CREATING HOMOGENOUS LANDFIRE VEGETATION CLASSES FOR FOREST INVENTORY APPLICATIONS IN THE INTERIOR WEST

Madelon R.K. Basil[1], Salma Huque[2], Kelly S. McConville[3], Gretchen G. Moisen[4], and Tracey S. Frescino[5]

Abstract

This project involved collapsing the existing vegetation type (EVT) classes created by the LANDFIRE Program into a smaller set of structure metrics based on four key forest attributes: (1) aboveground biomass, (2) net volume of trees, (3) count of trees per acre, and (4) basal area. Our goal was to create categories that were internally homogenous with respect to these four variables and heterogeneous between categories so that the resulting categories could be effectively used in estimation, mapping, and analysis applications for forest inventory. We used k-means clustering as the main classification method and made refinements to the algorithm in order to make the final categories as intuitive and useful as possible. The result was two classification schemes (1) one with five categories, and (2) one with 10 categories. We used data from one cycle of the Interior West region of the Forest Inventory and Analysis Program (FIA) but intend our method to be applicable for all FIA administered regions.

Keywords: Classification algorithm, FIA, k-means clustering, LANDFIRE, U.S. Geological Survey (USGS).

INTRODUCTION

The U.S. Department of Agriculture (USDA), Forest Service, Forest Inventory and Analysis Program (FIA) estimates numerous forest attributes at a variety of domains across the United States. Supplementing the data from ground plots with remote imagery, FIA uses auxiliary remotely sensed data to increase precision in estimates (Bechtold and Patterson 2005), build predictive models for mapping forest attributes (Moisen and Frescino 2002), and serve as a backdrop for analyses (Shaw and others 2017). Often, however, remotely-sensed data comes in the form of classified maps with an unwieldy number of classes such that collapsing is necessary to make a workable dataset for estimation, mapping, and analytical applications. This project illustrates an objective collapsing process applied to classified remotely sensed data to determine how these data could be transformed to support FIA's varied uses of remotely sensed data.

LANDFIRE (Landscape Fire and Resource Management Planning Tools) is a joint program between the USDA, Forest Service and U.S. Department of the Interior that provides geographic information to assist with wildland fire management (Rollins 2009). The LANDFIRE data products include categories such as disturbance, fuel, and vegetation. The vegetation type products include existing vegetation layers and potential vegetation layers. LANDFIRE offers three existing vegetation layers: (1) Existing Vegetation Type (EVT), (2) Existing Vegetation Canopy Cover, and (3) Existing Vegetation Height (USGS 2013, 2016). The layers were created using satellite imagery and classification and regression trees.

An EVT Class depicts the vegetation and land cover type, based on a classification of NatureServe's terrestrial ecological systems for the western Hemisphere. These types are groups of plant communities that tend to occur together (NatureServe 2018). Each class is further represented by a more generic physiognomic

[1] Research Assistant, Swarthmore College, Swarthmore, PA, 19081 and Reed College, Portland, OR 97202.
[2] Research Assistant, Reed College, Portland, OR 97202.
[3] Assistant Professor of Statistics, Reed College, Portland, OR 97202.
[4] Research Forester, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT 84401.
[5] Forester, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT 84401.

Citation for proceedings: Brandeis, Thomas J., comp. 2020 (Revised). Celebrating progress, possibilities, and partnerships: Proceedings of the 2019 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; November 19–21, 2019; Knoxville, TN. e-Gen. Tech. Rep. SRS–256. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 276 p.

class describing the overall characteristic of the community, such as Conifer, Shrubland, and Grassland. The EVT classification also aligns with the upper physiognomic levels of the National Vegetation Classification (NVC) System, which provides standards for vegetation classifications (http://usnvc.org/). The NVC upper physiognomic levels include three variables representing general physiognomic features (*order*), such as Tree-dominated and Shrub-dominated; and more detailed physiognomic characteristics (*class*), such as Open tree canopy and Closed tree canopy; and subclasses (*subclass*), such as Evergreen closed tree canopy and Deciduous closed tree canopy. The mapped values of EVT used in these analyses are predicted values from an empirical model that naturally has a classification error.

Our primary task was the re-categorization of LANDFIRE's EVT variable. The EVT variable currently encompasses over 635 distinct classes such as "Western Cool Temperate Undeveloped Ruderal Grassland", "Western Great Plains Shortgrass Prairie", and "Madrean Pinyon-Juniper Woodland; 182 of these classes can be found in the Interior West. LANDFIRE releases the data at this level of detail to accommodate all users and their potential analysis. LANDFIRE provides users with examples and tutorials on how land managers can collapse classes to suite their analysis questions and to conduct a related assessment of class agreements. We illustrate an alternative solution by collapsing the current EVT classes into a smaller set of categories based on four forest structure attributes (live trees only) that form the foundation of many FIA analyses: aboveground biomass (pounds), net volume (cubic feet) of trees, count of trees per acre, and basal area (square feet). Understanding that EVT classes are based on species composition and the target variables are based on forest structure, our goal is to create new structure metric categories that are internally homogenous with respect to these four variables and heterogeneous between categories.

This report explains the methodology used to create the structure metrics for collapsing the EVT variable. In the Data section, we discuss the data used to build the new classification and in the Methods section, we explain the techniques used to create the classification scheme. In the Results section, we present our refined process, final schemes, and visualizations. In the Discussion section, we review the effectiveness of our final schemes, offer limitations of our work, and outline potential uses of the final schemes.

DATA

The data used include the most current 10-year measurement cycle for each Interior-West State, downloaded from the FIA database on February 6, 2019 from version FIADB_1.8.9.99 (last updated December 3, 2018). The data were collected between 2007 and 2017 and represent 86,085 ground plots. All States included a complete 10-year cycle of data except Wyoming, which had only 7 of 10 years of data collected at the time of download. These data were collected using a quasi-systematic sampling design (Bechtold and Patterson 2005) wherein the observational unit is a single ground plot. We merged a dataset containing the four FIA collected variables. These variables were extrapolated and summed to an acre-plot size for live trees only (aboveground biomass, net volume of trees, count of trees per acre, and basal area) with a dataset containing the EVT variable that was identified by the center point of each plot location. A clustering algorithm was applied to the data at (1) the plot level, wherein each row of the dataset represented one of the 86,085 ground plots, and (2) at the EVT level, wherein each row of the dataset represented one of the 182 EVT classes found in the Interior West. To create the EVT level dataset, we recorded the mean of each of the four forest attribute variables (mean biomass, mean net volume, mean tree count, and mean basal area) across all the observations within a specific EVT class.

