Image Processing for the Development of Detailed Quantitative Land Cover Information - 'Discrete Classification'

By
Ken Stumpf and John Koltun
Geographic Resource Solutions
1125 16th Street, Ste 213
Arcata, CA
(707)822-8005
grs@grsgis.com

Where are we going today ?

- Land Cover Mapping Data
- The GRS Discrete Classification methodology -
  - What is different compared to typical methods ?
  - What obstacles are overcome ?
- Classification data products
  - Concerns with traditional “mythods”
  - Benefits of this methodology

Why Us ?

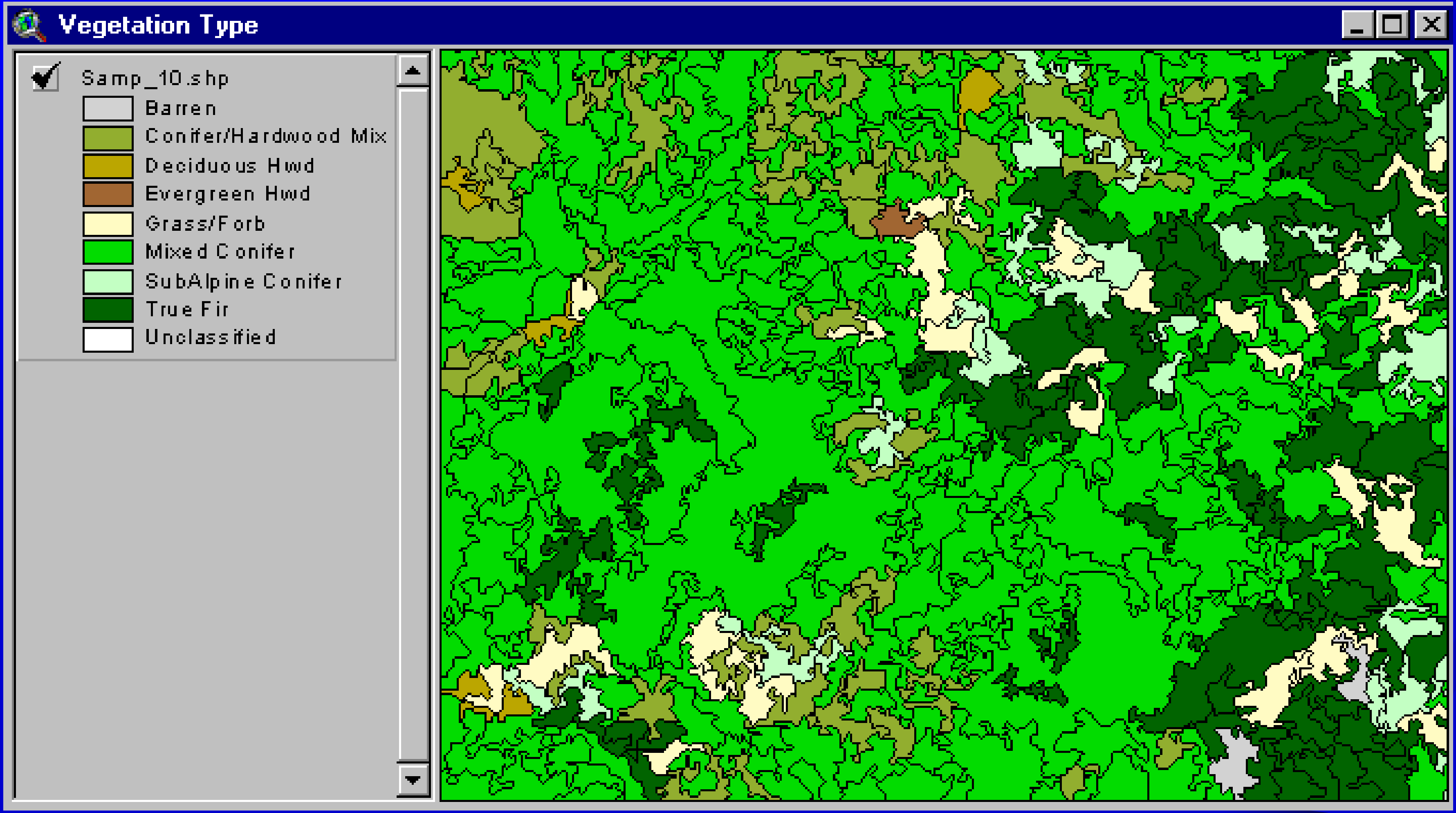
- Background
- Healthy skepticism ...

Participant Issues ...

Land Cover Classification Map Products

- Categorical Maps - general types
  - **general cover-type description**
    - **Forb, Barren, Shrub, Conifer, Hardwood, ...**

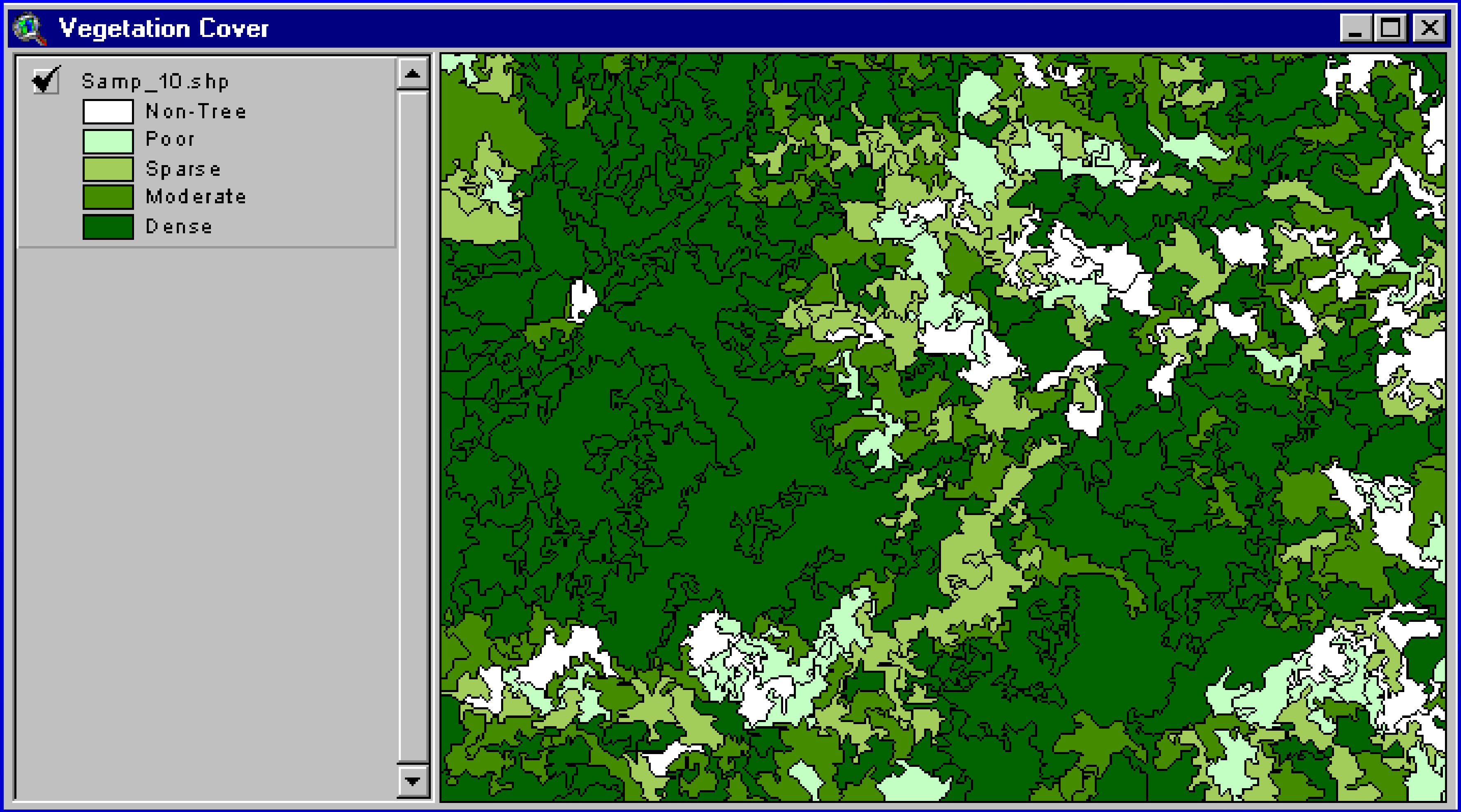
Categorical Map Data – Cover-Type



Land Cover Classification Map Products

- Categorical Maps - general types
  - – general cover-type description
    - Forb, Barren, Shrub, Conifer, Hardwood, …
  - – **general density class values**
    - **Sparse, Poor, Moderate, Dense ….**

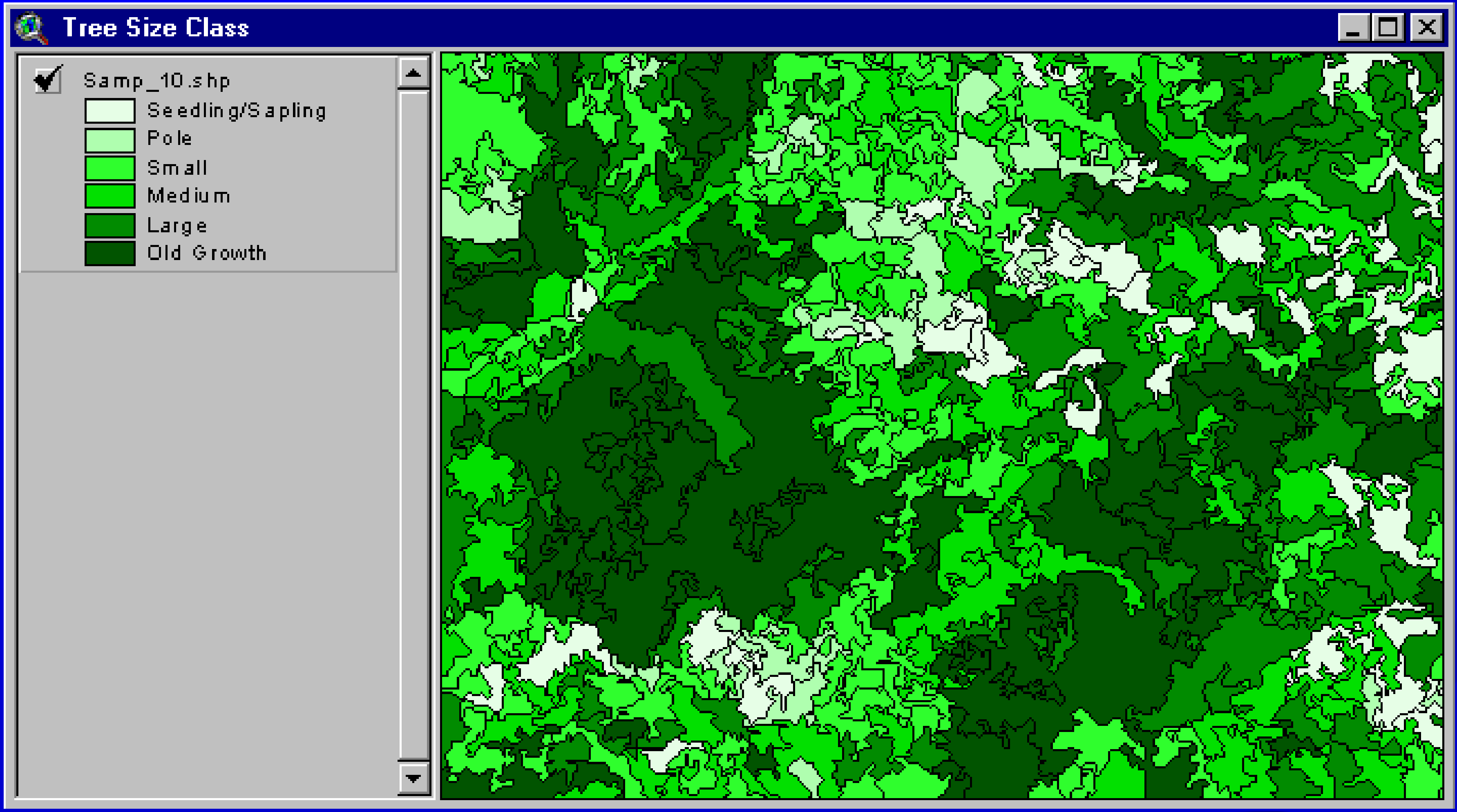
Categorical Map Data - Density



Land Cover Classification Map Products

- Categorical Maps - general types
  - general cover-type description
    - Forb, Barren, Shrub, Conifer, Hardwood, …
  - general density class values
    - Sparse, Poor, Moderate, Dense ….
  - **general size class values**
    - **Sapling, Pole, Small, Medium, Large, …**

Categorical Map Data – Tree Size



Land Cover Classification Map Products

- Categorical Maps
  - general cover-type description
    - Forb, Barren, Shrub, Conifer, Hardwood, …
  - general density class values
    - Sparse, Poor, Moderate, Dense ….
  - general size class values
    - Sapling, Pole, Small, Medium, Large, …
- Ready for use, analysis, and distribution

But then … problems !!

What is wrong with the map ?

- It’s not 90% correct like I thought, but more like 60-70% correct.
- My data are too general and not very useful !

- I used the manual/documentation …
- I followed the steps in the right order …
- I hit all the right buttons …

What Happened ?

Image Classification doesn’t work …

“The Pitfalls of Image Classification”

or

What I learned in school that I now need to forget !

The Obvious Answers …

- The mapping project is much more complex …
  - Processes that may work in small localized areas can't be applied well over large areas
  - Larger area leads to greater complexity and confusion
    - Terrain – slope and aspect
    - Ecological regions
    - Elevation differences
- We need detail to create detail !
- We experience glitches in the results due to the *pitfalls* of image classification and land cover mapping methodologies.

The Most Common Pitfalls

- The data
  - Imagery
    - resampling algorithm
    - differential illumination
  - "Ground-truth" - field data
  - Training site selection
- Training set development
- Classification techniques
- Pixel map cleanup and modeling
- Accuracy assessment

Why is the map data not more useful ?

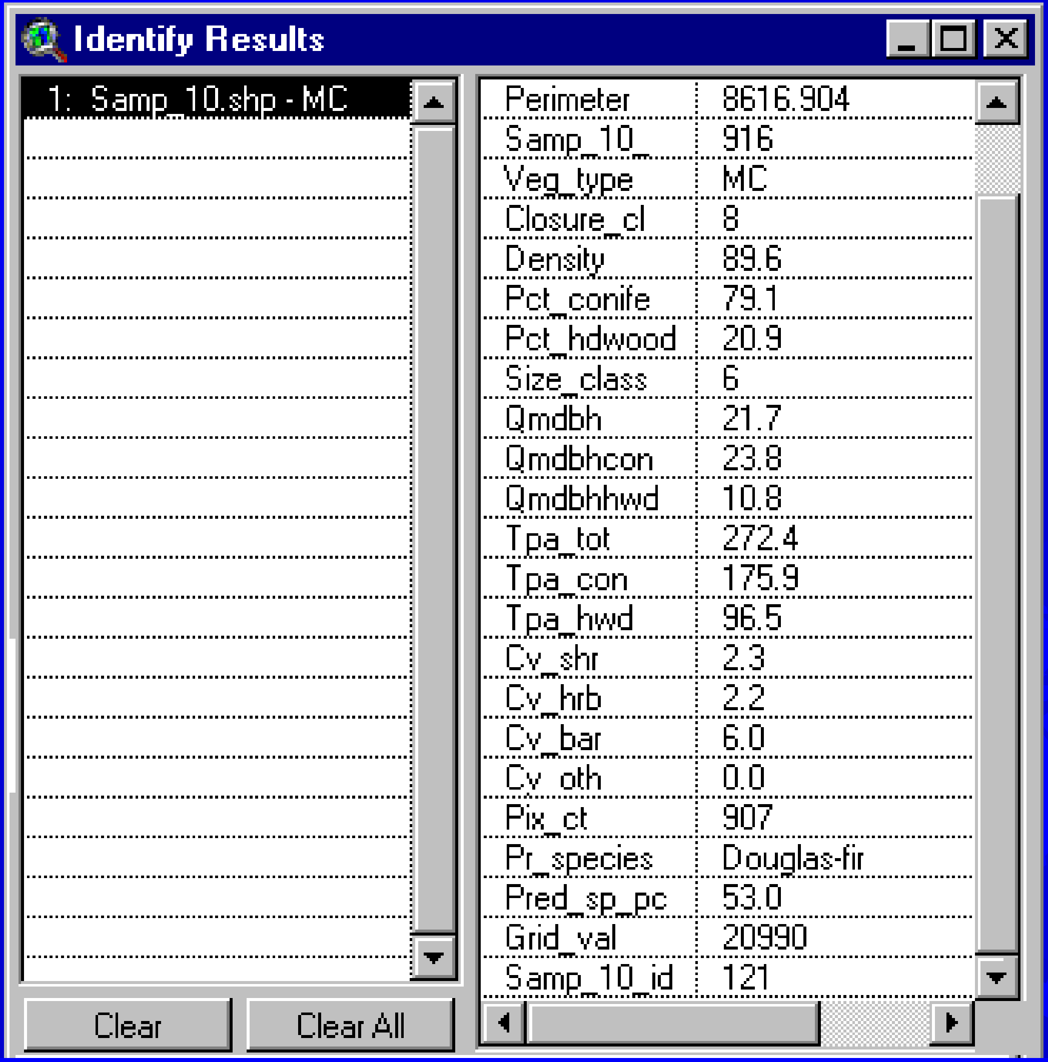
Categorical Data ….

Classification Map Products

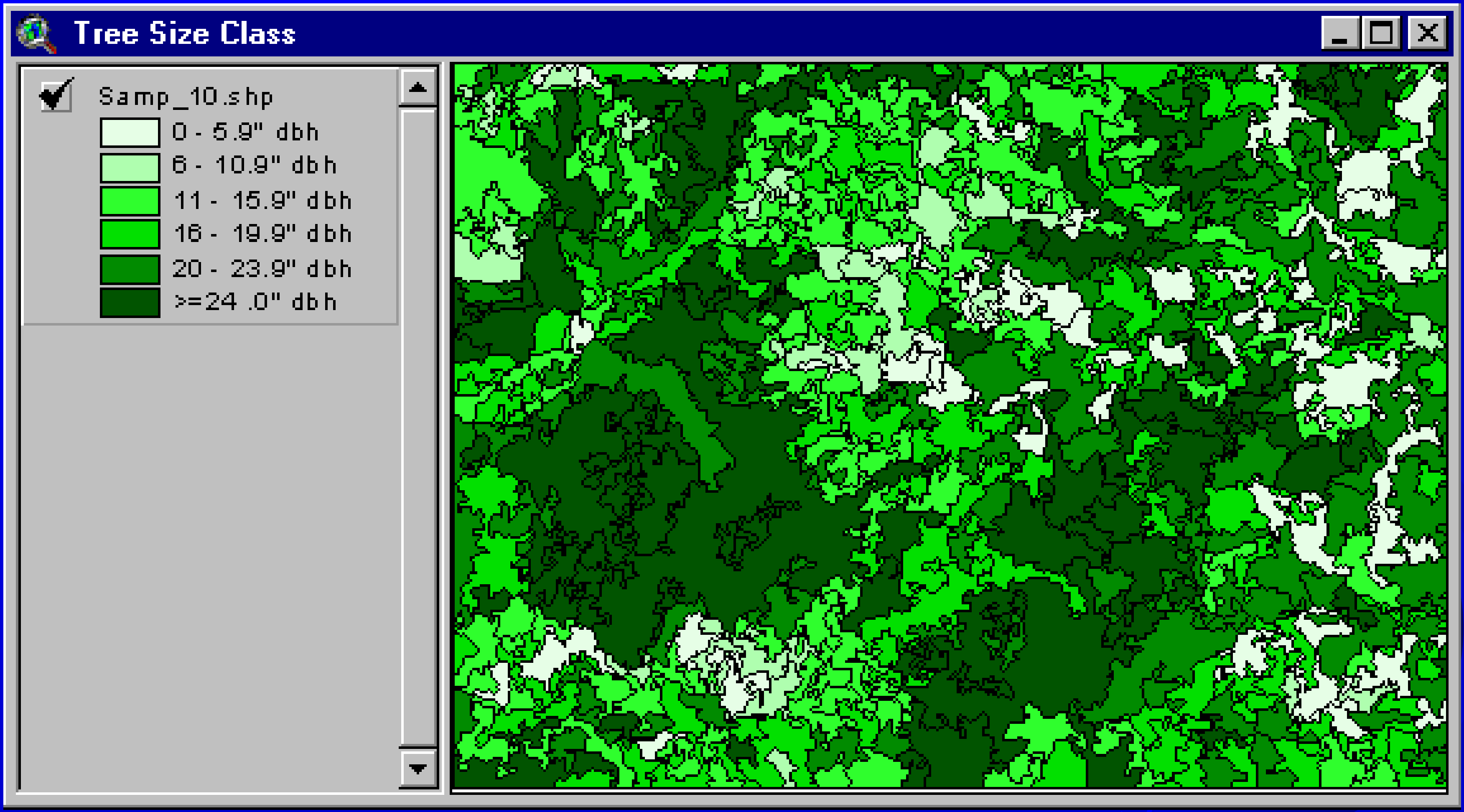
- Quantitative - Accurate - Usable Map Data
  - Cover by Species
    - 29 % cover redwood, 35 % cover tanoak, 10 % fern,

      and 26% litter and duff
  - Average Tree Size
    - Conifer = 23.8” qmd
    - Hardwood = 10.8” qmd
  - Specific Type Names
    - Mixed Conifer/Douglas-fir
    - Alder-willow: low shrub: open

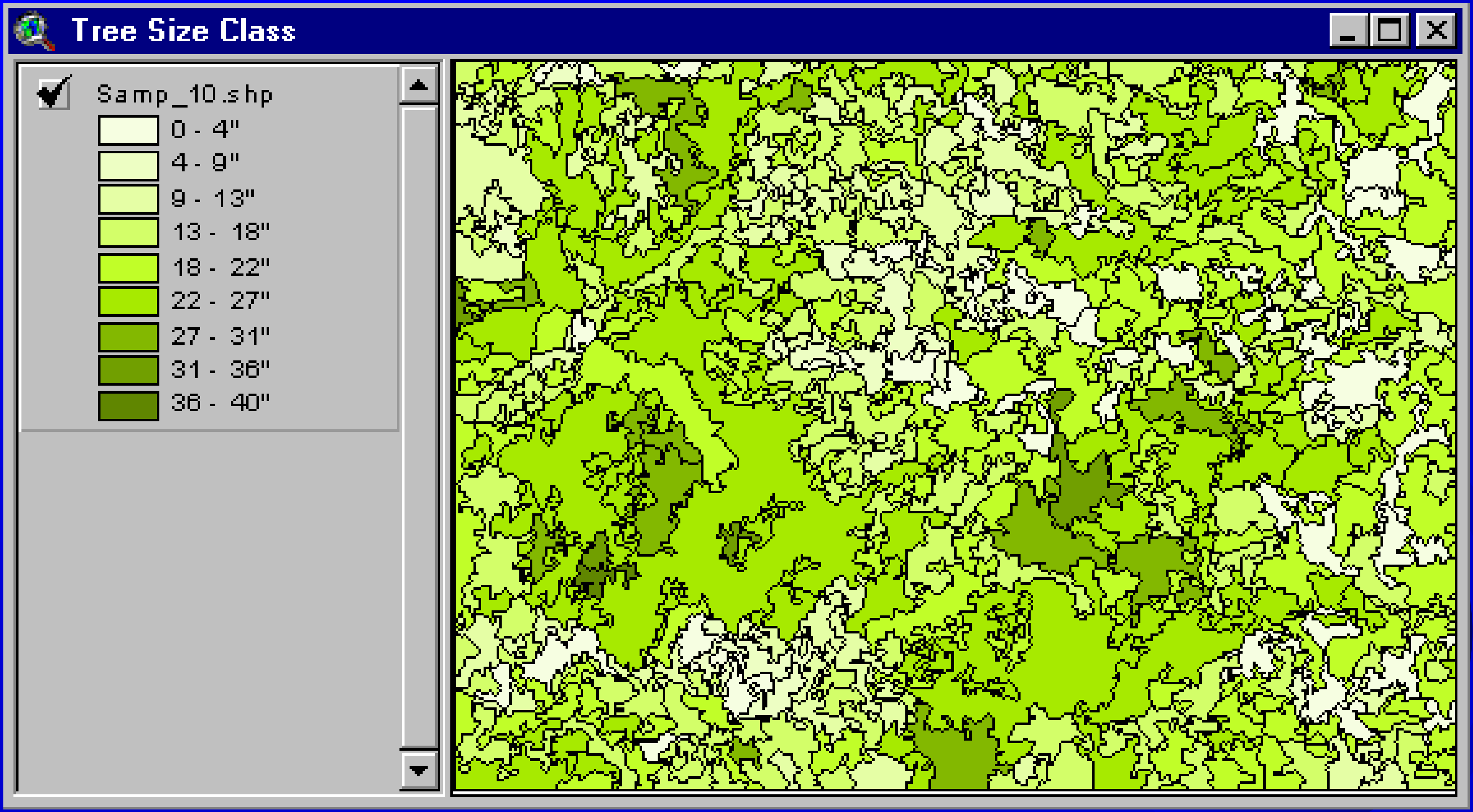
Quantitative Map Data



Quantitative Map Data



Quantitative Map Data



Typical Land Cover Mapping Project

- Acquire imagery
- Collect some “ground truth” - training
- Classify the imagery using training data
- Clean up the pixel map data
- Generate the final data set
- Develop estimates of map accuracy

GRS Uses A Different Approach

- The ‘Best’ Imagery
- Illumination Correction
- Training Site Selection
- Quantitative Ground-truth
- Training Set Development
- Hybrid Classification
- Rule-based Pixel Aggregation
- Accuracy Assessment

The Best Imagery – Potential Problems

- Striping ?
- Saturation ?
- Resampling Algorithm
  - Cubic Convolution or Nearest Neighbor ?

Federal procurement standard (USDI) calls for resampling using cubic convolution algorithm.

Different Results

Do you want smoothed (distorted) data or the 'same' values ?

What would happen to a checker board pattern ?

One method changes the distribution of the data !

