Title: Utilizing citizen science to model the distribution of *Aedes aegypti* ([Diptera]: [Culicidae])
in West Africa

Authors: Elizabeth A. Freeman^{a,b}, Elizabeth J. Carlton^b, Sara Paul^b, Samuel Dadzie^c, Andrea Buchwald^d

Affiliations:

a. Colorado School of Public Health, Colorado State University, Fort Collins, CO, 80523

b. Department of Environmental and Occupational Health, Colorado School of Public Health,
University of Colorado, Denver, CO 80045

c. Parasitology Department, Noguchi Memorial Institute for Medical Research, University of
Ghana, Legon, Ghana

d. Department of Biostatistics and Informatics, Colorado School of Public Health, University of
Colorado, Denver, CO 80045

Corresponding author: Elizabeth Freeman, efreeman.mph@gmail.com

Abstract: In a rapidly urbanizing region such as West Africa, *Aedes* mosquitoes pose an emerging threat of infectious disease that is compounded by limited vector surveillance. Citizen science has been proposed as a way to fill surveillance gaps by training local residents to collect and share information on disease vectors. Increasing citizen science efforts can begin to bridge the gaps in our current knowledge of *Aedes* distribution while engaging locals with mosquito

control and public health efforts. Understanding the distribution of disease vectors in West Africa can inform researchers and public health officials on where to conduct disease surveillance and focus public health interventions. We aimed to compare citizen science data to published literature observations of *Aedes* mosquitoes and to quantify how incorporating citizen science changes our understanding of *Aedes* mosquito distribution in West Africa. We utilized citizen science data collected through NASA's GLOBE Observer mobile phone application and data from a previously published literature review on *Aedes* mosquito distribution to examine the contribution of citizen science to understanding the distribution of *Ae. aegypti* in West Africa using Maximum Entropy modeling. Combining citizen science and literature-derived observations improved the fit of the model compared to models created by each data source alone, but did not alleviate location bias within the models, likely due to lack of widespread observations. Understanding *Ae. aegypti* distribution will require greater investment in *Aedes* mosquito surveillance in the region, and citizen science should be utilized as a tool in this mission to increase the reach of surveillance.

Keywords: *Aedes*, mosquito distribution, Maximum Entropy, West Africa, niche modeling, citizen science

Introduction

Arboviral diseases transmitted by *Aedes* mosquitoes are an emerging threat to the health of populations globally (Leta et al. 2018, Ryan et al. 2019), however there are data gaps regarding the distribution of key mosquito vectors in many regions, such as West Africa. West Africa has experienced multiple outbreaks of Zika, dengue and chikungunya over the past 15 years, yet the distribution of *Aedes* vectors is poorly defined (Buchwald et al. 2020). Vector surveillance can help public health officials to better predict where and when *Aedes*-borne diseases will emerge, as disease follows patterns of vector abundance (Carlson et al. 2016, Messina et al. 2016, Leta et al. 2018). Global environmental models of *Aedes aegypti* and *Ae. albopictus* have been created (Kraemer et al. 2015, Kamal et al. 2019, Ryan et al. 2019), but these lack the fine-scale spatial resolution desired by decision-makers and are subject to climate bias from global collinearity of environmental variables (Dormann et al. 2013). In many regions around the world where systematic arbovirus surveillance infrastructure is lacking, there is a need for enhanced local surveillance in order to improve epidemic preparedness and early warning systems. Given recurrent yellow fever outbreaks and the increasing public health burden due to dengue and chikungunya, West Africa is a priority region for strengthening the public health entomology capacities around *Aedes* surveillance and control. Improved *Aedes* surveillance and modelling will provide essential information for the formulation of evidence-based vector control strategies and also improve preventive actions to protect human populations from the diseases transmitted by *Aedes* mosquitoes.

Citizen science has been proposed as a way to bolster local *Aedes* mosquito surveillance (Walther and Kampen 2017, de Souza Silva et al. 2018, Eritja et al. 2019, Amos et al. 2020).

Citizen science is generally defined as members of a local or global community- many without formal training- participating in scientific activities, such as data collection and species surveillance (Kullenberg and Kasperowski 2016). Citizen science approaches have been used to enhance surveillance for multiple mosquito-borne diseases (Lwin et al. 2014, Hardy and Barrington 2017, Palmer et al. 2017, Richman et al. 2018). Utilizing citizen scientists can improve *Aedes* surveillance: residents may have preexisting knowledge of where mosquitoes reside in their community and can help scientists decide where to focus research (Mwangungulu et al. 2016). Participating in citizen science also benefits the community by increasing knowledge of mosquito-borne diseases and can spark behavioral changes to mitigate risk, such as engaging in household mosquito breeding site reduction (Toledo Romani et al. 2007, Stefopoulou et al. 2018, Asingizwe et al. 2020). However, the utility of citizen science programs can be doubtful due to challenges including recruitment and training of citizen scientists, as well as issues of data integrity (Johnson et al. 2018).

