A STUDY OF FOUR RASTER-BASED DATA
GENERALIZATION PROCEDURES

by

Margaret Grace Mayers

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree
of

MASTER OF SCIENCE

in

Geography

APPROVED:

Dr. James B. Campbell Chairman

Dr. Laurence W. Carstensen

Dr. James L. Smith

May, 1987

Blacksburg, Virginia

RASTER-BASED GENERALIZATION STRATEGIES FOR
LAND USE/LAND COVER NOMINAL DATA

by

Margaret Grace Mayers

Committee Chairman: James B. Campbell
Geography

(ABSTRACT)

Strategies for generalizing raster-based land cover data were investigated. Generalization strategies were studied as part of the U.S. Geological Survey's (USGS) Federal Land Information System (FLIS). Area filtering, category aggregation, resampling, and modal search and replace comprised the strategies tested on USGS Alaska Interim Land Cover data.

Generalization of the land cover data was deemed necessary for two reasons: 1) reduction in the volume of homogeneous land cover regions required for computer memory storage, and 2) simplification of the highly complex land cover map. The generalization strategies were evaluated based upon how well they maintained the integrity of land cover information while minimizing the number of homogeneous land cover regions. Maintenance of land cover information after application of a generalization strategy was measured by omission and comission errors, percent unchanged, Cohen's statistic, and the number of land cover regions.

Area filtering, category aggregation, modal search and replace, and resampling all reduced the number of gridded land cover regions significantly, but it was found that aggregation of categories to lower levels of precision gave the highest amount of region reduction with no category error. Area filtering and area filtering in combination with category aggregation resulted in the lowest amount of category specific errors relative to the amount of overall region reduction. Consideration of pixel contiguity for each land cover category by the area filtering algorithm produced the most desirable results.

ACKNOWLEDGEMENTS

This thesis was largely supported by TGS Technology, Inc., contractor to the U.S. Geological Survey, in support of the U.S. Geological Survey's Federal Land Information System, contract 14-08-0001-22521. Support in the form of encouragement, suggestions, and patience was given from many scientists at the EROS Data Center, especially . I also thank for editing numerous abstract drafts, and for helping me pull through the South Dakota blues. deserves a large thanks for his suggestions and ideas using area filtering for generalization purposes. help with understanding the resampling results was greatly appreciated. Thanks also go out to for his help with my slides and for his literature suggestions. I am grateful to for editing my rough drafts and for his strong moral support. Finally, I would like to thank for his help with LAS.

Gratitude is also extended to my graduate committee, Dr. James B. Campbell, Dr. Laurence W. Carstensen, and Dr. James L. Smith for their guidance, suggestions, and encouragement.

The author would like to thank her parents, , for their unwavering love and support. This

thesis is dedicated to them with eternal gratitude. Education is only one of many gifts they have given their daughter.

TABLE OF CONTENTS

LIST OF FIGURES

(LIST OF FIGURES CONTINUED)

(LIST OF FIGURES CONTINUED)

(LIST OF FIGURES CONTINUED)

(LIST OF FIGURES CONTINUED)

LIST OF TABLES

(LIST OF TABLES CONTINUED)

(LIST OF TABLES CONTINUED)

(LIST OF TABLES CONTINUED)

(LIST OF TABLES CONTINUED)

(LIST OF TABLES CONTINUED)

(LIST OF TABLES CONTINUED)

INTRODUCTION

The Alaska Federal Land Information System

The U.S. Geological Survey (USGS) began investigating the feasibility of developing a nationwide digital data base for assessing mineral potential on Federal Lands in 1983 under a program called the Federal Mineral Land Information System (FMLIS). In 1986 the program was renamed to the Federal Land Information System (FLIS) and the scope was broadened to include other sources and types of resource information in addition to the original mineral data layers (USGS FLIS Program Workplan, July 1986).

Alaska was selected as the starting state for the FMLIS Program because of the large amount of existing Federal Land in Alaska, and the need for resource data by policy makers to effectively manage this land. The Alaska FMLIS data base was expanded in 1986 to produce and demonstrate a FLIS prototype (USGS FLIS Program Workplan, July 1986).

Recently, FLIS personnel incorporated USGS Valdez Interim Land Cover data to the Alaska FLIS database and will continue to incorporate other quadrangles as they become available. Incorporation of the Valdez Land Cover data initiated many problems for storage and analysis of this data set. The USGS Valdez, Alaska 1:250,000 scale land cover map contained approximately 98,000 land use and land cover gridded homogeneous areas, or regions. Generalization was deemed necessary for reduction in the volume of homogeneous land cover regions required for computer memory storage, and for simplification of the highly complex land cover map.

Statement of Problem and Objectives

The purpose of this research project was to study the impact of different generalization strategies on the USGS Valdez Interim Land Cover data. The study focused on overall data reduction for the entire area as well as changes occurring within each land cover category after generalization. Four specific objectives were chosen: 1) to compare the four generalization strategies against one another for overall region reduction and overall similarity to the original land cover data, 2) to compare the impact of these different strategies had on individual land cover categories, 3) to select one of the four strategies for use in generalizing forthcoming Alaska land cover data sets, and 4) to identify those aspects of the selected strategy that need further study.

The research was exploratory in nature and the results provide information for further studies which will tailor the generalized data to a specific user as well as provide a guideline for preprocessing future raster data into the Alaska FLIS database. Those who have different land use/land cover data may benefit by utilizing the strategies conducted in this study for generalizing raster data.

USGS Alaska Interim Land Cover Mapping Program

Landsat satellite data covering over two-thirds of Alaska have been classified for land cover. Conventional land use/land cover data for Alaska has been produced for only the Valdez 1:250,000 scale quadrangle. Shasby and Carneggie (1986) state that this is due to the limited availability of source material, the diversity of classification schemes required by user agencies, and the difficulty of applying the Anderson (Anderson et al., 1976) classification system to Alaska. To resolve this situation, the USGS designed the Interim Land Cover Mapping Program for Alaska. The Interim Program will be carried out by reclassifying existing raster classifications, and by implementation of new classification efforts for unmapped areas.

Processing of USGS Interim Land Cover Data

The USGS Interim Land Cover data for the Alaska Copper River area were developed from Landsat multispectral scanner data in five major steps. The first step included data base development. This step reformatted the Landsat multispectral data and registered it to the desired spatial configuration. The second step included training and clustering. In this step the selection of spectral data samples and definition of distinct spectral groups was accomplished. Step three involved spectral class evaluation. The vegetation group associated with a particular spectral response evaluated for homogeneity during this step. Post-classifcation stratification comprised step four. Terrain variables and winter Landsat data were studied to reduce spectral class confusion. The last step involved final classification refinement. During this step, the categories were aggregated into a more general classification system, spatial smoothing was applied, and the final land cover maps for the Copper River area were produced (USGS Copper River Area Land Cover Mapping Project, July 1984).

USGS Valdez Interim Land Cover Data

For the Valdez quadrangle, second order transformation and nearest neighbor resampling was conducted for data base development. Cluster analysis was used to define spectral classes for each scene using training samples. Elevation, slope, solar illumination, and winter Landsat data were used for post-classification stratification. The final refinement was accomplished by mosaicking the Landsat scenes for Valdez, and smoothing in order to generalize the spatial detail and to remove random misclassifications (USGS Copper River Land Cover Mapping Project, July 1984). The resulting USGS Alaska Interim land cover categories are a combination of the Anderson level I and II map classification system, and the Viereck and Dyrness vegetation classification system (Shasby and Carneggie, 1986). The USGS Alaska Interim Land Cover Categories and their definitions are listed in Table 1.

Data Reduction Requirements

Data reduction is not always a necessary step for incorporating classified Landsat data into a vector-based GIS. If a computer memory system can adequately store in its memory a coverage with many line segments which define

TABLE 1: USGS Interim Land Cover Classification for Alaska

Level I	Level II
I. Forest (forest canopy cover of one-third or more)	A. Needleleaf Forest (over two-thirds of tree cover contributed by needleleaf species)
	B. Broadleaf Forest (over two-thirds of tree cover contributed by broadleaf species)
	C. Mixed Forest (broadleaf or needleleaf species contribute one-third to two-thirds of the tree cover)
II. Shrub (forest canopy cover of less than one-third and shrub canopy cover of one-third or more)	A. Shrublands
	B. Dwarf Shrublands and Related Communities (rarely exceeding 50 cm in height)
III. Herbaceous (vegetation with 5% or more of vascular and non-vascular (mosses and lichens) cover and less than one-third cover of woody plants)	A. Dry or moist
	B. Wet (seasonally or systematically submerged)
	C. Very wet-aquatic (sites are typically large ponds in the lower elevations)
	D. Mosses
	E. Lichens
IV. Agriculture	
V. Urban	
VI. Barren	A. Sparse vegetation
	B. Non-vegetated (rock, soil)

(TABLE 1 Continued)

VII. Water	A. Clear/Deep (water bodies typically deeper than 1 meter with low turbidity)
	B. Turbid/Shallow (water bodies typically less than 1 meter deep or that have significant amounts of suspended sediments)
VIII. Ice and Snow (includes snowfields and glaciers)	
IX. Shadow (areas in rugged terrain in which incident solar illumination is too low to permit the discrimination of vegetation types)	
X. Clouds	

(Shasby and Carneggie, 1986 and USGS Copper River Area Land Cover Mapping Project, July 1984) polygons, then generalization may not be necessary. Generalization may also not be necessary if the user wants data for a small area. In this instance, the land cover map may be quite detailed with many polygons but because of the small spatial extent, the computer systems memory and processing time is not significantly impacted. Because the pixel size of the classified land cover data used in this study is only 50 meters on one side, and FLIS personnel wish to obtain a quadrangle-by-quadrangle coverage of the entire state of Alaska, generalization is necessary in order to reduce the number of arcs or line segments which define land use boundaries primarily for two reasons. First, to reduce the amount of computer memory storage space required for the land cover data, and second to simplify complex land cover maps for broad applications in policy making and land management.

REVIEW OF THE LITERATURE

Land Use Studies in Remote Sensing

Lintz and Simonett (1976) have identified three problems concerned with land use studies in remote sensing; 1)procedures for identifying land use from various types of remote sensor imagery, 2)classification and categorization, and 3)mapping land use traits. Classification and categorization of remotely sensed images for applications by resource managers has often resulted in land use/land cover maps (Lauer 1986, Campbell 1983). These maps have been generated manually, as well as with the help of digital image processing and geographic information system software (Lauer, 1986).

Early Satellite Remote Sensing Applications

Satellite remote sensing resource applications have evolved through three general phases; 1)manual analysis, 2)digital analysis, and 3)digital analysis and information processing (Lauer, 1986). Traditional remote sensing applications have utilized raster, or grid cell, data because of the compatibility of sensors, such as Landsat, with image processing systems (Lauer, 1986).

Also, applications usually require an uninterrupted layer of data, such as land use, which were easily obtainable from satellite data (Lauer, 1986). With the evolution of vector-formatted pen plotters and geographic information systems, compatibility between different software systems for cartographic products derived from remote sensing imagery became a problem. Lauer's discussion concluded with the problems associated with the processing, analysis and generation of earth science products with a single data format (Lauer, 1986).

Vector Format for Storing Image Data

Peuquet (1984) has also discussed the advantages and disadvantages of raster and vector-formatted data with respect to topologic analysis capabilities and image processing. Vector data models were among the first efforts for producing automated cartographic products (Kolassa 1982, Peuquet 1984). These methods model manual map making processes by digitally recording a series of linear chains, based on the DIME concept (Peuquet 1984, Kolassa 1982). Many cartographic products derived from remotely sensed data, such as land use/land cover maps, are generated from vector-based geographic information systems (Selden and Domaratz, 1983).

Problems with Vector Format

Peuquet (1984) stated that the main problem with the vector data models is that individual line segments do not occur in any particular sequence order. To retrieve any particular line segment, a sequential exhaustive search must be performed on the entire file. To retrieve all line segments which define the boundary of a polygon, an exhaustive search must be made as many times as there are line segments in the polygon boundary (Peuquet, 1984).

Raster Format for Storing Image Data

During the early 1960s, several raster scanners were developed for sampling and reproducing cartographic source material (Kolassa, 1982). Raster data models digitally record cartpgraphic data in an array of uniform regions, or pixels. Raster scanners automatically recorded and stored pixeled information from the source data within short periods of time, often within minutes. These scanners significantly increased data capture rates compared to the manual digitizing process which involved many man hours of data entry and editing.

Problems with Raster Models

According to Kolassa (1982), a major obstacle in the use of raster data for cartographic applications was the difficulty in performing feature or segment manipulations in a fashion similar to most cartographers. Peuquet (1984) also stated that spatial relationships are built in for raster-type models but that they are not to be very compact.

Raster to Vector Data Conversion

Vector to raster data conversion has been common since the early days of automated cartography (Nichols, 1981). Conversion from raster-to-vector data is not utilized commonly. Marble and Nichols (1984, 1981), however, have stated several justifications for raster to vector data conversion, such as reduction of digitizing costs, frequent temporal data updates by satellites, and the availability of raster-based data for vector-based geographic information systems (Marble 1984, Nichols 1981). On the other hand, the raster-to-vector conversion process has typically encountered two problems; 1) a high processing overhead, and 2) difficulty of associating attributes with newly created vector chains (Kolassa 1982).

Rosenfeld's Raster to Vector Conversion Algorithm

The most commonly used procedure for direct raster to vector conversion, without any generalization, was developed by Rosenfeld (1978). Rosenfeld's procedure employed a process of skeletonization to reduce lines to a single unit of width (a raster cell side), upon which line extraction and topology reconstruction was performed. The final vector-based polygon map consisted of edges enclosing the classified regions and the interiors which represented the characteristics of the regions (Rosenfeld, 1978).

Need for Digital Map Generalization

Direct raster-to-vector data conversion was easily accomplished using Rosenfeld's procedure; however, data complexity and large data volumes resulting from this process soon became a problem. Each pixel in raster-formatted data was converted into a small polygon if the pixel was not contiguous with a similarly classified neighboring pixel. Not only were single pixel polygons and small clumps of homogeneous pixel areas difficult to examine, the topology (once converted) of these small areas took up enormous amounts of computer memory (Peuquet, 1984). Selden and Domaratz (1983) stated that digital map generalization and procedures for cartographic manipulation of digital data were the two most rapidly advancing aspects of the cartographic process.

Generalization has become an important issue because of the large accumulation of cumulating digital Earth science information in both raster and vector formats (Lauer 1986, Monmonier 1983, and Marble 1984). Marble (1984) has stated the need to aggregate and generalize spatial data for data reduction purposes as data bases grow and approach global coverage. Hole and Campbell (1985) have expressed the need for generalization in order to reduce complex data into a form suitable for comprehension of essential elements. Generalization simplifies complex detail presented on aerial images into a form convenient for use by the map reader (Campbell, 1983).

Considerations for Generalization

Many different factors have been listed by authors pertaining to the essential considerations one must make for effective generalization. A frequently stated consideration is thorough knowledge of the data to be generalized. Hole and Campbell (1985) listed 1) knowledge of the pattern to be generalized and 2) knowledge of the cartographic model used to represent the pattern for effective generalization. Monmonier (1983) stated that the principal considerations in data generalization are; 1) type of feature (points, line, or area), 2) feature density, 3) degree of scale reduction and 4)type of display device. Knowledge of scale is also listed as a primary consideration by Crapper, et al (1986). Logan and Woodcock (1984) selected diagonal pixel connections and a comprehension of the spatial displacement and class frequency change that occurs as a result of generalization as their primary considerations.

Systematic Sampling Generalization

An early generalization technique developed for scale reduction was systematic sampling. This was a fast and efficient technique but was found to systematically favor certain kinds of parcels on the basis of sizes, shapes, and locations of delineations (Nichols 1975). For example, long thin, linear features such as streams and rivers may be eliminated from the generalized image is the sampling intervals are spaced far enough apart.

Evolution of Spatial Filtering

A second early technique used for generalization purposes was spatial filtering. Many of the filtering procedures used for spatial generalization have evolved from techniques first used in electrical engineering and communications science (Guptill, 1978). Early filters designed for nominal data generalization were largely centered around binary-valued functions, such as those encountered in picture processing (Guptill, 1978). Examples of binary-valued filtering functions may be found in Rosenfeld (1964) and Andrews (1970).

Research in Spatial Filtering

Research in spatial filtering of classified images has often stemmed from two different motivations and their associated assumptions concerning the accuracy of classification. In one case the classification was assumed to have errors and the use of spatial filters was utilized to improve accuracy (Guptill 1976, Goldberg and Goodenough 1978). These models were based on probability functions derived from each pixel's neighboring values. Examples of early probability filters have been done by MacDogall (1972), Strong and Rosenfeld (1973), and Guptill (1975).

The contrasting situation assumed accurate classification and spatial filtering improved spatial coherence (Logan and Woodcock, 1986).

Davis and Peet's Area Filtering Algorithm

In addition to removing groups of pixels which fall below a minimum mapping unit, as in area filtering, Davis and Peet (1976) developed a procedure which varied the minimal acceptable size from feature to feature. Numerous area filtering procedures had been published which changed the class of a pixel if the pixel in question was too different from its neighbors, according to a predetermined set of criteria (Goldberg et al 1975, Kan and Lo 1975). However, Davis and Peet reported that the majority of these procedures did not have the capability of altering the minimum area by category (Davis and Peet, 1976). Although Davis and Peet's area filtering procedure proved to be flexible, it was designed for raster data only and the algorithm did not include conversion of data to vector format upon completion of filtering.

Nichols' Area Filtering Algorithm

One of the first raster to vector transfer algorithms which incorporated an area filtering generalization strategy was the one developed by Nichols (1981). Nichols' procedure provided the capability of eliminating polygons which did not satisfy a given area threshold requirement, however it was not adjustable from feature to feature like the algorithm developed by Davis and Peet (1976). Nichols based his algorithm on the raster to vector procedure described by Rosenfeld (1978). Although Nichols' algorithm incorporated a generalization strategy in converting raster data to vector data, the area filtering procedure did not have the flexibility of altering the minimum threshold size by category as in the procedure developed by Davis and Peet (1976).

Logan and Woodcock's Utilization of a Pairwise Weight Table

Logan and Woodcock (1984) tried several approaches designed to reduce high spatial variance when generalizing timber maps. Their techniques included: 1) Davis and Peet's minimum area spatial filter modified to accept apriori class conversion weights, 2) the IBIS procedure which was actually several techniques involving image magnification, a modal filter and individual class conversion weights, and 3)a voting rule technique for existing manually delineated timber stand maps (Logan and Woodcock, 1984). Their raster to vector conversion algorithm was based on Nichols' (1981) algorithm but incorporated the ability to modify minimum threshold sizes by category and assign a weighting factor to each individual category. Logan and Woodcock (1984) found that class conversion weights were useful for tailoring a generalization technique to a particular user.

Monmonier's Area Filtering Algorithm

Monmonier (1983) implemented generalization techniques on vector land use/land cover data after transferring the information into a raster format. Upon completion of the generalization process, Monmonier converted the data back to its original vector format. His generalization approaches included; 1)suppression of land cover categories, and 2)display priority weighting of land cover categories. Monmonier applied these two generalization approaches in gap bridging, local erode operators, and halo-bias coefficients for tailoring land use/land cover maps to a particular user.

Gap bridging required a list of visual-dominance rankings or a pairwise weight table. When two occurrences of a higher ranked category are separated by one or more lower ranked categories, the middle cells are reassigned to the higher ranked category. An erode operator consisted of a moving 3 x 3 window. The center of the window was placed over each cell of the data. The category in the center was compared with its neighbors in the window. Erosion took place by reassigning the category at the center if two or more mutually contiguous cells did not have this same category. The reassignment was determined by the frequency and weights of surrounding cells. Finally, halo-bias coefficients consisted of a weighting scheme based on distance from the interior of a moving window (Monmonier, 1983).

Monmonier (1983) justified his conversion of vector data to raster prior to generalization, and back to vector after generalization for two reasons; 1) the increasing demand for the cartographic products of remote sensing, and 2) the likely replacement of the pen plotter by raster mode display technology.

Other Vector-based Generalization Strategies

Other generalization strategies have been applied to vector data in order to reduce data volume. Several procedures have been published on automated smoothing of digitized lines, especially those procedures which retain points that make significant changes in direction and those that fit a smoothed interpolation of the original line, as well as others (Douglas and Peucker, 1973). Line smoothing would further reduce data volume if applied after area filtering, resampling or category aggregation.

MATERIALS AND METHODS

Computer Hardware and Software

This study was conducted using the EROS Data Center Land Analysis System for generalizing the USGS Interim Land Cover Data. LAS was run on a Digital Equipment Corporation (DEC) VAX 11/780 computer with eight megabytes of main memory. The resultant images were viewed on a DeAnza color monitor with a trackball for manipulation and display of image data. The Applicon AP5500 color plotter was used for paper copies of the land cover data (EDC LAS User's Manual, 1986).

Valdez Land Cover Data

Generalization strategies were operated on a portion of the USGS Valdez Interim Land Cover Data. The entire USGS Valdez land cover data consisted of a classified image of 2370 by 3500 pixels. The Valdez Interim Land Cover data covered the entire USGS 1:250,000 scale quadrangle and occupied the area between 144 and 147 degrees W longitude and 61 to 62 degrees N latitude. Each pixel occupied 50 meters on a side, for a total of 0.618 acres (0.250 hectares) ground area. The entire Valdez classified image was displayed on a Universal Transverse Mercator (UTM) projection.

Valdez Study Area

A subscene from the Valdez Interim Land Cover data was selected for testing purposes in order to reduce computer processing time and memory storage. This test area occupied the upper right hand corner of the USGS 1:250,000 series Valdez quadrangle. The entire test area comprised one-eighth of the Valdez quadrangle (Figure 1). The selected subscene was a 1185 by 875 classified pixel image, and contained the Interim Land Cover categories summarized in Table 2. Although the test area did not contain all the USGS Interim Land Cover categories, it was chosen for generalization testing because it contained the only significant coverage of the mixed forest category, which was deemed a critical category by resource personnel for effective land management.

Generalization Strategies

Four strategies were chosen for generalizing the Valdez land cover test area: 1) area filtering, 2) category aggregation, 3) resampling, and 4) modal search and replace. These strategies were selected based on procedures which are currently being implemented by other scientists for data reduction, and on the availability of LAS functions which lend themselves to generalizing digital data.

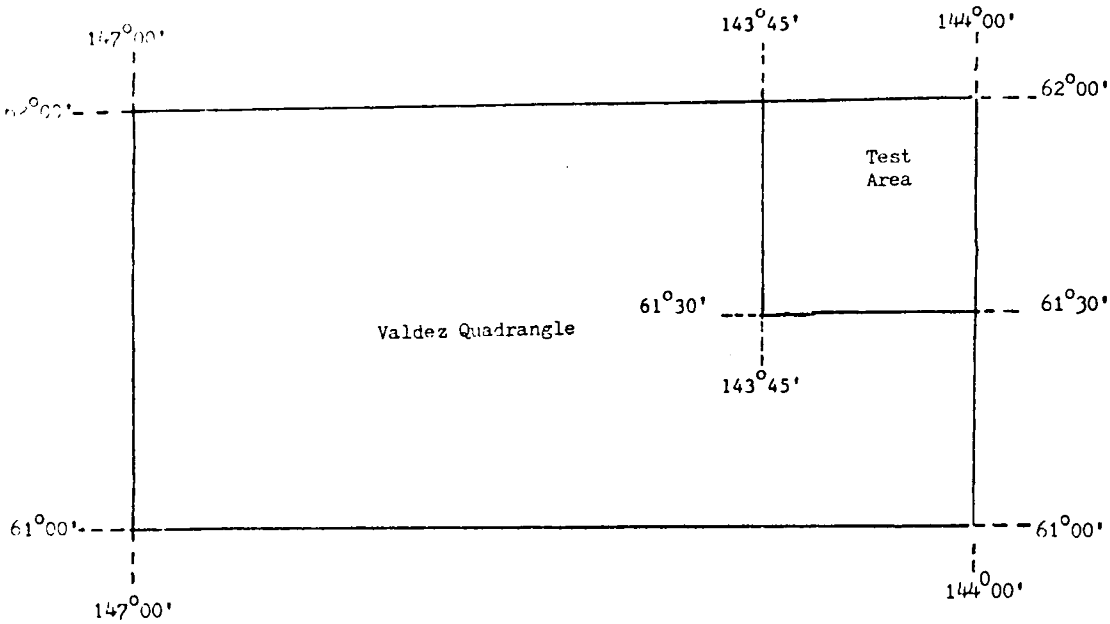


FIGURE 1

Valdez Generalization Test Area

TABLE 2

Composition of Ungeneralized
USGS Interim Land Cover Test Area

Category		Percent of Classified Image
IA	Needleleaf Forest	29.78
IB	Broadleaf Forest	1.41
IC	Mixed Forest	4.79
IA	Shrublands	21.08
IIB	Dwarf Shrublands	13.44
VIA	Sparse Vegetation	1.28
VIB	Non-vegetated	10.70
VIIA	Clear/Deep Water	0.42
VIIB	Turbid/Shallow Water	1.73
VIII	Ice and Snow	10.33
IX	Shadow	5.04
		TOTAL = 100.00

Types of Generalization Strategies

Hole and Campbell (1985) list eight different operations which may be performed in order to generalize a map: 1) smoothing, 2) enlargement of parcels, 3) deletion of boundary segments, 4) combination of mapping units, 5) deletion of parcels, 6) addition of boundaries, 7) assignment of ranks to boundaries, and 8) shifts in boundary position. These operations may be used singularly or in combination to perform four different generalization strategies given by Hole and Campbell (1985): 1) graphic generalization, 2) taxonomic generalization, 3) spatial generalization, and 4) generalization by sampling.

