COMPUTER GRAPHICS CENTER
GIS • REMOTE SENSING • GPS

COMPUTER GRAPHICS CENTER

CGC REPORT No. 101

ACCURACY ASSESSMENT OF LAND COVER CHANGE DETECTION

BY

Authors are listed alphabetically except the chair:
Siamak Khorram, (Chair), North Carolina State University, Raleigh, NC
Gregory S. Biging, University of California, Berkeley, CA
Nicholas R. Chrisman, University of Washington, Seattle, WA
David R. Colby, National Oceanic and Atmospheric Administration, Beaufort, NC
Russell G. Congalton, University of New Hampshire, Durham, NH
Jerome E. Dobson, Oak Ridge National Laboratory, Oak Ridge, TN
Randolph L. Ferguson, National Oceanic and Atmospheric Administration, Beaufort, NC
Michael F. Goodchild, University of California, Santa Barbara, CA
John R. Jensen, University of South Carolina, Columbia, SC
Thomas H. Mace, Environmental Protection Agency, Research Triangle Park, NC

PREPARED FOR:

COASTAL CHANGE ANALYSIS PROGRAM
COASTAL OCEAN PROGRAM
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
BEAUFORT LABORATORY

May 4, 1994

BOX 7106, 5112 JORDAN HALL, NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NC 27695-7106 (919) 515-3430

CGC REPORT No. 101

ACCURACY ASSESSMENT
OF LAND COVER CHANGE DETECTION

BY

Authors are listed alphabetically except the chair:
Siamak Khorram, (Chair), North Carolina State University, Raleigh, NC
Gregory S. Biging, University of California, Berkeley, CA
Nicholas R. Chrisman, University of Washington, Seattle, WA
David R. Colby, National Oceanic and Atmospheric Administration, Beaufort, NC
Russell G. Congalton, University of New Hampshire, Durham, NH
Jerome E. Dobson, Oak Ridge National Laboratory, Oak Ridge, TN
Randolph L. Ferguson, National Oceanic and Atmospheric Administration, Beaufort, NC
Michael F. Goodchild, University of California, Santa Barbara, CA
John R. Jensen, University of South Carolina, Columbia, SC
Thomas H. Mace, Environmental Protection Agency, Research Triangle Park, NC

PREPARED FOR:

COASTAL CHANGE ANALYSIS PROGRAM
COASTAL OCEAN PROGRAM
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
BEAUFORT LABORATORY

May 4, 1994

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Accuracy Assessment of Land Cover Change Detection

1.0 EXECUTIVE SUMMARY

Effective management of coastal resources requires knowledge of change in coastal land cover with national scope, regional continuity, local reliability, and reasonable repeat cycles. National Oceanic and Atmospheric Aadministration's Coastal Ocean Program, through its Coastal Change Analysis Program, C-CAP, has developed a methodology and is actively producing retrospective and progressive regional land cover change databases from remotely sensed data for monitoring coastal land cover on a repeat cycle of 3 to 5 years. C-CAP has identified accuracy assessment as a key constraint to the full development and application of these regional land cover change databases. Determination of the level of accuracy of these change detection databases is essential to refine change detection methodologies and to apply the data appropriately. Land cover change databases require, therefore, the collection of reference data not only for calibration but also for accuracy assessment purposes. Quantification of the degree of error and determination of the level of statistical significance of land cover change in regional land cover change databases is required. Unfortunately, the remote sensing and statistical literature has not provided guidelines for land cover change detection and accuracy assessment appropriate for areas of regional extent.

The objective of this report is to recommend a procedure to NOAA's C-CAP for assessing the accuracy of land cover change detection datasets derived from remotely sensed data. This paper focuses on the classification, mapping and verification of the reported extent of coastal wetlands, adjacent uplands, and submerghed lands. The primary data are satellite imagery (for wetlands and uplands), aerial photography (for submerged aquatic vegetation) and ancillary field data. Land cover classes within the classification scheme are exhaustive, mutually exclusive, and can be identified with remotely-sensed data combined with site observations. **The recommended procedure is the result of four workshops held during 1993 and attended by a panel of experts in spatial statistics, remote sensing, environmental monitoring, and geographic information systems (GIS).**

Two change detection algorithms are recommended to C-CAP: post classification comparisons and a procedure which utilize a binary change masking. Post classification change detection is considered suitable for both photo interpretation and spectral imagery while the binary change mask method is suitable for spectral imagery only. Post classification change detection involves complete classification of the remotely sensed data at two time periods. A comparison of land cover type is then conducted on a pixel by pixel basis to identify areas of change. The binary change mask compares spectral data directly and identifies locations of spectral change between the two discrete time periods. The base time period is classified completely but only the locations of change are classified in the other time period for cost saving purpose.

This report differentiates the concepts of error and generalization and then deals exclusively with error. Both error and generalization are inconsistencies between the subject area and the database. Inconsistencies that are beyond the temporal, spatial and categorical resolution of the database are attributable to generalization and not to error. C-CAP has in its Guidance for Regional Implementation, defined levels of generalization in the database. These were approved by consensus in five regional workshops and were determined to be necessary and acceptable concessions to achieve the regionally synoptic data essential for short term (3-5 year) change detection. The degree of generalization inherent in the database, therefore, may or may not be acceptable to all potential users of the databases.

Assessing the error of regional change detection databases requires attention to the classification system, the nature of land cover and land cover change data, problems inherent to regional databases, and selection of error evaluation algorithms. Existing land cover databases may not be considered as suitable reference data for calibration or accuracy assessment due to not being synchronous and contemporary with C-CAP data, not addressing the same classification scheme, and not having a measured level of reliability. Accuracy assessment of regional change detection databases is a prime illustration that methods for remote sensing contemporary in analysis of an area of local extent do not apply for areas of regional extent or for past and multiple periods of time.

Errors of many types enter the database during acquisition, processing, analysis and conversion of the data. In addition, errors occur as a result of error assessment and final product presentation. Systematic or random and positional or attribute errors can occur. Some positional errors, such as registration error, can be estimated with reference to well defined and accurately positioned features visible in the database. Attribute error is qualitative and can be detected by comparison with local observation of land cover categories. Errors in change detection databases include errors of omission (change occurred but was not detected) and commission (change was detected but did not occur). Assessment of change error requires accurate and independent data for local categorization of land cover at both times of remote observation.

In this report, certain statistical procedures appropriate to the unique problem of accuracy assessment of regional land cover change databases are discussed. The procedures include sampling strategies, data processing, presentation approaches and formulae to compute diagnostic statistics. For change detection accuracy assessment traditional sampling techniques cannot usually be employed This is largely due to the fact that the change polygons cover only a small portion of the original image and would not be well detected with, say, random sampling. Because of these limitations, partitioning of the image into two strata - **Change** and **No Change** is proposed. The (predicted) No Change Stratum by far is the largest stratum generally encompassing over 90% of the original image, or thematic map. The sampling of these two components can be done following standard techniques such as random, stratified random and systematic unaligned designs provided the second case is not a truly rare event (say <10%). The real advantage of partitioning the image is that different types and intensities of inventories on these two strata may be conducted.

Creation of a reference data set for accuracy assessment depends upon accurate location of stratified random sample locations and accurate local data for the **"from"** and **"to"** categories at the location of change. Global Positioning System (GPS) technology with real time differential correction is the method of choice for positioning in the field. Once located, an objective procedure for observing features and recording data relevant to identification of the land cover categories and types of change is conducted with the aid of a detailed data recording form. Such forms need to be generated regionally.

Generation of an error matrix and performance of a KAPPA analysis are recommended for quantification and interpretation of change detection errors and testing for statistical significance of change. The error matrix is an excellent starting point for a number of descriptive and analytical statistical techniques. The matrix identifies errors of omission and commission and the frequency of all of the possible combinations of "from" and "to" categories. Error tabulation in the change matrix is recommended for error summation and for discrete multivariate techniques. The change detection error matrix has the same characteristics as the traditional classification error matrix but will assess errors in change between time periods and not simply errors in a single classification. The KAPPA analysis is a useful measure of agreement or accuracy and provides a test for statistical significance of observed changes.

2.0 INTRODUCTION

One of the "grand challenges" for scientists and policy makers in the 1990's is to achieve a deeper understanding of global processes with particular attention to human interactions (Office of Science and Technology Policy, Executive Office of the President). In the current decade, global environmental change has become a major national and international policy issue. Regional change has been recognized as an important aspect of resource management and environmental restoration. Coastal habitats are crucial resources in jeopardy in the United States and throughout the world. Many of the world's greatest cities (e.g. Amsterdam, Boston, London, San Francisco, Tokyo, Washington D.C., etc.) threaten wetlands by virtue of being built on rivers, estuaries, and coastal areas. The impetus to measure change in coastal land cover results from the high environmental and economic value of vegetated uplands, wetlands, and submerged land. There is widespread consensus among citizens, resources managers, and scientists that unacceptable losses are occurring. Policy and management decisions demand regionally comprehensive, rigorous data on changes in the quantity, quality, and distribution of coastal resources.

For more than a century land cover, the visible surface of the earth, has been recognized as the principal indicator of changes associated with human activity and with natural processes that alter physical and cultural environments (Hartshorne, 1939; Stamp, 1948). Land cover change databases, quantitatively indicating differences from one time period to another at a determined level of accuracy, are essential to the scientific analysis and management of environmental processes. Unfortunately, previous monotemporal and multitemporal programs based on field collected <u>or</u> remotely sensed data have experienced mixed success. Since the 1930's remote sensing, first from aircraft and later from satellites, has been the principal means by which land cover is observed, categorized, recorded, and quantified (Khorram et al., 1991; Dobson et. al., 1993). Remote sensing provides the most feasible approach to regional land cover change detection. Thematic maps or images obtained at different times can be compared to identify the location, extent, and type of change. Remote sensing products consist of analog or digital images that may be expressed as static photographs, maps, or digital databases for a single time period or compared to show change between two or more times.

The accuracy of land cover maps produced from remotely sensed data is an essential element required by both environmental researchers and managers. This need has been recognized by the remote sensing community for over a decade (Hord and Brooner, 1976; Van Genderen and Lock, 1977; Ginevan, 1979; Rosenfield and Malley, 1980; Congalton and Mead, 1983; Czaplewski and Catts, 1992; Congalton, 1991; Khorram et. al., 1992a; Goodchild et. al., 1992; Ferguson et al., 1993). Techniques developed for accuracy assessment of interpretations of aerial photographs have generally been extended to the evaluation of land cover data sets derived from non-photographic sensors (Hellden and Stern, 1980; Roller and Visser, 1980; Khorram et al., 1992b). Several studies have incorporated more sophisticated statistical procedures in the analysis of error matrices compiled for monotemporal land cover data sets generated from remotely sensed digital data (Congalton et al., 1983; Aronoff, 1985; Congalton, 1988). In dealing with change detection data sets, research has shown that, in addition to factors traditionally held to affect accuracy assessment of monotemporal datasets, one must also consider factors such as image registration and boundary problems (Wickware and Howarth, 1981; Haley, 1985; Corr et al., 1989; Singh, 1989).

NOAA's Coastal Ocean Program, through its Coastal Change Analysis Program (C-CAP) has developed protocols for regional land cover change analysis, developed a Coastal Land Cover Classification System (CLCCS) suitable for use with satellite data and aerial photography, and has identified accuracy assessment as a key factor in large-area change analysis methodology.

The purpose of the C-CAP project is to build a digital data base that, when integrated with other data within a GIS, may ultimately enable scientists to link development in the coastal regions to the ecological and economic productivity of the coastal and the marine environment. To accomplish this, C-CAP has developed a comprehensive, nationally standardized and locally managed or implemented, information system for land cover and habitat change in the coastal regions of the United States (Dobson et al., 1993). The system will emphasize a geographic approach including the use of geographic information systems (GIS), field data, and remotely sensed data. Data from satellite scanners, and aerial photography will be interpreted and classified, assessed for change, and for accuracy of the change data base. Output products will include: (l) spatially registered digital images, (2) hardcopy maps, and (3) tabular summaries. However, current remote sensing literature suggests that techniques of change detection are well developed for areas of local extent, poorly defined for areas of regional extent, and non-existent for areas of continental or global extent. **This report focuses on accuracy assessment for large-area static and change detection databases developed from satellite remote sensing of uplands and wetlands and from aerial photographs of submerged land.** Readers should refer to the "Guidelines for Regional Implementation" (Dobson et al., 1993 and Klemas et al., 1993) for C-CAP definitions of land cover categories and more detailed descriptions of methods pertinent to the development of change data bases and accuracy assessment techniques. For satellite data, land cover change will be detected in a pixel by pixel comparison from different time periods. For photographic data, change will be detected in a post classification comparison of the spatial data. The resulting information will enhance conceptual and predictive models and support coastal resource policy analysis.

Currently, researchers and resource managers using remotely sensed data for mapping and monitoring change in land cover rely on *ad hoc* approaches to accuracy assessment. Many of the techniques employed are based directly on approaches developed for one-time inventories or for more simplistic technologies or data types. Little attention has given to establishing step-by-step methodologies or standards for data accuracy in change detection and reference data sets. Most research has focused on methodologies for producing the change detection data sets, and, while data accuracies are reported for the different methodologies, little or no detail is reported on how the reference data for accuracy assessment were derived.

NOAA C-CAP is a cooperative interagency and state federal effort (Cross and Thomas, 1992), which is currently developing regional databases on land cover change for coastal regions of the United States. The change detection and analysis cycle is primarily based on a regular cycle ranging from one to five years may vary according to the rate and magnitude of change in each region. The broad land cover categories of interest in the paper are composed of coastal wetlands, adjacent uplands, and submerged lands. The primary data are satellite imagery (for wetlands and uplands), aerial photography (for submerged aquatic vegetation) and necessary field data. **The recommended procedure reported here is the result of four workshops held during 1993 attended by the authors.**

3.0 ACCURACY ASSESSMENT ISSUES

Assessing the accuracy of change detection products derived from remotely sensed data requires careful attention to several issues, including: 1) the land cover classification system, 2) the nature of land cover change, 3) problems specific to regional land cover change databases, 4) creation of reference databases, and 5) selection of error evaluation algorithms.

3.1 The Coastal Land Cover Classification System

Change analysis begins with the classification system. The classification system brings with it a necessity for arbitrary subdivision of continua into categories and polygons with discrete boundaries. It also represents a subjective description of the environment, as we see it. Change detection is relative to the classification system being used. Thus, it is vital that the classification system be consistently applied to all data sets undergoing change analysis. It is also important that the classification system being used have classes that are exhaustive and mutually exclusive. Just as all measurements have some error, there will be some inconsistency in application of the classification system; there will be some overlap between classes; and, there will be some classes that are not anticipated by the taxonomic system. Tests for consistency, independence and completeness can be made, but, for the purposes of making practical recommendations to C-CAP on accuracy assessment for change detection, we assume that the classification system is valid and that it is uniformly applied to all remote and in situ data sets.

To satisfy the nationwide coastal focus of the C-CAP information system, the CLCCS for coastal uplands, wetlands, and submerged ecosystems was developed to be hierarchical, respect ecological relationships, optimize the potential for discrimination by remote sensors, be usable with GIS, and be compatible with other data bases (Klemas et al., 1993; Dobson et al., 1993). CLCCS emphasizes wetlands and photic deep water habitats, critical habitats which support the living marine organisms (Wilen, 1990). However, upland categories also are included because upland land cover and land use influence water quality and the distribution and productivity of wetlands and submersed habitat within the photic zone.

The CLCCS includes three Level I superclasses; 1.0 Uplands, 2.0 Wetlands, and 3.0 Water and Submerged Lands. These superclasses are broken down into classes and subclasses at Level II and Level III, respectively. The CLCCS focuses on both upland and coastal wetland land cover which can be derived primarily from remotely sensed information and then used to produce wetland ecosystem change information. While the categories of the CLCCS, discussed briefly below, are generally compatible with Anderson et al. (1976) and Cowardin et al. (1979), system definitions, some modifications were necessary to reconcile inconsistencies between Anderson and Cowardin, to move toward land cover and away from land use categories, and to accommodate remotely sensed data and detailed definitions of all classes and subclasses are presented in the C-CAP protocol document (Dobson et al., 1993).

3.1.1 Uplands

The definitions of Uplands classes and subclasses are similar to those in Anderson et al. (1976) and USGS Technical Instructions (1992). The Uplands superclass consists of seven classes: Developed Land, Cultivated Land, Grassland, Woody Land, Bare Land, Tundra, and Snow/Ice. Upland classes were modified from Level I classes in the USGS Land Use/Land Cover Classification System (Anderson et al., 1976; USGS, 1992). For detailed information see Klemas et al., (1993).

3.1.2 Wetlands

According to the Cowardin classification system, wetlands are lands where ground water saturation is the dominant factor determining soil development and the types of plant and animal communities living in the soil and on its surface. The single feature that all wetlands share is soil or substrate that is at least periodically saturated with or covered by water.

The upland limit of wetland is designated as: (1) the boundary between land with predominantly hydrophytic cover and land with predominantely mesophytic or xerophytic cover; (2) the boundary between soil that is predominantly hydric and soil that is predominantly nonhydric; or (3) in the case of the wetlands without vegetation or soil, the boundary between land that is flooded or saturated at some time during the growing season each year and land that is not (Cowardin et at., 1979). The majority of all wetlands are vegetated and are found on soil.

In the CLCCS, "Wetland" includes all areas considered wetland by Cowardin et al., (1979), except for wetland Bottoms, wetland Reefs, wetland Aquatic Beds, and Nonpersistent Emergent Wetlands. The class breakdown under Wetlands was adopted from the Cowardin system. At Level II, CLCCS employs certain Cowardin classes (e.g., Rocky Shore, Unconsolidated Shore, Emergent Wetland) or groups of the Cowardin classes (e.g., Woody Wetland composed of Scrub-Shrub and Forested Wetland), in combination with Cowardin systems (i.e., Marine, Estuarine, Riverine, Lacustrine, Palustrine). Thus a typical Level II class in the CLCCS might be Palustrine Woody Wetland.

Salinity displays a horizontal gradient in marshes typical of coastal plain estuaries. This is evident not only through the direct measurement of salinity but in the horizontal distribution of marsh plants (Daiber, 1986). Therefore, Estuarine Emergent Wetlands are partitioned into Haline (Salt) and Mixohaline (Brackish) Marshes. CLCCS uses the definitions shown in the Cowardin system.

3.1.3 Water and Submerged Land

All areas of open water are assigned to the Superclass Water and Submerged Land. This superclass includes the open water of both wetland and deepwater habitats of the Cowardin et al. (1979) classification, i.e. with <30% cover of trees, shrubs, persistent emergency plants, emergent mosses, or lichens. The Superclass is comprised of six Level II classes, of which two are the current emphasis of C-CAP monitoring: Water and Marine/Estuarine Aquatic Bed. The class Water includes Cowardin et al.'s (1979) Rock Bottom, Unconsolidated Bottom, and Nonpersistent Emergent Wetlands, as well as undetected Reefs and Aquatic Beds. Most C-CAP products will not subdivide water into types. C-CAP recognizes the subclasses: Marine/Estuarine, Riverine, Lacustrine, Palustrine, of the class Water. However, the C-CAP system identifies the subclass data as optional in the database. Having the water subclasses also makes the C-CAP scheme more compatible with the Cowardin et al. (1979) system.

The class Marine/Estuarine Aquatic Bed includes two subclasses. The current C-CAP focus is the Subclass: Rooted Vascular (SRV for submersed rooted vascular) which, as an option in the data base, is subdivided into High Salinity (25 ppt) and Low Salinity (< 5 ppt). At >5 ppt salinity, habitats defined by the occurrence of true seagrasses are separated from those at <5 ppt, defined by submersed grasses and forbs that tolerate or require low salinity. The exception is euryhaline grasses (e.g. Rupia maritima) which thrive above and below 5 ppt. Aquatic beds above and below 5 ppt are very different in terms of flora and fauna but both are important. High Salinity includes mesosaline, polysaline, eusaline, and hypersaline salinity categories of Cowardin et al. (1979). Low Salinity includes oligosaline and fresh categories. The Subclass: Algal is optional in the data base awaiting establishment of guidelines for remote sensing of this diverse and regionally variable resource.

