Accuracy Assessment, "The Emperor's New Clothes", and Other Fairy Tales

Ken Stumpf
Geographic Resource Solutions
Arcata, CA
stumpfk@grsgis.com
www.grsgis.com

Presented at ASPRS Conference 2012 , Sacramento, CA. March 22, 2012

Accuracy Assessment Design Problems

- Past ASPRS Conference Presentations
- Land Cover Mapping Project Reports
- Requests for Proposals
- User Interaction/Discussion

I found myself exasperated at the lack of understanding of statistics and sampling design involving Accuracy Assessment!

Geographic Resource Solutions

“The Emperor Has No Clothes!”

I am here to play the part of the innocent young boy who cried out

“The Emperor Has No Clothes!”

regarding

“Withheld training site data being used as the basis of an Accuracy Assessment”

GEOGRAPHIC RESOURCE SOLUTIONS™

Use of Training Data for an Accuracy Assessment?

Rationale ????

- "It's too difficult or costly to do it right !"
- "What's wrong with it ?"
- "But everybody else does it ?"
- "I'm new to the project and was told to do it this way!"

Lillesand & Kiefer Say Otherwise …

Note that the error matrix in Table 7.3 indicates an overall accuracy of 84 percent. However, producer's accuracies range from just 51 percent ("urban") to 100 percent ("water") and user's accuracies vary from 72 percent ("sand") to 99 percent ("water"). Furthermore, this error matrix is based on training data. *It should be remembered that such procedures only indicate how well the statistics extracted from these areas can be used to categorize the same areas! If the results are good, it means nothing more than the training areas are homogeneous, the training classes are spectrally separable, and the classification strategy being employed works well in the training areas. This aids in the training set refinement process, but it indicates little about how the classifier performs elsewhere in a scene. One should expect training area accuracies to be overly optimistic, especially if they are derived from limited data sets.* (Nevertheless, training area accuracies are sometimes used in the literature as an indication of overall accuracy. They should not be!)

(Lillesand & Kiefer, 1994)

Hunsaker Says … Optimistic Bias …

"Note that there are a number of studies in the remote sensing where the reported classification accuracy is based on back-classification of the training data. *This procedure can strongly bias results towards an optimistic assessment*."

.

.

.

"A more valid approach, therefore is to partition available data for training and testing at the stand or site level. That is, *randomly sample entire stands* from the elements …"

(Hunsaker, 2001)

GEOGRAPHIC RESOURCE SOLUTIONS

Classification Assessment is NOT an Accuracy Assessment

- **Classification Accuracy describes how accurate your training data are at reclassifying the map.**
- **Classification Accuracy is a measure of the validity of your training data statistics.**
- **Classification Accuracy indicates how much confusion is present in your training data set.**
- **Classification Accuracy should be close to 100% as a lower value indicates there are other confused "map types" that share the same spectral statistics as classes in your training data set.**

GEOGRAPHIC RESOURCE SOLUTIONS

Consider 2 Training Areas

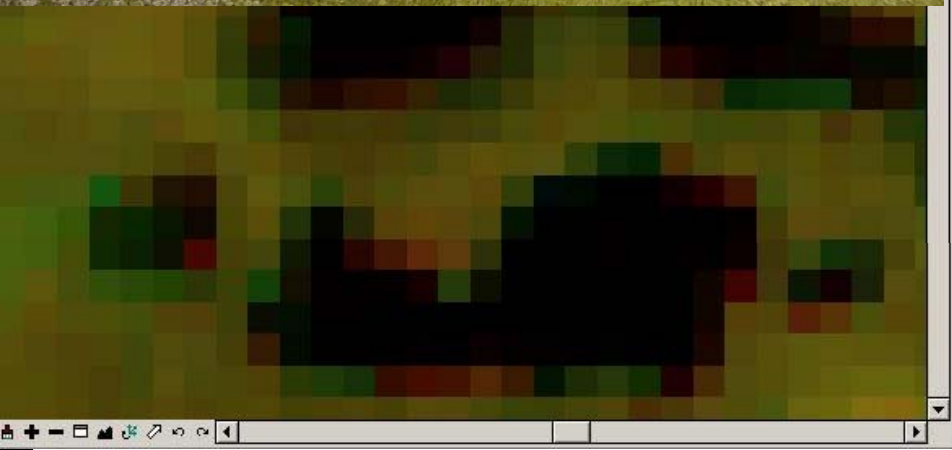
- One represents a senescent wet marsh
- One represents an open White Spruce Forest

