A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP10287.

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,[1] *Justin K. Davis,*[1] *Michael B. Hildreth,*[2] *and Joshua L. Clayton*[3]

[1]Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, Oklahoma, USA
[2]Department of Biology and Microbiology, South Dakota State University, Brookings, South Dakota, USA
[3]South Dakota Department of Health, Pierre, South Dakota, USA

Background: West Nile virus (WNV), a global arbovirus, is the most prevalent mosquito-transmitted infection in the United States. Forecasts of WNV risk during the upcoming transmission season could provide the basis for targeted mosquito control and disease prevention efforts. We developed the Arbovirus Mapping and Prediction (ArboMAP) WNV forecasting system and used it in South Dakota from 2016 to 2019. This study reports a post hoc forecast validation and model comparison.

Objectives: Our objective was to validate historical predictions of WNV cases with independent data that were not used for model calibration. We tested the hypothesis that predictive models based on mosquito surveillance data combined with meteorological variables were more accurate than models based on mosquito or meteorological data alone.

Methods: The ArboMAP system incorporated models that predicted the weekly probability of observing one or more human WNV cases in each county. We compared alternative models with different predictors including *a*) a baseline model based only on historical WNV cases, *b*) mosquito models based on seasonal patterns of infection rates, *c*) environmental models based on lagged meteorological variables, including temperature and vapor pressure deficit, *d*) combined models with mosquito infection rates and lagged meteorological variables, and *e*) ensembles of two or more combined models. During the WNV season, models were calibrated using data from previous years and weekly predictions were made using data from the current year. Forecasts were compared with observed cases to calculate the area under the receiver operating characteristic curve (AUC) and other metrics of spatial and temporal prediction error.

Results: Mosquito and environmental models outperformed the baseline model that included county-level averages and seasonal trends of WNV cases. Combined models were more accurate than models based only on meteorological or mosquito infection variables. The most accurate model was a simple ensemble mean of the two best combined models. Forecast accuracy increased rapidly from early June through early July and was stable thereafter, with a maximum AUC of 0.85. The model predictions captured the seasonal pattern of WNV as well as year-to-year variation in case numbers and the geographic pattern of cases.

Discussion: The predictions reached maximum accuracy early enough in the WNV season to allow public health responses before the peak of human cases in August. This early warning is necessary because other indicators of WNV risk, including early reports of human cases and mosquito abundance, are poor predictors of case numbers later in the season. https://doi.org/10.1289/EHP10287

Introduction

West Nile virus (WNV), a mosquito-transmitted zoonotic arbovirus with a global distribution, poses significant public health threats in many parts of the world (Kramer et al. 2008). Although most infected humans are asymptomatic or have only mild symptoms, ~25% develop West Nile fever and <1% develop severe neuroinvasive disease that can lead to death (Petersen et al. 2013). In North America, the virus was first detected in New York City in 1999 and spread throughout the entire continent over the next several years (Kramer et al. 2019). WNV is now endemic in the conterminous United States, but the occurrence of human cases varies considerably in space and time. Annual incidence rates have been consistently highest in the northern Great Plains states, including North Dakota, South Dakota, and Nebraska (Ronca et al. 2019). Between 2004 and 2019, the total number of annually reported human WNV cases in the United States varied from 712 to 5,674, and there was even more interannual variation in individual states (CDC 2021). Because of these fluctuations, predictions of future WNV risk are desirable for targeting disease prevention and mosquito control activities at the times and locations where transmission risk is highest (Nasci and Mutebi 2019). This article presents a retrospective validation of the Arbovirus Mapping and Prediction (ArboMAP) system, which was developed to carry out WNV forecasting in South Dakota and implemented in 2016.

Mosquitoes in the genus *Culex* are the primary vectors of WNV, and multiple species of wild birds are the reservoir hosts. Humans are dead-end hosts that can be infected when bitten by a mosquito that has previously bitten an infected bird. The population dynamics and vector competence of mosquitoes are sensitive to meteorological variables, such as temperature and humidity (Shocket et al. 2020). The availability of aquatic habitats for development of mosquito larvae and the resulting densities of adult mosquitoes are affected by rainfall patterns (Gardner et al. 2012; Karki et al. 2016). Bird populations are impacted by mortality caused by extreme weather events, such as heat waves and droughts (Albright et al. 2010). In temperate climates, the timing of egg laying and hatching is influenced by the rate of warming in the spring (Burkett-Cadena et al. 2012; Shave et al. 2019). All of these environmentally mediated factors influence mosquito populations and the rate of virus amplification in mosquitoes and avian hosts, which in turn affect the density of infected mosquitoes and the potential for WNV transmission to humans (Paz 2019; Reisen 2013).

In the United States, reported human WNV cases are typically not confirmed until weeks to months after they occur, providing a lagging indicator of disease risk during the transmission season (DeFelice et al. 2019). Vector mosquitoes are often trapped and tested for WNV to obtain more timely data about where the virus is currently circulating and where humans are at risk (Hadler et al. 2015; Kilpatrick and Pape 2013). The

Address correspondence to Michael C. Wimberly, 100 E. Boyd St., Sarkeys Energy Center, Norman, OK 73019 USA. Email: mcwimberly@ou.edu

Supplemental Material is available online (https://doi.org/10.1289/EHP10287).
The authors declare they have nothing to disclose.
Received 9 September 2021; Revised 29 July 2022; Accepted 1 August 2022; Published 16 August 2022.

amount of mosquito surveillance that can be conducted is limited by available resources, and there are inevitably many locations with no mosquito data. In contrast, meteorological variables and other environmental risk factors can be mapped continuously and monitored through time using freely available geospatial data sets, such as interpolated climate grids and satellite remote sensing products. However, meteorological data alone are seldom sufficient to predict WNV cases with high accuracy. Other unmeasured factors related to vector populations, avian host communities, and human activities can decouple disease transmission cycles from their environmental potential (Allan et al. 2009; Johnson et al. 2012; Kwan et al. 2012). Therefore, we adopted a forecasting approach that used both mosquito surveillance data and meteorological observations with an aim of leveraging the strengths and minimizing the limitations of each type of information.

Various approaches have been used to develop predictive models for WNV (Barker 2019; Keyel et al. 2021). Most of these efforts have used a correlative approach, where mathematical functions or machine learning algorithms are fitted to historical observations. In these models, the dependent variable is an index of risk derived from either human WNV cases or mosquito infection rates. The independent variables are measurements of temperature, humidity, and precipitation from meteorological stations or related indices, such as land surface temperature and the normalized difference vegetation index (NDVI), derived from Earth-observing satellites. A variety of analytical techniques have been applied, ranging from generalized linear and mixed effects models (Hahn et al. 2015; Karki et al. 2020; Manore et al. 2014; Shand et al. 2016) to more flexible curve-fitting techniques, including generalized additive models (Chuang and Wimberly 2012; Smith et al. 2020) and machine learning methods, such as random forests (Keyel et al. 2019; Skaff et al. 2020). Mechanistic models, including epidemiological compartment models based on the susceptible–infected–recovered (SIR) framework, have also been used to incorporate temperature and mosquito infection rates as predictors of human WNV cases (DeFelice et al. 2017, 2018).

Although these modeling studies have been conducted in diverse locations using different data and modeling techniques, they have consistently found that high temperature before and during the transmission season increases mosquito infection rates and human WNV cases. However, the specific temperature thresholds for WNV transmission can vary geographically and over time (Skaff et al. 2020). In contrast to these temperature effects, there is much less consistency in the effects of precipitation on WNV, and both positive and negative effects have been found in different geographic settings (Hahn et al. 2015; Wimberly et al. 2014). Fewer studies have incorporated other moisture-related variables, such as humidity and soil moisture, making it difficult to generalize their effects. Recent reviews have highlighted several significant gaps in the development and implementation of forecasting models for WNV. Although multiple models have been created using a variety of data sources at a wide range of spatial and temporal scales, few have been applied to make real-time forecasts and connect the predictions with public health decision-making (Keyel et al. 2021). Further, most predictions have not been rigorously validated against independent data that were not used in the model-fitting process (Barker 2019).

The present study aims to address these gaps through a retrospective validation of 4 y of WNV forecasting in South Dakota, a state with a high burden and considerable interannual variation of WNV. Between 2009 and 2018, South Dakota had the second-highest average annual incidence of WNV neuroinvasive disease in the United States (3.06/100,000), with annual incidence rates ranging from zero to 7.44/100,000 (McDonald et al. 2021). *Culex tarsalis* has been implicated as the primary mosquito vector of WNV in South Dakota and surrounding states (Dunphy et al. 2019; Vincent et al. 2020). This species is primarily associated with rural habitats (Chuang et al. 2011), and as a result, the incidence of WNV is generally higher in small towns and rural areas than in larger cities (Chuang et al. 2012; Wimberly et al. 2008). Within South Dakota, the incidence of WNV is highest in the northern James River Valley, a glacial lowland dominated by relatively flat terrain with poorly drained soils, shallow water tables, and extensive wetlands that provide habitat for vectors and hosts (Hess et al. 2018).

