arXiv:2302.06747v1 [stat.AP] 13 Feb 2023

# 利用 INLA 进行哥斯达黎加登革热风险预测的贝叶斯时空模型

Shu-Wei Chou-Chen *、Luis A. Barboza、Paola V´asquez、Yury E. Garc´ıa、Juan G. Calvo、Hugo G. Hidalgo、Fabio Sanchez

2023年2月15日

**摘要**

由于伊蚊 的地理分布迅速扩展，且登革热发病率不断上升，登革热已日益成为全球热带和亚热带国家公共卫生部门关注的问题。气候变化、卫生系统承受的负担以及杀虫剂抗药性加剧等重大挑战，凸显了开发新的、经济有效的工具以制定公共卫生干预措施的必要性。用于开发基于气候的预警系统的各种本地化统计方法，日益成为全球研究与关注的领域。哥斯达黎加拥有多种微气候，自1993年以来登革病毒（DENV）持续地方性传播，为开发预测模型提供了理想条件；此类模型有望帮助指导公共卫生工作和干预措施，以控制并监测未来的登革热暴发。

关键词：公共卫生、贝叶斯推断、时空模型、气候、媒介传播疾病

## 1 引言

登革热是全球最常见的媒介传播疾病之一，可影响各年龄段人群。感染可能无症状，也可能引起范围广泛的临床表现，从非特异性且具有自限性的病毒综合征，到伴有出血表现和多系统损害、可能导致患者死亡的疾病 [1]。感染由四种登革病毒血清型之一（登革病毒1–4型）引起，通过受感染雌蚊叮咬传播给人类，主要传播媒介为埃及伊蚊，而白纹伊蚊则为次要传播媒介。

包括全球化、贸易、旅行、人口趋势和气温升高在内的多种因素相互作用，与蚊虫的传播有关。这种蚊虫如今已遍布除南极洲以外的所有大陆 [2, 3], , 因此成为全球最严重的100种入侵物种之一 [4]。这也导致该疾病在此前并不存在的地区出现 [5, 6, 7, 8, 9], , 使全球一半以上的人口面临感染风险，主要集中在热带和亚热带地区 [1]。在这种背景下，如果没有有效的预防和控制措施，预计登革热将继续扩大其地理分布范围 [10, 11]。由于缺乏疫苗和抗病毒药物，预防和

*通讯作者：shuwei.chou@ucr.ac.cr

控制暴发疫情的关键仍在于通过化学和生物干预减少孳生地，并积极动员社区，而这主要取决于受影响国家可用的资源。

鉴于病毒的生态特性与传播登革热的蚊子的生态特性密不可分 [12], ，气候和环境条件会改变媒介生态的时空动态，尤其是温度、降雨量和相对湿度 [13, 14]。温度会影响病毒扩增 [15], ，并提高媒介存活率、繁殖能力和叮咬率 [16, 17]。卵、幼虫和蛹的长期孳生栖息地可能受到降雨后留下的废水或人们将水储存在容器中的行为影响 [18, 19]。登革热发病率也与植被指数、树木覆盖、住房质量以及周边土地覆盖有关 [20, 21]。

了解这些气候变量对不同地区疾病发病率的影响，有助于及早检测疾病进展、指导资源分配，并实施适当的卫生干预 [1]。为此，人们开发了多种监测系统方法 [22, 23, 24]。然而，由于疾病本身复杂且动态多变，成功的早期预警策略仍然有限。生物、社会经济、环境和气候因素之间的复杂相互作用，导致登革热发病强度呈现显著的时空异质性。这给监测与控制系统的建立带来了挑战。

哥斯达黎加是一个位于中美洲的国家，国土面积为 51,179 $km^2$，拥有多种微气候。自1993年以来，登革热一直在该国流行，并从那时起成为一项公共卫生负担。据卫生部统计，过去28年间报告的病例超过398,000例 [25]。登革病毒1型、登革病毒2型和登革病毒3型一直是主要流行血清型。然而，值得特别指出的是，2022年在该国多个市镇发现了血清型4的传播。由于这种血清型历来未在国家领土内出现，其出现以及该国缺乏相应免疫力可能导致病例发病率上升，这进一步凸显了加强预防、检测和及时治疗工作的必要性，从而避免病例发病率上升以及疾病发展为更严重的形式。

在本研究中，我们提出一种用于预测哥斯达黎加登革热的时空模型，纳入气候变量和地理信息，以捕捉影响该国登革热发病率时空变化的因素；这些因素的数据无法获取，例如人口流动、社会经济信息和人口统计信息。本研究是 [26, 27], 一系列工作的组成部分，这些工作旨在开发早期预警系统，以监测该国的登革热风险。我们的目标是为哥斯达黎加卫生部门提供指导工具，以便实施和验证这些工具，并优化和分配用于登革热预防与防控的资源。

哥斯达黎加卫生当局根据登革热的历史发病负担和适宜的疾病传播环境，确定了 32 个重点市镇（全国共划分为 83 个市镇）作为研究地点。入选的市镇主要位于太平洋和加勒比海沿岸地区，以及哥斯达黎加城市化程度和人口密度最高的大都会区。此次遴选也考虑了决策者需要纳入新的、经济高效的工具，以指导全年资源分配。本文结构如下：第 2 节介绍数据、统计模型、假设及所采用的方法。第 3 节展示 32 个重点市镇的研究结果。最后，第 4 节讨论这一建模方法的结果、局限性、优势以及

未来工作。

# 2 方法

## 2.1 数据描述

### 2.1.1 登革热病例

本分析采用了哥斯达黎加公共卫生部门关注的 32 个市镇的月度登革热病例数据。数据涵盖 2000—2021 年，来源为哥斯达黎加卫生部 [28]。

### 2.1.2 气候变量

1. 每日降水量估计值（$P_{i,\ t}$）用于表征地表降雨。数据取自气候灾害红外降水站点数据集（CHIRPs）；见 [29]。由于该数据集具有较高的空间分辨率（5 千米 × 5 千米），我们能够通过汇总较小行政区划（区）内的精确估计值，计算各市镇的月累计降雨量估计值。
2. 厄尔尼诺—南方涛动（ENSO，$S_{i,\ t}$）变化采用海表温度异常（SSTA）指数进行表征，所涉区域称为 Niño 3.4 区（5°N–5°S，120°W–170°W）。月度数据取自美国国家海洋和大气管理局（NOAA）的气候预测中心（CPC）（见 [30]）。3. 归一化植被指数（NDVI，$N_{i,\ t}$）用于表征植被绿度，时间分辨率为 16 天，空间分辨率为 250 米。该指数由中分辨率成像光谱仪（MODIS）卫星获取，并可通过 MODISTools R 程序包获得（见 [31]）。
4. 白天地表温度（LST，$L_{i,\ t}$）以开尔文为单位，时间分辨率为 8 天，空间分辨率为 1 千米；其数据来源与 NDVI 协变量相同。5. 热带北大西洋指数（TNA，$TN_{i,\ t}$），即东部热带北大西洋海表温度异常（SSTA）指数（见 [32]）。该地区以往的研究 [33] 表明，与仅使用太平洋厄尔尼诺—南方涛动条件所做的预测相比，纳入加勒比海/大西洋的 SSTA 信息能够提高对中美洲地表降雨和温度的预测效果。

## 2.2 模型

我们采用分布滞后非线性模型（DLNM）框架[34, 35]，纳入气候协变量的历史暴露以及过去相对风险的变化。该方法采用二维函数空间来定义暴露-滞后-反应函数 $f \cdot w(x,\ l)$，该函数以综合方式刻画预测变量 $x$ 在各滞后期 $l$ 上的关系。这种组合刻画了非线性且存在延迟的关联

气候协变量与登革热发病率之间的关系。对于每个协变量，我们考虑至少 3 个月的滞后期和最长 12 个月的暴露期，以获得最长提前三个月的预测估计。这种滞后期组合是根据各序列之间的交叉相关和小波行为确定的（见 [27]）。我们还使用不同的最大和最小滞后期组合对模型进行了测试。

为比较非线性设定相较于线性设定带来的预测增益，我们比较了两类模型：一类使用自然三次B样条函数，为每个协变量及其滞后效应设置2个节点，以刻画暴露的滞后效应与结局之间的非线性关系；另一类则采用暴露的滞后效应与结局之间的线性关系。

在设定协变量的历史暴露后，我们采用时空贝叶斯层级模型，以各市镇每月登革热病例数作为响应变量 $i$，$i = 1.....32$，$t = 1.....264$，如下所示：

$$Y_{it}|\mu_{it}, \kappa \sim NegBin(\mu_{it}, \kappa), \tag{1}$$
$$\log(\mu_{it}) = \log(E_{it}) + \log(RR_{it}), \tag{2}$$

and

$$\log RR_{it} = \alpha + f_1(RR_t) + f_2(P_t) + f_3(S_t) + f_4(N_t) + f_5(L_t) + f_6(TN_t) + f_7(M_t) + \phi_{i,(month)} + \theta_{i,(year)},$$

其中，$f_k$，$k = 1.....7$ 是暴露-滞后-反应函数，对滞后期3至12的每个气候协变量施加线性效应；$\phi_{i,\ (month)}$ 是市镇特异的月度随机效应，其先验服从一阶循环随机游走，即 $\phi_{i,\ (month)} - \phi_{i,\ (month-1)} \sim N\ 0$ （，$\sigma_\phi^2$）；$\theta_{i,\ (year)}$ 是随机空间效应。

对于空间效应，定义了两种邻近矩阵 $\boldsymbol{W}$：

1. 通常的邻接矩阵定义为 $\boldsymbol{W} = \{\boldsymbol{W}\}_{ij} = 1$，如果市镇 $i$ 和 $j$ 是邻居，则取该值；否则取0。

2. 另一种距离矩阵基于各市镇市中心之间的主要道路距离（单位：千米）定义，即 $\boldsymbol{W} = \{\boldsymbol{W}\}_{\boldsymbol{ij}} = 1$，如果距离小于总体中位数，则取该值；否则取 0。我们引入这一距离，以更真实地衡量社会动态之间的邻近程度。

本文采用四种空间结构。首先假设空间效应相互独立。随后，我们采用带非正规先验的内在条件自回归（CAR）设定、带正规先验的CAR模型，以及Besag-York-Mollie（BYM）模型 [36]。

具体　　地，特定年份空间效应的CAR设定定义如下：

$$\theta_{i,(year)}|\theta_{j,(year),\tau_\theta} \sim N\left(\frac{1}{n_i}\sum_{j\sim i}\theta_{j,(year)}, \frac{1}{\tau_\theta n_i}\right),$$

其中，$\tau_\theta$ 是条件精度，$j \sim i$ 表示 $\boldsymbol{W}_{ij} = 1$，$n_i$ 是根据两种邻近矩阵的定义得出的邻居数量。通过向 $n_i$添加一个正值 $d$，可得到适当条件自回归模型；而BYM模型则是为每个市镇添加一个非结构化随机效应。详情请参见 [37]。所有计算均使用R软件包dlnm [38] 和INLA [39] 完成。

最后，我们采用了两种简单的预测程序，并从预测能力方面将其与所提出的模型进行比较。这两种方法均使用最近五年（2015—2019年）的训练集。首先，使用最近五年的简单月均值 $RR_{it}$进行朴素预测。该方法易于计算，但只有在作出独立性等强假设的前提下，才能计算预测区间；因此，只能计算 $NMRSE$。