This study examines the value of incorporating data from NASA's Global Learning and Observations to Benefit the Environment (GLOBE) Observer Mosquito Habitat Mapper (MHM) application (<https://observer.globe.gov/toolkit/mosquito-habitat-mapper-toolkit>) (Amos et al. 2020) to estimate the distribution of *Aedes* mosquitoes in West Africa. The GLOBE Program has regional offices around the world and conducts outreach to students and teachers; providing them curriculum and protocols for conducting citizen science- however people who do not receive training can also use the app. In 2018, the GLOBE Zika Education and Prevention Project led an effort in 30 countries to bolster MHM observations (Aikpon 2019). The MHM was designed to address two goals; to categorize and map water sources used as mosquito breeding

grounds and to understand the distribution of mosquitoes. The app primarily captures data on the larval stage of mosquito development, but additionally has users input information about other life stages present at the habitat (Amos et al. 2020). GLOBE Observer walks the user through a series of questions to identify habitat types and the genus of mosquitoes found. This process requires the user to have access to mosquito larvae sampling equipment and magnifying lens for their phone camera (Muñoz et al. 2020). From 2016 to 2019, there were 320,000 observations submitted in the app by 38,000 users (Amos et al. 2020).

In this study, we aim to evaluate the utility of the citizen science data for enhancing the characterization of environmental suitability for *Aedes* mosquitoes across West Africa. To do this, we examine citizen science observations from GLOBE MHM and data from a literature review to estimate distribution of *Ae. aegypti* in West Africa. We use Maximum Entropy modeling to evaluate the extent to which citizen science data can improve estimates of the spatial distribution of the primary *Aedes*-borne disease vector.

Methods

We modeled distribution of *Ae. aegypti* mosquitoes in West Africa using two different data sources: i) observations extracted from GLOBE MHM, and ii) a compilation of the published literature containing scientific reports on *Aedes* surveillance from Buchwald et al. 2020 (Buchwald et al. 2020). There are multiple methods to model species distribution (Vezzani et al. 2005, Kearney et al. 2009, Richman et al. 2018), we chose Maximum Entropy modeling (Maxent) because it is well suited to the small numbers of observations, and presence-only data characterizing our dataset. Maxent modeling combines species observation data with key

climate predictors of species presence to predict the probability of species presence (Phillips and Dudík 2008). We compared the ranges for *Ae. aegypti* predicted by each data source alone and in combination to evaluate whether inclusion of citizen science data can improve estimates of species distribution in a region lacking data.

Study Area

We defined the study area of West Africa as between the latitude of 20 and -30 and between the longitude of 20 and -10. This area encompasses Burkina Faso, Benin, Cameroon, Cape Verde, Chad, Côte d'Ivoire, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Sao Tome and Principe, Sierra Leone, and Togo. These boundaries were chosen because they contain all *Aedes* presence points in both GLOBE MHM and the literature data used in the study.

Mosquito Presence data

Data from GLOBE MHM were obtained via the GLOBE Observer's Advanced Data Access Tool (<https://datasearch.globe.gov/>). We used data from the Mosquito Habitat Mapper and Mosquito Larvae protocols and searched by site name for all countries in our study area. Analysis was done in Rstudio (RCoreTeam 2021). Data were accessed on XXXX December 8th, 2020 and observations ranged from July 5th, 2017 to November 22nd, 2020. We selected observations that marked "true" for mosquito eggs, pupae, or adults or a non-zero larvae count for inclusion in further analysis. Most observations in the GLOBE dataset

lacked genus and species data. Observations were filtered to include only unique locations through the removal of duplicates.

Mosquito observations from the published literature were collected as part of a previous study by Buchwald et al. covering the time period from 1999 through October 2018 (Buchwald et al. 2020). For this study, we restricted our main analysis to *Aedes* observations published from 2009-2018 to overlap with the time period of GLOBE MHM observations. We include a sensitivity analysis to examine the impact of including *Aedes* observations from the literature from 1999-2008.

Climate Data

We obtained climate data on 19 bioclimatic variables in raster form from WorldClim version 2.1 (worldclim.org) which provides averages of data from the years 1970 through 2000. We began our analysis with all 19 variables at 2.5 minute resolution. Raster data was restricted to our defined study area. WorldClim variables describe aspects of either temperature or precipitation- both of which influence mosquito ecology- however, there is no consensus on which variables most effect *Ae. aegypti* distribution. To choose variables to model, we analyzed the variable inflation factor of the variables to assess collinearity (De Marco and Nóbrega 2018). We used the R package usdm (Toxopeus 2014) and selected variables for inclusion in the maximum entropy models which had VIF < 5 (Supplement Table 1.). We chose not to include land cover and population data as climatic variables are assumed to be the main drivers at the continental scale (the scope of this study)(Pearson and Dawson 2003, Bogawski et al. 2019).

Additionally, urbanization in West Africa has progressed rapidly during the study period of 2008 to 2020 and may be difficult to capture in a single time-point.

Maximum Entropy Modeling

We modeled *Ae. aegypti* distribution using Maximum Entropy (Maxent) version 3.4.4, an approach that creates habitat suitability models by fitting species presence data to climatic variables in raster grid cells (Phillips and Dudík 2008). The software utilizes the climatic pattern in presence areas to predict other locations with suitable habitat for the study species (Phillips and Dudík 2008). To adjust for bias in observation data, a bias file was created using the closely related species method (Phillips et al. 2009). All *Aedes* observations including *Ae. aegypti*, *Ae. albopictus*, *Ae. incerta*, and unidentified *Aedes* from GLOBE and *Ae. aegypti* and *Ae. albopictus* observations from the literature were projected onto a raster and analyzed using two-dimensional kernel density estimation from the R package MASS to use as the bias layer (Ripley 2002, Hijmans 2020).

