Species Distribution Modeling of North American Odonates

Rebecca Beilinson
Yale University

Follow this and additional works at: https://elischolar.library.yale.edu/library_map_prize

Part of the Biodiversity Commons, Biology Commons, Entomology Commons, and the Terrestrial and Aquatic Ecology Commons

Recommended Citation
https://elischolar.library.yale.edu/library_map_prize/5

This Article is brought to you for free and open access by the Library Prizes at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Library Map Prize by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Species Distribution Modeling of North American Odonates

A Senior Research Project for Ecology and Evolutionary Biology

Yale University

Rebecca Beilinson

Advisor: Dr. Walter Jetz

29 April 2016
Abstract

Species distribution modeling, which uses species-environment relationships to predict species’ geographic ranges, is a powerful technique for biogeographical and ecological analysis as well as for conservation planning. However, it has typically been underutilized for invertebrates and for freshwater species, as modeling these groups can pose unique challenges. Here I present methods for modeling odonate distributions and demonstrate these methods for four North American species. I show that incorporating expert-derived range maps as spatial priors can significantly improve model performance as compared to MaxEnt models. I also introduce a new distance to freshwater layer as an environmental variable, and show that this is an important predictor of presence for all tested species. The methods and example cases presented here contribute to a greater understanding of this ecologically important but understudied taxonomic group.

Introduction

Environmental change caused by human activity poses a severe threat to biodiversity worldwide. Current extinction rates are orders of magnitude higher than background rates (Vitousek et al. 1997), and freshwater systems are especially vulnerable to biodiversity loss (Dudgeon et al. 2006). In order to conserve and protect species, it is crucial to understand their biogeography and spatial distribution (Jetz et al. 2012).

One key method used to study species’ biogeography is species distribution modeling. Species distribution models (SDMs) use species-environment relationships to predict areas that contain suitable habitat for a given species (Pearson 2007, Elith and Leathwick 2009). These predictions are generated by combining species occurrence records (observations of individuals) with background climatic or topographic
information. They have a wide range of applications, including predicting species response to climate change, designing reserves for conservation purposes, quantifying biodiversity hotspots, and predicting species geographic ranges when occurrence data is scarce (Elith and Leathwick 2009, Collins and McIntyre 2015).

While species distribution modeling has become a popular method of ecological analysis and conservation planning, relatively little work has been done on insects as compared to vertebrates and flowering plants (Ballesteros-Mejia et al. 2013), despite the fact that insects make up the majority of the world’s biodiversity. Species-level occurrence records, which are necessary for SDMs, are often lacking for invertebrates and there has not been as much interest in insects as there has been in more charismatic species. Where insect species records are available, they are often collected opportunistically rather than as part of systematic surveys, which can lead to biased data. However, one group of insects that has drawn considerable attention is Order Odonata—dragonflies and damselflies. Because they are charismatic insects that are relatively easy to identify by sight, there is significantly more occurrence data for odonates than other invertebrates (Collins and McIntyre 2015).

Odonates have also drawn attention because they are very sensitive to environmental change (such as changes in temperature and water quality) and therefore serve as indicators of freshwater ecosystem health and biodiversity (Clausnitzer et al. 2012, Simaika et al. 2012). Dragonflies are effective indicators of threat status for birds, mammals, and amphibians when making conservation decisions, but those groups are not always accurate indicators for dragonflies or other invertebrates (Simaika et al. 2012).
Therefore, identifying important areas for dragonfly conservation can also provide valuable insights for the effective conservation of numerous other freshwater species.

A number of studies, reviewed by Collins and McIntyre (2015), have modeled Odonata distributions in recent years. Modeling odonate distributions presents some unique challenges. Odonates have an aquatic larval stage and terrestrial adult stage; they therefore require freshwater habitats, but adults may be observed in areas that are not suitable for larvae (Kalkman et al. 2008). This separation of niches between adults and juveniles makes odonate distribution particularly difficult to model, and it may be necessary to model adults and larvae separately to gain a full understanding of their distribution (Patten et al. 2015). Though interest in modeling odonate distributions has increased recently, few studies (three of the thirty reviewed by Collins and McIntyre 2015) have focused on North American odonates, and most only modeled a few species.