METHODS

Defining Observations

The overarching objective of this analysis was to collapse the EVT classes into a meaningful number of larger classes for our analysis purposes by assigning EVT classes to clusters identified through the k-means process. The process was run in two distinctly different ways. First, the individual plots were treated as the observations and run through the clustering process, referred to as "plot-level analyses" in the RESULTS. Second, the EVT classes themselves were treated as the observations, referred to as "EVT class level analysis" in the RESULTS.

K-Means Clustering

For much of our research, we utilized the k-means clustering algorithm. K-means clustering is an unsupervised learning algorithm that helps to reveal distinct structures in a dataset by partitioning the observations into non-hierarchical groups called clusters (James et al. 2013). To create the clusters, the algorithm requires the user to specify the number of groups they want to obtain from the data; this value is denoted as *k*. In step one, the algorithm randomly assigns each observation to one of the *k* groups. Then, for each of these groups, a mean point known as a centroid is computed. In step two, each observation is reassigned a cluster based on its Euclidean distance to the closest centroid. The mean of each new cluster is found and the centroid is repositioned to once again represent the mean. Step two is iterated repeatedly until the centroids no longer shift. To implement this method, we used the R function *kmeans()* (R Core Team 2019), which implements the Hartigan and Wong (1979) algorithm. When used appropriately, k-means clustering results in observations neatly separated into clusters that are internally similar with respect to the response variables of interest, yet dissimilar from other clusters.

Elbow Method

In conjunction with our use of k-means, we consulted the elbow method, a process used to verify that a user of the k-means clustering algorithm has selected a suitable value for *k*. To determine the optimal number for *k*, the elbow method uses the within sum of squares measure, which is given by

$$WSS = \sum_{j=1}^{k}\sum_{i=1}^{nj}(\boldsymbol{x}_{j,i} - \mathbf{c}_j)^t(\boldsymbol{x}_{j,i} - \mathbf{c}_j)$$

where $\boldsymbol{x}_{j,i} = (x_{j,i,1}, x_{j,i,2}, x_{j,i,3}, x_{j,i,4})^t$ is the vector containing the four forest attribute values for the ith observation in cluster j, $\mathbf{c}_j = (c_{j,1}, c_{j,2}, c_{j,3}, c_{j,4})^t$ is the vector containing the forest attribute values for the centroid of cluster *j*, and *nj* is the number of observations in cluster *j* (James and others 2013). This measure is the sum of the squared distances between each observation and its respective centroid. When calculated for various k-means trials, all with different *k* values, it demonstrates how tightly clustered the resulting k-means groupings are, as *k* grows. In this sense, the elbow method allowed us to determine the smallest *k* value in which the resulting clusters are still very tightly grouped and name this our optimal *k* value. We produced an elbow curve that displays the relationship between k-means trials with a different *k* and the calculated corresponding WSS. This visualization indicated that beginning with four clusters ($k = 4$), the WSS only exhibits small decreases as *k* increases for both the plot level dataset and the EVT class level dataset and therefore $k = 4$ would be the best starting option for our analyses.

RESULTS

Zero Response Cluster

Seventy-two percent of the plots contained a zero for biomass, net volume, count of trees, and basal area. The vast majority of these plots, 96 percent, belonged to the Interior West non-forest stratum. When the k-means algorithm was applied to the dataset that included the zero observations, it resulted in the creation of a very large cluster that represented plots having small values for the four response variables. In our 5-cluster scheme, this "Low" cluster was composed of over 72,000 of the 86,085 observations and combined different EVT classes that were not to be conflated with one another. For example, sparsely vegetated and shrub-like classes were grouped together with observations labeled "Snow-Ice", "Barren", and "Water". Because we aimed to create clusters guided by ecological principles (as well as being homogenous in terms of our responses), the zero observations were automatically placed into their own cluster, denoted *Cluster 0*, and excluded from further analyses.

Increasing the Number of Categories

Another technique we used to further refine our k-means clusters involved increasing our k value and evaluating whether any new patterns were drawn out from our cloud of observations. Altering the number of clusters we wanted to obtain from our data helped us identify layering effects we were not able to see when there were only five clusters. The $k = 10$ appeared to bring out subtle patterns that may prove beneficial to some applications. Therefore, we decided our final product should consist of two schemes for condensing EVT: one coarser scheme where $k = 5$ and one finer scheme where $k = 10$.

Mapping EVT Classes to Clusters

***Plot level analysis*—**Once each ground plot had a cluster assignment, the final task was to assign each EVT class to a particular cluster. This proved difficult because in many cases, a given EVT class had membership in more than one of our clusters. For example, we would anticipate that most "Northern Rocky Mountain Mesic Montane Mixed Conifer Forest" observations would be sorted into the one of our four k-means clusters that ensured the most within-cluster homogeneity and between-cluster heterogeneity in terms of our four responses of interest. However, what we found when we plotted the division of the four clusters within "Northern Rocky Mountain Mesic Montane Mixed Conifer Forest", as shown in figure 1, was that this EVT class was included in all four of our distinct k-means clusters. This phenomenon was not exclusive to only one or two

