Assessing Climate Driven West Nile Virus Risk in Washington State

Gabriella E. LaBazzo

A thesis

submitted in partial fulfillment

requirements for the degree of

Master of Public Health

University of Washington

2021

Committee:

Christine Khosropour

Cory Morin

Ruanne Barnabas

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Assessing Climate Driven West Nile Virus Risk in Washington State

Gabriella E. LaBazzo

Chair of the Supervisory Committee:

Christine Khosropour

Department of Epidemiology

Background: West Nile virus (WNV) is becoming a significant public health concern in Washington State (WA). Climate factors likely impact WNV case numbers in WA, but this has not been widely explored.

Methods: We examined the relationship between weekly averaged metrological conditions and reported WNV cases from 2006 to 2019 for two counties in Eastern WA (Yakima and Benton). We used univariate analyses to identify the time-lag (1-6 weeks) for each climate variable (maximum temperature, minimum temperature, vapor pressure deficit [VPD], specific humidity, and total precipitation) that was associated with the presence of a WNV case with the highest statistical significance. The five climate variables were subsequently included in two different predictive models: a multi-variate logistic regression model and a decision tree model, a type of machine learning method. The results of the models were then qualitatively and quantitively compared based on their accuracy, sensitivity, and specificity.

Results: The five variables most significantly associated with WNV in the univariate analysis were Maximum Temperature Lag Week 2, Minimum Temperature Lag Week 2, Specific Humidity Lag Week 2, Total Precipitation Lag Week 4, and Vapor Pressure Deficit Lag Week 3. In both the logistic regression and decision tree model, minimum temperature was the most important climate variable for determining the probability of a WNV case in Yakima & Benton county, and Vapor Pressure Deficit was identified as the second most important predicative variable. The two models had a similar accuracy, but the decision tree model was more sensitive than the logistic regression model.

Conclusion: The decision tree model and logistic regression model yielded similar results, but the decision tree model may be a better approach to predict WNV as it is more sensitive and may be more intuitive for public health practitioners. The findings of this study can be used to develop early warning systems or seasonal forecast systems that could improve predictions of WNV cases in endemic and non-endemic areas.

Introduction

West Nile Virus (WNV), is the leading cause of mosquito-borne disease in the continental United States (US). It was introduced to the Western Hemisphere in 1999 and has spread across most of North America (Wimberly et al., 2014). Most WNV human cases are undiagnosed, as only 20% of cases experience a febrile illness often accompanied by headaches, body aches, joint pains, gastrointestinal upset, and rash. However, approximately 1% of infected individuals develop severe, and potentially fatal, neuroinvasive disease which can include encephalitis, meningitis, and acute flaccid paralysis. (Marcantonio et al., 2015; Wiesman, 2019). As of 2019, there were 51,801 reported human cases of WNV in the United States (US), of which 25,290 cases were classified as severe neuroinvasive disease. WNV has caused 2,390 deaths in the US (CDC, 2019).

Climate conditions can strongly influence the ecology of WNV. In order for WNV to be transmitted to humans, environmental conditions must be suitable for the mosquitoes and virus to complete their transmission cycle. The mosquito must feed on an infected avian host, and then retransmit the virus to a susceptible human host. Mammalian species are considered dead-end hosts, and cannot retransmit the virus to other mosquitoes. For the virus, ambient temperature can affect viral replication and alter the length of the extrinsic incubation period, which is the time it takes for the virus to replicate to a high enough viral load to make the mosquito infectious (Paz et al., 2013). Not only are the mosquito and virus susceptible to climate conditions, host species may also change behaviors when certain environmental factors shift (e.g. humans spend less time outside in unseasonably warmer/colder temperatures; climate may influence bird migration) (Paz, 2015).

High temperature may increase the risk of WNV transmission (Davis et al., 2017). Elevated temperatures accelerate mosquito development and increase viral infection, dissemination, and transmission by the mosquito. Moreover, elevated temperatures increase mosquito reproduction rates, the number of blood meals, and prolong the mosquito breeding season (Paz et al., 2013). Precipitation and relative humidity may also be important predictors of shifts in the geographic pattern of WNV transmission (Hess et al., 2018) and WNV incidence (Chuang & Wimberly, 2012; Hahn et al., 2015; Marcantonio et al., 2015; Smith et al., 2020; Wimberly et al., 2014).

The first locally-acquired human case of WNV in Washington State (WA) was reported in 2006. As of 2019, there were 117 cases of WNV and 71 cases of neuroinvasive disease caused by WNV. Although these numbers are relatively low compared to high burden states such as the Dakotas, Nebraska, and eastern Montana and Wyoming, there has been a recent shift in the epidemiology of WNV in WA State, with detection of WNV in areas of the state where it had not been previously detected (Wimberly et al., 2013; Weisman, 2019). This spread to previously unaffected parts of WA may be due to climate change altering patterns of WNV suitability in WA. Gaining a better understanding of how climate influence disease risk in WA state can inform where to implement WNV prevention and surveillance strategies (Hess et al., 2018).

The objective of this study was to better understand how climate variables such as temperature, specific humidity, total precipitation, and vapor pressure affect the incidence and seasonality of WNV disease in WA state, in order to aid public health professionals in determining, managing and preventing cases of WNV across WA state.

Methods

Data Sources and Population

West Nile Virus Data: West Nile Virus is a notifiable condition in WA State. All healthcare providers, laboratories, and veterinarians are required to report cases of WNV to the local health department, which subsequently reports this information to the Washington State Department of Health (WA DOH). For this analysis, WA DOH provided weekly human case counts of WNV at the county level from 2006-2019. We only included serologically confirmed human case records of West Nile fever, West Nile neuroinvasive disease, and WNV infected blood donors. We did not include data collected from testing mosquito pools or avian species. We aggregated data into four county groups: Eastern WA, Western WA, Yakima County, and Benton County. Yakima and Benton county accounted for 69% of WNV cases in WA State during the study period and have a similar climate, therefore the remainder of the analysis was completed with only data from these two counties.

