

# Leveraging VectorBase/VEuPathDB Mosquito Data for Enhanced Modeling

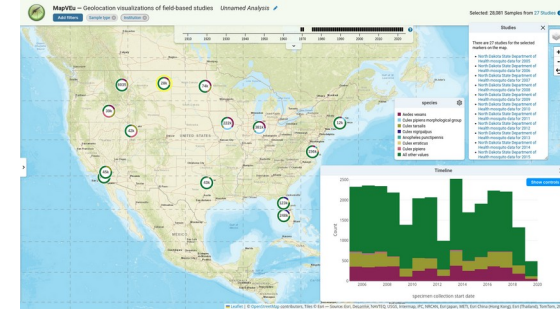
Bob MacCallum, Imperial College London

# What we do/did for the modelling community

Predominantly the work of Sarah Kelly and Sam Rund

## Large-scale curation of US mosquito surveillance data

- 1.4 million samples from 20 states
- 26 mosquito control districts have provided 2 or more years' worth of data
- Excellent coverage in Florida (*Aedes albopictus* & *aegypti*)



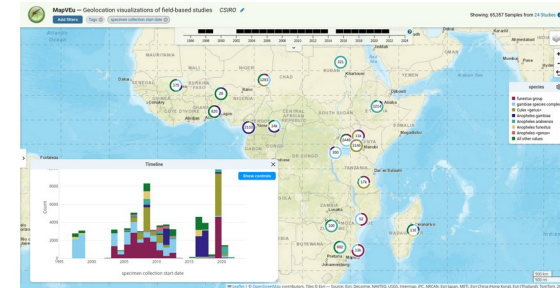
## Targeted curation for CSIRO modellers

- 65k samples from 25k collections from 1,800 sites in 18 countries from 24 datasets

## Genotype & phenotype data

- 100s of datasets

Curation policy...



# Exploratory data analysis tool development

Not focus for today!

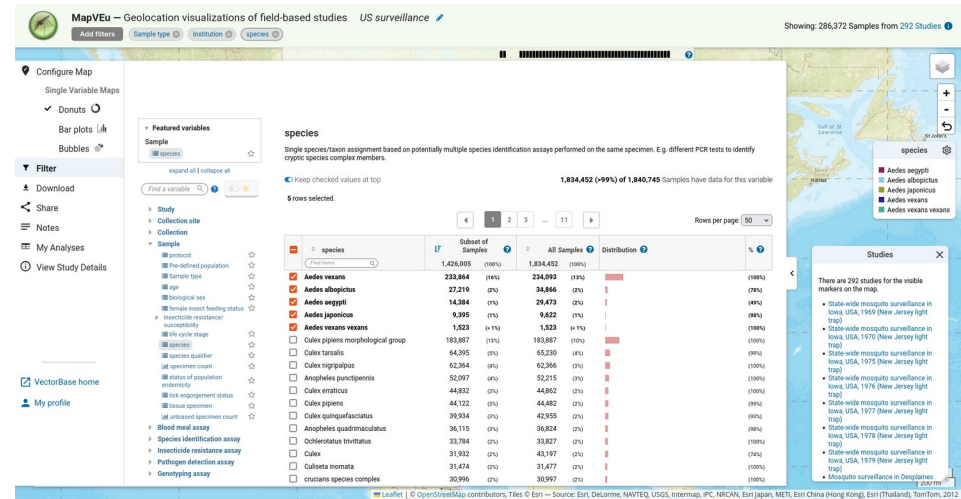
But it's (re)designed for

- browse
- filter
- visualize
- download

Emphasis on showing data and metadata in space and time

Not hard-coded to any data types, columns, etc

Powerful, difficult to deploy



# Mosquito abundance - more than just numbers...

Sam Rund, Sarah Kelly and I have worked on this for some time, among others...

Normalised abundance requires

- **C**: Total number of mosquitoes collected (Count).
- **D**: Total duration of trapping effort (Duration in nights).
- **T**: Total number of traps deployed.

$$\text{Abundance} = \frac{C}{D \times T}$$

Article | [Open access](#) | Published: 25 April 2019

## **MIReAD, a minimum information standard for reporting arthropod abundance data**

[Samuel S. C. Rund](#) , [Kyle Braak](#), [Lauren Cator](#), [Kyle Copas](#), [Scott J. Emrich](#), [Gloria I. Giraldo-Calderón](#), [Michael A. Johansson](#), [Naveed Heydari](#), [Donald Hobern](#), [Sarah A. Kelly](#), [Daniel Lawson](#), [Cynthia Lord](#), [Robert M. MacCallum](#), [Dominique G. Roche](#), [Sadie J. Ryan](#), [Dmitry Schigel](#), [Kurt Vandegrift](#), [Matthew Watts](#), [Jennifer M. Zaspel](#) & [Samraat Pawar](#)

*Scientific Data* **6**, Article number: 40 (2019) | [Cite this article](#)

**3254** Accesses | **22** Citations | **41** Altmetric | [Metrics](#)

# Absence makes the disk grow fuller

Ideally, providers would report zero counts for every species of interest.