其次，作为一种替代方案，我们估计了模型（1），并采用

$$\log RR_{it} = \alpha,$$

负二项零模型，以便在这种情况下计算预测不确定性。执行这两种预测程序的主要目的是比较所提出的模型与这两种简单算法的表现，后者采用由协变量和空间随机效应构成的更复杂结构。

### 2.3 模型选择与预测

以往研究 [27, 40] 表明，这些气候协变量对于预测登革热发病率至关重要。为开始校准，选取一个训练期来拟合模型1，并采用不同的协变量组合和模型时空配置。首先，DLNM 框架允许我们选择协变量历史信息的最大和最小滞后的不同组合。此外，所有协变量的基函数均选择为非线性或线性。最后，拟合了四种空间结构：独立、条件自回归、适当条件自回归和BYM模型。

通过采用不同标准比较所有拟合模型，选出最佳模型。针对每个模型计算偏差信息准则（DIC）和平均交叉验证对数评分。DIC衡量模型的拟合精度和复杂度；而交叉验证对数评分则衡量模型的预测能力，每次留出一个数据点。

最后，计算两个指标来比较各模型的预测性能。归一化均方误差（$NRMSE$）：

$$NRMSE = \sqrt{\frac{1}{m\overline{RR}}\sum_{t=1}^{m}(RR_t - \widehat{RR}_t)^2},$$

其中， $m$ 是测试期的月数， $RR$ 是同一时期内的平均相对风险，而 $RR$ 是根据两个模型中任一模型估计的相对风险。置信水平为 $\alpha$ （$NIS_\alpha$）时的归一化区间评分，是区间评分的归一化版本（参见 [41] 和 [42]）：

$$NIS_\alpha = \frac{1}{m\overline{RR}}\sum_{t=1}^{m}\left[(U_t - L_t) + \frac{2}{1-\alpha}(L_t - RR_t)\cdot 1_{RR_t<L_t} + \frac{2}{1-\alpha}(RR_t - U_t)\cdot 1_{RR_t>U_t}\right],$$

其中， $U_t$ 和 $L_t$ 分别是预测区间的上限和下限。当通过预测区间 [42]概括不确定性时，后者指标在评估模型的预测能力方面更为全面。该指标已用于此前针对哥斯达黎加登革热的预测研究（见 [26, 40]）。我们使用标准化后的均方根误差和区间评分，因为无论各地点相对风险的尺度如何，都可以进行比较。

## 3 结果

我们使用所有气候协变量，为登革热数据拟合了具有不同结构的模型（1）。训练期包括2000年1月至2020年12月的月度数据，测试期为2021年1月至2021年3月。

关于 DLNM 框架，指定用于纳入气候协变量历史暴露的最大和最小滞后期颇具挑战性。我们将最大和最小滞后期分别设为 12 和 3 个滞后期，以便预测未来最多三个月的登革热数据。我们还使用最大和最小滞后期分别为 12 和 0 的设定拟合了模型，与第一种方案相比，预测精度差异并不显著。表 1 列出了所有模型的 DIC 和 CV对数评分，各项指标中最小的值以粗体显示。首先，显然就拟合优度而言，具有

表 1：根据偏差-信息准则（DIC）和平均交叉验证（CV）对数评分对模型进行比较。

| DLNM | 邻近矩阵 | 空间结构 | DIC | CV对数评分 |
|---|---|---|---|---|
| 线性 * | 邻居 | 独立 | 57135.37 | 3.8710 |
| | | CAR | 54256.47 | 3.6872 |
| | | 适当条件自回归 | **52628.40** | **3.5774** |
| | | BYM | 52632.24 | 3.5784 |
| | 距离 | CAR | 53416.92 | 3.6264 |
| | | 适当条件自回归 | 52633.29 | 3.5787 |
| | | BYM | 52636.92 | 3.5787 |
| 非线性 * | 邻居 | 独立 | 53429.63 | 3.8756 |
| | | CAR | 50640.66 | 3.6838 |
| | | 适当条件自回归 | **49438.81** | **3.5954** |
| | | BYM | 49461.01 | 3.5977 |
| | 距离 | CAR | 54674.34 | 3.9653 |
| | | 适当条件自回归 | 49468.89 | 3.5985 |
| | | BYM | 49456.27 | 3.5971 |

*每种 DLNM 规格的最佳模型以粗体标出。

空间结构优于假设空间结构相互独立的模型。接着可以看到，比较线性与非线性（B样条）DLNM 框架时，DIC 和交叉验证对数得分的差异很小。

最后，为了在模型复杂度与所有地点的预测精度之间取得可接受的平衡，并综合考虑 DIC 和 CV对数评分，我们选择了采用 linearDLNM 和邻近矩阵的适当条件自回归模型作为拟合效果最佳的模型。

表 2汇总了训练期和测试期各市镇的预测指标（$NRMSE$和 $NIS_{0.05}$），并将最佳模型与具有独立空间结构的基线模型进行了比较。

比较这两种建模方案，可以看出空间信息有助于提高预测精度。最引人注目的市镇 Parrita 表现出独特的特征。具有独立空间结构的基线模型

Ta表 2：所选模型训练数据集和测试数据集的预测指标 l

| 市镇 | 训练集 | | | | 测试集 | | | |
|---|---|---|---|---|---|---|---|---|
| | 独立 | | 适当条件自回归 | | 独立 | | 适当条件自回归 | |
| | $NRMSE$ | $NIS_{95}$ | $NRMSE$ | $NIS_{0.05}$ | $NRMSE$ | $NIS_{0.05}$ | $NRMSE$ | $NIS_{0.05}$ |
| 阿拉胡埃拉 | 0.7815 | 14.5297 | 0.3982 | 5.3710 | 0.0750 | 1.0905 | 0.0416 | 1.6417 |
| 阿拉胡埃利塔 | 0.3090 | 23.2238 | 0.2175 | 13.1177 | 0.4515 | 22.8135 | 0.0515 | 2.0922 |
| 阿特纳斯 | 4.6100 | 24.7736 | 2.5706 | 10.1453 | 5.1453 | 87.6173 | 0.3283 | 10.2199 |
| 卡˙尼亚斯 | 10.1008 | 24.3243 | 5.7069 | 12.3277 | 5.8803 | 60.7800 | 0.4829 | 6.8129 |
| 卡里略 | 4.3786 | 15.9598 | 3.5404 | 9.1421 | 0.9860 | 13.2041 | 0.4175 | 2.5945 |
| 科雷多雷斯 | 5.0620 | 25.1830 | 2.5155 | 8.6126 | 1.6824 | 17.0115 | 0.5697 | 1.9358 |
| 德萨姆帕拉多斯 | 0.2351 | 18.6300 | 0.1466 | 8.2922 | 0.0282 | 3.0978 | 0.0142 | 3.0085 |
| 埃斯帕尔萨 | 4.7287 | 17.4775 | 2.6296 | 8.4081 | 1.1899 | 16.1291 | 0.1849 | 2.1748 |
| Garabito | 38.4940 | 29.0191 | 5.2545 | 9.8071 | 28.8475 | 232.7807 | 0.4258 | 12.9528 |
| Golfito | 3.7245 | 23.5563 | 1.7734 | 8.2174 | 2.0078 | 35.7736 | 0.3499 | 8.5601 |
| Guacimo | 1.9057 | 13.8120 | 1.0844 | 4.4132 | 3.0709 | 18.1805 | 2.0633 | 4.8628 |
| La Cruz | 6.2484 | 26.6755 | 4.2779 | 14.3090 | 6.0351 | 118.6721 | 0.5029 | 14.7439 |
| Liberia | 3.6829 | 23.1675 | 2.0260 | 10.2626 | 4.0913 | 92.4202 | 0.5693 | 16.2507 |
| Limon | 2.9640 | 17.1259 | 1.6025 | 6.7593 | 1.3724 | 11.2322 | 0.1034 | 1.7042 |
| Matina | 5.1143 | 19.9086 | 2.7162 | 7.4890 | 58.0750 | 192.2903 | 0.1232 | 3.0869 |
| Montes de Oro | 5.9220 | 21.8260 | 4.1289 | 12.2332 | 11.4909 | 126.5836 | 1.2106 | 19.5139 |
| Nicoya | 2.8036 | 20.5059 | 2.0672 | 10.1305 | 3.6262 | 101.6524 | 0.1894 | 10.2497 |
| Orotina | 13.7161 | 17.5376 | 4.0681 | 6.0472 | 1.0599 | 6.7613 | 0.4888 | 1.5221 |
| Osa | 5.9295 | 33.7053 | 4.2208 | 17.2559 | 1.1235 | 13.0806 | 0.5758 | 1.7891 |
| Parrita | $1.24 \times 10^9$ | 24483.9181 | 13.7622 | 9.8340 | 4.4446 | 46.3670 | 0.6773 | 7.9705 |
| Perez Zeled´on | 2.0103 | 30.0358 | 0.8733 | 9.4268 | 1.2543 | 20.3380 | 0.1783 | 1.3942 |
| Pococ´ı | 2.2524 | 12.7658 | 1.3001 | 6.1328 | 1.0524 | 12.0212 | 0.1284 | 1.8534 |
| 蓬塔雷纳斯 | 1.5802 | 12.8416 | 0.9303 | 4.6933 | 1.3263 | 23.9665 | 0.1646 | 1.7646 |
| 克波斯 | 56.8344 | 37.6499 | 10.0908 | 12.9868 | 36.9331 | 257.7509 | 1.1780 | 23.2153 |
| 圣何塞 | 0.2105 | 12.6814 | 0.1328 | 5.2650 | 0.0215 | 1.2409 | 0.0155 | 2.3248 |
| 圣安娜 | 0.7837 | 21.9311 | 0.5860 | 12.6213 | 0.6759 | 43.5264 | 0.3576 | 15.4854 |
| 圣克鲁斯 | 35.1198 | 37.3329 | 8.4762 | 12.7070 | 6.5905 | 91.5896 | 0.3027 | 3.0635 |
| 萨拉皮基´ı | 7.9218 | 22.2392 | 2.8096 | 6.9807 | 0.3880 | 4.9820 | 0.1047 | 1.9942 |
| 锡基雷斯 | 2.1184 | 13.9371 | 1.4077 | 6.6140 | 2.7214 | 19.0646 | 0.3564 | 1.7977 |
| 塔拉曼卡 | 4.9193 | 21.1751 | 2.2374 | 8.0522 | 0.1113 | 0.7642 | 0.7989 | 1.3152 |
| 图里亚尔瓦 | 2.5905 | 25.2386 | 1.9572 | 15.8268 | 1.0122 | 20.0694 | 0.2614 | 1.8489 |
| 乌帕拉[1] | 1.2744 | 21.7159 | 0.9203 | 12.2963 | - | - | - | - |

[1] $NRMSE$乌帕拉测试集的 $NIS_{0.05}$未显示，因为观测到的相对风险为零。

无法预测训练期和测试期。相比之下，采用适当条件自回归空间结构的选定模型能够显著降低其预测指标。此外，所提出模型的精确度优于简单的预测程序（见补充材料中的表A.1）。