The area filtering and modal search and replace algorithms may be considered a form of graphic generalization which has the objective of reducing visual complexity by reducing the total line length on the map. The category aggregation algorithm may be thought of as a form of taxonomic generalization which is accomplished by the removal of boundaries between neighboring parcels that are taxonomically similar. Finally, the systematic resampling algorithm may be considered as a form of generalization by sampling (Hole and Campbell, 1985).

Area Filtering

Area filtering was the first generalization strategy implemented on the Valdez land cover data. Filtering may be used for two very different purposes. First, a classified image may have an unknown accuracy and filtering may be used to reduce the number of pixels thought to contain error. On the other hand, the classified image may be assumed to be correct and filtering is used for data reduction. This research implemented area filtering in order to reclassify the many small pixel homogeneous areas, or regions. Area filtering may be defined as the reclassification of homogeneous regions which do not contribute to the information needed to solve problems defined by the user of the filtered image (Fosnight, 1985). Because this research was exploratory in nature, several different minimum threshold areas were selected for implementation and the impacts on the land cover categories and total number of regions were studied.

LAS Area Filtering Algorithm

The LAS area filtering strategy used in this project was based on a procedure first outlined by Nichols (Nichols, 1981). First, regions, or groups of pixels with homogeneous land use/land cover categories, were identified. These regions were labeled with a unique region number and stored with the pixel value comprising that region and the number of pixels present within that region. Regions which fell below the specified minimum area threshold, were flagged by assigning those pixels a negative value. Finally, the negative regions were replaced by the most frequently occurring land use/land cover value which surrounded each negative region (Figure 2).

Variations in Area Filtering Algorithm

Variation in the area filtering process may take place in two places. First, the minimum threshold area value may be altered for the entire classification system, or by land cover category. This threshold value may be said to specify a minimum size requirement for a land cover type to be left unfiltered. Regions smaller than the size specified, are flagged with a negative value. The threshold values come into play after the region labeling procedure has been complete, but before the flagging procedure initiates (Figure 2).

Another variation may take place after the flagged regions are given negative values and before the smoothing procedure begins. This filtering variation involves a

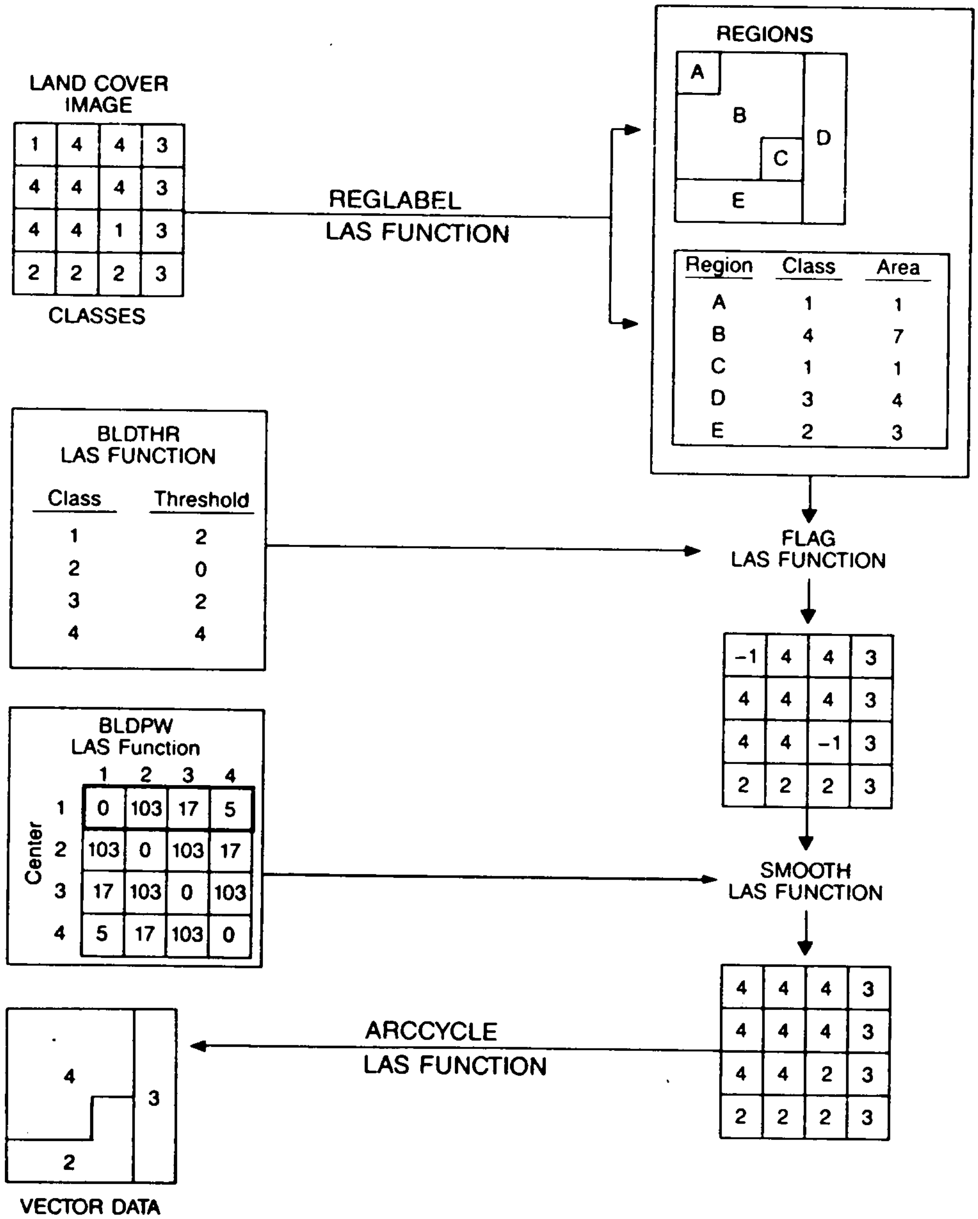


FIGURE 2

Area Filtering Procedure

For Alaska Land Cover Data Reduction

Prior to Incorporation in the ARC/INFO GIS pairwise weight table. A pairwise weight table consists of a list of every possible category pair and its associated numeric weight which is used to reclassify a flagged pixel (Table 3). A pairwise weight table may also cause filtering to be one-way (unidirectional) or two-way (bidirectional). One way filtering allows a specific category to be replaced in the smoothing process by another category but the reverse replacement is not allowed (i.e., water may be replaced by forest but forest cannot be replaced by water). A value of zero is assigned to the pairwise weight table for those categories which are wished to be excluded from the reassignment of a flagged region, while all other categories are given a value greater than zero. Thus, when the values in the pairwise weight table are multiplied with the classified pixels for reassignment of a flagged region, the zero values in the pairwise weight table remove that category from consideration (Table 3). Two-way filtering gives the same weighting for a pair of categories in either direction.

Minimum Threshold Sizes

A number of different minimum threshold sizes were run on the unfiltered land cover data. Table 4 shows the different threshold or minimum pixel areas used. These

TABLE 3
Pairwise Weight Table for a Unidirectional Shadow Filter

	NF	BF	MF	SV	NV	SL	DS	DW	SW	IS	SD
NF	1	1	1	1	1	1	1	1	1	1	0
BF	1	1	1	1	1	1	1	1	1	1	0
MF	1	1	1	1	1	1	1	1	1	1	0
SV	1	1	1	1	1	1	1	1	1	1	0
NV	1	1	1	1	1	1	1	1	1	1	0
SL	1	1	1	1	1	1	1	1	1	1	0
DS	1	1	1	1	1	1	1	1	1	1	0
DW	1	1	1	1	1	1	1	1	1	1	0
SW	1	1	1	1	1	1	1	1	1	1	0
IS	1	1	1	1	1	1	1	1	1	1	0
SD	1	1	1	1	1	1	1	1	1	1	1

Rows = Category which has been flagged for removal
Columns = Category which will replace flagged region

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SV = Sparse Vegetation
NV = Non-Vegetated
SL = Shrublands
DS = Dwarf Shrublands
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 4
Minimum Threshold Sizes Selected for Area Filtering

Threshold Size (Pixels)	Ground Area (Hectares)
2	0.5
16	4
68	17
100	25
128	32
256	64
512	128

sizes were selected based on initial studies done by Sturdevant (Sturdevant, 1986).

As each area filtering test was completed, a pixel count and a contingency table were generated in order to observe the change in the overall image and within category change. Each filtered image was compared to the original unfiltered Valdez land cover test area. Finally, the LAS region labeling function was run on each resultant image in order to observe the change in the number of homogeneous pixel regions.

Statistical Measurements

Five different calculations were made on each area filtered image: 1) percent unchanged with respect to the original unfiltered image, 2) errors of omission, 3) errors of comission, 4) Cohen's KHAT statistic of agreement, and 5) Monmonier's fragmentation index (Table 5). An accuracy assessment was never carried out on the USGS Interim Land Cover data after it was classified and processed from Landsat data. Therefore, the statistical measurements made on the generalized data are based upon the original, ungeneralized land cover data and not the actual ground area.

TABLE 5
Statistical Measurements

1. PERCENT UNCHANGED = $\sum Xij/N$

$\sum Xij$ = the sum of all diagonal elements in a contingency table
N = the total number of elements

2. PROBABILITY OF NON-OMISSION = (1 - (Xij/Xi)) x 100

Xij = the number of elements in a particular generalized category which correctly represent the original data
Xi = the total number of elements in a particular category in the original image

3. PROBABILITY OF NON-COMISSION = (1 - (Xij/Xj)) x 100

Xij = the number of elements in a particular generalized category which correctly represent the original data

Xj = the total number of elements in a particular category in the generalized image

4. COHEN'S KHAT STATISTIC = ((N* Xij)-Xi+X+i)/(N -Xi+X+i)

N = the total number of elements in a contingency table
$\sum Xij$ = the sum of all the diagonal elements in a contingency table
Xi+X+i = the total sum of each category row and column product

5. FRAGMENTATION INDEX (F) = ((m-1)/(n-1))/Fs

m = the number of regions or parcels
n = the number of map units or cells
Fs = (m-1)/(n-1) Monmonier's fragmentation index for the initial, ungeneralized image

Percent Unchanged Vs Cohen's KHAT Statistic

Overall percent unchanged, or more commonly known as percent unchanged, was calculated because it is a common, familiar term to many resource managers and scientists. Both percent unchanged and KHAT are reported in this study, however, since KHAT is a discrete multivariate analysis technique used to test the overall agreement between generalization techniques (Congalton and Mead, 1983). KHAT is an appropriate statistical measurement because the land cover data are discrete and multinomially distributed. KHAT is the maximum likelihood estimate from the multinomial distribution and is the measure of the actual agreement minus the chance agreement. The actual agreement is the cell value itself while the chance agreement is defined as the product of the marginals. The approximate large sample variance of KHAT can then be used to construct a hypothesis test for significance difference between generalized error matrices. Cohen's KHAT statistic was developed from Cohen's Kappa for nominal scales which measured the relationship of chance agreement to actual agreement (Cohen, 1960). Rosenfield and Fitzpatrick-Lins state that Cohen's KHAT measure of agreement is a better estimator because it incorporates column and row totals of a contingency matrix, not just the diagonal elements (Rosenfield and Fitzpatrick-Lins, 1986).

Omission and Comission Errors

Errors of omission, sometimes referred to as producer's accuracy, were calculated for each category in order to determine the likelihood than an unfiltered pixel will be correctly represented in the filtered image. Errors of comission, sometimes referred to a user's accuracy, were calculated for each category in order to determine the probability that a filtered pixel will correctly represent the corresponding unfiltered pixel (Story and Congalton, 1986). The probability of non-omission and probability of non-comission were derived by subtracting the omission and comission errors from one, and then multiplying the result by 100 to obtain a percent probability.

Fragmentation Index

Many different diversity indices have been developed for digital grid cell maps to measure their complexity (Robinove, 1986). Monmonier's fragmentation index was calculated at the conclusion of each area filtering test in order to measure the complexity of the generalized data (Monmonier, 1974). Monmonier's fragmentation index (F) ranges from zero for a simple map to 1 for a highly complex map (Table 5). After calculating Monmonier's fragmentation index, the indices were normalized by dividing each F by the fragmentation index calculated for the initial, ungeneralized image. This was done in order to make the results comparable to the ungeneralized image, rather than a map in which every pixel contained a different classification value. The resulting normalized fragmentation values range from 1 (ungeneralized), to 0 (very generalized).

Unidirectional Shadow Area Filters

A second set of area filtering tests was carried out which utilized the same minimum threshold sizes as listed in Figure 3, however the pairwise weight table was manipulated to create a unidirectional shadow filter. Shadow was allowed to be reclassified to any other land cover type but other land cover types were not allowed to be changed to shadow. All shadow to category pairs will be given a value of 1 while all category to shadow pairs will be given a value of 0 (Table 3). The purpose of conducting the unidirectional shadow area filtering tests was to reduce the number of homogeneous regions while restricting the increase of an non-informational category (shadow).

		Generalized Image (j)								Xi
Original Image	i	1	4	7	9	1	6	6	2	36
		9	5	9	1	1	8	3	5	41
		6	3	9	7	9	1	4	2	41
		9	7	9	5	6	3	2	1	42
		8	1	1	5	4	4	3	9	35
		7	6	0	7	4	4	8	1	37
		0	5	9	8	2	3	1	7	35
		6	8	7	5	5	9	2	3	45
	Xj	46	39	51	47	32	38	29	30	312

N = 312

FIGURE 3
Example Contingency Table

Category Aggregation

The second strategy tested for generalization was category aggregation. As already mentioned, the USGS Interim Land Cover data are composed of level I and II category precision. Precision, in this case, may be thought of as the amount of detail present in a map or image. The level II categories have a higher precision, or level of detail as compared to the level I categories. All level II categories present in the Valdez test area were aggregated to a more general level I category. This was possible because of the hierarchical nature of the classification system. Table 6 illustrates the aggregation strategy. Once the categories were aggregated, the LAS region labeling function was run to determine the reduction in number of homogeneous pixel areas, or regions. Finally, the aggregated categories were filtered using the same minimum threshold pixel sizes listed in Table 4. The area filtered, aggregated image results were compared against the unfiltered level I image, rather than the original land cover data. The five statistical measurements were calculated, as before (Table 5) on the resultant images.

TABLE 6
Category Aggregation Strategy

Level II Categories	Aggregated Level I Categories
Needleleaf Forest Broadleaf Forest Mixed Forest	Forest
Shrublands Dwarf Shrublands	Shrub
Sparse Vegetation Non-Vegetated	Barren
Deep/Clear Water Shallow/Turbid Water	Water
	*Ice and Snow
	*Shadow

*Category which was already at Level I precision and did not require aggregation.

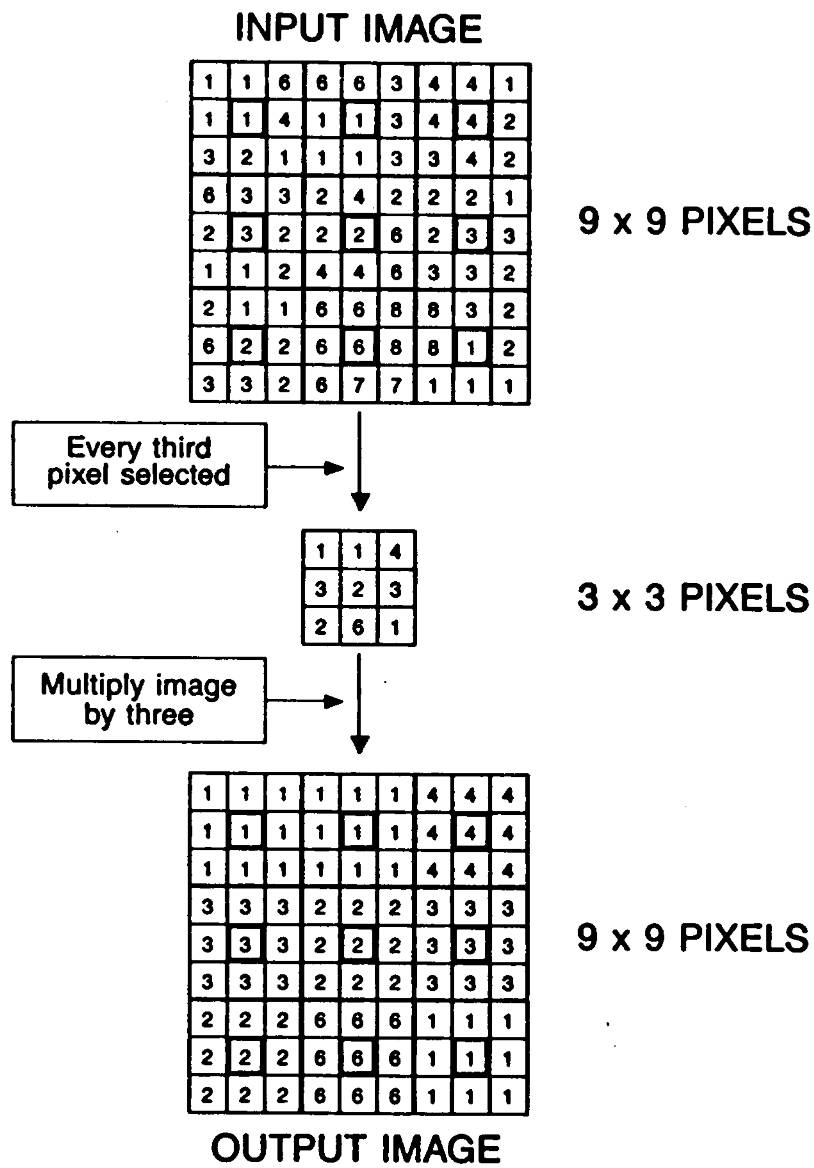
Systematic Resampling

The third generalization strategy examined in this study was systematic resampling. Systematic resampling was accomplished by imposing larger area grids over the land cover cells and selecting the central land cover cell within each imposed grid (Figure 4). After resampling, the resultant pixels were multiplied by a constant in order to return the land cover images back to their original size (Figure 4). Table 7 gives the imposed grid sizes used in the resampling tests.

Overall accuracy, errors of omission, errors of comission, and Cohen's KHAT statistic were calculated for each of the three resultant resampled images (Table 5). Contingency tables were generated comparing the resampled image to the original Valdez test area land cover data, as before.

Modal Search and Replace

The final generalization strategy examined in this study was modal search and replace. This strategy utilized the LAS cartographic neighborhood technique. For a given input pixel radius, each pixel is used as the center of a modal tabulation. The most frequently occurring land cover



Example of Resampled Data with an Imposed

Grid Size of 150 meters (3 pixels)

FIGURE 4

Systematic Resampling Algonthm

Systematic Resampling

TABLE 7
Grid Sizes Used in Resampling Generalization Strategy

Ground Size Ground Size	Every Nth Pixel Selected	Resultant Image Multiplication Factor
Grid1 = 150 meters	N = 3	3
Grid2 = 250 meters	N = 5	5
Grid3 = 350 meters	N = 7	7

category within the specified radius is used to replace, if necessary, the existing center pixel (Figure 5). Table 8 lists the radii used for the cartographic neighborhood function.

Overall accuracy, omission errors, comission errors and Cohen's KHAT statistic were all calculated for each of the three modal search and replace tests. Contingency tables were generated comparing the resultant images against the original Valdez test area land cover data, as before.

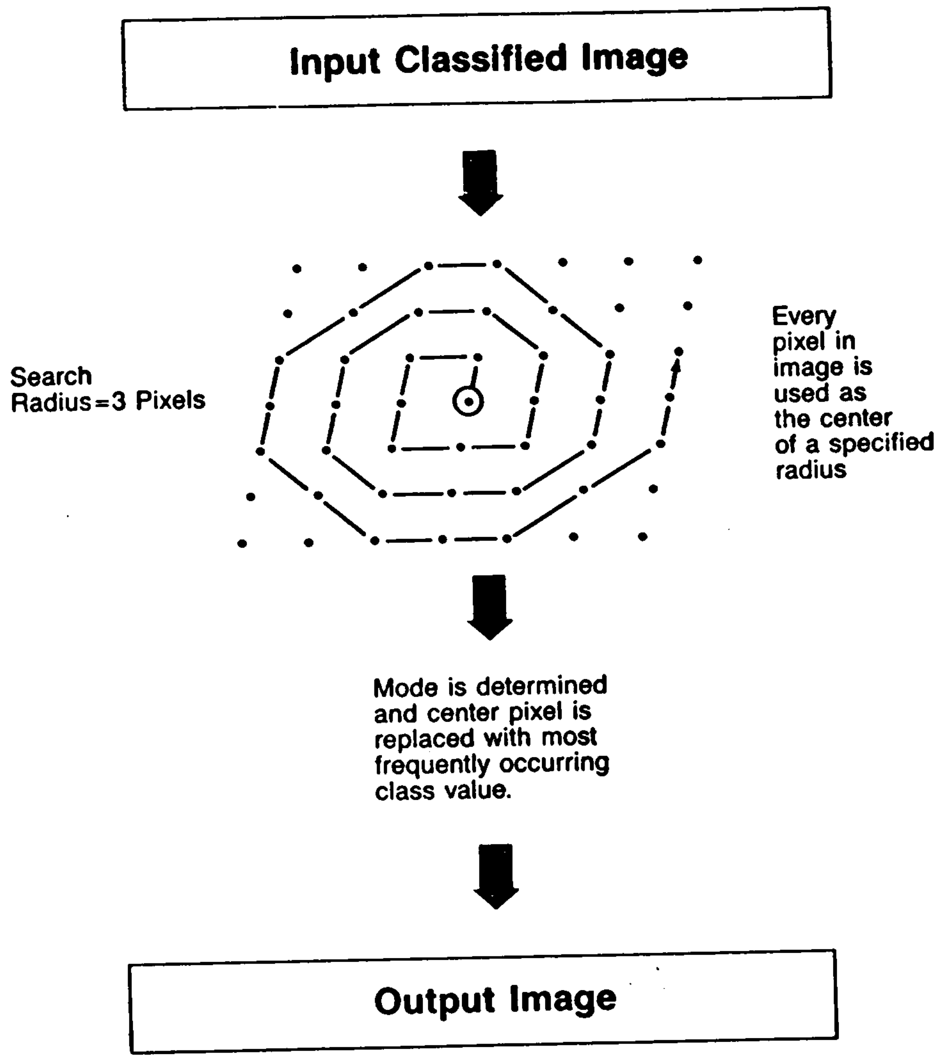


FIGURE 5

Modal Search and Replace Algonthm

TABLE 8

Radii Used for LAS NEIGHBOR Function

Radius (Pixels)	Search Area (Pixels)	Ground Area (Hectares)
1	3 x 3	2.25
2	5 x 5	6.25
3	7 x 7	12.25

RESULTS

Results from the statistical analyses applied at the conclusion of each generalization strategy are listed in the following tables. The tables are grouped in the following order according to each of the four generalization strategies studied: 1) Area Filtering, 2) Category Aggregation, 3) Modal Search and replace, and 4) Systematic Resampling. Corresponding to each strategy, there are seven tables which list the results from the applied statistical tests: 1) Percent of Total Image by Category, 2) Region Reduction, 3) Percent Unchanged with Original Image, 4) Probability of Non-Omission, 5) Probability of Non-Comission, 6) Cohen's KHAT Statistic, and 7) Fragmentation Index.

