Measuring Impact of Climate Change on Tree Species: analysis of JSDM on FIA data

Hyun Choi, Sergio Marconi, Ali Sadeghian, Ethan White, Daisy Zhe Wang
University of Florida
Gainesville, FL
{hchyun, asadeghian, daisyw}@ufl.edu, {ethan, sergio.marconi}@weecology.org

Abstract

One of the first beings affected by changes in the climate are trees, one of our most vital resources. In this study tree species interaction and the response to climate in different ecological environments is observed by applying a joint species distribution model to different ecological domains in the United States. Joint species distribution models are useful to learn inter-species relationships and species response to the environment. The climates' impact on the tree species is measured through species abundance in an area. We compare the model's performance across all ecological domains and study the sensitivity of the climate variables. With the prediction of abundances, tree species populations can be predicted in the future and measure the impact of climate change on tree populations.

1 Introduction

Ecologist expect the rapid change in climate to have unprecedented impacts on species geographical redistribution. Shifts in economically important species will likely have significant impacts on human well-being, forest management strategies and conservation planning [1, 9, 13]. Reliably predicting future patterns in species distributions has become a major goal in ecology, and hence numerous species distribution models have been developed [1, 10, 3]. Modern implementations use historical climate data and a few topographic variables to predict individual species abundance, i.e. number of individuals per targeted area [5].

Most of these algorithms are univariate approaches and focus on individual key species separately [10]. However, most species interact with others and their correlation can hold crucial information. This is particularly true for rare species, whose records may be limited in number and borrowing strength from co-occurring species may be fundamental to improve predictions and reduce their uncertainty [5]. This has motivated the use of, Joint Species Distribution models (JSDM), a group of multivariate approaches that generally use the partial correlation matrix among responses to estimate their interactions [5].

Another key challenge in species distribution modeling is how to address the role of scale on the ecological drivers controlling species distribution [6]. For example, both SDM and JSDM are generally trained on continental datasets. The resulting models may fail when used to make predictions for local areas, because the same environmental features may have different effects on species abundance locally [8]. Understanding and integrating the scaling rules behind changes in features effects will potentially unlock the opportunity to explicitly implement SDMs universally applicable. Yet, to date we don't even have a complete analysis of the role of scale on joint species distribution modeling.

To address this limit, we studied the relationship of species distributions and species interactions across multiple ecological domains and scales utilizing the generalized joint attribute model (GJAM)

[6]. GJAM is a cutting-edge and open source JSDM. Using GJAM for the purpose of this multi-scale analysis has several advantages: (1) it is one of the few joint multilevel model of species abundance available for ecologists, (2) it integrates partial correlations among species and estimates structure among their interactions [14]; (3) changes in models coefficients can easily be interpreted; (4) it has been used extensively in the last few years, including for predicting tree species abundance at continental scale.

By applying GJAM across different scales we find that species have different interaction patterns and sensitivity to environmental features and we find similar results for different US ecological domains.

2 Methodology

In this section we will demonstrate a method of modeling tree species behaviours in different environments. We use GJAM, a JSDM that is well suited for handling multifarious data [6].

Let $Y_i \in \mathbb{R}^S$ represent the abundances of S unique species in the ith plots, and $X_i \in \mathbb{R}^Q$ be the feature vectors consisting of the Q climate variables corresponding to each plot, for $1 \leq i \leq n$. We model (for the sake of space and simplicity assume all features are continuous):

$$Y_i \sim \mathcal{N}(B^\top X_i, \Sigma) \tag{1}$$

Where $\mathcal{N}(\cdot, \cdot)$ is the multivariate normal distribution, $B \in \mathbb{R}^{Q \times S}$ symbolizes the species response to the climate variables, and $\Sigma \in \mathbb{R}^{S \times S}$ is a covariance matrix that resembles the species interaction between each other. Given observations for X and Y, GJAM estimates B and Σ using equation (1), we refer the reader to [6] for further details. With the distribution of species and climate variables the model can predict the species response in a plot.

