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

FINAL REPORT

Prepared by

NatureServe

For the NPS Vegetation Inventory Program & LANDFIRE

30 June 2012

© Jim Drake. Saint Croix River, Wisconsin.

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 Jim Drake, Shannon Menard, Milo Pyne, and Carl Nordman; NatureServe Regional Vegetation Ecologists.

Table of Contents

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.

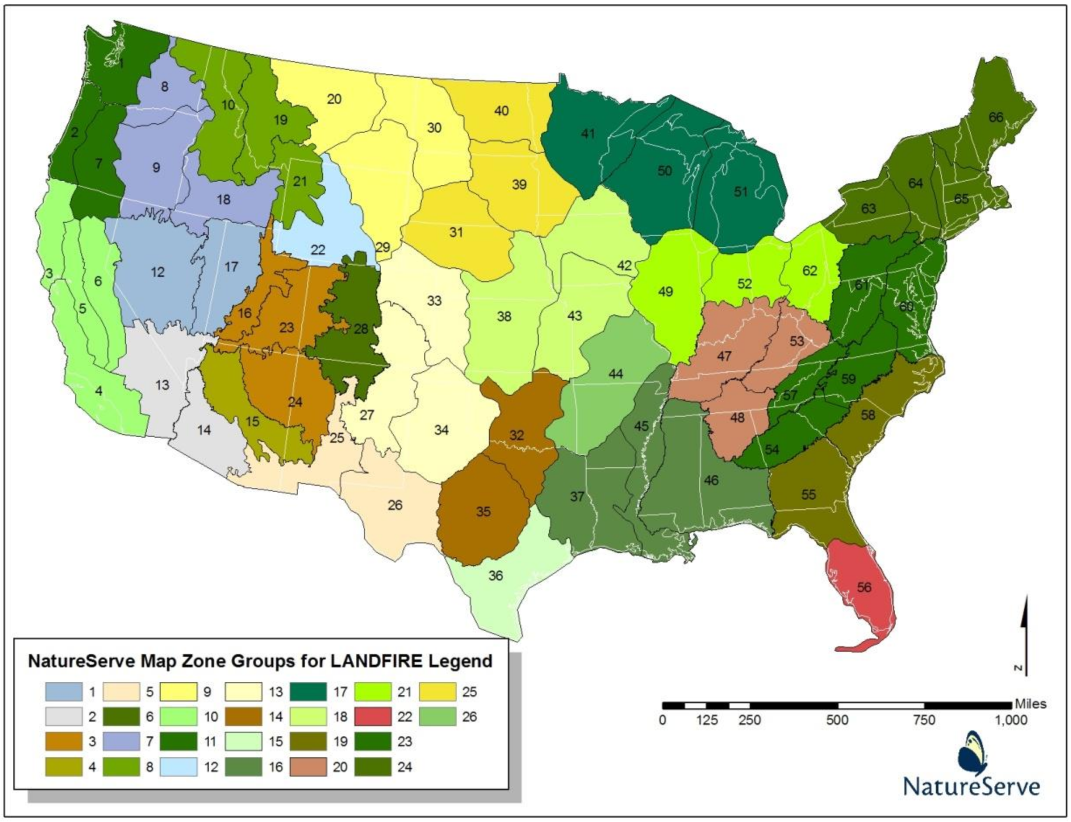
LANDFIRE's short-term needs, and long-term plans, required a repeatable methodology, consistently applied rules to categorize each reference sample, and documentation of the criteria applied. In essence, sequence tables codify the criteria and methods for keying geo-referenced vegetation data to a land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

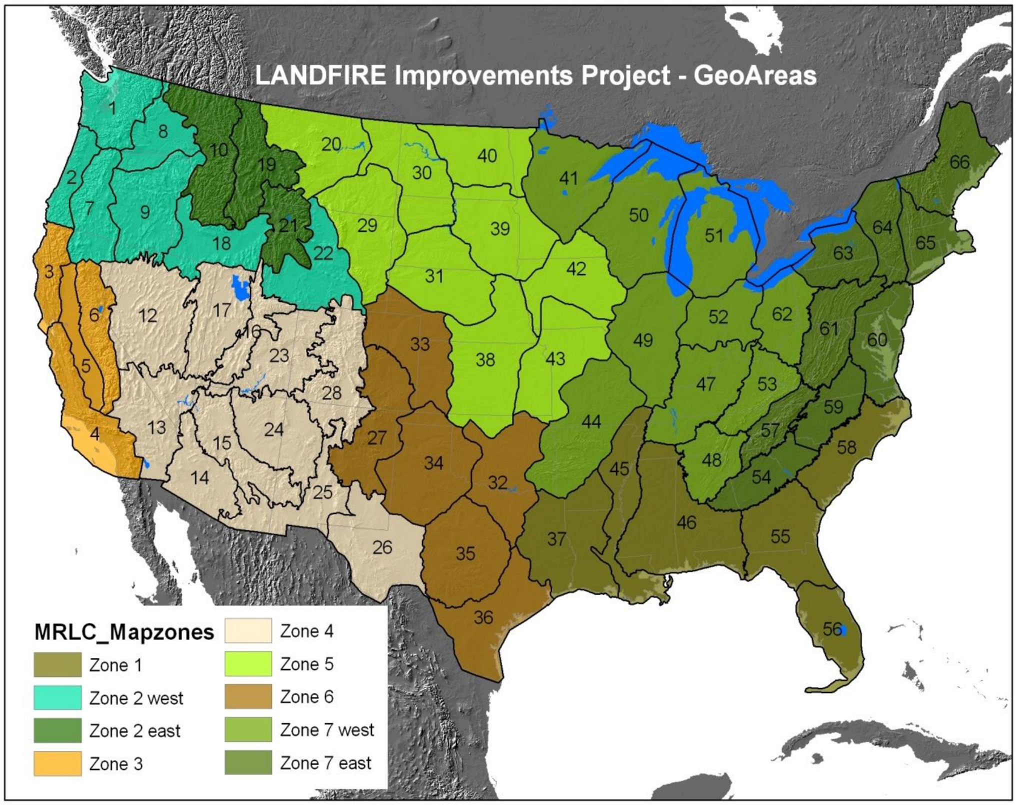
Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in the coterminous U.S.



