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

FINAL REPORT

Prepared by
NatureServe
For the NPS Vegetation Inventory Program & LANDFIRE

30 June 2012

© Keith Schulz. Waikolu Valley, Island of Molokai'i, Hawai'i.

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 Keith Schulz, NatureServe Regional Vegetation Ecologist.

Table of Contents

Figures

Tables

Introduction

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

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

Project Goals

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

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

Background on LANDFIRE Auto-keys

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

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

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

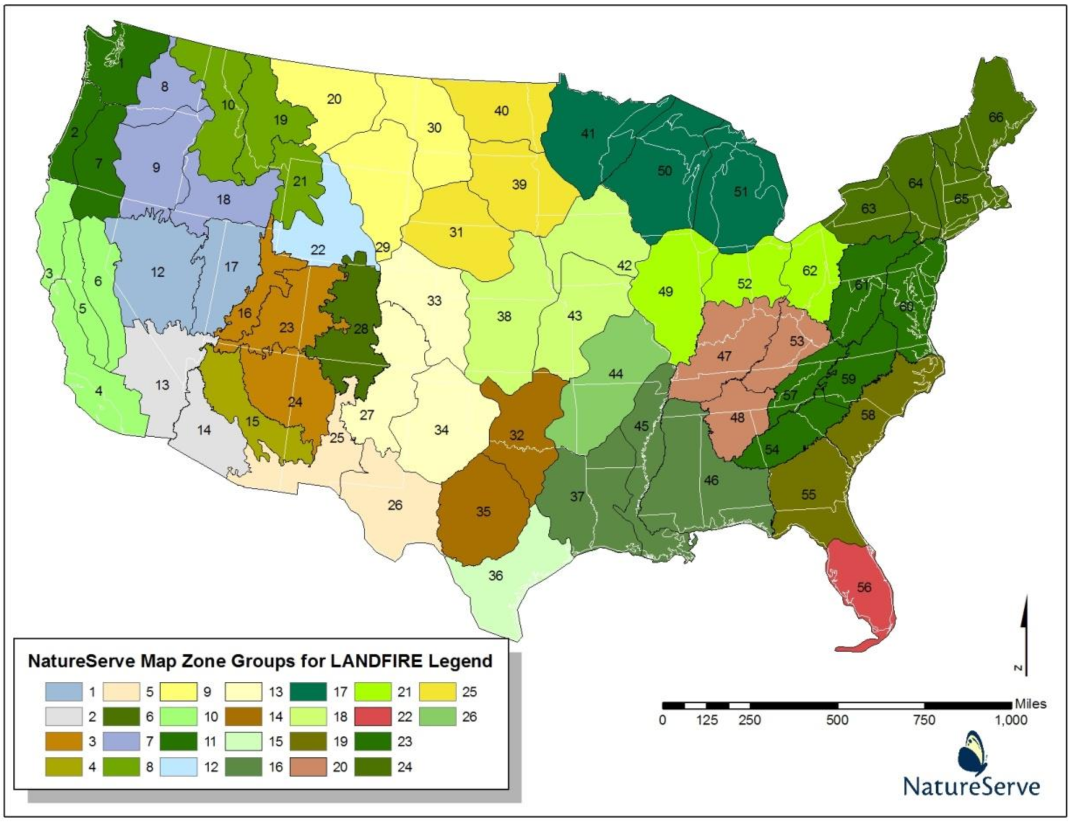
LANDFIRE's short-term needs, and long-term plans, required a repeatable methodology, consistently applied rules to categorize each reference sample, and documentation of the criteria applied. In essence, sequence tables codify the criteria and methods for keying geo-referenced vegetation data to a land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

More details about this methodology include:

