The Macroscope Meets the Microscope: Integrating Earth Science Data with Disease Surveillance for Outbreak Forecasting

Michael C. Wimberly

Department of Geography and Environmental Sustainability

University of Oklahoma, Norman OK USA

COLLEGE OF ATMOSPHERIC AND GEOGRAPHIC SCIENCES
DEPARTMENT OF GEOGRAPHY
AND ENVIRONMENTAL SUSTAINABILITY
The UNIVERSITY of OKLAHOMA

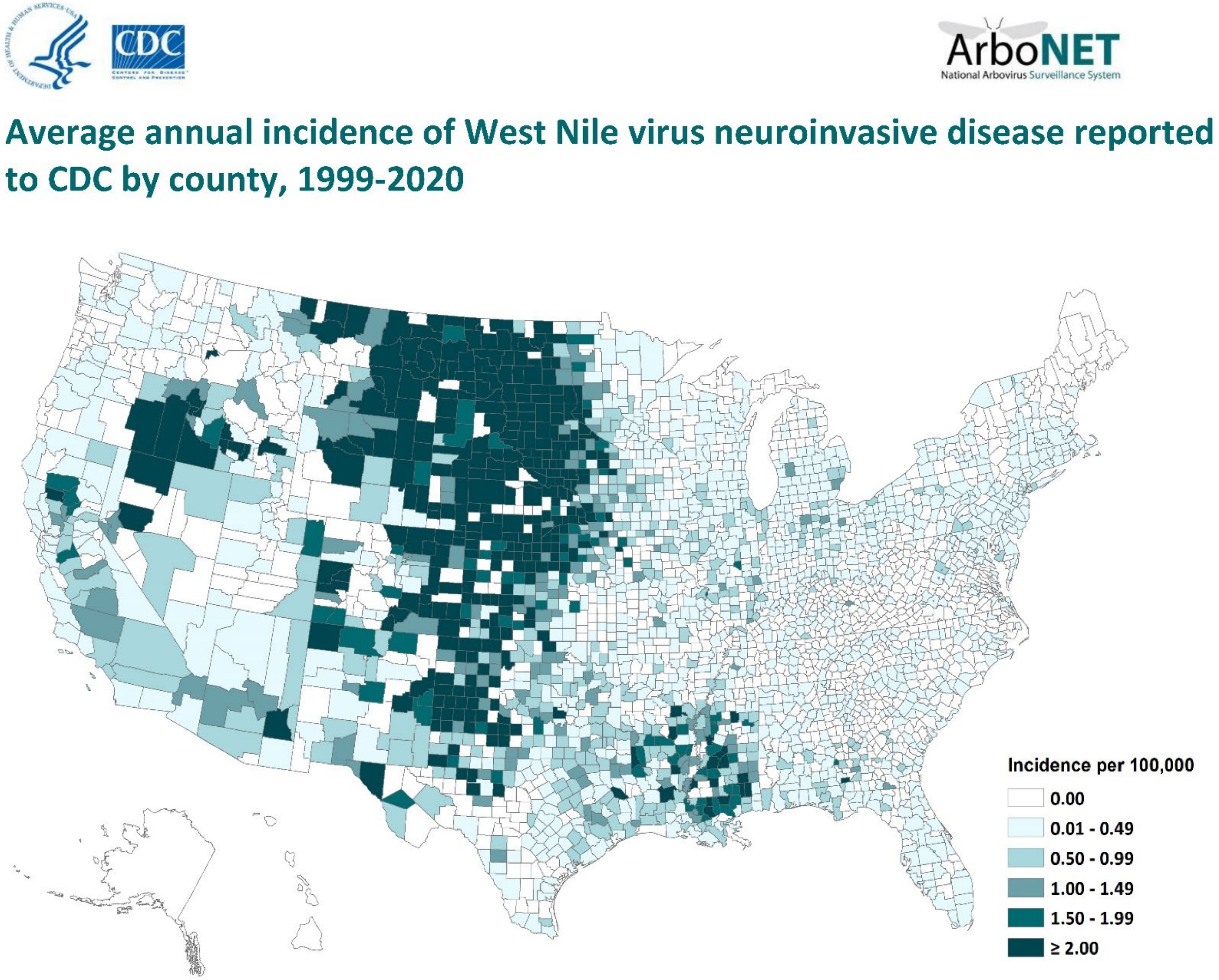
This talk will focus on West Nile virus forecasting

- A brief **introduction** to West Nile virus
- **Background research** into the environmental determinants of West Nile virus in South Dakota
- A **modeling framework** for West Nile virus forecasting
- **Implementation** a West Nile virus forecasting system
- **Validation** of West Nile virus forecasts in South Dakota
- **Extensions** to other states

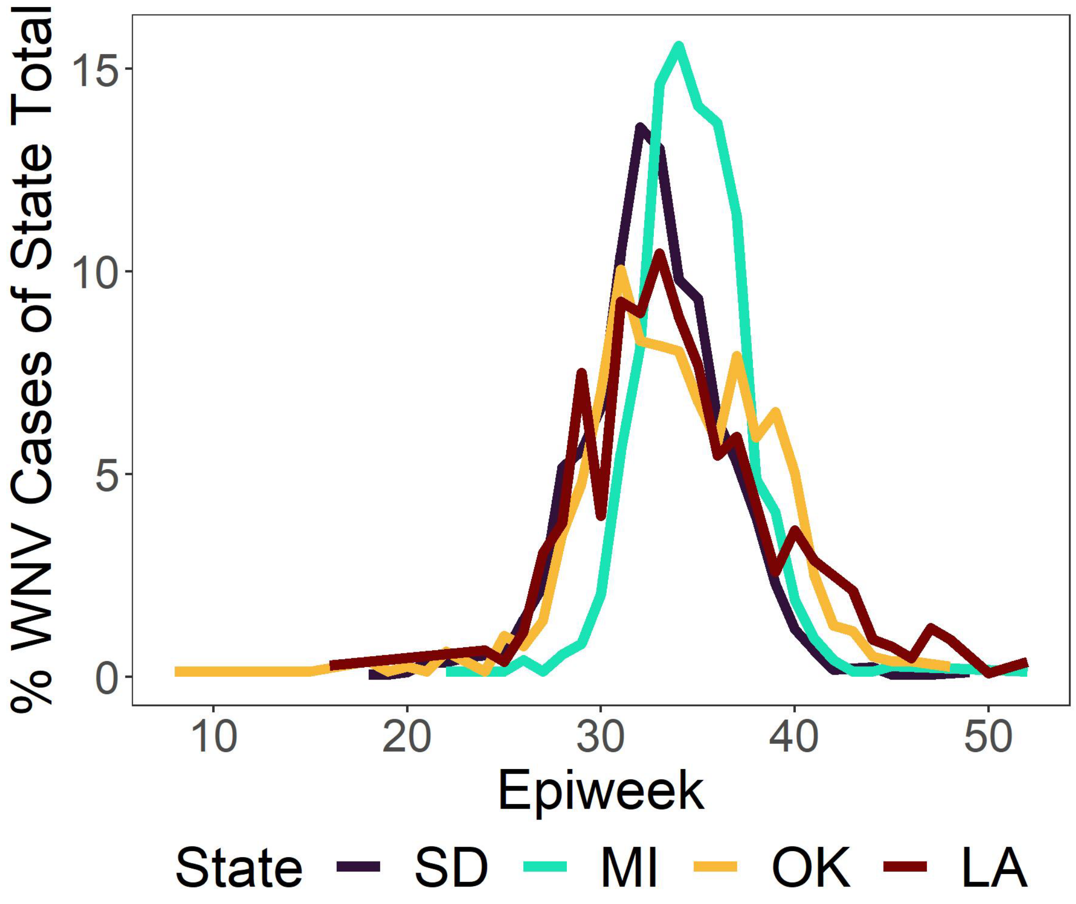
Introduction

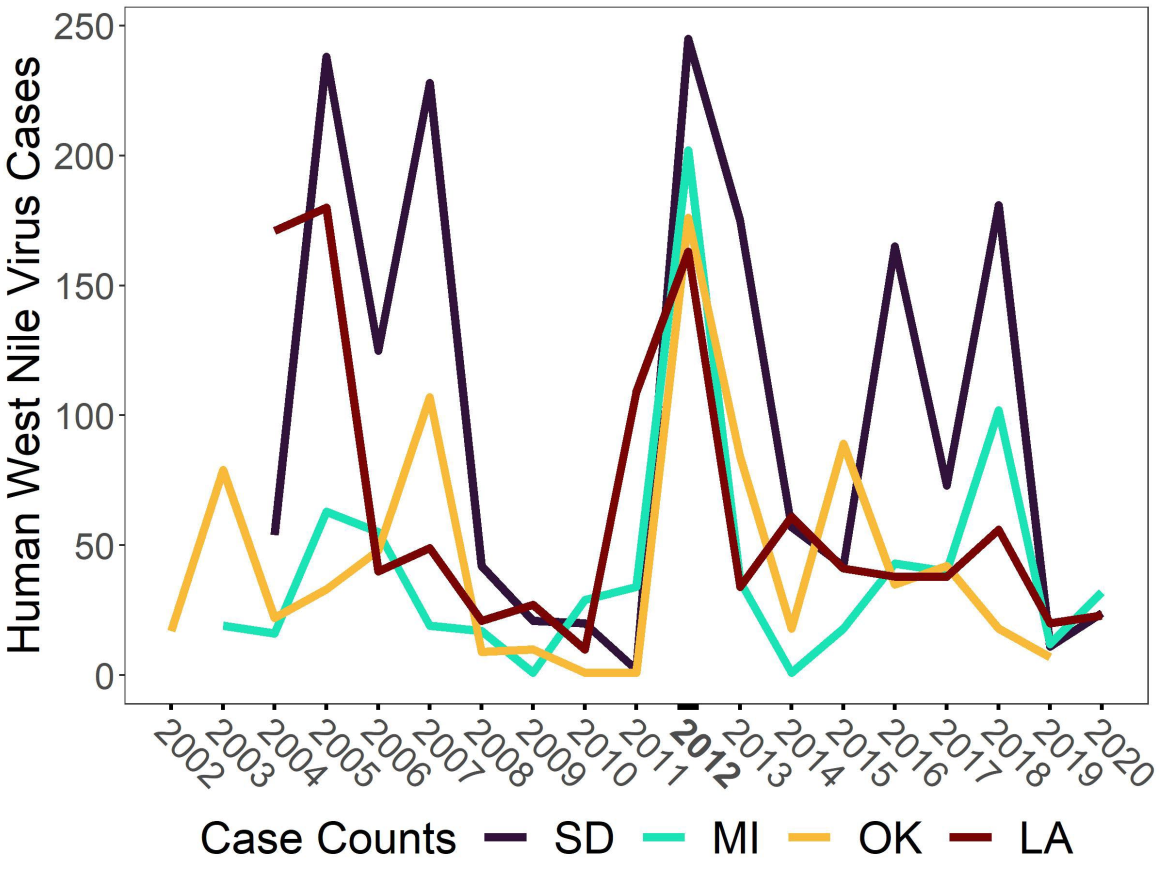
West Nile virus in the US

- Arrived in the US in 1999, spread west across N. America by 2004
- Most common mosquito-borne disease in the U.S.
  - ~2,400 cases per year
    - Highly variable – range: 21 to 9,862
    - 70-80% of infections asymptomatic
    - Severe disease in ~1 in 150
- Wild birds are the primary reservoir hosts
- Humans are dead-end hosts

