Expert Attribution for Auto-Key Improvements (LANDFIRE) and Advancing Methods for integration with the revised US-National Vegetation Classification Standard: GeoArea 7E

FINAL REPORT

Prepared by

NatureServe

For the NPS Vegetation Inventory Program & LANDFIRE

30 June 2012

© Mark Hall, Appalachian Trail, Maine.

NatureServe

A Network Connecting Science With Conservation

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

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

Table of Contents

Figures

Tables

Introduction

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

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

Project Goals

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

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

Background on LANDFIRE Auto-keys

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

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

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

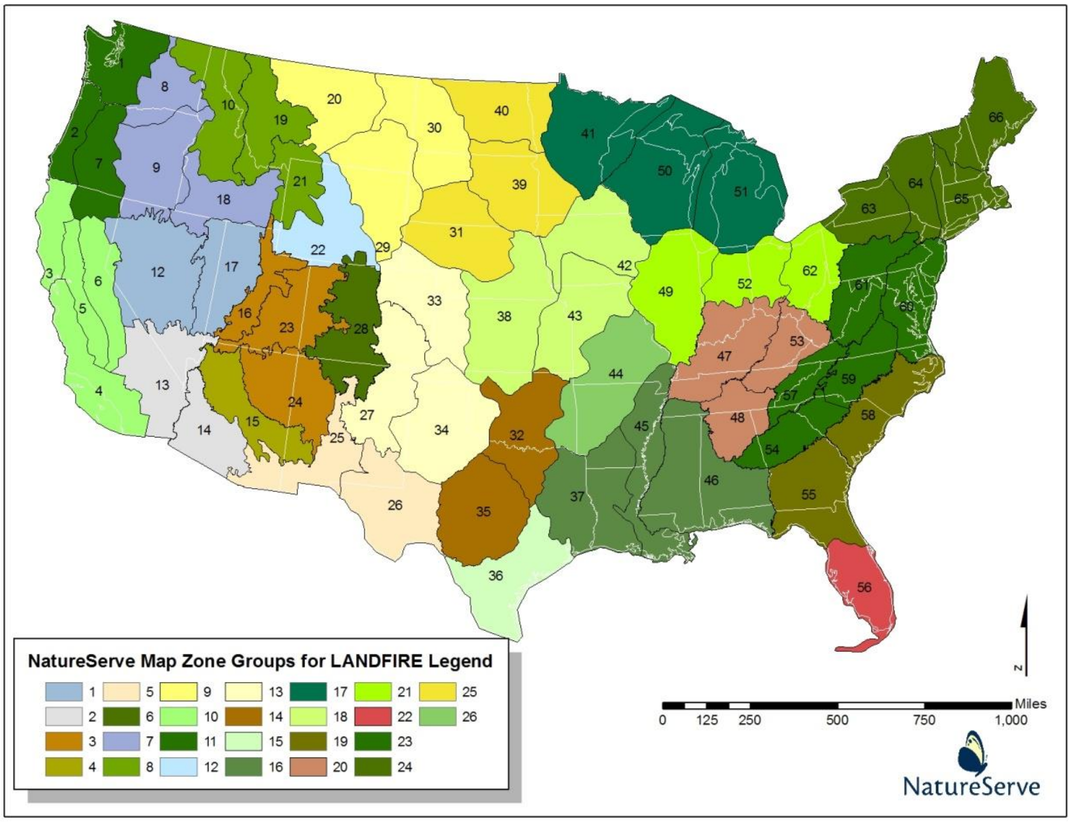
LANDFIRE's short-term needs, and long-term plans, required a repeatable methodology, consistently applied rules to categorize each reference sample, and documentation of the criteria applied. In essence, sequence tables codify the criteria and methods for keying geo-referenced vegetation data to a land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

More details about this methodology include:

