

TEMPORAL ANALYSIS AND SPATIAL MODELING OF THE DISTRIBUTION AND  
ABUNDANCE OF *CS. MELANURA*, EASTERN EQUINE ENCEPHALITIS VECTOR:

CONNECTICUT, 1997-2012

By

Chelsi White

---

A Thesis Submitted to the Faculty of the

DEPARTMENT OF EPIDEMIOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2016

## STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Chelsi White

## APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

---

Heidi Brown  
Assistant Professor  
Epidemiology & Biostatistics

---

2/17/2016  
Date

## ACKNOWLEDGEMENTS

I owe thanks to many people who directly or indirectly made this work possible. The data, on which all this work has been based was made possible by Theodore Andreadis. Moral support was afforded me during my long journey by my friends, in particular Dr. Jayleen Gun, Dr. Kathrine Center, Intan Yokelson, Jesse Jefferson, and the Al-Salehs'. I also wish to acknowledge the encouragement given to me by Dr. Clayton Loehn and Dr. George Gehrels during the initial stages of my graduate research. Many thanks go out to Dr. Gary Christopherson and Dr. Phil Geurtin who provided me with assistance and support even when time was not on their side. I also wish to acknowledge my committee of Dr. Heidi Brown, Dr. Kacey Ernst, and Dr. Gregg Garfin for all of their time and contribution to the work.

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b> .....	<b>5</b>
<b>LIST OF TABLES</b> .....	<b>5</b>
<b>Abstract</b> .....	<b>6</b>
<b>Introduction &amp; Background</b> .....	<b>7</b>
<i>Transmission Cycle</i> .....	8
<i>Ecological Niche</i> .....	9
<i>Seasonality</i> .....	10
<i>Connecticut</i> .....	12
<b>Objectives</b> .....	<b>13</b>
<b>Methods</b> .....	<b>13</b>
<i>Vector Data</i> .....	14
<i>Environmental Variable Sources</i> .....	19
Analysis I Variables.....	19
Analysis II Variables.....	19
Analysis I & II Variables.....	21
<i>Analysis I: Temporal</i> .....	21
<i>Analysis II: Spatial</i> .....	23
Spatial Variable Extraction.....	23
Spatial Regression Model Development.....	23
Spatial Regression Model Validation.....	26
<b>Results</b> .....	<b>27</b>
<i>Analysis I: Temporal</i> .....	27
<i>Analysis II: Spatial Regression Modeling</i> .....	32
<b>Discussion</b> .....	<b>35</b>
<i>Strengths &amp; Limitations</i> .....	38
<i>Recommendations</i> .....	40
<i>Conclusion</i> .....	42
<b>References</b> .....	<b>44</b>

## LIST OF FIGURES

1	Conceptual model for <i>Cs. melanura</i> abundance related to meteorological and groundwater factors.....	11
2	Distribution of 83 Centers for Disease Control and Prevention (CDC) light 1 traps with data from 1997-2012 used in temporal and spatial analyses. ....	15
3A-D	Comparison of average <i>Cs. melanura</i> per trap night per year and a: percent of mosquito pools positive per trapnight b: average groundwater level and winter groundwater levels differenced from average (ft) c: the previous year precipitation totals and current year spring precipitation total (mm) d: average winter temperature (°C).....	28
4	Probability of high <i>Cs. melanura</i> abundance in Connecticut.....	34

## LIST OF TABLES

1	Average <i>Cs. melanura</i> per trap night and percent of pools tested positive for EEEV by year.....	17
2	Data sources used and variables extracted in the study.....	18
3	Analysis I Pearson's partial correlation coefficients of explanatory temporal variables for <i>Cs. melanura</i> abundances.....	30
4	Analysis II logistic regression models.....	33

**Title:** Temporal Analysis and Spatial Modeling of the Distribution and Abundance of *Cs. melanura*, Eastern Equine Encephalitis vector: Connecticut, 1997-2012

**Abstract**

Eastern Equine Encephalitis virus is a vector-borne virus amplified by the *Culiseta melanura* mosquito in an enzootic avian cycle, causing high morbidity and mortality to horses and humans when contracted as incidental hosts. The virus is distributed across most of the eastern United States, Canada, and Gulf coast, and has been expanding in geographic range and season of activity over time. Spatial-temporal trends in *Cs. melanura* abundance were correlated with available meteorological (temperature and precipitation) and remotely sensed environmental data for the period of 1997-2012 in Connecticut. The effects of inter-annual changes in precipitation, temperature, and groundwater levels on *Cs. melanura* abundances using time-series linear regression and cross-correlation analyses were inconclusive. Habitat modeling using logistic regression and landscape-based predictive variables demonstrated strong efficiency (46.2%) and acceptable sensitivity and specificity (65.6 and 78.6%, respectively) using NDVI difference and distance from palustrine areas as predictive factors. Remotely sensed data can improve the understanding of vector abundance patterns, helping to forecast future outbreaks and regional expansions by guiding surveillance efforts.

## **Introduction & Background**

Eastern Equine Encephalitis virus (EEEV) is a highly pathogenic, mosquito-borne virus that circulates among birds but is known to cause serious disease in horses and humans (Reimann et al., 2008). A member of the genus *Alphavirus*, EEEV is found along the Atlantic region of the eastern United States and Canada and the Gulf Coast states; it has several alternate viral lineages in Central and South America (Centers for Disease Control and Prevention (CDC) 2015; Center for Food Security and Public Health & Institute for International Cooperation in Animal Biologics (CFSPH-IICAB), 2008-2015). Eastern Equine Encephalitis was first diagnosed in humans in 1933 in the U.S., but was seen in horses in as early as the late 1800's (Armstrong and Andreadis 2013; Nadalur and Urban 2014). Eastern Equine Encephalitis virus is relatively rare compared to other mosquito-borne diseases in the U.S. For example, Arizona accumulated 133 cases of West Nile Virus in 2012 (CDC-USGS 2012). In comparison, over the past 50 years an average of 5 to 10 EEE cases are reported each year in the U.S. (CFSPH-IICAP 2008-2015). Incidence appears to be rising. From 1970 to 1989, reported human cases ranged from 0 to 15, with an average of approximately 5 cases a year (Letson, Bailey, Person and Tsai 1993), and, between 2004 and 2013, 3 to 21 human cases were reported per year, averaging 8.5 cases (CDC 2015). Although EEE is relatively rare in the U.S., its high case-fatality rates (35-75%) place it as the most pathogenic mosquito-borne pathogen in the North American region (Deresiewicz et al., 1997; Reimann et al., 2008). More than half of those that survive EEE sustain severe neurological impairment, with long-term care costing ~\$3 million per person (Villari et al., 1995).

Activity for EEEV is seasonal, ranging from late spring to early fall. In the southeast, particularly Florida, EEEV is now active year-round due to optimal climatic factors for vector activity (Burkett-Cadena et al. 2015). From 2004-2013, 15 cases were reported in Florida

(CFSPH-IICAP 2008-2015; CDC 2015). Other southeastern states with cases during this period were Louisiana (n=3), Alabama (n=4), Arkansas (n=1), Georgia (n=4), North Carolina (n=7), and South Carolina (n=2) (CDC 2015). For the northeastern U.S., cases from 2004-2013 were reported for Massachusetts (n=24), Michigan (n=3), New Hampshire (n=9), New York (n=3), Rhode Island (n=1), Vermont (n=2), and Connecticut (n=1) (CDC 2015). Though EEE is rare, because of its persistent presence in the eastern United States and the high costs associated with mortality and after-care, EEE remains a health concern warranting further investigation (USDA 2013; CDC 2015).

### *Transmission Cycle*

The EEE virus is maintained and amplified in an enzootic cycle between birds—typically passerine—and *Culiseta melanura* mosquitos, where the virus is transferred by the bite of an infected adult female mosquito (USDA 2004; Nasci and Edman 1981). The fundamental primary bridge vectors in epizootic transmission, affecting dead-end horse and human host, is less clearly understood, as it varies among geographic location and time of year. Implicated mosquito vectors include *Aedes vexans*, *Coquillettidia perturbans*, and *Aedes canadensis* in inland areas and *Aedes sollicitans* and *Culex salinarius* in more coastal areas (Nasci and Edman 1981; Muul, Johnson, & Harrison 1975; Crans 1977; Freemaster and Getting 1941; Mangarelli 1977). Other studies have shown infrequent, but potentially important mammalian blood meals by *Cs. melanura* later in the season, suggesting *Cs. melanura* contributes both to the enzootic and epizootic cycle (Armstrong and Andreadis 2010; Estep et al. 2010; Molaei et al. 2013; Molaei et al. 2015).

A number of factors influence a mosquito's capacity to be a vector of EEEV including: 1) feeding behaviors, 2) habitat preferences and 3) proximity to hosts. Though *Cs. melanura*

primarily feeds on avian species, *Cs. melanura* has been shown to feed on mammals preceding and following bird migrations (Estep et al. 2010). New strains of EEEV are suggested to be linked to migrating avian hosts that frequent endemic regions, such as Florida (Armstrong and Andreadis 2013; Burkett-Cadena et al. 2015). In the case of EEEV infection in humans and horses, more opportunistic feeding mosquito species like *Ae. cantator* and—to a much lesser extent—*Cs. melanura*, cause the greatest incidence because they transition between birds and mammals depending on availability (Magnarelli 1977; USDA 2004; Molaei et al. 2013; Molaei et al. 2015).

### *Ecological niche*

*Cs. melanura*'s eastern geographic range is from southern Florida to Quebec, Canada (Mahmood and Crans, 1998; Woods, Dang, and Ellis 1979). It is generally found in lowlands, specifically freshwater hardwood swamps, and lakes (USDA 2004). Forested freshwater wetlands and swamp-like areas are a key component of *Cs. melanura* habitat preference and the environment in which the enzootic cycle of EEEV is most likely amplified (Mahmood and Crans 1998). *Cs. melanura* lays eggs in aquatic subterranean habitats, where its larvae then develop among the deeply shaded cavities of upturned tree roots and in holes under pads of sphagnum (Mahmood and Crans 1998). These root system and plant growth patterns provide the high concentration of organic matter needed to support oviposition and immature aquatic development (USDA 2004). In the northeast the larval development can take up to two or three months, and during that time tree root systems generally provide a stable, cool (<20°C), and acidic environment (Mahmood and Crans 1998; Andreadis et al. 2012). In the southern regions, such as Florida, development time is much quicker (Mahmood and Crans 1998), contributing to year round amplification of the virus and epizootic transmission (Burkett-Cadena et al. 2015).

