APPENDIX C

Oak Openings Ecological Niche Model

From a thesis entitled “A Regional Management Strategy for Invasive Plants in the Oak Openings”
by Sara N. Guiher, 2017
APPENDIX C Ecological Niche Modeling of Invasive Plants in the Oak Openings Region

From a Thesis entitled A Regional Management Strategy for Invasive Plants in the Oak Openings by Sara N. Guiher, The University of Toledo 2017

Abstract

Many efforts exist to manage the negative effects of invasive plant species. Since invasions are easier to treat in early stages, management strategies often focus on early detection and rapid response (EDRR). Using species occurrence data and environmental variables, we developed ecological niche models with Maxent to predict suitable habitat for seven invasive plant species (Alliaria petiolata, Berberis thunbergii, Celastrus orbiculatus, Elaeagnus umbellata, Frangula alnus, Rhamnus cathartica, and Rosa multiflora) within the Oak Openings region. Models were evaluated based on Maxent outputs as well as transect surveys for species presence. All species models performed well initially, with AUC values ranging from 0.85 to 0.93. Relationships between data from transect surveys and predicted suitable habitat were investigated using logistic regression. Probability of species presence increased significantly with predicted suitable habitat for all species except B. thunbergii and R. multiflora. Our results suggest that ecological niche models can be used to plan monitoring and management activities that focus on EDRR by predicting areas likely to be invaded by individual invasive plants.

Introduction

Natural areas are degraded by invasive plants due to a wide range of impacts that differ among invading species and the systems they colonize (Pyšek et al. 2012). Invaders of wetland communities can alter hydrological processes through changes in community structure, impeding benefits such as water storage and nutrient processing (Windham and Ehrenfeld 2003). Monocultures established by many invasive plants reduce the diversity of plant communities, which can in turn reduce the diversity present in higher trophic levels (Vila et al. 2011). In some cases, invasive plants increase risk to human health by providing habitat for disease vectors such as mice, decreasing the value of natural areas as recreational resources (Williams et al. 2009). To preserve these resources and native diversity, communities and conservation agencies must allocate time and money for invasive plant control.

Invasive plant management is often a major component of a restoration or conservation plan, and it may also be the most expensive and time consuming activity (Drucker et al. 2008). Management typically focuses on invasive species that pose the highest threat to the habitat being managed, as well as species that are emerging. These target species are given top priority in terms of available resources and are the focus of monitoring efforts aimed at early detection and rapid response (EDRR) (Carlson et al. 2008, Heibert and Stubbendieck 2008). While EDRR has proven to be an effective strategy, it is often difficult to implement because many invasive plants establish large or widespread populations before they are detected or recognized as a threat by land managers. Early detection is commonly achieved through monitoring of natural areas by conservation agencies and volunteers; time spent on this activity will be reduced by prioritizing areas for monitoring.

Knowledge of habitat requirements and vectors of spread for invasive plant species are essential components of a management strategy that incorporates targeted monitoring. Habitat characteristics that commonly influence invasive plant establishment include light and moisture availability, existing community type, and topography (Rouget 2003). Natural vectors of spread such as waterways and
animals, as well as anthropogenic vectors such as roadways and development, also help to determine where an invasion may occur (Hansen and Clevenger 2005). Many land managers and management staff develop an extensive inventory of where these factors occur on the lands under their care and use this knowledge to plan monitoring. However, these inventories are often anecdotal and are not accompanied by measurements or specific geographic coordinates. While this method may be effective with a small staff working in a limited area, larger agencies or multiple agencies working at a regional scale are not able to effectively compile or communicate this information. Because of this, monitoring efforts at a large scale are often difficult to coordinate.

