Expert Attribution for Auto-Key Improvements (LANDFIRE) and Advancing Methods for integration with the revised US-National Vegetation Classification Standard: GeoArea 5

FINAL REPORT

Prepared by

NatureServe

For the NPS Vegetation Inventory Program & LANDFIRE

30 June 2012

© Shannon Menard, Horsethief Canyon, KS

NatureServe

A Network Connecting Science With Conservation

Contact information:
Prepared by Patrick J. Comer, Chief Ecologist
Email: **pat_comer@natureserve.org**
Phone: 703.797.4802
Mailing address: NatureServe
4001 Discovery Dr. Suite 2110
Boulder, CO 80303

Acknowledgements
This work was completed with funding provided by the inter-agency LANDFIRE Program through the National Park Service's Vegetation Inventory Program. It is the final report as required for Task Order J2340100052 under Cooperative Agreement H2380-04-0002 between NatureServe and the NPS. Additional funding was provided by The Nature Conservancy, under the auspices of their North American Science program. The Interagency LANDFIRE Program provided all sample data used here. An inter-agency team, including representatives from NatureServe, The Nature Conservancy, Forest Service Rocky Mountain Research Station, Forest Service FIA, and USGS- Gap Analysis Program provided project design, support, and analysis. Results reported here emphasize expert evaluations provided primarily by Shannon Menard and Jim Drake, NatureServe Regional Vegetation Ecologists.

Table of Contents

Figures

Tables

Introduction

The Inter-agency LANDFIRE Program implemented a series of new procedures and tools for processing vegetation sample plot data to rapidly supply geo-referenced samples for dynamics modeling and vegetation mapping. This effort made substantial advances in processing several hundred thousand vegetation plots nationwide, including standardizing many sample attributes (species taxonomy, structural classes, etc.) and applying labels reflecting the LANDFIRE map legend. However, given the pace of project activity, there was limited time to identify systematic error within the processing *auto-keys* and internalize lessons learned to improve technical procedures. There was also limited ability to develop an expert-reviewed, independent sample data set for use in map accuracy assessment. Additionally, given recent developments, there is a desire to adopt the revised US-National Vegetation Classification (US-NVC) for future mapping of existing vegetation types as part of the LANDFIRE effort.

This project represents a cooperative research effort with federal agency partners to systematically review the results of automated sample plot labeling (*auto-keys*), identify sources of systematic error, and clarify needs for technical improvements. Through this review process, comparisons between the existing LANDFIRE map legend and new types described the US-NVC were evaluated and documented. The effort has also generated an expert-reviewed, independent sample data set for use in map accuracy assessment nationwide.

Project Goals

- Identify “accuracy” issues with the existing auto-keys and resultant labels.
- Identify spatial or thematic gaps in the current LANDFIRE national reference database.
- Develop recommended solutions/approaches to issues encountered.
- Build an independent data set that could be used in other applicable mapping projects (GAP, regional wildlife, state habitat maps, etc.).
- Identify issues specific to labeling training data based on the newly adopted National Vegetation Classification Standard hierarchy.
- Identify and document appropriate updates to NPS vegetation field methods documentation.

In-kind contributions to this effort have come from federal agency partners, including USGS (Gap Analysis Program and Earth Resources Observation and Science (EROS) Data Center), US Forest Service Rocky Mountain Research Station (RMRS) and Forest Inventory Analysis (FIA)), among others. The National Park Service retains considerable expertise in the use of project outputs and benefits directly from project outcomes. NatureServe ecologists have contributed expertise in U.S. vegetation types and processing procedures, and development of the LANDFIRE *auto-key* tools.

Background on LANDFIRE Auto-keys

A major need and hence objective of LANDFIRE was to compile geo-referenced vegetation data for the entire United States. These data needed to be combined into a single database and attributed in a consistent, repeatable fashion to NatureServe’s Terrestrial Ecological Systems or a set of land use or land cover classes. Once attributed with ecological systems, the geo-referenced samples were used as training data in a mapping effort that utilized a modeling process whereby the samples were only one of several inputs to the model. Systems for Environmental Management (SEM), based in Missoula MT, was contracted by LANDFIRE to compile the LANDFIRE Reference Database, or LFRDB, of all relatively recent, geo-referenced vegetation samples (also called “plots”) that could be obtained and processed.

LANDFIRE contracted with NatureServe to work with the LANDFIRE team to develop a methodology to automate attribution of the samples contained in the LFRDB to ecological systems or the other standardized land use/land cover classes. Prototyping and testing of this methodology evolved over several months in 2004 into a process involving two components: a set of floristic and structural rules for each vegetation type, and a computer application to use the plots from the LFRDB and the rules as inputs to generate results useable by LANDFIRE's mapping teams. The sets of floristic rules or criteria are now known as Sequence Tables, and the software application is called the Auto-key.

One of the main requirements for LANDFIRE map units was that they be differentiated floristically. Since abiotic variables were not consistently available for every plot, contextual landscape or abiotic information could not be used to differentiate vegetation types represented by the plots. In addition, sequence tables were intended to work with regional-scale patterns, as opposed to more local-scales. Thus keying each plot using only the required floristic data was the best way to assign a map unit to each plot.

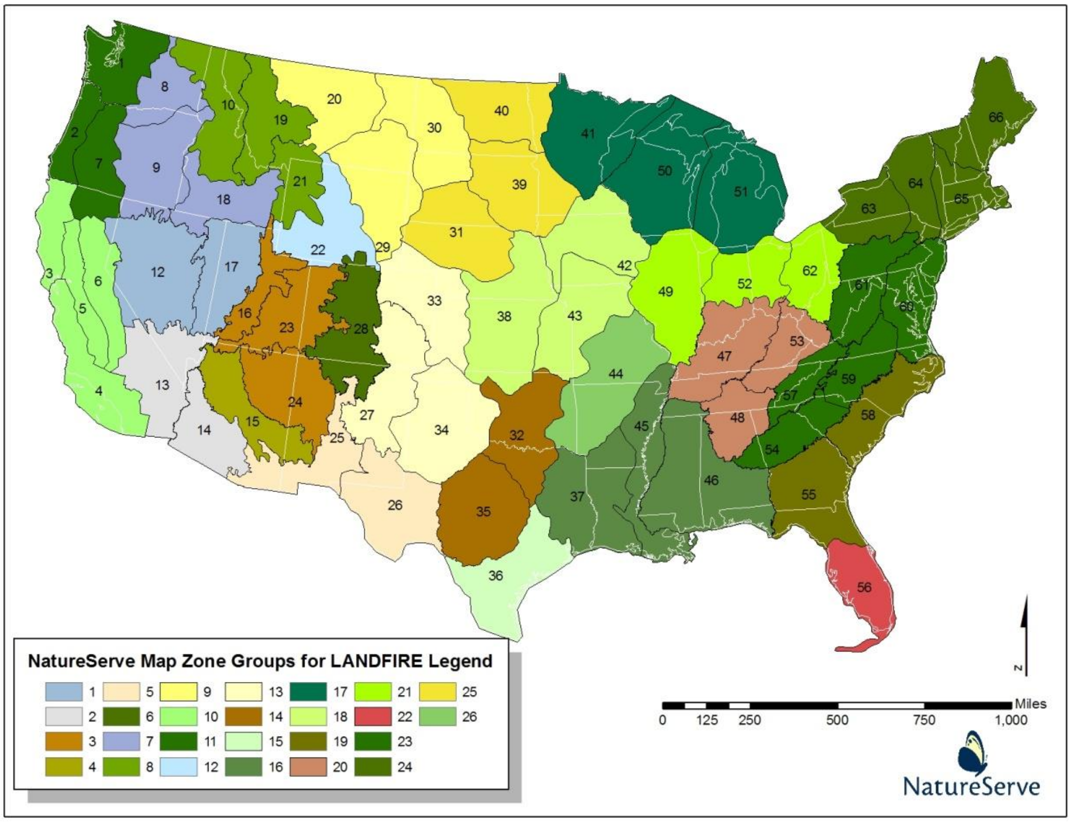
LANDFIRE's short-term needs, and long-term plans, required a repeatable methodology, consistently applied rules to categorize each reference sample, and documentation of the criteria applied. In essence, sequence tables codify the criteria and methods for keying geo-referenced vegetation data to a land cover class, whether it's an ecological system or some other vegetation category. Because of this, the methods are repeatable by anyone who may not necessarily be familiar with the vegetation of the region covered by a particular sequence table.