3.2 The Nature of Land Cover Change

Many kinds of changes occur on the surface of the Earth, but this report is concerned exclusively with those that appear as changes of land cover. While some of these are caused by smoothly changing underlying environmental conditions, such as global climate regimes or increased salinity, others reflect more or less catastrophic events, such as flooding, urban development, or infilling. When detected and recorded from space, change of land cover may show a variety of geographic signatures. Some changes may affect entire areas uniformly and instantaneously, while others may take the form of slow advances or retreats of boundaries between land cover classes, and other changes may have very complex spatial textures.

As discussed by Jensen and Toll (1982) and Chrisman (1993), the limitations of the above definition include: 1) change through time could be gradual due to slow trends, or abrupt due to catastrophic or episodal events and difficult to assess with arbitrarily timed "snapshots" of remote data; 2) change in categories does not elucidate the process causing the change; 3) change in category over time is confounded with error at both of the two times of observation.

Characterization of change involves establishment of the "from" and "to" categories for a location on the map. It also involves establishing the change in a spatial context. Change at a specific location may be qualitative, from one category to another. In a spatial context, change can be of several types:

1) a polygon becomes a different land cover category

2) a polygon expands, shrinks or changes shape

3) a polygon shifts position

4) a polygon fragments or proximal polygons of the same type coalesce.

Significance of change from an environmental management and research perspective is a function of the relative values of the "to" and "from" categories and of the implications the change has regarding underlying processes and predictions of future changes. The significance of change is coupled to the spatial context of the change because in different places the same land cover category may have different values and changes, when placed in a spatial context, can be powerful sources of testable hypotheses.

Detectability of change is a function of the nature of the "from" and "to" categories and the spatial extent and context of the change. If a "from" category is broadly defined in terms of spectral distribution, substantial spectral change may occur without a change in category. This is particularly troublesome if the spectral distributions of two land cover categories overlap. At what point does one switch a land cover category when the spectral data is intermediate between the "from" and potential "to" category? Change over a large spatial extent is less likely to be missed than change over a small spatial extent. Regionally related changes of the same type are less likely to be ignored but are likely to be partitioned between "error" and "change".

One major task in change detection involves separating the apparent differences into those likely to arise from error sources and those produced by real change processes. While errors are both positional and attribute, change processes are also of the same two forms. Some changes are incremental and connected in space, as when a polygon border moves. But such boundary movement may be indistinguishable from positional error. A change of greatest interest to management and research is one with a specific direction, explicable by environmental forces at work. Positional errors or changes due to cyclic interval forces may weave back and forth in a more or less random manner. A test based on the runs test was proposed by Goodchild (1978) that might help in separating these two forms in some cases.

Change may also convert large objects in a more instantaneous manner. These changes may be distinguishable from some classification errors by looking at the distance between the changes in the spectral space of the remote sensing images, by looking at the likelihood of the environmental processes which could cause the apparent change, or by examining the historical record of previous changes.

3.3 Problems Specific to Regional Land Cover Change Databases

Maps, whether analog or digital, are models of features found on the Earth's surface. Map accuracies vary depending on the methods and care used in producing them (Mailing, 1989). The thematic accuracy of a land cover map is constrained by several factors, including the land cover classification scheme, quality of data sources, size of minimum detection and minimum mapping units, scale of presentation, and expertise of the photointerpreters or image analysts, and cartographers producing the map. As with all maps, land cover maps contain errors that should be quantified before they can be used with confidence.

Congalton (1991) suggests that until recently the idea of assessing the classification accuracy of remotely sensed data was treated more as an after-thought than as an integral part of any project. In fact, throughout the 1980's, studies would simply report a single number to express the accuracy of a classification. In addition, many assessments were conducted using the same data set as was used to train the classifier. This training and testing on the same data set obviously results in overestimates of classification accuracy.

Several methods have been used for accuracy assessment of land cover mapping results based on remotely sensed data. Rosenfield (1980) proposed the use of analysis of variance techniques for accuracy assessment. However, violation of the normal theory assumption and independence assumption when applying this technique to remotely sensed data has limited its application. Aronoff (1985) suggested the use of a minimum accuracy value as an index of classification accuracy. This approach is based on the binomial distribution of the data and is therefore appropriate for remotely sensed data. The major disadvantage of this approach is that it is limited to a single overall accuracy value rather than using the entire error matrix. However, it is useful in that this index does express statistically the uncertainty involved in any accuracy assessment. Skidmore and Turner (1989) have begun work on techniques for assessing error as it accumulates through many spatial layers of information in a GIS, including remotely sensed data. These techniques have included using a line sampling method for accuracy assessment as well as probability theory to accumulate error from layer to layer.

The methods mentioned above focus on the attribute accuracy of classifications. There is also need to consider the positional accuracy of remote sensing at least in the process of registration of images. The average error in identifying well-defined points is often below the pixel size using modern sensors, but the registration of images is not the end of spatial error effects that will influence change detection. While both positional and attribute errors occur in the processing of remotely sensed information, the positional accuracy is often considered to be separate from the classification accuracy. Many procedures for assessing classification accuracy explicitly avoid samples near boundaries where a positional effect might be detected. Some research (Goodchild and Dubuc, 1987; Chrisman, 1989; Chrisman and Lester, 1991; Goodchild et al., 1992) has attempted to model and test these errors in a more integrated manner.

3.3.1 Large Area

Current guidance does not work well for large areas. There is an underlying assumption that reference data of higher accuracy can be developed through field work and/or aerial photographic interpretation. Differences among individuals are well documented (McGwire, 1992), but they are usually ignored when the area is small enough to be covered by a single individual and revisited to resolve disagreements. For large areas, variability among observers is a major concern. Field verification in the Chesapeake Bay (Burgess et al., 1991, Shapiro, 1993), for example, has involved more than 50 individuals from at least two state governments, three universities and six separate federal agencies. All these individuals have spent time in the field, but some have relied more heavily than others on aerial photographs which introduce yet another perceptual filter. In addition, there is an assumption that training of field workers can overcome the problem of observer variability. Our experience suggests that the instruction set required to ensure consistency over large areas, large numbers of classes, and large teams of field workers greatly exceeds the resources available in most projects and the tolerance of even the most committed analysts.

3.3.2 Large Number of Classes in Matrix

The potential number of categories ("no change," "from," and "to") in a change database for two time periods is the square of the number in a single time period. Conceptually, each "from" and "to" category can be treated as a separate category for sampling. This creates a very large number of potential classes which must be sampled when performing error evaluation.

3.3.3 Impossibility of Field Verification for Past Time Periods

Current methods do not work well for past time periods. There is an underlying assumption that reference data of higher accuracy can be developed from aerial photographs. In reality, contemporaneous (same year, same season) photographs are seldom available, and differences of a year or more can cause great uncertainty in continuous change (e.g. forest regrowth from clear-cutting) or abrupt change (e.g. precise date of construction). Generally, data derived through aerial photographic interpretation are not always more reliable than data derived through classification of digital satellite data. The two approaches result in different types of error because each has its strengths and weaknesses. Many analysts feel more comfortable with aerial photographs because the images appear more like the visual images familiar to human observers in the field. However, a photograph actually contains less attribute information than a multispectral scanner image due to the large number of spectral bands, several extending beyond the visual spectrum, in the digital image. The perception of greater information in the photograph comes from the higher spatial resolution of most photographs and the habit of employing spatial pattern recognition to interpret features. Much of the recent improvement in satellite classification has come because the image processing systems have become interactive enough to support that same kind of spatial pattern recognition by the human analyst viewing a workstation screen. The finer resolution of many aerial photographs permits the photointerpreter to recognize patterns at a more human scale, but the 30 m X 30 m resolution of Thematic Mapper data is also sufficient for many pattern recognition tasks. For example, this resolution is generally adequate and may even be preferred for the task of ensuring that most marshes fall in the proper hydrologic relationship to streams and waterbodies. Digital satellite data also can offer a significant advantage in positional accuracy and precision.

3.3.4 Fuzzy, Continuous Phenomena

Current methods treat land cover categories as discrete sets, but change detection accentuates the fact that land cover is spatially, temporally, and categorically continuous (Foody, 1992). Land cover change can be distinct as when a forest is cut, the land is graded, and buildings are constructed. Such changes are widespread and also of interest to scientists and policy makers because they substantially alter physical, economic, and cultural conditions on a relatively long-term basis. More common, however, are the gradual changes that occur at the indistinct boundaries of polygons, at the imprecise definitional boundaries among classes, and in the temporal waxing and waning of key indicators (e.g. moisture, vegetation, permafrost). These changes occur naturally in response to numerous physical factors (e.g. weather, erosion, sedimentation) and may be of little consequence in a healthy ecosystem where growth is equivalent to decline. In contrast, they may be crucial indicators when basic conditions alter, as anticipated in global climate change. For change detection, *per se*, it is difficult to define an objective boundary distinguishing between minor and major shifts along the land cover continuum. New methods will have to be developed for accuracy assessment of land cover change databases, and the guidance must explicitly address the fuzziness of spatial, temporal, and categorical dimensions as well as the fuzzy perceptions of observers.

3.3.5 Error in Regional Reference Databases

The terms "error" and "accuracy" are frequently used in regard to attributes and boundaries of polygons obtained from remotely sensed data. Error and accuracy estimation requires comparison to reference data of higher accuracy and reliability. Conceptually, "accuracy" can only be determined on the basis of a highly specific set of criteria that effectively generalizes for large areas the resolution and accuracy that is attainable in field surveys. As noted previously (Sec. 3.3.1), it is essential that criteria are devised to consistently assign each class and the spatial extent of that land cover class whether the data on which the judgement is based is remotely or locally obtained.

The design and accuracy of the regional reference database are of paramount importance. Land cover classes are abstractions intended to generalize and simplify complex real world phenomena. The abstract classes are discrete, yet the data surfaces they represent are continuous. Real world land cover is, in fact, a prime example of fuzzy set composition. Spatially, natural land cover grades from one type to another through transition zones that clearly are not distinct boundaries. Furthermore, classes and zones change fractally in accordance with the resolution of observation. Categorically, land cover types are distinguished from one another by characteristic configurations of water, minerals, and woody and herbaceous biomass in an infinite variety of proportions and two-dimensional and three-dimensional arrangements. Definitions are frequently attempted, but it is conceptually impossible to define all possible variations. Temporally, land cover at a specific place continuously changes from one type to another due to weather, climatic change, hydrologic alteration, other natural processes, and human activity. Usually, change is imperceptible over brief periods of observation, but substantial change can be expected from year to year, and massive changes can be expected over long periods of time. Observationally, land cover is fuzzy because each observer is forced to assimilate all of the spatial, categorical, and temporal fuzziness into a discrete judgement. Thus, the designation of static land cover for any given site, even in the field, is probabilistic, not deterministic. Thus, for land cover change, the designation is even more complicated because of the additional need to decide what constitutes a significant change among the many changes continuously occurring in a real world landscape while referencing these apparent changes to a static database.

3.4 Selection of Error Evaluation Algorithms

Accuracy assessment is a prime illustration that numerous methods and techniques, well established for remote sensing of small areas and single time periods, do not serve for large areas, past time periods, or change databases. Current guidance (Chrisman, 1991a; Congalton, 1991; NIST, 1992) does not work well for change detection because attribute and positional accuracy are measured independently. Sample points for measuring attribute accuracy typically are taken at the interiors of class polygons and are separate from the samples used to measure positional accuracy (Root Mean Square Error-RMSE). Even for the static land cover of a single time period, the large areal extent of regional databases creates many problems that have not been resolved in remote sensing literature. It is infeasible, for example, to conduct field investigations of such a large area with a sampling pattern as dense as that employed in an area of local extent. Hence, the sample to population ratio is lower and, in effect, each sample site stands for a much larger portion of the total population. The magnitude of the field effort is compounded in a change detection database by the large number of classes.

Change databases usually are expressed as a change matrix in which each axis represents the quantity of area in each land cover class in each time period. The challenge is to develop statistics that will adequately capture the relationship of such small samples to the larger regional population (area stratified by land cover class). The problem may be stated: Given a region of size A and a matrix of dimension X, how large as a the stratified sample is required to adequately represent regional accuracy? Once this sample size and distribution are determined, the question shifts to practical considerations, principally the aggregate time and cost of sampling compared to project resources. Section 6 makes recommendations on the proper selection and application of statistics for assessing the accuracy of change detection products.

4.0 ERROR CHARACTERIZATION

Broadly speaking, the simple definition of change is difference - difference in the landscape between two times. Unfortunately, the definition of error is quite close. Error can be roughly defined as the difference between the data recorded and a "true" value that might have been obtained with greater expense, effort, etc. While the concept of a true value is a useful construct for statistics, it is rather difficult in actual practice. Errors are unexplained variations, whether they are estimated over time or by other repetitions. One of the most important issues in change detection is ensuring that the changes reported are not confused with errors. This is a difficult proposition, and unlikely to be handled in a totally satisfactory manner.

In the treatment of spatial data, it is common to consider as errors all the differences between the information recorded and the "truth". The concept of truth here is decidedly less than absolute. In practice, as compared with a source of higher accuracy, the one with a RMSE of 1/3 of the RMSE of the source in question, may be adequate to perform tests with reasonable success. Other measurements can also give an estimate of the variability of a data source using sampling procedures or repeated measurement as in surveying. Such an approach would focus on the errors introduced by each technical process in preparing the information. It could be termed an "error budget" approach.

Another attitude towards spatial data considers the variability to be inherent in the landscape as well as in our procedures used to measure it. The clean and tidy models that we apply to land cover are really only approximations for ecological systems that are characterized by a whole hierarchy of processes. These processes vary in time from minutes to thousands of years and in space from small local areas to large regions. Hence, we can speak of a landscape that has no enduring "truth", but more of a probability density function, a stochastic process.

Whatever our attitude about error, errors can be tracked back to the various technical steps in information handling. Errors will have a specific form depending on the nature of the technical processes involved. This chapter of the report will proceed with a review of the error sources that should be expected (section 4.1), and a general characterization of error types (section 4.2). Section 4.3 will deal with an error budget approach, although not taken for this study.

4.1 Error Sources

The procedures used in constructing a data product have an influence on error. Some procedures increase the error, so that it accumulates, while other procedures are designed to reduce error. A basic step in accuracy assessment is to understand the nature of the specific procedures used in the particular case. Since there are so many possible production processes, this section will only consider the generic circumstances of satellite remote sensing, as an example.

There are many sources of errors that are associated with classifications derived from remotely sensed data. These sources are well documented in a paper by Lunetta et al. (1991) and include data acquisition, data processing, data analysis, and data conversion error. In addition, errors occur as a result of the error assessment process and the final product presentation. Figure 1 presents a summary of these error sources. The error for a single datum is represented in a figure from Lunetta et al. (1991). A second date has been added and the errors due to change have also been incorporated into the new figure. It should be noted that the errors are likely to correlate at the two dates. It is hoped that some learning occurred during the first classification that would minimize the errors in the second date. However, it is possible that the two classifications would be performed by separate groups at different times, and in this case, the errors would be independent.

The major steps that introduce errors start with the basic source material. The sensor systems have limitations, and are influenced by environmental conditions such as atmosphere, soil moisture, and plant phenology. The imagery is connected to the map projection by correlation to ground control, another source of information. This step contributes to positional error, considered in the next section below. The raw material is then processed, first through a set of rectifications, then typically through statistical classification procedures.

As mentioned above, the error assessment process itself can be a source of error and confusion. Congalton and Green (1993) document many of sources of confusion between the remotely sensed classification and the reference data used. In this case, the errors would be independent.

1. Registration differences between the reference data and the remotely sensed map classification.

2. Delineation error encountered when the sites chosen for accuracy assessment are digitized.

3. Data entry error when the reference data is entered into the accuracy assessment data base.

4. Error in interpretation and delineation of the reference data (e.g. photo interpretation error or field observation error).

5. Changes in land cover between the date of the remotely sensed data and the date of the reference data (temporal error). For example, changes due to fires or urban development or harvesting.

SOURCES OF ERROR USING REMOTELY SENSED DATA FOR CHANGE ANALYSIS

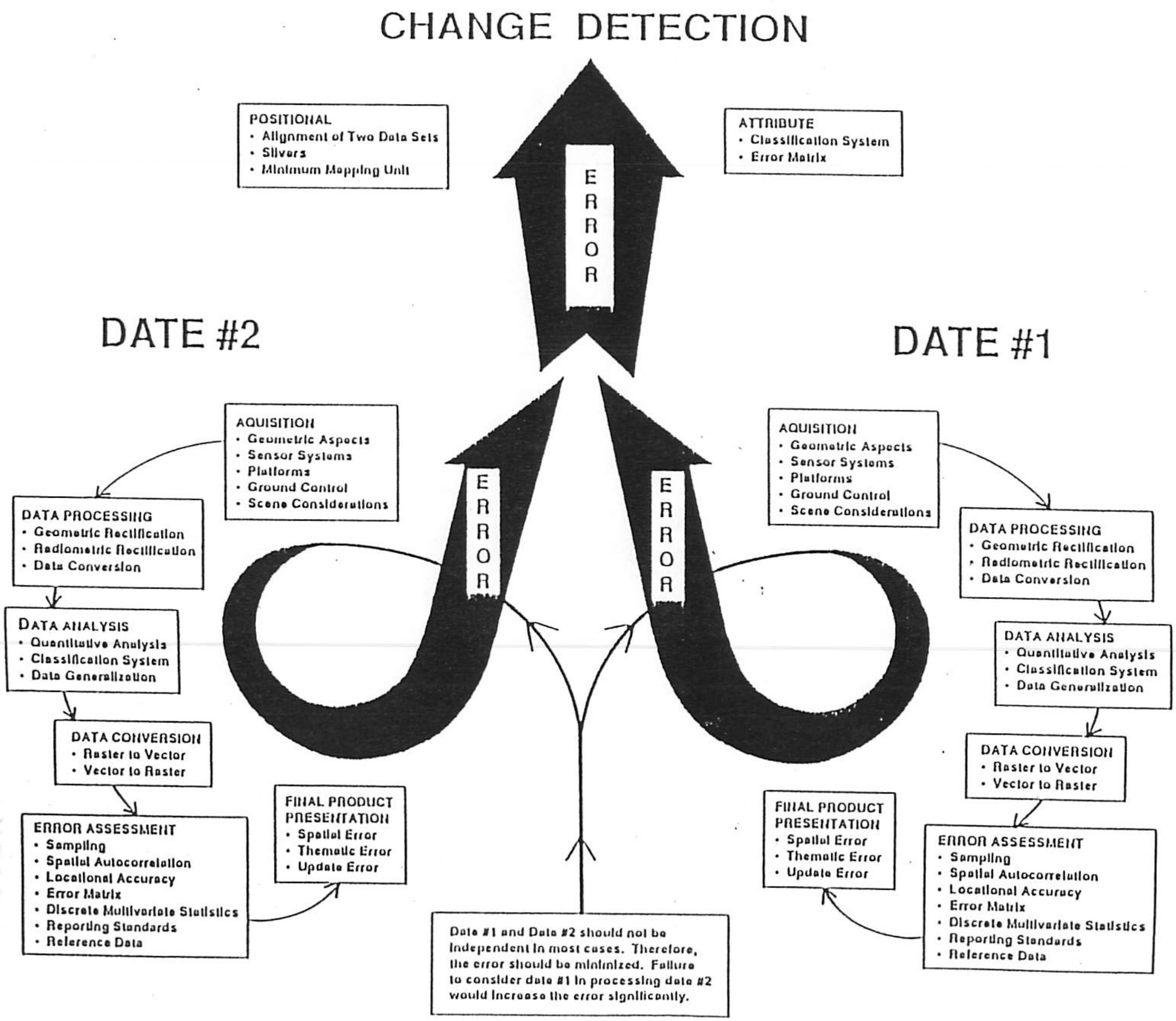

Figure 1 Sources of Error Using Remotely Sensed Data for Changte Analysis

6. Variation in classification and delineation of the reference data due to inconsistencies in human interpretation of heterogeneous vegetation.

