Species Distribution Modelling of Odonata to Investigate Climatic Habitat

ANUJ SHRESTHA & ANNE DENTON, North Dakota State University

ND-EPSCOR

ABSTRACT

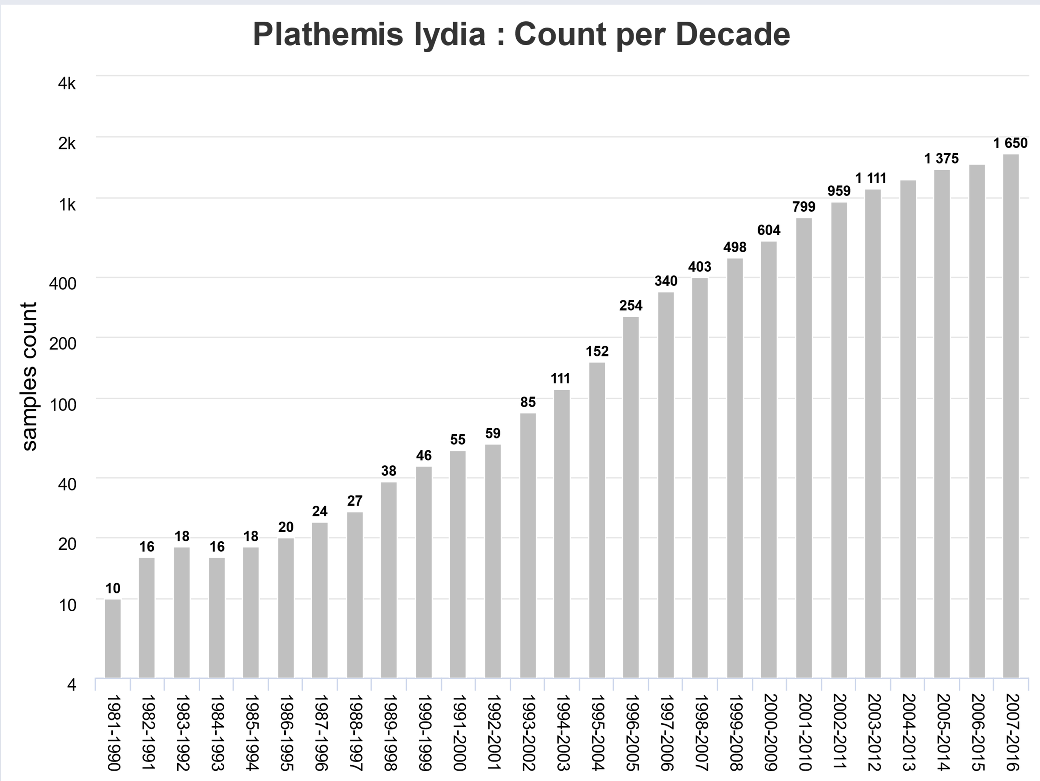
Local climate could affect the abundance of species. In this study, we create SDM models for known occurrence of interested Odonata species (source: Odonata Central), using popular machine learning and regression methods by utilizing climate data as predictors (source: PRISM), to investigate past and present distributions. We perform this analysis on a “moving-decade” framework. The results allowed us to interpret the climatic variables of interest and understand the gradual change in their distributions over time.

INTRODUCTION

In this study, we are considering samples of 4 odonata species (for this presentation, we’re considering Plathemis Lydia only). These samples were gathered from the Odonata Central database. The idea was to find the bio-climatic conditions at the habitat of these species, and to model the likelihood of the presence of these species in a given area (mainland USA). We used climate data from PRISM Climate Group. The main problem we faced was to make sure that the model's prediction was consistent. To check this we employed "moving-decade" framework to utilize a sample in multiple instances and to increase the sample size as well. We then utilized popular methods, that are usually employed for SDM, and checked for consistency among them.

DATASET

- Odonata Central : 1981 - 2016



BIO-CLIMATIC VARIABLES

- PRISM : 1981-2016 (Monthly dataset as raw data)
- Biologically meaningful variables
- Derived from monthly temperature and rainfall values
- Represents: Annual-trends / Seasonality / Extreme factor

BIO1 = Annual Mean Temp.
BIO2 = Mean Diurnal Range
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temp. Seasonality
BIO5 = Max Temp. of Warmest Month
BIO6 = Min Temp. of Coldest Month
BIO7 = Temp. Annual Range (BIO5-BIO6)
BIO8 = Mean Temp. of Wettest Quarter
BIO9 = Mean Temp. of Driest Quarter
BIO10 = Mean Temp. of Warmest Quarter
BIO11 = Mean Temp. of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

