial AT / \partial c$ 作為衡量離群值的指標發生在時空擴散範圍內任意可變區間 $range_i$ (方程式 26)，代表每增加一革熱病例數時空擴散範圍的成長幅度，並且依造成長幅度分成兩群，第一群為有效預測，另一群為無效預測。最大時空擴散範圍對病例數一階導數是標準斷點特徵，斷點之後的最大時空擴散範圍對病例數一階導數必定是無效預測。

根據時空擴散範圍對病例數一階導數是標準斷點特徵，向之前推論趨勢類似

標準斷點特徵時空擴散範圍對病例數一階導數的進行非監督的群聚，本研究根據標準斷點特徵進行時空擴散範圍對病例數二階導數，並向前迭代透過時空擴散範圍對病例數二階導數，直到找不到類似標準斷點特徵的時空擴散範圍對病例數二階導數作為有效預測點，這一點病例數與時空擴散範圍對應的預估涵蓋率定義為門檻。

方程式 27 是門檻等於 $range_k$ 化簡過後的基本式，其中 $\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c}(range_i)$ 代表在小於門檻特徵，反之 $\sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i)$ 是大於門檻特徵，門檻的獲得主要有兩個重要變數 $range_{max}$ 與 $range_k$ ，其中 $range_{max}$ 是斷點特徵(方程式 28)；而 $range_k$ 是尋找門檻遞迴的過程找到的門檻。不等式說明門檻值帶入時空擴散範圍對病例數一階導數會小於有效預測與無效預測的時空擴散範圍對病例數二次積分，小於關係說明涵蓋病例數與時空擴散範圍在門檻值發生時時空擴散範圍對病例數一階導數曲線有下凹現象，因此從 $range_k$ 之前的趨勢互斥 $range_{max}$ 的趨勢，方程式 29 加強說明門檻具有明顯區分有效預測與無效預測的重要特徵，其他關於門檻的數學性質證明，請參考附錄五。

$$\frac{\partial AT}{\partial c}(range_i) = AT' \quad (\text{方程式 27})$$

$$\max\left(\frac{\partial AT}{\partial c}(range_{max})\right) = AT'_{max} \quad (\text{方程式 28})$$

$$AT^{-1} \left(\int \int \frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2}(range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i) \right) \partial c^2 \right) \geq range_k = p - 0 = threshold \quad (\text{方程式 29})$$

$range_i$ 預估涵蓋率範圍在任意點 i 、 $range_{max}$ 是斷點特徵的預估涵蓋率點、 AT 是時空擴散範圍、 AT' 是時空擴散範圍對病例數的一階導數、 AT'_{max} 時空擴散範圍對病例數的一階導數發生最大值的現象、 k 是門檻特徵的預估涵蓋率點、 $threshold$ 是門檻特徵等於預估涵蓋率上界 p 。



(4) 取得門檻的難處與替代方案

取得門檻的困難在於預測未來發生的病例，無法得知未來的病例數，再加上登革熱每日的疫情變化多樣，無法透過短時間得知病例數與時空擴散範圍的趨勢，因此本研究提供以下替代方案。

當年度登革熱發生疫情當下，已經存在過去群聚傳染的特徵，可以應用於訓練各種登革熱時空散模式，透過這些登革熱時空擴散模式訓練且驗證的結果觀察病例數對時空擴散範圍的趨勢，透過尋找門檻的演算法進行施作獲得門檻，以此門檻作為預測未來登革熱時空擴散區的設置門檻。以本研究作法為例採取訓練資料集驗證過後運算的門檻值，直接應用在測試階段做每日滾動預測設定的門檻，作為模式預測未來登革熱時空擴散區的預估涵蓋率參數。

四、滾動預測循環

我們在傳遞鍵集合階段取得 $Hist_{RAD}$ 與 $Hist_{SI}$ ，分別是傳染半徑與發病間距的直方圖，又在設定門檻值階段取得 ec 預估涵蓋率，定義 q 預估涵蓋率下界、 p 預估涵蓋率上界，預設 ec 預估涵蓋率，讓 p 等於 ec 且讓 q 等於0，會得到 $range$ 介於0至 ec 的預估涵蓋間距(方程式 30)，我們再從設定門檻值的程序取得門檻值不等式，我們運算門檻值 $threshold$ 後帶入模式中(方程式 31)。

求取任意 x 病例，來自於測試資料集，病例的傳染半徑 $x.RAD$ 與 x 病例的發病間距 $x.SI$ ，透過帶入 $range$ 的 x 病例的傳染半徑直方圖 $Hist_{RAD}$ 的累積機率密度的反函數與帶入 $range$ 的 x 病例的傳染半徑直方圖 $Hist_{SI}$ 的累積機率密度的反函數，得到 x 病例的傳染半徑與發病間距(方程式 32)。

$$\begin{aligned}
0 < q < p \leq 1, p - q &= range \\
let p = ec, q &= 0 \\
range &= [0, ec]
\end{aligned}$$

(方程式 30)

$$range \leftarrow threshold = AT^{-1} \left(\int \int \frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2}(range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i) \right) \partial c^2 \right)$$

(方程式 31)

$$x \in TD$$

$$x.RAD, x.SI \leftarrow cdf^{-1}(Hist_{RAD}, range), cdf^{-1}(Hist_{SI}, range) \quad (\text{方程式 } 32)$$

q 預估涵蓋率下界、 p 預估涵蓋率上界、 ec 預估涵蓋率、 $range$ 預估涵蓋間距、 $x.RAD$ 是 x 病例的傳染半徑、 $x.SI$ 是 x 病例的發病間距、 cdf^{-1} 累積機率密度的反函數。

傳染半徑與發病間距的直方圖來自於登革熱疫情前期，訓練資料集由疫情樹構成的傳遞鏈，因此所有滾動預測的預測傳染半徑與預測發病間距，每個病例在當年度的滾動預測結果是固定的傳染半徑與發病間距。

第六節 模式三

模式三是實行訓練模式，從疫情前期的訓練資料集重建疫情樹取得傳遞鏈集合，透過自組織類神經網路進行時空擴散範圍的分類，分類各別病例發病位置座標，根據各別登革熱病例的發病位置，預測屬於該登革熱病例獨特的傳染半徑與發病間距，生成登革熱的時空擴散區(圖 15)。

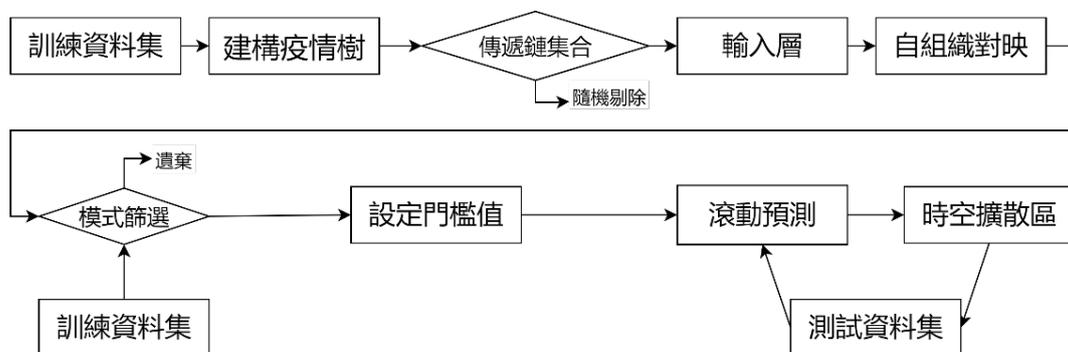


圖 15、模式三流程圖。模式三比起模式一複雜，必須設置訓練階段而採用訓練資料集，並建構疫情樹，之後生成傳遞鏈集合，又必須經過隨機剔除過程，放入傳遞鏈與病例位置作為輸入層，後輸入訓練資料集進行模式驗證與篩選，選擇最佳模式，設定門檻值，才能進入滾動預測循環。



一、建構疫情樹

承襲模式二流程，從疫情前期的訓練資料集 TrD 從中選出所有的登革熱病例，透過 MST-DBSCAN 嘗試建構可能傳染關係，生成疫情樹傳遞鏈(方程式 33)。

$$MST - DBSCAN(\forall(i, j) \in TrD) \quad (\text{方程式 } 33)$$

二、自組織對映作業流程

(1) 傳遞鏈集合

從傳遞鏈集合隨機剔除 20% 的傳遞鏈製作成輸入層向量，本模組為了減少運算量，以單一被傳染者可以傳染給多個被傳染者為前提，僅選取傳染半徑最長、中位數與最短的傳遞鏈組合，又發病間距最長、中位數與最短的組合，單一傳染者最多共六組傳遞鏈，所有選取出來的傳遞鏈集合隨機剔除 20% 的傳遞鏈後生成訓練資料集，根據隨機性的剔除 20% 的傳遞鏈，預設產生 10 組獨立的訓練資料集，這 10 組訓練資料集都獨立完成以下的自組織對映訓練(方程式 34)。

$$trc' = \text{Random truncate } 20\% (\forall(i, j) \in \text{Epidemic tree}) \quad (\text{方程式 } 34)$$

$Epidemic tree$ 疫情樹、 trc' 根據隨機性的剔除 20% 的傳遞鏈產生的任一訓練資料集。

(2) 輸入層

從疫情樹輸出的疫情樹的連結中擷取傳遞鏈集合 trc' ，輸入層向量以所有傳遞鏈 u 的傳染半徑、發病間距、傳遞鏈中傳染病例 i 的度與傳遞鏈中傳染病例 i 的緯度放置輸入層(方程式 35)，整體而言傳遞鏈集合結合各個被傳染病例的位置自組織對映的輸入層(圖 16)。

$$\begin{aligned} & \text{foreach } u = trc' \\ & \text{InputLayer} \leftarrow [RAD(u), SI(u), u.i.lon, u.i.lat] \end{aligned} \quad (\text{方程式 } 35)$$

trc' 根據隨機性的剔除 20% 的傳遞鏈產生的任一訓練資料集、InputLayer 輸入層、 RAD 函數是擷取所有疫情樹的傳遞鏈的傳染半徑、 SI 函數是擷取所有疫情樹



的傳遞鏈的發病間距、 i 傳遞鏈傳染病例、 lon 經度、 lat 緯度此處經緯度是病例 i 的發病位置經緯度座標。

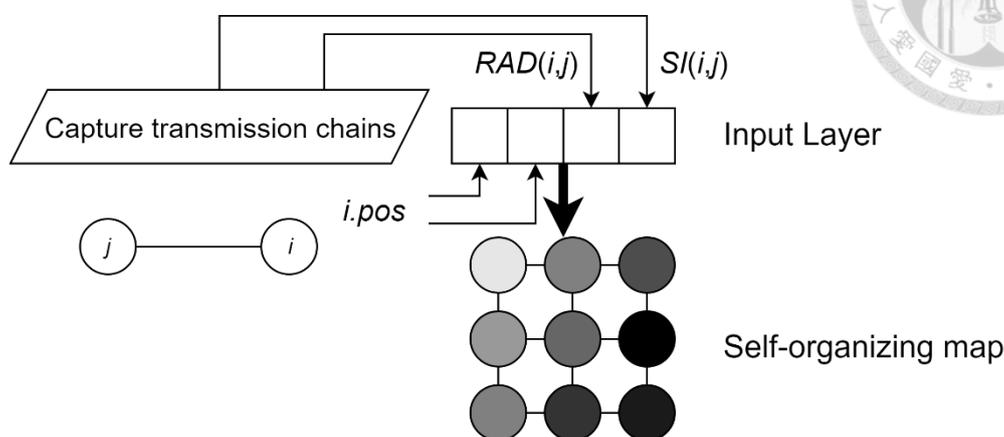


圖 16、輸入層示意圖。 RAD 函數是擷取所有疫情樹的傳遞鏈的傳染半徑、 SI 函數是擷取所有疫情樹的傳遞鏈的發病間距、 i 與 j 分別是傳染病例與被傳染病例構成傳遞鏈、 $i.pos$ 病例 i 的發病位置、Capture transmission chains 擷取傳遞鏈、Input Layer 輸入層、Self-organizing map 自組織對映。

(3) 自組織對映

採用基礎的自組織對映演算法(Kohonen, 2001)，自組織對映生成自組織拓樸圖，在拓樸圖上第 i 個 m 神經元³， x 是輸入層一筆資料的向量，於拓樸途中尋找 x 歐幾里得距離最接近的活化神經元 m_i ，找到之後輸出活化神經元 c (方程式 36)，則計算其他神經元位於拓樸的鄰近性； γ_c 代表活化神經元在拓樸的位置； γ_i 代表其他任意選定神經元在拓樸的位置， $\|\gamma_c - \gamma_i\|$ 兩者相對歐幾里得距離； $h_{ci}(t)$ 鄰近性函數又名高斯方程式，預設 $\sigma^2(t)$ 隨迭代 t 遞減鄰近性相除取負號指數， $\alpha(t)$ 為學習率隨迭代數遞減(方程式 38)，整個拓樸圖依照鄰近性 $h_{ci}(t)$ 更改接近 x 向量拓樸中任意選定的神經元 m_i 的權重(方程式 37)，以上敘述式基礎的自組織對映演算原理。

$$c = \underset{i}{\operatorname{argmin}} \{\|x - m_i\|\}$$

(方程式 36)

³ 此處的 i 不是登革熱傳染者，是代表選定的自組織對映的神經元，以上為了尊重原文的變數名稱命名。

$$m_i(t + 1) = m_i(t) + h_{ci}(t)[x - m_i(t)]$$

(方程式 37)

$$h_{ci}(t) = \alpha(t) \exp\left(-\frac{\|\gamma_c - \gamma_i\|}{2\sigma^2(t)}\right)$$

(方程式 38)

m 神經元權重、 $argmin$ 運算最小(距離)、 i 選定的神經元、 c 活化神經元、 x 輸入層一筆資料的向量、 t 迭代數、 h_{ci} 鄰近性函數又名高斯方程式、 $\alpha(t)$ 學習率、 $\sigma(t)$ 核心寬度。

本模組運作的架構利用疫情樹生成的疫情樹連結選取傳遞鏈，傳遞鏈具有傳染半徑與發病間距性質；傳染者與被傳染者的傳染半徑與發病間距，傳遞鏈當中的被傳染者透過登革熱的病例輸入模組根據模式需求輸入因子，本模式三的輸入層僅有傳染半徑、發病間距、傳染病例的經緯度座標，輸入層的向量經標準化，啟動執行續池平行化訓練基礎自組織對映網路。

以下敘述本模組實作基礎自組織對映的細節，輸入層輸入向量經標準化，由於標準化資料可以增加類神經網路訓練收斂的速度(Ioffe and Szegedy, 2015)，本拓樸圖的邊長是 20 乘 20 個神經元網路，設定時期(Epoch)總共 300 次，設定臨近參數(Sigma)為 3 下降至 1 的區間，運算方式參考鄰近參數下降方程式 39；學習率為 0.01 下降至 0.002，運算方式參考方程式 40，本模組的自組織對映在尋找活化神經元與計算臨近權重的函數採用 C++ Boost 的執行序池(Thread pool)，進行多核平行化排程訓練自組織對映，實作過程請參考原始碼與之詳解(附錄三之二自組織對映模組)。

$$\text{Let } \lambda = 2 \frac{\text{Epoch}}{\text{Toposize}}, \text{Epoch} = 300, \text{Toposize} = 20$$

鄰近參數下降方程式:

$$\sigma(t) = (\sigma_0 - \sigma_{final}) \exp\left(-\frac{t}{\lambda}\right) + \sigma_{final}, \sigma_0 = 3, \sigma_{final} = 1$$

(方程式 39)

學習率下降方程式:

$$\alpha(t) = (\eta_0 - \eta_{final}) \exp\left(-\frac{t}{\lambda}\right) + \eta_{final}, \eta_0 = 0.01, \eta_{final} = 0.02$$

(方程式 40)



$Epoch$ 總時期數、 $Toposize$ 拓樸圖的邊長、 λ 衰減率、 t 時期數、 σ 鄰近參數方程式、 σ_0 初始鄰近參數、 σ_{final} 終止鄰近參數、 α 學習率、 η_0 初始學習率、 η_{final} 終止學習率。

完成自組織對映的訓練後，提取 trc' 根據隨機性的剔除 20%的傳遞鏈產生的任一訓練資料集所有傳遞鏈做尋找最佳配飾單元的函數，輸入向量 x 的歐幾里得距離最短或者全部的神經元歐幾里得距離最短的神經元，此處採用全部的神經元找到輸入最短的神經元進行輸出，輸出的 c 是輸入 x 的活化神經元(方程式 41)，根據 c 的神經元產生傳染半徑的直方圖與發病間距的直方圖， x 是傳遞鏈具有的傳染半徑與發病間距分別加入 bmu 的神經元產生傳染半徑的直方圖與發病間距的直方圖，最後把 c 紀錄成活化神經元放入活化神經元集(方程式 42)，完成所有傳遞鏈的活化神經元與時空直方圖(圖 17)，以上結束本模式運算，存取自組織對映、活化神經元集合、傳染半徑與發病間距的直方圖。

$$\text{foreach } u = trc'$$

$$x \leftarrow [RAD(u), SI(u), u.i. lon, u.i. lat]$$

$$bmu = \underset{q}{\operatorname{argmin}} \{ \|x - m_q\| \}$$

(方程式 41)

$$bmu.Hist_{RAD} \leftarrow RAD(u)$$

$$bmu.Hist_{SI} \leftarrow SI(u)$$

(方程式 42)

trc' 根據隨機性的剔除 20%的傳遞鏈產生的任一訓練資料集、 m 神經元權重、 q 選定的神經元⁴、 bmu 活化神經元、 $Hist_{RAD}$ 傳染半徑的直方圖、 $Hist_{SI}$ 發病間距的直方圖、*Epidemic tree* 疫情樹是 DBSCAN 演算法完成深度搜尋後的承接變數、 RAD 函數是擷取所有疫情樹的傳遞鏈的傳染半徑、 SI 函數是擷取所有疫情樹的傳遞鏈的發病間距。

⁴ 此處的 q 是代表選定的自組織對映的神經元，之前使用 i 是為了尊重原文的變數名稱命名，但本處是模式變數，為了不搞混此處採用 q 。

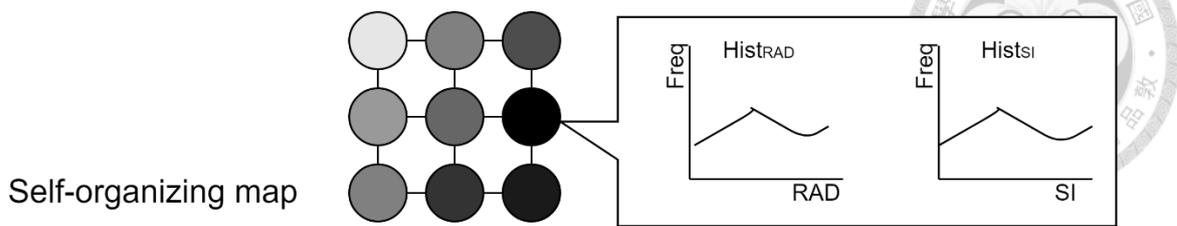


圖 17、自組織對映運作示意圖。Hist 直方圖、RAD 傳染半徑、SI 發病間距、Freq 頻率圖，Self-organizing map 自組織對映。

三、模式篩選

前面輸入層前置作業項傳遞鏈集合款提到根據隨機性的剔除 20% 的傳遞鏈，預設產生 10 組獨立的訓練資料集，這 10 組訓練資料集都獨立完成以下的自組織對映訓練，本項對這 10 組訓練資料集必須進行模式篩選，挑選出最適合的模式進行說明。

模式篩選的方法必須透過滾動預測的輪詢過程，請參考本結第五項滾動預測，滾動預測輪詢結果，唯一與滾動預測不同的點在，我們放的是訓練資料集，因此我的的病例 x 是屬於訓練資料集，得到病例的傳染半徑共識決直方圖 $x.Hist_{RAD}$ 與 $x.Hist_{SI}$ 病例的發病間距共識決直方圖，我們不斷嘗試 ec 預估涵蓋率從 0 每次增加 0.05 到 1 的過程，觀察時空擴散區 AT 與病例數之間的關係(方程式 43)，然而我們模式篩選有興趣的是時空擴散區密度與時空擴散區內病例數量之間的關係， k 是一個觀察點當時空擴散區的密度與時空擴散區內的病例數， $\frac{c_k}{AT_k}$ 是時空擴散區病例密度除時空擴散區涵蓋病例數結果， c_k 是時空擴散區涵蓋病例數，我們藉由切成 1 到 n 個片段累加時空擴散區內病例數與時空擴散區的密度，但由於時空擴散區內的病例數越多，則時空擴散區的病例密度會有衰減的現象，因此我們利用 u 時空擴散區密度的放大參數、 v 時空擴散區內涵蓋病例數的放大參數進行修正，最後取得 BM 最佳模式值(方程式 44)， BM 最佳模式值越大代表模式表現越好，關於最佳模式值證明請參考附錄六，我們挑選最佳模式值作為代表當年度本模式的模式。



Let $ec = [0, 1]$

$x \in TrD$

for ec plus 0.05 $x.RAD, x.SI \leftarrow cdf^{-1}(x.Hist_{RAD}, ec), cdf^{-1}(x.Hist_{SI}, ec)$

$$AT = \bigcup_{i=1}^n |x.RAD|^2 \pi |x.SI|$$

(方程式 43)

$$BM = \sum_{k=1}^n 0.5 \left(\frac{c_{k+1}}{AT_{k+1}} + \frac{c_k}{AT_k} \right) (c_k - c_{k+1}) e^u e^v$$

(方程式 44)

ec 預估涵蓋率、 $Hist_{RAD}$ 傳染半徑的直方圖、 $Hist_{SI}$ 發病間距的直方圖、 u 時空擴散區密度的放大參數、 v 時空擴散區內涵蓋病例數的放大參數、 BM 最佳模式值。

四、設定門檻值

當我們挑選出最佳模式，我們必須設定門檻值的方法必須透過滾動預測的輪詢過程，請參考本結第五項滾動預測，滾動預測輪詢結果，唯一與滾動預測不同的點在，我們放的是訓練資料集，因此我的病例 x 是屬於訓練資料集，得到病例的傳染半徑共識直方圖 $x.Hist_{RAD}$ 與 $x.Hist_{SI}$ 病例的發病間距共識直方圖，與上一項方程式 43 相同作法，我們不斷嘗試 ec 預估涵蓋率從 0 每次增加 0.05 到 1 的過程，觀察在每個預估涵蓋率時空擴散範圍對病例數一階導數的成長趨勢，其餘過程不在贅述，可參考本章第五節第三項，設定門檻值的細節。

$$range = AT^{-1} \left(\iint \frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2} (range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2} (range_i) \right) \partial c^2 \right)$$

(方程式 45)

$range_i$ 預估涵蓋率範圍在任意點 i 、 $range_{max}$ 是斷點特徵的預估涵蓋率點、 AT 是時空擴散範圍、 AT' 是時空擴散範圍對病例數的一階導數、 AT'_{max} 時空擴散範圍對病例數的一階導數發生最大值的現象、 k 是門檻特徵的預估涵蓋率點、 $threshold$ 是門檻特徵等於預估涵蓋率上界 p 。



五、滾動預測

時空擴散區模組的運作流程，回復自組織對映的訓練結果、對應的時空直方圖與活化神經元，時空擴散區模式是輸入任意 x 病例，對自組織對映進行輪詢 (Polling)，每次輪詢都採用不同的活化神經元的傳染半徑與發病間距組合，並限制在訓練階段的活化神經元尋找最佳配飾單元 (Best match unit)，所有時空半徑組合產生的最佳配飾單元集合 $BMUs$ ，累加屬於最佳配飾單元集合 $BMUs$ 中在訓練階段的第 k 個神經元傳染半徑直方圖與第 k 個神經元發病間距直方圖，累加屬於最佳配飾單元的時空直方圖是共識 (Consensus) 程序，之後生成 x 病例的傳染半徑直方圖的累積密度函數與發病間距累積密度函數，根據預估涵蓋率 ec 帶入 x 病例的傳染半徑累積密度反函數與發病間距累積密度反函數取得傳染半徑區間與發病間距區間，產生 x 病例的時空擴散環域 $x.RAD$ 與 $x.SI$ ，最後蒐集並累加多個 x 病例的時空擴散環域累加形成時空擴散範圍的直方圖分布 (圖 18)，虛擬碼請參考附錄三之三時空擴散區模組。

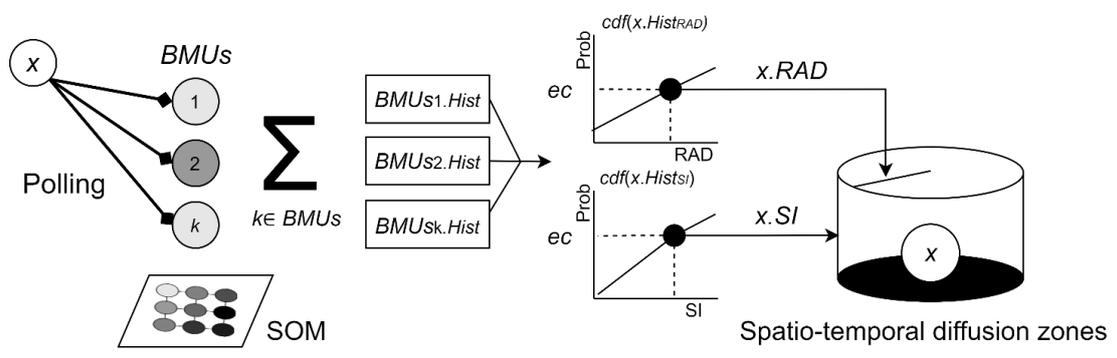


圖 18、時空擴散區模式示意圖。 x 登革熱病例、Polling 輪詢過程、SOM 自組織對映、 $BMUs$ 最佳配飾單元複數或最佳配飾單元集合 (Best match units)、 k 神經元編號、 $Hist$ 直方圖、 RAD 傳染半徑、 SI 發病間距、 $Prob$ 累積密度、 ec 預估涵蓋率、Spatio-temporal diffusion zones 時空擴散區。

我們當下無法得知任意 x 病例，預測傳染半徑與發病間距，輪詢過去自組織模式的活化神經元，過去在訓練階段自組織對映模組取得活化神經元集合 $Actnn$ ，活化神經元集合的神經元數量是 m 個，我們以 x 病例具有發病位置，根

據模式輸入因子，這些模式因子我們簡稱 $x.factors$ ，我們輪詢的技巧是 $x.factors$ 結合 $Actnn$ 活化神經元的 1 至 m 的傳染半徑 Rad 與發病間距 SI ，產生 m 組的 x 病例的輸入資料，另外一方面採用 $findtheBestMatchUnit$ 函數限制所有神經元 $\{m_q\}$ 在活化神經元集合 $Actnn$ 中尋找最佳配飾單元(方程式 46)，因此在此運算等於 $x.factors$ 結合 $Actnn$ 活化神經元的 1 至 m 的傳染半徑 Rad 與發病間距 SI 對 $Actnn.weights$ 做笛卡爾積，完成笛卡爾積找到所有排列組合 x 病例的組合歐幾里得距離最短的神經元 $argmin$ 的 $BMUs$ 活化神經元集合(方程式 47)。根據 $BMUs$ 累加的最佳配飾單位集合傳染半徑直方圖與發病間距直方圖，產生 x 病例的傳染半徑直方圖 $x.Hist_{RAD}$ (方程式 48)與發病間距直方圖 $x.Hist_{SI}$ (方程式 49)。

$$Actnn = \{active\ neurons\}, u = number\ of\ Actnn$$

Polling:

$$Actnn \leftarrow \{m_q\} \cap Actnn$$

(方程式 46)

$BMUs$

$$= \underset{bmu}{argmin} (\{[Actnn_1.Rad, Actnn_1.SI, x.factors], \dots [Actnn_u.Rad, Actnn_u.SI, x.factors]\} \times Actnn.weights)$$

(方程式 47)

Expert modeling:

$$x.Hist_{RAD} = \sum_{k \in BMUs} Hist_{RAD}[k]$$

(方程式 48)

$$x.Hist_{SI} = \sum_{k \in BMUs} Hist_{SI}[k]$$

(方程式 49)

$Actnn$ 活化神經元集合、 u 是 $Actnn$ 的活化神經元數量、 $\{m_q\}$ 所有自組織對映可以選定的神經元 m 第 q 個，集合符號代表任選 q 是自組織對映全部神經元的集合、 $x.factors$ 是 x 病例採用的因子、 $Actnn.weights$ 活化神經元集合的每個神經元權重集合、 $BMUs$ 最佳配飾單元集合、 \times 笛卡爾積(Cartesian product)、 $Hist_{RAD}$ 傳染半徑的直方圖、 $Hist_{SI}$ 發病間距的直方圖。



x 病例根據輪詢的過程得到 x 病例的 $x.Hist_{RAD}$ 與 $x.Hist_{SI}$ ，我們利用前一項設定門檻值運算出來的 $threshold$ 設定 $range$ 預估涵蓋間距 $range \leftarrow threshold = AT^{-1} \left(\int \int \frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2} (range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2} (range_i) \right) \partial c^2 \right)$ (方程式 50)，求取 x 病例的傳染半徑 $x.RAD$ 與 x 病例的發病間距 $x.SI$ ，透過帶入 $range$ 的 x 病例的傳染半徑直方圖 $x.Hist_{RAD}$ 的累積機率密度的反函數與帶入 $range$ 的 x 病例的傳染半徑直方圖 $x.Hist_{SI}$ 的累積機率密度的反函數，得到 x 病例的傳染半徑與發病間距(方程式 51)。

Define:

$$range \leftarrow threshold = AT^{-1} \left(\int \int \frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2} (range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2} (range_i) \right) \partial c^2 \right) \quad (方程式 50)$$

$$x.RAD, x.SI \leftarrow cdf^{-1}(x.Hist_{RAD}, range), cdf^{-1}(x.Hist_{SI}, range) \quad (方程式 51)$$

