Acta Tropica 178 (2018) 155–162

Contents lists available at ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica

Mapping the spatial distribution of *Aedes aegypti* and *Aedes albopictus*

Fangyu Ding[a,b,1], Jingying Fu[a,b,1], Dong Jiang[a,b,*], Mengmeng Hao[a,b], Gang Lin[a]

[a] Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[b] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

A R T I C L E I N F O

Keywords:
Global distribution
Aedes aegypti
Aedes albopictus
Multidisciplinary datasets
Machine learning models

A B S T R A C T

Mosquito-borne infectious diseases, such as Rift Valley fever, Dengue, Chikungunya and Zika, have caused mass human death with the transnational expansion fueled by economic globalization. Simulating the distribution of the disease vectors is of great importance in formulating public health planning and disease control strategies. In the present study, we simulated the global distribution of *Aedes aegypti* and *Aedes albopictus* at a 5 × 5 km spatial resolution with high-dimensional multidisciplinary datasets and machine learning methods Three relatively popular and robust machine learning models, including support vector machine (SVM), gradient boosting machine (GBM) and random forest (RF), were used. During the fine-tuning process based on training datasets of *A. aegypti* and *A. albopictus*, RF models achieved the highest performance with an area under the curve (AUC) of 0.973 and 0.974, respectively, followed by GBM (AUC of 0.971 and 0.972, respectively) and SVM (AUC of 0.963 and 0.964, respectively) models. The simulation difference between RF and GBM models was not statistically significant ($p > 0.05$) based on the validation datasets, whereas statistically significant differences ($p < 0.05$) were observed for RF and GBM simulations compared with SVM simulations. From the simulated maps derived from RF models, we observed that the distribution of *A. albopictus* was wider than that of *A. aegypti* along a latitudinal gradient. The discriminatory power of each factor in simulating the global distribution of the two species was also analyzed. Our results provided fundamental information for further study on disease transmission simulation and risk assessment.

1. Introduction

Arbovirus epidemics, such as Dengue, Chikungunya and Zika, are spreading widely with human movement and trade (Nunes et al., 2014; Brockmann and Helbing, 2013; Kyeongah et al., 2016; Weaver and Lecuit, 2015). For example, the Zika virus, first discovered in a rhesus monkey from the Zika forest of Uganda in 1947 (Dick et al., 1952), has caused several major epidemics in France and Brazil during the past five years (Campos et al., 2015; Cauchemez et al., 2016). Moreover, scientists have shown that the Zika virus disrupts the neural progenitor development and leads to microcephaly (Li et al., 2016) and the development of Guillain–Barré syndrome (GBS) (Brasil et al., 2016). With the increase in public health and safety awareness, the mosquitoes *Aedes aegypti* and *Aedes albopictus* have raised public health concerns because they are regarded as the major vectors for the transmission of these viral diseases to humans (Bargielowski et al., 2015; Vega-Rúa et al., 2014; Liu et al., 2017; Roslan et al., 2013; Nuckols et al., 2015). Understanding the distribution of *Aedes* mosquitoes can help prevent these human arbovirus infections(Santos and Meneses, 2017). Based on small-scale field trials and oviposition monitoring, some local studies have revealed that the most common habitats of the mosquitoes are artificial and natural containers, including discarded tins, used tires, tree holes and rock pools (Dutta et al., 1998; Chan et al., 1971; Simard et al., 2005; Hornby et al., 1994).

