

# **Expert Attribution for Auto-Key Improvements (LANDFIRE) and Advancing Methods for integration with the revised US-National Vegetation Classification Standard: GeoArea 2E**

FINAL REPORT

Prepared by

NatureServe

For the NPS Vegetation Inventory Program & LANDFIRE

**June 2012**



*© Marion Reid. Glacier National Park, MT*



## LANDFIRE Improvements – Autokey Analysis

### **Contact information:**

Prepared by Patrick J. Comer, Chief Ecologist

Email: [pat\\_comer@natureserve.org](mailto: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 Anne Davidson of the National Gap Analysis Program, and Gwen Kittel, NatureServe Regional Vegetation Ecologist

# Table of Contents

<b>INTRODUCTION.....</b>	<b>4</b>
PROJECT GOALS .....	4
<i>Background on LANDFIRE Auto-keys.....</i>	<i>4</i>
METHODS.....	6
INTENDED PRODUCTS OF THIS EFFORT .....	10
<b>RESULTS .....</b>	<b>10</b>
GEOAREA 2E: NORTHERN ROCKY MOUNTAINS.....	10
COMPARISON OF AUTO-KEY AND EXPERT ASSIGNMENTS.....	11
EXPERT ASSIGNMENTS .....	16
IMPROVING THE AUTO-KEY PROCESS.....	18
ADAPTING AUTO-KEYS FOR NVC GROUPS, MACROGROUPS, AND DIVISIONS .....	19
<i>US-NVC Groups.....</i>	<i>19</i>
<i>US-NVC Macrogroups.....</i>	<i>22</i>
<i>US-NVC Divisions .....</i>	<i>23</i>
<b>DISCUSSION .....</b>	<b>23</b>
A.    QUALITY OF VEGETATION DATA.....	23
B.    CONSTRAINTS WITHIN SEQUENCE TABLE.....	24
C.    DEVELOPING AUTOMATED KEYS FOR LARGE GEOGRAPHIC AREAS .....	25
D.    COST/BENEFIT & EFFICIENCY .....	26
<b>RECOMMENDATIONS (DRAFT).....</b>	<b>26</b>

## *Figures*

Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in the coterminous U.S. ....	6
Figure 2. Modified LANDFIRE GeoAreas in the conterminous U.S. for use in this project. ....	7
Figure 3. Example of an image clip for one plot in GeoArea 4. ....	8
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.....	17

## *Tables*

Table 1. Categories & fields of data provided to expert during review process.....	8
Table 2. Under-sampled types within GeoArea 2E.....	11
Table 3. Summary of types with adequate samples where agreement between auto-key and expert was below 80%.....	12
Table 4. A selection of expert comments related to labeling sample plots for types where their confidence was reported as moderate or low.....	18
Table 5. Ecological Systems of GeoArea 2E that have complex relationships with NVC Groups.....	20
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.....	22

### 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 Improvements – Autokey Analysis

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:

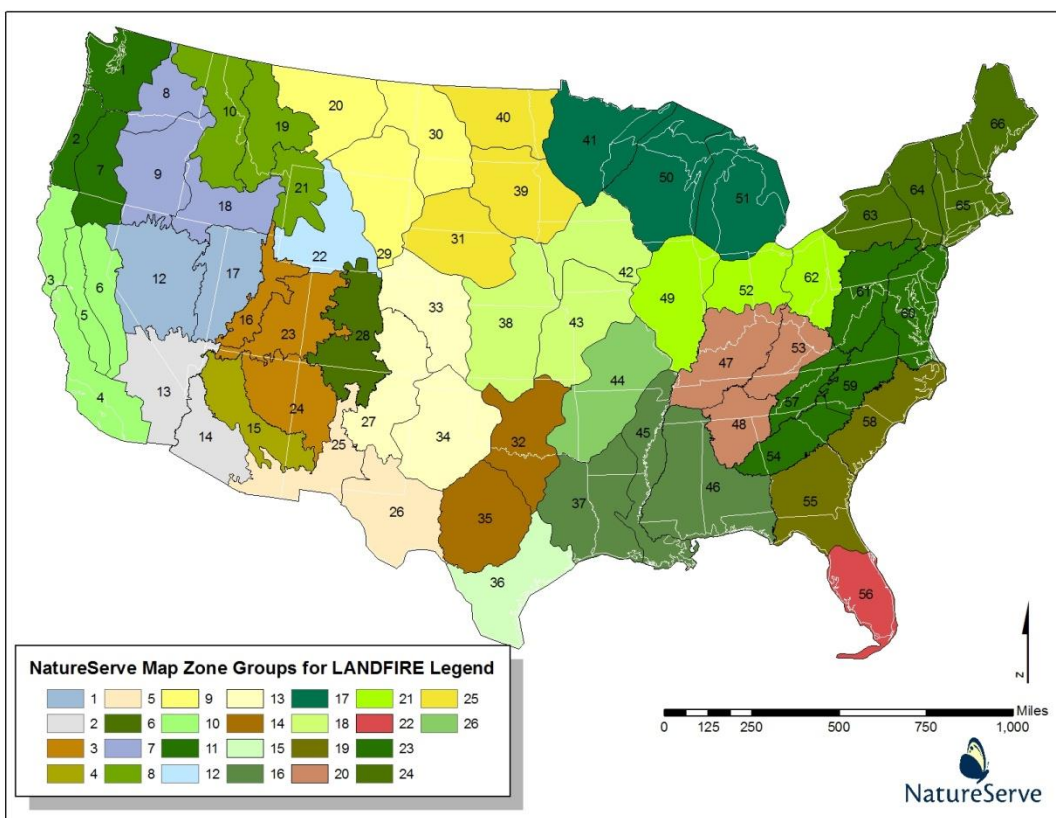
1. Each LANDFIRE sequence table was designed to efficiently automate keying of thousands to 10's of thousands geo-referenced vegetation samples to the LANDFIRE map units, which included both Ecological Systems for the 'natural' portions of the landscape, and a variety of land use or land cover classes 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

