An Image Classification Sampling Methodology Based on the Integration of IP/GIS Capabilities

By
Ken Stumpf
Geographic Resource Solutions
Arcata, CA
grs@grsgis.com

Techniques for Developing a Comprehensive and Cost Effective Land Cover Mapping Training Data Set

I will briefly present …

- The need for this methodology
- The sampling methodology
  - Image Stratification
  - Candidate Site Database Development
  - Candidate Site Refinement
  - Sample Plan Development and Administration
- Benefits of this methodology

Training Data are Extremely Important

Image processing training data collection issues:

- Foundation for accurate and detailed land cover mapping
  - represent diversity of land cover
  - represent area of interest
- Significant cost component
  - large number of "types"
  - travel time and equipment
  - number of samples
- Limited field sampling opportunity

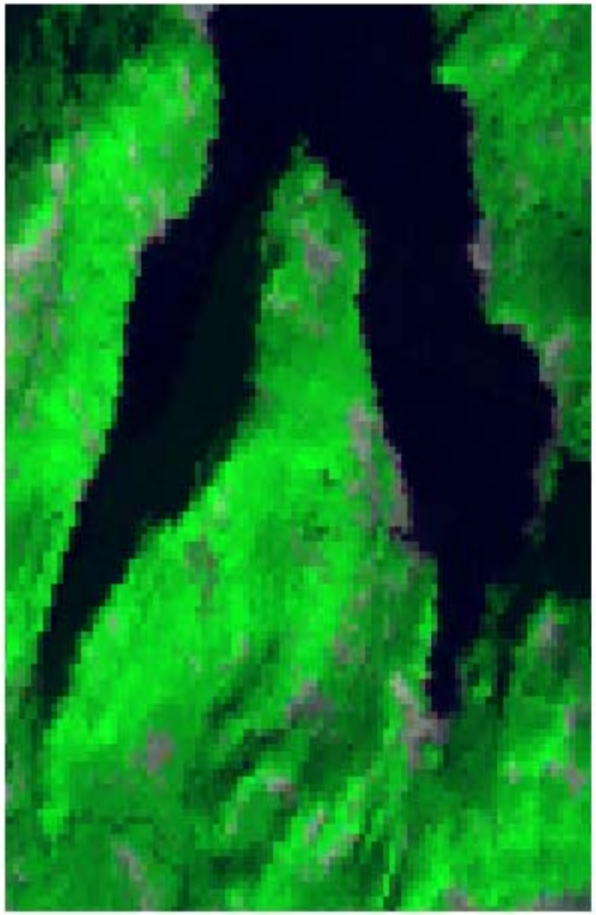
Control Effort and Cost

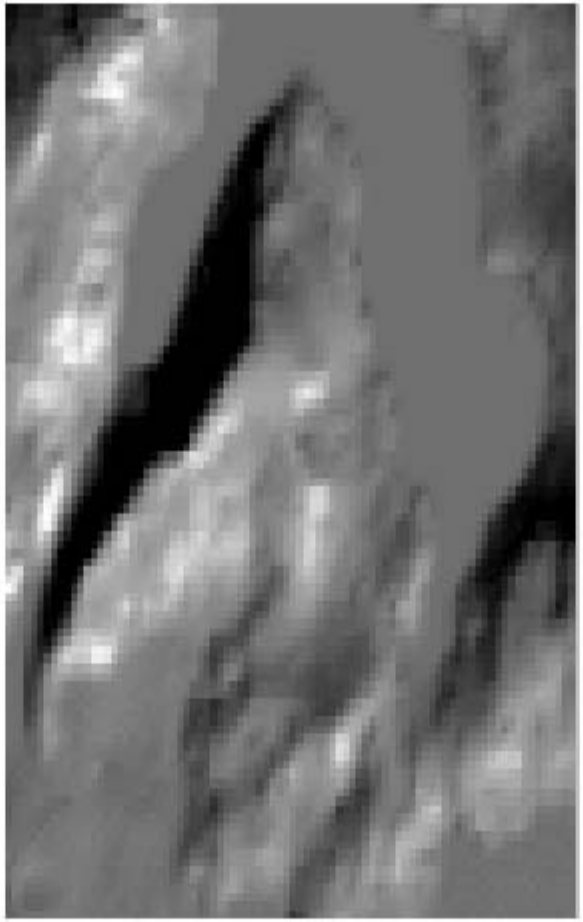
Reduce the number of data collection sites while still describing the diversity and geographic range of the project area

