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

FINAL REPORT

Prepared by
NatureServe
For the NPS Vegetation Inventory Program & LANDFIRE

30 June 2012

© SWReGAP. Eastern New Mexico.

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 Keith Schulz and Judy Teague, 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.

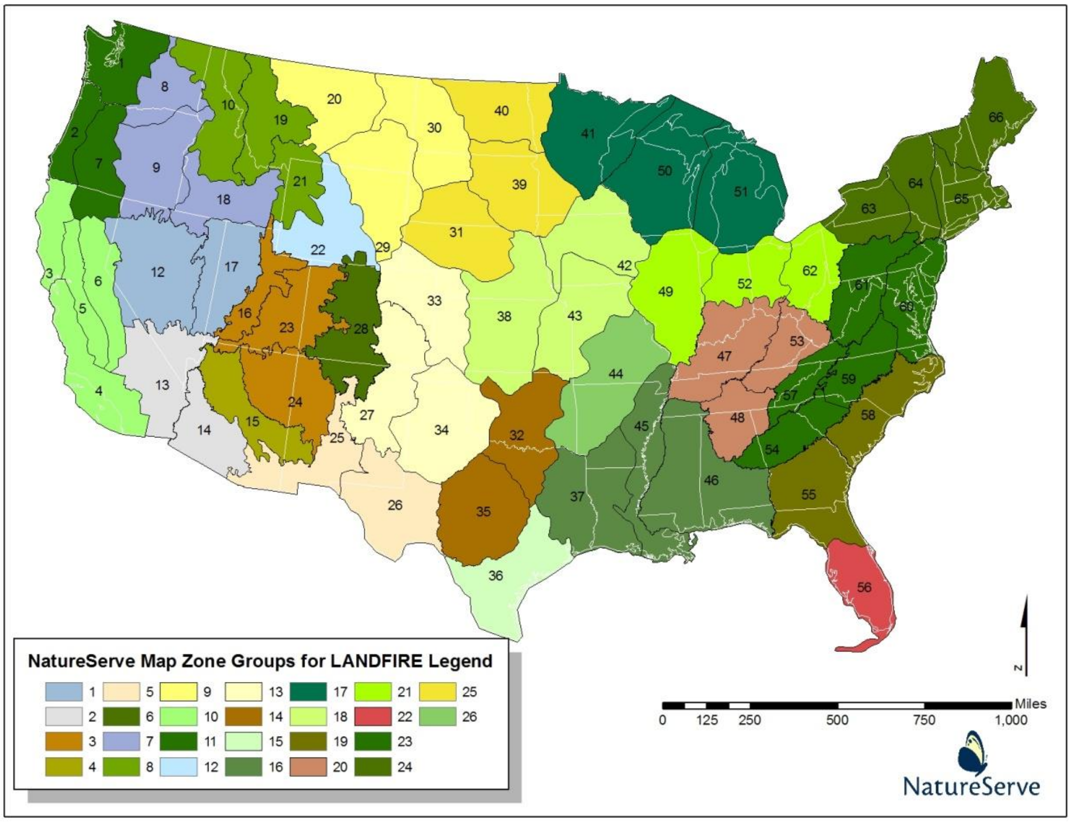
LANDFIRE's short-term needs, and long-term plans, required a repeatable methodology, consistently applied rules to categorize each reference sample, and documentation of the criteria applied. In essence, sequence tables codify the criteria and methods for keying geo-referenced vegetation data to a land 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.

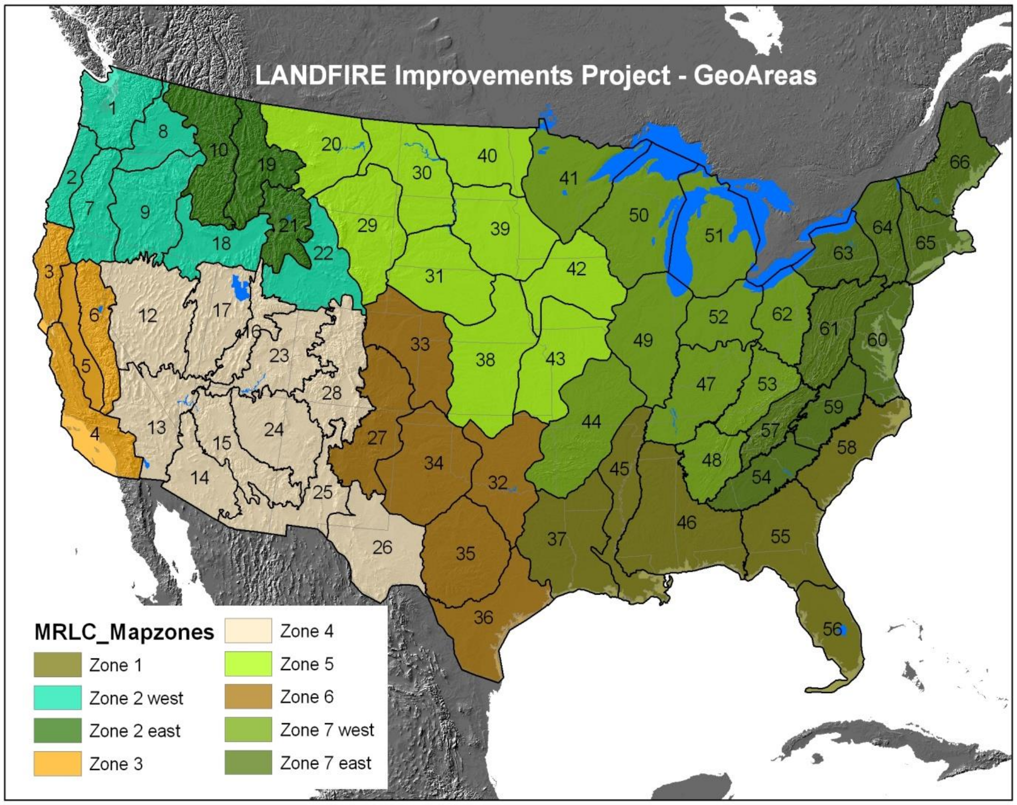
Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in the coterminous U.S.



