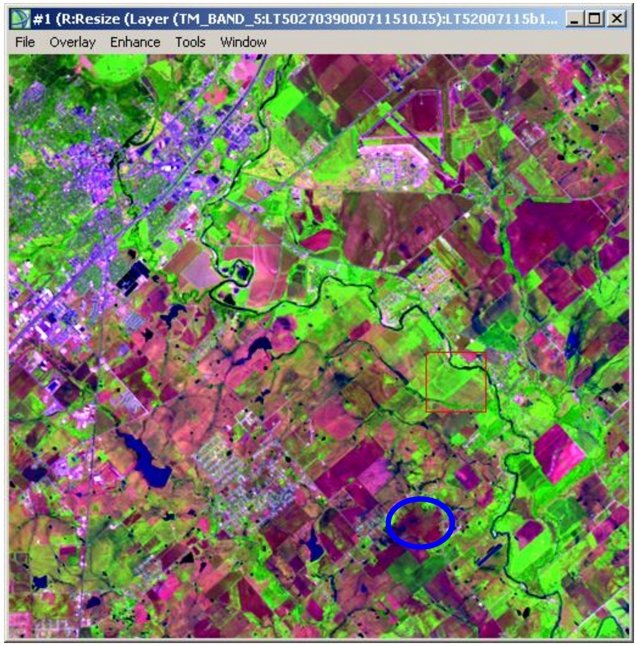
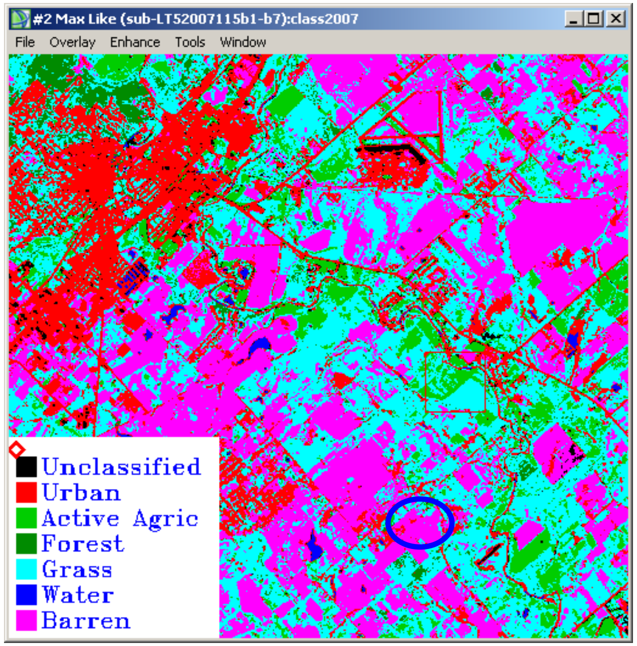
Classification Accuracy Assessment

Lecture 9

1. Introduction: Classification or thematic map

- Remote Sensing is becoming more and more important information/data source, like for GIS and for many applications. Classification is a efficient way extracting information from image.
- The classification map is also called thematic map.





2. Error source of classification

- RS thematic map contains lots of errors, because of:
  - Geometric error
  - Un-complete atmospheric correction
  - Clusters incorrectly labeled after unsupervised classification
  - Training sites incorrectly labeled before supervised classification
  - Un-distinguishable classes
  - None of classification method is perfect
- We should identify the sources of the errors, minimize it, do accuracy assessment, quantify the uncertainties, create metadata before being used in scientific investigations and policy decisions.

3. Accuracy Assessment of Classification

- **Based on ground or known reference pixels**
  - These sites are not used to train the classification algorithm and therefore represent unbiased reference information
  - It is possible to collect some ground sites prior to the classification, perhaps at the same time as the training data
  - But majority of test reference is often collected after classification.
  - Landscape often change rapidly. Therefore, it is best to collect the ground reference as close to the date of remote sensing data acquisition as possible.

3.1 Sampling design

- How many samples do we need?
  1. Based on binomial probability theory:

$$N = \frac{z^2(p)(q)}{E^2}$$

$$= \frac{2^2(85((15)}{5^2} = 203$$

N is the sample size, p is the expected percent accuracy of the entire map, q=100-p, E is the allowable error, and Z=2 (σ) covering 95.4% of image. If your expected accuracy is 85% at an allowable error of 5%, the number of points for a reliable results is 203.

3.1 Sampling design

- How many samples do we need?

2. Based on multinomial probability theory:

$$N = \frac{B\Pi_i(1-\Pi_i)}{b_i^2}$$

Π_i is the proportion of a population in the ith class out of k classes that has the proportion closest to 50%, b_i is the desired precision for this class, B is the upper (α/k) x 100th percentile of the chi square (χ^2) distribution with 1 degree of freedom, and k is the number of classes.

For example, k=8, and we know Π_i is closing 30% of the total population, we desire a level of 95% confidence and a precision (b_i) of 5%,

1- α/k = 1- 0.05/8 = 0.99375, so B= $\chi^2_{(1,\ 0.99375)}$ = 7.568

$$N = \frac{B\Pi_i(1-\Pi_i)}{b_i^2} = \frac{7.568(0.3)(1-0.3)}{0.05^2} = 636$$

samples

So randomly 80 samples per class are required (8 x 80 = 640)

3.1 Sampling design

- How many samples do we need?

3. No idea of the class distribution

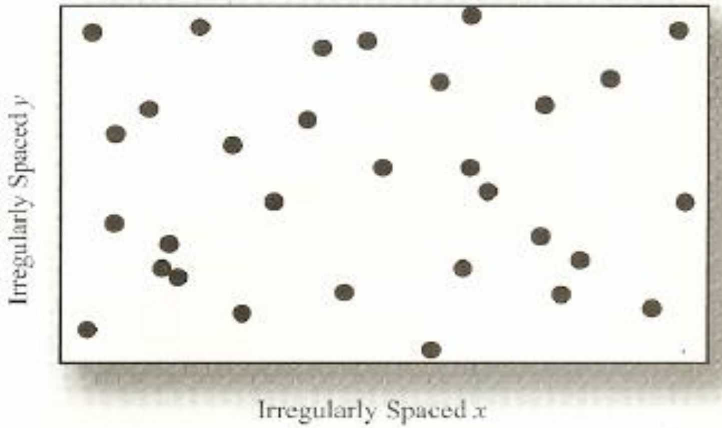
If we have no idea about the proportion of any of the classes in the image, then we can use the worst-case multinomial distribtuion algorithm where we assume that one class occupies 50% of the study area:

$$N = \frac{B}{4 \bullet b_i^2} = \frac{7.568}{4 \bullet (0.05^2)} = 757 \text{ samples}$$