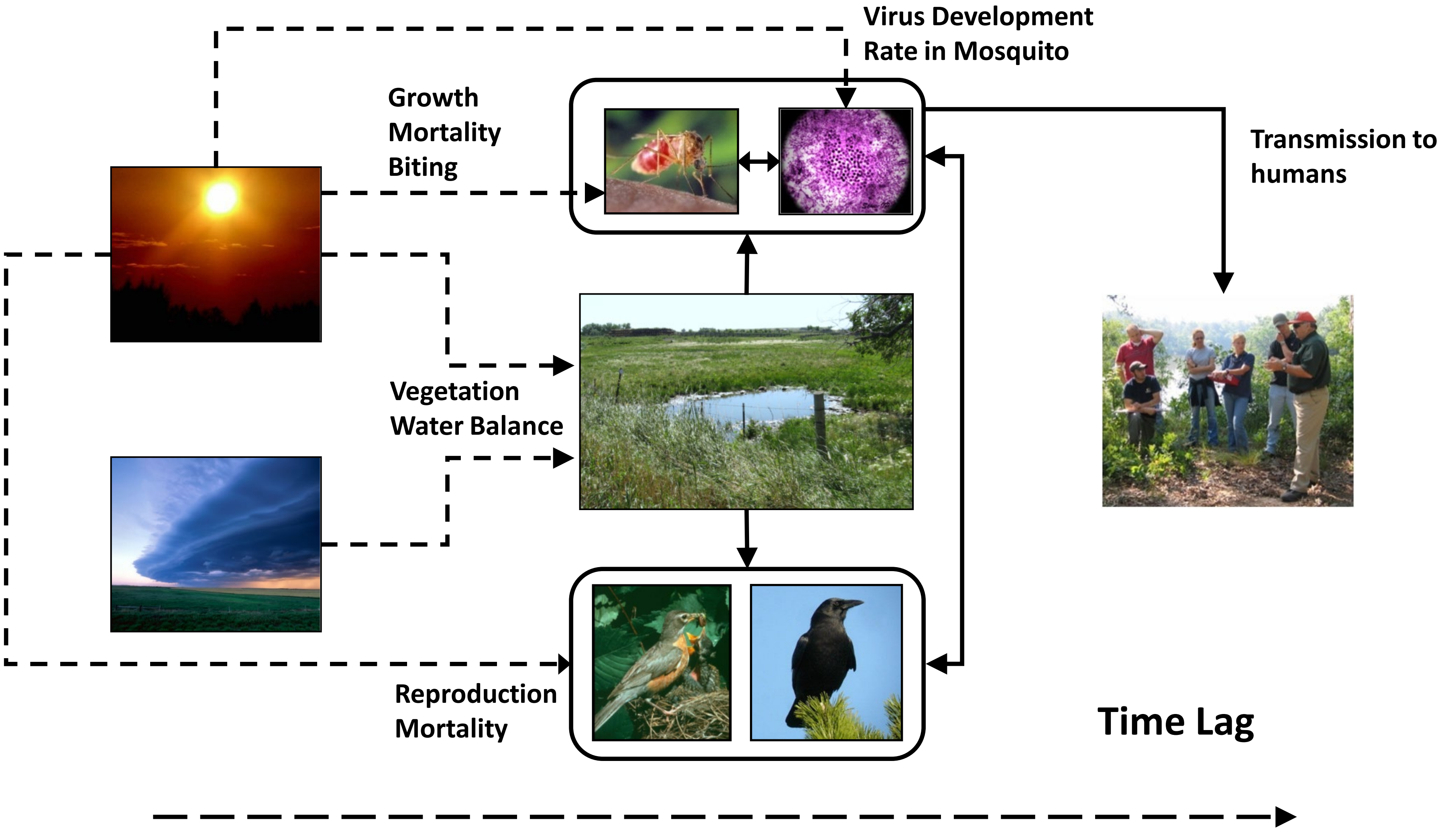


Seasonal and interannual variation of WNV cases





Environmental factors influence WNV transmission through multiple pathways.



Background Research

Two predominant mosquito species in South Dakota

Aedes vexans

Inland Floodwater Mosquito

Eggs laid in flood-prone areas and hatch simultaneously when inundated

Nuisance mosquito

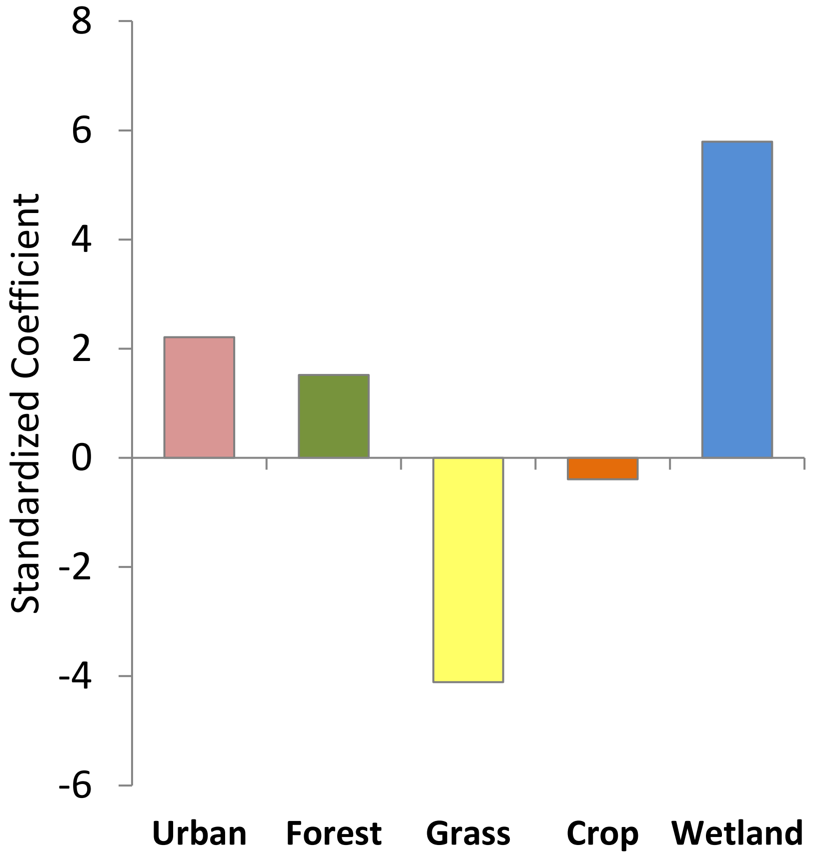
Culex tarsalis

Western Encephalitis Mosquito

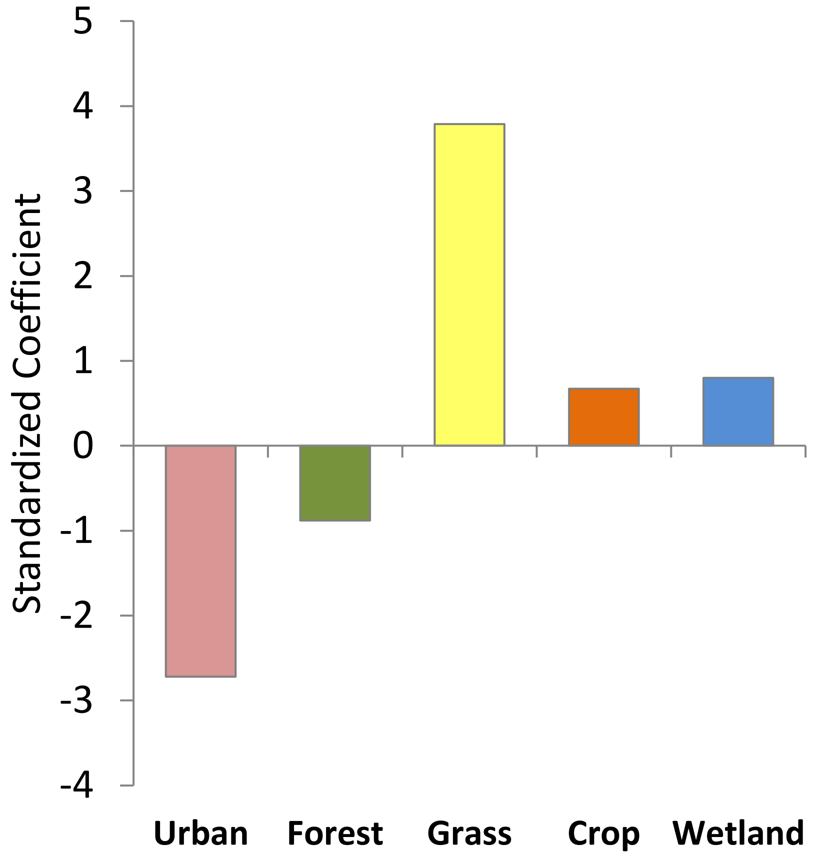
Standing water mosquito, breeds in natural and anthropogenic habitats with high organic content

Vector of West Nile virus

Aedes vexans is positively associated with wetlands and negatively associated with grasslands



Culex tarsalis is positively associated with grasslands and negatively associated with developed areas



Weather and Land Cover Influences on Mosquito Populations in Sioux Falls, South Dakota

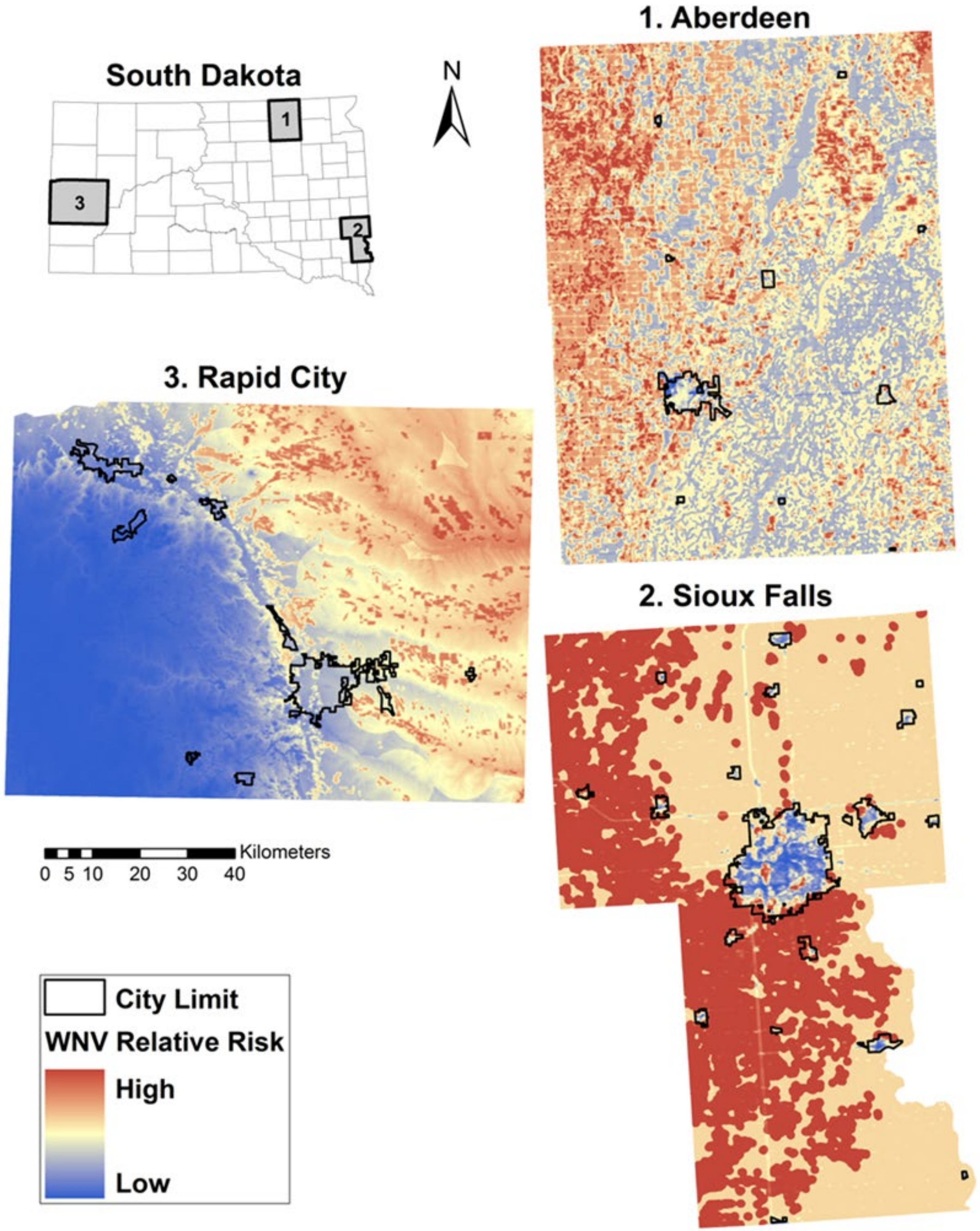
TING-WU CHUANG,[1,2] MICHAEL B. HILDRETH,[3,4] DENISE L. VANROEKEL,[5] AND MICHAEL C. WIMBERLY[1]

J. Med. Entomol. 48(3): 669–679 (2011); DOI: 10.1603/ME10246

Relationships between *Culex tarsalis* and land cover affect the geographic distribution of human West Nile virus cases.