Weather and Climate Data: The meteorological data used to identify climate variables associated with WNV incidence in WA State was obtained from gridMET, a dataset of daily high-spatial resolution (~4km, 1/24th degree) surface meteorological data covering the contiguous United States (Abatzoglou, 2013). It combines the PRISM gridded dataset (Daly et al., 2018) and the NLDAS-2 reanalysis dataset (Mitchell et al., 2004). Variables available from gridMET included weekly maximum and minimum temperature (°C), total precipitation (mm), vapor pressure deficit (kPa), and specific humidity (kg/kg). Vapor Pressure Deficit (VPD), is the difference between the water vapor pressure at saturation and the actual water vapor pressure for a given temperature. As VPD increases, the rate at which an organism can lose water to the atmosphere increases, especially mosquitoes, that are highly susceptible to water loss (Holmes & Benoit, 2019). Due to terrain differences across counties in WA state, we decided to use the UN WPP (World Populations Prospects)- Adjusted Population Data from the NASA Socioeconomic Data and Application Center to re-grid the 2.5 minute (~5 km) gridded population data to the gridMET grid using nearest neighbor interpolation. Once we re-gridded the population data to the gridMET grid, we computed the weather averages for only the grid cells with a population of 100 people or more..

The impact of climate on pathogen ecology is not always immediate; therefore, climate lag variables were constructed for the analysis. The lag variables were based on extrinsic incubation period of WNV in the two most common species of mosquito vectors in WA state, *Culex pipiens* and *Culex tarsalis* (Anderson et al., 2008) as well as the intrinsic incubation period in humans*.* To assess time lag effects in weekly variables, we created lag-week variables of 1-week, 2-week, 3-week, 4-week, 5-week, and 6-weeks for the five aforementioned climate variables (Davis et al., 2017).

Statistical Analysis

Our final dataset was limited to the weeks of the year when WNV cases were reported during the period from 2006 to 2019 in Yakima and Benton counties. This was done to enhance the statistical power and predictability of the model. Therefore, the start date of the analysis is Week 33 in 2008 and our final analysis only included weeks 27 though 46 for each subsequent year through 2019.

In our descriptive analyses, we present the number of WNV cases by county and region in WA (Figure 1), the weekly climate patterns for Yakima and Benton Counties (Figure 2), and the distribution of climate variables included in the analysis (Appendix Figure 1).

Univariate Analysis: We used a Welch two-sample t-test with unequal variance (α=0.05) to test the univariate association between our climate variables and presence/absence of a WNV case, to determine which variables were most strongly associated with WNV case presence. The climate variables included in the univariate analyses are county averaged maximum and minimum temperature, total precipitation, specific humidity, and vapor pressure with six time-lags for each variable. The time-lag with the smallest p-value for each climate variable was included in the predictive models described below. If more than one variable had an identically low p-value, we selected the variable that was most plausibly associated with WNV case presence. Only one variable per climate category was included in the final model (Table 1).

Table 1. Univariate analysis of the association between climate variables (mean and standard deviation [SD]) and weeks of a WNV case, Yakima and Benton Counties, WA State, 2008-2019 (annual weeks 27-46)

	Weeks when WNV Case Diagnosed (N = 56)	Weeks when no WNV Cases diagnosed (N = 405)	p-value**
	Mean ± SD	Mean ± SD	
Max temperature °C	30.19 ± 4.70	24.58 ± 7.95	5.315e-14
1-week lag	31.70 ± 3.98	25.40 ± 7.13	< 2.2e-16
***2-week lag**	**31.69 ± 3.61**	**26.17 ± 6.47**	**< 2.2e-16**
3-week lag	31.72 ± 3.06	26.84 ± 5.85	< 2.2e-16
4-week lag	32.16 ± 3.41	27.36 ± 5.32	< 2.2e-16
5-week lag	31.98 ± 3.40	27.78 ± 4.89	2.538e-15
6-week lag	31.26 ± 3.43	28.12 ± 4.53	3.99e-10
Min temperature °C	13.21 ± 3.69	8.9 ± 5.43	4.742e-14
1-week lag	14.35 ± 3.14	9.49 ± 4.90	<2.2e-16
***2-week lag**	**14.45 ± 2.73**	**9.94 ± 2.73**	**<2.2e-16**
3-week lag	14.13 ± 2.50	10.34 ± 4.27	< 2.2e-16
4-week lag	14.68 ± 2.92	10.69 ± 3.88	< 2.2e-16
5-week lag	14.40 ± 2.91	11.03 ± 3.55	2.48e-14
6-week lag	13.92 ± 2.56	11.26 ± 3.33	3.855e-12
Specific Humidity (kg/kg)	0.0071 ± 0.0014	0.0061 ± 0.0015	4.499e-07
1-week lag	0.0071 ± 0.0013	0.0063 ± 0.0014	5.766e-06
***2-week lag**	**0.0073 ± 0.0012**	**0.0063 ± 0.0014**	**1.449e-08**
3-week lag	0.0069 ± 0.0012	0.0065 ± 0.0014	0.00153
4-week lag	0.0070 ± 0.0012	0.0065 ± 0.0013	0.00115
5-week lag	0.0071 ± 0.0014	0.0065 ± 0.0012	0.004281
6-week lag	0.0066 ± 0.0012	0.0066 ± 0.0012	0.8788
Total Precipitation (mm)	1.31 ± 2.01	2.56 ± 4.68	0.0001725
1-week lag	1.19 ± 2.12	2.56 ± 4.74	7.777e-05
2-week lag	1.28 ± 2.99	2.46 ± 4.57	0.005395
3-week lag	0.75 ± 1.69	2.21 ± 4.31	9.14e-07
***4-week lag**	**0.72 ± 1.37**	**2.24 ± 4.53**	**5.543e-08**
5-week lag	1.40 ± 2.91	1.94 ± 3.66	0.1644
6-week lag	0.87 ± 1.47	2.18 ± 4.35	2.47e-06
Vapor Pressure Deficit (kPa)	2.02 ± 0.61	1.49 ± 0.71	1.384e-09
1-week lag	2.03 ± 0.64	1.54 ± 0.67	4.157e-08
2-week lag	1.95 ± 0.62	1.58 ± 0.66	1.136e-05
***3-week lag**	**2.05 ± 0.59**	**1.59 ± 0.64**	**1.646e-08**
4-week lag	2.00 ± 0.67	1.60 ± 0.62	8.347e-06
5-week lag	1.86 ± 0.63	1.61 ± 0.63	0.002458
6-week lag	1.82 ± 0.69	1.60 ± 0.69	0.01035

