The Past and Present Distribution of *Liatris scariosa* var. *novae-angliae*

Lauren Ash

May 5, 2016

Abstract

Few studies are able to use historic data of sites with known extinctions for comparison to current populations, due to the rarity of such information. This study uses such data to compare the past and present distribution of the rare New England flower, *Liatris novae-angliae*, the northern blazing star, which has disappeared from 69% of its former range. I used 19 bioclimatic ('BioClim') variables and ANOVA, discriminant, logistic regression, and regression tree analyses to determine if there were significant differences between sites with current populations of *L. novae-angliae* and sites with known extinctions of the flower. Using the regression tree analysis, *L. novae-angliae* occurrence could be accurately predicted 78.46% of the time. Four variables were found to be consistently statistically significant: Mean Temperature of the Wettest Month, Precipitation of the Wettest Month, Precipitation of the Wettest Quarter, and Precipitation of the Coldest Quarter. All analyses show that with increased precipitation at a site, the probability of the northern blazing star being extant at that site decreases. Conversely, increased temperatures at a site lead to increased probabilities of *L. novae-angliae* being extant. These findings correspond to previous knowledge of the plant, specifically preference for dry, warm grassland habitats. However, it was also shown that the northern blazing star might be susceptible to high seasonality and temperatures too warm, which is not promising in this ever-warming and fluctuating environment. Conservation efforts concerning *L. novae-angliae* should include climatic data, specifically temperature and precipitation, along with habitat loss factors when creating conservation plans. Using such data, conservation efforts could be preemptively prioritized, depending on the temperature and precipitation values at a particular site, which could potentially aid this struggling species.

Introduction

Detecting differences between current habitats of species and areas where they have gone extinct is critical for the conservation of rare species. However, this is often rare, due to the lack of data accurately pinpointing areas where species are missing but have historically occurred. Studies instead tend to concentrate on pair-wise comparisons of rare and common species (Kelly and Woodward 1996; Murray et al. 2002; Farnsworth 2007) or estimating species ranges to find correlations between distribution and habitat variables (Gaston 2003; Rushton et al. 2004).

One such rare plant is *Liatris scariosa* var. *novae-angliae*, or the northern blazing star. The genus *Liatris* (Family Asteraceae) is a genus of around 40 North American herb species, including hybrids, and is found across the Midwest (Kane and Schmitt 2001). However, *Liatris novae-angliae* is a perennial plant endemic to New England, with only 82 extant occurrences at present (Kane and Schmitt 2001). The northern blazing star has disappeared from 69% of sites where it had historically occurred (Vickery 2002), and 85% of its currently extant populations have fewer than 500 individuals (Kane and Schmitt 2001; Gravuer et al. 2005). Once distributed in the sandplain grasslands of southern Maine to northern New Jersey, these small populations now occur in Maine, New Hampshire, Rhode Island, New York, Connecticut, and Massachusetts.

Liatris novae-angliae prefers early-successional, dry, open woods, clearings, and barrens, with sandy low-nutrient soil. They can be found in coastal plains, morainal grasslands, and even roadsides (Kane and Schmitt 2001). The loss of *L. novae-angliae*'s preferred habitat has been implicated in the species' decline, with over 90% of sandplain grassland and coastal heathland being lost since European settlement (Vickery 2002). Sandplain grassland communities are now considered globally endangered (Kane and Schmitt 2001). It is not only natural habitat loss impacting *L. novae-angliae* populations. In the mid-ninteenth century, less than 40% of land in Connecticut, Rhode Island, Massachusetts, and Vermont was forested (Kane and Schmitt 2001). However, loss of agricultural lands led to forest-dominated landscapes, undesirable habitat for *Liatris novae-angliae*.

This study aims to determine whether factors outside of habitat loss, such as climatic variables, can be used to predict *Liatris scariosa* var. *novae-angliae* occurrence. I use current presence data and historical records from Natural Heritage Programs' field surveys and herbarium records, allowing for comparison between sites where it currently exists and where it had previously existed. I hypothesize that the presence of *Liatris novae-angliae* can be predicted using temperature and precipitation data. I expect to see significant differences in the values of these variables at points where *L. novae-angliae* persists, compared to points where they have gone extinct. The findings of this study will elucidate the factors associated with the persistence of current Liatris populations, and understanding current distributions and traits associated with rare species' resilience are crucial for designing conservation strategies that target extinction-prone taxa (Farnsworth and Ogurcak 2006).

