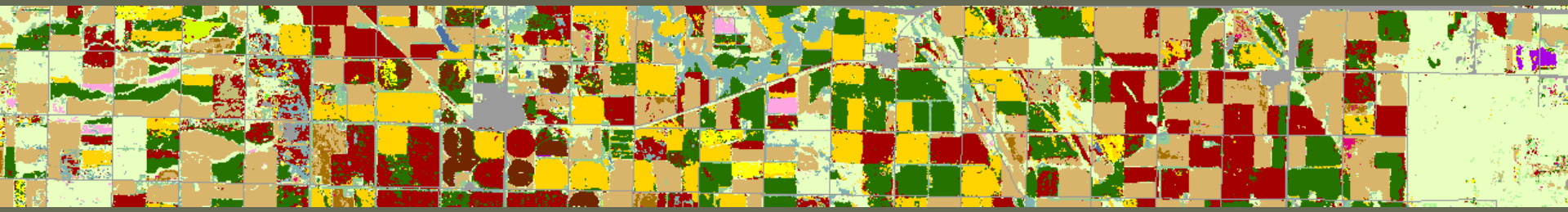
# Feasibility of spatial resolution and herbaceous category improvements to the Cropland Data Layer

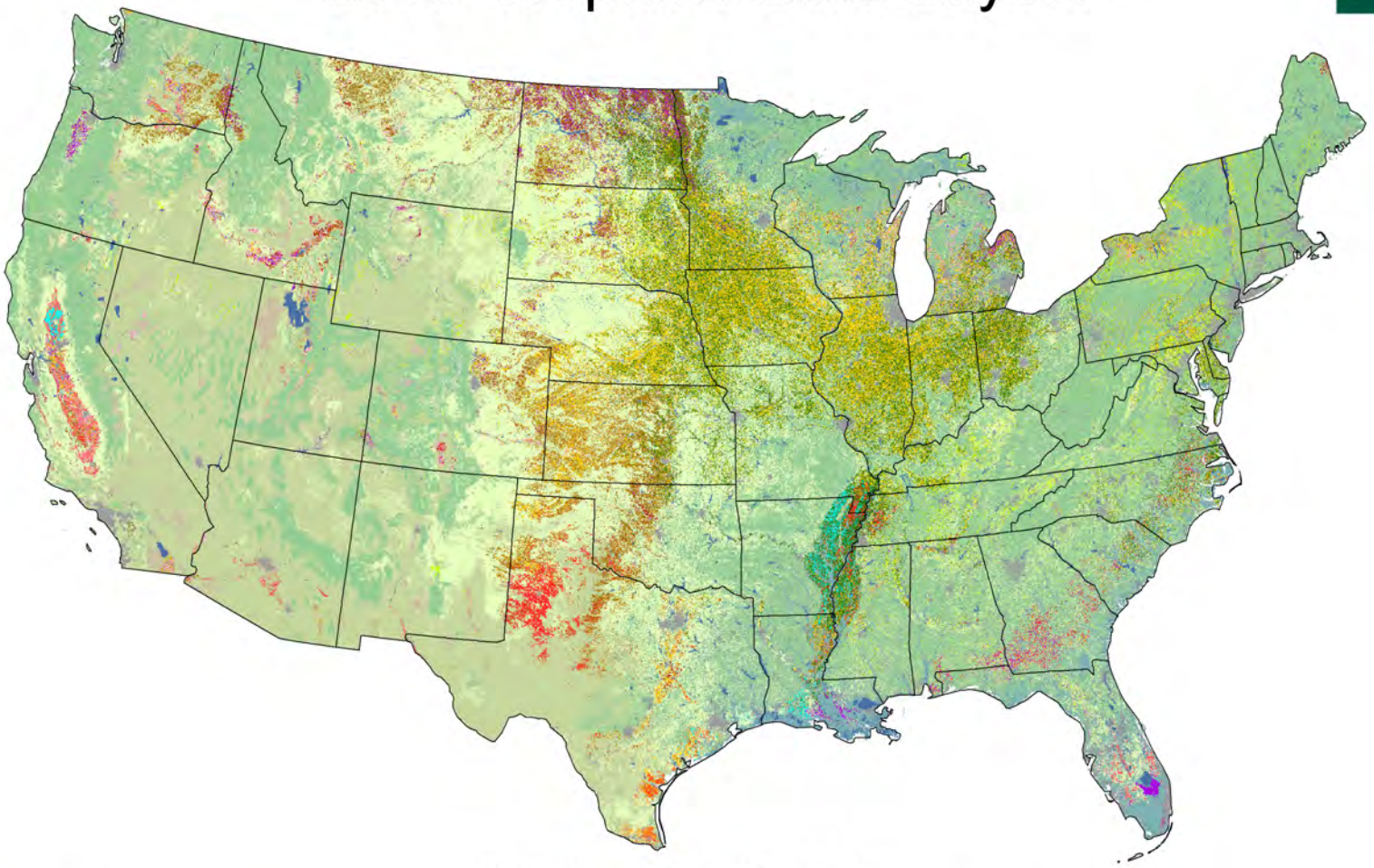


David M. Johnson
Geographer

National Agricultural Statistics Service
United States Department of Agriculture

May 3, 2011
ASPRS Annual Conference
Milwaukee, Wisconsin

# 2010 Cropland Data Layers



**Major Land Cover Categories**

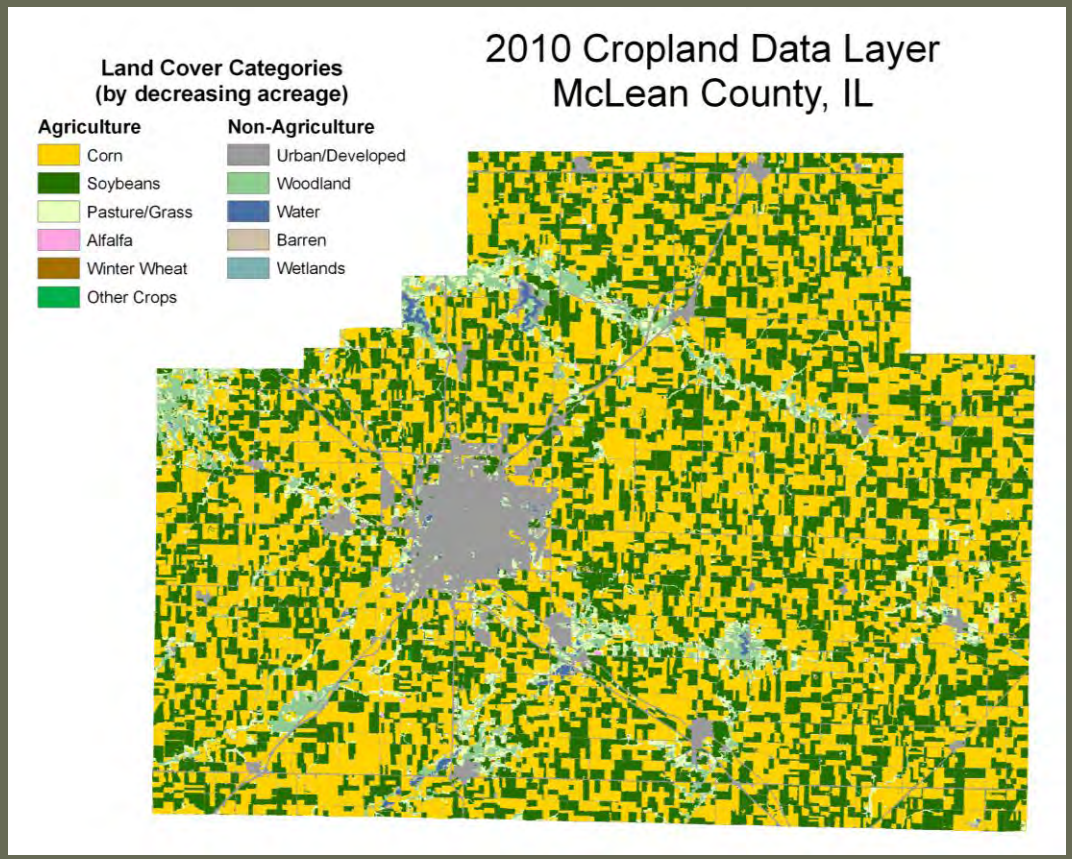
**Agriculture**

- Pasture/Grass
- Corn
- Soybeans
- All Wheat
- Other Hay
- Fallow/Idle Cropland
- Alfalfa
- Cotton
- Other Crops
- Vegetables/Fruits/Nuts
- Sorghum
- Other Small Grains
- Rice

**Non-Agriculture**

- Woodland
- Shrubland
- Urban/Developed
- Wetlands
- Water
- Barren
- Perennial Ice/Snow

