Modeling Effects of Temperature, Precipitation, and Vegetation on West Nile Virus Infection Rates

Aarush Gupta
Sadhana Kumar
Evan Shephard

1. Abstract

The West Nile Virus (WNV) is one of the most common vector-borne diseases in the United States, with approximately 2,500 cases reported in 2018. Previous research has primarily relied on hydrologic conditions to model WNV transmission and has been unable to find a link between WNV transmission and vegetation, precipitation, or temperature. Statistically analyzed data from the Centers for Disease Control and Prevention (CDC), Global Learning and Observation to Benefit the Environment (GLOBE), and the National Oceanic and Atmospheric Association (NOAA) finds the relationship between the number of WNV cases and three variables (precipitation, vegetation index, and temperature) in the United States, coming up with a comprehensive model for the nation from 2016 to 2019. Visualized vegetation data was extracted from the GLOBE API interface into an interactive ArcGIS dashboard. A significant relationship was found between precipitation and WNV cases in 2019 ($p = 0.0001$) using multiple linear regression and determined that there is a positive correlation between WNV cases and precipitation and a negative correlation with both vegetation index and temperature. The findings indicate that WNV transmission is affected by multiple confounding variables rather than a single environmental factor.

Keywords: West Nile Virus, precipitation, temperature, vegetation index, statistical analysis

2. Research Question

How do vegetation, temperature, and humidity play a role in West Nile Virus outbreaks across the United States?

The purpose of this question is to find patterns for future West Nile Virus (WNV) outbreaks and a correlation between certain environmental factors such as vegetation, precipitation, and temperature which have not been addressed in previous studies. West Nile Virus spreads through the bite of an infected *Culex* mosquito and is one of the most common vector-borne diseases in the United States. Some states in the United States are susceptible to more WNV outbreaks than others and regularly receive more than 100 cases annually. Average values of temperature and humidity between the months of June and September in each of the 48 continental US states were collected along with the vegetation data to find a relationship between these variables and the number of outbreaks. Our hypothesis is that there is a significant relationship between the variables and the number of WNV cases.

3. Introduction

Data for the average temperature and precipitation in each of the 48 continental US states were gathered from the National Oceanic and Atmospheric Administration's Global Time Series tool. Using data from across the continental United States allowed for the analysis of both extremes of all three variables: average temperature, humidity, and precipitation. Temperature and precipitation data were averaged for each state from June through September when the highest transmission of West Nile Virus occurs. This process was repeated for each of the four years, 2016 through 2019. The use of four consecutive years also allowed for a deeper analysis of the relationship between West Nile Virus transmission and environmental factors. 2017 and 2018, for the most part, had more reported West Nile Virus cases than 2016 and 2019, with California being an exception to this trend.

As stated earlier, datasets downloaded from NOAA were averaged in R using the mean() function. The characteristic 'na.rm' was set to 'true', meaning that all rows without a value for a specific data variable were excluded from the average. The function as. numeric() was used when necessary to convert values from characters to numbers if they had not been defined as numbers in the original .csv file. The R package 'tidyverse' was used for data filtering and subsetting. The plot() function in the base R library was used to create the plots; the function abline() was used to create the best-fit line via linear regression, along with axis(), mtext(), and legend() to create the axis labels and legend for the plots.

After the plots were created, a linear regression test was performed on all relationships for all years to determine the significance of the results. A test each was performed for precipitation and temperature for each of the four years, resulting in 8 total linear regression models. Including the two vegetation analyses conducted leads to a total of 10 linear regression tests conducted. The dependent variable, cases, was inputted with each independent variable (precipitation, temperature, and vegetation) into the lm() function, which returned the p-value representing the significance between a perceived positive or negative relationship between two factors.

Vegetation data was extracted from the GLOBE API Interface and uploaded onto an interactive ArcGIS map. Only the average annual values for 2018 and 2019 were available, and for specific counties in the United States. Land cover classifications uploaded onto GLOBE from citizen scientists were being utilized in this project and, therefore, there wasn't an equal distribution of data across the United States like the temperature and humidity variables. The ArcGIS Map was then compared with the ArcGIS CDC map of WNV cases for the years 2018 and 2019. Shown below is the vegetation map, with bigger circles representing denser amounts of vegetation and smaller circles representing less vegetation. For the WNV map, darker greens represent a higher number of cases in the particular region. Although no statistical or further analysis was conducted on this data, a general trend was identified between the variables.