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

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

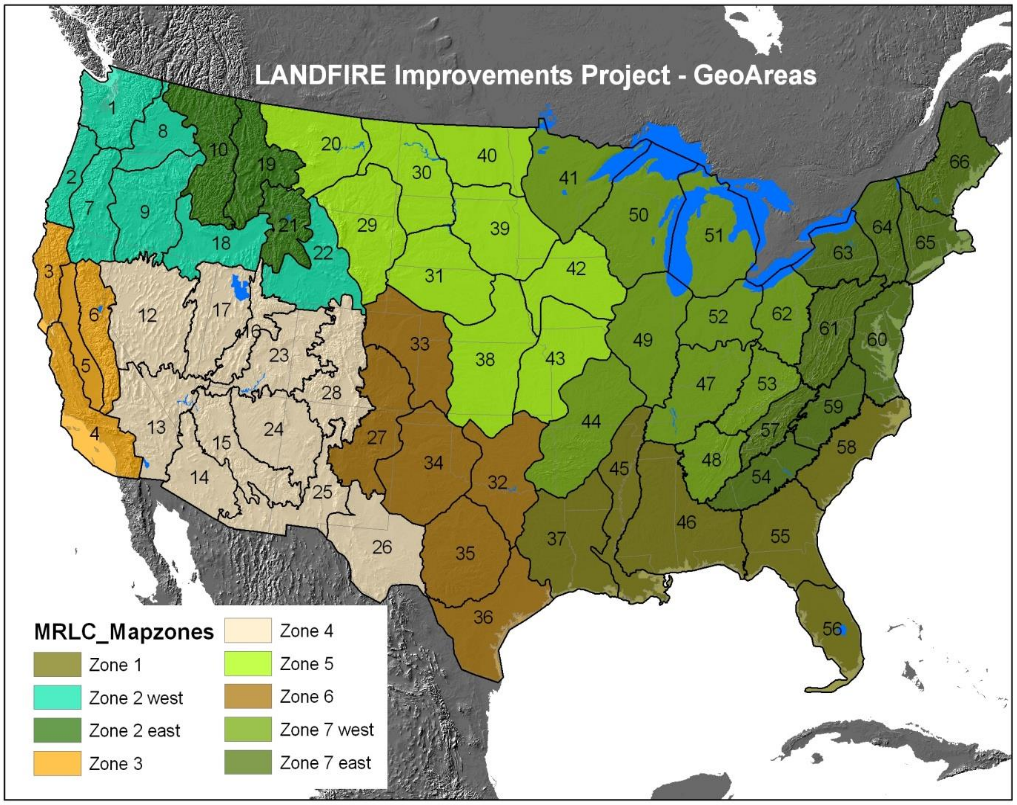
Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in the coterminous U.S.



Methods

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

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



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

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

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

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

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

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

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

Figure 3. Example of an image clip for one plot in GeoArea 7E

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 7E 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 7E: Northeast

GeoArea 7E encompasses the southern, central and northern Appalachian, Blue Ridge, Piedmont, High Allegheny Plateau ecoregions. This GeoArea includes a total of 9 map zones (54, 57, 59, 60, 61, 63, 64, 65, 66; Figure 2) originally clustered for purposes of designing and implementing auto-keys. The total number of plots in this Geo Area analysis was 1,713. A total of 35 natural ecological system types were assigned to a total of 1,233 plots by the auto-keys. A total of 71 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 or wetland/riparian types).

An additional 17 types were assigned by the auto-key but were not assigned by experts, including 3 standard systems:

- Acadian-Appalachian Subalpine Woodland and Heath-Krummholz
- Central Appalachian Alkaline Glade and Woodland
- Eastern Serpentine Woodland

and 14 broader groups of systems ("aggregates"):

- Boreal Acidic Peatland Systems
- Central Interior and Appalachian Floodplain Systems
- Central Interior and Appalachian Riparian Systems
- Central Interior and Appalachian Sparsely Vegetated Systems
- Central Interior and Appalachian Swamp Systems
- Gulf and Atlantic Coastal Plain Floodplain Systems
- Gulf and Atlantic Coastal Plain Small Stream Riparian Systems
- Gulf and Atlantic Coastal Plain Sparsely Vegetated Systems
- Gulf and Atlantic Coastal Plain Swamp Systems
- Gulf and Atlantic Coastal Plain Tidal Marsh Systems
- Laurentian-Acadian Floodplain Systems
- Laurentian-Acadian Shrub-Herbaceous Wetland Systems
- Laurentian-Acadian Sparsely Vegetated Systems
- Laurentian-Acadian Swamp Systems

Comparison of Auto-key and Expert Assignments

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

Of the 35 natural types assigned labels by the auto-keys, 9 types (26%) had fewer than 10 samples available for this analysis (Table 2). Seven of these undersampled types are relatively uncommon and occur naturally in small patches rarely exceeding a few hundred acres in extent: Central Appalachian Alkaline Glade and Woodland [n=8]; Southern Appalachian Grass and Shrub Bald [n=6]; Northern Atlantic Coastal Plain Maritime Forest [n=6]; Northern Atlantic Coastal Plain Dune and Swale [n=4]; Eastern Serpentine Woodland [n=2]; Acadian-Appalachian Subalpine Woodland and Heath-Krummholz [n=1]; and Northeastern Interior Pine Barrens [n=1]. The remaining two types are at the edge of their range in this GeoArea: Southern Atlantic Coastal Plain Nonriverine Swamp and Wet Hardwood Forest [n=7], and Southern Atlantic Coastal Plain Dry and Dry-Mesic Oak Forest [n=5]. These under-sampled types were excluded from further analysis.