SD, standard deviation; WNV, West Nile Virus *Bold variables were selected for final models based on smallest p-value in each climate category **P-values derived from Welch Two Sample t-test

We developed a decision tree model and logistic regression model, described below. For these analyses, we split the data randomly 75:25 into a training dataset and a test dataset. The same training and test dataset were used for both the decision tree and logistic regression model.

Decision Tree Model: Decision tree analyses are a type of machine learning that use a tree-like model of decisions to calculate the probability of an outcome (Navada et al., 2011). The decision tree model for this study calculated the probability of a WNV case given the meteorological conditions based on the variables selected in the univariate analysis. The model was run in the R statistical environment, Version 1.3.1093 using the *tree* package.

We used the training dataset to build the decision tree model, using the five variables identified in the univariate analysis as being most strongly associated with the presence of a WNV case. The only growth condition initially specified for the model was a restriction on the minimum number of samples in leaf nodes. The tree did not continue to branch if five or fewer observations remained in a node. To avoid overfitting, the tree was then pruned using K-fold cross validation (Barnard et al., 2019). Using this method, we determined the number of terminal nodes that would result in the lowest test error. This parameter was then applied to the model and used to build the final decision tree featured in Figure 3.

Multivariate Logistic Regression: Using the training dataset, we developed a multivariate logistic regression model using the same five variables from the univariate analysis that we used in the decision tree model. We then used the test dataset to examine the accuracy of the model to identify the presence of a WNV case (outcome) given our set of five climate variables identified in the univariate analysis (predictors). The model was run in the R statistical environment using the *glm* function.

We used a qualitative approach to compare the results of the decision tree model and multivariate logistic regression model based on their calculated probability of WNV detection. We determined the accuracy, sensitivity, and specificity of each model at different probability thresholds using the test dataset (Table 3). The thresholds were established using the probabilities displayed by the terminal nodes in the decision tree (0.35 and 0.81). The default threshold for normalized predicted probabilities is 0.5. Therefore, we evaluated the two models using the 0.5 threshold in addition to 0.35 and 0.81. At each threshold, we determined the number of true positive cases (i.e., weeks when a WNV case was detected) and true negative cases (i.e., weeks where there were no WNV cases detected) from the test dataset. We calculated sensitivity as the number of model-predicted positive cases divided by the number of true positive cases, and calculated sensitivity as the number of model-predicted negative cases divided by the number of true negative cases. To calculate the accuracy for each model, at each threshold, we determined the number of true positive cases (i.e., true model-predicted cases and actual cases) and the number of true negative cases (i.e., true model-predicted non-cases and actual non-cases) and then divided them by the total number of observations in the test dataset.

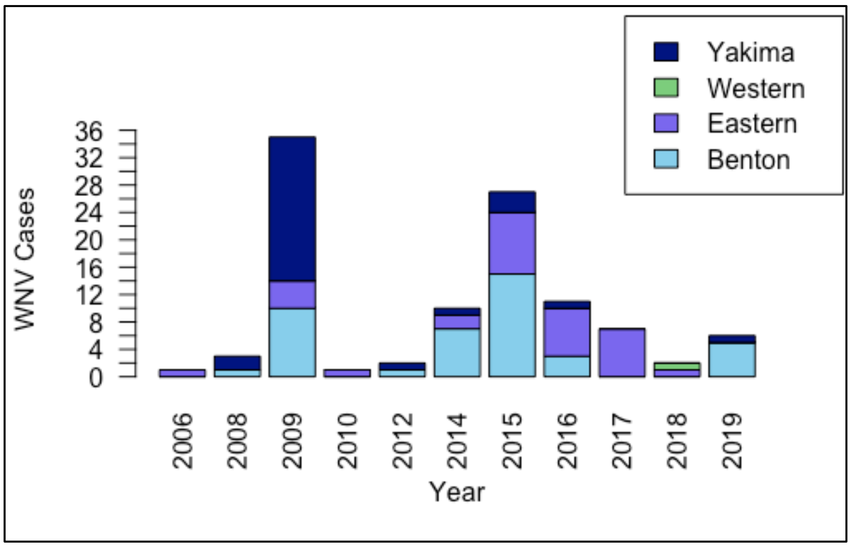
Results

There were 151 human cases of WNV reported to WA DOH between 2006 and 2019. Of these, we excluded 46 cases from this analysis because the exposure location was missing, unknown, located outside the WA State, listed with multiple states including or excluding WA, or multiple counties were listed. Of the remaining 105 cases, 72 (69%) were selected for the final analytic analyses because they originated in either Yakima or Benton counties. Climate data from only these two counties were used in the final analysis.