These tables are included for the sake of replication validation and future applications of this study in generalization. For a more complete discussion of the impact these generalization strategies had on the land use data, see the Discussion and Conclusion sections of this document.

AREA FILTERING

TABLE 9
Percent of Total Image by Category
After Area Filtering

	Minimum Threshold Size (Pixels)							
CATEGORY	0	2	16	68	100	128	256	512
NF	29.78	29.80	30.26	31.11	31.35	31.42	32.07	33.72
BF	1.41	1.39	1.21	1.00	0.86	0.85	0.46	0.00
MF	4.79	4.75	4.15	3.11	2.70	2.43	1.72	0.93
SL	21.08	21.10	21.45	22.04	22.28	22.57	22.96	24.07
DS	13.44	13.46	13.76	14.59	15.02	15.39	15.97	15.57
SV	1.28	1.27	0.92	0.39	0.27	0.14	0.00	0.00
NV	10.70	10.70	10.66	10.27	10.29	10.48	10.52	10.57
DW	0.42	0.42	0.38	0.30	0.26	0.22	0.19	0.08
SW	1.73	1.74	1.81	1.97	2.04	2.03	2.19	2.16
IS	10.33	10.33	10.26	10.13	10.15	9.99	9.95	10.10
SD	5.04	5.04	5.14	5.09	4.78	4.48	3.97	2.80

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DW = Deep/Clear Water
SW = Shallow/Turbid Water
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
IS = Ice and Snow
SD = Shadow

TABLE 10
Region Reduction After Area Filtering

Minimum Threshold Area (Pixels)	Number of Regions	Percent Reduction in Number of Regions
0	9278	0.00
2	7138	23.07
16	1916	79.35
68	466	94.98
100	310	96.66
128	249	97.32
256	121	98.70
512	59	99.36

TABLE 11
Percent Unchanged With Original Unfiltered Image

Minimum Threshold Area (Pixels)	Percent Unchanged
2	99.69
16	95.33
68	88.54
100	86.65
128	85.47
256	81.63
512	77.48

TABLE 12
Probability of Non-Omission
for Area Filtered Categories

(Probability that an Unfiltered Pixel will be Correctly Represented in the Filtered Image)

	Minimum Threshold Area (Pixels)						
	2	16	68	100	128	256	512
CATEGORY							
NF	99.81	97.41	94.03	92.97	92.39	90.96	90.10
BF	98.58	80.80	58.44	48.02	46.76	24.26	0.00
MF	98.66	79.58	51.39	44.13	39.54	27.85	15.36
SL	99.84	97.63	94.12	93.40	93.03	90.30	87.80
DS	99.78	96.63	91.71	90.75	90.35	87.08	80.85
SV	97.92	65.48	24.03	15.79	8.01	0.00	0.00
NV	99.62	93.68	82.69	80.04	79.02	73.44	68.82
DW	99.43	88.69	69.62	60.16	52.04	45.57	19.82
SW	99.91	98.76	96.82	96.27	96.27	96.47	93.50
IS	99.88	97.89	94.78	94.43	93.44	92.05	91.21
SD	99.66	94.28	82.77	76.51	70.64	57.59	40.80

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 13
Probability of Non-Comission
for Area Filtered Categories

(Probability that a Filtered Pixel will Correctly Represent the Corresponding Unfiltered Pixel)

	Minimum Threshold Area (Pixels)						
	2	16	68	100	128	256	512
CATEGORY							
NF	99.71	95.85	89.99	88.31	87.56	84.44	79.55
BF	99.70	94.00	82.46	78.63	77.63	74.35	0.00
MF	99.43	91.74	79.24	78.40	77.92	77.74	78.77
SL	99.74	95.95	90.00	88.36	86.95	82.91	76.91
DS	99.64	94.43	84.47	81.19	78.95	73.29	69.83
SV	99.40	91.03	79.55	75.57	72.82	0.00	0.00
NV	99.63	94.05	86.23	83.28	80.74	74.70	69.68
DW	99.57	97.46	96.55	97.24	99.86	99.84	99.64
SW	99.66	94.64	85.24	81.68	80.69	76.37	75.01
IS	99.92	98.50	96.59	96.04	96.60	95.56	93.27
SD	99.56	92.47	81.91	80.61	79.55	73.07	73.47

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 14
Cohen's KHAT Statistic for Area Filtered Images

(Measure of Overall Agreement Between Filtered and Unfiltered Images)

Minimum Threshold Area (Pixels)	KHAT (Percent)
2	99.63
16	94.30
68	85.94
100	83.60
128	82.13
256	77.32
512	71.29

TABLE 15
Fragmentation Index (F) for Area Filtered Images

Minimum Threshold Area (Pixels)	F
2	0.76928
16	0.20669
68	0.05021
100	0.03347
128	0.02692
256	0.01310
512	0.00655

Monmonier's Fragmentation Index for initial, ungeneralized image = 0.01374

UNIDIRECTIONAL SHADOW AREA FILTERING RESULTS

TABLE 16
Percent of Total Image by Category After Area Filtering with a Unidirectional Shadow Filter

	Minimum Threshold Size (Pixels)							
	0	2	16	68	100	128	256	512
CATEGORY								
NF	29.78	29.81	30.31	31.23	31.46	31.52	32.21	33.80
BF	1.41	1.39	1.21	1.00	0.86	0.85	0.46	0.00
MF	4.79	4.75	4.20	3.18	2.76	2.48	1.73	0.95
SL	21.08	21.10	21.49	22.18	22.42	22.70	23.08	24.10
DS	13.44	13.47	13.86	14.95	15.43	15.78	16.49	15.77
SV	1.28	1.26	0.94	0.40	0.29	0.17	0.00	0.00
NV	10.70	10.71	10.75	10.46	10.45	10.66	10.81	10.90
DW	0.42	0.42	0.38	0.30	0.26	0.22	0.19	0.08
SW	1.73	1.74	1.81	1.97	2.04	2.07	2.19	2.16
IS	10.33	10.33	10.30	10.18	10.20	10.02	9.97	10.21
SD	5.04	5.02	4.75	4.14	3.83	3.53	2.87	2.03

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
SD = Shadow

DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
IS = Ice and Snow

TABLE 17
Region Reduction After Area Filtering with a Unidirectional Shadow Filter

Minimum Threshold Area (Pixels)	Number of Regions	Percent Reduction in Number of Regions
0	9278	0.00
2	7158	22.85
16	2015	78.28
68	562	93.94
100	401	95.68
128	336	96.38
256	199	97.86
512	107	98.85

TABLE 18
Percent Unchanged With Original Unfiltered Image

Minimum Threshold Area (Pixels)	Percent Unchanged
2	99.69
16	95.39
68	88.65
100	86.76
128	85.60
256	81.79
512	77.65

TABLE 19
Probability of Non-Omission for Unidirectional Shadow Area Filtered Images

(Probability that an Unfiltered Pixel will be Correctly Represented in the Filtered Image)

	Minimum Threshold Area (Pixels) (Unidirectional Shadow Filter)						
	2	16	68	100	128	256	512
CATEGORY							
NF	99.81	97.45	94.11	93.05	92.47	91.09	90.11
BF	98.58	80.81	58.42	47.98	46.80	24.25	0.00
MF	98.67	79.80	51.82	44.60	39.98	28.13	15.59
SL	99.84	97.64	94.16	93.44	93.07	90.31	87.80
DS	99.78	96.66	91.86	90.92	90.62	87.49	80.97
SV	97.92	65.68	24.21	16.07	8.36	0.14	0.14
NV	99.62	93.83	83.13	80.42	79.46	74.11	70.16
DW	99.43	88.69	69.62	60.16	52.05	45.57	19.82
SW	99.91	98.79	96.87	96.32	96.32	96.48	93.52
IS	99.88	97.98	94.91	94.56	93.55	92.12	91.35
SD	99.66	94.27	82.43	75.99	70.15	56.91	40.30

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 20
Probability of Non-Comission for Unidirectional Shadow Area Filtered Categories

(Probability that a Filtered Pixel will Correctly Represent the Corresponding Unfiltered Pixel)

	Minimum Threshold Area (Unidirectional Shadow Filter)						
	2	16	68	100	128	256	512
CATEGORY							
NF	99.70	95.72	89.71	88.05	87.36	84.21	79.37
BF	99.64	94.05	82.50	78.64	77.66	73.49	0.00
MF	99.37	90.83	78.11	77.38	77.25	77.63	78.99
SL	99.73	95.75	89.50	87.86	86.42	82.49	76.80
DS	99.59	93.77	82.59	79.22	77.16	71.32	69.03
SV	99.34	89.93	76.84	71.42	64.76	100.00	100.00
NV	99.60	93.45	85.07	82.37	79.76	73.41	68.87
DW	99.61	97.19	96.45	97.30	99.86	99.84	99.64
SW	99.66	94.63	85.25	81.69	80.63	76.40	75.02
IS	99.92	98.28	96.32	95.76	96.45	95.44	92.46
SD	100.00	100.00	100.00	100.00	100.00	100.00	100.00

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 21
Cohen's KHAT Statistic for Unidirectional Shadow Area Filtered Images

(Measure of Overall Agreement Between Filtered and Unfiltered Images)

Minimum Threshold Area (Pixels)	KHAT (Percent)
2	99.63
16	94.36
68	86.07
100	83.72
128	82.26
256	77.48
512	72.18

TABLE 22
Fragmentation Index (F) for Unidirectional Shadow Area Filtered Images

Minimum Threshold Area (Pixels)	F
2	0.77147
16	0.21688
68	0.06040
100	0.04294
128	0.03566
256	0.02110
512	0.01091

Monmonier's Fragmentation Index for initial, ungeneralized image = 0.01374

CATEGORY AGGREGATION TO LEVEL I PRECISION AND AREA FILTERING

TABLE 23
Percent of Total Image by Category After Category Aggregation and Area Filtering

	Minimum Threshold Area (Pixels)							
	0	2	16	68	100	128	256	512
CATEGORY								
FT	58.32	58.32	58.39	58.56	58.49	58.62	58.49	58.65
SH	22.48	22.48	22.61	22.98	23.26	23.42	23.99	24.35
BN	7.80	7.80	7.65	7.38	7.33	7.34	7.25	7.39
WT	1.40	1.40	1.41	1.43	1.45	1.44	1.50	1.42
IS	6.72	6.72	6.67	6.57	6.57	6.48	6.43	6.49
SD	3.28	3.28	3.27	3.08	2.91	2.70	2.34	1.70

LEGEND

FT = Forest
SH = Shrub
BN = Barren
WT = Water
IS = Ice and Snow
SD = Shadow

TABLE 24
Region Reduction After Category Aggregation and Area Filtering

Minimum Threshold Area (Pixels)	Number of Regions	*Percent Reduction in Number of Regions
0	5141	0.00
2	4050	21.22
16	1222	76.33
68	332	93.54
100	219	95.74
128	177	96.56
256	84	98.37
512	41	99.20

*Percent reduction in number of regions is based on comparisons with the aggregated, unfiltered image.

TABLE 25
Percent Unchanged with Unfiltered Aggregated Image

Minimum Threshold Area (Pixels)	Percent Unchanged
2	99.84
16	97.48
68	93.22
100	91.85
128	91.10
256	88.44
512	85.57

TABLE 26
Probability of Non-Omission for Aggregated Area Filtered Categories

(Probability that an Unfiltered Pixel will be Correctly Represented in the Aggregated Filtered Image)

	Minimum Threshold Area (Pixels)						
	2	16	68	100	128	256	512
CATEGORY							
FT	99.86	98.03	94.67	93.44	93.19	91.35	89.60
SH	99.91	98.46	95.76	95.10	94.92	93.44	91.14
BN	99.66	94.14	85.05	82.62	81.22	75.64	72.69
WT	99.82	96.89	91.60	89.31	87.76	86.66	79.28
IS	99.88	97.89	94.78	94.42	93.42	92.03	91.14
SD	99.66	94.27	82.49	76.10	70.29	57.44	40.58

Legend

FT = Forest
SH = Shrublands
BN = Barren
WT =Water
IS = Ice and Snow
SD = Shadow

TABLE 27
Probability of Non-Comission for Aggregated
Area Filtered Categories

(Probability that a Filtered Pixel will Correctly Represent the Corresponding Aggregated Unfiltered Pixel)

	Minimum Threshold Area (Pixels)						
CATEGORY	2	16	68	100	128	256	512
FT	99.86	97.73	93.78	92.87	92.16	91.07	88.74
SH	99.87	97.83	93.64	91.85	91.02	87.32	84.05
BN	99.74	96.06	89.83	87.77	86.12	81.24	76.00
WT	99.66	96.16	89.49	86.43	85.38	80.67	78.20
IS	99.94	98.68	96.98	96.57	96.99	96.25	94.22
SD	99.66	94.69	87.64	85.94	85.05	79.56	78.29

LEGEND

FT = Forest
SH = Shrubs
BN = Barren
WT = Water
IS = Ice and Snowater
SD = Shadow

TABLE 28
Cohen's KHAT Statistic for Aggregated
Area Filtered Images

(Measure of Overall Agreement Between Filtered and Unfiltered Aggregated Images)

Minimum Threshold Area (Pixels)	KHAT (Percent)
2	99.78
16	96.51
68	90.59
100	88.68
128	87.61
256	83.87
512	79.80

TABLE 29
Fragmentation Index (F) for Aggregated Area Filtered Images

Minimum Threshold Area (Pixels)	F
2	0.78843
16	0.23784
68	0.06438
100	0.04204
128	0.03416
256	0.01576
512	0.00788

Monmonier's Fragmentation Index for initial, aggregated image = 0.00761

RESAMPLING RESULTS

TABLE 30
Percent of Total Image by Category After Resampling

	Imposed Grid Size (Pixels)			
	0	3	5	7
CATEGORY				
NF	29.78	29.73	29.74	30.73
BF	1.41	1.39	1.42	0.92
MF	4.79	4.80	4.85	3.73
SL	21.08	21.05	20.99	21.85
DS	13.44	13.49	13.36	14.77
SV	1.28	1.29	1.32	0.64
NV	10.70	10.77	10.79	10.18
DW	0.42	0.42	0.44	0.29
SW	1.73	1.74	1.70	1.90
IS	10.33	10.30	10.37	10.13
SD	5.04	5.02	5.02	4.86

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands

DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
IS = Ice and Snow
SD = Shadow

TABLE 31
Region Reduction After Resampling

Imposed Grid Size (Pixels)	Number of Regions	Percent Reduction in Number of Regions
0	9278	0.00
3	6359	31.46
5	3425	63.08
7	2093	77.44

TABLE 32
Percent Unchanged With Original Unresampled Image

Imposed Grid Size (Pixels)	Percent Unchanged
3	83.99
5	75.33
7	70.37

TABLE 33
Probability of Non-Omission for Resampled Categories

(Probability that an Unresampled Pixel will be Correctly Represented in the Resampled Image)

	Imposed Grid Size (Pixels)		
	3	5	7
CATEGORY			
NF	88.56	82.69	79.55
BF	60.45	42.56	34.88
MF	61.59	44.64	36.52
SL	88.04	80.96	76.30
DS	83.65	73.83	68.60
SV	51.28	31.19	23.59
NV	77.06	64.78	57.10
DW	76.38	64.14	51.90
SW	87.60	78.92	73.10
IS	92.00	87.60	84.89
SD	74.96	59.02	49.14

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 34
Probability of Non-Comission for Resampled Categories

(Probability that a Resampled Pixel will Correctly Represent the Corresponding Unresampled Pixel)

	Imposed Grid Size (Pixels)		
	3	5	7
CATEGORY			
NF	88.61	82.64	79.25
BF	60.95	42.01	34.04
MF	61.37	43.98	36.33
SL	88.08	81.18	76.40
DS	83.31	74.15	67.35
SV	51.08	30.26	24.02
NV	76.54	64.14	57.85
DW	76.71	61.64	51.59
SW	87.43	79.95	74.15
IS	92.16	87.10	85.01
SD	75.17	59.18	48.75

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 35
Cohen's KHAT Statistic for Resampled Images

(Measure of Overall Agreement Between Resampled and Unresampled Images)

Imposed Grid Size (Pixels)	KHAT (Percent)
3	80.51
5	69.97
7	63.90

TABLE 36
Fragmentation Index (F) for Resampled Images

Imposed Grid Size (Pixels)	F
3	0.68543
5	0.36913
7	0.22556

Monmonier's Fragmentation Index for initial, ungeneralized image = 0.01374

MODAL SEARCH AND REPLACE RESULTS

TABLE 37
Percent of Total Image by Category After Modal Search and Replace

	Search Radius (Pixels)			
	0	1	2	3
CATEGORY				
NF	29.78	30.31	30.62	30.92
BF	1.41	1.30	1.18	1.09
MF	4.79	4.52	4.14	3.84
SL	21.08	21.18	21.43	21.65
DS	13.44	13.55	13.77	13.95
SV	1.28	1.11	0.93	0.76
NV	10.70	10.61	10.49	10.35
DW	0.42	0.39	0.37	0.34
SW	1.73	1.75	1.80	1.85
IS	10.33	10.31	10.28	10.26
SD	5.04	4.97	4.99	4.99

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
SD = Shadow

DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
IS = Ice and Snow

TABLE 38
Region Reduction After Modal Search and Replace

Search Radius (Pixels)	Number of Regions	Percent Reduction in Number of Regions
0	9278	0.00
1	5414	41.65
2	3333	64.08
3	2340	74.78

TABLE 39
Percent Unchanged With Original Unreplaced Image

Search Radius (Pixels)	Percent Unchanged
1	96.40
2	92.53
3	89.27

TABLE 40
Probability of Non-Omission for Replaced Categories

(Probability that an Ungeneralized Pixel will be Correctly Represented in the Generalized Image)

	Search Radius (Pixels)		
	1	2	3
CATEGORY			
NF	98.14	95.84	94.00
BF	86.41	72.00	60.97
MF	87.04	72.84	61.65
SL	97.69	95.41	93.44
DS	96.87	93.86	91.15
SV	80.62	60.85	44.90
NV	94.63	88.48	83.07
DW	93.17	84.53	77.13
SW	98.32	96.47	95.30
IS	98.22	96.35	94.77
SD	95.03	89.99	85.35

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 41
Probability of Non-Comission for Replaced Categories

(Probability that a Generalized Pixel will Correctly Represent the Corresponding Ungeneralized Pixel)

	Search Radius (Pixels)		
	1	2	3
CATEGORY			
NF	96.41	93.23	90.58
BF	93.25	85.83	78.88
MF	92.19	84.32	76.93
SL	97.27	93.89	91.04
DS	96.08	91.66	87.89
SV	92.92	84.19	75.64
NV	95.53	90.31	86.00
DW	99.13	96.27	95.05
SW	97.38	92.77	89.11
IS	98.47	96.83	95.43
SD	96.41	90.96	86.18

LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

TABLE 42
Cohen's KHAT Statistic for Replaced Images

(Measure of Overall Agreement Between Generalized and Ungeneralized Images)

Search Radius (Pixels)	KHAT (Percent)
1	95.61
2	90.87
3	86.87

TABLE 43
Fragmentation Index (F) for Replaced Images

Search Radius (Pixels)	F
1	0.58342
2	0.35909
3	0.25212

Monmonier's Fragmentation Index for initial, ungeneralized image = 0.1374

DISCUSSION

Area Filtering Generalization

Effects on Total Generalized Image

As the minimum threshold area was increased, KHAT, percent unchanged, and the number of homogeneous regions decreased. The KHAT agreement values and the percent unchanged values had identical patterns, although the KHAT agreement values tended to be slightly lower as expected (Figure 6). The sharpest decline in KHAT values, percent unchanged values, and number of homogeneous regions occurred between the 2 and 68 pixel filters (Figures 6, 7, 8). As mentioned before, these results will vary by the area selected for analysis.

Because the sharpest decline in region numbers after filtering (Figure 8) occurred between the 2 and 68 pixel filters, it may be concluded that the majority of homogeneous regions in the Valdez test area fell within this size range. The software used for analysis in this study was unable to list the region, the corresponding pixel size, and the associated land cover category for all regions under analysis. If this information had been available, further tests may have been conducted which

