

Application Notes

The Arbovirus Mapping and Prediction (ArboMAP) system for West Nile virus forecasting

Dawn M. Nekorchuk, PhD¹, Anita Bharadwaja, MPH², Sean Simonson, MPH³,
Emma Ortega, MPH³, Caio M.B. França, PhD^{4,5}, Emily Dinh, PhD⁶, Rebecca Reik, MPH⁶,
Rachel Burkholder, MPH⁶, Michael C. Wimberly  PhD^{1,*}

¹Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK 73019, United States, ²South Dakota Department of Health, Pierre, SD 57501, United States, ³Louisiana Department of Health, New Orleans, LA 70112, United States, ⁴Department of Biology, Southern Nazarene University, Bethany, OK 73008, United States, ⁵Quetzal Education and Research Center, Southern Nazarene University, San Gerardo de Dota, 11911, Costa Rica, ⁶Michigan Department of Health and Human Services, Lansing, MI 48909, United States

*Corresponding author: Michael C. Wimberly, PhD, Department of Geography and Environmental Sustainability, University of Oklahoma, 100 East Boyd Street, Norman, OK 73019, United States (mcwimberly@ou.edu)

Abstract

Objectives: West Nile virus (WNV) is the most common mosquito-borne disease in the United States. Predicting the location and timing of outbreaks would allow targeting of disease prevention and mosquito control activities. Our objective was to develop software (ArboMAP) for routine WNV forecasting using public health surveillance data and meteorological observations.

Materials and Methods: ArboMAP was implemented using an R markdown script for data processing, modeling, and report generation. A Google Earth Engine application was developed to summarize and download weather data. Generalized additive models were used to make county-level predictions of WNV cases.

Results: ArboMAP minimized the number of manual steps required to make weekly forecasts, generated information that was useful for decision-makers, and has been tested and implemented in multiple public health institutions.

Discussion and Conclusion: Routine prediction of mosquito-borne disease risk is feasible and can be implemented by public health departments using ArboMAP.

Lay Summary

West Nile virus (WNV) is the most common mosquito-borne disease in the United States. To reduce the risk of WNV, public health agencies distribute information about how to avoid mosquito bites and use insecticides to reduce the abundances of disease-transmitting mosquitoes. Information about when and where the risk of getting WNV is highest would help these agencies to target their activities and use limited resources more efficiently. To support this goal, we developed the ArboMAP software system for predicting the risk of WNV disease in humans. ArboMAP uses information about recent weather combined with data obtained from trapping mosquitoes and testing them for presence of WNV to predict how many human cases will occur in future weeks. Predictions extend throughout the current WNV season (typically May-September) and are made for each county within a state. The system is implemented as a set of free software tools that can be used by epidemiologists in state and municipal departments of health. Feedback from public health agencies in South Dakota, Louisiana, Oklahoma, and Michigan has been incorporated to enhance the usability of the system and design visualizations that summarize the forecasts.

Key words: mosquito; weather; surveillance; software; outbreak.

Background and significance

Diseases caused by mosquito-transmitted arboviruses are a global health threat. In the United States, West Nile virus (WNV) is the most common mosquito-borne disease. This virus is transmitted primarily by mosquitoes in the genus *Culex*, and wild birds are the zoonotic reservoir hosts.¹ Most human infections are asymptomatic or cause only mild symptoms, but ~25% cause West Nile fever and <1% result in severe neuroinvasive disease that can be fatal.² The burden of human WNV disease is highly variable. In the conterminous United States between 2009 and 2018, total annual cases ranged from 712 to 5674 and average annual incidence of WNV neuroinvasive disease varied from 0.02 cases/100 000

in Maine to 3.16 cases/100 000 in North Dakota.³ Public health responses to WNV include prevention messaging to encourage behaviors that prevent mosquito bites and vector control activities to reduce vector abundance.⁴ Prediction of WNV outbreaks would allow proactive targeting of disease prevention and mosquito control activities to reduce transmission.

