Using Maxent to Predict Species Distribution and Compare Niche Overlap Between *Aptinothrips rufus* and *Aptinotrhips stylifer*

Mitchell McKesson

Department of Entomology - Purdue University - West Lafayette, IN 47907

Abstract

In 1968, while studying Aptiothrips, Dr. Lewis J. Stannard hypothesized that the two species found in Illinois, *A. rufus* and *A. stylifer,* are allopatric. During the time of his publication, he did not have the means to put this idea to the test. Thankfully, modern advancements in technology have made it much easier to predict species distributions from a sample population. Using the maximum entropy (maxent) method of niche modeling, and niche equivalency testing, we seek to put Stannard's hypothesis to the test and see if *A. rufus* and *A. stylifer* do indeed inhabit different niches. Our results show that the two species have a considerable amount of niche overlap, and thus cannot be considered allopatric. These or similar techniques may be used in the future to ask further questions about *A. rufus* and *A. stylifer*, or other thrips species.

Introduction

Thrips is the common name assigned to insects in the order of Thysanoptera. The word Thysanoptera can be roughly translated into English to mean "fringed wings." This is quite an accurate description, as most adult thrips can be easily recognized by the presence of two pairs of fringed, or feather-like, wings. Another key characteristic of thrips is their asymmetrical mouthparts. The right mandible of thrips is reduced, while the left is modified into a stylet for piercing-sucking feeding (Chisholm & Lewis, 1984). Aside from these two unique characteristics, thrips can also be identified by their small size, typically less than 1mm, and slender body shape.

Thrips are quite unique in terms of feeding behaviors. Many species are plant feeders, particularly on the leaves, flowers, and fruits of a wide variety of plants. There are, however, quite a few examples of predatory thrips that will feed on other thrips, mites, and smaller insects. The damage caused by many of the plant-feeding thrips has been studied and is considered to be economically significant (Sathe et al., 2015). Along with the physical damage caused directly by their feeding, many thrips act as key vectors to a number of dangerous plant diseases.

The objective of this study is not to study all thrips species, but rather two specific ones: *Aptinothrips rufus* and *Aptinothrips stylifer*. The genus *Aptinothrips* contains four distinct species of grass feeding thrips. All four of these species are believed to have originated from Europe, but our two of focus, *A. rufus* and *A. stylifer*, have since been introduced to the United States (Palmer,1974). The economic significance of these two thrips has not yet been studied in great detail, however, invasive species is a topic of great interest, and learning more about their ecology may also further our knowledge on their biology.

Stannard's (1968) "The Thrips, or Thysanoptera of Illinois," is a landmark in terms of thrips literature. Within this book, he describes the order Thysanoptera in great detail, but more importantly, he described every species found within the state of Illinois, and provides locality data as to where these species were found in the state. As thrips are not as studied as many other insects, this paper still remains an invaluable guide. Important to our study is a special note made while discussing *Aptionthrips*. Within his samples, he observed that *A. stylifer* was primarily found towards the northern part of the state, and that *A. rufus* had quite a few samples towards the southern part of the state. This led him to hypothesis that these two species are allopatric. That is to say, he expected that these species inhabit different niches, with *A. stylifer* favoring more northern climates, and *A. rufus* favoring more southern climates.

At the time of his publication, Stannard did not have the means to fully test this hypothesis. However, thanks to advancements in technology, we now have the means to put this idea of his to the test. There now exist many ways to model species distribution, but one that is very widely used, especially for insects, is the maximum entropy (maxent) niche model (Phillips et al., 2017). Using maxent, we can create distribution maps for our two species of *Aptinothrips*, and by analyzing the overlap of these models, we can compare and contrast the niches of our two species to see if Stannard was correct when he hypothesized that the two species are allopatric.

Materials and Methods

Maxent

Maxent is one of the many regularly used niche modeling programs. We chose this software primarily because it is free, easy to use, has been highly supported, and works well with smaller sample sizes. Combining species locality data and environmental data, maxent subsamples your data, using the majority of the points to train the models, and then a smaller proportion is used to test against these models. This is then replicated several times to build the best possible model. We mostly used the default settings of the program, but changed the settings to create response curves, show variable importance (Figure 9), and write background predictions. *A. rufus* was replicated 5 times through cross-validation, but because we had less samples for *A. stylifer*, it was replicated 10 times through bootstrapping.

