# Prediction accuracy not improved for most species using spatially explicit models at local scales: a case study with Oklahoma grassland birds

Claire M. Curry

Jeremy D. Ross

Andrea J. Contina

Eli S. Bridge

Oklahoma Biological Survey, University of Oklahoma, Norman, OK

Corresponding author: [curryclairem@gmail.com](mailto:curryclairem@gmail.com), phone: 405-325-4034, fax 405-325-7702, 111 E. Chesapeake Street, Norman, OK 73019

Running title: Prediction accuracy of spatially explicit models

Word count total: must not exceed 7000 total.

Summary:

Main text:

Acknowledgements:

References:

Tables:

Figure legends:

Number of tables and figures:

Number of references:

# Summary

1. Species distribution models can be made more accurate by use of new “Spatiotemporal Exploratory Models”, a type of spatially explicit ensemble model developed at the continental scale that uses regional models and averages them by pixel to create more accurate predictions of species distributions. However, they are computationally expensive, so we examined how much computational effort it takes to get increases in accuracy that come from using such a model, and whether it improves accuracy at smaller-than-continental scales. We also develop a technique for mapping variable importance in these spatially explicit models using a mosaic of random forest variable rankings for variables of interest. 2. We used a combination of survey data and citizen science data to test a spatially explicit ensemble model at a smaller scale for its effects on accuracy of current models and variation in future models of climate change. We used grassland bird species in Oklahoma to test these models at the state level. 3. We found that spatially explicit ensemble models had longer runtimes without a concomitant increase in accuracy. Only a few species deviated from this overall pattern. 4. We found that climate change estimates of range loss or gain were not significantly different among scales. **Policy implications:** We therefore recommend use of standard species distribution models for local species prediction models and further study to understand at what point computationally expensive spatially explicit models become necessary for accurate modeling.

Key words: random forest, machine learning, spatiotemporal exploratory models, species distribution models

# Introduction

To help understand environmental and land use factors that drive declines, species distribution modeling (SDM) is a tool that uses environmental and geographic variables to predict what areas are suitable for a species and to better understand what factors constrain species’ ranges (Elith & Leathwick 2009). SDM can also be used to predict potential impacts of climate and land use change (Beaumont *et al.* 2007; Lipsey *et al.* 2015). Newer regression and machine learning techniques incorporated into SDM continue to increase prediction accuracy (Phillips, Dudík & Schapire 2004; Elith *et al.* 2006; Cutler *et al.* 2007; Elith, Leathwick & Hastie 2008; Lorena *et al.* 2011). To increase accuracy in the face of variation in regional drivers, a new method called Spatiotemporal Exploratory Modeling (STEM) uses smaller, overlapping subsets of data to generate regional predictions that are combined into an average (Fink *et al.* 2010). This averaging of overlapping smaller models allows the local models to correctly model patterns that may not occur in all parts of the study area, resulting in an overall map with more accurate predictions. The original STEM was used on continent-scale survey data and can be used with any base model (Fink et al. 2010, Fink et al. 2013). Additional uses of this approach also cover continent- to hemisphere-wide scales (Fink, Damoulas & Dave 2013). However, STEMs are computationally expensive because of predicting numerous supporting maps followed by averaging them to create the final model, instead of predicting just one map. Additionally, the relative increase in accuracy has not been compared to the relative expense of computational time nor have STEMs been tested at scales at which much species management occurs, such as state or regional initiatives (Brennan, Kuvlesky & Morrison 2005).

Grasslands are one of the world’s most endangered ecosystems, with declines of 82.6-99.9% of tallgrass prairie, 30-99.9% of mixed-grass prairie, and 20-85.8% of short-grass prairie in the plains states and provinces of North America (Samson & Knopf 1994), and as such could benefit from increased knowledge of distributional drivers. Drivers of decline include land use conversion via agriculture and changes in fire and grazing regimes (Samson, Knopf & Ostlie 2004), but specifics vary by region (Askins *et al.* 2007). The already tenuous status of grassland birds is further threatened by conversion to new crops resulting in permanent land use changes (Wright & Wimberly 2013), generational changes in land use (Higgins, Naugle & Forman 2002), changes in conservation programs for grassland habitats (Klute, Robel & Kemp 1997), and alterations to vegetation (Alward 1999) and ecosystem structure (Brown, Valone & Curtin 1997) from climate change (McCarty 2001). Grassland bird species are declining faster than other groups of birds as well (Knopf 1994; Askins *et al.* 2007) and thus continued to be imperiled by these new threats to their habitat. Range-wide species distribution predictions have been made for grassland birds but some species with smaller ranges are not accurately modeled (O’Connor et al. 1999), perhaps because different drivers of distribution vary by region (Bakker, Naugle & Higgins 2002; Askins *et al.* 2007).

A suitable test area for STEMs and grassland birds should have regional variation in habitat and climate. In the southern Great Plains, the U.S. state of Oklahoma has over a third of its land area as grasslands and over 15% as croplands, containing grassland birds characteristic of habitats ranging from southeastern pine savannahs to tallgrass, mixed-grass, and shortgrass prairies (Askins *et al.* 2007; Diamond & Elliott 2015). Additionally, climate change is forecast to impact the Great Plains strongly (National Assessment Synthesis Team (U.S.) 2001; Loarie *et al.* 2009) and the climate of temperate grasslands are predicted to change rapidly relative to other biomes (Loarie *et al.* 2009). Oklahoma’s grassland habitats, agricultural importance, and forecast impact by climate change makes it an ideal region to test relative efficacy of different methods for modeling current and forecast species distributions.

