NASA Technical Memorandum 101042

Accuracy Assessment, Using Stratified Plurality Sampling, of Portions of a Landsat Classification of the Arctic National Wildlife Refuge Coastal Plain

Don H. Card and
Laurence L. Strong

(NASA-TM-101042) ACCURACY ASSESSMENT, USING STRATIFIED PLURALITY SAMPLING, OF PORTIONS OF A LANDSAT CLASSIFICATION OF THE ARCTIC NATIONAL WILDLIFE REFUGE COASTAL PLAIN (NASA) 48 p CSCL 08B G3/43

N89-17339

Unclas
0191241

January 1989

NASA

National Aeronautics and
Space Administration

NASA Technical Memorandum 101042

Accuracy Assessment, Using Stratified Plurality Sampling, of Portions of a Landsat Classification of the Arctic National Wildlife Refuge Coastal Plain

Don H. Card, Ames Research Center, Moffett Field, California
Laurence L. Strong, TGS Technology, Inc., Moffett Field, California

January 1989

NASA

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

TABLE OF CONTENTS

SUMMARY

This report describes an application of a classification accuracy assessment procedure for a vegetation and land cover map prepared by digital image processing of Landsat multispectral scanner data. A statistical sampling procedure called Stratified Plurality Sampling was used to assess the accuracy of portions of a map of the Arctic National Wildlife Refuge coastal plain (Walker et al. 1982). Results are tabulated as percent correct classification overall as well as per category with associated confidence intervals. Although values of percent correct were disappointingly low for most categories, the study was useful in highlighting sources of classification error and demonstrating shortcomings of the plurality sampling method.

INTRODUCTION

With the development of the Landsat satellite series, the availability of vegetation and land cover maps of large areas has greatly increased. The satellite's synoptic view, and high temporal frequency can provide timely high resolution imagery that can contribute to improved understanding of the spatial structure of vegetation. However, the use of spectral reflectance to classify vegetation types is subject to the errors inherent in all remote sensing techniques, so effective use of the vegetation maps requires estimates of their accuracy. There are two major obstacles to an easy accuracy assessment of vegetation maps; the true class identity is not only difficult to obtain, but is often difficult even to define and vegetation cover is highly variable, both in space and in time. Total enumeration of errors of classification either by ground checking or by comparison with an independent map base is impossible; in the first case, because of the enormous cost it would entail, and in the second case, available maps are usually either out of date or nonexistent, because the original purpose of the mapping project was to develop improved maps. This means that to assess the accuracy of the map, field checking, either by random sampling or otherwise, is necessary.

Literature on the efficiency of different sampling designs for accuracy assessment of Landsat-derived vegetation maps is meager (Mead and Szajgin 1982); a recent study by Congalton (1988) discussed bias in estimates of accuracy for a few sampling designs, but did not delve into questions of precision. Most of the literature on accuracy assessment addresses the use of the binomial probability distribution for the analysis of data collected as a simple random sample or a stratified random sample of single pixels (Card 1982; Fitzpatrick-Lins 1980). In practice, it is often difficult to locate single pixels in the field for comparison with the vegetation map, and as a result cluster sampling has been recommended by some investigators (Pettinger 1982; Todd, Gehring, and Haman 1980; Forbes, Fox, and Mayer 1980). Unfortunately, investigators have often failed to realize that the analysis of the data is dependent upon the sampling scheme used to collect the data, and they have incorrectly used statistical formulas appropriate to single pixel, simple random sampling for proportions, particularly with regard to sample size determinations for a specified precision (Card 1982). Cluster sampling should be advantageous relative to single pixel sampling when the cost to locate sample units is greater than the cost to identify the vegetation type at the individual pixel level (Cochran 1977), although no demonstration of this is available in the remote sensing literature.

Stratified random sampling, either of clusters or single pixels, in other contexts has been shown to be more efficient than simple random sampling when the sampling variable has strata means which vary significantly between strata, because the sampling results in estimates that are weighted according to relative strata sizes. It stands to reason that similar efficiencies might be realized in sampling for accuracy assessment. This is a source of some confusion in the accuracy assessment literature as it is often assumed that stratifying by proportion of vegetative type is sufficient. However, because the variable of interest is accuracy (proportion correct), and not proportion of type, the stratification may not be appropriate for accuracy assessment. Since the appropriate stratification variable, accuracy, is unknown, it is difficult to suggest ways to stratify a classification map for accuracy evaluation. This chicken-or-egg situation is common in stratified sampling problems and is discussed by Cochran (1977, p. 101). It seems to be particularly difficult in this context because accuracy is not obviously correlated with any observable variables related to Landsat imagery.

The subject of this report, plurality cluster sampling, is one attempt to deal with this problem by letting the classification results drive the sampling; that is, the strata are formed by partitioning the classified image, rather than the relative radiance image. Linden and Szajgin (1981) seem to have been the first to discuss this use of cluster sampling in stratified designs to estimate land cover map accuracy. They coined the term "stratified plurality sampling" (SPS) to describe the approach. In SPS, clusters of pixels are assigned to strata on the basis of their most frequent map category; e.g., if most of the pixels (a plurality) in a cluster belong to category i according to the classified image, the cluster is assigned to plurality stratum i. These strata then provide the population from which a stratified sample is taken and field checked for accuracy. Linden and Szajgin (1981) recommended the SPS method for estimating commission errors for individual land cover classes. The motivation behind this procedure is that pixels in low-occurrence classes are more likely to be sampled in their plurality class than they would be in ordinary stratified cluster sampling, where the strata are usually selected by visual inspection of imagery or base maps.

Our thanks to Len Gaydos, for supervising the project; to William Acevedo for collecting field data and conducting photointerpretation of the accuracy assessment sample units; and to Susan Benjamin for image processing in support of plurality sampling and preparing photos for field work.

PROJECT BACKGROUND

Section 1002c of the Alaska National Interest Lands Conservation Act (ANILCA) provides for baseline and continuing study of wildlife and their habitats on the coastal plain of the Arctic National Wildlife Refuge (ANWR).The spatial distribution of vegetation and land cover is an important component of wildlife habitats. Because of its large size and remote location, the lack of detailed existing vegetation and land cover maps for the ANWR and a 2-year period in which to prepare an Environmental Impact Statement, a 1:250,000-scale vegetation and land cover map of the coastal plain of the ANWR was prepared from digital image processing of portions of three Landsat Multispectral Scanner (MSS) scenes (Walker et al. 1982). Cooperators in preparation of the vegetation and land cover map included the Institute for Arctic and Alpine Research (INSTAAR), University of Colorado; Geographic Investigations Office (GIO), Western Mapping Center, U.S. Geological Survey; Institute of Polar Studies, Ohio State University; Earth Sciences Branch, Research Division, U.S. Army Cold Regions Research and Engineering Laboratory (CRREL); and Technology Applications Branch, Aircraft Missions and Applications Division, NASA Ames Research Center.

As part of the continuing study of wildlife habitats of the ANWR provided for by Section 1002c of ANILCA, the U.S. Fish and Wildlife Service (USFWS), CRREL, and INSTAAR agreed to conduct an accuracy assessment of the Landsat-derived vegetation and land cover map and to collect species lists and environmental data for permanent study sites in the summer of 1982. INSTAAR and the USGS GIO agreed to cooperate in the accuracy assessment.

This paper reports the use of SPS sampling design for estimating overall, commission, and omission errors at six study areas on the ANWR vegetation and land cover map, the methods used to collect the field data, the results of the comparison of ground data with the Landsat-derived vegetation and land cover map, and recommendations for future accuracy assessment.

METHODS

Classification of Landsat Data

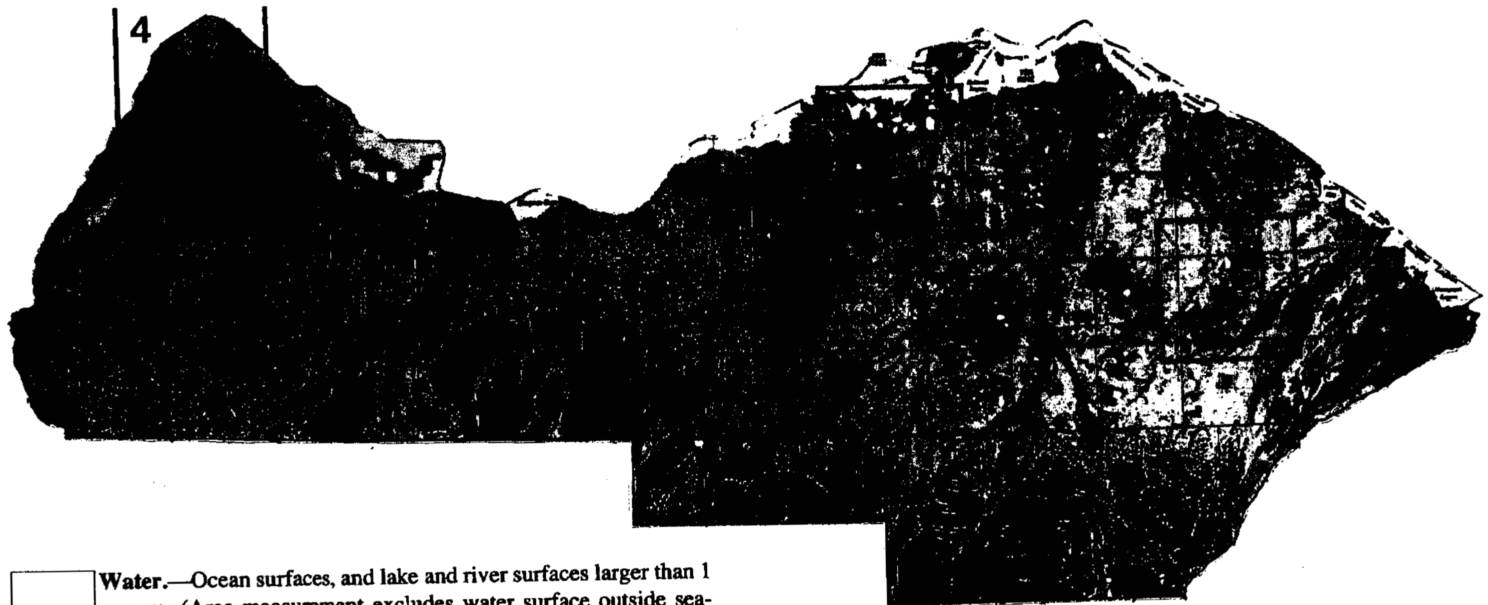
A land cover classification of 5700 km^2 of the coastal plain of the ANWR was prepared as part of an Environmental Impact Statement in anticipation of seismic oil exploration. The land cover classification recognized 12 map categories (fig. 1) as a result of a 7-day reconnaissance survey in August 1981 that concentrated on four townships within the coastal plain of the ANWR. Ground reference data, classification system, and Landsat data processing and classification procedures used in preparation of the land cover classification are described in Walker et al. (1982). Digital image data used in the accuracy assessment was the land cover classification before spatial filtering for cartographic presentation as USGS map MPI-1443, Vegetation and Land Cover, Arctic NWR, Coastal Plain, Alaska.

Sample Design

Because of travel cost constraints, six study areas representing four terrain types (thaw-lake plains, river floodplains, hilly coastal plains, and foothills) on the ANWR coastal plain (see Walker et al. 1982) were selected for this accuracy study (fig. 1). This kind of sampling is subjective rather than random, and therefore does not result in unbiased estimates of accuracy of known precision for the map as a whole. It was decided that, given the resources available, estimates of classification accuracy for vegetation and land cover within the different terrain types would be useful in evaluating the map. An additional complicating factor in the interpretation of the results is that four of the six study areas (sites 1, 2, 3, and 4) were used in labeling spectral classes and therefore accuracy for those areas is probably inflated above what would be expected for study areas independent of training areas. However, this does not affect the validity of accuracy estimates for those four areas per se, since within each of the six study areas probability sampling was used.

