METACOMMUNITY ECOLOGY LINKS ENVIRONMENTAL DRIVERS TO *CULICOIDES*
SPP. COMMUNITIES AND HEMORRHAGIC DISEASE REPORTS IN THE
SOUTHEASTERN UNITED STATES.

by

CHRISTOPHER CLEVELAND

(Under the Direction of Daniel Mead and Andrew Park)

ABSTRACT

Hemorrhagic disease (HD), one of the most important diseases of white tailed deer
(WTD: *Odocoileus virginianus*) in North America, is vector-borne, and caused by closely related
orbiviruses, epizootic hemorrhagic disease viruses (EHDV) and bluetongue viruses (BTV)
(family: Reoviridae). In North America, it is generally accepted that a biting midge (*Culicoides
sonorensis*) is the vector for transmission to domestic livestock. However, the role of *C.
sonorensis* in virus transmission to WTD is not well defined and evidence suggests that other
*Culicoides* species may be involved. I identified discrete *Culicoides*-communities in the
southeastern US and environmental covariates potentially contributing to community
composition. I also evaluated relationships between environmentally-driven *Culicoides-
communities and HD reports in the southeastern US and found composition of *Culicoides-
communities predicted by temperature, precipitation, and elevation. *Culicoides* spp. not known
to vector HD were found to occur in areas of high disease reporting, whereas *C. sonorensis* was
rarely present.

INDEX WORDS: *Culicoides*, Hemorrhagic disease, metacommunity, vector diversity
METACOMMUNITY ECOLOGY LINKS ENVIRONMENTAL DRIVERS TO *CULICOIDES*
SPP. COMMUNITIES AND HEMORRHAGIC DISEASE REPORTS IN THE
SOUTHEASTERN UNITED STATES.

by

CHRISTOPHER CLEVELAND

B.S., University of North Carolina at Asheville, 2010

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2015
METACOMMUNITY ECOLOGY LINKS ENVIRONMENTAL DRIVERS TO *CULICOIDES* SPP. COMMUNITIES AND HEMORRHAGIC DISEASE REPORTS IN THE SOUTHEASTERN UNITED STATES.

by

CHRISTOPHER CLEVELAND

Co-Major Professor: Daniel Mead
Co-Major Professor: Andrew W. Park
Committee: Joseph L. Corn, David Stallknecht

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015
DEDICATION

This work would not have been possible without the support, insight, and love of my beautiful wife Ania. Always ready to help, to talk, and to tell me I’m probably wrong; I’m forever indebted to you. Maybe someday I’ll get to be your field tech again? To my mother and brother, who both have brought me various dead animals to dissect, to listen to me blather about things that probably make no sense, and to find the humor in the darkest situations: you have my eternal gratitude. To my daughter, Daphne Lydia Fox: you are my soul and fire, and although too young to know why I am up at one in the morning trying to write a dedication, figure out why my code isn’t running, or finish all that’s on the to do list, just know that everything I do in this life I do out of love for you.
ACKNOWLEDGEMENTS

The Southeastern Cooperative Wildlife Disease Study (SCWDS) has been the most incredible learning opportunity for my career, and no matter the future I will hold in highest esteem the crew that has lent me their time, their energy, and their willingness to teach and share (in spite of the fact that I can be thicker than mud). Dr. Joseph Corn, you are a most gracious, patient, and supportive mentor. Thank you for allowing me the opportunity to work at SCWDS and take on this project, and for the multiple walks in our outside office. Dr. Andrew Park, the countless hours spent in your lab working on code, trying to understand error messages, pestering you at all hours…and yet you always smile and seem happy to see me. Unbelievable. Your teachings far surpass those that you might be aware of, and I am so fortunate to have had the opportunity to work with you. Dr. David Stallknecht, you keep me on my toes, that’s for damn sure. Your questions give me heartburn, but I know that my work and my career are stronger given the lessons you’ve shared and how you’ve taught me to critique, analyze, and defend. Thanks for taking the time. Dr. Daniel Mead, you are a balanced voice of reason for so many questions I’ve juggled over the years. Your willingness to guide me and offer direction at SCWDS above and beyond this project, to teach me and lend a hand at the drop of a hat is a gift I will forever cherish. Tad, you really helped put the nail in the coffin, and made me envious of your super hero like coding skills. To the SCWDS techs, staff, and all others at bldg. 1082, thanks for putting up with me, showing me how to work the fax machine, and for keeping that coffee black, strong, and nasty enough to strip the meat off your bones.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2  LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>CAUSATIVE AGENTS</td>
<td>4</td>
</tr>
<tr>
<td>HEMORRHAGIC DISEASE IN WHITE TAILED DEER</td>
<td>5</td>
</tr>
<tr>
<td>HEMORRHAGIC DISEASE REPORTING</td>
<td>6</td>
</tr>
<tr>
<td>CULICOIDES BIOLOGY</td>
<td>6</td>
</tr>
<tr>
<td>CULICOIDES AS VECTORS</td>
<td>7</td>
</tr>
<tr>
<td>CULICOIDES AS VECTORS OF HD TO LIVESTOCK</td>
<td>7</td>
</tr>
<tr>
<td>CULICOIDES AS VECTORS OF HD TO WTD</td>
<td>8</td>
</tr>
<tr>
<td>METACOMMUNITY ECOLOGY</td>
<td>9</td>
</tr>
<tr>
<td>ENVIRONMENTAL AND SPATIAL COVARIATES</td>
<td>10</td>
</tr>
<tr>
<td>BOOSTED REGRESSION TREES</td>
<td>11</td>
</tr>
<tr>
<td>3  METACOMMUNITY ECOLOGY LINKS ENVIRONMENTAL DRIVERS TO</td>
<td></td>
</tr>
<tr>
<td>CULICOIDES SPP. COMMUNITIES AND HEMORRHAGIC DISEASE REPORTS</td>
<td></td>
</tr>
<tr>
<td>IN THE SOUTHEASTERN UNITED STATES</td>
<td>12</td>
</tr>
<tr>
<td>4  CONCLUSIONS AND FUTURE DIRECTIONS</td>
<td>33</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>35</td>
</tr>
</tbody>
</table>
APPENDICES

A SUPPLEMENTARY MATERIAL .................................................................44
LIST OF FIGURES

Figure 3.1: The geographic locations of sites sampled for Culicoides from 2007-2012 .............29

Figure 3.2: Culicoides species distributions among the wildlife management areas sampled .....30

Figure 3.3: Relative contribution of spatial and environmental variables to the species occurrence of Culicoides species communities ....................................................................................31

Figure 3.4: Synthesis of the relationship between environmental drivers, Culicoides community composition and level of disease reporting .................................................................32

Figure A.1: Barber’s modularity ................................................................................................47
CHAPTER 1
INTRODUCTION

Hemorrhagic disease (HD) is a vector-borne disease caused by two groups of closely related orbiviruses, epizootic hemorrhagic disease viruses (EHDV) and bluetongue viruses (BTV) in the family Reoviridae. HD is considered to be one of the most important diseases of white-tailed deer (Odocoileus virginianus), has been diagnosed in other wild ruminants as well as domestic cattle (Dulac et al. 1988, St. George et al. 1983) and occurs with greatest frequency in the southeastern US (Nettles and Stallknecht 1992). EHDV was first described in 1955, when EHDV serotype 1 was isolated from a white-tailed deer (WTD) during a high mortality event (Shope et al. 1955, 1960). In 1966, BTV-10 was diagnosed in a WTD in Texas (Stair et al. 1968). Subsequently, three EHDV serotypes (-1,-2,-6) and six BTV serotypes (-1, -2, -10, -11, -13, -17) have been reported to cause HD in WTD in the United States (Pearson et al. 1992, Stallknecht 2008).