The objectives of our study were threefold. First, we estimated the distribution of Oklahoma grassland birds to understand current distribution statewide with standard species distribution modeling methods. Next, these statewide current distribution predictions allowed us to compare the statewide species distribution model for each species with ensemble spatiotemporally explicit models to see whether this approach is suitable in terms of increased accuracy relative to increased computational power required for spatially explicit species distribution modeling. Finally, we ask how estimates of distributional changes from climate change may be altered by use of different spatial scale support sets. Our results will allow managers to make decisions on whether increased accuracy in modeling is worth the additional computational effort required by newer modeling techniques

# Methods

## Study area

Oklahoma contains diverse vegetation and climate, making it a suitable region to examine effects of spatially explicit models. There are ca. 165 vegetation types (based on soil and vegetation composition) in 15 land cover types (Diamond & Elliott 2015), with over a third of the vegetation in grasslands. Historically, grasslands were tallgrass (3 million ha), mixed grass (2.5 million ha), and shortgrass (1.3 million ha) (Samson & Knopf 1994). Rainfall and temperature vary across the state (Oklahoma Climatological Survey), with annual precipitation of about 17” of rain in the western portion to 56” in the eastern part of the state; mean annual temperature ranges from approximately 62°F in the southeast to about 56°F in the northwestern part of the state. Summer temperatures over 90°F can occur from 60-115 days out of the year varying statewide. Agriculture accounts for over $2.8 billion in the state’s gross domestic product in the study years (US Department of Commerce 2016) and includes crops such as winter wheat, hay, corn for grain, soybeans, and cotton, as well as cattle and hog ranching (*Oklahoma Agricultural Statistics 2016* 2016), ranking in the top 5 of US acreage for wheat for grain, winter wheat for grain, and forage land for hay (USDA/NASS 2016).

## Response data

We focused on 10 species of grassland birds found during our general surveys, plus Brown-headed Cowbirds (an obligate brood parasite with presence often depending on land use factors) (Benson, Chiavacci & Ward 2013), for a total of 11 species. Many of these species are declining at the state or North American level and none are increasing in population (Table 2). We collected data 1-4 times each at 339 8-min roadside point counts (0.13 hr and 0 km long) and at 87 cross-country transects 0.3-3.1 hrs and 0.3-4.3 km long (mean±SD: 1.2±0.6 hrs and 1.8±0.8 km). Each survey was conducted stationary (point counts) or walking at an even pace (transects). An observer noted species, number of individuals, and distance and angle to each sighting. Perpendicular distance from the transect line was calculated using distance and angle for transects. We had 14 observers total (6 in 2013 and 8 in 2014).

We supplemented our survey data with citizen science data from the eBird Reference Dataset (Munson *et al.* 2014). We used complete data from 2011-2014 and excluded casual counts. In the case of multiple checklists for a given sampling event, we used the primary checklist. Our maximum survey transect effort was 4.3 km length and 3.1 hrs time, so we restricted use of eBird samples to ≤4.3km and ≤3.1 hours to have comparable effort in all checklists. When combining our survey data with eBird data for use in species distribution models, we used the point count center or the transect midpoint as the count location to have comparable precision to eBird coordinates (Fink *et al.* 2010). We also only used survey sightings within 500 m of the point or transect for the species distribution models. Because some observers entered sightings from before and during our surveys into eBird, we eliminated 14 counts from 2013 and 2014 that were within two hours of the actual survey start time and within 15 km of the survey start location. For model evaluation data, we used datasets from 2011 and 2012. These filters resulted in a training (2711) and evaluation (2711) dataset of 5423 complete checklists (158 transect sampling events, 614 point count sampling events, and 4651 eBird sampling events); this was split randomly by half for training and evaluation by each species using the caret package (Kuhn 2017) createDataPartition function, which samples such that both training and evaluation splits have similar distributions of presence and absences. The temporally independent evaluation dataset from 2011-2012 consisted of 2541 complete checklists.

## Predictors

We used climatic variables, conservation easement status (O’Connor *et al.* 1999), and land cover variables to predict bird distribution (Table 1). We also used effort (length of observation in distance and time) and time of day. Mean effort (distance and time) and time of day rasters were created for prediction, such that all predictions are created assuming effort that does not vary geographically and amounts of effort that are typical for both surveys and citizen science efforts in 2013 and 2014 (mean time: 0.73 hr; mean distance: 0.75 km). The time of day raster for prediction was given a value of 7:00am (Fink et al. 2010). Neighborhood predictors were calculated by the values in rectangular areas around each point, at the scale of 5 x 5 pixels (150 x 150 m) and 15 x 15 pixels (450 x 450 m) (Fink et al. 2010). We looked at proportion of each land cover class and proportion of summed open space land covers (grasslands, hay/pasture, cropland, herbaceous wetlands, and barren land) since grassland bird occupancy can be influenced by the total non-structural cover (McDonald 2017). Neighborhoods were created in QGIS 2.16 with the GRASS r.neighbors processing tool (Quantum GIS Development Team 2016).

We chose only one climate model as the purpose of our climatic predictions is to give a range of estimates based on STEM scaling. We used the HadGEM2-ES (The HadGEM2 Development Team: G. M. Martin *et al.* 2011) downscaled Bioclim rasters for an intermediate climate model (4.5).

We created lower resolution (4.3 x 4.3 km resolution, to match the maximum size of transects and data in our response variable) rasters using the gdalwarp function in QGIS 2.16 (Quantum GIS Development Team 2016). Rasters were resampled using the “near” function which uses nearest neighbor resampling for categorical values and using “bilinear” which uses bilinear interpolation for continuous rasters.

## Current distribution

### Statewide species distribution model

All models were created using random forest classification trees (Breiman 2001) in the R package randomForest (Liaw & Wiener 2002). Random forest gives results competitive to those from other machine learning techniques such as boosted regression trees and bagged decision trees (used in Fink et al. 2010 for the non-spatially-explicit comparison model). Minimal tuning parameters are required (Caruana & Niculescu-Mizil 2006; Cutler *et al.* 2007; Guo *et al.* 2010). They are suitable for species distribution models (Prasad, Iverson & Liaw 2006; Lorena *et al.* 2011) and can use small sample sizes for presence records (Mi *et al.* 2017). The random forest algorithm bootstraps a subset of the data using only a set proportion of the predictor variables. It then calculates the error rate on training data using the “out of bag” sample (the portion of data not used in the bootstrap for each tree) (Hastie, Tibshirani & Friedman 2001). The trees are then averaged for a final model. The use of only some variables in each bootstrapped tree allows estimation of variable importance. A random forest model was created for the statewide scale for each species using all training data, with a prediction raster created using the predict function in the R package raster (Hijmans 2016). All random forests (including those in later models) were generated with 500 trees which is generally suitable for stable, accurate trees (Cutler *et al.* 2007). We used the default number of variables per bootstrap tree (default ‘mtry’=the square root of the number of predictor variables) for all trees because this is known to result in accurate predictions (Cutler et al. 2007).

