# Introduction

Grasslands are very endangered, with only ?% remaining. Rate of development is… ? Grassland report (Askin et al?). WRITE MORE HERE ABOUT GRASSLANDS and WHICH TYPES ARE THREATENED. The already tenuous status of grassland birds is now further threatened by conversion to new crops resulting in permanent land use changes (Wright and Wimberly 2013), generational changes in land use (Higgins et al. 2002), changes in conservation programs for grassland habitats (Klute et al. 1997), and climate change (citation). Expand on these topics.

Understanding impacts of changing land use and climate in a region with much ecosystem variety requires a dynamic approach to species distribution modeling. Species distribution modeling uses predictors, usually climatic variables () but now expanding to biotic variables (), to predict what areas are most suitable for a given organism. A new method called Spatio-temporal Exploratory Modeling adds additional accuracy by using an ensemble of local models to account for differing variable importances in different regions. It has been used at a national scale (Fink et al. 2010 and 2013), but we will compare its effectiveness with typical SDM at a smaller regional scale.

Our study region, the U.S. state of Oklahoma, contains a wide variety of grassland birds as its ecoregions range from tallgrass prairies in the east on the edge of eastern deciduous forests to several types of grasslands in the central part of the state and westward. Agriculture accounts for over $2.8 billion in the state’s gross domestic product in the study years (U.S. Bureau of Economic Analysis, 2017). Major crops in the state use. Land use is also for ranching, with ?? animals. State plans for biofuels GOOGLE. This combination of agricultural importance and impact by climate change makes Oklahoma’s grassland birds vulnerable to a changing world.

Thus, the objectives of our study are to find out the current distribution of Oklahoma grassland birds and understand what variables are important in their distribution. These data will allow managers to make decisions on what areas are important for populations, what land use practices and trends may impact populations, and how climate change interacts with these. We will compare use of a typical SDM and a STEM as well.

* Our specific objectives are to examine:
  + What is the current distribution of Oklahoma grassland songbirds?
    - Density estimates use point count and transect distance sampling
      * Compare estimates from transect and PC sampling (PC along roads, transects cross-country)
    - Species distribution/STE model maps
  + What landcover (including crops, conservation easements?), vegetation (from 2014 transects only), and climatic variables predict the distributions of the study species?
    - Response variables: Use presence/absence (from our surveys and from ebird) in species distribution models
    - Predictor variables: climate (bioclim/worldclim), vegetation types (NASS crop raster layer includes switchgrass and other crop types), 2014 transect vegetation surveys
    - Compare STEM and statewide models
  + How will distributions of selected species move with climate change and land use changes?
    - Distribution changes with predicted climate change (bioclim/worldclim predict layers)
    - IF CAN FIND DATA: predicted landuse/crop cover changes, using soil types possibly (predict where switchgrass and other crops can be grown? Found very detailed soil types maps). Need to inquire with Todd.
      * Downloaded a paper by Nicky et al. on predicting land use changes (based on soil maps) for Sprague’s pipit that might be a good model to use here.

# Methods

## Study area

Brief discussion of ecoregions in Oklahoma, range of precipitation and temperature across the state, and what types of grasslands (and what areas exist) are here.

## Response data

* Survey methods
  + Point counts
  + Transects
* Citizen science data
  + eBird: All complete data (points and transects) for 2013 and 2014.
  + Still waiting on response from ebird about whether the dataset I downloaded is “complete counts” only. Worst-case scenario, download their other dataset, filter by complete, and re-incorporate (probably take ~8 hours if formatting different.) If they don’t email back by Feb. 15 I will try again.
  + Because some observers entered sightings from before and during our surveys into eBird.org, we eliminated counts that were within two hours of the actual survey start time and within 15 km of the survey start location. This eliminated ?? of ?? ebird entries.

## Predictors

We used climatic variables, land use variables, and land cover variables to predict bird distribution. Table 1 shows the variables and their sources and resolutions.

* Bioclim (get through R or from website)
* Data I have downloaded
  + NRCS Conservation Easement Areas by State  
       Size: 0.40 megabytes (46 files).  Download compressed size: 0.19 megabytes (1 map).  
       <http://gws.ftw.nrcs.usda.gov/GWDL/3276698/easements_EASEAREA_ok_3276698_01.zip>  
     National Land Cover Dataset  by State  
       Size: 35.18 megabytes (7 files).  Download compressed size: 29.96 megabytes (1 map).  
       <http://gws.ftw.nrcs.usda.gov/GWDL/3276698/land_use_land_cover_NLCD_ok_3276698_02.zip>  
     Cropland Data Layer by State  
       Size: 235.53 megabytes (3 files).  Download compressed size: 235.57 megabytes (1 map).  
       <http://gws.ftw.nrcs.usda.gov/GWDL/3276698/land_use_land_cover_NASS_CDL_ok_3276698_03.zip>

Additionally, we used neighborhood predictors about the values in rectangular areas around each point. We did it at the scale of 5 x 5 pixels (150 x 150 m) and 15 x 15 pixels (450 x 450 m) (Fink et al. 2010). These were proportion of a given land cover type from the NLCD 2011 raster. We looked at proportion of each land cover class and proportion of several summed variables: open space (grasslands, hay/pasture, cropland, herbaceous wetlands, and barren land) since grassland bird occupancy can be influenced by the total non-structural cover (McDonald 2017). We also used NLCD’s canopy cover and impervious surface layers.