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:

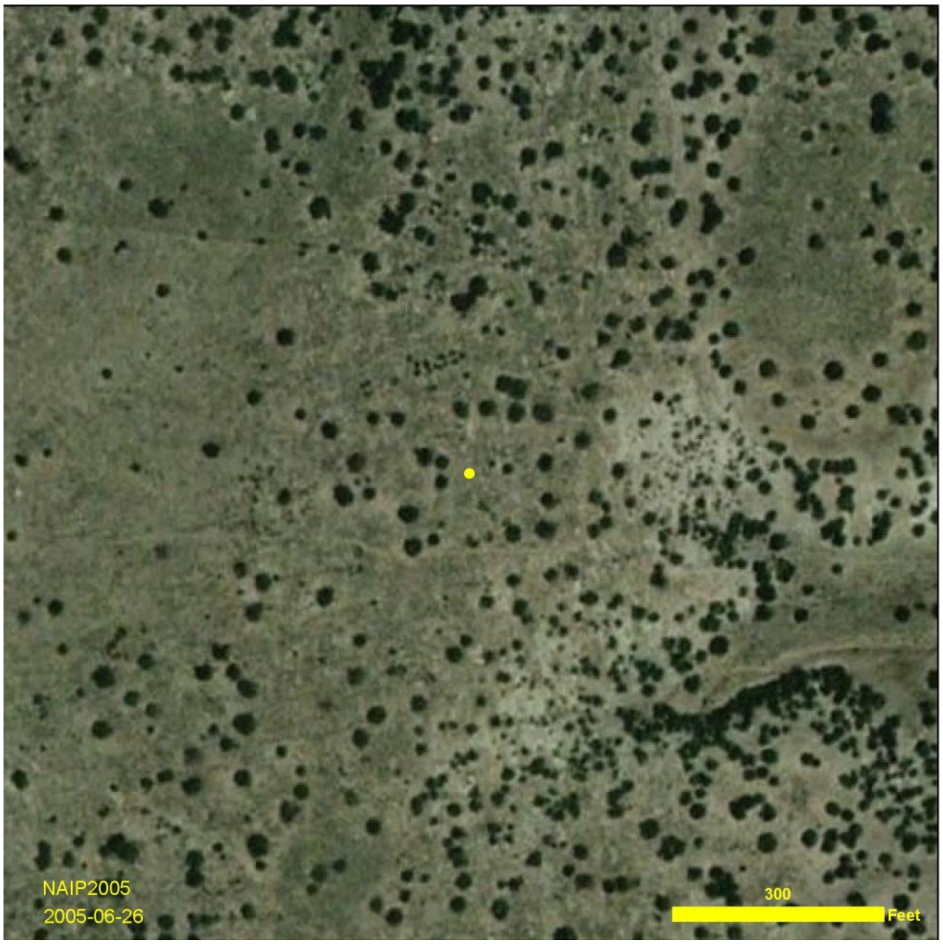
1. Providing the same data as are used in the auto-key procedures
2. Providing additional data that were not originally used in the auto-keys, and
3. Maintaining the "privacy" of the FIA data, ensuring the experts could not determine which plots were FIA vs not

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 6



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 6 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 6: Southern Plains

GeoArea 6 encompasses the Great Plains Tablelands, Southeastern Great Plains, Western Great Plains, Southern Great Plains, Edwards Plateau and Western Gulf Plains (Figure 2). This GeoArea includes a total of 6 map zones (27, 32, 33, 34, 35, and 36), originally clustered for purposes of designing and implementing auto-keys (Figure 1). The total number of plots in this Geo Area analysis was 1,295. A total of 52 natural ecological system types were assigned to a total of 930 plots by the auto-keys. A total of 54 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). Other "additional" systems are characteristic of peripheral areas such as Chihuahuan Desert, which is transitional to the Southern Plains and high Rocky Mountain forest systems which may occur on isolated mountains on the western edge of the GeoArea. These systems were added to the sequence tables because of the potential inclusion of plots from these areas.

An additional 22 types were assigned by the auto-key but were not assigned by experts:

- Chihuahuan Loamy Plains Desert Grassland
- Chihuahuan Sandy Plains Semi-Desert Grassland
- Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland
- Inter-Mountain Basins Montane Sagebrush Steppe
- Madrean Pinyon-Juniper Woodland
- Northwestern Great Plains Canyon
- Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna
- Ozark-Ouachita Shortleaf Pine-Oak Forest and Woodland
- Rocky Mountain Foothill Limber Pine-Juniper Woodland
- Rocky Mountain Lodgepole Pine Forest
- Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland
- Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland
- West Gulf Coastal Plain Pine-Hardwood Forest
- Central Interior and Appalachian Floodplain Systems
- Central Interior and Appalachian Riparian Systems
- Gulf and Atlantic Coastal Plain Floodplain Systems
- Gulf and Atlantic Coastal Plain Small Stream Riparian Systems
- Inter-Mountain Basins Sparsely Vegetated Systems
- Rocky Mountain Montane Riparian Systems
- Western Great Plains Depressional Wetland Systems
- Western Great Plains Floodplain Systems
- Western Great Plains Sparsely Vegetated Systems

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 52 natural types assigned labels by the auto-keys, 23 types (44%) had fewer than 10 samples available for this analysis (Table 2). These under-sampled types tended to include types that are found on the geographic periphery of their range within this GeoArea (e.g., Chihuahuan Loamy Plains Desert Grassland, Chihuahuan Mixed Desert and Thornscrub, Chihuahuan Sandy Plains Semi-Desert Grassland, Chihuahuan Stabilized Coppice Dune and Sand Flat Scrub, Madrean Encinal, Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna), or with the inclusion of atypical environments such as a mountain range in a primarily grassland mapzone. These systems include Rocky Mountain

Subalpine-Montane Limber-Bristlecone Pine Woodland, Southern Rocky Mountain Montane-Subalpine Grassland, and Inter-Mountain Basins Montane Sagebrush Steppe. Other undersampled systems are generally within this range, but are less common types, or simply have had inadequate sampling effort across this region. These include Central and South Texas Coastal Fringe Forest and Woodland, Edwards Plateau Mesic Canyon, Llano Uplift Acidic Forest-Woodland-Glade, Ozark-Ouachita Shortleaf Pine-Oak Forest and Woodland, Southeastern Great Plains Tallgrass Prairie, and West Gulf Coastal Plain Pine-Hardwood Forest.