To determine which predictors were important in current species distributions, for each species we ranked variables using the statewide model. We used the mean decrease in accuracy given by randomForest R package to rank variable importance. Mean decrease in accuracy (unscaled) with no sampling replaced is the least biased variable importance selection method available in randomForest (Strobl *et al.* 2007) We created partial dependence plots of the top variables for each statewide model to show how each variable increases or decreases probability of presence (Hastie, Tibshirani & Friedman 2001; Cutler *et al.* 2007). Variable importance can be measured more accurately in other packages at the expense of computational power (Strobl *et al.* 2007), so we tested variable importance again in the party package (Strobl *et al.* 2007, 2008). Due to the substantial increase in raster prediction time, we were unable to create predict rasters with cforest() in party, but this allowed us to compare variable importance in an unbiased manner with the multiple measurements scales of our rasters.

## Spatially explicit model accuracy and computational time

To model current species distributions based on our predictors, we created four models for each species at varying spatial scales: the previous statewide model and three spatially explicit ensemble models. The statewide model allowed use to compare the performance of the more computationally intense STEMs. The remaining models are three STEMs at varying support set scales, with some modifications from Fink et al. (2010). These locally dynamic models should provide more accurate maps than the statewide model for predicting species distribution both currently and in the future. Because they consist of numerous models (each which could generate its own set of variable importance rankings), they are too complex to use for variable importance ranking without a focal subregion. It is typical that models with increasing local accuracy are harder to interpret (because more accurate methods are often not very transparent) and generalize (because they typically fit the training dataset very well without regards to its ability to extrapolate) (James *et al.* 2013). With the diverse habitats and climatic variables found across Oklahoma, the models created by an ensemble of regional support sets should provide better predictions than the statewide model because the regional support sets will allow regional differences in important variables for prediction. Using both the statewide and spatially-explicit models give us complementary information on factors affecting species distribution in Oklahoma in addition to testing the usefulness of STEMs at this scale.

We adapted the STEM approach for our study in several ways. First, the scale of our support sets reflects the state extent (i.e. our small, medium, and large scales are relatively smaller than those needed for a continent-wide scale). As our survey goals are to determine breeding distribution only, we used a broader temporal window (April-July in all years) for our model (unlike Fink et al. 2010, who used single date windows). Secondly, for all base models, we used random forest classification trees (Breiman 2001) as described above. Finally, our geographic sampling of the training and evaluation datasets, described in more detail in the next paragraph, reflects the differing nature of our base models. Fink et al. (2010) sample 63% of each support set to imitate bootstrapping sampling, but because each support set is being bootstrapped by the random forest algorithm, we used the full support set for each region.

For the spatially explicit models, we created stratified random points in the study area (with the spsample function from the R package sp (Pebesma & Bivand 2005; Bivand, Pebesma & Gomez-Rubio 2013) and created a square of size small (100 points with 120 x 120 km), medium (37 points with 200 x 200 km), or large (12 points with 450 x 450 km) around these points, which resulted in no significant difference in pixel coverage (F2,147=0.63, p=0.53) before removing support sets with too few checklists. Each support set included all checklists from the training dataset located within its boundaries. The support set was discarded if it contained fewer than 25 checklists or did not contain both presences and absences (as the trees models failed in these conditions). We used the same random forest settings and created a prediction raster for each support set as for the statewide model. All support set rasters for a given scale (small, medium, or large) were stacked using the raster mosaic() function to get the mean value of each pixel, creating the spatially explicit ensemble (Hastie, Tibshirani & Friedman 2001; Fink *et al.* 2010; Oppel *et al.* 2012) made of many regional random forest predictions. This process was repeated at the three support set spatial scales, resulting in three spatially explicit, stacked ensembles per species for current distribution.

The statewide model was trained with the full training set (5423 checklists, as described previously). We tested the current distribution models with eBird data from 2011 and 2012 (2541 checklists, as described previously), sampled repeatedly for spatial uniformity (see below). Using data from different years results in a more accurate evaluation of whether the model generalizes well beyond its training timeframe (Araújo & Guisan 2006; Cutler *et al.* 2007). To evaluate model performance and choose between the four models for each species, we tested the models on the temporally independent evaluation dataset. To ensure spatially uniform testing (Fink *et al.* 2010), we created a statewide grid of 10 x 10 km cells. We randomly sampled no more than 10 observations from each grid cell in the evaluation dataset. We repeated the spatial sampling procedure randomly 50 times to get a distribution of performance for each model (Fink et al. 2010). We did this twice; once with the same year (2013-2014) evaluation dataset and again with the different year (2011-2012) evaluation dataset.

We tested four models in this way for each species and noted the scale (small, medium, large, statewide) with best performance measures for each species. Performance measures were root mean square error (RMSE) and area under the receiver operating curve (AUC). RMSE is calculated from the model residuals, taking the squared value of observed minus expected values, then taking the square root to return to original units; thus a larger value indicates the model deviates farther from expected values (Kuhn & Johnson 2013). AUC is a summary of model performance measuring how often the model misclassifies individual test observations; AUC ranges from 0 to 1, with 1 being perfect and 0.5 being a model that performs no better than random chance (Hanley & McNeil 1982; James *et al.* 2013).

Finally, to compare times to create trees and models, we used the R package microbenchmark to measure the time to run for each of the four models of the 11 species, resulting in 11 samples in each of the four model scales to compare computing efficiency. The time included all randomForest trees and the raster prediction. We included the mosaic creation time for the ensembles. We used linear mixed models to compare means for species of both error types for both years, with scale as the fixed effect and species as the random effect.

## Variation in future distribution predictions with use of STEM

We used the future models to test how much variation in range predictions is caused by the use of statewide versus spatiotemporally explicit models. We estimated percent loss of range area from the comparable-scale (statewide, small, medium, large) current distribution. As in previous models, species was included as a random effect.

# Results

## Current distribution

Current statewide distributions are shown in panel (a) of Figs. 1-4 and Figs. S1-7. The top 10 important variables (as ranked by the cforest tree implementation) are shown with partial dependence plots (based on the randomForest implementation) in Fig. S8-18.