Maxent gives a number of outputs with its models. The one we are most interested in while is the AUC. AUC stands for area under the curve, and in this case, the curve is the sensitivity vs specificity of our model. This acts a way to quantify how good your models are. In essence, AUC scores represent the probability of goodness for a model. An AUC of 0.5 means your model is essentially operating randomly, and is not good. Typically, an AUC of at least 0.7 is considered good, but the closer you get to 1, the better.

Species Presence Records

Records for both species were initially taken directly from Stannard's book. Here, he listed each of the counties of Illinois in which the species were found. 33 samples of *A. rufus* and 11 samples of *A. stylifer* were provided. This level of data can be used to create distribution

maps, however, we questioned if this broad sample area of an entire county and small number of samples would result in high levels of uncertainty and thus lessen the strength of our models. In an attempt to strengthen our models, we got in contact with the University of Illinois, and got access to the samples Stannard used when writing his book. These samples included much more precise locality data compared to the county levels used in his book. 31 samples of *A. rufus* and 15 samples of *A. stylifer* were provided. The number of samples from book to slide data did not change significantly, but we decided to run models for both of these data sets to see if the increase in precision would actually result in stronger models.

Given that all of our data points are from the state of Illinois, we chose to focus on the Northern Midwest United States as our study area. This includes the states of Illinois, Indiana, Michigan, Ohio, and Wisconsin.

Environmental Parameters

Environmental variables used in creating our models were downloaded from the Global Climate Data website (Flick & Hijmans, 2017). We used the 19 bioclimatic variables as well as elevation. Due to the different levels of precision in our species records, we had to use two different resolutions for our environmental variables. The models created using the less precise data from Stannard's book had to be made using the 2.5 minutes resolution, and the models using the more precise data collected from the slides used the 30 seconds resolution.

Recognizing that a high level of correlation likely existed among these environmental variables, we decided to test for this, and remove any highly correlated variables. To do this, we did a simple pairwise analysis in R (R Core Team, 2021), and by examining Pearson's correlation coefficient, had R remove any of the highly correlated layers (r >0.70). This left us with 8 variables to be used in creating our models (elevation, Bio2, Bio3, Bio5, Bio7, Bio8, Bio9, and Bio18).

Figure 1: 19 bioclimatic variables and elevation taken from worldclim.org. Variables in bold and marked with asterix were chosen as representatives of highly correlated variables.

Abbreviation	Environmental Variable
bio1	Annual Mean Temperature
*bio2	**Mean Diurnal Range (Mean of monthly (max temp - min temp))**
*bio3	**Isothermality (BIO2/BIO7) (×100)**
bio4	Temperature Seasonality (standard deviation ×100)
*bio5	**Max Temperature of Warmest Month**
bio6	Min Temperature of Coldest Month
*bio7	**Temperature Annual Range (BIO5-BIO6)**
*bio8	**Mean Temperature of Wettest Quarter**
*bio9	**Mean Temperature of Driest Quarter**
bio10	Mean Temperature of Warmest Quarter
bio11	Mean Temperature of Coldest Quarter
bio12	Annual Precipitation
bio13	Precipitation of Wettest Month
bio14	Precipitation of Driest Month
bio15	Precipitation Seasonality (Coefficient of Variation)
bio16	Precipitation of Wettest Quarter
bio17	Precipitation of Driest Quarter
*bio18	**Precipitation of Warmest Quarter**
bio19	Precipitation of Coldest Quarter
*elev	**Elevation**