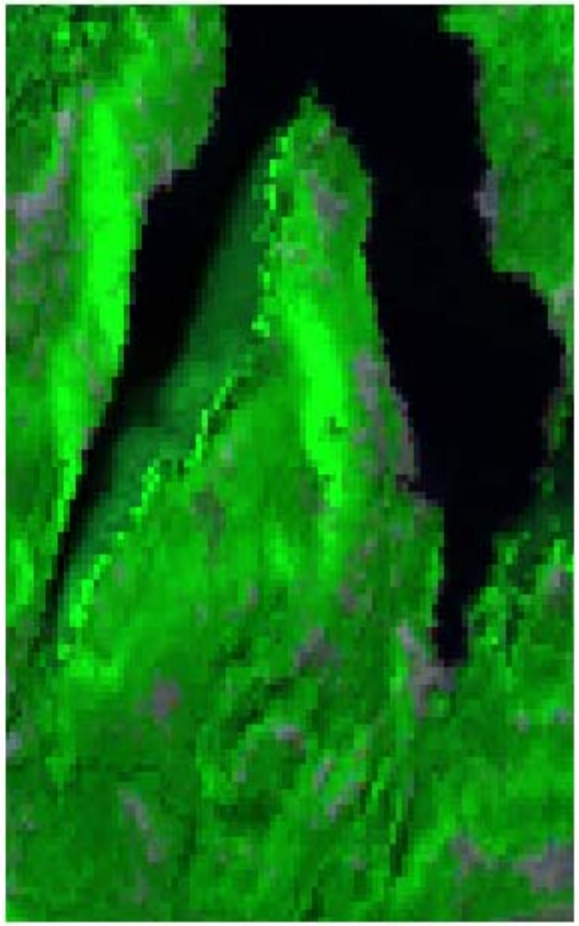
- perform illumination correction
- apply classification training site sampling methodology
- use existing data

Geographic Resource Solutions

Illumination Correction







Reduce Sampling Effort

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

Let's preprocess the data so we only collect data only at 'valid' training sites.

Characteristics of 'Valid' Training Areas

- Spectrally homogeneous and normally distributed
- Accessible
- Sufficient size
  - for an adequate sample of pixels
  - to locate in the field
  - to distinguish from neighboring types

The Norm is Opportunistic Sampling

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

A different approach ….

- Let’s use the data and our IP/GIS tools to guide and direct our training data collection efforts -

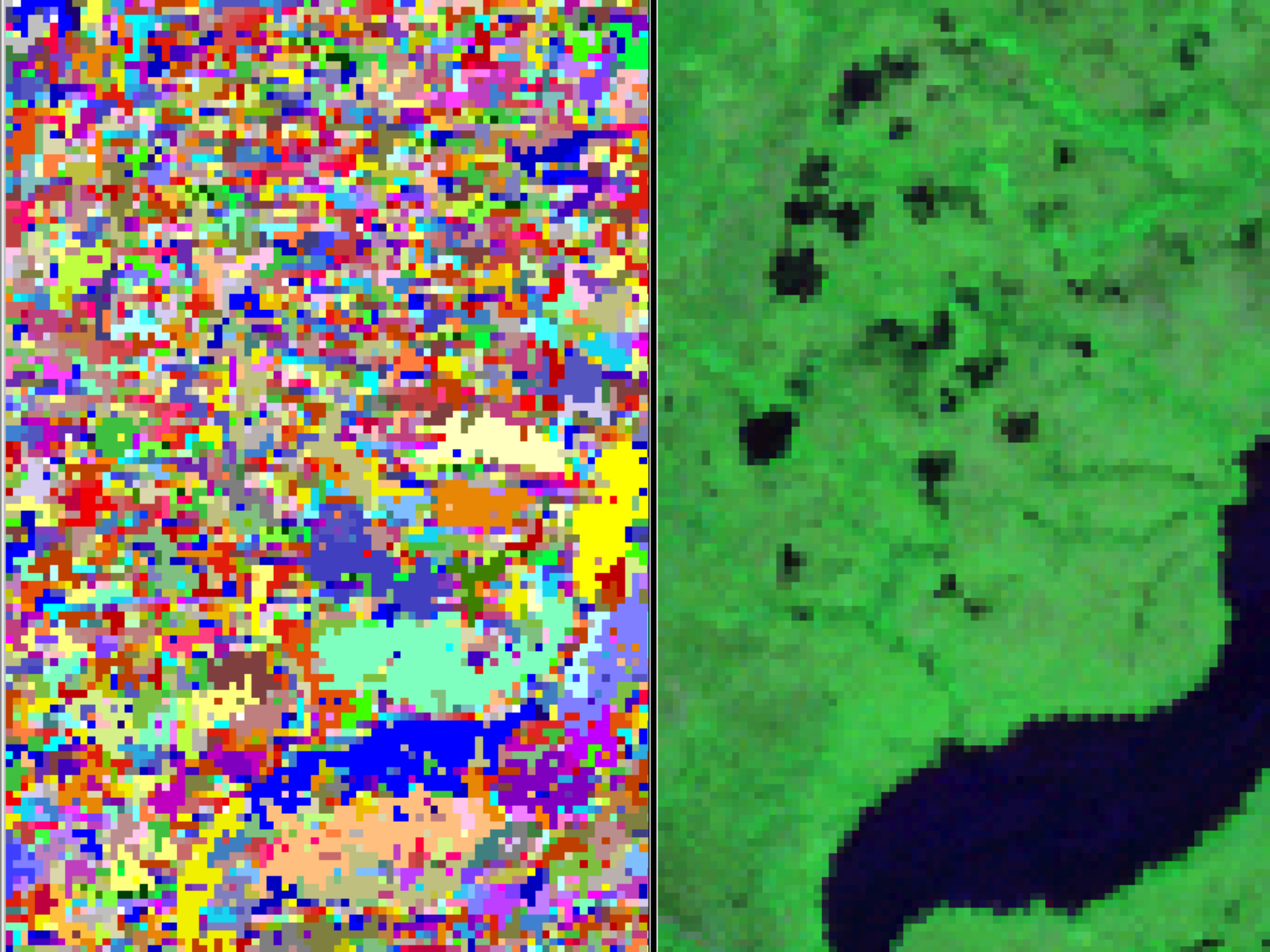
  avoid invalid sites and focus on those sites necessary to build a detailed training data set that accurately represents the range of land cover characteristics over the entire project area.