CROSS TABULATION REPORT

Rows represent grid file : c:\mgeprojects\pifw\imagery\6545sr_4.tif
Columns represent grid file : c:\mgeprojects\pifw\imagery\6545sn_4.tif

	9	10	11	12	13	14	15	16	17	18	19
1	0	0	0	0	0	0	0	0	1	0	0
2	0	0	1	0	1	2	0	0	2	0	0
4	93	56	12	5	3	2	1	1	1	1	3
5	346	6220	2242	193	53	25	10	8	5	2	1
6	32	8967	45726	19677	2206	444	102	59	31	9	9
8	2	523	39589	147512	90268	15174	2136	428	160	82	42
9	0	44	4596	117078	417638	317475	51136	6153	1155	360	175
10	0	3	672	20706	288472	1145874	750584	90120	11019	2073	670
12	0	2	75	2705	43643	460976	1965864	914524	108650	14635	3318
13	0	0	17	321	4906	51966	425443	2007563	868179	106208	17772
14	0	0	4	49	649	6137	49607	344372	1812155	690515	97294
16	0	0	1	19	133	1024	7404	46614	300060	1457567	551353
17	0	0	1	5	55	302	1672	8790	46745	262949	1331385
19	0	0	0	3	20	94	537	2213	9885	41919	248370
20	0	0	0	2	20	60	198	840	2979	10233	44421
21	0	0	1	4	10	37	114	336	1174	3462	12693
23	0	0	0	0	5	17	67	158	539	1509	4528
24	0	0	0	0	3	15	29	80	283	753	2109
25	0	0	0	0	1	6	32	51	151	396	1032
27	0	0	0	0	1	8	9	34	102	256	577
28	0	0	0	0	0	2	11	22	60	151	354
29	0	0	0	0	0	1	5	10	39	104	208
31	0	0	0	0	0	1	1	8	20	64	136
32	0	0	0	0	0	0	2	8	16	37	101
33	0	0	0	0	1	0	2	3	13	24	67
35	0	0	0	0	0	1	2	3	4	20	52

Illumination/Topographic Correction

- Differential illumination in imagery is caused by:
  - topography (slope and aspect)
  - sun angle at time of acquisition
  - direction (azimuth) of sensor
- Differential illumination causes:
  - Confusion of training data
  - Increased variances
  - Reduction in accuracy

Illumination/Topographic Correction

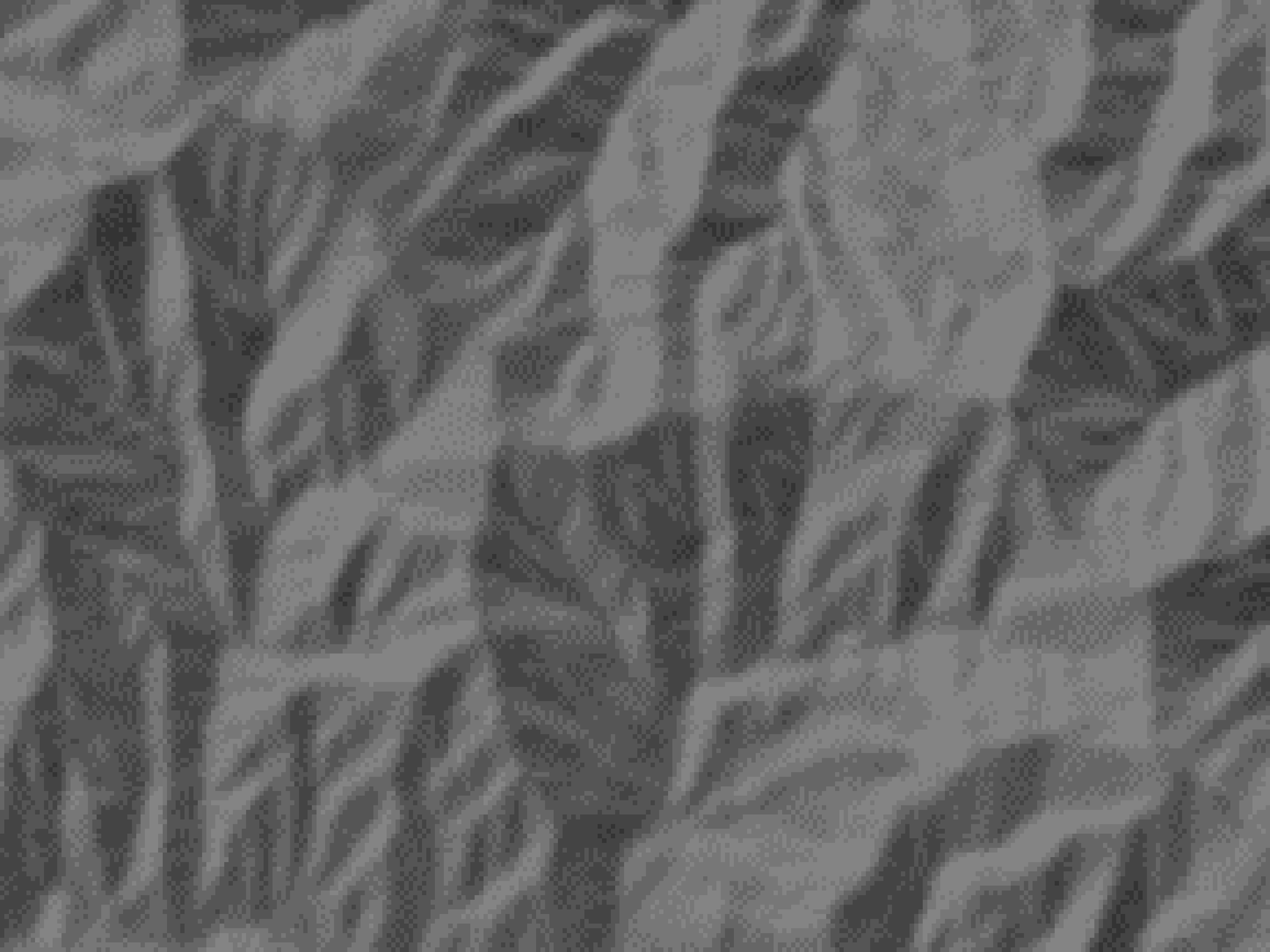
- Major classification issues
  - Same land cover type - different spectral data requires additional training sites
  - Different land cover type - same spectral data causes confusion
- Illumination Correction leads to
  - Fewer Training Sites
  - Reduced Variance Within Type Strata
  - Higher Classification Accuracy

Errors Related to Differential Illumination

Errors in Size Class Related to Aspect

Aspect	Sample Size	Percent Correct	Kappa	Z Score
N,NW,W	119	58%	.4651	
				2.74
NE,E,SE,S,SW	177	77%	.6603	

Z score > 1.96 indicate a significant difference @ 95% probability level



Training Site Selection

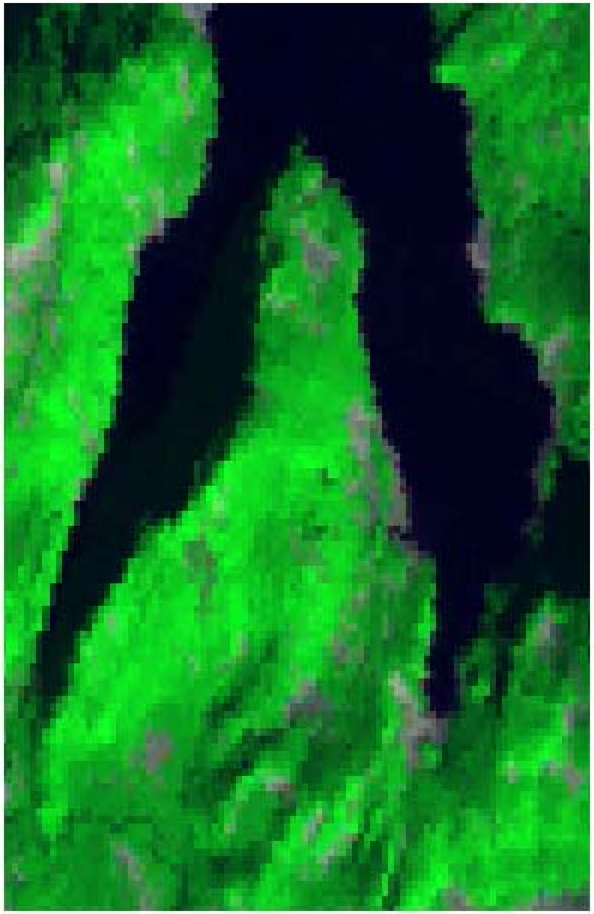
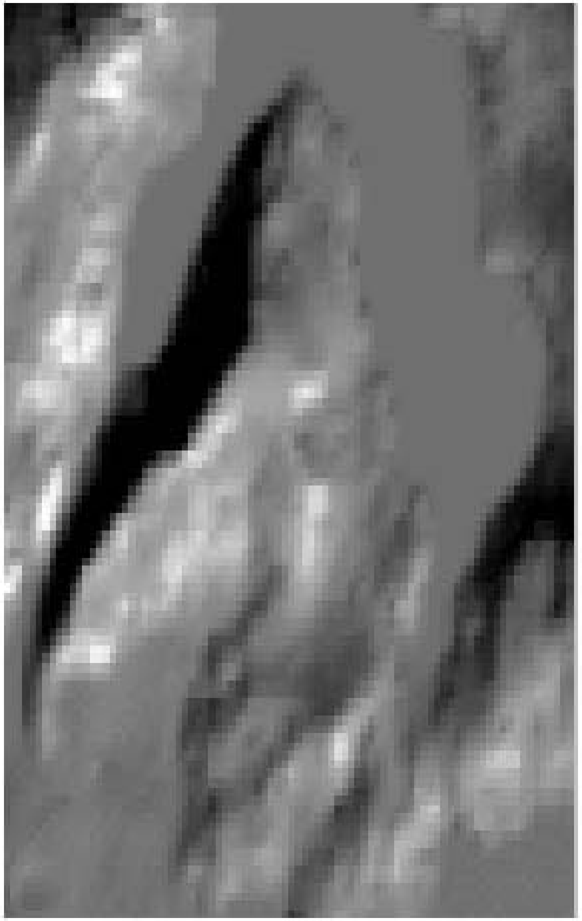
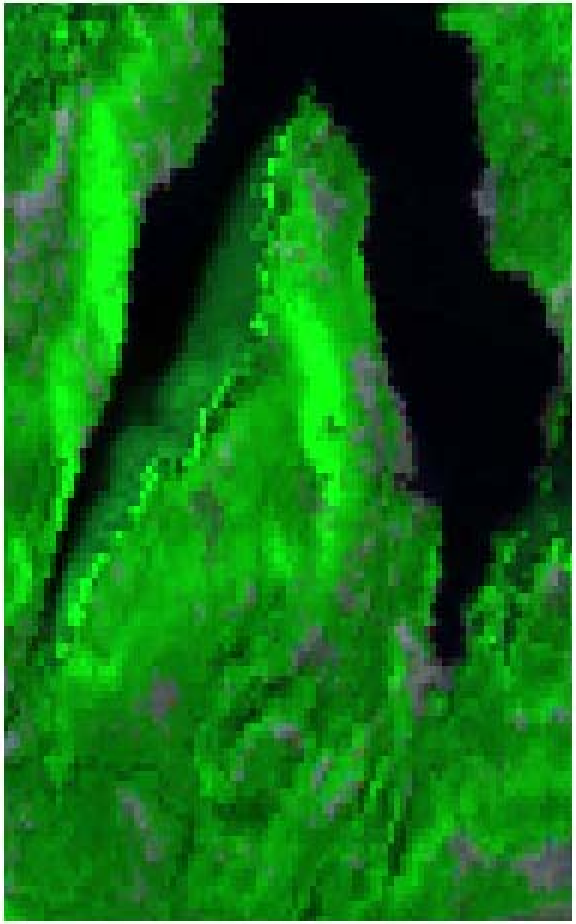
- Image processing training data collection issues
  - Provide foundation for accurate and detailed land cover mapping
    - represent diversity of land cover
    - represent area of interest
  - Reasonable cost
    - number of ‘types’
    - travel time and equipment
    - number of samples
  - Data collection window of opportunity

Cost/Time Reduction

Goal is to reduce the number of data collection sites while still describing the diversity and geographic range of the project area

- Two ways ….
  - Illumination correction
  - Classification training site selection methodology

Illumination Correction

Reduce Sampling Effort

- Eliminate collection of erroneous data
- Eliminate collection of redundant data

Training Area Characteristics

- Spectrally homogeneous and normally distributed (may be floristically heterogeneous !)
- Accessible
- Large enough
  - for an adequate sample
  - to locate in the field
  - to distinguish from neighboring spectral types

The Norm

- Overview project area
- Visually select sites
- Visit and collect data
- Build the training sample set as you go

The Problem(s) with the Norm

- Incorporate 'bad" data into process
  - Visual acceptance rather than spectral
  - Group sites by categorical values
  - Sample non-normally distributed spectral data
- May leave out data
  - Leave spectral holes in the training set

A different approach ….

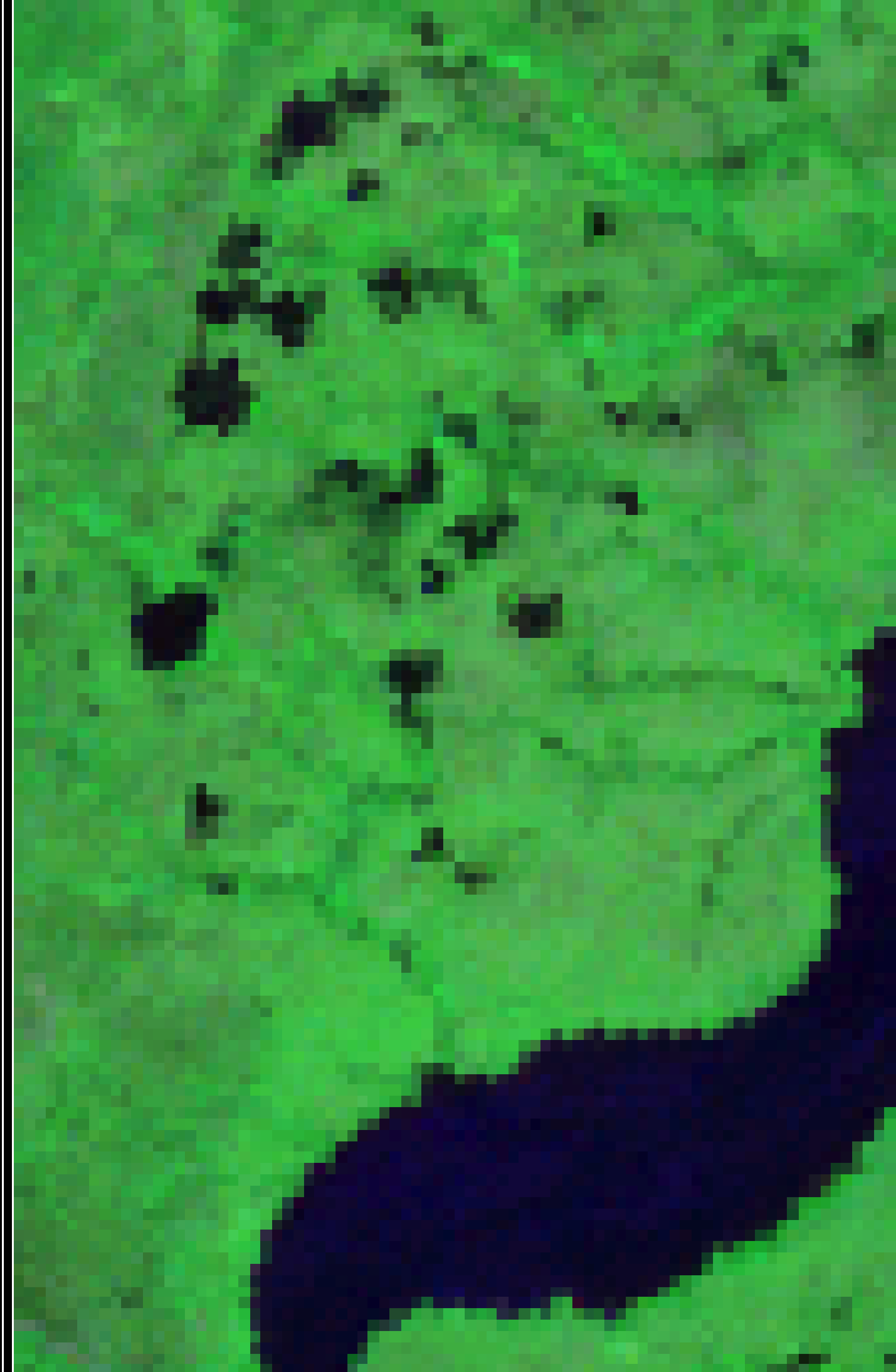
Let’s use the data and our Image Processing and GIS tools to guide and direct our data collection efforts, staying away from invalid sites and focusing on those sites necessary to build an accurate training data set that represents the range of land cover types over the entire project area.

GRS Sampling Methodology

- Image Stratification
- Candidate Site Database Development
- Candidate Site Refinement
- Sample Plan Development and Administration

Image Stratification

- Use unsupervised classification methods to generate spectrally homogeneous classes
  - Identify diversity of the project area
  - Identify area/frequency and relative magnitude of 'types'
  - Break project area into sub-regions (NDVI) or ecotypes to increase diversity of classes



Class Area and Relative Magnitude

```
     Histogram      d:\mgeprojects\npsa03\grd\7219cls.grd
  Value Frequency       %  Cum. %   Area (sq_m)     (Each * represents 1%)
---------------------------------------------------------------------------
............
  13001      97827    1.22    1.22   88044300.0        *
  13002       8525    0.01    1.23    7672500.0        *
  13003     188309    2.35    3.58  169478100.0        ****
  13004     242908    3.00    6.58  218617200.0        ***
  13005     205191    2.56    9.14  184671900.0        ***
  13006     221868    2.77   11.91  199681200.0        ***
  13007     354165    4.43   16.34  318748500.0        ****
  13008      34564    0.43   16.77   31107600.0
  13009     307886    3.85   20.62  277097400.0        ****
  13010     236191    2.95   23.57  212571900.0        ***
  13011     187121    2.34   25.91  168408900.0        **
  13012      66805    0.84   26.75   60124500.0        *
  13013     147286    1.84   28.59  132557400.0        **
  13014     181647    2.27   30.86  163482300.0        **
  13015     199983    2.50   33.36  179984700.0        **
  13016      85332    1.07   34.43   76798800.0        *
  13017     130294    1.63   36.05  117264600.0        **
............
```

Isodata Classmap Database

id	iso_class	#pixels
...		
24971	13024	14
24972	13003	1
24973	13020	1
24974	13021	1
24975	13003	3
24976	13009	134
24977	13024	3
24978	13003	9
24979	13007	2
24980	13010	1
24981	13010	12
24982	13024	1
24983	13019	5
24984	13010	3
24985	13010	1
24986	13024	70
24987	13027	1
24988	13011	1
...		

Candidate Training Site Database Development

- Apply minimum size limit of 60 pixels or 13 acres to the area listing and create a new set of candidate training site locations

select id, iso_class from grid_val where pix_count >= 60

Reduced 8.6 million ‘areas’ to 36,833 areas

Characterize Candidate Sites - Frequency by Class

iso_class	freq	pixels	ave_size
13001	56	5699	101
13002	2	8497	4248
13003	175	22232	127
13004	96	10802	112
13005	44	4262	96
13006	64	6561	102
13007	428	73239	171
13008	87	17090	196
13009	393	77351	196
13010	278	37048	133
13011	90	9730	108
13012	25	60551	2422
13013	176	27261	154
13014	130	16639	127
13015	10	192441	19244
13016	104	18261	175
13017	138	19150	138
13018	20	1809	90
13019	56	6002	107
13020	148	24548	165

Sample Site Selection

- Determine missing or rare classes

  select distinct iso_class from grid_val where iso_class not in

  (select distinct iso_class from candidate_trsite)

  - identified 0 missing isodata classes

  select iso_class,count(*) from candidate_trsite group by iso_class

  having count(*) < 5 order by iso_class

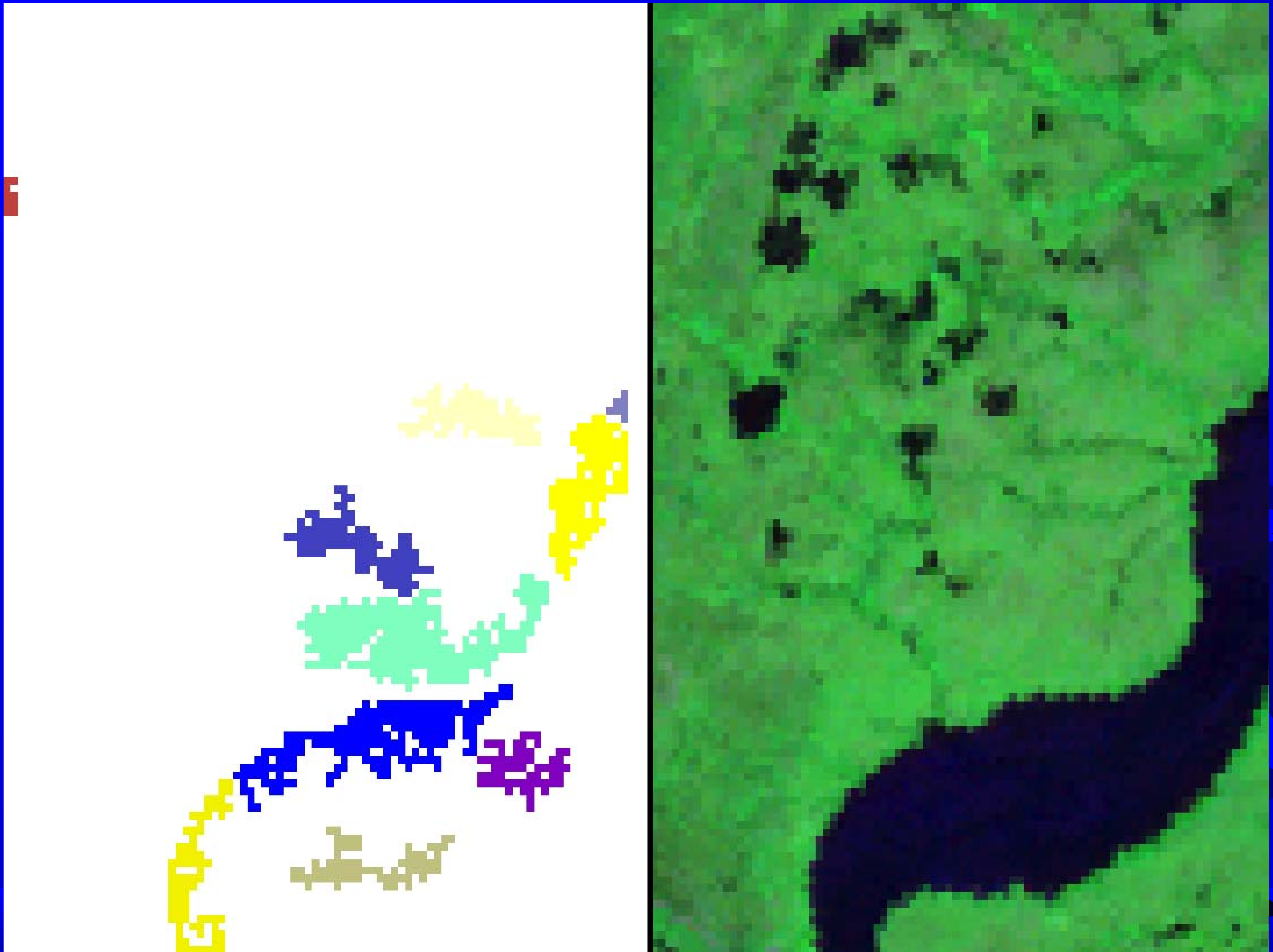
  - identified 42 scarce isodata classes

- Add additional candidate areas to supplement scarce classes by lowering minimum size limit to 45 pixels or 10 acres
  - added 305 sites to these 42 classes

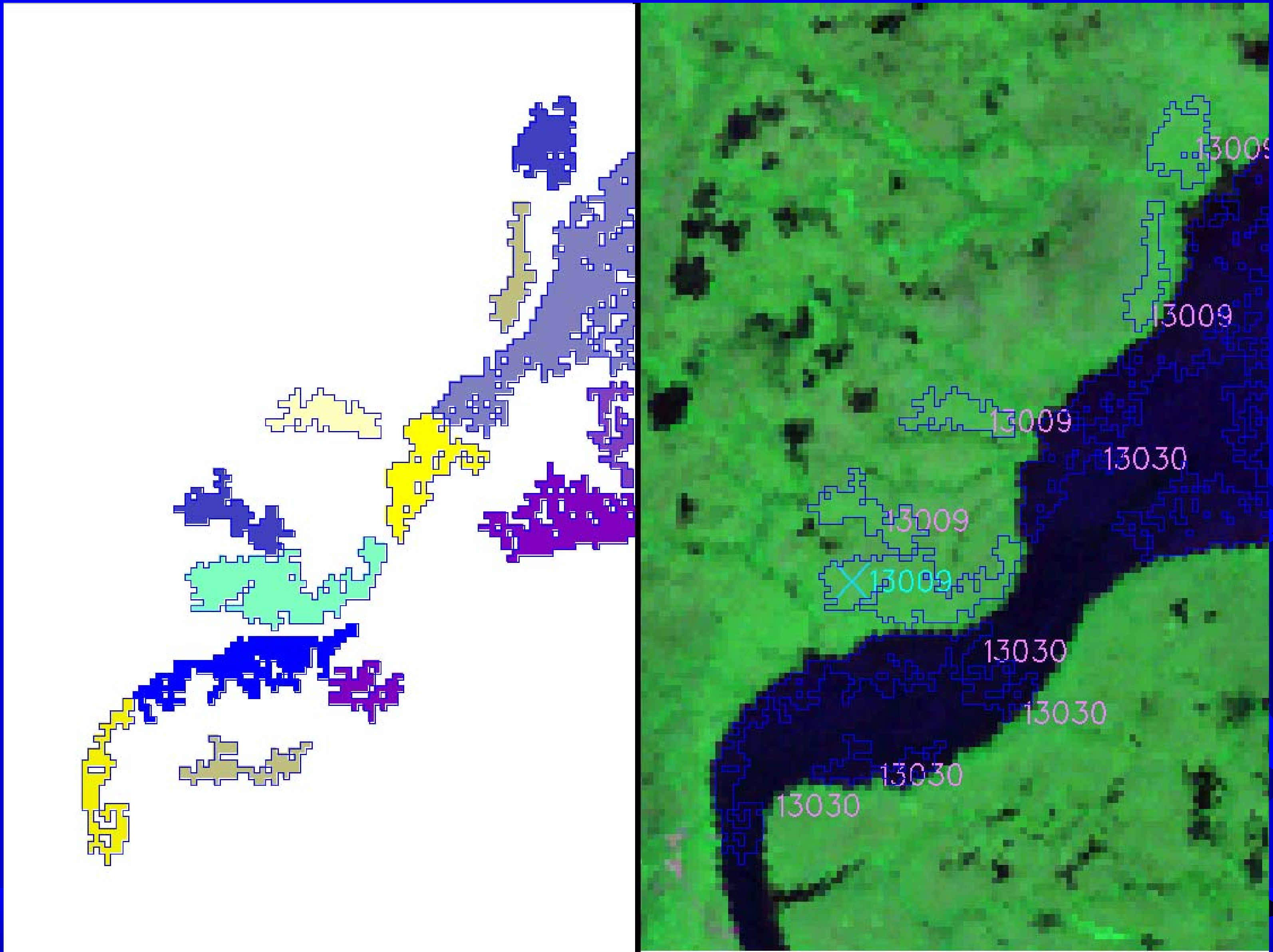
Generate GIS Database

- Reclass all pixels of groups with size/area less than the specified minimum size(s) to a value of 0
- Vectorize the remaining pixel groups and relate to unique area number

Reclass Areas Too Small to '0'



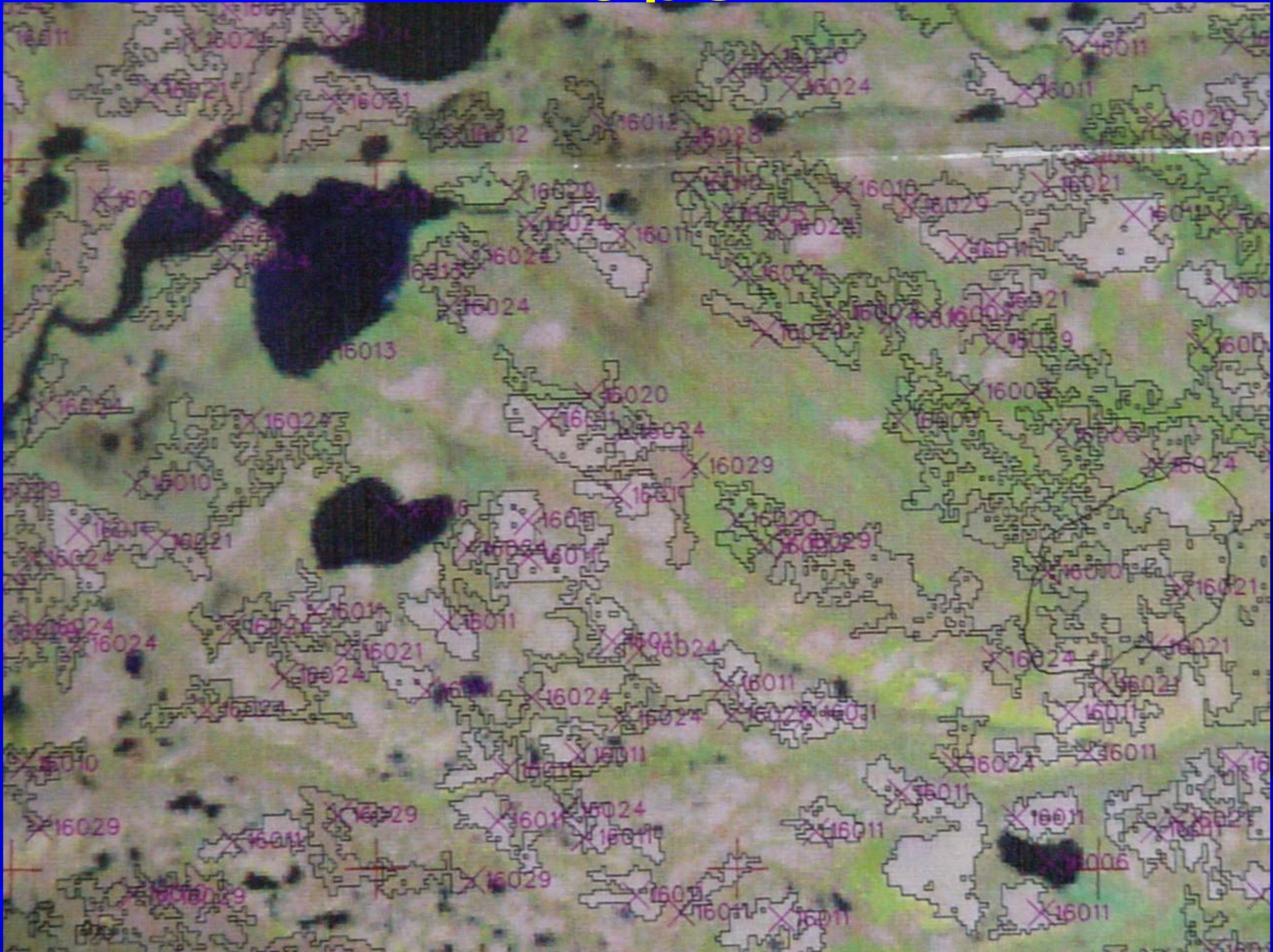
Vectorize and Label Candidate Areas



Candidate Training Site Database Contains ...

