

<https://doi.org/10.52288/jbi.26636204.2020.01.10>

登革熱空間分佈研究-基於循證共識資料 Spatial Distribution of Dengue Fever Based on Evidence-based Consensus Data

劉穎^{1*} 楊國樑² 胡玥³
Ying Liu Grant G.L. Yang Yue Hu

摘要

本文以氣候因素及人口密度為預測變數，以系統抽樣的方式選擇未爆發的地區與已知的登革熱爆發地區組成因變數，使用機械學習中的增強回歸樹模型，找出對該疾病流行影響最大的因素。結果顯示影響登革熱爆發的最重要因素是水汽壓和溫度；不同偽安全區域資料的選擇方法確實會對模型擬合產生影響；基於循證共識資料的預測可能是對真實風險圖的更好估計，因為它們可以消除由非報告偏差和近距離選擇造成的潛在氣候偏差影響。

關鍵字：登革熱、增強回歸樹模型、氣候因素、循證共識資料

Abstract

Objective of this paper is to use evidence-based consensus data to construct a spatial distribution model of global dengue fever, to identify the important factors affecting the epidemic of dengue fever, and to examine the impact of different geographical distance selection on the model results. In this study, climatic factors and population density were used as predictors. The unexploded areas were selected by systematic sampling, and the dependent outbreaks were used to form dependent variables. Boosted regression trees in mechanical learning were used to find out the disease's greatest impact on the epidemic. Results show that the most important factors affecting the outbreak of dengue are water vapor pressure and temperature; different PA selection methods do have an impact on model fitting; ECS-based predictions might be a better estimate of real risk maps because they are based on pseudo-safety regional consensus data to eliminate the effects of non-reported bias and avoid close-range selection that could lead to climate bias.

Keywords: Dengue Fever, Boosted Regression Trees Model, Climatic Factors, Evidence-based Consensus Data

1. 前言

登革熱是一種廣泛傳播的媒介傳播病毒感染性疾病，它通過蚊子在人類之間傳播。登革熱的症狀包括高燒、頭痛、關節和肌肉疼痛、嘔吐和皮疹（Srivastava 等，1990），並且會發展成為一種具有危及人類生命的併發症的出血熱。世界衛生組織報

¹ 廈門大學嘉庚學院國際商務學院講師 yingliu722@163.com*通訊作者

² 廈門大學嘉庚學院國際商務學院副教授

³ 廈門大學嘉庚學院會計與金融學院本科生

告報導，儘管沒有針對登革熱的具體治療方法，但通過早期發現和適當的醫療護理，死亡率可降至 1% 以下 (WHO, 2012)。登革熱病媒和病毒向新地區的傳播依賴於人類的運輸 (例如遷徙) 和流動網路 (例如航空交通網) (Huang, 2012; Wilder-Smith 與 Gubler, 2008; Rogers, 2012; Randolph 與 Rogers, 2010)。由於疾病向城市地區的有效傳播和適應，登革熱的全球負擔變得越來越大。據世界衛生組織估計，全世界每年大約有 5,000 萬至 1 億人感染，然而，新的估計值大約高出 4 倍 (Gubler, 2006)。研究登革熱至為重要，一些研究試圖調查和概述其空間分佈和引起疾病的風險區域 (Gubler, 2006; Bhatt 等, 2013; Simmons 等, 2012; Astrom 等, 2012; Brady 等, 2012)。由於登革熱由蚊子傳播，它對氣候很敏感，因此更好地瞭解氣候和氣候變化，對登革熱擴展到未感染地區的影響相對重要 (Simmons 等, 2012; Brady 等, 2012)。

