Coalescing disparate data sources for the geospatial prediction of mosquito abundance, using Brazil as a motivating case study

Anwar Musah*, Ella Browning, Aisha Aldosery, Iuri Valerio Graciano Borges, Tercio Ambrizzi, Merve Tunali, Selma Basibuyuk, Orhan Yenigun, Giselle Machado Magalhaes Moreno, Clarisse Lins de Lima, Ana Clara Gomes da Silva, Wellington Pinheiro dos Santos, Tiago Massoni, Luiza Cintra Campos & Patty Kostkova

UCL Department of Geography (UK), UCL IRDR Centre for Digital Public Health in Emergencies (dPHE) (UK), UCL Centre for Biodiversity & Environmental Research (UK), University of Sao Paulo (Brazil), Bogazici University (Turkey), Federal University of Pernumbuco (Brazil), Federal University of Campina Grande (Brazil), UCL Department of Civil Environmental & Geomatic Engineering (UK)

INTRODUCTION

One of the barriers for performing geospatial surveillance for mosquito occupancy or infestation anywhere in the world is the paucity of primary entomologic survey data geolocated at a residential property-level matched to important risk factor information (e.g., anthropogenic, environmental and climate) that enables the spatial risk prediction of mosquito occupancy or infestation.

The reality is such data remains illusive in these low-resource settings, and where available, such high quality data that contains both individual and spatial characteristics to inform the geospatial descriptive and risk patterning of infestation remains rare.

This poster aims to:

- Point out to readers where these reliable open-source data can be acquired and how they can be used as risk factors for making spatial prediction for mosquito occupancy in general.
- Use Brazil as a case study to demonstrate how these datasets can be brought together to predict the presence of arboviruses through the use of Ecological Niche Modelling using the Maximum Entropy (MAXENT) algorithm
- To briefly summaries the benefits of using bespoke applications beyond these open-source online options so as to build a case to explain how it can be the new "gold-standard" approach for gathering primary entomologic survey data.

Readers should bear in mind that the scope of this poster was mainly limited to a Brazilian context due to the fact that it builds on an existing partnership with academics and stakeholders from environmental surveillance agencies in Recife (State of Pernambuco) and Campina Grande (State of Paraiba).

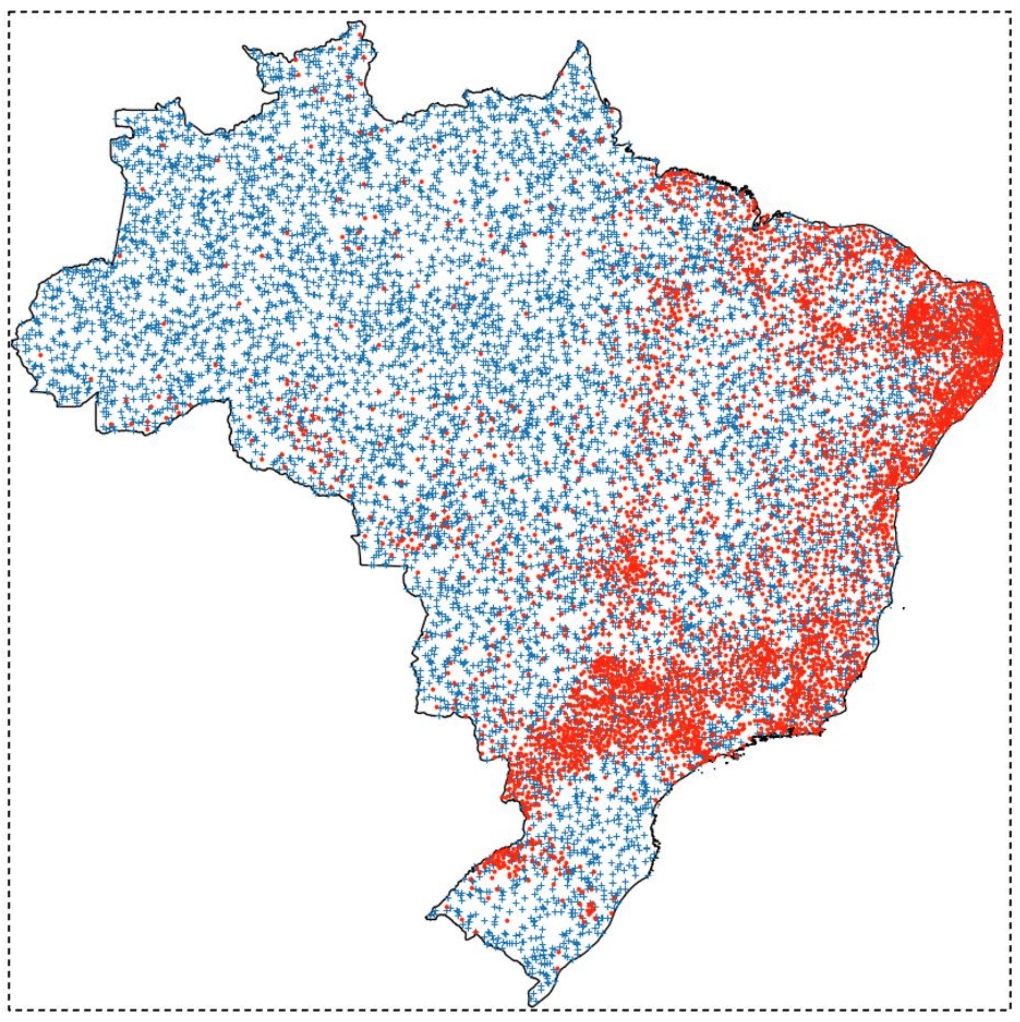
The analysis presented in this poster was also limited to a specific mosquito species under the Aedes family i.e., *Aedes aegypti* due to its endemicity status in Brazil.