## LANDFIRE Improvements – Autokey Analysis

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

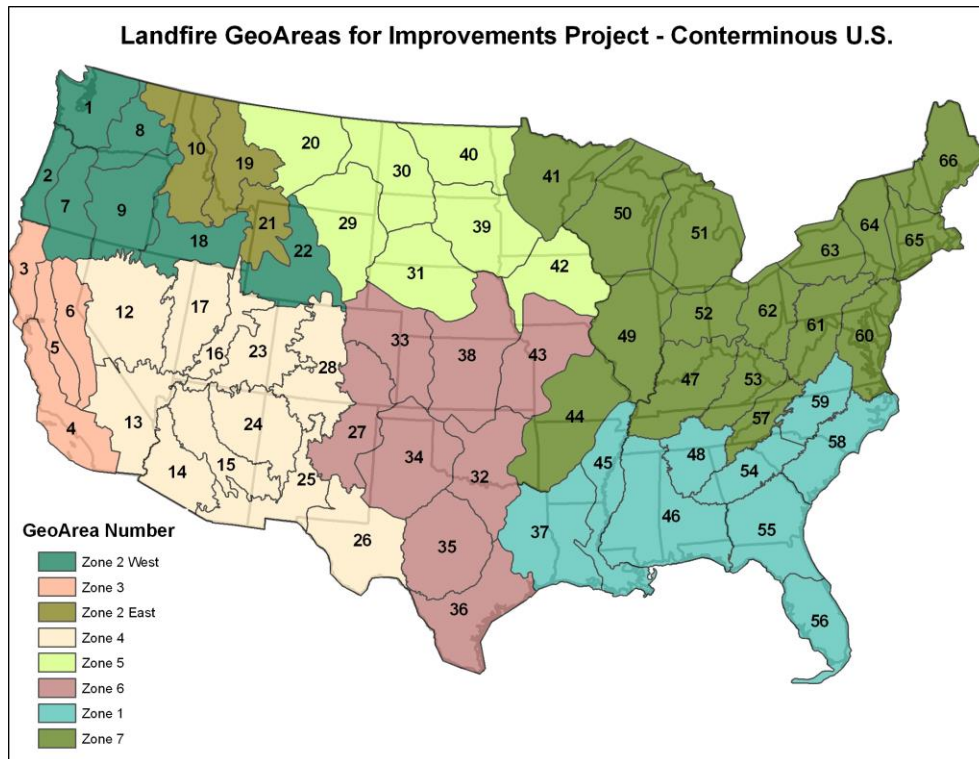
For the LANDFIRE effort, both dichotomous field keys and auto-keys were developed for map legend classes and organized in a series of 17 map zone groupings that spanned the nation. For ongoing maintenance of national map products, the map zone groups have been further aggregated by LANDFIRE into larger 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.



## LANDFIRE Improvements – Autokey Analysis

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 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

## LANDFIRE Improvements – Autokey Analysis

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:

