A spatially resolved and environmentally informed forecast model of West Nile virus in Coachella Valley, California.

Matthew J. Ward[1], Meytar Sorek-Hamer[2], Jennifer A. Henke[3], Eliza Little[4], Aman Patel[1], Jeffery Shaman[5,6], Krishna Vemuri[1], and Nicholas B. DeFelice[1]

[1]Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, NY, NY 10029. [2]Universities Space Research Association (USRA) at NASA Ames Research Center, Moffett Field, CA 94043. [3]Coachella Valley Mosquito & Vector Control District, Indio, CA 92201. [4]Connecticut Department of Public Health, Hartford, CT 06103. [5]Columbia Climate School, New York, NY 10025. [6]Mailman School of Public Health, New York, NY 10032.

Corresponding author: Matthew J. Ward, PhD, MSPH (matthew.ward@mssm.edu)

Key Points:

- Environmentally informed West Nile virus forecast model.
- Our environmentally informed early season forecast shows that a dry winter, followed by a wet, warm spring, and a cool summer promotes WNV.
- Early season forecasts are a potential decision tool to inform public health and mosquito abatement intervention.

Abstract

West Nile virus is the most significant arbovirus in the United States in terms of both morbidity and mortality. West Nile exists in a complex transmission cycle between avian hosts and the arthropod vector, *Culex spp.* mosquitoes. Human spillover events occur when humans are in close proximity to vector populations with high rates of infection. Predicting these rates of infection and therefore the risk to humans is not straightforward. In this study, we evaluate the hydrological and meteorological drivers associated with mosquito biology and viral development to determine if these associations can be used to forecast seasonal West Nile risk in the Coachella Valley of California. To test this, we developed and tested a spatially-resolved ensemble forecast model of West Nile virus transmission in the Coachella Valley using 17 years of mosquito surveillance data and NLDAS-2 environmental data. Our multi-model inference system indicated that the combination of a cooler and dryer winter followed by a wetter and warmer spring and a cooler than normal summer was most predictive of West Nile positive mosquitoes in the Coachella Valley. The ability to make accurate early season predictions of West Nile risk could allow local abatement districts and public health entities to implement early season interventions such as targeted adulticiding and public messaging before human transmission occurs. Such early and targeted interventions could better mitigate the risk of West Nile virus to humans in the Coachella Valley.

1 Introduction

The largest county-level arboviral neuroinvasive disease outbreak ever recorded in the United States occurred in a desert climate in 2021 with 1,427 recorded human West Nile virus (WNV) neuroinvasive cases resulting in 100 deaths. The extent of this outbreak was over 1.5 times greater than the total neuroinvasive cases reported in Arizona between 1999 - 2016 but only represent a fraction of the overall infections (Ronca et al., 2019). However these 1,427 cases do capture the high morbidity related to this WNV outbreak, which based on previous studies may be estimated to have cost nearly one billion dollars in hospitalizations, follow-up care, and work lost (Ronca et al., 2019). The morbidity and mortality of this outbreak emphasizes how critical it is to gain a better understanding of the environmental drivers associated with the complex transmission cycle between avian reservoir hosts and Culex genus mosquitoes, which can result in incidental zoonotic spillover to humans (Colpitts et al., 2012; Nasci & Mutebi, 2019).

WNV transmission is not only dependent on factors such as bird immunity, the success of mosquito overwintering, and mosquito feeding behavior but is also substantially driven by environmental drivers such as meteorological and hydrological conditions (Davis et al., 2017; DeFelice et al., 2017; Kilpatrick et al., 2006; Paull et al., 2017; Shaman et al., 2005; Wimberly et al., 2022). Much work has been undertaken to produce accurate forecast models for WNV transmission, however, there remains significant variability and little consensus between these products (Barker, 2019; DeFelice et al., 2017; Keyel et al., 2021; Little et al., 2016; Wimberly et al., 2022). Temperature, humidity, and available water affect the development and survival of WNV mosquito vectors (sp. *Culex*), along with the extrinsic incubation period (EIP) of the virus (Epstein, 2001; Reisen et al., 2008; Shaman et al., 2005; Wegbreit & Reisen, 2000). Warmer temperatures accelerate population growth by providing conditions suitable for earlier season larval development, completion of the gonotrophic cycle, and shortened EIP of the virus (Ciota

& Kramer, 2013). Conversely, extreme temperature events can negatively affect mosquito population growth resulting in slower development at low temperatures and significant mortality at high temperatures (Mordecai et al., 2019). Temperature driven shifts in mosquito activity and WNV infections have been observed in areas that experience seasonal temperature variations, such as coastal Los Angeles, where during the extreme hot periods of summer WNV incidence increases in the Coastal Zones that during the rest of the year are typically too cool for transmission (Skaff et al., 2020).

Additionally, the availability of suitable aquatic breeding sites is equally important for mosquito population success, but less conclusive is how to most effectively measure this habitat. Precipitation has been demonstrated to have contradictory associations with WNV incidence. On one hand, heavy rainfall may increase available water for mosquito breeding sites, but it may also wash out more sustained breeding habitats, such as ditches or catchment basins favored by *Culex spp.* (Koenraadt & Harrington, 2008). Conversely, drought periods increase the proximity and contact between infected mosquitoes and avian hosts, accelerating enzootic amplification of WNV throughout the mosquito population and increasing the risk of spillover to humans (Shaman et al., 2005). Using precipitation as an indicator for mosquito population success also presents a challenge in arid desert environments with little rainfall. However, evapotranspiration (ET), the measure of the amount of water transferred from the ground or plants to the atmosphere, offers a unique measure of the available water in an ecosystem (Fisher et al., 2020). ET may be a better hydrological indicator of mosquito population success and viral amplification in a desert region where the availability of surface water is primarily driven by agricultural and recreational practices rather than precipitation.

Warmer, wetter environments favor mosquito development, whereas dryer than usual conditions in a geographic location may favor WNV amplification and transmission between the mosquito vector and bird hosts. However, unusual weather patterns must happen at specific points in a mosquito population's development to result in WNV amplification in mosquitoes and spillover events to humans. Lastly, the effect of these environmental deviations on the mosquito infection rate and therefore the potential risk of a human WNV outbreak may not be apparent until weeks or months after the environmental events. We hypothesize that extreme heat events seen in Coachella Valley (CV) during the summer months have the effect of depressing the mosquito population, thereby slowing WNV transmission, and that cooler-than-average summers lead to increased rates of WNV transmission. Additionally, we postulate that changes to the late-season climatology of the CV may play a role in the magnitude of WNV transmission the following year.

In this study, we evaluate the hydrological and meteorological drivers that are associated with mosquito biology and viral development to understand if these associations may be used to forecast seasonal WNV risk. We developed and tested a spatially resolved ensemble forecast model of WNV transmission in the CV using 17 years of mosquito surveillance data and environmental data from the North American Land Data Assimilation System (NLDAS) - 2. We then cross-validated the ensemble model using a leave-one-year-out strategy and produced real-time monthly forecasts for 2022 with analyses of agreement with observations at the end of the season. Our work indicates that we can accurately forecast annual mosquito infection rates using early-season meteorological conditions. The ability to make accurate early season predictions of

West Nile risk could better mitigate the risk of West Nile virus to humans in the Coachella Valley by allowing local public health entities to implement early season interventions such as targeted adulticiding, larvaciding and public messaging before human transmission occurs.

2 Materials and Methods

2.1 Study Location

This study was carried out in the Coachella Valley of California (CV). CV has unique ecotypes that vary across a North-South gradient with the Salton Sea and its associated marsh lands to the South, agricultural and recreational land along the length of the valley, and desert to the North. The center of the valley is largely agricultural with urban areas intermixed. Hydrology and, by association, viable mosquito habitat, are dictated by agricultural and recreational water use. This is evident by the geographic seasonality of WNV in the CV which has historically followed a South to North track - originating around the Salton Sea then distributing Northward along irrigated agricultural land and golf courses (Figure 1).