WNV surveillance commonly involves trapping and testing of vector mosquitoes, and the presence of WNV-infected mosquitoes is a strong indicator of the local risk of human disease.⁵ The vectors and hosts of WNV are also sensitive to habitat availability, and WNV cases exhibit lagged responses to meteorological factors such as temperature and

Received: September 17, 2023; Revised: December 4, 2023; Editorial Decision: December 9, 2023; Accepted: December 20, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the American Medical Informatics Association.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

humidity.^{6–8} Mosquito infection rates and environmental variables have been used to develop predictive models to forecast human cases throughout the transmission season. These models accurately predict seasonal outbreaks early enough in the year to allow public health responses prior to the annual peak in cases.^{9–11} However, many public health agencies lack the software and expertise that is needed to implement disease forecasting.

Objectives

The objective of this project was to develop and implement the Arbovirus Mapping and Prediction (ArboMAP) software for WNV forecasting by epidemiologists working in state health departments in the United States.

Methods

System overview

ArboMAP is implemented in the R programming language using the RStudio interactive development environment with all code stored in an R Markdown script. The forecasting process begins with ingestion of new data and harmonization of multiple data sources into a unified format suitable for modeling (Figure 1). Models are calibrated using data from prior years, and recent observations are used to inform predictions of WNV cases during the current transmission season. A report containing summaries of the data and the forecasts is automatically generated. Forecasts are usually made for all counties within a US state and are produced by an epidemiologist or other public health professional working in a government agency that conducts vector-borne disease surveillance.

Input data

Three sources of data are required. The first is de-identified human case data from surveillance databases, with each case referenced by the date of symptom onset and the county of residence. These data are converted to a weekly indicator variable for the occurrence of one or more human cases in each county. The second data source is mosquito testing results, also from surveillance databases. These data include one record for each pool of mosquitoes tested referenced by test result (positive or negative), the date of collection, and the county of collection. ArboMAP calculates indices of mosquito infection from these data, including the mosquito infection growth rate, which has been shown to be an effective predictor of human WNV cases in South Dakota.¹⁰ The third data source includes environmental variables that fluctuate throughout the transmission season. These data can be county-level summaries of daily meteorological variables such as temperature, humidity, precipitation, and windspeed or remotely sensed variables such as land surface temperature and spectral indices.¹²

To facilitate access to meteorological data, we developed a version of the Retrieving Environmental Analytics for Climate and Health (REACH) app¹³ to access meteorological data for WNV forecasting. We used the gridMET meteorological dataset, which contains interpolated weather station data aggregated to daily summaries and downscaled to a 4 km grid.¹⁴ All processing and summarization of the meteorological grids takes place in the cloud using Google Earth Engine (GEE),¹⁵ and the user downloads daily county-level

summaries. The app includes a graphical user interface that displays the raw meteorological data and allows the user to specify a date range and location to download. Code to implement this app in GEE is provided with the ArboMAP distribution, or it can be accessed directly at (<https://daw-neko.users.earthengine.app/view/arbomap-gridmet>).

Forecasting models

ArboMAP uses generalized additive models (GAMs) that predict whether a county will have one or more human WNV cases in a week. These are implemented as “big additive models,” computationally efficient GAMs designed to work with large datasets, using the bam() function¹⁶ from the R mgcv library.¹⁷ Predictors include mosquito infection indices and meteorological variables summarized as distributed lags, where the lagged effects are modeled as smoothed functions of the number of days before the current week. Maximum lag length is a user specified parameter and varied from 151 days in Michigan and South Dakota to 181 days in Louisiana and Oklahoma. Several options are available for model specification, including (1) different indices for summarizing the mosquito infection data, (2) different combinations of environmental variables, (3) untransformed environmental data versus environmental anomalies, (4) a single, fixed set of distributed lags versus time-varying lags that change over the course of the WNV season, and (5) different spline functions for modeling the smoothed responses. Multiple models can be combined to generate predictions based on model ensembles. Model selection is carried out using an information theoretic approach in which alternative models are compared using Akaike’s Information Criterion.¹⁰

To predict WNV cases during the current year, the models are first calibrated using data from previous years. Then, all available current-year environmental and mosquito data are used to generate predictions for every week of the transmission season, including backcasts for past weeks and forecasts for future weeks. Generating backcasts as well as forecasts is essential because there are delays in the diagnosis and reporting of WNV cases, and the reported numbers of cases from recent weeks are usually incomplete. Predictions are validated by calibrating the model with historical data and comparing predictions of human case occurrence to observations that were not used in the fitting process.¹¹