Bias File

When dealing with species occurrence data, especially for specimens as small as thrips, a level of sampling bias is impossible to avoid. Our assumption is that areas with more samples are not inherently more environmentally significant, but rather where there was more sampling effort. Maxent is able to account for this sampling bias with the aid of a sample bias file (FIGURE). To create this bias file, we coded a two-dimensional kernel density estimation in R, using the kernsmooth (Wand, 2020) package to optimize the amount of density. In this case, the density being measure is for our species locality data. The bias file assigns different weights to areas depending on the density of samples in that area. Areas with less samples are assigned a higher weight than areas with more in an attempt to balance out our data and give each location equal environmental importance. We assume that both of our species were gathered with equal sampling effort, so we can use one bias file for both species. However, we did need to make two bias files for the different environmental layer resolutions.
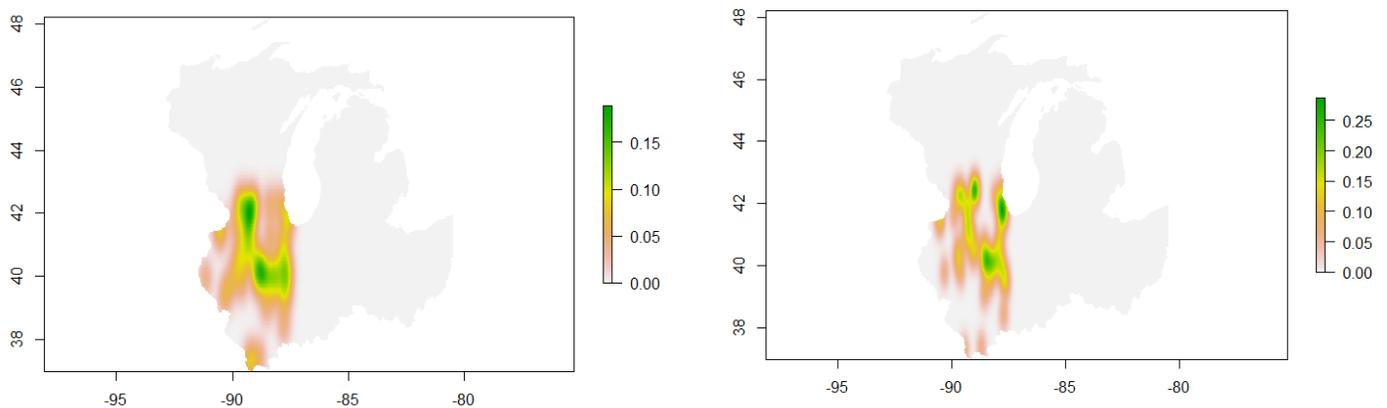

Figure 2: Bias files used to correct for sampling bias in our models. File on the left used for data from the book at 2.5m resolution. File on the right used for data from the slides at 30s resolution.

Niche Overlap

A simple way to examine the differences between two species' distributions is doing an overlap test. For this, we took the distribution maps of our species from our maxent models, overlayed them in R, and subtracted to highlight the differences (Figure 7). As this is a little difficult to interpret, we then transformed this into a simpler binary map to highlight specifically the areas of similarities and differences. Both of these maps provide a way to visualize our data, but alone can not quantify the overlap, which is needed in order to answer our question.

Niche Equivalency

To actually quantify the amount of overlap between these two species' distribution, we used a niche equivalency test, as proposed by Warren et al. (2008) and Suárez-Mota et al. (2015). To accomplish this test, we first did an overlap test between the two models we created using the species slide data. The outputs of this test include Schoeners D (Schoener, 1968) and Hellinger