7. Errors in the remotely sensed image classification.

8. Errors in the remotely sensed image delineation.

4.2 Error Partitioning

Whatever the particular sources of error, spatial information has distinct forms of differences. At the most primitive level, all spatial data involves recording something concerning time, space and some "attribute". Different forms of collecting information will lead to different choices about which element is emphasized. Sinton (1978) described some of the basic choices that lie behind most map sources. In monitoring certain processes, such as tides and river flow, it is common to record the height of some body of water at a specific gauge location. These heights can be tracked as a continuous trace on analog equipment, or discretized to readings at specific times. For this kind of data, the spatial location is the least flexible. The measurements are not comparable if the gauge is relocated to another position. The time also serves as a form of control, and the height attribute is the element actually measured. These gauges are not the type of data recorded on most maps. If they are used on a map, the fixed locations must be used to estimate a new distribution that emphasizes the spatial component. To do this, typically, the time element must be reduced in prominence. For instance, it is common to show Mean Low or Mean High tides as a contour, interpolated between the heights estimated at each gauge. Sometimes tide-coordinated photography can assist in this process by timing the photograph at the time of estimated tides. In any case, the data are transformed from a form of temporal control with measurement of height to a measurement in the horizontal plane with time more fixed and less prominent.

In land cover mapping, the standard approach has fixed the time component by mapping a "snapshot". Sinton discussed the two distinct approaches to land cover mapping characterized by one focused on the locations, and the other on the attribute. In traditional field survey techniques (dating back to the early part of this century), it is common to go into the field (or into the photo interpretation process) with a specific set of classes to map. The mapping process involves distinguishing objects on the basis of this key, and drawing boundaries to divide them. The measurements are actually of the boundaries. In the other approach, some set of spatial objects is taken as the unit of analysis - most typically an exhaustive set of pixels. With some basic measurements for these objects, a transformation is sought that matches the desired classification as closely as possible. Due to cartographic conventions and expectations established long ago, the remotely sensed products are often filtered and smoothed to make them appear to look like results of the former method. These procedures actually confuse the basis for measurements. In practice, then, most land cover mapping is performed under a relatively consistent set of conventions, despite divergences in the technology.

A land cover map portrays a standardized set of exhaustive categories covering some region at a fixed time. Although some classifications permit categories of "mixed" types, the basic convention is that each place on the map belongs to one and only one category. In such a map, there are three kinds of error, though they will be inevitably entwined. All error in a map will consist of identifying the wrong category at a particular point. There are distinct ways in which this may occur, just as there are distinct forms of change (see above). First, the map can be of the wrong time, or of variable time across its surface. However, the principal indication that if it is of the wrong time is through the other two forms of error. Some errors in category are spatial in character. Since the distinctions form discrete boundaries, these boundaries could be mislocated through a variety of processes. These errors will be considered as "positional", even though some of them are clearly relative to such problems as "fuzzy" boundaries that involve the classification process as well. Finally, other errors occur basically outside the spatial framework of the data. These involve misclassification of whole objects. These will be called "attribute" error.

4.2.1 Positional Error

The basic model of a land cover map relies upon identifying each place in the landscape as a corresponding place in the model. This process involves a simplification or generalization. One aspect of that simplification is that the concept of "point" is not really as dimensionless as the mathematical model of a point. If this spatial transformation is not carried out correctly, then the model will make incorrect statements about some places. Since the standard is that the categories are different, not all differences in position will be perceptible. If a land cover map is misregistered to the north, then the differences will only appear along the north and south sides of each clump of points (polygon) with the same classification. In a vector representation, classifications are bounded by boundaries. These boundaries enclose polygons, regions determined to be sufficiently uniform to merit inclusion in the particular class. This decision involves many components, some of them dealing with the attribute and classification issues, and some of them positional. A boundary in some cases may be perfectly sharp, reflecting a clear change in the phenomenon, as in the transition from a farm field to a woodlot, or from a lake to a sand beach (at a specific time). Other transitions may be quite imperceptible, as in the gradation from a dry forest to a forested wetland.

One element of classifying land cover involves "inclusions", small isolated pockets of another class enclosed in a larger matrix. Many ecological processes depend on the texture of these relationships. As Goodchild et al. (1992) points out, in a vector representation, the edges may actually have a higher chance of representing the correct local conditions. In a raster representation, positional effects are just as inescapable. Traditional photointerpretation integrates a spatial component into the classification step, and produces a vector map. The raster method, by classifying pixels independent of their neighbors (as usually done), then forces a subsidiary spatial filtering step to simplify the spatial structure of the results. This step is usually performed using the final classification, not the original spectral values.

Positional error can be divided into a number of categories. The"observational" and "inherent" error relate to error associated with the methods or to the landscape. While this partitioning deals with cause, other divisions deal with what can be done to model and remove certain errors. An error that can be modeled is systematic. The most simple systematic positional errors involve misregistration of a map (e.g., offset or rotated). Systematic errors often can be removed in quality control steps, if tests are performed and analyzed appropriately. Most statistical models leave the residual error in the category of "random", or invidual perturbations. In some cases, it is important to recognize that these errors can come from decidedly distinct populations. Some large errors (as in incorrectly locating a registration tick mark) should be termed "blunders" and dissociated from the kinds of errors more usually considered random. Positional errors of the random nature can derive from the sources, listed in section 4.1.

Positional accuracy is best discussed within the framework of measurement by thinking of position as a combination of two measurements (x,y), representing the easting and northing of a pair of UTM coordinates, respectively. Each measurement is subject to error. For example, if 100 people were asked to measure the coordinates of a road intersection on a topographic map, the results would show variation in both coordinates. In practice, it is likely that the variation in both coordinates would follow a normal distribution or bell curve, with most measurements clustered, and a few extremes in both positive and negative directions. The amount of variation is likely to be similar in both coordinates, and, moreover, errors are likely to be uncorrelated, in the sense that the average of all 100 measurements will be very close to the true location. If these assumptions are true, and there is no obvious reason why they should not be, the errors in both coordinates can be visualized as a three-dimensional bell, or circular normal distribution.

Several statistics of the circular normal distribution are in common use to describe positional accuracy. Perhaps the most commonly used is the Circular Map Accuracy Standard, or CMAS, defined as the 90th percentile of the circular normal distribution, or 2.146 times its standard deviation. More conceptually, it forms a circle about the true location of the point, within which the observed location is expected to lie 90% of the time. Using the example of the topographic map, it might turn out that 90% of the 100 people determined the road intersection's coordinates to within 0.5mm of their true location at the scale of the map, leaving 10 people with positional errors of more than 0.5mm. National map accuracy standards require a CMAS of 1/40 inch, or 0.64mm, while a typical CMAS for a map digitizer is 1mm.

CMAS is a well-defined statistic, and useful when positional accuracy is assessed based on a sample of points. For general purposes, however, it is necessary to use a less rigorous approach. Unless otherwise specified in this section of the report, the term "positional accuracy" should be interpreted as a linear measure approximately equal to the CMAS, and based on the same assumptions, but not implying the same rigor of definition.

CMAS can be used to describe the positional accuracy of a point. The accuracy with which two points can be positioned relative to each other depends on the positional accuracy of both. If it is possible to assume that errors in the two points are independent, then a simple calculation can be used to determine the error in relative position, by taking the square root of the sum of the squares of the two positional accuracies. For example, if one point has a positional accuracy of 1mm and a second point has a positional accuracy of 2mm, a rough estimate of the error in their relative positions is:

$$\sqrt{1^2 + 2^2} = \sqrt{5} = 2.236 \text{ mm}$$

Most of the information in a land cover map does not consist of well-defined points that can be measured with CMAS or similar approaches. Some positional errors in the whole database, such as registration, can be estimated from those of a few well-defined points in the whole database. However, the positional errors in boundaries of classes are not totally captured by such a measure. The issue of “fuzzy boundaries” (not to be confused with a more general “fuzzy set” theory for land cover maps) creates position-like error from what is really a difficulty in carrying out the classification system for attributes in the environmental conditions of ecotones or transition zones.

4.2.2 Attribute Error

In contrast to positional error, attribute error involves primarily the taxonomy of the classification, not the spatial expression of the land cover map. Of course, these two components are intermingled, so it may be hard to distinguish them. Some of the most devastating attribute errors involve difficulties with time and change. In making a land cover map, it is common to use multiple sources, such as field samples and satellite images, photo keys, and aerial photographs. If these derive from different dates, a true change may be designated as an error.

In most procedures to construct a land cover map, a major source of error is connecting the source materials to the classification system. When an unsupervised classifier, the problem arises when regrouping and naming the clusters. When a supervised classification, the problem arises when selecting and naming the supervised training samples as representative of the class.

The most classic attribute error will distinguish a particular spatial object (i.e., one spectrally distinct from its neighbors) and then classify it in the wrong category. These mistakes occur when, for example, plowed fields are classified as developed because the soil is spectrally confused with concrete.

In other cases, the idea that there is a single classification may be the problem. Land cover may be mixed, down below the level of resolution in the sensing system or the mapping filter or mixed above the aerial perception of the field observer. Land cover maps come with written or unwritten expectations about spatial texture, often phrased in terms of minimum mapping units. If the actual variability is smaller, then there are inclusions that are intentionally generalized into their surrounding zones.

Unlike positional error that occurs in the continuous metric of space (where distance provides a useful measure of error), attribute error is often discrete. Though there might be applications in which classes are at least partially ordered on some scale, land cover classes are fundamentally unordered. A misclassification matrix (see below) must be used to represent the amount of error between each pair of classes. When there are a large number of classes, the misclassification matrix (increasing as the square of the number of categories) is very large, and often, as a result, rather sparse. Due to the size of the matrix, it is unreasonable to attempt to estimate the quantity in each cell with the same level of confidence. Sampling efforts would be misapplied in the case of some rare events. A spatial stratification is more likely to estimate prevalence in a cost-effective manner. It is also true that some categories are less likely to be involved in errors, either because they are spatially distinct or because they are distinct taxonomically (i.e., spectrally).

Current procedures used to estimate classification accuracy are directed towards the centers of homogeneous polygons. By avoiding errors associated with the positional accuracy of the borders, the result is focused on attributes, but at the expense of understanding other types of error. A spatial framework for accuracy estimation should ensure that the various sources of error will be included in the assessment.

In other disciplines, there has been substantial development of the statistical procedures to treat cross-classifications such as these error matrices. Notice that a change detection analysis produces a similar matrix of cross-classification. However, there are many difficulties in applying the procedures directly from other disciplines. Spatial data does not have discrete "events" or cases, because the land cover map is an exhaustive classification of a whole region. Rather than the traditional sampling error based on underlying population prevalence, spatial error is much more attuned to local aberrations and autocorrelations. Without the discrete metric of a "case", the Poisson sampling model simply may not address the kind of error likely in spatial classifications. Hence, the fit provided in most estimation procedures does not relate to the actual error circumstances. Yet, the analytical perspective of these techniques, in decomposing the effects, may be useful in change detection studies, if they can be estimated with confidence.

5.0 ACCURACY IMPLICATIONS OF IMAGE PROCESSING AND CHANGE DETECTION METHODS

Successful remote sensing change detection, especially those changes in uplands and wetlands in the coastal regions, requires careful attention to sensor systems, environmental characteristics, and geodetic control. Failure to understand the impact of the various parameters on the change detection process can lead to inaccurate results. Ideally, the remotely sensed data used to perform C-CAP change detection are acquired by a remote sensor system which holds the following factors constant: temporal, spatial (including look angle), spectral, and radiometric. It is instructive to review each of these parameters and identify how they impact the accuracy of C-CAP remote sensing change detection projects. The issues involved in change detection of uplands and wetlands using satellite remote sensor data and issues in common with aerial photographic change detection are discussed, and issues specific to aerial photographic interpretation of submerged rooted vascular plants (SRV) are addressed in section 5.2.

5.1 Satellite Remote Sensing

5.1.1 System Considerations

There are two important temporal resolutions which should be held constant when performing coastal change detection using multiple dates of remotely sensed data. First, the data should be obtained from a sensor system which acquires data at approximately the same time of day (e.g., Landsat Thematic Mapper data are acquired before 9:45 am for most of the conterminous United States). This eliminates diurnal sun angle effects which can cause anomalous differences in the reflectance properties of the remotely sensed data. Second, whenever possible it is desirable to use remotely sensed data acquired on anniversary dates, e.g., October 1, 1988 and October 1, 1993. Using anniversary date imagery removes seasonal sun angle differences which can make change detection difficult and unreliable. Usually precise anniversary date imagery is not available. The determination of acceptable near-anniversary dates then depends on local and regional factors such as phenological cycles and annual climatic regimes.

Accurate spatial registration of at least two images is essential for digital change detection. Ideally, the remotely sensed data are acquired by a sensor system which collects data with the same instantaneous-field-of-view (IFOV) on each date. For example, Landsat Thematic Mapper data collected at 30 x 30 m spatial resolution (Table 1) on two dates are relatively easy to register to one another.

Table 1 Selected Satellite Remote Sensing System Characteristics

Remote Sensor System	Spectral Resolution In Micrometers	Spatial Resolution In Meters	Temporal Resolution In Days	Radiometric Resolution In Bits
Landsat MSS 1, 2, 3	Band 1 (.50-.60)	80 X 80	18	7
Landsat MSS 1, 2, 3	Band 2 (.60-.70)	80 X 80	18	7
Landsat MSS 1, 2, 3	Band 3 (.70-.80	80 X 80	18	7
Landsat MSS 1, 2, 3	Band 4 (.80-1.1)	80 X 80	18	7
Landsat Thematic Mapper 4, 5	Band 1 (.45-.52)	30 X 30	16	8
Landsat Thematic Mapper 4, 5	Band 2 (.52-.60)	30 X 30	16	8
Landsat Thematic Mapper 4, 5	Band 3 (.63-.69)	30 X 30	16	8
Landsat Thematic Mapper 4, 5	Band 4 (.76-.90)	30 X 30	16	8
Landsat Thematic Mapper 4, 5	Band 5 (1.55-1.75)	30 X 30	16	8
Landsat Thematic Mapper 4, 5	Band 7 (2.08-2.35)	30 X 30	16	8
Landsat Thematic Mapper 4, 5	Band 6 (10.4-12.5)	120 X 120	16	8
SPOT HRV XS	Band 1 (.50-.59)	20 X 20	pointable	8
SPOT HRV XS	Band 1 (.61-.68)	20 X 20	pointable	8
SPOT HRV XS	Band 3 (.79-.89)	20 X 20	pointable	8
SPOT HRV PAN	Pan (.51-.73)	10 X 10	pointable	8

Geometric rectification algorithms (Jensen, 1986; Novak, 1992) are used to register the images to a standard map projection (Universal Transverse Mercator - UTM, for most U.S. projects). Rectification should result in the two images having a root mean square error (RMSE) of $\leq \pm 0.5$ pixel. RMSE > ± 0.5 pixel may result in the identification of spurious areas of change between the two data sets.

It is possible to perform change detection using data collected by two different sensor systems with different IFOVs, e.g. Landsat TM data (30 x 30 m) for date 1 and SPOT HRV data (20 x 20 m) for date 2. In such cases, it is necessary to decide upon a representative minimum mapping unit (e.g. 20 x 20 m) and then resample both data sets to this uniform pixel size. This does not present a significant problem as long as one remembers that the information content of the resampled data can never be greater than the IFOV of the original sensor system (i.e. even if the Landsat TM data are resampled to 20 x 20 m pixels, the information was still acquired at a 30 x 30m nominal resolution and one should not expect it to be able to extract additional spatial detail in the dataset).

Some remote sensing systems like SPOT collect data at off-nadir look angles as much as $\pm$ 20° (Table 1), i.e. the sensors obtain data of an area on the ground from an 'oblique' vantage point. Two images with significantly different look angles can cause problems when used for change detection purposes due to bidirectional reflectance factor (BRDF) difference. For example, consider a maple forest consisting of very large, randomly spaced trees. A SPOT image acquired at 0° off-nadir will look directly down upon the 'top' of the canopy. Conversely, a SPOT image acquired at 20° off-nadir will record reflectance information from the 'side' (near the top) of the canopy. Differences in BRDF from the two datasets can cause spurious change detection results. Therefore, the data used in a remote sensing digital change detection should be acquired with approximately the same look angle whenever possible.

A fundamental assumption of digital change detection is that there should exist a difference in the spectral response of a pixel on two dates if the biophysical materials within the IFOV have changed between dates. Ideally, the spectral resolution of the remote sensor system is sufficient to record reflected radiant flux in spectral regions that best capture the most descriptive spectral attributes of the object. Unfortunately, different sensor systems do not record energy in exactly the same portions of the electromagnetic spectrum, i.e. bandwidths (Table 1). For example, the Landsat multispectral scanner system (MSS) records energy in four relatively broad bands, SPOT HRV sensors record in three relatively coarse multispectral bands and one panchromatic band, and the Thematic Mapper in six relatively narrow optical bands and one broad thermal band (Table 1). Ideally, the same sensor system is used to acquire imagery on multiple dates. When this is not possible, the analyst should select bands which approximate one another. For example, SPOT bands 1 (green), 2 (red), and 3 (near-infrared) can be considered with Landsat TM bands 2 (green), 3 (red), and 4 (near-infrared) or Landsat MSS bands 1 (green), 2 (red), and 4 (near-infrared). Many of the change detection algorithms to be discussed do not function well when bands from one sensor system do not match those of another sensor system, e.g. utilizing the Landsat TM band 1 (blue) with either SPOT or Landsat MSS data is not wise.

An analog to digital conversion of the satellite remote sensor data usually results in 8-bit brightness values with values ranging from 0 through 255 (Table 1). Ideally, the sensor systems collect the data at the same radiometric precision on both dates. When the radiometric resolution of data acquired by one system (e.g., Landsat MSS 1 with 7-bit data) are compared with data acquired by a higher radiometric resolution instrument (e.g., Landsat TM with 8-bit data) then the lower resolution data (e.g., 7-bit) should be 'decompressed' to 8-bits for change detection purposes. However, the precision of decompressed brightness values can never be better than the original, uncompressed data.

5.1.2 The Preferred C-CAP Satellite Sensor System

The Landsat Thematic Mapper (TM) is currently the primary sensor recommended for C-CAP image acquisition and change analysis for all land cover except submerged aquatic vegetation. A Landsat TM image, although its spatial resolution is not as high as that of a SPOT satellite or aircraft MSS image, is generally better suited to the C-CAP mission and less expensive to acquire and process for large-area coverage. Compared to SPOT imagery, TM has better spectral resolution and specific spectral bands that are more applicable to wetlands delineation (bands 5 and 7). In addition, TM is preferred over SPOT because TM has collected data for a longer time (since 1982 as opposed to SPOT since 1986) and because many TM scenes of the United States coastal regions were systematically collected on a routine basis.

There are advantages and disadvantages to using other sensors. Aircraft multispectral scanners are more expensive and complex to utilize over large regions (Jensen et al., 1987). However, good algorithms are now available for georeferencing, and in certain cases (e.g., when higher spectral or spatial resolution is needed and when unfavorable climactic conditions for satellite sensors exist) aircraft sensors may be optimum. The SPOT sensor has a greater temporal coverage because the satellite can collect data off-nadir. However, if off-nadir SPOT imagery is used for C-CAP change analyses, the data must be normalized to compensate for different look angles that may preclude pixel-to-pixel spectral-change analysis. Nevertheless, SPOT imagery may be a reasonable alternative in certain areas due to cloud cover or other impediments to TM data availability. C-CAP should remain flexible in order to take advantage of new sensors and other technologies that become operational during the lifetime of the program.

5.1.3 Environmental Considerations

Failure to understand the impact of various environmental characteristics on the remote sensing change detection process can also lead to inaccurate C-CAP results. When performing change detection it is desirable to hold environmental variables as nearly constant as possible. Specific environmental variables and their potential impacts are described below.

Ideally, there should be no clouds, haze, or extreme humidity on the days remote sensing data are collected. Even a thin layer of haze can alter spectral signatures in satellite images enough to create the false impression of spectral change between two dates Obviously, 0% cloud cover is preferred for satellite imagery and aerial photography. At the upper limit, cloud cover > 20% is usually unacceptable. It should also be remembered that clouds not only obscure terrain but the cloud shadow also causes major image classification problems. Any area obscured by clouds or affected by cloud shadow may impact the entire change detection process, severely limiting the utility of the final change detection product. Therefore, regional analysts must use good professional judgement in evaluating such factors as the criticality of the specific locations affected by cloud cover and shadow, and the availability of timely surrogate data for those areas obscured (e.g. perhaps substituting aerial photography interpretation for a critical area). Even when the stated cloud cover is 0%, it is advisable to 'browse' the proposed image on microfiche at the National Cartographic Information Center in each region to confirm that the cloud cover estimate is correct. Cooperators are also referred to the Aerial Photographers Clear Day Map, U.S. Department of Commerce, Environmental Data Service for monthly probabilities of clear air.

