Expert Attribution for Auto-Key Improvements (LANDFIRE) and Advancing Methods for integration with the revised US-National Vegetation Classification Standard: GeoArea 3

FINAL REPORT

Prepared by

NatureServe

For the NPS Vegetation Inventory Program & LANDFIRE

30 June 2012

© Marion Reid. Fire near Redding CA.

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 Gwen Kittel, NatureServe Regional Vegetation Ecologist.

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.

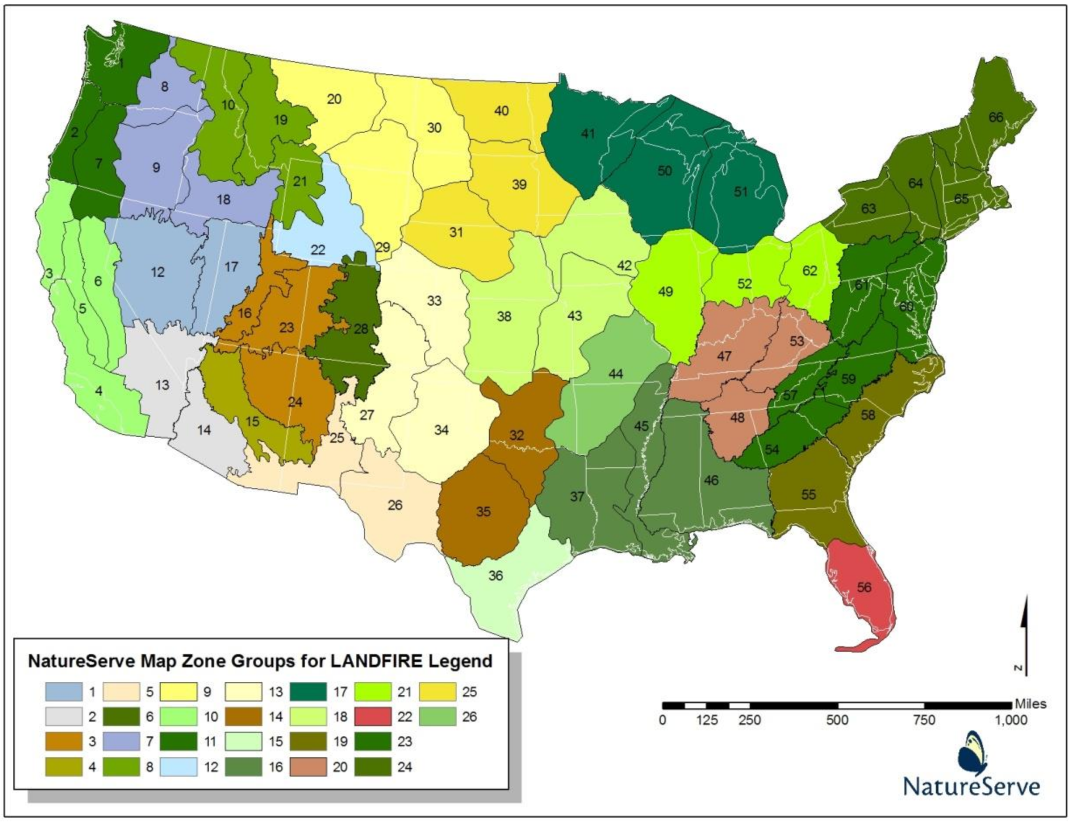
LANDFIRE's short-term needs, and long-term plans, required a repeatable methodology, consistently applied rules to categorize each reference sample, and documentation of the criteria applied. In essence, sequence tables codify the criteria and methods for keying geo-referenced vegetation data to a land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

