University of South Florida

Digital Commons @ University of
South Florida

2020

Utilizing Ecological Niche Modelling to Predict Habitat Suitability of Eastern Equine Encephalitis in Florida

Claire Burch
University of Oklahoma

Rebecca Loraamm
University of Oklahoma

Thomas R. Unnasch
University of South Florida, tunnasch@health.usf.edu

Joni A. Downs
University of South Florida, downs@usf.edu

Utilizing ecological niche modelling to predict habitat suitability of eastern equine encephalitis in Florida

Claire Burch , Rebecca Loraamm , Thomas Unnasch & Joni Downs

Published online: 25 Feb 2020.

Submit your article to this journal ⬀

Article views: 465

View related articles ⬀

View Crossmark data ⬀

Taylor & Francis
Taylor & Francis Group

Utilizing ecological niche modelling to predict habitat suitability of eastern equine encephalitis in Florida

Claire Burch [iD][a], Rebecca Loraamm [iD][a], Thomas Unnasch[b] and Joni Downs[c]

[a]Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK, USA; [b]College of Public Health, University of South Florida, Tampa, FL, USA; [c]School of Geosciences, University of South Florida, Tampa, FL, USA

ABSTRACT

Eastern Equine Encephalitis virus (EEEV) is a virus found predominantly east of the Mississippi River in the United States that can be fatal to both equines and humans. The disease has previously been most prolific in states like Florida, but there has been an increase in the prevalence in other states further up north on the east coast of the United States in recent years. The purpose of this research is to use the ecological niche modelling program Maxent to model EEEV habitat suitability probability. This research utilized data of fatality incidence in equine hosts, versus sentinel chicken infection data, the spatial data traditionally utilized for mapping EEEV. This research produced a map of habitat suitability, which expanded on previous risk models by utilizing additional environmental factors. It confirmed areas of higher probability identified by previous models but identified more narrow areas of higher probability as well. This model adds to the literature applying ecological modelling techniques to spatial epidemiology. It highlights spaces that represent the culmination of environmental factors for the transmission of EEEV. Considering these environmental factors identified can assist in identifying places where there is a higher risk of EEEV as new cases begin to appear.

1. Introduction

Eastern Equine Encephalitis virus (EEEV) is a disease that occurs east of the Mississippi River within the United States, which typically impacts horses but can infect humans as well. Both mosquitoes and birds are carriers for the virus; however, birds are considered 'dead-end' hosts as they cannot transmit the disease further (CDC 2016). The inclusion of 'equine' in this virus' name is misleading, as the cycle of transmission to humans and equines occurs via spillover from birds, which are the major enzootic hosts of the virus. In birds, the major vector transmitting EEEV is *Culiseta melanura*; when it is transmitted to the so-called bridge vector mosquito species including *Aedes* spp., *Coquillettidia* spp., and *Culex* spp, it can be further transmitted to equines and humans (Vander Kelen et al. 2012a). The vectors for EEEV in both the enzootic and epidemic cycles have been found to be associated with forested wetlands and swamplands as well as tree plantations, and it has been hypothesized that these habitats not only have positive environmental conditions for *Cs. melanura* but also for the avian reservoirs and the bridge vectors (such as *Aedes spp.*) (Vander Kelen et al. 2012a; Jacob et al. 2010; Scott and Weaver 1989). When a human contracts EEE, it is often fatal as there is no specific treatment for the disease – according to the Florida Department of Health (FDOH, 2019), about one-half of EEEV cases die of the disease. In horses, the infection is thought to be almost uniformly fatal in unvaccinated animals (FDOH, 2019). When the viral infection is detected in the equine host, it suggests that there is a high intensity of infection in the local bird population, suggesting that there is a higher risk of it also being transmitted to humans. As climates continue to warm and weather patterns change, these changes could lead to further proliferation of the mosquito vectors and increased risk of EEEV infection (Vander Kelen et al. 2012a). In addition, as urbanization continues to spread to undeveloped land, it is important to understand which natural environments, conditions, or assemblages promote transmission in order to better prepare medical professionals for potential cases. Currently, in 2019, almost 30 cases of human infection have been reported east of the Mississippi, in Michigan, Massachusetts, Rhode Island, Tennessee, and North Carolina. This is contrasted with only 6 cases of human infection reported in 2018, in Florida, Georgia, Pennsylvania, and Michigan (CDC ArboNet 2018, 2019). It is therefore critical to understand the contexts which promote viral spread.

CONTACT Claire Burch ✉ burchcm@ou.edu ⬚ Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK, USA

The state of Florida is one of the most extensively studied areas for EEEV as well as the state with the largest proportion of cases every year. Compared to other states, it has the most reported human cases – one to two per year – and horse cases, having roughly 70 reported per year (Bingham et al. 2014). Typically, in other states along the east coast and into the northeast, the EEEV viral cycle becomes dormant during the winter months. For example, in Maryland, EEEV has a distinct transmission season that begins with transmission detection in July and stretching until November or December. In Florida, however, transmission can occur year-round, with the peak season occurring from May to August (Scott and Weaver 1989; Bigler et al. 1976). Within Florida, the virus circulates most intensively in the Florida panhandle and central Florida versus the southern half of the state (Vander Kelen et al. 2012b). Research has proposed that Florida, because of its year-long transmission period, could potentially be a reservoir for the disease within the US – birds contract EEEV before migrating and then proceed on to other areas within the country (Tan et al. 2018). Understanding the spatial distribution and risk potential of EEEV could assist in mitigating not only transmission risk within the state of Florida but also the risk of continual spread to the rest of the country. Conditions found within Florida where the virus can be transmitted to equines and humans must be identified, towards mitigating transmission within the state. Identification of areas favourable to the transmission where cases have not yet been reported is also a primary concern for mitigation and planning.

Previous work has modelled EEEV distribution and perpetuation, typically utilizing information on mosquito vector location (Cupp et al. 2003; Bigler et al. 1976; Bingham et al. 2014; Jacob et al. 2010) as well as mapping locations or counts of potential bird vectors (Jacob et al. 2010; Bingham et al. 2014). Most of the modelling performed has been completed utilizing spatial distribution modelling such as spatial linear prediction models predicting abundance (Jacob et al. 2010), density-based spatial clustering to identify clusters of habitats associated with the disease (Vander Kelen et al. 2012b), and spatial risk modelling to identify areas of high risk for transmission (Vander Kelen et al. 2014). This research proposes an application for ecological niche modelling capable of creating a habitat suitability model for viral transmission. The areas identified as favourable to sustaining the EEEV viral cycle in this research can assist mitigation efforts employing methods from spatial epidemiology and risk modelling.

Studies employing ecological niche modelling in spatial epidemiology routinely use the Maxent software package, exposing a presence-only-based species distribution model (Phillips, Anderson, and Schapire 2006), to predict potential distribution for a multitude of diseases, including flu outbreak risk in California, US. (Belkhiria et al., 2018), Chagas disease in Brazil (Costa et al. 2014), and malaria in Zimbabwe (Gwitira et al. 2015). Maxent is a software package which operationalizes a modelling approach often used for estimating habitat suitability, species distribution and the probability of occurrence given presence-only data and environmental (climatic) variables in coordinate space (Elith et al. 2011; Phillips, Anderson, and Schapire 2006; Merow, Smith, and Silander 2013). Essentially, the Maxent suite takes a series of confirmed animal/disease case locations as input, along with a series of environmental predictors (bioclimatic co-variates thought to influence the rate of occurrence of animals/disease cases in the environment) recorded as continuous values assigned to a gridded space representing the study area. Maxent analyzes these data to estimate the probability of occurrence for the species in question, conditioned on the environmental predictors and a number of model parameters set by the user. This estimate relating species occurrence to bioclimatic conditions is then leveraged to suggest the probability of animal occurrence across the study area.

Conceptually, Maxent can be thought of as using an iterative approach to find a function minimizing the distance between two multivariate probability distributions. These are the distribution of bioclimatic variables at the species occurrence locations and the 'background' distribution of bioclimatic variables at locations where species were not observed (Elith et al. 2011). Maxent compares the set of environmental covariates at the presence locations against a random sample of the 'background' locations as part of this process.