Accurate predictive models of potential habitat for invasive plant species help to direct monitoring and management efforts to areas likely to support specific species (Zimmermann et al. 2011, Crall et al. 2013). Ecological niche models incorporate environmental characteristics and occurrence data to determine suitable habitat. The use of niche modeling to predict establishment and spread of invasive plants has been widely investigated (Thuiller et al. 2005, Jarnevich and Reynolds 2011). This method has been especially successful for specialized species or habitats; modeling of generalist species can be useful as well, although factors such as insufficient occurrence data (Hernandez et al. 2006) or a wider range of potential habitat (Gibson et al. 2004) likely affect model performance. Likewise, Evangelista et al. (2008) found that models for Tamarix chinensis, a habitat specialist, consistently outperformed those for the generalist species Bromus tectorum. While model performance may vary between species, increased occurrence and environmental data should increase model accuracy.

This study aims to examine potential distributions and vectors of spread for invasive plant species in the Oak Openings region of southeast Michigan and northwest Ohio. To achieve this goal, we first collected species occurrence data through a series of transect surveys. Because this region is lacking in large data sets for invasive species presence (with absence data rarely collected), methods that effectively analyze small presence-only datasets are particularly valuable. Maxent (Phillips et al. 2006) is effective with this type of data and has been used to direct monitoring efforts for invasive plants (Gormley et al. 2011). We used Maxent to create ecological niche models for seven invasive plant species. Resulting models were validated with additional transect surveys. Using these methods, we have developed a framework that can be used by land management agencies to plan monitoring and management activities for invasive plants.

Methods

Study Area

The Oak Openings region of southeast Michigan and northwest Ohio is a 3300 km² area of globally rare ecosystems driven by soils which are remnants of the last period of glaciation (Figure 1). Glacial clay is overlaid by a varying depth (< 1m to > 2m) of sand deposited by a series of glacial lakes. This soil structure results in distinctive habitat types ranging from upland sand barrens and oak savannas to sand flatwoods and wet prairies (Schetter and Root 2011). The wide variety of conditions present in the region leads to high levels of native biodiversity; this heterogeneity can also provide habitats that are suitable for a high diversity of invasive species (Shea and Chesson 2002).

Figure 1 The Oak Openings region as defined by soil characteristics

2
Modeling methods

Ecological niche models were created with Maxent software (version 3.3.3k) to predict suitable habitat for invasive plant species (Phillips et al. 2006). Models are created using species occurrence data and environmental data layers to identify suitable habitat. Maxent was selected because it has been used effectively as a tool to predict the potential distribution of plants using presence only data (Lemke et al. 2011, Marcer et al. 2012). This method was also chosen because Maxent is user-friendly and the environmental data required is readily available. Ecological niche models can then be incorporated with other invasive plant management techniques by conservation staff with a minimal need for additional training.

Species Occurrence Data

To address the small amount of invasive plant occurrence data in the region, transect surveys were completed at sites with varying management histories and habitat characteristics. Using these parameters, we chose 12 sites that were either accessible by the public or private lands that we were permitted to visit. Surveys focused on natural areas that may be targeted for management activities; we did not sample agricultural, suburban or urban areas. Transects were oriented within each site to intersect the full gradient of environmental conditions (e.g. vegetation cover, elevation, dominant community type). The number and length of transects per site varied depending on the size of the site and the range of conditions present; when multiple transects were completed in a site, each transect was placed a minimum of 30m away from any other transect. Plots were established at 30m intervals along the transect by marking a waypoint with a Garmin eTrex Venture HC handheld GPS unit. All invasive plants within a 5m radius of the waypoint were identified and recorded with the GPS unit. Species presence in the radius was marked once regardless of the population size.

