Statistical Evaluation of the Wyoming and Colorado Landcover Map Thematic Accuracy Using Aerial Videography Techniques

Final Report

May 2000

William A. Reiners
Kenneth L. Driese
Department of Botany
University of Wyoming
Laramie, WY 82070

Donald Schrupp
Colorado Division of Wildlife

EXECUTIVE SUMMARY

Accuracy assessment of the Colorado and Wyoming Gap Analysis land cover maps was initiated in 1996 as a cooperative effort between the Biological Resource Division (BRD) of the U.S. Geological Survey (USGS), the University of Wyoming (UW), and the Colorado Division of Wildlife (CDOW). The primary objective was to provide statistical assessment of the thematic accuracy of the Wyoming and Colorado Gap landcover maps at level 4 (the "Anderson level") for the maps overall, and for individual cover types using airborne videography as proxy for "ground truth". Secondary objectives included (1) exploration of the application of fuzzy accuracy assessment techniques to statewide land cover maps and (2) evaluation of airborne videography as a tool for statewide accuracy assessments in the Rocky Mountain Region. This assessment provides tools and *caveats* for researchers embarking on similar assessments in the future.

Assessment of the Colorado land cover map was undertaken first because 6 of 8 air video transects had been acquired in Colorado by Donald Schrupp of CDOW prior to initiation of project funding. Two additional transects were eventually flown to provide 8 total transects, each extending from the northern to the southern border of Colorado. Video frames were linked to differentially corrected GPS data collected during the flights. These data allowed creation of a GIS layer depicting the location of individual video frames. Attributes describing land cover, elevation, and other frame-specific information were added to the GIS frame attributes, allowing selection of frames for ground training and accuracy assessment. Ground visits to locations corresponding to video frames for each Colorado cover type captured on video (not all cover types were "visited" by video transects) allowed assembly of an interpretation key that included prints of zoom and wide angle video frames, site photos, cover type and field notes. This key was used for interpreter training and as a reference during interpretation.

A total of 7139 video frames interpreted by two trained interpreters characterized 591 polygons from the Colorado land cover map. Because some mapped cover types occupied relatively small areas the video transects encountered only 38 of the 52 mapped types in Colorado. A simulation study examined diminishing returns with video sampling and suggested that collecting additional cover types by increasing transect length was impractical. Consequently, accuracy statistics for the Colorado land cover map are based on analysis of the 591 sampled polygons representing the 38 sampled cover types only. Sampled polygons were used to calculate traditional map accuracy (overall, User's, Producer's, Kappa) and the fuzzy operators (MAX, RIGHT, CONFUSION, AMBIGUITY, DIFFERENCE, MEMBERSHIP) described by Gopal and Woodcock (1994). Accuracy statistics describe both level 4 (referred to in this report as the "Anderson Level") and level 5 (referred to as the "Gap Level") in the classification hierarchy, and address other issues, including primary and secondary polygon labels and interpreter confidence.

Gap level overall map accuracy for the Colorado map was 31% (Kappa = 28%) based on the video interpretation while User's and Producer's accuracies for individual types varied widely. The fuzzy operator "RIGHT", which counts a polygon correct if the confusion is not serious for habitat modeling, suggested overall accuracy at the Gap level of 76%. At the Anderson level, overall map accuracy was 58% (Kappa = 48%), again with a wide range of User's and Producer's accuracy for individual types. The fuzzy "RIGHT" operator at the Anderson level suggested overall map accuracy of 79%. Detailed analysis of the assessment data suggest that much of the map confusion occurred within, rather than between physiognomic types. As emphasized below, it is likely that the source of some of this confusion is the interpretation of the video reference data, rather than confusion in the map. Simulations are underway to explore the propagation of confusion in reference data through the assessment process.

Assessment of the Wyoming Gap land cover map was initiated using identical methodology as used in Colorado but was not completed. Airborne videography was collected in Wyoming along 8 north-south transects. Video data were linked to a GIS point data layer using GPS as in Colorado and video frames representing sampled cover types were selected and visited on the ground to assemble an interpretation key. Cost of the Colorado portion of the assessment, however, proved higher than anticipated and effort was concentrated in that state to ensure that an assessment product and exploration of important issues related to that assessment were completed. Wyoming air video on analog high-8mm video cassettes and related GIS data are archived at the University of Wyoming. These data are available for use during future "reGap" of Wyoming.

Airborne videography as a tool for accuracy assessment provides relatively inexpensive "ground truth" for land cover evaluation and has other advantages over traditional ground-based assessment. Airborne sampling allows characterization of large polygons that would be difficult to see comprehensively on the ground and it allows visits to sample units in inaccessible areas. Video also provides a large set of potential sample points (frames) for accuracy assessment. With these advantages however, come important limitations that should be considered by anyone designing statewide assessments. First, it is impractical to visit all mapped landcover types across a state with video transects when some types occupy small areas. Second, our experience in Wyoming and Colorado suggests that many important cover types at level 5 are indistinguishable from one another on analog video. While every geographic region has unique cover with different interpretability, in general we found that many shrub species and most grassland types are not distinguishable in the videography that we used. Finally, it is difficult to quantify error in video interpretation, which we feel can be substantial in some cover types. Unassessed error in video interpretation and error in the map being assessed cannot be disentangled, resulting in unquantified uncertainty in map accuracy statistics. Some of these limitations may be overcome by improvement in video systems (Dana Slaymaker - personal communication), in particular with digital videography and stereo image pairs. Even so, videography should be evaluated realistically for each application and geographic region to which it will be applied, and alternative assessment methods should be considered for some cover types.

Acknowledgments

We would like to thank several people who contributed substantially to this project. First, Thomas Owens, of the USGS did most of the interpretation of air video frames in Colorado. This work was obviously central to the assessment. Tom's experience in photointerpretation, and his thoughtful evaluation of video as a tool for distinguishing vegetation were extremely valuable. Thanks also to the National Gap Program, in particular Patrick Crist and Michael Jennings, for encouragement and financial support. The National Gap Program administrative staff also provided invaluable support.

The Colorado Division of Wildlife was an essential cooperator for this project. Colorado air videography was acquired using a CDOW aircraft and pilot, with video equipment assembled by Donald Schrupp of CDOW. CDOW also provided office space and computer and administrative support for the Colorado portion of the project. The videography equipment was modified and used for video acquisition in Wyoming.

Dana Slaymaker of the University of Massachusetts has been a central figure in any Gap study using aerial videography and ours is no exception. Building on his Gap experience in Arizona, Dana has continued to improve systems and methodology for using air video in land cover mapping applications. He has made several trips to Colorado to help us assemble our videography system, and to instruct us in methods for processing and analyzing the resulting video data. We are extremely grateful to Dana for his patience and hard work.

We would like to thank the U.S. Forest Service for providing an aircraft and pilot for the video flights in Wyoming. In particular, William Snyder was essential for helping us through the bureaucratic web resulting from using a USFS aircraft for a USGS project. Coleen Roberts, the USFS pilot for the Wyoming flights was excellent, and helped us to capture a quality set of video transect data.

Finally we would like to thank Ken Gerow of the University of Wyoming Department of Statistics, and Steve Stehman, of the State University of New York, for valuable consultation on statistical issues. Many other people, especially from the Gap Analysis program in individual states have contributed with ideas and support.

TABLE OF CONTENTS

Executive Summary
Acknowledgments

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Background

Uncertainty is an inherent quality of spatial data in general and of maps in particular. It is crucial, especially when important decisions are based on spatial data, to assess the accuracy of these data so that users can understand the types, sources and nature of uncertainty and their implications. The primary purpose of map accuracy assessment is to explore uncertainty in the map being assessed and to present the results of this exploration in a manner that can be interpreted by map users.

The Gap landcover map of Wyoming (WYGAP) was completed in 1994 (Merrill et al. 1996, Driese et al. 1997) and of Colorado (COGAP) in 1997 (Schrupp et al. 2000). While these maps were produced specifically for Gap vertebrate habitat modeling, they are used for other applications by scientists (e.g. Oleson et al. 1997) and land managers (e.g. Davis et al. 1994). To ensure that users understand the strengths and limitations of the maps, accuracy assessment based on reference data from aerial videography was initiated in 1996. This report describes the accuracy assessment methods, provides complementary interpretations of the results, and evaluates airborne videography as a tool for accuracy assessment in the Rocky Mountain Region. The Wyoming assessment was not completed although videography was acquired, processed and archived and a sampling scheme for assessment was devised. Consequently, this report focuses on methods, results and research issues related to the assessment of the Colorado landcover map. Similar methodology could be applied to the Wyoming map, which was produced using identical mapping procedures as Colorado.

Accuracy assessment of statewide landcover maps is difficult, especially in the Western U.S., because the maps cover extensive areas. Rugged terrain, private land access and the sheer size of map units limit our ability to gather reference data. Airborne videography is one tool that provides relatively inexpensive sampling of large areas (like states) at high resolution. Video allows the capture of continuous swaths of imagery along flight transects, and allows simultaneous collection of highly detailed zoom video and wide angle video for context. Airborne videography has been used in Gap Analysis for both designation of training sites for classification (the "front end") and for accuracy assessment (the "back end"). Any remote sensing tool, including videography, has strengths and weaknesses specific to the environment in which it is applied. Readers of this report should be cognizant of this, and should understand that successful use of videography depends on the physical characteristics of the system in which they work, and on their specific objectives.

1.2 Objectives of Thematic Map Accuracy Assessment

Hay (1979) proposed five general questions that should be addressed to understand thematic map classification accuracy. These include:

1. What proportion of the classification is correct?
2. What proportion of the assignments to a particular class are correct?
3. What proportion of a given class is correctly classified?
4. Is a given class over- or under-estimated?
5. Are errors randomly distributed?

Although providing a framework for accuracy assessment, more detailed information can be determined by combining traditional accuracy assessment based on contingency table analysis and fuzzy accuracy assessment that recognizes multiple levels of right and wrong. We expand Hay's list of basic questions into a set of general objectives at a broad level and a more extensive set of specific objectives at the state level. These objectives raise research questions about sampling strategies, efficiency considerations, and the viability of large area assessments using air videography. They are organized here into a hierarchy that defines the scope and goals of this project.

1. Develop a sampling scheme that maintains statistical rigor while: a) estimating the dominant land cover in each sample unit (polygon) and b) representing the collection of mapped cover types.

    A. Explore the efficiency of video sampling for achieving these sampling goals by modeling flight transect scenarios.
    B. Collect airborne videography for each state based on the results of the sampling simulations.

2. Develop methods for interpreter training and explore differences among interpreters.

3. Use contingency table analyses based on reference data from interpreted air videography to calculate overall map accuracy, producer's accuracy, user's accuracy and the Kappa statistic for the Colorado land cover map.

    A. Apply to both Gap level and Anderson level landcover types.

4. Modify the fuzzy accuracy measures proposed by Gopal and Woodcock (1994) for application to statewide accuracy assessments.

    A. Calculate and interpret the fuzzy operators MAX, RIGHT, DIFFERENCE, MEMBERSHIP, CONFUSION and AMBIGUITY.

    B. Apply to both Gap level and Anderson level landcover types.

5. Evaluate the utility of aerial videography as a tool for compiling reference and training data in Colorado and other analogous Rocky Mountain and northern plains areas.

    A. Explore the accuracy of the videography interpretation itself for each of the land cover types.
    B. Compare interpreter confidence in interpretation of individual video frames with the accuracy of corresponding mapped polygons.
    C. Make recommendations for future use of air video in similar geographic regions.

6. Present the results of the accuracy assessment so that they are easily understood and interpreted by map users.

1.3 Description of the Maps

Landcover mapping in Wyoming and Colorado followed the methods developed for the California Gap project (Davis et al. 1995). Map units (vector polygons) were digitized directly on the computer screen using enhanced Landsat Thematic Mapper (TM) imagery as a base for interpretation. The minimum mapping unit (MMU) followed the Gap standard of 1 km^2, although many polygons in both states were much larger, representing extensive units of single vegetation types. The Colorado landcover map depicts 52 cover types in 16,618 polygons. The Wyoming map includes 41 cover types in 14,732 polygons. In both maps, polygons were assigned attributes describing the land cover cover within their boundaries. Attributes included the primary and secondary dominant cover types (by proportional area), other cover types occurring within the polygon, wetland type codes, and codes describing disturbance, if any. Detailed description of mapping methods and map characteristics for each map are found for Wyoming in Merrill et al. (1996) and Driese et al. (1997) and for Colorado in (Schrupp et al. 2000).

Several map characteristics deserve emphasis because of their pertinence to the accuracy assessment. First, the primary map units in these maps are vector polygons, not raster pixels. The map polygons that designate landcover types were *not* derived from a per-pixel classification, but were digitized directly as polygons. For this reason, polygons were the basic sample unit for which map accuracy was assessed. Secondly, although polygon attributes describe both the dominant ("primary") and subdominant ("secondary") land cover within each polygon, we assessed only the accuracy of the mapped primary cover type. In Wyoming, the secondary cover type attribute was largely complete, but in Colorado attribution of the secondary type was partial. The relationship of the secondary cover to map accuracy was explored in some cases where data were available. The remaining polygon attributes (i.e. wetlands, disturbance type) were not assessed.

1.4 Sampling Considerations

Large, irregular polygons present challenges for accuracy assessment. Determining the dominant landcover within each polygon is a formidable problem when polygons contain heterogeneous mixtures of other landcover types. Sampling on the ground is impractical because it is difficult to "see" enough of a polygon to ascertain dominant cover or even to gain access to the polygon. Despite these difficulties, sampling should confidently estimate the dominant cover within polygons before they are used for accuracy assessment.

Assessment also requires sufficient sample size (enough sample polygons) to support confident accuracy estimates for individual cover types. In part this sample size is dependent on the mapped accuracy for each cover type, requiring either *a priori* estimates or samples sufficiently large to account for "worst case" scenarios. An efficient sampling scheme takes the former approach with *a priori* accuracy estimates for each cover type ideally based on the experience of the interpreter who assigned the polygon attributes.

Sampling issues are of obvious importance for accuracy assessment and should be considered in specific detail at the beginning of the assessment process. The sampling strategy used in this assessment is described in detail in Chapter 3 of this report. Many alternative sampling strategies are possible, and each project should consider goals, data requirements and field characteristics when choosing a sampling regime. It is important to balance statistical rigor with practical considerations and to recognize the trade-offs between them.

1.5 Traditional Map Accuracy Assessment

Map error includes boundary (error in the location of boundaries) and thematic error (error in the labels assigned to map units) (Bolstad and Smith 1992), each requiring unique assessment techniques. The work described in this report addresses only the latter; assessment of the thematic accuracy of polygon attributes.

Traditional map accuracy assessment is based on the premise that a map unit is mapped either correctly (absolutely right) or incorrectly (absolutely wrong). This binary scale does not consider degrees of map error or ambiguity in map classification systems, both of which have received increased attention in recent years (Woodcock and Gopal 1992, Gopal and Woodcock 1994). However, many accuracy assessment projects use these traditional binary measures, and this approach provides the opportunity for comparison of the Colorado landcover map to other maps. Such assessments are also easier to interpret for map users not familiar with fuzzy accuracy measures. The traditional approach is not inferior to fuzzy assessments, but additional information can be gathered by considering fuzzy measures of accuracy along with traditional ones. In fact, a successful accuracy assessment addresses map error from many angles to help users understand the strengths and weaknesses of a map for a particular use.