In this paper, I develop methods to model Odonate distributions for all species across North America, and present full results for four case species. I build on modeling methods using generalized linear models (GLMs) and MaxEnt methods, and then present a novel method for modeling Odonate distributions using spatial priors derived from expert range maps. I also introduce two new environmental layers based on remote-sensing imagery, which may be useful for modeling many freshwater species. By improving SDM methods for dragonfly and damselfly species, I hope to open the door to fine-grain analyses of these important insects over large spatial extents.
Methods

Species Occurrence Data

Species occurrence records were compiled from several databases: the Global Biodiversity Information Facility (gbif.org 2015), iNaturalist (inaturalist.org 2015), Odonata Central (Abbott 2006-2016), and Biodiversity Information Serving Our Nation (BISON) (U.S.G.S. 2015).

The data from iNaturalist and Odonata Central is citizen science data; the data has been collected by amateur enthusiasts who upload their species observations to online databases. Citizen science has become an increasingly widespread source of biodiversity data for scientists in the past decade and has been successfully used to augment species occurrence records (Barve 2014, Silvertown 2009). The records from citizen science data in this study were only included if they were designated as “research-grade” or vetted by experts.

Any occurrence records without geographic coordinates or dates were excluded. Records were only included if they dated from 1950 onwards, as this is the time scale of the available environmental data. Where coordinate accuracy information was available, I only included points with coordinate accuracy of under 1 km, which is the grain of the environmental data. Only records of adult specimens were included. Lastly, all duplicate records were removed from the combined dataset. The combined and cleaned data yielded a total of 181,386 records of 786 distinct species or taxonomic units.

Environmental Data

Environmental variable layers were compiled from a variety of sources (Table 1). Variables were selected that were assumed to be ecologically significant for dragonfly
species, and preliminary models confirmed their significance for test species. All layers covered the extent of North America. Hydro1K layers were re-sampled from a 1 km equal area grid to a 30 arcsec grid using nearest neighbor resampling. All other layers had a 30 arcsec (approximately 1 km) spatial grain.

Because odonates have aquatic larval stages, hydrological variables are likely to influence their distributions. Most freshwater species distribution models use hydrological variables that are derived from digital elevation models (DEMs), which model stream networks using topographic features as opposed to observed water availability. However, these models may not always accurately predict the location of water. I therefore introduce two new layers, derived from satellite-based Landsat imagery (Hansen et al. 2014; Woodcock et al. 2008), which identify water using direct remote sensing observation. The first of these layers is a proportion freshwater layer, derived from 30 m resolution satellite imagery aggregated to 1 km cells (Amatulli, unpublished, based on Hansen et al. 2014). The second layer is a distance to freshwater layer, which measures the distance to the nearest grid cell containing any freshwater. Both DEM-derived and satellite imagery-based sets of variables were included in the models.
Table 1. Environmental variables used as predictors in species distribution models for North American Odonata.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Name</th>
<th>Source of Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean annual temperature</td>
<td>Bio_1</td>
<td>WorldClim (Hijmans et al. 2005)</td>
</tr>
<tr>
<td>Mean annual precipitation</td>
<td>Bio_12</td>
<td>WorldClim (Hijmans et al. 2005)</td>
</tr>
<tr>
<td>Precipitation seasonality</td>
<td>Bio_15</td>
<td>WorldClim (Hijmans et al. 2005)</td>
</tr>
<tr>
<td>Land cover of mixed/other trees</td>
<td>Lc_class4</td>
<td>EarthEnv (Tuanmu and Jetz 2014)</td>
</tr>
<tr>
<td>Land cover of cultivated/managed land</td>
<td>Lc_class7</td>
<td>EarthEnv (Tuanmu and Jetz 2014)</td>
</tr>
<tr>
<td>CTI ground wetness index</td>
<td>cti</td>
<td>HYDRO1K (USGS)</td>
</tr>
<tr>
<td>Flow accumulation</td>
<td>fa</td>
<td>HYDRO1K (USGS)</td>
</tr>
<tr>
<td>Proportion land vs. proportion freshwater</td>
<td>Proportionland</td>
<td>Giuseppe Amutulli, unpublished, based on Hansen et al. 2014</td>
</tr>
<tr>
<td>Distance to nearest cell containing freshwater</td>
<td>Waterdistance</td>
<td>New layer, based on Hansen et al. 2014</td>
</tr>
</tbody>
</table>