Parrita附近有两个预测指标中等偏高的市镇，即Garabito和克波斯，它们位于该国的中太平洋海岸。我们推测，难以预测这些特定地区的主要原因是它们属于旅游区，而这些地区的登革热病例可能存在漏报。

除了气候协变量对模型的贡献外，时间和空间随机效$\phi_{i、(month)}$和$\theta_{i、(year)}$应（）也是登革热变化建模中的重要因素，因为它们提供了通过选定协变量无法观测到的时间或空间信息。图1展示了市镇特定的月度随机效应。具有相似时间行为的市镇被划分为第1至第7组，最后一组由不具有特定行为的市镇组成。随后，图2展示了这些市镇组的地理位置。我们观察到，时间随机效应提供的信息与该国微气候导致的时间变化大体一致。

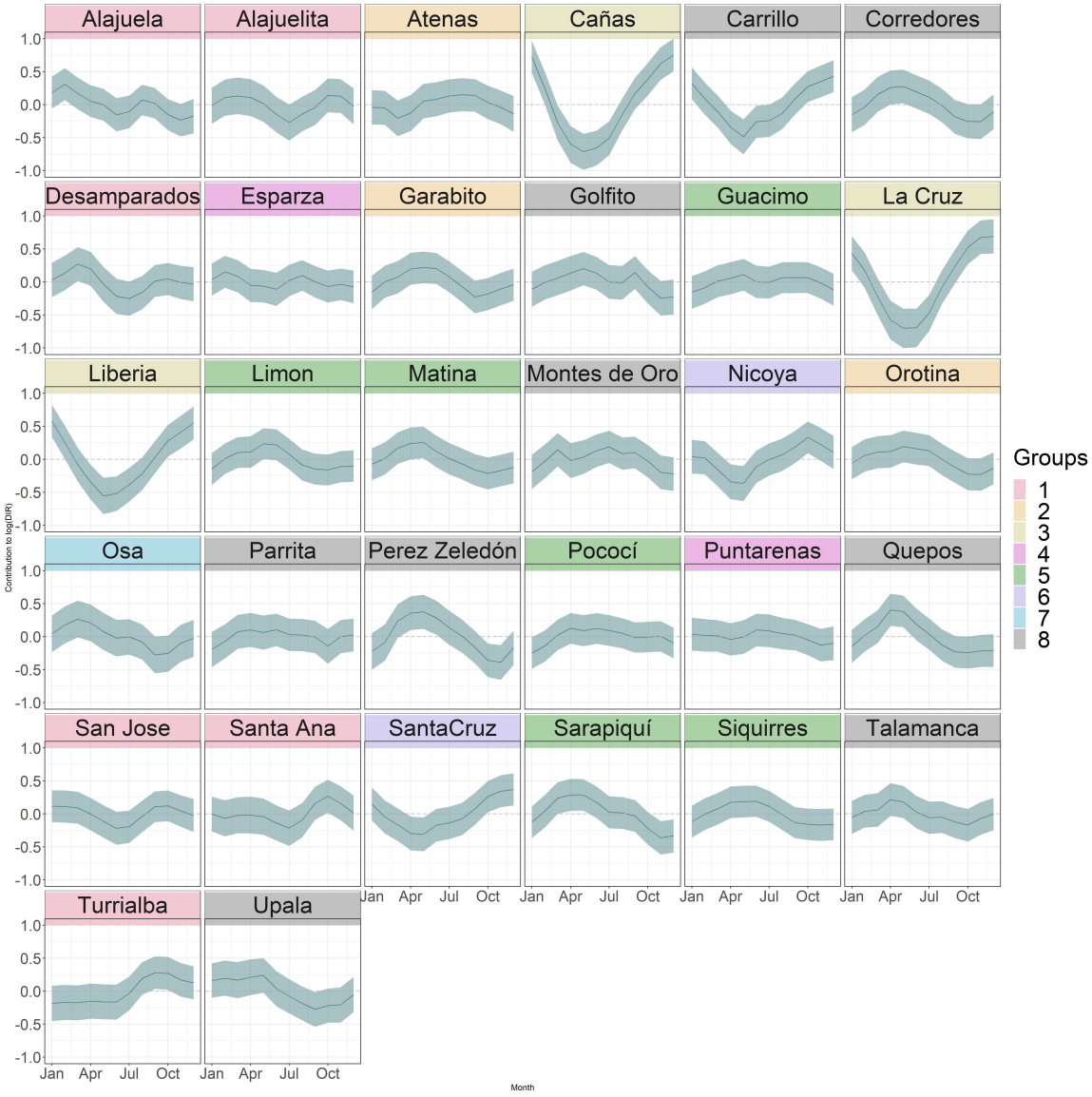



图1：市镇特定的月度随机效应的后验均值和95%可信区间。

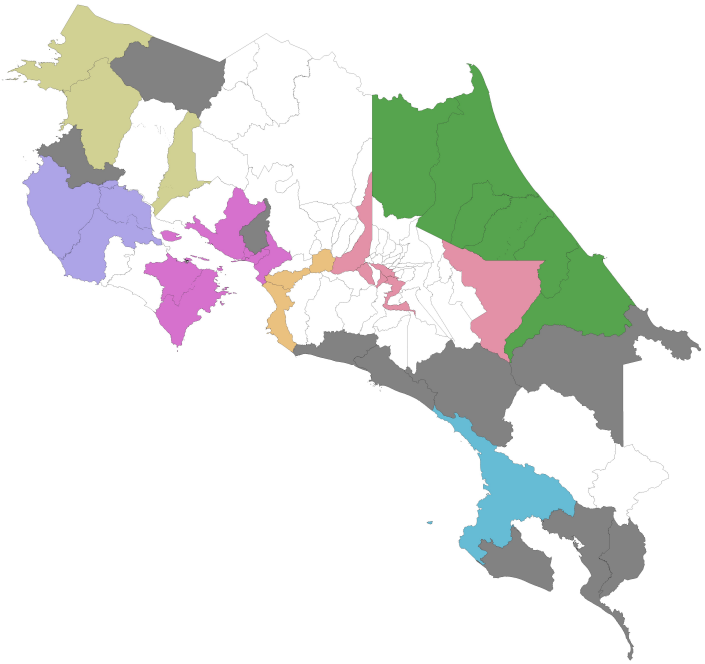


图2：具有相似时间行为的八个组的示意图。

另一方面，每年的空间随机效应如图3所示。我们观察到，不同年份存在一些州集群，其相对风险的对数较高或较低。例如，2002年该国南太平洋地区以及北太平洋地区的相对风险对数贡献较低，而2009年中太平洋地区的相对风险对数贡献较高。

需要注意的是，这种空间随机效应表示气候协变量未能解释的相对风险对数变异。尽管如此，在某种程度上，可以通过邻居的空间结构对其进行建模。众所周知，登革热是一种复杂现象，不仅涉及气候因素，还涉及社会经济因素和人口流动。此外，不同地区的某些市镇在不同时期也会发生疫情暴发，而本研究中的气候协变量无法解释这些疫情。因此，与假设空间结构相互独立的基线模型相比，该模型更加全面。

登革热预测可从时间或空间维度进行可视化。对于时间维度，图4 展示了在训练期内，根据$NIS_{0.05}$ ，表现最佳的六个市镇和表现最差的三个市镇的95%后验预测登革热相对风险。在测试期内，图5 展示了相同市镇在测试期内的%后验预测登革热相对风险。总体而言，登革热相对风险的变化可以得到精确建模，预测分布也能够涵盖测试期内观测到的RR。需要指出的是，由于相对风险本身具有非对称性，且模型以相对风险的对数为建模对象，因此预测不确定性呈正偏态。

对于空间维度，图6 展示了后验预测登革热相对风险

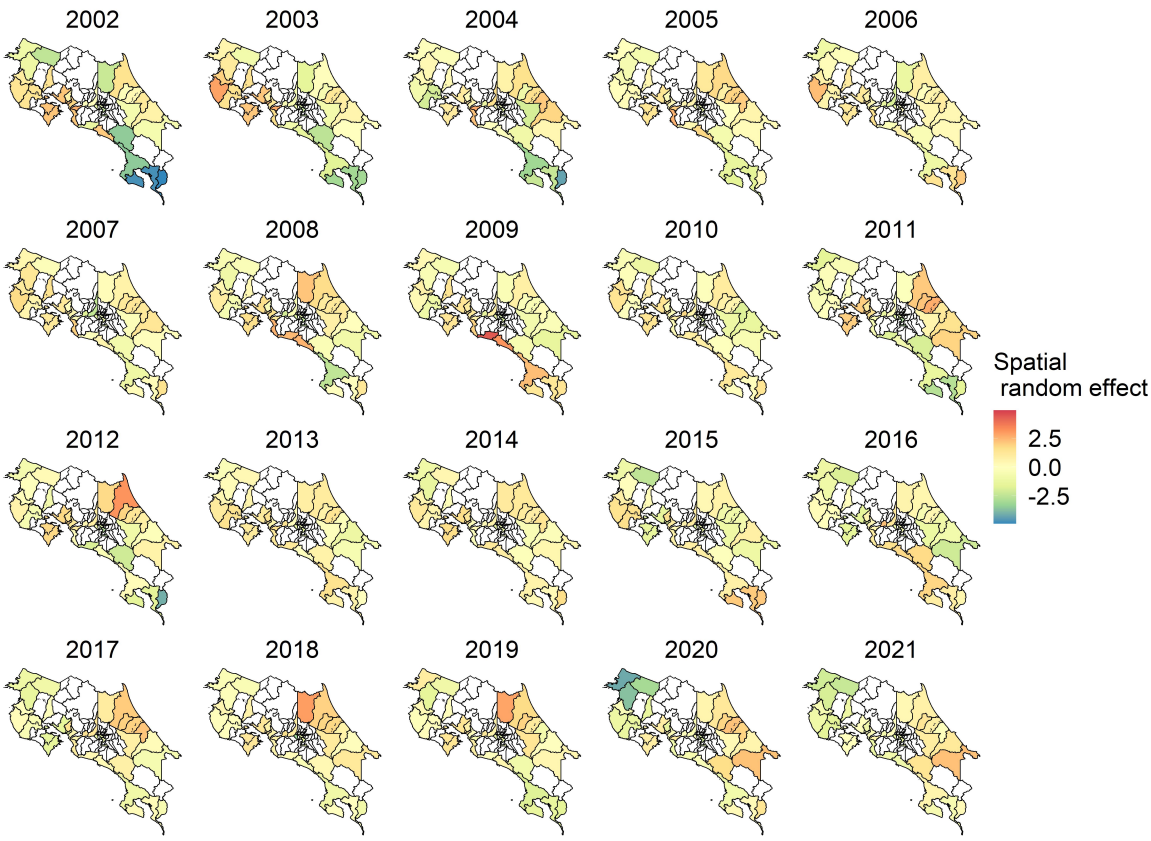



图3: 特定年份的空间随机效应对登革热相对风险对数值的贡献 k.

