

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

Journal:	<i>Ecology and Evolution</i>
Manuscript ID	ECE-2018-08-00994.R1
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; University of Oklahoma, Corix Plains Institute 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.

SCHOLARONE™
Manuscripts

For Review Only

1
2 1 VARYING DATASET RESOLUTION ALTERS PREDICTIVE ACCURACY OF
3 2 SPATIALLY EXPLICIT ENSEMBLE MODELS FOR AVIAN SPECIES DISTRIBUTION
4
5

6 3 Claire M. Curry^{1,2}
7
8 4 Jeremy D. Ross²
9
10 5 Andrea J. Contina^{2,3}
11
12 6 Eli S. Bridge²
13
14

15 7 ¹University Libraries, University of Oklahoma, 401 W. Brooks St, Norman, OK, USA 73019
16
17

18 8 ²Oklahoma Biological Survey, University of Oklahoma, 111 E. Chesapeake Street, Norman, OK,
19
20 9 USA 73019
21
22

23 10 ³Corix Plains Institute, University of Oklahoma, Five Partners Place, 201 Stephenson Parkway,
24
25 11 Norman, OK, USA 73019
26
27

28 12
29
30

31 13 Corresponding author: Claire M. Curry; address: University Libraries, University of Oklahoma,
32
33 14 401 W. Brooks St, Norman, OK, USA 73019, email: cumcurry@ou.edu
34

35 15 Running title: Prediction accuracy of spatially explicit ensemble models
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

16

SUMMARY

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

34

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

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 predictor-response relationships 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 relative to the study area, with variation in spatial and temporal processes at scales intermediate to the study area and predictor resolution, 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

1
2
3 68 particular, the grassland birds of Oklahoma inhabit diverse grassland types and climatic
4 69 extremes. The open habitats of Oklahoma, which contains over a third of its land area as
5 70 grasslands and an additional 15% as croplands (Diamond & Elliott, 2015), contain grassland
6 71 birds characteristic of habitats ranging from southeastern pine savannahs to tallgrass, mixed-
7 72 grass, and shortgrass prairies (Askins et al., 2007; Diamond & Elliott, 2015). Grassland species
8 73 in areas half the size of Oklahoma in a single ecoregion have shown spatial and temporal
9 74 differences in variable importance (Ethier, Koper, & Nudds, 2017). Forest species, which
10 75 likewise occupy a single habitat type, also show spatial and temporal variation in predictor
11 76 importance (Zuckerberg et al., 2016). Similarly, such a technique has been used on shorebirds in
12 77 habitats with structural similarity to grasslands at a statewide scale (Johnston et al., 2015).
13 78 Finally, Oklahoma occurs on a strong east-west climatic gradient (Oklahoma Climatological
14 79 Survey 2017) that has had profound impacts on the ecosystems of the region (Kukal & Irmak,
15 80 2016; Seager et al., 2018). Physiological balances in animals can change in importance with
16 81 other environmental variables (Kearney, Simpson, Raubenheimer, & Kooijman, 2013), therefore
17 82 variable importance may be expected to change for at least some species across climatic
18 83 gradients. Oklahoma's grassland habitats, agricultural importance, and susceptibility to climate
19 84 change (Loarie et al., 2009; National Assessment Synthesis Team (U.S.), 2001) make it an ideal
20 85 and important region to test relative efficacy of different methods for modeling species
21 86 distributions.
22
23

24
25 87 Grasslands are one of the world's most endangered ecosystems, with declines of 82.6-99.9% of
26 88 tallgrass prairie, 30-99.9% of mixed-grass prairie, and 20-85.8% of short-grass prairie in the
27 89 plains states and provinces of North America (F. Samson & Knopf, 1994), and as such could
28 90 benefit from increased knowledge of distributional drivers. Drivers of decline include land use
29 91 conversion via agriculture and changes in fire and grazing regimes (Samson, Knopf & Ostlie
30 92 2004), although specifics vary by region (Askins et al., 2007). The already tenuous status of
31 93 grassland birds is further threatened by conversion to new crops resulting in permanent land use
32 94 changes (Wright & Wimberly, 2013), generational changes in land use (Higgins, Naugle, &
33 95 Forman, 2002), changes in conservation programs for grassland habitats (Klute, Robel, & Kemp,
34 96 1997), alterations to vegetation (Alward, 1999) and ecosystem structure (Brown, Valone, &
35 97 Curtin, 1997; Hamer, Flather, & Noon, 2006), and climate change (McCarty, 2001). Grassland
36 98 bird species are declining faster than other groups of birds (Askins et al., 2007; Hill, Egan,
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 99 Stauffer, & Diefenbach, 2014; Knopf, 1994) and continue to be imperiled by ongoing and
4 expanding threats to their habitat. Range-wide species distribution predictions have been made
5 for grassland birds but some species with smaller ranges are not accurately modeled (O'Connor
6 et al. 1999), perhaps because some drivers of distribution vary regionally (Askins et al., 2007;
7 Bakker, Naugle, & Higgins, 2002; Ethier et al., 2017), at a scale smaller than the study region.
8 Additionally, spatial and temporal variation in habitat needs and selection pressures (Davis 2005;
9 Winter, Johnson & Shaffer 2005) or interactions with weather events (Pipher, Curry, & Koper,
10 2016) are known to be important in grassland birds, therefore they are particularly suitable as a
11 106 testing ground for a spatially explicit approach to modeling.
12 107
13
14
15
16
17
18

19 108 The objectives of our study were threefold. First, we estimated the distribution of Oklahoma
20 grassland birds to understand current distribution statewide with standard species distribution
21 modeling methods. Next, these statewide current distribution predictions allowed us to compare
22 the statewide species distribution model for each species with SEEMs to evaluate whether this
23 approach is suitable at the scale of our region. Finally, we compared each approach's accuracy
24 when using fine- or coarse-resolution predictor sets. Although our approach is at a smaller scale
25 than originally envisioned for SEEMs, it is important to test their potential applicability at the
26 smaller scales at which most management decisions are made. Our results will allow others to
27 make decisions on whether increased accuracy in modeling is worth the additional computational
28 effort required by newer modeling techniques and provide guidance for future work into where
29 given modeling applications are useful.
30 114
31 115
32 116
33 117
34 118
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

119 METHODS

120 Study area

121 Oklahoma contains diverse vegetation and climate, making it a suitable region to examine effects
122 of spatially explicit models. There are ca. 165 vegetation types (based on soil and vegetation
123 composition) in 15 land cover types (Diamond & Elliott, 2015), with over a third of the
124 vegetation in grasslands. Rainfall and temperature vary across the state (Oklahoma
125 Climatological Survey, 2017), with annual precipitation ranging from ~ 43cm of rain in the
126 northwest to 142 cm in the southeast and mean annual temperature ranging from ~13°C in the
127 northwest to ~17°C in the southeast. Summer temperatures over 32°C can occur from 60-115

1
2
3 128 days out of the year varying statewide. Agriculture in Oklahoma is dominated by livestock
4 ranching and row crops (USDA/NASS, 2016), and accounted for over \$2.8 billion of the state's
5 gross domestic product in the study years (US Bureau of Economic Analysis, 2016); Oklahoma
6 ranks in the top 5 of US acreage for grain wheat and forage land for hay (USDA/NASS, 2016).
7
8
9
10
11 132 **Bird surveys**

12 133 We collected data 1-4 times each at 339 8-min roadside point counts (0.13 hr) and at 87 non-
13 roadside transects 0.3-3.1 hrs and 0.3-4.3 km long (mean \pm SD: 1.2 \pm 0.6 hrs and 1.8 \pm 0.8 km).
14 135 Each survey was conducted stationary (point counts) or walking at an even pace (transects). We
15 had 14 observers total (6 in 2013 and 8 in 2014). We only used sightings within 500 m of the
16 observer to preserve identification accuracy and recognize that detection is imperfect; however,
17 all models compared use similar data and as such it should not impact our comparison of models.
18 139 A zero (absence) or 1 (presence) was assigned for each combination of date and time and
19 species. We focused on 10 species of grassland birds found during our general surveys
20 [Northern Bobwhite (*Colinus virginianus*); Upland Sandpiper (*Bartramia longicauda*); Horned
21 Lark (*Eremophila alpestris*); Cassin's Sparrow (*Peucaea cassinii*); Field Sparrow (*Spizella*
22 *pusilla*); Lark Sparrow (*Chondestes grammacus*); Grasshopper Sparrow (*Ammodramus*
23 *savannarum*); Dickcissel (*Spiza americana*); Eastern Meadowlark (*Sturnella magna*); and
24 145 Western Meadowlark (*Sturnella neglecta*)], plus the obligate brood parasite Brown-headed
25 Cowbirds (*Molothrus ater*) for which presence often depends on land use factors (Benson,
26 Chiavacci, & Ward, 2013), for a total of 11 species. Many of these species are declining at the
27 state or North American level; none are increasing in population (Sauer et al., 2017).

28 149 We supplemented our survey data for the 11 focal species with citizen science data from the
29 eBird Reference Dataset (Munson et al., 2014) during the months of April, May, June, and July,
30 to match the surveys we conducted. We used complete primary checklist data from 2013-2014
31 and excluded casual counts. Complete checklists contain all birds sighted by the observer;
32 primary checklists are the main checklist submitted when more than one observer submitted
33 checklists for the same observations. We restricted use of eBird samples to \leq 4.3km and \leq 3.1
34 hours to be comparable to our surveys. We used the point count center or the transect midpoint
35 as the count location for our surveys to have comparable precision to eBird coordinates (Fink et
36 al., 2010). Likewise, because some eBird sightings will be from similar locations, we used all
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 158 replicates of our point counts and transects. Because some of our observers entered sightings
4 from before and during our surveys into eBird, we eliminated 14 counts from 2013 and 2014 that
5 were within two hours of the actual survey start time and within 15 km of the survey start
6 location. The combined dataset contained 5422 complete checklists (158 transect sampling
7 events, 613 point count sampling events, and 4651 eBird sampling events). Data points are
8 shown in Fig. 1.
9
10
11
12
13

14 164 To partition training and evaluation datasets, the combined dataset was split randomly for each
15 species using the `createDataPartition` function in the CARET package (Kuhn, 2017), which
16 samples such that both training and evaluation splits have similar distributions of presence and
17 absences.
18
19
20

21 168 **Predictors**

22 169 We used bioclimatic variables from WorldClim at 30 second resolution (Hijmans *et al.* 2005),
23 conservation easement status (O'Connor, Jones, Boone, & Lauber, 1999), and land cover
24 variables (USDA/NRCS - National Geospatial Center of Excellence, 2011) to predict bird
25 distribution (Table S1). We also included effort (length of observation in distance and time) and
26 time of day in the analysis to control for differences in bird activity and observer effort that may
27 influence species checklists. Neighborhood predictors were calculated by the values in
28 rectangular areas around each point, at the scale of 5 x 5 pixels (150 x 150 m) and 15 x 15 pixels
29 (450 x 450 m) (Fink *et al.* 2010). Although the 15 x 15 pixel unit is smaller than our 500 m
30 cutoff, most sightings are from even larger areas with the maximum length being under 4.3 km,
31 an area comparable to Fink *et al.* 2010. Additionally, using a neighborhood value centered at the
32 location point still provides information about the neighborhood, whether or not it overlaps or
33 surrounds the sighting. We looked at proportion of each land cover class and proportion of
34 summed open space land covers (grasslands, hay/pasture, cropland, herbaceous wetlands, and
35 barren land) since grassland bird occupancy can be influenced by the total non-structural cover
36 (McDonald 2017). Neighborhoods were created in QGIS 2.16 with the GRASS `r.neighbors`
37 processing tool (Quantum GIS Development Team, 2016).

38 185 We tested for the effects of using coarser (lower resolution) rasters to see if matching predictor
39 and response variable scale affected accuracy. This is applicable as lowering raster resolution
40 could be a route to making potentially more accurate models available to more researchers and
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 188 managers. We scaled our previously created predictor rasters from their native or previously-
4 resampled 30 m resolution to the approximate scale of our largest response data resolution, by
5 decreasing cell size 144-fold to 4.32 km using means in the ‘aggregate’ function in the R
6 package RASTER (Hijmans, 2016). Using these coarser predictor sets trimmed the 2013-2014
7 dataset slightly down to 5327 checklists (2664 for training and 2663 for evaluation).
8
9 192
10
11 193
12
13
14

15 194 **Species distribution models**

16
17 195 We ran models on Amazon Web Services (AWS) Elastic Cloud Computing (EC2) m4.4xlarge
18 instances (16 vCPU and 64 GiB memory).

19
20 197 *Base model*

21
22 198 To create all species distribution models, we used random forest regression trees (Breiman,
23 199 2001) in the R package RANDOMFOREST (Liaw & Wiener, 2002). Random forest gives results
24 competitive to those from other machine learning techniques such as boosted regression trees
25 and bagged decision trees (used in Fink et al. 2010 for the non-spatially-explicit comparison
26 model). Minimal tuning parameters are required (Caruana & Niculescu-Mizil, 2006; Cutler et
27 al., 2007; Guo, Graber, McBurney, & Balasubramanian, 2010). Random Forests are suitable for
28 species distribution models (Lorena et al., 2011; Prasad, Iverson, & Liaw, 2006) even with few
29 presence records (Mi, Huettmann, Guo, Han, & Wen, 2017). The random forest algorithm
30 bootstraps a subset of the data using only a set proportion of the predictor variables. It then
31 calculates the error rate on training data using the “out of bag” sample (the portion of data not
32 used in the bootstrap for each tree) (Hastie, Tibshirani, & Friedman, 2001). The trees are then
33 averaged for a final model (Prasad et al., 2006). All random forests (both support set and
34 statewide models) were generated with 500 trees which is generally suitable to achieve stability
35 and accuracy (Cutler et al., 2007). We used the default number of variables per bootstrap tree
36 (default ‘mtry’=the square root of the number of predictor variables) for all trees because this is
37 known to result in accurate predictions (Cutler et al., 2007).

38
39 214 Maps were created using the predict function in RASTER at the resolution of the original predictor
40 datasets (30m and 4.32 km). For the maps, we assumed a uniform effort and time of day by
41 creating constants for prediction: mean effort (distance and time) and time of day rasters. Thus,
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 217 all predicted distribution models are generated assuming survey effort does not vary
4 218 geographically and survey effort is typical for both surveys and citizen science efforts in 2013
5 219 and 2014 (mean time: 0.73 hr; mean distance: 0.75 km). The time of day raster for prediction
6 220 was given a value of 7:00am (Fink et al. 2010). Prediction values for evaluation did not use
7 221 these constants.
8
9
10
11
12 222 *Statewide and SEE Models*
13
14 223 We created four models for each species at varying spatial scales: a statewide model and three
15 224 SEEMs. The statewide model allowed us to compare the performance to SEEMs. A random
16 225 forest model was created for the statewide scale for each species using all training data. The
17 226 three remaining models are at varying support set scales, with some modifications from Fink et
18 227 al. (2010). First, the scale of our support sets reflects the state extent (i.e. our small, medium,
19 228 and large scales are relatively smaller than those needed for a continent-wide scale). As our
20 229 survey goals are to determine breeding distribution only, we used a broader temporal window
21 229 (April-July in all years) for our model. Secondly, for all base models, we used random forest
22 230 classification trees (Breiman, 2001) as described above. Finally, our geographic sampling of the
23 231 training and evaluation datasets, described in more detail in the next paragraph, reflects the
24 232 differing nature of our base models. Fink et al. (2010) sampled 63% of each support set to
25 233 imitate bootstrapping sampling, but we used the full data set for each support set region.
26
27
28
29
30
31
32
33
34
35
36 235 Building a SEEM consists of creating random support sets, generating trees and predictions for
37 236 each support set, and then combining each support set model predictions into the final overall
38 237 prediction. We created stratified random points in the study area to create support sets (Fig. 2).
39
40
41 238 The randomization of the support set center is important to fit ensemble models with low bias
42 239 and high variance (Kuncheva & Whitaker, 2003). We used the ‘spsample’ function from the R
43 240 package SP (Bivand, Pebesma, & Gomez-Rubio, 2013; Pebesma & Bivand, 2005) and created
44 241 squares of size small (100 boxes of 120 x 120 km), medium (37 boxes of 200 x 200 km), or large
45 242 (12 boxes of 450 x 450 km) around these points, which resulted in no significant difference in
46 243 pixel coverage ($F_{2,147}=0.63$, $p=0.53$; small mean: 6.9, median 7, range 2-10; medium mean: 6.3,
47 244 median 7, range 2-11; large mean: 6.6, median 7, range 2-10) before removing support sets with
48 245 too few (<25) or uniform (all presence or all absence) checklists (models cannot run with
49 246 uniform values). Using a larger number of base model pixel coverage is ideal to reduce
50
51
52
53
54
55
56
57
58
59
60

1
2
3 247 “blockiness” in final ensemble maps and prediction coverage, but we were limited by
4 computational costs. Each support set included all checklists from the training dataset located
5 within its boundaries. All remaining support set rasters for a given scale (small, medium, or
6 large) were combined into one larger raster using the RASTER mosaic() function to get the mean
7 value of each pixel (ranging from 0 to 1), creating the spatially explicit ensemble (Fink et al.,
8 2010; Hastie et al., 2001; Oppel et al., 2012) made of regional random forest predictions. This
9 process was repeated at the three spatial scales, resulting in three SEEMs per species.
10
11
12
13
14

15 254 *Model evaluation and error*
16
17

18 255 To evaluate model performance, we created a statewide grid of 10 x 10 km cells and randomly
19 sampled no more than 10 observations from each grid cell for spatial uniformity (Fink et al.,
20 2010) using the held back data. The actual presence or absence from each checklist is compared
21 to predicted values at each cell with the date and time of the sighting (instead of the uniform date
22 and time used to create the maps). These sampling grid cells are larger than either predictor size
23 and are used to ensure that we do not weight the accuracy of the models towards regions with
24 more reports or surveys. We repeated the evaluation sampling 50 times to create a performance
25 distribution for each model and error type (Fink et al. 2010). We noted the scale (small,
26 medium, large, statewide) with best performance measures for each species and compared
27 performance with notched box plots (Chambers, Cleveland, Kleiner, & Tukey, 1983).
28
29
30
31
32
33
34

35 265 Performance measures were root mean square error (RMSE) and area under the curve (AUC).
36 RMSE is calculated from the model residuals, taking the squared value of observed minus
37 expected values, then taking the square root to return to original units; a larger value indicates
38 the model deviates further from expected (Kuhn & Johnson, 2013). AUC is a summary of model
39 performance measuring how often the model misclassifies individual test observations; AUC
40 ranges from 0 to 1, with 1 being perfect and 0.5 being a model that performs no better than
41 random chance (Hanley & McNeil, 1982; James, Witten, Hastie, & Tibshirani, 2013).
42
43
44
45
46
47

48 272 To compare computing efficiency, we used the R package MICROBENCHMARK to measure
49 runtimes. All runtimes included RANDOMFOREST trees and RASTER prediction; ensembles also
50 included mosaic creation time. We compared runtimes with a ratio of scaled model runtime to
51 statewide model runtime as computational times will differ by the user’s available machines.
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9

RESULTS

276 Current statewide distributions are shown in panel (a) of Figs. 3-6 and Figs. S1-7. SEEMs took
277 2.7-12.7 times longer (with fine resolution predictors) or 2.6-22.7 times longer (with coarse
278 resolution predictors) to run than a statewide model, depending on species.

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

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

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

290
291
292
293
294
295
296
297
298
299
300
301
302
303
304

DISCUSSION

291 Although SEEMs increase model accuracy over continental scales (Fink et al., 2013, 2010), our
292 study found their performance differed by species and predictor resolution even in a state with
293 variable climate and diverse ecoregions. Two species were often better represented by SEEMs,
294 suggesting their distributional processes may vary regionally. There were few obvious
295 commonalities among these species that would lead to SEEMs being more accurate for them.
296 One species is non-passerine (Northern Bobwhite), and the other is a common grassland
297 passerine (Western Meadowlark). Two species were always better with statewide models
298 (Brown-headed Cowbird and Dickcissel). The cowbird is strongly dependent on habitat
299 structure (Benson et al., 2013; Bernath-Plaisted, Nenninger, & Koper, 2017), but these variables
300 are not what is measured by the predictor layers that we used. Dickcissel is known for its semi-
301 nomadic movement patterns (Temple, 2002); as such, neither species may be as dependent on
302 local climatic variation mapped by the BioClim predictor inputs. The inconsistencies in the
303 remainder of the species suggest that a larger sample of species and predictor resolutions is
304 needed to compare why models are appropriate for given situations. On our original models, the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 305 predictors are consistently finer-scaled (30 m) than some, but not all, response location data
4 (ranging from exact point count locations to aggregate sightings along a 4.3 km transect).
5
6 307 However, Fink *et al.* (2010) used transects almost twice as long as ours (up to 8.1 vs 4.3 km)
7 with 30 m resolution predictor data, so that should not account for differences between our
8 results.
9
10 309

11 310 A potential mechanism for variation between species includes whether species' distributions
12 depend more upon bioclimatic versus ecological variables, as bioclimatic variables should
13 change more smoothly over a larger area (potentially reducing the need for adaptive local
14 models). It could also be that species-specific processes determine whether SEEMs are required.
15 However, one benefit of random forest models and other machine learning methods is minimal
16 tuning and expert opinion required to generate an accurate map (Fink *et al.*, 2010). Requiring
17 researchers to choose spatial scale based on expert opinion of variable importance negates this
18 benefit. However, the fact that most species showed different model performance based on
19 whether we used fine or coarse predictor resolution suggests that model performance depends at
20 least partially on dataset resolutions. Researchers who suspect that a SEE model is appropriate
21 for their dataset and system can compare a small number of base models for different regions or
22 times and see if relationships vary among the test models.
23
24 317

25 322 An alternative approach for modelers seeking increased accuracy is the use of non-spatially
26 explicit ensemble models, where different base models (predicting for the whole study area) are
27 combined to produce a single prediction map (Araújo & New, 2007; Oppel *et al.*, 2012). We
28 recommend this approach as more efficient for regional managers. Multiple maps will still be
29 generated for the whole study area (n = number of base models used), but typically fewer than
30 the number of support sets created in a SEEM or STEM. These types of ensembles are known to
31 increase accuracy relative to a single base model (Araujo & New 2007; Oppel *et al.* 2012).
32 Although large-scale solutions to conserve grasslands are needed (Samson, Knopf & Ostlie
33 2004), local and regional conservation and management efforts also have critical impacts
34 (Brennan, Kuvlesky & Morrison 2005). We expected that SEEMs would be most accurate and
35 therefore relevant to wildlife management in this state with diverse ecotypes that occur at scales
36 larger than predictors but smaller than our study region. However, based on our study, we
37
38 333

1
2
3 334 recommend that when using a single base model type, all distribution model types should be run
4 335 (statewide and at least one or more scales of SEEM) if computing capacity is available.
5
6

7 336 Accurate species distribution models can help us understand what factors, both environmental
8 337 and land use, drive species declines (Elith & Leathwick 2009), but we need to conduct modeling
9 338 with predictors and responses at the appropriate spatial scale. Further research is needed to
10 339 elucidate at what study scale and data resolution SEEMs become appropriate. In fact, we found
11 340 a modern laptop or desktop unable to handle fine resolution SEEMs and turned to cloud
12 341 computing to complete them, so the length of time and computing expense involved can be
13 342 substantial. Coarser predictor models were much quicker to run (less than an hour of increase
14 343 relative to statewide models on the high-speed cloud computing), but they were still many times
15 344 longer in runtime than the comparable statewide model. At the continental and temporally fine-
16 345 grained scales, Fink *et al.* (2010)'s result still stands; it is at intermediate scales where more
17 346 research is needed.
18
19

20 347 **AUTHORS' CONTRIBUTIONS**

21 348 ESB, JDR, AJC, and CMC conceived the ideas and designed methodology. ESB and JDR
22 349 collected data. CMC analyzed the data. All authors contributed critically to the drafts and gave
23 350 final approval for publication.
24
25