Species Selection

In order to determine how the models perform for groups of species with different ecological and geographical requirements, species were divided into groups. Only species with a minimum of 1000 presence records were considered for modeling. These groups included ecological differences (lentic vs. lotic) and taxonomic differences (dragonflies vs. damselflies) which are also associated with distinct ecological niches. Four focal species were selected from groups with ecological and phylogenetic differences: *Argia vivida* (1769 presence records), *Erythemis simplicicollis* (5278 presence records), *Lestes rectangularis* (1135 presence records), and *Libellula vibrans* (1231 presence records).
*Lestes rectangularis* and *Libellula vibrans* are both lentic species, whereas *Erythemis simplicicollis* is lotic and *Argia vivida* is a generalist that can use both lentic and lotic water. While *Argia vivida* and *Lestes rectangularis* are damselflies (suborder Zygoptera), the other two species are dragonflies (suborder Anisoptera). The ranges of these four species (based on expert range maps) also vary in size and location, and they therefore serve as a representative sample to test the performance of the models across the continent.

**Modeling**

All analyses were performed in R (R Development Core Team 2016). Preliminary testing and variable selection was performed using generalized linear models (GLMs), because they are computationally simpler and faster to run than the final models. Various combinations of variables were tested using GLMs, and the model formula with the lowest average area under the curve (AUC; Swets, 1988) was selected for the final models. This final model formula was:

\[
\text{Presence} \sim \text{bio}_1 + \text{bio}_12 + \text{bio}_15 + \text{lc}_\text{class}4 + \text{lc}_\text{class}7 + \text{cti} + \text{fa} + \text{proportionland} + \text{waterdistance}
\]

Once this model formula was selected, two methods of modeling were compared: MaxEnt (without spatial priors), and spatial priors models. MaxEnt models, or maximum entropy models, are multinomial logistic regressions that can be used for modeling species distributions given presence-only data (Merow et al. 2013). My simple MaxEnt models did not include priors and served as a baseline model against which to compare results from the spatial priors models. The spatial priors models, in contrast, take into
account expert range maps as priors (Domisch et al. 2015). The methods used here for spatial priors follow methods developed by Merow et al. (Manuscript).

Both models were run for the four focal species. The available data was presence-only, so I used 10,000 randomly selected pseudo-absences (also known as background sample points) for each model. The data was split with 70% of the data for model fitting and 30% for validation.

Expert range maps from Paulson (2009, 2011) were downloaded from Map of Life (mol.org) to be used for priors in the spatial priors model. The modeling domain for each species was set as the range map for that species plus a five-degree bounding box in each direction to allow sufficient space for background samples. The model with the lowest Akaike Information Criterion (AIC) was selected as the best model, and compared to a MaxEnt model with no priors. The performance of both the MaxEnt and the best spatial priors was then tested by AUC.

Thresholds

The output of spatial priors models are given as a relative occurrence rate across the modeling extent. I applied several thresholds to this output in order to determine where presence was most likely to occur according to multiple measures (Table 2). For each measure, a specific threshold value was determined based on model outputs, and these thresholds were use to make binary maps where a value of one indicated that the given grid cell was above the threshold and a value of zero indicated that it did not meet the threshold. These binary threshold maps were then summed to produce a map where
each cell could have a value between zero and seven, with zero indicating that no thresholds were passed and seven indicating that all seven thresholds were passed. Thus areas with high threshold sums had a high probability of presence according to several unique measures. This allowed for a simple visualization of where species were most likely to occur.

<table>
<thead>
<tr>
<th>Threshold Name</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSS</td>
<td>Maximum True Skills Statistic (TSS), indicating the minimal difference between sensitivity and specificity</td>
</tr>
<tr>
<td>Min.occurrence.prediction</td>
<td>Minimum prediction for presence records</td>
</tr>
<tr>
<td>Mean.occurrence.prediction</td>
<td>Mean prediction for the presence records</td>
</tr>
<tr>
<td>Max.kappa</td>
<td>Threshold at which kappa is maximized</td>
</tr>
<tr>
<td>Spec_sens</td>
<td>Threshold at which the sum of sensitivity and specificity is maximized</td>
</tr>
<tr>
<td>Prevalence</td>
<td>Threshold at which modeled prevalence is closest to observed prevalence</td>
</tr>
<tr>
<td>Sens</td>
<td>Fixed (specified) sensitivity</td>
</tr>
</tbody>
</table>

**Model Comparisons**

The AUC of MaxEnt models and spatial priors models were compared to determine which performed better. Additionally, the prediction map generated by the MaxEnt model was subtracted from the prediction map generated by the spatial priors model in order to determine where the two modeling techniques differed in their predictions.