Figure 1. Average temperature and humidity were analyzed for each of the 48 continental United States through the years 2016 and 2019.

4. Results

By analyzing all 4 plots made in R for each of the years of 2016, 2017, 2018, and 2019, the correlation of precipitation and temperature on West Nile Virus cases remained inconclusive. Using multiple linear regression and statistical analysis, it was determined that neither precipitation nor temperature were statistically significant factors in WNV cases as the calculated p-value was above 0.05. There was, however, a significant relationship between precipitation and WNV cases specifically in 2019 ($p = 0.0001$). Figures 2, 3, 4, and 5 show the relationship between each variable and the number of cases for each state. These figures do not show a strong association between the variables.

Although a statistical analysis between vegetation and the number of cases was not conducted, a general trend was noticed by comparing the ArcGIS vegetation map with the West Nile Virus cases map provided by the CDC, as seen in Fig 6 and 7. It was found that locations with relatively higher amounts of vegetation led to fewer cases and areas with fewer amounts of vegetation generally had more cases. Due to time constraints, this was not analyzed further.

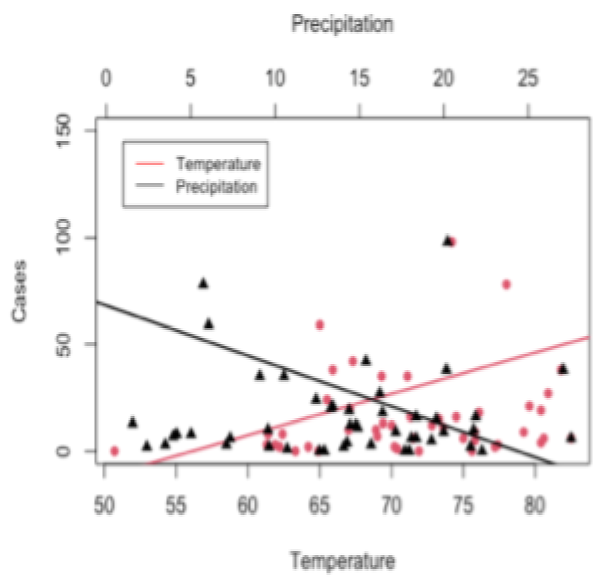


Figure 2. Plot representing the relationship between average temperature, average precipitation, and WNV cases in 2016.

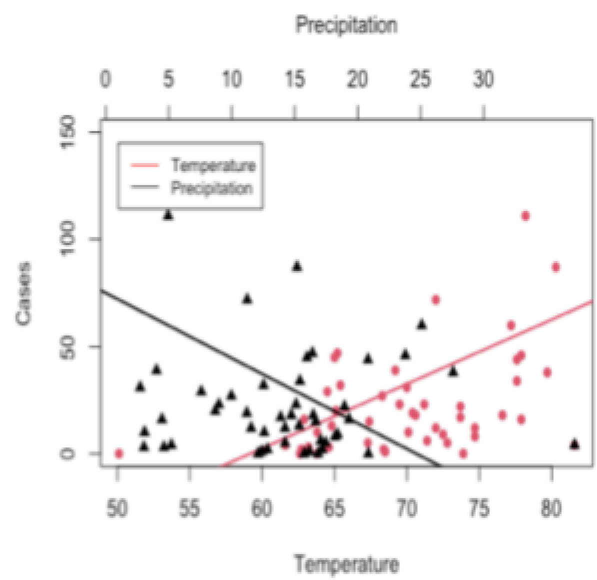


Figure 3. Plot representing the relationship between average temperature, average precipitation, and WNV cases in 2017.