*Cs. melanura* and EEEV activity have not always overlapped. Although *Cs. melanura* is present across all of the eastern United States and some of the southern regions of Canada (Wood et al. 1979; Mahmood and Crans 1998; USDA 2004; Andreadis et al. 2012), in the past decade, the virus has expanded its spatial range northward and lengthened its time of activity (Armstrong and Andreadis 2013). The disease moved into northern New England in the early 21<sup>st</sup> century, where an outbreak was recorded in 2005 in New Hampshire (Armstrong and Andreadis 2013). Also, later in 2012, Vermont experienced its first human case of EEE (Armstrong and Andreadis 2013). Additionally, a large number of EEE cases have been reported as far north as Maine and the virus is now active year-round in Florida (Freier 1993; USDA 2004; Gibney et al. 2011). Researchers have suggested Florida as the reservoir for EEEV. It is suggested that migrating birds reintroduce the virus to the northeast after winter each year, due to its constant activity in the south (Armstrong et al. 2008). Some suggest that this disease's range extension is due to climatic changes, which have increased temperature and precipitation and affected vector abundance (Armstrong et al. 2008; Andreadis et al. 2012; USDA 2004; Freier 1993).

### *Seasonality*

Eastern Equine Encephalitis virus transmission is extremely seasonal and weather dependent due to its dependence on mosquito vectors (USDA 2013; Armstrong and Andreadis 2013). Higher than average precipitation, mild winters, and elevated spring and summer temperatures, play a large role in vector abundance and dispersal patterns (USDA 2004; Armstrong and Andreadis 2013; Mahmood and Crans 1998 [Figure 1]). Higher levels of precipitation affect water tables and humidity (Day and Shaman 2008). Higher water tables can increase swampy areas where larval crypts are typically found (Mahmood and Crans 1998; Day and Shaman 2008 [Figure 1]). Therefore, transmission in both horses and humans occur during

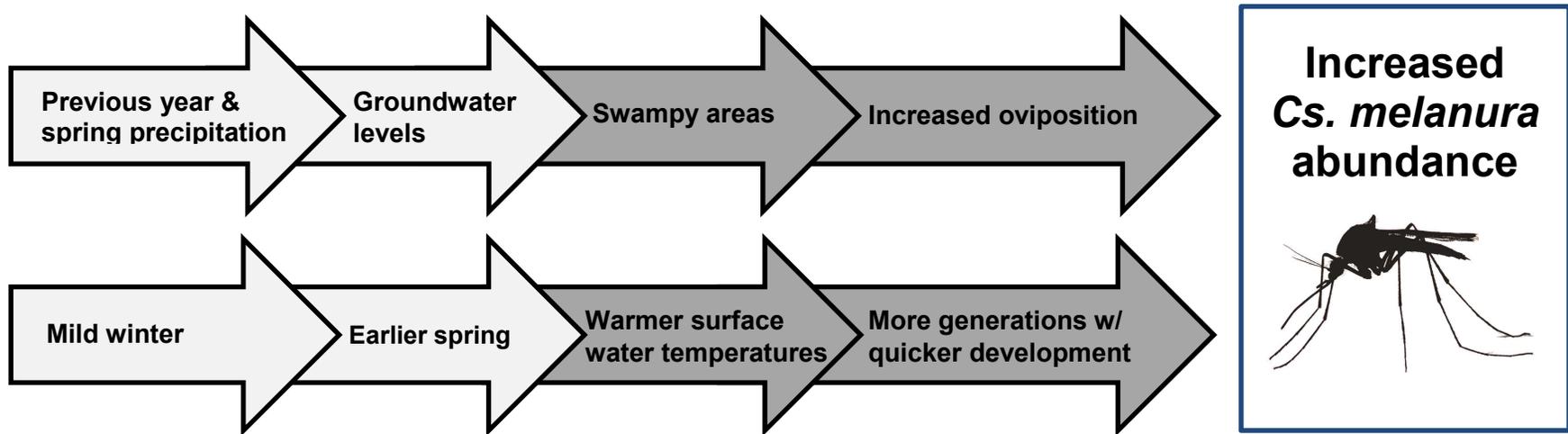


Figure 1: Conceptual model for *Cs. melanura* abundance related to meteorological and groundwater factors. Increased precipitation increases groundwater levels which increases surface water areas ideal for oviposition. Warmer winters lead to earlier springs and warm temperatures, which increases water temperatures where larvae reside. This would diminish developmental time and makes the possibility for more generations of adult females more likely.

warm rainy seasons, generally in mid-to-late summer and even fall (USDA 2004; CFSPH-IICAB 2008-2015). The epizootic cycle is then late in the enzootic season for *Cs. melanura*—around August and September—in northern states such as Vermont or Michigan (USDA 2004; Molaei et al. 2015). In southern states, specifically Florida, equine cases peak around June and July, due to an earlier and longer EEEV amplification cycle (USDA 2004). This has been linked to availability of avian hosts and year-round adult mosquito activity; both are a reflection of seasonal factors, like increased temperature and humidity.

Unfortunately, long-term trends in temperature and extreme precipitation have been shown to be increasing in most of North America, as indicated by the National Climate Assessment 2014 report (Kunkel et al. 1999; Luber et al. 2014; Horton et al. 2014). Recurring extreme, multi-day precipitation events have a great effect on flood frequencies and hydrologic cycles (Kunkel et al. 1999; Horton et al. 2014), which can facilitate habitat suitability for the vector *Cs. melanura* (Mahmood and Crans 1998; Day and Shaman 2008). Moreover, warmer than average seasonal temperatures can speed up development time, which can increase the opportunity for multiple generations of *Cs. melanura* each season (Mahmood and Crans 1998; Andreadis et al. 2012). Temporal pattern examination and modeling may link *Cs. melanura* abundance and EEEV activity to meteorological events and would facilitate future resource management and control. A diagram of proposed weather mechanisms leading to increased *Cs. melanura* abundance are depicted in Figure 1.

### *Connecticut*

Connecticut (CT) is the third smallest state in the U.S. (5,543 mi<sup>2</sup>) with a high horse-to-land ratio (>20 horses per 1000 mi<sup>2</sup>), fifth in population density (~738.1 people/mi<sup>2</sup>), and a well-established *Cs. melanura* population—ideal for EEE transmission (OHorse 2015; United States

Census Bureau, 2010). Also, 9% of the landscape includes forested and non-forested wetlands, mostly in the coastal south, based on 2006 land cover estimates (CLEAR 2012). These forested freshwater wetlands and swamp-like areas are a key component of *Cs. melanura* habitat, and the environment in which the enzootic cycle of EEEV is most likely amplified (Mahmood and Crans 1998). Additionally, CT has robust EEEV vector data spanning over 16 years and across much of the southern part of the state. These data supported the longitudinal studies of density and distribution of *Cs. melanura* in both time and space that were examined in this paper's analyses.

### **Objectives**

The goal of this investigation was to test for a northward trend of *Cs. melanura* abundance and EEE disease incidence in CT from 1997-2012, to see if there was a link with meteorological and well water data. We also sought to determine the geographic distribution of *Cs. melanura* abundance using habitat analysis methods detailed in Diuk-Wasser et al. (2006). Both datasets used CDC light trap data.

Temporal patterns of abundance were examined against winter average temperatures, groundwater levels, and various precipitation values, with the goal of generating a predictive linear regression model in Analysis I. The habitat model of Diuk-Wasser et al (2006) was expanded from the two southwestern counties to the whole state, and we expanded the time frame from 1997 to 2012 for Analysis II. Modeling temporal and spatial patterns may identify when and where human and horse dead-end hosts are at greatest risk for exposure to EEEV infected *Cs. melanura*, and where prevention efforts can be concentrated.

### **Methods**

Methods used to explore temporal and spatial trends in abundance in this investigation included meteorological and landscape-based variables and regression techniques. Research in

Analysis I used time-series analysis to assess trends in *Cs. melanura* abundance over 16 years, based on meteorological data as an alternative to standard linear regression methods typically used on continuous data. Analysis II focused on expanding the spatial analysis method of Duik-Wasser et al. (2006) that used landscape components, based on vector data in two southern counties of Connecticut, to create vector habitat models.

### ***Vector data***

Mosquito collection sites were in the majority of southern CT and specific areas in the central and northern regions of the state. Sites pertaining to this investigation were the fixed CDC light traps (N=91) maintained by the Connecticut Agricultural Experiment Station. The majority of traps were concentrated in the southern part of the state, coinciding with wetlands. Trapping sites were placed in high-risk mosquito areas and areas that experienced nuisance complaints made by residents (See Figure 2). Trap sites included, but were not limited to: parks, greenways, golf courses, sewage treatments plants, landfills, undeveloped wood lots, and wetlands (Diuk-Wasser et al. 2006).