Methods

Data Collection

Liatris novae-angliae occurrence data were collected through the Natural Heritage Programs of Maine, Massachusetts, Connecticut, Rhode Island, and New York (Figure 1). The data totaled 71 presence points and 59 absence points. I went through the individual Element Occurrence Records (EOR) summaries to assign a point as extant or extinct, based on the field surveys conducted. For those records that did not include latitude and longitude information, the QGIS plugin, Coordinate Capture, was used to obtain the coordinates at those points (QGIS Development Team 2015). The latitude and longitude of records whose formats were shapefile polygons were taken at the polygon centroids.

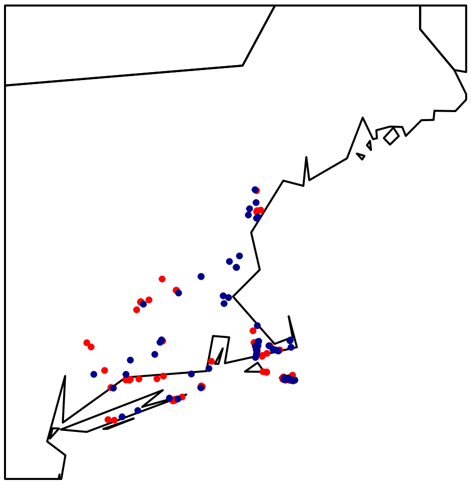

Figure 1 A map of the Northeastern United States with the extant points of *Liatris scariosa* var. *novae-angliae* in red and the extinct points in blue.

Bioclimatic data were obtained using the getData function in the package raster (Hijmans 2015) in R (R Core Team 2015), which retrieves the 19 'BioClim' variables from WorldClim (Hijmans et al. 2005) and had a spatial resolution of 2.5 minutes (or 4.5 kilometers at the equator). The variables included Annual Mean Temperature, Mean Diurnal Range, Isothermality, Temperature Seasonality, Maximum Temperature of the Warmest Month, Minimum Temperature of the Coldest Month, Temperature Annual Range, Mean Temperature of the Wettest Quarter, Mean Temperature of the Driest Quarter, Mean Temperature of the Warmest Quarter, Mean Temperature of the Coldest Quarter, Annual Precipitation, Precipitation of the Wettest Month, Precipitation of the Driest Month, Precipitation Seasonality, Precipitation of the Wettest Quarter, Precipitation of the Driest Quarter, Precipitation of the Warmest Quarter, and Precipitation of the Coldest Quarter. Temperature was recorded in degrees Celsius (°C), while precipitation was recorded in millimeters (mm). Using the latitude and longitude coordinates, climate data values were extracted at those points. Units of temperature were converted, since the original values were multiplied by 10 in order to allow for the storage of integer values.

Data Analysis

To determine whether data points were spatially clustered or spatially independent, two convex hulls were created, using the adehabitatHR package in R (Calenge 2006) that surrounded 95% of the extant and extinct points. The areas of the convex hulls were obtained using the rgeos package (Bivand and Rundel 2015). A thousand convex hull simulations were repeated using shuffled extant identities to determine where the area of the initial observed hull fell in that distribution.

One-way ANOVAs were performed to determine if any of the 19 variables were significantly different between the extant and exinct sites. Boxplots were created to visualize the values of the significant variables. To determine if the extant and extinct groups could be discriminated, a discriminant analysis using all 19 variables and a discriminant analysis using the four significant variables from the ANOVA were conducted. The analyses were performed using the candisc package in R (Friendly and Fox 2015), and the scores were plotted for the analysis with 4 variables.

A multiple logistic regression was conducted with all 19 variables, and subsequently, individual logistic regressions were conducted on the significant variables. All regressions used the glm function in the stats package, and their curves were plotted using the predict fuction, also in the stats package (R Core Team 2015).

A predictive modeling approach was also used through the construction of a regression tree, using the rpart package (Therneau et al. 2015), along with the predict function. The minimum number of observations existing in a node before a split could be attempted was set to 20.

Results

Convex Hull

The convex hull of the presence points was plotted as an yellow polygon with 95% of the red presence points plotted inside (Figure 2). Similarly, 95% of the extinct points were plotted as blue points inside of the orange polygon, representing the extinct convex hull. The orange polygon overlaps with most of the yellow polygon range, but they were of similar size, with the presence hull area of 5.3758 and extinct hull area of 4.9035.

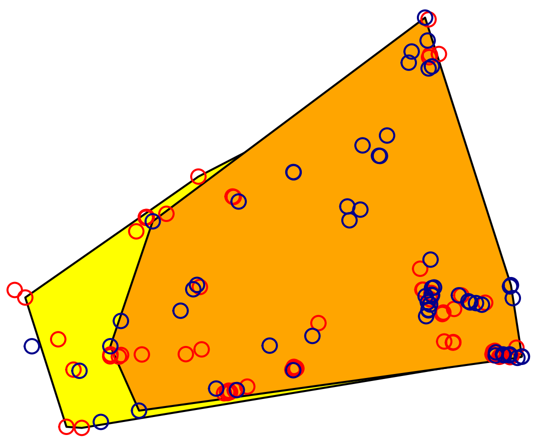

Figure 2 Overlapping convex hulls, which incorporate 95% of extant points (yellow polygon with red points) and extinct points (orange polygon with blue points).