$threshold$ 門檻值、 $range$ 預估涵蓋間距、 $x.RAD$ 是 x 病例的傳染半徑、 $x.SI$ 是 x 病例的發病間距、 cdf^{-1} 累積機率密度的反函數。

第七節 模式四

模式四是實行訓練模式，架構與模式三幾乎一致，從疫情前期的訓練資料集重建疫情樹取得傳遞鏈集合，然而不同之處，在於透過自組織類神經網路進行時空擴散範圍的分類，分類各別病例發病位置座標、氣候與環境因子，根據各別登革熱病例的發病位置，預測屬於該登革熱病例獨特的傳染半徑與發病間距，生成登革熱的時空擴散區。

由於模式四與模式三的演算法架構類似，因此我們針對模式四與模式三不同之處進行論述，不在贅述期演算法細節，兩者模式最大的差別是輸入層階段，模式三僅採用病例的傳遞鏈的傳染半徑與發病間距，加上發病病例的經緯度座標，然而本模式在輸入層擴充環境取氣候因子，同樣放入自組織對映類神經網路進行訓練，其他運算細節與模式三一致。

另一方面還有些微的滾動預測的流程與模式三產生差異，由於本模式在訓練階段結合環境與氣候因子，因此在滾動預測階段必須加入環境與位置因子的資

料，同時氣候因子是一個具有時間性特色的變數，由於氣候是每日氣象共同構成的變數，因此我們必須要模擬真實滾動預測的情境，我們在滾動預測階段輸入層做了一些修改，針對結合氣候因子的模式四在滾動預測過程，輸入測試資料集的形式，來確保模式輸入具有時間的氣候因子，可以合理的符合實際滾動預測擴散的情境。

一、輸入層的修改

過去的模式三僅輸入因子僅有傳遞鏈的傳染半徑與發病間距加上病例病例 i 的經度與傳遞鏈中傳染病例 i 的緯度，僅四個維度的因子輸入輸入層，然而我們擴充輸入層的維度來添加環境與氣候因子，我們必須根據被傳染病例的發病日與發病位置的環境氣候作為輸入因子。

(1) 氣候與環境因子的選擇

氣候因子我們採用病例發病日前 140 日共 20 周，每日的最低溫度與日累積雨量作為我們的模式輸入層，由於文獻指出 20 周以前的溫度與雨量延遲會影響登革熱的發生風險(Hii, Rocklöv, et al., 2012; Hii, Zhu, et al., 2012; Hii et al., 2009)，臺灣登革熱研究結果依然發現發病日前 20 周的溫度與雨量樣態會提升登革熱風險(Chien and Yu, 2014)。

我們採用每日最低溫度，根據溫度擾動會顯著導致病媒蚊存活與族群消長(Lambrechts et al., 2011)，平均溫度降低會增長登革熱潛伏期(Watts et al., 1987)，臺灣地區 1998 至 2011 年南高屏研究區周最低溫度比起周平均溫度適合應用於評估研究登革熱疫情風險(Chien and Yu, 2014)，低溫會導致病媒蚊生存受到威脅，導致病媒蚊無法生存超過外生潛伏期。

我們採用每日累積雨量，探討延遲雨量因子對疫情的影響效果，周最大雨量對於臺灣登革熱風險影響最大(Chien and Yu, 2014)，但近年研究成果顯示累積雨量會優於最大雨量預測登革熱風險(Yuan et al., 2020; Hii, Zhu, et al., 2012; Cheong et al., 2013)，最大雨量考量的觀點是極端雨量沖刷導致病媒蚊喪失孳生源，但長期累積雨量的觀點是考慮氣候變遷導致長期雨量變化影響病媒蚊喪失孳生源，因此本研究採取周累積雨量作為因子，周累積雨量可以涵蓋季風氣候導致的長期雨量變化同時涵蓋極端降雨的特徵。



環境因子我們採用人口密度，人類是登革熱病毒傳染過程的重要宿主，人口數量與登革熱混合度是影響受感人口比的重要因子(Adams and Kapan, 2009)，人口密度提高登革熱暴露的場域，容易造成疫情向外擴散的現象(Wen, Lin, and Fang, 2012; Laureano-Rosario et al., 2018)，人類宿主提升受感與暴露的可能性，因此採用人口密度作為場域的宿主受感與暴露的因子。

(2) 適當降低運算量

我們的氣候與環境因子，環境因子只有人口密度一個因子，而氣候如採用前 20 周的每日的最低溫度與日累積雨量，會產生 140 個因子，輸入層的維度會非常高，運算量會非常龐大，我們發現過去的研究是採用周作為單位，研究氣候對於登革熱疫情地影響，因此我們嘗試將 140 日的每日累積雨量與最低溫度，簡化成 20 個輸入維度的輸入因子(方程式 52)。

其中必須注意一個細節，我們取得周最低溫度是透過取每周最低值的方法，等於每七日取一個日最低溫的最低溫，我們取周累積雨量，累加七日的每日雨量(方程式 52)。

$$\begin{aligned} & \{min_week(mintemp)_1, min_week(mintemp)_2 \dots min_week(mintemp)_{20}\} \\ & \leftarrow \{mintemp_1, mintemp_2 \dots mintemp_{140}\} \\ & \{cul_week(culprecp)_1, cul_week(culprecp)_2 \dots cul_week(culprecp)_{20}\} \\ & \leftarrow \{culprecp_1, culprecp_2 \dots culprecp_{140}\} \end{aligned}$$

(方程式 52)

mintemp 每日最低溫度、*culprecp* 日累積雨量、*min_week* 起每周最低值、*cul_week* 取每周累積值。

(3) 輸入層修改

從疫情樹輸出的疫情樹的連結中擷取傳遞鏈集合 *trc'*，輸入層向量以所有傳遞鏈 *u* 的傳染半徑、發病間距、傳遞鏈中傳染病例 *i* 的經度與傳遞鏈中傳染病例 *i* 的緯度放置輸入層，另外我們添加 20 周的周最低溫度與 20 周的周累積雨量，外加人口密度，總共獲得 45 個維度的輸入層(方程式 53)，整體而言傳遞鏈集合結合各個被傳染病例的位置輸入自組織對映的輸入層(圖 19)。



foreach $u = trc'$

InputLayer $\leftarrow [RAD(u), SI(u), u. i. lon, u. i. lat,$
 $\{min_week(mintemp)_1, min_week(mintemp)_2 \dots min_week(mintemp)_{20}\},$
 $\{cul_week(culprecp)_1, cul_week(culprecp)_2 \dots cul_week(culprecp)_{20}\},$
 $popden]$

(方程式 53)

trc' 根據隨機性的剔除 20% 的傳遞鏈產生的任一訓練資料集、InputLayer 輸入層、 RAD 函數是擷取所有疫情樹的傳遞鏈的傳染半徑、 SI 函數是擷取所有疫情樹的傳遞鏈的發病間距、 i 傳遞鏈傳染病例、 lon 經度、 lat 緯度此處經緯度是病例 i 的發病位置經緯度座標、 $mintemp$ 每日最低溫度、 $culprecp$ 日累積雨量、 min_week 起每周最低值、 cul_week 取每周累積值。

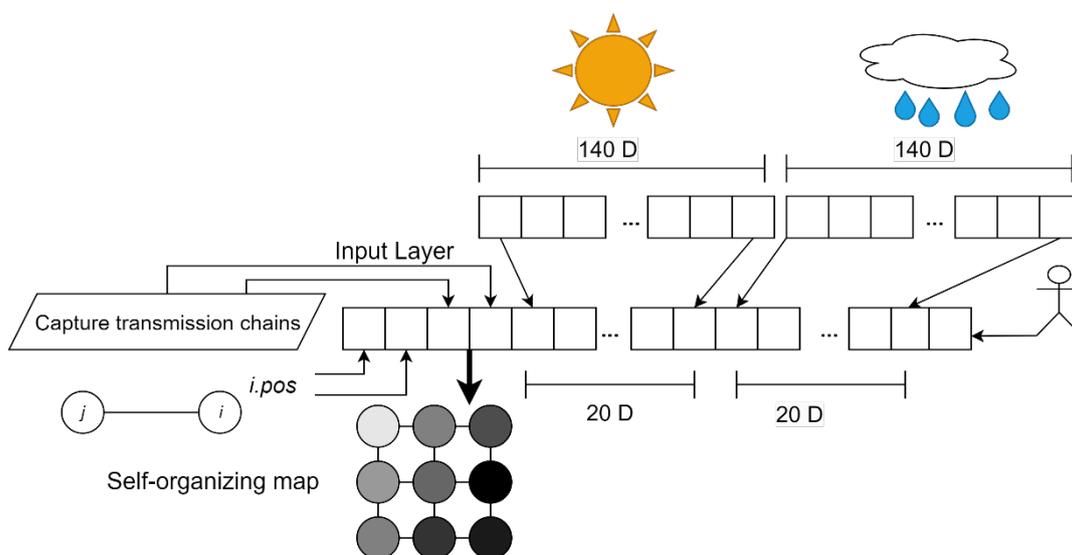


圖 19、模式四輸入層示意圖。 RAD 函數是擷取所有疫情樹的傳遞鏈的傳染半徑、 SI 函數是擷取所有疫情樹的傳遞鏈的發病間距、 i 與 j 分別是傳染病例與被傳染病例構成傳遞鏈、 $i.pos$ 病例 i 的發病位置、Capture transmission chains 擷取傳遞鏈、Input Layer 輸入層、Self-organizing map 自組織對映、 D 的是輸入維度，主要模式從原本的經緯度因子與傳遞鏈的傳染半徑與發病間距因子，我們還擴充氣候環境因子，氣候因子輸入 140 日最低溫度與日累積雨量，然而經過簡化降為 20 維度的輸入層，環境因子是人口密度一個因子。

二、滾動預測模式修改

由於我們模式在訓練階段有 45 個維度的輸入層，包含傳染半徑、發病間距、傳遞鏈中傳染病例 i 的經度與傳遞鏈中傳染病例 i 的緯度放置輸入層，另外我們添加前 20 周的周最低溫度與前 20 周的周累積雨量，外加人口密度因子，因此我們在滾動預測階段，測試資料集輸入滾動預測生成時空擴散區，實際上我們面臨一個問題，我們當初訓練是病例發病日是過去的氣象資料，然而我們無法得知下一個病例發病日期間的氣象資料，根據我們模式的預設，登革熱最短發病日是 6 日，另一方面我們可以在病例發病日當日，獲得約七日的氣象預報資料。

七日的氣象預報資料，在模式中由於我們具有完整的每日氣象預報資料，因此我們以病例發病日向後七日的氣象資料作為七日的氣象預報資料。

因此我們嘗試透過氣象預測資料輸入模式，我們將氣象資料共 140 日的日最低溫度與日累積雨量，加入 7 日的氣象預報資料，如方程式 54 的 $mintemp_{-1}$ 是病例發病日隔日的日最低溫度， $mintemp_{-7}$ 是病例發病日 7 日後的日最低溫度， $min_week(mintemp)_{-1}$ 是 7 日的預報日最低溫經過維度簡化取周最低溫後生成的輸入層因子項目，而雨量的部分以此類推，依然採用預報每日的累積雨量簡化成周累積雨量 $cul_week(culprecp)_{-1}$ ，並且將原本採用病例 140 日的氣象資料，向後推 7 日，依然維持 140 個維度的氣象資料(方程式 55)，然而經過簡化降為 20 維度的輸入層(圖 20)，再做輪詢(Polling)的程序接續原本的滾動預測模式(方程式 55)。

$$\begin{aligned} & \{min_week(mintemp)_{-1}, min_week(mintemp)_2 \dots min_week(mintemp)_{19}\} \\ & \leftarrow \{mintemp_{-7} \dots mintemp_{-1}, mintemp_1, mintemp_2 \dots mintemp_{133}\} \\ & \{cul_week(culprecp)_{-1}, cul_week(culprecp)_2 \dots cul_week(culprecp)_{19}\} \\ & \leftarrow \{culprecp_{-7} \dots culprecp_{-1}, culprecp_1, culprecp_2 \dots culprecp_{133}\} \end{aligned} \quad (\text{方程式 54})$$

$$\begin{aligned} \text{Polling: InputLayer} & \leftarrow [\text{Actnn}_u.\text{Rad}, \text{Actnn}_u.\text{SI}, u.i.\text{lon}, u.i.\text{lat}, \\ & \{min_week(mintemp)_{-1}, min_week(mintemp)_2 \dots min_week(mintemp)_{19}\}, \\ & \{cul_week(culprecp)_{-1}, cul_week(culprecp)_2 \dots cul_week(culprecp)_{19}\}, \\ & \text{popden}] \end{aligned} \quad (\text{方程式 55})$$

$mintemp$ 每日最低溫度、 $culprecp$ 日累積雨量、 min_week 起每周最低值、

cul_week 取每周累積值、Polling 輪詢過程、InputLayer輸入層、 $Actnn$ 活化神經元集合、 u 任意活化神經元。

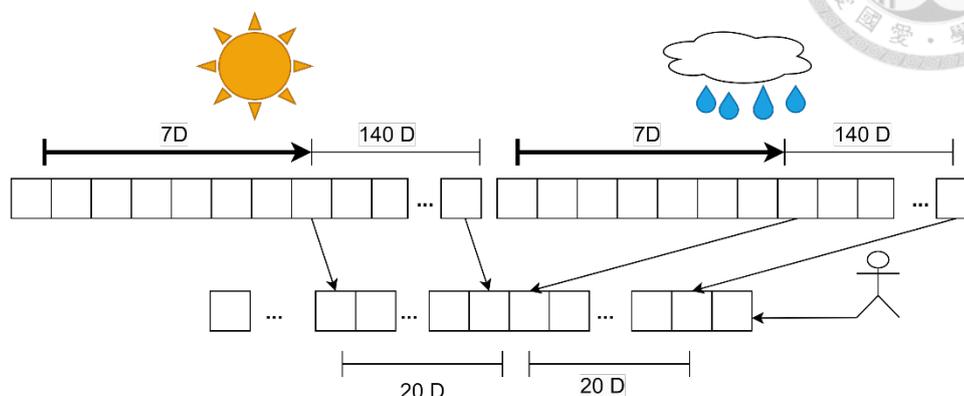


圖 20、測試資料集輸入滾動預測的輸入層向量示意圖。D 是代表維度，測試資料集數入輸入層，由於模式四應用氣象資料，在模式訓練階段，我們可以得知所有過去每日的氣象資料，然而在預測階段，面臨一個嚴重的挑戰，我們無法知道未來病例發病日，與未來病例的發病日之間的氣象資料，因此我們嘗試透過氣象預測資料輸入模式，本模式當中應用的氣象資料包含日最低溫度與日累積雨量，我們將氣象資料共 140 日的日最低溫度與日累積雨量，加入 7 日的氣象預報資料，並且將原本採用病例 140 日的氣象資料，向後推 7 日，依然維持 140 個維度的氣象資料，然而經過簡化降為 20 維度的輸入層，環境因子是人口密度一個因子。

第八節 驗證研究成果

一、指標定義

我們期望定義綜合指標(Multiple index)是檢核任何模式的時空擴散區預測的能力，然而我們認為所謂擴散模式的優劣，來自兩個觀點，第一預測時空擴散區能涵蓋病例，因此我們定義涵蓋率(Coverage)，第二點時空擴散範圍適當規模，模式不會過份預測時空擴散區得時空擴散範圍，模式認定過大的傳染半徑或者過長的發病間距，導致防疫資源的浪費，因此我們定義合適度(Fit)，如果一個指標可以同時考慮涵蓋率與合適度，我們定義為綜合指標(Comprehensive indicator)。

我們定義過去 i 病例有 1 到 n 個病例，但我們現在要觀察 τ 時間間距的時空擴散區與病例之間的關係。

涵蓋率是一段時間 τ 的病例數量 C 是屬於該段時間(方程式 56)，而一段時間 τ 的病例數 c 是空擴散區所涵蓋，他必須屬於任何過去預測的傳染半徑與發病間距內，因此在這段時間我們得到涵蓋率，是病例數 c 是空擴散區所涵蓋除以全部病例數量 C (方程式 57)。如果以單日作為一段時間 τ ， y_1 至 y_4 是當日的的所有登革熱病例共 4 個病例， y_1 至 y_3 被時空擴散區 Spz_1 與 Spz_2 所涵蓋共 3 個病例，因此當日涵蓋率是 0.75 (圖 21)。

合適度是一段時間 τ 的時空擴散區，所有任意 x 會預測出病例 i 的 1 到 n 個時空擴散區聯集得到時空擴散範圍，然而僅有部分的模式預測時空擴散區涵蓋登革熱病例，因此我們挑出涵蓋病例的時空擴散區取連集運算時空擴散範圍，這些時空擴散區我們簡稱有效時空擴散區，我們將有效時空擴散區除以聯集時空擴散區，得到合適度。如果以單日作為一段時間 τ ，總共有 Spz_1 至 Spz_3 共 3 個單日預測的時空擴散區，然而有效的時空擴散區僅有 Spz_1 與 Spz_2 (圖 21)，我們運算 Spz_1 至 Spz_3 共 3 個單日預測的時空擴散範圍即面積，與 Spz_1 與 Spz_2 單日預測的有效時空擴散範圍，並且將有效時空擴散範圍除以時空擴散範圍獲得當日合適度。

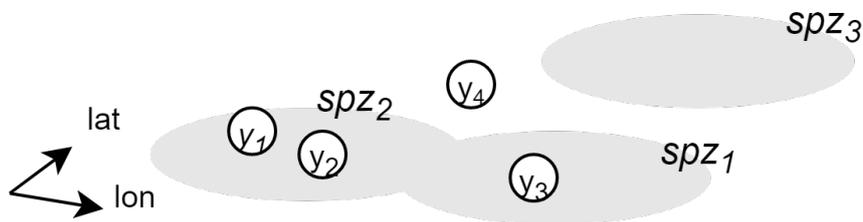


圖 21、單日滾動預測時空擴散區。綜合指標範例說明。 Spz_1 至 Spz_3 模式單日預測的時空擴散區、 y_1 至 y_4 當日的的所有登革熱病例、lat 緯度、lon 經度。

綜合指標的定義認為涵蓋率與合適度一樣重要，因此我們借助 F1-score 的概念，給予相等權重發展出綜合指標 $Midx$ ，一個帶入觀察一段時間的 τ 就能結合涵蓋率與合適度的檢核模式函數(方程式 60)。

定義:

$$\text{Let } x = \forall i, i = 1 \sim n$$

$$T = \tau$$

(方程式 56)

涵蓋率:

$$Cov(\tau) = \frac{c \in \tau \in x.RAD \in x.SI}{C \in \tau}$$



時空擴散範圍:

$$at_x = |x.RAD|^2 \pi |x.SI|$$

$$AT = \bigcup_{i=1}^n at_i$$

(方程式 58)

合適度:

$$Fit(\tau) = \frac{\bigcup_{i=1}^n |x.RAD|^2 \pi |x.SI \in \tau| \exists c \in \tau \in x.RAD \in x.SI}{\bigcup_{i=1}^n |x.RAD|^2 \pi |x.SI \in \tau|}$$

(方程式 59)

綜合指標定義:

$$Idc(\tau) = 2 \frac{Cov(\tau) \times Fit(\tau)}{Cov(\tau) + Fit(\tau)}$$

(方程式 60)

每日滾動預測的綜合指標:

Let τ is date of the start of rolling prediction
Define m is the date of the end of rolling prediction

$$Idc_Rl(\tau) = \frac{\sum_{p=0}^m Idc(\tau + p)}{m}$$

(方程式 61)

at_i 第 i 個病例的時空擴散範圍、 $i.RAD$ 是 i 病例的預測傳染半徑、 $i.SI$ 是 i 病例的預測發病間距、 AT 聯集時空擴散範圍、 n 病例總數、 c 時空擴散區涵蓋的病例、 C 總病例數、 $x.RAD$ 任意的登革熱病例 x 的預測傳染半徑、 $x.SI$ 任意的登革熱病例 x 的預測發病間距、 T 當前運算時期、 τ 任意時期、 Cov 涵蓋率函數、 Fit 合適度函

數、 Idc 綜合指標定義、 Idc_Rl 每日滾動預測的綜合指標、 m 滾動預測結束的日期。



綜合指標一段時間的 τ ，可以檢核一段時間的模式發展優劣，然而我們的每日滾動預測時空擴散區，必須每日預測時空擴散區，因此一段時間的 τ 只有每一日這種選擇，除了即時預測功能，然而我們又希望能觀察並評估及時預測功能的長期成效，因此我們稍微修正綜合指標的形式， $MIdx_Rl$ 定義綜合指標函數的變形，將 τ 作為起始日，將 m 作為終止日，計算每日的綜合指標後取平均，以此替代方案評量滾動預測模式的優劣(方程式 61)。

二、正規化均方根誤差

本研究設計四種模式，他們的輸入因子數可能不一樣，因此利用方均跟誤差的方法難以比較模式優劣，因此正規畫方均根誤差，是一個勉強能評估不同模式的誤差比較方法，然而我們不傾向以此作為指標，原因來自於實質上的登革熱病例空間分布的綜合指標，比起模式運算殘差的過程更直覺，同時更能模擬實際的情況。

然而本項討論方均跟誤差的目的，從另一方面只是要說明模式三與模式四，透過自組織類神經網路迭代過程，誤差降低並且有收斂的現象，以此正規化方均跟誤差隨迭代變化過程，間接證明相關的環境、區位與氣候因子可能有助於模式提升擴散預測能力。

(1) 比較模式的一致性

四種模式面對多樣性的輸出與輸入層的因子維度，根據過去文獻提及的比較各種模式產生的不同維度的輸出成果，本研究依照影像辨識的概念定義正規化均方根誤差，文獻提到正規化均方根誤差的概念來自產出的影像向量與參考影像向量差異除以參考影像向量的通則，應用於計算空間分辨率(Yamaguchi et al., 2020)，應用於比較輸入與輸出結果的一致性(Sun et al., 2021)。

以下說明四種模式如何透過正規化均方根誤差進行比較，模式一共有 2 個輸出，模式二有 2 個輸出，模式三有 4 個輸出與輸入因子，模式四有 45 個輸出與輸入因子，模式三與模式四由於是屬於訓練模式，因此可以藉由輸出與輸入維度

對位估計均方根誤差，反之模式一輸出的 2 個維度是預設傳染半徑與發病間距；模式二輸出的 2 個維度是歸納傳染半徑與發病間距，因此我們反向思考，如果某病例位在時空擴散區內，到底某病例的發病日與發病位置與模式認知的時空擴散區發病間距上界即最長發病日與傳染半徑上界的差異性，因此我們可以將模式一與模式二的發病間距上界與傳染半徑上界作為 I_{rj} ，而病例位在時空擴散區的發病日與傳染半徑做為 O_{rj} 。

正規化均方根誤差方程式，結合各模式輸入層向量的 j 個維度因子，第 r 個輸入層向量 I_{rj} 經過各模式進行運算後輸出層產生的向量 O_{rj} 的差絕對值平方進行累加除以 I_{rj} 向量絕對值平方累加，後續將絕對值平方相除結果開根號，觀察輸出向量與輸入向量達到一致性(方程式 62)。

$$NRMSE = \sqrt{\frac{\sum_r \sum_j |O_{rj} - I_{rj}|^2}{\sum_r \sum_j |I_{rj}|^2}}$$

(方程式 62)

r 是資料比數、 j 是輸入層向量的每個維度、 I_{rj} 是代表所有輸入層的向量 r 組 j 個向量維度、 O_{rj} 是代表所有輸出層的向量 r 組 j 個向量維度、 $NRMSE$ 正規化均方根誤差是結果比對輸出與輸入的一致性。

正規化均方根誤差方程式的概念，可觀察模式在自組織階段輸出的向量與輸入的向量達到一致性，如果一致性的現象發生正規化均方根誤差會相對的縮小，所以根據各種模式正規化均方根誤差縮小，可以用於評定模式的訓練結果與實際的結果一致的現象，根據本研究的假設密度方程式當中的系統性誤差項，會受到模式分類的優劣所影響，因此系統性誤差可透過正規化均方根誤差代換，系統性誤差越低的模式再預測相同病例數密度會相對有上升的趨勢，因此輸入向量與輸出向量一致性越高的模式，預測登革熱擴散區病例密度會有明顯的提升(方程式 63)，密度方程式的證明參考附錄四。



$$\frac{c}{\|ub(x + \Delta + \delta) - lb(x + \Delta + \delta)\|}$$

$$\text{let } \Delta = NRMSE \frac{c}{\|ub(x + NRMSE + \delta) - lb(x + NRMSE + \delta)\|}$$

$$\text{if } NRMSE \rightarrow 0, x = \frac{(ub(x + \Delta + \delta) + lb(x + \Delta + \delta))}{2}$$

$$\frac{c}{|a(p+q)|} \geq \frac{c}{\|ub(x + \Delta + \delta) - lb(x + \Delta + \delta)\|} \geq \frac{c}{\left|(e^p - \frac{1}{e^q})e^{x+\delta}\right|} \geq \frac{c}{\left|(e^p - \frac{1}{e^q})e^{x+\Delta+\delta}\right|}$$

(方程式 63)

Δ 是密度方程式系統性誤差，此處模擬NRMSE等於系統性誤差 Δ ，假設NRMSE趨近於零代表模式預測的結果優異，密度方程是可以得到密度最大化。

(2) 模式訓練達到收斂

正規化均方根誤差在模式內所代表的意義，是均方根誤差除以輸入層的向量絕對值平均 $\overline{|I|}$ 等價於均方根誤差在類神經訓練過程的誤差，誤差越低代表模式的預測力越佳，另一方面均方根誤差在上一個時期(Epoch)與未來任意 m 個時期類神經的訓練達到收斂的現象(方程式 64)，因此可藉由均方根誤差的變化過程觀察是否單一模式達到收斂的現象。

$$NRMSE = \frac{RMSE}{\overline{|I|}}$$

$$NRMSE_{epoch} = NRMSE_{epoch+m}, m \in N$$

(方程式 64)

NRMSE正規畫均方根誤差、RMSE均方根誤差、 $\overline{|I|}$ 輸入層的向量絕對值平均、epoch是類神經網路的時期、 m 屬於任意自然數。

第五章 研究成果

研究成果產製 2010 至 2015 年登革熱疫情爆發與非爆發年度，設計四種模式預測登革熱時空擴散區，在各個年度疫情前期的四個月分作為訓練資料集，接續疫情中後期兩個月作為測試資料集，疫情前期四個月的驗證資料集，提供訓練與驗證階段的運算，接續疫情中後期兩個月作為測試資料集做測試評估模式，在 2013 年度臺灣登革熱病例數稀少且無明顯的登革熱群聚傳染特徵，無法進行訓練與預測所以不列入成果，無法比較當年度的預測時空擴散區的表現。

登革熱疫情非爆發期間是 2010 至 2012 年與登革熱疫情爆發期間 2014 至 2015 年，這五年之間四種模式在各項指標評估結果具有明顯的差異，本章根據登革熱疫情非爆發之 2011 年為例與登革熱疫情爆發之 2015 年作為研究成果說明範例，選取這兩個年份的作為說明範例的理由，是 2011 年與 2015 年登革熱病例的空間分布分別跨越臺南市與高雄市兩大行政區，且在兩大行政區皆有發生群聚傳染的現象。

本研究成果分成五節，分別說明模式生成時空擴散區的性質，挑選每個年度最佳的模式，第一節綜合指標評量是根據建構的四種時空擴散模式，透過綜合指標評估長期的模式滾動預測表現，其中必須注意個模式在驗證每日滾動預測過程中的平均表現，例如綜合指標的細節反映模式生成時空擴散區的涵蓋率與合適度的特色，本章重點在挑選最佳的模式，預測時空擴散區應用於阻斷疫情傳遞鏈，防堵登革熱孳生源。

第二節時空擴散區滾動預測，我們根據上述綜合指標找到最佳的模式，透過視覺化平台展示滾動預測與驗證每日的滾動預測過程，觀察模式過去預測的時空擴散區是否涵蓋當日的病例空間分布，同時長期而言登革熱的時空擴散區的移動軌跡，是否有呈現疫情擴散的現象，我們可透過視覺化的過程呈現。

第三節訓練過程正規化均方根誤差，是觀察四種模式訓練過程正規化均方根誤差的結果與模式三與模式四的收斂過程，其中四種模式的整體方均跟誤差，評量的是各種模式在機器學習領域中，數值化評估模式的表現能力，是否能反映真實世界的結果，迭代模式的正規化方均跟誤差是反映經由類神經網路訓練模式，在訓練過程當中收斂的過程，說明在模式上的誤差收斂過程與結果，是否能反應模式的預測能力。



第四節時空擴散區的細節，我們討論各種模式生成時空擴散區的傳染半徑與發病間距細節，主要針對於應用於登革熱孳生源清除的實務。時空擴散區的時空擴散範圍，我們必須討論如果採用各種模式，可成產生的時空擴散範圍，這些時空擴散範圍可能是應用於登革熱孳生源清除的範圍，另一方面時空擴散區的病例密度是討論長期滾動預測的過程，一平方公里的時空擴散區能涵蓋的病例數，說明採取模式的優劣，然後談到時空擴散區傳染半徑與發病間距的細節，是觀察各種模式長期滾動預測，所認定登革熱病例的傳染半徑與發病間距結果，最後我們觀察長期滾動預測的傳染半徑與發病間距在空間當中分布的趨勢，觀察城鄉差異或空間差異性是否對於疫情的擴散過程產生變化，本節我們對於登革熱擴散的趨勢關注，並說明個模式對於登革熱疫情擴散的觀點，應用於防疫的操作。

