Modeling *Culex tarsalis* Habitat Suitability in the Great Plains of Montana

Submitted in partial fulfillment of the requirements for graduation with honors from the Department of Life and Environmental Sciences at Carroll College, Helena, MT

Teal Bullick
Carroll College, Helena, MT
April 24, 2018

This thesis for honors recognition has been approved for the Department of Life and Environmental Sciences by:

Dr. Grant Hokit, Director,
Department of Life and
Environmental Sciences

Dr. Dan Gretch, Professor
Department of Life and
Environmental Sciences,
Department of Chemistry and
Physics

Dr. Gerald Schafer, Professor
Department of Health Sciences

April 24, 2018

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank Dr. Grant Hokit for his guidance and support throughout this project. I am very grateful for the opportunity to participate in this ongoing research, and will cherish the experience and knowledge shared with me during the process. I would also like to thank Dr. Dan Gretch and Dr. Gerald Schafer for reviewing my research, and my fellow peers at Carroll College for all their work with mosquito processing, especially my fieldwork buddy Rachel Hopkins. Additionally, this work was made possible by the mosquito trapping contributions by MSU County Extension agents, Aaniiih Nakoda College, Chief Dull Knife College, Little Big Horn College, and Salish Kootenai College, personnel working with county health departments, weed and pest districts, and private citizens. This research was supported by Montana Department of Public Health and Human Services, the Howard Hughes Medical Institute through the Precollege and Undergraduate Science Education Program (grant number 52007534), and by the National Institute of General Medical Sciences of the National Institutes of Health (Award Number P20GM103474).

ABSTRACT

As the leading cause of arbovirus encephalitis in the United States, West Nile Virus (WNV) poses a public health risk in the state of Montana where infection rates in mosquitoes can be as high as 15%. Spatial modeling can serve as a tool for predicting outbreaks and directing prevention measures. Models for the entire state of Montana currently exist that predict WNV risk and habitat suitability of the predominant vector species *Culex tarsalis*. After collecting additional mosquito vector samples in the summer of 2017, I used niche modeling techniques with historic and new presence-only data to build a *Cx. tarsalis* habitat suitability model covering only the Great Plains region of Eastern Montana. An average area under the curve (AUC) value of 0.815 after replicate cross-validation testing indicated that my model performed better than a random model (AUC = 0.5) in predicting habitat suitability of *Cx. tarsalis*. Jackknife analyses indicated that land cover type, presence of virulent competent birds, early spring mean temperature, and early spring precipitation were the four most influential environmental variables in predicting optimal habitat for *Cx. tarsalis*. Additionally, my new data was used to perform validations on the previous model. A two-way analysis of variance (ANOVA) was calculated on the statewide model and Great Plains model's predicted levels of *Cx. tarsalis* habitat suitability, and was found to be not significant (p = 0.373), suggesting that both

models are equally good predictors of *Cx. tarsalis* distribution patterns across the state.

INTRODUCTION

As an arbovirus, West Nile Virus (WNV) persists through a mosquito-bird-mosquito life cycle where the bird host acts as a reservoir for virus replication (Lindsey et al., 2010). While humans do not provide the necessary environment to form a viral reservoir and are therefore considered dead-end hosts, WNV and the associated health risks can still be transmitted to humans through mosquito bites (Lindsey et al., 2010). WNV was first detected in the United States in 1999 in New York (Dauphin et al., 2004). By 2003, the virus reached the west coast of the country, including establishment in the state of Montana (Dauphin et al., 2004). Montana is home to 50 species of mosquitoes (Rolston et al., 2016); however, WNV is primarily transmitted to humans by *Culex tarsalis* in this region (Hayes et al., 2005). While 80% of human infections of WNV are asymptomatic, roughly 20% of infected humans develop flu-like symptoms, while less than 1% of cases result in neuroinvasive disease. Therefore, WNV is the leading cause of arbovirus encephalitis in the United States (Lindsey et al., 2010). Particularly in older patients, encephalitis can be fatal (Dauphin et al., 2004). Thus, WNV poses a serious public health risk. Since an effective vaccine has yet to be developed (Dauphin et al., 2004), monitoring and prevention tactics are currently the best methods for reducing this risk.

Spatial epidemiology uses maps and models of disease and vector presence in order to help quantify the risk of disease, and better direct prevention efforts. These models can also be used in predicting future risk (Ostfeld et al., 2005). Hokit et al. (unpublished) created a niche envelope model for *Cx. tarsalis* for the state of Montana based on a variety of environmental factors and mosquito sampling data collected from 2003 to 2015. A WNV risk model was developed as well. Continued validation of these models is crucial to maintaining their usefulness as predictive tools.

