THESIS

CULICOIDES SPECIES AND LIVESTOCK OVERLAY ANALYSIS: A HABITAT SUITABILITY FRAMEWORK FOR *CULICOIDES INSIGNIS, STELLIFER, AND VENUSTUS* AND POTENTIAL BLUETONGUE VIRUS PRESENCE USING ENVIRONMENTAL AND METEOROLOGICAL VARIABLES TO ENHANCE TRAP DETECTION

Submitted by

Peter James Kessinger

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2022

Master's Committee:

Advisor: Sheryl Magzamen
Co-Advisor: Angela Bosco-Lauth

Joshua Schaeffer
Christie Mayo

ABSTRACT

CULICOIDES SPECIES AND LIVESTOCK OVERLAY ANALYSIS: A HABITAT SUITABILITY FRAMEWORK FOR *CULICOIDES INSIGNIS, STELLIFER, AND VENUSTUS* AND POTENTIAL BLUETONGUE VIRUS PRESENCE USING ENVIRONMENTAL AND METEOROLOGICAL VARIABLES TO ENHANCE TRAP DETECTION

Culicoides spp. midges are blood feeding insects capable of transmitting a variety of pathogens. Of particular concern are Bluetongue virus and Epizootic Hemorrhagic Disease virus. Bluetongue virus is extremely dangerous for ruminants, infecting mainly sheep and cattle, and is a growing concern for areas like the United States. There is little known about the range and habitat preference for *Culicoides* midges, especially in the United States. Our study focuses on predicting habitat suitability for three species of concern: *Culicoides insignis, Culicoides. stellifer,* and *Culicoides. venustus*. Each of these species are linked to the spread or potential spread of Bluetongue virus.

We obtained data from the Southeastern Cooperative Wildlife Disease Study that included the presence and absence data from midge traps for each of the species of interest from 2008-2020. We combined these data with meteorological data and environmental data to generate a habitat suitability model. The maps were then used to predict the probability of midge species presence in that area and create an overlay analysis for each species of midge and livestock of interest: goats, sheep, and cattle. For the statistical analysis, we used both generalized linear models with binomial regression and random forest models to predict potential midge habitat suitability. We then used the AUC scores to determine model fit using both training and test datasets.

Our results indicated that environmental and meteorological variables of significance vary between the species of interest. Most variables were significant for the species of interest, with the most common exception being wind direction. The generalized linear models performed better than the

random forest model overall, with *C. insignis, C. stellifer,* and *C. venustus* having AUC scores of 0.86, 0.70, and 0.71, for generalized linear models respectively.

Overall, prediction models were successful in visualizing and predicting midge presence on the provided environmental and meteorological variables. However, further sampling should be conducted, and variables reassessed for suitability.

ACKNOWLEDGEMENTS

I want to take a moment and thank everyone who has been helping me during this process. First and foremost are my advisors: Dr. Sheryl Magzamen and Dr. Angela Bosco-Lauth. Both have been incredible during this process and have helped me to develop not only my thesis but also my abilities as a researcher. I also want to say thank you to my committee members Dr. Joshua Schaeffer and Dr. Christie Mayo for sitting on my committee and offering feedback and advice during this process.

I also want to say thank you to my mentors at the USDA. Dr. Angela James and Dr. Kelly Patyk for their guidance and mentorship.

I would also like to take this time to say thank you to my friends and family, who have been great supporters during my thesis. I appreciate the time they took to offer advice, hang out, or just listen to me vent.

TABLE OF CONTENTS

CHAPTER 1: BACKGROUND

Culicoides midge species

Culicoides midges are blood feeding insects of the Diptera Ceratopogonidae fly family. They are incredibly widespread and versatile, existing throughout much of the globe with more than 1300 species worldwide[1,2]. These small insects, approximately 1-4 mm in size, can inflict a painful bite on animals and humans and transmit a variety of viruses, protozoa, and filarial nematodes[2,3]. While only a handful of these cause serious illness, *Culicoides* can have serious health and economic outcomes, such as Bluetongue Virus (BTV).

Bluetongue Virus

Bluetongue virus (BTV) is a hemorrhagic orbivirus[1,2] and primarily affects ruminants, with a particular affinity for sheep[1,2]. BTV transmission is mainly through vectors, specifically certain species of *Culicoides* midges, though there has been evidence of sexual transmission from infected bulls[1–3]. In most ruminants, clinical disease signs include lesions of the mouth and nasal cavity, with more severe disease resulting in fever, muscle weakness, and death[1,3]. Mortality and morbidity varies by ruminant species[1]. A signature sign of BTV is the cyanosis that can occur due to a lack of oxygen[1]. This is due to inflammation and fluid buildup in the lungs, contributing to the dark blue tongue often seen in BTV afflicted animals[1]. Deaths caused by BTV result in significant economic impact for agricultural communities and industries[1,3]. One main effect is the economic impact faced by owners that lose animals and the resulting downstream impact on the industry. Even when an animal survives BTV infection, it often has a significant recovery time and has reduced production of goods, such as milk, resulting in economic disruption even without the death of the animal[1,3]. There is no treatment or cure for BTV. Therefore, control measures such as travel bans and restrictions during at-risk times are used to reduce the potential for infection[1].

Environmental Variables

The environment and meteorology of an area plays a significant role in the presence and prevalence of *Culicoides* midges. Environmental factors that contribute to midge abundance include land type, host proximity, soil type, water, and vegetation. Land type or use can influence the midges' feeding habits[4–6]. Factors such as proximity to pastures with grazing ruminants, or forests with wild deer, can influence midge presence[5,6]. Soil type can impact the midge breeding cycle, with certain soils and characteristics of soil (sand fraction, soil pH, and organic content) being conducive to larval stages[7]. This works in tandem with water proximity, with bodies of water influencing soil moisture, soil pH, water pH, and midge prevalence[7]. Vegetation, specifically vegetation cover and biomass production and is often correlated with soil moisture and rainfall[7,8]. Normalized difference vegetation index (NDVI) can be used as a proxy for vegetation and is often associated with *Culicoides* species of interest[7,8].

Environmental factors are also influenced by meteorological variables. Meteorological variables that influence presence and abundance of *Culicoides* spp. include temperature, precipitation, wind speed, and wind direction. Temperature is associated with species survivability, with both upper and lower ranges negatively impacting *Culicoides* species survivability and ideal temperatures for species survivability and reproduction[7–9]. These temperature ranges can also influence feeding rate. For example, higher temperatures are associated with shorter midge life spans, but an increase in blood meals[10]. Temperature can also influence how quickly a midge can become a vector for BTV[10]. Higher temperatures facilitate BTV incubation and can result in uninfected midges becoming vectors sooner after having an infected blood meal[10]. Precipitation is often used as a standard predictor, along with temperature, for *Culicoides* presence[8]. This, in combination with NDVI and soil moisture, can create a comprehensive representation of the influence of moisture on midge presence. Wind speed can also influence successful trapping and surveillance, though the effect of wind direction is still unclear[9]. Overall knowledge of the influence of environmental and meteorological effects on *Culilcoides* species is limited, in part due to challenges within field studies to understand species presence, abundance and life cycle. One major factor that influences field collection of *Culicoides* are methods used to trap species. Once environmental and meteorological variables are properly quantified and linked to midge species presence, their influence on trap characteristics can be adjusted and enhance trapping strategies to allow for better surveillance of species of concern.

Trap Characteristics

Trap characteristics serve important roles in determining the success of a study as well as the potential bias that the study may face. Factors include the type of bait used (light, chemical attractants, host), location, and height, with sub-factors for each of those categories. There are multiple types of trap baits. Light is one of the most common ones for surveillance, especially for *Culicoides*[11]. The type of light can be set to appropriately target the population of interest, though research is ongoing in determining species preference[11–16]. Certain populations of *Culicoides* are more attracted to green wavelength while others prefer blue or ultraviolet (UV) light[12,14]. One study suggested that *Culicoides* could be divided into two groups: UV attracted and green-attracted[12,17]. Furthermore, there are some species, such as *Culicoides sonorensis*, that are more dependent upon wavelength rather than intensity[17]. This contrasts with one recommendation for biting flies, which focuses on high-intensity, short wave-length light-emitting diodes (LEDs)[17].

Further, some studies have indicated that disease status can influence trap attractiveness. *Culicoides sonorensis* studies have shown some preliminary data that BTV status changes how the midge species interacts with traps[15]. Typically attracted to UV light, a *Culicoides sonorensis* midge appears to develop an aversion to UV light once positive for BTV[15]. This finding still needs further research and has interesting implications for future surveillance studies, especially if disease prevalence in vector populations is the outcome of interest.

The success of a light trap can also be influenced by its background, such as foliage. Since the insects have poor vision and operate during twilight and nocturnal hours, increasing the contrast between the trap and its background can aid in successful trapping[11,16,17]. Depending on the species that is being targeted, certain colors can help to increase the contrast and increase the attractiveness of the trap, while other colors can reduce visibility and subsequently decrease collection[17].