- Isodata class value
- X,Y coordinates
- Area - number of pixels
- Slope, aspect, and elevation
- Scene indicator
- Scarcity flag
- Sampling status
- Group/vicinity value

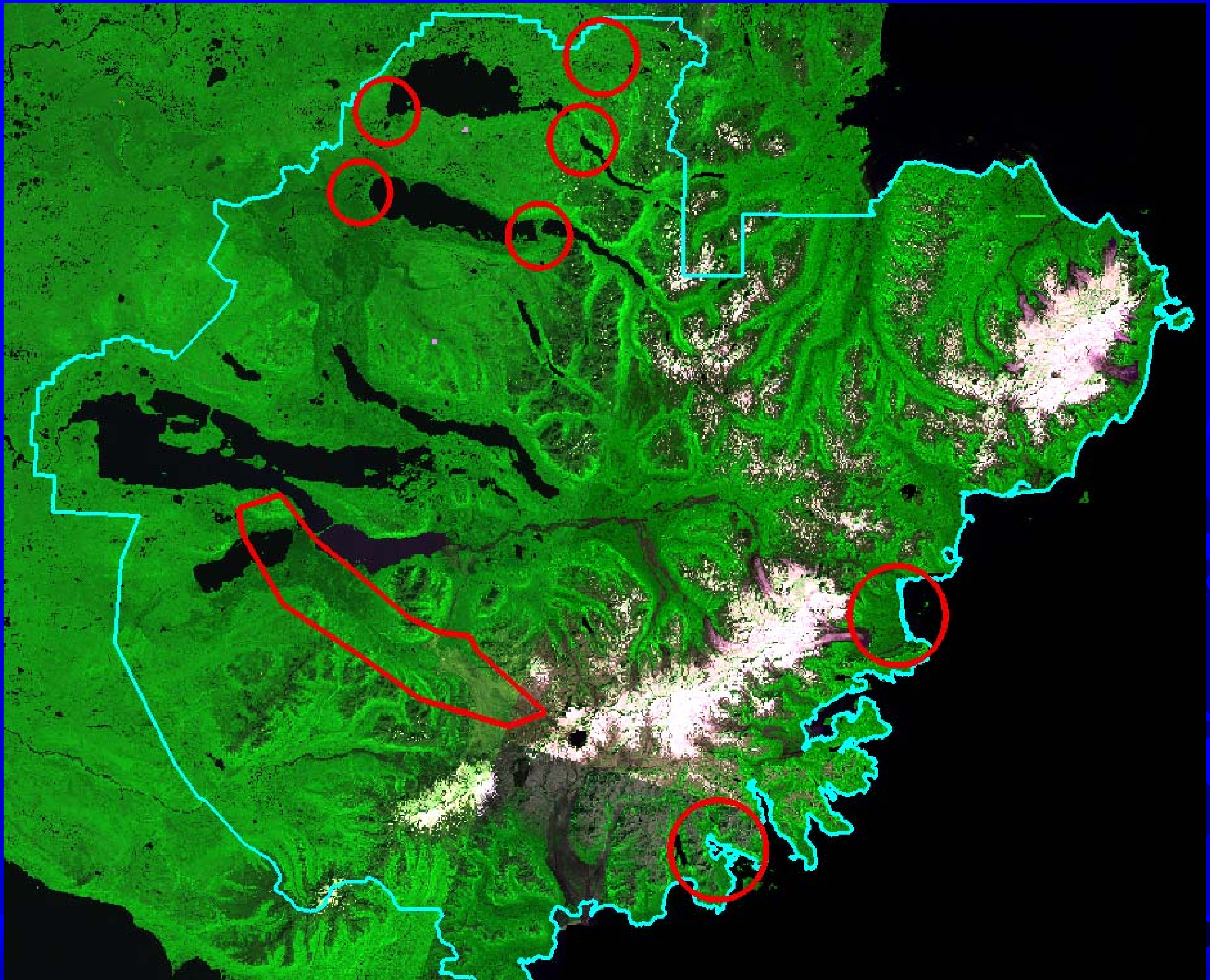
Generate Plots and Field Maps



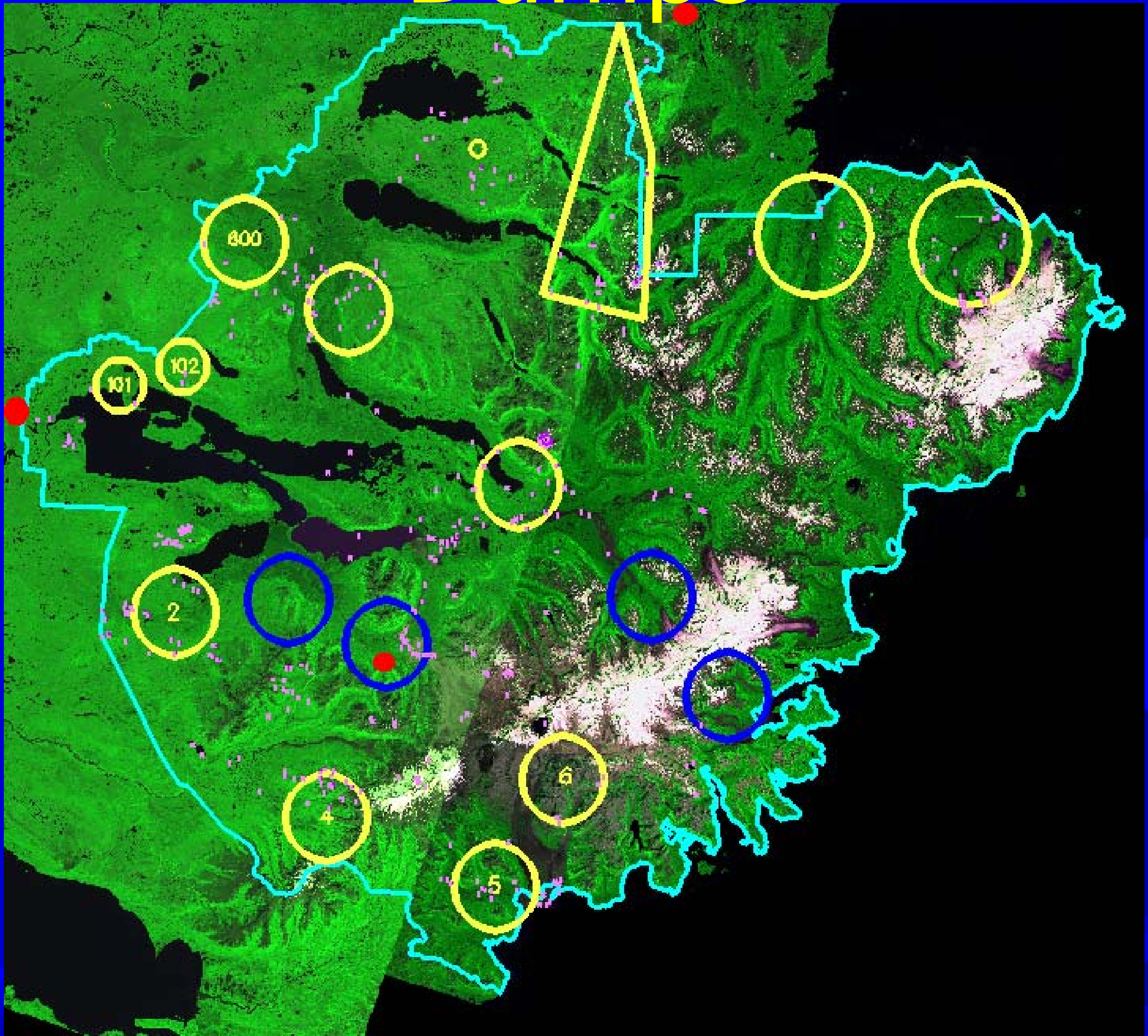
Candidate Site Selection Criteria

- Access
- Distance traveled
- Scarce isodata classes
- Proximity of candidate training sites to each other
- Overlap areas/number of scenes

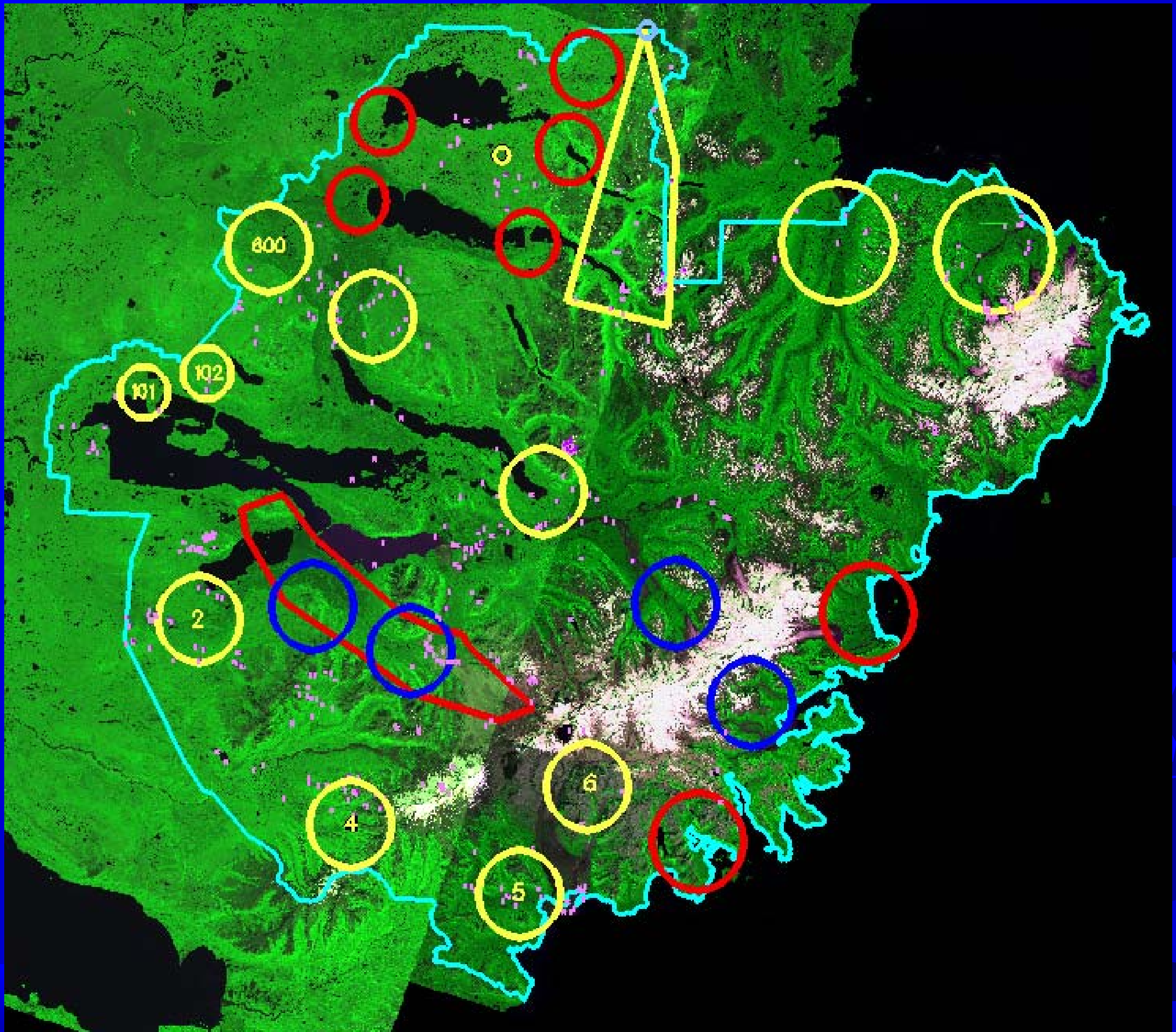
'No-Fly' Zones



Areas of Interest/Fuel Dumps



‘No-Fly’ Zones AND AOIs



Sampling Plan Development and Administration

- Daily Plan Development
  - fulfill sampling needs of scarce isodata classes
  - fulfill daily plan
  - fulfill overall plan requirements
  - multi-scene samples
- Field Maps
- Upload target sites to GPS
- Monitor progress

Area Candidate Site Report

iso_class	tr_group	#pixels	count
13001	600	1122	9
13003	600	4294	41
13005	600	216	2
13006	600	248	3
13007	600	1000	9
13008	600	934	8
13009	600	516	6
13010	600	8486	59
13011	600	1779	17
13013	600	2384	23
13014	600	368	4
13016	600	1216	13
13017	600	2589	22
13018	600	350	4
13019	600	211	3
13020	600	2227	16
13021	600	1190	11
13023	600	5337	40
13024	600	342	5

Daily Plan Report

tr_group	trsite_id	iso_class	lat	lat_min83	long	long_min83	aspect	slope	elevft	map
206		13006	58	39.28299	-156	23.015442	8	2	144	D 2
206		13023	58	38.98407	-156	23.386230	172	2	177	D 2
1100		16007	59	2.6213837	-155	25.258484	200	7	928	B 5
1100		16010	59	6.1978912	-155	20.719299	270	2	1066	B 5
1100		16028	59	4.6220398	-155	22.118225	288	1	1239	B 5
1100		16003	59	7.8076172	-155	13.487549	355	8	826	B 6
1100		16011	59	11.286163	-155	9.3164063	0	0	820	B 6
1100		16019	59	14.15657	-155	6.8536377	306	3	1246	B 6
1100		16020	59	11.508865	-155	7.1667480	207	3	862	B 6
1100		16021	59	14.164581	-155	8.3312988	153	1	1164	B 6
1100		16024	59	13.615494	-155	5.7000732	113	2	1289	B 6
1100		16024	59	6.322403	-155	19.465942	270	1	1239	B 6
1100		16027	59	13.79631	-155	6.2145996	104	3	1348	B 6
1100		16029	59	14.400101	-155	8.3322144	180	1	1184	B 6
1100		13049	58	45.376511	-156	2.9278564	292	2	1121	C 3
1100		6001	58	53.482819	-155	43.914185	45	1	259	C 4
1100		6002	58	48.600311	-155	41.022949	0	0	108	C 4
1100		6004	58	59.102097	-155	46.408081	180	1	495	C 4
1100		6005	58	55.098724	-155	43.585510	225	1	206	C 4
1100		6005	58	59.135742	-155	44.056091	153	2	531	C 4
1100		6003	58	52.462921	-155	38.461304	201	2	160	C 5
1100		6006	58	51.665497	-155	38.291931	315	1	157	C 5
1100		6007	58	51.786575	-155	40.390320	270	0	137	C 5

Plots and Field Maps

Sample Plan Status by Area

iso_class	tr_group	visit_status	#pixels	count - isos/areas
13001	600	0	738	7
13001	600	comp	384	2
13003	600	0	3992	39
13003	600	comp	302	2
13005	600	0	84	1
13005	600	comp	132	1
13006	600	0	146	2
13006	600	comp	102	1
13007	600	0	905	8
13007	600	comp	95	1
13008	600	0	829	7
13008	600	comp	105	1
13009	600	0	422	5
13009	600	comp	94	1
13010	600	0	8361	58
13010	600	comp	125	1

Sample Plan Status - Overall

iso_class	count
...	
13001	1
13003	1
13004	1
13005	1
13006	2
13007	2
13008	2
13009	1
13010	1
13011	3
13013	1
13014	1
13016	2
13018	1
13020	2
13021	1
13023	3
13024	2
13025	3

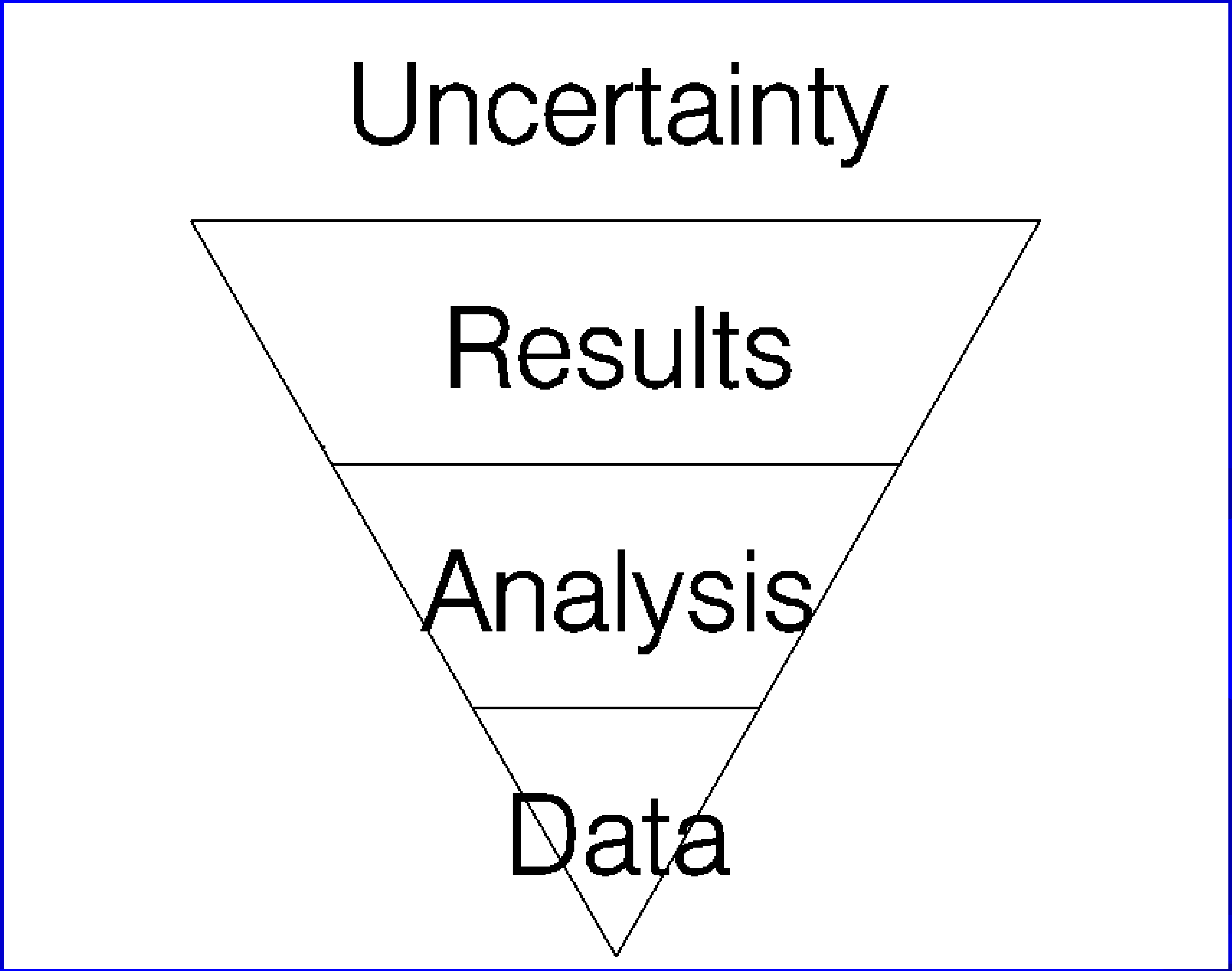
Benefits ...

- Better spectral data - less confusion
- Fewer rejected areas
- Fewer redundant samples
- Diversity has been sampled
  - Sample significant types
  - Sample scarce types
- Less speculation/seat-of-pants judgement
- Lower cost and/or less time

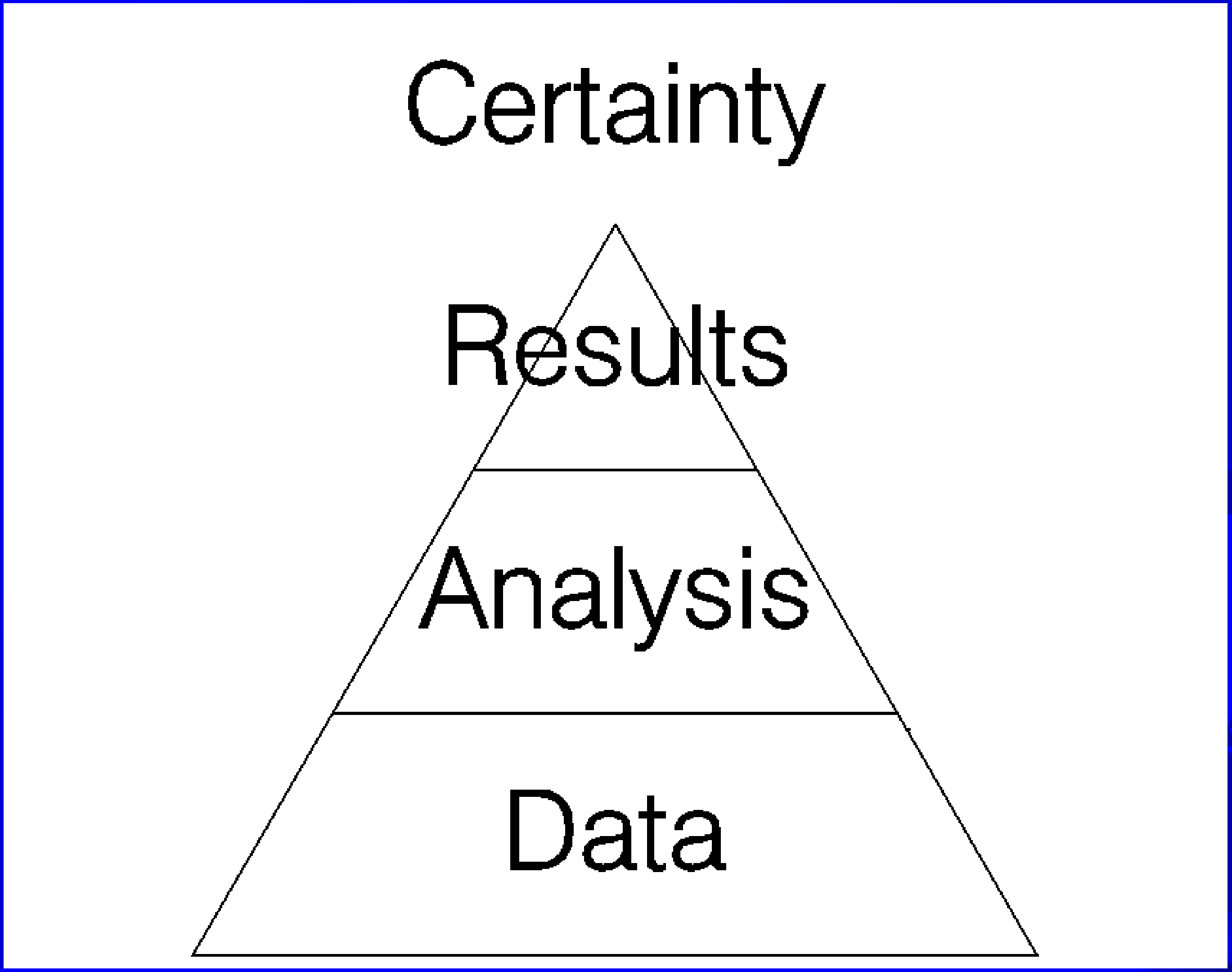
Quantitative Field Data Collection

- Field Data Collection
  - The most important part of any project!
  - The largest cost component
  - Must cover the range of land cover types
  - Do not collect any data you cannot use
  - Do not collect redundant data
- Be a Splitter not a Lumper
  - Develop and retain details from the start
  - Lump it and lose it

Project Data Model(s)



Project Data Model(s)



“Ground Truth” - Problems

- Data Sources
- Data Estimates
- Data Location

“Ground Truth” - Problems

- Use low cost data sources
  - Interpretation of aerial photography
  - Ocular estimates

Photo-Interpretation a Poor Substitute for Field Data

- Studies have proven photo-interpretation to be 50-70% accurate
- Photo-interpretation is subjective
  - Results are different between observers
  - Results may differ for the same interpreter at different times

Ocular Estimates a Poor Substitute for Measurement

- Measured data are objective and repeatable
- Ocular estimates are subjective and vary between observers
- Ocular Estimates Vs. Measured Data

  65% agreement for Canopy Closure and
  Tree Size

  75% agreement for Species Cover Type

  Sample size = 597
- Make sure you check your data

More Potential Data Problems

- Estimate categorical values that allow grouping of training data rather than detailed information that enables assignment of class values
  - Type
  - Density class
  - Size Class
- Categorical values are not easy to estimate
  - cliff thresholds for continuous variables
  - bias

More Potential Data Problems

- Wrong location - ‘true’ position in the image/data space relative to
  - mapped location or
  - GPS location

Be careful - field data collection errors are persistent and errors you make will show up over and over again !