## Spatially explicit model accuracy and computational time

Small (β=23.31±1.28, df=30, t=18.16, *P*<0.001), medium (β=20.04±1.28, df =30, t=15.62, *P*<0.001), and large (β=16.84±1.28, df =30, t=13.12, *P*<0.001) models took half a day to a day longer than statewide models to run. As computing times vary between machines, this means that used scaled models can take 2.7-12.7 times longer than a statewide model depending on species.

When accuracy was compared to runtimes, there were not significant relationships for AUC in same years (Fig. 5a; β=-0.00011±0.00013, df =32.01, t=-0.871, *P*=0.39), AUC in different years (Fig. 5b; β=-0.000049±0.00013, df =32.08, t=-0.374, *P*=0.71), or RMSE in same years (Fig. 5c; β=0.00025±0.00013, df =32.01, t=1.94, *P*=0.062). RMSE increased with longer runtime for different year evaluations (Fig. 5d; β=0.00026±0.000091, df =32.03, t=2.89, *P*=0.0068).

When examined by species, statewide models were usually more accurate, with less error than the scaled models (Table 3). Only three species had any evaluation measures better for non-statewide models. Northern Bobwhite (Fig. 1) had lower RMSE for medium models, but performed better with AUC for statewide models. Western Meadowlark (Fig. 3) had higher AUC for same-year evaluation large models. Brown-headed Cowbird (Fig. 4) had higher AUC for small models and lower RMSE for medium models (Table 3). The remaining species had equal or significantly better performance with statewide models relative to the scale models. All species’ distributions were predicted more accurately with data from the same years than with data from previous years for both AUC (β=0.35±0.011, df =76, t=31.64, *P*<0.0001) and RMSE (β=-0.06±0.0063, df =76, t=-9.59, *P*<0.0001).

## Variation in future distribution predictions with use of STEM

The estimated range change based on a future climate model did not vary significantly by scale when comparing spatially explicit models to statewide for large (β=-16.69±10.72, df =30, t=-1.56, *P*=0.13), medium (β=-20.36±10.72, df =30, t=-1.90, *P*=0.067), or small (β=-21.04±10.72, df =30, t=-1.96, *P*=0.059).

# Discussion

The use of spatially explicit ensemble modelling generally did not provide increased accuracy in predictions of species distribution in our study. While spatially explicit models increase model accuracy over continental scales (Fink *et al.* 2010; Fink, Damoulas & Dave 2013), our study found them mostly unnecessary at a small scale even in a state with variable climate and diverse ecoregions. In fact, we found a modern laptop or desktop unable to handle the spatially explicit models and turned to cloud computing to complete them, so the length of time and computing expense involved is not unsubstantial.

While most of our study species performed equal or better on most measures of accuracy with statewide models, three species did not. With one non-passerine (Northern Bobwhite), one common grassland passerine (Western Meadowlark), and one obligate brood parasite (Brown-headed Cowbird), there are no immediately obvious commonalities between the three species. A larger sample of species could indicate why certain species are modeled better with spatially explicit models versus statewide models. A potential variable includes whether species’ distributions depend more upon bioclimatic versus ecological variables, as bioclimatic variables should change more smoothly over a larger area (potentially reducing the need for adaptive local models). However, one benefit of random forest models and other machine learning methods is minimal tuning and minimal expert opinion required to generate an accurate map (Fink *et al.* 2010). One approach solution might be to run the faster statewide map, and then run spatially explicit models additionally only if certain conditions are met, but more study will be required to understand at what scale spatially explicit maps are necessary.

Finally, use of spatially explicit models did not add significant differences to climate change distribution predictions, but that the spatially explicit models increase computation time (as for the current distribution models). Climate change models may be overall accurate but poorly forecast a given region without additional interpretation by local experts (Risbey *et al.* 2002). Additionally, climate models are generated at much coarser scales than our response dataset (Harris *et al.* 2014) and models may fail to predict accurately when future climates do not have current-day analogs (Boiffin, Badeau & Bréda 2017).

Accurate species distribution models can help us understand what factors, both environmental and land use, that drive species declines (Elith & Leathwick 2009), but we need to conduct modeling with predictors and responses at the appropriate spatial scale. While large-scale solutions to conserve grasslands are needed (Samson, Knopf & Ostlie 2004), local and regional conservation and management efforts can also have critical impacts (Brennan, Kuvlesky & Morrison 2005). Based on our study, we recommend the use of standard species distribution models at the statewide scale as the use of spatially explicit models increases computation time dramatically for no gain or even a loss in accuracy. Instead, if local modelers seek increased accuracy, we recommend use of non-spatially explicit ensemble models, where different types of distribution models are combined over the whole study area to produce more refined prediction maps (Araujo & New 2007; Oppel *et al.* 2012). Future work should examine at what scale spatially explicit models become necessary between a diverse state versus a diverse continent (Fink *et al.* 2010). whether we can determine based on habitat values and scales of spatial heterogeneity whether a spatially explicit model is appropriate for given species and prediction range. It may be that species-specific processes determine whether spatially explicit models are required (as suggested by only the obligate brood parasite in our study doing better with small scaled spatially explicit models).

# Authors’ contributions

ESB, JDR, AJC, and CMC conceived the ideas and designed methodology. ESB and JDR collected data. CMC analyzed the data. All authors contributed critically to the drafts and gave final approval for publication.

# Acknowledgements

This work was funded by USDA-NIFA grant #2013-67009-20369 to ESB and supported by the AWS Cloud Credits for Research program. This study was conducted under OU IACUC R12-019. We thank technicians M. Buschow, R. Feliciano, M. Fuirst, J. Haughawout, J. Hightower, J. Kruk, C. Myers, K. Oliver, T. Smith, R. Soto, and J. Tibbits for help in gathering survey data. T. Fagin and W.T. Honeycutt provided technical assistance and comments. CMC was also supported by National Science Foundation grant IDBR 1014891 to ESB.

# References

Alward, R.D. (1999) Grassland Vegetation Changes and Nocturnal Global Warming. *Science*, **283**, 229–231.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Araujo, M. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

Askins, R.A., Chávez-Ramírez, F., Dale, B.C., Haas, C.A., Herkert, J.R., Knopf, F.L. & Vickery, P.D. (2007) Conservation of grassland birds in North America: understanding ecological processes in different regions: ‘Report of the AOU Committee on Conservation’. *Ornithological Monographs*, **64**, iii-46.