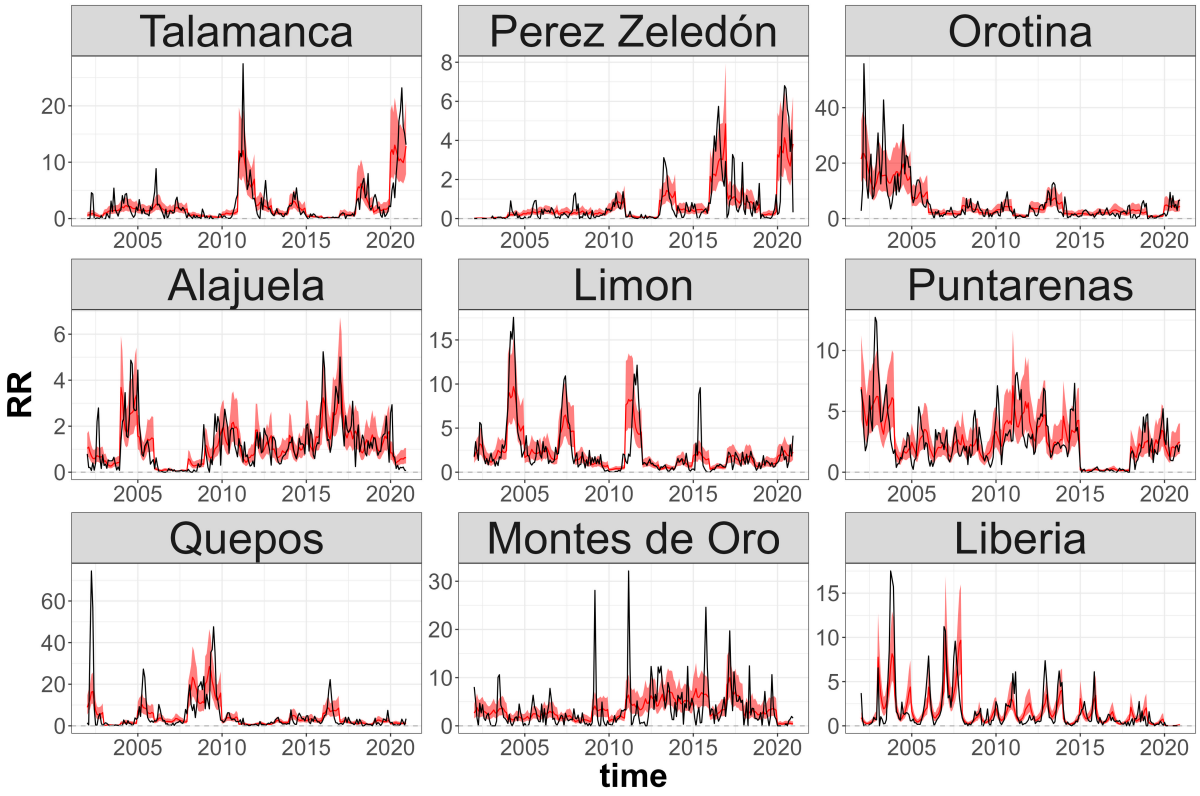



图4：训练期内观测到的登革热相对风险（黑色）和95%后验预测登革热相对风险（红色）。上方六个面板：根据 NIS 指标评出的表现最佳的市镇。下方三个面板：根据 NIS 指标评出的表现最差的市镇。

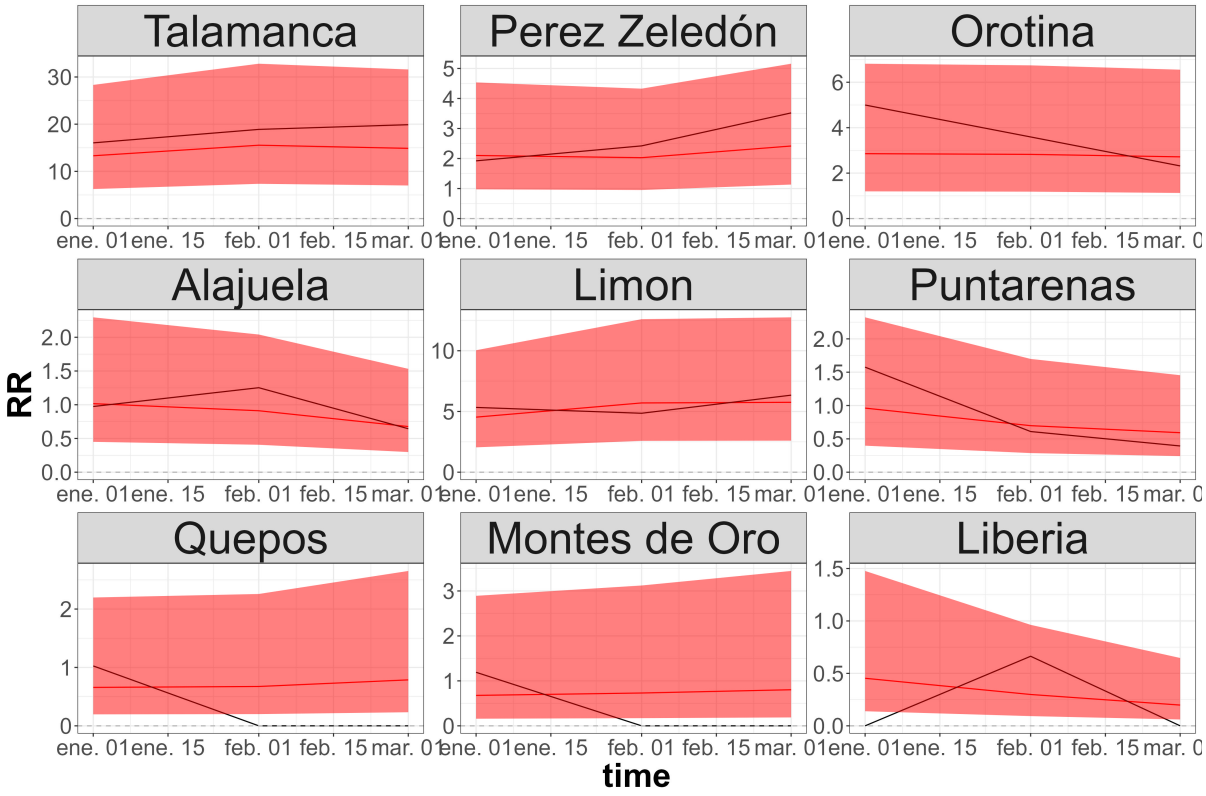
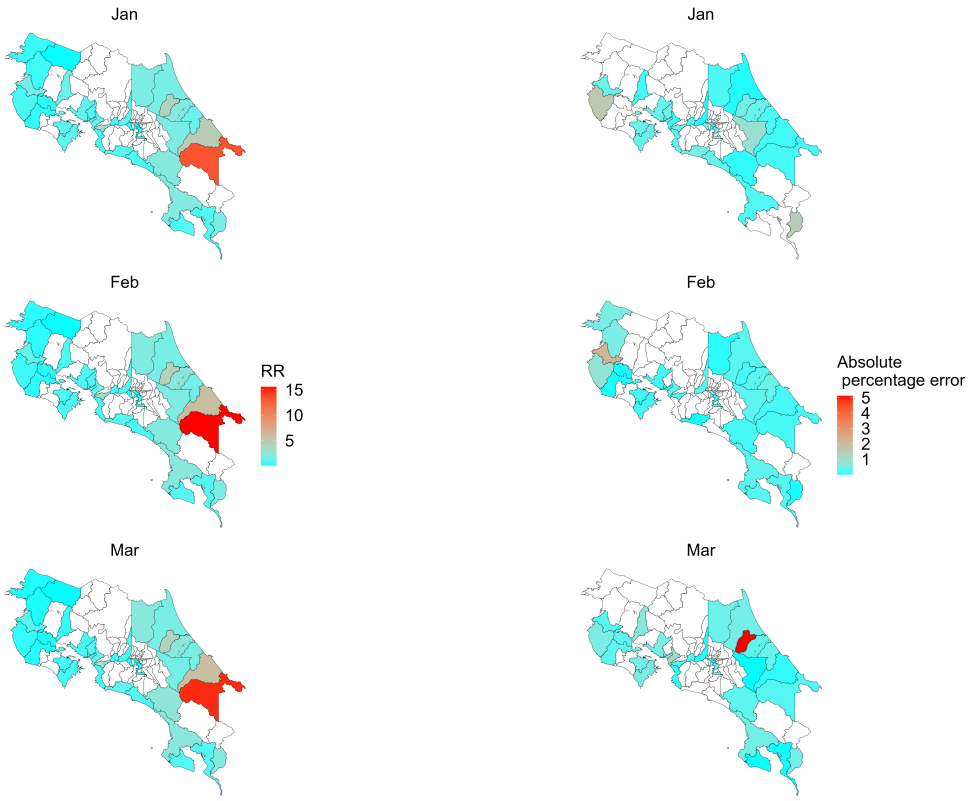



图 5：测试期内观测到的登革热相对风险（黑色）和95%后验预测登革热相对风险（红色）。上方六个面板：根据 NIS 指标评估的最佳市镇。下方三个面板：根据 NIS 指标评估的最差市镇。

均值及其在各市镇和月份（一月、二月和三月）的绝对百分比误差。这些地图可以帮助我们直观地了解特定月份登革热相对风险较高的区域。此外，图6b 显示，在测试期内，绝对百分比误差在整个区域内分布均匀。




(a) 相对风险预测。　　(b) 绝对百分比误差。

图6：2020年一月至三月哥斯达黎加81个市镇的相对风险预测及其绝对百分比误差。

对于训练集，也观察到类似的情况。补充材料的C节和D节展示了三个选定年份（2002年、2011年、2020年）的相对风险后验均值地图，以及同年绝对百分比误差地图。

# 4 讨论

流行病学模型是帮助公共卫生部门和政策制定者有效、高效地分配有限资源的重要工具。使用这些预测模型有助于了解疾病动态、疫情暴发可能对人口造成的影响，以及卫生系统的应对措施。纳入人口统计学特征、遗传学和环境变量，可改进这些模型，从而提高预测的准确性。

开发和使用流行病学模型面临的挑战之一，是获取高质量数据。稳健而全面的数据收集系统对于为模型提供准确的输入至关重要。此外，必须以一致、及时的方式整合和处理数据，以确保模型准确且保持最新。数据源

包括健康记录、实验室结果、监测系统和家庭调查。这些模型的输出有助于制定与防控目的和监测方法相关的决策，也有望成为有效的预测工具。预测是监测系统的一部分，更具体地说，是预警系统的一部分 [43]。

气候变异和气候变化对登革热传播的影响复杂、非线性，而且通常会延迟数周至数月 [13, 44]。在数学统计模型中结合登革热传播的时空模式，以及与气候和其他社会生态变化长期数据相契合的信息，能够改善登革热风险预测和疾病防控。

根据将模型（1）拟合登革热数据的结果可知，纳入所有气候协变量显著提升了模型准确预测登革热病例数的表现。模型采用不同的结构，有助于深入理解登革热传播与气候因素之间的关系。模型结果表明，它有望成为疾病防控和监测方法中决策过程的实用工具。此外，模型成功应用于2000年一月至2021年3月的月度数据，凸显了其在未来预测和预警系统中的潜力。