研究物種和疾病分佈可以借助多種分析方法 (Hales 等, 2002; 列汝良等, 2018; 鄭爛等, 2018; 李春敏等, 2018; 任紅豔等, 2019; 易彬檜等, 2003; 嶽玉娟等, 2016)。這些分析方法的實現要依靠已經存在 (presence only-PO) 資料，此類資料是由地理編碼確認的登革熱觀測值；有時這些分析方法的實現也需要依靠真實缺失觀測值 (true absence observation) 或偽缺失資料 (pseudo absence-PA)。PO 資料 (即已經存在資料) 對應的是已報告物種或疾病的位置，而登革熱缺失資料 (或稱為安全區域資料) 是指一些尚未被報告有登革熱爆發，但有可能存在潛在但未確認的登革熱疫情的地理位置。因此，這些登革熱缺失資料觀察結果被歸為偽缺失資料這個資料類別。如果 PA 資料的選擇包含了關於真實缺失資料或者用於形成模型的預測因數的系統偏差，則由這些模型得到的風險圖將是不正確且有偏差的。循證共識法指遵循證據達到共識的方法，運用循證共識法則可以適度避免偏差產生。

本研究目的在於：1. 使用循證共識法選擇資料，構建登革熱的全球空間分佈回歸樹模型，找到對登革熱流行影響最大的影響因素；2. 通過比較估計的回歸模型的預測變數貢獻解釋力，研究不同的選擇未爆發地區觀測值的方法，對登革熱疫情空間分佈的貢獻程度造成的偏差。

2. 資料與方法

2.1 研究資料

本文用以建立增強回歸樹模型的資料包括引數 (或預測變數) 資料和因變數資料 2 個部分。引數資料中有 1 個人口密度變數和 10 個氣候變數，其中每個氣候變數都含有最大、最小和中間值。對於這 11 個主要預測變數的具體統計描述見表 1。

因變數資料，即登革熱資料，是登革熱是否爆發的地點觀測值。每個觀測值都是在以 0.5×0.5 弧度 (arc degree) 為基本單位的全球經緯度地理網格上取得的，每個觀測值都帶有經度和緯度座標數值 (Brady 等, 2012)。因變數資料共有 67,420 個觀測值，其中全球登革熱已爆發地區的觀測值有 1,537 個，其他地區均為未爆發地區 (Climatic Research Unit, 2013)。該資料是由歐洲疾病控制和預防中心 European Centre of Disease Control and Prevention (ECDC) 提供的。因變數 y 服從二項分佈，當 $y=1$ 時表示該地區爆發了登革熱， $y=0$ 時表示該被觀測地區沒有爆發登革熱。

表1. 主要預測變數描述分析表

變數	變數描述	均值	最小值	最大值	單位
cld	雲層覆蓋	57.11	11.45	92	%
dtr	溫度日較差	11.33	2.68	29	°C
frs	霜日頻率	14.64	0	30	days
pet	潛在蒸散量	2.73	0.35	8	millimeters
pre	降水量	54.61	0	617	millimeters
tmp	日平均氣溫	8.57	-27.61	31	°C
tmn	月平均日最低氣溫	2.91	-41.97	26	°C
tmx	月平均日最高氣溫	14.25	-23.16	38	°C
vap	水汽壓	10.71	0.1	32	hPa
wet	潮濕天的頻率	9.02	0	30	days
pop	人口密度	92,158.78	0	17,259,910	people

資料來源：歐洲疾病控制和預防中心 European Centre of Disease Control and Prevention (ECDC)

2.1.1 循證共識數據

Brady等人(2012)根據全球已確認關於登革熱爆發的所有可用資訊，對全球各地區的爆發情況進行評分，分值從0到200分，描繪了登革熱爆發與否的確定性及不確定性。他們創設了9大證據一致的類別：完全未爆發 (complete absence)、高度不流行 (good absence)、中度不流行 (moderate absence)、輕度不流行 (poor absence)、流行性不明確 (indeterminate)、輕度流行 (poor presence)、中度流行 (moderate presence)、高度流行 (good presence) 和完全爆發 (complete presence)。本文在此基礎上把這9個類別進一步分成完全未爆發、完全爆發和不確定3個組別 (見表2)，並使用完全未爆發 (Absence) 這個組別作為PA資料，建立循證共識資料庫 Evidence-based consensus data (ECS data) 來進行統計分析。