Field Data Cost !

- Field Data accounted for 50% of total project costs from our previous mapping efforts.
- 50-60% field data costs are associated with travel.
- With the tremendous cost of acquiring field data, its worth the effort to do it right !

Field Data Collection/Estimation

- At some point you must be in touch with reality
  - Know what is there
  - Know where you are
- We should use reliable field data collection techniques that are <u>objective</u>, <u>repeatable</u>, and provide <u>quantitative</u> information consistent with our project goals

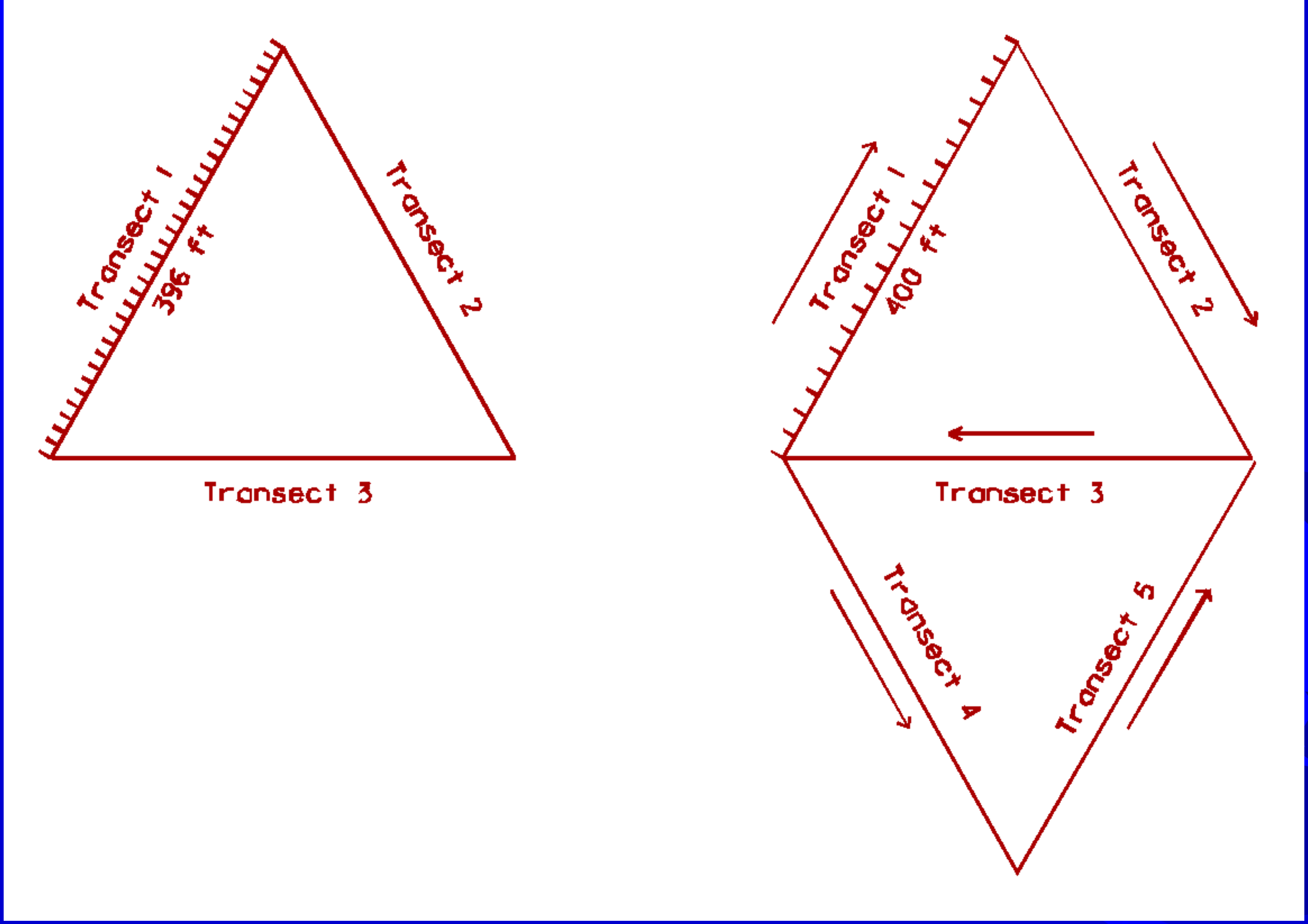
Data Collection Methods

- Preferred
  - Develop quantitative estimates
    - Transects across the landscape
    - Densitometer - estimate cover by type characteristic for different (vertical) layers of the type
- In a pinch ..
  - Ocular estimates of values by trained ("grounded") staff
- IP Analyst participates in efforts

Transect Methodology

- Sample a training area/type
  - Transect represents a sample across the type
    - record presence of different characteristics at each point along the transect
      - species or characteristic
      - size/dimension
      - height
      - crown radius
      - class/status
      - layer

Transect Configuration



Transect Methodology

- Sample a training area/type
  - Point characteristics represent a vertical sample through the type
    - record the layer of the characteristics being sampled at each point
      - top layer (bird’s-eye view from above)
      - subordinate/over-topped layer
      - pole/sapling near-ground layer
      - seedling ground layer
      - ground surface condition
  - relate features of different layers

Transect Field Data Form

VEGETATION TRANSECT

Cluster: . Transect: (1) - 2 - 3 Page: 1 of 4

Location: T 41S R8E S 5 ___ BM Poly#: TR#4

Date 9/26/94 Crew: CA/CW Azimuth: 340°, 100°, 220°

Elevation: 4200 ft Owner: FS

Pt #	SP	dbh	CD	SC	Pt #	SP	dbh	CD	SC
1	351	—	—	4	15	01	3	12	1
2	11	3	7	1	15	230	—	—	4
2	94	1	4	—	16	01	3	8	1
2	360	—	—	4	16	01	2	8	2
3	01	35	27	1	16	230	—	—	4
3	11	3	9	2	17	361	—	—	4
3	360	—	—	4	18	01	40	34	1
4	01	2		1	18	11	2	5	2
4	230	—		4	18	150	—	—	3
5	01	4	9	1	18	360	—	—	4
5	230	—	—	4	19	01	40	34	1
6	1	0	2	1	19	150	—	—	3
6	230	—	—	4	19	361	—	—	4
7	01	39	25	1	20	01	40	34	1
7	01	1	4	2	20	150	—	—	3
7	351	—	—	4	20	361	—		4
8	01	39	25	1	21	150	—	—	3
8	150	—	—	3	21	360	—	—	4
8	351	—	—	4	22	93	2	4	1
9	01	39	25	1	22	01	0	2	2
9	94	1	5	2	22	360	—	—	4
9	360	—	—	4	23	93	0	1	1
16	01	39	25	1	24	351	—	—	4
16	150	—	—	3	25	01	26	24	1
10	230	—	—	4	25	31	8	17	2
11	11	4	10	1	25	351	—	—	4
11	150	—	—	3	26	01	20	24	1
11	230	—	—	4	26	150	—	—	3
12	150	—	—	3	26	351	—	—	4
12	230	—		4	27	01	31	33	1
13	81	3	10	1	27	351	—	—	4
13	01	3	7	2	28	11	4	8	1
13	230	—	—	4	28	31	5	10	2
14	01	2	8	1	28	230	—	—	4
14	230	—	—	4	29	11	6	10	1

SP: Species/Cover Code
dbh: Diameter Breast Height, 1" size classes
CD: Crown Diameter to nearest 1'
SC: Y (yes), P (potential), or N (no) for Spectral Contribution

GRS v1.5

Transect Methodology

- Field data may be used to characterize the horizontal and vertical nature of each field sample area
  - cover matrix by layer
    - bird’s-eye view
    - understory
    - ground condition
    - …

Transect Methodology

- Data may also characterize unique features, as well as related features
  - indicator features
    - plants
    - snag(s)
    - water
    - coarse woody debris
  - isolated features
  - rare/endangered features
  - feature associations

Sample Data => Cover Matrix

Stand Cover Density Summary:
Stand: 23
Total Number of Pixels: 1

Size Class:	0-4"	5-8"	9-12"	13-16"	17-20"	21-25"	26-31"	32-47"	48"+	Tree Cover	Non-Tree Cover	Total Cover
Species												
Douglas-fir	2.0%	7.0%	4.0%	5.0%	4.0%	4.0%				26.0%		26.0%
ponderosa pine	5.0%	13.0%	6.0%	1.0%						25.0%		25.0%
cedar	3.0%	4.0%	1.0%	1.0%						9.0%		9.0%
hardwoodC	1.0%		1.0%							2.0%		2.0%
madrone	1.0%	1.0%				1.0%	5.0%			8.0%		8.0%
shrub/brush											7.0%	7.0%
prairie											5.0%	5.0%
rock											7.0%	7.0%
exposed soil											4.0%	4.0%
duff/debris											7.0%	7.0%
Total Cover	12.0%	25.0%	12.0%	7.0%	4.0%	5.0%	5.0%	0.0%	0.0%	70.0%	30.0%	100.0%
Total Tree Cover												70.0%

Stand Tree Density Summary:

Size Class:	0-4"	5- 8"	9-12"	13-16"	17-20"	21-25"	26-31"	32-47"	48"+	All Sizes
Species										
Douglas-fir	2.9%	10.0%	5.7%	7.1%	5.7%	5.7%				37.1%
ponderosa pine	7.1%	18.6%	8.6%	1.4%						35.7%
cedar	4.3%	5.7%	1.4%	1.4%						12.9%
hardwoodC	1.4%		1.4%							2.9%
madrone	1.4%	1.4%				1.4%	7.1%			11.4%
Total Tree Cover	17.1%	35.7%	17.1%	10.0%	5.7%	7.1%	7.1%	0.0%	0.0%	100.0%

Stand Quadratic Mean DBH (by Cover) Summary:

Size Class:	0-4"	5- 8"	9-12"	13-16"	17-20"	21-25"	26-31"	32-47"	48"+	All Sizes
Species										
Douglas-fir	2.1"	6.3"	10.0"	14.6"	18.0"	24.0"	0.0"	0.0"	0.0"	14.4"
	2.0pts	7.0pts	4.0pts	5.0pts	4.0pts	4.0pts	0.0pts	0.0pts	0.0pts	26.0pts
ponderosa pine	2.9"	6.3"	11.0"	14.0"	0.0"	0.0"	0.0"	0.0"	0.0"	7.7"
	5.0pts	13.0pts	6.0pts	1.0pts	0.0pts	0.0pts	0.0pts	0.0pts	0.0pts	25.0pts
cedar	4.0"	6.6"	12.0"	14.0"	0.0"	0.0"	0.0"	0.0"	0.0"	7.9"
	3.0pts	4.0pts	1.0pts	1.0pts	0.0pts	0.0pts	0.0pts	0.0pts	0.0pts	9.0pts
hardwoodC	3.0"	0.0"	11.0"	0.0"	0.0"	0.0"	0.0"	0.0"	0.0"	8.1"
	1.0pts	0.0pts	1.0pts	0.0pts	0.0pts	0.0pts	0.0pts	0.0pts	0.0pts	2.0pts
madrone	3.0"	8.0"	0.0"	0.0"	0.0"	24.0"	29.5"	0.0"	0.0"	25.0"
	1.0pts	1.0pts	0.0pts	0.0pts	0.0pts	1.0pts	5.0pts	0.0pts	0.0pts	8.0pts
QMean DBH	3.1"	6.4"	10.8"	14.4"	18.0"	24.0"	29.5"	0.0"	0.0"	13.4"
	12.0pts	25.0pts	12.0pts	7.0pts	4.0pts	5.0pts	5.0pts	0.0pts	0.0pts	70.0pts
OMean DBH - Con	3.1"	6.4"	10.8"	14.4"	18.0"	24.0"	0.0"	0.0"	0.0"	11.1"
	10.0pts	24.0pts	11.0pts	7.0pts	4.0pts	4.0pts	0.0pts	0.0pts	0.0pts	60.0pts
OMean DBH - Hwd	3.0"	8.0"	11.0"	0.0"	0.0"	24.0"	29.5"	0.0"	0.0"	22.6"
	2.0pts	1.0pts	1.0pts	0.0pts	0.0pts	1.0pts	5.0pts	0.0pts	0.0pts	10.0pts

Sample Data => Categorical Values

Categorical values are developed from the quantitative data estimates

Attributes of Land Cover - Polygons with 10Acmmu

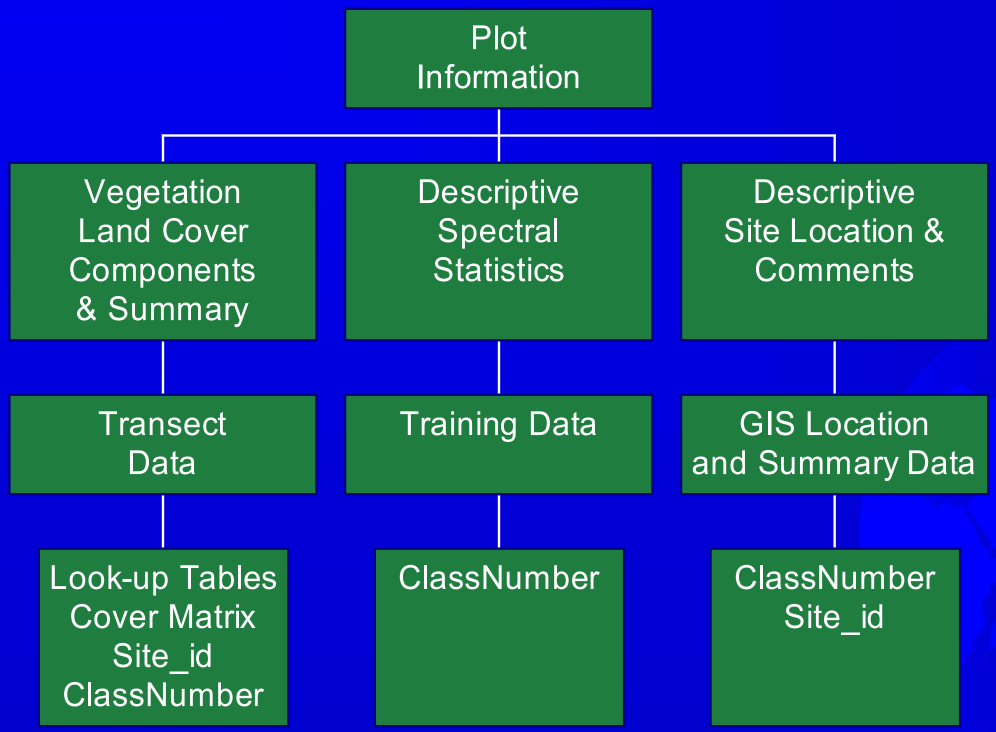
APPLE_10	VEG_TY	DENSITY	CLOSURE_CLASS	PCT_CONIFER	PCT_HDWOOD	CV_SHR	CV_HRB	CV_BAR	QMDBH	SIZE_CLASS	QMDBHCON
37	MC	88.5	8	78.1	21.9	3.9	1.5	6.1	16.5	4	17.6
43	CH	79.8	7	69.3	30.7	8.1	2.1	10.1	15.6	4	17.4
122	CH	72.4	7	45.9	54.1	13.6	1.2	12.8	14.2	4	18.4
147	CH	85.5	8	69.5	30.5	4.9	1.3	8.2	15.8	4	17.6
182	CH	70.7	7	50.2	49.8	13.8	3.4	12.2	13.7	4	16.8
117	MC	86.3	8	74.4	25.6	4.6	2	7.1	16.7	4	18.2
241	CH	64.2	6	49.6	50.4	15.6	4.8	15.5	12.7	3	15.5
306	CH	68.2	6	50.5	49.5	13.2	2.6	16	12.2	3	14.7
338	MC	87.5	8	81.3	18.7	4	1.8	6.7	16.8	4	17.8
368	MC	69.3	6	83.1	16.9	5.9	7.6	17.1	14.4	4	15
397	CH	48.8	4	45.6	54.4	25.5	6.4	19.4	10.9	3	12.9
453	MC	84	8	77.4	22.6	4.8	2.6	8.6	16.7	4	17.9
531	CH	72.3	7	61.9	38.1	9	4.7	13.9	13.9	4	15.9
543	CH	82.9	8	66.9	33.1	7	3	7.1	16.6	4	18.8
580	GF	1.7	0	0	0	7.5	81	9.8	0	0	0
603	CH	57.1	5	54.8	45.2	16.7	8.6	17.6	12	3	13.6
618	DH	45.6	4	18.6	81.4	28	11.2	15.2	10.6	3	14.6
671	MC	88.4	8	79.4	20.6	3.6	1.9	6.1	17.3	5	18.5
720	MC	61.4	6	77.6	22.4	7.9	11.5	19.1	12.4	3	13.2
756	CH	55.9	5	60.6	39.4	11.4	13.8	18.9	11.2	3	12.7

Location ? - Use GPS to Locate Field Data Sites

- Despite what the salesman told you - GPS has limitations - work within the GPS receiver’s limits
- Don’t waste time with an inaccurate receiver - use a good one
- Check GPS with reference points that are identifiable in the imagery
- Let it run … collect data all day long if you can

Storing the Field Data
A GIS Approach

Managing Field Data in a GIS



Training Set Development Traditional Approach

- Group training data into training classes that are representative of the categorical values
- Generate separate training sets for each categorical map value
  - Size
  - Cover/Density
  - Type
- Classify
- Merge individual categorical maps to form final map

Myth - One classification for each aspect of vegetation & land cover

- Many existing vegetation mapping methodologies classify the same image multiple times for individual vegetation characteristics.
  - One classification for canopy closure
  - One classification for tree size
  - One classification for species cover type
- Later during the polygon formation processes these multiple classifications are combined.

Reality - Spectral response is related to the combined influence all vegetation characteristics.

These two stands have very different spectral properties

- Douglas-fir
- **70% canopy closure**
- 32” Average Tree Diameter

- Ponderosa Pine
- **70% canopy closure**
- 8” Average Tree Diameter

GRS Approach - One Training Site is a Spectral Class

- Limit statistical range of spectral statistics – region growing
- Build training sets with many, many classes that each have small statistical variances – 2 SEs.
- Process training sites as individual classes for all categorical values
- Limit application by eco-regions
- All data available in each class map

GRS Approach …

- Katmai National Park - over 600 training classes in the final map !
- Wrangell St.-Elias National Park - over 1300 training classes in the final map !

Training Site Evaluation - are the Data Good ?

- Spectral/Confusion report
  - Good versus bad confusion
- Fidelity Report

Spectral/Confusion Analysis

CONFUSION SUMMARY FOR TRANSECT#: 10

TR#	VEG TYPE	PR SPECIES	DENSITY	QM DBH	ELEV	ASPECT	SLOPE CLASS	JM DIST
	MC	Douglas-fir	91.0%	26.3	3448	NE	C	
552	MC	Douglas-fir	77.0%	30.7	4005	E	S	1.12050
31	MC	Douglas-fir	99.0%	30.3	3776	SE	S	1.14567
572	MC	Douglas-fir	85.0%	51.7	4639	NE	M	1.27500
16	MC	Douglas-fir	80.0%	25.1	5101	N	S	1.29330

Fidelity Evaluation

- Self-classification of training area
- Comparison of land cover characteristics by training area to determine ‘match’

Fidelity Report

Scene	trsite id	Match Type	Match Type(Se	pct pure	pixel count	type	pr_comp	pr_comp cover	cover class	tree cover
7119_ypgW	80529			78.1%	178	PHw	White Spruce	24.1	Open	35.2
7219_ypg	80529			36.9%	111	PGl	White Spruce	22.3	Open	27.7
7219_ypgW	80529			36.9%	111	PGl	White Spruce	22.3	Open	27.7
TrainCalc_Data_Match	80529	M	m	55.3%						
TrainCalc'ed_032602	80529	M	m		1	PGl	White Spruce	25.0	Open	30.0
TrainCalc'ed_050102	80529	M	m		1	PGl	White Spruce	25.0	Open	30.0
TrainCalc'ed_072302	80529					White Spruce:Open				
TrainCall	80529	M	m		1	Spruce:Open	White Spruce	25.0	Open	30.0

other veg	conf cover	hdwd cover	shr cover	tsh cover	lsh cover	dsh cover	hrb cover	bar cover	oth cover	same pixels
64.6	74.3	25.7	53.0	1.0	4.9	47.2	11.6	0.0	0.3	139
70.1	82.5	17.5	52.8	0.1	6.4	46.3	17.3	2.0	0.1	41
70.1	82.5	17.5	52.8	0.1	6.4	46.3	17.3	2.0	0.1	41
70.0	83.3	16.7	60.0	0.0	5.0	55.0	10.0	0.0	0.0	1
70.0	83.3	16.7	60.0	0.0	5.0	55.0	10.0	0.0	0.0	1
70.0	83.3	16.7	60.0	0.0	5.0	55.0	10.0	0.0	0.0	1

When Can We Classify ?

When is training data

- acceptable ?
- sufficient ?

When problems have been identified and resolved ...

Hybrid Classification

Combine both supervised and unsupervised classification methods ...

Unsupervised vs. Supervised Classification

- Supervised Approach
  - Areas of interest are defined in the imagery; and spectral classes are developed for those areas.
- Unsupervised Approach
  - Statistical parameters are defined; the image is sampled; and the classification process determines the spectral classes.

Supervised - Pros

- Spectral classes are spatially homogeneous
- Vegetation characteristics are highly correlated to spectral statistics.
- May be able to develop classes with small variances

Supervised - Cons

- May not develop “clean” statistics.
- Spectral classes may or may not classify the entire project area leaving unclassified areas.

Unsupervised - Pros

- You can develop many statistically "clean" spectral classes with little effort.
- The spectral classes can be used to classify almost all of your project area.
- Relatively fast and inexpensive (quick and dirty ?).

Unsupervised - Cons

- Too easy to use
- Classes may or may not be spatially homogeneous.
- Vegetation & land cover characteristics may or may not be correlated to directly to classes.
- “How does it know?” or “it’s a black box !”
- What is it ?