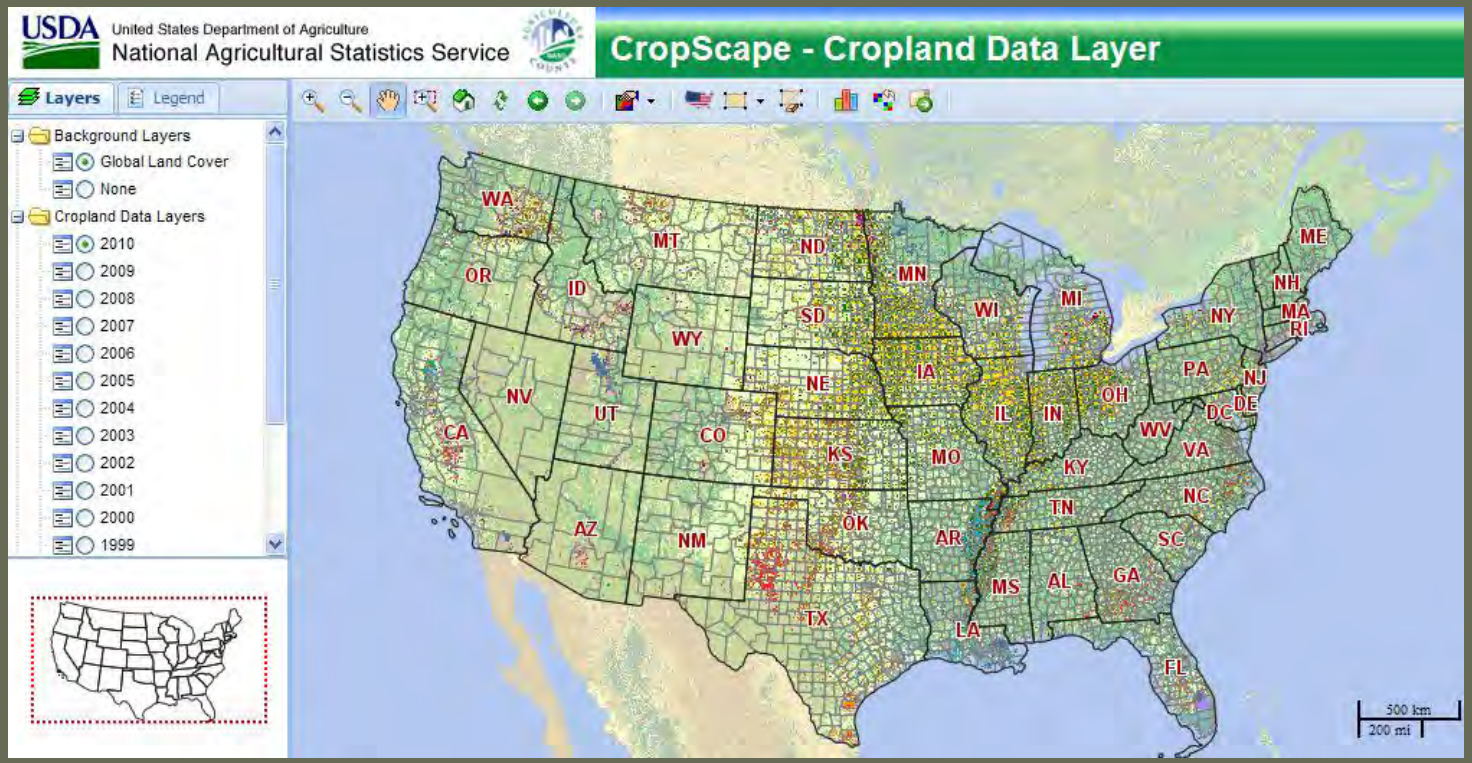
# CDL generalities




- Annual land cover classification targeted to identifying *circa* summer cultivated crops
- Encompasses (since 2009) all of conterminous USA
- 56m or 30m resolution
  - Depending on year
- Derived primarily from
  - Landsat-5 TM
  - Resourcesat-1 AWiFS
- Built with a supervised classification tree methodology
  - Implemented with See 5.0 via ERDAS Imagine
- Utilizes ground/training data from USDA Farm Service Agency and ancillary data from National Land Cover Database
- Highly robust for dominant crop types
  - corn, soybeans, wheat, rice, cotton, etc.
- Used internally by NASS to estimate state and county–level acreage

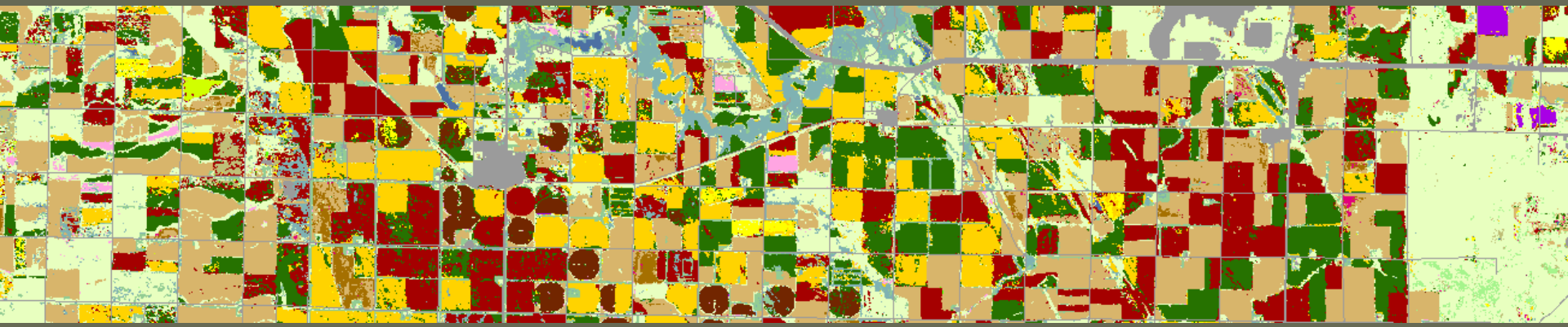
# CDL access

- Traditionally through DVD, FTP, or the USDA Geospatial Data Gateway
- Now easiest is the “CropScape” portal
  - nassgeodata.gmu.edu/CropScape
  - or just Google “CropScape”



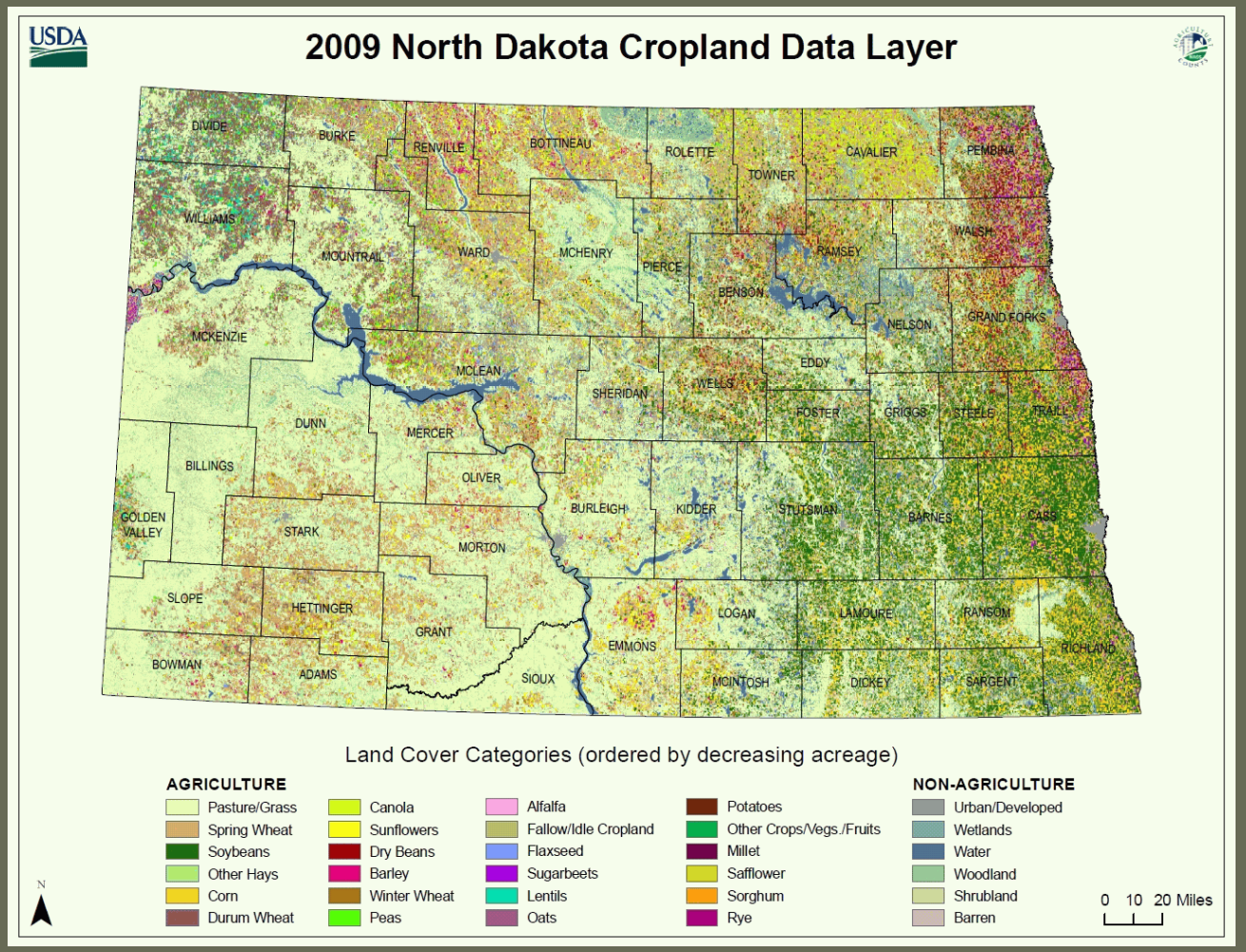
# Two of the bigger user questions about CDL

- Can you improve the spatial resolution?
  - 56m is too coarse
  - 56m is not“standard”
- Can you better define and accurately map herbaceous categories?
  - Pasture, hay, grasslands, Cropland Reserve Program (CRP), native grasses, rangeland, etc…



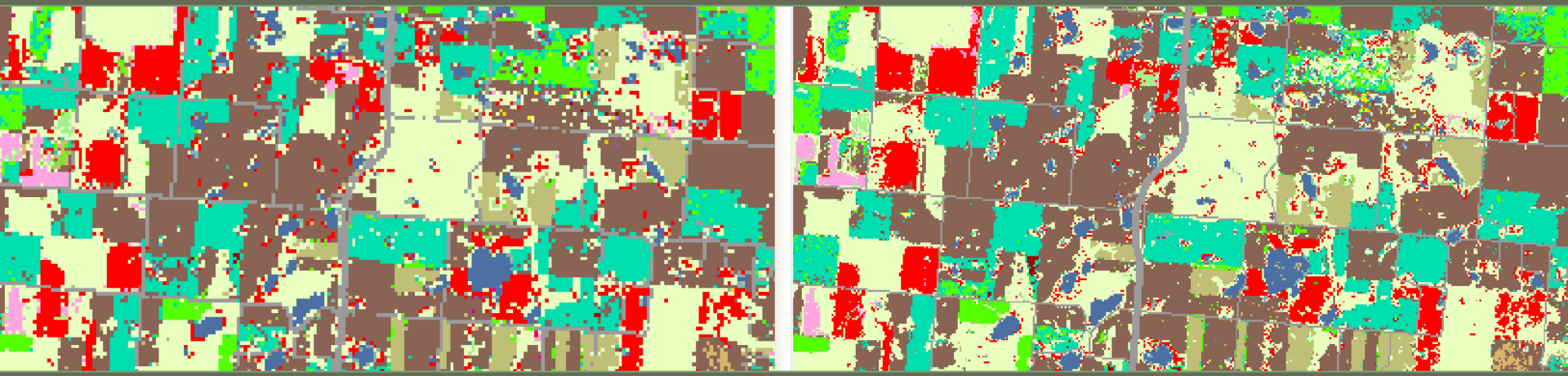
# North Dakota as a study area

- ND is nice to analyze because
  - Large diversity of major commodity crops
  - Large diversity of grassland usages



# Question #1: Finer spatial resolution?

- All CDLs, to date, are either 30m or 56m resolution
  - Based on the primary data source
    - If Landsat TM, or ETM, then 30m
    - If Resourcesat AWiFS then 56m
  - Earlier CDLs (early 2000s) were 30m
  - More recent CDLs (late 2000s) were 56m