More details about this methodology include:

1. Each LANDFIRE sequence table was designed to efficiently automate keying of thousands to 10's of thousands geo-referenced vegetation samples to the LANDFIRE map units, which included both Ecological Systems for the 'natural' portions of the landscape, and a variety of land use or land cover classes for the remainder. The objective was to accurately key as many samples as possible, not to attempt to key all samples.
2. Each sequence table was created to key to systems and mappable US-NVC alliances in an ecologically-related geographic area, utilizing the MRLC map zones. There are 66 map zones for the conterminous US. NatureServe developed 26 sequence tables for these 66 map zones (Figure 1).
3. LANDFIRE also contracted with NatureServe to have dichotomous field keys written for all of the U.S. map zones. These keys were developed to cover the same map zones clusters as the sequence tables, and are available in MS Word documents for all of the U.S.
4. From a data processing standpoint, the vegetation samples first had to be formatted to match the specifications of the auto-key program created by USFS Missoula Fire Lab staff. We do not detail these formatting requirements here, as they are rather complex, and are completed by LANDFIRE contractors.
5. The sequence tables and vegetation samples are run through an automated Python application, developed by staff at the Missoula Fire Lab, called the "auto-key". The auto-key program sequentially compares each vegetation sample against criteria contained in the sequence table. Each ecological system type is represented in the sequence table via a set of vegetation composition criteria, which are organized in a particular order, or "sequence" (hence Sequence Table, or SQT). Each plot or point must meet all of the criteria for a particular ecological system, as represented by one sequence. If the sample meets all the criteria, the auto-key attributes the plot with the ecological system code and name. Samples which do not meet the criteria for a system can be attributed either with a more generic label, such as "unclassified forest and woodland", or else go through the entire SQT without keying and are attributed with "none".

Other land cover classes, such as introduced annual grasslands, or introduced riparian woody vegetation, are also included in a SQT to appropriately attribute any vegetation samples representing those land cover classes.

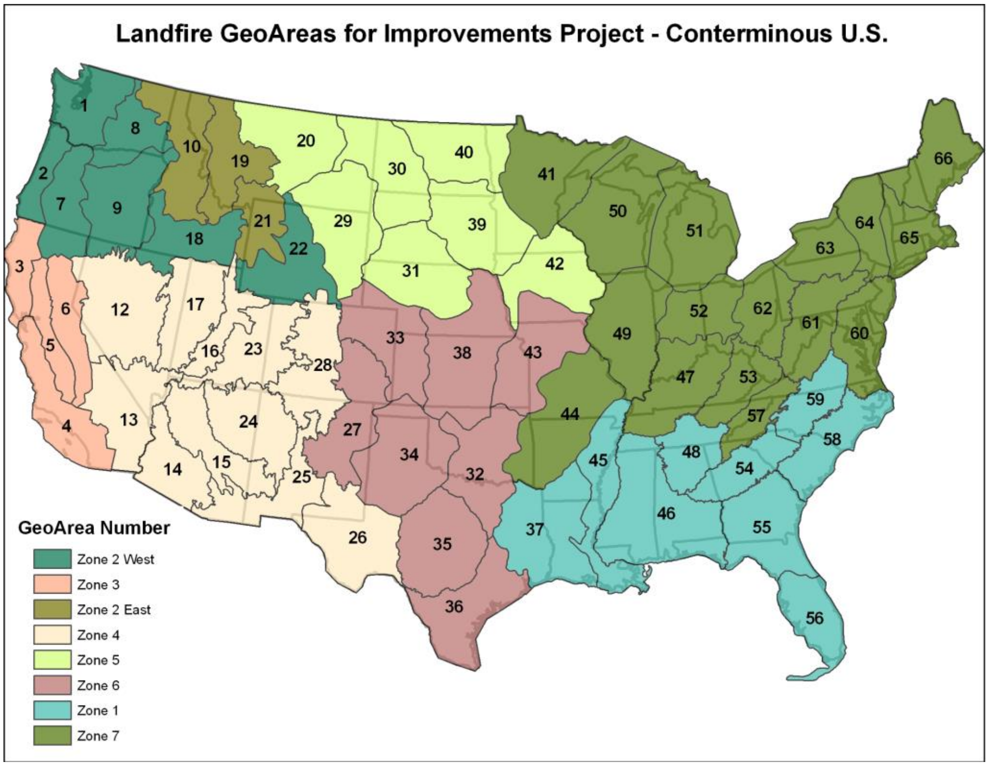
Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in the coterminous U.S.



Methods

For the LANDFIRE effort, both dichotomous field keys and auto-keys were developed for map legend classes and organized in a series of 17 map zone groupings that spanned the nation. For ongoing maintenance of national map products, the map zone groups have been further aggregated by LANDFIRE into larger geographic areas (GeoAreas). This project was organized around a modified form of these LANDFIRE GeoAreas (Figure 2). Within each GeoArea, project ecologists were provided with a subset of sample data for each relevant LANDFIRE map class (up to 30 sample plots). Using sample data on vegetation composition and structure, along with limited mapped ancillary data (for general orientation and ecological context), ecologists applied a map legend label to each sample. They documented their expert process for making label assignments, highlighting key pieces of information they used to arrive at their determination. The expert assignments were then compared to those previously applied through the LANDFIRE auto-keys assignments on spatially located field plots. Contingency tables were developed, analyzed, and documented. Key outcomes from each expert analysis include the contingency table, systematic discrepancies between expert and auto-key labels, and recommended changes to the auto-keys and technical procedures.

Figure 2. Modified LANDFIRE GeoAreas in the conterminous U.S. for use in this project.



Sample data were segmented by those that were used directly in LANDFIRE map production versus those that were held aside for use in accuracy assessment. Therefore, an expert-reviewed, independent sample data set for accuracy assessment was an additional project outcome. Expert ecologists were also be well-positioned to evaluate the results of auto-key assignments for LANDFIRE map legend classes in light of the related NVC Group and Macrogroup vegetation concepts that have been established and described.

For the expert reviews, the team needed to first determine the plot data available for use in the project and the sample design for selecting a subset of those plots. Secondly an evaluation was required of what kinds of data are contained in the plots that could be used for the expert review. The analysis team obtained counts of plots by map zone, GeoArea and system or land cover type, as well as counts of how many were used as training data in the mapping effort, or were withheld and used as the initial accuracy assessment plots. Additional counts were obtained for the number of plots acquired after the LANDFIRE mapping effort was completed in each GeoArea. A series of calls were held to discuss the number and distribution of plots by system type to be used in a “sample draw” for the expert review. Once the number of plots by system type by GeoArea was decided upon, the sample draw was completed by TNC and EROS team members, by selecting plots for each system randomly across all map zones in the GeoArea, with "independent" plots (not used in the original mapping effort) given selection priority.

The analysis team then reviewed in detail the available data tables and fields that are stored and managed in the LANDFIRE Reference Database (LFRDB). The data in the LFRDB is derived from many source datasets of varying quality and completeness. In addition, many plots in the LFRDB for forest types were provided by the Forest Inventory and Analysis (FIA) program, which has restrictions on sharing of their data. The discussions about what data to provide the experts for use in the labeling centered around:

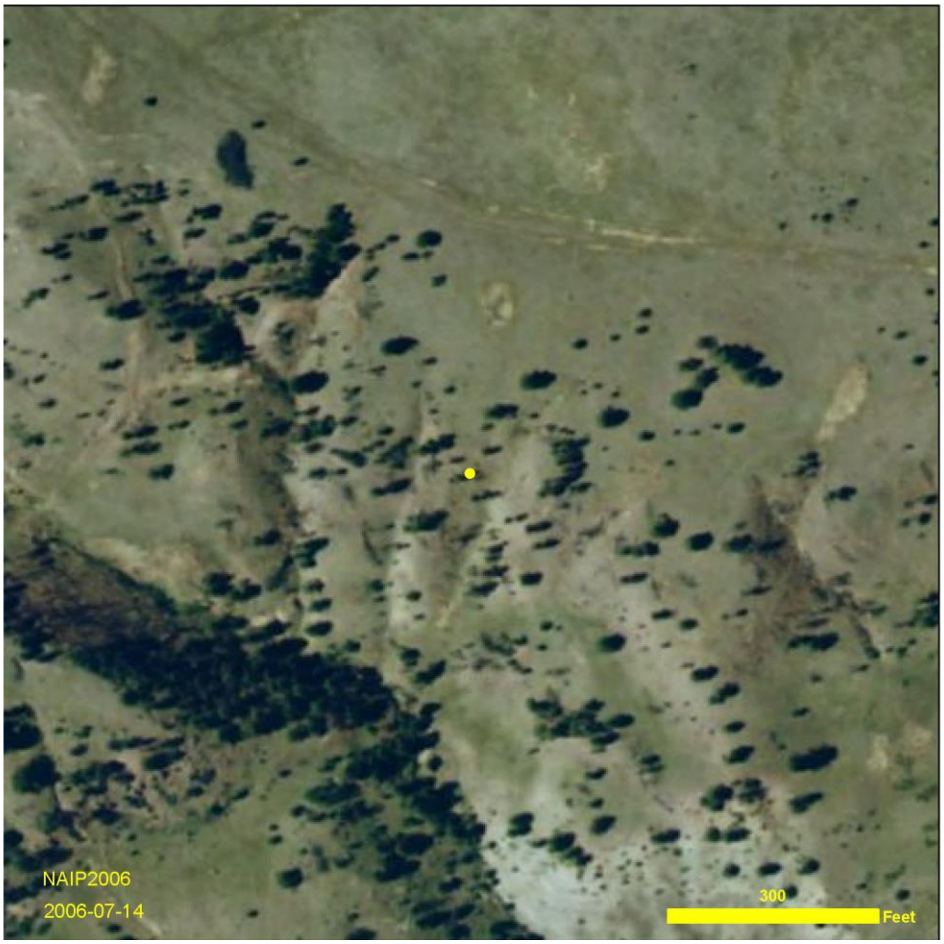
1. Providing the same data as are used in the auto-key procedures
2. Providing additional data that were not originally used in the auto-keys, and
3. Maintaining the "privacy" of the FIA data, ensuring the experts could not determine which plots were FIA vs not

Table 1 is a list of the general categories of data that were extracted from the LFRDB and provided to the experts for use in their review. After much discussion, it was also determined to provide a remotely-sensed image clip for each plot, as well as between 1 and 3 on-the-ground photos for the plot if such were available from the original data providers. These images provide some context for the expert reviewer, without revealing the exact location of the plot. The image clips were created automatically from the plot coordinates, and in the lower 48 were from NAIP imagery. All images were of the same scale, with the plot location a dot in the center of the image (Figure 3 is an example).