A Hybrid Approach

Exploit the strengths and limit the weaknesses of supervised and unsupervised techniques

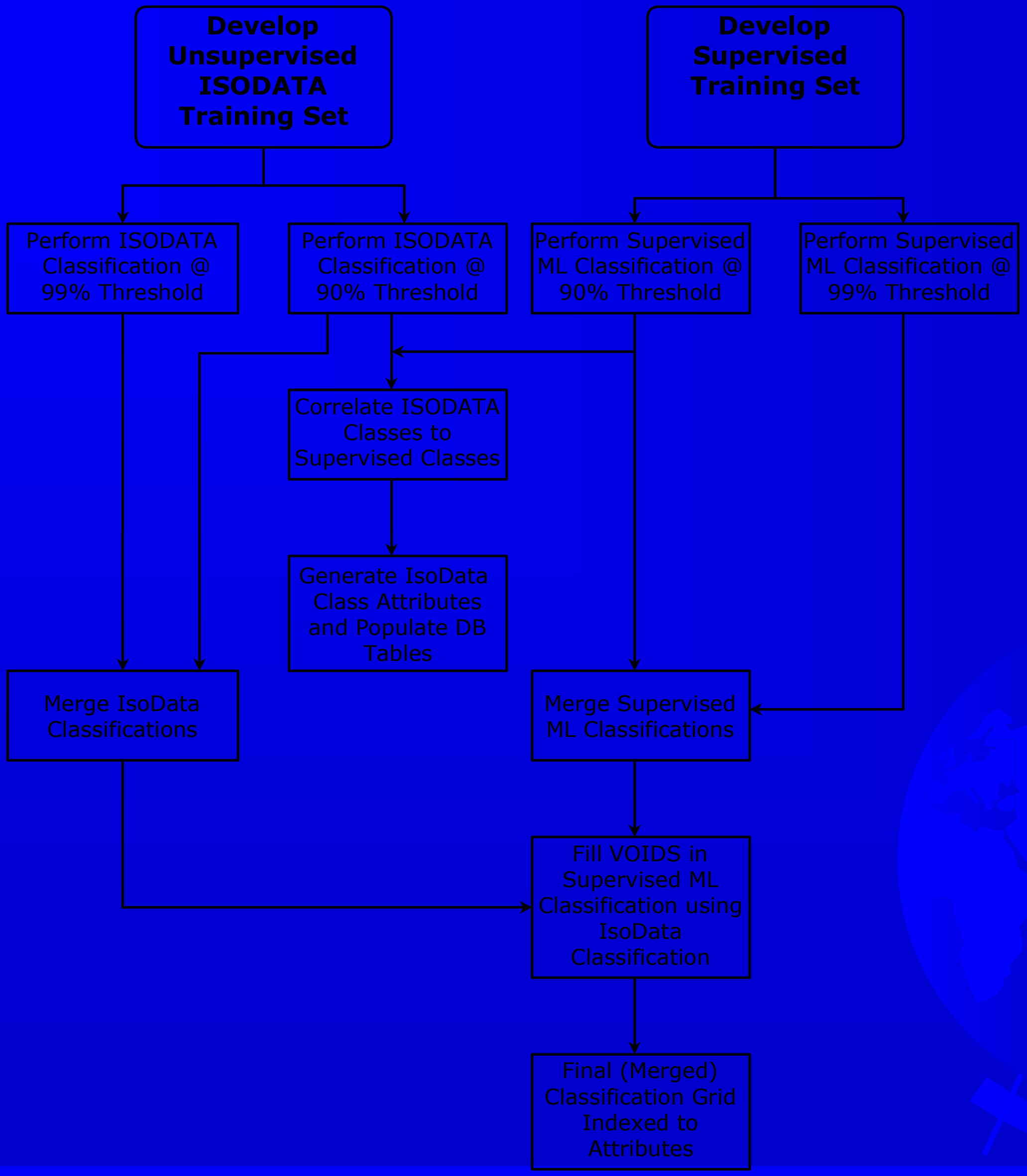
Supervised Classification

- Supervised classification
  - Used to classify 95% of the project area
  - Keyed to individual training classes
- Resulting classification pixels
  - Indexed to training site id’s
  - Indexed to specific training statistics and quantitative values

Unsupervised Classification

- Unsupervised classification
  - Used to fill-in unclassified areas
  - spectrally driven
  - Correlate unsupervised classes to supervised classes
- Resulting classification pixels
  - Indexed to training site id's
  - Indexed to derived training statistics

Hybrid Classification Workflow



Classification Results

Vectorize This

Polygon Formation

Rule-based Pixel Aggregation

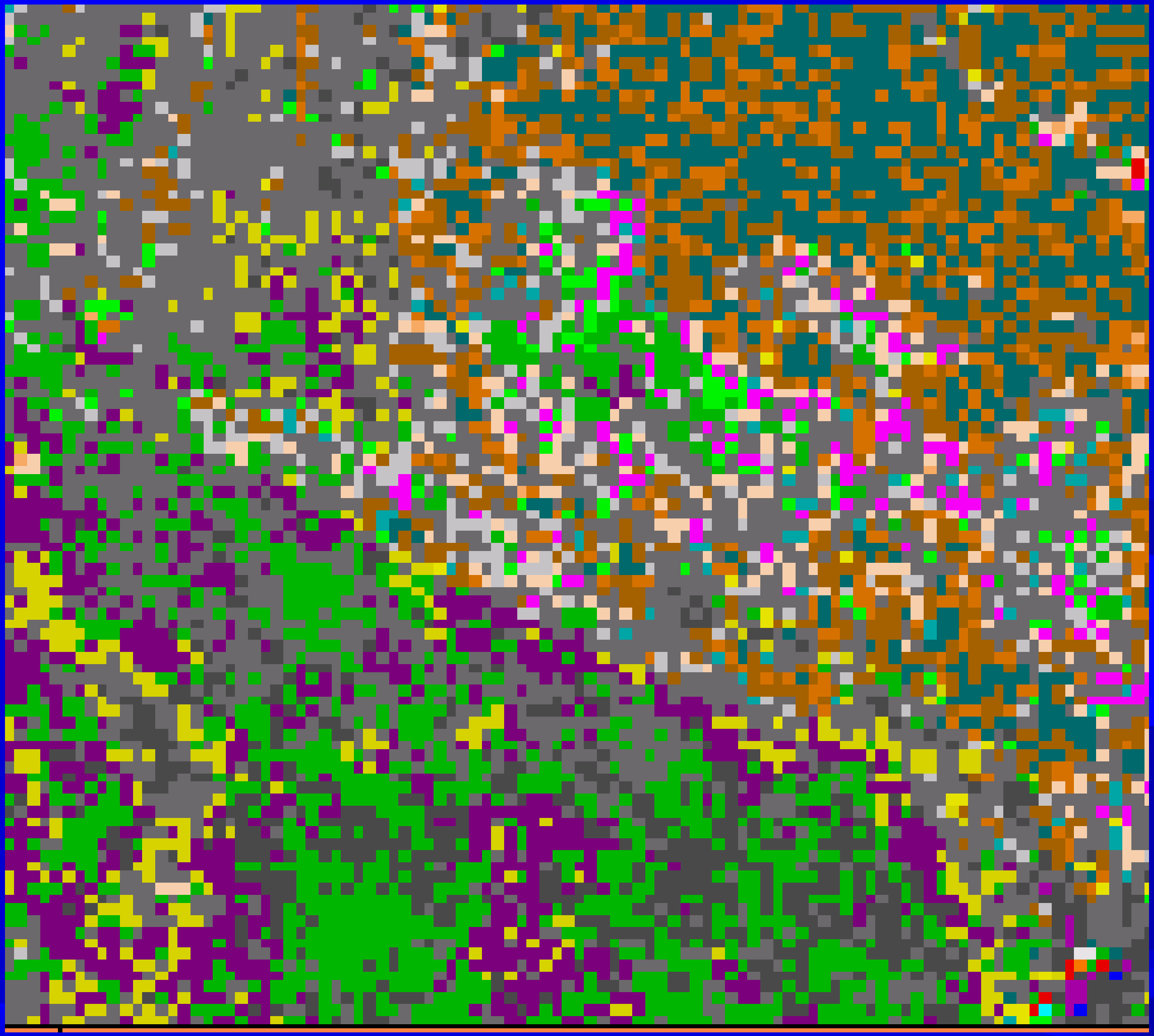
The Aggregation of Pixel Data into Mapped Area Features

The Formation of Polygons from a Pixel Classification Database

A Case Study based on the Wrangell-St Elias Mapping Project

Project Classification Results
Vectorize This !

Or how about this !



Unfortunately, most land cover is not distributed homogeneously in large areas !

Conversion to vectors is difficult.

Why don’t we just leave the data set in pixel format ??

? Image Processing Myth ?

- Pixels are just fine.
- Polygons are unnecessary -
  we can live without them!

Reality - Use Both
Pixel heterogeneity confounds the user

- Excessive information and detail - resolution is often too small for most projects
- Represents data at a level that is not manageable
  - Difficult to process and query
  - Difficult to summarize and evaluate
  - Difficult to evaluate for accuracy

Why We Like Polygons ?

- Filters/Averages Pixel Attributes
- Enable area queries
- Easier to visualize, understand, and map
- Can describe spatial relationships of types - corridors, buffers, adjacencies, and edges
- Results in ‘type’ level information that may not be present at the pixel level
- Easier to test accuracy - we can develop and use statistics at the polygon level

Polygon Formation

- Image Interpretation
- Pixel Processing
  - Segmentation
  - Smoothing
  - Aggregation

Image Interpretation

- Human views image data and draws lines
  - Subjective nature of the work …
  - Consistency of results …
  - Repeatability
  - Adaptiveness to modification

Segmentation

- Based on spectral data relationships
- Groups/clusters pixels based on user defined thresholds
- Minimum Mapping Unit (mmu) Specification ?
- You trust that spectral ‘closeness’ represents your classification objectives.

Traditional Approach to Pixel Cleaning or Polygon Formation

Filtering, Scanning, Smoothing,

and Merging

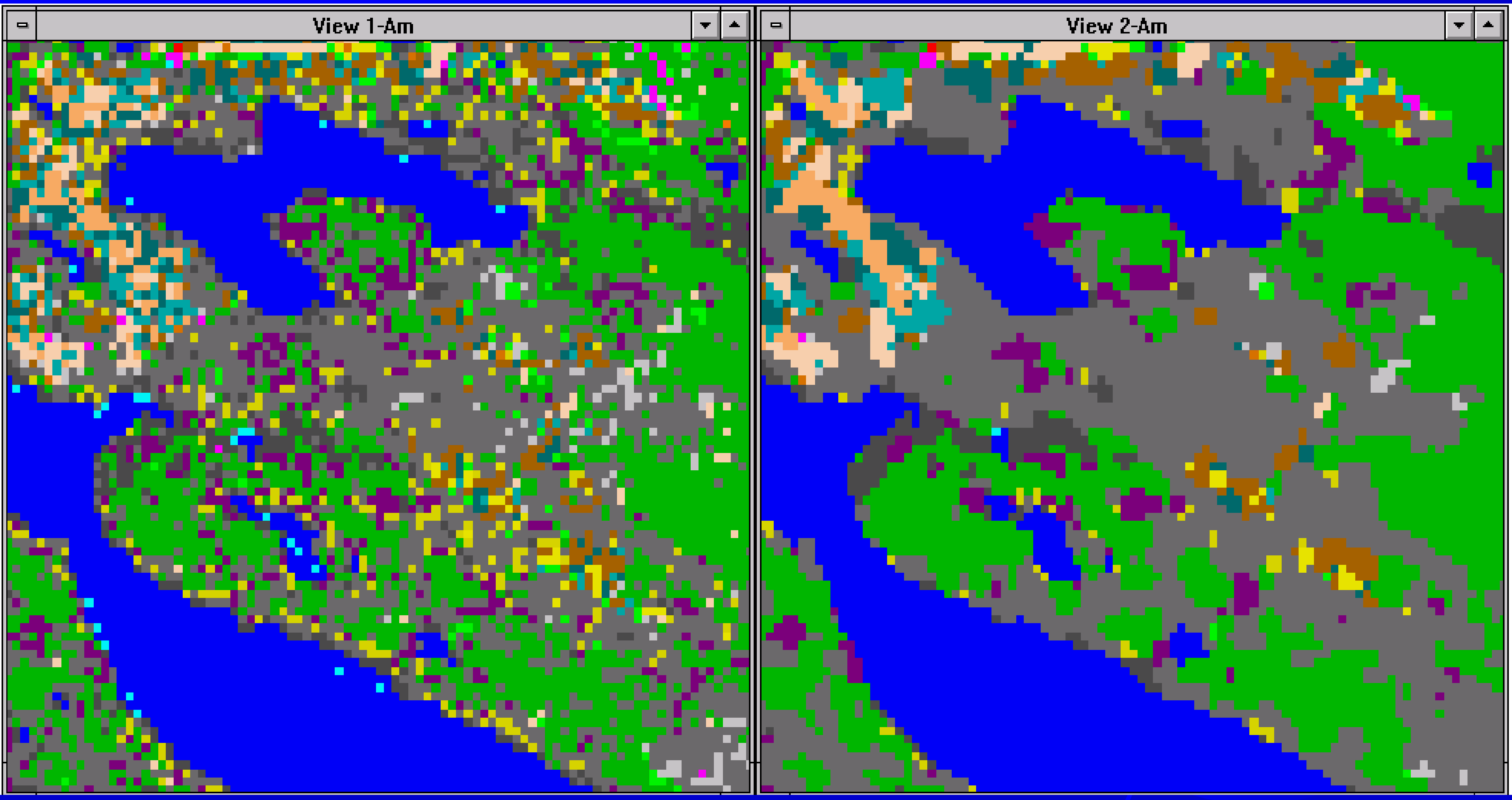
or

“How to distort your data”

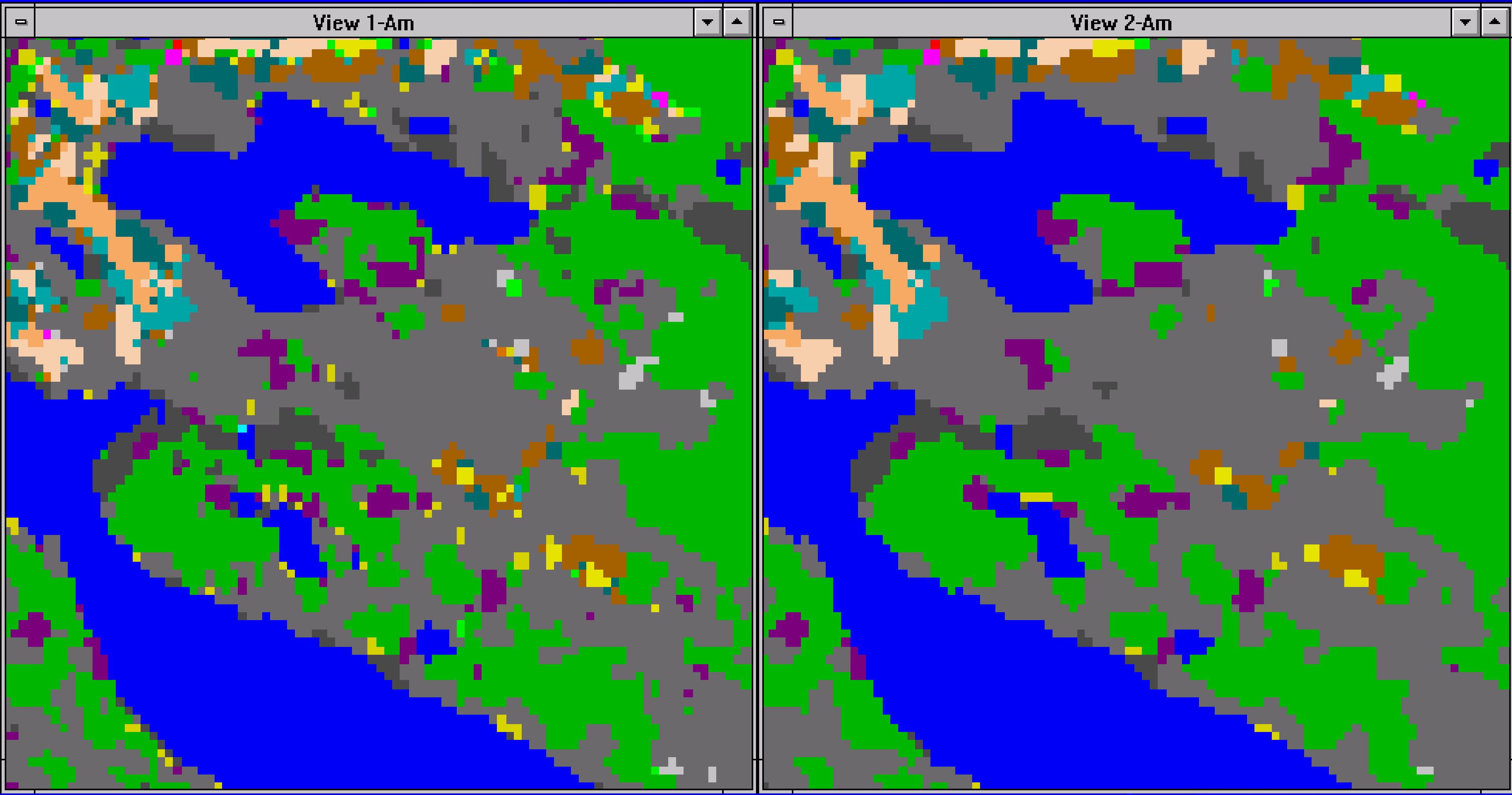
? More Myths ?

- Modal or Majority (mathematical) filters are useful tools for forming polygons
- Cleanup and develop separate themes which we can then merge to form a final land cover map