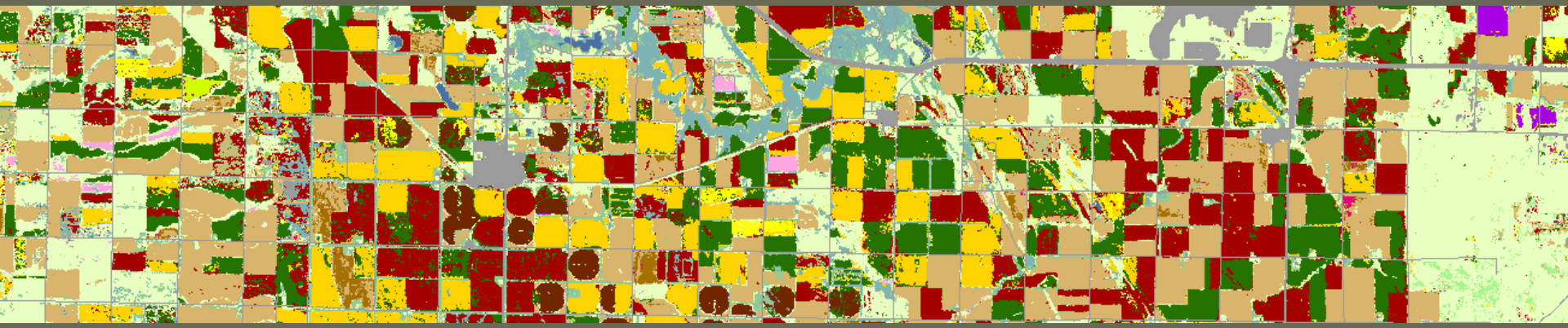


56m NW North Dakota 30m

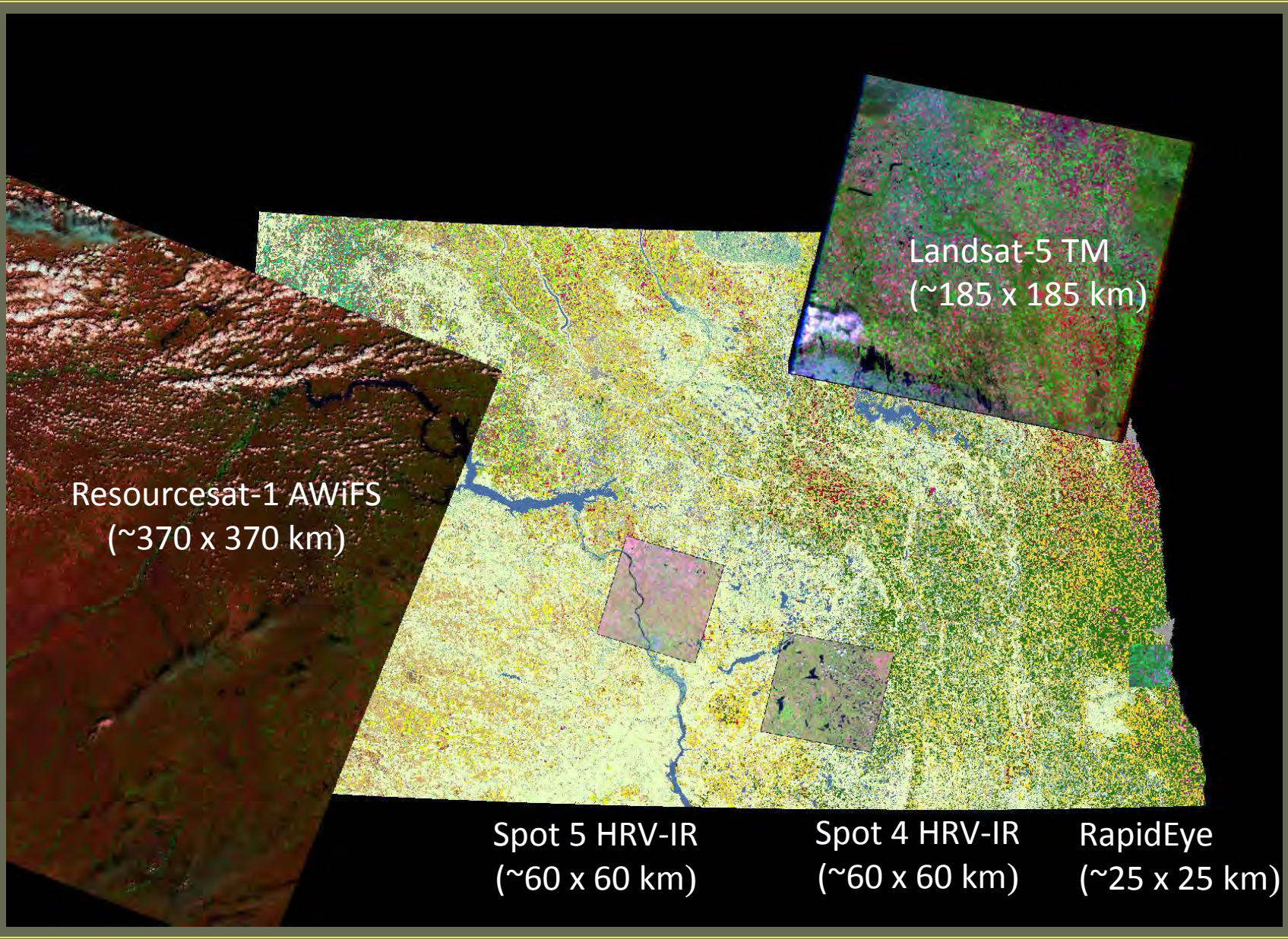
- **2010 originally planned and processed at 56m but ultimately reprocessed and released at 30m!**
- Accuracy (pixel level) of cropped areas: 56m = 78.1%, 30m = 80.5%
- In terms of NASS derived area estimates 56m v. 30m perform about the same

# Finer than 30m resolution?

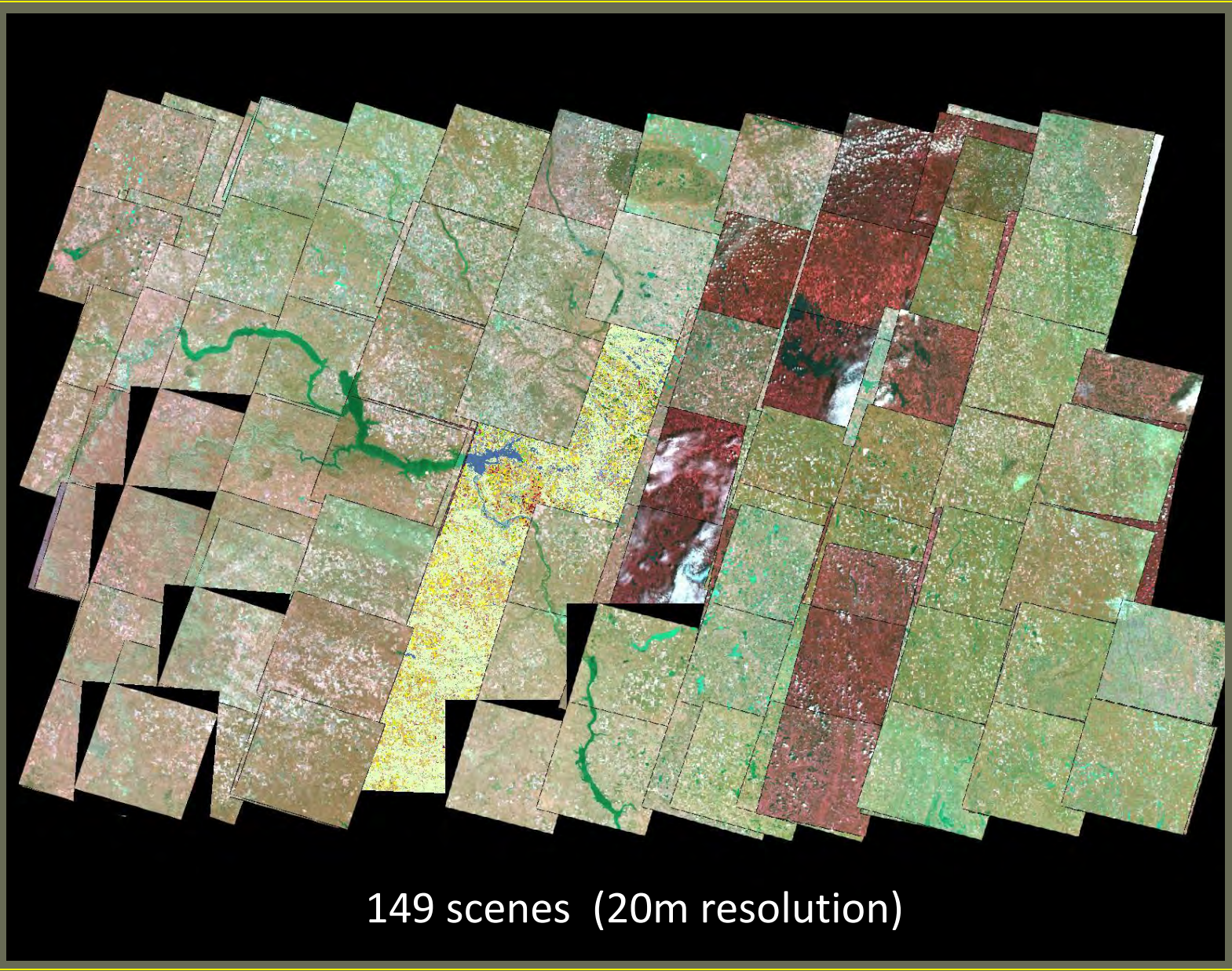
- What about 20m, or 10m, or 5m resolution?
  - Number of sensors currently producing finer resolution data with multispectral capability
    - SPOT, LISS-III, DMCii, DEIMOS, RapidEye etc…
- For 2010 large US Federal data buy of SPOT imagery
  - Targeted over growing areas but covering most of US.
  - Freely distributable to Federal agencies via USGS EarthExplorer
  - Imagery not ortho-rectified but reasonably well geo-registered
- So SPOT best candidate for higher resolution land cover mapping…