A contingency table, also called an ‘error’ or ‘confusion’ matrix, whose columns are the mapped cover types and whose rows are the reference cover types (i.e. ground truth”or ground control) forms the basis of traditional assessment. Each cell in the matrix tabulates the correspondence between mapped and reference landcover. Cells along the diagonal of the confusion matrix represent correctly mapped sample units. Off-diagonal cells represent map errors. In many instances, it is the examination of the off-diagonal matrix elements that provides the most information about map accuracy (Congalton and Green 1993, 1999) because they illuminate types of error and their causes.

Common measures of map accuracy generated using a confusion matrix include overall, producer's and user's accuracies. Overall map accuracy considers the map as a whole by summing the cells on the matrix diagonal (correctly mapped sample units) and dividing by the total number of sample units (Congalton and Mead 1983). This quantity, usually expressed as percent correct, is the broadest measure of map accuracy used for comparison between maps and is perhaps the most easily understood by map users. However, overall map accuracy does not consider correspondence between mapped and reference data due to chance and can be misleading. An area composed of 75% sagebrush, if completely mapped as sagebrush, would have 75% overall map accuracy though this is largely a result of chance.

Chance is considered in several accuracy statistics, including Kappa (Cohen 1960), discussed later in this report and the Tau statistic, proposed as an improvement on Kappa by Ma and Redmond (1995). Kappa has been widely used in the remote sensing literature as a measure of overall accuracy and will be reported here. The Kappa statistic gives the overall map accuracy *not* resulting from chance, so a Kappa of .71 indicates that the map accuracy is 71% higher than the expected accuracy.

Producer's and user's accuracy describe map accuracy for individual cover types and provide specific information to the map makers (producers) and the map users. For a cover type, producer's accuracy is the number of units on the ground that are correctly depicted on the map. User's accuracy describes the proportion of a cover type on the map that matches that type on the ground. Together these describe different aspects of the overall map accuracy and supply more detailed information about types of error.

Traditional assessment is important because it is intuitively straightforward and easily interpreted by map users. The reporting of traditional measures of map accuracy has a historical precedent, which allows comparison between maps based on similar assessment methodology. Finally, traditional accuracy, considered with other accuracy measures create a clearer picture of the map's strengths and limitations. Traditional accuracy measures are discussed in detail in Chapter 4 of this report.

1.6 Fuzzy Map Accuracy Assessment

Some class attribution errors that occur on thematic maps are more serious than others. Classification of lodgepole pine as Douglas fir might be of less consequence than classification of the same forest as a greasewood flat, depending on the user's objectives. Fuzzy accuracy assessment recognizes this and uses degrees of right and wrong to quantify accuracy in a "fuzzy" sense. Gopal and Woodcock (1994) presented this concept for assessment of spatial data and developed a collection of operators that measure aspects of fuzzy accuracy. We based the fuzzy analysis of the Colorado land cover map on their work.

Researchers commonly use a 5-level scale (Chapter 5) that assigns codes to degrees of right and wrong to rate the seriousness of matches or mismatches between reference data (in our case interpreted video) and mapped data (the landcover map), thus recognizing that mismatches may not be completely wrong or right. By assigning a code from the verbal scale to each match or mismatch, fuzzy accuracy can be quantified using a collection of indices designed to highlight different aspects of map error.

Gopal and Woodcock (1994) suggest that information about map errors should address the frequency, magnitude, nature and source of the errors. Their operators are designed to highlight these aspects of error. Frequency of errors are described using functions called MAX and RIGHT to distinguish best matches from acceptable matches between reference and map data. MAX is identical to user's accuracy (section 1.5) because only polygons that perfectly match the reference data are considered correct. RIGHT considers all polygons that correspond to reasonable or acceptable (Code 3, Table 5.1) or better as correct, and recognizes that some mismatches are acceptable.

The magnitude of errors are measured by the DIFFERENCE operator, which calculates the difference between the score given for match (or mismatch) between the mapped polygon and the reference data and the score of the best match (code 5 from Table 5.1). Mean DIFFERENCE is an index of how far the mapped type is from the ideal type for that polygon.

Sources of error are explored using the MEMBERSHIP operator which counts, for each land cover class, the number of other classes that could have been mapped while still meeting or exceeding an acceptability threshold. Lodgepole pine forest, for example, may be mapped as many other evergreen classes and receive a score of ‘reasonable or acceptable’ or better and would have a high MEMBERSHIP. Open water, in contrast, may have only a few types with which confusion is not considered a serious error.

Finally, Gopal and Woodcock quantify the ‘nature of errors using operators called CONFUSION and AMBIGUITY. CONFUSION lists categories for which a correctness rank is higher than the category that was actually mapped, and identifies better answers than the mapped class. For example, if 15 cover types would have provided a higher rank from table 1.1 than the type that was acually mapped, the polygon would show higher CONFUSION than a polygon for which only 10 cover types were better answers. Similarly, AMBIGUITY identifies classes with the same correctness rank as the mapped class. A polygon with 15 equivalent answers has higher AMBIGUITY than one with 10 equivalent answers. Together these functions highlight the types of mismatches that occur. For example, they may indicate that in sites where there is ambiguity between two or more types, the site is most often mapped as only one of the choices, indicating interpreter bias towards one of several equivalent choices.

Fuzzy accuracy measures provide new tools for describing the accuracy of thematic maps. Together with the traditional measures, they give users a picture of the types of uncertainty in the Colorado Gap land cover map, and their seriousness for habitat modeling. Fuzzy accuracy of the Colorado landcover map is discussed in detail in Chapter 5 of this report.

1.7 Airborne Videography as Proxy for Ground Truth

Comparison of reference data to mapped thematic units is the basis of accuracy assessment. Reference data gathered in the field ("ground truth") have historically been used to assess maps or classifications created from remotely-sensed imagery. While this makes sense for some maps, practical difficulties, especially in large, sometimes remote, areas diminish its usefulness. In fact, even under ideal conditions, pixels can be difficult to locate on the ground. The problem is compounded when sample units are large and heterogeneous. To overcome this, remotely sensed data at finer scale than that of the map can serve as proxy for ground truth, eliminating the problem of access in remote terrain or on private land, and providing a statistically unbiased sample for estimating the cover where heterogeneity may be difficult to untangle on the ground. Aerial photography, and more recently, aerial videography have been used for this purpose (Slaymaker et al. 1996, Marsh et al. 1994).

Remotely sensed reference data, while overcoming sampling and logistical problems for large projects, present challenges of their own. Interpretation, even at fine resolution, injects uncertainty into accuracy assessment when reference data are misinterpreted. All accuracy assessments assume that the reference data are true, and although this assumption is questionable even for ground-based reference, it is more questionable when reference data are interpreted remotely. It is important to understand, to the extent possible, the consequences of errors in reference data and their effect on the results of the assessment of the landcover map. These problems are discussed in Chapter 6 of this report along with other *caveats* and limitations.

1.8 How This Report Is Organized

This report is organized into chapters detailing the key aspects of the accuracy assessment. After this introductory chapter, we discuss video acquisition (Chapter 2) and development of the sampling scheme and a simulation study on the efficiency of video transect sampling (Chapter 3). Work through Chapter 3 was completed in both Colorado and Wyoming, and both states are discussed. The remaining chapters pertain primarily to the Colorado assessment. Discussion of traditional (Chapter 4) and fuzzy (Chapter 5) accuracy assessments, forming the core of the project, follow. Chapter 6 highlights limitations of the assessment, and specific uncertainties encountered in the Colorado work. It also includes an evaluation of airborne videography as a tool for accuracy assessment in the Northern and Central Rockies. Several appendices include large data tables and matrices, including the contingency tables, used in the assessment. We hope that this report serves both as a description of the accuracy of the Colorado landcover map and as a reference for future assessments using videography.

CHAPTER 2

Acquisition of Air Videography

2.1 Introduction

Acquisition of airborne videography includes the assembly of video equipment, the logistics of flying multiple transects to collect video data across each state, and the processing of the resulting data to prepare them for use in the accuracy assessment. For Wyoming and Colorado, these steps were accomplished with the support of both the Colorado Division of Wildlife (Donald Schrupp) and Dana Slaymaker, of the University of Massachusetts. Their expertise has been central to Gap videography efforts. The technical challenges that must be overcome to capture airborne videography are substantial and to start from scratch would be formidable.

Many aspects of the assessment relate to the acquisition of videography. In particular, the design of flight transects is determined by sampling requirements. We save the discussion of sampling for Chapter 3 and concentrate here on the technical aspects of assembling and deploying the video cameras, and on the preliminary data processing.

2.2 Videography Equipment

Videography equipment used in Colorado and Wyoming mimicked the system described by Slaymaker et al. (1996) (Fig. 2.1) and was purchased by CDOW. Two parallel high-8mm analog video cameras formed the core of the system. Cameras were mounted for aircraft with "belly ports", allowing containment of the system within the aircraft, rather than extended from a window. By setting one camera at zoom and the other at wide angle, a detailed view of the ground and a wider contextual perspective were captured simultaneously. On-the-ground "footprint" size of the zoom view was about 30 m for a flight altitude of 610 m above ground level (AGL). The corresponding wide angle "footprint" was about 500 m. The cameras capture video at 30 frames/second.

A GPS unit recorded the position of the aircraft once per second. Position and GPS time (transmitted from the GPS satellites) were recorded on a laptop computer using Geolink software. Simultaneously, GPS time was written to each video frame by Horita time stamp recorders. Time stamps allowed determination of frame position using the time/coordinate file stored on the laptop. No correction was made for aircraft pitch or roll, both of which contribute to positional inaccuracy of the video frames on the ground. Recent improvements in airborne video systems (Slaymaker - personal communication) incorporate digital gyroscopes (Watson boxes) that correct for this problem. We feel that for assessing large polygons, like those in the Colorado and Wyoming maps, the positional inaccuracies in our system were acceptable, although they may contribute to uncertainty near polygon edges. For projects that require locating individual pixels, higher positional accuracy would be essential. GPS positional accuracy was enhanced by differential correction (Section 2.4.1).

Two portable 8 mm video recorders (for zoom and wide angle cameras) recorded the video on Sony ProMP 120 minute high-8mm tapes during flight. Portable monitors allowed viewing of the video in real time to ensure that the cameras were operating properly and that the time stamp was recorded. The recording decks and monitors were strapped to a platform behind the passenger seat of the aircraft so that they could be accessed while in the air. Cameras were less accessible during flight, and were started on the tarmac before take-off.

Figure 2.1 Schematic diagram of the video equipment used to record analog videography for Wyoming and Colorado. This system mimics the equipment originally assembled by

Slaymaker et al. (1996).

2.3 Flight Logistics

2.3.1 Colorado

Colorado videography was acquired from a CDOW Cessna 172 along 8 systematic transects across the north-south extent of the state (Fig. 2.2). Individual transects were about 450 km long and required 2 hours of flight time, meaning that data for each transect could be recorded on a single video cassette (for each of the zoom and wide angle cameras). Total transect length for all 8 flight lines was 3600 km in Colorado. Flight altitude (AGL) was maintained as close to 2000' (610 m) as possible, though variation was unavoidable in high-relief terrain. In Colorado, the CDOW pilot gained altitude in anticipation of mountain crossings by spiraling where possible, rather than by ramping altitude along the transect, as was done in Wyoming (see below).

Undiagnosed problems with one of the video cameras caused significant blurring of the imagery along flight line 7 in Eastern Colorado. Because landcover along this line was largely redundant with lines 6 and 8 (prairie/agriculture mixture in the Great Plains) the line was not reflown. Other flight lines were generally of high quality, though occasional short sections of unusable video were discovered during interpretation. Sufficient samples were available without these "video drop out" areas.

2.3.2 Wyoming

Wyoming videography was acquired from a USFS Cessna 206 from 17 - 20 June, 1997. As in Colorado, 8 systematically spaced, north-south transects were flown (Fig. 2.3). Also as in Colorado, each transect required about 2 hours of flight time and covered approximately 450 km, with 3600 km of total transect length from all 8 flight lines. During flights, altitude (AGL) of approximately 2000' (610 m) was maintained, although the plane increased altitude in anticipation of mountain ranges. In Wyoming, altitude was gained by ramping along transect lines, rather than by spiraling upwards near the foothills as was done in Colorado. Transect lines were followed using an onboard GPS that was part of the plane's navigation system. Occasional small deviations from transects were necessary to avoid storm cells, although the weather was generally good during all transect flights. Videography from all 8 transects flown in Wyoming were of high quality and cameras functioned consistently.

Figure 2.2. Videography transects flown in Colorado. Lines 1 and 2 were flown on 10 October, 1996, lines 3 and 4 on 11 October 1996 and line 6 on 9 October, 1996. Line 5 was flown on 22 July 1997. Lines 7 and 8, were flown on 7 January, 1997. The latter flight took advantage of a snow-free period for these Great Plains transects.

Figure 2.3. Videography transects for Wyoming. Transects were flown starting in eastern Wyoming (line 8) on 17 June, 1997 and finishing in western Wyoming (line 1) on 20 June, with 2 transects flown each day.

2.4 Processing of Videography and Associated Data Files

Video flights provide two products: (1) the analog, time-stamped video tapes and (2) the associated data file containing positional data and time stamps from the GPS. Interpreter training, sampling of individual video frames for accuracy assessment, and statistical analysis of the samples all required manipulation of point representations of individual frames in a GIS. To accomplish this, the original data files captured on the laptop during transect flights were processed to create Arc/Info GIS point coverages for Wyoming and Colorado, with each point representing a single video frame corresponding to a GPS position. Because the GPS recorded positions once each second and because video was recorded at 30 frames/second, only 1/30th of the available video was used. This still represents a very large (30,000 frames/state) pool of frames from which to sample. Video frames recorded every 1/30th of a second overlap significantly, and sampling every 30th frame ensures that samples are not redundant.

Processing included differential correction of the GPS data; addition of textual information describing transect number, flight date and differential correction status (corrected or not corrected); and finally, creation of the arc/info point layer using the modified text file and attributes. Once this GIS point layer was created it was combined and overlaid with other spatial data (e.g., the land cover map and road data) using standard GIS techniques to generate samples for ground truthing and interpretation. The latter steps will be discussed in Chapter 3 (Sampling) and we will limit the discussion here to conversion of text files collected in flight to GIS point coverages for each state.