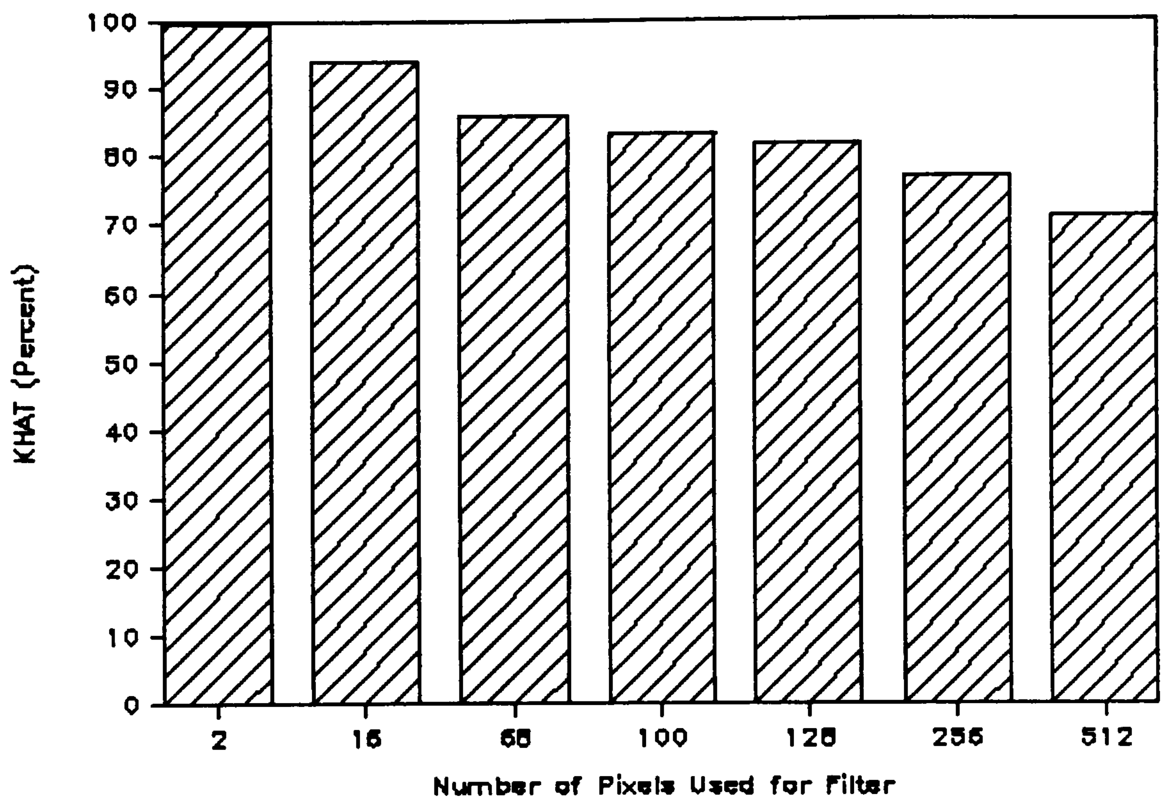


FIGURE 6
Cohen's KHAT Statistics
for Area Filtered Images

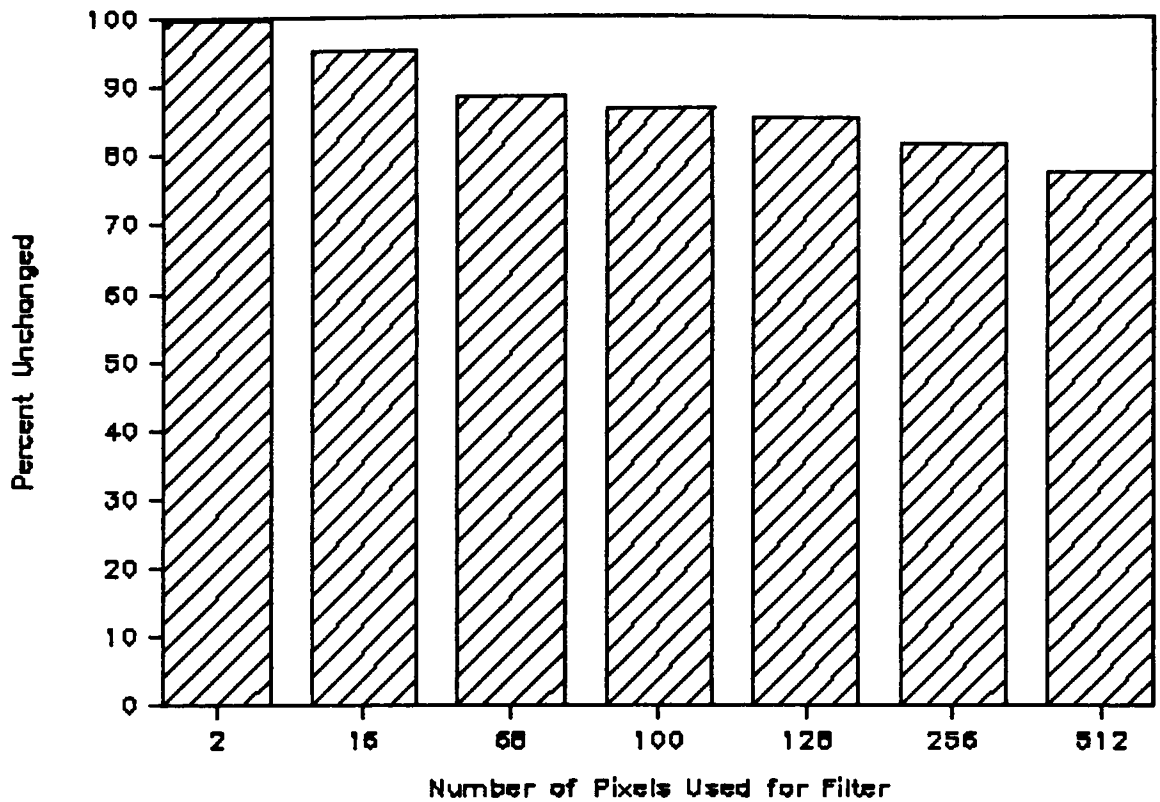


FIGURE 7
Percent Unchanged With Original Unfiltered Image

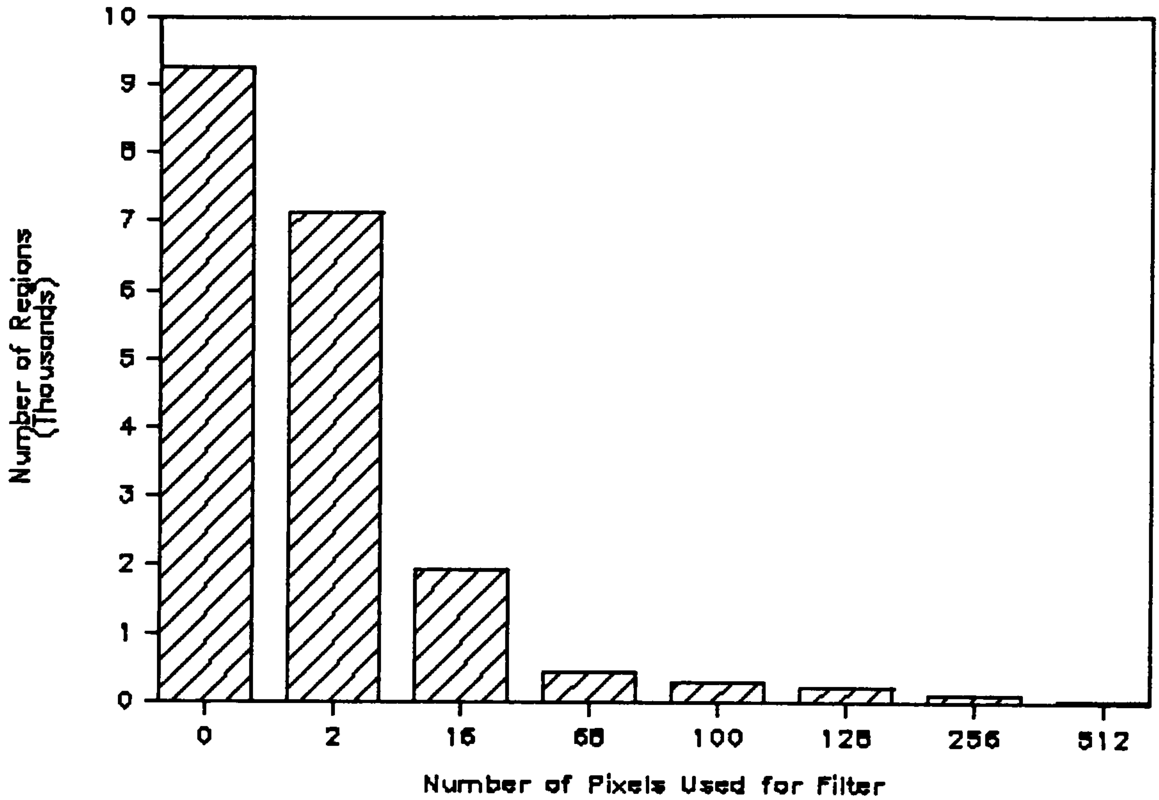


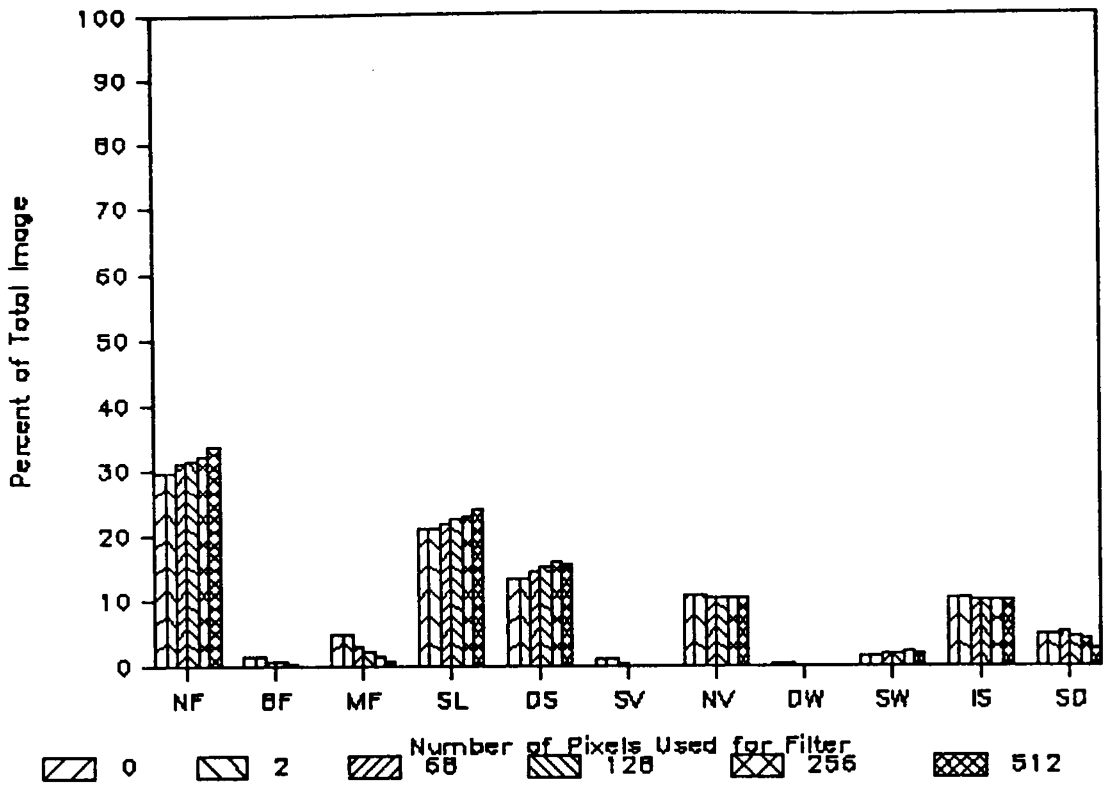
FIGURE 8
Number of Regions after Area Filtering

utilized the maximum and minimum region sizes for each individual land cover category.

Although only 59 regions remained after area filtering with a 512 pixel minimum threshold size (Table 10), the percent unchanged and KHAT agreement values (77.48 and 71.29 respectively) were both very high (Table 11, 14). This may have resulted from large areas of the image being occupied by very homogeneous land cover categories, such as Needleleaf Forest and Shrublands (Figure 9).

While using area filtering for data reduction, the number of homogeneous regions tended to level off after the 68 pixel threshold size (Figure 8). This trend did not occur with the percent unchanged or KHAT values. Thus, using minimum threshold sizes larger than 68 pixels does little to decrease data volume (number of regions) but results in decreasing agreement values.

The fragmentation index steadily decreased with the sharpest decline between the 2 and 68 pixel filters. Monmonier's fragmentation index for the ungeneralized image was very small, (Table 15) which indicated that the raw land use/land cover data was relatively simple compared to the most complex state possible where each pixel would be a region, similar to the appearance of a checkerboard. The probability of such a complex map was extremely small (1.32 x 10 -7). Therefore, normalizing these values by dividing



LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DW = Deep/Clear Water
SW = Shallow/Turbid Water

DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
IS = Ice and Snow
SD = Shadow

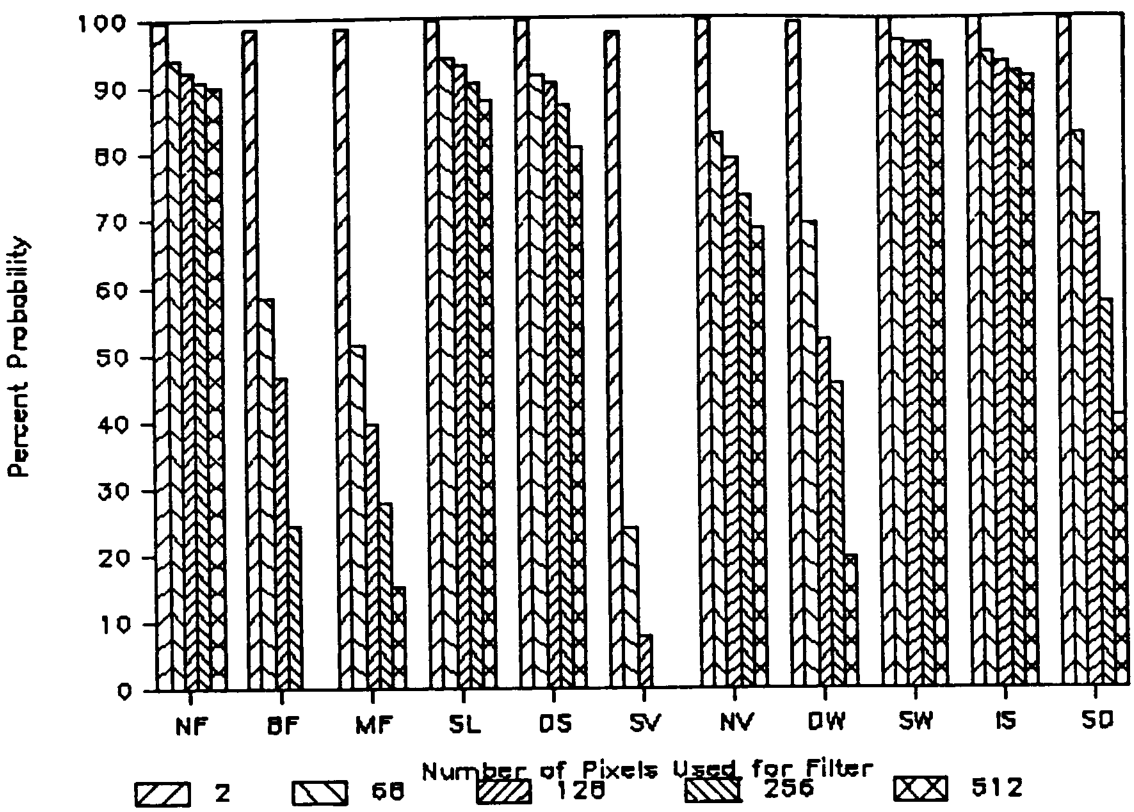
FIGURE 9
Percent of Total Image by Category
After Area Filtering

all fragmentation indices by the ungeneralized image's value resulted in a better comparison of relative image complexity. A value of 1 indicates no generalization (complexity) while a value of 0 indicates high generalization (simplicity). The fragmentation index decreased with increasing threshold area, with the sharpest decline occurring between the 2 and 16 pixel threshold sizes (Table 15).

Effects on Individual Generalized Categories

Area filtering increased the percent of total image occupied by Needleleaf Forest, Shrublands, and Dwarf Shrublands and decreased the percent of Broadleaf Forest, Mixed Forest, Sparse Vegetation, Shadow and Deep Water categories (Figure 9). The declining categories tended to have high omission errors while the increasing categories had low omission errors (Figure 10). Thus, as categories with small regions were eliminated by reassignment, other categories expanded.

The probability of Broadleaf Forest, Mixed Forest, Sparse Vegetation, Deep Water and Shadow pixel in the original data was low for correct representation in the filtered, generalized image. Omission errors increased with increased minimum threshold sizes (Figure 10). However the low error, high probability categories such as



LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

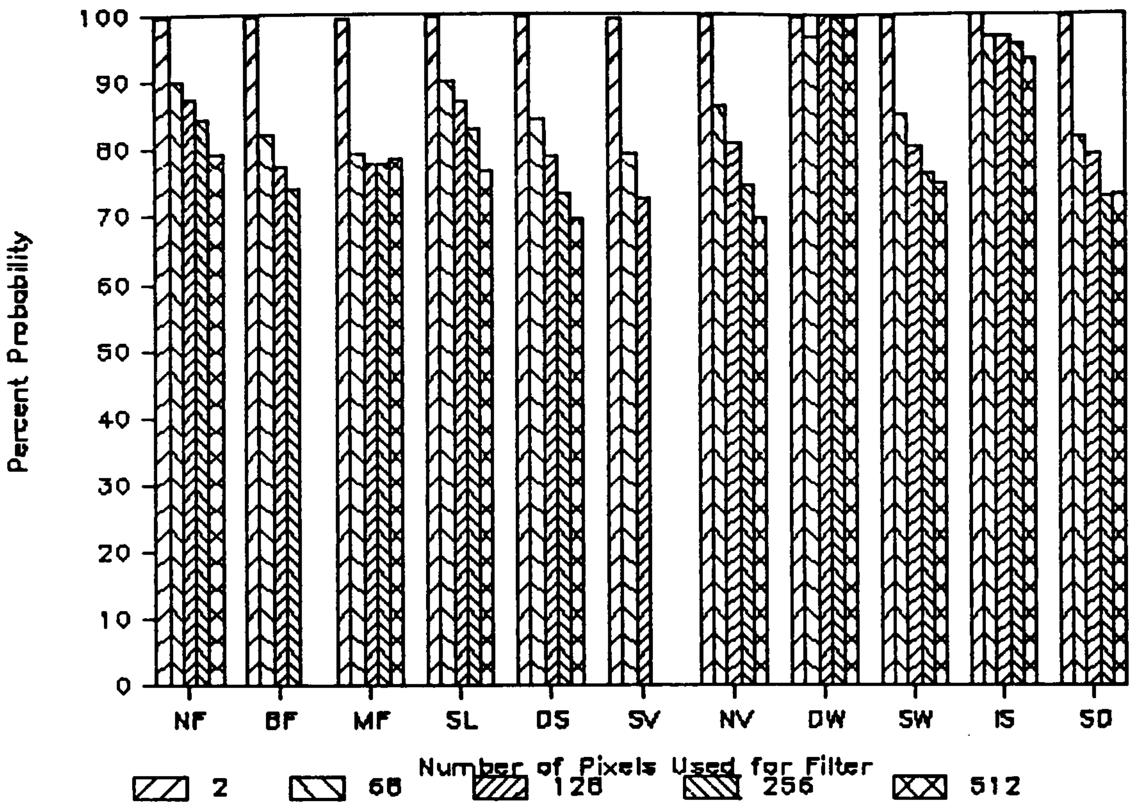
FIGURE 10
Probability of Non-Omission
for Area Filtered Categories

Needleleaf Forest, Shrublands, Dwarf Shrublands, Non-Vegetated and Ice and Snow tended to be less effected by increasing minimum threshold sizes, even at the 512 pixel minimum threshold size (Figure 10).

The probability of non-comission for the filtered categories represented the probability that a filtered pixel will correctly represent the corresponding unfiltered pixel's category. Needleleaf Forest, Deep water and Ice and Snow and to some extent Shrublands, Dwarf Shrublands and Shallow water tended to have low comission errors while broadleaf Forest and Sparse Vegetation tended to have high errors (Figure 11).

The Deep Water and Mixed Forest had high omission errors, low comission errors, and a decreasing percent of total image (Figures 9, 10, 11). It was concluded that several of the Deep Water pixels formed small regions which were eliminated as the minimum threshold size was increased. However, those regions which were labeled as Deep Water in the filtered images were very homogeneous and contained a strong similarity to the original data.

The Needleleaf Forest, Shrublands, Dwarf Shrublands, and Ice and Snow categories had low omission and comission errors and all tended to increase in percent of total image (Figures 9, 10, 11). The Broadleaf Forest and Sparse Vegetation categories had high omission and comission



LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

FIGURE 11
Probability of Non-Comission
for Area Filtered Categories

errors and declining percents of total image (Figures 9, 10, 11). Thus, it was concluded that these categories were composed of small scattered regions which tended to be omitted or to be poorly represented in the generalized image.

Unidirectional Shadow Filtering

The most obvious effect was the decline in percent of shadow from 5.04 in the original data to 2.03 in the 512 pixel filtered image (Table 16). The overall percent unchanged and KHAT values (Tables 18 and 21, respectively) were only slightly higher with the imposed unidirectional shadow filter as compared with the initial filtering results. Omission errors in the Sparse Vegetation and Non-Vegetated categories decreased slightly while the errors of comission for Shadow were nonexistent because of the nature of the pairwise weight table (Table 20).

Although the shadow category was reduced as much as 3 percent by imposing a unidirectional shadow filter, the number of regions was slightly higher compared to the absence of a shadow filter on area filtered images. This resulted from the small size and relative homogeneity of the shadow areas on the image. The purpose of imposing a unidirectional shadow filter on the land cover data was to decrease a noninformational category. The 3 percent reduction was offset by a one percent gain in the number of homogeneous regions. Thus, the usefulness of the filter in combination with an area filtering generalization strategy, ultimately depends upon the use of the data.

Category Aggregation Generalization

Effects on Total Generalized Image

When level II precision categories were aggregated to level I precision, the number of homogeneous regions was reduced to almost half of the original 9278 regions (Figure 12). Strict category aggregation does not inflict the categories with any omission or comission error, but simply redefines the categories to a more general level. Category aggregation was facilitated by the hierarchical structure of the USGS Interim land cover data.

Percent unchanged, KHAT and number of homogeneous regions all decreased with increasing minimum threshold sizes (Figures 12, 13, 14). The percent unchanged and KHAT agreement values were both much higher compared to the unaggregated filtered data (Figures 12, 13) at the same minimum threshold levels. The percent reduction after area filtering the aggregated data was similar to the initial

