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

FINAL REPORT

Prepared by
NatureServe
For the NPS Vegetation Inventory Program & LANDFIRE

30 June 2012

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 Mark Hall, NatureServe Regional Vegetation Ecologist.

Table of Contents

Figures

Tables

Introduction

Comment [kls1]: As far as I can tell the only thing in the Intro/Methods sections that Mark updated is Fig. 2, so I updated everything else from latest CONUS version.

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.

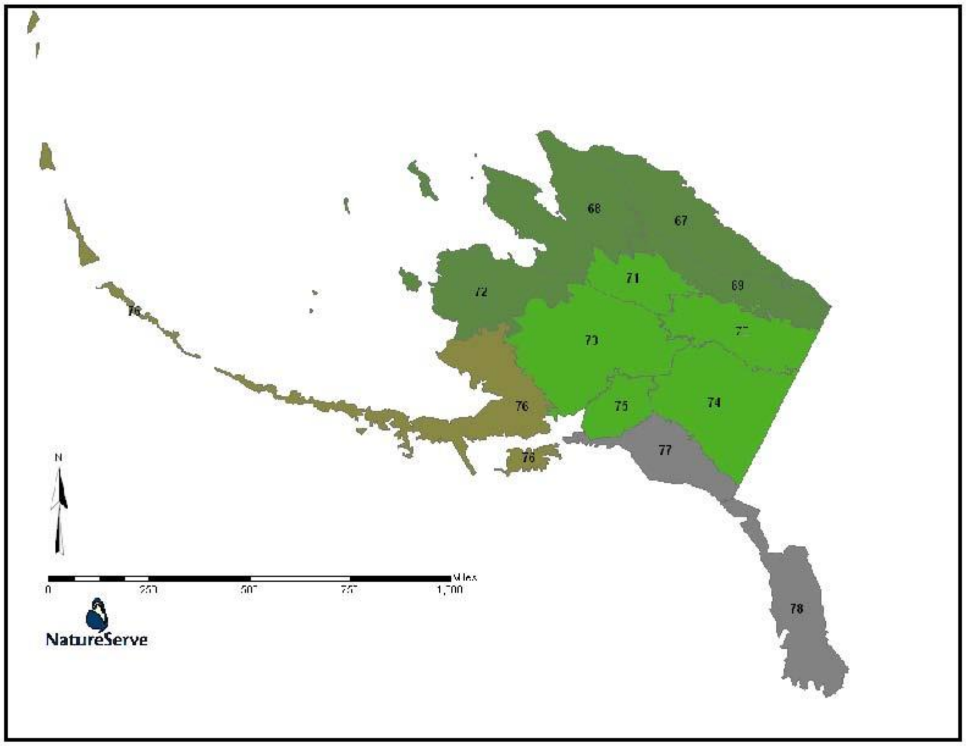
LANDFIRE's short-term needs, and long-term plans, required a repeatable methodology, consistently applied rules to categorize each reference sample, and documentation of the criteria applied. In essence, sequence tables codify the criteria and methods for keying geo-referenced vegetation data to a land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

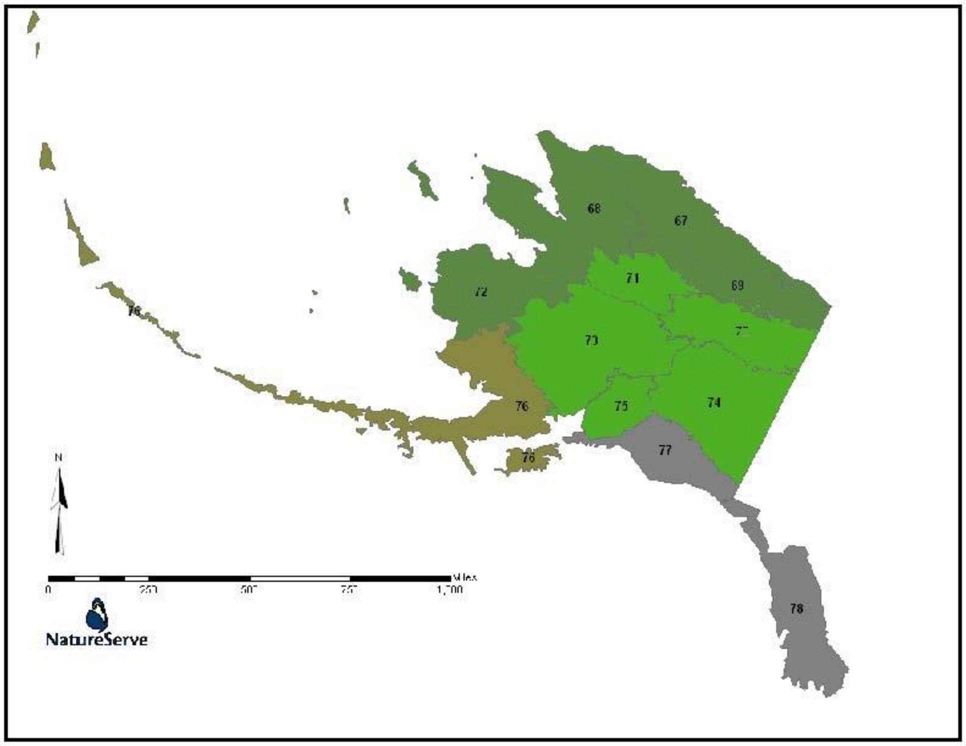
Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in Alaska.



