Vegetation Data: How much is enough?

B.K. Schulz*, USDA Forest Service PNW Research Station and
R.L. DeVelice, USDA Forest Service Chugach National Forest

Introduction:

Recognition of the value of forest vegetation data has increased in recent years, especially when it is collected using consistent methods over many forest types. However, the cost of collecting data is often perceived to be prohibitive. Managers must balance the cost of collecting data with the utility of the conclusions that may be drawn from analyses of the data collected.

One of the basic uses of vegetation data is to classify plant community types. Classification is the process of grouping similar entities together based on shared characteristics. Species composition data and environmental features are used to describe units that are useful for management applications. Hierarchical clustering is used as the first step in classification.

This study compares the results of hierarchical cluster analysis of full data sets (all vascular plant species) versus "trimmed" data sets for three distinct geographic areas of the United States. Trimmed data sets are often considered to be less expensive to collect.

Data sets:
Three "community types"—defined by forest type and ecological province—were selected based on an adequate number of Forest Inventory and Analysis (FIA) Phase 3 (P3) plots with vegetation indicator data in distinct geographic regions.

Methods:

A ***full*** species matrix was built for each community type from FIA P3 vegetation indicator data, with species abundance averaged to the plot level from raw subplot data.

A ***trimmed*** species matrix was created for each community type by pulling only subplot species recorded with a canopy cover of at least 3%, averaging each surviving species to the plot level.

Each matrix was then subjected to several steps for reducing noise, as per typical clustering analyses:

•Cleaned for clustering: Species that were only recorded on one plot within the community and all unknown species were deleted.

•Species abundance (canopy cover) were relativized to the maximum value.

Hierarchical cluster analysis was performed using PC-ORD® software using flexible beta linkage method (beta value set to -0.25) and Sorenson (Bray-Curtis) distance measure.

Results: Differences in datasets

Full species data sets

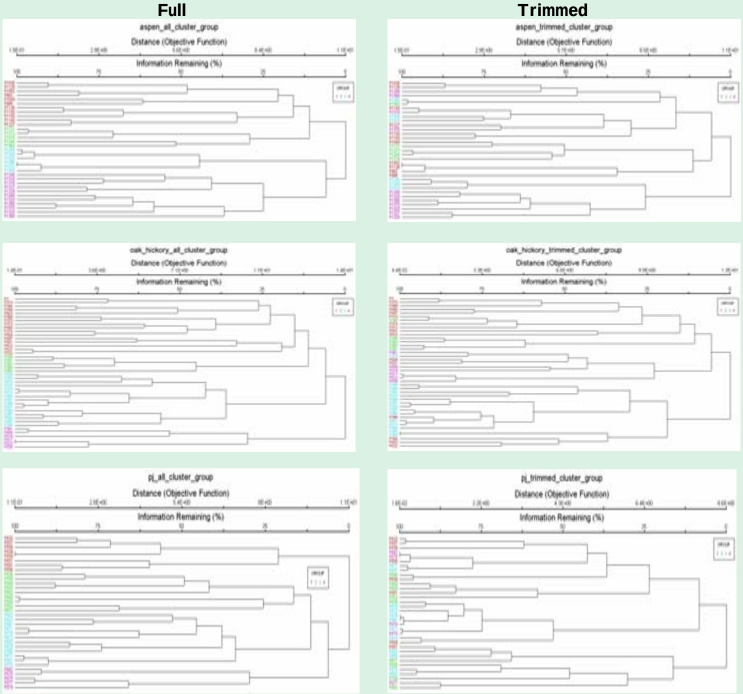
Community		Raw data		Cleaned for clustering	
ECOProvince— Forest Type	n	Gamma	Mean plot alpha	Gamma	Mean plot alpha
212—Aspen	33	438	46.6	227	39.6
222—Oak-Hickory	42	470	50.0	272	44.2
341—Pinyon-Juniper	34	248	20.1	97	14.9

Trimmed species data sets

Community		Raw data		Cleaned for clustering	
ECOProvince— Forest Type	n	Gamma	Mean plot alpha	Gamma	Mean plot alpha
212—Aspen	33	249	21.7	121	17.6
222—Oak-Hickory	42	212	19.7	102	17.0
341—Pinyon-Juniper	34	78	5.2	23	4.0

Gamma = the total number of unique species in data set
Mean plot alpha = the average number of species per plot

Cluster dendrograms:



Plot numbers are colored to represent 4 basic groups defined in the full data sets at the ~12.5% information remaining level, as displayed at the far right side of the dendrogram. Original colors are maintained in the trimmed data dendrograms to display shifts in order. The number of moves from the full basic group order in the trimmed data dendrograms are tallied in the next table.

Shifts in basic cluster order between full and trimmed data sets

Number of moves	Aspen	Oak-Hickory	Pinyon-Juniper
		number of plots	
0	17(51.5%)	15 (35.7%)	11 (32.3%)
1	5 (15%)	22 (52.8%)*	14 (41.2%)
2	7 (21.2%)	2 (5%)	8 (23.5%)
3	3 (9%)	4 (9.5%)	1 (3%)

* Two large groups shifted in order

Discussion points:

•Sparsely vegetated forest communities lose the highest percentage of species data when criteria for species data collection require species be present at a minimum percentage canopy cover.

•Cursory examination of the species "lost" in trimmed data set indicate many have high constancy but low abundance. Indicator species analysis may reveal these species are important to distinguish site qualities defining distinct plant associations.

•Classification of forest plant communities beyond forest type is one of the stated objectives for collecting "most abundant" species data on forest inventory plots. This preliminary study suggests resulting classifications may not remain stable when important indicator species are not recorded due to low abundance.

•Complete inventory of all vascular plants on standard sized plots allows for assessments of species diversity (richness) and species distributions (frequency). These assessments are not possible with inventories that only include "most abundant" species.

Further research:

•Trim the data sets using different criteria, such as including several "most abundant species" from each growth habit without a minimum percent canopy cover limit, to see if clustering will remain stable.

•Examine hierarchy structure to see where stability patterns change.

•Use ordination (e.g. NMMDS) to evaluate the magnitude of shifts in multidimensional space.

•Perform indicator species analysis on selected species that were lost in the trimming exercise to test their importance for defining plant associations.

***Contact Information:**
Beth Schulz, FIA
Anchorage Forestry Science Lab
Telephone: 907-743-9424
E-mail: bschulz@fs.fed.us