DATA SOURCES

There are various sources of data which can be obtained online. We document its usage which can be used for causal inference and predictive analytics for mosquito occupancy. These included shapefiles for countries as well as raster grids for weather and environmental information. For raster data, we particularly highlighted the most reliable and updated sources available at a high spatial resolution:

- **GADM** (https://gadm.org/index.html) for obtaining spatial boundaries (National & Administrative etc.,) as shapefiles for countries in the world.
- **Vegetation (NDVI) cover** from USGS Earth Explorer (https//earthexplorer.usgs.gov/) using satellite (e.g., MODIS products such as MODI31A1 (500m) and MOD13Q1 (250m))
- **STRM Digital Elevation Database** for gridded land surface elevation above sea-level at 90m resolution (https://srtm.csi.cigar.org)
- **Worldpop.org** (https://www.worldpop.org/) for high resolution gridded data (100m) for anthropogenic indicators such as population density, natural lighting and urbanisation.
- **WorldClim.org** (https://www.worldclim.org) contains comprehensive high resolution weather (maximum and minimum temperature, and climate on a global scale.
- **OpenWeatherMap API** (https://openweathermap.org/api) an online meteorological API service for pulling weather information for stations.[1]

RESULTS



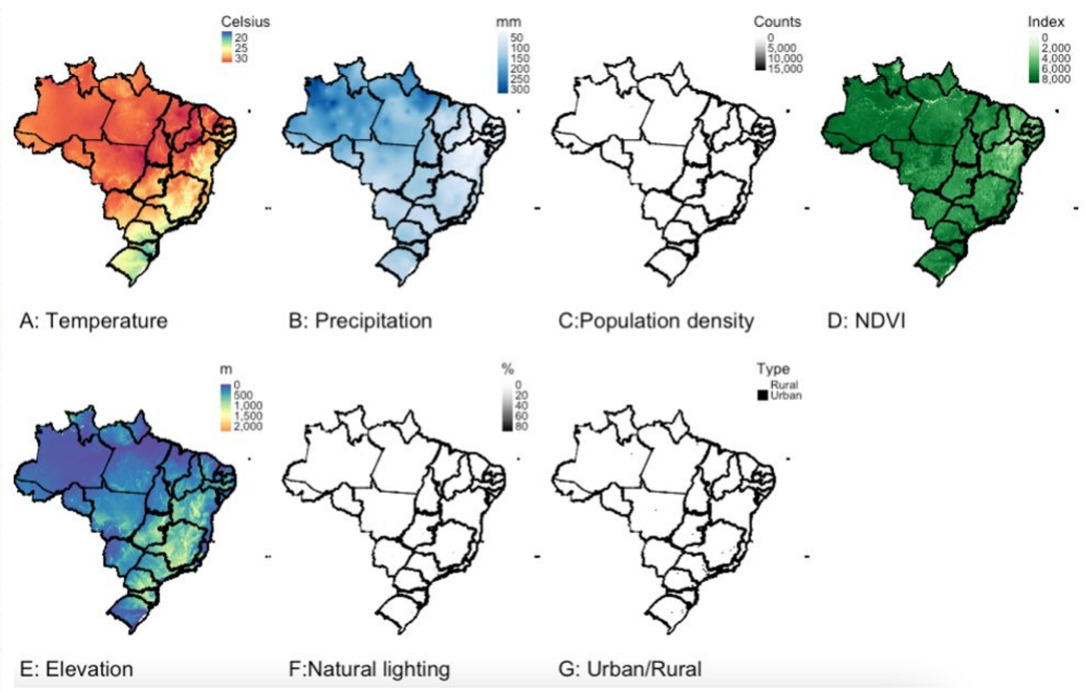


Figure 1: Right panel show point locations with known presence (red dots) with Aedes aegypti in 2013 in Brazil while the blue crosses correspond to pseudo-absence points; Left panel shows the following Brazilian covariate data measured for 2013 - (A) annual temperature; (B) annual precipitation; (C) Population density; (D) NDVI; (E) land surface elevation; (F) Natural lighting; and (G) Urban/rural classification

