Vegetation Classification

The original AIM sample design for the field office utilized stratified random sampling within classified vegetation types which the plots would make statistical inference to. The vegetation types were composed of alliances, communities, and associations, from the GAP-Landfire National Terrestrial Ecosystems spatial data set (Program (2016)). These units were aggregated, by an expert at the field office, to form broader vegetation groups which could have a greater number of replicates per stratum for statistical purposes.

However, the GAP data set erroneously classifies many vegetation types at a non-negligible rate. Accordingly, many areas stratified as a certain vegetation type through the project did not, or seldom, featured the intended target vegetation. Several of the stratified sampling areas are in error, and while we cannot change the area associated with each plot, in certain instances the actual stratum which the plot was sampled in, and the actual cover of that stratum across the landscape will be considered as necessary.

For example, of the nine vegetation types which our AIM project was stratified across, a couple were seldom seen, such as 'Mixed Conifer' vegetation. The study area does contain this vegetation type, however it represents only a fraction of a percent of the field office, and the designated stratified area seldom coincided with it's actual presence. As the mapped areas did not correlate with the vegetation type, neither could the random plots drawn within it. Accordingly the acreage of these areas should, in several instances, be considered within actual vegetation types, which the plots were sampled in, in order that these data are interpreted in the appropriate context.

Additional problems were inherited with the vegetation types known as 'Other'. This is an aggregate developed from the *need* to classify the entirety of the field office. It functions as a catch-all designation for vegetation types which do not have adequate cover to form a broader stratification area and which have little semblance to other eight units. For example, a small patch of Blackbrush (*Coleogyne ramossisima*) on highway 90, gypsum terraces in the Paradox Valley, and escarpment vegetation with Stansbury's Cliffrose (*Purshia stansburyiana*) across the entire study area were placed into this designation. In hindsight as 'Other' is not a group wherein the members have any inherit similarities to themselves, we argue they should align with other groups which they share, even if only weak, affinities. Affinities between the gypsum terraces, to the salt desert, both soils which reduce the availability of free water for plant usage and result in barren to salt-tolerant vegetation are evident. Similarities between the escarpments of mesa's and the Pinyon-Juniper which occupies both the thin soil at the edges of the mesa, and the Pinyon-Juniper woodlands on the rocky soils at the toes of the escarpments, as well as those which are scattered throughout the Stansbury Cliffrose areas make this a tangible target for placement of these 'other' plots.

A final problem is associated with the need to classify bodies of water. Our field office the 'Uncompahgre', is named after a word of Ute origin which has various translations, but a central element of them is a reference to 'Red Rocks' and 'flowing water'. Our design stratum had 4% of the survey area designated as riparian, in part to hold surface rivers. However, given the allowance to shift plots 50m in the cardinal directions, the tri-spoke design of AIM plots requiring a 60m diameter, and the deployment of Lotic AIM during the sampling period, few to none plots remained in the riparian vegetation type. Given that this is a significant amount of cover it is addressed here.

In order to resolve these issues with the analysis of the AIM 2018-2022 sample design, we reclassify the field office into four major, and one very minor, vegetation groups which may accommodate a major swath of the lands in the field office. To accomplish this we use over 1600 random points across the entire extent of the field office, classified in Google Earth, with National Aerial Imagery Program (NAIP) aerial imagery and a couple simple spatial data products, as inputs to a simple random forest model which is projected onto the aerial extent of the field office.

Methods

NAIP Imagery from 2019 were downloaded from the official repository at 'Box' in Fall 2022. While decoding from MrSID (multiresolution seamless image database) 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, and all computing performed on Linux Ubuntu (20 & 22) on multiple hardware. The NAIP 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 10 meter Digital Elevation Model (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 (Haralick et al. (1973), Zvoleff (2020)). 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 imagery. Texture statistics: mean, variance, homogeneity, contrast, dissimilarity, with windows of 5 in both direction, shifts in all directions (i.e. *Queen's case*), and the default value of 32 grey levels. NAIP data were processed to derive another preditive layer a Normalized Difference Vegetation Index (NDVI) band.

$$NDVI = (NearIR - Red)/(NearIR + Red)$$

NDVI is well suited for identifying sparsely vegetated areas, it was useful in distinguishing salt desert from all other strata, and help in distinguishing between Mixed-Mountain Shrub and Pinyon-Juniper.