26 351 **ACKNOWLEDGEMENTS**

27 352 This work was funded by U.S. Department of Agriculture (USDA) NIFA grant 2013-67009-
28 353 20369 to ESB and supported by the AWS Cloud Credits for Research program. Our survey data
29 354 were collected under OU IACUC R12-019. We thank technicians M. Buschow, R. Feliciano, M.
30 355 Furst, J. Haughawout, J. Hightower, J. Kruk, C. Myers, K. Oliver, T. Smith, R. Soto, and J.
31 356 Tibbits for help in gathering survey data. T. Auer, T. Fagin, W. Hochachka, W.T. Honeycutt, A.
32 357 Johnston, Steffen Oppel, M.A. Patten, one anonymous reviewer, and five anonymous reviewers
33 358 from University of Cambridge journal club provided discussion, technical assistance, and
34 359 comments. CMC was supported by National Science Foundation (NSF) grants IDBR 1014891
35 360 and ABI 1458402 to ESB and Oklahoma Department of Wildlife Conservation grant
36 361 F17AF01294 (W-194-R-1) to M.A. Patten. AJC was supported by NSF grants IDBR 1014891,
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 362 DGE 1545261, and DEB 0946685 and by USDA grant NIFA-AFRI-003536. The authors have
4
5 363 no conflicts of interest to declare.
6
7
8 364 DATA ACCESSIBILITY
9
10 365 Model code and survey data will be archived on datadryad.org upon acceptance. eBird data are
11
12 366 available from eBird.org (Munson et al., 2014).
13
14
15 367 REFERENCES
16
17 368 Alward, R.D. (1999) Grassland vegetation changes and nocturnal global warming. *Science*, **283**,
18
19 369 229–231.
20
21
22 370 Araújo, M. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology
23 & Evolution*, **22**, 42–47.
24
25
26 372 Askins, R.A., Chávez-Ramírez, F., Dale, B.C., Haas, C.A., Herkert, J.R., Knopf, F.L. & Vickery,
27
28 373 P.D. (2007) Conservation of grassland birds in North America: understanding ecological
29
30 374 processes in different regions. *Ornithological Monographs*, **64**, iii-46.
31
32
33 375 Bakker, K.K., Naugle, D.E. & Higgins, K.F. (2002) Incorporating landscape attributes into
34
35 376 models for migratory grassland bird conservation. *Conservation Biology*, **16**, 1638–1646.
36
37
38 377 Beaumont, L.J., Pitman, A.J., Poulsen, M. & Hughes, L. (2007) Where will species go?
39
40 378 Incorporating new advances in climate modelling into projections of species
41
42 379 distributions. *Global Change Biology*, **13**, 1368–1385.
43
44
45 380 Benson, T.J., Chiavacci, S.J. & Ward, M.P. (2013) Patch size and edge proximity are useful
46
47 381 predictors of brood parasitism but not nest survival of grassland birds. *Ecological
48 Applications*, **23**, 879–887.
49
50
51 383 Bivand, R.S., Pebesma, E. & Gomez-Rubio, V. (2013) *Applied Spatial Data Analysis with R*,
52
53 384 2nd ed. Springer, New York.
54
55
56 385 Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.
57
58
59
60

- 1
2
3 386 Brennan, L.A., Kuvlesky, W.P. & Morrison. (2005) North American grassland birds: an
4
5 387 unfolding conservation crisis? *Journal of Wildlife Management*, **69**, 1–13.
6
7
8 388 Brown, J.H., Valone, T.J. & Curtin, C.G. (1997) Reorganization of an arid ecosystem in response
9
10 389 to recent climate change. *Proceedings of the National Academy of Sciences*, **94**, 9729–
11 390 9733.
12
13
14 391 Caruana, R. & Niculescu-Mizil, A. (2006) An empirical comparison of supervised learning
15
16 392 algorithms. *Proceedings of the 23rd International Conference on Machine Learning*, pp.
17
18 393 161–168. ACM.
19
20 394 Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J.
21
22 395 (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783–2792.
23
24 396 Davis, S.K. (2005) Nest-Site Selection Patterns and the Influence of Vegetation on Nest Survival
25
26 397 of Mixed-Grass Prairie Passerines. *The Condor*, **107**, 605–616.
27
28
29 398 Diamond, D.D. & Elliott, L.F. (2015) *Oklahoma Ecological Systems Mapping Interpretive
30 Booklet: Methods, Short Type Descriptions, and Summary Results*. Oklahoma
31
32 400 Department of Wildlife Conservation, Norman, OK.
33
34
35 401 Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J.,
36
37 402 Huettmann, F., R. Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A.,
38
39 403 Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T.,
40
41 404 Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams,
42
43 405 S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of
44
45 406 species' distributions from occurrence data. *Ecography*, **29**, 129–151.
46
47 407 Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and
48
49 408 prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*,
50
51 409 **40**, 677–697.
52
53 410 Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees.
54
55 411 *Journal of Animal Ecology*, **77**, 802–813.
56
57
58
59
60

- 1
2
3 412 Ethier, D.M., Koper, N. & Nudds, T.D. (2017) Spatiotemporal variation in mechanisms driving
4 413 regional-scale population dynamics of a Threatened grassland bird. *Ecology and*
5 414 *Evolution.*
- 6
7
8
9 415 Fink, D., Auer, T., Ruiz-Gutierrez, V., Hochachka, W.M., Johnston, A., Sorte, F.A.L. & Kelling,
10 416 S. (2018) Modeling Avian Full Annual Cycle Distribution and Population Trends with
11 417 Citizen Science Data. *bioRxiv*, 251868.
- 12
13
14
15 418 Fink, D., Damoulas, T. & Dave, J. (2013) Adaptive Spatio-Temporal Exploratory Models:
16 419 Hemisphere-wide species distributions from massively crowdsourced eBird data.
17 420 *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, p. 7.
18 421 AAAI Press, Palo Alto, CA.
- 19
20
21
22
23 422 Fink, D., Hochachka, W.M., Zuckerberg, B., Winkler, D.W., Shaby, B., Munson, M.A., Hooker,
24 423 G., Riedewald, M., Sheldon, D. & Kelling, S. (2010) Spatiotemporal exploratory models
25 424 for broad-scale survey data. *Ecological Applications*, **20**, 2131–2147.
- 26
27
28
29
30 425 Guo, Y., Gruber, A., McBurney, R.N. & Balasubramanian, R. (2010) Sample size and statistical
31 426 power considerations in high-dimensionality data settings: a comparative study of
32 427 classification algorithms. *BMC Bioinformatics*, **11**, 447.
- 33
34
35
36 428 Hamer, T.L., Flather, C.H. & Noon, B.R. (2006) Factors associated with grassland bird species
37 429 richness: the relative roles of grassland area, landscape structure, and prey. *Landscape*
38 430 *Ecology*, **21**, 569–583.
- 39
40
41
42 431 Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating
43 432 characteristic (roc) curve. *Radiology* 143: 29–36, **143**, 29–36.
- 44
45
46 433 Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The Elements of Statistical Learning*. Springer
47 434 series in statistics Springer, Berlin.
- 48
49
50
51 435 Higgins, K.F., Naugle, D.E. & Forman, K.J. (2002) A case study of changing land use practices
52 436 in the northern Great Plains, U.S.A.: an uncertain future for waterbird conservation.
53 437 *Waterbirds*, **25**, 42–50.
- 54
55
56
57
58
59
60

1
2
3 438 Hijmans, R.J. (2016) *Raster: Geographic Data Analysis and Modeling*. R package version 2.5-8.
4 439 Available at: <https://CRAN.R-project.org/package=raster>.
5
6
7
8
9
10
11
12
13

14 440 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution
15 441 interpolated climate surfaces for global land areas. *International Journal of Climatology*,
16 442 25, 1965–1978.
17
18
19

20 443 Hill, J.M., Egan, J.F., Stauffer, G.E. & Diefenbach, D.R. (2014) Habitat availability is a more
21 444 plausible explanation than insecticide acute toxicity for U.S. grassland bird species
22 445 dDeclines ed R.M. Brigham. *PLoS ONE*, 9, e98064.
23
24
25

26 446 James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An Introduction to Statistical Learning*.
27 447 Springer New York, New York, NY.
28
29
30
31

32 448 Johnston, A., Fink, D., Reynolds, M.D., Hochachka, W.M., Sullivan, B.L., Bruns, N.E.,
33 449 Hallstein, E., Merrifield, M.S., Matsumoto, S. & Kelling, S. (2015) Abundance models
34 450 improve spatial and temporal prioritization of conservation resources. *Ecological
35 451 Applications*, 25, 1749–1756.
36
37

38 452 Kearney, M.R., Simpson, S.J., Raubenheimer, D. & Kooijman, S.A.L.M. (2013) Balancing heat,
39 453 water and nutrients under environmental change: a thermodynamic niche framework ed
40 454 S. Chown. *Functional Ecology*, 27, 950–966.
41
42
43

44 455 Klute, D.S., Robel, R.J. & Kemp, K.E. (1997) Will Conversion of Conservation Reserve
45 456 Program (CRP) Lands to Pasture be Detrimental for Grassland Birds in Kansas?
46 457 *American Midland Naturalist*, 137, 206.
47
48

49 458 Knopf, F.L. (1994) Avian assemblages on altered grasslands. *Studies in Avian Biology*, 15, 247–
50 459 257.
51
52

53 460 Kuhn, M. (2017) *Caret: Classification and Regression Training*. R package version 6.0-76.
54 461 Available at: <https://CRAN.R-project.org/package=caret>.
55
56

57 462 Kuhn, M. & Johnson, K. (2013) *Applied Predictive Modeling*. Springer New York, New York,
58 463 NY.
59
60

- 1
2
3 464 Kuncheva, L.I. & Whitaker, C.J. (2003) Measures of Diversity in Classifier Ensembles and Their
4 465 Relationship with the Ensemble Accuracy. *Machine Learning*, **51**, 181–207.
5
6
7 466 Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. *R News*, **2**, 18–22.
8
9
10 467 Lipsey, M.K., Doherty, K.E., Naugle, D.E., Fields, S., Evans, J.S., Davis, S.K. & Koper, N.
11 468 (2015) One step ahead of the plow: Using cropland conversion risk to guide Sprague's
12 469 Pipit conservation in the northern Great Plains. *Biological Conservation*, **191**, 739–749.
13
14
15
16 470 Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The
17 471 velocity of climate change. *Nature*, **462**, 1052–1055.
18
19
20
21 472 Lorena, A.C., Jacintho, L.F.O., Siqueira, M.F., Giovanni, R.D., Lohmann, L.G., de Carvalho,
22 473 A.C.P.L.F. & Yamamoto, M. (2011) Comparing machine learning classifiers in potential
23 474 distribution modelling. *Expert Systems with Applications*, **38**, 5268–5275.
24
25
26
27 475 McCarty, J.P. (2001) Ecological consequences of recent climate change. *Conservation Biology*,
28 476 **15**, 320–331.
29
30
31 477 Mi, C., Huettmann, F., Guo, Y., Han, X. & Wen, L. (2017) Why choose Random Forest to
32 478 predict rare species distribution with few samples in large undersampled areas? Three
33 479 Asian crane species models provide supporting evidence. *PeerJ*, **5**, e2849.
34
35
36
37
38 480 Munson, M.A., Webb, K., Sheldon, D., Fink, D., Hochachka, W.M., Iliff, M., Riedewald, M.,
39 481 Sorokina, D., Sullivan, B., Wood, C. & Kelling, S. (2014) *The eBird Reference Dataset*,
40 482 Version 2014. Cornell Lab of Ornithology, Ithaca, NY.
41
42
43
44 483 National Assessment Synthesis Team (U.S.) (ed). (2001) *Climate Change Impacts on the United*
45 484 *States: The Potential Consequences of Climate Variability and Change: Foundation*.
46 485 Cambridge University Press, Cambridge ; New York.
47
48
49
50 486 O'Connor, R.J., Jones, M.T., Boone, R.B. & Lauber, T.B. (1999) Linking continental climate,
51 487 land use, and land patterns with grassland bird distribution across the conterminous United
52 488 States. *Studies in Avian Biology*, **19**, 45–59.
53
54
55
56
57
58
59
60

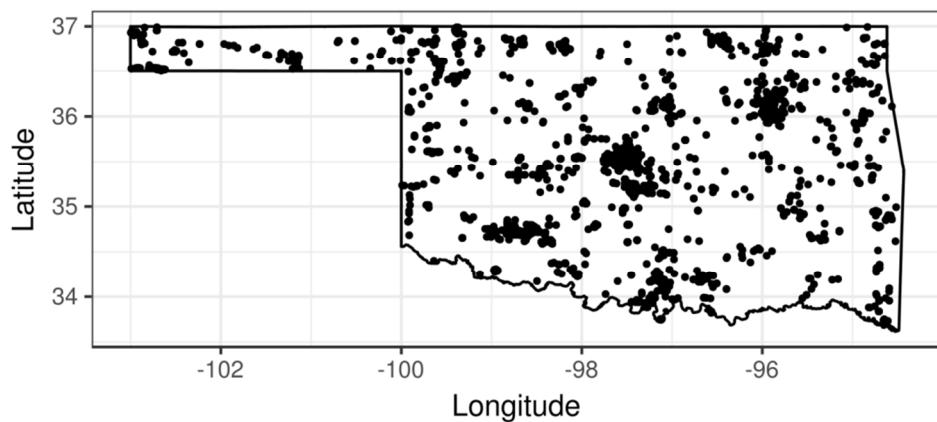
- 1
2
3 489 Oklahoma Climatological Survey. (2017) Climate of Oklahoma,
4 490 http://climate.ok.gov/index.php/site/page/climate_of_oklahoma
5
6
7 491 Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I. & Louzao, M.
8 492 (2012) Comparison of five modelling techniques to predict the spatial distribution and
9 493 abundance of seabirds. *Biological Conservation*, **156**, 94–104.
10
11
12
13 494 Pebesma, E. & Bivand, R.S. (2005) Classes and methods for spatial data in R. *R News*, **5**.
14
15
16 495 Phillips, S.J., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species
17 496 distribution modeling. *Proceedings of the Twenty-First International Conference on*
18 497 *Machine Learning*, p. 83. ACM.
19
20
21
22 498 Pipher, E.N., Curry, C.M. & Koper, N. (2016) Cattle Grazing Intensity and Duration Have
23 499 Varied Effects on Songbird Nest Survival in Mixed-Grass Prairies. *Rangeland Ecology &*
24 500 *Management*, **69**, 437–443.
25
26
27
28 501 Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer classification and regression tree
29 502 techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181–
30 503 199.
31
32
33
34
35 504 Quantum GIS Development Team. (2016) *Quantum GIS Geographic Information System*. Open
36 505 Source Geospatial Foundation Project.
37
38
39 506 Samson, F. & Knopf, F.L. (1994) Prairie conservation in North America. *BioScience*, **44**, 418–
40 507 421.
41
42
43
44 508 Samson, F.B., Knopf, F.L. & Ostlie, W. (2004) Great Plains ecosystems: past, present, and
45 509 future. *Wildlife Society Bulletin*, **32**, 6–15.
46
47
48 510 Sauer, J.R., Niven, D.K., Hines, J.E., Ziolkowski, Jr., D.J., Pardieck, K.L., Fallon, J.E. & Link,
49 511 W.A. (2017) *The North American Breeding Bird Survey, Results and Analysis 1966 -*
50 512 *2015, Version 2.07.2017*. USGS Patuxent Wildlife Research Center, Laurel, MD.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 513 US Bureau of Economic Analysis. (2016) Gross domestic product (GDP) by state (millions of
4 current dollars),
5 514 https://www.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=10&isuri=1
6 &7003=200&7035=-1&7004=naics&7005=4&7006=40000&7036=-
7 &7001=1200&7002=1&7090=70&7007=2014,2013&7093=levels
8
9 517
10
11
12 518 USDA/NASS. (2016) USDA/NASS 2016 State Agriculture Overview for Oklahoma,
13 https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=OKLA
14
15 520 HOMA
16
17
18
19 521 USDA/NRCS - National Geospatial Center of Excellence. (2011) 2011 National Land Cover
20 Dataset (NLCD): Oklahoma.
21
22
23 523 Winter, M., Johnson, D.H. & Shaffer, J.A. (2005) Variability in Vegetation Effects on Density
24 and Nesting Success of Grassland Birds. *The Journal of Wildlife Management*, **69**, 185–
25 524
26 525 197.
27
28
29 526 Wright, C.K. & Wimberly, M.C. (2013) Recent land use change in the Western Corn Belt
30 threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences*,
31 527
32 528 **110**, 4134–4139.
33
34
35 529 Zuckerberg, B., Fink, D., La Sorte, F.A., Hochachka, W.M. & Kelling, S. (2016) Novel seasonal
36 land cover associations for eastern North American forest birds identified through
37 530
38 531 dynamic species distribution modelling. *Diversity and Distributions*, **22**, 717–730.
39
40
41 532
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9

FIGURES

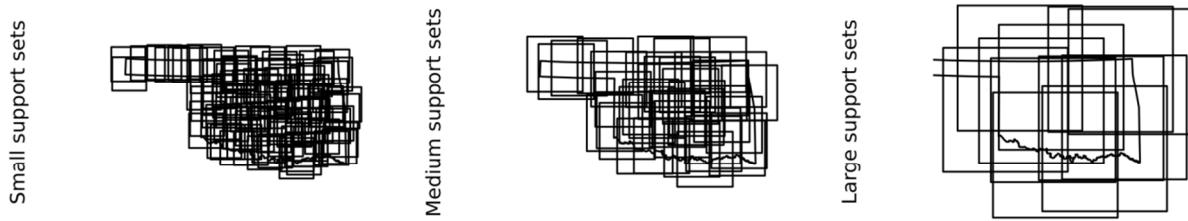
533
534 Fig. 1. The complete dataset used in this study from eBird and surveys by the authors in 2013
535 and 2014 in the central U.S. state of Oklahoma in the Great Plains. The dataset was sampled
536 such that half each were used for model training and model evaluation.



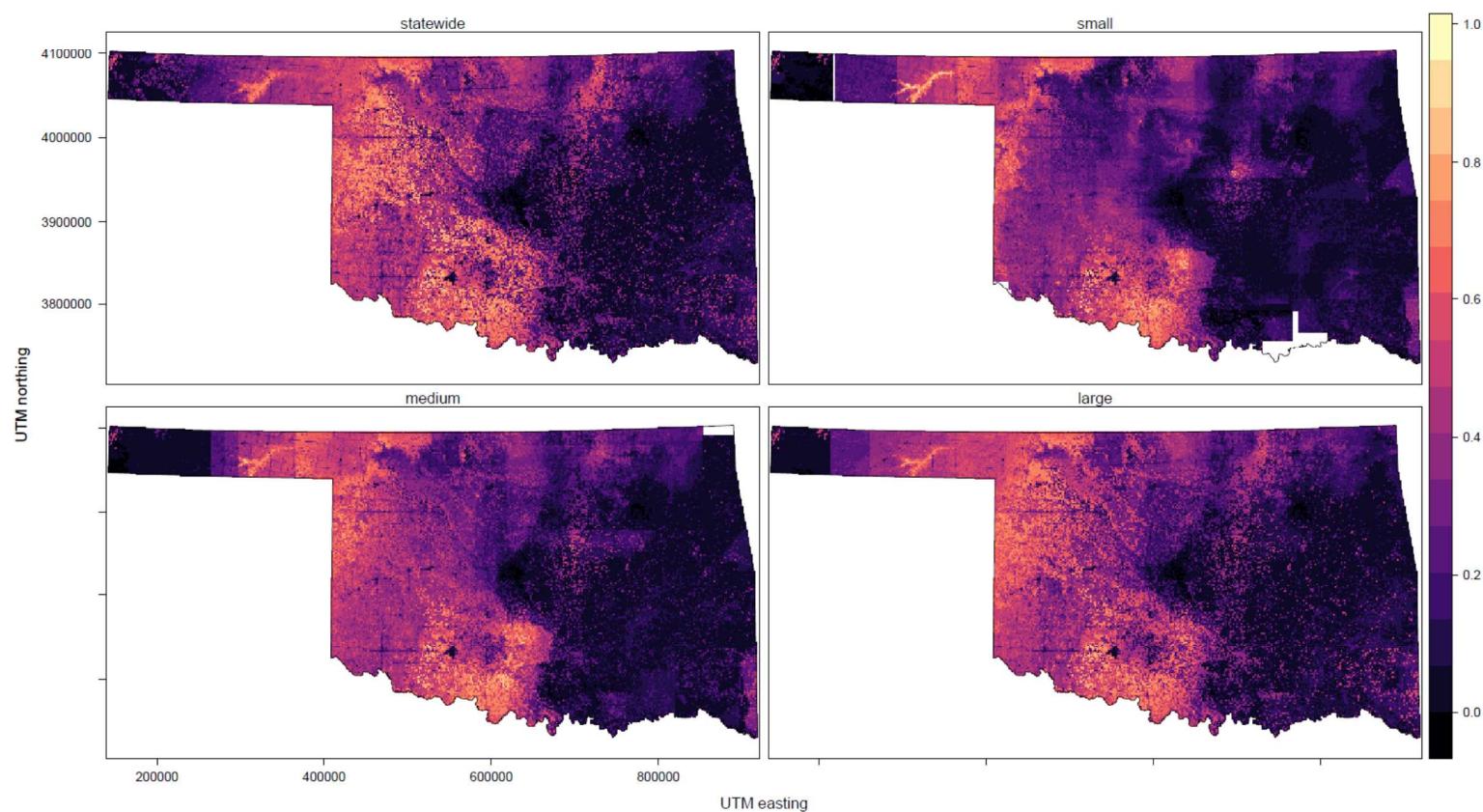
537

538

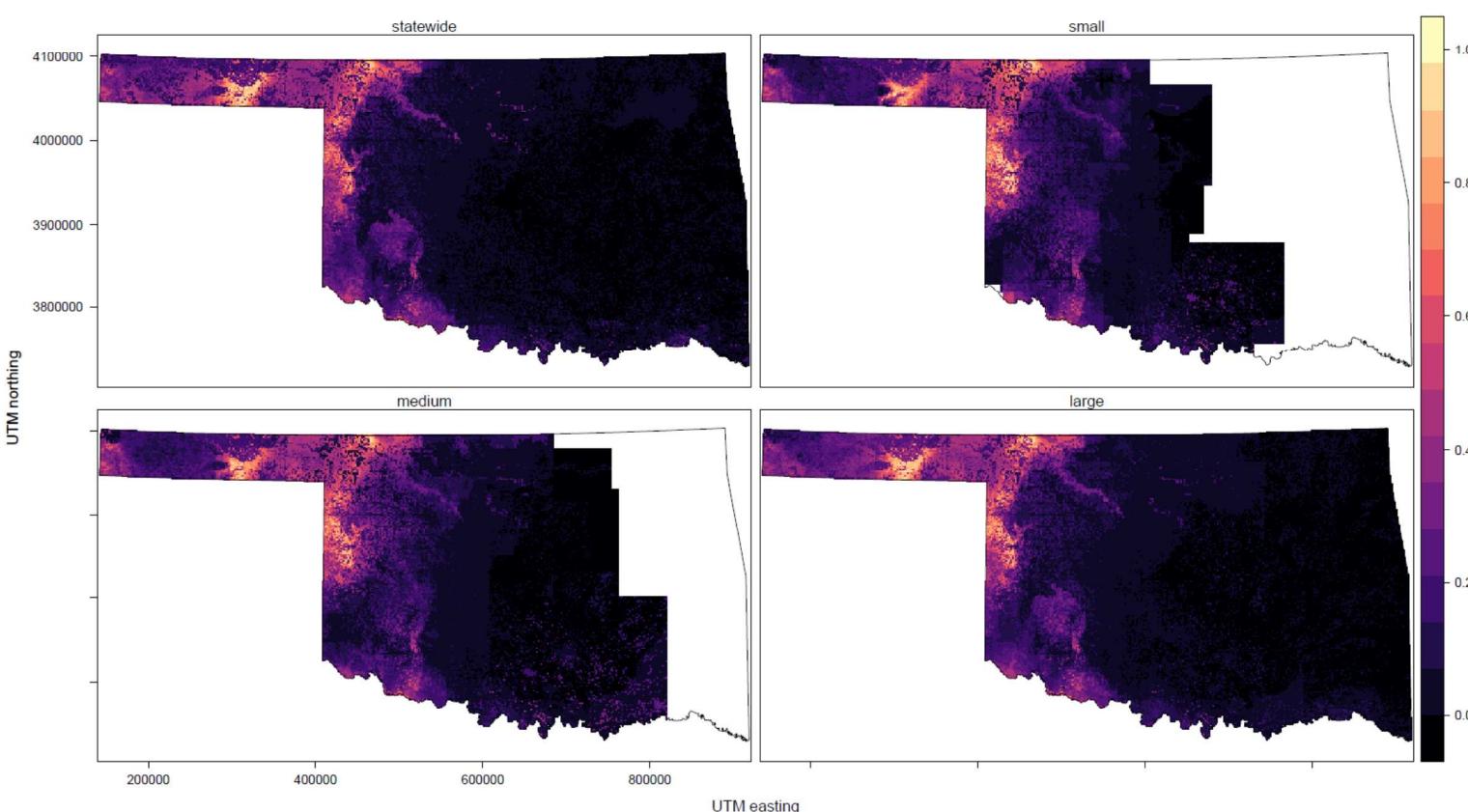
1
2
3 539 Fig. 2. Support sets of small (left), medium (middle), and large (right) scale overlaid over the
4
5 540 study area of Oklahoma, USA.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34 541
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3 542 Fig. 3. Species distribution model for Northern Bobwhite generated at four scales (statewide and three spatially explicit ensemble
4 543 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 544 occurrence from 0-1. Blank areas (in white within the state boundaries) were not able to calculate a model.
6
7