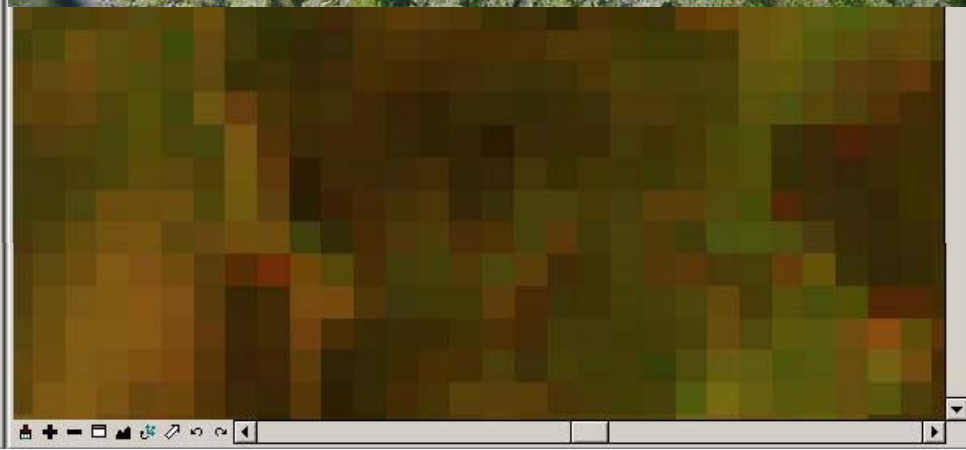
Now I happen to know based on having processed all these spectral data that these two sites have spectral statistics that are very confused ! Virtually the same means and very similar variances band for band …

Geographic Resource Solutions™


12 12:07PM


16 11:37AM





If we only include the marsh in the training data set and not the White Spruce Forest area …

GEOGRAPHIC RESOURCE SOLUTIONS™


Marsh
71257
PicGla Open Forest
70822

If we only include the marsh in the training data set and not the White Spruce Forest area …

Marsh pixels will placed in the White Spruce forest training area !

GEOGRAPHIC RESOURCE SOLUTIONS


Marsh
71257
PicGla Open Forest
70822

We should include the marsh in the training data set and the White Spruce Forest area …

Marsh pixels will placed in the White Spruce forest training area and vice versa as these two sites are confused !

However, we want them to be classified correctly as marsh and forest …

GEOGRAPHIC RESOURCE SOLUTIONS


Marsh
71257
PicGla Open Forest
70822

How Do We Know They Are Confused ?

- We don't.
- If we leave the White Spruce forest site out of the training data set, then it does not show up confused with the Marsh site.
- We do not realize it is confused until after we make our land cover map and check the left out White Spruce site against the map and find out the area mapped as marsh should be a White Spruce forest area.

Confusion Report

Trsite_id:	71257		Hrb	Wet Herbaceous	0	0	0	0	60	25	15

Training Set	Confused trsite_id	J-M Dist	Cover Type	Calculated Class	Tree Cvr	Pct Conf	Shr Cvr	Dsh Cvr	Forb Cvr	Bar Cvr	Oth Cvr

How Do We Know They Are Confused ?

- If we include both sites in our training data set the confusion between the two sites is identified in the confusion report.
- We correct the confused training sites so that we can properly maps these different land cover types.
- We develop a map that is more correct than if we withheld training data.

GEOGRAPHIC RESOURCE SOLUTIONS™

Confusion Report

Trsite_id:	71257		Hrb	Wet Herbaceous	0	0	0	0	60	25	15
Training Set	Confused trsite_id	J-M Dist	Cover Type	Calculated Class	Tree Cvr	Pct Conf	Shr Cvr	Dsh Cvr	Forb Cvr	Bar Cvr	Oth Cvr
6467aSup	70822	0.973	PGl	White Spruce:Open	32	100	25	5	7	9	22

This report identifies “bad” confusion that indicates spectral confusion between different types. This needs to be fixed before making our map.

GEOGRAPHIC RESOURCE SOLUTIONS

"But I randomly selected my Accuracy Assessment sites from my training sites …"

- The random selection of stands from the map would be the equivalent of throwing darts at the map to select your Accuracy Assessment sample sites. Stands hit by darts would be sampled.
- The random selection of training sites would be the equivalent of throwing darts at a list of homogeneous training sites to select your sample sites. This list has nothing to do with the map.

The random selection of homogenous training site stands do not reflect the overall nature of the map!