Assuming no cloud cover, the use of anniversary dates helps to ensure general, seasonal agreement between the atmospheric conditions on the two dates. However, if dramatic differences exist in the atmospheric conditions present on the "n" dates of imagery to be used in the change detection process, it may be necessary to remove the atmospheric attenuation in the imagery. Two alternatives are available. First, sophisticated atmospheric transmission models can be used to correct the remote sensor data if substantial in situ data are available on the day of the overflights. Second, an alternative empirical method may be used to remove atmospheric effects. A detailed description of one empirical method of image to image normalization is found in "Guidance for Regional Implementation," (Dobson et al., 1993).

Ideally, the soil moisture conditions should be identical for the *n* dates of imagery used in a change detection project. Extremely wet or dry conditions on one of the dates can cause serious change detection problems. Therefore, when selecting the remotely sensed data to be used for change detection it is very important not only to look for anniversary dates, but also to review precipitation records to determine how much rain or snow fell in the days and weeks prior to remote sensing data collection.

When soil moisture differences between dates are significant for only certain parts of the study area (perhaps due to a local thunderstorm), it may be necessary to stratify (mask out) those affected areas and perform a separate analysis which can be added back in the final stages of the project.

Vegetation grows according to seasonal and annual phenological cycles. Obtaining near-anniversary images greatly minimizes the effects of wetland seasonal phenological differences which may cause spurious change to be detected in the imagery. One must also be careful about two other factors when dealing with man-made upland seasonal agricultural crops. First, many monoculture crops (e.g. corn) normally are planted at approximately the same time of year. A month lag in planting date between fields having the same crop can cause serious change detection error. Second, many monoculture crops are comprised of different species (or strains) of the same crop which can cause the crop to reflect energy differently on multiple dates of anniversary imagery. These observations suggest that the analyst must know the biophysical characteristics of the vegetation as well as the cultural land-tenure practices in the study area so that imagery which meets most of these characteristics can be selected for change detection. Many other factors also come into play such as hydrological and climatological factors (i.e., a wet versus dry year), which are recommended to be taken into account to the extent possible.

The choice of image date is best determined by mutual agreement among remote sensing specialists, biologists, ecologists, and local experts. The selection of the acceptable window of acquisition will be made independently by participants in each region. No single season will serve for all areas because of substantial latitudinal variation extending from temperate to tropical regions. For example, coastal marshes in the Mid Atlantic Region are best inventoried from June through October while submerged habitats in southern Florida may be inventoried best in November. Even within regions, some cover types will be more easily distinguished in different seasons. For example, in the Caribbean, estuarine seagrasses can be detected best in early January, yet marine seagrasses can be detected best in May or June. The best time of the year to acquire photography is during the season of maximum biomass or flowering of dominant species, considering the phenologic overlap for the entire community. This is June for submerged vegetation of the Pacific northwest and Atlantic northeast, April and May in North Carolina, and September for most of the other species in the eastern U.S. Technically, these vegetation patterns could be monitored throughout the year, but cost limitations usually limit the analyst to a single date.

Tidal stage is a crucial factor in satellite image scene selection and the timing of aerial surveys. Ideally, tides should be constant between time periods, but this would rule out synoptic satellite sensors since tide stages are not synchronized within a region or even with a single image. Alternatively, analysts should avoid selecting the highest tides and should take into account the tide stages occurring throughout each scene. Tidal effect varies greatly among regions. In the Northwest, for example, when all of the temporal, atmospheric, and tidal criteria are taken into account the number of acceptable scenes may be quite small. In some regions it may be necessary to seek alternative data such as SPOT satellite data, aerial photographs, or other land cover databases. For most regions, mean low tide (MLT) or lower will be preferred, one or two feet above MLT will be acceptable, and three feet or more will be unacceptable. Ideally, tides for aerial surveys should approach low tide as predicted in NOAA, National Ocean Survey (NOS) tide tables, but optimal visualization of the subtidal bottom depends on water clarity as well as depth. In addition, some estuaries experience significant tidal lag which must be considered even within a single satellite scene.

5.1.4 Image Processing

With the classification scheme developed and the appropriate remote sensor data selected, it is possible to process the data to extract upland and wetland change information. This involves geometric and radiometric correction, classification (if necessary) selection of an appropriate change detection algorithm, creation of change detection products, and error evaluation.

5.1.4.1 Rectification

Georeferencing (spatial registration of a remotely sensed image to a standard map projection) is a necessary step in digital change detection and cartographic representation. The following C-CAP recommendations should be followed when rectifying the base image to a standard basemap:

- Geocoded base TM images can be purchased if preferred by regional analysts. However, participants should be aware that some analysts have reported undocumented variations in commercial products that can lead to poor registration in certain regions, especially where local relief requires substantial terrain correction. Additional registration may be necessary to achieve the C-CAP standard precision of RMSE$<\pm 0.5$ pixel. Therefore, it is recommended that each regional project perform its own base image to map rectification using the radiometrically corrected but not geocoded data.

- Ground control points (GCPs) used to compute rectification transformation coefficients should be relatively static features in the landscape (e.g. road intersections) and, whenever possible, based on new Global Positioning System (GPS) measurements taken in the field. When GCPs are digitized from USGS 7.5' (1:24,000) maps, analysts should use the marginal information and updates available to improve location of the control points. GCPs should be extracted from mylar copies of the USGS maps whenever possible to minimize system produced digitizing error. Traditional paper maps expand and contract with changes in relative humidity and should not be used for digitizing GCPs.

• C-CAP recommends the use of the current NAD '83 North American Datum. Unfortunately, most existing map series are based on the NAD '27 datum. NAD '27 will be acceptable on a region by region basis until published maps based on NAD '83 are universally available.

• In all but the flattest coastal regions, terrain correction of imagery may be necessary to reduce image distortion caused by local relief.

• The required coordinate system is Universal Transverse Mercator (UTM). If another coordinate system is used (e.g. State Plane), it is the responsibility of the regional analyst to provide complete documentation and conversion equations.

• It is the responsibility of the regional analyst to understand (or seek advice concerning) the variety of rectification resampling algorithms (e.g., bilinear interpolation, nearest neighbor, cubic convolution) and their impact on the data. Nearest neighbor resampling is recommended.

Rectification of an earlier date (T_{b-1}) or later date (T_{b+1}) to the base image (T_b) can be accomplished in several ways. The primary concern is to accomplish the most exact co-registration of pixels from each time period and thus reduce a potentially minimum recommendations and requirements:

• Geocoded and terrain-corrected TM data can be ordered from commercial vendors. Two separate images can be overlaid according to like coordinates, but this technique may introduce error if prior geocoding was not precisely the same in both images. The regional analyst has no control in this process, but if high precision is accomplished by the vendor, the analyst can significantly reduce image processing effort at the regional facility.

• The regional analyst can geocode the image to geographic coordinates as was done with the base image. If this technique is adopted, it is important to use the identical GCPs and resampling algorithm used to rectify the base image.

• For multiple images, the preferred technique is to rectify non-geocoded images directly to the geocoded base image. This technique may have the advantage of reducing or better controlling co-registration error among images. Selection and consistency of control points and rectification algorithms are important to the success of this technique. Cubic convolution algorithms normally yield the most precise spatial fit, but cubic convolution and bilinear interpolation algorithms suffer from the disadvantage of averaging pixel brightness values. Nearest neighbor algorithms are spatially less precise, but they offer the advantage of retaining pixel brightness values through the processes of rectification and registration.

5.1.4.2 Radiometric Normalization of Multi-Date Images

The use of remotely sensed data to classify coastal and upland land cover on individual dates is contingent upon there being a robust relationship between remotely sensed brightness values (BVs) and actual surface conditions. However, factors such as sun angle, Earth/sun distance, detector calibration differences between the various sensor systems, atmospheric condition, and sun/target/sensor geometry (phase angle) will also affect pixel brightness value. Differences in direct beam solar radiation due to variation in sun angle and Earth/sun distance can be calculated accurately, as can variation in pixel BVs due to detector calibration differences between sensor systems. Removal of atmospheric and phase angle effects requires information about the gaseous and aerosol composition of the atmosphere and the bi-directional reflectance characteristics of elements within the scene. However, atmospheric and bi-directional reflectance information are rarely available for historical remotely sensed data. Also, some analysts may not have the necessary expertise to perform a theoretically based atmospheric path radiance correction on remotely sensed data. Hence, it is suggested that a relatively straightforward 'empirical scene normalization' be employed to match the detector calibration, astronomic, atmospheric, and phase angle conditions present in a reference scene.

Image normalization reduces pixel BV variation caused by non-surface factors so variations in pixel BVs between dates can be related to actual changes in surface conditions. Normalization enables the use of image analysis logic developed for a base year scene to be applied to the other scenes. This can be accomplished using techniques pioneered by personnel of the U.S. Bureau of Land Management (Eckhardt et al., 1990). Image normalization is achieved by developing simple regression equations between the brightness values of 'normalization targets' present in the base scene (Date T_b) and the scene to be normalized (Date $T_{b\pm1}$). Normalization targets are assumed to be constant reflectors, therefore any changes in their brightness values are attributed to detector calibration, astronomic, atmospheric, and phase angle differences. Once these variations are removed, changes in BV may be related to changes in surface conditions.

Acceptance criteria for potential 'normalization targets' are (Eckhardt et al., 1990):

- The targets must be at approximately the same elevation as the land cover of primary interest within the scene. Most aerosols in the atmosphere occur < 1000 m above ground level (AGL). Selecting a mountain top normalization target, thus, would be of little use in estimating atmospheric conditions near sea level. Although C-CAP projects are on the coast, many regions include areas of substantial local relief.

• The targets should contain only minimal amounts of vegetation. Vegetation spectral reflectance can change over time due to environmental stresses and plant phenology. Good targets include bare soil fields and deep, non-turbid water bodies.

• The targets must be on relatively flat terrain so that incremental changes in sun angle between dates will have the same proportional increase or decrease in direct beam sunlight for all normalization targets.

• The normalization targets should have approximately the same texture over time. Changing textural patterns indicate variability within the target which could mean that the reflectance of the target as a whole may not be constant over time. For example, a mottled pattern on what had previously been a uniformly gray dry lake bed indicates changing surface moisture conditions, which would eliminate the dry lake bed from consideration as a normalization target.

Besides selecting targets that meet the four conditions listed above, efforts are often made to select a set of wet and dry normalization targets exhibiting a range of pixel brightness values. The greater the time period between the base image (e.g. 1980) and an earlier year image (e.g. 1972), the more difficult it is to locate unvegetated, dry normalization targets. For this reason, analysts sometimes use man-made, 'pseudo-invariant' normalization targets such as concrete, asphalt, rooftops, parking lots, and roads when normalizing historical remotely sensed data (Schott et al., 1988; Caselles and Garcia, 1989). Hall et.al, (1991) also suggest that the members of the radiometric control sets do not have to be the same pixels from image to image in contrast to geometric control points for spatial image rectification, which are composed of identical elements in each scene. Furthermore, they suggest that 'using fixed elements inevitably requires manual selection of sufficient numbers of image-to-image pairs of suitable pixels, which can be prohibitively labor intensive, particularly when several images from a number of years are being considered.'

The mean brightness values of the base image targets are regressed against the mean brightness values of the Date T_{b+1} targets for the *n* bands used in the classification of the remote sensor data (e.g. TM bands 2, 3, and 4). The slope and y-intercept of the *n* equations are then used to normalize the $T_{b\pm1}$ Landsat TM data to the T_b Landsat TM data. Each regression model contains an additive component (y-intercept) that corrects for the difference in atmospheric path radiance between dates, and a multiplicative term (slope) that corrects for the difference in detector calibration, sun angle, Earth/sun distance, atmospheric attenuation, and phase angle between dates.

It is customary to first normalize the remote sensor data and then perform image rectification (using nearest-neighbor resampling if image classification is to take place). These data are then ready for individual date classification or the application of various multi-image change detection algorithms. Most studies that attempt to monitor biophysical properties such as vegetation biomass, chlorophyll absorption, health, and other biophysical properties require atmospheric correction.

5.1.5 Change Detection Algorithms

The selection of an appropriate change detection algorithm is very important. First, it will have a direct impact on the type of image classification to be performed (if any). Second, it will dictate whether important "from-to" information can be extracted from the imagery. C-CAP requires that the "from-to" information be readily available in digital form suitable for geographic analysis and for producing maps and tabular summaries. At least seven change detection algorithms are commonly used by the remote sensing community, including:

1. Change Detection Using Write Function Memory Insertion
2. Multi-date Composite Image Change Detection
3. Image Algebra Change Detection (Band Differencing or Band Ratioing)
4. Post-classification Comparison Change Detection
5. Multi-date Change Detection Using A Binary Mask Applied to Date 2
6. Multi-date Change Detection Using Ancillary Data Source as Date 1
7. Manual, On-screen Digitization of Change
8. Other methods (e.g., regression-based)

The C-CAP protocol (Dobson et al., 1993) recommends the use of methods 4 primarily for submerged land and 5 primarily for uplands and wetlands. Accuracy assessment implications of these two methods are discussed in the following sections.

5.1.5.1 Post-Classification Comparison Change Detection

This is the most commonly used quantitative method of change detection and may be used in regional C-CAP projects under certain conditions. It requires rectification and classification of each of the remotely sensed images. These two maps are then compared on a pixel by pixel basis using a 'change detection matrix' to be discussed. Unfortunately, every error in each individual date classification map will also be present in the final change detection map (Rutchey and Velcheck, in press). Therefore, it is imperative that the individual classification maps used in the post-classification change detection method be as accurate as possible (Augenstein et al., 1991; Price et al., 1992).

Post classification change detection is the method of choice with aerial photographic data. The C-CAP objective of site specific change detection places greater emphasis on accuracy and precision of spatial data than required in one-time inventories or regional summaries of change. Unless historical photography meets C-CAP requirements and is supported by surface level data, historical presence or absence of SRV at a given location may remain an open question. Methodology for monitoring site specific change on a statewide or regional scale is a recent development (Ferguson et al., 1993).

A change detection matrix, constructed similarly as an error matrix, can be developed with T_b at the top (instead of reference data in the error matrix) and $T_{b\pm1}$ on the side (instead of classified image data in the error matrix). This change detection matrix can then be used to summarize specific "from-to" classes for display in a change detection map. In this way, (n^2-n) off-diagonal possible erroneous change classes can be generated for display on the change detection map.

Post-classification comparison change detection is widely used and easy to understand. When conducted by skilled image analysts it represents a viable technique for the creation of C-CAP change detection products. Advantages include the detailed “from-to” information and the classification map for each year. Unfortunately, the accuracy of the change detection is heavily dependent on the accuracy of the two separate classifications. The post-classification comparison is recommended for C-CAP regional projects when different sensors are involved, when two separate organizations are classifying the same region at different times, and when change detection is based on aerial photographic interpretation.

5.1.5.2. Multi-Date Change Detection Using A Binary Change Mask Applied to Date 2

This method of change detection is highly recommended for C-CAP regional projects. First, the analyst selects the base image (Date T_b). Date $T_{b\pm1}$ may be an earlier image (T_{b-1}) or a later image (T_{b+1}). A traditional classification of Date T_b is performed using rectified remote sensor data. Next, one of the bands from both dates of imagery is placed in a new dataset. The two band dataset is then analyzed using various image algebra functions (e.g. band ratio, image differencing, even principal components) which produces a new image file. The analyst usually selects a "threshold" value to identify areas of change and no-change in the new image. The change image is then recoded into a binary mask file, consisting of areas which have changed between the two dates. Great care must be exercised when creating the "change/no-change" binary mask (Dobson and Bright, forthcoming). The change mask is then overlaid onto Date $T_{b\pm1}$ of the analysis and only those pixels which were detected as having changed are classified in the Date $T_{b\pm1}$ imagery.

This method may reduce change detection errors (omission and commission) and provide detailed "from-to" change class information. The technique reduces effort by allowing analysts to focus on the small amount of area that has changed. In most regional projects, the amount of actual change over a five-year period will probably be no greater than 10% of the total surface area. The method is complex, requiring a number of steps, and the final outcome is dependent on the quality of the "change/no-change" binary mask used in the analysis and the reclassification of the masked area for $T_{b\pm1}$. Nevertheless, this is a highly recommended C-CAP change detection algorithm.

5.1.6 Classification Algorithms

The previous section indicated that two of the eight most commonly used change detection algorithms are acceptable for C-CAP regional projects:

- Post-Classification Comparison
- Change Detection Using A Binary Change Mask Applied to Date $T_{b\pm1}$

Each of these requires a complete pixel by pixel classification of one date of imagery and, at least, a partial classification of an additional date. Regardless which approach is used, the end result is a complete classification for each time period and a matrix of changes by class ("from" and "to"). Hence, it is instructive to review the C-CAP approved image classification logic which may be used in the regional projects.

5.1.6.1 Supervised and Unsupervised Image Classification Logic

The primary reason for employing digital image classification algorithms is to reduce human labor and improve consistency. It is expected that regional analysts will have sufficient expertise to assess the advantages of alternative classification algorithms and to recognize when human pattern recognition and other types of intervention are necessary. In practice, it may be necessary to employ a suite of algorithms including both supervised and unsupervised statistical pattern recognition approaches. Currently maximum-likelihood classifiers often serve as a good first step, but new statistical approaches are being developed and implemented on a routine basis (Jensen et al., 1987; Hodgson and Plews, 1989; Foody, 1992). It is important for analysts to remain flexible with regard to procedures and algorithms.

In an **unsupervised** classification, the computer is allowed to query the multispectral properties of the scene based on user specified criteria and to identify *x* mutually exclusive clusters in n-dimensional feature space (Chuvieco and Congalton, 1988). The analyst must then convert (label) the *x* spectral clusters into information classes such as those found in the C-CAP Coastal Land Cover Classification System.

In a **supervised** classification, the analyst 'trains' the classifier by extracting mean and co-variance statistics for known phenomena in a single date of remotely sensed data (Gong and Howarth, 1990). These statistical patterns are then passed to either a minimum-distance-to-means algorithm where unknown pixels are assigned to the class nearest in *n*- dimensional feature space, or to a maximum-likelihood classification algorithm which assigns an unknown pixel to the class in which it has the highest probability of being a member. Great care must be exercised when selecting training samples (Mausel et al., 1990).

The most common operational technique for multispectral image digital interpretation is to perform an unsupervised signature extraction, augmented by supervised training, followed by a maximum likelihood classification of up to 5 bands. The resulting spectral radiance map is labelled by interpreters using field observation, aerial photography, and ancillary data. The resulting reclassified data are usually smoothed using a local operator (perhaps iterative) and a majority neighborhood rule. The results are very sensitive to operator selections in the clustering and merging phases of signature extraction and to the optimization strategies used in the software for the maximum likelihood classifier (maximum liklihood routines in different commercial packages are not always the same). EPA has developed a version of the maximum likelihood classifier that is roughly an order of magnitude faster than other methods, with no significant change in results from a full calculation (Weerackoon and Mace, 1990). C-CAP might consider its adoption. In addition to the care that should be taken during the labelling process (where most of the errors will occur), C-CAP should avoid multiple algorithms on different dates for the same area, as method produced changes will occur.

C-CAP should also consider segmenting the images into broad landscape categories prior to running the automated signature extraction and classification routines. Elevation, land use, and spatial frequency of land cover are all useful stratification and regionalization tools. In any case, the technique followed should be uniformly applied across all dates for a region.

Only training sites which were actually visited on the ground by experienced professionals or very clearly identified on large scale photographs by experts should be selected for extracting the multispectral statistical 'signature' of a specific class when performing a supervised classification or identifying clusters in an unsupervised classification. It is suggested that a minimum of five training sites per land cover class be collected. This creates a representative training set when performing supervised classification and makes labeling clusters much easier in an unsupervised classification. In addition to the image analysts, the field team should contain specialists in ecology, biology, forestry, geography, statistics, and other pertinent fields such as agronomy. Field samples should be stratified by land cover type and by various physical factors such as slope, elevation, vegetation density, species mix, season, and latitude. The polygonal boundary of all field sites should be measured using global positioning systems whenever possible, and the locational, temporal, and categorical information should be archived.