Methods

For the LANDFIRE effort, both dichotomous field keys and auto-keys were developed for map legend classes and organized in a series of 4 map zone groupings for Alaska (maritime, boreal, arctic and Aleutian). 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, Alaska map zones are all one GeoArea (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 GeoArea in Alaska 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 plot selection for Alaska was modified slightly from the above procedure, because several ecological systems lacked plots all together, or had very low numbers of plots concentrated in a relatively small area. Some 73 ecological systems keyed by auto-keys had at least 20 plots for expert review; the other systems all had fewer than 20, and were excluded from selection for the expert review.

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

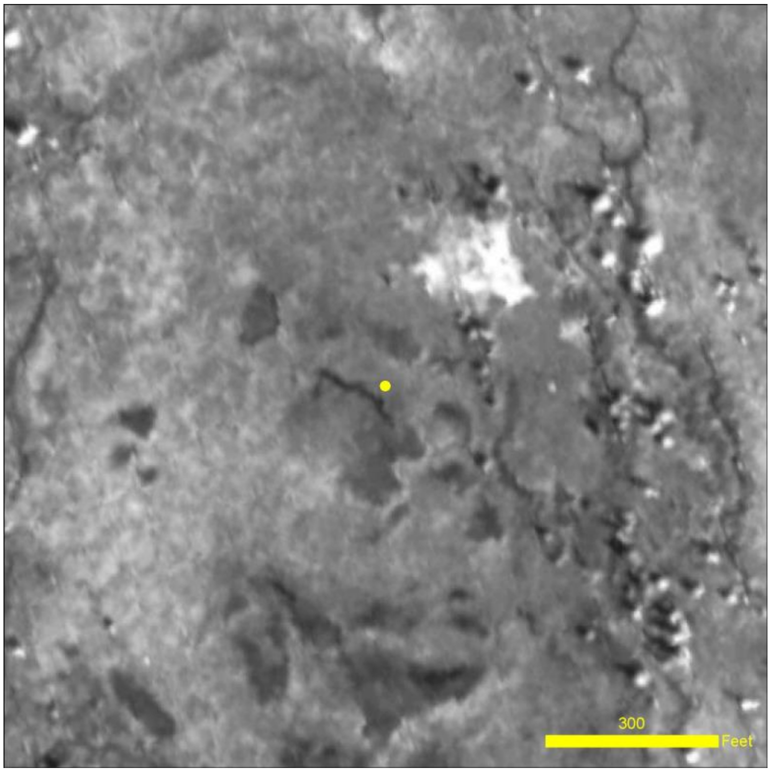
Table 1 is a list of the general categories of data that were extracted from the LFRDB and provided to the experts for use in their review. After much discussion, it was also determined to provide a remotely-sensed image clip for each plot, as well as between 1 and 3 on-the-ground photos for the plot 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). In Alaska, many of the image clips were blank (indicating no imagery for the area), or gray tones that were not very helpful. Figure 3 is an example of the later.

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)

Data category	Fields	Notes
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 8



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 8 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 8: Alaska

GeoArea 8 encompasses the entire state of Alaska (Map zones 67-78, Figure 2). This GeoArea includes a total of 12 map zones originally clustered into 4 groups for purposes of designing and implementing auto-keys. The total number of plots in this Geo Area analysis was 1,454. A total of 73 natural ecological system types were assigned to a total of 1,454 plots by the auto-keys. A total of 93 systems 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 12 types were assigned by the auto-key but were not assigned by experts. Those types included:

- Alaska Arctic Floodplain
- Alaska Arctic Large River Floodplain
- Alaska Sub-boreal Avalanche Slope Shrubland
- Alaskan Pacific Maritime Alpine Floodplain
- Alaskan Pacific Maritime Coastal Meadow and Slough-Levee
- Alaskan Pacific Maritime Mountain Hemlock Peatland
- Aleutian Shrub-Sedge Peatland
- Western North American Boreal Alpine Dwarf-Shrub Summit
- Western North American Boreal Alpine Floodplain
- Western North American Boreal Lowland Large River Floodplain Forest and Shrubland
- Western North American Boreal Montane Floodplain Forest and Shrubland
- Western North American Boreal Shrub and Herbaceous Floodplain Wetland

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 73 natural types assigned labels by the auto-keys, none were considered under-sampled (represented by <10 plots). A total of 10 types (or ~7% of 73 types) had >80% agreement between expert and auto-key assignments. All of these types had at least 20 sample plots. Expert self-assessment of confidence for these types were predominantly 'high' although Western North American Boreal Tussock Tundra had 5 plots which were considered to have 'moderate' confidence.