Table 2. Under-sampled types within GeoArea 7E

EVT Code	EVT Name	System elcode	total Plots
2400	Central Appalachian Alkaline Glade and Woodland	CES202.602	8

EVT Code	EVT Name	System elcode	total Plots
2501	Southern Atlantic Coastal Plain Nonriverine Swamp and Wet Hardwood Forest	CES203.304	7
2414	Southern Appalachian Grass and Shrub Bald	CES202.294	6
2379	Northern Atlantic Coastal Plain Maritime Forest	CES203.302	6
2335	Southern Atlantic Coastal Plain Dry and Dry-Mesic Oak Forest	CES203.241	5
2436	Northern Atlantic Coastal Plain Dune and Swale	CES203.264	4
2375	Eastern Serpentine Woodland	CES202.347	2
2389	Acadian-Appalachian Subalpine Woodland and Heath-Krummholz	CES201.568	1
2354	Northeastern Interior Pine Barrens	CES202.590	1

A total of 2 types having more than 20 sample plots (or nearly 6% of 35 types) had >80% agreement between expert and auto-key assignments. These types were Southeastern Interior Longleaf Pine Woodland [n=24; 88% agreement between expert and auto-key assignment] and Northern Atlantic Coastal Plain Pitch Pine Barrens [n=49; 84% agreement]. Self-assessments of confidence for these types were predominantly 'high' or 'moderate'.

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 24, or nearly 69% of the total types assigned. 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. It appears, however, that the greatest percentage of errors occurred among floristically similar groups in areas of geographic transition. In only XXX systems were there no apparent floristic similarity between the assignments when they did not agree. Here we summarize a cross-section of results from GeoArea 7.

There was 72% (36 of 50 plots) where the expert assignment and the auto-key matched for the Southern Appalachian Oak Forest. The experts had high confidence in their assignments for 82% of the plots, and indicated that they used oaks species composition and elevation to label their plots. Of the 14 mis-matches, 6 had been labeled as Southern Appalachian Cove Forest, 3 as Southern Appalachian Low Elevation Pine, and 2 as Allegheny-Cumberland Dry Oak Forest and Woodland.

Agreement in the Southern Piedmont Dry Oak Forest was 69% (35 of 50 plots). The experts were highly confident in their assignment for 80% of the plots and identified species composition as a major indicator of the system. Experts labeled 4 of the 15 mismatched plots as Southern Piedmont Mesic Forest, 4 plots as Southern Appalachian Oak Forest and 2 plots as Southern Appalachian Low Elevation Pine Forest. The experts were likely to be incorporating specific species in calling mesic forest and on the ecoregional and elevation information on labeling the Southern Appalachian types.

Those systems primarily characterized by dry-site oaks were confused with each other. Significant numbers of plots assigned by the auto-key to Central Appalachian Dry Oak-Pine Forest were assigned to either Allegheny-Cumberland Dry Oak Forest and Woodland (n=22]; Northeastern Interior Dry-Mesic Oak Forest [n=24]; or Central Appalachian Pine-Oak Rocky Woodland [n=21] by expert attribution. Only 5% of plots assigned by experts were not either high or medium confidence. All three forested types are characterized by widespread oak species such as white oak (*Quercus alba*) or chestnut oak (*Quercus prinus)*, but have different geographic ranges. Central Appalachian Pine -Oak Rocky Woodland shares many of the same characteristic tree species but has an open canopy and a steeper environmental setting. Reviewer comments indicated that in some cases, assignment could be equally applied to two floristically related systems, and that sufficient species information for high confidence was lacking. Experts used subsection of occurrence to discriminate in geographic transition areas, and where the subsection spanned two related types, confidence in assignment was lower.

Floristically related spruce-fir and northern hardwood systems of the central and southern Appalachians were also confused with each other. For example, a number of Central and Southern Appalachian Spruce-Fir Forest plots were erroneously identified as Southern Appalachian Northern Hardwood Forest or Appalachian (Hemlock)-Northern Hardwood Forest by the auto-key. A significant number of plots [n=27] auto-keyed to South-Central Interior Mesophytic Forest were assigned to Appalachian (Hemlock)-Northern Hardwood Forest by experts. This is likely because both systems share characteristic species such as sugar maple (*Acer saccharum)*, beech (*Fagus grandifolia),*and hemlock (*Tsuga canadensis).* Expert confidence in attributions to Appalachian (Hemlock)-Northern Hardwood Forest indicates a higher degree of uncertainty.