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

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

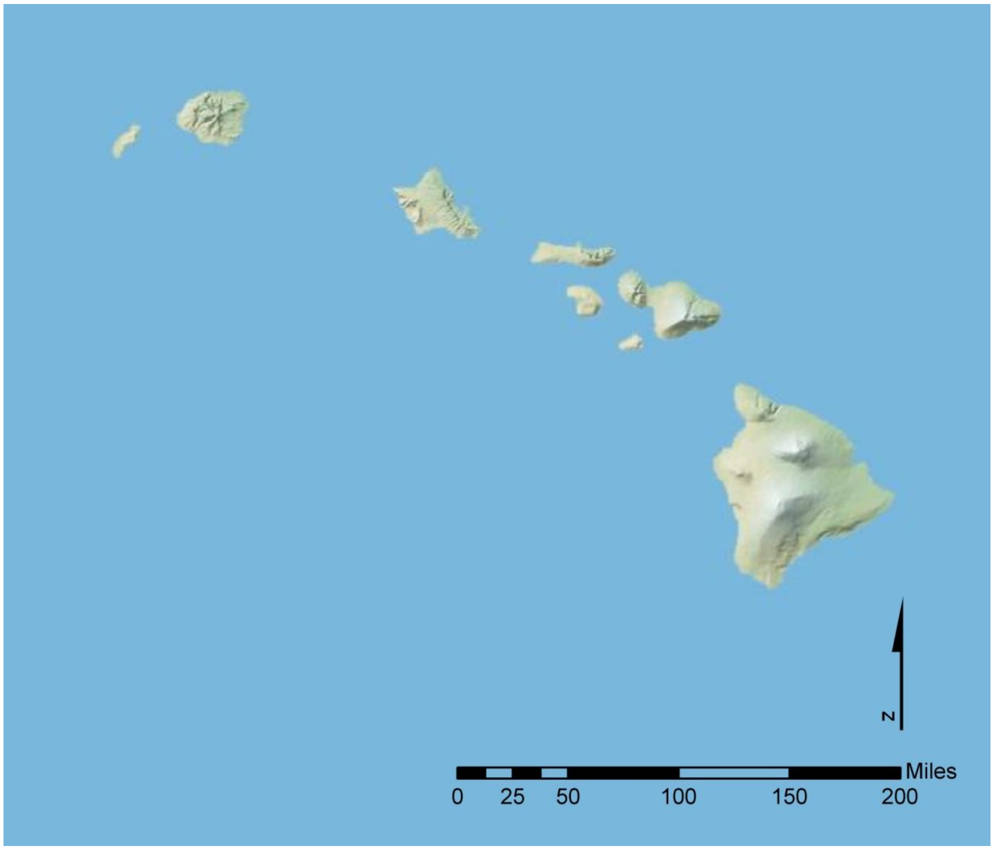
Figure 1. Groups of MRLC map zones that were the analysis units for the LANDFIRE sequence tables in the coterminous U.S.



Methods

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

Figure 2. LANDFIRE Hawai'i GeoArea 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 from NAIP imagery, however the quality of available NAIP imagery was so poor in the Hawaiian GeoArea (low resolution and clouds) that the decision was made to provide actual plot coordinates and use Google-Earth or other satellite imagery that could be viewed with GIS. Figure 3a is an example of the original NAIP imagery provided that was unusable. Figure 3b is an example of the ESRI World View imagery which used GeoEye IKONOS 1m resolution imagery for Hawai'i.

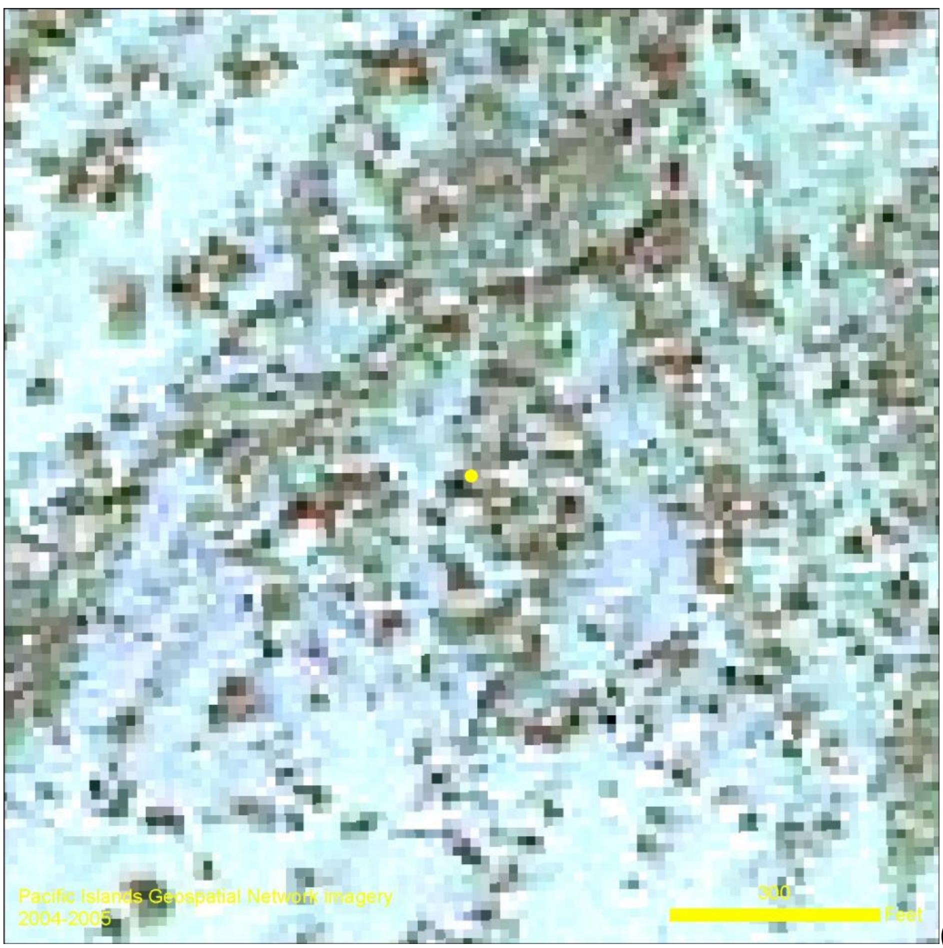
In addition, NatureServe used moisture zones developed by Price et al. 2007 to help define several Ecological Systems in the Hawaiian GeoArea. These moisture zones were used to label plots both in the auto-key and during expert review. The Price et al. (2007) moisture zones were not included in the HI GeoArea database, so they had to be viewed while reviewing the imagery using a GIS. These necessary modifications in methodology were a small departure from standard methodology used in GeoAreas from the lower 48 states.