Descriptive Analysis of WNV Cases and Climate

Figure 1 shows the distribution of the 105 WNV cases in all counties across Washington between 2006-2019. The majority of cases occurred in Yakima and Benton county, with only one human case reported in Western Washington during this time period (in 2018).

Figure 1. Distribution of year of serologically confirmed human WNV cases in Washington State, by county and region, 2006-2019.



Eastern WA Counties: Adams, Asotin, Chelan, Columbia, Douglas, Ferry, Franklin, Garfield, Grant, Kittitas, Klickitat, Lincoln, Okanogan, Pend Oreille, Spokane, Stevens, Walla Walla, and Whitman
Western WA Counties: Clallam, Clark, Cowlitz, Grays Harbor, Island, Jefferson, King, Kitsap, Lewis, Mason, Pacific, Pierce, San Juan, Skagit, Skamania, Snohomish, Thurston, Wahkiakum, Whatcom

In Yakima and Benton counties, during the study period, the maximum and minimum temperatures gradually increased in the summer months and peaked around the end of July, in week 31 (Figure 2; Panels 1 and 2). Precipitation levels followed an inverse trend and peaked in the winter months (weeks 50-03). The majority of the WNV cases in Yakima and Benton occurred between weeks 27-39 (beginning of July through the end of September), following a similar trend to the maximum and minimum temperature (Figure 2; Panel 3). No WNV cases have ever been reported between weeks 47-26 in Yakima or Benton County.

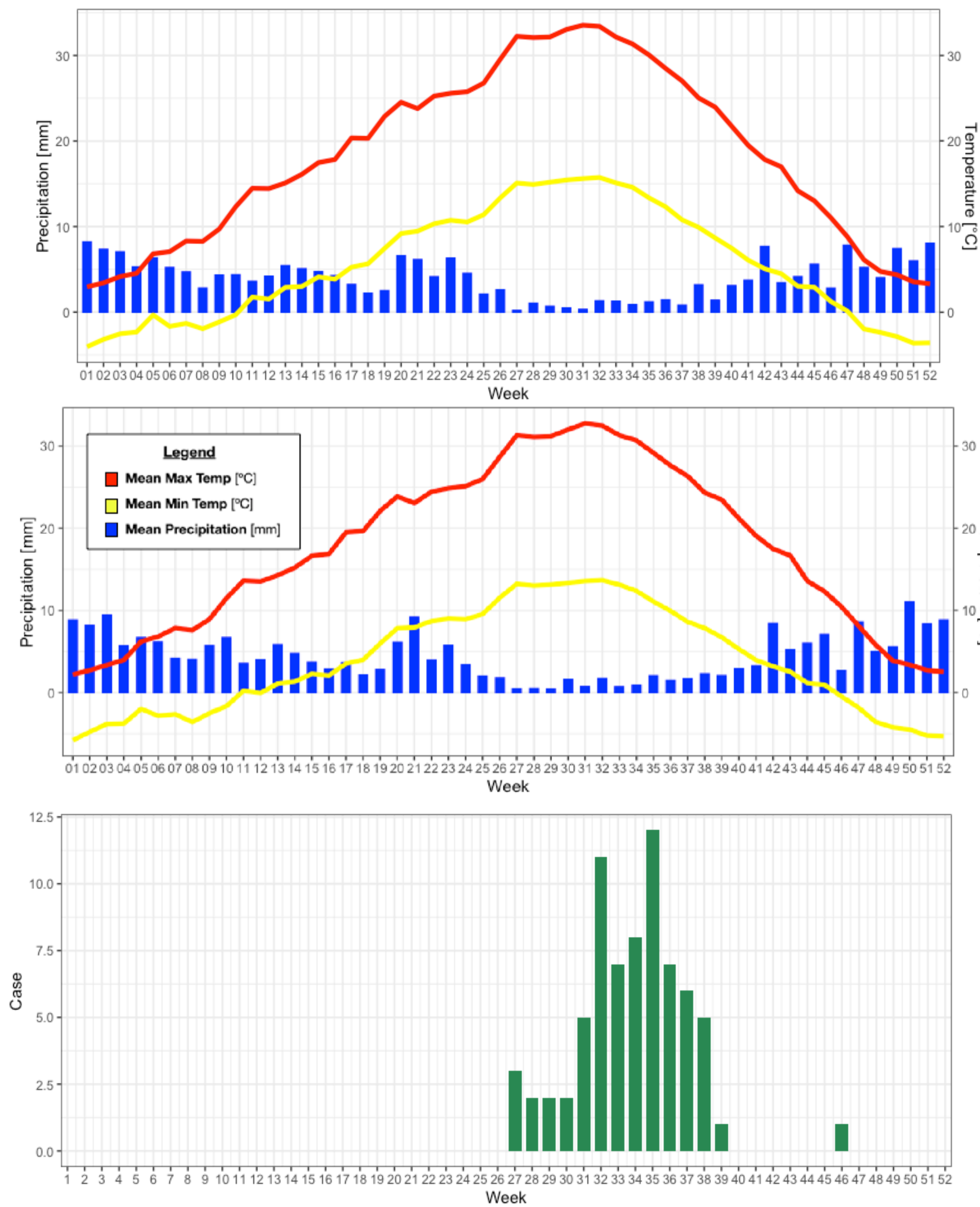


Figure 2. Top two figures are Climographs representing the weekly mean maximum temperature °C (Red), minimum temperature °C (Yellow) and precipitation [mm] (Blue) for Yakima County and Benton Counties, WA State, 2008-2019. Bottom figure in green represents the weekly mean cases of WNV for Yakima & Benton counties, WA State, 2008-2019.