MOVING DECADE

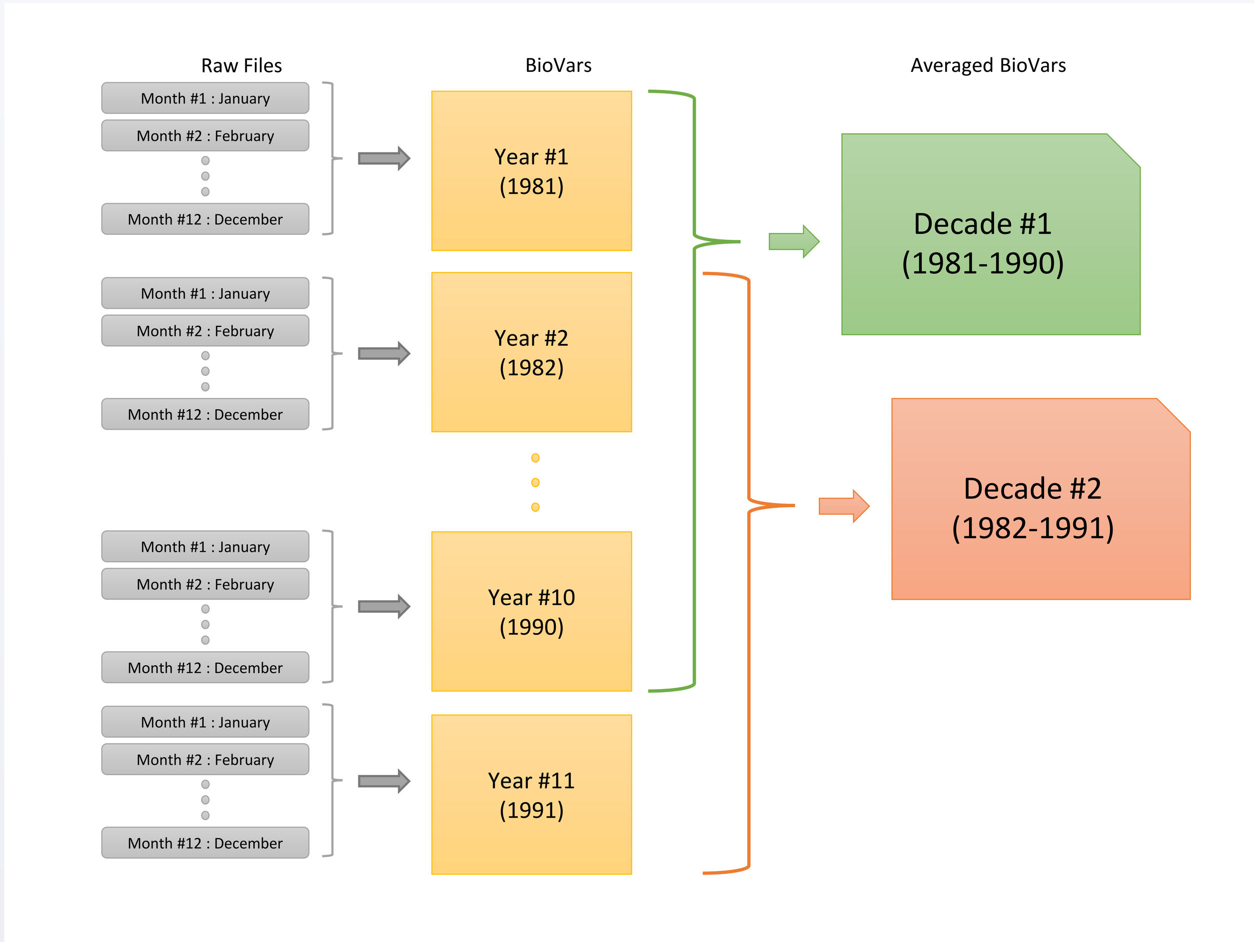


Figure 1 : Decade Climatic Layers Formulation

Monthly data from PRISM were downloaded from 1981-2016. These monthly data were then stacked and converted to bio-climatic variables for each year. Then, for each 10 consecutive years, the bio-climatic variables were averaged to form the variables for the decade. There were 27 such decades, starting with 1981-1990 and ending with 2007-2016.

EVALUATION

The popular methods for SDM were employed to model the habitat of the species. These methods were checked for consistency, based on the importance given to each bio-climatic variables. The methods are:

Methods that appear consistent
Maximum Entropy - MAXENT
Support Vector Machine - SVM
Random Forest - RF
Boosted regression trees - BRT

Methods that do not appear consistent
Classification and Regression Tree - CART
General Additive Models - GAM
Multivariate Adaptive Regression Splines - MARS
Generalized linear models - GLM
Mixture discriminant analysis - MDA
Flexible discriminant analysis - FDA
Maximum Likelihood - MAXLIKE
Regression Partition and Regression tree - RPART

The “consistent” methods showed least anomaly in its report of variable importance, & is shown in figure 2 and 3.