In order to visualize these differences in a single map, the delta map generated by subtracting MaxEnt from spatial priors was log-transformed. Then, spatial priors predictions were separately subtracted from MaxEnt predictions, and this delta map was also log-transformed. The purpose of this was to visualize cases in which each model was
greater than the other, since logs of negative numbers are undefined. These delta maps were then combined and scaled such that positive delta values indicate that the spatial priors probability prediction was greater than the MaxEnt probability prediction while negative values indicate that the MaxEnt prediction was greater than the spatial priors prediction.

**Results**

The spatial priors models performed better than MaxEnt for all test species in terms of AUCs (Table 3). Both MaxEnt and spatial priors also performed better than preliminary GLMs (see appendices A1-A3 for all preliminary GLM results).

<table>
<thead>
<tr>
<th>Table 3. Area Under the Curve (AUC) values for models with the same predictors using MaxEnt without spatial priors and spatial priors methods.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species</strong></td>
</tr>
<tr>
<td>Argia vivida</td>
</tr>
<tr>
<td>Erythemis simplicicollis</td>
</tr>
<tr>
<td>Lestes rectangularis</td>
</tr>
<tr>
<td>Libellula vibrans</td>
</tr>
</tbody>
</table>

Coefficient estimates for each species were determined for the spatial priors models (Table 4). For *Argia vivida*, cultivated land cover and distance to water had large negative effects, whereas temperature and seasonality had a strong positive effect. For *Erythemis simplicicollis*, temperature was an important positive driver and distance to water and precipitation were both strong negative ones. For *Lestes rectangularis*, seasonality and distance to water had the strongest negative effects, whereas temperature had a strong positive effect. In *Libellula vibrans*, distance to water, precipitation, and seasonality all had strong negative effects and temperature had a strong positive effect. Flow accumulation, topographic wetness index, and land cover variables tended to be the
least important variables for all species, with the exception of cultivated land cover in *Argia vivida*.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Argia vivida</th>
<th>Erythemis simplicicollis</th>
<th>Lestes rectangularis</th>
<th>Libellula vibrans</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.719835467</td>
<td>5.394058364</td>
<td>4.90406216</td>
<td>3.650751601</td>
</tr>
<tr>
<td>bio 12</td>
<td>-0.227768375</td>
<td>-1.366490796</td>
<td>-0.486034827</td>
<td>-1.627590735</td>
</tr>
<tr>
<td>bio 15</td>
<td>0.985424715</td>
<td>-0.850480632</td>
<td>-0.791079565</td>
<td>-0.971990097</td>
</tr>
<tr>
<td>bio 1</td>
<td>1.897974277</td>
<td>2.007185306</td>
<td>1.051944479</td>
<td>1.917675254</td>
</tr>
<tr>
<td>cti</td>
<td>-0.196699439</td>
<td>0.316278148</td>
<td>0.07710956</td>
<td>0.521870309</td>
</tr>
<tr>
<td>fa</td>
<td>0.183389496</td>
<td>-0.046646125</td>
<td>0.022552808</td>
<td>-0.520483719</td>
</tr>
<tr>
<td>lc_class4</td>
<td>0.032982277</td>
<td>-0.139619314</td>
<td>-0.510247587</td>
<td>0.050140474</td>
</tr>
<tr>
<td>lc_class7</td>
<td>-0.922406972</td>
<td>-0.421923232</td>
<td>-0.269289012</td>
<td>-0.482198979</td>
</tr>
<tr>
<td>proportionland</td>
<td>0.425740793</td>
<td>0.195256971</td>
<td>0.679360178</td>
<td>0.361675919</td>
</tr>
<tr>
<td>waterdistance</td>
<td>-0.73548929</td>
<td>-3.388699457</td>
<td>-2.565479775</td>
<td>-3.295014004</td>
</tr>
</tbody>
</table>

