Sacrificing the Ecological Resolution at the Altar of Thematic Accuracy: Map Accuracies of Vegetation Classifications in Eastern Nevada for the Southwest Regional Gap Analysis Project (SW ReGAP)

Todd D. Sajwaj[1], William G. Kepner[2], David F. Bradford[2], and Christine T. Herndon[2]

1. U.S. Army Corps of Engineers co-located at U.S. EPA, 944 E. Harmon Ave., Las Vegas, NV 89109; 702-798-2279; Sajwaj.Todd@epa.gov 2. U.S. Environmental Protection Agency, 944 E. Harmon Ave., Las Vegas, NV 89019

ABSTRACT

The Southwest Regional Gap Analysis Project (SW ReGAP) improves upon previous GAP projects conducted in Arizona, Colorado, Nevada, New Mexico, and Utah to provide a consistent, seamless vegetation map for this large and ecologically diverse geographic region. Nevada's component of the land cover mapping effort comprises 15 mapzones, or 291,700 km^2. As of October 2003, preliminary field sampling has been completed via road-based sampling and backpacking surveys in all 15 of Nevada's mapzones yielding a data set of 17,000+ sites. Based on plant community data collected in the field, each site is labeled with NVCS alliance and ecological system labels, and a National Land Cover Database (NLCD) label. Site polygons were intersected with 40+ spectral, topographic, climatic, and edaphic datalayers. A set of decision rules (or land cover models) was generated by the application of a classification/regression tree (CART) algorithm to the plant community label and its associated dependent variables. Land cover models were implemented in Imagine 8.6 image processing software to create classified vegetation maps. Three maps were constructed for each mapping unit at increasing levels of ecological resolution: an NLCD level map (coarsest), and ecological systems map (intermediate), and an alliance level map (finest). Maps have been constructed for the Mojave, Eastern Great Basin, and Lahontan Basin mapping units. Final vegetation maps were assessed for thematic accuracy at each of the three levels of ecological resolution. The NLCD level maps produced the highest thematic accuracy while the alliance level map produced the lowest thematic accuracy. The procedures used in field data collection, land cover modeling, accuracy assessment, and edge-matching adjacent mapping units are illustrated with examples from the east Great Basin mapping unit of east central Nevada.

INTRODUCTION

GAP Program Overview:

A "gap" is the lack of representation or underrepresentation of an element of biodiversity (plant community or animal species) in an area intended for its long-term maintenance. Gap analysis is a process to keep common species common by plugging the gaps in our network of lands managed for biodiversity. The GAP methodology is straightforward: 1) map the distributions of natural plant communities, 2) map predicted distributions of native terrestrial vertebrate species, 3) map the degree of management for biodiversity maintenance, 4) analyze the representation of vegetation and animal species distributions in the conservation network to identify "gaps" in long-term security. This type of coarse-filter approach can be used for research, land management, and conservation planning purposes by land managers, scientists and policy makers at both regional and ecosystem levels.

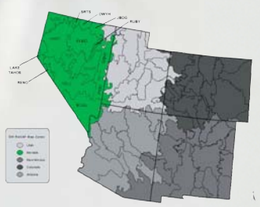

Figure 1

MATERIALS AND METHODS

Field Data Collection:

Field crews select training sites opportunistically based on homogeneity of plant species composition, landform, and spectral characteristics. Three essential steps are performed at each site: plant community characterization, site delineation, and photographic documentation.

Current Field Data Collection:

To date, 15,808 sites have been collected in Nevada's mapzones as depicted in figure 2. Fieldwork has been completed in 12 of Nevada's 15 mapzones. All field data collection is anticipated to be completed by October 31, 2003, and will comprise a dataset of 18,000 data points.

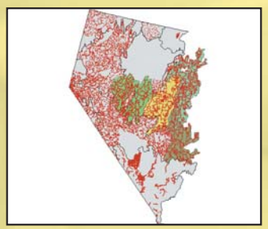

Figure 2

Plant Community Classification:

The National Vegetation Classification System (NVCS), developed by NatureServe, is the basis for plant community classification for the SW ReGAP project. Based on the plant community characterization data collected in the field, each site is assigned an alliance, ecological system, and National Land Cover Data (NLCD) label. (See Table 1.)