However, they often do not.

Even if they did, storing zeroes would typically require 20 times more storage space.

Zeroes are crucial when calculating **mean abundance** across multiple collection sites and dates.

For example, consider four one-night collections in July that record counts of 10, 20, 0, and 10 mosquitoes of species X.

- The mean abundance is calculated as

$$\text{Mean Abundance} = \frac{10 + 20 + 0 + 10}{1 + 1 + 1 + 1} = 10 \text{ mosquitoes/trap/night}$$

- Even if we don't store the zero count, it is essential to know which collections were actively sampling for species X to ensure the correct denominator is used.

# Assumptions for zero counts

Unless otherwise communicated, the list of *observed* species is used to infer zeroes.

Collections (site + date) with at least one non-zero count are assumed to have zero counts for all other species.

Providers must provide information about collections where *nothing at all* was collected.

(We do not add zeroes for fictional "out of season" collections.)

Legacy map had "include zeroes" option in downloads. Current version does not.

# Mosquito abundance - more than just numbers...

In order to download consistent data, further metadata is desirable

- type of trap or collection method
- attractants
- trap placement, where relevant
- pathogen assay results
- species names are consistent and typo-free

**Ontologies** help with this

VectorBase's data loading procedures use(d) ontologies mainly for standardisation of column names and values

# The human touch

AI might help with data processing but historically we have learnt that

1. almost every dataset requires to-and-fro with data provider
2. disaggregation (e.g. summary data with %WNV & %YFV)
3. species abbreviation clarification
4. "hybrid" species clarification



# No more NIH/NIAID funding

After 20 years, NIH will no longer fund the VEuPathDB Bioinformatics Resource Center.

VEuPathDB is the umbrella for VectorBase, PlasmoDB, TriTrypDB, CryptoDB and several more pathogen/vector interest areas.

Also ClinEpiDB, MicrobiomeDB, OrthoMCL

*New awardee will almost certainly not continue to work with the community to clean, process, manage and disseminate field-based data.*

# Future of VectorBase/VEuPathDB

US NIH awarded renewal to another team.

Websites are currently down.

Emergency funding has been found to bring back the sites in static form

Longer-term sustainability to be developed.

Specific additional support will be needed for data curation activities, especially outside of genomics.



# Continuity

Curation expertise continues with Sarah Kelly working with Lauren Cator and Samraat Pawar for VectorByte and *The One Health Vector-Borne Diseases Hub*

US surveillance → VecDyn (part of VectorByte)

Hope to build self-service data management capacity within VEuPathDB for mosquito surveillance efforts (including the interactive map)

**A new £1.5m hub will support the sharing of data and findings on 'vector-borne' diseases among the UK research community and with policymakers.**

The One Health Vector-Borne Diseases Hub has been awarded £1.5m funding from Defra and UK Research and Innovation, via the [Biotechnology and Biological Sciences Research Council](#). Imperial College London researchers will lead the UK-wide hub with partners at [The Pirbright Institute](#), the [London School of Hygiene & Tropical Medicine](#), the [University of Liverpool](#) and the [UK Centre for Hydrology & Ecology](#).

# Thanks

All the VectorBase and VEuPathDB team including

- Dan Lawson
- Ioannis Kiritizoglou
- Sarah Kelly
- Sam Rund

And the community for engaging with and guiding our work.

# Success stories



RESEARCH

INSECT POPULATIONS

## Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances

Roel van Klink<sup>1,2,3\*</sup>, Diana E. Bowler<sup>1,4,5</sup>, Konstantin B. Gengalsky<sup>6,7</sup>, Ann B. Swengel<sup>8</sup>, Alessandro Gentile<sup>1</sup>, Jonathan M. Chase<sup>1,9</sup>

Recent case studies showing substantial declines of insect abundances have raised alarm, but how widespread such patterns are remains unclear. We compiled data from 166 long-term surveys of insect assemblages across 1676 sites to investigate trends in insect abundances over time. Overall, we found considerable variation in trends even among adjacent sites but an average decline of terrestrial insect abundance by -9% per decade and an increase of freshwater insect abundance by +11% per decade. Both patterns were largely driven by strong trends in North America and some European regions. We found some associations with potential drivers (e.g., land-use drivers), and trends in protected areas tended to be weaker. Our findings provide a more nuanced view of spatiotemporal patterns of insect abundance trends than previously suggested.