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

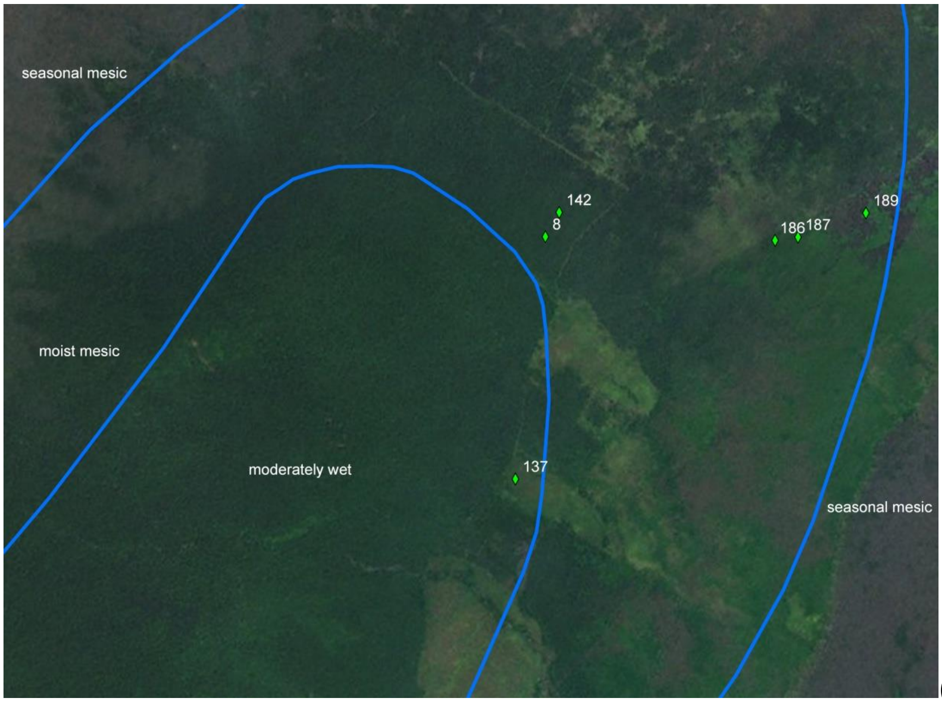
Data category	Fields	Notes
Vegetation Structure	% cover of trees, shrubs, herbs, trees per acre, height of trees or shrubs	Values are calculated from source data & stored in LFRDB
Floristic composition	complete species list, % cover by species, nativity, height if available	Species list & % cover values are from the original source data, but other fields were derived by LANDFIRE
Dominant species	the 2 most dominant species within the major lifeform of the plot	The dominant and codominant species are provided, with % cover; the species are drawn from the dominant lifeform category of the plot (e.g. shrub dominated plots will have shrub species listed)

Data category	Fields	Notes
Geographic setting	map zone, USFS subsection, TNC ecoregion	These are derived by LANDFIRE from the coordinates of the plot
Landscape setting	elevation, aspect, slope	Values are derived form a DEM for the coordinates of the plot
Moisture zone (HI only)	GIS layer used when viewing aerial imagery and plot location in GIS	The seven moisture zones developed by Price et al. 2007 were simplified to 3 and used to help define Ecological Systems for Hawai'i.
Field notes	comments from field crew	Original field crew comments, if available
Imagery with clips (HI only)	Single image, same areal extent/scale for all plots	Google Earth or ESRI World View imagery was used for Hawaiian plots;
Image clips **(not used in HI)**	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 (a) a NAIP image clip in GeoArea HI and (b) a WorldView image clip for plot in GeoArea HI



(a)



(b)

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 HI 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 HI: Hawai'i

GeoArea HI encompasses the Hawaiian Island system, map zone 79. The total number of plots in this Geo Area analysis was 248. A total of 12 natural ecological system types were assigned to a total of 248 plots by the auto-keys. A total of 11 system types were assigned by experts. In this GeoArea there were no aggregated systems types for either sparsely vegetated types or wetland/riparian types.