2.1 Modeling across Different Ecological Domains

GJAM has been previously used to learn continental joint species distribution of trees [6] using FIA. In this study we further investigate how climate interacts with scale at different geographical regions by applying GJAM to 20 independent ecological domains defined by Neon [12]. We use B to study the climate variables' impact on the species across different ecological domains by comparing it from one domain to another and across scales. Similarly, Σ can be used to study species interaction across all domains. Also, having the climate variables the model can predict the species response for plots with no samples or at another point in time.

The environmental predictors we used are slope, aspect (direction a tree faces due to the slope of the land), elevation, average day length (March, April, October), precipitation (Summer months), radiation (August), maximum temperature (August), and minimum temperature (January).

3 Preliminary Evaluation and Results

Data. For this study we used the Forest Inventory and Analysis (FIA)[1], an openly available inventory of forests collected at a fine grid nationwide [2]. Plots are distant from each other (~5km), and cover an area of 673 m^2. Plots are sampled every 5 to 10 years depending on the State. We used tree species identities from the latest census to have an estimate of the abundances for any target species, at each plot, for current climate.

The climate data source, Daymet [15], contains daily information at a 1×1 (km) resolution for the climate such as max and min temperature, radiation, and precipitation.

Ecologists divided the United States into different ecological regions for NEON. The NEON data [12] has airborne remotely sense images and ground observations of NEON plots.

3.1 Experimental set up

To investigate JDSM across scales, we trained an independent model for each individual ecological region and the entire United States. Finally, we compared species-to-species and species-to-environment correlations, for all those pairs of species co-occurring in different regions (n=262).

---

[1]FIA data available at `https://apps.fs.usda.gov/fia/datamart/datamart.html`.

We created 20 separate response and predictor matrices for each of the domains by assigning each FIA plot to its respective domain. To address sampling errors driven by small plot size, we group neighbouring plots in clusters of 16 and treat them as contiguous ~1ha plots using a same size K-means clustering method based on the coordinates. With the clustered plots, we also aggregated the response matrix to align each observation with the climate observations.

We randomly split 80/20 for train/test sets each of the domains. The evaluation metric for the model's prediction is the coefficient of determination.

3.2 Assumptions

For this experiment, species that do not occur or sparsely occur are not predicted, which may inflate the accuracy of the model. The reason for the exclusion of these species is due to the smaller impact that these species may have on the overall region.

3.3 Preliminary results

We have obtain preliminary results for all domains, except Hawaii, Puerto Rico, and Tundra and Taiga in Alaska domains due to the lack of FIA data in those regions.
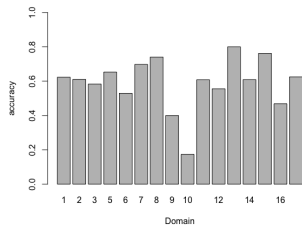

Figure 1: Accuracy per domain, where accuracy is the percent of species where GJAM performed better than the sample mean.
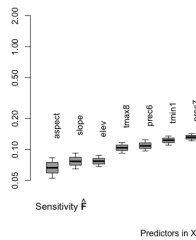

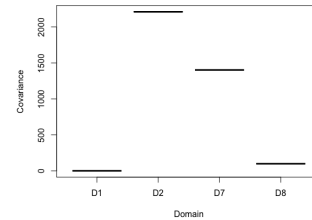Figure 2: Sensitivity of predictors for Domain 3.

Figure 3: Covariance of Liriodendron tulipifera and Acer Rubrum across Northeast (D1), Mid Atlantic (D2), Appalachians (D7), Ozarks Complex (D8).

Figure 2 displays that the day lengths are the most sensitive predictors for domain 3, this is true across all domains. The high importance of `day length` can be explained through tree phenology, as trees adapt their growing and reproduction cycles over time [4]. The variation of the impact with other predictors in different domains can be explained by the adaptation to the unique climates, for example elevation has higher impact in mountainous regions such as the Northern Rockies domain compared to the flatter Southeast domain. Species performance varies from domain to domain. The reason for this is because a species may occur centrally in one domain while it occurs marginally in another, the model generally performs better when a species occurs centrally.