Clinical signs of HD are highly variable, and many infected animals appear normal or only show mild signs of disease. In WTD, 3 patterns of disease have been described: peracute, acute, and chronic (Prestwood et al. 1974). Pulmonary congestion, edema, and hemorrhaging are typically observed in peracute cases; hemorrhages of the rumen, buccal papillae and base of the pulmonary artery, hydropericardium, severe oral lesions, pulmonary edema, and recumbency are associated with acute cases of HD (Thomas 1981, Howerth et al. 2001, Couvillion et al. 1981). In its chronic form, HD may result in interrupted hoof growth (or hoof sloughing), scarring of the
rumen, healing lesions, and abrasions on carpal joints and sternum from periods of recumbency (Thomas 1981, Howarth et al. 2001).

Historically, HD in the United States has been reported from the southeastern corner of the country (Florida) to the northwest (Washington) (Nettles et al. 1992). In general, as latitude increases, the frequency of HD reports decreases but there are increasing reports of mortality. In contrast, the disease is rarely reported from the extreme southern range, most likely associated with areas of enzootic stability (high rate of infection with minimal or no clinical disease) (Stallknecht et al. 1996, Willhelm 1966, Gaydos et al. 2002).

As with all vector-borne pathogens, the occurrence of natural cycles is largely dependent on the distribution of the vectors, changes in the vectors’ distribution or populations, or introduction of pathogens into an area where competent vectors exist (Roy 2008). In North America, Culicoides sonorensis (previously C. variipennis sonorensis) is recognized to be the primary vector of BTV and EHDV to sheep and cattle (Tabachnick 1996, Greiner et al. 1985). However, no strong evidence suggests that C. sonorensis is the vector, or sole vector, of EHDV and BTV to white-tailed deer (Smith et al. 1996). In fact, records from field collections of Culicoides communities using penned or captive WTD as bait revealed that C. sonorensis was absent or underrepresented from Culicoides collections. This suggests that within communities of Culicoides, species other than or in addition to C. sonorensis are likely involved in transmission of EHDV and BTV to WTD (Smith et al. 1996, Mullen et al. 1985, Kramer et al. 1985, Gerhardt 1986). Recently, HD has been documented outside of the historical range as well as beyond the historical range of C. sonorensis. For example, EHDV-6 was first detected in the US in 2006, and has since been isolated in multiple mid-western states. Furthermore, between 2007 and 2012 there has been an increase in the number of HD reports from more northeastern
states (NJ, PA, IN, MI). These states previously reported HD on rare occasions or not at all (Stallknecht 2012). Since environmental impacts, primarily land cover and climate, have a strong influence on arthropod ecology (Patz et al. 2000, Altizer et al. 2013), the resulting range expansion could be climate driven.

There is clearly a lack in the understanding of whether discrete *Culicoides* communities present in the southeastern US are affected by climate, and whether certain species within these communities are related to occurrences of HD. Therefore, the objectives of this research are to investigate communities of *Culicoides* in the southeastern US, identify environmental variables that may predict those communities, and then determine the extent to which environmentally driven *Culicoides* communities are related to HD reports in WTD.
Hemorrhagic Disease (HD) is a vector-borne viral disease affecting white-tailed deer in North America and occurs with greatest frequency in the southeastern United States (Nettles and Stallknecht 1992). HD is caused by genetically related orbiviruses epizootic hemorrhagic disease viruses (EHDV) and bluetongue viruses (BTV) in the family Reoviridae (Howerth et al. 2001). EHDV and BTV were identified as causative agents in 1955 and 1966, respectively (Shope et al. 1955, 1960, Stair et al. 1968). In North America, *Culicoides sonorensis* (previously *C. variipennis sonorensis*) is recognized to be the primary vector of HD to domestic livestock (Tabachnick 1996, Greiner et al. 1985). However, no strong evidence suggests that *C. sonorensis* is the vector, or sole vector, of HD to white-tailed deer (Smith et al. 1996).

**Causative Agents:**

The epizootic hemorrhagic disease viruses (EHDV) and bluetongue viruses (BTV) are *Orbiviruses* in the Reoviridae family. The Orbivirus genus contain 20 species, 12 tentative species, and greater than 169 serotypes (Mertens 2004, Mertens et al. 2005). In the United States, three EHDV serotypes (-1,-2,-6) and six BTV serotypes (-1, -2, -10, -11, -13, -17) have been reported (Pearson et al. 1992, Stallknecht 2008). The Orbivirus virion is composed of an outer and inner shell surrounding the genome (Roy 2007) with the outer shell consisting of two major proteins, VP2 and VP5 (Verwoerd et al. 1972, Roy 1996). The structural protein VP2 is responsible for virion bonding with mammalian cells, resulting in receptor-mediated endocytosis, and is used for determining virus serotype (Mertens 2004).
**Hemorrhagic Disease in White-tailed deer:**

Clinical responses of WTD to infection with BTV or EHDV can vary from inapparent to death. The viruses replicate in vascular endothelial cells and have been documented in hemopoietic stem cells (Thomas 1981). When disease manifests, it is typically characterized by high temperature, respiratory distress, edema located in the head and neck, mucosal congestion, lethargy and inappetence, and cyanotic mucous membranes (Thomas 1981; Nettles et al. 1992).

HD can be classified as peracute, acute, or chronic with the respective clinical findings (Prestwood et al. 1974): pulmonary congestion, edema, and hemorrhaging are observed in peracute cases; hemorrhages of the rumen, buccal papillae and base of the pulmonary artery, hydropericardium, severe oral lesions, pulmonary edema, and recumbency are associated with acute cases of HD (Thomas 1981; Howerth et al. 2001; Couvillion et al. 1981). Chronically infected animals may display hoof sloughing in response to interruption of hoof growth caused by necrosis due to capillary damage (Howerth et al. 2001). Stomatitis, rumenitis, bacterial pneumonia, healing lesions, and abrasions on carpal joints and sternum from periods of recumbency may also be observed (Thomas 1981). Emaciation and death as a result of chronic infections are often seen in the winter and spring following initial infection (Howerth et al. 2001).