Ambient light can also reduce the effectiveness of both contrast and traps by reducing the effective range of light-baited traps[11,17]. Even natural sources, such as a full moon, increase *C. sonorensis* activity but can reduce trap effectiveness due to the increased ambient light[11,17]. Considering ambient light and competing sources, light pollution is a growing problem. Areas that need increased surveillance but are in or close to an urban setting can have increased levels of ambient light and increased amounts of competing light sources that can make surveillance difficult[11,17]. Light pollution is increasing by approximately six percent per year, and as that trend continues, the effectiveness of light traps will continue to decrease[11].

Range is also a contributing factor to light traps effectiveness and is often influenced by the type of light source, intensity, wavelength, contrast, and ambient light. It is estimated that the Centers for Disease Control (CDC) miniature light trap has a range of approximately fifteen meters and the onderstepoort light trap has a range of thirty meters[11,18,19]. However, one study found that the estimated range for the onderstepoort could be as low as two to four meters for *Culicoides*[11].

Trap location plays an important role in trap effectiveness. Understanding the breeding grounds, the feeding habits, and the ideal habitat of the specific *Culicoides* species of interest will aid in the setup and success of traps[16]. Trap rates can be influenced by proximity to hosts, such as sheep or cattle[16].

Trap height is another factor with important consequences for successful and representative midge trapping. Before studies began to investigate the possible relationship between trap success and trap height, there was no rationale as to why traps were set at a chosen height. However, new research has found that height does play a role that varies by species of interest[16,20,21]. An example would be *Culicoides insignis*, which appears to favor tree canopies over the ground level[21]. While still attacking ground hosts to feed, traps set in the canopy had better trap rates than traps set on the ground[21]. Traps were set at 1.37 meters for ground level and 6 meters and 9 meters for 2016 and 2017 respectively[21]. Traps at 6 meters and 9 meters were set up to sample the tree canopy[21]. However, there are still species that favor ground traps over canopy traps[20,21]. Additional research using onderstepoort light traps found that a height of 2.8 meters was most effective when compared to lower heights[22].

Understanding trap characteristics and how they influence successful surveillance campaigns is important for proper analysis of both species' presence and potential disease risk. Future studies and Southeastern Cooperative Wildlife Disease Study (SCWDS) surveillance campaigns should have duplicate traps with differing characteristics, such as modulating height, to allow for comparison and analysis of the influence of trap characteristics on species surveillance. However, before deciding where to place traps for this next study, modeling needs to be conducted to determine suitable locations.

Habitat Suitability Model

Habitat suitability models (HSM), or species distribution models (SDM), are models that use a set of predictors to predict an outcome using machine learning and statistical techniques[8,23,24]. Some common options include generalized linear models (GLM), random forest (RF), boosted regression trees (BRT), maximum entropy models (Maxent), and multivariate adaptive regression splines (MARS)[24]. The common goal of a habitat suitability model is to investigate the potential relationship between species occurrence and environmental variables[24]. These models use environmental and climate variables to predict a species presence as the outcome. This outcome data is usually classified as presence and absence data and uses predictor variables for these presence and absence points to extrapolate the probability of presence at unsampled sites[23,24]. Habitat suitability models are important tools for researchers looking to understand what variables are important for species presence, as well as a tool to predict where a species may occur, making it invaluable for surveillance.

Culicoides midges are understudied but are growing in importance. These midges cause significant livestock and wildlife morbidity and mortality. However, even with these devastating impacts, we have limited research on these midges in comparison to other vectors. Similar to other vectors, *Culicoides spp.* are impacted by both environmental and meteorological variables. Field studies investigating these effects are limited due to a lack of surveillance methods and resources, especially around trapping methods. Further study on habitat suitability could help to prepare better surveillance network.

CHAPTER 2: MANUSCRIPT

Introduction

Culicoides midges are small, blood-feeding arthropods that are part of the Diptera Ceratopogonidae fly family[1,2]. These small insects, approximately 1-4 mm in size, can inflict a painful bite on animals and humans[2,3]. Along with this bite, these flies are also vectors for a plethora of diseases, the main ones being Bluetongue virus (BTV) and Epizootic Hemorrhagic Disease virus (EDHV)[2,3]. Certain species of *Culicoides* have come under increased scrutiny for their ability to transmit these dangerous viruses[1,2]. The damage to both wild and domestic ruminants can be severe, resulting in economic and ecological losses[1].

Bluetongue virus (BTV) is a hemorrhagic orbivirus[1,2] that primarily affects ruminants, with a particular affinity for sheep[1,2]. Transmission is mainly through insect vectors, specifically certain species of *Culicoides* midges, though there has been evidence of sexual transmission from infected bulls[1–3]. Typically, ruminants experience symptoms mainly in the mouth and nasal cavity, resulting in fever, muscle weakness, and death[1,3]. A signature sign of BTV is the cyanosis that can occur due to a lack of oxygen[1]. This is due to inflammation and fluid buildup in the lungs, contributing to the dark blue tongue often seen in BTV afflicted animals[1]. Deaths caused by BTV result in significant economic impact for agricultural communities and industries[1,3]. One main effect is the economic impact faced by owners that lose animals and the resulting downstream impact on the industry. Even when an animal survives BTV infection, it often has a significant recovery time and has reduced production of goods, such as milk, resulting in economic disruption even without the death of the animal[1,3]. There is no treatment or cure for BTV. Therefore, control measures such as travel bans and restrictions during at-risk times are used to reduce the potential for infection[1].

BTV is widespread throughout the United States and is typically found in the southeast, central, and western regions[1]. However, there has been a recent increase in the number of outbreaks moving north[1]. This is related to the change in vector range, as BTV is only transmitted by certain *Culicoides* species. Currently, North America's primary BTV vector is *Culicoides sonorensis*. However, there are several species that are of concern and implicated in possible BTV transmission. These species include *Culicoides venustus, Culicoides stellifer,* and *Culicoides insignis*[9,25]. *C. venstus* and *C. stellifer* are suspected vectors for BTV but are not confirmed [9,25]. *C. insignis* is a confirmed vector for BTV in the Caribbean and South America. There are concerns that it will be able to transmit BTV in the United States as it expands North[9]. This further exemplifies the need for better understanding of meteorological and environmental data for use in midge analysis as changing variables such as temperature and precipitation could change the effective range of these species.

The environment and meteorology of an area plays a definitive role in the presence and prevalence of *Culicoides* midges. Environmental factors that contribute to midge abundance include land type, host proximity, soil type, water, and vegetation. Land type or use can influence the midges' feeding habits[4–6]. Factors such as proximity to pastures with grazing ruminants, or forests with wild deer, can influence midge presence[5,6]. Soil type can impact the midge breeding cycle, with certain soils being conducive to larval stages[7]. This works in tandem with water proximity, with bodies of water influencing soil moisture, soil pH, water pH, and midge prevalence[7]. Vegetation represents vegetation cover and biomass production and is often correlated with soil moisture and rainfall[7,8]. Normalized difference vegetation index (NDVI) can be used as a proxy for vegetation and is often associated with *Culicoides* species of interest[7,8].

These environmental factors are also influenced by meteorological variables. Meteorological variables include temperature, precipitation, wind speed, and wind direction. Temperature is associated with species survivability, with both upper and lower ranges negatively impacting *Culicoides* species survivability and ideal temperatures for species survivability and reproduction[7–9]. These temperature ranges can also influence feeding rate. For example, higher temperatures are associated with shorter midge life spans, but an increase in blood meals[10]. Temperature can also influence how quickly a midge can become a vector for BTV[10]. Higher temperatures facilitate BTV incubation and can result in uninfected midges becoming vectors sooner after having an infected blood meal[10]. Precipitation is often used as a standard predictor, along with temperature, for *Culicoides* presence[8]. Wind speed can also influence successful trapping and surveillance, though the effect of wind direction is still up for debate[9]. Once there is a better understanding of the environmental and meteorological variables, their influence on trap characteristics can be adjusted and enhance trapping strategies and allow better surveillance of species of concern. These species include *C. insignis*, *C. stellifer*, and *C. venustus*, discussed in previous paragraphs, that have been implicated as potential BTV vectors in North America. Both *C. stellifer* and *C. venustus* have been confirmed through laboratory testing to be competent BTV vectors but are unconfirmed in natural environments as playing a major role in BTV transmission[7,25]. *C. insignis* is slightly different, as it is a confirmed BTV vector in the Caribbean and South America[26]. However, it is unknown whether *C. insignis* is a significant vector of BTV in the United States[7]. *C. insignis* began as an exotic species in the United States, but quickly became established in Florida and is moving northward into more southeastern states[7]. Due to all the above, each of these species has become a concern and there is a growing need for improved surveillance on these species to understand their potential role in BTV transmission. To fill in the gaps, it is necessary to properly quantify environmental and meteorological variables to produce a habitat suitability model that can predict midge presence probabilities. Once these areas have been identified, trap characteristics can be modulated to enhance surveillance efforts.

Our goal is to create a habitat suitability model and livestock-midge overlay analysis that can be used to determine a successful surveillance campaign before placing any traps. This habitat suitability would allow users to determine the appropriate place to set up traps for the best possible results for midge surveillance.

Methods

Study Area

The study area was determined by the Southeastern Cooperative Wildlife Disease Study (SCWDS) and focus primarily on the southeastern United States. The states included are Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Missouri, North Carolina, South Carolina, and Tennessee. Trap data were provided at the county level; only counties with trap sites are represented in the study area (Figure 1). Study locations used sites selected for their ability to successfully detect the species of interest, with the goal of capturing the wildlife-livestock interface. However, most sites contained more wildlife area than livestock area due to availability. United States does not have an official midge surveillance network.