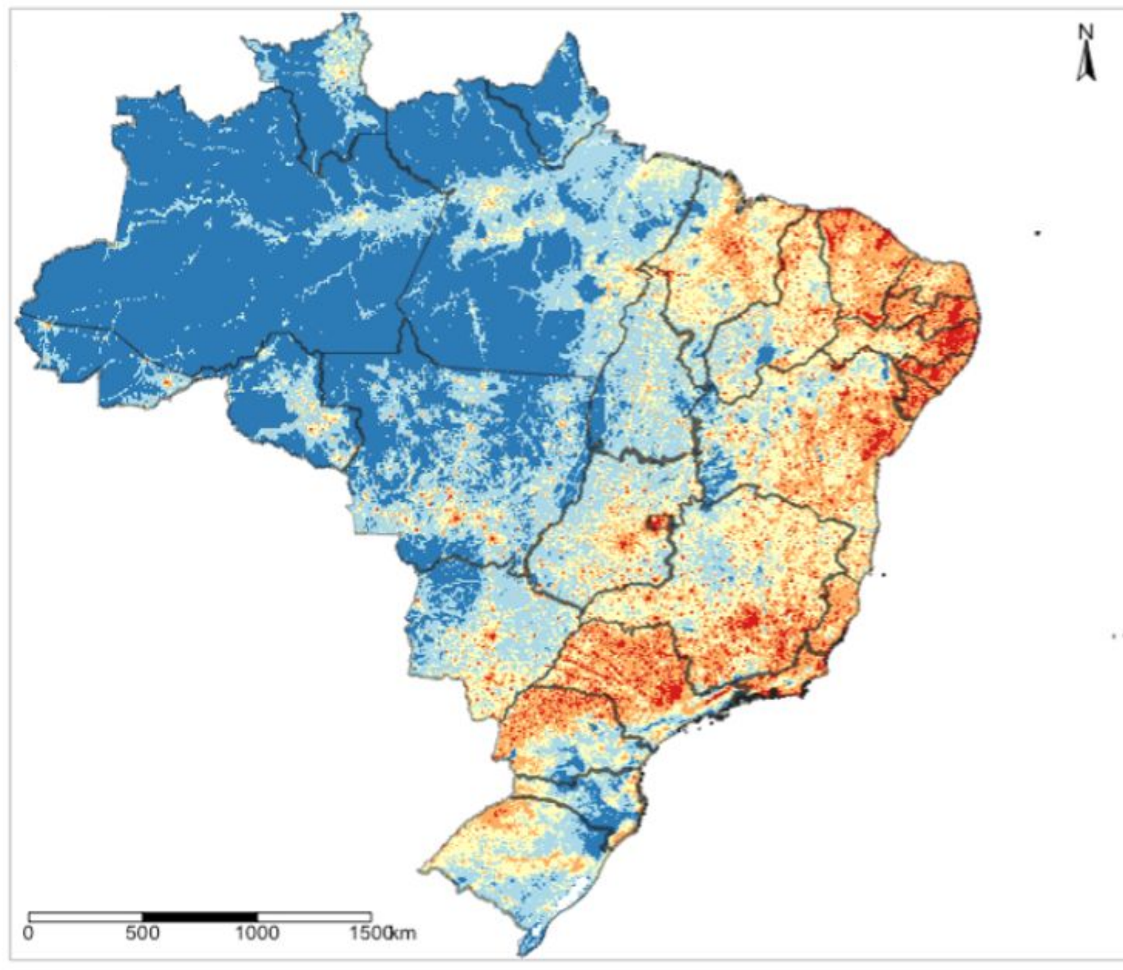


Figure 2: MAXENT modelling results show the predicted probability (as a heat map) of the mosquito occupancy i.e., Aedes aegypti species, in Brazil in 2013; The regions steering towards red colours indicates a very high probability of occurrence – which means that those are areas where mosquitos occupy and likely to thrive. Note that we performed a 4-fold cross-validation analysis yielding an overall AUC estimate of 0.837 (83.7%) which was obtained after averaging AUC-specific estimates for each 4-fold cross-validation i.e., 0.8435 (84.35%), 0.8403 (84.03%), 0.833 (83.3%) and 0.8331 (83.31%). This value is greater than 0.5 thus indicating the model's predictive reliability

