

A Matrix-Based Approach to Fuzzy Set Accuracy Assessment

RS/GIS Lab White Paper

Remote Sensing/GIS Laboratory
College of Natural Resources
Utah State University, Logan, UT 84322-5275

25 May, 2007

John H. Lowry, Jr.^{1*}, Lisa A. Langs¹, R. Douglas Ramsey¹, Jessica Kirby² and Keith Schulz³

¹ RS/GIS Laboratory, College of Natural Resources, Utah State University, Logan, Utah

² State Institutional Trust Lands Administration (SITLA), State of Utah, Salt Lake City, Utah

³ NatureServe, Boulder, Colorado

* Corresponding author: J. Lowry is to be contacted at jlowry@gis.usu.edu, tel. 435-797-0653.

Abstract

The goal of many remote sensing-based land cover mapping projects is to map complex landscape features with high thematic resolution. Complex classification schemes often contain land cover descriptions that do not have clearly defined spectral, or at times, ecological breakpoints. Recognizing gradations in land cover classes may more accurately portray reality, but it also complicates quantitative assessments of map quality based on classical set theory and the deterministic error matrix approach. Using fuzzy sets for accuracy assessment provides an effective means to address this problem by recognizing varying levels of set membership for multiple map classes. In addition, fuzzy set assessments provide metrics for map quality assessment beyond those provided by the deterministic error matrix. These metrics include assessments of the frequency, distribution, magnitude, and source of map errors. We present an innovative approach to fuzzy set map accuracy assessment that uses a systematic framework based on ecological criteria. The methodology we present builds on the theoretical framework first presented by Gopal and Woodcock (1994). In this white paper we demonstrate how our approach can be used to assess the MAX, RIGHT, DIFFERENCE and MEMBERSHIP operators. A primary objective is to incorporate greater objectivity in the error evaluation process, thus allowing multiple map experts to conduct fuzzy accuracy assessments for multiple maps in a standardized fashion. Our approach differs from more common approaches to fuzzy set assessment by focusing on a holistic evaluation of the deterministic error matrix to derive fuzzy set metrics.

1.0 Introduction

Assessing land cover map accuracy is a significant concern for remote sensing-based mapping projects. While various approaches to map accuracy assessment are recognized (Foody, 2002), making an assessment helpful to the map user should be of primary importance (Smits et al., 1999). The error matrix and kappa statistic have emerged as the de facto standard for evaluating map accuracy and presenting this information to the map user (Foody, 2002). Appeal for the error matrix lies partly in its simplicity, but also in its utility (Congalton, 1991; Congalton and Green, 1999). The error matrix provides information about errors of commission and omission for individual map classes, as well as a quantitative measure of overall map accuracy in a single table.

Other approaches aimed at delivering additional information to map users have been explored. Most notable is information gained through fuzzy set analysis based on methods first presented by Gopal and Woodcock (1994). Fuzzy set analysis provides additional information to the map user because it recognizes that land cover does not fit unambiguously into single membership classes. The deterministic, or “traditional”, error matrix, based on classical set theory, assumes that all locations on the map can be unambiguously assigned to a map category. By recognizing varying levels of set membership into multiple map categories, it is possible to provide information not only on the frequency of errors, but also on the nature, magnitude and source of the errors (Gopal and Woodcock, 1994).

A fundamental component of fuzzy set assessment is the construction of a “linguistic measurement scale” to assign degrees of correctness for classification errors. Gopal and Woodcock (1994) suggest five levels of linguistic values that experts can use when evaluating a map product relative to reference samples, these are: *absolutely wrong*, *understandable but wrong*, *reasonable or acceptable*, *good*, and *absolutely right*. Determining the appropriate error level for any given reference site is subject to the judgment of the error assessment “expert.” This subjectivity can be problematic when more than one expert is involved in the fuzzy set assessment process (Woodcock and Gopal, 1992). Determining “good” versus “reasonable or acceptable” is not only subject to the judgment of each expert, but the intended application of the map. Establishing objective criteria for assigning the level of error, therefore, is a critical component of a meaningful fuzzy set assessment.

To address the need for a more objective approach to evaluating map errors in a fuzzy set analysis, Townsend (2000) developed a quantitative evaluation method focused explicitly on biotic similarity among mapped vegetation classes. Using a modified version of the Bray-Curtis coefficient of community (Bray and Curtis, 1957), Townsend (2000) evaluated similarity between reference sites and each mapped class to identify varying measures of agreement.

Townsend's (2000) technique represents an important contribution to fuzzy set assessment procedures and the Gopal-Woodcock methodological framework. The primary limitation to the approach, however, lies in the need for extensive amounts of survey information required to generate the quantitative similarity index (Townsend, 2000). Part of Townsend's study examined how limited, or generalized, site data affects the reliability of the similarity index. Sample sites providing sufficient information on species presence and abundance to reliably measure similarity among sample sites took between 1 and 4 hours to survey (Townsend, 2000). For typical remote-sensing based mapping efforts this level of detail is cost-prohibitive given the large number of samples required for both training and map validation.

This paper presents a methodology for fuzzy set assessment based on the framework suggested by Gopal and Woodcock (1994). Like the work of Townsend (2000), our methodology focuses on incorporating greater objectivity in the error evaluation process. We assert that the first requirement for a meaningful fuzzy set assessment is to explicitly define the context within which evaluation criteria are established. The second requirement is to establish specific criteria, or rules, for the error evaluation process within the defined context. In this paper we define the context for error evaluation to be gradations of ecological similarity. Within this context we present a systematic framework for establishing ecological criteria identifying set membership.

2.0 Background

The methods presented in this paper were developed, in part, through our effort to assess map quality for the Southwest Regional Gap Analysis Project (SWReGAP). SWReGAP was a multi-state collaborative effort involving five land cover mapping teams and a team of vegetation ecologists. The five-state region covers approximately 1.4 million square kilometers and includes the states of Arizona, Colorado, Nevada, New Mexico and Utah. The region was divided into 25 spectral-physiographic mapping zones, which

facilitated image classification and assisted in project tracking and management (Manis et al., 2000). We used a decision tree classifier with multi-season Landsat ETM+ imagery and digital elevation model derivatives to predict land cover (Lowry et al., in review). The land cover legend was based on the Terrestrial Ecological System Classification framework developed by NatureServe (Comer et al., 2003). Ecological Systems are defined as a groups of plant communities with similar ecological processes, substrates, and or environmental gradients, manifesting themselves in geographic scales of tens to thousands of hectares and persisting for 50 years or more (Comer et al., 2003). For the five-state region, approximately 93,000 ground truth samples were available for training and map validation.