The covariates within the Maxent modelling process (and other species distribution modelling methods) are the environmental factors affecting the suitability of habitat. Maxent begins by randomly generating a background sample data set from the landscape of interest utilizing the covariates inputted by the user (the Maxent default is 10,000 locations from the grid) – these locations do not take into account the present location data and could contain some of these data points. Comparing the ratio of the density of the covariates in presence locations against the density of these covariates across the landscape (informed by the background sample generation), Maxent's raw output provides the insight for determining which environmental factors are important. The model is also programmed to allow for generalization of results (Elith et al. 2011). If the sample utilized in Maxent represents a sample largely unimpacted by sampling bias, the output of the model

can be interpreted as an index of habitat suitability (Merow, Smith, and Silander 2013). Merow, Smith, and Silander (2013) recognize this interpretation can only be made if some sampling assumptions are ignored (when considering sampling bias, Maxent output is more aligned with the relative occurrence rate of a species), but they emphasize that interpreting output as habitat suitability can lead to further analysis based on results.

Because Maxent may not adequately compensate for sampling bias given presence-only data in the way presence-absence data would eliminate biases (because the bias may be present in both the presence and absence data points), it is reliant on collection of an extensive and comprehensive set of presence location data points. Also, the landscape (geographic area) being utilized should accurately portray the range of the species and exclude regions that have not been surveyed. Environmental factors that are understood to impact the species based on previous ecological research should be utilized; limiting the number of covariates to those that are known to have an impact allows for smoother and better-fit results. It is also important to consider the chance of observing a species in areas of interest – the Maxent default is set at a '50% chance of observing the species', but this setting may not be appropriate if working with data from a species rarely observed unless long-term, extensive surveying methods are implemented prior to modelling (Elith et al. 2011).

The purpose of this research is to apply Maxent to create a model of disease distribution for EEEV in the state of Florida. As mentioned, previous spatial epidemiology work involving EEEV has utilized GIS to map clusters of EEEV risk-based largely on vector or reservoir host data but did not create a predictive model of distribution based on incidence data. Vander Kelen et al. (2014) created a risk model utilizing risk variables associated with specific habitat land cover and land cover traits (i.e. habitat, proximity to wetlands, wetland composition, tree plantation proximity, and plantation composition). This research proposes utilizing bioclimatic factors in addition to land cover to create a more expansive risk model. Modelling utilizing bioclimatic variables and land cover may identify additional areas of risk not indicated by land cover characteristics alone.

Work performed by Vander Kelen et al. (2012b) utilized a novel approach in mapping spatial distribution of EEEV by focusing on incidence in a dead-end host – the horse – versus the vector or reservoir host. They proposed utilizing this data set because little is known about the habitat dynamics of the epizootic transmission areas, homing in on the areas where mammal infections occur. While there is evidence indicating suitable habitats for the arthropod vectors of the disease, infection incidence data allow us to model the portion of the

cycle of EEEV that results in potentially fatal infections of equines and humans. The incidence data of equine fatalities due to EEEV enable the creation of more effective risk modelling tools, as there is a multitude of bridge vectors that could infect equines and humans. Mapping the distribution of one or a few of these vectors is not an inclusive model, whereas the fatality incidence data provide information characterizing the potential location of the portion of the cycle with the highest risk to human populations. This research will expand on utilizing the equine fatality incidence data set with the objective of creating a potential distribution model of EEEV based on the incidence rate of the disease in horses in the state of Florida using bioclimatic variables, land cover data, and elevation models.

2. Study area and data

2.1. Study area

The state of Florida comprises the study area for this research (Figure 1). Transmission of EEEV occurs year-round in Florida, where transmission peaks between May and August (FDOH 2019). EEEV transmission and infection have presented in 64 of the 67 counties in Florida as well, a prevalence rate not observed in any other state where the infection is known to occur (Vander Kelen et al. 2012a). Counties within Florida also have active monitoring methods in place to track the virus as well as attempt to reduce disease prevalence. These activities represent millions of dollars in surveillance expenditure each year, a significant expenditure for the state (Jaing 2010). Since incidence data and environmental variables utilized in this research could be processed at a scope appropriate to understanding EEEV transmission potential over a large area, this research was performed at the state-level to provide a comprehensive model of EEEV transmission risk.

2.2. Incidence data

Disease incidence data used for this research represent confirmed (reported) equine EEE cases in Florida occurring from 2005–2010; these points can be seen in Figure 1. EEEV is a reportable disease, which means that doctors and laboratories must report cases to the CDC. Prior research efforts involving EEEV have often been based on presence of EEEV in mosquito vectors (Cupp et al. 2003) or on seroconversion in bird species (sentinel chickens are often used to track arboviruses in Florida and other southeastern states) (Vander Kelen et al. 2014). While tracking vector and reservoir hosts provide valuable models of the disease, incidence rates in dead-end hosts can assist in creating accurate risk assessment models for treatment and/or
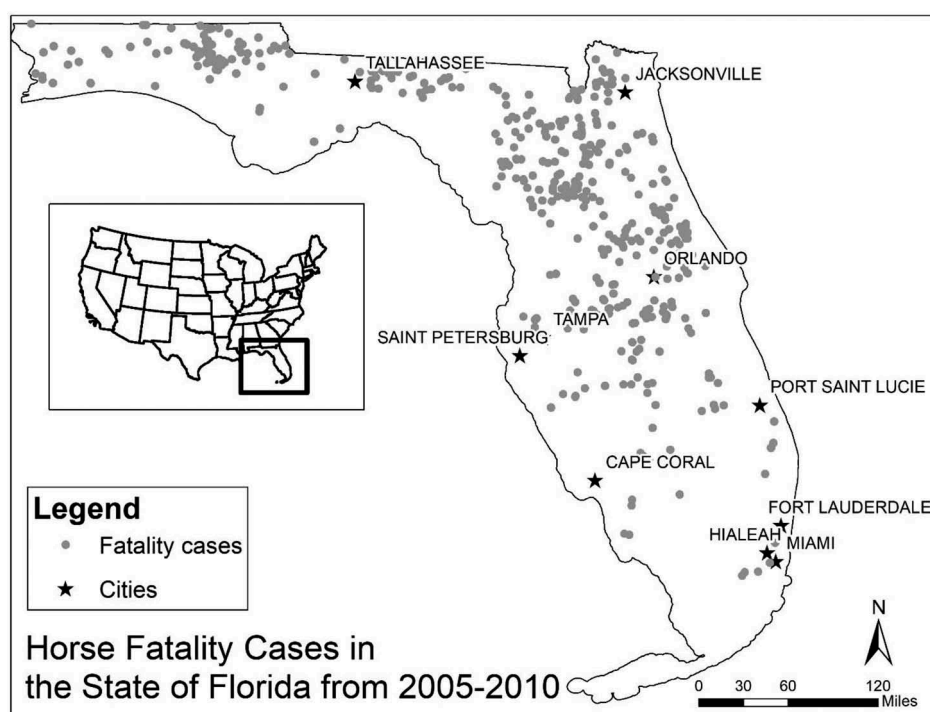
Figure 1. Map of Florida, United States of America, the area of interest for the study. The study area map includes the incidence points of EEE equine fatality from the Vander Kelen et al. (2012b) as well the top 10 most populated cities in the state of Florida (obtained from Florida Geographic Data Library, with data from the National Atlas of the United States (2007)).

vaccinations (in horses). The data set utilized in this research represents EEE fatal cases in horses and was developed previously by Vander Kelen et al. (2012b) towards mapping the spatial distribution of the disease in Florida. Vander Kelen et al. (2012b) prepared this data set by obtaining GPS location reports of equine EEE cases along with additional data from the Florida Department of Health. The final incidence data set used in modelling for this research contains 438 equine cases of EEE along with their respective GPS locations.

2.3. Environmental data

Elevation and land cover data were obtained through the United States Geological Survey (USGS). Bioclimatic variables capturing temperature and precipitation conditions were obtained from the National Climatic (NCDC) Data Center for weather stations in the state of Florida. The dataset obtained for this research is based on the 1980–2010 normals provided by NCDC, across a total of 19 bioclimatic covariates collected at point weather stations statewide. These bioclimatic variables were calculated to match parameters utilized in previous Maxent studies – it was suggested by the Maxent modelling package (Phillips, Anderson, and Schapire 2006) that bioclimatic variables identified by WorldClim (worldclim.org) be utilized as climate variables for modelling purposes. These variables were

utilized in other Maxent studies, either all 19 or a selected group determined by methods such as a principal component analysis (Belkhiria et al., 2018; Costa et al. 2014; González, Paz, and Ferro 2014; Gwitira et al. 2015; Rose and Wall 2011; Signorini et al. 2014).