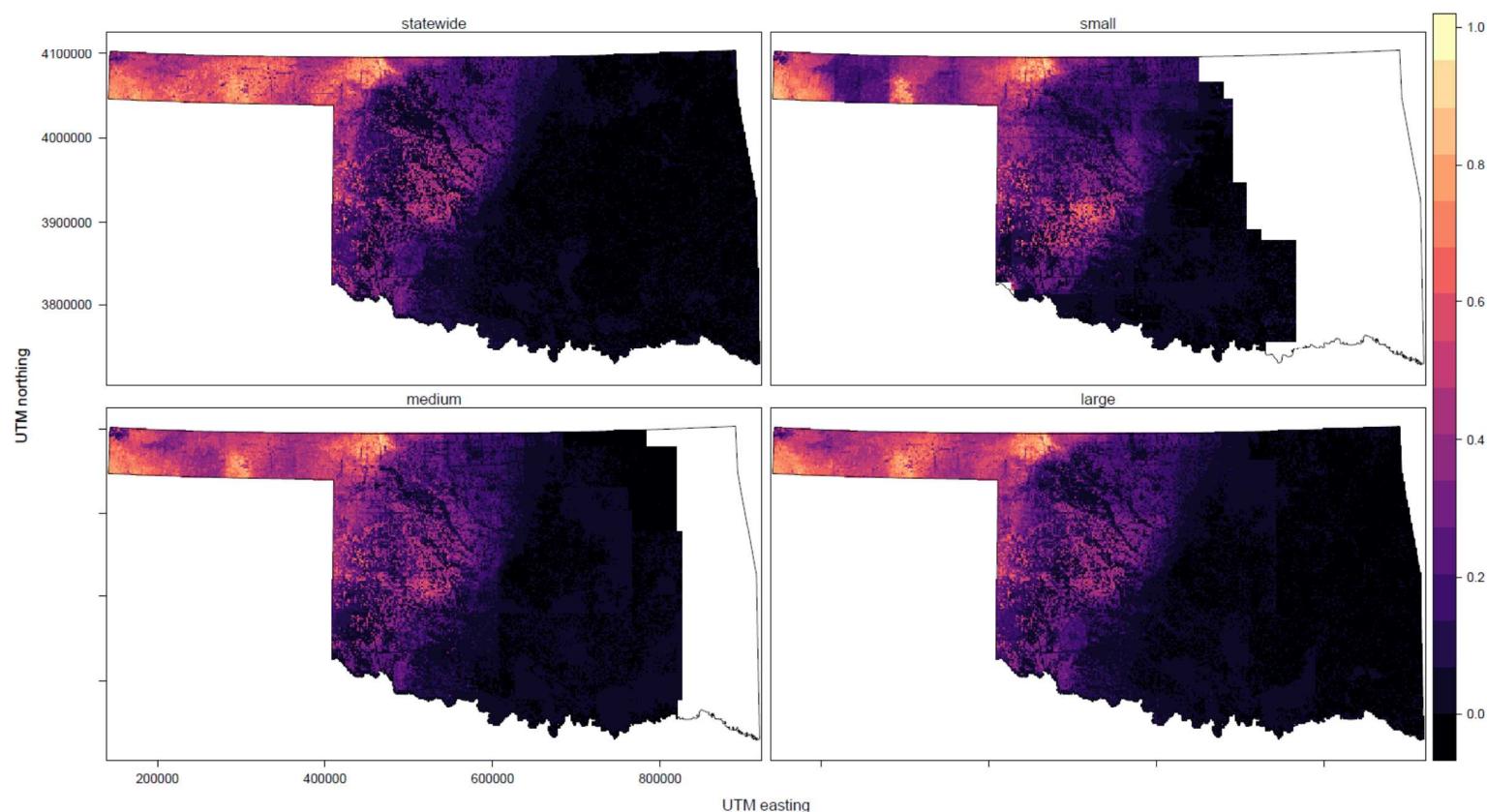


1
2
3 546 Fig. 4. Species distribution model for Cassin's Sparrow generated at four scales (statewide and three spatially explicit ensemble
4 547 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 548 occurrence from 0-1. Blank areas (in white within the state boundaries) were not able to calculate a model.
6
7

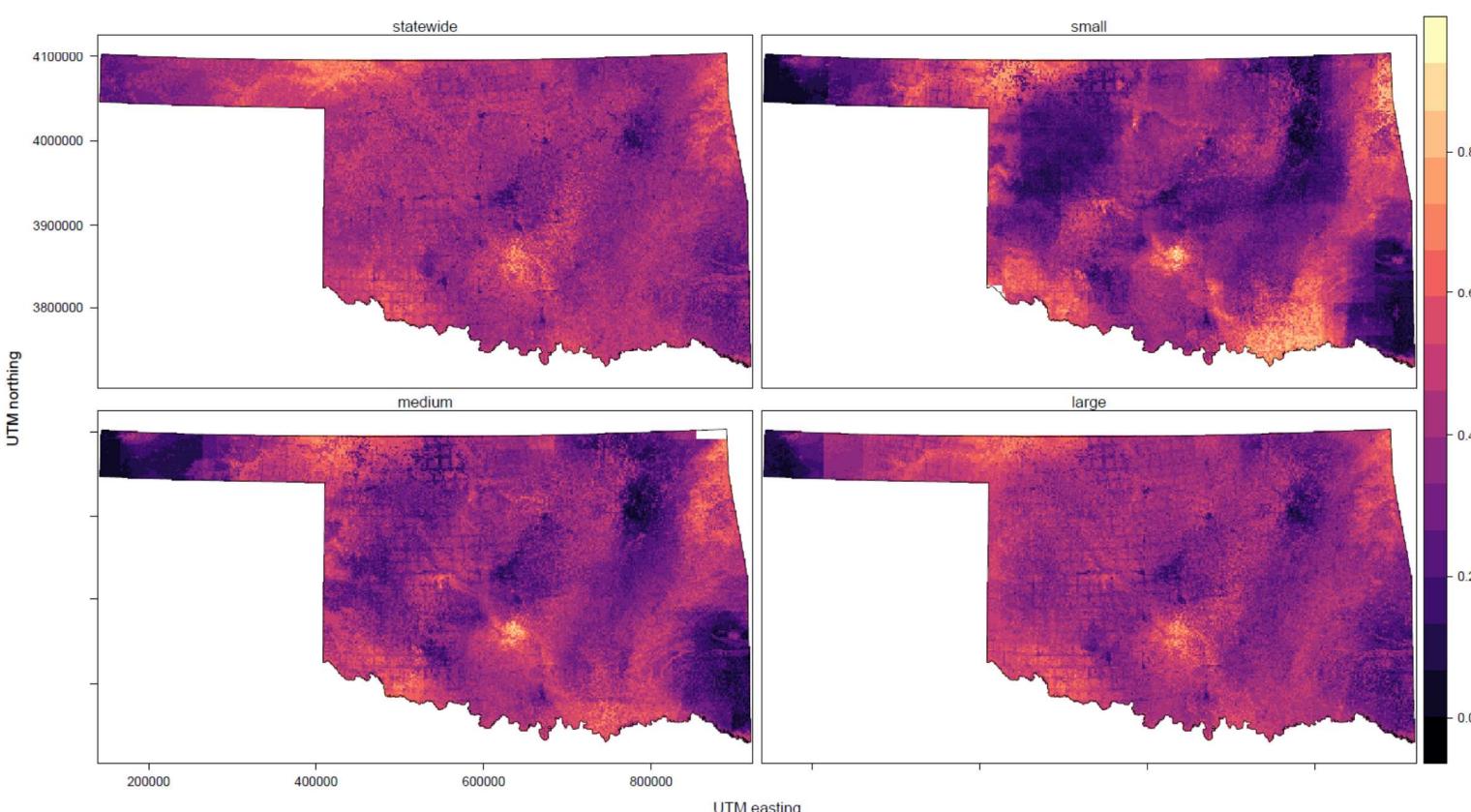


549

1
2
3 550 Fig. 5. Species distribution model for Western Meadowlark generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas (in white within the state boundaries) were not able to calculate a model.
6
7

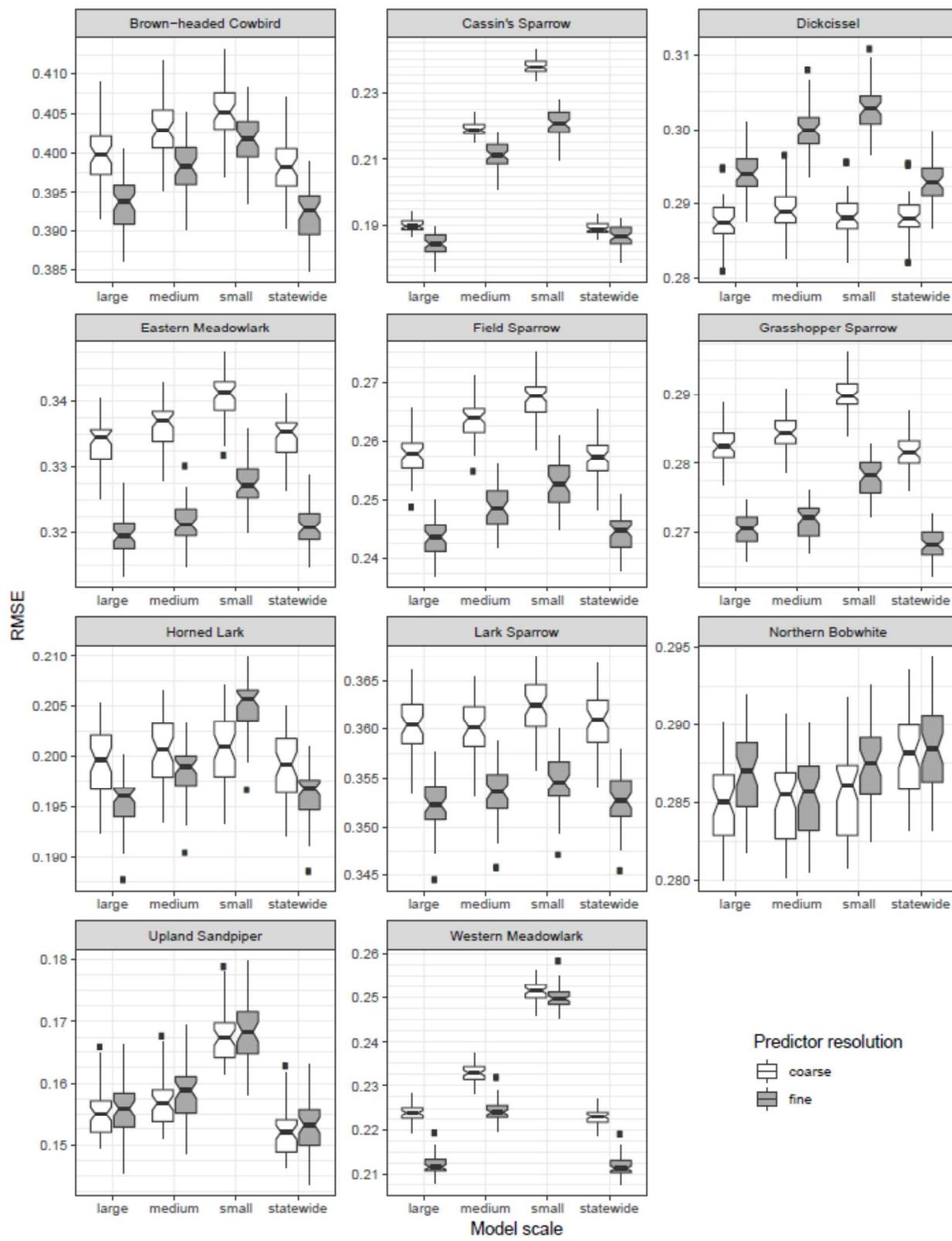


1
2
3 554 Fig. 6. Species distribution model for Brown-headed Cowbird generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas (in white within the state boundaries) were not able to calculate a model.
6
7



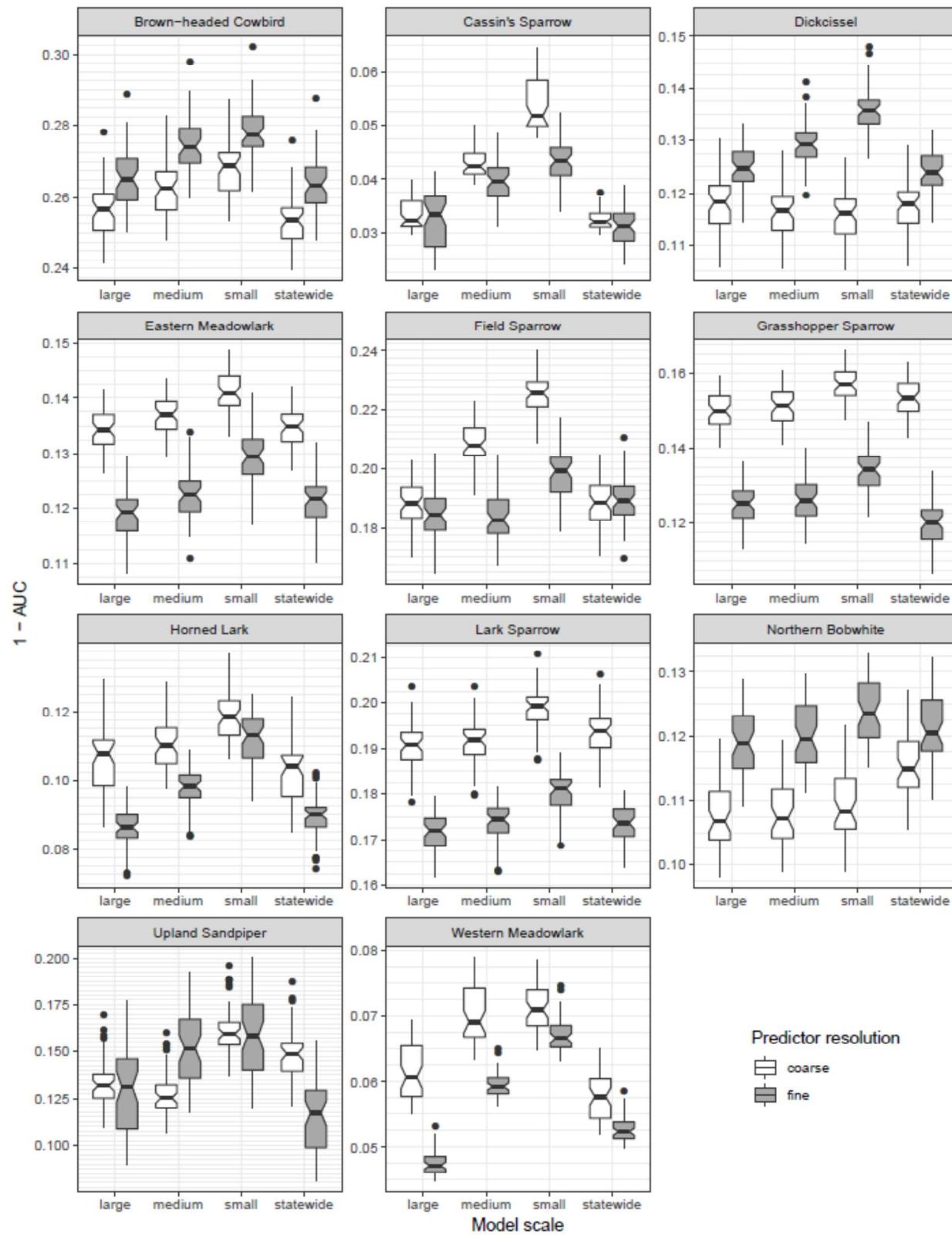
1
2
3 558 Fig. 7. RMSE evaluations for all 44 models compared by predictor resolution. Each panel shows
4 559 one species. Overlapping notches on boxplots show no difference; non-overlapping notches
5 560 show a significant difference in medians. Center line represents median. Fine grid lines are
6 561 shown to facilitate notch comparison.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only



1
2
3 563 Fig. 8. AUC evaluations for all 44 models compared by predictor resolution. Each panel shows
4 564 one species. AUC = 0.5, where prediction is random, and above which prediction is better than
5 565 random. We show the y axis as 1 – AUC so that a lower value is better prediction to facilitate
6 566 comparison with RMSE in Fig. 7. Overlapping notches on boxplots show no difference; non-
7 567 overlapping notches show a significant difference in medians. Center line represents median.
8 568 Fine grid lines are shown to facilitate notch comparison.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only



Supporting Information

Table S1. 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	Undeveloped open space (11, 31, 71, 81, 82, 95)	
undevopenspace_15cell_okmask	Open water (11)	
openwater11_5cell_okmask		
openwater11_15cell_okmask		
devOpenspace21_5cell_okmask	Developed open	

1		
2		
3	dev_openspace21_15cell_okmask	space (21)
4	dev_low22_5cell_okmask	Low intensity
5		development (22)
6	dev_med23_5cell_okmask	Medium intensity
7		development (23)
8	dev_high24_5cell_okmask	High intensity
9		development (24)
10	barren31_5cell_okmask	Barren (31)
11	barren31_15cell_okmask	
12	forest41to43_5cell_okmask	Forest (41, 42, 43)
13	scrub52_5cell_okmask	Scrub and
14		shrubland (52)
15	grasslands71_5cell_okmask	Grasslands (71)
16	pasturehay81_5cell_okmask	Pasture and hay
17		(81)
18	croplands82_5cell_okmask	Croplands (82)
19	woodywetlands90_5cell_okmask	Woody wetlands
20		(90)
21	herbwetlands95_5cell_okmask	Herbaceous
22		wetlands (95)
23	census_utm_30m	Human population density in number per km ²
24		(U.S. Department of Commerce/U.S. Census Bureau 2010)
25		
26	bio1_12_OK	BIO1 = Annual Mean Temperature
27		variables from
28	bio_12_OK	BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
29		Hijmans <i>et al.</i> 2005)
30		
31	bio3_12_OK	BIO3 = Isothermality (BIO2/BIO7) (* 100)
32		
33	bio4_12_OK	BIO4 = Temperature Seasonality (standard deviation *100)
34		
35	bio5_12_OK	BIO5 = Max Temperature of Warmest Month
36		
37	bio6_12_OK	BIO6 = Min Temperature of
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

bio7_12_OK	Coldest Month 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

1
2
3 Fig. S1. Species distribution model for Upland Sandpiper generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas (in white within the state boundaries) were not able to calculate a model.
6
7

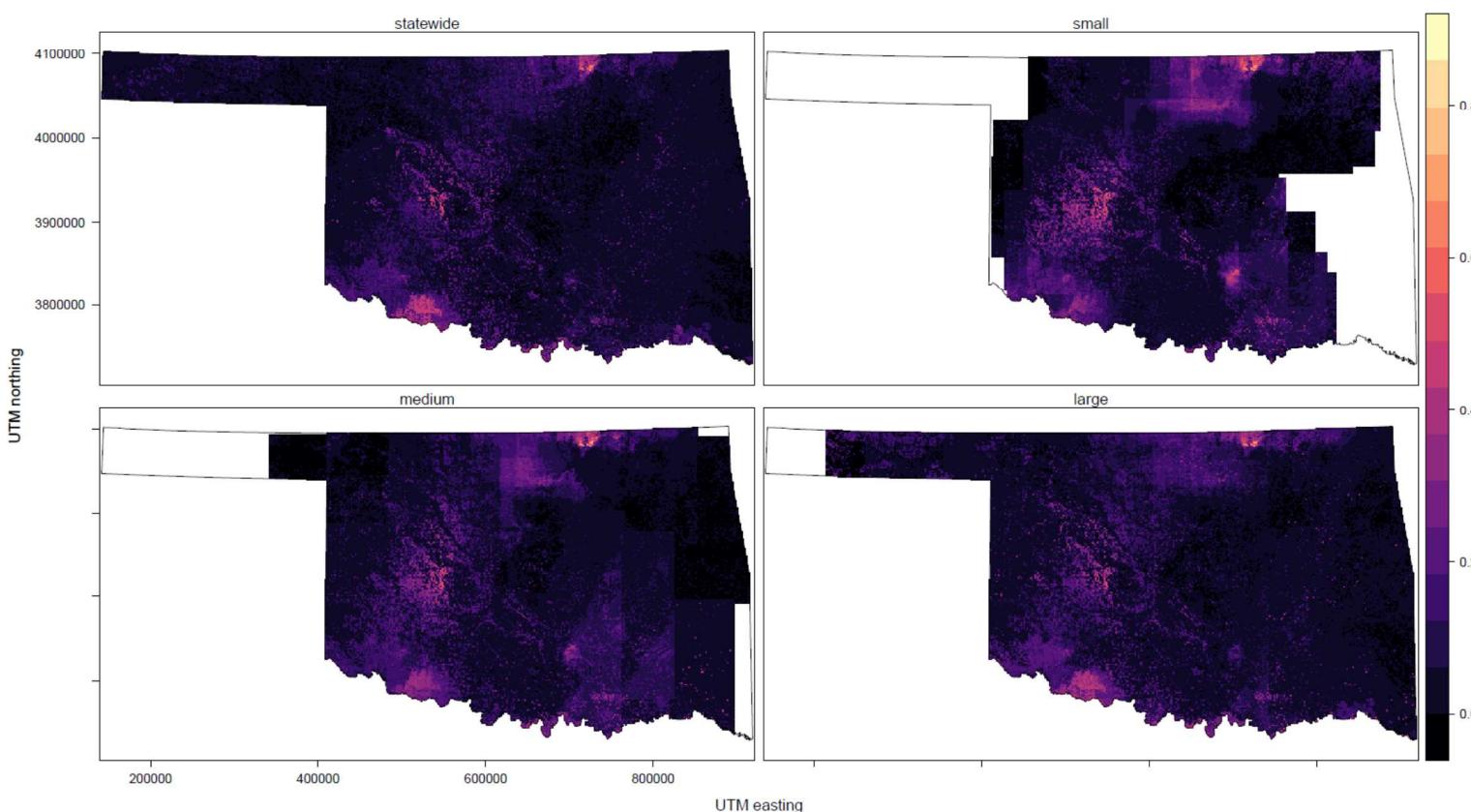


Fig. S2. Species distribution model for Horned Lark generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

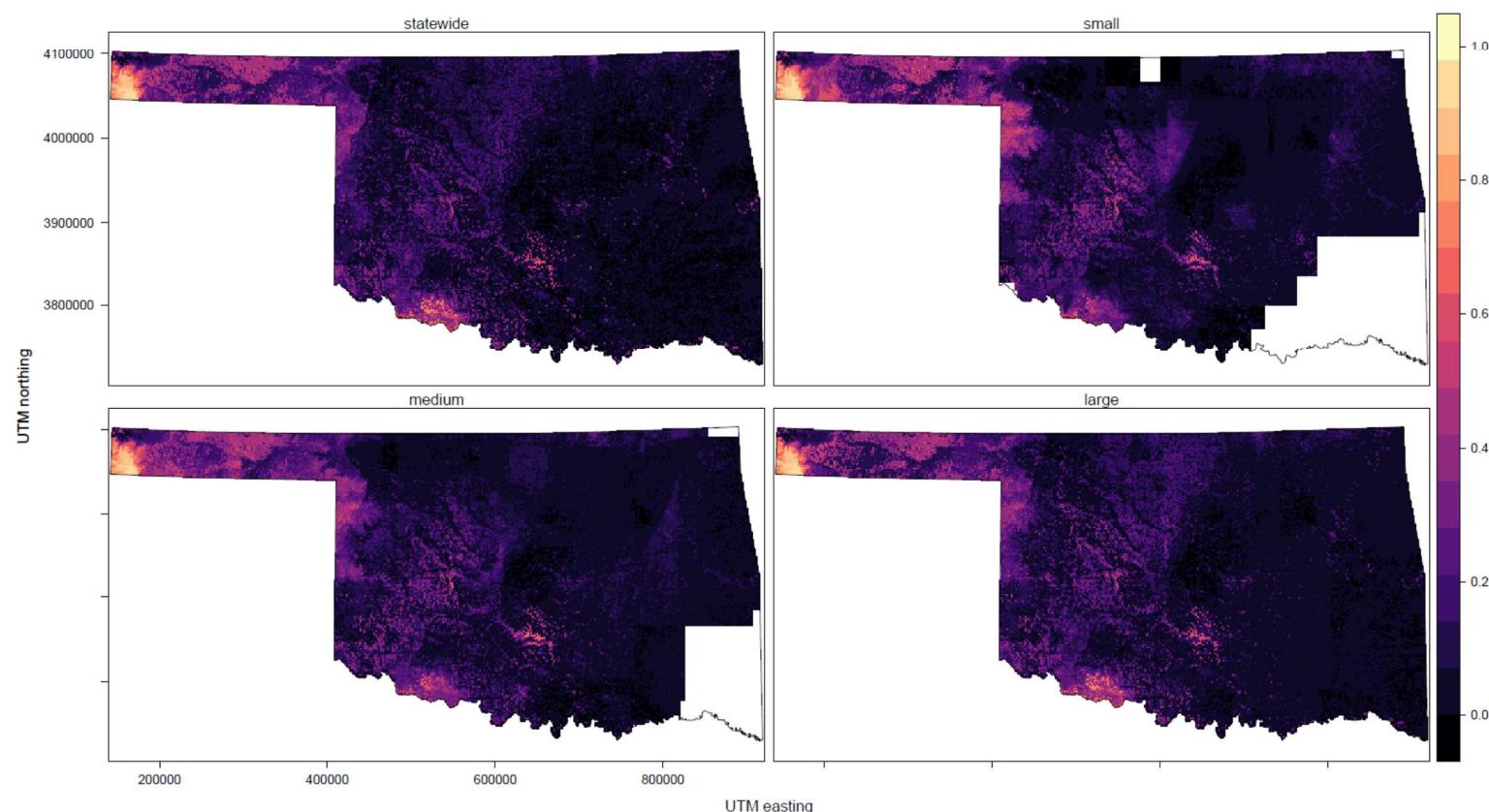


Fig. S3. Species distribution model for Field Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