The sampling method employed in each study area was single stage cluster sampling with plurality stratification. Single stage cluster sampling is well documented in the statistical literature (Cochran 1977; Hansen, Hurwitz, and Madow 1953). The idea of plurality stratification does not seem to have been documented in the published statistical literature, and was apparently first discussed by Linden and Szajgin (1981) and Szajgin et al. (1982). It is simply a special kind of stratification of clusters in which strata are defined not by geographical location, but by the dominant cover type as determined by the classification. More concretely, a cluster is assigned to the stratum associated with a cover class if the number of pixels in that class and in that cluster, as determined from the classification image, is a plurality. For example, if the class or category having the most pixels in the cluster is water, the cluster is assigned to the water-plurality stratum. Ties are broken by an arbitrary rule. This usage of the term plurality is not quite accurate according to the dictionary definition, since if the dominant proportion is greater than a half, the term majority should be used. However, we will use the term plurality in both cases to be consistent with its usage in Linden and Szajgin (1981).

The purpose behind this kind of stratification is to ensure, by independent random sampling of clusters within plurality strata, that a sufficient number of pixels of rare cover types occur in the sample for accuracy estimates to be made for the rare categories. Obviously, this approach will work well only if there is a sufficient number of clusters having the rare category pixels in plurality, and if commission errors for the rare classes is not excessive. It is usually not possible to check these assumptions in an operational setting, since they depend on ground data, and this is not available prior to sampling.



1 **Water.**—Ocean surfaces, and lake and river surfaces larger than 1 acre. (Area measurement excludes water surface outside seaward limit of study area.)

2 **Pond/Sedge Tundra Complex; Aquatic Tundra; or shallow water.**—Very wet tundra areas with ponds and/or emergent communities of *Carex* spp. or *Arctophila;* and up to 50% moist or wet tundra.

3 **Wet Sedge Tundra.**—Wet tundra with little standing water or up to half of surface area water-covered or emergent vegetation, or coastal areas periodically covered with salt water.

4 **Moist/Wet Sedge Tundra Complex; or Dry Prostrate Shrub, Forb Tundra (*Dryas* river Terraces).**—Moist sedge tundra with up to 40% wet sedge tundra; or dense prostrate mat of *Dryas* on river terraces.

5 **Moist Sedge, Prostrate Shrub Tundra; or Moist Sedge/Barren Tundra Complex (frost-scar tundra).**—Better drained areas on rolling terrain sometimes with tussocks; or sparsely vegetated frost-scar tundra.

6 **Moist Sedge Tussock, Dwarf Shrub Tundra.**—Well drained upland tussock tundra in foothills with high percentage of cotton-grass tussocks and dwarf or prostrate shrubs.

7 **Moist Dwarf Shrub, Sedge Tussock Tundra; or Moist Sedge Tussock, Dwarf Shrub/Wet Dwarf Shrub Complex (water track complex).**—Upland tundra with shrubs to 50 cm high; or upland tussock tundra with shrubs in water tracks.

8 **Shrub-Tundra.**—South-facing slopes in foothills or subalpine, with willow, birch, alder to 2m; or dense shrubs in water tracks.

9 **Partially vegetated areas.**—Diverse habitats including river bars, alpine tundra and moss mats with barren rock and talus, lichen-covered, sorted stone-nets and beach or mud flats.

10 **Barren gravel or rock.**—Bare light-colored river gravel, gravel and sand spits, alpine barrens (especially dolomite), and cultural barrens (road or runway), often with rich but sparse floras.

11 **Wet gravel or mud.**—Extensive barren mud in river deltas and wet or dark-colored gravel on beaches or river beds, or dark-colored barren rock in mountains.

12 **Ice.**—River icings in the braided stream channels of most larger rivers.

Figure 1. Land cover classification of Arctic National Wildlife Refuge coastal plain by Walker et. al. (1982).

However, after sampling, if sufficient ground data are available, the validity of these assumptions can be examined if not statistically tested.

A sampling frame consisting of a grid of mutually exclusive and exhaustive clusters was defined for each of the six study areas as follows: a 1:60,000-scale photograph covering the study area was selected, and the corresponding area was extracted from the classified digital image. The image was gridded into a regular array of blocks (clusters) of 25 pixels each, five on a side. The gridding of the image and counting of pixels in each category within the clusters to define plurality strata was performed by a set of programs within the Interactive Digital Image Manipulation System (IDIMS) called SAMPLET (ESL 1988). Plurality class formation and random sampling were performed by a set of programs developed by D. Linden to augment SAMPLET (personal communication).

The number of sample clusters was allocated to plurality strata in an ad hoc manner, rather than according to statistical criteria. Unfortunately, several examples of sample size determination for cluster sampling described in the literature (Pettinger 1982; Todd, Gehring, and Haman 1980) are incorrect, because they are based on statistical formulas appropriate to single pixel, simple random sampling for proportions. One possible approach to sample size determination for cluster sampling based on "design effect" variables is described by Cochran (1977); however, this method requires knowledge of per-class accuracy for a pilot sample of data from previous similar accuracy studies, and was not considered feasible in this study. Therefore, the number of sample clusters allocated to each plurality class was based pragmatically on available helicopter time. One day of helicopter time was available for each study area, and it was estimated that a maximum of 30 clusters could be visited in a day. Allocation of the 30 clusters to plurality strata was based roughly on the total number of clusters in the stratum and the importance of the category defining the stratum (table 1).

TABLE 1.– SAMPLE AND POPULATION SIZES FOR CLUSTER BLOCKS BY SITE AND PLURALITY STRATA.

Plurality strata	Site 1		Site 2		Site 3		Site 4		Site 5		Site 6	
	Population	Sample	Population	Sample	Population	Sample	Population	Sample	Population	Sample	Population	Sample
1	2		37		1		729		578			
2	0		5	3	0		54	4	65	4		
3	4	2	894	4	1		1741	4	1427	4	77	1
4	495	4	755	4	24	4	484	4	1535	4	287	2
5	1299	2	2373	4	134	4	0		600	4	529	1
6	971	5	0		2778	4	0		0		2251	2
7	110	5	0		687	4	0		0		120	2
8	1	1	0		19	4	0		0		38	3
9	0		33	4	10	4	42	4	135	4	358	3
10	142	4	56	3	4		27	4	77	4	132	1
11	0		0		0		280	4	206	4	3	2
12	0		5		0		3		0		24	
Total	3024	23	4158	22	3657	24	3360	24	4623	28	3819	20

PRECEDING PAGE BLANK NOT FILMED

PAGE 6 INTENTIONALLY BLANK

Field Investigations

In preparation for field work, line printer maps of the vegetation and land cover classification were superimposed, using a Zoom Transfer Scope, on 9- by 9-in., 1:60,000-scale color infrared aerial photographs and sample clusters were delineated on mylar overlays. Sample cluster locations were visually located from a helicopter using the 1:60,000-scale CIR photographs. As the helicopter circled the site, a sketch map for each 25-pixel cluster was prepared which showed the relative locations of observer-identified land cover categories (fig. 2). Category labels for polygons delineated on the sketch map were determined from ground observations of species composition, and ocular estimates of canopy coverage by plant growth forms and relative amounts of standing and subsurface water. Identification of boundaries for individual pixels was impossible, and this prevented determining category labels on a pixel-by-pixel basis in the field.

Laboratory Studies

To determine boundaries of individual pixels within each cluster, enlargements were obtained of the 1:60,000-scale photographs, the typical scale being 1:15,250. Line printer maps were generated at the appropriate scale and overlaid on the enlargements. Each cluster was located on the enlargement using patterns seen on both line printer maps and photographs in the local neighborhood of the cluster. Comparison of the location of clusters on the 1:60,000-scale photographs and the 1:15,250-scale enlargements showed that some cluster locations were different between the two photographs. Where clusters occurred adjacent to water or other features with distinct signatures and patterns, the location of clusters between photographs was similar. In contrast, the Jago River study area (fig. 1) had little water to aid registration and significant location errors occurred when overlaying line printer maps on the 1:60,000-scale photographs using the Zoom Transfer Scope. In the extreme case, the location of two clusters differed about 800 m between the 1:60,000-scale photographs and the enlargements.

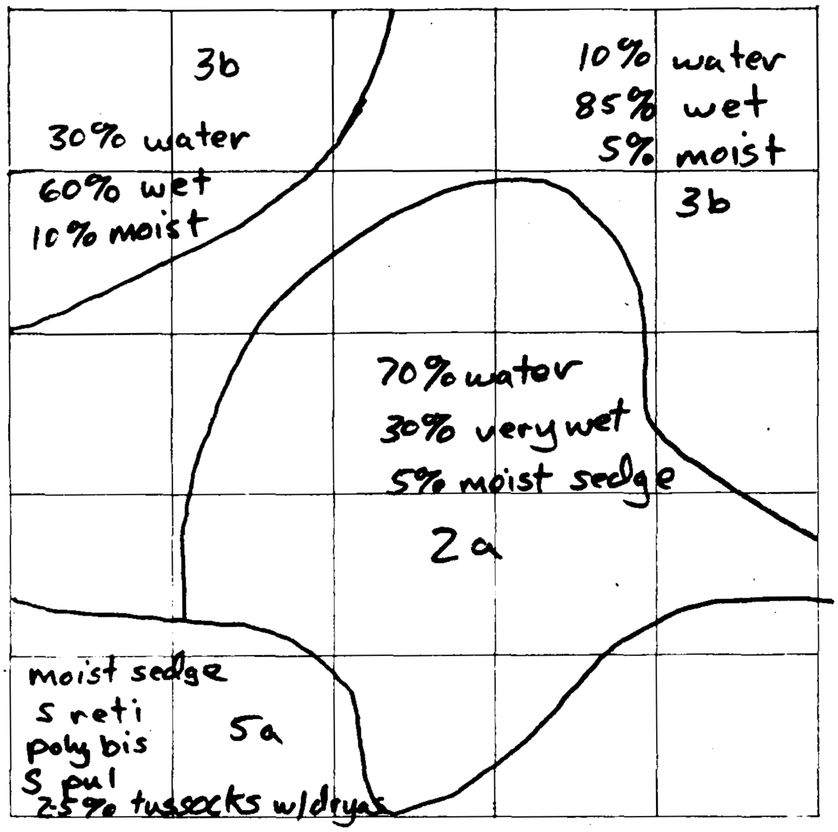
A 5- by 5-pixel grid was placed over the cluster boundary delineated on the enlargement. The sketch map and ground observations were used to identify the land cover category for each pixel within the cluster. If the pixel was a mixture of land cover categories, the observer noted the dominant categories (greater than 25% of the pixel area) and assigned the pixel to the category occupying the greatest area. For clusters with large differences in their location on the enlargement and 1:60,000-scale photographs, field notes of clusters having similar photo characteristics were used to guide the labeling of individual pixels.

Accuracy was estimated by comparing the classified image land cover category designation of each pixel in each sample cluster with the "true" designation of the pixel. The true designation was determined using a combination of on-site ocular estimation and aerial photo interpretation as described above. This definition of true implies that the verification data base (ocular/photo designation) is 100% accurate—which may not be the case.

RESULTS AND DISCUSSION

Classification error matrices for the six study areas are presented in tables 2-7. The diagonals in the error matrices contain the number of pixels for which Landsat and ground determinations of land cover category agree. Off-diagonal elements C (i,j) for row i and column j in the error matrix represent omission and commission errors: omission error for row category i and commission error for column category j. Interpretation of the errors is difficult because data are observational and not experimental; therefore bias is difficult to detect with the small number of occurrences. However, we suggest a number of explanations for observed errors which should be considered in planning future accuracy assessments.