An additional 2 types were assigned by the auto-key but were not assigned by experts:

- Hawai'i Dry Cliff
- Hawai'i Coastal Mesic Forest

Both types unassigned by experts are of very limited extent in Hawai’i.

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 12 natural types assigned labels by the auto-keys, 6 types (50%) had fewer than 10 samples available for this analysis (Table 2). These under-sampled types tended to include types that are less common such as Northern Polynesia Tidal Salt Marsh and Hawai'i Coastal Mesic Forest, types that are difficult to sample such as Hawai'i Dry Cliff, and other types that simply have had inadequate sampling effort across this region such as Hawai'i Dry-Site Lava Flow. These include Hawai'i Lowland Dry Shrubland and Hawai'i Lowland Mesic Shrubland. The extent and condition of native lowland vegetation has been significantly impacted by development and invasive, non-native species.

Table 2. Under-sampled types within GeoArea HI

EVTCode	EVT Name	System elcode	total Plots
2817	Hawai'i Lowland Dry Shrubland	CES412.409	9
2818	Hawai'i Lowland Mesic Shrubland	CES412.412	7
2812	Hawai'i Coastal Mesic Forest	CES412.417	7
2831	Hawai'i Dry-Site Lava Flow	CES412.416	6
2807	Northern Polynesia Tidal Salt Marsh	CES412.224	4
2825	Hawai'i Dry Cliff	CES412.414	1

No types had >47% agreement between expert and auto-key assignments. This is surprising how poorly the auto-key performed when compared to other GeoAreas.

Table 3 provides a summary of adequately-sampled types where agreement between expert and auto-key ranged from just below 50% down to 6%. These six types represent 50% of the total types assigned. Further analysis of those grouped within the 50-6% agreement range suggests subtleties within types that left the expert with greater or lesser confidence in their assignment. This confusion is a result of many factors, some having to do with the fact that much of the dominant species in Hawai'i occur in multiple systems, so differentiation is based on indicator species and environmental factors such as moisture zone and elevation. For example, the widespread often dominant tree, *Metrosideros polymorpha* occurs from near sea level to subalpine elevations and on dry to wet sites. Using indicator species is problematic when floristic composition is poorly known, species lists in the samples are incomplete, or with indicator species that are uncommon enough that they are not consistently present in every plot. Another big issue, especially in lowland tropical vegetation (<1000 m elevation), is introduced species which complicate the use of auto-keys to label plots but dilute the relative cover of key native species.

Table 3. Summary of types with adequate samples where agreement between auto-key and expert was below 50%

				Plots with Expert Matches				
EVT Code	EVT Name	System elcode	total Plots	total	%	High conf	Med conf	Low conf
2826	Hawai'i Dry Coastal Strand	CES412.418	17	8	47%	4	3	1

			Plots with Expert Matches					
EVT Code	EVT Name	System elcode	total Plots	total	%	High conf	Med conf	Low conf
2816	Hawai'i Montane-Subalpine Mesic Forest	CES412.406	78	31	40%	0	31	0
2813	Hawai'i Lowland Dry Forest	CES412.408	18	6	33%	2	3	1
2814	Hawai'i Lowland Mesic Forest	CES412.411	34	5	15%	0	4	1
2808	Hawai'i Lowland Rainforest	CES412.226	36	3	8%	1	1	1
2810	Hawai'i Montane Rainforest	CES412.215	31	2	6%	0	2	0

Analysis of the contingency table (Results Workbook) for these types with lesser levels of agreement reveals the many ongoing challenges with finding agreement between experts and auto-keys for complex vegetation types. Here we summarize the results from GeoArea HI

Hawai'i Montane-Subalpine Mesic Forest – was most often confused with Hawai'i Lowland Mesic Forest (32), followed by Hawai'i Montane Rainforest (6), Hawai'i Lowland Dry Forest (4), and Hawai'i Lowland Rainforest (4). This confusion is the result of the elevation break used in the auto-key. The lowland mesic forest was allowed to include plots with <1100 m, when the expert generally used a 1000 m break unless there were ecological justification for increasing it. The confusion between lowland dry forest and lowland rainforest were likely the result of borderline issues with the moisture zone breaks.