Range predictions generated by each model are presented in Figures 1-4. In *Argia vivida*, the MaxEnt model (Figure 1A) predicted a large hotspot on the southwest coast of North America, where no observations have been recorded and which lies outside of the species’ expert range map. The model that used the expert range map as a prior (Figure 1B) predicted a much lower occurrence rate in this area. A similar pattern was observed in predictions for *Lestes rectangularis* (Figure 3). In this species, the MaxEnt model predicted similar occurrence rates across northern North America, with high values in locations without records. The spatial priors model, however, differentiated patches with high numbers of presences, predicting presence at a higher rate in the central-northern areas and lower values in the northwest and northeast sections (Figure 3). For *Erythemis simplicicollis* and *Libellula vibrans*, the two dragonfly species, there was not one isolated area outside the expert range maps where the MaxEnt model erroneously predicted high presence (Figures 2 and 4). However, the MaxEnt models predicted generally high
probability of presence outside of the expert range map, where few observed records occur, which the spatial priors model did not.

For all species, the areas predicted by the spatial priors model as having the highest probability of presence—demonstrated by areas that exceeded several thresholds—fell nearly entirely within the expert range maps (Figures 1C, 2C, 3C, 4C).

The delta maps showing the differences between the spatial priors predictions and MaxEnt predictions generally revealed more pronounced differences within the species range maps than outside of them (Figures 1D, 2D, 3D, 4D). MaxEnt models tended to predict higher probabilities of presence outside of the expert range maps than spatial priors models did, while spatial priors models tended to predict higher probabilities of presence inside the expert range maps than MaxEnt models did.
Figure 1. Species distribution prediction maps for *Argia vivida*. Points represent observed presences. Lines represent expert range maps. 

A. Prediction from MaxEnt with no priors. Color scale represents a relative probability of occurrence for the given extent. 

B. Prediction from spatial priors models. Color scale represents a relative probability of occurrence for the given extent. 

C. Thresholded map of the Spatial Priors model showing how many out of seven thresholds were met. Colors represent the sum of how many thresholds were met in each grid cell; higher values represent a higher probability of presence. 

D. Delta map of the log-transformed differences between the spatial priors prediction and MaxEnt prediction. Positive values indicate spatial priors prediction > MaxEnt prediction and negative values indicate MaxEnt prediction < spatial priors prediction.
Figure 2. Species distribution prediction maps for *Erythemis simplicicollis*. Points represent observed presences. Lines represent expert range maps. 

A. Prediction from MaxEnt with no priors. Color scale represents a relative probability of occurrence for the given extent. 

B. Prediction from spatial priors models. Color scale represents a relative probability of occurrence for the given extent. 

C. Thresholded map of the Spatial Priors model showing how many out of seven thresholds were met. Colors represent the sum of how many thresholds were met in each grid cell; higher values represent a higher probability of presence. 

D. Delta map of the log-transformed differences between the spatial priors prediction and MaxEnt prediction. Positive values indicate spatial priors prediction > MaxEnt prediction and negative values indicate MaxEnt prediction < spatial priors prediction.
**Figure 3.** Species distribution prediction maps for *Lestes rectangularis*. Points represent observed presences. Lines represent expert range maps. A. Prediction from MaxEnt with no priors. Color scale represents a relative probability of occurrence for the given extent. B. Prediction from spatial priors models. Color scale represents a relative probability of occurrence for the given extent. C. Thresholded map of the Spatial Priors model showing how many out of seven various thresholds were met. Colors represent the sum of how many thresholds were met in each grid cell; higher values represent a higher probability of presence. D. Delta map of the log-transformed differences between the spatial priors prediction and MaxEnt prediction. Positive values indicate spatial priors prediction > MaxEnt prediction and negative values indicate MaxEnt prediction < spatial priors prediction.
Figure 4. Species distribution prediction maps for *Libellula vibrans*. Points represent observed presences. Lines represent expert range maps. 

A. Prediction from MaxEnt with no priors. Color scale represents a relative probability of occurrence for the given extent. 

B. Prediction from spatial priors models. Color scale represents a relative probability of occurrence for the given extent. 

C. Thresholded map of the Spatial Priors model showing how many out of seven various thresholds were met. Colors represent the sum of how many thresholds were met in each grid cell; higher values represent a higher probability of presence. 