Distances [I] (van der Vaart, 1998). Both of these variables measure overlap using different calculations. It is standard to calculate and report both, so we will do just that. Next, we created 100 replicates for each species by randomly scrambling the names assigned to the data points, and ran this randomized data through maxent under the same procedure as our non-random models. By testing our non-random models against all of these random replicates, we created histograms (Figure 5) to show the frequency of random overlap values. These histograms then allowed us to test our observed D and I values to the average D and I values from our random replicates, and finally calculate the probability that our two species inhabit different niches.
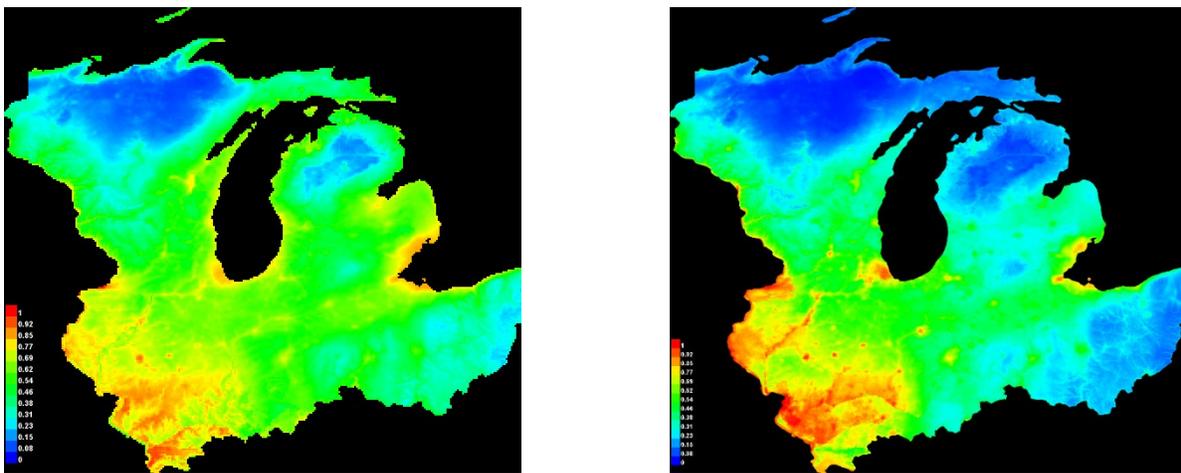
Results

Distribution Models

Figure 3.1: Maxent distribution models for *Aptinothrips rufus*. Model on right created with data from book and 2.5m resolution environmental layers. Model on left created with data from slides and 30s resolution environmental layers.

Figure 3.2: Maxent distribution models for *Aptinothrips stylifer*. Model on right created with data from and 2.5m resolution environmental layers. Model on left created with data from slides and 30s resolution environmental layers.
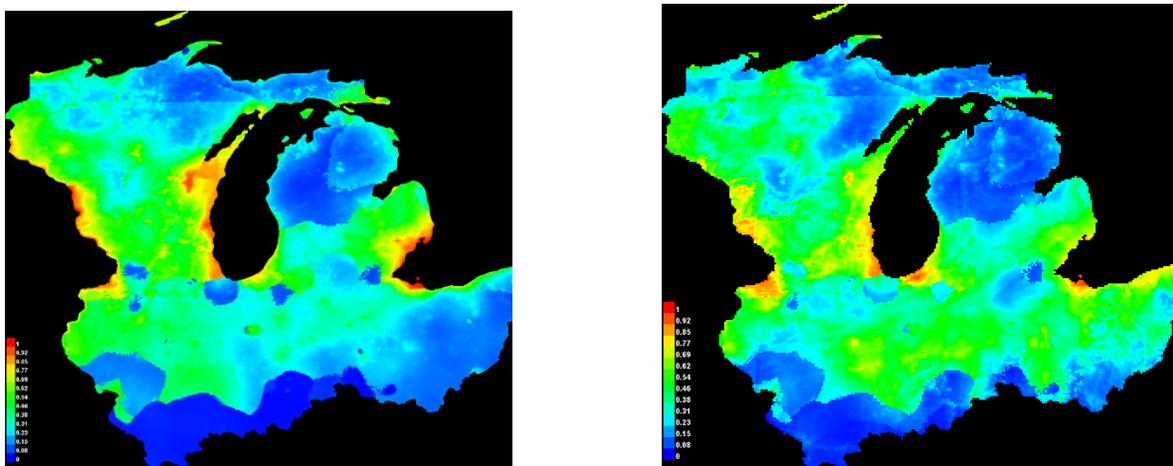
Our models were evaluated using the area under the curve (AUC), as explained by Phillips et al. (2006). The training and testing AUC scores for all of our models are reported in figure 4. For all four of our models, maxent reported relatively high AUC scores, suggesting that all of these models are significant statistically.

Figure 4: Area under the curve (AUC) scores for both *Aptinothrips rufus* and *Aptinothrips stylifer* using both the data from book and data from the slides. A high AUC score (AUC>70) is generally considered good and statistically significant.