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:

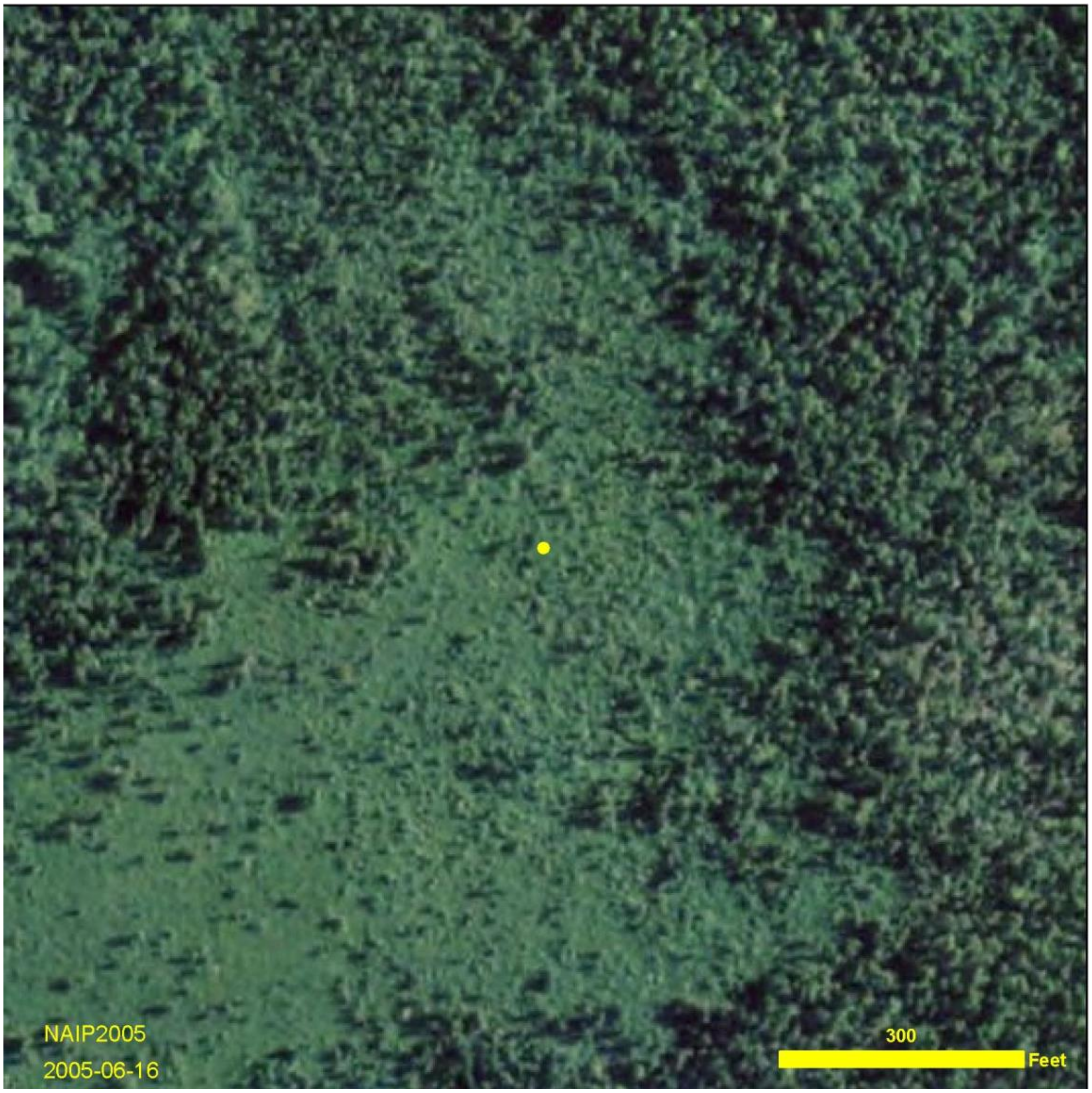
1. Providing the same data as are used in the auto-key procedures
2. Providing additional data that were not originally used in the auto-keys, and
3. Maintaining the "privacy" of the FIA data, ensuring the experts could not determine which plots were FIA vs not

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 7W.