The WTD range extends from southern Canada throughout most of the United States and south into the neotropics (Gallina 2008). Within the United States, HD has been documented in the southeast, midwest, northwest, and pacific coast (Nettles et al., 1992), yet reports of disease are rare in the southeast. Deer mortality events similar to HD epizootics have been reported as far back as 1890 and 1901 (Shope et al. 1959; Trainer 1964; Gaydos et al. 2002; Nettles and Stallknecht 1992).
In extreme southern latitudes of the United States (southern Florida and Texas) HD is enzootically stable (high levels of infection with little to no clinical disease) (Stallknecht et al. 1996; Willhelm 1966; Gaydos et al. 2002). Increasing in latitude, epizootics occur more frequently, and are typically associated with mild or inapparent disease. HD epizootic cases in more northern latitudes are less frequent, and are associated with severe clinical disease and large-scale mortality (Stallknecht et al. 1996).

**Hemorrhagic Disease Reporting:**

The Southeastern Cooperative Wildlife Disease Study has compiled nationwide county level HD reports of morbidity and mortality in white-tailed deer since 1980. Confirmation of HD depends on four criteria: 1) Sudden, unexplained accounts of deer mortality in high numbers during late summer and early fall; 2) necropsy diagnosis of hemorrhagic disease conducted by a trained wildlife biologist or diagnostician at a State diagnostic laboratory or veterinary college; 3) virus isolation from an infected deer; 4) hunter killed deer that presented with clinical symptomology i.e. sloughing of hooves, oral ulcers, or scars on the rumen. Criteria 1-3 indicate a report for deer mortality with the associated county, whereas criterion 4 represents morbidity (Stallknecht and Howerth 2004). As a result, a broad data-set has been compiled containing information on HD viruses, serotype, morbidity or mortality, and location.

**Culicoides Biology**

*Culicoides* biting midges (Diptera: Ceratopogonidae) are among the most abundant hematophagous insects in the world. Over 1400 species have been described and of these, 151 species are known to occur in North America (Borkent and Grogan 2009, Vigil et al. 2014). Adult midges are small (1-3mm), crepuscular, holometabolous insects (Mellor et al. 2000, 2009). Females oviposit in a variety of sites, including but not limited to tree-holes, wetlands, pools and
streams, irrigation pipe leaks, saturated soil and animal dung (Mellor et al. 2000). Development of eggs laid in suitable substrate occurs between two to seven days (Meiswinkel et al. 2004; Mellor et al. 2000). The vermiform larvae develop through four larval stages, and certain species may overwinter as fourth instar larvae in diapause (Mellor et al. 2000).

**Culicoides as Vectors**

Several pathogens with implications for both domestic and wildlife species are transmitted by biting midges. The bluetongue viruses (BTV) and epizootic hemorrhagic disease viruses (EHDV) are transmitted to domestic livestock and potentially to WTD by *C. sonorensis* (Griener 1985). This species is also a suspected vector of vesicular stomatitis New Jersey virus (Drolet et al. 2005). African horse sickness virus (AHSV) is transmitted by *C. imicola* (Sanders et al. 2010; Mellor et al. 2000). In the Amazon region of South America, Panama, and the Caribbean, Oropouche virus is transmitted to humans by *C. paraensis* (Yanase et al. 2005). *C. arboricola* transmits protozoal parasites (*Haemoproteus*) to birds (Levine 1961, Borkent and Grogan 2009) and *C. variipennis* transmits filarial nematodes (*Onchocerca cervicalis*) to equines (Collins and Jones 1978).

**Culicoides as Vectors of HD to livestock**

Bluetongue, caused by the bluetongue viruses, is an economically important disease worldwide and affects a wide variety of domestic ruminant species (Mellor et al. 2000). Severe disease usually only occurs in certain breeds of sheep (McLaughlin et al. 2003), whereas infection in cattle and goats is most often inapparent. Cattle are considered to be suitable reservoir hosts for BTV due to extended periods (16-20 weeks) of viremia (MacLachlan et al. 1994). Even though BTV infections in cattle may not always result in disease (Hourrigan and Klingsporn 1975), there can still be significant economic impact due to the restriction of animal
movement or sale. For example, it is estimated that the US cattle industry has annual losses nearing $125 million due to BTV (Tabachnick 1996).

BTV serotypes -1, -2, -10, -11, -13, -17 have been detected in the US (Mullen et al. 1999, Stallknecht 2008). Within the United States, *Culicoides sonorensis* has been implicated as a BTV vector, using field caught specimens to infect sheep (McKercher et al. 1953, Price 1954). Additional studies verified *C. sonorensis* as a biological vector by allowing *C. sonorensis* midges to feed on infected sheep, undergo an incubation period of 10-15 days, and then take a blood meal on non-infected sheep which subsequently became infected (Foster et al. 1963). *C. insignis*, a vector of BTV serotypes -1, -3, -4, -6, -8, -12, -14, and -17 within Central America and the Caribbean, has also been shown to transmit BTV-2 in southern Florida (Mecham 2003, Tabachnick 1996).

The epizootic hemorrhagic disease viruses are also known to infect and cause disease in domestic cattle. Ten serotypes have been identified (St. George et al. 1983, Gorman 1992). The Ibaraki strain (EHDV-2) transmitted by *C. oxystoma* in Japan (Omori et al. 1969) and EHDV-7 transmitted by *C. imicola* in Israel (Yadin et al. 2008; Savini et al. 2011) are both documented cases involving outbreaks of HD resulting in varying levels of morbidity in domestic cattle with serious economic impact (Savini et al. 2011). Within the US, EHDV serotypes -1,-2,-6 have been reported (Stallknecht 2008), and the primary vector is *C. sonorensis* (Greiner et al. 1985).

**Culicoides as Vectors of HD to WTD**

While it is generally accepted that within the United States, *C. sonorensis* is the vector of the BTVs and EHDVs to livestock, their role in transmitting these viruses to WTD is not well understood. In field studies conducted using WTD as bait, and in studies where HD has been detected in WTD, *C. sonorensis* was largely absent in collections. Over 950,000 host seeking
Culicoides were collected using light traps and a WTD fawn as bait in areas where the viruses are known to be enzootic. *C. sonorensis* represented < 0.01% of the total catch.