Association of Climate Variables and WNV Cases

The mean weekly values of the seven lag variables for each climate category are displayed in Table 1. For maximum and minimum temperature, the smallest p-values were the same for the variables describing lag weeks 1-4. Based on the biological plausibility of the course of WNV infection, we selected the lag week 2 variable for minimum and maximum temperature for inclusion in our decision tree analysis and multivariate model. The remaining climate variables were selected for inclusion in our subsequent analyses based on the lowest p-value. These included: specific humidity lag week 2, total precipitation lag week 4, and vapor pressure deficit lag week 3.

Decision Tree Analysis

The decision tree model of best fit is displayed in Figure 3. The root node variable of the decision tree was minimum temperature (lag week 2) and the branching criteria was less than 13.22 °C. The final decision node branches with the criterion of a vapor pressure deficit of less than 2.07 kPa. The terminal node indicating a vapor pressure deficit between 1.9 kPa and 2.07 kPa and a minimum temperature above or equal to 13.21 °C, predicted a probability of 81%. The alternate terminal node predicted a 35% probability of a WNV case occurring in Yakima or Benton county if the following climate conditions were met; a vapor pressure deficit above or equal to 2.07 kPa and a minimum temperature above or equal to 13.22 °C.

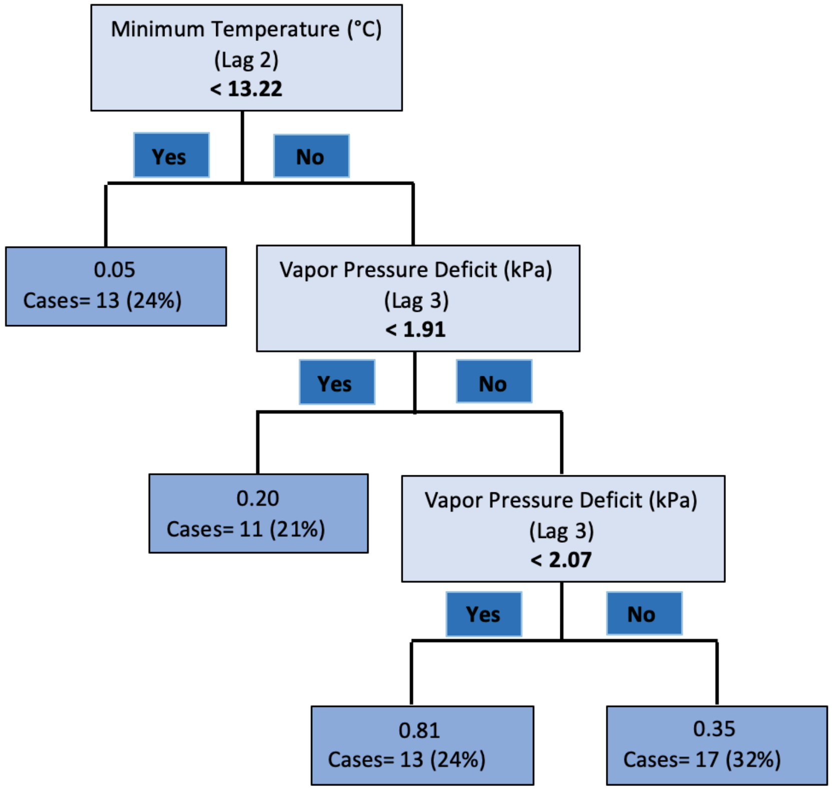


Figure 3. Decision Tree model of the probability of a WNV case occurring in Benton and Yakima Counties, given a set of meteorological conditions, N= 369 (315 non-cases, 54 cases).

*The light blue boxes with bold text provide the specific meteorological conditions. The dark blue boxes represent the probability of a WNV case occurring if this condition was met (ie. If the minimum temperature is below 13.22 °C, then the probability of a case occurring is 5%). The number and percent of cases provided in each dark blue box represents the number of cases (and percent out of the total number) that occurred in the weeks that met the corresponding set of meteorological conditions.

Multivariate Logistic Regression

In the multivariate logistic regression model, the two variables that were most strongly associated with presence of a WNV case were minimum temperature and vapor pressure deficit (Table 2). As the minimum temperature and vapor pressure deficit increase, the probability of a WNV case occurring in Yakima and Benton Counties increases. The remaining three climate variables were not significantly associated with the presence of a WNV case, when adjusting for the other climate variables.

Table 2. Multivariate logistic regression model results of the association between specific climate variables and WNV cases in Yakima & Benton counties, WA State, 2008-2019*		
	β Coefficient (95% CI)	**p-value**
Intercept	-5.84 (-47.81, -10.60)	0.002
Max temperature °C (Lag week 2)	-0.04 (-1.11, 0.69)	0.65
Min temperature °C (Lag week 2)	0.31 (0.15, 2.93)	0.03
Specific Humidity (kg/kg) (Lag week 2)	179.79 (-658.49, 2456.42)	0.37
Total Precipitation (mm) (Lag week 4)	-0.17 (-1.79, 0.08)	0.26
Vapor Pressure Deficit (kPa) (Lag week 3)	0.26 (-1.53, 4.14)	0.07
CI, confidence interval; WNV, West Nile Virus *Climate variables were selected as part of a univariate analysis (see Table 1)		

Comparison of the Decision Tree Model and Multivariate Logistic Regression Model

A comparison of the performance of the decision tree model and logistic regression model are in Table 3. Overall, both models had relatively similar accuracies for each probability threshold. The sensitivity for both models decreased as the probability threshold increased. Overall the decision tree model had a higher sensitivity than the logistic regression model, but the logistic regression model had a higher specificity.