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

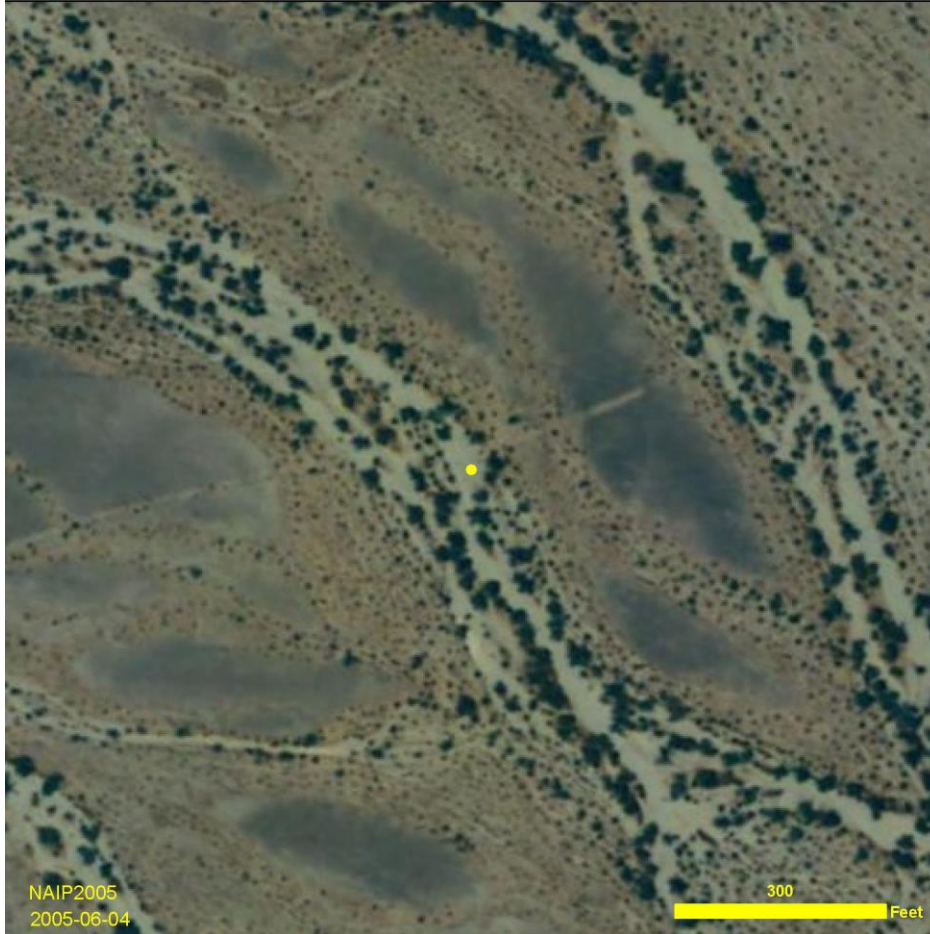
Table 1 is a list of the general categories of data that were extracted from the LFRDB and provided to the experts for use in their review. After much discussion, it was also determined to provide a remotely-sensed image clip for each plot, as well as between 1 and 3 on-the-ground photos for the plot 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.

<b>Data category</b>	<b>Fields</b>	<b>Notes</b>
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 from 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 4.





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

## LANDFIRE Improvements – Autokey Analysis

- 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***

- 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).
- A report by each GeoArea detailing processes and results, specifically identifying how they made individual assignments.
- A report that documents procedures and data elements to improve the auto-key process in each GeoArea.
- A report that documents technical procedures to adapt auto-keys for labeling NVCS group, Macrogroup, and Division concepts.
- Full data sets with independent assignments for each GeoArea in standard LFRDB format.
- A single overall report with recommendations for all GeoAreas, including commonalities and unique issues.

### **Results**

The following results for GeoArea 2E are organized according to these primary product deliverable categories:

- 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).
- A report by each GeoArea detailing processes and results, specifically identifying how they made individual assignments.
- A report that documents procedures and data elements to improve the auto-key process in each GeoArea.
- A report that documents technical procedures to adapt auto-keys for labeling NVCS group, macrogroup, and division concepts.

### ***GeoArea 2E: Northern Rocky Mountains***

GeoArea 2E encompasses 3 map zones (Figure 2): the Northwestern Rocky Mountains (10), Northern Rocky Mountains (19), and Middle Rocky Mountains (21). These map zones were originally clustered for purposes of designing and implementing auto-keys (**Error! Reference source not found.**). The total number of plots in this GeoArea analysis was 1,971. A total of 40 natural ecological system types were assigned to a total of 1,532 plots by the auto-keys. A total of 49 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).

## LANDFIRE Improvements – Autokey Analysis

An additional 9 types were assigned by the auto-key but were not assigned by experts:

- Columbia Basin Palouse Prairie
- Columbia Plateau Scabland Shrubland
- Great Basin Xeric Mixed Sagebrush Shrubland
- Rocky Mountain Bigtooth Maple Ravine Woodland
- Inter-Mountain Basins Montane Riparian Systems
- Inter-Mountain Basins Sparsely Vegetated Systems
- Rocky Mountain Alpine/Montane Sparsely Vegetated Systems
- Rocky Mountain Montane Riparian Systems
- Rocky Mountain Subalpine/Upper Montane Riparian Systems
- 

Five of these types are the aggregated types used by the LANDFIRE but the first four are Ecological Systems that could have been selected by the experts. The concepts and descriptions for these types may need to be revisited, with the likelihood of occurrence in the GeoArea reevaluated. If the type is still expected to occur additional guidance on how to apply the system relative to this GeoArea may need to be incorporated into the descriptions.

•

### ***Comparison of Auto-key and Expert Assignments***

- 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 40 natural types assigned labels by the auto-keys, 8 types (20%) 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. Great Basin Xeric Mixed Sagebrush Shrubland, Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland , Columbia Basin Foothill and Canyon Dry Grassland, and Rocky Mountain Bigtooth Maple Ravine Woodland), others are within this range but are relatively rare types (Rocky Mountain Alpine Fell-Field, Rocky Mountain Alpine Turf, Inter-Mountain Basins Greasewood Flat). The Columbia Basin Palouse Prairie was historically extensive but the vast majority of its historic range has been converted to agriculture and representative plots sites are very difficult to find.

**Table 2. Under-sampled types within GeoArea 2E.**

<b>EVTCode</b>	<b>EVT Name</b>	<b>System elcode</b>	<b>Total Plots</b>
2153	Inter-Mountain Basins Greasewood Flat	CES304.780	8
2143	Rocky Mountain Alpine Fell-Field	CES306.811	6
2012	Rocky Mountain Bigtooth Maple Ravine Woodland	CES306.814	6
2144	Rocky Mountain Alpine Turf	CES306.816	5
2079	Great Basin Xeric Mixed Sagebrush Shrubland	CES304.774	4
2134	Columbia Basin Foothill and Canyon Dry Grassland	CES304.993	4
2142	Columbia Basin Palouse Prairie	CES304.792	1
2057	Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland	CES306.819	1

## LANDFIRE Improvements – Autokey Analysis

**Of the 40 types, 6 had >80% agreement between expert and auto-key assignments.** 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 26, or 72% 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.

