West Nile Virus: Relation to Population Density, Precipitation, and Elevation

Diya Deepak, Ananya Gummadi, Asher Jiang, and Harry Wu

2022

Abstract

Vector-borne diseases are spread to humans by a variety of arthropods. The Culex mosquito stands out as one of the most potent transmitters of many diseases, such as the West Nile Virus. Past efforts to map a correlation between population density and the rate of this disease spread have mostly reported inconsistent results. Therefore, socioeconomic class and type of human settlements must be taken into account in order to confidently establish a correlation, particularly in areas that contain urban settlements. Despite past inconsistent results, determining a correlation between environmental factors and the rate of spread of vector-borne diseases is crucial for the safety of humans living in many regions around the world. This project will focus first on West Nile Virus data from the state of Texas, before extending the methods used to other regions and diseases. Using data obtained from various sources and a symbolic regression tool in the form of HeuristicLab, this research project aims to find definitive correlations between the amount of WNV cases and the population density, precipitation level, elevation, and temperature of Texas counties. After obtaining the necessary data and running the symbolic regression tool, the equation generated displayed a very strong correlation between population density and WNV cases, and a much weaker correlation between WNV frequency, precipitation levels, and elevation. There was no correlation with temperature. In the future, with greater availability of West Nile Virus data, including non-environmental factors that affect an individual's susceptibility to this disease, more accurate predictions could be used to issue warnings and mitigate the effects of WNV transmission. Additionally, by modifying the equation to work with different programs and programming languages, it can be used in a similar fashion to predict outbreaks and identify correlations between other diseases.

Keywords: West Nile Virus, symbolic regression, outbreak, population density

Introduction and Literature Review

Although seen as small and insignificant, mosquitos are largely considered one of the most dangerous animals on the planet and it is largely advised that individuals take precautions against mosquitoes with "the use of insect repellant, ...long-sleeved shirts and pants, and [controlling] mosquitoes outdoors and indoors" (CDC, 2020). Of the over 3,000 species of mosquitoes and although "many have the potential for carrying and spreading various diseases", 3 species stand out: Aedes, Anopheles, and Culex (Mosquito-Authority, n.d.).

Aedes mosquitos are common vectors of large-scale diseases like Yellow Fever, Dengue, and Zika and are characterized by their black and white striped legs and biting during the day. Anopheles mosquitos are recognized as being "responsible for the spread of probably the most well-known mosquito-borne illness in the world" called malaria and typically bite late at night (Mosquito-Authority, n.d.). Last but not least, Culex mosquitos are the species at the center of this research project due to their transmission of the West Nile Virus. The purpose of this research is to identify a possible link between population density and the transmission of this disease which may or may not be anthropogenically influenced by different types of environmental factors.

With more than 55% of the world's population living in urban areas, the possibility of a relation existing between population density and the transmission of mosquito-transmitted vector-borne disease cannot be overlooked due to the future implications that a possible, existing correlation could have. Due to predictions of a larger disparity between the world's population that lives in urban or rural areas, identifying existing influences on the transmission of the disease is critical to mitigating mosquito oviposition that could lead to the large-scale spread of illnesses. This research project aims to act as a preemptive form of disease control by identifying the conditions that support mosquito-transmitted disease and preventing conditions from being like so. In that way, "the best way to prevent West Nile is [not] to protect yourself from mosquito bites" but to use similar modeling tools to prevent the spread of the disease before outbreaks can occur (CDC, 2020). Precipitation and elevation are variables that this research project placed great focus on as by collecting information on both of these factors and the species of mosquitoes found there, certain regions can be recognized as being more or less susceptible to certain diseases for housing certain types of mosquitos. In addition, this research places a focus on population density as the number of people living in a certain area in close proximity results in a different rate in the spread of disease than those who live in rural areas. By identifying any correlation between mosquito-transmitted vector-borne diseases and said environmental factors, the effects of fluctuations in population density and possible implications of disease outbreaks can be recognized and prepared for in areas of varying population size.