表2. Brady 資料庫

名稱	類別	評分	頻數
完全未爆發	Complete, good, moderate (absence)	0-43	41,197
完全爆發	Complete, good, moderate (presence)	157-200	13,427
不確定	Poor absence, indeterminate, poor presence	44-156	6,871
Total			61,495

資料來源：Brady, O. J., Gething, P. W., Bhatt, S. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl. Trop. Dis., 6(8), e1760.

2.1.2 PA 資料選擇策略

用於建立模型的登革熱安全區域資料，是從不同的偽安全區域資料選擇策略中選擇出的，以此來評估和描述不同 PA 選擇策略對登革熱空間分佈的影響。在採用循證共識資料挑選方法的條件下，共有 41,197 條記錄屬於安全區域資料 (Presence data) 被用於 ECS 資料庫，作為偽安全區域資料。根據這個 ECS 資料庫並結合不同地理距離的選擇，創建 5 個樣本資料用來擬合 BRT 模型。樣本資料的創建一共分 3 個步驟：1.分別計算這些 PA 資料到 PO 資料的距離；2.從中挑選出其距離 PO 資料不超過 5、7.5、10 和 12.5 度的 PA 數據；3.運用隨機抽樣法在這些挑選出來的資料裡，選取與

PO 資料相等數量的 PA，並和 PO 資料組成新的樣本資料用以創建模型。讓 PA 和 PO 資料的總數相等是因為這樣可以使模型具有更好的模型精度 (Wisz 與 Guisan, 2009; Barbet-Massin 等, 2012; Rogers 等, 2014; McPherson 等, 2004)。對不同的選擇策略所做出的描述如表 3 所示。

表 3. 樣本資料概覽及 PA 資料選擇方法的描述

樣本資料名稱	PA 的選擇方法
ECS	從 ECS 資料庫中隨機抽取
ECS5	從 ECS 資料庫中選擇與 PO 距離不超過 5° 的資料
ECS7.5	從 ECS 資料庫中選擇與 PO 距離不超過 7.5° 的資料
ECS10	從 ECS 資料庫中選擇與 PO 距離不超過 10° 的資料
ECS12.5	從 ECS 資料庫中選擇與 PO 距離不超過 12.5° 的資料

2.2 研究方法

增強回歸樹 Boosted Regression Trees (BRT) 是一種強大的機器學習方法，已應用於預測全球疾病風險圖。該方法依賴於登革熱爆發與否的資料，來將分類演算法整理為最優判別。用 BRT 預測物種分佈優於其他方法，例如廣義加性模型 Generalized Additive Models (GAM) 和回歸模型 (Elith 等, 2006)。BRT 模型根據變數重要性測量其相對分數，從而厘清每個預測變數的貢獻；數位越大，表明既定的預測變數對回應變數的影響越大。模型的驗證對於未用於構建模型的資料最有效，可防止過度擬合數據。為此，可以優先使用交叉驗證方法，將資料分成兩組，並使用一組擬合模型，另一組驗證它 (Hastie 等, 2009)。根據不同的 PA 選擇策略為每個策略在 ECS 資料裡隨機抽取 100 組 PA 資料，並和 PO 資料合併構建 100 個 BRT 模型，之後對這 100 個模型的預測值求均值，用以繪製全球風險預測圖。