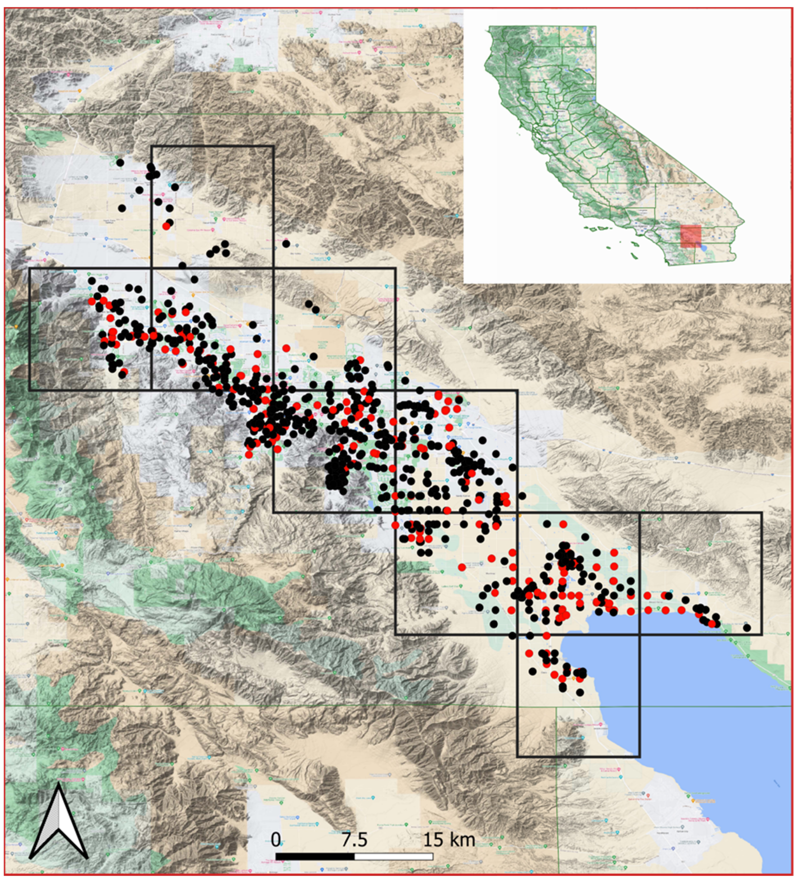

Figure 1. NLDAS grid cells with all historical trap locations that have ever been positive (red) and negative (black) overlaid the Coachella Valley. Insert: Location of Coachella Valley within California.

2.3 Meteorological and Hydrological Data

We used meteorological variables from the North American Land Data Assimilation System (NLDAS) - 2 (https://ldas.gsfc.nasa.gov/nldas/nldas-get-data) for CV. Correlation testing was done on 10 monthly estimates of environmental conditions (Table 1, Figure S1). August and September of the year in question were omitted as these months were the months we were most interested in predicting. We chose monthly estimates of atmospheric temperature (ATMP) [Kelvin, 2 m above ground level] and the mosaic hydrology model simulation estimates of evapotranspiration (ET) [kg/m^2, monthly accumulated] with a spatial resolution of 0.125° (≈13 km^2 grid cells) creating ten grid cells over CV (Figure 1).

Table 1. Monthly NLDAS environmental variables.

Environmental Variable	NLDAS Abbreviation	Scale/Unit
Total precipitation	apcpsfc	$\frac{kg}{m^2}$
0 - 40 cm soil moisture	soilm40	$\frac{kg}{m^2}$
0 - 10 cm soil moisture	soilm10	$\frac{kg}{m^2}$
Surface pressure	pressfc	Pa
2 m above ground specific humidity	spfh2m	$\frac{kg}{kg}$
Average surface skin temperature	avsftsfc	K
2 m above ground temperature	tmp2m	K
Total evapotranspiration	evpsfc	$\frac{kg}{m^2}$

2.3 Mosquito Data

Mosquito data was obtained for 17 years (2006 - 2022) from the Coachella Valley Mosquito and Vector Control District. Trap data was analyzed to evaluate mosquito population abundance per trap night by CDC week. These data were averaged over the CV. To understand the annual WNV infection rate in the mosquito population, the control district tests trapped mosquitoes in pools of one – 50.

We determined infection rates of mosquitoes at the NLDAS scale using the presence/absence of mosquito pools and a statistical method [maximum likelihood estimator (MLE)] to estimate the annual infection rate of mosquitoes per 1,000 mosquitoes tested (I_M). Briefly, all traps were assigned an NLDAS grid, all pools within each grid were combined from CDC weeks 20 - 45, and the annual infection rate was calculated (Table S1.). For methodology on calculating the annual I_M using the MLE see (Ward et al., 2023). The Centers for Disease Control and

Prevention's Morbidity and Mortality Weekly Report (CDC week) was used as a temporal reference for this study.

To address the uncertainty of the observed WNV mosquito infection rates, we set a threshold of a minimum of 500 mosquitoes tested within a year to establish a valid annual observation (DeFelice et al., 2017). This resulted in the inclusion of ten NLDAS grids (Figure 1). Data were also evaluated for seasonality and completeness before being included as a valid observation. Seasonality was addressed to include 97.5% of all positive mosquito pools historically were collected in the Coachella Valley, CDC weeks 20 - 45 (Figure S2). These weeks also correspond to when the daily low temperature exceeds (14.3 ºC), high enough for viral amplification and 3 weeks after the daily low temperature goes below (Figure 2) (Reisen et al., 2006). I_M was calculated for each of these NLDAS grids for which there were both meteorological and hydrological data available. Aggregating mosquito data by NLDAS grid cell discounts more local scale environmental factors that may bias trap collections and allows for analysis of how climate conditions influence relative mosquito infection rates.

2.4 Model selection

We built a hierarchical negative binomial model with annual infection rates at the NLDAS grid cell scale as the outcome variable and standardized monthly ATMP and ET as independent variables. The hierarchical negative binomial model was chosen due to high zero-inflation within the mosquito infection rate data (Figure S3). We included random intercepts for each NLDAS grid cell, and the annual infection rate of mosquitoes quantified using a maximum likelihood of all pools collected over the year for each grid cell. We used ATMP and ET data from November - December of the year prior, and from January to July of the current year, as independent variables in the regression.

Models where all variables were significant were selected for use in an ensemble, and this inference system was then calibrated using data from 2006 - 2018. Retrospective forecasts were then generated and compared against the 2019 - 2021 observations. Forecasts were generated using early-season environmental observations to predict areas of concern, which were defined as grids with I_M greater than one infected mosquito per 1,000 tested, which equates to ≥ the 75th percentile of the historical annual infection rate.

To identify the best combination of models to include in our environmentally informed forecast system, we developed a multi-model inference system to identify the combination of environmental conditions over time and how these combinations of environmental events are associated with WNV mosquito infection rates. Equations 1-3 are used to identify the best combination of environmental parameters and calculate model-averaged predictions with unconditional confidence intervals. Specifically, we identified the best models using whole model goodness-of-fit estimated from the second order estimation of the Akaike Information Criterion (AICc) which is a better estimation of model fit when the ratio of parameters (n) to observations (k) is small $\left(\frac{n}{k} < 40\right)$ (Burnham & Anderson, 2002). Following evaluation of numerous model structures, a mixed effects negative binomial model with grid cell as the random effect produced the best AICc and was used to make predictions with four environmental variables of ATMP and ET in each model.

The ensemble defined as the weighted average of the combination of the best fitting models, is used to improve forecast accuracy and account for the uncertainty of their competing predictions. To rank goodness-of-fit among the models tested, we calculated an AICc score and the weight of model i, ω_i, relative to the best model, equation 1:

$$\omega_i = \frac{e^{-\frac{1}{2}\Delta_i}}{\sum_{i=1}^{R} e^{-\frac{1}{2}\Delta_i}} \quad (1)$$

where $\Delta_i = AICc_i - AICc_{min}$, $AICc_{min}$ is the AICc of the best-fit model, and R is the number of models where all parameters were statistically significant. We used a subset (N) of models whose weights summed to 0.95, and after identifying the combination of best fitting models, the Akaike weights were re-normalized to sum to 1.

The model averaged prediction was generated from a weighted average, equation 2:

$$\bar{\hat{\theta}} = \sum_{i=1}^{N} \tilde{\omega}_i \cdot \hat{\theta}_i \quad (2)$$

where $\hat{\theta}_i$ is the estimate of model i, $\tilde{\omega}_i$ is the re-normalized Akaike weight for model i in the ensemble and $\bar{\hat{\theta}}$ is the mean ensemble prediction calculated from the N best fitting models identified in the step above. The predictions from the models are calculated on a logarithmic scale. We first converted them to the natural scale, obtained the weighted ensemble prediction, and used these values to calculate unconditional model-averaged variance, equation 3:

$$Var = \left[\sum_{i=1}^{N} \tilde{\omega}_i \cdot \sqrt{Var\left(\hat{\theta}_i\right) + \left(\hat{\theta}_i - \bar{\hat{\theta}}\right)^2}\right]^2 \quad (3)$$

Where $Var\left(\hat{\theta}_i\right)$ is the variance of the prediction from model i and $\bar{\hat{\theta}}$ is defined as above.

This unconditional estimator takes into account the variation within and between each model in the model set (i.e. the model selection uncertainty) and was used to estimate unconditional confidence intervals around each model-averaged prediction.

We performed leave-one-year-out (LOYO) temporal cross validation analysis for each year from 2006 - 2018, with outcome data for that year excluded from the input data, and generated predictions for that year based on the ensemble model identified from a large number of candidate models. We identified an ensemble model using data from all years 2006 - 2018 combined, which we used to make predictions retrospectively for 2019, 2020, and 2021. We also applied the same methodology to data from 2006 - 2021 to develop ensemble models to produce real-time monthly predictions for 2022.

