Increase Variation Around Narrow ESD Concepts

The quantitative benchmarks of ESDs 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 relevant MLRA. 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. 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 site, and not across multiple time points, 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.

Well that approach is prudent, it is not prudent for us to assume such narrow ranges of variation. These may unduly penalize estimates of the amount of land under analysis which are meeting condition benchmarks. 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 79 values were removed from the initial data set. The remaining 46 observations were used as **training** data for the linear model: lm(Range $\sim$ Mean + Functional Group) (r formula notation).

Mean	Range	Percent*
1 - 10	< 3	30.0 %
10 - 20	< 4	26.7 %
20 - 30	< 5	20.0 %
30 - 50	< 6	15.0 %
50 - 100	< 7.5	10.0 %

- This calculated as the *range* divided by the midpoint of the *mean* multipled by 100 to gather a percentile, $\frac{max(Range)}{\text{lower Mean + upper Mean)}} * 100$

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.

This simple model served to explain a moderate amount ($r^2 = 0.27$) of the variation in the *response variables...* It found that functional type has a very strong effect on the slope of the trend line (p = 0.0218), and there was strong evidence (p = 0.0496) that increasing the mean of the values had a negative effect on the observed variation. For every one percent increase in the mean of the benchmark values the percent range in variation of the estimates decreased by 0.013 percent.

Once the linear model was *fit*, the removed data points had estimates of their values recorded.

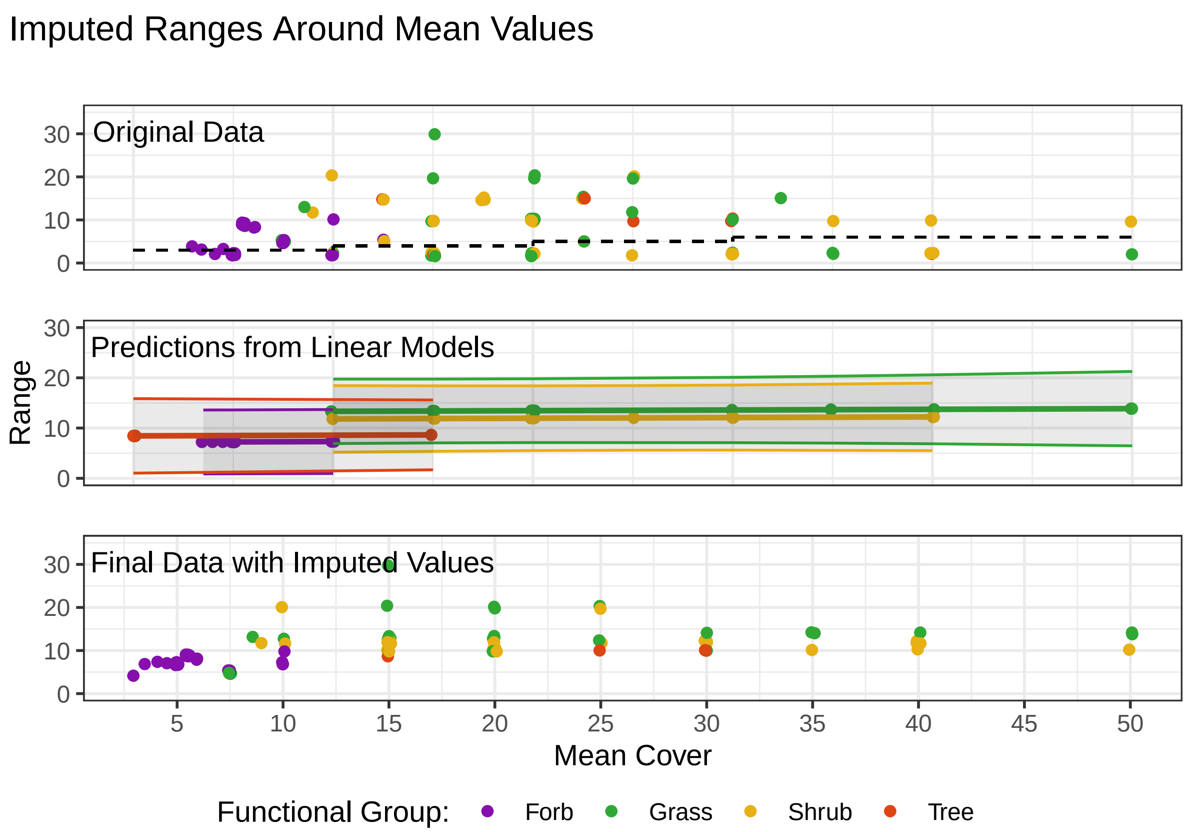

Figure 1: Imputing the values for