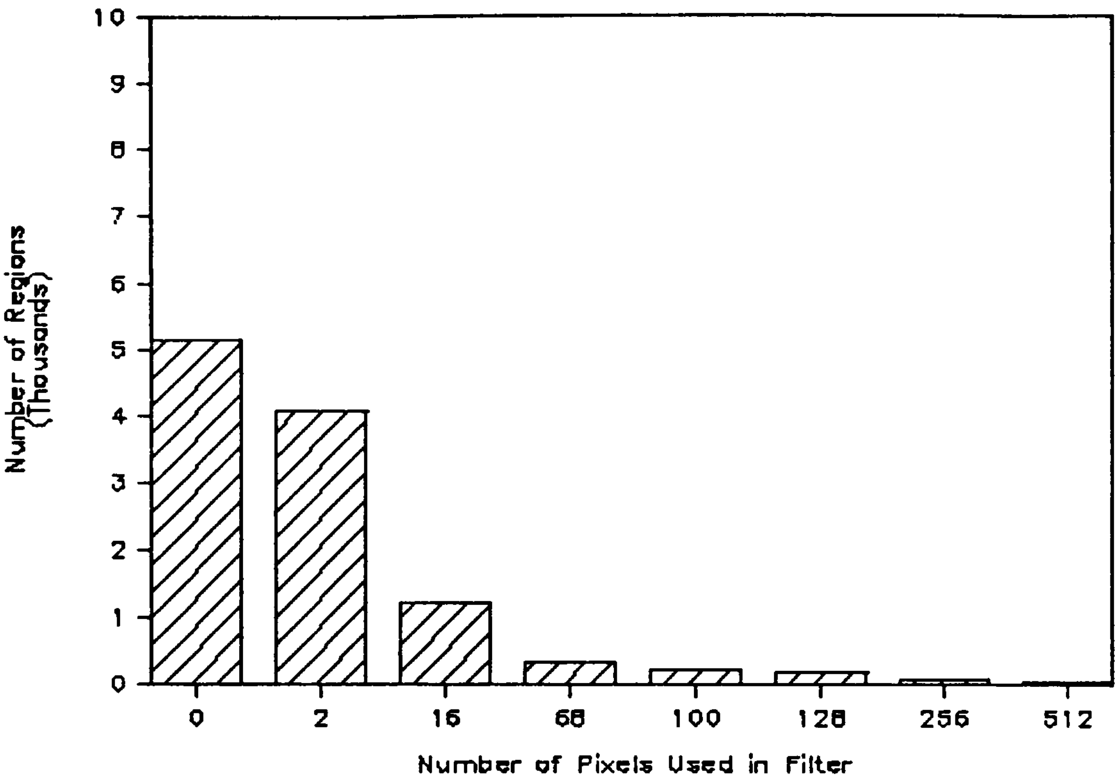


FIGURE 12
Number of Regions After Category
Aggregation and Area Filtering

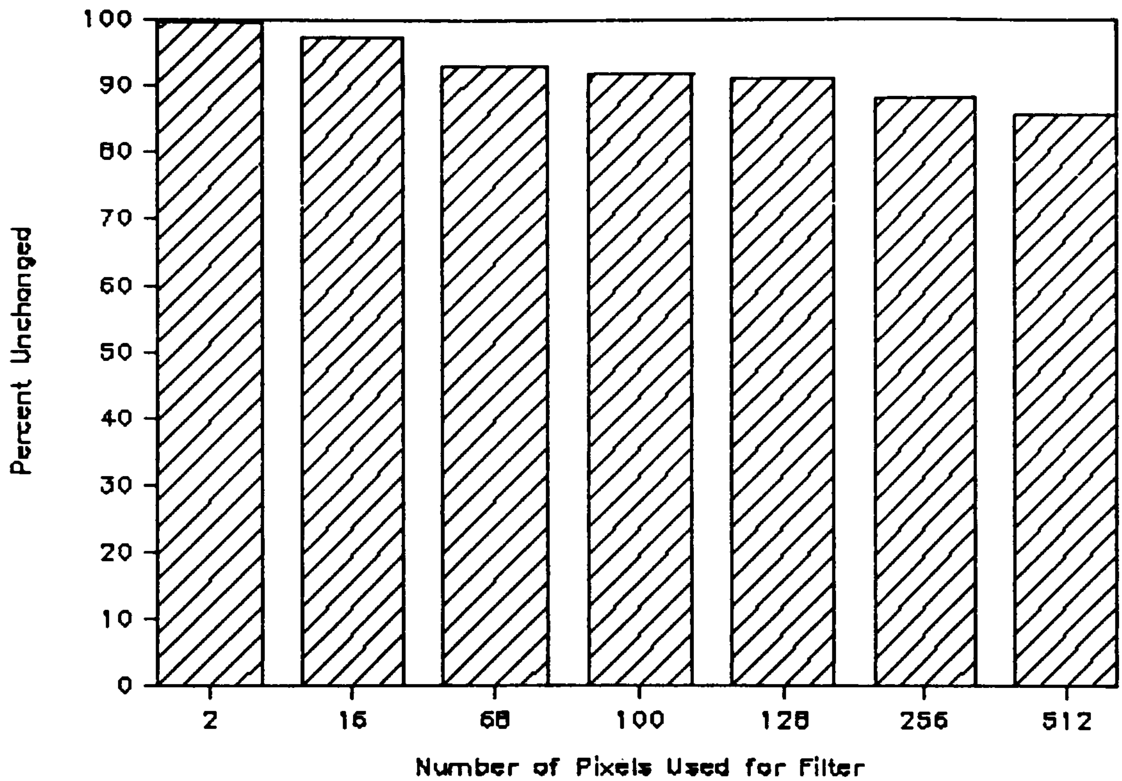


FIGURE 13
Percent Unchanged
with Original Aggregated Image

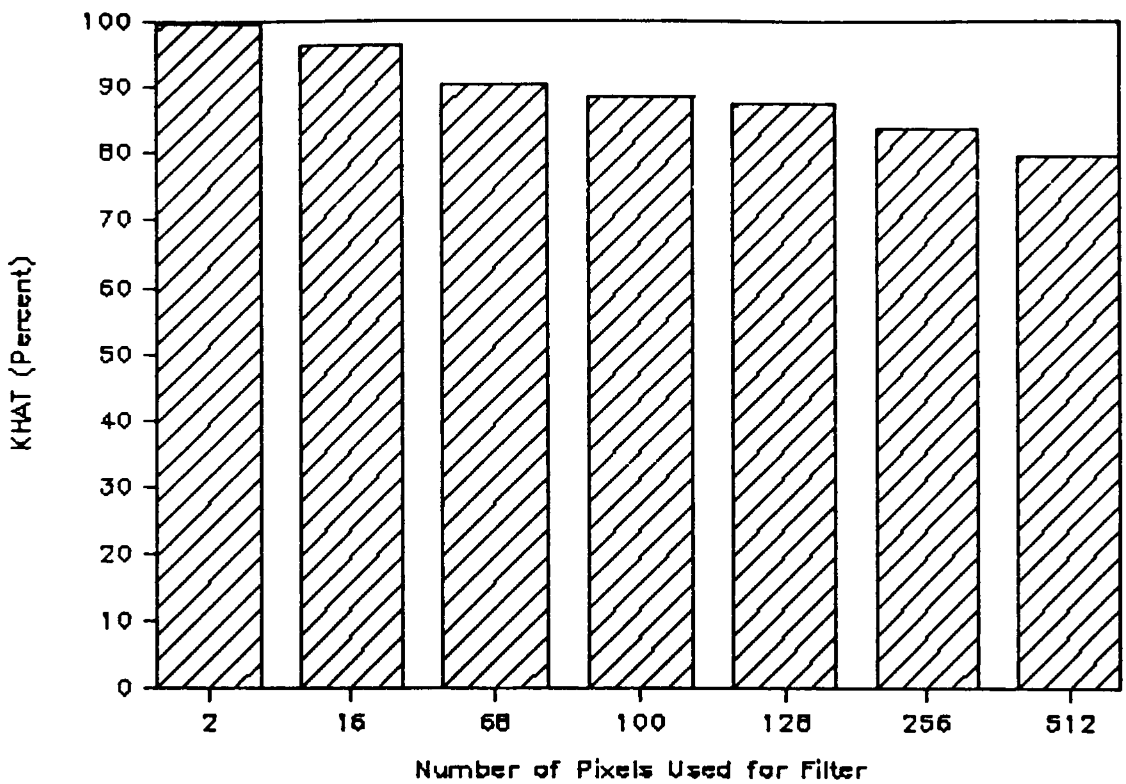


FIGURE 14
Cohen's KHAT Statistic for Category
Aggregated, Area Filtered Images

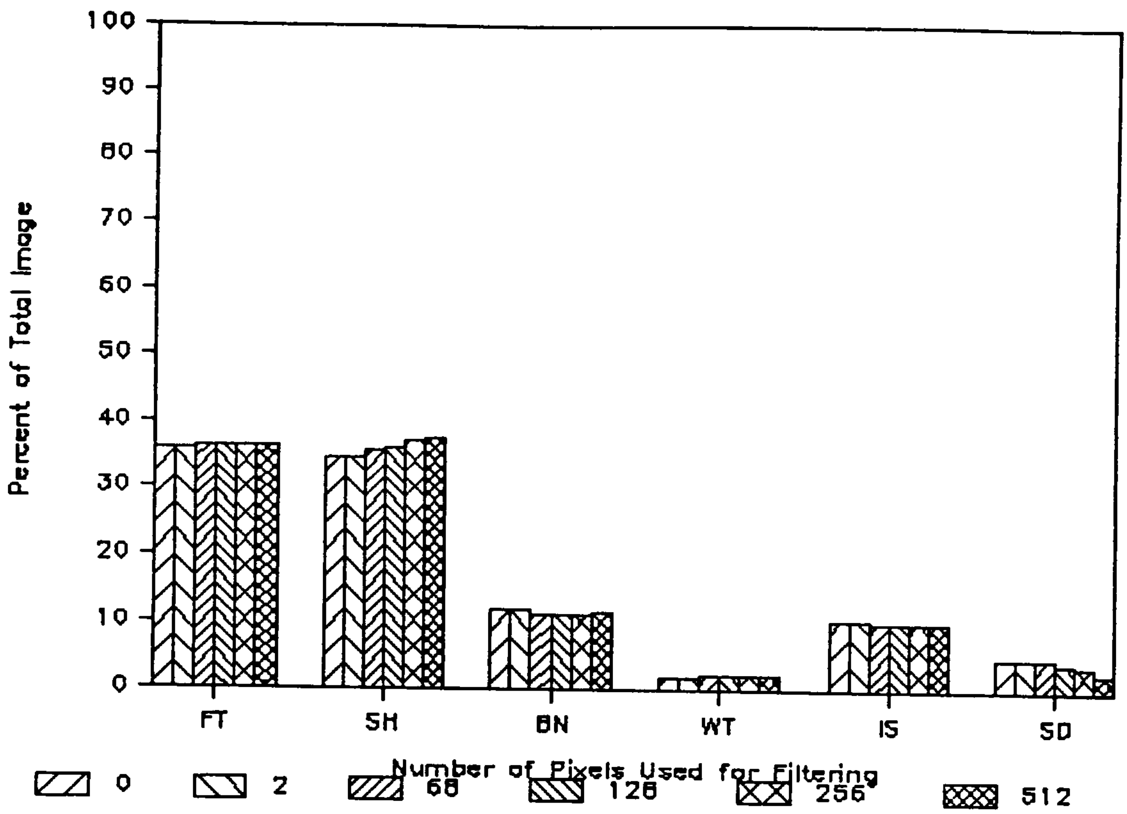
filtering results. The steepest reduction in regions occurred between the 2 and 68 pixel filters (Figure 12) and tended to level off with larger sized filters.

The fragmentation indices for aggregated, filtered data were much lower compared to the unaggregated, filtered data. Thus, the aggregated data tended to have simpler spatial patterns. The steepest decline in the fragmentation index occurred between the 2 and 16 pixel minimum threshold area. The fragmentation values leveled off revealing little increase of map simplicity with minimum area thresholds above the 68 pixel size (Table 29).

Effects on Individual Generalized Categories

There was little change in percent of total image for the aggregated level I categories with increasing minimum threshold sizes. The Shrub category increased in total percent while the Shadow category decreased (Figure 15). This resulted from the increased size of the aggregated categories.

Omission errors overall were very low for the aggregated categories, thus they had a relatively high probability of being correctly reassigned in the filtering process. The Shadow category encountered the highest amount of omission error, resulting from its small, unaggregated



LEGEND

FT = Forest
SH = Shrub
BN = Barren
WT = Water
IS = Ice and Snow
SD = Shadow

FIGURE 15
Percent of Total Image by Category
After Aggregation and Area Filtering

unaggregated regions. As the minimum threshold size was increased, omission errors increased (Figure 16). However, the Forest, Shrub, and Ice and Snow categories maintained a high probability with increasing threshold size. This may be attributed to the increased size occupied by these categories (Figure 16).

Errors of comission also tended to be low for the aggregated, filtered categories compared to the unaggregated filtered data. As the threshold size increased, so did the comission error. The Forest and Ice and Snow categories maintained high probabilities, even at the 512 pixel threshold size (Figure 17). This may be attributed to the relative homogeneity of these large regions.

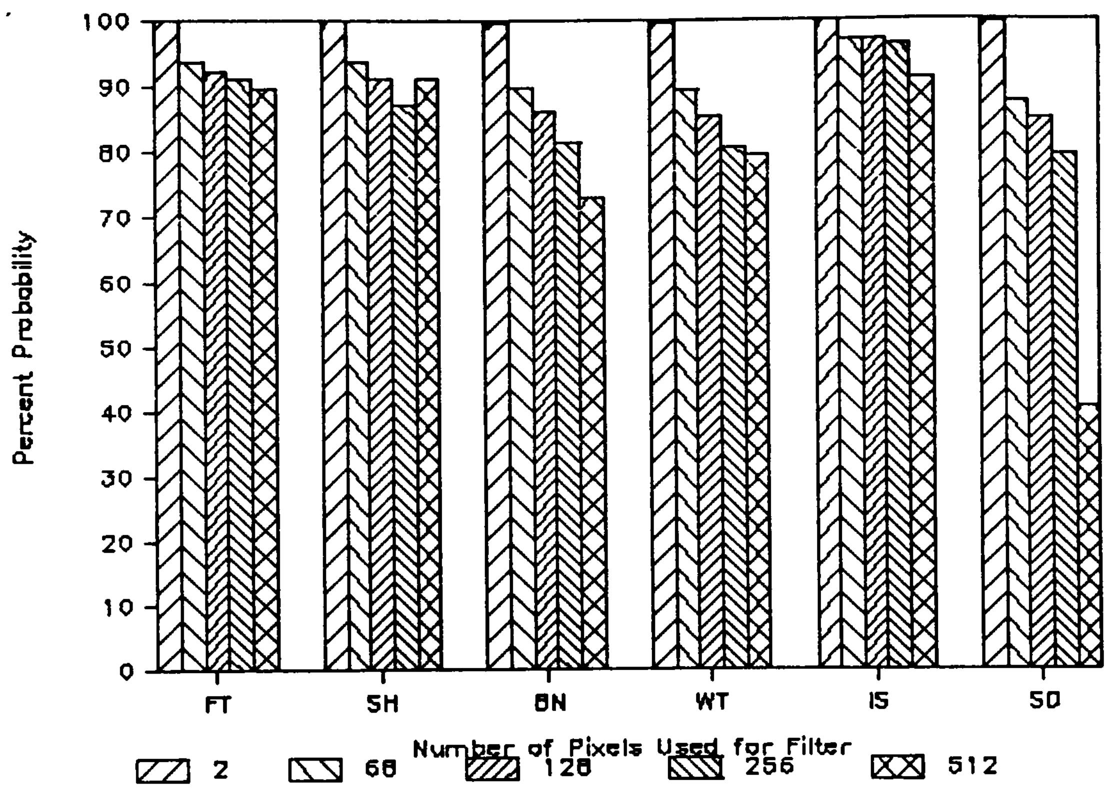
There was little difference between the omission and comission errors for the categories except for the Shadow category. Shadow achieved a high omission error but a much lower comission error. Therefore, the few regions which were reclassified to Shadow in the generalization process maintained at least a 78 percent probability or greater of correctly representing a Shadow pixel in the original aggregated data.

LEGEND

FT = Forest
SH = Shrubs
BN = Barren

WT = Water
IS = Ice and Snowater
SD = Shadow

FIGURE 16
Probability of Non-Omission for
Aggregated, Area Filtered Categories



LEGEND

FT = Forest
SH = Shrub
BN = Barren

WT = Water
IS = Ice and Snow
SD = Shadow

FIGURE 17
Probability of Non-Comission for
Aggregated, Area Filtered Categories

Resampling Generalization

Effects on Total Generalized Image

Percent unchanged, KHAT, and the number of homogeneous regions decreased with an increase in imposed grid size (Figures 18, 19, 20). The KHAT values, as before, were slightly less than the percent unchanged values (Figures 18, 19). While the decline in homogeneous regions was not as great as in the area filtered data, the KHAT and percent unchanged values were much lower compared to the area filtered images. Thus, resampling results in a low reduction in regions as well as low agreement values with the original data compared to the area filtered images.

Effects on Individual Generalized Categories

There was almost no change in percent of total image by category with an increase in grid size (Figure 21). This resulted from the small imposed grid sizes selected for resampling. Because of the large testing area selected for study, the occurrence of small regions or lone pixels occupying the center of a grid cell were compensated with those regions being occupied by spatially extensive categories. This also resulted in the similarity between the omission and comission errors (Figures 22, 23).

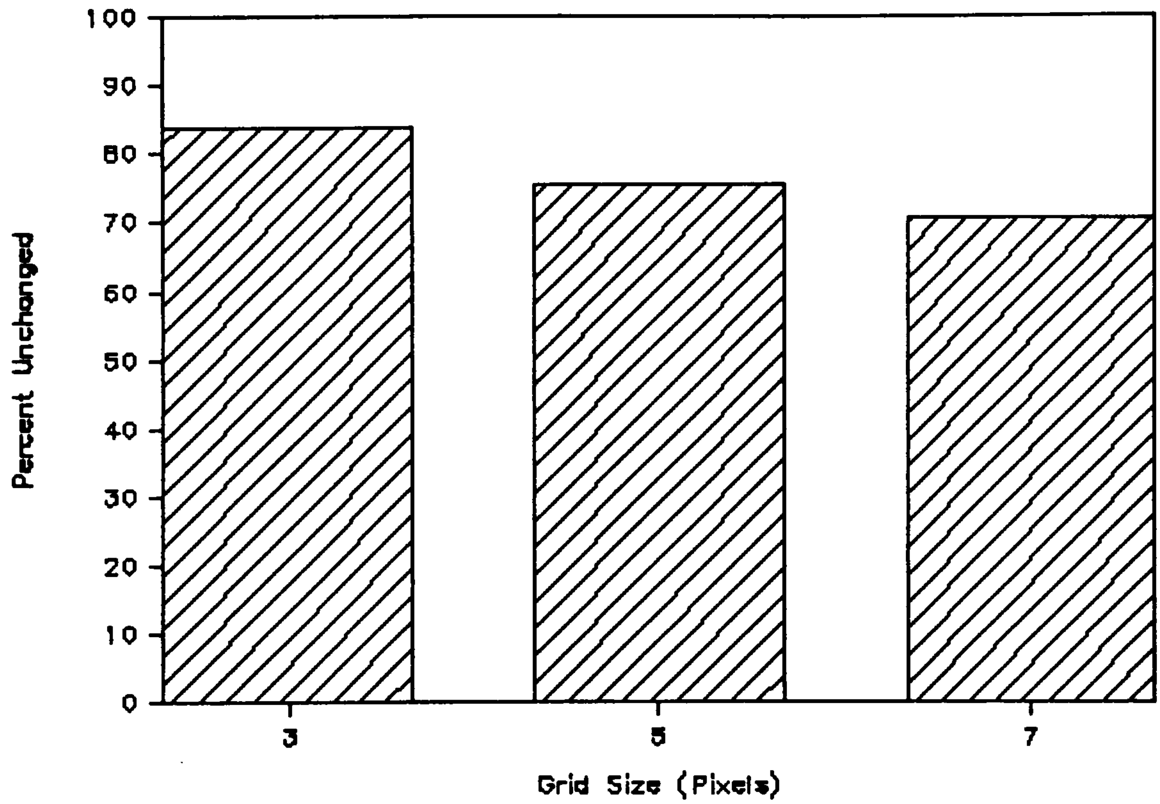


FIGURE 18
Percent Unchanged with
Original Unresampled Image

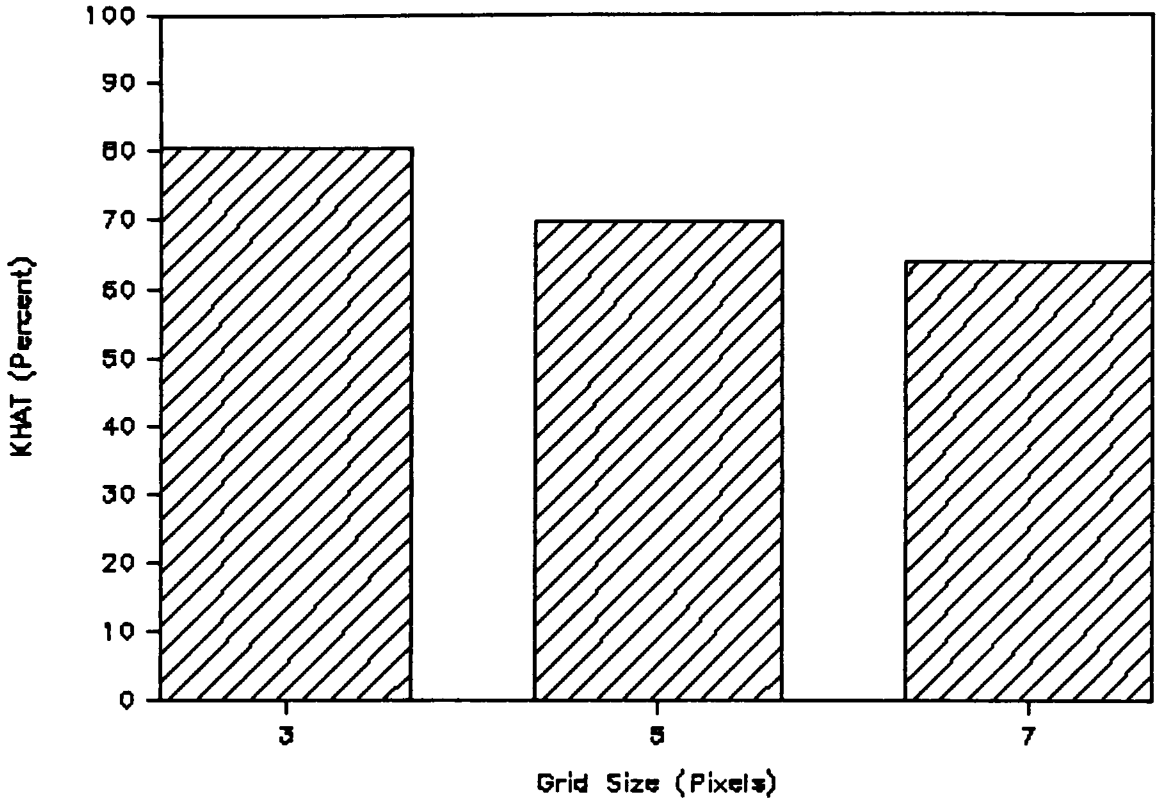


FIGURE 19
Cohen's KHAT Statistic
for Resampled Images

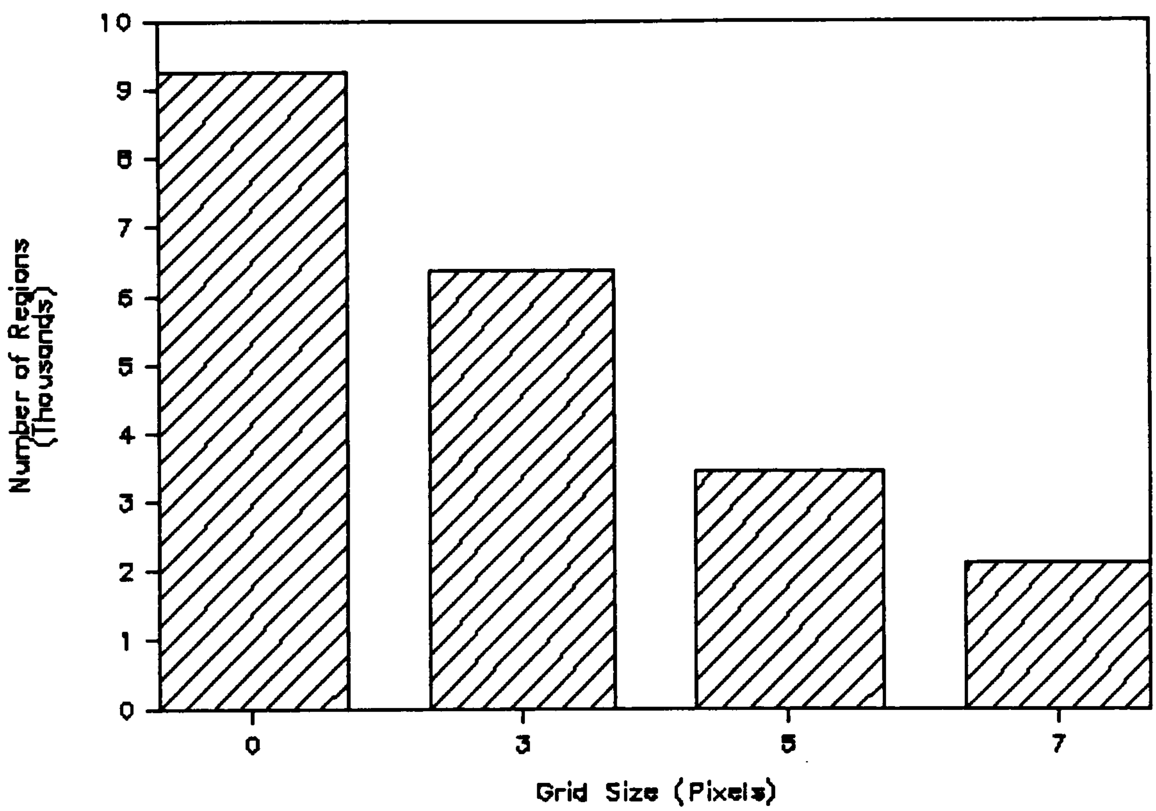
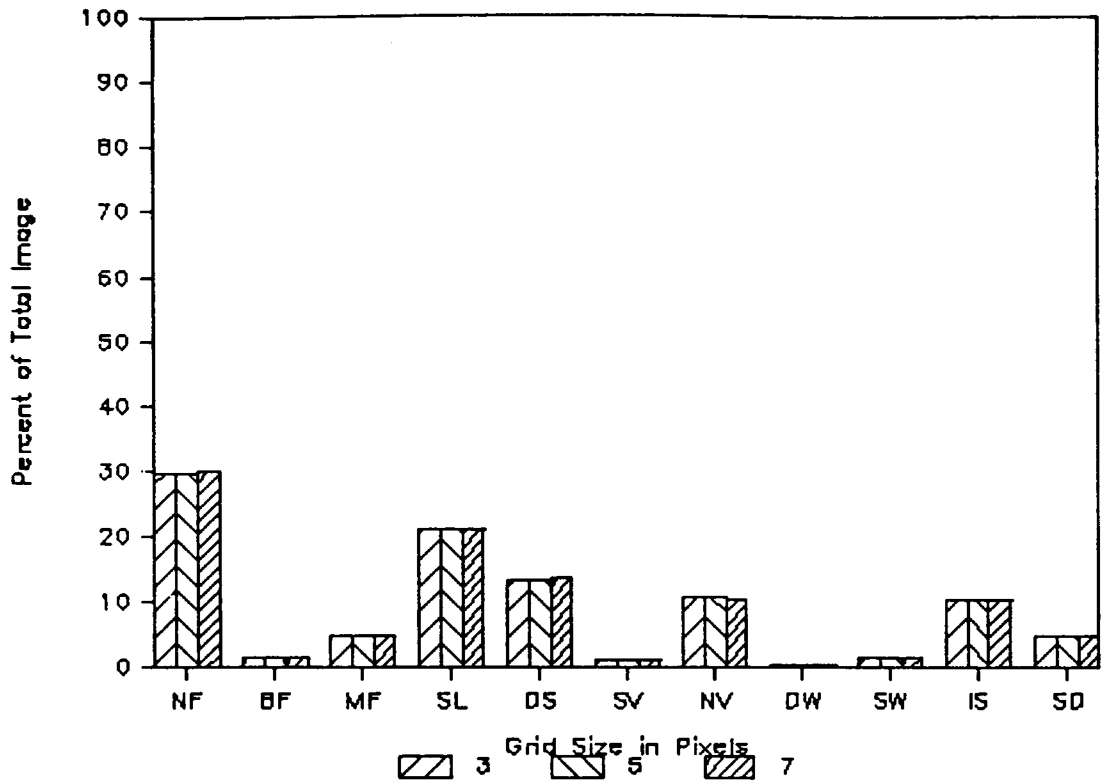