Three models were created: 1) using GLOBE MHM data, 2) using literature data from 2009 to 2018, and 3) using GLOBE MHM data combined with literature data from 2009 to 2018. The first two models were used to compare estimated species distributions between models using citizen science data and models using recent data from the literature. The third model was used to determine the change in estimated distribution when adding citizen science data to the literature-derived data. Three additional models were created as a sensitivity analysis, 4) using literature data from 1999 to 2008, 5) using all literature data from 1999 to 2018, and 6) using GLOBE MHM data combined with all literature data (1999 to 2018). We used *Ae. aegypti*

observations from 1999-2008 literature as our sensitivity analysis because the time frame in that dataset overlaps with the WorldClim bioclimate variables whereas the data in our main three models do not. The climatic data used was the same for every model. The regularization multiplier (RM) parameter was set to 1.5 to combat potential overfitting issues due to our small datasets. Each model was replicated 20 times and the mean of the replications was examined. All other parameters within Maxent were kept at their default values. Cross-validation compared the runs of the models. Cross-validation used all of the data for validation of models by creating smaller subsets of data as test data for the model that changes each run and using the rest as training data. Models were evaluated based on the area under the receiver operating curve (AUC) for the test data and the prevalence. The AUC is used to estimate how well the model fits the presence and climate data across our study area. The prevalence is the average estimated probability of species present in that area over background sites across all twenty models. Jackknife analysis was run to assess the variable importance of climatic predictor variables for each of the models.

Results

Mosquito Presence Data

Figure 1. All observations from GLOBE MHM downloaded from the GLOBE Observer Advanced Data Access Tool on November 22nd, 2020.

There were 4,900 observations downloaded from GLOBE Observer’s Advanced Data Access Tool. Of those observations, 3,607 had either larvae, pupae, eggs, or adult mosquitoes present and 307 of those were identified as the genus *Aedes*. Within the *Aedes* identifications, 69 had an accompanying species. After filtering for unique location points, the data contained 13 unique sites for *Ae. albopictus*, 12 unique sites for *Ae. incerta*, and 24 unique sites for *Ae. aegypti*. *Ae. aegypti* was the only species with sufficient observations to create Maxent models. *Ae. aegypti* were recorded in four countries with most observations (87.3%) located in Senegal (Figure 1).

Senegal was a major contributor of data in the literature as well. Data from literature published from 2009-2018 contained *Aedes* observations from Cameroon, Cape Verde, Côte d’Ivoire, Gabon, Ghana, Mali, Nigeria, Senegal, and Sao Tome (Figure 2, Table 1). Nigeria and Senegal contained the most *Ae. aegypti* observations with 17 and 16, respectively (Table 1). Literature data published from 1999-2008 contained *Aedes* observations from Burkina Faso, Cameroon, Côte d’Ivoire, Gabon, Gambia, Ghana, Guinea, Nigeria, Senegal, and Mali (Figure 2).

Figure 2. *Aedes* mosquito observations from literature and GLOBE MHM. Top panel: All *Aedes* observations by source dataset. Bottom panel: *Ae. aegypti* observations by source dataset.

Table 1. Distribution of *Aedes* mosquito observations in West Africa from literature review and GLOBE Mosquito Habitat Mapper App by species, country, and decade of publication. GLOBE MHM observations were collected 2017 – 2020.

Aedes

	<i>Aedes aegypti</i>		<i>Aedes albopictus</i>		<i>incerta</i>	Overall	
Data Source	09-18	GLOBE	09-18	GLOBE	GLOBE	09-18	GLOBE
	Literature	MHM	Literature	MHM	MHM	Literature	MHM
N Observations	55	32	22	17	20	77	307
Country							
Cameroon	2 (3.6%)	0 (0%)	4 (18.2%)	0 (0%)	0 (0%)	6 (7.8%)	0 (0%)
Cape Verde	2 (3.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.6%)	0 (0%)
Côte d'Ivoire	7 (12.7%)	0 (0%)	2 (9.1%)	0 (0%)	0 (0%)	9 (11.7%)	0 (0%)
Gabon	5 (9.1%)	0 (0%)	3 (13.6%)	0 (0%)	0 (0%)	8 (10.4%)	0 (0%)
Ghana	4 (7.3%)	1 (3.1%)	1 (4.5%)	0 (0%)	0 (0%)	5 (6.5%)	1 (0.3%)
Mali	2 (3.6%)	0 (0%)	3 (13.6%)	0 (0%)	0 (0%)	5 (6.5%)	0 (0%)
Nigeria	17 (30.9%)	0 (0%)	7 (31.8%)	1 (5.9%)	0 (0%)	24 (31.2%)	1 (0.3%)
Senegal	16 (29.1%)	21 (65.6%)	0 (0%)	12 (70.6%)	15 (75.0%)	16 (20.8%)	268 (87.3%)
Sao Tome	0 (0%)	0 (0%)	2 (9.1%)	0 (0%)	0 (0%)	2 (2.6%)	0 (0%)
Benin	0 (0%)	3 (9.4%)	0 (0%)	2 (11.8%)	4 (20.0%)	0 (0%)	25 (8.1%)
Burkina Faso	0 (0%)	7 (21.9%)	0 (0%)	2 (11.8%)	0 (0%)	0 (0%)	10 (3.3%)
Guinea	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (5.0%)	0 (0%)	1 (0.3%)
Togo	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.3%)

