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

# 登革热空间分布研究-基于循证共识数据
# Spatial Distribution of Dengue Fever Based on Evidence-based Consensus Data

刘颖[1*] 杨国梁[2] 胡玥[3]
Ying Liu Grant G.L. Yang Yue Hu

## 摘要

本文以气候因素及人口密度为预测变量，以系统抽样的方式选择未爆发的地区与已知的登革热爆发地区组成因变量，使用机械学习中的增强回归树模型，找出对该疾病流行影响最大的因素。结果显示影响登革热爆发的最重要因素是水汽压和温度；不同伪安全区域数据的选择方法确实会对模型拟合产生影响；基于循证共识数据的预测可能是对真实风险图的更好估计，因为它们可以消除由非报告偏差和近距离选择造成的潜在气候偏差影响。

**关键词：**登革热、增强回归树模型、气候因素、循证共识数据

## 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

登革热是一种广泛传播的媒介传播病毒感染性疾病，它通过蚊子在人类之间传播。登革热的症状包括高烧、头痛、关节和肌肉疼痛、呕吐和皮疹（Srivastava 等，1990），并且会发展成为一种具有危及人类生命的并发症的出血热。世界卫生组织报

[1] 厦门大学嘉庚学院国际商务学院讲师 yingliu722@163.com*通讯作者
[2] 厦门大学嘉庚学院国际商务学院副教授
[3] 厦门大学嘉庚学院会计与金融学院本科生

告报道，尽管没有针对登革热的具体治疗方法，但通过早期发现和适当的医疗护理，死亡率可降至 1%以下（WHO，2012）。登革热病媒和病毒向新地区的传播依赖于人类的运输（例如迁徙）和流动网络（例如航空交通网）（Huang，2012；Wilder-Smith 与 Gubler，2008；Rogers，2012；Randolph 与 Rogers，2010）。由于疾病向城市地区的有效传播和适应，登革热的全球负担变得越来越大。据世界卫生组织估计，全世界每年大约有 5,000 万至 1 亿人感染，然而，新的估计值大约高出 4 倍（Gubler，2006）。研究登革热至为重要，一些研究试图调查和概述其空间分布和引起疾病的风险区域（Gubler，2006；Bhatt 等，2013；Simmons 等，2012；Astrom 等，2012；Brady 等，2012）。由于登革热由蚊子传播，它对气候很敏感，因此更好地了解气候和气候变化，对登革热扩展到未感染地区的影响相对重要（Simmons 等，2012；Brady 等，2012）。

研究物种和疾病分布可以借助多种分析方法（Hales 等，2002；列汝良等，2018；郑斓等，2018；李春敏等，2018；任红艳等，2019；易彬樘等，2003；岳玉娟等，2016）。这些分析方法的实现要依靠已经存在（presence only-PO）数据，此类数据是由地理编码确认的登革热观测值；有时这些分析方法的实现也需要依靠真实缺失观测值（true absence observation）或伪缺失数据（pseudo absence-PA）。PO 数据（即已经存在数据）对应的是已报告物种或疾病的位置，而登革热缺失数据（或称为安全区域数据）是指一些尚未被报告有登革热爆发，但有可能存在潜在但未确认的登革热疫情的地理位置。因此，这些登革热缺失数据观察结果被归为伪缺失数据这个数据类别。如果 PA 数据的选择包含了关于真实缺失数据或者用于形成模型的预测因子的系统偏差，则由这些模型得到的风险图将是不正确且有偏差的。循证共识法指遵循证据达到共识的方法，运用循证共识法则可以适度避免偏差产生。

本研究目的在于：1.使用循证共识法选择数据，构建登革热的全球空间分布回归树模型，找到对登革热流行影响最大的影响因素；2.通过比较估计的回归模型的预测变量贡献解释力，研究不同的选择未爆发地区观测值的方法，对登革热疫情空间分布的贡献程度造成的偏差。

## 2. 资料与方法

### 2.1 研究资料

本文用以建立增强回归树模型的数据包括自变量（或预测变量）数据和因变量数据 2 个部分。自变量数据中有 1 个人口密度变量和 10 个气候变量，其中每个气候变量都含有最大、最小和中间值。对于这 11 个主要预测变量的具体统计描述见表 1。