Environmental Data

We developed niche models using six environmental variables: elevation, slope, aspect, distance to road, distance to stream, and Normalized Difference Vegetation Index (NDVI). All data layers were downloaded at a 30m grain, as this scale is readily available for a number of variables, facilitating continued use and addition of variables in the future. This scale is also appropriate to plan monitoring within individual sites. Elevation, slope, and aspect were interpolated from the US Geological Survey (USGS) National Elevation Dataset (USGS 2016) using ArcMap 10.3 (Environmental Systems Research Institute [ESRI] 2014). Aspect was transformed by cosine(aspect) to a linear variable, “northness.” Road data was obtained from the U.S. Census Bureau online catalog of TIGER shapefiles (U.S. Census Bureau 2014) and stream layers from the USGS National Hydrography Dataset (USGS 2016). Distance to road and distance to stream were calculated using the Euclidean distance tool in ArcMap 10.3. Landsat 8 satellite images collected by the USGS on June 19, 2016 were used to calculate NDVI. This index uses near-infrared radiation (NIR) and visible radiation (VIS) to calculate a range of plant density from -1 to +1, using the formula:

\[ \text{NDVI} = \frac{\text{NIR}-\text{VIS}}{\text{NIR} + \text{VIS}} \]

Soil composition strongly influences plant distribution in the region; however, available soil surveys do not reflect the fine scale variation in soil composition characteristic of Oak Openings habitats, and
therefore were not included in the models. Management intensity also has a significant impact on invasive plant populations, yet data on this variable is not available throughout the region.

Ecological Niche Models

Of the 20 species identified during surveys, seven species were recorded in sufficient numbers (> 10) to be used for niche models (Phillips et al. 2006): Alliaria petiolata, Berberis thunbergii, Celastrus orbiculatus, Elaeagnus umbellata, Frangula alnus, Rhamnus cathartica, and Rosa multiflora. Due to the relatively small sample sizes (n=24-72) in our study area, the Replication option (using cross validation) was used in Maxent, and 10 replicates were run for each species. Other Maxent settings were left as the default as recommended by Phillips and Dudik (2008). The average replicate output was used for the final models. Model performance was initially assessed with the area under the receiver operating characteristic curve (AUC) values produced by Maxent. A value of 0.5 suggests that the model performs as well as random, and a value of 1 suggests excellent predictive power, although AUC values can be misleading with small presence-only datasets (Raes et al. 2009). Maps of suitable habitat were created using model outputs; suitable habitat was delineated by the 10th percentile training presence logistic threshold provided by Maxent. This threshold was chosen to allow for possible sampling errors by basing the amount of suitable habitat on 90% of the data collected (Raes et al. 2009, Johnson et al. 2016).

Model Validation

We also tested models by surveying five sites that we had not previously surveyed. These sites were chosen because of varying management histories and accessibility as well as a broad range of suitable habitat predicted by Maxent. Surveys were structured to obtain occurrences of invasive species in an efficient manner, thus at each site a zig-zag path was walked constantly searching for invasive species (Thieme et al 2015). The seven target species were mapped using a handheld GPS unit whenever encountered, with recorded occurrences for the same species separated by a minimum of 30m. Tracks, automatically recorded as a series of points by the handheld GPS unit while walking transects, were used to identify absence locations. After data collection, occurrence points and track points were uploaded to ArcMap 10.3. Each occurrence point was given a value of 100, and each track point a value of 1. The track points were somewhat irregular; however the maximum number of track points in any given cell was 22. Values of these points were summed in each 30m cell using the Point to Raster tool in ArcMap. This allowed us to assign a value of zero (absence) to each cell that had a sum lower than 100, and a value of one (presence) to each cell with a sum of 100 or higher using R version 3.3.3 (R Core Team 2017). Relationships between test survey occurrence data and Maxent probability of suitable habitat were examined with logistic regression (significance threshold = 0.05) using the glm function in R.