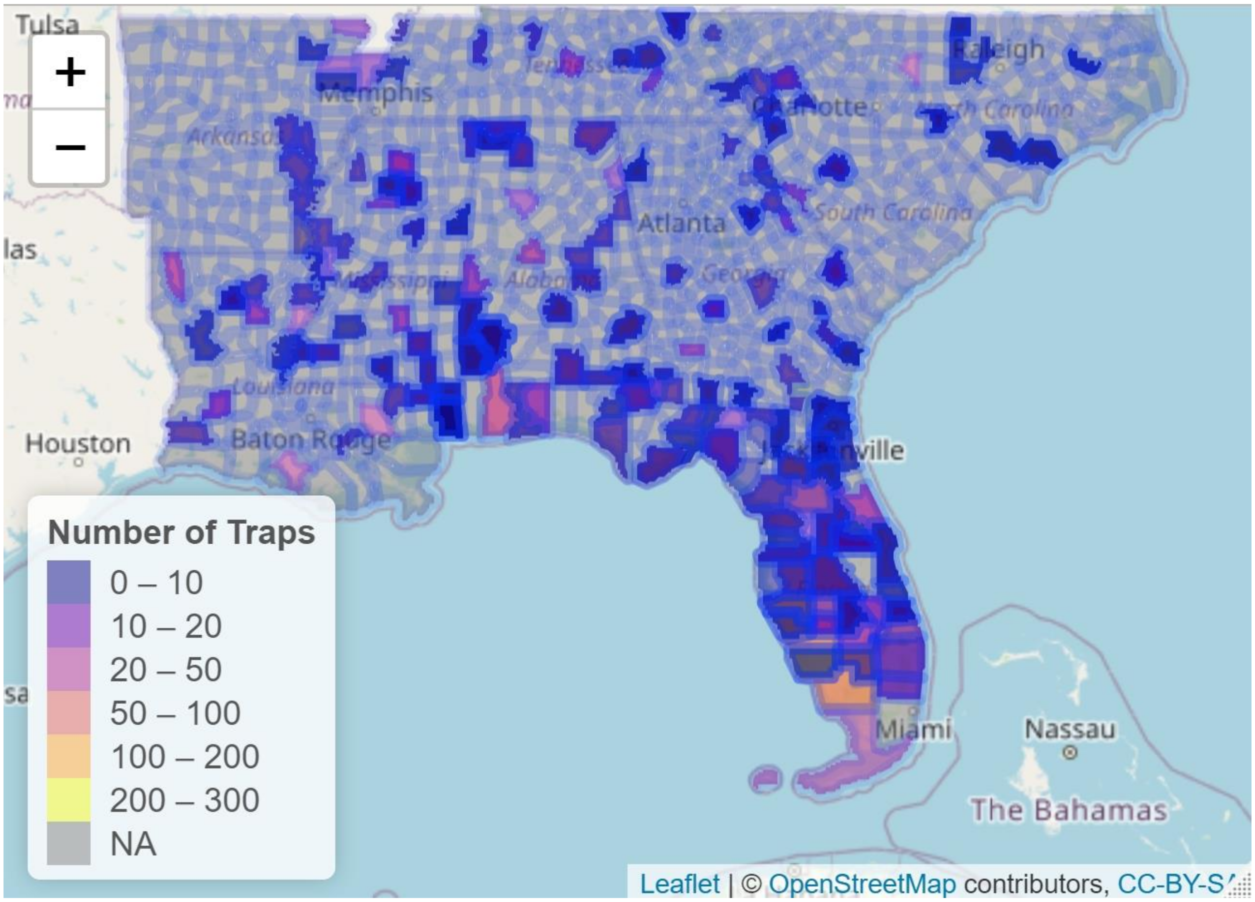

Figure 1: Study area with SCWDS midge trap site locations by county

Midge Population

Our population of interest includes three different species: *Culicoides insignis, Culicoides stellifer,* and *Culicoides venustus*. *Culicoides insignis* is a confirmed vector for BTV, while *C. stellifer* and *C. venustus* have been shown to be competent vectors but are not confirmed. We selected these populations due to a growing concern over exotic midge species and their expansion into new territory in the United States. These observations are counts of traps that detected a certain species, not of species in the traps. Species were collected as presence or absence data. Traps were initially incandescent light for 2008 and part of 2009 but were all changed to ultraviolet light from that point forward. Trap height ranged from 1.5 to 2 meters. Data collection was from July 2008 to October 2020 and included the above study area. Not all traps were constant per location, and more details about the trap placement and function can be found in Vigil et al. 2014 and Vigil et al. 2018.

Covariate selection

The covariates included in this study were selected based on the previous literature and used to create a directed acyclic graph (DAG) (Figure 2). Variables included temperature (maximum and minimum) in degrees Celsius, precipitation in millimeters, wind speed in meters per second, wind direction in degrees, organic soil content, organic soil carbon content, sand fraction, top layer of soil pH, and maximum, minimum, and phase normalized difference vegetation index (NDVI). Each of these variables was involved in previous literature and was used to generate the habitat suitability model. These variables could also influence successful trapping strategies. Furthermore, in reference to the DAG, the soil, pH, and NDVI variables would fall under the habitat node. Data was obtained for each of these variables using the following sources: meteorological data from GridMET and soil and NDVI from ISRIC. Data types and sources of background data to indicate relevant exposures are summarized in Table 1. All data is included at a 4 kilometer resolution.

Table 1: Covariates and Data Sources

Variable	**Source**	**Description**	**Citations**
Temperature	GridMET	1 = Minimum temperature (˚C) 2 = Maximum temperature (˚C)	Sloyer et al. 2019 Zuliana et al. 2015 Mayo et al. 2020
Precipitation	GridMET	1 = annual precipitation (in)	Zuliana et al. 2015
Wind	GridMET	1 = Wind speed (m/s) 2 = Wind direction (degrees)	Mayo et al. 2020

NDVI	ISRIC	1 = minimum NDVI 2 = maximum NDVI 3 = phase NDVI	Steinke et al. 2016 Erram et al. 2019 McGregor et al. 2021 Sloyer et al. 2019 Zuliana et al. 2015 Mayo et al. 2020
Soil	ISRIC	1 = Organic carbon 2 = organic soil content 3 = pH top layer 4 = sand fraction	Steinke et al. 2016 Erram et al. 2019 McGregor et al. 2021 Sloyer et al. 2019 Zuliana et al. 2015 Mayo et al. 2020