Climate Data

Of the 19 WorldClim bioclimate variables, 8 variables were included in our Maxent models (Figure 3, Supplemental Table 1). Jackknife analysis of climatic predictor variables shows the change in AUC with keeping or removing predictive climatic variables in the models, giving us information about the relative importance that each climatic variable plays in each of the models. The relative importance for each climatic variable varied across the models. In Model 1

(Figure 3A), mean temperature of the wettest quarter (Bio8) lead to the model with the highest AUC when used alone as a predictive variable, whereas precipitation of the coldest quarter (Bio19) lead to the largest decrease in AUC when removed from the full model. In Model 2 (Figure 3B), mean temperature of the coldest quarter (Bio11) alone gave the model with the highest AUC and when removed lead to the largest decrease in predictive accuracy. In Model 3 (Figure 3C), Bio8 led to the highest AUC when used alone and mean diurnal range (Bio2) led to the greatest decrease in AUC when removed.

Figure 3. A: Jackknife plot of climatic variables for Model 1; Maximum Entropy output using GLOBE MHM data. **B:** Jackknife plot of climatic variables for Model 2; Maximum Entropy output using 2009-2018 literature data. **C:** Jackknife plot of climatic variables for Model 3; Maximum Entropy output using a combination of GLOBE MHM data and 2009-2018 literature data.

Maximum Entropy Modeling

The suitable habitat predicted by each model varied in prevalence (Table 2). Model 1, using data from GLOBE MHM estimated average prevalence relatively lower than models that used literature data, with predicted high prevalence of *Aedes aegypti* in Senegal and scattered presence across southern Mali, Burkina Faso, and Nigeria (Figure 4A). Model 2, using the 2009-2018 literature data (Figure 4B) predicted widespread suitable habitat for *Ae. aegypti*, with high presence probability estimated for an area including the coast of Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, and Gabon. Areas across Senegal and Burkina Faso were also predicted to have high probability of *Ae. aegypti* presence. Model 3 (Figure 4C) used

data combined from the GLOBE MHM and the 2009-2018 literature data and predicted a lower average prevalence than Model 2, but the test AUC increased compared to Model 2. Using the combined dataset, Maxent predicted more *Ae. aegypti* habitat suitability than Model 1, covering regions in Senegal, Burkina Faso, Nigeria, and along the coasts of Côte d'Ivoire, Ghana and Togo.

Figure 4. Estimated *Ae. Aegypti* habitat suitability in West Africa using Maximum Entropy Modeling. A) Models fit to *Ae. Aegypti* observations from GLOBE MHM; B) Models fit to data from the literature published from 2009 to 2018; C) Models fit to data from GLOBE MHM and the 2009-2018 literature.

Table 2. Model output for prevalence of *Ae. aegypti*, area under the curve (AUC) of the ROC, and the standard deviation of the AUC for all combinations of data ran in Maxent. Prevalence describes the proportion of the area modeled estimated to be suitable for *Ae. aegypti* and *Ae. spp.* AUC and AUC standard deviation are reported for the test data in each model.

Model	Data Source	N Observations	Prevalence	Test AUC	Test AUC SD
1	GLOBE MHM	24	0.1901	0.9417	0.056
2	09-18 Literature	55	0.3713	0.8157	0.167
3	GLOBE MHM and 09-18 Literature	79	0.293	0.8669	0.109
4	99-08 Literature	179	0.1693	0.7631	0.078

5	99-18 Literature	234	0.2374	0.7124	0.0838
6	GLOBE MHM and 99-18 Literature	258	0.1999	0.7601	0.068

Sensitivity Analyses

The data from literature published 1999-2008 was the largest individual data set (n=179), however Model 4 both predicted lower prevalence of *Ae. aegypti* and had a lower AUC than any of Models 1-3. Model 5 (Supplement figure 1B), using all literature data published from 1999 to 2018, similarly estimated lower prevalence than from the model including only recent data and had decreased fit to the data. Model 6 (Supplemental Figure 1C), combining all *Ae. aegypti* data from the literature and from GLOBE MHM, similarly estimated decreased prevalence compared to Model 2 and had decreased fit to the data (Table 2).

Discussion

We compared predictions of *Ae. Aegypti* mosquito distribution in West Africa from models using literature-derived and citizen science-derived data. Addition of citizen science data improved the model fit to the data, suggesting an ongoing role for citizen science in improving our understanding of *Aedes* mosquito distribution in the region. However, all of our models were likely subject to location bias, highlighting the need for increased surveillance of *Aedes* species mosquitoes throughout West Africa. Despite the lack of observations and potential location bias in the data, all models including observations from the literature predicted widespread distribution of *Ae. aegypti* in the region.

Previous global mapping efforts have estimated the distribution of *Ae. aegypti* in West Africa using a variety of methods (Kraemer et al. 2015, Gloria-Soria et al. 2016, Kamal et al. 2019). All our models estimated lower prevalence of *Ae. aegypti* in West Africa compared to previous global models of *Ae. aegypti*. Few presence points in our study may have contributed to the lower prevalence values that we found (van Proosdij et al. 2016). However, all previous models have been at a wider spatial scale (Kraemer et al. 2015, Gloria-Soria et al. 2016, Kamal et al. 2019, Ryan et al. 2019). To our knowledge, this study is the first model of *Ae. aegypti* distribution focusing solely on the region of West Africa. Aedes behavior and habitat selection may vary regionally, due to subspecies variation, local-level competition, or other factors- potentially biasing estimates from global models.