A similar pattern was noted in northern analogs of spruce-fir and northern hardwood systems: a number of plots keyed as Acadian Low-Elevation Spruce-Fir-Hardwood Forest were assigned to Acadian-Appalachian Montane Spruce-Fir Forest by experts. Discrepancies were also noted between low-elevation and montane spruce-fir systems; 9 plots assigned by auto-key to Acadian-Appalachian Montane Spruce-Fir Forest were attributed to Acadian Low-Elevation Spruce-Fir-Hardwood Forest by experts. It's possible these disagreements have to do with the elevation rules used in the auto-keys within the GeoArea versus a more local-scale elevation break that the expert reviewer may have used (perhaps even adjusting for latitude).

Discrepancies between auto-key and expert attribution were also common in coastal plain hardwood systems. The auto-key and experts agreed on the assignment of 19 plots to the Northern Atlantic Coastal Plain Hardwood Forest, but experts attributed an additional 31 plots to this type that had been auto-keyed to Southern Atlantic Coastal Plain Mesic Hardwood Forest. These two systems are described as overlapping in the coastal plain with GeoArea7E. It is likely the experts relied heavily on biogeography in addition to species composition, while the 2 sequence tables drafted for this GeoArea relied more on species composition. It's possible that the expert assigning the plots had a different understanding of the ranges of the two types than had been initially incorporated in the sequence table.

North-Central Interior Wet Flatwoods showed 27% agreement (4 of the 15 plots). This system has little floristic similarity to other systems, so errors were caused by other factors. A high proportion of the total number of plots (9 out of 15) could not be assigned due to lack of information; experts had medium confidence in most of the plots assigned, but confidence was not high in any of the assignments. Three of the plots that could not be assigned had attributed by the auto-key to an aggregated wetland system (Central Interior and Appalachian Swamp Systems, Laurentian-Acadian Floodplain Systems, and Gulf and Atlantic Coastal Plain Swamp Systems). Two of the plots that had been assigned to a single system had been assigned by experts to two different upland types. With so few plots finding patterns in the disagreement is a challenge and each of the plots should be explored to help identify potential refinements in characterizing this small patch partially isolated wetland system.

Southern Ridge and Valley / Cumberland Dry Calcareous Forest showed little agreement (14%) with 30 of the 37 plots being assigned by experts to five different systems (South Central Interior Mesophytic Forest (5 plots), Appalachian (Hemlock)-Northern Hardwood Forest (1 plot), Allegheny-Cumberland Dry

Oak Forest and Woodland (7 plots), Central Appalachian Dry Oak-Pine Forest (3 plots), Northeastern Interior Dry-Mesic Oak Forest (12 plots), Southern Appalachian Oak Forest (2 plots). The remaining two mismatched plots were labeled can't assign and "other". The experts actually only labeled 6 plots as the Southern Ridge and Valley/ Cumberland Dry Calcareous Forest and when they did they had high confidence (83%). It is likely the sequence table was labeled plots as this system because species composition is similar across a variety of systems in the GeoArea with biographic boundaries constraining the concept. If the sequence table drafted for the entire GeoArea did not take into account ecoregional boundaries that could explain the low agreement and the assignment by experts to a diversity of systems.

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**
2315	Southern Appalachian Oak Forest	CES202.886	50	36	72%	28	5	3
2368	Southern Piedmont Dry Oak(-Pine) Forest	CES202.339	50	35	70%	29	6	0
2369	Central Appalachian Dry Oak-Pine Forest	CES202.591	50	33	66%	16	17	0
2350	Central and Southern Appalachian Spruce-Fir Forest	CES202.028	45	28	62%	27	0	1
2302	Laurentian-Acadian Northern Hardwoods Forest	CES201.564	50	31	62%	11	18	2
2374	Acadian-Appalachian Montane Spruce-Fir Forest	CES201.566	50	29	58%	19	9	1
2370	Appalachian (Hemlock)-Northern Hardwood Forest	CES202.593	51	28	55%	6	14	8
2373	Acadian Low-Elevation Spruce-Fir-Hardwood Forest	CES201.565	50	26	52%	0	26	0
2318	Southern and Central Appalachian Cove Forest	CES202.373	50	26	52%	17	9	0
2303	Northeastern Interior Dry-Mesic Oak Forest	CES202.592	48	19	40%	5	11	3
2324	Northern Atlantic Coastal Plain Hardwood Forest	CES203.475	51	19	37%	7	11	1
2309	Southern Appalachian Northern Hardwood Forest	CES202.029	27	10	37%	6	4	0
2353	Southern Appalachian Low-Elevation Pine Forest	CES202.332	50	16	32%	9	3	4
2366	Laurentian-Acadian Pine-Hemlock-Hardwood Forest	CES201.563	44	14	32%	2	12	0
2317	Allegheny-Cumberland Dry Oak Forest and Woodland	CES202.359	50	14	28%	10	4	0