BRT 模型包含兩種方法：Boosting 演算法以及回歸樹。BRT 具有三個重要特徵，每個特徵都可以影響模型擬合：樹的數量 (nt)，學習率 (lr) 和樹的複雜度 (tc)。學習率用來降低每一個樹被增加到模型時所產生的影響。例如，若模型具有 2,000 個樹且 $lr = 0.05$ ，則生成的預測是來自 2,000 個樹中的每一個樹的預測的總和乘以 0.05 (Jun, 2013)。樹的複雜度控制樹中的節點數，因此，它們之間具有一定的交互作用程度。當 $tc = 1$ 時，模型只產生主要影響；當 $tc = 2$ 或 3 時，可以達到雙向或三向交互作用等 (Schonlau, 2005)。優化模型預測性的 nt 由參數 lr 和 tc 確定，因為通過增加適合 lr 和 tc 的信號樹，從而實現偏差最小化，可以估計出最佳的 nt 。

本文實證使用統計軟體為 RStudio 中的 gbm package，設定參數 $lr = 0.005$ 、 $tc = 5$ 來建立增強回歸樹 (Abeare, 2009)。BRT 有各種分佈模式，例如高斯分佈、伯努利分佈、泊松分佈、AdaBoost 演算法、拉普拉斯變換和比例風險回歸模型 (Elith 等, 2008)。本文中使用的伯努利 BRT 模型，因為反應變數登革熱有兩個值：0 和 1。模型的預測性能由受試者工作特徵曲線 (Receiver Operating Characteristic Curve-ROC) 下的面積測量得出，ROC 曲線下的面積 (Area under the Receiver Operating Characteristic Curve) 是一個名為 AUC 的統計量。ROC 曲線是真陽性率與假陽性率的關係圖，在這種情況下，它描述了區分存在資料和缺失資料的能力。AUC 的取值範圍是從 0 到 1，它用於描述模型的預測準確性。如果 AUC 值在 0.9 和 1 之間，則模型的預測性良好；如果它的值在 0.7 和 0.9 之間，則表明預測性中等；如果該值介於 0.5 和 0.7 之間，表示預測性較差；如果該值等於 0.5，表示該模型具有隨機性。

(Schonlau, 2005)。

3. 結果

對於不同地理距離選擇策略的每個樣本資料都分別構建了100個BRT模型，再計算出模型擬合結果的均值，作為該選擇策略的最終模型估計結果。最終的BRT模型對應的可變重要性和AUC，也是用全部100個擬合BRT模型的平均值得到的。從表4中可以發現，不同的地理距離在ECS資料庫中選擇PA資料，確實會影響預測變數的重要性以及模型擬合。將ECS選擇與地理距離策略相結合，對模型ECS5和ECS7.5來說，產生最大貢獻的因素是最低日平均溫度，而對ECS10和ECS12.5模型來說則為蒸汽壓力。隨著所選擇的PA與既定PO的距離從5弧度增加到12.5弧度，最低日平均溫度的重要值從33.3%降低至12.19%；人口密度的重要性值從21.56%下降到13.1%。在ECS資料庫中選擇PA資料時，所有BRT模型的AUC值都等於1，表明從該選擇策略具有最佳模型精度。

表4. 擬合BRT模型

模型	資料集	重要性 (%)					AUC
		1	2	3	4	5	
A	ECS	vap (81.4)	tmx1 (3.71)	Pop (3)	tmn1 (2.72)	tmp1 (2.36)	1
A1	ECS5	Tmpl (33.3)	Vap (21.83)	Pop (21.56)	tmx1 (8.85)	pet3 (1.81)	1
A2	ECS 7.5	Tmpl (24.17)	Pop (19.75)	Vap (5.51)	vap1 (11.77)	tmx1 (11.43)	1
A3	ECS10	vap (37.99)	tmp1 (15.35)	Pop (14.91)	tmx1 (8.64)	pet3 (5.71)	1
A4	ECS 12.5	vap (40.41)	Pop (13.1)	tmp1 (12.19)	vap1 (6.95)	tmx1 (6.89)	1