Table 2. Under-sampled types within GeoArea 8

There are no under-sampled types within GeoArea 8.

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 63, or nearly 86% 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 autokey were attributed to the Western North American Boreal White Spruce Forest were mistaken for Western North American Boreal White Spruce-Hardwood Forest. Another example was confusion between the Alaskan Pacific Maritime Periglacial Woodland and Shrubland and the other maritime forest systems, especially the Alaskan Pacific Maritime Sitka Spruce Forest and the Alaskan Pacific Maritime Floodplain Forest and Shrubland. All three of these have sitka spruce as a major component, but are distinguished by floodplain settings or by early successional status in areas of recent glacial melt. Clarification of floristics (between the floodplain type and the other 2), or how to distinguish primary successional areas for the periglacial type would help to improve the sequence table as well as expert understanding of the types.

Many 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
2178	North Pacific Hypermaritime Western Red-cedar-Western Hemlock Forest	CES204.842	20	15	75%	14	0	1
2611	Western North American Sub-boreal Mesic Bluejoint Meadow	CES105.114	20	14	70%	11	2	1
2648	Alaskan Pacific Maritime Mountain Hemlock Forest	CES204.142	20	14	70%	14	0	0
2644	Alaskan Pacific Maritime Sitka Spruce Forest	CES204.151	20	14	70%	11	1	2
2698	Alaska Arctic Wet Sedge Meadow	CES102.185	20	13	65%	5	5	3
2638	Alaska Arctic Mesic Alder Shrubland	CES104.168	20	13	65%	10	3	0
2604	Western North American Boreal Mesic Black Spruce Forest	CES105.107	20	13	65%	9	4	0
2650	Alaskan Pacific Maritime Periglacial Woodland and Shrubland	CES204.311	20	13	65%	13	0	0
2600	Western North American Boreal White Spruce Forest	CES105.104	20	12	60%	9	2	1

				Plots with Expert Matches				
EVT Code	**EVT Name**	**System elcode**	**Total Plots**	**Total**	**%**	**High conf**	**Med conf**	**Low conf**
2679	Alaska Sub-boreal White Spruce-Hardwood Forest	CES105.136	20	12	60%	9	3	0
2689	Alaska Arctic Non-Acidic Dryas Dwarf-Shrubland	CES104.174	20	11	55%	10	0	1
2635	Western North American Boreal Alpine Ericaceous Dwarf-Shrubland	CES105.133	20	11	55%	9	1	1
2677	Alaska Sub-boreal White-Lutz Spruce Forest and Woodland	CES105.102	20	10	50%	9	1	0
2639	Alaska Arctic Mesic-Wet Willow Shrubland	CES104.169	16	8	50%	5	3	0
2610	Western North American Boreal Mesic Scrub Birch-Willow Shrubland	CES105.113	20	9	45%	6	0	3
2682	Alaska Arctic Scrub Birch-Ericaceous Shrubland	CES104.170	20	8	40%	7	0	1
2603	Western North American Boreal White Spruce-Hardwood Forest	CES105.106	20	8	40%	8	0	0
2040	North Pacific Mesic Western Hemlock-Yellow-cedar Forest	CES204.843	20	8	40%	8	0	0
2693	Alaska Arctic Shrub-Tussock Tundra	CES102.180	20	7	35%	6	1	0
2705	Alaska Arctic Sedge Freshwater Marsh	CES102.184	20	7	35%	2	2	3
2701	Alaska Arctic Coastal Sedge-Dwarf-Shrubland	CES102.211	20	7	35%	4	3	0
2683	Alaska Arctic Mesic Sedge-Willow Tundra	CES102.187	20	6	30%	6	0	0
2712	Alaska Arctic Coastal Brackish Meadow	CES102.210	20	6	30%	6	0	0
2688	Alaska Arctic Acidic Dryas Dwarf-Shrubland	CES104.173	20	6	30%	6	0	0
2628	Western North American Boreal Low Shrub-Tussock Tundra	CES105.126	20	6	30%	6	0	0
2660	Alaskan Pacific Maritime Wet Low Shrubland	CES204.157	20	6	30%	5	0	1
2605	Western North American Boreal Mesic Birch-Aspen Forest	CES105.108	20	5	25%	5	0	0
2623	Western North American Boreal Black Spruce-Tamarack Fen	CES105.121	20	5	25%	3	2	0
2645	Alaska Sub-boreal and Maritime Alpine Mesic Herbaceous Meadow	CES204.145	20	5	25%	5	0	0