模型中的时空随机成分提醒我们，可以考虑更多结构性信息，例如人口流动和社会经济因素，以获得更好的预测结果。获取这些数据颇具挑战，我们将在未来的研究中继续努力。

最后，通过提供准确的预测，决策者能够更有效地应对暴发疫情，并实施策略以减轻疾病对人口的影响，从而帮助卫生部门优化通常较为稀缺的资源。为最大限度地发挥模型的效用，模型必须以高质量数据为基础，持续更新，并与公共卫生官员共同验证，以反映环境变化及其他可能影响疾病传播的因素，包括社会决定因素。持续开发此类模型将在抗击登革热及其他传染病的过程中发挥关键作用。

## 参考文献

[1] 世界卫生组织（WHO）。登革热和重症登革热，2022年8月。[2] Jolyon M Medlock、David Avenell、Iain Barrass 和 Steve Leach。分析白纹伊蚊（双翅目：蚊科）在英国存活和季节性活动的潜力。*Vector Ecology*，31(2):292–304，2006年。[3] Roberto Romi、Francesco Severini 和 Luciano Toma。罗马白纹伊蚊雌蚊的低温驯化和越冬。*Journal of the American Mosquito ControlAssociation*，22(1):149–151，2006年。[4] Abdelkrim Outammassine、Said Zouhair 和 Souad Loqman。当前和未来气候条件下三种未受充分重视的虫媒病毒媒介（日本伊蚊、骚扰伊蚊和斑翅伊蚊）的全球潜在分布。*Transboundary and Emerging Diseases*，69(4):e1160–e1171，2022年。[5] Natasha Evelyn Anne Murray、Mikkel B Quam 和 Annelies Wilder-Smith。登革热流行病学：过去、现在与未来展望。*Clinical Epidemiology*，5:299，2013年。

[6] Eduardo Massad、Marcos Amaku、Francisco Antonio Bezerra Coutinho、Claudio Jos´e Struchiner、Marcelo Nascimento Burattini、Kamran Khan、Jing Liu-Helmersson、Joacim Rockl¨ov、Moritz UG Kraemer 和 Annelies Wilder-Smith。估算登革病毒传入欧洲及出现继发本地病例的概率。*Scientific Reports*，8(1):1–12，2018年。[7] Luis Fernandez Lopez、Marcos Amaku、Francisco Antonio Bezerra Coutinho、Mikkel Quam、Marcelo Nascimento Burattini、Claudio Jos´e Struchiner、Annelies Wilder-Smith 和 Eduardo Massad。经由旅行者传播的传染病输入与输出建模。*Bulletin of Mathematical Biology*，78(2):185–209，2016年。[8] Duane J Gubler。登革热的经济负担。*The American Journal of Tropical Medicine and Hygiene*，86(5):743，2012年。[9] Hao Wang、Shaohua Zhao、Shengjun Wang、Yue Zheng、Shaohua Wang、Hui Chen、Jiao- jiao Pang、Juan Ma、Xiaorong Yang 和 Yuguo Chen。过去30年来脑炎负担的全球规模及其演变模式。*Journal of Infection*，84(6):777–787，2022年。[10] Jane P Messina、Oliver J Brady、Nick Golding、Moritz UG Kraemer、GR Wint、Sarah E Ray、David M Pigott、Freya M Shearer、Kimberly Johnson、Lucas Earl 等。登革热当前和未来的全球分布及风险人口。*Nature Microbiology*，4(9):1508–1515，2019年。[11] Xiaorong Yang、Mikkel BM Quam、Tongchao Zhang 和 Shaowei Sang。过去30年来登革热的全球负担及其演变模式。*Journal of Travel Medicine*，28(8):taab146，2021年。[12] Cory W Morin、Andrew C Comrie 和 Kacey Ernst。气候与登革热传播：证据及启示。*Environmental Health Perspectives*，121(11-12):1264–1272，2013年。[13] Suchithra Naish、Pat Dale、John S Mackenzie、John McBride、Kerrie Mengersen 和 Shilu Tong。气候变化与登革热：对定量建模方法的批判性系统综述。*BMC Infectious Diseases*，14(1):1–14，2014年。[14]Karen M Campbell、Kristin Haldeman、Chris Lehnig、Cesar V Munayco、Eric S Halsey、V Alberto Laguna-Torres、Mart´ín Yagui、Amy C Morrison、Chii-Dean Lin 和 Thomas W Scott。天气调节登革病毒在人与蚊子之间传播的地点、时间和强度。*PLoS Neglected Tropical Diseases*，9(7):e0003957，2015年。[15] Douglas M Watts、Donald S Burke、Bruce A Harrison、Richard E Whitmire 和 Ananda Nisalak。温度对白纹伊蚊传播登革病毒2型媒介效能的影响。技术报告，美国陆军传染病医学研究所（马里兰州德特里克堡），1986年。[16] W Tun-Lin、TR Burkot 和 BH Kay。温度和幼虫食物对白纹伊蚊发育速度及存活的影响：以澳大利亚昆士兰北部为例。*Medical and Veterinary Entomology*，14(1):31–37，2000年。

[17] LM Rueda、KJ Patel、RC Axtell 和 RE Stinner。温度对致倦库蚊和埃及伊蚊（双翅目：蚊科）发育与存活率的影响。医学昆虫学杂志，27(5):892–898，1990年。[18] Muhammad Shahzad Sarfraz、Nitin K Tripathi、Taravudh Tipdecho、Thawisak Thongbu、Pornsuk Kerdthong 和 Marc Souris。运用因子分析和空间环形制图分析登革热媒介幼虫密度与土地利用之间的时空关系。*BMC*公共卫生，12(1):1–19，2012年。[19] Hamady Dieng、Abu Hassan Ahmad、Jazem A Mahyoub、Abdulhafis M Turkistani、Hatabbi Mesed、Salah Koshike、Tomomitsu Satho、MR Che Salmah、Hamdan Ahmad、Wan Fatma Zuharah 等。沙特阿拉伯麦加市家户调查：容器孳生蚊及影响埃及伊蚊（双翅目：蚊科）流行率的气候因素。亚太热带生物医学杂志，2(11):849–857，2012年。[20]Birgit HB Van Benthem、Sophie O Vanwambeke、Nardlada Khantikul、Chantal Burghoorn-Maas、Kamolwan Panart、Linda Oskam、Eric F Lambin 和 Pradya Somboon。登革热感染血清阳性的空间分布模式及风险因素。《美国热带医学与卫生杂志》，72(2):201–208，2005年。[21] Roberto Barrera、Manuel Amador 和 Gary G Clark。影响波多黎各萨利纳斯人工容器中埃及伊蚊（双翅目：蚊科）产出率的生态因素。医学昆虫学杂志，43(3):484–492，2006年。[22] Julio C´esar Mateus 和 Gabriel Carrasquilla。当地疟疾暴发的预测因素：哥伦比亚早期预警系统的开发方法。*Mem´orias doInstituto Oswaldo Cruz*，106:107–113，2011年。[23] Millicent Eidson、Laura Kramer、Ward Stone、Yoichiro Hagiwara、Kate Schmit、纽约州西尼罗病毒禽类监测小组等。死鸟监测作为西尼罗病毒早期预警系统。新发传染病，7(4):631，2001年。[24] 欧洲疾病预防控制中心。Vbornet——欧洲人类公共卫生节肢动物媒介监测网络，2022年8月。[25] 哥斯达黎加卫生部。哥斯达黎加卫生部网站。欢迎，2022年。[26] Paola V´asquez、Antonio Lor´ıa、Fabio Sanchez 和 Luis A. Barboza。气候驱动统计模型可有效预测哥斯达黎加当地登革热发病率：广义加性模型与随机森林方法。*Revista de Matematica:Teoria y Aplicaciones*，27(1):1–21，2020年。[27] Yury E. Garcia、Shu-Wei Chou-Chen、Luis A. Barboza、Maria L. Daza-Torres、J. Cricelio Montesinos-Lopez、Paola Vasquez、Juan G. Calvo、Miriam Nuno 和 Fabio Sanchez。哥斯达黎加登革热病例、气候与当地环境变量之间的共同模式：一种小波方法，2023年。[28] 哥斯达黎加卫生部。哥斯达黎加卫生部网站。欢迎，2022年。[29] Chris Funk、Pete Peterson、Martin Landsfeld、Diego Pedreros、James Verdin、Shraddhanand Shukla、Gregory Husak、James Rowland、Laura Harrison、Andrew Hoell 和

Joel Michaelsen。站点红外降水气候危害——一种新的环境……
用于监测极端事件的环境记录。Scientific Data, 2(1):150066–150066, 2015.[30] 美国国家海洋和大气管理局。气候预测中心。https://www.cpc.ncep.noaa.gov/data/indices/ersst5.nino.mth.91-20.ascii。访问日期：2022-05-01。[31] Sean L. Tuck、Helen R.P. Phillips、Rogier E. Hintzen、J”orn P.W. Scharlemann、Andy Purvis 和 Lawrence N. Hudson。Modistools——在 R语言 中下载和处理 MODIS 遥感数据。Ecology and Evolution, 4(24):4658–4668, 2014.[32] David B Enfield、Alberto M Mestas-Nu˜nez、Dennis A Mayer 和 Luis Cid-Serrano。热带大西洋海表温度中的偶极关系有多普遍? Journal of Geophysical Research: Oceans, 104(C4):7841–7848, 1999.[33] H G Hidalgo、E J Alfaro 和 B Quesada-Montano。利用高分辨率气象数据集观测中美洲（1970–1999年）的气候变率及其对气候变化研究的意义。Clim. Change, 141(1):13–28, 2017.[34] A. Gasparrini、B. Armstrong 和 M. G. Kenward。分布滞后非线性模型。Statistics in Medicine, 29(21):2224–2234, 2010.[35] Antonio Gasparrini。利用分布滞后非线性模型对暴露–滞后–反应关联进行建模。Statistics in Medicine, 33(5):881–899, 2014.[36] Julian Besag、Jeremy York 和 Annie Molli´e。贝叶斯图像复原及其在空间统计学中的两个应用。Annals of the Institute of Statistical Mathematics, 43(1):1–20.[37] Roger Bivand、Virgilio G´omez-Rubio 和 H˚avard Rue。使用 R-INLA 进行空间数据分析及其扩展。Journal of Statistical Software, 63(20):1–31, 2015.[38] A. Gasparrini。R语言中的分布滞后线性和非线性模型：dlnm 软件包。Journal of Statistical Software, 43(8):1–20, 2011.[39] H̆avard Rue、Sara Martino 和 Nicolas Chopin。使用积分嵌套拉普拉斯近似对潜在高斯模型进行近似贝叶斯推断。Journal of the Royal Statistical Society . Series B ( Methodological ), 71(2):319–392, 2009.[40] Luis A. Barboza、Shu-Wei Chou-Chen、Paola V´asquez、Yury E. Garc´ıa、Juan G. Calvo、Hugo G. Hidalgo 和 Fabio Sanchez。利用气候变量和机器学习技术评估哥斯达黎加的登革热风险。PLOS Neglected Tropical Diseases, 17(1):1–13, 01 2023.[41] Robert L Winkler 和 Allan H Murphy。“优秀的”概率评估者。Journal of Applied Meteorology and Climatology, 7(5):751–758, 1968.[42] Tilmann Gneiting 和 Adrian E Raftery。严格适当评分规则、预测与估计。Journal of the American Statistical Association, 102(477):359–378, 2007年3月。[43] Vanessa Racloz、Rebecca Ramsey、Shilu Tong 和 Wenbiao Hu。登革热病毒监测：流行病学模型与预警系统综述。PLoS neglected tropical diseases, 6(5):e1648, 2012。