After collecting more sampling data during the summer of 2017 in new locations throughout eastern Montana, I incorporated this new information into a habitat suitability model of just the Great Plains region of eastern Montana. Since Montana's mosquito habitat differs depending on the landscape geography, modeling data for the Great Plains separately from the Rocky Mountain and intermountain valley regions, may generate more accurate predictions of *Cx. tarsalis* abundance. After generating my Great Plains model using tools in ArcGIS and the machine learning tool MaxEnt, I aimed to compare the two models in RStudio. I preformed a two-way analysis of variance (ANOVA) on the statewide model and Great Plains model's predicted levels of

Cx. tarsalis habitat suitability in order to test for significant differences in predictive power. My validations and model modifications may improve the current *Cx. tarsalis* habitat suitability and WNV prediction models.

MATERIALS AND METHODS

I. Sample Collection and Processing

During the summer of 2017, I trapped mosquitoes in previously unstudied locations within the Great Plains of Montana. Each trapping site had two CDC miniature light traps baited with CO_2 from compressed gas tanks (CDC, 2003). The traps were set one hour before sunset and ran through the night until collection 1-2 hours after sunrise. The mosquitoes were then transported to Carroll College in coolers for processing. Mosquito samples were sorted for identification of *Cx. tarsalis* and mosquitoes were tested for the presence of WNV using the protocols described in Hokit et al. (2013). A total of 256 sites were sampled and tested in the Great Plains (Figure 1).

II. Model Rebuilding

To build a new niche envelope model specifically for the Great Plains of Montana, I individually extracted 11 environmental data layers from the original *Cx. tarsalis* habitat suitability model (Hokit et al., unpublished) using the Great Plains region as a mask in ArcGIS 10.5. I used the new environmental variable layers and 235 locations of the target species presence, to generate a new *Cx.*

tarsalis habitat suitability model specific to the Great Plains using the machine learning tool MaxEnt (Phillips, 2010). MaxEnt uses cross-validation such that 90% of my data was used for training the model, and 10% was used to test the performance of the model (Elith et al., 2011). This modeling technique was replicated 10 times, and the average results of each replicate produced the final, overall Great Plains model. The performance of the model was evaluated using the area under the curve (AUC) statistic from receiver operating characteristic (ROC) analysis, and jackknife tests demonstrated which environmental variables most influenced the predictive power of the model (Phillips, 2010). Response curves showed the associations between the environmental variables and the presence of *Cx. tarsalis*. MaxEnt generated these response curves by building models using each variable individually (Phillips, 2010). Comparing the AUC values from the original niche envelope model (Hokit et al., unpublished) and the new Great Plains model determined which model had a better predictive performance.

III. Further Model Validation

Using newly collected data from the summers of 2016 and 2017, I performed a two-way analysis of variance (ANOVA) in RStudio to compare the predictive power of the statewide model and the Great Plains model. The average number of *Cx. tarsalis* per trapping night was used as the dependent variable, while habitat suitability and model type served as independent variables. The *Cx. tarsalis* habitat suitability factor contained three levels: low, moderate, and high suitability, which served as threshold values predicted to correspond to trap night mosquito counts. The statewide model and the Great Plains model served as the two levels of the model type variable. The ANOVA tested whether habitat suitability and model type were significantly associated with *Cx. tarsalis* trap counts. Additionally, the ANOVA tested for an interaction between the two independent variables.

RESULTS

The average AUC value of the 10 replicates of the *Cx. tarsalis* habitat suitability model for the Great Plains was 0.815. This value is much greater than the 0.5 value of a random model, which indicates that the predictability performance of the Great Plains model was better than random. The average prediction results of the 10 replicates were depicted geographically to model thresholds of habitat suitability for *Cx. tarsalis* in the Great Plains region (Figure 2).

The jackknife analysis results determined which environmental variables were the most prominent contributors to predicting habitat suitability for the Great Plains model. According to these results, the four most influential environmental variables were land cover type, presence of virulent competent birds, early spring mean temperature, and early spring precipitation (Table 1).