2.4.1 Differential Correction

When Colorado (and Wyoming) transects were flown, selective availability (SA) was *not* turned off and GPS positional accuracy could be improved significantly by post-processing using base station data. Colorado GPS data were corrected using two base stations. In eastern Colorado, the CompassCom Trimble Community base station located in Littleton, Colorado (39^o 37' 9.15" N Lat., 104^o 59' 22.63" W Lon.) was used. This base station is maintained by CompassCom as a public service to the GIS/GPS community in Colorado, and can be used to achieve sub-meter accuracy within a 300-mile radius of Denver. In western Colorado, we used a Trimble community base station maintained by Mesa County and located atop the Mesa County Courthouse in Grand Junction, CO (39^o 04' 7.55" N Lat, 108^o 33', 47.42" W Lon.). Differential correction was accomplished using the Pathfinder Office software (pfinder) from Trimble. Failure of the GPS unit to capture ephemeris data caused some uncorrected video frame locations. This problem was partially solved by copying ephemeris data from base station files, but still prevented correction of all points (Table 2.1). Experience in the field suggested that positional accuracy was on the order of +/- 100m, largely due to uncorrected aircraft pitch and roll.

Table 2.1. Percentage of GPS points corrected for each videography transect in Colorado. The 8 transects are numbered from west to east, and correspond to the transect numbers in figure 2.2.

Transect Number	***Proportion of successful differential correction (%)***
1	100
2	44

Transect Number	*Proportion of successful differential correction (%)*
3	100
4	100
5	84
6	100
7	93
8	99

Wyoming differential correction was accomplished using the University of Wyoming - BLM GPS community base station located on the roof of the BLM Casper District Office in Casper, Wyoming (42^o 51' 23.43" N Lat., 106^o 18' 11.02" W Lon.). The base station uses a Trimble Pathfinder PROXL single frequency mapping GPS receiver, and due to its central location in Wyoming, allows correction of data throughout the state, all of which is within a 300 mile radius of the base station. As in Colorado, failure of the GPS unit to capture proper ephemeris data caused some points to be uncorrected. Copying appropriate ephemerides from the base station data was less effective in solving this problem in Wyoming (Table 2.2). Similar positional accuracy as seen in Colorado (+/- 100m) was noted during Wyoming field work, again due to uncorrected aircraft pitch and roll.

Table 2.2. Percentage of GPS points corrected for each videography transect in Wyoming. The 8 transects are numbered from west to east, and correspond to the transect numbers in figure 2.3.

Transect Number	*Proportion of successful differential correction (%)*
1	65
2	83
3	97
4	13
5	0
6	0
7	62
8	100

2.4.2 Conversion to Arc/Info Point Layer

After differential correction, columns were added to the GPS files to indicate whether differential correction had been successful for each point, the flight line from which the data were collected, and the date of the flight. The resulting delimited text file was used to create a point coverage in Arc/Info, with the original attribute columns imported into the 'Info' database associated with the point locations. Coverage points represent the aircraft position when each frame was acquired projected onto the ground. Differentially corrected coordinates were used when possible and uncorrected coordinates otherwise. The result of the post flight processing was an Arc/Info coverage of video frame locations with attributes for each frame describing position in UTM, flight date, flight line, and differential correction status.

2.5 Archiving

Analog video tape deteriorates with use. When video is used for interpretation, frequent "frame freezing" accelerates the deterioration as the heads in the video deck spin in one place on the video tape. For this reason, the original video cassettes for each flight line in both Wyoming and Colorado were copied to duplicate analog video tapes. The "master" tapes were archived and the copies used for all subsequent work. The Colorado Division of Wildlife (Schrupp) holds the original tapes for Colorado. Wyoming masters are stored at the University of Wyoming Department of Botany (Reiners and Driese). Although analog video loses substantial quality when copied, we felt that the value of the originals dictated preservation. Copies are still of sufficient quality for interpretation. Digital video, of course, does not deteriorate with use or copying, and should provide a solution to this problem in the future.

CHAPTER 3

Sampling for Accuracy Assessment and Interpreter Training

3.1 Introduction

This chapter discusses sampling in four contexts. First, we explored the effect of various, simulated flight transect scenarios on sample sizes of landcover classes. Second, we sampled to select a set of frames that were visited on the ground to produce an interpretation key that was used for interpreter training and reference. Third, we chose a set of map polygons for the accuracy assessment itself. Since polygons were the fundamental assessment unit, we required an unbiased sample of polygons representing the mapped landcover types. Fourth, because mapped polygons were large and contained mixtures of dominant landcover and less important inclusions, we selected a sample of video frames from within each sample polygon for interpretation. This was required in order to provide a confident estimate of the dominant landcover within each sample polygon. The chapter also describes interpreter training procedures followed prior to interpretation of videography data.

3.2 Transect Sampling and Diminishing Returns: A Simulation Experiment

The distributions of areas occupied by individual cover types in state landcover maps are highly skewed (Fig. 3.1). With much of the state areas in a few types and many less common types occupying small areas, it is difficult to visit some cover types using videography transects without large flight time and cost. To explore this issue, we simulated hypothetical flight lines of different orientations (north-south, east-west, random), lengths (450km - 4500km) and configurations (random vs. systematic) across Colorado, Wyoming and Arkansas to test maps of different areas and different mapping methods. The numbers of cover types encountered for each scenario were tallied and compared. The simulation was designed to answer several questions about video transect sampling: 1) Is there a critical transect length beyond which additional flight distance adds few types to the sample?, 2) Does transect orientation affect the relationship between flight length and number of types visited?, and 3) Does transect configuration affect this relationship? To answer these questions, we plotted curves describing the relationship between total transect length and number of types visited for the different combinations of transect orientation, configuration and length.

Figure 3.1. The distribution of land cover proportional areas in (a) Colorado, (b) Wyoming and (c) Arkansas. The highly skewed distributions make it difficult to visit all types with videography transects.

Wyoming and Colorado were mapped using identical methodology (Merrill et al. 1996, **Schrupp et al - Colorado final report**) but their maps differ in the number of landcover types (52 in Colorado; 41 in Wyoming) and in the broad distribution of landcover. Colorado has strong east-west zonation due to the natural break between Great Plains vegetation in the eastern third of the state and alternating mountains and mountain valleys in the western two-thirds. In Wyoming, zonation is less pronounced, with mountain ranges throughout the state separated by broad basins (Knight 1994). The Arkansas Gap landcover map was created from an unsupervised classification of enhanced and stratified TM data (Smith et al. 1998) and aggregation of the result to achieve the 100ha Gap MMU. The Arkansas map depicts 36 landcover classes and landcover itself is different in Arkansas than in Wyoming and Colorado, where climate is more continental and elevations are higher. In Arkansas, zonation between the highlands of the north and west and lowlands in the south and east is not as strong as the zonation in Colorado, and human disturbance is more widespread.

3.2.1 Transect Simulation Methods

Sets of simulated video transects were generated for each of a variety of orientations (Table 3.1). For each transect orientation, configuration was either random or systematic. Randomly spaced transects were oriented east-west, north-south, or randomly within each state map. Systematically spaced transects were either east-west or north-south. For each combination of transect spacing and orientation, sets of one to ten transects each were generated to simulate increasing sampling effort (450 to 4500 km total). Finally, for the randomly spaced transects, 10 sets of each orientation and length were generated to characterize variability. Because the systematic transects were placed in a regular pattern across the extent of each landcover map, only one set per length of these transects was generated.

Table 3.1. Video transect scenarios used for modeling in Colorado, Wyoming and Arkansas. For each combination of random transect placement and orientation, 10 sets of transects were generated for each of 1 to 10 transect lines per set, resulting in 100 total sets for the random transect placements. One set of from 1 to 10 transects was generated for each systematic orientation, resulting in 10 total transect sets for these simulations.

Placement	Orientation	Number of lines	Number of Sets per Scenario	Total Number of Sets
Random	East - West	1 - 10	10	100
Random	North - South	1 - 10	10	100
Random	Random	1 - 10	10	100

Placement	Orientation	Number of lines	Number of Sets per Scenario	Total Number of Sets
Systematic	East-West	1 - 10	1	10
Systematic	North-South	1 - 10	1	10

A GIS (Arc/Info) allowed tabulation of the total transect length within the landcover maps for each of the scenarios. The number of landcover types intersected by each set of transects was determined by overlaying the transects on the landcover maps. This allowed construction of characteristic curves describing the change in total number of intersected types as sampling effort (total transect length) increased.

Total transect length was plotted against the number of cover types intersected by each set of transects for each scenario and 2-parameter exponential curves were fit to the resulting scatter plots. Curve-fitting with additional parameters improved the fit slightly, but at the expense of interpretability and so was not used. All curve-fitting was done using SigmaPlot 4.0 (SPSS, Inc., Chicago, IL). The resulting exponential curves are of the form:

$$(1) \qquad y = a(1 - e^{-bx})$$

where y is the number of land cover types intersected, x is the total transect length (km), and ‘a’ and ‘b’ are model parameters unique for each transect scenario and determined empirically by iteration to achieve a best fit curve for each scenario. The exponential curves describing the relationship between total transect length and the number of cover types encountered were analyzed to identify: 1) the transect length beyond which increasing length provided little increase in the number of cover types intersected (the level of diminishing returns), 2) differences for the various curve orientations, and 3) differences in curve shape and asymptote between systematic and random placement of video transects. To provide a consistent basis for comparison, we defined the level of diminishing returns as the point on the mean exponential curve where its slope reached 0.001. Physically, this was where 1000 km of additional transect length was required to add 1 landcover type to the sample, a level beyond which video sampling is arguably not cost effective.

3.2.2 Results of Transect Simulations

Imposing exponential curves (Equation (1)) onto the data scatter is artificial in the sense that sufficient transect length would eventually capture all mapped types in a state (e.g., 41 for Wyoming), but for the maps analyzed here, the asymptote of the fitted curves is an index of the level of diminishing returns beyond which additional sampling produces only small gains in the number of types encountered. Because many mapped types in these states are rare (Fig. 3.1), prohibitively long transect lengths are required on average before these types are encountered (Fig. 3.2).

Figure 3.2. Best fit exponential curves representing the relationships between total transect length and the number of landcover types encountered for (a) Colorado, (b) Wyoming and (c) Arkansas for the various transect orientation and configuration scenarios.

For Colorado, diminishing returns defined by our slope criterion of 0.001 types/km occurs at 46 land cover types (75%), corresponding to a total transect length of 3355 km (Fig. 3.2a). For Wyoming, diminishing returns occurs at 33 cover types (Fig. 3.2b) representing 80% of all mapped types and corresponding to a total transect length of 2495 km. Diminishing returns in Arkansas occurs at 25 cover types (69%) and 1769 km of transect length. These differences between states are substantial, and suggest that simulation exercises are worthwhile for planning sampling strategies for large accuracy assessment projects. Beyond the level of diminishing returns, extension of airborne videography transects adds little to the sample size, at least in the number of sampled cover types (though additional samples are added for already visited types).

Exponential curves do not quantify the rate at which the rare types not encountered early in transect sampling are added. Total transect lengths at the level of diminishing returns are approximately half of the longest transects generated in our simulation for each state (Fig. 3.2). Doubling transect length in Wyoming from 2500 km to 5000 km increases the number of cover types encountered from 33 to about 35, an increase that might be better achieved by using other sampling strategies for rare types. Similar small increases after diminishing returns are observed in Colorado. Despite this, we flew significantly beyond the transect length corresponding to diminishing returns in Wyoming and Colorado during sampling to improve sample size of cover types that were visited.

Our second research question addresses the orientation of the transect lines. Results (Fig. 3.2) reveal only small differences between the best fit exponential curves for each state for the various transect orientations. E-W, N-S and randomly oriented curves all converge on approximately the same asymptotic number of types for a state, and within states all orientations reach this asymptote at approximately the same transect length. Colorado, where zonation between the eastern plains and the western mountains is more distinct than in Wyoming, shows a faster initial increase in captured cover types for E-W transects than for N-S transects. Nevertheless, the curves converge at a transect length of about 4000 km (Fig. 3.2), approximating the calculated point of diminishing returns. For all three states, randomly oriented transects achieved a slightly lower level of diminishing returns than did E-W or N-S transects , but again differences are small.

Lastly, systematic transects show a slightly higher rate of increase to the asymptotic level and achieve a small advantage in the number of captured types at the level of diminishing returns (Fig. 3.2) than do randomly spaced transects. The advantage of systematic transects is probably the increased probability that they will sample all broad landcover zones because the transects by design are spread across the latitudinal or longitudinal extent of each map. There is no guarantee that randomly placed transects will sample all zones, although at long transect lengths (longer than used in our simulations) the results should converge. Differences appear related to broad-scale zonation in land cover, but even in Colorado, where the differences between results for systematic and randomly placed transects are greatest, the difference in transect length between the most efficient and least efficient rates of capture represent only 1000 km of transect length in a state that requires 4000 km of video transect to achieve diminishing returns. These results, coupled with the practical considerations of planning flight lines, may dictate the use of systematic rather than random transects, but the advantages in sampling efficiency are small and alternatives for particular sampling strategies should not be discarded.

3.3 Sampling for Ground Truthing of Videography and Interpreter Training

Gap landcover types are abstractions like many other ecological entities and geographically defined features. A necessary initial step is to ensure that the interpreter accurately relates the images displayed in videography with this abstraction. This is not straightforward, since many Gap cover types can be distinguished on video only by recognizing subtle features, and some types cannot be distinguished at all (Chapter 6). One way to improve interpretation accuracy is to build a visual key with which to train the interpreter before interpretation begins and to which the interpreter can refer during interpretation. Visual keys for Colorado and Wyoming were constructed by visiting sites on the ground corresponding to video frames from as many landcover types as possible. We visited several examples of some types in each state to capture within-type variability. Finally, we used these site visits to construct a key, containing hardcopy prints of zoom and wide angle video for each site, annotated with notes from the field and accompanied by site photos and descriptions. In this section we describe the process used to construct this key.

3.3.1 Ground Truth Sampling

Construction of a photointerpretation key requires site visits on the ground but is not dependent on statistical rigor and unbiased sampling because the samples are not used in inferential statistics. Efficiency in the field is important, however, to ensure visitation of a large selection of sites for capturing variability in cover types. In Wyoming and Colorado, we used access from roads as a criterion for site selection..

Road access was a primary criterion for site selection in Wyoming and Colorado. U.S. Census Bureau TIGER road data were used to select sample video frames from the GIS point data corresponding to the video transects (Chapter 2). The TIGER data contain attributes describing road type (e.g., limited-access, secondary, etc.). In the GIS, we selected only roads accessible using 2-WD vehicles, and created 500m buffers on either side. These buffered road data allowed selection of a subset of the available video frame data falling inside the buffers. A total of 155 sites were chosen for Colorado and 163 for Wyoming representing 47 and 27 cover types respectively. Because some cover types were common to Colorado and Wyoming, ground truth training sites from Colorado were occasionally used in the Wyoming interpretation key.

3.3.2 Field Visits

For each of the field sites chosen for ground truth visits, photographs were taken of the zoom and wide angle video frames and labeled with video frame number, location in UTM, a north arrow, and transect number. Additionally, the wide angle video was viewed for the area on either side of each target frame and a sketch maps drawn on "post-its" were affixed to each photo pair as an aid to finding the frame in the field. Dirt roads and landmarks were often visible in the videography that did not show up on topographic maps, and the sketch maps were valuable field tools. Sample frame locations were transcribed onto 1:100,000 scale USGS or BLM topographic maps and labeled with the frame numbers.