Insects are the most ubiquitous and diverse | 20 years. Because our main focus was on the

Used data from many sources, including VectorBase to reconstruct long term insect population trends to investigate the ‘insect apocalypse’

ev  
es  
w  
in  
in  
Fi  
di  
20  
2,  
fo  
te  
se  
pe  
[e  
ix  
m  
er  
pe  
ch



OPEN

# Climate, landscape, and life history jointly predict multidecadal community mosquito phenology

Lindsay P. Campbell<sup>1✉</sup>, Mohamed F. Sallam<sup>2</sup>, Amely M. Bauer<sup>1</sup>, Yasmin Tavares<sup>2</sup> & Robert P. Guralnick<sup>3</sup>

Phenology of adult host-seeking female mosquitoes is a critical component for understanding potential for vector-borne pathogen maintenance and amplification in the natural environment. Despite this importance, long-term multi-species investigations of mosquito phenologies across environments and differing species' life history traits are rare. Here we leverage long-term mosquito control district monitoring data to characterize annual phenologies of 7 host-seeking female mosquito species over a 20-year time period in suburban Illinois, USA. We also assembled data on landscape context, categorized into low and medium development, climate variables including precipitation, temperature and humidity, and key life history traits, i.e. overwintering stage and Spring–Summer versus Summer–mid-Fallseason fliers. We then fit linear mixed models separately for adult onset, peak abundances, and flight termination with landscape, climate and trait variables as predictors with species as a random effect. Model results supported some expectations, including warmer spring temperatures leading to earlier onset, warmer temperatures and lower humidity leading to earlier peak abundances, and warmer and wetter fall conditions leading to later termination. However, we also found sometimes complex interactions and responses contrary to our predictions. For example, temperature had generally weak support on its own, impacting onset and peak abundance timing; rather temperature has interacting effects with humidity or precipitation. We also found higher spring precipitation, especially in low development contexts, generally delayed adult onset, counter to expectations. These results emphasize the need to consider how traits, landscape and climatic factors all interact to determine mosquito phenology, when planning management strategies for vector control and public health protection.

Mosquito control surveillance trap data were downloaded from VectorBase PopBio data repository (<https://vectorbase.org/popbio-map/web/>)<sup>36</sup>. The data set included multiple fields, which are outlined in an extensive data dictionary provided through the data repository. For the purpose of this study, we included latitude and longitude coordinates of the collection site, the collection date, trap type, attractant, and an abundance value for individual species. Data were filtered to include only New Jersey light traps with visible light as the attractant between the years 2000–2020.





# Random Forest Analysis of Impact of Abiotic Factors on *Culex pipiens* and *Culex quinquefasciatus* Occurrence

Arinder K. Arora<sup>1</sup>, Cheolho Sim<sup>2</sup>, David W. Severson<sup>3</sup> and David S. Kang<sup>4\*</sup>

<sup>1</sup>Department of Entomology, University of Florida, Gainesville, FL, United States, <sup>2</sup>Department of Biology, Baylor University, Waco, TX, United States, <sup>3</sup>Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN, United States, <sup>4</sup>USDA Agricultural Research Service, Biological Control of Insects Research Laboratory, Columbia, MD, United States

The *Culex pipiens* complex of mosquitoes are significant vectors of several pathogens resulting in infectious human diseases in North America, including but not limited to West Nile encephalitis, Rift Valley Fever, and Lymphatic filariasis. Among this complex are *C. pipiens* form *pipiens* and *Culex quinquefasciatus*. While morphologically similar, the mosquitoes exhibit unique life histories that suit them uniquely to divergent niches, wherein *C. pipiens* can thrive despite the cold winters of the northern United States and *C. quinquefasciatus* is able to survive periods of drought typical in the southern states. Here, Random Forests machine-learning algorithms were employed to model and explore which environmental parameters best explain mosquito occurrence in historical trapping data across the continental United States of America, and test correlation with abundance data. The models explained between 71 and 97% of the presence or absence of the two mosquitoes based on historical climatic data. The results of this study will improve vector management programs by explaining which environmental variables will provide the most accurate predictions of mosquito presence at a given site.