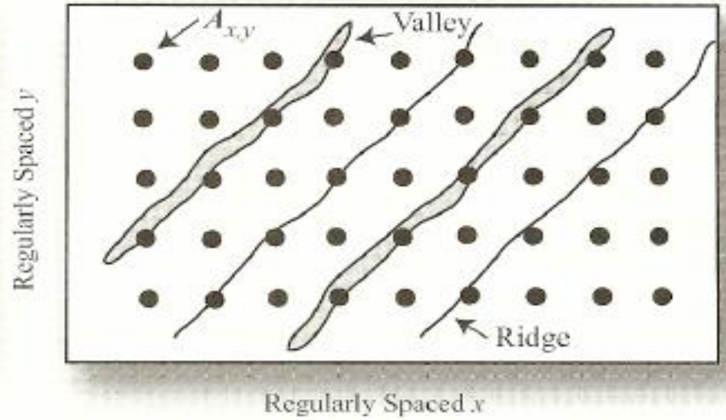
So randomly 95 samples per class are required (8 x 95 = 760)

3.1 Sampling design

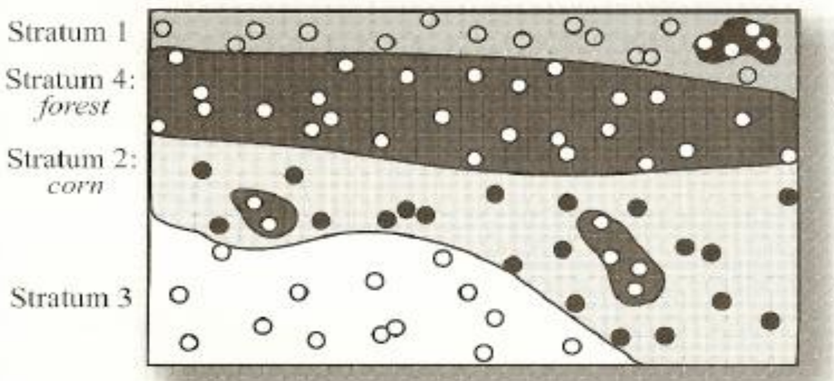
- Where should we go sampling?
  - Random sampling
  - Systematic sampling
  - Stratified random sampling
  - Stratified systematic unaligned samplingd
  - Cluster sampling



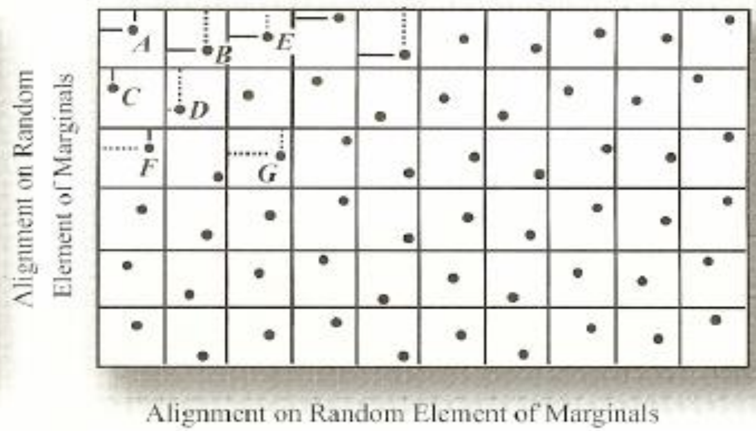
a. Random sampling.



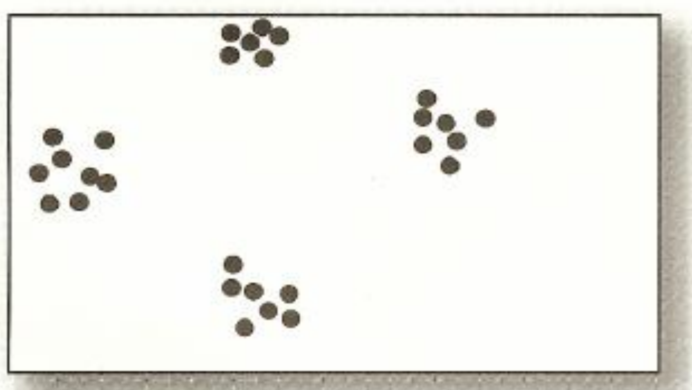
b. Systematic sampling.



c. Stratified random sampling.

