Impact of Clouds on Late-Season Agricultural Land-Cover Classifications in Kentucky

Mary Lindsey
University of Maryland, Department of Geography
United States Department of Agriculture, National Agricultural Statistics Service Graduate Research Fellow

Introduction

Each year, the United States Department of Agriculture National Agricultural Statistics Service (USDA/NASS) prepares an agricultural land-cover classification product called the Cropland Data Layer (CDL) for major agricultural regions of the US. NASS uses the CDL to make operational, in-season acreage estimates for decision-maker support.

Initially, NASS produced CDLs for two states per year, classifying only three crop types and using a limited number of Landsat scenes (Allen and Hanuschak 1988). Advances in commercial software, newer multispectral satellites, robust ground truth, and the ability to integrate ancillary layers have allowed NASS to produce classifications multiple times throughout the growing season (Mueller and Seffrin 2006), in an expanding program that included 21 states in 2008.

Objective

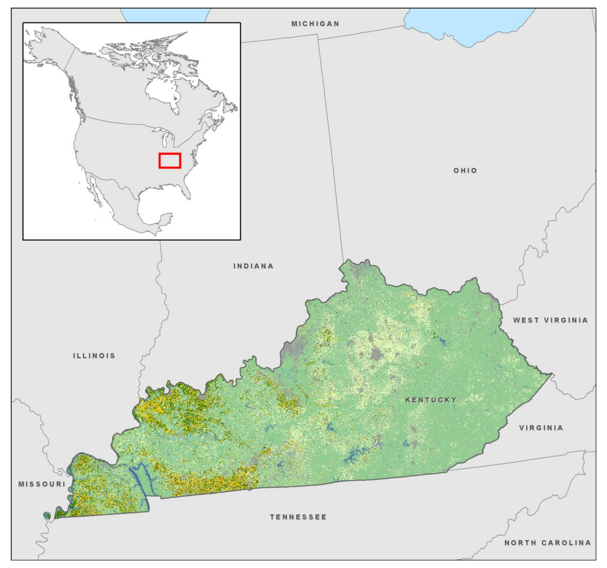
One of the more time-consuming steps in the classification is the manual examination of AWiFS imagery to determine if a scene is too cloudy to include without pre-processing, and to winnow down the total number of scenes.

This study tested the feasibility of an automated classification process that does not include manual examination or winnowing of imagery, a time-saving measure that would allow for an annual, national CDL product.

Study Area

Kentucky was selected for this study because it was not previously included in the CDL program, and it has approximately 13.7 million acres of farmland – 54% of the state's total acres (USDA NASS 2007).

Additionally, some of Kentucky's most crop-intensive areas are at least partly cloudy 70% of the year (NOAA NCDC 2008), making it an ideal test for the effect of clouds on the NASS classification methodology.



USDA NASS Classification Overview

Parameters
- All datasets resampled to 56m, projected to Albers Equal Area Conic, and clipped to a 10km-buffered Kentucky border

Inputs
- Resourcesat-1 Advanced Wide Field Sensor (AWiFS) 56m time series images consisting of four bands – green (0.52 – 0.59µm), red (0.62 – 0.68µm), near IR (0.77 – 0.86µm), and mid IR (1.55 – 1.70µm)
- NASA MODIS 250m 16-day Normalized Difference Vegetation Index (NDVI) composites spanning the growing season
- USGS National Elevation Dataset, National Land Cover Database (NLCD) forest canopy and impervious surface data sets

Ground Truth
- USDA Farm Service Agency (FSA) field-level farmer-reported tabular data linked to FSA Common Land Unit (CLU) shapefiles for agriculture (ag) land cover categories
- NLCD 2001 for non-agriculture (non-ag) land cover categories

Classification Method
- Decision tree by Rulequest See5 Data Mining Tool
- Sampling and classification by NLCD Mapping Tool and ERDAS Imagine