Table 1. Categories & fields of data provided to expert during review process.

Data category	Fields	Notes
Vegetation Structure	% cover of trees, shrubs, herbs, trees per acre, height of trees or shrubs	Values are calculated from source data & stored in LFRDB
Floristic composition	complete species list, % cover by species, nativity, height if available	Species list & % cover values are from the original source data, but other fields were derived by LANDFIRE
Dominant species	the 2 most dominant species within the major lifeform of the plot	The dominant and codominant species are provided, with % cover; the species are drawn from the dominant lifeform category of the plot (e.g. shrub dominated plots will have shrub species listed)
Geographic setting	map zone, USFS subsection, TNC ecoregion	These are derived by LANDFIRE from the coordinates of the plot
Landscape setting	elevation, aspect, slope	Values are derived form a DEM for the coordinates of the plot
Field notes	comments from field crew	Original field crew comments, if available
Image clips	Single image, same areal extent/scale for all plots	NAIP imagery was used for coterminous U.S. plots; coordinates in center of the image; no other locational information provided.

Figure 3. Example of an image clip for one plot in GeoArea 5.



NatureServe developed a MS Access 2007 relational database (the Expert Attribution Database, EADB) for use in the project. A user interface was designed to link to the above LFRDB data (provided by EROS in a separate LFRDB), the image clip, and any ground-photos in easily navigated forms for review by the expert. An additional form allowed the expert to select from a subset of system types when labeling plots. The reviewer was required to select from the ecological systems known or highly probable to occur in the GeoArea. If the expert could not label the plot with a system type, then "can't assign" was an additional option. All plots also required a confidence in label assignment (high, medium, low) and the expert was asked to document in comments why they assigned that confidence, or why they could not assign it to an ecological system.

After the expert reviews were completed for a particular GeoArea, the results were run through several quality control procedures to check for plots missing labels, or other discrepancies in the resulting data. Then a number of queries were run in the Access database, to generate summary statistics for each GeoArea, comparing labels on plots from the auto-keys and the experts.

Analysis Team

- Patrick Comer, NatureServe
- NatureServe Regional Ecologists (Marion Reid, Kristin Snow, Mary Harkness, Gwen Kittel, Keith Schulz, Mark Hall, Milo Pyne, Carl Nordman, Judy Teague, Lesley Sneddon, Jim Drake, Shannon Menard)
- Anne Davidson, GAP

- Don Long, USFS RMRS
- Brenda Lundberg, EROS
- Chris Toney, USFS FIA
- Alexa McKerrow, GAP
- Gretchen Meier, EROS
- Chris Lea, NPS
- Jim Smith, TNC, Overall Coordinator

Intended Products of this Effort

2.1 Tabular comparisons (as contingency tables) between LANDFIRE auto-key assignment and expert assignment for each GeoArea data set with an associated interpretation of the outcomes (systematic discrepancies between expert and auto-key labels, and recommended changes).

2.2 A report by each GeoArea detailing processes and results, specifically identifying how they made individual assignments.

2.3 A report that documents procedures and data elements to improve the auto-key process in each GeoArea.

2.4 A report that documents technical procedures to adapt auto-keys for labeling NVCS group, Macrogroup, and Division concepts.

2.5 Full data sets with independent assignments for each GeoArea in standard LFRDB format.

2.6 A single overall report with recommendations for all GeoAreas, including commonalities and unique issues.

Results

The following results for GeoArea 5 are organized according to these primary product deliverable categories:

2.1 Tabular comparisons (as contingency tables) between LF auto-key assignment and expert assignment for each GeoArea data set with an analysis and reports document (identified, systematic discrepancies between expert and auto-key labels, and recommended changes).

2.2 A report by each GeoArea detailing processes and results, specifically identifying how they made individual assignments.

2.3 A report that documents procedures and data elements to improve the auto-key process in each GeoArea.

2.4 A report that documents technical procedures to adapt auto-keys for labeling NVCS group, macrogroup, and division concepts.

GeoArea 5: Northern Plains

GeoArea 5 encompasses the northern Great Plains Steppe, Black Hills, Dakota Mixed Grass Prairie, Northern Tall Grass Prairie and Central Tall Grass Prairie ecoregions. This GeoArea includes a total of 9 map zones (20, 29, 30, 31, 38, 39, 40, 42, and 43; Figure 2) originally clustered into grouping for purposes of designing and implementing auto-keys (**Error! Reference source not found.**). The total umber of plots in this Geo Area analysis was 1,456. A total of 55 natural ecological system types were assigned to a total of 1,044 plots by the auto-keys. A total of 66 system types were assigned by experts (i.e., these included individual types that had been aggregated to broader classes by LANDFIRE for either sparsely vegetated types or wetland/riparian types).

An additional 20 types were assigned by the auto-key but were not assigned by experts:

- Boreal Aspen-Birch Forest
- Boreal White Spruce-Fir-Hardwood Forest
- Central Interior Highlands Calcareous Glade and Barrens
- Central Interior Highlands Dry Acidic Glade and Barrens
- Inter-Mountain Basins Semi-Desert Grassland
- Northern Rocky Mountain Foothill Conifer Wooded Steppe
- Northwestern Great Plains Canyon
- Ozark-Ouachita Dry-Mesic Oak Forest
- Western Great Plains Tallgrass Prairie
- Boreal Acidic Peatland Systems
- Central Interior and Appalachian Floodplain Systems
- Central Interior and Appalachian Shrub-Herbaceous Wetland Systems
- Central Interior and Appalachian Swamp Systems
- Eastern Great Plains Floodplain Systems
- Rocky Mountain Alpine/Montane Sparsely Vegetated Systems
- Rocky Mountain Montane Riparian Systems
- Rocky Mountain Subalpine/Upper Montane Riparian Systems
- Western Great Plains Depressional Wetland Systems
- Western Great Plains Floodplain Systems
- Western Great Plains Sparsely Vegetated Systems

The first nine types are uncommon in this GeoArea because it is at the edge of their geographic range, they are uncommon throughout their entire range, or both. The last 11 types are aggregates of individual Systems used by Landfire for mapping. The expert reviewers attributed sites to individual Systems and so would not have used these units in their review process.

Comparison of Auto-key and Expert Assignments

2.1 Tabular comparisons (as contingency tables) between LF auto-key assignment and expert assignment for each GeoArea data set with an analysis and reports document (identified, systematic discrepancies between expert and auto-key labels, and recommended changes).

Of the 55 natural types assigned labels by the auto-keys, 27 types (49%) had fewer than 10 samples available for this analysis (Table 2). All but five, Northwestern Great Plains Canyon, Rocky Mountain Foothill Limber Pine-Juniper Woodland, Paleozoic Plateau Bluff and Talus, Central Tallgrass Prairie, and North-Central Interior Oak Savanna, are at the edge of their geographic range in this GeoArea and are more common elsewhere. Most of these are types found more commonly in the Rocky Mountains or inter-mountain basins further west but that also occur in this GeoArea in the isolated mountain ranges and scattered drier habitats of MapZones 20 and 29 (essentially central and southeastern Montana and northeastern Wyoming). Of the five types that are concentrated within this GeoArea but still have <10 plots, four are truly rare across the landscape and the other, Paleozoic Plateau Bluff and Talus, is restricted to a relatively small area on the eastern edge of GeoArea 5 and has some of its occurrences in the adjacent GeoArea 7W.

Table 2. Under-sampled types within GeoArea 5.