For this investigation 83 (91%) of the 91 traps established in the study period were utilized to compensate for inconsistencies in overlapping data layers, trap type used and legitimacy of each site to represent a 'good' *Cs. melanura* location. Good traps are considered sites where at least one *Cs. melanura* was found for over 50% of the trapping period. Light trap sites were established across CT, with the majority of sites in the sites in the southern counties (N=61). Vector data available ranged from 1997 through 2012 with approximately 1159 trapping events per year (minimum 446 and maximum 1515). Traps were excluded for the following reasons: being a gravid trap rather than a light trap (N=1); position along an appendage line of Landsat data (N=1); and if no *Cs. melanura* were collected for over half of the sampling period

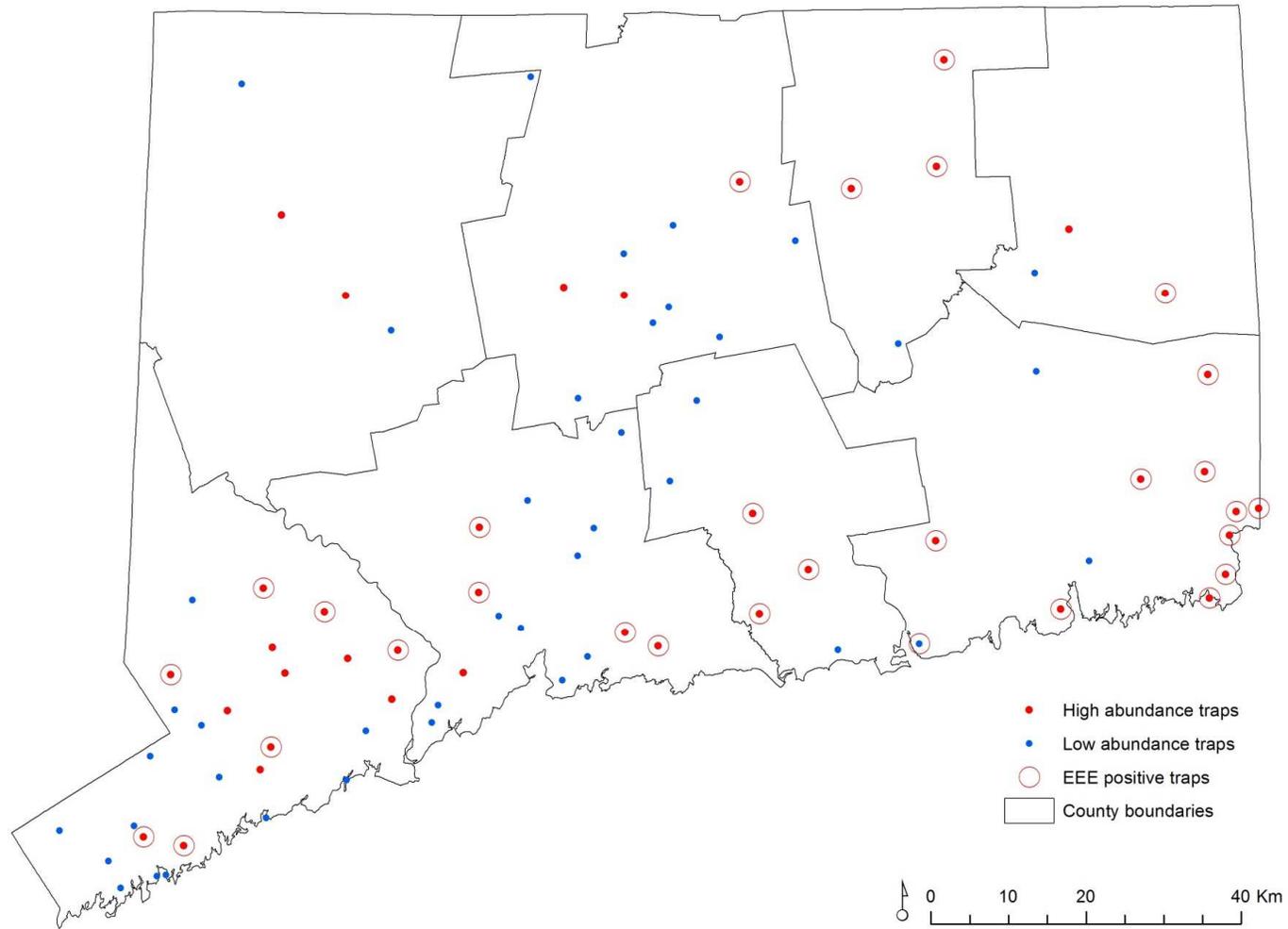


Figure 2: Distribution of 83 Centers for Disease Control and Prevention (CDC) light traps with data from 1997-2012 used in temporal and spatial analyses. High abundance sites are red and low abundance sites are blue, as used in the logistic regression. CDC light trap sites that recorded a EEE positive pool of *Cs. melanura* are indicated by a hollow red circle. County boundaries are provided.

(N=6). These traps were distributed across the southern portion of the state with the majority of exclusions in Fairfield (N=4) and New Haven (N=3) counties with an already robust number of trap sites.

Data from each trap site included the number of mosquitoes, mosquito species, and the presence or absence of EEEV within mosquito pools. Data were collected at a periodicity of twice a month, from September to May, to once every week, in June to October. To test for EEEV, mosquitoes were pooled by species in groups up to 50. If traps contained more than 50 mosquitoes, multiple pooled tests were analyzed and added together for trap night positivity. Mosquito collections were standardized by trap night (mosquito number/trap night) across the study area for each year--across all years. (Table1).

Table 1: Average *Cs. melanura* per Trap Night and Percent of Pools Positive for EEEV by Year

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
<i>Cs. melanura</i>																
#	11.82	13.01	2.93	7.63	5.01	2.51	10.27	5.73	3.72	5.44	4.18	5.17	13.92	4.05	10.3	5.77
std. dev.	(14.97)	(18.53)	(4.18)	(21.23)	(7.8)	(5.49)	(18.06)	(10.47)	(8.87)	(10.14)	(8.1)	(14.09)	(20.98)	(7.03)	(17.48)	(9.88)
Positive																
pools %	0.19	1.61	0	0.52	1.17	0	4.26	1.86	0	0.77	0.99	0	5.16	0.17	0.30	
std. dev	(0.86)	(6.28)		(1.64)	(2.69)		(5.73)	(3.90)		(2.10)	(2.77)		(7.28)	(0.79)	(0.79)	1.0 (2.20)

Table 2: Data Sources Used and Variables Extracted in the Study

<b>Data Source</b>	<b>Variables</b>	<b>Description</b>	<b>Format</b>	<b>Temporal Resolution</b>
<i>Temporal</i>				
PRISM Climate Group	Precipitation	Current and prior year and spring and winter averaged from maximum and minimum values	raster	1996-present
	Temperature	Current winter averaged from maximum and minimum values	raster	1996-present
<i>Spatial</i>				
U.S. Fish and Wildlife Service National Inventory	Wetlands (Palustrine)	Non-tidal wetlands dominated by trees, shrubs, emergence, mosses or lichens or wetlands that occur in tidal areas where salinity is below 0.5%. (Cowardin et al. 1992)	polygon	2006
USGS National Land Cover Database (NLCD)	LULC change	Percent change between 2006 and 2011 LULC classifications	raster	2006 to 2011 comparison
USGS Landsat 5 Thematic Mapper (TM)	NDVI April	NDVI from TM April 2009 scene	raster	2009
	NDVI September	NDVI from TM September 2009 scene	raster	2009
	NDVI difference	Difference in NDVI between the September 2009 and April 2009 scenes	raster	2009
<i>Temporal and Spatial</i>				
USGS National Water Information System	Groundwater level (GWL)	Current and prior year averages and winter difference from average	point	1996- present

### ***Environmental Variable Sources***

Analysis I explanatory variables were selected based on key features of *Cs. melanura* biology and reproductive cycle (Mahmood and Crans 1998; Day and Shaman 2008; Andreadis et al. 2012; Armstrong and Andreadis 2013). Analysis II variables were chosen based on Diuk-Wasser et al. (2006). All variables were trimmed to the spatial extent of CT state boundary. The data sources are provided in Table 2 and include the following:

#### *Analysis I Variables*

1. Monthly averaged meteorological data from 1996-2013 were obtained from the PRISM Climate Group at Oregon State University (<http://www.prism.oregonstate.edu/>). These include interpolated values for temperature and precipitation that were at a spatial resolution of 2 km. Average temperature and precipitation values were extracted at individual grid cells as proxies for points of each trap site across the span of the study period data for each month (June 1997 to Dec 2012), derived from downloaded minimum and maximum temperatures. Spring and winter designations were generated for temperature and precipitation. Spring was defined as April and May and winter was defined as the previous year's December through current year's March, based on the region's typical seasons used by both Andreadis et al. (2012) and Mahmood and Crans (1998). Season measurements were the result of averaging across all months that define the season.

#### *Analysis II Variables*

2. U.S. Fish and Wildlife Service National Wetlands Inventory (NWI): Data downloaded from this inventory followed the Cowardin et al. (1979) method for wetland classification with remotely sensed data across the United States. The classification system from 2006

was used. All final data were projected from UTM to the Albers Equal Area map projection, and quadrangle boundaries were merged

([http://www.fws.gov/wetlands/Data/metadata/FWS\\_Wetlands.xml](http://www.fws.gov/wetlands/Data/metadata/FWS_Wetlands.xml)). Palustrine areas designated for this investigation are a combination of freshwater emergent wetlands, freshwater forested shrub/wetlands, and freshwater ponds (Cowardin et al. 1979).

3. Landsat 5 TM imagery: Landsat 5 TM imagery from April 28, 2009 (leaf-off) and September 19, 2009 (leaf-on) were acquired from United States Geological Survey (USGS) Earth Explorer (<http://earthexplorer.usgs.gov/>). The dates selected had less than 0.0 - 0.5% cloud cover in the same year that had the appropriate spatial extent in the study period's date range. These remotely sensed data were collected at 30 m x 30 m resolution in reflective wavelengths. Dates chosen were used to calculate the Normalized Difference Vegetation Index (NDVI) difference. High NDVI difference represents a deciduous forest area where there was a large contrast between leaf-on leaf-off between September and March/April. Basic image requirements included minimal cloud cover (ideally <1%: Diuk-Wasser et al. 2006) and spatial extent; therefore, the selected images covered the desired geographic.

4. Land Use/Land Cover: Land Use/Land Cover (LULC) data were downloaded from the National Land Cover Database (NLCD) using the 2006 reproduction. Previous examination of these data has shown little alteration in urban and wetlands classifications between 2006 and 2011 (Jin et al. 2013). This indicates that current wetland areas are fairly consistent across the time period of this analysis and between this and the previous assessment by Diuk-Wasser et al. (2006). Additionally, some large acreages of inland wetlands are protected by agencies such as The Nature Conservancy, though in 1992 an

upper estimate of 5% percent of all wetlands are were considered at risk for change in the coming years (Metzler and Tiner 1992).

### *Analysis I & II*

5. Hydrological data were accessed from the USGS National Water Information System Mapper and Time Series Creator (<http://maps.waterdata.usgs.gov/mapper/index.html>). Active groundwater well sites were selected. Properties considered in the investigation from these sites included depth to water and shallow well whole depth (<100ft), such that the well would more closely reflect surface water levels. Time frames available for this information with currently active sites were from the start of 1997 to 2015. All data were georeferenced.