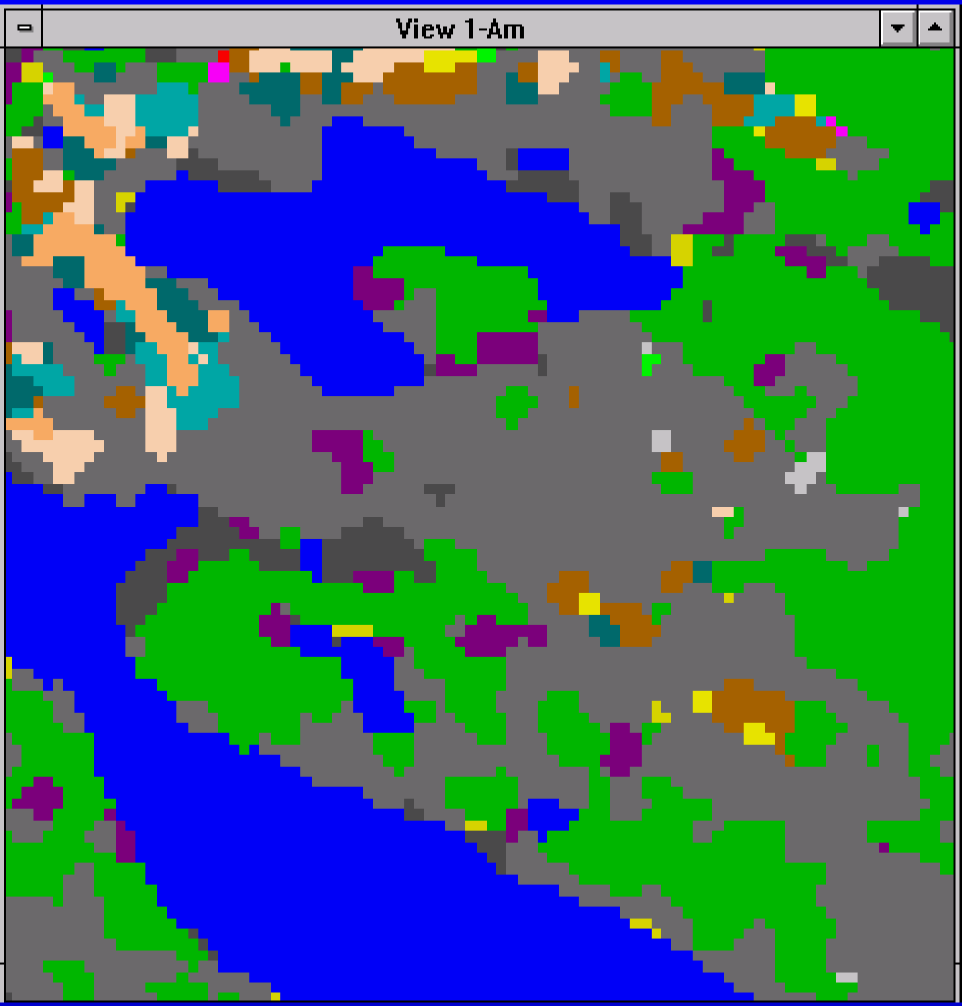
Modal Filtering - 1st Pass

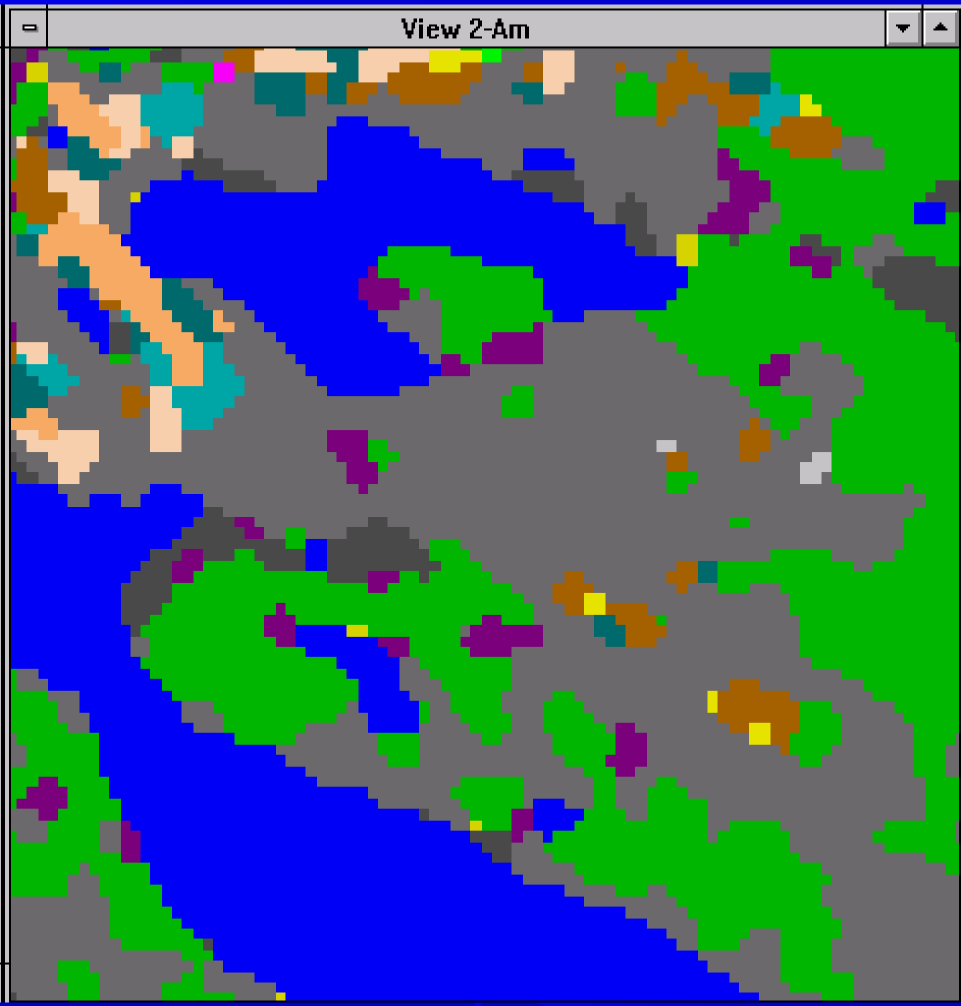


Modal Filtering - 2nd Pass

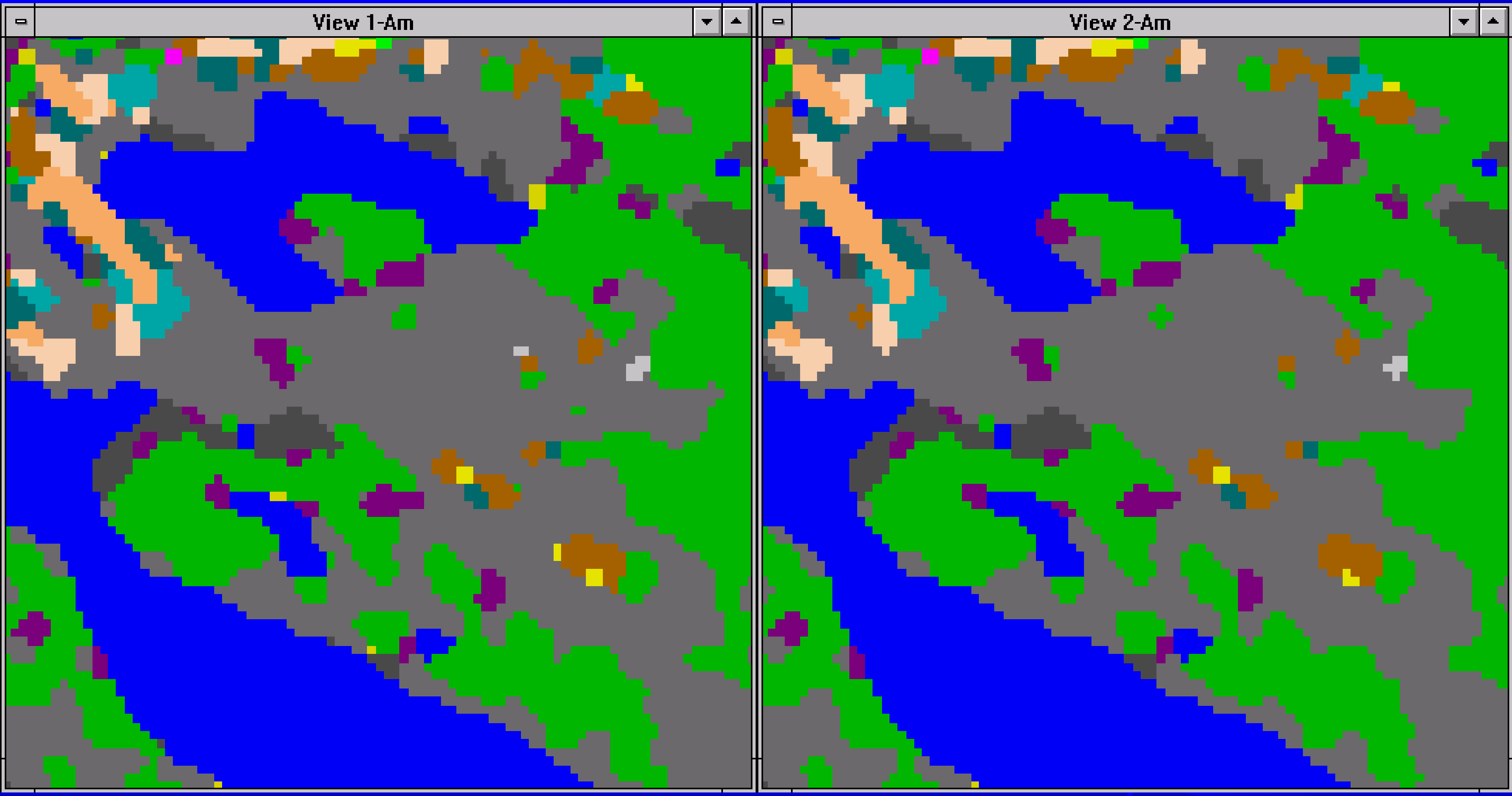


Modal Filtering - 3rd Pass

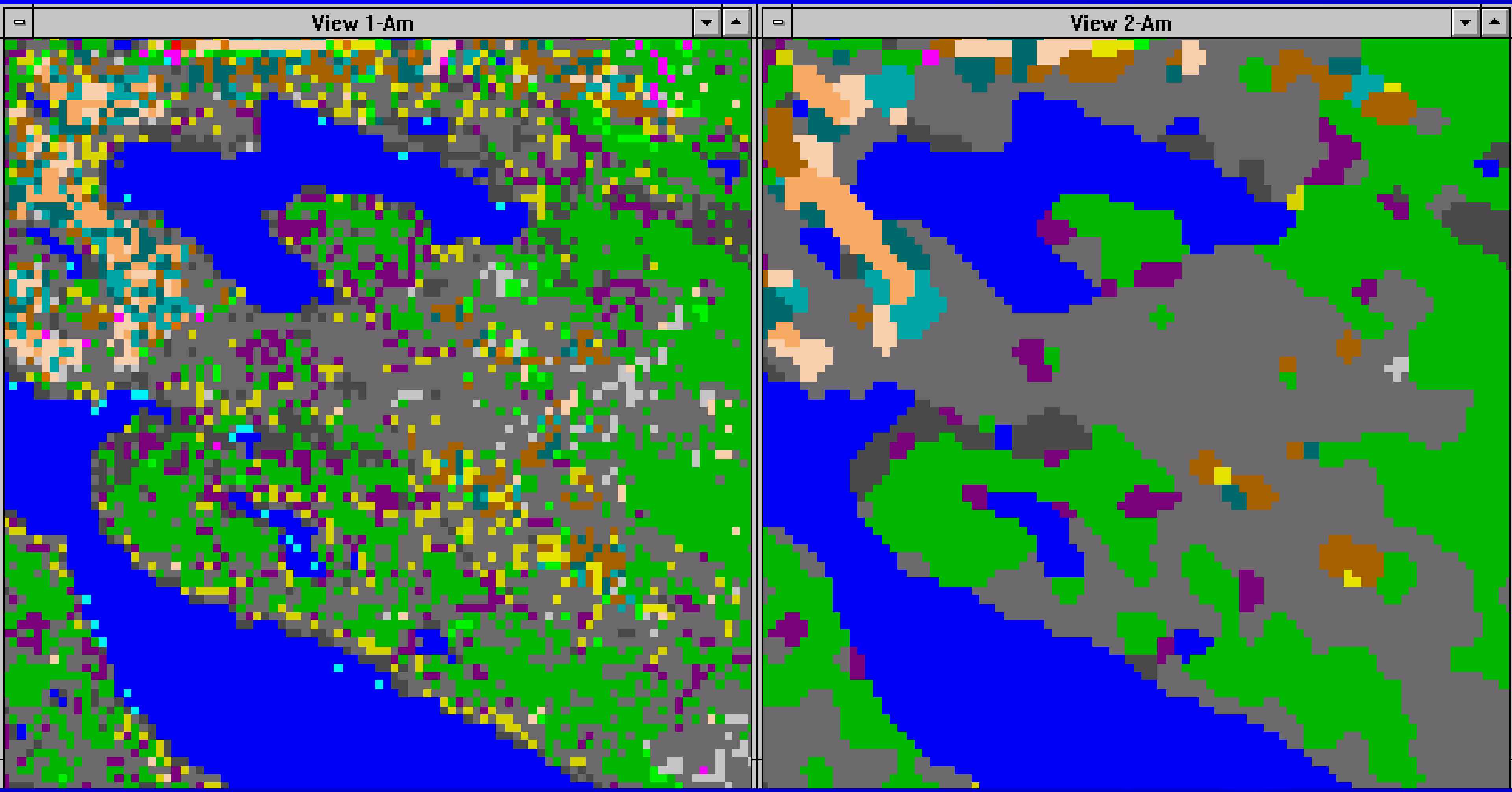




Modal Filtering - 4th Pass



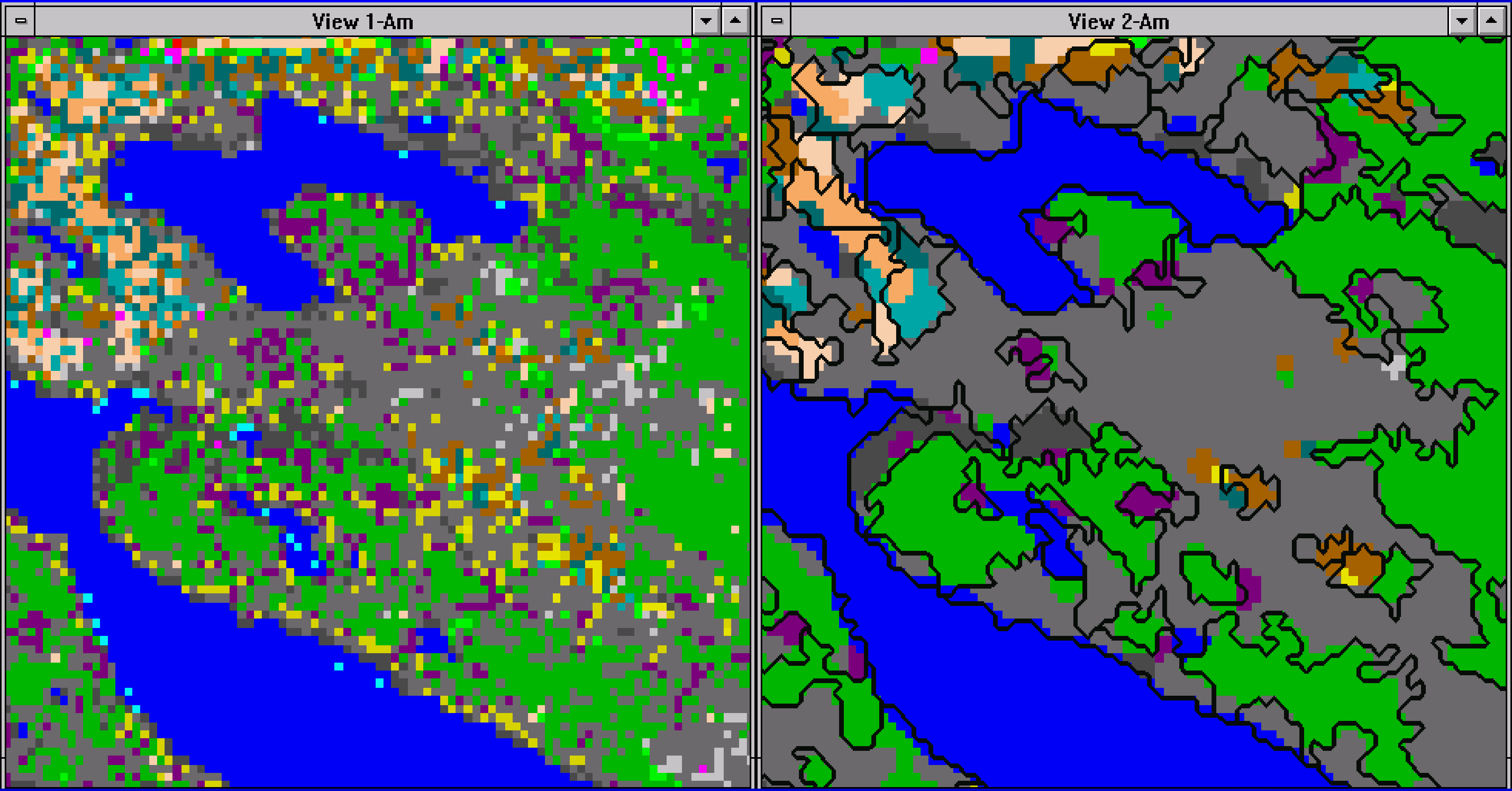
Modal Filtering - 4 Passes



Reality
Mathematical Filters Do Not Approximate Ecological Relationships and Morphological Differences

- Feast or famine solution - e.g. shrub, grassland, or tree when mixes should be developed
- Linear feature removal
- Edge degradation/creep
- Minimum size problems - “When do you know you can stop filtering?”

Filtering Problems



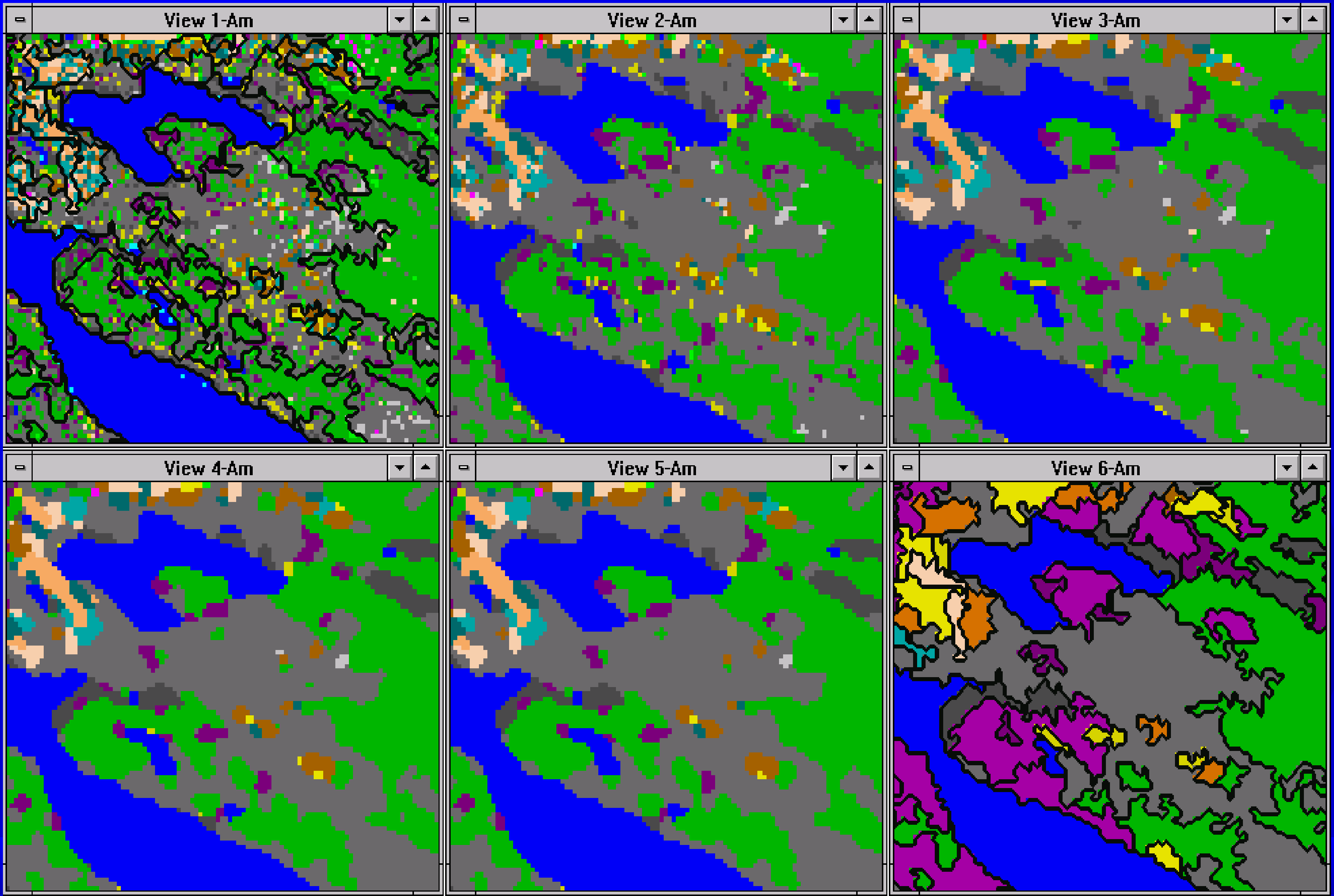
Reality - Vegetation/land cover Characteristics are Interrelated

- Should not build separate themes and merge
  - make wrong decision about type boundaries
  - massive sliver problems
- Polygon attributes must be computed as weighted averages of attributes represented by pixels, not pixel values
- Polygons may yield new types not present in the classification

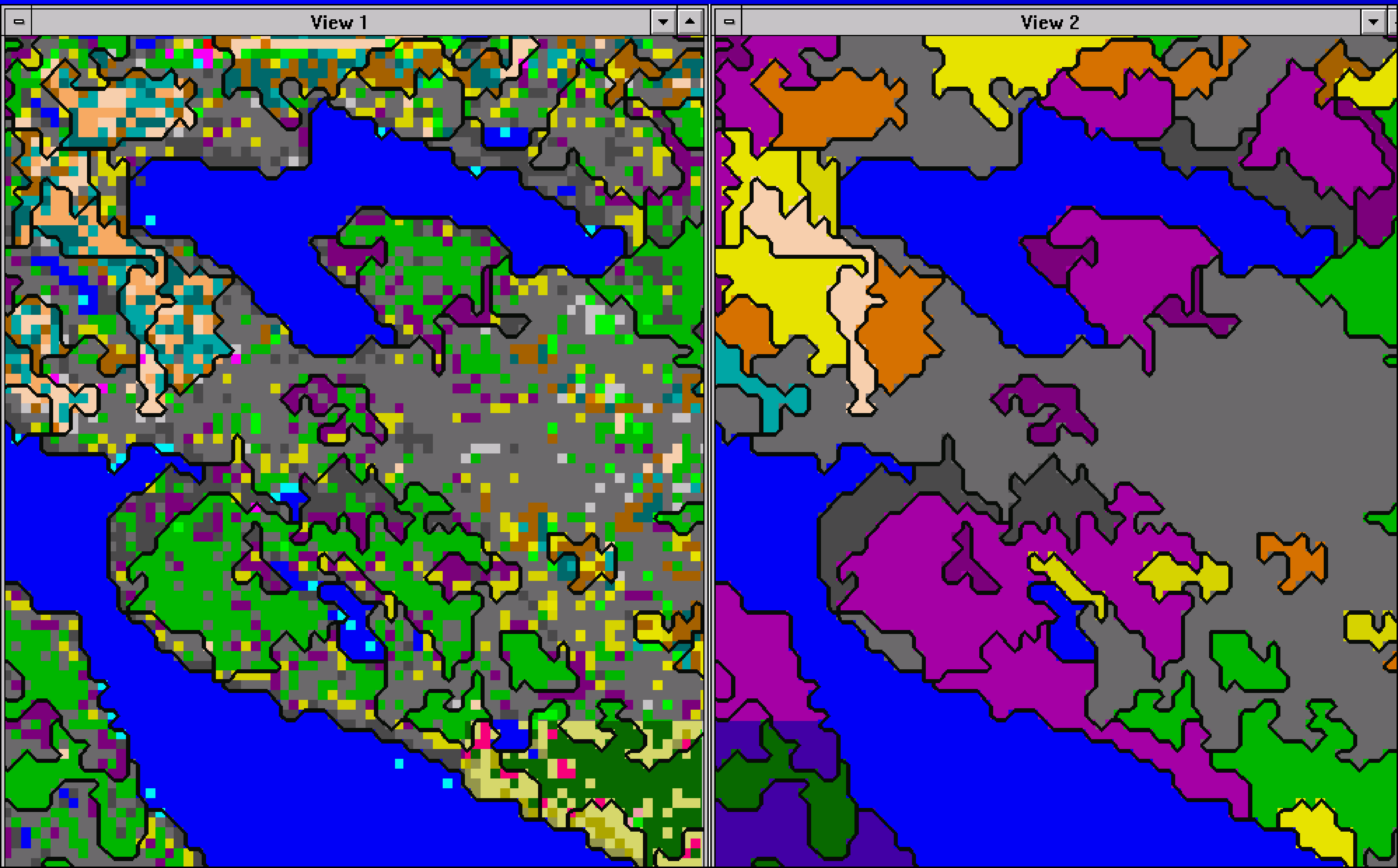
Solution: Ecological Rule-based Pixel Aggregation

- The classified pixel is a stratum in a stratification
- Each stratum represents a distinct set of cover, size, and species descriptions that are based on ground data collection efforts or other data descriptions
- Polygons are formed by grouping areas that have the most similar or related vegetation/land-cover characteristics
- Process data until all polygons meet minimum mapping unit size limits

An Alternative Solution …



… From Pixels to Polygons



Ecological Rule-Based Pixel Aggregation

- Based on:
  - Class Attribute Data (cover matrix)
  - Rules and Relationships

Step 1: Develop Data Sets Representative of Pixel Data

- Ground truth
- Classification information and associated data

Data must represent all components of all types

Class: 2 **Name: White Spruce - Open**

	% Cover	%Con/Hwd	% Species Cover
Trees: 42.50% cover comprised of:			
W Spruce	87.50%	87.50%	32.50%
B Spruce	87.50%	12.50%	4.50%
Total Conifer			37.00%
Hardwood	12.50%	100.00%	5.50%
Total Tree			42.50%

Shrubs:

Tall Shrub	45.00%
Low Shrub	5.00%
Dwarf Shrub	0.00%
Total Shrub	50.00%

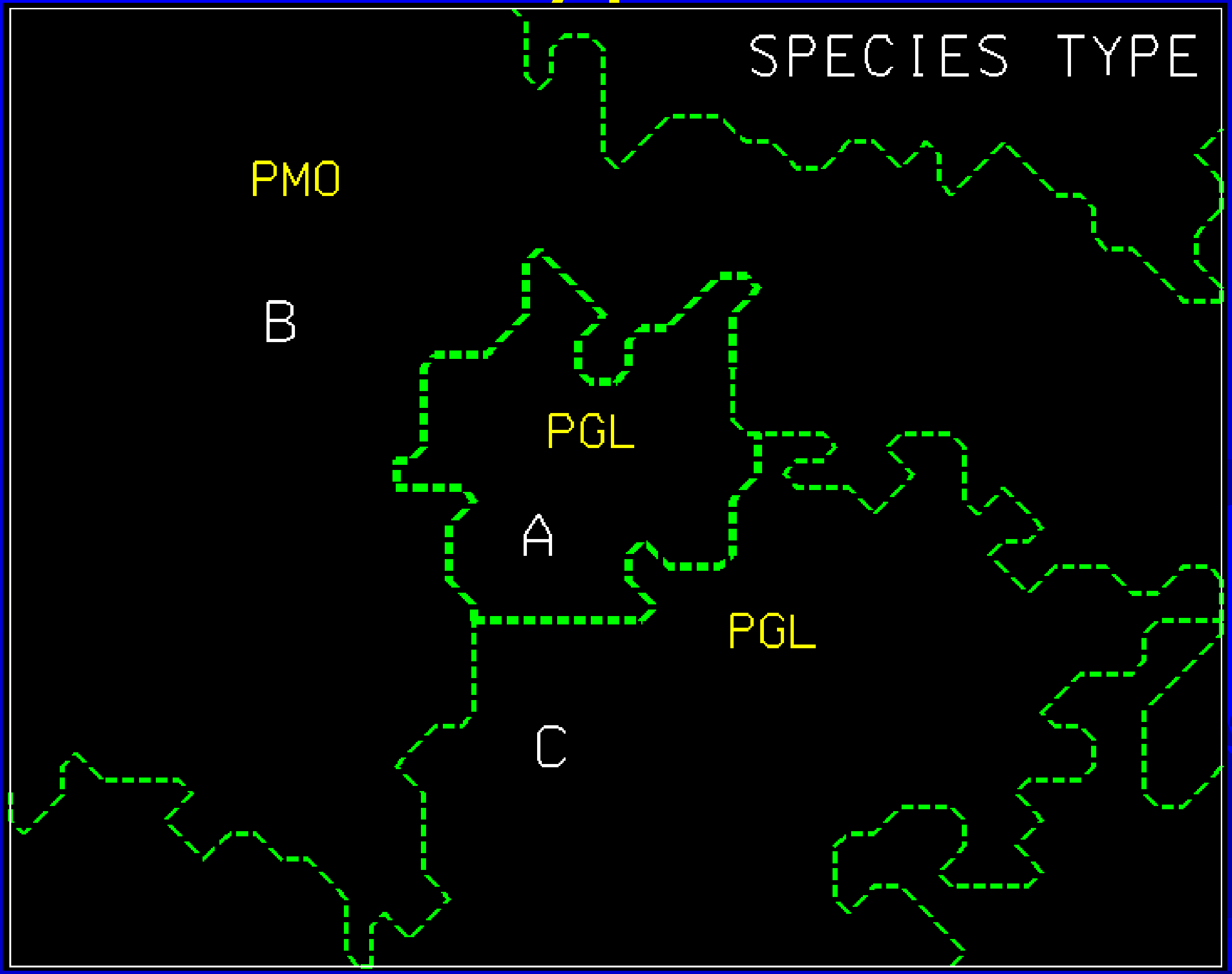
Forb:

Graminoid	0.00%
Forb	0.00%
Dry	7.50%
Wet	0.00%
Total Forb	7.50%

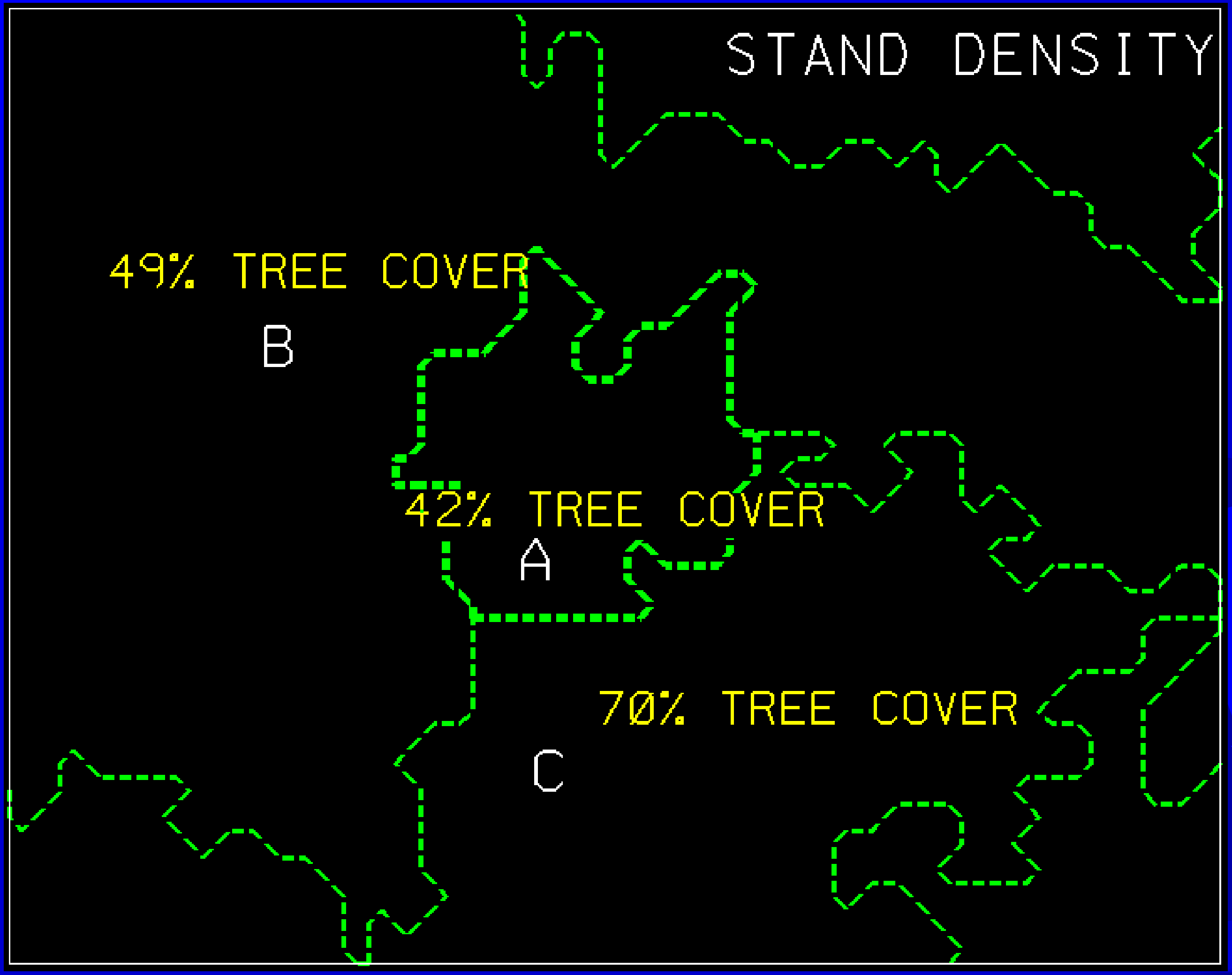
Step 2: Define Rules to Guide the Aggregation Process

- Vegetation classification definitions, relationships, and relative importance
- Minimum mapping unit size (by characteristic)