EVT Code	EVT Name	System elcode	total Plots
2308	Crosstimbers Oak Forest and Woodland	CES205.682	9
2125	Inter-Mountain Basins Big Sagebrush Steppe	CES304.778	9
2066	Inter-Mountain Basins Mat Saltbush Shrubland	CES304.783	9
2051	Southern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest and Woodland	CES306.823	9
2049	Rocky Mountain Foothill Limber Pine-Juniper Woodland	CES306.955	9
2311	North-Central Interior Dry Oak Forest and Woodland	CES202.047	8
2341	Northwestern Great Plains Canyon	CES303.658	8
2062	Inter-Mountain Basins Curl-leaf Mountain-mahogany Woodland and Shrubland	CES304.772	8
2106	Northern Rocky Mountain Montane-Foothill Deciduous Shrubland	CES306.994	7
2150	Western Great Plains Tallgrass Prairie	CES303.673	6
2086	Rocky Mountain Lower Montane-Foothill Shrubland	CES306.822	6
2518	North-Central Interior Wet Flatwoods	CES202.700	5
2401	Central Interior Highlands Calcareous Glade and Barrens	CES202.691	4
2517	Paleozoic Plateau Bluff and Talus	CES202.704	4
2132	Central Mixedgrass Prairie	CES303.659	4
2135	Inter-Mountain Basins Semi-Desert Grassland	CES304.787	4
2117	Southern Rocky Mountain Ponderosa Pine Savanna	CES306.649	4
2140	Northern Rocky Mountain Subalpine-Upper Montane Grassland	CES306.806	4
2363	Central Interior Highlands Dry Acidic Glade and Barrens	CES202.692	3
2421	Central Tallgrass Prairie	CES205.683	3
2009	Northwestern Great Plains Aspen Forest and Parkland	CES303.681	3
2081	Inter-Mountain Basins Mixed Salt Desert Scrub	CES304.784	3
2057	Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland	CES306.819	3
2056	Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodland	CES306.830	3
2394	North-Central Interior Oak Savanna	CES202.698	2
2147	Western Great Plains Foothill and Piedmont Grassland	CES303.817	2
2365	Boreal White Spruce-Fir-Hardwood Forest	CES103.021	1

A total of 5 types (or 9% of 55 types) had >80% agreement between expert and auto-key assignments. All of these types had >10 sample plots. Expert self-assessment of confidence for these types was predominantly 'moderate' to 'high' although plots assigned to Rocky Mountain Lodgepole Pine Forest had a nearly evenly distributed confidence rating between 'low' to 'high'. This was because lodgepole pine can be a component of several other montane types, particularly in early successional stages, so the abundance of lodgepole pine was not always a clear determinant of the type.

Table 3 provides a summary of adequately-sampled types where agreement between expert and auto-key ranged from just below 80% down to zero. These types total 23, or 41% of the total types assigned. Analysis of the nature of the disagreements between expert and auto-key attribution reveals some of the sources of these disagreements and suggests some methods to help reduce these in the future. The following are some specific examples of levels of disagreement and possible explanations based on interpretations from the contingency table.

Northwestern Great Plains Mixedgrass Prairie – 6 sites (12%) were confused with floristically similar shrub or shrub-steppe types, Inter-Mountain Basins Big Sagebrush Shrubland or Inter-Mountain Basins Big Sagebrush Steppe. These types can grade into each other with the cover of shrubs determining which type fits best. Careful determination of shrub cover is important for consistent assignment of sites among these types.

Middle Rocky Mountain Montane Douglas-fir Forest and Woodland – This type was confused with three others with Inter-Mountain Basins Big Sagebrush Steppe being the most common (13% of the time). This is likely due to different interpretations of whether there was enough tree cover to fit the sites into a forest/woodland type versus a shrub type.

North-Central Interior Dry-Mesic Oak Forest and Woodland – This type was confused with two other upland forests, North-Central Interior Maple-Basswood Forest (8%) and North-Central Interior Dry Oak Forest and Woodland (2%), and two savanna/open woodland types, Eastern Great Plains Tallgrass Aspen Parkland (2%) and North-Central Interior Oak Savanna (2%).

North-Central Interior Maple-Basswood Forest – Seven plots (14%) percent of the sites auto-keyed to this type were assigned to the “Can’t assign” by the expert reviewer indicating not enough data to assign the type or that the site did not fit any natural type. The auto-key may have been too aggressive in assigning sites based on limited data and may have included non-natural sites.

Western Great Plains Dry Bur Oak Forest and Woodland – This type was confused with Western Great Plains Wooded Draw and Ravine most commonly. These two types often occur in similar environmental settings in the Great Plains. Dominance by bur oak is a good characteristic for this type but where there is a mix of species such as American basswood, green ash, bur oak, and elm then classification is difficult.

Eastern Great Plains Tallgrass Aspen Parkland – This type was not confused with any other natural type but 24% of the sites were not assigned by the expert reviewer. The dominant tree species in this type – quaking aspen – is also a common early successional species in old fields, logged, or burned areas so the dominance of just that species is not enough to assign a site to this type. Ground layer data and aerial photos are helpful in distinguishing disturbed sites from natural sites that fit this type.

Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland – Nearly all the disagreements between this type and others was in the assignment of sites auto-keyed to this type into the Rocky Mountain Lodgepole Pine Forest. Thirty-eight percent of the auto-keyed plots were assigned by the expert to the Rocky Mountain Lodgepole Pine Forest. Lodgepole pine can be a strong component of early successional stages of this Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland. In this case, sites with some subalpine fir and/or Englemann spruce were assigned to this type even if lodgepole pine was dominant but the auto-key placed those plots in the Rocky Mountain Lodgepole Pine Forest.

Inter-Mountain Basins Big Sagebrush Shrubland – Every site auto-keyed to this type and expert-assigned to another was to the Inter-Mountain Basins Big Sagebrush Steppe. The primary difference between these two is the cover of shrubs. A difference in interpretation of this cover resulted in this discrepancy in this case.

Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest – This type was confused primarily with two other conifer forests in the area. This type is typically dominated by Ponderosa pine, Douglas fir, or a mix of the two. It was confused with the more pure ponderosa pine type of the lower elevations (Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna) and the more pure Douglas fir type also found nearby at low elevations (Middle Rocky Mountain Montane Douglas-fir Forest and Woodland). The interface of these types is poorly differentiated unless prairie understory is present.

Inter-Mountain Basins Montane Sagebrush Steppe – This type was confused almost entirely (54%) with the Inter-Mountain Basins Big Sagebrush Steppe. Careful identification of the sub-species of *Artemisia tridentata* is important to differentiate these types. In some cases, the auto-key would have been keying on other floristic components (others shrubs, or the bunch grasses) suggestive of the more montane system, but the expert might have keyed based more on elevation.

Northwestern Great Plains Shrubland – This type was confused with Northwestern Great Plains Mixedgrass Prairie most often (25%). Since the associated species and geographic ranges overlap extensively, this confusion is based on different interpretations of the cover of shrubs. If a site has >25% shrub cover it will fit this type best but <25% shrub cover it will likely fit the mixedgrass prairie type best.

Western Great Plains Sandhill Steppe – This type was confused with four floristically related grassland types – Central Mixedgrass Prairie, Northwestern Great Plains Mixedgrass Prairie, Western Great Plains Sand Prairie, and Western Great Plains Shortgrass Prairie. These disagreements in assignments were due to differences in interpretation of the cover of shrubs. The auto-key calculated a higher cover for these sites and placed them in a shrubland type whereas the expert reviewer thought they fit a grassland type.

Western Great Plains Sand Prairie – Most of the disagreements (38%) with assignments of sites auto-keyed to this type were with assignments made by the expert reviewer to the Northwestern Great Plains Mixedgrass Prairie. Some of the species common to the Western Great Plains Sand Prairie can increase under grazing of the Northwestern Great Plains Mixedgrass Prairie. The expert reviewer judged some of these sites to fit the Northewestern Great Plains Mixedgrass Prairie best even though they had a relatively high cover of junegrass, *Sporobolus* spp., etc.

Western Great Plains Wooded Draw and Ravine – This type was confused with multiple other natural types and 6 plots (12%) were expert labeled with “Can’t assign” categories indicating not enough data to assign the type or that the site did not fit any natural type. The auto-key may have been too aggressive in assigning sites based on limited data and may have included non-natural sites.

Boreal Aspen-Birch Forest – This type was not assigned by the expert reviewer to any sites. It is not supposed to occur in GeoArea 5 and should be removed with a better geographic delineation of its distribution in the auto-key. All sites auto-keyed to this type were assigned by the expert reviewer to the Eastern Great Plains Tallgrass Aspen Parkland.

Table 3. Summary of types with adequate samples where agreement between auto-key and expert was below 80%.