### ***Analysis I: Temporal***

Time-series regression methods were used to identify groundwater levels and meteorological predictors for *Cs. melanura* abundance each year. Standard regression methods cannot always portray the characteristics of time-series because lagged linear relations in the dependent variable may contribute to correlation into the series (Shumway and Stoffer 2006). In using an autoregressive integrated moving average (ARIMA) regression model, the dependent variable is rendered stationary and the independent variable can then be used as a lagged version of the dependent variable, which is necessary for assessing validity of future predictions (Shumway and Stoffer 2006). The autoregressive term and error, or moving average can help determine the lags needed.

The ARIMA model assumes the time series is stationary or can be modified to achieve stationarity (Shumway and Stoffer 2006). Trend in the series was assessed using the Dickey-Fuller test for stationarity. For the trend series, the null hypothesis was that the data series has a

unit root and was not stationarity. Unit roots occur when 1 is the root of the characteristic equation of the process. Stationarity was considered a lack of trend in the data being used. Investigations of stationarity were important, because if trends are not taken into account or incorporated in a linear regression this can cause erroneous coefficients and reduction in model validity. Additionally, the white noise Q test was performed on *Cs. melanura*, as a second test for autocorrelation or serial correlation in the data. The null hypothesis for the Q test was that there is no serial correlation, i.e. the data are characteristic of a white noise process. As with spatial autocorrelation, temporal autocorrelation refers to a significant relationship between points in time.

Cross-correlations between mean numbers of *Cs. melanura* per trap night/per year and meteorological, and also groundwater variables for up to 5 lags were tested. Five-year lags were assessed across the 16 years of available data. Lag selections were based on: the Akaike Information Criterion (AIC), agreement among Schwarz's Bayesian information criterion (SBIC), and Hannan and Quinn information criterion (HQIC), as well as agreement among variables (Ivanov and Kilian 2001). Correlated lag years ( $r > 0.50$ ) with the standardized *Cs. melanura* data were tested using univariate linear regression models, as were their original, unlagged versions, for comparison of significance level changes. The dependent variable was tested for autocorrelation using correlograms to help specify an autoregressive integrated moving average (ARIMA) model.

Automated forward stepwise variable selection was used to generate a final model. Variables were included based on parsimony (lower AIC). The final linear model was tested for homoscedasticity using the Breusch-Pagan/Cook-Weisberg test, and sensitivity using the leave-one-out-method (Breusch and Pagan 1979). The leave-one-out-method generated an average  $R^2$

and 95% confidence intervals (CI). Statistical significance was set at  $p < 0.05$ . All data were analyzed using STATA version 13 (StataCorp LP, College Station, TX).

## ***Analysis II: Spatial***

### *Spatial Variable Extraction*

All thematic layers taken from the data sources above were converted into ArcMap (ESRI Inc., Redlands, CA) GRID format and re-projected using the Universal Transverse Mercator, zone 18N projection, World Geodetic System 1984 datum (matching original Landsat imagery) at a 30 m x 30 m resolution. Land Use Land Cover (LULC) variables were transformed to a raster format representing the Euclidean distance from the target cell to the nearest variable pixel of choice (palustrine, developed, coniferous forest). Using the “focal statistics” tool in ArcMap 10.2, the mean value of 100m circular intervals were created out to 800m, resulting in 8 new rasters for each variable. Well data were averaged each year across 7 active well sites in CT. Data were transformed to Thiessen polygons and then rasterized (30 m x 30 m cell resolution) for sampling at point locations. Using the “Sample Points to Table” tool in ArcMap, all variables were sampled to trap site points at all buffer sizes and then put into a comprehensive table. The data were then combined with mosquito data using the “Join” feature in ArcMap 10.2 in preparation to build a logistic regression model.

### *Spatial Regression Model Development*

Logistic regression models were used to identify habitat characteristics that compared high versus low abundance *Cs. melanura* trap sites, creating probability models for mosquito presence. The final response variable was the median split of abundance averaged across 16 years (1997-2012) for 83 trap sites across CT, 42 classified as high abundance trap sites and 41 classified as low abundance sites (Figure 2). For the spatial models, the median split of high

versus low abundances was used as the response variable. Linear regression was not used, because data were highly skewed and predictive models in ArcGIS are more suitable to develop with logistic models (Pfeiffer et al. 2012). Moreover, the dichotomy created by logistic regression is useful when trying to identify high abundance or risk areas for enhancing surveillance (Pfeiffer et al. 2012).

Spatial autocorrelation violates the assumption of independent observations necessary for a logistic regression. To determine the extent of spatial autocorrelation of *Cs. melanura* abundance for all years of the study, a Moran's I spatial autocorrelation test was performed in ArcMap with a squared inverse distance weighting and a distance threshold of approximately 27,000 meters. Moran's I ranges in value from -1 to 1, with 1 being complete positive spatial autocorrelation and -1 being complete negative spatial autocorrelation. Monte Carlo randomization (999 permutations) was used to test validity and generate a p-value for the z-score. A 27 km threshold was chosen, as it is close to the median between neighboring points and also ensures each trap site has a neighbor for comparison.

Univariate logistic regression was conducted on standardized mosquito abundance (*Cs. melanura* [all years] /trap night [all years]), with each predictor variable at each buffer size. Any predictor variable that had a significant p-value ( $p < 0.05$ ) at over half the radius sizes was retained in multivariate analysis, then fit for any combination of up to two predictors. Buffer size most appropriate for each variable was selected based on the smallest AIC. Collinearity was assessed by using Spearman's rank correlation; if two variables were considered highly collinear ( $r_{\text{Spearman}} > 0.75$ ), those with higher significance were retained. This method has been previously used by Diuk-Wasser et al. (2006) and Sokal and Rohlf (1989).

The final model was selected based on variable significance and AIC (Akaike 1974) to rank candidate buffer sizes for each variable with significance. AIC determines the most parsimonious model for the given data by balancing the number of model parameters with the fit of the model, in order to reduce over-fitting (Akaike 1974).

The logistic regression models' underlying assumptions were tested. The final model was assessed for linearity in the log-odds, using locally weighted scatterplot smoothing. This is created by running a regression of the dependent variable on the independent variable, using a certain data point and data near the point and weighting points less as they move further away from the target data point (Vittinghoff et al. 2005). This was repeated through the whole data set, creating a smoothing regression for each data point (Vittinghoff et al. 2005). Locally weighted scatterplot smoothing was performed for each independent variable to determine best fitting buffer size. Influential observations in the dataset were determined using Pregibon's Delta-Beta influence statistic, as they can have a large effect on estimated coefficients. Influential observations were removed if they resulted in a greater than 10 percent change in coefficients. Outliers were expected in the high abundance range, due to skewness in the vector trapping data. Goodness-of-fit was evaluated using the Hosmer-Lemeshow test (Hosmer and Lemeshow 1989; Hosmer and Hjort 2002); a significant result ( $p < 0.05$ ) indicates the model does not fit the data well. A summary measure of predictive accuracy (area under Receiver Operator Characteristic [ROC] curve) was also calculated for model predictability.

Sensitivity/specificity probability cutoffs were determined using the STATA command "lsens", which graphs sensitivity and specificity versus a probability cutoff. The probability cutoff that maximized both sensitivity and specificity was chosen for the output of the final

model. All statistical analyses were conducted using STATA 13.0 (Stata Corporation, College Station, TX).

### *Spatial Regression Model Validation*

To determine the predicted abundance of *Cs. melanura* across CT, logistic regression equations were fed back into ArcMap using the Raster Calculator (Spatial Analysis Tool) with the “smoothed” predictor variables of interest. The inverse transformation of the logistic regression:

$$\text{Probability (of high abundance)} = \exp \left[ \frac{Y}{1 + \exp(Y)} \right]$$

*or simplified: 1/(1 + exp(-Model))*

was applied to the set of linear predictors (Y) to transform the data from the logit values to values between 0 and 1, in order to determine probabilities.

Using a conditional statement in Raster Calculator, the efficiency of the model was tested by creating a binary raster surface derived from the cutoff points determined from optimal specificity/sensitivity tests. This classifies areas as 1, high abundance, if the probability model was above the cut point, or 0, low abundance, if the probability model was below the cutoff. Best sensitivity/specificity was capitalized at approximately 0.6, using a sensitivity and probability plot.

This binary model was used to determine efficiency. Efficiency gives an understanding of percent of the correctly classified high abundance sites in relation to the study area. The equation to for the model’s efficiency was:

$$\text{efficiency} = 1 - \left( \frac{\text{percentage of total area within most likely category}}{\text{percentage of total sites within most likely category}} \right)$$

The efficiency measures model goodness of fit to the data, where values near 0 indicate random change and values near 1 indicate high efficiency or near-perfect fit. Efficiency of 0.5 and above was considered very efficient (O'Sullivan and Unwin 2010). All final spatial model representations and efficiencies were performed in ArcMap version 10.2 (Environmental Systems Resource Institute [ESRI], Redlands, CA).

## **Results**

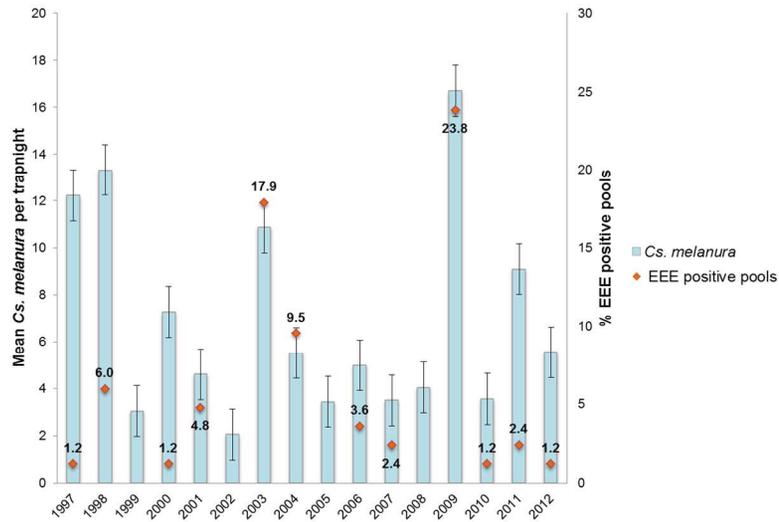
### *Analysis I: Temporal*

*Cs. melanura* abundance varies across time from approximately 3 to 14 mosquitoes per trap night with a considerable amount of between trap variance as well (Table 1). Peak years for both *Cs. melanura* abundance and percent of light trap sites with EEEV positivity are 1998, 2003, and 2009 (Figure 3A). EEE horse cases in the state do not occur till the end of the data period in 2009 and 2012: 2012 being a relatively low abundance year (Figure 3A).