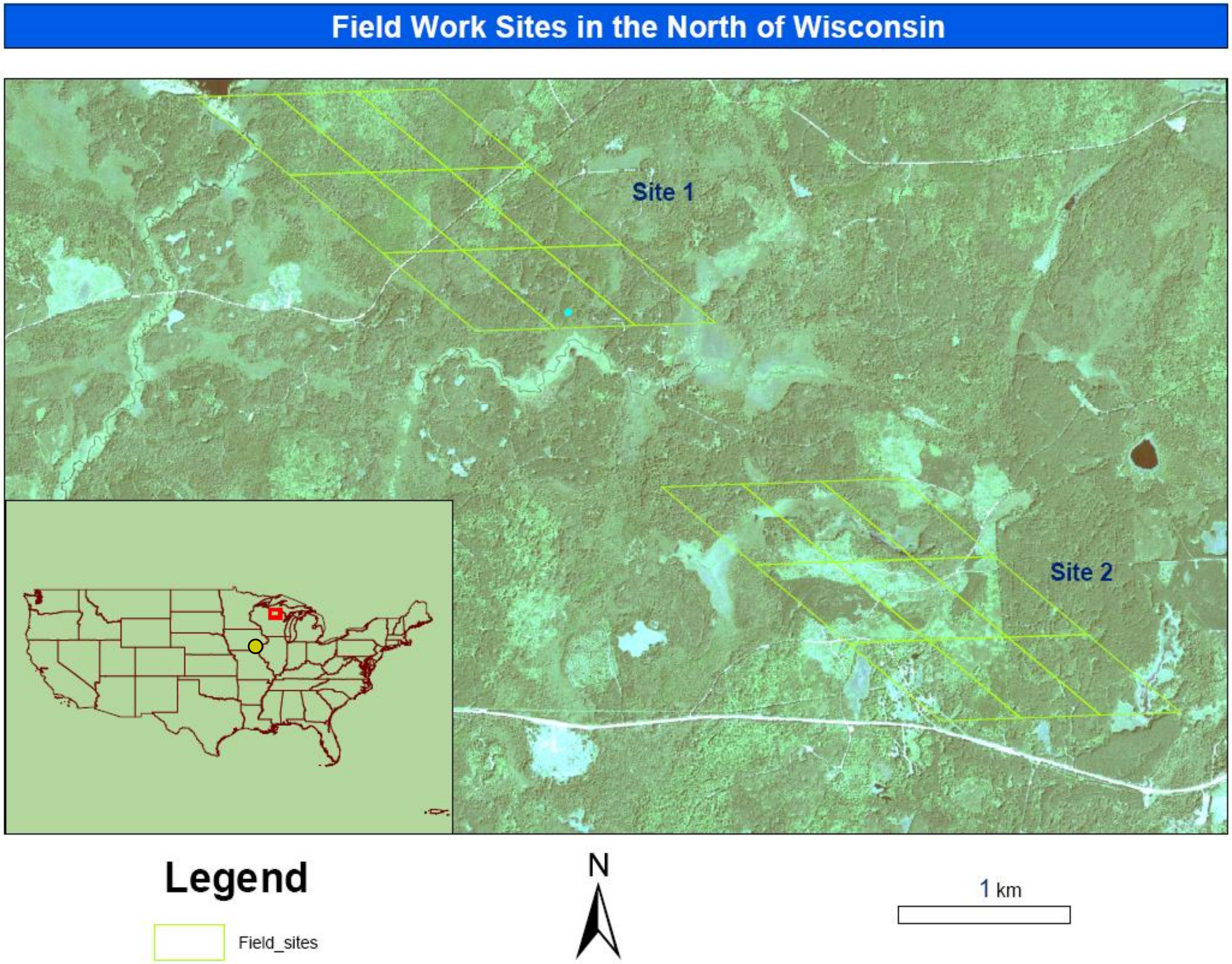


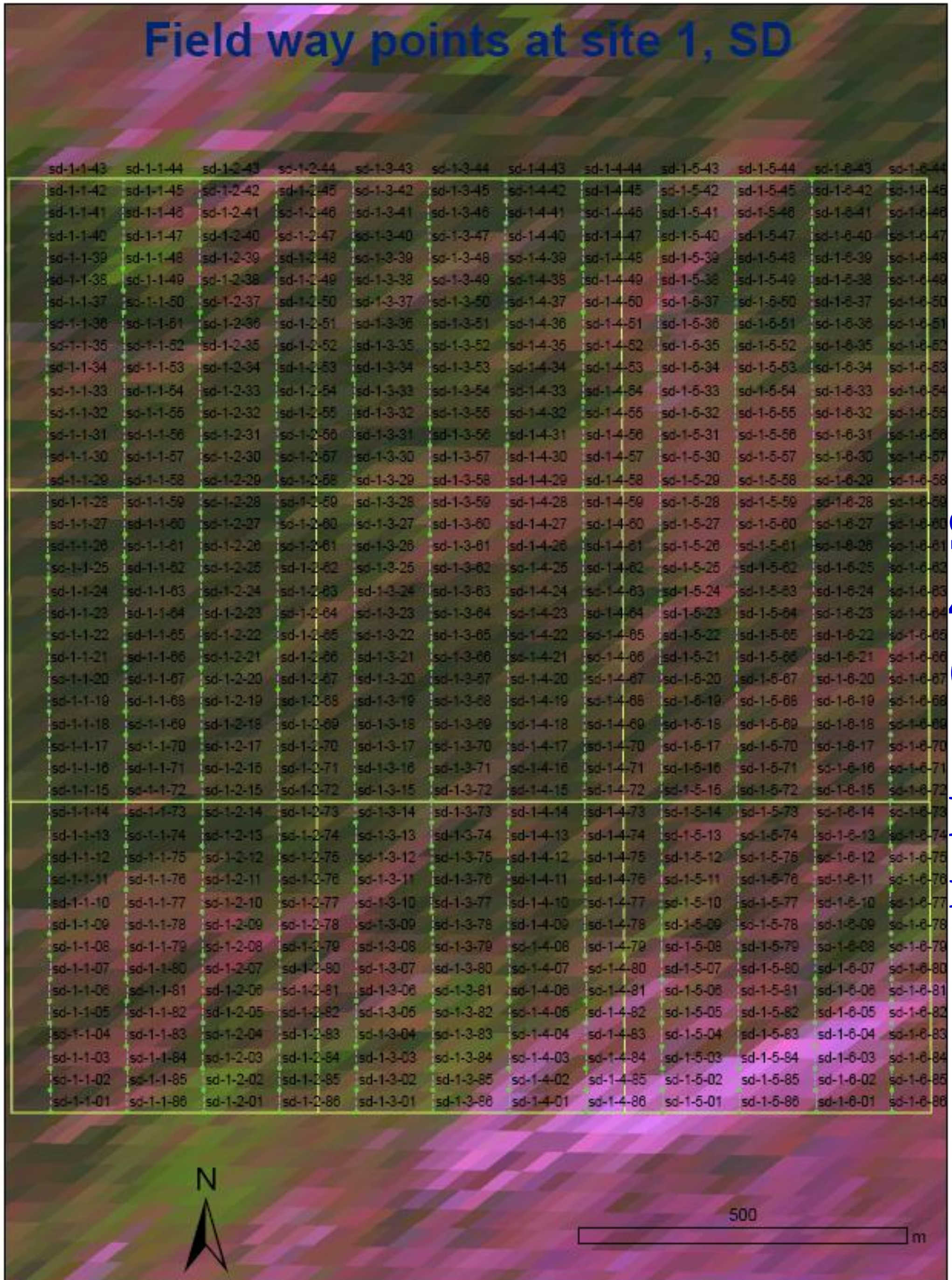
d. Stratified systematic unaligned sampling.



e. Cluster sampling.