In 2016, we developed and implemented an integrated WNV modeling approach in South Dakota that combined meteorological data with mosquito infection data. Prospective forecasts for 2016 correctly predicted a resurgence in WNV in contrast with relatively low case numbers in 2014 and 2015 (Davis et al. 2017). A subsequent model comparison study found that temperature and vapor pressure deficit were the most important meteorological predictors, showed that their effects varied over the course of the WNV transmission season, and confirmed that including infection rates from mosquito surveillance data improved model fit (Davis et al. 2018). These techniques were incorporated into a computer program for automatically generating WNV forecasts that we named ArboMAP. Here, we present a retrospective validation of WNV forecast accuracy in South Dakota using the ArboMAP system from 2016 to 2019. We tested the hypothesis that predictions based on meteorological variables combined with mosquito infection data were more accurate than models based on meteorological or mosquito data alone. These results, combined with our experience forecasting WNV risk in near real time and using the forecasts to support WNV control and prevention in South Dakota, can inform efforts to implement disease early warning systems and use them to support public health decision-making.

Methods

Data Sources

Records for 1,679 human WNV infections from 2004 to 2019 were provided by the South Dakota Department of Health (SDDOH). The initial epidemic years of 2002 and 2003 had large numbers of cases that reflected the unique conditions when WNV was first introduced into the region (Wimberly et al. 2013). These atypical years were excluded, and we analyzed endemic WNV from 2004 to 2019. Human WNV records included all instances of confirmed WNV fever cases (67%) and neuroinvasive disease cases (23%), as well as records of incidental discovery during blood donation (10%). Data on travel status were not available, and each case was referenced to the person's county of residence. The project was determined to be exempt from review by the institutional review board because the human case data were collected as a normal part of the SDDOH's surveillance process and were county-level summaries that did not include any personally identifying information.

The time of each case was referenced to the date of symptom onset or, for viremic blood donors, the date when the blood was donated. The median incubation period for WNV is 2.6 d, and 95% of individuals who develop symptoms do so within 7.0 d (Rudolph et al. 2014). Thus, the majority of symptom onset dates are within 1 wk of the infection date. Blood donors were screened using nucleic acid–amplification tests that detect active infections. WNV RNA is expected to be detectable up to 8 d after the infectious mosquito bite and would be detected prior to symptom onset for donors who eventually experience symptoms (Busch et al. 2008; Zou et al. 2010).

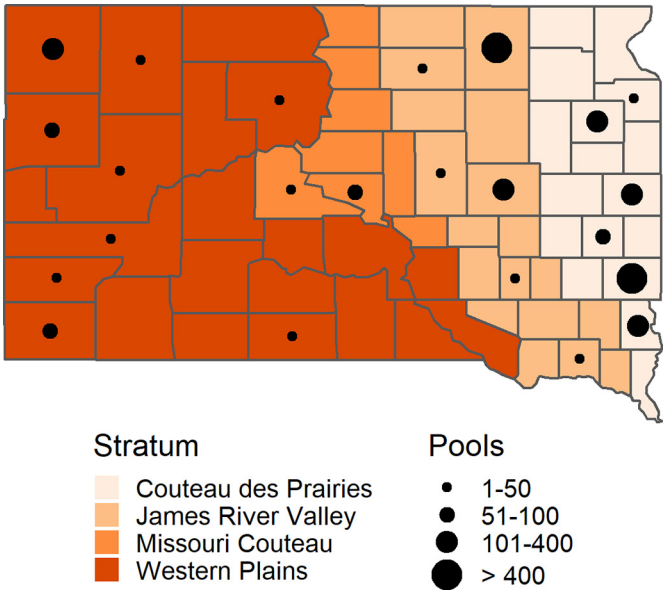

Figure 1. South Dakota counties with total number of tested mosquito pools from 2004 to 2019 and geographic strata used for modeling the mosquito infection growth rate (MIGR). Strata were delineated based on patterns of mosquito surveillance along with climate and physiography, and their names refer to the major physiographic zones with which they are associated. Map drawn using R (version 4.2.1; R Development Core Team).

Mosquito infection data for 2004–2019 were collected by state, county, city, and tribal entities. We used data from 162 trapping sites in 23 of the 66 South Dakota counties (Figure 1). All mosquitoes were collected with carbon dioxide–baited Centers for Disease Control and Prevention (CDC) light traps equipped with photo switches and air-activated gates. Mosquito trapping is generally conducted weekly for at least one 24-h period, but the timing can vary depending on the municipality carrying out the surveillance. A detailed description of mosquito surveillance activities in South Dakota is provided by Vincent et al. (2020). Mosquitoes were identified by morphological characteristics by trained individuals within local agencies and separated into single-species pools containing no more than 50 mosquitoes. Most of the pools (81%) were submitted to the SDDOH for testing using polymerase chain reaction according to the recommendations of Lanciotti et al. (2000). Mosquito pools collected by city agencies in Sioux Falls and Aberdeen were tested using the RAMP test (ADAPCO). Data from pools of *Culex tarsalis*, the primary vector of WNV in South Dakota, were used for statistical modeling. Of the 33,315 pools tested from 2004 to 2019, 1,414 (4.2%) were positive for WNV.

Meteorological variables were obtained from the Gridded Surface Meteorological (gridMET) data set (Abatzoglou 2013), including gridded daily estimates of mean temperature, mean relative humidity, mean vapor pressure deficit, and total precipitation at a 4-km spatial resolution. Daily data for 2004–2019 were summarized by county as zonal means.

Models

Because of delays in diagnosis and reporting, human cases are a lagging indicator of WNV transmission, and the total WNV case burden is not known until after the WNV season. In contrast, meteorological data and mosquito surveillance data can be obtained with latencies of a week or less. The objective of modeling was therefore to predict current and future human cases as a lagged function of meteorological observations and infection rates from mosquito surveillance data.

The dependent variable was human WNV cases summarized as county-weeks, with a value of one if any number of cases occurred in a county during a particular week (positive) and zero if no cases occurred (negative). Over the 16 y of data used in the study, 82% of the positive county-weeks included a single WNV case, and an additional 16% included two or three WNV cases. When summarized by county and year, the total number of positive county-weeks was strongly associated with the total number of WNV cases based on Spearman's rank correlation (rho = 0.98, $n = 532$, $p < 0.001$). Based on these results, we determined that the dichotomous variable for human case occurrence within a county-week provided an adequate representation of the spatiotemporal dynamics of WNV in South Dakota, and the additional complexities of modeling heavily zero-inflated data were not warranted. Spatial and temporal variability in positive county-weeks was therefore modeled as a dichotomous variable using logistic regression.

Meteorological effects were incorporated as distributed lags (Gasparrini et al. 2010), which are commonly used in epidemiology to model the delayed effects of environmental exposures on health outcomes (Braga et al. 2001; Schwartz 2000). Using this approach, each county-week was associated with a 120-d environmental history extending from the day the week began (lag 0) back to 120 d in the past (lag 120). Incorporating each daily lag as a different independent variable with a separate parameter would result in severe overfitting. The distributed lag approach instead models the effect of the independent variable as a smoothed function of lag date using a thin-plate spline. The effect of each independent variable in the logistic regression model is then estimated as

$$\sum_{l=0}^{L} s(l) u_{t-l},$$

where l is the lag in days, L is the maximum lag in days, $s(l)$ is the estimated coefficient for lag l modeled with a thin-plate spline, and u_{t-l} is the independent variable observed l days before time t, where t represents the first day of the current week. An advantage of this approach is that the modeling process automatically determines the range of lags over which the meteorological effects are strongest.

Because mosquito infection data were not available for every county, we partitioned the state into four geographic strata that reflected the patterns of mosquito surveillance along with climate and physiography (Figure 1). The strata were associated with major physiographic zones, including the Couteau des Prairies plateau, James River Valley, and Missouri Couteau in eastern South Dakota and the drier Great Plains landscapes west of the Missouri River. Mosquito testing data for all traps within each geographic stratum were aggregated and used to estimate weekly mosquito infection rates with a mixed effects logistic regression

$$logit(I_{s,y,t}) = (\gamma_0 + v_{0,y} + u_{0,s,y}) + (\gamma_1 + v_{1,y} + u_{1,s,y}) doy_t,$$

where $logit(I)$ was the probability that a mosquito pool was infected with WNV for stratum s, year y, and day t; γ_0 was the global intercept; γ_1 was the global slope; v were random effects on intercept and slope per year; u were random effects on intercept and slope per stratum within year; and doy was the standardized day of year (Davis et al. 2018).

The mosquito infection growth rate (MIGR) was estimated for every combination of stratum and year as

$$m_{s,y} = \widehat{\gamma_1} + \widehat{v_{1,y}} + \widehat{u_{1,s,y}},$$

where the hats indicated posterior modes from the fitted model. The rationale behind this model is that a higher MIGR is an

Table 1. Description of models used for West Nile virus forecasting in South Dakota.