Those area values were not significantly different from the means of the 1000 simulated convex hulls, since they fell within the 95% confidence interval around the simulated presence and extinct distribution means, 5.0207 and 4.9674, respectively (Figure 3).

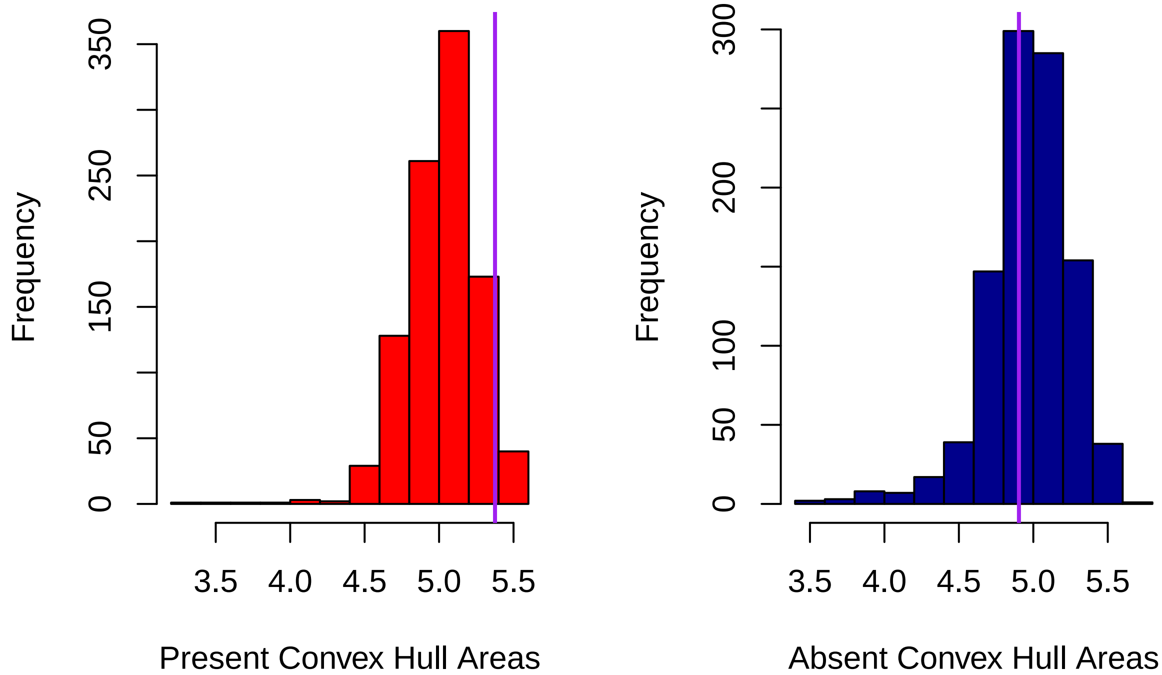

Figure 3 Histograms of the distribution of 1000 simulated convex hulls of presence sites (red) and extinct sites (blue) with the observed area plotted as a purple line for both present (5.3758) and extinct (4.9035) hulls.

ANOVA

The Analysis of Variance (ANOVA) showed four variables significantly differed among sites where *L. novae-angliae* is currently still present, as opposed to where it has gone extinct. These four variables were Mean Temperature of the Wettest Month (p-value 0.0112), Precipitation of the Wettest Month (p-value 0.000949), Precipitation of the Wettest Quarter (p-value 0.00239), and Precipitation of the Coldest Quarter (p-value 0.0104). The rest of the variables had a p-value greater than 0.05. The significant variables were plotted as boxplots (Figure 4). For each of the precipitation variables, the sites where *L. novae-angliae* has gone extinct have higher precipitation values (mm), with the wettest month of extinct sites having a mean of 166.59 mm, coldest quarter having a mean of 297.1 mm, and wettest quarter having a mean of 324.1. The extant sites had means of 112.56 mm, 289.65 mm, 317.03 mm, respectively. However, the extant sites have higher mean temperatures of the wettest month (5.6°C) compared to extinct sites (3.9°C).