	Training AUC	Testing AUC
Book Data		
Aptionthrips rufus	0.85	0.80
Aptinothrips stylifer	0.82	0.82
Slide Data		
Aptionthrips rufus	0.79	0.87
Aptinothrips stylifer	0.85	0.78

Niche Equivelancy

The results of our niche equivalency test are represented in figure 5. Both graphs show the frequency of overlap values of our 200 random replications. On both graphs, the solid line represents the observed D and I value of our non-random overlap test (D=0.31, I=0.39). The dashed line is then the mean of our random tests. By comparing our observed values to these mean values, we get p values of 0.91 for D and 0.98 for I. In both cases, a large p value signifies a low probability that these two species inhabit different niches.
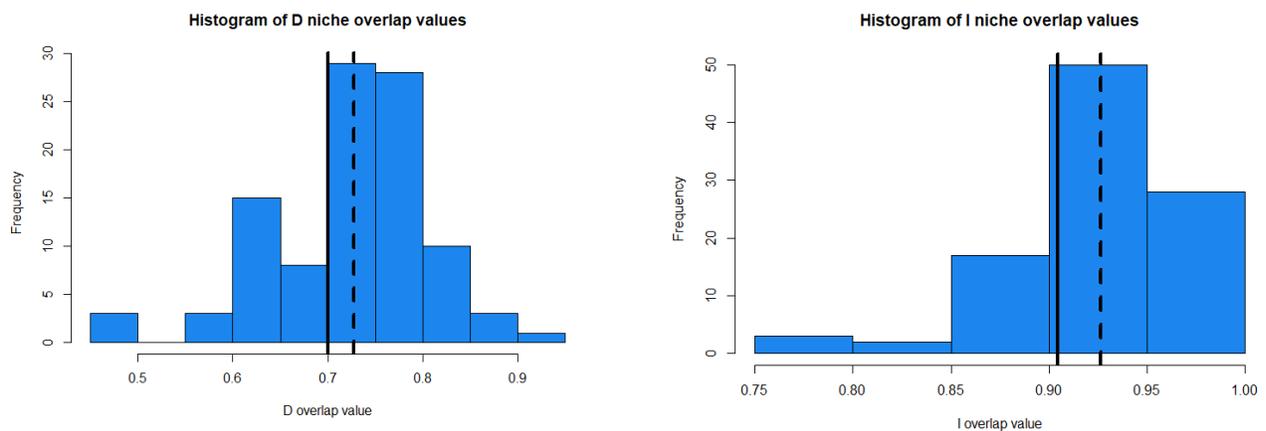

Figure 5: Histograms for the result of our niche equivalency test. Both Schoeners D and Hellinger Distances calculated. For both graphs, the bars represent frequency of overlap for random replicate overlap tests, solid line is our observed D and I values for non-random overlap test, and dashed line is mean value of random replicates.

Discussion

From our results we can conclude that Stannard's initial prediction was incorrect, and *Aptinothrips rufus* and *Aptinothrips stylifer* likely do not inhabit different niches. Our models, however, do show that there are indeed differences among the ranges of these two species. Despite the high level of overlap, *A. rufus* does seem to be better adapted for more southern regions, and *A. stylifer* better adapted for the north. We can begin to question why this is, or what factors may be driving it. Looking at the percent contribution of each environmental variable as provided by maxent (figure 9), we can already begin to start making predictions to this question. We can see that isothermality contributed much more to our *A. stylifer* model than it did for *A. rufus*. It is worth reminding that isothermaility was chosen as a representative of the correlated layers. If we look at the results of our pairwise analysis, we see that specifically it is highly correlated to Bio4 (temperature seasonality) and Bio15 (Precipitation Seasonality). Examining all three of these factors in greater detail may help us learn more about what is driving the differences between these two species.

Another thing that may help us learn more about these two species of *Aptinothrips* is compare what we are seeing here in their introduced range to what is occurring in their native range. We have an idea as to what environmental factors are significant to the species here in the United States, so it may be worth examining how these may compare and contrast to their native environments.

One last thing that may be considered in the future is future predictions. Using expected future environmental conditions such as increase in CO_2 emissions and temperature, maxent is able to then predict changes to a species distribution. By implementing this sort of technique to our *Aptinothrips species*, we could see if their niche overlap will increase, decrease, or stay the same as the world begins to change. Stannard may have been incorrect in predicting that the two species are allopatric, but there is still an undeniably unique relationship between them that should absolutely be further examined.