Abbreviation	Description
Base	Baseline model with seasonal cyclical function and no mosquito or meteorological data
Mosq	Mosquito infection rate model with seasonal cyclical function
Env	Environmental model with distributed lags of untransformed meteorological variables
EnvAnom	Environmental model with seasonal cyclical function and distributed lags of meteorological anomalies
EnvSv	Environmental model with seasonally varying distributed lags of untransformed meteorological variables
EnvSvAnom	Environmental model with seasonal cyclical function and seasonally varying distributed lags of meteorological anomalies
Comb	Combined mosquito–environment model with distributed lags of untransformed meteorological variables
CombAnom	Combined mosquito–environment model with seasonal cyclical function and distributed lags of meteorological anomalies
CombSv	Combined mosquito–environment model with seasonally varying distributed lags of untransformed meteorological variables
CombSvAnom	Combined mosquito–environment model with seasonal cyclical function and seasonally varying distributed lags of meteorological anomalies
Ens2	Mean of CombAnom and CombSvAnom
Ens3	Mean of CombAnom, CombSvAnom, and Comb
Ens4	Mean of CombAnom, CombSvAnom, Comb, and CombSv

Note: Each row describes 1 of 13 models. The abbreviation for each model is used in the article text, tables, and figures. The description includes the distinctive components of each model. Each model contains a county-level fixed effect term in addition to the components listed in the description.

Table 2. List of model equations used for West Nile virus forecasting in South Dakota.

Model	Equation
Base	$\text{logit}(p_{i,t}) = c_{0i} + cyc(w)$
Mosq	$\text{logit}(p_{i,t}) = c_{0i} + c_1 m_{i,t} + cyc(w)$
Env	$\text{logit}(p_{i,t}) = c_{0i} + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l) u_{i,t-l}$
EnvAnom	$\text{logit}(p_{i,t}) = c_{0i} + cyc(w) + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l) a_{i,t-l}$
EnvSv	$\text{logit}(p_{i,t}) = c_{0i} + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l,w) u_{i,t-l}$
EnvSvAnom	$\text{logit}(p_{i,t}) = c_{0i} + cyc(w) + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l,w) a_{i,t-l}$
Comb	$\text{logit}(p_{i,t}) = c_{0i} + c_1 m_{i,t} + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l) u_{i,t-l}$
CombAnom	$\text{logit}(p_{i,t}) = c_{0i} + c_1 m_{i,t} + cyc(w) + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l) a_{i,t-l}$
CombSv	$\text{logit}(p_{i,t}) = c_{0i} + c_1 m_{i,t} + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l,w) u_{i,t-l}$
CombSvAnom	$\text{logit}(p_{i,t}) = c_{0i} + c_1 m_{i,t} + cyc(w) + \sum_{k=1}^{2}\sum_{l=0}^{L} s_k(l,w) a_{i,t-l}$

Note: $p_{i,t}$ is the probability of one or more WNV cases in county i at time t; c_{oi} is a county-level fixed effect; $m_{i,t}$ is the mosquito infection growth rate (MIGR) in county i at time t; c_1 is the parameter for MIGR; $cyc(w)$ is a cyclical seasonal function; k indexes two meteorological variables; l indexes daily lags up to $L = 120$ days in the past; $s_k(l)$ is a thin-plate spline function of lag days; $s_k(l, w)$ is a thin-plate spline function of lag days and week of the year; $u_{i,t-l}$ is an untransformed meteorological variable for county i at time $t–l$ and $a_{i,t-l}$ is an anomalized meteorological variable for county i at time $t–l$. Base, baseline; Comb, combined; CombAnom, combined with anomalies; CombSv, combined with seasonally varying distributed lags; CombSvAnom, combined with anomalies and seasonally varying distributed lags; Ens2, two model ensemble; Ens3, three model ensemble; Ens4, four model ensemble; Env, environmental; EnvAnom, environmental with anomalies; EnvSv, environmental with seasonally varying distributed lags; EnvSvAnom, environmental with anomalies and seasonally varying distributed lags; Mosq, mosquito.

indicator of rapid WNV amplification in the early transmission season, which results in larger proportions of infected mosquitoes and higher risk of human cases during the peak season (Davis et al. 2017).

We compared five types of models (Table 1). The baseline model included only a cyclical seasonal function and a fixed effect for each county. The cyclical function captured the annual epidemic curve resulting from seasonal climate patterns. The county-level effects captured geographic variation caused by differences in human population and ecological effects of land cover and hydrology (Hess et al. 2018). This model used only historical human case data as predictor variables, thus providing a benchmark for determining if incorporating mosquito and meteorological predictors can improve accuracy.

The mosquito model included the cyclical and fixed effect terms plus the MIGR variable. A set of four environmental models included cyclical and fixed effect terms plus distributed lags for temperature and vapor pressure deficit. These variables were selected based on a previous model comparison exercise that found they were the strongest predictors of historical WNV occurrence (Davis et al. 2018). The environmental models incorporated meteorological data in two different ways: *a*) as untransformed variables, and *b*) as anomalies calculated by subtracting the seasonal expectation for each day of the year. For each type of data transformation, relationships with WNV cases were modeled using two approaches: *a*) a fixed set of distributed lags that assumed environmental relationships were constant throughout the WNV season, and *b*) seasonally varying distributed lags that allowed environmental relationships to vary over the course of the WNV season (Davis et al. 2018). A set of four combined models incorporated the MIGR variable into each of the four environmental models. The mathematical equations for these models are specified in Table 2. Finally, a set of three model ensembles were implemented as simple means of the predictions of the best two combined models, the best three combined models, and all four combined models.

Forecasts

We used the terms "forecast year" to refer to the year in which a forecast was made, "forecast week" to identify the week of the year when a forecast was produced, and "predicted week" to indicate a week for which human cases were predicted. For a given forecast year, the model was calibrated using data from all previous years back to 2004 (Figure 2). For example, the models used for the 2016 predictions were calibrated using data from 2004 to 2015, and the models used for the 2017 predictions were calibrated using data from 2004 to 2016. There were 19 forecast weeks during each WNV season, starting in early June (week 22, where weeks are specified as CDC epidemiological weeks) and continuing through the end of September (week 40). During each forecast week, forecasts were generated for every predicted week of the WNV season (weeks 22–40), estimating the probability of one or more WNV cases occurring in each county during the week. Thus, the predictions generated by the ArboMAP system encompassed past weeks of the WNV season and the current forecast week, as well as future weeks. However, we still refer to these predictions as forecasts because the actual weekly case counts are typically not known until after the end of the WNV season.

To generate the prediction for a forecast week, current-year mosquito infection data and meteorological variables were used as independent variables to predict county-level WNV probabilities (Figure 2). The current-year MIGR values were estimated for all strata using the available mosquito surveillance data up to the first day of the forecast week. For forecast weeks early in the

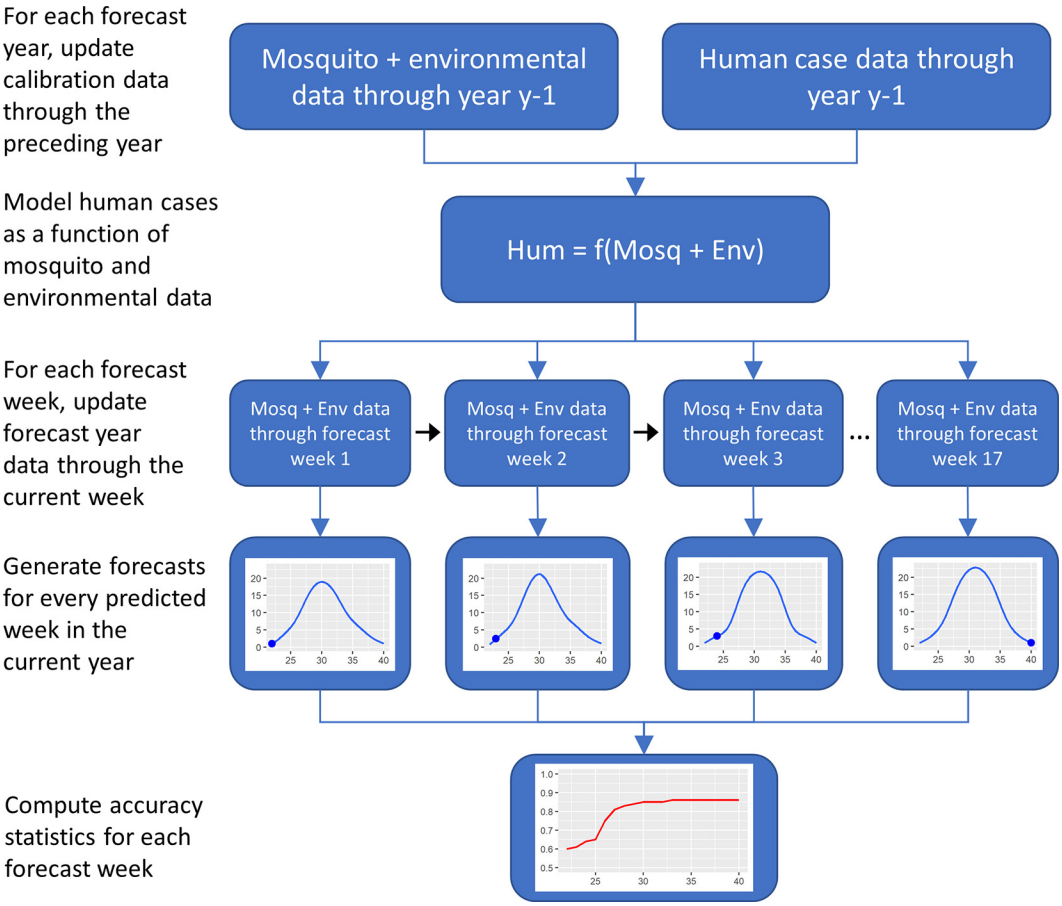

Figure 2. Major steps for calibrating the Arbovirus Monitoring and Prediction (ArboMAP) system, generating weekly forecasts, and computing accuracy statistics. Note: Env, environmental; Mosq, mosquito.

season when no positive mosquito pools had been detected, the MIGR could not be estimated, and the average historical MIGR for that forecast week taken over all historical years was used instead. All available meteorological observations were combined with average values from the historical climatology for future dates. The MIGR estimates and observed meteorological data were then used as independent variables to predict county-level WNV probabilities. Thus, the amount of current-year data that was incorporated into the forecasts increased throughout the year as the forecast week advanced. The predicted probabilities could be summed across counties to compute the expected number of positive counties per prediction week, across predicted weeks to compute the expected number of positive weeks per county, and across counties and predicted weeks to estimate the total number of positive county-weeks in the year.