第五節成果統整，統整模式的特性，說明應該採用的模式，以及各種評估觀點是否合乎滾動預測時空擴散區的宗旨，最後統整根據最佳模式模擬的登革熱擴散的趨勢與特色。

本研究成果聚焦在模式的選擇、模式的模擬過程與模式的細節，最後說明模擬登革熱擴散過程，是否能對於阻斷疫情傳遞鏈，登革熱孳生源清除，產生實質的助益，是本研究成果期望能達成的目的。

第一節 綜合指標評量

每日滾動預測的成果，綜合指標結合涵蓋率與合適度的概念，評估各個模式的表現成果，綜合指標是長期的滾動預測成果表現，我們可以透過綜合指標的量值選擇各年份的登革熱時空擴散區的模式，挑選出適合模擬登革熱擴散的過程，預測未來登革熱的高風險區域。

一、綜合指標評量

我們根據各年度測試資料集，在疫情中後期的兩個月進行滾動預測過程，綜合指標在各模式的預測成果，我們經過長期滾動預測綜合指標的數值(表 5)，在登革熱非爆發年度的 2010 至 2012 年的結果顯示，模式四採用位置、環境與氣候因子的預測成果，綜合指標的評論分別為 0.34、0.45 與 0.27(表 5)，由此結果認定模式四優於或等於其他模式的表現，然而在登革熱爆發的 2014 與 2015 年模式三採用位置因子的預測成果優於其他模式，顯然通過訓練的模式表現，優於其他



未經過訓練的模式，然而可能會因為疫情的嚴重程度，導致預測模式所採用的因子產生差異，登革熱爆發年份應該採取登革熱發病位置作為滾動預測因子，在登革熱非爆發年份，位置、環境與氣候因子可能對於登革熱的時空擴散區產生較大的影響。

在另一方面模式二未採用任何訓練模式，僅透過歸納傳染半徑與發病間距的過程，卻意外在綜合指標的考核中，獲得穩定甚至與最優的模式齊平的綜合指標數值，模式二從登革熱非爆發年份的 2010 年至 2012 年綜合指標分別是 0.34、0.41 與 0.24(表 5)，幾乎接近於模式四的綜合指標表現，另外在疫情爆發的 2014 至 2015 年比起其他模式，在綜合指標的表現接近於模式二，可見模式二在缺乏登革熱病例的訓練模式，或者缺乏位置、氣候與環境資料的情況下，僅採用模式二歸納登革熱的時空擴散區的傳染半徑與發病間距，其時空擴散區的滾動預測效果並不會產生過大的偏差。

模式一根據過去研究臺灣地區的疫情擴散過程所預設的傳染半徑與發病間距，就實作過程上中心性較強，較能還原登革熱的擴散過程，然而實際上長期滾動預測的成果，綜合指標卻遠遠不如其他三種模式，其各研究年份綜合指標值分別是 0.11、0.14、0.08、0.39 與 0.34(表 5)，不管在登革熱爆發年份與非爆發年份，其表現遠不如其他模式。

表 5、測試資料集綜合指標各模式各研究年份預測成果表。

年分\模式	一	二	三	四
2010	0.11	0.34	0.25	0.34
2011	0.14	0.41	0.35	0.45
2012	0.08	0.24	0.16	0.27
2014	0.39	0.52	0.57	0.51
2015	0.34	0.59	0.61	0.56

我們再根據各年度訓練資料集，在疫情前期的四個月滾動預測兼具驗證的過程，綜合指標在各模式的預測成果我們經過長期滾動預測綜合指標的數值(表 6)，登革熱非爆發年度的 2010 至 2012 年的結果顯示，模式四採用位置、環境與氣候因子的預測成果，綜合指標的評論分別為 0.21、0.23 與 0.21，在 2012 年度

的滾動預期成果卻略低於模式二的 0.22，然而在登革熱爆發的 2014 與 2015 年模式三採用位置因子的預測成果，在 2014 年 0.35 略高於模式二，在 2015 年卻低於模式二 0.38(表 6)，顯然通過訓練的模式的表現，在訓練資料集的滾動預測成果反而不如模式二歸納時空擴散區的傳染半徑與發病間距，與測試資料集的综合指標結果產生些微差異。

模式二未採用任何訓練模式，僅透過歸納傳染半徑與發病間距的過程，卻意外在綜合指標的考核中，獲得穩定甚至與最優的模式齊平的综合指標數值，以模式二從登革熱非爆發年份的 2010 年至 2012 年綜合指標分別是 0.20、0.19 與 0.22，幾乎接近於模式四的综合指標表現，另外在疫情爆發的 2014 至 2015 年比起其他模式，在綜合指標的表現接近於模式二甚至大於模式三(表 6)，可見模式二在缺乏登革熱病例的訓練模式，或者缺乏位置、氣候與環境資料的情況下，僅採用模式二歸納登革熱的時空擴散區的傳染半徑與發病間距，其時空擴散區的滾動預測效果除了不會產生過大的偏差，甚至優於模式三與模式四得採用因子經過訓練模式。

模式一根據過去研究臺灣地區的疫情擴散過程所預設的傳染半徑與發病間距，依然其各研究年份綜合指標值分別是 0.10、0.08、0.10、0.18 與 0.18(表 6)，不管在任何年份預測，其表現遠不如其他模式。

表 6、訓練資料集綜合指標各模式各研究年份預測成果表。

年分\模式	一	二	三	四
2010	0.10	0.20	0.17	0.21
2011	0.08	0.19	0.16	0.23
2012	0.10	0.22	0.17	0.21
2014	0.18	0.34	0.35	0.32
2015	0.18	0.38	0.36	0.34

二、綜合指標細節

每日滾動預測的综合指標，來自於涵蓋率與合適度，涵蓋率指的是單日所有社區感染病例與時空擴散區涵蓋病例數的比例，合適度來自於時空擴散區是否涵蓋當日的病例，如時空擴散區涵蓋當日病例的面積則是有效預測，反之則是無效



的時空擴散區，合適度觀察有效時空擴散區與全部的時空擴散區的比例，因病例比例與時空擴散區的比例發展綜合指標，其中本項需要關注在各模式中涵蓋率與合適度的成果對於綜合指標的影響。

本項我們觀察測試資料集各年度的綜合指標細節，由於訓練資料集是屬於模式建立的資料，因此無法有效檢核預測模式是否能應用於滾動預測過程，而訓練資料集的綜合指標細節放置於附錄一之一訓練資料集綜合指標細節，本項僅討論測試資料集對於登革熱疫情的中後期的綜合指標細節。

(1) 涵蓋率

涵蓋率指的是單日所有社區感染病例與時空擴散區涵蓋病例數的比例，我們首先觀察模式一的每日滾動預測涵蓋率分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.61、0.63、0.60、0.92 和 0.94(圖 22a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，每日僅包含約 60%的病例數，每日滾動預測的過程遺漏 40%左右的病例，在疫情爆發年份的 2014 與 2015 年卻可以包含 90%以上的病例，雖然涵蓋率中位數不及其他模式，然而也達到 90%以上，在登革熱的爆發年分模式依然有參考價值，然而我們觀察模式一的四分位距幾乎大於其他模式，接近於模式三，這可能反應每日滾動預測較為不精準，隨病例的分布位置是否接近時空擴散區影響涵蓋率的浮動。

模式二的每日滾動預測涵蓋率分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.87、0.88、0.71、0.95 和 0.97 (圖 22a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，每日包含約 70%的病例數，每日滾動預測的過程遺漏 30%左右的病例，在疫情爆發年份的 2014 與 2015 年可以包含 95%以上的病例。各年度每日涵蓋率中位數穩定，然而我們觀察模式二的四分位距小於其他模式，我們發現模式三透過歸納傳染半徑與發病間距的過程，可能產生的時空擴散區僅能捕捉部分的病例，涵蓋率整體表現而言普通偏高。

模式三的每日滾動預測涵蓋率分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.69、0.71、0.46、0.93 和 0.95 (圖 22a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，每日包含約 70%的病例數，然而在 2012 年卻只有 46%遠低於其他模式，每日滾動預測的過程遺



漏 30%左右的病例，甚至在 2012 年每日遺漏高達五成以上的病例數，在疫情爆發年份的 2014 與 2015 年可以大致包含 90%以上的病例。各年度每日涵蓋率中位數不穩定，尤其在登革熱疫情非爆發年度有明顯偏低的情形，整體而言在涵蓋率的表現並不理想，除模式一之外涵蓋率整體趨勢低於模式二與模式四。然而我們觀察模式三的涵蓋率四分位距接近模式一大於模式二與模式四，這個說明本模式針對病例的預測顯得相當不穩定，每日產生的時空擴散區似乎無法捕捉到真實的登革熱病例。

模式四的每日滾動預測涵蓋率分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.78、0.82、0.82、0.97 和 0.97 (圖 22a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，每日包含約 80%的病例數，每日滾動預測的過程僅遺漏 20%左右的病例，在疫情爆發年份的 2014 與 2015 年可以大致包含 95%以上的病例。各年度每日涵蓋率中位數穩定偏高的跡象，在涵蓋率的表現理想，尤其是針對登革熱疫情爆發年份的 2014 與 2015 年整體涵蓋率高於其他模式。然而我們觀察模式四的涵蓋率四分位距比起其他模式小，在 2012 年涵蓋率的四分位距寬大，然而集中於 0.71 至 1.00 之間因此在涵蓋率的滾動預測過程明顯優於其他模式，這個說明本模式針對病例的預測顯得相當穩定，每日產生的時空擴散區似乎比起其他模式最能捕捉到真實的登革熱病例。

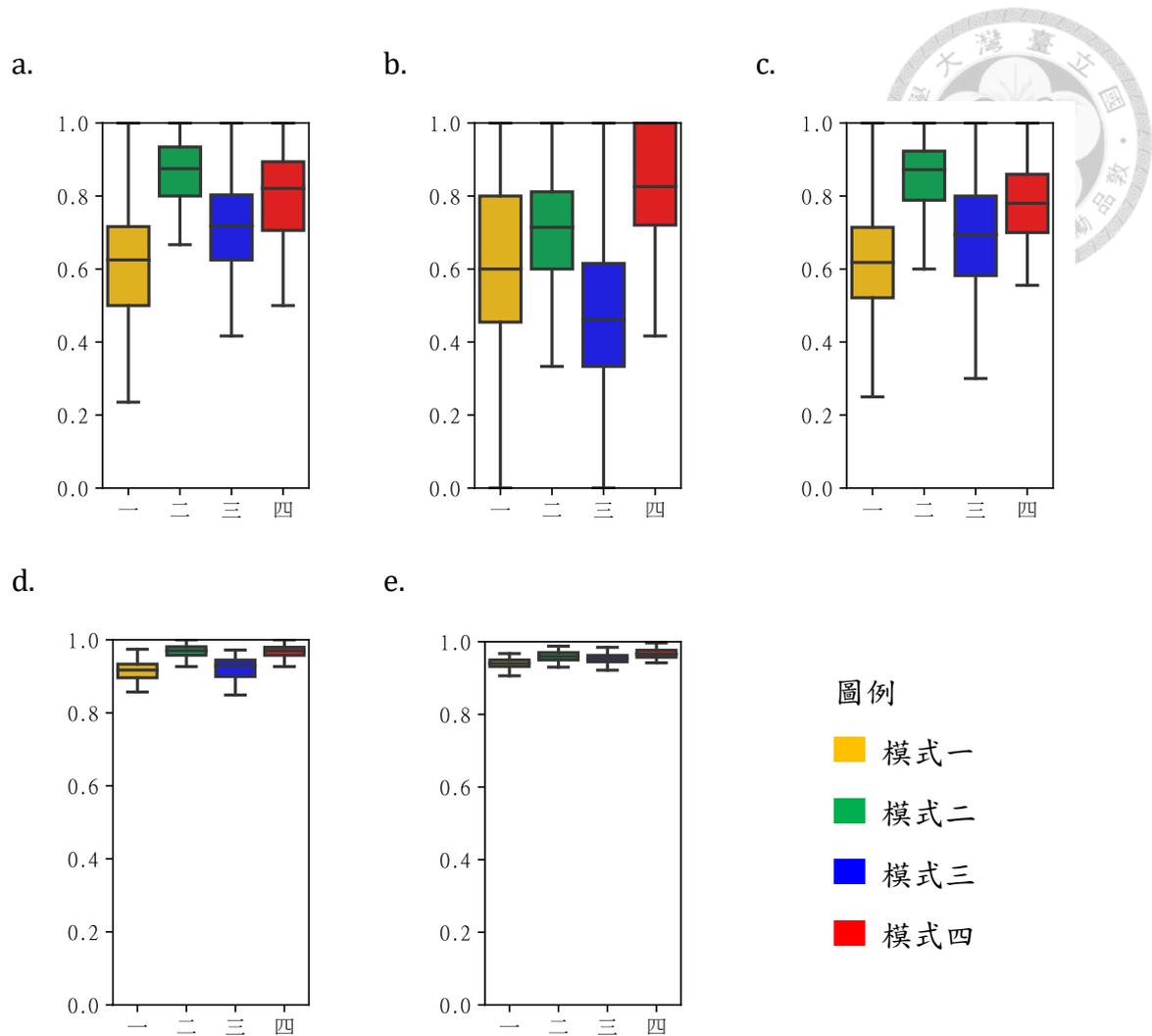


圖 22、預測資料集的每日涵蓋率分布盒狀圖。a-e 圖分別是 2010、2011、2012、2014 與 2015 年的四種模式涵蓋率，模式一至模式四分別由黃、綠藍和紅設色。

(2) 合適度

合適度觀察有效時空擴散區與全部的時空擴散區的比例，我們首先觀察模式一的每日滾動預測合適度的分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.06、0.08、0.05、0.25 和 0.21 (圖 23a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，每日僅低於 10% 的時空擴散範圍能涵蓋病例，高達 90% 以上的時空擴散範圍是屬於無效預測，在疫情爆發年份的 2014 與 2015 年僅可以包含約 20% 的時空擴散範圍是屬於有效的預測，其涵蓋率中位數皆不及其他模式，模式一所產生的時空擴散區，就合適度長期每日的觀察而言是最多的無效時空擴散範圍預測。其時空擴散區的合適度四分位距非常小而且第一與第三四分位距偏低的情況，幾乎可推測此模式產生的時空擴散區幾



乎無法涵蓋更多的登革熱病例。

觀察模式二的每日滾動預測合適度的分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.25、0.28、0.15、0.35 和 0.41 (圖 23a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，每日有 15% 的時空擴散範圍能涵蓋約病例，其中位數在 2010 年與 2011 年位居第二名，在 2012 年位居第一，在疫情爆發年份的 2014 與 2015 位居第二名，其涵蓋率中位數大致中等，模式二所產生的時空擴散區，就合適度長期每日的觀察而言是穩定生成時空擴散範圍。其時空擴散區的合適度四分位距幾乎與模式三和模式四類似，幾乎可推測此模式產生的有效的時空擴散範圍介於模式三與模式四之間。

觀察模式三的每日滾動預測合適度的分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.15、0.24、0.11、0.44 和 0.44 (圖 23a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，僅能保證每日有 10% 的時空擴散範圍能涵蓋約病例，尤其在 2012 年整體的合適度低落，其中位數在 2010 年與 2011 年位居三名，在疫情爆發年份的 2014 與 2015 卻位居第一名，其合適度中位數高達 40% 以上。其涵蓋率中位數浮動，受到登革熱疫情的爆發與非爆發年份所影響，模式三所產生的時空擴散區，就合適度長期每日的觀察而言僅適合應用於登革熱爆發的年份進行滾動預測。其時空擴散區的合適度四分位距幾乎與模式二和模式四類似。

觀察模式四的每日滾動預測合適度的分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的涵蓋率中位數分別是 0.21、0.32、0.17、0.34 和 0.38 (圖 23a-e)，在登革熱非爆發年度的 2010 至 2012 年產生的時空擴散區，僅能保證每日有 17% 的時空擴散範圍能涵蓋約病例，在合適度 2010 與 2011 年高於其他模式，僅於 2012 年低於模式二，在疫情爆發年份的 2014 與 2015 在合適度中位數表現位居第三名，其合適度中位數僅有 30% 以上。其涵蓋率中位數浮動，依然受到登革熱疫情的爆發與非爆發年份所影響，模式四所產生的時空擴散區，就合適度長期每日的觀察而言僅適合應用於登革熱爆發的年份進行滾動預測。其時空擴散區的合適度四分位距幾乎與模式二和模式三類似，然而在登革熱的非爆發年份其合適度第一四分位距與第三四分位距皆高於其他模式，因此我們認為根據合適度的考核結果，本模式適合應用於登革熱的非爆發年份。

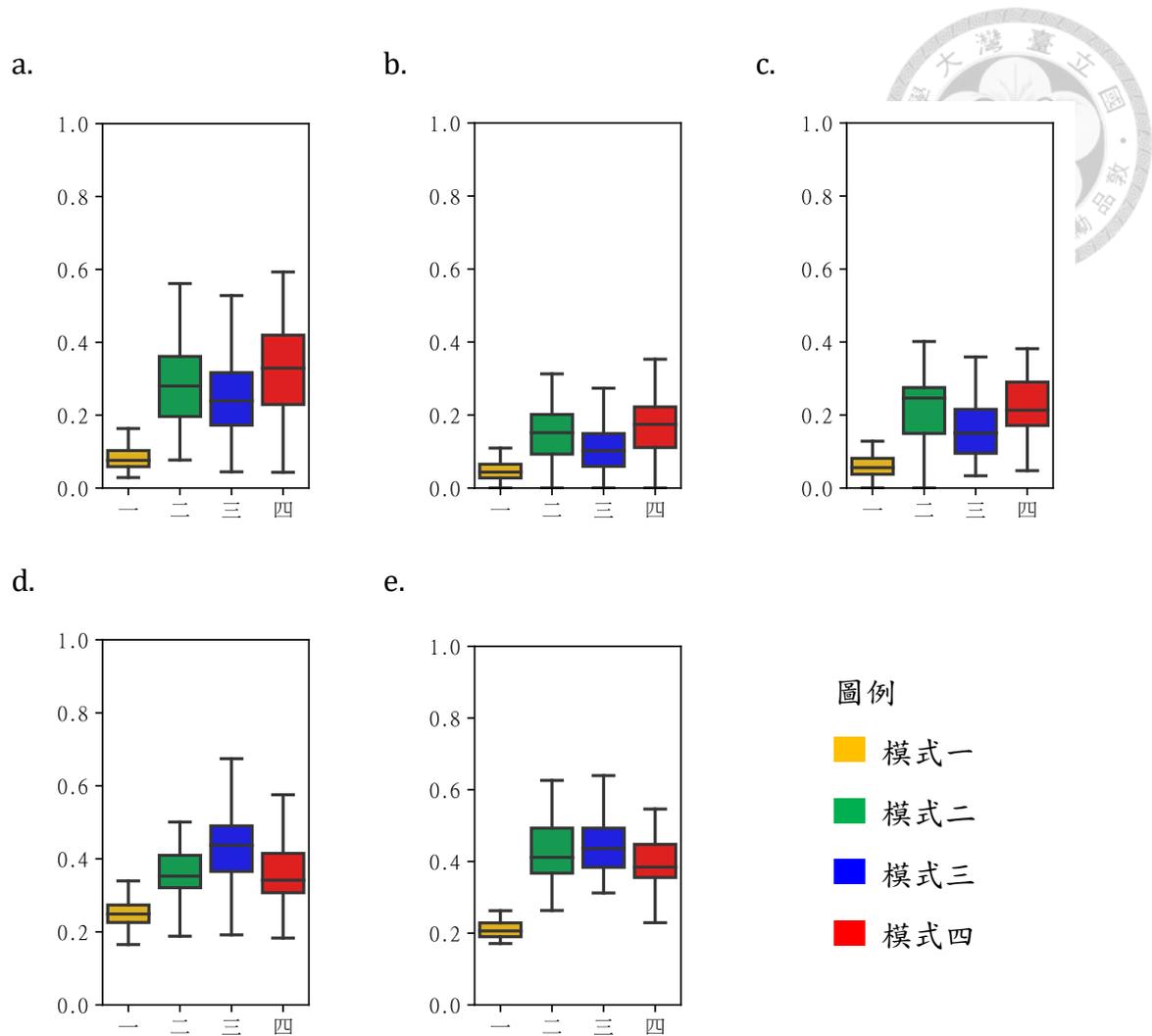


圖 23、預測資料集的每日合適度分布盒狀圖。a 至 e 圖分別是 2010、2011、2012、2014 與 2015 年的四種模式合適度，模式一至模式四分別由黃、綠藍和紅設色。

第二節 時空擴散區滾動預測

滾動預測是產生每日的時空擴散區與當日發病的病例，透過 Google 地圖的 API 與 C# 平台繪製滾動預測與驗證過程，滾動預測圖繪製每日病例發生的位置與模式產生的時空擴散區，每日病例發病位置標記在地圖上，病例標記分成綠色的登革熱本土病例位於時空擴散區內，藍色是登革熱本土病例未涵蓋病例與紫色登革熱境外移入病例。紅色的圓形緩衝區是有效時空擴散區，是過去登革熱病例透過模式預測的有效時空擴散區，有效時空擴散區是該過去預測的時空擴散區確實涵蓋當日病例的時空擴散區，而黃色的圓形緩衝區是無效時空擴散區，是過去登革熱病例透過模式預測的時空擴散區，在當日未涵蓋任何病例。

我們根據綜合指標的定義與細節，計算每日時空擴散區的時空擴散範圍量值，與計算有效時空擴散範圍，同時每日滾動預測模式可以偵測每日時空擴散區涵蓋的病例與所有的當日發病的病例，因此我們能因此得到每日的涵蓋率與合適度，並且加以演算每日的綜合指標值，以下每日綜合指標的細節已繪製於滾動預測的組圖中(圖 24)。

滾動預測時空擴散區與即時運算綜合指標的過程，我們挑選登革熱疫情非爆發年與登革熱爆發年各一年的疫情中後期，展示登革熱時空擴散區滾動預測的過程，我們展示各年度擁有最佳指標的模式在疫情中後期的滾動預測時空擴散區的成果。

滾動預測圖在疫情非爆發年採用 2011 年最佳模式的模式三的測試資料集，當滾動預測組圖屬於當年疫情中後期的 10 月 1 日至 10 月 8 日每日的過程與其運算綜合指標細節，主要聚焦於高雄市濱海都會地區的疫情發展敘述滾動預測的過程，由於當年度的登革熱病例好發於此地區。

滾動預測圖在疫情爆發年採用 2015 年最佳模式的模式三的測試資料集，當滾動預測組圖屬於當年疫情中後期的 9 月 1 日至 9 月 8 日每日的過程與其運算綜合指標細節，主要聚焦於臺南市濱海都會地區的疫情發展敘述滾動預測的過程，由於當年度臺南登革熱疫情情況是空前的嚴重，因此我們特別觀察該地區的時空擴散區發展動態。

我們首先觀察 2011 年測試資料集滾動預測時空擴散區與即時運算綜合指標的過程組圖，Google 地圖的位置位在高雄市的北半邊包含鼓山區、左營區、前鎮區、鳳山區與鳥松區，當期大多數的病例多集中在左營區，從 10 月 1 日的涵蓋病例分布主要集中於左營區，而且大部分的病例是屬於時空擴散區所涵蓋的病例，在圖中是呈現綠色的點，僅有一例在鳳山區北部是屬於未涵蓋病例，另外過去預測的時空擴散區大致分布於左營區，有效的時空擴散區紅色緩衝區部分圍繞左營區的輪廓，無效的時空擴散區集中於左營區中心與周遭的黃色緩衝區(圖 24a)。隔日 10 月 2 日涵蓋病例分布於左營區的南北鄰近區域，然而無效的時空擴散區分布於左營區中心，在左營區北邊出現一個時空擴散區未涵蓋病例(圖 24b)，隔日 10 月 3 日時空擴散區涵蓋的病例出現在左營區中心，因此有效時空擴散區分布於左營區內，然而疫情向東南方蔓延的趨勢依舊持續(圖 24c)。10 月 4 日有效病例與有效時空擴散區像左營區周圍蔓延，左營區中心的多是無效的時空



擴散區(圖 24d)。10月5日在左營區市中心又有許多有效時空擴散區與時空擴散區涵蓋病例，在北方的仁武區出現有效的時空擴散區，同時一個時空擴散區的涵蓋病例(圖 24e)。10月6日在左營區市中心又有許多有效時空擴散區與時空擴散區涵蓋病例，前鎮區也出現有效時空擴散區與時空擴散區涵蓋的病例，而無效時空擴散區沿著左營與前鎮區的外圍輪廓發展，北邊的仁武區也出現有效時空擴散區與時空擴散區涵蓋病例(圖 24f)。然而到了10月7日疫情有趨緩的現象，僅有少數的有效時空擴散區與時空擴散區涵蓋病例坐落於左營區，其他區域的時空擴散區是無效時空擴散區(圖 24g)。然而到了10月8日位於左營區的有效時空擴散區又擴大，無效的時空擴散區沿左營區的輪廓發展，然而在仁武區的時空擴散區是屬於無效的，在10月7日以後在仁武區疫情沒有復發的情況(圖 24h)。

然而我們觀察每日滾動預測時空擴散區綜合指標與其細節，這是當日全部研究區的彙整，我們發現從10月1日至10月8日涵蓋率介於0.5至1.0的區間(圖 24i)，在10月7日模式僅能找到病例當日發病的50%的病例，然而在隔日卻可以找到100%的病例在當日發病(圖 24i)，我們可以藉由觀察圖 24h也可以了解到所有時空擴散區皆涵蓋所有當日的病例。從10月1日至10月8日合適度介於0.09至0.48的區間(圖 24j)，10月7日發生合適度出現極低值的情形(圖 24j)，我們根據圖 24g推斷由於當日病例數非常少，導致滾動預測的時空擴散區多屬於無效預測，而疫情在當日緩和的情況，然而模式認定這些時空擴散區依然存在風險，而在10月2日出現極高合適度的情形(圖 24j)，是由於在左營區東南邊出現病例，因此過去預測的時空擴散區確實屬於有效預測。從10月1日至10月8日綜合指標介於0.15至0.62之間(圖 24k)，其中10月7日出現極低值(圖 24k)，原因可能出自於當日的合適度過低的情況，綜合指標的結果不彰，也反映出嚴重高估時空擴散區的現象，另一方面在10月5日綜合指標最高的情況(圖 24k)，我們可以從圖 24e找到線索，由於地圖中產生大量的時空擴散區涵蓋病例，因此產生大量有效的時空擴散區，兩個原因加成的結果導致當日綜合指標最高的跡象，根據圖 24i與圖 24j的數值圖，我們依然可以明由於涵蓋率與合適度高導致當日的滾動預測綜合指標高的情形。

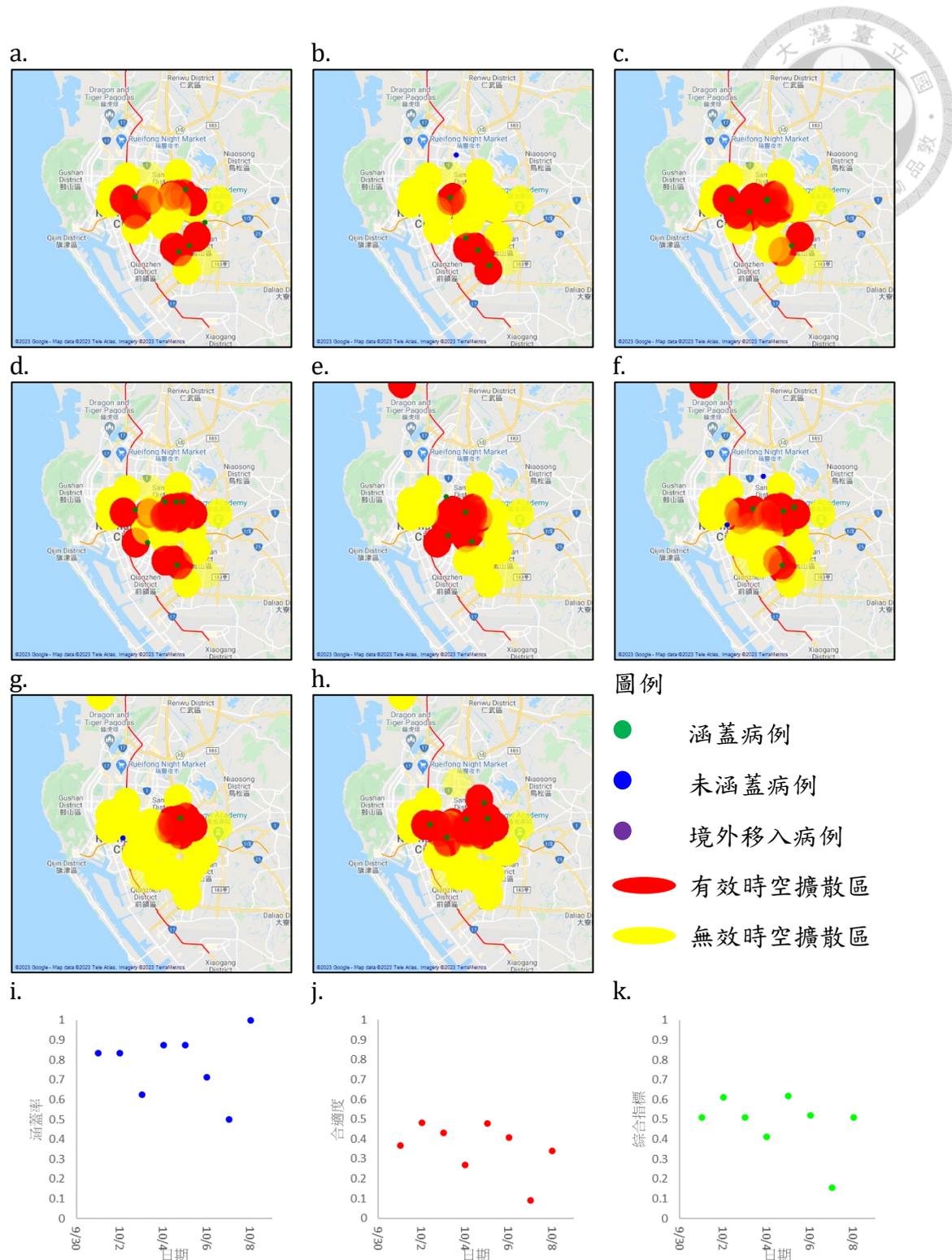


圖 24、2011 年測試資料集滾動預測時空擴散區與即時運算綜合指標的過程。圖 a-h 是 2011 年 10 月 1 日至 2011 年 10 月 8 日的每日滾動預測過程模擬，共八日八張滾動預測圖，以聚焦高雄市為例。圖 i-k 是研究區 2011 年 10 月 1 日至 2011 年 10 月 8 日每日全部的綜合指標與其細節，其中 i 圖是涵蓋率、j 圖示合適度、而 k 圖是綜合指標每日驗證運算數值圖。