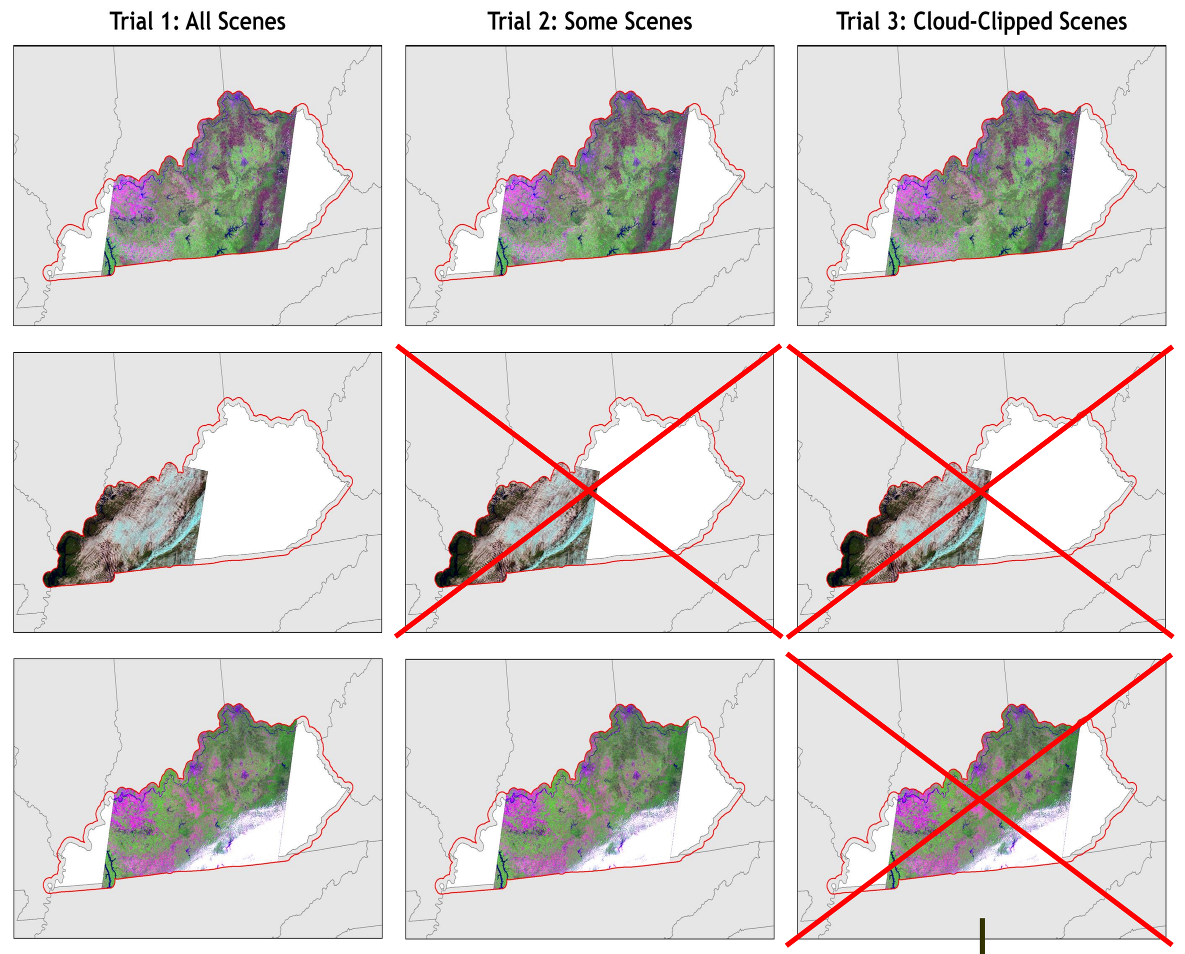
Validation
- Ground truth split into two independent sets for training and validation
- Accuracy assessed by confusion matrix and by regressing classification acres against reported acres in NASS June Area Survey (JAS) segments

Methodology: AWiFS Preprocessing

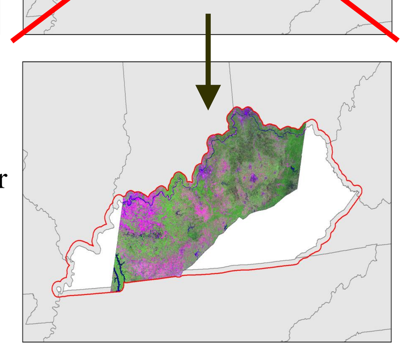
AWiFS scenes covering the growing season from 01 May, 2008 through 21 September, 2008 were clipped to the Kentucky border.

Trial 1: Scenes that did not intersect with Kentucky were discarded. All other scenes, regardless of cloud contamination, were kept. Forty-five scenes used.

Trial 2: Predominately cloudy scenes were excluded. Thirty-nine scenes used.



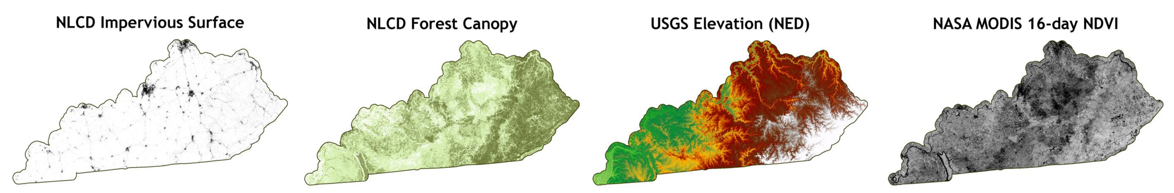
Trial 3: Remaining 39 scenes were examined for clouds and haze. Areas containing continuous clouds, 'popcorn' clouds, or haze were clipped from 31 scenes.