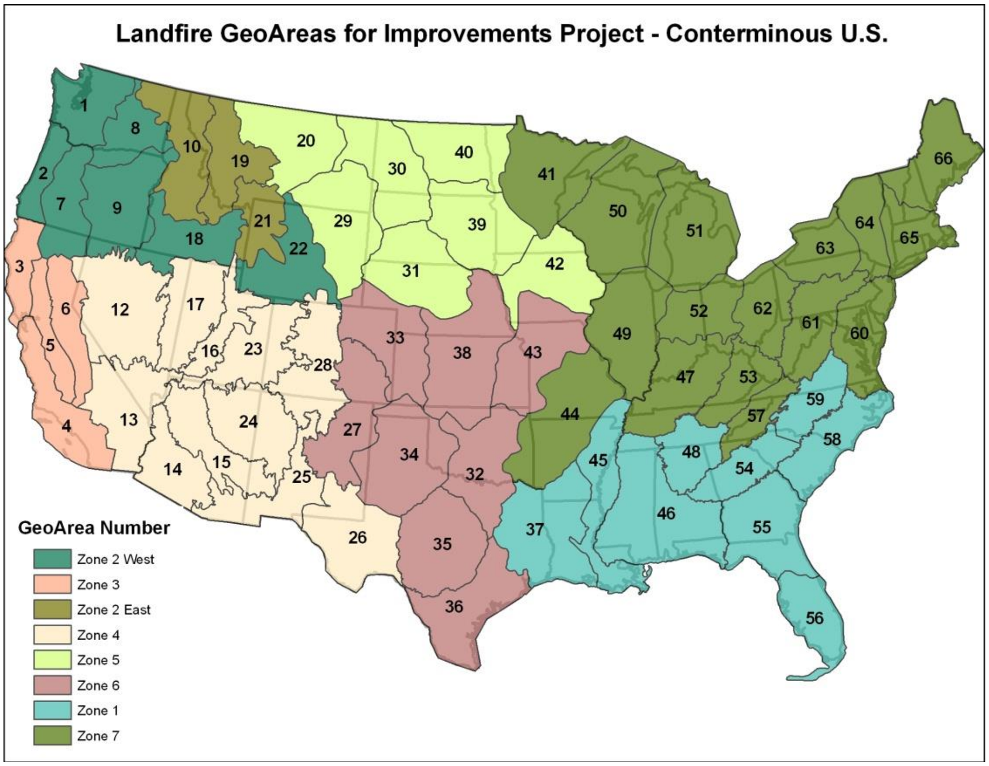
Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in the coterminous U.S.



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:

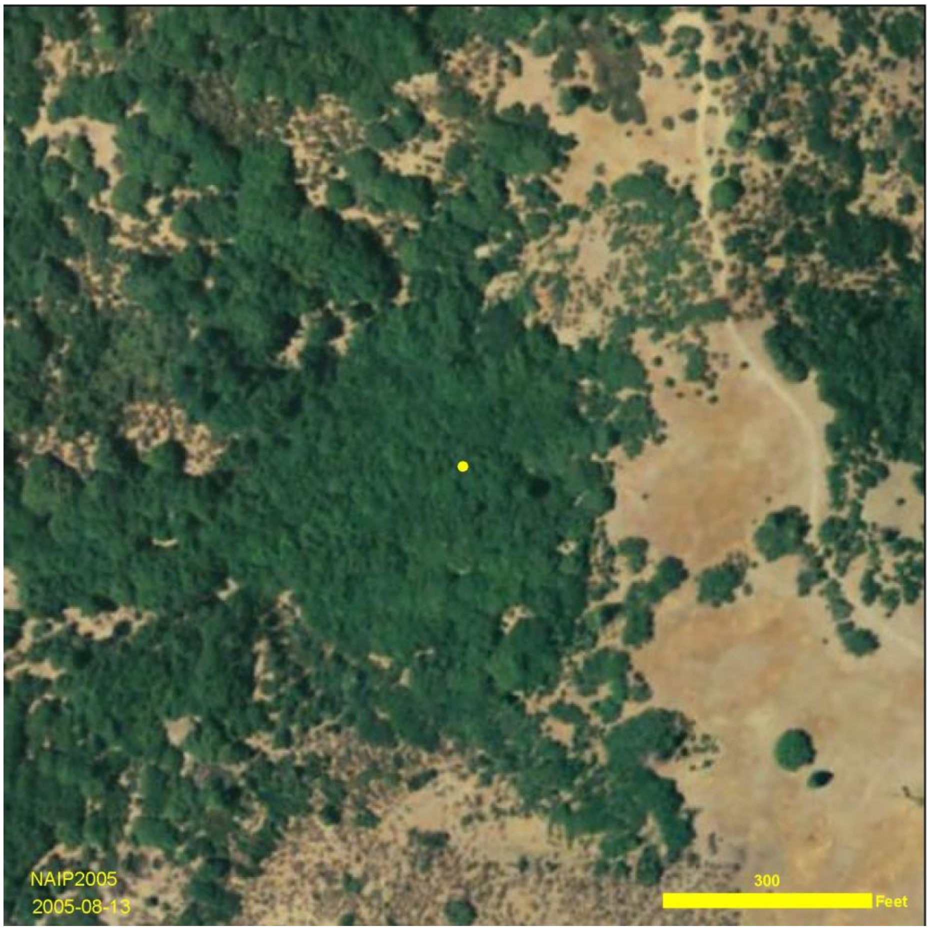
1. Providing the same data as are used in the auto-key procedures
2. Providing additional data that were not originally used in the auto-keys, and
3. Maintaining the “privacy” of the FIA data, ensuring the experts could not determine which plots were FIA vs not