*括弧內的數值為預測變數貢獻解釋度。

3.1 登革熱的全球風險圖

圖 1 至圖 5 為在不同地理距離選擇策略條件下，使用 ECS 資料構建的 BRT 模型所預測的全球登革熱風險圖。根據圖 1 至圖 5 和基於 ECS 的預測顯示，在從無登革熱疫情傳播的地區循證共識資料選擇的偽安全區域樣本中，與預測登革熱的發生具有更高的適應性的結果相當一致。基於 ECS 和對交叉驗證模型概率的全域預測，表明從現有的觀察中選擇出的距離（隨機、10 度和 12.5 度）之間存在巨大差異。圖 1 至圖 5 表明了在這預測中，人口因素產生的影響較少，並且輕微減少了重要的環境和氣候因素作為疾病擴散誘因對登革熱疫情的影響。

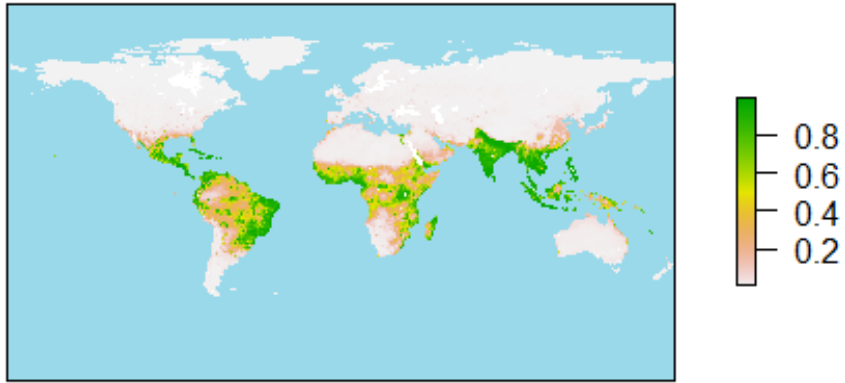


圖 1. ECS 選擇的全球預測 預測的登革熱傳播概率為 0-1。深綠色的位置是登革熱發生風險最高的區域（概率=1）。白色的位置表示該地沒有登革熱的風險。

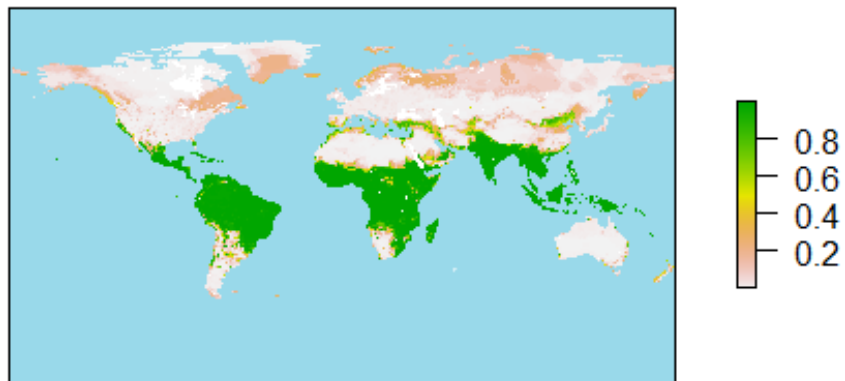


圖 2. ECS5 選擇的全球預測 預測登革熱爆發的概率為 0 至 1。深綠色的位置是登革熱發生風險最高的位置（概率=1）。白色的位置是沒有登革熱風險的區域。

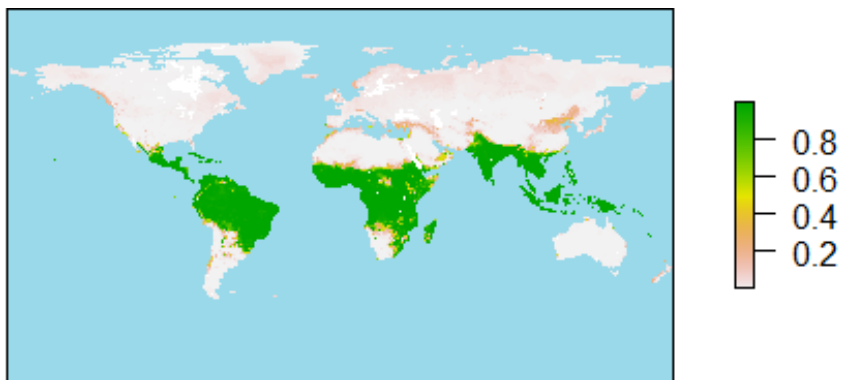