Previous studies have also attempted to evaluate the distribution of *A. aegypti* and *A. albopictus* at a regional or global scale. Based on larval surveillance datasets, Kobayashi et al. (2002) used geographic information systems (GIS) to analyze the relationship between climate conditions and the distribution of *A. albopictus* in northern Japan, illustrating that the accumulated temperature over 1350 °C-days may be close to the development zero of *A. albopictus* (Kobayashi et al., 2002). From a global viewpoint, Brady et al. (2014) adopted a biologically relevant temperature-based model to produce temperature suitability maps for these arbovirus vectors that considered the seasonal and diurnal variations in temperature and their cumulative effects on *Aedes* mosquitoes(Brady et al., 2014). Campbell et al. (2015) combined an ecological niche model with six climate models to predict possible changes in the distribution patterns of these mosquitoes in the future;

* Corresponding author at: Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.
E-mail addresses: dingfy.14s@igsnrr.ac.cn (F. Ding), fujy@igsnrr.ac.cn (J. Fu), jiangd@igsnrr.ac.cn (D. Jiang), haomm.16b@igsnrr.ac.cn (M. Hao), ling@lreis.ac.cn (G. Lin).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.actatropica.2017.11.020
Received 23 August 2017; Received in revised form 31 October 2017; Accepted 26 November 2017
Available online 27 November 2017

Table 1
Multidisciplinary datasets used to map the spatial distribution of the arbovirus vectors.

Factors	Scale	Format	Data sources
Min Temperature	1 × 1 km	Grid	WorldClim database, version 2.0
Max Temperature	1 × 1 km	Grid	
Precipitation	1 × 1 km	Grid	
NDVI	8 × 8 km	Grid	Global Inventory Modeling and Mapping Studies Group
Relative humidity	–	Shapefile	Surface Meteorology and Solar Energy, NASA
Urbanicity	–	Shapefile	Socioeconomic Data and Applications Center, NASA
Population Density	1 × 1 km	Grid	
Nighttime Lights	1 × 1 km	Grid	The Earth Observation Group, NOAA
Urban Accessibility	1 × 1 km	Grid	European Commission Joint Research Centre Global Environment Monitoring Unit

the results indicated that the potential range shifts in response to climate were slight (Campbell et al., 2015). To overcome the limitations of environmental variables, a global land-cover layer was used by Kraemer et al. (2015a,b) for the first time to produce a probabilistic species distribution of *A. aegypti* and *A. albopictus* with area under the curve (AUC) values of 0.87 and 0.90, respectively (Kraemer et al., 2015a). However, these studies were primarily based on climate and land-cover variables, and the impact of social-economic factors reflecting human movements and urbanization on the spread of the arbovirus vectors were neglected. Moreover, literature on the comparison of significant differences among models in the simulation process was relatively scarce.

Given the availability of data, multidisciplinary datasets (meteorological, environmental and social) and comprehensive species occurrence records were collected (Table 1). The objectives of this study were to simulate the global probabilistic species distribution of *A. aegypti* and *A. albopictus* using three relatively popular machine learning models based on the high-dimensional datasets. In addition, statistically significant differences between these models at a 5% confidence level were analyzed for the first time.

2. Materials and methods

All datasets used in this research were transformed into a unique geographic coordinate system, i.e., WGS-84. The global spatial distribution of *A. aegypti* and *A. albopictus* at a 5 × 5 km spatial resolution was produced using GIS software and the C++ programming language, including ArcMap 10.2 (http://www.esrichina.com.cn/), GDAL 2.1.0 (http://www.gdal.org/) and Proj. 4 (https://github.com/OSGeo/proj.4). The technical flow chart of the study is shown in Fig. 1.

2.1. Data acquisition

2.1.1. Climatic conditions

Temperature plays an important role in the growth of *Aedes* mosquitoes affecting key physiological processes in these vectors, including adult female survivorship and length of the first gonotrophic cycle (Brady et al., 2014; Nasci, 1986). For example, adult mosquito mortality increases markedly as the mean temperature exceeds 35 °C, and if mosquitoes cannot survive the first gonotrophic cycle, they are less likely to oviposit (Martens, 1998). Based on mark-release-recapture experiments and under controlled conditions in the laboratory, the thermal limits for both species were evaluated, showing that *A. aegypti* can tolerate a broader range of temperatures than *A. albopictus* (Brady et al., 2013).