Bakker, K.K., Naugle, D.E. & Higgins, K.F. (2002) Incorporating Landscape Attributes into Models for Migratory Grassland Bird Conservation. *Conservation Biology*, **16**, 1638–1646.

Beaumont, L.J., Pitman, A.J., Poulsen, M. & Hughes, L. (2007) Where will species go? Incorporating new advances in climate modelling into projections of species distributions. *Global Change Biology*, **13**, 1368–1385.

Benson, T.J., Chiavacci, S.J. & Ward, M.P. (2013) Patch size and edge proximity are useful predictors of brood parasitism but not nest survival of grassland birds. *Ecological Applications*, **23**, 879–887.

Bivand, R.S., Pebesma, E. & Gomez-Rubio, V. (2013) *Applied Spatial Data Analysis with R*, 2nd ed. Springer, New York.

Boiffin, J., Badeau, V. & Bréda, N. (2017) Species distribution models may misdirect assisted migration: insights from the introduction of Douglas-fir to Europe. *Ecological Applications*, **27**, 446–457.

Breiman, L. (2001) Random forests. *Machine learning*, **45**, 5–32.

Brennan, L.A., Kuvlesky, W.P. & Morrison. (2005) Invited paper: North American grassland birds: an unfolding conservation crisis? *Journal of Wildlife Management*, **69**, 1–13.

Brown, J.H., Valone, T.J. & Curtin, C.G. (1997) Reorganization of an arid ecosystem in response to recent climate change. *Proceedings of the National Academy of Sciences*, **94**, 9729–9733.

Caruana, R. & Niculescu-Mizil, A. (2006) An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning*, pp. 161–168. ACM.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J. (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783–2792.

Diamond, D.D. & Elliott, L.F. (2015) *Oklahoma Ecological Systems Mapping Interpretive Booklet: Methods, Short Type Descriptions, and Summary Results*. Oklahoma Department of Wildlife Conservation, Norman, OK.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., R. Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species’ distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Fink, D., Damoulas, T. & Dave, J. (2013) Adaptive Spatio-Temporal Exploratory Models: Hemisphere-wide species distributions from massively crowdsourced eBird data. *AAAI*

Fink, D., Hochachka, W.M., Zuckerberg, B., Winkler, D.W., Shaby, B., Munson, M.A., Hooker, G., Riedewald, M., Sheldon, D. & Kelling, S. (2010) Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications*, **20**, 2131–2147.

Guo, Y., Graber, A., McBurney, R.N. & Balasubramanian, R. (2010) Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. *BMC Bioinformatics*, **11**, 447.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology 143: 29–36*, **143**, 29–36.

Harris, R.M.B., Grose, M.R., Lee, G., Bindoff, N.L., Porfirio, L.L. & Fox-Hughes, P. (2014) Climate projections for ecologists. *Wiley Interdisciplinary Reviews: Climate Change*, **5**, 621–637.

Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The Elements of Statistical Learning*. Springer series in statistics Springer, Berlin.

Higgins, K.F., Naugle, D.E. & Forman, K.J. (2002) A Case Study of Changing Land Use Practices in the Northern Great Plains, U.S.A.: An Uncertain Future for Waterbird Conservation. *Waterbirds: The International Journal of Waterbird Biology*, **25**, 42–50.

Hijmans, R.J. (2016) *Raster: Geographic Data Analysis and Modeling*.

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

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An Introduction to Statistical Learning*. Springer New York, New York, NY.

Klute, D.S., Robel, R.J. & Kemp, K.E. (1997) Will Conversion of Conservation Reserve Program (CRP) Lands to Pasture be Detrimental for Grassland Birds in Kansas? *American Midland Naturalist*, **137**, 206.

Knopf, F.L. (1994) Avian assemblages on altered grasslands. *Studies in Avian Biology*, **15**, 247–257.

Kuhn, M. Predictive Modeling with R and the caret Package.

Kuhn, M. & Johnson, K. (2013) *Applied Predictive Modeling*. Springer New York, New York, NY.

Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. *R News*, **2**, 18–22.

Lipsey, M.K., Doherty, K.E., Naugle, D.E., Fields, S., Evans, J.S., Davis, S.K. & Koper, N. (2015) One step ahead of the plow: Using cropland conversion risk to guide Sprague’s Pipit conservation in the northern Great Plains. *Biological Conservation*, **191**, 739–749.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The velocity of climate change. *Nature*, **462**, 1052–1055.

Lorena, A.C., Jacintho, L.F.O., Siqueira, M.F., Giovanni, R.D., Lohmann, L.G., de Carvalho, A.C.P.L.F. & Yamamoto, M. (2011) Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*, **38**, 5268–5275.

McCarty. (2001) 2001\_McCarty\_ecological\_consequences\_of\_climate\_change.pdf.

Mi, C., Huettmann, F., Guo, Y., Han, X. & Wen, L. (2017) Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*, **5**, e2849.

Munson, M.A., Webb, K., Sheldon, D., Fink, D., Hochachka, W.M., Iliff, M., Riedewald, M., Sorokina, D., Sullivan, B., Wood, C. & Kelling, S. (2014) *The eBird Reference Dataset, Version 2014.* Cornell Lab of Ornithology, Ithaca, NY.

National Assessment Synthesis Team (U.S.) (ed). (2001) *Climate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change: Foundation*. Cambridge University Press, Cambridge ; New York.

O’Connor, R.J., Jones, M.T., Boone, R.B. & Lauber, T.B. (1999) Linking continental climate, land use, and land patterns with grassland bird distribution across the conterminus United States. *Studies in Avian Biology*, **19**, 45–59.

*Oklahoma Agricultural Statistics 2016*. (2016)

Oklahoma Climatological Survey. Climate of Oklahoma, http://climate.ok.gov/index.php/site/page/climate\_of\_oklahoma

Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O’Connell, A.F., Miller, P.I. & Louzao, M. (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation*, **156**, 94–104.

Pebesma, E. & Bivand, R.S. (2005) Classes and methods for spatial data in R. *R News*, **5**.

Phillips, S.J., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the twenty-first international conference on Machine learning*, p. 83. ACM.

Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, **9**, 181–199.

Quantum GIS Development Team. (2016) *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project.