				Plots with Expert Matches				
EVT Code	EVT Name	System Elcode	Total Plots	Total	%	High conf	Med conf	Low conf
2141	Northwestern Great Plains Mixedgrass Prairie	CES303.674	50	37	74%	15	13	9
2166	Middle Rocky Mountain Montane Douglas-fir Forest and Woodland	CES306.959	23	17	74%	6	9	2
2310	North-Central Interior Dry-Mesic Oak Forest and Woodland	CES202.046	50	36	72%	13	19	4
2314	North-Central Interior Maple-Basswood Forest	CES202.696	50	31	62%	22	6	3
2013	Western Great Plains Dry Bur Oak Forest and Woodland	CES303.667	39	23	59%	9	11	3
2331	Eastern Great Plains Tallgrass Aspen Parkland	CES205.688	23	13	57%	1	6	6
2055	Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	CES306.828	50	28	56%	10	13	5
2080	Inter-Mountain Basins Big Sagebrush Shrubland	CES304.777	46	25	54%	14	10	1
2045	Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest	CES306.805	50	27	54%	15	7	5
2126	Inter-Mountain Basins Montane Sagebrush Steppe	CES304.785	50	20	40%	10	6	4
2085	Northwestern Great Plains Shrubland	CES303.662	28	10	36%	6	3	1
2094	Western Great Plains Sandhill Steppe	CES303.671	15	5	33%	1	4	0
2148	Western Great Plains Sand Prairie	CES303.670	49	16	33%	10	3	3
2153	Inter-Mountain Basins Greasewood Flat	CES304.780	28	8	29%	1	4	3
2385	Western Great Plains Wooded Draw and Ravine	CES303.680	50	12	24%	6	5	1
2054	Southern Rocky Mountain Ponderosa Pine Woodland	CES306.648	50	10	20%	4	5	1
2139	Northern Rocky Mountain Lower Montane-Foothill-Valley Grassland	CES306.040	23	4	17%	1	3	0
2167	Rocky Mountain Poor-Site Lodgepole Pine Forest	CES306.960	10	1	10%	0	1	0
2145	Rocky Mountain Subalpine-Montane Mesic Meadow	CES306.829	13	1	8%	0	1	0

				Plots with Expert Matches				
EVT Code	**EVT Name**	**System Elcode**	**Total Plots**	**Total**	**%**	**High conf**	**Med conf**	**Low conf**
2061	Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland	CES304.776	14	1	7%	0	1	0
2301	Boreal Aspen-Birch Forest	CES103.020	19	0	0%	0	0	0
2165	Northern Rocky Mountain Foothill Conifer Wooded Steppe	CES306.958	13	0	0%	0	0	0
2304	Ozark-Ouachita Dry-Mesic Oak Forest	CES202.708	11	0	0%	0	0	0

Expert Assignments

2.2 A report by each GeoArea detailing processes and results, specifically identifying how they made individual assignments.

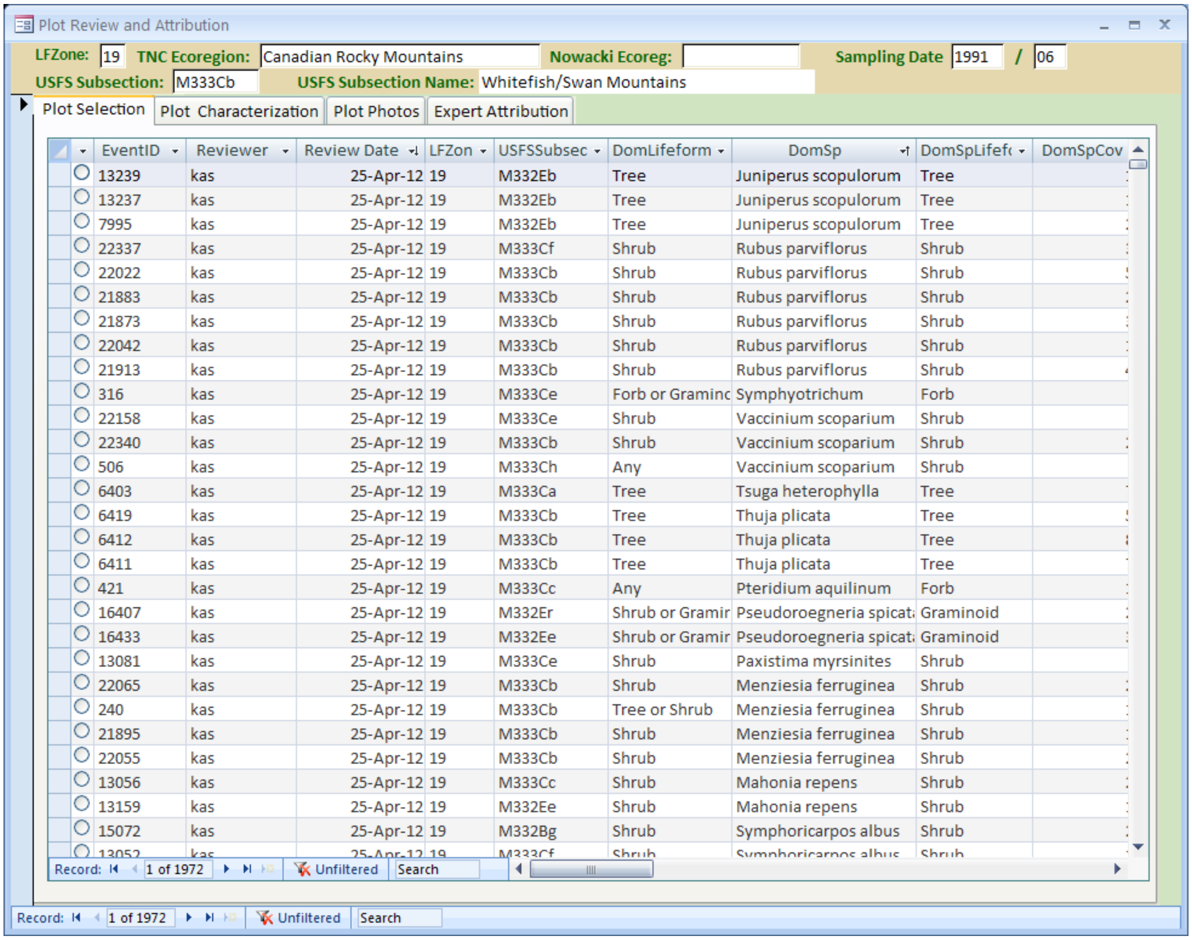
As described in the methods section above, the expert reviewers worked directly in the expert attribution database (EADB). Since GeoArea 5 had almost 1,500 plots to review, a systematic, efficient process for reviewing and labeling plots was required. The forms provided in the EADB allowed the reviewer to sort and filter on subsets of plots to select groups of them with similar characteristics. For instance, the reviewer could select all plots found within a particular USFS Section or MapZone, then select all plots dominated by trees, then sort alphabetically by the dominant species. The reviewer could also select all treed plots, then select all plots with the same dominant tree species (such as *Pinus edulis*), then sort by % cover of that species, from high to low. Figure 4 shows the main form in the EADB which has these data fields. Additional fields were provided from which to select or sort plots, such as elevation, aspect, slope, and total cover by lifeform in the plot.

Once the reviewer had selected a subset of plots for reviewing, the next step was to select an individual plot to review and label. If the expert was working on treed plots first, then they had a further option of selecting the set of ecological systems from which to pick a label for the plots. This was accomplished via a filter on the NLCD land cover class applied to all systems (such as forest and woodland, shrubland, herbaceous, woody wetlands, and so on).

For each plot, the expert reviewed environmental and geographic setting, as well as the floristic and vegetation structural characteristics of the plot. In many cases the expert could then assign an ecological system label with no further information. However, in some cases the reviewer might consult the descriptions for a group of similar ecological systems to clarify their understanding of differences in concept, geographic distribution, floristics, or structural characteristics.

For example, in the lower montane areas of central Montana, forests dominated by ponderosa pine, Douglas fir, or a combination can occur. The reviewer would need to determine if those plots represented Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna, Northern Rocky Mountain Ponderosa Pine Woodland and Savanna, Middle Rocky Mountain Montane Douglas-fir Forest and Woodland, or Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest.

Figure 4. Screen shot of EADB form, showing some of the data the expert reviewer could select from or sort on to efficiently review similar plots.



Plot Review and Attribution

LFZone: 19 TNC Ecoregion: Canadian Rocky Mountains Nowacki Ecoreg: Sampling Date 1991 / 06