3.2 Field sites in the North of Wisconsin





9 MODIS pixels

4 tracts (per pixel)/125m

One measurement/35m

Background is ETM+ image

3.2 Collect the reference classes

1. In situ visit and measurement **(of tree canopy coverage)**

3.2 Collect the reference classes

2. Permanent ground stations

i.g, **SNOTEL station (right): The SNOTEL** network uses sonar and underground scales to measure snow pack at over 660 remote sites throughout the 11 western states installed, operated, and maintained by the National Resources Conservation Service (NCRS), USDA. The stations transmit hourly data to the Internet via radio signals. The SNOTEL data sets are available on the public domain at NRSC's homepage

(http://www.wcc.nrcs.usda.gov/snotel/).

3.2 Collect the reference classes

3. Use high spatial resolution image at almost the same day/week for agriculture, same month or season for forest, desert, and urban area.

- When compare reference image of higher resolution with the classification image of lower resolution, we must choose a threshold to decide the primary class in the reference image.

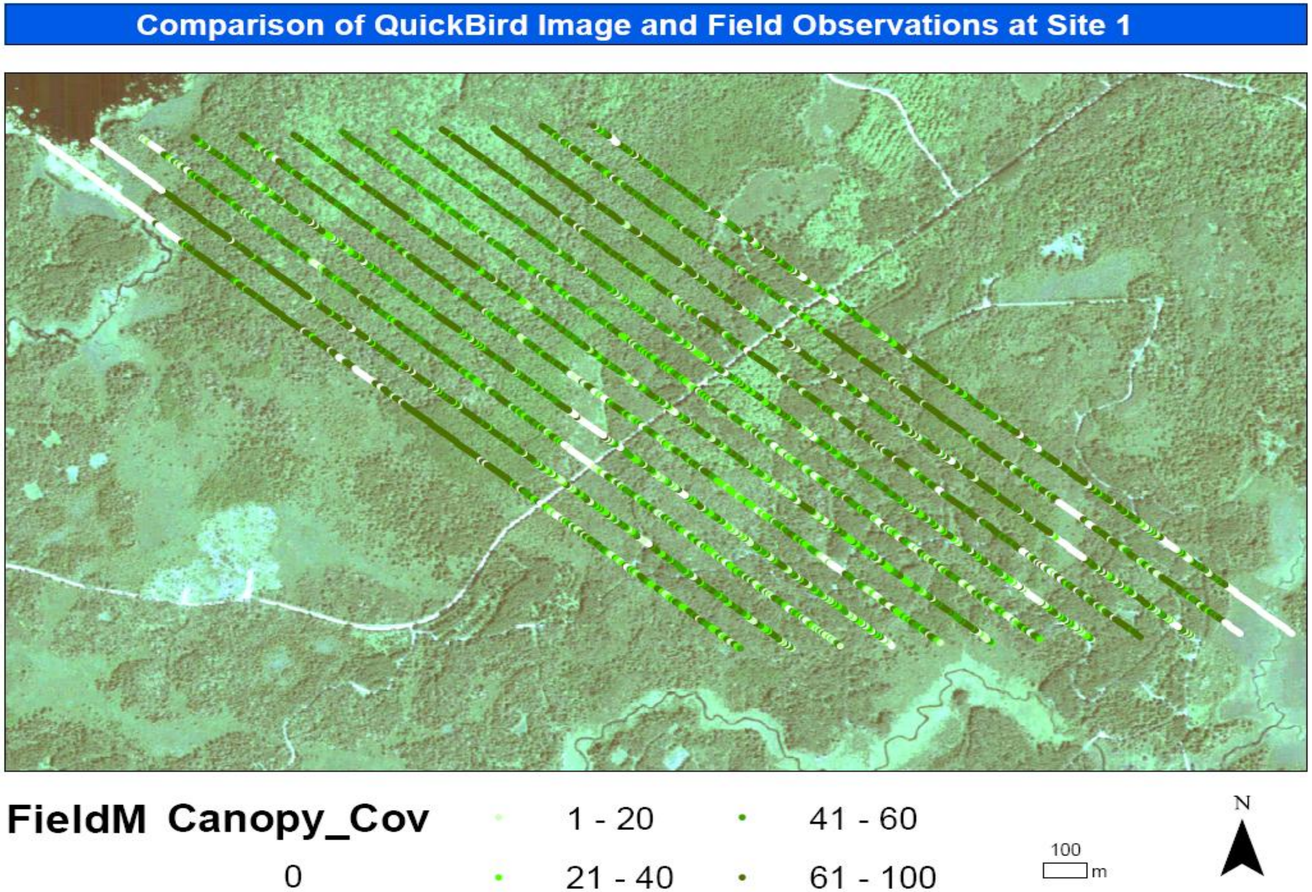
i.e., 1 MODIS pixel(250m)= 64 ETM+ (pixels 30m)

When compare them, we need to aggregate the 64 ETM+ pixels into one MODIS pixel, and a 75% threshold is used to decide the primary class.