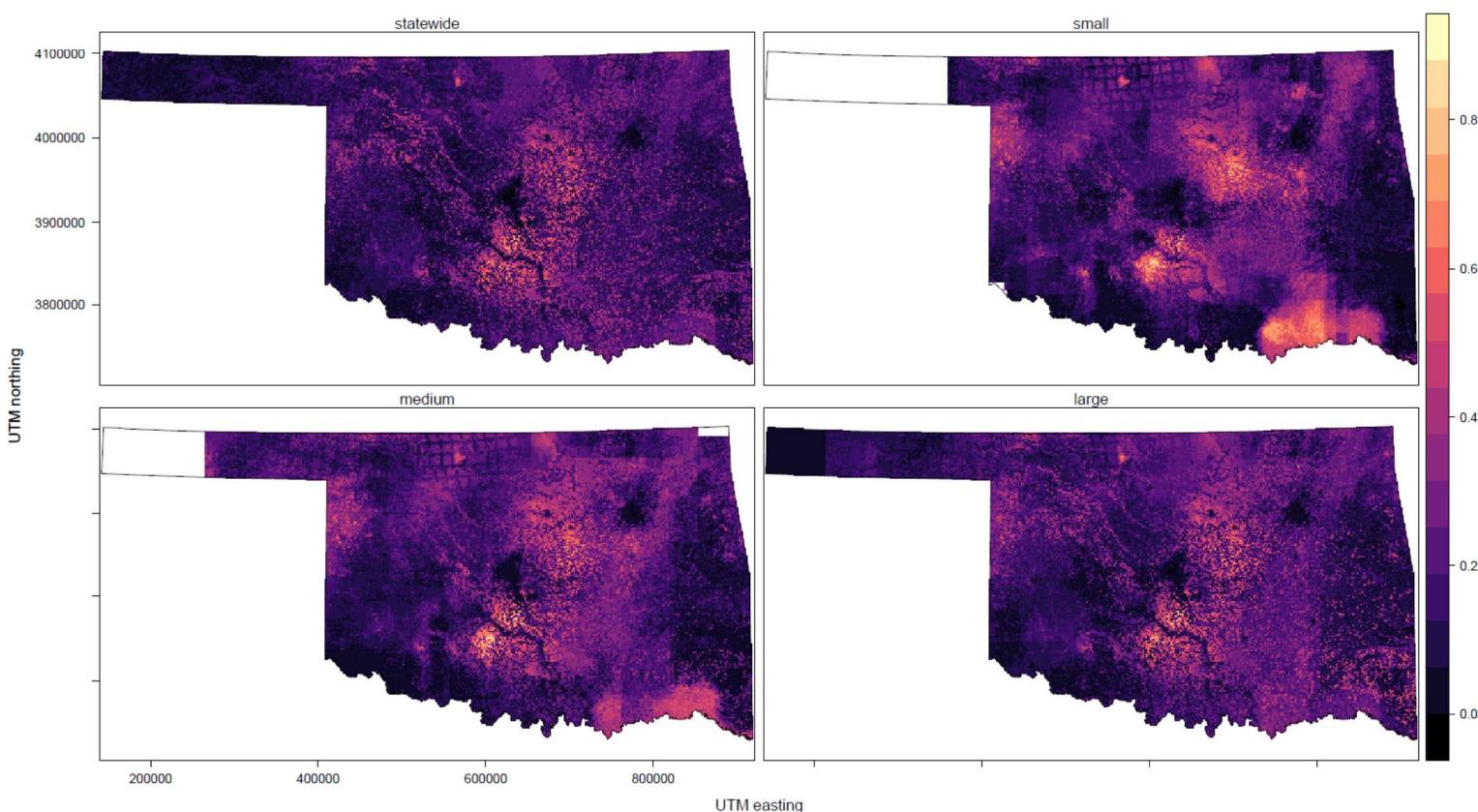


Fig. S4. Species distribution model for Lark Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

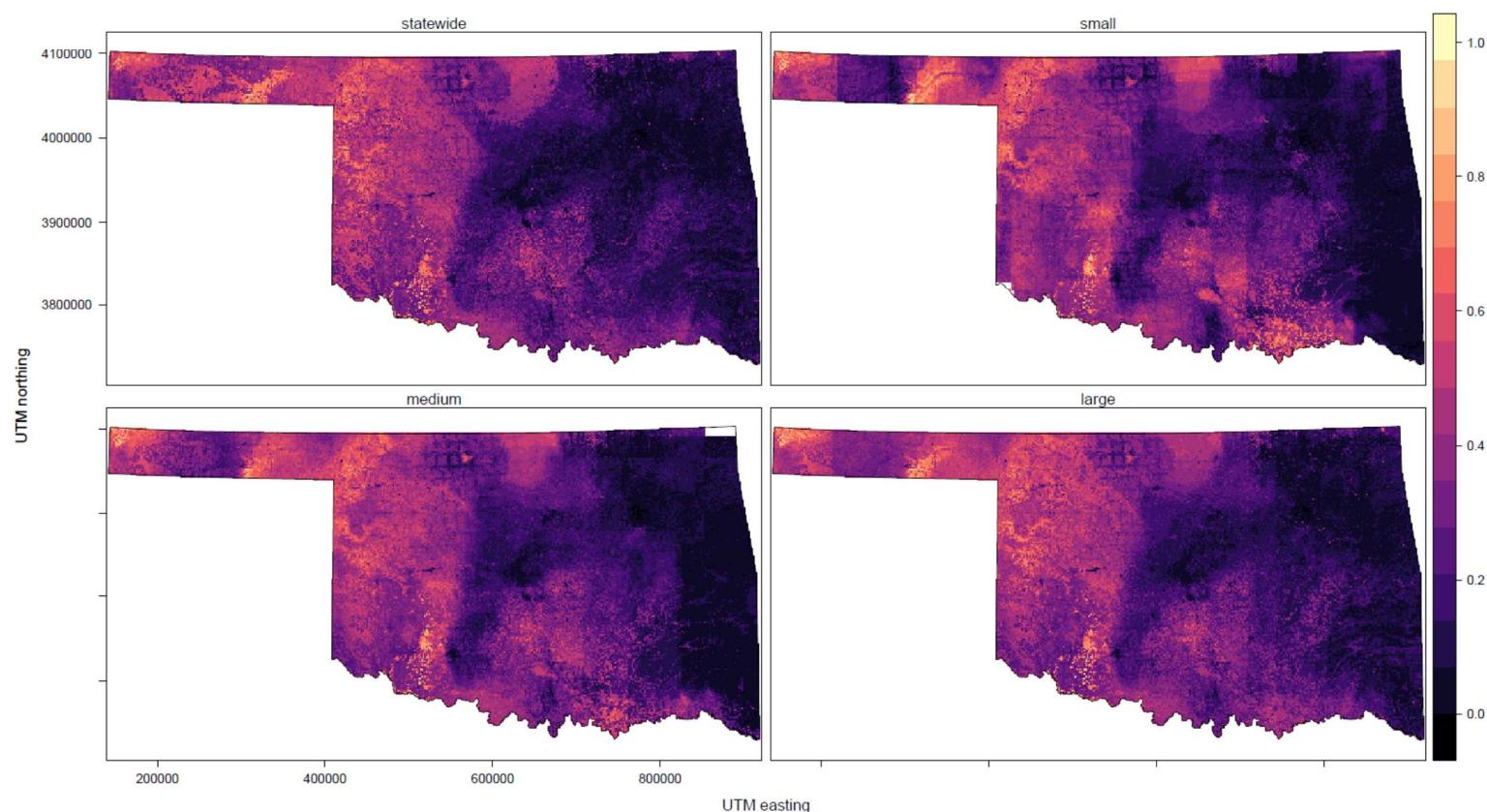


Fig. S5. Species distribution model for Grasshopper Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white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

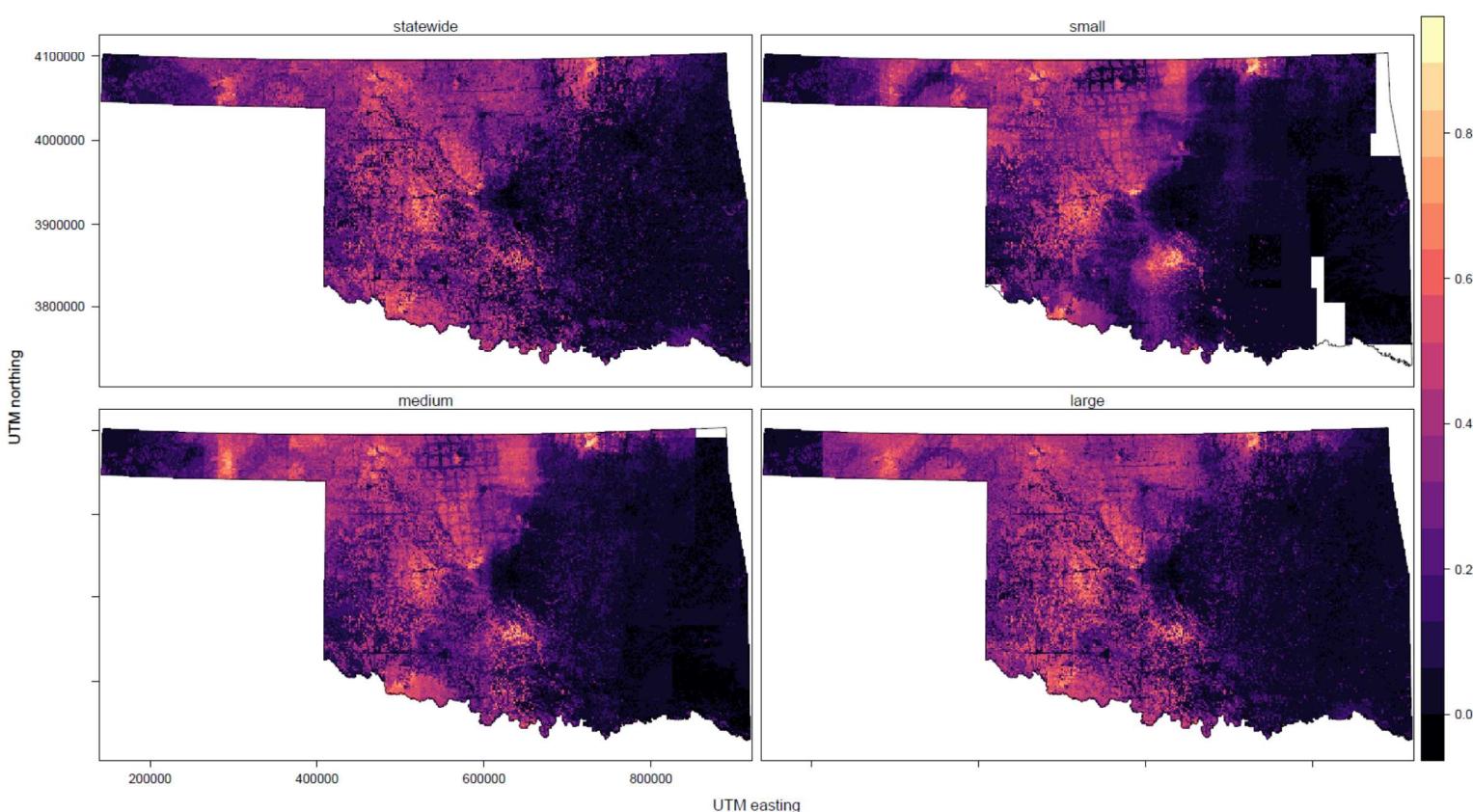
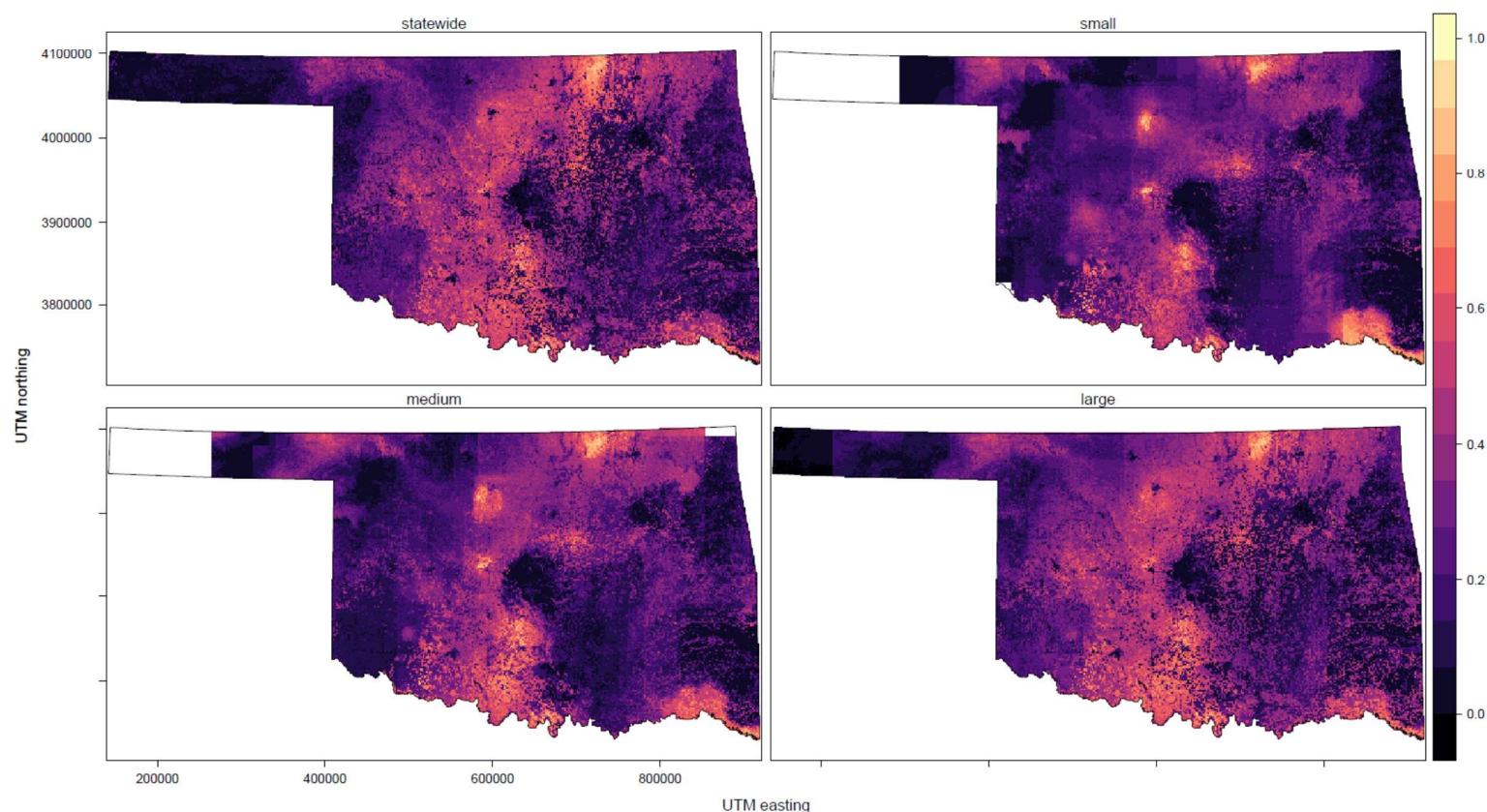
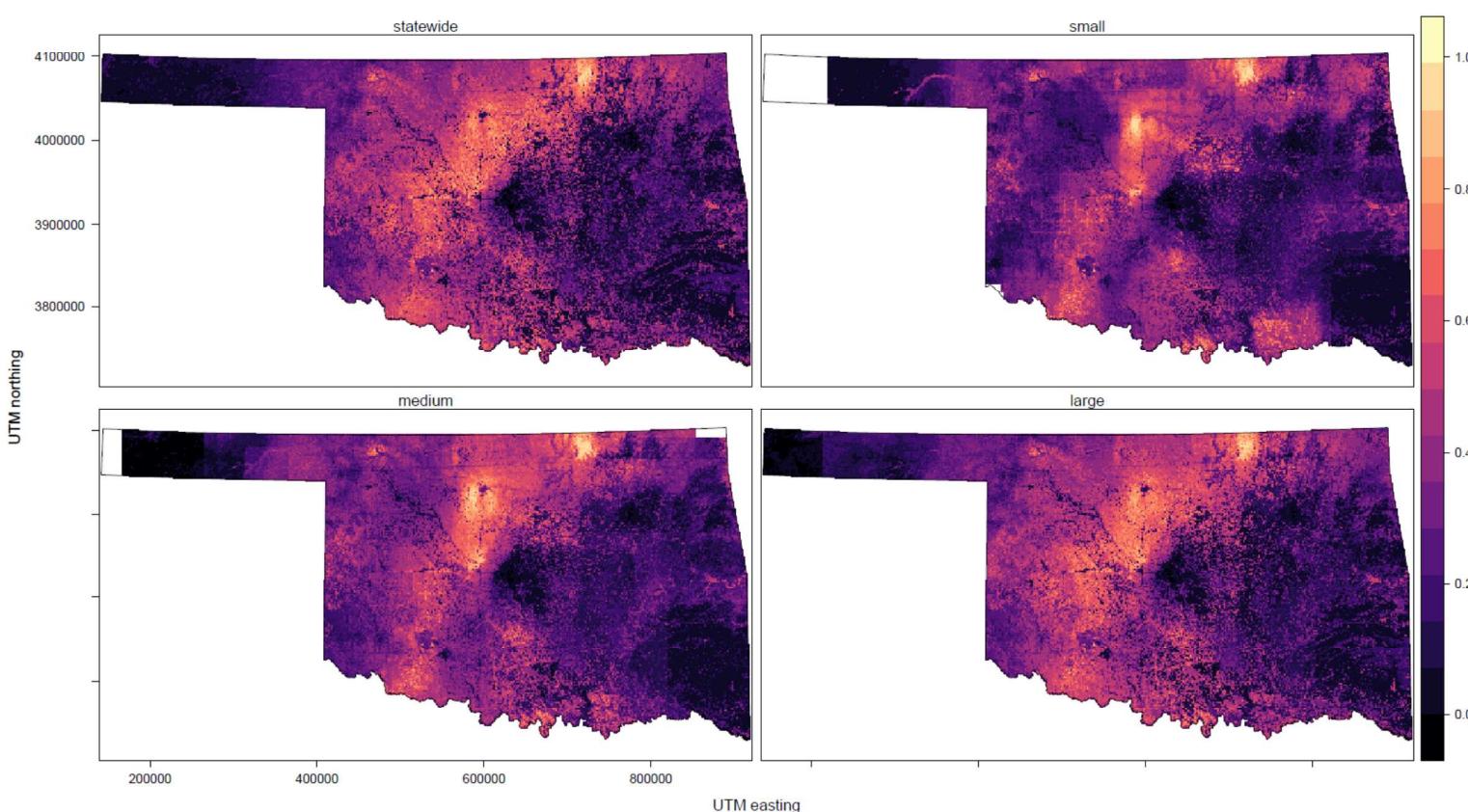


Fig. S6. Species distribution model for Dickcissel generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within the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



1
2
3 Fig. S7. Species distribution model for Eastern Meadowlark generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas (in white within the state boundaries) were not able to calculate a model.
6
7



1
2 1 VARYING DATASET RESOLUTION ALTERS PREDICTIVE ACCURACY OF
3 2 SPATIALLY EXPLICIT ENSEMBLE MODELS FOR AVIAN SPECIES DISTRIBUTION
4
5

6 3 Claire M. Curry^{1,2}
7
8

9 4 Jeremy D. Ross²
10
11

12 5 Andrea J. Contina^{2,3}
13
14

15 6 Eli S. Bridge²
16
17

18 7 ¹University Libraries, University of Oklahoma, 401 W. Brooks St, Norman, OK, [USA](#) 73019
19
20

21 8 ²Oklahoma Biological Survey, University of Oklahoma, 111 E. Chesapeake Street, Norman, OK,
22 9 [USA](#) 73019
23
24

25 10 [³Corix Plains Institute, University of Oklahoma, Five Partners Place, 201 Stephenson Parkway,](#)
26 11 [Norman, OK, USA 73019](#)
27
28

29 12
30
31 13 Corresponding author: Claire M. Curry; address: University Libraries, University of Oklahoma,
32 14 401 W. Brooks St, Norman, OK, [USA](#) 73019, email: cumcurry@ou.edu
33
34

35 15 Running title: Prediction accuracy of spatially explicit ensemble models
36
37

38 16
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

17

SUMMARY

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

35

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 predictor-response relationshipspatterns 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 relative to the study area, with variation in spatial and temporal processes at scales intermediate to the study area and predictor resolution, 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

1
2
3 69 particular, the grassland birds of Oklahoma inhabit diverse grassland types and climatic
4 70 extremes. The open habitats of Oklahoma, which contains over a third of its land area as
5 71 grasslands and an additional 15% as croplands (Diamond & Elliott, 2015), contain grassland
6 72 birds characteristic of habitats ranging from southeastern pine savannahs to tallgrass, mixed-
7 73 grass, and shortgrass prairies (Askins et al., 2007; Diamond & Elliott, 2015). Grassland species
8 74 in areas half the size of Oklahoma in a single ecoregion have shown spatial and temporal
9 75 differences in variable importance (Ethier, Koper, & Nudds, 2017). Forest species, which
10 76 likewise occupy a single habitat type, also show spatial and temporal variation in predictor
11 77 importance (Zuckerberg et al., 2016). Similarly, such a technique has been used on shorebirds in
12 78 habitats with structural similarity to grasslands at a statewide scale (Johnston et al., 2015).
13 79 Finally, Oklahoma occurs on a strong east-west climatic gradient (Oklahoma Climatological
14 80 Survey 2017) that has had profound impacts on the ecosystems of the region (Kukal & Irmak,
15 81 2016; Seager et al., 2018). Physiological balances in animals can change in importance with
16 82 other environmental variables (Kearney, Simpson, Raubenheimer, & Kooijman, 2013), therefore
17 83 variable importance may be expected to change for at least some species across climatic
18 84 gradients. Oklahoma's grassland habitats, agricultural importance, and susceptibility to climate
19 85 change (Loarie et al., 2009; National Assessment Synthesis Team (U.S.), 2001) make it an ideal
20 86 and important region to test relative efficacy of different methods for modeling species
21 87 distributions.
22
23

24
25 88 Grasslands are one of the world's most endangered ecosystems, with declines of 82.6-99.9% of
26 89 tallgrass prairie, 30-99.9% of mixed-grass prairie, and 20-85.8% of short-grass prairie in the
27 90 plains states and provinces of North America (F. Samson & Knopf, 1994), and as such could
28 91 benefit from increased knowledge of distributional drivers. Drivers of decline include land use
29 92 conversion via agriculture and changes in fire and grazing regimes (Samson, Knopf & Ostlie
30 93 2004), although specifics vary by region (Askins et al., 2007). The already tenuous status of
31 94 grassland birds is further threatened by conversion to new crops resulting in permanent land use
32 95 changes (Wright & Wimberly, 2013), generational changes in land use (Higgins, Naugle, &
33 96 Forman, 2002), changes in conservation programs for grassland habitats (Klute, Robel, & Kemp,
34 97 1997), alterations to vegetation (Alward, 1999) and ecosystem structure (Brown, Valone, &
35 98 Curtin, 1997; Hamer, Flather, & Noon, 2006), and climate change (McCarty, 2001). Grassland
36 99 bird species are declining faster than other groups of birds (Askins et al., 2007; Hill, Egan,
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 100 Stauffer, & Diefenbach, 2014; Knopf, 1994) and continue to be imperiled by ongoing and
4 expanding threats to their habitat. Range-wide species distribution predictions have been made
5 for grassland birds but some species with smaller ranges are not accurately modeled (O'Connor
6 et al. 1999), perhaps because some drivers of distribution vary regionally (Askins et al., 2007;
7 Bakker, Naugle, & Higgins, 2002; Ethier et al., 2017). at a scale smaller than the study region.
8 Additionally, spatial and temporal variation in habitat needs and selection pressures (Davis 2005;
9 Winter, Johnson & Shaffer 2005) or interactions with weather events (Pipher, Curry, & Koper,
10 2016) are known to be important in grassland birds, therefore they are particularly suitable as a
11 testing ground for a spatially explicit approach to modeling.
12
13
14
15
16
17
18