[44] Tzai-Hung Wen、Neal H Lin、Chun-Hung Lin、Chwan-Chuen King 和 Ming-Daw Su。绘制时间风险特征的空间分布，以改进环境健康风险识别：台湾登革热疫情案例研究。*Science of the Total Environment*, 367(2-3):631–640, 2006。

# 补充材料：使用 INLA 的贝叶斯时空模型，用于预测哥斯达黎加的登革热风险

## 测试期（三月2021至三月2021）的朴素方法预测指标

表A.1：朴素预测方法和负二项分布零模型的测试数据集预测指标

| 市镇 | 朴素方法 | 负二项分布零模型 | |
|---|---|---|---|
| | $NRMSE$ | $NRMSE$ | $NIS_{95}$ |
| 阿拉胡埃拉 | 2.6506 | 1.4229 | 42.9843 |
| 阿拉胡埃利塔 | 0.4765 | 2.1888 | 59.8866 |
| 阿特纳斯 | 89.8765 | 7.3082 | 156.7569 |
| 卡˜纳斯 | 3.5234 | 2.9727 | 69.7955 |
| 卡里略 | 0.8118 | 0.9751 | 26.7633 |
| 科雷多雷斯 | 0.3051 | 0.7459 | 20.5897 |
| 德萨姆帕拉多斯 | 0.2598 | 22.3581 | 436.3609 |
| 埃斯帕尔萨 | 0.6795 | 0.9124 | 26.0958 |
| Garabito | 15.6450 | 9.0487 | 189.7400 |
| Golfito | 0.9871 | 3.6079 | 82.9118 |
| Guacimo | 3.6451 | 3.2932 | 28.8848 |
| La Cruz | 3.5519 | 17.4492 | 339.4951 |
| Liberia | 0.4927 | 16.3574 | 319.1370 |
| Limon | 3.4131 | 2.1836 | 23.9121 |
| Matina | 2.6768 | 0.8195 | 30.0298 |
| Montes de Oro | 41.8550 | 8.0778 | 160.1058 |
| Nicoya | 9.2749 | 17.0730 | 338.3457 |
| Orotina | 2.2117 | 0.9807 | 15.6226 |
| Osa | 1.2030 | 0.6291 | 13.9754 |
| Parrita | 15.7029 | 2.4511 | 54.8356 |
| 佩雷斯·塞莱东 | 1.5211 | 0.2745 | 7.9940 |
| 波科西 | 1.0359 | 0.2630 | 10.7672 |
| 蓬塔雷纳斯 | 0.7206 | 2.0884 | 52.3785 |
| 克波斯 | 10.8189 | 9.6902 | 192.1924 |
| 圣何塞 | 0.0267 | 11.4795 | 237.4494 |
| 圣安娜 | 0.4281 | 19.8108 | 383.3144 |
| 圣克鲁斯 | 2.3943 | 6.6282 | 144.8392 |
| 萨拉皮基´ | 13.7461 | 0.0603 | 3.2807 |
| 锡基雷斯 | 1.8468 | 0.2815 | 9.1988 |
| 塔拉曼卡 | 13.8805 | 14.4673 | 35.1501 |
| 图里亚尔瓦 | 1.4088 | 0.3812 | 14.9645 |
| 乌帕拉[1] | - | - | - |

[1] $NRMSE$ 和 $NIS_{95}$ 的乌帕拉结果未显示，因为观测到的相对风险为零。

# B 登革热病例建模与预测

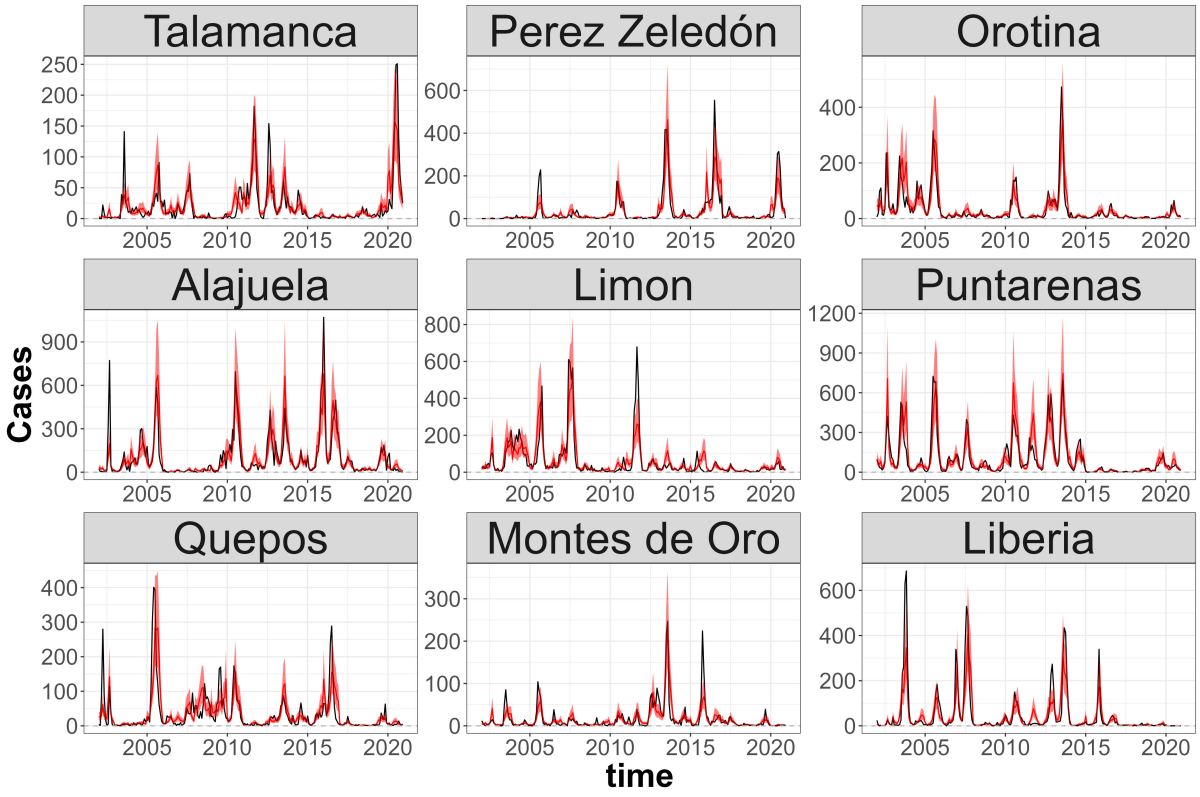



图 B.1：训练期内观测到的登革热病例（黑色）和 95% 后验预测登革热病例（红色）。上方六个面板：根据 NIS 指标表现最佳的市镇。下方三个面板：根据 NIS 指标表现最差的市镇。

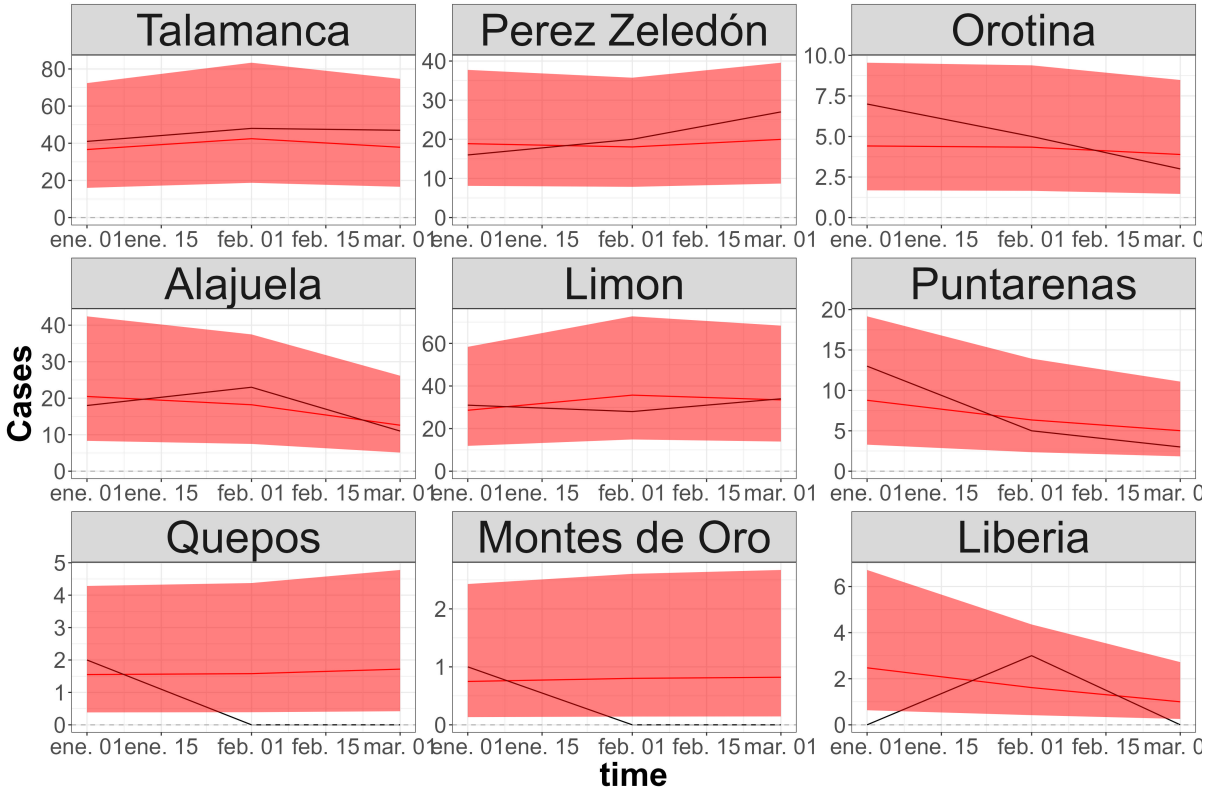
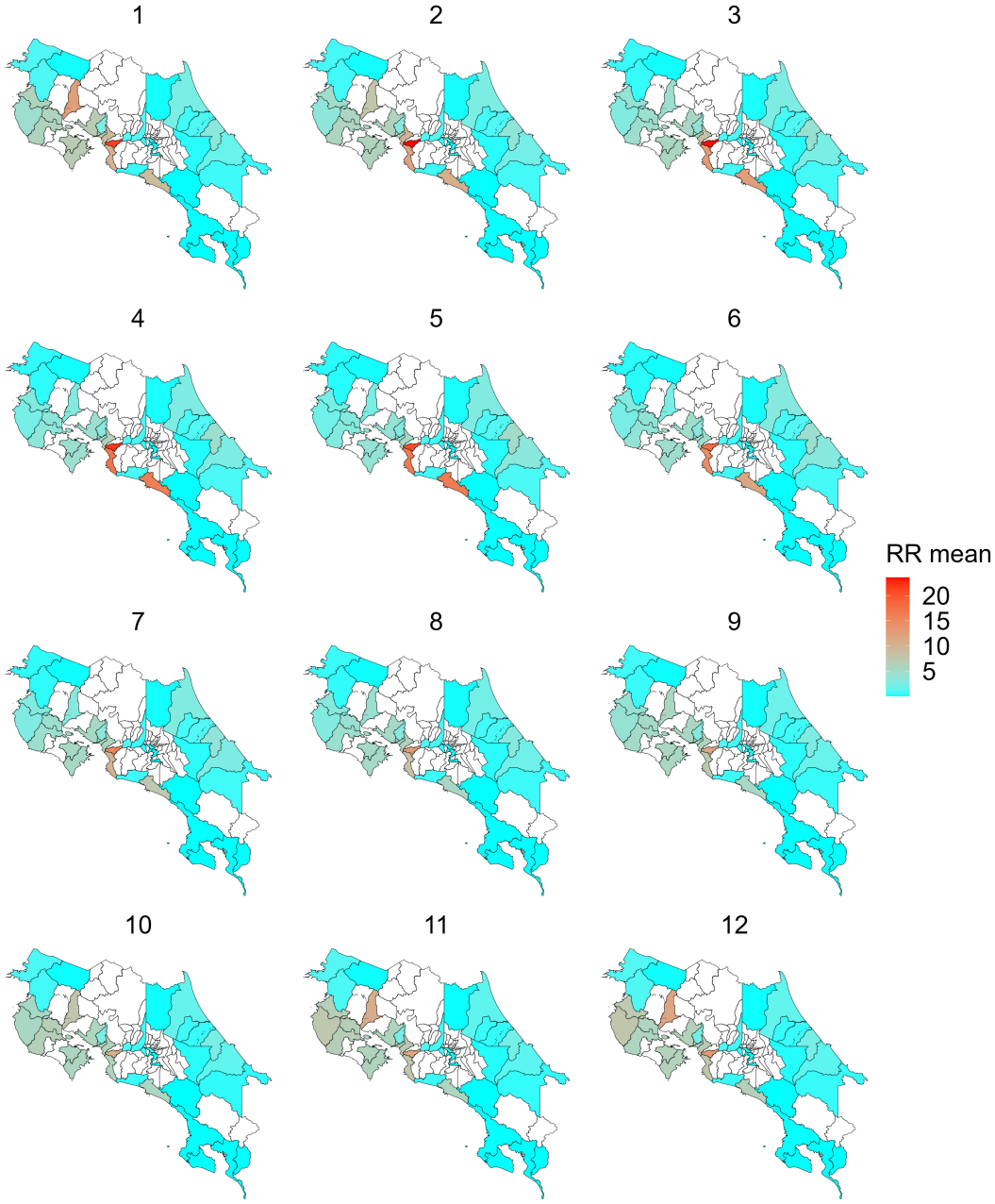