GRS's Training Data Development Methodology

- Stratify Project Area/Imagery
- Build Candidate Site Database
- Refine Candidate Sites
- Develop and Administer Training Set Sample Plan

Image Stratification

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



Determine 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	**
.....					

Unique Area Isodata Class 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 Refinement

- 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 unique areas to 36,833 candidate areas

Geographic Resource Solutions

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 Identification

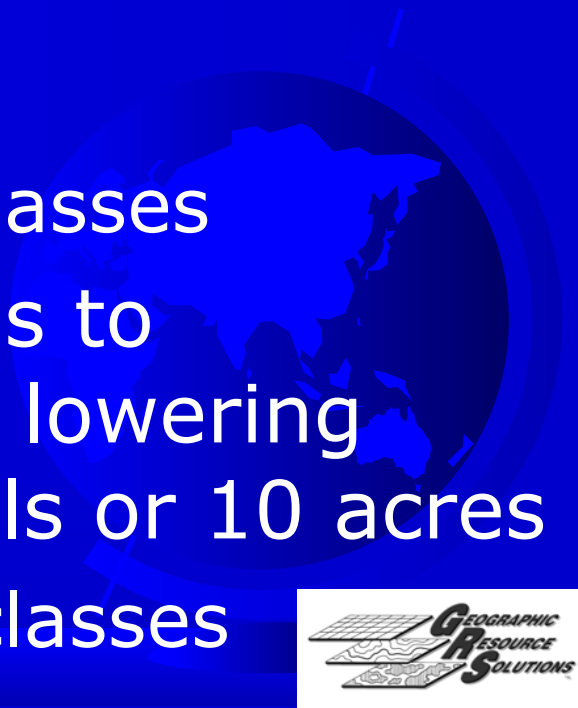
- Determine missing or rare classes

  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

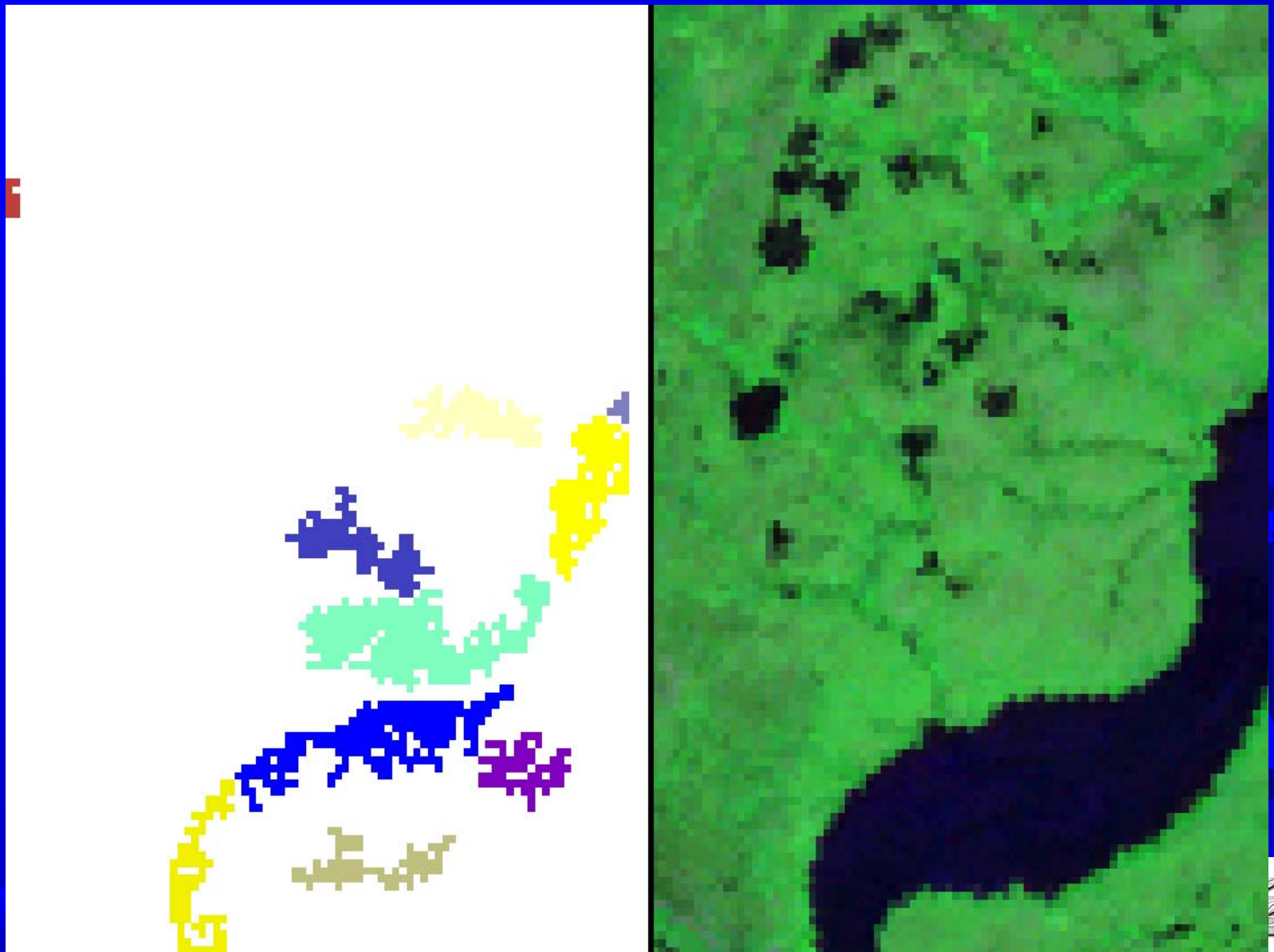
  - identified 42 scarce isodata classes
- Add additional candidate areas to supplement scarce classes by lowering minimum size limit to 45 pixels or 10 acres
  - added 305 sites to these 42 classes