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 7W 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 7W: Great Lakes & Central Highlands

GeoArea 7W encompasses 10 map zones (Figure 2): Northern Lake Country (41), Ozark Highlands (44), Appalachia Bluegrass Hills (47), Cumberland Highlands (48), Central Till Plains (49), Central Great Lakes Uplands (50), Great Lakes Plains (51), Eastern Till Plains (52), Appalachia (53), and Allegheny Plateau (62). These map zones were originally clustered for purposes of designing and implementing auto-keys (Figure 1). The total number of plots in this Geo Area analysis was 1,908. A total of 43 natural ecological system types were assigned to a total of 1,403 plots by the auto-keys. A total of 69 system types were assigned by experts (i.e., these included individual types that had been aggregated to broader classes by LANDFIRE for either sparsely vegetated types or wetland/riparian types).

An additional 11 types were assigned by the auto-key but were not assigned by experts:

- Crosstimbers Oak Forest and Woodland
- Lower Mississippi River Dune Woodland and Forest
- Mississippi Delta Maritime Forest
- Southern Atlantic Coastal Plain Dune and Maritime Grassland
- Texas Saline Coastal Prairie
- Texas-Louisiana Coastal Prairie Pondshore
- Caribbean Swamp Systems
- Gulf and Atlantic Coastal Plain Floodplain Systems
- Gulf and Atlantic Coastal Plain Small Stream Riparian Systems
- Gulf and Atlantic Coastal Plain Swamp Systems
- Gulf and Atlantic Coastal Plain Tidal Marsh Systems

The first type is possibly in the GeoArea. The next five types do not occur in GeoArea 7W and should not be attributed in it. The final five types are aggregates of individual Systems used by Landfire for mapping. The expert reviewers attributed sites to individual Systems and so would not have used these units in their review process.

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 43 natural types assigned labels by the auto-keys, 10 types (23%) had fewer than 10 samples available for this analysis (Table 2). Six of these types are probably truly rare in the GeoArea either because it is on the edge of their geographic range, because they are uncommon types throughout their range, or both. The East Gulf Coastal Plain Northern Loess Plain Oak-Hickory Upland, West Gulf Coastal Plain Nonriverine Wet Hardwood Flatwoods, Southeastern Interior Longleaf Pine Woodland, West Gulf Coastal Plain Pine-Hardwood Flatwoods, South-Central Interior/Upper Coastal Plain Wet Flatwoods, and North-Central Interior Oak Savanna types fit one or both of these criteria. The other four types - Great Lakes Wooded Dune and Swale, South-Central Interior/Upper Coastal Plain Flatwoods, Central Appalachian Alkaline Glade and Woodland, and Paleozoic Plateau Bluff and Talus are not abundant in the GeoArea but are probably underrepresented in the data compared to their abundance in the GeoArea.

Table 2. Under-sampled types within GeoArea 7W

EVT Code	EVT Name	System elcode	total Plots
2306	East Gulf Coastal Plain Northern Loess Plain Oak-Hickory Upland	CES203.482	9
2466	Great Lakes Wooded Dune and Swale	CES201.726	8
2326	South-Central Interior/Upper Coastal Plain Flatwoods	CES203.479	8
2506	West Gulf Coastal Plain Nonriverine Wet Hardwood Flatwoods	CES203.548	7
2351	Southeastern Interior Longleaf Pine Woodland	CES202.319	5
2458	West Gulf Coastal Plain Pine-Hardwood Flatwoods	CES203.278	5
2457	South-Central Interior/Upper Coastal Plain Wet Flatwoods	CES203.480	5
2400	Central Appalachian Alkaline Glade and Woodland	CES202.602	3
2517	Paleozoic Plateau Bluff and Talus	CES202.704	3
2394	North-Central Interior Oak Savanna	CES202.698	1

A total of 4 types (9% of 43 types) had >80% agreement between expert and auto-key assignments. All of these types had at least 10 sample plots. Expert self-assessment of confidence for these types was predominantly 'high' although Laurentian-Acadian Alkaline Conifer-Hardwood Swamp resulted more in 'moderate' and 'low' confidence levels. This type is very common in the northern Great Lakes states and northeastern US but can be difficult to attribute confidently based only on overstory data. Two common overstory species, northern white-cedar (*Thuja occidentalis*) and red maple (*Acer rubrum*), can also dominate in upland forests so hydrology, soil, or understory data are needed to confidently assign sites to this type.

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 29, or 67% of the total types assigned. Analysis of the nature of the disagreements between expert and auto-key attribution reveals some of the sources of these disagreements and suggests some methods to help reduce these in the future. The following are some specific examples of levels of disagreement and possible explanations based on interpretations from the contingency table.

Four common types had 66%-75% agreement. Three of these - North-Central Interior Dry-Mesic Oak Forest and Woodland, Southern Interior Low Plateau Dry-Mesic Oak Forest, and Laurentian-Acadian Northern Pine(-Oak) Forest - were most often confused with floristically similar types based on the apparent abundance of dominant species (oaks or pines). North-Central Interior Beech-Maple Forest was most often confused with a similar type that replaces it to the south – South-Central Interior Mesophytic Forest.