19 109 The objectives of our study were threefold. First, we estimated the distribution of Oklahoma
20 grassland birds to understand current distribution statewide with standard species distribution
21 modeling methods. Next, these statewide current distribution predictions allowed us to compare
22 the statewide species distribution model for each species with SEEMs to evaluate whether this
23 approach is suitable at the scale of our region. Finally, we compared each approach's accuracy
24 when using fine- or coarse-resolution predictor sets. Although our approach is at a smaller scale
25 than originally envisioned for SEEMs, it is important to test their potential applicability at the
26 smaller scales at which most management decisions are made. Our results will allow others to
27 make decisions on whether increased accuracy in modeling is worth the additional computational
28 effort required by newer modeling techniques and provide guidance for future work into where
29 given modeling applications are useful.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

121 Study area

122 Oklahoma contains diverse vegetation and climate, making it a suitable region to examine effects
123 of spatially explicit models. There are ca. 165 vegetation types (based on soil and vegetation
124 composition) in 15 land cover types (Diamond & Elliott, 2015), with over a third of the
125 vegetation in grasslands. Rainfall and temperature vary across the state (Oklahoma
126 Climatological Survey, 2017), with annual precipitation ranging from ~ 43cm of rain in the
127 northwest to 142 cm in the southeast and mean annual temperature ranging from ~13°C in the
128 northwest to ~17°C in the southeast. Summer temperatures over 32°C can occur from 60-115

1
2
3 129 days out of the year varying statewide. Agriculture in Oklahoma is dominated by livestock
4 ranching and row crops (USDA/NASS, 2016), and accounted for over \$2.8 billion of the state's
5 gross domestic product in the study years (US Bureau of Economic Analysis, 2016); Oklahoma
6 ranks in the top 5 of US acreage for grain wheat and forage land for hay (USDA/NASS, 2016).
7
8
9
10
11 133 **Bird surveys**

12 134 We collected data 1-4 times each at 339 8-min roadside point counts (0.13 hr) and at 87 non-
13 roadside transects 0.3-3.1 hrs and 0.3-4.3 km long (mean \pm SD: 1.2 \pm 0.6 hrs and 1.8 \pm 0.8 km).
14 135 Each survey was conducted stationary (point counts) or walking at an even pace (transects). We
15 had 14 observers total (6 in 2013 and 8 in 2014). We only used sightings within 500 m of the
16 observer to preserve identification accuracy and recognize that detection is imperfect; however,
17 all models compared use similar data and as such it should not impact our comparison of models.
18 137 A zero (absence) or 1 (presence) was assigned for each combination of date and time and
19 species. We focused on 10 species of grassland birds found during our general surveys
20 [Northern Bobwhite (*Colinus virginianus*); Upland Sandpiper (*Bartramia longicauda*); Horned
21 Lark (*Eremophila alpestris*); Cassin's Sparrow (*Peucaea cassinii*); Field Sparrow (*Spizella*
22 *pusilla*); Lark Sparrow (*Chondestes grammacus*); Grasshopper Sparrow (*Ammodramus*
23 *savannarum*); Dickcissel (*Spiza americana*); Eastern Meadowlark (*Sturnella magna*); and
24 140 Western Meadowlark (*Sturnella neglecta*)], plus the obligate brood parasite Brown-headed
25 Cowbirds (*Molothrus ater*) for which presence often depends on land use factors (Benson,
26 Chiavacci, & Ward, 2013), for a total of 11 species. Many of these species are declining at the
27 state or North American level; none are increasing in population (Sauer et al., 2017).

28 141 We supplemented our survey data for the 11 focal species with citizen science data from the
29 eBird Reference Dataset (Munson et al., 2014) during the months of April, May, June, and July,
30 to match the surveys we conducted. We used complete primary checklist data from 2013-2014
31 and excluded casual counts. Complete checklists contain all birds sighted by the observer;
32 primary checklists are the main checklist submitted when more than one observer submitted
33 checklists for the same observations. We restricted use of eBird samples to \leq 4.3km and \leq 3.1
34 hours to be comparable to our surveys. We used the point count center or the transect midpoint
35 as the count location for our surveys to have comparable precision to eBird coordinates (Fink et
36 al., 2010). Likewise, because some eBird sightings will be from similar locations, we used all
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 159 replicates of our point counts and transects. Because some of our observers entered sightings
4 from before and during our surveys into eBird, we eliminated 14 counts from 2013 and 2014 that
5 were within two hours of the actual survey start time and within 15 km of the survey start
6 location. The combined dataset contained 5422 complete checklists (158 transect sampling
7 events, 613 point count sampling events, and 4651 eBird sampling events). Data points are
8 shown in Fig. 1.
9
10
11
12
13

14 165 To partition training and evaluation datasets, the combined dataset was split randomly for each
15 species using the `createDataPartition` function in the CARET package (Kuhn, 2017), which
16 samples such that both training and evaluation splits have similar distributions of presence and
17 absences.
18
19
20

21 169 **Predictors**

22

23 170 We used bioclimatic variables from WorldClim at 30 second resolution (Hijmans *et al.* 2005),
24 conservation easement status (O'Connor, Jones, Boone, & Lauber, 1999), and land cover
25 variables (USDA/NRCS - National Geospatial Center of Excellence, 2011) to predict bird
26 distribution (Table S1). We also included effort (length of observation in distance and time) and
27 time of day in the analysis to control for differences in bird activity and observer effort that may
28 influence species checklists. Neighborhood predictors were calculated by the values in
29 rectangular areas around each point, at the scale of 5 x 5 pixels (150 x 150 m) and 15 x 15 pixels
30 (450 x 450 m) (Fink *et al.* 2010). Although the 15 x 15 pixel unit is smaller than our 500 m
31 cutoff, most sightings are from even larger areas with the maximum length being under 4.3 km,
32 an area comparable to Fink *et al.* 2010. Additionally, using a neighborhood value centered at the
33 location point still provides information about the neighborhood, whether or not it overlaps or
34 surrounds the sighting. We looked at proportion of each land cover class and proportion of
35 summed open space land covers (grasslands, hay/pasture, cropland, herbaceous wetlands, and
36 barren land) since grassland bird occupancy can be influenced by the total non-structural cover
37 (McDonald 2017). Neighborhoods were created in QGIS 2.16 with the GRASS `r.neighbors`
38 processing tool (Quantum GIS Development Team, 2016).

39 186 We tested for the effects of using coarser (lower resolution) rasters to see if matching predictor
40 and response variable scale affected accuracy. This is applicable as lowering raster resolution
41 could be a route to making potentially more accurate models available to more researchers and
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 189 managers. We scaled our previously created predictor rasters from their native or previously-
4 resampled 30 m resolution to the approximate scale of our largest response data resolution, by
5 decreasing cell size 144-fold to 4.32 km using means in the ‘aggregate’ function in the R
6 package RASTER (Hijmans, 2016). Using these coarser predictor sets trimmed the 2013-2014
7 dataset slightly down to 5327 checklists (2664 for training and 2663 for evaluation).
8
9 193
10
11 194
12
13
14

15 **Species distribution models**

16
17 196 We ran models on Amazon Web Services (AWS) Elastic Cloud Computing (EC2) m4.4xlarge
18 instances (16 vCPU and 64 GiB memory).
19
20

21 *Base model*

22
23 199 To create all species distribution models, we used random forest regression trees (Breiman,
24 200 2001) in the R package RANDOMFOREST (Liaw & Wiener, 2002). Random forest gives results
25 competitive to those from other machine learning techniques such as boosted regression trees
26 and bagged decision trees (used in Fink et al. 2010 for the non-spatially-explicit comparison
27 model). Minimal tuning parameters are required (Caruana & Niculescu-Mizil, 2006; Cutler et
28 al., 2007; Guo, Graber, McBurney, & Balasubramanian, 2010). Random Forests are suitable for
29 species distribution models (Lorena et al., 2011; Prasad, Iverson, & Liaw, 2006) even with few
30 presence records (Mi, Huettmann, Guo, Han, & Wen, 2017). The random forest algorithm
31 bootstraps a subset of the data using only a set proportion of the predictor variables. It then
32 calculates the error rate on training data using the “out of bag” sample (the portion of data not
33 used in the bootstrap for each tree) (Hastie, Tibshirani, & Friedman, 2001). The trees are then
34 averaged for a final model (Prasad et al., 2006). All random forests (both support set and
35 statewide models) were generated with 500 trees which is generally suitable to achieve stability
36 and accuracy (Cutler et al., 2007). We used the default number of variables per bootstrap tree
37 (default ‘mtry’=the square root of the number of predictor variables) for all trees because this is
38 known to result in accurate predictions (Cutler et al., 2007).
39
40 214
41
42
43
44
45
46
47
48
49

50
51 215 Maps were created using the predict function in RASTER at the resolution of the original predictor
52 datasets (30m and 4.32 km). For the maps, we assumed a uniform effort and time of day by
53 creating constants for prediction: mean effort (distance and time) and time of day rasters. Thus,
54
55
56
57
58
59
60

1
2
3 218 all predicted distribution models are generated assuming survey effort does not vary
4 219 geographically and survey effort is typical for both surveys and citizen science efforts in 2013
5 220 and 2014 (mean time: 0.73 hr; mean distance: 0.75 km). The time of day raster for prediction
6 221 was given a value of 7:00am (Fink et al. 2010). Prediction values for evaluation did not use
7 222 these constants.
8
9
10
11

12 223 | *Statewide and SEE M~~m~~odels*

13 224 | We created four models for each species at varying spatial scales: a statewide model and three
14 225 | SEEMs. The statewide model allowed us to compare the performance to SEEMs. A random
15 226 | forest model was created for the statewide scale for each species using all training data. The
16 227 | three remaining models are at varying support set scales, with some modifications from Fink et
17 228 | al. (2010). First, the scale of our support sets reflects the state extent (i.e. our small, medium,
18 229 | and large scales are relatively smaller than those needed for a continent-wide scale). As our
19 230 | survey goals are to determine breeding distribution only, we used a broader temporal window
20 231 | (April-July in all years) for our model. Secondly, for all base models, we used random forest
21 232 | classification trees (Breiman, 2001) as described above. Finally, our geographic sampling of the
22 233 | training and evaluation datasets, described in more detail in the next paragraph, reflects the
23 234 | differing nature of our base models. Fink et al. (2010) sampled 63% of each support set to
24 235 | imitate bootstrapping sampling, but ~~because each of our support sets was being bootstrapped by~~
25 236 | ~~the random forest algorithm~~, we used the full data set for each support set region.

26 237 | Building a SEEM consists of creating random support sets, generating trees and predictions for
27 238 | each support set, and then combining each support set model predictions into the final overall
28 239 | prediction. We created stratified random points in the study area to create support sets (Fig. 2).
29 240 | The randomization of the support set center is important to ~~produce fit useful~~ ensemble models
30 241 | with low bias and high variance (Kuncheva & Whitaker, 2003). We used the ‘spsample’
31 242 | function from the R package SP (Bivand, Pebesma, & Gomez-Rubio, 2013; Pebesma & Bivand,
32 243 | 2005) and created squares of size small (100 boxes of 120 x 120 km), medium (37 boxes of 200
33 244 | x 200 km), or large (12 boxes of 450 x 450 km) around these points, which resulted in no
34 245 | significant difference in pixel coverage ($F_{2,147}=0.63$, $p=0.53$; small mean: 6.9, median 7, range 2-
35 246 | 10; medium mean: 6.3, median 7, range 2-11; large mean: 6.6, median 7, range 2-10) before
36 247 | removing support sets with too few (<25) or uniform (all presence or all absence) checklists
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 248 | (models cannot run with uniform values). [Using a larger number of base model pixel coverage is](#)
4 [ideal to reduce “blockiness” in final ensemble maps and prediction coverage, but we were](#)
5 [limited by computational costs.](#) Each support set included all checklists from the training dataset
6 located within its boundaries. All remaining support set rasters for a given scale (small, medium,
7 or large) were combined into one larger raster using the RASTER mosaic() function to get the
8 mean value of each pixel (ranging from 0 to 1), creating the spatially explicit ensemble (Fink et
9 al., 2010; Hastie et al., 2001; Oppel et al., 2012) made of regional random forest predictions.
10
11 251 | This process was repeated at the three spatial scales, resulting in three SEEMs per species.
12
13
14
15
16
17
18 256 | *Model evaluation and error*

19
20 257 | To evaluate model performance, we created a statewide grid of 10 x 10 km cells and randomly
21 sampled no more than 10 observations from each grid cell for spatial uniformity (Fink et al.,
22 259 | 2010) using the held back data. The actual presence or absence from each checklist is compared
23 to predicted values at each cell with the date and time of the sighting (instead of the uniform date
24 and time used to create the maps). These sampling grid cells are larger than either predictor size
25 and are used to ensure that we do not weight the accuracy of the models towards regions with
26 more reports or surveys. We repeated the evaluation sampling 50 times to create a performance
27 distribution for each model and error type (Fink et al. 2010). We noted the scale (small,
28 medium, large, statewide) with best performance measures for each species and compared
29 performance with notched box plots (Chambers, Cleveland, Kleiner, & Tukey, 1983)[1983](#).
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

267 | Performance measures were root mean square error (RMSE) and area under the curve (AUC).
268 RMSE is calculated from the model residuals, taking the squared value of observed minus
269 expected values, then taking the square root to return to original units; a larger value indicates
270 the model deviates further from expected (Kuhn & Johnson, 2013). AUC is a summary of model
271 performance measuring how often the model misclassifies individual test observations; AUC
272 ranges from 0 to 1, with 1 being perfect and 0.5 being a model that performs no better than
273 random chance (Hanley & McNeil, 1982; James, Witten, Hastie, & Tibshirani, 2013).

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

1
2
3
4
5
6
7
8
9

RESULTS

278 Current statewide distributions are shown in panel (a) of Figs. 3-6 and Figs. S1-7. SEEMs took
279 2.7-12.7 times longer (with fine resolution predictors) or 2.6-22.7 times longer (with coarse
280 resolution predictors) to run than a statewide model, depending on species.

281
282 SEEMs outperformed statewide models for only Northern Bobwhite and Western Meadowlark
283 within each data resolution for both AUC (Fig. 7) and RMSE (Fig. 8). Statewide models
284 outperformed or equaled SEEMs within each data resolution for Brown-headed Cowbird and
285 Dickcissel for both AUC and RMSE.

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

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

292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360

DISCUSSION

293 Although SEEMs increase model accuracy over continental scales (Fink et al., 2013, 2010), our
294 study found their performance differed by species and predictor resolution even in a state with
295 variable climate and diverse ecoregions. Two species were often better represented by SEEMs,
296 suggesting their distributional processes may vary regionally. There were few obvious
297 commonalities among these species that would lead to SEEMs being more accurate for them.
298 One species is non-passerine (Northern Bobwhite), and the other is a common grassland
299 passerine (Western Meadowlark). Two species were always better with statewide models
300 (Brown-headed Cowbird and Dickcissel). The cowbird is strongly dependent on habitat
301 structure (Benson et al., 2013; Bernath-Plaisted, Nenninger, & Koper, 2017), but these variables
302 are not what is measured by the predictor layers that we used. Dickcissel is known for its semi-
303 nomadic movement patterns (Temple, 2002); as such, neither species may be as dependent on
304 local climatic variation mapped by the BioClim predictor inputs. The inconsistencies in the
305 remainder of the species suggest that a larger sample of species and predictor resolutions is
306 needed to compare why models are appropriate for given situations. On our original models, the

1
2
3 307 predictors are consistently finer-scaled (30 m) than some, but not all, response location data
4 (ranging from exact point count locations to aggregate sightings along a 4.3 km transect).
5
6 309 However, Fink *et al.* (2010) used transects almost twice as long as ours (up to 8.1 vs 4.3 km)
7 with 30 m resolution predictor data, so that should not account for differences between our
8 results.
9
10 311

12 312 A potential mechanism for variation between species includes whether species' distributions
13 depend more upon bioclimatic versus ecological variables, as bioclimatic variables should
14 change more smoothly over a larger area (potentially reducing the need for adaptive local
15 models). It could also be that species-specific processes determine whether SEEMs are required.
16 However, one benefit of random forest models and other machine learning methods is minimal
17 tuning and expert opinion required to generate an accurate map (Fink *et al.*, 2010). Requiring
18 researchers to choose spatial scale based on expert opinion of variable importance negates this
19 benefit. However, the fact that most species showed different model performance based on
20 whether we used fine or coarse predictor resolution suggests that model performance depends at
21 least partially on dataset resolutions. Researchers who suspect that a SEE model is appropriate
22 for their dataset and system can compare a small number of base models for different regions or
23 times and see if relationships vary among the test models.
24
25 321

322 An alternative approach for modelers seeking increased accuracy is the use of non-spatially
323 explicit ensemble models, where different base models (predicting for the whole study area) are
324 combined to produce a single prediction map (Araújo & New, 2007; Oppel *et al.*, 2012). We
325 recommend this approach as more efficient for regional managers. Multiple maps will still be
326 generated for the whole study area (n = number of base models used), but typically fewer than
327 the number of support sets created in a SEEM or STEM. These types of ensembles are known to
328 increase accuracy relative to a single base model (Araujo & New 2007; Oppel *et al.* 2012).
329 Although large-scale solutions to conserve grasslands are needed (Samson, Knopf & Ostlie
330 2004), local and regional conservation and management efforts also have critical impacts
331 (Brennan, Kuvlesky & Morrison 2005). We expected that SEEMs would be most accurate and
332 therefore relevant to wildlife management in this state with diverse ecotypes that occur at scales
333 larger than predictors but smaller than our study region. However, based on our study, we
334
335

1
2
3 336 recommend that when using a single base model type, all distribution model types should be run
4 337 (statewide and at least one or more scales of SEEM) if computing capacity is available.
5
6

7 338 Accurate species distribution models can help us understand what factors, both environmental
8 339 and land use, drive species declines (Elith & Leathwick 2009), but we need to conduct modeling
9 340 with predictors and responses at the appropriate spatial scale. Further research is needed to
10 341 elucidate at what study scale and data resolution SEEMs become appropriate. In fact, we found
11 342 a modern laptop or desktop unable to handle fine resolution SEEMs and turned to cloud
12 343 computing to complete them, so the length of time and computing expense involved can be
13 344 substantial. Coarser predictor models were much quicker to run (less than an hour of increase
14 345 relative to statewide models on the high-speed cloud computing), but they were still many times
15 346 longer in runtime than the comparable statewide model. At the continental and temporally fine-
16 347 grained scales, Fink *et al.* (2010)'s result still stands; it is at intermediate scales where more
17 348 research is needed.
18
19

20 349 **AUTHORS' CONTRIBUTIONS**

21 350 ESB, JDR, AJC, and CMC conceived the ideas and designed methodology. ESB and JDR
22 351 collected data. CMC analyzed the data. All authors contributed critically to the drafts and gave
23 352 final approval for publication.
24
25

26 353 **ACKNOWLEDGEMENTS**

27 354 This work was funded by U.S. Department of Agriculture ([USDA](#)) NIFA grant 2013-67009-
28 355 20369 to ESB and supported by the AWS Cloud Credits for Research program. Our survey data
29 356 were collected under OU IACUC R12-019. We thank technicians M. Buschow, R. Feliciano, M.
30 357 Furst, J. Haughawout, J. Hightower, J. Kruk, C. Myers, K. Oliver, T. Smith, R. Soto, and J.
31 358 Tibbits for help in gathering survey data. T. Auer, T. Fagin, W. Hochachka, W.T. Honeycutt, A.
32 359 Johnston, Steffen Oppel, M.A. Patten, one anonymous reviewer, and five anonymous reviewers
33 360 from University of Cambridge journal club provided discussion, technical assistance, and
34 361 comments. CMC was ~~also~~-supported by National Science Foundation ([NSF](#)) grants IDBR
35 362 1014891 and ABI 1458402 to ESB and Oklahoma Department of Wildlife Conservation grant
36 363 F17AF01294 (W-194-R-1) to M.A. Patten. [AJC was supported by NSF grants IDBR 1014891,](#)
37
38

1
2
3 364 | [DGE 1545261](#), and [DEB 0946685](#) and by USDA grant [NIFA-AFRI-003536](#). The authors have
4
5 365 no conflicts of interest to declare.
6
7
8 366 DATA ACCESSIBILITY
9
10 367 Model code and survey data will be archived on datadryad.org upon acceptance. eBird data are
11
12 368 available from eBird.org (Munson et al., 2014).
13
14
15 369 REFERENCES
16
17 370 Alward, R.D. (1999) Grassland vegetation changes and nocturnal global warming. *Science*, **283**,
18
19 371 229–231.
20
21
22 372 Araújo, M. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology
& Evolution*, **22**, 42–47.
23
24
25 374 Askins, R.A., Chávez-Ramírez, F., Dale, B.C., Haas, C.A., Herkert, J.R., Knopf, F.L. & Vickery,
26
27 375 P.D. (2007) Conservation of grassland birds in North America: understanding ecological
28
29 376 processes in different regions. *Ornithological Monographs*, **64**, iii-46.
30
31
32 377 Bakker, K.K., Naugle, D.E. & Higgins, K.F. (2002) Incorporating landscape attributes into
33
34 378 models for migratory grassland bird conservation. *Conservation Biology*, **16**, 1638–1646.
35
36
37 379 Beaumont, L.J., Pitman, A.J., Poulsen, M. & Hughes, L. (2007) Where will species go?
38
39 380 Incorporating new advances in climate modelling into projections of species
40
41 381 distributions. *Global Change Biology*, **13**, 1368–1385.
42
43
44 382 Benson, T.J., Chiavacci, S.J. & Ward, M.P. (2013) Patch size and edge proximity are useful
45
46 383 predictors of brood parasitism but not nest survival of grassland birds. *Ecological
47 384 Applications*, **23**, 879–887.
48
49
50 385 Bivand, R.S., Pebesma, E. & Gomez-Rubio, V. (2013) *Applied Spatial Data Analysis with R*,
51
52 386 2nd ed. Springer, New York.
53
54 387 Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.
55
56
57
58
59
60

- 1
2
3 388 Brennan, L.A., Kuvlesky, W.P. & Morrison. (2005) North American grassland birds: an
4
5 389 unfolding conservation crisis? *Journal of Wildlife Management*, **69**, 1–13.
6
7
8 390 Brown, J.H., Valone, T.J. & Curtin, C.G. (1997) Reorganization of an arid ecosystem in response
9
10 391 to recent climate change. *Proceedings of the National Academy of Sciences*, **94**, 9729–
11 392 9733.
12
13
14 393 Caruana, R. & Niculescu-Mizil, A. (2006) An empirical comparison of supervised learning
15
16 394 algorithms. *Proceedings of the 23rd International Conference on Machine Learning*, pp.
17
18 395 161–168. ACM.
19
20 396 Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J.
21
22 397 (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783–2792.
23
24 398 Davis, S.K. (2005) Nest-Site Selection Patterns and the Influence of Vegetation on Nest Survival
25
26 399 of Mixed-Grass Prairie Passerines. *The Condor*, **107**, 605–616.
27
28
29 400 Diamond, D.D. & Elliott, L.F. (2015) *Oklahoma Ecological Systems Mapping Interpretive
30 Booklet: Methods, Short Type Descriptions, and Summary Results*. Oklahoma
31
32 401 Department of Wildlife Conservation, Norman, OK.
33
34
35 403 Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J.,
36
37 404 Huettmann, F., R. Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A.,
38
39 405 Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T.,
40
41 406 Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams,
42
43 407 S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of
44
45 408 species' distributions from occurrence data. *Ecography*, **29**, 129–151.
46
47 409 Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and
48
49 410 prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*,
50
51 411 **40**, 677–697.
52
53 412 Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees.
54
55 413 *Journal of Animal Ecology*, **77**, 802–813.
56
57
58
59
60

- 1
2
3 414 Ethier, D.M., Koper, N. & Nudds, T.D. (2017) Spatiotemporal variation in mechanisms driving
4 415 regional-scale population dynamics of a Threatened grassland bird. *Ecology and*
5 416 *Evolution.*
- 6
7
8
9 417 Fink, D., Auer, T., Ruiz-Gutierrez, V., Hochachka, W.M., Johnston, A., Sorte, F.A.L. & Kelling,
10 418 S. (2018) Modeling Avian Full Annual Cycle Distribution and Population Trends with
11 419 Citizen Science Data. *bioRxiv*, 251868.
- 12
13
14
15 420 Fink, D., Damoulas, T. & Dave, J. (2013) Adaptive Spatio-Temporal Exploratory Models:
16 421 Hemisphere-wide species distributions from massively crowdsourced eBird data.
17 422 *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, p. 7.
18 423 AAAI Press, Palo Alto, CA.
- 19
20
21
22
23 424 Fink, D., Hochachka, W.M., Zuckerberg, B., Winkler, D.W., Shaby, B., Munson, M.A., Hooker,
24 425 G., Riedewald, M., Sheldon, D. & Kelling, S. (2010) Spatiotemporal exploratory models
25 426 for broad-scale survey data. *Ecological Applications*, **20**, 2131–2147.
- 26
27
28
29
30 427 Guo, Y., Gruber, A., McBurney, R.N. & Balasubramanian, R. (2010) Sample size and statistical
31 428 power considerations in high-dimensionality data settings: a comparative study of
32 429 classification algorithms. *BMC Bioinformatics*, **11**, 447.
- 33
34
35
36 430 Hamer, T.L., Flather, C.H. & Noon, B.R. (2006) Factors associated with grassland bird species
37 431 richness: the relative roles of grassland area, landscape structure, and prey. *Landscape*
38 432 *Ecology*, **21**, 569–583.
- 39
40
41
42 433 Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating
43 434 characteristic (roc) curve. *Radiology* 143: 29–36, **143**, 29–36.
- 44
45
46 435 Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The Elements of Statistical Learning*. Springer
47 436 series in statistics Springer, Berlin.
- 48
49
50
51 437 Higgins, K.F., Naugle, D.E. & Forman, K.J. (2002) A case study of changing land use practices
52 53 in the northern Great Plains, U.S.A.: an uncertain future for waterbird conservation.
54 439 *Waterbirds*, **25**, 42–50.
- 55
56
57
58
59
60

1
2
3 440 Hijmans, R.J. (2016) *Raster: Geographic Data Analysis and Modeling*. R package version 2.5-8.
4 441 Available at: <https://CRAN.R-project.org/package=raster>.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

442 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution
443 interpolated climate surfaces for global land areas. *International Journal of Climatology*,
444 25, 1965–1978.