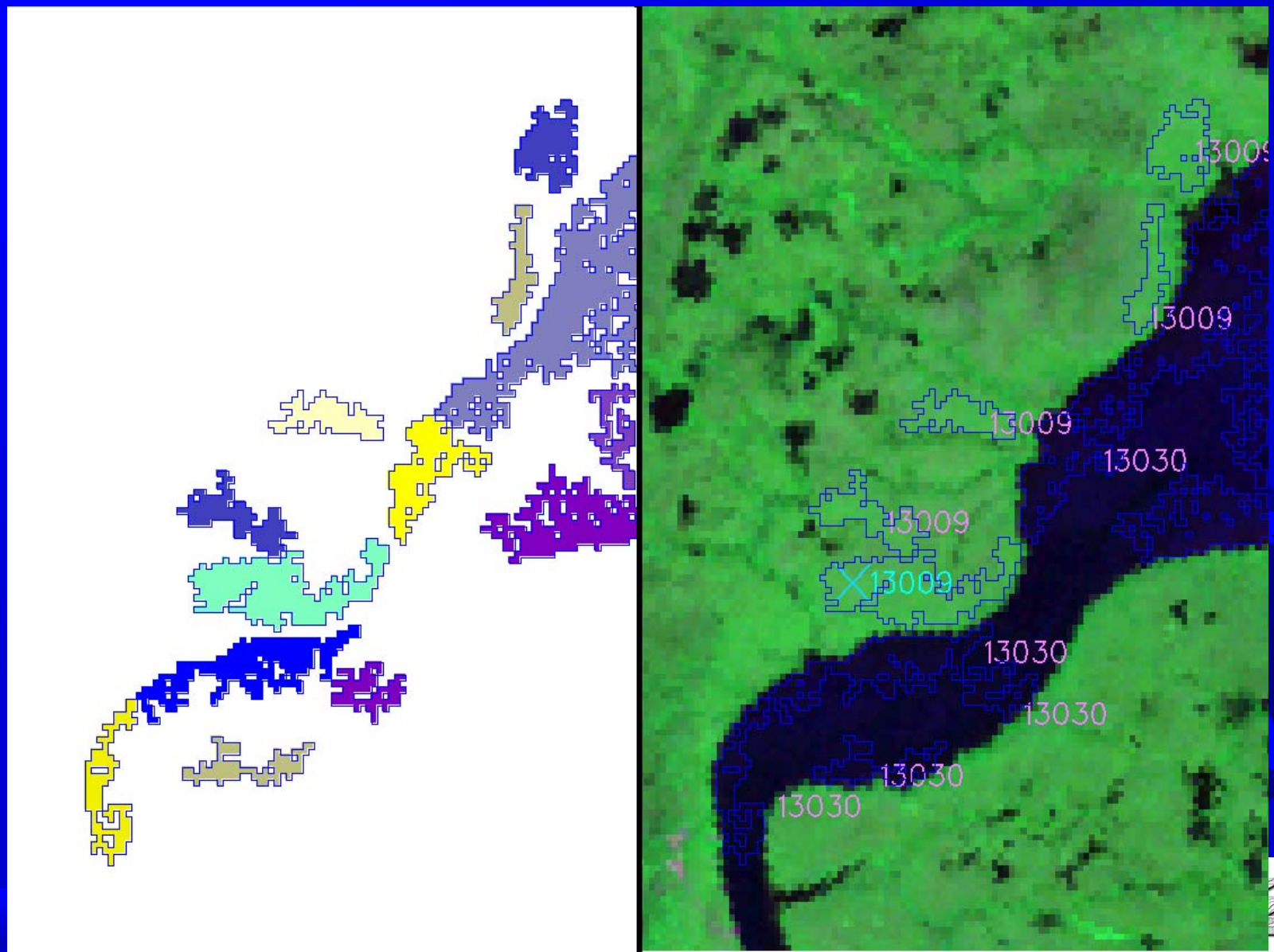
Generate GIS Database

- Reclass all areas (pixels) with size less than the specified minimum size(s) to a value of 0
- Vectorize the remaining areas and identify by unique area number

Reclass Areas ‘Too Small’ to ‘0’



