Ecology and Evolution
Open Access

VARYING DATASET RESOLUTION ALTERS PREDICTIVE ACCURACY OF SPATIALLY EXPLICIT ENSEMBLE MODELS FOR AVIAN SPECIES DISTRIBUTION

Journal:	*Ecology and Evolution*
Manuscript ID	Draft
Wiley - Manuscript type:	Original Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Curry, Claire M.; University of Oklahoma, University Libraries; University of Oklahoma, Oklahoma Biological Survey Ross, Jeremy; University of Oklahoma, Oklahoma Biological Survey Contina, Andrea; University of Oklahoma, Oklahoma Biological Survey Bridge, Eli; University of Oklahoma, Oklahoma Biological Survey
Category:	Population Ecology
Habitat:	Terrestrial, Theoretical, Ecosystem
Organism:	Vertebrate
Approach:	Comparative, Method Development
Abstract:	1. Species distribution models can be made more accurate by use of new “Spatiotemporal Exploratory Models” (STEMs), a type of spatially explicit ensemble model (SEEM) developed at the continental scale that averages regional models pixel by pixel. Although SEEMs can generate more accurate predictions of species distributions, they are computationally expensive. We compared the accuracies of each model for 11 grassland bird species, and examined whether they improve accuracy at a statewide scale for fine and coarse predictor resolutions. 2. We used a combination of survey data and citizen science data for 11 grassland bird species in Oklahoma to test a spatially explicit ensemble model at a smaller scale for its effects on accuracy of current models. 3. We found that only four species performed best with either a statewide model or SEEM; the most accurate model for the remaining seven species varied with data resolution and performance measure. 5. Policy implications: Determination of non-heterogeneity may depend on the spatial resolution of the examined dataset. Managers should be cautious if any regional differences are expected when developing policy from rangewide results that show a single model or timeframe. We recommend use of standard species distribution models or other types of non-spatially explicit ensemble models for local species prediction models. Further study is necessary to understand at what point SEEMs become necessary with varying dataset resolutions.

VARYING DATASET RESOLUTION ALTERS PREDICTIVE ACCURACY OF SPATIALLY EXPLICIT ENSEMBLE MODELS FOR AVIAN SPECIES DISTRIBUTION

Claire M. Curry[1,2]

Jeremy D. Ross[2]

Andrea J. Contina[2]

Eli S. Bridge[2]

[1]University Libraries, University of Oklahoma, 401 W. Brooks St, Norman, OK 73019

[2]Oklahoma Biological Survey, University of Oklahoma, 111 E. Chesapeake Street, Norman, OK 73019

Corresponding author: Claire M. Curry; address: University Libraries, University of Oklahoma, 401 W. Brooks St, Norman, OK 73019, email: cmcurry@ou.edu

Running title: Prediction accuracy of spatially explicit ensemble models

Summary

1. Species distribution models can be made more accurate by use of new "Spatiotemporal Exploratory Models" (STEMs), a type of spatially explicit ensemble model (SEEM) developed at the continental scale that averages regional models pixel by pixel. Although SEEMs can generate more accurate predictions of species distributions, they are computationally expensive. We compared the accuracies of each model for 11 grassland bird species, and examined whether they improve accuracy at a statewide scale for fine and coarse predictor resolutions. 2. We used a combination of survey data and citizen science data for 11 grassland bird species in Oklahoma to test a spatially explicit ensemble model at a smaller scale for its effects on accuracy of current models. 3. We found that only four species performed best with either a statewide model or SEEM; the most accurate model for the remaining seven species varied with data resolution and performance measure. 5. **Policy implications:** Determination of non-heterogeneity may depend on the spatial resolution of the examined dataset. Managers should be cautious if any regional differences are expected when developing policy from rangewide results that show a single model or timeframe. We recommend use of standard species distribution models or other types of non-spatially explicit ensemble models for local species prediction models. Further study is necessary to understand at what point SEEMs become necessary with varying dataset resolutions.

Key words: random forest, machine learning, spatiotemporal exploratory models, Oklahoma, grassland birds, landscape ecology, data resolution

INTRODUCTION

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 (Jane 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 (Cutler et al., 2007; J. Elith, Leathwick, & Hastie, 2008; Jane Elith et al., 2006; Lorena et al., 2011; Phillips, Dudík, & Schapire, 2004). One such method, Spatiotemporal Exploratory Modeling (STEM), has recently been introduced as a means of coping with variation in regional drivers. 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 (the model type used here is referred to as the base model) 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 ensemble technique of combining overlapping predictions can be used with almost any model type (Fink et al. 2010, Fink et al. 2013), and can cover continent- to hemisphere-wide scales (Fink et al., 2018; Fink, Damoulas, & Dave, 2013). Unfortunately, these spatially explicit ensemble models (SEEMs) are computationally expensive, because instead of predicting just one map they must predict numerous supporting maps followed by averaging them to create the final model. Additionally, the relative increase in accuracy has not been compared to the relative expense of computational time nor have SEEMs been tested at scales at which much species management occurs, such as state or regional initiatives (Brennan, Kuvlesky, & Morrison, 2005).

STEMs have been developed for continental scale analyses because such a broad scale provides enough habitat and climate variation to require such a model. However, there are cases in which even a regional scale dataset can provide a wide range of bioclimatic heterogeneity and therefore can be suitable for this application (Johnston et al., 2015; Zuckerberg, Fink, La Sorte, Hochachka, & Kelling, 2016). The state of Oklahoma in the United States (U.S.) provides a such case because of its high biodiversity, ranking 9th for bird species richness, 15th for total species richness, and above the median in species richness for reptiles, amphibians, freshwater fish, vascular plants, and mammals in the U.S. (Stein, 2002). In particular, the grassland birds of Oklahoma inhabit diverse grassland types and climatic extremes. The open habitats of

Oklahoma, which contains over a third of its land area as grasslands and an additional 15% as croplands (Diamond & Elliott, 2015), contain grassland birds characteristic of habitats ranging from southeastern pine savannahs to tallgrass, mixed-grass, and shortgrass prairies (Askins et al., 2007; Diamond & Elliott, 2015). Grassland species in areas half the size of Oklahoma in a single ecoregion have shown spatial and temporal differences in variable importance (Ethier, Koper, & Nudds, 2017). Forest species, which likewise occupy a single habitat type, also show spatial and temporal variation in predictor importance (Zuckerberg et al., 2016). Similarly, such a technique has been used on shorebirds in habitats with structural similarity to grasslands at a statewide scale (Johnston et al., 2015). Finally, Oklahoma occurs on a strong east-west climatic gradient (Oklahoma Climatological Survey 2017) that has had profound impacts on the ecosystems of the region (Kukal & Irmak, 2016; Seager et al., 2018). Physiological balances in animals can change in importance with other environmental variables (Kearney, Simpson, Raubenheimer, & Kooijman, 2013), therefore variable importance may be expected to change for at least some species across climatic gradients. Oklahoma's grassland habitats, agricultural importance, and susceptibility to climate change (Loarie et al., 2009; National Assessment Synthesis Team (U.S.), 2001) make it an ideal and important region to test relative efficacy of different methods for modeling species distributions.

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 (F. 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), although 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), alterations to vegetation (Alward, 1999) and ecosystem structure (Brown, Valone, & Curtin, 1997; Hamer, Flather, & Noon, 2006), and climate change (McCarty, 2001). Grassland bird species are declining faster than other groups of birds (Askins et al., 2007; Hill, Egan, Stauffer, & Diefenbach, 2014; Knopf, 1994) and continue to be imperiled by ongoing and expanding 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 some drivers of distribution vary regionally (Askins et al., 2007; Bakker, Naugle, & Higgins, 2002; Ethier et al., 2017). Additionally, spatial and temporal variation in habitat needs and selection pressures (Davis 2005; Winter, Johnson & Shaffer 2005) or interactions with weather events (Pipher, Curry, & Koper, 2016) are known to be important in grassland birds, therefore they are particularly suitable as a testing ground for a spatially explicit approach to modeling.

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 SEEMs to evaluate whether this approach is suitable at the scale of our region. Finally, we compared each approach's accuracy when using fine- or coarse-resolution predictor sets. Although our approach is at a smaller scale than originally envisioned for SEEMs, it is important to test their potential applicability at the smaller scales at which most management decisions are made. Our results will allow others to make decisions on whether increased accuracy in modeling is worth the additional computational effort required by newer modeling techniques and provide guidance for future work into where given modeling applications are useful.

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. Rainfall and temperature vary across the state (Oklahoma Climatological Survey, 2017), with annual precipitation ranging from ~ 43cm of rain in the northwest to 142 cm in the southeast and mean annual temperature ranging from ~13°C in the northwest to ~17°C in the southeast. Summer temperatures over 32°C can occur from 60-115 days out of the year varying statewide. Agriculture in Oklahoma is dominated by livestock ranching and row crops (USDA/NASS, 2016), and accounted for over $2.8 billion of the state's gross domestic product in the study years (US Bureau of Economic Analysis, 2016); Oklahoma ranks in the top 5 of US acreage for grain wheat and forage land for hay (USDA/NASS, 2016).