**Keywords:** random forest, *Culex*, mosquito, climate, machine learning

## INTRODUCTION

*Culex pipiens* complex mosquitoes are major vectors of several pathogens that cause human diseases such as West Nile encephalitis, Rift Valley Fever, and Lymphatic filariasis (Morgan et al., 1980; Monath, 1988; Lari et al., 2009; Diamond, 2009). One of the mosquitoes belonging to this group, the southern house mosquito *Culex quinquefasciatus*, is found in warmer regions with its distribution limited below a latitude of 39°N in the continental United States America (Alanz et al., 2019). Its sister species, the northern house mosquito *Culex pipiens* form *pipiens*, are found in the northern temperate regions of North America. However, the hybrid zone generally exists at latitudes between 30 and 40°N (Barr, 1957) between these two sister species although some population genetics studies showed that introgression among the *C. pipiens* complex species is much more widespread (Kobers et al., 2009, 2012; Huang et al., 2011). Despite these contradictory results of introgression in *C. pipiens* complex, each species exhibits unique life histories adapted to their specific ecological niches (Cornell et al., 2003; McAbee et al., 2008; Kobers et al., 2012; Kang et al., 2021) and purifying selection appears to maintain the unique genetic identities wherever the two biotypes are found

(Fonseca et al., 2004; Arensburg et al., 2010; Amraoui et al., 2012; Hickner et al., 2013; Kang and Sim, 2013).

Environmental conditions are considered a significant effector of *Culex* population structure and shape differences in life history, including geographical distribution, behavior, reproduction, host preference, and diapause (overwintering) (Barr, 1957; Spielman, 1967; Harbach et al., 1984; Clements, 1992). Crucially, temperature profoundly influences the life cycle of this mosquitoes complex and is likely a major factor in its geographic distribution and consequently its ability to disseminate diseases (Chaves and Kitron, 2011; Chuang et al., 2011). A distinct temperature-related evolutionary adaptive feature of the *C. pipiens* complex is the presence or absence of overwintering diapause. Diapause is an anticipated, preprogrammed response of insects to the shortening of days and lower temperatures at the onset of winter and is characterized by a developmental arrest resulting in reduced metabolism, enhanced stress tolerance, and fat hypotrophy (Mori et al., 2007; Sim and Denlinger, 2013; Kang et al., 2014, 2016; Sim et al., 2015). This reallocation of resources toward survival is a critical adaptation for the northern house mosquito, *C. pipiens*. The absence of the diapausing phenotype would cull the southern mosquitoes from *C. pipiens* populations, yet fails to fully explain the low level of introgression of *C. pipiens* into the southern *C. quinquefasciatus* populations in North America.

Understanding how the changing climate influences the distribution of these mosquitoes is essential for management and control efforts. The interaction between climatic and local factors makes for complex processes. Multiple processes likely drive species distribution at different scales, and often simple models will fail to account for hidden relationships (Wagner and Fortin, 2005). This is further complicated by the nature of geospatial data, which often present complex, non-linear, high dimensional datasets (Cutler et al., 2007). This can present a problem for traditional methods such as generalized linear models (De'ath and Fabricius, 2000).

Random forests is a machine-learned method of pattern recognition that is very well suited to ecological classification and regression problems, and a chief use of the Random forests algorithm is to determine the relative importance of each feature in explaining a phenomenon (Cutler et al., 2007). Random forests is an algorithm in which multiple decision trees are combined (bagging/bootstrap aggregation) to solve classification or regression problems. This approach can compensate for errors in unbalanced datasets, which are common in ecological data (Breiman, 2001). This approach helps overcome the problem individual decision trees have with overfitting. Random forests do not need assumptions about the distribution of the data, are relatively robust when the quality of the data is inconsistent, are adept at handling outliers, and require no scaling or transformation of datasets (Ruiz et al., 2010). The algorithm reduces data into subsets and can handle high-dimensional datasets. Indeed, the potential of Random forests has already yielded promising results at a small scale in a landmark study considering the impacts of rainfall and temperature data on infection rates of *C. pipiens* and *Culex restuans* mosquitoes in the greater Chicago area by West Nile Virus (Ruiz et al., 2010).

Additionally, machine learning has been gaining popularity in other entomological studies to explain the abiotic drivers of insect trap catch (Midgarden and Lira, 2006; Enkerlin et al., 2016; Bekker et al., 2019).

Here, we test whether a random forest trained on historical climatic and mosquito collection data can accurately explain future mosquito trapping occurrence within the continental United States. We examine various environmental factors, including minimum, maximum, and mean temperature, rainfall, elevation, and dew point, to determine the relative importance of each environmental factor to model accuracy. We explained above that resilience to the cold is of utmost biological relevance to these mosquitoes. Here, we investigate whether minimum temperature can explain mosquito trap events. Further, we examine whether different combinations of climatic factors can be combined to improve the accuracy of our model.