因变量数据，即登革热数据，是登革热是否爆发的地点观测值。每个观测值都是在以 0.5×0.5 弧度（arc degree）为基本单位的全球经纬度地理网格上取得的，每个观测值都带有经度和纬度坐标数值（Brady 等，2012）。因变量数据共有 67,420 个观测值，其中全球登革热已爆发地区的观测值有 1,537 个，其他地区均为未爆发地区（Climatic Research Unit，2013）。该数据是由欧洲疾病控制和预防中心 European Centre of Disease Control and Prevention（ECDC）提供的。因变量 y 服从二项分布，当 y=1 时表示该地区爆发了登革热， y=0 时表示该被观测地区没有爆发登革热。

表1. 主要预测变量描述分析表

| 变量 | 变量描述 | 均值 | 最小值 | 最大值 | 单位 |
|---|---|---|---|---|---|
| cld | 云层覆盖 | 57.11 | 11.45 | 92 | % |
| dtr | 温度日较差 | 11.33 | 2.68 | 29 | °C |
| frs | 霜日频率 | 14.64 | 0 | 30 | days |
| pet | 潜在蒸散量 | 2.73 | 0.35 | 8 | millimeters |
| pre | 降水量 | 54.61 | 0 | 617 | millimeters |
| tmp | 日平均气温 | 8.57 | -27.61 | 31 | °C |
| tmn | 月平均日最低气温 | 2.91 | -41.97 | 26 | °C |
| tmx | 月平均日最高气温 | 14.25 | -23.16 | 38 | °C |
| vap | 水汽压 | 10.71 | 0.1 | 32 | hPa |
| wet | 潮湿天的频率 | 9.02 | 0 | 30 | days |
| pop | 人口密度 | 92,158.78 | 0 | 17,259,910 | people |

数据来源：欧洲疾病控制和预防中心 European Centre of Disease Control and Prevention（ECDC）

### 2.1.1 循证共识数据

Brady等人（2012）根据全球已确认关于登革热爆发的所有可用信息，对全球各地区的爆发情况进行评分，分值从0到200分，描绘了登革热爆发与否的确定性及不确定性。他们创设了9大证据一致的类别：完全未爆发（complete absence）、高度不流行（good absence）、中度不流行（moderate absence）、轻度不流行（poor absence）、流行性不明确（indeterminate）、轻度流行（poor presence）、中度流行（moderate presence）、高度流行（good presence）和完全爆发（complete presence）。本文在此基础上把这9个类别进一步分成完全未爆发、完全爆发和不确定3个组别（见表2），并使用完全未爆发（Absence）这个组别作为PA数据，建立循证共识数据库 Evidence-based consensus data（ECS data）来进行统计分析。

表2. Brady数据库

| 名称 | 类别 | 评分 | 频数 |
|---|---|---|---|
| 完全未爆发 | Complete, good, moderate (absence) | 0-43 | 41,197 |
| 完全爆发 | Complete, good, moderate (presence) | 157-200 | 13,427 |
| 不确定 | Poor absence, indeterminate, poor presence | 44-156 | 6,871 |
| Total | | | 61,495 |

数据来源：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 数据选择策略

用于建立模型的登革热安全区域数据，是从不同的伪安全区域数据选择策略中选择出的，以此来评估和描述不同 PA 选择策略对登革热空间分布的影响。在采用循证共识数据挑选方法的条件下，共有 41,197 条记录属于安全区域数据（Presence data）被用于 ECS 数据库，作为伪安全区域数据。根据这个 ECS 数据库并结合不同地理距离的选择，创建 5 个样本数据用来拟合 BRT 模型。样本数据的创建一共分 3 个步骤：1.分别计算这些 PA 数据到 PO 数据的距离；2.从中挑选出其距离 PO 数据不超过 5、7.5、10 和 12.5 度的 PA 数据；3.运用随机抽样法在这些挑选出来的数据里，选取与

PO 数据相等数量的 PA，并和 PO 数据组成新的样本数据用以创建模型。让 PA 和 PO 数据的总数相等是因为这样可以使模型具有更好的模型精度(Wisz 与 Guisan，2009；Barbet-Massin 等，2012；Rogers 等，2014；McPherson 等，2004)。对不同的选择策略所做出的描述如表 3 所示。