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

EVT Code	EVT Name	System Elcode	Total Plots	Plots with Expert Matches				
				Total	%	High conf	Med conf	Low conf
2106	Northern Rocky Mountain Montane-Foothill Deciduous Shrubland	CES306.994	50	39	78%	24	13	2
2124	Columbia Plateau Low Sagebrush Steppe	CES304.080	50	38	76%	33	4	1
2047	Northern Rocky Mountain Mesic Montane Mixed Conifer Forest	CES306.802	51	37	73%	30	7	0
2011	Rocky Mountain Aspen Forest and Woodland	CES306.813	50	35	70%	25	8	2
2080	Inter-Mountain Basins Big Sagebrush Shrubland	CES304.777	50	34	68%	16	18	0
2139	Northern Rocky Mountain Lower Montane-Foothill-Valley Grassland	CES306.040	50	28	56%	19	9	0
2168	Northern Rocky Mountain Avalanche Chute Shrubland	CES306.801	50	26	52%	9	16	1
2126	Inter-Mountain Basins Montane Sagebrush Steppe	CES304.785	50	24	48%	20	4	0
2045	Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest	CES306.805	50	21	42%	14	6	1
2169	Northern Rocky Mountain Subalpine Deciduous Shrubland	CES306.961	50	21	42%	10	10	1
2140	Northern Rocky Mountain Subalpine-Upper Montane Grassland	CES306.806	50	20	40%	17	3	0
2056	Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodland	CES306.830	48	18	38%	10	8	0
2061	Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland	CES304.776	50	18	36%	9	7	2
2125	Inter-Mountain Basins Big Sagebrush Steppe	CES304.778	55	16	29%	9	7	0

## LANDFIRE Improvements – Autokey Analysis

EVT Code	EVT Name	System Elcode	Total Plots	Plots with Expert Matches				
				Total	%	High conf	Med conf	Low conf
2145	Rocky Mountain Subalpine-Montane Mesic Meadow	CES306.829	98	27	28%	14	10	3
2115	Inter-Mountain Basins Juniper Savanna	CES304.782	11	3	27%	1	2	0
2046	Northern Rocky Mountain Subalpine Woodland and Parkland	CES306.807	50	11	22%	10	1	0
2049	Rocky Mountain Foothill Limber Pine-Juniper Woodland	CES306.955	50	11	22%	5	6	0
2127	Inter-Mountain Basins Semi-Desert Shrub-Steppe	CES304.788	50	7	14%	4	1	2
2161	Northern Rocky Mountain Conifer Swamp	CES306.803	50	7	14%	3	3	1
2166	Middle Rocky Mountain Montane Douglas-fir Forest and Woodland	CES306.959	50	7	14%	6	1	0
2165	Northern Rocky Mountain Foothill Conifer Wooded Steppe	CES306.958	50	0	0%	0	0	0
2167	Rocky Mountain Poor-Site Lodgepole Pine Forest	CES306.960	48	0	0%	0	0	0
2009	Northwestern Great Plains Aspen Forest and Parkland	CES303.681	36	0	0%	0	0	0
2123	Columbia Plateau Steppe and Grassland	CES304.083	19	0	0%	0	0	0
2065	Columbia Plateau Scabland Shrubland	CES304.770	14	0	0%	0	0	0

Analysis of the contingency table (in the Results Workbook for GA 2E) 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. Three types of disagreement between somewhat floristically similar types in the plot assignments became apparent through this analysis: where change occurs along an elevation gradient or along a moisture gradient and where types have different geographic ranges.

- **Confusion amongst systems determined along an elevation gradient reduced agreement** - In the mountainous Northern Rocky Mountain GeoArea there are a number of Ecological Systems that grade into another somewhat similar, but higher, elevation system.
  - Northern Rocky Mountain Lower Montane, Foothill and Valley Grassland and
  - Northern Rocky Mountain Subalpine-Upper Montane Grassland
  - 
  - Northern Rocky Mountain Montane-Foothill Deciduous Shrubland and
  - Northern Rocky Mountain Subalpine Deciduous Shrubland
  - 
  - Inter-Mountain Basins Big Sagebrush Steppe and
  - Inter-Mountain Basin Montane Sagebrush Steppe
  -

## LANDFIRE Improvements – Autokey Analysis

- Rocky Mountain Limber Pine-Juniper Woodland and
- Rocky Mountain Subalpine Montane Limber-Bristlecone Pine Woodland

Confusion between some of these pairs of systems is very high and reduced the agreement between auto-key and expert assignments, for example including plots assigned to the Northern Rocky Mountain Subalpine-Upper Montane Grassland to the totals for the Northern Rocky Mountain Lower Montane Foothill and Valley Grassland increases agreement between the autokey and expert assignments from 56 to 88 percent. Some assignment overlap between these types is to be expected and is likely unavoidable due to their occurrence along an ecological gradient and the many species that they share.