我們再觀察 2015 年測試資料集滾動預測時空擴散區與即時運算綜合指標的過程組圖，Google 地圖的位置位在臺南市的西南邊與高雄市北邊，包含七股區、安南區、中西區、永康區、歸仁區、仁德區、中西區、南區、東區與高雄市的湖內區，當期大多數的病例多集中在高雄都會區與臺南都會區的安南區、中西區、南區與東區，疫情向周圍區域瀰漫至七股區、安南區、中西區、永康區、歸仁區、仁德區與高雄市的湖內區。

從 9 月 1 日的涵蓋病例分布主要集中於臺南都會區的安南區、中西區、南區與東區，而且幾乎所有的病例是屬於時空擴散區所涵蓋的病例，在圖中是呈現綠色的點，另外過去預測的時空擴散區大致分布於該區，而且在該區都是有效的時空擴散區，僅有零星的無效時空擴散區位於臺南市都會區的外圍(圖 25a)。隔日 9 月 2 日涵蓋病例與有效時空擴散區除臺南市都會區外，向北延伸永康區(圖 25b)，隔日 9 月 3 日時空擴散區涵蓋的病例與時空擴散區依然位在臺南市都會區，然而周遭產生一些無效的時空擴散區，位於鄰近都會區疫情有緩和的趨勢(圖 25c)。9 月 4 日依然時空擴散區與前一日類似的情況，但再臺南市都會區南邊與高雄市湖內區出現有效時空擴散區與時空擴散區涵蓋病例，在臺南市南邊與高雄市湖內區的疫情升溫，我們的滾動預測模式已經提前預警(圖 25d)。9 月 5 日依然沿承襲昨日的時空擴散區(圖 25e)。9 月 6 日在除臺南市都會區，在都會區南邊出現大量的有效時空擴散區，疫情確實蔓延至臺南市都會區的外圍，甚至都會區的東北方永康區也出現成堆的有效時空擴散區(圖 25f)。然而到了 9 月 7 日疫情有升溫的跡象，有效的時空擴散區已經遍布臺南市都會區的周圍(圖 25g)。然而到了 9 月 8 日位於臺南市都會區的有效時空擴散區又擴大，疫情沒有趨緩的跡象，同時在西北的七股區出現時空擴散區未涵蓋病例，因此可能在該地區可能未來爆發零星的登革熱群聚傳染(圖 25h)。

然而我們觀察每日滾動預測時空擴散區綜合指標與其細節，這是當日全部研究區的彙整，我們發現從 9 月 1 日至 9 月 8 日涵蓋率介於 0.91 至 0.94 的區間(圖 25i)，在疫情期每日時空擴散區可以找到 91% 以上的病例(圖 25i)，我們可以藉由觀察圖 25a-h 可以了解到所有時空擴散區皆涵蓋所有當日的病例。從 9 月 1 日至 9 月 8 日合適度介於 0.31 至 0.52 的區間(圖 25j)，9 月 7 日發生合適度出現極低值的情形(圖 25j)，我們根據圖 25g 推斷由於當日疫情趨緩，導致滾動預測的時空擴散區多屬於無效預測，而疫情在當日僅稍微餘臺南市都會區周遭地區緩和



的情況，然而模式認定這些時空擴散區依然存在風險，在 9 月 5 日出現極高合適度的情形(圖 25e)，是由於在臺南市都會區南邊與高雄市湖內區出現時空擴散區的涵蓋病例，因此產生過去預測的有效時空擴散區。從 10 月 1 日至 10 月 8 日綜合指標介於 0.47 至 0.67 之間(圖 25k)，其中 9 月 7 日出現極低值(圖 25k)，原因可能出自於當日的合適度過低的情況，綜合指標的結果不彰，也反映出高估時空擴散區的現象，另一方面在 9 月 5 日綜合指標最高的情況(圖 25k)，我們可以從圖 25e 找到線索，由於地圖中產生大量的時空擴散區涵蓋病例，因此產生大量有效的時空擴散區，兩個原因加成的結果導致當日綜合指標最高的跡象，根據圖 25i 與圖 25j 的數值圖，我們依然可以明由於涵蓋率與合適度高導致當日的滾動預測綜合指標高的情形。

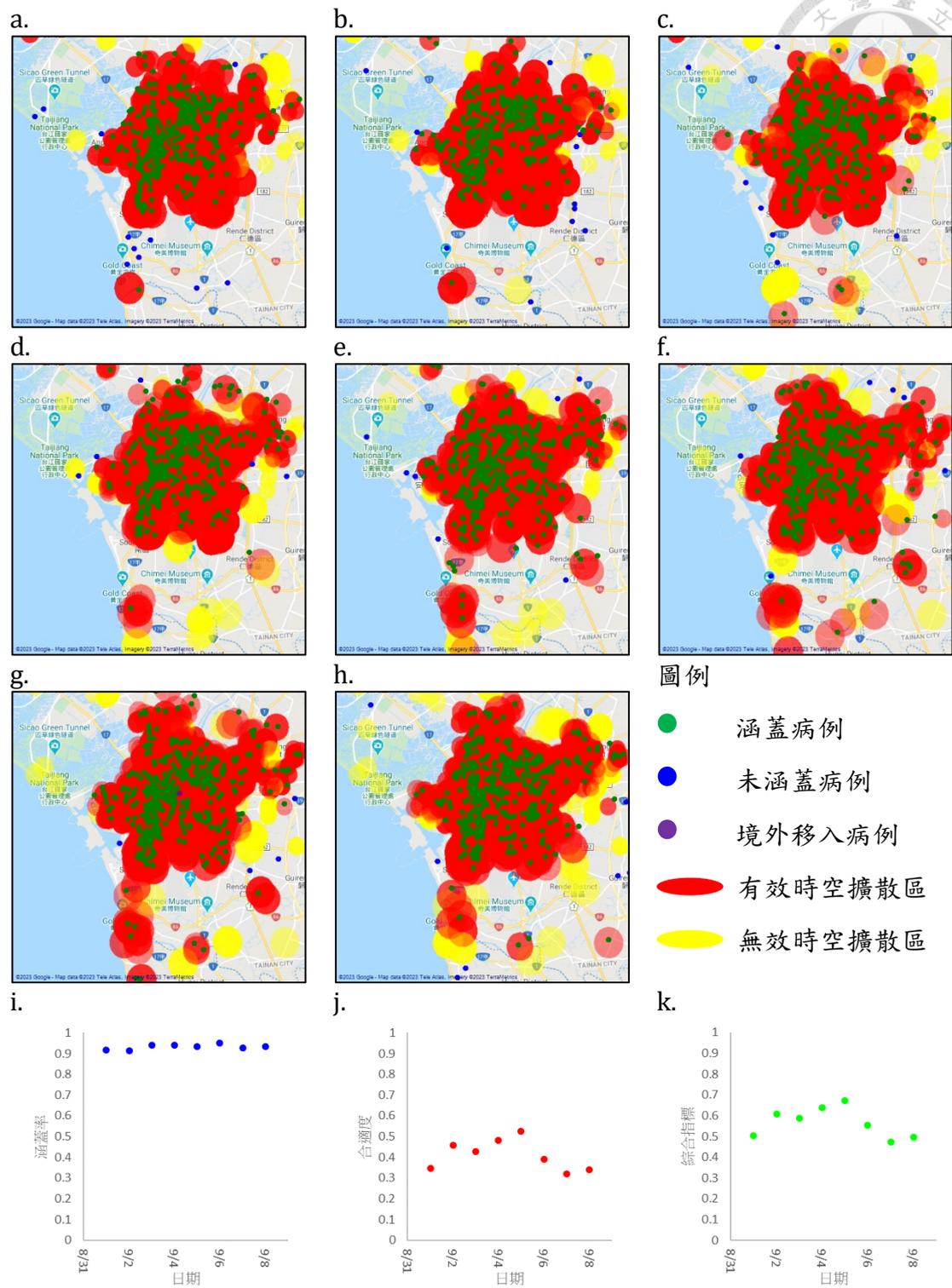


圖 25、2015 年測試資料集滾動預測時空擴散區與即時運算綜合指標的過程。圖 a-h 是 2015 年 9 月 1 日至 2011 年 9 月 8 日的每日滾動預測過程模擬，共八日八張滾動預測圖，以聚焦臺南市為例。圖 i-k 是研究區 2011 年 9 月 1 日至 2011 年 9 月 8 日每日全部的綜合指標與其細節，其中 i 圖是涵蓋率、j 圖示合適度、而 k 圖是綜合指標每日驗證運算數值圖。



第三節 正規化均方根誤差

各種模式有多樣化的輸入因子數，模式一預測傳染半徑與發病間距因此僅有 2 個輸入因子，模式二歸納傳染半徑與發病間距僅有 2 個輸入因子，然而模式三輸入除傳染半徑與發病間距外，還有傳染病例的座標經緯度共有 4 個輸入因子，模式四除模式三的四個因子外，外加氣候與環境因子共 45 個因子數，因此我們面臨比較模式誤差的困難，方均跟誤差是機器學習常見的誤差表示方法，然而其數值受到輸入因子的數量影響，幸好有正規化誤差的演算法，透過方均跟誤差與因子數的相除過程，得到一個一致標準的誤差比較模式，本研究由於建立四個模式，彼此的輸入因子數量差異，因此採用正規化的均方根誤差，是可以初步比較個模式在預測結果與過程中的誤差量值。

整體方均跟誤差的章節考慮各模式，在經過建模或訓練的結果，與輸入資料因子的誤差量值，作為模式參考價值的評量方式。然而迭代模式的正規化方均跟誤差衰減，是聚焦於模式三與模式四經由訓練的過程誤差的變化，進一步證明是否模式達到收斂的現象。

一、整體方均根誤差

整體方均跟誤差於測試資料集的情況，模式一高於其他模式，可見得根據預設傳染半徑與發病間距的模式，在誤差的數值觀點表現並不佳，這可能反應在綜合指標中的表現不佳的情形。模式二歸納傳染半徑與發病間距的模式，在 2010 年、與 2011 年居於最低的情形，也是最佳模式，但在 2012 年位居第二低，在 2014 年位居第二低，在 2015 年與模式四相同處於 0.9 最低的情況，或許這反映模式二在綜合指標的評估表現，在各年份位居第二名的特色。模式三的方均跟誤差在各年份的表現處於第三名與第四名，其中在 2012 年甚至是第四名高達 2.67 的方均跟誤差值，這似乎也反映當年綜合指標的評比僅勝過模式一，但綜合指標僅 0.16 非常的不理想的情況。模式四在方均跟誤差的表現上，在 2010 年與 2011 是第二名的低，在 2012 年是最低的情況，因此可推斷當年度的綜合指標成果最佳，經查證高達 0.27 是當年度最佳的模式，在 2014 與 2015 年是所有模式中誤差最低的模式，但實際上的綜合指標考核卻不如模式二 (表 7)。



表 7、測試資料集的模式正規化方均根誤差。

年分\模式	一	二	三	四
2010	1.95	1.16	1.54	1.30
2011	1.64	1.12	1.38	1.27
2012	1.91	1.75	2.67	1.55
2014	1.01	0.93	1.00	0.91
2015	0.96	0.90	0.94	0.90

整體正規化方均跟誤差於訓練資料集的情況，模式一依然高於其他模式，可見得根據預設傳染半徑與發病間距的模式，在誤差的數值觀點表現並不佳，這可能反應在綜合指標中的表現依然不佳的情形。模式二歸納傳染半徑與發病間距的模式，在 2010 與 2011 年位居最低，於 2012、2014 與 2015 年次低，或許這反映模式二在綜合指標的評估表現穩定的特色。模式三的方均根誤差在各年份的表現，誤差在 2010 至 2014 年最高，2015 年僅小於模式一之 0.1 的標準化方均跟誤差，這個趨勢可以反映綜合指標於 2010 至 2012 年與 2014 年的綜合指標表現僅位居第三名。模式四在方均跟誤差的表現上，在 2010 是第二名的低，在 2011 與 2012 年是最低的情況，因此可推斷當年度的綜合指標成果佳，經查證在 2011 年是 0.23 是當年度最佳的模式，然而在 2012 年是 0.21 位居第二名，在 2014 與 2015 年是所有模式中誤差最低的模式，但實際上的綜合指標考核卻不如其他模式 (表 8)。

表 8、訓練資料集的模式正規化方均根誤差。

年分\模式	一	二	三	四
2010	1.76	1.58	1.69	1.68
2011	1.74	1.47	1.59	1.47
2012	1.51	1.66	1.77	1.61
2014	1.38	1.27	1.47	1.26
2015	1.23	1.17	1.22	1.15



二、迭代模式的正規化方均根誤差衰減

迭代模式的正規化方均根誤差衰減，是聚焦於模式三與模式四經由訓練的過程誤差的變化，進一步證明是否模式達到收斂的現象，然而收斂的過程應該符合底下方程式 65 的推論，模式訓練收斂的情形在正規化方均根誤差前一期的迭代與往後的迭代 m 次的正規化方均根誤差一樣，但實際上在類神經網路運算的過程中是浮動的現象，我們僅能追求模式在迭代過程中正規化方均根誤差浮動的現象較小，則接近收斂的現象，然而這個過程無法透過任何數學公式進行運算，僅能透過多次的模擬，找尋浮動較小的正規化方均根誤差變化的迭代過程，以此說明模是達到收斂並且具備預測能力，其中我們經過多次的嘗試，我們設定任意模式三與模式四的迭代次數為 300 次，可以達到收斂的情形，所有訓練模式的標準一致觀察其方均跟誤差衰減的情形。

$$NRMSE_{epoch} = NRMSE_{epoch+m}, m \in N$$

(方程式 65)

$NRMSE$ 正規畫均方根誤差、 $RMSE$ 均方根誤差、 $\overline{|I|}$ 輸入層的向量絕對值平均、 $epoch$ 是類神經網路的時期、 m 屬於任意自然數。

我們觀察多個年份模式三在訓練資料集對應的疫情前期，訓練過程的方均跟誤差衰減現象，輸入因子有 4 個維度，理論上應能快速達到收斂的情形，尤其是學習率與鄰近學習率降低，應該能形成局部收斂的模式，獲得局部的最佳解，其中我們觀察圖 26a-e 的正規化方均跟誤差迭代情形，0 至 50 迭代數的過程中，正規化方均跟誤差快速的上揚，50 至 100 迭代數的過程正規化方均跟誤差開始緩升，100 至 200 迭代數開始較為緩降正規化方均跟誤差，然而到達 200 至 300 迭代數正規化方均跟誤差開始呈現穩定的跡象，然而在 2014 年與 2015 年的正規化方均跟誤差有小幅度上揚的情況，但整體而言模式的正規化方均跟誤差隨迭代過程穩定的浮動，因此我們可以推斷所有研究年份的模式三達到局部收斂的情形。

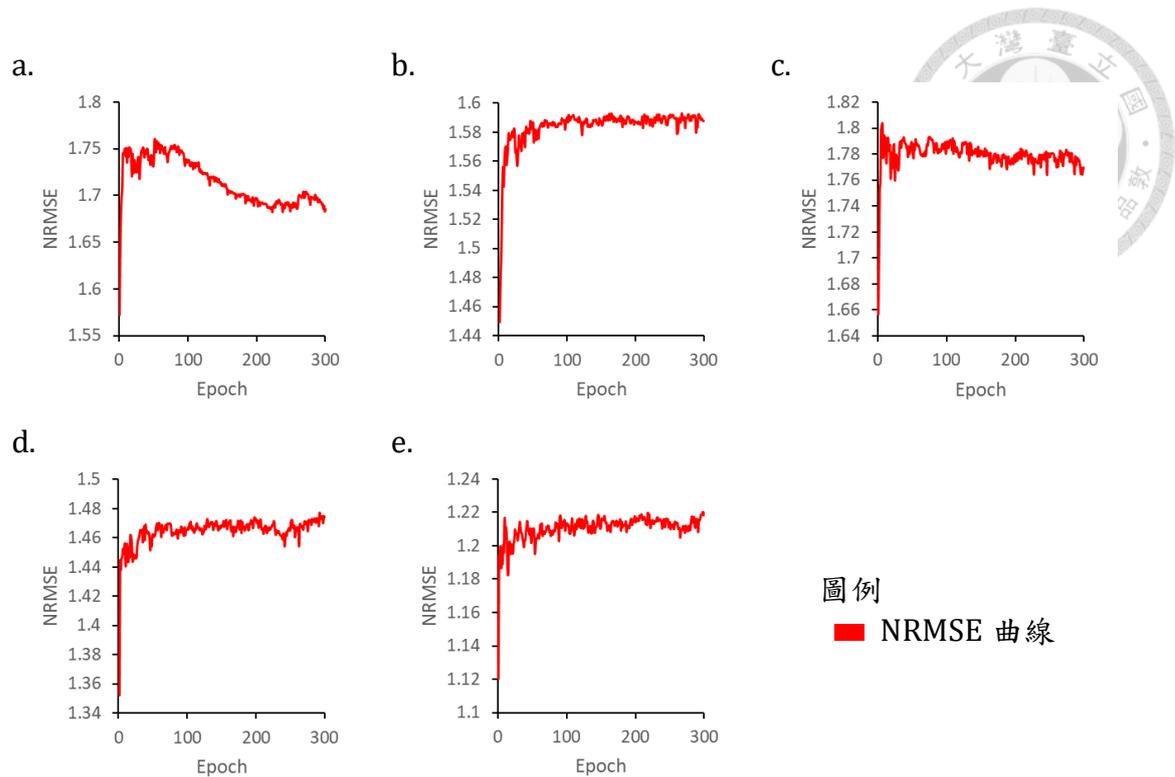


圖 26、各研究年份模式三的正規化方均根誤差衰減過程。a-c 是 2010 至 2012 年的訓練過程經過迭代過程的正規化方均跟誤差下降過程，d 與 e 分別是 2014 與 2015 年的訓練過程經過迭代過程的正規化方均跟誤差下降過程。NRMSE 曲線指的是正規化方均跟誤差曲線，橫軸是迭代數，縱軸是正規化方均跟誤差。

我們觀察多個年份模式四在訓練資料集對應的疫情前期，訓練過程的方均跟誤差衰減現象，輸入因子有 45 個維度，理論上較困難達到收斂的情形，必須長期的迭代過程獲得局部的最佳解，其中我們觀察圖 27a-e 的正規化方均跟誤差迭代情形，0 至 50 迭代數的過程中正規化方均跟誤差急遽的上揚，由於輸入因子非常多的情況，因此自組織類神經網路快速地調動權重，50 至 200 迭代數的過程正規化方均跟誤差開始急遽下降，證明多因子的學習模式從這個迭代過程快速歸納所有登革熱擴散有關的環境與氣候因子特徵，但是到了接近 200 至 250 迭代數區間 2010、2011、2014 與 2015 年卻發生正規化方均跟誤差局部急遽上揚的情形，實際上最佳化機器學習的方法，應該立即停損停止訓練模式，然而卻在 2012 年的正規化方均跟誤差衰減過程出現意外，由於 2012 年正規化方均跟誤差依在 200 至 250 區間不斷下降，直到接近 300 迭代數趨於穩定的特徵，因此我們的迭代數為求標準一致性，設定 300 迭代數終止訓練，而不採用 250 迭代數終止訓練，主要需要考量標準一致，整體需到達 300 迭代數才會趨於收斂的情形，然而犧牲的



其他年份的模式四的最佳化過程，但這種犧牲依然可以確保正規化方均跟誤差的值並不會是最差的情況，至少可以達成局部收斂與局部最佳解的情況。

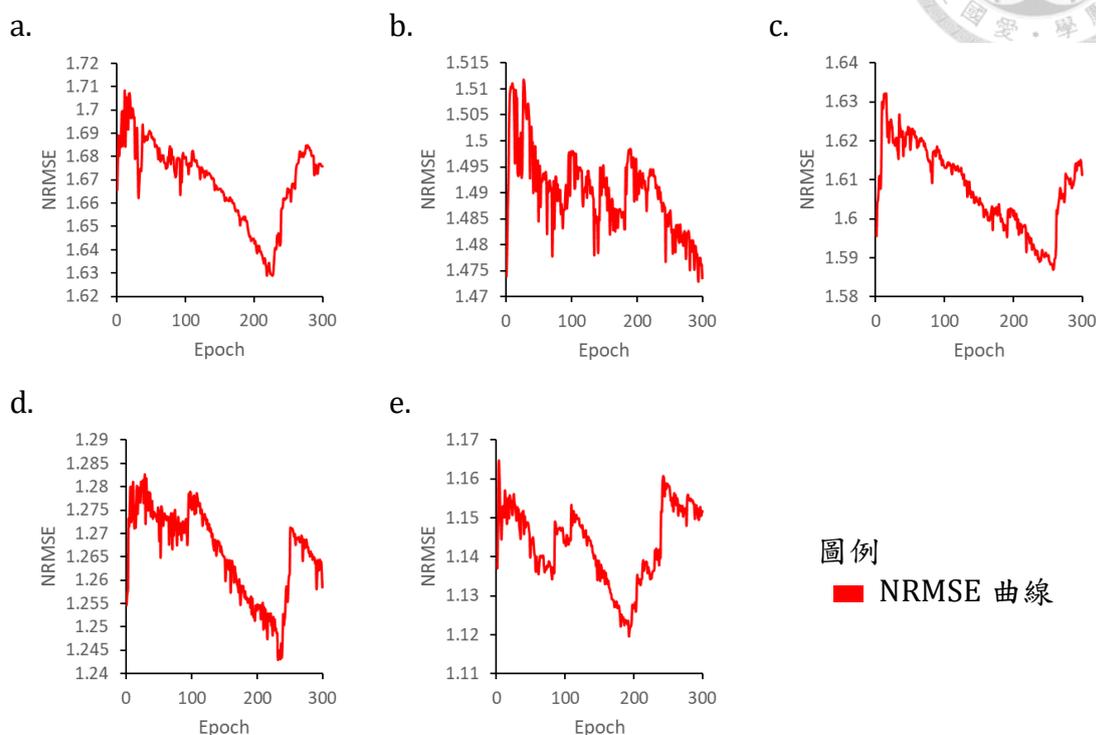


圖 27、各研究年份模式四的正規化方均根誤差衰減過程。a-c 是 2010 至 2012 年的訓練過程經過迭代過程的正規化方均跟誤差下降過程，d 與 e 分別是 2014 與 2015 年的訓練過程經過迭代過程的正規化方均跟誤差下降過程。

我們觀察正規化方均跟誤差隨迭代數的變化過程，儘管設置 300 迭代數無法達成最佳化的情況，但我們可以確保每個模式取得局部最佳解的情況，以上迭代數的設置是經過多次實驗後確立經驗值，我們根據正規化方均跟誤差衰減的過程中訂立的通則，可適用於研究登革熱時空擴散預測透過自組織類神經網路訓練過程的迭代數設置參考。

第四節 時空擴散區細節

時空過散區細節應用於阻絕登革熱疫情擴散的過程，其中時空擴散區的時空擴散範圍，說明所有模式在每日生成的時空擴散範圍應採取的防疫面積，時空擴散區的病例密度，說明一平方公里的時空擴散範圍在各模式各年份可以預測多少病例，時空擴散區的傳染半徑與發病間距細節是針對傳染半徑與發病間距在各研

究年分與模式，應該採取當日發病病例採取多大的傳染半徑，在未來的病例可能多長的最短發病日與最長發病日，會產生新的登革熱病例，應該如何制定阻絕登革熱擴散的空間與時間點，最後一項時空擴散區的傳染半徑與發病間距的空間分布，主要觀察最佳模式滾動預測模式針對發病間距與傳染半徑的空間分布，說明因地區與環境差異影響疫情擴散的趨勢。

一、時空擴散區的時空擴散範圍

時空擴散區的時空擴散範圍在疫情中後期的測試資料集(表 9)，我們觀察各個模式長期滾動預測的每日時空擴散範圍平均值，在模式一由於預設傳染半徑 300 公尺與發病間距 6 至 23 日，因此其時空擴散範圍面積每日小於其他模式，在登革熱疫情非爆發年度 2010 至 2012 年，每天僅有 10 平方公里的時空擴散範圍，而疫情爆發年份的 2014 年與 2015 年分別是每日 26.6 平方公里與每日 55 平方公里。

模式二是每年透過歸納疫情前期傳染半徑與發病間距生成時空擴散區，相對於其他模式，其生成的時空擴散範圍大於其他模式，在疫情非爆發年分的 2010、2011 與 2012 年分別是每日 36.6、31.8 與 23.1 平方公里，在疫情爆發年份的 2014 與 2015 年分別是每日 79.2 與 116.2 平方公里，是所有模式中滾動預測每日平均偏大的時空擴散範圍(表 9)。

模式三是每年透過空間分布差異，生成時空擴散區，相對於其他模式，其生成的時空擴散範圍偏低，在疫情非爆發年分的 2010、2011 與 2012 年分別是每日 22.6、15.5 與 7.4 平方公里，在疫情爆發年份的 2014 與 2015 年分別是每日 45.3 與 95.9 平方公里，預測的滾動預測每日平均最大的時空擴散範圍偏低的情形(表 9)。

模式四是每年透過空間、氣候與環境差異，生成時空擴散區，相對於其他模式，其生成的時空擴散範圍偏低，在疫情非爆發年分的 2010、2011 與 2012 年分別是每日 26.5、22.2 與 29 平方公里，在疫情爆發年份的 2014 與 2015 年分別是每日 76.8 與 130.8 平方公里，每日平均時空擴散範圍在 2010 與 2011 年位居第二名，在 2012 年居首，在 2014 年位居第二名，在 2015 年位居第五名(表 9)。



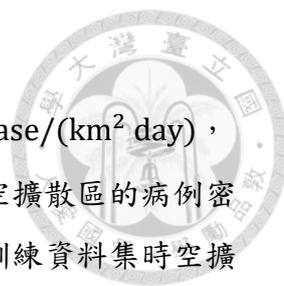
表 9、測試資料集的時空擴散區的時空擴散範圍每日平均值。值代表平均每日多少平方公里的時空擴散區於模式一至模式四，在每個研究年分。

年分\模式	一	二	三	四
2010	11.5	36.6	22.6	26.5
2011	9.0	31.8	15.5	22.2
2012	9.7	23.1	7.4	29.0
2014	26.6	79.2	45.3	76.8
2015	55.0	116.2	95.9	130.8

然而我們觀察時空擴散區的時空擴散範圍在疫情前期的驗證資料集(表 10)，我們觀察各個模式長期滾動預測的每日時空擴散範圍平均值，我們發現所有的模式在疫情前期的每日平均時空擴散範圍低於時空擴散範圍，我們根據綜合指標挑選出最佳模式進行比較，例如在登革熱疫情非爆發年份的 2010 至 2012 年模式四的平均每日時空擴散範圍的面積，分別是 14.7、10.3 與 20.3 平方公里(表 10)，遠低於疫情中後期的每日時空擴散範圍的面積，分別是 26.5、22.2 與 29 平方公里(表 10)，在疫情爆發年份的 2014 年與 2015 年在疫情前期每日時空擴散範圍僅 44.5 與 70.2 平方公里(表 10)，然而在疫情中後期卻高達 76.8 與 130.8 平方公里(表 10)，我們根據此規則推斷疫情前期的每日平均時空擴散範圍是疫情中後期的一半，因此我們事後推論在疫情前期如果可以預測出時空擴散區，並且採取防疫措施可能可以減少一半的防堵登革熱疫情擴散的資源。

表 10、測試資料集的時空擴散區的時空擴散範圍每日平均值。值代表平均每日多少平方公里的時空擴散區於模式一至模式四，在每個研究年分。

年分\模式	一	二	三	四
2010	5.4	23.6	13.7	14.7
2011	4	17.8	8.2	10.3
2012	6.3	16.9	5.4	20.3
2014	12.3	44.6	21.9	44.5
2015	23.6	58.3	48.6	70.2



二、時空擴散區的病例密度

時空擴散區的病例密度是每日每一平方公里涵蓋的病例數 $\text{case}/(\text{km}^2 \text{ day})$ ，我們以訓練資料集各研究年份與各模式每日滾動預測生成的時空擴散區的病例密度為例，訓練資料集時空擴散區的病例密度圖集在附錄一之二訓練資料集時空擴散區的病例密度。

我們首先觀察模式一的每日滾動預測時空擴散區的病例密度分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的時空擴散區的病例密度中位數分別是 0.31、0.48、0.26、2.68 和 2.61 $\text{case}/(\text{km}^2 \text{ day})$ (圖 28 a-e)，在登革熱非爆發年度的 2010 至 2011 年產生的時空擴散區的病例密度中位數，位居第一名在每一平方公里的時空擴散範圍可以涵蓋 0.3 個病例，然而在 2012 年位居第二名僅次於模式四，在疫情爆發年份的 2014 與 2015 年時空擴散區的病例密度中位數優於其他模式，然而我們觀察模式一時空擴散區的病例密度的盒狀圖的第一四分位距與第三四分位距區間大於其他模式，這可能反應預設傳染半徑與發病間距的模式僅涵蓋周遭的病例，導致整體時空擴散區的病例密度優於其他模式的特色。