Mapping responsibilities were divided among the 5 state teams by mapping zones, each state team having responsibility for 3 to 6 mapping zones. Each mapping zone functioned as a separate working unit and was independently assessed for map quality. Twenty percent of available samples for each mapping zone were randomly selected in proportion to the number of samples in each land cover class. Reference samples were intersected with the map and output formatted as a deterministic error matrix with a corresponding KHAT statistic (Congalton and Green 1999). The number of land cover classes varied with the size and complexity of each mapping zone, ranging from 13 classes in the least complex mapping zone to 53 classes in the largest and most complex mapping zone.

3.0 Methods and Results

A common approach to fuzzy set error assessment requires a “map expert” to assign multiple land cover labels to each reference site as acceptable alternatives to the most acceptable, or “true” call (Congalton and Green, 1999). The mapping analyst is unaware of the actual mapped class for that specific reference site, thus assuring independence between the alternate labels and the land cover map. Reference sites and the mapped land cover classes are compared, determining the level of correspondence, or “correctness”, between the reference data and the land cover map.

Our method differs from the common approach in its emphasis on a holistic evaluation of the deterministic error matrix rather than evaluation of individual reference sites. Instead of evaluating individual reference sites, the mapping analyst evaluates misclassification errors (i.e. off-diagonals) within the error matrix using a systematic framework that explicitly recognizes similarities among mapped land cover classes. An

important objective of this framework is to minimize the subjectivity inherent in the map expert's determination of acceptable alternatives to the most correct, or "true" label.

3.1 *Establishing a Systematic Framework*

We begin by recognizing two requirements for establishing a systematic framework for fuzzy set map assessment. First, we identify the context within which we consider similarities among land cover classes, and second we define explicit criteria for error evaluation given those recognized similarities. We suggest the context for recognized similarities be determined by the land cover classification scheme, or legend. For example, mapping projects dealing primarily with land use may choose to identify similarities among land uses as the context for establishing evaluation criteria. Because the land cover classification legend for SWReGAP focused on natural and semi-natural land cover classes utilizing NatureServe's Ecological Systems as the primary mapping unit, we identify an *ecological* context for our evaluation criteria.

We recognize four basic types of ecological similarity among mapped land cover classes (Table 1). Recognized *ecological similarity types* are defined as the possibility of a specific ecological condition shared by two distinct land cover classes. For example two land cover classes may have the same physiognomic structure (Type A), determined by how the Ecological System nests within the 2001 National Land Cover Data Set (NLCD) (Homer et al., 2003) classification legend. Some land cover classes share dominant or diagnostic species (Type B) as identified by the Ecological System class descriptions provided by NatureServe. Other classes may be commonly juxtaposed on the landscape, or may form a mosaic where patch or linear land cover classes commonly occur within a matrix of another land cover class (Type C). Finally some land cover classes may share similar specialized substrates (Type D).

Next we recognize that any two land cover classes may share multiple combinations of these basic ecological similarity types. Constructing a systematic ranking of all possible ecological similarity types provides a framework for assigning *ecological similarity categories* (Table 2). The idea of ecological similarity categories builds on Gopal and Woodcock's concept of "scaled linguistic values" to describe levels of possible membership within the fuzzy set framework. An important difference with our approach however, is that it explicitly defines an ecological context for determining membership using the combination of ecological similarity types to form criteria, or rules, for set membership (from Table 1).

Table 1. Names, codes and descriptions of four basic ecological similarity types.

Ecological Similarity Code	Ecological Similarity Type	Ecological Similarity Description
A	Physiognomic Structure	Reference and mapped classes share the same NLCD Class, such as:
		N30 Barren (Includes all Barren Lands)
		N40 Forest (Includes all Deciduous Forest, Evergreen Forest and Mixed Forest types)
		N50 Shrubland (Includes all Shrub, Dwarf Shrub and Shrub/Scrub types) Herbaceous (Includes all Grassland, Herbaceous, Savanna and Shrub-Steppe types)
		N70 Wetlands (Includes all Wetland, Riparian, Emergent Wetlands, Wet Meadows, and Greasewood Flats)
B	Dominant Species Composition	Reference and mapped classes share dominant/diagnostic species as specified in concept of Ecological Systems. For example, if systems share <i>dominant</i> or <i>codominant</i> species, then species composition is similar. If systems share species that are only <i>present</i> , then species composition is not similar. Would also apply between systems where the dominant/codominant species is common, but has been identified to a different subspecies (e.g. <i>Artemisia tridentata</i> spp.).
C	Juxtaposition	Reference and mapped classes commonly form a mosaic, such as patch or linear systems occurring within matrix systems, or where broad ecotonal boundaries between the classes occur with regularity. This often relates to minimum mapping unit (scale) issues with mosaics of similar landcover types. Refrain from using this code when the possibility of juxtaposition is only a rare occurrence.
D	Special Substrates	Reference and mapped classes share substrates with special properties that ecologically define each Ecological System. Apply with the following substrates only:
		- Eolian (sandsheets and dunes)
		- Bedrock (exposed weathering parent material); sparse vegetation (Barren) classes only
		- High Salinity (exposed marine shales, saline overflow /playas)

Another important difference is that our approach is designed to evaluate pairs of land cover types for recognized ecological similarities. For example, using this framework, an Inter-Mountain Basins Big Sagebrush Steppe land cover class is considered “moderately similar” to an Inter-Mountain Basins Big Sagebrush Shrubland land cover class, by virtue of two ecological similarity types: Type B (shared dominant or diagnostic species) and Type C (common juxtaposition). This differs from more common fuzzy set assessment methods that focus on assigning individual reference sites to membership categories. The importance of this difference becomes more apparent as we describe the computation of fuzzy set error matrices.