We evaluated forecast accuracy by grid cell and year for 2019 to 2021. Forecasts were deemed accurate if a prediction was above or below one infected mosquito per 1,000 tested in each grid cell. One infected mosquito per 1,000 tested annually was defined as high risk for transmission.

All analyses were conducted using the statistical software R (version 4.2.2) and the package 'glmmADMB' to build the hierarchical regression models (RCoreTeam, 2023).

3 Results

Here we present the results of applying an environmentally informed ensemble forecast of annual WNV infection rates in *Culex* mosquitoes at 13 km^2 spatial resolution. Retrospective forecasts were generated for 2019, 2020, and 2021 using data from 2006 to 2018. A forecast was then generated in real-time for Coachella Valley, CA, using a multi-model inference system calibrated using data from 2006 to 2021.

3.1 Mosquito & Climate Data

The collected mosquito population when normalized by trap night, generally exhibits a pronounced bimodal structure peaking in late spring to early summer (~CDC week 16) before drastically declining during the summer (~CDC week 30) during peak maximum daily ATMP and rebounding during the fall (~CDC week36) (Figure 2). Conversely, the I_M peaks during the summer (~CDC week 30) corresponding to the peak in average minimum daily temperature (Figure 2), however annual variability in these trends was observed.

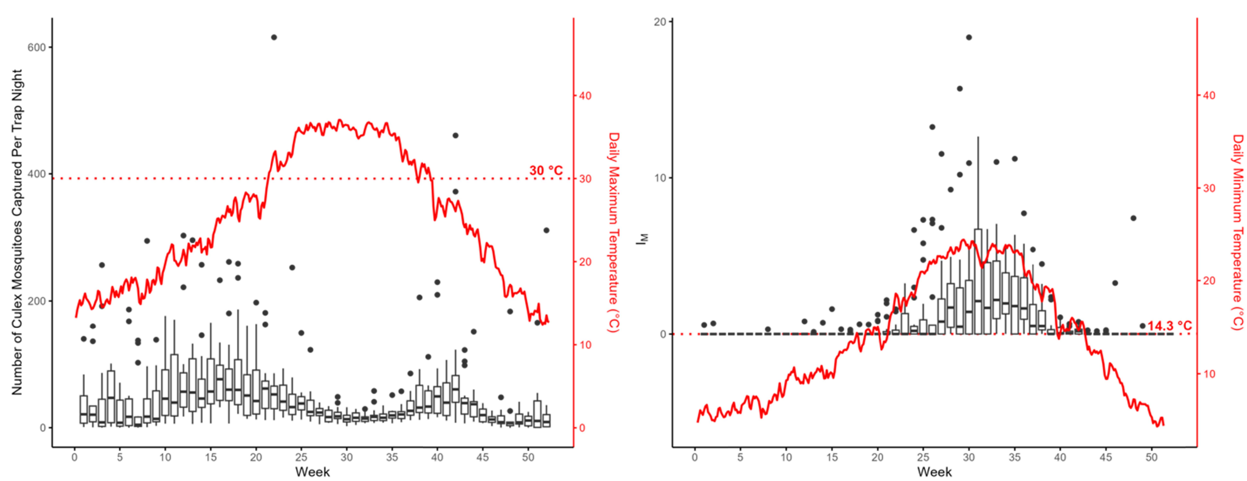

Figure 2. Historical CV *Culex* mosquito abundance, I_M and ATMP between 2006 - 2022. Left: Weekly mean number of mosquitoes trapped per night (boxplot, dots = outliers > 2*SD), maximum daily ATMP (red line), and temperature threshold for mosquito population decline (30 ºC, red dotted line). Right: Average weekly I_M (boxplot, dots = outliers > 2*SD), minimum daily ATMP (red line), and temperature threshold for viral amplification (14.3 ºC, black dashed line) (Reisen et al., 2006).

I_M at the NLDAS grid level for each year from 2006 - 2022 exhibited a relatively consistent pattern of WNV-positive mosquitoes focused near the Salton Sea in the South on typical years with dissemination North through the valley in outbreak years such as 2019 (Figure 3). Correlation analysis of available environmental variables led us to choose one environmental variable for temperature (ATMP) and one for hydrology (ET) for use in our models (Figure S1).

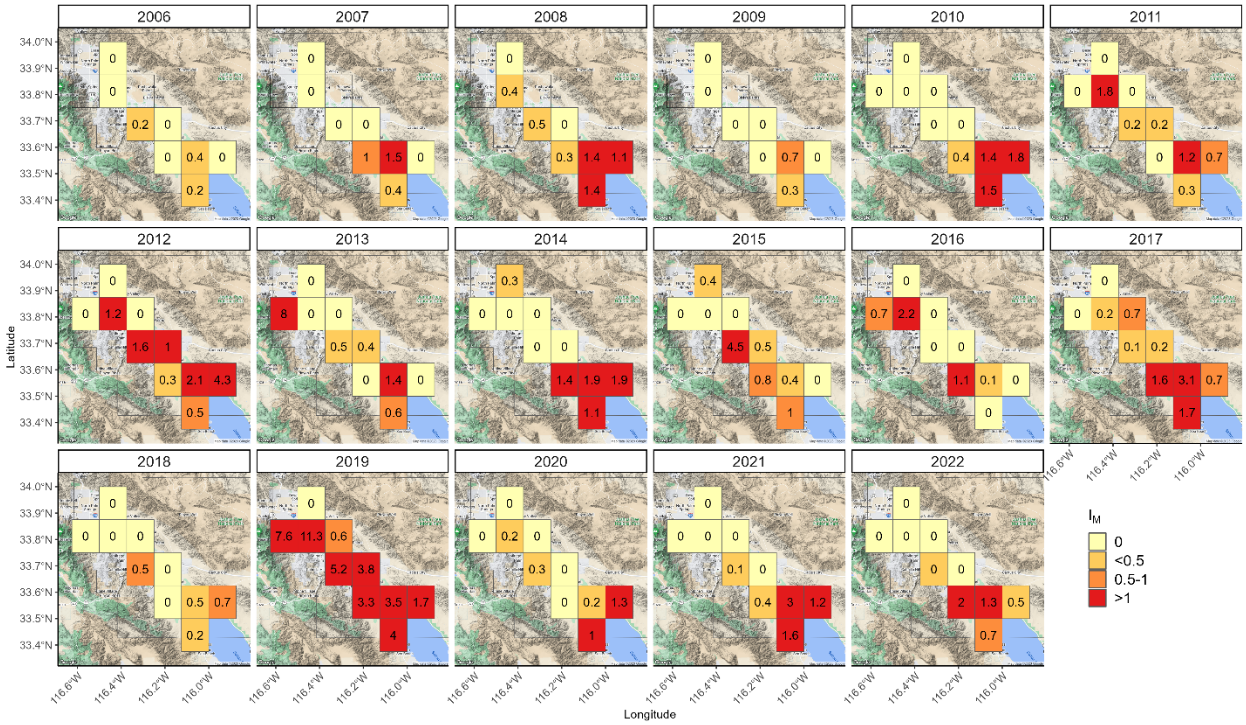

Figure 3. Annual I_M for CDC weeks 20 - 45 by NLDAS grid in CV from 2006 – 2022.

3.2 Model Selection

The regression analysis using all combinations of four independent variables from 2006 - 2021 yielded 104 significant models (0.979% of total models run, N = 10,626) with two models comprising the ensemble with weights summing to 95% (Table 2). The same analysis performed using environmental data from 2006 - 2018 yielded four significant models (0.038% of total models run, N = 10,626), all of which were included in the ensemble (Table 2). Four-predictor models for 2006 - 2021 suggest that a cooler and dryer winter followed by a wetter and warmer spring and a cooler than normal summer are most predictive of WNV I_M in the Coachella Valley (Figure 4).

Table 2. Significant models summing to 95% used in the ensemble models. Effects of each parameter are shown, including month, parameter estimate, and the standard error of the estimate in parentheses.