The 19 bioclimatic variable values were assigned to point weather station locations, and a spatial interpolation method was then applied to generate input surfaces used for modelling with the Maxent maximum entropy modelling software package (Phillips, Anderson, and Schapire 2006). An inverse distance weighting (IDW) approach was utilized to create raster datasets representing a continuous interpolated surface for each of the 19 variables. The IDW operation assumes that distance is the only factor influencing the magnitude of a variable's interpolated value across space. With increasing distance from a known location, values interpolated at unknown locations will exhibit a predictable linear decay in magnitude. Utilizing IDW as the spatial interpolator of choice for this research provides a straightforward estimation of bioclimatic variables at locations where they were not otherwise available.

3. Methods

3.1. Model parameterization

Default settings were utilized when running the Maxent model, with 25% of the data set aside as a test data set,

chosen at random utilizing the 'random seed' option (González, Paz, and Ferro 2014; Rose and Wall 2011; Signorini et al. 2014). Setting aside data as a test data set allows the model to run additional statistical analyses for robustness evaluation of the final product (Phillips 2017). The area under the curve (AUC) cut-off value utilized was 70%. AUC is an independent measure of predictive power – it is interpreted as how well the model distinguishes presence versus background points (Merow, Smith, and Silander 2013) Models with an AUC of 70% or greater are viewed as viable for predictive purposes (Baldwin 2009), with an AUC above 50% meaning the model predicts better than random choice (Rose and Wall 2011). The cloglog output format was selected, which gives a range of probability estimates between 0 and 1 – it is the easiest result format to interpret and the most useful format for the purposes of this research.

All 19 bioclimatic variables were utilized in the initial run of the model. Jackknife output data then allowed for the elimination of variables which did not contribute to the model. This jackknife approach represents a type of iterative evaluation of bioclimatic variables by permutation from the set of bioclimatic variables. For each iteration, each variable's contribution to the variance explained by the model is calculated, and the best explanatory variables for the model are made clear as these produce the highest AUC values. Jackknife tests in Maxent produce a bar graph where the bar at the bottom represents the AUC of the model with all variables – comparing this bar to the bars that represent model runs without particular variables, the following bioclimatic variables were eliminated from the model: 5, 6, 8, and 13 (Table 1). Multiple iterations of this process were completed to test for changes in AUC until the AUC value obtained was higher than the initial first run involving all 19 variables. The jackknife approach was used in this research as an alternative to applying dimensionality

reduction techniques such as principal component analysis (PCA) as pre-processing for the input bioclimatic variables. Applying PCA to eliminate variables by transforming their information to their principal components resulted in a final model with a lower AUC value than utilizing the jackknife approach. Therefore, the effective set of variables determined by jackknife was used instead of the final model. Table 1 lists the bioclimatic variables utilized in the final iteration of the model.

3.2. Model validation under current EEE incidence distribution

The habitat suitability (probability of occurrence) surface for EEE disease incidence prepared in MaxEnt was validated using a randomization technique, applied identically for each of four bioclimatic regions (Panhandle, North, Central and South, respectively) in the state of Florida (Figure 2). For validation, the MaxEnt surface developed using observed disease locations (occurring from 2005 to 2010) and bioclimatic covariates (collected for 1980 to 2010) was overlaid and compared with new observed EEE disease locations occurring during 2011–2018, evaluating the predictive power of the prepared surface for future viral seasons spanning 8 years. For each bioclimatic region in the state, 99 spatially random realizations of the n observed disease events for 2011–2018 were prepared and statistics were collected summarizing the probability values found at occurrence surface raster cells co-located with these simulated disease events. Each random simulation was spatially constrained to the boundary of its host bioclimatic region. Histograms visualizing the average probability of occurrence at observed disease locations versus simulation were prepared to demonstrate a pseudo-p value for significance assessment. Further comparisons among observed versus simulated statistics summarizing EEE disease distribution in bioclimatic regions are then discussed.

4. Results

4.1. EEEV habitat suitability model

The area under the curve (AUC) for the training data was .770 and 0.758 for the test data. Both AUC values exceed a threshold value of 0.5, which corresponds to the AUC expected from a model generated at random. When the initial model was run, training data AUC was 0.779, but test data AUC was 0.744, so eliminating variables increased the overall predictive power of the model. Generally, high habitat suitability area for EEE appears to be concentrated on the panhandle as well as northern and central Florida, with only a small hotspot in southeastern Florida (Figure 3).
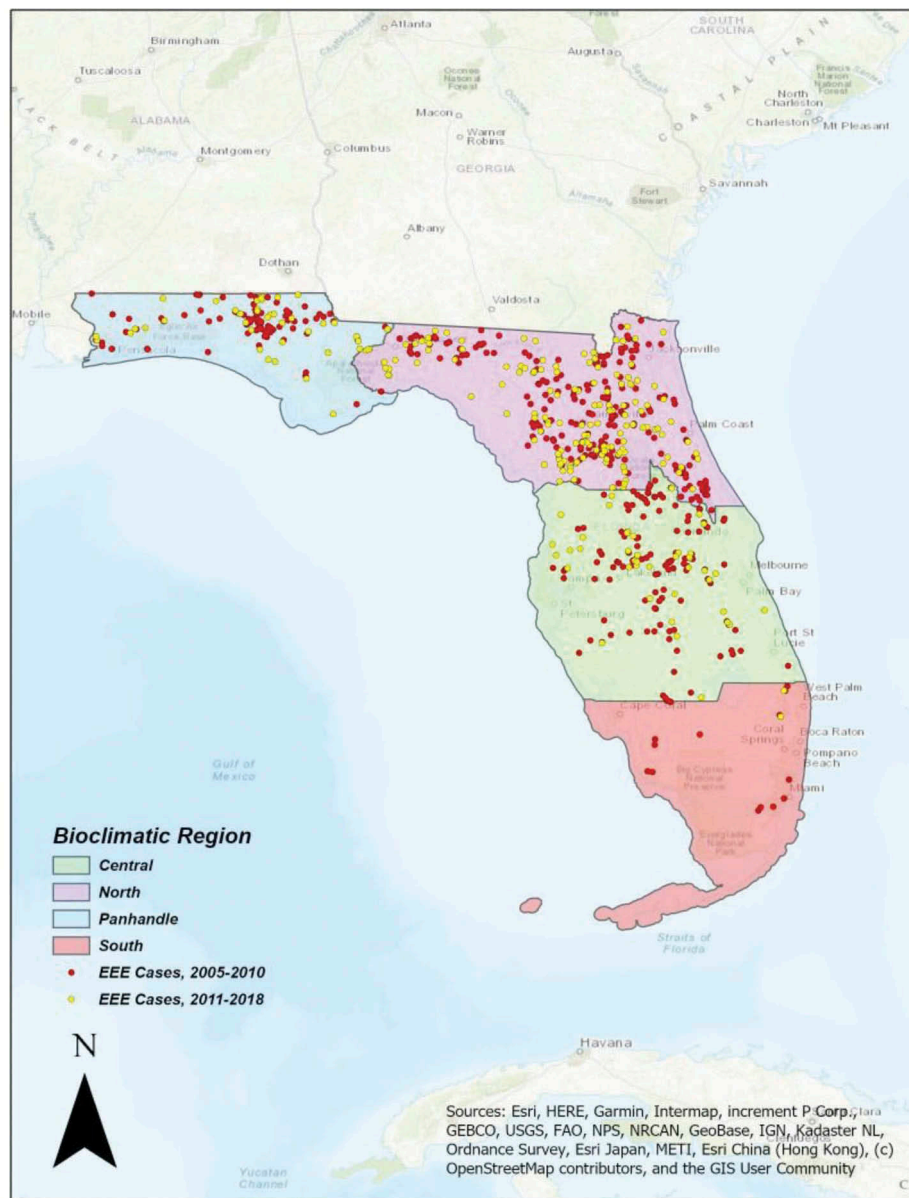
Table 1. Bioclimatic variables considered in the final model for this study.