Previous studies have reiterated this low presence of *C. sonorensis* (Mullen et al. 1985; Gerhart 1986). Given the low presence of *C. sonorensis* in field studies that utilize WTD, WTD facilities, or WTD as bait, as well as sampling performed in EHDV/BTV enzootic sites, *C. sonorensis* is now thought to be an unlikely, or not the sole vector of EHDV (Smith and Stallknecht 1996; Smith et al. 1996a). A list of suspected vectors for EHDV and BTV based off of midge abundance, host-seeking activity, feeding behavior, and seasonality found from studies (Smith 1996, 1996b; Mullen 1985) includes the following midges, *C. debilipalpus, C. stellifer, C. obsoletus, C. paraensis, C. biguttatus, C. niger,* and *C. spinosus*.

**Metacommunity Ecology**

Metacommunity ecology is a theoretical framework for studying responses of communities to ecological change. For this study, metacommunity techniques were used to analyze community composition, integrating spatial and environmental processes to characterize species gains to and losses from discrete communities (Holyoak 2005). The elements of metacommunity structure (EMS) are derived from a site-by-species matrix, which is a fundamental way of organizing data in community ecology (Leibold and Mikkelson 2002, Presley et al. 2010).

EMS analyses use three statistics (coherence, turnover, and boundary clumping) to classify metacommunities into a number of potential patterns or types. This is achieved by ordering sites and species based on ordination scores obtained via reciprocal averaging (Gauch 1982), which orders sites and species with similar communities and distributions, respectively, along the interaction matrix diagonal. The coherence, turnover, and boundary clumping statistics
of communities are compared to a null model, evaluating statistically significant community structures (Leibold and Mikkelson 2002). The fixed-row fixed-column null model provides a metric of modularity for comparison to metacommunity structure, and quantifies the extent to which the communities cluster into discrete groups (Barber 2007).

Communities can form six hypothetical patterns of species’ distribution among sites (checkerboards, nested subsets, Clementsian, Gleasonian, evenly-spaced gradients, and random) that typically result from at least one important biological or ecological factor (Presley 2010, Leibold and Mikkelson 2002), necessitating further evaluation of the abiotic factors involved with the vector community structure. In order to investigate the relationship between community composition, the community sites are ordered according to the metacommunity interaction matrix diagonal. This ordering creates a numerical gradient used for relating community site scores to environmental and spatial gradients.

**Environmental and Spatial Covariates:**

WorldClim is an interpolated set of global climate layers, compiling data from weather stations at a resolution of 1 square kilometer. In addition to standard measurements of abiotic factors (i.e. precipitation, temperature), WorldClim bioclimatic variables (BIOCLIM) combine standard abiotic measurements (monthly temperature and rainfall values) with biologically important variables associated with niche modeling (Hijmans et al. 2005). These variables provide a perspective on annual trends, seasonality, and extreme abiotic factors that may be overlooked when not viewed on a larger temporal scale (i.e. month vs. year).

While the WorldClim dataset has the advantages of reporting a large set of environmental covariates at fine (~1km) resolution, it is a temporal aggregate from 1950-2000. The PRISM data set (PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 4 Feb
2004) is a more recent compilation of three basic climate elements (precipitation, temperature, and dew point). Spatially resolved to ~800m, this data-set offers a strong basis for comparison with the WorldClim data, yet providing a more robust interpolation for more recent years (Daly et al. 2008). PRISM does not, however, contain a similar comparison to the WorldClim ‘BIOCLIM’ dataset. By confirming congruence between the two datasets in overlapping years in the southeastern US (Daly et al. 2008), the WorldClim data set is used, confident that it is a reliable measure of environmental drivers of the study period (2007-2012).

The measures of site position in the southeastern US are derived from latitude and longitude to investigate the spatial proximity of sites to one another. A way to account for spatial “relatedness” of one site to another is not to simply calculate distance based directly on latitude and longitude, but to translate the coordinates to measures of spatial distance between sites using principal coordinates analysis, PCoA (Gower 2005). This is similar to a principal components analysis on a distance matrix, in which distances represent Haversine distances between pairwise sites. Because space can be considered as either a confounding variable or a predictor in statistical modeling, developing an appropriate method for evaluation is essential, and can be done by comparing and weighting the spatial proximity or distance of one site to another, testing spatial autocorrelation among locations (Drey et al. 2006, Gower 1966).

**Boosted Regression Trees:**

Boosted regression is a statistical tool able to decipher multicollinearity among variables and evaluate nonlinear relationships between predictors and response (Elith et al. 2008). Boosted regression trees rely on decision trees, weighted by their predictive capability, in order to extract infrequent ("weak") decision factors for making a branch in the tree. The remaining frequent ("strong") decision rules combine to form a single predictive model (Elith et al. 2008).
CHAPTER 3

METACOMMUNITY ECOLOGY LINKS ENVIRONMENTAL DRIVERS TO *CULICOIDES* SPP. COMMUNITIES AND HEMORRHAGIC DISEASE REPORTS IN THE SOUTHEASTERN UNITED STATES

---

1 Cleveland, C., T.A. Dallas, S. Vigil, J.L. Corn, and A.W. Park
To be submitted to Ecology and Evolution
Abstract

Hemorrhagic disease (HD), one of the most important diseases of white tailed deer (WTD: Odocoileus virginianus) in North America, is vector-borne, and caused by closely related orbiviruses, epizootic hemorrhagic disease viruses (EHDV) and bluetongue viruses (BTV) (family: Reoviridae). In North America, it is generally accepted that a biting midge (Culicoides sonorensis) is the vector for transmission to domestic livestock. However, the role of C. sonorensis in virus transmission to WTD is not well defined and evidence suggests that other Culicoides species may be involved. Culicoides communities in the southeastern US were investigated, identifying discrete Culicoides-communities and environmental covariates potentially contributing to community composition. Relationships between environmentally driven Culicoides-communities and HD reports in the southeastern US were evaluated and the composition of Culicoides-communities are predicted by temperature, precipitation, and elevation. Culicoides spp. not known to vector HD were found to occur in areas of high disease reporting, whereas C. sonorensis was rarely present.
Introduction

It is estimated that around 30% of recent emerging infectious diseases are vector-borne (Jones et al. 2008). Due to the environmental sensitivity of arthropods, diseases vectored by arthropods (arboviruses) may be transmitted by more than one species (Gubler 2002). Growing evidence indicates that environmental impacts, primarily climate change and land cover, have a strong influence on arthropod distributions and phenologies (Patz et al. 2000, Daszak et al. 2001, Altizer et al. 2013). In spite of the obvious utility, predicting geographic hot-spots for disease is complicated by the fact that many arboviruses are transmitted by a suite of related arthropods. For practical purposes, studies of infectious diseases transmitted by such groups are often restricted to a single species (Lord 2010, Higa 2011, Allan et al. 2009). More detailed knowledge of environmental drivers of particular communities of arthropod species (including non-vector species that interact with vectors) may provide valuable links between underlying geographical variation and risk of infectious disease (Ostfeld et al. 2005, Blaustein et al. 2013).