Allegheny-Cumberland Dry Oak Forest and Woodland – Fifteen sites (30%) auto-keyed to this type were called Northeastern Interior Dry-Mesic Oak Forest by experts. These are very similar types distinguished primarily by the abundance of oaks favoring dry conditions versus dry-mesic conditions but there is significant overlap in the component species.

North-Central Interior Wet Flatwoods – Thirty sites were auto-keyed to this type and experts disagreed on 13 sites. Nine of these were listed as "Can't assign" or "Other"; designations that were used indicate not enough data to assign the type or that the site did not fit any natural System. The auto-key may have been too aggressive in assigning sites based on limited data and may have included non-natural sites.

Laurentian-Acadian Northern Hardwoods Forest – This type was confused with several other upland forests though no single one was common. The most common confusion was with the "Can't assign" or "Other" expert-assigned categories indicating not enough data to assign the type or that the site did not fit any natural System. The auto-key may have been too aggressive in assigning sites based on limited data and may have included non-natural sites.

Northeastern Interior Dry-Mesic Oak Forest – This type was confused primarily with two others – South-Central Interior Mesophytic Forest (15 times or 30%) and Appalachian (Hemlock-) Northern Hardwood Forest (7 times or 14%). Confusions with these types are likely based on the relative abundance of oaks vs. other mesic trees.

Ozark-Ouachita Dry Oak Woodland – Forty-four percent of the auto-keyed sites were assigned by experts to either Ozark-Ouachita Dry-Mesic Oak Forest (22%) or "Cant' assign" (22%). The Ozark-Ouachita Dry-Mesic Forest and Ozark-Ouachita Dry Oak Woodland grade into each other and the difference can be based on slight differences in cover of oak species.

South-Central Interior Mesophytic Forest – This type was confused with seven other types but most commonly with Southern Interior Low Plateau Dry-Mesic Oak Forest (25%). This is likely due to differences in the relative abundance of oak species.

North-Central Interior Dry Oak Forest and Woodland – This type was commonly confused with North-Central Interior Dry-Mesic Oak Forest and Woodland (36%). These two types grade into each other and the difference can be based on slight differences in cover of different oak species.

Ozark-Ouachita Dry-Mesic Oak Forest – This type was commonly (46%) confused with Ozark-Ouachita Dry Oak Woodland. These two types grade into each other and the difference can be based on slight changes in the cover of oak species.

Laurentian-Acadian Pine-Hemlock-Hardwood Forest – Most of the sites (56%) auto-keyed to this type were assigned to Laurentian-Acadian Northern Hardwoods Forest. This confusion is likely based on interpretations of the relative abundance of pines and hemlock.

Boreal White Spruce-Fir-Hardwood Forest – Forty percent of the sites auto-keyed to this type were assigned to either the Laurentian-Acadian Alkaline Conifer-Hardwood Swamp (20%) or "Can't assign" (20%). The Laurentian-Acadian Alkaline Conifer-Hardwood Swamp can have a strong component of balsam fir (*Abies balsamifera*), red maple (*Acer rubrum*), and quaking aspen (*Populus tremuloides*) as does this type. Understory data or other information that would indicate wetlands versus uplands would help differentiate these types.

North-Central Interior Maple-Basswood Forest – Thirty-eight percent of the sites auto-keyed to this type were classified as "Can't assign" by the expert reviewer. Most of these were sites that appeared on the aerial photo as not natural types. Of the sites assigned to natural types by the expert, most (18%) were confused with North-Central Interior Dry-Mesic Forest and Woodland. These types grade into each other, particularly from North-Central Interior Dry-Mesic Forest and Woodland into North-Central Interior Maple-Basswood Forest in the absence of fire.

Boreal Aspen-Birch Forest – This type was auto-keyed in areas it should not occur, areas outside the range of boreal types. Early successional sites in the northern Midwest are often dominated by quaking aspen (*Populus tremuloides*) or paper birch (*Betula papyrifera*) but sites must be in the boreal zone to fit this type. Thus, most sites auto-keyed to this type were assigned by experts to non-boreal types. A substantial number of auto-keyed sites (26%) were assigned by the expert to the “Can’t assign” category, indicating either not enough data were available or the site was not a natural type. Better delineation of the potential range of this type would improve the auto-key performance for this type.

Eastern Boreal Floodplain – No sites auto-keyed to this type matched the expert attribution. This type can be difficult to distinguish from surrounding uplands based solely on overstory data. Understory or hydrology data would help to identify sites of this type.

Central Interior Highlands Calcareous Glade and Barrens – All sites auto-keyed to this type were assigned by experts to upland forest and woodland types. Based solely on overstory characteristics, glades can easily be confused with dry or dry-mesic upland forests dominated by oaks. This can be seen in the fact that 67% of the sites auto-keyed to this type were assigned by experts to upland oak-dominated types. Aerial photos, understory data, and soils data could help clear up this confusion.

Central Interior Highlands Dry Acidic Glade and Barrens – All 10 of the sites auto-keyed to this type were assigned by experts to the Ozark-Ouachita Shortleaf Pine-Oak Forest and Woodland. This acidic glade can be similar floristically to this type but aerial photos, soil information, and understory data would help solve this confusion.