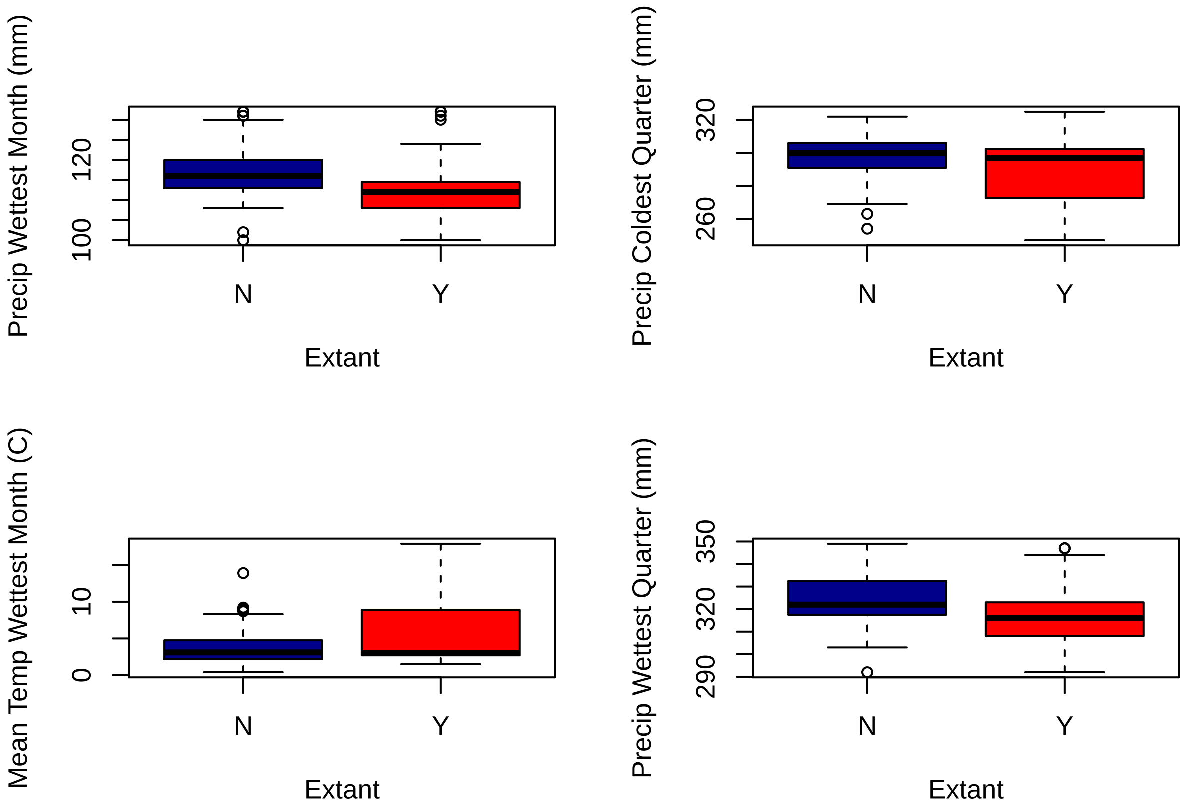

Figure 4 Boxplots of the variables deemed significant by the ANOVA. Precipitation of the wettest month in millimeters (top left), precipitation of the coldest quarter in mm (top right), mean temperature of the wettest month in Celsius (bottom left), and precipitation of the wettest quarter in mm (bottom right) in extinct (blue) versus extant (red) groups

Discriminant Analysis

Both analyses yielded similar results, so the analysis that used the 4 significant variables was reported and plotted. The canonical discriminant analysis yielded a p-value of 0.0004518, suggesting the canonical correlation value was not zero. Wilks Lambda was 0.908, with an approximate F statistic of 12.968, the eigenvalue was 0.1013, and R^2 was 0.09199. The calculated scores for the canonical discriminant function were plotted (Figure 5), but no clear discrimination between extant and extinct sites can be seen.