EVT Code	EVT Name	System elcode	Total Plots	Plots with Expert Matches: Total	%	High conf	Med conf	Low conf
2734	North Pacific Alpine and Subalpine Bedrock and Scree	CES204.853	20	5	25%	4	0	1
2622	Western North American Boreal Black Spruce Wet-Mesic Slope Woodland	CES105.120	20	4	20%	2	2	0
2699	Alaska Arctic Mesic Herbaceous Meadow	CES102.186	20	3	15%	3	0	0
2702	Alaska Arctic Wet Sedge-Sphagnum Peatland	CES102.200	20	3	15%	3	0	0
2717	Alaska Arctic Bedrock and Talus	CES102.228	20	3	15%	2	0	1
2601	Western North American Boreal Treeline White Spruce Woodland	CES105.137	20	3	15%	3	0	0
2690	Alaska Arctic Dwarf-Shrubland	CES104.175	19	2	11%	0	2	0
2700	Alaska Arctic Polygonal Ground Mesic Shrub Tundra	CES102.206	20	2	10%	2	0	0
2651	Aleutian Mesic Herbaceous Meadow	CES105.232	20	2	10%	2	0	0
2643	Alaskan Pacific Maritime Alpine Dwarf-Shrubland	CES204.310	20	2	10%	2	0	0
2621	Western North American Boreal Black Spruce Dwarf-Tree Peatland	CES105.139	21	2	10%	0	1	1
2684	Alaska Arctic Mesic Sedge-Dryas Tundra	CES102.199	20	1	5%	1	0	0
2703	Alaska Arctic Dwarf-Shrub-Sphagnum Peatland	CES102.201	20	1	5%	0	1	0
2707	Alaska Arctic Polygonal Ground Tussock Tundra	CES102.204	20	1	5%	1	0	0
2691	Alaska Arctic Acidic Dwarf-Shrub Lichen Tundra	CES104.177	20	1	5%	1	0	0
2606	Western North American Boreal Dry Aspen-Steppe Bluff	CES105.109	20	1	5%	0	0	1
2618	Western North American Boreal Herbaceous Fen	CES105.119	20	1	5%	1	0	0
2655	Alaskan Pacific Maritime Floodplain Forest and Shrubland	CES204.154	20	1	5%	1	0	0
2680	Alaskan Pacific Maritime Avalanche Slope Shrubland	CES204.162	20	1	5%	1	0	0
2681	Alaskan Pacific Maritime Poorly Drained Conifer Woodland	CES204.315	20	1	5%	1	0	0
2608	Alaska Sub-Boreal Avalanche Slope Shrubland	CES105.111	21	0	0%	0	0	0

				Plots with Expert Matches				
EVT Code	**EVT Name**	**System elcode**	**Total Plots**	**Total**	**%**	**High conf**	**Med conf**	**Low conf**
2714	Alaska Arctic Large River Floodplain	CES102.213	20	0	0%	0	0	0
2715	Alaska Arctic Floodplain	CES102.227	20	0	0%	0	0	0
2615	Western North American Boreal Lowland Large River Floodplain Forest and Shrubland	CES105.117	20	0	0%	0	0	0
2617	Western North American Boreal Shrub and Herbaceous Floodplain Wetland	CES105.118	20	0	0%	0	0	0
2626	Western North American Boreal Wet Meadow	CES105.124	20	0	0%	0	0	0
2631	Western North American Boreal Alpine Dwarf-Shrub Summit	CES105.129	20	0	0%	0	0	0
2620	Western North American Boreal Low Shrub Peatland	CES105.140	20	0	0%	0	0	0
2614	Western North American Boreal Montane Floodplain Forest and Shrubland	CES105.141	20	0	0%	0	0	0
2647	Aleutian Shrub-Sedge Peatland	CES105.238	20	0	0%	0	0	0
2659	Alaskan Pacific Maritime Mountain Hemlock Peatland	CES204.156	20	0	0%	0	0	0
2665	Alaskan Pacific Maritime Coastal Meadow and Slough-Levee	CES204.159	20	0	0%	0	0	0
2676	Alaskan Pacific Maritime Alpine Floodplain	CES204.161	20	0	0%	0	0	0
2637	Western North American Boreal Alpine Floodplain	CES105.135	17	0	0%	0	0	0

Analysis of the contingency table (Results Workbook) for these types with lesser levels of agreement reveals the many ongoing challenges with finding agreement between experts and auto-keys for complex vegetation types. The results for the Alaska GeoArea in particular show how difficult it is to distinguish plots when the overall floristic diversity is low (such as in arctic Alaska grading into the boreal region), and the difference between the ecological systems is based on often subtle differences in concept and floristics that are not well represented in the plot data. For example the suite of ericaceous and dwarf-shrub species that characterize shrubland communities in Alaska is limited. High elevation shrublands in the boreal regions will have much the same floristics as in the arctic, but are split into different ecological systems.

Boreal spruce and hardwood forests and woodlands have the same predominant tree species (black and white spruce, paper birch, aspen, and balsam poplar), and are important species in some dozen ecological systems, grading from sub-arctic spruce & lichen woodlands to black spruce peatlands. If adequate compositional data are not available in the plots, and the indicators of “wetlands” or “peatlands” or “floodplains” are not clearly understood, then neither the auto-key nor expert will be able to label the plots with confidence.

The contingency table shows this confusion and here we summarize a cross-section of results from GeoArea 8.

