The USDA-NASS Cropland Data Layer: Early Season Winter Wheat Identification Using Limited Ground Truth

ASPRS Annual Conference – Baltimore, 2009
NASS Overview

Provider of timely, accurate, and useful statistics in service to U.S. agriculture

United States Department of Agriculture
National Agricultural Statistics Service
Research and Development Division
Spatial Analysis Research Section

NASS Overview

Provider of timely, accurate, and useful statistics in service to U.S. agriculture

United States Department of Agriculture
National Agricultural Statistics Service
Research and Development Division
Spatial Analysis Research Section
Brown County, SD ‘08

Land Cover Categories
(Ordered by Decreasing Acreage)

Agricultural
- Corn
- Soybeans
- Spring Wheat
- Alfalfa
- Winter Wheat
- Dry Beans
- Sunflowers
- Oats
- Millet
- Sorghum
- Barley
- Rye
- Other Crops

Non-Agricultural
- Grass/Pasture/Non-Ag
- Urban/Developed
- Wetlands
- Water
- Woodland
- Fallow/Idle Cropland
- Barren
- Shrubland
Purpose of the Cropland Data Layer (CDL) Program

Typically, the CDL program goals are:

1) Combine remote sensing imagery, USDA/Farm Service Agency reported data and NASS survey data to produce supplemental, unbiased independent acreage estimates for the state’s major commodities.

2) Production of a crop-specific digital land cover data layer for distribution in industry standard formats.

Annual CDL states traditionally focused in the Midwest and Mississippi Delta States
- Corn, Cotton, Rice, Soybeans, Winter Wheat

Corn

Soybeans
Methodology

- “Stack” satellite imagery and ancillary data layers within a raster GIS
  - 56 meter grid cells, Albers Conic Equal Area projection
- Sample spatially from stack within known ground truth from FSA and NLCD
- Data-mine samples using Boosted Classification Tree Analysis to derive best fitting decision rules
- Apply derived decision rules back to input data stack
- Create land cover map
- Create probability map
- Assess map accuracy
- Derive acreage estimates
Methodology

• Ground Truth
  – Agricultural training & validation
    • Farm Service Agency (FSA) Common Land Unit (CLU)
  – Non-Agricultural training & validation
    • USGS 2001 National Land Cover Dataset (NLCD)

• Satellite Imagery
  – IRS Resourcesat-1 AWiFS
  – NASA Terra MODIS 16-day composite NDVI
  – Landsat 5

• Ancillary data layers
  – USGS National Elevation Dataset (NED)
  – USGS NLCD 2001 Impervious and Tree Canopy layers

• Software
  – Ground Truth Preparation: ESRI ArcGIS 9.2
  – Imagery Preparation: Leica Geosystems ERDAS Imagine 9.1
  – Decision-Tree Software: Rulequest See 5.0
  – Classification: NLCD Mapping Toolkit
  – Acreage Estimation: SAS
Agricultural Ground Truth
FSA Common Land Unit

- USDA programs (crop subsidy, disaster relief)
- Program crops (may under report specialty crops)
- GIS-ready (less labor intensive for NASS)
Non-Agricultural Ground Truth
USGS, National Land Cover Dataset 2001
Ground Truth Preparation

- Using the NLCD 2001 as a guide, we sample the NLCD non-ag proportionate to the amount of agricultural training data.

Example:

Expected statewide proportions:
- 60% cultivated crops
- 40% non-cultivated

Sample all FSA training data
60,000 pixels

Sample the corresponding proportion of NLCD non-ag
40,000 pixels
Expanding CDL Program Priorities

- Speculative Wheat States
- Speculative Soybean States
- Speculative Corn States
- Speculative Cotton States

Proposed 2009 CDL States
New Objective: Early Season
Winter Wheat Acreage Estimation

Historical:
Crop Production Annual Summary
CDL all crops/county estimates

2007 Crop Acreage Report
CDL winter wheat
The Challenge: Classification with Partial Ground Truth

FSA reporting deadlines
• Different reporting dates for different crops
• State dependent

Illinois example - early June 2008:

Available FSA data: ~ 14% of winter wheat reported
< 1% of any other crop

Classification inputs: AWiFS scenes
080405, 080406, 080415, 080420, 080430, 080504
MODIS 16 day NDVI composites
071016, 071101, 071117, 080406, 080422, 080508
Ancillary data
Elevation, NLCD Canopy, NLCD Imperviousness
The Challenge: Classification with Partial Ground Truth

June

September

http://www.nass.usda.gov/Charts_and_Maps/Crop_Progress_&_Condition/
Initial Classification

Normal CDL methodology

50% of the FSA data
(almost entirely winter wheat)

Proportionately sampled NLCD

Drastic winter wheat over classification
Initial Classification

Expected distribution during a normal year

Initial classification of winter wheat
Goal: Create a Historical ‘Other Crops’ Data Layer for Training

• Identify areas where:
  – All past CDL’s were cultivated crops
  – Exclude winter wheat

• Logic problem
  – Recode the cultivated ‘other crops’ to 1
  – Recode winter wheat and non-ag to 0
  – Multiply the layers together
Goal: Create a Historical ‘Other Crops’ Data Layer for Training

- Model created using Erdas Imagine Spatial Modeler
- Example for Illinois 1999-2007
Further refinements to the ‘Other Crops’ Mask

- More logic problems:
  - Mask out all winter wheat identified in the initial FSA-only classification.
  - Mask out all NLCD non-ag land cover.
- Applies to older CDLs—before the NLCD was used for non-ag training.
Add ‘Other Crops’ as Training

- Sample proportionately from:
  - The refined ‘other crops’ data layer
  - Available FSA data
  - Non-ag NLCD

2008 Illinois expected statewide proportions:
- 3.5% winter wheat
- 56.5% other crops
- 40% NLCD non-ag

Sample all of the early season FSA data
- 104,000 pixels of winter wheat

Sample the corresponding proportion of the ‘other crops’ mask
- 1,670,000 pixels of other crops

Sample the corresponding NLCD non-ag
- 1,180,000 pixels of NLCD non-ag
Resulting Classification

Classification using the historical ‘other crops’ as a training input
Results: Improved Acreage Estimation

• Compared to the official NASS 2008 Illinois harvested winter wheat estimate:
  
  – The initial classification using normal methodology over classified by 127,000 acres (+ 10%)

  – Using the historical ‘other crops’ technique under classified by 10,000 acres (- 1%)

• Acreage NOT just about counting pixels, regression based estimator
Regression-based Acreage Estimator

- June Survey summaries
- Area Sampling Frame
- CDLs
Improvements / Limitations

- Better understanding of crop rotation patterns
  - Claire Boryan, 9:15am Thursday, TS34 Session

- Limitations
  - Requires reliable historical data
  - Some knowledge of expected output

- Future testing:
  - Iterative classification approach for states with no historical data
Proposed 2009 CDL States

Additional Information and **Free** Downloads:
http://www.nass.usda.gov/research/