1	Annual mean temperature
2	Mean diurnal range
3	Isothermality
4	Temperature seasonality
7	Temperature annual range
9	Mean temperature of driest quarter
10	Mean temperature of warmest quarter
11	Mean temperature of coldest quarter
12	Annual precipitation
14	Precipitation of driest month
15	Precipitation seasonality
16	Precipitation of wettest quarter
17	Precipitation of driest quarter
18	Precipitation of warmest quarter
19	Precipitation of coldest quarter

Figure 2. Source and validation EEE case locations in context with Florida's bioclimatic regions.

The Maxent result also provides information on which variables were most influential via per cent contribution values. These represent the variables the Maxent algorithms suggest contributed most to the model formulation (Phillips 2017). The highest per cent contributions to the model were elevation at 37.3%, temperature seasonality at 14.8%, mean diurnal temperature range at 7.2%, and precipitation of the coldest quarter at 6.9%. While these values show a contribution to the predictive model created, contributions should be interpreted with caution because of possible correlation among these variables in their effect on EEEV transmission.

Elevation (fl_dem) appeared to be the most important variable according to the jackknife test as well (Figure 4). When treated as the only explanatory variable in the model, a test AUC result of approximately 0.68 was still obtained using just this elevation covariate. Mean diurnal range (bio2) was also an important variable according to the jackknife test, with a test AUC of approximately 0.66, as well as temperature seasonality (bio4). Also, given the jackknife test, the mean temperature of the coldest quarter (bio11) had a test AUC of approximately 0.66. Precipitation of the coldest quarter (bio19) was more in line with other variables in the jackknife test.
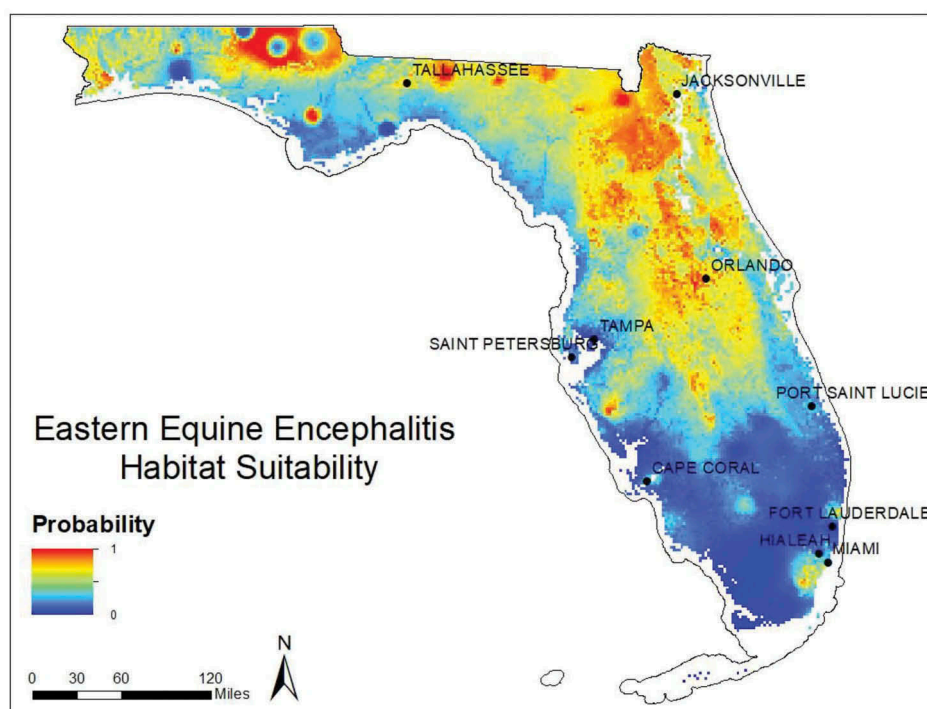
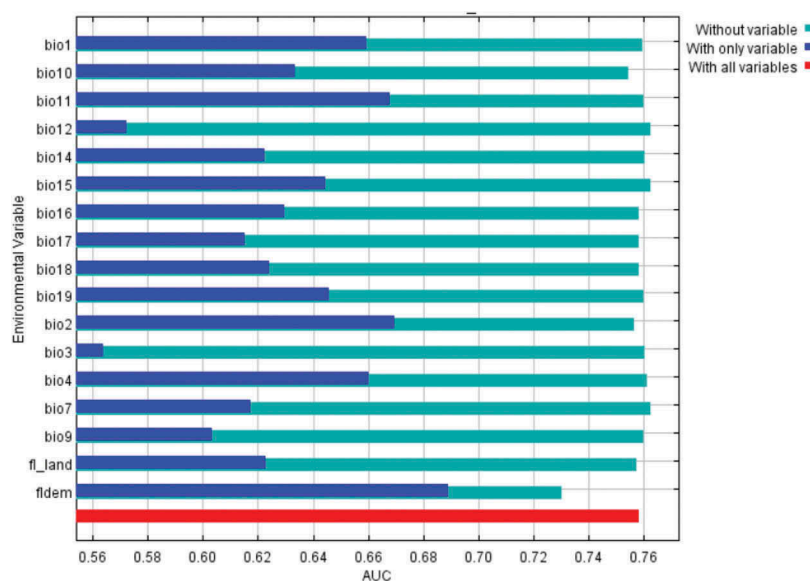**Figure 3.** EEE habitat suitability map, predicted by land cover, digital elevation, and bioclimatic variables.

Figure 4. Jackknife test results of final Maxent modelling output. See Table 1 for bioclimatic variables corresponding to numbered variables.

4.2. Results of model validation under 2011-2018 EEEV incidence distribution

Following application of the validation routine introduced in section 3.2, results suggest the MaxEnt probability of occurrence surface generated in consideration of 2005–2010 EEEV disease incidence has significant predictive power towards indicating risk for future year disease incidence (2011–2018) in the Panhandle ($n = 41$), North ($n = 156$) and Central Florida ($n = 44$) bioclimatic regions. Here, the MaxEnt surface is considered *predictive* provided the means of occurrence probabilities at randomly simulated disease patterns are lower than the mean occurrence probability at disease locations from the observed pattern. This dynamic establishes the MaxEnt surface is better at suggesting disease incidence than a random

surface, at a significance level roughly interpretable as *p = 0.01* given 99 simulations. Histograms depicting the mean probability of occurrence for the simulated reference distribution versus the observed pattern for the Panhandle (Figure 5), North Florida (Figure 6), Central Florida (Figure 7) and South Florida (Figure 8), respectively. For the South Florida region (where only *n = 2* disease cases were observed from 2011–2018), validation results do not suggest significantly higher observed occurrence probabilities from the MaxEnt surface are co-locating with (2011–2018) disease incidence locations. However, an alternative, additional simulation randomizing all available years of EEE disease incidence locations (2005–2018) for South Florida (*n = 18*) shown in Figure 9 suggests the MaxEnt surface is predictive when considering all available disease cases.