Previous studies have established a relationship between precipitation and infection risk of dengue virus (Bhatt et al., 2013; Restrepo et al., 2014; Sang et al., 2014; Hsieh et al., 2013). Considering that these two mosquito species are the primary vectors of the virus, we assumed precipitation would affect mosquito growth. Moreover, some local studies have shown that high precipitation areas are generally associated with vector abundance (Scott et al., 2000; Romero-Vivas and Falconar, 2005; Linthicum et al., 1985). Thus, an annual cumulative precipitation layer derived from a monthly mean precipitation dataset was chosen as one of the input data layers in the present study.

Global climate datasets were obtained from the WorldClim database version 2.0 (http://www.wclim.org/) at a 1 × 1 km spatial resolution using ANUSPLIN-SPLINA software based on world-wide weather stations (Hijmans et al., 2005). These datasets included monthly maximum temperature, monthly minimum temperature and monthly mean precipitation during the 1970–2000 period.

2.1.2. Environmental conditions

There is evidence that the abundance of reproductive mosquitoes is closely related to vegetation canopy greenness and relative humidity (Nihei et al., 2014; Thu et al., 1998). Vegetation canopy can protect mosquito habitats from direct sunlight, and relative humidity reflects the necessary moisture content for mosquito survival (Messina et al., 2016). To represent the vegetation canopy development, NDVI datasets have been usually adopted (Linthicum et al., 1999; Zhang et al., 2013). In the present study, an advanced very high resolution radiometer (AVHRR) NDVI dataset developed by the Global Inventory Modeling and Mapping Studies (GIMMS) group (http://glcf.umd.edu/) was used with an 8 × 8 km spatial resolution and a 15 day interval temporal resolution. Based on these datasets acquired from 1982 to 2015, we calculated the mean annual NDVI layer and used it as one of the input data layers for machine learning models. The global mean annual relative humidity dataset obtained from the NASA Surface Meteorology and Solar Energy (https://eosweb.larc.nasa.gov/) was converted from a shapefile to a raster layer and was also used as input data.

2.1.2.1. Social factors. The relationship between humans and these mosquito vectors has been thoroughly addressed in previous research (Bhatt et al., 2013; Brown et al., 2011; Powell and Tabachnick, 2013). In addition to natural rain-filled containers, artificial containers are a comfortable habitat for mosquitoes and suitable for larval development (Morrison et al., 2004). In the present study, we used global urban region, population density and nighttime light layers to characterize the temporal and geographic variation in human habitat.

The global urban region dataset was derived from the Global Urban Heat Island dataset (Center for International Earth Science Information Network – CIESIN – Columbia University, 2016a,b), which can be downloaded from the NASA Socioeconomic Data and Application Center (SEDAC) (http://sedac.ciesin.columbia.edu/). From the SEDAC website, we also obtained the UN-Adjusted Population Density, version 4 (GPWv4) (Center for International Earth Science Information Network – CIESIN – Columbia University, 2016a,b; Balk et al., 2006). Based on the population density layers for the years 2000, 2005, 2010 and 2015, we developed a global mean population density layer at a 5 × 5 km spatial resolution that represented the relative spatial distribution of population density. The nighttime light dataset was obtained from the NOAA Earth Observation Group (https://ngdc.noaa.gov/). The mean annual nighttime light layer at a 5 × 5 km spatial resolution was calculated based on stable light layers of nighttime light satellite imagery from 1992 to 2013.

There is a well-established link between international travel and trade routes and mosquito expansion (Kyeongah et al., 2016; Gloriasoria et al., 2014). With the rapid increase in connectivity among human populations, the global spread of associated pathogens has also become increasingly serious. Global datasets of human movement are often unavailable free of charge. However, a 1 × 1 km spatial resolution global urban accessibility dataset is freely available from the European Commission Joint Research Centrewebsite (http://forobs.jrc.