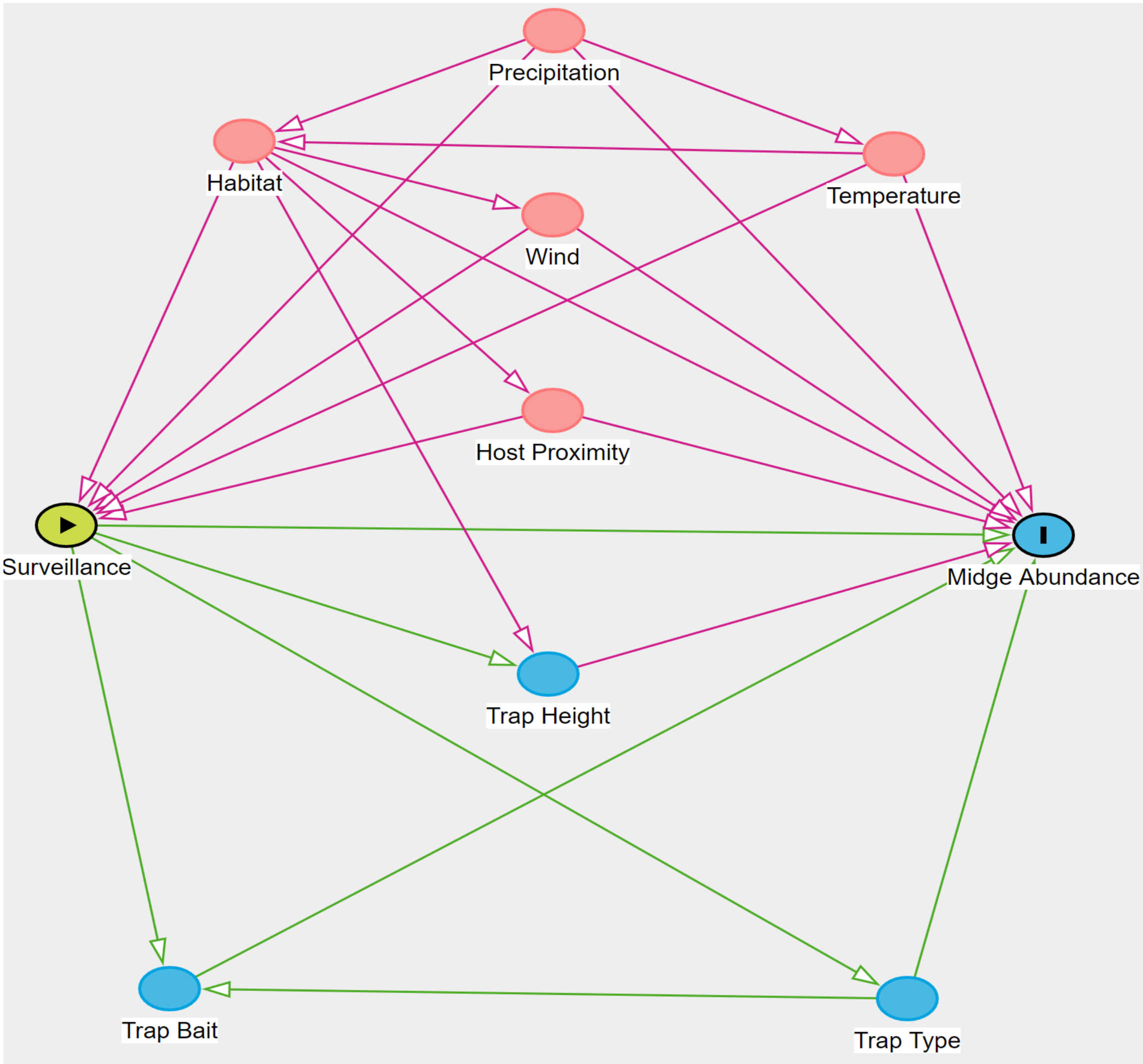

Figure 2: The DAG shows the causal paths of interest we were interested in based on the previous literature. The DAG looks at the relationship between the exposure (surveillance) and the outcome (Midge Abundance) and shows the potential causal paths and confounders that could influence that relationship. This DAG includes things like trap characteristics and host proximity as parts that will need future adjustment and characterization. In this study, we only focused on the environmental and meteorological paths to create the habitat suitability assessment.

Determination of outcome variable

Our outcome of interest for our study was the presence and absence of our species of interest. This was represented as a binary variable, with one being presence detection and zero being absence detection. We also created a fourth outcome of interest that was also a presence and absence binary variable, following the same format as the three species of interest. However, this variable, or the group variable, includes any species of detection, including the fourth species *C. sonorensis* that was provided in the dataset but not used as an individual outcome due to nearly no presence detection. Overall, this fourth variable is described as an "any" or "none" variable, with either any species or number of species being detected as presence, and no species detected as being absence. This group variable, as well as the three species of interest, were our final outcome variables included in the analysis.

Data Cleaning

All data was loaded into R and cleaned[27]. This included the midge dataset from SCWDS, the meteorological data (GridMET), the soil data, and NDVI data. GridMET data was pulled using the ClimateR package[28]. All other data was cleaned using packages from the R CRAN repository. All data was joined into one dataset, with the midge dataset being the parent dataset. Environmental and meteorological data was extracted from parent raster maps for each of the trap sites in the SCWDS midge dataset and used for model building. The parent rasters themselves were combined into a raster stack and we took the median value for each meteorological variable (temperature, precipitation, and wind) over the entire study period. These median values were then stacked with the NDVI and soil data to form our total raster stack for later prediction.

We ran univariable models to determine significance of the chosen variables. All variables had significance for each species in different combinations, with not all variables being significant per one species. Each variable was kept in the model due to the previous literature and significance among different species.

Statistical Model

There were four statistical models. The first three were a binomial model with individual species presence and absence data. These models provided basic statistical analysis and included presence and absence for each respective species as the outcome, with the covariates previously discussed as predictors. The fourth model was a binomial model using the “any” and “none” variable as our species outcome variable.

Prediction

We then make predictions using a habitat suitability framework. We used the SDM package in R to make our predictions[24]. We divided our data into two separate datasets using the caret package in R, a training dataset and a testing dataset[29]. The training dataset had 6181 observations, or 75 percent of the dataset, and our test dataset had 2060 observations (25 percent of the dataset). We were able to feed both the training and test datasets into the SDM framework and create an SDM model. Using this model, we could then use the raster stack we created from earlier that contains all the predictor variables to create a new predicted raster for each species, using a generalized linear model (glm) with the family set to binomial and a random forest classification model.

Overlay analysis

The overlay analysis used the predicted rasters from the habitat suitability model for each species in combination with livestock data from FLAPS (Farm location and Agricultural Production Simulator) to create categories of concern[30]. These categories include low concern, moderate concern, and high concern. We used criteria from both species probability and livestock count to create these categories at the county level. Across all three species, we had cutoffs at less than or equal to twenty percent (0.2), greater than twenty percent (0.2) and less than or equal to fifty percent (0.5), and greater than fifty percent. We combined these with cutoffs for each of the different livestock species: goats, sheep, and cattle. For goats, our cutoffs were less than or equal to 4000, greater than 4000 and less than or equal to 6000, and greater than 6000. For sheep, our cutoffs were less than or equal to 1000, greater than 1000 and less than or equal to 2000, and greater than 2000. For cattle, our cutoffs were less than or equal to 75,000, greater than 75,000 and less than or equal to 150,000, and greater than 150,000. Once we had these cutoffs, we combined them to create the levels of concern. These levels of concern are low, medium, and high concern. High concern indicates a need for better surveillance and prevention or intervention strategies. Low concern indicates no current action is necessary.

Results

Our population includes 8293 observations over the study area. Of those, no midges detected (none) was the largest category, with 5542 observations. *C. stellifer* was the next largest, with 1439 observations. *C. Insignis* had 854 observations.

Table 2: Number of traps that detected the labeled species. If multiple species were trapped, those traps were included in their own row.

Species	Trap Count
none	5542
C. stellifer	1439
C. insignis	854
C. insignis, C. stellifer	181
C. stellifer, C. venustus	172
C. venustus	48
C. sonorensis	19
C. sonorensis, C. stellifer	18
C. insignis, C. stellifer, C. venustus	10
C. sonorensis, C. stellifer, C. venustus	6
C. insignis, C. venustus	2
C. insignis, C. sonorensis	1
C. insignis, C. sonorensis, C. stellifer	1

Results from univariable models for variable selection, from top to bottom of the tables, include maximum temperature, minimum temperature, precipitation, wind speed, wind direction, maximum NDVI, minimum NDVI, phase NDVI, pH of top layer of soil, organic soil content, organic soil carbon content, and sand fraction. The “estimate” column is the exponentiated value from the model. As an example, in table three, a one degree increase in maximum temperature resulted in a approximate 3.4 percent decrease in the odds of *Culicoides insignis* presence. The variables were significant for *C. insignis* except for wind direction. *C. stellifer* had significant variables except for precipitation and wind direction, *C. venustus* had significant variables for all except wind speed, wind direction, and minimum NDVI. The “all” and “none” group had significance for all except precipitation, wind speed, and wind direction. All variable results are exponentiated.

Table 3: Univariable analysis results for ***Culicoides insignis***. Table combines the results of each individual univariable model. Digits were restricted to five significant figures, which resulted in some significant p-values becoming zero.

Term	Estimate	Conf.low	Conf.high
Maximum temperature	**0.96602**	**0.95230**	**0.98009**
Minimum temperature	**0.98381**	**0.97148**	**0.99652**
Precipitation	1.00875	1.00218	1.01497
Wind velocity	**1.11136**	**1.05636**	**1.16862**
Wind direction	0.99938	0.99867	1.00008
Maximum NDVI	**0.20017**	**0.17831**	**0.22427**
Minimum NDVI	**4.12308**	**3.48614**	**4.89341**
Phase NDVI	**2.56481**	**2.39371**	**2.75123**
Soil pH	**1.67584**	**1.53008**	**1.83448**
Organic soil content	**2.73386**	**2.55228**	**2.93233**
Organic soil carbon content	1.05546	0.98535	1.12828
Sand fraction	**4.79229**	**4.34035**	**5.31012**

Table 4: Univariable analysis results for ***Culicoides stellifer***. Table combines the results of each individual univariable model. Digits were restricted to five significant figures, which resulted in some significant p-values becoming zero.

Term	Estimate	Conf.low	Conf.high
Maximum temperature	**1.13105**	**1.11522**	**1.14736**
Minimum temperature	**1.08722**	**1.07362**	**1.10130**
Precipitation	0.99815	0.99203	1.00397
Wind velocity	**0.89226**	**0.84766**	**0.93847**
Wind direction	0.99968	0.99912	1.00025
Maximum NDVI	**2.51048**	**2.27091**	**2.77941**
Minimum NDVI	**1.15706**	**1.04388**	**1.28415**
Phase NDVI	**0.64159**	**0.60699**	**0.67773**

Soil pH	**0.33842**	**0.30294**	**0.37717**
Organic soil content	**0.42736**	**0.39868**	**0.45744**
Organic soil carbon content	**0.70325**	**0.65782**	**0.75104**
Sand fraction	**0.75673**	**0.72370**	**0.79109**

Table 5: Univariable analysis results for ***Culicoides venustus***. Table combines the results of each individual univariable model. Digits were restricted to five significant figures, which resulted in some significant p-values becoming zero.