Table 2. Under-sampled types within GeoArea 6

EVT Code	EVT Name	System elcode	total Plots
2503	Chihuahuan Loamy Plains Desert Grassland	CES302.061	9
2051	Southern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest and Woodland	CES306.823	8
2504	Chihuahuan-Sonoran Desert Bottomland and Swale Grassland	CES302.746	7
2410	Llano Uplift Acidic Forest-Woodland-Glade	CES303.657	7
2064	Colorado Plateau Mixed Low Sagebrush Shrubland	CES304.762	7
2371	West Gulf Coastal Plain Pine-Hardwood Forest	CES203.378	5
2117	Southern Rocky Mountain Ponderosa Pine Savanna	CES306.649	5
2179	Northwestern Great Plains-Black Hills Ponderosa Pine Woodland and Savanna	CES303.650	4
2341	Northwestern Great Plains Canyon	CES303.658	4
2061	Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland	CES304.776	4
2338	Central and South Texas Coastal Fringe Forest and Woodland	CES203.464	3
2423	Southeastern Great Plains Tallgrass Prairie	CES205.685	3
2390	Tamaulipan Mixed Deciduous Thornscrub	CES301.983	3
2100	Chihuahuan Mixed Desert and Thornscrub	CES302.734	3
2133	Chihuahuan Sandy Plains Semi-Desert Grassland	CES302.736	3
2150	Western Great Plains Tallgrass Prairie	CES303.673	3
2050	Rocky Mountain Lodgepole Pine Forest	CES306.820	2
2052	Southern Rocky Mountain Mesic Montane Mixed Conifer Forest and Woodland	CES306.825	2
2367	Ozark-Ouachita Shortleaf Pine-Oak Forest and Woodland	CES202.313	1
2122	Chihuahuan Gypsophilous Grassland and Steppe	CES302.732	1
2076	Chihuahuan Stabilized Coppice Dune and Sand Flat Scrub	CES302.737	1
2524	Edwards Plateau Mesic Canyon	CES303.038	1
2148	Western Great Plains Sand Prairie	CES303.670	1
2153	Inter-Mountain Basins Greasewood Flat	CES304.780	1
2126	Inter-Mountain Basins Montane Sagebrush Steppe	CES304.785	1
2023	Madrean Encinal	CES305.795	1
2057	Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland	CES306.819	1
2146	Southern Rocky Mountain Montane-Subalpine Grassland	CES306.824	1
2049	Rocky Mountain Foothill Limber Pine-Juniper Woodland	CES306.955	1

A total of 4 types (or nearly 8% of 52 types) had >80% agreement between expert and auto-key assignments. All of these types (Western Great Plains Sandhill Steppe, Southern Rocky Mountain

Juniper Woodland and Savanna, Southern Rocky Mountain Ponderosa Pine Woodland, and Western Great Plains Shortgrass Prairie) had at least 20 sample plots. Expert self-assessment of confidence for these types were predominantly 'high' although Southern Rocky Mountain Juniper Woodland and Savanna had nearly equal 'high' and 'moderate' attributions. This is largely because of transitional confusion between a tree savanna site and a grassland plot with scattered trees.

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 35, or nearly 47% of the total types assigned. Further analysis of those grouped within the 60-80% agreement range suggests subtleties within types that left the expert with greater or lesser confidence in their assignment. For example, some plots assigned by the auto-key to Edwards Plateau Dry-Mesic Slope Forest and Woodland were most frequently confused with Edwards Plateau Limestone Savanna and Woodland because differences in classification depend on dominance patterns within a group of shared species, some of which occur in the ground flora which tended to be poorly sampled in the available plot data. These types do transition into one another, so additional floristic indicators might be identified to better distinguish them. This same general pattern, one of carefully reviewing the dominant tree, shrub, or grass elements shared among related types, should be the focus of auto-key improvements for these types.

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**
2519	East-Central Texas Plains Post Oak Savanna and Woodland	CES205.679	50	37	74%	2	21	14
2308	Crosstimbers Oak Forest and Woodland	CES205.682	50	36	72%	5	15	16
2132	Central Mixedgrass Prairie	CES303.659	50	36	72%	8	25	3
2107	Rocky Mountain Gambel Oak-Mixed Montane Shrubland	CES306.818	14	10	71%	10	0	0
2383	Edwards Plateau Limestone Savanna and Woodland	CES303.660	50	34	68%	1	15	18
2095	Apacherian-Chihuahuan Mesquite Upland Scrub	CES302.733	45	29	64%	0	29	0
2081	Inter-Mountain Basins Mixed Salt Desert Scrub	CES304.784	15	9	60%	9	0	0
2523	Edwards Plateau Dry-Mesic Slope Forest and Woodland	CES303.656	50	29	58%	4	20	5
2059	Southern Rocky Mountain Pinyon-Juniper Woodland	CES306.835	50	26	52%	24	2	0
2393	Edwards Plateau Limestone Shrubland	CES303.041	12	2	17%	2	0	0
2111	Western Great Plains Mesquite Woodland and Shrubland	CES303.668	50	7	14%	2	5	0
2086	Rocky Mountain Lower Montane-Foothill Shrubland	CES306.822	20	2	10%	2	0	0

				Plots with Expert Matches				
EVT Code	EVT Name	System Elcode	Total Plots	Total	%	High conf	Med conf	Low conf
2121	Apacherian-Chihuahuan Semi-Desert Grassland and Steppe	CES302.735	21	2	10%	0	2	0
2391	Tamaulipan Mesquite Upland Scrub	CES301.984	50	4	8%	0	0	4
2525	Edwards Plateau Riparian	CES303.652	25	2	8%	0	0	2
2147	Western Great Plains Foothill and Piedmont Grassland	CES303.817	26	1	4%	1	0	0
2127	Inter-Mountain Basins Semi-Desert Shrub-Steppe	CES304.788	27	1	4%	0	0	1
2025	Madrean Pinyon-Juniper Woodland	CES305.797	50	0	0%	0	0	0
2055	Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	CES306.828	17	0	0%	0	0	0

Analysis of the contingency table (Results Workbook) for these types with lesser levels of agreement reveals the many ongoing challenges with finding agreement between experts and auto-keys for complex vegetation types. Here we summarize results for selected systems in GeoArea 6:

East-Central Texas Plains Post Oak Savanna and Woodland – low confidence in plot assignment was generally related to the low species diversity represented in the plot data.