Bird surveys

We collected data 1-4 times each at 339 8-min roadside point counts (0.13 hr) and at 87 non-roadside 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). We had 14 observers total (6 in 2013 and 8 in 2014). We only used sightings within 500 m of the observer to preserve identification accuracy and recognize that detection is imperfect; however, all models compared use similar data and as such it should not impact our comparison of models. A zero (absence) or 1 (presence) was assigned for each combination of date and time and species. We focused on 10 species of grassland birds found during our general surveys [Northern Bobwhite (*Colinus virginianus*); Upland Sandpiper (*Bartramia longicauda*); Horned Lark (*Eremophila alpestris*); Cassin's Sparrow (*Peucaea cassinii*); Field Sparrow (*Spizella pusilla*); Lark Sparrow (*Chondestes grammacus*); Grasshopper Sparrow (*Ammodramus savannarum*); Dickcissel (*Spiza americana*); Eastern Meadowlark (*Sturnella magna*); and Western Meadowlark (*Sturnella neglecta*)], plus the obligate brood parasite Brown-headed Cowbirds (*Molothrus ater*) for which presence often depends 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; none are increasing in population (Sauer et al., 2017).

We supplemented our survey data for the 11 focal species with citizen science data from the eBird Reference Dataset (Munson et al., 2014) during the months of April, May, June, and July, to match the surveys we conducted. We used complete primary checklist data from 2013-2014 and excluded casual counts. Complete checklists contain all birds sighted by the observer; primary checklists are the main checklist submitted when more than one observer submitted checklists for the same observations. We restricted use of eBird samples to ≤4.3km and ≤3.1 hours to be comparable to our surveys. We used the point count center or the transect midpoint as the count location for our surveys to have comparable precision to eBird coordinates (Fink et al., 2010). Likewise, because some eBird sightings will be from similar locations, we used all replicates of our point counts and transects. Because some of our 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. The combined dataset contained 5422 complete checklists (158 transect sampling events, 613 point count sampling events, and 4651 eBird sampling events). Data points are shown in Fig. 1.

To partition training and evaluation datasets, the combined dataset was split randomly for each species using the createDataPartition function in the CARET package (Kuhn, 2017), which samples such that both training and evaluation splits have similar distributions of presence and absences.

Predictors

We used bioclimatic variables from WorldClim at 30 second resolution (Hijmans *et al.* 2005), conservation easement status (O'Connor, Jones, Boone, & Lauber, 1999), and land cover variables (USDA/NRCS - National Geospatial Center of Excellence, 2011) to predict bird distribution (Table S1). We also included effort (length of observation in distance and time) and time of day in the analysis to control for differences in bird activity and observer effort that may influence species checklists. 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). Although the 15 x 15 pixel unit is smaller than our 500 m cutoff, most sightings are from even larger areas with the maximum length being under 4.3 km, an area comparable to Fink et al. 2010. Additionally, using a neighborhood value centered at the location point still provides information about the neighborhood, whether or not it overlaps or surrounds the sighting. 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 tested for the effects of using coarser (lower resolution) rasters to see if matching predictor and response variable scale affected accuracy. This is applicable as lowering raster resolution could be a route to making potentially more accurate models available to more researchers and managers. We scaled our previously created predictor rasters from their native or previously-resampled 30 m resolution to the approximate scale of our largest response data resolution, by decreasing cell size 144-fold to 4.32 km using means in the 'aggregate' function in the R package RASTER (Hijmans, 2016). Using these coarser predictor sets trimmed the 2013-2014 dataset slightly down to 5327 checklists (2664 for training and 2663 for evaluation).

Species distribution models

We ran models on Amazon Web Services (AWS) Elastic Cloud Computing (EC2) m4.4xlarge instances (16 vCPU and 64 GiB memory).

Base model

To create all species distribution models, we used random forest regression 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, Graber, McBurney, & Balasubramanian, 2010). Random Forests are suitable for species distribution models (Lorena et al., 2011; Prasad, Iverson, & Liaw, 2006) even with few presence records (Mi, Huettmann, Guo, Han, & Wen, 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 (Prasad et al., 2006). All random forests (both support set and statewide models) were generated with 500 trees which is generally suitable to achieve stability and accuracy (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).

Maps were created using the predict function in RASTER at the resolution of the original predictor datasets (30m and 4.32 km). For the maps, we assumed a uniform effort and time of day by creating constants for prediction: mean effort (distance and time) and time of day rasters. Thus, all predicted distribution models are generated assuming survey effort does not vary geographically and survey effort is 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). Prediction values for evaluation did not use these constants.

Statewide and SEE models

We created four models for each species at varying spatial scales: a statewide model and three SEEMs. The statewide model allowed us to compare the performance to SEEMs. A random forest model was created for the statewide scale for each species using all training data. The three remaining models are at varying support set scales, with some modifications from Fink et al. (2010). 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. 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) sampled 63% of each support set to imitate bootstrapping sampling, but because each of our support sets was being bootstrapped by the random forest algorithm, we used the full data set for each support set region.

Building a SEEM consists of creating random support sets, generating trees and predictions for each support set, and then combining each support set model predictions into the final overall prediction. We created stratified random points in the study area to create support sets (Fig. 2). The randomization of the support set center is important to produce useful ensemble models (Kuncheva & Whitaker, 2003). We used the 'spsample' function from the R package SP (Bivand, Pebesma, & Gomez-Rubio, 2013; Pebesma & Bivand, 2005) and created squares of size small (100 boxes of 120 x 120 km), medium (37 boxes of 200 x 200 km), or large (12 boxes of 450 x 450 km) around these points, which resulted in no significant difference in pixel coverage ($F_{2,147}$=0.63, p=0.53; small mean: 6.9, median 7, range 2-10; medium mean: 6.3, median 7, range 2-11; large mean: 6.6, median 7, range 2-10) before removing support sets with too few (<25) or uniform (all presence or all absence) checklists (models cannot run with uniform values). Each support set included all checklists from the training dataset located within its boundaries. All remaining support set rasters for a given scale (small, medium, or large) were combined into one larger raster using the RASTER mosaic() function to get the mean value of each pixel (ranging from 0 to 1), creating the spatially explicit ensemble (Fink et al., 2010; Hastie et al., 2001; Oppel et al., 2012) made of regional random forest predictions. This process was repeated at the three spatial scales, resulting in three SEEMs per species.

Model evaluation and error

To evaluate model performance, we created a statewide grid of 10 x 10 km cells and randomly sampled no more than 10 observations from each grid cell for spatial uniformity (Fink et al., 2010) using the held back data. The actual presence or absence from each checklist is compared to predicted values at each cell with the date and time of the sighting (instead of the uniform date and time used to create the maps). These sampling grid cells are larger than either predictor size and are used to ensure that we do not weight the accuracy of the models towards regions with more reports or surveys. We repeated the evaluation sampling 50 times to create a performance distribution for each model and error type (Fink et al. 2010). We noted the scale (small, medium, large, statewide) with best performance measures for each species and compared performance with notched box plots (Chambers, Cleveland, Kleiner, & Tukey, 1983) 1983)

Performance measures were root mean square error (RMSE) and area under the 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; a larger value indicates the model deviates further from expected (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, Witten, Hastie, & Tibshirani, 2013).

To compare computing efficiency, we used the R package MICROBENCHMARK to measure runtimes. All runtimes included RANDOMFOREST trees and RASTER prediction; ensembles also included mosaic creation time. We compared runtimes with a ratio of scaled model runtime to statewide model runtime as computational times will differ by the user's available machines.

Results

Current statewide distributions are shown in panel (a) of Figs. 3-6 and Figs. S1-7. SEEMs took 2.7-12.7 times longer (with fine resolution predictors) or 2.6-22.7 times longer (with coarse resolution predictors) to run than a statewide model, depending on species.

SEEMs outperformed statewide models for only Northern Bobwhite and Western Meadowlark within each data resolution for both AUC (Fig. 7) and RMSE (Fig. 8). Statewide models outperformed or equaled SEEMs within each data resolution for Brown-headed Cowbird and Dickcissel for both AUC and RMSE.

Coarse resolution models consistently outperformed fine resolution models in both AUC and RMSE for Dickcissel. Fine resolution models consistently outperformed coarse resolution models in both AUC and RMSE for Lark Sparrow, Grasshopper Sparrow, and Eastern Meadowlark.

The remaining species' best model (statewide or a SEEM) differed between resolutions or with choice of error evaluation.

Discussion

Although SEEMs increase model accuracy over continental scales (Fink et al., 2013, 2010), our study found their performance differed by species and predictor resolution even in a state with variable climate and diverse ecoregions. Two species were often better represented by SEEMs, suggesting their distributional processes may vary regionally. There were few obvious commonalities among these species that would lead to SEEMs being more accurate for them. One species is non-passerine (Northern Bobwhite), and the other is a common grassland passerine (Western Meadowlark). Two species were always better with statewide models (Brown-headed Cowbird and Dickcissel). The cowbird is strongly dependent on habitat structure (Benson et al., 2013; Bernath-Plaisted, Nenninger, & Koper, 2017), but these variables are not what is measured by the predictor layers that we used. Dickcissel is known for its semi-nomadic movement patterns (Temple, 2002); as such, neither species may be as dependent on local climatic variation mapped by the BioClim predictor inputs. The inconsistencies in the remainder of the species suggest that a larger sample of species and predictor resolutions is needed to compare why models are appropriate for given situations. On our original models, the predictors are consistently finer-scaled (30 m) than some, but not all, response location data (ranging from exact point count locations to aggregate sightings along a 4.3 km transect). However, Fink *et al.* (2010) used transects almost twice as long as ours (up to 8.1 vs 4.3 km) with 30 m resolution predictor data, so that should not account for differences between our results.

A potential mechanism for variation between species 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). It could also be that species-specific processes determine whether SEEMs are required. However, one benefit of random forest models and other machine learning methods is minimal tuning and expert opinion required to generate an accurate map (Fink et al., 2010). Requiring researchers to choose spatial scale based on expert opinion of variable importance negates this benefit. However, the fact that most species showed different model performance based on whether we used fine or coarse predictor resolution suggests that model performance depends at least partially on dataset resolutions.