觀察模式二的每日滾動預測時空擴散區的病例密度分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的時空擴散區的病例密度中位數分別是 0.17、0.23、0.13、1.00 和 1.45 $\text{case}/(\text{km}^2 \text{ day})$ (圖 28a-e)，在 2010 年、2011 年和 2014 年產生的時空擴散區的病例密度中位數低於其他模式，在 2012 年與 2015 年時空擴散區的病例密度中位數些微高於模式四的現象，然而我們觀察模式一時空擴散區的病例密度的盒狀圖的第一四分位距與第三四分位距區間在 2010 年、2011 年和 2014 年產生的時空擴散區的病例密度低於其他模式，在 2012 年與 2015 年時空擴散區的病例密度第一四分位距與第三四分位距區間依然略高於模式一的情況，這可能反應歸納傳染半徑與發病間距的模式可能過大劃設時空擴散範圍，導致整體時空擴散區病例密度表現不佳的現象。

模式三的每日滾動預測時空擴散區的病例密度分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的時空擴散區的病例密度中位數分別是 0.20、0.33、0.30、1.58 和 1.75 $\text{case}/(\text{km}^2 \text{ day})$ (圖 28 a-e)，在 2010 年、2011 年、2014 年和 2015 年產生的時空擴散區的病例密度中位數居次，在 2012 年時空擴散區的病例密度中位數高於模式一的現象，然而我們觀察模式一時空擴散區的病例密度的盒狀圖的四分位距大於模式二與模式四，我們可以推論模式二生成的時空擴散

區，受時空擴散區涵蓋病例的數量浮動所影響，可能受制於群聚傳染的嚴重程度導致時空擴散區的病例密度變化量較大。整體時空擴散區病例密度表現位居第二名，並且在 2012 年位居第一名，然而在 2012 年的綜合指標評量，表現卻非常不佳，因此我們推論當年度的時空擴散區病例密度，可能受到過小區列時空擴散範圍所影響。

模式四的每日滾動預測時空擴散區的病例密度分布情形，我們大致描述其 2010 至 2012 與 2014 和 2015 年的時空擴散區的病例密度中位數分別是 0.19、0.27、0.12、1.08 和 1.30case/(km² day) (圖 28a-e)，在 2010 年、2011 年、2014 年和 2015 年產生的時空擴散區的病例密度中位數僅優於模式二，在 2012 年與 2015 年時空擴散區的病例密度中位數位居最後，然而我們觀察模式一時空擴散區的病例密度的盒狀圖的第一四分位距與第三四分位距類似模式二，我們可以推論模式四生成的時空擴散區，其時空擴散範圍較大，縱使時空擴散區涵蓋更多的病例數，然而受到較大的時空擴散範圍，導致時空擴散區的病例密度表現不佳的情形。

最後我們統整各模式的時空擴散區的病例密度特色，在登革熱疫情非爆發年份的 2010 年至 2012 年每個模式幾乎低於 0.5case/(km² day)，然而在登革熱疫情爆發年份的 2014 與 2015 年任何模式大致可以高於 1case/(km² day)，這個結果也說明時空擴散區的病例密度，受到病例數的多寡與密集程度影響 (圖 28a-e)。

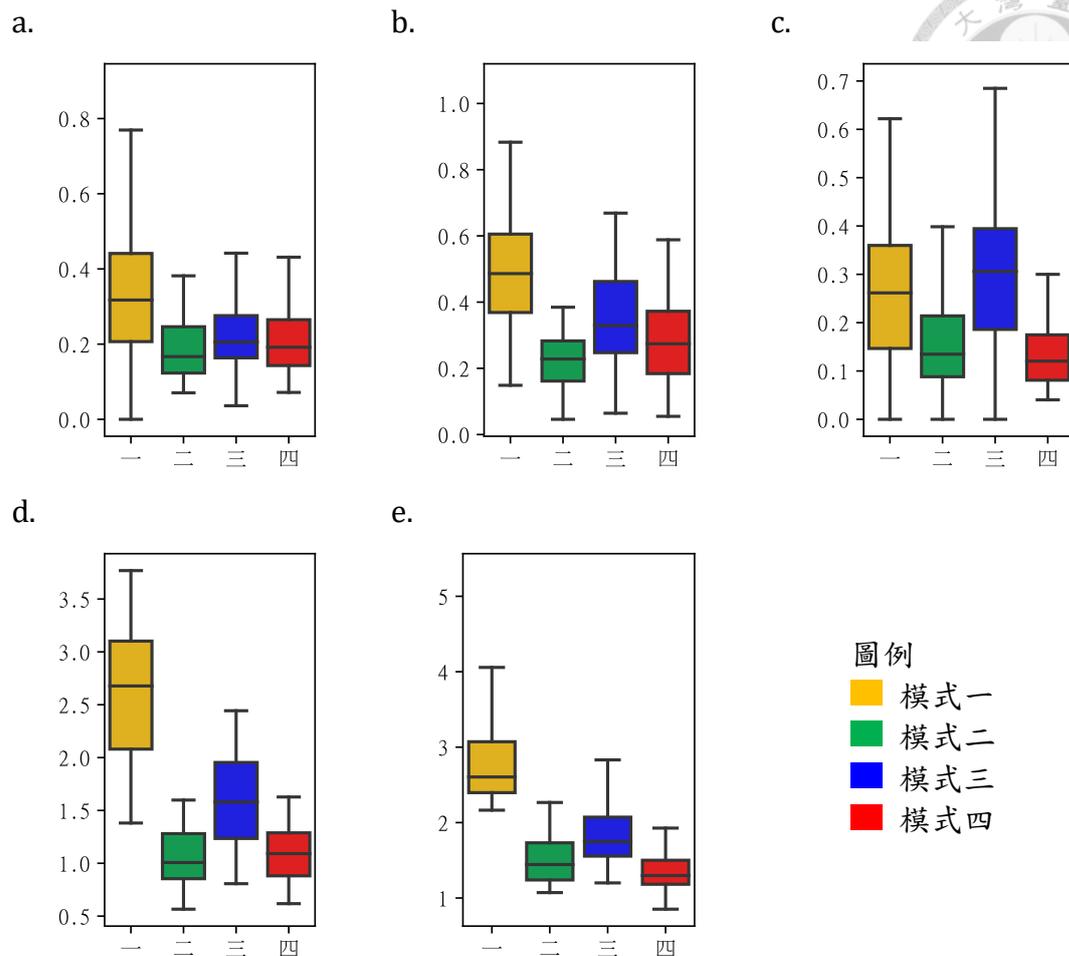


圖 28、測試資料集時空擴散區的病例密度。a-e 圖分別是 2010、2011、2012、2014 與 2015 年的四種模式的時空擴散區的病例密度，單位是每日每一平方公里涵蓋的病例數case/(km² day)，模式一至模式四分別由黃、綠藍和紅設色。

三、時空擴散區的傳染半徑與發病間距細節

時空擴散區的傳染半徑與發病間距細節，是觀察登革熱病例發生，應該採取當日發病病例採取多大的傳染半徑，在未來的病例可能多長的最短發病日與最長發病日，我們統整各個模式認定的傳染半徑，其傳染半徑上界是未來病例可能的傳染距離，與發病間距下界；即最短發病日，發病間距上界即最長發病日，我們統整時空擴散區的傳染半徑與發病間距細節，期望能透過模式模擬滾動預測的過程，了解時空擴散區時空特性。

時空擴散區的傳染半徑與發病間距細節，我們觀察疫情的中後期隨疫情的穩定發展，透過測試資料集滾動預測時空擴散區的過程，在研究年份與研究區內傳染半徑與發病間距細節，歸納每個年份任意模式對於登革熱擴散趨勢，然而疫情

前期的傳染半徑與發病間距細節請參考附錄一之三訓練資料集時空擴散區的傳染半徑與發病間距細節，由於登革熱疫情發展初期，變異性較大，比較難以看出時空擴散區的趨勢。

我們觀察模式一時空擴散區的傳染半徑與發病間距細節，由於傳染半徑與發病間距是根據過去文獻預設所以傳染半徑是介於 0 至 300 公尺，而發病間距之最短發病日是 6 日，最長發病日是 23 日(表 11)。

模式二時空擴散區的傳染半徑與發病間距細節，由於傳染半徑與發病間距是根據每個年份歸納訓練資料集，即疫情前期病例產生的傳遞鏈，其傳染半徑與發病間距，而傳染半徑皆由發病位置的 0 公尺，傳染半徑的上界因年份產生差異，2010、2011、2012、2014 與 2015 年分別是 920、960、810、970 和 980 公尺，每個年份採固定的傳染半徑繪製時空擴散區的傳染半徑，而發病間距之最短發病日是 6 日，而最長發病日的上界因年份產生差異，2010、2011、2012、2014 與 2015 年分別是 14、17、11、15 和 9 日，每個年份採固定的傳染半徑繪製時空擴散區的發病間距。(表 11)。

模式三時空擴散區的傳染半徑與發病間距細節，由於傳染半徑與發病間距是根據每個年份訓練資料集，結合病例的發病位置生成擴散滾動預測模式，其傳染半徑與發病間距，而傳染半徑皆由發病位置的 0 公尺，傳染半徑的上界因發病位置產生差異，因此 2010、2011、2012、2014 與 2015 年分別是 683 ± 185 、 608 ± 115 、 429 ± 38 、 770 ± 141 和 829 ± 145 公尺，皆採不固定的傳染半徑繪製時空擴散區的傳染半徑，而發病間距之最短發病日是 6 日，而最長發病日的上界因發病位置產生差異，2010、2011、2012、2014 與 2015 年分別是 11.7 ± 0.6 、 12.4 ± 5.0 、 9.5 ± 0.9 、 8.5 ± 1.5 和 8.9 ± 0.8 日，每個病例採不固定的傳染半徑繪製時空擴散區的發病間距。(表 11)。

模式四時空擴散區的傳染半徑與發病間距細節，由於傳染半徑與發病間距是根據每個年份訓練資料集，結合病例的發病位置、氣候與環境生成擴散滾動預測模式，其傳染半徑與發病間距，而傳染半徑皆由發病位置的 0 公尺，傳染半徑的上界因發病位置產生差異，因此 2010、2011、2012、2014 與 2015 年分別是 855 ± 75 、 906 ± 105 、 837 ± 72 、 914 ± 56 和 981 ± 22 公尺，皆採不固定的傳染半徑繪製時空擴散區的傳染半徑，而發病間距之最短發病日是 6 日，而最長發病日的上界因發病位置產生差異，2010、2011、2012、2014 與 2015 年分別是 1.4 ± 1.1 、



11.5±2.0、10.9±1.1、16.0±2.8 和 10.3±1.0 日，每個病例採不固定的傳染半徑繪製時空擴散區的發病間距。(表 11)。

我們根據綜合指標挑選出最佳模式，分別在登革熱疫情非爆發年份的 2010 年至 2012 年採用模式四，在登革熱疫情爆發年分的 2014 與 2015 年採用模式三觀察傳染半徑與發病間距細節，經過統整結果 2010 年至 2012 年傳染半徑上界平均是 855、906 與 837 公尺，發病間距上界平均是 11.4、10.5 與 10.9 日，大致歸納登革熱在疫情非爆發年份平均傳染半徑大於 800 公尺，平均發病間距上界約 11 日。

在 2014 與 2015 年採用模式三觀察傳染半徑與發病間距細節，經過統整結果 2014 年與 2015 年傳染半徑上界平均是 770 與 829 公尺，然而兩年度的標準差皆大於 140 公尺，發病間距上界平均是 8.5 與 8.9 日，大致歸納登革熱在疫情非爆發年份平均傳染半徑接近 800 可能甚至超過 1000 公尺，平均發病間距上界約 9 日。

表 11、測試資料集時空擴散區的傳染半徑與發病間距細節。各個模式在研究年份模式估計的傳染半徑與發病間距，其中傳染半徑上界是可能的傳染距離，發病間距下界是最短發病日，而上界是最長發病日，±符號是代表標準差，由於該模式每個時空擴散區是屬於不固定的時空擴散範圍，因此以此符號代表標準差。

模式	一				二			
年份	傳染半徑 (公尺)		發病間距 (日)		傳染半徑 (公尺)		發病間距 (日)	
	上界	下界	上界	下界	上界	下界	上界	下界
2010	300	0	23	6	920	0	14	6
2011	300	0	23	6	960	0	17	6
2012	300	0	23	6	810	0	11	6
2014	300	0	23	6	970	0	15	6
2015	300	0	23	6	980	0	9	6
模式	三				四			
年份	傳染半徑 (公尺)		發病間距 (日)		傳染半徑 (公尺)		發病間距 (日)	
	上界	下界	上界	下界	上界	下界	上界	下界
2010	683±185	0	11.7±0.6	6	855±75	0	11.4±1.1	6
2011	608±115	0	12.4±5.0	6	906±105	0	11.5±2.0	6
2012	429±38	0	9.5±0.9	6	837±72	0	10.9±1.1	6
2014	770±141	0	8.5±1.5	6	914±56	0	16.0±2.8	6
2015	829±145	0	8.9±0.8	6	981±22	0	10.3±1.0	6

我們推論在登革熱疫情非爆發年份的傳染半徑可能接近 800 公尺，而發病間距之最長發病日高達 11 日，疫情爆發年份的傳染半徑可能超過 1000 公尺，而發病間距之最長發病日僅有 9 日，所以登革熱在疫情爆發年份的疫情擴散現象是傳染半徑寬大，同時發病日短的現象，才會導致疫情爆發的結果。



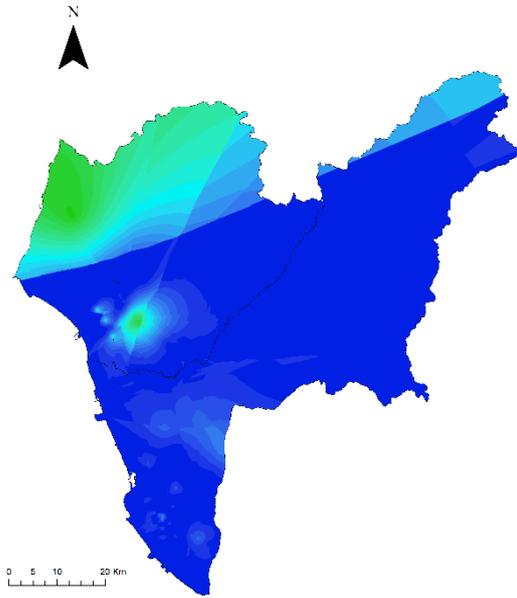
四、時空擴散區的傳染半徑與發病間距的空間分布

時空擴散區的傳染半徑與發病間距的空間分布，是根據滾動預測時空擴散區的過程，病例產生時空擴散區的傳染半徑與發病間距，其病例的發病位置的空間分布在不同的區域、環境與氣候影響下，導致擴散過程的傳染半徑與發病間距產生變化，我們試圖採用地圖繪製不同地區導致的不同的擴散特徵，並討論區域所產生的擴散特徵差異。

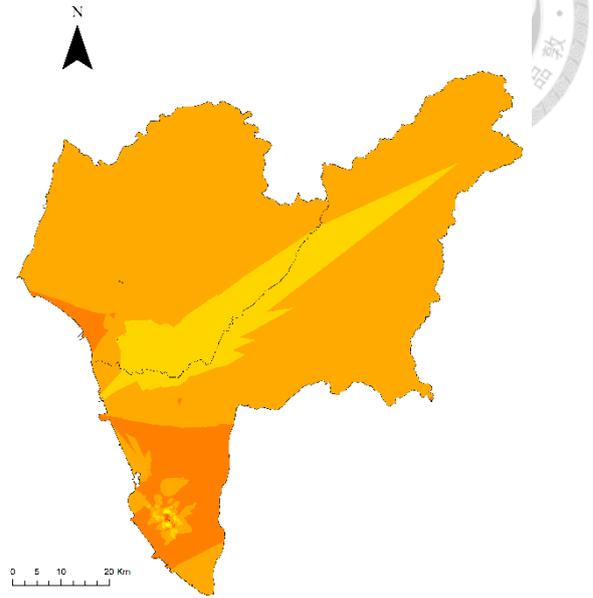
我們根據綜合指標挑選出登革熱疫情非爆發年分與爆發年分最佳的模式，我們嘗試繪製在疫情中後期採用測試資料集的擴散特徵，革熱疫情非爆發年份採用 2011 年模式四與革熱疫情爆發年份 2015 年疫情中後期模式三，測試資料集兩個月的擴散特徵位於空間的分布，其他年份 2010、2012 與 2014 年擴散特徵位於空間的分布圖在附錄一之四時空擴散區的傳染半徑與發病間距的空間分布，這兩種模式每個病例都是不固定的時空擴散區的傳染半徑與發病間距，同時又屬於該年度綜合指標評定最佳的模式，因此我們繪製傳染半徑上界與發病間距上界內插空間分布地圖⁵，期望能釐清登革熱的擴散特徵的區域差異性。

⁵ 繪製傳染半徑上界與發病間距上界內插空間分布地圖:由於模式設計，導致傳染半徑與發病間距的下界都是 0 公尺與 6 日，因此病例的預測傳染半徑與發病間距的下界都一樣，因此不需要繪製空間分布情形。

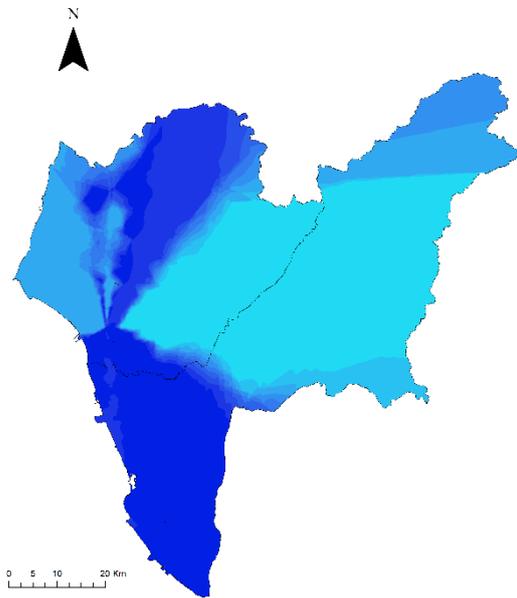
a.



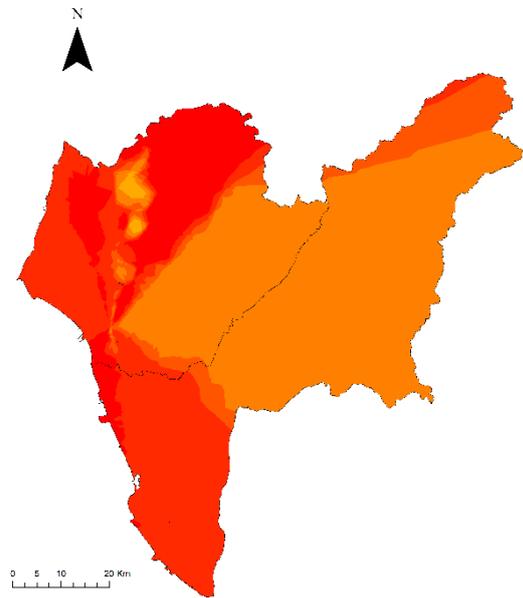
b.



c.



d.



續下頁

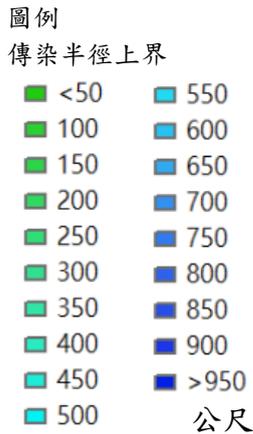


圖 29、疫情中後期時空擴散區的傳染半徑與發病間距上界組圖。a. 模式四 2011 年傳染半徑上界克利金內插空間分布、b. 模式四 2011 年發病間距上界克利金內插空間分布、c. 模式三 2015 年傳染半徑上界克利金內插空間分布、d. 模式三 2015 年發病間距上界克利金內插空間分布。a 與 c 圖是傳染半徑上界，因此應參考左圖例，單位是公尺，b 與 e 圖是發病間距上界，單位是日，請參考右圖例。

2011 年模式四時空擴散區傳染半徑(圖 29a)，病例分布在臺南市人口密集的地區、高雄市人口密集的地區、內插空間分布地圖顯示位於臺南市都會地區傳染半徑上界約 200 公尺向外的傳染半徑上界約 900 公尺以上，在臺南市西北邊傳染半徑上界僅約 100 公尺，高雄市大多數地區傳染半徑上界約 900 公尺以上，在高雄市東北方地區傳染半徑上界約 500 公尺，然而高雄市東南地區與西南沿海地區呈現傳染半徑上界約 700 至 800 公尺。

2011 年模式四擴散時空半徑的發病間距的空間分布(圖 29b)，從發病日向後推 6 日開始有新的病例發病的風險，發病間距圖所描繪的是發病間距上界，是新病例最長的發病日程，在臺南市人口密集區的中心發病間距上界僅 11 日，然而臺南市郊區大多是 12 日，疫情在臺南市區疫情相對的嚴峻，最長發病間距非常短暫，介於臺南市與高雄市邊界存在條帶約 13 日的發病間距上界，高雄市郊區大多數地區約 12 日的發病間距上界，然而在高雄市西南邊出現大片分布的發病間距上界 11 日，然而在高雄市西南沿海人口密集的都會區出現發病間距上界 14 日的現象，因此我們認為儘管同樣高雄市與臺南市人口密集區的情況，但疫情的發病間距上界卻呈現差異，臺南市在當年度的發病週期急促。

2015 年模式三時空擴散區傳染半徑(圖 29c)，病例分布在臺南市人口密集的地區、高雄市人口密集的地區、內插空間分布地圖顯示位於臺南市都會地區作為擴散傳染半徑上界分界中心，東半部約 540 公尺的傳染半徑上界，包含高雄市東北部地區，南半部傳染半徑上界約 540 公尺包含高雄市西南濱海地區，西北部地區傳染半徑上界約 640 公尺，正北方條帶傳染半徑上界約 800 公尺，其傳染半徑的空間分布，非常符合透過登革熱病例的發病位置經緯度產生的訓練模式，簡單而粗暴的歸納傳染半徑上界在地圖上的空間分布。

2015 年模式三擴散時空半徑的發病間距的空間分布(圖 29.d)，從發病日向後推 6 日開始有新的病例發病的風險，發病間距圖所描繪的是發病間距上界，是新病例最長的發病日程，在內插空間分布地圖顯示位於臺南市都會地區作為擴散發病間距的上界空間分界中心，東半部約 11 日的發病間距上界，包含高雄市東北部地區，南半部發病間距上界約 9 日包含高雄市西南濱海地區，西南部高雄與臺南市交界臨海地區僅 7 日發病間距上界，疫情可能在當地相當的急遽，西部地區靠近臺南濱海地區發病間距上界約 9 日，西北部地區條帶發病間距上界約 7 日，正北方條帶發病間距上界約 10 至 11 日，其東北邊喇叭狀的條帶發病間距上界僅約 7 日，整體而言西半部的發病間距上界極短，有部分地區的發病間距上界僅有 7 日，疫情非常的嚴峻且發病週期急促，然而研究區東半部大片區域呈現發病間距上界 11 日。

我們發現在登革熱非爆發年度 2011 年採用模式四的擴散特徵空間分布，確實能反映區位與環境氣候的差異性質，在都會地區捕捉到較短的傳染半徑上界，與較短的發病間距上界，而且會隨行政區不同，不同的都市型態獲得不同的傳染半徑與發病間距上界。我們發現在登革熱爆發年度 2015 年採用模式三的擴散特徵空間分布，重視將空間透過型的分割，臺南市作為當年度疫情的指標城市，因此模式三透過發病位置的經緯度訓練模式，粗暴的切割空間呈現放射狀，反映傳染半徑上界在臺南市都會區東、南、西、北部疫情的傳染半徑差異，與東西向的發病間距上界差異。

綜合指標挑選最佳模式，分別在登革熱疫情爆發與非爆發年份產生不同的最佳模式，登革熱非爆發時期，環境、氣候與區位會影響登革熱的擴散，然而疫情在爆發時其區位才是最重要的考量因子，這兩種不同觀點選擇，會因為疫情嚴重程度產生差異。

第五節 成果統整



表 12、模式特性與適用情境統整。

特性\模式	一	二	三	四
輸入因子	無	傳遞鏈	傳遞鏈、發病位置	傳遞鏈、發病位置、環境氣候因子
模式設計	預設傳染半徑與發病間距	歸納傳染半徑與發病間距	根據病例的發病位置	根據病例的發病位置氣候環境
誤差衰減過程	無	無	300 次迭代	300 次迭代
時空邊界特性	固定	固定	不固定	不固定
模式適用情境	不建議採用	缺乏環境氣候資料	登革熱疫情非爆發時期	登革熱疫情爆發時期

最後統整模式每日滾動預測時空擴散區的特性與適用情境，模式一是預設時空擴散區的傳染半徑與發病間距，模式沒有任何的輸入因子與訓練過程，因此沒有誤差的衰減過程，每個病例預測生成的時空擴散區是固定的傳染半徑與發病間距，然而其綜合指標結果並不理想，因此不建議在登革熱擴散預測議題採用此模式。

模式二是根據每年度疫情前期的訓練資料集歸納時空擴散區的邊界，輸入是疫情樹產生的傳遞鏈，藉由歸納傳遞鏈的傳染半徑與發病間距，沒有誤差的衰減過程，每個年份每個病例預測生成的時空擴散區是固定的傳染半徑與發病間距，然而其綜合指標結果在各年度位居二名，因此建議在缺乏登革熱環境氣候相關資料時，可藉由此模式生成擴散預測模式。

模式三是根據每年度疫情前期的訓練資料集，輸入疫情樹產生的傳遞鏈，傳遞鏈與病例的位置僅 4 個維度的輸入層，經過自組織對映類神經網路演算法 300 次迭代，因此有誤差衰減過程，其運算量略為龐大，每個病例預測生成的時空擴散區是不固定的傳染半徑與發病間距，然而其綜合指標結果在登革熱疫情爆發年分位居第三名，然而在疫情爆發年分居冠，因此建議在確認登革熱疫情爆發時期



則採用本模式，藉此模式生成擴散預測模式。

模式四是根據每年度疫情前期的訓練資料集，輸入疫情樹產生的傳遞鏈，傳遞鏈與病例的位置結合環境與氣候因子高達 45 個維度的輸入層，經過自組織對映類神經網路演算法 300 次迭代，因此有誤差衰減過程，其運算量非常龐大，每個病例預測生成的時空擴散區是不固定的傳染半徑與發病間距，然而其綜合指標結果在登革熱疫情爆發年分位居第三名，然而在疫情非爆發年分居冠，因此建議在確認登革熱疫情非爆發時期則採用本模式，藉此模式生成擴散預測模式。

根據綜合指標的評論結果，滾動預測生成時空擴散區，在登革熱疫情爆發時其應採用模式三、而在疫情非爆發時期應採用模式四，而缺乏病例相關背景資料，與有限運算資源可以採用模式二，以下模式特性與使用情境整理於表 12。

我們根據滾動預測挑選最佳模式，期的時空擴散區性質統整如下表 13，2010 年至 2012 年屬於登革熱疫情非爆發年份，採用模式四所預測的時空擴散區性質描述，大致每日的時空擴散區的時空擴散範圍小於 30 平方公里，傳染半徑上界略為 800 至 900 公尺的區間，然而發病間距上界，即發病日大多在 6 至 11 日發病區間，如採用模式四每日時空擴散區可以預測約 80% 的病例。

然而疫情爆發年分的 2014 年與 2015 年採用模式三所預測的時空擴散區性質描述，大致每日的時空擴散區的時空擴散範圍，於 2014 年高達 45.3 平方公里，2015 年高達 95.5 平方公里，傳染半徑上界大於 800 公尺，然而發病間距上界，即發病日大多在 6 至 9 日發病區間，如採用模式三每日時空擴散區約預測 90% 以上的病例。

表 13、最佳模式的時空擴散區性質統整。

年份\特性	每日時空擴散範圍	傳染半徑上界	發病間距	涵蓋率中位數
2010	26.5 平方公里	855±75 公尺	6-11.4±1.1 日	0.78
2011	22.2 平方公里	906±105 公尺	6-11.5±2.0 日	0.82
2012	29.0 平方公里	837±72 公尺	6-10.9±1.1 日	0.82
2014	45.3 平方公里	770±141 公尺	6-8.5±1.5 日	0.93
2015	95.9 平方公里	829±145 公尺	6-8.9±0.8 日	0.95

第六章 討論



第一節 擴散預測模式

一、預測時空擴散區的研究目的

預測時空擴散區的研究目的，比起重建擴散的研究可能更具有可信度，因為預測研究必須制定指標，檢核擴散的重建過程是否合理，本研究必須重新定義綜合指標評論擴散重建的優劣。

過去研究疫情擴散的過程，藉由核密度估計的模式或者藉由參數估計去找最配飾的傳染關係(Li et al., 2019; Chin et al., 2017; Guzzetta et al., 2018)，或者直接引用固定的傳染半徑與發病間距繪製疫情的時空擴散區(Kuo, Wen, and Sabel, 2018; Shah et al., 2013)，取得明確邊界的疫情擴散區與模擬傳染關係，這些研究雖然明確指出他們對疫情擴散重建的觀點與合理性，但我們難以比較這些模式適合那些地區與情境下使用。

所以我們將採用固定時空邊界環帶繪製時空擴散區(Kuo, Wen, and Sabel, 2018)，發現疫情擴散會放大、縮小、融合等現象，它的時空擴散區的設置觀點作為我們的研究的模式一，我們認為該研究區與研究時期最接近本研究，但實際應用於時空擴散的預測模式結果，根據綜合指標的檢核，表現並不如預期，該研究僅能解釋部分的登革熱群聚傳染導致疫情擴散的現象，因此我們推斷可能必須添加更多觀點認識登革熱的擴散過程。