				Plots with Expert Matches				
EVT Code	**EVT Name**	**System elcode**	**Total Plots**	**Total**	**%**	**High conf**	**Med conf**	**Low conf**
2377	Central Appalachian Pine-Oak Rocky Woodland	CES202.600	50	14	28%	5	8	1
2518	North-Central Interior Wet Flatwoods	CES202.700	15	4	27%	0	4	0
2352	Southern Appalachian Montane Pine Forest and Woodland	CES202.331	50	12	24%	8	3	1
2320	Central and Southern Appalachian Montane Oak Forest	CES202.596	50	12	24%	7	4	1
2316	Southern Piedmont Mesic Forest	CES202.342	52	12	23%	9	3	0
2362	Laurentian-Acadian Northern Pine(-Oak) Forest	CES201.719	50	9	18%	1	5	3
2376	Southern Ridge and Valley/Cumberland Dry Calcareous Forest	CES202.457	37	5	14%	4	0	1
2321	South-Central Interior Mesophytic Forest	CES202.887	50	6	12%	0	6	0
2343	Southern Atlantic Coastal Plain Mesic Hardwood Forest	CES203.242	50	4	8%	1	3	0

Expert Assignments

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

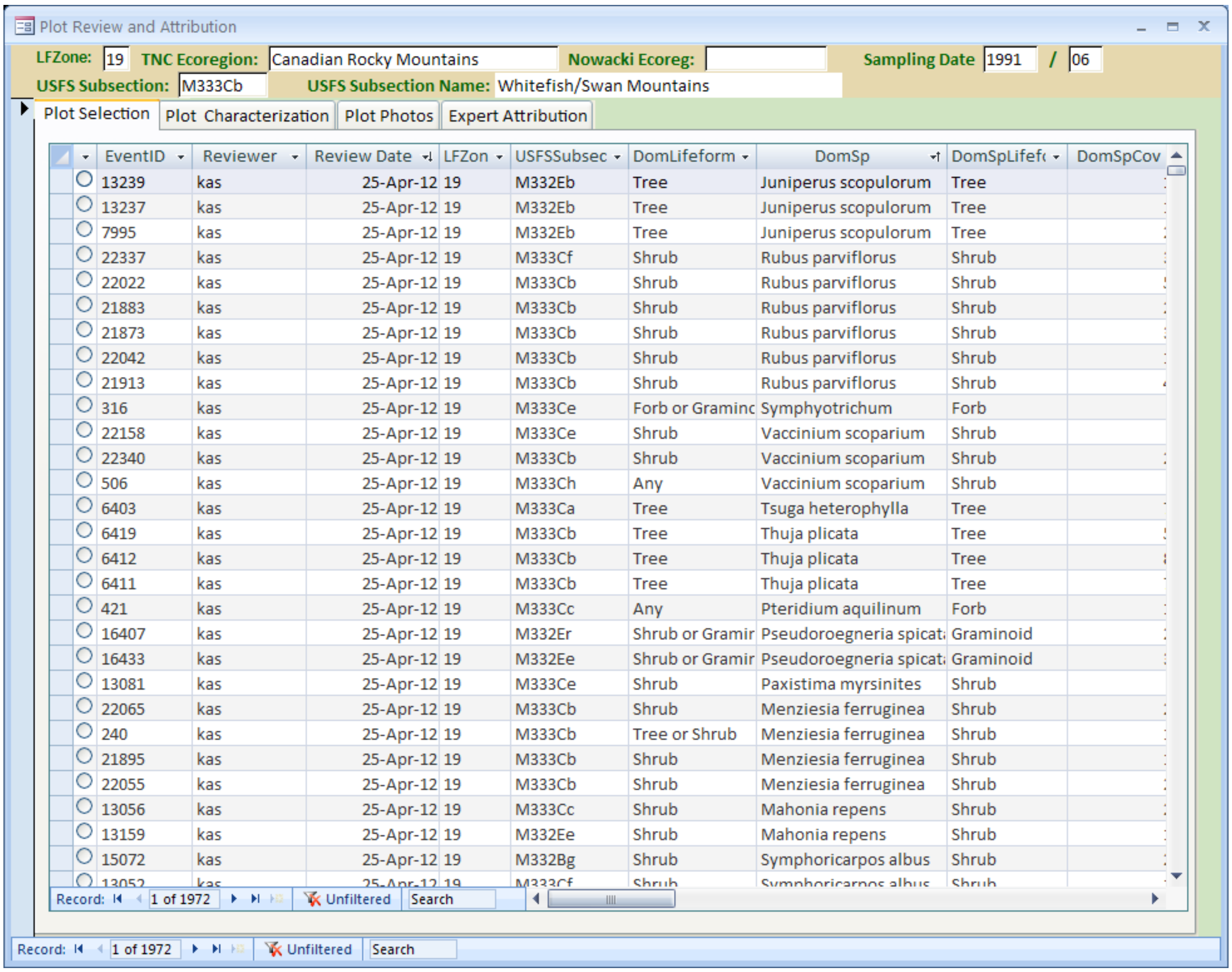
As described in the methods section above, the expert reviewers worked directly in the expert attribution database (EADB). Since GeoArea 7E had over 2,000 plots to review, a systematic, efficient process for reviewing and labeling plots was required. The forms provided in the EADB allowed the reviewer to sort and filter on subsets of plots to select groups of them with similar characteristics. For instance, the reviewer could select all plots found within a particular USFS Section or MapZone, then select all plots dominated by trees, then sort alphabetically by the dominant species. The reviewer could also select all treed plots, then select all plots with the same dominant tree species (such as *Pinus strobus*), then sort by % cover of that species, from high to low. Figure 4 shows the main form in the EADB which has these data fields. Additional fields were provided from which to select or sort plots, such as elevation, aspect, slope, and total cover by lifeform in the plot.