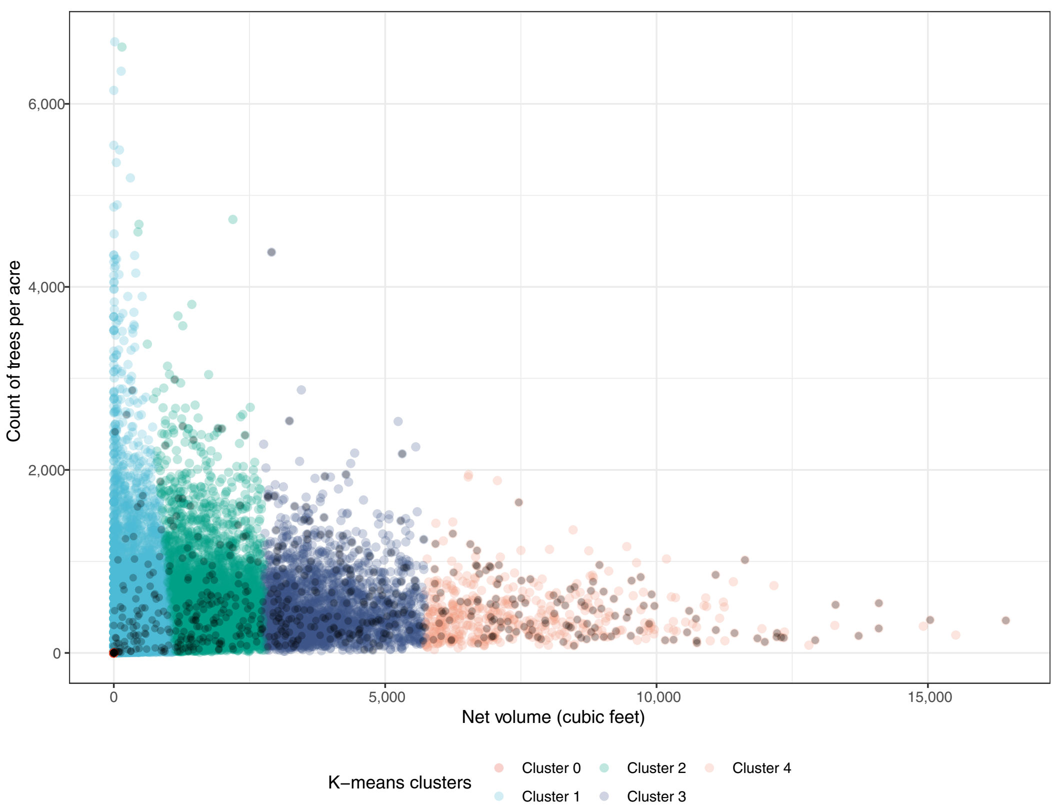

Figure 1—A scatterplot of volume and tree count, colored by the 5 k-means clusters. Although only two response variables are graphed, the clusters do account for all four response variables. Each dot represents a single ground plot and the black dots are those in the existing vegetation type (EVT) class "Northern Rocky Mountain Mesic Montane Mixed Conifer Forest."

EVT classes, but also appeared in many instances across the data. It was seen most commonly with observations marked "Conifer", "Hardwood", or "Riparian" in terms of plant physiognomy.

One possible explanation for the presence of these physiognomic classes in multiple clusters could be that certain types of vegetation are sorted into different clusters at different points in their growth. In the case of conifers, the trees included in this category grow very large across their lifespan. While conifer saplings may fall into a cluster characterized by low biomass and low net volume, fully grown trees will likely be clustered in a different group. On the other hand, this is less of an issue with EVTs generally composed of smaller trees and shrubs because they typically remain the same size (with limited aboveground biomass, net volume, etc.) over their lifespan relative to the conifers. Another possible explanation could be due to the natural variations in the vegetation communities within the EVT classes themselves. For example, plant communities at the fringes of membership in EVT classes can look very different in terms of canopy closure.

We attempted to resolve this issue by implementing our unique version of a classification algorithm. The process unfolded as follows: we first created a dataset with all observations grouped by *order*, with one row per cluster per order. We then determined which of the k-means clusters contained the majority of the data within each order and assigned all observations of that order to that majority cluster. In cases where there was no cluster that encompassed a clear majority, we divided the order into the finer category of *class* and repeated; for classes with no clear majority cluster, we went down to the *subclass* level and repeated. If a particular subclass did not have a clear majority cluster, we divided the subclass into EVT classes and repeated the process a final time. Note the "nested" nature of these categorical variables; all of the EVT classes in our data had membership in exactly one subclass, all subclasses had membership in exactly one class, and all classes has membership in exactly one order.

Ultimately, many of the EVT classes had no clear majority cluster, so we determined EVT class-level analyses to be most appropriate for this application. Exhausting this classification algorithm allowed us to return to the k-means algorithm with new knowledge about where our clusters intersected well with EVT classes and where the majority of observations within a particular EVT class refused to be sorted into a singular k-means cluster.

EVT class level analysis**—**We solved the multiple class problem by clustering at the EVT class level, instead of at the plot level. With only one observation per EVT class, it could only be placed into one cluster. As outlined in the Data section, we created the EVT level dataset by computing the mean response values (mean biomass, mean net volume, mean tree count, and mean basal area) across all observations within a specific EVT class. Then, instead of running k-means on all 86,085 ground plots, we were able to cluster each of the 182 EVT classes based on their mean response values. This eliminated the possibility of one EVT class inappropriately falling into multiple clusters. As with the schemes above, we included a *Cluster 0* composed of EVT classes in which the mean values for each forest structure attribute were all zero.

Final Results

Our final result was two classifications at the EVT level—one with five clusters and one with ten clusters. We used k-means to create the clusters and used the elbow method to determine that four clusters (plus a zero cluster) were the optimal number. As discussed in the Methods section, the decision to also increase the k value was made with the hope that doing so could reveal distinctions between different vegetation classes (particularly those with low volume and high-count values). Both of the resulting classifications were created using the mean values of the four response variables across all EVT observations. The result for $k = 5$ is shown in figure 2 and $k = 10$ in figure 3. A summary of each scheme is given in table 1 and table 2 and the entire look-up table can be found in the appendix table.

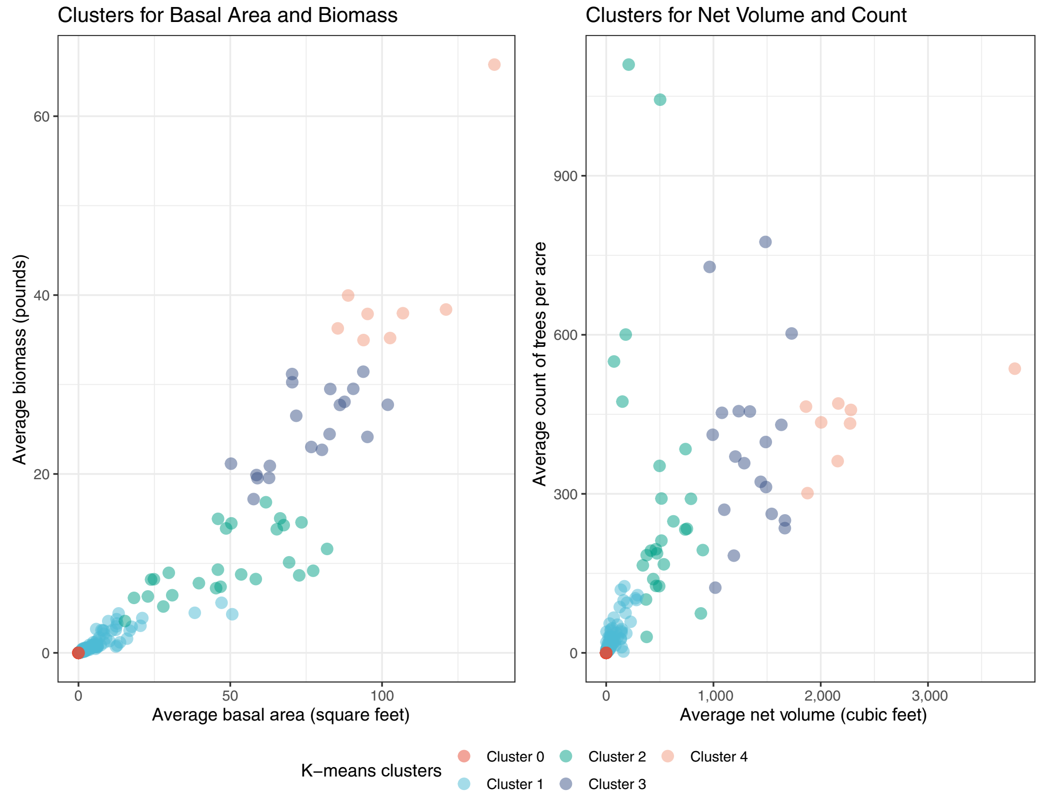

Figure 2—Scatterplots of the response variables, colored by the 5 k-means clusters at the existing vegetation type (EVT) classification level.