3.3 Evaluation of the classification

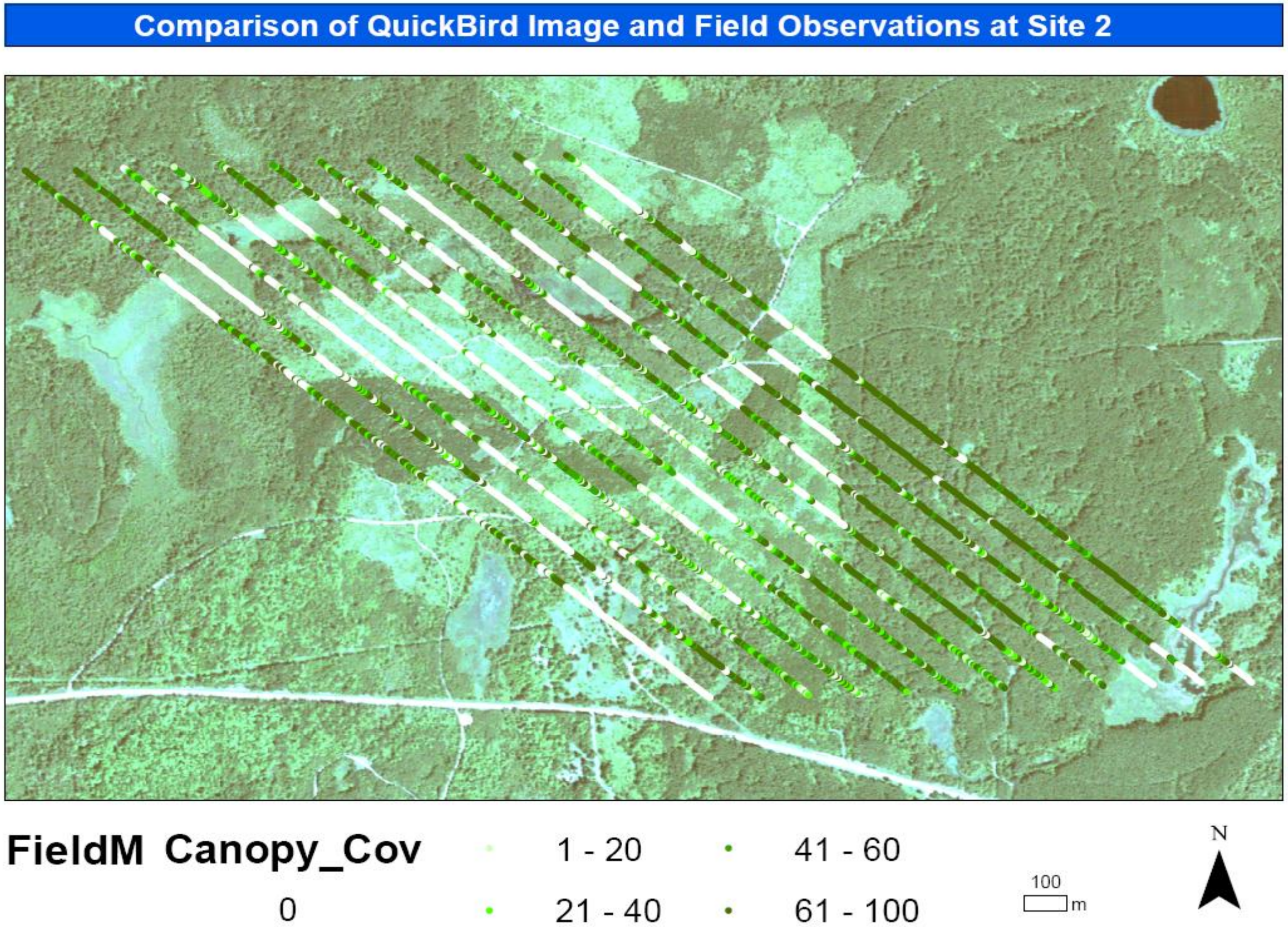
1. Visual/spatial comparison

site 1

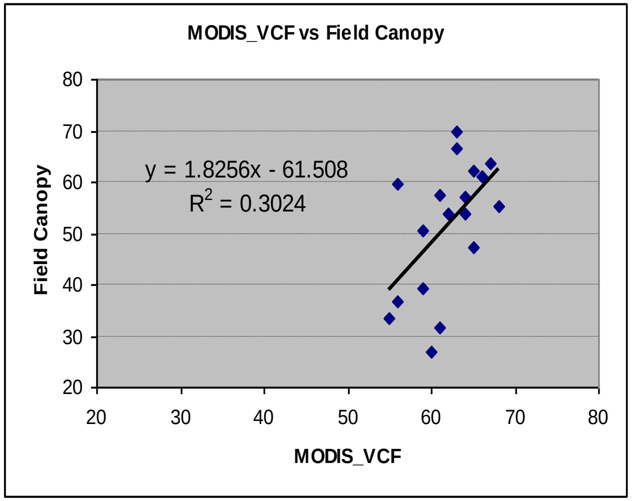


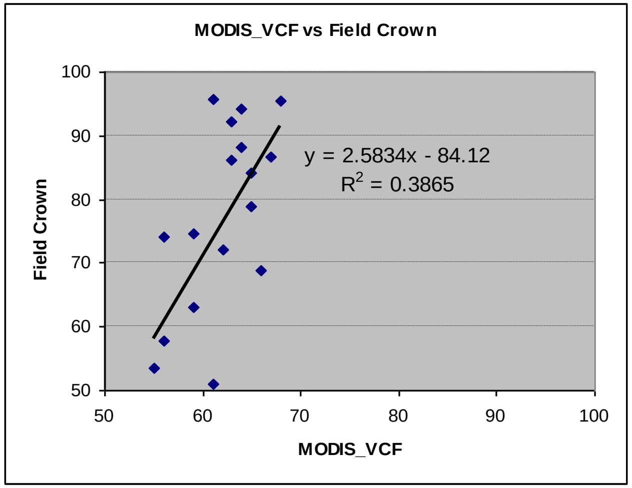
3.3 Evaluation of the classification

1. Visual/spatial comparison **site 2**



MODIS tree canopy and in situ canopy coverage

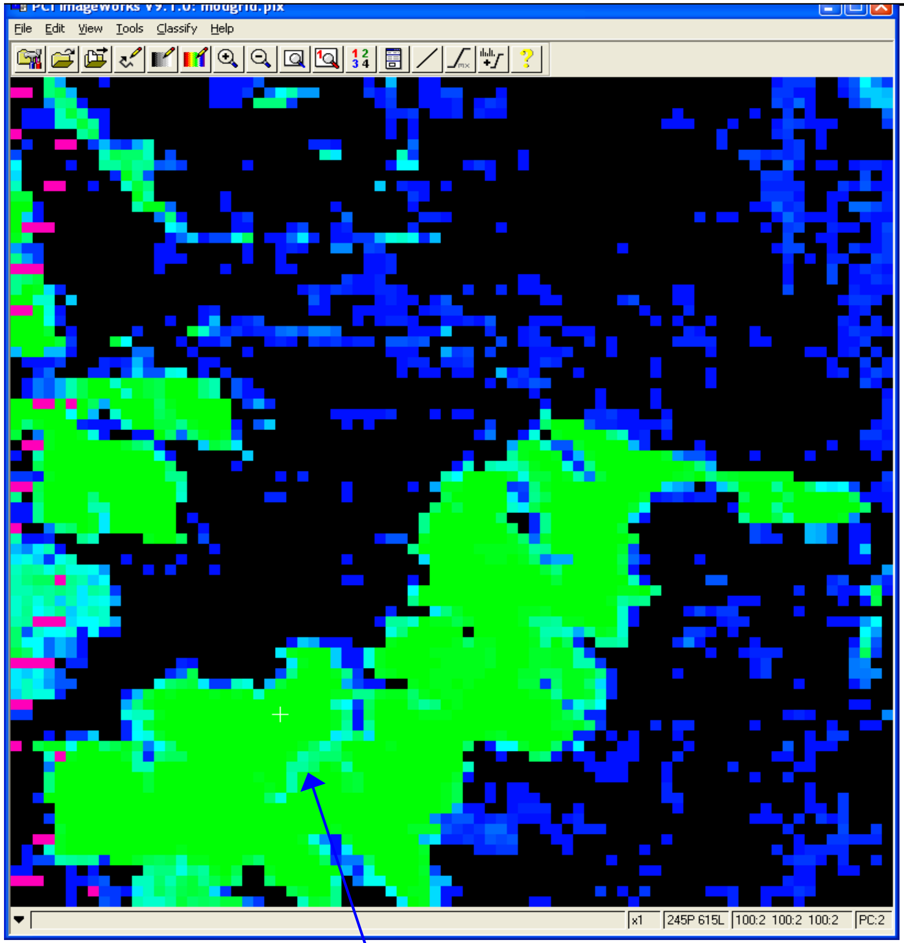