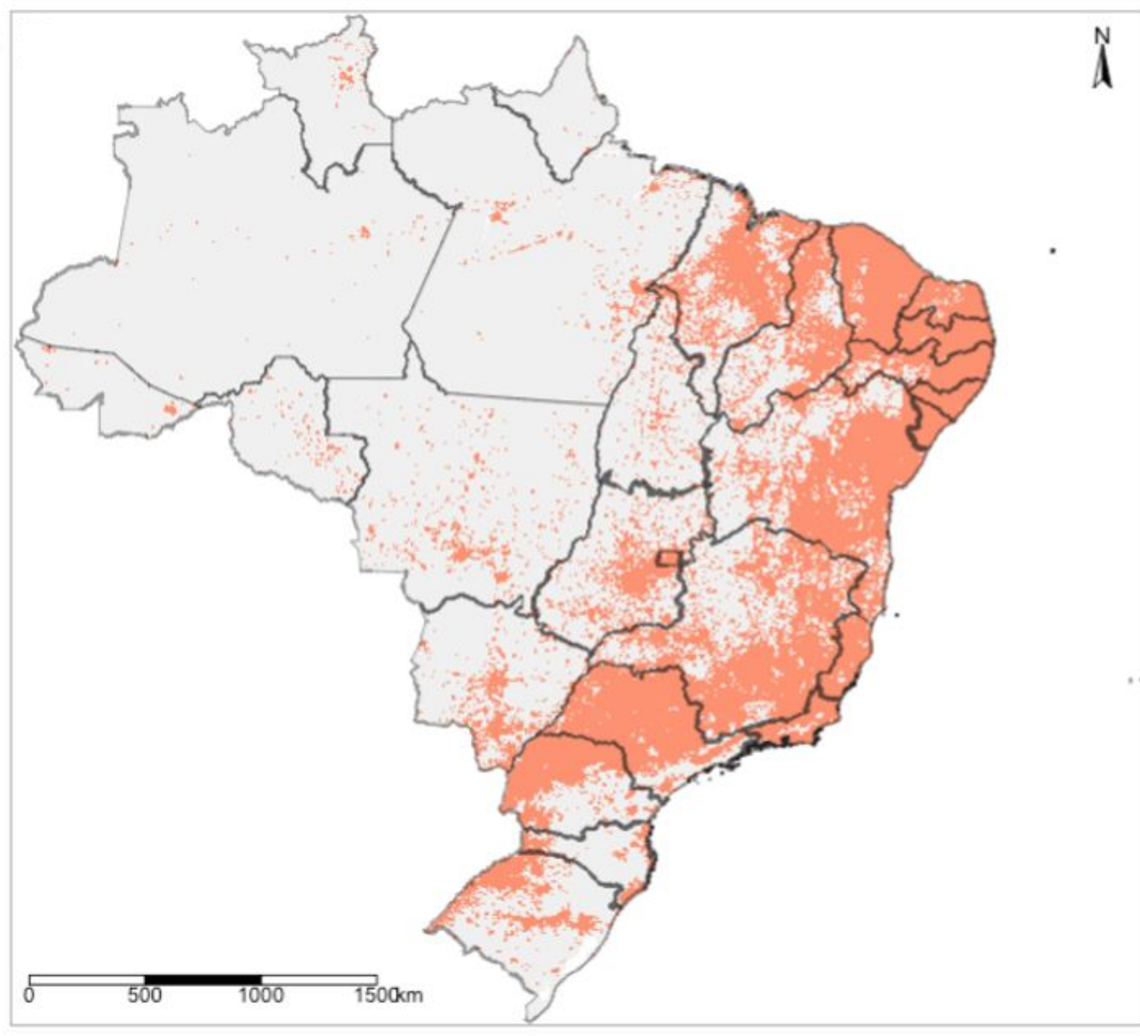


Figure 3: MAXENT modelling results show the suitability map based on the max TPR+TNR threshold (0.4953). This estimate was used to reclassify the output in figure 2 as "occupied" and "non-occupied" areas for where the *Aedes aegypti* species thrive in Brazil (based on 2013 data). These predictions were adjusted with seven different environmental covariates; here, we report the overall variable contributions: the population density, an anthropogenic indicator, has the highest contribution to the prediction was estimated at 75.75%; followed by natural lighting (10.29%); precipitation (6.79); NDVI (4.18%); temperature (2.55%); land surface elevation (0.421%) and urbanisation (negligible < 0.0001).

METHODOLOGY FOR CASE STUDY

Occurrence Data: The global compendium of the *Aedes* species[2] is an open source database which is accessible via Global Biodiversity Information Facility (GBIF) (https://www.gbif.org). A total of 4,410 occurrence locations in Brazil for the *Aedes Aegypti* was extracted from this database for the year 2013.

Study Design & Statistical Analysis: A country-scale ecological study design within a cross-sectional framework was used on 2013 data to retrospectively determine the following outcomes: 1.) the probability of the Aedes aegypti species being present at a location in Brazil; 2.) the likely areas which are environmentally suitable; and 3.) to report the set of restricted variables (i.e., temperature, precipitation, natural lighting, urbanisation, NDVI, population density and land surface elevation) that yields the highest contribution for mosquito prediction in a Brazilian context.

Ecological Niche Model such as the MAXENT was adopted. 4,410 location data for the *Aedes aegypti* were compiled and used as presence points. Background data i.e., twice the number of the presence points (i.e., 8,820) generated within the extent of the study area to serve as proxy locations for pseudo-absences for the Aedes aegypti. Bernoulli function modelled the probabilities in space.

SUMMARY

- We have identified a broad range of open-source data sources which can harnessed as risk factors for the spatial prediction of mosquito occupancy or infestation;