The remaining steps were handled identically for all trials.

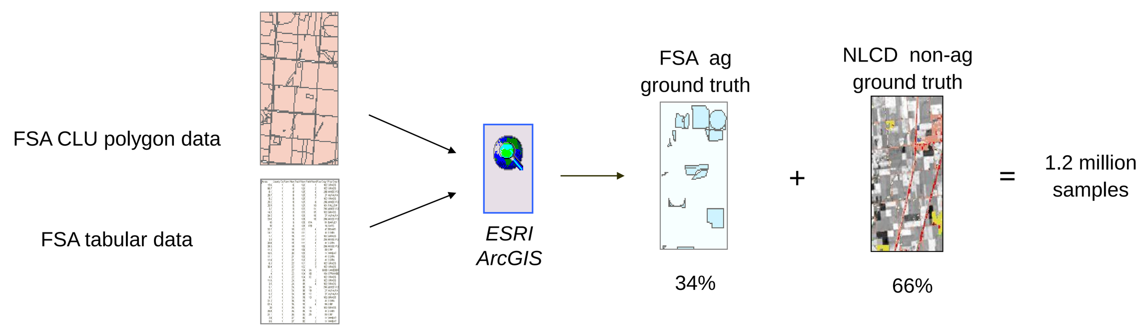
Methodology: Remaining Steps

Prepared remaining inputs

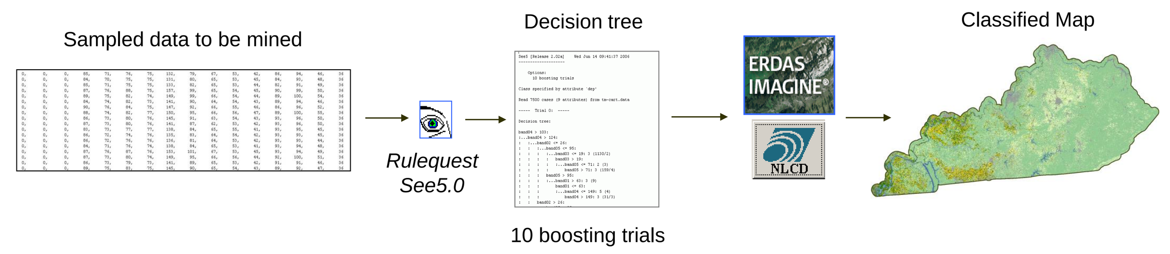


Prepared and sampled ground truth

USDA FSA Common Land Unit (CLU) shapefiles linked to field-level farmer-reported tabular data, buffered in by 56m, split into two sets for training and validation, and then rasterized. NLCD 2001 used for non-ag classes. All inputs were stacked and a stratified random sample was taken, proportionally balancing ag to non-ag samples to match the Kentucky ag to non-ag land cover ratio.



Created decision tree and classification



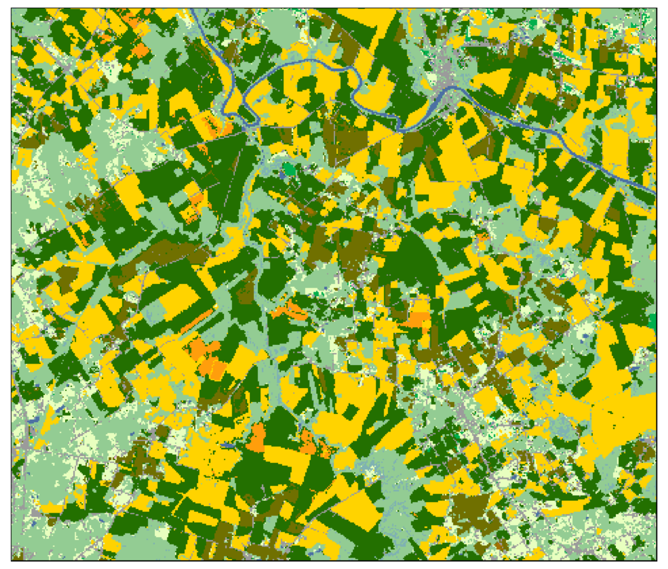
Validated

The accuracy of the classified image was validated by confusion matrix using withheld FSA ground truth. Additionally, acres in the classified image were regressed against reported acres from approximately 200 NASS June Area Survey segments to create acreage estimates.