An alternative approach for modelers seeking increased accuracy is the use of non-spatially explicit ensemble models, where different base models (predicting for the whole study area) are combined to produce a single prediction map (Araújo & New, 2007; Oppel et al., 2012). We recommend this approach as more efficient for regional managers. Multiple maps will still be generated for the whole study area (n = number of base models used), but typically fewer than the number of support sets created in a SEEM or STEM. These types of ensembles are known to increase accuracy relative to a single base model (Araujo & New 2007; Oppel *et al.* 2012). Although large-scale solutions to conserve grasslands are needed (Samson, Knopf & Ostlie 2004), local and regional conservation and management efforts also have critical impacts (Brennan, Kuvlesky & Morrison 2005). We expected that SEEMs would be most accurate and therefore relevant to wildlife management in a state with diverse ecotypes. However, based on our study, we recommend that when using a single base model type, all distribution model types should be run (statewide and at least one or more scales of SEEM) if computing capacity is available.

Accurate species distribution models can help us understand what factors, both environmental and land use, drive species declines (Elith & Leathwick 2009), but we need to conduct modeling with predictors and responses at the appropriate spatial scale. Further research is needed to elucidate at what study scale and data resolution SEEMs become appropriate. In fact, we found a modern laptop or desktop unable to handle fine resolution SEEMs and turned to cloud computing to complete them, so the length of time and computing expense involved can be substantial. Coarser predictor models were much quicker to run (less than an hour of increase relative to statewide models on the high-speed cloud computing), but they were still many times longer in runtime than the comparable statewide model. At the continental and temporally fine-grained scales, Fink *et al.* (2010)'s result still stands; it is at intermediate scales where more research is needed.

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 U.S. Department of Agriculture NIFA grant 2013-67009-20369 to ESB and supported by the AWS Cloud Credits for Research program. Our survey data were collected 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. Auer, T. Fagin, W. Hochachka, W.T. Honeycutt, A. Johnston, Steffen Oppel, M.A. Patten, one anonymous reviewer, and five anonymous reviewers from University of Cambridge journal club provided discussion, technical assistance, and comments. CMC was also supported by National Science Foundation grants IDBR 1014891 and ABI 1458402 to ESB and Oklahoma Department of Wildlife Conservation grant F17AF01294 (W-194-R-1) to M.A. Patten. The authors have no conflicts of interest to declare.

Data accessibility

Model code and survey data will be archived on datadryad.org upon acceptance. eBird data are available from eBird.org (Munson et al., 2014).

References

Alward, R.D. (1999) Grassland vegetation changes and nocturnal global warming. *Science*, **283**, 229–231.

Araújo, 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. *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.

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

Brennan, L.A., Kuvlesky, W.P. & Morrison. (2005) 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.

Davis, S.K. (2005) Nest-Site Selection Patterns and the Influence of Vegetation on Nest Survival of Mixed-Grass Prairie Passerines. *The Condor*, **107**, 605–616.

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.

Ethier, D.M., Koper, N. & Nudds, T.D. (2017) Spatiotemporal variation in mechanisms driving regional-scale population dynamics of a Threatened grassland bird. *Ecology and Evolution*.

Fink, D., Auer, T., Ruiz-Gutierrez, V., Hochachka, W.M., Johnston, A., Sorte, F.A.L. & Kelling, S. (2018) Modeling Avian Full Annual Cycle Distribution and Population Trends with Citizen Science Data. *bioRxiv*, 251868.

Fink, D., Damoulas, T. & Dave, J. (2013) Adaptive Spatio-Temporal Exploratory Models: Hemisphere-wide species distributions from massively crowdsourced eBird data. *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, p. 7. AAAI Press, Palo Alto, CA.

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.

Hamer, T.L., Flather, C.H. & Noon, B.R. (2006) Factors associated with grassland bird species richness: the relative roles of grassland area, landscape structure, and prey. *Landscape Ecology*, **21**, 569–583.

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.

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*, **25**, 42–50.

Hijmans, R.J. (2016) *Raster: Geographic Data Analysis and Modeling*. R package version 2.5-8. Available at: https://CRAN.R-project.org/package=raster.

431 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.

434 Hill, J.M., Egan, J.F., Stauffer, G.E. & Diefenbach, D.R. (2014) Habitat availability is a more plausible explanation than insecticide acute toxicity for U.S. grassland bird species dDeclines ed R.M. Brigham. *PLoS ONE*, **9**, e98064.

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

439 Johnston, A., Fink, D., Reynolds, M.D., Hochachka, W.M., Sullivan, B.L., Bruns, N.E., Hallstein, E., Merrifield, M.S., Matsumoto, S. & Kelling, S. (2015) Abundance models improve spatial and temporal prioritization of conservation resources. *Ecological Applications*, **25**, 1749–1756.

443 Kearney, M.R., Simpson, S.J., Raubenheimer, D. & Kooijman, S.A.L.M. (2013) Balancing heat, water and nutrients under environmental change: a thermodynamic niche framework ed S. Chown. *Functional Ecology*, **27**, 950–966.

446 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.

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

451 Kuhn, M. (2017) *Caret: Classification and Regression Training.* R package version 6.0-76. Available at: https://CRAN.R-project.org/package=caret.

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

455 Kuncheva, L.I. & Whitaker, C.J. (2003) Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning*, **51**, 181–207.

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, J.P. (2001) Ecological consequences of recent climate change. *Conservation Biology*, **15**, 320–331.

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 Climatological Survey. (2017) 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.

Pipher, E.N., Curry, C.M. & Koper, N. (2016) Cattle Grazing Intensity and Duration Have Varied Effects on Songbird Nest Survival in Mixed-Grass Prairies. *Rangeland Ecology & Management*, **69**, 437–443.

Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological orediction. *Ecosystems*, **9**, 181–199.

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

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.

US Bureau of Economic Analysis. (2016) Gross domestic product (GDP) by state (millions of current dollars), https://www.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=10&isuri=1

&7003=200&7035=-1&7004=naics&7005=4&7006=40000&7036=-1&7001=1200&7002=1&7090=70&7007=2014,2013&7093=levels

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. (2011) 2011 National Land Cover Dataset (NLCD): Oklahoma.

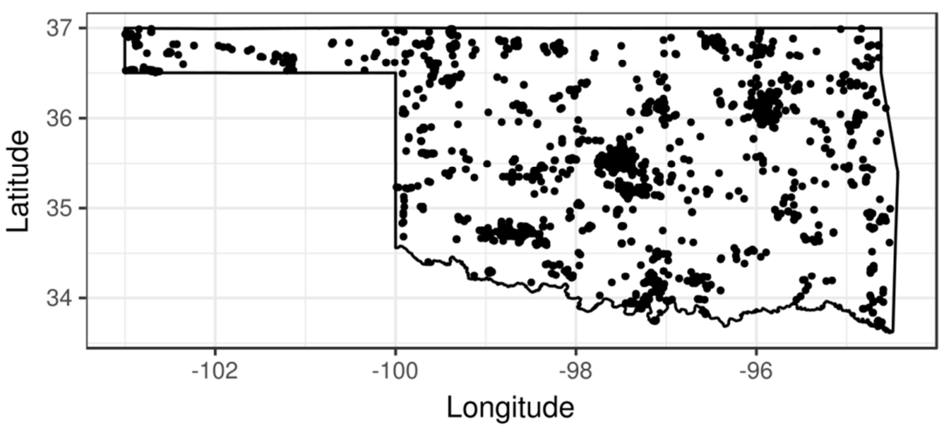
Winter, M., Johnson, D.H. & Shaffer, J.A. (2005) Variability in Vegetation Effects on Density and Nesting Success of Grassland Birds. *The Journal of Wildlife Management*, **69**, 185–197.

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.

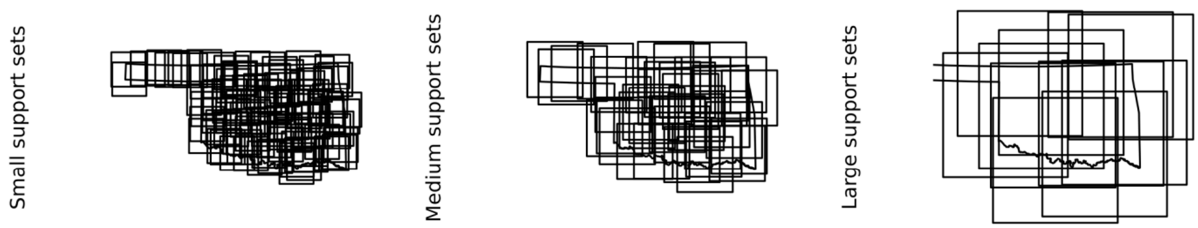
Zuckerberg, B., Fink, D., La Sorte, F.A., Hochachka, W.M. & Kelling, S. (2016) Novel seasonal land cover associations for eastern North American forest birds identified through dynamic species distribution modelling. *Diversity and Distributions*, **22**, 717–730.

Figures

Fig. 1. The complete dataset used in this study from eBird and surveys by the authors in 2013 and 2014 in the central U.S. state of Oklahoma in the Great Plains. The dataset was sampled such that half each were used for model training and model evaluation.