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 3



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 3 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 3: California

GeoArea 3 encompasses Northern California Coastal Range, Southern California Coastal Range, California Central Valley, and Sierra Nevada Mountain Range (Figure 2). This GeoArea includes a total of 4 map zones (Map zones 3-6), originally clustered for purposes of designing and implementing auto-keys (Figure 1). The total number of plots in this Geo Area analysis was 2,049. A total of 57 natural ecological system types were assigned to a total of 2,099 plots by the auto-keys. A total of 75 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 14 types were assigned by the auto-key but were not assigned by experts:

- Baja Semi-Desert Coastal Succulent Scrub
- California Mesic Serpentine Grassland
- Columbia Plateau Western Juniper Woodland and Savanna
- Mediterranean California Alpine Fell-Field
- North Pacific Maritime Mesic-Wet Douglas-fir-Western Hemlock Forest
- Northern California Mesic Subalpine Woodland
- California Montane Riparian Systems
- Inter-Mountain Basins Montane Riparian Systems
- Inter-Mountain Basins Sparsely Vegetated Systems
- Mediterranean California Sparsely Vegetated Systems
- North American Warm Desert Riparian Systems
- North American Warm Desert Sparsely Vegetated Systems
- Pacific Coastal Dunes and Other Sparsely Vegetated Systems
- Pacific Coastal Marsh 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 75 natural types assigned labels by the auto-keys, 17 types (24%) had fewer than 10 samples available for this analysis (Table 2). These under-sampled types tended to include types that are found on the periphery of their range within this GeoArea (e.g., Inter-Mountain Basins Greasewood Flat, Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland, Columbia Plateau Western Juniper Woodland and Savanna, Inter-Mountain Basins Big Sagebrush Steppe, Inter-Mountain Basins Mixed Salt Desert Scrub, Inter-Mountain Basins Semi-Desert Grassland, North Pacific Hypermaritime Seasonal Sitka Spruce Forest, North Pacific Maritime Mesic-Wet Douglas-fir-Western Hemlock Forest), while others are generally within this range, but are less common types, or simply have had inadequate sampling effort (for example were difficult to access) across this region. These include Mediterranean California Alpine Dry Tundra, Mediterranean California Alpine Fell-Field, Klamath-Siskiyou Upper Montane Serpentine Mixed Conifer Woodland, California Mesic Serpentine Grassland, Klamath-Siskiyou Lower Montane Serpentine Mixed Conifer Woodland, California Coastal Closed-Cone Conifer Forest and Woodland, Baja Semi-Desert Coastal Succulent Scrub and California Maritime Chaparral.