5. Discussion

5.1. Disease-case location as a proxy for presence-only data

Maxent provides a technique used heavily in ecology-enabling models of habitat suitability; this approach has great potential for utility in other fields. Utilizing Maxent, a habitat suitability predictor was produced for EEEV based on fatality data in horses collected from 2005–2010 in Florida, USA. This research utilized disease-case location data as a proxy for presence-only data – instead of tracking vector location, the model was created based on disease presence location. Rate of detection represented a substitution for survey methods used to gather data on species presence. Using the rate of detection in dead-end host versus tracing vector location could potentially eliminate the sampling bias issue heavily present in current presence-only data survey methods. Collecting dead-end host locations from veterinary professionals across the region of interest allow for more extensive surveying to create a presence-only data set versus relying on comprehensive (presence-absence) surveying methods. EEEV transmission and fatality are reported, and extensive records are kept of disease locations in this manner. This extensive record is kept by CDC and other entities allowing for a more comprehensive presence-only data set that has minimal skew due to sampling bias, because all known cases are reported. If utilizing disease-case data alleviates sampling bias (or at least produces data with less bias), then the qualms in interpreting the model raised by Merow, Smith, and Silander (2013) can be solved as the sampling bias prevents individuals from easily interpreting Maxent output as habitat suitability. Given less biased presence data,
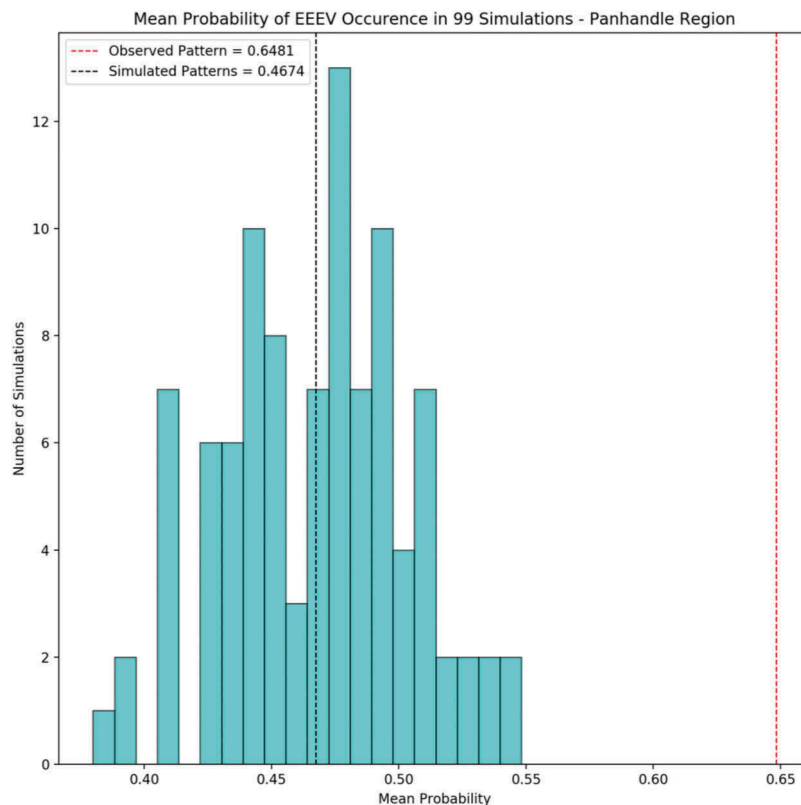

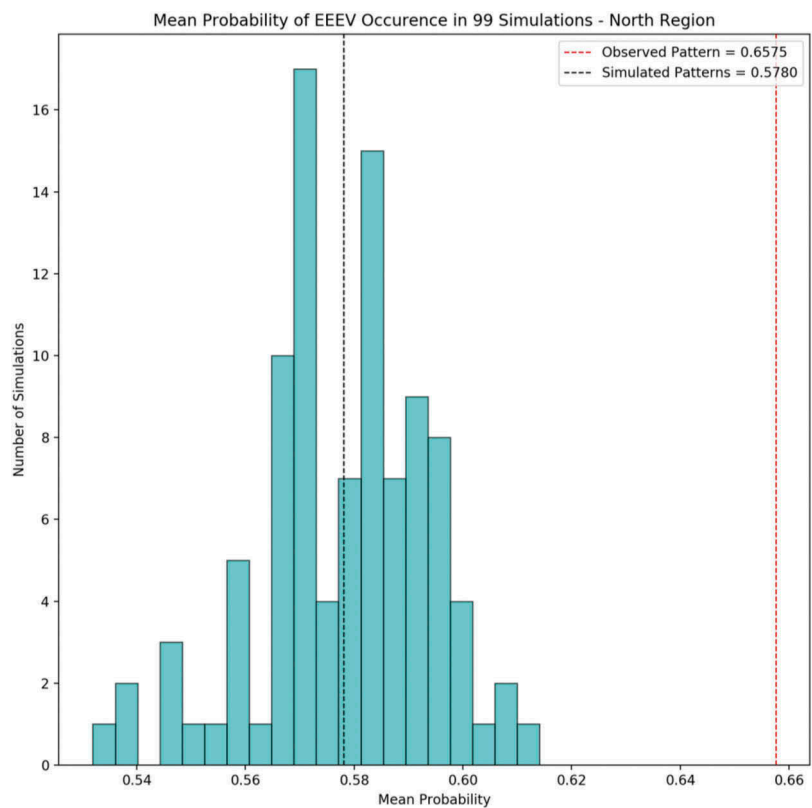**Figure 5.** Panhandle region validation result.
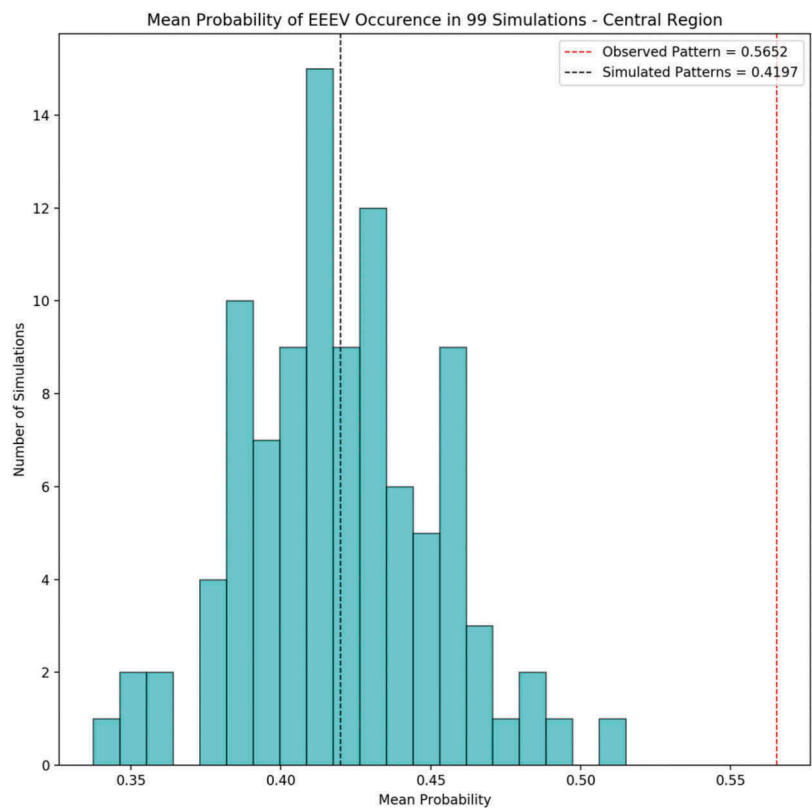
Figure 6. North Florida region validation result.

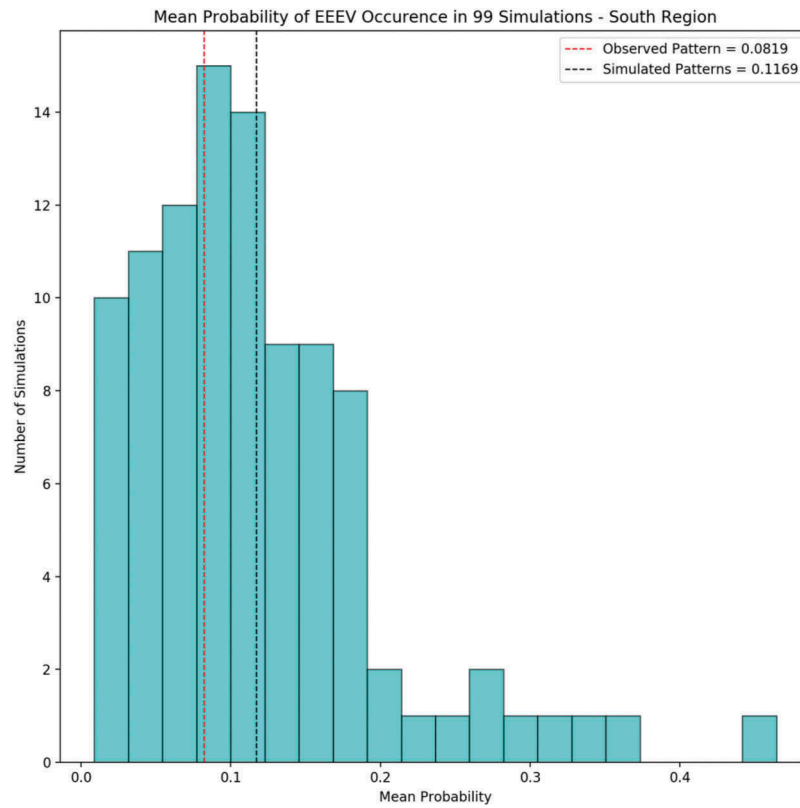**Figure 7.** Central Florida region validation result.