Metacommunity ecology represents a promising method of evaluating arthropod communities and relationships to infectious diseases, integrating spatial and environmental processes to characterize species gains to and losses from discrete communities (Cottenie 2005). Here, such an approach is used to study biting midge species in the genus Culicoides (Diptera: Ceratopogonidae), which are important vectors worldwide, capable of transmitting micro- and macro-parasites in human and animal populations and causing severe economic impacts to the domestic livestock industry (Jennings and Mellor 1989, Mellor et al. 2000a, Agbolade et al. 2006, Azevedo 2007, Rasmussen et al. 2012). Within the United States, Culicoides transmit closely related orbiviruses, causing Hemorrhagic disease (HD) in livestock and white-tailed deer (Odocoileus virginianus) populations. HD is considered one of the most important diseases of
WTD (Howarth et al. 2001) and provides an ideal system to simultaneously evaluate *Culicoides* community structure, environmental drivers, and the potential relationship of species within *Culicoides* communities to HD reports.

Using light trap data (from approximately 5500 trap nights; number of traps times number of nights) *Culicoides* communities at over 200 discrete sites were organized using a site-by-species presence/absence matrix. This matrix is then used when calculating metacommunity metrics: coherence (the number of embedded absences in the matrix, which can distinguish random, checkerboard and other patterns), turnover (which can distinguish patterns of nestedness, gradual turnover and block replacement of groups of species by other groups) and boundary clumping (the way in which species boundaries are distributed along an ordered dimension) (Leibold and Mikkelson 2002, Dallas 2014). Communities that emerged from the metacommunity analysis were evaluated via nineteen abiotic variables extracted using the WorldClim database (Hijmans et al. 2005) in addition to elevation, and distance proxies of latitude and longitude.

Finally, *Culicoides* community composition and corresponding site-specific abiotic variables were investigated for potential relationships to the number of hemorrhagic disease reports recorded for the corresponding area. This study provides a link between abiotic predictors and *Culicoides* community composition, with an interesting perspective on the absence of the known HD vector, *C. sonorensis*. Additionally, species of *Culicoides* not known to vector HD are present in areas of high HD reporting. Therefore, it may be relevant to shift from the examination of single vector species to a community of *Culicoides*, especially when the pathogen may be vectored by multiple species, as is the case with *Culicoides* species (Smith and Stallknecht 1996).
Methods

Culicoides Surveillance

Since 2007, the Southeastern Cooperative Wildlife Disease Study (SCWDS) has conducted surveillance for novel Culicoides spp. in the Atlantic and Gulf coastal regions of the southeast (Louisiana, Mississippi, Alabama, Georgia, and Florida) (Fig. 1). These surveys, typically occurring during peak HD activity (late July through September) and coinciding with HD reports, provide a presence/absence dataset of 50 Culicoides species in the U.S. and Canada. Surveys were conducted on public lands (wildlife management areas, state/county parks and forests) and multiple private sites using CDC light traps configured with UV lights and collection cylinders filled with 70% ethanol. Traps were hung 1.5m-2m from the ground in Culicoides oviposition and emergence sites (wetlands, streams, lakes, ditches, and forest cover). Eight to twelve traps at various locations were set each evening before dusk and collected the following morning within 1-2 hours of sunrise. Contents of each trap were stored in 70% ethanol for laboratory processing which entails sorting and removing Culicoides from trap contents and grouping by subgenus/species group. Species were identified, and voucher specimens were removed, dissected, and slide mounted. Previously identified individuals from collections of the Florida State Collection of Arthropods, Gainesville, Florida (FSCA) were used to verify identification and adhere to taxonomic indices per the Nearctic Culicoides Catalog (Borkent and Grogan 2009). For a thorough description of Culicoides sampling techniques and taxonomic identification referenced in this paper, see Vigil et al. 2014.

Hemorrhagic Disease Reports

HD is considered one of the most important diseases of WTD (Howerth et al. 2001) and occurs with greatest frequency in the southeastern United States (Nettles and Stallknecht 1992),
with areas of enzootic stability (high levels of infection with little to no clinical disease) suspected in the most southern latitudes of the US (Stallknecht et al. 1996; Willhelm 1966; Gaydos et al. 2002). HD is caused by epizootic hemorrhagic disease viruses and bluetongue viruses of the genus \textit{Orbivirus}, family \textit{Reoviridae} (Howerth et al. 2001). To evaluate the temporal and spatial extent of HD, the Southeastern Cooperative Wildlife Disease Study (SCWDS) has compiled nationwide county level reports of morbidity and mortality in white-tailed deer since 1980. For the purposes of this study, data was restricted to 2007-2012.

The case standards for reporting HD in WTD depend on four criteria: 1) Sudden, unexplained accounts of deer mortality in high numbers during late summer and early fall; 2) necropsy diagnosis of hemorrhagic disease conducted by a trained wildlife biologist or diagnostician at a State diagnostic laboratory or veterinary college; 3) virus isolation from an infected deer; 4) hunter killed deer that presented with clinical symptomology i.e. sloughing of hooves, oral ulcers, or scars on the rumen. Criteria 1-3 indicate a report for deer mortality with the associated county, whereas criterion 4 represents morbidity. For our purposes, a disease report is either morbidity or mortality.

The level of disease reporting at sites was determined in the following way: the containing county of each \textit{Culicoides} survey site was spatially aggregated with its four nearest county neighbors. This was to account for movement of WTD from areas of infection to areas of sampling, a factor that could yield inaccurate disease reporting. The number of disease reports was then established by counting the five aggregated counties with five years of HD reporting for a maximum of 25 possible HD reporting opportunities.
Metacommunity analysis

*Culicoides* metacommunity structure was analyzed in R (R Development Core Team 2012) using the package *metacom* (Dallas 2014). Because the *Culicoides* data are structured as presence-absence of species per specific site, a site by species matrix for my analyses. The site-by-species matrix is a fundamental way of organizing data in community ecology. The elements of metacommunity structure (EMS), (Leibold and Mikkelson 2002, Presley et al. 2010) analysis uses three statistics (coherence, turnover, and boundary clumping) to classify metacommunities into a number of potential patterns or types. Metacommunity metrics calculated on the empirical matrix were compared to 1000 null matrices, generated using the "swap" method (Gotelli and Entsminger 2003), keeping column and row totals fixed.

The analysis of coherence, turnover, and boundary clumping was achieved by sorting sites and species based on ordination scores obtained via reciprocal averaging (Gauch 1982), which orders sites and species with similar communities and distributions, respectively, along the interaction matrix diagonal. Coherence and turnover statistics were calculated against a collection of null model simulations that permute the interaction matrix while maintaining row and column totals (Leibold and Mikkelson 2002). The final statistic, boundary clumping, was established using Morisita's index, with significance determined relative to a chi-squared distribution. Apart from a simple ordering of sites and species, both sites and species were related to environmental and spatial covariates. Particular attention is paid to the score of sites (hereafter “*Culicoides* community score”).