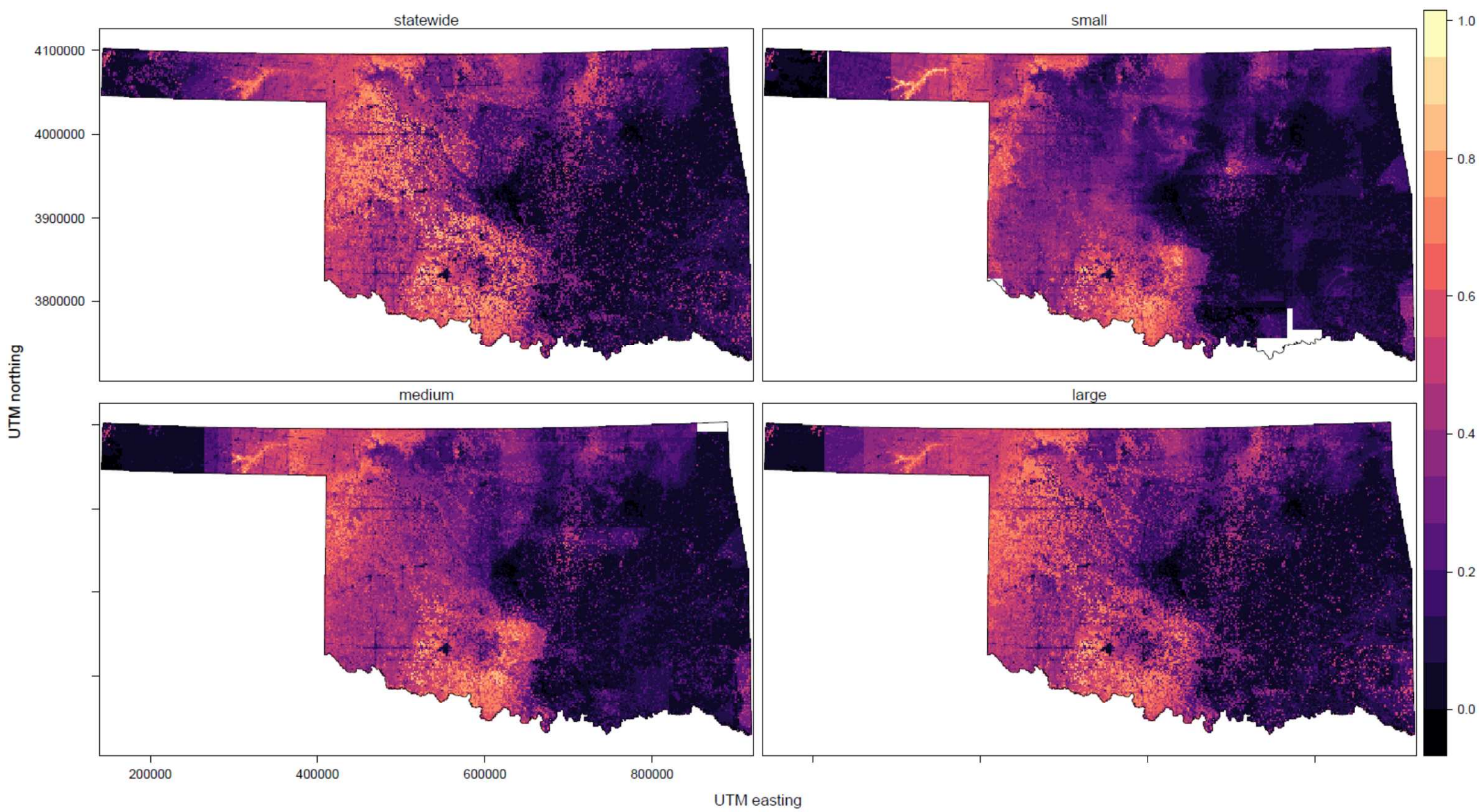


530 Fig. 2. Support sets of small (left), medium (middle), and large (right) scale overlaid over the
531 study area of Oklahoma, USA.

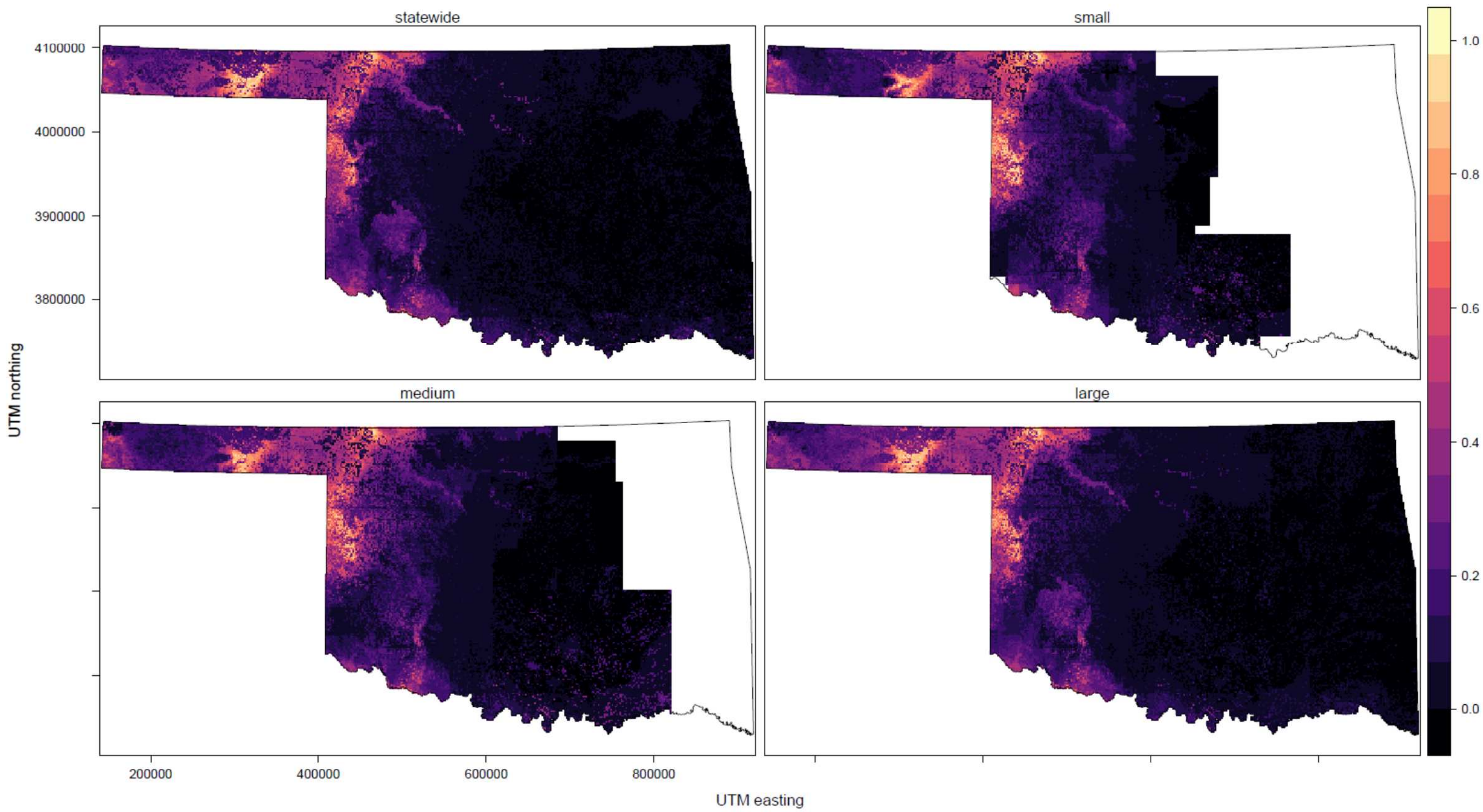


532

Fig. 3. Species distribution model for Northern Bobwhite generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