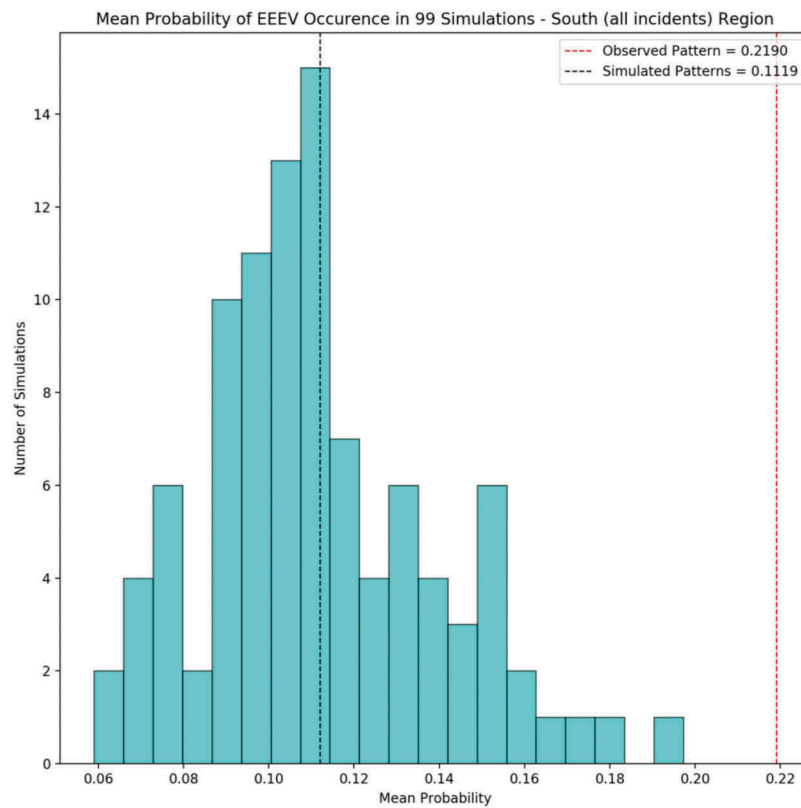**Figure 8.** South Florida region validation result.

Figure 9. South Florida region alternative validation result.

the model output may more accurately translate to relative occurrence rates. Decreasing sampling bias can create a model more robust to the assumptions associated with habitat suitability modelling overall. In addressing the concern in setting accurate parameters for the probability of observing a species, this becomes less of an issue since surveying for disease cases (in horses, these animals are large, usually domesticated and easy to locate) is easier than surveying for the presence of vectors (mosquitoes, which are small) in our case. EEEV presents obvious symptoms, which can be easily identified and reported in infected horses. While EEEV is a reportable disease, this data set is still limited; however, as it relies on individuals to report cases of the virus in their horses if they are not seen by a veterinarian. In addition, this case count data only reports equine fatalities. If horses are vaccinated, the disease is often not fatal (FDOH 2019) and that case point would, therefore, be excluded from this data set as it may not ever be evident that the equine was infected.

While disease-case location data have the potential to allow for accurate Maxent modelling in spatial epidemiology, a few questions arise concerning the collection and accuracy of the input data set. Specific to the case location data set used for this research, important questions surround the meaning of disease case spatial locations and their spatial accuracies. These locations reported could be the location in which the equine first presented with the symptoms of the disease but could also be the location of fatality or post-mortem examination. While disease-case location data could help to overcome the sampling bias issue prevalent in presence-only data, it is important to standardize spatial data collection to determine at which point in the disease life-cycle is the most appropriate to include in modelling suitability for future disease proliferation.

5.2. Model results

Based on the model produced, it appears that given the bioclimatic and environmental variables used (see Table 1), areas promoting disease transmission occur most often in the northern regions of Florida. This finding aligns with results from other studies of EEEV transmission, which found highest transmission clusters in the northern regions of Florida (Bigler et al. 1976; Heberlein-Larson et al. 2019; Vander Kelen et al. 2014, 2012a, 2012b).

According to Vander Kelen et al. (2012a), however, Walton County has one of the highest rates of transmission – on the Maxent model, Walton County is on the outskirts of the hotspots in the panhandle, with Holmes, Washington, and Jackson counties containing most of

the high habitat suitability values. Vander Kelen et al. (2012a) utilized sentinel chickens and rate of infection as a proxy for transmission, and the difference in results could thus reflect differences in detection of EEEV activity using horse cases versus chicken sentinel data. Bigler et al. (1976) identify two clusters of counties with mosquito pools often positive for EEEV – Leon-Jefferson-Madison and Polk-Orange (both clusters located in northern and central Florida) – and while these two clusters show higher probabilities of suitability, they do not contain hotspots comparable to central Florida.

While previous studies focused on utilizing land cover land use data as the basis for modelling (Vander Kelen et al. 2012b, 2014) land cover was only one factor considered within this model and was found to contribute relatively little information (4.4%) to the model compared to other variables. This model identified elevation as the major contributor to habitat suitability for EEEV, something not previously considered in EEEV modelling studies. Elevation could relate to hydrologic features, as these features are often found to play a significant role in EEEV transmission (Jacob et al. 2010). Vander Kelen et al.'s (2014) risk model using land cover-related risk variables showed a higher risk index stretching further into the panhandle than the final Maxent model presented in this research. While there seems to be some overlap in the areas of the high-risk index and high habitat suitability in the northern part of the state, there are differences between the two indices with respect to the panhandle. Vander Kelen et al. (2014) show the risk of transmission stretching into southern Florida, whereas the model produced in this study does not have high habitat suitability probability further south. It is important to note that Vander Kelen et al. (2014) used land cover data from the Florida Department of Environmental Protection, whereas this study utilized land cover data from the National Land Cover Database – the FL DEP data set provided more specific land cover identification classes. While the classification schemes differed, the general ecosystems presented should have remained fairly similar. Based on this comparison, this Maxent model presented in this research provides a more specific indication of risk than models created in 2014, which also utilized the equine fatality data. This could indicate that factors other than land cover play an important role in the transmission of the disease. Vander Kelen et al. (2014) also discuss the decrease in accuracy of model performance due to the lower incidence rates in the south – our model seems to have performed better in that regard as there is not as high suitability probability in the southern reaches of the state. However, while the modelling of lower risk areas in South Florida appears

consistent with an expectation for lower risk, validation for the available disease incidents in South Florida suggests the surface is not effective at predicting future disease incidence risk. The model presented in this research may have been able to predict presence more accurately due to the availability of more variables to compute probability with, versus prior work. The Maxent model produced in this study provides a more inclusive model that considered variables in addition to land cover.

Measures of precipitation and temperature were also found to contribute to the habitat suitability model, another factor often cited in studies on EEEV ecology dynamics (Vander Kelen et al. 2012a, 2012b; Jacob et al. 2010; Scott and Weaver 1989). Consideration of precipitation during the coldest quarter (as identified by per cent contribution) and mean temperature of the coldest quarter (as identified by the jackknife test) appear to contradict the idea of transmission peak being in the summer as coldest quarter values for climatic data were the last 4 months of the year, i.e. winter. The findings of this study do align, however, with Bingham et al. (2014), who evaluated winter transmission patterns of EEEV in Florida and found that, overall, vector populations were abundant during the winter months and readily feeding. While the data utilized in this paper's research was related to dead-end host fatality, there appears to be an alignment between the habitat suitability predicted by the model and the findings of high activity in the winter months by Bingham et al. (2014). This could indicate that EEEV activity in the peak season is partially determined by what the prior winter conditions were. If winter conditions are more favourable, the virus may get a 'head start' coming into peak season. The implications of this could be that more management is needed prior to peak season, and that management efforts should be focused on controlling the vector in the winter months to prevent proliferation in the following summer.

It was also of interest to examine the values of environmental variables in hotspots identified by the Maxent model to compare to prior results of habitat modelling for EEEV. Looking at temperature seasonality, areas of higher habitat suitability tended to have higher values for seasonality and were overall clustered in the panhandle and the north. It appears that a higher variation in temperature throughout the year favoured EEEV transmission. High mean diurnal range seemed to negatively impact habitat suitability – per Figure 3, fragments of very low suitability within high suitability areas in the panhandle are areas having high diurnal ranges. Considering precipitation during the coldest quarter, higher precipitation values were associated with higher

suitability values. Examining the mean temperature of the coldest quarter, the northern parts of Florida tended to have lower mean temperatures. When comparing this variable discussion to previous studies, there is no previous work evaluating seasonality and temperature diurnal range – while temperature is often mentioned as a factor, it is not evaluated in-depth. Precipitation, on the other hand, is often cited as a factor in the propagation of EEEV as it can lead to inundation in areas and therefore create breeding grounds for mosquito vector species (Jacob et al. 2010; Vander Kelen et al. 2012b; Cupp et al. 2003).

