Discussion: Combining Information from Survey and Non-Survey Data Sources for Policy Research

Linda J. Young
USDA National Agricultural Statistics Service
May 6, 2016

Outline

- Multiple Frames
- Bayesian Hierarchical Models
- Small Area Estimation

NASS List Frame

- List of agricultural producers and agribusinesses
  - Names, addresses, telephone numbers and e-mail addresses
  - Grouped by size and type of unit
- Kept as complete as possible
- Used as the sampling frame for numerous surveys and the Census of Agriculture

NASS Area Frame

- Covers all states but Alaska
- Has complete coverage
- 2009 Farm Numbers Research Project
  - Extensive misclassification during the screening prior to the June Area Survey
  - Resulted in undercount of the 2009 number of farms by an estimated 581,373
- Could no longer assume that NASS area frame could be used to fully account for undercoverage of NASS list frame

What Operations Are Being Missed?

- Both the NASS list frame and the NASS area frame provide good coverage for large producers
- Emerging sectors of agriculture
  - Organics, local foods, horticulture, urban agriculture
  - Operations tend to be
    - Smaller
    - Newer
    - More transient
    - More widely dispersed

  compared to more traditional agricultural operations in rural areas

MACE List Frame

Developed using web scraping

Data sources:

- Social media, e.g., facebook, twitter, etc.
- Information on permits from city and state governments
- Lists available from various sources

Result:

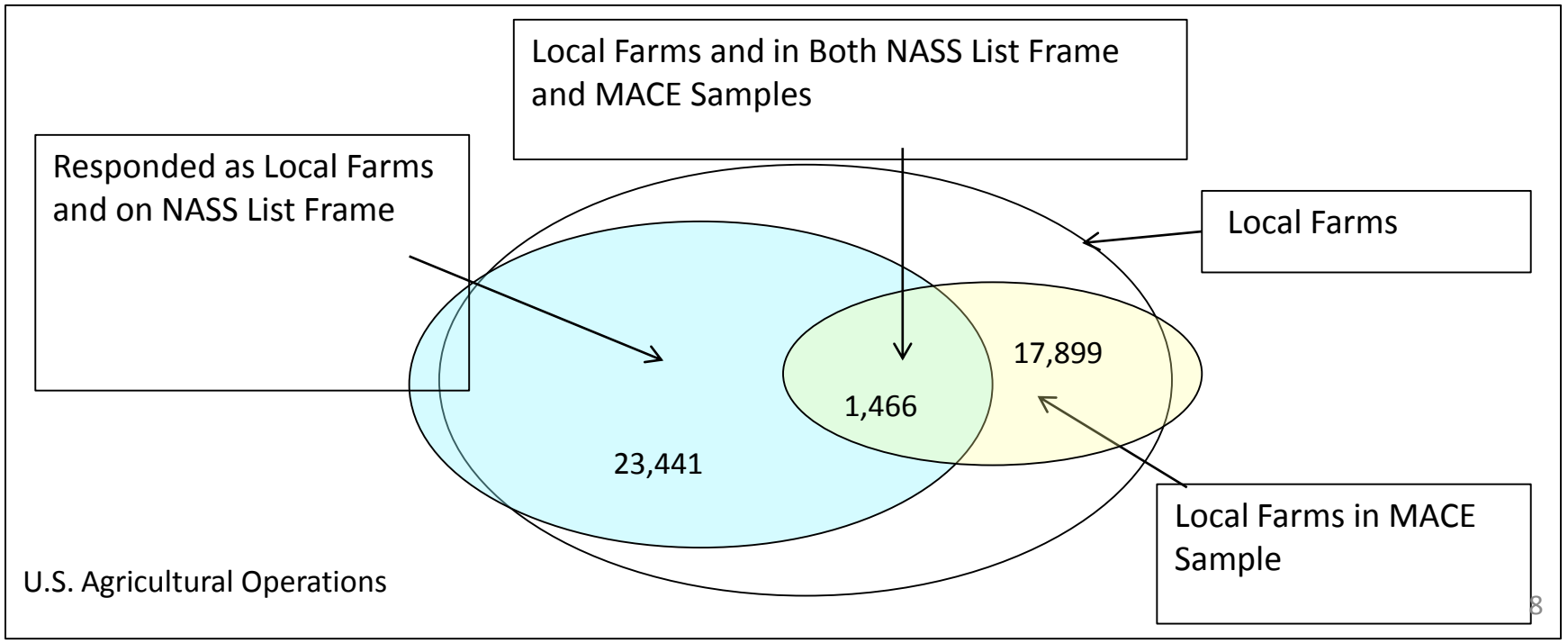
- List frame of potential operations having the agricultural activity of interest