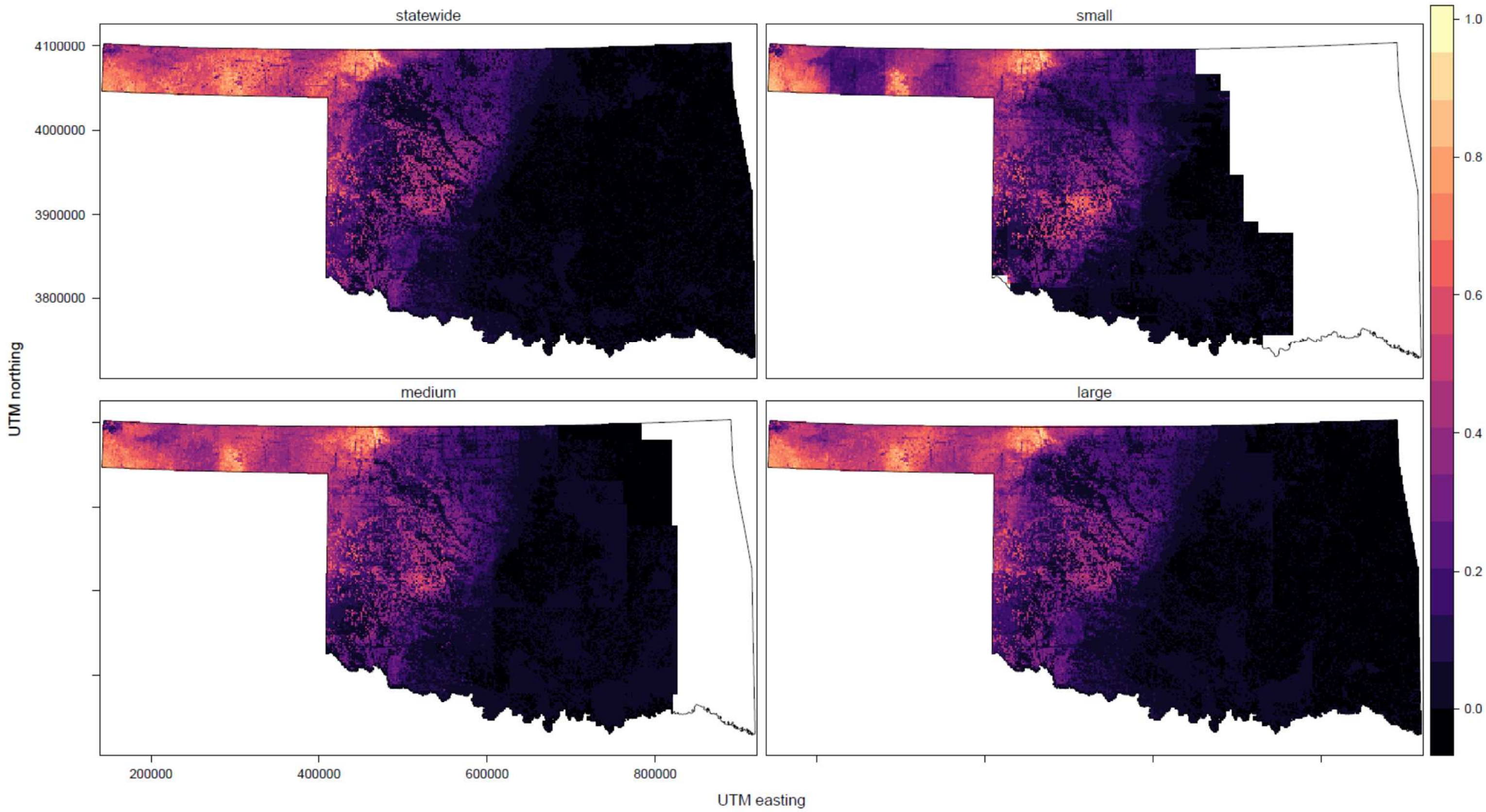


537 Fig. 4. Species distribution model for Cassin's Sparrow generated at four scales (statewide and three spatially explicit ensemble
538 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
539 occurrence from 0-1. Blank areas were not able to calculate a model.



540

541 Fig. 5. Species distribution model for Western Meadowlark generated at four scales (statewide and three spatially explicit ensemble
542 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
543 occurrence from 0-1. Blank areas were not able to calculate a model.



544

Fig. 6. Species distribution model for Brown-headed Cowbird generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

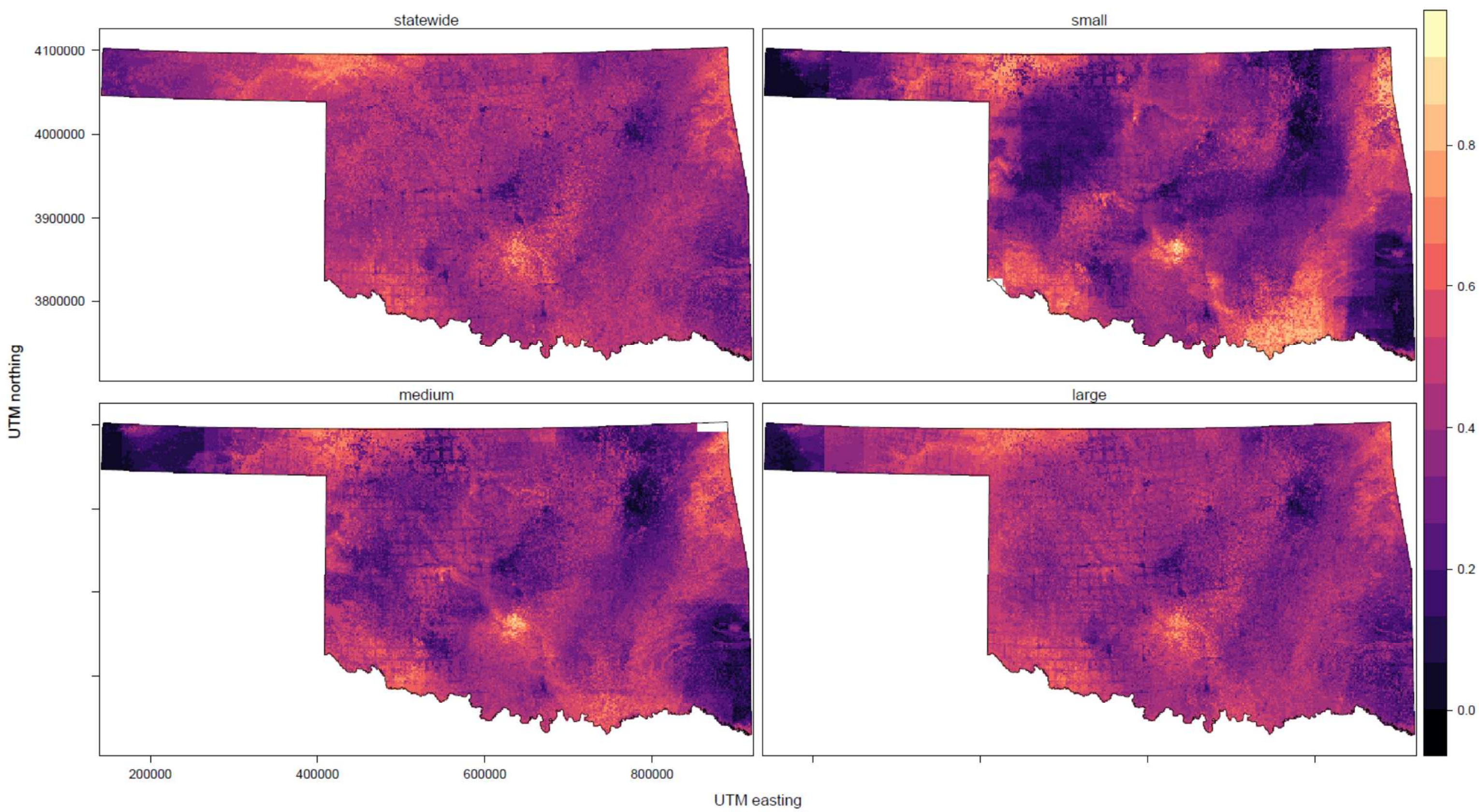


Fig. 7. RMSE evaluations for all 44 models compared by predictor resolution. Each panel shows one species. Overlapping notches on boxplots show no difference; non-overlapping notches show a significant difference in medians. Center line represents median. Fine grid lines are shown to facilitate notch comparison.