GEOGRAPHIC RESOURCE SOLUTIONS

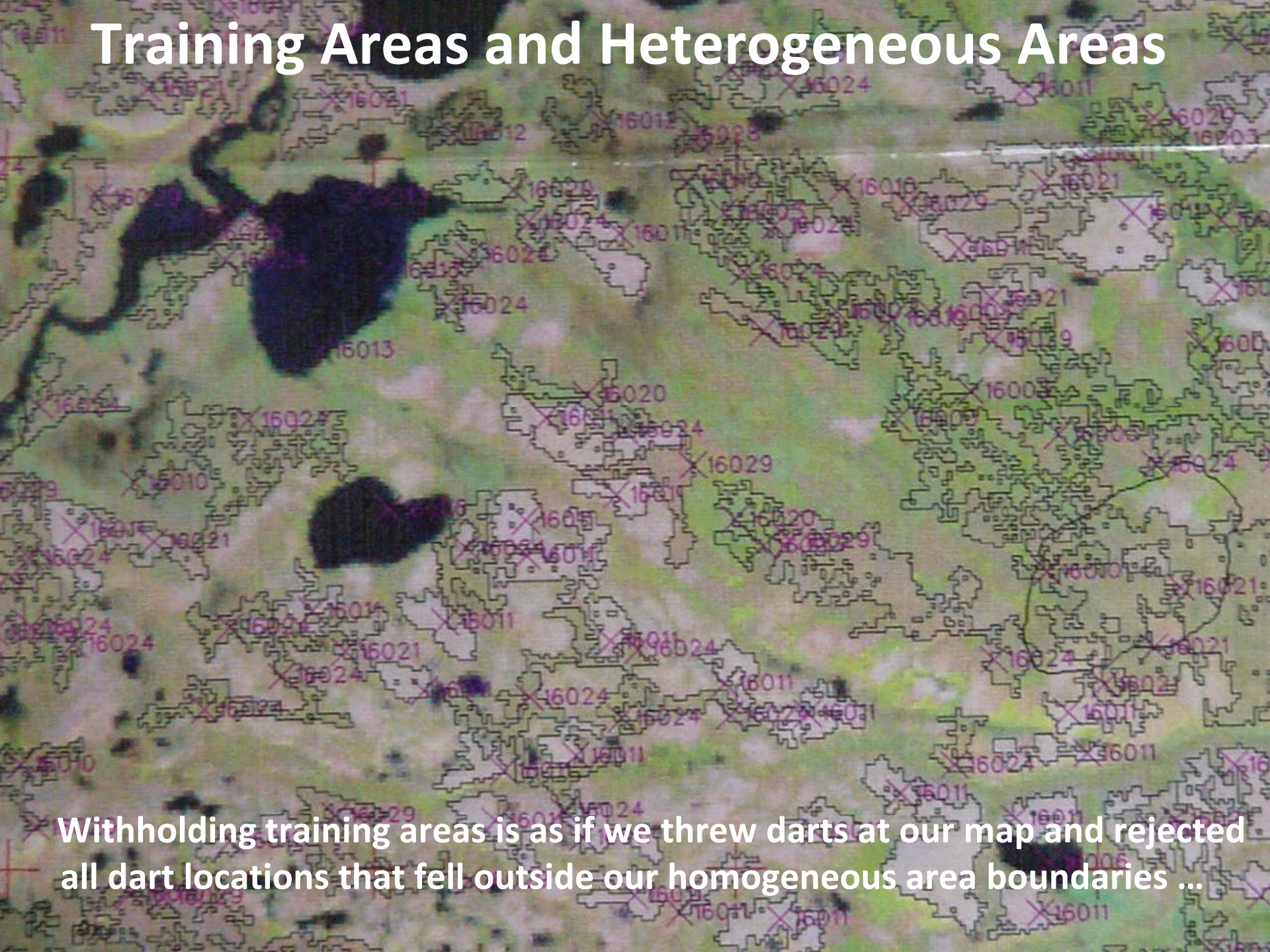
Not All Stands Are Training Areas

- Training areas do not represent the entire map. The entire map is comprised of stands or areas that are either spectrally homogeneous and heterogeneous.
- During a stratified random sample of the map all stands (in a stratum) have a probability proportionate to their area of being selected as part of the sample.
- Training areas are a subset of all stands and they have been selected based upon their homogeneity.

When do we sample the stands that would all be rejected as training areas ?