OBSERVER: W DATE: 8-18 U2 PHOTO 2393
CLUSTER BLOCK: 6 CONFIDENCE IN LOCATION: 5
(1-5 SCALE)



1 OPEN WATER (a) OCEAN (b) LAKE (c) RIVER
2 (a) POND SEDGE TUNDRA COMPLEX (b) AQUATIC TUNDRA
3 (a) WET SEDGE TUNDRA, NON-COMPLEX (b) VERY WET (c) MOIST (d) SALINE
4 (a) MOIST/WET SEDGE TUNDRA COMPLEX (b) DRY PROSTRATE SHRUB, FORB TUNDRA
5 (a) MOIST SEDGE, PROSTRATE SHRUB TUNDRA (b) MOIST SEDGE/BARREN TUNDRA COMPLEX
6 MOIST TUSSOCK SEDGE, DWARF SHRUB TUNDRA (a) ACIDIC (b) ALKALINE
7 (a) MOIST DWARF SHRUB, TUSSOCK SEDGE TUNDRA (b) BIRCH TUNDRA (c) MOIST TUSSOCK SEDGE, DWARF SHRUB/WET DWARF SHRUB TUNDRA COMPLEX
8 (a) SHRUB TUNDRA, NON-COMPLEX (b) SHRUB TUNDRA, WATER-TRACK COMPLEX
9 PARTIALLY VEGETATED (a) RIVER BARS (b) ALPINE TUNDRA (c) SORTED STONE NETS (d) BEACHES (e) SAND DUNES
10 BARREN GRAVEL OR ROCK
11 BARREN MUD OR WET GRAVEL
12 ICE

Figure 2.– Relative locations and areas of land cover categories at a cluster block.

TABLE 2.– CLASSIFICATION ERROR MATRIX FOR LAND COVER CATEGORIES AT THE KATAKTURUK RIVER STUDY AREA.

GROUND-DETERMINED LAND COVER CATEGORY	LANDSAT MAP CATEGORY 1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1													0
2													0
3			22	12	4								38
4			4	45						5			54
5				55	83	21		2		3			164
6					11	81	31	1					124
7					1	33	72	2					108
8					1	3	4	9		1			18
9				2	1	1				30			34
10				1		1				33			35
11													0
12													0
TOTAL	0	0	26	115	101	140	107	14	0	72	0	0	575

TABLE 3.– CLASSIFICATION ERROR MATRIX FOR LAND COVER CATEGORIES AT THE NIGUANAK RIVER STUDY AREA.

GROUND-DETERMINED LAND COVER CATEGORY	LANDSAT MAP CATEGORY 1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1	17	4	6	2	1				9	4			43
2	6	29	11	3	3								52
3		1	61	25	12				14	1			114
4			28	47	47				2	7		1	132
5		1	6	36	87				16	20		1	167
6			2		2								4
7				6	4								10
8													0
9													0
10									14	12		1	27
11													0
12			1										1
TOTAL	23	35	115	119	156	0	0	0	55	44	0	3	550

TABLE 4.– CLASSIFICATION ERROR MATRIX FOR LAND COVER CATEGORIES AT THE JAGO UPLANDS STUDY AREA.

GROUND-DETERMINED LAND COVER CATEGORY	LANDSAT MAP CATEGORY 1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1				3	2				14	3			22
2													0
3			1	8	6								15
4			5	25	29	14			4				77
5				10	10	2			1				23
6				3	37	94	95	28		1			258
7				2	10	41	39	19		3			114
8					1	2		5		3			11
9				11	12	1			40	6			70
10					1				3	6			10
11													0
12													0
TOTAL	0	0	6	62	108	154	134	52	62	22	0	0	600

TABLE 5.– CLASSIFICATION ERROR MATRIX FOR LAND COVER CATEGORIES AT THE CANNING RIVER STUDY AREA.

GROUND-DETERMINED LAND COVER CATEGORY	LANDSAT MAP CATEGORY 1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1	11	10	5						2		7		35
2	7	37	15								2		61
3	2	14	80	40						16			152
4		1	39	73						14	1		128
5													0
6													0
7													0
8													0
9	1		12	1						24	5		43
10	1		1						5		16		23
11	4		2						64	3	85		158
12													0
TOTAL	26	62	154	114	0	0	0	0	71	57	116	0	600

TABLE 6.– CLASSIFICATION ERROR MATRIX FOR LAND COVER CATEGORIES AT THE OKPILAK COASTAL STUDY AREA.

GROUND-DETERMINED LAND COVER CATEGORY \ LANDSAT MAP CATEGORY	1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1	29	6							6	11	27		79
2		19	1	1									21
3		2	76	19	7				3				107
4	4	8	26	65	51	2			9		5		170
5			7	15	39				2				63
6													0
7													0
8													0
9	5	9	3	7	6				75	11	10		126
10									8	22	8		38
11	3		2	2					11	37	41		96
12													0
TOTAL	41	44	115	109	103	2	0	0	114	81	91	0	700

TABLE 7.– CLASSIFICATION ERROR MATRIX FOR LAND COVER CATEGORIES AT THE SADLEROCHIT SPRING STUDY AREA.

GROUND-DETERMINED LAND COVER CATEGORY	LANDSAT MAP CATEGORY 1	2	3	4	5	6	7	8	9	10	11	12	TOTAL
1				3	1				2	1			7
2													0
3			5	7	1				3				16
4			8	20					2				30
5			4	7	10	11		1	4				37
6					3	53	13	10	4				83
7						1	8						9
8			1	7	6	18	18	35					85
9	2		8	5	12	5	2	2	86	1	20		143
10	2								19	17	8		46
11	7		3						34				44
12													0
TOTAL	11	0	29	49	33	88	41	48	154	19	28	0	500

Classification errors occurred most frequently between adjacent land cover categories along moisture gradients and shrub dominance gradients which are apparent from examination of figure 1. For example, classification errors for land cover category 7 (Moist Dwarf Shrub, Sedge Tussock Tundra; or Moist Sedge Tussock, Dwarf Shrub/Wet Dwarf Shrub complex) were identified as category 6 (Moist Sedge Tussock, Dwarf Shrub Tundra) and category 8 (Shrub Tundra). Category 6 accounted for 92, 55, and 100% of the omission errors for category 7 at the Katakturuk, Jago, and Sadlerochit study areas, respectively. Category 8 accounted for 25% of the omission errors for category 7 at the Jago study area. Commission errors for category 7 were identified as land cover categories 6 and 8.

Classification errors for land cover category 2 (Pond/Sedge Tundra complex; Aquatic Tundra; or Shallow Water) were most frequently identified as category 3 (Wet Sedge Tundra) and category 1 (Water). Categories 3 and 1 accounted for 74, 92, and 50% of the omission errors for category 2 at the Niguanak, Canning and Okpilak study areas, respectively. Categories 3 and 1 accounted for 83, 96 and 32% of the commission errors for category 2.

Errors between adjacent land cover categories along moisture and shrub dominance gradients are to be expected. Land cover categories of these moisture and shrub dominance gradients contain the same vegetation elements; the distinction among the categories is based upon the relative proportions of the elements. As a result, these land cover categories are spectrally similar, i.e., represented in spectral space by probability density functions with large regions of overlap. It is important to realize that precise structural uniformity of vegetation between sites within a land cover type does not exist. The primary relationship between plants within an aggregate of plants recognized as a land cover type is their common occupancy of an area (Rowe 1961). This spatial criterion of land cover type is given little importance in determinations of land cover category on a pixel by pixel basis. For example, the abundance of water tracks, which can only be determined by an assessment of the spatial extent of the land cover category, is the primary criterion used to distinguish between land cover category 7, Moist Sedge Tussock, Dwarf Shrub/Wet Dwarf Shrub complex and category 6, Moist Sedge Tussock, Dwarf Shrub Tundra.

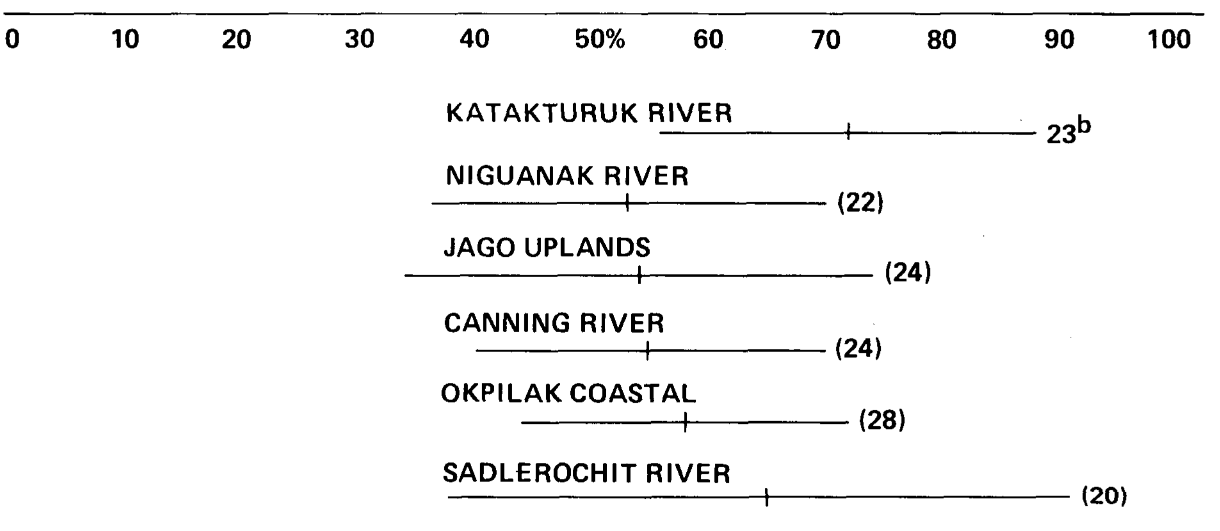
Similarly, spatial considerations are used to discriminate between land cover categories 2, Pond/Sedge Tundra complex; Aquatic Tundra; or Shallow Water and land cover categories 3 and 1, Wet Sedge Tundra and Water, respectively. All three land cover categories can have standing water and distinctions between cover types depend on the spatial interspersion of water and vegetation.

Observations for land cover category 9 (Partially Vegetated Areas) were obtained at all study areas. This map category contains a wide array of land cover types, including river bars, alpine tundra, sorted stone nets, beaches, and sand dunes. Omission and commission errors were common among land cover categories 9, 10 (Barren Gravel or Rock), and 11 (Wet Gravel or Mud). For the Katakturuk and Canning study areas, the omission error for category 9 was 100%. Category 10 accounted for 88 and 56%, respectively, of the omission errors for category 9 at the study areas. Commission errors for category 9 at the Canning River study area were identified as land cover category 11. These errors may be due to temporal changes in the river floodplains as the Landsat scene was acquired in 1979 and the ground data for the accuracy assessment were acquired in 1982. Temporal variation undoubtedly accounts for some of the omission and commission errors between land cover category 10 (Barren Rock or Gravel) and land cover category 11 (Wet Gravel or Mud) at the Canning and Okpilak study areas. These errors may also have resulted from inconsistent use of a vegetation canopy coverage criterion to distinguish between physiognomic descriptors, barren and partially vegetated.

Several land cover categories were misidentified as categories 9, 10, and 11. Some of these errors are due to the spatial complexity of riparian areas where land cover features are linear and have little width. For example, at the Jago River study area, the river is clearly observable on aerial photographs, but is seldom wider than 100 m. The river is omitted in the classified image presumably because of the small width of the river relative to the effective resolution of the scanner. Most of the omission errors for land cover category 1 (Water) at the Niguanak, Jago, Okpilak, and Sadlerochit study areas are due to a failure to discriminate rivers. Furthermore, the complex spatial juxtaposition of land cover patterns in the field prevents precise registration of land cover maps derived from Landsat data to aerial photographs. Omission errors for category 5 identified as categories 9 and 10 at the Niguanak study area are likely the result of complex spatial juxtaposition of land cover categories along the Niguanak and Sikrelurak rivers. Omission errors for category 8 identified as category 10 at the Jago study area are also believed to result from the complex spatial juxtaposition of land cover types along the Jago River.