Crosstimbers Oak Forest and Woodland – low confidence in plot assignment was generally related to the low species diversity represented in the plot data. In addition, the species recorded or the photo indicated a disturbed condition. In one case, the species identification was questioned.

Central Mixedgrass Prairie – was most often confused with Western Great Plains Shortgrass Prairie and Western Great Plains Foothill and Piedmont Grassland. There are several shared species between these systems and past grazing practices can influence composition and abundance of key species. For example, long-term heavy livestock grazing of Central Mixedgrass Prairie can reduce or eliminate indicator mid-grasses, leaving the site strongly dominated by grazing tolerant shortgrasses such as *Bouteloua gracilis*, which is characteristic of Western Great Plains Shortgrass Prairie. Western Great Plains Foothill and Piedmont Grassland is also characterized by mid-grasses, but these systems are typically geographically separate from each other (Central Plains versus Western Great Plains near the foothills). Clarifying geographic distribution may help with this one. Also, a better understanding of successional patterns under grazing pressure, and more complete species compositional data in the plots would help distinguish these types.

Rocky Mountain Gambel Oak-Mixed Montane Shrubland – was most often confused with Western Great Plains Sandhill Steppe, so identification of oak species may be an issue.

Edwards Plateau Limestone Savanna and Woodland – was most often confused with systems with which it interfingers. More complete species compositional data in the plots (more thorough data collection in the field) would help distinguish these types.

Apacherian-Chihuahuan Mesquite Upland Scrub – was mostly confused with Western Great Plains Mesquite Woodland and Shrubland. Both systems are dominated by *Prosopis glandulosa*, and transition into one another as the Chihuahuan Desert grades into the Southern Great Plains. Refining geographic distribution of the auto-key and more complete species compositional data in the plots with Chihuahuan Desert indicator species would help distinguish these two types. An additional complexity is that mesquite, while a native species, is invasive in some over-grazed upland areas, making it difficuylt to distinguish if a plot is a degraded example of a grassland system, or represents a more "natural" mesquite upland area.

Inter-Mountain Basins Mixed Salt Desert Scrub – was not consistently confused with any particular system. The *Atriplex* spp. that are characteristic of this systems also occur in other systems so more complete species compositional data in the plots would help distinguish these types (look for saline bottomland indicator species).

Edwards Plateau Dry-Mesic Slope Forest and Woodland – was most often confused with systems with which it interfingers. More complete species compositional data in the plots (more thorough data collection in the field) would help distinguish these types. In several cases, the identification of recorded species was questioned.

Southern Rocky Mountain Pinyon-Juniper Woodland – was frequently confused with the similar Southern Rocky Mountain Juniper Woodland and Savanna. Auto-key criteria and mis-labeled plots need to be reviewed to figure out how to better distinguish. The image clips were critical during expert review to determine if the plot is a juniper savanna or a more open area within a juniper woodland. This contextual information for the plot is vital and hard to replicate in the auto-key. Additionally, More complete species compositional data in the plots (more thorough data collection in the field) would help distinguish these types, as high cover of grasses is important for the savanna systems.

Tamaulipan Mesquite Upland Scrub – a large number of these plots (44) were not assigned because in many cases the only species listed was *Prosopis glandulosa*. Since this species is often an off-site invader, it was difficult to assign these plots to a system.

Western Great Plains Mesquite Woodland and Shrubland – a large number of these plots (31) were not assigned because in many cases the only species listed was *Prosopis glandulosa*. Since this species is often an off-site invader, and can dominate multiple systems, it was difficult to assign these plots to a system.

Rocky Mountain Lower Montane-Foothill Shrubland – was most often confused with Western Great Plains Shortgrass Prairie and Western Great Plains Foothill and Piedmont Grassland.
This system is often adjacent to these other systems and shareS many herbaceous species. Both of these grassland types may have some shrubs present so confusion may relate to amount of shrub cover needed to label to a shrubland system.

Western Great Plains Foothill and Piedmont Grassland – was most often confused with Central Mixedgrass Prairie and Western Great Plains Shortgrass Prairie. Central Mixedgrass Prairie is also characterized by mid-grasses, but systems are typically geographically separate from each other (Central Plain versus Western Great Plains near the foothills). This system often transitions at lower elevations to Western Great Plains Shortgrass Prairie, and is differentiated by the abundance of midgrasses. There are several shared species between these systems and past grazing practices can influence specie composition and reduce the abundance of key midgrass species. Clarifying geographic distribution may help distinguish from Central Mixedgrass Prairie. More complete species compositional data in the plots (more thorough data collection in the field) would help distinguish these types.

Inter-Mountain Basins Semi-Desert Shrub-Steppe – was most often confused with Western Great Plains Shortgrass Prairie. Review of sequence table to clarify that Western Great Plains Shortgrass Prairie may have also an open short shrub component composed of species of *Ericameria, Eriogonum*, *Gutierrezia Krascheninnikovia, Lycium,* or *Opuntia*, especially on disturbed sites.

Madrean Pinyon-Juniper Woodland – was most often confused with Southern Rocky Mountain Juniper Woodland and Savanna and Southern Rocky Mountain Pinyon-Juniper Woodland. Madrean Pinyon-Juniper Woodland is uncommon in this GeoArea other than transition areas with Chihuahuan Desert. Refined geographic distribution information and more complete species compositional data in the plots (more thorough data collection in the field) would help distinguish these types.

Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland – although these plots were codominated by *Picea engelmannii*, elevation was much too low. According to photos, some plots could be mesic or riparian so the spruce might be mis-identified *Picea pungens*, or the acronym PIEN could also be a typo for *Pinus engelmannii*, a Madrean species.

Expert Assignments

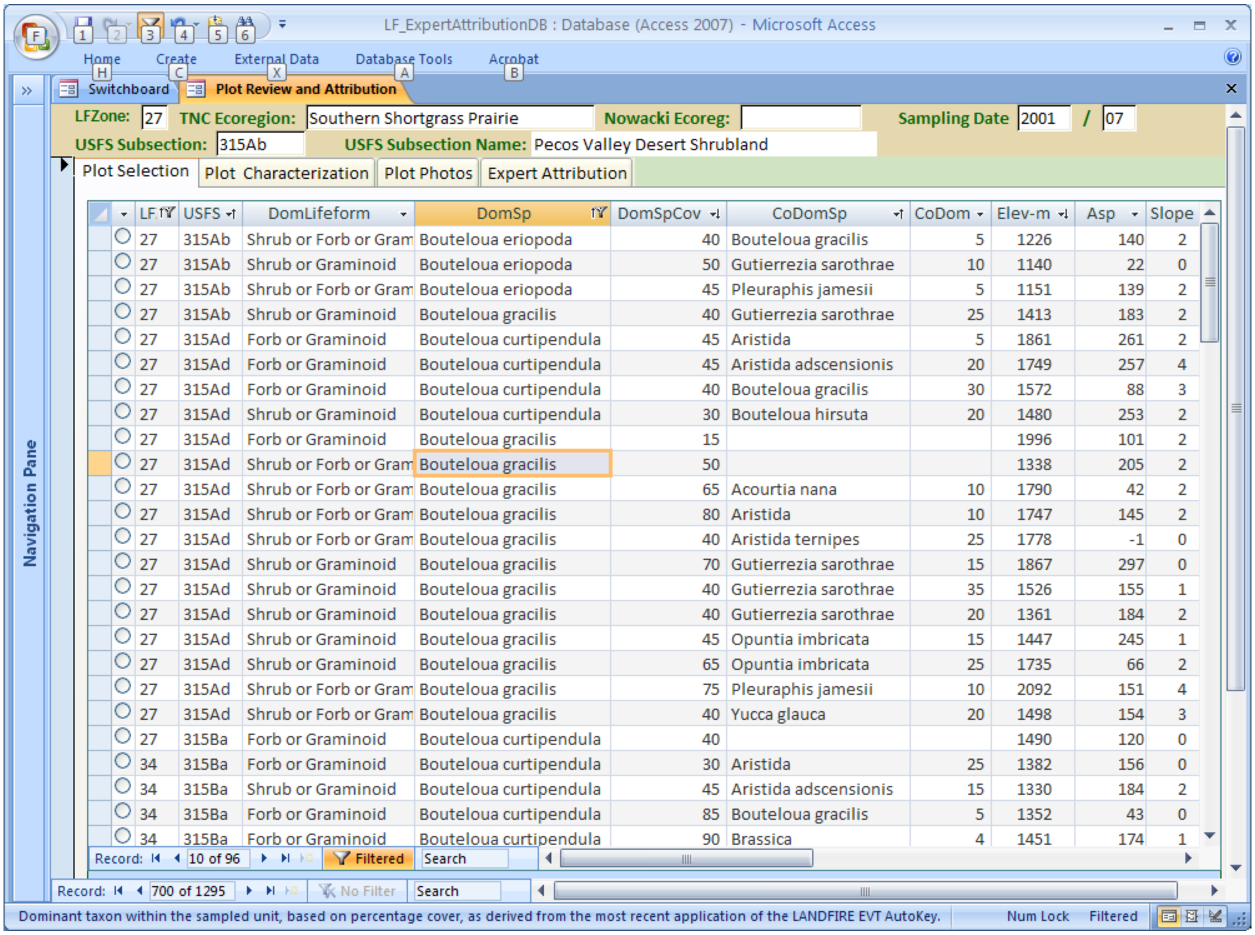
2.2 A report by each GeoArea detailing processes and results, specifically identifying how they made individual assignments.

As described in the methods section above, the expert reviewers worked directly in the expert attribution database (EADB). Since GeoArea 6 had over 2,000 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 graminoides, then sort alphabetically by the dominant and codominant species. The reviewer could also select all treed plots, then select all plots with the same dominant tree species (such as *Quercus stellata*), 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 transition zone between the southwestern portion of this GeoArea and the Chihuahuan Desert, labeling of grassland plots to a desert grassland system or a Great Plains grassland system requires an understanding of diagnostic species, floristic composition and geographic distribution of the systems in question. Complete species list is necessary to confidently distinguish these types.