GEOGRAPHIC RESOURCE SOLUTIONS

Training Areas and Heterogeneous Areas

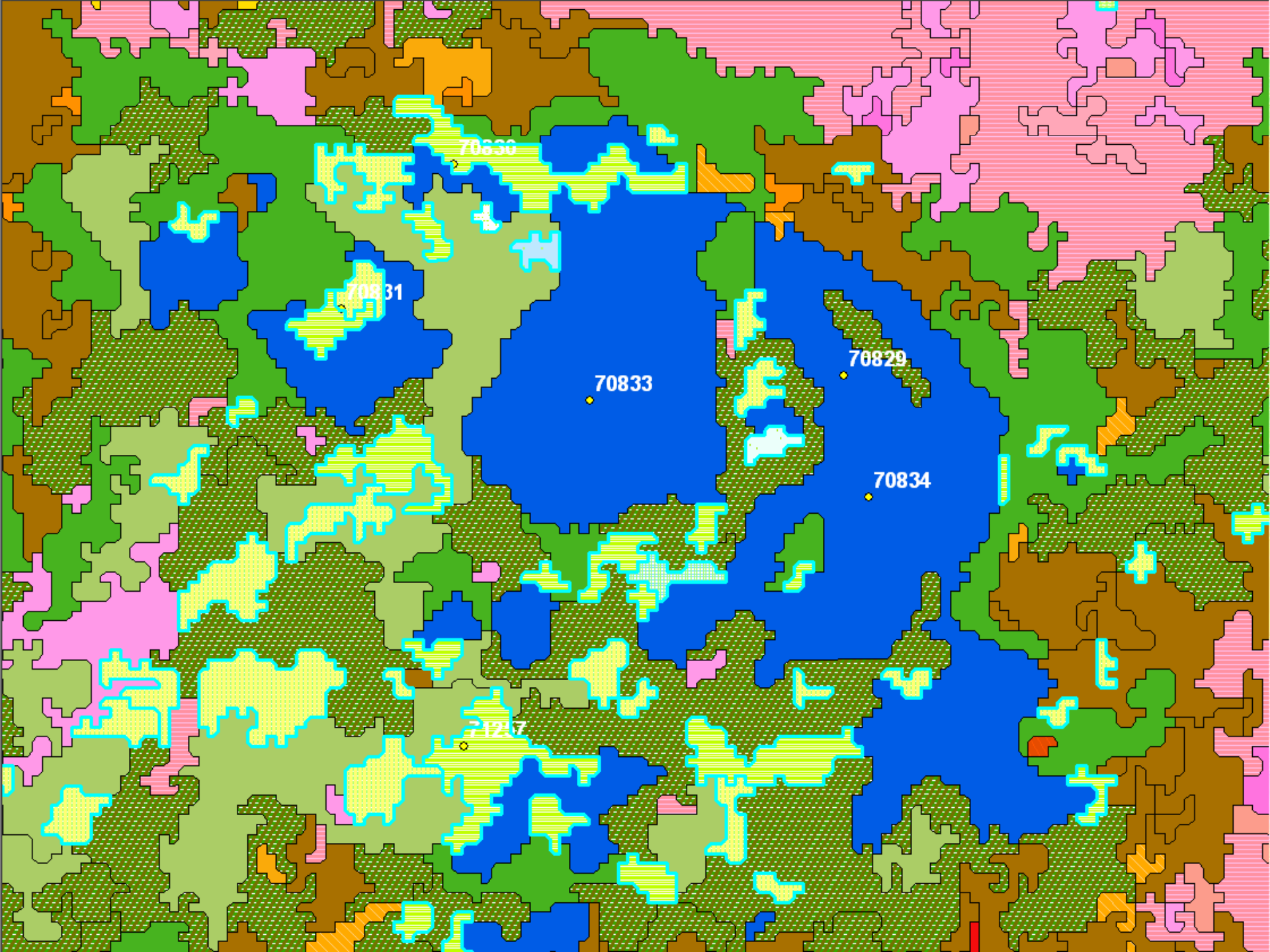


Withholding training areas is as if we threw darts at our map and rejected all dart locations that fell outside our homogeneous area boundaries …

Not All Training Sites Are Stands

- **The homogeneous area selected for sampling as a training site may only comprise a (small) portion of a larger stand.**

The smaller training area was not selected randomly, but picked because it had the characteristics of a GOOD Training Site, whereas the entire stand may include more variability that is not tested.


70830
70831
70833
70829
70834

“Do Not Use Training Areas in an Accuracy Assessment”

Randomly withholding sites from a biased sample of homogeneous training sites simply results in a smaller biased set of sample sites.

GEOGRAPHIC RESOURCE SOLUTIONS™

"Do Not Use Training Areas … Except"

- **This approach may be okay within maps developed using segmentation, if**
  - all stands boundaries are developed first and
  - Resulting stands are randomly sampled to determine their ground-truth characteristics based solely upon their participation in a particular "object class," that is unknown at the time of sample selection.
- **This one approach does <u>not</u> include selecting training sites from only homogeneous areas, but rather, from all areas of the map.**

GEOGRAPHIC RESOURCE SOLUTIONS

“Why Grandma, What ‘Nearby’ Sites You Sample ?”

- **Sample only those areas that are easily accessible.**
  - Only sample sites close to roads, trails, and other means of access.

