Increase Variation Around Narrow ESD Concepts

The quantitative benchmarks of Ecological Site Descriptions (ESD's) are meant to capture the variation inherent in a state and phase under multiple conditions, from consecutive drought to surpluses of moisture, and following multiple disturbances (CITE). They are intended to capture the variation that would be found in this state and phase combination across the geographic and climatic extents of the ESD in the MLRA for which it has been drafted or adapted. Some of the quantitative benchmarks, of the fractional cover of functional vegetation groups, for Ecological Sites which we collected from ESD's were very narrow (Figure 1, top panel). In many of these instances the reported values were more narrow than the uncertainty of the estimates of the true value of the population gleaned from a single AIM plot. It is apparent that several ESD developers did not emphasize the natural variability of the vegetation benchmarks while generating the cover estimates. This may be due to them only collecting quantitative vegetation data at a single point in time within the Ecological Site, accordingly it seems in multiple instances they may only have had a point of datum, and did not feel comfortable estimate the variation in the system.

While the approach descried above is prudent in the development of an ESD, it is not prudent for us to assume such narrow ranges of variation. These narrow ranges may unduly penalize estimates of the amount of land under analysis which are meeting condition benchmarks.

For example, if we seek to determine if plots within an ESD are meeting a standard - such as having from 20 - 30% grass cover, than given the time frame under which we sampled, plots with 20% grass cover have meet this condition. However, if this range of grass cover was reported as 24-26% than, a plot within the noted ranges of variation may not meet the standards. Here we seek to identify and broaden these estimates, we will use a simple method of *imputing* values in the context of *feature engineering*. A *linear model* will be fit to the benchmark values, which contain realistic ranges, and then the slope of this model will be used to fill in the missing values.

Ranges of estimated benchmark variation were estimated as being too low if they fell within the ranges in Table 1 & Figure 1 *top panel*. These 81 values were removed from the initial data set. The remaining 59 observations were used as **training** data for the linear model: lm(Range ∼ Mean + Functional Group) (r formula notation). We believed that the variation associated which each measurement of range would decrease as the mean cover increases. In other words, as the mean of the cover estimate get's larger, the percent variation of the range of them decrease.

Mean	Range	Variation*
1 - 10	< 3	30.0 %
11 - 20	< 4	26.7 %
21 - 30	< 5	20.0 %
31 - 50	< 6	15.0 %
51 - 100	< 7.5	10.0 %

* calculated as the *range* divided by the midpoint of the *mean* multiplied by 100 to gather a percentile, $\frac{max(Range)}{\text{midpoint(Mean)}} * 100$

This simple model served to explain a moderate amount ($r^2 = 0.28$) of the variation in the range of cover estimates. Based on these data it found moderate evidence that functional type has a effect on the slope of the trend line (p = 0.0056), and there was slight evidence (p = 0.0106) that increasing the mean of the values had a negative effect on the observed variation. However, this effect was much weaker than the additive co-variate of functional type; and the final results nearly exclude the term. For every one percent increase in the mean of the benchmark values the percent range in variation of the estimates increased by 0.001 percent.