Commission errors for category 4 (Moist/Wet Sedge Tundra Complex or Dry Prostrate Shrub, Forb Tundra) at the Niguanak, Jago, and Sadlerochit study areas identified from ground observations as land cover categories 6, 7, and 8 suggest another source of error. Land cover categories 6, 7, and 8 all have a dwarf shrub component of greater than 30% canopy coverage. Mike Spindler (USFWS-ANWR, personal communication) has identified areas of Wet Dwarf Shrub Tundra along the Jago River being mapped as category 4. This observation suggests that the description of land covers for category 4 is incomplete and should include a Wet Dwarf Shrub component.

Although the classification error matrices provide a useful overview of the results, showing the confusion between classes and the relationship between omission and commission errors, estimation of means and variances for the probability correct require statistical formulas which are dependent upon the sample design. The statistical formulas for estimating the mean and variance for a proportion correct from a simple random sample of clusters in stratified cluster sampling are presented in appendix A. Mean overall correct classification ranged from 53-72% for the six study areas (fig. 3).



(a) CONFIDENCE INTERVALS ARE BASED ON TCHEBYSHEFF'S INEQUALITY WHICH STATES THAT AT LEAST 75% OF THE MEANS FROM REPEATED SAMPLES WILL BE WITHIN THE INTERVAL REGARDLESS OF THE SHAPE OF THE DISTRIBUTION.

(b) CLUSTER SAMPLE SIZE.

Figure 3.– Means and confidence intervals[a] for percent overall classification agreement for six study areas on the Arctic National Wildlife Refuge coastal plain.

As discussed in the introduction, the formation of strata is a critical step in stratified sampling, and is especially so when one is dealing with multidimensional attributes; in our case classification accuracy for 12 land cover categories. In cluster sampling with plurality stratification, each cluster is assigned to a stratum on the basis of the most frequently occurring map category in the cluster. Strata developed by this method contain pixels belonging to more than one map category. Therefore, an estimate of each category in each stratum in which the category's pixels occurred is needed to estimate the accuracy of the land cover category.

For the estimation of commission errors, counts of map category pixels by stratum can be obtained from the digital land cover classification. The pixel counts are necessary to determine stratum weights for each land cover category. Several characteristics of cluster sampling with plurality stratification can be seen in table 8. First, in no case are all pixels belonging to the category contained in the stratum

TABLE 8.– ALLOCATION OF LAND COVER CATEGORY PIXELS TO STRATA FROM PLURALITY STRATIFICATION FOR TWO STUDY AREAS.

Land Cover Category*	Plurality Strata											
	Niguanak River Study Area											
	1	2	3	4	5	6	7	8	9	10	11	12
1	543	22	261	31	74	0	0	0	0	0	0	0
2	102	54	418	22	81	0	0	0	0	0	0	0
3	169	33	14833	2869	2287	0	0	0	53	46	0	4
4	52	4	4369	11374	8162	0	0	0	107	124	0	9
5	48	12	2118	4135	46653	0	0	0	143	304	0	21
6	0	0	0	0	78	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	11	0	264	222	772	0	0	0	359	157	0	7
10	0	0	18	113	1038	0	0	0	145	669	0	28
11	0	0	4	0	3	0	0	0	1	0	0	0
12	0	0	0	4	47	0	0	0	7	95	0	56
	Canning River Study Area											
1	14972	193	1124	185	0	0	0	0	2	32	791	22
2	684	692	1899	201	0	0	0	0	1	8	119	0
3	960	372	33251	2870	0	0	0	0	36	92	325	0
4	279	56	5697	8145	0	0	0	0	19	86	69	0
5	11	0	11	65	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	57	0	101	15	0	0	0	0	670	26	531	0
10	138	8	799	544	0	0	0	0	36	369	65	1
11	1072	29	643	75	0	0	0	0	286	62	5069	14
12	52	0	0	0	0	0	0	0	0	0	31	38

*According to Landsat.

corresponding to the category. Second, the majority of a map category's pixels are contained in two or three strata with a small percentage occurring in other strata. To estimate classification accuracy for a land cover category requires an estimate for the category in each stratum where the category's pixels occur. Random sampling from a stratum with a large population of clusters does not guarantee an estimate for each category. However with commission errors, selection of sample units can continue until an appropriate number of clusters for each land cover category occurring in the stratum has been obtained. To obtain estimates with good precision for all land cover categories occurring in the stratum would often result in sample sizes much larger than needed for the land cover category corresponding to the stratum.

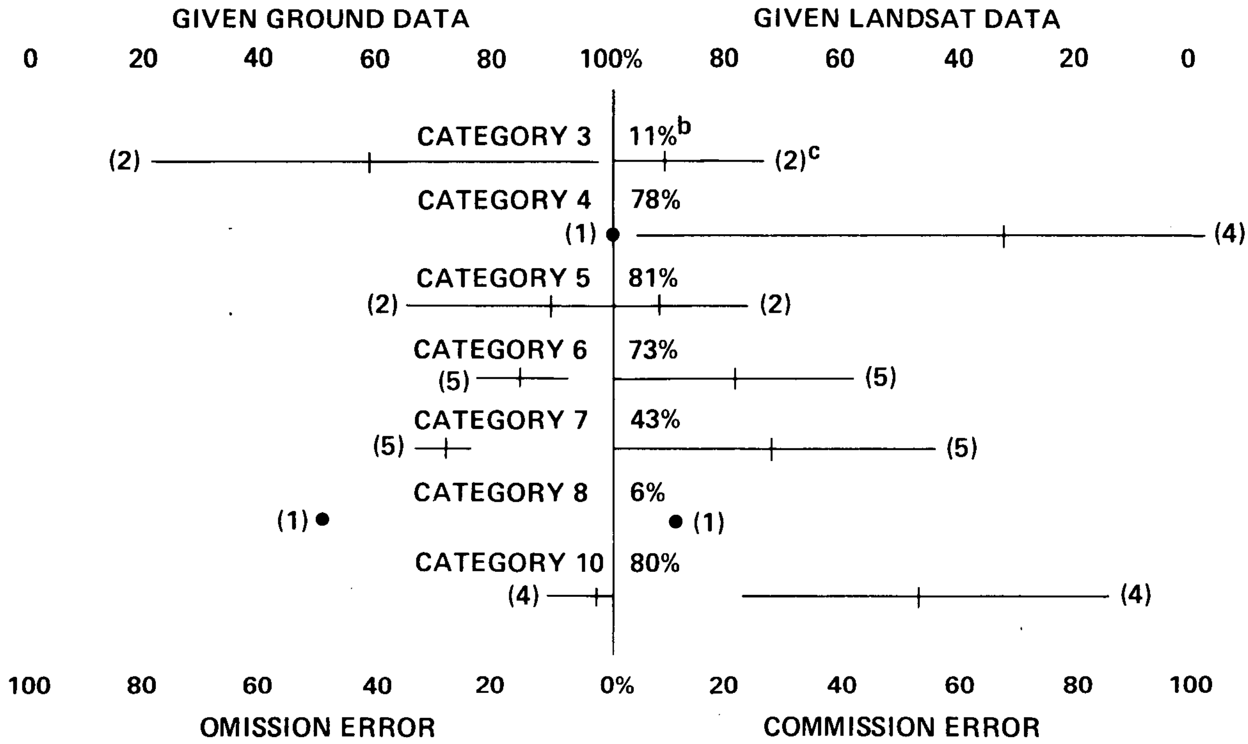
A serious shortcoming of plurality stratification occurs when the majority of a land cover category's pixels belong to a stratum other than its own. Examples of this problem for map category 2 can be found in table 8. The stratification in these cases is misleading. For greatest efficiency, the greatest sampling intensity for a category should be placed in the stratum where the majority of the category's pixels occur; however, in this case this is not the category's plurality stratum.

Obtaining stratified estimates of land cover category omission errors is even more difficult. As a map prepared from ground data is seldom available, strata weights must be determined from the sample data. Estimation of strata weights from small sample sizes may not reflect the true proportion of a category's pixels in each stratum. Furthermore, it is these weights which must be used to determine the importance of each stratum to each land cover category. Since the strata weights for omission errors can be calculated only after the sampling, little information is available to determine in which strata the sampling effort for a category should be concentrated.

Calculation of a stratified estimate for a land cover class often was not possible as estimates for the class in all strata which contained pixels of the class were seldom obtained with the small sample sizes of this study. Complete summaries of accuracy statistics for each land cover class in all strata at each study area are given in Appendix B as a set of tables. Means and confidence intervals for commission and omission errors for land cover categories in their plurality stratum are presented in figures 4-9. An important point should be made here. The estimates for a category in its plurality stratum are not valid for the population of pixels for the category. Figures 4-9 show that the proportion of a category's pixels that occurred in its plurality stratum ranged from only 6% for category 8 to 93% for category 6 at the Katakturuk and Sadlerochit study areas, respectively. The average percent of a category's pixels assigned to its plurality stratum ranged from 45 at the Niguanak study area to 56 at the Okpilak study area.

Category classification accuracy varied between study areas. For example, commission errors for category 9 ranged from 100% at the Niguanak and Canning study areas to 22% at the Okpilak study area. Classification accuracy appeared to vary by terrain types as sample clusters for category 9 were located in the hilly coastal plains and river flood plains terrain types at the Niguanak and Okpilak study areas, respectively. Other examples of classification accuracy varying by terrain types are present, but their detection in the sample statistics is prevented by the large variation associated with means and the stratification used in this study.

Design effects for comparing the relative efficiency of stratified cluster sampling and random cluster sampling for overall classification accuracy for the six study areas are presented in table 9. Design effect parameters are defined as the ratio of the variance of the study variable (in this case, overall proportion correct) for the more complex sampling method to that for a simple random sample (Cochran 1977). The design effects in table 9 show that plurality stratification did not reduce the variance for overall classification accuracy at any of the six study areas. In other words, the proportions of pixels that agreed were similar between strata. This was not surprising since the purpose of stratification was to increase the precision of individual category estimates. Also, cluster sampling was less efficient than single pixel sampling at all study areas. Hansen et al. (1953) emphasized that if cost were determined entirely by the number of elementary units in the sample, the design which would give the most precise results for a fixed number of elements would involve no clustering at all. However, cluster sampling may be more efficient than single pixel sampling if the cost to locate sample units is greater than the cost to identify the land