Results

Below are samples of each trial's classified map, accuracy measures from the confusion matrix, and charts showing the results from the regression estimation. *User's accuracy* is the percentage of times the classified pixels matched the ground truth. *Producer's accuracy* is the percentage of times the ground truth pixels were correctly mapped. *Kappa* is a modified accuracy measure that takes into account the probability of labeling a pixel correctly by chance (Congalton and Green 1999). The regression charts compare the acreage estimates from the classification to the official NASS acreage estimates.

Trial 1: All scenes



Trial 2: Some scenes

Trial 3: Cloud-clipped scenes

Crop Categories: Corn, Sorghum, Soybeans, Winter Wheat, Wntr. Wht./Soy Dbl. Cropped

Non-crop Categories: Pasture/Grass/Non-ag, Wetlands, Water, Developed, Forest

Confusion Matrix Accuracy Assessment

Trial 1

	Overall Accuracy (%)	Kappa Statistic (%)		
All FSA Crops	91.92	87.74		
Crop Category	Producer Accuracy (%)	Kappa Statistic (%)	User Accuracy (%)	Kappa Statistic (%)
Corn	95.43	94.72	95.21	94.48
Soybeans	92.28	91.37	91.51	90.52
Winter wheat	16.03	15.98	73.54	73.48
WW / Soybeans	95.44	95.18	89.76	89.20

Trial 2

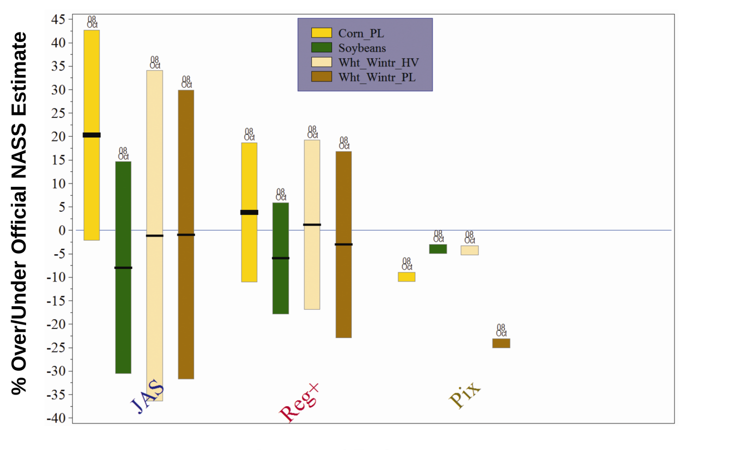
	Overall Accuracy (%)	Kappa Statistic (%)		
All FSA Crops	92.05	87.93		
Crop Category	Producer Accuracy (%)	Kappa Statistic (%)	User Accuracy (%)	Kappa Statistic (%)
Corn	95.39	94.68	95.39	94.68
Soybeans	92.50	91.61	91.63	90.66
Winter wheat	16.03	15.97	64.15	64.06
WW / Soybeans	95.25	94.97	89.86	89.30

Trial 3

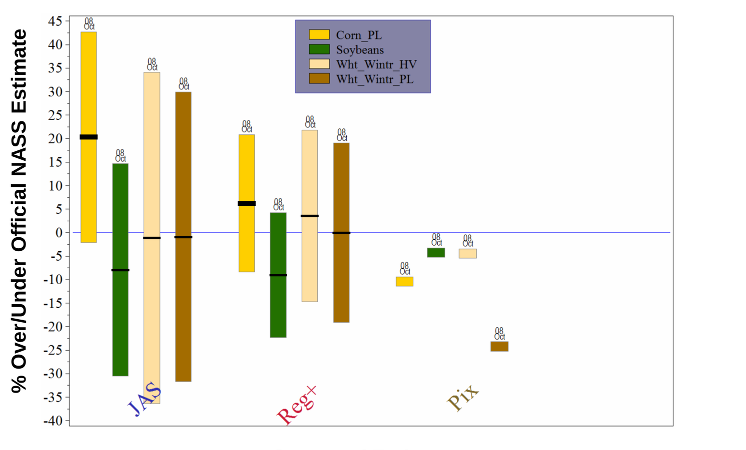
	Overall Accuracy (%)	Kappa Statistic (%)		
All FSA Crops	91.82	87.58		
Crop Category	Producer Accuracy (%)	Kappa Statistic (%)	User Accuracy (%)	Kappa Statistic (%)
Corn	95.33	94.61	95.15	94.40
Soybeans	92.15	91.23	91.42	90.42
Winter wheat	19.07	19.02	78.98	73.48
WW / Soybeans	95.16	94.88	89.86	89.30