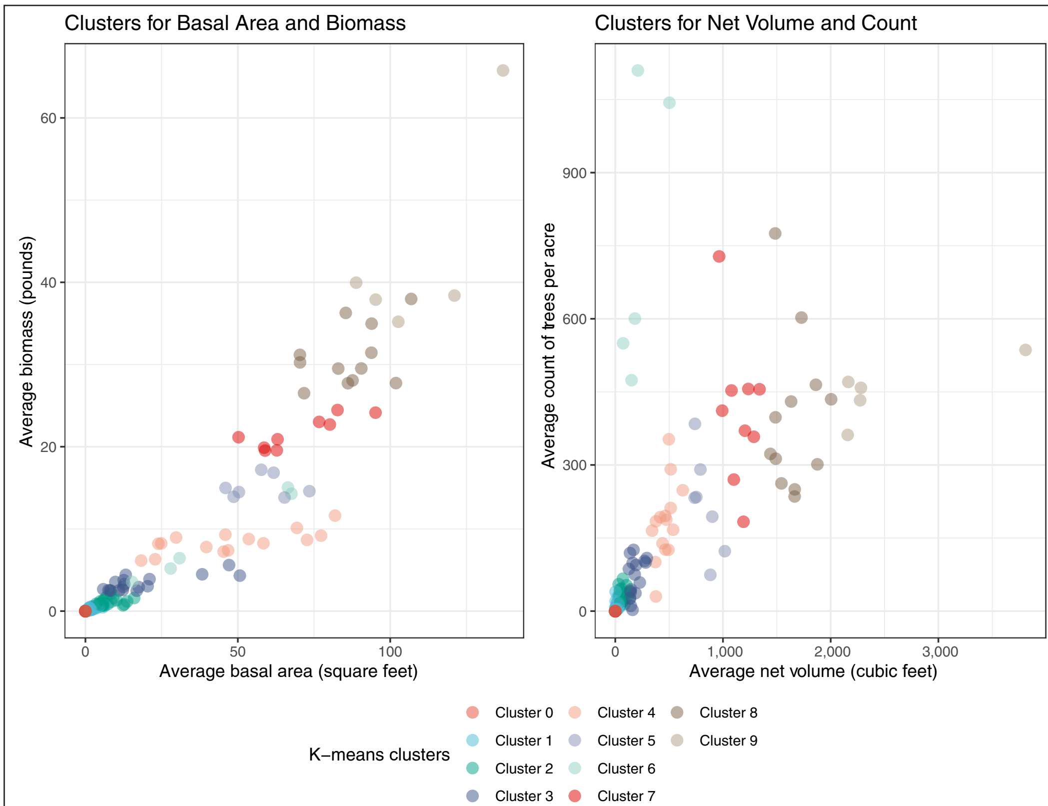

Figure 3—Scatterplots of the response variables, colored by the 10 k-means clusters at the existing vegetation type (EVT) classification level.

Table 1—Final existing vegetation type (EVT) cluster descriptions (*k* = 5)

Cluster	Cluster description	Most common EVT class names (count)
0	All response variables equal zero	• Western Cool Temperate Wheat (1,922) • Sonoran Desert Sparsely Vegetated (140) • Western Cool Temperate Undeveloped Ruderal Grassland (136)
1	Tends to be shrubland and prairie	• Inter-Mountain Basins Big Sagebrush Shrubland (7,323) • Northwestern Great Plains Mixedgrass Prairie (4,836) • Western Great Plains Shortgrass Prairie (4,519)
2	Tends to be Pinyon-Juniper Woodland	• Colorado Plateau Pinyon-Juniper Woodland (3,419) • Great Basin Pinyon-Juniper Woodland (1,911) • Madrean Pinyon-Juniper Woodland (681)
3	Tends to be conifers with moderately high levels of biomass, basal area, and net volume	• Southern Rocky Mountain Ponderosa Pine Woodland (1,462) • Rocky Mountain Aspen Forest and Woodland (1,039) • Rocky Mountain Lodgepole Pine Forest (1,024)
4	Tends to be conifers with moderately high levels of biomass, basal area, and net volume	• Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest (1,627) • Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest (1,012) • Northern Rocky Mountain Dry-Mesic Montane Mixed-Conifer Forest (918)

Final results are also illustrated in figures 2 and 3. They are the result of clustering on the EVT class level with the inclusion of a *Cluster 0* group. The coarser scheme (where $k = 5$) is more collapsed, but it is less representative of subtle patterning in the dataset, especially regarding observations with low volume values but higher tree count values. The finer scheme (where $k = 10$) is less collapsed but may be more appropriate for use in situations where it is important for the EVT variable to be able to distinguish between different types of vegetation with low volumes and high counts. Each cluster name is our attempt to broadly describe the EVT classes typically found in that cluster. However, these names don't perfectly define all the EVT classes represented.

Each scheme and the clusters included within them are outlined further in tables 1 and 2. They display a more extensive name for each cluster based on the various EVT classes that appear within it. Additionally, the tables list the top three most common, in terms of number of ground plots, EVT classes within each cluster.

DISCUSSION

The goal of this project was to demonstrate an objective process for using classified remotely sensed data to support FIA. We used FIA's commonly reported attributes for describing structure characteristics of forest land, with the understanding that the classified EVT map characterizes plant community types across all land. The difficulties in defining discrete clusters can be attributed to both how the map classification scheme was designed as well as with the prediction error of the map product. The alignment was naturally challenging because we were trying to coerce species composition-based classes into classes that specifically reflected forest structure. Also, prediction error is always an issue when trying to apply any classification scheme across large geographic extents. These are common issues FIA faces when trying to build relationships with ground data to classified map products. We are hoping this objective process can shed some light on this issue.