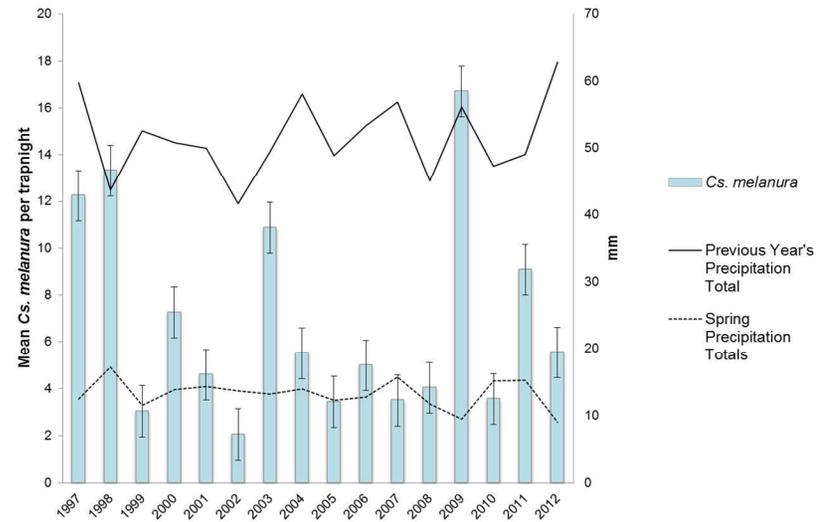
Yearly *Cs. melanura* abundance was plotted over 16 years of the available data period in comparison to average winter temperatures, previous year precipitation totals, previous spring only precipitation totals, average groundwater levels, difference of winter groundwater averages to years average, and the previous year's vector abundance (Figure 3A-D). Annual percent EEEV positivity of all sites and number of horse cases are also depicted in Figure 3A.

Comparing percentage of pools positive to CDC weeks each year, peaks occur from week 32 to 44. Looking at winter difference in groundwater levels from the average, in years 2003, 2008 and 2011 there is the greatest positive difference at approximately 1.5, 0.0 and 1.0 ft, which are a year behind peak *Cs. melanura* abundance years with the exception of 2011, though there is a moderate increase in *Cs. melanura* density (Figure 3C).

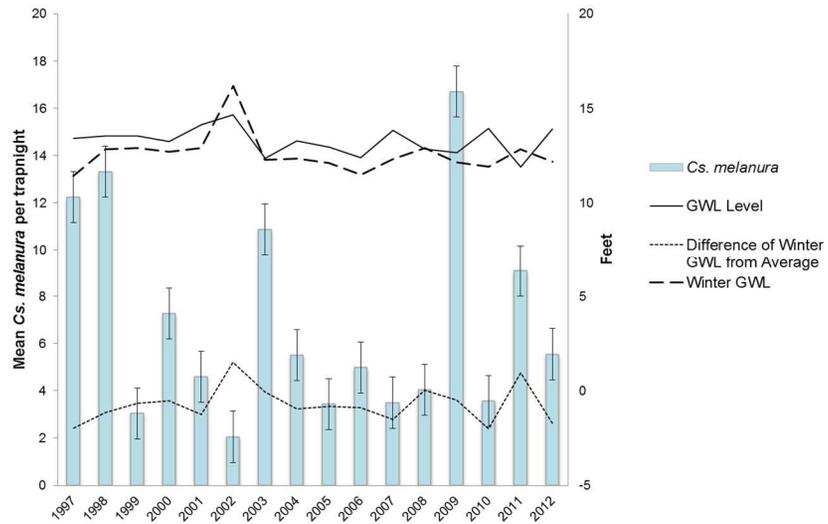
A



B



C



D

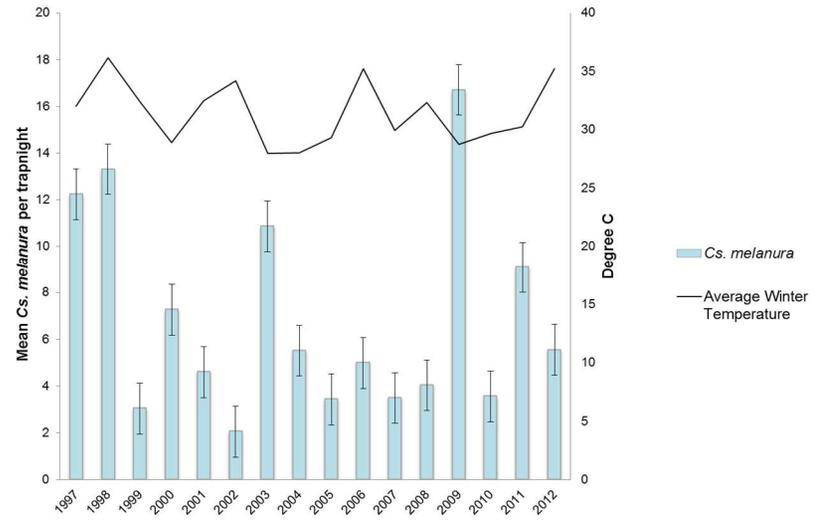


Figure 3: Comparison of average *Cs. melanura* per trap night per year and a: percent of mosquito pools positive per trapnight b: average groundwater level and winter groundwater levels differenced from average (ft) c: the previous year precipitation totals and current year spring precipitation total (mm) d: average winter temperature (°C).

Little pattern can be seen from groundwater levels alone, though dips in groundwater level seem to coincide with abundance going from low to high. Correlation coefficients also reflect a negative relationship between groundwater level increases and winter differences in groundwater levels from averages and *Cs. melanura* abundance (-0.85 and -0.71, respectively) (Table 3). This may mean that previous year levels being higher than average can give good indication of decreased abundance, as the correlation for lagged winter levels is 0.74 and significant (Table 3). Precipitation totals from spring of each year are more uniform compared the other variables, with a peak only in 1998 and a downward dip in 2009 and 2012 (Figure 3C). Previous year's totals are lower for higher *Cs. melanura* abundance years (Figure 3C) though this trend is not significant (Table 3). The winter temperature was the warmest of the study period in 1998. This year is also the year of the second highest *Cs. melanura* abundance and third highest EEEV positivity (Figure 3D; 2A). Long-term trends for this are not evident because there is a negative correlation between average winter temperatures and *Cs. melanura* abundance (-0.46), though this is not significant (Table 3).

Table 3: Pearson's Partial Correlation Coefficients of Explanatory Temporal Variables for *Cs. melanura* Abundance

<b>Variable</b>	<b>Correlation</b>	<b>p-value</b>
<b><i>GWL Averages (ft)</i></b>		
Current Year	-0.85	0.0039*
Prior Year Winter	0.74	0.0232*
Winter Difference from Average	-0.71	0.0312*
<b><i>Precipitation Totals (mm)</i></b>		
Current Year	-0.63	0.0701
Prior Year	-0.20	0.6144
Spring	-0.17	0.6592
<b><i>Temperature Averages (°C)</i></b>		
Winter	-0.46	0.2176

\* significant p-value ( $p < 0.05$ )

Linear regression was used to predict *Cs. melanura* abundance. Vector abundance had a skewed distribution, improved only by transforming to a square root or log; however, these transformations did not normalize the data distribution. Thus, predicted data were left in their original form and presented and interpolated here only for educational purposes. Forward stepwise methods for model selection were not successful, regardless of lag selection based on cross-correlations of predictor variable ( $p < 0.10$ ) and explanatory variables. An elevated p-value ( $p = 0.07$ ) was selected to increase predictive ability of the regression model by allowing more variables to be included.

Serial correlation, using the white noise Q test was not evident, with a p-value in acceptable bounds ( $p > 0.05$ ). However, testing for trend in the data with a Dickey-Fuller test for unit root was positive, indicating a trend was present and needed to be accounted for. Differencing *Cs. melanura* abundance by its first derivative did not alter the presence of trend at a lag of 5. However, by decreasing time lag to 3 years, the presence of trend was lessened slightly, with a p-value 0.06 rather than values  $> 0.60$ . Estimating autocorrelation with precision is necessary, since estimating *Cs. melanura* abundance is difficult and erroneous if there are internal dependencies not controlled by the predictive variables. Adding an autoregressive term or detrending data ensures that the mean and autocovariance functions meet the conditions of stationarity (Shumway and Stoffer 2006). Though there is overall no autocorrelation, there is nonstationarity in the data (trends over time) that could not be modeled properly. An autoregressive term could not be reached to create an optimized model for time-series such as an ARIMA, through examining lagged data. While a linear regression model could have been presented, without accounting for trend in the data, regression coefficient estimates would have

been insufficient to account for changes in abundance, thus producing sub-optimal forecasts (Granger and Newbold 1974).

### *Analysis II: Spatial Regression Modeling*

There was no significant spatial autocorrelation with a Moran's I of 0.20 ( $p = 0.123$ ), indicating a random pattern for the trap sites, weighted by *Cs. melanura* abundance, and the ability to proceed with traditional logistic regression methods for predictive modeling of *Cs. melanura*, assuming all traps to be spatially independent.

Similar to Diuk-Wasser et al (2006), the logistic regression models for *Cs. melanura* high/low abundance across CT included distance palustrine pixels and NDVI difference averaged across 800 m as their final predictors. Table 4 depicts the candidate logistic regression models, as well as their fit statistics. The best model was considered the one with the highest predictive ability and parsimony (AIC). Predictive variables for *Cs. melanura* across all radial averages (buffers) were significant ( $p < 0.05$ ) only in the case of distance to palustrine pixels in univariate regressions. However, NDVI difference (NDVI diff in Table 2) showed significance for over half of the buffers sampled (100 m and 500 m 800 m) and was left in the multivariate models. Groundwater levels (GWL in Table 4) lacked significance, but was left in to test if it modified the predictive ability and efficiency of Model 1.

The logistic regression equations are presented below (Table 4).