Regression Estimation Accuracy Assessment

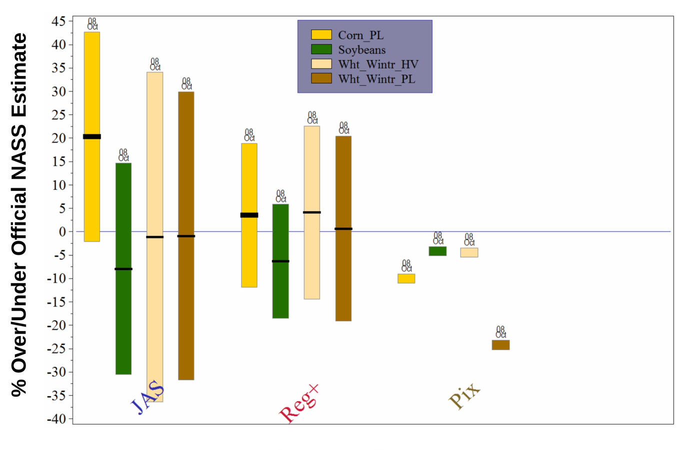
Regression Analysis and Variance +/- 2% CVs



Regression Analysis and Variance +/- 2% CVs



Regression Analysis and Variance +/- 2% CVs



JAS = June Area Survey Estimate, Reg+ = Classification Estimate, Pix = Pixel Counting, PL = Planted, HV = Harvested

Conclusions

Trial 3 represents the current NASS classification approach (most time consuming), while Trial 1 represents an approach that does *not* require the analyst to manually examine scenes or remove clouds (least time consuming). Trial 2 is an intermediate approach: predominantly cloudy scenes were removed (some time required), but no scenes had clouds clipped out.

All three classification methodologies performed well, with each classification's acreage estimates (Reg+ on the charts) lying well within the error bars of the estimates derived from the June Area Survey segments. All measures of accuracy show the methodological approach of Trial 1 – using all AWiFS scenes, regardless of cloud contamination – performed as well as Trial 3.

In fact, all three trials produced almost identical results. Therefore, this study demonstrates that an automated classification process is feasible. The efficiencies gained without loss of accuracy make an annual, national CDL product possible.

However, this study tested the effects of cloudy scenes on a *late-season* classification, when an abundance of imagery was available. NASS currently produces early-, mid-, and late-season classifications for 'speculative' states (states with high production of crops traded on commodity markets), and it is quite possible that the presence of cloudy scenes in an early-season classification, when less total imagery is available, would degrade the results. An automated classification may not be appropriate as a replacement of the current NASS process, but it offers a way to include the remaining non-speculative states once per year.

References

Allen, J. D., and Hanuschak, G. 1988. The Remote Sensing Applications Program of the National Agricultural Statistics Service: 1980-1987. SRB Staff Report Number SRB-88-08 USDA, NASS, Washington, DC-USA.

Congalton, R. G., and Green, K. 1999. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton: Lewis Publishers.

Mueller, R. and Seffrin R. 2006. New Methods and Satellites: A Program Update on the NASS Cropland Data Layer Acreage Program. *Proceedings of the Symposium on Remote Sensing Support to Crop Yield Forecast and Area Estimates*, ISPRS Commission VII, WG VI/4, Stresa, Italy.

NOAA National Climate Data Center. 2008. Cloudiness - Mean Number of Days data table. *NOAA NCDC Comparative Climatic Data Publication*. NOAA NCDC Website. Last accessed 1 March 2008. http://lwf.ncdc.noaa.gov/oa/climate/online/ccd/cldy.html

USDA NASS, Kentucky Field Office, 2007. Kentucky Agricultural Facts Pamphlet. USDA NASS Statistics by State, Kentucky Publications Website. Last accessed 1 March 2008. http://www.nass.usda.gov/Statistics_by_State/Kentucky/Publications/Pamphlets/

Acknowledgments

This research would not have been possible without the generous time, effort, and financial support of Dr. Chris Justice at the University of Maryland, Dr. Jessica McCarty at the University of Louisville, Dr. Matt Hansen at South Dakota State University GIScCE, and the NASS Spatial Analysis and Research Section – Rick Mueller (head), Claire Boryan, David M. Johnson, Patrick Willis, Mike Craig, Bob Seffrin, Audra Zakzeski, Zhengwei Yang, and Lee Ebinger.

Correspondence

Email: *lindsey@umd.edu*
Web: *www.nass.usda.gov/research*
www.geog.umd.edu