Citizen science can be used to increase *Aedes* surveillance data, but like any surveillance method, comes with limitations including recruitment, adherence, and concerns of data quality. Recruitment of citizen scientists to participate in surveillance projects is a challenge (Basso et al. 2015, Lwin et al. 2017), with participation unequal across socio-economic groups (Hobbs and White 2012). Lack of adherence to protocols may occur among citizen scientists (Cohnstaedt et al. 2016, Johnson et al. 2018), however similar problems can occur in traditional mosquito surveillance. There are known data biases in GLOBE MHM; participants only identified mosquito genus 14% of the time and 7.8% of observations from 2016-2019 were flagged as problematic previously (Amos et al. 2020). Some species may have been misclassified within the citizen science data used. Some data put into GLOBE MHM has been verified by scientists at organizations partnering with GLOBE, but much of the identification of mosquito genus and species in the app relies on morphological identifications by untrained citizen

scientists. GLOBE MHM attempts to mitigate this bias by walking the user through a series of larval identification. The technique presented requires the user to have a magnifying lens attachment for their smartphone camera, which not all users will have or use properly. Misidentification of species is not accounted for within Maxent and this may have influenced model estimates (Lozier et al. 2009).

Although there was likely bias created from GLOBE MHM observations in general and particularly due to the overweighting of observations from Senegal, the long-standing relationships between schools and the GLOBE program is thought to improve the quality of the data. Of the observations included in this study, 81% (N = 249) were recorded by lycee Thumakha, GLOBE's largest school partner in Senegal. They partnered with GLOBE in 2015 and have 100 students participating in the [program](#). Future research should focus on investigating the accuracy of GLOBE Observer data in Senegal alone where data may be sufficient to examine mosquito distribution on a local-level scale. Building similar partnerships throughout West Africa will improve data from citizen science projects, as well as expand local education and knowledge of the risks and prevention methods to control vector-borne diseases.

Multiple methods have been utilized for engaging citizen scientists in mosquito surveillance. Some have focused on collection and submittal of a biological specimen (Cohnstaedt et al. 2016, Walther and Kampen 2017), which are then identified by trained entomologists. Other projects have utilized citizen scientists to quantify mosquitoes caught in surveillance equipment to aid researchers in gathering data (Bazin and Williams 2018, Johnson et al. 2018). Recently, mobile applications have emerged as a popular tool for citizen science-

based mosquito surveillance (Lwin et al. 2017, Palmer et al. 2017, de Souza Silva et al. 2018) as applications can increase the accessibility of mosquito surveillance due to widespread smart phone use. Such applications have demonstrated mixed success as a surveillance tool (Lwin et al. 2017, Eritja et al. 2019). Use of the Mosquito Alert application (Palmer et al. 2017) led to the first detection of *Aedes japonicus* in Spain (Eritja et al. 2019). Others have been less successful due to inherent design issues (Lwin et al. 2014, Lwin et al. 2016, Lwin et al. 2017).

We recently introduced GLOBE to Ghanaian students at Junior High Schools near Accra. While there were some difficulties with implementation, including a lack of compatibility with widely-used varieties of tablets and phones available in low-resource settings, at the end of the study, the teachers and students were enthusiastic about the program and interested in seeing it continues. Following the training and data collection, students understood their role in mosquito control as a way to reduce mosquito-borne diseases, particularly *Aedes*-borne diseases. Teachers valued the hands-on instruction in the app and expressed enthusiasm for integrating education on mosquitoes and arboviruses in the schools. The ability for students to do “field work” in their own back yards was engaging and highlighted the potential for the GLOBE program to expand and contribute to mosquito surveillance in the future.

The estimates of *Ae. aegypti* mosquito distribution in West Africa produced herein are subject to a number of important limitations: Observations used in these models were locationally biased due to clustering in the datasets used. Location bias may have biased estimates of environmentally suitable habitats as well as have led to over-fitting of models with few presence points. While we attempted to adjust for location bias with both bias file offsets and tweaking the RM parameter, this may not have been sufficient. We additionally were

unable to adjust for land cover due to fluctuations in land cover over the (long) study period. Future estimates may benefit from using predicted land cover data to account for rapid urbanization. Additionally, while WorldClim is used widely in Maxent models (Lubinda et al. 2019, Tiffin et al. 2019, Ibáñez-Justicia et al. 2020), the dates of the WorldClim data (1970-2000) do not overlap with the dates of the *Aedes* presence data used in Models 1-3. Climate change has been noted on a short time scale- 2000 to 2009- in West Africa (Barry et al. 2018) and in such a dynamic region, both mosquito distribution and local climate patterns are likely in constant flux, requiring real-time data to detect changes. Finally, we note that growing evidence suggests *Ae. albopictus* also plays an important role in arboviral infections in the region, however there was not sufficient data to examine *Ae. albopictus* distribution

Conclusions

There is a need for expanded *Aedes* mosquito surveillance throughout West Africa. Despite some heterogeneity in implementation and though quality of data collected by citizen scientists is a concern, these issues are outweighed by the potential benefits of increased data in mosquito-borne disease research. The models presented in this manuscript suggest data from these mobile applications can be useful in supplementing our understanding of vector distribution. While both traditional mosquito surveillance and citizen science have pros and cons, using them in combination can improve understanding of mosquito distributions and *Aedes*-borne disease risk.