Table 2. Under-sampled types within GeoArea 3

EVT Code	EVT Name	System elcode	total Plots
2136	Mediterranean California Alpine Dry Tundra	CES206.939	8
2153	Inter-Mountain Basins Greasewood Flat	CES304.780	8
2061	Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland	CES304.776	7

EVT Code	EVT Name	System elcode	total Plots
2022	Klamath-Siskiyou Upper Montane Serpentine Mixed Conifer Woodland	CES206.914	6
2073	Baja Semi-Desert Coastal Succulent Scrub	CES206.934	6
2131	California Northern Coastal Grassland	CES206.941	5
2130	California Mesic Serpentine Grassland	CES206.943	5
2067	Mediterranean California Alpine Fell-Field	CES206.900	4
2021	Klamath-Siskiyou Lower Montane Serpentine Mixed Conifer Woodland	CES206.917	4
2177	California Coastal Closed-Cone Conifer Forest and Woodland	CES206.922	3
2017	Columbia Plateau Western Juniper Woodland and Savanna	CES304.082	3
2125	Inter-Mountain Basins Big Sagebrush Steppe	CES304.778	3
2036	North Pacific Hypermaritime Seasonal Sitka Spruce Forest	CES204.841	2
2039	North Pacific Maritime Mesic-Wet Douglas-fir-Western Hemlock Forest	CES204.002	1
2096	California Maritime Chaparral	CES206.929	1
2081	Inter-Mountain Basins Mixed Salt Desert Scrub	CES304.784	1
2135	Inter-Mountain Basins Semi-Desert Grassland	CES304.787	1

A total of 18 types (or nearly 32% of 57 types) had >80% agreement between expert and auto-key assignments. All of these types had at least 25 sample plots. Expert self-assessment of confidence for these types were predominantly 'high'.

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 22, or nearly 39% 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 Mediterranean California Mixed Evergreen Forest were most frequently mistaken for Mediterranean California Lower Montane Black Oak-Conifer Forest and Woodland probably because they included *Pseudotsuga menziesii* and *Quercus chrysolepis,* which is common to both ecosystems. 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
2043	Mediterranean California Mixed Evergreen Forest	CES206.919	49	38	78%	37	1	0
2034	Mediterranean California Mesic Serpentine Woodland and Chaparral	CES206.928	50	38	76%	37	1	0
2097	California Mesic Chaparral	CES206.926	50	36	72%	34	2	0
2029	Mediterranean California Mixed Oak Woodland	CES206.909	50	32	64%	32	0	0

2011	Rocky Mountain Aspen Forest and Woodland	CES306.813	33	21	64%	21	0	0
2030	Mediterranean California Lower Montane Black Oak-Conifer Forest and Woodland	CES206.923	38	24	63%	24	0	0
2126	Inter-Mountain Basins Montane Sagebrush Steppe	CES304.785	36	20	56%	20	0	0
2088	Sonora-Mojave Mixed Salt Desert Scrub	CES302.749	24	13	54%	13	0	0
2028	Mediterranean California Mesic Mixed Conifer Forest and Woodland	CES206.915	50	25	50%	25	0	0
2099	California Xeric Serpentine Chaparral	CES206.927	47	23	49%	22	1	0
2098	California Montane Woodland and Chaparral	CES206.925	59	28	47%	28	0	0
2027	Mediterranean California Dry-Mesic Mixed Conifer Forest and Woodland	CES206.916	50	19	38%	19	0	0
2033	Mediterranean California Subalpine Woodland	CES206.910	49	18	37%	18	0	0
2118	Southern California Oak Woodland and Savanna	CES206.938	50	17	34%	17	0	0
2112	California Central Valley Mixed Oak Savanna	CES206.935	47	15	32%	15	0	0
2014	Central and Southern California Mixed Evergreen Woodland	CES206.920	50	14	28%	14	0	0
2105	Northern and Central California Dry-Mesic Chaparral	CES206.931	50	14	28%	14	0	0
2108	Sonora-Mojave Semi-Desert Chaparral	CES302.757	50	13	26%	13	0	0
2008	North Pacific Oak Woodland	CES204.852	16	3	19%	3	0	0
2082	Mojave Mid-Elevation Mixed Desert Scrub	CES302.742	50	7	14%	6	1	0
2103	Great Basin Semi-Desert Chaparral	CES304.001	16	2	13%	2	0	0
2044	Northern California Mesic Subalpine Woodland	CES206.911	26	0	0%	0	0	0

Analysis of the contingency table (see 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 a cross-section of results from GeoArea 3. It's important to note that the sequence table for these California map zones did not include any ability to key plots based on an elevation criterion. Some of the disagreements between the auto-keyed results and the expert results might be resolved in the future by the addition of elevation rules in the sequence table.

