

# Summary report: Rusty Blackbird habitat suitability on wintering grounds and during spring migration

*Brian S. Evans*

*October 1, 2016*

## Overview of methods

Assessment of habitat suitability for winter and spring blitz periods required several processing steps, including: processing bird observation data, model development and calibration, processing of land cover and climatic data, model evaluation, and statistical interpretation of models. Because habitat suitability predictions and their interpretation are highly dependent on each of these steps, below I provide a brief overview of each of these steps.

- **Processing bird observation data:** Observations from the Spring and Winter Blitz as well as from the larger eBird database were used throughout this analysis. To ensure that collection of eBird and Blitz observations used similar protocols, eBird data were subset such that only travelling counts were considered for this analysis. Because the spatial extent of a study area can bias suitability models and large extents can lead to exaggerated measures model performance, observations were subset to those within the 99th percentile spatial distribution of Rusty Blackbird Observations. Observation data were grouped into sampling periods, with Winter Blitz observations grouped by year and Spring Blitz observations grouped into two-week intervals for each year. To maximize independence of samples (i.e., to avoid double-counting individuals) observations were aggregated to a 4 km resolution grid (see below), with grid cells representing the maximum count of Rusty Blackbirds during each sampling period. Finally, because the collection of eBird data is biased towards certain environments (e.g., warmer temperatures and close to human development), we evaluated habitat suitability of presence locations relative to locations in which eBird lists did not contain Rusty Blackbird observations.
- **Processing land cover and climatic data:** Land cover data were obtained from the USGS National Gap Analysis Program (<http://gapanalysis.usgs.gov>). These data are provided at a resolution of 30 m with classification of land cover in natural areas dependent on the plant community present at a given site. Land cover data were re-classified to habitat types expected to be predictive of Rusty Blackbird distribution. For each land cover type, we then aggregated the proportional cover to a resolution of 4 km, a spatial resolution thought to be representative of the spatial habitat use by wintering Rusty Blackbirds. Climatic data were obtained from the PRISM Climate Group (<http://prism.oregonstate.edu/>) and included average daily precipitation and minimum daily temperatures at a resolution of 4 km. Precipitation for both winter and spring migration analyses was evaluated as the mean daily precipitation across each sampling period (i.e., 1 Jan - 14 Feb 2009-2011; X Mar - X Apr 2014-2016). For the winter analysis, we calculated the average minimum temperature across the period for each year in the study, whereas for the spring analysis we determined the minimum temperature associated with each bird observation. After processing, land cover and climatic data were extracted to bird observation locations.
- **Model development and calibration:** Model development involves fitting the observed distribution of Rusty Blackbirds and observations (eBird lists) that do not include Rusty Blackbirds to the distribution of environmental covariates in a maximum entropy (MaxEnt) modeling framework. Importantly, because these models are developed in environmental rather than geographic space, these models do not predict the probability that a Rusty Blackbird will be present at a given location but rather that the environment at a given location is suitable for Rusty Blackbirds as a function of the distribution of Rusty Blackbird observations. Models were developed separately for Rusty Blackbird observations that fell into one of three flock size classes, including: small (1-19 individuals per observation), medium (20

- 99 individuals per observation), and large flocks (100 or more individuals per observation). For all variables with the exception of minimum temperature, only linear models were used to describe Rusty Blackbird habitat suitability. A quadratic term for minimum temperature was used to address the expected hump-shaped response of Rusties to minimum temperatures. Models were calibrated to avoid overfitting by excluding variables that are highly correlated and choosing models with the greatest explanatory power of the models relative to the number of variables used in their construction.

- **Interpretation of models:** Our goals in model interpretation were to determine: 1) How do Rusty Blackbird suitability models perform and did the inclusion of Blitz data improve model performance (formally tested for the Winter Blitz)? 2) Where is Rusty Blackbird habitat most suitable?; 3) Which environmental variables best predict Rusty Blackbird distributions?; and 4) Do large flocks of Rusty Blackbirds occupy a different environmental niche than small and medium size flocks (formally tested for the Winter Blitz)? Model performance is evaluated by comparing model sensitivity (i.e., the true positive rate – the proportion of correctly identified sampled at a given threshold of habitat suitability) and specificity (the false positive rate – in species distribution modeling this is described the proportional predicted area for the model at a given threshold of habitat suitability). To address the suitability of habitat, we used the relationship between Rusty Blackbird distributions in environmental space to predict suitability models for small, medium, and large flocks of Rusty Blackbirds in environmental space. We then evaluated the contribution of the environmental covariates to model performance to estimate which variable (habitat and climatic) best explains observed Rusty Blackbird distributions. Finally, to address whether different flock size classes occupy a different environmental niche, we tested whether there are differences in relation to the environment in which each flock size class are observed and whether large flocks occupy a narrower niche space than small and medium flocks (i.e., are more constrained in their distribution about the environmental covariates). Further details for each step of this analysis are provided in the results below.

## Summary of results

### 1. Model evaluation

To evaluate models, we assessed the area under the receiver operator curve (ROC, AUC, see Fig. 1A), a measure of model performance. While the details of AUC are outside of the goals of this document, AUC describes the sensitivity and specificity of models, with values of 0.5 representing equivalent model performance relative to random, 0.7 - 0.8 representing “fair” performance, 0.8 - 0.9 “good” performance, and 0.9 - 1.0 “excellent” model performance. AUC was calculated by separating samples into 5 “folds” of training and test samples and examining how well training samples predict test samples (cross-validation with k-fold partitioning). We found strong evidence that model performance increased with increased flock size, with fair performance for small flocks and good performance for medium and large flocks (across observation methods, Fig. 1B). Because the number of samples used to develop models decreased with flock size, this suggests that larger flocks are observed across a narrower range of the environmental covariates (tested explicitly in Number 4, below). We tested whether the predictive capacity of models varied by flock size by comparing training AUC across folds for a given flock size class against a null distribution developed by randomizing two flock size classes (e.g., large vs. small flocks). Similarly, to determine whether Blitz data improved estimates evaluate the performance of models relative to observation methods (eBird or Blitz, Winter Blitz only), we compared observed AUC against a null distribution developed by randomizing eBird and Blitz samples. Our findings show that models developed using eBird and Blitz data, combined, performed better than either eBird and Blitz alone (Fig. 1B). There was not evidence for a difference in predictive capacity between Blitz and eBird samples, however there was strong evidence that the inclusion of Blitz data (i.e., in combined Blitz and eBird observation methods) greatly improved model performance for large and medium flock size classes.