Risbey, J.S., Lamb, P.J., Miller, R.L., Morgan, M.C. & Roe, G.H. (2002) Exploring the structure of regional climate scenarios by combining synoptic and dynamic guidance and GCM output. *Journal of climate*, **15**, 1036–1050.

Samson, F. & Knopf, F.L. (1994) Prairie conservation in North America. *BioScience*, **44**, 418–421.

Samson, F.B., Knopf, F.L. & Ostlie, W. (2004) Great Plains ecosystems: past, present, and future. *Wildlife Society Bulletin*, **32**, 6–15.

Sauer, J.R., Niven, D.K., Hines, J.E., Ziolkowski, Jr., D.J., Pardieck, K.L., Fallon, J.E. & Link, W.A. (2017) *The North American Breeding Bird Survey, Results and Analysis 1966 - 2015, Version 2.07.2017*. USGS Patuxent Wildlife Research Center, Laurel, MD.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. (2008) Conditional variable importance for random forests. *BMC Bioinformatics*, **9**, 307.

Strobl, C., Boulesteix, A.-L., Zeileis, A. & Hothorn, T. (2007) Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, **8**, 25.

The HadGEM2 Development Team: G. M. Martin, Bellouin, N., Collins, W.J., Culverwell, I.D., Halloran, P.R., Hardiman, S.C., Hinton, T.J., Jones, C.D., McDonald, R.E., McLaren, A.J., O’Connor, F.M., Roberts, M.J., Rodriguez, J.M., Woodward, S., Best, M.J., Brooks, M.E., Brown, A.R., Butchart, N., Dearden, C., Derbyshire, S.H., Dharssi, I., Doutriaux-Boucher, M., Edwards, J.M., Falloon, P.D., Gedney, N., Gray, L.J., Hewitt, H.T., Hobson, M., Huddleston, M.R., Hughes, J., Ineson, S., Ingram, W.J., James, P.M., Johns, T.C., Johnson, C.E., Jones, A., Jones, C.P., Joshi, M.M., Keen, A.B., Liddicoat, S., Lock, A.P., Maidens, A.V., Manners, J.C., Milton, S.F., Rae, J.G.L., Ridley, J.K., Sellar, A., Senior, C.A., Totterdell, I.J., Verhoef, A., Vidale, P.L. & Wiltshire, A. (2011) The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, **4**, 723–757.

US Department of Commerce. (2016) Bureau of Economic Analysis, https://www.bea.gov/iTable/index\_regional.cfm

U.S. Department of Commerce/U.S. Census Bureau. (2010) TIGER/Line Shapefile, 2010, 2010 state, Oklahoma, 2010 Census Block State-based.

USDA/NASS. (2016) USDA/NASS 2016 State Agriculture Overview for Oklahoma, https://www.nass.usda.gov/Quick\_Stats/Ag\_Overview/stateOverview.php?state=OKLAHOMA

USDA/NRCS - National Geospatial Center of Excellence. (2010) USDA-NRCS Easement Areas: Oklahoma.

USDA/NRCS - National Geospatial Center of Excellence. (2011) 2011 National Land Cover Dataset (NLCD): Oklahoma.

Wright, C.K. & Wimberly, M.C. (2013) Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences*, **110**, 4134–4139.

# Tables

Table 1. Predictors used in models.

|  |  |  |
| --- | --- | --- |
| Predictor variable name | Definition | Source |
| effort\_length | Length of survey (km) | Survey data |
| effort\_time | Duration of survey (hr) | Survey data |
| time\_of\_day | Time of survey | Survey data |
| conservation\_easements\_presenceabsence | Presence (1) or absence (0) of a conservation easement | (USDA/NRCS - National Geospatial Center of Excellence 2010) |
| conservation\_easements\_CalcArea | Area of the conservation easement in which a given pixel exists (acres) |
| nlcd\_ok\_utm14\_okmask | NLCD2011 Landcover Classes | (USDA/NRCS - National Geospatial Center of Excellence 2011)  Neighborhoods modified from NLCD landcover classes. |
|  | NLCD 2.25 ha (5x5 cells) and 20.25 ha (15 x 15 cells) neighborhoods: proportion of neighborhood with the named land cover classes (values range from 0 to 1). Definitions described the land cover type and list the category numbers included in each neighborhood. |
| undevopenspace\_5cell\_okmask  undevopenspace\_15cell\_okmask | Undeveloped open space (11, 31, 71, 81, 82, 95) |
| openwater11\_5cell\_okmask  openwater11\_15cell\_okmask | Open water (11) |
| dev\_openspace21\_5cell\_okmask  dev\_openspace21\_15cell\_okmask | Developed open space (21) |
| dev\_low22\_5cell\_okmask | Low intensity development (22) |
| dev\_med23\_5cell\_okmask | Medium intensity development (23) |
| dev\_high24\_5cell\_okmask | High intensity development (24) |
| barren31\_5cell\_okmask  barren31\_15cell\_okmask | Barren (31) |
| forest41to43\_5cell\_okmask | Forest (41, 42, 43) |
| scrub52\_5cell\_okmask | Scrub and shrubland (52) |
| grasslands71\_5cell\_okmask | Grasslands (71) |
| pasturehay81\_5cell\_okmask | Pasture and hay (81) |
| croplands82\_5cell\_okmask | Croplands (82) |
| woodywetlands90\_5cell\_okmask | Woody wetlands (90) |
| herbwetlands95\_5cell\_okmask | Herbaceous wetlands (95) |
| census\_utm\_30m | Human population density in number per km2 | (U.S. Department of Commerce/U.S. Census Bureau 2010) |
| bio1\_12\_OK | BIO1 = Annual Mean Temperature | Bioclim variables from Worldclim (Hijmans *et al.* 2005) |
| bio\_12\_OK | BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)) |
| bio3\_12\_OK | BIO3 = Isothermality (BIO2/BIO7) (\* 100) |
| bio4\_12\_OK | BIO4 = Temperature Seasonality (standard deviation \*100) |
| bio5\_12\_OK | BIO5 = Max Temperature of Warmest Month |
| bio6\_12\_OK | BIO6 = Min Temperature of Coldest Month |
| bio7\_12\_OK | BIO7 = Temperature Annual Range (BIO5-BIO6) |
| bio8\_12\_OK | BIO8 = Mean Temperature of Wettest Quarter |
| bio9\_12\_OK | BIO9 = Mean Temperature of Driest Quarter |
| bio10\_12\_OK | BIO10 = Mean Temperature of Warmest Quarter |
| bio11\_12\_OK | BIO11 = Mean Temperature of Coldest Quarter |
| bio12\_12\_OK | BIO12 = Annual Precipitation |
| bio13\_12\_OK | BIO13 = Precipitation of Wettest Month |
| bio14\_12\_OK | BIO14 = Precipitation of Driest Month |
| bio15\_12\_OK | BIO15 = Precipitation Seasonality (Coefficient of Variation) |
| bio16\_12\_OK | BIO16 = Precipitation of Wettest Quarter |
| bio17\_12\_OK | BIO17 = Precipitation of Driest Quarter |
| bio18\_12\_OK | BIO18 = Precipitation of Warmest Quarter |
| bio19\_12\_OK | BIO19 = Precipitation of Coldest Quarter |