Western North American Boreal White Spruce Forest is confused with Western North American Boreal White Spruce-Hardwood Forest, Alaska Sub-boreal White-Lutz Spruce Forest and Woodland, Western North American Boreal Treeline White Spruce Woodland, Western North American Boreal Mesic Black Spruce Forest, Western North American Boreal Lowland Large River Floodplain Forest and Shrubland, and Western North American Boreal Montane Floodplain Forest and Shrubland. All of these have somewhat subtle differences in the structural characteristics, such as the percent of hardwood versus ocnfier in the canopy or relative proportions of white spruce versus black or lutz spruce. They also share a suite of similar shrub and forb species, even the systems found in floodplains will have much the same set of low or dwarf-shrubs and perennial forbs as the adjacent upland systems.

Western North American Boreal Mesic Black Spruce Forest is confused with Western North American Boreal Black Spruce Wet-Mesic Slope Woodland, Western North American Boreal Black Spruce Dwarf-Tree Peatland, Western North American Boreal Black Spruce-Tamarack Fen. All of these systems are dominated by black spruce but are separated by a relative moisture gradient as well as a gradient of peat development. If plots are lacking information on the non-vascular components (which are used as indicators for the fen or peat systems), or similarly lacking the shrub or sedge indicators of peatlands versus fens, then keying them can be especially difficult.

Western North American Boreal Mesic Scrub Birch-Willow Shrubland is confused with many other shrubland types, but especially Western North American Boreal Low Shrub-Tussock Tundra, Western North American Boreal Alpine Floodplain, Alaskan Pacific Maritime Alpine Floodplain, and Western North American Boreal Low Shrub Peatland. Scrub birch (Betula nana or B. glandulosa) occurs across all of Alaska, and ranges from dry uplands, the understory of spruce woodlands, to wetlands and fens (herbaceus indicators are critical to distinguishing across these types). But this example highlights the issue that floodplains in Alaska often do not have clear floristic indicators, as are usually found in drier areas of the western U.S.

Alaskan Pacific Maritime Sitka Spruce Forest is confused with Alaskan Pacific Maritime Periglacial Woodland and Shrubland and Alaskan Pacific Maritime Mountain Hemlock Forest. And the Alaskan Pacific Maritime Poorly Drained Conifer Woodland was confused with North Pacific Mesic Western Hemlock-Yellow-cedar Forest, which in turn was confused with the Alaskan Pacific Maritime Western Hemlock Forest. In coastal maritime Alaska, with very high amounts of precipitation, the shrub and herb composition within the forest and woodland ecological systems is not highly variable. Distinguishing between poorly drained conifer woodlands (swamps), forested uplands, and floodplains is difficult without information as to micro-topographic characteristics of the plot.

Western North American Boreal Alpine Mesic Herbaceous Meadow – This system is characterized by herbaceous species which are found in both meadow and marsh systems. Many of these plots were difficult to assign as a result of a lack of environmental data to determine the level of hydrologic inundation.

Alaska Arctic Non-Acidic Dwarf-Shrub Lichen Tundra – Since this system is defined by high lichen cover, but as non-vascular cover was often not recorded the expert would have to default to other shrub systems such as Alaska Arctic Scrub Birch-Ericaceous Shrubland.

Alaskan Pacific Maritime Avalanche Slope Shrubland – This system was often labeled by experts as North Pacific Alpine and Subalpine Bedrock and Scree resulting from similar environmental and floristic characteristics.

Western North American Boreal Herbaceous Fen – This system is characterized by peat soils, but was difficult for experts to attribute due to lack of environmental data. It was not possible to determine if soils were peat. Therefore it was difficult to determine if these plots represented fen or meadow systems due to similar floristics; if the fen indicators listed in the description were not present or recorded on the plot it would not be possible to know if the plot represented a fen or wet meadow.

Western North American Boreal Spruce-Lichen Woodland – Non-vascular cover was not always recorded in plots making it difficult to attribute to this system.