User interactions

ArboMAP settings are controlled by parameters, with default values provided in the R markdown script. The parameters determine how the models will be implemented, specify the time periods of historical data used for model calibration and the current-year data used for forecasts, and indicate how results will be presented. This script can be directly edited and run in RStudio, or a small R script can be run to invoke a graphical user interface (GUI) using the built-in Shiny interface. The GUI can then be used to modify the parameters and launch ArboMAP. The software automatically calibrates the models, uses them to generate forecasts, and produces formatted results in HTML or PDF format.

Results

Forecast outputs

Model outputs are presented in a formatted report that was co-developed with partners in state health departments

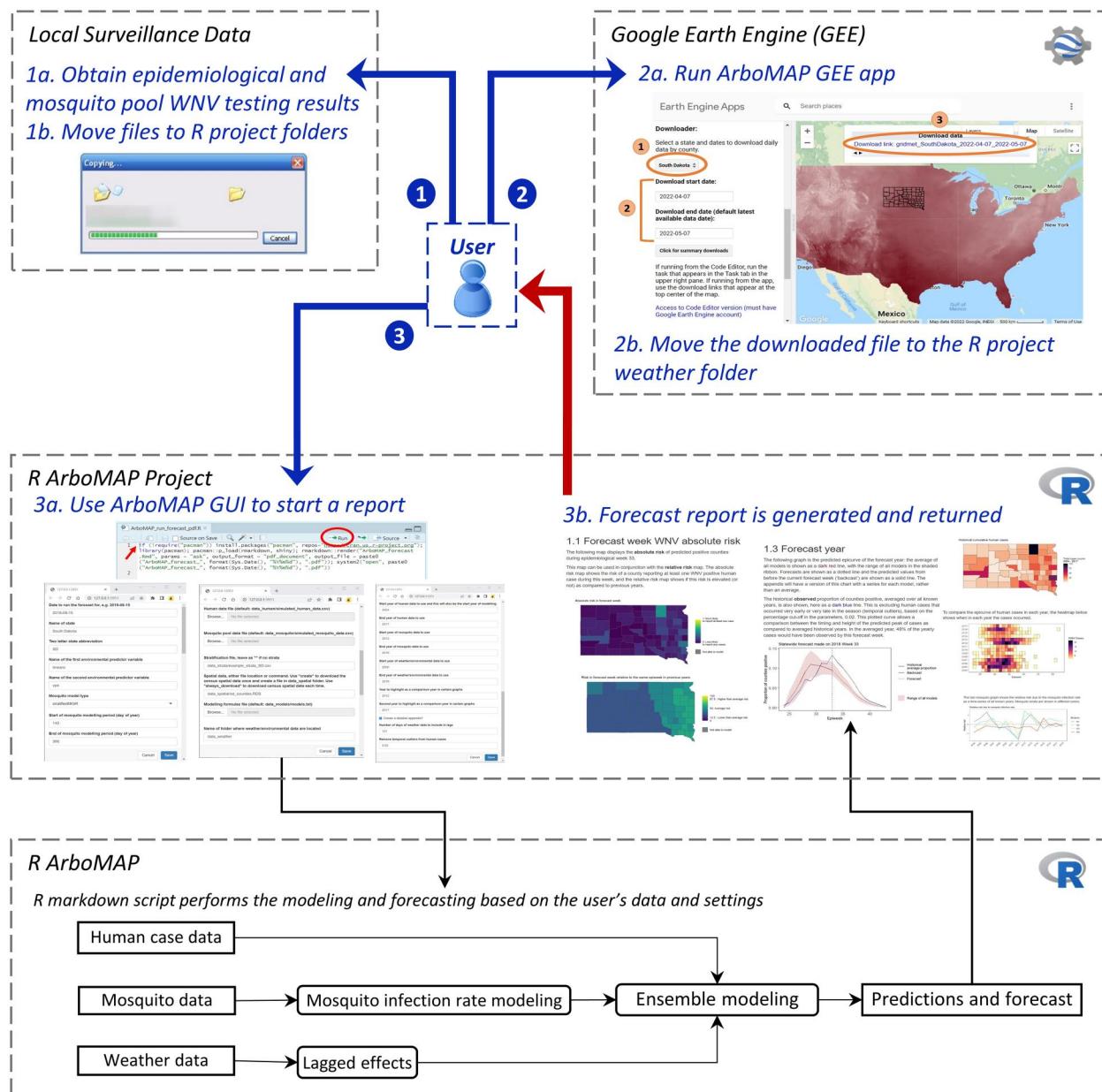


Figure 1. User-centered diagram showing the workflow for WNV forecasting. Step 1: Acquire updated entomological surveillance data, Step 2: Use the GEE app to update meteorological data, and Step 3: Use the RStudio GUI to generate a report. The system diagram on the bottom shows the high-level processes for modeling and forecasting.

(Figure 2). Because of the large amount of information, it is essential to present the most important components at the beginning where they are accessible. Thus, forecast results are provided first followed by summaries of the input data and an optional appendix containing diagnostic information about the models. The forecast results section includes predictions for the current week followed by summaries of forecasts and backcasts over the entire transmission season and comparisons of the current year predictions with the historical time series. When more than one model is used, only the ensemble mean of the predictions is presented in the forecast results section for simplicity. However, details on the individual models are available in the appendix and other ensemble metrics such as the median can also be calculated. Most of the outputs are shown as maps or graphs for ease of

interpretation and communication (Figure 2). Descriptive text is provided throughout to aid in interpreting the results.

Operational use

Before the beginning of the WNV transmission season, the data on human cases, mosquito infection, and meteorological variables must be brought up to date for all previous years. These historical data are used by ArboMAP for model calibration during the upcoming year. Decisions must also be made about the types of models and the predictor variables that will be used in the WNV forecasts. Evaluations of model fit and validations of model predictions in previous years can be conducted to inform these decisions.

The ArboMAP software was designed to minimize the number of manual steps required for a weekly forecast