In the field, a Trimble GeoExplorer II GPS unit with an external antenna, along with the frame photographs, sketches and photographic maps were used to locate frames. The GPS antenna was attached to the roof of the field vehicle using a magnetic mount so that a relatively clear view of the sky was maintained while driving. These tools enabled one to drive along roads to within close vicinity of the target site. The GPS was then carried away from the road and used to get close to the actual frame location, which was pinpointed when possible by identifying objects in the video frame that were visible on the ground. The resolution of the zoom frames was such that individual shrubs, downed logs, and small boulders could be identified and used to find frames.

Once frame locations were located, one or more site photographs were taken and objects in the zoom and wide angle photographs of the videography were labeled directly on the photographs. These photo labels were valuable interpretation tools. Notes describing the site and its surroundings were recorded and in some cases, photographs emphasizing the growth habit of dominant individuals were taken as a reference for particular cover types. In total, 99 of the 155 sample sites chosen for Colorado were accessed (64%). In Wyoming, 114 of 163 sites were visited (70%). Private land or poor roads were the reason for most missed sites.

3.3.3 Construction of Interpretation Keys

For each state, field results were organized into 3-ring binders to serve as interpretation keys. Binders were organized by cover type, and the sections for each type included the hardcopy prints of zoom and wide angle video images, annotated site photos taken in the field, field notes, photographs of growth form for dominant species, and the formal cover type descriptions developed for the original mapping in each state. The single copies of the interpretation keys for each state are held at the Dept. of Botany, University of Wyoming (Driese and Reiners) and are available for loan upon request.

3.3.4 Interpreter Training

For Colorado, video interpretation was performed by vegetation mappers and photointerpreters (Thomas Owens, USGS and Kenneth Driese, UW Dept. of Botany) with knowledge of the vegetation of the state. Thus, interpreter training required only study of the interpretation key and examination of the field sites on the analog video viewed on-screen. Interpreters not familiar with the landcover of Colorado would have required a more rigorous training program than was needed for this study.

3.4 Sampling of Polygons for Assessment of Cover Types

As emphasized previously, polygons are the fundamental map unit in the Colorado and Wyoming landcover maps and form the basis of the accuracy assessment of these maps. The number of sample polygons needed to achieve a given standard error for accuracy estimates depends on the proportion of correctly mapped polygons available for each cover type--a number that cannot be known with confidence until after the accuracy assessment is complete. To estimate these sample sizes (n), we made *a priori* estimates of individual cover type accuracy (Tables 3.2, 3.3) based on the experience of the photointerpreter who did polygon attribute assignment for Wyoming (Driese). Using a standard error of 0.08 and the equation:

$$n = (N*p*(1-p))/(N*s^2+p*(1-p))$$

where N = the total number of mapped polygons of each type in each state
p = the *a priori* estimated accuracy for that type
s = the desired standard error (0.08)

we estimated the number of polygons necessary for the initial sample (Tables 3.2, 3.3). Some cover types did not have enough polygons containing video to achieve these minimum sample numbers and some had no video coverage. For these types, all possible polygons were sampled. Sample polygons were chosen randomly. In Colorado, 593 polygons were sampled for the cover types that had video coverage (Table 3.2). In Wyoming 552 polygons were sampled (Table 3.3).

Table 3.2. Colorado land cover type (see Appendix 1 for legend), *a priori* estimated accuracy, estimated number of polygons to achieve standard error of 0.08, available polygons with video data, and the actual number of polygons sampled for interpretation in Colorado.

Colorado Cover Type	***A priori Est. Accuracy***	***Estimated Sample Size***	***Available Polygons***	***Actual Sample Size***
1	0.75	29	102	27
2	0.80	7	0	0
3	0.75	28	38	27
4	0.80	25	53	17
5	0.80	12	2	1
6	0.75	4	0	0

Colorado Cover Type	***A priori Est. Accuracy***	***Estimated Sample Size***	***Available Polygons***	***Actual Sample Size***
7	0.80	25	59	25
8	0.80	2	0	0
9	0.70	4	0	0
10	0.75	27	14	14
11	0.75	29	138	29
12	0.70	14	2	2
13	0.80	21	11	11
14	0.90	13	4	4
15	0.75	28	54	28
16	0.80	25	76	27
17	0.80	12	10	10
18	0.70	24	6	5
19	0.70	22	13	11
20	0.70	20	6	6
21	0.70	21	0	0
22	0.75	13	1	1
23	0.70	32	71	32
24	0.75	27	14	6
25	0.75	26	15	11
26	0.75	25	15	15
27	0.80	22	17	17
28	0.70	1	0	0
29	0.70	26	4	4
30	0.80	23	23	23
31	0.80	15	4	4
32	0.70	28	23	21
33	0.70	31	34	31
34	0.70	32	90	46

Colorado Cover Type	*A priori Est. Accuracy*	*Estimated Sample Size*	*Available Polygons*	*Actual Sample Size*
35	0.80	16	10	10
36	0.70	31	29	28
37	0.80	21	18	16
38	0.70	28	8	7
39	0.80	21	7	2
40	0.80	25	66	29
41	0.85	20	118	25
42	0.80	1	0	0
43	0.70	2	0	0
44	0.80	22	1	0
45	0.85	15	7	6
46	0.85	11	1	0
47	0.85	7	0	0
48	0.85	12	1	0
49	0.85	1	0	0
50	0.90	13	4	4
51	0.80	9	1	0
52	0.90	14	9	9

Table 3.3. Wyoming landcover types (see Appendix 1 for legend), *a priori* estimated accuracy, estimated number of polygons to achieve standard error of 0.08, the number of polygons with video coverage, and the number of polygons sampled in Wyoming.

Wyoming Cover Type	*A priori Est. Accuracy*	*Estimated Sample Size*	*Available Polygons*	*Actual Sample Size*
1	0.80	24	25	24
2	0.70	30	35	30
3	0.85	20	91	20
4	0.80	21	11	11

Wyoming Cover Type	***A priori Est. Accuracy***	***Estimated Sample Size***	***Available Polygons***	***Actual Sample Size***
5	0.60	22	2	2
6	0.65	27	11	11
7	0.80	24	37	24
8	0.70	31	31	31
9	0.85	16	6	6
10	0.60	7	0	0
11	0.75	27	22	22
12	0.85	19	33	19
13	0.70	12	4	4
14	0.70	27	11	11
15	0.60	4	3	3
16	0.75	28	50	28
17	0.80	25	366	25
18	0.70	1	0	0
19	0.70	16	0	0
20	0.70	31	53	31
21	0.75	27	27	27
22	0.75	27	23	23
23	0.80	12	5	5
24	0.80	16	4	0
25	0.80	24	32	24
26	0.75	29	214	29
27	0.70	9	1	1
28	0.75	13	0	0
29	0.75	28	30	28
30	0.70	17	2	2
31	0.70	26	4	4
32	0.80	24	37	24

Wyoming Cover Type	***A priori Est. Accuracy***	***Estimated Sample Size***	***Available Polygons***	***Actual Sample Size***
33	0.80	25	43	25
34	0.75	27	15	15
35	0.70	30	21	21
36	0.70	11	0	0
37	0.85	10	2	2
38	0.95	8	11	8
39	0.80	18	2	2
40	0.90	12	6	6
41	0.85	3	0	0

3.5 Sampling for Assessment of Vector Polygons

The number of samples (video frames) necessary to confidently estimate the dominant landcover *within* heterogeneous sample polygons depends on the proportional area of the dominant cover type. A polygon with relatively equal areas of dominant and subdominant landcover requires a larger sample than a polygon having the majority of its area in the attributed landcover type. Unfortunately, there is no consistent *a priori* measure of polygon heterogeneity for the Colorado map. Attributes describing the percent occupancy of each polygon by the primary and secondary cover types were candidates for such a measure, but these attributes were not consistently recorded in the Colorado map. In Wyoming, proportional occupancy of the dominant type was more consistently included in the map attributes. We calculated the number of samples (Table 3.4) required for several levels of proportional area of the dominant cover type in each polygon using the equation:

$$n = p(1-p)/s^2$$

where p = percent cover of dominant type (>50% by definition)
s = the desired standard error (0.08)

Table 3.4. The number of video frames necessary to identify the dominant land cover with a standard error of 0.08 for polygons with varying percent occupancy by the dominant cover type.

Percent Cover of Dominant Type	Number of Video Frames per Poly
60	37
70	33
80	25

Percent Cover of Dominant Type	Number of Video Frames per Poly
90	14

Because of the large number of sample polygons necessary for accuracy assessment of each cover type and the large number of cover types, we interpreted 15 frames per polygon. This level of within-polygon sampling results in lower confidence in estimates of landcover type within heterogeneous polygons, but still requires interpretation of more than 6000 video frames for each state. Many polygons in Wyoming and Colorado are homogeneous in the dominant cover type.

For each sample polygon (Section 3.4), individual video frames represented by the point GIS data were selected randomly when 15 or more frames were available. For sample polygons with less than 15 frames, all available frames were added to the sample. For Colorado, a total of approximately 6600 video frames were interpreted, although some of these frames were rejected during interpretation due to poor video quality. The set of sample frames was presented to the interpreter as a text file containing flight line, frame number, GPS time and corresponding ground elevation. Sample points for each flight line were sorted by GPS time to allow the interpreter to efficiently scroll through the tapes to find the sample, which was identified by the GPS time stamp in the video frame. Elevation data were included to aid the interpreter in making simple distinctions, such as between alpine grasslands and prairies.

CHAPTER 4

Traditional Accuracy Assessment

4.1 Introduction

Analysis of contingency tables (error matrices) is the core of traditional accuracy assessment. Contingency table analysis was performed on the Colorado data following methods described by Congalton and Green (1999), Campbell (1996) and others. To build contingency tables, interpreted video frames were used to estimate the dominant landcover type in each sample polygon. These polygons ("reference data") were compared to the mapped polygons (map data) and the count of each combination of map and reference cover types was stored in the corresponding cell in the contingency table (Appendix 2). The resulting matrices are used in this chapter to calculate and discuss overall map accuracy, user's accuracy, producer's accuracy, errors of omission and commission, and the Kappa statistic at the Gap- (level 5) and Anderson-levels (level 4) for Colorado. We also examine the mapped secondary (subdominant) cover type for each sample polygon to identify confusion between dominant and subdominant types in the map.

Readers are cautioned that these assessments are based on the assumption that the reference data (video interpretations of dominant cover) are "true". This assumption may not be valid in some cases, especially at the Gap-level where differences between types are difficult to distinguish in air videography. Essentially, a product produced with one form of remotely sensed data (Landsat TM) is being compared with another product produced with a different form of remotely sensed data (airborne videography). Errors in reference data are impossible to identify without ground visits and their effects on calculated map accuracy are difficult to quantify. These limitations are discussed in Chapter 6 and work is underway to understand how errors in reference data affect the results of accuracy assessment.

4.2 Overall Map Accuracy and the Kappa Statistic

Overall map accuracy is the sum of the cell counts along the diagonal of the contingency table (Appendix 2) divided by the total number of sample polygons--the proportion of perfect matches between map and reference data. Of all the map accuracy descriptors, this is probably the easiest to understand and the most widely reported, but offers the least information to map users. Instead, it is the off-diagonal elements of contingency tables that are informative because they highlight the types of confusion that occur in the map (Congalton and Green 1999) (see Section 4.3). In Colorado, overall map accuracy was 31% at the Gap level and 58% at the Anderson level.

Overall map accuracy does not consider matches between map and reference units that occur purely by chance. Chance agreement is quantified by the Kappa statistic, which measures map accuracy *not* expected by chance alone and provides a better measure of true accuracy than does overall map accuracy. The kappa statistic, calculated as:

$$k = (\text{observed} - \text{expected})/(1 - \text{expected})$$

where "observed" is the overall accuracy and "expected" is the accuracy expected by chance for the Colorado map was 28% at the Gap level and 48% at the Anderson level. While these overall accuracies are low, they should not be reported in isolation since fuzzy accuracy assessment (Chapter 5), analysis of the off-diagonal elements of the contingency tables (Section 4.3), and possible difficulty distinguishing some types in the air videography (Chapter 6) all paint a clearer picture of map uncertainty.

4.3 User's and Producer's Accuracy

4.3.1 Definitions

User's accuracy (consumer's accuracy) is map accuracy from the perspective of a map user interested in knowing how often a unit designated as a particular type on the map is indeed that type on the ground (Campbell 1996). For example, if 75% of the polygons labeled Douglas fir on a map are Douglas fir in the field, user's accuracy is 75%. In the contingency table, user's accuracy for a cover type is the number of perfect matches between mapped and reference polygons (the number on the diagonal) divided by the total number of polygons of that type in the map being evaluated.

Producer's accuracy is map accuracy from the perspective of a map maker (producer) interested in knowing how often a type on the ground is mis-labeled on the map. In other words, producer's accuracy is the proportion of the time that, for example, a person standing in Douglas fir on the ground would find that unit of Douglas fir on the map. In the contingency table, producer's accuracy is the number of correctly classified units of a particular cover type divided by the total number of reference units of that type.

Accuracy can also be expressed as errors of omission and commission. Errors of omission and commission are always reported with respect to a particular cover type and every error of omission for one type corresponds to an error of commission for another. Errors of omission are the occurrence of cover types on the ground that are *omitted* from the map. In the contingency table, omission errors are calculated by dividing the number of misclassified (off-diagonal) polygons of a particular cover type by the total number of polygons of that type in the reference data. Omission errors and producer's accuracy sum to 100%. Commission errors occur when a type on the ground is assigned the wrong type on the map. The erroneous type on the map is *committed* -- it is mapped where it should not be found. Errors of commission are the number of misclassified units of a particular type from the map, divided by the total number of units of that type in the reference data. Note that commission errors can be greater than 100% if a cover type appears in more map units than reference units and that commission errors and user's accuracy do *not* sum to 100%.

4.3.2 Gap-level Accuracy

User's and producer's accuracy and errors of omission and commission for the Colorado landcover map are reported in table 4.1 along with the number of polygons sampled for each cover type ("actual sample size" from Table 3.3). Discussion of these results for each of the individual cover types is difficult, and perhaps tedious. Instead, we have sorted the table in order of decreasing user's accuracy, and draw examples from the table for ranges of user's accuracy (UA). Categories are alternately shaded in table 4.1. To supplement this discussion, we urge map users to study the types of confusion in the map by examining off-diagonal elements in the contingency tables (Appendix 2).

Table 4.1. User's (UA) and producer's (PA) accuracy (% correct), errors of omission (EO) and commission (EC) (%) and the number of sample polygons (from Table 3.3) for each Colorado Gap-level landcover type. "NA" in the producer's accuracy column represents cover types that were never assigned to video frames by the interpreter. "NA" in the user's accuracy column represents cover types from the map that were not sampled by video transects. Cover type code numbers are defined in Appendix 1. The table is sorted by increasing user's accuracy and shading reflects categories of user's accuracy discussed in the text.

