*Incorporating migratory data into species distribution models improves conservation of American Woodcock habitat*

Target: Journal of Applied Ecology

*Abstract*

1. ~~Recent studies (esp 3 billion birds) have made it clear that we need to be expanding migratory bird conservation in the United States. Bird declines are due not only to the loss of breeding and winter habitat, but also migratory stopover habitat, which often has different characteristics.~~
2. ~~Our most frequent tool for prioritizing conservation areas (species distribution models) is typically only used to model habitat for a single migratory stage at a time. However, decisionmakers will usually only develop a single conservation prioritization plan for a species, and so they need a way to combine migratory and breeding season habitat use into a single prioritization framework.~~
3. ~~We use a joint residential/migratory species distribution model developed for managing American Woodcock in Pennsylvania to demonstrate how decision makers can incorporate migratory and breeding season habitat models into a single decision-making framework, allowing local land managers to decide whether to prioritize migratory or breeding season habitat management based on an area’s suitability.~~

*Introduction*

Recent studies have reinforced the concept that bird population stability is dependent, not just on breeding season habitat, but also habitat throughout the winter and migratory seasons (cite full annual cycle tk). Bird habitat use during these seasons can affect bird populations through either carryover effects (cite tk) or directly reduced survival. Therefore, preserving these habitats has become a conservation priority (cite tk). Recent efforts have included a focus on migratory stopover habitat, which is habitat used by migratory birds to rest and refuel between migratory flights (cite tk). Stopover habitat can have differing quality, and differing value to migratory birds. Mehlman et al. (2005tk) defined a stopover site as any place where a bird can land and survive until the next migratory flight. However, those sites which have more resources can either improve or be less detrimental to a migratory bird’s condition and can increase the bird’s probability of successfully completing migration. As migratory survival is believed to be a limiting factor for many species of birds (cite tk), conserving stopover habitat has become a popular topic in ornithological literature (cite tk).

As research on stopover habitat improves, there are increasing opportunities to incorporate stopover habitat management into conservation efforts. Here, we explore the best methodologies for incorporating stopover habitat into migratory bird habitat management, with a focus on American Woodcock (*Scolopax minor*) in the state of Pennsylvania. American Woodcock are a short distance migrant, most of whose migratory stopover habitat overlaps with their breeding and wintering range (eBird range figure tk). Pennsylvania provides both breeding habitat and stopover habitat for birds migrating to the northern extent of their range (migratory connectivity work tk). However, a past study suggests that woodcock stopover habitat does not always completely overlap with habitat used during other seasons (Allen et al. tk). Therefore, management for breeding season habitat alone may not be enough to protect woodcock stopover habitat.

To determine whether additional management for stopover habitat is necessary, we first model migratory and breeding season woodcock habitat in Pennsylvania and examine the overlap between the two. If there is significant overlap between migratory and breeding season habitat, then existing management strategies for woodcock in Pennsylvania (which have been based on breeding season data) may be adequate to protect migratory stopover habitat. If not, then additional management may need to take place to additionally protect stopover habitat. We then demonstrate one possible method for incorporating stopover habitat into a habitat management framework by combining migratory and breeding season habitat suitability models into a single habitat prioritization tool. This tool allows users to manually decide their priorities for migratory and breeding season management and identify areas that provide woodcock habitat during both seasons, allowing users to incorporate full annual cycle conservation into a straightforward conservation prioritization framework.

**Methods**

*Breeding season species distribution model*

Breeding season species distribution models are the most frequently used tool for planning bird conservation (cite past work tk). As such, they can be created using most traditional survey datasets. For the Pennsylvania example, we used federal Woodcock Singing Ground Surveys and similar state-level surveys conducted by the Pennsylvania Game Commission. SGS surveys consist of tk mile long survey routes consisting of tk evenly-spaced points, with presence-absence determined at each point based on whether male displays were visible during a tk minute interval at dusk. SGS survey routes were randomly distributed through the state? In 19tk? And each route has been run annually since tk. Pennsylvania Game Commission woodcock surveys were run using the same methodology, but their routes were intentionally placed near state gamelands or in areas where managers believed woodcock occupancy was likely. We converted state and federal survey data from 20tk-20tk to a presence-absence dataset by marking each survey point as present if there were woodcock observed at that point at least once during the 5 year interval, and absent if they were not. These presence-absence locations were then used as the response variable in the breeding season species distribution model. The explanatory variables in the species distribution model included several suites of variables presumed to be relevant to woodcock habitat. These included variables representing land use/land cover (NLCD 20tk), forest successional class (LANDFIRE 20tk), elevation (source tk?), slope (source tk?), ecoregions (EPA level 3 tk), soil drainage (source?), and topographic wetness index (cite methodology tk). We additionally added landscape metrics from the landscapemetrics R package (cite tk) representing landscape composition (% forest, % agricultural, % developed) and configuration (aggregation index, cohesion, edge density). We ran each of these landscape metrics at multiple scales, represented by radii from 90m pixels. The radii used were 500m, 1km, 5km, and 10km. To generate these landscape metrics, we cropped a binary forest/non-forest layer to the extent of a circle of the given radius from each 90m pixel, and then ran the appropriate function from the landscapemetrics package on each cropped raster. We then assigned the output value from the function to the appropriate 90m pixel. These were then used as explanatory variables in the species distribution model. The species distribution model used a random forest classifier designed for clustered data (MixRF package tk) to predict whether woodcock would be present or absent at survey points. The survey route id was used as the clustering variable to compensate for autocorrelation between points on the same survey route. We also included federal/state as an explanatory variable in the analysis to account for bias in the state survey route allocation. To avoid overwhelming the model with highly correlated variables, we elected to use a backwards variable-selection approach (VSURF package tk) to determine which set of variables should be used in the final model. This approach uses a three-step process, with each step producing a more parsimonious model. The first step eliminates variables that have little importance to prediction, the second step removes variables that have some relevance but are not critical for prediction, and third step eliminates variables that are redundant. We calculated the AUCs for the models produced by each of the three steps and used the set of variables produced by the step with the highest AUC in the final model. We used a k-fold cross validation approach with 10 folds to evaluate our final model, using 90% of the data in each fold as a training dataset and the remaining 10% as a testing dataset. We averaged together the AUCs calculated for each of the 10 folds to find the AUC for the final model. We then calculated a predictive layer using the models calculated for each of the 10 folds and averaged those layers together to create a final predictive layer for the breeding season model. For the predictive layer, the survey type predictive variable was set to “federal” to exclude bias resulting from the placement of state survey locations near state gamelands.