Table 2—Final existing vegetation type (EVT) cluster descriptions (*k* = 10)

Cluster	Cluster description	Most common EVT class names (count)
0	All response variables equal zero	• Western Cool Temperate Wheat (1,922) • Sonoran Desert Sparsely Vegetated (140) • Western Cool Temperate Undeveloped Ruderal Grassland (136)
1	Tends to be shrubland and prairie	• Inter-Mountain Basins Big Sagebrush Shrubland (7,323) • Northwestern Great Plains Mixedgrass Prairie (4,836) • Western Great Plains Shortgrass Prairie (4,519)
2	Tends to be grassland and shrubland	• Inter-Mountain Basins Semi-Desert Grassland (1,417) • Artemisia Tridentata ssp. Vasenya Shrubland Alliance (1.413) • Inter-Mountain Basins Montane Sagebrush Steppe (889)
3	Tends to be grassland	• Northern Rocky Mountain Lower Montane-Foothill-Valley Grassland (811) • Rocky Mountain Subalpine-Montane Mesic Meadow (582) • Introduced Upland Vegetation-Perennial Grassland and Forbland (473)
4	Tends to be Pinyon-Juniper Woodland	• Colorado Plateau Pinyon-Juniper Woodland (3,419) • Great Basin Pinyon-Juniper Woodland (1,911) • Madrean Pinyon-Juniper Woodland (681)
5	Tends to be conifer woodlands	• Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna (267) • North American Warm Desert Riparian Forest and Woodland (105) • Southern Rocky Mountain Ponderosa Pine Savanna (99)
6	Tends to be shrubland with a high tree count	• *Quercus Gambelii* Shrubland Alliance (419) • Rocky Mountain Gambel Oak-Mixed Montane Shrubland (169) • Rocky Mountain Bigtooth Maple Ravine Woodland (76)
7	Tends to be conifers with moderately high levels of biomass, basal area, and net volume	• Southern Rocky Mountain Ponderosa Pine Woodland (1,462) • Rocky Mountain Aspen Forest and Woodland (1,039) • Northern Rocky Mountain Subalpine Woodland and Parkland (605)
8	Tends to be conifers with high levels of biomass, basal area, and net volume	• Rocky Mountain Lodgepole Pine Forest (1,024) • Dry-Mesic Montane Douglas-Fir Forest (894) • Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland (782)
9	Tends to be Spruce-Fir Forest with high levels of biomass, basal area, and net volume	• Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest (1,627) • Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest (1,012) • Northern Rocky Mountain Dry-Mesic Montane Mixed-Conifer Forest (918)

Our hope is that these schemes will allow for a more convenient use of the EVT variable by FIA. Rather than having to deal with an overwhelming 182 EVT classes in the Interior West, researchers will now be able to utilize the new classes we created in order to incorporate information from the EVT class into estimation, mapping, and analysis applications. Of course, caution must be exercised when using collapsed classes in model-assisted estimators to ensure the data used to create new clusters are not the same data used in estimation equations. Similarly when building predictive models for mapping, care should be taken not to use the same data in cluster creation as is used in map validation.

The biggest difficulty we faced when collapsing the EVT variable revolved around trying to generate clusters which made sense from both a mathematical and an ecological standpoint. Our priority throughout the research process was to create groupings that were internally homogenous for our forest structure attributes, but fairly dissimilar from the other clusters created. However, it was also important for us to consider that our clusters were able to reconcile these statistically driven criteria with ecological distinctions. For example, we thought it was inappropriate for our clusters to group things like "Snow-Ice", "Barren", and "Water" observations with sparsely vegetated plant communities. In the same way, it was not useful to have Pinyon-Juniper type trees in different clusters. However, we were able to strike this balance between the ecological and the statistical through our various refinements of the k-means process. These refinements included constructing the *Cluster 0*, increasing the k value to expose patterns that were otherwise obscured, and clustering on the EVT class level rather than the plot level.

Researchers looking to continue refining the re-categorization of the EVT variable may want to turn to hierarchical clustering. Running a complete hierarchical clustering algorithm on the multitude of EVT classes may yield interesting results. It would be compelling to see if our proposed k-means schemes would be enhanced, corroborated, or contradicted by a procedure that arranges the EVT classes hierarchically. Another avenue for further inquiry may be to explore the geography of our clusters. Since many of the original EVT classes seemed to be geographically informed (especially with class names like "Colorado Plateau Pinyon-Juniper Woodland" and "Southern Rocky Mountain Ponderosa Pine Woodland"), it would be valuable to investigate if our suggested clusters are concentrated in particular regions of the Interior West or if they follow any notable geographic patterns.

ACKNOWLEDGMENTS

We would like to thank Wendy Goetz with the Geospatial Technology and Applications Center, Chris Toney with the USDA Forest Service, Rocky Mountain Research Station, Karen Schleeweis with the USDA Forest Service, Rocky Mountain Research Station, and Karen Short with the USDA Forest Service, Rocky Mountain Research Station for their very careful and thoughtful review of the manuscript. This work was conducted as part of the educational experience provided by the Forestry Data Science Lab at Reed College. The lab provides undergraduate students with real-world statistical work experience and brings diverse student perspectives to statistical and analytical challenges facing forest inventory. This work was supported by the USDA Forest Service Forest Inventory and Analysis Program, Reed College, and Swarthmore College. The authors would also like to thank the USDA Forest Service Forest Inventory and Analysis Program as well as the USDA Forest Service Remote Sensing Applications Center for the data.

LITERATURE CITED

Bechtold, W.A.; Patterson, P.L. 2005. The Enhanced Forest Inventory and Analysis program—National sampling design and estimation procedures. Gen. Tech. Rep. SRS–80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 85 p. Vol. 080.

Hartigan, J.A.; Wong, M.A. 1979. Algorithm AS 136: A K-means clustering algorithm. Applied Statistics. 28: 100–108. doi: 10.2307/2346830.

James, G.; Witten, D.; Hastie, T.; Tibshirani, R. 2013. An introduction to statistical learning. New York: Springer. 426 p. Vol. 112.

Moisen, G.G.; Frescino, T.S. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling. 157: 209–225.

NatureServe. 2018. International ecological classification standard: Terrestrial ecological classifications. NatureServe Central Databases. Arlington, VA. https://landfire.gov/documents/LANDFIRE_Natural_NVC_Groups_Descriptions_EasternUS.pdf. [Date accessed: October 7, 2019].

R Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. [Date accessed: June 1, 2019].

Rollins, M.G. 2009. LANDFIRE: A nationally consistent vegetation, wildland fire, and fuel assessment. International Journal of Wildland Fire. 18(3): 235–249.

Shaw, J.D.; Goeking, S.A.; Menlove, J.; Werstak, C.E., Jr. 2017. Assessment of fire effects based on Forest Inventory and Analysis data and a long-term fire mapping data set. Journal of Forestry. 115(4): 258–269.

U.S. Geological Survey (USGS). 2013. LANDFIRE.US_ASP. Wildland Fire Science, Earth Resources Observation and Science Center, U.S. Geological Survey. Sioux Falls, SD: USGS EROS. https://www.landfire.gov/topographic.php. [Date accessed: July 2, 2020].