Cover Type	***UA (%)***	***PA (%)***	***EO(%)***	***EC (%)***	***No. Samp.***
10	0.00	NA	NA	NA	14
12	0.00	0.00	100.00	66.67	2
13	0.00	0.00	100.00	110.00	11
14	0.00	0.00	100.00	15.38	4
18	0.00	0.00	100.00	250.00	5
20	0.00	0.00	100.00	300.00	6
22	0.00	NA	NA	NA	1
24	0.00	0.00	100.00	75.00	6
29	0.00	0.00	100.00	400.00	4
32	0.00	NA	NA	NA	21
35	0.00	NA	NA	NA	10
30	4.30	100.00	0.00	2200.00	23
27	5.90	50.00	50.00	800.00	17
25	9.10	100.00	0.00	1000.00	11

Cover Type	***UA (%)***	***PA (%)***	***EO(%)***	***EC (%)***	***No. Samp.***
3	18.50	62.50	37.50	275.00	27
33	22.60	11.86	88.14	40.68	31
31	25.00	25.00	75.00	75.00	4
7	28.00	35.00	65.00	90.00	25
23	28.10	23.68	76.32	60.53	32
38	28.60	11.11	88.89	27.78	7
4	29.40	27.78	72.22	66.67	17
34	32.60	31.91	68.09	65.96	46
45	33.30	20.00	80.00	40.00	6
15	35.70	58.82	41.18	105.88	28
17	40.00	66.67	33.33	100.00	10
26	40.00	85.71	14.29	128.57	15
40	41.40	66.67	33.33	94.44	29
37	43.80	29.17	70.83	37.50	16
41	44.00	57.89	42.11	73.68	25
19	45.00	35.71	64.29	42.86	11
1	48.10	52.00	48.00	56.00	27
39	50.00	5.26	94.74	5.26	2
11	51.70	27.78	72.22	25.93	29
36	53.60	34.88	65.12	30.23	28
16	59.30	57.14	42.86	39.29	27
52	77.80	58.33	41.67	16.67	9
5	100.00	50.00	50.00	0.00	1
50	100.00	100.00	0.00	0.00	4
9	NA	NA	NA	NA	0
8	NA	NA	NA	NA	0
6	NA	NA	NA	NA	0
51	NA	NA	NA	NA	0
49	NA	0.00	100.00	NA	0

Cover Type	***UA (%)***	***PA (%)***	***EO(%)***	***EC (%)***	***No. Samp.***
48	NA	NA	NA	NA	0
47	NA	NA	NA	NA	0
46	NA	NA	NA	NA	0
44	NA	NA	NA	NA	0
43	NA	NA	NA	NA	0
42	NA	NA	NA	NA	0
28	NA	0.00	100.00	0.00	0
21	NA	0.00	100.00	0.00	0
2	NA	0.00	100.00	NA	0

Landcover types with user's accuracy of 0% also had producer's accuracy of 0% and large errors of omission and commission (Table 4.1). Some of these types (12, 14, 18, 22 and 29) also had low sample sizes (N <= 5), meaning that accuracy estimates were based on little data. Four types (10, 13, 32 and 35) had accuracies based on larger samples (N = 13, 11, 15 and 10 respectively). All of these except type 11 were never assigned to video frames by the interpreters. Inspection of the contingency table (Appendix 2) for type 10 (juniper woodland) shows that 8 of the 14 samples were classified as pinon-juniper woodland. This suggests that either juniper was confused with pinon-juniper consistently on the Colorado landcover map, or that they were confused during the interpretation of the air videography. In either case, the error is not as serious as user's and producer's accuracy suggest at first glance, since confusion of juniper and pinon-juniper has only a small impact on the habitat relationship models used in Gap (seriousness of errors is considered in the fuzzy assessment discussed in Chapter 5). Cover type 13 (mixed conifer) was confused 6 of 11 times with the mixed forest type, again an understandable error with small impact on habitat modeling. Type 32, tall-grass prairie, was most frequently confused with mid-grass prairie, and type 35, sand dune complex (grass) was most frequently confused with mid- and shortgrass prairie types.

Landcover types 25, 27 and 30 had low (<10%) user's accuracy but high producer's accuracy and very high errors of commission, suggesting that these types appear on the map a much higher proportion of the time than they are found on the ground. Inspection of the contingency table (Appendix 2) shows that of the polygons chosen for assessment, 11 appear as type 25 (saltbush fans and flats) on the landcover map but only 1 was given this assignment in the video interpretation. Most of the polygons mapped as saltbush in the landcover map were assigned to the basin big sagebrush class in the videography, a potentially serious map error given the structural difference between the two communities. Type 27 (shrub sand dune complex) was rarely assigned to video frames (only 2 polygons) and was most frequently confused with grassland types. Type 30 (mixed tundra) was mostly confused with subalpine meadow and meadow tundra types.

Cover types with low user's accuracy (<= 50%), and a range of producer's accuracies require individual examination to untangle the nature of the errors associated with them. Two examples serve to highlight some common themes. Within this group some types have high user's accuracy relative to producer's accuracy, meaning that if they appear on the map they are likely to actually exist on the ground, but if you are standing in the type on the ground, it is less likely to appear on the map. An example, type 39 (graminoid/forb dominated wetland/riparian), has a user's accuracy of 50% and a producer's accuracy of 5.25%. This is interesting because of the importance of wetlands for biodiversity in the semi-arid western U.S., and because it highlights an issue with first-generation Gap landcover maps. Among all of the polygons sampled for the accuracy assessment of the Colorado landcover map, only 2 corresponded to graminoid/forb wetlands on the map, but 19 were assigned to this type in the video interpretation. Because riparian zones are distinctive, we assume that interpretation error is infrequent. Examination of the contingency table (Appendix 2) shows that polygons assigned to this type from video interpretation were frequently mapped as greasewood fans and flats, a cover type often *associated* with riparian areas. We believe that this confusion highlights a map resolution issue in the landcover map. Riparian zones are long, narrow features that by definition had to be 200m wide to be mapped in Colorado. The video, in contrast, is of higher resolution and can distinguish narrow riparian networks that were overlooked in the map. It was due to this problem that riparian areas were given special attention independently of the landcover maps during the Gap process.

The second example from this group is from types that have low user's accuracy relative to producer's accuracy. Type 3 (Douglas fir) is an example of this, with user's accuracy of 18.5% and producer's accuracy of 62.5% (Table 4.1). In contrast to the riparian example, Douglas fir on the landcover map rarely corresponds to Douglas fir on the ground, but Douglas fir on the ground is frequently Douglas fir on the map. Airborne videography, due to its high resolution, is particularly good at distinguishing conifer species based on canopy shape. In contrast, Landsat TM is poor at distinguishing conifer species because resolution is too low to distinguish structural characters of individual trees and spectral differences among conifer species are subtle. The contingency table (Appendix 2) again affords some clues. Most areas mapped as Douglas fir in Colorado were actually other conifer or mixed conifer cover types according to the video interpretation. The low user's accuracy suggests that there is confusion between conifer types on the map, while the relatively high producer's accuracy suggests that Douglas fir may be over-mapped in the land cover map, and therefore includes real areas of the type on the ground, as well as areas of other conifers. This is supported by low producer's accuracies for common conifer types such as lodgepole pine (type 4) which was frequently mapped as Douglas fir.

Two types (5 and 50) had 100% user's accuracy. As with the types at the other extreme of the accuracy range, these accuracies are drawn from very low sample sizes (N < 5) and consequently are of low confidence.

These examples emphasize that although map accuracy expressed by summary numbers (like user's or producer's accuracy) are frequently used, it is their combination with analysis of the off-diagonal elements in the contingency table itself that provide a more realistic picture of the strengths and weaknesses of the map being assessed or of the assessment itself.

4.3.3 Anderson-level Accuracy

At the coarser Anderson-level (Anderson et. al. 1976) of the classification hierarchy (Appendix 1) cover types are distinguished by physiognomy rather than dominant species. At this level we expect higher classification accuracy in both the landcover map and the video interpretation than for the Gap-level. As emphasized in section 4.3.2 above, we stress the importance of using the contingency table (Appendix 2) in combination with the indices presented here. User's and producer's accuracy, errors of omission and commission, and sample size are presented in table 4.2.

Table 4.2. User's (UA) and producer's (PA) accuracy (% correct), errors of omission (EO) and commission (EC) (%) and the number of sample polygons for each Colorado Anderson-level landcover type. "NA" in the producer's accuracy column represents cover types that were never assigned to video frames by the interpreter. "NA"’ in the user's accuracy column represents cover types from the map that were not sampled by video transects. Cover type codes are defined in Appendix 1.

Cover Type	***UA (%)***	***PA (%)***	***EO(%)***	***EC (%)***	***No. Samp.***
1	75.16	71.52	28.48	23.64	157
2	52.73	61.70	38.30	55.32	55
3	40.00	80.00	20.00	120.00	10
4	46.08	52.22	47.78	61.11	102
5	3.70	50.00	50.00	1300.00	27
6	25.00	20.00	80.00	60.00	4
7	65.33	56.65	43.35	30.06	150
8	28.57	11.11	88.89	27.78	7
9	50.00	5.26	94.74	5.26	2
10	46.30	69.44	30.56	80.56	54
11	33.33	20.00	80.00	40.00	6
12	0.00	0.00	100.00	0.00	0
13	100.00	100.00	0.00	0.00	3

Cover Type	***UA (%)***	***PA (%)***	***EO(%)***	***EC (%)***	***No. Samp.***
14	NA	NA	NA	NA	0
15	77.78	63.64	36.36	18.18	9
16	NA	NA	NA	NA	0

Contingency table analysis of the map at the Anderson classification level was conducted using the same procedures discussed for the Gap-level classification. User's accuracy spanned the range from 0% for type 12 (basin bare rock and soil) to 100% for type 13 (open water). Again, these extreme values were supported by few samples, although high accuracy in the open water type was expected because of its distinct appearance in TM imagery.

Analysis of individual types at this level should follow the same procedure discussed for Gap-types in section 4.3.2. Use of the contingency table (Appendix 2) to highlight the reasons for differences in user's and producer's accuracies reported in table 4.2 yields information not obvious when only summary numbers are examined. Three Anderson-level cover types deserve mention here and serve as examples at this level. First, type 1 (Needleleaf evergreen forest) showed high user's and producer's accuracy (Table 4.2), suggesting that this type was accurately represented in the Colorado land cover map. In fact, conifers are easily distinguished in satellite imagery from other physiognomic types. The contingency table shows that of the polygons mapped as evergreen forest, the relatively infrequent confusion that did occur was with deciduous forest, shrub and brush rangeland, and herbaceous rangeland. Deciduous forest is often mixed with evergreen forest in the field. The latter two types are quite different structurally from evergreen forest and would be considered serious errors from a habitat modeling standpoint.

Anderson landcover type 5 (shrub and brush tundra) has very low (3.7%) user's accuracy and relatively high producer's accuracy (50%), suggesting that areas of this type on the map rarely corresponded to areas interpreted as this type from the videography. Inspection of the contingency table (Appendix 2) reveals that only 2 polygons were assigned shrub and brush tundra by the video interpreters. Other map polygons of this type were most commonly assigned to Anderson types 7, 8 and 11 (herbaceous rangeland, herbaceous tundra and alpine unvegetated, respectively) using the videography. The confusion between alpine types is not of great concern, but confusion with herbaceous rangeland is more serious. It is probable that this inaccuracy is a function of video misinterpretation as well as error in the landcover map. Herbaceous types are virtually impossible to distinguish from one another on video, and the difference between alpine and basin types is largely dependent on knowledge of context. It is likely that map confusion between herbaceous and shrub-dominated alpine types is compounded by the difficulty of distinguishing alpine from basin shrubs in the videography if contextual information is not present. Obviously, the context was clearly available during production of the original map suggesting that these types of errors lie entirely with the video determinations.

Cover types with intermediate accuracy, such as Anderson type 2 (deciduous forest), are easily distinguished in air videography and less distinct in satellite imagery, due to confusion with evergreen forest, with which it is often mixed. This is reflected in the contingency user's and producer's accuracies for this type (Table 4.2) and in the contingency table (Appendix 2), where most of the confusion is with evergreen forest.

4.4 Secondary Class Labels and Map Accuracy

Secondary class labels are one way to increase the information content of landcover maps (Woodcock et al. 1996). Polygon maps, like the Wyoming and Colorado Gap landcover maps, are particularly suited for secondary labels because large MMU polygons almost always contain inclusions of types beside the dominant cover. Secondary class labels were assigned during the original mapping based on relative area in each polygon. Primary (dominant) landcover is the cover occupying the largest area within each polygon and secondary cover is the type occupying the next largest area. Most polygons received secondary class labels. According to Woodcock et al. (1996), "One of the key issues confronting the use of secondary labels will be whether or not they can be provided with sufficient accuracy to be useful."

We did not perform a formal accuracy assessment of secondary attributes in the Colorado Gap landcover map because of insufficient data, but instead explored the confusion between primary and secondary labels where they were available. To do this, we tabulated the proportion of secondary label matches with reference data when the primary label did not match. In other words, if the primary label was "wrong", was the secondary label "right"? These proportions, expressed as percentages (Table 4.3), provide insight into map confusion and the value of secondary cover type designations. This analysis is at the Gap-level only.

Table 4.3. The proportion (%) of secondary class labels that match the video reference data when the primary label does not match for the Colorado landcover map. User's accuracy is from Table 4.1. The proportion of primary mismatch with secondary matches is the percentage of the time that the primary label did not match the video reference data but the secondary label did, indicating confusion between dominant and subdominant land cover types.

Cover Type	***User's Accuracy (%)***	***Proportion of Primary Mismatch w/ Secondary Match (%)***
1	48.10	23.10
2	NA	NA

Cover Type	***User's Accuracy (%)***	***Proportion of Primary Mismatch w/ Secondary Match (%)***
3	18.50	33.30
4	29.40	55.60
5	100.00	NA
6	NA	NA
7	28.00	6.70
8	NA	NA
9	NA	NA
10	0.00	10.00
11	51.70	22.20
12	0.00	50.00
13	0.00	18.20
14	0.00	0.00
15	35.70	20.00
16	59.30	11.10
17	40.00	25.00
18	0.00	20.00
19	45.00	50.00
20	0.00	50.00
21	NA	NA
22	0.00	0.00
23	28.10	35.30
24	0.00	33.30
25	9.10	50.00
26	40.00	62.50
27	5.90	33.30
28	NA	NA
29	0.00	0.00

Cover Type	***User's Accuracy (%)***	***Proportion of Primary Mismatch w/ Secondary Match (%)***
30	4.30	0.00
31	25.00	0.00
32	0.00	31.60
33	22.60	35.00
34	32.60	25.00
35	0.00	0.00
36	53.60	12.50
37	43.80	33.30
38	28.60	20.00
39	50.00	0.00
40	41.40	10.00
41	44.00	0.00
42	NA	NA
43	NA	NA
44	NA	NA
45	33.30	25.00
46	NA	NA
47	NA	NA
48	NA	NA
49	NA	NA
50	100.00	NA
51	NA	NA
52	77.80	0.00

Comparison of the secondary labels to reference data reveals that for many of the Colorado cover types there was some confusion between primary and secondary mapped landcover (Table 4.3). For example, polygons labeled as greasewood fans and flats (type 26) had user's accuracy of 40%, but of the 60% that were mislabeled (from a user's perspective), 62.5% would have been correct had the secondary type been assigned to the primary attribute. This suggests that greasewood polygons in the Colorado landcover map frequently contained mixtures of landcover types whose relative proportions were difficult to estimate using the TM imagery, a scenario that is supported by the examination of the contingency table discussed in section 4.3.2 above. In contrast, for aspen forest (type 16), with 59.3% user's accuracy, only 11.1% of the mislabeled polygons would have been corrected by assigning the mapped secondary cover type to the primary polygon attribute. For aspen polygons, confusion between dominant and subdominant landcover does not appear to be a serious concern, perhaps because this cover type was more easily distinguished in TM imagery than greasewood.