Term	Estimate	Conf.low	Conf.high
Maximum temperature	**1.04695**	**1.01478**	**1.08112**
Minimum temperature	**1.06603**	**1.03404**	**1.10110**
Precipitation	**1.01228**	**1.00039**	**1.02239**
Wind velocity	0.90961	0.77836	1.05391
Wind direction	0.99986	0.99847	1.00126
Maximum NDVI	**3.37517**	**2.57870**	**4.46366**
Minimum NDVI	0.95236	0.74592	1.22905
Phase NDVI	**0.65892**	**0.57438**	**0.75395**
Soil pH	**0.36788**	**0.27590**	**0.48323**
Organic soil content	**0.40412**	**0.33288**	**0.48502**
Organic soil carbon content	**0.75851**	**0.64201**	**0.89061**
Sand fraction	**0.71221**	**0.63589**	**0.79626**

Table 6: Univariable analysis results for **"any"** midge species in the dataset present versus "none" of the species of interest present. The table combines the results of each individual univariable model. Digits were restricted to five significant figures, which resulted in some significant p-values becoming zero.

Term	Estimate	Conf.low	Conf.high
Maximum temperature	**1.07214**	**1.06010**	**1.08446**
Minimum temperature	**1.05075**	**1.04036**	**1.06137**
Precipitation	1.00257	0.99748	1.00759
Wind velocity	0.99002	0.95090	1.03050
Wind direction	0.99951	0.99901	1.00001
Maximum NDVI	**0.87766**	**0.81416**	**0.94621**
Minimum NDVI	**1.68839**	**1.53487**	**1.85950**
Phase NDVI	**1.05009**	**1.00491**	**1.09726**
Soil pH	**0.79121**	**0.73455**	**0.85159**
Organic soil content	**0.92976**	**0.88924**	**0.97189**
Organic soil carbon content	**0.81025**	**0.76729**	**0.85486**
Sand fraction	**1.21862**	**1.17223**	**1.26704**

We ran four binomial models on species presence for the three species of interest and the grouped binomial variable. These were multivariable binomial models for each individual outcome of interest. All model outputs were exponentiated, producing the percent odds of species detection per one unit increase in that predictor. For example, in *C. insignis*, there was a 4.5% increase in the odds of *C. insignis* presence per one degree Celsius increase in maximum temperature.

Table 7: Multivariable binomial model with C. insignis presence as the outcome of interest and with the environmental and meteorological variables as predictors

Term	Estimate	Conf.low	Conf.high
(Intercept)	0.00330	0.00123	0.00868
Maximum temperature	**1.04548**	**1.00926**	**1.08317**
Minimum temperature	0.98236	0.95730	1.00811
Precipitation	1.00727	0.99822	1.01615
Wind velocity	1.02493	0.96309	1.09050
Wind direction	**0.99896**	**0.99814**	**0.99978**
Maximum NDVI	**0.44023**	**0.33674**	**0.57383**
Minimum NDVI	1.33950	0.91778	1.96751
Phase NDVI	**0.62752**	**0.55984**	**0.70311**
Soil pH	**1.68317**	**1.43702**	**1.97243**
Organic soil carbon content	**1.11140**	**1.01008**	**1.22128**
Organic soil content	**1.43274**	**1.28236**	**1.60107**
Sand fraction	**4.41670**	**3.68799**	**5.31824**

Table 8: Multivariable binomial model with C. stellifer presence as the outcome of interest and with the environmental and meteorological variables as predictors

Term	Estimate	Conf.low	Conf.high
(Intercept)	0.00465	0.00226	0.00942
Maximum temperature	**1.12710**	**1.09820**	**1.15703**
Minimum temperature	0.99095	0.96834	1.01422
Precipitation	1.00555	0.99816	1.01265
Wind velocity	**1.08484**	**1.02141**	**1.15178**
Wind direction	**0.99871**	**0.99794**	**0.99949**
Maximum NDVI	1.08102	0.84017	1.39525

Term	Estimate	Conf.low	Conf.high
Minimum NDVI	0.83415	0.66796	1.04152
Phase NDVI	**0.82626**	**0.73669**	**0.92735**
Soil pH	**0.40733**	**0.32869**	**0.50172**
Organic soil carbon content	**0.72801**	**0.65645**	**0.80637**
Organic soil content	**0.62712**	**0.56147**	**0.69935**
Sand fraction	1.05461	0.92559	1.20218

Table 9: Multivariable binomial model with C. stellifer presence as the outcome of interest and with the environmental and meteorological variables as predictors

Term	Estimate	Conf.low	Conf.high
(Intercept)	0.00365	0.00048	0.02594
Maximum temperature	1.00399	0.93278	1.08033
Minimum temperature	1.04941	0.97889	1.12797
Precipitation	1.01355	0.99707	1.02813
Wind velocity	1.03022	0.85978	1.22452
Wind direction	0.99800	0.99570	1.00032
Maximum NDVI	1.74238	0.77774	3.96218
Minimum NDVI	0.55334	0.27531	1.10795
Phase NDVI	**0.67581**	**0.47446**	**0.96530**
Soil pH	**0.30799**	**0.13934**	**0.63755**
Organic soil carbon content	**0.68553**	**0.48989**	**0.94513**
Organic soil content	**0.63433**	**0.44849**	**0.88046**
Sand fraction	1.34074	0.89688	2.01855

The “any” and “none” model included presence for any detected species per trap and “none was the complete absence of species in the dataset.

Table 10: Multivariable binomial model with any midge presence as the outcome of interest and with the environmental and meteorological variables as predictors

Term	Estimate	Conf.low	Conf.high
(Intercept)	0.01736	0.00976	0.03067
Maximum temperature	**1.10677**	**1.08308**	**1.13116**
Minimum temperature	**0.97519**	**0.95809**	**0.99262**
Precipitation	1.00539	0.99932	1.01136
Wind velocity	**1.04746**	**1.00026**	**1.09679**

Term	Estimate	Conf.low	Conf.high
Wind direction	**0.99880**	**0.99821**	**0.99939**
Maximum NDVI	**0.77697**	**0.65264**	**0.92561**
Minimum NDVI	**1.34727**	**1.13236**	**1.60458**
Phase NDVI	**0.79674**	**0.73297**	**0.86589**
Soil pH	0.92269	0.81658	1.04210
Organic soil carbon content	**0.92824**	**0.86481**	**0.99534**
Organic soil content	0.93100	0.86174	1.00553
Sand fraction	**1.44379**	**1.30576**	**1.59777**

These multivariable models were made with the overall dataset since there was no prediction used for these models. Models were consistent with the significance demonstrated from the univariable models, except that precipitation did not reach the threshold for significance in any of the multivariable models. However, precipitation was kept in the model due to previous literature and use for predictive models.

The predictive models had AUC scores for each model type (binomial generalized linear model or classification random forest). Generalized linear models had higher AUC scores except for *C. insignis*. *C. stellifer* had the lowest model performance overall and *C insignis* was the highest AUC for the three species.

Table 11: AUC score by model type and species. AUC scores above 0.8 are considered good performances. AUC scores of equal to or greater than 0.7 are considered ok performances.

Species	Model Type	AUC Score
C. insignis	glm	0.86
	rf	0.87
C. stellifer	glm	0.70
	rf	0.62
C. venustus	glm	0.71
	rf	0.66

The predicted models then produced rasters with probability of that species presence denoted by color. We used the glm prediction maps due to better AUC scores.

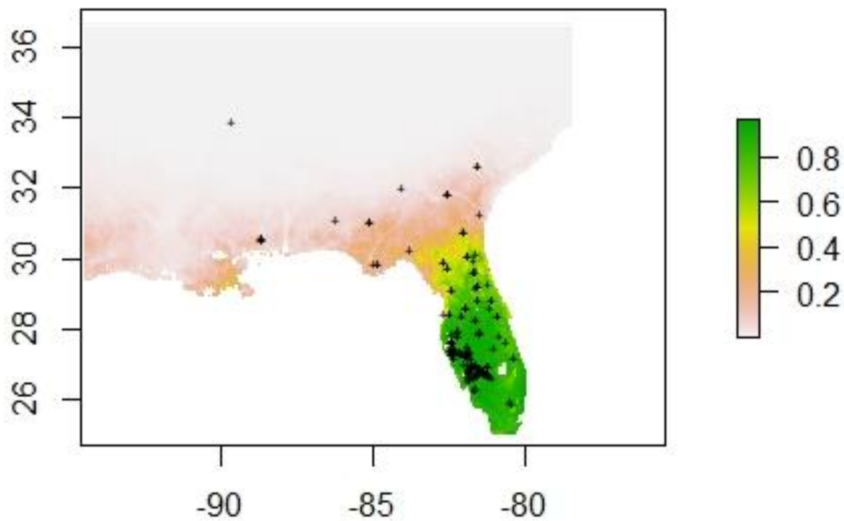

Figure 2: C. insignis SDM habitat suitability prediction. Presence is for C. insignis detection and absence is if there was no detection. Values are probabilities of species in that area, based on the fitted SDM model and the raster used for prediction. Points represent presence detections from the test dataset.

The *C. insignis* outcome variable was the strongest performing of each of the models, with the glm model reaching an AUC score 0.86. The predicted raster indicated stronger probabilities to the south and decreasing as the map moves North, with a stronger predicted presence in southern and central Florida. *C. stellifer* and *C. venustus* are included in figures 6 and 8 in the supplemental material.