Hawai'i Lowland Rainforest – was most often confused with Hawai'i Montane-Subalpine Mesic Forest (16), followed by Hawai'i Montane Rainforest (9), and Hawai'i Lowland Mesic Forest (7). This indicates issues with the elevation zones and moisture zones where the auto-key labels the plots <1100 m first, but the expert thinks these borderline plots are montane. The moisture zones are also an issue which may hinge on Moisture Zone 5 being transitional between wet forest and mesic forest systems.

Hawai'i Lowland Mesic Forest – was most often confused with Hawai'i Lowland Dry Forest (14), followed by Hawai'i Montane-Subalpine Mesic Forest (9) and Hawai'i Montane Rainforest (5). Again, this indicates an issue with the elevation zones and moisture zones where the auto-key labels the plots <1100 m first, but the expert thinks these borderline plots are montane. The moisture zones are also an issue which may hinge on Moisture Zone 5 being transitional between wet forest and mesic forest systems.

Hawai'i Montane Rainforest– was most often confused with Hawai'i Montane-Subalpine Mesic Forest (14), followed by Hawai'i Lowland Dry Forest (5), Hawai'i Lowland Rainforest (2), and other non auto-key systems (8). The biggest issue is the level 5 moisture zone (seasonally mesic) and its confusion with the montane rainforest system. Moisture Zone 5 may be transitional between wet forest and mesic forest systems, rather than solely mesic.

Hawai'i Lowland Dry Forest– was confused with Hawai'i Lowland Mesic Forest (3), followed by Hawai'i Lowland Rainforest (2), Hawai'i Montane-Subalpine Mesic Forest (1), and other, non auto-key systems (6). The moisture zones in the auto-key need to be reviewed.

Hawai'i Dry Coastal Strand – was most often confused Northern Polynesia Tidal Salt Marsh (4). Both systems occur along coast. The indicator species may need to be restricted more or the order in the auto-key swapped so that the more restrictive brackish water species of the tidal marsh get labeled first.

Expert Assignments

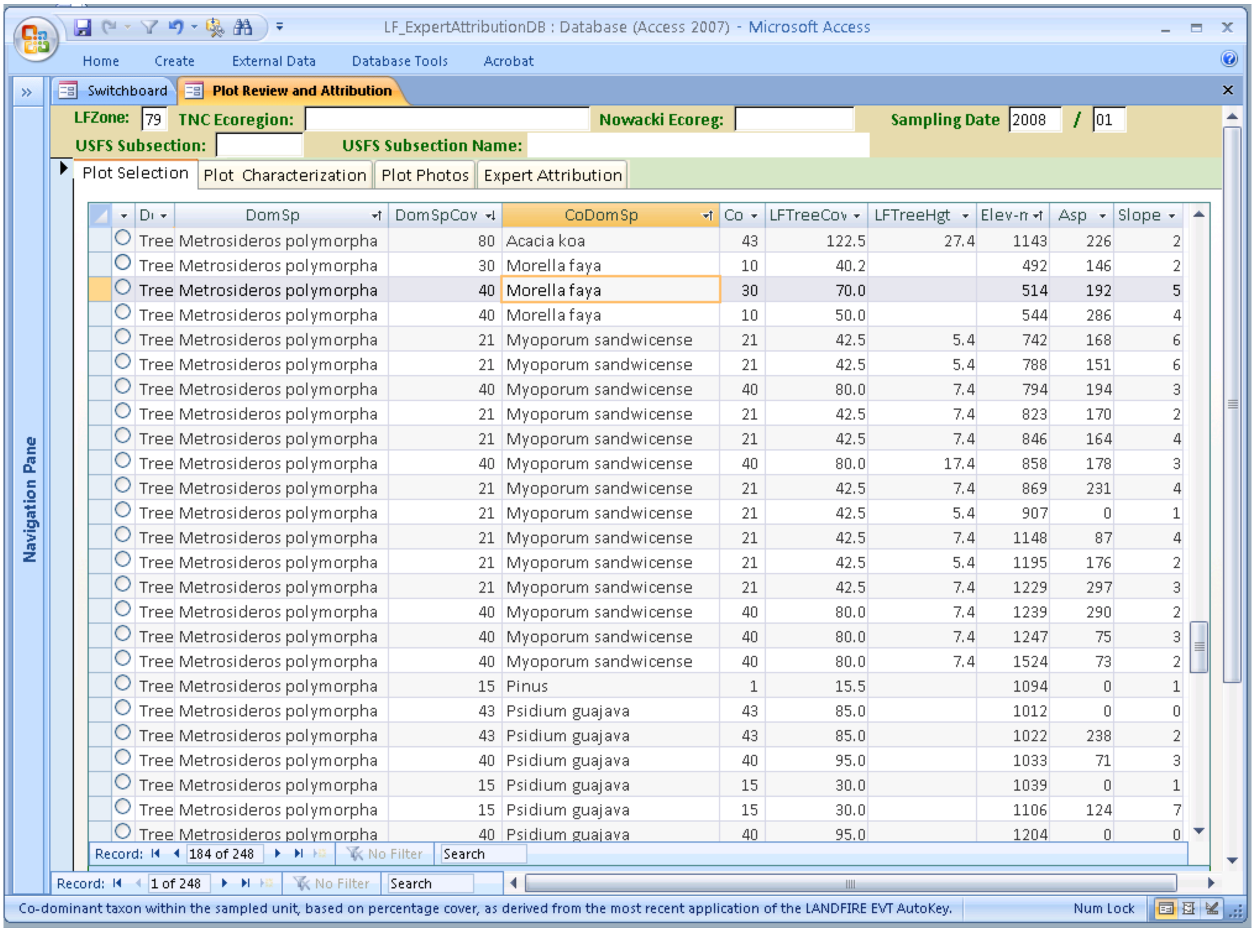
2.2 A report by each GeoArea detailing processes and results, specifically identifying how they made individual assignments.