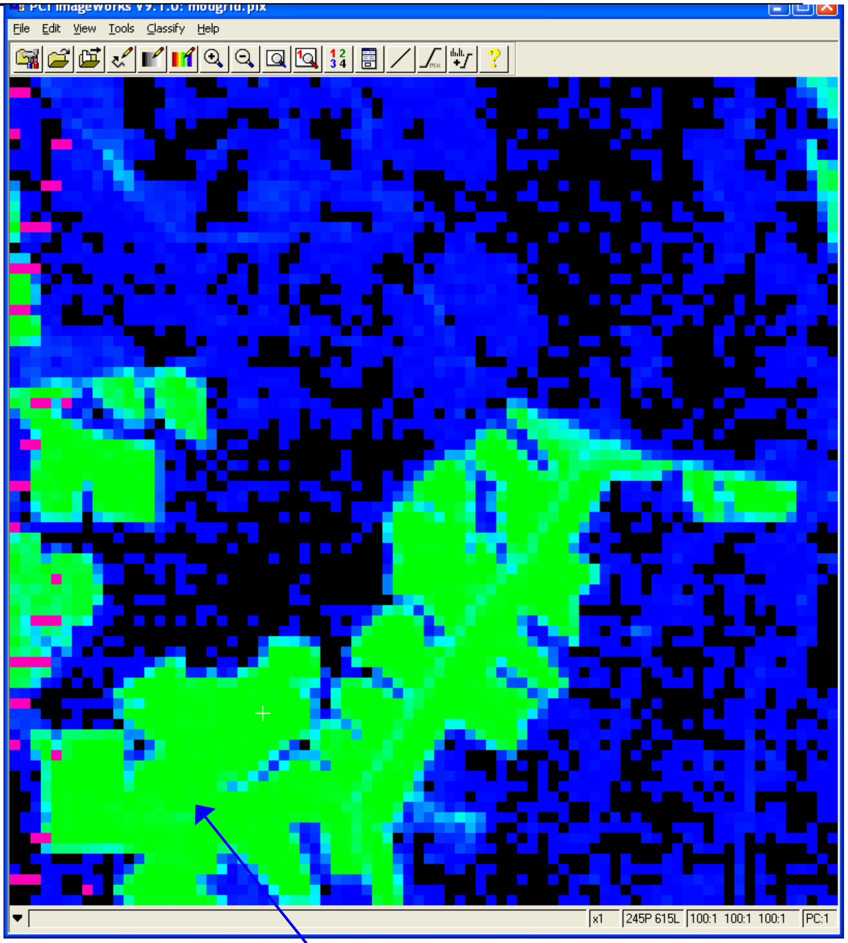
site 1 site 2

Scatter plot of field canopy density versus MODIS VCF at both sites in Wisconsin.

Comparison of Deforestation detected by MODIS and LandSat (250m grid)



MODIS deforestation: 250m grid



LandSat deforestation

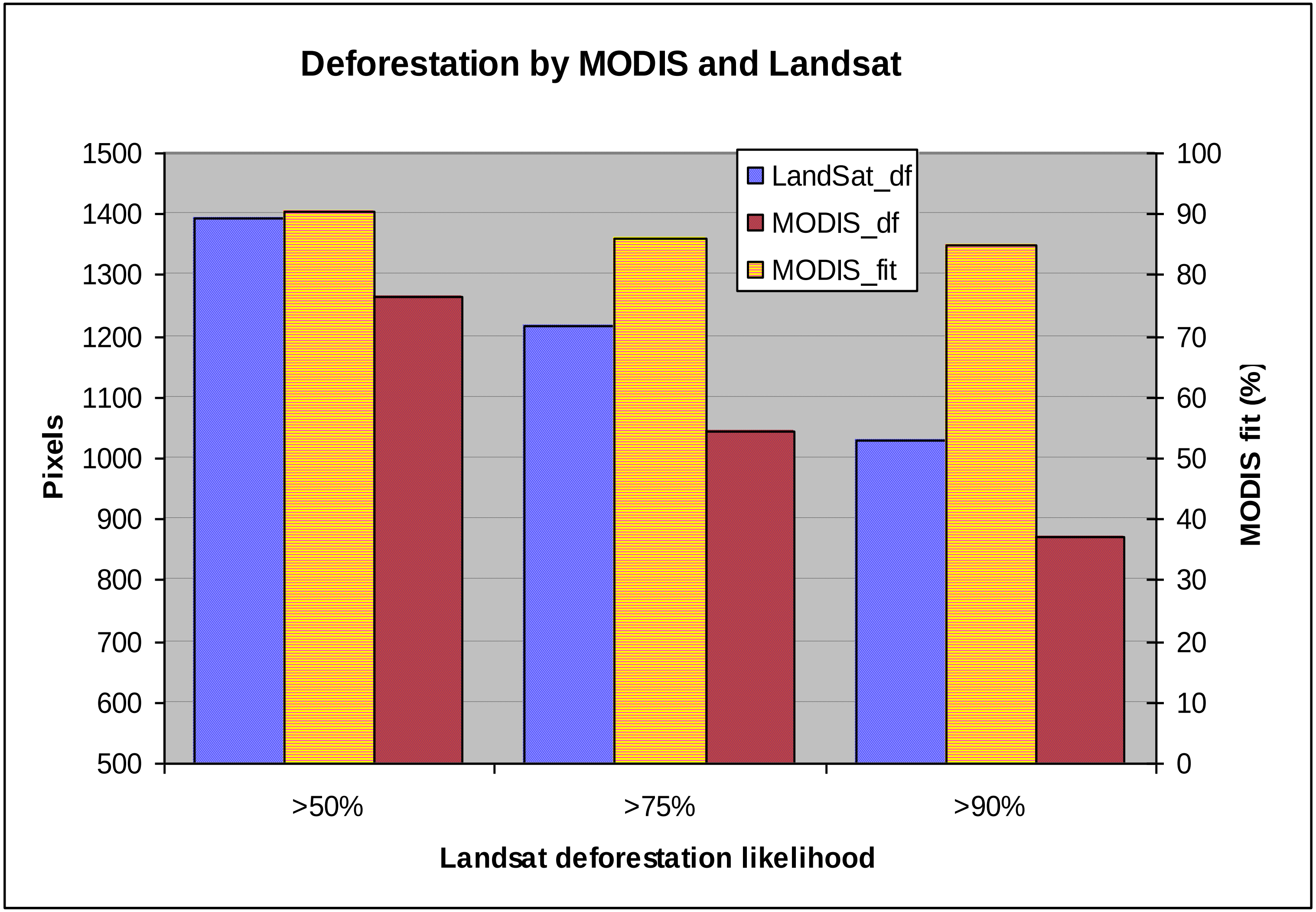
Comparison of Deforestation detected by MODIS and LandSat (250m grid)