Validation

The ArboMAP system was used in a real-time, prospective mode to generate forecasts from 2016 to 2019. Various updates were made to the model every year, including changes to the specification of the statistical models, the numbers of spatial strata used in the MIGR model, and the specific meteorological variables used as predictors. To facilitate a systematic accuracy assessment and model comparison, we undertook a retrospective validation in which we evaluated 13 alternative models across the entire 4-y period. The meteorological variables and mosquito surveillance data were censored at each time step so that only data available prior to each forecast week were used to make the predictions. This approach replicated the operational forecasting process as described in the preceding section and allowed us to compare multiple forecasting scenarios with different models.

We computed the receiver operating characteristic curve (AUC) based on the predicted probability for each county-week and the observations of county-weeks. We computed the temporal error (TE) by summing the predicted probabilities and observed county-weeks across counties and computing the mean absolute error (MAE) for all prediction weeks. We computed the spatial error (SE) by summing the predicted probabilities and observed county-weeks across predicted weeks and computing the MAE for all counties. These accuracy statistics were computed separately for each forecast week to compare forecasts made at different times during the WNV season. Results from different models were compared to assess how accuracy varied in relation to the types of data used for the predictions and the specifications of the predictive models.

This validation exercise determined the accuracy of each model when making seasonal predictions of WNV occurrence for the entire transmission season. We also examined the accuracy of lagged weekly predictions by computing the AUC of predictions made zero through 6 weeks in the past for predicted weeks 28–40. Finally, we carried out a sensitivity analysis to determine whether the presence of a mosquito trap in a county had a significant effect on model accuracy. We removed the 23 counties with mosquito infection data from the statewide data set and carried out the same validation process that was used on the full data set.

Software

The ArboMAP system was implemented using R (version 4.2.1; R Development Core Team). The application was coded as an R

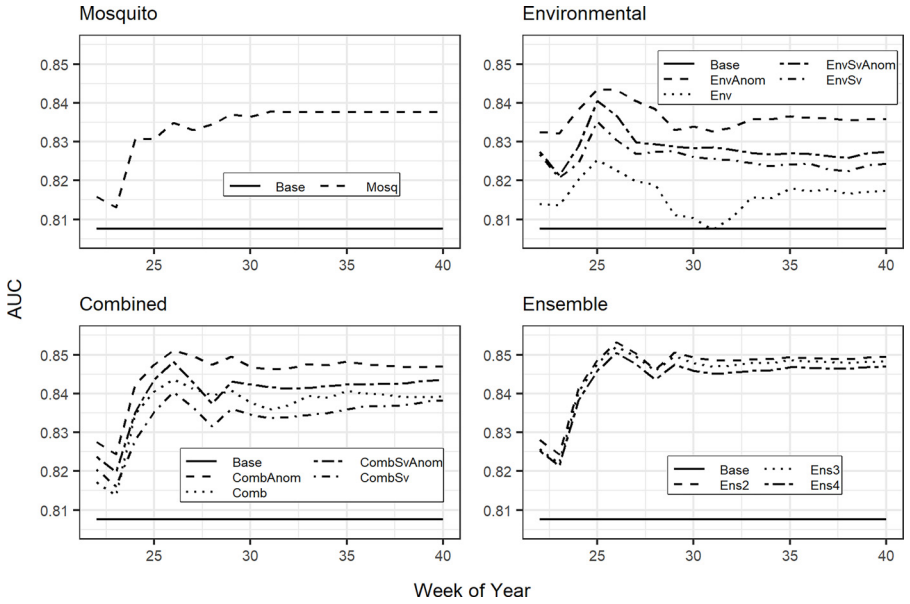

Figure 3. Accuracy of South Dakota West Nile virus predictions from 2016 to 2019 measured as the area under the receiver operating characteristic curve (AUC). Predictions were generated using the Arbovirus Monitoring and Prediction (ArboMAP) software with 13 different model formulations. Each line represents the variation in AUC by forecast week for one model. The logistic regression models incorporated various combinations of mosquito infection variables and lagged meteorological variables (temperature and vapor pressure deficit) with different variable transformations and model structures. Individual models are defined in Table 1 and mathematical forms are provided in Table 2. Note: Base, baseline; Comb, combined; CombAnom, combined with anomalies; CombSv, combined with seasonally varying distributed lags; CombSvAnom, combined with anomalies and seasonally varying distributed lags; Ens2, two model ensemble; Ens3, three model ensemble; Ens4, four model ensemble; Env, environmental; EnvAnom, environmental with anomalies; EnvSv, environmental with seasonally varying distributed lags; EnvSvAnom, environmental with anomalies and seasonally varying distributed lags; Mosq, mosquito.

markdown file that automatically executed all data processing and modeling steps and generated a formatted report with model output. The code used to carry out the analyses in this article is included in the Supplemental Material ("ArboMap_Main_Code.Rmd"). The ArboMAP system is under active development, and the most recent version of the code along with a user guide and simulated data for testing can be found on our GitHub archive (https://github.com/EcoGRAPH/ArboMAP).

Results

The AUC statistics characterized the overall accuracy of county-week predictions over both space and time, with larger values indicating higher accuracy (Figure 3). The two-model ensemble (Ens2) had the highest AUCs in most of the forecast weeks. AUC values were only slightly lower for the combined model with climate anomalies and without seasonally varying environmental effects (CombAnom) and for the three- and four-model ensembles (Ens3 and Ens4). The AUCs for this group of models increased rapidly through forecast week 26 and reached peak values of 0.85 or higher. After forecast week 26, the AUCs declined slightly but remained relatively stable with values close to 0.85 through forecast week 40.

At the beginning of the WNV season in forecast weeks 22–23, the AUCs of the environmental models were the same or higher than the corresponding combined models (Figure 3). However, beginning in forecast week 26, the AUC values of the combined models were always higher than the environmental models. The AUC of the mosquito model (Mosq) was lowest at the beginning of the WNV season, increased through forecast week 31, and was stable thereafter, but it was lower than all three ensemble models and the best combined model (CombAnom) in every forecast week. All models had the same or higher AUCs than the baseline model, which included only a cyclical seasonal term and a fixed effect for each county.

The TE statistic characterized the accuracy of the predicted weekly number of counties with one or more WNV cases, with lower values indicating higher accuracy (Figure 4). Results were generally comparable to those based on the AUC statistic. The two-model ensemble (Ens2) had the lowest error in all forecast weeks, with TE values <2.0 in forecast weeks 26–40. The TE values of the other ensemble models (Ens3 and Ens4) and two of the combined models (CombAnom and CombSvAnom) were only slightly higher. The environmental model with climate anomalies and without seasonally varying environmental effects (EnvAnom) had lower TEs than the mosquito model (Mosq) and was nearly as low as the ensemble models and the two best combined models (CombAnom and CombSvAnom). All of the ensemble and combined models, as well as the mosquito models and two of the four environmental models, had lower TEs than the baseline model.

Considering the weekly time series of results from the two-model ensemble (Ens2), the fitted values from 2004 to 2015 showed that the meteorological data and mosquito infection data were able to predict much of the seasonal and interannual fluctuations in historical WNV cases (Figure 5). Although there was overprediction in some years (2006, 2010–2011, and 2015) and underprediction in other years (2004–2005), the fitted values clearly distinguished the lowest-transmission years (2004, 2008–2011, and 2014–2015) from the highest-transmission years (2005, 2007, and 2012–2013). The predicted seasonal curves for 2016–2019 changed with the forecast week as more meteorological and mosquito data from the current year were incorporated as predictors. However, the predictions generated during the early weeks of the WNV season (forecast weeks 26–30) were able to distinguish between years with high cases (2016 and 2018) vs. those with low cases (2017 and 2019).