The model hotspots, with respect to their land cover covariates, generally lie in areas of cultivated crops, hay, and pastureland, evergreen forest, and small areas of shrub/scrub. In the areas that do not have as high of a probability within northern Florida stretching into central Florida, the land cover is comparatively the same but with less cultivated cropland cover. Interestingly, the areas of 0 probability nestled in the hotspots in the panhandle are centred around bodies of water in that region. The land cover comparison here is largely consistent with Vander Kelen et al.'s (2012b) spatial epidemiology study utilizing the same EEEV horse cases data set, where clusters of incidence were associated with crop and pastureland. The land cover data set utilized for this research did not distinguish cropland areas by major crop, such as tree plantations, which Vander Kelen et al.'s (2012a, 2012b) studies found to be a significant area for EEEV activity. It is possible that part of the tree cover associated with this land cover dataset could be tree plantations and not natural growth. The cropland and pastureland results presented here do generally match with the studies mentioned above. In addition, there do not appear to be large concentrations of wetland land cover categories, which had often been associated with EEEV proliferation in the past. In line with recent studies, this model furthers the notion that wetland land cover is not the optimal place for EEEV proliferation. It is important to highlight again, however, that land cover contributed only 4.4% to the predictive power of the model, which may indicate that changes in environmental variables in other parts of the state to match environmental conditions in hotspots could increase risk without the land cover has changed. This is important to note, as previous models indicated a higher risk associated with tree plantation land cover (Heberlein-Larson et al. 2019; Vander Kelen et al. 2012a, 2012b), but this model indicates land cover is not a strong determinant.

Overall, the model created in this study further refines the risk index models created previously to predict EEEV transmission, as Maxent is better equipped to handle the

lack of data points in southern Florida. In addition, while risk modelling can be utilized to evaluate areas where there is a perceived risk of future transmission (Downs et al. 2019), Maxent has the capability to do this at a larger scale. The model aligns with previous studies that emphasize the Florida panhandle as a hotspot for EEEV transmission, and this hotspot location has now been verified by both sentinel chicken infection data and equine fatality data. Heberlein-Larson et al. (2019) emphasize in their study the theory that Florida represents a region that seeds the virus for transmission in other states in the Northeast – birds migrating through Florida to the Northeast may carry the virus with them, for example, as suggested by Bingham et al. (2014). The panhandle, according to the model, represents a series of natural spaces where conditions are optimal not only for spread to avian hosts to continue the epizootic cycle but to equines, which are dead-end hosts where the infection can be fatal. As indicated earlier, once the disease breaks the avian cycle and infects an equine host, there is a higher risk of infection in humans as well. The hotspot in the panhandle as well as the areas of higher suitability probability in northern Florida both represent areas of higher risk for human infection.

5.3. Limitations to the model

While Maxent allows for spatial epidemiology modelling that includes environmental variables, it is not an all-inclusive model. Other factors could be impacting the spread of EEEV, such as biological factors of vectors, reservoir hosts, and dead-end hosts, small-scale ecological system differences not captured by a broad range model and human factors such as pest control measures and strong vaccination regimes in certain areas that decrease equine fatality. There is a question of whether the fenced-in nature of the equine environment impacts proliferation of the disease – dense populations of equines in an area where they are limited in range, could encourage spread of the disease as the vector would have access to multiple hosts. While equines are dead-end hosts and cannot transmit, the large concentration of equines may lead to more cases present. The environmental factors, therefore, may not determine habitat suitability as intensely as the density of hosts would, as this model relies on case counts and their association with environmental factors; the spatial distribution of hosts is not explicitly considered as an influence on the result.

In addition, while Maxent can produce viable models with minimal data points and can, therefore, be useful when only a small amount of data is available to input for EEEV risk assessment in other regions, more extensive data could provide a stronger model. This efficacy is demonstrated in this research under a validation routine evaluating the predictive power of the model surface given new disease incident locations occurring after the years considered for model construction. Current efficacy notwithstanding, additional actionable information for future modelling efforts is needed. Another limitation lies in co-linearity between bioclimatic variables – this was addressed in this model using an iterative jackknifing process. However, even with proper identification of possible co-linearity between input variables, confounding effects associated with the use of similar explanatory variables can still occur. A more sophisticated evaluation of the input variables for co-linearity, spatial or temporal correlation before modelling would benefit this effort in future research. Alternative interpolation methods to Inverse Distance Weighting for the preparation of input variables from point weather station data may also be explored.

Additionally, evaluation of the model for specific regions of Florida could be compared with statewide results, as EEEV transmission is, of course, a process which does not respect political borders. While Florida has a strong base in research on EEEV, it is understudied in other regions where it is still a risk to human and equine populations (Cupp et al. 2003) and Maxent modelling could extend to these areas. Maxent modelling could prove to be a valuable tool for risk assessment and mitigation. Future work utilizing Maxent as a modelling tool could include creating models with vector or reservoir host locational data, as has been done in previous GIS-based modelling studies.

Lastly, the model presented is limited in terms of spatial resolution, as the spatial resolution of the resulting modelled surface can be no finer than the coarsest of the input covariates. The original data sets for the land cover data as well as the elevation models had relatively high-resolution values (30 m x 30 m and 1/3 arc seconds, respectively). In order to provide for corresponding climate data, however, an interpolation approach was necessary to estimate surfaces representing climate covariates from point bioclimatic observations. Per the coarsest spatial resolution among the covariates, the final resolution of the interpolated climate variables was approximately 3000 m x 3000 m. Because of this, the resolution of the NLCD data, as well as the DEM data, was generalized to match the resolution of the interpolated climate variables utilized. Maxent requires an identical extent and resolution among input covariates in order for the model to run properly. If additional climate data points were available when interpolating bioclimatic input surfaces, this could have yielded worthwhile interpolated surfaces, and final model results, at higher spatial resolutions.

6. Conclusion

Utilizing Maxent for spatial epidemiology provides a method for medical, veterinary and public health professionals to predict future areas of infection. Studies such as Gwitira et al.'s (2015) work modelling malaria with Maxent have helped to guide medical and public health professionals in designing and implementing preventative methods. Pursuant to the development of preventative measures, a comprehensive model could be used to guide medical professionals as well as veterinary professionals towards addressing risk areas for EEEV infection. This model has shown the panhandle of Florida as well as other parts of northern Florida to be spaces where there is a higher probability of habitat suitability for the disease. Not only do these hotspots represent risks to local populations but also have the potential to assist in seeding the virus in other regions. Natural spaces have suitable conditions for virus proliferation, including characteristic elevation, precipitation, and temperature. Previous studies had focused on the risk associated with land cover, but the model created in this research suggests that land cover is a small factor in the overall risk assessment for an area. Risk of EEEV transmission in Florida and other states needs to be evaluated more holistically based on the variables identified here as well as additional climatic variables. As the climate changes, more regions may become suitable habitats for EEEV. In addition, as urbanization continues, there is the risk increased contact of humans with habitats where conditions are suitable for the virus cycle and where humans are more likely to become exposed.

There is a vaccine for equines, but the virus is usually fatal in unvaccinated equines. While Florida is no longer as competitive in equine breeding as it has been historically (Wilson et al. 1986), EEEV represents a viable threat to horse stocks and therefore local economies. EEEV is also often fatal in humans, and there is no vaccine or current treatment available for it once an individual has contracted the disease (Vander Kelen et al. 2014). Case fatality for EEEV in humans is around 50% and cases that are not fatal still cause permanent and serious neurological damage (FDOH 2019). Awareness of high-risk areas will allow veterinary professionals to ensure that all at-risk animals in that region have been properly vaccinated and will also provide a potential warning system for doctors. In general, successful modelling of EEEV using Maxent will add to the growing body of literature using Maxent in spatial epidemiology and help highlight its practical applications. Further research could utilize this modelling technique to evaluate habitat suitability probability in states who are beginning to experience an uptick in EEEV activity.

