

PREDICTIVE NICHE MODELING FOR THE IDENTIFICATION OF MAIZE
PATHOGENS OF GREATEST CONCERN IN THE UNITED STATES

A Thesis

Presented to

The Faculty of the Department of Biological Sciences
Sam Houston State University

In Partial Fulfillment

of the Requirements for the Degree of
Master of Science

by

Nicholas Reger

August 2022

PREDICTIVE NICHE MODELING FOR THE IDENTIFICATION OF MAIZE
PATHOGENS OF GREATEST CONCERN IN THE UNITED STATES

by

Nicholas Reger

APPROVED:

Chris Randle, PhD
Committee Director

Justin Williams, PhD
Committee Member

Amber Ulseth, PhD
Committee Member

John B. Pascarella, PhD
Dean, College of Science and
Engineering Technology

DEDICATION

I dedicate this thesis to my family: My wife, Shelby Reger, for always being the light of my life, my father, Gregg Reger, for showing me joys of learning, and my mother, Azita Reger, who I have always looked to for inspiration.

ABSTRACT

Reger, Nicholas, *Predictive niche modeling for the identification of maize pathogens of greatest concern in the United States*. Master of Science (Biology), May, 2022, Sam Houston State University, Huntsville, Texas.

Maize is one of the world's most valuable food crops with 717 million metric tons produced annually. Its economic significance worldwide is second only to rice. Given the importance of maize, it is crucial to understand the potential range of pests and pathogens that pose a significant risk to the crop. Ecological niche modeling is used to identify the environmental requirements of these pests and pathogens. Models can be built using existing occurrence data and records of environmental conditions such as vegetative coverage, isothermality, altitude, temperature, and precipitation. In this study, I use pest occurrence location data from the Global Biodiversity Information Facility and bioclimatic variables from WorldClim to create generalized additive models (GAM), maximum entropy (MaxENT) models, boosted regression trees (BRT), ensemble, and CLIMEX models to predict suitable habitat for maize pests and pathogens in the US. Distribution models were made of insect pests of highest concern, including Lepidopterans *Autographa gamma*, *Chilo partellus*, *Helicoverpa armigera*, *Spodoptera litura*, and *Thaumatotibia leucotreta*, Coleopterans *Diabrotica speciosa* and *Heteronychus arator*, and the Hemipteran *Laodelphax striatellus*. Each of the forty models were then used to make maps of the potential geographical range that highlights areas that would be most suitable to the greatest number of pests. Coastal areas are susceptible to most maize pests and these maps convey the levels of risk associated with land near an ocean. These maps can be used to efficiently direct preventative action to high-risk areas.

KEY WORDS: Maize; Ecological niche modeling; MaxEnt, Conservation; Boosted regression trees; Ensemble modeling; Sam Houston State University; Texas; Graduate school.

ACKNOWLEDGEMENTS

I would like to acknowledge my team at the Randle lab at Sam Houston State University; specifically, Kayla Hankins, Jesus Castillo, and Kenneth Nobleza for their assistance in developing an R template. I would also acknowledge the most important person in my project, Dr. Christopher Randle for his knowledge, leadership, patience, and guidance. I would also like to specifically thank Laura Bianchi, for helping me first learn ecological niche modeling and demonstrating java usage in R. This research would not have been possible without the grant awards AP20PPQS&T00C153 and AP21PPQS&T00C019 from the USDA-APHIS 7721Farm Bill.

TABLE OF CONTENTS

	Page
DEDICATION	iii
ABSTRACT.....	iv
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS.....	vii
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER I: INTRODUCTION	1
Pests	2
Ecological Niche Drives Pest Distributions	3
Niche Modeling	3
Occurrence Data	5
Bioclimatic Predictors and Collinearity.....	7
Assumptions of Niche Models.....	8
GAM	10
MaxEnt	11
BRT.....	12
CLIMEX	12
Study Goals.....	13
CHAPTER II: METHODS	15
Bioclimatic Predictors and Collinearity.....	17
Correlative Model Training and Prediction	18

Correlative Model Evaluation.....	19
Ensemble Modeling and Evaluation	20
Summary Analysis.....	20
CLIMEX Training and Prediction	20
CHAPTER III: RESULTS	23
Results of each Species.....	23
CLIMEX	38
Differences in Model Projections	41
Overall Trends	41
CHAPTER IV: DISCUSSION	49
Modeling Invasive Species	50
Human Impact as a Predictor.....	50
Differences in Model Projections	50
Potential Model Performance Erosion.....	51
Model Usage and Improvement.....	52
REFERENCES	53
VITA.....	71

LIST OF TABLES

Table	Page
1 AUC Values of GAM, MaxEnt, BRT, and Ensemble Models	43
2 Predictors in Descending Order of Importance	44

LIST OF FIGURES

Figure	Page
1 Occurrence Records	16
2 Habitat Suitability of <i>Autographa Gamma</i>	24
3 Habitat Suitability of <i>Chilo Partellus</i>	26
4 Habitat Suitability of <i>Diabrotica Speciosa</i>	28
5 Habitat Suitability of <i>Helicoverpa Armigera</i>	30
6 Habitat Suitability of <i>Heteronychus Arator</i>	32
7 Habitat Suitability of <i>Laodelphax Striatellus</i>	34
8 Habitat Suitability of <i>Spodoptera Litura</i>	36
9 Habitat Suitability of <i>Thaumatotibia Leucotreta</i>	38
10 CLIMEX Maps	39
11 Weighted Average Pest Risk Map	47
12 Corn in High-Risk Areas Map	48

CHAPTER I

Introduction

Maize is crucial to both the diets of humans and domesticated animals, which places it at the center of global food security (Mahuku et al., 2015). In the United States, 90 million acres are planted with maize, which is then partially used to produce 95% of the feed grain used in the United States (USDA United States Department of Agriculture, 2020). The U.S. produced 347 million metric tons of maize in 2019 and approximately 41.8% of the crop yield was used to produce ethanol (FAO Food and Agriculture Organization of the United Nations, 2019; Klopfenstein et al., 2013). Soybeans are the only agricultural commodity in the United States that contribute more to GDP than corn (Agricultural Exchange Rate Data set. USDA ERS - Home, n.d.). Corn is an ingredient that has been used in used in items such as toothpaste, cosmetics, shampoo, diapers, paper, shoes, spark plugs, tires, fireworks, deodorant, textiles, and plastics (Agricultural Exchange Rate Data set. USDA ERS - Home, n.d.).

Pests and pathogens of maize threaten food security. The economic damage from insects alone costs at least 70.0 billion dollars per year globally (Bradshaw et al, 2016). Strategies have been developed to control or contain domestic pests and there are also native predators that limit domestic pest population size and range expansions. However, there are no strategies nor native predators to control non-native corn pests. Alien crop pests have been referred to as the “single most formidable natural disaster threat of the 21st century” (Schnase et al., 2002). Pests destroy between five to forty percent of crop yields, or two to sixteen billion dollars of corn domestically per year (Sarkozi, 2021). Native and established nonnative pests are the reason the domestic pest control industry

generates almost ten billion dollars in revenue a year (Industry Market Research, 2022).

The economic costs will continue to worsen as invasive nonnative species are being introduced domestically through international trade and human travel at an increasing rate (Hulme, 2009).

Pests

The United States Department of Agriculture has identified maize pests of the highest concern (USDA Cooperative Agricultural Pest Survey, 2021). The pests include: *Autographa gamma* (Silver Y moth), *Chilo partellus* (spotted stem borer), *Heteronychus arator* (African black beetle), *Diabrotica speciosa* (cucurbit beetle), *Spodoptera litura* (tobacco cutworm), *Thaumatotibia leucotreta* (false codling moth), *Helicoverpa armigera* (cotton bollworm), and *Laodelphax striatellus* (small brown planthopper).

Autographa gamma is dangerous because the larvae eat around the leaf plate and transfer viruses (Sullivan & Molet, 2007). *Chilo partellus* is a threat because as young larvae feed on leaf whorl and older larvae tunnel into stems (PPQ, 2019). *Heteronychus arator* chews the cortex of stems underground (Sullivan & Molet, 2007). *Diabrotica speciose* poses a risk to maize because larvae feed on roots adults feed on most parts of the plant (CPHST Pest Datasheet, 2022). *Spodoptera litura* is difficult to contain because as a flying adult it can migrate up to 1.5 kilometers and it does damage by chewing leaves (Sullivan, 2007). *Thaumatotibia leucotreta* creates a cavity in the stem that is almost a perfect circle; it does little initial damage but eventually allows molds entry into the stem (Sullivan, 2007). *Helicoverpa armigera* is particularly a nuisance because it attacks plants reproductive systems (Sullivan & Molet, 2007). *Laodelphax striatellus* is a threat because it is a vector for numerous plant viruses (Isiguro & Yanase, 1991).

Ecological Niche Drives Pest Distributions

Organisms can only persist within a geographical range that provides an environment that supports population growth. The set of resources and environmental conditions that allow the species to maintain a viable population is called a niche. A species distribution is determined in part by its ecological niche, which comprises abiotic conditions such as climate and biotic interactions such as competition or prey availability, but also a species ability to disperse into available niche space (Soberón & Nakamura, 2009).

There are several ways to conceptualize niche and the distinctions are important because different understandings of niche correspond to diverse ways of modeling species distributions. This is because species distribution is constrained by the species niche. The Hutchinsonian niche is an n-dimensional hypervolume, where the dimensions are the environmental conditions and resources, that define the requirements of a species to persist (Levin et al., 2009). The Grinnellian niche is the sum of the habitat requirements and behaviors that allow a species to persist (Grinnell, 1917). A fundamental niche is the entirety of abiotic conditions under which a species can experience positive population growth. The realized niche, or the set of ecological conditions occupied by a species, is further limited by interactions with other organisms, and dispersal ability.

Niche Modeling

The habitability of any given location for a particular species can be discerned through ecological niche modeling. Niche models are mathematical functions that describe the relationship between environmental conditions and species presence, and

which can be used to make predictions of the probability of species presence based on these environmental variables. While the environmental niche that constrains a species distribution is made of many factors other than just abiotic climate data, niche modeling is often based solely on the environmental requirements of climate interactions because biotic interactions and dispersal are often too variable and poorly quantified for modeling. Such niche models are called bioclimatic models. The strong relationship between species distributions and environmental requirements allows for ecological niche modeling to identify the environmental requirements of pests and pathogens using presence-only data (Elith et al., 2006). This assumes that a niche can be represented by an n-dimensional hypervolume of environmental conditions (Grinnell, 1917).

Niche modeling is a proactive approach to invasive pest control; being able to predict where a pest is most likely to become established so that preventive measures can be employed is more efficient than the eradication of an already-established population (Peterson & Viegla 2001). Models of invasive species are most predictive when they are constructed based on observations of species occurrence locations from both invaded and native ranges (Crimmins et al., 2006). This is likely because when a species is introduced to a nonnative range there are different environmental pressures limiting its distribution that may not exist in both ranges (Pérez et al., 2006).

Most bioclimatic models are correlative as they link environmental conditions with the probability of species occurrence. On the other hand, mechanistic models link ecophysiological responses to environmental covariates (Webber et al., 2011). Correlative models recover the Hutchinsonian niche as they are constructed using occurrences to infer the n dimensions of the Hutchinsonian niche and mechanistic models

predict distributions closer to the realized Grinnellian climate niche (Rodda et al., 2011). This is because mechanistic models are built using only physiological constraints rather than being built from functions linking presence to environmental conditions like a correlative model.

There is almost always a difference between realized niche and the fundamental niche, and as models are built using occurrences from the realized niche no model will represent the full fundamental niche (Phillips et al., 2006). This is because current models do not have the capacity to incorporate all the information required to represent the fundamental niche. To increase the portion of the fundamental niche represented by the model, occurrence records from a as large geographic range as possible must be incorporated (Peterson & Holt 2003).

Occurrence Data

Niche models are most often built using environmental data gathered from occurrence records. Occurrence records represent sets of coordinates where a species has been observed to occur. These are recorded mostly from museum and herbarium specimens (Ponder et al., 2001). Rarely do occurrence data represent a random sample of the species distribution. Rather, collections are made largely based on convenience—from places that are easily accessible to the collector: roadsides, natural areas, and locations within a half-day's drive of a biological collection facility. Convenience sampling will result in a non-random sample of a species realized niche, with some environments being over-represented by occurrence records, while others are under-represented. Such a sample can be said to be autocorrelated. In an autocorrelated sample, occurrences can be predicted by past occurrences. Autocorrelated points that result from

convenience sampling artificially inflate the significance of environmental conditions of overrepresented environmental conditions. This is known as model distortion.

To estimate a probability distribution of habitat suitability on a set of environmental conditions associated with occurrence points, models need information on all environmental conditions that may be available to a species in a landscape. Background data are coordinates that represents where a species has not been observed. They are used to teach a model what environmental conditions may not be suitable for a species. If a model is built with background data that characterize environments in the study region with the same bias as the occurrence data, model distortion is avoided and there is improved predictive performance (Phillips et al., 2009). A distinction should be made between background and absence data, though both are largely used the same way in niche modeling. Absence data are records of where a species was sought out and was not found. Absence records are more likely to represent true species absence than background data because the absence is supported by survey effort. Absence data is also directly correlated with the quality of the survey and the technical expertise of the individuals doing species identification. An individual can record false absences for many reasons such as making observations at the wrong time or not waiting long enough for the species to appear. While presence data is readily available and publicly accessible, absence data is not often collected or available (Phillips et al., 2006). When absence data is available, it may suffer from inadequate sampling that results in false negatives of species presence (Anderson, 2003).

Presence data is collected and databased by people and is also subject to human error (Wieczorek et al., 2004). Misidentification errors are common with species that

resemble other species. This type of error can be minimized either by using only occurrences from highly trained individuals or modeling particularly distinctive species. Transcription errors arise when records were entered into a database or from an unspecified geographical datum. Some of these erroneous points can be caught in a data cleaning process (Hijmans et al., 1999). Nonetheless, remaining errors degrade the predictive capacity of most models.

Bioclimatic Predictors and Collinearity

Predictor variables are the ecological conditions associated with occurrence points. Predictors are abiotic factors associated with the environment, such as vegetative coverage, moisture, and temperature, which constrain a species' niche. In a simple sense, conditions outside of a species physiological tolerance result in the absence of that species. Different species have different physiological tolerances based on their individual adaptations. For example, a species can be well adapted to cold temperatures while another species cannot handle cold temperatures but is instead well adapted to high temperatures. It is important to consider how predictors affect species in conjunction with other predictors as many predictors have interactions with each other. One of the most straightforward interactions is how high temperature can be offset by a high precipitation that protects the individuals from the desiccating properties of high heat.