USFS Subsection: M333Cb USFS Subsection Name: Whitefish/Swan Mountains

Plot Selection | Plot Characterization | Plot Photos | Expert Attribution

EventID	Reviewer	Review Date	LFZon	USFSSubsec	DomLifeform	DomSp	DomSpLifef	DomSpCov
13239	kas	25-Apr-12	19	M332Eb	Tree	Juniperus scopulorum	Tree	
13237	kas	25-Apr-12	19	M332Eb	Tree	Juniperus scopulorum	Tree	
7995	kas	25-Apr-12	19	M332Eb	Tree	Juniperus scopulorum	Tree	
22337	kas	25-Apr-12	19	M333Cf	Shrub	Rubus parviflorus	Shrub	
22022	kas	25-Apr-12	19	M333Cb	Shrub	Rubus parviflorus	Shrub	
21883	kas	25-Apr-12	19	M333Cb	Shrub	Rubus parviflorus	Shrub	
21873	kas	25-Apr-12	19	M333Cb	Shrub	Rubus parviflorus	Shrub	
22042	kas	25-Apr-12	19	M333Cb	Shrub	Rubus parviflorus	Shrub	
21913	kas	25-Apr-12	19	M333Cb	Shrub	Rubus parviflorus	Shrub	
316	kas	25-Apr-12	19	M333Ce	Forb or Graminc	Symphyotrichum	Forb	
22158	kas	25-Apr-12	19	M333Ce	Shrub	Vaccinium scoparium	Shrub	
22340	kas	25-Apr-12	19	M333Cb	Shrub	Vaccinium scoparium	Shrub	
506	kas	25-Apr-12	19	M333Ch	Any	Vaccinium scoparium	Shrub	
6403	kas	25-Apr-12	19	M333Ca	Tree	Tsuga heterophylla	Tree	
6419	kas	25-Apr-12	19	M333Cb	Tree	Thuja plicata	Tree	
6412	kas	25-Apr-12	19	M333Cb	Tree	Thuja plicata	Tree	
6411	kas	25-Apr-12	19	M333Cb	Tree	Thuja plicata	Tree	
421	kas	25-Apr-12	19	M333Cc	Any	Pteridium aquilinum	Forb	
16407	kas	25-Apr-12	19	M332Er	Shrub or Gramir	Pseudoroegneria spicat	Graminoid	
16433	kas	25-Apr-12	19	M332Ee	Shrub or Gramir	Pseudoroegneria spicat	Graminoid	
13081	kas	25-Apr-12	19	M333Ce	Shrub	Paxistima myrsinites	Shrub	
22065	kas	25-Apr-12	19	M333Cb	Shrub	Menziesia ferruginea	Shrub	
240	kas	25-Apr-12	19	M333Cb	Tree or Shrub	Menziesia ferruginea	Shrub	
21895	kas	25-Apr-12	19	M333Cb	Shrub	Menziesia ferruginea	Shrub	
22055	kas	25-Apr-12	19	M333Cb	Shrub	Menziesia ferruginea	Shrub	
13056	kas	25-Apr-12	19	M333Cc	Shrub	Mahonia repens	Shrub	
13159	kas	25-Apr-12	19	M332Ee	Shrub	Mahonia repens	Shrub	
15072	kas	25-Apr-12	19	M332Bg	Shrub	Symphoricarpos albus	Shrub	
13052	kas	25-Apr-12	19	M333Cf	Shrub	Symphoricarpos albus	Shrub	

Record: 1 of 1972 Unfiltered Search

Record: 1 of 1972 Unfiltered Search

In cases like this, the determination of which system type to assign to the plot might require:

a) review of the image clip for the context of the plot,
b) review of where the plot was located geographically (USFS Subsections provide local scale geographic location), may help distinguish the Great Plains-Black Hills type from the Rocky Mountain types,
c) consideration of topographic setting (e.g. north-facing slopes at lower elevations could support ponderosa pine woodlands),
d) consideration of any available height data for the plot (e.g. were the ponderosa pines all tall, apparently mature trees; or were they short),
e) careful consideration of the full floristic composition of the plot and cover for each species.
f) awareness of possible errors in the plot data, such as misidentification of juniper species by the field crews, unevenness in how the cover values were estimated in the field or converted into the LFRDB (e.g. cover for trees estimated by a person standing on the ground vs. an aerial view of the plot).

Below are some examples of comments relevant to the above example:

- Tree cover almost too low for a treed System but photo and Source data appear to show enough tree cover so site not a grassland.

- Little P. ponderosa but this site is too low and within a Great Plains grassland landscape so it can't be Rocky Mountain Foothills Limber Pine-Juniper System.
- There are several dry-mesic forest Systems with mixed conifer canopies possible in this area but this seems the best fit.
- Probably enough Pseudotsuga menziesii to fit this System better than the Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer System.
- Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest can also be dominated by P. menziesii but since there is little besides that species in the canopy it seems to fit the assigned System best.
- Moderately low cover of Pseudotsuga and high cover of Juniperus but still seems to fit this System fairly well. Dry site.

Given all of the above, the reviewer had to make a decision for the plot, and assign an ecological system label. In cases where the assignment was not made with high confidence, the reviewer was requested to provide comments as to the factors they used to assign a label to the plot, or what the alternative assignment could be. Report Section 2.3 below discusses some of the results pertinent to confidence of assignment.

Improving the auto-key process

2.3 A report that documents procedures and data elements to improve the auto-key process in each GeoArea.

Of the 66 types assigned to plots by experts, 35 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 31 types, the numbers of samples labeled to a given type ranged from 138 (Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna) down to 10 (Southern Rocky Mountain Ponderosa Pine Woodland). For all (100%) of these types, experts reported moderate or high confidence in their labels for at least 20% of the type's plots. 15 types indicated low confidence for at least 20% of the type's plots. These statistics are listed in the Appendix. A small sampling of expert comments related to moderate or low confidence plots are included in Table 4.

Table 4. A selection of expert comments related to labeling sample plots for types where their confidence was reported as moderate or low.

Type Name	Expert Comment
Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna	Little P. ponderosa but this site is too low and within a Great Plains
Northwestern Great Plains Mixedgrass Prairie	Lots of exotic cover but the few native grasses indicate this System.
North-Central Interior Floodplain	Floristics and photo look like a floodplain but very little floristic data to go by.
Inter-Mountain Basins Big Sagebrush Steppe	Borderline herbaceous cover for this System but fits here best.
Rocky Mountain Lodgepole Pine Forest	Low cover in all three strata but fits a treed System best.
North-Central Interior Dry-Mesic Oak Forest and Woodland	Quercus ellipsoidalis seems odd in this mix but, if present, pushes this to an upland System.

These and other comments point to several important aspects for consideration. First, some ecological systems concepts are better known and understood than others. Therefore, a certain degree of classification refinement is likely needed in order to improve auto-keys. In this GeoArea, this is particularly true in the transition from lower montane/foothill types to Great Plains types. As currently defined there is no clear differentiation between these types. Second, the inclusion of some limited landform, soil, and or landscape context information could assist with some determinations within the key, or by a subsequent expert reviewer. Similarly, repeated references to photos further indicates the need for expert review of many types where moderate-low confidence of experts suggest that auto-keys might be prone to error. Third, additional floristic information is cited in some cases where their suspected limitations provide the primary source of expert uncertainty in labeling.

Other samples were those labeled by auto-keys to aggregates of multiple ecological system types. This was because LANDFIRE had mapping objectives focused on uplands where fire regimes are prevalent. That meant that many individual wetland and sparsely-vegetated ecological system types were not treated within the auto-keys. Expert labeling of these samples, however, provides an indication of the feasibility of their inclusion in updated auto-keys. Of 279 samples, experts were able to assign 237 (85%) to an individual ecological system type; a total of 32 individual ecological system types were assigned to these samples. This result indicates the potential for inclusion of these types within subsequent mapping efforts. We cannot yet comment on the issues associated including these types within future regional auto-keys, but this appears to be an issue worthy of exploration.

Another class of samples did not contain enough information for the auto-keys to assign a system or system aggregate; these samples were labeled with broad "unclassified" types, such as "Unclassified Grassland" or "None". Of 133 samples, experts were able to assign 94 (71%) to an individual ecological system type; a total of 25 individual ecological system types were assigned to these samples.

Adapting auto-keys for NVC Groups, Macrogroups, and Divisions

2.4 A report that documents technical procedures to adapt auto-keys for labeling NVCS group, macrogroup, and division concepts.

US-NVC Groups

In an effort to understand the potential implications of adapting LANDFIRE autokeys for use with the revised US-NVC, we first compared the mapped ecological system types within this GeoArea to their related US-NVC Group concepts. These two classification concepts, with the NVC designed solely using existing vegetation, and ecological systems combining existing vegetation and biophysical factors, are most closely related at the Group level of the revised US-NVC hierarchy. Since these two classifications have been thoroughly related to each other, these relationships should provide insight for the task of updating autokeys for use with the NVC.

Within this GeoArea, some 104 terrestrial ecological system types could occur. Of these, 43 have a practical 1:1 relationship with NVC Group concepts, and 53 nest cleanly within 34 NVC Group concepts (1:many group:system relationship), for a total of 96 or 92% of ecological system concepts with a clean relationship to an NVC Group. There is some potential for slight differences among floristic elements among these NVC Groups relative to ecological systems. For example, one or more associations linked to a given terrestrial ecological system type may now be linked to a different NVC Group concept. There is some limited potential that the floristic information found within the autokey would need to be revisited to account for this, but within this GeoArea, we believe that this instance is quite limited.

Where the relationship between ecological systems and NVC Groups is more complex, there is potential need for substantive changes to existing autokeys. Within this GeoArea, 6 (6%) ecological system types have a more complex relationship with NVC Group concepts (Table 5). The number of system-Group inter-relationships is relatively small and the relationships are easily handled by the existing auto-keys or small alterations to them. Here we provide additional commentary on the implications for autokey adjustment brought by these types.

Table 5. Ecological Systems of GeoArea 5 that have complex relationships with NVC Groups. Interrelated Systems and Groups are shown in the heavy-outline boxes. The number of NVC Groups each system is related to is shown in the Groups column, and the number of Ecological Systems each NVC Group is related to is shown in the Systems column.