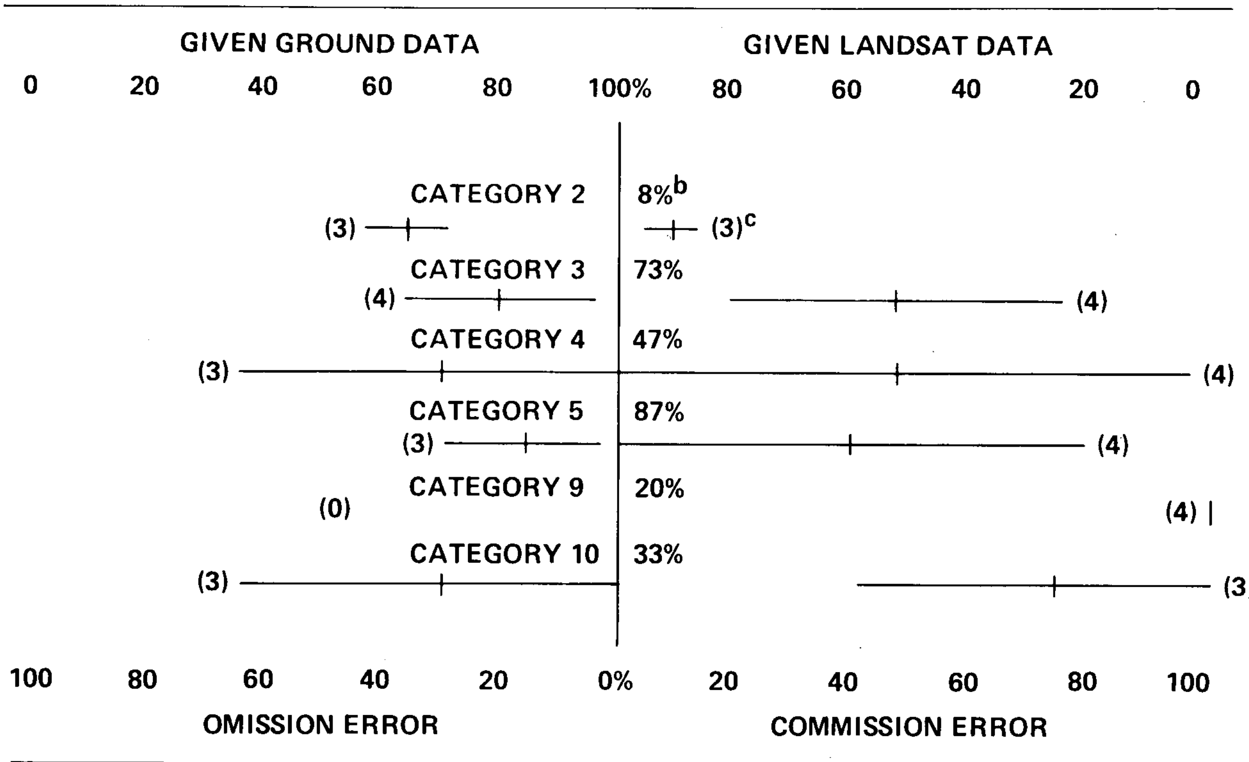


(a) CONFIDENCE INTERVALS ARE BASED ON TCHEBYSHEFF'S INEQUALITY WHICH STATES THAT AT LEAST 75% OF THE MEANS FROM REPEATED SAMPLES WILL BE WITHIN THE INTERVAL REGARDLESS OF THE SHAPE OF THE DISTRIBUTION.

(b) PERCENT OF THE MAP CATEGORY'S PIXELS WHICH OCCURRED IN THE MAP CATEGORY'S PLURALITY STRATUM.

(c) CLUSTER SAMPLE SIZE.

Figure 4.– Means and confidence[a] intervals for percent classification agreement for a category in its plurality stratum for the Katakturuk River study area.

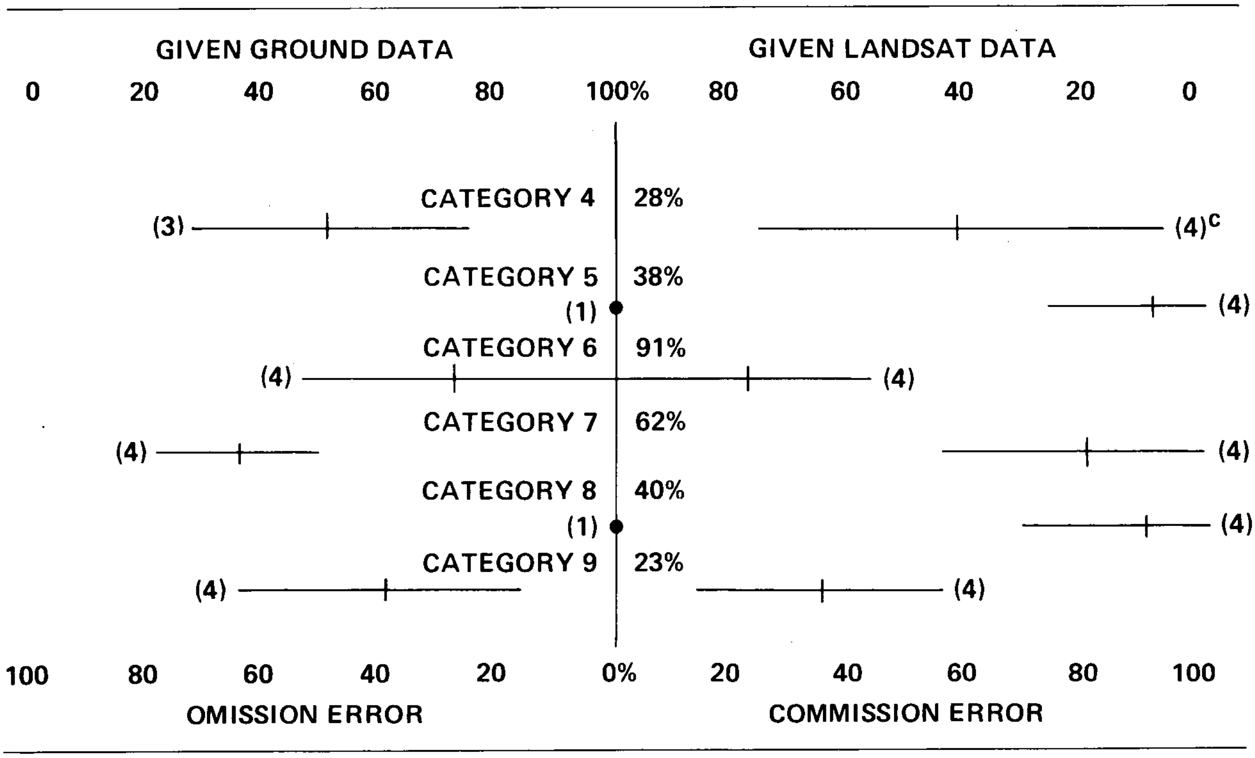


(a) CONFIDENCE INTERVALS ARE BASED ON TCHEBYSHEFF'S INEQUALITY WHICH STATES THAT AT LEAST 75% OF THE MEANS FROM REPEATED SAMPLES WILL BE WITHIN THE INTERVAL REGARDLESS OF THE SHAPE OF THE DISTRIBUTION.

(b) PERCENT OF THE MAP CATEGORY'S PIXELS WHICH OCCURRED IN THE MAP CATEGORY'S PLURALITY STRATUM.

(c) CLUSTER SAMPLE SIZE.

Figure 5.– Means and confidence[a] intervals for percent classification agreement for a category in its plurality stratum for the Niguanak River study area.

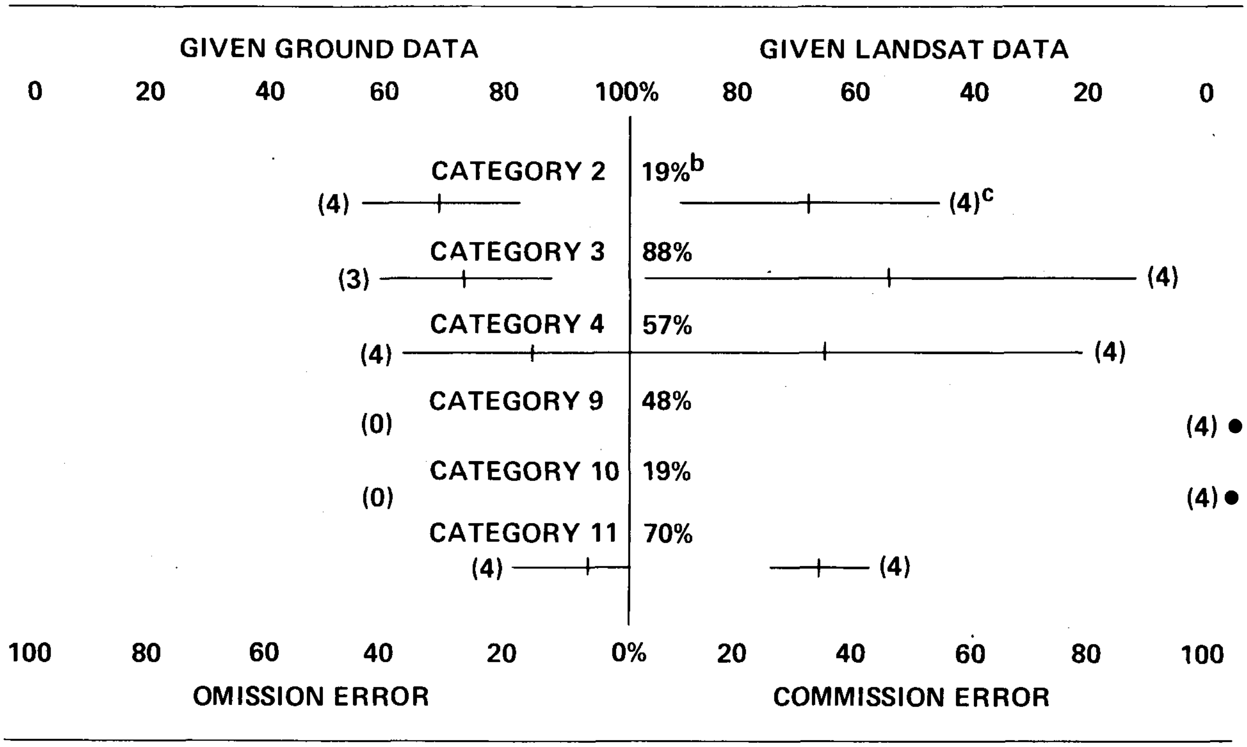


(a) CONFIDENCE INTERVALS ARE BASED ON TCHEBYSHEFF'S INEQUALITY WHICH STATES THAT AT LEAST 75% OF THE MEANS FROM REPEATED SAMPLES WILL BE WITHIN THE INTERVAL REGARDLESS OF THE SHAPE OF THE DISTRIBUTION.

(b) PERCENT OF THE MAP CATEGORY'S PIXELS WHICH OCCURRED IN THE MAP CATEGORY'S PLURALITY STRATUM.

(c) CLUSTER SAMPLE SIZE.

Figure 6.– Means and confidence[a] intervals for percent classification agreement for a category in its plurality stratum for the Jago Uplands study area.

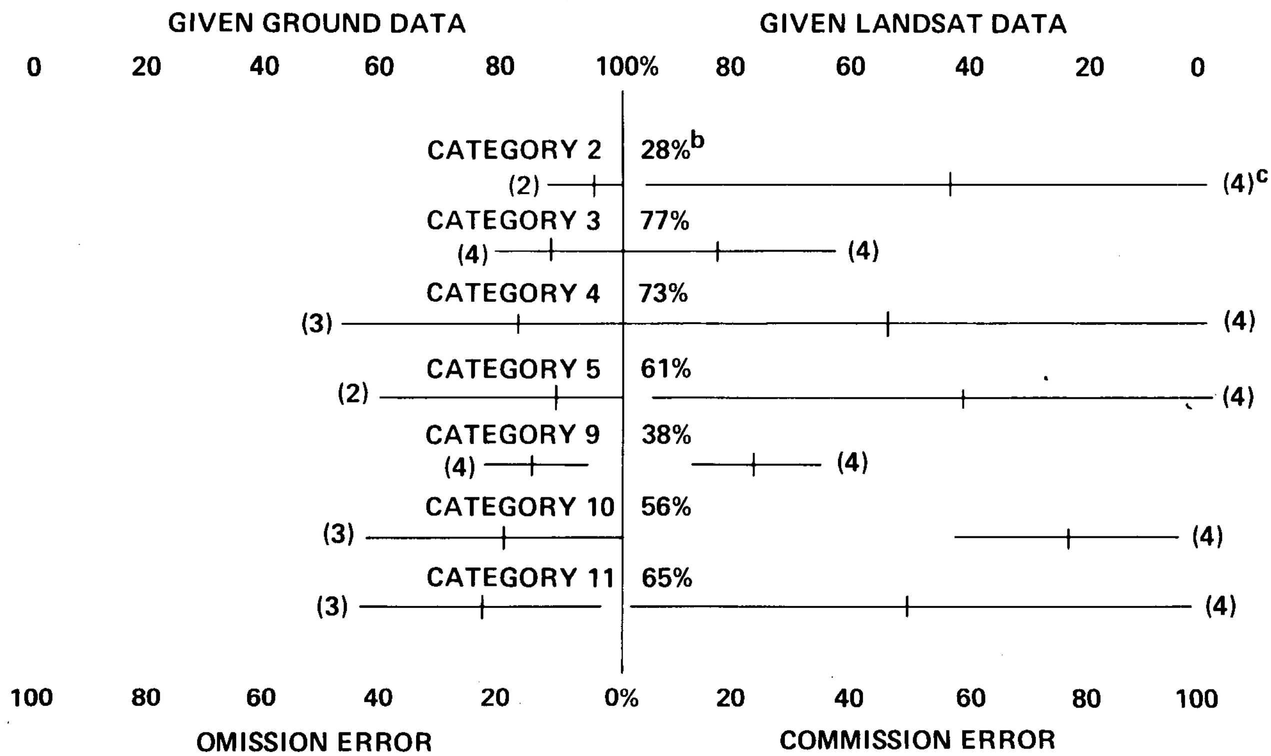


(a) CONFIDENCE INTERVALS ARE BASED ON TCHEBYSHEFF'S INEQUALITY WHICH STATES THAT AT LEAST 75% OF THE MEANS FROM REPEATED SAMPLES WILL BE WITHIN THE INTERVAL REGARDLESS OF THE SHAPE OF THE DISTRIBUTION.