Examination of secondary labels gives users information about types of confusion present in the Colorado landcover map and should be used in conjunction with the other accuracy measures described in this report. Note that for this assessment we are concerned with the accuracy of the primary cover type labels (dominant land cover in each polygon) and did not assess the accuracy of secondary class labels for secondary (subdominant) landcover.

CHAPTER 5

Fuzzy Accuracy Assessment

5.1 Introduction

Fuzzy set theory (Zadeh 1965) recognizes that the human tradition of classifying objects or ideas into discrete or "crisp" classes (e.g., sagebrush vs. grass; correct vs. incorrect), is not always realistic. The assumption of classical (crisp) set theory for traditional accuracy assessment (Chapter 4) is that classified map units are either completely and unambiguously correct or equally incorrect (Woodcock and Gopal 2000). Fuzzy set theory, in contrast, uses degrees of set membership, and in the context of accuracy assessment, *levels* of right and wrong are assigned to map units based on their comparison to reference data.

Fuzzy accuracy assessment can follow two approaches with regard to thematic landcover maps. The first recognizes that landcover is a continuum with "fuzzy" regions between mapped classes that have characteristics of two or more cover types. This approach is site-specific in the sense that degrees of error are assigned based on the landcover characteristics of each reference sample location. For example, a map unit classified as either sagebrush or grassland and corresponding to a reference site with characteristics of both sagebrush and grassland would be considered partially correct for either label. This approach was developed by Gopal and Woodcock (1994). A second approach assigns degrees of right and wrong to map units based on the seriousness of the error with regard to a specific map application. In the case of the Gap Analysis, the landcover maps are used as a spatial variable in animal habitat relation models and the degree of error depends on the effect of landcover misclassifications on habitat prediction (Scott et al. 1991). This second approach is used for the assessment of the Colorado landcover map.

Gopal and Woodcock (1994) introduced six fuzzy operators designed to highlight different aspects of map accuracy (see Section 1.6). They applied these operators using the site-specific approach discussed above. For the assessment of the Gap landcover maps, we modified their methods to implement the second approach and assigned codes representing a verbal "correctness" scale (Table 5.1) (Gopal and Woodcock 1994) by judging the seriousness of mismatches between map and reference data in the context of the habitat models. In this way the six fuzzy operators describe aspects of map accuracy with regard to the Gap application. We emphasize this because map users interested in different applications may assign the verbal correctness scale differently, thus producing a different outcome. Map users are urged to study the fuzzy matrices (Appendix 3) to understand the assignment of the correctness scale (Table 5.1).

Table 5.1. The verbal "correctness" scale and associated codes used for fuzzy accuracy assessment based on the work of Gopal and Woodcock (1994).

Code	Description
5	Absolutely right
4	Good answer
3	Reasonable or acceptable answer
2	Understandable, but wrong
1	Absolutely wrong

5.2 Creation of Fuzzy Matrices

Matrices of correctness codes from table 5.1 were assigned to each combination of reference and map cover types by considering their effect on animal habitat modeling. These "fuzzy matrices" for both the Gap- and Anderson-level classifications in Colorado are included as Appendix 3. Fuzzy matrices were reviewed by the habitat modelers at CDOW who confirmed the correctness codes.

At the Gap-level, we use Douglas fir as an example for the assignment of correctness codes. If a polygon from the Colorado landcover map labeled as type 3 (Douglas fir) is Douglas fir-dominated according to the video interpretation, the polygon receives a confidence code of 5 (absolutely right) in the fuzzy matrix (Appendix 3). Thus, elements on the diagonal of the fuzzy matrix are perfect matches between map and reference interpretations. But if the same Douglas fir polygon from the map should have been mapped as lodgepole pine (type 4) according to the reference data, it receives a correctness code of 4 (good answer), indicating that for habitat modeling, confusion between Douglas fir and lodgepole pine is not serious, since many animals can use the two cover types interchangeably. However, if the Douglas fir polygon should have been mapped as greasewood (type 26) the polygon would receive a 1 (absolutely wrong) since few if any animals use Douglas fir and greasewood interchangeably for habitat.

The Anderson-level fuzzy matrix (Appendix 3) was constructed using the same philosophy as used to build the Gap-level matrix. Anderson-level mismatches are generally more serious errors for habitat modeling because the physiognomically-based types are often less equivalent for animal use than some Gap-level mismatches. The Anderson-level classification was not actually used during Gap habitat modeling but the fuzzy analysis highlights aspects of map accuracy at this level that are instructive.

5.3 Fuzzy Accuracy Statistics

The fuzzy matrices (Appendix 3) that assigned codes from the verbal scale (Table 5.1) to all possible combinations of map and reference polygons were used to calculate the six fuzzy operators (Gopal and Woodcock 1994) for both the Gap- and Anderson-level classifications. To accomplish this, contingency tables (Appendix 2) and fuzzy matrices (Appendix 3) were combined to compute tables and matrices describing the results of MAX, RIGHT, DIFFERENCE, MEMBERSHIP, AMBIGUITY and CONFUSION (Chapter 1 and below). We do not describe the programming details used to produce these results, but instead discuss calculations for each operator in general terms for each operator, while providing examples. In short, the analyses were done by converting the fuzzy matrices and contingency tables to arc/info GRIDS (matrices in arc/info format), which were manipulated using programs written in the arc macro language (aml) to produce the output tables and matrices published here. AML programs are carefully documented and are easily generalized for use by other projects. Code is available upon request from the Department of Botany at UW (Driese).

5.4 Gap-level Fuzzy Accuracy

5.4.1 The MAX Operator

The MAX operator records a match only for perfect correspondence between map and reference and therefore is identical to user's accuracy calculated during the traditional accuracy assessment (Chapter 4). Consequently, we will not discuss MAX here, but instead refer to user's accuracy from the traditional assessment (Tables 4.1, 4.2).

5.4.2 The RIGHT Operator

The RIGHT operator (Gopal and Woodcock 1994) is perhaps the most intuitive of the fuzzy operators that we used. RIGHT counts mapped polygons as correct for all combinations of map and reference corresponding to a correctness score of 3 (reasonable or acceptable answer) (Table 5.1, Appendix 3) or better. The proportion of matches by this criterion is the measure of accuracy of the mapped cover types when errors considered acceptable for habitat modeling are counted as correct, even if they are not perfect answers. The RIGHT operator considers the off-diagonal elements from the traditional confusion matrix and counts them as matches if they are not serious errors from the point of view of animal habitat modeling.

Because it uses a less stringent criterion for counting mapped polygons correct, the RIGHT operator results in higher accuracies for individual cover types (Table 5.2) and for the map overall than does traditional user's accuracy (Table 4.1). Overall map accuracy for the Colorado landcover map using the RIGHT criterion is 75.7% at the Gap-level. This accuracy, however, reflects information about the frequency of errors with regard to animal habitat modeling that is not expressed in user's accuracy (Gopal and Woodcock 1994). Comparison of the results of the RIGHT operator for individual types with traditional user's accuracy is instructive in this regard, and we supply two examples to illustrate.

Mixed tundra (type 30) has very low (4.3%) user's accuracy (Table 4.1) but much higher accuracy by the RIGHT criterion (69.6%). This suggests that although mixed tundra polygons on the Colorado landcover map rarely correspond to mixed tundra on the ground, they often correspond to cover types that are considered equivalent to mixed tundra from the point of view of the animal habitat models. This is confirmed by examination of the contingency table (Appendix 2) for the Gap level classification which shows that most confusion for this type is with either subalpine meadow (type 37) or meadow tundra (type 38), both of which are considered "acceptable" answers (Appendix 3). Some of the confusion, though, is with alpine exposed rock (type 45), which is not considered an acceptable alternative for mixed tundra, and consequently lowers the RIGHT accuracy.

In contrast to mixed tundra, type 25 (saltbush fans and flats) shows no change in accuracy (9.1%) for the RIGHT operator compared to user's accuracy (Tables 4.1, 5.2), suggesting that map errors for saltbush are frequent and have serious consequences for animal habitat modeling. Again, examination of the contingency table is instructive. For saltbush, most of the confusion is with basin big sagebrush (type 23), which corresponds to a correctness rating of 2 (understandable but wrong) from table 5.1. Because basin big sagebrush is probably not good habitat for animals who prefer saltbush flats, this is considered a serious error and the RIGHT operator reflects this.

Similar analyses can be done for other cover types of particular interest to map users and we again urge that map accuracy indices (like the RIGHT operator) be considered in combination with other measures and with the contingency table. Generally speaking, cover types that have high accuracies (> 75%) by the RIGHT criterion, such as evergreen forest types, are adequately represented on the landcover map for the purpose of animal habitat mapping, even when the Gap-level cover type designation is not perfect. Types with low RIGHT scores, such as many of the basin shrub types may be less reliable.

Table 5.2. Results of the RIGHT fuzzy operator. Percent matches are the proportion of map polygons for which the correspondence with the video reference data is considered "reasonable or acceptable" (Table 5.1, Appendix 3) for animal habitat models.

Gap Cover Type	*Percent Match*	*Gap Cover Type*	*Percent Match*	*Gap Cover Type*	*Percent Match*
1	92.59	19	63.64	37	93.75
2	NA	20	0.00	38	85.71
3	92.59	21	NA	39	50.00
4	82.35	22	100.00	40	82.76
5	100.00	23	68.75	41	72.00
6	NA	24	66.67	42	NA
7	84.00	25	9.10	43	NA
8	NA	26	53.33	44	NA
9	NA	27	94.12	45	50.00
10	85.71	28	NA	46	NA
11	72.41	29	25.00	47	NA
12	20.00	30	69.57	48	NA
13	100.00	31	75.00	49	NA

Gap Cover Type	*Percent Match*	*Gap Cover Type*	*Percent Match*	*Gap Cover Type*	*Percent Match*
14	50.00	32	80.95	50	100.00
15	75.00	33	74.19	51	NA
16	88.89	34	71.74	52	77.78
17	60.00	35	90.00		
18	40.00	36	75.00		

5.4.3 The DIFFERENCE Operator

The DIFFERENCE operator describes the *magnitude* of map errors (Gopal and Woodcock 1994). It is the difference between the correctness code for the cover type that was actually mapped and the code for the cover type that is the best answer (correctness code 5 from table 5.1)for a sample polygon from among all possible cover types. For example, if a mapped polygon on the Colorado landcover map matches the video interpretation perfectly, it receives a correctness code of 5 (absolutely right). If the highest correctness code for all alternative cover types that could have been mapped for that polygon was 3, DIFFERENCE would equal five minus three, or positive two. Positive values of DIFFERENCE indicate that the mapped type was a better answer than all other possibilities. This only occurs when a mapped polygon matches the video interpretation perfectly because otherwise some other answer would receive a perfect score (5) and result in a negative DIFFERENCE. Negative DIFFERENCE indicates that the mapped cover type and the video interpretation did not match perfectly, and the magnitude of that negative number is a measure of the seriousness of the error. We use the mean of DIFFERENCE for each Colorado cover type to indicate the relative magnitude of errors for that type (Table 5.3).

Table 5.3. Results of the DIFFERENCE operator for the Colorado landcover map. Mean difference is the average difference between the correctness code for the mapped polygon and that of a perfect match between map and reference. "NA" indicates that no data were available for that type.

Gap Cover Type	*Mean Difference*	*Gap Cover Type*	*Mean Difference*	*Gap Cover Type*	*Mean Difference*
1	-0.44	19	-0.11	37	-0.40
2	NA	20	-3.00	38	-0.71
3	-1.22	21	NA	39	-1.00
4	-1.24	22	-1.00	40	-0.86
5	1.00	23	-1.59	41	-1.12
6	NA	24	-2.50	42	NA
7	-1.08	25	-2.91	43	NA
8	NA	26	-1.67	44	NA
9	NA	27	-1.94	45	-1.20
10	-1.57	28	NA	46	NA
11	-0.79	29	-2.75	47	NA

Gap Cover Type	*Mean Difference*	*Gap Cover Type*	*Mean Difference*	*Gap Cover Type*	*Mean Difference*
12	-3.00	30	-0.91	48	NA
13	-1.27	31	-0.33	49	NA
14	-2.50	32	-1.21	50	2.00
15	-1.11	33	-1.10	51	NA
16	-0.08	34	-0.95	52	1.38
17	-1.00	35	-1.20		
18	-2.60	36	-0.76		

Positive mean DIFFERENCE occurs for open water (type 50) (2.0) and urban (52) (1.38) types only, indicating that most mapped polygons of these types match the video interpretation and when they don't errors are not serious.

Examples of types with small negative mean DIFFERENCE include aspen forest (type 16) (-0.08), spruce-fir forest (1) (-0.44) and xeric upland shrub (19) (-0.11). Mixed tundra, discussed in Section 5.4.2, has a difference of -0.91 (Table 5.3), confirming that most confusion for this type is of low seriousness for animal habitat models.

Types with more serious errors include bitterbrush shrub step (type 20) (-3.0) mesic upland shrub (18) (-2.6), desert shrub (24) (-2.5) and saltbush fans and flats (25) (-2.91). Shrub types are difficult to distinguish both on Landsat imagery and in the video and based on the DIFFERENCE operator, errors in classification can be substantial in terms of animal habitat considerations. The DIFFERENCE operator assumes that all error is in the map and does not account for possible video misinterpretation.

5.4.4 The MEMBERSHIP Operator

The MEMBERSHIP operator (Table 5.4) highlights *sources* of map error which are primarily of interest to map makers (producers) because they give information about the circumstances in which misclassifications were made (Gopal and Woodcock 1994). For each landcover type, MEMBERSHIP counts the number of alternative classes exceeding a user-defined correctness threshold, (we used "3", a reasonable or acceptable answer), that could have been mapped instead of the type occurring on the map. By considering MEMBERSHIP with user's accuracy and the RIGHT operator, one gains insight into the situations likely to cause map errors.

MEMBERSHIP is most appropriate for Gopal and Woodcock's (1994) site-specific approach (Section 5.1) to fuzzy assessment because it recognizes heterogeneous sites (e.g. sites in transition zones) as high membership sites and rates them accordingly. For their approach, if high MEMBERSHIP sites have low accuracy, and low MEMBERSHIP sites have high accuracy, it suggests that map errors are occurring primarily in heterogeneous areas. When the verbal correctness scale is assigned based only on the seriousness of errors as we have done, errors occurring more frequently in high MEMBERSHIP types than in low MEMBERSHIP types suggest that confusion is among equivalent habitat, and that other classification approaches may be necessary to separate those habitats during mapping. Our approach to MEMBERSHIP is less directly based on the spectral and physical characteristics of the sites, and therefore more removed from the mapping process itself.