The collection of field training sites often requires multiple visits to the field. Some of the field sites may be used to train a classifier or label a cluster while a certain proportion of the field sample sites should be held back to be used for classification error assessment to be discussed.

The following materials are indispensable to a successful field exercise:

- Imagery geocorrected to a standard map projection
- Topographic Maps at 1:24,000 or largest available scale
- Global Positioning System (GPS)
- Aerial photographs

It is advisable to perform, at least, a cursory classification before initiating fieldwork. In this case, both raw and classified data should be taken to the field. The primary function of the cursory classification is to guide field workers in targeting the covers and signatures that are most difficult and confusing. Keep in mind that the vast majority of all cover will be easy to identify on the ground and on the imagery. Efficient use of field time as documented on field survey sheets provides for quick verification of easy cover types and maximum attention to difficult, unusual, and ecologically critical cover types.

5.1.6.2 Collateral Data for Training and Accuracy Assessment

There are many potential sources of collateral data including soil maps, NOAA coastlines (T-sheets), timber surveys, USGS digital line graphs, and digital elevation models (for elevation, slope, and aspect). These can be incorporated by masking, filtering, probability weighing, or inclusion in the signature file (Ryerson, 1989; Baker et al., 1991). Depending on the importance of each category, analysts may use certain categories to overrule others.

The National Wetlands Inventory (NWI) is a collateral database that may be of value when classifying wetlands. Regional analysts should incorporate NWI data to the maximum extent possible. NWI data are recognized as the most authoritative and complete source of wetlands land cover data (Wilen, 1990). However, NWI maps are not temporally synchronized in each region and are not in a digital format for many regions. An approach based on complementary use of NWI and imagery will be an asset to both C-CAP and NWI. At a minimum, NWI maps and/or digital data should be used to make defined training samples, to check intermediate results, and to aid in the final verification of the wetlands portion of the C-CAP maps.

Both primary data (field observations, imagery, monitoring network measurements) and interpreted data (maps, statistical summaries, reports) are useful. There are many catalogs for primary and interpreted data (i.e. the National Technical Information Service, the USGS Global Land Information System, the NASA/Global Change Master Directory, etc.). There are hundreds of federal, state (i.e. State Cartographers Office), and local (i.e. cadastral and regional planning systems) directories where detailed and summary information can be found. Sorting through this plethora of data sets is a laborious task, and mining the results can be disappointing. Imagery searches should start with the National Aerial Photography Program (NAPP) and Agricultural Stabilization and Conservation Service (ASCS) for aerial photography, including the National Archives for older imagery. Beyond that, universities, states, and private aerial survey firms all maintain high resolution photography. The list is only bounded by the searcher's energy and resources.

In addition to the obvious uses of USGS Land Use/Land Cover Data and USFWS NWI, a few programs within the federal government are worthy of special attention, but some have data confidentiality issues associated with their use. The National Agricultural Statistical Service (NASS) has data on agricultural land cover and use that would be potentially useful to C-CAP. However, the detailed field data sheets are protected, and serious data confidentiality issues need to be resolved before NOAA could profitably make use of these data sets. Also, within the Department of Agriculture, the U.S. Forest Service has detailed forest plot data on thousands of sites throughout the U.S. as part of the Forest Inventory and Assessment Program. These data are held as confidential, but they are not protected by special legislation, and arrangements may be made for their use. Additionally, the Soil Conservation Service maintains aerial photography (usually oblique, 35mm slides) at the DOA county offices as part of its responsibilities under the "swampbuster" and "sodbuster" legislation. These are very useful in documenting conversion of wetlands to agriculture. These slides are publicly available.

EPA also is conducting environmental monitoring surveys of ecosystems as part of the Environmental Monitoring and Assessment Program (EMAP). This is primarily a field survey program in which assessment of ecological condition is made from sampling indicators using a systematic sample with nodes spaced at approximately 27 km intervals. Local intensification of the grid may occur at selected regions, and contiguous coverage of land cover will be made through the Landscape Characterization and Landscape Ecology subgroups. Additionally, the Forest, Agriculture, Wetland, Great Lakes, and Nearcoastal Ecosystem subgroups may have important field observations for use by C-CAP. Access to the EMAP data will be on-line through the EMAP Information System, although it is possible to develop special data sharing arrangements.

EPA, NASA, and USGS are participating in the Global Change Research Program's (GCRP) North American Landscape Characterization Pathfinder Project (NALC), The NALC is producing Landsat MSS time series triplets (70's, 80's, and 90's) of all of North America and the Caribbean for remote sensing of change at an Anderson Level I (approximation).

Another data set is also being developed using Landsat TM (by EMAP, GAP, and others) in an automated, wall-to-wall classification of the U.S. Partners are welcomed, and the results will be available through the Earth Observing System Data and Information System (EOSDIS). Although this effort is at a much coarser resolution (categorically and cartographically), than C-CAP's, a substantial amount of ancillary data will be developed during the interpretation process, and some of the change conditions may also be relevant to C-CAP.

The intelligence community can make some of its data available for environmental purposes, under special circumstances. A task force of scientists, called the Environmental Task Force, has made some recommendations on the potential use of intelligence assets for environmental issues. Currently, some (obviously restricted) uses are permitted. The USGS/Reston, Virginia, can assist properly cleared personnel in obtaining access to these data. C-CAP should contact them for further discussions on applicability and access.

Finally, the whole issue of data access in the federal government is being addressed through a number of mechanisms. One of the most comprehensive is the Global Change Data and Information System (GCDIS). GCDIS will start with elements such as the NASA Master Directory and GLIS (which exist now) and add EODIS from NASA and implementations from each participating agency. NOAA is participating in this process through the Interagency Working Group on Data Management for Global Change and the GCRP Working Group, Data and Information. Through this effort, online search, browse, and ordering of imagery, model results, publications, and field data will be possible by the end of this decade, and in prototype by 1996. Many search tools, such as WAIS, the CIESIN Greenpages, WWW, and Gopher, are available now. It is recommended that C-CAP track this development and plan to use it in the future as part of their long range program.

5.2 Accuracy Assessment Issues Specific to Change Detection of Water and Submerged Land Based on Aerial Photography

The C-CAP Coastal Land Cover Classification identifies Marine/Estuarine Aquatic Beds, specifically, submerged Rooted Vascular Plants (SRV) of primary importance to be inventoried and placed in the C-CAP database (Klemas et al., 1993). Many of the steps discussed in Section 4 to monitor Uplands and Wetlands are pertinent to monitor SRV. However, there are significant differences which cannot be ignored, including:

- mapping SRV is primarily a photogrammetric task rather than a satellite task requiring an entirely different sensor system (aircraft, camera filter and film);
- aerial photography is not normally radiometrically (except for color balance between photographs) or geometrically corrected;
- time of day, sensor altitude, and flightline placement are very flexible, unlike fixed orbit satellite sensor systems;
- numerous environmental conditions can be considered (sea state, water clarity, water depth, low altitude atmospheric conditions), to optimize photography, and
- aerial photographs are in analog format.

These differences are so significant that it is instructive to focus on aerial photography of SRV. Some successes have been reported with satellite imagery and a number of other technologies. At the present time, these technologies may supplement and eventually they may replace aerial photography for change detection in SRV.

5.2.1 Metric Photography and Photographic Scale

The recommended film is Aerocolor 2445 color negative film. Second choices are Aerochrome 2448 color reversal film and Aerographic 2405 black and white negative film. We do not recommend infrared film for delineating SRV habitat except when SRV are intertidal. In our experience in North Carolina with tandem cameras, Aerochrome 2443 false color infrared film, was much less effective than color film at recording benthic features in shallow, moderately turbid water. Metric quality aerial photographs ($\leq$ 3 degrees of tilt off-nadir) are essential and should be acquired with a protocol similar to that employed by NOAA's Photogrammetry Branch (1980) to produce the highest quality data possible. Photography should be obtained at a scale appropriate to the areal extent of habitat, local water conditions, type of habitat being studied and resolution requirements for the resultant data. Photographic scale should normally range from 1:12,000 to 1:24,000. For chronically turbid estuarine or brackish water areas, 1:12,000 or larger scale photographs taken at times of minimal turbidity, may be required for acceptable visualization of submerged features.

Flightlines are planned with reference to aeronautical and nautical charts to include all areas known to have or which could have SRV. Ideally, each photograph in a flightline records land features sufficient to accurately locate images relative to the base map, about 1/3 of the exposure. This permits correction of photographic scale and registration to the external reference system. Sequential aerial photographs in a flightline should be obtained with 60% endlap to allow for stereoscopic interpretation and to compensate for loss of coverage due to sun glint in the photographs. Sidelap should be 30% to ensure contiguous coverage of adjacent flightlines and to produce a block of aerial photographs which may be subjected to photogrammetric bundle adjustment if necessary.

Knowledge of the study important to a successful project includes: plant species comprising SRV, morphology and phenology of these plants, depth range and location of known habitat, types and locations of benthic features that may confuse photointerpretation of SRV, seasonality of turbidity, weather, and haze, daily patterns in wind speed and direction, and progression of sun angle through the day.

5.2.2 Photo interpretation

Habitat defined by the presence of SRV can be interpreted from metric quality aerial photographs exposed as recommended in the previous section. The designation of a given area as SRV is a function of minimum detection unit, minimum mapping unit, and its proximity to other SRV. Assuming a photographic scale of 1:24,000, high quality optics, high resolution film and ideal conditions (e.g. dense seagrass growing on light-colored sediment in shallow, clear, calm water), it is usually possible to have a minimum detection unit of 1 meter for SRV. All detected SRV which appear to be in a continuum with adjacent SRV in an area which exceeds 0.03 hectare will be mapped in a single polygon. The minimum mapping unit is the smallest area to be mapped as habitat. At the C-CAP map scale of 1:24,000, the C-CAP standard minimum mapping unit is 0.03 hectare for SRV (i.e. a diameter of about 0.8 mm on the map represents a diameter of about 20 meters on the ground). Under some conditions and photographic scales it may be possible and desirable to exceed the 0.03 hectare minimum mapping unit. In any case, the minimum mapping unit for each coverage should be specified.

Two types of surveys are required within one year of the photography to characterize errors of omission and commission. Stratified random samples of potential habitat are observed for accuracy assessment (omission). Other sample sites are selected from the photographs to verify the photointerpretation (commission). Surface level data are intended to augment the photointerpreted data based on differentially corrected GPS positioning to a CEP of $\leq$5 meters. Accurate and current planimetric base maps of coastal land features are essential for georeferencing (establishment of geographic location) and scaling polygons of habitat interpreted from aerial photographs. C-CAP recommends the most accurate and current base map available for the study area and the most cost effective technology to apply local horizontal control of aerial photographs by comparison to the base maps. The base map and technology may vary regionally.

The accuracy of the base map used for local horizontal control places a limit on the accuracy of the C-CAP product. The two base maps broadly available are NOAA shoreline and USGS 7.5' topographic maps. NOAA, National Ocean Survey produces highly accurate shoreline maps based on tide-coordinated and fully rectified photography (Ellis, 1978; Slamma, 1980; NOAA Photogrammetry Branch, 1989; Crowell et al., 1991). When available and current, NOAA shoreline and coastal data should be used for C-CAP projects.

5.2.3 Mapping, Digitization, and Change Detection

Polygons of habitat interpreted from aerial photographs are mapped into a standard map projection coordinate system. The Universal Transverse Mercator Projection is recommended. C-CAP protocol allows the polygons interpreted from aerial photography to be transferred onto plainmetrically accurate basemaps using three approaches:

1. Stereoscopically interpret the photographs and optically scale the polygons and photographic image to fit planimetric horizontal control in the basemap with a zoom transfer scope.

2. Process the aerial photographs to become plainmetrically accurate orthophotographs, interpret and directly trace habitat polygons onto the planimetric base map.

3. Delineate and simultaneously rectify and digitize habitat polygons using an analytical stereo plotter.

Habitat polygons that have been transferred to the planimetric base map according to procedures 1 or 2 above, require digitization to be incorporated into the C-CAP spatial database. Digitization normally is accomplished using a digitizing tablet. The C-CAP objective of site specific change detection places greater emphasis on accuracy and precision of spatial data than required in one-time inventories or regional summaries of change. Methodology for monitoring site specific change on a statewide or regional scale is a recent development (Ferguson and Wood, 1990; Orth et al., 1991; Ferguson et al., 1993). Quantitative historical data, with possible exceptions in Chesapeake Bay or spatially limited study sites, does not exist. C-CAP recommends post-classification change detection for SRV. Post-classification change detection can be accomplished graphically or polygons are digitized and compared using a geographic information system to detect spatial displacement and quantify change.

6.0 RECOMMENDATIONS FOR ACCURACY ASSESSMENT OF CHANGE DETECTION

Sampling to determine the accuracy of a change map is inherently different from sampling for accuracy assessment of a one-point-in-time thematic map. The fundamental difference between these two situations occurs because the change categories usually represent a small portion of the original image, or thematic map. For example, in Dobson and Bright's (forthcoming) study of the Chesapeake Bay, approximately 3% of the area had experienced change over a five year period. The relative scarcity of change polygons in the change map implies that these polygons can be considered to be "rare" events that would only occasionally be detected using traditional sampling techniques such as random sampling. This means that if we used traditional sampling techniques to verify change, the confidence intervals on the accuracy rates would be unacceptably large due to the small number of samples in any category. In contrast, for assessing the accuracy of a one-point-in-time thematic map, random, stratified random and systematic unaligned sampling schemes have proven to be accurate means of estimating map accuracy (Congalton, 1988 and 1991). This section discusses the statistical methodology needed for assessing the accuracy of land cover change detection.

6.1 General Considerations on the Applicability of Statistical Theory to Accuracy Assessment of Land Cover Change Products

When one is dealing with a clearly defined, objective, sampling process, in which units are selected at random with known probabilities from some specified statistical population, statistical theory can be valuable in selecting formulae for estimating population parameters such as sums or means and their respective variances. However, C-CAP's proposed approach to generation of land cover change products is inherently complex and subjective. Subjectivity enters the process in the selection of scenes considered suitable for inclusion in an analysis, in the development of a land cover classification system, in the assignment of classes to pixels or polygons, in the assignment of classes in field verification operations, in the specification of spectral boundaries between land cover classes, and for that matter, boundaries among land cover classes on the ground. Moreover, selections among myriad alternative approaches for dealing with small areas of a given land cover type, co-registration errors, and a host of other problems are intrinsically arbitrary and subjective.

Whereas the subjectivity of aerial photointerpretation has generally been recognized and accepted, the subjectivity of the process of developing land cover products from digital data from satellite sensors has not. Perhaps this difference in perception arises because photo interpretation is more of an analog, "Gestalt" process, as opposed to the digital, "hard number" nature of the other. Indeed, the quality of an interpretation of an aerial photograph is a direct reflection of the expertise of the interpreter and his or her familiarity, knowledge and understanding of land covers for the region within the photograph. It should also be recognized that the quality of a land cover product generated from satellite data is no less a direct reflection of the expertise of the individual(s) producing the product. Thus, a land cover product is necessarily a personal accomplishment. This in turn means that estimates of accuracy do not generalize beyond the specific product in any kind of a rigorous way.

Other considerations also suggest the idea of restricting accuracy to individual products. For example, the geometrical size and complexity of land covers and the heterogeneity of their spatial context can be expected to have a major influence on the total number of misclassification errors, as well as, of course, the number of possible types of such errors. Obviously, regional-specific characteristics of land cover preclude generalization of accuracy estimates beyond the boundaries of the region sharing those characteristics.

Often, a goal of statistical analysis is to make inferences about one or more parameters of a population on the basis of a sample. Normally, one doesn't know and can't feasibly determine the true value of a parameter being estimated and thus can't know how accurate a given estimate is. Statistical theory can however be used to establish whether or not the formula used to compute the estimate is unbiased; i.e. whether the expected value of the estimator equals the parameter. However, in assessing the accuracy of a land cover product developed from remotely sensed data, we are faced with the fact that we cannot compute expected values of estimates, because additional complexities in the process of generating estimates preclude this. These complexities include the intrusion of subjectivity into decision processes alluded to above, but additionally include uncertainties that arise from the measurement processes themselves, as discussed elsewhere in this document. However, probably the most troublesome problem is the frequent absence of reference data that might serve as unequivocal standards against which accuracy could be measured. While field verification exercises might seem to provide such data, practical experience indicates they often have not, because even trained professionals on site don't always agree on what land cover class should be assigned to the corresponding pixel/polygon on a land cover product. Further, assignments of classes based upon on-site quantitative measurements of land cover are often at variance with assignments based upon non-quantitative on-site assessment.

Errors in a land cover change product can arise through failure to detect change from one class to another (errors of omission, "false negative") or through false implication of change that did not in reality take place (errors of commission, "false positives"). Both types of error are important, but errors of commission are easier to deal with simply because they will normally constitute a much smaller population of pixels, a subset of those for which change in land cover was indicated. Errors of omission on the other hand reside hidden within the larger population of pixels for which no change was detected.

Allocation of field sampling effort between the two types of classification error will depend upon the relative areas of the "changed" and "no change" stratum. When a product implies little change in land cover, it may be feasible to visit every site of apparent change and ascertain the accuracy. A probability sampling design would then be developed to assess errors of omission, the sampling frame consisting of those pixels/polygons for which no change was implied. Apparent current land cover classes could be used to delineate stratum. Allocation of sampling effort among strata could be weighted to reflect relative "importance" of the difference classes to the objectives of the overall program.

The design of the field efforts to estimate errors of omission in the change detection should take advantage of ancillary information to stratify the sampling frame. In the present case the frame is the population of pixels for which no change was detected. After completion of the change analysis, those conducting the accuracy assessment can seek ancillary information about where change either occurred or was likely to have occurred during the relevant time period, and use this information for subdividing the population of pixels into sets with varying likelihoods of having experienced land cover change. Then allocation of sampling effort to the strata can be accomplished to reflect the varying likelihoods of change, so that most of the effort is expended on pixels most likely to have changed. Very substantial gains in efficiency can be expected from such an approach, because most of the sampling effort will be expended where the changes in land cover are most likely.

6.2 Sampling Design Considerations for Field Work

Before discussing statistical technicalities about sampling design for the field work associated with accuracy assessment, it would seem appropriate to consider a conceptual framework of the types of error that can arise in a post-classification comparison to detect change. For simplicity our hypothetical classification of land covers contain only three classes A, B, and C. In the schematic below a capital letter refers to the true land cover of a pixel within a scene. The corresponding small letters refer to the class assigned to that pixel. Column headings of the matrix refer to the base scene (T_b), whereas row headings refer to the same scene at $T_{b\pm1}$. The diagonal, (upper left to lower right), 9-element cells of the matrix thus denote those cases in which there was no true change in the land cover within the boundaries of a hypothetical pixel. The six, off-diagonal, 9-element cells then represent those cases in which there was real change in the land cover of the pixel.

T_b

			A			B			C		
			a	b	c	a	b	c	a	b	c
T_{b+1}	A	a	1	4	4	5	2	6	5	6	2
		b	4	3	4	6	5	6	6	5	6
		c	4	4	3	6	6	5	6	6	5
	B	a	5	6	6	3	4	4	5	6	6
		b	2	5	6	4	1	4	6	5	2
		c	6	6	5	4	4	3	6	6	5
	C	a	5	6	6	5	6	6	3	4	4
		b	6	5	6	6	5	6	4	3	4
		c	2	6	5	6	2	5	4	4	1

Within the matrix:

1 = Correct no change	3
2 = Correct change	6
3 = Incorrect no change	24
4 = Incorrect change	48

With N (in this example, 3) classes of land cover, there are N^4 (in this example, 81) different possible outcomes of which N represent correct measurement of no change in land cover and N^2-N (in this example, 6) represent correct measurement of change. Of the remaining possible outcomes, N^3-N (in this example, 24) are errors of omission (false negatives), and the rest (48) are errors of commission (false positives).