(b) PERCENT OF THE MAP CATEGORY'S PIXELS WHICH OCCURRED IN THE MAP CATEGORY'S PLURALITY STRATUM.

(c) CLUSTER SAMPLE SIZE.

Figure 7.– Means and confidence[a] intervals for percent classification agreement for a category in its plurality stratum for the Canning River study area.

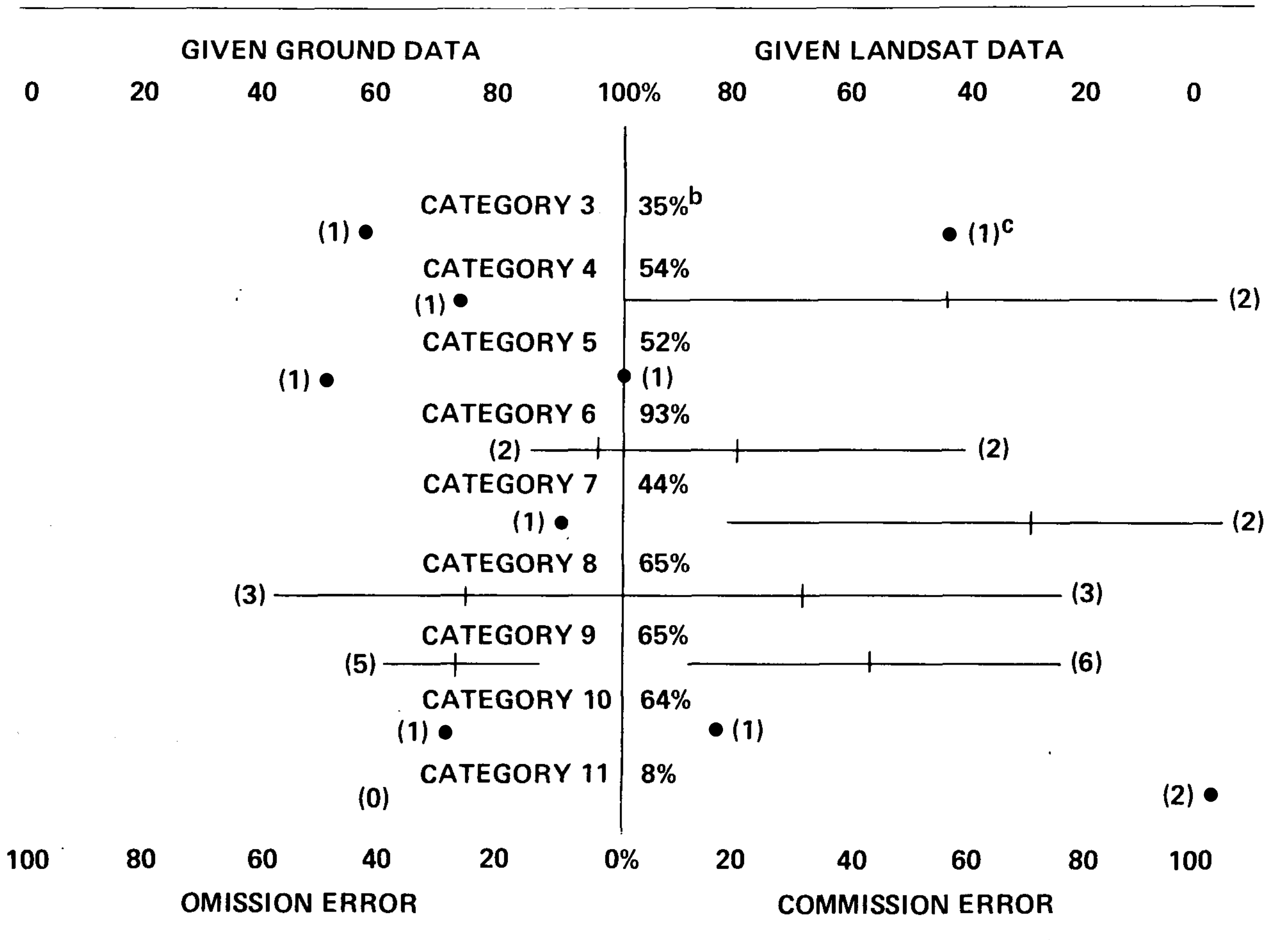


(a) CONFIDENCE INTERVALS ARE BASED ON TCHEBYSHEFF'S INEQUALITY WHICH STATES THAT AT LEAST 75% OF THE MEANS FROM REPEATED SAMPLES WILL BE WITHIN THE INTERVAL REGARDLESS OF THE SHAPE OF THE DISTRIBUTION.

(b) PERCENT OF THE MAP CATEGORY'S PIXELS WHICH OCCURRED IN THE MAP CATEGORY'S PLURALITY STRATUM.

(c) CLUSTER SAMPLE SIZE.

Figure 8.– Means and confidence[a] intervals for percent classification agreement for a category in its plurality stratum for the Okpilak coastal study area.



(a) CONFIDENCE INTERVALS ARE BASED ON TCHEBYSHEFF'S INEQUALITY WHICH STATES THAT AT LEAST 75% OF THE MEANS FROM REPEATED SAMPLES WILL BE WITHIN THE INTERVAL REGARDLESS OF THE SHAPE OF THE DISTRIBUTION.

(b) PERCENT OF THE MAP CATEGORY'S PIXELS WHICH OCCURRED IN THE MAP CATEGORY'S PLURALITY STRATUM.

(c) CLUSTER SAMPLE SIZE.

Figure 9.– Means and confidence[a] intervals for percent classification agreement for a category in its plurality stratum for the Sadlerochit Spring study area.

TABLE 9.– DESIGN EFFECTS FOR COMPARISON OF THE EFFICIENCY OF STRATIFIED CLUSTER, RANDOM CLUSTER, AND SINGLE PIXEL SAMPLING METHODS FOR OVERALL CLASSIFICATION ACCURACY.

Study area	Stratified cluster variance/ random cluster variance	Random cluster variance/ random pixel variance
Katakturuk	2.0	8.7
Niguanak	2.6	5.7
Jago	1.5	16.2
Canning	2.6	5.6
Okpilak	2.1	8.0
Sadlerochit	3.2	11.8

cover category at the sample unit. The choice between the two sampling methods requires understanding both cost and precision.

Comparison of stratified and random cluster sampling for omission and commission errors of individual categories could be made for few categories because of difficulties in their calculation, described above. However, when sample size was sufficient to allow calculation of an estimate for a category in all strata containing pixels of the category, plurality stratification provided greater precision. For example, plurality stratification reduced the variance for category 5 commission errors at the Katakturuk river study area by a factor of five compared to the variance expected for random cluster sampling. In other words, the proportion of pixels for which the ground determination of land cover category agreed with the map category varied considerably between stratum. This variation in classification accuracy by stratum is further evidence of the possible shortcomings of plurality stratification when estimates for a category are not available for all strata in which the category's pixels occurred.

CONCLUSIONS

Spectral confusion among land cover categories, spatially complex mixtures of land cover types, incomplete land cover category descriptions, and changes in land cover between the time of the classification and the time of the accuracy assessment were sources of classification error. In areas such as the Arctic Coastal Plain, which consists of spatially complex mixtures of land cover types, the fixed pixel size and shape in the Landsat data format and the binary choice of "correct" or "incorrect" for each pixel are sources of much of the error. The appropriate definition of a category is relative to the scale at which the definition makes sense, and this scale may not coincide with the pixel scale. The dependence of category definition on scale results in classification errors such as inclusions, mixed pixel confusions, and others, and can only be dealt with by using classifiers more sophisticated than per-pixel classifiers.

The reader of this report will not be able to ask the question, "how accurate is the map?" and expect a simple one-number answer, such as X%. First, a single number such as overall percent correct does not convey all of the information needed by planners. Per-class accuracies (equivalently, per-class omission and commission errors) are usually of more interest. In a complex classification process, one usually needs information about what categories are being confused, which can only be obtained from contingency tables. Second, as described in the Sample Design section, sampling was conducted at six areas rather than the entire coastal plain because of travel cost constraints. As a result of this decision, accuracies can be estimated only for individual study areas rather than the map as a whole. Furthermore, given the time resources for the accuracy assessment, this sampling plan resulted in small sample sizes, 20 to 24 clusters, at each study area. The sample provided accuracy estimates of low precision for a land cover category in its plurality stratum, but the sample sizes were not sufficient to obtain estimates of accuracy for all categories in all strata and hence prevented the calculation of stratified estimates. Calculation of estimates for a category using observations from only its plurality stratum introduces biases of unknown magnitude in accuracy estimates.

Stratified plurality sampling (SPS) is a complex sampling design that can provide unbiased and precise estimates of overall accuracy and also individual category commission and omission errors, given sufficient resources. Often, as in this study, resources are not sufficient to obtain large enough sample sizes to ensure that each category is represented in the sample of clusters from each plurality stratum. For these categories, unbiased stratified estimates of commission error and omission error are not possible because of one or more strata having no samples. Sample size required to obtain estimates of specified precision can be calculated, and cluster sample selection in each stratum continued until observations for all categories in all strata are obtained. For commission errors, selection of cluster sample units can be accomplished in the lab prior to field work. Estimation of omission errors is more difficult because the field data will have to be inspected to ensure samples of all categories in all strata are obtained. Results of this study suggest SPS can be an efficient sampling method for estimating commission errors, although more study of its efficiency relative to other sampling designs is needed.

REFERENCES

Card, D. H.: Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy. Photogrammetric Engineering and Remote Sensing, vol. 48, 1982, pp. 431-439.

Cochran, W. G.: Sampling Techniques. Third ed., John Wiley and Sons, Inc., 1977.

Congalton, R. G.: A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data. Photogrammetric Engineering and Remote Sensing, vol. 54, 1988, pp. 593-600.

ESL: IDIMS Functional Guide, vols. 1 and 2. Electromagnetic Systems Lab, Sunnyvale, CA, 1981.

Fitzpatrick-Lins, K.: The Accuracy of Selected Land Use and Land Cover Maps at Scales of 1:250,000 and 1:100,000. Geological Survey Circular 829, 1980.

Forbes, A. R.; Fox, L. III; and Mayer, K. E.: Forest Resource Classification of the McCloud District, McCloud, California, Using Landsat Digital Data. USDA U.S. Forest Service Final Report EM-7145-2, 1980.