圖 3. ECS7.5 選擇的全球預測 預測登革熱爆發的概率為 0 至 1。深綠色的位置是登革熱發生風險最高的位置（概率=1）。白色的位置為沒有登革熱風險的區域。

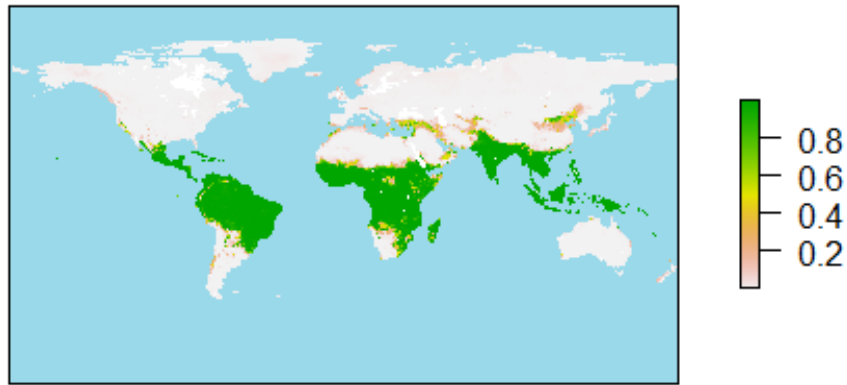


圖 4. ECS10 選擇的全球預測 預測登革熱爆發的概率為 0 至 1。深綠色的位置是登革熱發生風險最高的位置（概率=1）。白色的位置為沒有登革熱風險的區域。

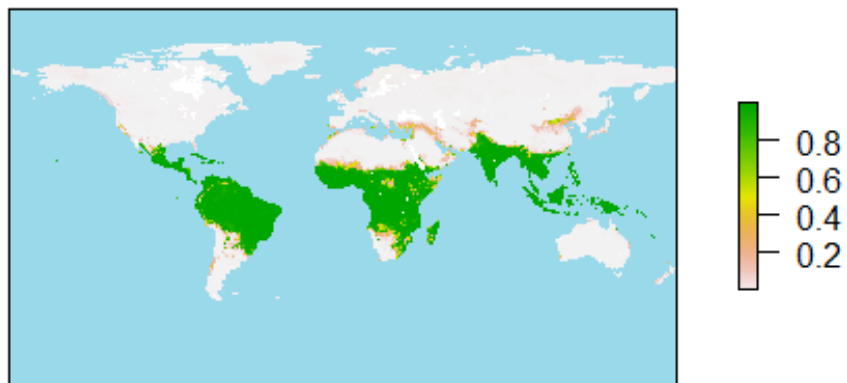


圖 5. ECS12.5 選擇的全球預測 預測登革熱爆發的概率為 0 至 1。深綠色的位置是登革熱發生風險最高的位置（概率=1）。白色的位置為沒有登革熱風險的區域。

4. 討論

總體而言，根據基於循證共識的證據，使用未確認不流行或可能不流行的局部選擇的控制區域（或偽安全區域），在全球預測中，對傳播的高適應性預測區域具有極大影響。部分原因可能是由於當地的選擇不適合進行全球預測，因為強制疾病擴散的全球氣候被忽視，地方特徵被賦予最高權重。此外，使用基於循證共識的證據來選擇偽安全區域，對預測產生極大的影響；在這種方法下，從現有的觀察中產生的選擇距離顯得不太重要。這表明，在比較基於距離和基於共識或隨機選擇的證據時，預測中的誤差/偏差表明距離不是那麼重要，因為偽安全資料是在瞭解部分地區傳播證據的條件下被選擇出的，特別是在根據氣候和社會經濟變化對疾病風險領域的全球氣候變化進行預測時，這一點非常重要。