Acknowledgements

We would like to thank all schools and citizen scientists who contributed mosquito observations to the GLOBE MHM. This work was funded by a grant through the GLOBE Engaging Citizens in the Forecasting and Observation of Mosquito Threats Program from the United States Department of State and the University Corporation for Atmospheric Research (SLMAQM17CA2074). The GLOBE Implementation Office (GIO) managed the activities in the citizen and mosquito program. The opinions, findings and conclusions stated herein are those of the authors and do not necessarily reflect those of GIO, UCAR, or the US State Department.

Data Availability Statement

All data and code used in this manuscript are available at https://github.com/eafreeman/aedes_globe.

References

- Aïkpon, R., Klotoé, J.R., Dramane, G., Brettenny, M., Lawani, Y. 2019.** Larval breeding characteristics and distribution of *Aedes* mosquito species in the economic capital of Benin: A public health concern. *International Journal of Entomology Research* 4: 57-62.
- Amos, H. M., M. J. Starke, T. M. Rogerson, M. Colón Robles, T. Andersen, R. Boger, B. A. Campbell, R. D. Low, P. Nelson, D. Overoye, J. E. Taylor, K. L. Weaver, T. M. Ferrell, H. Kohl, and T. G. Schwerin. 2020.** GLOBE Observer Data: 2016-2019. *Earth Space Sci* 7: e2020EA001175.
- Asingizwe, D., P. M. Poortvliet, A. J. H. van Vliet, C. J. M. Koenraadt, C. M. Ingabire, L. Mutesa, and C. Leeuwis. 2020.** What do people benefit from a citizen science programme? Evidence from a Rwandan citizen science programme on malaria control. *Malar J* 19: 283.
- Barry, A., J. Caesar, A. Klein Tank, E. Aguilar, C. McSweeney, A. M. Cyrille, M. Nikiema, K. Narcisse, F. Sima, and G. Stafford. 2018.** West Africa climate extremes and climate change indices. *International Journal of Climatology* 38: e921-e938.
- Basso, C., E. García da Rosa, S. Romero, C. González, R. Lairihoy, I. Roche, R. M. Caffera, R. da Rosa, M. Calfani, E. Alfonso-Sierra, M. Petzold, A. Kroeger, and J. Sommerfeld. 2015.** Improved dengue fever prevention through innovative intervention methods in the city of Salto, Uruguay. *Trans R Soc Trop Med Hyg* 109: 134-142.
- Bazin, M., and C. R. Williams. 2018.** Mosquito traps for urban surveillance: collection efficacy and potential for use by citizen scientists. *J Vector Ecol* 43: 98-103.
- Bogawski, P., T. Damen, M. M. Nowak, K. Pędziwiatr, P. Wilkin, G. Mwachala, J. Pierzchalska, and J. Wiland-Szymańska. 2019.** Current and future potential distributions of three *Dracaena* Vand. ex L. species under two contrasting climate change scenarios in Africa. *Ecol Evol* 9: 6833-6848.
- Buchwald, A. G., M. H. Hayden, S. K. Dadzie, S. H. Paull, and E. J. Carlton. 2020.** *Aedes*-borne disease outbreaks in West Africa: A call for enhanced surveillance. *Acta Trop* 209: 105468.
- Carlson, C. J., E. R. Dougherty, and W. Getz. 2016.** An Ecological Assessment of the Pandemic Threat of Zika Virus. *PLOS Neglected Tropical Diseases* 10: e0004968.
- Cohnstaedt, L. W., D. Snyder, E. Maki, and S. Schafer. 2016.** Crowdsourcing methodology: establishing the Cervid Disease Network and the North American Mosquito Project. *Vet Ital* 52: 195-200.
- De Marco, P. J., and C. C. Nóbrega. 2018.** Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS One* 13: e0202403.
- de Souza Silva, G. C., L. M. Peltonen, L. Pruinelli, H. Yoshikazu Shishido, and G. Jacklin Eler. 2018.** Technologies to Combat *Aedes* Mosquitoes: A Model Based on Smart City. *Stud Health Technol Inform* 250: 129-133.
- Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, and P. J. Leitao. 2013.** Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27-46.

Eritja, R., I. Ruiz-Arrondo, S. Delacour-Estrella, F. Schaffner, J. Álvarez-Chachero, M. Bengoa, M. Puig, R. Melero-Alcibar, A. Oltra, and F. Bartumeus. 2019. First detection of *Aedes japonicus* in Spain: an unexpected finding triggered by citizen science. *Parasit Vectors* 12: 53.

Gloria-Soria, A., D. Ayala, A. Bheecarry, O. Calderon-Arguedas, D. D. Chadee, M. Chiappero, M. Coetzee, K. B. Elahee, I. Fernandez-Salas, and H. A. Kamal. 2016. Global genetic diversity of *Aedes aegypti*. *Molecular ecology* 25: 5377-5395.