To create data set for training a random forest model, all 469 sampled AIM and LMF points, from 2018-2022, as well as all drawn 2022 AIM points, were exported to Google Earth and 440 were classified. 400 random points were generated across the field office 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 were the second closely located plot, but under distinct record elements in the TerraDat database. In all instances points were buffered by a 30m radius, to create the dimensions of an AIM plot, and the single most dominant vegetation type was recorded.

To develop a random forest model, the data set of 1657 classified points were partitioned using a split of 0.7:0.3 for the training and testing sets (n = 1146, n = 488) using 'caret' (Kuhn (2022)). The data set was not balanced (number of plots per stratum: AS-7, MC-6, MMS-124, PJ-631, SD-235, SS-143) due to the natural varying presence of these vegetation types in the study area.

The number of mtry in the random forest model were tuned using the function 'tuneRF' with the number of try's 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 with parellel processing (Liaw & Wiener (2002), R Core Team (2022), Corporation & Weston (2022)).

Four simple diagnostic criteria were used to evaulate the predictions from out simple model.

$$Sensitivity = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Which can be thought of as the probability of the method giving a positive result when the test subject is positive.

$$Specificity = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

Which can be thought of as the probability of the method giving a negative result when the test subject is negative

$$Accuracy = \frac{\text{correct classifications}}{\text{all classifications}}$$

Which can be thought of as the probability of the method giving the correct result under all conditions. *Balanced accuracy* accounts for differences in the number of individuals in groups, because if the number of individuals in groups differ widely, accuracy may gave a false indication of the models performance. This happens when one group, with many records, is classified very well at the expense of other groups.

$$kappa = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}}$$

For our model the overall accuracy was 0.773 (range 0.73 - 0.81 95 % confidence interval), due to an imbalance in the number of observations per group a more appropriate measure for evaluating the overall performance of the model, is the kappa, 0.628, which indicates substantial agreement.

Notably, the model has high rates of Specificity (median = 0.95, less Aspen and Mixed-Conifer), showing that when it predicts a vegetation type onto a pixel cell, the prediction is almost always correct; less so for prediction of Pinyon-Juniper (0.766). However, the model suffers from low Sensitivity (median = 0.71, less AS and MC), 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 Pinyon-Juniper, at the expense of Mixed Mountain Shrub and Sagebrush-steppe. 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 Pinyon-Juniper plots did not significantly increase the quality of predictions.

Further indicating, that the features used are not adequate for distinguishing between the transitional points where both Sagebrush-steppe phases into Pinyon-Juniper, and where shrubs increase in Pinyon-Juniper and it phases into Mixed-Mountain Shrub. In part we suspect this is an issue of the difficulty in splitting hairs required to generate the training data sets, and regrowth of historic vegetation in fire scars. While we believe this may be readily accomplished, given the objectives and goals of this process, we believe these are out of the current scope.
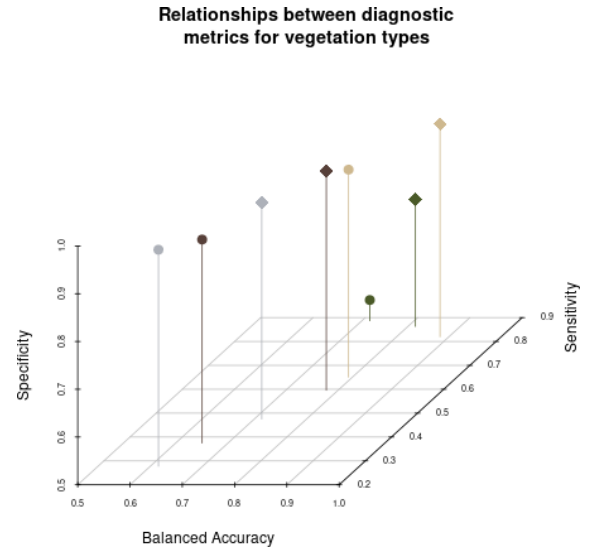

Figure 1: Three dimensional scatter-plot showing the relationship between three common metrics for evaluating the predictions of models. Circles represent the performance of the original stratification, and diamonds the new classification. Positions towards the upper right corner indicate more desirable qualities.