Model Rank	AICc	Weight	Temperature			Evapotranspiration		
Ensemble models 2006 - 2021								
1	321.32	0.76	April 0.66	July -0.35	-	January -0.53	February 1.05	-

			(0.15)	(0.15)		(0.22)	(0.19)	
2	323.75	0.23	December -0.26 (0.13)	April 0.85 (0.175)	July -0.59 (0.134)	February 0.71 (0.11)	-	-
Ensemble models 2006 - 2018								
1	244.77	0.68	April 0.79 (0.24)	July -0.38 (0.18)	-	January -0.76 (0.33)	February 1.24 (0.39)	-
2	248.19	0.12	May 0.53 (0.23)	June 0.51 (0.19)	July -0.72 (0.22)	February 0.58 (0.22)	-	-
3	248.42	0.11	June 0.44 (0.18)	-	-	January -0.67 (0.29)	February 0.96 (0.32)	June -0.45 (0.18)
4	248.95	0.08	June 0.36 (0.17)	July -0.35 (0.16)	-	February 0.41 (0.18)	June -0.39 (0.19)	-

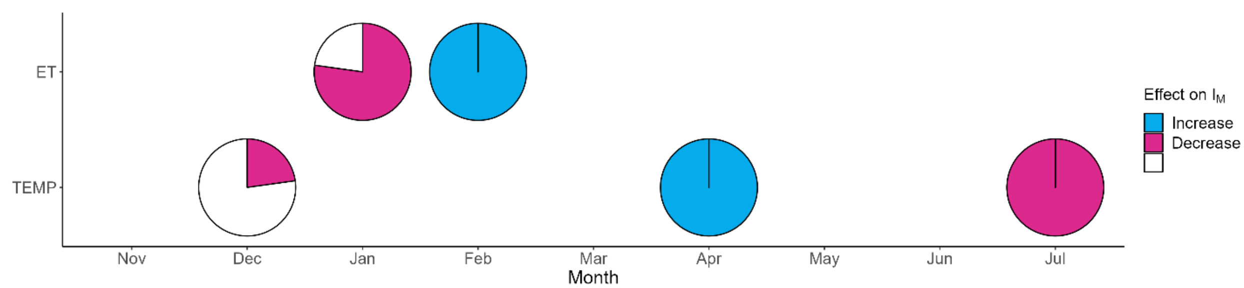

Figure 4. Effect and contribution of ET and ATMP to ensemble model of I_M at the monthly NLDAS scale for 2006 - 2021. Pies indicate the weight and direction of deviation of each environmental variable in the ensemble from the average that increases I_M ($^{+Blue}/_{-Red}$).

3.3 LOYO 2006 – 2021

We found the LOYO model error was small comparable to the omitted years of data, with a Root Mean Square Error (RMSE) of 2 I_M, indicating that the model can predict future years. Additionally, there was no single year that dominated the ensemble and the effect/contribution of the environmental parameters across years was consistent (Figure 5).

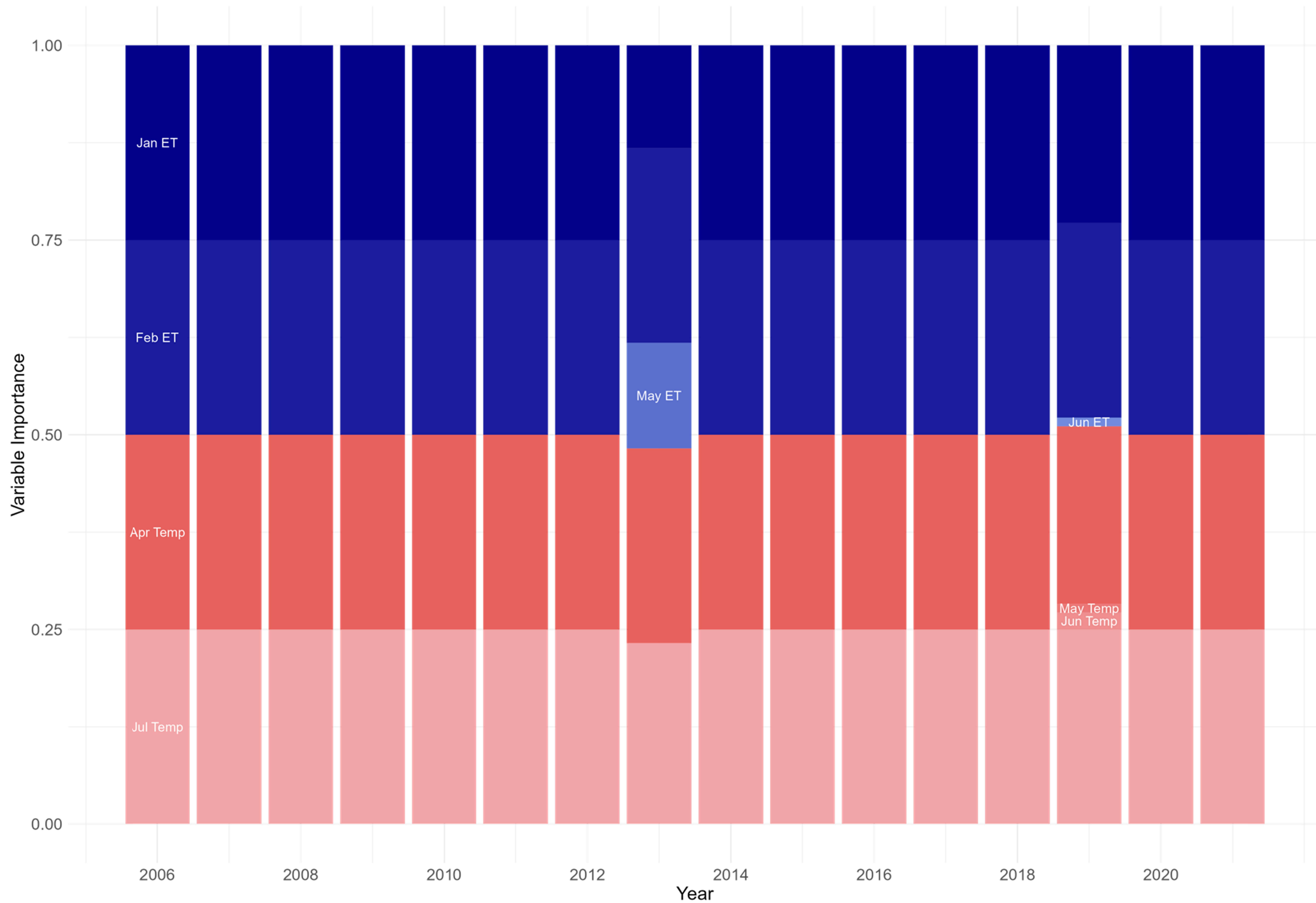

Figure 5. Variable weights contributing to LOYO models for 2006 - 2021 (Red = ATMP, Blue = ET). Year is the annual data removed from the LOYO model.

The sensitivity and specificity of the ensembles' ability to predict above average years ($I_M > 1$; 75th percentile) was 0.81 and 0.57, respectively for 2006 - 2021 (Table 3). This is comparable to the best model which was 0.83 and 0.58 (Table 3). When evaluating the observed vs. predicted LOYO (Figure S4), the ensemble forecast system was better able than the best model to identify years having above average WNV infection rates closer to the observed value.

Table 3. Sensitivity and specificity of observed versus predictions for the best model and the ensemble model for 2006 - 2021.

Observed versus	Sensitivity	Specificity	RMSE (I_M)
Best Model	0.83	0.58	2.39
Ensemble	0.81	0.57	1.94

3.4 Retrospective Forecast

Retrospectively, the ensemble forecast correctly predicted if an areas' annual mosquito infection rate was above or below one infectious mosquito per 1,000 tested 83% of the time from 2019 to 2021 (Figure 6) using the fit model from 2006 to 2018.

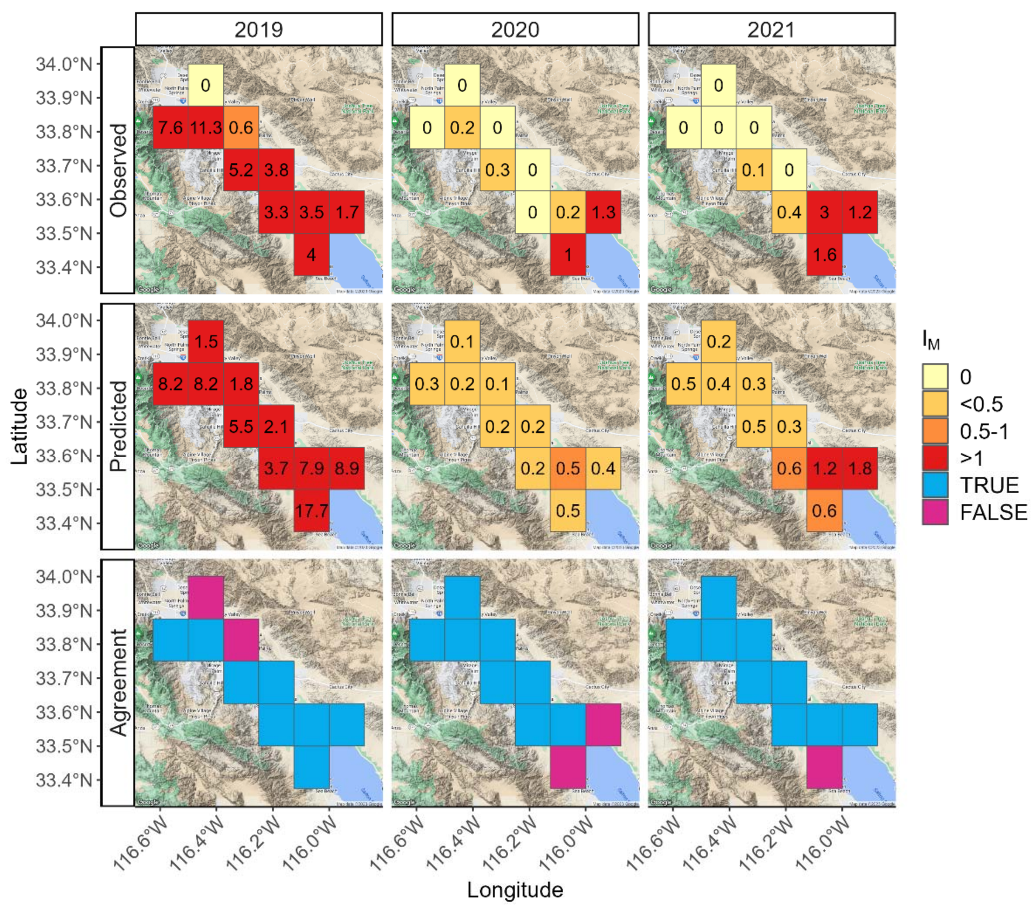

Figure 6. Agreement between observed and predicted I_M stratified by NLDAS grid in the CV for 2019 - 2021. Top Row: Observed I_M in 2019, 2020, and 2021. Middle Row: Predicted I_M in