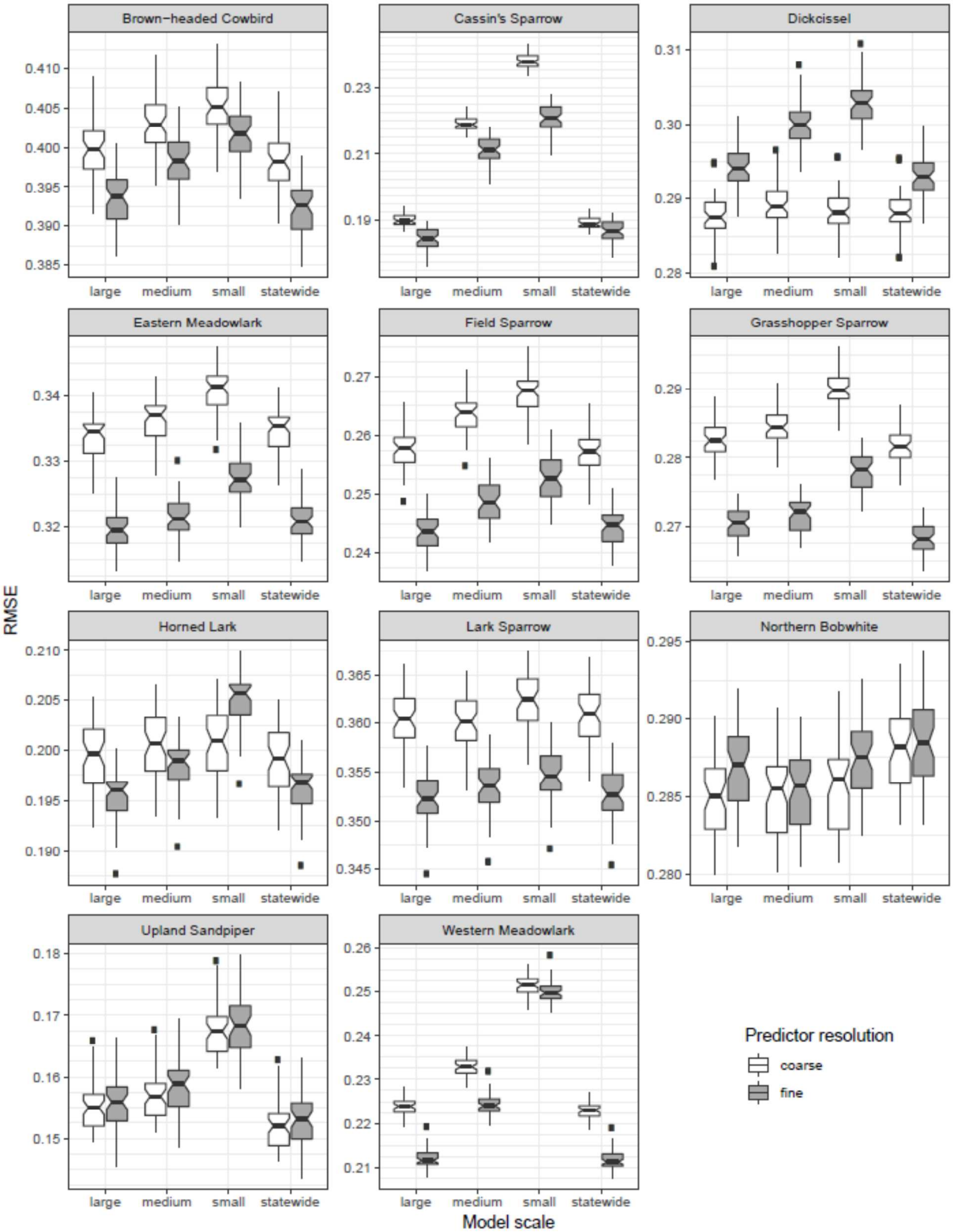
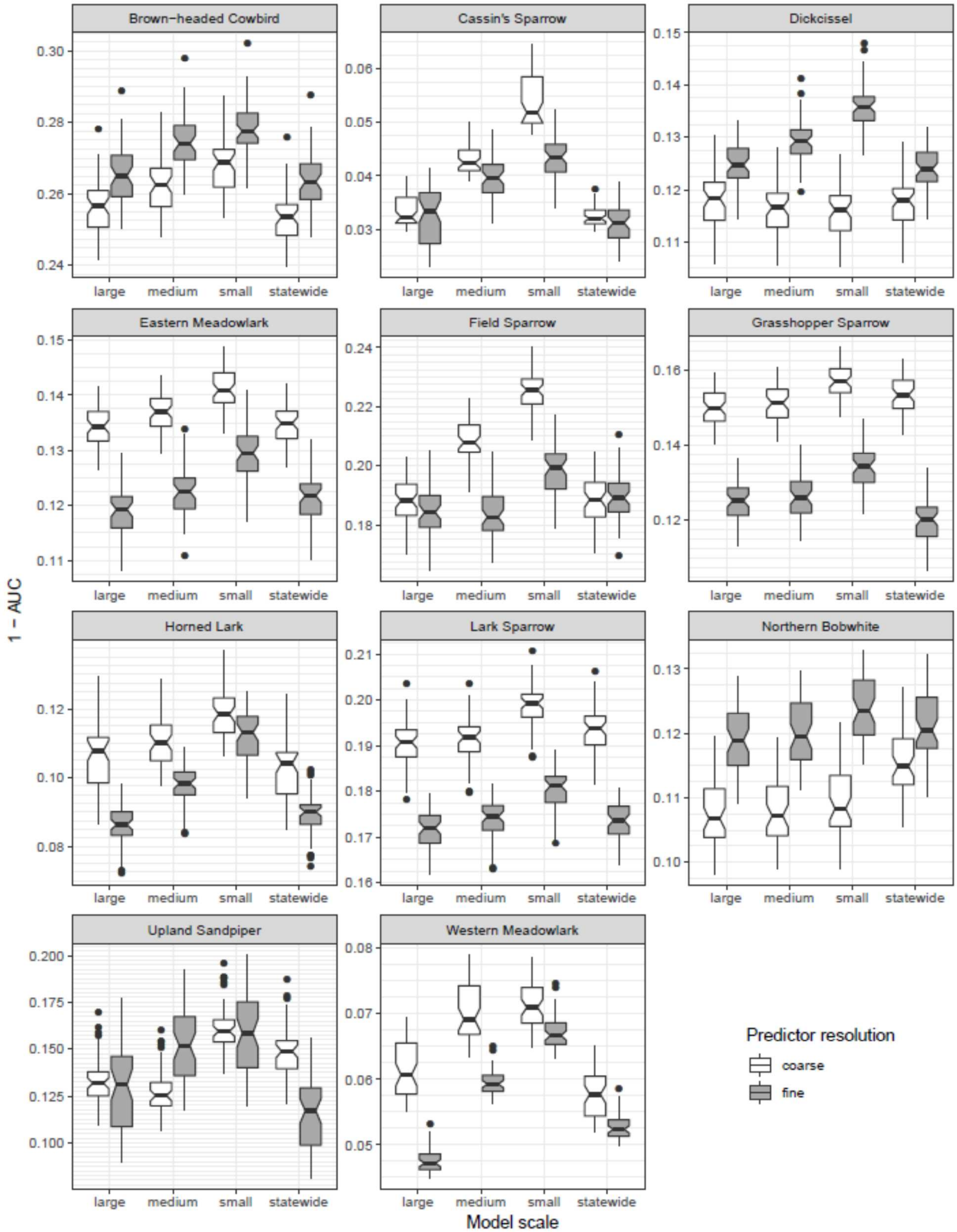

Brown−headed Cowbird
Cassin's Sparrow
Dickcissel
Eastern Meadowlark
Field Sparrow
Grasshopper Sparrow
Horned Lark
Lark Sparrow
Northern Bobwhite
Upland Sandpiper
Western Meadowlark
RMSE
large
medium
small
statewide
Model scale
Predictor resolution
coarse
fine

Fig. 8. AUC evaluations for all 44 models compared by predictor resolution. Each panel shows one species. AUC = 0.5, where prediction is random, and above which prediction is better than random. We show the y axis as 1 – AUC so that a lower value is better prediction to facilitate comparison with RMSE in Fig. 7. Overlapping notches on boxplots show no difference; non-overlapping notches show a significant difference in medians. Center line represents median. Fine grid lines are shown to facilitate notch comparison.


Brown−headed Cowbird
Cassin's Sparrow
Dickcissel
Eastern Meadowlark
Field Sparrow
Grasshopper Sparrow
Horned Lark
Lark Sparrow
Northern Bobwhite
Upland Sandpiper
Western Meadowlark
large
medium
small
statewide
1 − AUC
Model scale
Predictor resolution
coarse
fine

Supporting Information

Table S*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 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.	(USDA/NRCS - National Geospatial Center of Excellence 2011) Neighborhoods modified from NLCD landcover classes.
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	Developed open	

dev_openspace21_15cell_okmask	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 km^2	(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

Fig. S1. Species distribution model for Upland Sandpiper generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

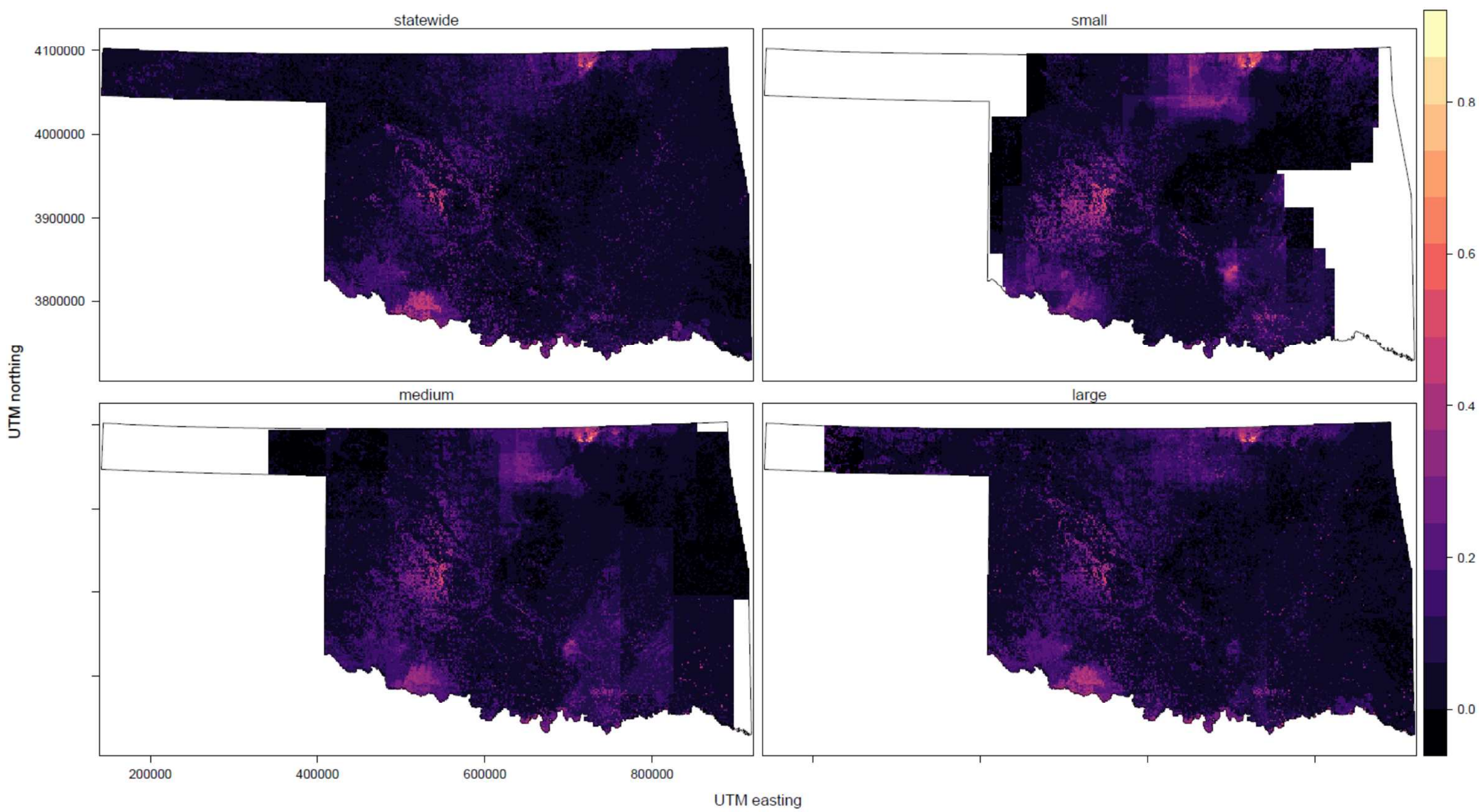


Fig. S2. Species distribution model for Horned Lark generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