3.2 Computing Fuzzy Set Matrices

The ecological context and explicit ecological criteria established in Tables 1 and 2 form the foundation upon which we performed fuzzy set assessments for each of the 25 mapping zones in SWReGAP. Using these ecological criteria, mapping analysts constructed

a series of *supporting* matrices (described below) used in conjunction with the deterministic (traditional) error matrix to compute fuzzy set error matrix statistics. Thus, the same ecological criteria from the framework were applied across all 25 mapping zones.

Table 2. Examples and explanations of ranked ecological similarity categories based on ecological similarity types.

Ecological Similarity Code	Relative Similarity Category	Example	Explanation	Relative Similarity Score (RSS)
No Similarity (blank)	ABSOLUTELY INCORRECT	<i>Inter-Mountain Basins Cliff and Canyon (S009)</i> CONFUSED WITH <i>Great Basin Xeric Mixed Sagebrush Shrubland (S055)</i>	No Types of Ecological Similarity are shared between these two Ecological Systems.	1
A C D	SOMEWHAT SIMILAR	<i>Rocky Mountain Aspen Forest and Woodland (S023)</i> CONFUSED WITH <i>Great Basin Pinyon-Juniper Woodland (S040)</i>	These two Ecological Systems are nested within the same NLCD Class for shrub/scrub and therefore share A- Physiognomy. No other Types of Ecological Similarity are shared.	2
B AB AC AD BC BD CD	MODERATELY SIMILAR	<i>Inter-Mountain Basins Big Sagebrush Shrubland (S054)</i> CONFUSED WITH <i>Inter-Mountain Basins Montane Sagebrush Steppe (S055)</i>	These two Ecological Systems are similar in terms of B- shared dominant/diagnostic species and C- Juxtaposition.	3
ABC ABD ACD BCD ABCD	VERY SIMILAR	<i>Inter-Mountain Basins Big Sagebrush Shrubland (S054)</i> CONFUSED WITH <i>Great Basin Xeric Mixed Sagebrush Shrubland (S055)</i>	These two Ecological Systems are similar relative to A- Physiognomic Structure, B- Dominant Species Composition and C- Juxtaposition.	4
Diagonal Cell	ABSOLUTELY CORRECT	<i>Great Basin Pinyon-Juniper Woodland (S040)</i> MAPPED AS <i>Great Basin Pinyon-Juniper Woodland (S040)</i>	The reference and mapped classes are identical.	5

The initial supporting matrix is the *ecological similarity code* matrix (Table 3). This matrix has the same structure as a traditional error matrix, with an equal number of columns and rows for the number of land cover classes mapped. The analyst evaluates each possible paired combination of confused land cover classes, and assigns ecological similarity code(s) to each cell in the matrix based on recognized ecological similarity types (see Table 1). Types

of ecological similarity are determined by examining the description for each Ecological System from the classification legend and applying the criteria specified in Tables 1 and 2. Multiple combinations of ecological similarity codes representing different types of ecological similarity are possible. Empty cells indicate no ecological similarity between paired land cover classes, and self-similar land cover classes (i.e. diagonal cells) are given a code of 'X'.

Table 3. Ecological similarity code matrix derived from Tables 1 and 2.

LAND COVER CLASS	Map Code	S009	S023	S028	S040	S050	S054	S055	S065	S071	S078	S090	S096	S118
Inter-Mountain Basins Cliff and Canyon	S009	X		C	C	C				C		C		
Rocky Mountain Aspen Forest and Woodland	S023		X	AC	A					C				C
Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	S028	C	AC	X	AC					C				
Great Basin Pinyon-Juniper Woodland	S040	C	A	AC	X	C	C	C		C	C	C		C
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	S050	C			C	X	AC	A	A	C	C			
Inter-Mountain Basins Big Sagebrush Shrubland	S054				C	AC	X	ABC	AC	BC	BC	C	C	C
Great Basin Xeric Mixed Sagebrush Shrubland	S055				C	A	ABC	X	AC	C	BC	C	C	
Inter-Mountain Basins Mixed Salt Desert Scrub	S065					A	AC	AC	X		C	C	BCD	
Inter-Mountain Basins Montane Sagebrush Steppe	S071	C	C	C	C	C	BC	C		X	ABC	A		C
Inter-Mountain Basins Big Sagebrush Steppe	S078				C	C	BC	BC	C	ABC	X	AC	C	C
Inter-Mountain Basins Semi-Desert Grassland	S090	C			C		C	C	C	A	AC	X	C	
Inter-Mountain Basins Greasewood Flat	S096						C	C	BCD		C	C	X	AC
Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland	S118		C	C	C		C			C	C		AC	X

The analyst creates the second supporting matrix by translating the nominal ecological similarity codes in Table 3 to ordinal *relative similarity scores* (Table 4). Relative similarity scores (RSS) are numeric values ranging from 1 to 5 and correspond to the linguistic ecological similarity categories in Table 2. The *relative similarity score* matrix is a fundamental component of our methodology, and will be referenced throughout the remainder of this paper.

The deterministic (traditional) error matrix is a cross-tabulation of agreement between reference sites (columns) and mapped land cover classes (rows), based on classical

set theory (Congalton and Green, 1999). The matrix shows the number of reference sites that match the mapped land cover in the diagonal cells, and the number of reference sites that do not match the mapped land cover in the off-diagonal cells. The ratio of matched sites to total sites is represented as a percentage indicating individual and overall class accuracies. Producer accuracies are the ratio of matched sites to total number of reference sites for a given land cover class (columns) and provide a measure of omission error. User accuracies are the ratio of matched sites to total number of sites mapped in a given land cover class (rows) and provide a measure of commission error (Table 5).

Table 4. Relative Similarity Score (RSS) matrix.