Hardy, M. C., and D. J. Barrington. 2017. A Transdisciplinary Approach to Managing Emerging and Resurging Mosquito-Borne Diseases in the Western Pacific Region. *Trop Med Infect Dis* 2.

Hijmans, R. J. 2020. raster: Geographic Data Analysis and Modeling computer program, version By Hijmans, R. J.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Hobbs, S. J., and P. C. L. White. 2012. Motivations and barriers in relation to community participation in biodiversity recording. *Journal for Nature Conservation* 20: 364-373.

Ibáñez-Justicia, A., J. D. Alcaraz-Hernández, R. van Lammeren, C. J. M. Koenraadt, A. Bergsma, L. Delucchi, A. Rizzoli, and W. Takken. 2020. Habitat suitability modelling to assess the introductions of *Aedes albopictus* (Diptera: Culicidae) in the Netherlands. *Parasites & Vectors* 13: 217.

Johnson, B. J., D. Brosch, A. Christiansen, E. Wells, M. Wells, A. F. Bhandoola, A. Milne, S. Garrison, and D. M. Fonseca. 2018. Neighbors help neighbors control urban mosquitoes. *Sci Rep* 8: 15797.

Kamal, M., M. A. Kenawy, M. H. Rady, A. S. Khaled, and A. M. Samy. 2019. Mapping the global potential distributions of two arboviral vectors *Aedes aegypti* and *Ae. albopictus* under changing climate. *PLOS ONE* 13: e0210122.

Kearney, M., W. P. Porter, C. Williams, S. Ritchie, and A. A. Hoffmann. 2009. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Functional Ecology* 23: 528-538.

Kraemer, M. U., M. E. Sinka, K. A. Duda, A. Q. Mylne, F. M. Shearer, C. M. Barker, C. G. Moore, R. G. Carvalho, G. E. Coelho, and W. Van Bortel. 2015. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *elife* 4: e08347.

Kullenberg, C., and D. Kasperowski. 2016. What Is Citizen Science?--A Scientometric Meta-Analysis. *PLoS One* 11: e0147152.

Leta, S., T. J. Beyene, E. M. De Clercq, K. Amenu, M. U. Kraemer, and C. W. Revie. 2018. Global risk mapping for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*. *International Journal of Infectious Diseases* 67: 25-35.

Lozier, J., P. Aniello, and M. Hickerson. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. Wiley Online Library.

Lubinda, J., J. A. Treviño C, M. R. Walsh, A. J. Moore, A. A. Hanafi-Bojd, S. Akgun, B. Zhao, A. S. Barro, M. M. Begum, H. Jamal, A. Angulo-Molina, and U. Haque. 2019. Environmental

suitability for *Aedes aegypti* and *Aedes albopictus* and the spatial distribution of major arboviral infections in Mexico. *Parasite Epidemiology and Control* 6: e00116.

Lwin, M. O., S. Vijaykumar, G. Lim, O. N. Fernando, V. S. Rathnayake, and S. Foo. 2016. Baseline Evaluation of a Participatory Mobile Health Intervention for Dengue Prevention in Sri Lanka. *Health Educ Behav* 43: 471-479.

Lwin, M. O., K. Jayasundar, A. Sheldenkar, R. Wijayamuni, P. Wimalaratne, K. C. Ernst, and S. Foo. 2017. Lessons From the Implementation of Mo-Buzz, a Mobile Pandemic Surveillance System for Dengue. *JMIR Public Health Surveill* 3: e65.

Lwin, M. O., S. Vijaykumar, O. N. Fernando, S. A. Cheong, V. S. Rathnayake, G. Lim, Y. L. Theng, S. Chaudhuri, and S. Foo. 2014. A 21st century approach to tackling dengue: Crowdsourced surveillance, predictive mapping and tailored communication. *Acta Trop* 130: 100-107.

Messina, J. P., M. U. G. Kraemer, O. J. Brady, D. M. Pigott, F. M. Shearer, D. J. Weiss, N. Golding, C. W. Ruktanonchai, P. W. Gething, E. Cohn, J. S. Brownstein, K. Khan, A. J. Tatem, T. Jaenisch, C. J. L. Murray, F. Marinho, T. W. Scott, and S. I. Hay. 2016. Mapping global environmental suitability for Zika virus. *eLife* 5: e15272.

Muñoz, J. P., R. Boger, S. Dexter, and R. Low. 2020. Mosquitoes and Public Health: Improving Data Validation of Citizen Science Contributions Using Computer Vision, pp. 469-493. In N. Wickramasinghe and F. Bodendorf (eds.), *Delivering Superior Health and Wellness Management with IoT and Analytics*. Springer International Publishing, Cham.

Mwangungulu, S. P., R. D. Sumaye, A. J. Limwagu, D. J. Siria, E. W. Kaindoa, and F. O. Okumu. 2016. Crowdsourcing Vector Surveillance: Using Community Knowledge and Experiences to Predict Densities and Distribution of Outdoor-Biting Mosquitoes in Rural Tanzania. *PLoS One* 11: e0156388.

Palmer, J., A. Oltra, F. Collantes, J. Delgado, J. Lucientes, S. Delacour, M. Bengoa, R. Eritja, and F. Bartumeus. 2017. Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes. *Nature Communications* 8.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global ecology and biogeography* 12: 361-371.

Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161-175.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications* 19: 181-197.