As described in the methods section above, the expert reviewers worked directly in the expert attribution database (EADB). Since GeoArea HI had over 200 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 *Metrosideros polymorpha*), then sort by codominant species, from high to low. Figure 4 shows the main form in the EADB which has these data fields. The filter is also useful to narrow selections. Additional fields were provided from which to select or sort plots, such as % cover of species, 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 Hawai'i there are several species such as *Metrosideros polymorpha* which occur widely and dominate the canopy of several ecological systems.

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



Di	DomSp	DomSpCov	CoDomSp	Co	LFTreeCov	LFTreeHgt	Elev-m	Asp	Slope
Tree	Metrosideros polymorpha	80	Acacia koa	43	122.5	27.4	1143	226	2
Tree	Metrosideros polymorpha	30	Morella faya	10	40.2		492	146	2
Tree	Metrosideros polymorpha	40	Morella faya	30	70.0		514	192	5
Tree	Metrosideros polymorpha	40	Morella faya	10	50.0		544	286	4
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	5.4	742	168	6
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	5.4	788	151	6
Tree	Metrosideros polymorpha	40	Myoporum sandwicense	40	80.0	7.4	794	194	3
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	7.4	823	170	2
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	7.4	846	164	4
Tree	Metrosideros polymorpha	40	Myoporum sandwicense	40	80.0	17.4	858	178	3
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	7.4	869	231	4
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	5.4	907	0	1
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	7.4	1148	87	4
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	5.4	1195	176	2
Tree	Metrosideros polymorpha	21	Myoporum sandwicense	21	42.5	7.4	1229	297	3
Tree	Metrosideros polymorpha	40	Myoporum sandwicense	40	80.0	7.4	1239	290	2
Tree	Metrosideros polymorpha	40	Myoporum sandwicense	40	80.0	7.4	1247	75	3
Tree	Metrosideros polymorpha	40	Myoporum sandwicense	40	80.0	7.4	1524	73	2
Tree	Metrosideros polymorpha	15	Pinus	1	15.5		1094	0	1
Tree	Metrosideros polymorpha	43	Psidium guajava	43	85.0		1012	0	0
Tree	Metrosideros polymorpha	43	Psidium guajava	43	85.0		1022	238	2
Tree	Metrosideros polymorpha	40	Psidium guajava	40	95.0		1033	71	3
Tree	Metrosideros polymorpha	15	Psidium guajava	15	30.0		1039	0	1
Tree	Metrosideros polymorpha	15	Psidium guajava	15	30.0		1106	124	7
Tree	Metrosideros polymorpha	40	Psidium guajava	40	95.0		1204	0	0

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

a) review of the image clip or photo for the context of the plot,
b) review of where the plot was located geographically (wet windward side of island or dryer leeward side of island. In Hawai'i the USFS Subsections and do not vary so not useful.
c) consideration of topographic setting (e.g. alpine, subalpine, montane, lowland and coastal zones all support different vegetation),
d) consideration of any available height data for the plot (e.g. were trees normal size or dwarfed from being exposure to extreme weather on a ridge in the cloud forest)
e) careful consideration of the full floristic composition of the plot and cover for each species. This is especially important in Hawai'i where several species such as *Metrosideros polymorpha* occur widely and dominate the canopy of several ecological systems. Introduced species are a huge problem in lowland Hawai'i, so complete species composition is necessary to determine if the plot represents a disturbed natural system or has converted to a ruderal system dominated by introduced species.
f) awareness of possible errors in the plot data, such as mis-identification of species by the field crews, unevenness in how the cover values were estimated in the field or converted into the LFRDB (e.g. cover for trees estimated by a person standing on the ground vs an aerial view of the plot).