Higher WNV risk in grasslands, at lower elevations, and on poorly drained soils.

Lower WNV risk in forests, at higher elevations, and in cities



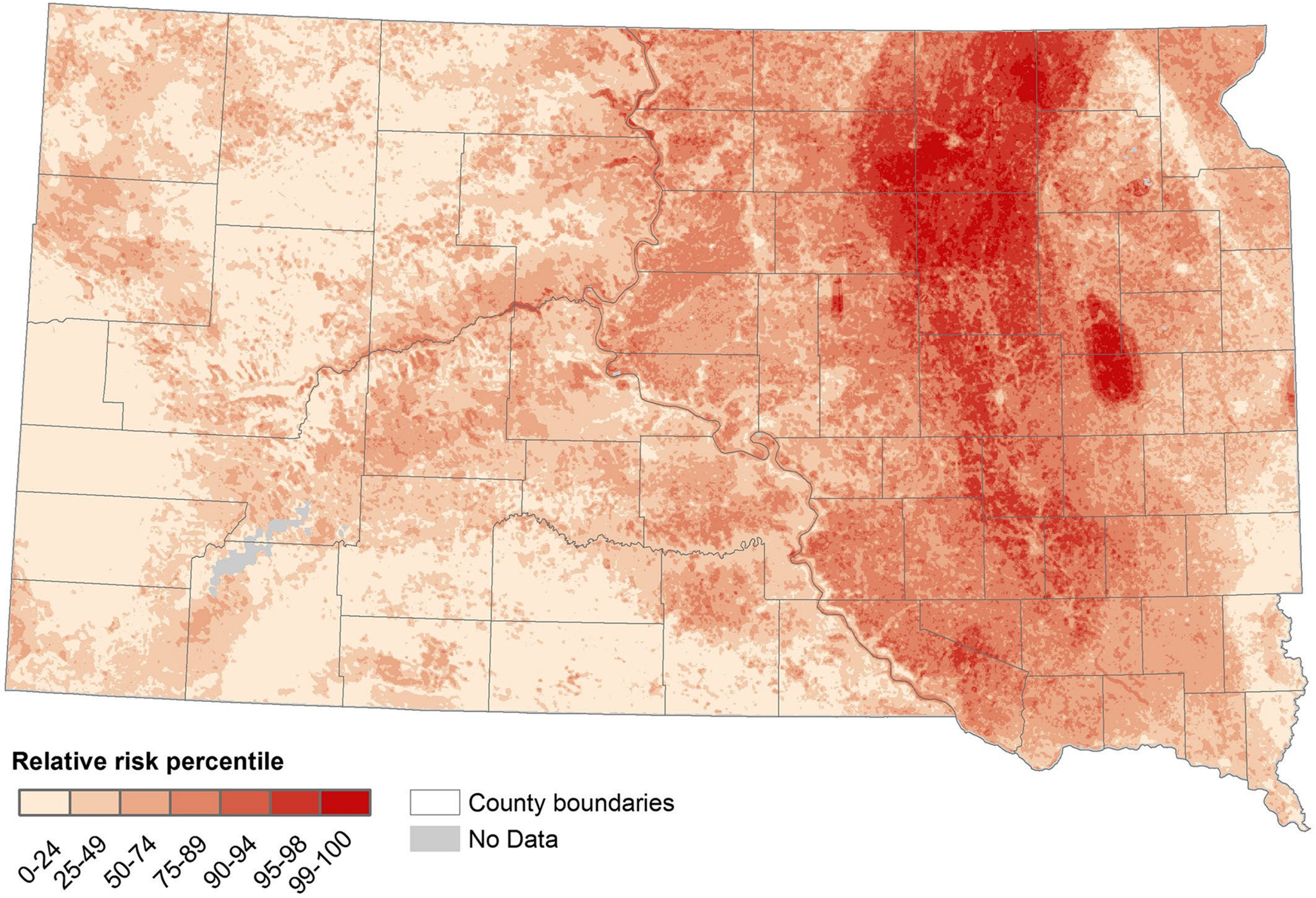
Am. J. Trop. Med. Hyg., 86(4), 2012, pp. 724–731
doi:10.4269/ajtmh.2012.11-0515

Landscape-Level Spatial Patterns of West Nile Virus Risk in the Northern Great Plains

Ting-Wu Chuang,* Christine W. Hockett, Lon Kightlinger, and Michael C. Wimberly
Geographic Information Science Center of Excellence, South Dakota State University, Brookings, South Dakota; South Dakota Department of Health, Pierre, South Dakota

We were able to map the statewide patterns of WNV risk using environmental variables from multiple sources:

- Elevation (digital elevation model)
- Humidity (interpolated climate data)
- Precipitation (interpolated climate data)
- Wetness index (MODIS satellite imagery)
- Land cover classification (Landsat satellite imagery)
- Soil drainage (soil survey data)



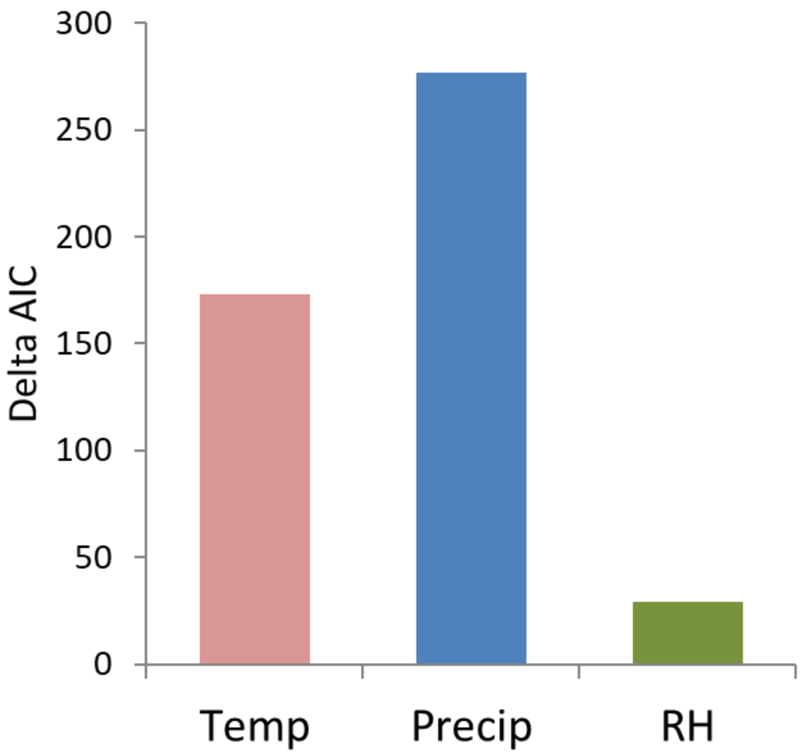
GeoHealth

Research Article | Open Access

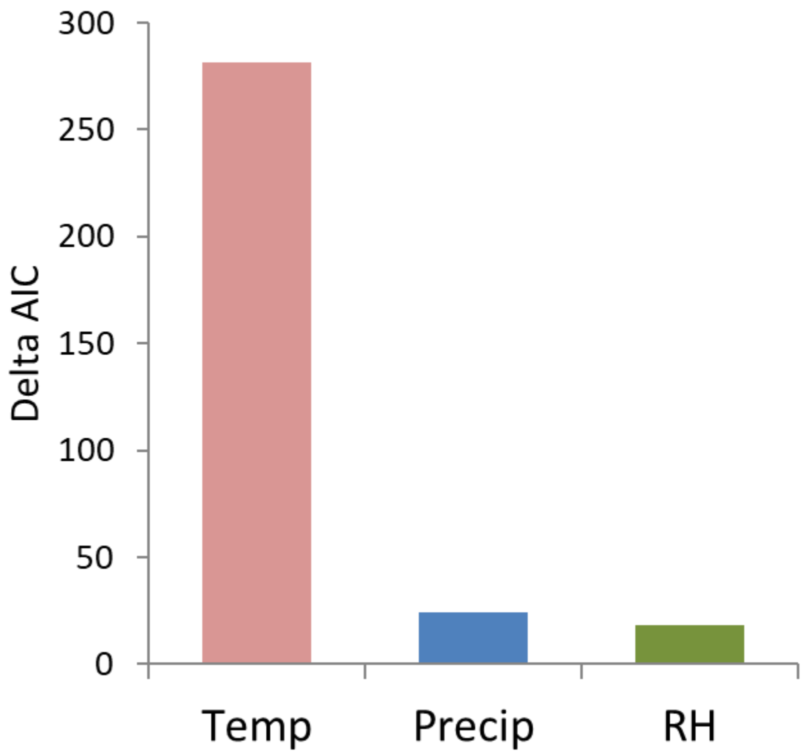
Identifying Environmental Risk Factors and Mapping the Distribution of West Nile Virus in an Endemic Region of North America

A. Hess, J. K. Davis, M. C. Wimberly

Aedes vexans was positively influenced by current temperature and lagged precipitation (2-3 weeks in the past)



Culex tarsalis was positively influenced by current temperature and lagged temperature (1-2 weeks in the past)



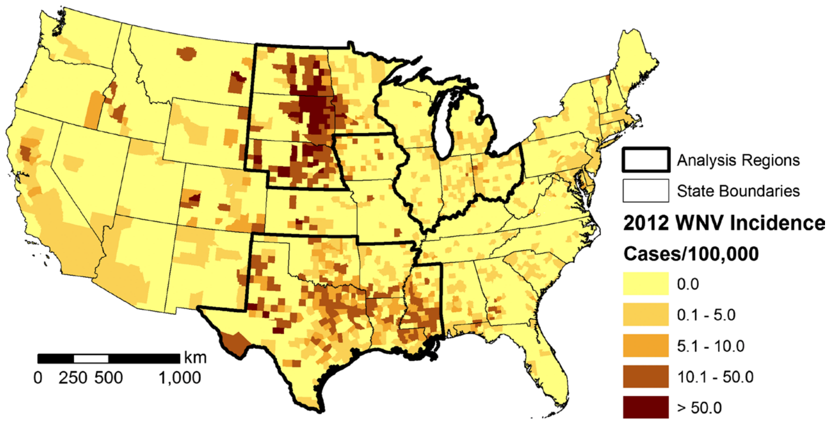
Weather and Land Cover Influences on Mosquito Populations in Sioux Falls, South Dakota

TING-WU CHUANG,[1,2] MICHAEL B. HILDRETH,[3,4] DENISE L. VANROEKEL,[5] AND MICHAEL C. WIMBERLY[1]

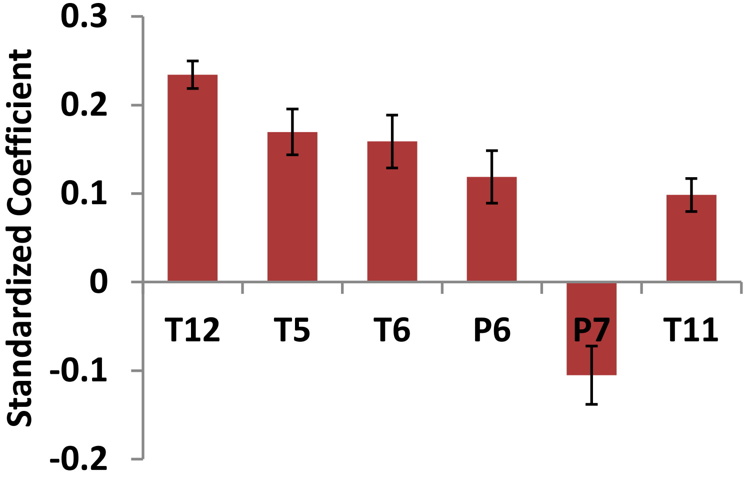
J. Med. Entomol. 48(3): 669–679 (2011); DOI: 10.1603/ME10246

WNV outbreaks are associated with positive temperature anomalies during the WNV season and the preceding spring/winter.