表 3. 样本数据概览及 PA 数据选择方法的描述

| 样本数据名称 | PA 的选择方法 |
|---|---|
| ECS | 从 ECS 数据库中随机抽取 |
| ECS5 | 从 ECS 数据库中选择与 PO 距离不超过 5° 的数据 |
| ECS7.5 | 从 ECS 数据库中选择与 PO 距离不超过 7.5° 的数据 |
| ECS10 | 从 ECS 数据库中选择与 PO 距离不超过 10° 的数据 |
| ECS12.5 | 从 ECS 数据库中选择与 PO 距离不超过 12.5° 的数据 |

**2.2 研究方法**

增强回归树 Boosted Regression Trees（BRT）是一种强大的机器学习方法，已应用于预测全球疾病风险图。该方法依赖于登革热爆发与否的数据，来将分类算法整理为最优判别。用 BRT 预测物种分布优于其他方法，例如广义加性模型 Generalized Additive Models（GAM）和回归模型（Elith 等，2006)。BRT 模型根据变量重要性测量其相对分数，从而厘清每个预测变量的贡献；数字越大，表明既定的预测变量对响应变量的影响越大。模型的验证对于未用于构建模型的数据最有效，可防止过度拟合数据。为此，可以优先使用交叉验证方法，将数据分成两组，并使用一组拟合模型，另一组验证它（Hastie 等，2009)。根据不同的 PA 选择策略为每个策略在 ECS 数据里随机抽取 100 组 PA 数据，并和 PO 数据合并构建 100 个 BRT 模型，之后对这 100 个模型的预测值求均值，用以绘制全球风险预测图。

BRT 模型包含两种方法：Boosting 算法以及回归树。BRT 具有三个重要特征，每个特征都可以影响模型拟合：树的数量（*nt*)，学习率（*lr*）和树的复杂度（*tc*)。学习率用来降低每一个树被增加到模型时所产生的影响。例如，若模型具有 2,000 个树且 *lr* = 0.05，则生成的预测是来自 2,000 个树中的每一个树的预测的总和乘以 0.05（Jun，2013)。树的复杂度控制树中的节点数，因此，它们之间具有一定的交互作用程度。当 *tc* = 1 时，模型只产生主要影响；当 *tc* = 2 或 3 时，可以达到双向或三向交互作用等（Schonlau，2005)。优化模型预测性的 *nt* 由参数 *lr* 和 *tc* 确定，因为通过增加适合 *lr* 和 *tc* 的信号树，从而实现偏差最小化，可以估计出最佳的 *nt*。

本文实证使用统计软件为 RStudio 中的 gbm package，设定参数 *lr* =0.005、*tc* =5 来建立增强回归树（Abeare，2009)。BRT 有各种分布模式，例如高斯分布、伯努利分布、泊松分布、AdaBoost 算法、拉普拉斯变换和比例风险回归模型(Elith 等，2008)。本文中使用的是伯努利 BRT 模型，因为反应变量登革热有两个值：0 和 1。模型的预测性能由受试者工作特征曲线（Receiver Operating Characteristic Curve-ROC）下的面积测量得出，ROC 曲线下的面积(Area under the Receiver Operating Characteristic Curve）是一个名为 AUC 的统计量。ROC 曲线是真阳性率与假阳性率的关系图，在这种情况下，它描述了区分存在数据和缺失数据的能力。AUC 的取值范围是从 0 到 1，它用于描述模型的预测准确性。如果 AUC 值在 0.9 和 1 之间，则模型的预测性良好；如果它的值在 0.7 和 0.9 之间，则表明预测性中等；如果该值介于 0.5 和 0.7 之间，表示预测性较差；如果该值等于 0.5，表示该模型具有随机性（Schonlau，2005)。

## 3. 结果

对于不同地理距离选择策略的每个样本数据都分别构建了100个BRT模型，再计算出模型拟合结果的均值，作为该选择策略的最终模型估计结果。最终的BRT模型对应的可变重要性和AUC，也是用全部100个拟合BRT模型的平均值得到的。从表4中可以发现，不同的地理距离在ECS数据库中选择PA数据，确实会影响预测变量的重要性以及模型拟合。将ECS选择与地理距离策略相结合，对模型ECS5和ECS7.5来说，产生最大贡献的因素是最低日平均温度，而对ECS10和ECS12.5模型来说则为蒸汽压力。随着所选择的PA与既定PO的距离从5弧度增加到12.5弧度，最低日平均温度的重要值从33.3%降低至12.19%；人口密度的重要性值从21.56%下降到13.1%。在ECS数据库中选择PA数据时，所有BRT模型的AUC值都等于1，表明从该选择策略具有最佳模型精度。