Table 1. Modified NVCS for the SW ReGAP Project

Level	Primary Basis for Classification	Examples
NLCD	Coarse land use/land cover classes	•Barren •Deciduous/Evergreen Forest •Short/Tall Shrublands •Shrub Herbaceous •Grassland •Woody/Herbaceous Wetland
Ecological System	Aggregation of plant communities that occur in similar ecological settings	•Sonora-Chihuahua Desert Riparian Mesquite Bosque •Sonora-Mojave Creosotebush-White Bursage Desert Scrub •Inter-Mountain Basins Mixed Salt Desert Scrub •Great Basin Pinyon-Juniper Woodland
Alliance	Dominant/diagnostic species of the uppermost stratum	*Populus deltoides* Temporarily-flooded Woodland Alliance

Modeling:

The SW ReGAP project is faced with a daunting task, namely constructing predictive vegetation models and subsequent vegetation maps for a diverse region that is almost as large as Alaska, or 530,000 square miles. Classification and regression trees can handle large numbers of predictor variables and are relatively insensitive to "noisy" data. Therefore, the SW ReGAP modeling approach is centered on the use of classification trees. Once all of the training site polygons for a mapping unit are collected, they are intersected through various digital datalayers. The geospatial data layers include Landsat 7 Enhanced Thematic Mapper Plus (ETM+) imagery acquired between 1999 and 2001 for 3 seasons (spring, summer, fall), digital elevation model data, and STATSGO soils data.

Classification and Regression Trees: Classification trees recursively partition a dataset into increasingly "pure" subsets based on a multitude of predictor variables. In the case of SW ReGAP, the pure subsets are groups of field sites that belong to the same alliance or system. The output of a classification tree is a set of decision rules.

Accuracy Assessment: As final predictive vegetation maps are completed, it will be subjected to various accuracy assessment procedures. Our methods include withholding a proportion of the training dataset to use in a conventional accuracy assessment and review of draft vegetation maps by regional experts.

PRELIMINARY RESULTS

Field Data Collection:

Preliminary maps for the East Great Basin Mapping Unit: The location of the East Great Basin mapping unit is shown in figure 3, and is composed of four mapzones (Duckwater, Goshute, Pioche, and Toiyabe). Of the 6041 training sites collected in the East Great Basin mapping unit, 4825 sites were used to construct classification rules while 1216 sites were used to assess the accuracy of the final land cover maps.

Table 2

Level of Ecological Detail	# Plant Communities from Field Data	# Modeled/Mapped Plant Communities	Internal Validation	Accuracy
NLCD	10	7 (70%)	83.3%	71.0%
Ecological Systems	46	21 (46/%)	80.4%	54.1%
Alliances	172	76 (44/%)	45.6%	22.6%

Figure 3

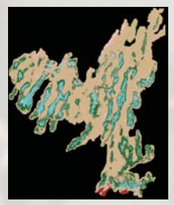

Figure 4

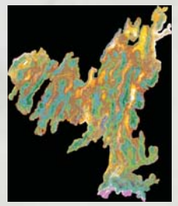

Figure 5

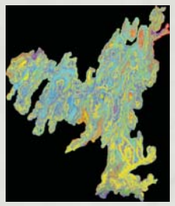

Figure 6

DISCUSSION

Use of a high degree of ecological detail for broad ecoregional maps is not a worthwhile effort. As the level of ecological resolution increases (see Table 2), the number of modeled classes does not increase at a similar rate; those classes that are modeled are less accurate. Rare classes (e.g. alpine meadows) are the first to be lost due to low sample numbers. If rare plant communities (and their associated vertebrate species) are not mapped they may not receive needed protection. Closely-related map classes (e.g. *Sarcobatus vermiculatus* Shrubland and *Sarcobatus vermiculatus* Intermittently-Flooded Shrubland) must typically be combined due to broad overlap of spectral and ecological setting characteristics.

FUTURE DIRECTIONS

Simultaneously, specific methodologies for producing predictive landcover maps, including the CART variants of "boosted" trees and random forests, are being developed. The completion of landcover mapping in spring 2004 will be followed by an intensive period of in which all mapping units will be edge-matched to complete the 5-state vegetation map. By the end of 2005, a regionwide digital vegetation map and database will be produced enabling land managers, scientists, and policy makers to make informed decisions regarding land use.

3181ab03.DB