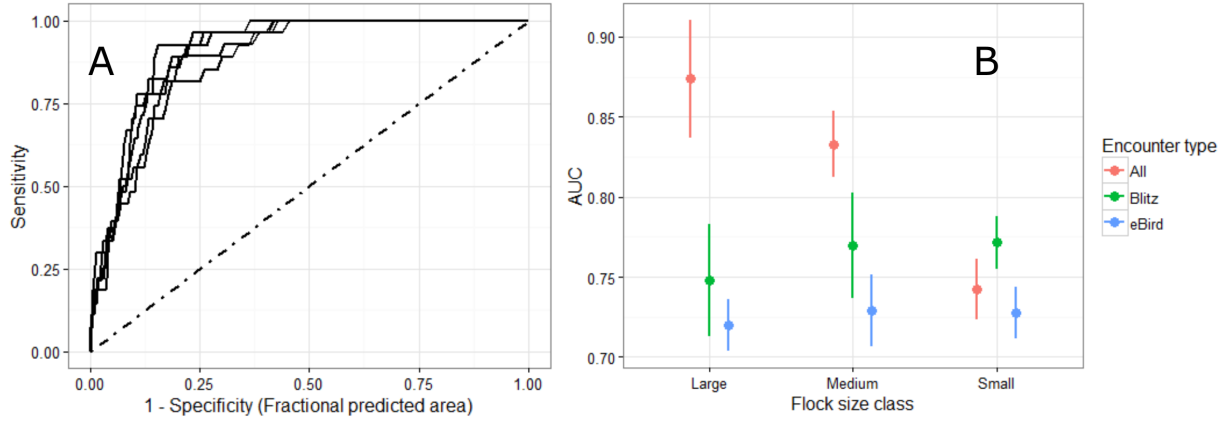


Figure 1: A: Example receiver operator curves (ROC) for cross-validation runs of large flocks during the Winter Blitz period, 2009-2011. The dotted line represents the null model of random model performance. B: Area under the receiver operator curve (AUC) across flock size and observation classes for the Winter Blitz period, 2009-2011. Error bars represent the 95% confidence interval about mean AUC values across cross-validation runs.

## 2. Habitat suitability for small, medium, and large flocks of Rusty Blackbirds

Winter suitability maps generated for each flock size class (Fig. 2) show potential hotspots in the southern Mississippi River Basin, the Blackbelt region of Alabama and Mississippi, and the coastal plains of the southeastern states. Note that the predicted suitability appears to decrease with increasing flock size (formally tested in Number 4, below). Surprisingly, the Piedmont region of the southeastern states appears to be of moderate suitability for large flocks, however, low sample sizes for this flock size class mean that some samples (e.g., an observation in Piedmont forest) may have a strong effect on habitat suitability estimates.

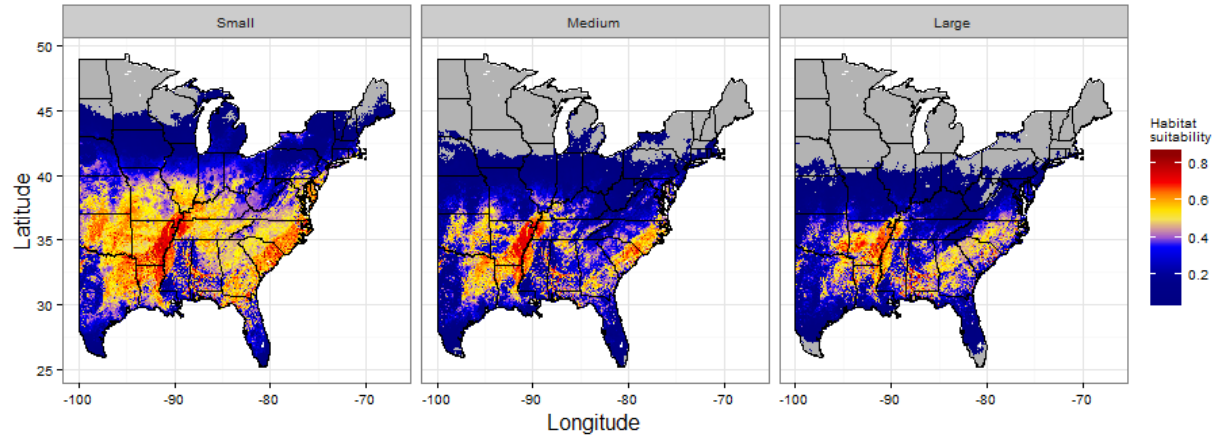


Figure 2: Winter prediction maps for small medium, and large flock size classes across observation methods.

## 3. The contribution of environmental variables to habitat suitability

To determine which environmental variables best predict Rusty Blackbird distributions for each flock size class, we compared the contribution of each variable to model performance after calibrating models (see above). To assess differences in Rusty Blackbird's environmental distribution across flock size classes, we

compared prevalence (i.e., the average predicted suitability for each map pixel), predicted area