D. Delta map of the log-transformed differences between the spatial priors prediction and MaxEnt prediction. Positive values indicate spatial priors prediction > MaxEnt prediction and negative values indicate MaxEnt prediction < spatial priors prediction.
Discussion

Model Performance

Models that used expert range maps as spatial priors performed better than simpler models for every test species. These models take into account important prior knowledge about species ranges and may correct inaccuracies caused by factors not explained by the climatic variables used in simpler models (Domisch et al. 2015). Expert range maps overestimate species’ geographic distribution on their own (Hurlbert and Jetz 2007) but should not be neglected in SDMs as they can provide informative priors (Domisch et al. 2015). The MaxEnt models used in this study predict probabilities in environmental space, not geographic space: they predict habitat suitability of each grid cell, but do not necessarily indicate presence because they do not account for factors like dispersal barriers or biotic interactions, both of which constrain species ranges (Elith and Leathwick 2009). This is particularly clear in the GLM (Figure A3) and MaxEnt (Figure 1A) maps for *Argia vivida*. While the models without spatial priors predict high presence in one spot along Mexico’s west coast, there are no occurrence records here. The Sonoran Desert, located between the more northern range of this species and the southern area of suitable habitat, may act as a dispersal barrier, constraining the range to the northern habitats. Because the dispersal abilities of *Argia vivida* are not explicitly included in the model, using expert range maps can help to account for this dispersal limitation. This is seen as well in the other test species, where MaxEnt predicts high occurrence in areas that are geographically distant from any occurrence records.

The threshold sum values for spatial priors models indicate that the highest probability of species occurrence is generally limited to a subset of the expert range map.
In *Argia vivida*, for example, the highest threshold sums occurred along the west coast, though the expert range map covered most of the western United States. Similarly, in both *Lestes rectangularis* and *Libellula vibrans* the highest sums were concentrated in the southern portions of the expert range maps, while the northern portions of the range map had identical threshold sums to large swaths of space outside of the range map. This result emphasizes the overall importance of species distribution modeling in determining fine-scale species ranges: while expert range maps may delineate the outer limits of a species’ range, species preferentially occupy only subsets of those ranges. Detailed knowledge of species ranges is therefore best captured by combining coarse expert range maps with finer scale environmental data (Hurlbert and Jetz 2007).

**Ecological Predictions**

The importance of temperature in predicting all species ranges in this study is expected for ectothermic species like odonates, which rely on external temperatures for thermoregulation and development processes (Collins and McIntyre 2015). Precipitation seasonality was also among the most important predictors, which is in line with the finding of Collins et al. (2014) that seasonality affects hydroperiod of lentic waters, which is an important predictor for many freshwater species. Land cover variables were generally less important than other variables. Odonate species may be less affected by human-dominated landscapes than vertebrates or other species because they can use water sources even in human-dominated areas and feed primarily on other insects, which are abundant in virtually all landscapes.

Distance to water had a large negative coefficient in the spatial priors models for all species, which is expected given that odonates rely on water for reproduction
(Kalkman et al. 2008). However, distance to water had the least negative effect on Argia vivida (coefficient = -0.73548929) compared to other species, which may be explained by the fact that Argia vivida is a generalist species, which can breed in both lotic and lentic waters (Hof et al. 2006). Since this species can take advantage of a wider range of water sources, it may not be as limited by proximity to large bodies of water or certain types of bodies of water as other species are. Although lentic species are expected to have greater dispersal abilities than lotic species (Hof et al. 2006), distance to water had a similar effect on the lentic species presented here (Lestes rectangularis and Libellula vibrans) as the lotic species (Erythemis simplicicollis).

For all test species, the two new remote sensing imagery hydrological variables (proportion freshwater and distance to freshwater) were more important predictors of species occurrence than either of the elevation-derived hydrological variables (flow accumulation and topographic wetness index). This supports the idea that layers based on image-derived detection of water may map freshwater more accurately than digital elevation models, and can therefore be more useful for species distribution modeling in semi-aquatic species like odonates. As a next step in exploring this idea, it may be useful to model larval odonates. This study was limited to adults, as species-level data for adults is much more widely available; however, larvae are likely to be more influenced by hydrological variables and should be modeled separately from adults (Patten et al. 2015). Distance to freshwater and proportion of freshwater are likely to predict adult distributions well, because these variables relate to adult dispersal ability. In contrast, variables like flow accumulation and wetness index describe qualities of the water rather
than the availability of the water, and are therefore more likely to be important for aquatic larvae than for terrestrial adults.