Results

Surveys

Surveys for species occurrence data were completed along 26 transects totaling 7.1 km within 12 sites. Twenty invasive plant species were identified and mapped, adding a total of 368 additional presence points to regional data. Test surveys were completed at five sites and added 251 presence points for the seven species modeled.
Ecological Niche Models

Models predicted between 18 to 23% suitable habitat within the region for the species modeled: A. petiolata 18%, B. thunbergii 21%, C. orbiculatus 12%, E. umbellata 23%, F. alnus 22%, R. cathartica 22%, R. multiflora 19% (Figure 2). Suitable habitat for all species was closely associated with areas of higher NDVI value. Areas of predicted presence were concentrated where cover type such as meadows or forests persisted, with little suitable habitat found in agricultural or suburban settings. Differences in suitable habitat between species depended on the variables aside from NDVI that influenced each species. For example, R. cathartica habitat responded to stream presence more than E. umbellata, which was more dependent on the presence of roads.

Figure 2 Maxent predicted suitable habitat within the Oak Openings region for: a) A. Petiolata b) B. Thunbergii c) C. orbiculatus d) E. umbellata e) F. alnus f) R. cathartica g) R. multiflora

Environmental Variables

After initial model runs, aspect was omitted as an environmental variable because it decreased the predictive power in six of seven models. NDVI was the variable of highest percent contribution in each model, from 33.6% for C. orbiculatus to 71.7% for R. Multiflora. Differences between NDVI and the 2nd most important variable were substantial for every species except C. orbiculatus; for the remaining six species, NDVI contributed an average of 4.5 times more than the 2nd most important variable (Table 1). The relationship between suitable habitat and NDVI was generally positive, with a slight decrease in suitability for A. petiolata, B. thunbergii, F. alnus, and R. multiflora at the highest NDVI values (Figure 3).
Table 1 Number of occurrences, Maxent AUC value, and contribution of each variable for seven species modeled

<table>
<thead>
<tr>
<th>Species</th>
<th>n_{avg}</th>
<th>AUC_{avg}</th>
<th>Elevation</th>
<th>Slope</th>
<th>NDVI</th>
<th>Distance to Road</th>
<th>Distance to stream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alliaria petiolata</td>
<td>31</td>
<td>0.86</td>
<td>5.9</td>
<td>25.6</td>
<td>48.7</td>
<td>16.1</td>
<td>3.8</td>
</tr>
<tr>
<td>Berberis thunbergii</td>
<td>22</td>
<td>0.91</td>
<td>4.3</td>
<td>2.6</td>
<td>65.1</td>
<td>27.9</td>
<td>0.2</td>
</tr>
<tr>
<td>Celastrus orbiculatus</td>
<td>27</td>
<td>0.93</td>
<td>5.3</td>
<td>9.6</td>
<td>33.6</td>
<td>27.1</td>
<td>24.5</td>
</tr>
<tr>
<td>Elaeagnus umbellata</td>
<td>40</td>
<td>0.85</td>
<td>4.5</td>
<td>12.9</td>
<td>67.3</td>
<td>11.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Frangula alnus</td>
<td>58</td>
<td>0.90</td>
<td>14.2</td>
<td>1.5</td>
<td>64.4</td>
<td>14.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Rhamnus cathartica</td>
<td>22</td>
<td>0.85</td>
<td>5.5</td>
<td>8.3</td>
<td>66</td>
<td>9.5</td>
<td>10.7</td>
</tr>
<tr>
<td>Rosa multiflora</td>
<td>64</td>
<td>0.91</td>
<td>3.6</td>
<td>3.6</td>
<td>71.7</td>
<td>10.5</td>
<td>10.6</td>
</tr>
</tbody>
</table>

Figure 3 Maxent response curve for A. Petiolata to NDVI. A similar pattern was observed for B. thunbergii, F. alnus, and R. multiflora.