The SE statistic characterized the accuracy of predicted total weeks with one or more WNV cases at the county level, with lower values indicating higher accuracy (Figure 6). Overall, SE

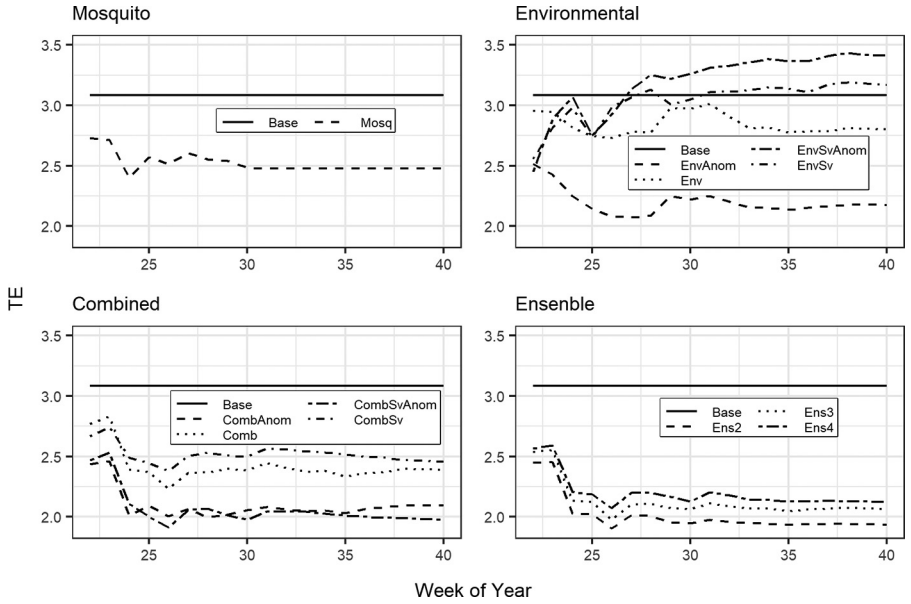

Figure 4. Accuracy of South Dakota West Nile virus predictions from 2016 to 2019 measured as temporal error (TE), which is the mean absolute error of predicted positive counties per week. Predictions were generated using the Arbovirus Monitoring and Prediction (ArboMAP) software with 13 different model formulations. Each line represents the variation in TE by forecast week for one model. The logistic regression models incorporated various combinations of mosquito infection variables and lagged meteorological variables (temperature and vapor pressure deficit) with different variable transformations and model structures. Individual models are defined in Table 1 and mathematical forms are provided in Table 2. Note: Base, baseline; Comb, combined; CombAnom, combined with anomalies; CombSv, combined with seasonally varying distributed lags; CombSvAnom, combined with anomalies and seasonally varying distributed lags; Ens2, two model ensemble; Ens3, three model ensemble; Ens4, four model ensemble; Env, environmental; EnvAnom, environmental with anomalies; EnvSv, environmental with seasonally varying distributed lags; EnvSvAnom, environmental with anomalies and seasonally varying distributed lags; Mosq, mosquito.

was lower than TE, which was expected because the statewide pattern of WNV cases remains consistent from year to year and the county-level fixed effect is a strong predictor of this pattern. Results for all the ensemble and combined models were similar, with SE values decreasing rapidly in forecast weeks 22–26 and remaining relatively stable between 0.9 and 0.95 in forecast weeks 26–40. The mosquito model (Mosq) had SE values only slightly higher than the ensemble and combined models, and it generally had lower SEs than the environmental models. All of the ensemble and combined models, as well as the mosquito model and three of the environmental models, had lower SE than the baseline model.

Maps of observed WNV cases showed similar patterns in each year, with more WNV-positive weeks in eastern than in western South Dakota (Figure 7). However, some shifts in the spatial pattern were evident, such as the concentration of WNV

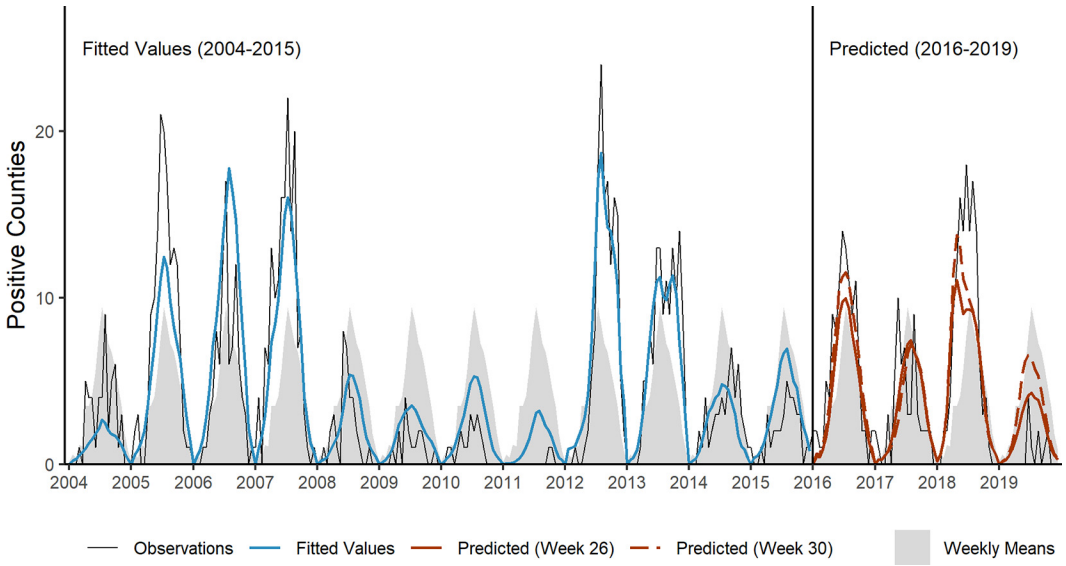

Figure 5. Observed numbers of counties with one or more West Nile virus cases per week from 2004 to 2019 in South Dakota and predictions from the two-model ensemble (Ens2). Ensemble predictions are the mean of predictions from the combined mosquito–environment model with distributed lags of meteorological anomalies (CombAnom) and the combined mosquito–environment model with seasonally varying distributed lags of meteorological anomalies (CombSvAnom). Fitted values are shown from 2004 to 2015 and predictions outside of the training data set from 2016 to 2019. Predictions are shown for two forecast weeks: 26 (late June) and 30 (late July). Shaded areas represent the weekly means of observed positive counties across all years.

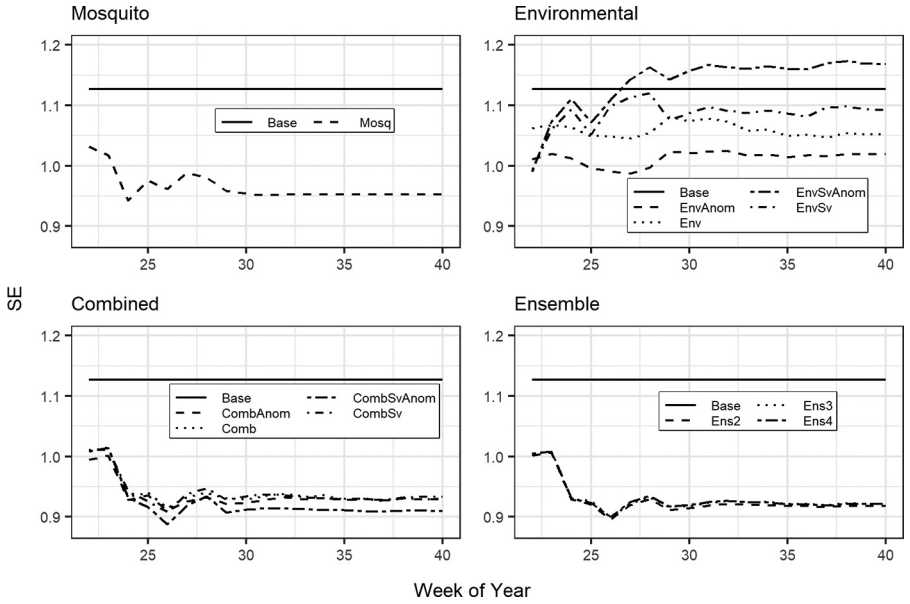

Figure 6. Accuracy of South Dakota West Nile virus predictions from 2016 to 2019 measured as spatial error (SE), which is the mean absolute error of predicted positive weeks per county. Predictions were generated using the Arbovirus Monitoring and Prediction (ArboMAP) software with 13 different model formulations. Each line represents the variation in SE by forecast week for one model. The logistic regression models incorporated various combinations of mosquito infection variables and lagged meteorological variables (temperature and vapor pressure deficit) with different variable transformations and model structures. Individual models are defined in Table 1 and mathematical forms are provided in Table 2. Note: Base, baseline; Comb, combined; CombAnom, combined with anomalies; CombSv, combined with seasonally varying distributed lags; CombSvAnom, combined with anomalies and seasonally varying distributed lags; Ens2, two model ensemble; Ens3, three model ensemble; Ens4, four model ensemble; Env, environmental; EnvAnom, environmental with anomalies; EnvSv, environmental with seasonally varying distributed lags; EnvSvAnom, environmental with anomalies and seasonally varying distributed lags; Mosq, mosquito.

in southeastern South Dakota in 2017 and the westward expansion of WNV in 2018. The largest prediction errors were concentrated in the southeastern and north-central portions of the state, where WNV-positive weeks were underpredicted in 2016 and 2018.

