Vegetation\_Classification

steppe

2022-10-03

The initial AIM sample design for each office was based upon small stratified areas - (Primary Sampling Units) PSU’s which up to 3 plots per each of are to make inference of. The primary sampling units were based on older vegetation classification projects which do not perform as high as other mechanisms.

We have used the in field AIM data to reclassify a number of plots in unusual strata such as: ‘Other’, ‘Mixed Conifer’, and ‘Riparian’. We will use all of this reclassified AIM data to determine whether the PSU on the whole is classified as the same vegetation type as the points which were sampled within it.

To do this a very simple vegetation classification will be performed using a handful of spatial products. The central product will be the National Aerial Imagery Programs flight data. Supplemental data includes a 10m DEM which was previously re-processed using WhiteBoxTools, a 10m resolution slope dataset, and a 10m landform classification dataset.

# Methods

Colorado 2019 NAIP Imagery were downloaded from the official repository at Box in Fall 2022. While decoding from MrSID to tif file formats, their resolution was reduced by a factor of four using the ‘mrsiddecode’ program (Vers. 9.5.1.4427) from LizardTech. This effectively reducing their resolution to 9.6 meters. Following decompression of MrSID data, all spatial data processing occurred using R version 4.2.1, all computing performed on linux Ubuntu (20 & 22) on multiple hardware. These raster tiles were united via mosaic, cropped to the extent of the Field Office’s ownership, and masked to BLM administered surface areas. These data were then aligned with previously generated raster data sets derived from a 10m DEM; when required all re-sampling of these tiles were achieved using cubic spline interpolation.

A Gray Level Co-occurrence Matrix (GLCM) was created using the glcm package to create a texture raster layer to aid in classification. Texture bands are, among other properties, capable of indicating the amounts of heterogeneity of habitat types across the landscape. Texture layers were produced using both NAIP natural color and infrared data. Texture statistics: mean, variance, homogeneity, contrast, dissimilarity, with windows of 5 in both direction, shifts in all directions (**Queen’s case**), and the default value of 32 grey levels.

NAIP data were processed to create an NDVI band. NDVI is well suited for identifying sparsely vegetated areas, it may be useful in distinguishing salt desert from all other strata, and help in distinguishing between MMS and PJ.

To create a more equally balanced training data set, all classified AIM points were exported to Google Earth. 150 random points were generated across the focal BLM district and classified in Google Earth via the vegetation ecologist which lead the AIM sampling in 2022. Additional records for each stratum, less Aspen forest, without enough points for balanced sampling were found by the vegetation ecologist and marked via Google Earth.

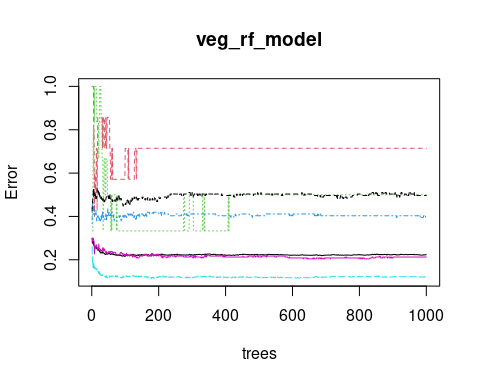
To create a larger training data set, all 469 AIM and LMF points were exported to Google Earth and 440 were classified. 400 random points were generated across the focal BLM district and 377 classified in Google Earth via the vegetation ecologist which lead the AIM sampling in 2022. An additional 885 regularly placed plots were drawn across the extent of the field office and 854 were classified. Unclassified computer generated points were generally those that fell upon a wide road, or were outside BLM Ownership. Unclassified AIM/LMF points were LMF points which must have represented the revisitation of a single plot.

Classified plots were randomly sampled to ensure an equal number of points per stratum, less Aspen and Mixed Conifer.

The dataset of 1657 classified points were partioned into a 0.3 test and training set 0.7 using caret, the dataset was not balanced, see table XX for sample sizes.

Variable Importance in Random Forest Model

| Variable | Importance |
| --- | --- |
| GLCM\_PC2 | 176.028 |
| DEM | 173.738 |
| NDVI | 109.820 |
| Slope | 91.389 |
| NAIP\_blue | 49.249 |
| NAIP\_red | 38.504 |
| GLCM\_PC1 | 32.012 |
| NAIP\_IR | 25.261 |
| NAIP\_green | 22.819 |



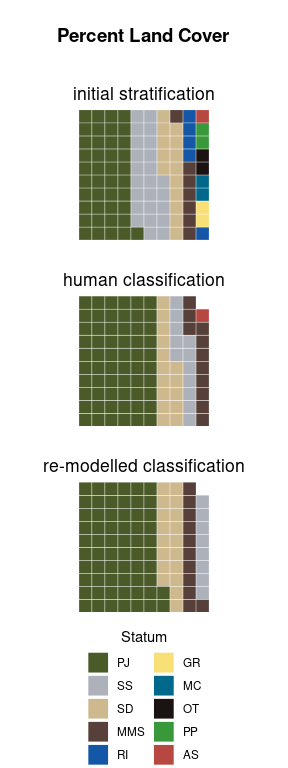
The number of mtry in the random forest model were tuned using the function tuneRF with the number of trys set to 1000, a step factor of 1.5 and a relative minimum improvement in Out of Bag (OOB) error rate set at 0.01. The random forest model was then trained using 4 mtry and 1000 trees, all using the RandomForest package.