2015 Local Foods Survey

- 2,007,110 on NASS List Frame
  - Includes all (not just local foods) operations
  - Consists both of confirmed farms and potential farms
- 33,394 on MACE List Frame, which only has potential local foods farms
  - Are not confirmed to be farms
  - In Baltimore urban ag pilot study about half had agricultural activity

Primary Assumptions for Capture-Recapture Analysis

- Two Independent Samples:
  - NASS List Frame
  - MACE List Frame
- Proportion of MACE local foods farms captured by the NASS list frame sample is equal to the proportion of the US local foods farms captured by the NASS list frame sample

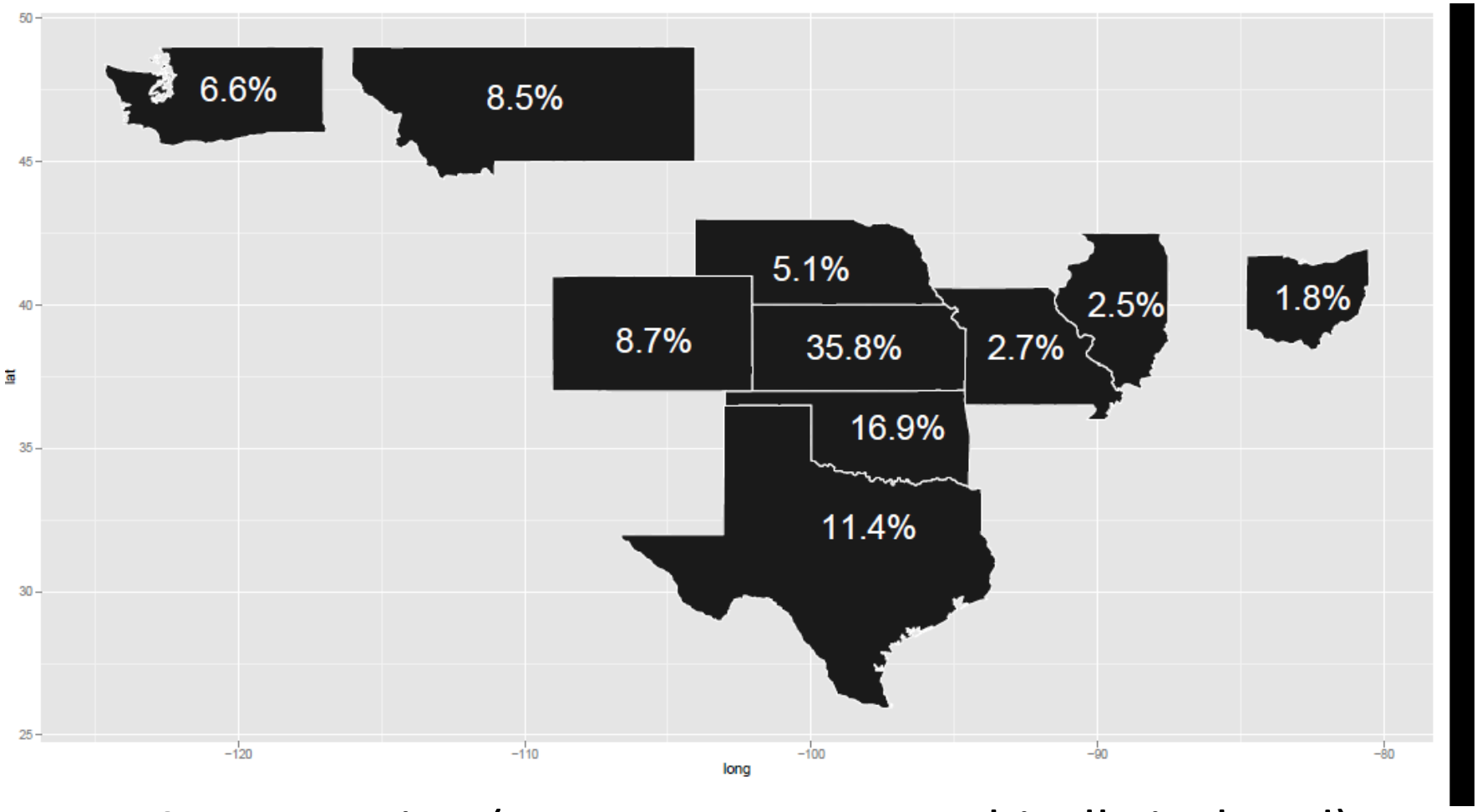


Multiple Frames and Coverage

- Goal: Estimate totals
- Multiple frames do not provide complete coverage
- Capture-recapture is one approach to adjusting for coverage, but more research is needed

Winter Wheat