# Areal comparison



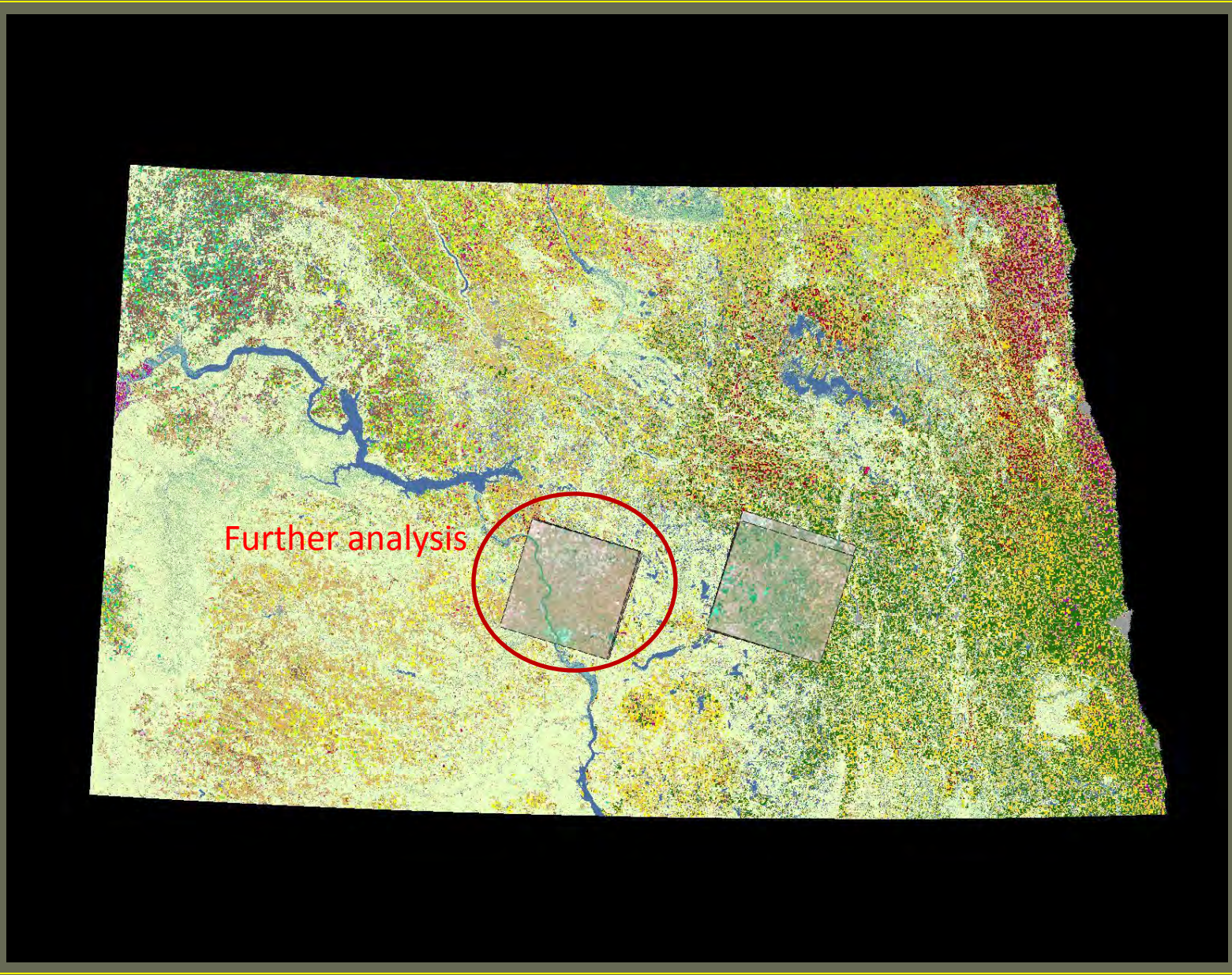
# SPOT 4 North Dakota 2010 summer collects



# SPOT 5 North Dakota 2010 summer collects

# Areas with at least two SPOT 5 scenes

# Areas with at least three SPOT 5 scenes

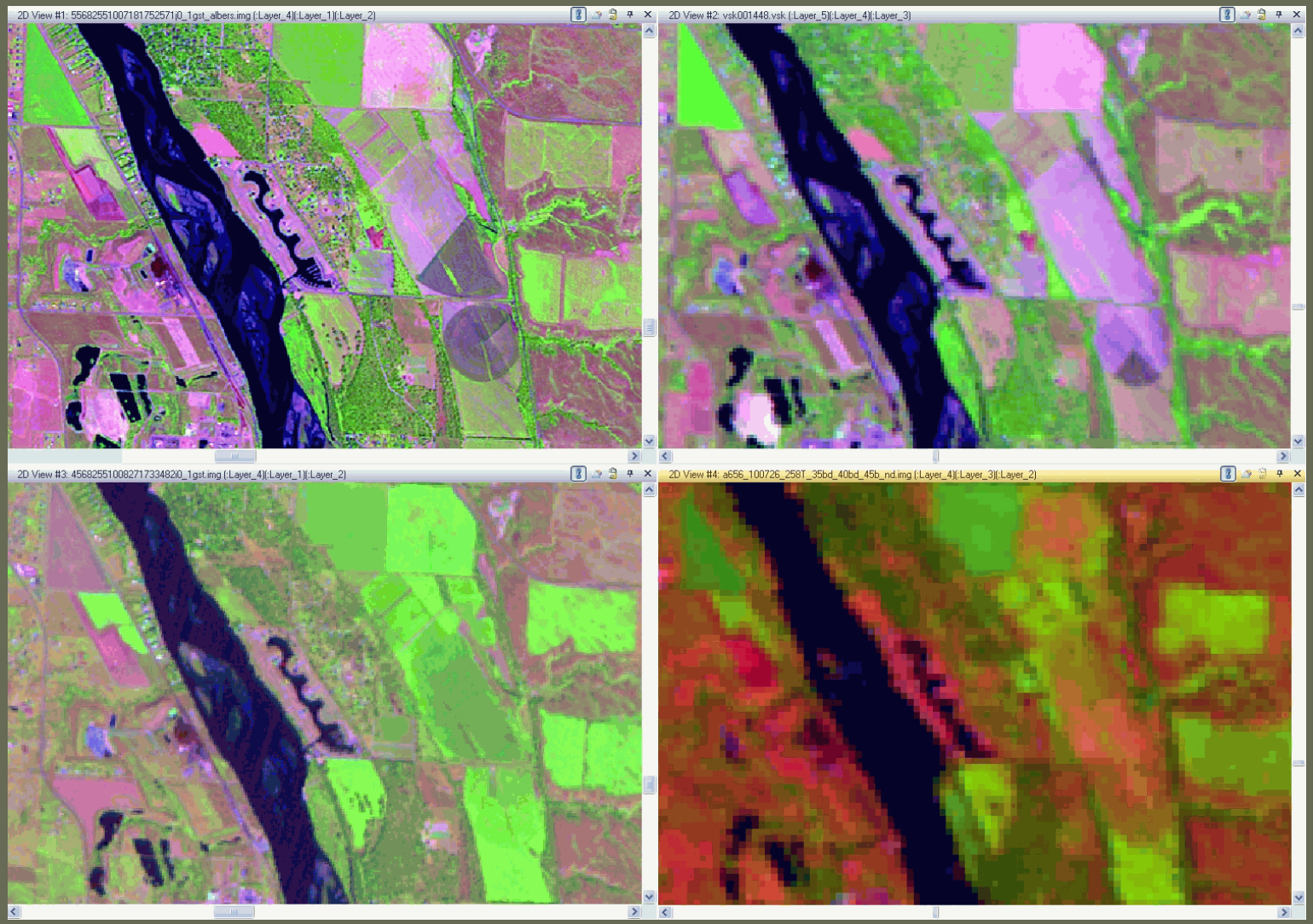


# Visual comparison (“false natural color”)

SPOT 5

Landsat TM

July 18



Aug. 27

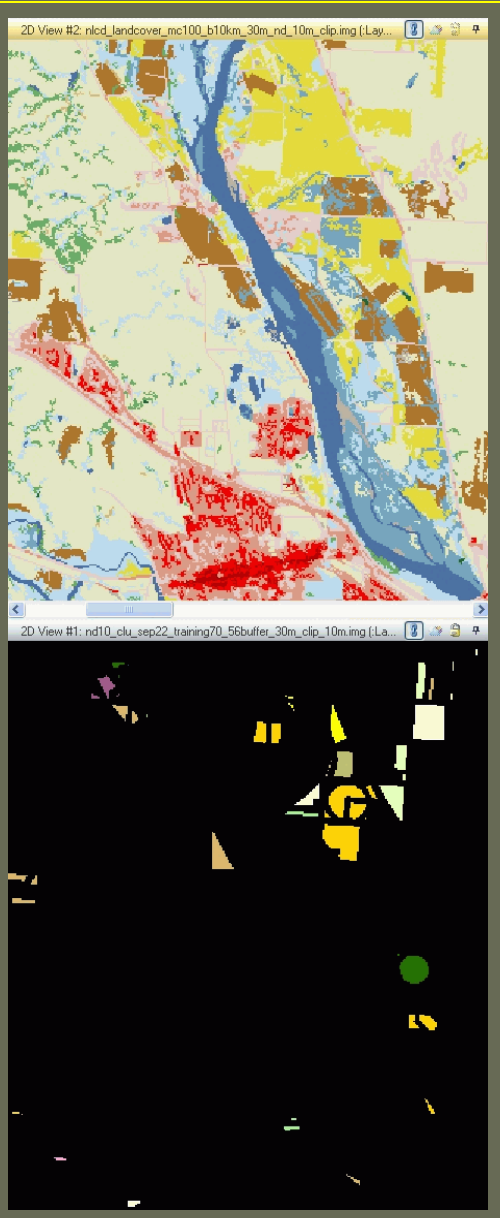
July 26

SPOT 4

Resourcesat-1 AWiFS

# Notes about ground data

- Agricultural categories
  - Use Farm Service Agency farmer reported "578" program crop tied to Common Land Unit (CLU) polygon data
  - Early in the season this information is thin
- Non-agricultural categories
  - Draw samples from the NLCD to act as proxy ground truth
- Ultimately draw hundreds of thousands of sample from both across a state and use in supervised decision tree type classification
  - Perhaps 1 – 10 % agricultural ground truth coverage early in the season
  - Grows to 20, 30, 40%, or more, as season evolves
- Made even more robust with knowledge that classifier is tolerant of outliers/errors