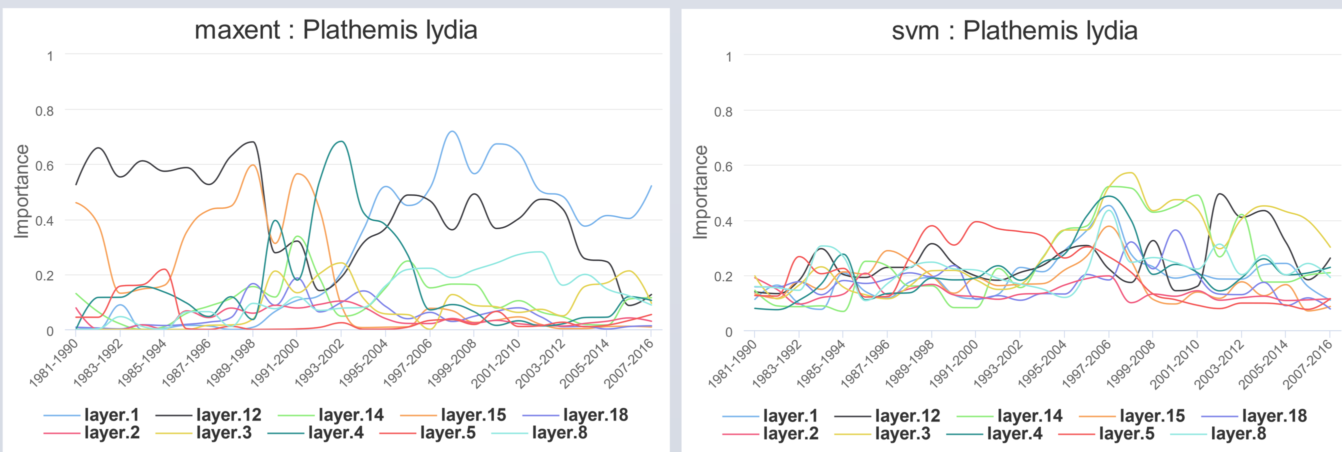


Figure 2 : (left) MAXENT (right) SVM

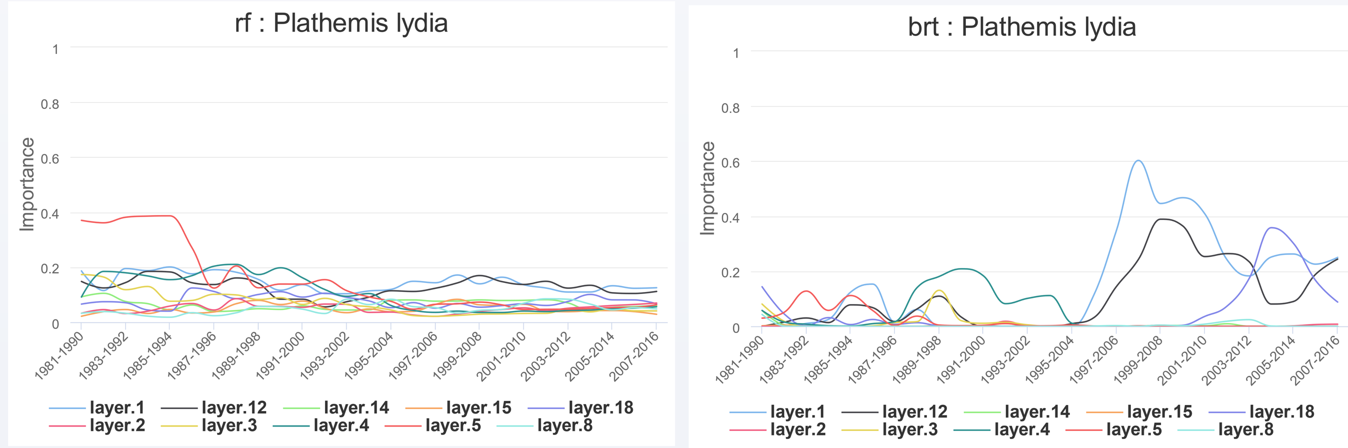


Figure 3 : (left) RF (right) BRT

8 methods were founds to have inconsistent results. Such inconsistency can be seen as shown in figure 4 for methods CART and MAXLIKE.