## MATERIALS AND METHODS

### Collection of Data

Mosquito abundance data of the United States comprising all the parameters were downloaded from Vectorbase<sup>1</sup> on Dec 28, 2019 (Supplementary Table 1). The trap data ranged from 2010 to 2019. The collection data from Vectorbase was comprised of 642,812 *C. pipiens* and 209,658 *C. quinquefasciatus* collected from various locations in the United States. Because entries in varied from samples trapped per day and per month, the data were combined into per year counts per trap site. The data were rearranged and aggregated to obtain the sum of each mosquito species collected per year irrespective of the trap used for their collection. After curation there were 544,040 *C. pipiens* and 139,054 *C. quinquefasciatus* for the dataset leading up to 2016 and 98,772 *C. pipiens* and 70,601 *C. quinquefasciatus* for the post-2017 dataset. After combining the data per year in the data leading up to 2016, we had 2,336 zeroes for *C. pipiens* and 2,009 zeroes for *C. quinquefasciatus*. For the post-2017 data we have 486 zeroes for *C. pipiens* and 560 zeroes for *C. quinquefasciatus*. We included all the trap available trap and scored trap sites collecting mosquitoes other than *C. pipiens* and *C. quinquefasciatus* as negative trapping events.

In addition to species abundance and year of collection, other retained parameters were the longitude and latitude of each trap. Environmental data including minimum, maximum, and mean temperature, rainfall, elevation, and dew point were retrieved from the PRISM climate group.<sup>2</sup> The available environmental data was 30-year average for all the parameters.

### Collating Mosquito Abundance Data and Environmental Data

The coordinates of environmental data were matched using QGIS<sup>3</sup> (QGIS, 2020) to align the mosquito abundance with environmental parameters. Subsequently, the mosquito

<sup>1</sup><https://www.vectorbase.org/popbio/>

<sup>2</sup><http://www.prism.oregonstate.edu/>

<sup>3</sup><https://qgis.org/en/site/>

## OPEN ACCESS

### Edited by:

Bradley S. Callie,  
Auckland University of Technology,  
New Zealand

### Reviewed by:

Miguel Moreno-García,  
Servicios de Salud de Morelos,  
Coordinador de Enfermedades  
Transmisibles por Vector y Zoonosis,  
Mexico

Camilo E. Khachatryan,  
Dartmouth College, United States

### \*Correspondence:

David S. Kang  
Dave.Kang@usda.gov

### Specialty section:

This article was submitted to  
Models in Ecology and Evolution,  
a section of the journal  
Frontiers in Ecology and Evolution

Received: 09 September 2021

Accepted: 20 December 2021

Published: 27 January 2022

### Citation:

Arora AK, Sim C, Severson DW  
and Kang DS (2022) Random Forest  
Analysis of Impact of Abiotic Factors  
on *Culex pipiens* and *Culex  
quinquefasciatus* Occurrence.  
Front. Ecol. Evol. 9:773360.  
doi: 10.3389/fevo.2021.773360





# Effects of landscape on mosquito community composition and abundance in Manatee County, Florida

Amely M. Bauer<sup>1,2,\*</sup>, Shelly Whitehead<sup>3</sup>, Narayani Barve<sup>4</sup>, Mohamed Sallam<sup>5</sup>, Robert Guralnick<sup>4</sup>, Julie M. Allen<sup>3</sup>, Kenneth Linthicum<sup>6</sup>, Seth Britch<sup>7</sup>, Samuel Rund<sup>7,8</sup> and Lindsay P. Campbell<sup>1,2</sup>

<sup>1</sup>Florida Medical Entomology Laboratory, University of Florida, Vero Beach, FL, <sup>2</sup>Department of Entomology and Nematology | IFAS, University of Florida, Gainesville, FL, <sup>3</sup>Manatee County Mosquito Control District, Palmetto, FL, <sup>4</sup>Department of Natural History and the Florida Museum of Natural History, University of Florida, Gainesville, FL, <sup>5</sup>Department of Biology, University of Nevada, Reno, Reno, NV, <sup>6</sup>United States Department of Agriculture, Agricultural Research Service Center for Medical, Agricultural, and Veterinary Entomology, Gainesville, FL, <sup>7</sup>Department of Biological Sciences, Galvan Life Science Center, University of Notre Dame, Notre Dame, IN, <sup>8</sup>EC Institute for Global Health, University of Notre Dame, Notre Dame, IN

## INTRODUCTION

- Mosquitoes are the most important vectors of arthropod-borne disease [1]
- Vector-borne disease transmission requires vector, host and pathogen to be present and overlap in space and time [2]
- Abiotic and biotic factors can affect the number of mosquito species present at a location and overall abundance, which can impact pathogen circulation and opportunities for transmission [3]
- Joint species distribution models are a powerful tool for predicting the probability of occurrence of multiple mosquito species across a geographic area [4]

## OBJECTIVE

- Quantify effects of landscape on mosquito community composition, abundances, and diversity
- Generate prediction maps for West Nile virus vector competent mosquitoes in Manatee County, Florida, USA



## CONCLUSIONS & RELEVANCE

- Increased percentages of developed landcover associated with decreased species richness, but proportion of competent WNV vector species increased
- Community-weighted mean traits can provide insights relevant to public health, vector control management, and research investigating disease systems and ecology
- Predicting distributions of individual species and mosquito communities across different landscapes can support informed land use planning

## CONTACT

amelybauer@ufl.edu

## ACKNOWLEDGEMENTS

We thank the Manatee County Mosquito Control District and personnel for mosquito collection, identification, and digitization, the University of Florida Biodiversity Institute, David Fine Aerial, and the University of Florida Entomology and Nematology Department for awarding AME a Graduate School Funding Award.