2019, 2020 and 2021 using a four-predictor model ensemble trained on years 2006 - 2018. Bottom Row: Proportion of cells in agreement with the ensemble model using one infected mosquito per 1,000 tested as a cutoff value/threshold.

Generally, these models indicate that a cooler drier winter followed by a wetting period and a warm spring with a cooler than normal summer, increase the risk of WNV and are the best predictors of WNV rates in the Coachella Valley.

3.5 Real-Time Forecast 2022

We conducted real-time monthly forecasting between March and August 2022 using our four - predictor ensemble model based on data from 2006 - 2021 at the NLDAS grid scale. In 2022, infected mosquito pools were identified in only four grids with three of the four grids becoming positive in July and August (Figure 7, Row 2). Our ensemble model predicted I_M changed in May and August as additional environmental data were added to the system. These predictions agreed with the observed values 80% of the time early in the season (March & April) and 90% of the time for the remainder of the season (Figure 7, Row 4). Furthermore, our model correctly predicted when I_M was greater than one for a 13 km^2 grid cell for 75% of the grid cells that were identified as having WNV during the 2022 season. Lastly, our 2022 forecast, when compared with the observed values, had a sensitivity of 0.5, a specificity of 1, and a RMSE of 1.902 I_M.

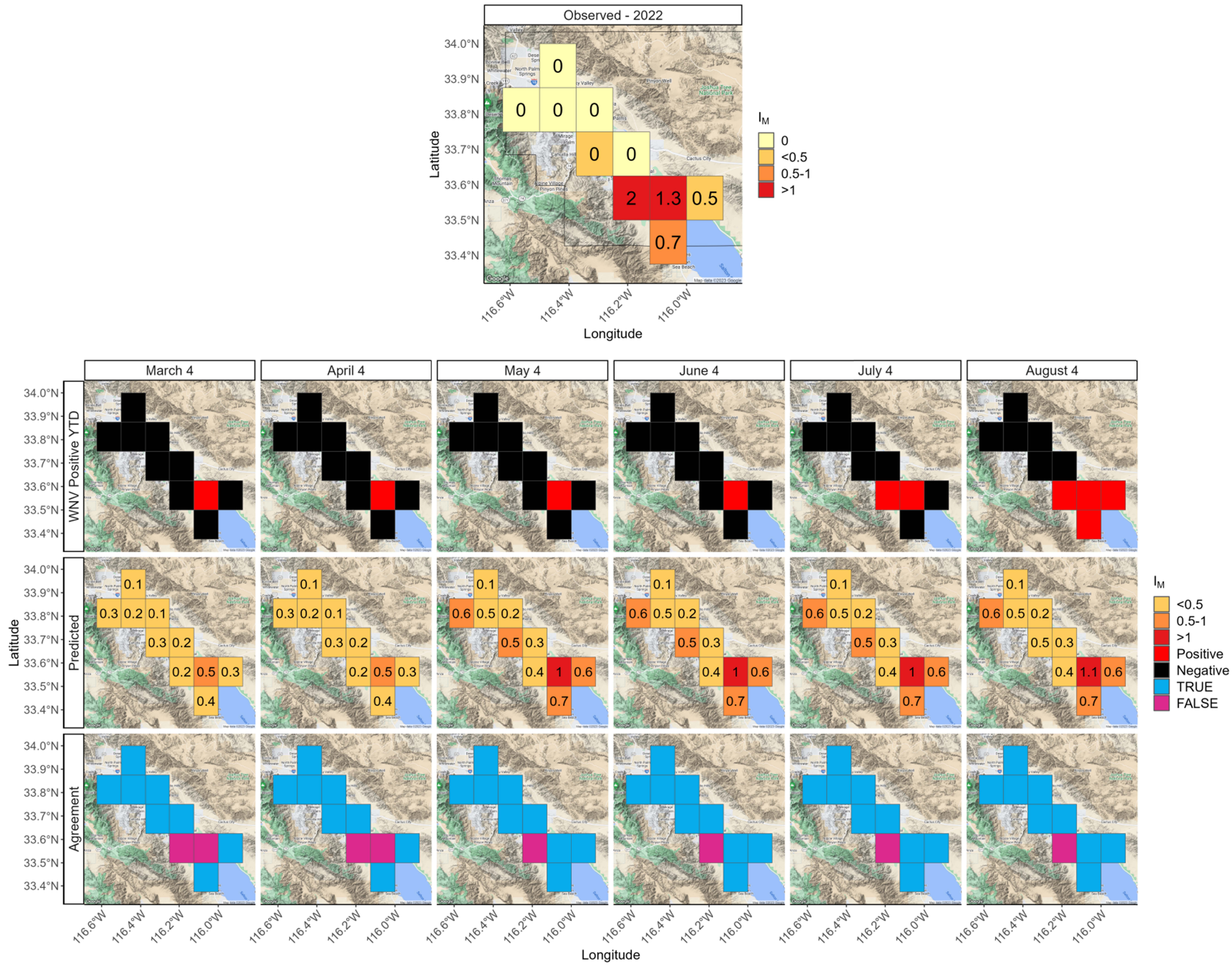
Observed - 2022
Latitude
Longitude
I_M
0
<0.5
0.5-1
>1
March 4
April 4
May 4
June 4
July 4
August 4
WNV Positive YTD
Predicted
Agreement
Positive
Negative
TRUE
FALSE

Figure 7. Real-time forecasting of WNV infected mosquito pools from March - August of 2022 using available NLDAS environmental data (13 km^2) and the four - predictor ensemble model. Row 1: Observed I_M in 2022. Row 2: Grid cells in which WNV was identified in the CV mosquito population year to date (Red = WNV+, Black = WNV not identified). Row 3: Predicted annual I_M. Row 4: Proportion of cells in agreement with the ensemble model using one infected mosquito per 1,000 tested as a cutoff value/threshold.

4 Discussion

This study aims to understand the association between hydrological and meteorological conditions in the desert climate and their relationship with annual mosquito WNV infection rates in the Coachella Valley of California. Our multi-model inference system trained on 13 years of data was able to retrospectively predict I_M greater than one for a 13 km^2 grid cell 83% of the time (2019 - 2021); and in real time for 2022 trained on 16 years of data, was able to predict correctly nine out of 10 grid cells as of May 4th, 2022. This multi-model inference system indicated that the combination of a cooler and dryer winter followed by a wetter and warmer spring and a cooler than normal summer was best predictive of the I_M of WNV in the Coachella Valley.

Desert climates such as those of Southern California or the U.S. Southwest pose unique challenges for arthropods with aquatic larval stages such as mosquitoes. These challenges drive mosquito populations in desert climates to have unique ecologies and population dynamics (Reisen et al., 1992). While mosquitoes benefit from mild winters compared to populations at more northern latitudes, they must weather extreme heat and drought events during the summer months. The hydrology of desert regions means there is far less larval habitat available, and populations are restricted to areas with sufficient available water for larval development, especially during the warmest periods of the year (Chew & Gunstream, 1970; Reisen et al., 1992). Additionally, mosquito populations are further geographically restricted to areas with sufficient blood meals to facilitate egg development. Paradoxically, this geographic restriction concentrates both mosquito and avian hosts in the same location facilitating transmission of WNV between birds and mosquito, and viral amplification in both populations. However, if the summer is too warm it is detrimental to the mosquito population, causing die offs. During these years we see average or even below average annual I_M in the Coachella Valley. Conversely, and as demonstrated by our model, during years where the average temperature during the typically warmest months of the year (July - August) are slightly cooler than normal, mosquito populations are less stressed, and transmission and amplification more readily occurs leading to increased annual I_M in the Coachella Valley.