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

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

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

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

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

460 Knopf, F.L. (1994) Avian assemblages on altered grasslands. *Studies in Avian Biology*, 15, 247–
461 257.

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

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

- 1
2
3 466 Kuncheva, L.I. & Whitaker, C.J. (2003) Measures of Diversity in Classifier Ensembles and Their
4 467 Relationship with the Ensemble Accuracy. *Machine Learning*, **51**, 181–207.
5
6
7 468 Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. *R News*, **2**, 18–22.
8
9
10 469 Lipsey, M.K., Doherty, K.E., Naugle, D.E., Fields, S., Evans, J.S., Davis, S.K. & Koper, N.
11 470 (2015) One step ahead of the plow: Using cropland conversion risk to guide Sprague's
12 471 Pipit conservation in the northern Great Plains. *Biological Conservation*, **191**, 739–749.
13
14
15
16 472 Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The
17 473 velocity of climate change. *Nature*, **462**, 1052–1055.
18
19
20
21 474 Lorena, A.C., Jacintho, L.F.O., Siqueira, M.F., Giovanni, R.D., Lohmann, L.G., de Carvalho,
22 475 A.C.P.L.F. & Yamamoto, M. (2011) Comparing machine learning classifiers in potential
23 476 distribution modelling. *Expert Systems with Applications*, **38**, 5268–5275.
24
25
26
27 477 McCarty, J.P. (2001) Ecological consequences of recent climate change. *Conservation Biology*,
28 478 **15**, 320–331.
29
30
31 479 Mi, C., Huettmann, F., Guo, Y., Han, X. & Wen, L. (2017) Why choose Random Forest to
32 480 predict rare species distribution with few samples in large undersampled areas? Three
33 481 Asian crane species models provide supporting evidence. *PeerJ*, **5**, e2849.
34
35
36
37 482 Munson, M.A., Webb, K., Sheldon, D., Fink, D., Hochachka, W.M., Iliff, M., Riedewald, M.,
38 483 Sorokina, D., Sullivan, B., Wood, C. & Kelling, S. (2014) *The eBird Reference Dataset*,
39 484 Version 2014. Cornell Lab of Ornithology, Ithaca, NY.
40
41
42
43
44 485 National Assessment Synthesis Team (U.S.) (ed). (2001) *Climate Change Impacts on the United*
45 486 *States: The Potential Consequences of Climate Variability and Change: Foundation*.
46 487 Cambridge University Press, Cambridge ; New York.
47
48
49
50 488 O'Connor, R.J., Jones, M.T., Boone, R.B. & Lauber, T.B. (1999) Linking continental climate,
51 489 land use, and land patterns with grassland bird distribution across the conterminous United
52 490 States. *Studies in Avian Biology*, **19**, 45–59.
53
54
55
56
57
58
59
60

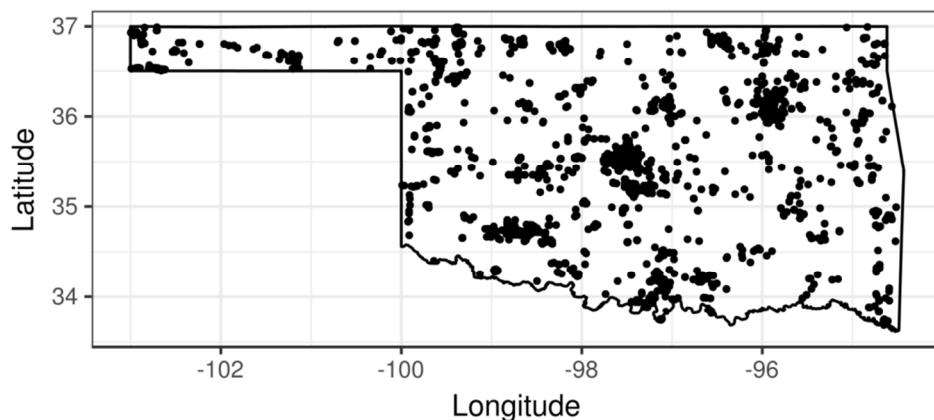
- 1
2
3 491 Oklahoma Climatological Survey. (2017) Climate of Oklahoma,
4 492 http://climate.ok.gov/index.php/site/page/climate_of_oklahoma
5
6
7 493 Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I. & Louzao, M.
8 494 (2012) Comparison of five modelling techniques to predict the spatial distribution and
9 495 abundance of seabirds. *Biological Conservation*, **156**, 94–104.
10
11
12
13 496 Pebesma, E. & Bivand, R.S. (2005) Classes and methods for spatial data in R. *R News*, **5**.
14
15
16 497 Phillips, S.J., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species
17 498 distribution modeling. *Proceedings of the Twenty-First International Conference on*
18 499 *Machine Learning*, p. 83. ACM.
19
20
21
22 500 Pipher, E.N., Curry, C.M. & Koper, N. (2016) Cattle Grazing Intensity and Duration Have
23 501 Varied Effects on Songbird Nest Survival in Mixed-Grass Prairies. *Rangeland Ecology &*
24 502 *Management*, **69**, 437–443.
25
26
27
28 503 Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer classification and regression tree
29 504 techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181–
30 505 199.
31
32
33
34 506 Quantum GIS Development Team. (2016) *Quantum GIS Geographic Information System*. Open
35 507 Source Geospatial Foundation Project.
36
37
38
39 508 Samson, F. & Knopf, F.L. (1994) Prairie conservation in North America. *BioScience*, **44**, 418–
40 509 421.
41
42
43
44 510 Samson, F.B., Knopf, F.L. & Ostlie, W. (2004) Great Plains ecosystems: past, present, and
45 511 future. *Wildlife Society Bulletin*, **32**, 6–15.
46
47
48 512 Sauer, J.R., Niven, D.K., Hines, J.E., Ziolkowski, Jr., D.J., Pardieck, K.L., Fallon, J.E. & Link,
49 513 W.A. (2017) *The North American Breeding Bird Survey, Results and Analysis 1966 -*
50 514 *2015, Version 2.07.2017*. USGS Patuxent Wildlife Research Center, Laurel, MD.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 515 US Bureau of Economic Analysis. (2016) Gross domestic product (GDP) by state (millions of
4 current dollars),
5 516 https://www.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=10&isuri=1
6 &7003=200&7035=-1&7004=naics&7005=4&7006=40000&7036=-
7 &7001=1200&7002=1&7090=70&7007=2014,2013&7093=levels
8
9 519
10
11
12 520 USDA/NASS. (2016) USDA/NASS 2016 State Agriculture Overview for Oklahoma,
13 https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=OKLA
14
15 521 HOMA
16
17
18 522
19 523 USDA/NRCS - National Geospatial Center of Excellence. (2011) 2011 National Land Cover
20 Dataset (NLCD): Oklahoma.
21
22
23 525 Winter, M., Johnson, D.H. & Shaffer, J.A. (2005) Variability in Vegetation Effects on Density
24 and Nesting Success of Grassland Birds. *The Journal of Wildlife Management*, **69**, 185–
25 526
26 527 197.
27
28
29 528 Wright, C.K. & Wimberly, M.C. (2013) Recent land use change in the Western Corn Belt
30 threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences*,
31 529
32 530 **110**, 4134–4139.
33
34
35 531 Zuckerberg, B., Fink, D., La Sorte, F.A., Hochachka, W.M. & Kelling, S. (2016) Novel seasonal
36 land cover associations for eastern North American forest birds identified through
37 532
38 533 dynamic species distribution modelling. *Diversity and Distributions*, **22**, 717–730.
39
40
41 534
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9

FIGURES

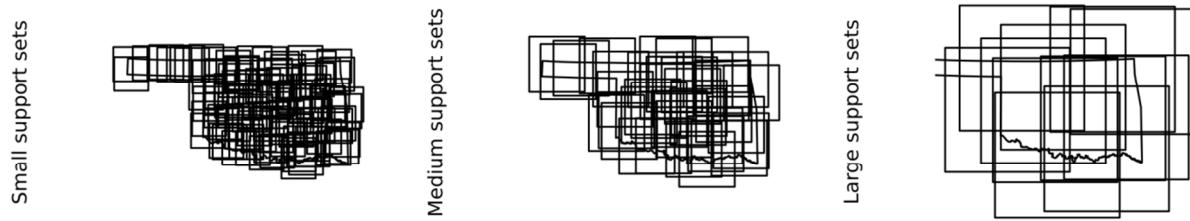
535
536 Fig. 1. The complete dataset used in this study from eBird and surveys by the authors in 2013
537 and 2014 in the central U.S. state of Oklahoma in the Great Plains. The dataset was sampled
538 such that half each were used for model training and model evaluation.



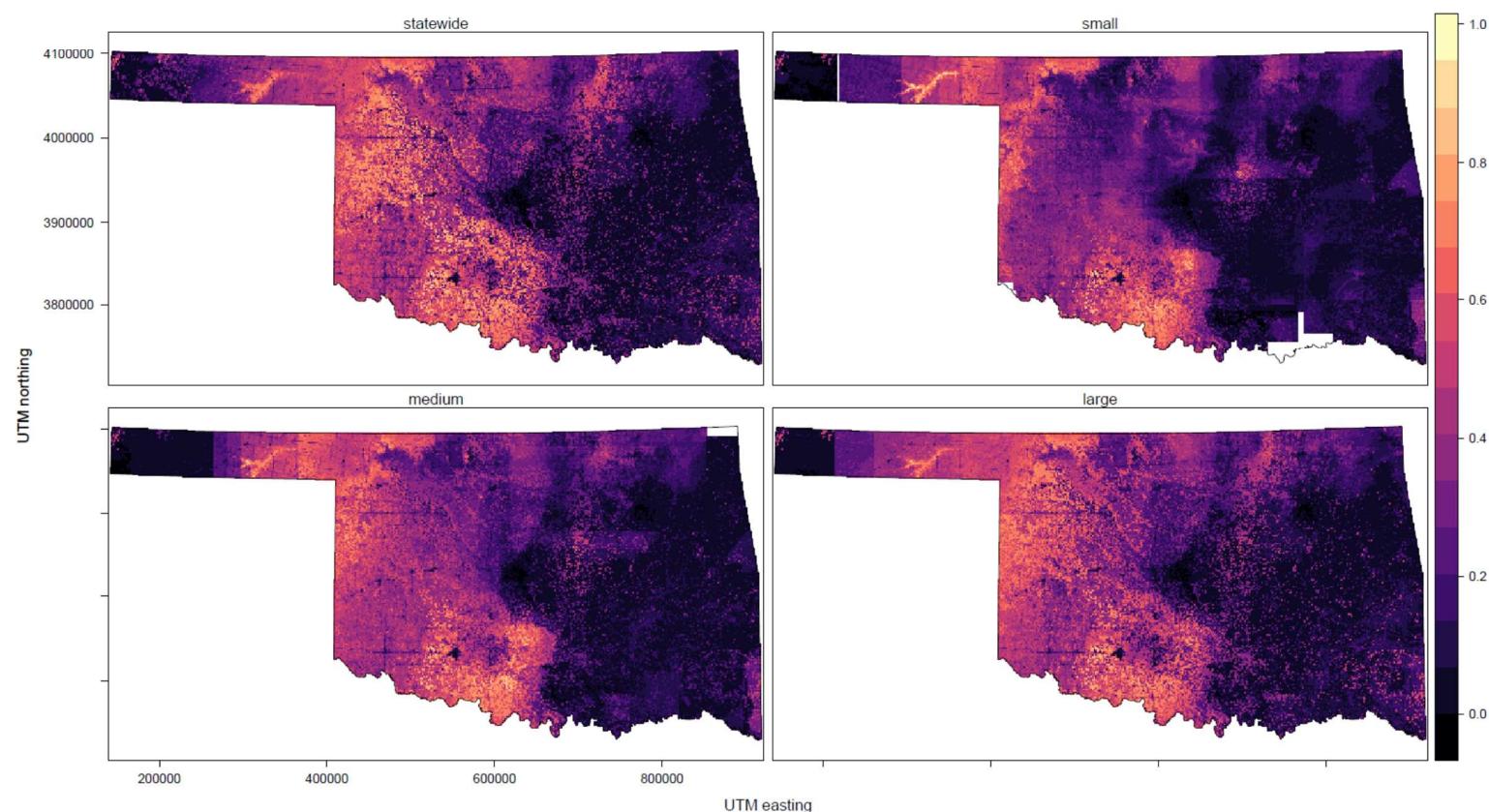
539

540

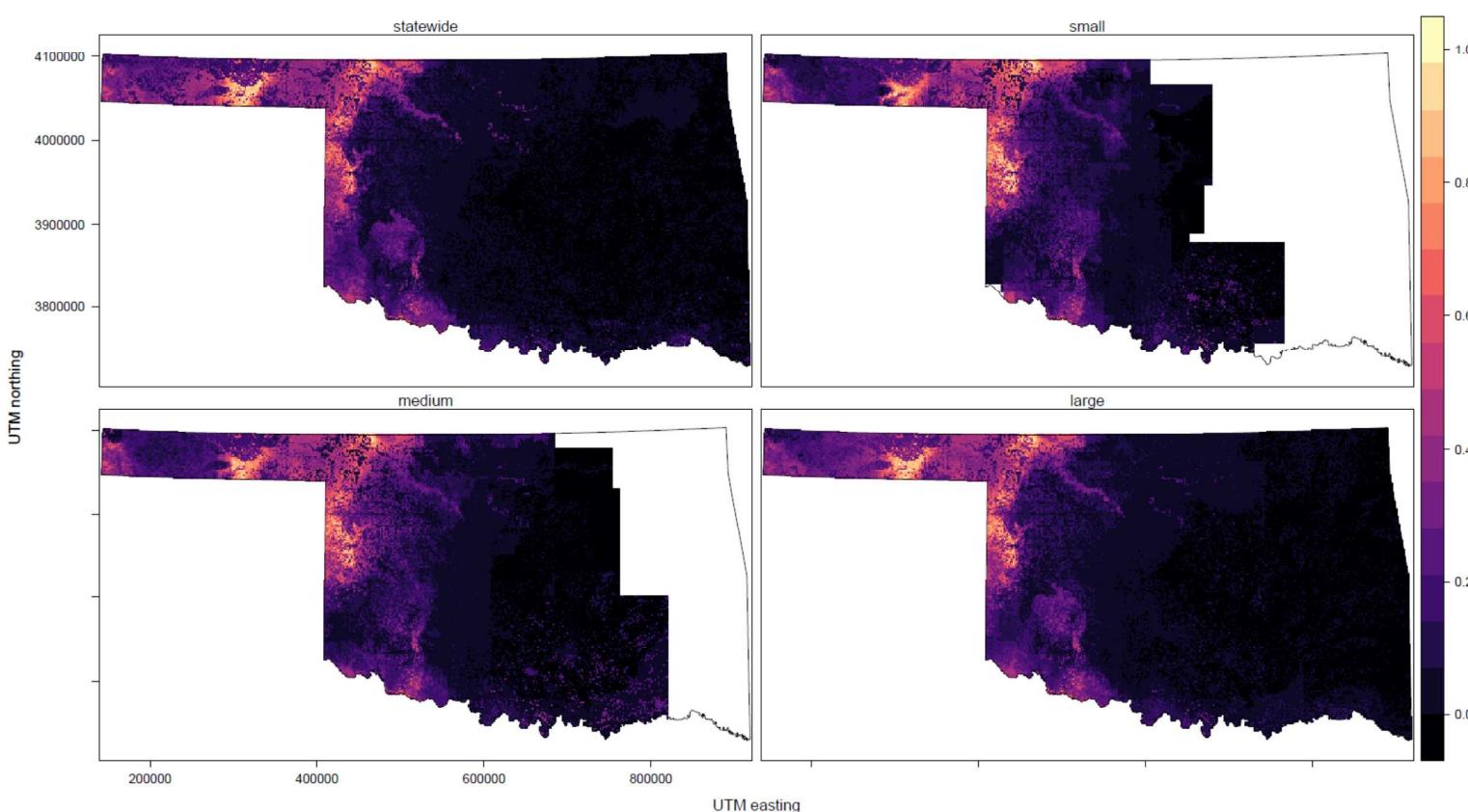
1
2
3 541 Fig. 2. Support sets of small (left), medium (middle), and large (right) scale overlaid over the
4
5 542 study area of Oklahoma, USA.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34 543
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



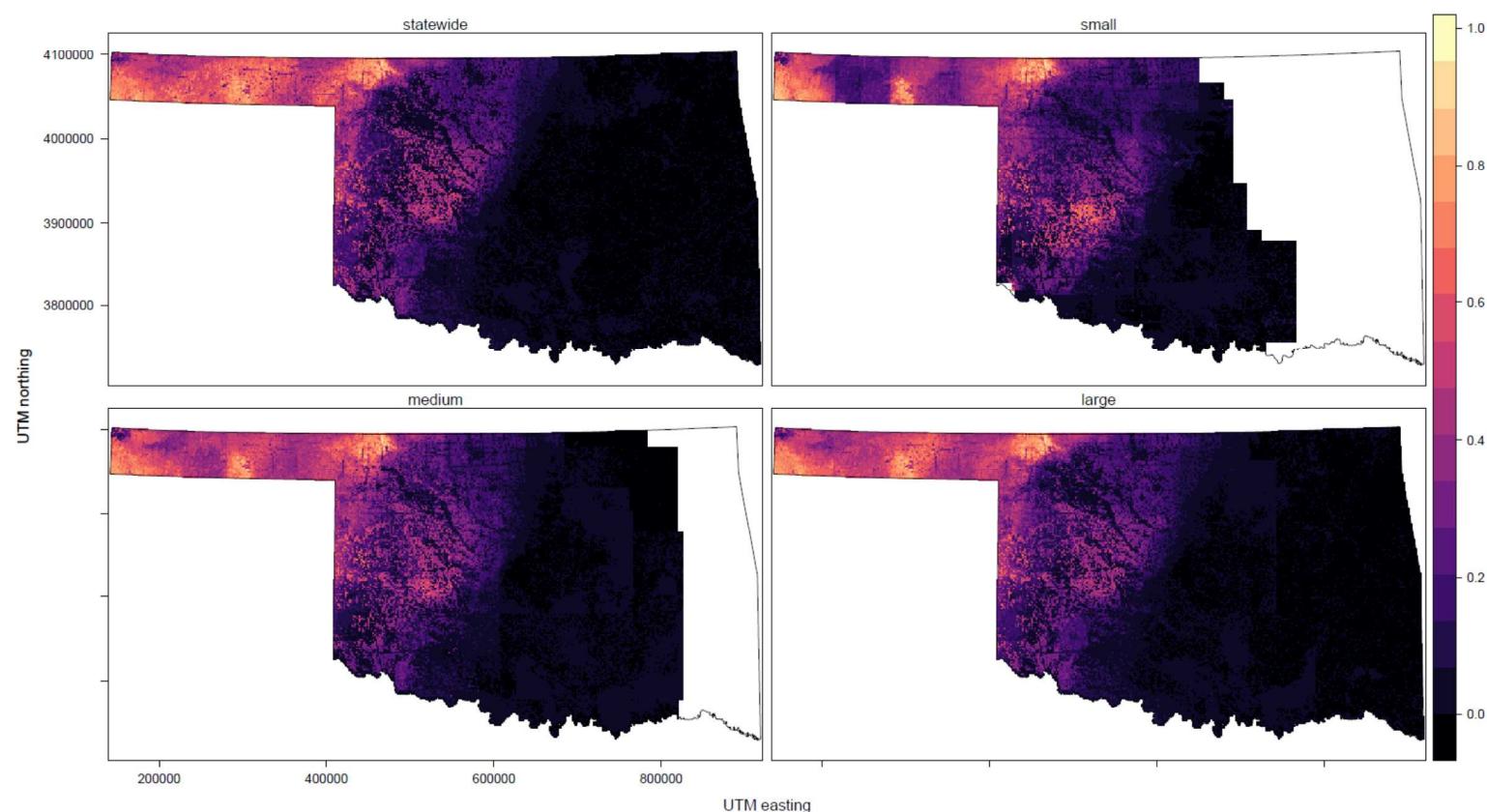
1
2
3 544 Fig. 3. Species distribution model for Northern Bobwhite generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas [\(in white within the state boundaries\)](#) were not able to calculate a model.
6
7



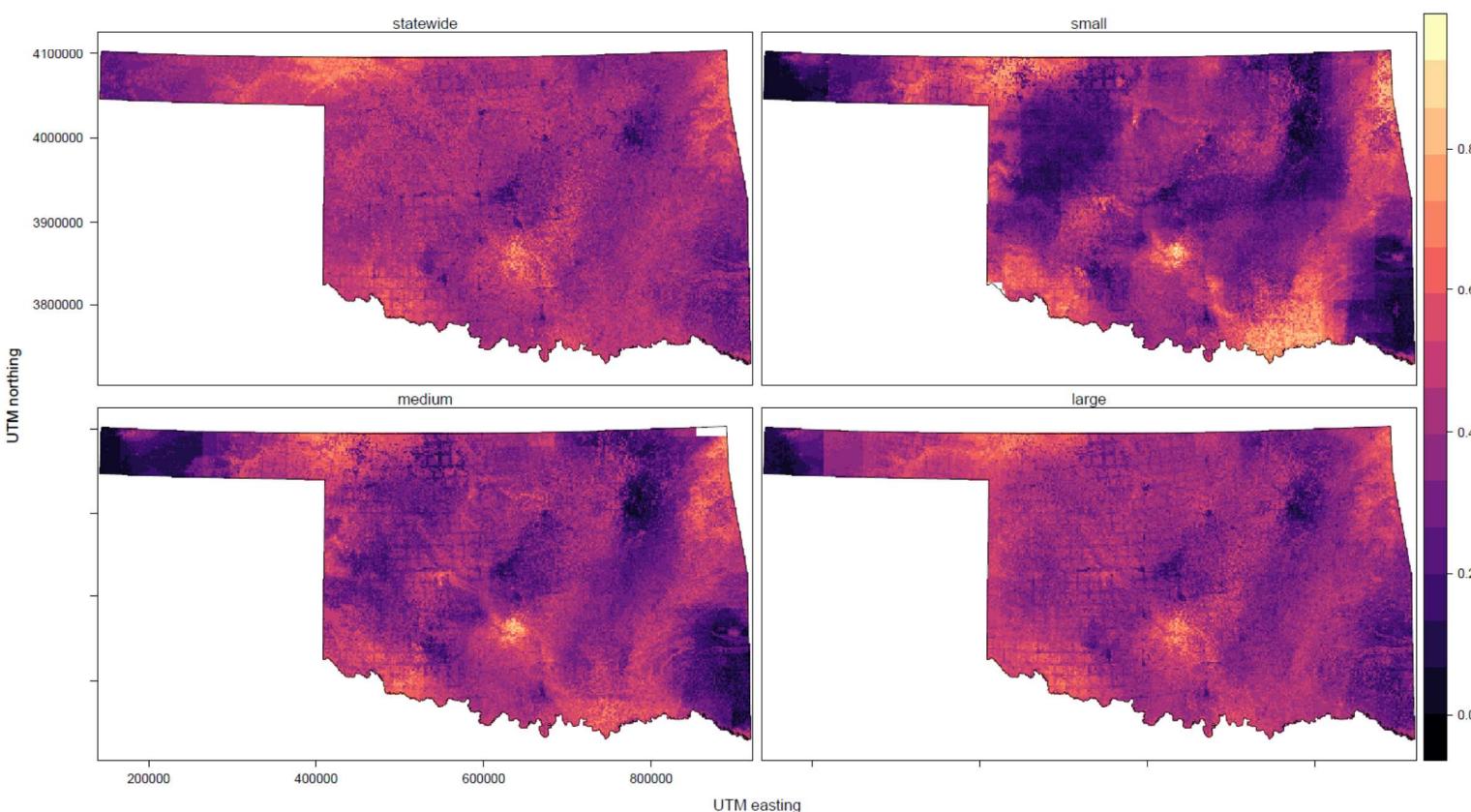
1
2
3 548 Fig. 4. Species distribution model for Cassin's Sparrow generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas [\(in white within the state boundaries\)](#)-were not able to calculate a model.
6
7