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**
2310	North-Central Interior Dry-Mesic Oak Forest and Woodland	CES202.046	50	37	74%	15	13	9
2305	Southern Interior Low Plateau Dry-Mesic Oak Forest	CES202.898	50	35	70%	29	5	1
2362	Laurentian-Acadian Northern Pine(-Oak) Forest	CES201.719	50	34	68%	9	9	16
2313	North-Central Interior Beech-Maple Forest	CES202.693	50	34	68%	10	12	12
2317	Allegheny-Cumberland Dry Oak Forest and Woodland	CES202.359	49	28	57%	20	7	1
2518	North-Central Interior Wet Flatwoods	CES202.700	30	17	57%	0	7	10
2302	Laurentian-Acadian Northern Hardwoods Forest	CES201.564	50	28	56%	11	6	11
2303	Northeastern Interior Dry-Mesic Oak Forest	CES202.592	50	22	44%	11	8	3
2364	Ozark-Ouachita Dry Oak Woodland	CES202.707	50	21	42%	11	7	3

				Plots with Expert Matches				
EVT Code	**EVT Name**	**System Elcode**	**Total Plots**	**Total**	**%**	**High conf**	**Med conf**	**Low conf**
2321	South-Central Interior Mesophytic Forest	CES202.887	50	20	40%	16	2	2
2311	North-Central Interior Dry Oak Forest and Woodland	CES202.047	50	18	36%	9	4	5
2304	Ozark-Ouachita Dry-Mesic Oak Forest	CES202.708	50	17	34%	13	3	1
2307	East Gulf Coastal Plain Northern Dry Upland Hardwood Forest	CES203.483	30	10	33%	10	0	0
2366	Laurentian-Acadian Pine-Hemlock-Hardwood Forest	CES201.563	50	15	30%	5	5	5
2365	Boreal White Spruce-Fir-Hardwood Forest	CES103.021	50	14	28%	2	1	11
2308	Crosstimbers Oak Forest and Woodland	CES205.682	15	3	20%	1	2	0
2314	North-Central Interior Maple-Basswood Forest	CES202.696	51	9	18%	2	6	1
2407	Laurentian Pine-Oak Barrens	CES201.718	49	7	14%	1	3	3
2323	West Gulf Coastal Plain Mesic Hardwood Forest	CES203.280	47	6	13%	2	3	1
2301	Boreal Aspen-Birch Forest	CES103.020	50	6	12%	0	0	6
2370	Appalachian (Hemlock-) Northern Hardwood Forest	CES202.593	50	6	12%	2	3	1
2334	Ozark-Ouachita Mesic Hardwood Forest	CES202.043	37	4	11%	3	1	0
2444	Eastern Boreal Floodplain	CES103.588	50	0	0%	0	0	0
2315	Southern Appalachian Oak Forest	CES202.886	49	0	0%	0	0	0
2309	Southern Appalachian Northern Hardwood Forest	CES202.029	46	0	0%	0	0	0
2401	Central Interior Highlands Calcareous Glade and Barrens	CES202.691	24	0	0%	0	0	0
2344	Boreal Jack Pine-Black Spruce Forest	CES103.022	14	0	0%	0	0	0
2409	Great Lakes Alvar	CES201.721	13	0	0%	0	0	0
2363	Central Interior Highlands Dry Acidic Glade and Barrens	CES202.692	10	0	0%	0	0	0