Table 3. Comparison of sensitivity, specificity, and accuracy of the logistic regression and decision tree model to predict the presence of a WNV case in a given week using the test dataset (N=18 cases and 106 non-cases)		
	Logistic Regression	**Decision Tree**
Probability Threshold = 0.35		
Sensitivity	0.33	0.66
Specificity	0.91	0.85
Accuracy	0.82	0.82
Probability Threshold = 0.50		
Sensitivity	0.16	0.50
Specificity	0.99	0.87
Accuracy	0.87	0.83
Probability Threshold = 0.81		
Sensitivity	0.00	0.22
Specificity	1.00	0.98
Accuracy	0.85	0.87

Discussion

In this analysis using climate and WNV data from two counties in Eastern WA, we used two different models that predicted the presence of a WNV case from a set of climate variables. We found that minimum temperature and vapor pressure deficit were the two most indicative climate predictors of a WNV case occurrence in Yakima and Benton Counties. When comparing the performance of the two models, we found that the decision tree model has a higher sensitivity for predicating WNV cases. Given the relative ease of generating decision tree models and their intuitiveness, decision tree models may be useful tools for public health practitioners to prepare for WNV cases in the wake of a changing climate.

Minimum temperature was identified as the most important predictive factor of a WNV case in both models. This finding is in line with other studies, which have similarly found that as the minimum temperature increases, the probability of a WNV case also increases (Davis et al., 2018; Shipp & Gillespie, 1993). This finding makes intuitive sense, because the extrinsic incubation period is strongly temperature dependent, with the rate of viral replication within a susceptible mosquito vector increasing with increasing temperature (Lockaby et al., 2016; Paz et al., 2013). In this study, we noted that all WNV cases occurred between a minimum temperature of 10-14°C, which is consistent with studies suggesting that the thermal threshold for replication of WNV may lie between 10-14°C (Reisen et al., 2006). Outside of this threshold, it is unlikely for WNV cases to be reported in humans. Under a changing climate, the number of weeks per year that facilitate WNV replication in mosquitoes, may increase, extending the season of potential transmission.

Both the logistic regression and decision tree model predicted that as vapor pressure deficit (VPD) increases, the incidence of WNV increases. This is inconsistent with other studies. In a study in North Dakota, Davis et al. found that human WNV disease risk was highest following periods of high temperature and high (negative) vapor pressure deficit. However, that study used different methodology and was conducted in a dissimilar climate compared to WA state. In WA, WNV cases are generally reported in late summer and early fall, this is also the time of year when VPD is highest. Therefore, the VPD results of this study are consistent with known climate trends in Yakima and Benton county. Mosquitoes are highly susceptible to water loss due to high rates of evaporation through respiratory spiracles (Holmes & Benoit, 2019). Evaporation is dependent upon atmospheric moisture conditions, and can be most precisely measured by VPD. As the VPD increases, the rate at which mosquitoes lose water to the atmosphere increases (Davis et al., 2018; Holmes & Benoit, 2019), which may lead to a delayed decrease in the number of mosquitoes and therefore WNV cases after the peak of VPD (early September).

Interestingly, we did not find precipitation or relative humidity to be predictors of WNV in these two counties. This is somewhat inconsistent with other studies (Hess et al., 2018). In some parts of Europe and the US, precipitation has been found to be associated with WNV incidence ((Marcantonio et al., 2015) (Paz, 2015; Soverow et al., 2009). Other studies have found that precipitation may affect WNV incidence differently depending on factors such as land use, land cover, climate, and prevalence of animal hosts and mosquito vectors (Chuang & Wimberly, 2012; Hahn et al., 2015; Smith et al., 2020; Wimberly et al., 2014). It is possible that these two climate factors were not predictive in our study because they are closely correlated with minimum temperature and VPD, and thus their effects were minimal once minimum temperature and VPD were accounted for in the analysis.

Both models identified the same two climate variables (minimum temperature and VPD) as being strongly predictive of WNV cases, but their predictive performance differed somewhat. Both models performed well and attained high accuracy, but the decision tree model had a higher sensitivity, whereas the logistic regression had a higher specificity. From a public health standpoint, it may be preferable to correctly predict the presence of a case, rather than predict the absence of a case; thus, the decision tree model may be the preferred approach. When comparing interpretability, the graphical representation of the decision tree can be followed by most audiences and is more intuitive than the logistic regression model, which may require some degree of statistical knowledge.

We believe that the results from this study will assist public health practitioners with developing WNV guidelines and adaptation and mitigation strategies for the two counties included in this analysis, as well as for other counties throughout WA and outside of WA. Very few studies, to our knowledge, have used decision tree modeling to predict vector-borne disease incidence. But it is becoming increasingly used in other settings, and may aid public health officials to distribute warning messages about WNV to the public and provide safety strategies to community members who live and recreate in high risk areas. Not only can educational information be distributed prophylactically, but these models may help mosquito control districts in WA state (MCDs) determine how to adequately control mosquito populations.

This study has a number of strengths. The two counties analyzed in this study had very similar climates which allowed us to include a large sample of WNV cases. We used time-varying lag variables to develop the most accurate models using biological principals regarding the incubation period of WNV in mosquitoes (extrinsic) and humans (extrinsic). There are several limitations to this study. First, this study did not account for bird or human behavior. As the climate changes and temperature increase, bird species have been migrating to their breeding grounds earlier in the season (Horton et al., 2020). This phenomenon might influence the appearance and timing of disease in locations near migratory routes (Paz, 2015). Likewise, human behaviors may differ in response to continually fluctuating climate patterns (Wimberly et al., 2014). Second, there is an unknown degree of spatial uncertainty because some infected humans may have been infected with WNV at a location away from their residence and thus incorrectly report their exposure location. Third, although the climates in both Yakima & Benton county are very similar, the climate values were averaged over both counties and may not have always been representative of the conditions the mosquito who transmitted the virus actually experienced. Fourth, there are few reported cases of WNV in many of the counties in WA, which limits the ability of the model to accurately predict WNV presence. Since only two counties in WA were analyzed, the results cannot be extrapolated to other parts of WA, specifically Western WA where the mean temperature is lower and precipitation is higher. Fifth, this study only reported clinically diagnosed human WNV cases between 2006-2019 and did not include positive mosquito counts from trapping data. However, human disease surveillance can still provide valuable information that complements other studies that include information about mosquito trapping surveillance, dead bird surveillance, and environmental monitoring data (Wimberly et al., 2013).