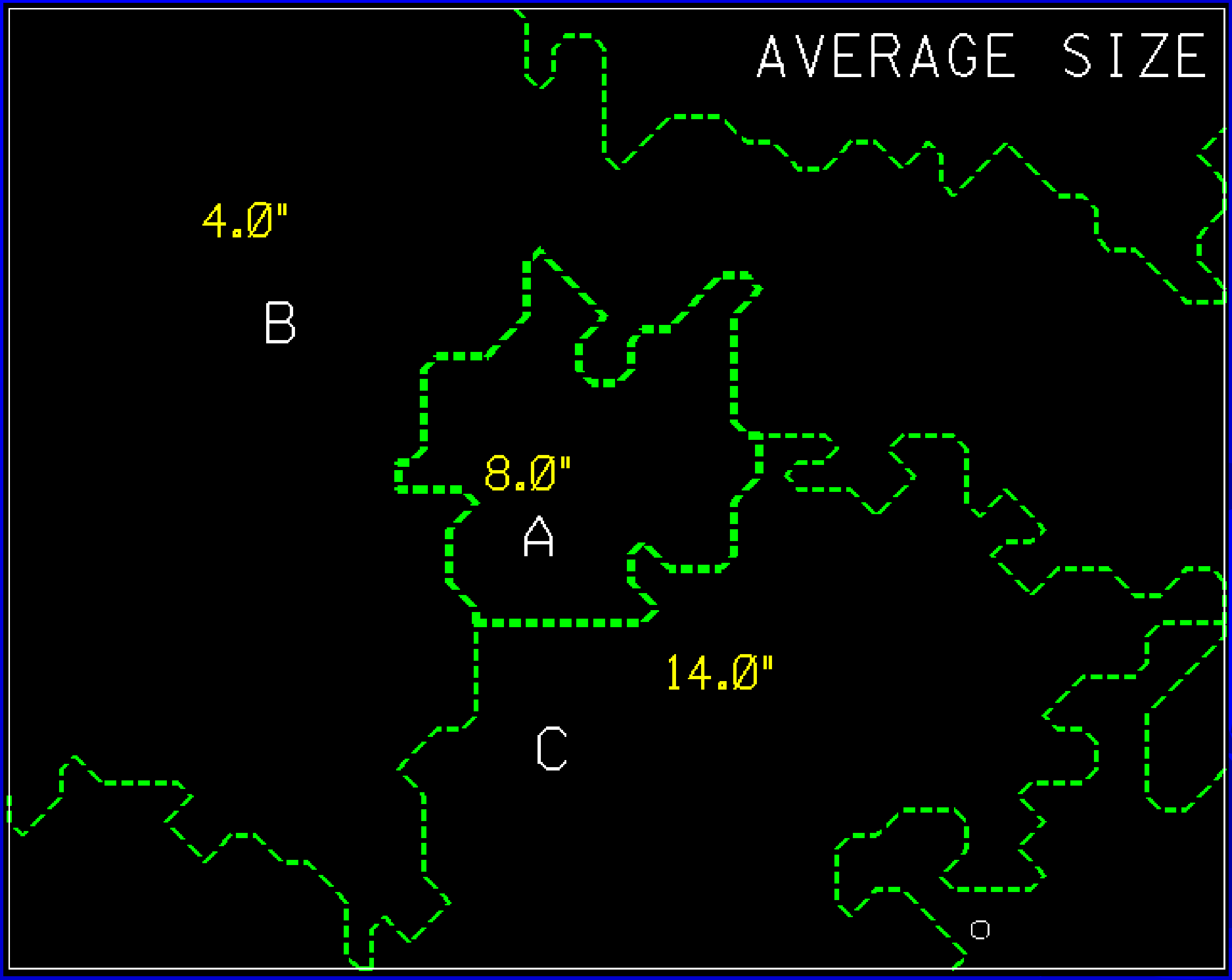
Similarity Of Features - Type



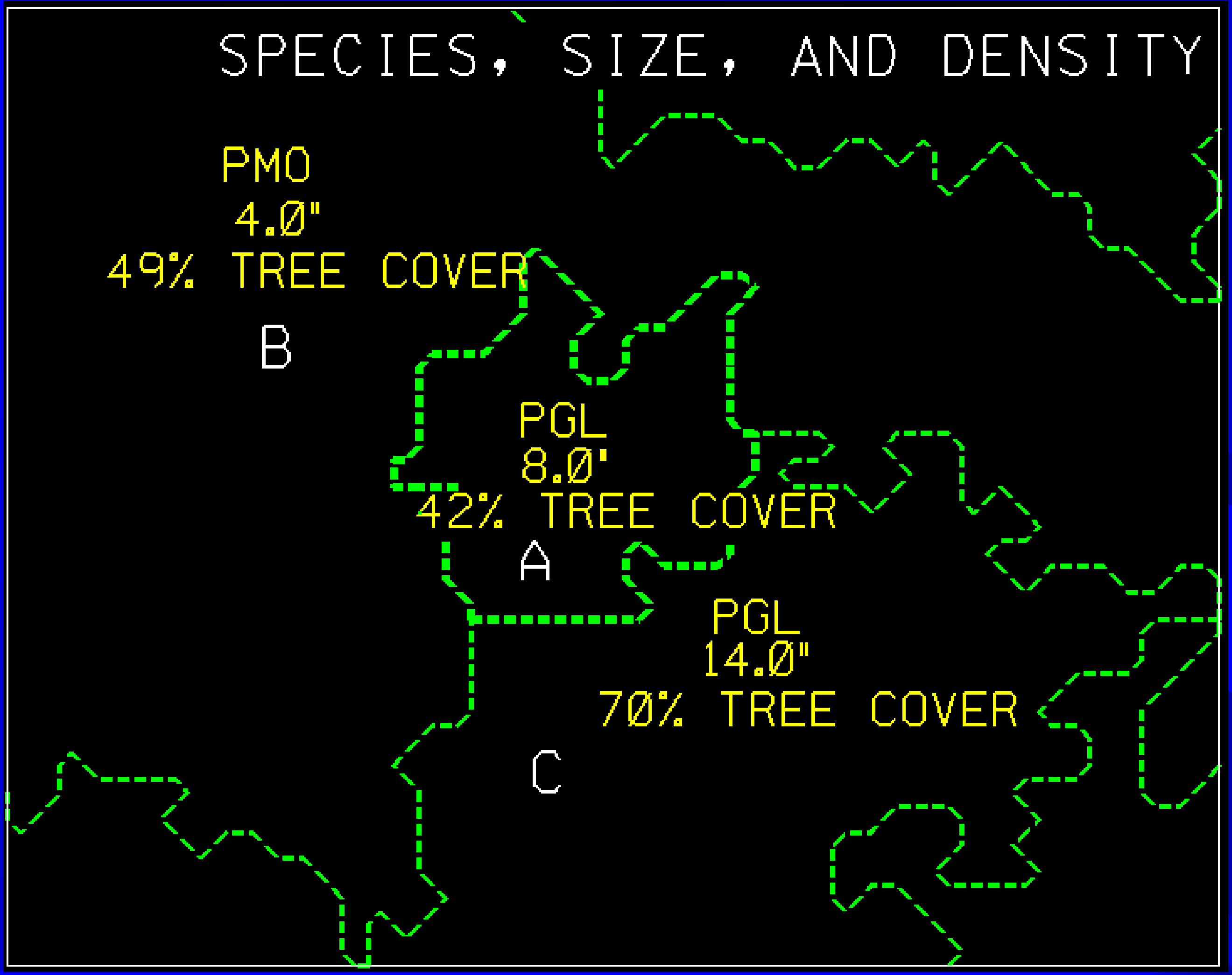
Similarity Of Features - Cover



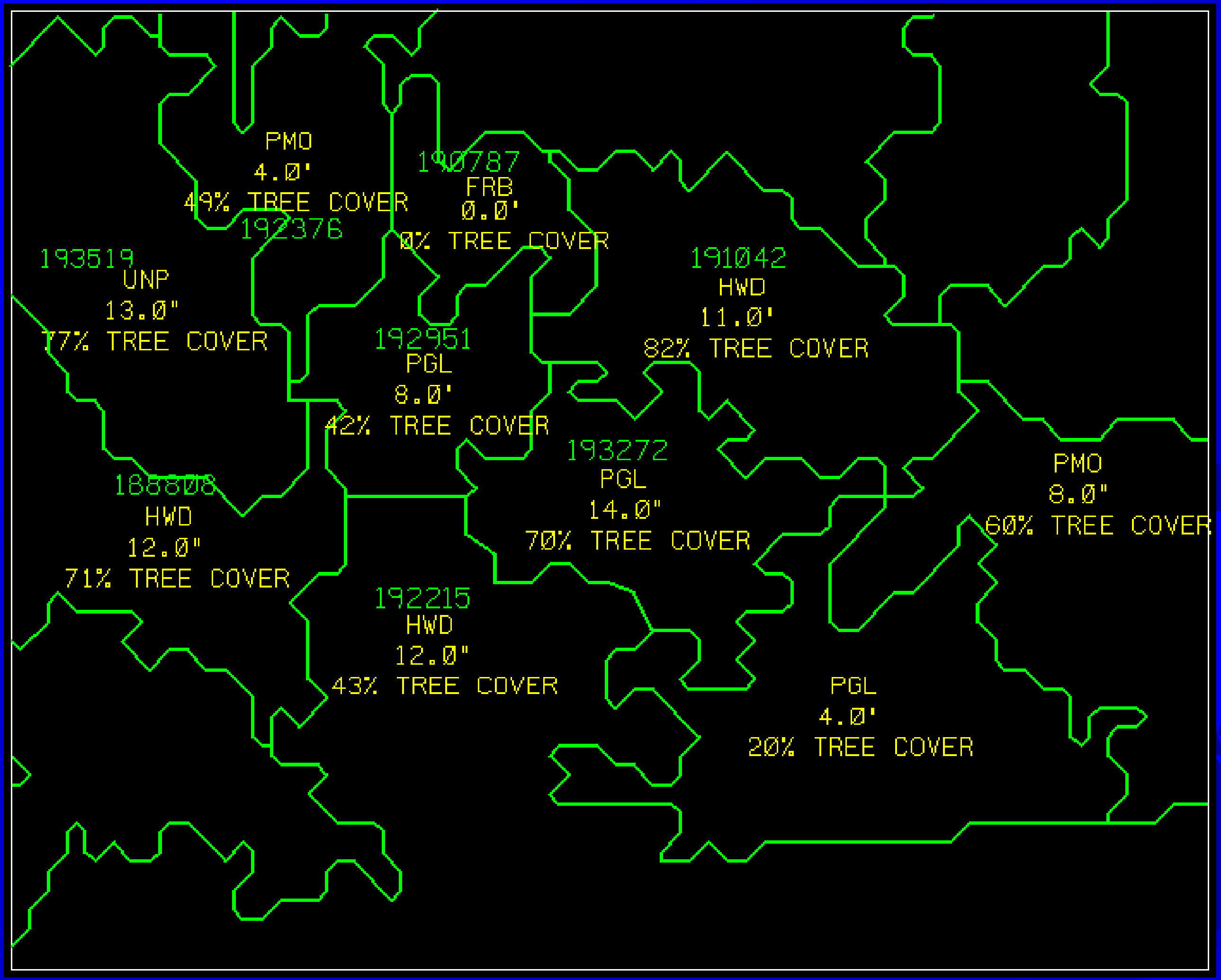
Similarity Of Features - Size



Similarity Of Multiple Features



Similarity Of Multiple Stands



Minimum Size Mapping Unit

- Degree of similarity - similar vs dissimilar
  - Desirable limits
  - Critical limits

Minimum Acreage Limits

Land Cover Type	Desirable Minimum Size	Critical Minimum Size
PGl	20.0 acres	3.0 acres
PMo	20.0 acres	3.0 acres
UnP	20.0 acres	3.0 acres
PHw	20.0 acres	3.0 acres
Hwd	20.0 acres	3.0 acres
TSh	20.0 acres	5.0 acres
LSh	20.0 acres	5.0 acres
DSh	20.0 acres	5.0 acres
MSh	20.0 acres	5.0 acres
Frb	20.0 acres	5.0 acres
Lch	5.0 acres	3.0 acres
H2O	5.0 acres	3.0 acres
…..		

Step 3: Evaluate Similarity and Merge With Most Similar

- Represent the rules as a function and attempt to quantify similarity
- For each subject area evaluate all adjacent areas and determine the most similar area
  - Merge the subject area into the most similar area
  - Recompute merged area attributes
- Stop when minimum mapping unit thresholds are met

Sample Similarity Estimates

Stand = 81373

----

stand#	iw	ip	cover	pctcon	shr	hrb	mtype	psp	ltype	pixels
81373	PGl	PGl	43	82	0.0	0.0	0	11	12	134
84939	PHw	PGl	35	57	0.0	0.0	0	11	20	55
* 10.5			2.0	2.5	0.0	0.0	0.0	0.0	6.0	

----

stand#	iw	ip	cover	pctcon	shr	hrb	mtype	psp	ltype	pixels
81373	**PGl**	**PGl**	**43**	**82**	**0.0**	**0.0**	**0**	**11**	**12**	**134**
82936	**PGl**	**PGl**	**55**	**85**	**0.0**	**0.0**	**0**	**11**	**12**	**55**
*** 3.3**			**3.0**	**0.3**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	

----

stand#	iw	ip	cover	pctcon	shr	hrb	mtype	psp	ltype	pixels
81373	PGl	PGl	43	82	0.0	0.0	0	11	12	134
85658	Hwd	Hwd	39	22	0.0	0.0	0	22	25	20
* 22.3			1.0	6.0	0.0	0.0	0.0	5.5	9.8	

Aggregate stand 81373 with stand 82936

Step 4: Report Polygon Attributes

- Summarize weighted averages of pixel characteristics within the polygon boundary
- Develop discrete estimates and variances from weighted averages as polygon attributes
  - variance of tree cover is related to spatial distribution of cover
  - variance of tree size is related to stand structure
- Develop single theme maps from polygon map through reclassification of database characteristics
- Develop categorical estimates from discrete estimates - no need to jaywalk anymore

Land Cover Density Summary:
Stand ID: 1789
Total Number of Pixels: 50
Contributing Pixels: 50

Size Class:	0-4"	5-8"	9-12"	13"+	Total	Other	Total
White Spruce	0.0%	0.0%	27.1%	0.0%	27.1%		27.1%
Black Spruce	0.0%	0.0%	10.2%	0.0%	10.2%		10.2%
Hardwood	0.0%	0.0%	6.4%	0.0%	6.4%		6.4%
Tall shrub						39.8%	39.8%
Low shrub						8.1%	8.1%
Wet moss						1.8%	1.8%
Lichen						5.3%	5.3%
Total Cover	0.0%	0.0%	43.7%	0.0%	43.7%	56.3%	100.0%
Total Tree Cover					**43.7%**		

Stand Tree Composition Summary:
Stand: 1789

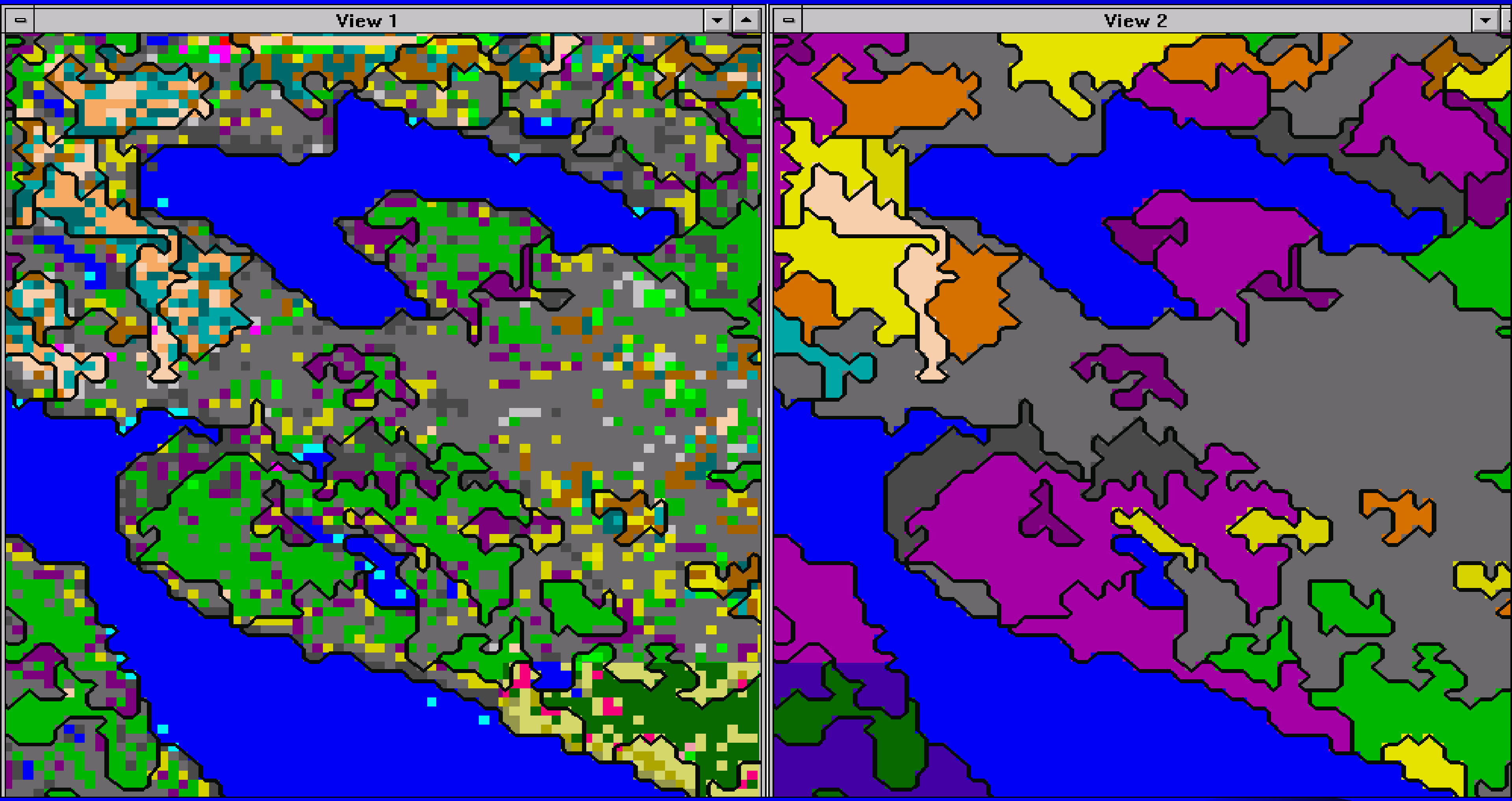
Size Class:	0-4"	5-8"	9-12"	13"+	Total
White Spruce	0.0%	0.0%	62.0%	0.0%	62.0%
Black Spruce	0.0%	0.0%	23.3%	0.0%	23.3%
Hardwood	0.0%	0.0%	14.7%	0.0%	14.7%
Total Tree Cover	0.0%	0.0%	100.0%	0.0%	100.0%

Table 7: Polygon Cover Description

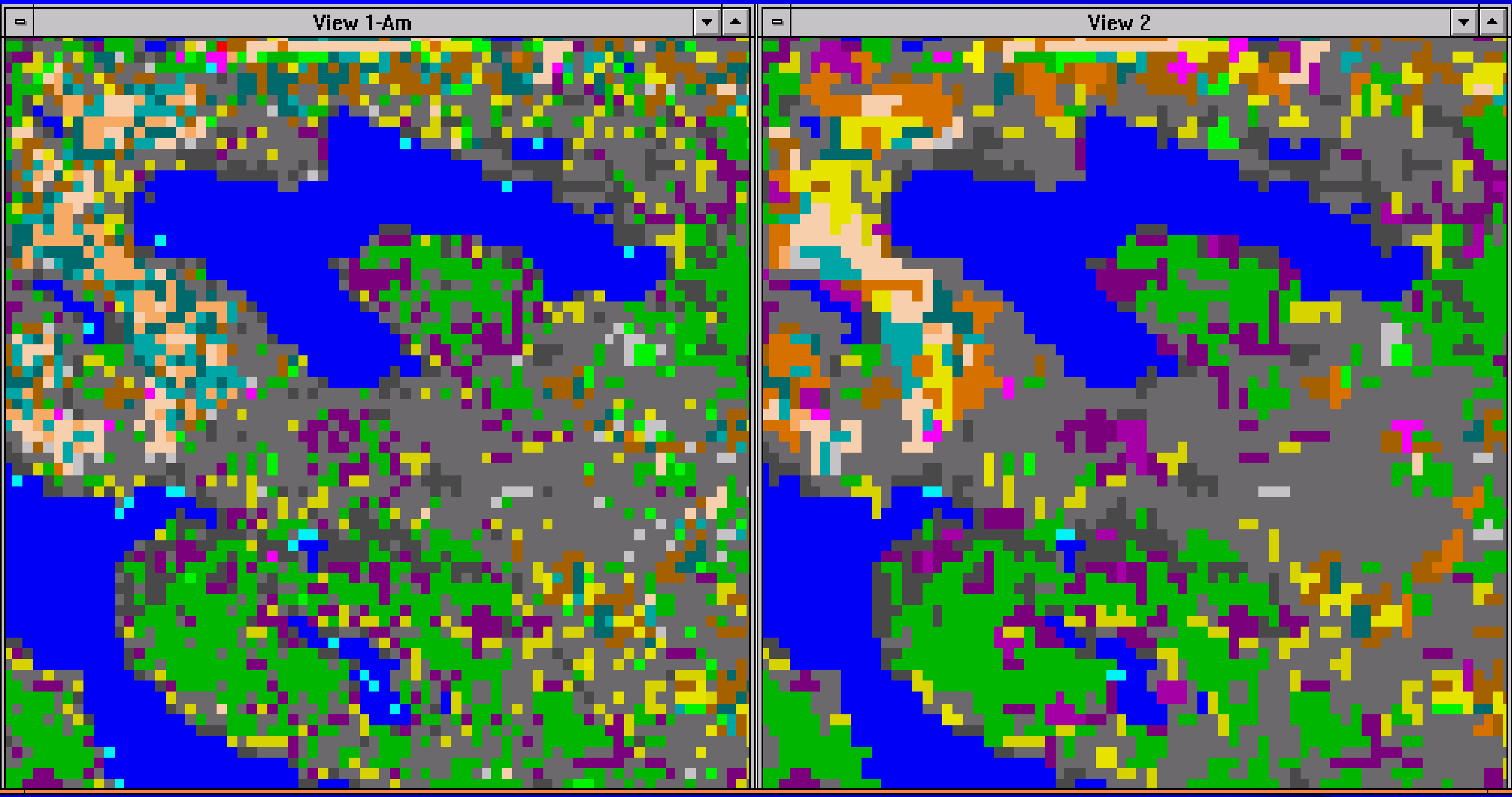
Table 9: Database Record Listing

wrangle_-id	[172598]
mapid	[100064]
lform	[s]
ltype	[UnP]
closure_class	[2]
density	[43.7]
pct_conifer	[85.3]
pct_hdwood	[14.7]
pr_species	[White Spruce]
pred_sp_pct	[62.0]
other_cover	[56.0]
cv_shr	[48.3]
cv_hrb	[7.6]
cv_bar	[0.3]
cv_oth	[0.0]
pix_ct	[50]
grid_val	[**1789**]
class_status	[8]
acreage	[11.1504]