Expert Assignments

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

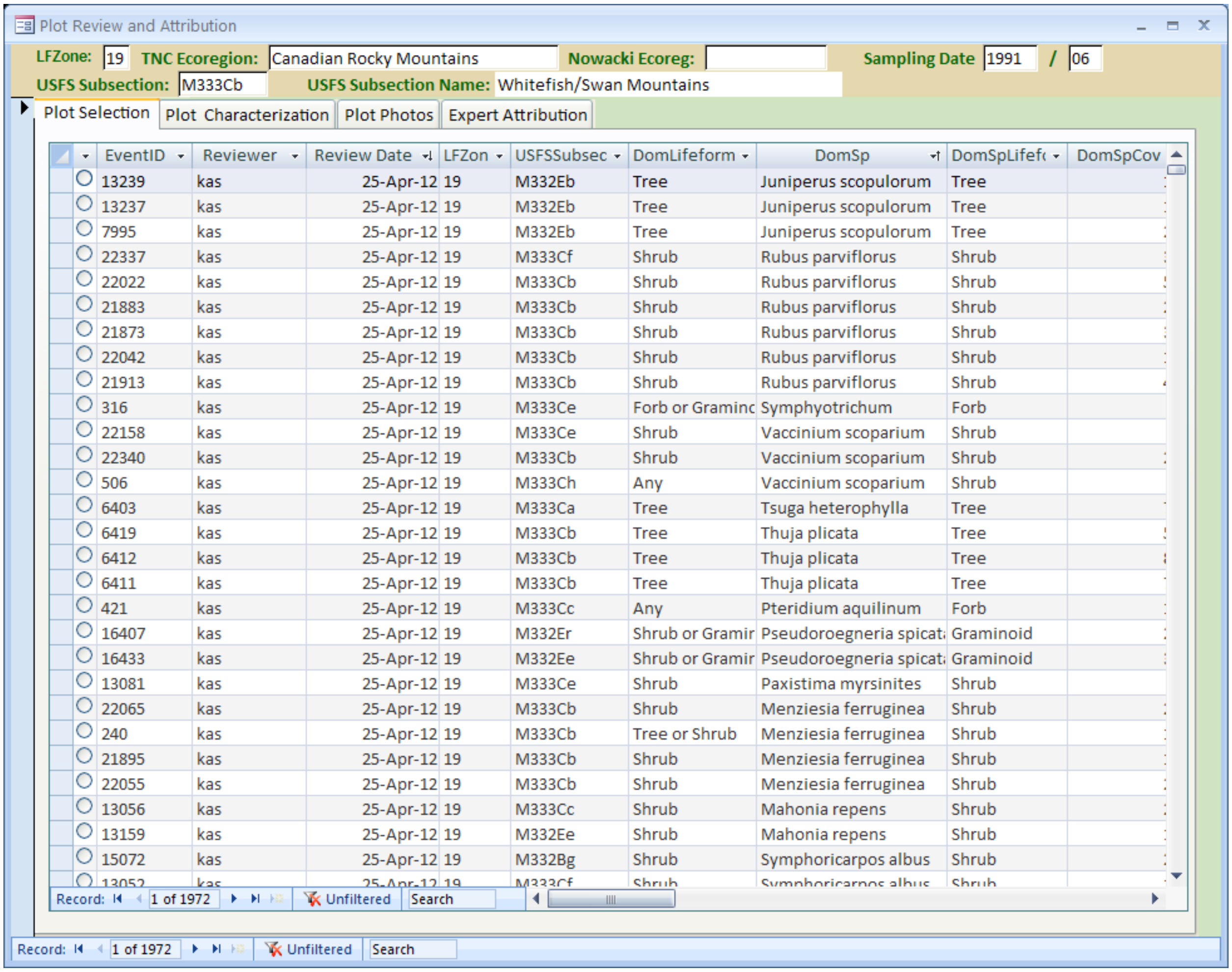
As described in the methods section above, the expert reviewers worked directly in the expert attribution database (EADB). Since GeoArea 7W had almost 2,000 plots to review, a systematic, efficient process for reviewing and labeling plots was required. The forms provided in the EADB allowed the reviewer 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 central, glaciated areas of this GeoArea, forests can transition from dry-mesic oak-dominated to mesic sugar maple-dominated. In these areas, the tree canopy could be a mix of white oak, red oak, and sugar maple in just about any combination. The reviewer would have to determine whether plots represented North-Central Interior Maple-Basswood Forest, North-Central Interior Beech-Maple Forest, or North-Central Interior Dry-Mesic Forest and Woodland.

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



Plot Review and Attribution

LFZone: 19 TNC Ecoregion: Canadian Rocky Mountains Nowacki Ecoreg: Sampling Date 1991 / 06

USFS Subsection: M333Cb USFS Subsection Name: Whitefish/Swan Mountains

Plot Selection | Plot Characterization | Plot Photos | Expert Attribution

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

Record: 1 of 1972 Unfiltered Search

Record: 1 of 1972 Unfiltered Search

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

a) review of the image clip for the context of the plot,
b) review of where the plot was located geographically (USFS Subsections provide local scale geographic location), to distinguish North-Central Interior Beech-Maple Forest from North-Central Interior Maple-Basswood Forest.
c) consideration of topographic setting (e.g. north-facing slopes indicate the more mesic maple-dominated types versus south-facing or steep slopes indicating the dry-mesic oak type)
d) consideration of any available height data for the plot (e.g. a canopy of taller oaks over maples indicates the oak type, though in transition to one of the maple types),
e) careful consideration of the full floristic composition of the plot and cover for each species.
f) awareness of possible errors in the plot data, such as mis-identification of oak species by the field crews, unevenness in how the cover values were estimated in the field or converted into the LFRDB (e.g. cover for trees estimated by a person standing on the ground vs. an aerial view of the plot).

Below are some examples of comments relevant to the above example:

- Composition is not typical (no *Quercus* spp.) and area looks disturbed on photo. Possibly not a natural System.
- Not much *Quercus* spp. and this could be too dry for this System. Lots of disturbance nearby so this site could be a non-natural System, too.
- Only *Quercus macrocarpa* is characteristic of this System. Other species not strongly characteristic but still fit moderately well.
- Composition could fit this System or NCI Maple-Basswood Forest. Moderately steep SW-facing slope fits this *Quercus*-dominated System better.
- This could also be one of the mesic maple systems; it is in the range for North Central Interior Beech Maple forest (look for *Fagus* in rest of polygon). Although unlikely, it is possible that it could be North Central Interior Maple Basswood Forest.
- Almost enough *Quercus* spp. (particularly *Q. muehlenbergii*) to fit NCI Dry-Mesic Oak Forest and Woodland.
- Dominated by early successional trees but some trees characteristic of this System. Possibly too disturbed to be a natural System.
- Little *Quercus* spp. present but *Carya ovata* can be part of this System. Site is likely quite disturbed and may not be a natural System.
- *Quercus ellipsoidalis* tends to indicate a drier setting than this System but probably not enough *Q. ellipsoidalis* to make this not fit.
- Mix of *Acer saccharum* and *Quercus alba* almost allow this to fit NCI Beech-Maple.
- Area is near the border of the NCI Dry-Mesic Oak Forest and Woodland and Southern Interior Low Plateau Dry-Mesic Oak Forest but this Section should be just NCI Dry-Mesic Oak Forest and Woodland. Also, quite a bit of *Acer saccharum* for this System