Geographic Resource Solutions™

“Why Grandma, What Gentle Sites You Sample ?”

- **Sample only those areas that are easily accessible.**
  - Only sample sites close to roads, trails, and other means of access.
- **Sample only those sites on slopes less than 40%.**

GEOGRAPHIC RESOURCE SOLUTIONS™

"Why Grandma, What Biased Sites You Sample ?"

- **Sample only those areas that are easily accessible.**
  - Only sample sites close to roads, trails, and other means of access.
- **Sample only those sites on slopes less than 40%.**
- **Sample only the interior/center of an area.**

GEOGRAPHIC RESOURCE SOLUTIONS™

"Why Grandma, What Generalized Data You Collect ?"

- **Sample only those areas that are easily accessible.**
  - Only sample sites close to roads, trails, and other means of access.
- **Sample only those sites on slopes less than 40%.**
- **Sample only the interior/center of an area.**
- **Develop just a "type call" at the site.**
  - Do not collect data to support a questionable "type call"
  - Do not develop alternative "type calls" when borderline or threshold situations are encountered.

GEOGRAPHIC RESOURCE SOLUTIONS

"You're Not My Grandma !!!"

- **If sample sites are not representative of the mapped area, then the Accuracy Assessment may not be a representative sample of the mapped area.**
  - Need to sample the same or more area than where training data were collected.
- **If the appropriate data are not collected at the AA sites, then the ability to determine whether or not the field AA site information matches the map information may be misleading or incomplete.**
  - Need to collect data and develop alternative "type calls" where field crews have questionable answers for decision rule thresholds.
  - Do we have a "Match" or not ?

Geographic Resource Solutions

That Beautiful Red Apple May Be Poisonous !

- **Error Matrix**
  - % Correct describes the expected value of the probability of randomly selecting any map polygon from a list of polygons regardless of area.
- **Error Matrix - Area Weighted**
  - % Correct describes the expected value of the probability of randomly selecting any unit of area (hectare, acre, and so forth).
- **Visual representation of Mapping Error(s)**
  - Add % Correct as a database table item and populate with the values for different types.

GEOGRAPHIC RESOURCE SOLUTIONS

Accuracy Assessment Error Matrix

- Based on treating all stands as equal observations.
- Record the observation as “1” in the appropriate cell of the matrix.
- Area or weighting of strata are not considered.

Geographic Resource Solutions™

Land Cover Type Error Matrix

		REFERENCE DATA									
		BA	CH	DH	EH	GF	MC	SC	TF	WA	TOTAL	PERCENT CORRECT
MAP DATA	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

GEOGRAPHIC RESOURCE SOLUTIONS™

Let Results Reflect Stratification

- Add area to the matrix.
- Apply percent correct mapped to stratum area to generate estimates of “area correctly mapped.”
- Overall percent correct is now weighted by area of the different strata.

Geographic Resource Solutions™

Area Weighted Land Cover Type Error Matrix

		REFERENCE DATA											
		BA	CH	DH	EH	GF	MC	SC	TF	WA	TOTAL	PERCENT CORRECT	ACRES	CORRECT ACRES
MAP DATA	BA	2									2	100%	3,871	3,871.0
	CH		5		1						6	83%	230,481	191,299.2
	DH		2	4				1			7	57%	32,331	18,428.7
	EH		3		2						5	40%	33,410	13,364.0
	GF					20					20	100%	41,035	41,035.0
	MC						14		1		15	93%	193,733	180,171.7
	SC					1		16			17	94%	21,429	20,143.3
	TF							1	14		15	93%	37,480	34,856.4
	WA	1								4	5	80%	1,391	1,112.8
	TOTAL	3	10	4	3	21	14	18	15	4	92		595,161	504,282.1
	PERCENT	67%	50%	100%	67%	95%	100%	89%	93%	100%		88%		

Total Percent Correct Acres = 85%

Kappa= 0.8589

GEOGRAPHIC RESOURCE SOLUTIONS

Are Stands Created Equal ?

- **Error matrix based on 1's and 0's does not reflect the area of the stands being tested.**
- **Possible relationship of stand size to error ?**
  - At Lassen National Park, 50% of all error stands were stands <= 2 acres in size.
- **Store acres of stand in Error Matrix, rather than equal values of 1.**

GEOGRAPHIC RESOURCE SOLUTIONS

Land Cover Type Area Error Matrix

		REFERENCE DATA									
		BA	CH	DH	EH	GF	MC	SC	TF	WA	TOTAL	PERCENT CORRECT
MAP DATA	BA	25									25	100%
	CH		105		15						120	88%
	DH		12	88				8			108	81%
	EH		21		23						44	52%
	GF					323					323	100%
	MC						341		7		348	98%
	SC					6		160			166	96%
	TF							9	456		465	98%
	WA	5								88	93	95%
	TOTAL	30	138	88	38	329	341	177	463	88	1692	
	%Correct	83%	76%	100%	61%	98%	100%	90%	98%	100%		95%
									Kappa			.9404

Let Results Reflect Stratification

- Add area to the matrix.
- Apply percent correct mapped to stratum area to generate estimates of “area correctly mapped.”
- Overall percent correct is now weighted by area of the different strata.