表4. 拟合BRT模型

| 模型 | 数据集 | 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 | Tmp1 (33.3) | Vap (21.83) | Pop (21.56) | tmx1 (8.85) | pet3 (1.81) | 1 |
| A2 | ECS 7.5 | Tmp1 (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 |

*括号内的数值为预测变量贡献解释度。

### 3.1 登革热的全球风险图

图 1 至图 5 为在不同地理距离选择策略条件下，使用 ECS 数据构建的 BRT 模型所预测的全球登革热风险图。根据图 1 至图 5 和基于 ECS 的预测显示，在从无登革热疫情传播的地区循证共识数据选择的伪安全区域样本中，与预测登革热的发生具有更高的适应性的结果相当一致。基于 ECS 和对交叉验证模型概率的全局预测，表明从现有的观察中选择出的距离（随机、10 度和 12.5 度）之间存在巨大差异。图 1 至图 5 表明了在这些预测中，人口因素产生的影响较少，并且轻微减少了重要的环境和气候因素作为疾病扩散诱因对登革热疫情的影响。

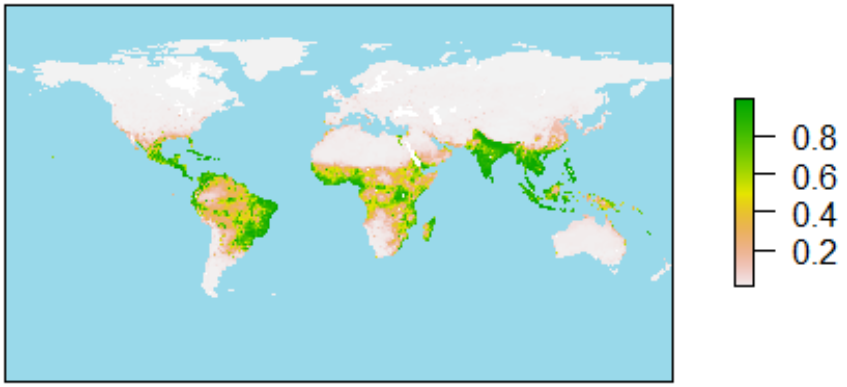


图 1. ECS 选择的全球预测 预测的登革热传播概率为 0-1。深绿色的位置是登革热发生风险最高的区域（概率=1）。白色的位置表示该地没有登革热的风险。

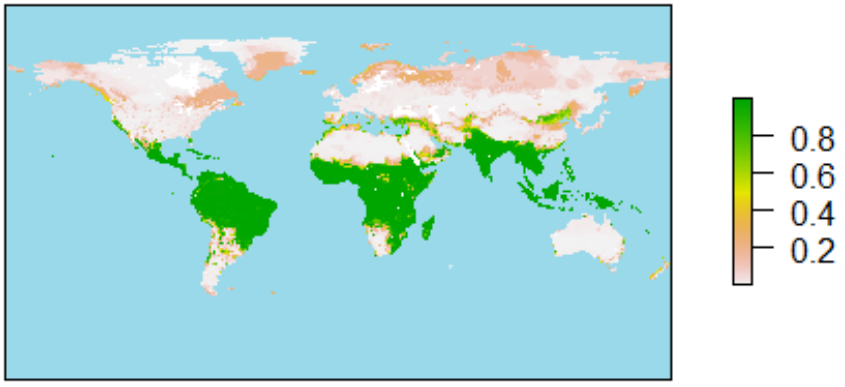