U.S. Geological Survey (USGS). 2016. LANDFIRE Existing Vegetation Type. Wildland Fire Science, Earth Resources Observation and Science Center, U.S. Geological Survey. Sioux Falls, SD: USGS EROS. https://www.landfire.gov/topographic.php. [Date accessed: October 7, 2019].

APPENDIX

Look-up table for Interior West existing vegetation type (EVT) classification

evtLF	Class name	Coarse clusters	Fine clusters	Count
3968	Western Cool Temperate Wheat	Cluster 0	Cluster 0	1922
3223	Sonoran Desert Sparsely Vegetated	Cluster 0	Cluster 0	140
3944	Western Cool Temperate Undeveloped Ruderal Grassland	Cluster 0	Cluster 0	136
3985	Western Warm Temperate Close Grown Crop	Cluster 0	Cluster 0	81
3298	Developed-High Intensity	Cluster 0	Cluster 0	44
3004	North American Warm Desert Sparsely Vegetated Systems	Cluster 0	Cluster 0	35
3913	Western Warm Temperate Urban Herbaceous	Cluster 0	Cluster 0	26
3504	Chihuahuan-Sonoran Desert Bottomland and Swale Grassland	Cluster 0	Cluster 0	22
3077	Chihuahuan Succulent Desert Scrub	Cluster 0	Cluster 0	21
3088	Sonora-Mojave Mixed Salt Desert Scrub	Cluster 0	Cluster 0	21
3094	Western Great Plains Sandhill Shrubland	Cluster 0	Cluster 0	21
3963	Western Cool Temperate Row Crop - Close Grown Crop	Cluster 0	Cluster 0	20
3988	Western Warm Temperate Wheat	Cluster 0	Cluster 0	20
3943	Western Cool Temperate Undeveloped Ruderal Shrubland	Cluster 0	Cluster 0	19
3980	Western Warm Temperate Orchard	Cluster 0	Cluster 0	13
3255	Inter-Mountain Basins Montane Riparian Shrubland	Cluster 0	Cluster 0	11
3987	Western Warm Temperate Pasture and Hayland	Cluster 0	Cluster 0	11
3211	*Grayia spinosa* Shrubland Alliance	Cluster 0	Cluster 0	10
3983	Western Warm Temperate Row Crop - Close Grown Crop	Cluster 0	Cluster 0	10
3134	Columbia Basin Foothill and Canyon Dry Grassland	Cluster 0	Cluster 0	7
3090	Sonoran Granite Outcrop Desert Scrub	Cluster 0	Cluster 0	4
3065	Columbia Plateau Scabland Shrubland	Cluster 0	Cluster 0	3
3925	Western Warm Temperate Developed Ruderal Deciduous Forest	Cluster 0	Cluster 0	2

Note: The first and second column represent the original EVT code and classification name. Column three is based on $k = 5$ and column 4 is based on $k = 10$. The last column represents the number of Interior West ground plots in that EVT class.

(continued)

Look-up table for Interior West existing vegetation type (EVT) classification (continued)

evtLF	Class name	Coarse clusters	Fine clusters	Count
3027	Mediterranean California Dry-Mesic Mixed Conifer Forest and Woodland	Cluster 0	Cluster 0	1
3138	North Pacific Montane Grassland	Cluster 0	Cluster 0	1
3926	Western Warm Temperate Developed Ruderal Evergreen Forest	Cluster 0	Cluster 0	1
3948	Western Warm Temperate Undeveloped Ruderal Shrubland	Cluster 0	Cluster 0	1
3911	Western Warm Temperate Urban Evergreen Forest	Cluster 0	Cluster 0	1
3080	Inter-Mountain Basins Big Sagebrush Shrubland	Cluster 1	Cluster 1	7323
3141	Northwestern Great Plains Mixedgrass Prairie	Cluster 1	Cluster 1	4836
3149	Western Great Plains Shortgrass Prairie	Cluster 1	Cluster 1	4519
3125	Inter-Mountain Basins Big Sagebrush Steppe	Cluster 1	Cluster 1	3100
3081	Inter-Mountain Basins Mixed Salt Desert Scrub	Cluster 1	Cluster 1	3006
3109	Sonoran Paloverde-Mixed Cacti Desert Scrub	Cluster 1	Cluster 1	2072
3294	Barren	Cluster 1	Cluster 1	2039
3181	Introduced Upland Vegetation-Annual Grassland	Cluster 1	Cluster 1	1949
3153	Inter-Mountain Basins Greasewood Flat	Cluster 1	Cluster 1	1910
3256	Apacherian-Chihuahuan Semi-Desert Grassland	Cluster 1	Cluster 1	1480
3079	Great Basin Xeric Mixed Sagebrush Shrubland	Cluster 1	Cluster 1	1458
3135	Inter-Mountain Basins Semi-Desert Grassland	Cluster 1	Cluster 2	1417
3220	Artemisia tridentata ssp. vaseyana Shrubland Alliance	Cluster 1	Cluster 2	1413
3087	Sonora-Mojave Creosotebush-White Bursage Desert Scrub	Cluster 1	Cluster 1	1253
3095	Apacherian-Chihuahuan Mesquite Upland Scrub	Cluster 1	Cluster 1	1124
3924	Western Cool Temperate Developed Ruderal Grassland	Cluster 1	Cluster 1	1094
3965	Western Cool Temperate Close Grown Crop	Cluster 1	Cluster 1	1064
3966	Western Cool Temperate Fallow/Idle Cropland	Cluster 1	Cluster 1	1060
3082	Mojave Mid-Elevation Mixed Desert Scrub	Cluster 1	Cluster 1	989
3126	Inter-Mountain Basins Montane Sagebrush Steppe	Cluster 1	Cluster 2	889
3139	Northern Rocky Mountain Lower Montane-Foothill-Valley Grassland	Cluster 1	Cluster 3	811
3292	Open Water	Cluster 1	Cluster 1	782
3210	Coleogyne ramosissima Shrubland Alliance	Cluster 1	Cluster 1	731
3074	Chihuahuan Creosotebush Desert Scrub	Cluster 1	Cluster 1	706
3964	Western Cool Temperate Row Crop	Cluster 1	Cluster 1	696
3299	Developed-Roads	Cluster 1	Cluster 1	692
3219	Inter-Mountain Basins Sparsely Vegetated Systems II	Cluster 1	Cluster 2	673
3093	Southern Colorado Plateau Sand Shrubland	Cluster 1	Cluster 1	622
3145	Rocky Mountain Subalpine-Montane Mesic Meadow	Cluster 1	Cluster 3	582
3148	Western Great Plains Sand Prairie Grassland	Cluster 1	Cluster 1	548
3967	Western Cool Temperate Pasture and Hayland	Cluster 1	Cluster 1	515
3212	Western Great Plains Sandhill Grassland	Cluster 1	Cluster 1	498
3100	Chihuahuan Mixed Desert and Thornscrub	Cluster 1	Cluster 1	497
3182	Introduced Upland Vegetation-Perennial Grassland and Forbland	Cluster 1	Cluster 3	473
3127	Inter-Mountain Basins Semi-Desert Shrub-Steppe	Cluster 1	Cluster 2	464

Note: The first and second column represent the original EVT code and classification name. Column three is based on $k = 5$ and column 4 is based on $k = 10$. The last column represents the number of Interior West ground plots in that EVT class.