Geographic Resource Solutions™

Area Weighted Land Cover Type Area Error Matrix

		REFERENCE DATA											
		BA	CH	DH	EH	GF	MC	SC	TF	WA	TOTAL	PERCENT CORRECT		
MAP DATA	BA	25									25	100%	3,871	3,871.0
	CH		105		15						120	88%	230,481	201,670.9
	DH		12	88				8			108	81%	32,331	26,343.8
	EH		21		23						44	52%	33,410	17,464.3
	GF					323					323	100%	41,035	41,035.0
	MC						341		7		348	98%	193,733	189,836.1
	SC					6		160			166	96%	21,429	20,654.5
	TF							9	456		465	98%	37,480	36,754.6
	WA	5								88	93	95%	1,391	1,316.2
	TOTAL	30	138	88	38	329	341	177	463	88	1692		595,161	538,946.3
	%Correct	83%	76%	100%	61%	98%	100%	90%	98%	100%		95%		
											Total Percent Correct Acres =	91%		
											Kappa	.9404		

View Accuracy Results …

Visualize by incorporating Accuracy values by class into the map data set table attributes.

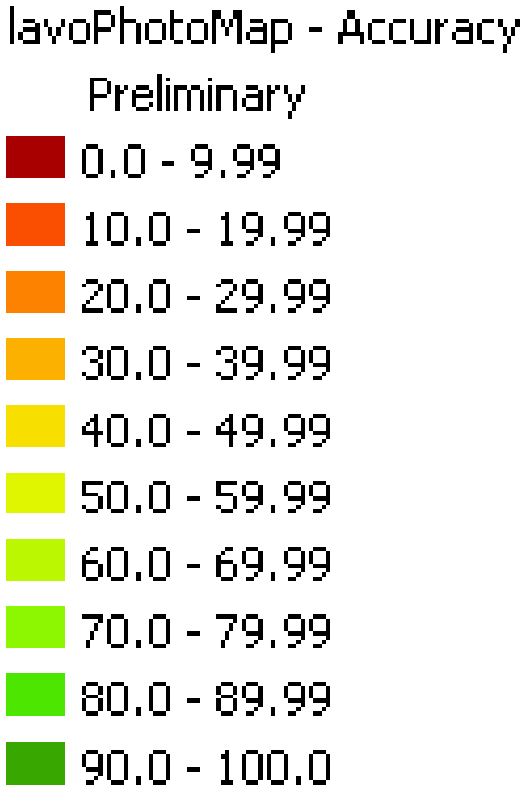
A value of 69 indicates that this class is mapped correctly 69% of the time.

Field	Value
FID	6581
Shape	Polygon
AREA	9725301.11445
PERIMETER	170690.404784
UNIT_ID	7260
TYPE_ID	13
ACREAGE	2403.1599
DENSITY_MO	> 60%
DENSITY_CL	9
ASSOCIATION_GENERALIZED	LupObt:forb
ALLIANCE_GENERALIZED	LupObt:forb
ASSOCVEG_GENERALIZED	
ASSOCIATION_DETAILED	LoPd:forb
ALLIANCE_DETAILED	LoPd:forb
ASSOCVEG_DETAILED	
AAMAP_GENERALIZED	69
AAMAP_DETAILED	69