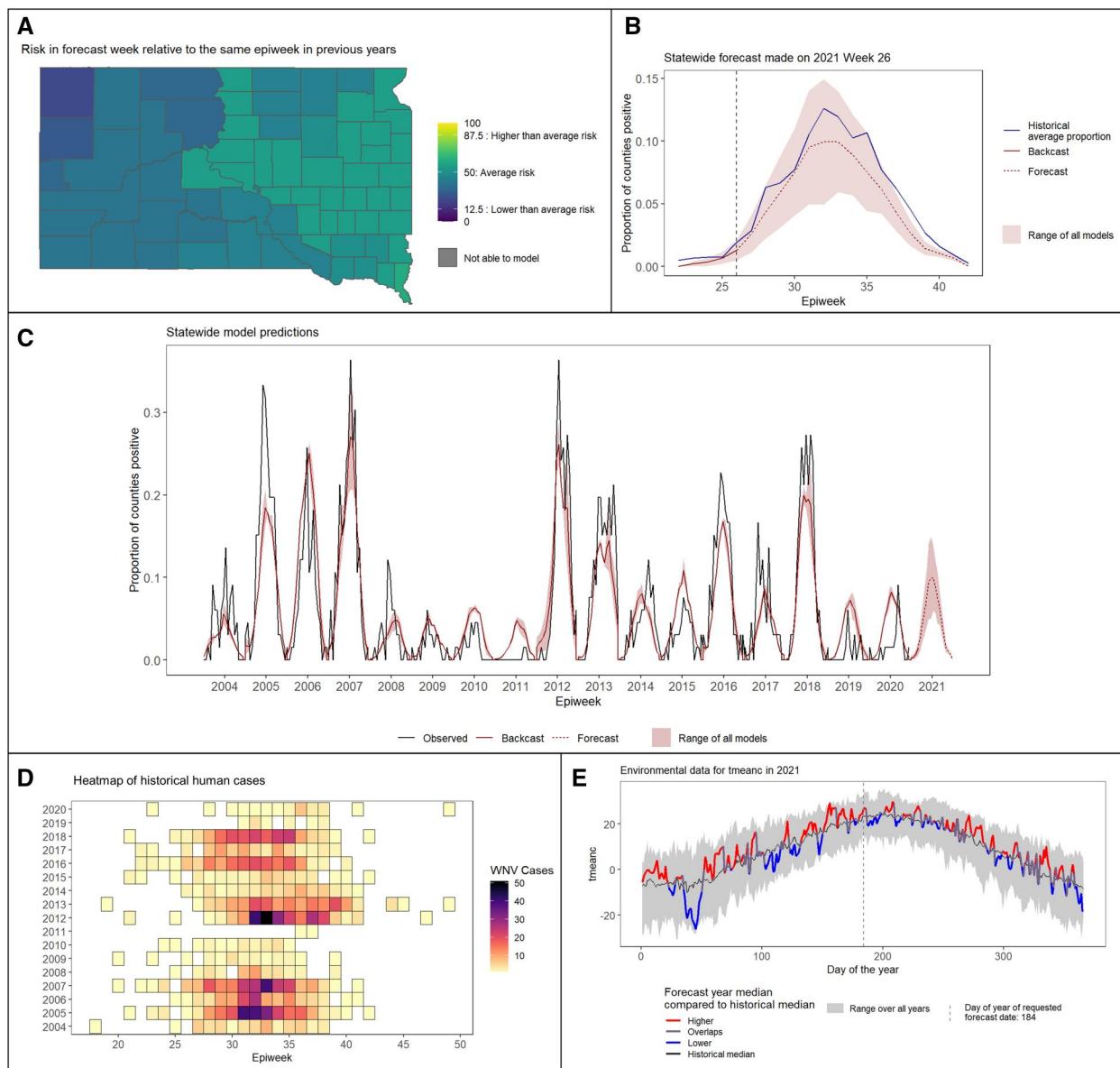


Figure 2. Examples of charts from an ArboMAP report for 2021 week 26 in South Dakota. (A) The relative risk of a county having at least one positive human West Nile virus. (B) The modeled epidemiological curve for the current year, including backcasts (historical predictions prior to the current week) and forecasts (future predictions after the current week). (C) Modeled epidemiological curves for all years, including fitted values in historical years (2004-2020) and forecasts for the current year (2021). (D) The weekly proportions of counties with at least one human case from historical years. (E) Daily temperatures in the current year compared to historical averages.

(Figure 1). First, new mosquito data collected since the previous forecast are obtained by querying the organization's surveillance database. (Step 1a). A mosquito data template is provided with the ArboMAP software, and a single CSV file containing all mosquito data for the current year is copied to the mosquito data folder in the ArboMAP RStudio project (Step 1b). Then, new environmental data are obtained from the ArboMAP GEE app (Step 2a). The results are downloaded as CSV files that are copied directly into the environmental data folder in the ArboMAP RStudio project and automatically ingested and combined by the software (Step 2b). The user can now start the ArboMAP application and modify default parameters using the GUI (Step 3a). In most cases, the same set of parameters are used to generate forecasts throughout an entire season and the only required

change is the date of the current forecast week. At this point, the run is initiated, and the modeling and report generation (Step 3b) are automated.