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

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

For example, in the northern Appalachian and sub-boreal region of the northeast, transitions from northern hardwoods to spruce-fir forests occur. In these areas, the tree canopy might be a mix of beech, sugar maple, yellow birch, red spruce, white pine, hemlock, paper birch, and balsam fir, often with little variation in tree canopy cover. In these cases, the reviewer would encounter plots of mixed composition, and need to determine whether those plots represented*Acadian Low-Elevation Spruce-Fir-Hardwood Forest, Laurentian-Acadian Pine-Hemlock-Hardwood Forest, Laurentian-Acadian Northern Hardwoods Forest, Acadian Low-Elevation Spruce-Fir-Hardwood Forest, Acadian-Appalachian Subalpine Woodland and Heath-Krummholz*, *Northern Appalachian-Acadian Conifer-Hardwood Acidic Swamp*, or *Acadian-Appalachian Montane Spruce-Fir Forest.*

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



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

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

a) review of the image clip for the context of the plot,
b) review of where the plot was located geographically (USFS Subsections provide local scale geographic location), to distinguish Acadian vs. Appalachian systems
c) consideration of topographic setting (e.g. highest elevations near summits support woodlands and Krummholz; valley bottoms and topographic depressions support swamps or spruce flats),
d) consideration of any available height data for the plot (e.g. were the conifers tall, apparently mature trees; or were they short and stunted),
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 spruce species by the field crews, unevenness in how the cover values were estimated in the field or converted into the LFRDB (e.g. cover for trees estimated by a person standing on the ground vs. an aerial view of the plot).

Below are some examples of comments relevant to the above spruce – fir and northern hardwood example:

- Near absence of species composition data, however the two co-dominants, the elevation, and subsection of occurrence suggest this is a northern hardwood forest
- This could also be a hemlock swamp; wetland status unknown
- Difficult to determine whether this is a wetland or not. Also, elevation of 1500 feet is close to the transition between montane and lowland
- somewhat sloping and has a small amount of Larix, so wetland status is questionable.
- hemlock not present in plot and land use not known; could be recovering pine plantation
- successional? Near altered land, Fraxinus abundant
- steep slope and relatively high elevation
- Subsection of occurrence, elevation and presence of Abies

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 71 types assigned to plots by experts, 37 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 34 types, the numbers of samples labeled to a given type ranged from 152 (Central Appalachian Dry Oak-Pine Forest) down to 10 (Northern Atlantic Coastal Plain Dune and Swale). For 35 (95%) of these types, experts reported moderate confidence in their labels for at least 20% of the type's plots. None 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