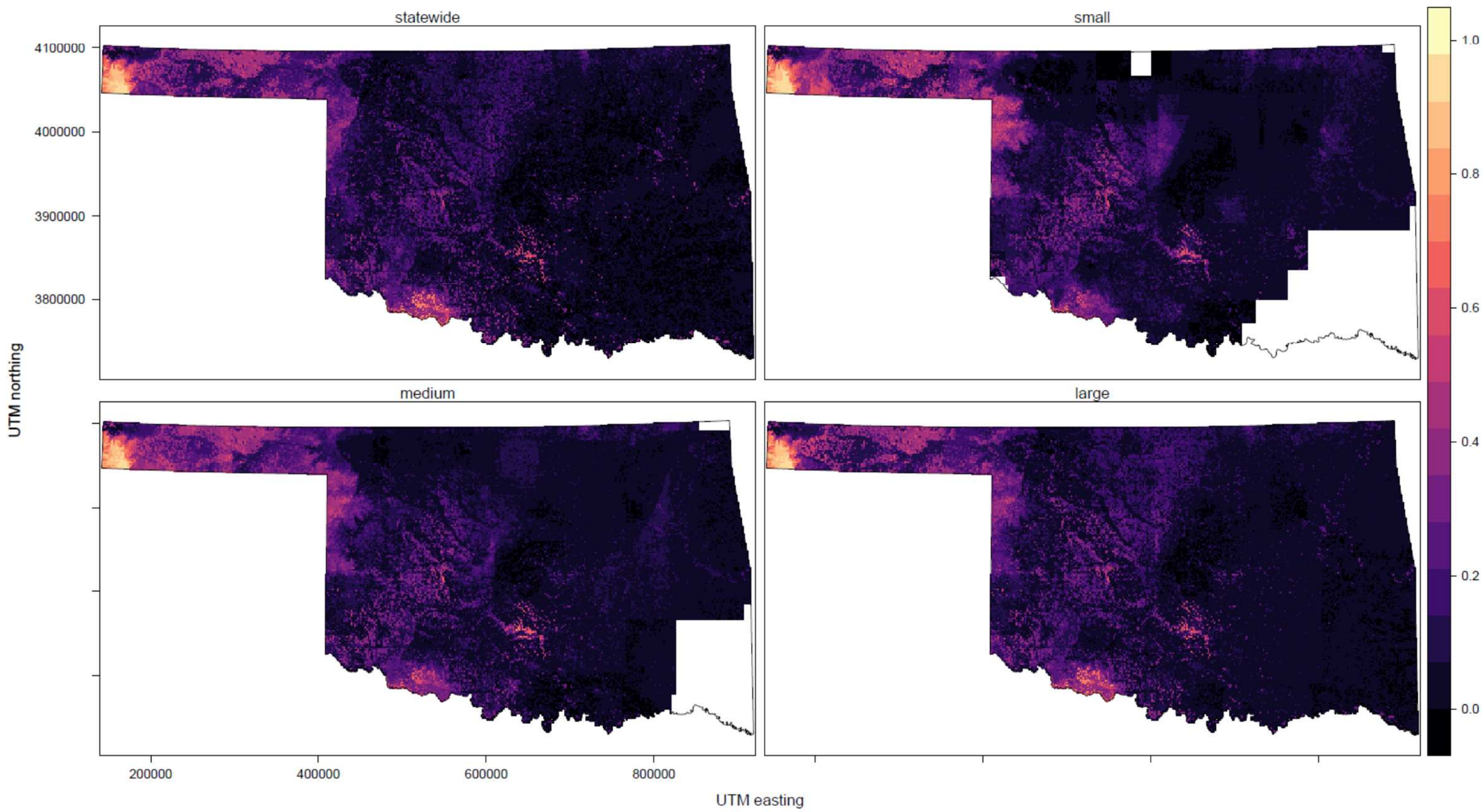


Fig. S3. Species distribution model for Field Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

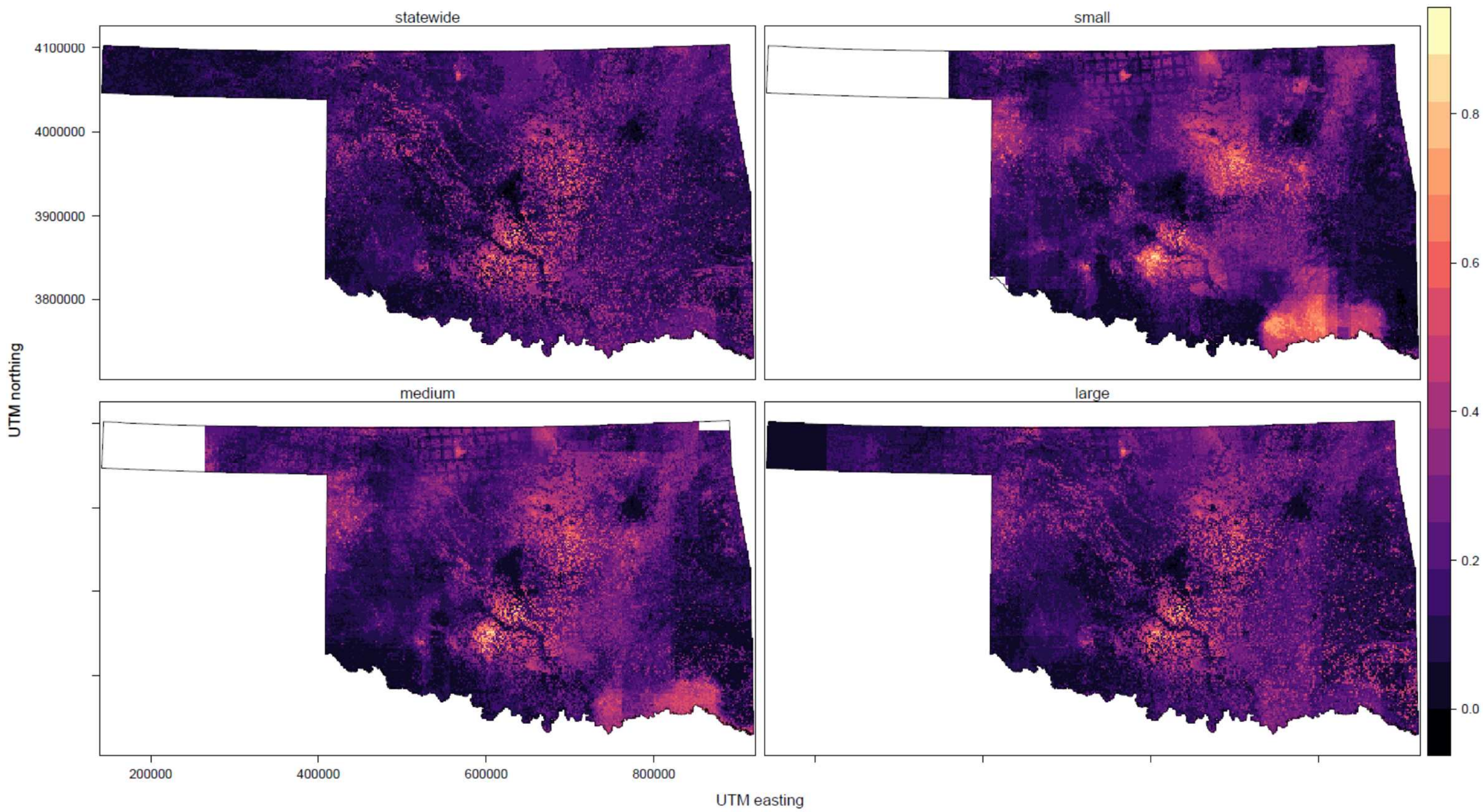


Fig. S4. Species distribution model for Lark Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