An earlier version of ArboMAP was first implemented by the South Dakota Department of Health in 2016, and the tool has been used there since. Following several years of collaboration with the developers, the Louisiana Department of Health began using ArboMAP independently in 2022. Southern Nazarene University collaborated with the Oklahoma City County Health Department to generate forecasts beginning in 2022. The Michigan Department of Health and Human Services began generating forecasts with ArboMAP in 2023. In South Dakota and Michigan, ArboMAP forecasts have been incorporated into online WNV dashboards and communicated with stakeholders via statewide email.

listservs. Because reported human cases are often delayed by weeks or months and observed mosquito abundance is a poor indicator of transmission risk, predictions from ArboMAP have been useful for highlighting WNV risk and targeting mosquito control and disease prevention activities prior to the seasonal peak in transmission.¹¹

Discussion

There is considerable interest in developing and testing new approaches for modeling and forecasting outbreaks of WNV and other infectious diseases.^{18–21} If these techniques are combined with improved systems for timely and accurate collection of relevant data, they have the potential to improve public health responses to outbreaks.²² The importance of having robust software to operationalize disease early warning systems has been recognized,²³ but this topic has not been widely addressed in the scientific literature.²⁴ The ArboMAP software system has been successfully implemented for routine forecasting of WNV. It can be used to forecast WNV in other locations where sufficient data are available and could also be adapted to work with other climate-sensitive vector-borne diseases.