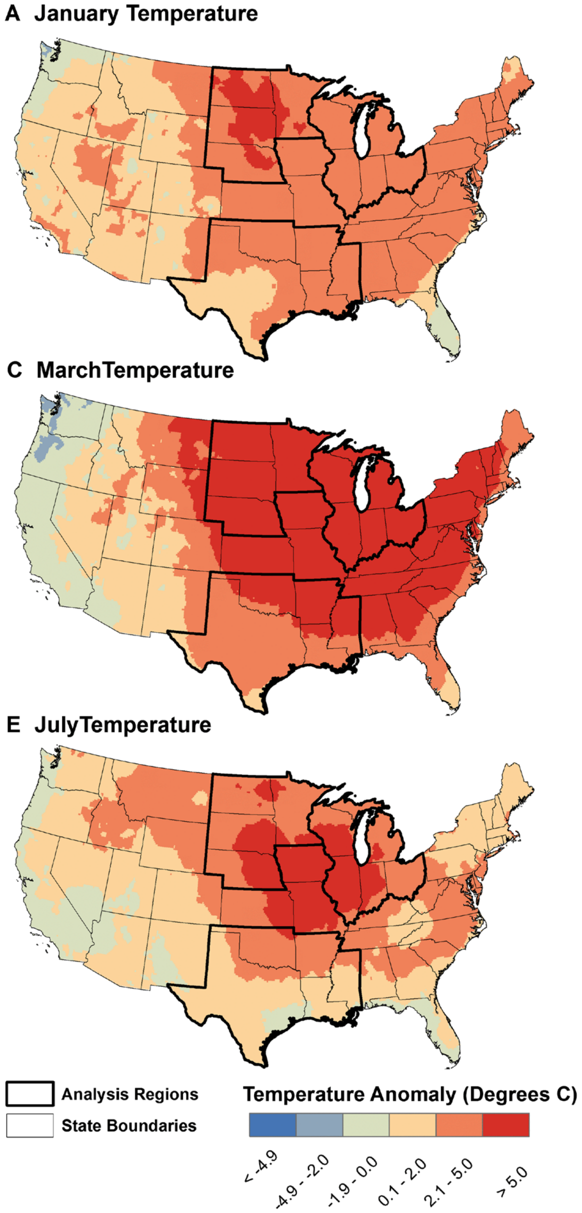
2012 WNV Incidence



PLS Coefficients for 2004-2012 outbreak model



2012 Climate Anomalies



Am. J. Trop. Med. Hyg., 91(4), 2014, pp. 677–684
doi:10.4269/ajtmh.14-0239

Regional Variation of Climatic Influences on West Nile Virus Outbreaks in the United States

Michael C. Wimberly,* Aashis Lamsal, Paolla Giacomo, and Ting-Wu Chuang
Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, South Dakota; Department of Parasitology and Center for International Tropical Medicine, Taipei Medical University, Taipei, Taiwan

Modeling Framework

Forecasting and surveillance of West Nile virus risk

Early warning would give time for intervention efforts

- Risk varies geographically and over time
- Target prevention messages and vector control

Human case surveillance

- Reporting delayed by weeks or months
- Lagging indicator of risk during the WNV season

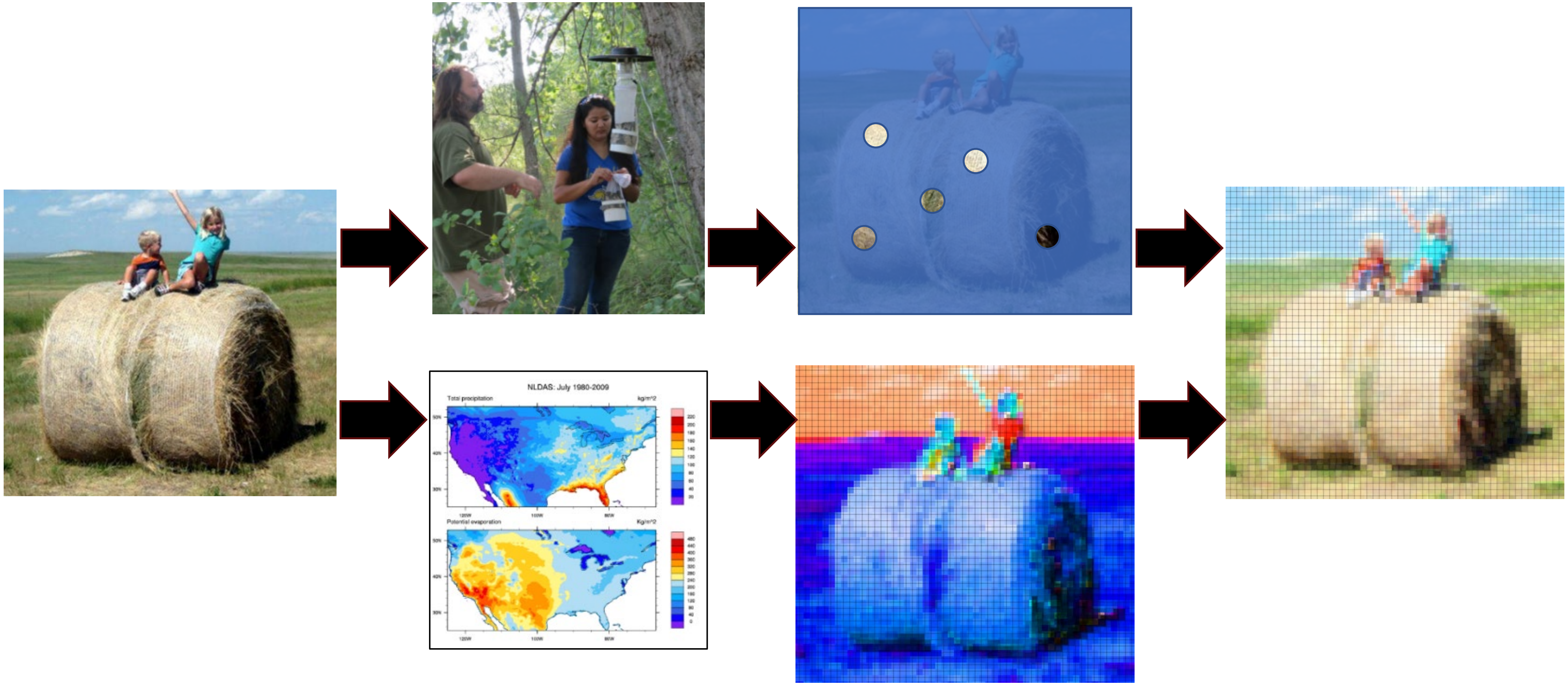
Mosquito surveillance

- Conducted weekly throughout the WNV season
- WNV-infected mosquito pools are associated with transmission risk

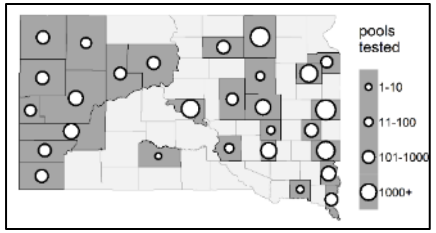
Environmental monitoring

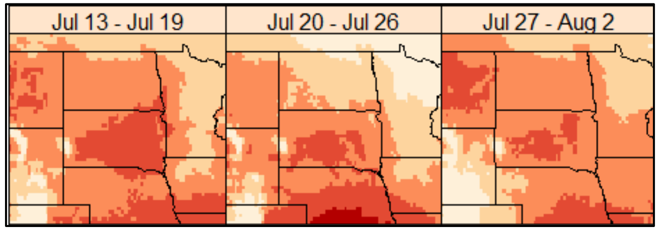
- Data on temperature, humidity, and precipitation available daily
- Weather affects mosquito and bird populations and virus development in mosquitoes

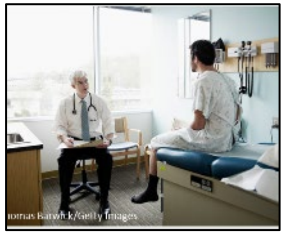
Our approach combines mosquito surveillance data with environmental monitoring data to predict human WNV cases



Forecasting models are calibrated with data from preceding years and used to predict the weekly probability of human WNV cases by county in the current year.







Mosquito infection data (2004-2021)

Mosquito infection data (2022)

GridMET daily weather (2004-2021)

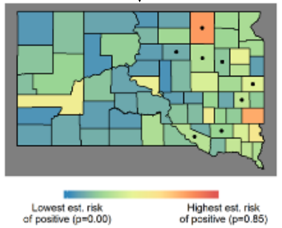
GridMET daily weather (2022)

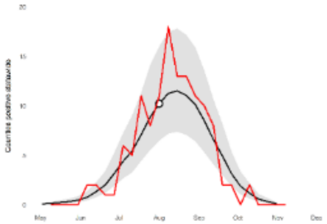
Human WNV cases (2004-2021)

Mixed-effects model of mosquito infection growth

Distributed lag functions

"Big additive model" of human WNV cases by county and week





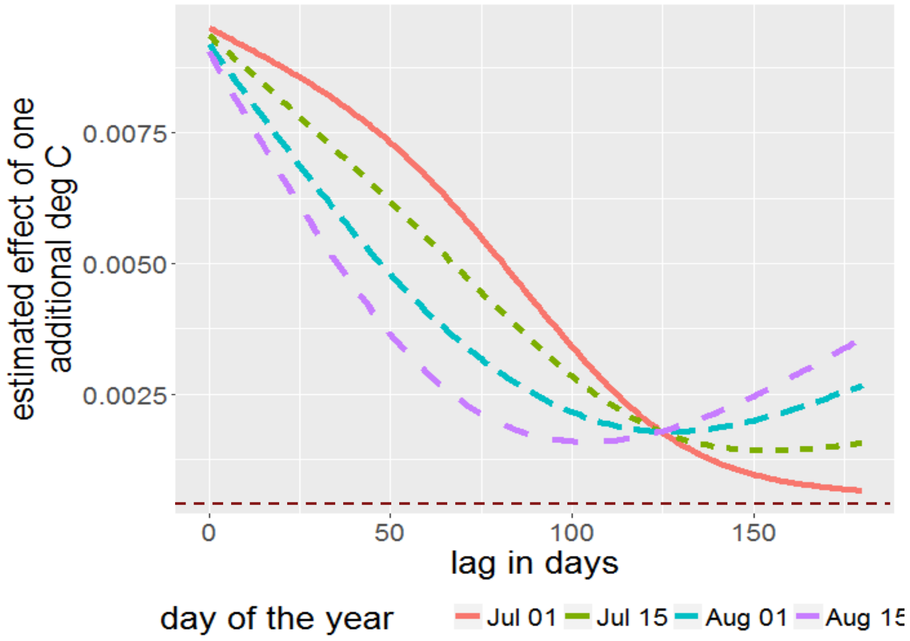
Calibration

Prediction

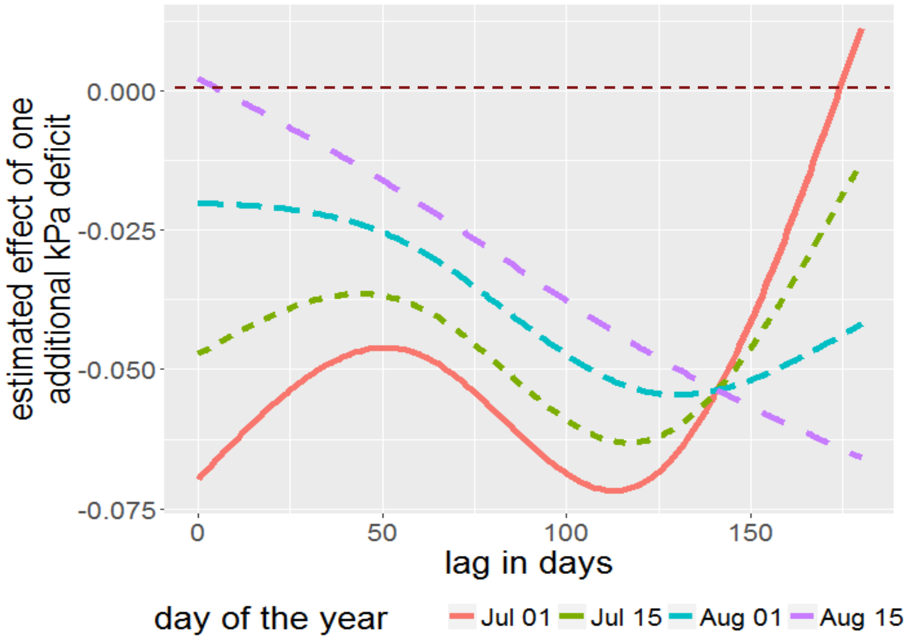
Distributed lags identify lagged meteorological effects over the past six months and can vary over the course of the transmission season.