Additionally, in desert climates, where winters are mild compared to other regions of the United States and where Culex species do not fully diapause, the magnitude of WNV transmission the previous year may also have a residual effect on the following season (Chew & Gunstream, 1970; Reisen et al., 1992). For instance, in the CV the mosquito population exhibits a substantial drop during the hottest part of the summer that is usually followed by a rebound period in the fall going into the winter months. The magnitude of this population recovery may affect the magnitude of WNV transmission the following season as greater numbers of adult mosquitoes successfully overwintering may more rapidly build the population the following season. Furthermore, a cooler than normal winter such as indicated by our model may increase the success of mosquito overwintering by allowing more mosquitoes to quiescence (Diniz et al., 2017). The ensuing rapid population expansion produces more naive mosquitoes that can contribute to the WNV transmission cycle. Similarly, greater numbers of successfully overwintering adult mosquitoes increase the potential that some of those mosquitoes are already infected with WNV and thus able to transmit the virus to naive birds early in the season. These factors may be especially important should environmental conditions be ideal for rapid population growth, such as a wetter and warmer spring.

The ability of our models to accurately predict WNV infection rates at the beginning of May, four months prior to the historical peak of infections in the CV (July - August), provides a valuable decision-making tool for local public health entities and mosquito control districts. Mosquito control intervention in the Coachella Valley is predominantly aerial application of adulticide. The forecast model presented here may provide the abatement district with further spatially refined information on how changes in meteorological and hydrological conditions early in the year may lead to higher WNV I_M. Additionally, it provides empirical data to support early targeted adulticide applications to mitigate WNV amplification and transmission in the valley. Furthermore, early season forecasts could provide public health agencies with the information necessary to target messaging to at-risk populations. This may be especially important because mosquito populations are paradoxically at their lowest when infection rates are at their highest. This lower population may skew people's perception resulting in a perception of lower risk due to fewer nuisance biting mosquitoes. This, in turn, may reduce the proclivity to take protective measures, such as wearing repellant or long clothes, at these times when mosquito infection rates are highest. Early season messaging may help avoid this, but caution must be used, too, not over-saturate residents and cause message fatigue, whereby residents become desensitized to risk-related messaging (Eppler & Mengis, 2002; So et al., 2017).

A major component of industry in the Coachella Valley is agriculture which requires laborers to work outdoors, potentially in proximity to WNV hot spots. Spatially resolved forecasts, such as our multi-model inference system, provide local public health entities additional information on where and when to target educational outreach and risk reduction efforts such as signage, collaborating with local farm-worker organizations, and providing protective measures such as DEET products to those at risk. Additionally, due to migrant status and other social-economic factors such as barriers to care or accessing care in Mexico, it is likely that many, if not a majority of non-neuroinvasive WNV cases go unreported in migrant farm-worker populations (Horton & Cole, 2011; Seid et al., 2003; Villarejo, 2003). Early season, targeted outreach could aid in closing this reporting gap by increasing knowledge about WNV and care-seeking behaviors, while decreasing barriers to care.

Our use of NLDAS data have both advantages and disadvantages. First, NLDAS data are freely downloadable and easily accessible; therefore, this model and ensemble strategy could be applied to other geographical regions. However, data sets of this scale and complexity do require a computational skillset to manipulate and make full use of mosquito trapping data. Additionally, an advantage of NLDAS is its scale. At 13 km^2, the grid cells are suitably large to not only potentially match the scale of aerial abatement interventions, but also are large enough to capture the effects of meteorology on infection rates over the background of land use, which is essential for a climatically driven model.

5 Conclusions

This study showed the potential for accurate prediction of the annual risk of WNV in Coachella Valley at the NLDAS scale utilizing an ensemble model, mosquito pool data, and freely available meteorological data. The ensemble model performed the best and indicated that cooler and dryer than usual winters followed by a wetter and warmer spring with a cooler than usual summer are the sequence of conditions best supporting WNV amplification in *Culex* mosquitoes. The ability to make accurate early season predictions of WNV risk could allow local abatement districts and public health entities to implement early season interventions such as targeted adulticiding and public messaging before human transmission occurs, thereby better mitigating the risk of WNV to humans in the Coachella Valley. This approach could be applied to other localities with similar ecologies and patterns of WNV transmission.

Acknowledgments

We would like to thank NASA (ECOSTRESS18-0046 and the Health and Air Quality Program), Universities Space Research Association, the Pacific Southwest Regional Center of Excellence in Vector-Borne Diseases and the Centers for Disease Control and Prevention (1U01CK000649-01), the Coachella Valley Mosquito and Vector Control District, and the Department of Environmental Medicine and Public Health at the Icahn School of Medicine at Mount Sinai (NIEHS P30ES02351, K25 HD109509-01) for their support of this research.

Open Research

Mosquito and West Nile virus data used in this research can be obtained by contacting the Coachella Valley Mosquito and Vector Control District (https://www.cvmosquito.org), the NLDAS data is freely available from NASA (https://ldas.gsfc.nasa.gov/nldas), our code, figures, data, and methodology can be accessed via GitHub (https://github.com/nbd15/CV-MMI-WNV-Forecasting-), and the statistical software R and cited packages can be obtained from the R Project for Statistical Computing (https://www.r-project.org).

References

Barker, C. M. (2019). Models and Surveillance Systems to Detect and Predict West Nile Virus Outbreaks. *J Med Entomol*, *56*(6), 1508-1515. https://doi.org/10.1093/jme/tjz150

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference. In *a practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag New York, Inc.

Chew, R., & Gunstream, S. (1970). Geographical and seasonal distribution of mosquito species in southeastern California. *Mosquito News*, *30*(4).

Ciota, A. T., & Kramer, L. D. (2013). Vector-virus interactions and transmission dynamics of West Nile virus. *Viruses*, *5*(12), 3021-3047. https://doi.org/10.3390/v5123021

Colpitts, T. M., Conway, M. J., Montgomery, R. R., & Fikrig, E. (2012). West Nile Virus: biology, transmission, and human infection. *Clin Microbiol Rev*, *25*(4), 635-648. https://doi.org/10.1128/CMR.00045-12

Davis, J. K., Vincent, G., Hildreth, M. B., Kightlinger, L., Carlson, C., & Wimberly, M. C. (2017). Integrating Environmental Monitoring and Mosquito Surveillance to Predict Vector-borne Disease: Prospective Forecasts of a West Nile Virus Outbreak. *PLoS Curr*, *9*. https://doi.org/10.1371/currents.outbreaks.90e80717c4e67e1a830f17feeaaf85de

DeFelice, N. B., Little, E., Campbell, S. R., & Shaman, J. (2017). Ensemble forecast of human West Nile virus cases and mosquito infection rates. *Nat Commun*, *8*, 14592. https://doi.org/10.1038/ncomms14592

Diniz, D. F. A., de Albuquerque, C. M. R., Oliva, L. O., de Melo-Santos, M. A. V., & Ayres, C. F. J. (2017). Diapause and quiescence: dormancy mechanisms that contribute to the geographical expansion of mosquitoes and their evolutionary success. *Parasites & Vectors*, *10*(1), 310. https://doi.org/10.1186/s13071-017-2235-0

Eppler, M. J., & Mengis, J. (2002). The Concept of Information Overload. *MCM Research Paper*.

Epstein, P. R. (2001). West Nile virus and the climate. *J Urban Health*, *78*(2), 367-371. https://doi.org/10.1093/jurban/78.2.367

Fisher, J. B., Lee, B., Purdy, A. J., Halverson, G. H., Dohlen, M. B., Cawse-Nicholson, K., . . . Hook, S. (2020). ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. *Water Resources Research*, *56*(4), e2019WR026058. https://doi.org/https://doi.org/10.1029/2019WR026058

Horton, S., & Cole, S. (2011). Medical returns: Seeking health care in Mexico. *Social Science & Medicine*, *72*(11), 1846-1852. https://doi.org/https://doi.org/10.1016/j.socscimed.2011.03.035

Keyel, A. C., Gorris, M. E., Rochlin, I., Uelmen, J. A., Chaves, L. F., Hamer, G. L., . . . Smith, R. L. (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. https://doi.org/10.1371/journal.pntd.0009653

Kilpatrick, A. M., Kramer, L. D., Jones, M. J., Marra, P. P., & Daszak, P. (2006). West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. *PLoS Biol*, *4*(4), e82. https://doi.org/10.1371/journal.pbio.0040082

Koenraadt, C. J., & Harrington, L. C. (2008). Flushing effect of rain on container-inhabiting mosquitoes Aedes aegypti and Culex pipiens (Diptera: Culicidae). *J Med Entomol*, *45*(1), 28-35. https://doi.org/10.1603/0022-2585(2008)45[28:feoroc]2.0.co;2