The descriptions for most of these types are already relatively detailed with extensive lists of characteristic understory species. The problem when classifying plots to these systems arises when a plot has some species that are characteristic of each system. In this case the expert weighs the coverage of each species and attempts to determine which of the two system descriptions the plot fits most closely. The addition of elevation variables to the sequence table process should help to improve classification of these types and the development of more nuanced rules determining which species presence (or prevalence) trumps the presence of other more generalist species would help provide more consistency in how these systems are assigned through either process.

While most of the systems where confusion along an elevation gradient are well described, distinctions between the Rocky Mountain Limber Pine Juniper Woodland and the Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland are not as clear as they could be. Better description of the range of the Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland outside the zone of occurrence of Bristlecone Pine would also be helpful. Experts assigned the Rocky Mountain Subalpine-Montane Limber-Bristlecone Pine Woodland to 14 plots in this GeoArea while the AutoKey only used it once. This may indicate some confusion on how common this system should be in the GeoArea.

- 
- **Confusion amongst systems determined along a moisture gradient reduced agreement-**  
Evaluation of the contingency table indicates that confusion between drier systems and a similar but more mesic system was also a factor in reducing agreement between the auto-key and expert assignments. Examples of these pairs of systems include
  - Northern Rocky Mountain Mesic Montane Mixed Conifer Forest
  - Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest
- 
- Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland
- Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodlands
- 
- Northern Rocky Mountain Subalpine-Upper Montane Grassland
- Rocky Mountain Subalpine-Montane Mesic Meadow

As with the elevation gradient systems, some assignment overlap between these types is to be expected. Again the development of clearer rules on how to handle plots that contain elements of both the drier and more mesic systems would also be helpful in increasing system assignment consistency for these types of systems. Additionally, guidance on the likely relative abundance of one system compared to the other in a pair would be helpful in making a call. For example the auto-key only slightly favored the assignment of the Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland to the Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodlands (44–38 plots). However, the experts assigned the Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland much more



## LANDFIRE Improvements – Autokey Analysis

frequently than the more mesic system (60-21 plots). At least from the point of view of one expert, this was done consciously with the understanding that in this GeoArea the Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland would be the more prevalent system.

- **Confusion amongst floristically similar systems with different ranges reduced agreement-** Another type of disagreement between the auto-key and expert assignments appears to have arisen due to the application of different geographic ranges to determine the assignment of two somewhat floristically similar systems. Examples of this type of confusion included.
  - Northwestern Great Plains Aspen Forest and Woodlands
  - Rocky Mountain Aspen Forest and Woodland
  - 
  - Middle Rocky Mountain Douglas-Fir Forest and Woodland
  - Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest
  - 
  - Intermountain Basins Juniper Savanna
  - Rocky Mountain Foothill Limber Pine-Juniper Woodland

One example of this type of confusion is that 23 of the 63 plots assigned to the Rocky Mountain Aspen Forest and Woodland by the experts were assigned to the Northwestern Great Plains Aspen and Forest and Parkland by the auto key and both of the plots assigned by the experts to the Northwestern Great Plains Aspen Forest and Woodland were assigned by the Auto key to the Rocky Mountain Aspen Forest and Woodland.

Another example is the Middle Rocky Mountain Montane Douglas-fir Forest and Woodland. This system was used much more extensively by the auto-key than by the experts (50-16), a difference that can be at least partially attributed to the different ranges applied in the two attribution processes. The Middle Rocky Mountain Montane Douglas-fir Forest and Woodland is a forest or woodland system dominated by *Pseudotsuga menziesii* that occurs outside of the range of true firs such as *Abies concolor*, *Abies grandis*, and *Abies lasiocarpa*. The auto-key process assigned plots across the GeoArea to this system if they were dominated by *Pseudotsuga menziesii* (or closely associated species) and did not contain fir (or closely associated species). This tended to work well for plots that contained sufficiently detailed lists of species but resulted in the inclusion of some *Pseudotsuga menziesii* dominated plots occurring within the true fir zone that did not include fir species in their species lists. The experts were able to use the FS sections to identify these plots occurring within the true fir zone systems and assign them to more suitable systems.

Better application of reviewed and established ranges documenting where each system occurs and the other does not would greatly reduce this type of disagreement. In areas where two similar systems both occur, the development of more detailed guidance on each systems occurrence and the use of elevation, soil and other non-floristic variable to make plot assignments would increase plot assignment accuracy.

- Other points of interest in contingency table evaluation
- **Inter-Mountain Basins Big Sagebrush Shrubland and Intermountain Basins Big Sagebrush Steppe-** There was considerable disagreement between expert and auto-key plots assigned to Inter-Mountain Basins Big Sagebrush Shrubland and Inter-Mountain Big Sagebrush Steppe. Developing greater clarification on how to distinguish these two systems should help to reduce this disagreement.