The relevance of mosquitos has been recognized largely due to their abilities to act as vectors for the transmission of diseases such as West Nile Virus - which "is spread to people through the bite of an infected mosquito" (CDC, 2020). The researchers placed mosquitoes at the center of the project due to the complex way that "mosquitoes contract and transmit illnesses... [in] five steps" (Mosquito-Authority, n.d.). These steps - starting from "a female mosquito taking a blood meal" to ending with another blood meal highlight the simplicity of the process of infection by mosquitoes and provides a clue as to why diseases are spread so quickly when mosquitoes act as vectors (Mosquito-Authority, n.d.). The data obtained will be used to determine the prevalence of West Nile Virus, a type of vector-borne disease, in each county in the state of Texas. By narrowing the data collection to only one state, and only one disease, it is possible to obtain more accurate models and in the future, develop a system to expand the results found today.

In the case of West Nile Virus disease, the importance of staying diligent in regards to mosquito bite prevention is especially critical as "there is no vaccine to prevent WNV infection" (CDC, 2020). The research question hopes to identify a connection between attributes like population density, precipitation, and elevation with the spread of disease: How do various environmental factors affect the spread of West Nile Virus in Texas counties? The mapping of a relationship between population density and the rate of disease spread in the past has largely produced mixed results. Therefore, to firmly establish a correlation, especially in areas with urban settlements, socioeconomic class and kind of human settlements must be taken into consideration. The safety of people living in many parts of the world depends on establishing a link between environmental conditions and the rate at which vector-borne diseases spread, notwithstanding the conflicting results that have previously been reported. Finding a correlation between population density, type of settlements, and other environmental factors, like precipitation and elevation, and the rate at which diseases spread, is especially important for nations that are rapidly industrializing, as many areas may experience rapid population increases without a proportionate improvement in medical treatment and prevention facilities.

Research Methods

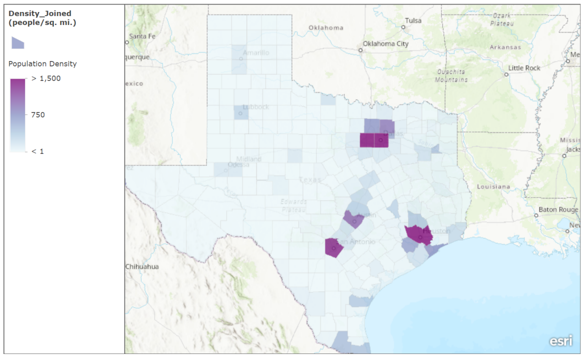
This research project was conducted by collecting field data from several relevant sources. First, data from the CDC was compiled on the amount of West Nile Virus infections in each county in Texas from 2004-2018. The West Nile Virus was chosen specifically because of the vast amount of data available in comparison with other mosquito-transmitted diseases. The state of Texas was chosen in particular as it has a large amount of West Nile Virus cases, second only to the state of California. Below is an excerpt from the CDC data that shows the amount of West Nile Virus cases by county.

County	WNV								COUNTY TOTAL
	M	A	E	SC	H				
					WNF	WNND	PVD‡	TOTAL	
Angelina					1			1	1
Bailey						1		1	1
Bell	5							0	5
Bexar	2					1		1	3
Brazos	1				1	1		2	3
Briscoe					1			1	1
Cameron	7					1		1	8
Carson						1		1	1
Castro						1		1	1
Chambers					1			1	1
Collin	23				2	2		4	27
Cooke	1		1					0	2
Dallas	203		2		2	11		13	218
Deaf Smith			1					0	1
Denton	23		2			2		2	27
Ector						1		1	1
El Paso	20				3	3	1	6	26
Ellis						1		1	1
Fort Bend	3					2		2	5
Galveston						1		1	1
Grayson	2							0	2
Gregg						2		2	2
Guadalupe					1			1	1
Hale					1			1	1
Hansford			1		1			1	2
Harris	310	6	1		8	31	13	39	356
Haskell						1		1	1
Hidalgo			1		2	2	3	4	5
Hunt			1			1		1	2
Jefferson	8							0	8
Johnson	5					2		2	7
Jones			2					0	2
Kaufman						1		1	1
Lavaca			1					0	1
Lubbock	5				1	3		4	9
Mason			1					0	1
Midland					1	1	1	2	2
Montague			1					0	1
Montgomery	105				3	8	2	11	116
Navarro						1		1	1
Nueces	1							0	1
Orange	3					1		1	4
Potter	1						2	0	1
Randall			1		1		1	1	2
Roberts						1		1	1
Rockwall	1							0	1
Runnels						1		1	1
Smith					1	3		4	4
Tarrant	272		1		6	12		18	291
Taylor			1					0	1
Tom Green						1		1	1
Travis	8				1	3		4	12
Trinity						1		1	1
Walker						1		1	1
Webb						1		1	1
Wharton						1	1	1	1
Wichita	1		1					0	2
Williamson	11							0	11
Total Number of Reports	**1,021**	**6**	**19**	**0**	**38**	**108**	**24**	**146**	**1,192**