Little, E., Campbell, S. R., & Shaman, J. (2016). Development and validation of a climate-based ensemble prediction model for West Nile Virus infection rates in Culex mosquitoes, Suffolk County, New York. *Parasit Vectors*, *9*(1), 443. https://doi.org/10.1186/s13071-016-1720-1

Mordecai, E. A., Caldwell, J. M., Grossman, M. K., Lippi, C. A., Johnson, L. R., Neira, M., . . . Villena, O. (2019). Thermal biology of mosquito-borne disease. *Ecol Lett*, *22*(10), 1690-1708. https://doi.org/10.1111/ele.13335

Nasci, R. S., & Mutebi, J. P. (2019). Reducing West Nile Virus Risk Through Vector Management. *J Med Entomol*, *56*(6), 1516-1521. https://doi.org/10.1093/jme/tjz083

Paull, S. H., Horton, D. E., Ashfaq, M., Rastogi, D., Kramer, L. D., Diffenbaugh, N. S., & Kilpatrick, A. M. (2017). Drought and immunity determine the intensity of West Nile virus epidemics and climate change impacts. *Proc Biol Sci*, *284*(1848). https://doi.org/10.1098/rspb.2016.2078

RCoreTeam. (2023). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

Reisen, W. K., Cayan, D., Tyree, M., Barker, C. M., Eldridge, B., & Dettinger, M. (2008). Impact of climate variation on mosquito abundance in California. *J Vector Ecol*, *33*(1), 89-98. https://doi.org/10.3376/1081-1710(2008)33[89:iocvom]2.0.co;2

Reisen, W. K., Fang, Y., & Martinez, V. M. (2006). Effects of temperature on the transmission of west nile virus by Culex tarsalis (Diptera: Culicidae). *J Med Entomol*, *43*(2), 309-317. https://doi.org/10.1603/0022-2585(2006)043[0309:EOTOTT]2.0.CO;2

Reisen, W. K., Hardy, J. L., Presser, S. B., Milby, M. M., Meyer, R. P., Durso, S. L., . . . Gordon, E. (1992). Mosquito and arbovirus ecology in southeastern California, 1986-1990. *J Med Entomol*, *29*(3), 512-524. https://doi.org/10.1093/jmedent/29.3.512

Ronca, S. E., Murray, K. O., & Nolan, M. S. (2019). Cumulative Incidence of West Nile Virus Infection, Continental United States, 1999-2016. *Emerg Infect Dis*, *25*(2), 325-327. https://doi.org/10.3201/eid2502.180765

Seid, M., Castañeda, D., Mize, R., Zivkovic, M., & Varni, J. W. (2003). Crossing the Border for Health Care: Access and Primary Care Characteristics for Young Children of Latino Farm Workers Along the US-Mexico Border. *Ambulatory Pediatrics*, *3*(3), 121-130. https://doi.org/https://doi.org/10.1367/1539-4409(2003)003<0121:CTBFHC>2.0.CO;2

Shaman, J., Day, J. F., & Stieglitz, M. (2005). Drought-induced amplification and epidemic transmission of West Nile virus in southern Florida. *J Med Entomol*, *42*(2), 134-141. https://doi.org/10.1093/jmedent/42.2.134

Skaff, N. K., Cheng, Q., Clemesha, R. E. S., Collender, P. A., Gershunov, A., Head, J. R., . . . Remais, J. V. (2020). Thermal thresholds heighten sensitivity of West Nile virus transmission to changing temperatures in coastal California. *Proc Biol Sci*, *287*(1932), 20201065. https://doi.org/10.1098/rspb.2020.1065

So, J., Kim, S., & Cohen, H. (2017). Message fatigue: Conceptual definition, operationalization, and correlates. *Communication Monographs*, *84*(1), 5-29. https://doi.org/10.1080/03637751.2016.1250429

Villarejo, D. (2003). The Health of U.S. Hired Farm Workers. *Annual Review of Public Health*, *24*(1), 175-193. https://doi.org/10.1146/annurev.publhealth.24.100901.140901

Ward, M. J., Sorek-Hamer, M., Vemuri, K. K., & DeFelice, N. B. (2023). Statistical Tools for West Nile Virus Disease Analysis. *Methods Mol Biol*, *2585*, 171-191. https://doi.org/10.1007/978-1-0716-2760-0_16

Wegbreit, J., & Reisen, W. K. (2000). Relationships among weather, mosquito abundance, and encephalitis virus activity in California: Kern County 1990-98. *J Am Mosq Control Assoc*, *16*(1), 22-27.

Wimberly, M. C., Davis, J. K., Hildreth, M. B., & Clayton, J. L. (2022). Integrated Forecasts Based on Public Health Surveillance and Meteorological Data Predict West Nile Virus in a High-Risk Region of North America. *Environ Health Perspect*, *130*(8), 87006. https://doi.org/10.1289/EHP10287

Supplementary Information

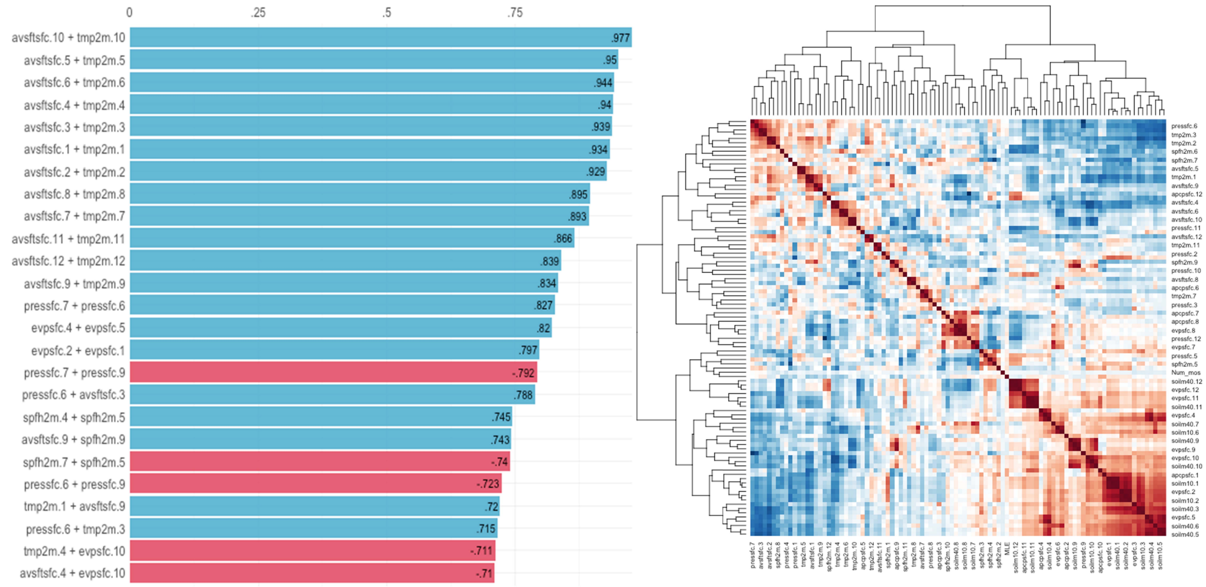

Figure S1. Correlation analysis of potential environmental predictors at the monthly NLDAS scale. Left: Ranked cross-correlations of 25 most relevant variables tested ($p < 0.05$). Right: Correlation dendogram of 25 most relevant variables tested ($p < 0.05$).