Table 5.4. Results of the MEMBERSHIP operator for the Gap-level types. MEMBERSHIP is the number of alternative classes for each cover type that receive an ‘"acceptable or reasonable"’or better score (correctness code ≥ 3).

Gap Cover Type	*Membership*	*Gap Cover Type*	*Membership*	*Gap Cover Type*	*Membership*
1	14	19	15	37	19
2	11	20	7	38	6
3	17	21	12	39	8
4	15	22	15	40	9
5	9	23	14	41	9
6	14	24	14	42	5
7	15	25	8	43	7
8	14	26	14	44	6
9	13	27	14	45	9
10	8	28	16	46	19
11	11	29	5	47	19
12	13	30	5	48	8
13	18	31	15	49	8
14	17	32	6	50	4
15	9	33	17	51	9
16	14	34	15	52	5
17	16	35	8		
18	11	36	15		

Douglas fir forest (type 3) has high MEMBERSHIP (Table 5.4), low user's accuracy (Table 4.1) and high RIGHT accuracy (Table 5.2), indicating that this type is probably mis-mapped frequently at the Gap-level because there are many similar alternatives (other evergreen forest types). We expect other evergreen forest types to show a similar pattern. In fact, spruce-fir (type 1), lodgepole pine (4) and ponderosa pine (7) forests, representing much of the evergreen forest in Colorado, all have high MEMBERSHIP (Table 5.4), low user's accuracy (Table 4.1) and high RIGHT accuracy (Table 5.2). Map producers might consider separating spectrally similar members of this physiognomic type by introducing a modeling component.

For types with low MEMBERSHIP, if user's accuracy is low and RIGHT is considerably higher, as is the case for shrub tundra (type 29), mixed tundra (30) and meadow tundra (38) we surmise that there is significant confusion among small groups of similar types on the map. In Colorado, these examples suggest that tundra types are unique from other types with similar physiognomy, but not well distinguished from one another, perhaps because they grow in the same environment (high alpine areas above treeline). Map makers might consider lumping tundra types or finding alternative sensors or modeling approaches if separation of tundra varieties is a goal.

5.4.5 The Confusion and Ambiguity Operators

The CONFUSION and AMBIGUITY operators provide information about the *nature* of errors--the types of confusion that occurs in the map being assessed (Gopal and Woodcock 1994). These two operators are another way of examining the off-diagonal elements in the contingency table as we did for the traditional assessment, but from a "fuzzy" perspective. Gopal and Woodcock based these operators on their site-based approach to assignment of correctness codes. Because we modified this approach to address the seriousness of mismatches from a habitat modeling perspective, we also modified the formulation of these operators. The CONFUSION and AMBIGUITY operators are perhaps less informative from this perspective, but practical considerations prevented their calculation using the Gopal and Woodcock approach.

For our analysis, CONFUSION is the number of alternative landcover choices for each occupied cell in the contingency table (Appendix 2) that would have been a *better* answer than the label that was given to the polygons represented by that cell. AMBIGUITY is the number of alternative landcover choices with the *same* correctness score as the polygon label given. The results of CONFUSION and AMBIGUITY are matrices (Appendix 4) of the same dimensions as the original contingency tables. Cells in these matrices are non-zero only if the corresponding cells in the traditional contingency matrix (Appendix 2) was occupied, indicating that the corresponding combination of map and reference data actually occurred in Colorado. To understand these results requires consideration of both the traditional contingency table to determine the number of occurrences of each map-reference combination, and the CONFUSION and AMBIGUITY matrices to determine the number of better or equal alternative answers for that combination. An example from the Colorado map demonstrates analysis of these operators.

Examination of the traditional contingency table (Appendix 2) for the Gap-level classification in Colorado shows that lodgepole pine forest (type 4) was mis-mapped as basin big sagebrush for 1 sample polygon. The corresponding cell in the CONFUSION matrix (Appendix 4) shows that for this combination of map and reference data, there were 22 better labels for the mis-mapped polygon than the one given. The AMBIGUITY matrix shows that 29 polygon labels were equivalent. On the other hand, no basin big sagebrush polygons (from the reference data) were ever mapped as lodgepole pine forest. So, in the Colorado landcover map, lodgepole pine was sometimes mapped as basin big sagebrush even though there were 22 other choices that would have been better and 29 other choices that would have been equivalent from an animal habitat mapping standpoint but basin big sagebrush was rarely if ever mapped as lodgepole pine. One explanation for this is that sparse lodgepole pine with a sagebrush understory may be mapped as sagebrush even when lodgepole is dense enough to fit the definition of lodgepole pine forest. The resolution of Landsat TM causes the lodgepole pine signal to be overwhelmed by the understory. Map users might expect relatively sparse lodgepole forest to be mis-mapped as understory types. Users would also note that these errors may have serious consequences for habitat models, since there were many better or equivalent answers than the one applied on the map.

Similar analyses, though complicated, should be performed by map users interested in particular cover types and their confusion on the Colorado landcover map. These analyses help point out situations (like the failure in transitional areas between forest and shrubland) where confusion might be found on the map.

5.5 Anderson-Level Fuzzy Accuracy

Fuzzy operators RIGHT, DIFFERENCE, MEMBERSHIP, CONFUSION and AMBIGUITY were calculated for the Anderson-level classification for Colorado using the same procedures described for the Gap-level classification above. Calculations are based on a fuzzy matrix (Appendix 3) in which codes from the verbal scale (Table 5.1) are assigned to each possible combination of map polygon and video interpretation. Because the Anderson-level types are less similar to one another in general than many of the Gap level types, confusion between mapped and reference types is generally considered more serious here.

5.5.1 The RIGHT Operator

Results of the RIGHT operator (Table 5.5) show that some polygons mapped at the Anderson-level have higher accuracy using the RIGHT criterion than user's accuracy from traditional assessment (Table 4.2). Accuracy of herbaceous wetlands (type 9), for example, improved from 50% user's accuracy to 100% RIGHT accuracy because all types mapped as herbaceous wetland were in fact one of the wetland types according to the video reference data. Other types, like alpine unvegetated (type 11) showed no improvement. Overall map accuracy using the RIGHT criterion at the Anderson-level was 78.84%.

Table 5.5. Results of the RIGHT operator for Anderson-level classes. The percent match is the proportion of polygons for which the Anderson-level designation matched the video interpretation at a verbal correctness (Table 5.1) corresponding to "reasonable or acceptable" (3)’or better.

Cover Type	*Percent Match*	*Cover Type*	*Percent Match*
1	82.80	9	100.00
2	72.73	10	83.33
3	40.00	11	33.33
4	77.45	12	NA

Cover Type	*Percent Match*	*Cover Type*	*Percent Match*
5	66.67	13	100.00
6	50.00	14	NA
7	82.67	15	77.78
8	85.71	16	NA

5.5.2 The DIFFERENCE Operator

The DIFFERENCE operator for the Anderson level classification is presented in Table 5.6. The largest negative mean differences occur in types 5, 11, 3 and 6, indicating that these types are mapped with the highest magnitude of error. Shrub and brush tundra (type 5) is frequently confused with herbaceous rangeland and with herbaceous wetland/riparian. This is most likely a result of misinterpretation of the video, rather than of map errors because the map makers always knew the elevational context in which they were working. Elevation data for sample sites was also available to the video interpreters, but may not have been used in every case. Alpine tundra resembles rangeland in video and can easily be misinterpreted if the alpine context is not clearly visible in the wide angle video or if elevational information is not used. Alpine unvegetated (11) may experience the same problem. Forested wetlands (type 3) are frequently confused with shrub rangelands, shrub wetlands and herbaceous rangeland, all of which are serious errors for animal habitat modeling. This confusion may occur because of mixtures of different cover types that occur in areas mapped as riparian. The resolution of the landcover map is coarse enough that riparian polygons include significant areas occupied by uplands that border the riparian zone proper. The same may be said for shrub and brush wetlands (type 6). Careful analysis of the contingency table, the matrix of correctness scores, and mean DIFFERENCE can help pull apart the reasons for these map (or video interpretation) errors.

Table 5.6. Mean DIFFERENCE for the Anderson level classes. This is the mean difference between the correctness rating (Table 5.1) for a landcover type and the best rating (5) for that type.

Cover Type	*Mean Difference*	*Cover Type*	*Mean Difference*
1	-0.06	9	0.00
2	-0.96	10	-0.33
3	-1.90	11	-2.00
4	-0.58	12	NA
5	-2.15	13	2.00
6	-1.75	14	NA
7	0.29	15	0.78
8	-1.14	16	NA

5.5.3 The MEMBERSHIP Operator

The results of the MEMBERSHIP operator (Table 5.7) show that all of the Anderson cover types have several acceptable alternatives and that there is not a strong relationship with user's accuracy (Table 4.2) or RIGHT percent matches (Table 5.5). Needleleaf evergreen forest (Anderson-level type 1) has user's accuracy of 75.2%, RIGHT accuracy of 82.8% and 3 alternative classes that would be reasonable map alternatives for habitat modeling. Little improvement of RIGHT accuracy over user's accuracy for conifers suggests that confusion that does occur is not among types that are equivalent for habitat modeling. Herbaceous rangeland (type 7), which has the most acceptable alternatives (7) among the Anderson types, has user's accuracy of 65.3% (Table 4.2) and RIGHT accuracy of 82.7%, suggesting that confusion for this type is likely to be with types that are acceptable alternatives.

Table 5.7. Results of the MEMBERSHIP operator for the Anderson-level classes in Colorado. MEMBERSHIP is the number of alternative classes to the mapped class that are "reasonable or acceptable"’or better (correctness code 3 from Table 5.1) answers for that class.

Cover Type	*Member-ship*	*Cover Type*	*Member-ship*
1	3	9	6
2	3	10	3
3	6	11	5
4	4	12	5
5	5	13	4
6	3	14	5
7	7	15	2
8	4	16	3

5.5.4 CONFUSION and AMBIGUITY Operators

CONFUSION and AMBIGUITY for the Anderson-level classes (Appendix 4) highlight off-diagonal contingency table elements as discussed for the Gap-level types. Using analogous examples to the Gap-level discussion (Section 5.4.5), at the Anderson-level, evergreen forests (type 1) are mis-mapped as shrub and brush rangeland at 10 sample sites though there were 7 better alternative answers and 8 equivalent answers. At the Anderson-level, shrub and brush rangeland was confused with lodgepole pine at 12 sample sites though again there were 7 better and 8 equivalent alternatives. At the Gap-level, it appeared that lodgepole pine forest was sometimes being overwhelmed on the Colorado landcover map by understory shrubs when pine was sparse. At the Anderson-level, this confusion is probably more broadly related to misclassification of transitional zones between shrublands and evergreen forests, regardless of species. User's should note though, that this confusion is not common, with evergreens being confused as shrub rangeland for only 7.6% of all evergreen sample polygons. The Anderson-level reinforces the observation that sparse conifers may be confused with understory species in Colorado. The CONFUSION and AMBIGUITY operators suggest that this confusion is independent of the similarity of the confused cover types for habitat modeling.

5.6 Summary of Fuzzy Assessment

In summary, fuzzy assessment provides new tools for examining different aspects of map accuracy while considering specific map applications. For the Colorado Gap landcover map, we conducted fuzzy accuracy assessment by recognizing that some map errors have more serious consequences for animal habitat models than others. By quantifying this "seriousness" using a coded verbal scale, we were able to measure several aspects of accuracy by modifying the operators of Gopal and Woodcock (1994).

In general, fuzzy assessment increases the accuracy of the landcover map for the Gap-level cover types by loosening the criteria for counting a mapped polygon as correct. Anderson-level cover types are less affected because confusion among them is less frequent on the map than at the Gap-level in the first place, and when Anderson-level types *are* confused, the confusion is likely to be serious from a habitat modeling perspective.

CHAPTER 6

Evaluation of Air Videography and Limitations of Accuracy Assessment

6.1 Introduction

The assumption of accurate reference data is central to accuracy assessment of thematic maps. For the assessment of the Colorado landcover map, reference data are polygons whose dominant landcover was estimated by interpreting videography frames. Because airborne videography is a form of remote sensing, and because interpretation of remotely sensed data is rarely without error, the assumption of accurate reference data has certainly been violated to some degree for the assessment described in this report. For this reason, map users should use the accuracy measures described here cautiously as guidelines for deciding if the Colorado landcover map is appropriate for their particular application.

Recognizing that interpretation of videography is imperfect, we devote this chapter to 1) identifying cover types for which videography is well suited, and those for which it is not, 2) examining the relationship between interpreter confidence in video interpretation and map accuracy, 3) discussing the potential effect of mis-interpretation of reference data on the accuracy measures described previously and 4) evaluating airborne videography as a tool for the assessment of statewide landcover maps in the Rocky Mountain region (and as a source of training data for classification). These evaluations are based on interpreter experience and on informal studies conducted during the assessment.

It should also be mentioned that advances in airborne videography equipment and methodology since the Colorado and Wyoming assessments were begun may improve the utility of this tool for identifying some cover types. Digital videography and manipulation of the resulting data to permit stereo viewing of image mosaics may enhance the ability of interpreters to discriminate between some cover types. A digital system has been used with success in the northeastern U.S. and in the tropics by Slaymaker (personal communication) but remains untested in the more arid shrublands Colorado and Wyoming at the time of this writing.

6.2 Evaluation of Airborne Videography for Colorado Gap-level Landcover Types

The utility of airborne videography for accuracy assessment and collection of training data depends on the interpretability of Gap-level classes in the video. We discuss interpretability here based on the experience of the interpreters who analyzed the Colorado video (Owens and Driese) and on an informal test using ground-truthed examples of Colorado cover types collected in Wyoming for the construction of the interpretation key for that state. Notes describing key features used for the interpretation of specific Gap-level types in Colorado are included as Appendix 5, as a reference for others doing interpretation in similar environments.

Videography worked well for the Gap-level evergreen forest types found in Colorado because its resolution allowed the interpreter to compare the distinctive canopy shapes of individual trees species (Appendix 5). Spruce-fir and Douglas fir forest were occasionally confused because both had pointed crowns, but this confusion was not a serious problem since these trees generally grow at different elevations. Lodgepole pine was distinguished from other conifer species by its even canopy with a yellowish-green color in the video and small, rounded crowns. Ponderosa pine had larger rounded crowns whose aggregate canopy was less even than that of lodgepole pine. Ponderosa frequently occurred with Douglas fir in Colorado but the two species were distinguishable because of their crown shapes. Blue spruce, white fir and Rocky mountain bristlecone pine were uncommon and not identified in the Colorado video. We suspect that distinguishing blue spruce and white fir from Engelman spruce and subalpine fir forests would be difficult. Mixed forest types were distinguished by combinations of evergreen species described above or combinations of evergreen and deciduous tree species. No pure juniper woodlands were assigned to video frames, suggesting that there was a problem distinguishing pinon pine from juniper, since both juniper woodland and pinon-juniper occur commonly in Colorado. In summary, video works well for coniferous tree species having distinctive crown shapes but does not work well for those separable only on the basis of color or texture.