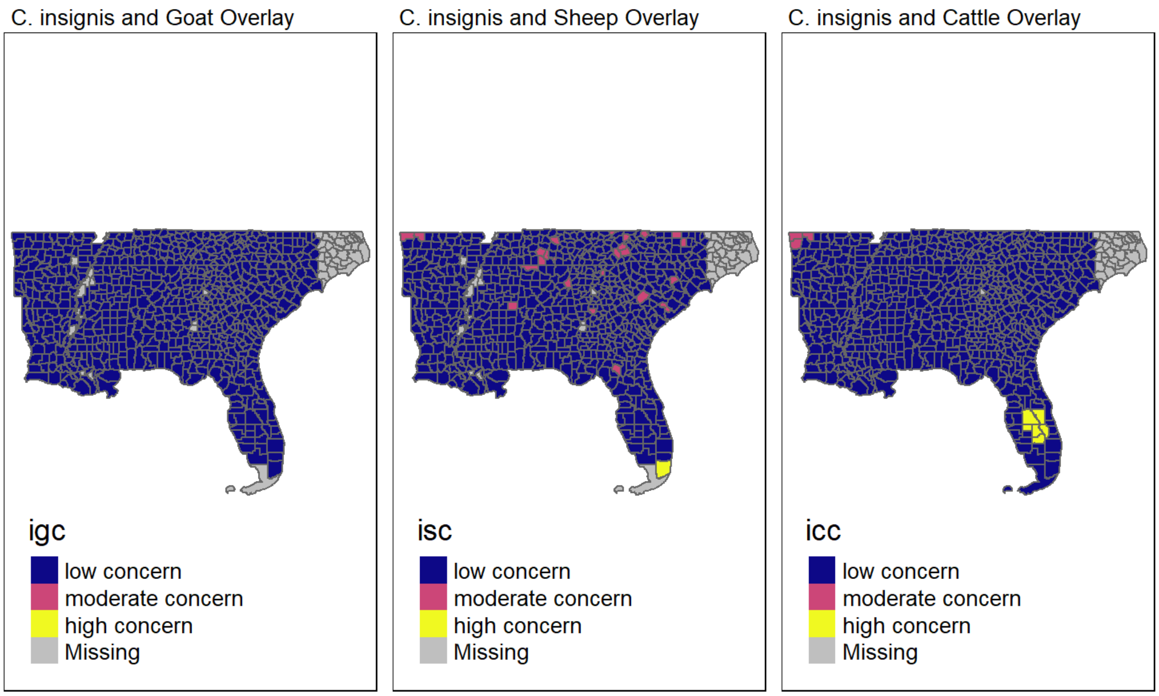

Figure 3: C. insignis overlay analysis with goat, sheep, and cattle. Using the predicted raster for C. insignis for probabilities in combination with cutoffs livestock data

The overlay analysis had more emphasis on livestock count than purely midge species probability. As livestock population increased, levels of concern increased, with areas that had higher *Culicoides* midge probability having higher levels of concern. For *C. insignis*, higher levels of concern were present for both sheep and cattle. All areas of high concern were in Florida. There were also areas of moderate concern for both sheep and cattle, with sheep moderate concern dispersed throughout the study area, and with cattle moderate concern concentrated in the northwest corner of the study area. Goats, with respect to *C. insignis*, only had low concern. Both *C. stellifer* and *C. venustus* overlays are included in figures 7 and 9 in the supplemental material.

Conclusion

Our study showed that *C. insignis, C. stellifer,* and *C. venustus* have many intricate variables associated with their environments that can influence their predicted ranges. We saw what was expected with respect to the trap site locations and species density. *C. insignis* was mostly contained to Florida, but the other two species were mostly north of Florida. Temperature continued to be one of the main variables for species prediction, especially in the multivariable models.

However, more data are required. *C. stellifer* was our largest species category but had one of the lower model performances, indicating that the model should be reevaluated, and predicator variables reassessed. Future studies should consider including relative humidity as a better way to predict environment than precipitation. Precipitation should also be reevaluated due to holding low significance for most of the species present. However, this could be due to how precipitation was used for modeling, as while we extracted point data for overall model building, we used median values over the study period for the rasters used for prediction. This could influence the model due to precipitation having irregular distributions and removing the model's ability to differentiate between values. Relative humidity may be more consistent and provide a better measure of how moisture influences *Culicoides* midge habitat.

Another issue of concern is the midge survey data. More samples should be collected over a wider area with more consistent trap placement and timing. The midge data we did have was also a biased dataset, with traps being placed where they could be and where midge presence was anticipated to aid in collection. This resulted in traps being placed in wild areas, state parks, and other nature areas, rather than areas with livestock nearby. Having traps with livestock nearby would be better to analyze the potential relationship between midge presence and risk of BTV. Furthermore, by not having as many traps in areas with livestock, we could be potentially missing important variables for our species' habitat, as livestock yards and pens have been shown to be potential living and breeding grounds for several *Culicoides spp.*

Another limitation is time as a covariate in this data. While we did have the date that these traps were placed and data was recorded, we did not include this in our model. Time was difficult to incorporate correctly into the habitat suitability model and the predicting raster. Furthermore, traps were placed in increments, with trapping beginning in Florida in 2008 and gradually expanding out as resources allowed for it. Also, some traps were discontinued in one area and relocated to another to allow for a broader sampling area. Both of these made incorporating time difficult and time should be considered for future studies.

Another consideration is that more traps should be placed overall to generate more data for better modeling. *C. stellifer* was one of our most abundant species, while *C. venustus* was one of our least abundant. Having more records, especially for lower abundance species, would help produce better models and better predictions. Traps should also be placed with trap variation in mind, as this dataset does not provide a contrast between important trap characteristics mentioned in the literature and the introduction of this paper. Being able to include trap characteristics as predictors in the model would enhance surveillance strategies for future use. Future studies should also include lag times and seasonal analysis for environmental and meteorological variables regarding midge presence. Seasonal changes and performance of previous seasons could potentially influence midge presence in the current season.

Our future direction will focus on improving our model to provide an in-depth overlay analysis. The goal is to provide actionable data that could be used for trap placement and *Culicoides* midge surveillance. On top of that, understanding areas that have higher probabilities of vector presence, along with the presence of high numbers of hosts, such as ruminant livestock, could be used to better inform prevention strategies to prevent an outbreak.

CHAPTER 3: FUTURE DIRECTIONS

This study provided information on environmental and meteorological variables and their influence on species distribution for *C. insignis, C. stellifer,* and *C. venustus*. However, additional information is required to prepare for future surveillance campaigns. A significant group of factors missing from this study is trap characteristics. While this study expands on the understanding of environmental variables, it lacks information on trapping strategies and characteristics. Traps were set at approximately uniform heights and had identical baits. This prevents us from making comparisons on trap efficacy and reveals a gap in the data that is required for creating a comprehensive sampling campaign. Therefore, future studies need to consider a variety of trap factors.

Trap characteristics serve important roles in determining the success of a study as well as the potential bias that the study may face. Factors include the type of bait used (light, chemical attractants, host), location, and height, with sub-factors for each of those categories.

There are multiple types of trap baits. Light is one of the most common traps for surveillance, especially for *Culicoides*[11]. The type of light can be set to appropriately target the population of interest, though research is ongoing in determining species preference[11–16]. Certain populations of *Culicoides* are more attracted to green wavelength while others prefer blue or ultraviolet (UV) light[12,14]. One study suggested that *Culicoides* could be divided into two groups: UV attracted and green-attracted[12,17]. Furthermore, there are some species, such as *Culicoides sonorensis*, that are more dependent upon wavelength rather than intensity[17]. This contrasts with one recommendation for biting flies, which focuses on high-intensity, short wave-length light-emitting diodes (LEDs)[17].

Some studies have indicated that disease status can influence trap attractiveness. *Culicoides sonorensis* studies have shown some preliminary data that BTV status changes how the midge species interacts with traps[15]. Typically attracted to UV light, a *Culicoides sonorensis* midge appears to develop an aversion to UV light once positive for BTV[15]. This finding still needs further research and has interesting implications for future surveillance studies, especially if disease prevalence in vector populations is the outcome of interest.

The success of a light trap can also be influenced by its background, such as foliage. Since the insects have poor vision and operate during twilight and nocturnal hours, increasing the contrast between the trap and its background can aid in successful trapping[11,16,17]. Depending on the species that is being targeted, certain colors can help to increase the contrast and increase the attractiveness of the trap, while other colors can reduce visibility and subsequently hurt collection[17].

Ambient light can also reduce the effectiveness of both contrast and traps,[11,17]. Even natural sources, such as a full moon, increase *C. sonorensis* activity but can reduce trap effectiveness due to the increased ambient light[11,17]. Considering ambient light and competing sources, light pollution is a growing problem. Areas that need increased surveillance but are in or close to an urban setting can have increased levels of ambient light and increased amounts of competing light sources that can make surveillance difficult[11,17]. Light pollution is increasing by approximately six percent per year, and as that trend continues, the effectiveness of light traps will continue to decrease[11].

Range is also a contributing factor to light trap effectiveness and is often influenced by the type of light source, intensity, wavelength, contrast, and ambient light. It is estimated that the Centers for Disease Control (CDC) miniature light trap has a range of approximately fifteen meters and the onderstepoort light trap has a range of thirty meters[11,18,19]. However, one study found that the estimated range for the onderstepoort could be as low as two to four meters for *Culicoides*[11].