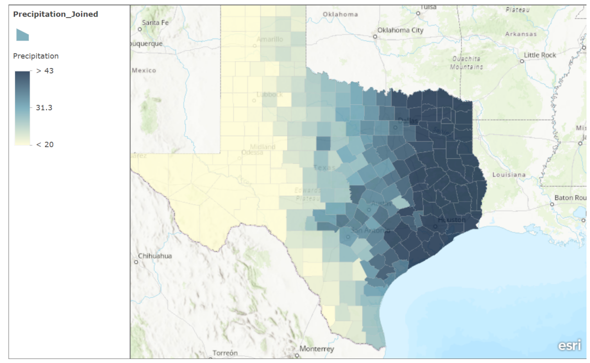
M-Mosquito A-Avian E-Equine SC-Sentinel Chicken H-Human
WNV-West Nile Virus WNF-West Nile Fever WNND-West Nile Neuroinvasive Disease
PVD-Presumptive Viremic Blood Donor
‡PVDs are not included in any of the "Total" columns.

[Texas Department of State Health Services. (n.d.). *DSHS Arbovirus Weekly Activity reports*. Texas Department of State Health Services. Retrieved July 28, 2022, from https://www.dshs.texas.gov/idcu/disease/arboviral/westNile/reports/weekly.aspx]

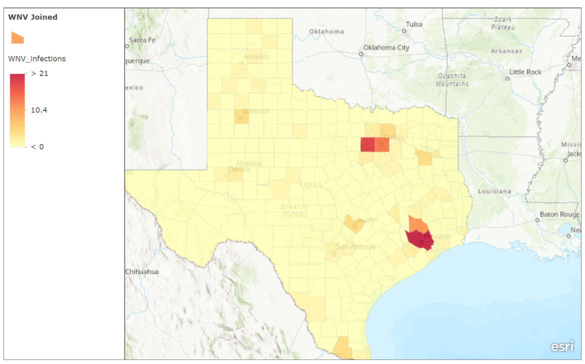
Next, GLOBE data was incorporated in the form of an ArcGIS map that could be overlaid over our final map. Other online sources provided the environmental data that the project needed on a variety of factors, such as the amount of precipitation or altitude of each county. After preliminary analysis of these sources: the four variables were determined: Population Density (D) in people/mi^2, High temperature in July (T) in Fahrenheit, Annual Precipitation (P) in inches, and Elevation (E) in feet. Population density was chosen because it was hypothesized by researchers that more people in a particular area would lead to more potential hosts, increasing the probability of disease spread. Climate data (such as precipitation and temperature) were compiled because our preliminary research showed that mosquito population is impacted by both temperature and precipitation (Kumar, P., & Sarthi, P., 2022). In particular, data for the max temperature per year and mean yearly rainfall were collected from the years 1971-2000. Lastly, elevation was chosen as a variable as it has been proven earlier that most mosquito species prefer lower altitude locations (AMCA, 2022). Below are maps created with the data compiled from these sources.