- Classes (MODIS and Landsat thematic map):
  - Cloud mask:0 (no decision)
  - No change: 0 (likelihood)
  - Deforestation: 1-100% (likelihood)
- How to distinguish from no change (=0) and cloud mask (=0)?

$$p = \frac{\sum_{i=1}^{N} p_i}{N - N_c}$$

P is the likelihood of deforestation in the aggregated image; P_i is likelihood of deforestation for pixel i; N is the total ETM+ Pixels at one MODIS pixel (64); Nc the total pixels of cloud mask at one MODIS pixel.

Comparison of Deforestation detected by MODIS and LandSat (250m grid)



3.3 Evaluation of the classification

2. Base error matrix

Reference or ground truth classes

Classification

	Residential	Commercial	Wetland	Forest	Water	Row total
Residential	**70**	5	0	13	0	88
Commercial	3	**55**	0	0	0	58
Wetland	0	0	**99**	0	0	99
Forest	0	0	4	**37**	0	41
Water	0	0	0	0	**121**	121
Column total	73	60	103	50	121	407

3.3 Evaluation of the classification

- **Base error matrix**
  - **Agreement/accuracy:** the probability (%) that the classifier has labeled an image pixel into the ground truth Class. It is the probability of a reference pixel being correctly classified.
  - **Overall accuracy:** the total classification accuracy
  - **Commission error:** represent pixels that belong to another class but are labeled as belonging to the class.
  - **Omission error:** represent pixels that belong to the truth class but fail to be classified into the proper class.
  - **Kappa coefficient(K_{hat}):** a discrete multivariate technique of use in accuracy assessment. K_{hat}>0.80 represent strong agreement and good accuracy. 0.40-0.80 is middle, <0.40 is poor.

3.3 Evaluation of the classification

- Base error matrix

Classification

	Residential	Commercial	Wetland	Forest	Water	Row total
Residential	70	5	0	13	0	88
Commercial	3	55	0	0	0	58
Wetland	0	0	99	0	0	99
Forest	0	0	4	37	0	41
Water	0	0	0	0	121	121
Column total	73	60	103	50	121	407

Agreement/accuracy: 70/73=96% 55/60 99/103 37/50 121/121

Omission error: 3/73=4% 5/60 ……

Commission error: (5+13)/88=20% …….

Overall accuracy: (70+55+99+37+121)/(73+60+103+50+121)=94%

3.3 Evaluation of the classification

Computation of K_{hat} Coefficient of Agreement

$$\hat{K} = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+i})}$$

Example: they took 407 samples (pixels) based on the stratified random sampling after classification. First made 5 files (each contain one class), using a random number generator to get points.

where $N = 407$

$$\sum_{i=1}^{k} x_{ii} = (70 + 55 + 99 + 37 + 121) = 382$$

$$\sum_{i=1}^{k}(x_{i+} \times x_{+i}) = (88 \times 73) + (58 \times 60) + (99 \times 103) + (41 \times 50) + (121 \times 121) = 36{,}792$$

$$\text{therefore } \hat{K} = \frac{407(382) - 36792}{407^2 - 36792} = \frac{155474 - 36792}{165649 - 36792} = \frac{118682}{128857} = 92.1\%$$

3.3 Evaluation of the classification

Conditional K_{hat} coefficient of agreement

$$\hat{K}_i = \frac{N \bullet x_{ii} - x_{i+} \times x_{+i}}{N \bullet x_{i+} - x_{i+} \times x_{+i}}$$

$$\hat{K}_{resid} = \frac{407 \bullet 70 - 88 \bullet 73}{407 \bullet 88 - 88 \bullet 73} = 75\%$$

3.3 Evaluation of the classification

Table 1. Error matrix between MOD10A1 and ground measurements in the 2003-04 hydorlogic year at 20 stations in Northern Xinjiang, China

Ground observations	Total	MOD10A1			Accuracy after cloud removed*
	6496	Snow	Land	Cloud	
Land (Snow depth = 0)	4247	13	2764	1441	99%
	65%	1%	65%	34%	
Fractional Snow (Snow Depth =1-3 cm)	252	31	31	190	50%
	4%	12%	12%	75%	
Snow (Snow depth ≥ 4 cm)	1997	868	21	1108	98%
	31%	43%	1%	55%	

Image classification accuracy $\underbrace{\text{in all sky}}$ (Snow, Land, Cloud columns) $\underbrace{\text{in clear sky}}$ (Accuracy after cloud removed column)