Environmental Covariates

To examine the relationship between *Culicoides* communities and abiotic variables, a set of 22 covariates (Fig. 3) available via WorldClim (Hijmans et al. 2005) were analyzed using a
boosted regression tree analysis via gbm (Ridgeway 2013). The abiotic data were obtained from WorldClim at a spatial resolution of 30 seconds (~1km), and interpolated to estimate average monthly climate data from weather stations. Bioclimatic variables (BIOCLIM), such as precipitation and temperature seasonality and isothermality, were combined with standard abiotic measurements (monthly temperature and rainfall values) to develop biologically important variables typically associated with niche modeling (Hijmans et al. 2005). These variables provide a perspective on annual trends that may be overlooked when not viewed on a larger temporal scale (i.e. month vs. year). Elevation was also sourced via WorldClim for each sampling site.

While the WorldClim dataset has the advantage of reporting a large set of environmental covariates at fine (~1km) resolution, it is a temporal aggregate from 1950-2000. As many of these covariates are related to temperature and precipitation, the correlation between the limited overlap of the WorldClim dataset and the 2009 Prism data set (PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 4 Feb 2004) (the approximate temporal mid-point of our sampling period) was tested. Testing the correlation between WorldClim and Prism provided a comparison between precipitation, mean temperature, and annual temperature range from the two datasets, which were all found to be tightly correlated (Pearson’s r= 0.76, 0.99, 0.97, respectively). Additionally, it has been previously noted that the Prism and WorldClim data sets are in general good agreement, especially in the southeastern US (Daly et al. 2008). Having established the congruence between the datasets, the WorldClim data was used as it reports many more environmental covariates per site.

Latitude and longitude were used to investigate the spatial proximity of sites to one another. Latitude and longitude coordinates were translated to measures of spatial distance between sites using principal coordinates analysis, PCoA (Gower 2005). The PCoA is similar to
a principal components analysis on a distance matrix, in which distances represent Haversine
distances between pairwise sites, and was obtained using R package *ape* (Paradis et al. 2004).
The first two PCoA vectors were used in the boosted regression tree analysis, as they constituted
97.8% of the corrected eigenvalues obtained from the decomposition of the distance matrix.

**Linking *Culicoides* Community Scores and Covariates via Boosted Regression**

*Culicoides* community scores were related to covariates using boosted regression (Elith et
al. 2008). Boosted regression trees rely on decision trees, weighted by their predictive capability,
in order to extract infrequent (“weak”) decision factors for making a branch in the tree. The
remaining frequent (“strong”) decision rules are combined to form a single predictive model. The
optimal number of trees was determined using five-fold cross validation (i.e. 80% training data,
20% testing data) (Elith et al. 2008). In addition, model accuracy was also assessed by a cross-
validation method, training the model on 80% of the data, and using this model to predict
*Culicoides* community score for the remaining 20% of the data. A linear model on predicted
versus actual *Culicoides* community scores was used to provide a measure of model accuracy.

The final model provides relative contribution (RC) values for each predictor variable,
calculated based on the number of decision trees in which a variable was included, weighted by
the improvement in model performance as a result of the inclusion of the predictor. All RC
values are bound between 0 and 100, where low scores correspond to variables with little
contribution to *Culicoides* community composition, and high scores indicate highly predictive
variables. Lastly, *Culicoides* community composition was related to disease reporting using
simple linear models.
Results

*Culicoides* metacommunity structure

*Culicoides* communities were significantly coherent ($z=2.93$, $p=0.003$), had higher species turnover ($z=-3.95$, $p<0.0001$), and more clumped species range boundaries (Morisita's index = 3.89, $p < 0.0001$) than expected under the fixed-row fixed-column null model. This is indicative of a *Clementsian* structure, in which species replace each other in discrete communities (Leibold and Mikkelson 2002). This suggests that the site-by-species matrix should be significantly modular, in which site and species combinations cluster, resulting in dense connections between subsets of sites and species, but sparse connections between modules (Barber 2007). The network was significantly modular (see Supplementary Materials), supporting the main analysis, and suggesting that species and sites cluster, such that discrete *Culicoides* communities replace each other along some environmental or spatial gradient.

Environmental and spatial covariates of metacommunity structure

The gradient along which *Culicoides* communities form discrete communities was related to environmental and spatial variables via Boosted Regression Trees (BRT), using the *Culicoides* community scores as a measure of the latent gradient along which the *Culicoides* community was structured. Precipitation in the wettest month (RC = 18.87), annual temperature range (RC = 15.95), elevation (RC = 14.27), and annual mean temperature (RC = 10.47) were all indicators of *Culicoides* community composition (Fig. 3.3, Fig. 3.4). The remaining variables had RC values below 10, and were not considered to be influential factors. The boosted regression tree model, trained on a random subset of the data (80% training set), predicted site scores of the remaining 20% with high accuracy (linear model of predicted versus actual scores; adjusted $R^2 = 0.838$).
**Culicoides community scores and Hemorrhagic disease reporting**

The number of HD reports for counties in this study were between 0-8 instances of HD occurrence. The *Culicoides* community score was a predictor of the level of disease reporting (linear model; adjusted $R^2 = 0.258$) (Fig. 3.2, Fig. 3.4). High levels of disease reporting were associated with *Culicoides* communities that occur at relatively low precipitation in the wettest month, at relatively low temperatures (but with large temperature swings between summer and winter) and at higher elevation compared to total sites sampled, although absolute elevation is moderate across study sites (~ 0-250m above sea level).

**Discussion**

The SCWDS *Culicoides* surveillance efforts, focusing on the southeastern US, provided presence data on 50 of the 151 North American species of *Culicoides* (Vigil et al. 2014). Using the *Culicoides* surveillance data in the metacommunity analysis, *Culicoides* communities in the southeastern US were found to be discrete (coherence ($z=2.93$, $p=0.003$, species turnover ($z=-3.95$, $p<0.0001$), boundary clumping (Morisita's index = 3.89, $p < 0.0001$)). These results indicate that *Culicoides* communities are composed of species within specific areas that do not readily disperse to other locations. The three statistics used to calculate *Culicoides* community scores (coherence, boundary clumping, and turnover) also indicate that certain species of *Culicoides* cluster at sampled sites, rather than all species being widespread across the study region.