## LANDFIRE Improvements – Autokey Analysis

- **Rocky Mountain Poor-Site Lodgepole Pine and Rocky Mountain Lodgepole Pine Forest-** Forty-seven of the 48 plots assigned to the Rocky Mountain Poor-Site Lodgepole Pine system by the auto-key were assigned to the more generic Rocky Mountain Lodgepole Pine Forest by the experts. The Rocky Mountain Poor-Site Lodgepole pine system occurs on very specific soil types and because soil information was not available the experts may not have felt comfortable making this distinction. Currently the Rocky Mountain Poor-Site Lodgepole system description focuses on where concentrations of this system can be expected to occur in Oregon, describing occurrences in this GeoArea 2E as patchy. More details on where concentrations of this system may occur in this GeoArea would aid in the assignment of plots to this system by either method.
- **Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland-** Only thirty-six percent of the auto-key plots assigned this system were assigned it by the experts, while 86% of the plots assigned to this system by the experts were also assigned by the auto-key. The majority of the extra plots assigned this system by the auto-key were assigned to a conifer forest type by the experts. This indicates that the experts were looking for a higher coverage of *Populus tremuloides* before assigning this system and/or a more restricted range than the auto-key used. This is an example of a system where the concept of its geographic range and composition may not be consistently applied by the auto-keys or experts, and requires review.
- **Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest –** The experts assigned plots to this system much more often than the auto-key (132-46). Some of this difference can be explained by the previously discussed differences in the application of the Middle Rocky Mountain Montane Douglas-fir Forest and Woodland but much of it cannot. Expert confidence in assigning this system was generally high with 83% high 15% moderate and 2% low. Because this is a widespread system in this GeoArea further analysis of the discrepancy may be warranted before another classification effort is initiated.

### Expert Assignments

- 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 2E had nearly 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 *Picea engelmannii*), 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

## LANDFIRE Improvements – Autokey Analysis

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.

As an example, in GeoArea 2E spruce and fir species may occur in a large variety of ecological systems including Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland, Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodland, Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest, Northern Rocky Mountain Mesic Montane Mixed Conifer Forest, Northern Rocky Mountain Subalpine Woodland and Parkland, Rocky Mountain Lodgepole Pine Forest, Northern Rocky Mountain Conifer Swamp, and Rocky Mountain Subalpine-Montane Riparian Woodland. The experts must select the best of these choices using information on a sites elevation, slope, species dominance, tree canopy cover, presence of other tree species, mesic or xeric understory species, photographs, hydrology and soil and geologic information if available

**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.**

EventID	Reviewer	Review Date	LFZon	USFS Subsec	DomLifeform	DomSp	DomSpLife	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 Gramin	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		

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

- review of the image clip for the context of the plot,

## LANDFIRE Improvements – Autokey Analysis

- review of where the plot was located geographically (USFS Subsections provide local scale geographic location), to distinguish Columbia Plateau from Northern Rocky Mountain systems for example..
- consideration of topographic setting (e.g. north-facing slopes at lower elevations could support ponderosa pine woodlands),
- consideration of any available height data for the plot (e.g. were the ponderosa pines all tall, apparently mature trees; or were they short),
- careful consideration of the full floristic composition of the plot and cover for each species.
- 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).

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***

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

Of the 49 types assigned to plots by experts, 13 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 36 types, the numbers of samples labeled to a given type ranged from 132 (Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest) down to 12 (Northern Rocky Mountain Conifer Swamp). For 41 (87%) of these types, experts reported moderate confidence in their labels for at least 20% of the type's plots. 2 types indicated low confidence for at least 20% of the type's plots. These statistics are listed in the Appendix. A small sampling of expert comments related to moderate or low confidence plots are included in Table 4.

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

Type Name	Expert Comment
Rocky Mountain Cliff, Canyon and Massive Bedrock	Steep borderline sparsely vegetated site with 14% total vegetation cover, with nothing dominant
Northern Rocky Mountain Montane-Foothill Deciduous Shrubland	Plot is sparsely vegetated for this system
Northern Rocky Mountain Lower Montane, Foothill and Valley Grassland	Very limited species information reduced confidence
Rocky Mountain Subalpine Mesic-Wet Spruce-Fir Forest and Woodland	Hydrology information is not provided, if plot is riparian then assignment should be CES306.833
Northern Rocky Mountain Avalanche Chute Shrubland	Cannot tell snowpack conditions from information given

## LANDFIRE Improvements – Autokey Analysis

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 those labeled by auto-keys to aggregates of multiple ecological system types. This was because LANDFIRE had mapping objectives focused on uplands where fire regimes are prevalent. That meant that many individual wetland and sparsely-vegetated ecological system types were not treated within the auto-keys. Expert labeling of these samples, however, provides an indication of the feasibility of their inclusion in updated auto-keys. Of 219 samples, experts were able to assign 181 (83%) to an individual ecological system type; a total of 31 individual ecological system types were assigned to these samples. This result indicates the potential for inclusion of these types within subsequent mapping efforts. We cannot yet comment on the issues associated including these types within future regional auto-keys, but this appears to be an issue worthy of exploration.

Another class of samples did not contain enough information for the auto-keys to assign a system or system aggregate; these samples were labeled with broad "unclassified" types, such as "Unclassified Grassland" or "None". Of 220 samples, experts were able to assign 172 (78%) to an individual ecological system type; a total of 31 individual ecological system types were assigned to these samples.

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

- 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 68 terrestrial ecological system types could occur. Of these, 26 have a practical 1:1 relationship with NVC Group concepts, and 25 nest cleanly within 15 NVC Group concepts (1:many group:system relationship), for a total of 51 or 75% 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.