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 69 types assigned to plots by experts, 32 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 37 types, the numbers of samples labeled to a given type ranged from 143 (for Laurentian-Acadian Alkaline Conifer-Hardwood Swamp) down to 10 (for West Gulf Coastal Plain Small Stream and River Forest). For 34 (91%) of these types, experts reported at least moderate confidence in their labels for at least 20% of the type's plots. 14 had 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
Laurentian-Acadian Alkaline Conifer-	This may be a seep or other wet area within larger Laurentian

Hardwood Swamp	Acadian Northern Hardwood matrix
Laurentian-Acadian Northern Hardwoods Forest	Dominated by aspen-birch but not boreal, rather it is early successional.
Ozark-Ouachita Dry Oak Woodland	Canopy coverage is high for this system
Boreal-Laurentian Conifer Acidic Swamp	This is a little out of range but seems best choice
North-Central Interior Wet Flatwoods	The composition matches this System but could also indicate a floodplain.
Boreal White Spruce-Fir-Hardwood Forest	This borders what appears to be wet area and could lean more towards Laurentian - Acadian Alkaline Hardwood Conifer Swamp

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. For example, elucidating a brighter line between characteristics of the various dry-mesic and dry oak forests would help solve much of the confusion between those 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. In this GeoArea there were many sites that were auto-keyed to a natural type but which the expert reviewer assigned to a non-natural type, often based on the aerial photograph. Small woodlots in an agricultural field, old fields with scattered tree regeneration, or fencerows might have overstory composition that fits a natural type but they are clearly ruderal or cultural based on their origin, size, and surroundings. Similarly, repeated references to photos further indicates the need for expert review of many types where moderate-low confidence of experts suggest that auto-keys might be prone to error. Third, additional floristic information is cited in some cases where their suspected limitations provide the primary source of expert uncertainty in labeling.

Other samples were labeled by auto-keys to aggregates of multiple ecological system types. This was because LANDFIRE had mapping objectives focused on uplands where fire regimes are prevalent. That meant that many individual wetland and sparsely-vegetated ecological system types were not treated within the auto-keys. Expert labeling of these samples, however, provides an indication of the feasibility of their inclusion in updated auto-keys. Of 286 samples, experts were able to assign 261 (68%) to an individual ecological system type; a total of 36 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 Shrubland" or "None". Of 119 samples, experts were able to assign 40 (34%) to an individual ecological system type; a total of 18 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 139 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 6 with no NVC Group assignment) nest cleanly within 45 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 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. However, within this GeoArea, no ecological system types have a more complex relationship with NVC Group concepts (Table 5).

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

There are no GeoArea 7W 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, and also compared the percent of expert-autokey matches at the system level versus the Macrogroup level (Table 6).

There are 12 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, rolling up plot attribution to the Central Oak-Hardwood & Pine Forest Macrogroup increases the percent of matches of auto-key versus expert reviewer assignments from 45% to 75%. That Macrogroup includes nearly all the common, oak-dominated dry and dry-mesic forests from the western Appalachians west through the glaciated Midwest and unglaciated Interior Low Plateau to the oak woodlands scattered in the eastern Great Plains. Tightening up the geographic range parameters and characteristic relative abundances of the dominant oaks would yield significant improvements in the attribution correspondence while maintaining the ecological detail available using Systems. There are five cases where rolling up results to the Macrogroup level gains little in terms of increasing the percentage of matches between auto-key and expert attributions but still results in a loss of ecological detail. Many of the disagreements between auto-keyed and expert assigned attributions were due to the expert not being able to assign sites to a natural System due to a lack of sufficient data or because the site was highly disturbed and fell within a ruderal or cultural type. Adding some kind of check on the data to reduce these occurrences would increase the agreement in attribution.

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
M008 Southern Mixed Deciduous-Evergreen Broadleaf Forest	1	47	13%	13%
M012 Central Oak-Hardwood & Pine Forest	6	261	45%	75%
M014 Acer saccharum - Betula alleghaniensis - Pinus strobus - Tsuga canadensis Forest	5	199	25%	38%
M016 Southern Hardwood & Pine Forest	8	259	52%	79%
M153 Acer (barbatum, saccharum) - Tilia americana - Fagus grandifolia - (Liriodendron tulipifera) Forest	5	209	40%	49%
M159 Northern & Eastern Pine - Oak Forest & Barrens	2	99	41%	63%
M030 Northern & Central Swamp Forest	5	102	62%	65%
M033 Southern Coastal Plain Basin Swamp	2	12	25%	50%
M037 Eastern & Central North American Boreal Conifer & Hardwood Forest	3	114	18%	21%
M300 North American Boreal Flooded Forest	1	50	0%	0%
M054 Great Plains Tallgrass Prairie, Savanna & Shrubland	1	1	0%	0%
M124 Northern & Central Alvar & Glade	4	50	0%	0%