Table 2. Study species with their Breeding Bird Survey trends for Oklahoma and North America (Sauer et al. 2017). Values are reported as trends with 95% credible intervals (lower, upper). Significant declines are marked with **boldface**. Sample size is reported as number of routes. Species with a decline at any region or timeframe are in italics.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species | Breeding bird survey trend in Oklahoma 1966-2015 | 2005-2015 | n | BBS trend North America 1966-2015 | 2005-2015 | n |
| *Northern Bobwhite* | **-1.57**  **(-2.24, -0.89)** | **-3.03**  **(-5.50, -0.57)** | 60 | **-3.48**  **(-3.77, -3.24)** | **-2.79**  **(-3.51, -2.07)** | 1983 |
| Upland Sandpiper | 0.20  (-2.71, 2.51) | 3.21  (-6.07, 14.80) | 32 | 0.40  (-0.14, 0.87) | 1.50  (0.33, 2.77) | 994 |
| *Horned Lark* | 0.61  (-1.47, 2.78) | 1.64  (-2.20, 7.89) | 53 | **-2.46**  **(-2.84, -2.12)** | **-2.08**  **(-2.66, -1.48)** | 2598 |
| Cassin’s Sparrow | 0.01  (-4.03, 4.13) | 2.29  (-6.71, 13.46) | 25 | -0.75  (-2.30, 0.33) | -1.72  (-4.27, 1.20) | 333 |
| *Field Sparrow* | **-2.94**  **(-3.59, -2.29)** | **-6.63**  **(-9.05, -4.26)** | 53 | **-2.33**  **(-2.60, -2.17)** | **-2.18**  **(-2.71, -1.67)** | 2192 |
| *Lark Sparrow* | **-1.20**  **(-1.97, -0.39)** | -0.88  (-2.44, 0.88) | 60 | **-0.78**  **(-1.19, -0.39)** | 0.07  (-0.62, 0.77) | 1628 |
| *Grasshopper Sparrow* | 0.20  (-1.22, 1.61) | 0.87  (-2.30, 4.73) | 56 | **-2.52**  **(-3.69, -2.04)** | -0.77  (-1.98, 0.45) | 2161 |
| Dickcissel | 0.65  (-0.22, 1.55) | 1.18  (-1.60, 4.30) | 60 | -0.36  (-0.86, 0.05) | 0.74  (-0.27, 1.72) | 1269 |
| *Eastern Meadowlark* | **-2.09**  **(-2.63, -1.53)** | **-2.64**  **(-4.18, -1.03)** | 55 | **-3.28**  **(-3.60, -3.03)** | **-3.05**  **(-3.56, -2.29)** | 2526 |
| *Western Meadowlark* | -1.24  (-2.47, 0.04) | -1.84  (-4.28, 0.96) | 38 | **-1.29**  **(-1.54, -1.11)** | **-1.21**  **(-1.61, -0.83)** | 2085 |
| *Brown-headed Cowbird* | **-1.34**  **(-1.87, -0.77)** | -1.40  (-2.89, 0.22) | 60 | **-0.66**  **(-0.81, -0.51)** | -0.07  (-0.51, 0.41) | 4409 |