## LANDFIRE Improvements – Autokey Analysis

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, 16 (24%) ecological system types have a more complex relationship with NVC Group concepts (Table 5). A review of these more complex ecological system to NVC Group relationships indicates that, while revisions to the autokey would be necessary to use the NVC Groups for plot assignments the changes required for GeoArea 2E do not appear to be prohibitively substantive and would likely improve the plot classification process in this GeoArea. The simplification of the geographic modifiers from the riparian types would help to clear up some confusion in the classification of these types that often share many species. The addition of a riparian forest Group would help to differentiate two rather distinct habitats that have been previously been combined (i.e. shrubland and forest). The splitting of the Northern Rocky Mountain Avalanche Chute Shrubland into a more general high elevation shrubland Group and a riparian Group should help to make the classification of this type less difficult and error prone. Similarly the change from the North Pacific Avalanche Chute Shrubland to a more general Vancouverian Alder - Salmonberry - Willow Shrubland Group should be an improvement as accurate classification of the Avalanche Chute Shrubland systems typically required expert photo review and the changes made moving to the NVC Group should reduce the need for this time intensive review. Splitting the Northern Rocky Mountain Mesic Montane Mixed Conifer Forest into two NVC Groups one that recognizes the unique Western Red-cedar - Western Hemlock Forest is a welcome change that should be relatively easy to accomplish with minor revisions to the auto-key. The inability to distinguish these biologically important forest types from the more common mesic grand-fir forests is a common criticism of the current land cover maps for the region. Breaking the Rocky Mountain Alpine-Montane Wet Meadow into two Groups distinguished by elevation also seems beneficial. Currently this system is used to map a very wide variety of wet alpine and montane habitats and more definition in these areas seems warranted.

One area of concern is the combination of many of the sagebrush systems into only two NVC-Groups. [Note: there are other sagebrush Groups for the dwarf-sages, such as black sage, low sage; and montane sagebrush steppe remains as a separate NVC Group]. While assigning sagebrush plots to systems is difficult and agreement between auto key and experts assignments was often low for these types, they represent complex habitats and caution should be used to avoid oversimplification in these types. While the geographic modifiers associated with the sagebrush systems can cause confusion, usually because it is difficult to decide where one type ends and the other begins, many reviewers of the land cover maps suggest that these projects require more detailed information on the sagebrush types than is currently provided. We should be careful to make sure the NVC-Groups provide an adequate level of definition for these diverse types.

**Table 5. Ecological Systems of GeoArea 2E 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
Columbia Basin 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
Columbia Plateau Silver Sagebrush Seasonally Flooded Shrub-Steppe	G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland	1	5



## LANDFIRE Improvements – Autokey Analysis

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
Northern Rocky Mountain Lower Montane Riparian Woodland and Shrubland	G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland	2	5
	G503 Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest	2	4
Rocky Mountain Lower Montane-Foothill Riparian Woodland and Shrubland	G526 Rocky Mountain & Great Basin Lowland & Foothill Riparian & Seep Shrubland	2	5
	G503 Rocky Mountain & Great Basin Lowland & Foothill Riparian Forest	2	4
Columbia Basin Palouse Prairie	G273 Northern Rocky Mountain Lower Montane, Foothill & Valley Grassland	2	2
	G275 Northern Rocky Mountain Montane-Foothill Mesic Deciduous Shrubland	2	2
Northern Rocky Mountain Lower Montane, Foothill and Valley Grassland	G273 Northern Rocky Mountain Lower Montane, Foothill & Valley Grassland	1	2
Northern Rocky Mountain Montane-Foothill Deciduous Shrubland	G272 Northern Rocky Mountain Montane-Foothill Dry Deciduous Shrubland	2	1
	G275 Northern Rocky Mountain Montane-Foothill Mesic Deciduous Shrubland	2	2
Columbia Plateau Steppe and Grassland	G302 Intermountain Mesic Tall Sagebrush Shrubland & Steppe	1	3
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
Northern Rocky Mountain Avalanche Chute Shrubland	G305 Northern Rocky Mountain High Montane Mesic Shrubland [Provisional]	2	3
	G504 Rocky Mountain & Great Basin Montane Alder & Birch Riparian Shrubland	2	1
Northern Rocky Mountain Subalpine Deciduous Shrubland	G305 Northern Rocky Mountain High Montane Mesic Shrubland [Provisional]	1	3
Northern Rocky Mountain Mesic Montane Mixed Conifer Forest	G217 Central Rocky Mountain Interior Western Red-cedar - Western Hemlock Forest	2	2
	G211 Central Rocky Mountain Mesic Grand Fir - Douglas-fir Forest	2	2
Rocky Mountain Alpine-Montane Wet Meadow	G521 Vancouverian & Rocky Mountain Montane Wet Meadow	2	2

## LANDFIRE Improvements – Autokey Analysis

	G520 Vancouverian & Rocky Mountain Subalpine Snowbed, Wet Meadow & Dwarf-Shrubland	2	1
--	---	---	---

### 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 in the Results Workbook, and also compared the percent of expert-autokey matches at the system level versus the Macrogroup level (Table 6).

There are 14 US-NVC Macrogroups represented among natural mapped classes in this GeoArea. Rolling up to the Macrogroup improves auto-key and expert agreement for six of these 14 Macrogroups when compared to agreement for the ecological system level. For five of these Macrogroups that show increased levels of agreement, improvement is by 20% or more.