1
2
3 552 Fig. 5. Species distribution model for Western Meadowlark generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas ([in white within the state boundaries](#)) were not able to calculate a model.
6
7

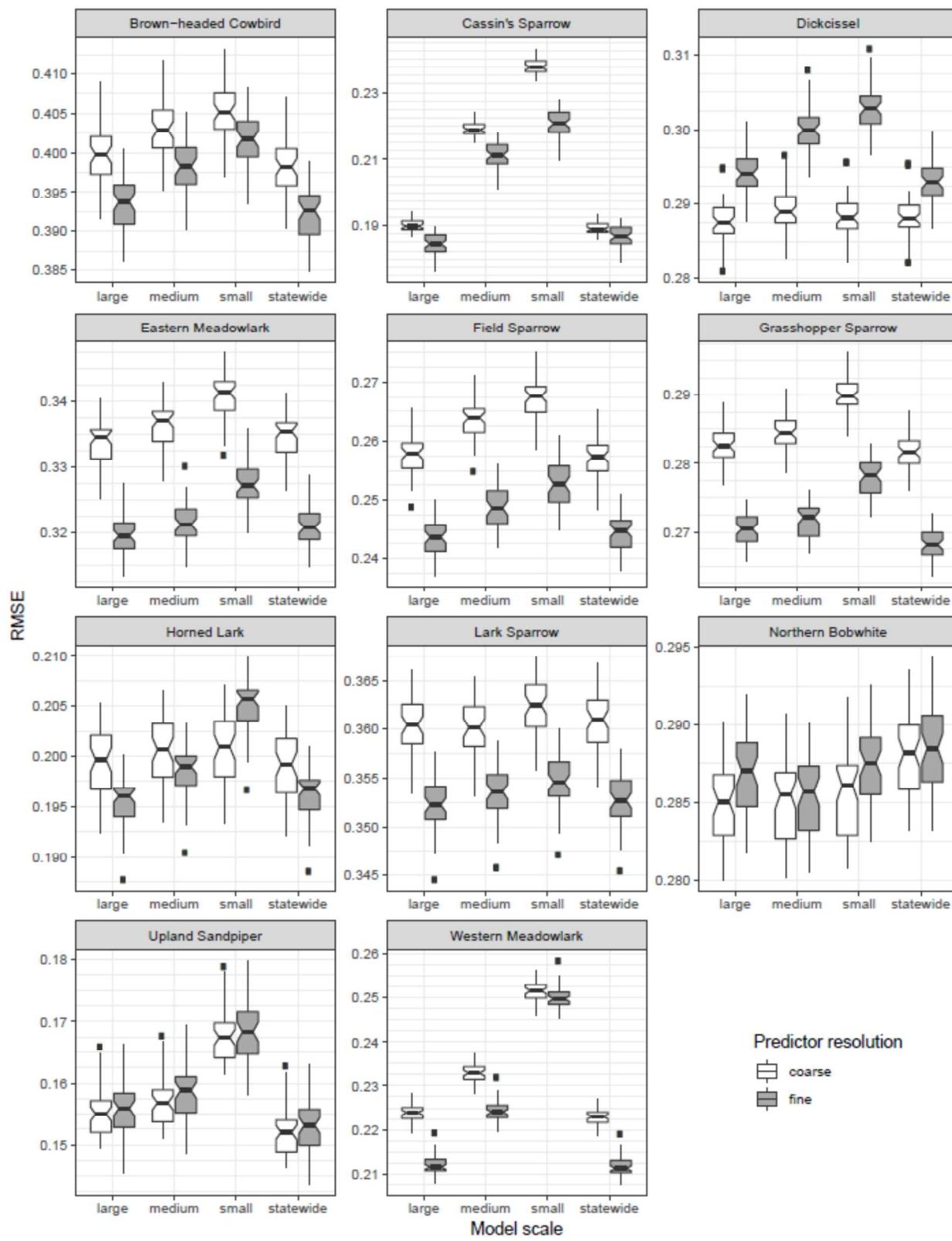


1
2
3 556 Fig. 6. Species distribution model for Brown-headed Cowbird generated at four scales (statewide and three spatially explicit ensemble
4 models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of
5 occurrence from 0-1. Blank areas ([in white within the state boundaries](#)) were not able to calculate a model.
6
7



1
2
3 560 Fig. 7. RMSE evaluations for all 44 models compared by predictor resolution. Each panel shows
4 561 one species. Overlapping notches on boxplots show no difference; non-overlapping notches
5 562 show a significant difference in medians. Center line represents median. Fine grid lines are
6 563 shown to facilitate notch comparison.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

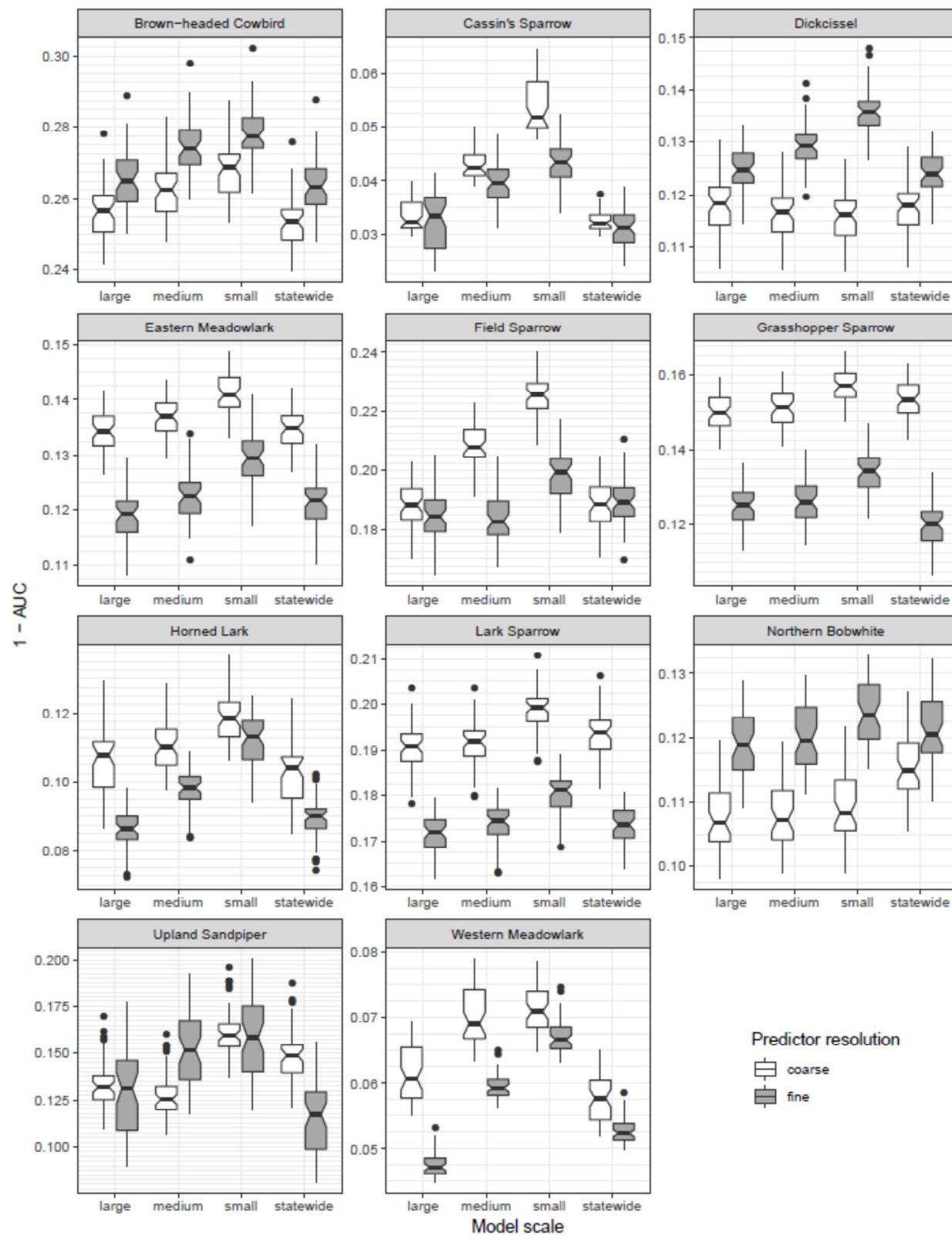
For Review Only



1
2
3 565 Fig. 8. AUC evaluations for all 44 models compared by predictor resolution. Each panel shows
4 one species. AUC = 0.5, where prediction is random, and above which prediction is better than
5 random. We show the y axis as 1 – AUC so that a lower value is better prediction to facilitate
6 comparison with RMSE in Fig. 7. Overlapping notches on boxplots show no difference; non-
7 overlapping notches show a significant difference in medians. Center line represents median.
8 569
9 570 Fine grid lines are shown to facilitate notch comparison.

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only



1
2
3
4 Associate Editor Comments to Author:
5
6

7 Associate Editor
8

9 Comments to the Author:
10

11 We have received a review on this revised version of your manuscript from one of the original
12 reviewers. They find your re-worked version to be much improved, an opinion I share. I think
13 addressing the remaining comments they have should be relatively straightforward and I would
14 anticipate accepting a revised version of your paper.
15

16 Response: We appreciate the positive remarks upon our revision and hope that the additional
17 changes will address the remaining comments.
18

19 Reviewer(s)' Comments to Author:
20

21 Reviewer: 2
22

23 Comments to the Author
24

25 Ecology and Evolution
26

27 ECE-2018-01-00114, Revision September 2018
28

Title: SPATIALLY EXPLICIT ENSEMBLE MODELS DO NOT ALWAYS IMPROVE SPECIES
29 DISTRIBUTION MODEL ACCURACY
30

31 Summary:
32

In this revision the authors have improved several technical aspects of the study including the
33 test-set sampling and model evaluation. They have also improved the description of the
ensemble of support set and how they are created.
34

35 Response: We thank you for the positive comments on our revised manuscript.
36

37 When do you need SEEM / STEM?
38

The manuscript does not accurately and completely communicate when these models are
needed and how to test for this. This is important so that readers can assess these methods in
other use cases.
40

41 In general, non-stationarity in predictor-response relationships (e.g. environment-occurrence or
environment-abundance relationships) become a concern when the:
42

- 43 1) The ratio of the size of the study extent to the size of the predictor resolution or grain is
relatively large, and
44
- 45 2) When the underlying processes driving species occurrence or abundance are expected to
vary at spatial &/or temporal scales between the size of the study extent and the resolution of
the predictors.
46

47 The introduction needs to be edited to communicate this information. Also, regional variation in
48 the predictors (e.g. land cover and climate) by itself is not enough to create non-stationarity in
49 predictor-response relationships. These lines in the introduction and discussion suggest that
50 variation in predictors alone is sufficient (e.g. Lines 59 -60, 324-25) and need to be edited.
51

52
53
54
55
56
57
58
59
60

1
2
3 Response: We have added wording in lines 47, 60-61, 101, and 329-330 to clarify that these
4 relationships and that variation in predictors alone is not sufficient.
5

6 Test for non-stationarity:
7

8 The authors are correct that fitting SEEM / STEM type ensembles requires additional
9 computational cost. So, it is natural to ask when is that extra expense worth it. It is easy to test
10 for spatial and temporal non-stationarity in predictor-response relationships: If the analyst
11 suspects non-stationarity is an issue, it can be tested by a small number of base models, each
12 fit different regions &/or seasons and then comparing the predictor-response relationships to
13 see if they vary. The discussion or introduction should describe this test to help equip readers
14 with the tools to evaluate when & where they need to account for non-stationarity in predictor
15 response relationships. This test does not require any expertise, beyond that necessary to run a
16 species distribution model, and it will only cost relatively little in terms of extra computations.
17 Lines 309-10 need to be revised accordingly.
18

19
20 Response: We have added wording in that paragraph to suggest this test in lines 318-320.
21

22
23 Creating Random Support Sets:
24

25 It's good to see more information about the ensemble of randomized support sets and their
26 creation.

- 27 • While randomization of the support set center is an important part of creating the ensemble
28 of support sets, the goal for constructing a good ensemble of base models is to fit low bias, high
29 variance models (Kuncheva & Whitaker 2003). This should be stated as the goal in line 235.
30

31 Response: We have edited now-lines 237-238 accordingly.
32

- 33 • Bootstrapping or subsampling is used to induce sampling variation among base models in
34 the ensemble. Thus, simply relying on the bootstrapping that is done within the base model will
35 not induce variation across base models as is implied in lines 230-231. These lines need to be
36 edited accordingly.
37

38 Response: We have edited line 232 to reflect this.
39

- 40 • Ensemble averaging Is used to reduce the sampling variation across base models. Thus, it
41 is important to average across a sufficient number of base models. If I understand the
42 description, the maximum number of randomized support sets covering any single prediction
43 was 10 or 11 (depending on the size of the support set) and on average closer to 7. This is an
44 extremely small number for this kind of averaging and can be seen in the blocky edges of the
45 estimated distribution maps. The number of base models should be at least 25, if not greater.
46

47 Response: We have added wording in lines 243-245 to suggest that future users of these
48 models repeat the sets more frequently to reduce the blockiness. We lack the funding to repeat
49 the models with larger sets of base models and would appreciate any suggestions on citations
50 to direct readers for how to choose an appropriate sampling size threshold to reach a suitably
51 low level of blockiness.
52

1
2
3
4 Figures – Blank areas should be shown in a different (non-black) color so that they can be
5 distinguished from predictions of 0.
6

7 Response: They are currently shown in white (blank) and black (0), and we have clarified all
8 captions so that this is clearer.
9

10 Line 218 - “SEE Models”
11

12 Response: Done.
13

14
15 Line 259 - Remove “1983)”
16

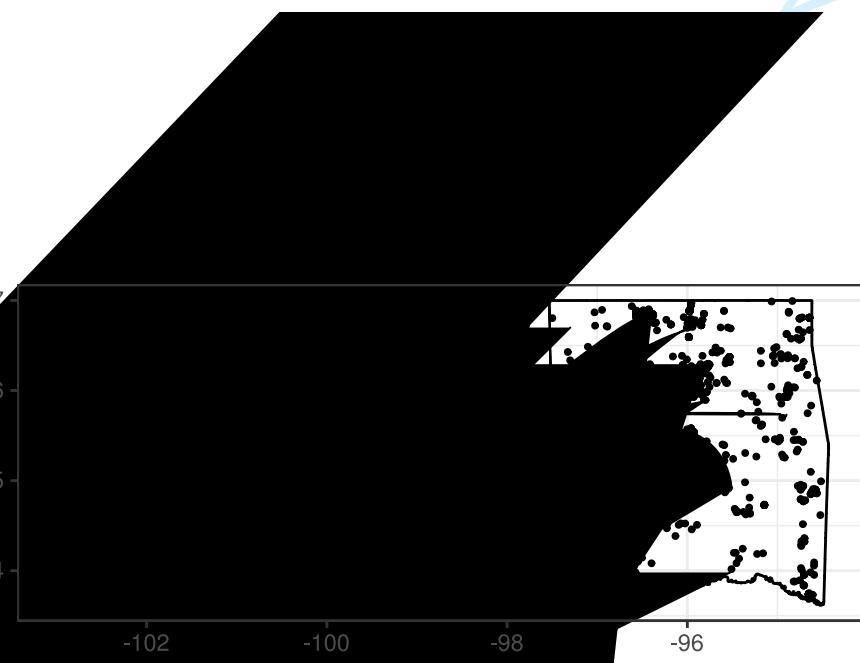
17 Response: Done.
18

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

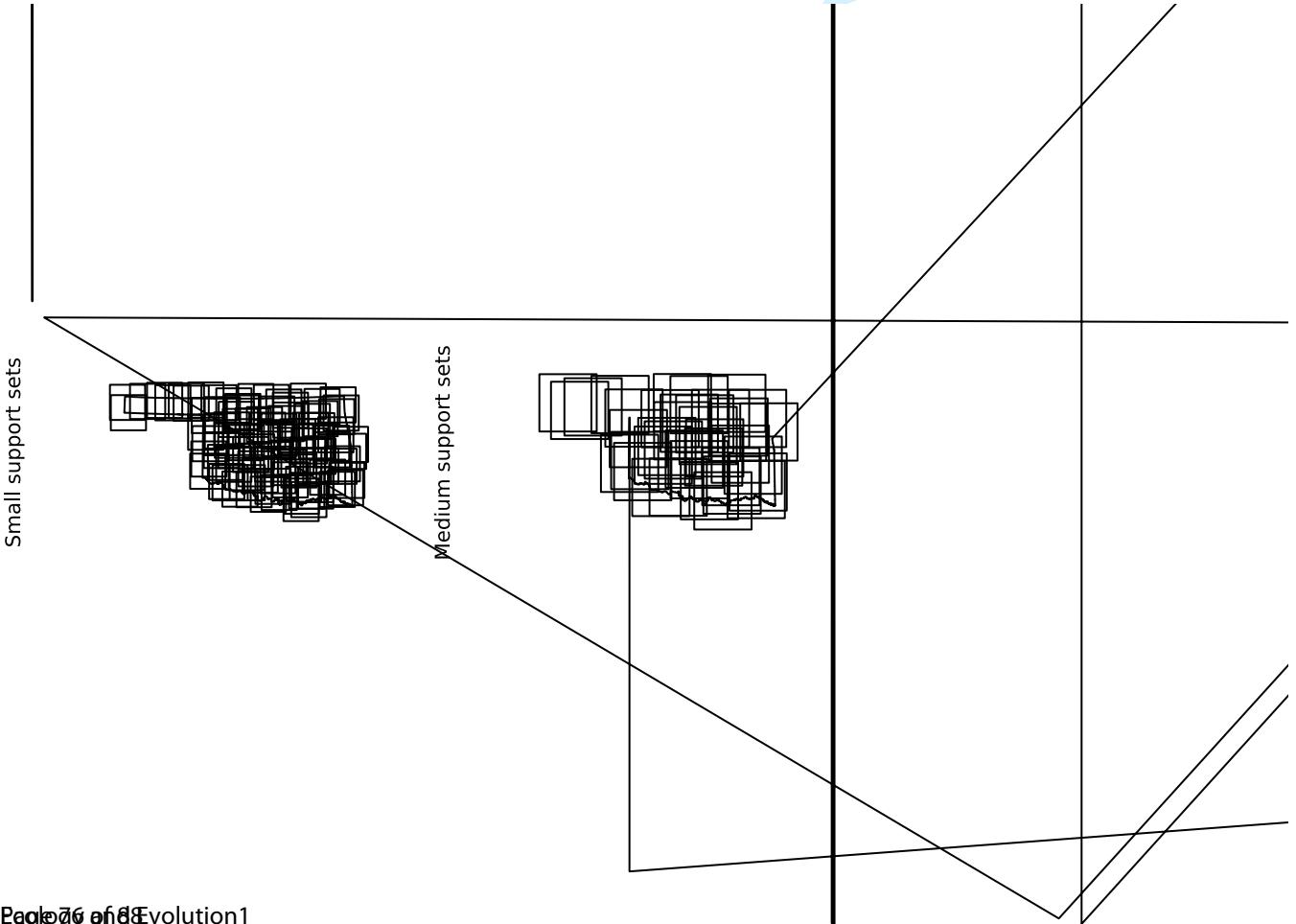
For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

For Review Only



For Review



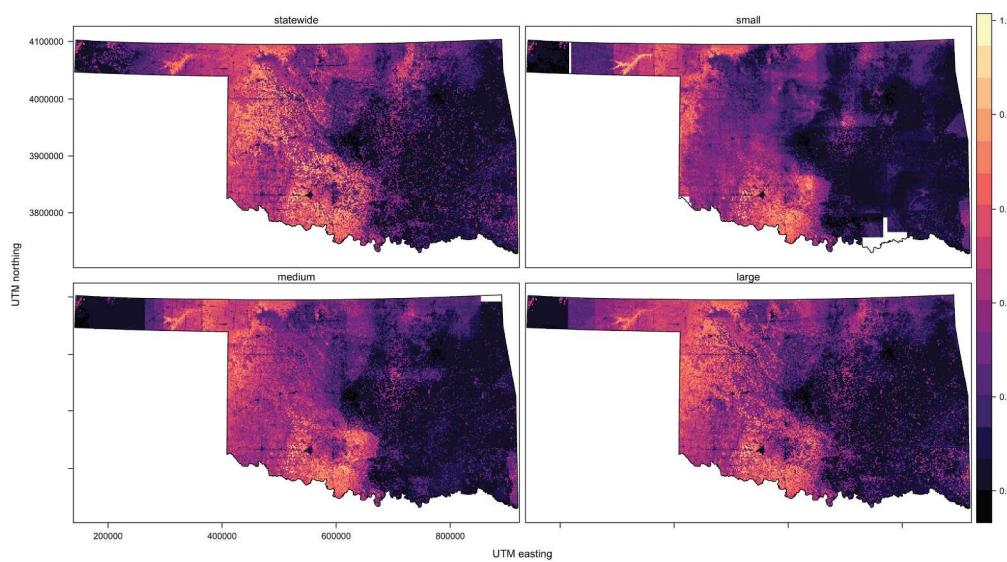


Fig. 3. Species distribution model for Northern Bobwhite generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

219x119mm (300 x 300 DPI)

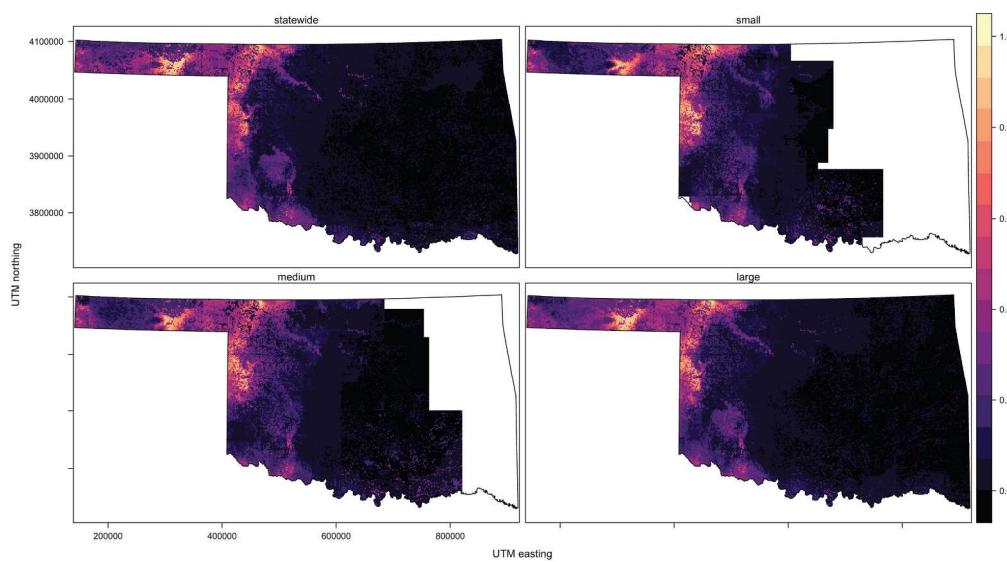


Fig. 4. Species distribution model for Cassin's Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

219x119mm (300 x 300 DPI)