**Future Applications**

The results presented here generally align with expectations regarding which environmental variables are important for odonate species. However, a more nuanced understanding of the differences between taxonomic and ecological groups could be gained by applying the methods presented here on a larger scale. With a greater sample of species, comparisons between lentic and lotic species, for example, could be more clearly elucidated. Modeling greater numbers of species using these methods could also help identify patterns of diversity and inform conservation planning on a continent-wide scale. In Africa, for example, continent-wide modeling of odonates has led to the identification of several biodiversity hotspots and major reserve gaps (Simaika et al. 2013, Clausnitzer et al. 2012). Similar analyses for North America would be informative, especially given that there has been relatively little modeling of North American odonates (Collins and McIntyre 2015).

The methods presented here provide a basis for increased and improved modeling of odonate species distributions. Modeling odonates and other invertebrate and freshwater species has only begun relatively recently, but is crucial to understanding and conserving the species that make up the vast majority of the world’s biodiversity. As modeling methods become more accurate and nuanced, their power as ecological tools will only grow.
Acknowledgements

I wish to especially thank Sami Domisch, who served as my mentor for this research and provided assistance with coding, guidance, and support for the entire project. This research was performed in the lab of Walter Jetz, whom I also wish to thank for his guidance and feedback throughout the project. Thanks to all members of the Jetz lab for their feedback on modeling methods and data sources. This project was supported in part by the facilities of the Yale University Faculty of Arts and Sciences High Performance Computing Center. Funding for a summer research project leading to this thesis was provided by the Yale College Dean’s Research Fellowship and the Trumbull Richter Fellowship.

Literature Cited


### Appendix

**A1.** Beta coefficients of the preliminary GLM models for all focal species.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Argia vivida</th>
<th>Erythemis simplicicollis</th>
<th>Lestes rectangularis</th>
<th>Libellula vibrans</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.750236807</td>
<td>-4.050159974</td>
<td>-4.085673432</td>
<td>-7.752062296</td>
</tr>
<tr>
<td>bio 1</td>
<td>1.507692571</td>
<td>2.533475777</td>
<td>0.8558958</td>
<td>3.411623432</td>
</tr>
<tr>
<td>bio 12</td>
<td>-0.310899443</td>
<td>-0.615508704</td>
<td>-0.605411484</td>
<td>-1.193138001</td>
</tr>
<tr>
<td>bio 15</td>
<td>0.792513475</td>
<td>-1.013614016</td>
<td>-1.065754297</td>
<td>-1.931059506</td>
</tr>
<tr>
<td>lc_class4</td>
<td>0.074284185</td>
<td>-0.24746213</td>
<td>-0.452678533</td>
<td>-0.008370234</td>
</tr>
<tr>
<td>lc_class7</td>
<td>-1.017805635</td>
<td>-0.208071741</td>
<td>-0.036442534</td>
<td>-0.37267116</td>
</tr>
<tr>
<td>proportionland</td>
<td>0.6306383</td>
<td>0.319872226</td>
<td>0.537270933</td>
<td>0.372033132</td>
</tr>
<tr>
<td>waterdistance</td>
<td>-0.833346071</td>
<td>-2.825002802</td>
<td>-2.323771539</td>
<td>-3.643959586</td>
</tr>
</tbody>
</table>

**A2.** Area under the curve (AUC) values for preliminary GLMs.

<table>
<thead>
<tr>
<th>Species</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argia vivida</td>
<td>0.857</td>
</tr>
<tr>
<td>Erythemis simplicicollis</td>
<td>0.887</td>
</tr>
<tr>
<td>Lestes rectangularis</td>
<td>0.782</td>
</tr>
<tr>
<td>Libellula vibrans</td>
<td>0.893</td>
</tr>
</tbody>
</table>
A3. Prediction maps generated by preliminary GLM outputs for *Argia vivida* (A), *Erythemis simplicicollis* (B), *Lestes rectangularis* (C), and *Libellula vibrans* (D). Color indicates the probability of presence at a given site.