*Migratory season species distribution model*

We used GPS data from the EWMRC to designate woodcock migratory stopover sites throughout the state of Pennsylvania. Sentence tk. These sites were used as present locations in the migratory season species distribution model. To create pseudo-absence locations, we randomly allocated 10,000?tk locations throughout the state. These two sets of data were combined to form the response variable in the model. The explanatory variables were the same as used in the breeding season model. We then used these data to create a species distribution model using a random forest classifier (SDMTune package tk). We used a k-fold cross validation approach with 10 folds to evaluate our final model, using 90% of the data in each fold as a training dataset and the remaining 10% as a testing dataset. We averaged together the AUCs calculated for each of the 10 folds to find the AUC for the final model. We then calculated a predictive layer using the models calculated for each of the 10 folds and averaged those layers together to create a final predictive layer for the migratory season model.

*Multi-season predictive layer*

To facilitate user choice in how and where to prioritize migratory and residential habitat, we generated a series of combined layers with different weights assigned to the migratory and breeding season layers. 0% migratory 100% breeding, 10% migratory 90% breeding, etc. We used an interactive application (e.g. shinyapps) to distribute this information and facilitate user weighting.

* + *Methods: We blended the model using a Shiny application (screenshot as a figure), which allowed the user to determine in 10% intervals how much they wanted to value residential vs. migratory habitat*
  + *Methods: In addition to the blending, the Shiny app enabled users (presumably managers) to examine how the layers overlayed with their gamelands, and the percent of their gamelands which fell into high (upper 33rd percentile of all pixels) or medium (66th to 33rd percentile of all cells) quality categories*

*Results*

The most informative residential model was the “tk” step model, for which all autocorrelated variables had been removed. This produced a model with an AUC of 0.tk, which was heavily informed by landscape variables at the 5 and 10 kilometer scales (Table tk). No variables at the finest landscape scale (500m) or in the suite of moisture variables were included in the most informative model. While random forest models do not provide coefficients that can be used to determine the impact of each covariate on the model, graphs of habitat suitability for each covariate show strong, non-linear relationships with several of the most informative variables. Suitability was highest for landscapes at the 10km scale with 0 – 25% developed land area, 0 – 50% agricultural land area, and aggregation index values of 80 – 100. At the 5km scale, the residential model also showed high suitability for landscapes with 30 – 100% forest cover (Fig. tk).

The most informative migratory model was the full model, including all landscape, land cover, geographic, and moisture covariates (Table tk). This produced a model with an AUC of 0.tk. Likely due to the wide array of covariates influencing the model, individual covariate graphs do not show clear visual patterns between migratory habitat suitability and any one covariate. However, the migratory model showed greater tolerance for developed and dis-aggregated landscapes at a 10km scale than the residential model (Fig. tk). The two models are also distinguished by the scale at which covariates influence habitat suitability. While the most informative residential model was not influenced by any landscape covariates at the 500m scale, and only 1 landscape covariate at the 1km scale, the most informative migratory model included all available small-scale landscape covariates. This caused the migratory model to provide predictions at a finer spatial scale than the residential species distribution map (Fig. tk).