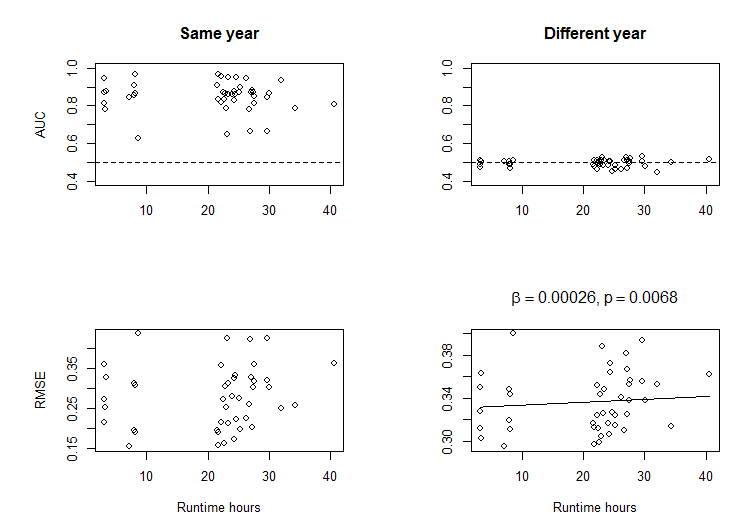
*Table 3.* A priori *comparisons between 50 evaluations each of large (L), medium (M), and small (S) scales to statewide models in linear models (species as random effect). Values shown are β±SE (P-value, Holm-corrected for multiple comparisons within each year and error type). Cells in* ***bold*** *have scaled models better than statewide. Cells in* italic *have scaled models worse than statewide.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | *β±SE (P)* | | | |
|  |  |  | Same year evaluations | | Different year evaluations | |
| Species | Presences out of 2711 checklists | Scale compared to statewide | AUC | RMSE | AUC | RMSE |
| Northern Bobwhite | 305 | large | 0.001±0.001 (1) | **-0.002±0.001 (0.0021)** | 0±0.004 (1) | 0.001±0.001 (1) |
| medium | 0.003±0.001 (0.1061) | **-0.005±0.001 (<0.0001)** | 0.005±0.004 (1) | **-0.005±0.001 (0.0004)** |
| small | 0±0.001 (1) | **-0.004±0.001 (<0.0001)** | -0.01±0.004 (0.2025) | **-0.005±0.001 (0.001)** |
| Upland Sandpiper | 52 | large | -0.008±0.003 (0.3093) | 0.002±0.001 (0.0681) | 0.008±0.003 (0.3205) | 0.002±0.002 (1) |
| medium | -0.008±0.003 (0.2524) | *0.007±0.001 (<0.0001)* | 0.001±0.003 (1) | 0.004±0.002 (0.0674) |
| small | *-0.013±0.003 (0.0043)* | *0.018±0.001 (<0.0001)* | 0.002±0.003 (1) | *0.011±0.002 (<0.0001)* |
| Horned Lark | 119 | large | -0.001±0.001 (1) | -0.001±0.001 (0.8859) | -0.001±0.004 (1) | -0.003±0.001 (0.4889) |
| medium | *-0.01±0.001 (<0.0001)* | 0.001±0.001 (0.2378) | *-0.025±0.004 (<0.0001)* | *0.005±0.001 (0.002)* |
| small | *-0.025±0.001 (<0.0001)* | *0.007±0.001 (<0.0001)* | *-0.021±0.004 (<0.0001)* | *0.006±0.001 (0.0006)* |
| Cassin’s Sparrow | 142 | large | 0.001±0.001 (1) | -0.001±0.001 (0.3232) | 0.011±0.004 (0.1024) | 0.002±0.001 (0.7755) |
| medium | *-0.009±0.001 (<0.0001)* | *0.024±0.001 (<0.0001)* | -0.003±0.004 (1) | *0.013±0.001 (<0.0001)* |
| small | *-0.014±0.001 (<0.0001)* | *0.032±0.001 (<0.0001)* | *-0.016±0.004 (0.0004)* | *0.016±0.001 (<0.0001)* |
| Field Sparrow | 162 | large | 0.006±0.002 (0.0956) | -0.001±0.001 (0.8859) | 0.004±0.003 (1) | 0.002±0.001 (0.2874) |
| medium | 0±0.002 (1) | *0.007±0.001 (<0.0001)* | 0.003±0.003 (1) | *0.008±0.001 (<0.0001)* |
| small | 0.006±0.002 (0.1245) | *0.006±0.001 (<0.0001)* | -0.006±0.003 (0.8383) | *0.011±0.001 (<0.0001)* |
| Lark Sparrow | 402 | large | 0.002±0.001 (1) | -0.001±0.001 (0.5436) | 0±0.003 (1) | 0.001±0.001 (0.9947) |
| medium | -0.002±0.001 (0.7297) | 0.001±0.001 (0.8859) | 0.01±0.003 (0.0807) | *0.006±0.001 (<0.0001)* |
| small | *-0.007±0.001 (<0.0001)* | *0.002±0.001 (0.0003)* | 0.008±0.003 (0.3286) | *0.011±0.001 (<0.0001)* |
| Grasshopper Sparrow | 206 | large | -0.001±0.001 (1) | 0.001±0.001 (0.3232) | 0.006±0.003 (0.9169) | 0.001±0.001 (1) |
| medium | *-0.004±0.001 (0.0023)* | *0.003±0.001 (<0.0001)* | -0.006±0.003 (1) | 0.003±0.001 (0.1643) |
| small | *-0.014±0.001 (<0.0001)* | *0.008±0.001 (<0.0001)* | -0.005±0.003 (1) | *0.005±0.001 (0.0099)* |
| Dickcissel | 403 | large | 0.001±0.001 (1) | 0±0.001 (0.8859) | 0.009±0.004 (0.2264) | 0±0.001 (1) |
| medium | *-0.006±0.001 (<0.0001)* | *0.005±0.001 (<0.0001)* | 0.002±0.004 (1) | *0.004±0.001 (0.0004)* |
| small | *-0.011±0.001 (<0.0001)* | *0.008±0.001 (<0.0001)* | 0.005±0.004 (1) | *0.007±0.001 (<0.0001)* |
| Eastern Meadowlark | 473 | large | 0.001±0.001 (1) | -0.001±0.001 (0.8859) | 0±0.003 (1) | 0±0.001 (1) |
| medium | *-0.003±0.001 (0.0028)* | 0.001±0.001 (0.3542) | 0.01±0.003 (0.1455) | *0.004±0.001 (0.0024)* |
| small | *-0.011±0.001 (<0.0001)* | *0.007±0.001 (<0.0001)* | 0±0.003 (1) | *0.009±0.001 (<0.0001)* |
| Western Meadowlark | 182 | large | **0.007±0.001 (<0.0001)** | -0.001±0.001 (0.2056) | 0.008±0.004 (0.6145) | -0.002±0.001 (0.9947) |
| medium | *-0.003±0.001 (<0.0001)* | *0.011±0.001 (<0.0001)* | -0.009±0.004 (0.2264) | *0.013±0.001 (<0.0001)* |
| small | *-0.013±0.001 (<0.0001)* | *0.036±0.001 (<0.0001)* | *-0.025±0.004 (<0.0001)* | *0.025±0.001 (<0.0001)* |
| Brown-headed Cowbird | 591 | large | **0.021±0.002 (<0.0001)** | **-0.011±0.001 (<0.0001)** | **0.012±0.004 (0.0221)** | **-0.012±0.001 (<0.0001)** |
| medium | **0.033±0.002 (<0.0001)** | **-0.016±0.001 (<0.0001)** | **0.016±0.004 (0.0004)** | **-0.018±0.001 (<0.0001)** |
| small | **0.037±0.002 (<0.0001)** | **-0.012±0.001 (<0.0001)** | **0.018±0.004 (<0.0001)** | **-0.006±0.001 (<0.0001)** |

# Figures

Figs 1-4 see pdfs.

Fig. 5. Error evaluations for all 44 models relative to runtime. Dashed line in a and b show AUC = 0.5, where prediction is random, and above which prediction is better than random. a. AUC in same years; b. AUC in different years; c. RMSE in same years; d. RMSE in different years.



a

b

c

d

# Supplemental figures S1-S18

See pdfs

# Appendix A.

Separate R code files to be added