While the accuracy for the model was 0.773 (range 0.73 - 0.81), due to an imbalance in the number of observations per group a more appropriate for evaluating the overall performance of the model, is the Kappa metric, 0.628.

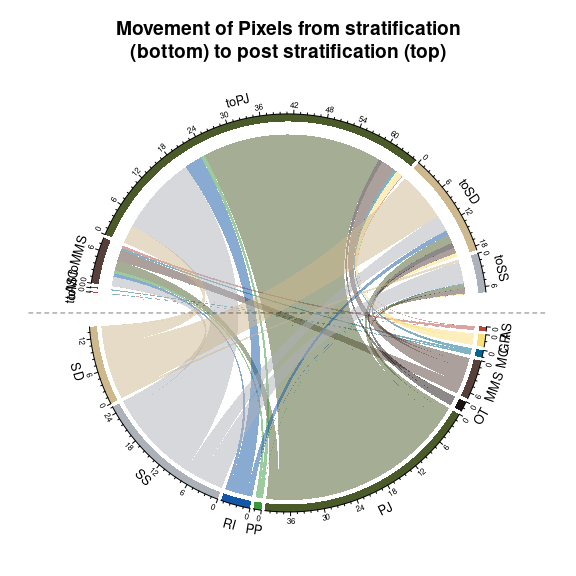
Notably, the model has high rates of Specificity (median = 0.96), showing that when it predicts a vegetation type onto a pixel cell, the prediction is usually correct; less so for prediction of Pinon-Juniper (0.766). However, the model suffers from low Sensitivity, indicating that it is unable to detect all occurrences of a vegetation type. For example, the model is only able to appropriately classify half of the occurrences of Sage Steppe (0.475 ) and Mixed Mountain Shrub (0.596). Accordingly this model is susceptible to over-predicting the occurrence of Pinon-Juniper, at the expense of Mixed Mountain Shrub and Sagesteppe. This is to be expected given the sample imbalance, which contained many more plots PJ than other types. However numerous trials of reducing the number of PJ plots did not significantly increase the quality of predictions. Further indicating, that the features used are not adequate for distinguishing between the transitional points enough Pinons and Junipers are present where SS phases into PJ, and where shrubs increase in PJ and it phases into MMS.

The Random Forest classification model was predicted onto a raster surface using the package Terra. This raster was then smooth using focal statistics, with a window of 5 cells, using the mode as the value to return.

# Comparison of New and Old vegetation classification models.



Changes in Estimated Land Cover between the initial stratification and post-stratification



This diagram shows how raster cells in the UFO are redistributed from the original sample design, to the the reclassified spatial product for the sample design

The initial AIM sample design was based upon a stratification which included r communities. This design was based on a manual reclassification of the GAP/LANDFIRE National Terrestrial Ecosystems 2011 data set.

# Citations

[[1]]  
Hijmans R (2022). \_terra: Spatial Data Analysis\_. R package version  
1.6-17, <https://CRAN.R-project.org/package=terra>.  
  
[[2]]  
Zvoleff A (2020). \_glcm: Calculate Textures from Grey-Level  
Co-Occurrence Matrices (GLCMs)\_. R package version 1.6.5,  
<https://CRAN.R-project.org/package=glcm>.  
  
[[3]]  
Leutner B, Horning N, Schwalb-Willmann J (2022). \_RStoolbox: Tools for  
Remote Sensing Data Analysis\_. R package version 0.3.0,  
<https://CRAN.R-project.org/package=RStoolbox>.  
  
[[4]]  
Liaw A, Wiener M (2002). "Classification and Regression by  
randomForest." \_R News\_, \*2\*(3), 18-22.  
<https://CRAN.R-project.org/doc/Rnews/>.  
  
[[5]]  
Kuhn M (2022). \_caret: Classification and Regression Training\_. R  
package version 6.0-93, <https://CRAN.R-project.org/package=caret>.  
  
[[6]]  
R Core Team (2022). \_R: A Language and Environment for Statistical  
Computing\_. R Foundation for Statistical Computing, Vienna, Austria.  
<https://www.R-project.org/>.  
  
[[7]]  
Corporation M, Weston S (2022). \_doParallel: Foreach Parallel Adaptor  
for the 'parallel' Package\_. R package version 1.0.17,  
<https://CRAN.R-project.org/package=doParallel>.