Response curves demonstrated the associations of each of the four most influential variables with the presence of *Cx. tarsalis.* Wetlands and developed land both had positive associations with the presence of the target species. Forests had a negative association. High levels of virulent competent birds had a positive association with the presence of the mosquito vector. The response curve for early spring mean temperature showed there is a positive association between increasing temperature and *Cx. tarsalis* in March and April. Early spring precipitation had a negative association with very dry conditions and very wet conditions, while moderate precipitation had a neutral association with *Cx. tarsalis* presence.

Results of the two-way ANOVA compared the predictive powers of the original statewide model and the Great Plains model (Table 2). There was no significant difference between the model types and their ability to predict average *Cx. tarsalis* counts per trapping night ($p = 0.373$). However, the p value of habitat suitability was significant, which indicates that the levels of habitat suitability predicted by the models are associated with significantly different trap count values. The levels of habitat suitability, low, moderate, and high, were associated with the mean values of trap counts, 13.8, 80.2, and 174.1, respectively.

Lastly, there was no interaction between model type and habitat suitability levels (p = 0.484).

DISCUSSION

The AUC value of the Great Plains *Cx. tarsalis* habitat suitability model (AUC = 0.815) was lower than the reported AUC value for the model of the entire state of Montana (AUC = 0.86; Hokit et al., unpublished). However, the lower AUC may be a result of the smaller sampling data size available for the Great Plains rather than an indication that the Great Plains model has less predictive power than the model of the entire state. Instead, the two-way ANOVA results indicate the two models are equivalent in predicting habitat suitability (Table 2). Comparing the visual representations of the new model to the plains region of the full state model showed that the new Great Plains model predicted a wider distribution of *Cx. tarsalis* in Eastern Montana than did the statewide model (Figure 2). Furthermore, the new Great Plains model predicts larger total areas of the highest and lowest thresholds of habitat suitability than the full state model, which predicts larger areas of moderate and high thresholds (Table 3). Montana is ecologically diverse with regions of plains, forested mountains, and intermountain valleys; the forests of Montana appear to be negatively associated with the presence of *Cx. tarsalis* (Hokit et al., unpublished). It is possible that, the full state model of *Cx. tarsalis* habitat suitability is underrepresenting the predicted presence of the vector species in the Great Plains due to forest data from Western Montana diluting the predictions for the vastly different climate of the plains in Eastern Montana.

Jackknife analysis and response curves results support the findings of previous studies that land cover type, temperature, and precipitation are helpful environmental variables in predicting habitat suitability of *Cx. tarsalis* (Hokit et al., unpublished). Since vector species presence helps indicate the possible spread of WNV, modeling these variables may aid public health prevention efforts for the disease. Although both the statewide and Great Plains models have equal power in predicting habitat suitability of *Cx. tarsalis*, the new Great Plains model may serve as a beneficial tool in building further WNV risk models.

REFERENCES

CDC. 2003. Epidemic/epizootic West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control. Fort Collins, CO, U.S.: Centers for Disease Control and Prevention. 1-78.

Chase JM, and Knight TM. 2003. Drought-induced mosquito outbreaks in wetlands. Ecology Letters. 6(11): 1017-24.

Dauphin G, Zientara S, Zeller H, and Murgue B. 2004. West Nile: worldwide current situation in animals and humans. Comparative Immunology Microbiology and Infectious Diseases. 27(3): 343-355.

Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17(1):43-57.

Environmental Protection Agency. 2004. Wetlands and West Nile Virus. NSCEP. Published online: https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=30005UPA.txt.

Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, and Campbell GL. 2005. Epidemiology and transmission dynamics of West Nile Virus disease. Emerging Infectious Disease. 11(8): 1167-73.

Hokit DG, Alvey S, Glowienka J, and Johnson GD. Climate dynamics drive outbreaks of West Nile Virus in Montana. Unpublished.

Hokit G, Alvey S, Geiger JMO, Johnson GD, Rolston MG, Kinsey DT, and Bear NT. 2013. Using undergraduate researchers to build vector and West Nile Virus surveillance capacity. International Journal of Environmental Research and Public Health. 10(8): 3192-202.

Lindsey NP, Staples JE, Lehman JA, and Fischer M. 2010. Surveillance for human West Nile Virus disease – United States, 1999-2008. MMWR Morbidity and Mortality Weekly Report. 59(SS02): 1-17.

Marra PP, Griffing S, Caffrey C, Kilpatrick AM, McLean R, Brand C, Saito E, Dupuis AP, Kramer L, and Novak R. 2004. West Nile Virus and wildlife. BioScience. 54(5): 393-402.

Ostfeld RS, Glass GE, and Kessing F. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. Trends in Ecology and Evolution. 20(6): 328-36.