FIGURE 20
Number of Regions
After Resampling

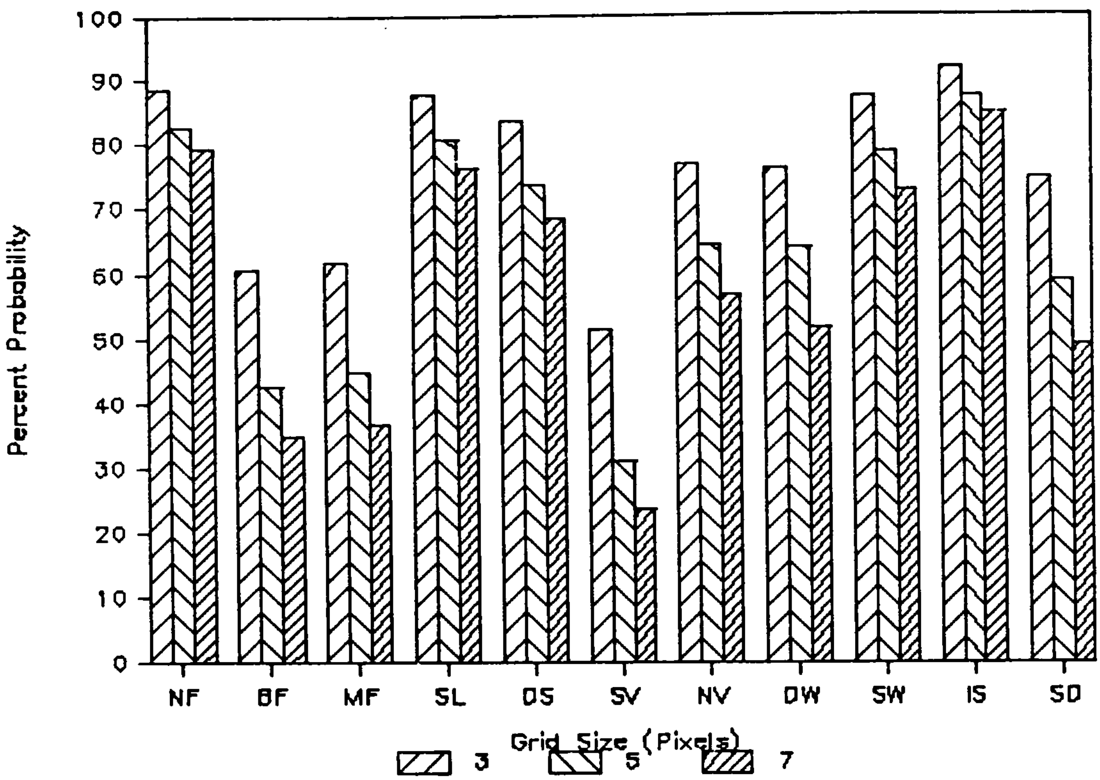


LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands

DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
IS = Ice and Snow
SD = Shadow

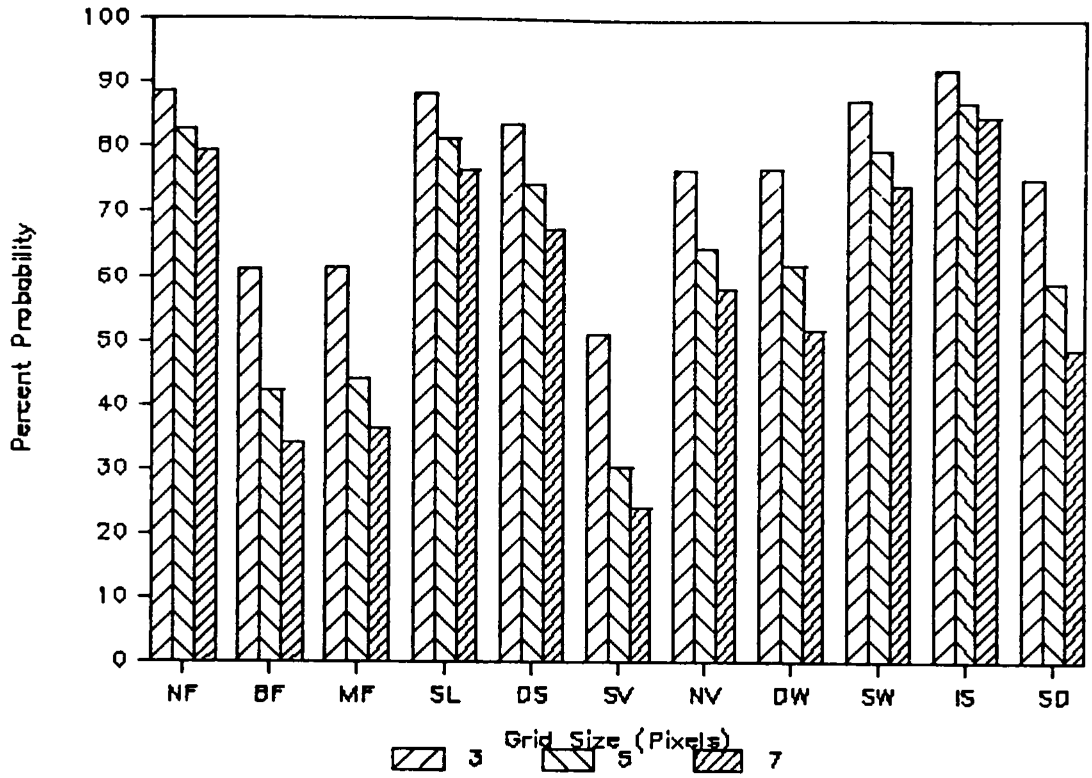
FIGURE 21
Percent of Total Image by Category
After Resampling



LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

FIGURE 22
Probability of Non-Omission
for Resampled Categories



LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

FIGURE 23
Probability of Non-Comission
for Resampled Categories

The omission and comission errors are largely based on the percent of total image a category occupies. Needleleaf Forest occupied almost 30 percent of the total image and had relatively low omission and comission errors while Broadleaf Forest which occupied less than two percent of the total image had relatively high omission and comission errors (Figures 22, 23).

Ice and Snow, Needleleaf Forest, Shrublands and Dwarf Shrublands all tended to have lower omission and comission errors compared with the Broadleaf Forest, Sparse Vegetation and Mixed Forest categories (Figures 22, 23). In general, omission and comission errors were much greater for all categories compared with the errors resulting from area filtering. Thus, resampling produces images with a larger number of regions, and higher category errors compared to area filtering.

Modal Search and Replace Generalization

Effects on Total Generalized Image

An increase in the search radius produced a decrease in percent unchanged, KHAT, and the number of regions (Figures 24, 25, 26). As before, percent unchanged values were greater compared to KHAT agreement values. KHAT, percent

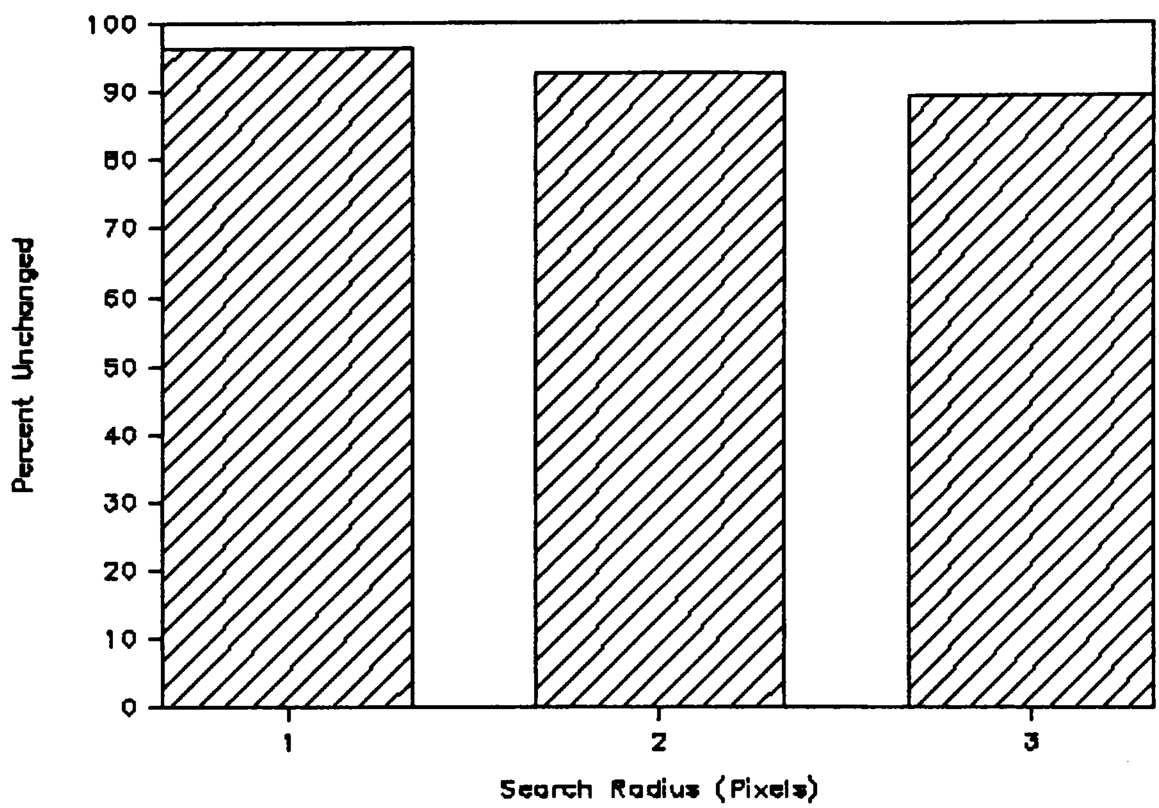


FIGURE 24
Percent Unchanged
with Original Unreplaced Images

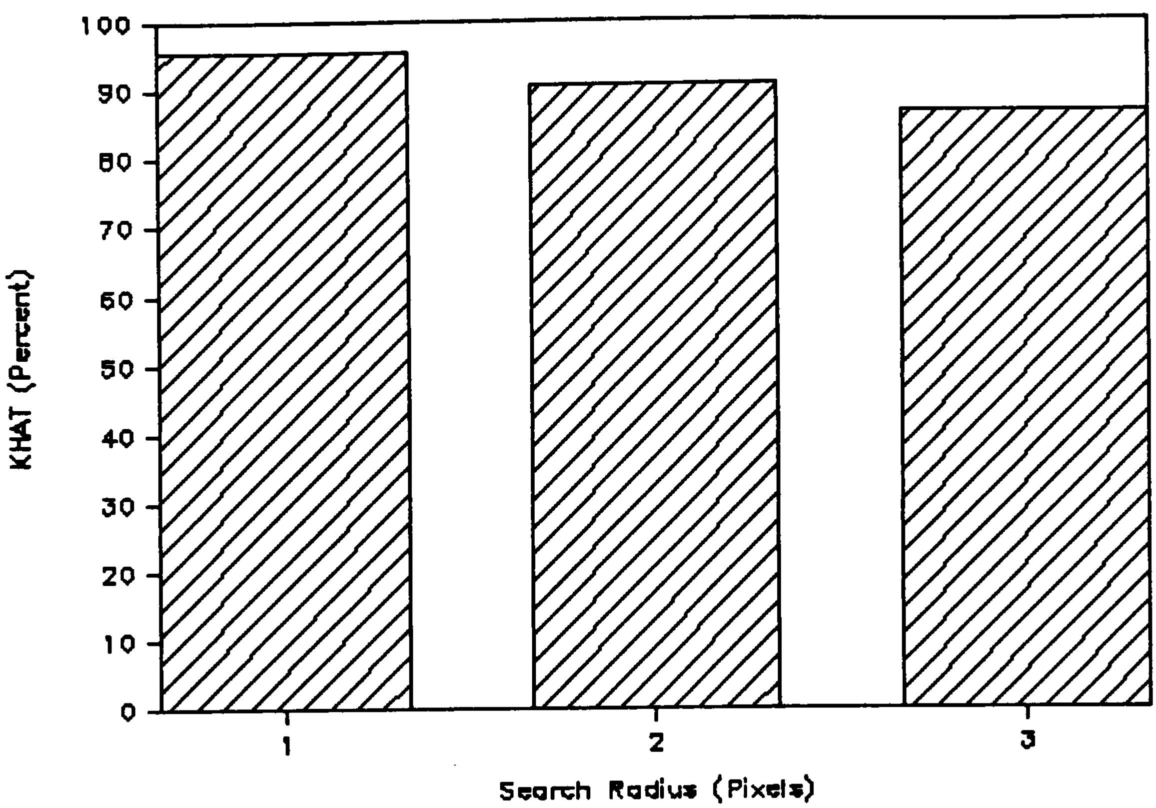


Figure 25
Cohen's KHAT Statistic
for Replaced Images

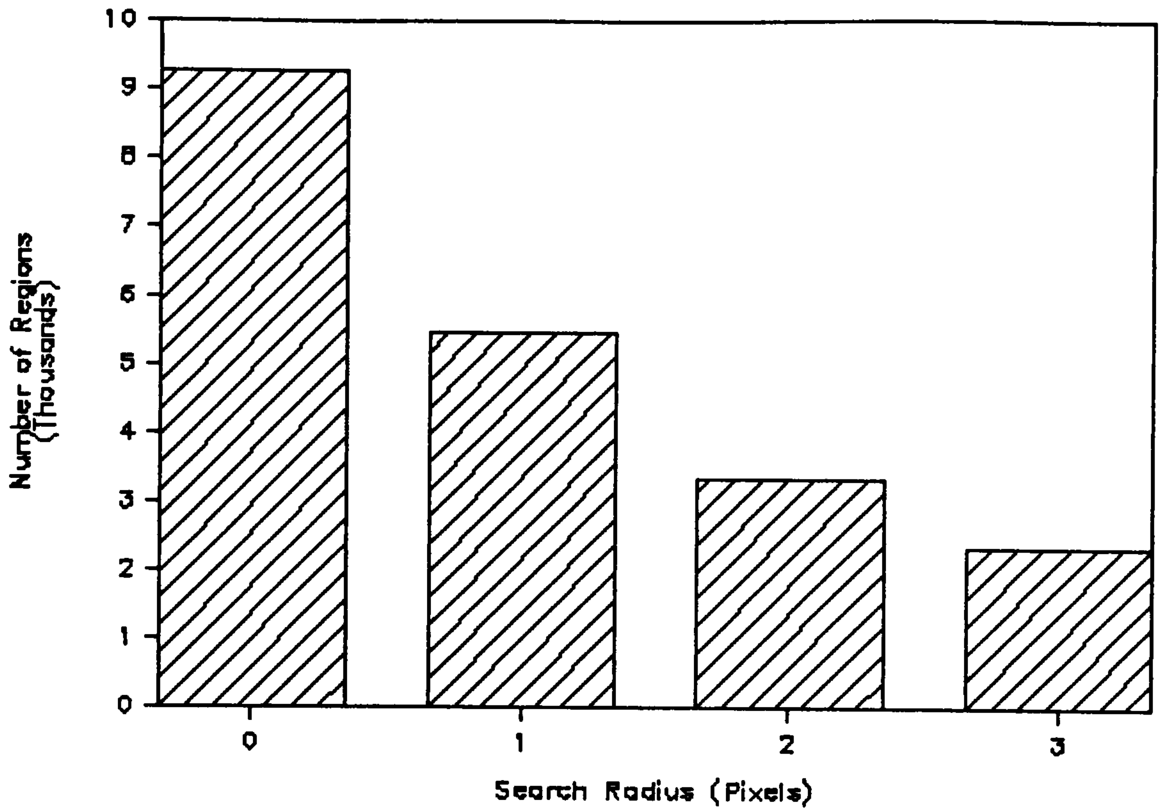
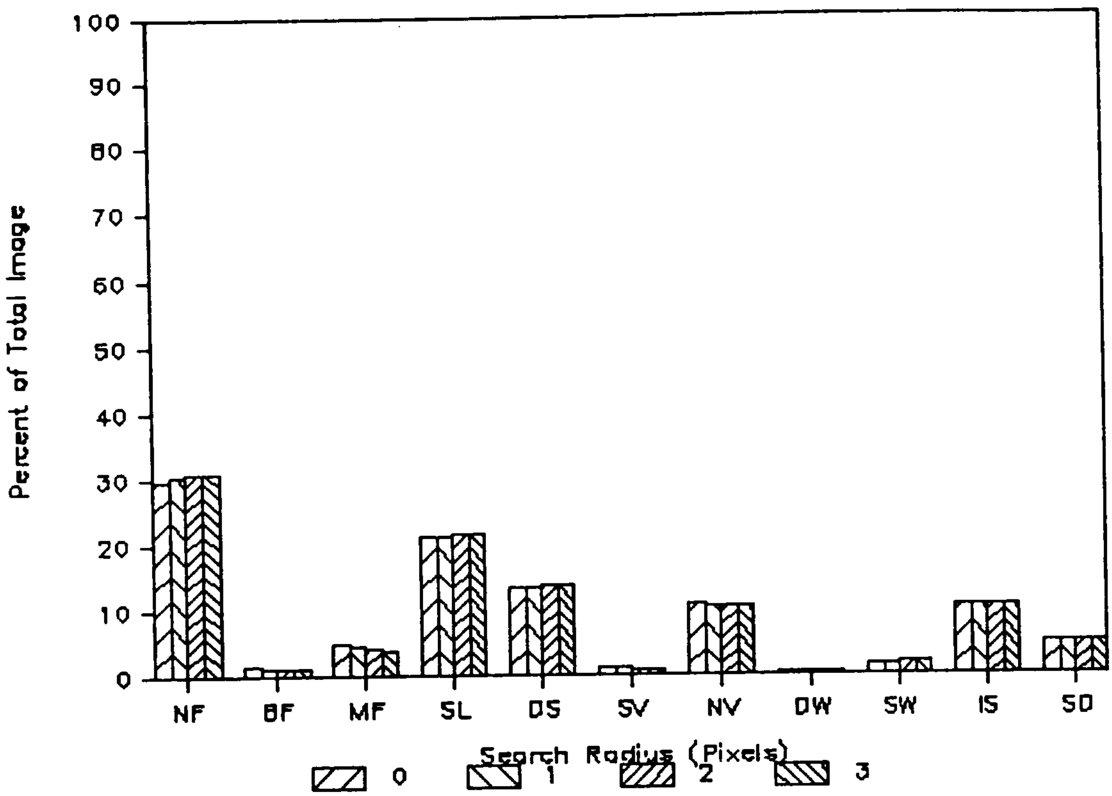


FIGURE 26
Number of Regions After Replacement

unchanged, and region reduction produced by the modal search and replace technique were higher compared to those values produced by the systematic resampling strategy (Figures 24, 25, 26). Although the percent reduction in homogeneous regions with a search radius of 2 and 3 were only slightly higher compared to the results obtained by resampling with a grid size of 5 and 7 pixels, the KHAT and percent unchanged were much higher in the modal search and replace images. The percent reduction in regions using a search radius of 3 pixels was similar to the percent reduction in regions produced with an area filter of 16 pixels. However, the 16 pixel area filter produced a higher KHAT and percent unchanged for the image compared to the modal search and replace technique.

Effects on Generalized Categories

The changes in percent of total image occupied by a category changed slightly with an increasing search radius (Figure 27). Needleleaf Forest, Shrublands, and Dwarf Shrublands categories increased in total image occupation, while Broadleaf Forest, Mixed Forest and Sparse Vegetation categories decreased (Figure 27).



LEGEND

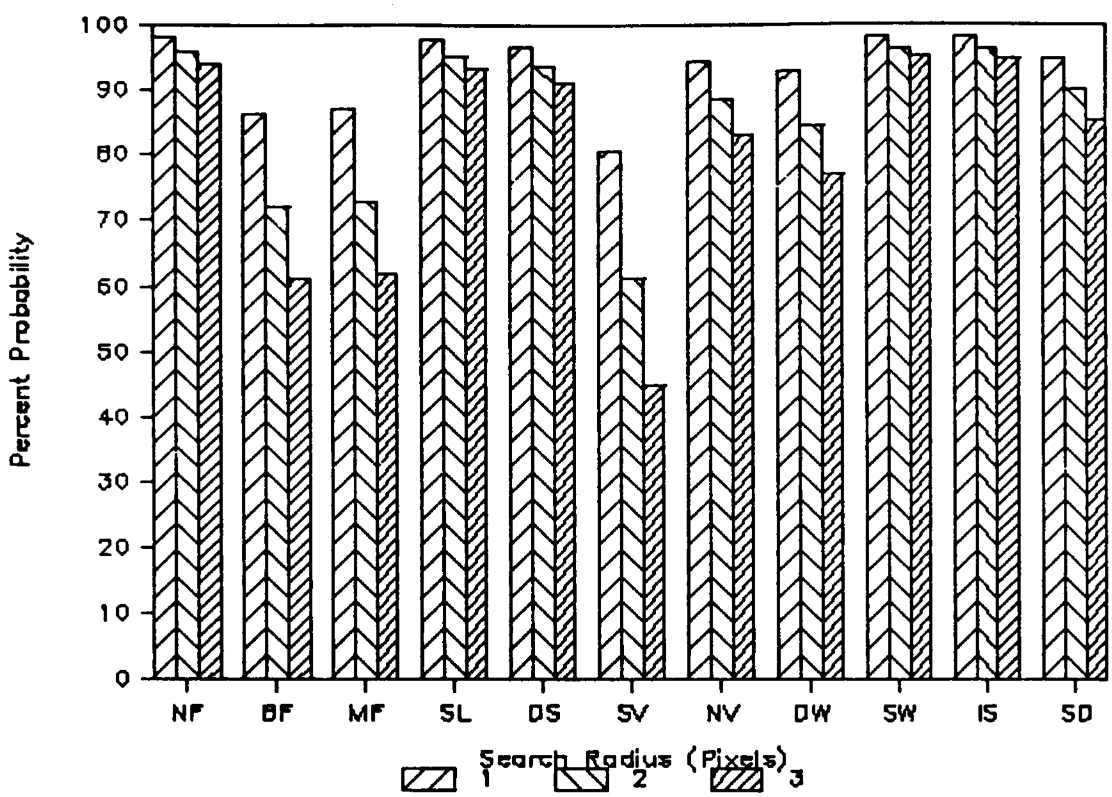
NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
SD = Shadow

DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
IS = Ice and Snow

FIGURE 27
Percent of Total Image by Category
After Replacement

In general, omission and comission errors were low compared to those achieved by resampling (Figures 28, 29). This resulted more from the large spatial extent of a category relative to the search radius, than from the category's homogeneity.

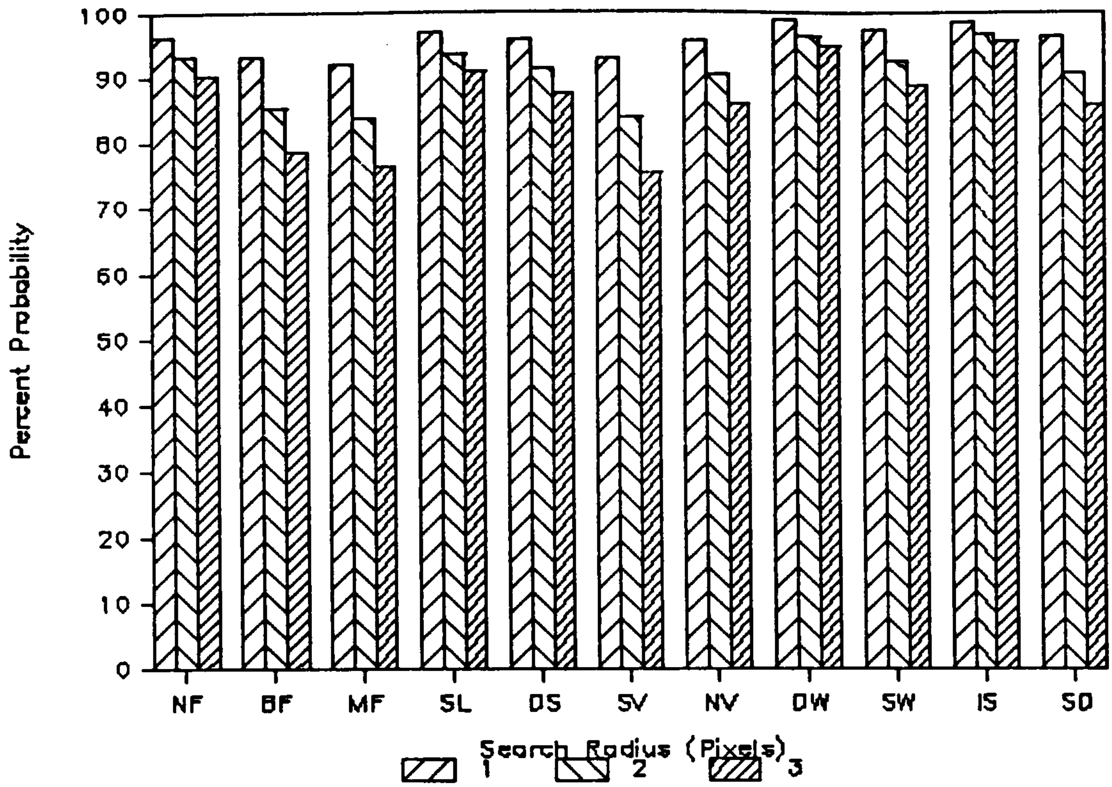
Broadleaf Forest, Mixed Forest, and Sparse Vegetation categories had the lowest probabilities of being correctly reassigned in the generalized image, while shrublands, Dwarf Shrublands, Needleleaf Forest, Ice and Snow and Shallow Water categories had higher probabilities (Figure 28). High comission errors were associated with those categories which contained a small spatial extent in the total image (Figure 29).