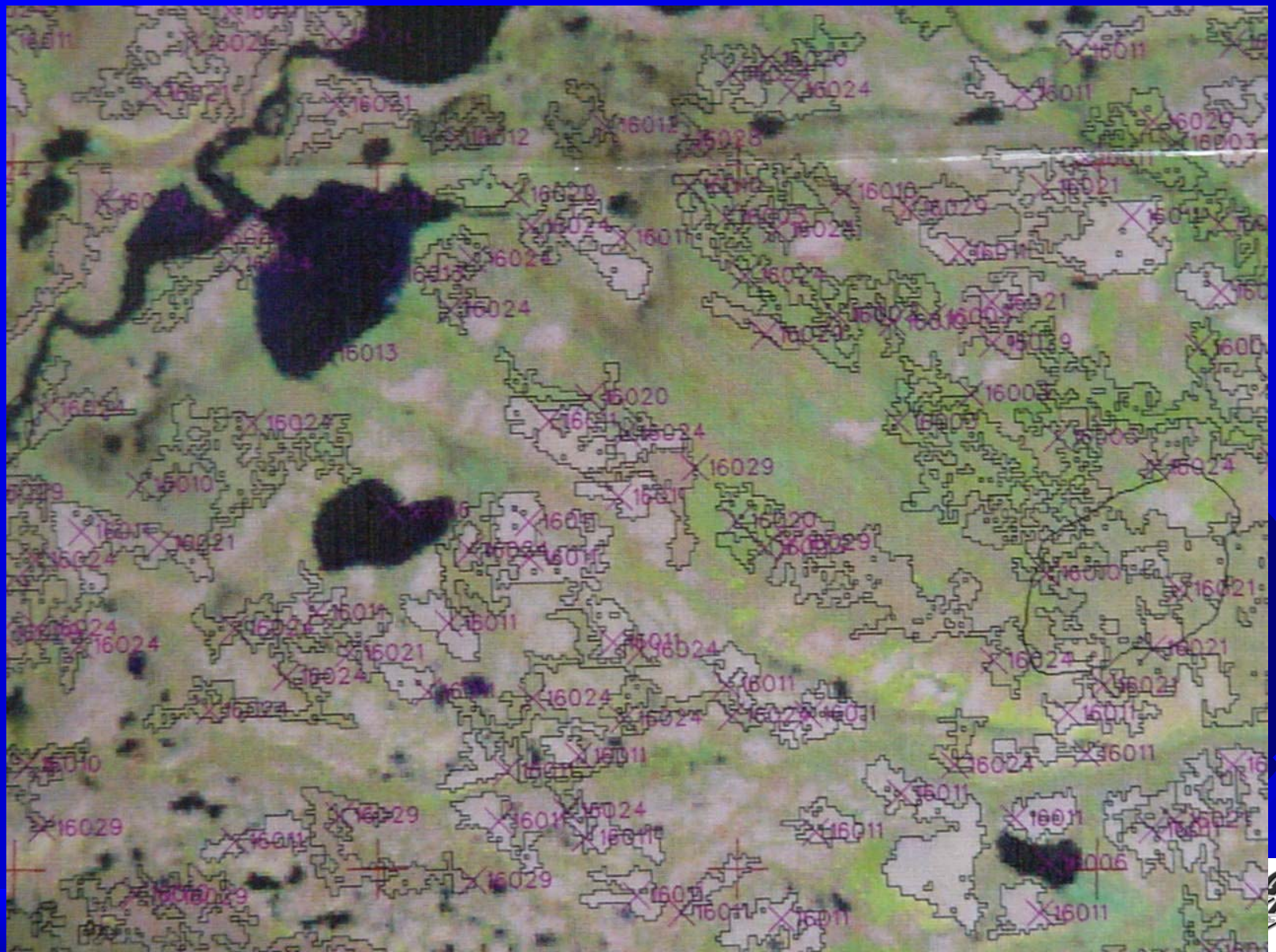
Vectorize and Label Candidate Areas



Candidate Training Site Database Contains ...

- Isodata class value
- Area - number of pixels
- X,Y coordinates
- Slope, aspect, and elevation
- Scene indicator
- Scarcity indicator
- Training group number