Figure 3.1 shows community scores of sites sampled, their relationship to other sites, and ultimately that *Culicoides* communities in the southern range of study area (Florida) are composed of different species than sites in more northern latitudes (AL, MS). Figure 3.2 illustrates this point on a species level, revealing that *C. insignis*, *edeni*, and *furens* occur more
readily in sites in FL, whereas C. stellifer, haematopotus, arboricola, debilipalpis, hinmani, and paraensis are present in more northern sites. The discrete structure of Culicoides communities in the southeastern US may be associated with habitat type, however, environmental impacts, primarily climate change and land cover, have a strong influence on arthropod distributions (Patz et al. 2000, Daszak et al. 2001, Altizer et al. 2013). It is therefore important to evaluate what environmental and spatial covariates may predict Culicoides community composition.

Culicoides community composition in the southeastern US was predicted via boosted regression trees, Culicoides community scores from metacommunity analysis, and 22 environmental and spatial covariates. The environmental gradients responsible for Culicoides community composition are precipitation of the wettest month, annual temperature range, elevation, and annual mean temperature (Fig. 3.3, Fig.3.4).

The southeastern United States contains many physiographic regions (Miller and Robinson 1994), including salt marsh, coastal plain, and piedmont. In spite of physiographic regions, the spatial coordinates of sites did not influence Culicoides community structure. The lack of spatial significance indicates that the discrete communities of Culicoides emerge more in response to abiotic variables and elevation than to spatial configuration (Fig. 3.1). Average temperature and temperature range will likely influence the emergence times and overwintering periods of Culicoides. Drought conditions play an important role in generating ecological niches of Culicoides, as receding water bodies create wet mud habitat thereby influencing ovipositing opportunities and population growth (Wilkening et al. 1985).

The results generated from linear models relating Culicoides communities to HD reports show an interesting trend in non-vector species emerging in areas with higher levels of HD reports (Fig. 3.2), at northern study sites with similar community scores (Fig. 3.1). Culicoides
*debilipalpis, guttipennis, haematopotus, hinmani, paraensis,* and *stellifer* are not known to vector HD, but are generalist blood feeders (Blanton 1979). These species have been collected from previous studies using penned WTD, the wildlife host for HD, as bait (Smith et al. 1996b, Mullen 1985). Due to their host seeking behavior and ability to take a blood meal, it is possible that these species are in some way involved in transmission of the HD viruses. Previous studies have shown that *C. debilipalpis* is a competent vector for BTV (Mullen 1985), but further research is needed on the vectorial competency of other hematophagous *Culicoides*.

The relationship between environmental drivers, *Culicoides* community structure and level of disease reporting (Fig. 3.4) illustrates that higher numbers of disease reports and certain *Culicoides* communities co-occur in areas where precipitation is lowest, with the greatest annual temperature range at higher elevations and lower mean temperature. This indicates that environmental variables are important within our study region, but likely to change if the geographic range is expanded to include more northern latitudes.

The *Culicoides* sampling sites and HD reporting counties are located in regions where enzootic stability of HD (high levels of infection with little to no clinical disease) has been demonstrated (Smith et al. 1996b). Contributes bias to the amount of disease reports generated, indicating that further *Culicoides* sampling at northern latitudes where HD results in epizootics is needed. Sampling at sites where HD is not enzootically stable would enhance the study by including reports of HD cases that result in clinical disease. HD reports also have inherent bias: the case criteria for reporting HD depends on state and county officials responsible for the management of wildlife populations. It is possible that certain locations and individuals may report more than others, thereby skewing the actual number of disease reports per county.
*Culicoides* surveillance is a presence absence based sampling effort occurring over short periods of time at sampling sites. As such, *Culicoides* sampling could result in inaccurate community sampling. Temporally, surveillance efforts coincide with emergence of *Culicoides* species; however, there is the potential to “miss” an accurate representation of the *Culicoides* communities in a given area during a narrow sampling window. This study included repeated sampling at sites over the study period (2007-2012), which decreased this bias. The southeastern US is a narrow geographic area of *Culicoides* spp. range. Expanding the geographic sampling area would potentially show different *Culicoides* communities that may include novel HD vectors, or increased presence of the previously identified vector.

*Culicoides sonorensis*, the known vector for HD in the United States, appears in this analysis rarely (Fig. 3.2). By comparison to other blood feeding *Culicoides, C. sonorensis* does not appear related to the level of HD reports referenced in this study. *C. insignis*, the documented vector for BTV in the Caribbean and Central and South America, occurs in this study frequently in areas of low disease reporting (Fig. 3.2). This is perhaps due to sampling locations in enzootically stable areas in southern Florida. Surveillance for *C. insignis* in more northern latitudes is needed, especially given recent documentation of this species occurring outside of its historical range (Vigil et al. 2014).

A list of suspected HD vectors based off of midge abundance, host-seeking activity, feeding behavior, and seasonality includes the following midges: *C. debilipalpis, C. paraensis,* and *C. stellifer* (Smith 1996, 1996b; Mullen 1985). This study has found that the generalist blood feeding midges, *C. debilipalpis, guttipennis, haematopotus, hinmani, paraensis,* and *stellifer* appear in sites related to higher levels of HD reporting. It is possible that the generalist blood feeding *Culicoides* species are involved in some facet of transmission of HD (i.e. mechanical
transmission), although further work with each species is needed to determine whether this is plausible. Additionally, investigation into the potential mechanical transmission of HD by other species of biting flies could help explain the spread of HD beyond the known vector range. The impact of climate change and the subsequent effects on arthropod communities warrants continued *Culicoides* surveillance for hematophagous species not only in the southeast where *C. insignis*, the vector for BTV-2 occurs, but into more northern latitudes where there are increasing reports of HD in the absence of *C. sonorensis*. This will provide a more accurate representation of the dynamic relationship between *Culicoides* communities, species that may be involved in HD transmission within those communities, and HD reports.

**Acknowledgements**

The *Culicoides* surveys were funded through a Cooperative Agreement with the United States Department of Agriculture, Animal and Plant Health Inspection Service Veterinary Services. AWP acknowledges funding from the McDonnell Foundation (grant number 220020193). We thank S. Budischak for comments on a draft of the manuscript. Additionally, we thank DE Stallknecht for sharing hemorrhagic disease reports and SCWDS technicians for support in *Culicoides* field collections. Funding for this project was provided through Cooperative Agreements 0691130808CA, 0791130808CA, 0891130808CA, 0991130808CA, 1091130808CA, 1191130808CA, and 1291130808CA, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture. Additional funds were provided through sponsorship from the fish and wildlife agencies of Alabama, Arkansas, Florida, Georgia, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Puerto Rico, South Carolina, Tennessee, Virginia, and West Virginia; through the Federal Aid to Wildlife
Restoration Act (50 Stat. 917) and Grant Agreement 06ERAG0005, Biological Resources Division, U.S. Geological Survey, U.S. Department of the Interior; and through Cooperative Agreements 0696130032CA, 0796130032CA, 0896130032CA, 0996130032CA, 1096130032CA, 1196130032CA, and 1296130032CA, Veterinary Services, Animal and Plant Health Inspection Service, U.S. Department of Agriculture.
Figure Legends

Figure 3.1: The geographic locations of sites sampled for *Culicoides* from 2007-2012 (ten light traps per site per night), with grey scale shading indicating *Culicoides* community score and circle diameter (small to large) corresponding to disease incidence reporting (0-8).