Hansen, M. H.; Hurwitz, W. N.; and Madow, W. G.: Sample Survey Methods and Theory. Vol. 1, Methods and Applications. John Wiley and Sons, Inc., 1953.

Linden, D. S.; and Szajgin, J.: Verification of Land Cover Maps from Landsat Data, Proceedings of the Western Regional Remote Sensing Conference—1981. NASA CP 2195, 1981.

Mead, R. A.; and Szajgin, J.: Landsat Classification Accuracy Assessment Procedures, Photogrammetric Engineering and Remote Sensing, vol. 48, 1982, pp. 139-141.

Pettinger, L. R.: Digital Classification of Landsat Data for Vegetation and Land-Cover Mapping in the Blackfoot River Watershed, Southeastern Idaho. Geological Survey Professional Paper 1219, 1982.

Rowe, J. S.: The Level-of-Integration Concept and Ecology. Ecology, vol. 42, 1961, pp. 420-427.

Szajgin, J.; Pettinger, L. R.; Linden, D. S.; and Ohlen, D. O.: Arizona Vegetation Resource Inventory (AVRI) Accuracy Assessment, Geological Survey Open File Report No. 82-814, 1982.

Todd, W. J.; Gehring, D. G.; and Haman, J. F.: Landsat Wildland Mapping Accuracy. Photogrammetric Engineering and Remote Sensing, vol. 46, 1980, pp. 509-520.

Walker, D. A.; Acevedo, W.; Everett, K. R.; Gaydos, L.; Brown, J.; and Webber, P. J.: Landsat-Assisted Environmental Mapping in the Arctic National Wildlife Refuge, Alaska, Cold Regions Research Engineering Laboratory Report 82-27, 1982.

APPENDIX A

STATISTICAL FORMULAS

In stratified cluster sampling for proportions, the formulas for estimating the mean and variance for a proportion correct from a simple random sample of clusters within each stratum are derived in Cochran (1977; pp.64-68).

Notation:

p_{ih} = Estimated proportion correct for category i, stratum h

a_{ijh} = Number of pixels in cluster j, stratum h, identified as category i by both ground data and Landsat (number of correct calls)

m_{ijh} = Number of pixels in cluster j, stratum h, identified as category i (for probability correct, given the true category, use ground data calls; for probability correct, given the Landsat category, use Landsat calls)

n_{ih} = Number of sample clusters in stratum h with at least 1 pixel called category i (for probability correct, given the true category, use ground data calls; for probability correct, given the Landsat category, use Landsat calls)

N_{ih} = Total number of clusters in stratum h with at least 1 pixel called map category i

$\bar{m}_{ih}$ = Average number of category i pixels per cluster in stratum h (for probability correct, given the true category, use ground data calls; for probability correct, given the Landsat category, use Landsat calls)

f_{ih} = n_{ih}/N_{ih} = Sampling fraction for category i, stratum h

N_i = ΣN_{ih} = Total number of clusters in all strata with at least 1 pixel called map category i

The comments in parentheses above refer to the two methods of computing proportion correct and its variance, depending upon whether omission error or commission error is being considered.

Omission error = 1-Prob (correct, given the true class)
Commission error = 1-Prob (correct, given the Landsat class)

An omission error occurs when when a sample unit (pixel) identified as a certain land-cover type from data collected on the ground is classified by Landsat as a different land-cover type. A commission error occurs when a sample unit is classified by Landsat as a certain land-cover type when it is not that land-cover type as determined from ground data. For example, suppose Landsat says Wet Sedge for a pixel, whereas the ground classification says that the pixel is actually Shrub. In this case the error is a commission error for Wet Sedge and an omission error for Shrub. In each case, the number of clusters, n, and the number of pixels, m, in the cluster may vary because not all classes occur in every cluster.

Category proportion correct (within stratum h):

$$p_{ih} = \sum_{j}^{n_{ih}} a_{ijh} \Big/ \sum_{j}^{n_{ih}} m_{ijh}$$

Approximate sampling variance (within stratum h):

$$\mathrm{var}(p_{ih}) = (1 - f_{ih})\left(\sum_{j}^{n_{ih}} a_{ijh}^2 - 2p_{ih} \sum_{j}^{n_{ih}} a_{ijh}m_{ijh} + p_{ih}^2 \sum_{j}^{n_{ih}} m_{ijh}^2\right) \Big/ \left[\bar{m}_{ih}^2 n_{ih}(n_{ih} - 1)\right]$$

Stratified estimator for proportion correct:

$$p_{sti} = \sum_{h}^{H_i} (N_{ih}/N_i)p_{ih}$$

Approximate sampling variance of p:

$$\mathrm{var}(p_{sti}) = \sum_{h}^{H_i} (N_{ih}/N_i)^2 \, \mathrm{var}(p_{ih})$$

where H_i is the number of strata containing at least k sample clusters with at least one pixel called category i (k is 1 for valid estimates p, and is 2 for valid estimates of variances).

The above formulas hold for overall proportion correct, but simplify somewhat, because we need not distinguish between the cases for omission and commission error (every pixel is either correct or incorrect, and is always counted). If we sum the a's over all categories and note that the cluster size m is now constant (25 pixels) and n is the total number of clusters sampled from stratum h, we have the following:

Overall proportion correct:

$$p_h = \sum_{j}^{n_h} a_{jh}/\bar{m}n_h$$

Approximate variance of overall proportion correct:

$$\mathrm{var}(p_h) = (1 - f_h)p_h(1 - p_h)/(n_h - 1)$$

Stratified estimate of overall proportion correct:

$$p_{st} = \sum_{h}^{H} (N_h/N)p_h$$

Approximate variance of p:

$$\text{var}(p_{st}) = \sum_{h}^{H} (N_h/N)^2(1 - f_h)p_h(1 - p_h)/(n_h - 1)$$

APPENDIX B

SITE 1—KATAKTURUK RIVER
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
GROUND AND AERIAL PHOTO DATA/OMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by photo interpretation ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
3	3	2	45	24	0.53	0.0105
4	3	1	5	5	1.00	.0000
	4	1	25	25	1.00	.0000
5	10	2	13	9	.69	.0353
	4	3	73	22	.30	.1146
	5	2	49	44	.90	.1040
6	6	3	30	15	.50	.0499
	10	2	17	7	.41	.1924
	4	1	2	2	1.00	.0000
	6	3	64	59	.92	.0209
7	7	4	40	22	.55	.0847
	6	4	31	8	.26	.0245
	7	5	85	73	.86	.0539
8	8	1	3	3	1.00	.0000
	8	1	18	10	.56	.0000
9	10	2	10	0	.00	.0000
10	8	1	4	1	.25	.0000
	10	4	60	58	.97	.0180
12	5	1	1	0	0.00	0.0000

PRECEDING PAGE BLANK NOT FILMED

SITE 1—KATAKTURUK RIVER
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
3	3	2	24	24	1.00	0.0000
	10	1	2	0	.00	.0000
4	3	2	22	5	.23	.1753
	4	4	73	25	.34	.3028
	10	3	20	9	.45	.1610
5	3	1	4	0	.00	.0000
	4	3	22	22	1.00	.0000
	5	2	44	44	1.00	.0000
	6	5	21	15	.71	.2167
	7	1	1	0	.00	.0000
	8	1	1	0	.00	.0000
	10	2	8	7	.88	.2172
6	4	1	3	2	.67	.0000
	5	1	6	0	.00	.0000
	6	5	94	59	.63	.2038
	7	5	31	22	.71	.2066
	8	1	16	0	.00	.0000
7	6	2	10	8	.80	.2797
	7	5	91	73	.80	.1153
	8	1	6	3	.50	.0000
8	4	1	2	0	.00	.0000
	7	1	2	0	.00	.0000
	8	1	10	10	1.00	.0000
10	8	1	2	1	.50	.0000
	10	4	70	58	0.83	0.0904

SITE 2—NIGUANAK RIVER
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
GROUND AND AERIAL PHOTO DATA/OMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by photo interpretation ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	2	2	11	11	1.00	0.0000
	3	2	18	10	.56	.0987
	9	2	5	0	.00	.0000
	10	2	2	0	.00	.0000
2	2	3	51	32	.63	.0430
	3	1	2	1	.50	.0000
	5	1	3	0	.00	.0000
	9	2	5	0	.00	.0000
3	2	2	12	6	.50	.2582
	3	4	66	55	.83	.0775
	4	1	12	7	.58	.0000
	5	3	17	6	.35	.2535
	9	2	16	3	.19	.0530
	10	1	1	1	1.00	.0000
4	3	3	11	3	.27	.1980
	4	3	46	34	.74	.1863
	5	4	54	21	.39	.0559
	9	1	6	4	.67	.0000
	10	3	15	10	.67	.1350
5	2	1	1	0	.00	.0000
	4	3	32	18	.56	.0986
	5	2	26	26	1.00	.0000
	9	3	34	15	.44	.0931
	10	3	40	19	.48	.1730
7	3	1	3	0	.00	.0000
	4	1	10	0	.00	.0000
8	9	2	2	0	.00	.0000
9	9	1	1	1	1.00	.0000
	10	1	1	0	.00	.0000
10	9	3	22	6	.27	.1380
	10	2	9	7	.78	.0485
11	9	1	1	0	.00	.0000
12	10	2	7	2	0.29	0.0802

SITE 2—NIGUANAK RIVER
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	2	2	13	11	0.85	0.0092
	3	2	10	10	1.00	.0000
2	2	3	32	32	1.00	.0000
	3	1	3	1	.33	.0000
3	2	3	20	6	.30	.1633
	3	4	70	55	.79	.1246
	4	1	8	7	.88	.0000
	5	2	7	6	.86	.0816
	9	2	4	3	.75	.1212
	10	1	6	1	.17	.0000
4	2	2	4	0	.00	.0000
	3	4	13	3	.23	.1002
	4	4	58	34	.59	.2237
	5	4	23	21	.91	.0828
	9	3	8	4	.50	.3575
	10	3	13	10	.77	.1868
5	2	2	6	0	.00	.0000
	3	2	4	0	.00	.0000
	4	4	34	18	.53	.3076
	5	4	70	26	.37	.2406
	9	3	23	15	.65	.2436
	10	3	19	19	1.00	.0000
9	9	4	48	1	.02	.0196
	10	3	7	0	.00	.0000
10	9	4	17	6	.35	.1276
	10	3	27	7	.26	.2158
12	10	1	3	2	0.67	0.0000

SITE 3—JAGO UPLANDS
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
GROUND AND AERIAL PHOTO DATA/OMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by photo interpretation ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	4	2	5	0	0.00	0.0000
	5	1	4	0	.00	.0000
	9	4	23	0	.00	.0000
3	4	2	23	2	.09	.0941
	5	1	7	0	.00	.0000
4	4	3	28	20	.71	.0516
	5	2	22	3	.14	.1230
	9	4	21	6	.29	.0463
5	4	2	10	0	.00	.0000
	8	1	19	0	.00	.0000
6	4	3	17	10	.59	.0257
	5	2	37	7	.19	.0928
	6	4	75	69	.92	.0475
	7	4	77	23	.30	.1334
	8	3	21	10	.48	.3456
7	5	1	13	0	.00	.0000
	6	2	21	17	.81	.1043
	7	3	21	20	.95	.0272
	8	4	48	26	.54	.0857
	9	1	1	1	1.00	.0000
8	4	2	9	0	.00	.0000
	7	1	2	0	.00	.0000
	8	1	12	10	.83	.0000
	9	3	13	0	.00	.0000
9	4	2	4	2	.50	.4787
	5	2	17	7	.41	.0824
	6	1	4	0	.00	.0000
	9	4	30	16	.53	.0895
10	4	2	4	1	.25	.3590
	9	3	12	6	0.50	0.1208

