# Introduction

* Background
  + Grasslands are very endangered, with only ?% remaining. Rate of development is… ? Grassland report (Askin et al?). Threats include agriculture and biofuels? Conservation easements?
  + Oklahoma is a state ranging from forests in the east to several types of grasslands in the central part of the state and westward. Agriculture is big, including xyz types of crops, and potential for biofuels and ? additional crop development.
  + Climate change is additionally forecast to affect Oklahoma in xyz ways. 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.
* 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
  + 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). Will be inquiring with Todd (Andrea says he probably has some data on this.) If cannot find maps, then the coefficients from models??

# Methods

## Study area

Brief description of Oklahoma vegetation and climate.

## Response data

* Survey methods
  + Point counts
  + Transects
    - lengths are not even. Some transects longer than others.
* eBird data
  + All complete data (points and transects)
  + Still waiting on response from ebird about whether the dataset I downloaded is “complete counts” only

## Predictors

* 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 useful?
* Bioclim (get through R or from website)
* Data I need to find if exists:
  + Forecast changes in landuse in OK
    - <http://tethys.dges.ou.edu/main/?cat=12>
* 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>

## 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

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 models is 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.

* Ensemble models
  + Compares models by weighting averages of each single model prediction “with weights assigned to each modelling technique based on its discriminatory power as measured by the area under the receiver-operated characteristic curve” (Oppel et al. 2012, seabird paper).
  + adaSTEM/STEM models
    - STEM is fixed model. In adaSTEM handout, they use GAM as base models and also linear models. So, I can work on making the STEM framework with ANY TYPE of model (though I don’t know if I can mix them). Unsure if can do with multiple types of models, like in Oppel paper? They had two regular models (linear and additive) and three machine learning but did not do spatiotemporal adaptive aspect. STEM is type of ensemble model with different bases, unsure if can incorporate multiple model types as bases.
    - Ensemble models of decision trees, used with “bagged decision trees” (a type of classification tree) as base models trees in Fink et al paper
    - 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.

# Results

# 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>