Model 1

$\log \text{odds } (Cs. melanura)$

$$= - 0.006 (\text{distance to palustrine } 800) + 9.12 (\text{NDVI difference } 800)$$

$$- 0.002 (\text{Groundwater levels}) - 1.87$$

Model 2 (final)

$$\log \text{odds} (Cs. melanura) = - 0.006(\text{distance to palustrine } 800) + 9.12 (\text{NDVI difference } 800) - 1.88$$

Table 4: Analysis II Logistic Regression Models

Variable	Coefficient	p-value	AIC	Area under ROC	Sensitivity (%)	Specificity (%)
<i>Diuk-Wasser et al. 2006</i>			-	-	87.5	81.3
NDVI diff 800	0.48				75	75
DPALU	-0.45				81.3	81.3
<i>Model 1</i>			1.224	0.786	65.6	78.6
NDVI diff 800	9.122	0.028*				
DPALU 800	-0.006	0.065				
GWL	-0.002	0.964				
<i>Model 2</i>			1.2	0.785	65.6	78.6
NDVI diff 800	9.12	0.028*				
DPALU 800	-0.006	0.064				

\*significant p-value ( $p < 0.05$ )

NDVI diff 800 represents NDVI difference averaged over 800m

DPALU 800 represents distance to palustrine averaged over 800 m

GWL represents groundwater level average for the study period of 1997-2012

Following model transformation to a probability scale, the preferred model (2) indicates the probability of high *Cs. melanura* abundance within a given pixel across CT (Figure 4). The best models for *Cs. melanura* comprised a slightly negative association with distance to palustrine pixels and a strongly positive association with NDVI difference that is, higher abundance occurred in deciduous forested sites closer to wetlands. The ROC was 78.5% indicating good model fit (Model 2). Including groundwater levels did not significantly improve the model (area under the ROC curve 78.6%, AIC =1.22) and showed a very small inverse relationship between increased levels and increased *Cs. melanura* levels (Model 1). Model 2, using only NDVI difference and distance to palustrine, good predictive efficiency at 49.2 %, and the sensitivity and specificity were moderately high at 78.6% and 66.7% and 72.6% correctly classified.

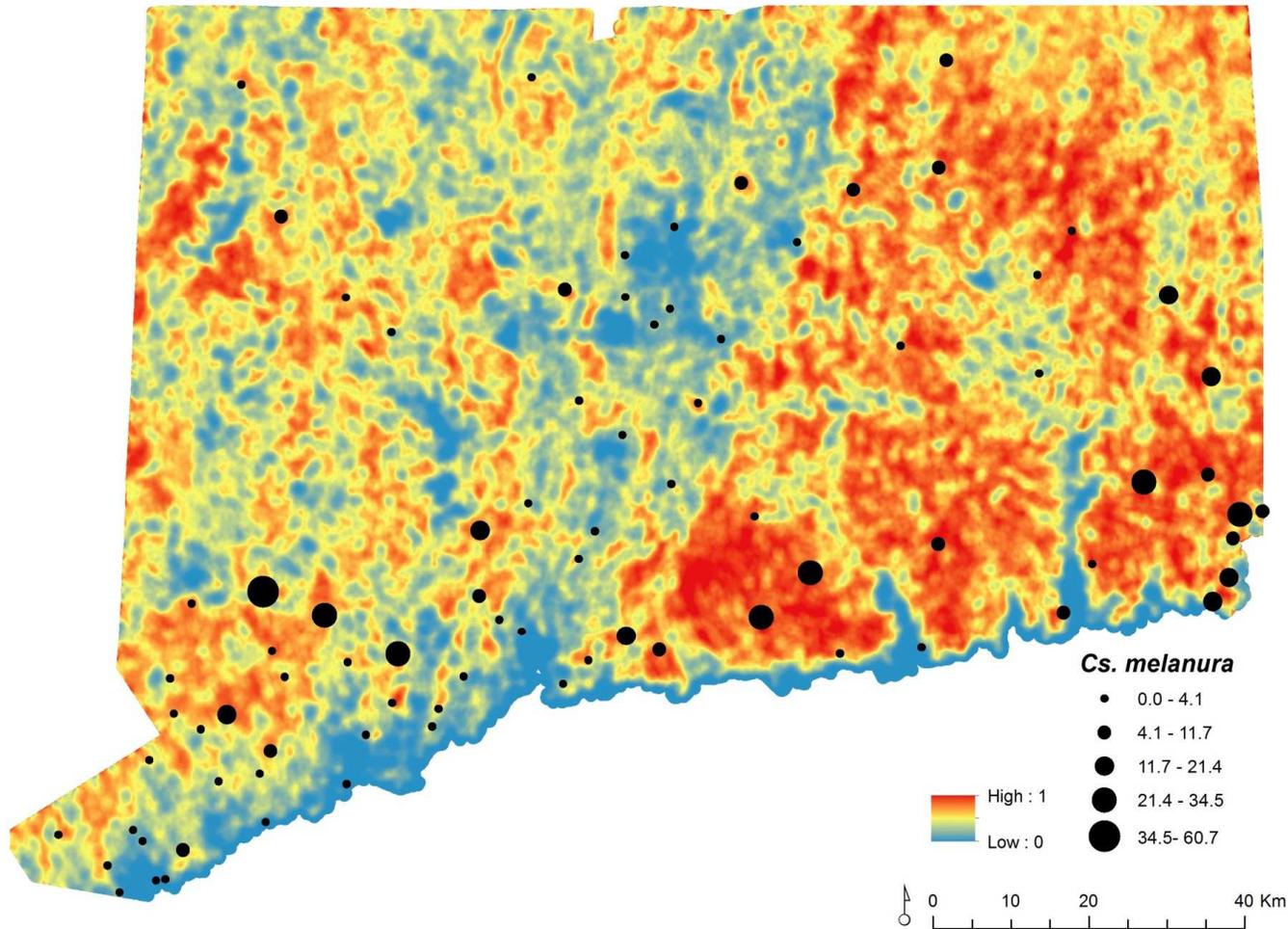


Figure 4: Probability of high *Cs. melanura* abundance in Connecticut. High values are depicted by red and low values are blue. These dark grey symbols show the average number of mosquitoes per trap night in CDC light traps between 1997 to 2012, using the Jenks Natural Breaks for categorization for 5 class.

Model 2 was used in the generation of a probability map presented in Figure 4. The chance of having a high abundance of *Cs. melanura* is what is being represented in the spatial model of probability. Observed high abundance sites are primarily in the southern most portions of the state. The probability map generated in Figure 4 shows the highest area of *Cs. melanura* abundance levels on the eastern side of CT, but also has high probability predicted for the western side. Overall, the predictive spatial model corresponds to observed values.

### **Discussion**

This study found high *Cs. melanura* abundance was predicted more often in the eastern portion of the state, especially in the south. This corresponds with known trap site data. Additionally, higher previous year winter groundwater levels were shown to help characterize high abundance years based on Pearson's correlations although graphical representation across time of this appears muffled. Geographic distribution results were achieved by transforming point-based trap site data to surface data, estimating *Cs. melanura* abundance at unsampled locations. This analysis was based on a previous method by Diuk-Wasser et al. (2006), but extended for the entire state of CT. Temporal distributions of *Cs. melanura* abundance were examined from 1997 through 2012 using meteorological factors, though ARIMA regression models could not be produced.

Remotely sensed data and wetland classifications resulted in a similar predictive power as the previous study of Diuk-Wasser et al. (2006) that used the same data from a shorter time period. The outcome variable, *Cs. melanura* abundance, had only slight spatial autocorrelation, which is consistent with Diuk-Wasser et al. (2006), even with a larger sample size and area. However, coefficient estimates were dissimilar (Table 4). Different estimates may be due to using the full state, expanding the study period, slight method modifications, or differences in source data.

The final model had two variables, which coincided with habitat preferences of *Cs. melanura*, demonstrating biological soundness (Mahmood and Crans 1998; Andreadis et al. 2012). The two predictor variables of importance were NDVI difference and distance to the nearest palustrine pixel at 800 m smoothed averages. NDVI difference is a component related to deciduous forest habitats and palustrine pixels are estimates of freshwater swamp-like areas. *Cs. melanura* is one of the most common mosquitoes in Atlantic white cedar, densely wooded red maple swamps, and sphagnum bogs of CT (Andreadis et al. 2004). The southwestern section of the state had some of the highest averaged *Cs. melanura* counts, but the probability map did not reflect the level of this trend as well as the eastern section of the state (Figure 4). This suggests that the three high abundance sites are outliers for values of NDVI difference, the variable with the greatest impact in the logistic regression model. Since high positive NDVI difference is associated with deciduous forest pixels, looking at this variable in more urbanized areas averaged across 800 m, as is the case in southwest CT, the NDVI difference value would diminish compared to more open, rural areas where forests can flourish. Based on NDVI difference in Figure 4, if the three high abundance sites in the west were next to heavily populated areas with fewer green spaces, changes in NDVI would be minimal and would have reduced probability of *Cs. melanura*. Had this section of the study area been isolated and run separately, the probability map would have rescaled a range of values that described highly populated areas.

Groundwater level was a new variable that was explored in addition to the previous predictors outlined in Diuk-Wasser et al. (2006) for examination of habitat preference. *Cs. melanura* lays eggs in aquatic subterranean habitats, where their larvae then develop among the deeply shaded cavities of upturned tree roots and also in holes under pads of sphagnum (Mahmood and Crans 1998). As such, groundwater levels were used as a proxy for increased

oviposition areas for eggs and larvae, because higher water tables potentially increase swampy areas where these cavities or ‘crypts’ might be found (Mahmood and Crans 1998; Day and Shaman 2008). Groundwater level data were not statistically significant in the spatial logistic regression models, but were still considered important for its perceived role in standing water.

Analysis I attempted to examine meteorological factors that indicated climate change and expansion northward, but such evidence could not be supported due to limited modeling. A few meteorological factors can be hypothesized to influence vector abundance and EEEV positivity when compared graphically. A mild winter is thought to shorten the amount of time required for larval development, as the spring development starts sooner it shortens larval stage hibernation. *Cs. melanura*, as a multivoltine species, can have two or three overlapping productions each year in places in the northeastern United States (Mahmood and Crans 1998, Andreadis et al. 2012). A warm winter, like the one recorded in 1998, may have caused the second largest increase in *Cs. melanura* abundance and third in EEE positivity (Figure 3A-D), since earlier generation of adults increases the chance of a third generation in the year. It also follows that increases in winter groundwater levels from average, especially in a mild winter, would expand flooded areas, subsequently increasing oviposition sites. This trend is not entirely realized by the data at the year resolution, though if the difference of winter groundwater levels were lagged one or two years, peaks would align.