The Mediterranean California Mesic Mixed Conifer Forest and Woodland (50%) and the Mediterranean California Dry-Mesic Mixed Conifer Forest and Woodland (38%) are very similar types distiguished on a moisture gradient, and were often confused with each other in the auto-key. Information on elevation, aspect and additional component species composition aided experts in teasing these two systems apart.

Auto-key often mistook Great Basin Xeric Mixed Sagebrush Shrubland for Inter-Mountain Basins Montane Sagebrush Steppe, which relies on the identification of the subspecies of several *Artemisia* species for proper classification; they also intergrade at higher elevation where they transition from one system to the other. Additional elevation and geographic information would help the auto-key process.

Mediterranean California Subalpine Woodland (37%) and Northern California Mesic Subalpine Woodland (0%) were often confused with lower elevation forest types. Elevation is a key factor that would aid the auto-key process.

Southern California Oak Woodland and Savanna (34%) and California Central Valley Mixed Oak Savanna (32%) were often confused with California Coastal Live Oak Woodland and Savanna as both can have *Querus agrifolia* mixed with other oak species. For stands with limited species compositional information, the addition of geographical location confirms the type of oak savanna.

Chaparral is complex and very diverse in California, so location and geography as well as species composition is critical to classify to the different types. The Auto-key and expert tended to disagree between the Northern and Central California Dry-Mesic Chaparral system and the Southern California Dry-Mesic Chaparral system. The species composition tends to intergrade where the two become adjacent, and correct identification of *Ceanothus* species becomes critical. Again location/geographic information and accurate species identification will be very helpful.