SITE 3—JAGO UPLANDS
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
3	4	2	6	2	0.33	0.4255
4	4	4	43	20	.47	.1670
	5	1	7	3	.43	.0000
	9	4	12	6	.50	.0913
5	4	3	10	0	.00	.0000
	5	4	68	0	.00	.0000
	6	2	7	0	.00	.0000
	9	4	23	0	.00	.0000
6	4	4	21	10	.48	.1356
	5	3	14	7	.50	.1618
	6	4	74	69	.93	.0579
	7	4	25	23	.92	.0677
	8	2	14	10	.71	.2317
	9	2	6	0	.00	.0000
7	4	1	2	0	.00	.0000
	6	2	19	17	.89	.0775
	7	4	75	20	.27	.1574
	8	4	34	26	.76	.1805
	9	1	4	1	.25	.0000
8	8	4	52	10	.19	.1846
9	4	2	10	2	.20	.1149
	5	2	11	7	.64	.2625
	9	4	41	16	.39	.0529
10	4	1	8	1	.13	.0000
	9	3	14	6	0.43	0.0970

SITE 4—CANNING RIVER
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
GROUND AND AERIAL PHOTO DATA/OMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by photo interpretation ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	2	3	11	3	0.27	0.1549
	10	2	4	1	.25	.2406
	11	2	15	7	.47	.1151
2	2	4	70	43	.61	.0469
	3	2	8	4	.50	.2499
	4	1	5	2	.40	.0000
	10	1	1	0	.00	.0000
3	2	4	14	12	.86	.0752
	3	3	52	42	.81	.0634
	4	2	37	22	.59	.0583
	10	3	34	8	.24	.1129
4	2	3	5	2	.40	.3563
	3	3	34	16	.47	.1791
	4	4	50	49	.98	.0136
	10	2	10	5	.50	.0962
	11	1	8	1	.13	.0000
5	4	1	8	0	.00	.0000
	10	1	10	0	.00	.0000
9	10	4	38	1	.03	.0249
	11	2	2	0	.00	.0000
10	3	1	6	1	.17	.0000
	9	2	7	0	.00	.0000
	10	1	2	2	1.00	.0000
	11	1	2	0	.00	.0000
11	9	4	93	31	.33	.0519
	10	1	1	0	.00	.0000
	11	4	73	70	0.96	0.0306

SITE 4—CANNING RIVER
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	2	4	16	3	0.19	0.1638
	10	1	1	1	1.00	.0000
	11	2	9	7	.78	.2952
2	2	4	52	43	.83	.0756
	3	2	5	4	.80	.0800
	4	1	3	2	.67	.0000
	10	1	1	0	.00	.0000
	11	1	1	0	.00	.0000
3	2	4	28	12	.43	.0893
	3	4	68	42	.62	.2292
	4	4	33	22	.67	.2957
	10	3	19	8	.42	.1028
	11	1	6	0	.00	.0000
4	2	1	2	2	1.00	.0000
	3	4	23	16	.70	.2274
	4	4	64	49	.77	.1705
	10	3	24	5	.21	.1553
	11	1	1	1	1.00	.0000
9	9	4	68	0	.00	.0000
	10	1	1	1	1.00	.0000
	11	1	2	0	.00	.0000
10	3	2	3	1	.33	.4442
	10	4	51	2	.04	.0353
	11	1	3	0	.00	.0000
11	2	1	2	0	.00	.0000
	3	1	1	0	.00	.0000
	9	4	32	31	.97	.0325
	10	2	3	0	.00	.0000
	11	4	78	70	0.90	0.0635

SITE 5—OKPILAK COASTAL
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
GROUND AND AERIAL PHOTO DATA/OMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by photo interpretation ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	1	6	106	97	0.92	0.0418
	2	3	22	14	.64	.1163
	3	1	2	0	.00	.0000
	9	2	11	3	.27	.1477
	10	4	13	0	.00	.0000
	11	4	20	7	.35	.1454
2	1	4	14	9	.64	.0477
	2	2	20	15	.75	.0492
	4	1	1	0	.00	.0000
3	1	3	20	15	.75	.0990
	2	2	14	10	.71	.1808
	3	4	76	66	.87	.0428
	4	3	32	17	.53	.0896
	5	1	8	0	.00	.0000
4	1	2	9	9	1.00	.0000
	2	1	3	1	.33	.0000
	3	2	11	7	.64	.0330
	4	4	25	24	.96	.0466
	5	3	57	17	.30	.0764
	9	1	1	0	.00	.0000
	10	1	3	0	.00	.0000
	11	3	10	2	.20	.2092
5	2	2	29	0	.00	.0000
	3	1	3	0	.00	.0000
	4	2	35	7	.20	.2627
	5	2	35	34	.97	.0408
	11	1	3	3	1.00	.0000
9	1	1	1	1	1.00	.0000
	2	1	12	2	.17	.0000
	3	1	5	1	.20	.0000
	4	1	6	4	.67	.0000
	9	4	58	50	.86	.0335
	10	3	19	14	.74	.1481
	11	2	12	6	.50	.0829

SITE 5—OKPILAK COASTAL (CONCLUDED)
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
GROUND AND AERIAL PHOTO DATA/OMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by photo interpretation ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
10	9	1	1	1	1.00	.0000
	10	3	14	12	.86	.1375
	11	2	9	4	.44	.3686
11	3	1	3	2	.67	.0000
	4	1	1	0	.00	.0000
	9	4	29	11	.38	.1775
	10	4	51	12	.24	.0639
	11	4	46	37	0.80	0.1057

SITE 5—OKPILAK COASTAL
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	1	6	100	97	0.97	0.0218
	2	4	22	14	.64	.2343
	9	1	3	3	1.00	.0000
	11	3	16	7	.44	.0242
2	1	4	15	9	.60	.2464
	2	4	42	15	.36	.2036
	9	1	2	0	.00	.0000
3	1	4	21	15	.71	.1238
	2	3	22	10	.45	.2619
	3	4	70	66	.94	.0302
	4	3	17	17	1.00	.0000
	5	1	1	0	.00	.0000
	9	2	5	0	.00	.0000
4	1	2	9	9	1.00	.0000
	2	1	2	1	.50	.0000
	3	4	16	7	.44	.1753
	4	4	67	24	.36	.1592
	5	3	21	17	.81	.1629
	11	1	3	2	.67	.0000
5	2	1	3	0	.00	.0000
	3	1	6	0	.00	.0000
	4	3	11	7	.64	.3609
	5	4	76	34	.45	.2735
6	5	1	2	0	.00	.0000
9	1	1	5	1	.20	.0000
	2	2	4	2	.50	.4922
	3	2	5	1	.20	.1599
	4	1	5	4	.80	.0000
	9	4	67	50	.75	.0393
	10	4	16	14	.88	.0815
	11	3	17	6	.35	.2838
10	10	4	64	12	.19	.1239
	11	2	9	4	.44	.1720

SITE 5—OKPILAK COASTAL (CONCLUDED)
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
11	2	1	5	0	.00	.0000
	3	1	3	2	.67	.0000
	9	2	15	11	.73	.2206
	10	4	20	12	.60	.1200
	11	4	48	37	0.77	0.1392

SITE 6—SADLEROCHIT SPRING
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
GROUND AND AERIAL PHOTO DATA/OMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by photo interpretation ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	4	1	5	0	0.00	0.0000
3	3	1	15	10	.67	.0000
	4	1	9	4	.44	.0000
4	3	1	10	8	.80	.0000
	4	1	16	15	.94	.0000
	5	1	5	5	1.00	.0000
5	5	1	13	9	.69	.0000
	6	1	7	0	.00	.0000
6	6	2	43	41	.95	.0541
	7	1	12	10	.83	.0000
	8	2	26	7	.27	.1037
	9	1	4	0	.00	.0000
7	7	2	17	15	.88	.1784
	8	2	3	3	1.00	.0000
8	4	1	20	0	.00	.0000
	7	1	19	5	.26	.0000
	8	3	37	30	.81	.1006
9	5	1	7	2	.29	.0000
	7	1	2	0	.00	.0000
	8	2	9	0	.00	.0000
	9	3	71	54	.76	.0640
	10	1	10	10	1.00	.0000
	11	2	23	16	.70	.0764
10	10	1	15	13	.87	.0000
	11	2	16	6	.38	.1894
11	11	1	11	8	0.73	0.0000

SITE 6—SADLEROCHIT SPRING
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
1	11	1	2	0	0.00	0.0000
2	5	1	1	0	.00	.0000
	11	1	4	0	.00	.0000
3	3	1	11	10	.91	.0000
	4	2	6	4	.67	.2214
	5	1	3	0	.00	.0000
	9	1	1	0	.00	.0000
4	3	1	9	8	.89	.0000
	4	2	29	15	.52	.3555
	5	1	8	5	.63	.0000
	9	2	3	0	.00	.0000
5	4	2	4	0	.00	.0000
	5	1	9	9	1.00	.0000
	6	1	2	0	.00	.0000
	8	3	8	0	.00	.0000
	9	2	10	0	.00	.0000
6	4	1	6	0	.00	.0000
	5	1	2	0	.00	.0000
	6	2	48	41	.85	.1518
	7	2	19	10	.53	.4944
	8	3	13	7	.54	.1456
7	4	1	4	0	.00	.0000
	7	2	26	15	.58	.3315
	8	3	11	3	.28	.1221
8	7	1	5	5	1.00	.0000
	8	3	43	30	.70	.2133
9	3	1	5	0	.00	.0000
	4	1	1	0	.00	.0000
	5	1	2	2	1.00	0.0000
	9	3	58	54	.93	.0694
	10	1	12	10	.83	.0000
	11	2	16	16	1.00	.0000

SITE 6—SADLEROCHIT SPRING (CONCLUDED)
PROPORTION OF PIXELS CORRECT FOR EACH CLASS BY STRATA
LANDSAT LAND COVER/COMMISSION

Class (i)	Strata (h)	Number of cluster blocks (n_{hi})	Number of pixels by Landsat ($\sum_j M_{hi}$)	Number of agreement pixels ($\sum_j a_{hi}$)	Proportion correct (p_{hi})	Standard error (p_{hi})
10	10	1	13	13	1.00	.0000
	11	1	6	6	1.00	.0000
11	9	1	3	0	.00	.0000
	11	2	22	8	0.37	0.2672

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-101042		

4. Title and Subtitle	5. Report Date
Accuracy Assessment, Using Stratified Plurality Sampling, of Portions of a Landsat Classification of the Arctic National Wildlife Refuge Coastal Plain	January 1989
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Don H. Card and Laurence L. Strong (TGS Technology, Inc., Moffett Field, CA)	A-88309
	10. Work Unit No.
9. Performing Organization Name and Address	11. Contract or Grant No.
Ames Research Center Moffett Field, CA 94035	1-7060-0450
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Washington, DC 20546-0001	Technical Memorandum
	14. Sponsoring Agency Code

15. Supplementary Notes

Point of Contact: James G. Lawless, Ames Research Center, MS 242-4, Moffett Field, CA 94035
(415) 694-5900 or FTS 464-5900

16. Abstract

This report describes an application of a classification accuracy assessment procedure for a vegetation and land cover map prepared by digital image processing of Landsat multispectral scanner data. A statistical sampling procedure called Stratified Plurality Sampling was used to assess the accuracy of portions of a map of the Arctic National Wildlife Refuge coastal plain (Walker et al. 1982). Results are tabulated as percent correct classification overall as well as per category with associated confidence intervals. Although values of percent correct were disappointingly low for most categories, the study was useful in highlighting sources of classification error and demonstrating shortcomings of the plurality sampling method.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Accuracy assessment Vegetation mapping Stratified sampling Computer classification	Unclassified – Unlimited Subject Category – 43

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	51	A04

 For sale by the National Technical Information Service, Springfield, Virginia 22161