图 2. ECS5 选择的全球预测 预测登革热爆发的概率为 0 至 1。深绿色的位置是登革热发生风险最高的位置（概率=1）。白色的位置是没有登革热风险的区域。

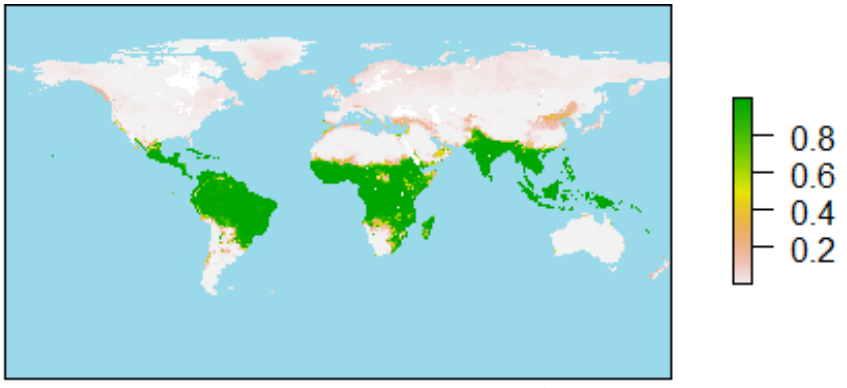


图 3. ECS7.5 选择的全球预测 预测登革热爆发的概率为 0 至 1。深绿色的位置是登革热发生风险最高的位置（概率=1）。白色的位置为没有登革热风险的区域。

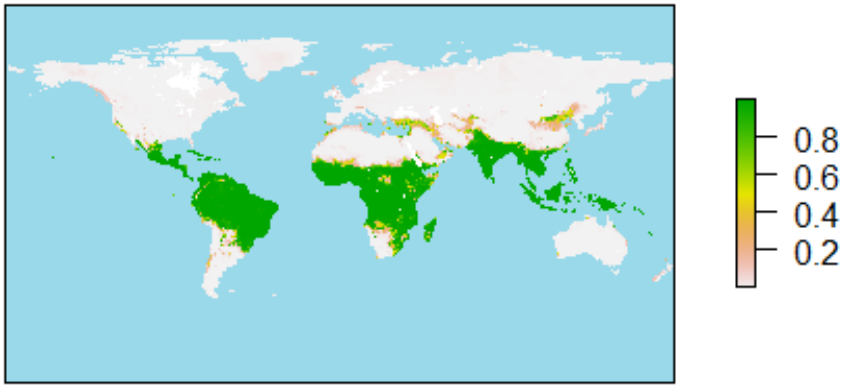


图 4. ECS10 选择的全球预测 预测登革热爆发的概率为 0 至 1。深绿色的位置是登革热发生风险最高的位置（概率=1）。白色的位置为没有登革热风险的区域。

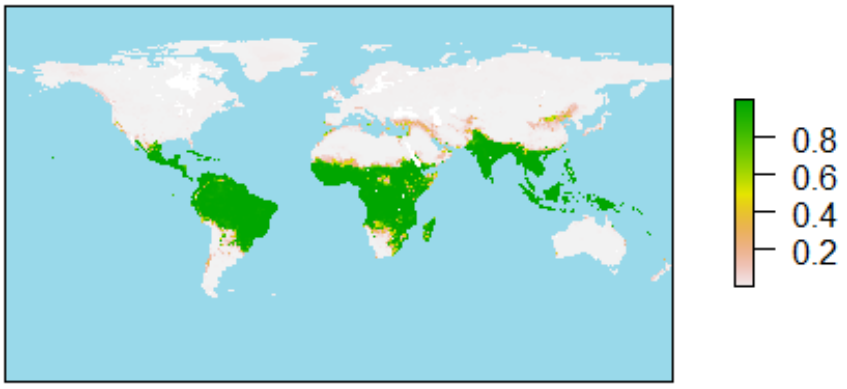


图 5. ECS12.5 选择的全球预测 预测登革热爆发的概率为 0 至 1。深绿色的位置是登革热发生风险最高的位置（概率=1）。白色的位置为没有登革热风险的区域。

## 4. 讨论

总体而言，根据基于循证共识的证据，使用未确认不流行或可能不流行的局部选择的控制区域（或伪安全区域），在全球预测中，对传播的高适应性预测区域具有极大影响。部分原因可能是由于当地的选择不适合进行全球预测，因为强制疾病扩散的全球气候被忽视，地方特征被赋予最高权重。此外，使用基于循证共识的证据来选择伪安全区域，对预测产生极大的影响；在这种方法下，从现有的观察中产生的选择距离显得不太重要。这表明，在比较基于距离和基于共识或随机选择的证据时，预测中的误差/偏差表明距离不是那么重要，因为伪安全数据是在了解部分地区传播证据的条件下被选择出的，特别是在根据气候和社会经济变化对疾病风险领域的全球气候变化进行预测时，这一点非常重要。