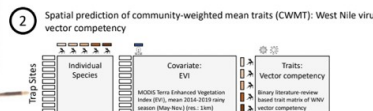
## REFERENCES

[1] Smith, G. A., Reagin, J. M., & Hagan, D. T. (2018). Arthropod-borne and zoonotic diseases in the United States. In H. Reagin (Ed.), *State and zoonotic diseases*. Elsevier Academic Press, London. [2] Reagin, H. (2018). *State and zoonotic diseases*. Elsevier Academic Press, London. [3] Reagin, H. (2018). *State and zoonotic diseases*. Elsevier Academic Press, London. [4] Reagin, H. (2018). *State and zoonotic diseases*. Elsevier Academic Press, London. [5] Reagin, H. (2018). *State and zoonotic diseases*. Elsevier Academic Press, London. [6] Reagin, H. (2018). *State and zoonotic diseases*. Elsevier Academic Press, London. [7] Reagin, H. (2018). *State and zoonotic diseases*. Elsevier Academic Press, London. [8] Reagin, H. (2018). *State and zoonotic diseases*. Elsevier Academic Press, London.

## METHODS

Weekly mosquito abundance data was collected across 56 mosquito control district surveillance sites in Manatee County, Florida using CO<sub>2</sub> baited CDC light traps from May to November 2020. Data was downloaded from vectorbase.org, and the mean number of mosquitoes per trap night for 40 species was calculated across the season at each trap location.

A joint species distribution modeling approach using the hierarchical modeling of species community (Hmsc) package in R was used to quantify landscape and vegetation greenness effects on community diversity, on individual species occurrence, and to predict West Nile virus vector distributions across the county.

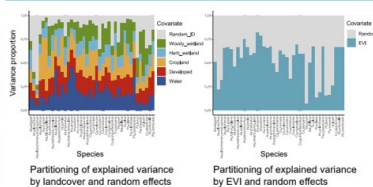


- Prediction**
- CWMT for West Nile virus vector competency calculated from Hmsc model output
  - Prediction of proportion of WNV vector competent mosquito species across Manatee County
  - Prediction of proportion of *Culex quinquefasciatus* and *Cx. nigripalpus* across Manatee County

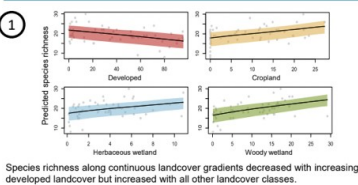
**Model parameterization & evaluation [1]**

- Hmsc models used species abundances as response variables, environmental covariates as fixed effects, a random effect by site, a spatial random effect to account for spatial autocorrelation, and a prior distribution.
- Default prior distributions, posterior distribution sampled using 3 MCMC chains, each with 10,000 sample iterations, thinned at 100 iterations, and a burn-in of 1,000 iterations
- Gelman's diagnostic indicated convergence across all parameter estimates for each species in both models with values close to or equal to 1

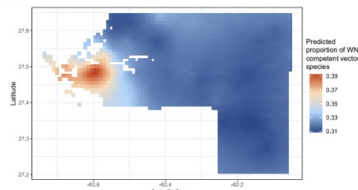
## RESULTS



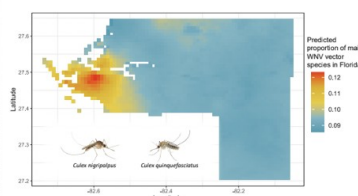
## RESULTS : EFFECTS & PREDICTIONS



Species richness along continuous landcover gradients decreased with increasing developed landcover but increased with all other landcover classes.



Concentration of vector competent species predicted in the most urbanized area in Manatee County.



West Nile virus transmission to humans in Florida is mainly driven by *Culex quinquefasciatus* and *Culex nigripalpus*. Their predicted distribution pattern resembles the pattern of all WNV competent vector species, including a focus area in northern Manatee County.