(continued)

Look-up table for Interior West existing vegetation type (EVT) classification (continued)

evtLF	Class name	Coarse clusters	Fine clusters	Count
3104	Mogollon Chaparral	Cluster 1	Cluster 2	441
3183	Introduced Upland Vegetation-Annual and Biennial Forbland	Cluster 1	Cluster 1	404
3066	Inter-Mountain Basins Mat Saltbush Shrubland	Cluster 1	Cluster 1	394
3086	Rocky Mountain Lower Montane-Foothill Shrubland	Cluster 1	Cluster 2	376
3923	Western Cool Temperate Developed Ruderal Shrubland	Cluster 1	Cluster 1	373
3218	North American Warm Desert Sparsely Vegetated Systems II	Cluster 1	Cluster 2	357
3124	Columbia Plateau Low Sagebrush Steppe	Cluster 1	Cluster 1	335
3006	Rocky Mountain Alpine/Montane Sparsely Vegetated Systems	Cluster 1	Cluster 3	330
3162	Western Great Plains Floodplain Forest and Woodland	Cluster 1	Cluster 3	325
3085	Northwestern Great Plains Shrubland	Cluster 1	Cluster 2	307
3001	Inter-Mountain Basins Sparsely Vegetated Systems	Cluster 1	Cluster 2	290
3503	Chihuahuan Loamy Plains Desert Grassland	Cluster 1	Cluster 1	262
3064	Colorado Plateau Mixed Low Sagebrush Shrubland	Cluster 1	Cluster 2	258
3296	Developed-Low Intensity	Cluster 1	Cluster 1	241
3106	Northern Rocky Mountain Montane-Foothill Deciduous Shrubland	Cluster 1	Cluster 2	199
3258	North American Warm Desert Riparian Herbaceous	Cluster 1	Cluster 1	193
3215	*Quercus turbinella* Shrubland Alliance	Cluster 1	Cluster 3	174
3123	Columbia Plateau Steppe and Grassland	Cluster 1	Cluster 1	144
3204	Western Great Plains Mesquite Shrubland	Cluster 1	Cluster 1	142
3132	Central Mixedgrass Prairie Grassland	Cluster 1	Cluster 1	139
3904	Western Cool Temperate Urban Shrubland	Cluster 1	Cluster 2	139
3076	Chihuahuan Stabilized Coppice Dune and Sand Flat Scrub	Cluster 1	Cluster 1	136
3986	Western Warm Temperate Fallow/Idle Cropland	Cluster 1	Cluster 1	136
3075	Chihuahuan Mixed Salt Desert Scrub	Cluster 1	Cluster 1	131
3297	Developed-Medium Intensity	Cluster 1	Cluster 1	128
3903	Western Cool Temperate Urban Herbaceous	Cluster 1	Cluster 1	126
3007	Western Great Plains Sparsely Vegetated Systems	Cluster 1	Cluster 1	113
3928	Western Warm Temperate Developed Ruderal Shrubland	Cluster 1	Cluster 1	94
3091	Sonoran Mid-Elevation Desert Scrub	Cluster 1	Cluster 1	93
3147	Western Great Plains Foothill and Piedmont Grassland	Cluster 1	Cluster 2	92
3213	*Quercus havardii* Shrubland Alliance	Cluster 1	Cluster 1	88
3133	Chihuahuan Sandy Plains Semi-Desert Grassland	Cluster 1	Cluster 2	86
3144	Rocky Mountain Alpine Turf	Cluster 1	Cluster 2	80
3072	Wyoming Basins Dwarf Sagebrush Shrubland and Steppe	Cluster 1	Cluster 1	75
3259	Introduced Riparian Shrubland	Cluster 1	Cluster 1	68
3164	Rocky Mountain Wetland-Herbaceous	Cluster 1	Cluster 2	64
3984	Western Warm Temperate Row Crop	Cluster 1	Cluster 1	61
3121	Apacherian-Chihuahuan Semi-Desert Shrubland	Cluster 1	Cluster 1	56
3116	Madrean Juniper Savanna	Cluster 1	Cluster 3	54
3295	Quarries-Strip Mines-Gravel Pits	Cluster 1	Cluster 1	51

Note: The first and second column represent the original EVT code and classification name. Column three is based on $k = 5$ and column 4 is based on $k = 10$. The last column represents the number of Interior West ground plots in that EVT class.

(continued)

Look-up table for Interior West existing vegetation type (EVT) classification (continued)