Acknowledgements

I would like to thank Dr. Stephen Cameron, Makani Fisher, and David Stanford-Beale for giving me the opportunity to join them on this study. You have been incredibly kind, supportive, and patient, and despite the hiccups we ran into, I truly did enjoy my time working on this project. Plus, you helped me find a passion for thrips I never expected to have.

I would also like to thank the entire Purdue Entomology department as a whole, including staff, faculty, and my classmates. As someone who transferred into the department at the start of my junior year, you have all been welcoming and supportive from the very beginning, and made my last two in a half years here at Purdue something special. Thank you.

Additional Models

Figure 6: Exact species locality data used in our models. Data from book shown on the left, data from the slides shown on the right.
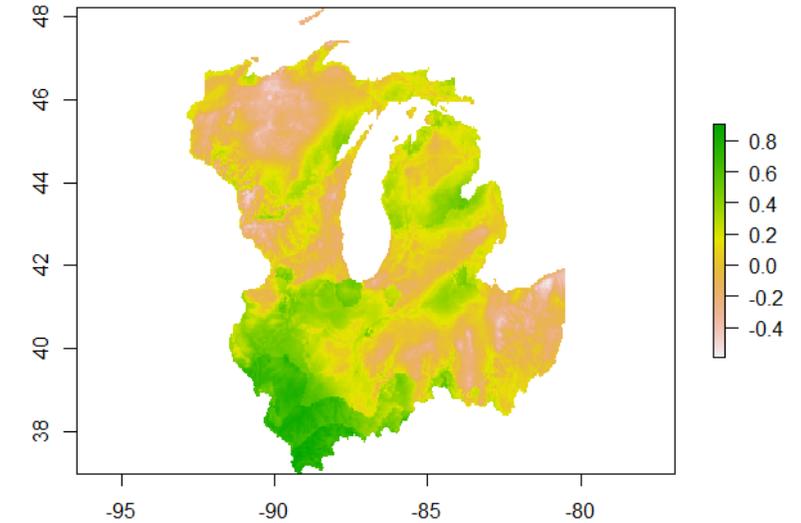

rufus	90.57	39.61011		rufus		
rufus	-90.5731	41.48917		rufus	-89.3004	41.74619
rufus	-87.6936	39.39278		rufus	-88.9616	39.86
rufus	-87.8997	42.03389		rufus	-89.9168	40.23965
rufus	-87.9344	40.93528		rufus	-88.8473	40.49089
rufus	-89.655	39.79944		rufus	-90.2015	39.71557
rufus	-87.6961	39.61111		rufus	-89.3207	42.04265

Figure 7: Map showing the overlap between *Aptinothrips rufus* and *stylifer* niches. Areas in green are no overlap areas, and *Aptinothrips* areas in orange are high overlap areas.

				78808
				01037
				46733
				75817
				35175
				50752
				47123
				18342
				44607
				75627
				33626
				78823
				35175
				33626
				32439
				.3333
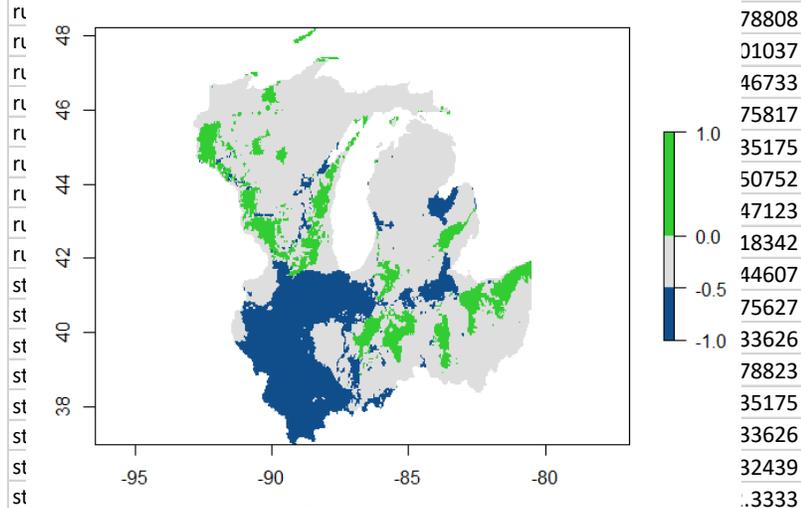
Figure 8: Binary overlap map for easier visualization of niche differences. Areas in green are where only *Aptinothrips stylifer* is expected to be found.