Gambel's oak and aspen were the only deciduous tree species comprising non-riparian forest types in Colorado and these were easily distinguished from one another in video by height (Gambel's oak forms short dense thickets and aspens are taller and more open) and canopy shape (Appendix 5). Also, because some of the video was flown in the fall, aspens were distinguished by yellow leaves. Oaks retain their dead leaves into the winter after aspen trees are bare. Video is well-suited for distinguishing these species.

Shrub species were nearly impossible to separate in videography and will probably require other tools for collection of reference data. Some Gap-level shrubland types are difficult to distinguish even on the ground, and the problem becomes greater when remotely sensed data are used. Mountain big sagebrush and Wyoming big sagebrush are important examples of this problem. Basin big sagebrush is taller and more easily distinguished, but its identification is still difficult. Bitterbrush was never assigned to video frames during interpretation, though it probably occurred in some of the frames. Bitterbrush tends to be mixed with other shrubs and is not distinctive in video. Desert shrub species have some distinguishing canopy shapes that help in their interpretation, but still were problematic. Some desert shrub types are mixtures of species that are difficult to distinguish on video. Other single-species types, like saltbush fans and flats cannot be separated from sparse sagebrush steppe. Shrub-dominated sand dune vegetation was often confused with non-dune shrub types unless the underlying dunes were obvious in the wide angle video frames. Dunes are, in fact, not always evident in video or even on the ground when they are fully vegetated. Otherwise these types resemble non-dune shrubs on low hills.

Colorado grassland types, like shrublands, were nearly impossible to distinguish from one another using videography except where environmental context was evident from elevation or the wide-angle video. Examples of the latter were foothills grasslands, subalpine meadows, and meadow tundra, all of which could sometimes be distinguished based on their positions relative to upper or lower treeline. Plains grassland types were essentially indistinguishable from one another in video, and probably will require ground-based sampling to acquire reference data confidently. Tall-, mid- and shortgrass prairie types are nearly identical in the video.

Riparian zones were distinguished in the Colorado Gap-level classification scheme by the physiognomy of their dominant vegetation. Airborne videography is an excellent tool at this level and its high resolution works well for distinguishing riparian types. In fact, videography resolved riparian zones that were too narrow to be treated consistently in the Colorado landcover map, where they were often lumped with surrounding upland landcover.

Alpine tundra is distinguishable from corresponding physiognomic types in basins by context, typically gathered from the wide angle video frames and with ancillary information on elevation. Where context was not clear, confusion between alpine and basin types occurred which almost certainly represented misinterpretation of video, since context was always available during the original mapping from Landsat data. Discriminating among tundra types depended on separating physiognomic types (herbs vs. shrubs), a task for which videography is well-suited. Problems due to steep slopes at various aspects caused problems with deep shadows in which plants composing the landcover were obscured.

Human-influenced cover types, including agriculture, were easily distinguished in video, mostly because they occur in distinctive patterns on the ground that are visible in the wide-angle frames. Irrigated and dryland agriculture were distinguished by the presence or absence of irrigation ditches or machinery. Other human-influenced types all had characteristic patterns.

6.3 Interpreter Confidence and Map Accuracy

We explored the relationship between interpreter confidence and map classification accuracy to see if there was a discernible relationship between the two. If there is a strong relationship, interpreter error in the reference data may be an important component of reported map accuracy. At the Gap-level especially, this was a concern due to the similarity of some cover types even when viewed on the ground (Section 6.2 above).

Video interpreters used a numerical scale (Table 6.1) to assign a "confidence level" to each sample video frame for both the Gap- and Anderson-level interpretations.

This scale quantified the interpreter's confidence in the assignment of a dominant cover type to each frame. Because many video frames were required to estimate dominant landcover in sample polygons, confidence scores from frames were first averaged for each sample polygon, and then average scores for the polygons were averaged for each landcover type. We also calculated mean confidence for all reference polygons that did not match the Colorado landcover map, and compared these to mean confidence for polygons that did, both at the Gap- and Anderson-levels.

Table 6.1. Numerical scale used to assign interpreter confidence in the assignment of a dominant Gap- or Anderson-level cover type to each interpreted video frame. This scale was used to explore the relationship between interpreter confidence and the frequency of matches and mismatches of reference data and mapped polygons.

Confidence Code	***Description***
1	Almost certain
2	Very confident by have some doubt
3	Probably correct but have significant doubt
4	Very low confidence but better than a guess
5	Essentially a best guess. Extremely low confidence

6.3.1 Gap-level Confidence Analysis

At the Gap-level, mean interpreter confidence for "correct" sample polygons was 1.422 (SE = 0.052, N = 184) and for "incorrect" sample polygons was 1.870 (SE = 0.045, N = 404), suggesting high interpreter confidence for both situations but statistically significant difference between the means at the 0.05 significance level.

For individual cover types, mean interpreter confidence ranged from 1.00 for open water (type 50) to 2.94 for saltbush fans and flats (type 25) (Table 6.2) but there was no meaningful relationship between mean confidence and either user's accuracy (R^2 = 0.14) or producer's accuracy (R^2 = 0.001).

Table 6.2. Mean interpreter confidence for each Gap-level type for which there were data from the Colorado interpretation. Cover types refer to Appendix 1. Mean confidences are from the scale in Table 6.1.

Cover Type	***Mean Confidence***	***Cover Type***	***Mean Confidence***
1	1.28	24	2.35
3	1.50	25	2.94
4	1.62	26	1.13
5	1.00	27	2.35
7	1.30	29	1.72
11	1.24	30	1.56
12	1.00	31	1.73
13	1.17	34	2.35
14	1.70	36	1.37
15	1.62	37	1.57
16	1.27	38	1.57
17	2.10	39	1.88
18	1.24	40	1.98
19	1.75	41	1.55
20	1.90	50	1.00
23	2.15		

These results suggest that although there is a difference in interpreter confidence on average for "correct" vs. "incorrect" map polygons at the Gap-level, the difference is small and the relationship with both user’s and producer’s accuracy of individual cover types is weak. We did not explore differences between the mean confidence of interpretation and the accuracy of polygons within types. Such an analysis may be done in the future to highlight particular cover types for which video interpretation accuracy may be an issue.

6.3.2 Anderson-level Confidence Analysis

Mean interpreter confidence at the Anderson-level for "correct" polygons was 1.074 (SE = 0.017, N = 340) and for "incorrect" polygons was 1.213 (SE = 0.035, N = 248). This difference is statistically significant at the 0.05 level. Mean interpreter confidence for individual Anderson-level cover types (Table 6.3) showed no relationship with user's or producer's accuracies.

Table 6.3. Mean interpreter confidence for Anderson-level dominant cover type designations. Cover types are from Appendix 1 and confidence values are from table 4.4.

Cover Type	*Mean Confidence*	*Cover Type*	*Mean Confidence*
1	1.06	8	1.42
2	1.08	9	1.50
3	1.26	10	1.13
4	1.18	11	1.23
5	1.39	13	1.00
6	1.00	15	1.06
7	1.13		

Overall interpreter confidence based on all interpreted polygons for the Anderson-level classes (1.132) was significantly higher than for the Gap-level classes (1.730), indicating the air videography is better suited as a source of reference data at the Anderson- than at the Gap-level.

6.4 Effects of Interpretation Errors on Accuracy Assessment

If there is error in both the map and the reference data, mismatches can be attributed to three possibilities: 1) the map is in error and the reference is correct, 2) the map is correct and the reference is in error, or 3) both the map and the reference are in error. Determining which of these possibilities is the case requires reference data for which confidence is high. Unfortunately, the very circumstances which encourage the use of remotely sensed reference data (large areas, poor access, etc.) discourage or prevent the acquisition of more reliable reference data. Accuracy assessment results are influenced differently by each of the three scenarios, resulting in either correct accounting of map error when reference data are correct, or faulty accounting when reference data are questionable.

Reference data (interpreted airborne videography) for the Colorado accuracy assessment undoubtedly is faulty to some degree. We suspect that, at the Gap-level, most reference error occurs within basin shrub and grassland types and emphasize that the reported accuracy measures, especially for these types, should be used with caution. A simulation study is planned to examine the effects of different types of reference data error on accuracy measures, but no on-the-ground evaluation of the Colorado video interpretation is planned.

6.5 Evaluation of Airborne Videography in the Northern Rocky Mountains

Airborne videography is one tool that can provide high resolution data about the earth's surface over large areas at relatively low cost. For these reasons, it is an appealing alternative to ground sampling for collection of reference data for accuracy assessment of statewide thematic landcover maps. Any tool, however, has strengths and limitations. Our experience in Colorado and Wyoming suggests that airborne videography is not adequate for distinguishing all landcover types at the Gap-level (level 5), and that it is more appropriate at the Anderson-level (level 4). This is particularly true for basin shrubland and grassland types, which are characteristic of large areas in the Greater Rocky Mountain region. Consequently, future missions to gather training data for remote sensing based classification should consider using airborne videography in combination with other (probably more expensive) strategies, like ground-based sampling. Airborne video works well for forested types, riparian areas and human-altered landscapes and is an efficient means of gathering data in those areas. Finally, we emphasize that landcover in different regions has unique characteristics that may add or detract from the usefulness of airborne videography as a sensing tool. Workers in these areas should consider this carefully when planning assessment strategies.

Vital for future research is detailed testing of interpreter’s accuracy for the full range of landcover types based on ground reconnaisance. Clearly, airborne videography has its own source of inherent error which requires measurement for it to be used as reference data for error analysis. Also, experimentation on the use of oblique rather than vertical videography might lead to a more effective usage that would be easily implemented.

We conclude that airborne videography has some appealing potential for characterizing landcover and for doing a limited range of attribute testing of maps created from other data, however, its use as reference data is inappropriate and the disappointing results reported here should not be used by themselves on to evaluate the accuracy of TM-based maps in Colorado or Wyoming.

Literature Cited

Anderson, James R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. A land use and land cover classification for use with remote sensor data. U.S. Geological Survey Professional Paper 964. Washington DC: U.S. Government Printing Office, 28 pp.

Bolstad, P.V. and J.L. Smith. 1992. Errors in GIS: assessing spatial data accuracy. *Journal of Forestry* 90:21-29.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1):37-46.

Congalton, Russell G. and Kass Green. 1993. A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing* 59(5):641-644.

Congalton, Russell G. and Kass Green. 1999. *Assessing the Accuracy of Remotely Sensed Data. Principles and Practices*. Lewis Publishers, Boca Raton, FL.

Congalton, Russell G. and Roy A. Mead. 1983. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing* 49(1):69-74.

Davis, F.W., P.A. Stine, and D.M. Stoms. 1994. Distribution and conservation status of coastal sage scrub in southwestern California. *Journal of Vegetation Science* 5:743-756.

Davis, F.W., P.A. Stine, D.M. Stoms, M.I. Borchert, and A.D. Hollinger. 1995. Gap analysis of the actual vegetation of California. The southwestern region. *Madrono* 42:40-78.

Driese, Kenneth L., William A. Reiners, Evelyn H. Merrill and Kenneth G. Gerow. 1997. A digital land cover map of Wyoming, USA: a tool for vegetation analysis. *Journal of Vegetation Science* 8:133-146.

Gopal, Sucharita and Curtis Woodcock. 1994. Theory and methods of accuracy assessment using fuzzy sets. *Photogrammetric Engineering and Remote Sensing* 60(2):181-188.

Hay, A. 1979. Sampling designs to test land use map accuracy. *Photogrammetric Engineering and Remote Sensing* 45:529-533.

Knight, Dennis H. 1994. *Mountains and Plains. The Ecology of Wyoming Landscapes*. Yale University Press, New Haven, CN. 338 pp.

Ma, Zhenkui and Roland Redmond. 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing* 61(4):435-439.

Marsh, Stuart E., James L. Walsh and Claudia Sobrevila. 1994. Evaluation of airborne video data for land-cover classification accuracy in an isolated Brazilian forest. *Remote Sensing of Environment* 48:61-69.

Merrill, Evellyn H., Thomas W. Kohley, Margo E. Herdendorf, William A. Reiners, Kenneth L. Driese, Ronald W. Marrs and Stanley H. Anderson. 1996. The Wyoming Gap Analysis Project. Final Report. USGS Biological Resources Division.

Oleson, K.W., K.L. Driese, J.A. Maslanik, W.J. Emery and W.A. Reiners. 1997. The sensitivity of a land surface parameterization scheme to the choice of remotely sensed land-cover datasets. *Monthly Weather Review* 125(7):1537-1555.

Schrupp, D.L., W.A. Reiners, T.G. Thompson, L.E. O'Brien, J.A Kindler, M.B. Wunder, J.F Lowsky, J.C. Buoy, L. Satcowitz, A.L. Cade, J.D. Stark, K.L. Driese, T.W. Owens, S.J. Russo and F. D'Erchia. 2000. Colorado Gap Analysis Program: A Geographic Approach to Planning for Biological Diversity – Final Report. USGS Biological Resources Division, Gap Analysis Program and Colorado Division of Wildlife, Denver, CO.

Scott, J.M., B Csuti, J.E. Estes and S. Caicco. 1991. Gap analysis of species richness and vegetation cover: an integrated biodiversity conservation strategy. Pages 282-297 in K. Kohm, ed. *Balancing on the Brink of Extinction: the Endangered Species Act and Lessons for the Future*. Island Press, Washington, DC. 318 pp.

Slaymaker, D.M., K.M.L. Jones, C.R. Griffin and J.T. Finn. 1996. Mapping deciduous forests in Southern New England using aerial videography and hyperclustered multi-temporal Landsat TM imagery. Pages 87-101 in *Gap Analysis: A landscape approach to biodiversity planning*. J.M. Scott, T.H. Tear and F.W. Davis (eds.). American Society of Photogrammetry and Remote Sensing, Bethesda MD, USA.

Smith, K.G., R.S. Dzur, D.G. Gatanzaro, M.E. Garner and W.F. Limp. 1998. Statewide biodiversity mapping for Arkansas: Final Report. Arkansas Cooperative Fish and Wildlife Research Unit, Center for Advanced Spatial Technologies (CAST), University of Arkansas, Fayatteville, AR.

Woodcock, Curtis and Sucharita Gopal. 1992. Accuracy assessment of the Stanislaus Forest vegetation map using fuzzy sets. Proceedings of the 4th Biennial Rmote Sensing Applications Conference. Orlando, Florida.

Woodcock, Curtis E. and Sucharita Gopal. 2000. Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *Int. J. Geographical Information Science* 14(2):153-172.

Woodcock, Curtis E., Sucharita Gopal and William Albert. 1996. Evaluation of the potential for providing secondary labels in vegetation maps. *Photogrammetric Engineering and Remote Sensing* 62(4):393-399.

Zadeh, L. 1965. Fuzzy Sets. *Information and Control* 8:338-353.