本研究通過比較不同的 PA 資料選擇方法，對登革熱空間分佈變數貢獻率的影響，來研究其如何影響模型擬合。基於 ECS 選擇，363 個和 957 個偽安全資料分別從 ECS5 和 ECS7.5 的資料集中產生。最重要的預測因素是在 PO 資料中 5 度和 7.5 度的地理距離內選擇 PA 資料時的最低日平均溫度，它表明選擇 PA 資料的數量也會影響模型擬合。隨著距離從 7.5 度降低到 5 度，最低日平均溫度的重要性從 24.17% 增加到 33.3%。當地理距離從 12.5 度降低到 10 度時，最重要的預測因數是蒸汽壓力，

其重要性從 38.1%增加到 40.41%，這表明當系統選擇的 PA 資料更接近 PO 資料時，真實氣候參數具有正偏差。相反地，當系統地選擇的 PA 資料更接近全球氣候資料集的 PO 資料—從 12.5 度到 5 度時，人口的重要性值從 45.78%下降到 44.56%，它表明真實的氣候參數具有極小的負面偏差。因此，使用全球氣候資料集來擬合本地地圖會更好。在這兩個資料集中，重要性值變化不大，因為氣候在 10 度到 5 度之間不會有太大變化。

所有模型的 AUC 值都較大，這表明使用 BRT 模型擬合登革熱的空間分佈是合適的。應用 ECS 資料來選擇以適應全域預測更好，因為它們具有極高的模型精度 (AUC = 1)。

一些研究採用了本文所應用的方法，並根據不同策略選擇 PA 資料 (Simmonds 等, 2012; Brady 等, 2012; Rogers 等, 2012; Bhatt 等, 2013)。本研究強調選擇 PA 方式的重要，它可能會導致基於氣候和人口統計預測的風險圖和預測的巨大差異；如果選擇的 PA 資料地理距離更接近 PO，則疾病風險區域間的氣候聯繫會減弱。如果擬合局部模型，這些方法可能是適當的。然而，這些模型也被用於預測登革熱的全球風險區域，例如 Simmonds 等人的研究 (2012)。這些風險圖可能存在偏差並受到當地因素的過度影響，並且低估了氣候驅動在全球範圍內的作用。本研究結果表明，基於 ECS 的預測可能是對真實風險圖的更好估計，因為它們基於偽安全區域的共識資料可以避免由非報告偏差和近距離選擇導致的潛在氣候偏差影響。

未來對全球風險圖的研究應考慮這些問題，以避免因選擇 PA 的方式不同而產生偏差。此外，在不同氣候變化情況下，對預測登革熱的研究應仔細考慮使用距離作為 PA 的選擇策略，因為距離過近可能導致地圖偏差。