- Temperature effects are strongest at the shortest lags, with longer-term effects (30-90 day lags) decreasing over the transmission season.
- Effects of vapor pressure deficit are strongest in June and shift to longer lags during the transmission season.

Temperature

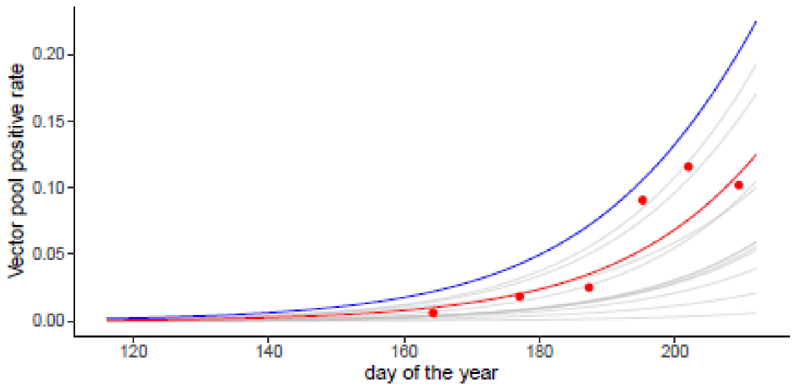


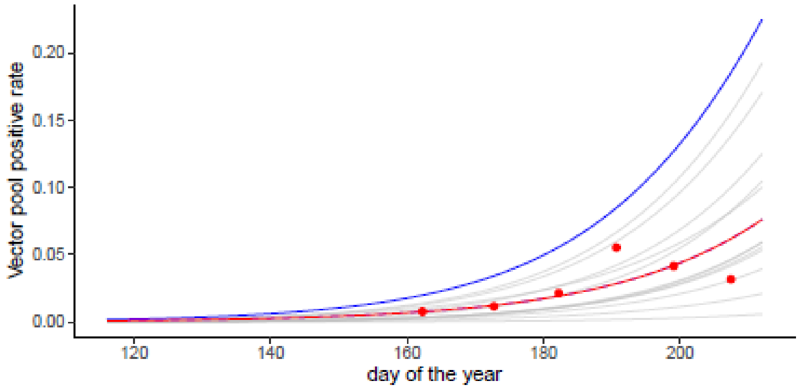
Vapor Pressure Deficit



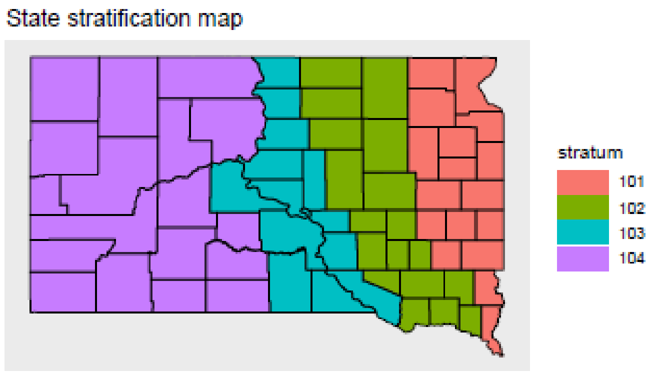
Seasonal increases in mosquito infection rates are modeled using a logistic growth function.

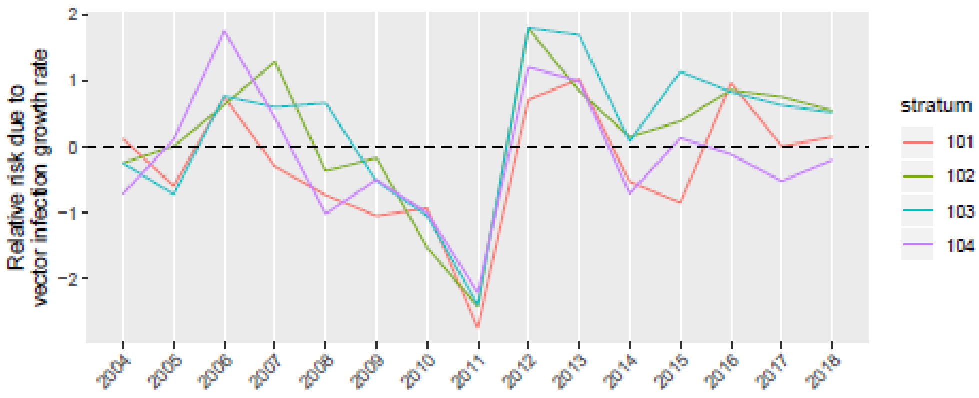
Interannual variation in these rates is associated with seasonal outbreak size.





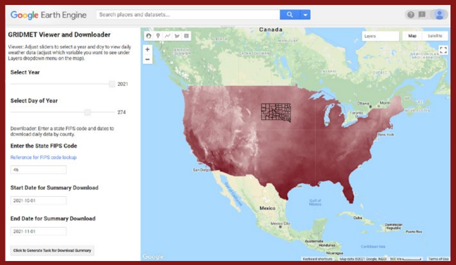
Spatial stratification captures geographic patterns of human cases



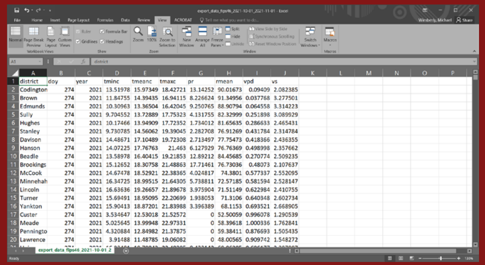


Implementation

Arbovirus Monitoring and Prediction (ArboMAP) system

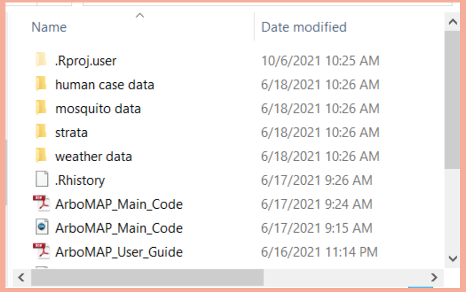


Web-based Google Earth Engine App

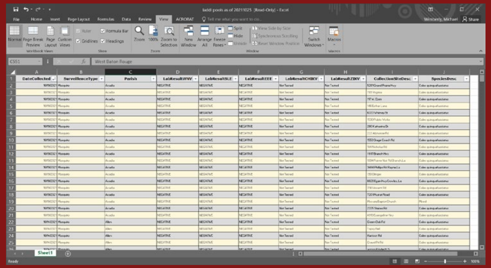


Update current-year weather data

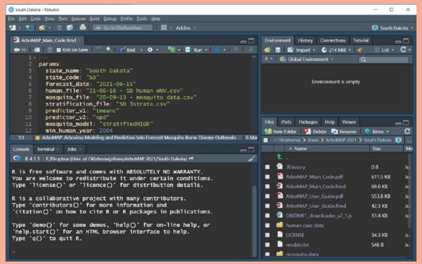
Fit predictive models with human, mosquito, and weather data from previous years



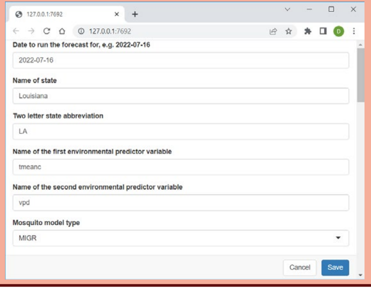
State surveillance databases



Update current-year mosquito data



Run ArboMAP script in Rstudio via web browser interface



Output

Designed to be used by state health departments for operational WNV forecasting

The UNIVERSITY of OKLAHOMA

Rationale for the design of ArboMAP

Completely automated, menu-driven, web-based system	Semi-automated system running on local workstations (ArboMAP)	Completely manual using software on local workstations

Completely automated, menu-driven, web-based system

- Easiest and quickest to use
- Centralized maintenance

- High cost for development and maintenance
- Data sharing limitations
- System integration challenges

Semi-automated system running on local workstations (ArboMAP)

- Considerably faster than a fully manual approach
- Implementable on a variety of computers
- Highly customizable

- Troubleshooting on diverse systems
- Requires some manual steps
- R and RStudio can be challenging for some users

Completely manual using software on local workstations

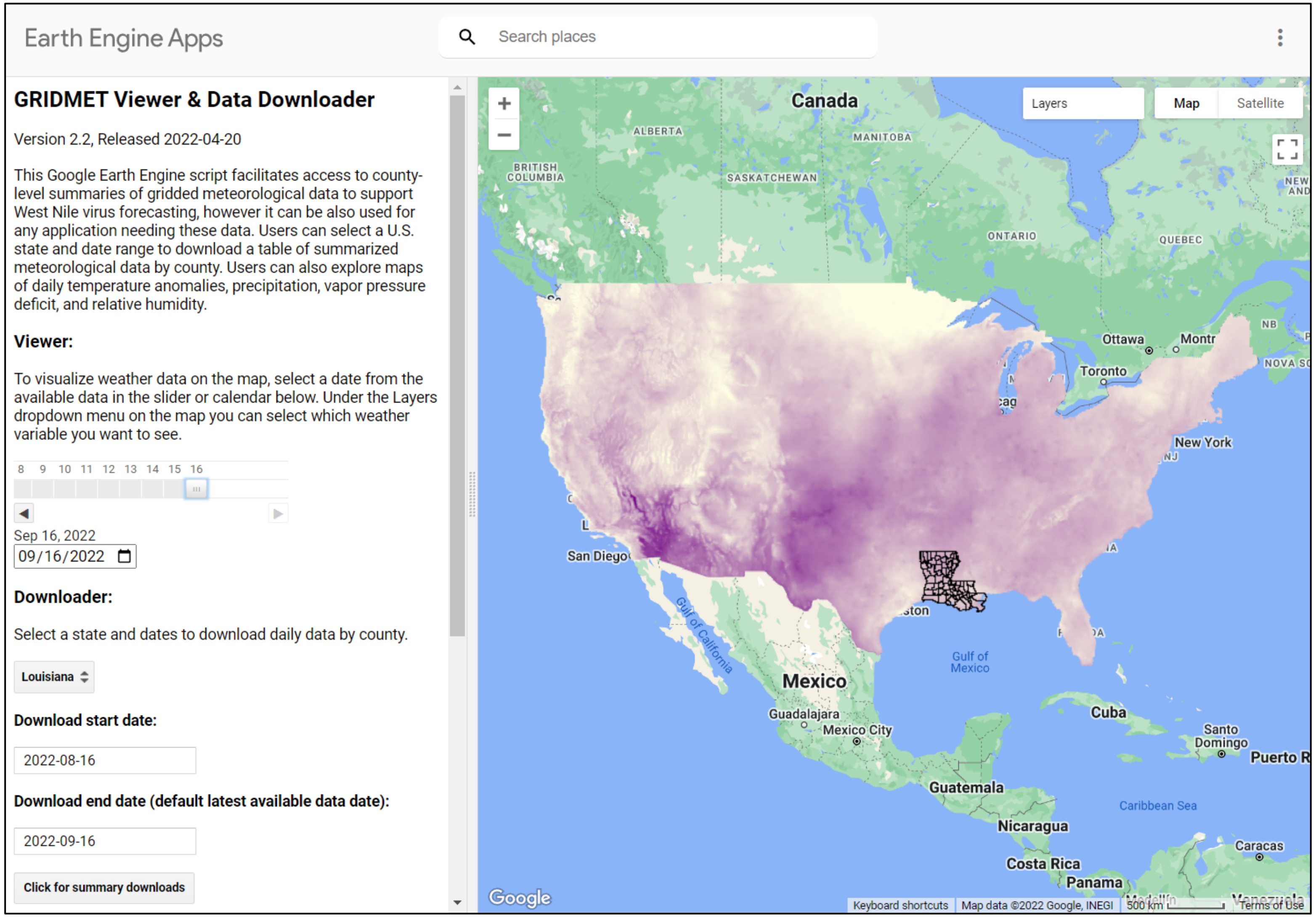
- Minimal development costs
- Leverages existing computer resources