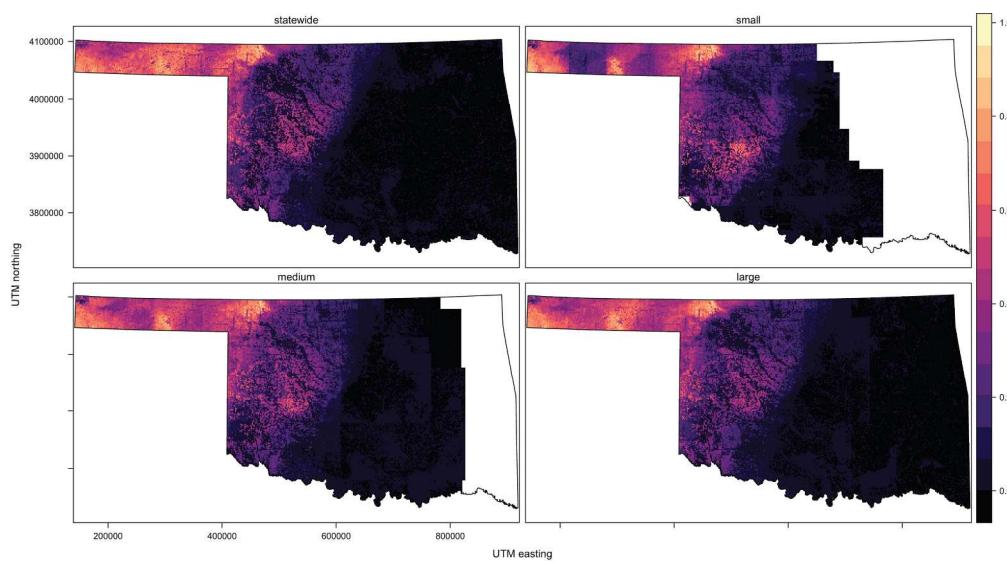


Fig. 5. Species distribution model for Western Meadowlark generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

219x119mm (300 x 300 DPI)

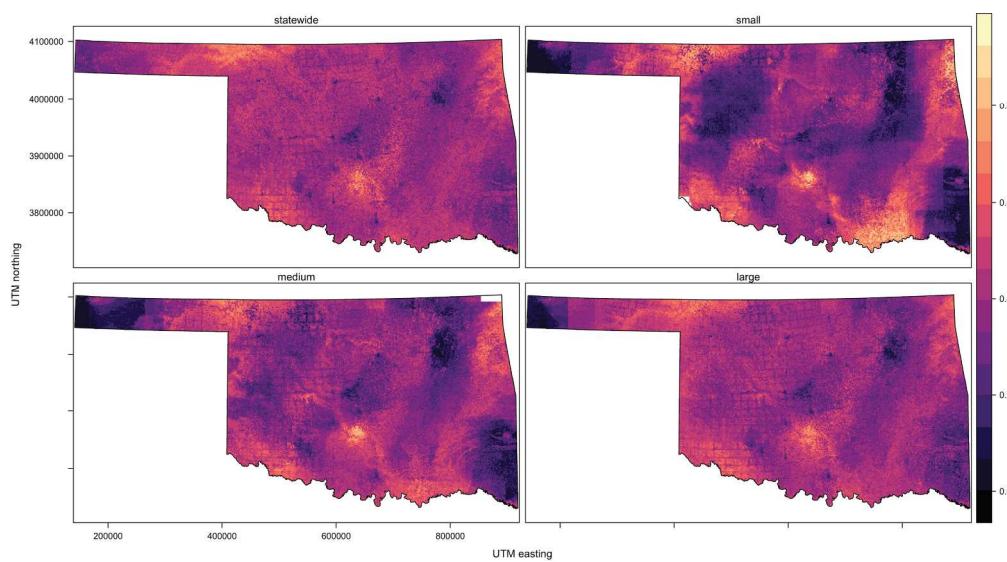


Fig. 6. Species distribution model for Brown-headed Cowbird generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white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

219x119mm (300 x 300 DPI)

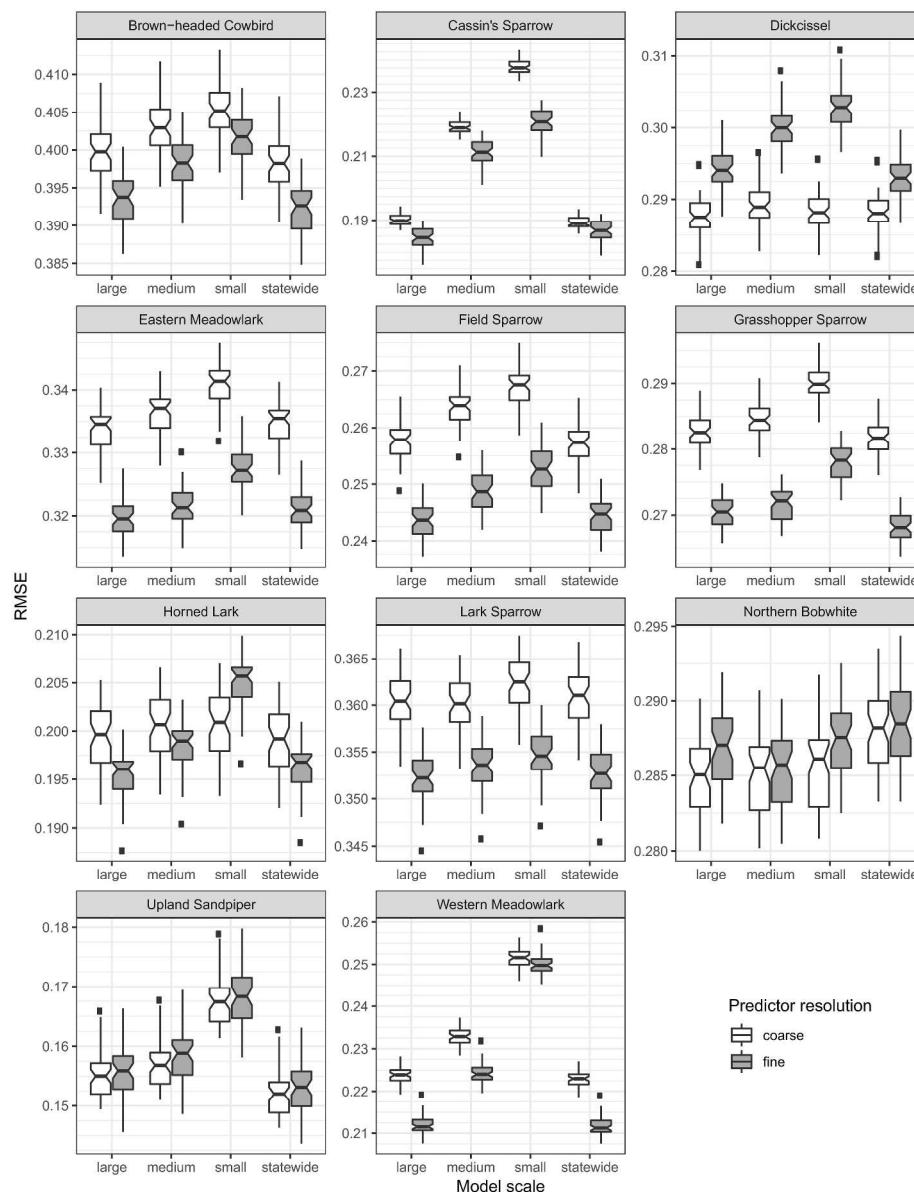


Fig. 7. RMSE evaluations for all 44 models compared by predictor resolution. Each panel shows one species. Overlapping notches on boxplots show no difference; non-overlapping notches show a significant difference in medians. Center line represents median. Fine grid lines are shown to facilitate notch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

279x361mm (300 x 300 DPI)

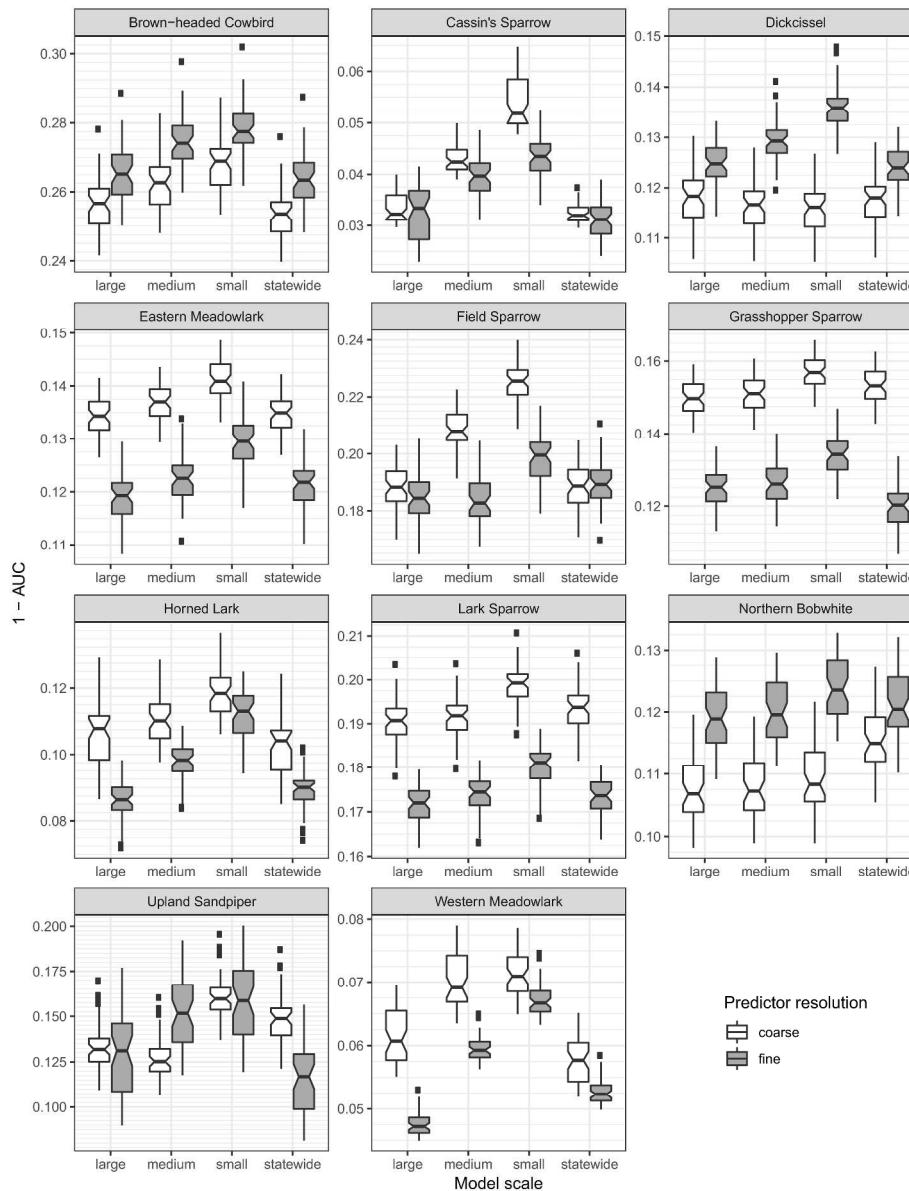


Fig. 8. AUC evaluations for all 44 models compared by predictor resolution. Each panel shows one species. $\text{AUC} = 0.5$, where prediction is random, and above which prediction is better than random. We show the y axis as $1 - \text{AUC}$ so that a lower value is better prediction to facilitate comparison 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.

279x361mm (300 x 300 DPI)

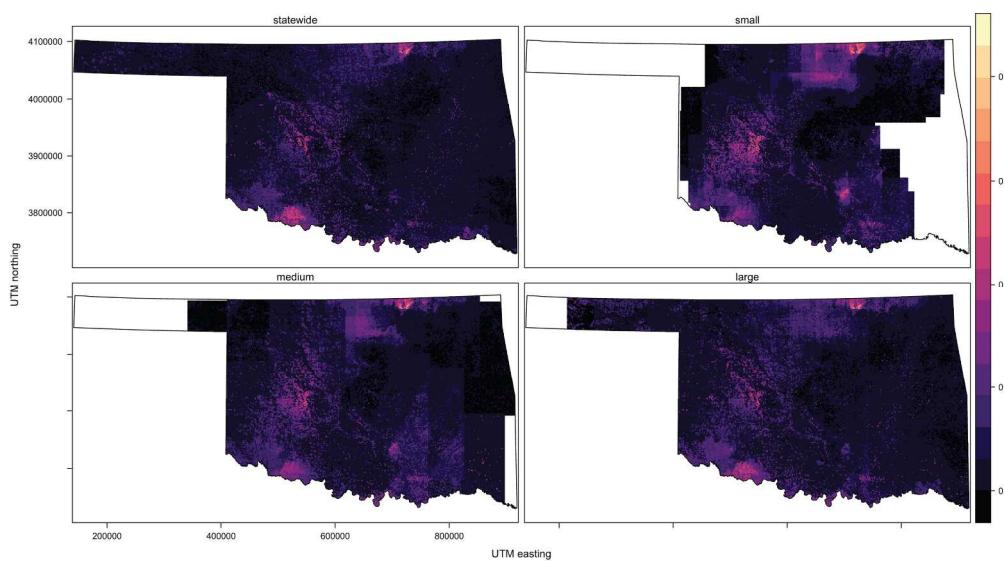


Fig. S1. Species distribution model for Upland Sandpiper generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

219x119mm (300 x 300 DPI)

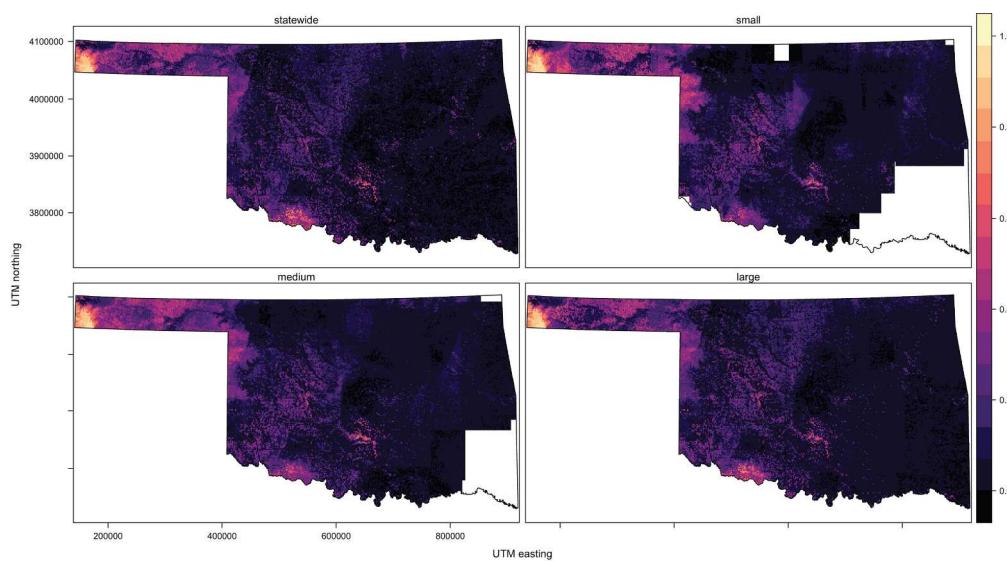


Fig. S2. Species distribution model for Horned Lark generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

219x119mm (300 x 300 DPI)

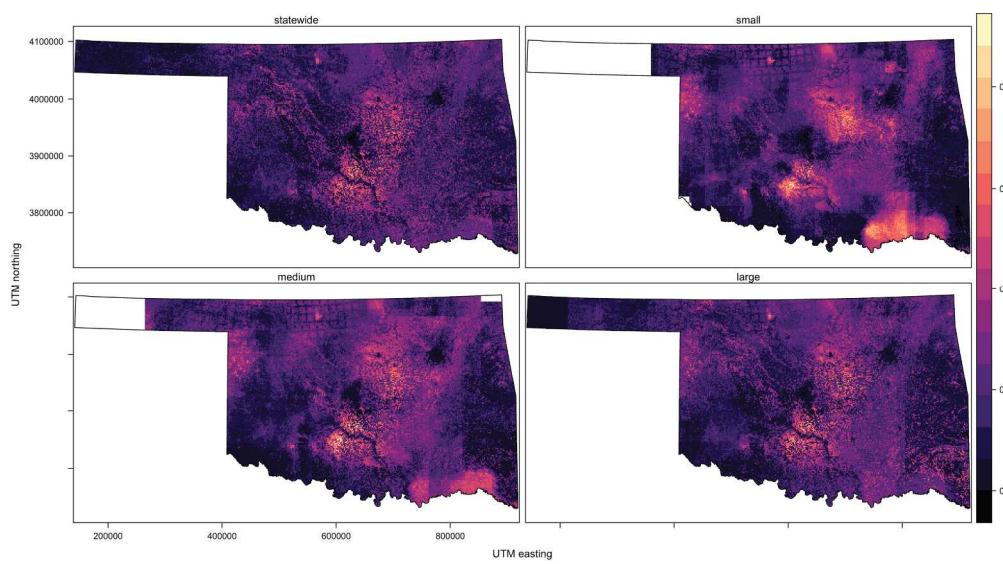


Fig. S3. Species distribution model for Field Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

219x119mm (300 x 300 DPI)

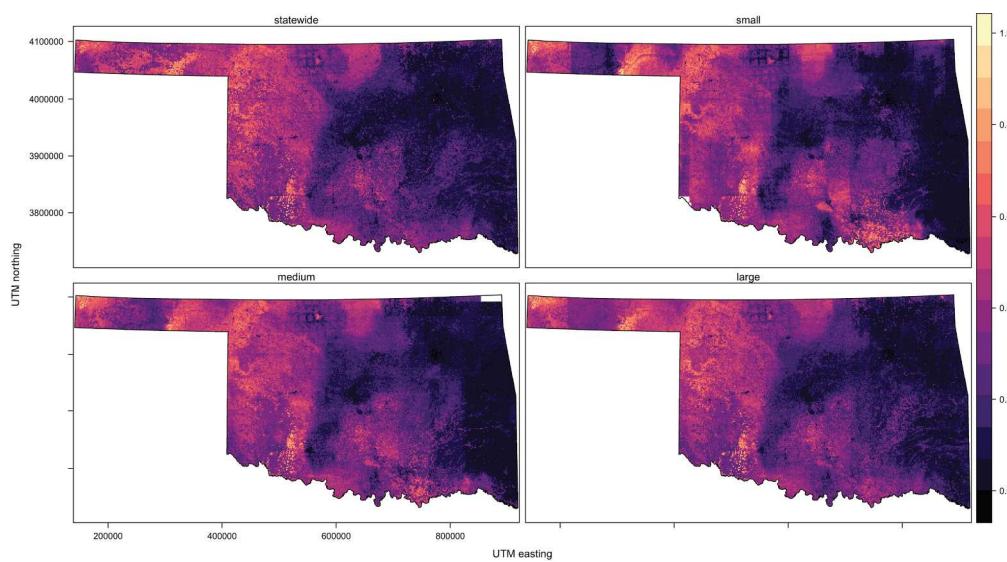


Fig. S4. Species distribution model for Lark Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within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

219x119mm (300 x 300 DPI)

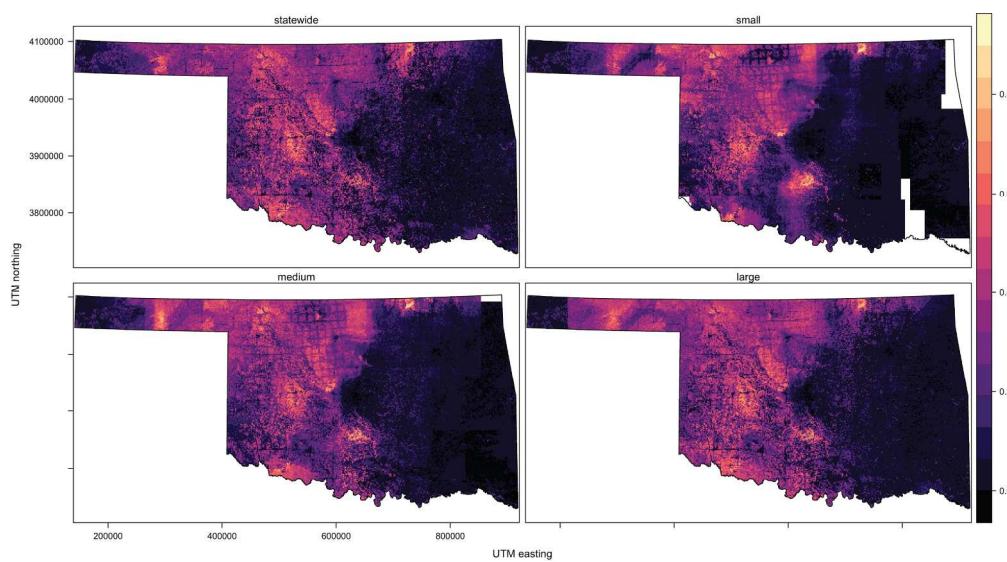


Fig. S5. Species distribution model for Grasshopper Sparrow generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white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

219x119mm (300 x 300 DPI)

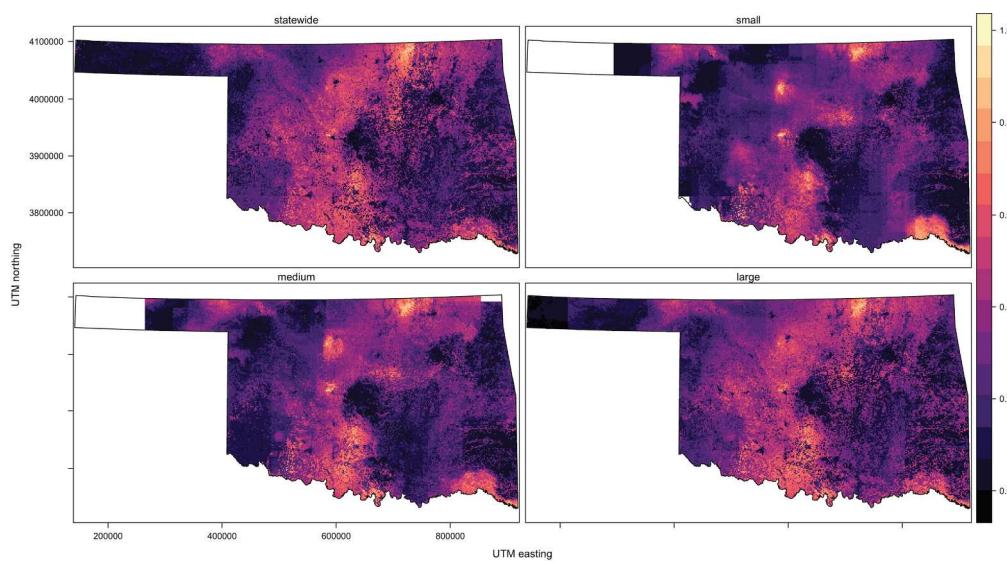


Fig. S6. Species distribution model for Dickcissel generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white within the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

219x119mm (300 x 300 DPI)

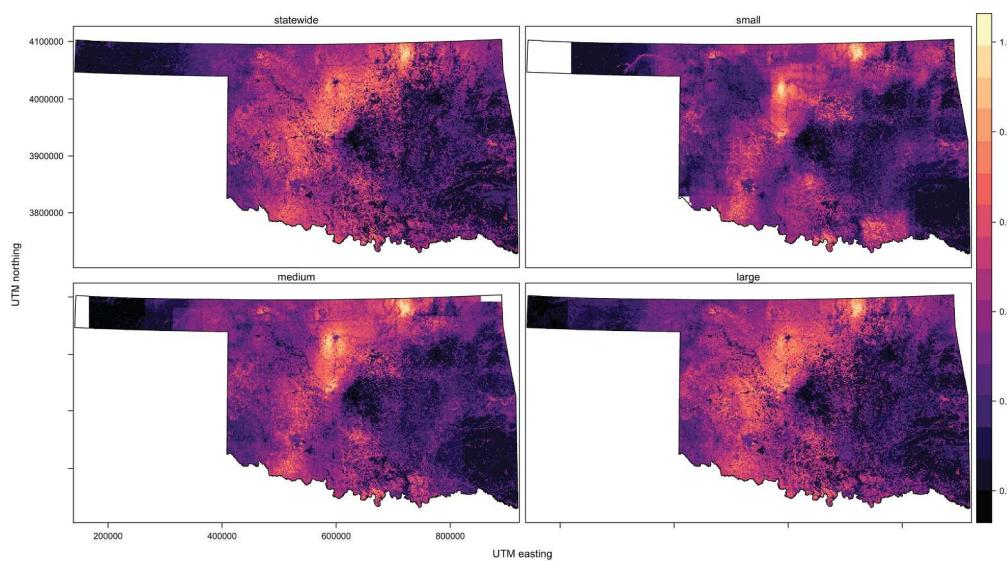


Fig. S7. Species distribution model for Eastern Meadowlark generated at four scales (statewide and three spatially explicit ensemble models at large, medium, and small support set sizes) with 30 m resolution in Oklahoma. Color scale indicates probability of occurrence from 0-1. Blank areas (in white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

219x119mm (300 x 300 DPI)