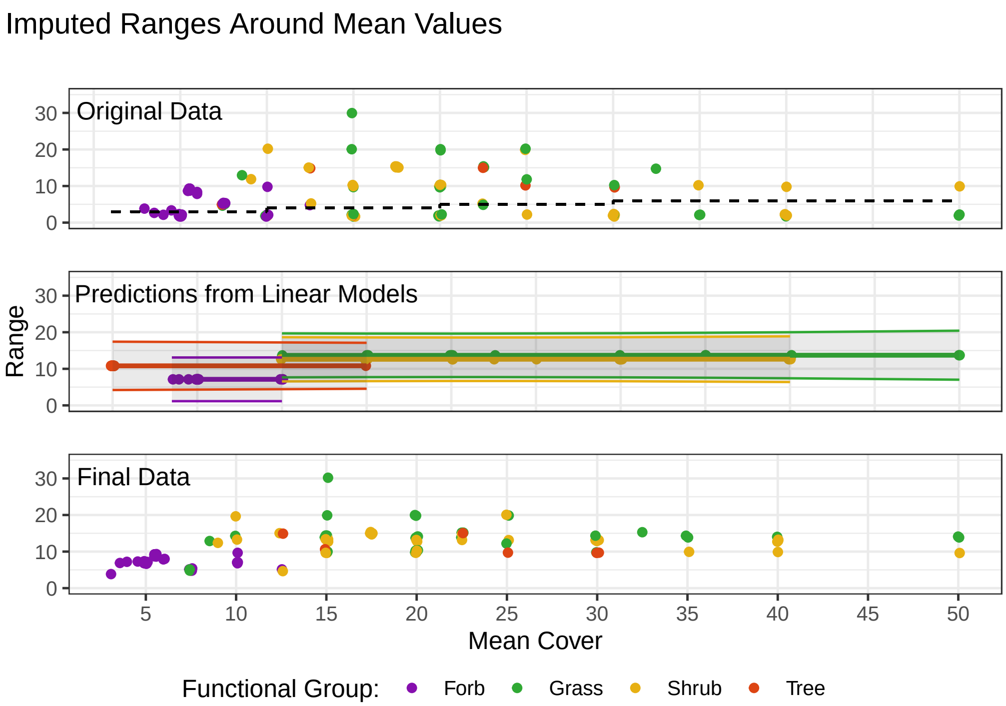

Figure 1: The process of imputing values to the benchmarks. The original data (top), contained many values without sufficient variability to capture the variation in an ES (beneath black dotted line). The predictions (middle) from linear models using the mean and functional types as predictors, the shaded areas representing an 80 percent prediction interval. The results (bottom) show the new estimates of the points from under the line in top, with the predictions from middle, with the data which was not altered.

Once the linear model was *fit*, the removed data points had both the Mean, or midpoint of their cover estimates, and functional types used together to predict what the Range of cover estimates would be for an ESD (Figure 1 middle panel). These estimates were then combined with the original data which meet the desired range of cover information to provide an altered quantitative benchmark dataset (figure 1 bottom panel & figure 2.

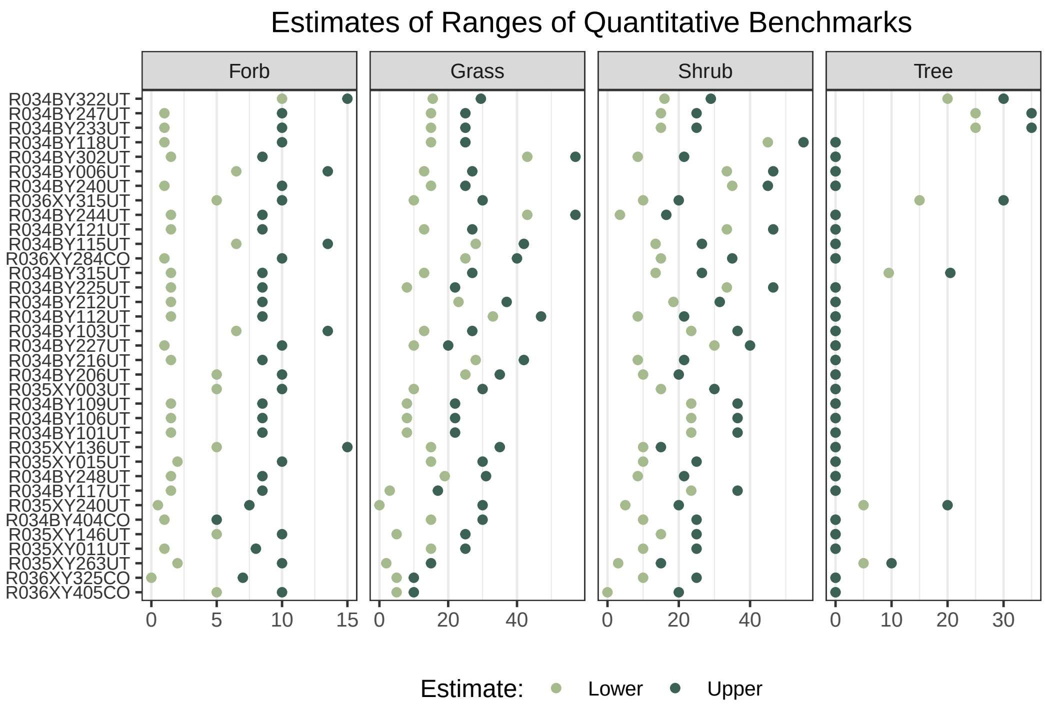

Figure 2: Initial and Calculated ESD Ranges