Expert Assignments

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

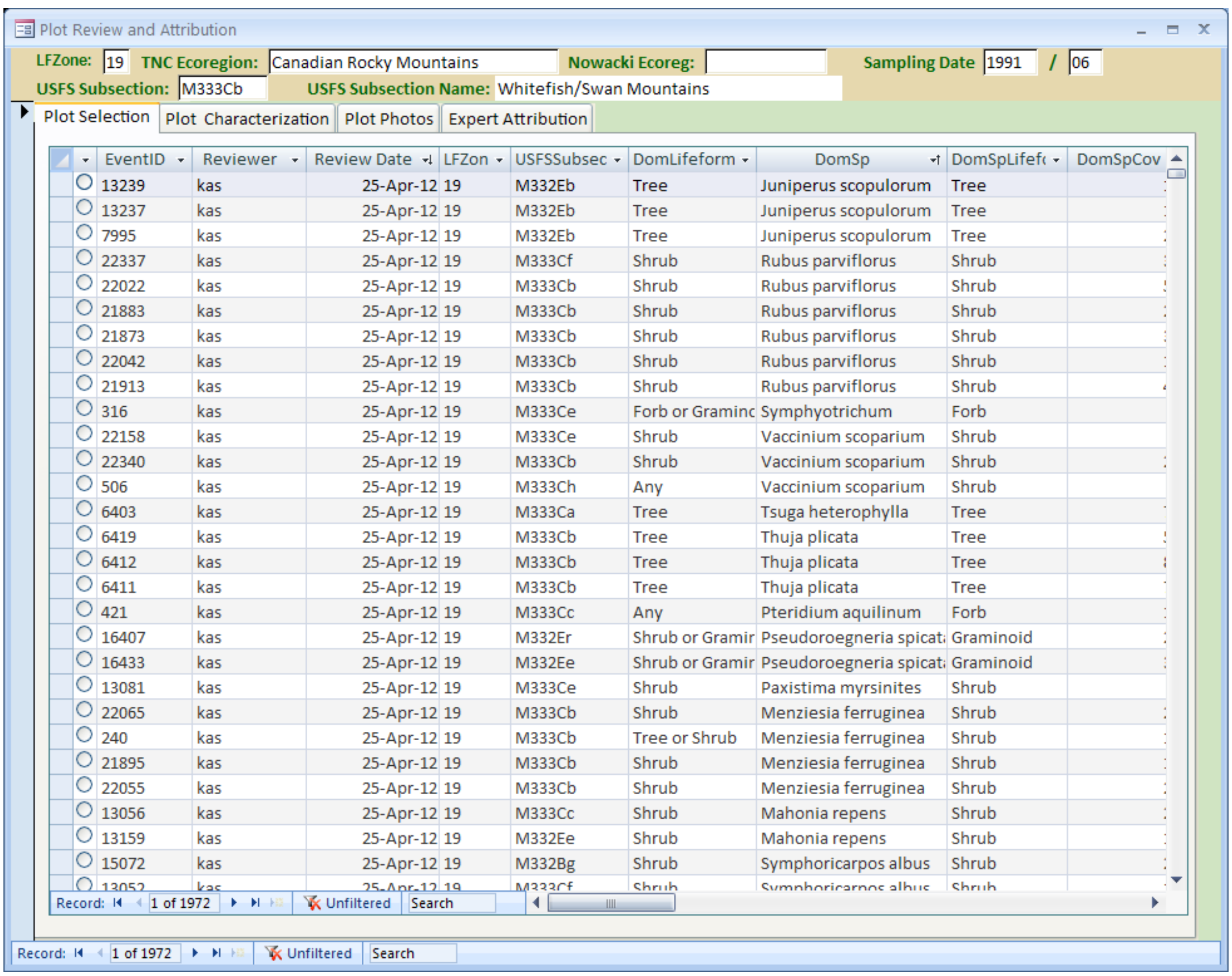
As described in the methods section above, the expert reviewers worked directly in the expert attribution database (EADB). Since GeoArea 3 had over 2,000 plots to review, a systematic, efficient process for reviewing and labeling plots was required. The forms provided in the EADB allowed the reviewer 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 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 contorta*), 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 foothills of the Sierra-Nevada there is a transition from Mediterranean California Lower Montane Black Oak-Conifer Forest and Woodland to Mediterranean California Dry-Mesic Mixed Conifer Forest and Woodland. In these areas, the tree canopy might be a mix of ponderosa, black oak (*Quercus kelloggii*), incense cedar (*Calocedrus decurrens)* and *Quercus chrysolepis*, along with a variable mixture of deciduous shrubs and grasses. Cover of the trees can vary from 10% to more than 80% . In these cases, the reviewer would encounter plots of mixed composition, and need to determine whether those plots represented Mediterranean California Dry-Mesic Mixed Conifer Forest and Woodland*,* the Mediterranean California Mesic Mixed Conifer Forest and Woodland or the Mediterranean California Mesic Serpentine Woodland and Chaparral.

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), to distinguish Southern California Coastal Mountains from Northern California Coast, and the Central Valley from Mojave Desert,
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 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:

- Coastal Chaparral species composition changes from north to south, but where northern and southern coastal California intergrade, it is helpful to know if the location is on the drier interior aspect or ocean facing part of the same mountain range. This is particularly useful if the full species composition is unavailable. Information or confirmation of soil types to identify serpentine areas would also be beneficial. Closely aligned chaparral includes California Xeric Serpentine Chaparral, California Maritime Chaparral, Southern California Dry-Mesic Chaparral, Northern and Central California Dry-Mesic Chaparral, and Southern California Coastal Scrub
- Coast live oak is the characteristic species of the California Coastal Live Oak Woodland and Savanna, however it is a wide-spread species in California and may occur in the foothills of the Sierra-Nevada or along the western edges of the central valley where it mixes with other oak species and is often a part of the Mediterranean California Mixed Oak Woodland, the California Central Valley Mixed Oak Savanna or the Mediterranean California Mixed Evergreen Forest. Again full species composition and geographic location aid the accuracy of point classification.