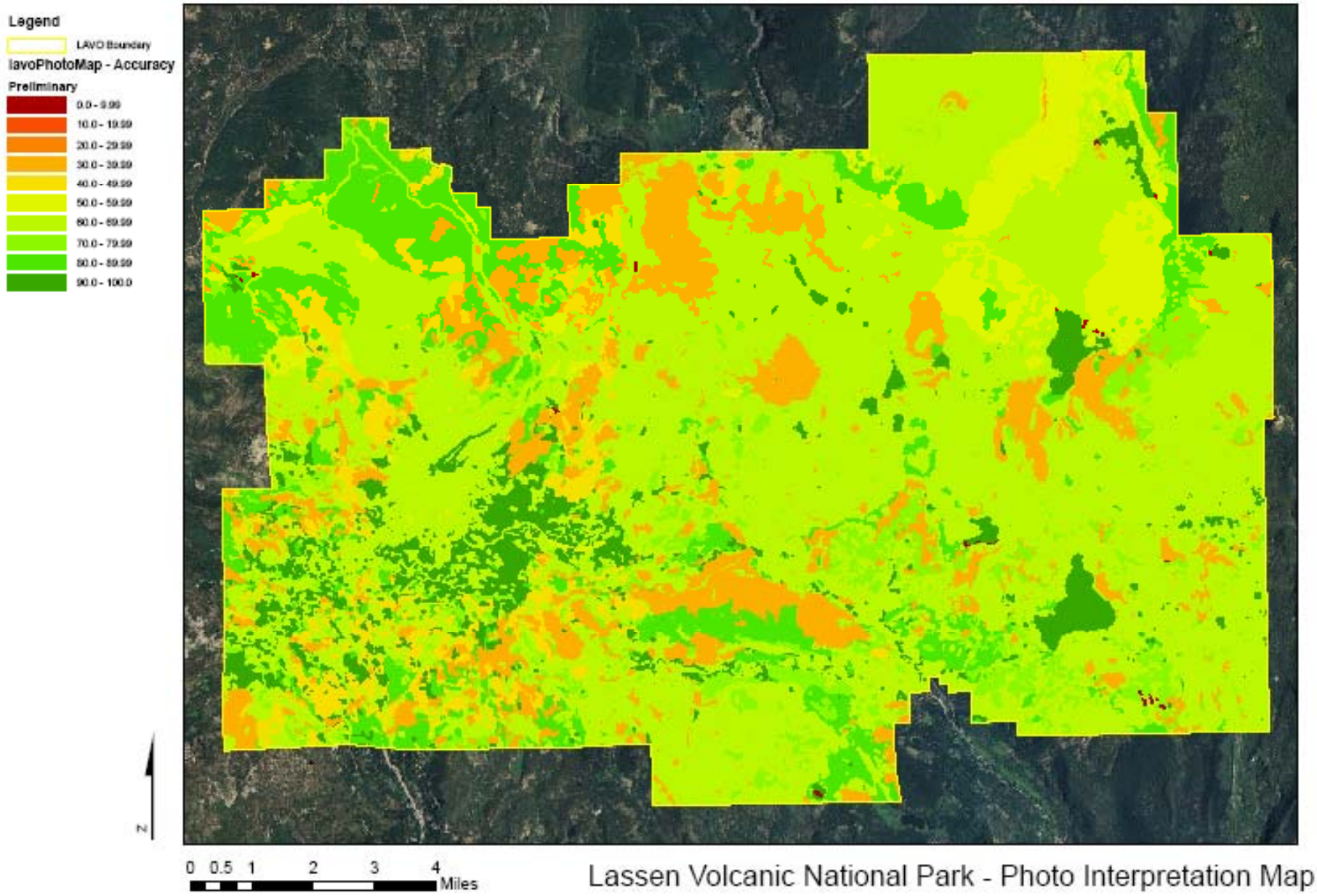
Map

Preliminary Photo Interp Map Accuracy

Accuracy is visually displayed using a color coded legend



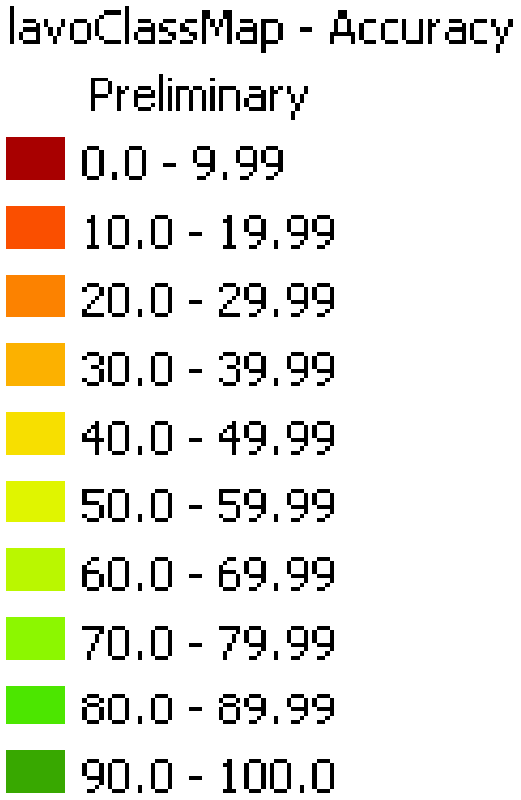
Lassen Volcanic National Park Comparative Mapping Project