- We have demonstrate in a reproducible way on how they can be brought together and implemented using the MAXENT algorithm within a Brazilian context. This approach should be utilised within a data sparse context.
- However, the usage of novel bespoke technologies such as mobile applications[3] which should be taken as the better and viable method for collecting primary entomologic data so as to address the problems of data paucity, as well as avoid potential biases that's typically found studies using open source datasets – doing so will improve a studies' internal and external validity.
- This poster was written in mind to build capacity and awareness of the various data sources, and reproducibility of the methods, and thus its applicable to different mosquito species and other areas in the Global South of similar circumstances. All resources, including codes, guidance notes and processed data for reproducing the analysis shown in this poster can be found at this GitHub repository: https://github.com/UCLPG-MSC-SGDS/Data-Sources

Reference

1. Musah A, Dutra LMM, Aldosery A, Browning E, Ambrizzi T, Borges IVG, Tunali M, Başibüyük S, Yenigün O, Moreno GMM, et al. An Evaluation of the OpenWeatherMap API versus INMET Using Weather Data from Two Brazilian Cities: Recife and Campina Grande. Data (2022) 7:106. doi: 10.3390/data7080106
2. Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, et al. The global compendium of Aedes aegypti and Ae. albopictus occurrence. Sci Data (2015) 2:150035. doi: 10.1038/sdata.2015.35
3. Aldosery A, Musah A, Birjovanu G, Moreno G, Boscor A, Dutra L, Santos G, Nunes V, Oliveira R, Ambrizzi T, et al. MEWAR: Development of a Cross-Platform Mobile Application and Web Dashboard System for Real-Time Mosquito Surveillance in Northeast Brazil. Front Public Health (2021) 9:1623. doi: 10.3389/fpubh.2021.754072

For further details, please contact: a.musah@ucl.ac.uk

Published work (in-press)
Musah A., Browning E, Aldosery A, Borges IVG, Ambrizzi T *et al.,* (2023). Coalescing disparate data sources for the geospatial predication of mosquito abundance, using Brazil as a motivating case study. *Frontiers in Tropical Diseases.* DOI: 10.3389/fitd.2023.1039735