In order to determine the relative performance of our model to the original GAP classification a confusion matrix was also generated. A similar number of test points, 486, were used for both data sets. These points were only selected from the computer generated points in order to be independent of the data product, which the AIM plots were derived from. Several metrics indicate the original model is less accurate than the second model. The accuracy of the original model was 0.578 (range 0.53 - 0.63 95% confidence interval), a difference of roughly 0.19 to the new model. This test data set was also unbalanced and it's kappa, 0.368, serves as a better predictor or overall model performance in this case, a difference of roughly 0.26. Considering only the four major vegetation classes also present in the re-stratified model the original has similar rates of specificity (median = 0.93, less AS and MC), but similar to the re stratified model but has lower rates of sensitivity (median = 0.51, less AS and MC) (Figure 1).

On the whole the new model outperforms the older model in all comparisons except for that the sensitivity of the original PJ classification is higher than that of the secondary PJ classification (0.887 to 0.863) (Figure

1). In general their are multiple trade offs in the comparison of models, however the new model is both likely to correctly identify a the strata of a location, and to identify it correctly.

Comparison of Old and New Vegetation Models

The initial sample design only classified around 41% of the field office as Pinyon-Juniper. Based on our field experiences we believe this is a serious underestimate. Despite the need to incorporate several extra - or 'Other' vegetation types into Pinyon-Juniper to fit the four model vegetation classification, manual classification resulted in 60% of the 1600 points being categorized as Pinyon-Juniper. This result is much more in line with our observations, and expectations. That smaller amounts of land were categorized as Pinyon-Juniper by the initial strata actually had large affects on plot placements. Because these woodlands were intended to be sampled at a lower rate than other strata, this roughly 20% of area which is PJ, but missed in the original stratification was sampled more than intended. Thus resulting in lower amounts of information on other vegetation types, such as sagebrush, which we wanted to know more about. Our model was slightly greedy with it's classification of Pinon-Juniper, raising it to 62% of the total area. An important caveat with our use of the models here is that, these areas include many late successional areas, which have not had disturbances to maintain their phases in mixed grasslands.

Estimates of the amount of Sagebrush habitat differed widely between stratification efforts. In our experience many areas which were classified as Sagebrush in the original classification were actually Pinyon-Juniper, and the age of the Pinyon-Juniper was great enough that this was unlikely to be a result of the different ages of aerial imagery utilized by both models. The original stratification had roughly 25% of the entire field office classified as Sagebrush, this value was too high. Our manual classification resulted in 12% of the field office being classified as Sagebrush, and our final classification from the model resulted in only 9% of land being classified as Sagebrush. We feel the true value is closer to 12% than the 9% predicted by the model. As mentioned, not as many plots were included in this strata as desired, due to the classification accuracy, rather many of the plots in this strata ended up being Pinyon-Juniper. Classifying sagebrush was difficult for us due to areas where it transitions into Mixed-Mountain Shrub and Pinyon-Juniper, i.e. what constitutes the transition point between these? Theoretically only small amounts of these two types are required to convert a land, and given Sage grouse preferences these make sense for our purposes.

The amount of Salt Desert is similar between all three methods, they agree the amount of salt desert in the field office is around 15% - 18%. In general, these predictions aligned well, and plots which were classified as salt-desert by the models actually were.
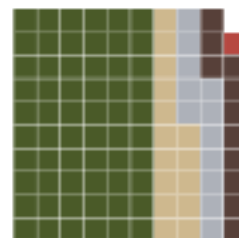
A notable incongruence is present with Mixed-Mountain Shrub between the systems. The original classification predicted very modest amounts of this stratum throughout the area, and predicted relatively high amounts of other forest types. Our manual classification, which retained Aspen and Mixed Conifer forest, detected nearly twice as much of
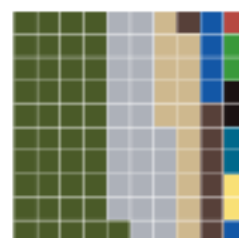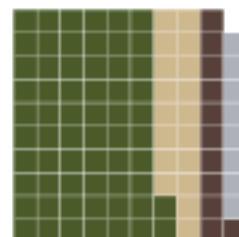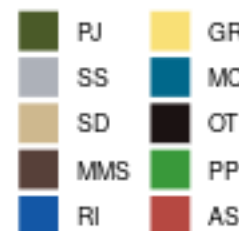