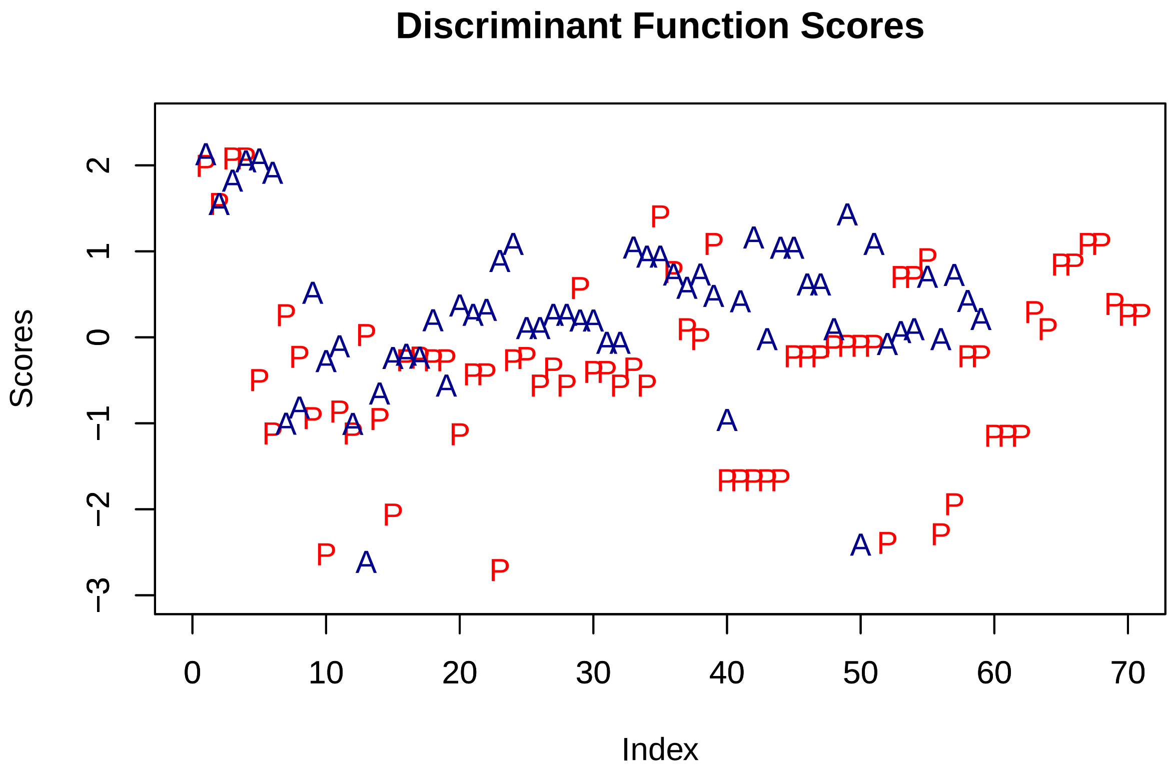

Figure 5 A plot of the canonical discriminant function scores for present sites (red) and extinct sites (blue). The discriminant function used four variables: Mean Temperature of the Wettest Month, Precipitation of the Wettest Month, Precipitation of the Wettest Quarter, and Precipitation of the Coldest Quarter. No clear discrimination can be seen.

Logistic Regression

The multiple logistic regression yielded seven significant variables, with only one overlapping with the significant ANOVA variables (Precipitation of the Coldest Quarter). However, when uni-variate logistic regressions were conducted, the significant variables matched with the significant variables from the previous ANOVA: Precipitation of the Wettest Month (logistic regression p-value 0.00196), Coldest Quarter (p-value 0.0127), and Wettest Quarter (p-value 0.00351), along with the Mean Temperature of the Wettest Month (p-value 0.0153).