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



LF	USFS	DomLifeform	DomSp	DomSpCov	CoDomSp	CoDom	Elev-m	Asp	Slope
27	315Ab	Shrub or Forb or Gram	Bouteloua eriopoda	40	Bouteloua gracilis	5	1226	140	2
27	315Ab	Shrub or Graminoid	Bouteloua eriopoda	50	Gutierrezia sarothrae	10	1140	22	0
27	315Ab	Shrub or Forb or Gram	Bouteloua eriopoda	45	Pleuraphis jamesii	5	1151	139	2
27	315Ab	Shrub or Graminoid	Bouteloua gracilis	40	Gutierrezia sarothrae	25	1413	183	2
27	315Ad	Forb or Graminoid	Bouteloua curtipendula	45	Aristida	5	1861	261	2
27	315Ad	Forb or Graminoid	Bouteloua curtipendula	45	Aristida adscensionis	20	1749	257	4
27	315Ad	Forb or Graminoid	Bouteloua curtipendula	40	Bouteloua gracilis	30	1572	88	3
27	315Ad	Shrub or Graminoid	Bouteloua curtipendula	30	Bouteloua hirsuta	20	1480	253	2
27	315Ad	Forb or Graminoid	Bouteloua gracilis	15			1996	101	2
27	315Ad	Shrub or Forb or Gram	Bouteloua gracilis	50			1338	205	2
27	315Ad	Shrub or Forb or Gram	Bouteloua gracilis	65	Acourtia nana	10	1790	42	2
27	315Ad	Shrub or Forb or Gram	Bouteloua gracilis	80	Aristida	10	1747	145	2
27	315Ad	Shrub or Forb or Gram	Bouteloua gracilis	40	Aristida ternipes	25	1778	-1	0
27	315Ad	Shrub or Graminoid	Bouteloua gracilis	70	Gutierrezia sarothrae	15	1867	297	0
27	315Ad	Shrub or Graminoid	Bouteloua gracilis	40	Gutierrezia sarothrae	35	1526	155	1
27	315Ad	Shrub or Graminoid	Bouteloua gracilis	40	Gutierrezia sarothrae	20	1361	184	2
27	315Ad	Shrub or Graminoid	Bouteloua gracilis	45	Opuntia imbricata	15	1447	245	1
27	315Ad	Shrub or Graminoid	Bouteloua gracilis	65	Opuntia imbricata	25	1735	66	2
27	315Ad	Shrub or Forb or Gram	Bouteloua gracilis	75	Pleuraphis jamesii	10	2092	151	4
27	315Ad	Shrub or Forb or Gram	Bouteloua gracilis	40	Yucca glauca	20	1498	154	3
27	315Ba	Forb or Graminoid	Bouteloua curtipendula	40			1490	120	0
34	315Ba	Forb or Graminoid	Bouteloua curtipendula	30	Aristida	25	1382	156	0
34	315Ba	Shrub or Graminoid	Bouteloua curtipendula	45	Aristida adscensionis	15	1330	184	2
34	315Ba	Forb or Graminoid	Bouteloua curtipendula	85	Bouteloua gracilis	5	1352	43	0
34	315Ba	Forb or Graminoid	Bouteloua curtipendula	90	Brassica	4	1451	174	1

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), to distinguish Western Great Plains from Central Great Plains or Chihuahuan Desert from Edwards Plateau,
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 mis-identification of oak or 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:

- This plot appears to be a small patch of Junipers in opening surrounded by PJ woodland. It is possible that it is a disturbed CES303.817 Western Great Plains Foothill and Piedmont Grassland or CES306.822 Rocky Mountain Lower Montane-Foothill Shrubland. The image clip is critical to determine tree savanna from open woodland
- Plot species list is incomplete. Plot photo shows *Artemisia filifolia* in foreground and *Aristida* sp. indicate similarities to CES303.671 Western Great Plains Sandhill Steppe
- Incomplete species list, Jumiper cover in photo is very low<5% not 10% so not savanna.
- Substrate information would improve assignment confidence.
- Significant cover of suffrutescent /dwarf-shrubs such as *Gutierrezia sarothrae*, *Krascheninnikovia lanata, Yucca glauca* and *Artemisia frigida* is included in concept of shortgrass steppe/prairie.
- This diverse plot is more similar to CES302.735 Apacherian-Chihuahuan Semi-Desert Grassland and Steppe than Western Great Plains Shortgrass Prairie. The desert grasslands extends up north into the Pecos Valley and then transitions to the CES303.672 Western Great Plains Shortgrass Prairie in the plains.
- Scattered PJ trees near bottom of canyon as indicated by relatively mesic *Juniperus scopulorum*. No other species reported so likely bare rock dominates ground surface. Could also be assigned to the Southwestern Canyons.
- Landform and hydrology information would improve confidence in assignment.
- If plot in far southern part of subsection MZ 27, then it could be desert scrub such as Apacherian-Chihuahuan Mesquite Upland Scrub or Western Great Plains Mesquite Woodland and Shrubland.
- Plot photo shows plot is in a transition zone between CES306.834 Southern Rocky Mountain Juniper Woodland and Savanna and CES306.822 Rocky Mountain Lower Montane-Foothill Shrubland with widely scattered juniper trees (5%) in a very dense *Cercocarpus montanus* shrubland.

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 54 types assigned to plots by experts, 31 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 23 types, the numbers of samples labeled to a given type ranged from 114 (for Western Great Plains Shortgrass Prairie) down to 10 (for Inter-Mountain Basins Mixed Salt Desert Scrub). For 27 (50%) of these types, experts reported moderate confidence in their labels for at least 20% of the type's plots and 11 indicated low confidence for at least 20% of the type's plots. These statistics are listed in the Results Workbook. 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
Edwards Plateau Limestone Savanna and Woodland	Could also be CES303.651 Edwards Plateau Floodplain
Southeastern Great Plains Riparian Forest	Vegetation supports this assignment, but photo does not indicate a creek
Crosstimbers Oak Forest and Woodland	Seems disturbed with a depauperate species list
Southeastern Great Plains Floodplain Forest	Could also be Southeastern Great Plains Riparian Forest
Llano Uplift Acidic Forest, Woodland and Glade	Limited species information makes attribution difficult
Ozark-Ouachita Dry-Mesic Oak Forest	CES202.707 Ozark-Ouachita Dry Oak Woodland might be a better choice
Llano Estacado Caprock Escarpment and Breaks Shrubland and Steppe	Questionable species id

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. This is perhaps most likely with desert scrub and related vegetation shared with northern Mexico. 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 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 228 samples, experts were able to assign 159 (70%) to an individual ecological system type; a total of 17 individual ecological system types were assigned to these samples. The ability to assign plots to wetland and riparian systems would be expected to increase if more information on landform, soil, and or landscape context could be provided, as suggested above.

The riparian systems groups are labeled to individual system with greater accuracy than the sparsely vegetated system groups in this GeoArea. Sparsely vegetated systems have low vegetation cover, variable species composition, and are often defined more based on substrate (e.g., sand dunes, rock outcrop) which is not currently used in the auto-key. 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 set 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 Shrubland" or "None". Of 137 samples, experts were able to assign 70 (51%) to an individual ecological system type; a total of 20 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 auto-keys 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 auto-keys for use with the NVC.