The validation of lagged weekly predictions found that the AUC statistics were nearly as high as in the seasonal validations, with a maximum AUC value from the two-model ensemble (Ens2) of just >0.84 (Figure S1). Furthermore, the AUC values were considerably higher than the baseline model and remained relatively stable as the forecast lead time increased from zero to 6 weeks. Thus, the modeling approach generated reliable weekly as well as seasonal forecasts.

AUC values were lower in the validation run that excluded all counties with mosquito infection data (Figure S2). However, the baseline predictions, which did not include mosquito data, were also considerably lower than in the main analysis. In general, the counties that did not provide mosquito data tended to have smaller towns, lower populations, and fewer WNV cases than the counties that did provide mosquito data. As a result, WNV occurrence is more heterogeneous in space and time and is more challenging to predict even when not using mosquito data. The accuracy of the mosquito model was still considerably higher than that of the baseline model, and the combined and ensemble models were more accurate than models based only on mosquito or meteorological data. Based on these results, we concluded that the mosquito data contributed substantially to the prediction accuracy even in counties where no surveillance was conducted.

Discussion

Statistical models based on air temperature and vapor pressure deficit combined with mosquito infection rates were able to predict spatial and temporal patterns of WNV cases in South Dakota. All models with either meteorological or mosquito data outperformed simple baseline models based only on county means and seasonal trends. The best combined models were more accurate than models based only on meteorological or mosquito data. Models based on anomalies were more accurate than models based on untransformed meteorological variables, and a simple ensemble of the two best combined models generated the most accurate predictions. Accuracy increased rapidly from forecasts made in early June (forecast week 22) through early July (week of year 26) and was relatively stable thereafter. These results were consistent for multiple accuracy statistics that measured the ability to discriminate county-weeks with and without WNV cases (AUC), error in the predicted time series of positive districts (TE), and error in the predicted spatial patterns of positive counties (SE).

Because human WNV cases in South Dakota do not peak until August (Kightlinger 2017; Wimberly et al. 2013), these early July predictions still allow sufficient lead time to expand mosquito control efforts and implement WNV prevention programs if there is an indication of high risk. The main technique for WNV vector control in South Dakota is ground-based spraying for adult mosquitoes. Many of the larger communities also use larvicides following major rain events. Initiating these control activities before there is widespread human transmission lowers the abundance of biting vector mosquitoes and may also reduce the mosquito infection rate (Nasci and Mutebi 2019). The SDDOH and some municipal governments also implement advertising campaigns that encourage prevention measures such as using mosquito repellants. Targeting this messaging when and where WNV risk is highest may help to limit the phenomenon of message fatigue, in which repeated warnings are eventually ignored because of overexposure (So et al. 2017).

In South Dakota and other states with high WNV burdens, linking control and prevention activities to WNV forecasts is

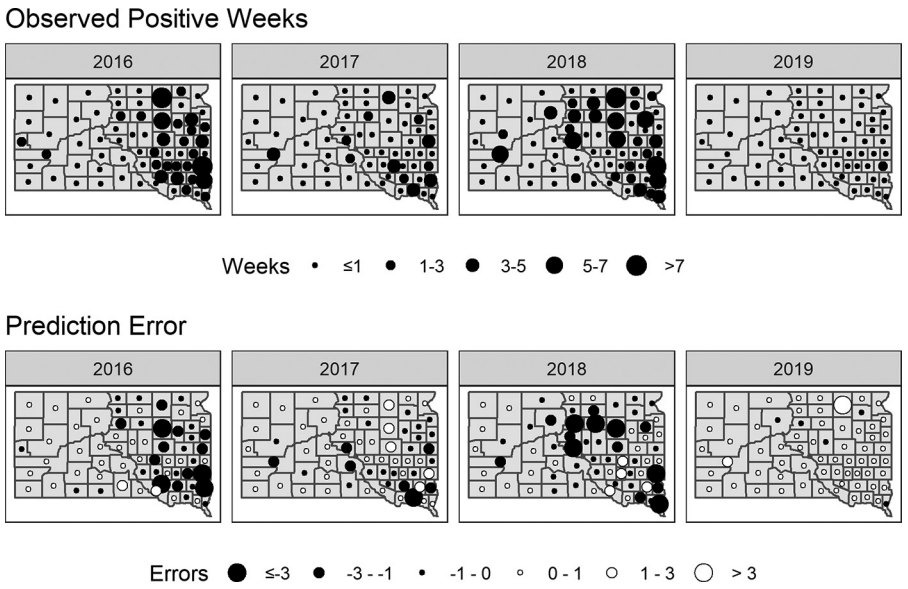

Figure 7. Observed numbers of weeks with one or more West Nile virus cases from 2016 to 2019 in South Dakota and prediction errors from the two-model ensemble (Ens2). Ensemble predictions are the mean of predictions from the combined mosquito–environment model with distributed lags of meteorological anomalies (CombAnom) and the combined mosquito–environment model with seasonally varying distributed lags of meteorological anomalies (CombSvAnom). Prediction error is the difference between the predicted number of positive weeks and the observed number of positive weeks for each county in each year. Map drawn using R (version 4.2.1; R Development Core Team).

important because other sources of information may not provide timely and reliable indication of disease risk. Because the reporting of human cases can take weeks to months, tracking human cases during the WNV season can provide misleading information about the current level of transmission. For example, in 2018, only eight confirmed cases were reported to the SDDOH by July 26. If viewed alone as an indicator of current WNV transmission, this number would have given a false impression of low risk. In reality, 50 cases had already occurred by this time, and there were 169 cases by the end of the 2018 season—the highest total since the large national outbreak in 2012. The ArboMAP system was able to predict a higher-than-normal level of WNV risk by early July in 2018 based on warm weather in the spring combined with rapid growth of the mosquito infection rate in early summer. The number of reported human cases did not increase substantially until August, after WNV had already peaked. Waiting this long to respond would be too late to have a substantial effect on transmission.

Mosquito control programs often respond to public perceptions of the mosquito nuisance and the resulting demands for action from local agencies (Dickinson and Paskewitz 2012). However, high mosquito abundances and frequent bites do not necessarily translate into high risk of WNV transmission. In South Dakota, the most important nuisance mosquito is *Aedes vexans*, which occurs in large numbers in spring and early summer after heavy rains flood their breeding habitats (Vincent et al. 2020). In contrast, the primary vector of WNV in South Dakota is *Culex tarsalis*. This species peaks during mid to late summer and has a lower abundance that is more stable within and between years. During the 2018 WNV season, local mosquito control programs in South Dakota were reporting low mosquito abundances and few public complaints through July. Without forecasts based on weather and mosquito infection data, these observations of low mosquito activity can be misinterpreted as indicating low WNV risk.

In the United States, mosquito surveillance and WNV testing programs are implemented at municipal, county, and state levels (Hadler et al. 2015). We have demonstrated how the data gathered from these programs can be enhanced by incorporating meteorological data as a component of WNV surveillance and using predictive models to forecast the risk of disease in humans. The GridMET data used in this study are freely available, and there are additional sources of free environmental monitoring data that could be used to model mosquito-transmitted diseases. However, expertise in geospatial data processing and modeling along with specialized software and sufficient computational resources are needed to acquire and use these data. To facilitate the use of the ArboMAP system by the SDDOH, we used the Google Earth Engine cloud-based platform (Gorelick et al. 2017) to build an app that automated the downloading of daily, county-level summaries of GridMET variables. The forecasting models were implemented in the free and open source R language and software environment as a script that automatically generated forecasts and output the results as a formatted portable document format (PDF) report. The continued development of informatics tools such as these will be essential to facilitate the integration of environmental monitoring data for disease surveillance and predictions.

The limitations of the ArboMAP forecasting approach are closely linked to the quality and scale of the data used to make the predictions. Some studies have analyzed only neuroinvasive cases based on the presumption that these severe disease cases are detected and reported more consistently than milder forms of disease (Hahn et al. 2015; Paull et al. 2017). However, we used all available WNV case data, including neuroinvasive cases, as well as fever cases and reports of viremic blood donors. Because South Dakota is a relatively small state where WNV is well known as a locally important disease in the medical community, we assumed that fever cases were also well diagnosed throughout the state. A previous study found that annual fever case counts were strongly correlated with neuroinvasive case counts, and that geographic patterns of total WNV incidence were similar to that of neuroinvasive disease incidence (Wimberly et al. 2013). Using all types of WNV cases in our models allowed us to detect more locations where WNV transmission to humans had occurred. The

results of our validation confirmed that WNV cases based on this broader definition were predictable using mosquito infection rates and environmental conditions. More generally, we suggest that improving the quality and consistency of human and mosquito surveillance data will increase the accuracy of forecasts for WNV and other infectious diseases.

There is also potential to improve prediction accuracy by exploring alternative sources of environmental data. We used meteorological data from GridMET because they were easy to access, easy to process with the Google Earth Engine, and were available at a lag of only a few days. Other prospective data sources include satellite Earth observations of temperature, vegetation greenness, soil moisture, and precipitation (Chuang and Wimberly 2012; Marcantonio et al. 2016), as well as modeled soil moisture and other hydrological outputs from land data assimilation systems (Davis et al. 2018; Shaman et al. 2010). Alternative spatial resolutions besides the county-level should also be explored to determine the scales at which environment–disease relationships are the strongest (Uelmen et al. 2021). Given the many possible approaches, developing and testing new environmental models of WNV remains an important area for future research.