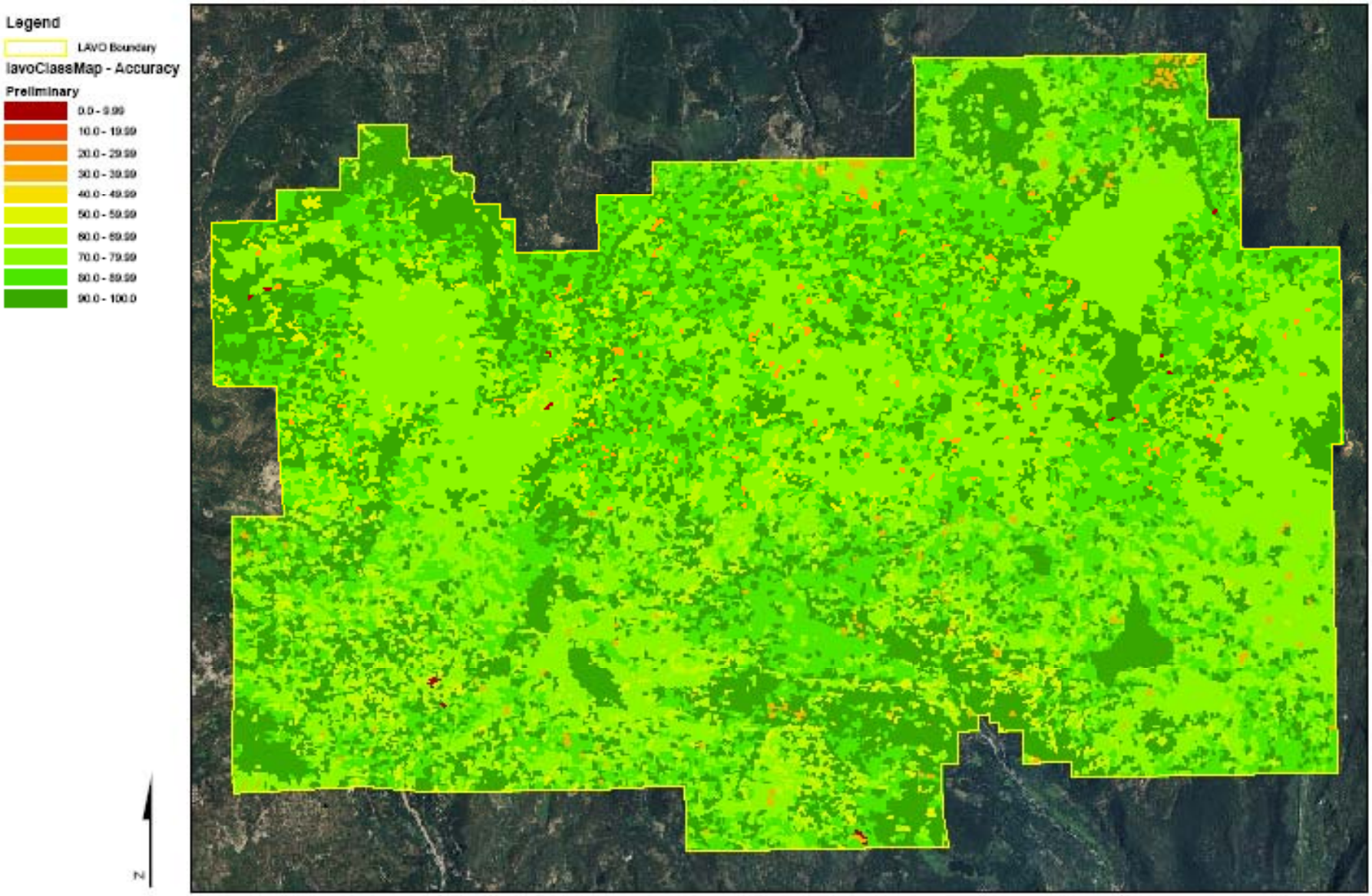
Lassen Volcanic National Park - Photo Interpretation Map

Preliminary Classification Map Accuracy

Accuracy is visually displayed using a color coded legend



Lassen Volcanic National Park Comparative Mapping Project



Lassen Volcanic National Park - Image Classification Map

Which Map Is Visually More Accurate ?

GEOGRAPHIC RESOURCE SOLUTIONS™

Summary

- Hopefully I have made you think about accuracy assessment and how it is used.
- Some things we do lead to optimistic assessments … some may lead to pessimistic assessments.
- At a minimum, we need to stop using biased methods to estimate map accuracy.
- Hopefully, we will see a movement away from Accuracy Assessment methods that do not accurately describe the real accuracy of the maps we develop.

GEOGRAPHIC RESOURCE SOLUTIONS

Questions and Comments

Geographic Resource Solutions™