Some predictors may predict the values of other predictors, a correlation called collinearity. For example, solar radiation can be used to predict temperature which makes the two predictors collinear. Collinearity is a frequent problem in species distribution modeling, reducing predictive capacity and leading to the misidentification of predictors in a model (Dormann et al., 2013). This problem is exacerbated when a model that was

built with one collinearity structure is extrapolated to another area where the collinearity can be different. When the correlation coefficients between predictor variables are greater than 0.7 collinearity begins to severely distort the model (Dormann et al., 2013). Through a strong familiarity with the underlying ecology behind the model, variable selection can be used to avoid collinearity.

Collinearity can be measured either between two predictors or measure how much one predictor contributes to overall collinearity. When a researcher has significant technical expertise about a species and its physiological tolerances, they know what variables are important and can do pairwise collinearity analysis to remove collinear predictors. Pairwise collinearity can be measured with Pearson's coefficient. In research with a large scope where there is a variety of diverse species incorporated, collinearity is usually measured based on each predictor's contribution to overall collinearity. Overall collinearity can be measured with variance inflation factor (VIF) (Gareth et al., 2017). VIF is a ratio of the variance of estimating a predictor in a model that includes multiple terms by the variance of a model constructed using only one term. The larger the VIF, the more colinear a predictor is with other predictors.

Assumptions of Niche Models

There are a few assumptions inherent in niche models, which when violated result in poor prediction. The most significant weakness results from the assumption that biotic interactions are constant over space and time. Models that use open-ended response curves predict species suitability in areas where the species cannot occur if they extrapolate globally from a limited dataset (Webber et al., 2011). An open-ended response curve is when a model recognizes the species survives at the highest values

available in the range of occurrences and so it predicts it will continue to survive in more extreme conditions than were used to train the model. Predictive capacity is also diminished if a model ignores sub-taxon structure because that violates the assumption that populations of a species have similar environmental requirements (Pearman et al., 2010; Shabani et al., 2017). Another assumption in niche modeling that species are in equilibrium with their environment (Guisan et al., 2021). This assumes that the species is no longer evolving in response to environmental conditions in its home range. Niche models constructed using a small portion of the known range do not account for the complete set of environmental conditions in which a species occurs, violating this assumption. Another assumption is that environmental factors important for determining a species distribution are available, which may be violated if the understanding of a species is not rich enough to include all the most important predictors (Guisan et al., 2021). Another assumption is that occurrences are samples from populations experiencing growth (Guisan et al., 2021). This assumption is violated when occurrences are from populations where the occurrence is only temporary, as when propagules disperse into environments in which they can survive but cannot reproduce. There is an assumption that predictor data is free of error, but it is possible that there are errors with the equipment used to record the measurements (Guisan et al., 2021). Another assumption is that occurrence data represents a randomly selected subset of the realized distribution (Guisan et al., 2021). This assumption is violated by convenience sampling, though background data with the same bias can mitigate some of the model degradation from this assumption violation. Despite the assumption violations, niche models can predict

habitable areas that were not used to train the model which makes them a useful tool in identifying potentially viable habitats.

This study will make use of three correlative models (General Additive Models, Maximum Entropy, and Boosted Regression Trees) and one mechanistic model (CLIMEX) to estimate the suitability of habitats in the conterminous US for USDA-CAPS-identified pests of maize.

GAM

Generalized additive models are a type of generalized linear model (GLM) (Hastie & Tibshirani, 1990). A GLM allows the magnitude of the variance of each measurement to be a function of its predicted value (Nelder & Wedderburn, 1972). It does so by relating a linear model to the response variable with a link function (Nelder & Wedderburn, 1972). A link function provides the relationship between the linear predictor and the mean of the distribution function (Nelder & Wedderburn, 1972). There are many well defined link functions that fit distributions such as normal, exponential, poisson, or other distributions (Nelder & Wedderburn, 1972). A GAM uses splines to fit together simple polynomials that likely fit the data in a limited range as a way of generating a link function that can closely fit a population (Hastie & Tibshirani, 1990). This is a simple model in comparison to MaxEnt and BRT models and is not as robust when confronted with occurrence errors. It creates a less conservative predicted distribution than the other models used in this research because splines cannot capture interactions between predictors.

MaxEnt

Maximum Entropy modeling (MaxEnt) is a machine learning technique that selects functions to build a model based on maximizing informational entropy (Guíasu & Shenitzer, 1985). Machine learning is a type of artificial intelligence that allows software to increase predictive accuracy without being manually programmed to increase accuracy (Mitchell, 1997). Informational entropy can act as a numerical measure describing how a particular probability distribution is uninformative (Guíasu & Shenitzer, 1985). The most informative probability distribution occurs when one of the propositions is known to be true, where information entropy is zero (Guíasu & Shenitzer, 1985). The least informative distribution is when there is no support for any propositions, where informational entropy is at its maximum value (Guíasu & Shenitzer, 1985). The distribution with the maximum entropy is the one that makes the fewest assumptions about the true distribution of data (Guíasu & Shenitzer, 1985). Therefore, when informational entropy is maximized, a model can be used to derive distributions from observational data while making the fewest implicit assumptions, and therefore makes predictions with no more precision than that required by the data (Guíasu & Shenitzer, 1985). The model is constructed using a uniform distribution as a starting point. MaxEnt simplifies functions used to describe the dataset to reduce overfitting, a process called regularization (Phillips et al., 2006), which is adjusted based on sample size because it has the most impact when sample sizes are small (Elith et al., 2006). Overfitting is when the model fits the training data so closely that it makes good predictions only about the data used to train the model and makes only negative predictions in environmental space outside of that used to train the model. MaxEnt models include interaction terms that

account for how an environmental variable changes the effect of another environmental variable. This modeling technique is unique in that model predictive performance is not impaired by the inclusion of a few errors in the occurrence records (Graham et al., 2008).

BRT

Boosted regression trees combine regression trees and boosting to create an additive regression model where the individual terms are simple trees, fitted forward and stagewise (Elith, 2008). The simple trees are graphs defined by branching points called nodes stagewise (Elith, 2008). A node is a threshold of an ecological condition where the tree forks into two directions called a binary split stagewise (Elith, 2008). For example, a tree may have a node of a temperature above 100°F where one direction is that the species cannot inhabit the environment and the other direction is that the species can inhabit the environment. On the branch where the species does survive there can be more nodes that convey other partitions that define probability of species survival. Boosting approximates the true response surface by combining simple trees that have misclassification problems thereby improving overall predictive accuracy (Elith et al., 2006). BRT models also account for interaction terms to a further extent than MaxEnt models (Elith et al., 2006). These models are slightly impaired when there is erroneous location data in the occurrence records. However, BRT models perform better than most other modeling techniques when occurrence records are flawed (Graham et al., 2008).

CLIMEX

CLIMEX is a mechanistic model that uses physiological observational data along with occurrence data which makes it ideal for pests that do not have abundant occurrence records. If the physiological tolerance data is not available, it can be inferred from

occurrence records. These models assume that climate is the main factor of species distributions (Kriticos et al., 2007). Mechanistic models such as CLIMEX are the most predictive model for novel climatic factors such as climate change predictions (Webber et al., 2011). The model incorporates population growth and physiological stress attributable to climatic variables. These effects are indexed as an Annual Growth Index that represents growth potential in the favorable season, and multiple Stress Indices that represent population reduction during an unfavorable season. The product of the Annual Growth Index and Stress Indices is an annual average index called the Ecoclimatic Index (EI). This index has a value from 0 to 100 that represents the suitability for species to occupy an area. Species may be established where EI is greater than 0, are marginally suitable where EI is 1 to 10, are moderately suitable where EI is 10 to 20, and are optimally suited where EI is above 20.

Study Goals

In this study I will be making predictions about most suitable habitat for corn pests in the United States. Maps of habitat suitability can be built using the models of physiological tolerances and worldwide climate data. Four different models will be used: MaxEnt, BRT, GAM, and CLIMEX. These models were chosen based on their high predictive capacity and because each is built on different statistical assumptions.

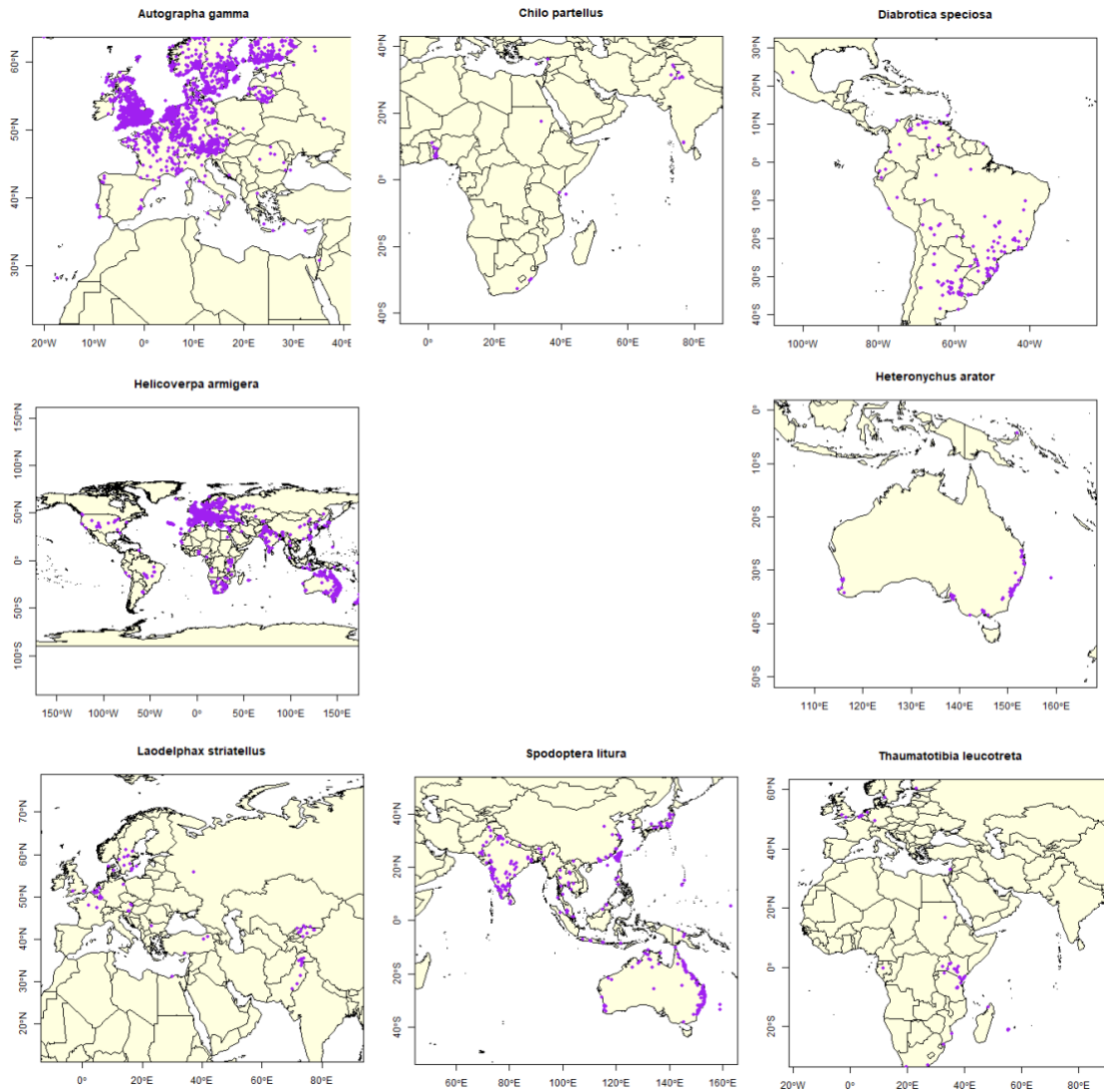
The first goal of this study is to create risk maps that convey what land in the United States is habitable to invasive pests. First predictive distribution maps will be generated of maize pests of the highest concern as identified by the USDA. Ensemble risk maps will then be generated using an ensemble approach that incorporates GAM, MaxEnt, and BRT models. A map of total risk in the United States was generated by

taking a weighted average of the eight ensemble maps that convey habitat suitability. A total risk map will overlay land where corn is grown and areas where three or more invasive pests have habitable environments. The second goal is to use CLIMEX, as a mechanistic model to make species distribution maps.

CHAPTER II

Methods

The eight pests that were modeled in this study were selected from the greater list of pests provided by the USDA based on occurrence records (USDA CAPS 2021). Pests that were not included, particularly molds, rusts, smuts, and blights, are rarely found in natural history collections, and therefore have few or no known geographical locations. Occurrence data from species home ranges were obtained from the Global Biodiversity Information Facility (Hijmans, 2021). The records include observations from museums over the last two hundred years as well as modern geotagged occurrence records. The records are in the World Geodetic System 1984 datum which was designated as a coordinates object in the R environment. Occurrence records were sanitized to remove errors such as duplications, flipped coordinates, and blank coordinates (Giannini et al., 2012). The data was also graphically projected for a visual inspection to verify there were no anomalous points in the ocean because such points are the result of data entry errors from GBIF. If there are more than 3000 occurrences a random subset of 3000 points was used as presence data. Most occurrence points were clustered around a home range except for *Helicoverpa armigera*, *Chilo partellus*, and *Spodoptera litura* which had occurrences from a broad distribution (Figure 1).

Figure 1*Occurrence Records*

A random subset of 20% of the occurrence data was withheld from training the models so that it could later be used to test the predictive capacity of the model (Fielding & Bell, 1997). This was done by assigning a random integer between one and five and segregating the points. Occurrences that received a value of one were separated from the rest of the dataset to use as testing data. To generate pseudo-absence points, I generated circles around each occurrence with a radius equal to the average distance between

occurrence points and removed any overlap between circles, creating a polygon from which I took a random sample of points. In some instances, this method created a hole where the polygon overlap would prevent processing. In such cases, the distance was iteratively lowered until there was no longer a hole.

Bioclimatic Predictors and Collinearity

Based on the occurrence records, I extracted the associated climate data from WorldClim Version 2 Bioclimatic variables (Hijmans, 2021), with a 10-min spatial resolution. The bioclimatic variables from this source all pertain to temperature and precipitation. I also included a human impact predictor from the Socioeconomic Data and Applications Center run by NASA (Wildlife Conservation Society, 2005) and an elevation raster from Worldclim (Fick & Hijmans, 2017). The human impact raster is made up of nine global data layers covering human population pressure (population density), human land use and infrastructure (built-up areas, nighttime lights, land use/land cover), and human access (coastlines, roads, railroads, navigable rivers) (Wildlife Conservation Society, 2005). To determine which predictors were not collinear, I used the VIF-step algorithm (Naimi et al., 2014). First, the VIF (variance inflation factors) was estimated jointly for all predictors sampled from occurrence training records. The predictors with the highest VIF value are considered colinear because they had the most predictive capacity towards other predictors. The most colinear predictors were sequentially removed until the variance inflation factor of all remaining predictors was below five (Chatterjee & Hadi, 2006). These remaining predictors were used to train the correlative models.