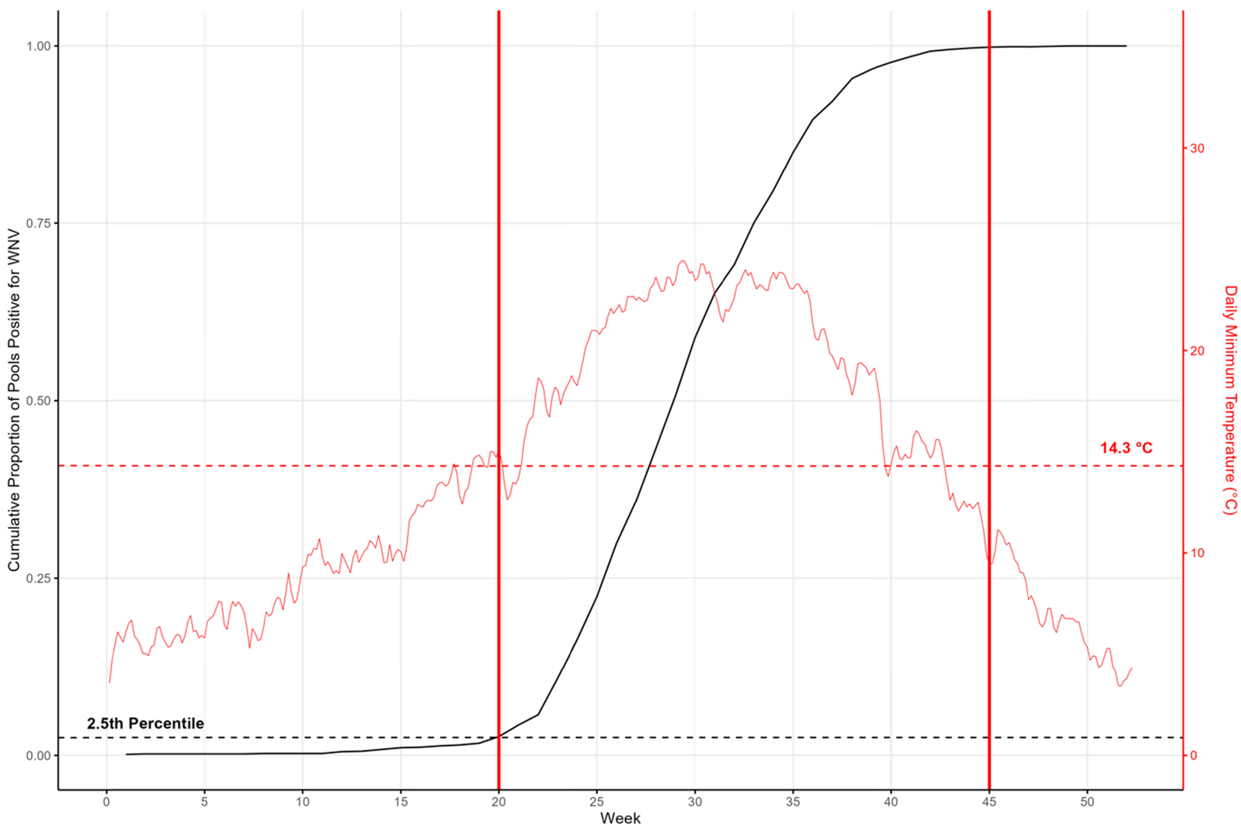

Figure S2. Seasonality of temperature (light red line) and the historic cumulative proportion of positive pools (black line) in the Coachella Valley, CA. 2.5% threshold (black dotted line) and CDC weeks 20 and 45 (vertical red lines), where week 20 corresponds to when the daily minimum temperature exceeds 14.3 ºC the threshold for viral amplification within a mosquito.

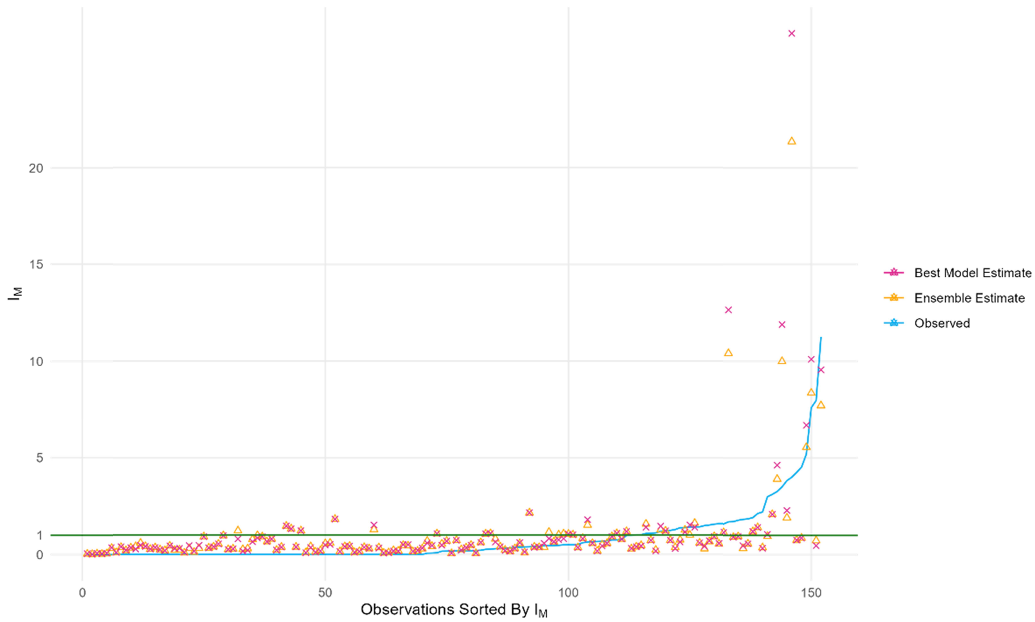

Figure S3. Comparison of observed I_M (blue) vs. best model (pink) and ensemble estimates (yellow). The green line represents the threshold, defined as the 75th percentile of historical infection rates (one I_M).

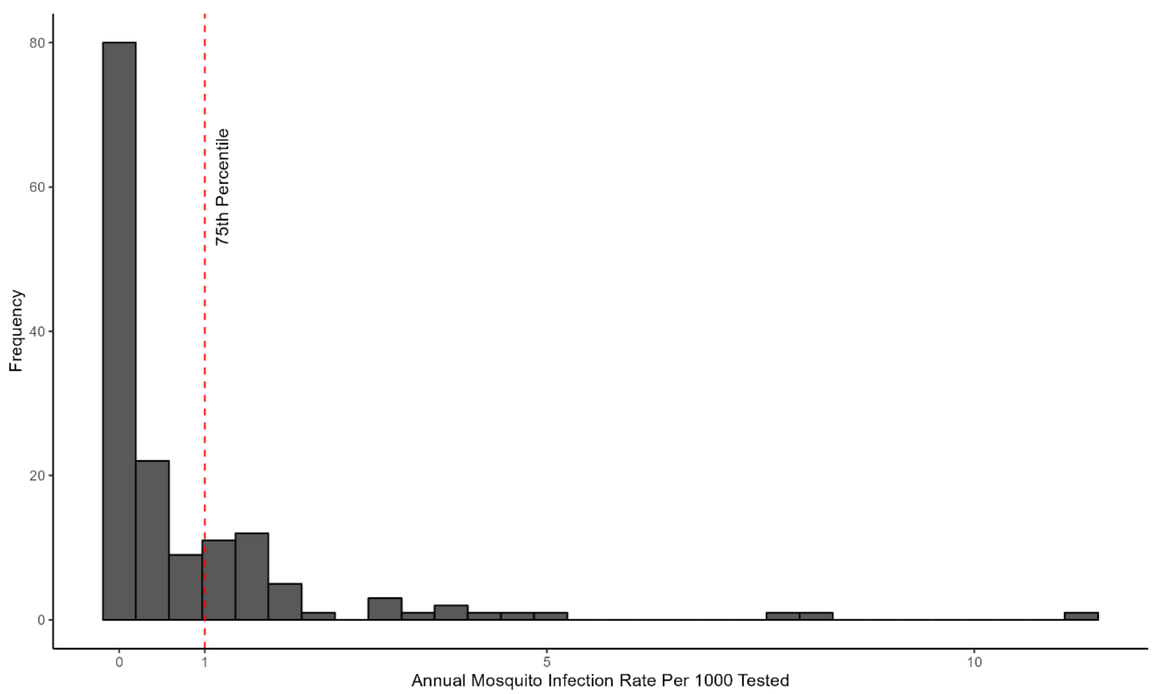

Figure S4. Distribution of WNV mosquito infection rates from 2006 - 2021 calculated at the NLDAS grid cell level. The 75th percentile comparison threshold of one I_M is shown.

Table S1. Summary statistics of yearly Culex mosquito trapping effort from CDC week 20 to 45 in Coachella Valley.

Year	Number of Pools	Positive Pools	Number of Mosquitoes Trapped	Estimated Infection Rate (I_M)
2006	2,553	24	97,269	0.25
2007	1,757	26	62,585	0.42
2008	2,084	54	79,902	0.69
2009	1,911	14	73,403	0.19
2010	3,320	69	131,572	0.53
2011	2,993	43	106,457	0.41
2012	3,412	115	127,970	0.91
2013	2,032	43	69,709	0.62
2014	2,130	66	70,823	0.95
2015	3,805	102	111,958	0.93
2016	4,889	19	157,348	0.12
2017	4,886	120	161,367	0.75
2018	7,667	45	245,134	0.18
2019	5,953	555	209,689	2.79
2020	4,938	55	166,119	0.33
2021	4,095	99	133,428	0.75
2022	10,855	148	360,937	0.41

Common Language Abstract

West Nile virus is the most significant arbovirus in the United States and is transmitted seasonally by mosquitoes. Humans are most at risk when they are in close proximity to infected mosquitoes. Predicting the risk to humans is not straightforward. In this study, we use deviations in climate associated with mosquito biology and viral development to forecast seasonal West Nile risk in the Coachella Valley of California. We developed a statistical model of West Nile virus transmission in the Coachella Valley using 17 years of mosquito surveillance data and environmental data. Our model indicated that the combination of a cooler and dryer winter followed by a wetter and warmer spring and a cooler than normal summer was most predictive of West Nile positive mosquitoes in the Coachella Valley. The ability to make accurate early season predictions of West Nile risk could assist local public health entities implement early season interventions to better mitigate the risk of West Nile virus to humans in the Coachella Valley.