Expert Assignments

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

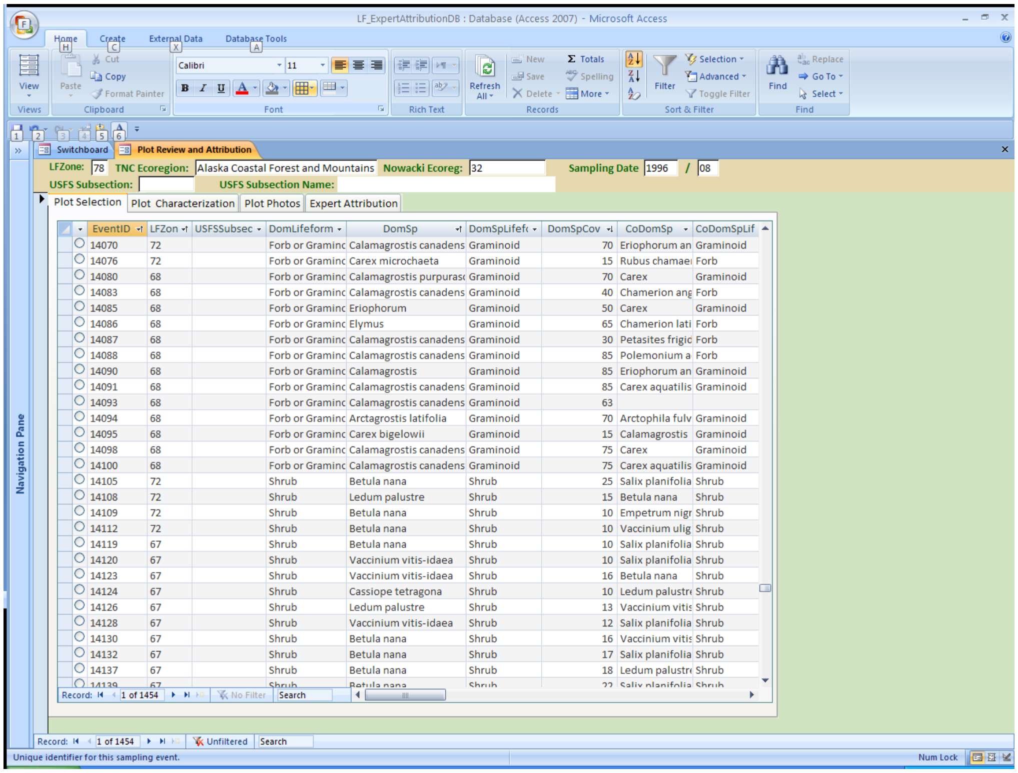
As described in the methods section above, the expert reviewers worked directly in the expert attribution database (EADB). Since GeoArea 8 had over 1,400 plots to review, a systematic, efficient process for reviewing and labeling plots was required. The forms provided in the EADB allowed the reviewer 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 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 mariana*), 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, within the group of 5 ecological systems where black spruce is the major tree component as described above differences between them are based on a gradient of moisture and peat development; indicators of such conditions either floristic ot biophysical settings, is necessary to accurately key the types. Since photos or image clips for many plots were either lacking or of poor quality, these were not available to assist the expert in the review and attribution. In addition a complete species list is necessary to confidently distinguish these types, especially in the non-vascular composition of the plot which is rarely collected.

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	LFZon	USFSSubsec	DomLifeform	DomSp	DomSpLifefc	DomSpCov	CoDomSp	CoDomSpLif
14070	72		Forb or Graminc	Calamagrostis canadens	Graminoid	70	Eriophorum an	Graminoid
14076	72		Forb or Graminc	Carex microchaeta	Graminoid	15	Rubus chamae	Forb
14080	68		Forb or Graminc	Calamagrostis purpuras	Graminoid	70	Carex	Graminoid
14083	68		Forb or Graminc	Calamagrostis canadens	Graminoid	40	Chamerion ang	Forb
14085	68		Forb or Graminc	Eriophorum	Graminoid	50	Carex	Graminoid
14086	68		Forb or Graminc	Elymus	Graminoid	65	Chamerion lati	Forb
14087	68		Forb or Graminc	Calamagrostis canadens	Graminoid	30	Petasites frigid	Forb
14088	68		Forb or Graminc	Calamagrostis canadens	Graminoid	85	Polemonium a	Forb
14090	68		Forb or Graminc	Calamagrostis	Graminoid	85	Eriophorum an	Graminoid
14091	68		Forb or Graminc	Calamagrostis canadens	Graminoid	85	Carex aquatilis	Graminoid
14093	68		Forb or Graminc	Calamagrostis canadens	Graminoid	63		
14094	68		Forb or Graminc	Arctagrostis latifolia	Graminoid	70	Arctophila fulv	Graminoid
14095	68		Forb or Graminc	Carex bigelowii	Graminoid	15	Calamagrostis	Graminoid
14098	68		Forb or Graminc	Calamagrostis canadens	Graminoid	75	Carex	Graminoid
14100	68		Forb or Graminc	Calamagrostis canadens	Graminoid	75	Carex aquatilis	Graminoid
14105	72		Shrub	Betula nana	Shrub	25	Salix planifolia	Shrub
14108	72		Shrub	Ledum palustre	Shrub	15	Betula nana	Shrub
14109	72		Shrub	Betula nana	Shrub	10	Empetrum nigr	Shrub
14112	72		Shrub	Betula nana	Shrub	10	Vaccinium ulig	Shrub
14119	67		Shrub	Betula nana	Shrub	10	Salix planifolia	Shrub
14120	67		Shrub	Vaccinium vitis-idaea	Shrub	10	Salix planifolia	Shrub
14123	67		Shrub	Vaccinium vitis-idaea	Shrub	16	Betula nana	Shrub
14124	67		Shrub	Cassiope tetragona	Shrub	10	Ledum palustr	Shrub
14126	67		Shrub	Ledum palustre	Shrub	13	Vaccinium vitis	Shrub
14128	67		Shrub	Vaccinium vitis-idaea	Shrub	12	Salix planifolia	Shrub
14130	67		Shrub	Betula nana	Shrub	16	Vaccinium vitis	Shrub
14132	67		Shrub	Betula nana	Shrub	17	Salix planifolia	Shrub
14137	67		Shrub	Betula nana	Shrub	18	Ledum palustr	Shrub