Figure 3.2: *Culicoides* distributions (columns) among the wildlife management areas (rows) sampled. A black rectangle indicates presence of that particular *Culicoides* species at that site in the matrix. Gray scale colors other than black indicate level of disease reporting (as indicated on gray-scale bar legend).

Figure 3.3: The relative contribution of spatial and environmental variables to the species occurrence of *Culicoides* communities. Bars in black contributed the most, having relative contribution values greater than 10. The boosted regression model predicted *Culicoides* community score on a subset of test data with relatively high predictive accuracy (inset: $R^2_{adjusted}=0.839$, $p < 0.0001$), validating the importance of environmental covariates on *Culicoides* community composition.

Figure 3.4: Synthesis of the relationship between environmental drivers, *Culicoides* community structure and level of disease reporting. In all panels, sampling sites are ordered along the x-axis according to vector community score, i.e. adjacent sites share similar community composition. Panel A shows the level of disease reporting at each site, panel B shows the extent of each *Culicoides* species range across sites, and panels C-F show how the most influential variables change across site. Lines through the data points are loess trend lines.
Figures

Figure 3.1
Figure 3.2
Figure 3.3

The figure shows a bar chart and a scatter plot with a regression line. The bar chart compares the relative contribution of various environmental variables, such as precipitation during different quarters, temperature seasonality, and elevation, to predicted community scores. The scatter plot plots predicted community scores against actual community scores, with a linear regression line indicating a high adjusted $R^2$ of 0.84 (p < 0.0001). Variables like mean temperature wettest quarter, annual temperature range, elevation, and annual mean temperature have significant contributions as indicated by the longer bars in the bar chart. The scatter plot also visually supports the strong correlation between predicted and actual community scores.
Figure 3.4

A: Number of disease reports

B: Species ranges

C: Precipitation (mm) in wettest month

D: Annual temperature range (°C)

E: Elevation (m) above sea level

F: Annual mean temperature (°C)
CHAPTER 4
CONCLUSIONS AND FUTURE DIRECTIONS

The goal of this study was to establish a link between environmentally driven *Culicoides* communities and Hemorrhagic disease reports in the southeastern United States. I investigated *Culicoides* communities’ composition in the southeastern US utilizing a metacommunity analysis that describes how groups of *Culicoides* species change over a widespread study region. After finding that community structure is not random, but contains discrete groups over large geographical ranges, it was necessary to determine what factors influenced community composition. I then determined relationships between environmental covariates and *Culicoides* communities to find what abiotic variables were significant in the species occurrence of *Culicoides* vector communities.

Upon finding environmentally-driven *Culicoides* communities, I used linear models to evaluate the relationships between these communities and the levels of disease reporting (0-8) to identify specific environmentally-controlled community types directly associated with high disease reporting. My research identifies several areas where future work with *Culicoides* and HD is needed. By establishing a relationship between HD reports and communities, I demonstrate that lines of inquiry further determining plausible HD vectors is needed. In the United States, accounts of HD transmission have been credited to the vector *Culicoides sonorensis* (Greiner 1985a, b, Mellor et al. 2000, Savini et al. 2011), typically surrounding livestock operations. *Culicoides* communities sampled in the southeast during this study focused on habitats associated with WTD, resulting in rare reports for the vector *C. sonorensis*. This
implies that there may be other potential hematophagous *Culicoides* capable of transmitting HD viruses. This coincides with previous studies conducted in regions where HD is enzootically stable (Smith and Stallknecht 1996, Mullens et al 1985).

Furthermore, as case reports of HD increase throughout the US, it is necessary to increase *Culicoides* surveillance in areas of HD to compliment the extensive sampling in the southeastern US. Doing so will expand the list of suspected vectors and their associated habitat types, providing valuable insight into how environmental gradients combine to change arthropod community structure and the capacity for pathogen transmission, allowing for better prediction, management, and prevention of diseases (Blaustein et al. 2013). Employing metacommunity analyses has shown that *Culicoides* communities in the southeastern US are composed of different species and predicted by specific environmental variables. The importance of identifying discrete communities via the metacommunity analysis and investigating *Culicoides spp.* community relationships to HD reports exhibits which species are occurring at areas with higher amounts of disease reporting. *Culicoides* spp. not known to vector HD were found to occur in areas of high disease reporting, whereas *C. sonorensis* was rarely present.
LITERATURE CITED


APPENDIX A
SUPPLEMENTARY MATERIAL

Modularity determination

The Elements of Metacommunity Structure analysis (EMS) classified the *Culicoides* vector metacommunity as being *Clementsian*. This suggests that species and sites within the interaction are forming discrete communities, such that some species are only found at a subset of sites (Figure S1). This mirrors a research theme in the understanding of complex system related to the detection of communities in networks (Newman and Girvan 2004, Fortunato 2010). While our analysis refers to community formation as *Clementsian*, the concept is also referred to as modularity. In order to provide additional support for our findings, we tested our empirical network for modularity, using Barber's $Q$ as our modularity measure (Barber 2007).

$$Q = \sum_{m=1}^{N} \left[ \frac{L_m}{L} - \left( \frac{K^A_m K^B_m}{L^2} \right) \right]$$

Equation 1: The form of Barber's modularity measure ($Q$) (taken from Thebault 2012). $K_m$ values are summed links for nodes of type $A$ and $B$, $L$ is the total number of links, $L_m$ is the number of links between a specific module and other nodes outside of the module.

To determine if the *Culicoides* metacommunity is significantly more modular than expected by chance, we generated 1000 null matrices maintaining the number of species per site, and the number of occurrences per species (i.e., a fixed-fixed null model was used), and compared the null distribution of $Q$ values to the empirical $Q$. The empirical interaction matrix was significantly more modular than expected under our null model simulations ($Q = 0.325, p <$
0.00001), providing support for our finding that *Culicoides* species replace each other across a gradient of sites. Perhaps interestingly, the ordering of *Culicoides* species based on modularity differed slightly from the ordination of species based on reciprocal averaging. However, they differed very slightly (only 4 species differed between the two methods), and this does not negatively affect the finding of community structure, especially as the reciprocal averaging ordination strives to fill the matrix diagonal, while ordering by modularity is an attempt to empty it. Further, it provides strong support for the presence of a structuring gradient in terms of site-level characteristics, as the ordination scores for sites matched ordering by modularity exactly.
Figure Legend

Figure A1: The two modules (in blue and red), as determined by the calculation of Barber's modularity ($Q$). Here, the matrix is ordinated to best view the modules (i.e., ordered by modularity score).