Type Name	Expert Comment
Appalachian (Hemlock)-Northern Hardwood Forest	Sounds like a transitional plot, not much to choose from.
Laurentian-Acadian Northern Hardwoods Forest	Seems early successional, disturbed area.
Southern Appalachian Oak Forest	Species mix describe in this type, but oaks not dominating.
Northeastern Interior Dry-Mesic Oak Forest	Appears to be transitional to Central Appalachian Dry Oak - Pine Forest
Southern and Central Appalachian Cove Forest	This is pretty high ~3500 feet, so is close to CES202.029
Laurentian-Acadian Pine-Hemlock-Hardwood Forest	Could also be low-elevation spruce fir forest

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, wetland status, and/or landscape context information could greatly assist with some determinations within the key, or by a subsequent expert reviewer. For example, spruce flats and acidic swamps share the same canopy dominants with several upland systems in the Acadian / Appalachian regions, and topographic position, wetland status, or soil type could easily improve auto-key results. Aerial photos can provide some additional information where the tree canopy is at least partially open, or the photography was taken under leaf-off conditions. In the heavily forested northeast, many of the photos were taken during the growing season and appeared as a solid green mat, providing little to no additional information.
Third, additional floristic information is cited in some cases where their suspected limitations provide the primary source of expert uncertainty in labeling.

Other samples were labeled by auto-keys to aggregates of multiple ecological system types. This was because LANDFIRE had mapping objectives focused on uplands where fire regimes are prevalent. That meant that many individual wetland and sparsely-vegetated ecological system types were not treated within the auto-keys. Expert labeling of these samples, however, provides an indication of the feasibility of their inclusion in updated auto-keys. Of 438 samples, experts were able to assign 306 (70%) to an individual ecological system type; a total of 49 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 set of samples did not contain enough information for the auto-keys to assign a system or system aggregate; these samples were labeled with broad "unclassified" types, such as "Unclassified Grassland" or "None". Of 42 samples, experts were able to assign 18 (43%) to an individual ecological system type; a total of 12 individual ecological system types were assigned to these samples.

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

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

US-NVC Groups

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

Within this GeoArea, some 138 terrestrial ecological system types could occur. Of these, 12 have a practical 1:1 relationship with NVC Group concepts, and the remaining 121 system concepts (except for 5 with no NVC Group assignment) nest cleanly within 49 NVC Group concepts (1:many group:system relationship). There is some potential for slight differences among floristic elements among these NVC Groups relative to ecological systems. For example, one or more associations linked to a given terrestrial ecological system type may now be linked to a different NVC Group concept. There is some limited potential that the floristic information found within the auto-key would need to be revisited to account for this, but within this GeoArea, we believe that this instance is quite limited.

Where the relationship between ecological systems and NVC Groups is more complex, there is potential need for substantive changes to existing auto-keys. Within this GeoArea, no ecological system types have a more complex relationship with NVC Group concepts (Table 5).

Table 5. Ecological Systems of GeoArea 7E that have complex relationships with NVC Groups

There are no GeoArea 7E Ecological Systems that have complex relationships with NVC Groups.

US-NVC Macrogroups

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

There are 11 US-NVC Macrogroups represented among natural mapped classes in this GeoArea. While the results in Table 6 suggest rolling up to Macrogroup would yield improved results, consideration must be given to the fact that many of these Macrogroups are in fact very broad concepts, and include ecologically diverse system types. For example, four macrogroups together encompass 23 ecological systems and a large proportion of the natural upland forests acreage of GeoArea 7E. These are:

- M014 Northern Mesic Hardwood & Conifer Forest (seven systems)
- M012 Central Oak-Hardwood & Pine Forest (seven systems)
- M159 Northern & Eastern Pine – Oak Forest, Woodland & Barren (five systems)
- M016 South-Central Oak – Hardwood & Pine Forest (four systems)

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
M012 Central Oak-Hardwood & Pine Forest	7	336	41%	79%
M014 Northern Mesic Hardwood & Conifer Forest	7	317	52%	77%
M016 South-Central Oak - Hardwood & Pine Forest	4	129	59%	81%
M153 Central Mesophytic Hardwood Forest	4	202	24%	25%
M159 Northern & Eastern Pine - Oak Forest, Woodland & Barrens	5	156	40%	42%
M030 Northern & Central Swamp Forest	1	15	27%	40%
M033 Southern Coastal Plain Basin Swamp	1	7	43%	43%
M122 Appalachian & Laurentian Rocky Scrub & Meadow	2	56	32%	48%
M124 Northern & Central Alvar & Glade	2	10	0%	0%
M127 Eastern North American Sub-Boreal Shrubland & Grassland	1	1	0%	0%
M057 Eastern North American Coastal Grassland & Shrubland	1	4	100%	100%