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 (recall, very few Alaska plots had useable image clips),
b) review of where the plot was located geographically (noting that for the GeoArea 8 plots, local geographic information such as USFS Subsection was not available, only very coarse-scale units such as the Map Zones, TNC or Nowacki ecoregions), to distinguish Arctic vs. Boreal for example,
c) consideration of topographic setting (e.g. north-facing slopes could logically represent the Western North American Boreal Black Spruce Wet-Mesic Slope Woodland),
d) consideration of any [rarely] available height data for the plot (e.g. were the black spruce trees all tall, apparently mature trees; or were they dwarfed),
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 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

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

Of the 93 types assigned to plots by experts, 45 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 48 types, the numbers of samples labeled to a given type ranged from 58 (for Western North American Boreal Mesic Scrub Birch-Willow Shrubland) down to 10 (for Western North American Boreal Low Shrub Peatland). For all of these types, experts reported high confidence in their labels for at least 50% of the type's plots. 4 types indicated low confidence for at least 20% of the type's plots. These statistics are listed in the Results Workbook. A small sampling of expert comments related to moderate or low confidence plots are included in Table 4.

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

Type Name	Expert Comment
Western North American Boreal Mesic Scrub Birch-Willow Shrubland	Betula glandulosa is not a dominant species in this system.
Western North American Sub-boreal Mesic Bluejoint Meadow	Unidentified sedge is problematic in assigning this system.
Alaska Arctic Polygonal Ground Shrub-Tussock Tundra	The presence of Carex utriculata suggests this is not an upland system, but tussock cover is very low.
Alaska Arctic Acidic Dryas Dwarf-Shrubland	Tussock sedges suggest this might be ecotonal
Alaskan Pacific Maritime Sitka Spruce Beach Ridge	Incomplete floristic data.

These and other comments point to several important aspects for consideration. First, some ecological systems concepts are better known and understood than others. Therefore, a certain degree of classification refinement is likely needed in order to improve auto-keys. This is particularly the case with a number of Alaskan ecological systems, where the literature and supporting plot data are often lacking complete floristic information, and where many species are characteristic in multiple system types across large areas of the state. Another issue is the taxonomic uncertainty for many groups of taxa, such as sedges, willows, and the dwarf or scrub birches; these tend to be difficult to distinguish correctly in the field which in turn leads to uncertainty as to the correct floristic composition for individual system types. Second, the inclusion of some limited landform, soil, and or landscape context information could assist with some determinations within the key, or by a subsequent expert reviewer. Third, additional floristic information is cited in some cases where their suspected limitations provide the primary source of expert uncertainty in labeling.

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 138 terrestrial ecological system types could occur. Of these 9 have a practical 1:1 relationship with NVC Group concepts, and 121 nest cleanly within 44 NVC Group concepts (1:many group:system relationship), for a total of 130 or 94% 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, just 3 (2%) of 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.

Table 5. Ecological Systems of GeoArea 8 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
North Pacific Shrub Swamp	G256 North Pacific Maritime Hardwood-Conifer Rich Swamp	3	2
North Pacific Shrub Swamp	G610 North Pacific Maritime Poor Swamp & Bog Forest	3	1
North Pacific Shrub Swamp	G322 Vancouverian Wet Shrubland	3	2
Alaskan Pacific Maritime Poorly Drained Conifer Woodland	G256 North Pacific Maritime Hardwood-Conifer Rich Swamp	1	2
Alaskan Pacific Maritime Subalpine Copperbush Shrubland	G322 Vancouverian Wet Shrubland	1	2

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 21 US-NVC Macrogroups represented among natural mapped classes in this GeoArea. There may be disagreements between expert and autokeys at the ecological systems level that would be resolved if the intention was to roll-up labeled classes to broader Macrogroup classes. The following types listed below are those for which this could be the case within this GeoArea, although for some of these the increase in percent of matches is small.

Tsuga heterophylla - Picea sitchensis - Sequoia sempervirens - Acer macrophyllum Forest Macrogroup
Vancouverian Flooded & Swamp Forest Macrogroup
Western North American Boreal Conifer & Hardwood Forest Macrogroup
Vancouverian Alpine Scrub, Forb Meadow & Grassland Macrogroup
Western Boreal Alpine Macrogroup
North American Arctic Tundra & Subarctic Alpine Macrogroup
Arctic Tundra Wet Meadow & Marsh Macrogroup
Northern Vancouverian Lowland & Montane Grassland & Shrubland Macrogroup
North American Boreal Shrubland & Grassland Macrogroup
North American Boreal Bog & Fen Macrogroup
Drosera rotundifolia - Comarum palustre - Brown mosses - *Sphagnum* spp. North Pacific Bog & Fen Macrogroup
Western North American Boreal Shrubland, Wet Meadow & Marsh Macrogroup
North American Pacific Coastal Salt Marsh Macrogroup