NLCD

FSA

# Three SPOT scenes and output classification

July 18

Aug 17

Sep 29

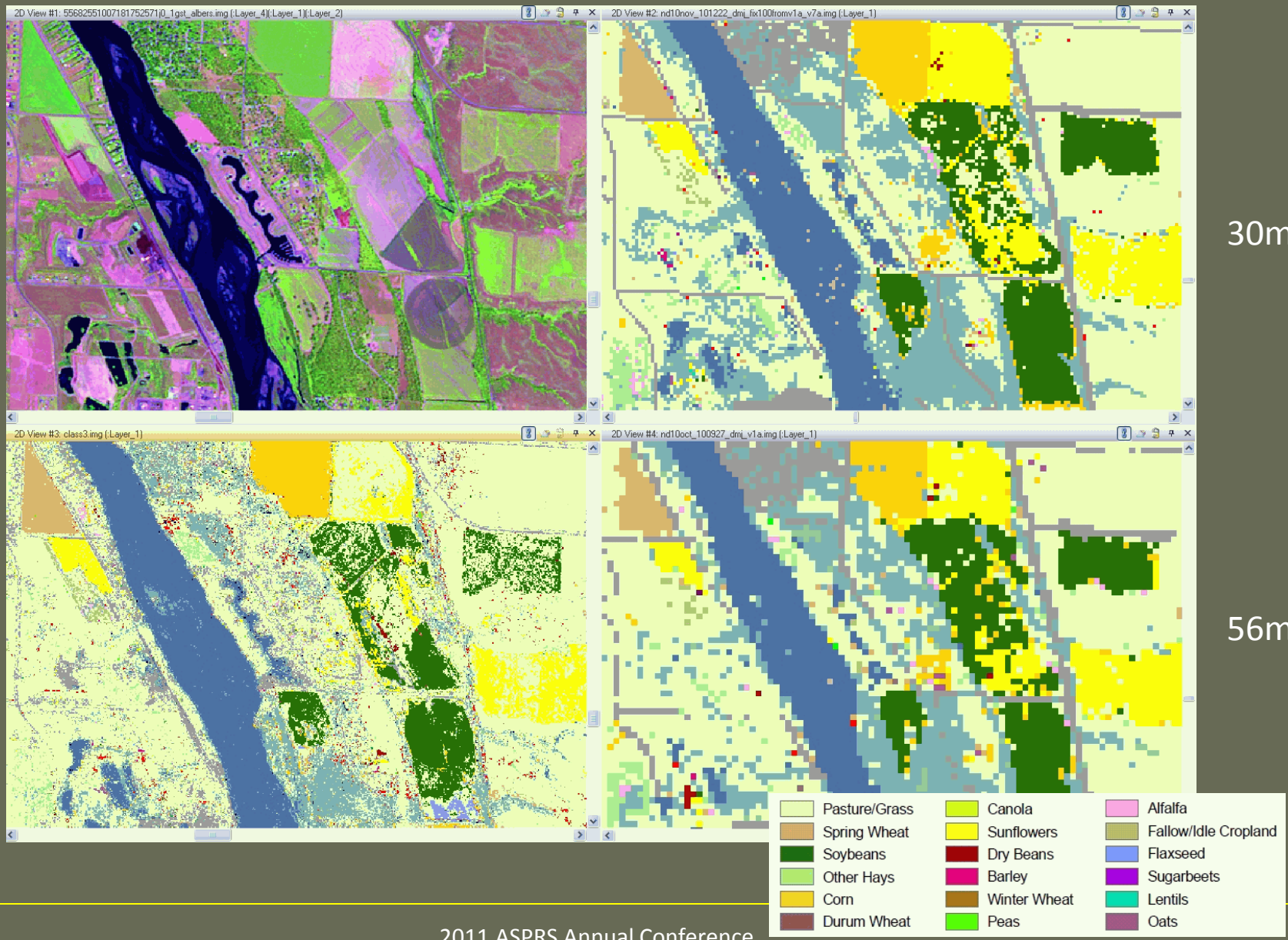
# Classifications compared

Raw 10m

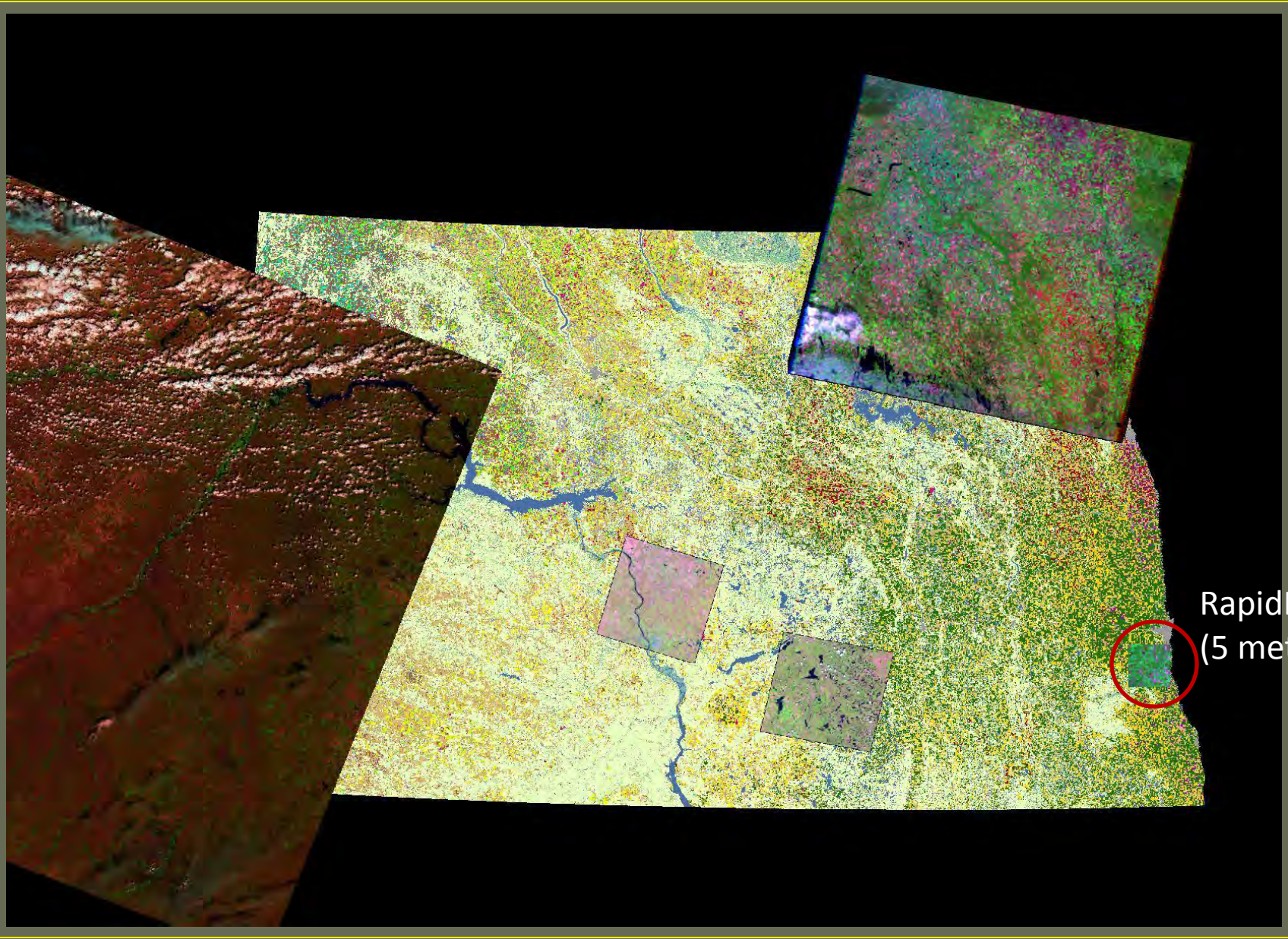
30m

10m

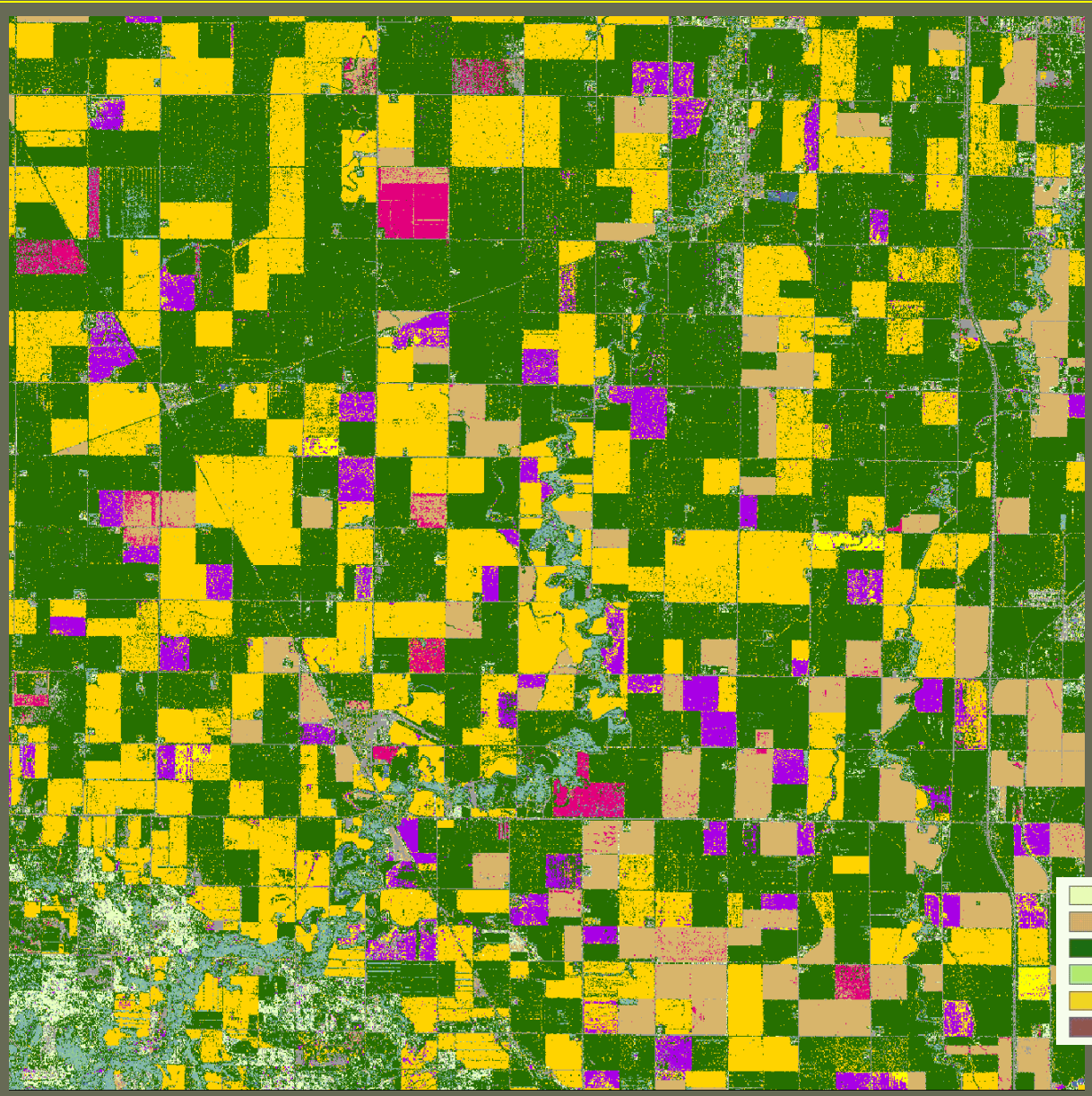
56m



# Second area - RapidEye analysis

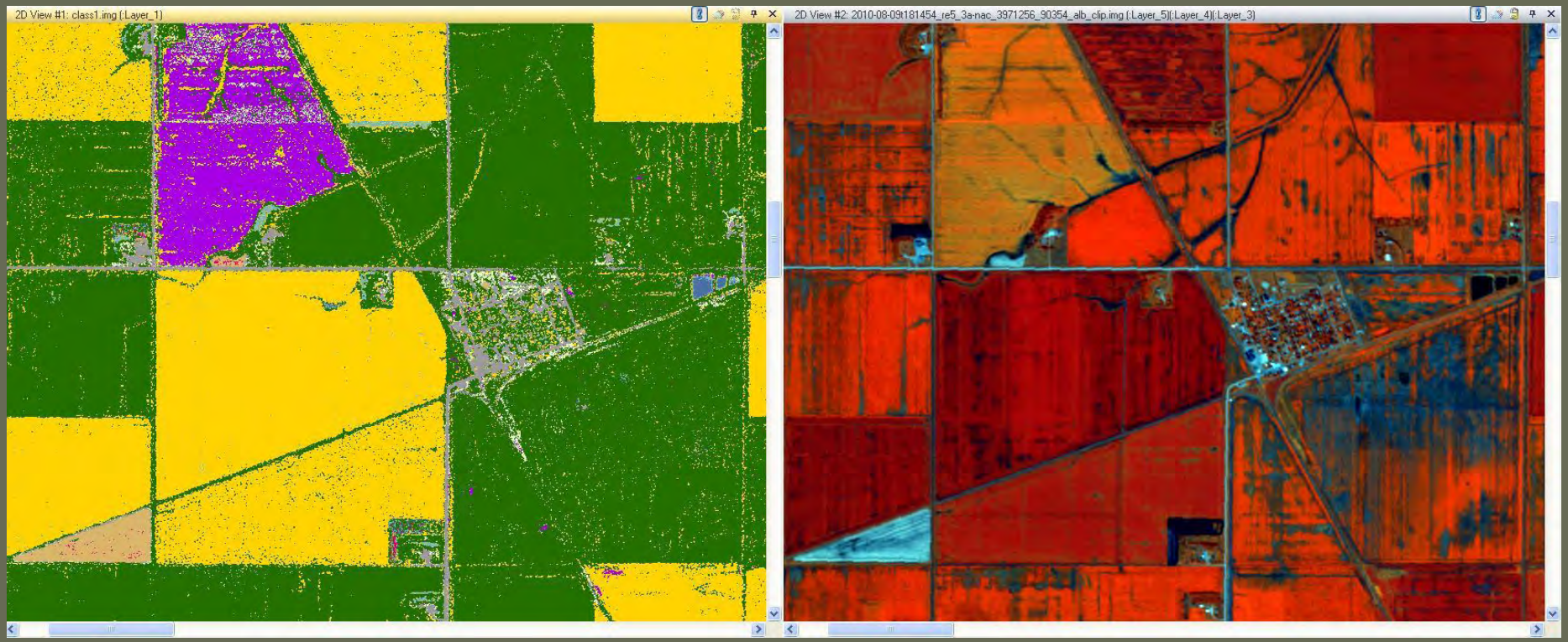


# RapidEye classification




- Single scene
  - August 9th
  - not "multi-temporal"
  - Reasonable nonetheless

# RapidEye 5m classification versus raw

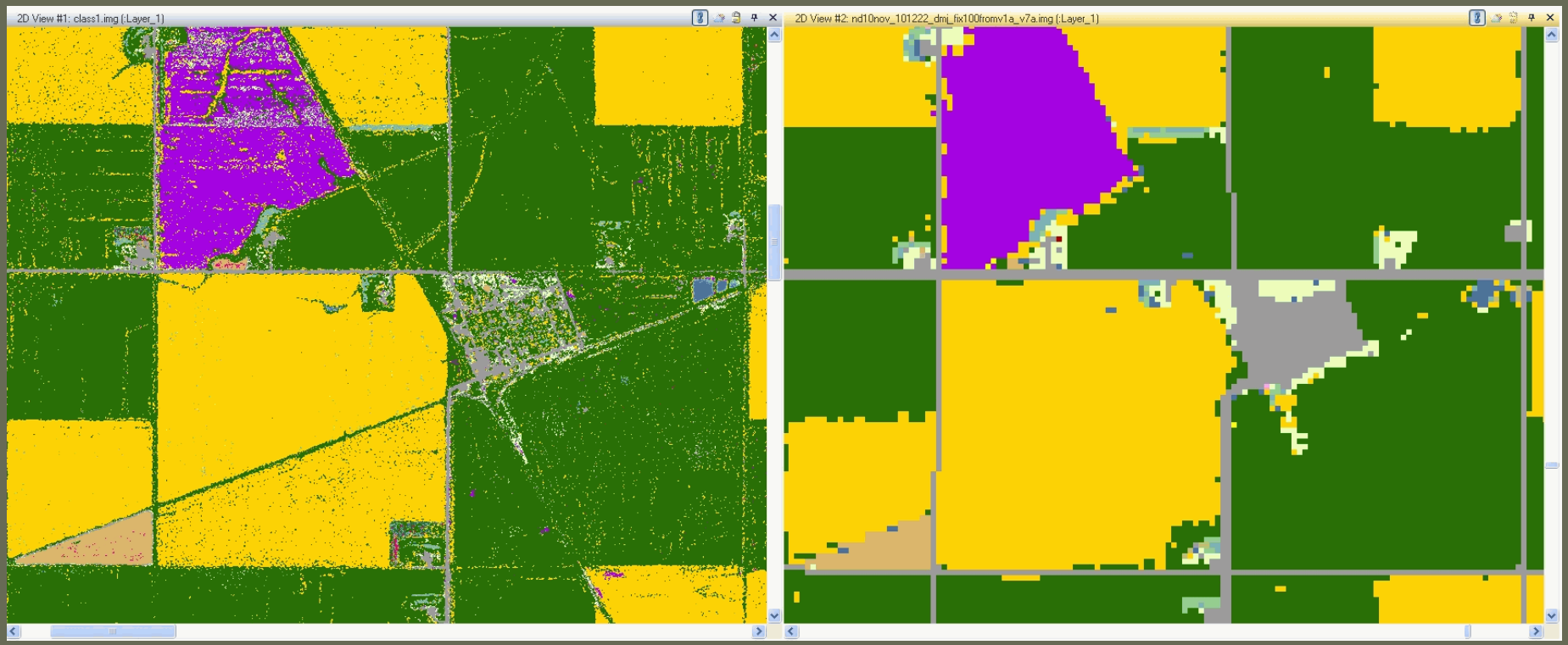



| | | |
|---|---|---|
| Pasture/Grass | Canola | Alfalfa |
| Spring Wheat | Sunflowers | Fallow/Idle Cropland |
| Soybeans | Dry Beans | Flaxseed |
| Other Hays | Barley | Sugarbeets |
| Corn | Winter Wheat | Lentils |
| Durum Wheat | Peas | Oats |

R = NIR band
G= “red edge” band
B= red band

# RapidEye classification versus 30m CDL

- Single date of 5m data seems to perform well
- Obvious improvements with detail
- “Red-edge” band may indeed be yielding additional information

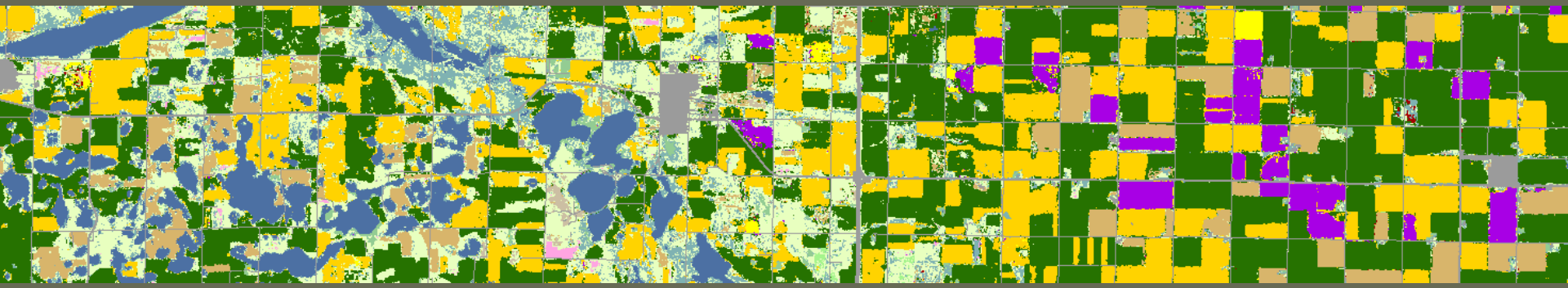



5m

30m

# Spatial resolution improvements conclusions

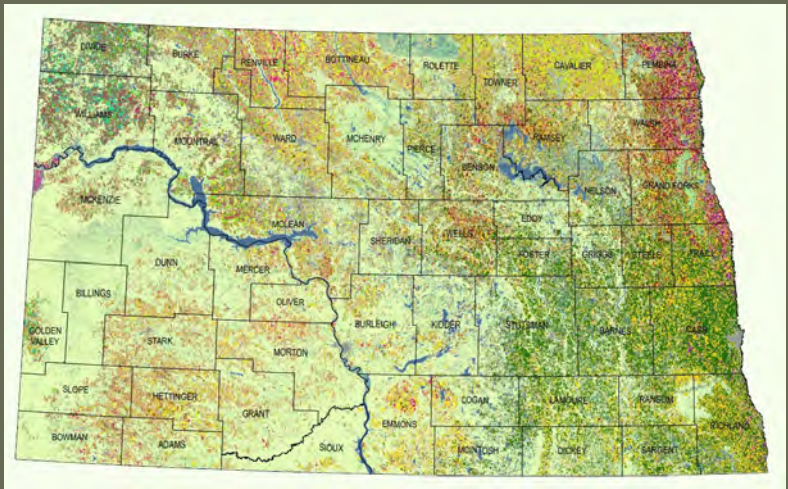
- National 30m, the new NASS standard
- May be possible to go finer but unlikely in the near future
  - Incomplete multi-temporal coverage
  - Exponential data handling needs
    - downloading, storage, and processing
  - Little to gain in area statistics improvement
  - Lack of ortho-registation with US SPOT data buy
    - Albeit SPOT 5 is probably close enough, especially if used in conjunction with coarser (e.g. Landsat) data.
  - Hindrance with no “bulk download” capability at USGS like for Landsat data



# Question 2: Herbaceous classification

- NASS CDL nomenclature would generally call these
  1. Hay (excluding alfalfa)
  2. Pasture
  3. Non-agriculture, grassland, waste, idle, …
     - These have been poorly defined and/or inconsistent through the years
- FSA provides information about 5 grass usage types
  1. “Forage”
  2. “Grazing”
  3. “Left Standing”
  4. “Seed”
  5. “Sod”
- Trying to force into cover type versus usage one could say
  1. Cut grass
  2. Chewed grass
  3. Undisturbed grass
  4. Seed grass
  5. Sod grass
- Also there is FSA information about CRP and cover type/usage
  – Not all CRP is grassy!