Phillips S. 2010. A brief tutorial on MaxEnt. Lessons in Conservation. 3: 108-35.

Rolston MG, Johnson GD, and Hokit DG. 2016. A taxonomic checklist of the mosquitoes of Montana with notes on new geographic distributions. The American Mosquito Control Association. 32(4): 326-28.

TABLES

Table 1. Relative contributions of the environmental variables to the Great Plains model.

Environmental Variable	Percent Contribution to Model
Land Cover	28.1
Virulent Competent Birds	20.3
Early Spring Mean Temperature	16
Early Spring Precipitation	9.5
Late Spring Precipitation	8.2
Shannon Bird Diversity Index	6.2
Late Spring Temperature Fluctuation	3.5
Summer Precipitation	3
Winter Mean Temperature	2.9
Average Heating Degree Days	1.6
Summer Temperature Fluctuation	0.7

Table 2. Results of Two-Way ANOVA with *Cx. tarsalis* count per trap night as the dependent variable.

Source of Variation	Degrees of Freedom	Summary Square	Mean Square	F value	*p* value
Model Type Effect	1	15267	15267	0.799	0.373
Habitat Suitability Effect	2	747542	373771	19.570	< 0.001 *
Model Type x Habitat Suitability	2	27926	13963	0.731	0.484
Residuals	121	2311014	19099		

* – Significant value, $p < 0.05$.

Table 3. Total areas of each habitat suitability threshold within the Great Plains model and the Great Plains portion of the original full Montana state model (Hokit et al., unpublished).

Prediction of Habitat Suitability	Habitat Suitability Thresholds	Area in New Model (sq km)	Area in Statewide Model[1] (sq km)
Lowest	< 0.07	89,474	23,341
Moderate	0.07 – 0.13	298,152	364,286
High	0.13 – 0.30	221,478	300,339
Highest	> 0.30	80,922	62,598

1 – Areas from the Great Plains portion of the model developed by Hokit et al., unpublished.

FIGURES

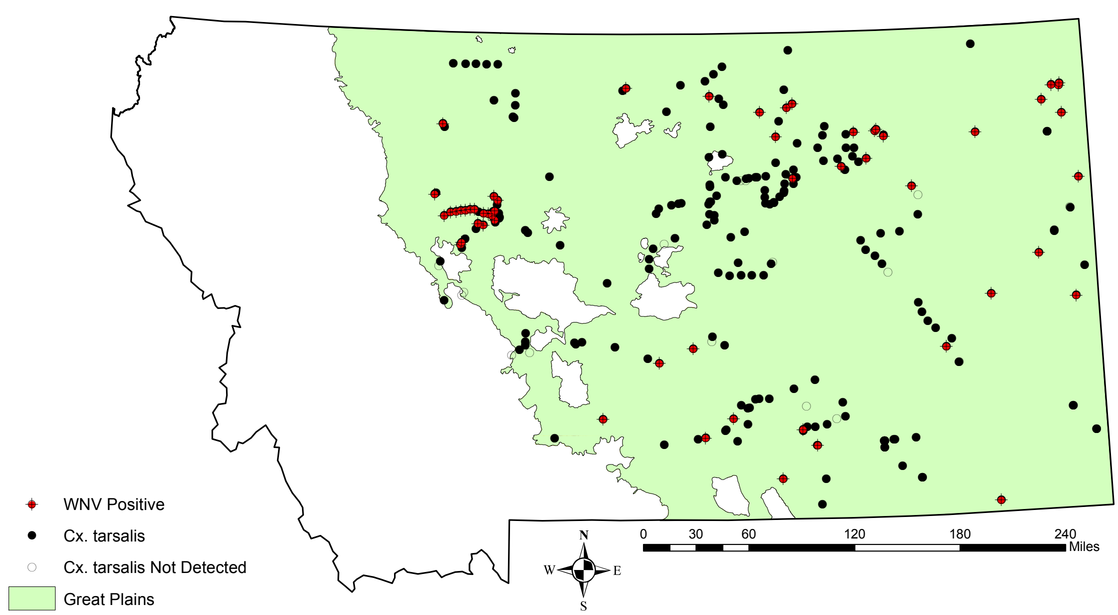

Figure 1. Locations of target species *Cx. tarsalis* (filled circles) and WNV presence (red targets) in the Great Plains of Montana. Sampling sites where *Cx. tarsalis* was not detected are represented with hollow circles.