LEGEND

NF = Needleleaf Forest
BF = Broadleaf Forest
MF = Mixed Forest
SL = Shrublands
DS = Dwarf Shrublands
SV = Sparse Vegetation
NV = Non-Vegetated
DW = Deep/Clear Water
SW = Shallow/Turbid Water
IS = Ice and Snow
SD = Shadow

FIGURE 28
Probability of Non-Omission
for Replaced Categories



LEGEND

NF = Needleleaf Forest	NV = Non-Vegetated
BF = Broadleaf Forest	DW = Deep/Clear Water
MF = Mixed Forest	SW = Shallow/Turbid Water
SL = Shrublands	IS = Ice and Snow
DS = Dwarf Shrublands	SD = Shadow
SV = Sparse Vegetation	

FIGURE 29
Probability of Non-Comission
for Replaced Categories

CONCLUSIONS

Evaluation of Generalization Techniques

The generalization techniques conducted in this study were evaluated based on the total reduction in homogeneous regions, the overall similarity between the generalized image and the original image and how well the generalization technique minimized category omission and comission errors. Because each technique required different inputs such as minimum threshold size, imposed grid size, and search radius, the results were not directly comparable. However, the strategies were evaluated by comparing all numerical measurements achieved by a particular strategy against all other strategies, based on the technique's relative reduction in region numbers.

Generalization results will vary depending on the quadrangle or area selected for analysis. However, the classification scheme used in this study was designed specifically for the land cover of Alaska and variations in generalization results should be small. The results from this study are intended for use as a guide for generalizing other Alaska quadrangles of land cover data and for further work with the area filtering strategy.

Category Aggregation

Category aggregation was achieved by clustering similar level II categories into less precise level I categories. Reduction in the number of regions from the original image was approximately 42 percent. This 42 percent reduction in region numbers was accomplished with 0 percent category omission and comission error, however a loss in category precision was experienced.

Category Aggregation and Area Filtering

The aggregated, filtered land cover data achieved the highest overall agreement and lowest category specific errors relative to the percent of region reduction compared to all the techniques examined in this study. However, as mentioned before, the land cover categories experienced a loss of precision from level II to level I.

Region Reduction

The majority of homogeneous regions which composed the aggregated, unfiltered data fell within the 1 to 68 pixel size range. Thus the decrease in regions using minimum threshold sizes larger than 68 pixels had only a minor effect in region reduction. In the unaggregated data, the majority of regions also fell within this size range, although aggregation reduced the number of regions within this range considerably. Region reduction in the aggregated data was very similar to reduction in the aggregated data after the 68 pixel minimum threshold size. Thus, category aggregation area filtering was only effective over the unaggregated data for simple region reduction in the 2 to 68 pixel range. This study was exploratory in nature, and points to a need for additional research in the alteration of minimum threshold sizes by individual land cover categories.

Overall Agreement

The KHAT and overall image agreements were the highest for the category aggregated, filtered data relative to the reduction in region numbers than any other generalized data in this study. Although region reduction leveled off after the 68 pixel minimum threshold size for the aggregated filtered data, the agreement measures continued to decline. Thus area filtered minimum threshold values above 68 tend to decrease overall image agreement but have only minor impacts on region reduction.

Individual Category Errors

In general, the omission and comission errors for the aggregated data were the lowest relative to the region reduction achieved. This study was aimed at identifying patterns among category errors for future use in tailoring land use maps for a particular user. The minimum allowable omission and comission errors were not defined for the purposes of this study because of the exploratory nature of this research. However, definite trends were identifiable.

As the minimum threshold size was increased, the aggregated, filtered data experienced an increase in omission and comission errors. Forest, Shrublands, and Ice and Snow categories had very low omission and comission errors, even with a 512 pixel area filter. Barren, Water, and Shadow, on the other hand, had high omission and comission errors relative to the other categories. Further region reduction with little increase in error might be accomplished by increasing the minimum threshold size for the low error categories and decreasing the minimum threshold size for the high error categories. Because every error in omission is also an error in comission, altering the minimum threshold sizes should be done with a particular user in mind.

Generalization by Area Filtering

Area filtering of level I and II categories was achieved by imposing a minimum threshold on all homogeneous regions. Regions which fell below this minimum size were reassigned to the most frequently occurring class value neighboring the region. Overall agreement values were very high and category specific errors were low relative to the percent of region reduction achieved with area filtering. Only the aggregated, filtered data resulted in higher overall agreement values and lower omission and comission errors relative to the reduction in the number of regions. Although the overall agrement and category specific probabilities were slightly lower, given the reduction in number of regions, the categories remained at a higher level of precision.

The implementation of a unidirectional shadow filter eliminated comission errors for the shadow category and decreased the presence of shadow by as much as 3 percent after area filtering. Although the unidirectional shadow filter did decrease the amount of shadow present on the images, it also resulted in a higher number of homogeneous regions compared with the absence of the unidirectional filter. In conclusion, the shadow filter decreased the effectiveness of the generalization strategy but may have increased the usefulness of the total image by reducing a noninformational category.

Region Reduction

The majority of homogeneous regions before filtering occurred in the 1 to 68 pixel size range. As in the aggregated data, the effects of filtering with minimum threshold sizes greater than 68 had a small effect on region reduction.

Overall Agreement

KHAT agreement values were consistently lower for the filtered images compared to the percent unchanged. Although region reduction tapered off after the 68 pixel minimum threshold size, KHAT and percent unchanged steadily decreased. Thus using a minimum threshold size above 68 decreased overall image agreement while having only a small increase in percent region reduction.

Individual Category Errors

The omission and comission errors increased with an increase in minimum threshold size. Filtering out lone pixels (minimum threshold size of 2) resulted in a region reduction of approximately 2000, while leaving all category probabilities in the 99 percentile. Errors of omission increased dramatically from the 2 to 16 sized filters and the 16 to 68 sized filters. This resulted from the large majority of regions falling between the 2 to 68 pixel range. Broadleaf Forest, Needleleaf Forest, Sparse Vegetation and Deep Water all encountered large omission errors, while shrublands, Dwarf Shrublands, Shallow Water, and Ice and Snow all encountered low errors, even at the 512 minimum pixel threshold range.

Comission errors were also very slight after using the 2 pixel area filter, leaving all categories in the 99 percentile probability. Broadleaf Forest, and Sparse Vegetation had high comission error and the categories were completely eliminated using a 512 pixel filter. Needleleaf Forest, Deep Water and Ice and Snow all retained low comission errors, even at the 512 pixel filter. The results from this technique have pointed to the need for further work with altering the minimum thresholds by category. Reduced error may result if categories with high error are given small threshold sizes while those categories with small error are given large minimum threshold areas.

Generalization by Modal Search and Replace

The USGS Interim Land Cover data was generalized using a modal search and replace strategy. This was accomplished by searching the area around each pixel based on a given radius, and reassigning the center pixel with the most frequently occuring class, or mode. The search strategy used every pixel as the center of a search area and wrote the reassigned category to a new file. The algorithm was simple in design but took a long time to implement because of the calculation of the mode for a given radius with each of the pixels in the center of that area. The modal search and replace strategy was also very inflexible. A small increase in the search radius of one pixel causes a large increase in the search area.

Region Reduction

Given a 3 x 3 search area (radius=1), the region reduction achieved was similar to that achieved by simple category aggregation. Although simple category aggregation resulted in no category errors, it did result in a lower category precision. Area filtering of unaggregated data also achieved higher percent unchanged and KHAT values based on the percent reduction in regions achieved by modal search and replace.

Overall Agreement

The KHAT values, as before, were consistently lower compared to the percent unchanged. There was a steady decrease in percent unchanged and KHAT values with an increase in search radius. The KHAT and percent unchanged tended to decreased linearly with an increase in search radius, although only three modal search tests were tried.

Individual Category Errors

Errors of omission for modal search had a similar trend when compared to the area filtered category. Broadleaf Forest, Mixed Forest and Sparse Vegetation categories had high omission errors while Needleleaf Forest, Shrublands, Dwarf Shrublands, Shallow Water, and Ice and Snow had low omission errors.

Errors of comission for modal search had a different trend among the categories compared to area filtering. The Mixed Forest and to some extent the Broadleaf Forest category had high comission errors in the modal search and replace technique but low comission errors in the area filtering results. This difference was due to the large circumference relative to the area of the mixed forest region boundary which were smoothed out in the modal search strategy.

Generalization by Resampling

Systematic resampling of the land cover data was accomplished by imposing different grid sizes over the image and selecting the pixels which fell within the center of each grid cess. The resampled data were then multiplied back to its original size. Of all the generalization strategies examined in this study, resampling produced the least desirable results. Although the simplicity of this technique allowed the algorithm to be implemented quickly, the poor overall agreement to the original data and the low category specific probabilities made this strategy highly undesirable.

Region Reduction

The number of homogeneous regions decreased with increasing grid cell sizes. However very low KHAT and percent unchanged values were obtained in relation to the overall agreement and category probabilities obtained with area filtering and modal search and replace.

Individual Category Errors

The omission and comission errors resulting from this strategy were largely based upon the percent of total image each category occupied. The omission and comission errors were nearly identical. This may have resulted from the small difference in imposed grid cell size as well as the large image size.

SUMMARY

Personnel associated with the Federal Land Information System recently acquired USGS Interim Land Cover raster data for the Valdez, Alaska quadrangle. This data set contained approximately 98,000 land use regions. The large number of regions occupied a sizeable amount of computer memory space, and caused resulting land cover maps to be extremely complex and difficult to read. Therefore, generalization was deemed necessary in order to reduce the data volume and to simplify the category spatial patterns before converting the data into a vector geographic information system format.

Because reduction of homogeneous regions effects the overall image similarity to the original data and has varying effects on individual categories, a generalization strategy was sought which maximized the reduction in homogeneous regions while minimizing the impact on individual categories. Four different generalization strategies were conducted on a portion of the Valdez Interim Land Cover data. The generalization strategies examined were; 1)category aggregation, 2)area filtering, 3)modal search and replace, and 4)systematic resampling.

Of the strategies tested, simple category aggregation to a lower precision gave the highest reduction in homogeneous regions. However, the data did experience a loss in

precision. Category aggregation and area filtered data also gave very high overall agreement with the original data and little individual category error relative to the region reduction. However, as before, the data did experience a loss in precision.

Of the three strategies conducted with unaggregated categories, area filtered data gave very high overall agreement with the original data and little individual category error relative to the region reduction. This resulted from the algorithm's consideration of contiguous pixels for reassignment based on a minimum threshold area. The modal search and replace generalized data resulted in lower overall agreement values and higher category specific errors. The algorithm's less desirable products were caused by the restriction of the search radius, and the loss of contiguously occurring pixels because of nearby domination of a larger category.

The strategy which produced the most undesirable results was the systematic resampling generalization technique. This technique produced data which was largely based upon the individual categories' total percent of an image. Overall agreement values were very low and individual category errors were very high based on the reduction in homogeneous regions. The biggest problem with this algorithm is the loss of natural category boundaries with the multiplication factor associated with this technique.

The results from the area filtering tests point to the need for additional research in the 2 to 68 minimum pixel threshold sizes, and for alteration of these threshold sizes by category. The optimum generalization strategy can only be selected based on a specific purpose for the particular land use/land cover map. Although this research was exploratory in nature and was not directed towards a particular use of land cover maps, the results have indicated the superiority of area filtering over the other techniques tested. Also, the resulting category error patterns from the generalization process have provided FLIS personnel a basis for continuing research with the area filtering technique.

LITERATURE CITED

Anderson, James R., Ernest E. Hardy, John T. Roach, and Richard E. Witmer. 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data, U.S. Geological Survey Professional Paper 964, U.S. Government Printing Office. Washington, D.C.: U.S. Geological Survey, 28 pp.

Andrews, H.C. 1970. Computer Techniques in Image Processing. New York: Academic Press, 187 pp.

Bishop, Yvonne M., Stephen E. Fienberg, and Paul W. Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. Cambridge: The MIT Press, pp. 395-397.

Bonnor, G. M. 1970. Newfoundland Forest Inventory Methods. Internal Report no. FMR-16 (unpublished). Newfoundland: Forestry Management Institute, 31 pp.

Cohen, J. 1960. A Coefficient of Agreement of Nominal Scales. Educational and Psychological Measurement. Vol. 20, No. 1, pp. 37-46.

Campbell, James B. 1983. Mapping the Land. Washington, D.C.: Association of American Geographers, 96 pp.

Crapper, P.F., P.A. Walker, P.M. Nanninga. 1986. Theoretical Prediction of the Effect of Aggregation on Grid Cell Data Sets. Geo-Processing. Vol. 3, pp. 155-166.

Davis, W.A. and F.G. Peet. 1976. The Identification and Reclassification of Small Regions on Digital Thematic Maps. Information Report FMR-X-90. Ottawa: Forest Management Institute, 22 pp.

Davis, W.A. and F.G. Peet. 1977. A Method of Smoothing Digital Thematic Maps. Remote Sensing of Environment. Vol. 6, pp. 45-49.

Douglas, D.H., and Peucker, T.K. 1973. Algorithms for the Reduction of the Numbers of Points Required to Represent a Digitized Line or its Caricature. The Canadian Cartographer. Vol. 10, pp. 112-122.

Earth Resource Observation System. 1986. <u>LAS User's Manual</u>. Sioux Falls, SD.

Environmental Systems Research Institute. 1986. ARC/INFO User's Manual. Redlands, CA.

Fabos, J.G., R. Careaga, C. Greene, and S. Williston. 1973. Model for landscape resource assessment-Part 1 of the <u>Metropolitan Landscape Planning Model</u> (METLAND). Research Bullentin No. 602. Amherst: University of Massachusetts Agricultural Experiment Station, 45 pp.

Goldberg, M. and D.G. Goodenough. 1978. Analysis of a Spatial Filter for Landsat Imagery. <u>Journal of Applied Photogrammetric Engineering</u>. Vol 4, No. 1., pp. 25-27.

Goldberg, M., W.A. Hallada, W. Lindboe, and R.F. Marcell. 1984. Topological Grid Structure: A Data Structure for Earth Science Modeling. <u>1984 Pecora Proceedings</u>. Robert M. Haralick, Chairperson. Sioux Falls, SD: IEEE Computer Society Press, pp. 83-86.

Hsu, Shin-yi and Xingyuan Huang. 1982. Raster-Vector Conversion Methods for Automated Cartography with Applications in Polygon Maps and Feature Analysis. <u>Auto-Carto 5 Proceedings</u>. Jack Foreman, Editor. Crystal City, VA: ASPRS and ACSM, pp. 407-418.

Fosnight, Eugene. 1985. Nominal Filtering and Automated Generation of Complex Mapping Units. U.S. Geological Survey Unpublished Project Plan. EROS Data Center, Sioux Falls, SD. 17 pp.

Guptill, Stephen C. 1978. An Optimal Filter for Maps Showing Nominal Data. <u>Journal of Research U.S. Geological Survey</u>. Vol 6, No. 2, pp. 161-167.

Guptill, Stephen C. 1978. An Evaluative Technique for Categorical Maps. <u>Geographical Analysis</u>. Vol. 10, No. 3, pp. 249-261.

Guptill, Stephen C. 1975. Spatial Filtering of Nominal Data: An Exploration. Ann Arbor, MI: University of Michigan unpublished Ph.D. dissertation, 210 pp.

Hole, Francis D. and James B. Campbell. 1985. <u>Soil Landscape Analysis</u>. Totowa, New Jersy: Rowman & Allanheld, pp. 125-142

Kalensky, Z. 1974. ERTS Thematic Map from Multidate Images. Proceedings from the Symposium on Remote Sensing. Banff, Alberta: International Socitey of Photogrammetry, Vol. 1, pp. 321-332.

Kan, E.P.F., and J.K. Lo. 1975. A New Image Enhancement Algorithm with Application to Forestry Stand Mapping. Lyndon B. Johnson Space Center, Houston, Texas: NASA Report No. LEC-6178, 16 pp.

Kolassa, David A. 1983. The Evolution of Raster Processing Technology within the Cartographic Environment. Auto-Carto 5 Proceedings, Editor Jack Foreman. Falls Chruch: ASP and ACSM, pp. 451-460.

Lauer, Donald T. 1986. Applications of Landsat Data and the Data Base Approach. Photogrammetric Engineering and Remote Sensing. Vol. 52, No. 8, pp. 1193-1199.

Lintz, Joesph, Jr., and David S. Simonett, editors. 1976. Remote Sensing of Environment. Reading, Mass: Addison-Wesley Publishing Co, 694 pp.

Logan, Thomas L. and Curtis E. Woodcock. 1982. User Alternatives in Post-Processing for Raster-to-Vector Conversion. ISPRS IV Proceedings, Editor Jack Foreman. Crystal City, VA: ASPRS and ACSM, pp. 397-418.

MacDougall, E. B. 1972. Optimal Generalization of Mosaic Maps. Geographical Analysis. Vol. 4, No. 4, pp. 416-423.

Marble, Duane F. 1984. Geographic Information Systems: An Overview. Pecora 1984 Proceedings, Robert M. Haralick, chairperson. Silver Spring, MD: IEEE Computer Society Press, pp. 18-24.

Monmonier, Mark Stephen. 1974. Measures of Pattern Complexity for Choropleth Maps. American Cartographer. Vol 1, No. 2, pp. 159-169.

Monmonier, Mark Stephen. 1983. Raster-Mode Area Generalization for Land Use and Land Cover Maps. Cartographica. Vol 20, No. 4, pp. 65-91.

Monmonier, Mark Stephen. 1985. Technical Transition in Cartography. Madison, WI: University of Wisconsin Board of Regents, 282 pp.

Nichols, D.A. 1981. Conversion of Raster Coded Images to Polygonal Data Structures. Proceedings of the 7th Annual William T. Pecora Memorial Symposium, Benjamin F. Richason, Editor. Sioux Falls, SD: ASP, pp. 508-515.

Nichols, J.D. 1975. Characteristics of Computerized Soil Maps. Soil Science Society of America Proceedings. Vol. 39, pp. 927-32.

Peuquet, Donna J. 1979. Raster Processing: An Alternative Approach to Automated Cartographic Data Handling. American Cartographer. Vol. 6, No. 2., pp. 129-139.

Peuquet, Donna J. 1984. A Conceptual Framework and Comparison of Spatial Data Models. Cartographica. Vol. 21, No. 4, pp. 66-113.

Robinove, Charles J. 1986. Spatial Diversity Index Mapping of Classes in Grid Cell Maps. Photogrammetric Engineering and Remote Sensing. Vol. 52, No. 8, pp. 1171-1173.

Rosenfeld, Azriel. 1969. Picture Processing by Computer. New York: Academic Press, 196 pp.

Rosenfeld, Azriel. 1978. Extraction of Topological Information From Digital Images. Proceedings from the First International Advanced Symposium on Topological Data Structures for Geographic Information Systems. Geoffery Duton, editor. Harvard: Harvard University Press, Vol. 6, pp. 790-805.

Rosenfield, George H., and Katherine Fitzpatrick-Lins. 1986. A Coefficient of Agreement as a Measure of Thematic Classification Accuracy. Photogrammetric Engineering and Remote Sensing. Vol. 52, No. 2, pp. 223-227.

Rowe, H. E. 1965. Signals and Noise in Communication Systems. Princeton, N.J: D. Van Nostrand Co., Inc., 257 pp.

Selden, David D. and Michael A. Domaratz. 1983. Digital Map Generalization and Production Techniques. Auto-Carto 5 Proceedings. Jack Foreman, editor. Falls Chruch, VA: ASP and ACSM, pp. 241-247.

Shasby, Mark and David Carneggie. 1986. Vegetation and Terrain Mapping in Alaska Using Landsat MSS and Digital Terrain Data. Photogrammetric Engineering and Remote Sensing. Vol. 52, No. 6, pp. 779-786.

Strong, J.P., and Azriel Rosenfeld. 1973. A Region Coloring Technique for Scene Analysis. Communication ACM. Vol. 16, No. 4, p. 237-246.

Story, Michael and Russell Congalton. 1986. Accuracy Assessment: A User's Perspective. Photogrammetric Engineering and Remote Sensing. Vol 52, No. 3, pp. 397-399.

Sturdevant, James. 1986. Notes on inital filtering tests for area threshold sizes. EROS Data Center, Sioux Falls, South Dakota.

U.S. Geological Survey. 1984. Copper River Area Land Cover Mapping Project. Internal Report, EROS Data Center. Sioux Falls, South Dakota, 110 pp.

U.S. Geological Survey. 1986. Alaska Federal Land Information System (FLIS) Program Workplan. Internal Report, EROS Data Center. Sioux Falls, South Dakota, 6 pp.

Viereck, L.A. and C.T. Dryness. 1980. A Preliminary Classification system for Vegetation of Alaska. U.S. Department of Agriculture, Forest Service General Technical Report PNW-106, 36 pp.