Correlative Model Training and Prediction

All three correlative models, GAM, MaxEnt, and BRT, were constructed in the R version 4.0.5 coding language (R Core Team, 2021) in RStudio version 1.3.959 (R Studio Team, 2020) including the packages *dismo* (Hijmans et al., 2021), *rgeos* (Bivand & Rundel, 2021), *maptools* (Bivand & Lewin-Koh, 2021), *shades* (Clayden, 2019), and *raster* (Hijmans, 2021). The additional package *mgcv* was used to create GAM models (Wood, 2017). The package *rJava* was used to generate the MaxEnt models using java within R (Hijmans, 2021). The package *sp* was used to work with the intermediate forms of data that went into the models such as data frames and polygons (Bivand et al., 2013).

The package *rgdal* was used to work with coordinate reference systems which needed to be associated with sets of occurrence records (Bivand et al., 2021). The package also allowed us to interpret the reference system and correct some errors. All three correlative models were initially run with all the noncolinear variables. The GAM and MaxEnt models were simply processed using their respective R packages. After the BRT model was first generated, tree complexity and learning rate were manually manipulated based so that adding more trees did not decrease the holdout deviance explained by the model. The ideal BRT model was then simplified by removing the 5 least predictive bioclimatic variables. Once the correlative models were generated and had had their efficacy tested, they were used to project distribution maps. This was done by comparing the environmental conditions of land in the United States at a 10-min spatial resolution to the environmental suitability proposed by the model. The default color palette for distribution maps in R is red and green which I modified to be a blue scale to make our results accessible to researchers with deuteranopia color blindness.

Correlative Model Evaluation

Predictive capacities were measured by the area under the receiver operating characteristic curve (AUC). This metric is based on the ratio of true positives to false positive predictions in the range of withheld occurrence records. A true positive is when the model predicted the species would occur in an area that does have an occurrence record for. A false positive is when the model makes a prediction that the species will occur in an area where it does not occur based on either absence or background data. Values of this metric range between 0 and 1 with a value of 1 meaning the model only produces true positives and never any false positives. A value of 1 is not often attained in niche modeling but instead values will usually fall between 0.5 and 1 with a value of 0.5 corresponding to a completely useless model that makes as many true positive predictions as false positive predictions.

One of the two biggest weaknesses of AUC is that a larger extent of absence data will correspond to a larger AUC value (Lobos et al., 2008). Similarly, AUC is based on false positive rates which are reliant of absence or background data which never is completely representative of true species absence and is rather an approximation. These weaknesses have created some contention over the use of AUC as a metric, but models with a high AUC value have made predictions that have been independently verified. Alternatives such as an ecological space index have been recently proposed but have yet to experience widespread use and peer review (Escobar et al., 2018). Until a better metric is introduced, AUC has become established as the standard tool for model evaluation when niche modeling.

Ensemble Modeling and Evaluation

Ensemble modeling is an approach that maximizes predictive capacity when niche modeling by mapping a distribution based on the aggregation of predictions from multiple models. Our ensemble model was developed using the AUC-weighted means of BRT, GAM, and MaxEnt predictions. This generated a model that leverages the advantages of each individual model while reducing the uncertainty present in an individual model (Beerman et al., 2021). Ensemble models usually have a higher predictive capacity, as measured by AUC, than any of the individual models because it does not suffer from the assumption violations of any one model while retaining the unique predictive capacities of each incorporated model (Marmion et al., 2009).

Summary Analysis

A map of total risk in the United States was generated by taking a weighted average based on the AUC value of the eight ensemble maps. Each individual raster was multiplied by its AUC score and added to the total risk map which was divided by the total of the AUC score values. The next risk map was generated by taking the first risk map and using an AUC threshold of 0.5 to represent areas of high habitability. The harvested area of corn in hectares from CHEMGRIDS was overlaid on this map to show corn crop land that is highly suitable to invasive pest establishment (Maggi et al., 2019).

CLIMEX Training and Prediction

The software CLIMEX Version 4 (Kriticos et al., 2015) was used to create CLIMEX models. The climatic requirements for each were inferred based on a genetic algorithm that is part of the CLIMEX software. The genetic algorithm uses occurrences

to create infer physiological tolerances of a species. Occurrences were input as ESRI point shapefiles was created in QGIS (QGIS, 2022). The point shapefile occurrence records were intersected with the CLIMEX CM10 map. Points outside the region used in CLIMEX to model distribution were manually deleted. If there are too many points for the model to run, a random subset of half of the occurrences were used in lieu of the full dataset.

Based on occurrences, the genetic algorithm quantifies how well the tolerance levels fit the occurrence record distribution. The genetic algorithm procedure was repeated for 100 generations. Each generation consists of eight sets of tolerances and the set that best fits the occurrence records is used to create the set of the next generation, resulting in stepwise improvement of model fit. A set that performs well is used as a parent set to create a new child set that is a slight modification of the parent. If the child fits the occurrences better than their parent then the parent is discarded, and the child is allowed to generate children of its own. If the child does worse than the parent, it is discarded, and the parent has more children until a child is able to surpass the parent. This procreation of parameter sets is why this is referred to as a genetic algorithm. Once physiological tolerances were estimated, the software was run in the “Compare Locations” mode where the tolerances were used to predict the species’ response to the average temperature and moisture across the United States.

CLIMEX is rarely used for niche modeling in comparison with correlative models. The main reason for it having limited use is that the way CLIMEX constructs models is private information. Therefore, more weight is given to the correlative models

that have become the current standard for niche modeling which are constructed based on established public methods.

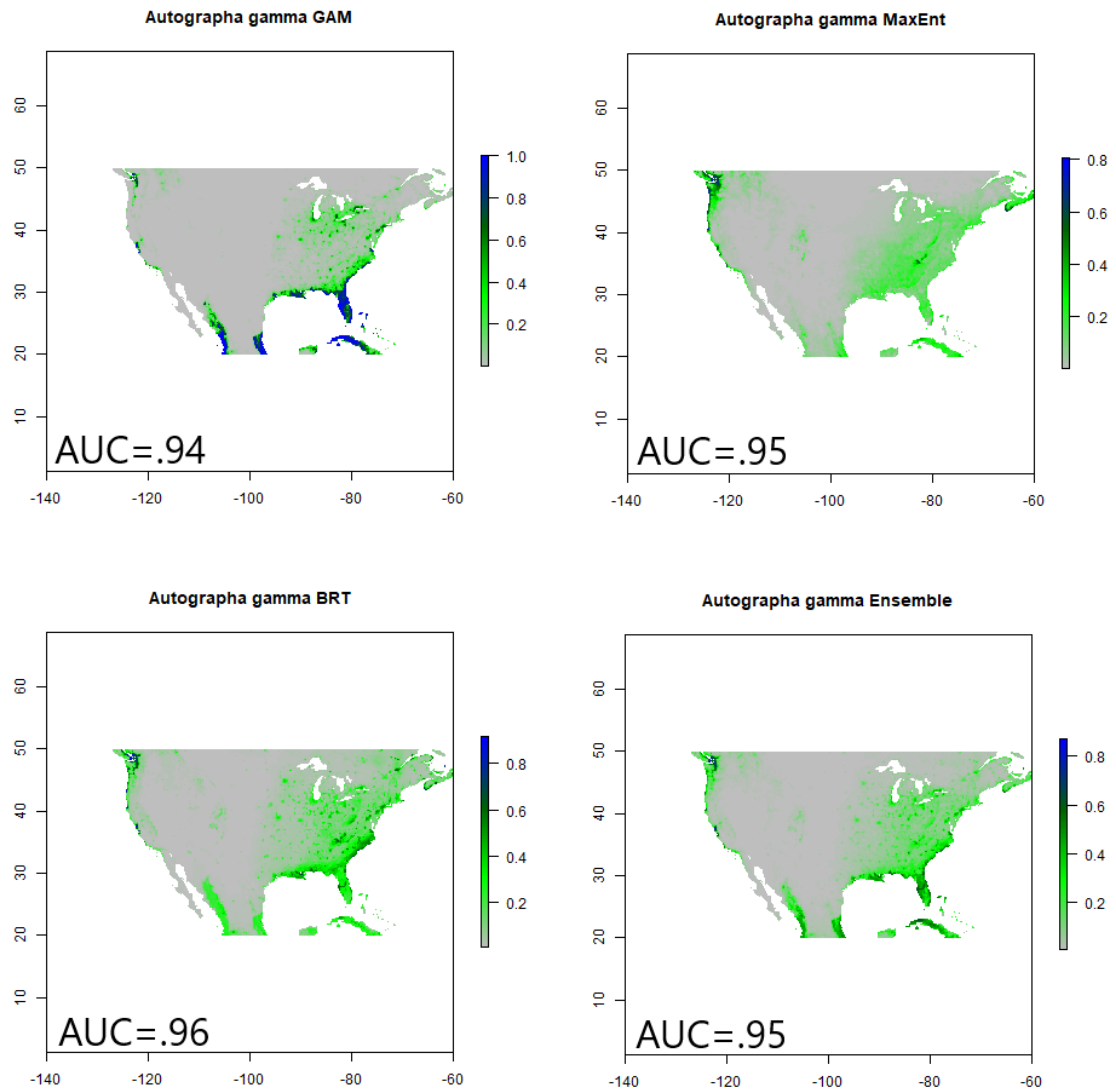
CHAPTER III

Results

Results of each Species

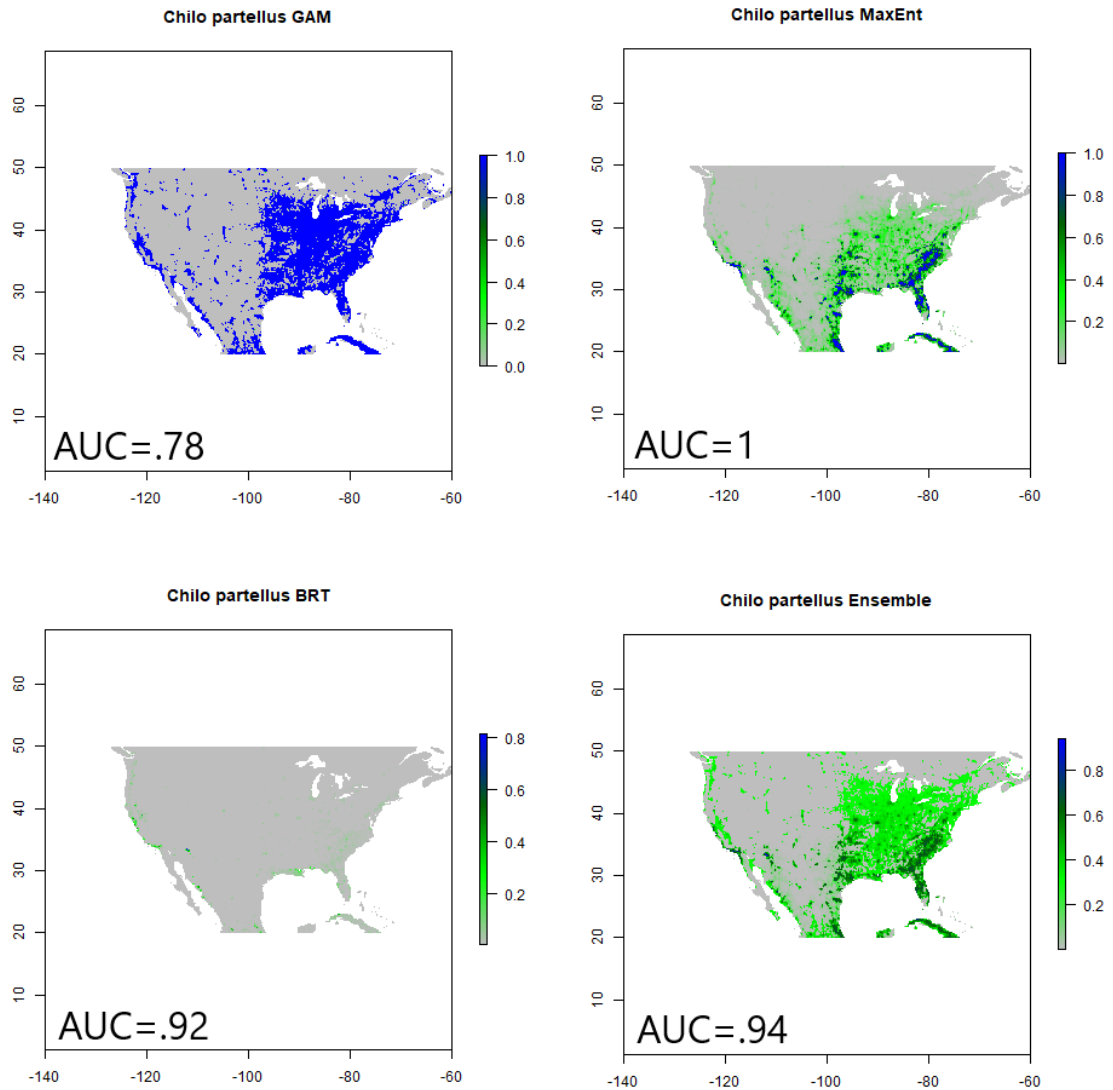
Autographa Gamma

Autographa gamma occurs everywhere in Europe as well as many occurrences in Asia, and northern Africa which was used as its home range (Venette et al., 2003). There are almost half a million occurrences on GBIF so a random subset of three thousand points had to be used so that the models could be run. The predictors that were most important in multiple models were human impact and temperature seasonality (Table 2). The east coast provides the most suitable habitat in the contiguous US under all three models (Figure 2). The MaxEnt distribution goes the farthest west Both MaxEnt and BRT project a small area of habitat suitability in the Rocky Mountains while GAM does not. All models projected habitat suitability across both coasts with much more continental penetration on the east coast. The high AUC values of these models indicate this model was able to correctly predict areas of habitat suitability and this map can be used with some confidence to predict potential *Autographa gamma* distribution.