LAND COVER CLASS	Code	S009	S023	S028	S040	S050	S054	S055	S065	S071	S078	S090	S096	S118
Inter-Mountain Basins Cliff and Canyon	S009	5	1	2	2	2	1	1	1	2	1	2	1	1
Rocky Mountain Aspen Forest and Woodland	S023	1	5	3	2	1	1	1	1	2	1	1	1	2
Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	S028	2	3	5	3	1	1	1	1	2	1	1	1	1
Great Basin Pinyon-Juniper Woodland	S040	2	2	3	5	2	2	2	1	2	2	2	1	2
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	S050	2	1	1	2	5	3	2	2	2	2	1	1	1
Inter-Mountain Basins Big Sagebrush Shrubland	S054	1	1	1	2	3	5	4	3	3	3	2	2	2
Great Basin Xeric Mixed Sagebrush Shrubland	S055	1	1	1	2	2	4	5	3	2	3	2	2	1
Inter-Mountain Basins Mixed Salt Desert Scrub	S065	1	1	1	1	2	3	3	5	1	2	2	4	1
Inter-Mountain Basins Montane Sagebrush Steppe	S071	2	2	2	2	2	3	2	1	5	4	2	1	2
Inter-Mountain Basins Big Sagebrush Steppe	S078	1	1	1	2	2	3	3	2	4	5	3	2	2
Inter-Mountain Basins Semi-Desert Grassland	S090	2	1	1	2	1	2	2	2	2	3	5	2	1
Inter-Mountain Basins Greasewood Flat	S096	1	1	1	1	1	2	2	4	1	2	2	5	3
Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland	S118	1	2	2	2	1	2	1	1	2	2	1	3	5

The key step in our methodology is a cell-by-cell evaluation of the deterministic error matrix in relation to corresponding cells in the relative similarity score matrix, which results in a computed “fuzzy set” matrix. During computation, misclassification errors in the off-diagonal cells of Table 5 are adjusted according their respective relative similarity score in Table 4. For example, in the “very similar” fuzzy set computation, cells with misclassification errors in the deterministic error matrix corresponding to a RSS of 4 (very

similar) are adjusted by adding the misclassification value (i.e. number of misclassified sites) to the diagonal cell in the computed fuzzy set matrix.

Table 5. Original error matrix.

LAND COVER CLASS		Code	REFERENCE											Total	User (%)		
			S009	S023	S028	S040	S050	S054	S055	S065	S071	S078	S090			S096	S118
MAPPED	Inter-Mountain Basins Cliff and Canyon	S009	5						1							6	83%
	Rocky Mountain Aspen Forest and Woodland	S023		4												4	100%
	Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	S028			5											5	100%
	Great Basin Pinyon-Juniper Woodland	S040				17							1			18	94%
	Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	S050					1									1	100%
	Inter-Mountain Basins Big Sagebrush Shrubland	S054				1		54	12	2	2	6	3	1		81	67%
	Great Basin Xeric Mixed Sagebrush Shrubland	S055						2	8	1	2	1				14	57%
	Inter-Mountain Basins Mixed Salt Desert Scrub	S065							1	2						3	67%
	Inter-Mountain Basins Montane Sagebrush Steppe	S071	1	2			1	1	3		18	2	1	1		30	60%
	Inter-Mountain Basins Big Sagebrush Steppe	S078						1				0		1		2	0%
	Inter-Mountain Basins Semi-Desert Grassland	S090						1					3			4	75%
	Inter-Mountain Basins Greasewood Flat	S096								1				1		2	50%
	Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland	S118													6	6	100%
	Total			6	6	5	18	2	59	25	6	22	9	8	4	6	176
Producers (%)			83%	67%	100%	94%	50%	92%	32%	33%	82%	0%	38%	25%	100%		70%
															KHAT= 0.63		

Fuzzy set matrices may be computed in one of two ways. Classification errors can be moved vertically to the diagonal “adjusting” errors of omission, or classification errors can be moved horizontally to the diagonal “adjusting” errors of commission (Table 6). New producer and user accuracies are computed for each land cover class reflecting a revised

assessment of errors given recognized ecological similarities between land cover classes and the new definition of what constitutes “misclassification.”

Separate fuzzy set matrices are produced for each of three relative similarity categories: very similar (4), moderately similar (3), and somewhat similar (2). Fuzzy set error matrices for levels 1 and 5 are not produced because they either provide the same information as the deterministic error matrix (level 5), or make all classes 100% accurate (level 1) which is not appropriate. The computation of each fuzzy matrix is facilitated using a script that compares the deterministic error matrix to the relative similarity score matrix.

Table 6. Revised "fuzzy set" matrix computed from the original error matrix (Table 5) and the Relative Similarity Score matrix (Table 4) using an RSS threshold of ≥ 4 . User and producer rates, and KHAT are recalculated. Values that have changed are in bold italics.

LAND COVER CLASS		Map Code	REFERENCE											Total	User (%)		
			S009	S023	S028	S040	S050	S054	S055	S065	S071	S078	S090			S096	S118
MAPPED	Inter-Mountain Basins Cliff and Canyon	S009	5						1							6	83%
	Rocky Mountain Aspen Forest and Woodland	S023		4												4	100%
	Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	S028			5											5	100%
	Great Basin Pinyon-Juniper Woodland	S040				17							1			18	94%
	Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	S050					1									1	100%
	Inter-Mountain Basins Big Sagebrush Shrubland	S054				1		56		2	2	6	3	1		71	79%
	Great Basin Xeric Mixed Sagebrush Shrubland	S055							20	1	2	1				#	83%
	Inter-Mountain Basins Mixed Salt Desert Scrub	S065							1	3						4	75%
	Inter-Mountain Basins Montane Sagebrush Steppe	S071	1	2			1	1	3		18		1	1		28	64%
	Inter-Mountain Basins Big Sagebrush Steppe	S078						1				2		1		4	50%
	Inter-Mountain Basins Semi-Desert Grassland	S090						1					3			4	75%
	Inter-Mountain Basins Greasewood Flat	S096												1		1	100%
	Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland	S118													6	6	100%
	Total			6	6	5	18	2	59	25	6	22	9	8	4	6	176
Producers (%)			83%	67%	100%	94%	50%	##	##	##	82%	##	38%	25%	100%		80%
															KHAT = 0.75		