图 B.2：测试期内观测到的登革热病例（黑色）和 95% 后验预测登革热病例（红色）。上方六个面板：根据 NIS 指标评选出的最佳市镇。下方三个面板：根据 NIS 指标评选出的最差市镇。

## C 2002年、2011年和2020年相对风险预测图




图C.1：哥斯达黎加81个市镇2002年1月至12月相对风险的后验均值。

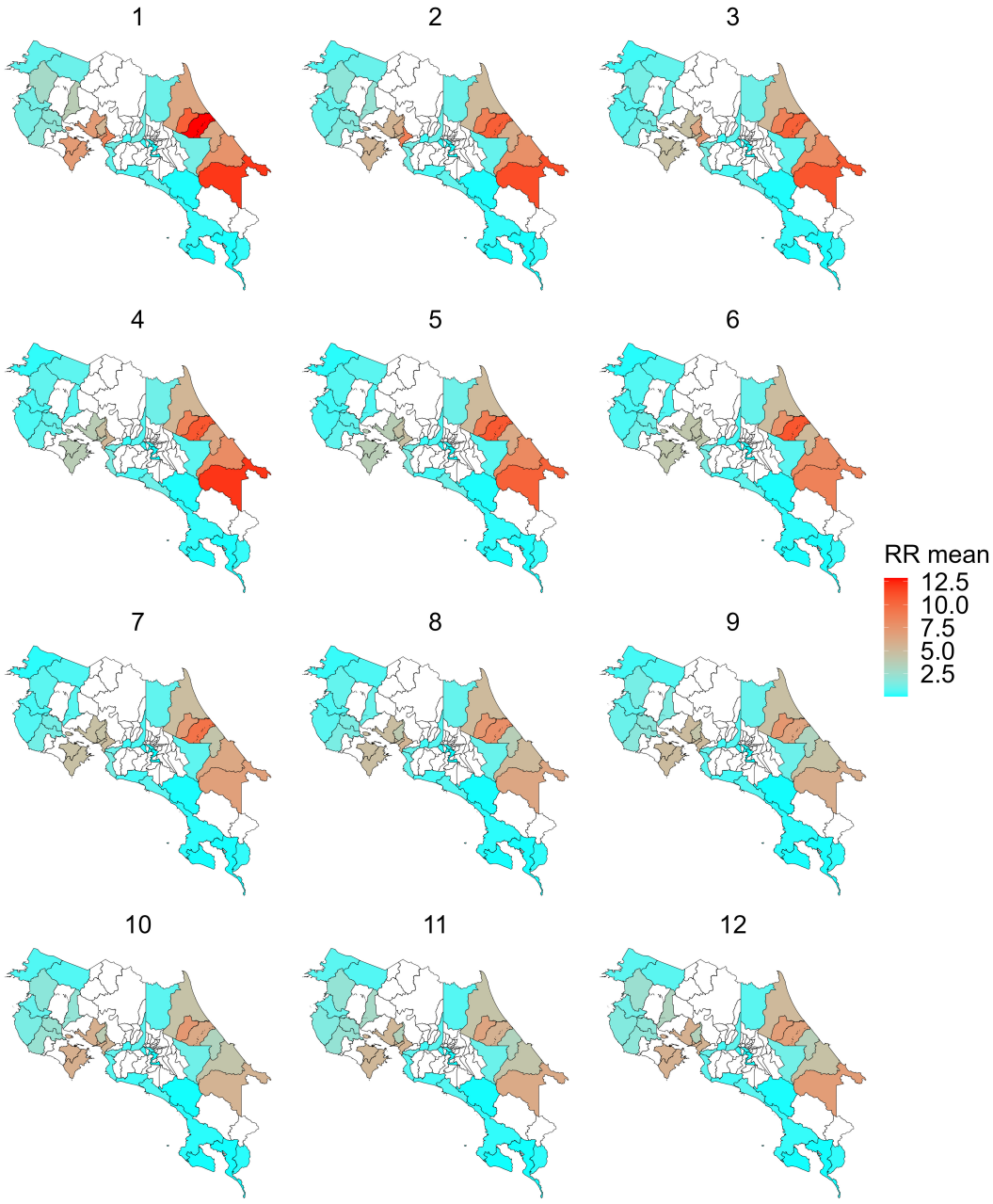



图 C.2：哥斯达黎加 81 个市镇 2011年1月至12月的相对风险后验均值。

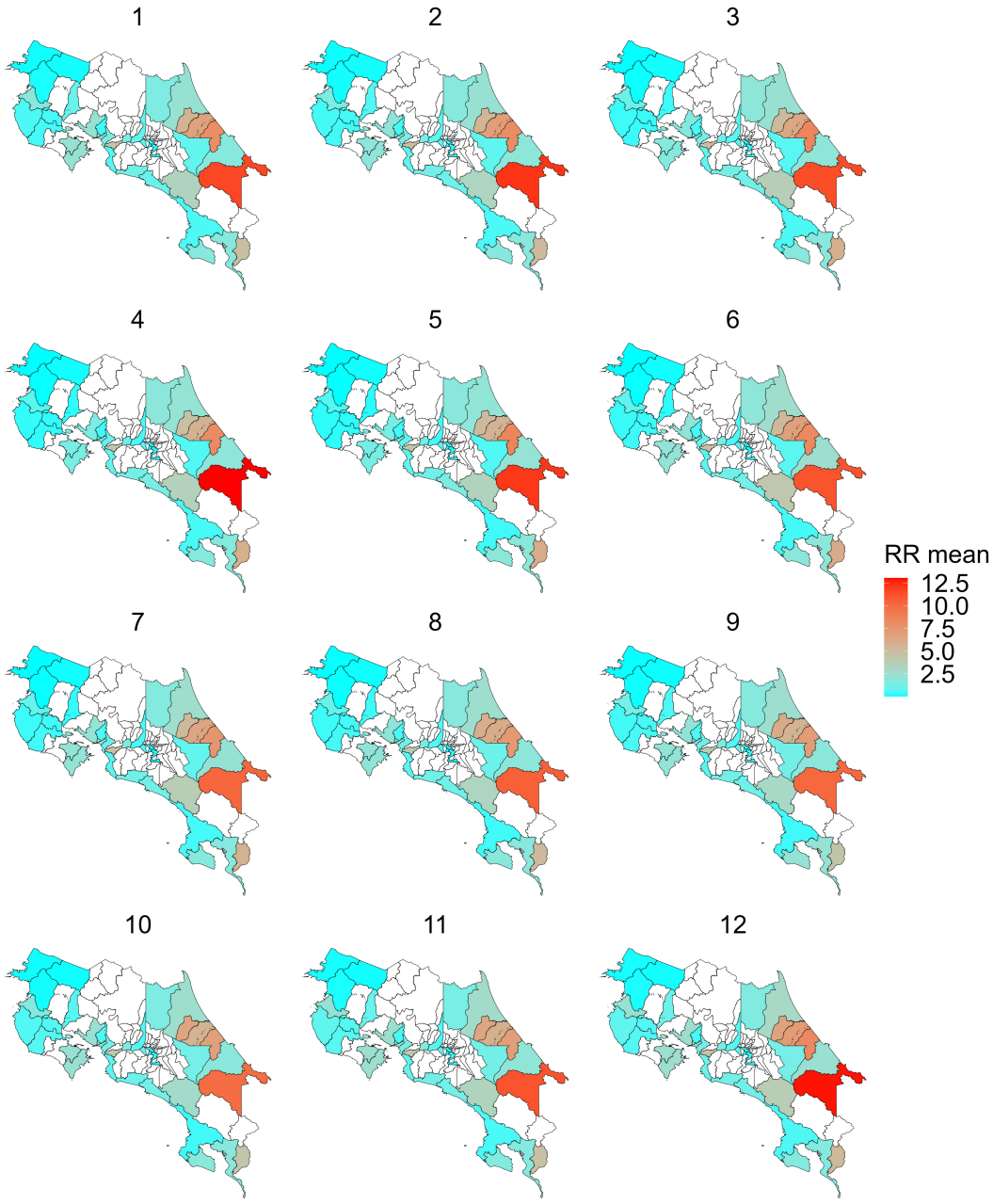
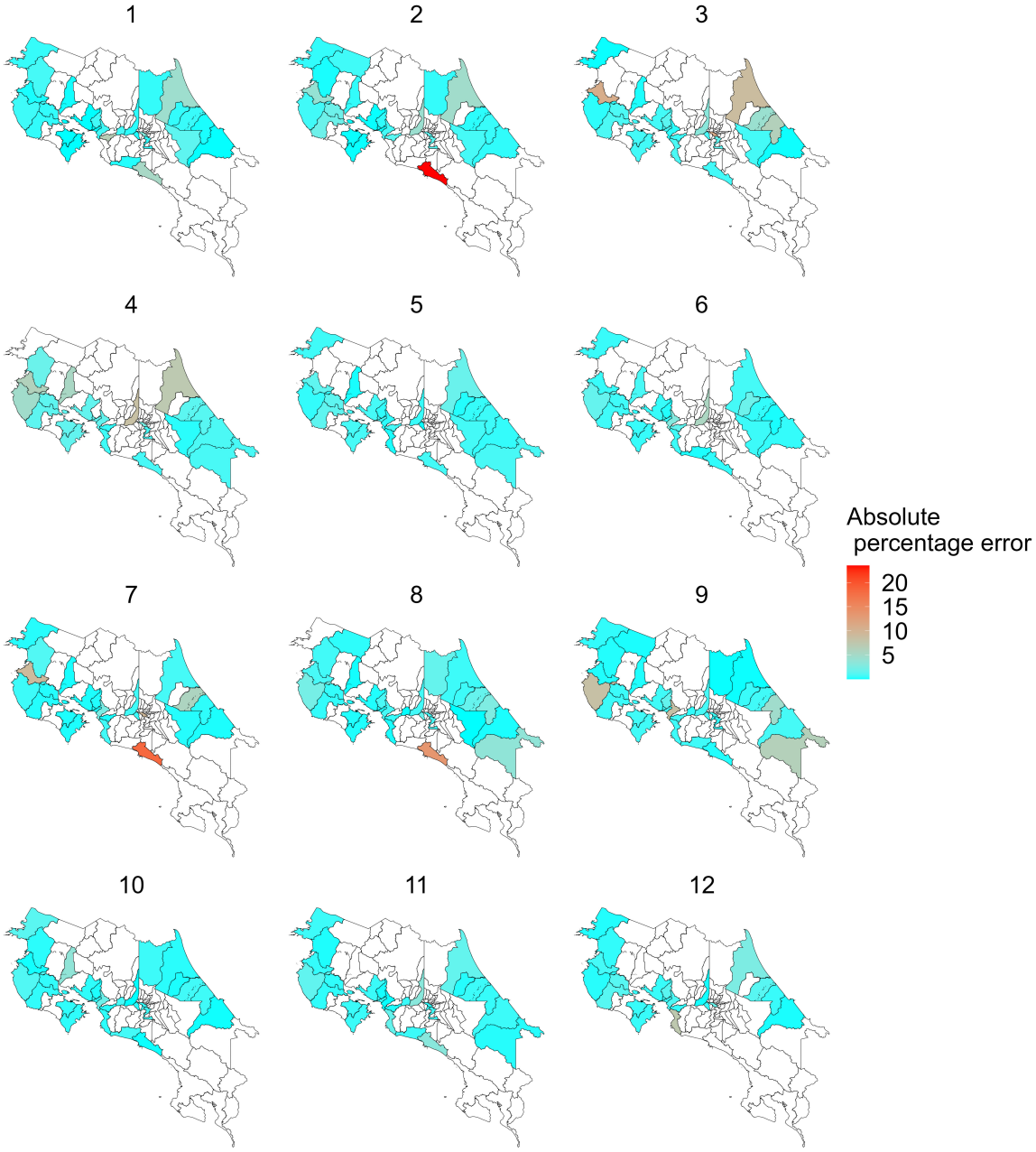