US-NVC Divisions

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

Discussion

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

Fundamentally, a sequence table as used by LANDFIRE is a set of criteria. Each vegetation sample has to meet some combination of criteria in the SQT to be labeled with an ecological system, or some other land cover class. Simply put, if the plot doesn't meet any criteria contained in the sequence table, then it may be mis-keyed, or not key to anything. Given our incomplete knowledge of the structural and floristic variability of each classification unit, it is nearly impossible to establish criteria in a sequence table - for regional application - to successfully and accurately key 100% of vegetation samples. However, with new field-based inventory and increasing ecological understanding, over time sequence tables can be revised and improved so as to accurately key increasing percentages of vegetation samples.

There are a number of reasons why a sequence table may not successfully key all samples run through it:

a) the unknown floristic quality of the vegetation data (how complete, how well collected, does it accurately represent the vegetation concept being keyed);
b) our limited knowledge of the variability in species composition, vegetation structure, and the distribution of ecological systems; and
c) the comprehensiveness (or lack thereof) in field inventory for any particular system (e.g., many from one small area, few to none from elsewhere in the region).

Each of these are discussed below.

A. Quality of vegetation data

First and foremost, the completeness and quality of the data as collected in the field, as well as the documentation of how the data were collected (the metadata) are primary issues for how well the sequence table process works. There are many different kinds of issues with the data collection, only a few of which are listed here as possible sources of problems:

- Was the species composition adequately sampled (complete species list)?
- Were only trees recorded (e.g., some FIA plots)? Only "dominant" or "most characteristic" species (e.g., SWReGAP training data)?
- Was the sample plotless, or within a plot or some other measured area?
- Or were the samples derived along transects?
- How was the cover or abundance data collected, or was it presence/absence?
- Was the sample area across an ecotone (for example across the transition from a wet valley bottom into the adjacent upland slope)?
- Does the sample adequately represent an occurrence of the vegetation type being sampled?
- Was the species taxonomy accurately recorded (many species are difficult for untrained crews to identify, such as *Carex* spp., or *Salix* spp.)?
- Were difficult species "lumped up" into broader taxon, such as genus, or even family?
- Was the sample location heavily or recently disturbed?

Many datasets obtained by the LANDFIRE team had inadequate metadata associated with them. Inadequate documentation of the sampling design or of what the values in the data tables represented, could result in incorrect processing of the data for use in the sequence tables.

The sampling design under which vegetation data was collected is an often neglected piece of metadata. A particular dataset could have many hundreds of plots in it, but the purpose(s) for which they were collected could be such as to negate their value for identifying floristically distinct vegetation types. For example, samples collected in a systematic grid without regard for sampling distinct vegetation types will often cross multiple ecological systems, and hence result in data that give erroneous results in an auto-key process.

An example of poor documentation of the collection protocols would include species names collected and provided as 4- or 6-letter acronyms, without a complete list of what species each acronym represents. The processing of the data into the LFRDB converts acronyms to full species utilizing the current NRCS PLANTS 'symbols'. So, POTR could be *Populus tremuloides*, *Poa tracyi,* or *Poa trivialis*, all valid species. But using PLANTS, POTR = *Poa tracyi*, while *Populus tremuloides* is POTR5. Each dataset has to be reviewed for its species taxonomy to ensure any acronyms are converted to the correct taxa, but without adequate metadata errors can creep in.

Another example would be where the species abundance data were collected in generalized "cover classes", and these had to be converted to "real cover" by using the mid-point of the class. If the metadata did not include documentation of what the classes represent, then the mid-points could be incorrectly converted, or even unobtainable. For example, cover class 3 could mean 5-25% cover (mid-point of 15%), or it could mean 25-35% cover (mid-point 30%). The sequence table process utilizes cover criteria for indicator species extensively, so incorrectly interpreted cover classes will lead to errors in the results.

B. Constraints within sequence table

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

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

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

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

C. Developing automated keys for large geographic areas

Each sequence table was constructed to work across relatively large geographic areas, on the order of 2-5 USFS Sections (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 is 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. In this GeoArea there are many highly disturbed or managed sites that may be dominated by native trees but that do not fit the criteria for a natural System. These include farm woodlots, old fields, pastures, tree plantations, drainage ditches, fencerows, wooded yards in suburbs/exurbs or cities, etc. These can all appear to fit the auto-key if it considers just dominant overstory species. Some kind of landscape analysis might screen these out without requiring a complete species list. Some kind of analysis incorporating proximity to a river/stream of a certain size might also assist in assigning a site to a floodplain System versus a basin wetland (the same associations can occur in either).

Narrowing vs. broadening the geographic application of the auto-key – FS Sections? In certain areas? Would this likely lead to greater accuracy? FS Section data would not be very helpful. Subsection data is much more useful. Some GeoAreas also used EPA Level IV Ecoregions.

Adjustment to auto-keys – additional requirements for vegetation sample data; primarily ground cover data. Auto-keys need to accurately reflect the potential geographic range of types. Several types were attributed using auto-keys outside of their range. Subsection and ecoregion data can help with this.

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

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

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

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