Generate Field Maps



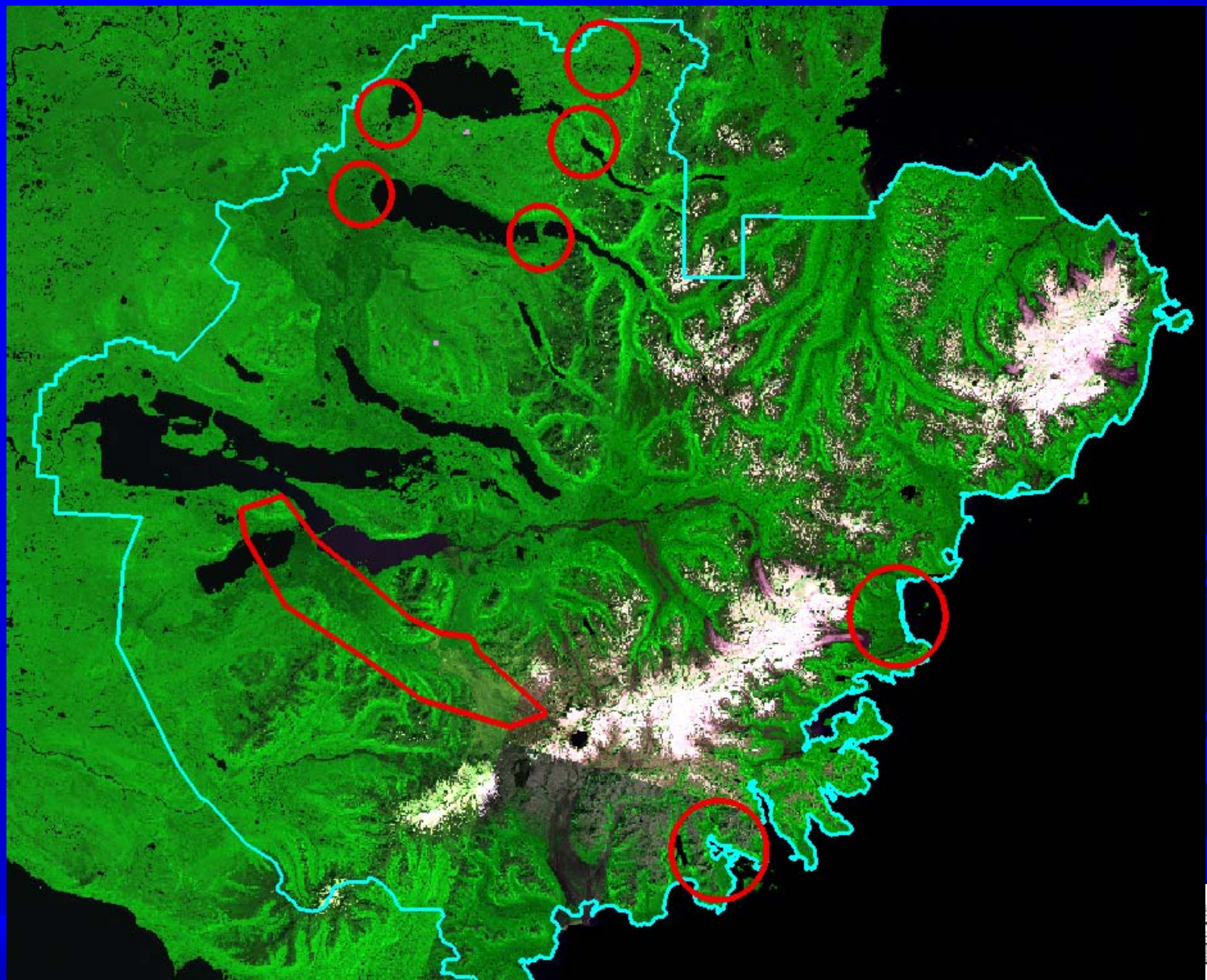
Geographic Resource Solutions

Candidate Site Selection Criteria

- Access
- Distance traveled
- Scarce isodata classes
- Proximity of candidate training sites to each other
- Areas of scene overlap