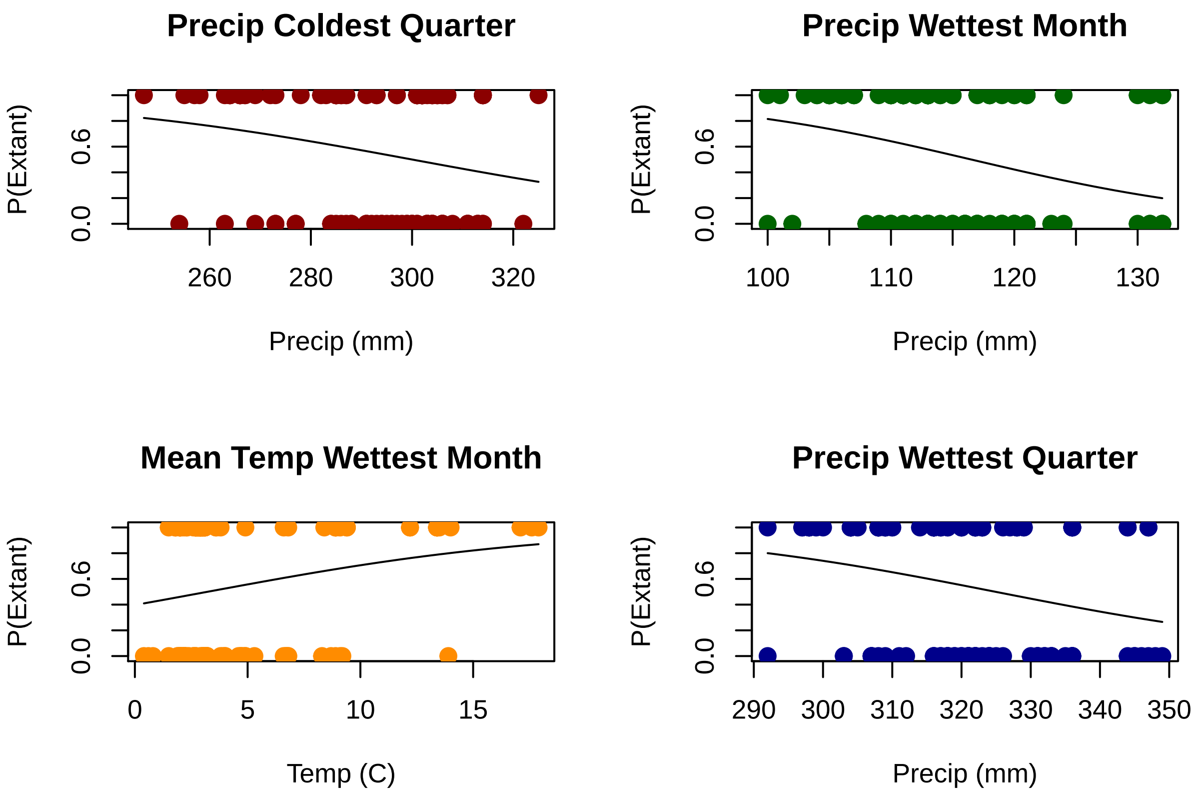

Figure 6 A plot of the individual logistic regressions for precipitation of the coldest quarter (top left), precipitation of the wettest month (top right), mean temperature of the wettest month (bottom left), and precipitation of the wettest quarter (bottom right) with the probability of *Liatris scariosa* var. *novae-angliae* being extant at a particular site on the y-axis.

The predicted curves of the logistic regression of individual variables were plotted (Figure 6). The precipitation curves all show that with increased millimeters of precipitation at a particular site, the probability of *L. novae-angliae* being extant at that site decreases. For the mean temperature of the wettest month, as degrees Celsius increase, the probability of *L. novae-angliae* being extant also increases.

Table 1 The confusion matrix generated from the regression tree, displaying the correct classification rates of extant and extinct predictions for *Liatris scariosa* var. *novae-angliae*.

	Predicted N	Predicted Y
No	45	14
Yes	14	57

Regression Tree

The correct classification matrix of the regression tree is displayed in Table 1. False positives and false negatives both occurred 14 times each. Predicting extinct sites had a correct classification rate of 76.27%, and extant sites had a correct classification rate of 80.28%. The overall correct classification rate was 78.46%.