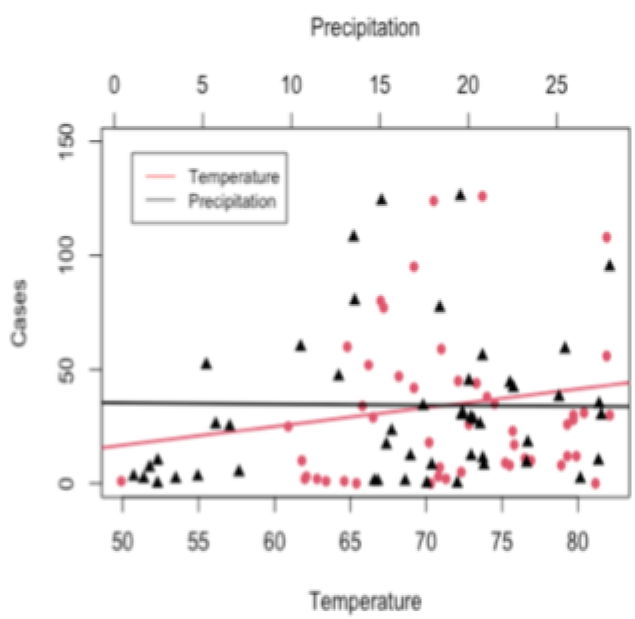


Figure 4. Plot representing the relationship between average temperature, average precipitation, and WNV cases in 2018.

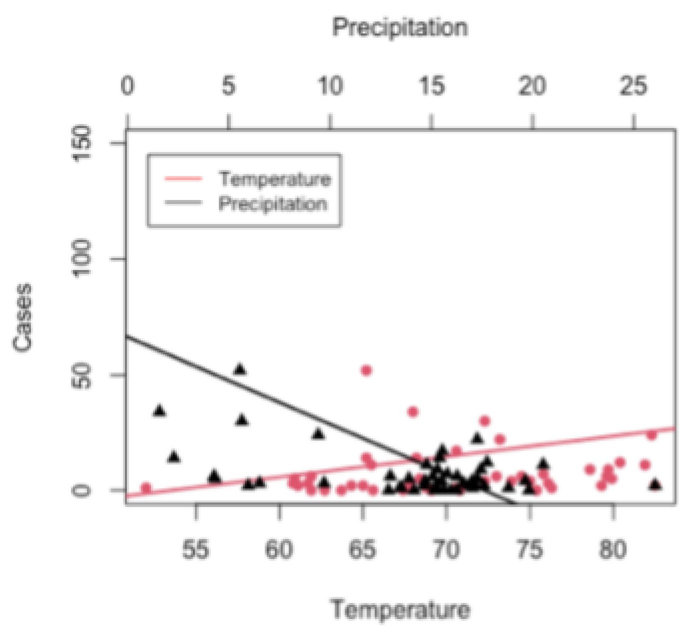
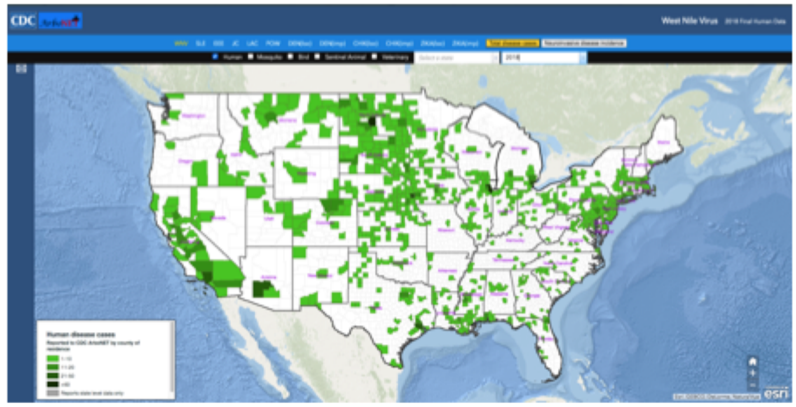


Figure 5. Plot representing the relationship between average temperature, average precipitation, and WNV cases in 2019.



Figures 6 & 7. Figure 6 (left) represents vegetation data from GLOBE in 2018. Figure 7 (right) represents the WNV map from CDC. It was noticed that a general trend exists during the years 2018 and 2019.

5. Discussion

This paper did not find any significant relationships between temperature or humidity with vegetation. With vegetation to be further analyzed, the results reject the hypothesis of finding a significant relationship between the variables and the number of cases. One potential error in this paper includes analyzing average humidity and temperature values which may not have accurately predicted change throughout the years. This may have resulted in an inaccurate analysis and skewed results. Another potential error is not accounting for additional variables such as population differences between states or the number of mosquitoes. The paper is also based on the inference that higher mosquito numbers lead to a greater number of outbreaks. This directly conflicts with the finding of Mori et al. stating that this inference may be ill-founded. Furthermore, Mori et al., finds relationships between mosquito populations and climate variables such as precipitation to be complex, suggesting more variables need to be analyzed. A study by Shaman et al. supports this and it also finds hydrologic conditions are strongly associated with the number of WNV cases in Colorado. Because this was a more concentrated study based on one location, it may have been simpler to find such relationships between hydrological conditions and WNV cases in Colorado as opposed to the entire United States.