Despite the difference in the resolution of predictions, there were some regional differences between the two layers (Fig tk. Ecoregion bar chart and ecoregion map, with asterisks for p-values). Residential suitability was significantly lower than migratory suitability for woodcock in Ecoregions 1, 6, and 11 (add t-test tk). Conversely, suitability was significantly higher in the residential season than the migratory season in Ecoregion 2. In other Pennsylvania ecoregions (tk), migratory and residential suitability was not significantly different.

|  |  |  |
| --- | --- | --- |
| Suite | Migratory | Residential |
| Landscape (500m) | Aggregation Index, Cohesion, Edge Density, % Forest, % Agricultural, % Developed |  |
| Landscape (1km) | Aggregation Index, Cohesion, Edge Density, % Forest, % Agricultural, % Developed | % Agricultural |
| Landscape (5km) | Aggregation Index, Cohesion, Edge Density, % Forest, % Agricultural, % Developed | Cohesion, % Forest, % Agricultural, % Developed |
| Landscape (10km) | Aggregation Index, Cohesion, Edge Density, % Forest, % Agricultural, % Developed | Aggregation Index, Cohesion, % Agricultural, % Developed |
| Land Cover | Forest, Successional Class |  |
| Geography | Elevation, Slope, Ecoregions | Elevation, Ecoregions |
| Moisture | Drainage, Topographic Wetness Index |  |

Table tk. Variables selected via backwards variable selection in VSURF for the migratory and residential models. The migratory model employs the full set of variables, while the residential model uses a subset of variables inclined towards coarse resolution landscape variables.

![Chart, scatter chart

Description automatically generated]()

Figure tk. Comparison of relationships between landscape variables and habitat suitability for migratory and residential models. During the residential season, woodcock habitat suitability is highest in highly aggregated landscapes with ~75% forest and ~25% agricultural cover. During the migratory season, however, woodcock become far more tolerant of landscapes that are unsuitable during the residential season, including landscapes with higher proportions of developed cover.

Map

Description automatically generated

Figure tk. The lack of correlation between residential and migratory habitat suitability (measured on a percentile scale) indicates that single-season species distribution models will likely not reflect the suitability of a landscape for woodcock during other portions of the year.

*Discussion*

These models produced migratory and residential layers that often predicted very different habitat use between the two seasons. The resolution of these predictions differed between the migratory and residential seasons, likely due to seasonal changes in the scale of selection. Someone (tk) has noticed changes in the scale of selection between the breeding and migratory seasons, which would cause an effect similar to the one we observed. We also noted that woodcock were far more likely to use developed landscapes during the migratory season than the residential season, and heavily utilized some ecoregions (tk) which are urbanized and has low residential habitat suitability. This fits with other research (tk) which suggests that many specialist bird species become generalists during migration and are more tolerant of certain habitat factors which would otherwise repel them during other seasons. Overall, our results show that woodcock use regions with a mix of forest and agricultural cover during the residential season (corroborating past research tk), but woodcock during the migratory season often choose sites based on a variety of local habitat characteristics.

The fact that woodcock habitat use changes drastically between seasons indicates that a “breeding only” strategy will not be enough to preserve woodcock stopover habitat. However, this does range the question of how best to conserve these habitats. To facilitate the incorporation of both migratory and residential habitat suitability into a single management framework, we used a Shiny tool to combine the two layers and left it up to the user to decide how these layers should be weighted. Leaving this decision to the user serves two purposes. First, it encourages discussion within the management agency regarding the agency’s priorities in conserving residential and stopover habitat, to meet both the objectives of their stakeholders and to achieve a stable woodcock population. Second, by allowing the user to make the weighting decision, we allow users to determine whether weighting should change based on regions. Some regions may be valuable as both stopover and residential habitat; in such a case, a balanced user weighting of the two layers may be a good way to determine which gamelands provide both types of habitat. Other regions (such as ecoregion tk) might provide only migratory habitat, and so a user weighting that favors migratory habitat might be best employed to determine where woodcock management would be best applied within that region. In addition, a map of migratory corridors for woodcock could be used to determine where woodcock are most likely to fly over the state and be used to determine the regions in which migratory habitat management would be most important (cite some component of my work tk).

As in any habitat analysis, we should be cautious about the spatial scale at which these results are applicable. For example, we found here that there was limited overlap between migratory and residential habitat at a landscape scale. However, Allen et al. (20tk) found considerably more overlap between migratory and residential habitat at a local scale. In this case, we may be able to identify areas of overlap between residential and migratory woodcock habitat at a landscape scale and assume that habitat managed for woodcock in these areas will support both breeding and migratory birds. However, this will not be the case for all migratory bird species, and multi-scalar studies into how selection differs between migratory and residential seasons will be necessary before appropriate stopover management can be implemented.

Agencies should also be prepared for the possibility that landscape-scale habitat suitability models for migratory habitat may show that some migratory stopover sites are focused in areas that aren’t traditionally managed for wildlife habitat. Our woodcock migratory stopover habitat model, for example, shows that several urban areas (including the greater Philadelphia area) are hotspots for migratory stopover activity. However, there may still be conservation opportunities in these areas through collaboration with other entities that maintain urban greenspace, such as parks departments and private landowners (cite tk). Partnerships between these organizations may provide the opportunity for wildlife management agencies to extend their reach into areas that are not traditionally managed for wildlife but may be important migratory stopover hotspots.

We hope that this first step into incorporating stopover habitat management into woodcock management plans for Pennsylvania provides an example of the type of methods that can be used for incorporating stopover habitat management into migratory bird conservation. This type of approach should be especially useful for species that are presumed to be limited by migratory survival.