APPENDIX

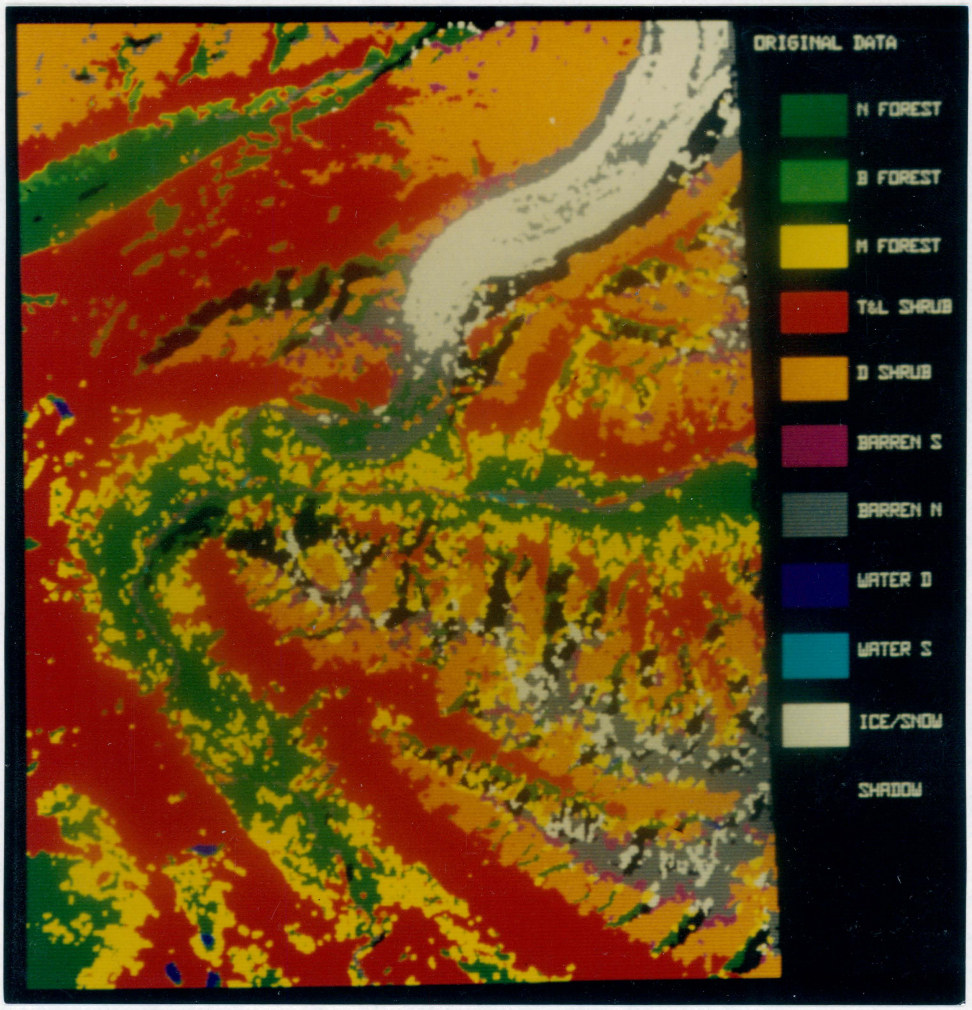


FIGURE 30
Original Valdez Land Cover Image

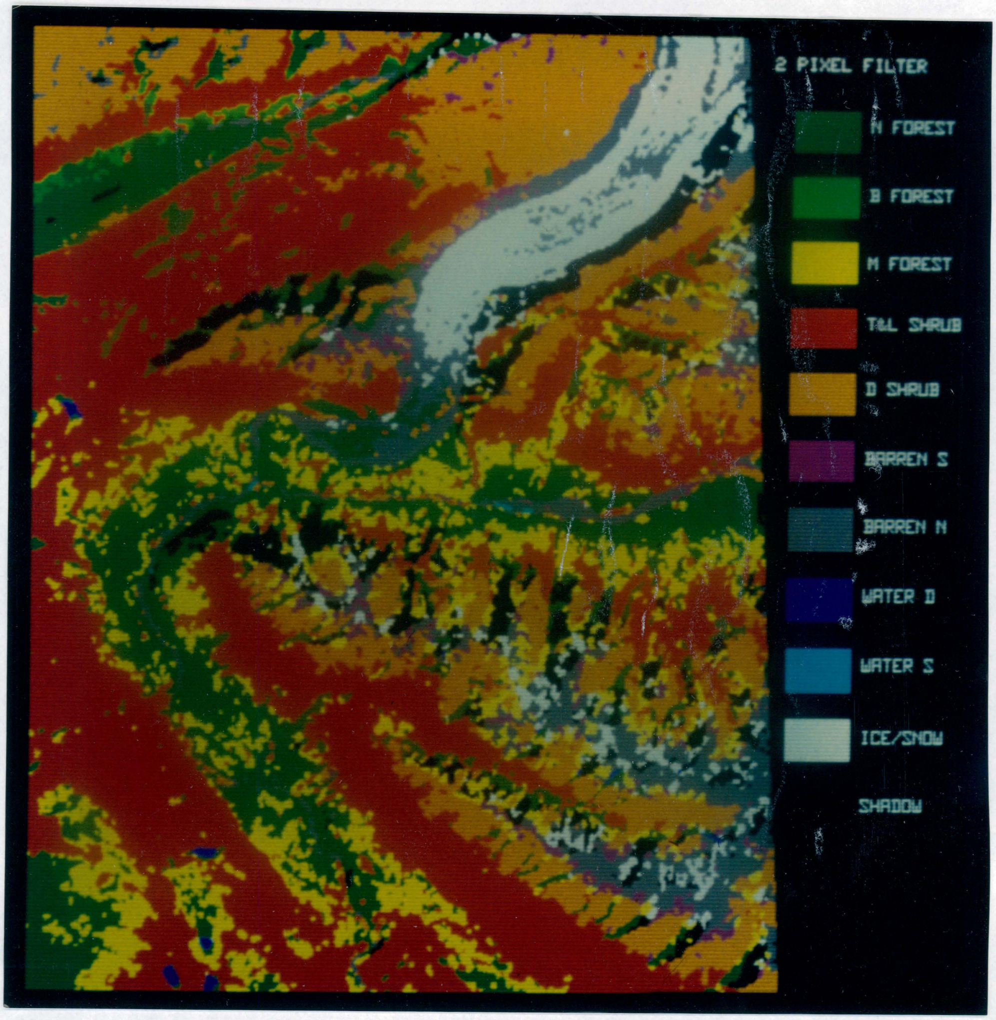


FIGURE 31
Two Pixel Area Filtered Image

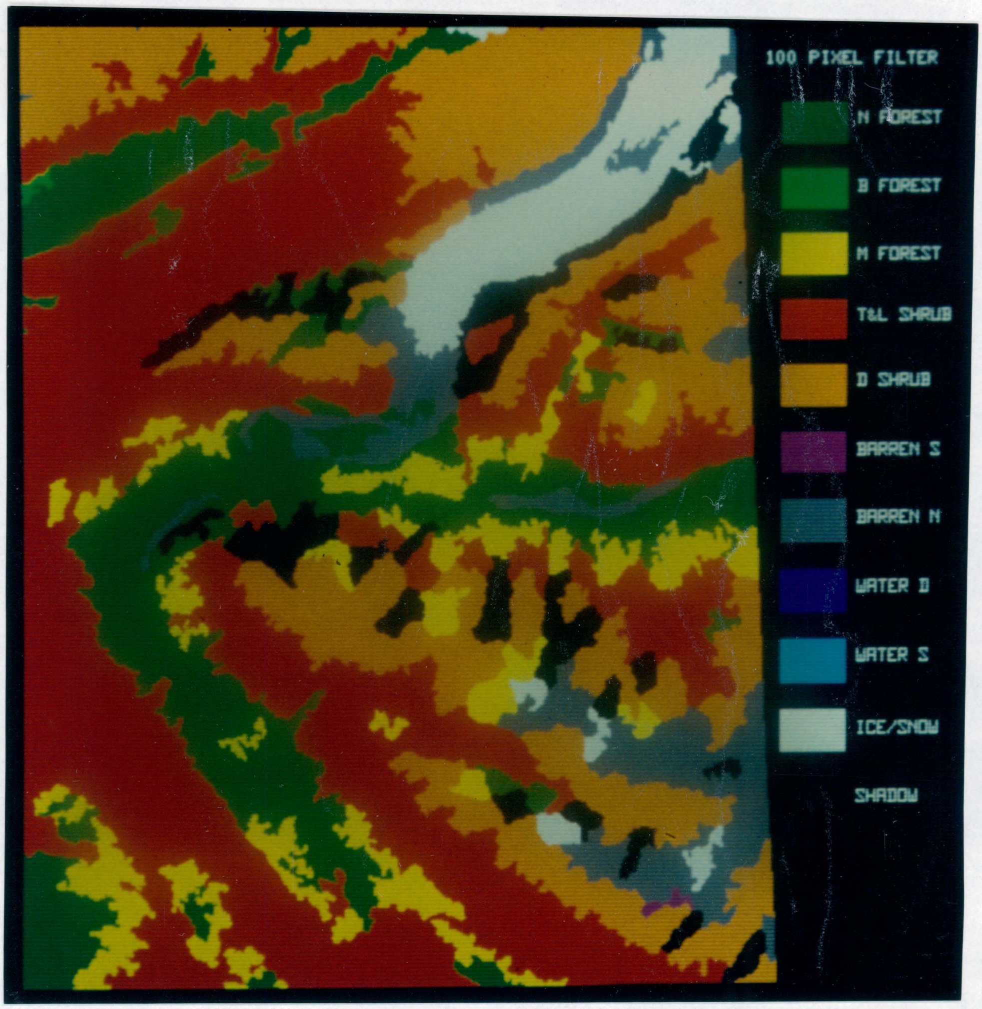


FIGURE 32
One Hundred Pixel Ara Filtered Image

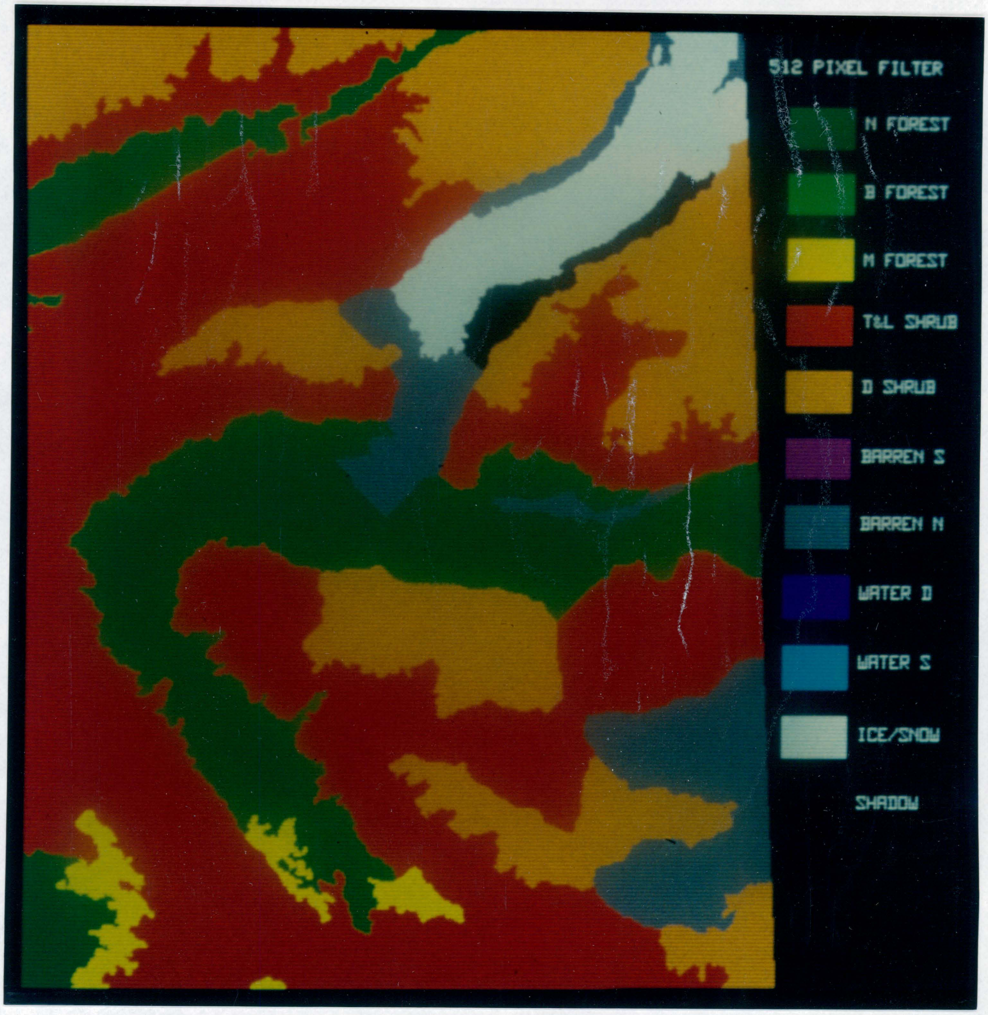


FIGURE 33
Five Hundred Twelve Pixel Area Filtered Image

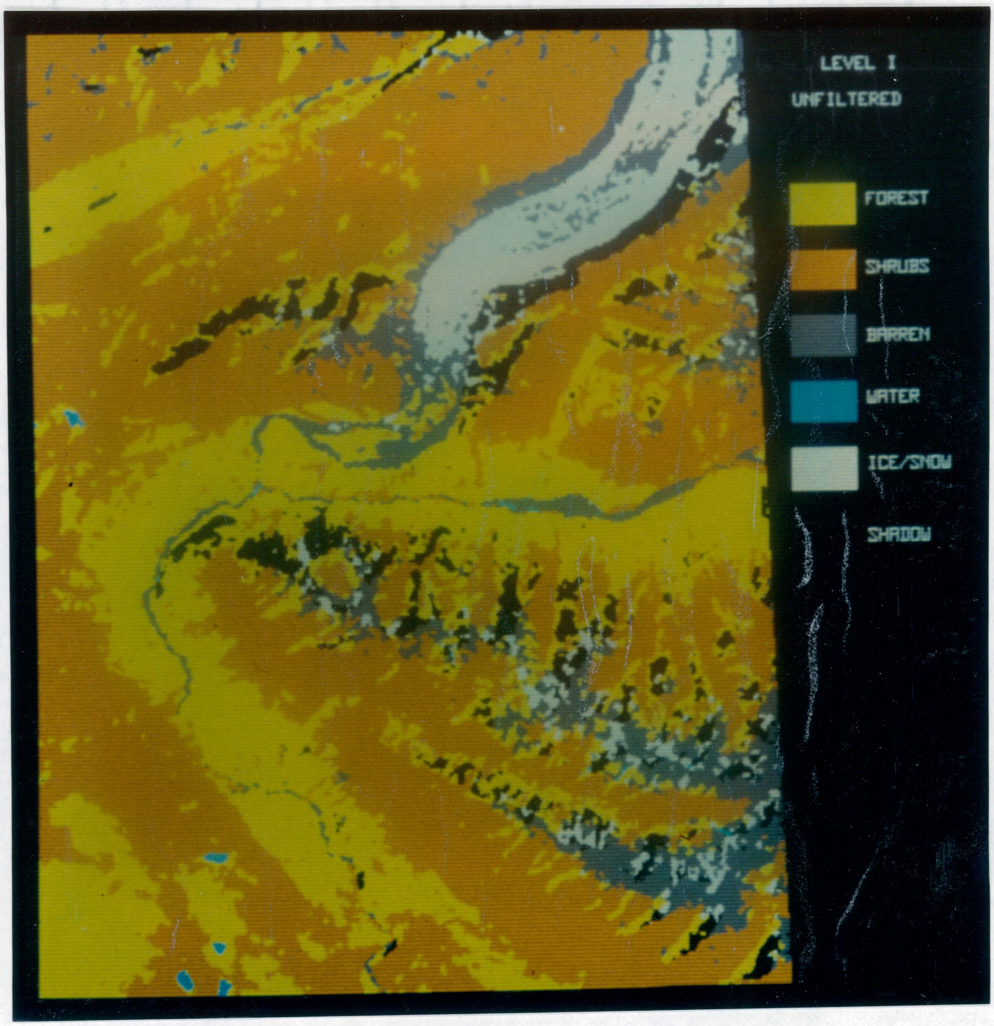


FIGURE 34
Aggregated Unfiltered Image

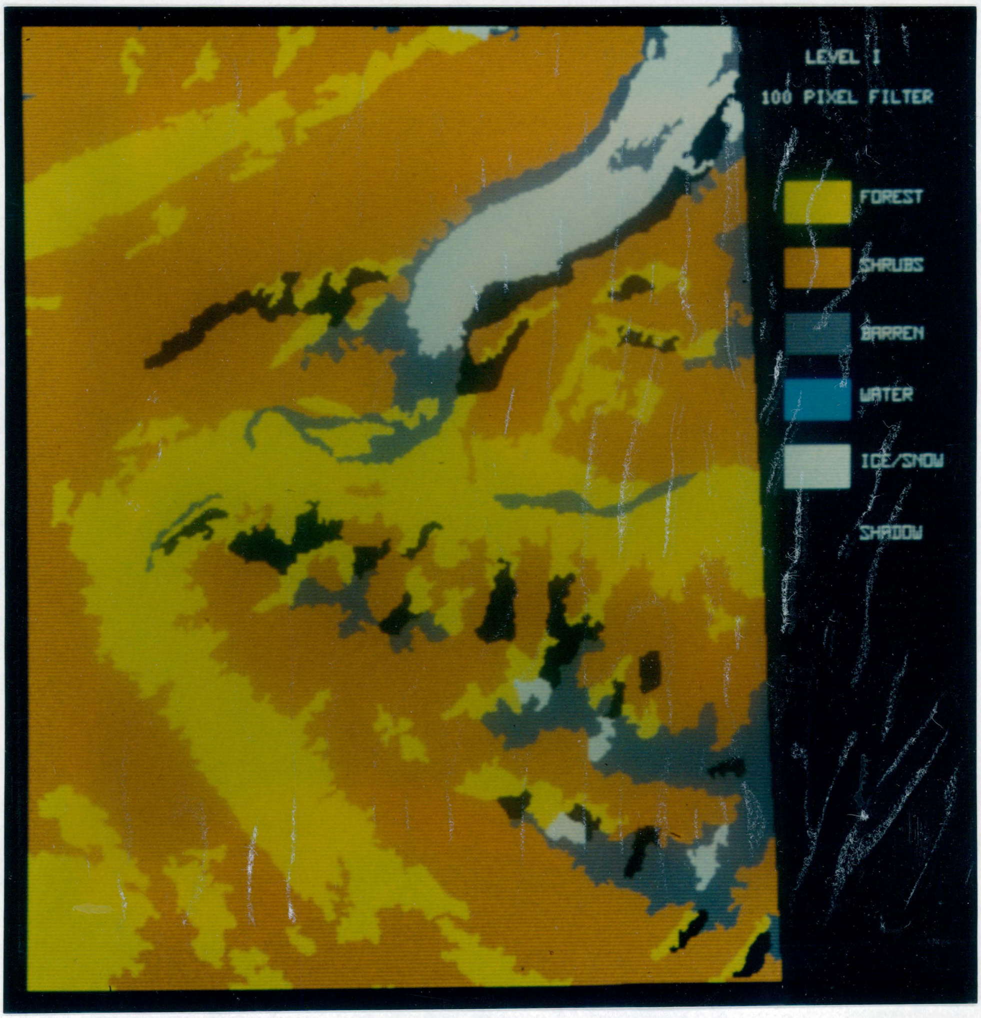


FIGURE 35
Aggregated One Hundred Pixel Filtered Image

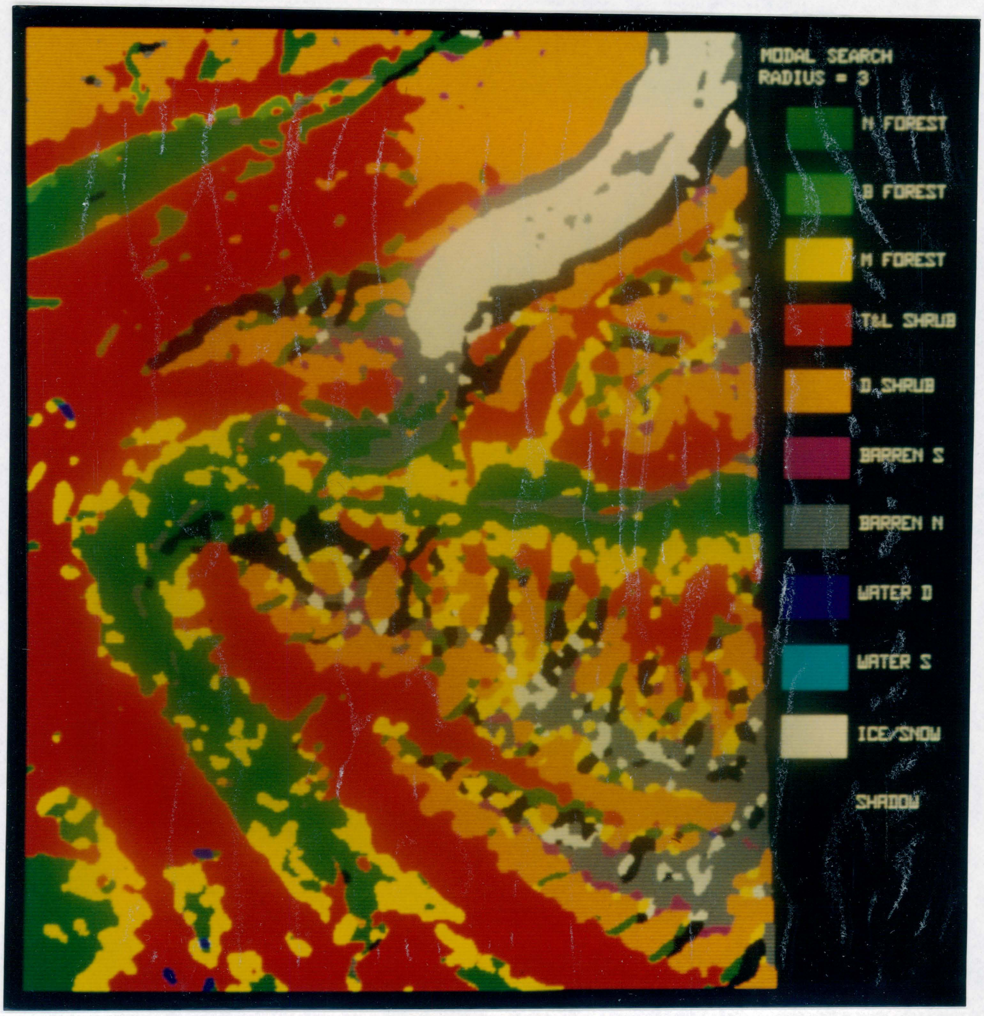


FIGURE 36
Replaced Image with a Search Radius of Three Pixels

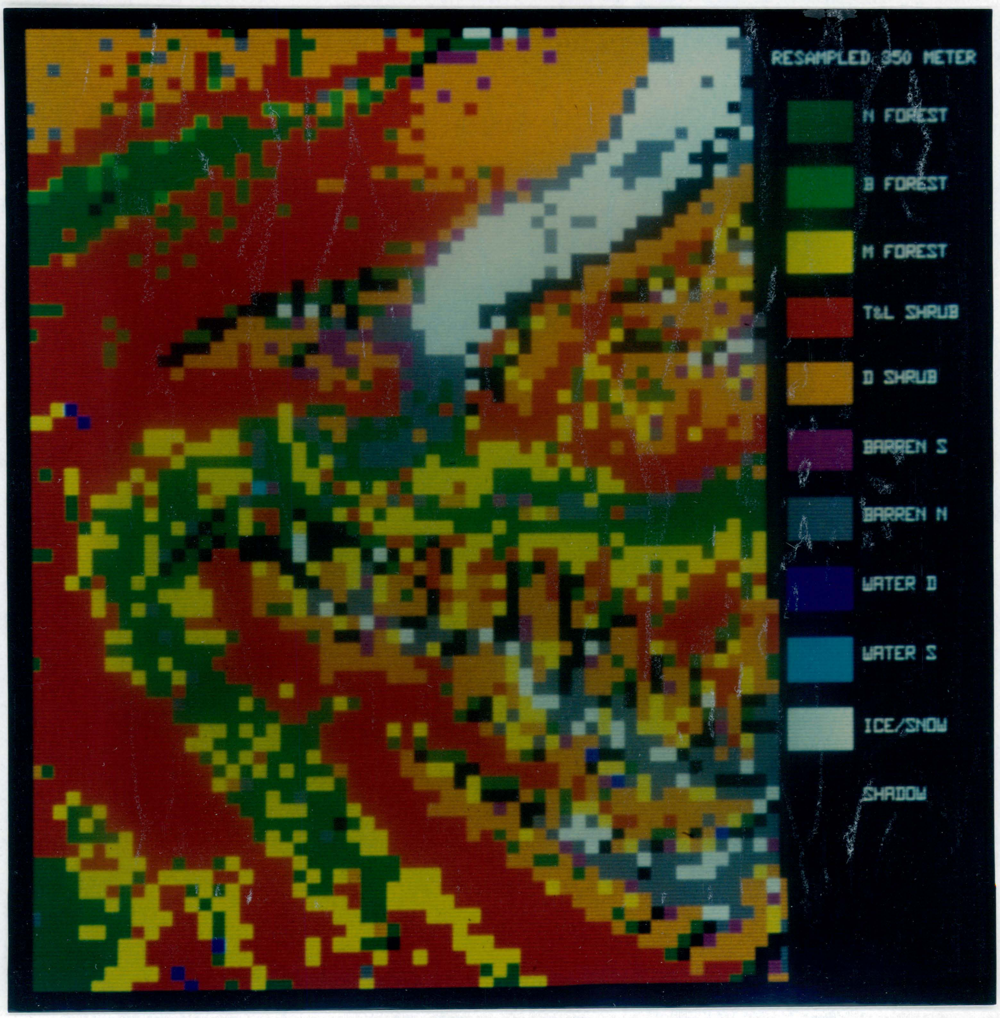


FIGURE 37
Resampled with a 350 Meter Grid Size

The vita has been removed from the scanned document