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

- *Metrosideros polymorpha* is dominant tree in several forest and woodland systems has low diagnostic value in determining which forest system.
- Generally, *Metrosideros polymorpha* dominated or codominated forests occurring above 1000 m are considered montane, and if occur in moisture zones 1-3 are thought to be dry forest, or if occur in moisture zones 4 or 5 are assumed to be mesic forest, and if occur in moisture zones 6 or 7 are wet forest (rainforest).
- Generally, *Metrosideros polymorpha* dominated or codominated forests occurring below 1000 m are considered lowland, and if occur in moisture zones 1-3 are thought to be dry forest, or if occur in moisture zones 4 or 5 are assumed to be mesic forest, and if occur in moisture zones 6 or 7 are wet forest (rainforest).
- Presence of indicator species works better in some systems than others as we are still refining them as we learn more.
- Introduced invasive species are converting native systems to ruderal systems so absolute and relative cover of species such as *Morella faya* and *Psidium* spp. are critical in determining system label.

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 11 types assigned to plots by experts, 3 had fewer than 10 samples, so are excluded from this particular analysis. From the remaining 8 types, the numbers of samples labeled to a given type ranged from 71 (Hawai'i Montane-Subalpine Mesic Forest) down to 11 (Hawai'i Lowland Rainforest). For all of these types, experts reported moderate confidence in their labels for at least 20% of the type's plots. Several (3) 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
Hawai'i Lowland Mesic Forest	Plot cannot be determined by available floristic data and elevation alone
Hawai'i Lowland Dry Forest	Plot occurs in the relatively recent 1868 lava flow so is dryer than the Price et al. 2007 Moisture Zone
Hawai'i Dry-Site Lava Flow	Plot is in moisture zone 5 so can't assign to CES412.416 Hawai'i Dry-Site Lava Flow.

These and other comments point to several important aspects for consideration. First, some ecological systems concepts are better known and understood than others. Therefore, a certain degree of classification refinement is likely needed in order to improve auto-keys. Second, the inclusion of some limited landform, soil, and or landscape context information could assist with some determinations within the key, or by a subsequent expert reviewer. Similarly, repeated references to photos further indicates the need for expert review of many types where moderate-low confidence of experts suggest that auto-keys might be prone to error. Third, additional floristic information is cited in some cases where their suspected limitations provide the primary source of expert uncertainty in labeling.

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 auto-keys 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 auto-keys for use with the NVC.

Within this GeoArea, some 32 terrestrial ecological system types could occur. Of these, 15 have a practical 1:1 relationship with NVC Group concepts, and the remaining 17 system concepts nest cleanly within 8 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 HI that have complex relationships with NVC Groups

There are no GeoArea HI 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 auto-keys for US-NVC Macrogroups instead of ecological systems could potentially resolve disagreements between experts and auto-keys found at the ecological systems level. To evaluate the potential effect of using the auto-key for Macrogroups, we arranged the ecological system types by US-NVC Macrogroup in the expert-auto-key contingency table, and also compared the percent of expert auto-key matches at the system level versus the Macrogroup level (Table 6).

There are 7 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, because of the confusion over lowland versus montane dry forests, using the macrogroups M210 Hawaiian Dry Forest combines Hawai'i Lowland Dry Forest and Hawai'i Montane-Subalpine Dry Forest and Woodland and significantly improves the percent of expert matches at MG level from 33% to 67%. This holds true for several other Macrogroups (Table 6).

However, combining Hawai'i Dry Cliff and Hawai'i Dry-Site Lava Flow into M265 Hawaiian Cliff, Scree & Rock Vegetation or combining Hawai'i Dry Coastal Strand and Hawai'i Wet-Mesic Coastal Strand into M231 Hawaiian Scrub & Herb Coastal Vegetation does not improve the percent matching. On the other hand, combining Hawai'i Lowland Dry Grassland and Hawai'i Lowland Dry Shrubland into M217 Hawaiian Lowland Shrubland, Grassland & Savanna does significantly improve the percent matching (6% to 31%), but at a cost of grouping grassland with shrublands, which are likely useful to keep separate for mapping and management purposes.