In conclusion, the results from the study show strong support for using decision tree models to predict WNV incidence in various climates and settings. We suggest that this type of model could be useful for improving environmental modeling of other mosquito-borne diseases, in addition to WNV virus. Furthermore, additional research should be conducted in other areas across the country with similar climates and higher WNV case counts. In future analyses, other modeling techniques that are known to reduce over-fitting, such as boosted regression trees, could be used and evaluated to determine their utility. As changes in weather patterns become more extreme, it will become particularly important to improve our ability to predict how climate factors impact mosquito ecology and their ability to transmit diseases. Overall, this study's findings could be used to develop an early warning system or seasonal forecast system that could allow for better predictions of future WNV outbreaks in endemic and non-endemic areas across various climates and geographic locations.

References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), 121–131. https://doi.org/10.1002/joc.3413

Anderson, J. F., Main, A. J., Delroux, K., & Fikrig, E. (2008). Extrinsic incubation periods for horizontal and vertical transmission of West Nile virus by culex pipiens pipiens (Diptera: Culicidae). *Journal of Medical Entomology*, *45*(3), 445–451. https://doi.org/10.1603/0022-2585(2008)45[445:EIPFHA]2.0.CO;2

Barnard, D. M., Germino, M. J., Pilliod, D. S., Arkle, R. S., Applestein, C., Davidson, B. E., & Fisk, M. R. (2019). Cannot see the random forest for the decision trees: selecting predictive models for restoration ecology. *Restoration Ecology*, *27*(5), 1053–1063. https://doi.org/10.1111/rec.12938

Chuang, T. W., & Wimberly, M. C. (2012). Remote Sensing of Climatic Anomalies and West Nile Virus Incidence in the Northern Great Plains of the United States. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0046882

Daly, C., Halbleib, M. D., Hannaway, D. B., & Eaton, L. M. (2018). Environmental limitation mapping of potential biomass resources across the conterminous United States. *GCB Bioenergy*, *10*(10), 717–734. https://doi.org/10.1111/gcbb.12496

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 Currents*, 1–24. https://doi.org/10.1371/currents.outbreaks.90e80717c4e67e1a830f17feeaaf85de

Davis, J. K., Vincent, G. P., Hildreth, M. B., Kightlinger, L., Carlson, C., & Wimberly, M. C. (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 Tropica*, *185*(January), 242–250. https://doi.org/10.1016/j.actatropica.2018.04.028

Hahn, M. B., Monaghan, A. J., Hayden, M. H., Eisen, R. J., Delorey, M. J., Lindsey, N. P., Nasci, R. S., & Fischer, M. (2015). Meteorological conditions associated with increased incidence of west nile virus disease in the United States, 2004-2012. *American Journal of Tropical Medicine and Hygiene*, *92*(5), 1013–1022. https://doi.org/10.4269/ajtmh.14-0737

Hess, A., Davis, J. K., & Wimberly, M. C. (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. https://doi.org/10.1029/2018gh000161

Holmes, C. J., & Benoit, J. B. (2019). Biological adaptations associated with dehydration in mosquitoes. *Insects*, *10*(11). https://doi.org/10.3390/insects10110375

Horton, K. G., La Sorte, F. A., Sheldon, D., Lin, T. Y., Winner, K., Bernstein, G., Maji, S., Hochachka, W. M., & Farnsworth, A. (2020). Phenology of nocturnal avian migration has shifted at the continental scale. *Nature Climate Change*, *10*(1), 63–68. https://doi.org/10.1038/s41558-019-0648-9

Lockaby, G., Noori, N., Morse, W., Zipperer, W., Kalin, L., Governo, R., Sawant, R., & Ricker, M. (2016). Climatic, ecological, and socioeconomic factors associated with West Nile virus incidence in Atlanta, Georgia, U.S.A. *Journal of Vector Ecology*, *41*(2), 232–243. https://doi.org/10.1111/jvec.12218

Marcantonio, M., Rizzoli, A., Metz, M., Rosa, R., Marini, G., Chadwick, E., & Neteler, M. (2015). Identifying the environmental conditions favouring West Nile Virus outbreaks in Europe. *PLoS ONE*, *10*(3), 1–18. https://doi.org/10.1371/journal.pone.0121158

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., … Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(7), 1–32. https://doi.org/10.1029/2003jd003823

Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. *Proceedings - 2011 IEEE Control and System Graduate Research Colloquium, ICSGRC 2011*, 37–42. https://doi.org/10.1109/ICSGRC.2011.5991826

Paz, S. (2015). Climate change impacts on West Nile virus transmission in a global context. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1665), 1–11. https://doi.org/10.1098/rstb.2013.0561

Paz, S., Malkinson, D., Green, M. S., Tsioni, G., Papa, A., Danis, K., Sirbu, A., Ceianu, C., Katalin, K., Ferenczi, E., Zeller, H., & Semenza, J. C. (2013). Permissive Summer Temperatures of the 2010 European West Nile Fever Upsurge. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056398