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 75 types assigned to plots by experts, 32 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 43 types, the numbers of samples labeled to a given type ranged from 152 (for Southern California Dry-Mesic Chaparral) down to 11 (for Mojave Mid-Elevation Mixed Desert Scrub). For all of these types, experts reported at least moderate confidence in their labels for at least 70% of the type's plots. In fact the vast majority of plot labels were given with high confidence, and almost no plots were given low confidence. 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
Great Basin Semi-Desert Chaparral	Need Quercus species identified
Mediterranean California Mesic Serpentine Woodland and Chaparral	Not sure soils are serpentine
Southern California Coastal Scrub	Has half Maritime and half dry Chaparral spp, so it depends on the location of the stand
Sierra Nevada Subalpine Lodgepole Pine Forest and Woodland	need var of lodgepole

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. 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 212 samples, experts were able to assign 209 (99%) to an individual ecological system type; a total of 26 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.

For example experts were able to differentiate Mediterranean California Foothill and Lower Montane Riparian Woodland, California Central Valley Riparian Woodland and Shrubland and Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland, based on species composition and geographic location, while the Landfire auto-key had these lumped into a single “California Montane Riparian Systems.”

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 Grassland" or "None". Of 177 samples, experts were able to assign 137 (77%) to an individual ecological system types; a total of 45 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 117 terrestrial ecological system types could occur. Of these, 33 have a practical 1:1 relationship with NVC Group concepts, and 72 nest cleanly within 37 NVC Group concepts (1:many group:system relationship), for a total of 117 or 90% 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, 10 (9%) ecological system types have a more complex relationship with NVC Group concepts (Table 5). Here we provide additional commentary on the implications for autokey adjustment brought by these types.

- Dominance of life form, already incorporated into the auto-keys, can account for and key the differences between G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland and G503 Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest
- Differentiating Steppe from Grassland is possible with the information on lifeform cover, which is incorporated into the auto-keys. These NVC Groups have further defined Ecological Systems along a moisture gradient, differentiating by detailed floristic composition in addition to relative life form cover will improve the key.
- Dominance by life form (woodland vs. shrubland), already incorporated into the auto-keys, and geographic location of each point, can account for and key the differences between G533 North American Warm Desert Riparian Low Bosque & Shrubland and G508 Sonoran-Chihuahuan Warm Desert Riparian Scrub.
- North Pacific Avalanche Chute Shrubland is a difficult system as it is more defined by the physical location than by species composition, as tree species can be maintained in a shrub-like form, and the species composition can be nearly identical to shrublands along riparian areas and in other wetland settings not associated with avalanche chutes. Slope, elevation and aspect, while extremely informative, are not adequate to confirm the landform characteristics of an avalanche chute. The distinction of 2 different NVC Groups defined more on geography and floristics will probably be straightforward to implement in the sequence table.
- North Pacific Shrub Swamp ecological system is very wide spread, such that several Groups now include the component associations. Geographic location information will be needed to key to each Group in turn.

Table 5. Ecological Systems of GeoArea 3 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 to which each NVC Group related is shown in the Systems column.

Ecological System	**NVC Group**	**Groups**	**Systems**
Great Basin Foothill and 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
Great Basin Xeric Mixed Sagebrush Shrubland	G303 Intermountain Dry Tall Sagebrush Shrubland	1	3
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
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
Sonoran Fan Palm Oasis	G508 Sonoran-Chihuahuan Warm Desert Riparian Scrub	1	3
North Pacific Avalanche Chute Shrubland	G305 Northern Rocky Mountain High Montane Mesic Shrubland [Provisional]	2	3
	G354 Vancouverian Alder - Salmonberry - Willow Shrubland	2	8
North Pacific Montane Grassland	G354 Vancouverian Alder - Salmonberry - Willow Shrubland	1	8
North Pacific Montane Riparian Woodland and Shrubland	G507 North Pacific Montane Riparian Woodland	2	1
	G322 Vancouverian Wet Shrubland	2	4
North Pacific Shrub Swamp	G256 North Pacific Maritime Hardwood-Conifer Rich Swamp	3	3
	G610 North Pacific Maritime Poor Swamp & Bog Forest	3	2
	G322 Vancouverian Wet Shrubland	3	4