Figure 2*Habitat Suitability of Autographa Gamma**Chilo Partellus*

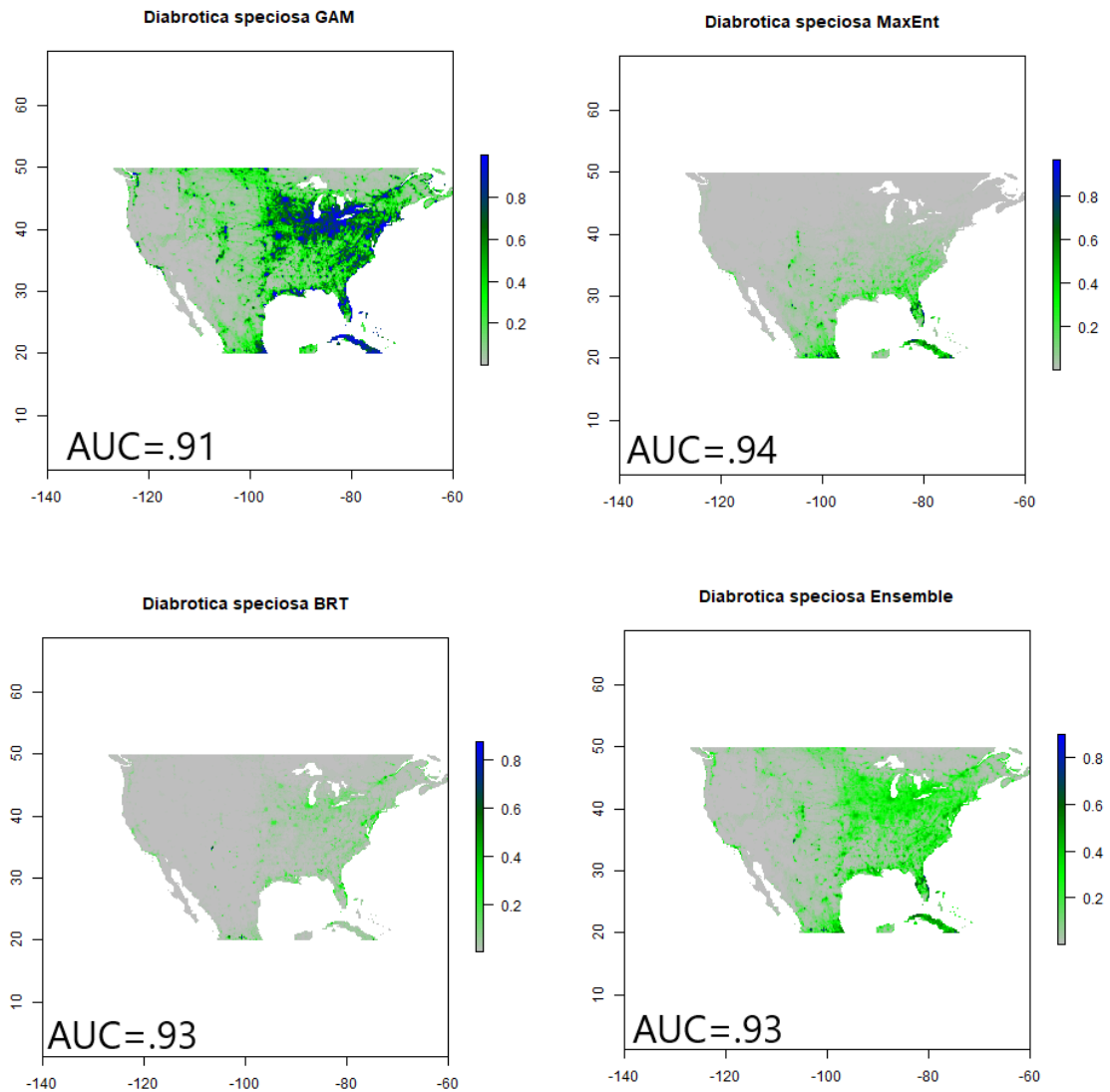
This species is native to the Indo-Australian region (Mohyuddin, 1971). The home range used to train the models was limited to Africa as there were occurrence records available. After data cleaning there were 45 occurrences records from GBIF remaining. The AUC value of 1 that is in figure 3 may be artificially high as only 35 occurrences were used to

train the MaxEnt model and 10 to test the model. The GAM model was the only projection with significant habitat suitability west of -100 longitude (Figure 3). Points in this region are not likely accurate as they have no consensus support. The exception to points in the west being inaccurate are the points along the west coast projected by the BRT model. The BRT model projected very little habitable area and all the areas it projected to be suitable are corroborated by all the other models. That corroboration makes the BRT projections the most likely to be accurate. If the BRT model was too conservative MaxEnt has a distribution that covers every latitude and goes as far west as -100 longitude with the unusually high AUC score of 1. Despite the high AUC values of these models, these predictions should be used cautiously as there was little data available to train and test this model. The stark differences between the BRT and MaxEnt model highlight the high level of model uncertainty. The model uncertainty is especially highlighted in how all predictors were equally important in the GAM model.

Figure 3*Habitat Suitability of Chilo Partellus**Diabrotica Speciosa*

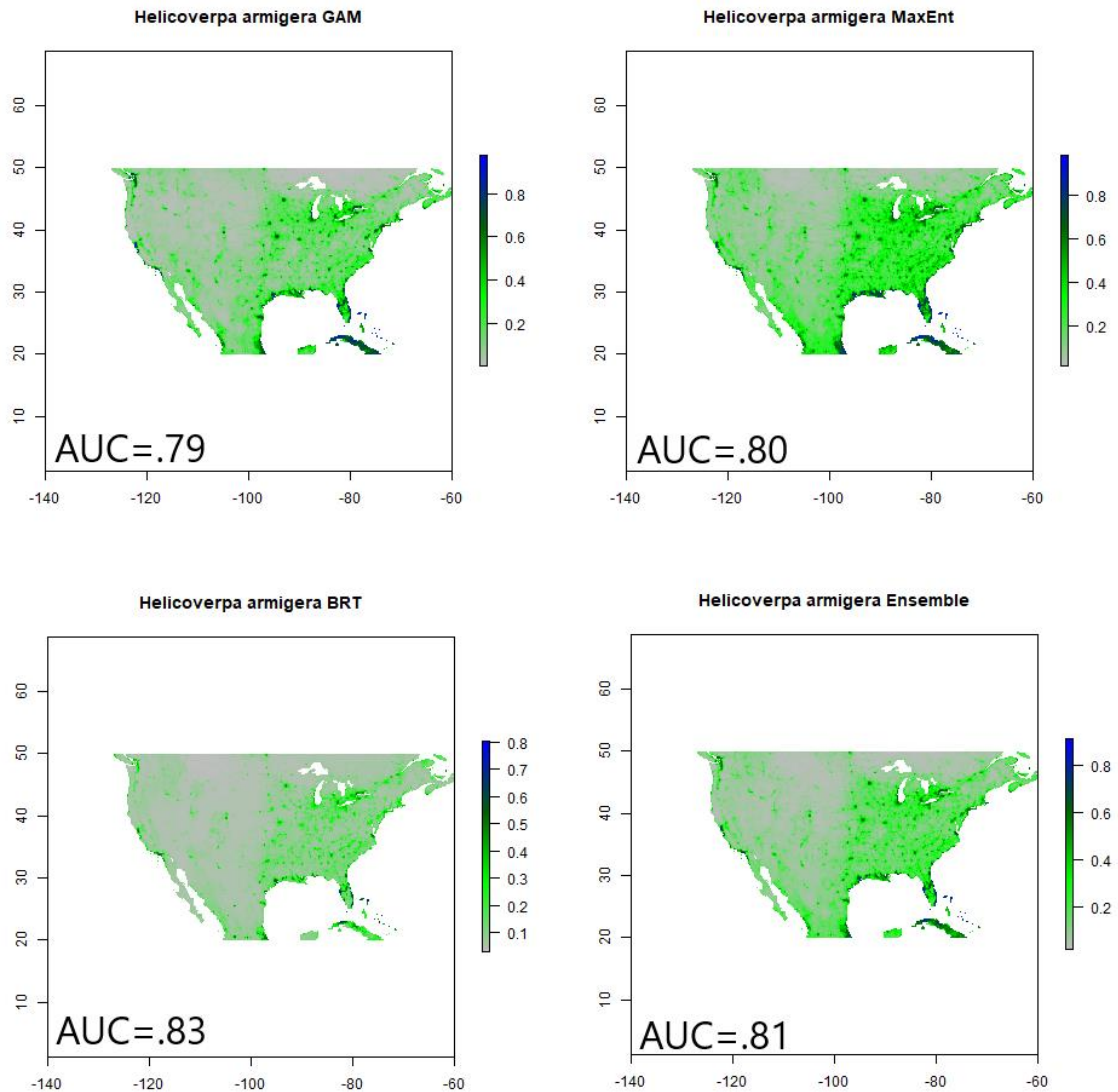
This species home range is in South America where it almost completely inhabits the southeastern coast (EPPO, 2005). There are over 2000 occurrences on GBIF but only 255 were left after the data cleaning process. Important predictors were human impact and radiation in the driest quarter. The projections of the BRT and MaxEnt models have a

high degree of agreement which is a strong indicator these models are representative of the true potential distribution (Figure 4). They project habitable areas mostly along the east coast with some suitability across the west coast as well. There was significant continental penetration projected by all models. The deep continental penetration is consistent with the existing distribution of *Diabrotica speciosa* which has become established deep into continental South America with a more condensed distribution around the coast. The models all projected some northeastern distribution in areas where significant areas of corn are grown which is especially noteworthy. The MaxEnt model was less conservative overall, except for the Great Lakes region. The GAM model projected many areas the other models did not corroborate such as in the Rocky Mountains and in the northwest United States. The areas projected solely by the GAM model are significantly less likely to represent the true distribution. This is because the GAM model is simple and does not account for interactions between predictors. The GAM model can make correct predictions about distribution that neither of the other two models project, but such predictions are usually on the borders of the MaxEnt and BRT projections rather than thousands of miles away. The highest habitat suitability based on model consensus is the east coast. More specifically Florida, the Carolinas, Delaware, and New Jersey all are states with high habitat suitability. Apart from the coast, the Great Lakes region is an area of high suitability according to the BRT and GAM models.

Figure 4*Habitat Suitability of Diabrotica Speciosa**Helicoverpa Armigera*

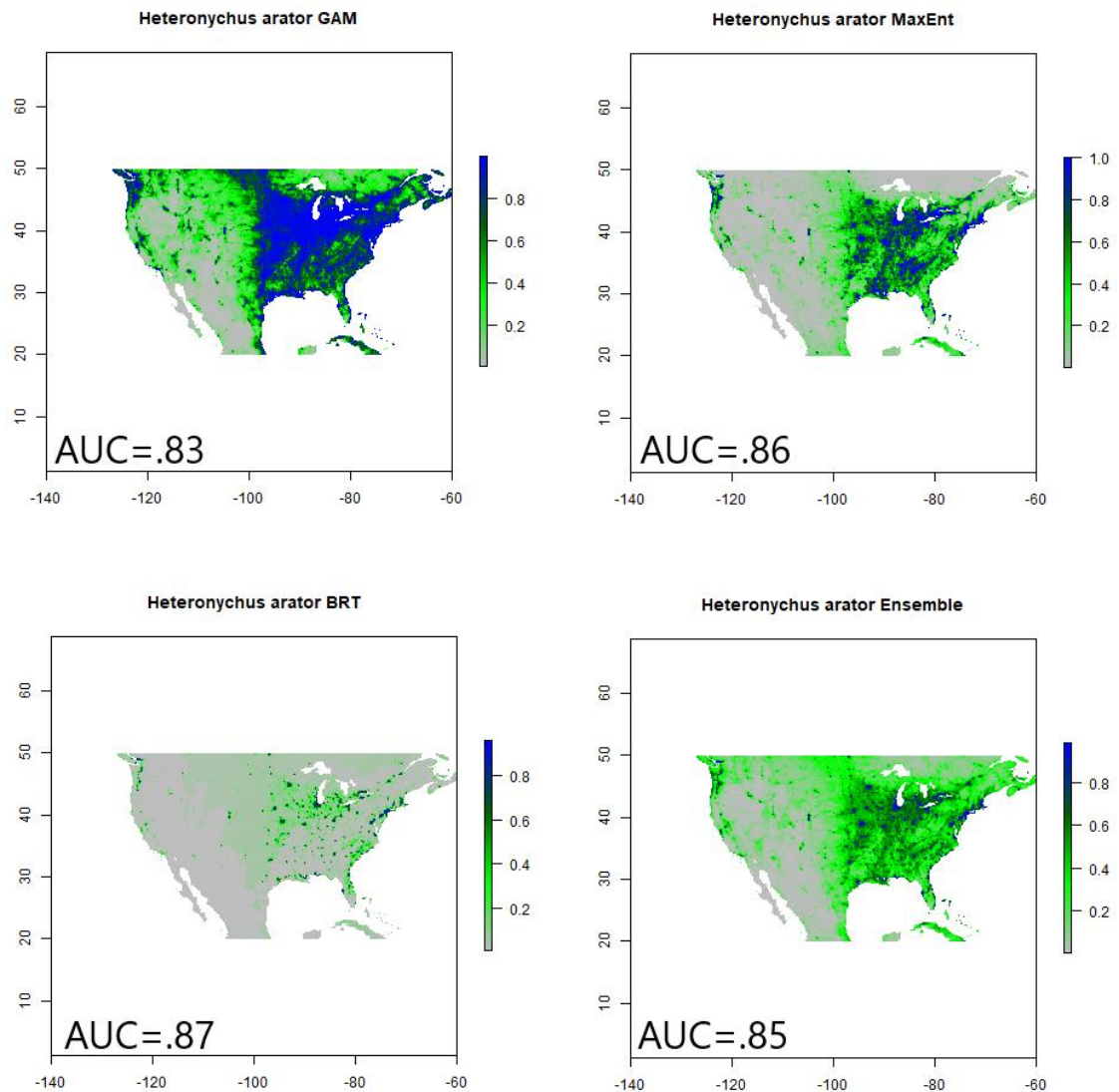
This species originated in South America but has spread across the world and this species has especially become established in Europe (Kriticos et al., 2015). Occurrences from all over the world were used to train this model and there were 2867 occurrences remaining to train the model after the data cleaning. The important predictors were mean

diurnal temperature range and human impact (Table 2). The models all had slightly lower AUC values than many of the other models (Figure 5, Table 1). Despite the slightly lower AUC values there are high levels of model agreement which indicates many of the areas it projects are likely to be accurate predictions. If additional predictors are identified in the future, it is possible a map with improved accuracy would be like these maps with only slight modifications.