二、影響登革熱疫情擴散的條件

在登革熱的疫情爆發期，造成登革熱疫情擴散的主要因素是區位，然而在登革熱的疫情非爆發期，造成登革熱疫情擴散的因素是環境、氣候與區位。我們透過滾動預測模式的綜合指標評論模式的時空擴散區優劣，我們得到一個複雜的結論，在登革熱疫情非爆發時期應採用模式四，然而登革熱疫情爆發時期應採用模式三。

在登革熱的疫情爆發期，造成登革熱疫情擴散的主要因素是區位，我們根據登革熱擴散研究的脈絡，我們發現群聚傳染是登革熱疫情的特色，病例的發病位置空間具有鄰近性，根據群聚傳染的特徵空間移動過程可以重建疫情的擴散模式(Wang, Wang, and Li, 2006; Wen and Tsai, 2015; Chin et al., 2017; Kuo, Wen, and

Sabel, 2018), 然而群聚傳染的過程病例彼此可能有類似的傳染半徑與發病間距, 原因可能來自於登革熱病例周遭存在病媒蚊族群(Yang et al., 2014), 病毒血清型的差異會造成不同的受感性 (Amâncio et al., 2014), 病媒蚊族群的區位分布與血清型可能是影響疫情擴散的主因。

在登革熱的疫情非爆發期, 造成登革熱疫情擴散的因素是環境、氣候與區位, 觀點來自於環境對登革熱疫情具有明顯的影響, 極端氣候高溫多雨的環境會促使病媒蚊的族群增長(Thavara et al., 2001; Watts et al., 1987), 高度人口密度地區提供病媒蚊叮咬機會, 是孕育登革熱病毒的溫床(Gubler, 2002; Sitepu et al., 2013), 影響登革熱擴散的空間特性包含人口密度、人口混合與人口流動性(Adams and Kapan, 2009), 有許多的環境影響登革熱疫情擴散過程的學說, 然而針對零星的病例或者小規模的群聚傳染, 氣候與人口條件影響登革熱擴散的因素較為顯著。

我們認為兩種觀點皆存在合理性, 疫情爆發時期大規模的病例發病位置分布區位鄰近並且於短時間內同時發生, 因此區位在疫情爆發扮演主要角色, 然而在非爆發時期由於病例的發病日不一致, 因此氣候與環境條件, 尤其是人口的型態, 可能反倒扮演重要的角色, 因此兩種觀點並不衝突。

三、登革熱疫情擴散的通則

登革熱的擴散存在通則性, 時空擴散區的傳染半徑與發病間距是可預測的, 我們發現模式二, 在各個研究年份的綜合指標穩定位居第二名, 該模式適根據疫情前期歸納傳染半徑與發病間距, 在滾動預測階段繪製各個年份每個病例固定的時空擴散區, 然而其擴散預測能力超乎預期。

我們根據文獻傳遞鏈的重建可以利用蒙地卡羅馬可夫參數估計模式估計最佳化的傳遞的傳染半徑與發病延遲(Guzzetta et al., 2018; Morelli et al., 2012), 顧名思義只要取得部分的真實傳遞鏈, 就能重建整個傳染關係, 這個觀點應該具有參考價值。

因此我們推論, 縱使缺乏太多的疫情相關背景資訊, 與相關的環境氣候資料, 我們僅獲得部分的真實登革熱傳染關係, 我們就能大致重建疫情的擴散過程, 甚至建立很好的預測擴散效果的模式。



第二節 時空擴散區的特色

一、預測登革熱疫擴散趨勢

登革熱的擴散是可以被預防的，我們只要選擇最佳預測時空擴散區模式，每日可以預測約 80% 的登革熱病例，我們可以觀察第五章成果統整表 13，參考涵蓋率中位數項目，根據模式的指引預測時空擴散區，並且按照每日時空擴散範圍的面積進行防疫措施，就可能在單日達到消除大規模病例的情形。

我們認為時序預測，準確利用氣候因子透過複迴歸加上波松迴歸預測病例數 (Hii, Zhu, et al., 2012) 缺乏空間分布，然而空間分布的研究採用年度 1 平方公里的網格 (Ong et al., 2018)，但每個網格代表當年度的風險，時間解析度過於粗糙，而臺灣的疫情期通常只有半年，時空擴散區模式的時空解析度是公尺與日，可以每日預測約 80% 的登革熱病例數的高風險區，儘管模式的預測能力無法與其他模式進行比較，可能略低於這些經典模式的預測能力。

然而在實際應用上，由於時空擴散區時空解析度精細，因此可以每日都立即預測登革熱擴散的趨勢，可以做到立即預防，阻絕登革熱傳遞的發生，短時間提升預測模式應用在疫情管控的實用性，有效縮小登革熱疫情打擊範圍。

二、登革熱的時空擴散區細節

時空擴散區的設置我們建議登革熱的孳生源清除空間半徑應該設定超過 400 公尺，時間半徑的發病間距依循過去文獻常見設定 6 至 20 日以內合適，世界衛生組織建議在登革熱發病病例的半徑 400 公尺內從事孳生源清除，採用內生潛伏期 4 至 10 日後關注新的病例 (WHO, 2009a)，但本研究實際運算出最佳模式的時空擴散區的傳染半徑，疫情爆發年份之 2014 年與 2015 年的傳染半徑上界是 800 公尺至 1000 公尺，疫情非爆發年份之 2010 至 2011 年的傳染半徑上界平均接近 800 公尺。

我們不排除風力與環境影響蚊子飛行的距離接近甚至超過 800 公尺的說法 (Regilme et al., 2021; Honório et al., 2003)，過去的捕蚊捉放法研究是建構，在環境外在因素干擾最小的情況處理病媒的飛行距離，得到病媒蚊的飛行非常短不易存活 (Shannon and Davis, 1930) 的結論，臺灣存在病媒蚊的種類 (Tsai and Teng, 2016)，病媒的飛行距離最遠 500 公尺 (Verdonschot and Besse-Lototskaya, 2014)，因此我們的傳染半徑框架始終被限制在 500 公尺內，但根據本研究成果我們推論



傳染半徑即等於清消半徑是 500 至 1000 公尺合適，本研究比較接受清除半徑 1 公里的相關文獻(表 14、清除半徑與相關文獻論點。)，風力可能稍微改變病媒蚊的飛行距離 (Honório et al., 2003)，擴散群聚限制在 1 公里內(Guzzetta et al., 2018; Kan et al., 2008)。

表 14、清除半徑與相關文獻論點。

清除半徑	相關文獻	論點
小於 500 公尺	Liew and Curtis, 2005 Harrington et al., 2005 Honório et al., 2009 Shah et al., 2013 Kuo, Wen, and Sabel, 2018 Juarez et al., 2020 Verdonschot and Besse-Lototskaya, 2014 Shannon and Davis, 1930	病媒蚊的飛行非常短，不易存活 (Shannon and Davis, 1930)。
小於或接近 1 公里	Verdonschot and Besse-Lototskaya 2014 Regilme et al., 2021 Honório et al., 2003 Guzzetta et al., 2018 Kan et al., 2008	風力與微氣候改變提高病媒蚊的飛行距離(Honório et al., 2003)。 擴散群聚限制在 1 公里內 (Guzzetta et al., 2018; Kan et al., 2008)。
遠大於 1 公里	Verdonschot and Besse-Lototskaya, 2014	迴歸分析理論值，傳染半徑是極端值的現象(Verdonschot and Besse-Lototskaya, 2014)。

三、時空擴散區的空間分布

登革熱的擴散時空半徑的空間分布具有空間與環境的異質性，都市的疫情比起郊區更容易控制，時空擴散區的傳染半徑與發病間距，位於地圖合乎自組織對映的拓撲圖模擬環境空間非線性的數值分布給予最合理的解釋(Vesanto and Alhoniemi, 2000; C. LourenÇO, Lobo, and Bação, 2018; Smith, 2002)。



我們觀察圖 29a、c 模式四 2011 年傳染半徑上界內插圖，我們發現傳染半徑會有伴隨人口密度高的地區縮短的現象。觀察圖 29b、d 發病間距的空間分布特徵，在高人口密度地區發病間距偏低。我們首先認為登革熱的疫情在人口密集區擴散的傳染半徑偏低與發病間距偏低的現象，來自於人造建物阻擋病媒蚊飛行的距離，導致登革熱疫情無法廣泛傳播，實際上根據登革熱水平與垂直傳播的研究發現，垂直的建物對於病媒蚊不構成阻擋效應；垂直建物各樓層的病媒蚊標定移動數量，與無遮擋的標定移動數量並不顯著(Liew and Curtis, 2005)，病媒蚊在實驗室條件下僅有 4 日壽命，推定是缺乏食物來源(Liew and Curtis, 2005)，轉向病媒蚊的生態學觀點觀察，比較合理的解釋情況農村地區的登革熱風險高於城市地區，缺乏自來水地區，儲水容器的使用為蚊子提供了繁殖場所，導致蚊子與人的比例很高(Schmidt et al., 2011)，水平建築地區登革熱風險高於垂直建築地區，縱使垂直建築地區高人口的現象(Barreto et al., 2008)，此結論依然指向水平建築地區提供病媒蚊孳生源，高人口密度地區，垂直性建物多，缺乏病媒蚊孳生源的病媒蚊導致族群縮小解釋傳染半徑變短現象。

傳染半徑與發病間距會有伴隨人口密度高的地區縮短，反映在防疫措施上郊區應該擴大孳生源清除範圍，而都市地區應該快速孳生源清除可能有效的避免登革熱疫情的擴散。

第三節 研究限制

缺乏實際的登革熱傳染關係，我們模式所建構的疫情樹產生的傳遞鏈，僅能透過 MST-DBSCAN 演算法實作，不是真實的傳染關係，導致無法做最佳化傳染半徑與發病間距的數值，無法重建真實的擴散過程，受制於 DBSCAN 的貪婪法缺陷，導致生成的時空擴散區無法估計發病間距的下界，如果有真實的傳染關係，相信模式二、模式三和模式四的滾動預測時空擴散區表現會更好。

缺乏精確的病例發病位置，由於個人資料的機敏性，導致我們無法取得病例的發病位置確切的經緯度座標，我們僅能取得發病位置最小統計區中心的座標，因此我們無法確保滾動預測生成時空擴散區的傳染半徑是否正確，然而幸運的是當我們選擇最佳的模式，幾乎能確保在疫情中後期時空擴散區涵蓋約 80% 的病例，因此本研究挑選出來的擴散預測模式，依然具備參考與應用的價值。

第七章 結論

我們確實建立的多個每日滾動預測時空擴散區的模式，並且也定義綜合指標驗證模式的優劣，我們也發現模式三與模式四透過自組織對映類神經網路共識決的模式優於其他滾動預測模式。

但我們也發現到疫情爆發時期，區位主要影響疫情擴散的趨勢，然而在非爆發時期氣候與環境條件包含人口型態影響擴散的趨勢，因此如果在登革熱爆發年份，應當採取模式三，根據發病病例的病例位置訓練模式，預測登革熱的擴散趨勢，反之在登革熱疫情非爆發時期，應當採取模式四，考慮區位、環境與氣候因子，預測登革熱的擴散趨勢。

儘管我們面臨缺乏登革熱的傳染關係與確切的發病位置，然而我們建立的每日滾動預測模式，其時空解析度細緻，可以即時的預測登革熱的擴散動向，同時在登革熱疫情的中後期，幾乎可以保證預測的時空擴散區涵蓋約 80% 以上的登革熱病例，我們更能肯定本研究能提升預測模式應用在疫情管控的實用性，有效縮小登革熱疫情打擊範圍。

參考文獻

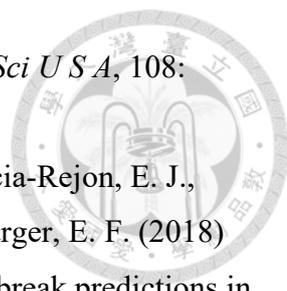


- 楊芝青 (2018) 因應氣候變遷氣候變遷之健康衝擊政策白皮書，台北市：衛生福利部。
- Aburas, H. M., Cetiner, B. G., and Sari, M. (2010) Dengue confirmed-cases prediction: A neural network model, *Expert Systems with Applications*, 37: 4256-4260.
- Ackley, D. H., Hinton, G. E., and Sejnowski, T. J.(1988).'A learning algorithm for boltzmann machines.' in, *Connectionist models and their implications: Readings from cognitive science* (Ablex Publishing Corp.).
- Adams, B., and Kapan, D. D. (2009) Man bites mosquito: Understanding the contribution of human movement to vector-borne disease dynamics, *PLoS One*, 4: e6763.
- Aggarwal, R., and Yonghua, S. (1998) Artificial neural networks in power systems. Iii. Examples of applications in power systems, *Power Engineering Journal*, 12: 279-287.
- Althouse, B. M., Ng, Y. Y., and Cummings, D. A. T. (2011) Prediction of dengue incidence using search query surveillance, *PLoS Negl Trop Dis*, 5: e1258.
- Amâncio, F. F., Ferraz, M. L., Almeida, M. C. d. M., Pessanha, J. E. M., Iani, F. C. M., Fraga, G. L., Lambertucci, J. R., and Carneiro, M. (2014) Dengue virus serotype 4 in a highly susceptible population in southeast brazil, *Journal of Infection and Public Health*, 7: 547-552.
- Bação, F., Lobo, V., and Painho, M. (2005) The self-organizing map, the geo-som, and relevant variants for geosciences, *Computers & Geosciences*, 31: 155-163.
- Baquero, O. S., Santana, L. M. R., and Chiaravalloti-Neto, F. (2018) Dengue forecasting in são paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models, *PLoS One*, 13: e0195065.
- Barreto, F. R., Teixeira, M. G., Costa, M. d. C. N., Carvalho, M. S., and Barreto, M. L. (2008) Spread pattern of the first dengue epidemic in the city of salvador, brazil, *BMC Public Health*, 8: 51.
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., and Hay, S. I. (2013) The global distribution and burden of dengue, *Nature*, 496:

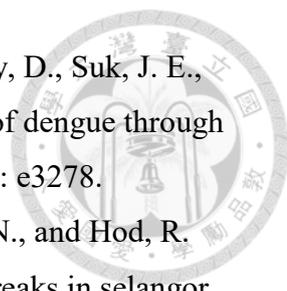
- 504.
- Birant, D., and Kut, A. (2007) St-dbscan: An algorithm for clustering spatial-temporal data 60: 208-221.
- C. LourenÇO, F., Lobo, V., and Bação, F.(2018) *Exploratory geospatial data analysis using self-organizing maps case study of portuguese mainland regions.*
- Calderón-Arguedas, O., Troyo, A., Solano, M. E., Avendaño, A., and Beier, J. C. (2009) Urban mosquito species (diptera: Culicidae) of dengue endemic communities in the greater puntarenas area, costa rica *revista de biología tropical*, 57: 1223-1234.
- Campbell, L. P., Luther, C., Moo-Llanes, D., Ramsey, J. M., Danis-Lozano, R., and Peterson, A. T. (2015) Climate change influences on global distributions of dengue and chikungunya virus vectors, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370: 20140135.
- Carvalho, S., Magalhães, M. d. A. F. M., and Medronho, R. d. A. (2017) Analysis of the spatial distribution of dengue cases in the city of rio de janeiro, 2011 and 2012, *Revista de saude publica*, 51: 79-79.
- Chan, M., and Johansson, M. A. (2012) The incubation periods of dengue viruses, *PLoS One*, 7: e50972-e50972.
- Chang, F.-J., Chang, L.-C., Kao, H.-S., and Wu, G.-R. (2010) Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network, *Journal of Hydrology*, 384: 118-129.
- Chang, L.-C., Shen, H.-Y., and Chang, F.-J. (2014) Regional flood inundation nowcast using hybrid som and dynamic neural networks, *Journal of Hydrology*, 519: 476-489.
- Changal, K. H., Raina, A. H., Raina, A., Raina, M., Bashir, R., Latief, M., Mir, T., and Changal, Q. H. (2016) Differentiating secondary from primary dengue using igg to igm ratio in early dengue: An observational hospital based clinico-serological study from north india, *BMC Infectious Diseases*, 16: 715-715.
- Cheong, Y. L., Burkart, K., Leitão, P. J., and Lakes, T. (2013) Assessing weather effects on dengue disease in malaysia, *International journal of environmental research and public health*, 10: 6319-6334.
- Chien, L. C., and Yu, H. L. (2014) Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence, *Environ Int*, 73: 46-56.
- Chin, W.-C.-B., Wen, T.-H., Sabel, C. E., and Wang, I. H. (2017) A geo-computational

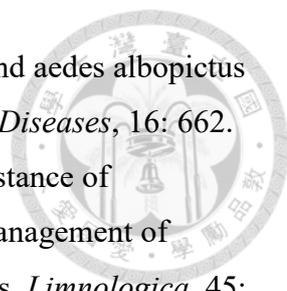
- 
- algorithm for exploring the structure of diffusion progression in time and space, *Scientific Reports*, 7: 12565.
- Codeço, C. T., Villela, D. A. M., and Coelho, F. C. (2018) Estimating the effective reproduction number of dengue considering temperature-dependent generation intervals, *Epidemics*, 25: 101-111.
- Ebi, K. L., and Nealon, J. (2016) Dengue in a changing climate, *Environmental Research*, 151: 115-123.
- Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–231. Portland, Oregon: AAAI Press.
- Genest, C., and Zidek, J. V. (1986) Combining probability distributions: A critique and an annotated bibliography, *Statistical Science*, 1: 114-135, 122.
- Gomide, J., Veloso, A., Wagner Meira, J., Almeida, V., Benevenuto, F., Ferraz, F., and Teixeira, M. (2011) Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *Proceedings of the 3rd International Web Science Conference*, 1-8. Koblenz, Germany: ACM.
- Gubler, D. J. (2002) Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century, *Trends in Microbiology*, 10: 100-103.
- Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., Luo, G., Li, Z., He, J., Zhang, Y., and Ma, W. (2017) Developing a dengue forecast model using machine learning: A case study in china, *PLoS Negl Trop Dis*, 11: e0005973.
- Guzzetta, G., Marques-Toledo, A., C., Rosà, R., Teixeira, M., and Merler, S. (2018) Quantifying the spatial spread of dengue in a non-endemic brazilian metropolis via transmission chain reconstruction, *Nature Communications*, 9: 2837.
- Haydon, D. T., Chase-Topping, M., Shaw, D. J., Matthews, L., Friar, J. K., Wilesmith, J., and Woolhouse, M. E. J. (2003) The construction and analysis of epidemic trees with reference to the 2001 uk foot-and-mouth outbreak, *Proceedings. Biological sciences*, 270: 121-127.
- Hii, Y. L., Rocklöv, J., Wall, S., Ng, L. C., Tang, C. S., and Ng, N. (2012) Optimal lead time for dengue forecast, *PLoS Negl Trop Dis*, 6: e1848.
- Hii, Y. L., Rocklöv, J., Ng, N., Tang, C. S., Pang, F. Y., and Sauerborn, R. (2009) Climate variability and increase in intensity and magnitude of dengue incidence

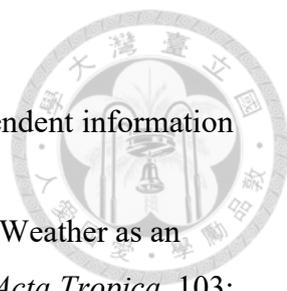
- 
- in singapore, *Global health action*, 2: 10.3402/gha.v3402i3400.2036.
- Hii, Y. L., Zhu, H., Ng, N., Ng, L. C., and Rocklöv, J. (2012) Forecast of dengue incidence using temperature and rainfall, *PLoS Negl Trop Dis*, 6: e1908.
- Ho, C. C., and Yee, T. (2015) *Time series analysis and forecasting of dengue using open data*.
- Honório, N. A., Silva Wda, C., Leite, P. J., Gonçalves, J. M., Lounibos, L. P., and Lourenço-de-Oliveira, R. (2003) Dispersal of *aedes aegypti* and *aedes albopictus* (diptera: Culicidae) in an urban endemic dengue area in the state of rio de janeiro, brazil, *Mem Inst Oswaldo Cruz*, 98: 191-198.
- Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences of the United States of America*, 79: 2554-2558.
- Hsieh, Y.-H. (2018) Dengue outbreaks in taiwan, 1998-2017: Importation, serotype and temporal pattern, *Asian Pacific Journal of Tropical Medicine*, 11: 460-466.
- Hsu, J. C., Hsieh, C.-L., and Lu, C. Y. (2017) Trend and geographic analysis of the prevalence of dengue in taiwan, 2010–2015, *International Journal of Infectious Diseases*, 54: 43-49.
- Ioffe, S., and Szegedy, C. (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, 448-456. Lille, France: JMLR.org.
- Kalayanarooj, S. (2011) Clinical manifestations and management of dengue/dhf/dss, *Trop Med Health*, 39: 83-87.
- Kan, C. C., Lee, P. F., Wen, T. H., Chao, D. Y., Wu, M. H., Lin, N. H., Huang, S. Y., Shang, C. S., Fan, I. C., Shu, P. Y., Huang, J. H., King, C. C., and Pai, L. (2008) Two clustering diffusion patterns identified from the 2001-2003 dengue epidemic, kaohsiung, taiwan, *Am J Trop Med Hyg*, 79: 344-352.
- Kohonen, T. (1990) The self-organizing map, *Proceedings of the IEEE*, 78: 1464-1480.
 —— (2001) Self-organizing maps, 3rd: 105-176.
- Kuo, F.-Y., Wen, T.-H., and Sabel, C. E. (2018) Characterizing diffusion dynamics of disease clustering: A modified space–time dbscan (mst-dbscan) algorithm, *Annals of the American Association of Geographers*, 108: 1168-1186.
- Lambrechts, L., Paaijmans, K. P., Fansiri, T., Carrington, L. B., Kramer, L. D., Thomas, M. B., and Scott, T. W. (2011) Impact of daily temperature fluctuations on

- 
- dengue virus transmission by aedes aegypti, *Proc Natl Acad Sci U S A*, 108: 7460-7465.
- Laureano-Rosario, E. A., Duncan, P. A., Mendez-Lazaro, A. P., Garcia-Rejon, E. J., Gomez-Carro, S., Farfan-Ale, J., Savic, A. D., and Muller-Karger, E. F. (2018) Application of artificial neural networks for dengue fever outbreak predictions in the northwest coast of yucatan, mexico and san juan, puerto rico, *Tropical Medicine and Infectious Disease*, 3.
- Li, M., Shi, X., Li, X., Ma, W., He, J., and Liu, T. (2019) Epidemic forest: A spatiotemporal model for communicable diseases, *Annals of the American Association of Geographers*, 109: 812-836.
- Liebig, J., Jansen, C., Paini, D., Gardner, L., and Jurdak, R. (2019) A global model for predicting the arrival of imported dengue infections, *PLoS One*, 14: e0225193.
- Liew, C., and Curtis, C. (2005) Horizontal and vertical dispersal of dengue vector mosquitoes, aedes aegypti and aedes albopictus, in singapore, *Medical and veterinary entomology*, 18: 351-360.
- Lippmann, R. (1987) An introduction to computing with neural nets, *IEEE ASSP Magazine*, 4: 4-22.
- Lowe, R., Bailey, T. C., Stephenson, D. B., Graham, R. J., Coelho, C. A. S., Sá Carvalho, M., and Barcellos, C. (2011) Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in brazil, *Computers & Geosciences*, 37: 371-381.
- Méndez-Lázaro, P., Muller-Karger, F. E., Otis, D., McCarthy, M. J., and Peña-Orellana, M. (2014) Assessing climate variability effects on dengue incidence in san juan, puerto rico, *International journal of environmental research and public health*, 11.
- Matangkasombut, P., Manopwisedjaroen, K., Pitabut, N., Thaloengsok, S., Suraamornkul, S., Yingtaweesak, T., Duong, V., Sakuntabhai, A., Paul, R., and Singhasivanon, P. (2020) Dengue viremia kinetics in asymptomatic and symptomatic infection, *International Journal of Infectious Diseases*, 101: 90-97.
- Morelli, M. J., Thébaud, G., Chadœuf, J., King, D. P., Haydon, D. T., and Soubeyrand, S. (2012) A bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data, *PLOS Computational Biology*, 8: e1002768.
- Morin, C. W., Comrie, A. C., and Ernst, K. (2013) Climate and dengue transmission:

- Evidence and implications, *Environ Health Perspect*, 121: 1264-1272.
- Mu-Chun, S., DeClaris, N., and Ta-Kang, L.(1997) Application of neural networks in cluster analysis. In *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 1-6 vol.1.
- Napp, S., Allepuz, A., Purse, B. V., Casal, J., García-Bocanegra, I., Burgin, L. E., and Searle, K. R. (2016) Understanding spatio-temporal variability in the reproduction ratio of the bluetongue (btv-1) epidemic in southern spain (andalusia) in 2007 using epidemic trees, *PLoS One*, 11: e0151151.
- Nishiura, H., and Halstead, S. B. (2007) Natural history of dengue virus (denv)—1 and denv—4 infections: Reanalysis of classic studies, *The Journal of Infectious Diseases*, 195: 1007-1013.
- Ong, J., Liu, X., Rajarethinam, J., Kok, S. Y., Liang, S., Tang, C. S., Cook, A. R., Ng, L. C., and Yap, G. (2018) Mapping dengue risk in singapore using random forest, *PLoS Negl Trop Dis*, 12: e0006587-e0006587.
- Othman, W. N., Ahmad Nazni, W., Noramiza, S., Shafa'ar-Ko'ohar, S. e., Sock Kiang, H., Halim Nor-Azlina, A., Khairul-Asuad, M., and Lim Lee, H.(2012) *Distribution of aedes mosquitoes in three selected localities in malaysia.*
- Pan, C.-Y., Liu, W.-L., Su, M.-P., Chang, T.-P., Ho, H.-P., Shu, P.-Y., Huang, J.-J., Lin, L.-J., and Chen, C.-H. (2020) Epidemiological analysis of the kaohsiung city strategy for dengue fever quarantine and epidemic prevention, *BMC Infectious Diseases*, 20: 347.
- Parimala, M., Lopez, D., and N C, S. (2011) A survey on density based clustering algorithms for mining large spatial databases, *International Journal of Advanced Science and Technology*, 31.
- Platt, K. B., Linthicum, K. J., Myint, K. S., Innis, B. L., Lerdthusnee, K., and Vaughn, D. W. (1997) Impact of dengue virus infection on feeding behavior of aedes aegypti, *Am J Trop Med Hyg*, 57: 119-125.
- Regilme, M. A. F., Carvajal, T. M., Honnen, A. C., Amalin, D. M., and Watanabe, K. (2021) The influence of roads on the fine-scale population genetic structure of the dengue vector aedes aegypti (linnaeus), *PLoS Negl Trop Dis*, 15: e0009139.
- Schmidt, W.-P., Suzuki, M., Dinh Thiem, V., White, R. G., Tsuzuki, A., Yoshida, L.-M., Yanai, H., Haque, U., Huu Tho, L., Anh, D. D., and Ariyoshi, K. (2011) Population density, water supply, and the risk of dengue fever in vietnam: Cohort study and spatial analysis, *PLOS Medicine*, 8: e1001082.

- 
- Semenza, J. C., Sudre, B., Miniota, J., Rossi, M., Hu, W., Kossowsky, D., Suk, J. E., Van Bortel, W., and Khan, K. (2014) International dispersal of dengue through air travel: Importation risk for Europe, *PLoS Negl Trop Dis*, 8: e3278.
- Shah, S. A., Sani, J. A. M., Hassan, M. R., Safian, N., Aizuddin, A. N., and Hod, R. (2013) Relationships between aedes indices and dengue outbreaks in Selangor, Malaysia, *Dengue Bulletin*, 36: 166-174.
- Shannon, R. C., and Davis, N. C. (1930) The flight of *Stegomyia aegypti* (L.)1, *American Journal of Tropical Medicine and Hygiene*, 10: 151-156.
- Shaukat, K., Masood, N., Shafaat, A. B., Jabbar, K., Shabbir, H., and Shabbir, S. J. C. (2015) Dengue fever in perspective of clustering algorithms, abs/1511.07353.
- Shaw, G. L. (1986) Donald Hebb: The organization of behavior. In *Brain Theory*, edited by Günther Palm and Ad Aertsen, 231-233. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Simmons, C. P., Farrar, J. J., van Vinh Chau, N., and Wills, B. (2012) Dengue, *New England Journal of Medicine*, 366: 1423-1432.
- Sipser, M. (2006) *Introduction to the theory of computation* (International Thomson Publishing: Boston, Massachusetts).
- Sitepu, M. S., Kaewkungwal, J., Luplerdlop, N., Soonthornworasiri, N., Silawan, T., Pongsombat, S., and Lawpoolsri, S. (2013) Temporal patterns and a disease forecasting model of dengue hemorrhagic fever in Jakarta based on 10 years of surveillance data, *Southeast Asian J Trop Med Public Health*, 44: 206-217.
- Smith, A. J. (2002) Applications of the self-organising map to reinforcement learning, *Neural Networks*, 15: 1107-1124.
- Sun, B., Wang, C., Yang, C., Xu, B., Zhou, G., Li, X., Xie, J., Xu, S., Liu, B., Xie, T., Kuai, J., and Zhang, J. (2021) Retrieval of rapeseed leaf area index using the prosail model with canopy coverage derived from UAV images as a correction parameter, *International Journal of Applied Earth Observation and Geoinformation*, 102: 102373.
- Sung-Bae, C. (1997) Neural-network classifiers for recognizing totally unconstrained handwritten numerals, *IEEE Transactions on Neural Networks*, 8: 43-53.
- Thavara, U., Tawatsin, A., Chansang, C., Kong-ngamsuk, W., Paosriwong, S., Boon-Long, J., Rongsriyam, Y., and Komalamisra, N. (2001) Larval occurrence, oviposition behavior and biting activity of potential mosquito vectors of dengue on Samui Island, Thailand, *J Vector Ecol*, 26: 172-180.