We analyze the covariance of Liriodendron tulipifera and Acer Rubrum which overlap in 4 domains, shown in Figure 3. The covariances in domains 2 and 7 are much higher than the other domains. The change in covariances across domains can be seen in other species as well and this variance will be tried to explained through ecological knowledge.

4 Conclusion and future work

We outline three publicly available data sources related to climate and tree species. We also apply a straightforward model to predict the abundance of species that takes into account both climate and inter species interactions. This prediction becomes more important with the drastic changes that climate change could have on global ecosystems. Further spatio-temporal analysis of the join species distribution model [7, 11], to study species distribution shift under climate change over time becomes more important with the drastic changes that climate change could have on global ecosystems. The extension of this work in progress will be enhancing species classification on NEON airborne remote sensing images with information learned from JSDM models. Furthermore we work closely with ecologist to both incorporate rules from domain knowledge in future models and to provide data driven insight back.

Acknowledgments

References

[1] M. B. Araújo and M. Luoto. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16(6):743–753, 2007.

[2] W. A. Bechtold, P. L. Patterson, et al. The enhanced forest inventory and analysis program-national sampling design and estimation procedures. *Gen. Tech. Rep. SRS-80. Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station. 85 p.*, 80, 2005.

[3] L. Boisvert-Marsh, C. Périé, and S. de Blois. Divergent responses to climate change and disturbance drive recruitment patterns underlying latitudinal shifts of tree species. *Journal of Ecology*, 2019.

[4] I. Chuine, J. Belmonte, and A. Mignot. A modelling analysis of the genetic variation of phenology between tree populations. *Journal of Ecology*, 88(4):561–570, 2000.

[5] J. S. Clark, A. E. Gelfand, C. W. Woodall, and K. Zhu. More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications*, 24(5): 990–999, 2014.

[6] J. S. Clark, D. Nemergut, B. Seyednasrollah, P. J. Turner, and S. Zhang. Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, 87(1):34–56, 2017.

[7] R. Fathony, A. Rezaei, M. A. Bashiri, X. Zhang, and B. Ziebart. Distributionally robust graphical models. In *Advances in Neural Information Processing Systems*, pages 8344–8355, 2018.

[8] A. E. Gelfand, S. Ghosh, J. S. Clark, et al. Scaling integral projection models for analyzing size demography. *Statistical Science*, 28(4):641–658, 2013.

[9] S. Hsiang, R. Kopp, A. Jina, J. Rising, M. Delgado, S. Mohan, D. Rasmussen, R. Muir-Wood, P. Wilson, M. Oppenheimer, et al. Estimating economic damage from climate change in the united states. *Science*, 356(6345):1362–1369, 2017.

[10] L. R. Iverson, M. P. Peters, A. M. Prasad, and S. N. Matthews. Analysis of climate change impacts on tree species of the eastern us: Results of distrib-ii modeling. *Forests*, 10(4):302, 2019.

[11] A. Mollalo, A. Sadeghian, G. D. Israel, P. Rashidi, A. Sofizadeh, and G. E. Glass. Machine learning approaches in gis-based ecological modeling of the sand fly phlebotomus papatasi, a vector of zoonotic cutaneous leishmaniasis in golestan province, iran. *Acta tropica*, 188: 187–194, 2018.

[12] National Ecological Observatory Network. Data products: Neon domains shapefile. `http://data.neonscience.org`, 2017. Provisional data downloaded on 19 Feb 2017. Battelle, Boulder, CO, USA.

[13] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019.

[14] D. Taylor-Rodriguez, K. Kaufeld, E. M. Schliep, J. S. Clark, A. E. Gelfand, et al. Joint species distribution modeling: dimension reduction using dirichlet processes. *Bayesian Analysis*, 12(4): 939–967, 2017.

[15] M. M. Thornton, P. E. Thornton, Y. Wei, B. W. Mayer, R. B. Cook, and R. S. Vose. Daymet: Monthly climate summaries on a 1-km grid for north america, version 3. *ORNL DAAC, Oak Ridge, Tennessee, USA.*, 2017. URL `https://doi.org/10.3334/ORNLDAAC/1345`.