Figure 5*Habitat Suitability of Helicoverpa Armigera**Heteronychus Arator*

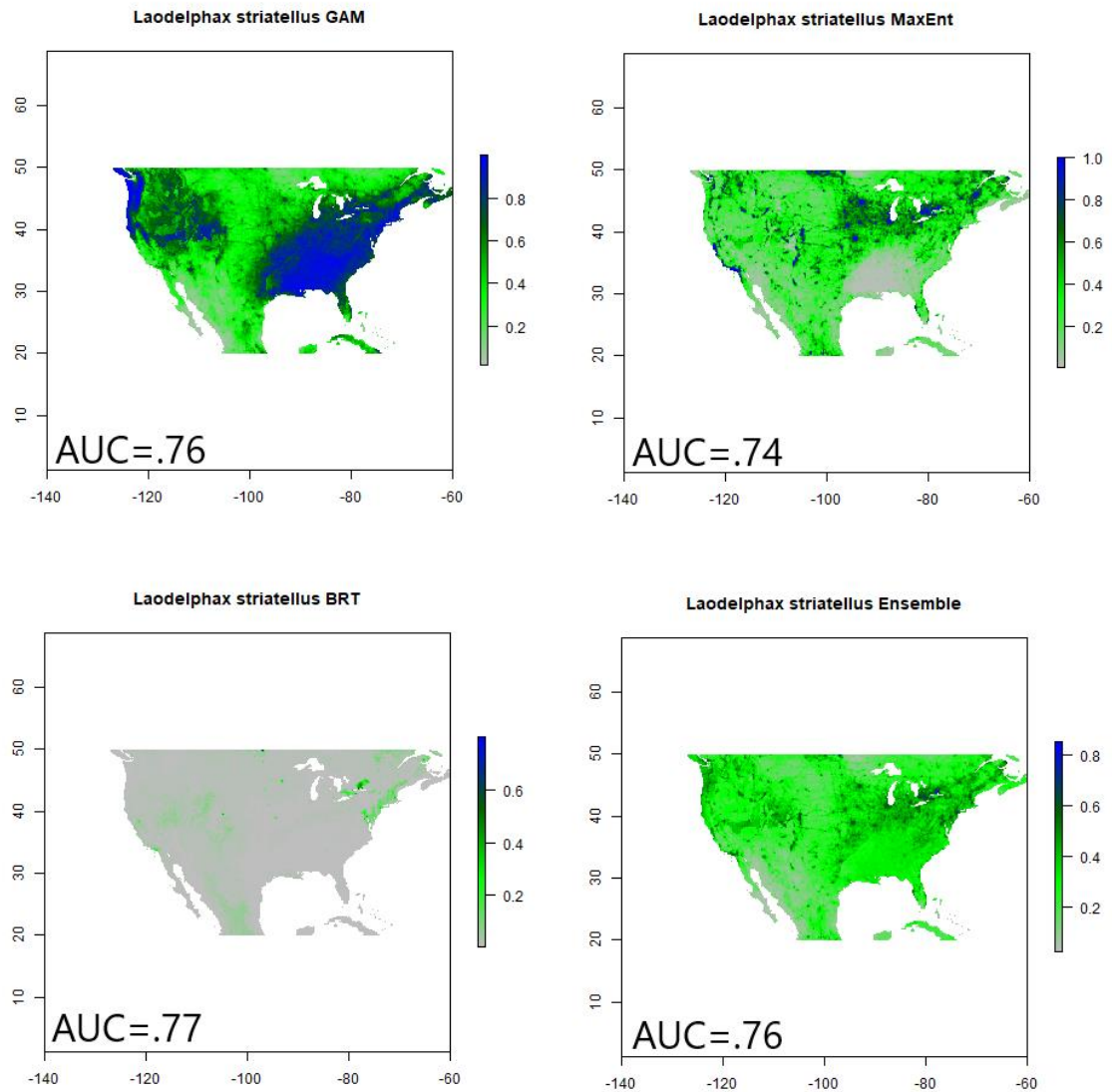
The home range used to train the model was southeastern Africa but there are occurrences in Australia as well (Learmonth, 2022). There were 195 occurrences left to train the model after the cleaning process. The most important predictors were human impact and isothermality (Table 2). These models are like *Helicoverpa armigera* in that

the AUC values are slightly low and there are high levels of model agreement (Figure 5 and Figure 6). The habitable area proposed by these models is broad and slightly patchy with the highest concentration of habitable area east of -100 longitude, surrounding the Great Lakes, and near both coasts. The broad distribution indicates *Heteronychus arator* would find the United States to be a particularly hospitable climate. States in the western United States such as Utah, Nevada, and Idaho have less projected habitability than the rest of the country but still have patches of significant suitability.

Figure 6*Habitat Suitability of Heteronychus Arator**Laodelphax Striatellus*

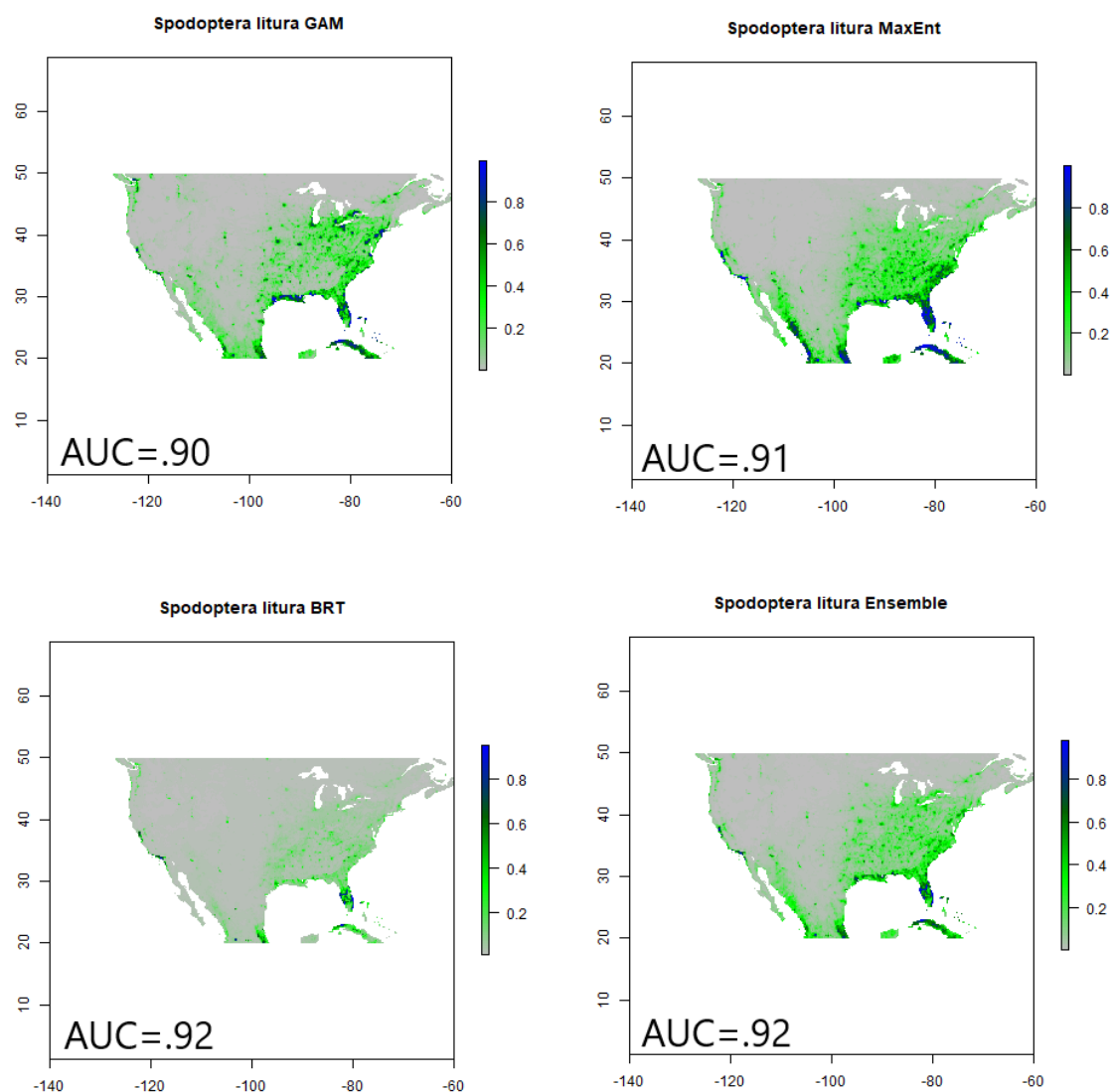
The home range used to train the models was Asia, which is its native range, but there were occurrences from Europe used as well where it has become widely distributed (Jeong et al., 2016). There were 167 occurrences remaining to train the model after the data cleaning process. The important predictors were human impact, radiation in driest

quarter, and precipitation seasonality (Table 2). Most of the models projected a high level of habitat suitability across the United States. The BRT model projected habitat suitability on the northeast coast, around Lake Ontario and Lake Erie, and around the Rocky Mountains. These areas are supported by the MaxEnt model, which makes them the most likely to be accurate predictions. The area is also supported by the GAM model which projects all land east of -105 longitude as extremely habitable. The land west of -105 not being projected to be particularly suitable, even by the most liberal model is indicative of low habitat suitability. In that region there are a few islands of suitability especially on the coast but the highest suitability in that region is in Canada on the coast.

Figure 7*Habitat Suitability of Laodelphax Striatellus**Spodoptera Litura*

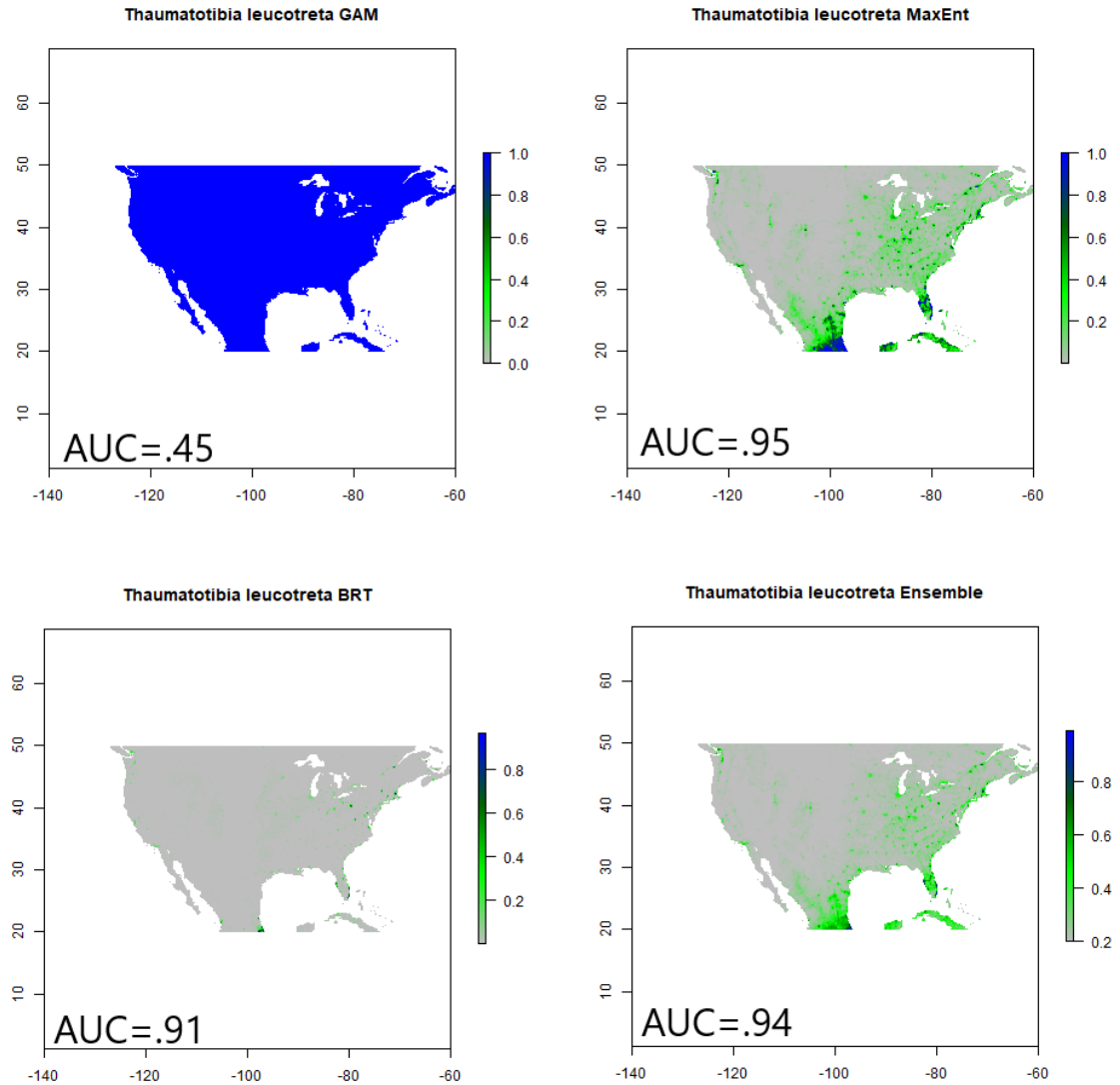
This pest was native to South America but has spread worldwide (Cui et al., 2020). Most occurrences used to train these models were from Asia, but there were occurrences in Australia as well where it has also become established (Thomson, 1965). There were 1139 occurrences remaining to train the model after data cleaning. The most

important predictors were human impact, temperature seasonality, and radiation in the driest quarter. All three models and their ensemble exhibited AUC values >0.9 which indicated this model does well at making accurate distribution projections (Figure 8). Florida is at very suitable and could act as a foothold for *Spodoptera litura* to become distributed through much of the land corn is grown domestically. This is especially dangerous as all of the models project habitat suitability in the Midwest which is the region where the most corn is grown.

Figure 8*Habitat Suitability of Spodoptera Litura**Thaumatotibia Leucotreta*

The home range is the Ethiopian zoogeographic province and most of the occurrences used to train these models were from Africa, but there are a few occurrences from Europe as well (Venette et al., 2003). There were 102 occurrences remaining after cleaning to train the models. Table 2 shows the important predictors were human impact, annual

precipitation, and radiation in wettest quarter. The GAM model could not discriminate any important predictors and predicts that everywhere in the United States is highly habitable (Figure 9). The GAM prediction results in a unusual AUC value of 0.45 because absence data was generated around occurrences and there is always a prediction of habitat suitability on every absence point. Because GAM predictions were no better than random, GAM was excluded from the ensemble. The ensemble demonstrates Indiana, Ohio, and Kentucky are all midwestern states with significant areas of habitat suitability. All models agree there is suitable habitat in the Midwest which makes this pest particularly dangerous to maize.