In the ArboMAP system, we used the seasonal rate of increase in the proportion of infected mosquito pools as an indicator of entomological risk (Davis et al. 2017, 2018). We chose this approach because it was relatively insensitive to changes in the numbers and locations of mosquito traps from year to year, allowing us to use all of our statewide mosquito data for model calibration and prediction. A prior study found that, in South Dakota, mosquito abundance is not a strong indicator of human WNV risk (Vincent et al. 2020). However, additional variables characterizing total mosquito abundance or vector index (the estimated abundance of infected mosquitoes) could be tested and incorporated into the model. Other types of predictor data could also be explored, including information about avian host communities, the timing and locations of outdoor gatherings that may increase human exposure, and the impacts of mosquito control activities.

In conclusion, this research showed that it is possible to predict the annual WNV case burden early in the season before most human transmission has occurred. The best predictions were obtained from models that included lagged weather variables, as well as mosquito infection data. Because environmental monitoring data are freely available from a variety of sources, it is feasible to integrate them with epidemiological and entomological surveillance data to implement routine forecasting of WNV. Our experience making real-time predictions in South Dakota demonstrated that these forecasts provide alerts of upcoming WNV outbreaks early in the season when other indicators of WNV risk, such as mosquito abundance and reported human WNV cases, are either ambiguous or misleading (Kilpatrick and Pape 2013; Winters et al. 2008). The general approach used in the ArboMAP system can be extended to new locations and different diseases and improved upon by testing alternative sources of data and novel predictive modeling techniques.

Acknowledgments

This work was funded by awards from the National Aeronautics and Space Administration (NSSC19K1233 and NN15AF74G, to M.C.W.).

References

Abatzoglou JT. 2013. Development of gridded surface meteorological data for ecological applications and modelling. Int J Climatol 33(1):121–131, https://doi.org/10.1002/joc.3413.

Albright TP, Pidgeon AM, Rittenhouse CD, Clayton MK, Wardlow BD, Flather CH, et al. 2010. Combined effects of heat waves and droughts on avian communities across the conterminous United States. Ecosphere 1(5):art12, https://doi.org/10.1890/ES10-00057.1.

Allan BF, Langerhans RB, Ryberg WA, Landesman WJ, Griffin NW, Katz RS, et al. 2009. Ecological correlates of risk and incidence of West Nile virus in the United States. Oecologia 158(4):699–708, PMID: 18941794, https://doi.org/10.1007/s00442-008-1169-9.

Barker CM. 2019. Models and surveillance systems to detect and predict West Nile virus outbreaks. J Med Entomol 56(6):1508–1515, PMID: 31549727, https://doi.org/10.1093/jme/tjz150.

Braga ALF, Zanobetti A, Schwartz J. 2001. The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities. J Occup Environ Med 43(11):927–933, PMID: 11725331, https://doi.org/10.1097/00043764-200111000-00001.

Burkett-Cadena ND, Hassan HK, Eubanks MD, Cupp EW, Unnasch TR. 2012. Winter severity predicts the timing of host shifts in the mosquito *Culex erraticus*. Biol Lett 8(4):567–569, PMID: 22399787, https://doi.org/10.1098/rsbl.2012.0075.

Busch MP, Kleinman SH, Tobler LH, Kamel HT, Norris PJ, Walsh I, et al. 2008. Virus and antibody dynamics in acute West Nile virus infection. J Infect Dis 198(7):984–993, PMID: 18729783, https://doi.org/10.1086/591467.

CDC (Centers for Disease Control and Prevention). 2021. West Nile Virus. Statistics & Maps. https://www.cdc.gov/westnile/statsmaps/ [accessed 1 August 2021].

Chuang TW, Hildreth MB, Vanroekel DL, Wimberly MC. 2011. Weather and land cover influences on mosquito populations in Sioux Falls, South Dakota. J Med Entomol 48(3):669–679, PMID: 21661329, https://doi.org/10.1603/me10246.

Chuang TW, Hockett CW, Kightlinger L, Wimberly MC. 2012. Landscape-level spatial patterns of West Nile virus risk in the northern Great Plains. Am J Trop Med Hyg 86(4):724–731, PMID: 22492161, https://doi.org/10.4269/ajtmh.2012.11-0515.

Chuang TW, Wimberly MC. 2012. Remote sensing of climatic anomalies and West Nile virus incidence in the northern Great Plains of the United States. PLoS One 7(10):e46882, PMID: 23071656, https://doi.org/10.1371/journal.pone.0046882.

Davis JK, Vincent G, Hildreth MB, Kightlinger L, Carlson C, Wimberly MC. 2017. Integrating environmental monitoring and mosquito surveillance to predict vector-borne disease: prospective forecasts of a West Nile virus outbreak. PLoS Curr 9:ecurrents.outbreaks.90e80717c4e67e1a830f17feeaaf85de, PMID: 28736681, https://doi.org/10.1371/currents.outbreaks.90e80717c4e67e1a830f17feeaaf85de.

Davis JK, Vincent GP, Hildreth MB, Kightlinger L, Carlson C, Wimberly MC. 2018. 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 185:242–250, PMID: 29727611, https://doi.org/10.1016/j.actatropica.2018.04.028.

DeFelice NB, Birger R, DeFelice N, Gagner A, Campbell SR, Romano C, et al. 2019. Modeling and surveillance of reporting delays of mosquitoes and humans infected with West Nile virus and associations with accuracy of West Nile virus forecasts. JAMA Netw Open 2(4):e193175, PMID: 31026036, https://doi.org/10.1001/jamanetworkopen.2019.3175.

DeFelice NB, Little E, Campbell SR, Shaman J. 2017. Ensemble forecast of human West Nile virus cases and mosquito infection rates. Nat Commun 8:14592, PMID: 28233783, https://doi.org/10.1038/ncomms14592.

DeFelice NB, Schneider ZD, Little E, Barker C, Caillouet KA, Campbell SR, et al. 2018. Use of temperature to improve West Nile virus forecasts. PLoS Comput Biol 14(3):e1006047, PMID: 29522514, https://doi.org/10.1371/journal.pcbi.1006047.

Dickinson K, Paskewitz S. 2012. Willingness to pay for mosquito control: how important is West Nile virus risk compared to the nuisance of mosquitoes? Vector Borne Zoonotic Dis 12(10):886–892, PMID: 22651384, https://doi.org/10.1089/vbz.2011.0810.

Dunphy BM, Kovach KB, Gehrke EJ, Field EN, Rowley WA, Bartholomay LC, et al. 2019. Long-term surveillance defines spatial and temporal patterns implicating *Culex tarsalis* as the primary vector of West Nile virus. Sci Rep 9(1):6637, PMID: 31036953, https://doi.org/10.1038/s41598-019-43246-y.

Gardner AM, Hamer GL, Hines AM, Newman CM, Walker ED, Ruiz MO. 2012. Weather variability affects abundance of larval *Culex* (Diptera: Culicidae) in storm water catch basins in suburban Chicago. J Med Entomol 49(2):270–276, PMID: 22493843, https://doi.org/10.1603/me11073.

Gasparrini A, Armstrong B, Kenward MG. 2010. Distributed lag non-linear models. Stat Med 29(21):2224–2234, PMID: 20812303, https://doi.org/10.1002/sim.3940.

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27, https://doi.org/10.1016/j.rse.2017.06.031.

Hadler JL, Patel D, Nasci RS, Petersen LR, Hughes JM, Bradley K, et al. 2015. Assessment of arbovirus surveillance 13 years after introduction of West Nile virus, United States. Emerg Infect Dis 21(7):1159–1166, PMID: 26079471, https://doi.org/10.3201/eid2107.140858.

Hahn MB, Monaghan AJ, Hayden MH, Eisen RJ, Delorey MJ, Lindsey NP, et al. 2015. Meteorological conditions associated with increased incidence of West Nile virus disease in the United States, 2004–2012. Am J Trop Med Hyg 92(5):1013–1022, PMID: 25802435, https://doi.org/10.4269/ajtmh.14-0737.

Hess A, Davis JK, Wimberly MC. 2018. Identifying environmental risk factors and mapping the distribution of West Nile virus in an endemic region of North America. Geohealth 2(12):395–409, PMID: 32159009, https://doi.org/10.1029/2018GH000161.

Johnson BJ, Munafo K, Shappell L, Tsipoura N, Robson M, Ehrenfeld J, et al. 2012. The roles of mosquito and bird communities on the prevalence of West Nile virus in urban wetland and residential habitats. Urban Ecosyst 15(3):513–531, PMID: 25484570, https://doi.org/10.1007/s11252-012-0248-1.

Karki S, Brown WM, Uelmen J, Ruiz MO, Smith RL. 2020. The drivers of West Nile virus human illness in the Chicago, Illinois, USA area: fine scale dynamic effects of weather, mosquito infection, social, and biological conditions. PLoS One 15(5):e0227160, PMID: 32437363, https://doi.org/10.1371/journal.pone.0227160.

Karki S, Hamer GL, Anderson TK, Goldberg TL, Kitron UD, Krebs BL, et al. 2016. Effect of trapping methods, weather, and landscape on estimates of the *Culex* vector mosquito abundance. Environ Health Insights 10:93–103, PMID: 27375359, https://doi.org/10.4137/EHI.S33384.