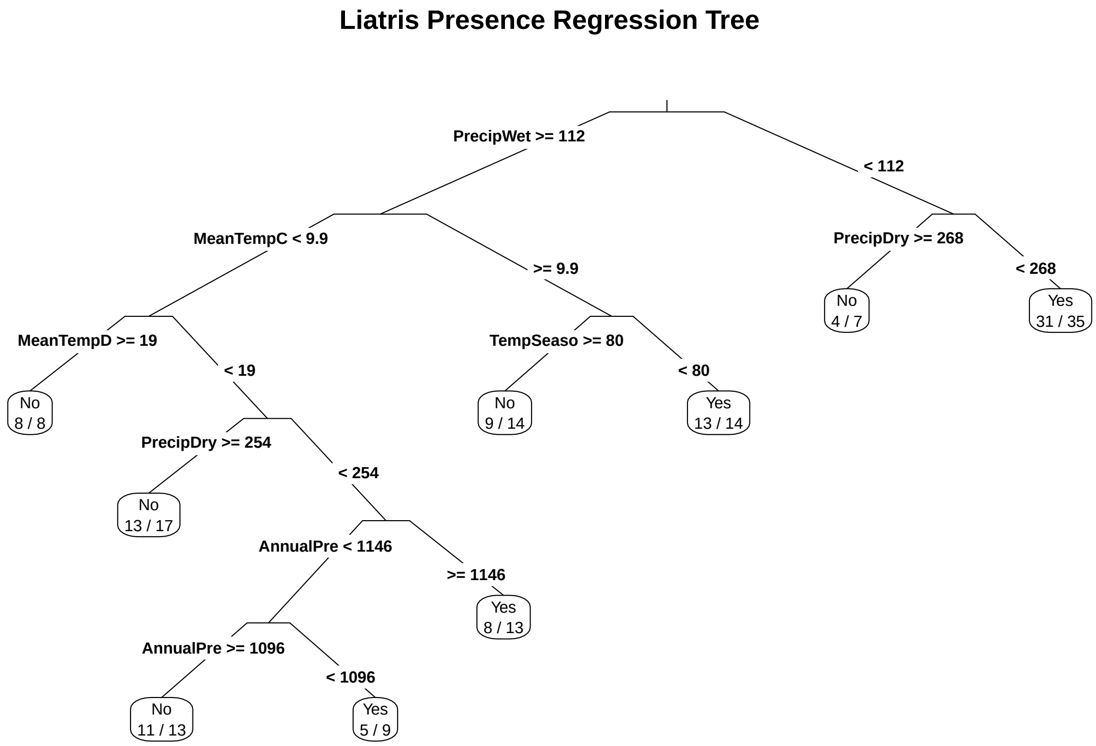

Figure 7 The regression tree predicting the presence of *Liatris scariosa* var. *novae-angliae* using the variables Precipitation of the Wettest Month (PrecipWet), Precipitation of the Driest Month (PrecipDry), Annual Mean Temperature (MeanTempC), Mean Temperature of the Driest Quarter (MeanTempD), Temperature Seasonality (TempSeaso), and Annual Precipitation (AnnualPre). Precipitation is in units of millimeters and temperature in degrees Celsius. The correct classification ratios are in the boxes of the terminal ends.

The regression tree used 6 of the 19 variables: Precipitation of the Wettest Month (PrecipWet), Precipitation of the Driest Month (PrecipDry), Annual Mean Temperature (MeanTempC), Mean Temperature of the Driest Quarter (MeanTempD), Temperature Seasonality (TempSeaso), and Annual Precipitation (AnnualPre). The tree had 8 splits that led to conclusions of whether *Liatris novae-angliae* was predicted to be present at a site, with the correct classification ratio on the ends of the branches (Figure 7). Looking at the first split, we see that the predictive ability of the variable Precipitation of the Driest Month is dependent on the Precipitation of the Wettest Month. For example, if precipitation of the wettest month is less than 112 mm, but also less than 268 mm in the driest month, *L. novae-angliae* is correctly predicted to be present 88.57% of the time.

Discussion

This study aimed to determine whether the presence of the rare, perennial flower, *Liatris scariosa* var. *novae-angliae*, could be predicted based on current occurrence data and data of historical presence. These preliminary findings suggest that, yes, *L. novae-angliae* occurrence can be accurately predicted based on bioclimatic data.

The similar size and spatial extent of the generated convex hulls (Figure 2) when compared to sites with randomly assigned 'present' or 'extinct' identities (Figure 3) suggest that neither the extant nor extinct sites are spatially clustered. Therefore, the bioclimatic variables between sites could be statistically analyzed.

From the ANOVA and individual logistic regressions (Figure 6), four variables are consistently significant: Mean Temperature of the Wettest Month, Precipitation of the Wettest Month, Precipitation of the Wettest Quarter, and Precipitation of the Coldest Quarter. All of the precipitation variables have higher values at sites where *L. novae-angliae* has historically occurred, but no longer persists (Figure 4). The logistic regressions of precipitation support this conclusion by showing the probability of *L. novae-angliae* being extant at a particular site decreasing as precipitation increases (Figure 6). Conversely, the temperature variable had higher values at sites where *L. novae-angliae* currently exists (Figure 4) and is supported by the logistic regression, which predicts higher probability of *L. novae-angliae* presence with warming temperatures (Figure 6). This suggests that sites that have been extirpated have had more precipitation and cooler temperatures, when looking at recent climatic conditions. This corresponds to known data of the plant, since dry and warm conditions are associated with grassland habitats (Kane and Schmitt 2001).

Both the discriminant analyses using all 19 variables and only the four significant variables from the ANOVA (Mean Temperature of the Wettest Month, Precipitation of the Wettest Month, Precipitation of the Wettest Quarter, and Precipitation of the Coldest Quarter) did not show clear discrimination between groups (Figure 5). This suggests the variables carry some complexity that cannot be captured with a linear discrimination analysis, so a regression tree was constructed (Figure 7). It had better success, with 78.46% correct classification of sites (Table 1). Looking at the regression tree, some conclusions correspond to the ANOVA and logistic regression findings. For example, less precipitation in the wettest and driest months lead to a correct classification of extant 88.75% of the time (Figure 7). This agrees with the ANOVA and logistic regressions that more precipitation is associated with extinct sites. An interesting aspect to note is that the regression tree only used one of the variables deemed significant through ANOVA and logistic regression: Precipitation of the Wettest Month. Instead of including the mean temperature of the wettest quarter and the precipitation of the wettest quarter, the opposing variables were used: Mean Temperature and Precipitation of the Driest Quarters. It seems the same variation is being captured with these opposite variables, and that they can still be used for classification, even if not deemed significant by ANOVA.

Looking ahead, the warming environment could have contrasting effects on the persistence of *Liatris scariosa* var. *novae-angliae*. On one hand, yes, *L. novae-angliae* prefers warmer conditions. However, the regression tree (Figure 7) shows that when the mean temperature of the driest month (MeanTempD) at a site is greater than 19°C (66.2°F), that site will be extinct (100% correct classification). Perhaps this species has an upper thermal limit and cannot persist in environments too warm. Climate change is also associated with extreme conditions. The regression tree (Figure 7) additionally shows that when the annual mean temperature (MeanTempC) is higher than 9.9°C (49.82°F), *L. novae-angliae* is predicted to be extant in locations with lower seasonality (TempSeaso; 92.86% correct classification). This implies extreme temperatures as a result of climate change may not bode well for *Liatris novae-angliae*.