# North Dakota 2010 accuracies




Crops average = 80.5%

| CLASS | | Correct | Producer | User | Bias |
|---|---|---|---|---|---|
| Corn | 1 | 668556 | 91.00% | 95.24% | -4.5% |
| Sorghum | 4 | 3197 | 45.02% | 82.76% | -45.6% |
| Soybeans | 5 | 1365715 | 95.43% | 95.89% | -0.5% |
| Sunflowers | 6 | 313862 | 89.05% | 93.17% | -4.4% |
| Barley | 21 | 107277 | 52.92% | 84.31% | -37.2% |
| Durum wheat | 22 | 829926 | 77.91% | 82.09% | -5.1% |
| Spring wheat | 23 | 2691720 | 90.99% | 88.53% | 2.8% |
| Winter wheat | 24 | 140939 | 84.35% | 91.78% | -8.1% |
| Oats | 28 | 36453 | 37.80% | 78.48% | -51.8% |
| Canola | 31 | 462926 | 94.10% | 97.28% | -3.3% |
| Flaxseed | 32 | 89174 | 67.56% | 87.57% | -22.9% |
| Alfalfa | 36 | 163644 | 49.71% | 75.04% | -33.8% |
| Other hays | 37 | 329416 | 41.26% | 71.61% | -42.4% |
| Beets | 41 | 24110 | 91.74% | 96.02% | -4.5% |
| Dry beans | 42 | 142142 | 84.76% | 88.18% | -3.9% |
| Potatoes | 43 | 25503 | 75.52% | 93.13% | -18.9% |
| Lentils | 52 | 129723 | 88.34% | 90.67% | -2.6% |
| Peas | 53 | 168639 | 84.74% | 89.25% | -5.1% |
| Idle / Fallow | 61 | 80903 | 46.30% | 87.81% | -47.3% |

# Hay/forage accuracies for all states 2010

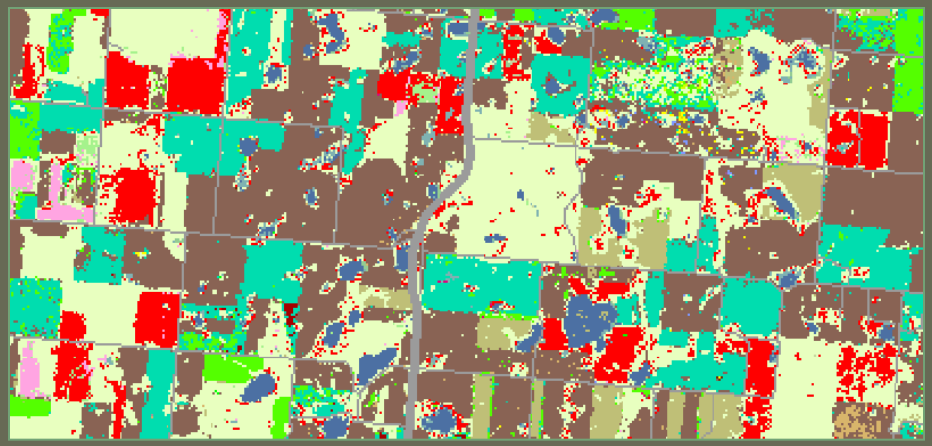
| STATE | PRODUCER'S ACCURACY | USER'S ACCURACY |
|---|---|---|
| AL | 43.0% | 29.5% |
| AZ | 47.0% | 80.5% |
| CA | 42.9% | 57.8% |
| CO | 39.9% | 80.8% |
| FL | 43.6% | 42.7% |
| GA | 33.7% | 51.3% |
| IA | 11.2% | 42.9% |
| ID | 38.3% | 67.5% |
| IL | 7.0% | 43.9% |
| IN | 10.4% | 43.7% |
| KY | 43.2% | 51.5% |
| MD,DE,NJ | 66.1% | 60.6% |
| ME,MA,CT,RI,VT,NH | 87.6% | 80.1% |
| MI | 59.7% | 69.8% |
| MN | 7.8% | 47.7% |
| MS | 60.0% | 20.5% |
| MT | 17.7% | 37.6% |
| NC | 39.0% | 50.5% |
| ND | 41.3% | 71.6% |
| NM | 23.7% | 88.7% |
| NV | 83.3% | 90.7% |
| NY | 75.5% | 65.8% |
| OH | 14.7% | 55.2% |
| OR | 47.0% | 59.8% |
| PA | 71.6% | 67.3% |
| SC | 36.8% | 55.7% |
| SD | 56.1% | 46.9% |
| TN | 56.7% | 41.1% |
| UT | 94.4% | 97.4% |
| VA,WV | 48.5% | 58.7% |
| WA | 47.5% | 85.6% |
| WI | 31.2% | 54.7% |
| WY | 88.5% | 95.4% |
| **unweighted average** | **45.9%** | **60.4%** |

# Pasture/Grazing accuracies all states 2010

| STATE | PRODUCER'S ACCURACY | USER'S ACCURACY |
|---|---|---|
| AL | 46.0% | 80.5% |
| CA | 54.6% | 92.2% |
| FL | 45.9% | 77.6% |
| GA | 71.8% | 70.9% |
| IA | 42.8% | 78.3% |
| ID | 62.4% | 93.8% |
| IL | 39.1% | 85.2% |
| IN | 23.0% | 72.5% |
| KY | 46.3% | 73.7% |
| MD,DE,NJ | 60.7% | 61.8% |
| ME,MA,CT,RI,VT,NH | 42.9% | 64.6% |
| MN | 65.3% | 86.2% |
| NC | 63.0% | 65.9% |
| NY | 29.7% | 50.3% |
| OH | 31.8% | 76.4% |
| OR | 30.2% | 95.2% |
| PA | 60.8% | 62.9% |
| SC | 62.5% | 73.6% |
| TN | 59.9% | 77.7% |
| VA,WV | 76.1% | 75.8% |
| WA | 53.0% | 94.7% |
| WI | 60.8% | 77.1% |
| **unweighted average** | **51.3%** | **76.7%** |

# Cropland Reserve Program accuracies

- Experimented adding CRP category for within season CDL products (56m, overall crop accuracy = 78.1%)
  - only used when know to be a grassland type cover
  - CRP results
    - 61.08% producer's accuracy
    - 74.37% user's accuracy
    - Not horrible

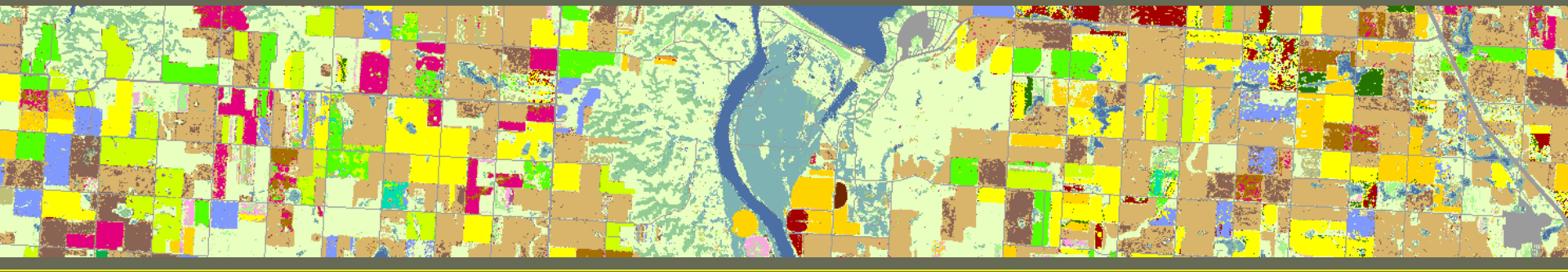


Red = CRP

| CLASS | | Correct | Producer | User | Bias |
|---|---|---|---|---|---|
| Corn | 1 | 141131 | 90.98% | 94.43% | -3.7% |
| Soybeans | 5 | 314465 | 95.33% | 95.38% | -0.1% |
| Sunflowers | 6 | 63206 | 87.43% | 89.18% | -2.0% |
| Barley | 21 | 15370 | 37.01% | 73.94% | -49.9% |
| Durum wheat | 22 | 147770 | 68.88% | 75.14% | -8.3% |
| Spring wheat | 23 | 571979 | 89.59% | 84.78% | 5.7% |
| Winter wheat | 24 | 25832 | 72.86% | 87.25% | -16.5% |
| Canola | 31 | 96998 | 94.26% | 96.57% | -2.4% |
| Flaxseed | 32 | 13210 | 51.82% | 81.49% | -36.4% |
| Mustard | 35 | 1330 | 51.25% | 86.76% | -40.9% |
| Alfalfa | 36 | 21761 | 38.88% | 62.68% | -38.0% |
| Other hays | 37 | 44939 | 33.81% | 57.88% | -41.6% |
| Beets | 41 | 5358 | 89.61% | 92.14% | -2.7% |
| Dry beans | 42 | 29806 | 77.91% | 84.18% | -7.4% |
| Lentils | 52 | 27772 | 86.93% | 88.03% | -1.2% |
| Peas | 53 | 33376 | 80.34% | 87.42% | -8.1% |
| Idle / Fallow | 61 | 11739 | 38.79% | 81.14% | -52.2% |
| CRP | 102 | 183859 | 61.08% | 74.37% | -17.9% |