Wang et al. 2007

Overall accuracy
= (2764+12+868)/6490
=56%

3.4 another sets of evaluation

Table 2. Confusion Matrix for Remote Sensing Image v.s. *in situ* Observations

	Image: Snow	Image: No Snow	Image: Cloud
Ground: Snow	a	b	e
Ground: No Snow	c	d	f

(a,b,c,d,e and f represent number of station-pixels in each particular classification category.)

$$IU = \frac{b}{a+b+c+d}$$ image underestimation

$$IO = \frac{c}{a+b+c+d}$$ image overestimation

$$S_c = \frac{a}{a+b}$$ Snow accuracy in clear sky

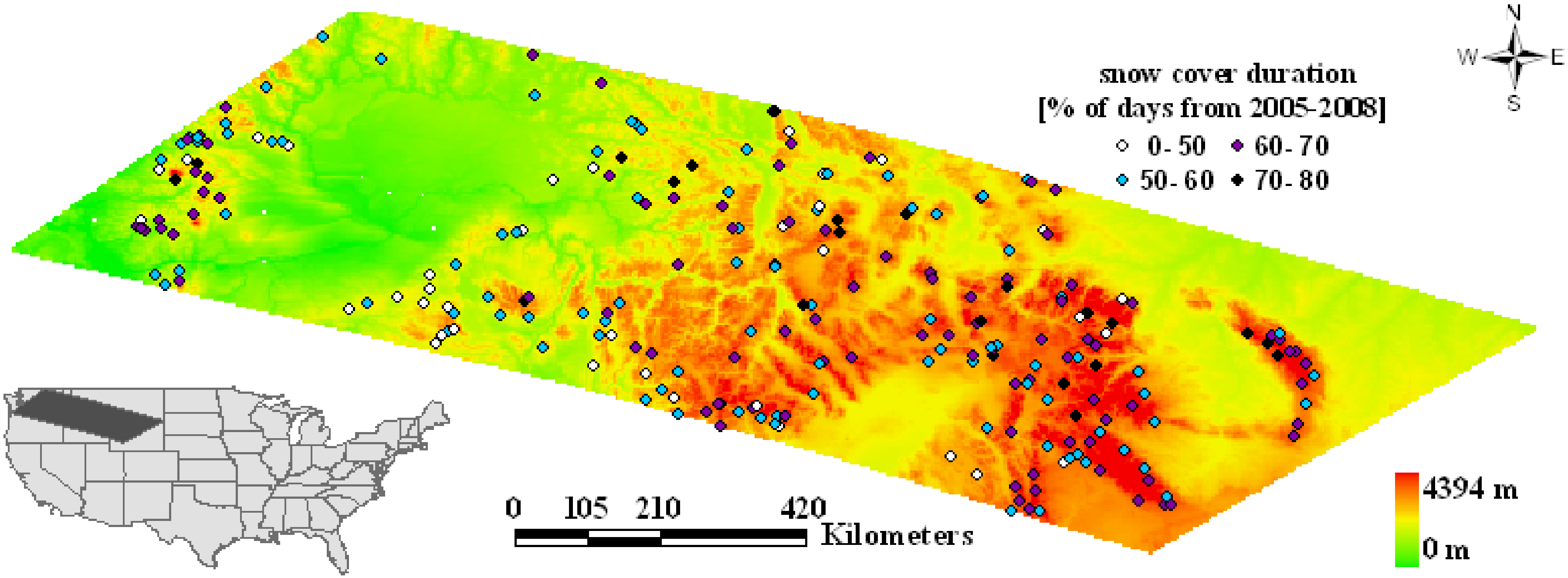
$$S_a = \frac{a}{a+b+e}$$ Snow accuracy in all sky

$$L_c = \frac{b}{a+b}$$ Land accuracy in clear sky

$$L_a = \frac{b}{a+b+e}$$ Land accuracy in all sky

Another sets

Regular sets



Gao et al. 2010

Table 4. Monthly and annual median image Under- (IU) and Over-estimation (IO) Errors (%) of the multi-temporal (fixed days and flexible multiday combined) snow cover products, evaluated for clear-sky condition, for the period of 2006-2008 hydrological years

Month	Fixed days combined								Multiday combined			
	2 DAYS		4 DAYS		6 DAYS		8 DAYS		MODISMC4		MODISMC8	
	IU	IO	IU	IO	IU	IO	IU	IO	IU	IO	IU	IO
Sep	2.3	4.9	2.2	9.3	2.2	13.4	2.0	17.8	2.5	7.3	2.6	8.4
Oct	9.9	15.9	8.1	21.5	7.0	28.1	6.5	31.5	9.5	18.4	10.8	20.7
Nov	12.4	5.4	8.9	7.3	7.1	7.8	4.7	9.7	10.8	6.4	6.7	8.3
Dec	4.5	0.3	3.5	0.3	3.3	0.3	2.1	0.3	3.3	0.3	2.9	0.3
Jan	2.8	0.0	2.5	0.1	2.0	0.1	1.3	0.1	2.4	0.1	1.4	0.1
Feb	3.0	0.2	1.9	0.3	1.8	0.3	1.3	0.3	2.2	0.3	1.6	0.3
Mar	3.9	0.5	3.1	0.4	2.4	0.5	1.9	0.5	3.1	0.4	1.8	0.5
Apr	9.7	1.2	6.8	1.9	5.8	2.4	4.3	2.6	9.1	1.6	8.3	1.6
May	24.7	3.9	18.9	5.6	17.5	6.9	15	8.5	22.8	4.4	22.1	5.5
Jun	8.4	3.3	8.0	5.6	7.3	7.8	6.8	10.1	8.4	4.3	8.3	5.4
Jul	0.2	1.2	0.2	2.2	0.2	3.1	0.1	4.3	0.2	1.0	0.2	1.0
Aug	0.0	1.5	0.0	2.9	0.0	4.4	0.0	5.4	0.0	1.7	0.0	1.7
Annual	**7.1**	**3.5**	**5.9**	**5.3**	**4.9**	**6.7**	**4.3**	**7.8**	**6.6**	**4.1**	**6.1**	**4.5**

Even in clear sky, image tends to underestimate snow cover, especially in transient months;
Annually, fixed daily, 2-day, 4-day, flexible multi-day images tend to underestimate snow cover;
Flexible multi-day images have similar behave as fixed 4-day.