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

Shipp, J. L., & Gillespie, T. J. (1993). Influence of temperature and water vapor pressure deficit on survival of Frankliniella occidentalis (Thysanoptera, Thripidae). *Environmental Entomology*, *22*(4), 726–732. https://doi.org/10.1093/ee/22.4.726

Smith, K. H., Tyre, A. J., Hamik, J., Hayes, M. J., Zhou, Y., & Dai, L. (2020). Using Climate to Explain and Predict West Nile Virus Risk in Nebraska. *GeoHealth*, *4*(9). https://doi.org/10.1029/2020GH000244

Soverow, J. E., Wellenius, G. A., Fisman, D. N., & Mittleman, M. A. (2009). Infectious disease in a warming world: How weather influenced West Nile virus in the United States (2001-2005). *Environmental Health Perspectives*, *117*(7), 1049–1052. https://doi.org/10.1289/ehp.0800487

Wiesman, J. L. K. L. S. (2019). *epi TRENDS*.

Wimberly, M. C., Giacomo, P., Kightlinger, L., & Hildreth, M. B. (2013). Spatio-temporal epidemiology of human west nile virus disease in South Dakota. *International Journal of Environmental Research and Public Health*, *10*(11), 5584–5602. https://doi.org/10.3390/ijerph10115584

Wimberly, M. C., Lamsal, A., Giacomo, P., & Chuang, T. W. (2014). Regional variation of climatic influences on West Nile virus outbreaks in the United States. *American Journal of Tropical Medicine and Hygiene*, *91*(4), 677–684. https://doi.org/10.4269/ajtmh.14-0239

Appendix

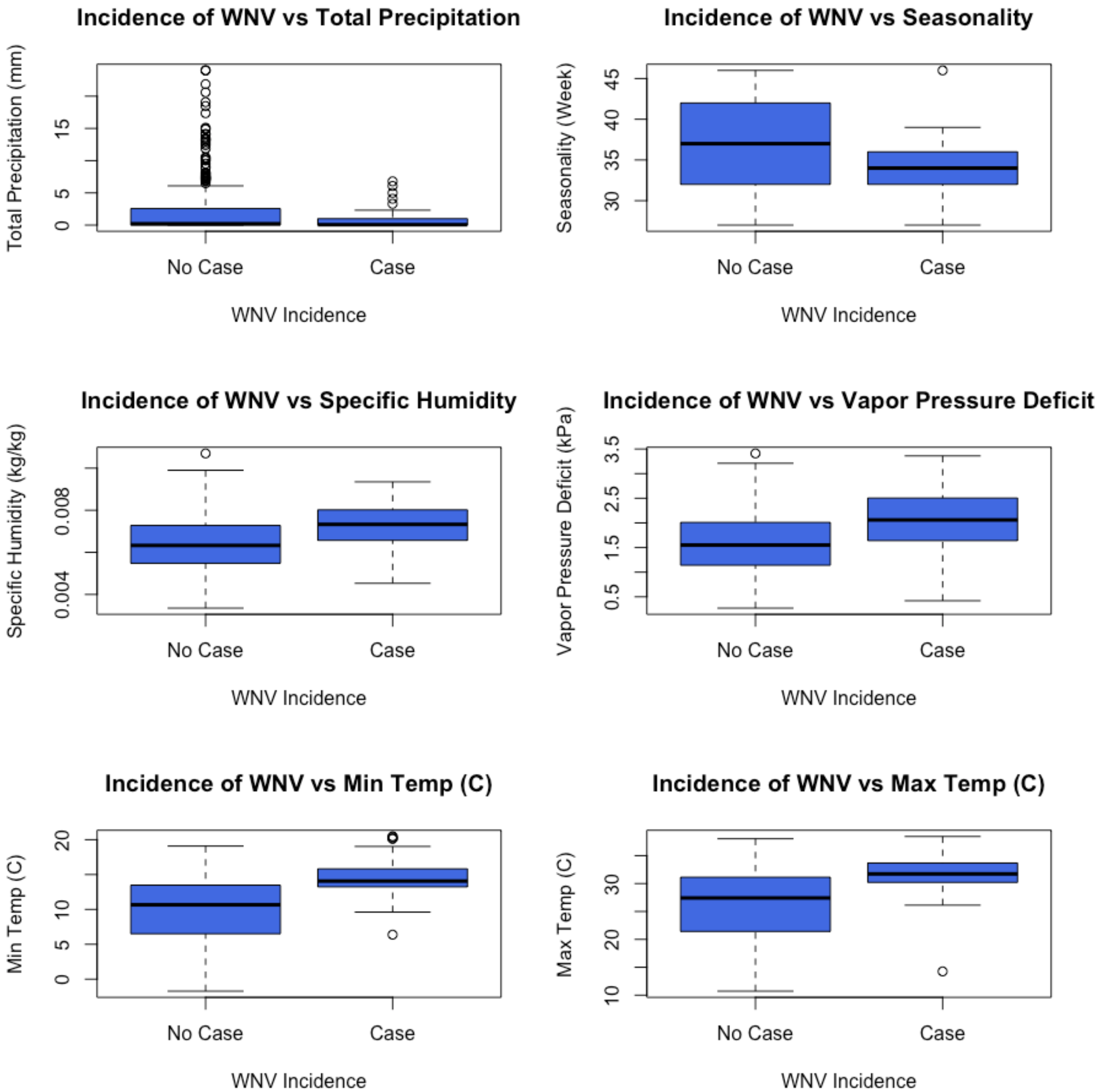


Figure A1. Boxplots represent all weekly mean climate variables from Table 1. and WNV incidence from Yakima and Benton county. Only the weeks where cases occurred between 2008-2019 were included (weeks 27-46).