3.3 Information Gained from Fuzzy Set Matrices

A chief advantage of the fuzzy set approach for accuracy assessment is that it provides the opportunity to extract additional information beyond that provided by the deterministic error matrix. Gopal and Woodcock (1994) suggest methods for exploring and presenting information on the frequency, magnitude, source and nature of errors through fuzzy set “operators.” These include the MAX and RIGHT operators, which provide information on the *frequency* and *distribution* of errors, and taken together provide a measure of “improvement” in accuracies based on fuzzy set theory. The DIFFERENCE operator also assesses improvement by identifying the *magnitude* or severity of errors for individual land cover classes. Exploring possible *sources* of error is accomplished through the MEMBERSHIP operator. Each of these modes of investigation is possible using the holistic matrix evaluation approach presented in this paper. Exploring the *nature* of errors using Gopal and Woodcock’s CONFUSION and AMBIGUITY operators requires individual reference site-based evaluations, and is therefore not possible with our approach. The remainder of this section describes how the frequency, distribution, magnitude and source of errors can be explored and presented using the fuzzy set matrices generated with our method.

3.3.1 Frequency and Distribution of Errors

The frequency and distribution of errors refers to the number of misclassifications distributed among all land cover classes. The deterministic error matrix provides this information through user and producer accuracies. What is of interest when performing a fuzzy set analysis of errors is how these accuracies change within the fuzzy set paradigm. In other words, recognizing that some errors may be considered “acceptable” because of recognized similarities among classes, how do the distribution and frequency of errors change?

Within the Gopal-Woodcock framework, the frequency and distribution of errors are presented using the MAX and RIGHT operators. The MAX operator is typically a more conservative metric of error and refers to the accuracies revealed by the deterministic error matrix. The more liberal RIGHT operator is defined by a specified threshold of acceptable fuzziness between classes. Using our terminology, this is a threshold of acceptable *ecological similarity* between classes. Numerically the threshold is defined by the relative similarity score, and any degree of similarity to the right of that score. For example, if the threshold

for the RIGHT operator is specified as “moderately similar,” classes with a relative similarity score ≥ 3 are considered matches rather than mismatches.

The holistic fuzzy set matrix approach we described in the previous section provides information based on the RIGHT operator. Each computed fuzzy set matrix is essentially a revised error matrix based on increasingly more liberal RIGHT thresholds (i.e. ≥ 4 , ≥ 3 and ≥ 2) of relative ecological similarity. Each of these matrices have value in and of themselves, however as Gopal and Woodcock (1994) suggest, information from the MAX and RIGHT operator presented together is more useful, as it provides a metric of improvement in the map’s accuracy given acceptable similarities between land cover classes.

Table 7 presents information on the MAX and RIGHT operators, using a RIGHT operator threshold of ≥ 4 . Using this threshold, mismatches identified by the ecological criteria as “very similar” are adjusted to be matches in the RIGHT operator. Individual class accuracies provided by the MAX operator represent only those cells with a relative similarity score of 5 (i.e. absolutely correct). We note that, given recognized ecological similarities at the ≥ 4 threshold level, improvements occur in 4 of the 13 classes. Mean improvement for the map was 10%, increasing from 70% to 80%. The most significant improvement occurred in the Great Basin Xeric Mixed Sagebrush Shrubland (S055) class, which increased from 32% to 80%. This can be interpreted to mean that if we are willing to accept the inclusion of land cover classes that are ecologically *very similar* to Great Basin Xeric Mixed Sagebrush Shrubland, the map has very good accuracy for this class.

Table 7: Results of MAX and RIGHT operators, identifying the frequency and distribution of errors, and "improvement" using a RSS threshold ≥ 4 (i.e. very similar).

LAND COVER CLASS	MapCode	Total # Sites	% of sites	MAX (M) - best answer		RIGHT (R) - acceptable		Improvement (R - M)
				Matches (RSS = 5)	Mismatches (RSS \neq 5)	Matches (RSS \geq 4)	Mismatches (RSS < 4)	
Inter-Mountain Basins Cliff and Canyon	S009	6	3%	5 (83%)	1 (17%)	5 (83%)	1 (17%)	0 (0%)
Rocky Mountain Aspen Forest and Woodland	S023	6	3%	4 (67%)	2 (33%)	4 (67%)	2 (33%)	0 (0%)
Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	S028	5	3%	5 (100%)	0 (0%)	5 (100%)	0 (0%)	0 (0%)
Great Basin Pinyon-Juniper Woodland	S040	18	10%	17 (94%)	1 (6%)	17 (94%)	1 (6%)	0 (0%)
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	S050	2	1%	1 (50%)	1 (50%)	1 (50%)	1 (50%)	0 (0%)
Inter-Mountain Basins Big Sagebrush Shrubland	S054	59	34%	54 (92%)	5 (8%)	56 (95%)	3 (5%)	2 (3%)
Great Basin Xeric Mixed Sagebrush Shrubland	S055	25	14%	8 (32%)	17 (68%)	20 (80%)	5 (20%)	12 (48%)

Inter-Mountain Basins Mixed Salt Desert Scrub	S065	6	3%	2 (33%)	4 (67%)	3 (50%)	3 (50%)	1 (17%)
Inter-Mountain Basins Montane Sagebrush Steppe	S071	22	13%	18 (82%)	4 (18%)	18 (82%)	4 (18%)	0 (0%)
Inter-Mountain Basins Big Sagebrush Steppe	S078	9	5%	0 (0%)	9 (100%)	2 (22%)	7 (78%)	2 (22%)
Inter-Mountain Basins Semi-Desert Grassland	S090	8	5%	3 (38%)	5 (62%)	3 (38%)	5(62%)	0 (0%)
Inter-Mountain Basins Greasewood Flat	S096	4	2%	1 (25%)	3 (75%)	1 (25%)	3 (75%)	0 (0%)
Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland	S118	6	3%	6 (100%)	0 (0%)	6 (100%)	0 (0%)	0 (0%)
TOTAL		176	100%	124 (70%)	53 (30%)	135 (80%)	35 (20%)	17 (10%)

3.3.2 Magnitude of Errors

Assessing the magnitude of errors reveals the severity of confusion among mismatches in the deterministic error matrix. This is particularly relevant with our ecological framework approach, as the severity of errors becomes explicitly tied to potential ecological consequences of misclassification. Here again, we borrow from Gopal and Woodcock (1994) the idea of a DIFFERENCE operator. Operationally this involves evaluating mismatches (i.e. off-diagonal cells) from the deterministic error matrix (Table 5) in relation to their corresponding relative similarity scores (RSS) in the relative similarity score matrix (Table 4). To reveal the magnitude of mismatches for each confused cell, we subtract the RSS for that cell from the highest possible similarity score (i.e. 5 or “absolutely correct”). The result is a range of difference scores between 0 and -4, where 0 indicates a perfect match ($5 - 5 = 0$), or no error, and -4 indicates a high-magnitude error ($1 - 5 = -4$).