本研究通过比较不同的 PA 数据选择方法，对登革热空间分布变量贡献率的影响，来研究其如何影响模型拟合。基于 ECS 选择，363 个和 957 个伪安全数据分别从 ECS5 和 ECS7.5 的数据集中产生。最重要的预测因素是在 PO 数据中 5 度和 7.5 度的地理距离内选择 PA 数据时的最低日平均温度，它表明选择 PA 数据的数量也会影响模型拟合。随着距离从 7.5 度降低到 5 度，最低日平均温度的重要性从 24.17% 增加到 33.3%。当地理距离从 12.5 度降低到 10 度时，最重要的预测因子是蒸汽压力，

其重要性从 38.1%增加到 40.41%，这表明当系统选择的 PA 数据更接近 PO 数据时，真实气候参数具有正偏差。相反地，当系统地选择的 PA 数据更接近全球气候数据集的 PO 数据–从 12.5 度到 5 度时，人口的重要性值从 45.78%下降到 44.56%，它表明真实的气候参数具有极小的负面偏差。因此，使用全球气候数据集来拟合本地地图会更好。在这两个数据集中，重要性值变化不大，因为气候在 10 度到 5 度之间不会有太大变化。

所有模型的 AUC 值都较大，这表明使用 BRT 模型拟合登革热的空间分布是合适的。应用 ECS 数据来选择以适应全局预测更好，因为它们具有极高的模型精度（AUC = 1）。

一些研究采用了本文所应用的方法，并根据不同策略选择 PA 数据（Simmonds 等，2012；Brady 等，2012；Rogers 等，2012；Bhatt 等，2013)。本研究强调选择 PA 方式的重要，它可能会导致基于气候和人口统计预测的风险图和预测的巨大差异；如果选择的 PA 数据地理距离更接近 PO，则疾病风险区域间的气候联系会减弱。如果拟合局部模型，这些方法可能是适当的。然而，这些模型也被用于预测登革热的全球风险区域，例如 Simmonds 等人的研究（2012）。这些风险图可能存在偏差并受到当地因素的过度影响，并且低估了气候驱动在全球范围内的作用。本研究结果表明，基于 ECS 的预测可能是对真实风险图的更好估计，因为它们基于伪安全区域的共识数据可以避免由非报告偏差和近距离选择导致的潜在气候偏差影响。

未来对全球风险图的研究应考虑这些问题，以避免因选择 PA 的方式不同而产生偏差。此外，在不同气候变化情况下，对预测登革热的研究应仔细考虑使用距离作为 PA 的选择策略，因为距离过近可能导致地图偏差。

## 参考文献


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., Gathing, 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、苏立贤、陈宝林、肖晓玲（2018）。广州市越秀区伊蚊密度监测与空间分布特点研究。深圳中西医结合杂志，28(22)，22-23+199。
14. 郑斓、李乔玄、任红艳、施润和、白开旭、鲁亮（2018）。基于土地利用回归模型的登革热疫情与社会环境要素的空间关系研究。中国媒介生物学及控制杂志，29(03)，226-230+234。
15. 李春敏、董学书、杨明东（2018）。云南省埃及伊蚊地理分布与季节消长。中国媒介生物学及控制杂志，29(04)，394-396+399。
16. 任红艳、吴伟、李乔玄、鲁亮（2019）。基于反向传播神经网络模型的登革热疫情时空分布特征和预测[J/OL]。中国媒介生物学及控制杂志，1-5[2019-01-21]。
17. 易彬樘、张治英、徐德忠、席云珍、付建国、罗军、袁明辉、刘少群、邝铿（2003）。广东省登革热及媒介种群的空间分布。第四军医大学学报，17，1623-1626。
18. 岳玉娟、吴海霞、李贵昌、刘起勇（2016）。2005-2013 年中国大陆登革热病例空间分析。现代预防医学，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.

收稿时间：2019-11-13
责任编辑、校对：沐园琳、程萌