Ecological System	NVC Group	Groups	Systems
Inter-Mountain Basins Big Sagebrush Shrubland	G303 Intermountain Dry Tall Sagebrush Shrubland	2	3
	G302 Intermountain Mesic Tall Sagebrush Shrubland & Steppe	2	3
Inter-Mountain Basins Big Sagebrush Steppe	G303 Intermountain Dry Tall Sagebrush Shrubland	2	3
	G302 Intermountain Mesic Tall Sagebrush Shrubland & Steppe	2	3
Northern Rocky Mountain Lower Montane Riparian Woodland and Shrubland	G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland	2	5
	G503 Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest	2	4
Rocky Mountain Lower Montane-Foothill Riparian Woodland and Shrubland	G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland	2	5
	G503 Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest	2	4
Northern Rocky Mountain Montane-Foothill Deciduous Shrubland	G272 Northern Rocky Mountain Montane-Foothill Dry Deciduous Shrubland	2	1
	G275 Northern Rocky Mountain Montane-Foothill Mesic Deciduous Shrubland	2	2
Rocky Mountain Alpine-Montane Wet Meadow	G521 Vancouverian & Rocky Mountain Montane Wet Meadow	2	2
	G520 Vancouverian & Rocky Mountain Subalpine Snowbed, Wet Meadow & Dwarf-Shrubland	2	1

Inter-Mountain Basins Big Sagebrush Shrubland - This System is related to two Groups. Differentiating between these Groups, G303 Intermountain Dry Tall Sagebrush Shrubland and G302 Intermountain Mesic Tall Sagebrush Shrubland & Steppe, should be relatively easy to accomplish based on associated species differences (dry vs. mesic species). These species may need to be parsed out in the existing auto-keys or characteristic species may need to be added.

Intermountain Basins Big Sagebrush Steppe – As with the previous System, this System is related to two Groups. Differentiating between these Groups, G303 Intermountain Dry Tall Sagebrush Shrubland and

G302 Intermountain Mesic Tall Sagebrush Shrubland & Steppe, should be relatively easy to accomplish based on associated species differences (dry vs. mesic species). These species may need to be parsed out in the existing auto-keys or characteristic species may need to be added.

Northern Rocky Mountain Lower Montane Riparian Woodland and Shrubland – This System is related to two Groups, G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland and G503 Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest. It should be possible to differentiate between these two Groups based on lifeform (shrubland vs. forest), which is already in the auto-key.

Rocky Mountain Lower Montane-Foothill Riparian Woodland and Shrubland – As with the previous System, this System is related to two Groups, G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland and G503 Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest. It should be possible to differentiate between these two Groups based on lifeform (shrubland vs. forest), which is already in the auto-key.

Northern Rocky Mountain Montane-Foothill Deciduous Shrubland – This System is related to two Groups, themselves related along a moisture gradient, G272 Northern Rocky Mountain Montane-Foothill Dry Deciduous Shrubland and G275 Northern Rocky Mountain Montane-Foothill Mesic Deciduous Shrubland. It should be possible to differentiate these Groups based on component species (dry vs. mesic).

Rocky Mountain Alpine-Montane Wet Meadow – This System is related to two Groups, G521 Vancouverian & Rocky Mountain Montane Wet Meadow and G520 Vancouverian & Rocky Mountain Subalpine Snowbed, Wet Meadow & Dwarf-Shrubland. These Groups are nominally distinguished by elevation though that may not always hold steady across the range of the Groups. It will have to be seen whether there are characteristic species that would reliably distinguish these two Groups.

US-NVC Macrogroups

Ecological Systems can be fairly comfortably rolled up to broader US-NVC Macrogroups, which cover the existing-vegetation component of their related ecological systems. Using LANDFIRE autokeys for US-NVC Macrogroups instead of ecological systems could potentially resolve disagreements between experts and autokeys found at the ecological systems level. To evaluate the potential effect of using the autokey for Macrogroups, we arranged the ecological system types by US-NVC Macrogroup in the expert-autokey contingency table, and also compared the percent of expert-autokey matches at the system level versus the Macrogroup level (Table 6).

There are 24 US-NVC Macrogroups represented among natural mapped classes in this GeoArea. While the results in Table 6 suggest rolling up to Macrogroup would yield improved results, in some areas, consideration must be given to the fact that many of these Macrogroups are in fact very broad concepts, and include ecologically diverse system types. Combining System results yields a small increase (<10%) in agreement for 20 of the 24 Macrogroups so the gain for these is likely not worth the loss in detail. Three of the remaining four Macrogroups that would see a significant gain are the among the most common in the dataset (and arguably on the landscape) however the loss in detail would be substantial. For example, rolling up the seven systems within M020 Rocky Mountain Subalpine & High Montane Conifer Forest would increase the agreement from 60% to 84%. However, this Macrogroup includes a wide variety of habitats from thin-soiled, harsh environments that support open lodgepole pine forests to aspen forests to mesic/wet-mesic spruce-fir forests. The reduction in ecological detail would be great and it would be wise to determine if sufficient gains in agreement could be made through revising the auto-keys, clearing up confusing classification concepts, and possibly using more thoroughly collected data to assign sites.

Table 6. Comparison of auto-keyed results when plots keyed to systems are rolled up to Macrogroups, showing percent of matches at the system level compared to Macrogroup level.

Macrogroup	# auto-keyed systems	# plots	% expert matches at system level	% expert matches at MG level
M093 Great Basin Saltbrush Scrub	2	12	83%	83%
M169 Great Basin & Intermountain Tall Sagebrush Shrubland & Steppe	3	105	50%	98%
M171 Great Basin & Intermountain Dry Shrubland & Grassland	1	4	0%	0%
M012 Central Oak-Hardwood & Pine Forest	2	58	71%	78%
M014 Acer saccharum - Betula alleghaniensis - Pinus strobus - Tsuga canadensis Forest	1	4	100%	100%
M016 Southern Hardwood & Pine Forest	2	20	40%	40%
M151 Northern Great Plains Woodland	5	143	43%	52%
M153 Acer (barbatum, saccharum) - Tilia americana - Fagus grandifolia - (Liriodendron tulipifera) Forest	1	50	62%	62%
M017 Northern Rocky Mountain Lower Montane & Foothill Forest	6	160	69%	92%
M020 Rocky Mountain Subalpine & High Montane Conifer Forest	7	136	60%	84%
M022 Southern Rocky Mountain Lower Montane Forest	3	63	25%	25%
M026 Intermountain Singleleaf Pinyon - Western Juniper Woodland	1	8	100%	100%
M030 Northern & Central Swamp Forest	1	5	20%	20%
M037 Eastern & Central North American Boreal Conifer & Hardwood Forest	2	20	0%	0%
M116 Great Plains Cliff, Scree & Rock Vegetation	1	8	0%	0%
M048 Northern Rocky Mountain Montane & Foothill Grassland & Shrubland	3	34	18%	47%
M049 Southern Rocky Mountain Montane Grassland & Shrubland	1	6	67%	67%
M168 Rocky Mountain-Vancouverian Subalpine & High Montane Mesic Grass & Forb Meadow	1	13	8%	8%
M051 Great Plains Mixedgrass Prairie & Shrubland	4	62	61%	71%
M052 Great Plains Sand Grassland & Shrubland	2	64	33%	39%
M053 Great Plains Shortgrass Prairie & Shrubland	1	29	93%	93%
M054 Great Plains Tallgrass Prairie, Savanna & Shrubland	2	5	80%	80%
M124 Northern & Central Alvar & Glade	2	7	0%	0%

M082 Cool Semi-Desert Alkali-Saline Wetland	1	28	29%	29%

US-NVC Divisions

NVC Divisions are substantially simplified vegetation concepts relative to terrestrial ecological system types, so autokeys designed for these concepts would be relatively simple to develop. For within this GeoArea, we would recommend starting from a new baseline starting point in order to adequately design one autokey to encompass the 11 natural US-NVC Division concepts that occur here.

Discussion

The LANDFIRE reference database is the first attempt by a single agency to compile comprehensive georeferenced vegetation data for the United States. As such it is a powerful tool for use in many different applications, but there are caveats that must be clearly understood by the user(s) of the data and the results. Sequence tables are an innovative method for rapidly and efficiently keying thousands of vegetation samples; for LANDFIRE they were developed to key to ecological systems and land cover classes, but could be modified to key to any floristically-based vegetation types, such as the Group level of the NVC hierarchy.

Fundamentally, a sequence table as used by LANDFIRE is a set of criteria. Each vegetation sample has to meet some combination of criteria in the SQT to be labeled with an ecological system, or some other land cover class. Simply put, if the plot doesn't meet any criteria contained in the sequence table, then it may be mis-keyed, or not key to anything. Given our incomplete knowledege of the structural and floristic variability of each classification unit, it is nearly impossible to establish criteria in a sequence table - for regional application - to successfully and accurately key 100% of vegetation samples. However, with new field-based inventory and increasing ecological understanding, over time sequence tables can be revised and improved so as to accurately key increasing percentages of vegetation samples.