Within this GeoArea, some 135 terrestrial ecological system types could occur. Of these, 47 have a practical 1:1 relationship with NVC Group concepts, and 76 nest cleanly within 45 NVC Group concepts (1:many group:system relationship), for a total of 123 or 91% 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 auto-key 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 auto-keys. Within this GeoArea, 7 (5%) ecological system types have a more complex relationship with NVC Group concepts (Table 5). Here we provide additional commentary on the implications for auto-key adjustment brought by these types.

Table 5. Ecological Systems of GeoArea 6 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

Ecological System	NVC Group	Groups	Systems
North American Warm Desert Riparian Woodland and Shrubland	G533 North American Warm Desert Riparian Low Bosque & Shrubland	2	2
	G508 Sonoran-Chihuahuan Warm Desert Riparian Scrub	2	3
North American Warm Desert Lower Montane Riparian Woodland and Shrubland	G508 Sonoran-Chihuahuan Warm Desert Riparian Scrub	1	3
North American Warm Desert Riparian Mesquite Bosque	G533 North American Warm Desert Riparian Low Bosque & Shrubland	1	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
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

Inter-Mountain Basins Big Sagebrush Steppe and Inter-Mountain Basins Big Sagebrush Shrubland. These two ecological system types each relate to two NVC Group concepts which themselves are related to other systems: Intermountain Mesic Tall Sagebrush Shrubland & Steppe Group and Intermountain Dry Tall Sagebrush Shrubland Group. Adjustments to auto-keys to capture these complications should be relatively easy, but require that plot data include more complete floristic information, which will form the primary indications that distinguish these two Groups. These are minor systems in this GeoArea. In fact, no samples were assessed with the auto-key or by experts so likely not an issue here.

North American Warm Desert Riparian Woodland and Shrubland. This type relates to parts of two NVC Group concepts which themselves are related to other systems: North American Warm Desert Riparian Low Bosque & Shrubland Group and Sonoran-Chihuahuan Warm Desert Riparian Scrub Group. The part of this system related to the North American Warm Desert Riparian Low Bosque & Shrubland Group is dominated by *Baccharis* spp., and *Fallugia paradoxa* growing along drainages. The part of this system related to the Sonoran-Chihuahuan Warm Desert Riparian Scrub Group is lower elevation (<1200 m) desert riparian sites that are dominated by *Acer negundo, Fraxinus velutina, Populus fremontii, Salix gooddingii, Salix lasiolepis, Celtis laevigata var. reticulata, Platanus racemosa,* or *Juglans major* growing in riparian areas.

North American Warm Desert Lower Montane Riparian Woodland and Shrubland. This type relates to part of the Sonoran-Chihuahuan Warm Desert Riparian Scrub Group, which is also related in part to North American Warm Desert Riparian Woodland and Shrubland system. The part of this system related to the Sonoran-Chihuahuan Warm Desert Riparian Scrub Group is higher elevation (1100 m) desert riparian sites that are dominated by *Populus angustifolia, Populus deltoides ssp. wislizeni, Populus fremontii, Platanus wrightii, Juglans major, Fraxinus velutina, Alnus oblongifolia,* or *Sapindus saponaria.* Other parts of this group relate to lower elevation North American Warm Desert Riparian Woodland and Shrubland, and Sonoran Fan Palm Oasis (does not occur in this GeoArea).

North American Warm Desert Riparian Mesquite Bosque. This type relates to part of the North American Warm Desert Riparian Low Bosque & Shrubland, which is also related in part to North American Warm Desert Riparian Woodland and Shrubland system. The part of this system related to North American Warm Desert Riparian Low Bosque & Shrubland Group is dominated by *Prosopis* spp growing along intermittently flooded drainages and riparian areas.

Rocky Mountain Alpine-Montane Wet Meadow. This type relates to two NVC Group concepts: Vancouverian & Rocky Mountain Subalpine Snowbed, Wet Meadow & Dwarf-Shrubland Group and Vancouverian & Rocky Mountain Montane Wet Meadow Group. The part of this system related to the Rocky Mountain Alpine-Montane Wet Meadow Group would occur in montane-subalpine elevation herbaceous wetlands in this GeoArea. The part of this system related to the Vancouverian & Rocky Mountain Subalpine Snowbed, Wet Meadow & Dwarf-Shrubland Group would include *Dasiphora fruticosa* ssp*. floribunda* Shrublands and other wet shrublands.

Rocky Mountain Lower Montane-Foothill Riparian Woodland and Shrubland. This type relates to two group concepts: Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland and Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest. This classification split between treed and shrub-dominated components of riparian areas may be relatively easy to identify from plot data within the auto-key, although they may or may not provide practical units for mapping, modeling, and analysis, since these components tend to represent distinct successional patches within the same riparian area.

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 auto-keys for US-NVC Macrogroups instead of ecological systems could potentially resolve disagreements between experts and auto-keys found at the ecological systems level. To evaluate the potential effect of using the auto-key for Macrogroups, we arranged the ecological system types by US-NVC Macrogroup in the expert-auto-key contingency table, and also compared the percent of expert auto-key matches at the system level versus the Macrogroup level (Table 6).

There are 26 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, especially Apacherian-Chihuahuan Semi-Desert Grassland & Steppe, Comanchian Forest & Woodland, and Rocky Mountain Two-needle Pinyon - Juniper Woodland, consideration must be given to the fact that many of these Macrogroups are in fact very broad concepts, and include ecologically diverse system types. For example, the Comanchian Forest & Woodland Macrogroup clusters a set of 4 system types in GeoArea 6: Edwards Plateau Mesic Canyon, Edwards Plateau Dry-Mesic Slope Forest and Woodland, Edwards Plateau Limestone Savanna and Woodland, and Llano Uplift Acidic Forest, Woodland and Glade. So, while rolling up from system to Macrogroup may improve the number of matches between auto-key and expert, the roll-up would conceal important vegetation differences grouping mesic bottomland forests and dry upland woodlands, and is probably not desirable.