The matrix thus makes it easy to consider the various possible land cover realities and corresponding land cover assignments for a pixel at the two times. For example, at the base time a given pixel might in reality be class B, but be misclassified as "a". In the other scene at a later time, the pixel might in reality be A, and correctly classified as "a". Our hypothetical post-classification comparison would erroneously indicate no change, "a" to "a", when in reality the land cover had changed from "B" to "A". This case then would represent one of the possible errors of omission.

In the matrix numbers have been entered to encode the nature of the various possibilities relating to the actual land cover within the boundaries of a map pixel over time and the assignments of the land cover to classes. A "1", for example, denotes a pixel where the land cover did not change, and the pixel was assigned the correct land cover class at both times. A "2" denotes a pixel where the land cover did change, and as before, the pixel was correctly classified on both occasions. A "3" denotes a case where there was no change in the land cover, but the pixel was assigned to the same incorrect land cover class on both occasions. A "4" refers to the situation in which there was no change in the land cover, but the pixel was mis-classified at one or both of the two times. A "5" refers to a case where the land cover changed but, because the pixel was assigned to the same incorrect class on both occasions, no change was detected. Finally, a "6" denotes a pixel where the land cover changed, and the assigned land cover classes changed, but the land cover class was incorrect on one or both occasions.

It is instructive to recall that if, for a given pixel, the probability of assigning it to its correct class is, say, 0.90 at time T_b and, say 0.95 at time T_{b+1} and if outcomes in the two time periods are statistically independent, then the probability that both assignments of land covers are correct is their product, 0.855, implying that the accuracy of a change detection product will be less than the accuracy of either of the land cover products used to generate it.

Rigorous accuracy assessment thus must be prospective and carefully planned with adequate allocation of resources to the work on the ground. The land cover measurement process on the ground must be one that will lead to unequivocal answers concerning the appropriate class for a given pixel, and be definable such that the classes are mutually exclusive and the set of classes is exhaustive.

It should be borne in mind that to some extent the goals of accuracy assessment and production of an accurate change detection product are contradictory. This is because accuracy of the change detection product could be increased by incorporating information gained during the on-site evaluations of classification accuracies. But, if this is done, then accuracy of the change detection product becomes an elusive moving target. Accuracy would increase as one attempted to estimate it and thus become a function of how much effort is put into the accuracy estimation process.

A case can be made for maintaining independence between the classification of the remotely sensed data and the operations on the ground to assess the accuracy of the classified scene. Otherwise, any attempt to detect change in land cover on the ground by repeated measurement of pixels at T_b and again at T_{b+1} (and thus directly measure accuracy of change detection by remote sensing) will be compromised by the special knowledge of those pixels by the individual(s) using the remotely sensed data for classification purposes.

As we have seen, inaccuracy can be subdivided into attribute error and positional error. Similarly, a case can be made that the ground (field) operations for assessing accuracy might be subdivided in a parallel fashion. For example, positional error involves mispositioning of boundaries. One might therefore argue that the most efficient sampling unit for assessing the accuracy of an ecotone boundary (one-dimensional) would be a line transect (one-dimensional). Conversely, a land cover is at least two-dimensional, and therefore a two dimensional sample unit would be more appropriate and efficient. When there are only two classes to deal with (e.g. SRV or its absence), then the problem is simplified to defining the boundaries between the two classes and therefore a line transect would be the sample unit of choice. A line transect would offer the accuracy assessment team in the field the additional opportunity to obtain data on the width of transitional ecotones where one class grades into another (a direct measure of fuzziness, if you like).

6.2.1 Sampling Design Guidelines for On-Site Determinations of Land Cover

1. The design should address both errors of commission and errors of omission at T_b and T_{b+1}, as well as for the change detection product.

2. Allocation of effort should ensure distribution of effort across the entire scene. This can be accomplished by subdividing the scene into geographically defined strata. Additional stratification factors might include physiographic characteristics, degree of urbanization, likelihood of future development, etc.

3. Random selection of sites on which land cover will be measured should be based upon unequal probability sampling to reflect overall priorities of the C-CAP program.

4. If the amount of change detected is relatively little, then assessment of errors of commission might efficiently proceed on the basis of including to the extent possible all change pixels for on-site land cover measurement. For large areas this task may become impossible.

5. Consideration should be given to selecting some sites/pixels for measurement at both T_b and $T_{b\pm1}$ to take advantage of the expected serial (temporal) correlation in land cover at a given site.

6. The importance of approximate concurrence in ground and satellite data or aerial photography cannot be over emphasized.

7. Thorough training of field crews must precede the collection of the reference data for accuracy assessment.

8. Rules must be developed and rigorously followed for inclusion/exclusion of selected sites which fall on ecotones and thus are mixtures of two or more land cover classes. We note that mixed pixels cannot be incorporated into the above matrix.

9. The accuracy assessment process is dependent upon the exact location of pixel/polygon boundaries on the ground.

6.3 Sample Design Description

As previously noted, sampling to determine the accuracy of a change map is inherently different from sampling for accuracy assessment of a one-point-in-time thematic map because the change categories usually represent a small portion of the original image, or thematic map. For change detection accuracy assessment traditional sampling techniques cannot usually be employed. The main reason that they are not used is because of the substantial cost of establishing "permanent" sample locations that can be revisited through time. In addition, the optimal allocation of samples for each of the change categories would be different than the optimal allocation of samples to the original thematic classes and additional sample points would be needed over the original base of "permanent" sample points. This is largely because the change polygons cover only a small portion of the original image and would not be well detected with, say, random sampling. Because of these limitations we are proposing another approach to sampling for the accuracy of a change map. To do so we need some simple notation as follows:

	Category		Definition
	T_c	=	true change
	T_n	=	true no change
	P_{nc}	=	predicted no change correct
	P_{ni}	=	predicted no change, but with incorrect class label
	P_{ci}	=	incorrectly predicted change
	P_{cc}	=	correctly predicted change
[1]	T_n-P_{nc}	=	true no change; correctly predicted no change **[correct no change]**
[2]	T_c-P_{cc}	=	true change; correctly predicted change **[correct change]**
[3]	T_n-P_{ni}	=	true no change; incorrectly predicted no change class **[incorrect no change]**
[4]	T_n-P_{ci}	=	true no change; incorrectly predicted change **[incorrect change]**
[5]	T_c-P_{ni}	=	true change; incorrectly predicted no change
[6]	T_c-P_{ci}	=	true change; incorrectly predicted change

Classes [1] through [6] of this table and the table in §6.2 are directly equivalent.

By following the procedure outlined in Section 5.5 for Change Detection Algorithms the analyst will produce a **Change Map** and a **No Change Map**. We can utilize this dichotomy to achieve real sampling efficiencies as will be shown. The following Table shows the components that can be estimated in each of the two strata types (Change and No Change).

Reference data

	True no change (T_n)	True change (T_c)	Strata
Predicted no change class correct (P_{nc})	[1] T_n-P_{nc}	--	No change
Predicted no change class incorrect (P_{ni})	[3] T_n-P_{ni} false negative	[5] T_c-P_{ni} false negative	No change
Predicted change class incorrect (P_{ci})	[4] T_n-P_{ci} false positive	[6] T_c-P_{ci} false positive	Change
Predicted change class correct (P_{cc})	--	[2] T_c-P_{cc}	Change

These sources of error can also be depicted using a change detection error matrix which is presented in more detail in sections 6.2 and 6.8

Change Detection Error Matrix and Sources of Error

		Reference Data								
		No Change			Change					
		AA	BB	CC	AB	AC	BA	BC	CA	CB
Predicted No Change	AA	1	3	3	5	5	5	5	5	5
	BB	3	1	3	5	5	5	5	5	5
	CC	3	3	1	5	5	5	5	5	5
Predicted Change	AB	4	4	4	2	6	6	6	6	6
	AC	4	4	4	6	2	6	6	6	6
	BA	4	4	4	6	6	2	6	6	6
	BC	4	4	4	6	6	6	2	6	6
	CA	4	4	4	6	6	6	6	2	6
	CB	4	4	4	6	6	6	6	6	2

6.3.1 The (predicted) No Change Strata

It is interesting to note that different components of change detection can be assessed within each of the two Strata - predicted **Change** and predicted **No Change**. The (predicted) No Change Stratum is by far the largest stratum, generally encompassing over 90% of the original image, or thematic map. For this dominant stratum it is possible to estimate three components of change denoted as [1], [3], and [5] in the above table. These components are for the majority case (predicted no change when there truly was no change (case[1]), and for the lesser occurrences of predicted no change, but of the wrong no change class (case[3]), and (case [5]) predicted no change when there was change.

The sampling of these two components of the predicted Change and the predicted No Change Strata can be done following standard techniques as outlined in Congalton (1988, 1991) including random, stratified random, and systematic unaligned designs provided that none of these components are rare events (say <10%). Guidelines on determining sample sizes and in sampling strategies are discussed in more detail below.

We are proposing two different sampling strategies - one for the Change Stratum and one for the No Change Stratum.

Because of the costs of estimating the rates of each of these components (to a given level of precision) it may make sense in some instances to combine classes. For example, we might decide that we really only care about components [1] and [5] in the No Change Stratum. If we did this, we would be essentially combining classes [3] and [1]. That is we would try to detect true no change when there was predicted no change, but the type of no change could be incorrect. In this simplified case we are able to utilize a binomial distribution (Cochran, 1977) of pixels within the No Change Stratum for determining variance and the associated sample size needed to achieve a desired precision level.

The binomial case is presented in the event that only 2 of the 3 components in either stratum are to be estimated. However, in most instances we would want to estimate all three components of the No Change Stratum (or the Change Stratum). Then the multinomial formulae presented for the Change Stratum in the next section can be used. Of course, the extension of the binomial to multiple categories is the multinomial distribution.

<u>Binomial Sampling and Sample Size Determination</u>

For the binomial distribution the probability that a sample of **n** units contains **a** units which are of true no change type is:

$$Pr(a) = \frac{p^a * q^{n-a} * n!}{a! * (n-a)!}$$

Let $p = a/n$ and $q = 1-p$

Then $$v(p) = \frac{p * q}{n-1}$$ and

$$[6] \quad n = \frac{t^2 * p * q}{\delta^2}$$

The variance function is greatest when the population is equally divided between the two classes and is symmetrical about this point. The standard error of p changes relatively little between $0.3 \leq p \leq 0.7$. However at $p < 0.3$ or $p > 0.7$ the sampling effort needed to reduce the standard error of the estimate to a desired level decreases rapidly. This is because the converse of a small p, say p=0.1 is a large q value (q=1-p) which in this case is 0.9. In equation [6] δ is the half width of the desired confidence interval, which we will also refer to as error tolerance.

For example if $t \approx 2$, $p = 0.9$ = the approximate probability of true no change in the (predicted) No Change Stratum, the half width of the confidence interval $\delta = 0.03$, and $q = .10$ then we need

$$n = \frac{2^2 * 0.9 * 0.1}{.03^2} = 400 \text{ samples}$$

Rather than trying to achieve one absolute tolerance limit it is often better to specify varying levels of precision for the different possible values of p. For example we might want to estimate $0.3 \leq p \leq 0.7$ within a tolerance δ=.05; $0.1 \leq p < 0.3$ within an error tolerance δ=.04, and for $p < 0.1$ within an error tolerance δ=0.03. The choice of error tolerances (δ) depends upon the relative importance of obtaining highly precise numbers given the importance of the ecological area under investigation and the cost of obtaining the samples. Another way to approach this problem is to estimate a population proportion with a coefficient of variation of **c** or less, rather than specifying a tolerance level as in equation [6]. The sample size required to achieve this is given by:

$$n \geq \frac{1-p}{p*c^2}$$

The sample size needed to achieve a specified coefficient of variation (**c**) decreases as p increases. At small values of p, a very large sample size is needed to meet the precision requirements. When $p \leq .10$ we usually consider the item to be rare in the population. Using a coefficient of variation (**c**) of 5% and p values of 0.9, and 0.1 respectively in equation [7] we obtain sample sizes of 44, and 3,600, respectively. By changing the coefficient of variation requirement to 15% the sample size decreases to 400 for p equal to 0.1. Obviously the choice of the coefficient of variation **c**, like the choice of the error tolerance δ, has a large impact on the required sample and needs to be judiciously chosen.

6.3.2 The (predicted) Change Stratum

The (predicted) Change Stratum is considerably smaller than the No Change Stratum and this gives us some advantages for sampling. For the Change Stratum we need to estimate three components of change denoted as [2], [4] and[6] in the previous table. The components associated with the (predicted) No Change Stratum are; incorrectly predicted change when there was no change (case [4]), correctly predicted change but of the wrong kind (case [6]), and correctly predicted change (case [2]).

The real advantage of partitioning the image, or thematic map, into Change and No Change Stratum is that we can conduct different kinds and intensities of inventories on these two Strata. The Change Stratum allows for concentrated sampling since the areas of change are usually clustered in specific locations within an image. It may be possible to conduct the sample survey for this Stratum using a helicopter to cover the Change Stratum in a minimum amount of time while visiting a maximum number of locations.

Because we need to estimate three components we no longer can use the binomial distribution, but rather the multinomial. It should be evident that in this simplistic appraisal we have tried to estimate only a total of 6 components of the Change Detection Error Matrix as specified in the two tables above. Of course, we are, in likelihood, interested in each of the 81 (9 by 9) components of the Change Detection Error Matrix presented in the example. In that matrix there are 9 elements contained in the No Change (predicted and actual) portion of the matrix (components [1] and [3]; 18 elements associated with component [5]; 18 elements associated with component [4]; and 36 elements associated with components [2] and [6]. We would utilize the multinomial distribution in deciding the sampling intensity needed to achieve a specified level of precision whether we are interested in all 27 elements of the predicted No Change matrix or in estimating only the 3 summary components of the predicted No Change matrix ([1], [3], and [5]). Likewise, in the predicted Change matrix in the example given, we would utilize the multinomial sample size formula in estimating the 54 elements of the Predicted Change matrix, or in estimating the 3 summary components ([2], [4], and [6]). The possible rationale for estimating only the summary components is that it requires significantly less sampling than for estimating each of the individual elements of the change matrix. It is expected that in most instances it is desired to estimate all elements of the Change Detection Error Matrix. In either event, we would use the formula presented below.

Tortura (1978) gives a procedure for determining the sample size required for simultaneous confidence intervals for parameters of the multinomial distribution which utilizes the approximate large sample equations for the confidence limits. The sample size n is given by:

$$n = \chi^2_{(1,1-\partial/k)} \frac{p * (1-p)}{\delta^2}$$

where δ is the half width of the desired confidence interval, and k=3 which is the number of categories in the Change Stratum (components [2], [4] and [6]). The Chi-Square value is approximately 4.6 for an α level of 0.10 with k=3 categories (components [2], [4], and [6] of the predicted Change Stratum). If the half-width (δ) of the confidence interval is set to δ=0.10 some sample sizes for different values of p and k are given below:

Sample sizes (n) as a function of the true probability of occurrence (p) of a category and the number of categories being estimated for α=0.10 and δ=0.10.

	Multinomial Sampling	Multinomial Sampling	Binomial Sampling
	k = 54 categories χ^2 =7.9	k = 3 categories χ^2 =4.6	k = 2 categories
p	n = sample size	n = sample size	n = sample size
0.1	71	41	36
0.3	166	97	84
0.5	198	115	100
0.7	166	97	84
0.9	71	41	36

When each p has a different half-width (δ) then a separate calculation is made for each p and the largest n is selected as the desired sample size. When only one half-width is required, then the sample size is calculated for the proportion nearest to 0.5. This provides for a conservative estimate of sample size since the largest sample size is required exactly at p=0.5. It is clear that sample size needed to achieve a given level of precision increases as the number of categories (elements of the matrix) being estimated increases. However, the increase in sample size is not proportional to the increase in the number of categories for which accuracy assessment are required.

We have already discussed how estimating all the elements of the predicted Change Stratum (or the predicted No Change Stratum) requires more sampling than if we are willing to ignore the specific "from-to" nature of the errors. For example if we only estimate the summary components of error ([1], [3], and [5] of the predicted No Change Stratum, and [2], [4], and [6] of the predicted Change Stratum) we would reduce the number of categories in the predicted No Change matrix from 27 to 3, and the number of categories in the predicted Change Stratum would be reduced from 54 to 3. This substantially reduces the amount of field sampling required. In the table above, the number of samples required to estimate a category with a true p=0.5 and a confidence interval half-width equal to 0.10 when there are 54 categories is 198, the sample size needed using the binomial distribution. If, as before, there are p=0.5 correctly classified pixels that we want to estimate with a confidence interval half-width of 0.10 we need n=100 samples with binomial sampling. Thus reducing from 3 to 2 categories only produces minor sampling efficiencies. The greatest gain is from reducing 54 to 3 categories under the Multinomial sampling model. Nonetheless we expect that the normal case will be that we want to estimate all elements of the Change Detection Error Matrix.

6.3.3 Overall Sample Design

To help ensure that the sampling is distributed throughout the region of study it is recommended that a geographically based multistage Stratified random sample (SRS) be taken of the entire area for estimating the accuracy of the No Change Stratum. In this setup the study area is broken up into several subregions along biophysical and ecological criteria. From within each subregion orthophoto quadrangle sheets, for example, could be chosen at random. Then from within a quadrangle, quarter-quadrangles are randomly selected and field plots are selected randomly or with a systematic unaligned sampling scheme. Plot sizes will vary according to the ecological class under investigation. However, given the registration errors in image processing the field plots should be approximately 3x3 pixels in size where possible to ensure that the sampled plot truly encompasses the designated sample selection point on the map or image.

For estimating the accuracy of the Change Stratum, we can do concentrated sampling using a helicopter to cover the Change Stratum in a minimum amount of time while visiting a maximum number of locations. Because the change areas tend to be concentrated in places such as urban centers, fringes of urban centers, or in rural forested areas in specific locales, it may be possible to designate a buffer region around these centers where we suspect change activity would take place. These buffers would then be added, so to speak, to the predicted change areas which are being more intensively sampled. We know that we can get better estimates of cases [1], [3] and [5] because they occur on a smaller area and we can sample them more intensively. By considering these buffer regions of "potential change" for more intensive sampling than they would normally achieve, we could improve on our ability to assess the accuracy of false negatives - finding true change when it has not been predicted. (case [5]).

6.4 Reference Data Collection

A critical component to any accuracy assessment is the need for accurate reference data. If the reference data are poor or improperly collected, then the entire assessment becomes meaningless. If the reference data has significant error associated with it, then it should not be used as reference data. Reference data should be collected using the same classification scheme that was used for the remotely sensed data. It should also be applied over the same minimum mapping unit as was applied to the remotely sensed data. In many instances, it is enough to simply make observations to determine the reference data; for example, the species of vegetation or the presence/absence of a certain factor. In other cases, actual measurements are required. In either case, sampling within the area of interest (polygon) is required in order to obtain the proper reference label. A sufficient number of samples must be acquired within each polygon to assure proper labeling. If too small a sample is taken then the label may be incorrect due to some inclusion in the polygon that was smaller than the minimum mapping unit, but was included in the sample. Finally, the reference data should be as objective as possible. An important mechanism for promoting objectivity is to use a reference data collection form to force all data collectors through the same collection process.

Reference data collection sheets are important for a number of reasons. As already mentioned, they should be designed to make the data collection as objective as possible. The form should lead the collector through a quantitative process to a definitive answer from the classification scheme. Figure 2 provides an example of a flow-chart that guides the field crews through the logic of the decisions necessary to arrive at the determination of the attribute label of the particular location under investigation. It also provides a means of performing a quality assessment/quality control check on the collection process. Obviously, the complexity of the reference data collection sheet is dependent on the level of the classification scheme.

Reference data collection sheets, regardless of their complexity, have some common components. These include: (1) the name of the collector and the date of the collection, (2) locational information about the site, (3) some type of table or logical progression that represents what the collector is seeing, (4) a place to fill in the actual category name from the classification scheme, and (5) a place to describe any anomalies, any variability, or interesting findings at the site. The project-specific list of attributes on field sheets needs to be developed on case-by-case basis.