- Time and cost of data analysis
- Lack of consistency
- Difficult to replicate

A Google Earth Engine application provides access to gridded meteorological data

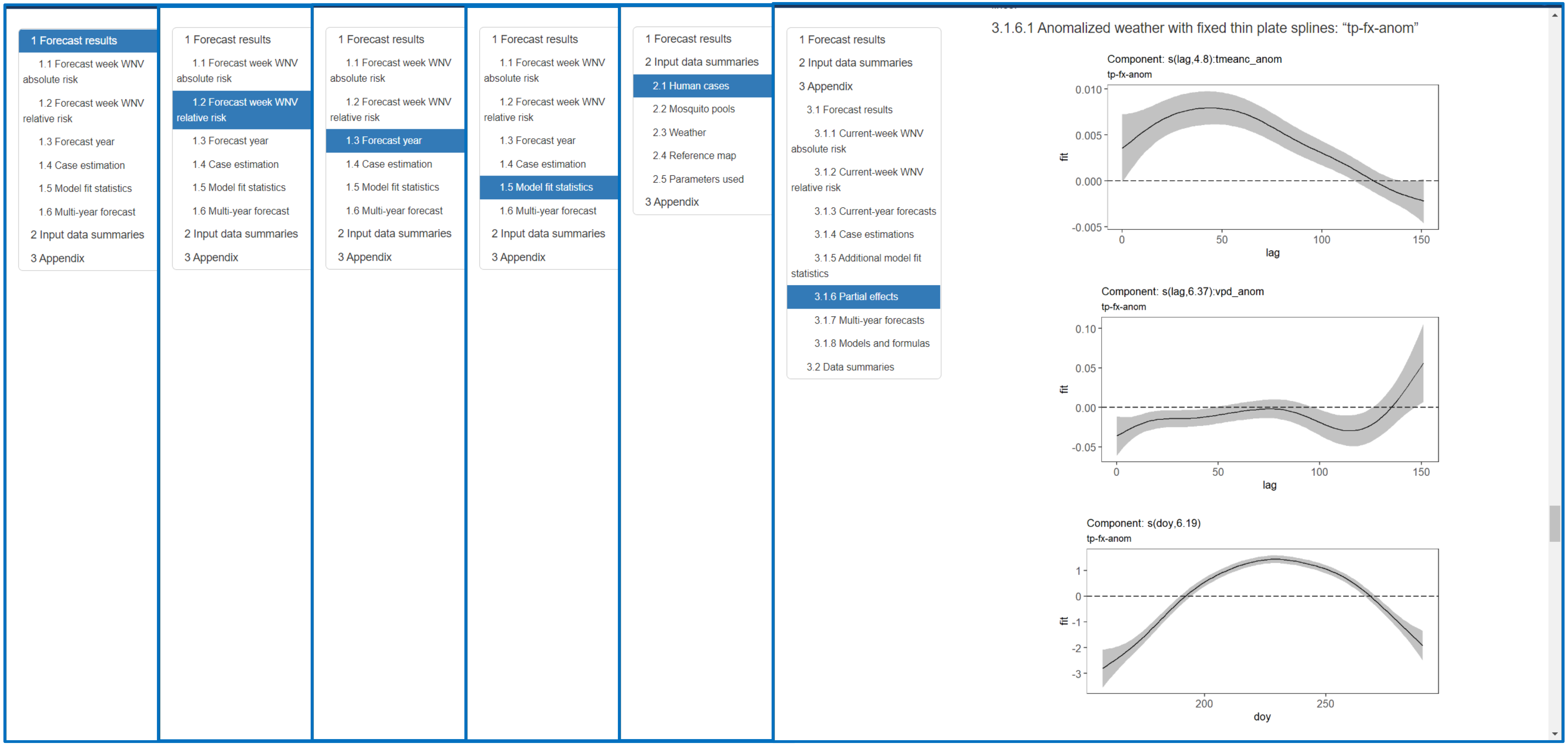
Automatically generates county-level tabular summaries that are read in by ArboMAP

district	fips	doy	year	tminc	tmeanc	tmaxc	pr	rmean	vpd	vs
Codington	46029	241	2022	10.6508	18.00361	25.35643	0	51.18069	1.199598	7.039887
Brown	46013	241	2022	10.22699	18.61046	26.99392	0	55.86343	1.253458	5.977255
Edmunds	46045	241	2022	10.629	18.8503	27.07159	0	53.58873	1.304932	5.862908
Sully	46119	241	2022	11.43182	20.32335	29.21488	0	49.80813	1.601839	5.110535
Hughes	46065	241	2022	11.62862	20.54096	29.4533	0	49.28324	1.642936	4.588199
Stanley	46117	241	2022	12.29954	21.22396	30.14838	0	47.53138	1.761037	4.670611
Davison	46035	241	2022	11.83324	21.04885	30.26445	0	42.61989	1.846022	5.144336
Hanson	46061	241	2022	11.56137	20.84423	30.1271	0	43.06917	1.822233	5.342723
Beadle	46005	241	2022	10.53975	19.84745	29.15514	0	48.85021	1.606574	5.598668
Brookings	46011	241	2022	10.31958	18.40501	26.49044	0	46.8909	1.384374	6.573551
McCook	46087	241	2022	10.85209	19.9863	29.12051	0	43.90171	1.696626	5.883674
Minnehaha	46099	241	2022	10.9022	19.63794	28.37367	0	43.29024	1.630269	6.182807
Lincoln	46083	241	2022	10.88606	20.14551	29.40495	0	45.47171	1.692342	5.475039
Turner	46125	241	2022	11.07187	20.48755	29.90323	0	44.76177	1.757634	5.494321
Yankton	46135	241	2022	11.28789	20.80838	30.32888	0	46.63622	1.763626	4.827213
Custer	46033	241	2022	11.56983	19.74449	27.91915	0	42.99915	1.667795	2.632135
Meade	46093	241	2022	12.02944	20.50135	28.97325	0	41.68096	1.782076	4.149048
Pennington	46103	241	2022	11.37736	19.89441	28.41146	0	43.34108	1.711369	3.528268
Lawrence	46081	241	2022	9.604772	17.34314	25.0815	0	43.83572	1.394663	4.252016
Union	46127	241	2022	10.48083	20.3885	30.29616	0	53.79768	1.614861	5.096313
Clay	46027	241	2022	10.9793	20.80764	30.63597	0	49.38417	1.740207	5.041141
Grant	46051	241	2022	10.98471	18.61747	26.25024	0	50.82275	1.278762	7.350341



https://dawneko.users.earthengine.app/view/arbomap-gridmet

Results are displayed as text, charts, and maps



Validation

Models were validated by making retrospective forecasts from 2016-2019

ehp Environmental Health Perspectives

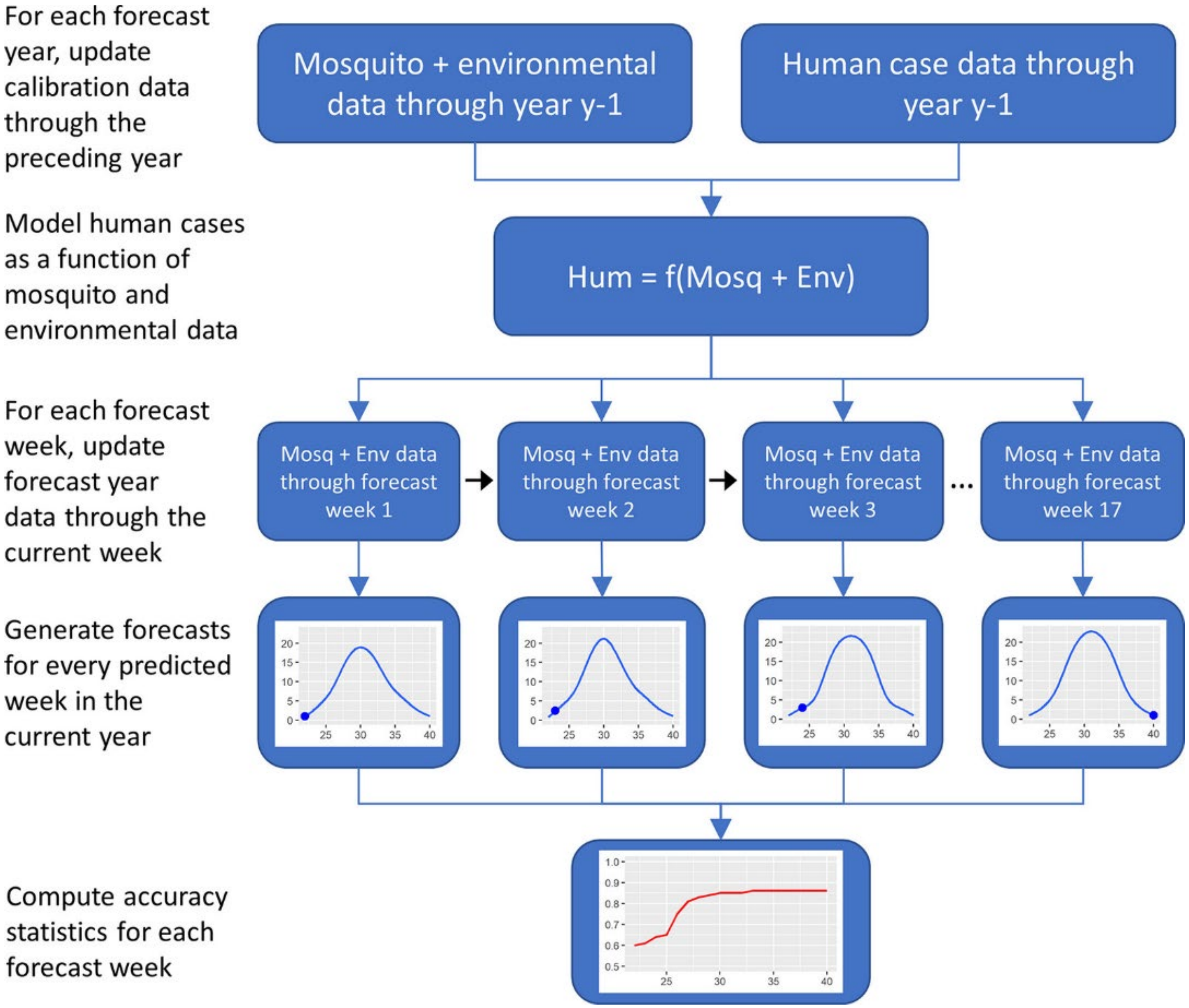
Open Access

Vol. 130, No. 8 | Research

Integrated Forecasts Based on Public Health Surveillance and Meteorological Data Predict West Nile Virus in a High-Risk Region of North America