There are a number of reasons why a sequence table may not successfully key all samples run through it:
a) the unknown floristic quality of the vegetation data (how complete, how well collected, does it accurately represent the vegetation concept being keyed);
b) our limited knowledge of the variability in species composition, vegetation structure, and the distribution of ecological systems; and
c) the comprehensiveness (or lack thereof) in field inventory for any particular system (e.g., many from one small area, few to none from elsewhere in the region).

Each of these are discussed below.

A. Quality of vegetation data

First and foremost, the completeness and quality of the data as collected in the field, as well as the documentation of how the data were collected (the metadata) are primary issues for how well the sequence table process works. There are many different kinds of issues with the data collection, only a few of which are listed here as possible sources of problems:

- Was the species composition adequately sampled (complete species list)?
- Were only trees recorded (e.g., some FIA plots)? Only "dominant" or "most characteristic" species (e.g., SWReGAP training data)?
- Was the sample plotless, or within a plot or some other measured area?
- Or were the samples derived along transects?
- How was the cover or abundance data collected, or was it presence/absence?
- Was the sample area across an ecotone (for example across the transition from a wet valley bottom into the adjacent upland slope)?
- Does the sample adequately represent an occurrence of the vegetation type being sampled?
- Was the species taxonomy accurately recorded (many species are difficult for untrained crews to identify, such as *Carex* spp., or *Salix* spp.)?
- Were difficult species "lumped up" into broader taxon, such as genus, or even family?
- Was the sample location heavily or recently disturbed?

Many datasets obtained by the LANDFIRE team had inadequate metadata associated with them. Inadequate documentation of the sampling design or of what the values in the data tables represented, could result in incorrect processing of the data for use in the sequence tables.

The sampling design under which vegetation data was collected is an often neglected piece of metadata. A particular dataset could have many hundreds of plots in it, but the purpose(s) for which they were collected could be such as to negate their value for identifying floristically distinct vegetation types. For example, samples collected in a systematic grid without regard for sampling distinct vegetation types will often cross multiple ecological systems, and hence result in data that give erroneous results in an auto-key process.

An example of poor documentation of the collection protocols would include species names collected and provided as 4- or 6-letter acronyms, without a complete list of what species each acronym represents. The processing of the data into the LFRDB converts acronyms to full species utilizing the current NRCS PLANTS 'symbols'. So, POTR could be *Populus tremuloides*, *Poa tracyi,* or *Poa trivialis*, all valid species. But using PLANTS, POTR = *Poa tracyi*, while *Populus tremuloides* is POTR5. Each dataset has to be reviewed for its species taxonomy to ensure any acronyms are converted to the correct taxa, but without adequate metadata errors can creep in.

Another example would be where the species abundance data were collected in generalized "cover classes", and these had to be converted to "real cover" by using the mid-point of the class. If the metadata did not include documentation of what the classes represent, then the mid-points could be incorrectly converted, or even unobtainable. For example, cover class 3 could mean 5-25% cover (mid-point of 15%), or it could mean 25-35% cover (mid-point 30%). The sequence table process utilizes cover criteria for indicator species extensively, so incorrectly interpreted cover classes will lead to errors in the results.

B. Constraints within sequence table

Ecological systems are classified using a multi-factorial approach, including environmental factors, ecological processes and vegetation structure and composition. However, the sequence table process as currently developed and used by LANDFIRE does not allow use of local-scale environmental factors which might assist with distinguishing among floristically similar ecological systems. For example, how would one use avalanche slopes in an automated plot keying process? Or high-gradient vs. low gradient stream flow-regime? These are diagnostic features of one or more ecological systems that facilitate ready recognition in the field, but if floristic information is limited there may be no way to identify individual plots that occur on these features.

The early versions of the auto-key only allowed use of vegetation structure and composition data. The most recent auto-key does allow the use of elevation data which is helpful in accurately labeling plots to ecological systems that can be readily distinguished by elevation zones. The auto-key allows use of regional-scale variables, such as occurrence in a TNC ecoregion, or a USFS Section. Beyond these 2 variables (elevation and general geographic location) the auto-key does not currently allow use of any other more local-scale environmental variables, such as aspect, slope, landforms, soils conditions, etc.

Over time, as our knowledge of the floristic composition and structure of vegetation in the United States becomes more complete, local-scale variables may not be needed. If the plot data themselves are complete (meaning the species composition has been adequately sampled and recorded for the plot) we can infer environmental setting and characteristic ecological dynamics through the use of indicator species. For example, *Heracleum maximum* to indicate mesic or wet understory conditions for wetland and riparian ecological systems or *Juncus drummondii* and *Caltha leptosepala* to indicate alpine wetland sites, or the predominance of *Festuca idahoensis* as a montane or subalpine grassland indicator. However, it's generally the combination of multiple species in varying abundance that are used in a sequence table to key plots; hence incomplete or poorly collected species compositional data generate poor results from the auto-keying process.

In comparison, dichotomous field keys to the ecological systems of a region do allow incorporation of the environmental or ecological "context" of a vegetation sample. In a field key, you can explicitly state "if you are in a marsh, then go to this part of the key…." or "if you are in the alpine, go here…", or "if this place is in the path of regular avalanches, go to this part of the key…". One of the LANDFIRE products is a set of dichotomous keys to be used in the field, for all ecological systems and land cover classes in groups of MRLC map zones.

C. Developing automated keys for large geographic areas

Each sequence table was constructed to work across relatively large geographic areas, on the order of 2-5 USFS Sections (**Error! Reference source not found.**). Hence each sequence table/auto-key included ens of ecological system types, and each system has some degree of compositional and structural variability across that region.

It's difficult to account for all compositional or structural variability that might occur in a single system type across a large geographic area. For example, western coniferous forests can vary from 25% tree cover to well over 90% cover, but in some patches may be less than 25%. Montane coniferous forests and woodlands on the Colorado Plateau are highly variable, with total tree cover ranging from 15% to >75%, with a diverse array of shrub associates, or sometimes no shrubs, and with little to no herbaceous component, or very high herbaceous cover. There are at least 4 different ecological systems for these montane forests; while the tree species are not particularly diverse, the possible shrub or herbaceous indicators are highly diverse. So, in this case the trees are not good indicators of the different ecological systems, and the shrubs are also only partially adequate. It is the herbaceous component that is particularly useful to key these systems, but when the plots are lacking in herbaceous data the task becomes much more difficult.

Another example is montane riparian shrublands of the southern Rocky Mountains, which are primarily placed into one ecological system. But to correctly key plots to the riparian system, the auto-key needs to account for every possible dominant shrub that might be found in a plot in the riparian zone (e.g., *Salix bebbiana, Salix geyeriana, Crataegus rivularis, Forestiera pubescens, Prunus virginiana, Rhus trilobata, Salix irrorata, Salix lucida, Shepherdia argentea, Betula occidentalis, Alnus incana, Salix exigua, Salix lasiolepis, Salix lutea, Salix ligulifolia,* etc.).

D. Cost/benefit & efficiency

The purpose of the auto-key process is to accurately key many hundreds of vegetation samples for each desired map class (ecological system or land cover) to feed into a mapping process. While a single georeferenced sample may be lacking in the complete floristics of an occurrence of an ecological system, the sequence table process aims to attribute many dozens to hundreds of plots to each ecological system or land cover class.

Auto-keys take a significant amount of time to develop for a region, and then to test, review, refine, and test again. A single auto-key for LANDFIRE typically took somewhere between 4 and 7 person days to create and refine. And, that assumes an agency such as SEM has already completed data compilation and processing for use. Some auto-keys for regions with large numbers of samples (for example map zones 1, 2, and 7 in the Pacific Northwest had over 100,000 plots) probably took closer to 10 person days to develop.

However, sequence tables can be refined over and over. The identification of combinations of species indicative of particular geographic or ecological settings is an ongoing effort amongst vegetation ecologists, and a repeatable and refine-able method such as this has distinct advantages. As we become more knowledgeable, field data becomes more comprehensive, and well collected datasets become more numerous, sequence tables can be improved until they successfully key 95% or more of the plots fed through them. This is a huge advantage for regional and national classification and mapping efforts, especially when it is desired to repeat them over some specified time frame with new imagery or new mapping methods.

Recommendations (draft)

This report section requires further development and interpretation; this is preliminary material. After other GeoAreas have been analyzed this section will be more completely written up. Recommendations may vary somewhat across the country, but we anticipate some general patterns relevant to all sequence tables and GeoAreas.

Adjustments to Auto-key procedures – inclusion of locational/biophysical information for pre-processing plots and/or inclusion of features in auto-keys

Narrowing vs. broadening the geographic application of the auto-key – FS Sections? In certain areas? Would this likely lead to greater accuracy?

Adjustment to auto-keys – additional requirements for vegetation sample data; primarily ground cover data

Expert review and labeling of plots for types of low confidence from auto-key.

Adjustments to Map Legends – moving to Group/Macrogroup concepts where systems level remains challenging – which ones?

Coping with uncertainty; what proportion of types could NOT be adequately handled through any of the above adjustments?

Careful review of the dominant tree, shrub, or grass elements shared among related types, should be the focus of auto-key improvements for these types.