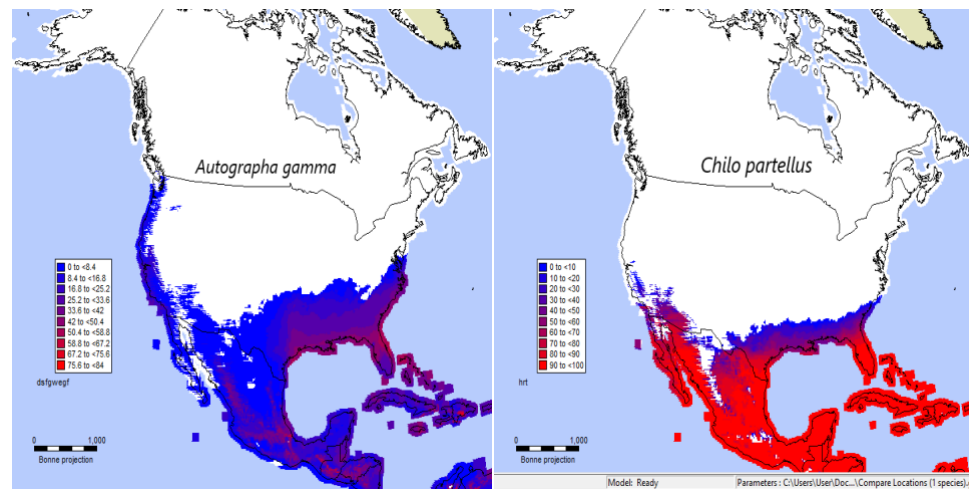
Figure 9*Habitat Suitability of Thaumatotibia Leucotreta***CLIMEX**

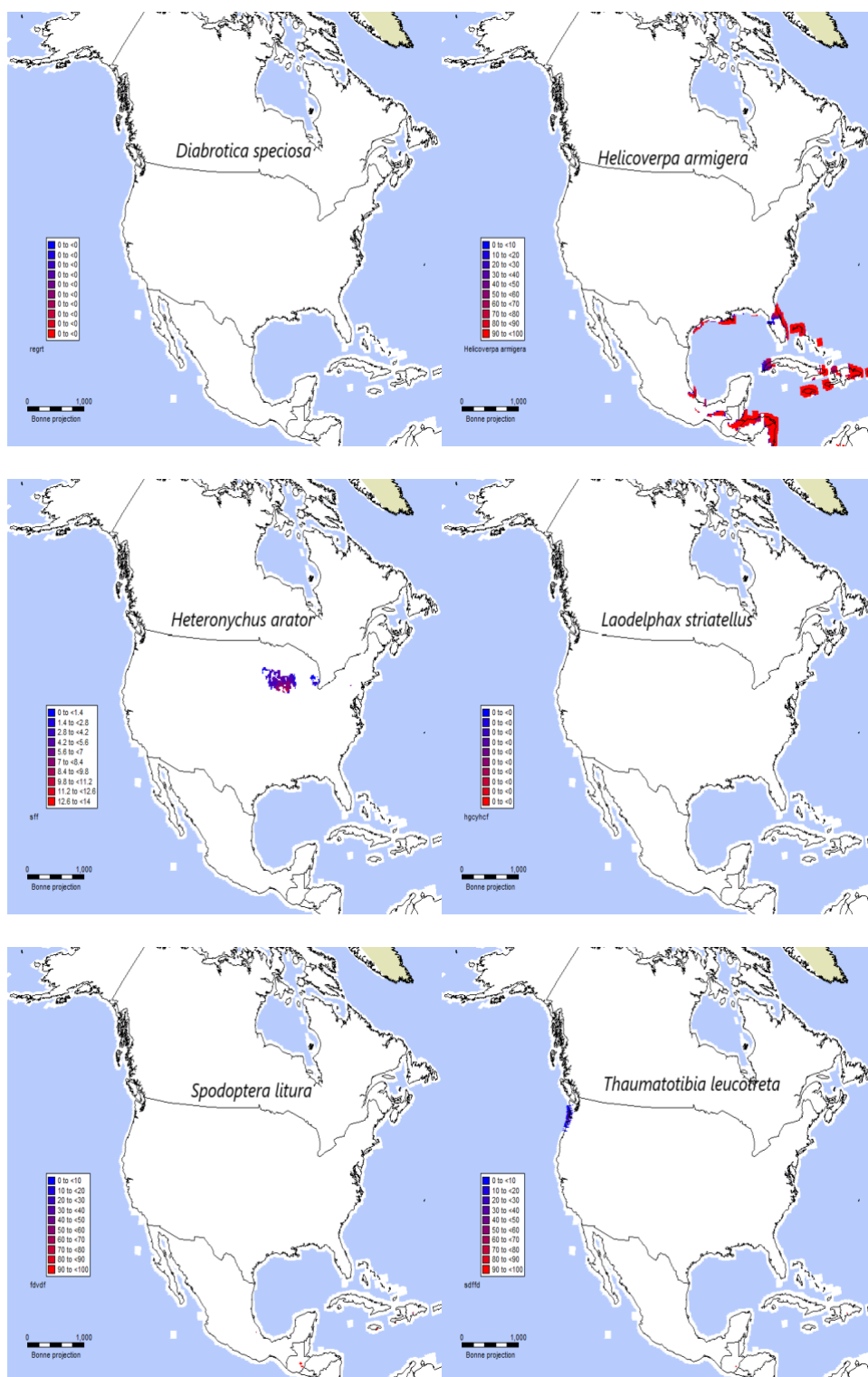
Five of the eight species did not have habitable areas within the contiguous United States excluding Hawaii (Figure 9). *Heteronychus arator* had a small region of habitat suitability near Iowa and Nebraska which are two of the states with the highest corn yield. *Chilo partellus* and *Autographa gamma* had large areas of habitat suitability

in the southern United States. *Chilo partellus* and *Autographa gamma* differ significantly from the correlative models in that the distribution projected by CLIMEX clearly is defined solely by a latitudinal gradient while the correlative models had high levels of suitability across the eastern US. The other six species differed from the correlative models in that there was little to no suitable habitat in the United States.

Figure 10

CLIMEX Maps





Note. The legend shows EI values which stands for Ecoclimatic Index. EI is a combination of a growth index and a stress index that base growth and stress on

physiological tolerances. An EI value between 0-10 is slightly habitable, 10-20 is habitable, and above 20 is highly habitable (Jung et al., 2016).

Differences in Model Projections

One of the trends apparent in the distribution maps is that the GAM model predictions were the most liberal by a significant margin. This could be the result of the model's simplicity not capturing as many of the predictor effects or interactions as that the more sophisticated models. If that is the case, the more liberal estimate is less accurate which is supported by lower AUC values (Table 1). *Spodoptera litura* and *Autographa gamma* had GAM maps that were more conservative than most of the other GAM maps and had a higher AUC value. These are the instances where GAM models are particularly useful. This is because MaxEnt and BRT models can be overly conservative, and the simple GAM model can improve total accuracy by supplementing complex models in an ensemble. The BRT models were almost always the most conservative and areas of highest habitat suitability projected by the BRT models are almost always also projected by the MaxEnt models. This is one of the best indicators these maps are accurate as the models are built using entirely different methods yet arrive similar predictions.

Overall Trends

The highest AUC values for any species alternated between MaxEnt and BRT, which was expected as no one model is the most predictive in all cases (Table 1). That is a primary reason an ensemble model is one of the most effective tools for creating accurate predictive maps. The ensemble AUC values were always close to the highest value and despite not having the highest AUC values, may be the most accurate of the predictions. This is because it is less eroded by the weaknesses of any one model and can

provide additional information about distribution that the model with the highest AUC value is missing. The most important predictor in the greatest number of models was human impact (Table 2) which is not surprising as that predictor entails many powerful drivers such as land use, crop locations, and pesticide use.

Table 1*AUC Values of GAM, MaxEnt, BRT, and Ensemble Models*

Species	GAM	MaxEnt	BRT	ENSEMBLE
<i>Autographa gamma</i>	0.94	0.95	0.96	0.95
<i>Chilo partellus</i>	0.78	1	0.92	0.94
<i>Diabrotica speciosa</i>	0.91	0.94	0.93	0.93
<i>Helicoverpa armigera</i>	0.79	0.80	0.83	0.81
<i>Heteronychus arator</i>	0.83	0.86	0.87	0.85
<i>Laodelphax striatellus</i>	0.76	0.74	0.77	0.76
<i>Spodoptera litura</i>	0.90	0.91	0.92	0.92
<i>Thaumatotibia leucotreta</i>	0.45	0.95	0.91	0.94

Table 2*Predictors in Descending Order of Importance*

Species	Most important Predictor	Most important Predictor	Most important Predictor
	GAM	MaxEnt	BRT
<i>Autographa gamma</i>	Temperature Seasonality	Human Impact	Human Impact
	Annual Precipitation	Temperature Seasonality	Maximum Temperature in Warmest Week
	Precipitation in Warmest Quarter		
	Human Impact		
<i>Chilo partellus</i>	All Predictors Equal	Human Impact	Precipitation in Coldest Quarter
<i>Diabrotica speciosa</i>	Human Impact	Human Impact	Radiation in Driest Quarter
	Radiation in Driest Quarter		
<i>Helicoverpa</i>	Mean Durnal Temperature Range	Human Impact	Mean Durnal Temperature Range
<i>armigera</i>	Temperature Seasonality	Mean Durnal Temperature Range	Human Impact
	Moisture Seasonality		Temperature Seasonality
	Human Impact		Radiation Seasonality
<i>Heteronychus</i>	Human Impact	Human Impact	Isothermality
<i>arator</i>			
<i>Spodoptera litura</i>	Radiation in Wettest Quarter	Human Impact	Radiation in Driest Quarter
	Isothermality	Temperature Seasonality	Temperature Seasonality

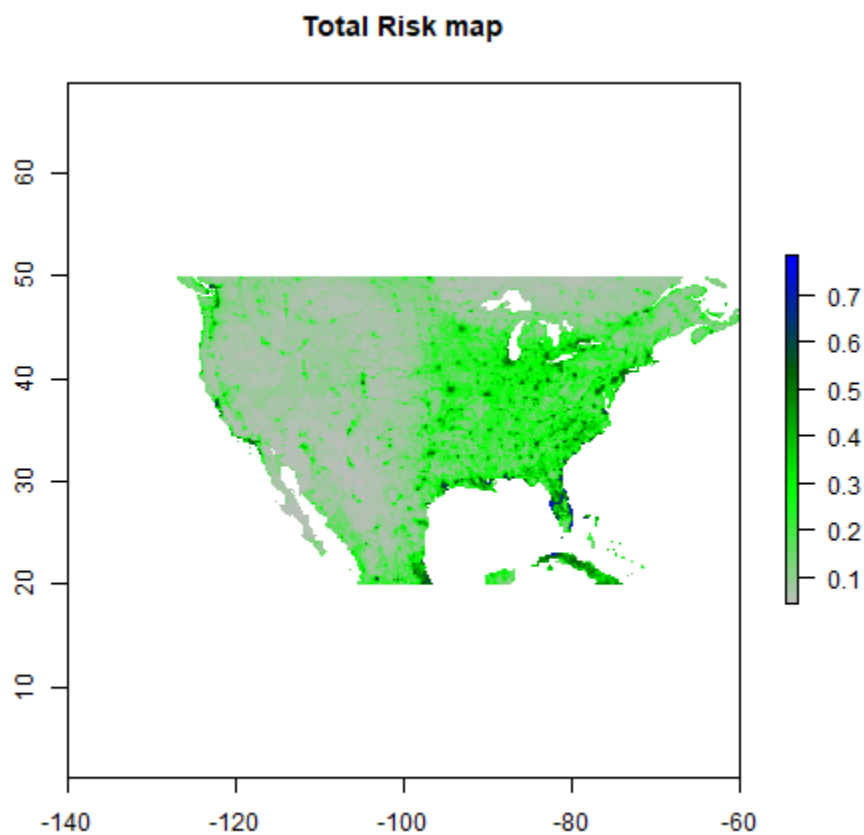
(continued)

Species	Most important Predictor	Most important Predictor	Most important Predictor
	GAM	MaxEnt	BRT
.	Temperature Seasonality		Precipitation in Warmest Quarter
	Human Impact		Mean Temperature in Driest Quarter
<i>Thaumatotibia</i>	All Predictors Equal	Annual Precipitation	Radiation Wettest Quarter
<i>leucotreta</i>			Human Impact

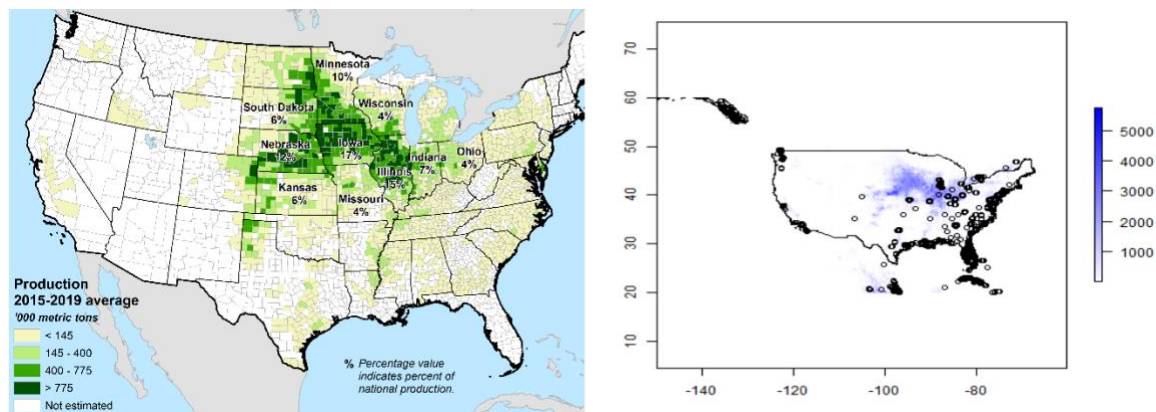
Everywhere in the United States is habitable to at least one pest species with an AUC threshold of .5, and there are some areas of extremely habitat suitability which can be seen on the map of habitability constructed from a weighted average of all eight ensemble models (Figure 11). The high suitability areas, especially in Florida have abiotic conditions suitable for up to six of the pest species. The east coast is more susceptible to corn pests than the west coast. Florida is particularly suitable as it has multiple regions with climate suitable to five or six pest species. There are areas that Corn is grown in areas that can host multiple of the pest species (Figure 12).

Figure 11

Weighted Average Pest Risk Map



Note. This figure was created using a weighted average of all eight ensemble models. Each raster was multiplied by its AUC score and added to the risk map which was divided by the total values of AUC scores.

Figure 12*Corn in High-Risk Areas Map*

Note. This figure shows the harvested area of corn in the United States. The map on the left shows where corn is grown domestically and is provided by the National Agriculture Statistics Service. The blue legend on the right represents land where corn was harvested in hectares based on the CHEMGRIDS crop harvest data. The black points on the right plot indicate habitability > 0.5 in one location for 3 or more pest species. This figure highlights areas that corn is growing in areas of high risk, where there are multiple invasive corn species that could infest the crop.

CHAPTER IV

Discussion

These models have identified suitable habitats in the United States for every maize pest. The pests that pose the most risk to land where maize is cultivated are *Autographa gamma*, *Diabrotica speciosa*, *Helicoverpa armigera*, *Heteronychus arator*, and *Spodoptera litura*. The remaining pests *Chilo partellus*, *Laodelphax striatellus* and *Thaumatotibia leucotreta* pose less risk because their suitable habitat does not coincide as much with land that is cultivated for maize. Of the pests that pose less risk some caution needs to be taken with *Chilo partellus* because there were not many occurrences available to train the model and it would be catastrophic to underestimate the threat posed by any of these invasive pests. The highest risk area is on the east coast. The west coast, Great Lakes, and Rocky Mountains are all also at risk of multiple corn pests. Some of the areas where corn pests find highly habitable do not have a lot of corn growing in the area, but these pests can eat many kinds of crop which would allow them to get a foothold while they make their way to where corn is grown. The lowest levels of habitat suitability were in the western continental United States but there are some patches of risk in that region as well. The most significant finding of this study is the identification of high-risk areas where corn is grown, such as the habitable areas in Iowa, Illinois, and Nebraska. These three states produce the most maize and all of them have ideal habitats to host an invasive maize pest (USDA, 2021).