On the other hand, Rocky Mountain Two-needle Pinyon - Juniper Woodland Macrogroup is composed of two similar ecological systems in this GeoArea (Southern Rocky Mountain Juniper Woodland and Savanna and Southern Rocky Mountain Pinyon-Juniper Woodland) that could be grouped to significantly improve the number of auto-key to expert matches from 64% to 96%, because confusion was almost entirely between these two systems.

Another example, Apacherian-Chihuahuan Semi-Desert Grassland & Steppe Macrogroup (41 matches) is composed of 5 ecological systems (12%) that could be grouped to improve auto-key accuracy. However, in GeoArea 6, these systems are still poorly auto-keyed and are often confused with Western Great Plains Shortgrass Prairie (in a different Macrogroup) so fixing auto-key would be more fruitful.

- Chihuahuan Loamy Plains Desert Grassland
- Chihuahuan Gypsophilous Grassland and Steppe
- Apacherian-Chihuahuan Semi-Desert Grassland and Steppe
- Chihuahuan Sandy Plains Semi-Desert Grassland
- Chihuahuan-Sonoran Desert Bottomland and Swale Grassland

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**
M086 Chihuahuan Desert Scrub	3	49	61%	67%
M087 Apacherian-Chihuahuan Semi-Desert Grassland & Steppe	5	41	12%	41%
M130 Tamaulipan Scrub & Grassland	2	53	13%	13%
M093 Great Basin Saltbrush Scrub	1	15	60%	60%
M169 Great Basin & Intermountain Tall Sagebrush Shrubland & Steppe	1	1	0%	0%
M170 Great Basin & Intermountain Dwarf Sage Shrubland & Steppe	1	7	100%	100%
M171 Great Basin & Intermountain Dry Shrubland & Grassland	1	27	4%	4%
M008 Southern Mixed Deciduous-Evergreen Broadleaf Forest	1	3	67%	67%
M157 Loblolly & Shortleaf Pine - Oak Forest & Woodland	1	5	0%	0%
M010 Madrean Warm Lowland Evergreen Woodland	2	51	2%	6%
M015 Comanchian Forest & Woodland	4	108	65%	86%
M158 Southern Plains Scrub Woodland & Shrubland	1	12	17%	17%
M012 Central Oak-Hardwood & Pine Forest	1	50	74%	74%
M016 Southern Hardwood & Pine Forest	2	51	71%	82%
M017 Northern Rocky Mountain Lower Montane & Foothill Forest	2	5	0%	0%
M020 Rocky Mountain Subalpine & High Montane Conifer Forest	4	24	0%	0%
M022 Southern Rocky Mountain Lower Montane Forest	4	62	69%	82%
M027 Rocky Mountain Two-needle Pinyon - Juniper Woodland	2	72	64%	96%
M154 Edwards Plateau Riparian Shrubland & Woodland	1	25	8%	48%
M116 Great Plains Cliff, Scree & Rock Vegetation	1	4	0%	0%
M049 Southern Rocky Mountain Montane Grassland & Shrubland	3	35	37%	37%
M051 Great Plains Mixedgrass Prairie & Shrubland	3	79	48%	58%
M052 Great Plains Sand Grassland & Shrubland	2	47	89%	91%

M053 Great Plains Shortgrass Prairie & Shrubland	2	100	47%	48%
M054 Great Plains Tallgrass Prairie, Savanna & Shrubland	1	3	100%	100%
M082 Cool Semi-Desert Alkali-Saline Wetland	1	1	100%	100%

US-NVC Divisions

NVC Divisions are substantially simplified vegetation concepts relative to terrestrial ecological system types, so auto-keys 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 auto-key to encompass the 10 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 in the Rocky Mountains, *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 (Figure 1). Hence each sequence table/auto-key included tens 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

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 (landform, soils, geology, landscape position) information for pre-processing plots and/or inclusion of features in auto-keys. Adding substrate type and geology information is especially relevant is sparsely vegetated systems are to be labeled by the auto-key. Sparsely vegetated systems typically have the low vegetation cover, variable species composition, and are often defined more based on substrate (e.g., sand dunes, rock outcrop). Percent cover of ground cover, such as bedrock, bare ground, litter would further help differentiate certain types.

Narrowing vs. broadening the geographic application of the auto-key – Using USFS SubSections would help when there is accurate distributional information on the system. In certain areas, such as transition zones between analogous systems such as Apacherian-Chihuahuan Mesquite Upland Scrub and Western

Great Plains Mesquite Woodland and Shrubland, labeling plots using complete floristic information may lead to greater accuracy.

Adjustment to auto-keys – additional requirements for a more complete species list of vegetation sample data; e.g., ground cover data, and a greater percentage of woody species (not just dominant and co-dominant species) would help improve accuracy for some systems, but limit the number of total plots for mapping. On the other hand some systems can be confidently labeled by a single dominant species in certain parts of their range, such as *Pinus edulis* of the Southern Rocky Mountain Pinyon-Juniper Woodland in Colorado.

A limitation of the simplified auto key methodology (and to a lesser extent Expert Labeling) is not consistently addressing disturbance, invasion (by native species, such as (mesquite), drought, seral stage, and grazing history - factors that affect species composition that the Autokey results are based on.

Expert review and labeling of plots for types of low confidence from auto-key, would reduce labeling errors with less expense. Using a similar expert review database, high confidence labeled plots could be identified and sorted out quickly to focus on the more difficult plots.

More thoroughly collected field data would help to distinguish a number of ecological system types; a lack of complete florisitic information for plots that could represent 2 or 3 differenty types makes it impossible to key them either in a sequence table, or for an expert to assign with high confidence.

Adjustments to Map Legends such as moving to Group/Macrogroup concepts where systems level remains challenging is an option, but has the risk of making map classes thematically too broad. This remedy could be reviewed and applied on a case-by-case basis.

Coping with uncertainty about what proportion of types could NOT be adequately handled through any of the above adjustments should to be addressed during auto-key improvement.

Finally, 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.