Trap location plays an important role in trap effectiveness. Understanding the breeding grounds, the feeding habits, and the ideal habitat of the specific *Culicoides* species of interest will aid in the setup and success of traps[16]. Trap rates can be influenced by proximity to hosts, such as sheep or cattle[16].

Trap height is another factor with important consequences for successful and representative midge trapping. Before studies began to investigate the possible relationship between trap success and height, there was no rationale as to why traps were set at a chosen height. However, studies have found that height does play a role that varies by species of interest[16,20,21]. An example would be *Culicoides insignis*, which appears to favor tree canopies over the ground level[21]. While still attacking ground hosts to feed, traps set in the canopy had better trap rates than traps set on the ground[21]. Traps were set at 1.37 meters for ground level and six meters and 9 meters for 2016 and 2017 respectively[21]. However, there are still species that favor ground traps over canopy traps[20,21]. Another study using onderstepoort light traps found that a height of 2.8 meters was most effective when compared to lower heights[22].

This reinforces the idea that understanding trap characteristics and how they influence successful surveillance campaigns is important for proper analysis of both species' presence and potential disease risk. Future studies and SCWDS surveillance campaigns should have duplicate traps with differing characteristics, such as modulating height, to allow for comparison and analysis of the influence of trap characteristics on species surveillance.

Future studies should also consider having consistent trapping timelines and placement. The data used for this analysis had inconsistent trap time and placement due to limited resources. Traps would sometimes be moved to different locations due a limited number of traps, which influenced both the location variables of certain traps, as well as time due to some traps having differing time frames than others. This made the comparison between traps difficult, and future studies should aim to have consistent placement and timing for all traps.

Finally, future studies should try to have an even distribution of traps across areas of interest. Traps should be set in livestock areas to provide a better representation of midge probability in areas of concern. Traps in this study were set in areas that were available. This created a biased dataset that focused primarily on parks, natural areas, and wildlife areas. This may not be representative of livestock areas and could bias the overlay analysis.

REFERENCES:

1. Rivera NA, Varga C, Ruder MG, et al. Bluetongue and Epizootic Hemorrhagic Disease in the United States of America at the Wildlife-Livestock Interface. *Pathogens*. 2021;10(8). doi:10.3390/PATHOGENS10080915

2. Sick F, Beer M, Kampen H, Wernike K. Culicoides Biting Midges-Underestimated Vectors for Arboviruses of Public Health and Veterinary Importance. *Viruses*. 2019;11(4). doi:10.3390/V11040376

3. Borkent A. *The Biting Midges, the Ceratopogonidae (Diptera)*.; 2005.

4. Steinke S, Lühken R, Balczun C, Kiel E. Emergence of Culicoides obsoletus group species from farm-associated habitats in Germany. *Medical and Veterinary Entomology*. 2016;30(2):174-184. doi:10.1111/mve.12159

5. Erram D, Blosser EM, Burkett-Cadena N. Habitat associations of Culicoides species (Diptera: Ceratopogonidae) abundant on a commercial cervid farm in Florida, USA. *Parasites and Vectors*. 2019;12(1). doi:10.1186/s13071-019-3626-1

6. McGregor BL, Blackburn JK, Wisely SM, Burkett-Cadena ND. Culicoides (Diptera: Ceratopogonidae) communities differ between a game preserve and nearby natural areas in northern Florida. *Journal of Medical Entomology*. 2021;58(1):450-457. doi:10.1093/jme/tjaa152

7. Sloyer KE, Burkett-Cadena ND, Yang A, et al. Ecological niche modeling the potential geographic distribution of four culicoides species of veterinary significance in florida, USA. *PLoS ONE*. 2019;14(2). doi:10.1371/journal.pone.0206648

8. Zuliani A, Massolo A, Lysyk T, et al. Modelling the northward expansion of culicoides sonorensis (Diptera: Ceratopogonidae) under future climate scenarios. *PLoS ONE*. 2015;10(8). doi:10.1371/journal.pone.0130294

9. Mayo C, McDermott E, Kopanke J, et al. Ecological Dynamics Impacting Bluetongue Virus Transmission in North America. *Frontiers in Veterinary Science*. 2020;7. doi:10.3389/fvets.2020.00186

10. Brand SPC, Keeling MJ. The impact of temperature changes on vector-borne disease transmission: Culicoides midges and bluetongue virus. *J R Soc Interface*. 2017;14(128). doi:10.1098/rsif.2016.0481

11. McDermott EG, Mullens BA. The dark side of light traps. *Journal of Medical Entomology*. 2018;55(2):251-261. doi:10.1093/jme/tjx207

12. Bishop AL, Bellis GA, McKenzie HJ, et al. Light trapping of biting midges Culicoides spp. (Diptera: Ceratopogonidae) with green light-emitting diodes. *Australian Journal of Entomology*. 2006;45(3):202-205. doi:10.1111/j.1440-6055.2006.00538.x

13. Spohr L, Bishop AL, Worrall RJ, Spohr LJ, Mckenzie HJ, Barchia IM. *Improving Light Trap Efficiency for Culicoides Spp. with Light-Emitting Diodes Improving Light-Trap Efficiency for Culicoides Spp. with Light-Emitting Diodes*. Vol 40.; 2004. https://www.researchgate.net/publication/43343984

14. Sloyer KE, Wisely SM, Burkett-Cadena ND. Effects of ultraviolet LED versus incandescent bulb and carbon dioxide for sampling abundance and diversity of Culicoides in Florida. *Journal of Medical Entomology*. 2019;56(2):353-361. doi:10.1093/jme/tjy195

15. McDermott EG, Mayo CE, Gerry AC, Laudier D, MacLachlan NJ, Mullens BA. Bluetongue virus infection creates light averse Culicoides vectors and serious errors in transmission risk estimates. *Parasites and Vectors*. 2015;8(1). doi:10.1186/s13071-015-1062-4

16. McDermott EG, Lysyk TJ. Sampling considerations for adult and immature culicoides (Diptera: Ceratopogonidae). *Journal of Insect Science*. 2020;20(6). doi:10.1093/jisesa/ieaa025

17. Wilson R, Wakefield A, Roberts N, Jones G. Artificial light and biting flies: the parallel development of attractive light traps and unattractive domestic lights. *Parasites and Vectors*. 2021;14(1). doi:10.1186/s13071-020-04530-3

18. Kirkeby C, Graesbøll K, Stockmarr A, Christiansen LE, Bødker R. *The Range of Attraction for Light Traps Catching Culicoides Biting Midges (Diptera: Ceratopogonidae)*. Vol 6.; 2013. http://www.parasitesandvectors.com/content/6/1/67

19. Rigot T, Gilbert M. Quantifying the spatial dependence of Culicoides midge samples collected by Onderstepoort-type blacklight traps: An experimental approach to infer the range of attraction of light traps. *Medical and Veterinary Entomology*. 2012;26(2):152-161. doi:10.1111/j.1365-2915.2011.00988.x

20. Swanson DA, Adler PH. Vertical distribution of haematophagous Diptera in temperate forests of the southeastern U.S.A. *Medical and Veterinary Entomology*. 2010;24(2):182-188. doi:10.1111/j.1365-2915.2010.00862.x

21. McGregor BL, Runkel AE, Wisely SM, Burkett-Cadena ND. Vertical stratification of Culicoides biting midges at a Florida big game preserve. *Parasit Vectors*. 2018;11(1):505. doi:10.1186/s13071-018-3080-5

22. Probst C, Gethmann JM, Kampen H, Werner D, Conraths FJ. A comparison of four light traps for collecting Culicoides biting midges. *Parasitology Research*. 2015;114(12):4717-4724. doi:10.1007/s00436-015-4720-8

23. Cuéllar AC, Kjær LJ, Baum A, et al. Modelling the monthly abundance of Culicoides biting midges in nine European countries using Random Forests machine learning. *Parasites and Vectors*. 2020;13(1). doi:10.1186/s13071-020-04053-x

24. Naimi B, Araújo MB. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*. 2016;39(4):368-375. doi:10.1111/ECOG.01881

25. McGregor BL, Sloyer KE, Sayler KA, et al. Field data implicating Culicoides stellifer and Culicoides venustus (Diptera: Ceratopogonidae) as vectors of epizootic hemorrhagic disease virus. *Parasites and Vectors*. 2019;12(1). doi:10.1186/s13071-019-3514-8

26. Aybar CAV, Juri MJD, Grosso MSL de, Spinelli GR. Spatial and temporal distribution of culicoides insignis and culicoides paraensis in the subtropical mountain forest of Tucumn, Northwestern Argentina. *Florida Entomologist*. 2011;94(4):1018-1025. doi:10.1653/024.094.0440

27. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

28. Johnson M. GitHub - mikejohnson51/climateR: An R 📦 for getting point and gridded climate data by AOI. Accessed March 28, 2022. https://github.com/mikejohnson51/climateR