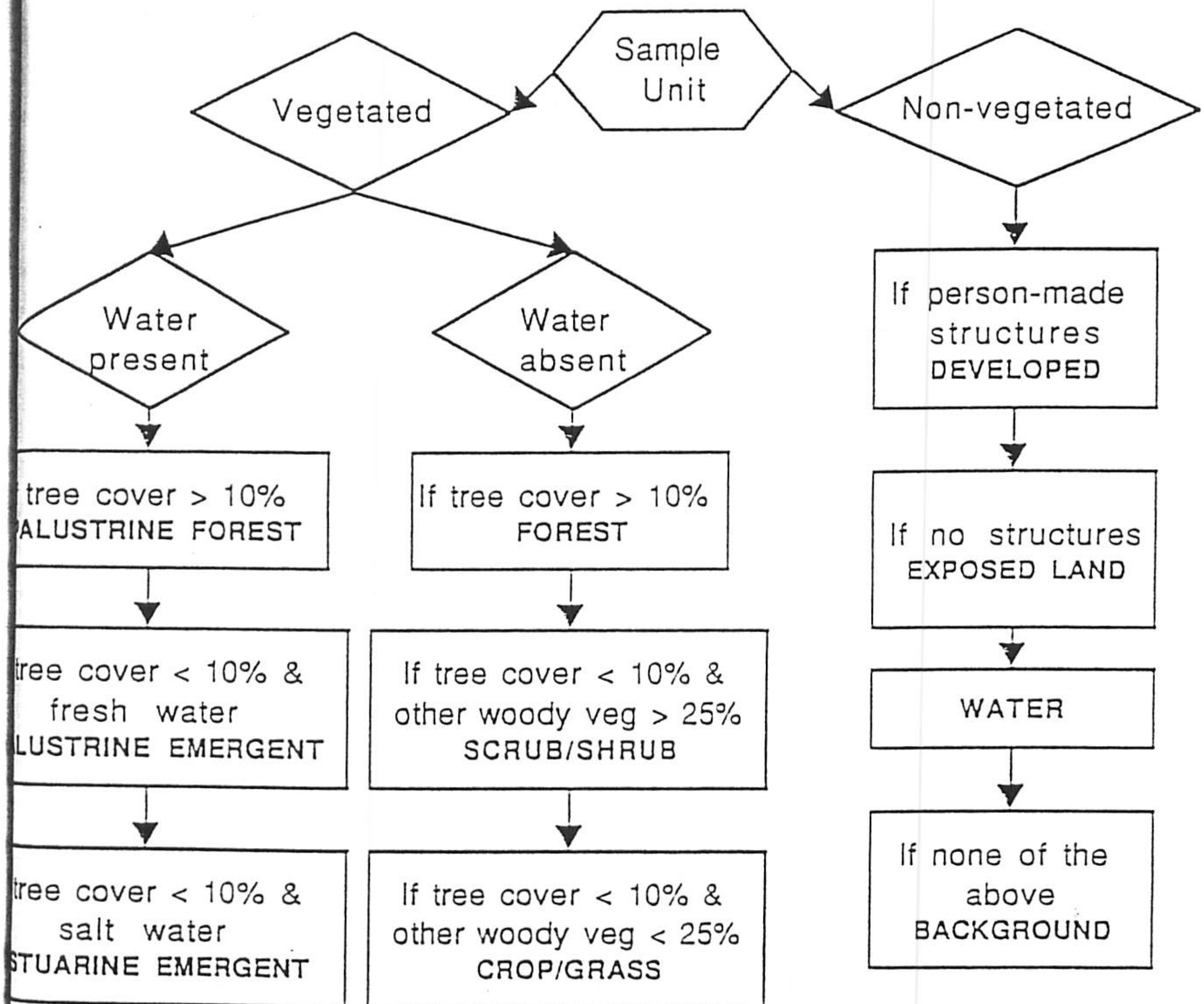

Figure 2 Flow Chart for Guiding Reference Data Collection.

It is important to realize that certain types of land cover changes can be confirmed easily and cheaply by direct visual inspection, while others require detailed quantitative field measures. For example, conversion of forest to a shopping center (urbanization) requires only visual inspection. Other types of changes will entail greater expense and effort because they will require on site measurement of characteristics such as vegetation cover percent, soil moisture, or salinity. For those parameters requiring quantitative measurement a design for determining the number of samples to take within a polygon (land cover class) and the exact measurements to be taken needs to be explicitly undertaken.

6.5 Positional Error Issues and Recommended Procedures

At several points in the change detection and accuracy assessment procedures, it will be necessary to identify the locations of specific points on the Earth's surface accurately. Specifically, accurate methods will be needed for determining the locations of: control points on TM scenes; the same control points on topographic maps; and pixels and polygon boundaries on the ground.

Errors in positioning can lead to false positives and false negatives in change detection, if a pixel at T_b is mismatched to a differently located pixel at another time. Errors in positioning can also lead to errors in accuracy assessment, if a pixel or polygon is compared to an incorrect location on the ground.

On a TM scene, all spatial variation within pixels is lost. At best, positional accuracy for a well-defined object such as a road intersection is 0.5 pixels, or 15m. However, for objects larger than 0.5 pixels, or for poorly defined or indistinct objects, positional accuracy can be much poorer. The physics of the detector can produce an apparent positional error of as much as 2 pixels, as the detector takes time to respond to sharp changes in contrast during its scan. In summary, the positional accuracy of an object detected in a TM scene varies between 15m and 60m, depending on the sharpness of definition and spatial and spectral extent of the object.

Errors due to misregistration of TM scenes can be reduced by careful selection of control points. These should be small, well-defined objects such as road intersections or buildings, and should be widely distributed over the scene. The more control points used, the greater the positional accuracy. It is helpful if software used for registration allows access to the residual positional errors at each control point, as these can often help to identify mistakes in identifying controls, or errors in positions.

The positional accuracy of a topographic map is established by the producing agency, normally in the form of a CMAS. The 1:24,000 topographic maps produced by the US Geological Survey are the most likely to be used for scene registration as sources of control point coordinates, because they are the largest scale base mapping available over the continental US. On topographic maps at a scale of 1:24,000 a

CMAS of 0.5mm corresponds to a distance of 12m on the ground. Thus, the accuracy of TM registration to a 1:24,000 map ranges from 17m to 72m depending on the degree of definition of the object being used as control point, assuming, of course, that no blunders have been made in identifying objects. In some regions of the country, section corners can be identified relatively easily to form a network of well-defined points for image registration. The presence of blunders can sometimes be detected by examining residuals, as suggested above.

When one TM scene is registered to another, some of the sources of positional error are not present, and it can be easier to achieve high accuracy. A positional accuracy of 0.5 pixels, or 15m, is routinely achieved in scene registration.

Accuracy assessment requires the identification of objects on the ground. As before, this is relatively straightforward if the object is comparatively unique and well-defined, such as a street intersection, although blunders are always possible because street names are invisible from space. But accuracy assessment requires the accurate location of randomly chosen pixels, and polygons which may have no obvious expression on the ground. In broad terms, two methods are available for accurate positioning of field check(s): (1) the global positioning system (GPS), a satellite-based system for direct measurement of position in Earth coordinates, including UTM; and (2) positioning relative to well-defined objects that can be found on the ground, such as road intersections or buildings. The use of GPS may be limited under forest canopy and when terrain is obscure.

At this time, the accuracy of GPS varies widely depending on the particular implementation of the technology.

a. Hand-held GPS routinely (95% of single point non-differentially corrected fixes with selective availability active) achieves accuracies of 100 m CEP (circular error probable).

b. Differential hand held GPS (95% of single point, differentially corrected fixes with SA active 10m CEP) is generally available in the coastal US and real time availability is scheduled via Coast Guard broadcasts.

c. The fixed receiver (12 channel upgrade of 6 channel) can be on a monument or positioned by survey to monuments and permit differentially corrected accuracies of 5m CEP (95% of single point fixes).

If field checks are positioned relative to well-defined field objects, positional accuracy depends on the positional accuracy of the reference objects, as well as on the system of measurement used to establish relative position. In a field situation, it is unlikely that the latter will be better than simple pacing, or ocular estimates. Moreover, blunders are common in this situation, such as when an object is misidentified, or the wrong point is found on a linear object such as a shoreline. Previous accuracy assessment by C-CAP found it almost impossible to achieve adequate positional accuracy of field checks in some areas, particularly coastal wetlands. Thus, the positional accuracy of a field check location established by this method is likely to be substantially worse than the positional accuracy of the topographic base map.

Because of these problems, the preferred method for locating field sites at this time is differential GPS, with the fixed receiver located at a geodetic control monument. Care should be taken to ensure that the datum used by the GPS positioning matches that of the topographic base map used for image registration; NAD83 is recommended.

6.6 Boundary Effects and Recommended Procedures for their Minimization

Implicit in the C-CAP change mapping program is the assumption that the surface of the Earth can be divided into areas of uniform land cover class, separated by sharp lines. As noted earlier, this assumption approximates the truth to varying degrees. Some classes, such as water, are relatively well-defined and spatially homogeneous. Other classes may grade continuously at their borders, and may include significant and substantial heterogeneity.

As noted in section 4, where the concept of error was introduced, there are various views of what constitutes "accuracy" in the context of land cover mapping. In the, "pixel" view, a land cover map provides an estimate of the land cover at every point, and its accuracy is determined by the proportion of pixels found to be correctly classified, based on a ground check or source of high accuracy. In the "polygon" view, a land cover map is a collection of classified polygons each enclosing a minimum number of pixels, the minimum mapping unit, and its accuracy is determined by the proportion of polygons found to be correctly classified. In this view, since a polygon can have only one class, inclusions or heterogeneities within polygons, less than the minimum mapping unit, and blurred boundaries, are generalization not errors.

The proposed method of accuracy assessment falls into the second, or "polygon" view. A sample of polygons will be chosen, and their classifications checked in the field. A polygon will be declared correct if the field check produces the same classification. Since many polygons will not have sharply defined boundaries, and will contain inclusions (less than the minimum mapping unit) of other classes, it will be necessary for the field procedure to generalize at boundaries and inclusions, with the minimum mapping unit in mind in making an assessment. Locations near polygon boundaries should be avoided as far as possible given the problems of field positional accuracy. Polygons should be assessed using appropriate averaging methods designed to generalize away small inclusions.

6.7 Error Matrix Generation and Analysis

Generation of an error matrix is recommended for quantification and interpretation of errors for change detection purposes.

An error matrix as defined by Congalton (1991) is a square array of numbers which express the number of sample units (i.e., pixels, clusters of pixels, or polygons) assigned to a particular category relative to the actual number in that category as verified on the ground. The columns usually represent the reference data and the rows indicate the classification generated from the remotely sensed data. An error matrix is a very effective way to represent accuracy in that the accuracies of each category are plainly described along with both the errors of inclusion (commission errors) and errors of exclusion (omission errors) present in the classification.

The error matrix can then be used as a starting point for a series of descriptive and analytical statistical techniques. Perhaps the simplest descriptive statistic is overall accuracy which is computed by dividing the total correct (i.e., the sum of the diagonal) by the total number of sample units in the error matrix. In addition, accuracies of individual categories can be computed in a similar manner. However, this case is a little more complex in that one has a choice of dividing the number of correct samples in that category by either the total number of samples in the corresponding row or the corresponding column. Traditionally, the total number of correct samples in a category is divided by the total number of samples of that category as derived from the reference data (i.e., the column total). This accuracy measure indicates the probability of a reference samples being correctly classified and is really a measure of omission error. This accuracy measure is often called "producer's accuracy" because the producer of a classification is interested in how well a certain area can be classified. On the other hand, if the total number of correct samples in a category is divided by the total number of samples that were classified in that category, then this result is a measure of commission error. This measure, called "user's accuracy" or reliability, is indicative of the probability that a sample classified on the map/image actually represents that category on the ground (Story and Congalton, 1986).

In addition to these descriptive techniques, an error matrix is an appropriate beginning for many analytical statistical techniques. This is especially true of the discrete multivariate techniques. Starting with Congalton et al. (1983), discrete multivariate techniques have been used for performing statistical tests on the classification accuracy of digital remotely sensed data. Since that time many others have adopted these techniques as standard accuracy assessment tools (e.g., Rosenfield and Fitzpatrick-Lins, 1986; Hudson and Ramm, 1987; Campbell, 1987). Discrete multivariate techniques are appropriate because remotely sensed data are discrete rather than continuous. The data are also binomially or multinomially distributed rather than normally distributed. Therefore, many common normal theory statistical techniques do not apply.

Another discrete multivariate technique of use in accuracy assessment is called KAPPA (Cohen, 1960). The result of performing a KAPPA analysis is a KHAT statistic (an estimate of KAPPA), which is another measure of agreement or accuracy. The KHAT statistic is computed as

$$\hat{K} = \frac{N \sum_{i=1}^{n} x_{ii} - \sum_{i=1}^{n} (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{n} (x_{i+} * x_{+i})}$$

where r is the number of rows in the matrix, x_{ii} is the number of observations in row i and column i, x_{i+} and x_{+i} are the marginal totals of row i and column i, respectively, and N is the total number of observations (Bishop et al., 1975). The equations for computing the variance of the KHAT statistic and the standard normal deviate can be found in Congalton et al. (1983), Rosenfield and Fitzpatrick-Lins (1986), and Hudson and Ramm (1987), to list just a few. It should be noted that the KHAT equation assumes a multinomial sampling model and that the variance is derived using the Delta method. In addition to being a third measure of accuracy, KAPPA is also a powerful technique in its ability to provide information about a single matrix as well as to statistically compare matrices.

6.8 Visualization and Interpretation of Change Error Matrix

The change error matrix will have the same characteristics as the traditional classification error matrix, but will assess errors in changes between two time periods and not simply a single classification. An example shown below in Figure 3 demonstrates the use of a change detection error matrix.

Figure 3 shows a single classification error matrix for three vegetation/land use categories (A, B, and C). The matrix is of dimension 3x3. The major diagonal of this matrix indicates correct classification. In other words, when the classification indicates the category was A and the reference data agrees that it is A, then the [A,A] cell in the matrix is tallied. The same logic follows for the other categories B and C. Off-diagonal elements in the matrix indicate the different types of confusion (called omission and commission error) that exist in the classification. This information is helpful in guiding the user to where the major problems exist in the classification. Figure 3 also shows a change detection error matrix for the same three vegetation/land use categories (A, B, and C). Note that the matrix is no longer of dimension 3x3 but rather 9x9. This is

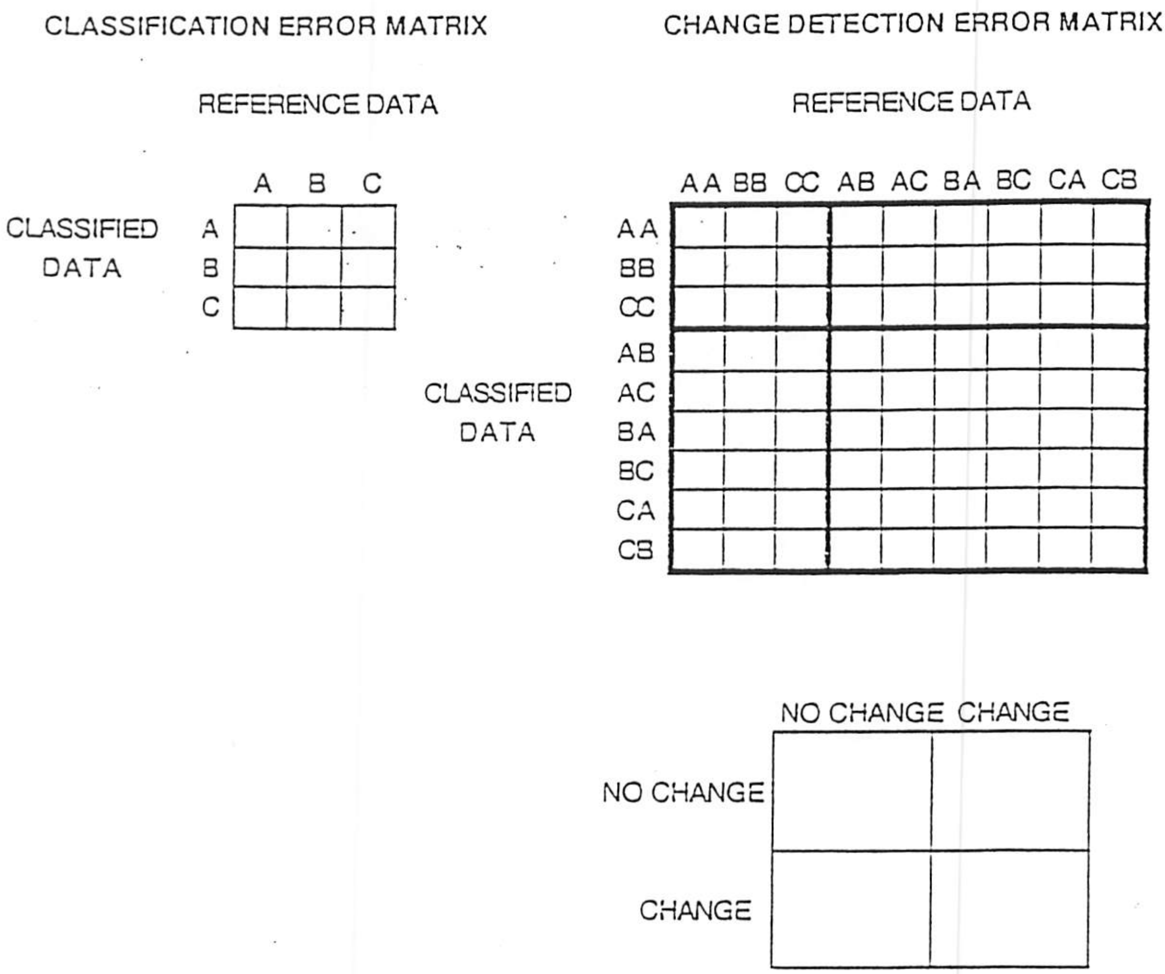

Figure 3 A comparison between a single classification error matrix and a change detection error matrix for the same vegetation/land use categories.

because we are no longer looking at a single classification but rather a change between two different classifications generated at different times. Therefore, the question of interest is, "What category was this area at time 1 and what is it at time 2?". The answer has 9 possible outcomes (A at time 1 and A at time 2, A at time 1 and B at time 2, A at time 1 and C at time 2,..., C at time 1 and C at time 2) all of which are indicated in the error matrix. It is then important to note what the remotely sensed data said about the change and compare it to what the reference data indicates. This comparison uses the exact same logic as for the single classification error matrix, it is just complicated by the two time periods (i.e., the change). Again, the major diagonal indicates correct classification while the off-diagonal elements indicate the errors or confusion.

Some other very interesting statistical difficulties result from the generation of the change detection error matrix. As noted in Figure 3, a three category, single classification error matrix means a nine category, change detection matrix. This problem is only compounded as the number of categories increases. Problems with sampling to build the matrix and especially sample size quickly become critical. See section 6.3 for the statistical formulae needed to calculate sample sizes as a rule of thumb for error matrix generation. The cost of collecting the reference data for these samples can quickly become prohibitively expensive. Therefore, it is obvious that all possible changes may not be included in the change detection error matrix. Fortunately, not all changes are of the same importance and the most critical ones can be included.

As described, the change detection error matrix can be an effective way of quantitatively assessing the accuracy of a change analysis data set. All the analysis techniques developed for the single classification error matrix are applicable, especially the Kappa analysis (Congalton et al., 1983) which will allow for the determination of statistical significance of the changes.

7.0 SUMMARY OF LIMITATIONS

The general limitations of the procedures recommended in this report include, but are not limited to the following:

Methods for tracking the 'lineage' of each file used in a change detection (two single date classifications and a change detection file) must be developed. Without such information, the error evaluation is incomplete because it may not be possible to replicate the results of the change detection analysis. The errors in the reference data also become a compounding factor.

There is relatively little experience dealing with large multi-temporal high resolution remote sensor datesets used for change detection. Additional work with such datasets will lead to improved perspectives on change detection error evaluation problems. We may not be addressing some of the more important problems simply because our scientific community does not have sufficient experience as of this writing.

Most remote sensing land-based change detection projects are based on radiometrically corrected satellite remote sensor data. Unfortunately, only water-related studies interested in monitoring suspended sediment, chlorophyll etc. correct the data. To conduct truly accurate digital change detection, it is probably necessary to apply radiative transfer atmospheric and bi-directional reflectance corrections (e.g. if off-nadir SPOT data are used) to the remote sensor data used in the change detection.