Precipitation was accounted for in this study as previous year’s totals and spring totals for the current year. These two representations do not have clear patterns Figure 3B. An increase in spring precipitation is thought to cause increased humidity and surface water accumulation, supporting second and third generation mosquitoes by increasing oviposition sites later in the year (Mahmood and Crans 1998; Andreadis et al. 2012). Previous year’s total rainfall and can be

associated with groundwater levels by increasing water tables for the current year, as suggested by previous research (Day and Shaman 2008). This, like spring rain totals, can produce and increase in the current year's available oviposition sites (Mahmood and Crans 1998; Day and Shaman 2008; Andreadis et al. 2012).

Other research has found that in the long-term, mosquito abundance does not seem to correlate with increased precipitation and only short-term examination has good predictability, corroborating Analysis I's results (Chase and Knight 2003). However, examining year-to-year trends can be useful if focusing on extreme events. Years following a drought year in both field and experimental investigation showed a marked increase in mosquito abundance for some species in semi-permanent wetland areas (semi-permanent meaning drying is not a regular event each year) (Chase and Knight 2003).

In these analyses we explored the potential link between climate variability and trends in *Cs. melanura* abundance over time as hypothesized in other research (Day and Shaman 2008; Andreadis et al. 2012; Armstrong and Andreadis 2013). However, at this time, within the scope of the research scale and variables used, a case for increasing vector abundance due to warmer temperatures and increased rainfall cannot be made. Future investigations might consider using spring/summer temperatures, and winter/fall precipitation in the linear regression analysis. Also a monthly resolution may increase the opportunity to address trend and patterns in the data both graphically and in time-series linear regression models.

### *Strengths & Limitations*

We chose to use ARIMA models in Analysis I because they impose stationarity in the dependent variables in order to more accurately assess trends in the explanatory variables. At the inter-annual scale, this model has a limited capacity to describe vector abundance in relation to

meteorological factors. In the future, methods and results provided by Jian et al. (2014) can be used to better select measurement periodicity needed, using Discrete Fourier Transform (DFT) determinations before modeling. In Analysis II a logistic regression was used to develop a spatial predictive model for *Cs. melanura* abundance to achieve a dichotomous outcome and to also handle the skewness in the data count by many low and zero mosquito counts. Other “count” modeling such as negative binomial regression and zero-inflated binomial regression could have more accurately dealt with over dispersion and biasing of parameter estimates. However, these methods do not as easily deal with spatial autocorrelation and its ramifications had they arose (Cheung 2002).

Lag selection in the temporal section, Analysis I, was an important component not fully developed. Too few lags may leave out relevant information, while too many may increase the error in the overall model forecasting (Ivanov and Kilian 2001). The failure to create a linear regression model due to undefined trend may be the mis-modeling of the lag structure in the data used. Also, the limited number of time-series components to examine for trend because of the incorporation of groundwater level data (16) may too harshly constrain the overall data structure.

Though LULC showed very little change from 2006 to 2011 products, there may still be inconsistencies that add uncertainty to the *Cs. melanura* abundance probability model created in Analysis II (Jin et al. 2013). Future modeling using this method should quantify classification changes between 1997 and 2006, as well as 2006 and 2011, as that is the expanse of the CDC light trap data. The novel approach of improving the size of buffers for each predictor variable, rather than a fixed sample size (usually 1 km) performed in other studies (Beck et al. 1997; Moncayo et al. 2000) was maintained. It was important to expand pixel values to reflect the surrounding habitat’s influence on *Cs. melanura* preference due to longer than average flight

range of 4-8 km (Howard et al. 1989; Estep et al. 2010). Due to *Cs. melanura*'s longer average flight range, increasing buffer sizes to even over 1 km may have been ideal for capturing predictor variable influence on habitat preference.

Lack of significance in groundwater data could have been the result of the original well data at insufficient resolution. Only 10 wells for the entire state met specifications and were located near trap site locations. Ideally, active well sites would occur at standardized distances and at high enough numbers that only a handful of site location would share the same groundwater level data. Stream flow data, a surface water and base water measurement, may capture the same habitat preference and be a more appropriate water level measurement, especially near runoff location near the highlands of CT where there is a greater chance of surface water collection (Day and Shaman 2008; Shaman et al. 2002). Another consideration is snow. Precipitation and snowmelt are both contributors to groundwater level measurements, though their impact on mosquito abundance differs significantly. Snow and snow-pack are more likely to increase development time in larval stage considering low to freezing temperatures. All of these could be considered in the future investigations.

### *Recommendations*

Quantifying standing water for estimating oviposition sites:

- Consider components of groundwater and contributions to surface water
  - Snowmelt vs. precipitation and times of year
- Wetland surveys (Chase and Knight 2003)
- Incorporation a dynamic hydrological model to produce truly saturated areas (surface wetness) and the shallow groundwater levels that supports them using (Shaman et al. 2002):

- Evaporation control factors
- Soil properties
- Vegetation
- Prior wetness of the watershed
- Topography
- Models at daily or hourly levels that capture oscillations

Other Variable Selections:

- Use a response function for initial selection and screening of seasons and variable like what is often used in tree-ring research for determining instrumental climatic variables that are associated with tree-ring width variability
  - Spring mildness in addition to winter mildness
  - Hard freezes and/or winter minimums and maximum
  - Total accumulations for spring and fall precipitation, minimums and maximums
  - Humidity
- Consider time of day in weather variables. Morning temperatures may be important when looking at virus transmission from bites (Burkett-Cadena et al. 2015)
  - Evening/ dusk temperatures
- Look at number of days of above or below average temperate and precipitation rather than just averaging or maximum and minimums. So accumulation of days of a certain type

Additional Analyses:

- Habitat preferences are likely static components that would not change over time. Instead it would be ideal to consider meteorological data in spatial context to discern inter-annual changes in climate leading changes in abundance patterns
- Consider space-time trends in a GIS environment
  - Probability model over each year
  - LISA each year →LISTA-VIZ
  - Spatial mean of abundance for each week, month or year for *Cs. melanura* abundance

Time resolution of data:

- Look at less aggregated time steps days up to month to month or season to season for time-series analysis
- Use an approach that takes into account the episodic nature of EEEV positivity and biology of *Cs. melanura* (5 days - 2 weeks) (Jian et al. 2014)

Other Considerations:

- El Niño versus La Niña years
- Include sentinel surveillance, examine blood meal type, or look at horse positivity

*Conclusion*

This study was important, as both temporal and spatial patterns of vector abundance are key components of understanding the EEEV transmission cycle and directing surveillance and control strategies to reduce EEE disease burden. Analysis II was able to examine spatial patterns of *Cs. melanura* and generate an efficient habitat suitability map, representing the probability of increased mosquito numbers at a high resolution across Connecticut. Creating continuous surface maps can improve upon point data surveillance and response efforts for EEEV with the addition of only a few outside layers for modeling. Analysis I was not successful at determining time

trends at the year resolution needed for groundwater level incorporation. However, it gives direction for future investigation of climatic factors, and their time scale, that can affect *Cs. melanura* abundance.

Future studies should consider maintaining more of the original data's temporal resolution in order to examine both intra-annual and inter-annual trends associated with climate. Also, integral factors for vector abundance response should be examined as extreme events or consecutive days of a certain characteristic(s), as well as those biologically relevant to the vector's life cycle. For example, consecutive days in winter or spring above a certain threshold or average measurement for temperature or precipitation may be more useful than using average temperatures, minimum or maximum temperatures, or simple accumulation of precipitation. Capturing variables or creating a factor that represents aquatic habitat for *Cs. melanura* is also important because of its preference for this habitat is invariant, like the palustrine areas explored in this analysis. A mosquito-borne viral disease necessitates the mosquito's presence and in adequate numbers. It follows that if we effectively quantify aquatic habitats at high resolution, we can work towards accurate and timely vector-disease system modeling to mitigate outbreaks and enhance surveillance efforts that are already in place.

## References

- Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;19(6):716-723.
- Andreadis TG, Anderson JF, Vossbrinck CR et al. Epidemiology of West Nile virus in Connecticut: a five-year analysis of mosquito data, 1990-2003. *Vector-Borne and Zoonotic Disease* 2004;4:360-378.
- Andreadis TG, Shepard JJ, and Thomas MC. Field observations on the overwintering ecology of *Culiseta melanura* in the northeastern USA. *Journal of the American Mosquito Control Association* 2012;28(4):286-91.
- Armstrong PM and Andreadis TG. Eastern equine encephalitis virus – Old enemy, new threat. *New England Journal of Medicine* 2013;368(18):1670-3
- Armstrong PM and Andreadis TG. Eastern Equine Encephalitis virus in mosquitoes and their role as bridge vectors. *Emerging Infectious Diseases* 2010;16(12):1869-74.
- Armstrong PM, Andreadis TG, Anderson JF, Stull JW et al. Tracking Eastern Equine Encephalitis virus perpetuation in the northeastern United States by phylogenetic analysis. *American Journal of Tropical Medicine and Hygiene* 2008; 79(2):291-296.
- Beck LR, Rodriguez MH, Dister SW, et al. Assessment of a remote-sensing as a landscape epidemiologic toll to identify villages at high risk for malaria transmission. *American Journal of Tropic Medicine and Hygiene* 1997;51:271-280.
- Breusch TS, and Pagan AR. A simple test for heteroskedasticity and random coefficient variation. *Econometrica* 1979;47(5):1287-1294.
- Burkett-Cadena ND, Bingham AM, Hunt B, Morse G, et al. Ecology of *Culiseta melanura* and other mosquitoes (Diptera:Culicidae) from Walton County, FL, during winter period 2013-2014. *Journal of Medial Entomology* 2015; 52(5);1074-1082.
- Center for Disease Control and Prevention (CDC) Eastern Equine Encephalitis: Epidemiology and geographic distribution 2015. Retrieved from [www.cdc.gov/EasternEquineEncephalitis/tech/epi.html](http://www.cdc.gov/EasternEquineEncephalitis/tech/epi.html)
- CDC and United States Geological Survey (USGS). *Dynamic Map Viewer*. West Nile Virus human: Final 2012 data. 2012. Retrieved from [diseasemaps.usgs.gov/mapviewer](http://diseasemaps.usgs.gov/mapviewer)
- Center for Food Security and Public Health & Institute for International Cooperation in Animal Biologics (CFSPH-IICAP). Eastern, Western, and Venezuelan Equine Encephalomyelitis. Iowa State University 2008-2015. Retrieved from [www.cfsph.iastate.edu/Factsheets/pdfs/easter\\_wester\\_venezuelan\\_equine\\_encephalomyelitis.pdf](http://www.cfsph.iastate.edu/Factsheets/pdfs/easter_wester_venezuelan_equine_encephalomyelitis.pdf)
- Center for Land Use Education Research (CLEAR). University of Connecticut. Statewide: Land Cover 2012. Retrieved from <http://clear.uconn.edu/projects/landscape/statewide.htm>