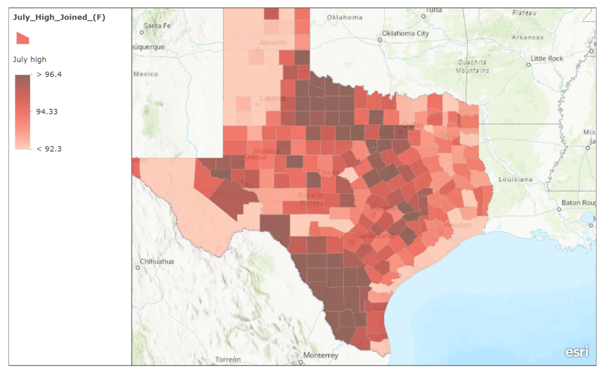
Population Density



Precipitation



WNV Cases

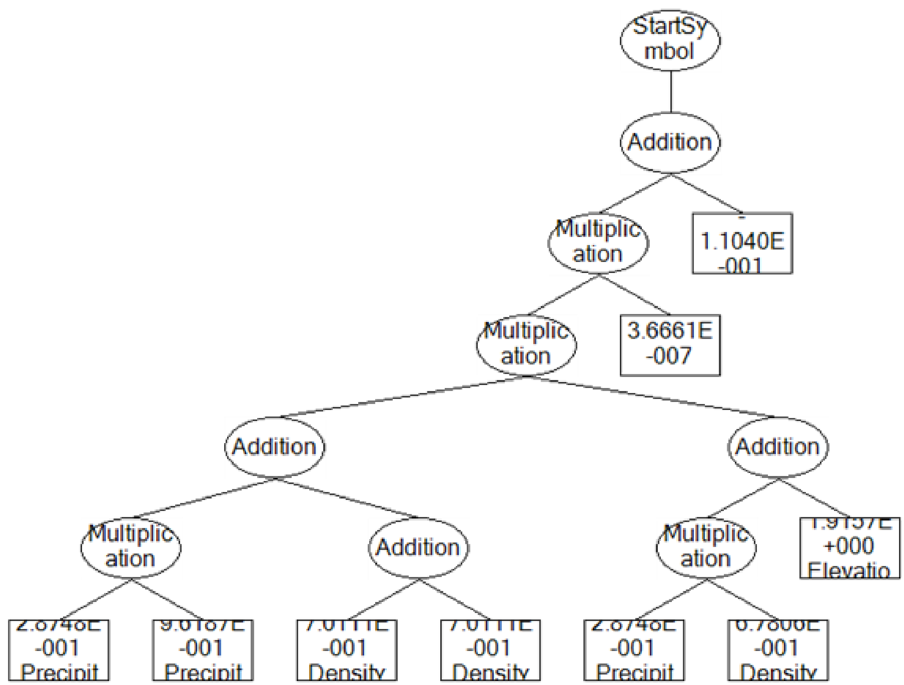


Temperature

Heuristic Labs' symbolic regression function (a modeling program) was then used to determine correlations between each environmental factor and WNV cases. Using population density and other climate variables as inputs, Heuristic Labs generated an equation that would allow for inputs of measurements of elevation, precipitation, and population density to produce predictions of West Nile Virus cases based on the specific conditions.

In order to compensate for unknown information and to simplify the model, several assumptions were made. First, the researchers assumed that the only factors affecting the spread of West Nile Virus were the ones taken into account and that the amount of WNV infections and population density of each region is roughly constant from 2004 to 2018. Next, it was assumed that the July high temperature is representative of temperature's effect on climate for the entire year. Additionally, climate change was effectively ignored and it was assumed that the maximum temperature and mean yearly rainfall between 1971-2000 would stay consistent every year after.

In order to prevent overfitting, the researchers used a 66-34 ratio for training test segments, as opposed to the industry standard which is 80-20. Additionally, to mitigate overfitting, the complexity of the model was limited to a Depth of 10 and a Length of 20. Below is an image of the decision tree that was used to produce the final equation.



Results

HeuristicLab returned the following expression for projected cases of WNV after the environmental data was plugged in. In the equation, P has been substituted for precipitation, D for population density, and E for elevation.