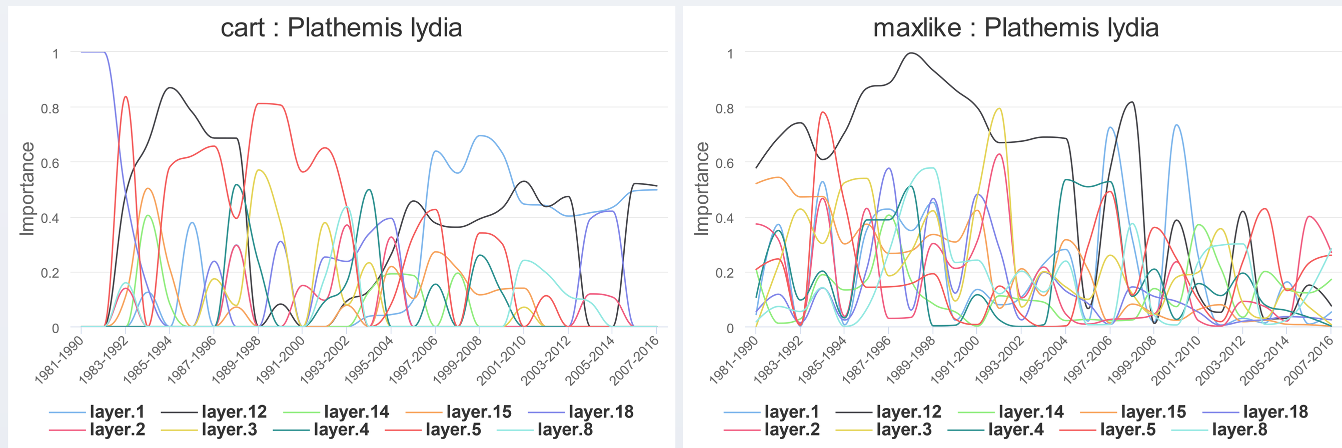


Figure 4 : (left) CART (right) MAXLIKE

Layer.1 and layer.12 was found to be important in all of the “consistent” methods. The importance values of the layers are consistent in all of the layers as shown in figure 5.

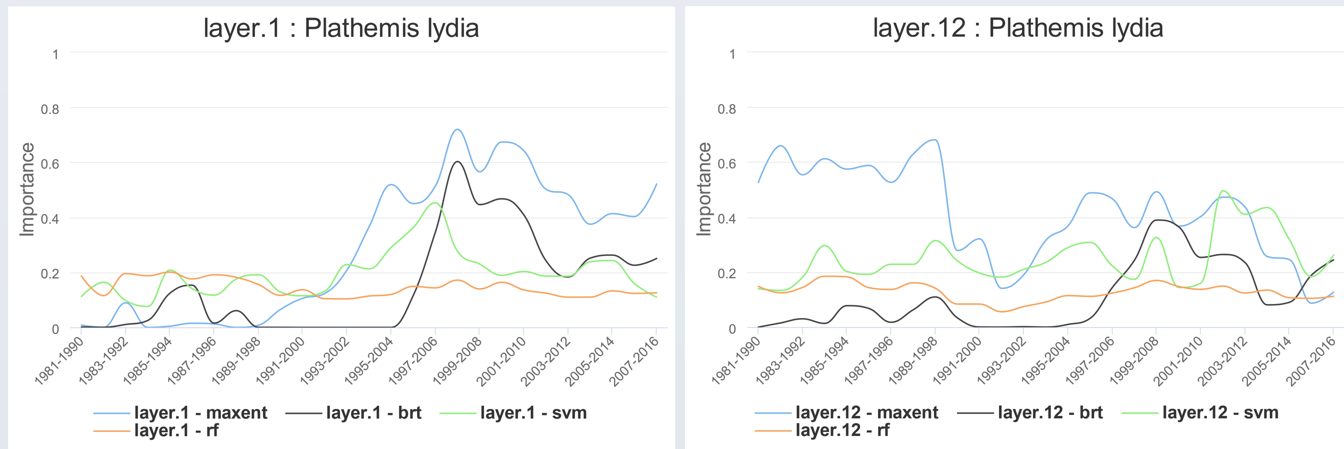


Figure 5 : (left) Layer.1 (right) Layer.12

CONCLUSION

The use of “moving decade” helped to find out the methods that produced consistent results and filtering out ones that produced highly irregular results. These “consistent” methods seems to agree with one another to a considerable degree, and thus, these methods could be further ensembled into a single model with weighted importance given to each model based on quality measures like AUC.

REFERENCES

[1] G. C. Vega, L. R. Pertierra and M. Á. Olalla-Tárraga, "MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling," Scientific Data 4:170078, 2017.

[2] B. Naimi and M. B. Araujo, "sdm: a reproducible and extensible R platform for species," Ecography, vol. 39, no. 10.1111/ecog.01881, pp. 368-375, 2016.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation through grant IIA-1355466.