ORCID

Claire Burch http://orcid.org/0000-0003-0068-0440
Rebecca Loraamm http://orcid.org/0000-0002-6829-0468

References

Baldwin, R. 2009. "Use of Maximum Entropy Modeling in Wildlife Research." *Entropy* 11 (4): 854–866. doi:10.3390/e11040854.

Belkhiria, J., R. J. Hijmans, W. Boyce, B. M. Crossley, and B. Martínez-López. 2018. "Identification of High Risk Areas for Avian Influenza Outbreaks in California Using Disease Distribution Models." *PloS One* 13 (1): e0190824. doi:10.1371/journal.pone.0190824.

Bigler, W. J., E. B. Lassing, E. E. Buff, E. C. Prather, E. C. Beck, and G. L. Hoff. 1976. "Endemic Eastern Equine Encephalomyelitis in Florida: A Twenty-year Analysis, 1955–1974." *The American Journal of Tropical Medicine and Hygiene* 25 (6): 884–890. doi:10.4269/ajtmh.1976.25.884.

Bingham, A. M., N. D. Burkett-Cadena, H. K. Hassan, C. J. McClure, and T. R. Unnasch. 2014. "Field Investigations of Winter Transmission of Eastern Equine Encephalitis Virus in Florida." *The American Journal of Tropical Medicine and Hygiene* 91 (4): 685–693. doi:10.4269/ajtmh.14-0081.

CDC ArboNet. 2018. *Eastern Equine Encephalitis – 2018 Final Human Data – USA*. Center for Disease Control. https://wwwn.cdc.gov/arbonet/maps/ADB_Diseases_Map/index.html

CDC ArboNet. 2019. *Eastern Equine Encephalitis – 2019 Provisional Human Data – USA*. Center for Disease Control. https://wwwn.cdc.gov/arbonet/maps/ADB_Diseases_Map/index.html

CDC (Centers for Disease Control and Prevention). 2016. *Eastern Equine Encephalitis*. April 6. https://www.cdc.gov/easternequineencephalitis/index.html

Costa, J., L. L. Dornak, C. E. Almeida, and A. T. Peterson. 2014. "Distributional Potential of the Triatoma Brasiliensis Species Complex at Present and under Scenarios of Future Climate Conditions." *Parasites & Vectors* 7 (1): 238. doi:10.1186/1756-3305-7-238.

Cupp, E. W., K. Klingler, H. K. Hassan, L. M. Viguers, and T. R. Unnasch. 2003. "Transmission of Eastern Equine Encephalomyelitis Virus in Central Alabama." *The American Journal of Tropical Medicine and Hygiene* 68 (4): 495–500. doi:10.4269/ajtmh.2003.68.495.

Downs, J., M. Vaziri, G. Deskins, W. Kellner, K. Miley, and T. R. Unnasch. 2019. "Optimizing Arbovirus Surveillance Using Risk Mapping and Coverage Modelling." *Annals of GIS* 1–11. doi:10.1080/19475683.2019.1688391.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. "A Statistical Explanation of MaxEnt for

Ecologists." *Diversity and Distributions* 17 (1): 43–57. doi:10.1111/ddi.2010.17.issue-1.

FDOH (Florida Department of Health). 2019. *Eastern Equine Encephalitis*. November 22. http://www.floridahealth.gov/diseases-and-conditions/eastern-equine-encephalitis/index.html

González, C., A. Paz, and C. Ferro. 2014. "Predicted Altitudinal Shifts and Reduced Spatial Distribution of Leishmania Infantum Vector Species under Climate Change Scenarios in Colombia." *Acta Tropica* 129: 83–90. doi:10.1016/j.actatropica.2013.08.014.

Gwitira, I., A. Murwira, F. M. Zengeya, M. Masocha, and S. Mutambu. 2015. "Modelled Habitat Suitability of a Malaria Causing Vector (Anopheles Arabiensis) Relates Well with Human Malaria Incidences in Zimbabwe." *Applied Geography* 60: 130–138. doi:10.1016/j.apgeog.2015.03.010.

Heberlein-Larson, L. A., Y. Tan, L. M. Stark, A. C. Cannons, M. H. Shilts, T. R. Unnasch, and S. R. Das. 2019. "Complex Epidemiological Dynamics of Eastern Equine Encephalitis Virus in Florida." *The American Journal of Tropical Medicine and Hygiene* 100 (5): 1266–1274. doi:10.4269/ajtmh.18-0783.

Jacob, B. G., N. D. Burkett-Cadena, J. C. Luvall, S. H. Parcak, C. J. McClure, L. K. Estep, … T. R. Unnasch. 2010. "Developing GIS-based Eastern Equine Encephalitis Vector-host Models in Tuskegee, Alabama." *International Journal of Health Geographics* 9 (1): 12. doi:10.1186/1476-072X-9-12.

Jaing, P. 2010. *Florida Department of Agriculture and Consumer Services*. Mosquito Control.

Merow, C., M. J. Smith, and J. A. Silander Jr. 2013. "A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36 (10): 1058–1069. doi:10.1111/ecog.2013.36.issue-10.

National Atlas of the United States. 2007. Cities and Towns of Florida Obtained from Florida Geographic Data Library [Data]. 1:2,000,000. "Cities and towns of Florida." June. https://fgdl.org/metadataexplorer/explorer.jsp

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3–4): 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

Phillips, S. J. 2017. *A Brief Tutorial on Maxent*. Accessed April 16 2019. http://biodiversityinformatics.amnh.org/open_source/maxent/

Phillips, S. J., M. Dudik, and R. E. Schapire Maxent software for modeling species niches and distributions (Version 3.4.1). http://biodiversityinformatics.amnh.org/open_source/maxent/

Rose, H., and R. Wall. 2011. "Modelling the Impact of Climate Change on Spatial Patterns of Disease Risk: Sheep Blowfly Strike by Lucilia Sericata in Great Britain." *International Journal for Parasitology* 41 (7): 739–746. doi:10.1016/j.ijpara.2011.01.012.

Scott, T. W., and S. C. Weaver. 1989. "Eastern Equine Encephalomyelitis Virus: Epidemiology and Evolution of Mosquito Transmission." In *Advances in Virus Research*. Vol. 37, 277–328. Cambridge, MA:Academic Press.

Signorini, M., R. Cassini, M. Drigo, A. F. Di Regalbono, M. Pietrobelli, F. Montarsi, and A. S. Stensgaard. 2014. "Ecological Niche Model of Phlebotomus Perniciosus, the Main Vector of Canine Leishmaniasis in North-eastern Italy." *Geospatial Health* 9 (1): 193–201. doi:10.4081/gh.2014.16.

Tan, Y., T. T. Y. Lam, L. A. Heberlein-Larson, S. C. Smole, A. J. Auguste, S. Hennigan, … M. H. Shilts. 2018. "Large-scale Complete-genome Sequencing and Phylodynamic Analysis of Eastern Equine Encephalitis Virus Reveals Source-sink Transmission Dynamics in the United States." *Journal of Virology* 92 (12): e00074–18. doi:10.1128/JVI.00074-18.

Vander Kelen, P., J. A. Downs, T. Unnasch, and L. Stark. 2014. "A Risk Index Model for Predicting Eastern Equine Encephalitis Virus Transmission to Horses in Florida." *Applied Geography* 48: 79–86. doi:10.1016/j.apgeog.2014.01.012.

Vander Kelen, P. T., J. A. Downs, N. D. Burkett-Cadena, C. L. Ottendorfer, K. Hill, S. Sickerman, … V. Hoover. 2012a. "Habitat Associations of Eastern Equine Encephalitis Transmission in Walton County Florida." *Journal of Medical Entomology* 49 (3): 746–756. doi:10.1603/ME11224.

Vander Kelen, P. T., J. A. Downs, L. M. Stark, R. W. Loraamm, J. H. Anderson, and T. R. Unnasch. 2012b. "Spatial Epidemiology of Eastern Equine Encephalitis in Florida." *International Journal of Health Geographics* 11 (1): 47. doi:10.1186/1476-072X-11-47.

Wilson, J. H., H. L. Rubin, T. J. Lane, and E. P. J. Gibbs. 1986. "A Survey of Eastern Equine Encephalomyelitis in Florida Horses: Prevalence, Economic Impact, and Management Practices, 1982–1983." *Preventive Veterinary Medicine* 4 (3): 261–271. doi:10.1016/0167-5877(86)90028-0.