- 
- Tsai, P.-J., and Teng, H.-J. (2016) Role of *aedes aegypti* (linnaeus) and *aedes albopictus* (skuse) in local dengue epidemics in taiwan, *BMC Infectious Diseases*, 16: 662.
- Verdonschot, P. F. M., and Besse-Lototskaya, A. A. (2014) Flight distance of mosquitoes (culicidae): A metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands, *Limnologica*, 45: 69-79.
- Vesanto, J., and Alhoniemi, E. (2000) Clustering of the self-organizing map, *IEEE Transactions on Neural Networks*, 11 3: 586-600.
- Wang, M., Wang, A., and Li, A.(2006) Mining spatial-temporal clusters from geodatabases. In *Proceedings of the Second international conference on Advanced Data Mining and Applications*, 263-270. Xi'an, China: Springer-Verlag.
- Wang, S.-F., Chang, K., Loh, E.-W., Wang, W.-H., Tseng, S.-P., Lu, P.-L., Chen, Y.-H., and Chen, Y.-M. A. (2016) Consecutive large dengue outbreaks in taiwan in 2014-2015, *Emerging microbes & infections*, 5: e123-e123.
- Watts, D. M., Burke, D. S., Harrison, B. A., Whitmire, R. E., and Nisalak, A. (1987) Effect of temperature on the vector efficiency of *aedes aegypti* for dengue 2 virus, *Am J Trop Med Hyg*, 36: 143-152.
- Wen, T.-H., Lin, M.-H., and Fang, C.-T. (2012) Population movement and vector-borne disease transmission: Differentiating spatial-temporal diffusion patterns of commuting and noncommuting dengue cases, *Annals of the Association of American Geographers*, 102: 1026-1037.
- Wen, T.-H., and Tsai, Y.-S.(2015). 'Analyzing the patterns of space-time distances for tracking the diffusion of an epidemic.' in Mei-Po Kwan, Douglas Richardson, Donggen Wang and Chenghu Zhou (eds.), *Space-time integration in geography and giscience: Research frontiers in the us and china* (Springer Netherlands: Dordrecht).
- WHO.(2009a) Dengue guidelines for diagnosis, treatment, prevention and control : New edition. In. Geneva: World Health Organization.
- . (2009b). 'Who guidelines approved by the guidelines review committee.' in, *Dengue: Guidelines for diagnosis, treatment, prevention and control: New edition* (World Health Organization
- World Health Organization.: Geneva).
- (2012) Global strategy for dengue prevention and control, 2012–2020, *Geneva*, 6: 43.

- 
- . (2015) Global alert and response—impact of dengue. In.
- Winkler, R. L. (1981) Combining probability distributions from dependent information sources, *Management Science*, 27: 479-488.
- Wu, P.-C., Guo, H.-R., Lung, S.-C., Lin, C.-Y., and Su, H.-J. (2007) Weather as an effective predictor for occurrence of dengue fever in taiwan, *Acta Tropica*, 103: 50-57.
- Wu, Y., Lee, G., Fu, X., and Hung, T. G. G. (2008) *Detect climatic factors contributing to dengue outbreak based on wavelet, support vector machines and genetic algorithm.*
- Yamaguchi, M., Akiyama, K., Tsukagoshi, T., Muto, T., Kataoka, A., Tazaki, F., Ikeda, S., Fukagawa, M., Honma, M., and Kawabe, R. (2020) Super-resolution imaging of the protoplanetary disk hd 142527 using sparse modeling, *The Astrophysical Journal*, 895: 84.
- Yang, C.-F., Hou, J.-N., Chen, T.-H., and Chen, W.-J. (2014) Discriminable roles of aedes aegypti and aedes albopictus in establishment of dengue outbreaks in taiwan, *Acta Tropica*, 130: 17-23.
- Yuan, H.-Y., Liang, J., Lin, P.-S., Sucipto, K., Tsegaye, M. M., Wen, T.-H., Pfeiffer, S., and Pfeiffer, D. (2020) The effects of seasonal climate variability on dengue annual incidence in hong kong: A modelling study, *Scientific Reports*, 10: 4297.

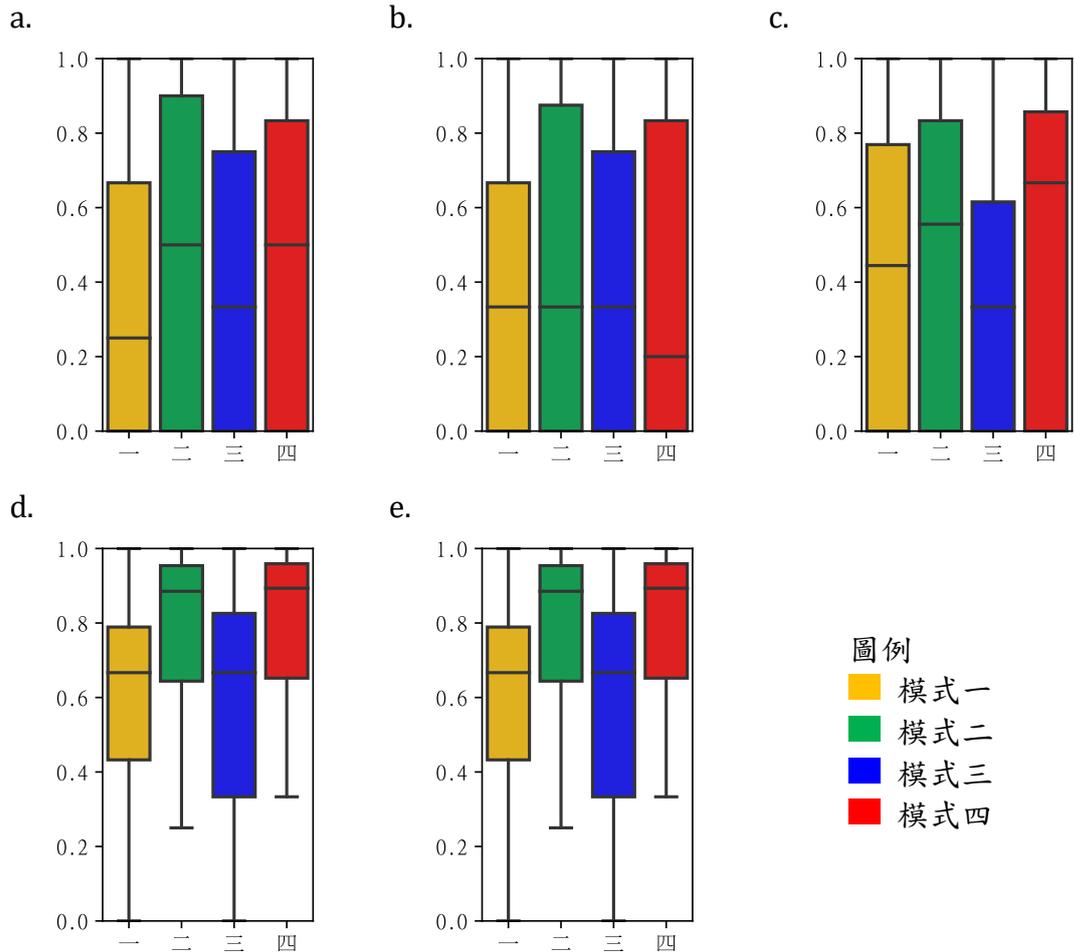
附錄



附錄一 研究成果補充

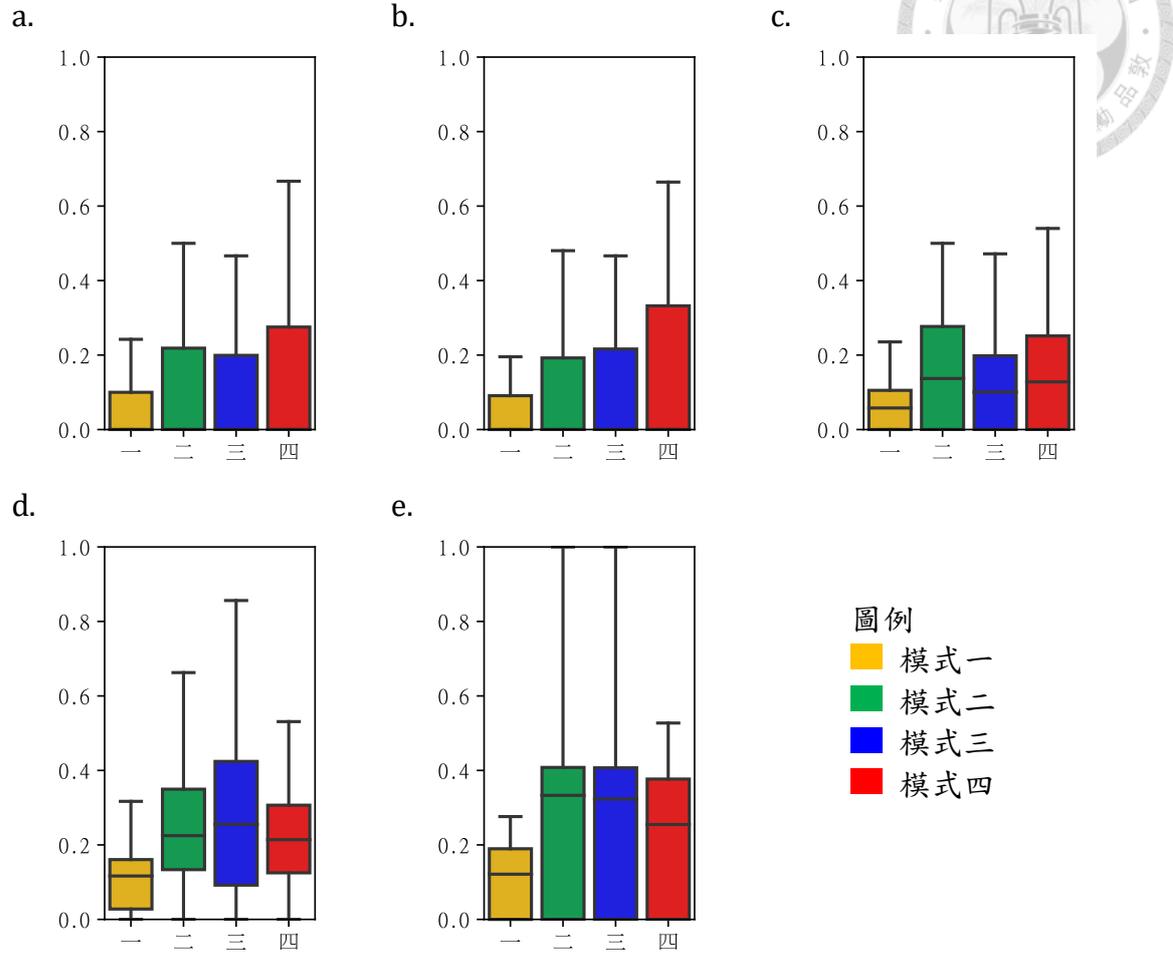
一、訓練資料集綜合指標細節

(1) 涵蓋率



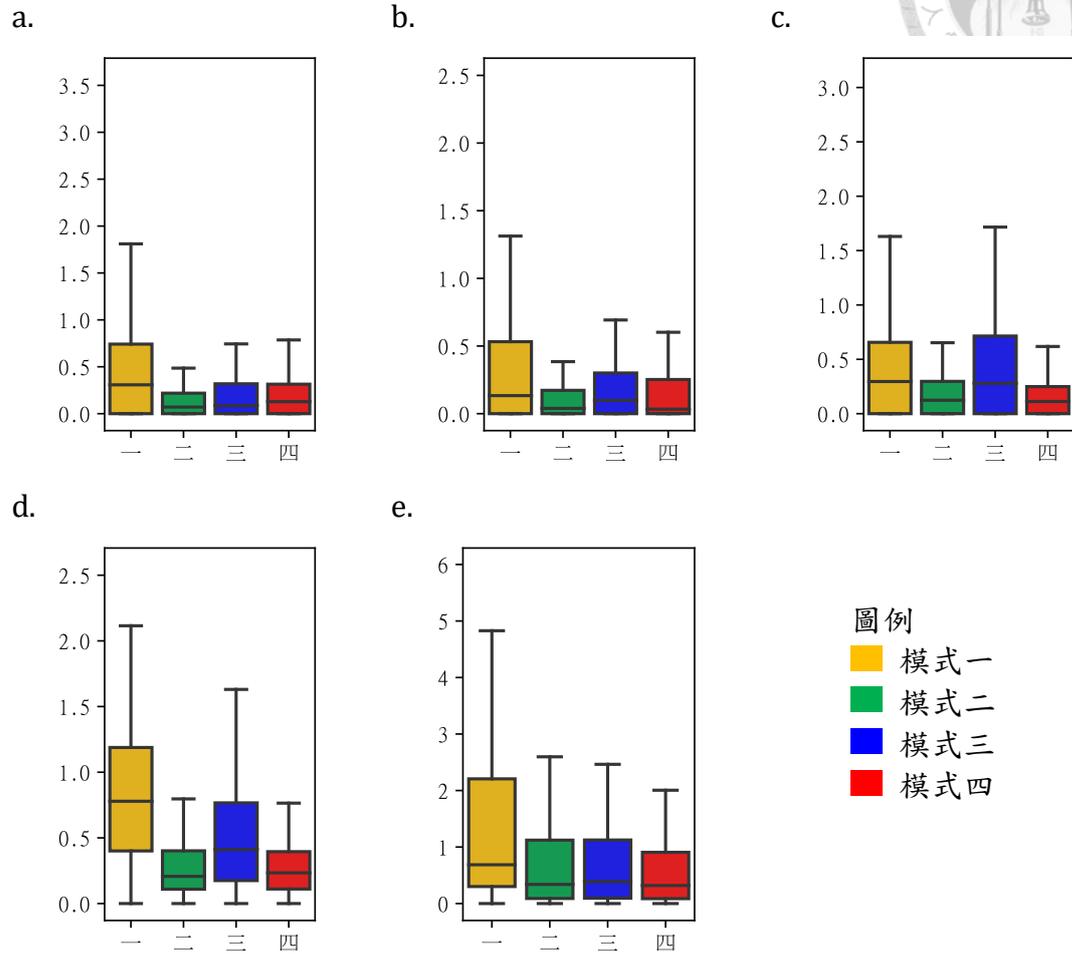
附錄圖 1、訓練資料集的每日涵蓋率分布盒狀圖。a 至 e 圖分別是 2010、2011、2012、2014 與 2015 年的四種模式合適度，模式一至模式四分別由黃、綠藍和紅設色。

(2) 合適度



附錄圖 2、訓練資料集的每日合適度分布盒狀圖。a 至 e 圖分別是 2010、2011、2012、2014 與 2015 年的四種模式合適度，模式一至模式四分別由黃、綠藍和紅設色。

二、訓練資料集時空擴散區的病例密度



附錄圖 3、訓練資料集時空擴散區的病例密度。a-e 圖分別是 2010、2011、2012、2014 與 2015 年的四種模式的時空擴散區的病例密度，單位是每日每一平方公里涵蓋的病例數 $\text{case}/(\text{km}^2 \text{ day})$ ，模式一至模式四分別由黃、綠藍和紅設色。



三、訓練資料集時空擴散區的傳染半徑與發病間距細節

附錄表 1、訓練資料集時空擴散區的傳染半徑與發病間距細節。各個模式在研究年份模式估計的傳染半徑與發病間距，其中傳染半徑上界是可能的傳染距離，發病間距下界是最短發病日，而上界是最長發病日，±符號是代表標準差，由於該模式每個時空擴散區是屬於不固定的時空擴散範圍，因此以此符號代表標準差。

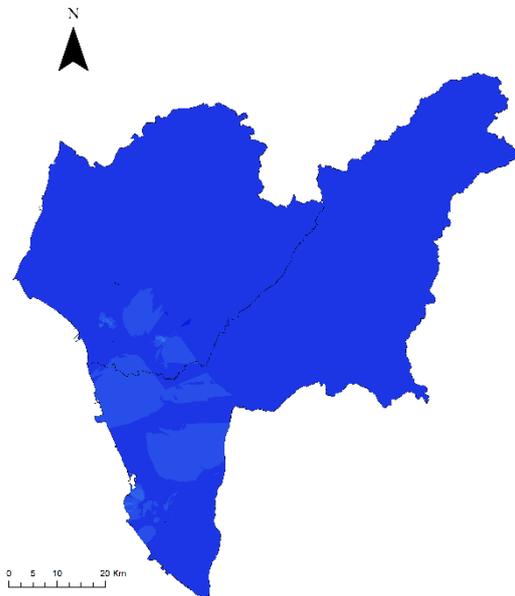
模式	一				二			
	傳染半徑 (公尺)		發病間距 (日)		傳染半徑 (公尺)		發病間距 (日)	
	上界	下界	上界	下界	上界	下界	上界	下界
2010	300	0	23	6	920	0	14	6
2011	300	0	23	6	960	0	17	6
2012	300	0	23	6	810	0	11	6
2014	300	0	23	6	970	0	15	6
2015	300	0	23	6	980	0	9	6

模式	三				四			
	傳染半徑 (公尺)		發病間距 (日)		傳染半徑 (公尺)		發病間距 (日)	
	上界	下界	上界	下界	上界	下界	上界	下界
2010	697±186	0	11.7±0.6	6	841±108	0	11.0±1.4	6
2011	597±108	0	11.8±5.8	6	830±169	0	12.0±3.2	6
2012	430±42	0	9.3±0.9	6	865±88	0	12.0±1.2	6
2014	742±146	0	8.7±1.5	6	966±36	0	14.3±4.5	6
2015	780±144	0	8.8±0.9	6	985±23	0	10.2±1.3	6

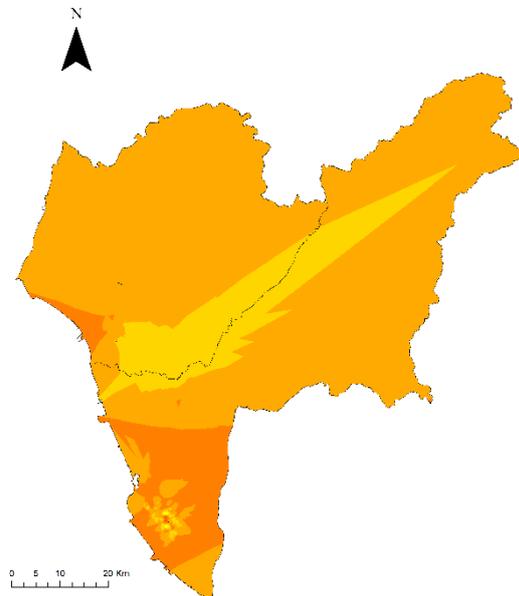
四、時空擴散區的傳染半徑與發病間距的空間分布



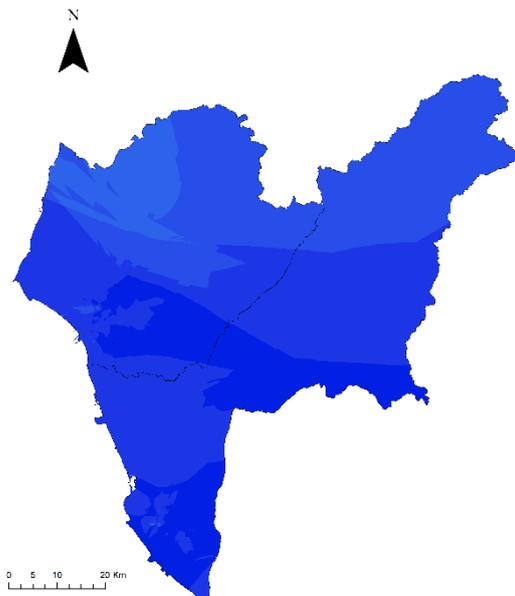
a.



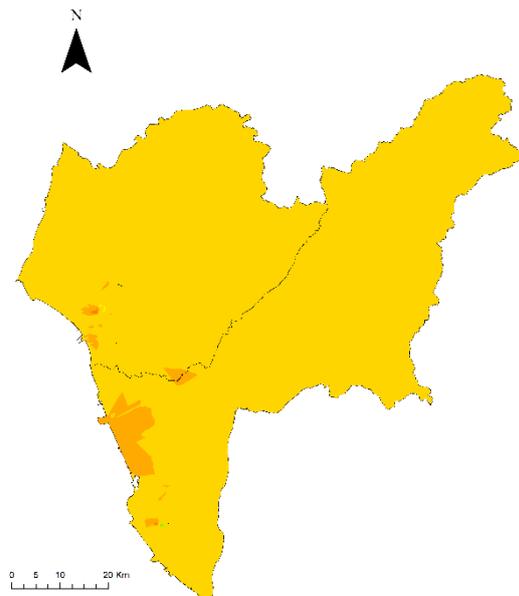
b.



c.

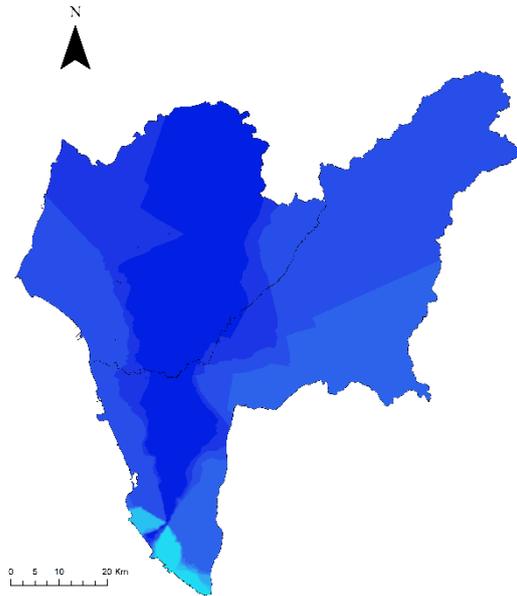


d.

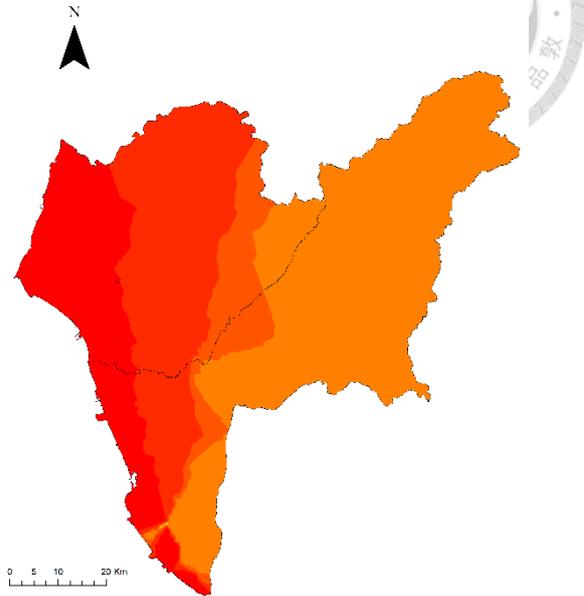


續下頁

e.

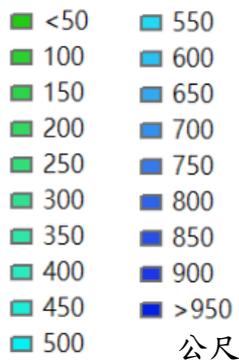


f.



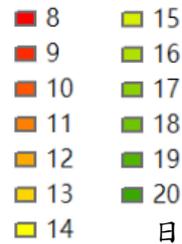
圖例

傳染半徑上界



圖例

發病間距上界



附錄圖 4、疫情中後期時空擴散區的傳染半徑與發病間距上界組圖。a. 模式四 2010 年傳染半徑上界克利金內插空間分布、b. 模式四 2010 年發病間距上界克利金內插空間分布、c. 模式四 2012 年傳染半徑上界克利金內插空間分布、d. 模式四 2012 年發病間距上界克利金內插空間分布、e. 模式三 2014 年傳染半徑上界克利金內插空間分布、f. 模式三 2014 年發病間距上界克利金內插空間分布。a、c 與 e 圖是傳染半徑上界，因此應參考左圖例，單位是公尺，b、d 與 f 圖是發病間距上界，單位是日，請參考右圖例。

附錄二 平台

主程式開發環境

執行檔: DengueDataProcessor.exe, DengueModel.exe, DengueModelVerify.exe,
DengueModelVerifyBuffer.exe, calculateDayAccuracy.exe

開發環境: Code Block ver. 20.03

編譯器: C++ 17 GCC X64 POSIX thread.

C++額外函式庫: boost/geometry.hpp, boost/thread_pool.hpp,

作業系統: Microsoft Windows 10 Enterprise.

中央處理器: Intel Xeon W2195, 18 cores 36 threads, 2.3 GHz.

運算環境: Asynchronous parallel executing with 18 threads.

記憶體空間: 128GB

每日滾動預測時空擴散區介面開發環境

執行檔: DenInterface.exe, DengueCheckInterface.exe, DengueGoogleMap.exe

開發環境: Microsoft visual studio C# 2017, Net Framework 4.6.1

額外函式庫: Microsoft DataVisualization.Charting

作業系統: Microsoft Windows 10 Enterprise.

中央處理器: Intel Xeon W2195, 18 cores 36 threads, 2.3 GHz.

運算環境: Asynchronous parallel executing with 18 threads.

記憶體空間: 128GB





附錄三 演算法

一、疫情樹採用 MST-DBSCAN 演算法

以下說明虛擬碼的運作過程， TrD 是訓練資料集， $minPts$ 中心性門檻、 $RadBound$ 是傳染半徑的閉集合、 $SIBound$ 是發病間距的閉集合、 $EpidemicTree$ 疫情樹是本模式目標的輸出。

起初所有驗證資料即的病例初始化為雜訊點， $checkNearPoints$ 函數的參數是測試資料集，從輸入登革熱模組要求任何傳染者與被傳染者的登革熱病例的屬性資料，從每個登革熱病例作為傳染者 i 病例與往後的發病日的每個病例作為疑似被傳染者 j 病例，觀察每個疑似被傳染者 j 病例是否落在時空半徑閉集合內，若 j 屬於時空半徑閉集合，則把 i 與 j 病例作為鄰近關係，且 i 與 j 病例賦予未分類點，回到主程序依發病日順序進入 dfs 函數， dfs 函數遞迴程序尋找每個未分類病例 i 的與病例 i 臨近性病例 j 且 j 病例不屬於雜訊點，透過 $CountNeighbors$ 函數計算病例 i 的鄰近病例數篩選中心性門檻，決定病例 i 是屬於分類點還是雜訊點，透過 $connect$ 函數作為連結 i 與 j 的傳染關係，逐發病日完成 dfs 函數， i 與 j 的傳染關係被記錄至 $EpidemicTree$ 變數，本模式輸出至疫情樹 $EpidemicTree$ 並結束前處理階段。

Epidemic tree module using modified spatial time DBSCAN algorithm

$VD, minPts=1, RadBound=[0,1000], SIBound=[7,50], EpidemicTree$

```
sub  $checkNearPoints(VD)$ 
```

```
  for each  $i$  in  $VD$  do
```

```
    require  $i$  from Inputs of dengue module
```

```
    for each  $j = i+1$  in  $VD$  do
```

```
      require  $j$  from Inputs of dengue module
```

```
      if(not  $j.imported$  &  $SI(i,j) \in SIBound$  &  $Rad(i,j) \in RadBound$ ) then
```

```
         $i \leftarrow$  NOTCLASSIFY
```

```
         $j \leftarrow$  NOTCLASSIFY
```

```
         $j$  is the neighbor of  $i$ 
```

```
      end if
```

```
    end for
```



```
    end for
end sub

sub dfs(i)
    if CountNeighbors(i) > minPts & i is NOTCLASSIFY then
        for each j in Neighbors of i & j is not NOISE do
            // j is not NOISE equal to j is NOTCLASSIFY or
            CLASSIFY
                connect(i,j)
                dfs(j)
            end for
            i ← CLASSIFY
        else if CountNeighbors(i) ≤ minPts & i is NOTCLASSIFY then
            i ← NOISE
        end if
    end sub

Input
for each i in TrD do
    i ← NOISE
end for
checkNearPoints(TrD)
for each i in TrD do
    EpidemicTree ← dfs(i)
end for
output EpidemicTree
end
```



二、自組織對映模組

以下說明虛擬碼的運作過程，*EpidemicTree* 疫情樹是疫情樹模組提供的推估傳染關係、*TrD* 是訓練資料集、*S* 是傳遞鏈的集合、*S'* 是傳遞鏈的集合的集合、*CoresNumber* 是中央處理器的核心數量、*ActNN* 活化神經元的集合、*Epoch* 是訓練的時期，預設值是 300 次、*Hist_{RAD}*[] 是傳染半徑對應活化神經元的直方圖、*Hist_{SI}*[] 是發病間距半徑對應活化神經元的直方圖、*trunProp* 是隨機消除傳遞鏈集合的比例，本處設定 20%、*NumOfTrD* 訓練資料集的產生數量，預設 10 組、 $\sigma(t)$ 鄰近參數隨 *t* 時期遞減，設定的鄰近參數是從 3 至 1、 $\alpha(t)$ 學習率隨 *t* 時期遞減、設定學習率從 0.01 至 0.002 遞減、*ModSig* 模式選擇器。