- Chase JM, Knight TM. Drought-induced mosquito outbreaks in wetlands. *Ecology Letters* 2003;6:1017-1024.
- Cheung YB. Zero-inflated models for regression analysis of count data: a study of growth and development. *Statistics in Medicine* 2002;21:1461-1469.
- Cowardin LM, Carter V, Golet FC, and LaRoe, ET. *Classification of Wetlands and Deepwater Habitats of the United States*. 1979. Washington D.C.: United States Department of the Interior Fish and Wildlife Service. Biological Services program.
- Crans WJ. *Aedes sollicitans* as an epidemic vector of Eastern Equine Encephalitis in New Jersey. *Mosquito News* 1977;37:85-89.
- Day JF, and Shaman J. Using hydrologic conditions to forecast the risk of focal and epidemic arboviral transmission in peninsular Florida. *Journal of Medical Entomology* 2008;45(3):458-465.
- Deresiewicz RL, Thaler SJ, Hsu L, Zamani AA. Clinical and neuroradiographic manifestations of Eastern Equine Encephalitis. *New England Journal of Medicine* 1997;336:1867-74.
- Diuk-Wasser Ma, Brown HE, Andreadis TG, Fish D. Modeling the spatial distribution of mosquito vectors for West Nile Virus in Connecticut, USA. *Vector-borne and Zoonotic Diseases* 2006;6(3):283-95.
- ESRI (Environmental Systems Resource Institute). 2013. ArcMap 10.2. ESRI, Redlands, California.
- Estep LK, Burkett-Cadena ND, Hill GE, Unnasch RS, et al. Estimation of the dispersal distances of *Culex erraticus* in a focus of eastern equine encephalitis virus in the southeastern United States. *Journal of Medical Entomology* 2010;47:977-86.
- Freemaster RF, and Getting GA. Distribution of the vectors of equine encephalomyelitis in Massachusetts. *American Journal of Public Health* 1941;31:791-802.
- Freier JE. Eastern Equine Encephalomyelitis. *The Lancet* (British edition) 1993. 342:(8882):1281-82.
- Gibney KB, Robinson S, Mutebi J, Hoenig DE, et al. Eastern Equine Encephalitis: An emerging arboviral disease threat, Maine, 2009. *Vector-Borne and Zoonotic Diseases* 2011;11(6):637-639.
- Granger CWJ and Newbold P. Spurious regression in econometrics. *Journal of Econometrics* 1974;2:111-120.
- Horton R, Yohe G, Esterling W, Kates R, et al. 2014: Ch.16: Northeast. *Climate Change Impacts in the United States: The Third National Assessment*, J.M. Melillo, Terese (T.C.) Richmond, and G.W. Yohe, Eds., U.S. Global Change Research Program, 372-395.
- Hosmer DW and Hjort NL. Goodness-of-fit processes for logistic regression: simulation results. *Statistical Methods* 2002;21:2723-2738.

- Hosmer DW and Lemeshow S. *Applied Logistic Regression*. 1989. New York, NY: Wiley.
- Howard JJ, White DJ and Muller SL. Mark-recapture studies on the *Culiseta* (Diptera:Culicidae) vectors of the eastern equine encephalitis virus. *Journal of Medical Entomology* 1989;26:190-9.
- Ivanov V and Kilian L. *A Practitioner's Guide to Lag-Order Selection for Vector Autoregressives*.2001. London, Centre for Economic Poly Research. Retrieved from [http://www.cepr.org/active/publications/discussion\\_papers/dp.php?dpno=2685](http://www.cepr.org/active/publications/discussion_papers/dp.php?dpno=2685)
- Jian Y, Silvestri S, Brown J, Hickman R., et al. The temporal spectrum of adult mosquito population fluctuations: Conceptual and modeling implications. *PLoS ONE* 2014;9(12): e114301 doi:10.1371/journal.pone.0114301.
- Jin S, Yang L, Danielson P, Homer C, Fry J, and Xia G. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sensing of the Environment* 2013;132: 159-175.
- Kunkel KE, Andsager K, and Easterling DR. Long-term trends in extreme precipitation events over the conterminous United States and Canada. *Journal of Climate* 1999;12:2515-27.
- Letson GW, Bailey RE, Pearson J, and Tsai TF. Eastern Equine Encephalitis (EEE): A description of the 1989 outbreak, recent epidemiologic trends, and the association of rainfall with EEE occurrence. *American Journal of Tropical Medicine and Hygiene* 1993;49(6):677-685.
- Luber G, Knowlton K, Balbus J, Frumkin H, et al. 2014: Ch.9: Human Health. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J.M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 220-356.
- Magnarelli LA. Host feeding patterns of Connecticut mosquitoes (Diptera:Culicidae). *The American Journal of Tropical Medicine and Hygiene* 1977;26(3);547-552.
- Mahmood F and Crans WJ. Effect of temperature on the development of *Culiseta melanura* (Diptera: Culicidae) and its impact on the amplification of Eastern Equine Encephalomyelitis virus in birds. *Journal of Medical Entomology* 1998;35(6):1007-12.
- Metzler KJ, and Tiner RW. *Wetlands of Connecticut*. National Wetlands Inventory. State Geological and Natural History Survey of Connecticut. Hartford, CT; 1992. Retrieved from [http://www.ct.gov/deep/lib/deep/water\\_inland/wetlands/wetlands\\_of\\_ct.pdf](http://www.ct.gov/deep/lib/deep/water_inland/wetlands/wetlands_of_ct.pdf)
- Molaei G, Andreadis TG, Armstrong PM, Thomas MC, et al. Vector-host interactions and epizootiology of Eastern Equine Encephalitis Virus in Massachusetts. *Vector-Borne and Zoonotic Diseases* 2013;13;1-13.

- Molaei G, Armstrong PM, Graham AC, Kramer LD, Andreadis TG. Insights into the recent emergence and expansion of Eastern Equine Encephalitis virus in a new focus in the Northern New England USA. *Parasites and Vectors* 2015;8:516-531.
- Moncayo AC, Edman JD, Finn JT. Application of geographic information technology in determining risk of Eastern Equine Encephalomyelitis virus transmission. *Journal of American Mosquito Control Association* 1991;7:608-615.
- Muul I, Johnson BK, and Harrison BA. Ecological studies of *Culiseta melanura* (Diptera:Culicidae) in relation to Eastern and Western Equine Encephalomyelitis virus on the eastern shore of Maryland. *Journal of Medical Entomology* 1975;11(6):739-748.
- Nadaular M and Urban AW. Epidemiology: Eastern Equine Encephalitis, September 2014. Retrieved from [emedicine.medscape.com/article/233442-overview#a6](http://emedicine.medscape.com/article/233442-overview#a6)
- Nasci RS, and Edman JD. Blood-feeding patterns of *Culiseta melanura* (Diptera:Culicidae) and associated sylvan mosquitoes in southeastern Massachusetts Eastern Equine Encephalitis enzootic foci. *Journal of Medical Entomology* 1981;18(6):493-500.
- O Horse! International Horse Directory. United States Horse Population Map 2015. Retrieved from [www.ohorse.com/ocal/north-america/united-states/map/](http://www.ohorse.com/ocal/north-america/united-states/map/)
- O'Sullivan D and Unwin DJ. *Geographic Information Analysis*. 2<sup>nd</sup> ed. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2010.
- Pfeiffer D, Robinson T, Stevenson M, Stevens K, et al. *Spatial Analysis in Epidemiology*. Oxford, UK: Oxford University Press; 2012.
- Reimann CA, Hayes EB, DiGuseppi C, Hoffman R, et al. Epidemiology of neuroinvasive arboviral disease in the United States, 1999-2007. *American Journal of Tropical Medicine and Hygiene* 2008;76(6):974-9.
- Shaman J, Stieglitz M, Stark C, Le Blancq SL, et al. Using a dynamic hydrology model to predict mosquito abundances in flood and swamp water. *Emerging Infectious Diseases* 2002;8(1):8-13.
- Shumway RH and Stoffer DS. 2006. *Time Series Analysis and Its Applications: With R Examples*. New York, NY; Springer.
- Sokal RR and Rohlf FJ. *Biometry: The Principles and Practice of Statistics in Biological Research*. San Francisco, CA: W.H. Freeman; 1989.
- StataCorp.2013. Stata Statistical Software: Release13. College Station, TX: StataCorp LP.
- United States Census Bureau (USCB). Population density estimates 2010. Retrieved from [www2.census.gov/geo/pdgs/maps-data/maps/thematic/us\\_popdensity\\_2010.pdf](http://www2.census.gov/geo/pdgs/maps-data/maps/thematic/us_popdensity_2010.pdf)
- United States Department of Agriculture (USDA). Animal and Plant Health Service. 2012 summary of eastern equine encephalitis equine cases in the United States 2013.
- USDA. Epidemiology and Ecology of Eastern Equine Encephalomyelitis. April 2004.

- Vallari P, Spielman A, Komar N, McDowell M, et al. The economic burden imposed by a residual case of Eastern Equine Encephalitis. *American Journal of Tropical Medicine and Hygiene* 1995;52:8-13.
- Vittenghoff E, Glidden DV, Shiboski, SC, and McCulloch CE. .2005. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. New York, NY: Springer Science+Buisness Media, Inc.
- Wood DM, Dang PT, and Ellis RA. Part 6: The mosquitoes of Canada. Diptera: Culicidae. *The insects and arachnids of Canada*. Biosystematics Research Institute, Ottawa. Publ. No. 1686:1-390.