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 23 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, 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, in the case of Macrogroup 009 (Quercus agrifolia - Quercus lobata - Umbellularia californica - Cupressus spp. - Pinus spp. Forest & Woodland) rolling up from systems to the MG would improve agreement from 52% to 78%, a significant increase. However, this Macrogroup contains 7 diverse ecological systems, ranging from Central Valley oak woodlands and savannas to coastal pine woodlands, mixed conifer oak types in the foothills of the Sierras and Coast Ranges, and the mixed evergreen forests found just inland of the fog belt close to the coast. Combining all of these into a single map class would result in the loss of much ecological information. A sequence table for this Macrogroup would not be difficult to construct, but investments would better be made in improving keying of the individual systems.

Macrogroup 025 (Abies magnifica - Abies X shastensis - Tsuga mertensiana - Pinus contorta var. murrayana Forest) is an interesting case, where rolling up from systems to Macrogroup improved agreement between expert and auto-key from 62% to 93% (Table 6). This suggests the Macrogroup concept is well understood, or can be clearly distinguished in individual plots, but the system types comprising the Macrogroup are somewhat difficult to separate in a sequence table. This Macrogroup includes the subalpine forests and woodlands of the Sierras (red fir, lodgepole, subalpine western juniper, mountain hemlock) which are often mixed conifer forests, with gradients of moisture and topographic position, along with fire history, determining species composition. Further work on clarifying the differences amongst these systems is needed.

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**
M088 Mojave-Sonoran Semi-Desert Scrub	2	71	38%	39%
M089 Viscaino-Baja California Desert Scrub	1	6	0%	0%
M093 Great Basin Saltbrush Scrub	2	25	56%	56%
M169 Great Basin & Intermountain Tall Sagebrush Shrubland & Steppe	4	137	80%	96%
M171 Great Basin & Intermountain Dry Shrubland & Grassland	1	1	0%	0%
M009 Quercus agrifolia - Quercus lobata - Umbellularia californica - Cupressus spp. - Pinus spp. Forest & Woodland	7	257	52%	78%
M019 Quercus garryana - Quercus kelloggii - Pseudotsuga menziesii - Arbutus menziesii Forest & Woodland	3	115	63%	68%
M020 Rocky Mountain Subalpine & High Montane Conifer Forest	2	40	63%	73%
M023 Calocedrus decurrens - Pinus (lambertiana, jeffreyi, monticola) - Abies concolor var. lowiana Forest	6	190	57%	71%

M024 Tsuga heterophylla - Picea sitchensis - Sequoia sempervirens - Acer macrophyllum Forest	3	53	96%	98%
M025 Abies magnifica - Abies X shastensis - Tsuga mertensiana - Pinus contorta var. murrayana Forest	4	175	62%	93%
M026 Intermountain Singleleaf Pinyon - Western Juniper Woodland	3	82	93%	98%
M036 Warm Mediterranean & Desert Riparian, Flooded & Swamp Forest	1	48	100%	100%
M101 Vancouverian Alpine Scrub, Forb Meadow & Grassland	3	30	57%	57%
M045 California Annual & Perennial Grassland	2	32	75%	81%
M043 California Chaparral	5	198	58%	92%
M044 California Coastal Scrub	1	50	90%	90%
M050 Southern Vancouverian Lowland Grassland & Shrubland	1	5	80%	80%
M168 Rocky Mountain-Vancouverian Subalpine & High Montane Mesic Grass & Forb Meadow	1	12	92%	92%
M091 Warm Interior Chaparral	1	50	26%	26%
M094 Cool Interior Chaparral	2	75	40%	61%
M058 Cool Pacific Coastal Beach, Dune & Bluff Vegetation	1	50	92%	92%
M082 Cool Semi-Desert Alkali-Saline Wetland	1	8	100%	100%

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 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, *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 <u>combination</u> <u>of multiple species</u> 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.