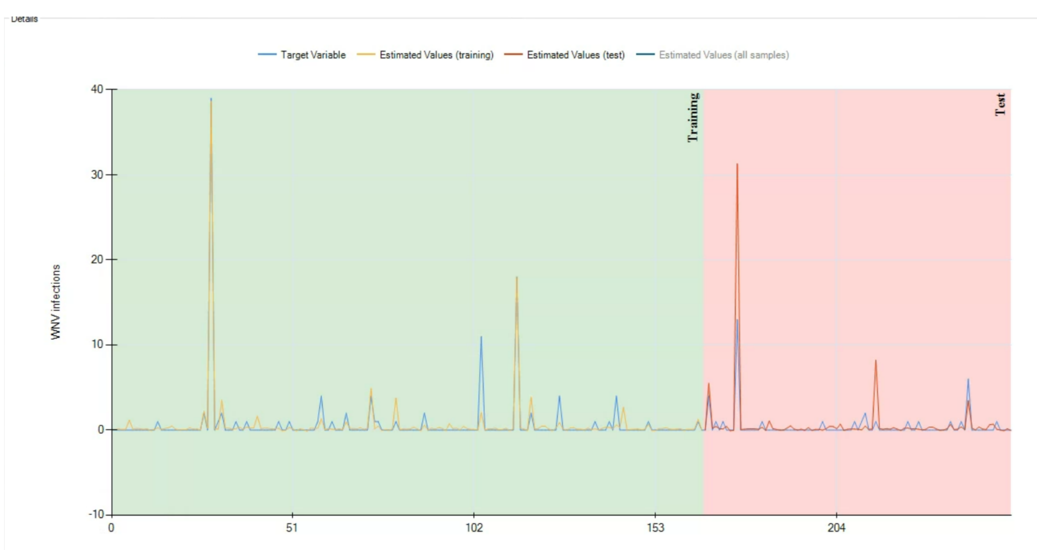
$$((0.2874 * P * 0.9618 * P + (0.7011 * D + 0.7011 * D)) * (0.2874 * P * 0.6780 * D + 1.9156 * E) * 3.6660 * 10^{-7} - 0.1103)$$

Discussion

The equation that we obtained, along with the variable impacts, shows a strong correlation between population density and West Nile Virus cases. However, a weaker correlation occurs with precipitation and elevation, and no correlation at all with temperature. This is seen in the correlation coefficients of all the variables where population density has a number close to 1, while precipitation and elevation have much lower numbers, and temperature has a coefficient of 0.

Density	0.92544560170148371
Precipitation	0.2496272955344746
Elevation (ft.)	0.011630938689099368
July High Temp	0

This is also modeled in the image below where the target variable and estimated values (test, training, and all samples) are plotted against the amount of West Nile Virus infections. It is important here to note that all the variable coefficients are positive, pointing to a direct correlation. These findings support the initial hypothesis in the sense that population density and West Nile Virus cases show a strong correlation. However, a much stronger correlation was expected for precipitation, elevation, and temperature than the given results of this study.



With an r-squared value of 0.925, the correlation between population density and number of West Nile Virus cases is very strong. This is consistent with our hypothesis that there would be a direct correlation between the aforementioned variables, due to many diseases spreading more rapidly in crowded environments. The correlations with precipitation and elevation are much weaker, most likely because there are too many environmental factors at play for a single factor to have an overwhelming impact on mosquito populations. The positive correlation with precipitation is consistent with the citizen science observations made by the GLOBE Mosquito Mappers Initiative. Precipitation often results in pools of standing water, which mosquitoes use to breed. The elevation difference between different Texas counties was also relatively minor, and therefore, it didn't have much of an impact on the final result. It is possible that the correlation found between WNV and elevation was a result of overfitting, which was a common issue that we ran into while running the data through HeuristicLab. The program found no correlation between West Nile Virus cases and temperature, which was most likely due to the minute temperature variations between the counties. Also, the July high temperature of each county doesn't necessarily reflect the average temperature. In addition, heat waves or other forms of extreme weather could have skewed the data. Potential sources of error include assuming a uniform density of people in each county, and that the precipitation and temperature data, which we got from 2000, would stay constant up until 2018.

Some of the limitations faced in this project were in regards to the time constraint and due to the sparsity of the data available to the researchers. There were only a few weeks allocated to conduct research on this project, but with more time, a different level of complexity could have been reached by working with factors beyond the current environmental variables identified as the focus of this study. In addition, the GLOBE data obtained and used in the analysis of this project was often from the same concentrated areas of Texas and lacked broader information from the more rural counties of Texas. This caused the researched to be influenced and centered primarily around the more populous areas of Texas (which had more submissions for the citizen science aspect of this study) rather than Texas as a whole.

Conclusion

Although it was necessary to work under the premise of certain assumptions and simplify the model due to limitations of time and sparse data, the equation that was obtained through the modeling program demonstrated a strong correlation between population density and West Nile Virus cases. The equation also indicated a weaker correlation of the cases with precipitation and elevation and did not identify any correlation with the temperature variable.