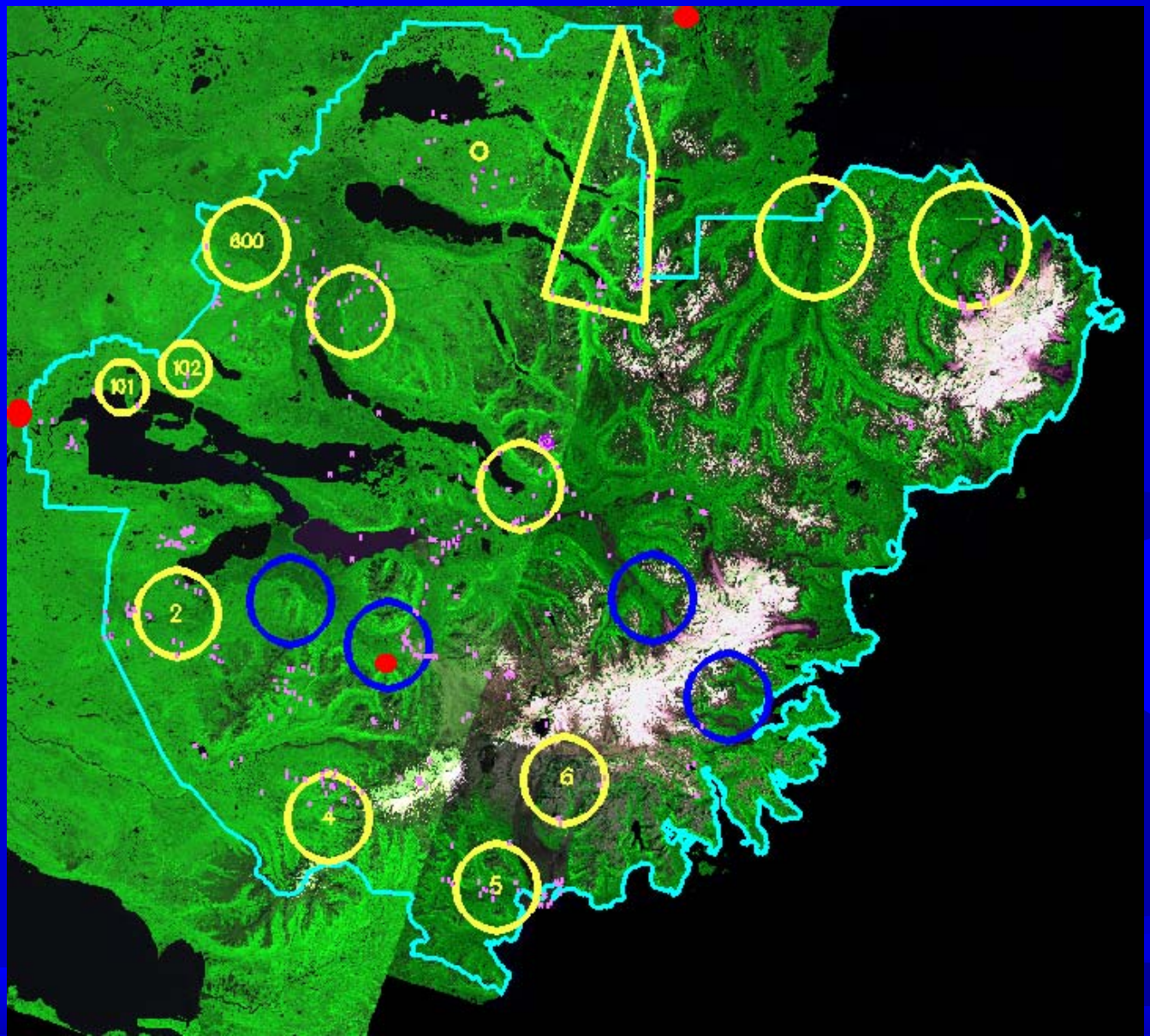
Geographic Resource Solutions

Constraints - ‘No-Fly’ Zones



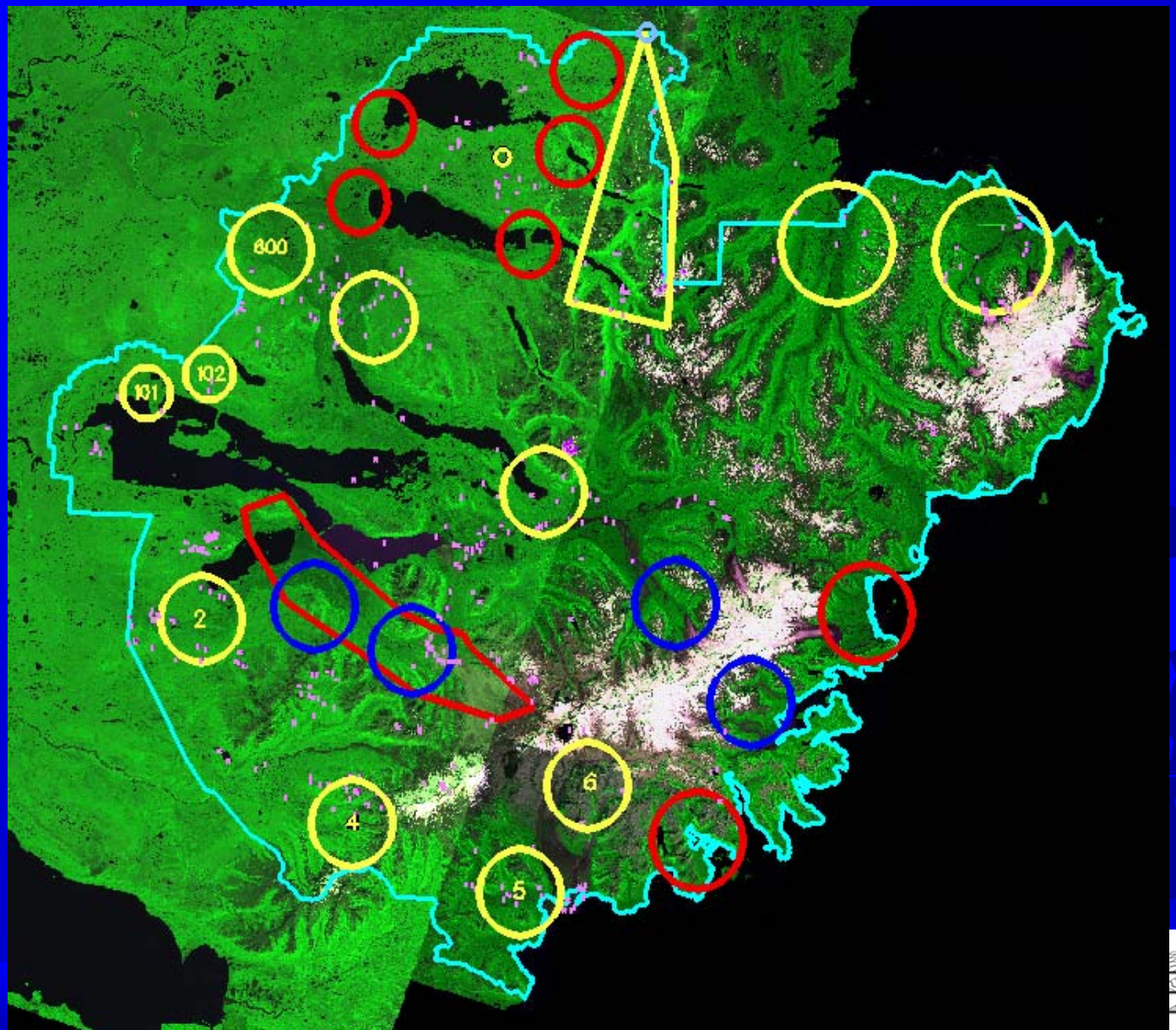
Geographic Resource Solutions

Areas of Interest/Fuel Dumps



Geographic Resource Solutions

‘No-Fly’ Zones AND AOIs



Geographic Resource Solutions

Sampling Plan Development and Administration

- Daily plan development
  - sample scarce isodata classes
  - fulfill daily plan requirements
  - fulfill overall plan requirements
  - obtain multi-scene samples
- Prepare field maps
- 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 Schedule

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

Sampling 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

Sampling 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

Geographic Resource Solutions

Results and Benefits

- Better signatures - less confusion
- Fewer rejected areas
- Project area has been sampled
  - significant types
  - scarce types
  - project area variation
- Less speculation/seat-of-pants judgement
- Lower cost and/or less time