evtLF	Class name	Coarse clusters	Fine clusters	Count
3495	Western Great Plains Depressional Wetland Systems	Cluster 1	Cluster 1	50
3115	Inter-Mountain Basins Juniper Savanna	Cluster 1	Cluster 3	47
3293	Snow-Ice	Cluster 1	Cluster 2	47
3222	Rocky Mountain Alpine/Montane Sparsely Vegetated Systems II	Cluster 1	Cluster 3	46
3914	Western Warm Temperate Urban Shrubland	Cluster 1	Cluster 1	41
3250	Inter-Mountain Basins Curl-leaf Mountain Mahogany Shrubland	Cluster 1	Cluster 1	37
3070	Rocky Mountain Alpine Dwarf-Shrubland	Cluster 1	Cluster 3	36
3119	Southern Rocky Mountain Juniper Woodland and Savanna	Cluster 1	Cluster 3	32
3900	Western Cool Temperate Urban Deciduous Forest	Cluster 1	Cluster 3	32
3254	Western Great Plains Floodplain Herbaceous	Cluster 1	Cluster 3	30
3251	Rocky Mountain Montane Riparian Shrubland	Cluster 1	Cluster 3	27
3180	Introduced Riparian Forest and Woodland	Cluster 1	Cluster 2	24
3078	Colorado Plateau Blackbrush-Mormon-tea Shrubland	Cluster 1	Cluster 1	21
3103	Great Basin Semi-Desert Chaparral	Cluster 1	Cluster 3	21
3108	Sonora-Mojave Semi-Desert Chaparral	Cluster 1	Cluster 1	21
3921	Western Cool Temperate Developed Ruderal Evergreen Forest	Cluster 1	Cluster 2	20
3253	Western Great Plains Floodplain Shrubland	Cluster 1	Cluster 1	15
3385	Western Great Plains Wooded Draw and Ravine	Cluster 1	Cluster 3	15
3101	Madrean Oriental Chaparral	Cluster 1	Cluster 2	12
3929	Western Warm Temperate Developed Ruderal Grassland	Cluster 1	Cluster 1	11
3216	*Cercocarpus montanus* Shrubland Alliance	Cluster 1	Cluster 2	8
3122	Chihuahuan *Gypsophilous* Grassland and Steppe	Cluster 1	Cluster 2	8
3902	Western Cool Temperate Urban Mixed Forest	Cluster 1	Cluster 3	6
3910	Western Warm Temperate Urban Deciduous Forest	Cluster 1	Cluster 1	5
3214	*Arctostaphylos patula* Shrubland Alliance	Cluster 1	Cluster 3	4
3221	Mediterranean California Sparsely Vegetated Systems II	Cluster 1	Cluster 1	4
3016	Colorado Plateau Pinyon-Juniper Woodland	Cluster 2	Cluster 4	3419
3019	Great Basin Pinyon-Juniper Woodland	Cluster 2	Cluster 4	1911
3025	Madrean Pinyon-Juniper Woodland	Cluster 2	Cluster 4	681
3217	*Quercus gambelii* Shrubland Alliance	Cluster 2	Cluster 6	419
3059	Southern Rocky Mountain Pinyon-Juniper Woodland	Cluster 2	Cluster 4	405
3179	Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna	Cluster 2	Cluster 5	267
3062	Inter-Mountain Basins Curl-leaf Mountain Mahogany Woodland	Cluster 2	Cluster 4	261
3146	Southern Rocky Mountain Montane-Subalpine Grassland	Cluster 2	Cluster 4	185
3140	Northern Rocky Mountain Subalpine-Upper Montane Grassland	Cluster 2	Cluster 4	178
3107	Rocky Mountain Gambel Oak-Mixed Montane Shrubland	Cluster 2	Cluster 6	169
3169	Northern Rocky Mountain Subalpine Deciduous Shrubland	Cluster 2	Cluster 4	157
3023	Madrean Encinal	Cluster 2	Cluster 4	127
3252	Rocky Mountain Subalpine/Upper Montane Riparian Shrubland	Cluster 2	Cluster 4	118
3155	North American Warm Desert Riparian Forest and Woodland	Cluster 2	Cluster 5	105

Note: The first and second column represent the original EVT code and classification name. Column three is based on $k = 5$ and column 4 is based on $k = 10$. The last column represents the number of Interior West ground plots in that EVT class.

(continued)

Look-up table for Interior West existing vegetation type (EVT) classification (continued)

evtLF	Class name	Coarse clusters	Fine clusters	Count
3117	Southern Rocky Mountain Ponderosa Pine Savanna	Cluster 2	Cluster 5	99
3049	Rocky Mountain Foothill Limber Pine-Juniper Woodland	Cluster 2	Cluster 4	94
3012	Rocky Mountain Bigtooth Maple Ravine Woodland	Cluster 2	Cluster 6	76
3154	Inter-Mountain Basins Montane Riparian Forest and Woodland	Cluster 2	Cluster 4	71
3017	Columbia Plateau Western Juniper Woodland and Savanna	Cluster 2	Cluster 4	51
3031	California Montane Jeffrey Pine(-Ponderosa Pine) Woodland	Cluster 2	Cluster 4	24
3901	Western Cool Temperate Urban Evergreen Forest	Cluster 2	Cluster 5	12
3920	Western Cool Temperate Developed Ruderal Deciduous Forest	Cluster 2	Cluster 6	11
3098	California Montane Woodland and Chaparral	Cluster 2	Cluster 5	4
3232	*Abies grandis* Forest Forest	Cluster 2	Cluster 4	2
3026	Madrean Upper Montane Conifer-Oak Forest and Woodland	Cluster 2	Cluster 5	2
3009	Northwestern Great Plains Aspen Forest and Parkland	Cluster 2	Cluster 6	1
3054	Southern Rocky Mountain Ponderosa Pine Woodland	Cluster 3	Cluster 7	1462
3011	Rocky Mountain Aspen Forest and Woodland	Cluster 3	Cluster 7	1039
3050	Rocky Mountain Lodgepole Pine Forest	Cluster 3	Cluster 8	1024
3061	Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland	Cluster 3	Cluster 8	782
3046	Northern Rocky Mountain Subalpine Woodland and Parkland	Cluster 3	Cluster 7	605
3235	Xeric Montane Douglas-fir Forest	Cluster 3	Cluster 8	603
3051	Southern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest and Woodland	Cluster 3	Cluster 7	545
3159	Rocky Mountain Montane Riparian Forest and Woodland	Cluster 3	Cluster 7	479
3166	Middle Rocky Mountain Montane Douglas-fir Forest and Woodland	Cluster 3	Cluster 8	298
3024	Madrean Lower Montane Pine-Oak Forest and Woodland	Cluster 3	Cluster 7	170
3053	Northern Rocky Mountain Ponderosa Pine Woodland and Savanna	Cluster 3	Cluster 8	157
3234	Mesic Montane Douglas-fir Forest	Cluster 3	Cluster 8	89
3167	Rocky Mountain Poor-Site Lodgepole Pine Forest	Cluster 3	Cluster 8	71
3208	*Abies concolor* Forest Alliance	Cluster 3	Cluster 8	58
3160	Rocky Mountain Subalpine/Upper Montane Riparian Forest and Woodland	Cluster 3	Cluster 7	46
3161	Northern Rocky Mountain Conifer Swamp	Cluster 3	Cluster 7	29
3057	Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland	Cluster 3	Cluster 8	23
3020	Inter-Mountain Basins Subalpine Limber-Bristlecone Pine Woodland	Cluster 3	Cluster 5	15
3228	Dry-mesic Montane Western Larch Forest	Cluster 3	Cluster 7	9
3055	Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	Cluster 4	Cluster 9	1627
3056	Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodland	Cluster 4	Cluster 9	1012
3045	Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest	Cluster 4	Cluster 9	918
3227	Dry-mesic Montane Douglas-fir Forest	Cluster 4	Cluster 8	894
3047	Northern Rocky Mountain Mesic Montane Mixed Conifer Forest	Cluster 4	Cluster 9	573
3052	Southern Rocky Mountain Mesic Montane Mixed Conifer Forest and Woodland	Cluster 4	Cluster 8	395
3233	Subalpine Douglas-fir Forest	Cluster 4	Cluster 8	237
3032	Mediterranean California Red Fir Forest	Cluster 4	Cluster 9	12

Note: The first and second column represent the original EVT code and classification name. Column three is based on $k = 5$ and column 4 is based on $k = 10$. The last column represents the number of Interior West ground plots in that EVT class.