Keyel AC, Elison Timm O, Backenson PB, Prussing C, Quinones S, McDonough KA, et al. 2019. Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in *Culex* mosquitoes and human case counts in New York and Connecticut. PLoS One 14(6):e0217854, PMID: 31158250, https://doi.org/10.1371/journal.pone.0217854.

Keyel AC, Gorris ME, Rochlin I, Uelmen JA, Chaves LF, Hamer GL, et al. 2021. 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 15(9):e0009653, PMID: 34499656, https://doi.org/10.1371/journal.pntd.0009653.

Kightlinger L. 2017. West Nile review: 15 years of human disease in South Dakota, 2002–2016. S D Med 70(8):346–351, PMID: 28813740.

Kilpatrick AM, Pape WJ. 2013. Predicting human West Nile virus infections with mosquito surveillance data. Am J Epidemiol 178(5):829–835, PMID: 23825164, https://doi.org/10.1093/aje/kwt046.

Kramer LD, Ciota AT, Kilpatrick AM. 2019. Introduction, spread, and establishment of West Nile virus in the Americas. J Med Entomol 56(6):1448–1455, PMID: 31549719, https://doi.org/10.1093/jme/tjz151.

Kramer LD, Styer LM, Ebel GD. 2008. A global perspective on the epidemiology of West Nile virus. Annu Rev Entomol 53:61–81, PMID: 17645411, https://doi.org/10.1146/annurev.ento.53.103106.093258.

Kwan JL, Kluh S, Reisen WK. 2012. Antecedent avian immunity limits tangential transmission of West Nile virus to humans. PLoS One 7(3):e34127, PMID: 22457819, https://doi.org/10.1371/journal.pone.0034127.

Lanciotti RS, Kerst AJ, Nasci RS, Godsey MS, Mitchell CJ, Savage HM, et al. 2000. Rapid detection of West Nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. J Clin Microbiol 38(11):4066–4071, PMID: 11060069, https://doi.org/10.1128/JCM.38.11.4066-4071.2000.

Manore CA, Davis JK, Christofferson RC, Wesson DM, Hyman JM, Mores CN. 2014. Towards an early warning system for forecasting human West Nile virus incidence. PLoS Curr 6:ecurrents.outbreaks.ed6f0f8a61d20ae5f32aaa5c2b8d3c23, PMID: 24611126, https://doi.org/10.1371/currents.outbreaks.ed6f0f8a61d20ae5f32aaa5c2b8d3c23.

Marcantonio M, Metz M, Baldacchino F, Arnoldi D, Montarsi F, Capelli G, et al. 2016. First assessment of potential distribution and dispersal capacity of the emerging invasive mosquito *Aedes koreicus* in Northeast Italy. Parasit Vectors 9:63, PMID: 26842546, https://doi.org/10.1186/s13071-016-1340-9.

McDonald E, Mathis S, Martin SW, Erin Staples J, Fischer M, Lindsey NP. 2021. Surveillance for West Nile virus disease—United States, 2009–2018. Am J Transplant 21(5):1959–1974, PMID: 33939278, https://doi.org/10.1111/ajt.16595.

Nasci RS, Mutebi JP. 2019. Reducing West Nile virus risk through vector management. J Med Entomol 56(6):1516–1521, PMID: 31549724, https://doi.org/10.1093/jme/tjz083.

Paull SH, Horton DE, Ashfaq M, Rastogi D, Kramer LD, Diffenbaugh NS, et al. 2017. Drought and immunity determine the intensity of West Nile virus epidemics and climate change impacts. Proc Biol Sci 284(1848):20162078, PMID: 28179512, https://doi.org/10.1098/rspb.2016.2078.

Paz S. 2019. Effects of climate change on vector-borne diseases: an updated focus on West Nile virus in humans. Emerg Top Life Sci 3(2):143–152, PMID: 33523144, https://doi.org/10.1042/ETLS20180124.

Petersen LR, Brault AC, Nasci RS. 2013. West Nile virus: review of the literature. JAMA 310(3):308–315, PMID: 23860989, https://doi.org/10.1001/jama.2013.8042.

Reisen WK. 2013. Ecology of West Nile virus in North America. Viruses 5(9):2079–2105, PMID: 24008376, https://doi.org/10.3390/v5092079.

Ronca SE, Murray KO, Nolan MS. 2019. Cumulative incidence of West Nile virus infection, continental United States, 1999–2016. Emerg Infect Dis 25(2):325–327, PMID: 30666940, https://doi.org/10.3201/eid2502.180765.

Rudolph KE, Lessler J, Moloney RM, Kmush B, Cummings DAT. 2014. Incubation periods of mosquito-borne viral infections: a systematic review. Am J Trop Med Hyg 90(5):882–891, PMID: 24639305, https://doi.org/10.4269/ajtmh.13-0403.

Schwartz J. 2000. The distributed lag between air pollution and daily deaths. Epidemiology 11(3):320–326, PMID: 10784251, https://doi.org/10.1097/00001648-200005000-00016.

Shaman J, Day JF, Komar N. 2010. Hydrologic conditions describe West Nile virus risk in Colorado. Int J Environ Res Public Health 7(2):494–508, PMID: 20616987, https://doi.org/10.3390/ijerph7020494.

Shand L, Brown WM, Chaves LF, Goldberg TL, Hamer GL, Haramis L, et al. 2016. Predicting West Nile virus infection risk from the synergistic effects of rainfall and temperature. J Med Entomol 53(4):935–944, PMID: 27113111, https://doi.org/10.1093/jme/tjw042.

Shave A, Garroway CJ, Siegrist J, Fraser KC. 2019. Timing to temperature: egg-laying dates respond to temperature and are under stronger selection at northern latitudes. Ecosphere 10(12):e02974, https://doi.org/10.1002/ecs2.2974.

Shocket MS, Verwillow AB, Numazu MG, Slamani H, Cohen JM, El Moustaid F, et al. 2020. Transmission of West Nile and five other temperate mosquito-borne viruses peaks at temperatures between 23°C and 26°C. Elife 9:e58511, PMID: 32930091, https://doi.org/10.7554/eLife.58511.

Skaff NK, Cheng Q, Clemesha RES, Collender PA, Gershunov A, Head JR, et al. 2020. Thermal thresholds heighten sensitivity of West Nile virus transmission to changing temperatures in coastal California. Proc Biol Sci 287(1932):20201065, PMID: 32752986, https://doi.org/10.1098/rspb.2020.1065.

Smith KH, Tyre AJ, Hamik J, Hayes MJ, Zhou Y, Dai L. 2020. Using climate to explain and predict West Nile virus risk in Nebraska. Geohealth 4(9):e2020GH000244, PMID: 32885112, https://doi.org/10.1029/2020GH000244.

So J, Kim S, Cohen H. 2017. Message fatigue: conceptual definition, operationalization, and correlates. Commun Monogr 84(1):5–29, https://doi.org/10.1080/03637751.2016.1250429.

Uelmen JA, Irwin P, Bartlett D, Brown W, Karki S, Ruiz MO, et al. 2021. Effects of scale on modeling West Nile virus disease risk. Am J Trop Med Hyg 104(1):151–165, PMID: 33146116, https://doi.org/10.4269/ajtmh.20-0416.

Vincent GP, Davis JK, Wittry MJ, Wimberly MC, Carlson CD, Patton DL, et al. 2020. Epidemic West Nile virus infection rates and endemic population dynamics among South Dakota mosquitoes: a 15-yr study from the United States northern Great Plains. J Med Entomol 57(3):862–871, PMID: 31799615, https://doi.org/10.1093/jme/tjz231.

Wimberly MC, Giacomo P, Kightlinger L, Hildreth MB. 2013. Spatio-temporal epidemiology of human West Nile virus disease in South Dakota. Int J Environ Res Public Health 10(11):5584–5602, PMID: 24173141, https://doi.org/10.3390/ijerph10115584.

Wimberly MC, Hildreth MB, Boyte SP, Lindquist E, Kightlinger L. 2008. Ecological niche of the 2003 West Nile virus epidemic in the northern Great Plains of the United States. PLoS One 3(12):e3744, PMID: 19057643, https://doi.org/10.1371/journal.pone.0003744.

Wimberly MC, Lamsal A, Giacomo P, Chuang TW. 2014. Regional variation of climatic influences on West Nile virus outbreaks in the United States. Am J Trop Med Hyg 91(4):677–684, PMID: 25092814, https://doi.org/10.4269/ajtmh.14-0239.

Winters AM, Bolling BG, Beaty BJ, Blair CD, Eisen RJ, Meyer AM, et al. 2008. Combining mosquito vector and human disease data for improved assessment of spatial West Nile virus disease risk. Am J Trop Med Hyg 78(4):654–665, PMID: 18385365, https://doi.org/10.4269/ajtmh.2008.78.654.

Zou S, Foster GA, Dodd RY, Petersen LR, Stramer SL. 2010. West Nile fever characteristics among viremic persons identified through blood donor screening. J Infect Dis 202(9):1354–1361, PMID: 20874087, https://doi.org/10.1086/656602.