## Analyses

### Density estimations

* Using distance sampling, possibly including detectability from repeated surveys
  + Comparison of point count vs transect effectiveness if sample size large enough for each and geographical overlap sufficient. However, point counts go along road and transects usually walking off-road. Alternative: comparison of estimations from road pcs vs “off road” transects?

### Species distribution models

Ensemble and spatio-temporally explicit ensemble models

To model species distributions based on our predictors, we created two sets of models for each species. The first is a statewide ensemble model using the base models which are known to give good predictions. This gives us interpretable models for which we can make specific predictions about what predictor variables are influencing distribution in what ways for each piece of the ensemble. The second are spatio-temporally weighted ensemble models (Fink et al. 2010). This second model, while it may give more accurate predictions, is harder to interpret (James et al 2013 ISLR book). Both strategies give us differing and complementary information on factors affecting species distribution in Oklahoma.

Both ensembles compare models by weighting averages of each single model prediction. We weighted each pixel by the sample size of models at each pixel. [Oppel et al. 2012 weighted each model by AUC but I’m not sure we need to.]

The statewide ensemble model for each species consists of a bagged decision tree, ?, and ?. These base models can each be interpreted. Ensembling predictions for all ? models is known to give more accurate predictions (citation).

The spatiotemporally explicit ensemble models include all ? types of base models, but merged over different spatial extents per Fink et al. 2010. The original STEM was designed for broad-scale survey data. We see whether it is useful at a smaller scale by adapting the scale of our support sets. With the diverse habitats and climatic variables found across Oklahoma, it should provide better predictions than the statewide model.

Models were evaluated by ??

To determine which predictors were important in species distributions, for each species we ranked variables. Hochaka et al 2007 article lists citations of Breiman 2001, Brieman et al. 1984, and Caruana et al. 2006 of how to rank important variables. GAMS are additive, adding up each line for each variable (doing a smoothing line for each one). Maybe not the best for wanting interactions because it doesn't do interactions. MaxEnt (generates curves for each type of thing). Need to read downloaded machine learning books more too.

### Study species

Table ? shows the species we analyzed.

# Results

Map for STEM-type, basic model, and distance sampling estimates for each species.

# Discussion

# Notes to self

* Data I have downloaded but not used at this time
  + Gridded Soil Survey Geographic (gSSURGO) by State  
       Size: 952.32 megabytes (4 files).  Download compressed size: 952.46 megabytes (1 map).  
       <http://gws.ftw.nrcs.usda.gov/GWDL/3273245/soils_GSSURGO_ok_3273245_01.zip>  
     Major Land Resource Areas by State  
       Size: 1.35 megabytes (46 files).  Download compressed size: 1.00 megabytes (1 map).  
       <http://gws.ftw.nrcs.usda.gov/GWDL/3276698/soils_MLRA_ok_3276698_05.zip>  
     Common Resource Areas by State  
       Size: 1.28 megabytes (45 files).  Download compressed size: 1.03 megabytes (1 map).  
       <http://gws.ftw.nrcs.usda.gov/GWDL/3276698/soils_CRA_ok_3276698_06.zip>
* Survey vegetation data from 2014 transects (none from 2013 or point counts in 2014?)… only can be used for that year and for 1/3 of 2014 transects. Not sure what it can be used for in this context of state-wide multiple years.
* Data I need to find if exists:
  + Forecast changes in landuse in OK
    - <http://tethys.dges.ou.edu/main/?cat=12>

Links for ensemble model making:

* + - How to implement
      * <http://machinelearningmastery.com/non-linear-classification-in-r-with-decision-trees/>
      * <https://cran.r-project.org/web/packages/ipred/vignettes/ipred-examples.pdf>
      * <https://cran.r-project.org/web/packages/adabag/adabag.pdf>
      * <https://onlinecourses.science.psu.edu/stat857/node/181>
      * <http://mlwave.com/kaggle-ensembling-guide/>
      * Simple averaging ensemble pseudocode: <http://www.kdnuggets.com/2016/02/ensemble-methods-techniques-produce-improved-machine-learning.html>
      * using caret to assemble ensembles?? <http://amunategui.github.io/blending-models/>
      * <http://www.overkillanalytics.net/more-is-always-better-the-power-of-simple-ensembles/>: has code, I think I can start from this.