首先進入模組根據 *NumOfTrD* 的數量產生訓練資料集，訓練資料集的生成透過 *CaptureTransmissionChains* 函數輸入 *EpidemicTree* 擷取傳遞鏈，*CaptureTransmissionChains* 僅擷取每個病例傳染者 *i* 配對 *j* 被傳染者最大、中位數、最小的傳染半徑、最大、中位數、最小的發病間距的傳遞鏈，因此最大、中位數、最小以反函數表示逆推傳遞鏈，並且隨機遺棄 *trunProp* 的比例的傳遞鏈回傳給 *S'* 附加，由於要生成多個隨機訓練資料集，因此根據 *NumOfTrD* 數量生成訓練資料集給 *S'*，由 *S'* 挑選出任一組 *S* 作為範例，由 *S* 建構訓練資料集，建構訓練資料集的過程向登革熱病例模組要求每一個傳遞鏈中的 *i* 與 *j* 病例，*j* 病例的其他因子，如果是模式一則要求 *j* 病例的氣候環境與位置因子屬性資料、如果模式三則要求 *j* 病例的位置因子屬性資料，並且根據傳遞鏈取得 *i* 與 *j* 病例的傳染半徑與發病間距組件成訓練資料集的輸入層向量，完成輸入層的向量組件之後初始化 *SOM* 自組織對映並呼叫 *ThreadPool* 帶入 *CoresNumber* 準備平行化訓練 *SOM*，再把訓練資料集 *TrD* 進行標準化，*TrD* 帶入 *SOM* 做平行化訓練，總共訓練 *Epoch* 時期每個 *TrD* 的資料都執行基礎自組織對映的訓練過程，以上程序透過 *Threadpool* 做排程，每完成一時期會輸出一個 RMSE，完成 *SOM* 的訓練後，從每一個 *TrD* 尋找的每一筆資料 *tr* 尋找活化神經元 *findtheBestMatchUnit*，*tr* 找到活化神經元 *actnn* 利用 *tr* 的傳染半徑與發病間距項分別附加入 *actnn* 的 *Hist_{RAD}* 與 *actnn* 的 *Hist_{SI}* 分別運算直方圖，再把 *actnn* 加入 *ActNN*，最後存取 *SOM*、*Hist_{RAD}*[]、*Hist_{SI}*[]、*ActNN* 提供給時空擴散區模組載入。



Self-Organizing map module

SOM, TrD, S, S', CoresNumber, ActNN, Epoch=300, Hist_{RAD}[], Hist_{SI}[], trunProp=20%, NumOfTrD=10, t, $\sigma(t) = [3, 1], \alpha(t) = [0.01, 0.002], ModSig$

sub *CaptureTransmissionChains(EpidemicTree)*

for each *i* in *EpidemicTree*

V.append [*max*(*RAD*(*i,j*)), *median*(*RAD*(*i,j*)), *min*(*RAD*(*i,j*)),
max(*SI*(*i,j*)), *median*(*SI*(*i,j*)), *min*(*SI*(*i,j*))]

end for

V ← *V* randomly truncate *trunProp* of transmission of chains

return *V*

end sub

sub *SOM(TrD)*

for *t=1* to *Epoch* do //Epoch

for each *x* in *TrD* do //iterator

Threadpool ← $c = \underset{q}{\operatorname{argmin}} \{ \|x - m_q\| \}$

Threadpool ← $m_q(t + 1) = m_q(t) + h_{cq}(t)[x - m_q(t)] // h_{cq}(t) =$

$\alpha(t) \exp\left(-\frac{\|r_c - r_q\|}{2\sigma^2(t)}\right)$

end for

output RMSE

end for

end sub

sub *findtheBestMatchUnit(SOM, x, {limnn})*

{*m_q*} ← $\forall SOM.neurons$

{*m_q*} ← {*m_q*} ∩ *limnn*

return $\underset{q}{\operatorname{argmin}} \{ \|x - m_q\| \}$

end sub

for *notrd=1* to *NumOfTrD* //*NumOfTrD=10*

S'.append(*CaptureTransmissionChains(EpidemicTree)*)

end for

S ← select each of *S'*

for each transmission chain of *i,j* in *S* do

```

require j from Inputs of dengue module
if (ModSig =4) then TrD.append [RAD(i,j),SI(i,j), j.pos,j.env]
else if (ModSig =3) then TrD.append [RAD(i,j),SI(i,j), j.pos]
end if
end for

```



```

Input
initial SOM
call ThreadPool(CoresNumber)
standardize TrD
parallel train SOM(TrD)
for each tr in TrD do
actnn ← findtheBestMatchUnit(SOM, tr,  $\forall$ )
HistRAD[actnn].append( tr[RAD(i, j)])
HistSI[actnn].append(tr [SI(i, j)])
  ActNN.append(actnn)
end for
save SOM, HistRAD[ ], HistSI[ ], ActNN
end

```



三、時空擴散區模組

以下說明本模組虛擬碼的運作過程， TD 測試資料集、 TrD 訓練資料集、 SOM 自組織對映、 X 病例的集合 X 、 $Hist_{RAD}[]$ 是傳染半徑對應活化神經元的直方圖、 $Hist_{SI}[]$ 是發病間距半徑對應活化神經元的直方圖、 $CoresNumber$ 是中央處理器的核心數量、 $ActNN$ 活化神經元的集合、 $BMUs$ 最佳配飾單元集合、 q 預估涵蓋率下界、 p 預估涵蓋率上界、 ec 預估涵蓋率、 $range$ 預估涵蓋間距、 $findtheBestMatchUnit$ 自組織對映限制活化神經元找到最佳配飾單元的函數。

進入本模式首先根據需求帶入測試資料集 TD 或訓練資料集 TrD 至病例的集合 X ，訓練開關 $TrSW$ 關閉，載入自組織對映模組儲存的 SOM 、 $CDF_{RAD}[]$ 、 $CDF_{SI}[]$ 與 $ActNN$ ，設定 $range$ 透過 q 、 p ，呼叫 $ThreadPool$ 帶入 $CoresNumber$ ，病例的集合 X 每個病例 x 做輪詢實作 $parallelPolling$ 函數， $parallelPolling$ 取得 x 的因子屬性資料且進行標準化，根據 $ActNN$ 的活化神經元集合中的第 u 個神經元重組 V 向量，呼叫 $ThreadPool$ 帶入 $findtheBestMatchUnit$ 找到活化神經元帶入 k ，把 k 附加至 $BMUs$ 直到所有 $ActNN$ 的活化神經元集合做完上述動作，取得 $BMUs$ 後根據 $BMUs$ 累加最佳配飾單元集合最佳配飾單元 k 的 $Hist_{RAD}[k]$ 與 $Hist_{SI}[k]$ 取得 $Hist_{RAD}$ 與 $Hist_{SI}$ 回傳主程序，主程序求取 x 病例的傳染半徑 $x.RAD$ 與 x 病例的發病間距 $x.SI$ ，透過帶入 $range$ 的 x 病例的傳染半徑直方圖 $x.Hist_{RAD}$ 的累積機率密度的反函數與帶入 $range$ 的 x 病例的傳染半徑直方圖 $x.Hist_{SI}$ 的累積機率密度的反函數，得到 x 病例的傳染半徑與發病間距，輸出任意 x 病例 $x.RAD$ $x.SI$ ，直到所有 x 病例屬於 X 在，預測階段輸出 AT 時空擴散區的時空擴散範圍，結束所有模組的運算過程。

Spatio-temporal diffusion zones modules

$TD, VD, SOM, X\{\}, Hist_{RAD}[], Hist_{SI}[], CoresNumber, ActNN, BMUs, q, p, ec, range,$
 $findtheBestMatchUnit$

sub $parallelPolling(x, SOM, ActNN)$

require x from Inputs of dengue module

standardize x

for each u in $ActNN$ do // Polling

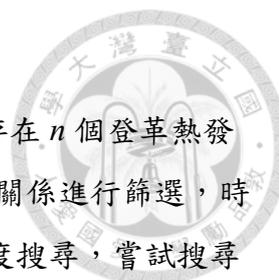
if ($ModSig = 1$) then $V \leftarrow [SOM[u][RAD(i,j)], SOM[u][SI(i,j)], x.pos, x.env]$

else if ($ModSig = 2$) then $V \leftarrow [SOM[u][RAD(i,j)], SOM[u][SI(i,j)], x.pos]$



```
end if
    Threadpool ← k ← findtheBestMatchUnit(SOM, V, ActNN)
    //Best Match unit only in
ActNN
    BMUs.append(k)
end for
     $Hist_{RAD} = \sum_{k \in BMUs} Hist_{RAD}[k]$ 
     $Hist_{SI} = \sum_{k \in BMUs} Hist_{SI}[k]$ 
    return  $Hist_{RAD}, Hist_{SI}$ 
end sub

Input
 $X \leftarrow TD$  or  $TrD$ 
 $TrSW \leftarrow false$ 
load  $SOM, CDF_{RAD}[\ ], CDF_{SI}[\ ], ActNN$ 
 $range = [q \leftarrow 0, p \leftarrow ec]$ 
call ThreadPool(CoresNumber)
for each  $x$  in  $X$  do
     $x.Hist_{RAD}, x.Hist_{SI} \leftarrow parallelPolling(x)$ 
     $x.RAD, x.SI \leftarrow cdf^{-1}(x.Hist_{RAD}, range), cdf^{-1}(x.Hist_{SI}, range)$ 
    Output  $x.RAD, x.SI$ 
end for
//Calculate the volume of Spatio-temporal diffusion zones
Prediction stage:
 $AT = \cup_{x \in X} |x.RAD^2 \Pi | \pi | x.SI |$ 
Output  $AT$ 
end
```



四、演算法複雜度

Modified Spatial Time DBSCAN 演算法的時空複雜度，在存在 n 個登革熱發病點，觀察鄰近點階段 $checkNearPoints(D)$ 所有可能存在的傳染關係進行篩選，時間複雜度: $O(n^2)$ ；空間複雜度: $O(n^2)$ ，針對鄰近點進行遞迴深度搜尋，嘗試搜尋發病節點與發病節點之間的所有連結推導時間複雜度，空間複雜度是深度搜尋遞迴最大深度 d 與發病節點的乘積，時間複雜度: $O(nodes + edges) = O(n + C_2^n) = O\left(\frac{n^2}{2} + \frac{n}{2}\right) = O(n^2)$ ；空間複雜度: $O(nodes * d) = O\left(\frac{n((n-n)+(n-1))}{2}\right) = O\left(\frac{n((n-n)+(n-1))}{2}\right) = O\left(\frac{n^2}{2} - \frac{n}{2}\right) = O(n^2)$ 。

訓練階段的自組織對映時空複雜度，捕捉傳遞鏈函數

CaptureTransmissionChain (*EpidemicTree*)根據 Modified Spatial Time DBSCAN 演算法產生的的連結數最多產生 n^2 個連結傳染關係，自組織對映的拓樸圖為邊長 s 乘 s 個神經元，因此輸入層輸入 n^2 個資料量又讓 $n^2 = N$ ，透過非平行化實作，時間複雜度: $O(Ns^2) = O(n^2s^2)$ ；空間複雜度: $O(N+s^2) = O(n^2+s^2)$ 。平行化實作以執行序 th 作為基礎，時間複雜度: $O\left(N\frac{s^2}{th}\right) = O\left(n^2\frac{s^2}{th}\right) = let_{th \rightarrow s^2} O(n^2) = O(n^2)$ ；空間複雜度: $O(N+s^2 + Nth) = let_{th \rightarrow s^2} O(n^2+s^2 + n^2s^2) = O(n^2s^2)$ 。下個步驟產生最佳配飾單位將輸入層 N 進行分類，非平行化實作時間複雜度: $O(Ns^2) = O(n^2s^2)$ ；空間複雜度: $O(N+s^2) = O(n^2+s^2)$ ，平行化實作時間複雜度: $O\left(N\frac{s^2}{th}\right) = let_{th \rightarrow s^2} O\left(N\frac{s^2}{th}\right) = O(n^2)$ ；空間複雜度: $O(Nth + s^2) = let_{th \rightarrow s^2} O(n^2s^2 + s^2) = O(n^2s^2)$ 。

最後應用在預測與驗證的輪詢階段產生擴散時空半徑 $parallelPolling(x)$ ，輸入層輸入 m 筆資料進行輪詢，將 m 進行模糊化產生 ms^2 組合時空複雜度皆 $O(ms^2)$ ，未進行平行化時間複雜度: $O(ms^2s^2) = O(ms^4)$ ；空間複雜度: $O(ms^2 + s^2)$ 。平行化實作以執行序 th 作為基礎，時間複雜度: $O\left(ms^2\frac{s^2}{th}\right) = let_{th \rightarrow s^2} O(ms^2) = O(ms^2)$ ；空間複雜度: $O(ms^2th + s^2) = let_{th \rightarrow s^2} O(ms^4 + s^2) = O(ms^4)$ 。

統整時空複雜度條件 $n > s$ ，我們假設 $m \cong n$ 訓練資料與驗證資料數量近似的
 研究模式，三個階段採用非平行化的時間複雜度： $O(n^2) + O(n^2s^2) + O(n^2s^2) +$
 $O(ns^4) = O(n^2s^2)$ ；空間複雜度： $O(n^2) + O(n^2+s^2) + O(n^2+s^2) + O(ns^2 +$
 $s^2) = O(n^2+s^2)$ 。採用理想平行化時間複雜度： $O(n^2) + O(n^2) + O(n^2) + O(ns^2) =$
 $O(n^2)$ ；空間複雜度： $O(n^2) + O(n^2s^2) + O(n^2s^2) + O(ns^4) = O(n^2s^2)$ 。本研究採
 用的折衷平行化，因硬體中央處理器的核心數有限因此採用執行序池技術， th 執
 行序數量時間複雜度： $O(n^2) + O\left(n^2 \frac{s^2}{th}\right) + O\left(n^2 \frac{s^2}{th}\right) + O(ms^2th + s^2) =$
 $O\left(n^2 \frac{s^2}{th}\right)$ ；空間複雜度： $O(n^2) + O(n^2+s^2 + n^2th) + O(n^2th + s^2) +$
 $O(ns^2th + s^2) = O(n^2th)$ ，本研究採取折衷模式因此時間複雜度分別介
 於： $O(n^2) \leq O\left(n^2 \frac{s^2}{th}\right) \leq O(n^2s^2)$ ；空間複雜度： $O(n^2+s^2) \leq O(n^2th) \leq O(n^2s^2)$ 。

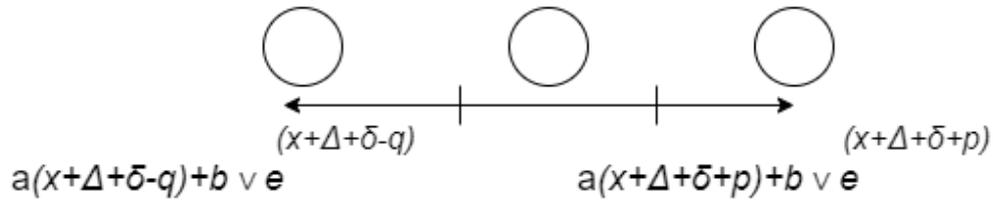
附錄四 證明最佳化病例密度方程式



Proof:

$$\frac{c}{\|ub(x + \Delta + \delta) - lb(x + \Delta + \delta)\|}$$

Method:



Let

$$ub(x + \Delta + \delta) = e^{x+\Delta+\delta+p}$$

$$lb(x + \Delta + \delta) = e^{x+\Delta+\delta-q}$$

$$\|ub(x + \Delta + \delta) - lb(x + \Delta + \delta)\| = \left| \left(e^p - \frac{1}{e^q} \right) e^{x+\Delta+\delta} \right|$$

Let

$$ub(x + \Delta + \delta) = a(x + \Delta + \delta + p) + b$$

$$lb(x + \Delta + \delta) = a(x + \Delta + \delta - q) + b$$

$$\|ub(x + \Delta + \delta) - lb(x + \Delta + \delta)\| = |a(p + q)|$$

$$\text{if } \Delta \rightarrow 0 \Rightarrow \left| \left(e^p - \frac{1}{e^q} \right) e^{x+\delta} \right|$$

$$|a(p + q)| \leq \|ub(x + \Delta + \delta) - lb(x + \Delta + \delta)\| \leq \left| \left(e^p - \frac{1}{e^q} \right) e^{x+\delta} \right|$$

$$\leq \left| \left(e^p - \frac{1}{e^q} \right) e^{x+\Delta+\delta} \right|$$

$$\frac{c}{|a(p + q)|} \geq \frac{c}{\|ub(x + \Delta + \delta) - lb(x + \Delta + \delta)\|} \geq \frac{c}{\left| \left(e^p - \frac{1}{e^q} \right) e^{x+\delta} \right|}$$

$$\geq \frac{c}{\left| \left(e^p - \frac{1}{e^q} \right) e^{x+\Delta+\delta} \right|}$$

附錄五 門檻

一、門檻證明



We define $\frac{\partial AT}{\partial c}(range_{max}) = AT'$

$$\max\left(\frac{\partial AT}{\partial c}(range_{max})\right) = AT'_{max}$$

$$\partial AT(range_{max}) = \int \max^{-1}(AT'_{max}) \partial c$$

$$range_{max} = \int \int \max^{-1}(AT'_{max}) \partial c$$

if $range_k \geq range_{max}$ then $range_k$: outlier, loop $k = k + 1$

if $range_k < range_{max}$

and $|\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2}(range_i) - \frac{\partial AT}{\partial c^2}(range_k)| < |\sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i) -$

$$\frac{\partial AT}{\partial c^2}(range_k)|$$

then $range_k$: outlier, loop $k = k - 1$

else $range_k$: not outlier, threshold = $range_k$

when $range_k = \text{threshold}$

$$|\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2}(range_i) - \frac{\partial AT}{\partial c^2}(range_k)| \geq |\sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i) - \frac{\partial AT}{\partial c^2}(range_k)|$$

$$\frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2}(range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i) \right) \geq \frac{\partial AT}{\partial c^2}(range_k)$$

$$\int \int \frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2}(range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i) \right) \partial c^2 \geq AT(range_k)$$

$$AT^{-1} \left(\int \int \frac{1}{2} \left(\sum_{i=k}^{range_{max}} \frac{\partial AT}{\partial c^2}(range_i) + \sum_{i=0}^{range_k} \frac{\partial AT}{\partial c^2}(range_i) \right) \partial c^2 \right) \geq range_k$$

= threshold



二、門檻值運算成果

附錄表 2、各研究年份訓練資料集門檻值。透過以門檻的定義，遞迴取得訓練資料集模式一至模式四的門檻值。

年分\模式	一	二	三	四
2010	-	0.95	0.95	0.85
2011	-	0.95	0.90	0.85
2012	-	0.85	0.70	0.90
2014	-	0.95	0.85	0.95
2015	-	0.90	0.95	0.95



附錄六 機器學習系統效能評估

一、驗證時空擴散區評估

除考量密度外時空擴散區的病例數量依然是重要的考量之一，作為效能系統評估的模式病例數量涵蓋能力代表被時空擴散區的病例數量，然而本最佳化模式病例數量與密度是個交易，密度高但病例數量過低情況可能導致預測能力不足無法阻斷登革熱的傳遞鏈，同時以系統效能的觀點模式涵蓋的能力越大代表解釋力越佳，另外一方面系統評估不可能模擬出所有情境或覆蓋參數 ec ，所以本模式採用各種等距的 $range$ 模擬各種預測情境描繪出預測涵蓋數量與密度的關係，並且將覆蓋參數通過的被模式預測點與平均密度進行曲線連結計算面積計算出預測面積 $Predictive\ area$ ，此處 i 是系統評估的抽樣點的、 m 代表本模式效能評估的抽樣點、 A_i 代表擴散的平均面積、 T_i 代表估計平均的發病時間間距與 c_i 代表此擴散面積與發病時間間距涵蓋到的病例數，作為第一個整體效能評估的參考數值。

$$Predictive\ area = \sum_{i=1}^m 0.5 \left(\frac{c_{i+1}}{AT_{i+1}} + \frac{c_i}{AT_i} \right) (c_i - c_{i+1})$$

Estimate the optimize of performances.

Predictive area:

$$\begin{aligned} \sum_{i=1}^m \frac{c_{i+1}}{AT_{i+1}} (c_i - c_{i+1}) &\leq \sum_{i=1}^m 0.5 \left(\frac{c_{i+1}}{AT_{i+1}} + \frac{c_i}{AT_i} \right) (c_i - c_{i+1}) \approx \int \frac{c_i}{AT_i} dc \\ &\leq \lim_{n \rightarrow \infty} \sum_i^n \frac{c_i}{AT_i} (c_i - c_{i+1}) \leq \sum_{i=1}^m \frac{c_i}{AT_i} (c_i - c_{i+1}) \leq \sum_{i=1}^2 \frac{c_i}{AT_i} (c_i - c_{i+1}) \\ &= \frac{c_1}{AT_1} (c_1 - c_2) \end{aligned}$$

$n > m > 2$ Let $c_i - c_{i+1} = c_j - c_{j+1}$ $i, j \in N$, because valid data \approx training data

二、驗證模式評估指標

預測面積 $Predictive\ area$ 代表整體的效能評估，但面對某些模式或演算法整體表現在發病病例數的預測與密度預測較佳，但缺乏極端大的密度最佳化值，僅利用預測面積 $Predictive\ area$ 可能低估其預測能力，或者少數模式其時空擴散區的病例數量較大導致密度衰減過度，

類神經網路透過隨機訓練的模式產生預測結果，隨機過程其預測效果參差不齊，因此會採用同時多組相同因子模式進行訓練，選取最佳的模式作為預測結果，其中選擇的方法針對數量給予較大權重放大時空擴散區的病例數量其次放大密度權重，涵蓋病例數量越大放大率越大；密度衰減越大放大率也提升，避免高估僅學習到少數特徵導致涵蓋數量不足但密度卻過高的情形，導致選擇的預測結果無法找到足夠的涵蓋病例數量。

$$\int \frac{dc_2 e^{v_2}}{dc_2} dc_2 > \int \frac{c_1}{AT_1} e^{u_1} dc_1, \text{ if } v_2 \rightarrow g, u_2 \rightarrow f, g > f, v_2 > v_1, u_2 > u_1, c_2 > c_1, c_2 - c_1 \rightarrow 0$$

修正預測面積：

$$\int \frac{c_i}{AT_i} dc, \text{ if } \frac{c_i}{AT_i} \text{ decreasing function}$$

let $e^u, u = 0 \sim f$ $e^v, v = 0 \sim g, f < g, f > 0, g > 0$

$$\text{let Best Model} = \int \frac{c_i}{AT_i} e^u dce^v = \int \frac{c_i}{AT_i} e^u \frac{dc dce^v}{1 dc} = \int \frac{c_i}{AT_i} e^u \frac{dce^v}{dc} dc$$

Find the model covering most cases:

$$\frac{\int \frac{c_i}{AT_i} e^u \frac{dce^0}{dc} dc}{\int \frac{c_i}{AT_i} e^u dc}, \text{ (if } v = 0) = 1 < \frac{\int \frac{c_i}{AT_i} e^u \frac{dce^g}{dc} dc}{\int \frac{c_i}{AT_i} e^u dc}, \text{ (if } v = g) \leq \int \frac{dce^g}{dc} dc$$

Find the model of highest density:

$$\frac{\int \frac{c_i}{AT_i} \frac{dce^v}{dc} dc}{\int \frac{dce^v}{dc} dc}, \text{ (if } u = 0) = \int \frac{c_i}{AT_i} dc < \frac{\int \frac{c_i}{AT_i} e^f \frac{dce^v}{dc} dc}{\int \frac{dce^v}{dc} dc}, \text{ (if } u = f) \leq \int \frac{c_i}{AT_i} e^f dc$$

Then find the model covering most cases priority than highest density models:

$$\text{if } c_2 > c_1, c_2 - c_1 \rightarrow 0, \text{ then } v_2 > v_1, u_2 > u_1, \text{ if } v \rightarrow g, u \rightarrow f, g > f$$

$$\int \frac{dc_2 e^{v_2}}{dc_2} dc_2 > \int \frac{dc_1 e^{v_1}}{dc_1} dc_1$$

$$\int \frac{c_2}{A_2 T_2} e^{u_2} dc_2 > \int \frac{c_1}{A_1 T_1} e^{u_1} dc_1$$



$$\int \frac{c_i}{AT_i} e^f \frac{dce^v}{dc} dc \leq \frac{\int \frac{c_i}{AT_i} e^f \frac{dce^v}{dc} dc}{\int \frac{dce^v}{dc} dc} \frac{\int \frac{c_i}{AT_i} e^u \frac{dce^v}{dc} dc}{\int \frac{c_i}{AT_i} e^u dc} \leq \int \frac{c_i}{AT_i} e^f dc * \int \frac{dce^v}{dc} dc$$

$$\int \frac{dc_2 e^{v_2}}{dc_2} dc_2 \int \frac{c_2}{AT_2} e^{u_2} dc_2 > \int \frac{dc_1 e^{v_1}}{dc_1} dc_1 \int \frac{c_1}{AT_1} e^{u_1} dc_1$$

because $\frac{c_2}{AT_2} < 1$, then $\int \frac{dc_2 e^{v_2}}{dc_2} dc_2 > \int \frac{dc_1 e^{v_1}}{dc_1} dc_1 > \int \frac{c_2}{AT_2} e^{u_2} dc_2$

$$> \int \frac{c_1}{AT_1} e^{u_1} dc_1$$

$$\int \frac{dc_2 e^{v_2}}{dc_2} dc_2 > \int \frac{c_1}{AT_1} e^{u_1} dc_1$$

選擇最佳模式的模式:

$$\sum_{j=1}^n 0.5 \left(\frac{c_{j+1}}{AT_{j+1}} + \frac{c_j}{AT_j} \right) (c_j - c_{j+1})$$

let $r \rightarrow 0$, $\frac{c_{j+1}}{AT_{j+1}} = \frac{c_{k+1}}{AT_{k+1}} e^u$, $\frac{c_j}{AT_j} = \frac{c_k}{AT_k} e^{u-r}$, $c_j = c_k e^v$, $c_{j+1} = c_{k+1} e^{v-r}$

$$\sum_{k=1}^n 0.5 \left(\frac{c_{k+1}}{AT_{k+1}} e^u + \frac{c_k}{AT_k} e^{u-r} \right) (c_k e^v - c_{k+1} e^{v-r})$$

$$\sum_{k=1}^n 0.5 e^u \left(\frac{c_{k+1}}{AT_{k+1}} + \frac{c_k}{AT_k} e^{-r} \right) e^v (c_k - c_{k+1} e^{-r})$$

$$BM = \sum_{k=1}^n 0.5 \left(\frac{c_{k+1}}{AT_{k+1}} + \frac{c_k}{AT_k} \right) (c_k - c_{k+1}) e^u e^v$$

為了每個模式有一致性的標準，因此我們運算 u 密度抗衰減參數、與 v 病例抗衰減參數，我們每個年份相同模式採用第五個子模式進行選擇最佳模式的預測面積過程，計算 BM 最佳模式值，我們挑最佳模式值最大的模式。



三、最佳模式值運算成果

附錄表 3、各研究年份訓練資料集模式三，多組獨立隨機訓練資料運算最佳模式值。運算訓練資料集模式三最佳模式值，取排名第一作為該年份該模式三的代表，以*表示， u 密度抗衰減參數、與 v 病例抗衰減參數，我們以模式四第五個子模式作為共同的抗衰減參數。

2010 年模式三										
模式編號	1	2	3	4	5*	6	7	8	9	10
最佳模式值	8105	5792	8395	8806	10796	8108	9363	7537	5749	8892
排名	7	9	5	4	1	6	2	8	10	3
$u=1.60, v=5.49$										
2011 年模式三										
模式編號	1	2*	3	4	5	6	7	8	9	10
最佳模式值	2960	3264	1121	2376	1662	1337	2451	2214	1528	1507
排名	2	1	10	4	6	9	3	5	7	8
$u=2.07, v=3.79$										
2012 年模式三										
模式編號	1	2	3	4	5	6	7	8	9*	10
最佳模式值	10025	10087	5930	7535	10647	9548	8681	6803	11839	10971
排名	5	4	10	8	3	6	7	9	1	2
$u=2.32, v=4.65$										
2014 年模式三										
模式編號	1	2	3	4	5	6	7	8*	9	10
最佳模式值	39529	31056	33644	30828	27286	39745	19953	40064	27235	37151
排名	3	6	5	7	8	2	10	1	9	4
$u=1.96, v=4.73$										
2015 年模式三										
模式編號	1	2	3	4	5	6	7*	8	9	10
最佳模式值	337600	341494	285563	228856	219146	340405	464070	304355	375284	321191
排名	5	3	8	9	10	4	1	7	2	6
$u=1.60, v=5.49$										



附錄表 4、研究年份訓練資料集模式四，多組獨立隨機訓練資料運算最佳模式值。運算訓練資料集模式四最佳模式值，取排名第一作為該年份該模式四的代表，以*表示， u 密度抗衰減參數、與 v 病例抗衰減參數，我們以模式四第五個子模式作為共同的抗衰減參數。

2010 年模式四										
模式編號	1	2	3	4	5	6	7*	8	9	10
最佳模式值	6651	6627	6865	6803	6702	6681	7028	6498	6955	6662
排名	8	9	3	4	5	6	1	10	2	7
$u=1.6, v=5.49$										
2011 年模式四										
模式編號	1	2	3	4*	5	6	7	8	9	10
最佳模式值	1565	1745	1671	1748	1548	1544	1437	1583	1582	1594
排名	7	2	3	1	8	9	10	5	6	4
$u=2.07, v=3.49$										
2012 年模式四										
模式編號	1	2	3	4	5	6*	7	8	9	10
最佳模式值	7390	8120	8235	8312	8237	8954	7871	7822	8789	8220
排名	10	7	5	3	4	1	8	9	2	6
$u=2.32, v=4.65$										
2014 年模式四										
模式編號	1*	2	3	4	5	6	7	8	9	10
最佳模式值	29917	29114	29202	28990	29548	28558	29336	28825	29011	29149
排名	1	6	4	8	2	10	3	9	7	5
$u=1.96, v=4.73$										
2015 年模式四										
模式編號	1	2	3	4	5	6*	7	8	9	10
最佳模式值	268866	289405	276261	289638	282994	301122	283296	278628	286872	260970
排名	9	3	8	2	6	1	5	7	4	10
$u=1.60, v=5.49$										