stylifer	-87.7533	41.715		stylifer	-89.3207	42.04265
stylifer	-90.5731	41.48917		stylifer	-87.6462	41.89342
stylifer	-88.1428	42.26694		stylifer	-90.5674	41.46733
stylifer	-88.5075	40.21278		stylifer	-89.2858	41.20447
stylifer	-89.0386	42.48417		stylifer	-87.7328	40.18342
stylifer	-89.0092	42.41333		stylifer	-88.1992	40.14008
stylifer	-87.8229	42.48239		stylifer	-88.5911	40.01037
				stylifer	-88.9041	40.17463

Areas in blue are where only *Aptinothrips rufus* are expected   to be found. Areas in grey are where the two species'
niches are expected to overlap.

Figure 9: Percent contribution tables for the environmental variables used in creating our models. Left is
the results from our *Aptinothrips rufus* model using slide data. Right is the results from our *Aptinothrips
stylifer* model using slide data. Refer to figure 1 for explanation of environmental variables.

Variable	Percent contribution	Permutation importance
thrips_Bio5	49.9	64
thrips_Bio8	40	7.7
thrips_Bio_elev	3.3	14
thrips_Bio9	3.1	0.3
thrips_Bio3	2.3	12.3
thrips_Bio2	1	1.1
thrips_Bio18	0.2	0.3
thrips_Bio7	0.2	0.4

Variable	Percent contribution	Permutation importance
thrips_Bio3	49.1	43.3
thrips_Bio8	41.5	49
thrips_Bio18	3.3	0
thrips_Bio2	2.1	2.1
thrips_Bio5	1.7	2.6
thrips_Bio_elev	1.2	3
thrips_Bio9	1	0
thrips_Bio7	0	0

References

Chisholm, I., & Lewis, T. (1984). A new look at thrips (Thysanoptera) mouthparts, their action and effects of feeding on plant tissue. Bulletin of Entomological Research, 74(4), 663-675. doi:10.1017/S0007485300014048

Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37 (12): 4302-4315.

Matt Wand (2020). KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995). R package version 2.23-18.
  https://CRAN.R-project.org/package=KernSmooth

Palmer, J. M. (1974). The grass-living genus Aptinothrips Haliday (Thysanoptera: Thripidae). Journal of Entomology Series B, Taxonomy, 44(2), 175–188. https://doi.org/10.1111/j.1365-3113.1975.tb00012.x

Phillips S.J., Anderson R.P., Shapire R.E. (2006) Maximum entropy modeling of species distributions. Ecological Modelling 190(3–4): 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sathe, Tukaram & Pranoti, Mithari & Patil, Sampatrao & Desai, Abhijit. (2015). Diversity, Economic Importance And Control Strategies of Thrips (Thysanoptera) on Crop Plants. Indian Journal of Applied Research. 5. 761-763.

Schoener T.W. (1968) The Anolis lizards of Bimini: resource partitioning in a complex fauna. Ecology 49(4): 704–726. https://doi.org/10.2307/1935534

Stannard, L. J. (1968). The thrips, or Thysanoptera, of Illinois. Illinois Natural History Survey Bulletin, 29(1-4), 215–552. https://doi.org/10.21900/j.inhs.v29.166

Steven J. Phillips, Miroslav Dudík, Robert E. Schapire. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/

Suárez-Mota M.E., Villaseñor R.J.L., López-Mata L. (2015) Ecological niche similarity between congeneric Mexican plant species. Plant Ecology and Evolution 148(3): 318–328. https://doi. org/10.5091/plecevo.2015.1147

VanDerWal J., Shoo L.P., Johnson C.N., Williams S.E. (2009) Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. The American Naturalist 174(2): 282–291. https://doi.org/10.1086/600087

Warren D.L., Glor R.E., Turelli M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62(11): 2868–2883. https://doi. org/10.1111/j.1558-5646.2008.00482.x