The design of ArboMAP represents a compromise in which most of the time-consuming steps required for data processing and harmonization, model fitting and prediction, and presentation of the forecast results have been automated. Other aspects of the software, such as the connections to external databases, have been implemented as loose couplings and require additional manual steps for data acquisition. ArboMAP was developed as a client-side application that is installed on a laptop or desktop workstation rather than a cloud-based application that can be remotely accessed. These decisions make it practical for multiple public health institutions to independently use ArboMAP. Because of the security and privacy issues associated with health surveillance data, it was not feasible for us to develop a tight coupling solution for connecting ArboMAP with these systems. These issues also limited our options for accessing surveillance data through the cloud. Although the design of ArboMAP has facilitated its use in multiple states, opportunities remain to further automate the process of data acquisition and decrease the user effort required to generate forecasts.

Co-development with public health partners has been essential in designing the ArboMAP system. The current forecasting reports were informed by design and evaluation workshops held in 2021 and 2022. Key design principles included emphasizing visualizations over written text, creating stand-alone figures that can be copied to other reports or websites to communicate the forecasts, using consistent color schemes and formatting throughout the report, and adjusting the order so that the most important results can be found on the first few pages of the report. User feedback has also informed the development of the user-specified parameters and the design of the GUI. All ArboMAP code along with comprehensive documentation and artificial datasets for demonstration and testing are available on GitHub (<https://github.com/EcoGRAPH/ArboMAP>). Users can run the software with their own datasets or customize the code to meet specific needs. Potential changes could include integrating novel streams of data, incorporating different predictive modeling techniques, or modifying the information provided in the forecasting reports.

Author contributions

D.M.N. and M.C.W. contributed to the initial conception and design of the ArboMAP software. D.M.N. programmed the software. A.B., S.S., E.O., C.M.B.F., E.D., R.R., and R.B. used and tested the software and contributed to the redesign of the user interface, forecasting reports, and software documentation.

Supplementary material

Supplementary material is available at *JAMIA Open* online.

Funding

This work was supported by the National Aeronautics and Space Administration (NASA) Health and Air Quality Program grants NSSC19K1233 and NNX15AF74G to the University of Oklahoma, and Oklahoma NASA Established Program to Stimulate Competitive Research (EPSCoR)/Space Grant Consortium grant 80NSSC19M0058 to Southern Nazarene University.

Conflicts of interests

The authors have no competing interests to declare.

Data availability

Public health surveillance data on human West Nile virus cases and mosquito infection are collected and maintained by state departments of health and are not publicly distributable. Meteorological data from GridMET are publicly available at <https://www.climatologylab.org/gridmet.html>. A Google Earth Engine app for accessing and downloading the GridMET data summarized by county can be accessed at <https://dawneko.users.earthengine.app/view/arbomap-gridmet>.