29. Kuhn M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*. 2008;28(5):1-26. doi:10.18637/JSS.V028.I05

30. Burdett CL, Kraus BR, Garza SJ, Miller RS, Bjork KE. Simulating the Distribution of Individual Livestock Farms and Their Populations in the United States: An Example Using Domestic Swine (Sus scrofa domesticus) Farms. *PLOS ONE*. 2015;10(11):e0140338. doi:10.1371/JOURNAL.PONE.0140338

APPENDIX

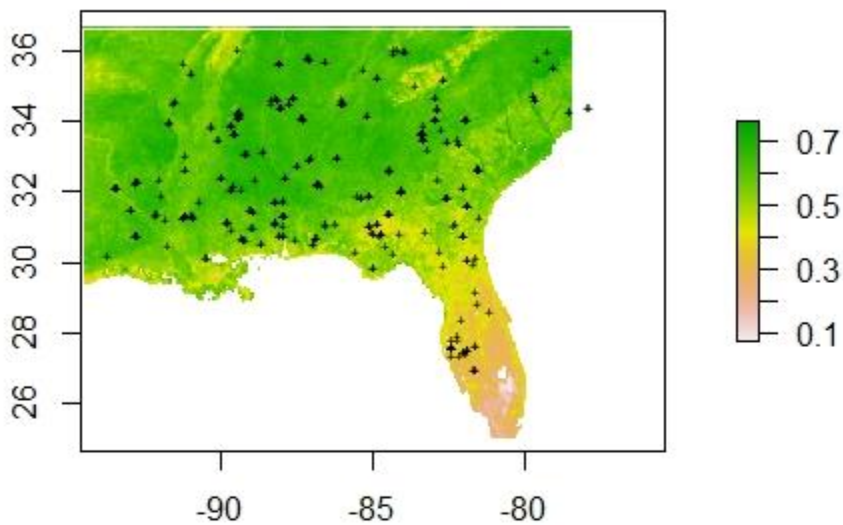

Figure 6: C. stellifer SDM habitat suitability prediction. Presence is for C. stellifer detection and absence is if there was no detection. Values are probabilities of species in that area, based on the fitted SDM model and the raster used for prediction. Points represent presence detections from the test dataset.

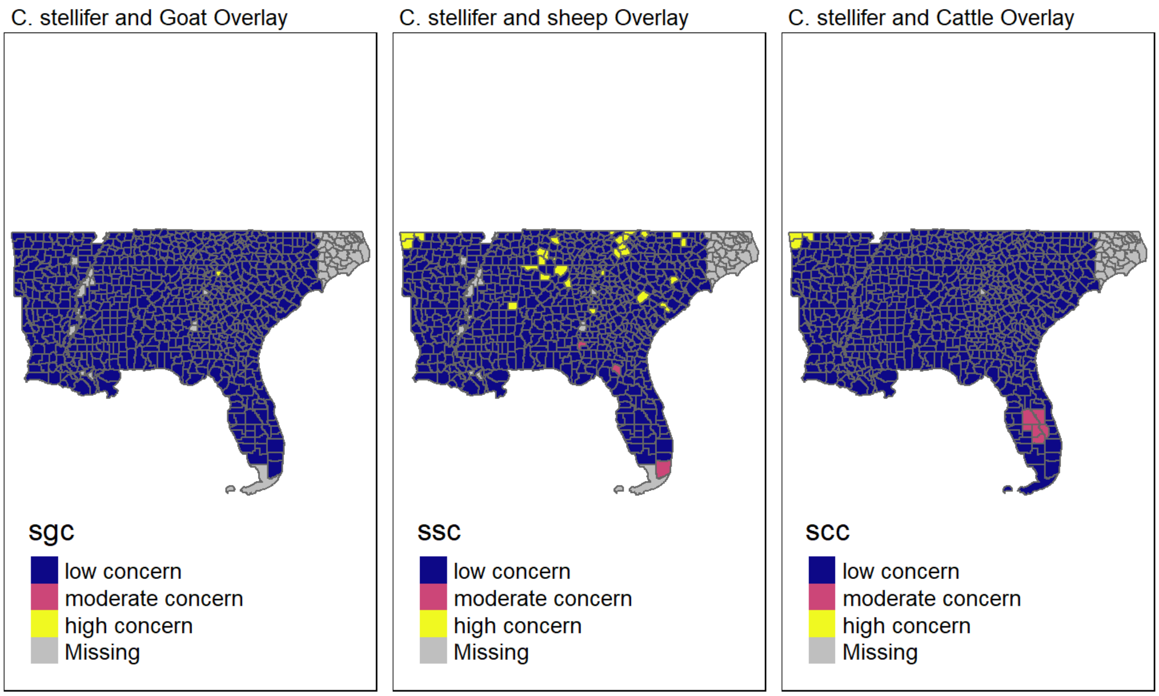

Figure 7: C. stellifer overlay analysis with goat, sheep, and cattle. Using the predicted raster for C. stellifer for probabilities in combination with cutoffs livestock data.

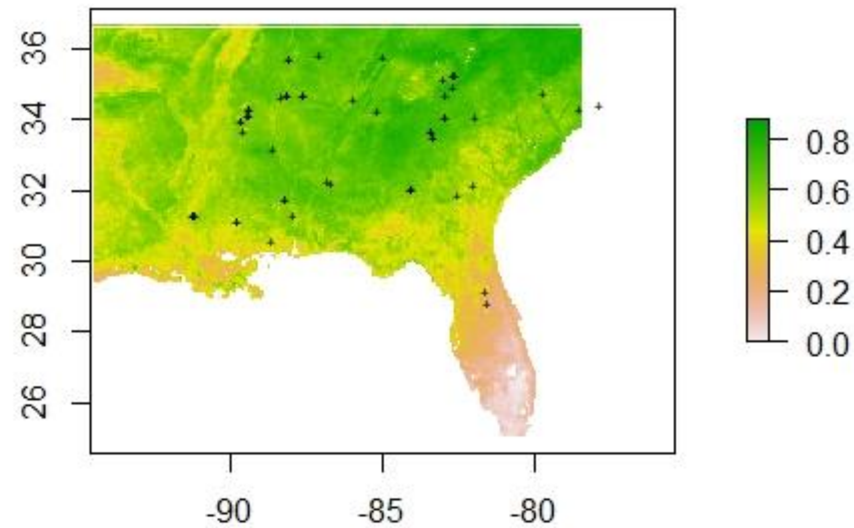

Figure 8: C. venustus SDM habitat suitability prediction. Presence is for C. venustus detection and absence is if there was no detection. Values are probabilities of species in that area, based on the fitted SDM model and the raster used for prediction. Points represent presence detections from the test dataset.

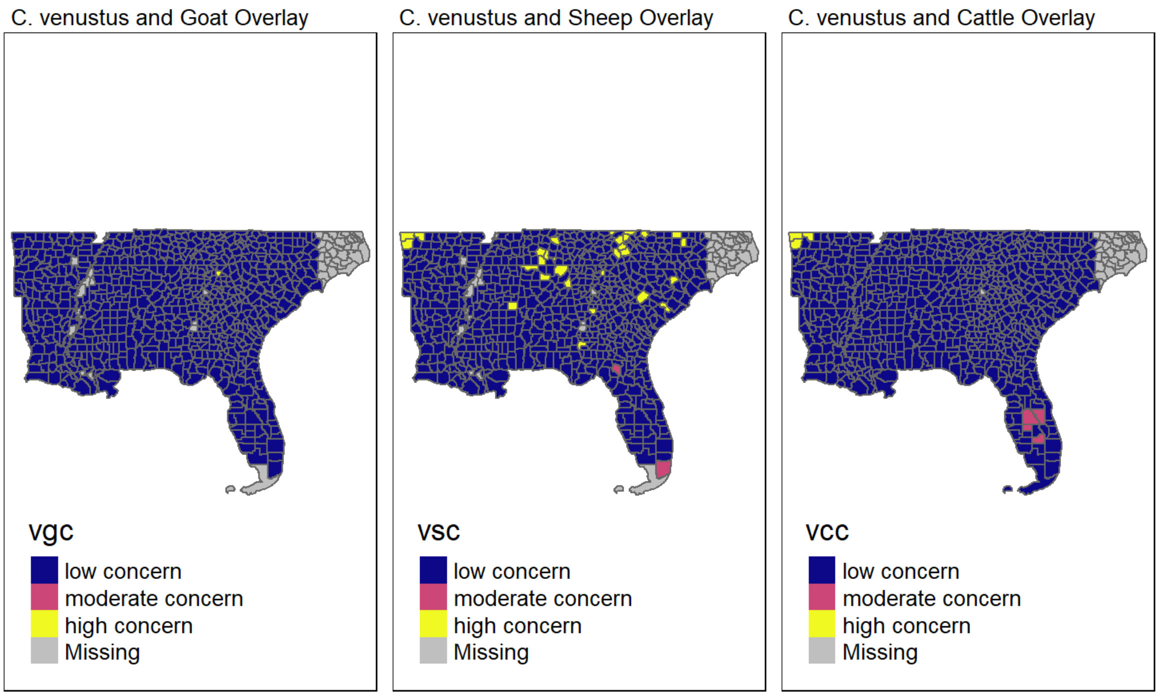

Figure 9: C. venustus overlay analysis with goat, sheep, and cattle. Using the predicted raster for C. venustus for probabilities in combination with cutoffs livestock data.