This information can be summarized and presented as in Table 8. Here the number of mismatches from the deterministic error matrix and their corresponding difference errors are presented under the DIFFERENCE heading. The *weighted mean of DIFFERENCE errors* for each class is calculated by multiplying the frequency of errors of each mismatch category by its score, summing the products, and dividing the sum by the total number of sites for the land cover class. Formally this is expressed:

$$D_j = \frac{\sum_{i=1}^c (n_m * m)}{s_j}$$

Where D_j is the weighted mean difference for land cover class j , n is the number of classification errors for mismatch score m (or weight), s is the total number of sites for class j and c is the total number of mismatch categories for class j . The index is very similar to the arithmetic mean suggested by Gopal and Woodcock (1994) and has the effect of scaling the

magnitude of errors within each land cover class relative to all other land cover classes. Weighted mean DIFFERENCE values approaching 0 represent low-magnitude errors, whereas increasingly negative values indicated high-magnitude errors.

Table 8. Results of the DIFFERENCE operator, indentifying the magnitude of error for each land cover class.

LAND COVER CLASS	Map Code	Total # sites	DIFFERENCE					Arithmetic Mean of DIFFERENCE
			Mismatches				Matches	
			-4	-3	-2	-1	0	
Inter-Mountain Basins Cliff and Canyon	S009	6		1			5	-0.50
Rocky Mountain Aspen Forest and Woodland	S023	6		2			4	-1.00
Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland	S028	5					5	0.00
Great Basin Pinyon-Juniper Woodland	S040	18		1			17	-0.17
Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	S050	2		1			1	-1.50
Inter-Mountain Basins Big Sagebrush Shrubland	S054	59		1	2	2	54	-0.15
Great Basin Xeric Mixed Sagebrush Shrubland	S055	25	1	3	1	12	8	-1.08
Inter-Mountain Basins Mixed Salt Desert Scrub	S065	6			3	1	2	-1.17
Inter-Mountain Basins Montane Sagebrush Steppe	S071	22		2	2		18	-0.45
Inter-Mountain Basins Big Sagebrush Steppe	S078	9			7	2	0	-1.78
Inter-Mountain Basins Semi-Desert Grassland	S090	8		5			3	-1.88
Inter-Mountain Basins Greasewood Flat	S096	4	1	2			1	-2.50
Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland	S118	6					6	0.00
TOTAL		176	2	18	15	17	124	
Percent of TOTAL		100%	1%	10%	9%	10%	70%	

In this example, the Inter-Mountain Basins Big Sagebrush Shrubland (S054) class has low-magnitude errors, with a mean DIFFERENCE value of -0.15. This reflects two errors that are considered “very similar” (mismatch of -1), two errors considered “moderately similar” (mismatch of -2) and 1 error considered “somewhat similar” (mismatch of -3). In contrast, the Inter-Mountain Basins Greasewood Flat class exhibits high-magnitude errors with a mean difference value of -2.50. This reflects two errors that are “somewhat similar” (mismatch of -3) and one error that is “absolutely incorrect” (mismatch

of -4). In summary, this operator provides a metric of land cover classification performance by measuring the severity of errors for individual land cover classes, based on the criteria provided by the ecological similarity framework. It does not however, indicate which classes are confused with any given land cover class, or in other words, the source of errors.

3.3.3 Source of Errors

An important advantage of fuzzy set analysis is the ability to recognize multiple set memberships and grades of set membership. The deterministic error matrix indicates the land cover class with which a particular land cover class may be confused (i.e. by interpreting the off-diagonal cells). However, it does not directly provide a metric identifying the frequency, or degree to which an individual class is confused with other land cover classes. The idea behind Gopal and Woodcock's MEMBERSHIP operator is that classification error may be considered "correct" given the acceptance of multiple set memberships (i.e. reference sites may not fit unambiguously into land cover classes). Gopal and Woodcock's method identifies the frequency of multiple set memberships for each reference site at a given fuzziness threshold (i.e. RIGHT operator threshold). Our approach adopts this concept, but applies it to an assessment of the misclassification errors presented by the off-diagonal cells within the deterministic error matrix. Thresholds of fuzziness are defined by recognized levels of relative ecological similarity, enumerated by the RSS (Table 4).

To assess the frequency of multiple set memberships relative to the producers accuracy at given thresholds of recognized ecological similarity, we return to an evaluation of the deterministic error matrix (Table 5) in relation to the RSS matrix (Table 4). The evaluation is facilitated by constructing a new matrix (Table 9a) that combines information from Tables 4 and 5. Here, the first numeral represents the number of reference sites in the cell (from Table 5) and the second numeral (in bold) indicates the relative similarity score for that cell (from Table 4). Table 9b summarizes information in the combined matrix providing the frequency of multiple set memberships, followed by the frequency represented as a proportion. The frequency of multiple set memberships for each RSS threshold provides information about the relative heterogeneity of the mapped land cover classes, based on gradations of recognized ecological similarity (i.e. RSS thresholds).

The *weighted mean RSS for errors* is the mean relative similarity score for all sites from the off-diagonal cells and provides a single metric for set membership errors. It is calculated by multiplying the frequency of each RSS score of all sites from the off-diagonal cells,

summing across all possible RSS scores and dividing by the total number of sites found off the diagonal (i.e. total number of errors). Formally it is expressed:

$$R_j = \frac{\sum_{i=1}^n (f_r * r)}{e_j}$$

Where R_j is the weighted mean RSS of errors for land cover class j , f is the frequency of errors for membership category r , e is the total number of errors for class j , and n is the total number errors for class j . The weighted mean RSS value therefore, is metric of set membership based on the observed errors (Table 5) and recognized similarities (Table 4). A mean value of 0 represents no overlap with any other class and a value of 4.00 shows significant ecological overlap with one or more land cover classes.