The analyses are important as it signifies the complex nature and relationship between environmental factors and infection rates. Finding a relationship between one of the variables may help scientists and health experts to quickly identify where future outbreaks may occur, allowing them to take the necessary precautions to prevent a serious health concern. Mosquitoes have been known to transmit some of the deadliest diseases (with West Nile Virus among them) and modeling data can help scientists to figure out a relationship between the number of cases and possible influential factors.

6. Conclusion

Our research included average values of temperature and humidity during the years 2016 through 2019 provided by the National Oceanic and Atmospheric Association (NOAA) and the number of WNV cases provided by the CDC. We also used hundreds of data points from the GLOBE API Interface to compare the ArcGIS map. As the results in this paper have shown to be insignificant, future work can include looking at a particular region or state and analyzing more environmental features, as mentioned in the discussion above. This concentrated approach is similar to Shaman et al. and may result in significant relationships as average data from the entire United States may not provide accurate results. Additionally, potential follow-up research may include statistically analyzing vegetation data with the number of cases to come up with a definitive analysis.

If a relationship were to be found, the data can be used to model future possible outbreaks. It is also worth noting that this research is based on the inference that higher numbers of mosquitoes lead to a greater number of cases, a theory that other studies have disproved. Therefore, it would be valuable to find the relationship between the number of mosquitoes in a particular region compared to the number of cases. Ideal conditions for mosquitoes to thrive may result in greater numbers of mosquitoes but this may not be associated with a greater number of vector-borne diseases.

We would like to give our thanks to the SEES mentors who were immensely knowledgeable and always ready for answering questions. They gave us the opportunity to conduct research on such an interesting and important topic. Always ready to help, the SEES mentors were an essential part of this research.

7. Acknowledgments

Thank you to Mentors: Dr. Russanne Low, Ms. Cassie Soeffing, Dr. Peder Nelson, Dr. Erika Podest, Dr. Becky Boger, Matteo Kimura

"The material contained in this poster is based upon work supported by National Aeronautics and Space Administration (NASA) cooperative agreements NNX16AE28A to the Institute for Global Environmental Strategies (IGES) for the NASA Earth Science Education Collaborative (NESEC) and NNX16AB89A to the University of Texas Austin for the STEM Enhancement in Earth Science (SEES). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA."

8. References

Shaman, J., Day, J. F., & Komar, N. (2010, February). Hydrologic Conditions Describe West Nile Virus Risk in Colorado. *International Journal of Environmental Research and Public Health*, 494–508. https://doi.org/10.3390/ijerph7020494

Mori, H., Wu, J., Ibaraki, M., & Schwartz, F. (2018, September). Key Factors Influencing the Incidence of West Nile Virus in Burleigh County, North Dakota. *International Journal of Environmental Research and Public Health*, 1928. https://doi.org/10.3390/ijerph15091928

Kilpatrick, A. M., Meola, M. A., Moudy, R. M., & Kramer, L. D. (2008, June). Temperature, Viral Genetics, and the Transmission of West Nile Virus by Culex pipiens Mosquitoes. In M. J. Buchmeier (Ed.), *PLoS Pathogens* (p. e1000092). Public Library of Science (PLoS). https://doi.org/10.1371/journal.ppat.1000092

Johnston, B. L., & Conly, J. M. (2000). West Nile Virus - Where did It Come from and Where Might It Go? *Canadian Journal of Infectious Diseases*, 175–178. https://doi.org/10.1155/2000/856598

Final Cumulative Maps & Data for 1999–2019. (1999–2019). [Dataset]. Centers for Disease Control and Prevention.

ArcGIS Web Application. (2003–2021). *Interactive West Nile Virus Map* [Dataset]. Centers for Disease Control and Prevention.

https://wwwn.cdc.gov/arbonet/Maps/ADB_Diseases_Map/index.html

NCEI.Monitoring.info@noaa.gov. (1880–2021). *Climate at a Glance | National Centers for Environmental Information (NCEI)* [Dataset]. National Oceanic and Atmospheric Association. https://www.ncdc.noaa.gov/cag/global/time-series