References

1. Reisen WK. Ecology of West Nile virus in North America. *Viruses*. 2013;5(9):2079–2105. <https://doi.org/10.3390/v5092079>
2. Petersen LR, Brault AC, Nasci RS. West Nile virus: review of the literature. *JAMA*. 2013;310(3):308–315. <https://doi.org/10.1001/jama.2013.8042>
3. McDonald E, Mathis S, Martin SW, Erin Staples J, Fischer M, Lindsey NP. Surveillance for West Nile virus disease—United States, 2009–2018. *MMWR Surveill Summ*. 2021;70(1):1–15.
4. Nasci RS, Mutebi J-P. Reducing West Nile virus risk through vector management. *J Med Entomol*. 2019;56(6):1516–1521.
5. Kilpatrick AM, Pape WJ. Predicting human West Nile virus infections with mosquito surveillance data. *Am J Epidemiol*. 2013;178(5):829–835. <https://doi.org/10.1093/aje/kwt046>
6. Hahn MB, Monaghan AJ, Hayden MH, et al. Meteorological conditions associated with increased incidence of West Nile virus disease in the United States, 2004–2012. *Am J Trop Med Hyg*. 2015;92(5):1013–1022.
7. Wimberly MC, Lamsal A, Giacomo P, Chuang TW. Regional variation of climatic influences on West Nile virus outbreaks in the United States. *Am J Trop Med Hyg*. 2014;91(4):677–684.
8. Gorris ME, Randerson JT, Coffield SR, et al. Assessing the influence of climate on the spatial pattern of West Nile virus incidence in the United States. *Environ Health Persp*. 2023;131(4):047016.
9. Davis JK, Vincent G, Hildreth MB, Kightlinger L, Carlson C, Wimberly MC. Integrating environmental monitoring and mosquito surveillance to predict vector-borne disease: prospective forecasts of a West Nile virus outbreak. *PLoS Curr*. 2017. <https://doi.org/10.1371/journal.poc.0000000>

doi.org/10.1371/currents.outbreaks.90e80717c4e67e1a830f17feeaaaf85de

- 10. Davis JK, Vincent GP, Hildreth MB, Kightlinger L, Carlson C, Wimberly MC. Improving the prediction of arbovirus outbreaks: a comparison of climate-driven models for West Nile virus in an endemic region of the United States. *Acta Trop.* 2018;185:242-250.
- 11. Wimberly MC, Davis JK, Hildreth MB, Clayton JL. Integrated forecasts based on public health surveillance and meteorological data predict West Nile virus in a high-risk region of North America. *Environ Health Persp.* 2022;130(8):087006.
- 12. Wimberly MC, De Beurs KM, Loboda TV, Pan WK. Satellite observations and malaria: New opportunities for research and applications. *Trends Parasitol.* 2021;37(6):525-537. <https://doi.org/10.1016/j.pt.2021.03.003>
- 13. Wimberly MC, Nekorchuk DM, Kankanala RR. Cloud-based applications for accessing satellite earth observations to support malaria early warning. *Sci Data.* 2022;9(1):208-211.
- 14. Abatzoglou JT. Development of gridded surface meteorological data for ecological applications and modelling. *Int J Climatol.* 2013;33(1):121-131. <https://doi.org/10.1002/joc.3413>
- 15. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens Environ.* 2017;202:18-27.
- 16. Wood SN, Goude Y, Shaw S. Generalized additive models for large data sets. *J R Stat Soc Series C Appl Stat.* 2015;64(1):139-155.
- 17. Wood SN. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation, Version 1.9-0. 2023. Accessed August 10, 2023. <http://CRAN.R-project.org/package=mgcv>
- 18. Keyel AC, Gorris ME, Rochlin I, et al. A proposed framework for the development and qualitative evaluation of West Nile virus models and their application to local public health decision-making. *PLoS Negl Trop Dis.* 2021;15(9):e0009653.
- 19. Kobres P-Y, Chretien J-P, Johansson MA, et al. A systematic review and evaluation of Zika virus forecasting and prediction research during a public health emergency of international concern. *PLoS Negl Trop Dis.* 2019;13(10):e0007451.
- 20. Sylvestre E, Joachim C, Cecilia-Joseph E, et al. Data-driven methods for dengue prediction and surveillance using real-world and big data: a systematic review. *PLoS Negl Trop Dis.* 2022;16(1):e0010056.
- 21. Holcomb KM, Mathis S, Staples JE, et al. Evaluation of an open forecasting challenge to assess skill of West Nile virus neuroinvasive disease prediction. *Parasit Vectors.* 2023;16(1):1-13.
- 22. Morgan O. How decision makers can use quantitative approaches to guide outbreak responses. *Philos Trans R Soc Lond B Biol Sci.* 2019;374(1776):20180365.
- 23. George DB, Taylor W, Shaman J, et al. Technology to advance infectious disease forecasting for outbreak management. *Nat Commun.* 2019;10(1):3932-3934.
- 24. Merkord CL, Liu Y, Mihretie A, et al. Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system. *Malar J.* 2017;16(1):89.