The intent of set membership analysis is to identify possible sources of error by revealing overlap among land cover classes. This is useful when interpreting the quality of the map product, but may have greater utility during early phases of the mapping process to determine whether land cover class definitions need refinement. For example, from Table 9a we see that classes S054 and S055 are highly confused with each other. From a producer's perspective (i.e. errors of omission, or errors along the vertical axis) we note that S054 is confused with four other classes: S055, S071, S078 and S090 (for class names, see Table 9a). The two-dimensional matrix format provides a visual impression of where overlap between classes exists. Including the relative similarity score (in bold) for each cell provides additional information on the ecological similarity of the errors for each mapped class. For example, of the five validation sites omitted from the correct classification of S054, two of them were mapped as S055, which is considered ecologically *very similar* to S054.

Further examination of the matrix reveals that class S055 is also confused with four other classes, and S054 is one of the four (from producer's perspective). Before we draw conclusions about confusion between these classes being attributable to ecological overlap, it is helpful to analyze the proportion of error for each class (from Table 9b). We note that for S054, two out of 59 samples, or 3% of the errors are confused with S055. On the other hand for S055, 12 out of 25 samples or 48% of the errors are confused with S054. If confusion between these two classes were attributable to ecological similarity we would expect the proportion of overlapping errors to be roughly the same. What we see is that S055 is being mapped almost half of the time as S054, but not the inverse. Therefore the explanation may

not be the similarity between the two classes, but a problem with the mapping algorithm. In this case, the decision tree classifier may have been heavily influenced by the disproportionate number of training samples available for the S054 class.

Table 9a. Results of MEMBERSHIP operator presented as a combined matrix (1st numeral from deterministic matrix, 2nd numeral (bold) from RSS matrix).

LAND COVER CLASS		REFERENCE													Total	User (%)	
		S009	S023	S028	S040	S050	S054	S055	S065	S071	S078	S090	S096	S118			
M A P P E D	Inter-Mountain Basins Cliff and Canyon S009	5:5						1:1								6	83%
	Rocky Mountain Aspen Forest and Woodland S023		4:5													4	100%
	Rocky Mountain Subalpine Dry-Mesic Spruce-Fir Forest and Woodland S028			5:5												5	100%
	Great Basin Pinyon-Juniper Woodland S040				17:5							1:2				18	94%
	Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland S050					1:5										1	100%
	Inter-Mountain Basins Big Sagebrush Shrubland S054				1:2		54:5	12:4	2:3	2:3	6:3	3:2	1:2			81	67%
	Great Basin Xeric Mixed Sagebrush Shrubland S055						2:4	8:5	1:3	2:2	1:3					14	57%
	Inter-Mountain Basins Mixed Salt Desert Scrub S065								1:3	2:5						3	67%
	Inter-Mountain Basins Montane Sagebrush Steppe S071	1:2	2:2			1:2	1:2	3:2			18:5	2:4	1:2	1:1		30	60%
	Inter-Mountain Basins Big Sagebrush Steppe S078						1:3					0:5		1:2		2	0%
	Inter-Mountain Basins Semi-Desert Grassland S090						1:2						3:5			4	75%
	Inter-Mountain Basins Greasewood Flat S096								1:4					1:5		2	50%
	Great Basin Foothill and Lower Montane Riparian Woodland and Shrubland S118														6:5	6	100%
Total		6	6	5	18	2	59	25	6	22	9	8	4	6	176		
Producers (%)		83%	67%	100%	94%	50%	92%	32%	33%	82%	0%	38%	25%	100%		70%	

Table 9b. Summary of combined matrix (Table 9a) from a producer's perspective.

SET MEMBERSHIP as Frequency	S009	S023	S028	S040	S050	S054	S055	S065	S071	S078	S090	S096	S118
No membership (RSS = 1)							1					1	
Multiple set membership (RSS ≥ 2)	1	2		1	1	2	3		2		5	2	
Multiple set membership (RSS ≥ 3)						1	1	3	2	6			
Multiple set membership (RSS ≥ 4)						2	12	1		2			
Single set membership (RSS = 5)	5	4	5	17	1	54	8	2	18		3	1	6

Total Sites	6	6	5	18	2	59	25	6	22	8	8	4	6
Total Errors	1	2	0	1	1	5	17	4	4	8	5	3	0
Weighted Mean RSS for Errors	2.00	2.00	0.00	2.00	2.00	3.00	3.41	3.25	2.50	3.25	2.00	1.67	0.00
SET MEMBERSHIP as a Proportion													
No membership (RSS = 1)							0.04						0.25
Multiple set membership (RSS ≥ 2)	0.17	0.33		0.06	0.50	0.03	0.12		0.09			0.63	0.50
Multiple set membership (RSS ≥ 3)						0.02	0.04	0.50	0.09	0.75			
Multiple set membership (RSS ≥ 4)						0.03	0.48	0.17		0.25			
Single set membership (RSS = 5)	0.83	0.67	1.00	0.94	0.50	0.92	0.32	0.33	0.82	0.00	0.38	0.25	1.00

In summary, combining the deterministic error matrix with the relative similarity score matrix presents additional information that can be used to better interpret the map product, as well as guide the mapping process. The overall goal is to identify potential sources of mapping confusion. In addition to more detailed analysis, the weighted mean RSS for errors provides a single metric of possible ecological overlap for each land cover class.

4.0 Discussion

Our methodology provides an innovative approach to fuzzy set map accuracy assessment that builds on the framework originally outlined by Gopal and Woodcock (1994). While we borrow concepts from Gopal and Woodcock (1994) and adhere to their underlying objectives for fuzzy set analysis, our approach differs in two fundamental ways. First, a central goal to our methodology was the creation of a standardized framework for fuzzy set evaluation that could be replicated by multiple map assessment experts. While the ecological framework approach does not completely eliminate subjectivity in the fuzzy set evaluation process (some subjectivity remains in the interpretation of the Ecological System descriptions) it provides a standardized structure within which decision rules can be constructed.