Modeling Invasive Species

Modeling invasive pests may be difficult when species do not occupy an easily identifiable home range (Gallien et al., 2012). Using occurrences from the invaded habitat decreases the probability of capturing the true fundamental niche. For example, in invaded habitat, species may exist in sink populations which occur where the invasive pest has spread but in which it cannot maintain positive population growth. These species may also be evolving in response to changes in environmental conditions either in response to being targeted for pest management or being introduced to a new area (Gallien et al., 2012). Those assumption violations weaken the predictive capacity of these niche models which is why multiple models need to be used to find model agreement.

Human Impact as a Predictor

This study demonstrates that ecological niche models can benefit from the inclusion of a human impact layer, which was overall the most important predictor. Note, however, that human impact layers may coincide with occurrence points because convenience sampling most likely occurs in areas of high human impact. This problem may be partially mitigated by our method of generating absences that are also concentrated near human impact but the association between human impact and convenience sampling is still a potential explanation for the importance of the predictor.

Differences in Model Projections

MaxEnt models usually had a similar and slightly more liberal prediction compared to their corresponding BRT model. The most significant exception was the

BRT projection of *Diabrotica speciosa* having more habitat suitability near the Great Lakes. The liberal MaxEnt projections may be stem from the models having unbound response curves. As a result of such response curves, habitats with predictor values above or below what the model has seen before can be artificially suitable and hence the overall predictions slightly more liberal. Another explanation is that the regularization of the MaxEnt models resulting in simpler functions which make the predictions less precise and hence more liberal. This trend of simple models generating liberal predictions is why the simple GAM models had the most liberal predictions by a significant margin. Extremely liberal projections such as *Autographa gamma* habitability in the Rocky Mountains are likely more liberal than the other models because GAM models do not account for the interactions between predictors, unlike the more sophisticated models. The GAM models are based on statistical methods while BRT and MaxEnt models use a combination of statistics and machine learning. This discrepancy results in BRT and MaxEnt models having higher predictive accuracy because machine learning allows iterative improvement of the model, something not included in GAM fitting (Kutywayo et al., 2013).

Potential Model Performance Erosion

A few models had relatively low AUC values in comparison to the models generated for other species. *Helicoverpa armigera*, *Heteronychus arator* and *Laodelphax striatellus* all had ensemble AUC values below 0.85. Such values indicate the models are still predictive as they are much greater than 0.5 but these models had some factor that impeded model performance compared to the other models in this study. The most likely explanations are that an important predictor that drives this distribution was not used to

train the models. Such a predictor may be an abiotic factor not included here, or alternatively may not be currently modellable such as dispersal or a biotic interaction with another species.

Model Usage and Improvement

This study has generated maps that can be used to direct surveys (Guisan et al., 2006). Survey effort is limited and should not be wasted in areas where a species cannot sustain a population. Knowledge about which locations have the potential to sustain a population should help triage limited funds and scientific resources (Guisan et al., 2006). Ideally these maps will result in sampling strategies that identify these pests soon after they are introduced, before there is much dispersal. This type of study also conveys novel information about how climactic conditions affect the distributions of these pest species (Kulhanek et al., 2011). The models should be updated regularly as more occurrences become available as each occurrence point could convey potentially new information that would improve model performance (Elith, 2006). Model performance would be significantly improved with the inclusion of any absences generated by sample effort especially in areas where pests have been introduced that are projected to be habitable.

REFERENCES

- Anderson, R. P. (2003). Real vs. artefactual absences in species distributions: Tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography*, 30(4), 591-605.
- Ashby, J., Moreno-Madriñán, M. J., Yiannoutsos, C. T., & Stanforth, A. (2017). Niche modeling of dengue fever using remotely sensed environmental factors and boosted regression trees. *Remote Sensing*, 9(4), 328.
- Beeman, S. P., Morrison, A. M., Unnasch, T. R., & Unnasch, R. S. (2021). Ensemble ecological niche modeling of West Nile virus probability in Florida. *Plos one*, 16(10), e0256868.
- Bellard, C., Jeschke, J. M., Leroy, B., & Mace, G. M. (2018). Insights from modeling studies on how climate change affects invasive alien species geography. *Ecology and Evolution*, 8(11), 5688-5700.
- Bivand R., & Rundel C. (2021). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). <https://CRAN.R-project.org/package=rgeos>
- Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R* (Vol. 747248717, pp. 237-268). Springer.
- Bivand, R., & Lewin-Koh, N. (2021). maptools: Tools for handling spatial objects. <https://CRAN.R-project.org/package=maptools>
- Bivand, R., Keitt, T., & Rowlingson, B. (2021). rgdal: Bindings for the 'geospatial' data abstraction library. <https://CRAN.R-project.org/package=rgdal>
- Bradshaw, C.J., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.M., Simard, F. and Courchamp, F., (2016). Massive yet grossly

underestimated global costs of invasive insects. *Nature Communications*, 7(1), 1-8.

Byeon, D. H., Jung, S., & Lee, W. H. (2018). Review of CLIMEX and MaxEnt for studying species distribution in South Korea. *Journal of Asia-Pacific Biodiversity*, 11(3), 325-333.

Cao, Q., Xu, W. Y., Gao, Q., Jiang, Z. H., Liu, S. Y., Fang, X. D., ... & Wang, X. B. (2018). Transmission characteristics of barley yellow striate mosaic virus in its planthopper vector *Laodelphax striatellus*. *Frontiers in Microbiology*, 9, 1419.

Cerasoli, F., Iannella, M., D'Alessandro, P., & Biondi, M. (2017). Comparing pseudo-absences generation techniques in Boosted Regression Trees models for conservation purposes: A case study on amphibians in a protected area. *PLoS One*, 12(11), e0187589.

Chatterjee, S. & Hadi, A. S. (2006). *Regression analysis by example*. John Wiley and Sons.

Ciss, M., Biteye, B., Fall, A. G., Fall, M., Gahn, M. C. B., Leroux, L., & Apolloni, A. (2019). Ecological niche modelling to estimate the distribution of *Culicoides*, potential vectors of bluetongue virus in Senegal. *BMC Ecology*, 19(1), 1-12.

Compton, T. J., Morrison, M. A., Leathwick, J. R., & Carbines, G. D. (2012). Ontogenetic habitat associations of a demersal fish species, *Pagrus auratus*, identified using boosted regression trees. *Marine Ecology Progress Series*, 462, 219-230.

Crop protection compendium: Global module. (2010). Commonwealth Agricultural Bureau International. <https://www.cabi.org/cpc>.

- Cui, Y., Ren, Y. D., Lyu, M., Zheng, S. C., Feng, Q. L., & Xiang, H. (2020). Genomic divergences between the two polyphagous *Spodoptera* relatives provide cues for successful invasion of the fall armyworm. *Insect Science*, 27(6), 1257-1265.
- Daiber, C. C. (1980). A study of the biology of the false codling moth [*Cryptophlebia leucotreta* (Meyr.)]: The adult and generations during the year. *Phytophylactica* 12: 187-193.
- Denno, R. F., & Roderick, G. K. (1990). Population biology of planthoppers. *Annual Review of Entomology*, 35(1), 489-520.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J. and Münkemüller, T. 2012. Collinearity: A review of methods to Deal with it and a simulation study evaluating their performance. *Ecography* 35, 001-020
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J. and Münkemüller, T. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46.
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., . . . E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>

- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43-57.
- EPPO (2005). *Diabrotica speciosa*. *OEPP/EPPO Bulletin* 35(3), 374-376.
<https://doi.org/10.1111/j.1365-2338.2005.00904.x>
- Escobar, L. E., Qiao, H., Cabello, J., & Peterson, A. T. (2018). Ecological niche modeling re-examined: A case study with the Darwin's fox. *Ecology and Evolution*, 8(10), 4757-4770.
- FAO Statistical Databases (2019). FAO Food and Agriculture Organization of the United Nations (FAO). <http://faostat.fao.org>.
- Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and distributions*, 13(3), 252-264.
- Fick, S.E. & R.J. Hijmans, (2017). WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 3(12): 4302-4315.
- Fielding, A. H. & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38-49.

- Fitzpatrick, M. C., Sanders N.J., Ferrier S., Longino J.T., Weiser M.D., & Dunn R.R. (2011). Forecasting the future of biodiversity: A test of single- and multi-species models for ants in North America. *Ecography* 34, 836–47.
- Fitzpatrick, M., & Lisk, M. (2016). Using the R package gdm to analyze and map biodiversity patterns. *R Package Version*, 1(11).
- Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E., & Thuiller, W. (2012). Invasive species distribution models—how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, 21(11), 1126-1136.
- GBIF. Global Biodiversity Information Facility. <http://www.gbif.org>
- Giannini, T. C., Siqueira, M. F., Acosta, A. L., Barreto, F. C., Saraiva, A. M., and Alvesdos-Santos, I. (2012). Current challenges of species distribution predictive modelling. *Rodrigu'esia*, 63(3), 733–749.
- Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Townsend Peterson, A., Loiselle, B. A., & NCEAS Predicting Species Distributions Working Group (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, 45(1), 239-247.
- Grinnell, J. (1917). Field tests of theories concerning distributional control. *The American Naturalist*, 51(602), 115-128.
- Gu, H., Wang, J., Ma, L., Shang, Z., & Zhang, Q. (2019). Insights into the BRT (Boosted Regression Trees) method in the study of the climate-growth relationship of Masson Pine in Subtropical China. *Forests*, 10(3), 228.
- Guiasu, S., & Shenitzer, A. (1985). The principle of maximum entropy. *The Mathematical Intelligencer*, 7(1), 42-48.

- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A., & Zimmermann, N. E. (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology*, 20(2), 501-511.
- Guisan, A., Zimmermann, N. E., Cola, V. D., Georges, D., Psomas, A., & Thuiller, W. (2021). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.
- Hallman, T. A., & Robinson, W. D. (2020). Comparing multi-and single-scale species distribution and abundance models built with the boosted regression tree algorithm. *Landscape Ecology*, 35(5), 1161-1174.
- Harington, J. S. (1953). Observations on the biology, the parasites and the taxonomic position of the black maize beetle-*Heteronychus sanctae-helenae* Blanch. *South African Journal of Science*, 50(1).
- Hastie, T. J.; Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall/CRC.
- Hijmans R.J., Phillips, S., Leathwick, J. and Elith J. (2021). dismo: Species Distribution Modeling. R package version 1.3-5. <https://CRAN.R-project.org/package=dismo>
- Hijmans, R. J. (2021). *Feed the future: The US government's global hunger and food security initiative*. Spatial Data Science. <http://www.rspatial.org>
- Hijmans, R. J. (2021). raster: Geographic data analysis and modeling. R package version 3.5-2. <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J., Schreuder, M., De la Cruz, J., & Guarino, L. (1999). Using GIS to check co-ordinates of genebank accessions. *Genetic Resources and Crop Evolution*, 46(3), 291-296.

- Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1), 10-18.
- Industry market research, reports, and statistics (2022, March 1). IBISWorld.
<https://www.ibisworld.com/united-states/market-research-reports/pest-control-industry/>
- Ireland, K. B., & Kriticos, D. J. (2019). Why are plant pathogens under-represented in eco-climatic niche modelling? *International Journal of Pest Management*, 65(3), 207-216
- Ireland, K. B., Hardy, G. E. S. J., & Kriticos, D. J. (2013). Combining inferential and deductive approaches to estimate the potential geographical range of the invasive plant pathogen, *Phytophthora ramorum*. *PLoS One*, 8(5), e63508.
- Isiguro, K., & Yanase M, (1991). Effects of feeding by white-backed planthopper, *Sogatella furcifera* Horvath, small brown planthopper, *Laodelphax striatellus* Fallen [L. striatella] and brown rice planthopper, *Nilaparvata lugens* Stal for quality on brown rice. *Annual Report of the Society of Plant Protection of North Japan*, 42, 88-90
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
- Jeong, I. H., Lee, S. W., Choi, B. R., Lee, S. H., & Kwon, D. H. (2016). Monitoring and evaluation of differential insecticide resistance profiles in the immigrant vs. indigenous populations of the small brown planthopper (*Laodelphax striatellus* Fallén) in Korea. *Journal of Asia-Pacific Entomology*, 19(1), 247-252.

Jon Clayden (2019). shades: Simple colour manipulation. R package version 1.4.0.

<https://CRAN.R-project.org/package=shades>

Jung, J. M., Lee, W. H., & Jung, S. (2016). Insect distribution in response to climate change based on a model: Review of function and use of CLIMEX.

Entomological Research, 46(4), 223-235.

Klopfenstein, T. J., Erickson, G. E., & Berger, L. L. (2013). Maize is a critically important source of food, feed, energy and forage in the USA. *Field Crops Research*, 153, 5-11.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., ... & Wilting, A. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, 19(11), 1366-1379.

Kriticos, D. J., Maywald, G. F., Yonow, T., Zurcher, E. J., Herrmann, N. I., & Sutherst, R. W. (2015). CLIMEX Version 4: Exploring the effects of climate on plants, animals and diseases. *Canberra: CSIRO*, 184.

Kriticos, D. J., Ota, N., Hutchison, W. D., Beddow, J., Walsh, T., Tay, W. T., Borchert, D. M., Paula-Moraes, S. V., Czapak, C., & Zalucki, M. P. (2015). The potential distribution of invading *Helicoverpa armigera* in North America: Is it just a matter of time? *PloS one*, 10(3), e0119618.

Kriticos, D. J., Potter, K. J., Alexander, N. S., Gibb, A. R., & Suckling, D. M. (2007). Using a pheromone lure survey to establish the native and potential distribution of an invasive Lepidopteran, *Uraba lugens*. *Journal of Applied Ecology*, 44(4), 853-863.

- Krysan, J.L (1986). *Introduction: Biology, distribution, and identification of pest Diabrotica*. Springer.
- Kulhanek, S. A., Leung, B., & Ricciardi, A. (2011). Using ecological niche models to predict the abundance and impact of invasive species: Application to the common carp. *Ecological Applications*, 21(1), 203-213.
- Kumar, S., Neven, L. G., Zhu, H., & Zhang, R. (2015). Assessing the global risk of establishment of *Cydia pomonella* (Lepidoptera: Tortricidae) using CLIMEX and MaxEnt niche models. *Journal of Economic Entomology*, 108(4), 1708-1719.
- Kumar, S., Yee, W. L., & Neven, L. G. (2016). Mapping global potential risk of establishment of *Rhagoletis pomonella* (Diptera: Tephritidae) using MaxEnt and CLIMEX niche models. *Journal of Economic Entomology*, 109(5), 2043-2053.
- Kutywayo, D., Chemura, A., Kusena, W., Chidoko, P., & Mahoya, C. (2013). The impact of climate change on the potential distribution of agricultural pests: The case of the coffee white stem borer (*Monochamus leuconotus* P.) in Zimbabwe. *PLoS One*, 8(8), e73432.
- Levin, S. A., Carpenter, S. R., Godfray, C. H. J., Kinzig, A. P., Loreau, M., Losos, J. B., Walker, B., & Wilcove, D. S. (2012). *The Princeton Guide to Ecology*. Princeton University Press.
- Lira-Noriega, A., & Peterson, A. T. (2014). Range-wide ecological niche comparisons of parasite, hosts and dispersers in a vector-borne plant parasite system. *Journal of biogeography*, 41(9), 1664-1673.