參考文獻

1. Srivastava, V. K., Suri, S., & Bhasin, A. (1990). An epidemic of dengue haemorrhagic fever and dengue shock syndrome in Delhi: A clinical study. *Ann. Trop. Paediatr*, 10, 329-334.
2. WHO (2012). Dengue and severe dengue. WHO Fact Sheet 1-4, at <www.who.int/mediacentre/factsheets/fs117/en/index.html>.
3. Huang, Z., Das, A., & Qiu, Y. (2012). Web-based GIS: The vector-borne disease airline importation risk (VBD-AIR) tool. *Int. J. Health Geogr*, 11, 33.
4. Wilder-Smith, A., & Gubler, D. J. (2008). Geographic expansion of dengue: The impact of international travel. *Medical Clinics of North America*, 92, 1377-1390.
5. Rogers, D. J. (2012). The climatic suitability for dengue transmission in continental Europe. doi:10.2900/62095.
6. Randolph, S. E., & Rogers, D. J. (2010). The arrival, establishment and spread of exotic diseases: Patterns and predictions. *Nat. Rev. Microbiol*, 8, 361-371.
7. Gubler, D. (2006). Dengue fever viruses. *eLS* 1-7, doi:10.1002/9780470015902.a0000412.pub2
8. Bhatt, S., Gething, P. W., & Brady, O. J. (2013). The global distribution and burden of dengue. *Nature*, 496, 504-507.
9. Simmons, C. P., Farrar, J. J., & Nguyen van, V. C. (2012). Dengue. *N. Engl. J. Med.*, 366, 1423-1432.
10. Astrom, C., Rocklov, J., & Hales, S. (2012). Potential distribution of dengue fever under scenarios of climate change and economic development. *Ecohealth*, 9, 448-454.

11. Brady, O. J., Gething, P. W., Bhatt, S. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl. Trop. Dis.*, 6(8), e1760.
12. Hales, S., De Wet, N., & Maindonald, J. (2002). Potential effect of population and climate changes on global distribution of dengue fever: An empirical model. *Lancet*, 360, 830-834.
13. 列汝良、蘇立賢、陳寶林、肖曉玲 (2018)。廣州市越秀區伊蚊密度監測與空間分佈特點研究。深圳中西醫結合雜誌，28(22)，22-23+199。
14. 鄭爛、李喬玄、任紅豔、施潤和、白開旭、魯亮 (2018)。基於土地利用回歸模型的登革熱疫情與社會環境要素的空間關係研究。中國媒介生物學及控制雜誌，29(03)，226-230+234。
15. 李春敏、董學書、楊明東 (2018)。雲南省埃及伊蚊地理分佈與季節消長。中國媒介生物學及控制雜誌，29(04)，394-396+399。
16. 任紅豔、吳偉、李喬玄、魯亮 (2019)。基於反向傳播神經網路模型的登革熱疫情時空分佈特徵和預測[J/OL]。中國媒介生物學及控制雜誌，1-5[2019-01-21]。
17. 易彬檉、張治英、徐德忠、席雲珍、付建國、羅軍、袁明輝、劉少群、鄺鏗 (2003)。廣東省登革熱及媒介種群的空間分佈。第四軍醫大學學報，17，1623-1626。
18. 岳玉娟、吳海霞、李貴昌、劉起勇 (2016)。2005-2013 年中國大陸登革熱病例空間分析。現代預防醫學，43(08)，1345-1348+1354。
19. Climatic Research Unit (2013). http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_1256223773328276。
20. Wisz, M. S., & Guisan, A. (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecol.*, 9, 8。
21. Barbet-Massin, M., Jiguet, F., & Albert, C. H. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods Ecol. Evol.*, 3, 327-338.
22. Rogers, D. J., Suk, J. E., & Semenza, J. C. (2014). Using global maps to predict the risk of dengue in Europe. *Acta Trop.*, 129, 1-14.
23. McPherson, J. M., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *J. Appl. Ecol.*, 41, 811-823.
24. Elith, J. H., Graham, C. P., & Anderson, R. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151.
25. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. books.google.com.
26. Jun, Sung-Hwan. (2013). Boosted regression trees and random forests.
27. Schonlau, M. (2005). Boosted regression (Boosting): An introductory tutorial and a stata plugin. *Stata J.*, 5, 330-354.
28. Abeare, S. (2009). Comparisons of boosted regression tree, GLM and GAM performance in the standardization of yellowfin tuna catch-rate data from the gulf of Mexico longline fishery. at <etd-11042009-152651>.
29. Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.*, 77, 802-813.

收稿時間：2019-11-13
責任編輯、校對：沐園琳、程萌