Second, our approach focuses on a holistic evaluation of the error matrix, rather than an evaluation of individual reference sites. Operationally, applying the decision rules based on ecological criteria results in the creation of the relative similarity score (RSS) matrix. From the relative similarity score matrix we have shown that it is possible to derive and present information on the frequency, distribution, magnitude, and source of errors. A drawback to this approach is that we cannot detect variability among reference sites. In other words, we must assume that all reference sites are labeled correctly. However, if an

important goal of fuzzy set assessments is to account for variation among land cover classes, our approach achieves results similar to those using more traditional fuzzy set assessment methods. A key advantage to the holistic matrix approach is that it can be performed quickly, either during or after the mapping process.

Through the course of our investigation we discovered new questions regarding the use of fuzzy sets in map accuracy assessment. These apply equally to our matrix-based evaluation and to more traditional reference site-based approaches. The first question deals with how unequal distributions of reference samples affect the outcome of fuzzy set operators. The hypothetical dataset presented by Gopal and Woodcock (1994) used an equal number of samples per class. In our dataset, the distribution of validation samples is not evenly distributed. Future research should investigate how fuzzy set operators are affected under conditions of unequal distribution of samples among classes.

A second question focuses on possible differences in outcome depending on whether the fuzzy set assessment is applied from a user's or producer's perspective. Conforming to previous fuzzy set assessment methods, we present our results in this paper using a producer's perspective (Gopal and Woodcock, 1994; Laba et al., 2002; Green and Congalton, 2005). However, our matrix-based approach opens the possibility of assessing multiple set memberships from a user's perspective as well. Future research investigating the differences, and possible utilities of fuzzy set assessment from both the user's and producer's perspective might prove interesting and worthwhile.

5.0 Conclusion

Classification error can be attributed to a number of sources, including poor discriminating power of the classification algorithm, inadequate training samples, mis-registration of training samples to imagery, and ambiguity among land cover classes. The focus of this research has been to address classification error due to ambiguity among land cover classes using fuzzy set theory.

Increasingly, remote sensing-based mapping approaches are mapping more complex landscape features with greater thematic detail. Green and Congalton (2005) point out that more complex classification schemes require more effective methods of assessing map quality than the traditional deterministic error matrix. The deterministic error matrix provides a good starting point for error interpretation because it provides useful information and is relatively easy to interpret. However, more complex classification schemes typically

do not have clearly defined differences between land cover mapping categories. Instead, classification schemes such as NatureServe's Ecological System classification scheme represent gradations of ecological homogeneity across the landscape. Taking into consideration ecological gradations in the landscape may more accurately portray reality, but also complicates making a quantitative assessment of map quality. Thus, as long there is demand for maps with complex classification schemes, there will be a need for better methods to assess their agreement with what is on the ground. Fuzzy set methods, such as the approach described in this paper, address this need.

6.0 References

- Bray, J.R. and J.T. Curtis, 1957. An ordination of upland forest communities of southern Wisconsin. *Ecological Monographs*, 27(4):325-349.
- Comer, P., D. Faber-Langendoen, R. Evans, S. Gawler, C. Josse, G. Kittel, S. Menard, M. Pyne, M. Reid, K. Schulz, K. Snow, and J. Teague, 2003. *Ecological Systems of the United States: A Working Classification of U.S. Terrestrial systems*. Arlington, Virginia: NatureServe.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37:35-46.
- Congalton, R.G. and K. Green, 1999. *Assessing the Accuracy of Remote Sensed Data: Principles and Practices*, Lewis Publishers, Boca Raton, Florida, 137 p.
- Foody, G., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80:185-201.
- Gopal, S. and C. Woodcock, 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, 60(2):181-188.
- Green, K. and R. Congalton, 2005. An error matrix approach to fuzzy accuracy assessment: The NIMA Geocover Project. In *Remote Sensing and GIS Accuracy Assessment*, R.S. Lunetta and J.G. Lyon (Editors). CRC Press, Boca Raton, Florida, 304 p.
- Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2003. Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, 40(7):829-840.
- Laba, M., S.K. Gregory, J. Braden, D. Oguracak, E. Hill, E. Fegraus, J. Fiore, S.D. DeGloria, 2002. Conventional and fuzzy accuracy assessment of the New York Gap Analysis Project land cover map. *Remote Sensing of Environment*, 81:443-455.
- Lowry, J. H, Jr., R. D. Ramsey, K. Boykin, D. Bradford, P. Comer, S. Falzarano, W. Kepner, J. Kirby, L. Langs, J. Prior-Magee, G. Manis, L. O'Brien, T. Sajwaj, K. A. Thomas, W. Rieth, S. Schrader, D. Schrupp, K. Schulz, B. Thompson, C. Velasquez, C. Wallace, E. Waller and B. Wolk, (in review). Mapping medium scale land cover over very large geographic areas within a collaborative framework: A case study of the Southwest Regional Gap Analysis Project (SWReGAP).
- Manis, G., C. Homer, R.D. Ramsey, J. Lowry, T. Sajwaj, and S. Graves, 2000. The development of mapping zones to assist in land cover mapping over large geographic areas: A case study of the Southwest ReGAP project. GAP Analysis Bulletin No. 9. U.S. Geological Survey, Biological Resources Division.

- Smits, P.C., S.G. Dellepiane, and R.A. Schowngerdt, 1999. Quality assessment of image classification algorithms for land-cover mapping a review and a proposal for a cost-based approach. *International Journal of Remote Sensing*, 20:1461-1486.
- Townsend, P.A., 2000. A quantitative fuzzy approach to assess mapped vegetation classifications for ecological applications. *Remote Sensing of Environment*, 72:253-267.
- Woodcock, C. and S. Gopal., 1992. Accuracy assessment of the Stanislaus Forest vegetation map using fuzzy sets. Proceedings of the 4th Biennial Remote Sensing Applications Conference, Orlando, FL, April, 17.