# Grassland mapping thoughts and conclusions

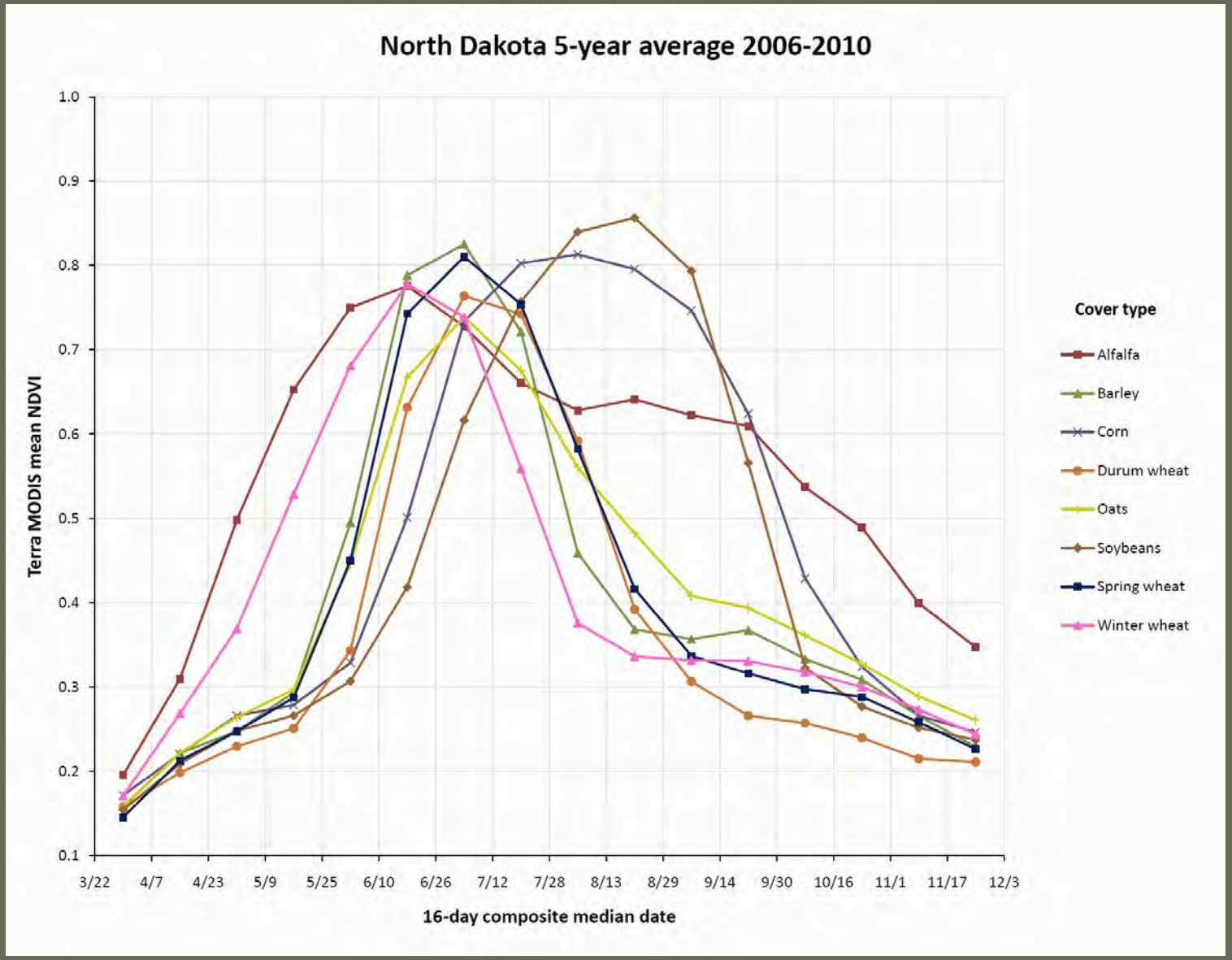
- Herbaceous differentiation
  - Difficult because typically land use (versus cover) designations
  - Little spectral difference between grass types
  - The data from FSA is not usually statistically representative of what is on the ground
  - CRP may be reasonable but we are trying to move away from any land use type categories and only focus on land cover
  - Will continue to haunt us
  - If someone can figure it out they will be a hero
- Recommendation to CDL use is to
  - Treat any grassy cover type cautiously and
  - Lump into a general grassland cover type

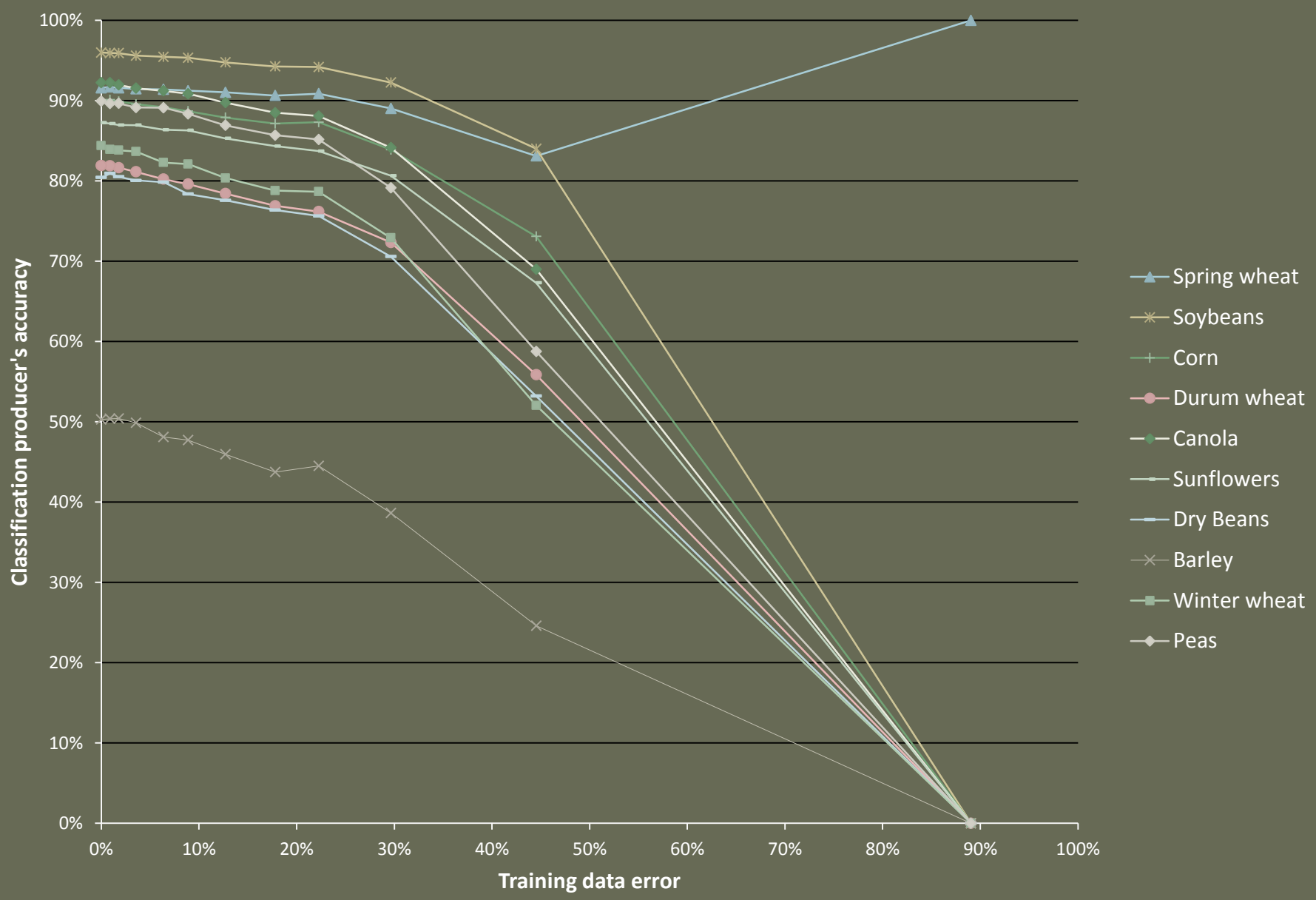


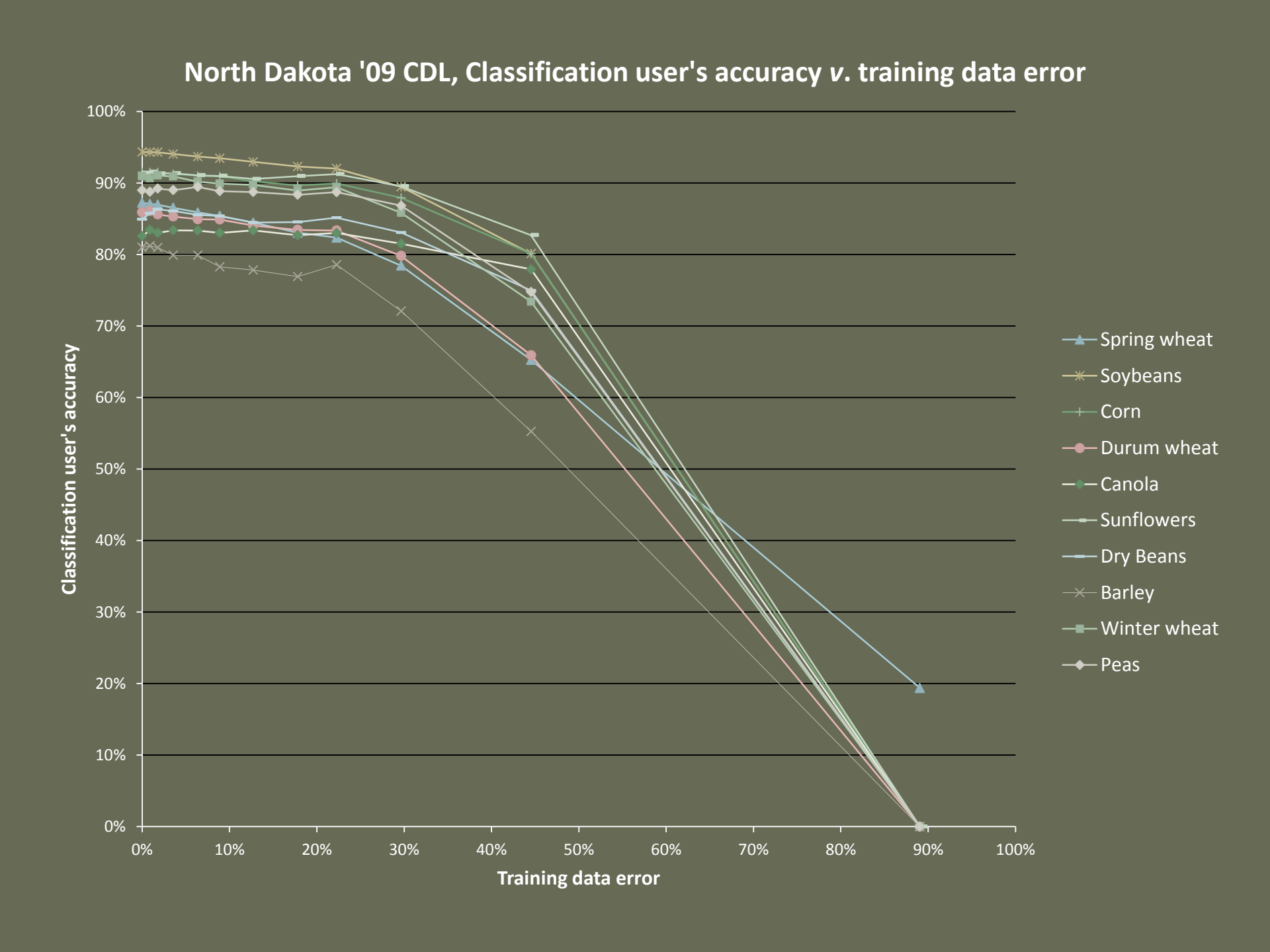
# 2011 CDL campaign

- National cover once again at 30m
- Considering clumping together all grass usage categories into one
  - Avoids perception that we can actually separate them
- Release to public early 2012
- Landsat TM is the primary data source
  - God willing
- DEIMOS and UK2 will also likely be incorporated
  - Native 22m, 3 band
  - But would be resampled to 30m

# North Dakota crop timing




North Dakota '09 CDL, Classification producer's accuracy v. training data error
Classification producer's accuracy
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Training data error
Spring wheat
Soybeans
Corn
Durum wheat
Canola
Sunflowers
Dry Beans
Barley
Winter wheat
Peas


North Dakota '09 CDL, Classification user's accuracy v. training data error
Classification user's accuracy
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Training data error
Spring wheat
Soybeans
Corn
Durum wheat
Canola
Sunflowers
Dry Beans
Barley
Winter wheat
Peas