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**
M210 Hawaiian Dry Forest	1	18	33%	67%
M187 Hawaiian Lowland Rainforest	3	77	10%	19%
M194 Hawaiian Montane & Cloud Forest	2	109	30%	49%
M265 Hawaiian Cliff, Scree & Rock Vegetation	2	7	57%	57%
M217 Hawaiian Lowland Shrubland, Grassland & Savanna	2	16	6%	31%
M231 Hawaiian Scrub & Herb Coastal Vegetation	1	17	47%	47%
M085 West Pacific Salt Marsh	1	4	0%	0%

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 7 natural US-NVC Division concepts that occur here.

Discussion

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

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

There are a number of reasons why a sequence table may not successfully key all samples run through it:
a) the unknown floristic quality of the vegetation data (how complete, how well collected, does it accurately represent the vegetation concept being keyed);
b) our limited knowledge of the variability in species composition, vegetation structure, and the distribution of ecological systems; and
c) the comprehensiveness (or lack thereof) in field inventory for any particular system (e.g., many from one small area, few to none from elsewhere in the region).

Each of these are discussed below.

A. Quality of vegetation data

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

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

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

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

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

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

B. Constraints within sequence table

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

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

Over time, as our knowledge of the floristic composition and structure of vegetation in the United States becomes more complete, local-scale variables may not be needed. If the plot data themselves are complete (meaning the species composition has been adequately sampled and recorded for the plot) we can infer environmental setting and characteristic ecological dynamics through the use of indicator species. For example in the Rocky Mountains, *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

The auto-key needs review and editing based on LF Improvement project analysis results:

Verify that the elevation and moisture zone indicator species are working in light of new information on species distribution in the Hawaiian Islands. Although ecologically there is a transition zone between lowland and montane vegetation, in general plots above 1000 m elevation are considered montane ecological systems and plots below 1000 m are considered lowland ecological systems. In auto-key, an overlapping elevation break was used to more represent the transition zone, however it would reduce confusion and auto-key to expert mis-matches if a 1000 m elevation break was used in the auto-key, at least until more reliable montane and lowland indicator species can be identified. If the transitional elevation criteria are to be used e.g., montane > 950 m elevation and lowland <1050 m elevation, then assigning montane –subalpine plots first in auto-key would work better. Another way would be to treat the elevation transition zone separately (900-1100 m) and use re-assessed montane and lowland indicator species to label plots.

Review the Price et al. 2007 moisture zones, in light of new information. Some plots occurring in moisture zone 5 (mesic) would be better labeled wet forest rather than mesic forest. Should we include plots in moisture zone 5 with the wet forest systems (moisture zones 6 and 7) under certain environmental conditions (elevation) or presence of certain wet habitat indicator species? Plots that occur in moisture zone 5 need to be addressed separately.

Introduced, invasive species are a significant problem in Hawai'i. Review how these invasive species were addressed in auto-key to assess stands degraded by introduced invasive species in upper canopy, but not converted to a non-native, ruderal type.

Additional adjustments to Auto-key procedures – inclusion of location/biophysical (landform, soils, geology, landscape position) information for pre-processing plots and/or inclusion of features in auto-keys need review. For example, soil salinity would help with Northern Polynesia Tidal Salt Marsh. Adding substrate type and geology information is especially relevant is sparsely vegetated systems are to be labeled by the auto-key. Sparsely vegetated systems typically have low vegetation cover, variable species composition, and are often defined more based on substrate (e.g., beach, coastal dunes, rock outcrop). Percent cover of ground cover, such as bedrock, bare ground, litter would further help differentiate certain types.

Adjustment to auto-keys – additional requirements for a more complete species list of vegetation sample data; e.g., ground cover data, and a greater percentage of woody species (not just dominant and co-dominant species) would help improve accuracy for some systems, but limit the number of total plots for mapping. On the other hand some systems can be confidently labeled by a single dominant species in Hawai'i such as *Marsilea villosa* indicating Hawai'i Ihiihiluakea Vernal Pool.

Expert review and labeling of plots for types of low confidence from auto-key would reduce labeling errors with less expense. Using a similar expert review database, high confidence labeled plots could be identified and sorted out quickly to focus on the more difficult plots.

Adjustments to Map Legends such as moving to Group/Macrogroup concepts where systems level remains challenging is an option, but has the risk of making map classes thematically too broad. This remedy could be reviewed and applied on a case-by-case basis.

Coping with uncertainty about what proportion of types could NOT be adequately handled through any of the above adjustments should to be addressed during auto-key improvement.

Finally, 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.