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**
M173 North American Arctic Tundra & Subarctic Alpine Macrogroup	14	272	22%	63%
M174 Arctic Tundra Wet Meadow & Marsh Macrogroup	10	200	44%	79%
M156 Western North American Boreal Conifer & Hardwood Forest Macrogroup	8	160	41%	67%
M024 Tsuga heterophylla - Picea sitchensis - Sequoia sempervirens - Acer macrophyllum Forest Macrogroup	5	100	67%	94%
M062 North American Boreal Bog & Fen Macrogroup	4	81	10%	12%
M172 Northern Vancouverian Lowland & Montane Grassland &	4	80	46%	73%

Shrubland Macrogroup				
M055 North American Boreal Shrubland & Grassland Macrogroup	3	61	38%	72%
M081 North American Pacific Coastal Salt Marsh Macrogroup	3	60	48%	80%
M404 Western Boreal Alpine Macrogroup	3	60	45%	55%
M063 Drosera rotundifolia - Comarum palustre - Brown mosses - Sphagnum spp. North Pacific Bog & Fen Macrogroup	3	60	10%	12%
M101 Vancouverian Alpine Scrub, Forb Meadow & Grassland Macrogroup	2	40	18%	40%
M035 Vancouverian Flooded & Swamp Forest Macrogroup	2	40	5%	8%
M300 North American Boreal Flooded Forest Macrogroup	2	40	0%	0%
M072 Western North American Boreal Shrubland, Wet Meadow & Marsh Macrogroup	2	40	0%	3%
M073 Western North American Temperate Lowland Wet Shrubland, Wet Meadow & Marsh Macrogroup	2	40	0%	0%
M109 Nuphar polysepala - Azolla filiculoides - Elodea nuttallii Western North American Freshwater Aquatic Macrogroup	1	20	85%	85%
M299 North American Boreal Swamp & Bog Forest Macrogroup	1	20	80%	80%
M025 Abies magnifica - Abies X shastensis - Tsuga mertensiana - Pinus contorta var. murrayana Forest Macrogroup	1	20	70%	70%
M120 Vancouverian Alpine Cliff, Scree & Rock Vegetation Macrogroup	1	20	25%	25%
M179 North American Subalpine & Subarctic Woodland Macrogroup	1	20	15%	15%
M175 Arctic & Boreal Cliff, Scree & Rock Vegetation [Placeholder] Macrogroup	1	20	15%	15%

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 combination of multiple species in varying abundance that are used in a sequence table to key plots; hence incomplete or poorly collected species compositional data generate poor results from the auto-keying process.

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

C. Developing automated keys for large geographic areas

Each sequence table was constructed to work across relatively large geographic areas, on the order of 2-5 USFS Sections (Figure 1). Hence each sequence table/auto-key included tens of ecological system types, and each system has some degree of compositional and structural variability across that region.

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

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

D. Cost/benefit & efficiency

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

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

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

Recommendations

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.

The ecological systems classification for Alaska would benefit from some additional investments in compilation of field-based vegetation data, improvement of the concepts of the systems based on these data, and possibly some revisions to the systems classification itself (combining some systems into more thematically broad types for example).

Latitudinal and elevational gradients from north to south in Alaska have not been clearly described in relation to the distribution of the ecological systems. Breaks between “arctic” and “boreal” vegetation are not well related to either the MRLC map zones, the Nowacki et al. (2001) level III ecoregions, or the TNC-defined ecoregions, which are the 3 ecoregional distribution units that were available for the auto-key and the expert to use. Further development of more locally-scaled distribution data for the Alaska ecological systems (e.g. Nowacki level 4 ecoregions, or USFS Sections if they were available) would help with this.

Adjustments to Auto-key procedures – inclusion of locational/biophysical information for pre-processing plots and/or inclusion of features in auto-keys. This is mentioned above in relation to the diffiiculty of keying black ecological systems that are distiguished along moisture and peat development gradients, when ocmplete floristics are lacking or not understood, micro- or meso-scale topographic information would help improve auto-key results.

Narrowing vs. broadening the geographic application of the auto-key – FS Sections? In certain areas? Would this likely lead to greater accuracy? This might be a helpful thing to consider for Alaska- the 4 sets of auto-keys of necessity had to include many ecological systems that were peripheral to the key. For example, the “boreal” map zones cover a huge area of interior Alaska, including the southern slopes of the Brooks Range on the north (with some arctic types), the northern slopes of the Alaska Range on the south (incuding some maritime types), the Cook Inlet region, and then extending west into western arctic areas.

Adjustment to auto-keys – additional requirements for vegetation sample data; primarily ground cover data. Again, for ecological systems where the bryophytes and other non-vascular species are particularly important indicators yet are rarely recorded for plots, this is an important issue.

Adjustments to Map Legends – moving to Group/Macrogroup concepts where systems level remains challenging – which ones? Many ecological systems in Alaska are grouped into a single NVC Group concept; 121 nest cleanly within 44 NVC Groups suggesting that some of the NVC Groups may be somewhat broader in concept, and might prove to be improved units for mapping and auto-keys.