Albeit the length of the research project was small, the implications and uses of the results collected are far-reaching. The idea that there exists a strong correlation between population density and West Nile Virus cases and other existing correlation between WNV cases and certain environmental factors like precipitation and elevation provides a focus for research into preventing outbreaks of other mosquito-transmitted vector-borne diseases like Malaria and Dengue Fever. One of the greatest attributes to this project can be recognized as the equation yielded by the models created with data from an array of online sources which together works as a method of identifying where outbreaks can occur based on the interactions of various environmental conditions. With the ability to mitigate the numbers of those infected by mosquitoes acting as vectors to transmit diseases, scientists gain a new perspective on the field of public health by gaining the skill to look into the future with the added bonus of understanding why and how.

With greater resources and more time, there is unlimited potential for this project in terms of the next steps as the equation currently only uses a limited subset of data available. With greater availability of West Nile Virus data, including factors other than environmental factors that make an individual more or less susceptible to this disease, more accurate predictions could be used to issue warnings and mitigate the effects of the transmission of this disease by mosquitoes. In addition, by modifying this equation to work with all different programs and programming languages, it can be used in a similar fashion to predict outbreaks and identify correlations between other diseases like "eastern equine encephalitis... and chikungunya virus... [that] do not have specific vaccines or treatments" (Mosquito-Authority, n.d.).

Acknowledgments

Thank you to the NASA SEES Earth System Explorer mentors: Dr. Rusanne Low, Ms. Cassie Soeffing, Mr. Peder Nelson, Dr. Erika Podest, Andrew Clark. A special acknowledgement to our peer mentor for this project, Paxton Calder. She played a vital role by guiding our group in choosing our focus, keeping our research progress on track, and providing helpful insights that ultimately contributed heavily to the success of our project. We're extremely grateful for all of the time and effort that she put in.

Bibliography/Citations

Authority, M. (n.d.). *How do mosquitoes transmit disease?* Mosquito Authority - Mosquito Service. Retrieved July 28, 2022, from https://www.mosquito-authority.com/blog/how-do-mosquitoes-transmit-disease/

Centers for Disease Control and Prevention. (2021, December 17). Final cumulative maps and data. Centers for Disease Control and Prevention. Retrieved July 22, 2022, from https://www.cdc.gov/westnile/statsmaps/cumMapsData.html

Centers for Disease Control and Prevention. (2020, December 7). Prevention. Centers for Disease Control and Prevention. Retrieved July 22, 2022, from https://www.cdc.gov/westnile/prevention/index.html

GLOBE, Globe Data User Guide. (n.d.). Retrieved July 23, 2022, from https://www.globe.gov/documents/10157/2592674/GLOBE+Data+User+Guide_v1_final.pdf

Kumar, P., & Sarthi, P. P. (2022, June). *Meteorological association for prevalence dynamics of Acute Encephalitis Syndrome: a modeling study*. ResearchGate. Retrieved July 28, 2022, from https://www.researchgate.net/publication/352787037_Meteorological_association_for_prevalence_dynamics_of_Acute_Encephalitis_Syndrome_a_modeling_study

Texas Department of State Health Services. (n.d.). *DSHS Arbovirus Weekly Activity reports*. Texas Department of State Health Services. Retrieved July 28, 2022, from https://www.dshs.texas.gov/idcu/disease/arboviral/westNile/reports/weekly.aspx

Texas State Library and Archives Commission. (n.d.). *Texas Counties: 2010 Population Density*. Texas Counties. Retrieved July 28, 2022, from http://www.texascounties.net/statistics/popdensity2010.htm

Texas State Library and Archives Commission. (n.d.). *Texas Counties: Elevation.* Texas Counties. Retrieved July 28, 2022, from http://www.texascounties.net/statistics/elevation.htm

Texas State Library and Archives Commission. (n.d.). *Texas Counties: Mean July High Temperature*. Texas Counties. Retrieved July 28, 2022, from http://www.texascounties.net/statistics/temperature-julyhigh.htm

The American Mosquito Control Association. (n.d.). *American Mosquito Control Association*. AMCA. Retrieved July 28, 2022, from https://www.mosquito.org/default.aspx

The GLOBE Program. (n.d.). Retrieved July 22, 2022, from https://www.globe.gov/