Michael C. Wimberly, Justin K. Davis, Michael B. Hildreth, and Joshua L. Clayton

Published: 16 August 2022 | CID: 087006 | https://doi.org/10.1289/EHP10287

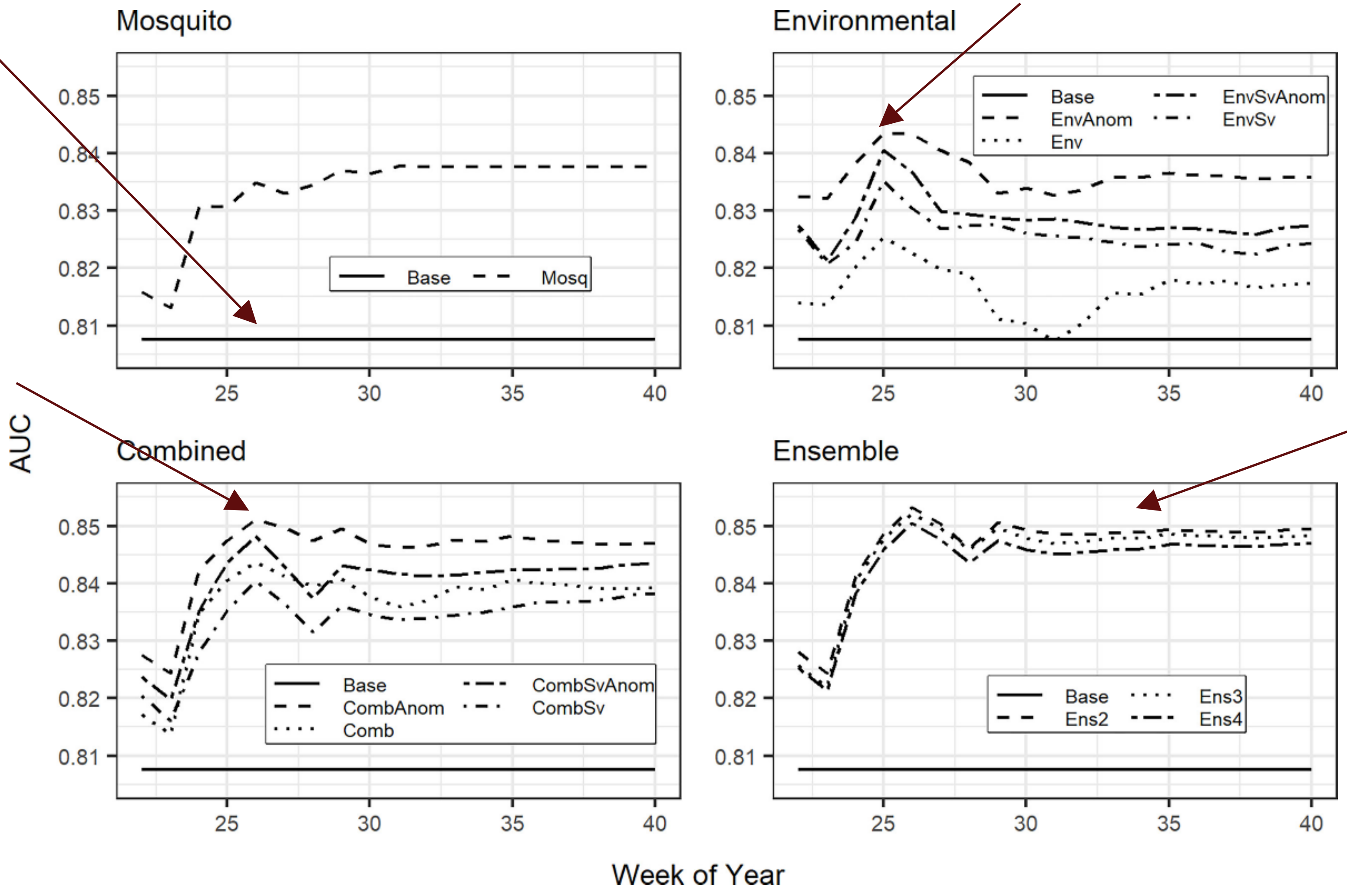


The baseline model included county-level means and a seasonal trend.

Environmental models outperformed mosquito models in the early season

Combined models outperformed models using only mosquito or environmental data and reached peak accuracy early in the WNV season.

Ensembles of combined models had the highest accuracies.

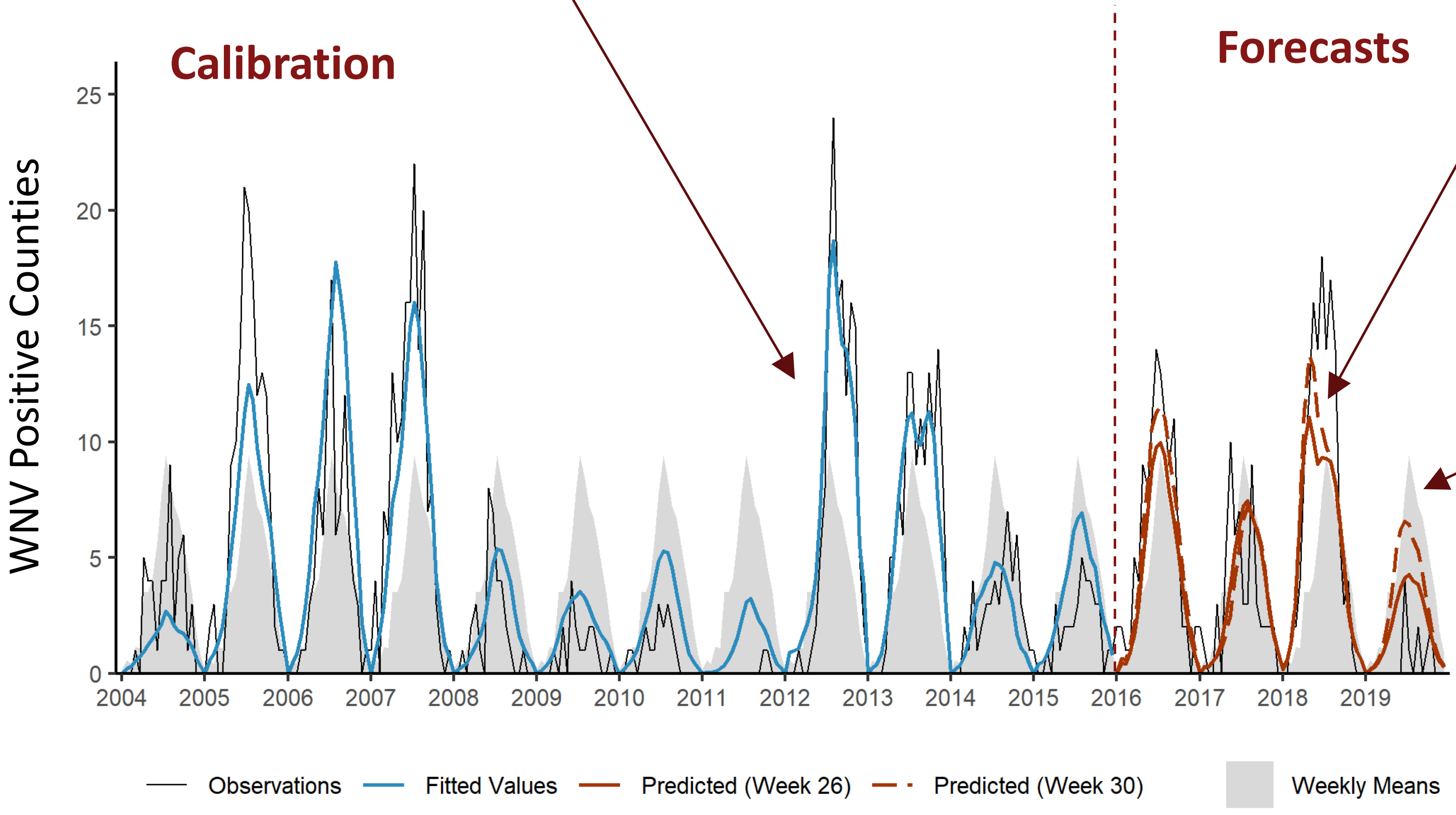


South Dakota Validation (2016-2019)

The best model, fitted to historical data, captures seasonal and interannual variations

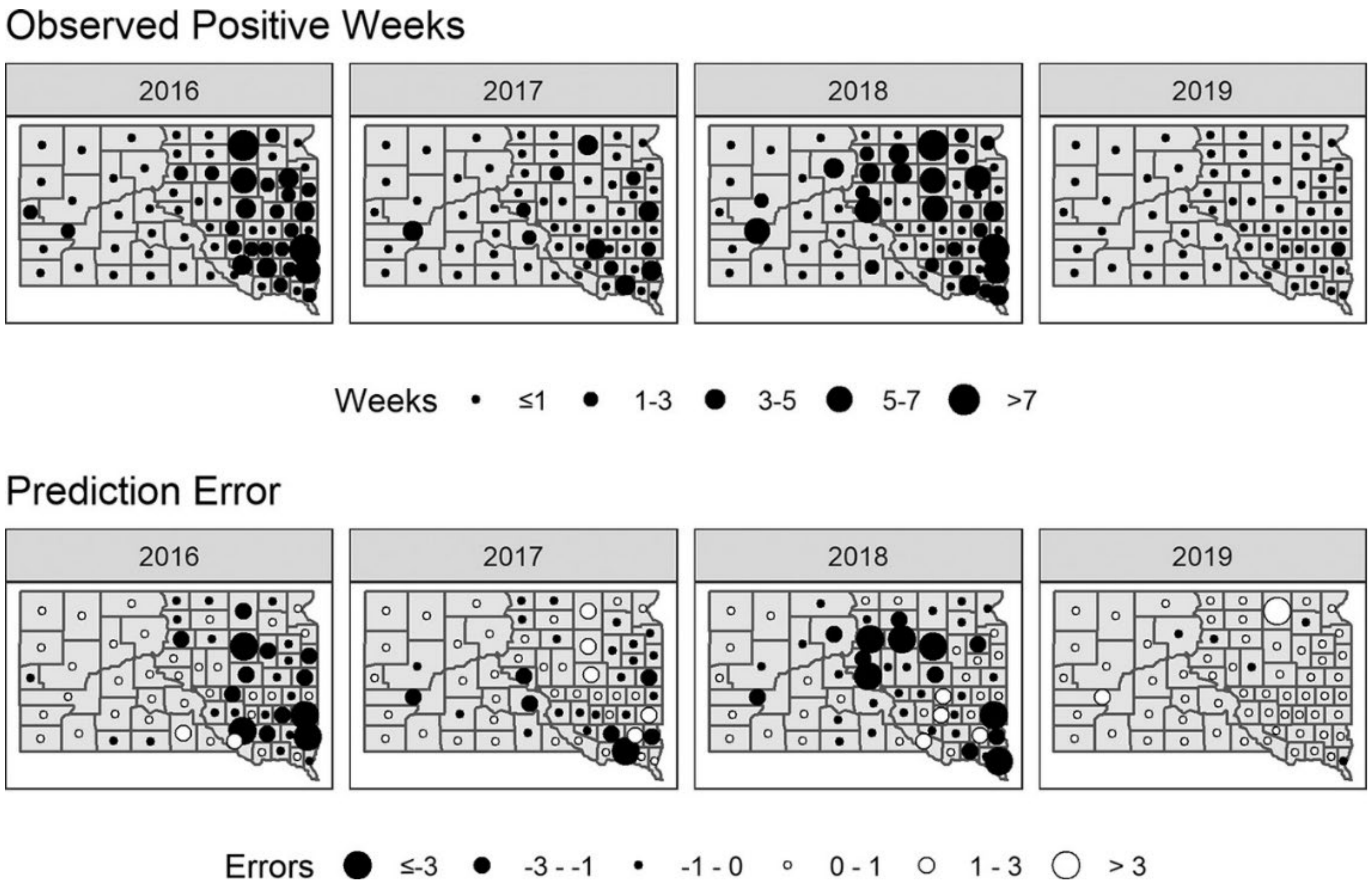
Solid and dashed lines show how predictions change over the course of the WNV season

Forecasts can distinguish years with high case numbers (2016 and 2018) from years with low case numbers (2017 and 2019)



The UNIVERSITY of OKLAHOMA

Forecasts based on environmental data predicted geographic patterns of WNV cases as well as their timing.