Model Validation
Maxent AUC values (average of replicates) ranged from 0.85 to 0.93 (Table 1) suggesting that the models have good to excellent predictive power. Logistic regression showed a significant positive relationship between Maxent probability of suitable habitat and probability of species occurrence for A. petiolata ($p < 0.01$), C. orbiculatus ($p < 0.01$), E. umbellata ($p = 0.03$), F. alnus ($p < 0.001$), and R. cathartica ($p < 0.001$). There was no relationship between model prediction and species occurrence for B. Thunbergii ($p = 0.5$) and R. multiflora had a significant negative relationship ($p = 0.03$). For the remaining five species, detection probability was two- to five times higher at suitable habitat probabilities above 0.8 than it was at a probability of 0.2 (Figure 4).
Figure 4 Logistic regression models for five species showing a positive relationship between Maxent predicted suitable habitat and probability of species presence
Discussion

Species presence had a significant positive relationship with predicted suitable habitat for five of seven invasive species, indicating that ecological niche models can be an effective component of management plans that focus on EDRR. The landscape features most influential to our model results were NDVI and distance to road. The ability to run models at a fine scale allows for planning at multiple scales, from regional to individual sites. Agencies can use this information to increase collaboration and implement invasive plant management that has maximum impact throughout the Oak Openings. Additionally, this methodology can easily transfer to other regions using the variables that are most important in each area.

Individual environmental variables had varying levels of impact on models for each species. NDVI had the greatest influence in each, with probability of suitable habitat increasing with vegetation density, then decreasing when approaching the maximum density. These results are not unexpected, since many plants are not able to persist in areas of dense cover. Vegetation cover is also of great importance for other reasons: Oak Openings community types with a low vegetation density are often maintained through management or hydrological regimes that are inhospitable to invasive plants. Monocultures formed by invasive plants also increase vegetation density, further supporting a strong positive relationship between NDVI and invasive species presence. Although this metric offers minimal detail about habitat, the influence of NDVI will help land managers begin to pinpoint areas of potential invasion.

Distance to road was also a major factor, as it was the 2nd highest contribution in three of the models, and the 3rd highest in the remaining four. We expected this variable to have higher percent contributions than it did in our results because roads are well known vectors of spread for invasive plants (Taylor et al. 2012, Ansong and Pickering 2013). These unexpected results may be attributed to the fact that the relationship between roads and invasive species is more complex, as well as a lack of sampling sites in suburban areas with a high density of roads. Although invasive plants are often found in home gardens, this is a reflection of horticulture rather than suitable habitat. For modeling purposes, areas of high human populations might provide valuable information as sources of invasive plant propagules.

Model discrimination did not directly increase with the number of occurrence points as expected. The model for C. orbiculatus (n=27) had the highest AUC of 0.93, while E. umbellata (n=40) had the lowest AUC of 0.85. This may be due to varying levels of specialization among plant species rather than variance in model performance; a generalist species usually requires a greater number of occurrences than a specialist to inform models (Evangelista et al. 2008). Species that are not as widespread, and therefore do not occupy their full niche, also result in less accurate predictive models (Crall et al. 2013). As ongoing surveys take place, we expect increasing presence points to increase the accuracy of model outputs for each species.

Ecological niche modeling will be more effective with an expanded database of presence points for invasive plant species in the region (Hernandez et al. 2006). Increasing the amount of regional data will require a concentrated monitoring effort among regional agencies. However, agency employees already spend ample time in the field observing invasive plant locations. These observations are quickly recorded using handheld GPS units and mobile applications such as the Midwest Invasive Species Information Network (MISIN). A minimal time investment recording presence data in the field will
provide more robust predictive models that will allow for targeted monitoring that focuses on EDRR. Monitoring efforts targeting native plants and animals can also be used to create ecological niche models. Areas in the Oak Openings with a high probability of invasion that are likely to support desirable or rare native species would be ideal places to concentrate management activities.

This study was designed to explore the potential distribution as well as the impact of individual environmental variables for selected species within the Oak Openings. Even with small datasets, models were able to predict suitable habitat for invasive plant species. The models produced showed that two known vectors, roads and streams, have varying influence on the distribution of individual species. Our results indicate that ecological niche models can be a valuable tool for land managers to apply to monitoring and management activities. Accurate predictive models, combined with knowledge of species requirements and land manager expertise, will increase the efficiency and impact of invasive plant management efforts in the Oak Openings.

References