Highly accurate change detection ideally requires calculation of at surface reflectances. This requires a highly accurate digital elevation model (especially in the coastal zone), a detailed library of bidirectional reflectance factors, a per-pixel atmospheric correction (including thin clouds) and precise knowledge of solar elevation and azimuth. Of these, only the solar angles are presently well characterized. One can expect, however, that as the EOS program develops, we will have both the appropriate satellite systems and the ancillary data (DEM's and BRDF's, etc.) to calculate at surface radiance values and reflectance with sufficient date-to-date precision for change detection purpose.

EOS, however, will not solve the problem of insufficient return frequency to enable to detect change in features with both high temporal and spatial frequencies. Those sensors (e.g. MODIS) which observe with high spectral and temporal frequencies (every 2 days) have relatively large pixel sizes (250 m to 1 km). Landsat 7 (and the follow-on advanced Land Remote Sensing System) is planned to remain on a 16-day cycle. Higher spatial resolution commercial systems are also proposed, but their swath widths preclude high repeat cycles. At this writing, no constellation of calibrated, nadir viewing, high spatial resolution systems have been funded. Unless there is a major change in funding for foreign or U.S. remote sensing satellites a major limitation for change detection will remain to be our inability to match plant phenology, cloud cover, atmospheric turbidity, and satellite overpass for multiple dates in major portions of the earth. This limitation is particularly important in the U.S. coastal zone areas involving the areas of interest to C-CAP.

Different sampling schemes have been proposed for the Change and No Change Strata. More intensive sampling of the predicted Change Stratum is recommended than for the No Change Stratum. Because more intensive, or even exhaustive, sampling in the Change Stratum is proposed a reasonably good algorithm for producing the Change Map is needed. If the Change Stratum is either too large or too small sampling efficiencies will be lost. Thus, the quality of the algorithm used to predict change has a direct impact on the success of the sampling protocol.

In the case of using a very conservative algorithm to identify potential change it is obvious that it would produce too few candidate change pixels. This will result in a very large No Change Stratum. If this happens then it will be very difficult to find false negatives (true change when there was no change predicted) because these can be considered rare events which are not well represented by simple random sampling or systematic sampling techniques.On the other hand, if a very liberal algorithm is utilized to identify potential change then a large Change Strata will be generated. Because intensive sampling for this Stratum is recommended the costs of sampling would become prohibitively large.

Limitations for change detection of submerged land cover have to do with quality of the photographic image (water penetration, contrast of vegetated and unvegetated bottom, and deviation of photographic axis from vertical) certainty of habitat signatures in the images, and accuracy of the base map for registration of the photographic data. Field data is limited by ambient conditions (which can seriously restrict the quality of field observations), accuracy of positioning in the field (differentially corrected GPS data is required) and appropriateness of the decision process from objective measurements to the assignment of land cover category for the station. Prospective change detection is the best way to go. Retrospective change analysis is limited due to lack of appropriate surface level data required to assure, or supplement the interpretation of the historical photography. Virtually all photography will have some limitations - particularly coverage and visibility of unvegetated bottom as depth increases to the limit of potentially vegetated bottom. Retrospective change detection not only will be spatially restricted but also biased in favor of loss. Historical loss will be more likely to be detected than historical gain because unequivocal signatures (those identifiable by an interpreter experienced in the study area and not requiring surface level verification) of vegetated bottom are more reliably obtained than are signatures of unvegetated bottoms from historical photographs, in the absence surface level verification.

8.0 ACKNOWLEDGEMENTS

The project supporting the workshops leading to generation of this report was funded by Grant #NA16RG0320-02 from the Coastal Ocean Program of the Naitonal Oceanic and Atmospheric Administration to the University of North Carolina Sea Grant College Program. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies.

The authors would also like to thank Heather Cheshire, John Brockhaus, Charlotte Swart, and Linda Babcock of Computer Graphics Center at North Carolina State University for their contributions to this report.

9.0 REFERENCES

Anderson, J. R., Hardy, E.E., Roach, J.T., and R. E. Witmer, 1976. "A Land Use and Land Cover Classification System for Use with Remote Sensor Data," Professional Paper 964. Washington, D.C.: U.S. Geological Survey.

Aronoff, S., 1985. "The Minimum Accuracy Value as an Index of Classification Accuracy," Photogrammetric Engineering and Remote Sensing, 51(1):99-111.

Augenstein, E.W., Stow, D.A., and A.S. Hope, 1991. "Evaluation of SPOT HRV-XS Data for Kelp Resource Inventories," Photogrammetric Engineering and Remote Sensing, 57(5):501-509.

Baker, J.R., Briggs, S.A., Gordon, V., Jones, A.R., Settle, J.J., Townsend, J.R., and B.K. Wyatt, 1991. "Advances in Classification for Land Cover Mapping Using SPOT HRV Imagery," International Journal of Remote Sensing, 12(5):1071-1085.

Biging, Greg S., Congalton, R.G., and E.C. Murphy, 1991. "A Comparison of Photo interpretation and Ground Measurements of Forest Structure," Proc., of the Fifty Seventh Annual Meeting of the American Society of Photogrammetry and Remote Sensing, Biltimore, MD, March 1991, 3:6-15.

Bishop, Y.M.M., Feinberg, S.E., and P.W. Holland, 1975. Discrete Multivariate Analysis-Theory and Practice, Cambridge, MA:MIT Press, 575 p.

Burgess, W., Christoffers, E., Dobson, J.E., Ferguson, R.L., Frisch, A., Lade, K.P., Thomas, J.P., and B. Wilen, 1991. "Results of a Field Reconnaissance of Remotely Sensed Land Cover Data," Maryland Department of Natural Resources, Annapolis, MD.

Campbell, J.B., 1987. Introduction to Remote Sensing, N.Y.: Guilford Press, 551 p.

Caselles, V. and M.J.L. Garcia, 1989. "An Alternative Simple Approach to Estimate Atmospheric Correction in Multitemporal Studies," International Journal of Remote Sensing, 10(6):1127-1134.

Chrisman, N.R., 1989. "Error in Categorical Maps: Testing Versus Simulation," Proc., AUTO-CARTO 9, pp. 521-529.

Chrisman, N. R., and M. K. Lester, 1991. "A Diagnostic Test for Error in Categorical Maps," Proc., AUTO-CARTO 10, pp. 330-348.

Chrisman, N.R., 1991a. The Error Component in Spatial Data in Maguire, David J.; Goodchild, Michael F.; and Rhind, David W., Geographical Information Systems: Principles and Applications. New York: John Wiley & Sons.

Chrisman, N.R., 1991b. " A Diagnostic Test for Error in Categorical Maps," Proc., AUTO-CARTO 10, pp. 330-348.

Chrisman, N.R., 1993. "Beyond the Snapshot: Changing the Approach to Change, Error and Process," Personal Communications.

Chuvieco, E., and R. Congalton. 1988. "Using Cluster Analysis to Improve the Selection of Training Statistics in Classifying Remotely Sensed Data," Photogrammetric Engineering and Remote Sensing, 54(9):1275-1281.

Cochran, W.G., 1977. Sampling Techniques, John Wiley & Sons, Inc., New York.

Cohen, J.A., 1960. "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, 20, pp. 37-46.

Congalton, R.G., and R.A. Mead, 1983. "A Quantitative Method to Test for Consistency and Correctness in Photo interpretation," Photogrammetric Engineering and Remote Sensing, 49(1):69-74.

Congalton, R.G., 1988. "Using Spatial Autocorrelation Analysis to Explore the Errors in Maps Generated from Remotely Sensed Data," Photogrammetric Engineering and Remote Sensing, 54(5):587-592.

Congalton, R.G., 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data," Remote Sensing of Environment, 37:35-46.

Congalton, R. and G. Biging, 1992. "A Pilot Study Evaluating Ground Reference Data Collection Efforts for Use in Forest Inventory," Photogrammetric Engineering and Remote Sensing, 58(12):1669-1671.

Congalton, R.G. 1993. "Accuracy Assessment of Remotely Sensed Data: Future Needs and Directions," Proc., Pecora 12 Symposium, Sioux Falls,

Congalton, R. and K. Green, 1993. "A Practical Look at the Sources of Confusion in Error Matrix Generation," Photogrammetric Engineering and Remote Sensing, 59(5):641-644.

Corr, D.G., Tailor, A.M., Cross, A., Hogg, D.C., Lawrence, D.H., Mason, D.C., and M. Petrou, 1989. "Progress in Automatic Analysis of Multi-Temporal Remotely-Sensed Data," International Journal of Remote Sensing, 10(7):1175-1195.

Cowardin, L.M., Carter, V., Golet, F.C., and E.T. LaRoe, 1979. "Classification of Wetlands and Deepwater Habitats of the United States," Washington: U.S. Fish and Wildlife Service, FWS/OBS-79/31, 103 p.

Cross, F.A. and J.P. Thomas, 1992, "CoastWatch Change Analysis Project (C-CAP) Chesapeake Bay Regional Project," Proc., ASPRS/ACSM 92 Vol. 1, Global Change and Education, Bethesda MD., p. 57.

Crowell, M., Leatherman, S.P., and M.K. Buckley, 1991. "Historical Shoreline Change: Error Analysis and Mapping Accuracy," Journal of Coastal Research, 7(3):839-852.

Czaplewski, R.L. and G.P. Catts, 1992. "Calibration of Remotely Sensed Proportion or Area Estimates for Misclassification Error," Remote Sensing of Environment, 39: 29-43.

Daiber, F.C., 1986. Conservation of Tidal Marshes, N.Y.: Van Nostrand Reinhold Inc., 341 p.

Dobson, J.E. and E.A. Bright, Forthcoming. "Large-Area Change Analysis: the CoastWatch Change Analysis Project (C-CAP)," Proc., Pecora 12 Conference.

Dobson, J.E., Bright, E.A., Ferguson, R.L., Field, D.W., Wood, L.L., Haddad, K.D., Jensen, J.R., Iredale,H., Klemas,V.V., Orth, R.J., and J.P. Thomas, 1993. "NOAA CoastWatch Change Analysis Project: Guidance for Regional Implemental," NOAA Coastal Ocean Program, (In Press).

Eckhardt, D.W., Verdin, J.P., and G.R. Lyford, 1990. "Automated Update of an Irrigated Lands GIS Using SPOT HRV Imagery," Photogrammetric Engineering and Remote Sensing, 56(11):1515-1522.

Ellis, J., 1978. Coastal Mapping Handbook, Washington: U.S. Government Printing Office, 200 p.

Ferguson, R.L., and L.L. Wood, 1990. " Mapping Submerged Aquatic Vegetation in North Carolina with Conventional Aerial Photography," In Federal Coastal Wetland Mapping Program. (S.J. Kiraly, F.A. Cross and J.D. Buffington, eds), U.S. Fish Wild. Serv. Biol. Rep. 90 (18):125-133.

Ferguson, R.L., Wood, L.L., and D.B. Graham, 1993. "Monitoring Spatial Change in Seagrass Habitat with Aerial Photography," Photogrammetric Engineering and Remote Sensing 59:1033-1038.

Foody, G. M., 1992. "A Fuzzy Sets Approach to the Representation of Vegetation Continua from Remotely Sensed Data: An Example from Lowland Hearth," Photogrammetric Engineering and Remote Sensing, 58(2):221-225.

Ginevan, M.E., 1979. "Testing Land-Use Map Accuracy: Another Look," Photogrammetric Engineering and Remote Sensing, 45(10):1371-1377.

Gong, P. and P.J. Howarth, 1990. "An Assessment of Some Factors Influencing Multispectral Land-Cover Classification Techniques," Photogrammetric Engineering and Remote Sensing, 56(5):597-603.

Goodchild, M.F., 1978. Statistical Aspects of the Polygon Overlay Problem in Dutton, G. (ed.) Harvard Papers of GIS, Addison-Wesley, Reading MA.

Goodchild, M.F., and O. Dubuc, 1987. "A Model of Error for Choropleth Maps with Applications to GIS," Proc., Auto-Carto 8, pp. 165-174.

Goodchild, M.F., Sun, G.Q., and S. Yang, 1992. "Development and Test of an Error Model for Categorical Data," International Journal of Geographical Information Systems, 6(2):87-104.

Haley, G.M., 1985. "Multispectral Change Detection Using Difference Classification and Bitemporal Classification, Proc., Eleventh International Symposium on Machine Processing of Remote Sensing Data," West Lafayette, IN., June 25-27, pp. 26-33.

Hall, F. G., Strebel, D. E., Nickeson, J.E., and S.J. Goetz, 1991. "Radiometric Rectification: Toward a Common Radiometric Response Among Multidate, Multisensor Images," Remote Sensing of Environment, 35:11-27.

Hartshorne, R., 1939. The Nature of Geography: A Critical Survey of Current Thought in Light of the Past, Association of American Geographers: Lancaster, PA.

Hellden, U. and M. Stern, 1980. "Evaluation of Landsat Imagery and Digital Data for Monitoring Decertification Indicators in Tunisia," Proc., Fourteenth International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Vol. III, pp. 1601-1610.

Hodgson, M.E. and R.W. Plews, 1989. "N-Dimensional Display of Cluster Means in Feature Space," Photogrammetric Engineering and Remote Sensing, 55(5):613-619.

Hord, R.M. and W. Brooner, 1976. "Land-Use Map Accuracy Criteria," Photogrammetric Engineering and Remote Sensing, 42(5):671-677.

Hudson, W.D., and C.W. Ramm, 1987. "Correct Formulation of the Kappa Coefficient of Agreement," Photogrammetric Engineering and Remote Sensing, 53(4):421-422.

Jensen, J.R., 1986. Introductory Digital Image Processing, Printice Hall, NJ, 389 p.

Jensen, J.R. and D. Toll, 1982. "Detecting Residential Land-Use Development athe the Urban Fringe," Photogrammetric Engineering and Remote Sensing, 48(4):629-643.

Jensen, J.R., Ramsey, E.W., Mackey, H.E., Christensen, E., and R. Sharitz, 1987. "Inland Wetland Change Detection Using Aircraft MSS Data," Photogrammetric Engineering and Remote Sensing, 53(5):521-529.

Khorram, S., Brockhaus, J.A., and A.L. Geraci, 1991. "A Regional Assessment of Land Use/Land Cover Types in Sicily with TM Data," International Journal of Remote Sensing, 12(1):69-78.

Khorram, S., Brockhaus, J.A., and H.M. Cheshire, 1992a. "Merger of Independently Derived Wetland Inventory Data Bases," Proc., 1992 ASPRS/ACSM Convention on Monitoring and Mapping Global Change, Washington, D.C., Vol. 1, pp. 111-119.

Khorram, S., Cheshire, H.M., Sidrellis, K., and Z. Nagy, 1992b. "Mapping and GIS Development of Land Use/Land Cover Categories for the Albemarle-Pamlico Drainage Basin," NC Dept. of Env., Health, and Natural Resources, Report No. 91-08, 55 p.

Klemas, V.V., Dobson, J.E., Ferguson, R.L., and K.D. Haddad, 1993. "A Coastal Land Cover Classification System for the NOAA CoastWatch Change Analysis Project," Journal of Coastal Research, 9(3):862-872.

Lunetta, R., Congalton, R., Fenstermaker, L., Jensen, J., McGwire, K., and L. Tinney, 1991. "Remote Sensing and Eographic Information System Data Integration: Error Sources and Research Issues," Photogrammetric Engineering and Remote Sensing, 57(6):677-687.

Mailing, D.H., 1989. Measurements from Maps, Principles and Methods of Cartometry, Pergamon Press, Oxford, 577 p.

Mausel, P.W.,. Kramber, W.J, and J.K. Lee, 1990. "Optimum Band Selection for Supervised Classification of Multispectral Data," Photogrammetric Engineering and RemoteSensing, 56(1):55-60.

McGwire, K.C., 1992. "Analyst Variability in Labeling of Unsupervised Classifications," Photogrammetric Engineering and Remote Sensing, 58(12):1673-1677.

National Institute of Standards and Technology, 1992. "Spatial Data Transfer Standard," FIPS Publication 173. Springfield, VA: National Technical Information Service.

NOAA Photogrammetry Branch, 1989. "Coastal Mapping Program Operations Manual," Section 3, Quality Program, Rockville, MD: NOAA.

Novak, J., 1992. "Rectification of Digital Imagery," Photogrammetric Engineering and Remote Sensing, 58(3):339-344.

Orth, R.J., Ferguson, R.L., and K.D. Haddad, 1991. "Monitoring Seagrass Distribution and Abundance Patterns," Coastal Wetlands Coastal Zone 1991 Conference, Long Beach CA.: American Society of Civil Engineers, pp. 281-300.

Price, K.P., Pyke, D.A., and L. Mendes, 1992. "Shrub Dieback in a Semiarid Ecosystem: The Integration of Remote Sensing and GIS for Detecting Vegetation Change," Photogrammetric Engineering and Remote Sensing, 58(4):455-463.

Roller, N.E.G. and L. Visser, 1980. "Accuracy of Landsat Forest Cover Type Mapping in the Lake States Region of the U.S.," Proc., Fourteenth International Symposium on Remote Sensing of Environment, San Jose, Costa Rica, pp. 1511-1520.

Rosenfield, G.H. and M.L. Melley, 1980. "Applications of Statistics to Thematic Mapping," Photogrammetric Engineering and Remote Sensing, 46(10):1287-1294.

Rosenfield, G.H., and K. Fitzpatrick-Lins, 1986. "A Coefficient of Agreement as a Measure of Thematic Classification Accuracy," Photogrammetric Engineering and Remote Sensing, 52(2):223-227.

Rutchey, K. and L. Velcheck, "Development of An Everglades Vegetation Map Using a SPOT Image and the Global Positioning System," Photogrammetric Engineering and Remote Sensing, in press.

Ryerson, R., 1989. "Image Interpretation Concerns for the 1990s and Lessons from the Past," Photogrammetric Engineering and Remote Sensing, 55(10):1427-1430.

Schott, J.R., Salvaggio, C., and W. J. Volchok, 1988. "Radiometric Scene Normalization Using Pseudoinvariant Features," Remote Sensing of Environment, 26:1-16.

Shapiro, C., 1993. "Coordination/Integration of Wetlands Data for Status and Trends and Inventory Estimates," (Wetlands Data Coordination Working Group, Wetlands Subcommittee, Federal Geographic Data Committee), Working draft. 2.0.

Singh, A., 1989. "Digital Change Detection Techniques Using Remotely Sensed Data," International Journal of Remote Sensing, 10(6):989-1003.

Sinton, D., 1978. The Inherent Structure of Information as a Constraint to Analysis: Mapped Thematic Data as a Case Study, Vol. 7, G. Dutton, ed., Harvard Papers on Geographic Information Systems, Addison Wesley.

Skidmore, A., and B. Turner, 1989. "Assessing the Accuracy of Resource Inventory Maps," Proc., Global Natural Resource Monitoring and Assessments: Preparing for the 21st Century, Venice, Italy, September, Vol. 2, pp. 524-535.

Slamma, C.C., 1980. "Manual of Photogrammetry," 4th Ed., Falls Church: American Society of Photogrammetry, 1056 p.

Stamp, D., 1948. The Land of Britain, its Use and Misuse, Longmans Green.

Story, M., and R.G. Congalton, 1986. "Accuracy Assessment: A User's Perspective," Photogrammetric Engineering and Remote Sensing, 52(3):397-399.

U.S. Geological Survey, National Mapping Program, 1992. "Standards for Digital Line Graphs for Land Use and Land Cover," Technical Instructions, Referral STO-1-2, (need page numbers)

Van Genderen, J.L., and B.F. Lock, 1977. "Testing Land-Use Map Accuracy," Photogrammetric Engineering and Remote Sensing, 43(9):1135-1137.

Weerackoon, R.D. and T. H. Mace, 1990. "A Method of Optimizing the Maximum Likelihood Classifier Using Mathematical Transformations," Proc., ACSM-ASPRS Annual Convention, Denver, Co., pp. 464-474.

Wickware, G.M. and P.J. Howarth, 1981. "Change Detection in the Peace-Athabasca Delta Using Digital Landsat Data," Remote Sensing of Environment, Vol. II, No. 1, pp. 9-25.

Wilen, B.O., 1990. "The U.S. Fish and Wildlife Service's National Wetlands Inventory," U.S. Fish and Wildlife Service Biological Report, 90(18):9-19.