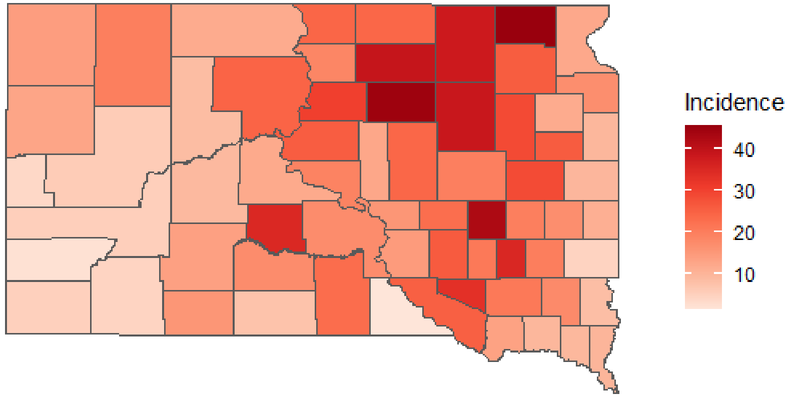


Extensions

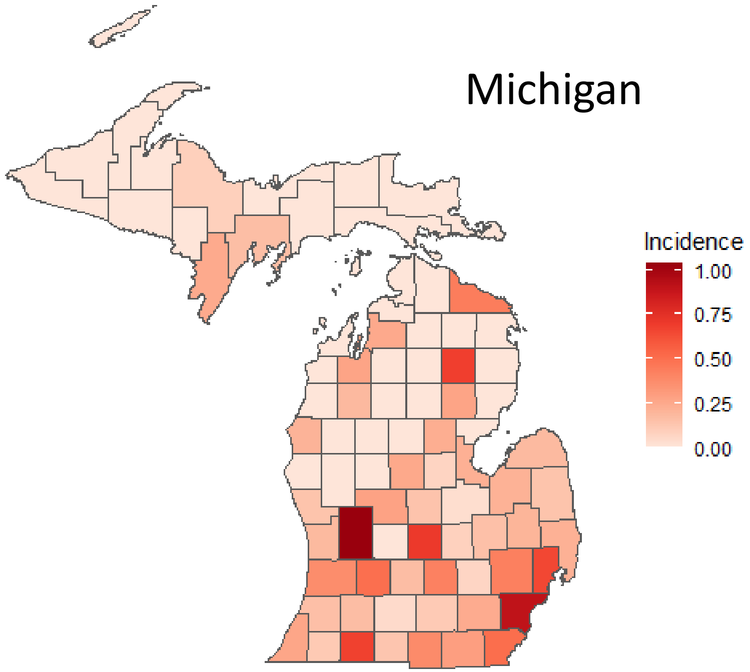
We have extended ArboMAP to other states

Average annual incidence rates in cases per 100,000

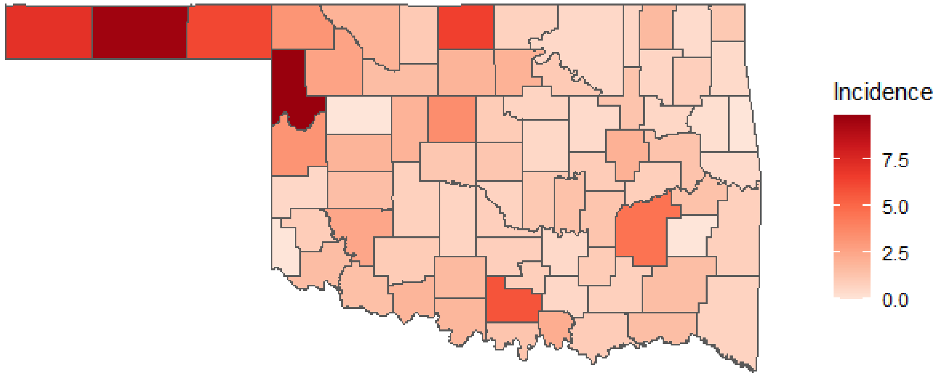
South Dakota



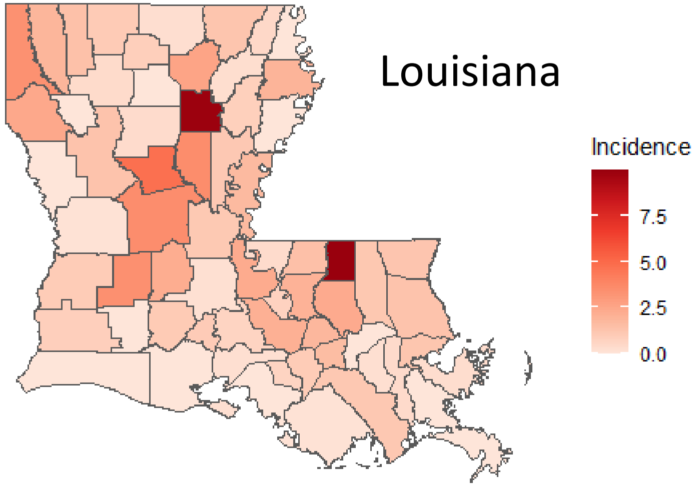
Michigan



Oklahoma



Louisiana



West Nile Virus Forecast Report for 2022-10-21
Louisiana

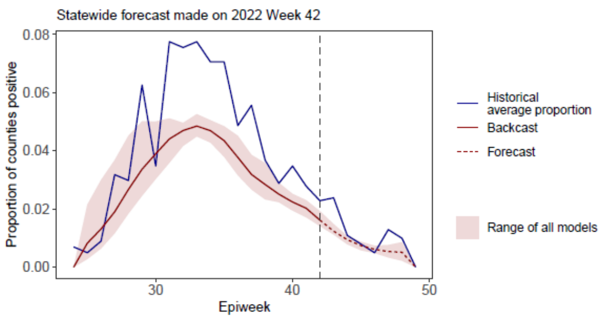
Arbovirus Modeling and Prediction (ArboMAP)

Dawn M. Nekorchuk, Justin K. Davis, and Michael C. Wimberly
(mcwimberly@ou.edu)

Geography and Envi

Rep

Contents



1.4 Case estimation

ArboMAP models are based on 'positive county-weeks', the probability that a county would have at least one human WNV case in a given week. These values can be used to predict a total number of **cases**, shown in the table below.

Table 1: Estimated number of WNV cases

Year	Predicted positive county-weeks	Average estimated cases (standard dev)	Range of estimated cases
2022	32	37 (+/-7)	30 - 45

1.5 Model fit statistics

The following table gives a summary of how well the model is fitting the historical years. The Area Under the ROC curve (AUC) is a statistic that ranges from 0 (model is right 0% of the time) to 1 (model is right 100% of the time). Scores above 0.5 are better than a random model, with >0.7 generally considered acceptable and >0.8 as good.

Table 2: Area Under Curve (AUC) statistics of all model fits

Model	Average AUC	Min AUC	Max AUC
Average of all models	0.88	0.88	0.89

West Nile Virus Forecast Report for 2022-07-19
Oklahoma

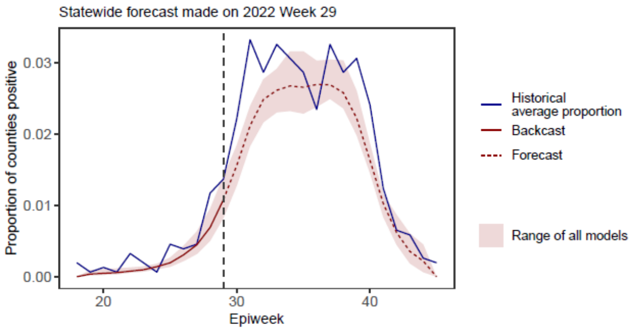
Arbovirus Modeling and Prediction (ArboMAP)

Dawn M. Nekorchuk, Justin K. Davis, and Michael C. Wimberly
(mcwimberly@ou.edu)

Geography and Enviro

Repo

Contents



1.4 Case estimation

ArboMAP models are based on 'positive county-weeks', the probability that a county would have at least one human WNV case in a given week. These values can be used to predict a total number of **cases**, shown in the table below.

Table 3: Estimated number of WNV cases

Year	Predicted positive county-weeks	Average estimated cases (standard dev)	Range of estimated cases
2022	23	20 (+/-4)	17 - 24

1.5 Model fit statistics

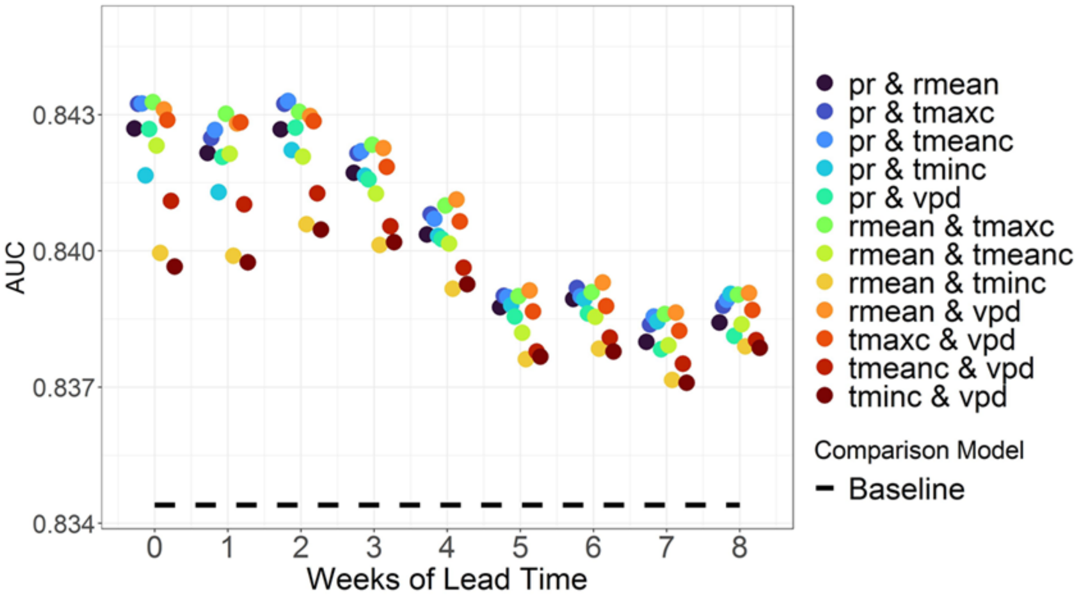
The following table gives a summary of how well the model is fitting the historical years. The Area Under the ROC curve (AUC) is a statistic that ranges from 0 (model is right 0% of the time) to 1 (model is right 100% of the time). Scores above 0.5 are better than a random model, with >0.7 generally considered acceptable and >0.8 as good.

Table 4: Area Under Curve (AUC) statistics of all model fits

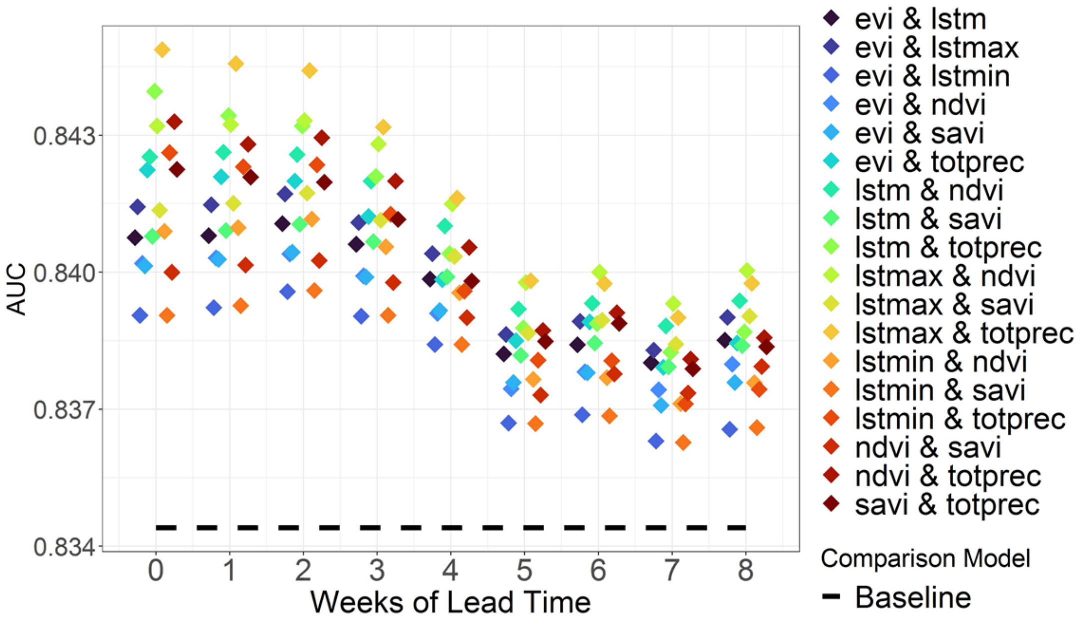
Model	Average AUC	Min AUC	Max AUC
Average of all models	0.9	0.9	0.91

Comparisons of forecast accuracy across based on different environmental data sources (2019-2021)

Gridded Weather Data: Interpolated meterological variables measured at ground-based stations.



Satellite Observations: Remotely sensed land surface temperature and vegetation greenness/moisture indicates.



Future prospects

- New modeling approaches
  - Other machine learning approaches
  - Dynamic (SIR) type epidemiological models
- New implementation approaches
  - More automated, cloud-based implementation
  - Need to address political/organizational as well as technical challenges
- New diseases
  - Could be applied to malaria and dengue as well as tick-borne diseases
  - Current forecasting approach better suited for endemic rather than emerging diseases
- For more information:
  - GitHub Archive: https://github.com/EcoGRAPH/ArboMAP
  - Lab Website: https://ecograph.net/
  - Contact: mcwimberly@ou.edu

SOUTH DAKOTA DEPARTMENT OF HEALTH

LOUISIANA DEPARTMENT OF HEALTH

MDHHS Michigan Department of Health & Human Services

OKC-COUNTY HEALTH DEPARTMENT

Supported by NASA Grants: NNX15AF74G, NSSC19K1233

NASA