Figure 2: Changes in Predicted Land Cover

this strata in the field office. The modelled cover was slightly less than this, and likely closer to a true value where 10% of the field office is Mixed-Mountain Shrub.

The relationships for the six remaining minor components is more complex, as is illustrated in Figure 3. This figure tracks the movement of each individual raster cell (or 'pixel') from the old to the new classification. As previously mentioned, the accuracy of the old system cannot be officially compared, but only discussed, because we reduced our classification process to the four major and two minor strata. For example, the original classification (lower half of circle), was correct in it's identification of Pinyon-Juniper, however it was unable to detect all of it. Hence the green line tracking from the bottom to the top, contains nearly every single pixel, except for a few which we were able to identify as Mixed-Mountain Shrub. Nearly all 4% of the Riparian stratum can be accommodated in Pinyon Juniper or the Salt desert stratum. This is sensible as nearly all ephemeral rocky canyon bottoms support Pinyon Juniper vegetation, and as the streams flow through the lower elevation portions of the field office, the riparian band is only a meter or two wide, and hence what an AIM crew would sample is Salt Desert. Most of the original Mixed Conifer was reassigned as Mixed-Mountain Shrub, with only a portion remaining in this stratum. Similarly, it appears roughly half of the Ponderosa Pine was assigned to Pinyon-Juniper and the other to Mixed-Mountain

Figure 3: Depiction of how raster cells in the UFO are redistributed from the original sample design (lower half), to the the reclassified spatial product for the sample design (upper half)

Shrub. Roughly half of the area classified as Aspen was maintained, and the other half was also reclassified as Mixed-Mountain Shrub. Most of the grassland stratum was transferred to Salt Desert
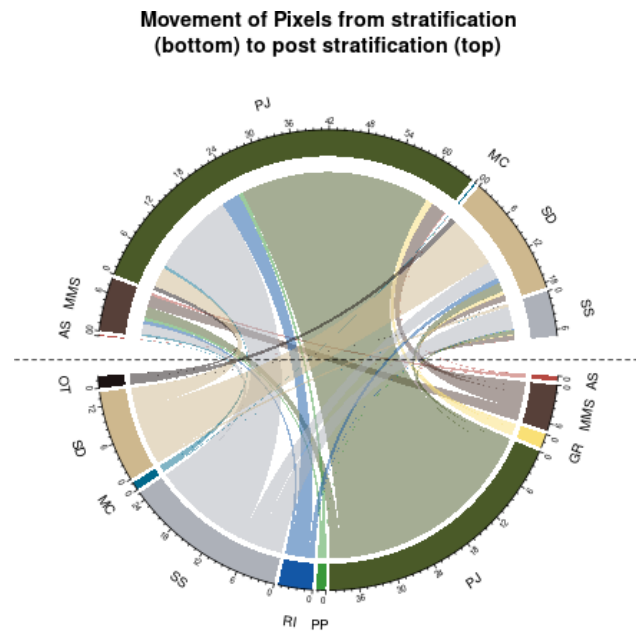
References

Corporation, M., & Weston, S. (2022). *doParallel: Foreach parallel adaptor for the 'parallel' package.* https://CRAN.R-project.org/package=doParallel

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *6*, 610–621.

Kuhn, M. (2022). *Caret: Classification and regression training.* https://CRAN.R-project.org/package=caret

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22. https://CRAN.R-project.org/doc/Rnews/

Program, U. G. S. G. A. (2016). *GAP/LANDFIRE national terrestrial ecosystems 2011: US geological survey.*

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Zvoleff, A. (2020). *Glcm: Calculate textures from grey-level co-occurrence matrices (GLCMs).* https://CRAN.R-project.org/package=glcm