RCoreTeam 2021. R: A language and environment for statistical computing computer program, version By RCoreTeam, Vienna, Austria.

Richman, R., D. Diallo, M. Diallo, A. A. Sall, O. Faye, C. T. Diagne, I. Dia, S. C. Weaver, K. A. Hanley, and M. Buenemann. 2018. Ecological niche modeling of *Aedes* mosquito vectors of chikungunya virus in southeastern Senegal. *Parasites & Vectors* 11: 255.

Ripley, W. N. V. a. B. D. 2002. Modern Applied Statistics with S computer program, version By Ripley, W. N. V. a. B. D.

- Ryan, S. J., C. J. Carlson, E. A. Mordecai, and L. R. Johnson. 2019.** Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS neglected tropical diseases* 13: e0007213.
- Stefopoulou, A., G. Balatsos, A. Petraki, S. LaDeau, D. Papachristos, and A. Michaelakis. 2018.** Reducing Aedes albopictus breeding sites through education: A study in urban area. *PLOS ONE* 13: e0202451.
- Tiffin, H. S., S. T. Peper, A. N. Wilson-Fallon, K. M. Haydett, G. Cao, and S. M. Presley. 2019.** The Influence of New Surveillance Data on Predictive Species Distribution Modeling of Aedes aegypti and Aedes albopictus in the United States. *Insects* 10: 400.
- Toledo Romani, M. E., V. Vanlerberghe, D. Perez, P. Lefevre, E. Ceballos, D. Bandera, A. Baly Gil, and P. Van der Stuyft. 2007.** Achieving sustainability of community-based dengue control in Santiago de Cuba. *Social Science & Medicine* 64: 976-988.
- Toxopeus, B. N. a. N. a. s. H. a. T. A. G. a. A. K. S. a. A. G. 2014.** Where is positional uncertainty a problem for species distribution modelling. *Ecography* 37: 191-203.
- van Proosdij, A. S. J., M. S. M. Sosef, J. J. Wieringa, and N. Raes. 2016.** Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39: 542-552.
- Vezzani, D., A. Rubio, S. M. Velázquez, N. Schweigmann, and T. Wiegand. 2005.** Detailed assessment of microhabitat suitability for Aedes aegypti (Diptera: Culicidae) in Buenos Aires, Argentina. *Acta Tropica* 95: 123-131.
- Walther, D., and H. Kampen. 2017.** The Citizen Science Project ‘Mueckenatlas’ Helps Monitor the Distribution and Spread of Invasive Mosquito Species in Germany. *Journal of Medical Entomology* 54: 1790-1794.

SUPPLEMENTAL DATA

Supplemental Table 1. Variable descriptions of all environmental variables in WorldClim dataset and examined in variable inflation factor analysis. Variables were chosen to be used in the Maxent models if they had a variable inflation factor of under 5.

Variable	Description	Selected for Maximum Entropy Modeling
Bio1	Annual Mean Temperature	No
Bio2	Mean Diurnal Range (Mean of monthly (max temp = min temp))	Yes
Bio3	Isothermality (Bio2/Bio7)x100	No
Bio4	Temperature Seasonality (Standard Deviation x 100)	No
Bio5	Max Temperature of Warmest Month	No
Bio6	Min Temperature of Coldest Month	No
Bio7	Temperature Annual Range (Bio5-Bio6)	No
Bio8	Mean Temperature of Wettest Quarter	Yes
Bio9	Mean Temperature of Driest Quarter	Yes
Bio10	Mean Temperature of Warmest Quarter	No
Bio11	Mean Temperature of Coldest Quarter	Yes
Bio12	Annual Precipitation	No
Bio13	Precipitation of Wettest Month	Yes
Bio14	Precipitation of Driest Month	Yes
Bio15 ⁺	Precipitation Seasonality (Coefficient of Variation)	Yes
Bio16	Precipitation of Wettest Quarter	No
Bio17	Precipitation of Driest Quarter	No
Bio18	Precipitation of Warmest Quarter	No
Bio19	Precipitation of Coldest Quarter	Yes

+ The coefficient of variation for precipitation seasonality is the ratio of the standard deviation of the monthly total precipitation to the mean monthly total precipitation expressed as a percentage (Hijmans et al. 2005).

Supplement Figure 1. A. Model 4; Maximum Entropy output of *Ae. aegypti* habitat suitability in West Africa using data from 1999-2008 literature data. **B.** Model 5; Maximum Entropy output of *Ae. aegypti* habitat suitability in West Africa using 1999-2018 literature data, **C.** Model 6; Maximum Entropy output of *Ae. aegypti* habitat suitability in West Africa using a combination of GLOBE Observer app data and 1999-2018 literature data.

- *Aedes aegypti*
- *Aedes albopictus*
- *Aedes incerta*
- Unidentified *Aedes*
- ▲ *Anopheles*
- ▲ *Culex*
- Unknown mosquito

bio11 Mean Temperature of Coldest Quarter
bio13 Precipitation of Wettest Month
bio14 Precipitation of Driest Month
bio15 Precipitation Seasonality
bio19 Precipitation of Coldest Quarter
bio2 Mean Diurnal Temperature Range
bio8 Mean Temperature of Wettest Quarter
bio9 Mean Temperature of Driest Quarter

A **B** **C**

A **B** **C**