Other parts of the classification tree do not make too much sense. For example, sites with annual precipitation (AnnualPre) higher than 1146 mm and lower than 1096 mm are predicted to have extant *Liatris* populations (low correct classifications of 61.54% and 55.56%, respectively). This could be due to a combination of the variation in the data, along with the possible over-fitting of the model, since a subset of the data was not used to train the model. Other sources of error could stem from the imprecision of the EOR summaries, as some populations have not been surveyed in decades. Additionally, for the purposes of this analysis, historic and recently extirpated populations were both lumped into the 'extinct' group, which could account for some variation of the data. Finally, although the means were shown to be significantly different in some variables, the medians were usually quite similar.

Further analyses with more data are required to support these preliminary findings. I would like to further separate the extinct group into recently extirpated and historical populations and use historical climate data to delve deeper. I would also like to add more predictor layers, such as land use and soil type, to determine if more accurate predictions can be made.

The New England Conservation and Research Plan (Kane and Schmitt 2001) for *Liatris scariosa* var. *novae-angliae* lists six major threats: habitat loss, destructive mowing regimes, deer grazing, seed predation, herbicide use, and collection and lack of awareness. These findings suggest that areas with higher precipitations and cooler temperatures are under greater threat of extirpation. With an ever-changing climate, we need to take these climatic variables into consideration when creating conservation guidelines, and possibly prioritize on a site-by-site basis.

Acknowledgements

I would like to thank my advisor, Dr. Nicholas Gotelli, and Elizabeth Farnsworth of the New England Wildflower Society for their extensive help with this project. I would also like to thank all of the New England Natural Heritage Program personnel (RI: David Gregg, MA: Tara Huguenin and Karro Frost, CT: Nelson DeBarros, ME: Lisa St. Hilaire, NY: Nicholas Conrad and Steve Young). Without their cooperation, this work could not have been done.

References

Bivand R. and Rundel C. 2015. rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-11. https://CRAN.R-project.org/package=rgeos

Calenge, C. (2006) The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. Ecological Modelling, 197, 516-519

Farnsworth E. J. 2006. Plant life history traits of rare versus frequent plant taxa of sandplains: Implications for research and management trials. Biological Conservation 136:44-52

Farnsworth E. J. and Ogurcak D. E. 2006. Biogeography and decline of rare plants in New England: Historical evidence and contemporary monitoring. 4:1327-1337

Friendly M. and Fox J. 2015. candisc: Visualizing Generalized Canonical Discriminant and Canonical Correlation Analysis. R packageversion 0.6-7. https://CRAN.R-project.org/package=candisc

Gaston, K. J. 2003. The structure and dynamics of geographic ranges. Oxford University Press, Oxford, U.K.

Gravuer K., von Wettberg E., and Schmitt J. 2005. Population differentiation and genetic variation inform translocation decisions for *Liatris scariosa* var. *novae-angliae*, a rare New England grassland perennial. Biological Conservation 124:155-167

Hijmans R. J. 2015. raster: Geographic Data Analysis and Modeling. R package version 2.4-20. https://CRAN.R-project.org/package=raster

Hijmans, R.J., Cameron S.E., Parra J.L., Jones P.G. and Jarvis A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Kane A. and Schmitt J. 2001. Liatris scariosa var. novae-angliae (northern blazing star) Conservation and Research Plan. New England Plant Conservation Program, Framingham, Massachusetts, USA.

Kelly, C. K., and F. I. Woodward. 1996. Ecological correlates of plant range size: Taxonomies and phylogenies in the study of plant commonness and rarity in Great Britain. Philosophical Transactions of the Royal Society of London, B, Biological Sciences 351: 1261–1269.

Murray, B. R., P. H. Thrall, and B. J. Lepschi. 2002. Relating species rarity to life history in plants of eastern Australia. Evolutionary Ecology Research 4: 937–950

QGIS Development Team. 2015. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rushton, S. P., S. J. Ormerod, and G. Kerby. 2004. New paradigms for modelling species distributions? Journal of Applied Ecology 41:193–200.

Therneau T., Atkinson B., and Ripley B. 2015. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-10. https://CRAN.R-project.org/package=rpart

Venables, W. N. and Ripley, B. D. 2002 Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Vickery P. D. 2002. Effects of the size of prescribed fire on insect predation of northern blazing star, a rare grassland perennial. Conservation Biology 16: 413-421