While the results in Table 6 suggest rolling up to Macrogroup would yield a higher level of agreement, 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 Northern Rocky Mountain Montane & Foothill Grassland & Shrubland contains 7 ecological systems and encompasses most (in terms of map area) of the non-forest habitats in the mountainous portions of the GeoArea. These types range from very dry foothill grasslands to very mesic sub-alpine shrublands and combining them into one class would not be advantageous for many applications.

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

Macrogroup	# auto-keyed systems	# plots	% expert matches at system level	% expert matches at MG level
M093 Great Basin Saltbrush Scrub	1	18	83%	83%
M169 Great Basin & Intermountain Tall Sagebrush Shrubland & Steppe	5	178	42%	79%
M170 Great Basin & Intermountain Dwarf Sage Shrubland & Steppe	2	64	59%	59%
M171 Great Basin & Intermountain Dry Shrubland & Grassland	1	50	14%	14%
M151 Northern Great Plains Woodland	1	36	0%	0%
M017 Northern Rocky Mountain Lower Montane & Foothill Forest	6	301	39%	83%
M020 Rocky Mountain Subalpine & High Montane Conifer Forest	8	347	49%	84%
M022 Southern Rocky Mountain Lower Montane Forest	1	6	0%	0%
M026 Intermountain Singleleaf Pinyon - Western Juniper Woodland	2	61	85%	85%
M034 Rocky Mountain & Great Basin Flooded & Swamp Forest	1	50	14%	50%
M099 Rocky Mountain Alpine Scrub, Forb Meadow & Grassland	3	60	73%	78%
M048 Northern Rocky Mountain Montane & Foothill Grassland & Shrubland	7	255	54%	77%

## LANDFIRE Improvements – Autokey Analysis

M168 Rocky Mountain-Vancouverian Subalpine & High Montane Mesic Grass & Forb Meadow	1	98	28%	28%
M082 Cool Semi-Desert Alkali-Saline Wetland	1	8	88%	88%

### 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 knowledge 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

## LANDFIRE Improvements – Autokey Analysis

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

## LANDFIRE Improvements – Autokey Analysis

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

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

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

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

### ***C. Developing automated keys for large geographic areas***

Each sequence table was constructed to work across relatively large geographic areas, on the order of 2-5 USFS Sections (**Error! Reference source not found.**). Hence each sequence table/auto-key included 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

## LANDFIRE Improvements – Autokey Analysis

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)**

[Note: yellow highlighted comments below are from the Report writer].

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

Can we include a pre-processing step that makes sure the physiognomic class indicated by the vegetation in the plot is consistent with what is currently on the ground? This should help identify areas where the vegetation has changed significantly since the data was collected through logging or other



## LANDFIRE Improvements – Autokey Analysis

methods and should help identify plots that may be inaccurate (due to location errors, data collection errors, or misunderstanding of field data when applying the auto-keys)

Some systems, for example the Rocky Mountain Poor-Site Lodgepole Pine Forest may be benefited from a two step auto-key process. Run the plots through the auto-key and see which key to the Rocky Mountain Lodgepole Pine Forest. Then for appropriate FS Sections where poor site conditions are thought to occur pull in other variables such as soil information and expert review to identify the poor site plots.

When first considering the addition of variables like slope and elevation to the auto-key process it seems that the addition of these variables would be beneficial but with further consideration it is difficult to envision general rules using these variables that would not end up eliminating correct points from type classes. In order to work the rules would need to be fairly location specific and would therefore be time intensive to develop. Perhaps it would be better to include these variables in a check process which would flag plots that seem outside of the normal range for the system for additional review.

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

It seem like the use of the FS Sections would help to improve auto-key accuracy in GeoArea 2E. There use would help to prevent points getting to assigned to systems outside their range of occurrence, for example Middle Rocky Mountain Montane Douglas-fir Forest and Woodland plots assigned adjacent to Northern Rocky Mountain Dry-Mesic Montane Mixed Conifer Forest in the Northern Rockies. The use of FS Sections would also help identify where two similar systems co-occur. In these areas additional variables could be pulled in to differentiate the two systems or additional expert review could be employed. Being able to more easily identify limited areas where two systems may be confused would help to conserve review resources and direct them to the most problematic areas.

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

It is important to make sure that when vegetation data is collected information on non-vegetative elements is also collected and considered. Obviously a site where aspen is the only plant species recorded and aspen covers 80% of the canopy is very different than a site where aspen is the only vegetation recorded but it provides 5% canopy coverage and the majority of the site is bedrock.

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

Expert review should be incorporated into the auto-key process where possible. Perhaps experts could review some percentage of all plots (5-20%) and look for patterns in where the auto-key may be mislabeling. Then the autokeys could be revised and improvement looked for. If improvement was not possible for certain systems or geographic areas could be flagged for more extensive review by experts. In GeoArea 2E systems that probably benefited the most from expert review were the alpine, wetland, riparian, bareground, avalanche, and subalpine parkland systems. These systems seem to present challenges to the auto-key that can often be resolved by experts through the review of photos and other context information.

## LANDFIRE Improvements – Autokey Analysis

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.