The project contingency table for this GeoArea arranged ecological system types according to the US-NVC Macrogroup that would encompass their existing vegetation components. This facilitates rapid analysis to evaluate the potential effect of using the auto-key for Macrogroups; i.e., there may be disagreements between expert and auto-keys at the ecological systems level that would be resolved if the intention was to roll-up labeled classes to broader Macrogroup classes.
Some examples of how generalizing up from the from ecological system to Macrogroup level would improve the agreement between the auto-key and expert assignments include:

- The Central Oak-Hardwood and Pine Forest Macrogroup is comprised of upland hardwood and pine forests of the Allegheny, Appalachian, and Southern Ridge and Valley / Cumberland Plateau ecoregions. Because the need to understand the biogeographic context for the plots would be reduced. rolling up to Macrogroup improves the matches between autokey and expert from 41 to 79%
- The Northern Mesic Hardwood and Conifer Forest Macrogroup is comprised of hemlock-northern hardwood forests and spruce-fir forests of the Laurentian and Acadian regions, as well as the central and southern Appalachians. Rolling up the macrogroup would increase agreement between experts and the existing sequence tables by 25% , going from 52-77% agreement.
- Combining four systems represented in the South-Central Oak-Hardwood & Pine Forest Macrogroup increases agreement from 59-81%. Scaling up to the macrogroup level would combine four relatively distinct ecological systems: Southeastern Interior Longleaf Pine Woodland, Southern Appalachian Low-Elevation Pine Forest, Southern Piedmont Dry Oak-(Pine) Forest, and Southern Atlantic Coastal Plain Dry and Dry-Mesic Oak Forest. In this GeoArea, only 5 plots were available for the Southern Atlantic Coastal Plain Dry and Dry-Mesic Oak Forest and 24 plots were included for Southeastern Interior Longleaf Pine Woodland. Those two systems

had relatively high agreement level with 4 of the 5 coastal plain forest plots and 21 of the 24 longleaf plots already in agreement. These favorable results would be obscured by rolling up to the Macrogroup level. However, an increase in agreement would be made by combining two systems within the macrogroup, the Southern Appalachian Low-Elevation Pine Forest and the Southern Piedmont Dry Oak – (Pine) Forest. There were 27 plots that experts had attributed to the Southern Piedmont Dry Oak – (Pine) Forest that had been labeled by the auto-key as the Southern Appalachian Low-Elevation Pine Forest.

- Little is gained by moving to the macrogroup level for the Central Mesophytic Hardwood Forest Macrogroup, comprised of mesic forests of the southern piedmont, southern and central Appalachians, and south-central interior. Most of the disagreement in labeling those systems was among macrogroups, and lay in the difficulty in distinguishing the mesic types from the dry upland oak types throughout the GeoArea.

It is important to recognize that an important source of disagreement in many of the systems was the result of a plot being assigned to "can't assign" or other. It is unclear how many of those plots could be easily assigned at the macrogroup level, when for many of the plots those calls were made because of a lack of information relative to species composition, or when the plot data and the photo indicated a change had occurred. Ruderal types will continue to be a challenge, and at the macrogroup level the expansion of the species list to accommodate all member systems may in fact make it more difficult to confidently label a plot.

US-NVC Divisions

NVC Divisions are substantially simplified vegetation concepts relative to terrestrial ecological system types, so auto-keys designed for these concepts would be relatively simple to develop. For within this GeoArea, we would recommend starting from a new baseline starting point in order to adequately design one auto-key to encompass the 10 natural US-NVC Division concepts that occur here.

Discussion

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

Fundamentally, a sequence table as used by LANDFIRE is a set of criteria. Each vegetation sample has to meet some combination of criteria in the SQT to be labeled with an ecological system, or some other land cover class. Simply put, if the plot doesn't meet any criteria contained in the sequence table, then it may be mis-keyed, or not key to anything. Given our incomplete 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 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 (draft)

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

Adjustments to Auto-key procedures – inclusion of locational/biophysical information for pre-processing plots and/or inclusion of features in auto-keys: these should include slope and aspect or other indication of topographic position; wetland vs. upland; whether the plot is representative of the area (i.e., whether a complex of different types, or representing a small inclusion within a matrix forest).

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

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

Expert review, labeling, or elimination of plots for types of low confidence from auto-key.

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

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

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

In the eastern U.S. – rapid land use change and past management are huge drivers of vegetation. In the the process of auto-keying it will be important to screen plots for potential changes in land use.

There is a need to standardize the concepts for ruderal and managed types.