**WHAT'S NEXT?**

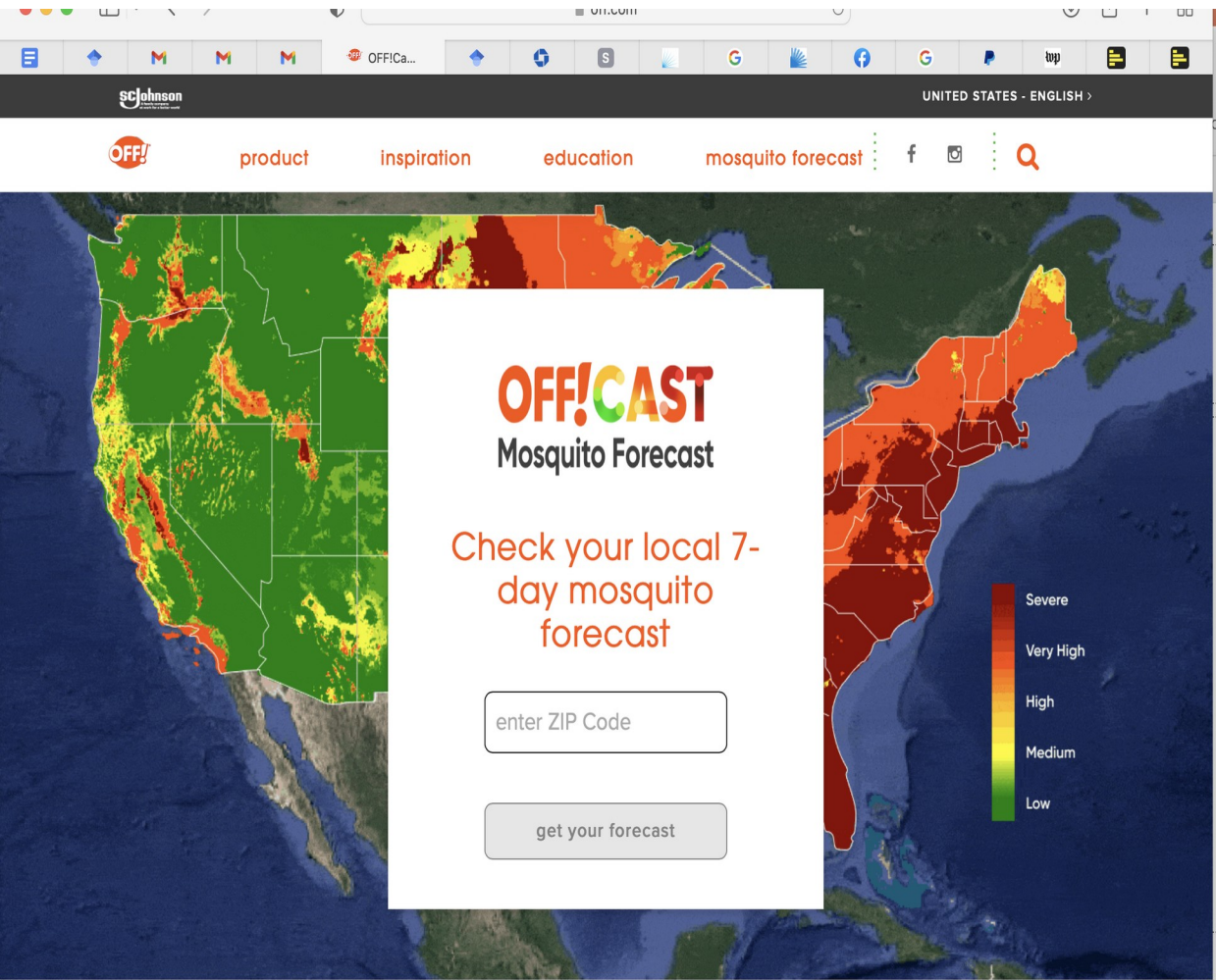
- Refine Hmsc model adding additional environmental and climatic data
- Expand predictions over space and time to predict impacts of land use and land cover change on mosquito assemblages and trait-based distributions
- Validation study to assess accuracy of the model

ARTICLE | Open Access

## Land use predicts proportion of West Nile virus vector-competent mosquitoes

Amely M. Bauer, Robert P. Guralnick, Shelley A. Whitehead, Narayani Barve, Julie M. Allen, Lindsay P. Campbell