- Maggi, F., Tang, F. H., la Cecilia, D., & McBratney, A. (2019). PEST-CHEMGRIDS, global gridded maps of the top 20 crop-specific pesticide application rates from 2015 to 2025. *Scientific Data*, 6(1), 1-20.
- Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., Cassone, B. J., Sevgan, S., Nyasani, J. O., Kusia, E., Kumar, P. L., Niblett, C. L., Kiggundu, A., Asea, G., Pappu, H. R., Wangai, A., Prasanna, B. M., & Redinbaugh, M. G. (2015). Maize Lethal Necrosis (MLN), an Emerging Threat to Maize-Based Food Security in Sub-Saharan Africa. *Phytopathology*®, 105(7), 956–965. <https://doi.org/10.1094/phyto-12-14-0367-fi>
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15(1), 59-69.
- Mau-Crimmins, T. M., Schussman, H. R., & Geiger, E. L. (2006). Can the invaded range of a species be predicted sufficiently using only native-range data?: Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling*, 193(3-4), 736-746.
- Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10), 1058-1069.
- Mitchell, T. (1997). *Machine learning*. McGraw Hill.
- Moffett, A., Shackelford, N., & Sarkar, S. (2007). Malaria in Africa: Vector species' niche models and relative risk maps. *PLoS One*, 2(9), e824.

- Mohyuddin, A. I. (1971). Comparative biology and ecology of *Apanteles flavipes* (Cam.) and *A. sesamiae* Cam. as parasites of graminaceous borers. *Bulletin of Entomological Research*, 61(1), 33-40.
- Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., & Toxopeus, A.G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37(2), 191-203.
- Nelder, J., Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)* 135(3), 370–384.
<https://doi.org/10.2307/2344614>
- Otuka, A. (2013). Migration of rice planthoppers and their vectored re-emerging and novel rice viruses in East Asia. *Frontiers in Microbiology*, 4, 309.
- Pearman, P. B., D'Amen, M., Graham, C. H., Thuiller, W., & Zimmermann, N. E. (2010). Within-taxon niche structure: Niche conservatism, divergence and predicted effects of climate change. *Ecography*, 33(6), 990-1003.
- Pebesma, E.J., Bivand, R.S. (2005). Classes and methods for spatial data. *The R Journal*, R News 5(2), 9-13. https://cran.r-project.org/doc/Rnews/Rnews_2005-2.pdf.
- Pérez, J. E., Nirchio, M., Alfonsi, C., & Muñoz, C. (2006). The biology of invasions: The genetic adaptation paradox. *Biological Invasions*, 8(5), 1115-1121.
- Peterson, A. T., & Holt, R. D. (2003). Niche differentiation in Mexican birds: Using point occurrences to detect ecological innovation. *Ecology Letters*, 6(8), 774-782.
- Peterson, A. T., & Vieglais, D. A. (2001). Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem: A new approach to ecological niche modeling, based on new tools drawn from

biodiversity informatics, is applied to the challenge of predicting potential species' invasions. *BioScience*, 51(5), 363-371.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4), 231-259.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181-197.

Plant Protection and Quarantine, Emergency and Domestic Programs (2010).

Plant Protection and Quarantine. (2019). Cooperative Agricultural Pest Survey (CAPS) Pest Datasheet for *Chilo partellus* (Crambidae): Spotted stem borer. United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine (PPQ), Raleigh, NC.

Plant Protection Service and Central Science Laboratory (2007). Joint Pest Risk Analysis for *Helicoverpa armigera*.

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwinmqT3mJX5AhUamWoFHczPBVkQFnoECAQQAQ&url=https%3A%2F%2Fpra.eppo.int%2Fgetfile%2F81ff20f2-652c-4bd6-a76c-4201e9eff214&usg=AOvVaw3f2Kw30eGL9qIfIza66Qpg>

Ponder, W. F., Carter, G. A., Flemons, P., & Chapman, R. R. (2001). Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, 15(3), 648-657.

QGIS Geographic Information System (2022). <http://www.qgis.org>

- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation, Vienna, Austria. <https://www.R-project.org>
- Rodda, G. H., Jarnevich, C. S., & Reed, R. N. (2011). Challenges in identifying sites climatically matched to the native ranges of animal invaders. *PloS one*, 6(2), e14670.
- Santana Jr, P. A., Kumar, L., Da Silva, R. S., Pereira, J. L., & Picanço, M. C. (2019). Assessing the impact of climate change on the worldwide distribution of *Dalbulus maidis* (DeLong) using MaxEnt. *Pest Management Science*, 75(10), 2706-2715.
- Sarkozi, A. (2021, June 2). *Climate change fans spread of pests and threatens plants and crops, new FAO Study*. FAO.
<https://www.fao.org/news/story/en/item/1402920/icode/>
- Schnase, J. L., Smith, J. A., Stohlgren, T. J., Graves, S., & Trees, C. (2002, June). Biological invasions: A challenge in ecological forecasting. In *IEEE International Geoscience and Remote Sensing Symposium* (Vol. 1, pp. 122-124). IEEE.
- Schöbel, C., & Carvalho, G. S. (2020). Niche modeling of economically important Mahanarva (Hemiptera, Cercopidae) species in South and Central America: Are Brazilian Spittlebug sugarcane pests potential invaders of South and Central America? *Journal of Economic Entomology*, 113(1), 115-125.
- Shabani, F., Kumar, L., & Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and Evolution*, 6(16), 5973-5986.

- Shabani, F., Kumar, L., & Solhjoui-fard, S. (2017). Variances in the projections, resulting from CLIMEX, boosted regression trees and random forests techniques. *Theoretical and Applied Climatology*, 129(3), 801-814.
- Shiferaw, B., Prasanna, B. M., Hellin, J., & Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*, 3(3), 307-327.
- Soberón, J., & Nakamura, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, 106(Supplement 2), 19644-19650.
- Socioeconomic Data and Applications Center (SEDAC).
<https://www.earthdata.nasa.gov/eosdis/daacs/sedac>
- Solano-Villarreal, E., Valdivia, W., Percy, M., Linard, C., Pasapera-Gonzales, J., Moreno-Gutierrez, D., ... & Rosas-Aguirre, A. (2019). Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon. *Scientific reports*, 9(1), 1-12.
- Spatial Data Science with R*. Spatial Data Science <https://rspatial.org/raster/index.html>
- Stibick, J. (2006). New pest response guidelines: False codling moth *Thaumatotibia leucotreta*. USDA–APHIS–PPQ–Emergency and Domestic Programs.
https://www.cdfa.ca.gov/plant/fcm/pdfs/publications/Stibick_2006-FCM_NPRG.pdf
- Sullivan, M. & Molet, T. (2007). CPHST pest datasheet for *Autographa gamma*. USDA–APHIS–PPQ–CPHST. Revised January 2014.

Sullivan, M. & Molet, T. (2007). CPHST pest datasheet for *Helicoverpa armigera*.

USDA-APHIS-PPQ-CPHST. Revised June 2018 by L. Morales and H. Moylett.

Sullivan, M. & Molet, T. (2007). CPHST pest datasheet for *Heteronychus arator*. USDA-

APHIS-PPQ-CPHST. Revised July 2014.

Sullivan, M. (2007). CPHST pest datasheet for *Spodoptera litura*. USDA-APHIS-

PPQCPHST. Revised April 2014.

Sullivan, M. (2007). CPHST pest datasheet for *Thaumatotibia leucotreta*. USDA-

APHISPPQ-CPHST. Revised January 2014.

Sutherst, R. W., Maywald, G. F., & Kriticos, D. J. (2007). CLIMEX version 3: User's guide.

Taylor, S., & Kumar, L. (2013). Potential distribution of an invasive species under climate change scenarios using CLIMEX and soil drainage: A case study of *Lantana camara* L. in Queensland, Australia. *Journal of Environmental Management*, 114, 414–422. <https://doi.org/10.1016/j.jenvman.2012.10.039>

Taylor, S., & Kumar, L. (2013). Potential distribution of an invasive species under climate change scenarios using CLIMEX and soil drainage: a case study of *Lantana camara* L. in Queensland, Australia. *Journal of environmental management*, 114, 414-422.

Thomson, N.J, (1965). The effect of nitrogen fertilizers on irrigated cotton at the Kimberley research station, north-western Australia. *Australian Journal of Experimental Agriculture and Animal Husbandry*. 218-226.

- Title, P. O., & Bemmels, J. B. (2018). ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, 41(2), 291-307.
- U.S. Department of Agriculture, Animal Plant Health Inspection Service.
- US Fish and Wildlife Service. (2013). National fish, wildlife, and plants climate adaptation strategy. <http://www.wildlifeadaptationstrategy.gov>
- USDA United States Department of Agriculture. (2022). *Agricultural Exchange Rate Data Set*. USDA, Economic Research Service. <https://www.ers.usda.gov/data-products/agricultural-exchange-rate-data-set>
- USDA United States Department of Agriculture (USDA). (2020). Cooperative Agricultural Pest Survey. <https://www.aphis.usda.gov/aphis/ourfocus/planthealth/caps>
- USDA United States Department of Agriculture (USDA). (2020). Economic Research Service. <https://www.ers.usda.gov>
- Venette, R. C., Davis, E. E., DaCosta, M., Heisler, H., & Larson, M. (2003). Mini risk assessment: False codling moth, *Thaumatotibia* (= *Cryptophlebia*) *leucotreta* (Meyrick)[Lepidoptera: Tortricidae]. *University of Minnesota, Department of Entomology, CAPS PRA*, 1-30.
- Venette, R. C., Davis, E. E., Heisler, H., & Larson, M. (2003). Mini risk assessment, silver Y moth, *Autographa gamma* (L.)[Lepidoptera: Noctuidae]. USDA APHIS PPQ.
- Webber, B.L., Yates, C.J., Le Maitre, D.C., Scott, J.K., Kriticos, D.J., Ota, N., McNeill, A., Le Roux, J.J. and Midgley, G.F. (2011). Modelling horses for novel climate

- courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models. *Diversity and Distributions*, 17(5), 978-1000.
- West, A. M., Kumar, S., Brown, C. S., Stohlgren, T. J., & Bromberg, J. (2016). Field validation of an invasive species Maxent model. *Ecological Informatics*, 36, 126-134.
- Wieczorek, J., Guo, Q., & Hijmans, R. (2004). The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International journal of geographical information science*, 18(8), 745-767.
- Wildlife Conservation Society (WCS) - WCS & Center for International Earth Science Information Network (CIESIN). (2005). Last of the wild project, V2, (LWP-2): Global Human Influence Index (HII) dataset (geographic). *NASA Socioeconomic Data and Applications Center (SEDAC)*.
<https://sedac.ciesin.columbia.edu/data/set/wildareas-v2-human-influence-index-geographic>
- Wood S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3-36.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2 ed.). Chapman and Hall/CRC.
- Wood, S.N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society* 65(1), 95-114.

- Wood, S.N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models.” *Journal of the American Statistical Association (B)*, 99(467), 673-686.
- Wood, S.N., Pya N. & Säfken B. (2016) Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111 (516), 1548-1563. <https://doi.org/10.1080/01621459.2016.1180986>
- WorldClim. (n.d.). <https://www.worldclim.org/>
- Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell, G., E. H., & Veran, S. (2013). Presence-only modelling using MAXENT: When can we trust the inferences? *Methods in Ecology and Evolution*, 4(3), 236-243.
- Yu, H., Cooper, A. R., & Infante, D. M. (2020). Improving species distribution model predictive accuracy using species abundance: Application with boosted regression trees. *Ecological Modelling*, 432, 109202.

VITA

Nicholas B. Reger

Communications

Sam Houston State University
Department Biological Sciences
2000 Avenue I Huntsville, Texas 77340

Major Research Interests

Plant evolution and systematics, parasitic plant evolution, and ecological niche modeling.

Education

B.A. (*Summa cum laude*) Biology, Sam Houston State College. 2020.

A.A. Lone Star Montgomery. 2017

Teaching Experience

Research Assistant. Sam Houston State University. 2018-2021.

Worked as a research assistant on a team lead by Dr. Randle to create MaxEnt and Boosted Regression Tree ecological niche models and maps for maize pathogens of greatest concern in the United States. These pathogens are *Autographa gamma*, *Chilo partellus*, *Diabrotica speciosa*, *Helicoverpa armigera*, *Heteronychus arator*, *Laodelphax striatellus*, *Spodoptera litura*, and *Thaumatotibia leucotreta*

Tutor. Sam Houston State University Academic Success Center. 2018-2020.

Tutored fellow students in calculus, physics, chemistry, and biology.

Honors and Awards

Fellows Scholar with a Capstone Independent Research Project in Italy. 2017.

Outstanding Delegation in the New York National Model United Nations. 2017.

Student Leadership Retreat. 2016.

Leadership

Leader in the Committee on Women's Rights for the Model United Nations 2017.

Organizer and Leadership behind the Capstone Tree Conservation Project. 2015.

Group Leader of Interact Woodlands Branch. 2015.

Presentations, Posters, and Abstracts

Randle, C. P., T. N. Bates, K. A. Bond, R. N. King, A. A. Martinez, N. B. Reger, S. Sadiq, A. P. Walker, D. I. Warren, and D. Haines. 2019 Is oak mistletoe host species distribution the result of the correlation between seasonal patterns of host tree growth and mistletoe phenology? Association of Southeastern Biologists conference, Memphis TN, April 3-6.

Martinez, A. A. N. B. Reger, D. C. Haines, and C. P. Randle. 2019. Variation in monoterpene emissions across potential hosts and over time as an explanation of local host preference in *Phorandendron leucarpum* (Raf.) Reveal & M.C. Johnst. (Viscaceae). Association of Southeastern Biologists conference, Memphis TN, April 3-6.

Randle, C.P., L. Bianchi*, J.Castillo*, N.B. Reger* and J.K.Williams. 2021. Predictive niche modeling for the identification of crop pests and pathogens of greatest concern in the US. Association of Southern Biologists Conference, Virtual, March 23-25.

Reger N. B., L. Bianchi, J. Castillo, J. K.Williams. and C.P. Randle. 2021. Predictive niche modeling for the identification of maize pathogens of greatest concern in the US. Association of Southern Biologists Conference, Virtual, March 23-25.

Reger, N. B., K. Hankins, and C. P. Randle. 2022. Predictive niche modeling for the identification of maize pathogens of greatest concern in the United States. Association of Southeastern Biologists conference, Little Rock AR, March 30- April 1.