图 C.3：哥斯达黎加81个市镇2020年1月至12月的相对风险后验均值。

# D　2002年、2011年和2020年的绝对百分比误差地图




图D.1：哥斯达黎加81个市镇2002年一月至十二月的绝对百分比误差。

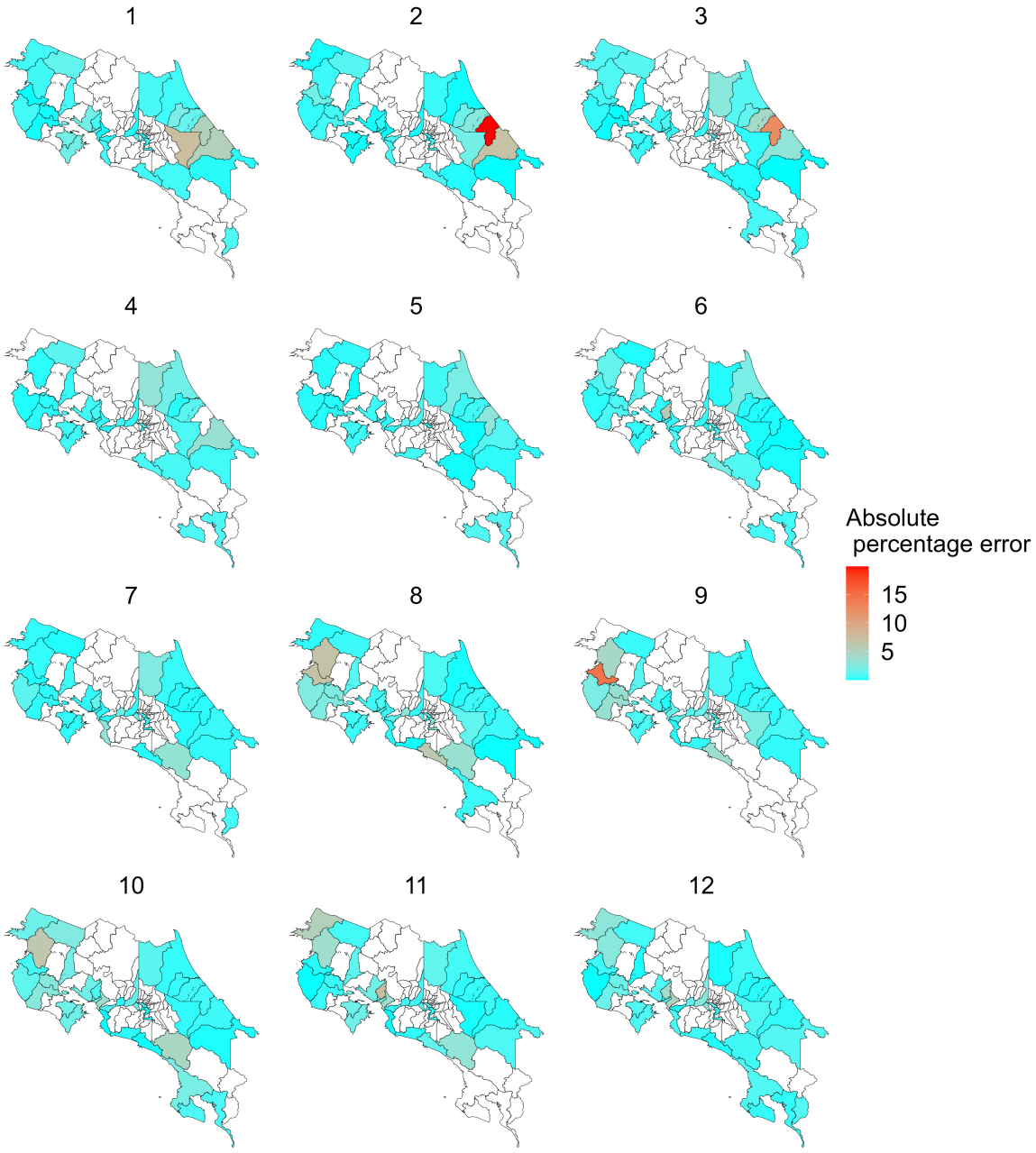



图 D.2：哥斯达黎加 81 个市镇 2011 年 1 月至 12 月的绝对百分比误差。

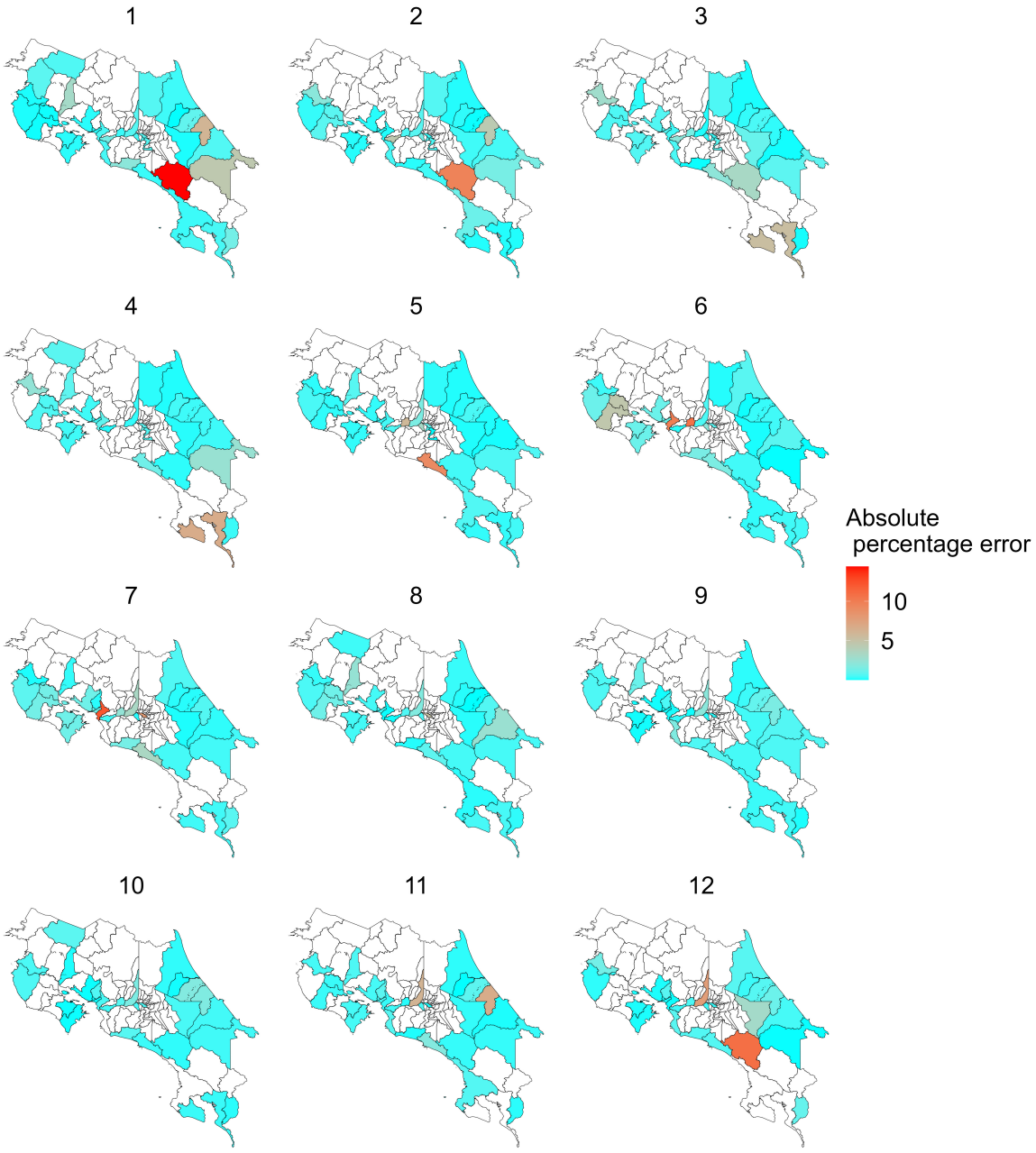



图 D.3：哥斯达黎加 81 个市镇 2020年1月至12月的绝对百分比误差。