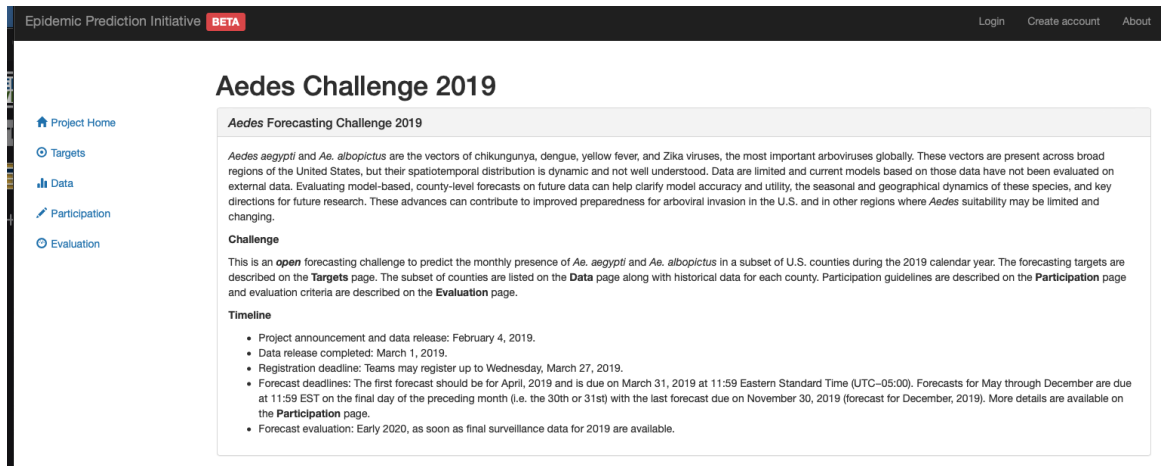
First published: 07 February 2024 | <https://doi.org/10.1002/ecs2.4771>



- Off mosquito forecast <https://off.com/en/mosquitoforecast>
- Validated using mapveu data
- Read more <https://www.cnet.com/health/medical/google-and-off-launch-mosquito-forecast-tool/>
- Google and SC Johnson project

# Species distribution & Forecasting

- Kamil Erguler (Cyprus) we know is using our data to do mosquito species distribution modeling
- We know some participants in CDC *Aedes* Forecasting Challenge 2019 used our data to make their models



The screenshot shows the 'Epidemic Prediction Initiative' website in BETA mode. The navigation menu includes 'Project Home', 'Targets', 'Data', 'Participation', and 'Evaluation'. The main content area is titled 'Aedes Challenge 2019' and contains the following text:

**Aedes Forecasting Challenge 2019**

*Aedes aegypti* and *Ae. albopictus* are the vectors of chikungunya, dengue, yellow fever, and Zika viruses, the most important arboviruses globally. These vectors are present across broad regions of the United States, but their spatiotemporal distribution is dynamic and not well understood. Data are limited and current models based on those data have not been evaluated on external data. Evaluating model-based, county-level forecasts on future data can help clarify model accuracy and utility, the seasonal and geographical dynamics of these species, and key directions for future research. These advances can contribute to improved preparedness for arboviral invasion in the U.S. and in other regions where *Aedes* suitability may be limited and changing.

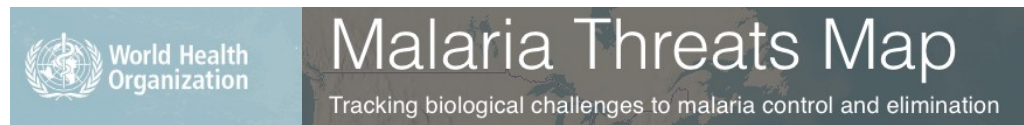
**Challenge**

This is an **open** forecasting challenge to predict the monthly presence of *Ae. aegypti* and *Ae. albopictus* in a subset of U.S. counties during the 2019 calendar year. The forecasting targets are described on the **Targets** page. The subset of counties are listed on the **Data** page along with historical data for each county. Participation guidelines are described on the **Participation** page and evaluation criteria are described on the **Evaluation** page.

**Timeline**

- Project announcement and data release: February 4, 2019.
- Data release completed: March 1, 2019.
- Registration deadline: Teams may register up to Wednesday, March 27, 2019.
- Forecast deadlines: The first forecast should be for April, 2019 and is due on March 31, 2019 at 11:59 Eastern Standard Time (JTC-05:00). Forecasts for May through December are due at 11:59 EST on the final day of the preceding month (i.e. the 30th or 31st) with the last forecast due on November 30, 2019 (forecast for December, 2019). More details are available on the **Participation** page.
- Forecast evaluation: Early 2020, as soon as final surveillance data for 2019 are available.

VB Dataflows  
out



GLOBI database of Global Biotic Interactions  
(what bites and cohabitates / parasitize what and  
what infects what)

GBIF has over 1 billion species occurrences records.  
Our records have been downloaded 2422 times via  
GBIF. Many of these users are likely doing species  
range modeling.

VectorBite works with modelers and mixes our data  
with non-vector species abundance profiles. Major  
component of new (sam's) NSF grant is to use this