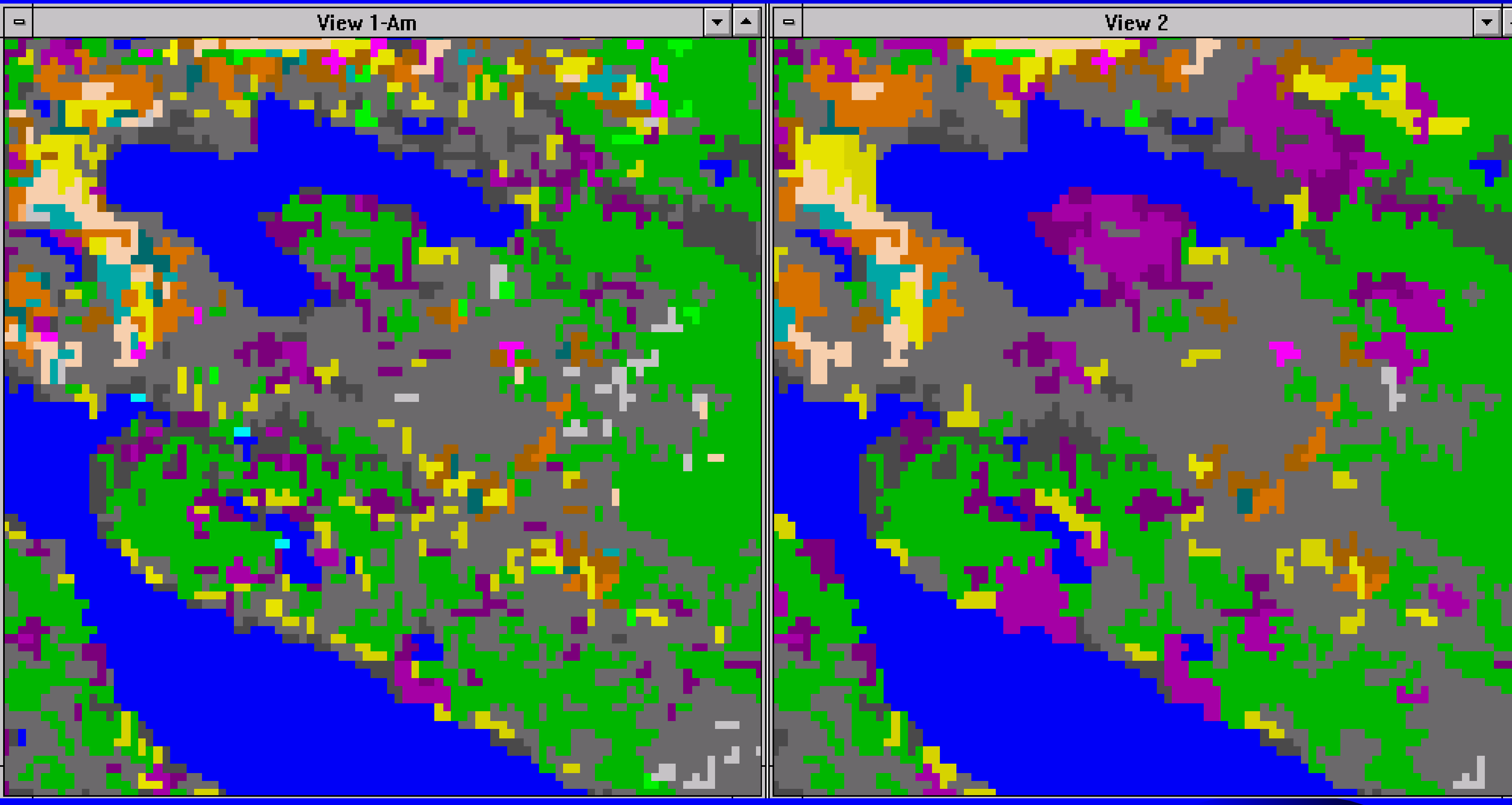
From Pixels to Polygons



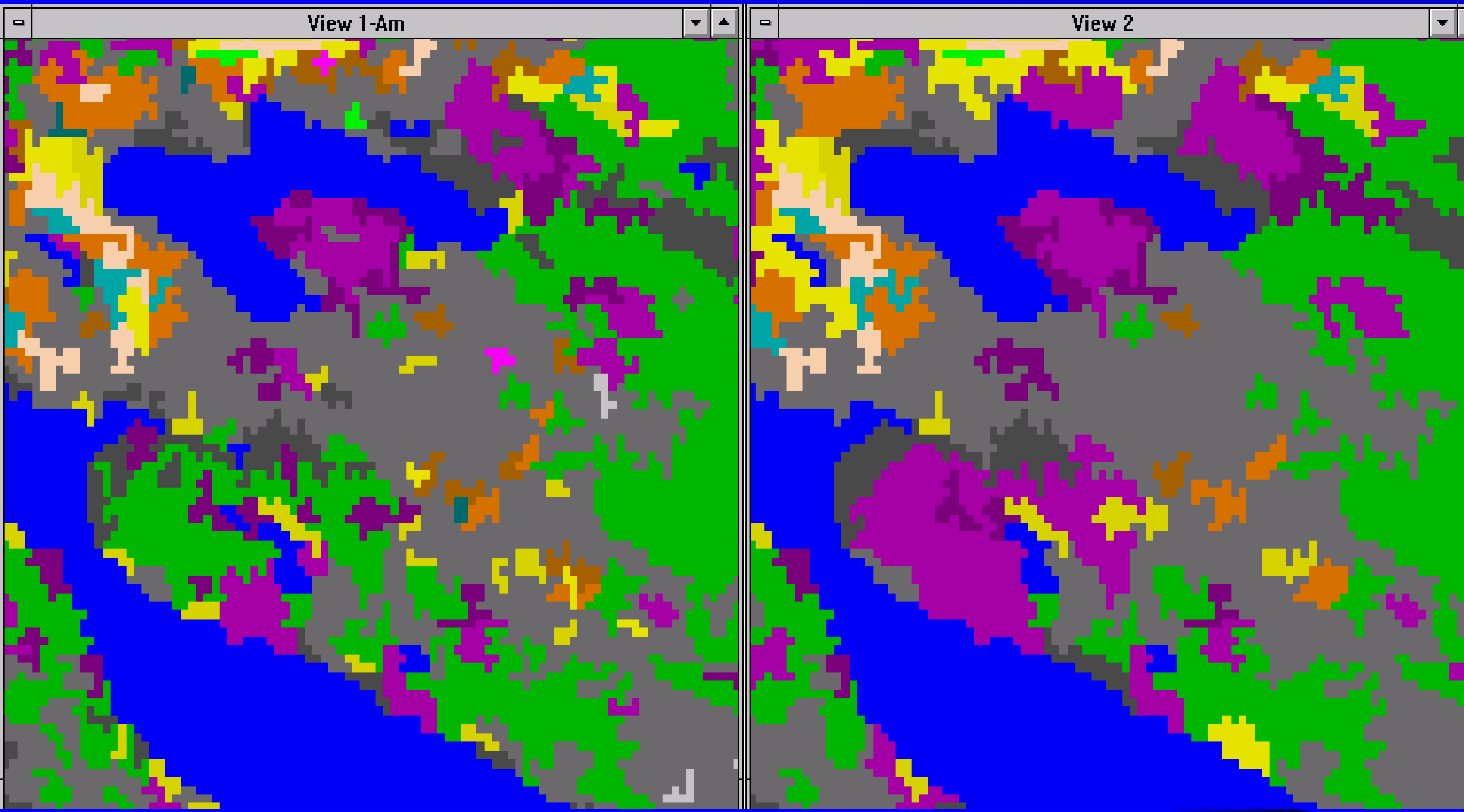
Iteration 1



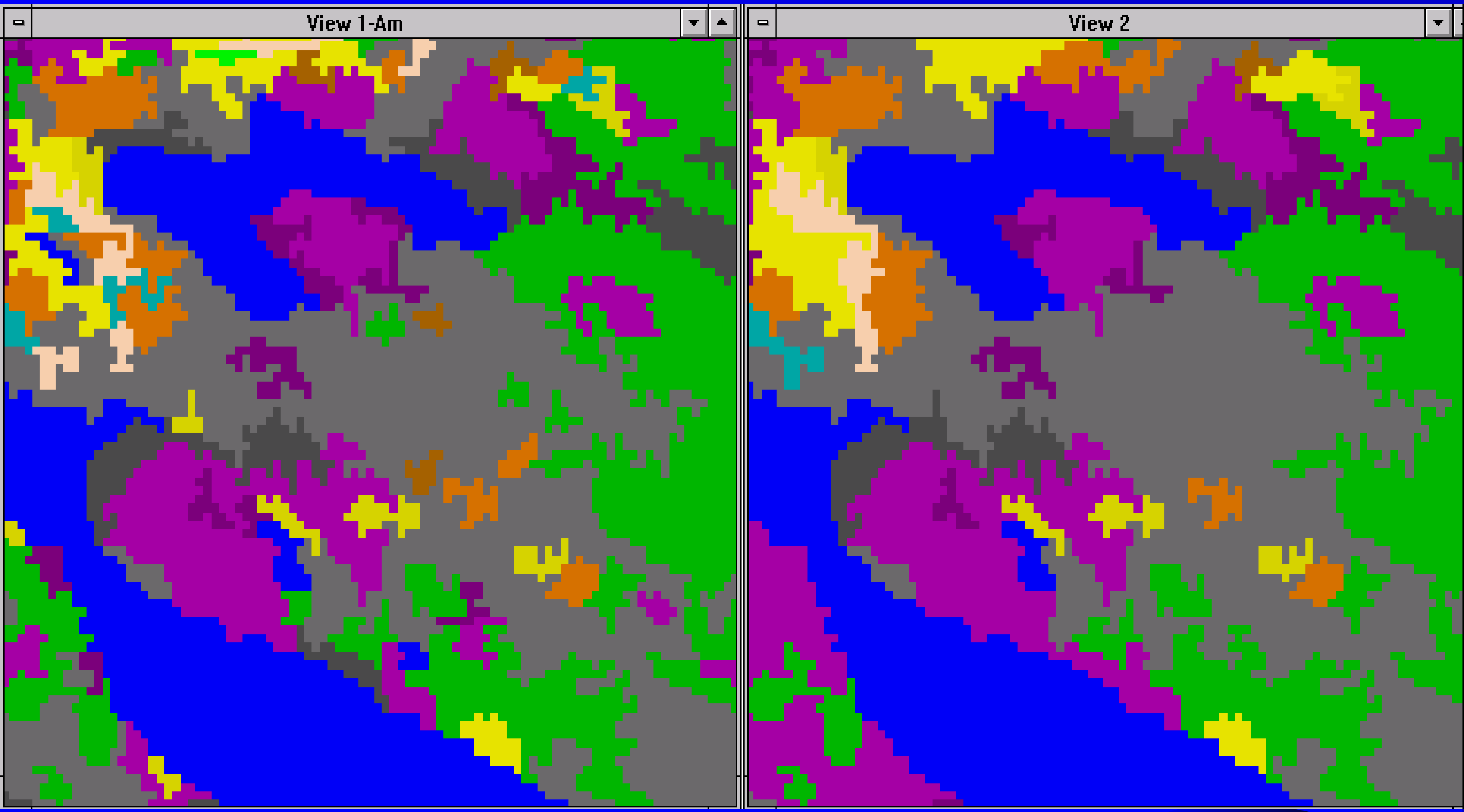
Iteration 2



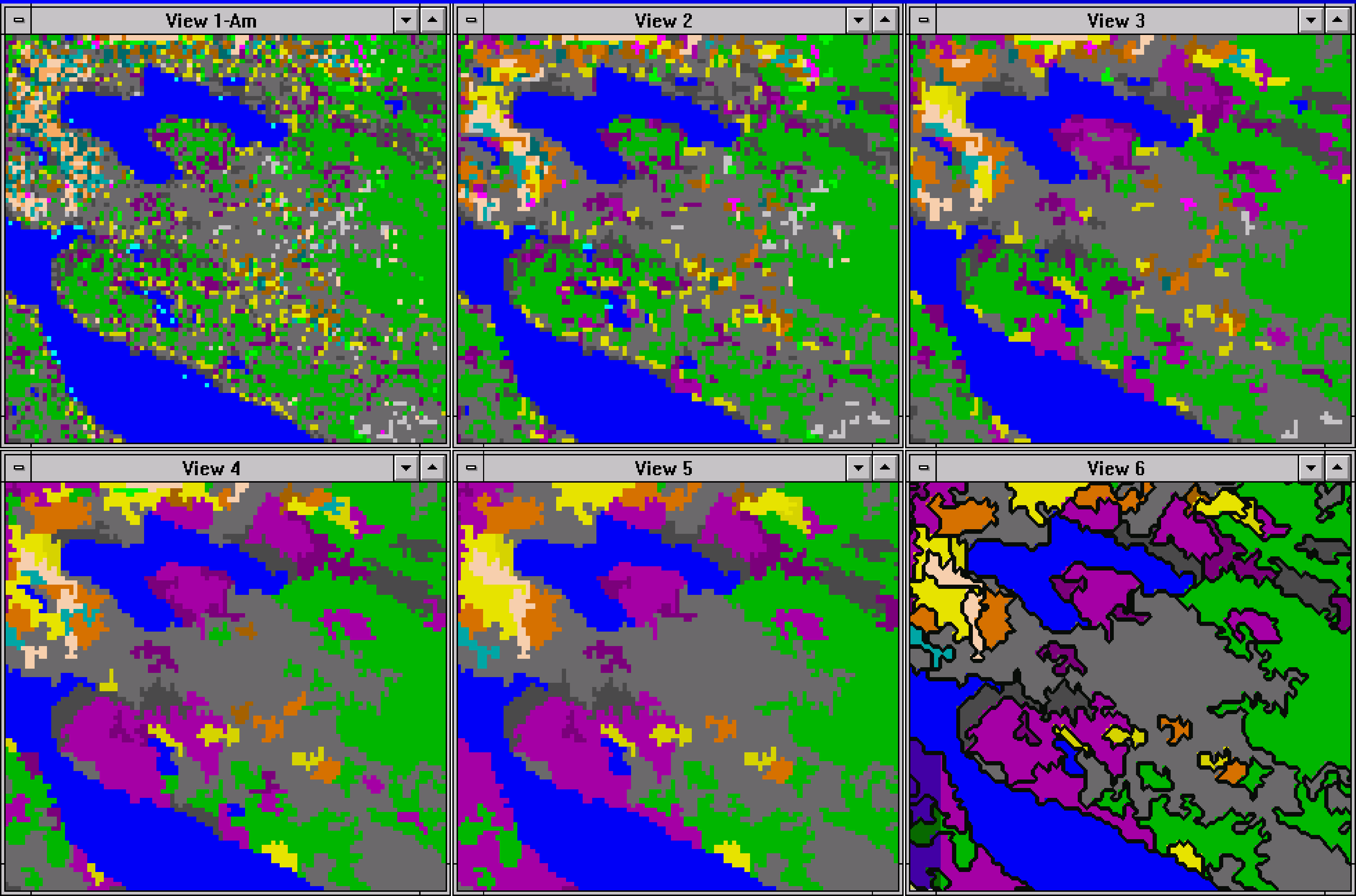
Iteration 3



Iteration 4



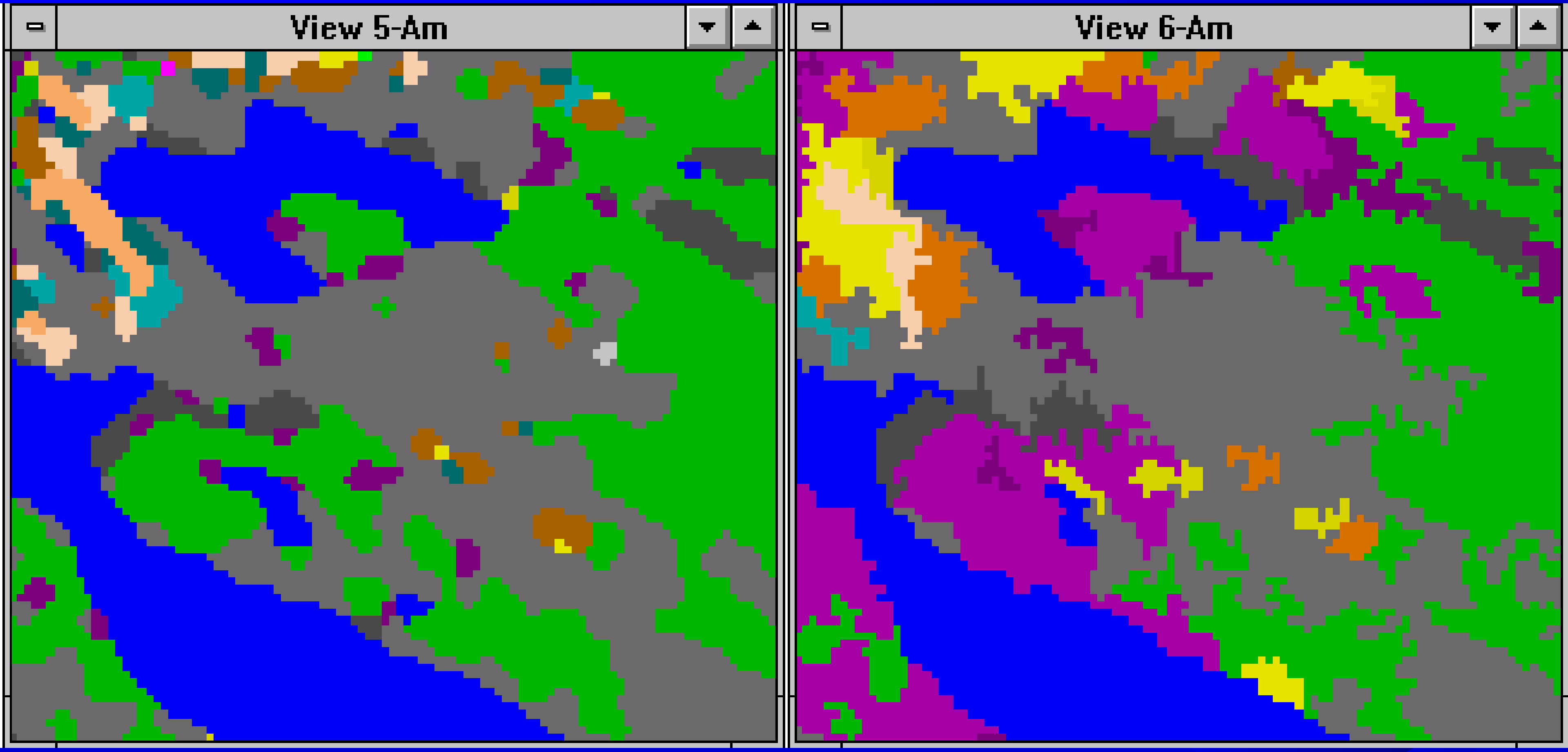
Start to Finish



Aggregation Process Results

- Overall balance of acreage by general type
- Movement towards mixed specie types
  - Unspecified Spruce
  - Spruce/Broadleaf
- Movement towards moderate density classes
- Development of new types
- Development of life form estimates based on attribute descriptions

Side by Side Results



Rule-based Aggregation Benefits

- Process millions of acres at one time
- Repeatable, consistent, and objective
- No human digitizing or editing of stand boundaries
- Can modify rules to change emphasis and produce different maps.
- Can aggregate using different vegetation classification schemes to develop different maps
- Similarity of values, not classes, yield polygons with lower within stand variation.
- Discrete estimates allow reclassification by user defined classes

Accuracy Assessment

The Determination of Map Accuracy

Checking the Map Database relative to …

Accuracy Assessment
Error Matrix

- Identify matches between 'mapped' estimates and 'ground truth' to estimate accuracy of the map
- Use to identify errors of omission and commission

Canopy Closure Error Matrix

		REFERENCE DATA						
		NON-TREE 0-20%	SPARSE 20-40%	OPEN 40-60%	MODERATE 60-80%	DENSE 80% +	TOTAL	PERCENT CORRECT
	NON-TREE	*20*					20	**100.0%**
M	SPARSE		*1*	1	3		5	**20.0%**
A								
P	OPEN			*3*		2	5	**60.0%**
	MODERATE			1	*11*	2	14	**78.6%**
D								
A	DENSE					*21*	21	**100.0%**
T								
A	TOTAL	20	1	5	14	25	65	
	PERCENT CORRECT	**100.0%**	**100.0%**	**60.0%**	**78.6%**	**84.0%**		**86.2%**

Kappa	**0.4574**
Var(Kappa)	0.0032

Tree Size Error Matrix

REFERENCE DATA

MAP DATA		0 non-forest	1 0-5"	2 5-13"	3 13-21"	4 21-32"	5 +32"	TOTAL	PERCENT CORRECT
	0	*20*						20	**100.0%**
	1		*1*					1	**100.0%**
	2			*13*	1			14	**92.9%**
	3	1		1	*13*			15	**86.7%**
	4				1	*5*	1	7	**71.4%**
	5						*2*	2	**100.0%**
	TOTALS	21	1	14	14	6	3	59	
	PERCENT	**95.2%**	**100.0%**	**92.9%**	**92.9%**	**83.3%**	**66.7%**		**91.5%**

Kappa **0.5126**
Var(Kappa) 0.0023

Land Cover Type Error Matrix

REFERENCE DATA

MAP DATA	BA	CH	DH	EH	GF	MC	SC	TF	WA	TOTAL	PERCENT CORRECT
BA	*2*									2	**100%**
CH		*5*		1						6	**83%**
DH		2	*4*				1			7	**57%**
EH		3		*2*						5	**40%**
GF					*20*					20	**100%**
MC						*14*		1		15	**93%**
SC					1		*16*			17	**94%**
TF							1	*14*		15	**93%**
WA	1								*4*	5	**80%**
TOTAL	3	10	4	3	21	14	18	15	4	92	
PERCENT	**67%**	**50%**	**100%**	**67%**	**95%**	**100%**	**89%**	**93%**	**100%**		**88%**

Kappa **0.8589**
Var(Kappa) 0.0015

What is a Match ?

- Map class boundaries are artificial and may not occur in the field.

  Be wary of problems.
- A choice between fuzziness or statistics?

	WHR	Pr_species	CC_Class	CC	SZ_Class	QMD	pct_con	pct_hwd
polygon	DFR	Douglas-fir	D	61%	4	25”	70%	30%
sample	MHC	Douglas-fir	M	59%	3	23”	65%	35%

Match Determination

- Use Statistical Parameters for Continuous Variables
  - Canopy Closure
  - Average Tree Size
  - Is the land cover type a continuous variable ?
- How do we develop the data or "ground truth"?

Accuracy Assessment

- Potential Problem
  - Photo-interpretation is a type of "Ground Truth"
- Solution
  - Ground Truth is "Ground Truth"
  - Reference Data
    - The data set used to test map accuracy is assumed to be 100% correct.
    - Must test all types in the map.

Agreement vs. Accuracy

- Sources of Reference Data
  - Existing Maps = Agreement
  - Photo-interpretation = Agreement
  - Ground data collected using ocular estimates or through the windshield = Agreement
  - Statistically valid sampling and measured field data = Accuracy!

Adequate Sampling of the Final Database

- Systematic Sampling
- Simple Random Sampling
- Stratified Random Sampling

Potential Bias ? Frequency vs. Area

- Errors may be related to polygon size
  - Larger polygons are more accurate as they are based on more homogeneous data
  - Smaller polygons are less accurate as they may be based on aggregations of types that meet the minimum mapping unit constraints.

Area Adjusted Accuracy

REFERENCE DATA

MAP DATA		0 non-forest	1 0-5"	2 5-13"	3 13-21"	4 21-32"	5 +32"	TOTAL	PERCENT CORRECT	ACRES	CORRECT ACRES
	0	20						20	100.0%	67,677	67,677
	1		1					1	100.0%	299	299
	2			13	1			14	92.9%	163,788	152,089
	3	1		1	13			15	86.7%	263,881	228,697
	4				1	5	1	7	71.4%	92,047	65,748
	5						2	2	100.0%	7,470	7,470
	TOTALS	21	1	14	14	6	3	59		595,162	521,979
	PERCENT	95.2%	100.0%	92.9%	92.9%	83.3%	66.7%		91.5%		

TOTAL PERCENT CORRECT ACRES 87.7%

Kappa 0.5126

Var(Kappa) 0.0023

Problems with Sample Bias ? Withheld Training Data

- If training area locations are not randomly selected, data collected in potential training data sites cannot be used for accuracy assessment data
  - Homogenous data collection site is abnormal with respect to mapped types - will only tests homogenous areas
  - Only tests 'known' types and none that may evolve during the mapping process
  - Really only a test of training data development procedures

Overestimation of Accuracy
Single theme vs. multiple themes

- Errors are multiplicative
  - 85% Canopy Closure
  - 85% Size
  - 85% Species Type
  - Real error rate for combined attributes
    - .85 x .85 x .85 = .61

Spatial Accuracies

- Data collection sites should be located without bias in order to
  - Verify the location of data collection sites.
  - Verify the location of polygon boundaries.

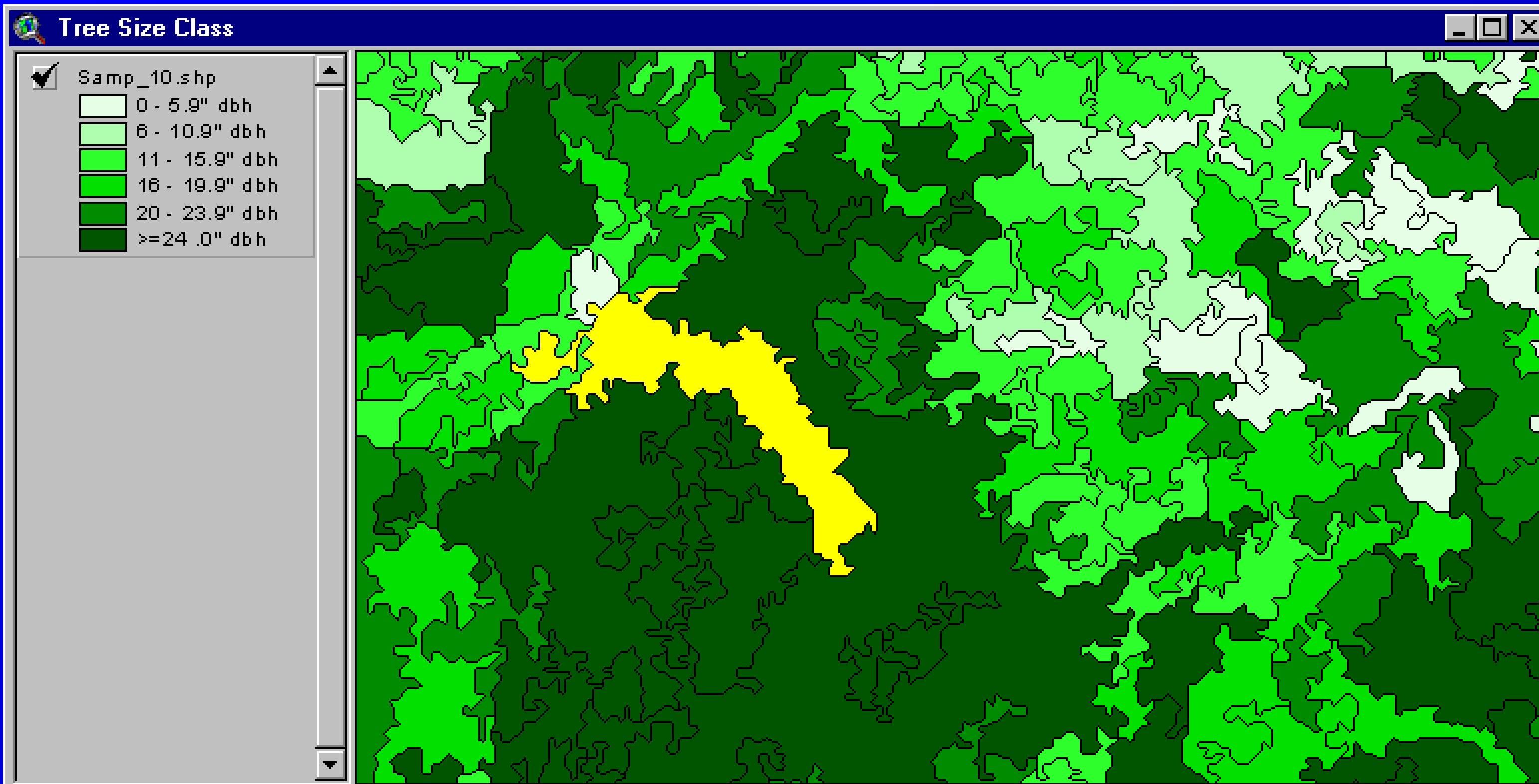
Accuracy Assessment Summary

- Need Repeatable, Objective, and Statistically Valid Mapping Tools
- Need to Understand the Features Being Mapped
- Need to Understand the Technical Issues

If You Avoid these Pitfalls …

- The data
  - Imagery and differential illumination
  - “Ground-truth” - field data
  - Training site selection and development
- Classification techniques
- Pixel cleanup and modeling
- Accuracy assessment

Results speak for themselves …



Attributes of Samp_10.shp

Samp_10	Veg_type	Closure_c	Density	Pct_conife	Pct_hdwood	Size_class	Qmdbh	Qmdbhcon	Qmdbhhwd	Tpa_tot	Tpa_con	Tpa_hwd	Cv1	Cv2
908	MC	7	74.3	96.5	3.5	7	27.2	27.6	11.2	522.1	502.1	20.0	8.5	7.1
909	TF	3	37.6	93.4	6.6	5	19.7	20.3	7.4	361.4	288.3	73.1	8.2	5.7
910	TF	6	62.3	97.2	2.8	5	20.2	20.4	10.2	262.9	243.7	19.2	4.4	8.4
911	TF	4	41.2	100.0	0.0	5	19.0	19.0	0.0	318.0	318.0	0.0	4.9	7.2
912	MC	7	78.0	98.5	1.5	7	28.4	28.6	13.1	333.2	329.0	4.2	5.4	8.4
913	MC	3	38.4	81.4	18.6	4	16.1	17.3	8.7	518.5	370.8	147.7	11.6	8.1
914	SC	0	18.4	59.3	40.7	2	5.6	6.3	4.3	834.6	448.5	386.0	15.7	1.6
915	MC	3	39.2	94.4	5.6	5	18.4	18.7	9.3	378.7	359.1	19.5	6.3	6.5
916	MC	8	89.6	79.1	20.9	6	21.7	23.8	10.8	272.4	175.9	96.5	5.7	16.4
917	CH	2	29.8	60.5	39.5	2	6.8	7.6	5.4	1006.1	478.4	527.7	21.8	4.2
918	GF	0	7.1	0.0	0.0	0	0.0	0.0	0.0	315.7	204.1	111.5	5.1	0.7

Other 'Derived' Applications Based on Detailed Land Cover Characteristics

- National Vegetation Classification System
  - Alliance/Association
- Fire Fuel Class Modeling
  - Fire Behavior Modeling
- Wildlife Habitat Modeling
  - Habitat Suitability

Citations

Biging, Greg S., and Edward C. Murphy. 1992. A comparison of photointerpretation and ground measurements of forest structure. In *Technical Papers. 1991 ACSM-ASPRS Annual Convention*. American Congress on Surveying and Mapping and American Society for Photogrammetry and Remote Sensing, Baltimore. Vol. 3 (Remote Sensing), 6-15.

Brown, G. and L. Fox, 1992. Digital Classification of Thematic Mapper Imagery for Recognition of Wildlife Habitat Characteristics. In: Proc. 1992 ASPRS/ACSM Convention, American Society of Photogrammetry and Remote Sensing, Bethesda, MD, (4):251-260

Civco, D.L., 1991: Topographic normalization of Landsat Thematic Mapper digital imagery. Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 9, pp. 1303-1309.

Colby, J.D., 1991: Topographic normalization in rugged terrain. Photogrammetric Engineering and Remote Sensing, Vol. 57, No. 5, pp. 531-537.

Stumpf, K., 1993. From Pixels to Polygons: The Rule-Based Aggregation of Satellite Image Classification Data Using Ecological Principles. In: Proc. Seventh Annual Symposium on GIS in Forestry, Environment and Natural Resources. Vancouver B.C. Canada. (2):939-945

Stumpf, K., 1993. The Estimation of Forest Vegetation Cover Descriptions Using a Vertical Densitometer. In: Proc. Society of American Forestry Inventory and Biometrics Working Groups, Indianapolis, IN. pp 163-169.

Stumpf, K. and J. Koltun, 1992. Rule Based Aggregation of Raster Image Classifications into Vector GIS Databases with Five- and Forty-acre Minimum Mapping Units. In: Proc. 1992 ASPRS/ACSM Convention, American Society of Photogrammetry and Remote Sensing, Bethesda, MD, (4):261-278

Papers and Presentations

See Publications page at www.grsgis.com

Sample Data Sets

See/Contact Ken Stumpf

Questions and Comments

Air Photo#: - Polygon code: Lat/Long: /

Landcover code: Landcover class name: Azimuth: Slope (0-100 °):

Surveyors: S.Klein K.Stumpf S.Henricy J.Grunblatt K.Boggs B.Koltun Photos: Roll #: Frame #s:

Dig photo: sess. #: photo #: azimuth:

Modifier: Riparian Alpine Subalpine Boreal Tussock Bog Meadow Videotape #:

Hydrologic Regime (circle one): Dry - Mesic - Wet - Aquatic % Tussocks: Landform:

Hgt (m)	**Tree**	**%**	**Tall shrub (>1.5m)**	**%**	**Low shrub (1.5m >20cm)**	**%**	**Dwarf shrub (<20 cm)**	**%**	**Herbaceous**	**%**
									Lichen	
									Moss	
									Sand/Gravel	
									Mud	
									Bedrock	
									Rock (Talus)	
									Bare	
									Water	
	Total		**Total**		**Total**		**Total**		**Total**	
									GRAND TOTAL	

	Spruce		**Mixed forest**		**Low shrub**		**Herbaceous**
1	Closed White Spruce	17	Closed Mixed	25	Closed Low Alder-Willow	40	Dry/Mesic Graminoid
2	Open White Spruce	18	Open Mixed	26	Closed Low Shrub Birch-Willow	41	Dry/Mesic Graminoid Forb
3	Open Black Spruce	19	Woodland Mixed	27	Closed Low Willow	42	Tussock
4	Open Stunted Spruce (black spruce)			28	Closed Low Shrub Birch-Ericaceous Shrub-Willow	43	Wet Salt Marsh Herbaceous
5	Open Spruce Lichen			29	Closed Low Shrub Birch	44	Wet Freshwater Herbaceous
6	Woodland Spruce			30	Open Low Alder-Willow	45	Aquatic
7	Woodland Stunted Spruce			31	Open Low Alder		
8	Woodland Spruce Lichen		**Tall shrub**	32	Open Low Shrub Birch-Ericaceous Shrub-Willow		**Other**
		20	Closed Tall Alder	33	Open Low Willow	46	Sparse Vegetation
			-Willow	34	Open Low Shrub Birch-Ericaceous Shrub	47	Marine Mudflat
	Deciduous forest	21	Closed Tall Alder			48	Sand/Gravel
9	Closed Birch	22	Closed Tall Willow			49	Bedrock
10	Closed Aspen	23	Open Tall Alder			50	Snow/Ice
11	Closed Balsam Poplar	24	Open Tall Willow		**Dwarf shrub**	51	Shadow/Indeterminate
12	Closed Mixed Deciduous			35	Dryas Dwarf Shrub	52	Silty Water
13	Open Birch			36	Mixed Dwarf Shrub	53	Clear Water
14	Open Aspen			37	Mixed Dwarf Shrub-Sedge	54	Marine Water
15	Open Balsam Poplar			38	Mixed Dwarf Shrub-Lichen	55	Human Development
16	Open Mixed Deciduous			39	Mixed Dwarf Shrub-Rock	56	Burn Site

INCLUSIONS/STRINGERS:

COMMENTS:

Land Cover Type Error Matrix

REFERENCE DATA

		BA	CH	DH	EH	GF	MC	SC	TF	WA	TOTAL	PERCENT CORRECT
	BA	*2*									2	**100%**
	CH		*5*		1						6	**83%**
M A P	DH		2	*4*				1			7	**57%**
	EH		3		*2*						5	**40%**
D A T A	GF					*20*					20	**100%**
	MC						*14*		1		15	**93%**
	SC					1		*16*			17	**94%**
	TF							1	*14*		15	**93%**
	WA	1								*4*	5	**80%**
	TOTAL	3	10	4	3	21	14	18	15	4	92	
	PERCENT	**67%**	**50%**	**100%**	**67%**	**95%**	**100%**	**89%**	**93%**	**100%**		**88%**

Kappa **0.8589**
Var(Kappa) 0.0015