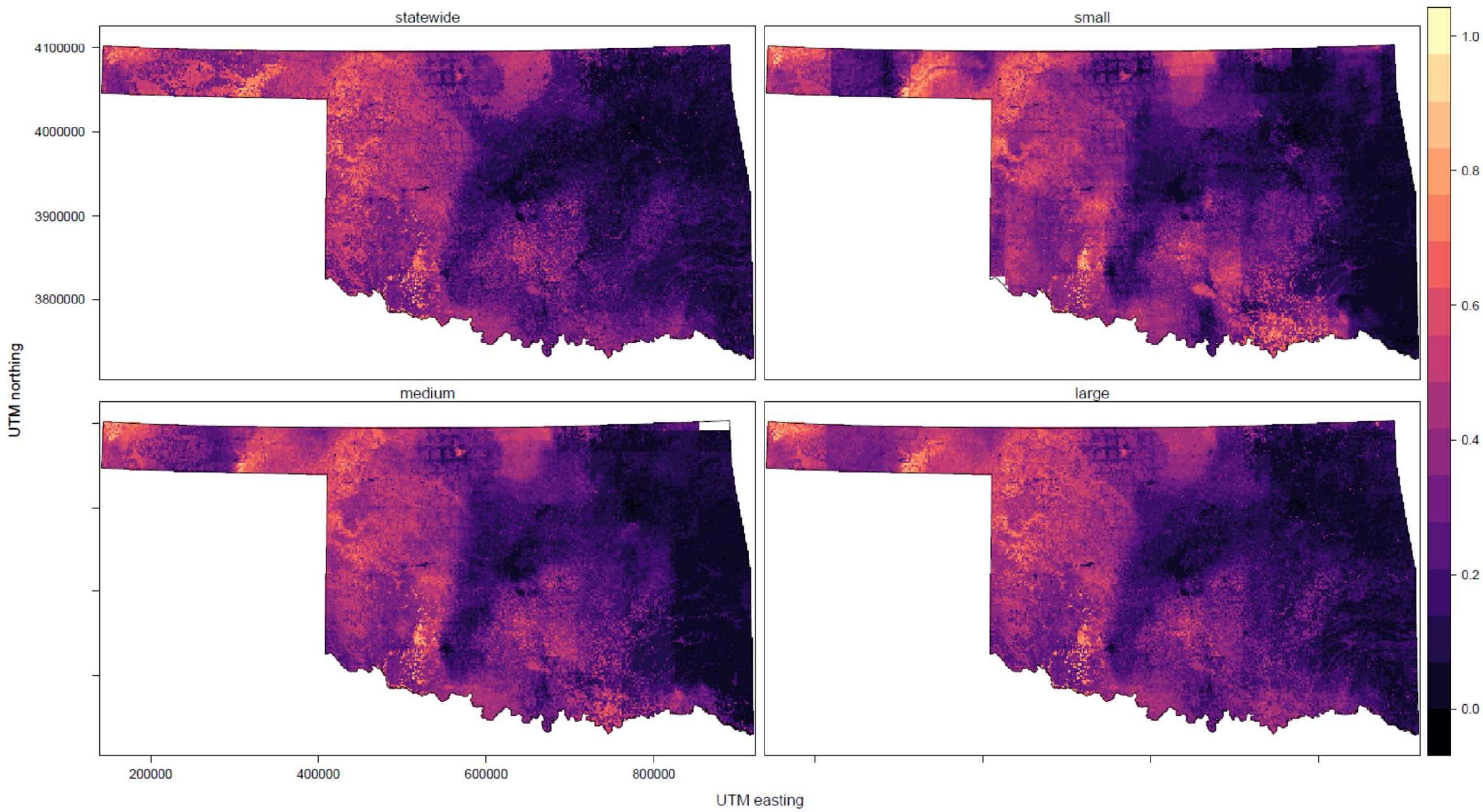


Fig. S5. Species distribution model for Grasshopper Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

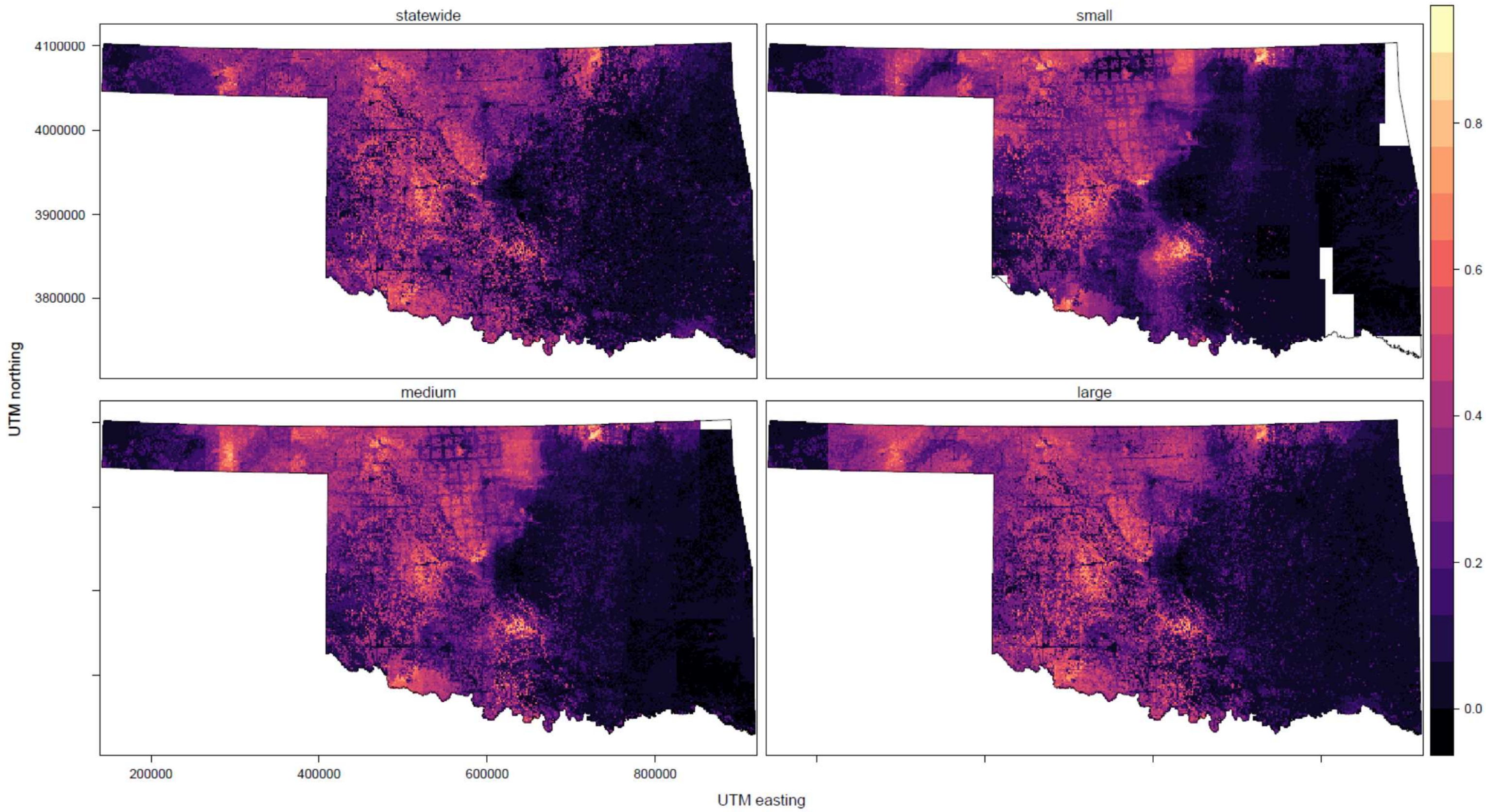


Fig. S6. Species distribution model for Dickcissel generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.

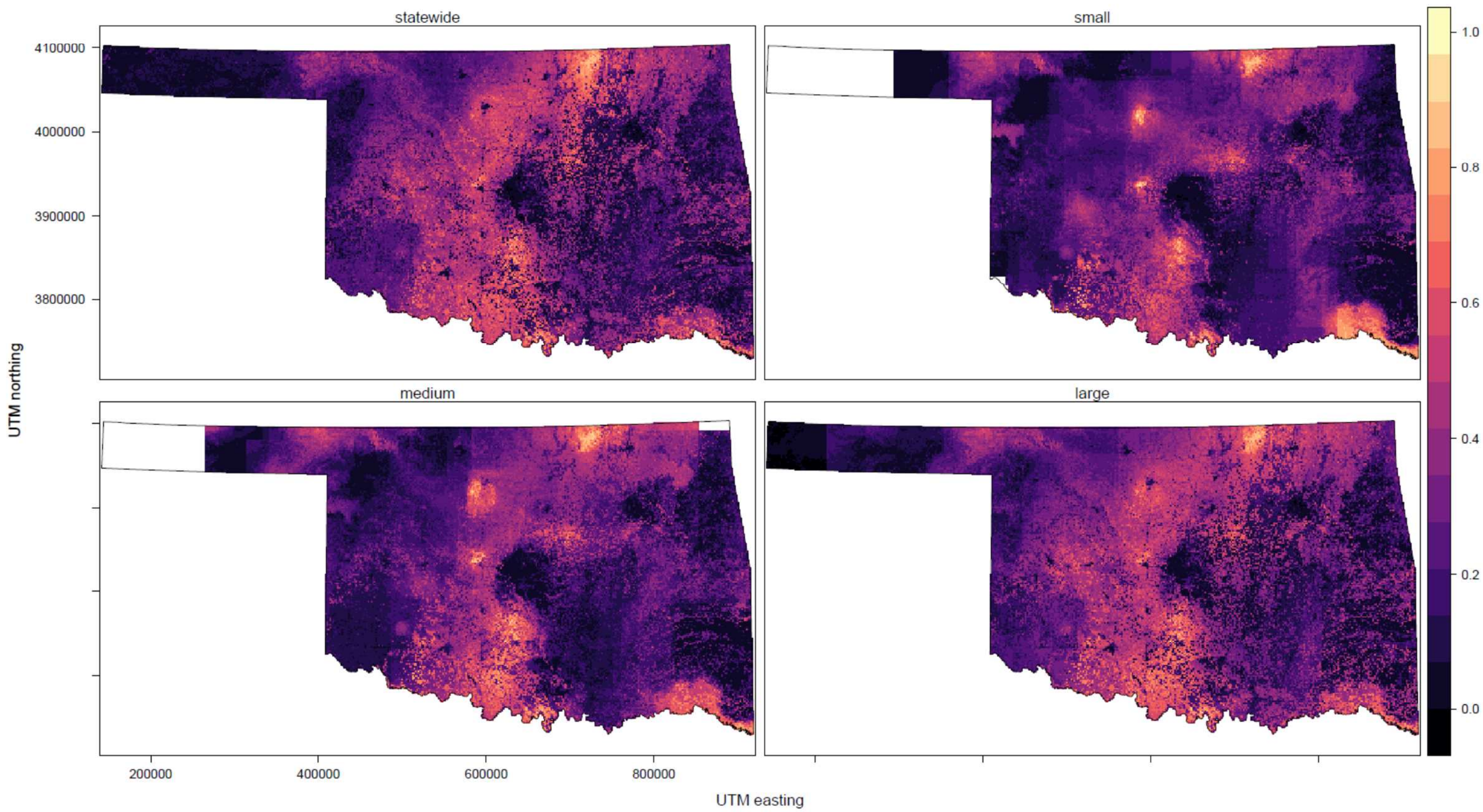


Fig. S7. Species distribution model for Eastern Meadowlark generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas were not able to calculate a model.