Speculative Region for Winter Wheat

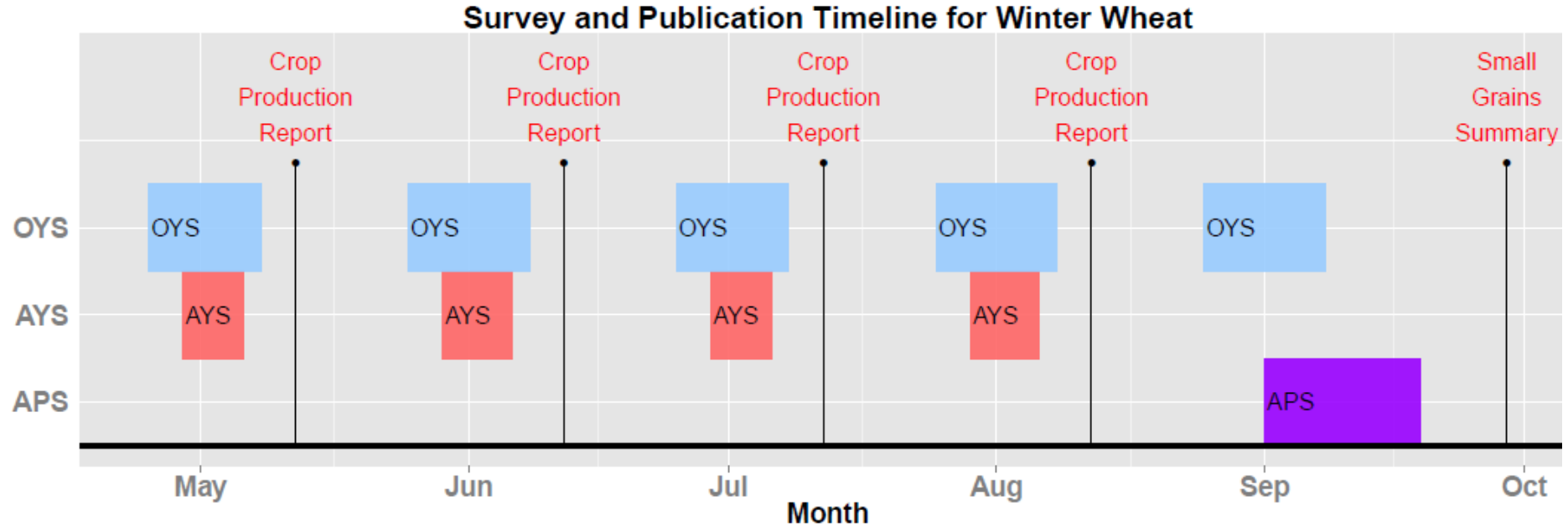


- 10 state region (some states geographically isolated)
- Kansas has major share of harvested acres

Predicting Crop Yields

NASS Crop Yield Surveys

- Objective Yield Survey (OYS): field measurements, conducted monthly in only speculative region states
- Agricultural Yield Survey (AYS): interview, conducted monthly with nationwide coverage
- Acreage, Production, and Stocks (APS): interview, end of season, large sample sizes, nationwide



Role of Agricultural Statistics Board (ASB)

- Expert panel of commodity specialists
- Current and historical survey estimates
- Other information, e.g., weather, crop condition, remote sensing
- Consensus on yield—set national and speculative region yield during deliberations
- See Bernard (2007) for additional description of ASB process

Publish State and National Estimates

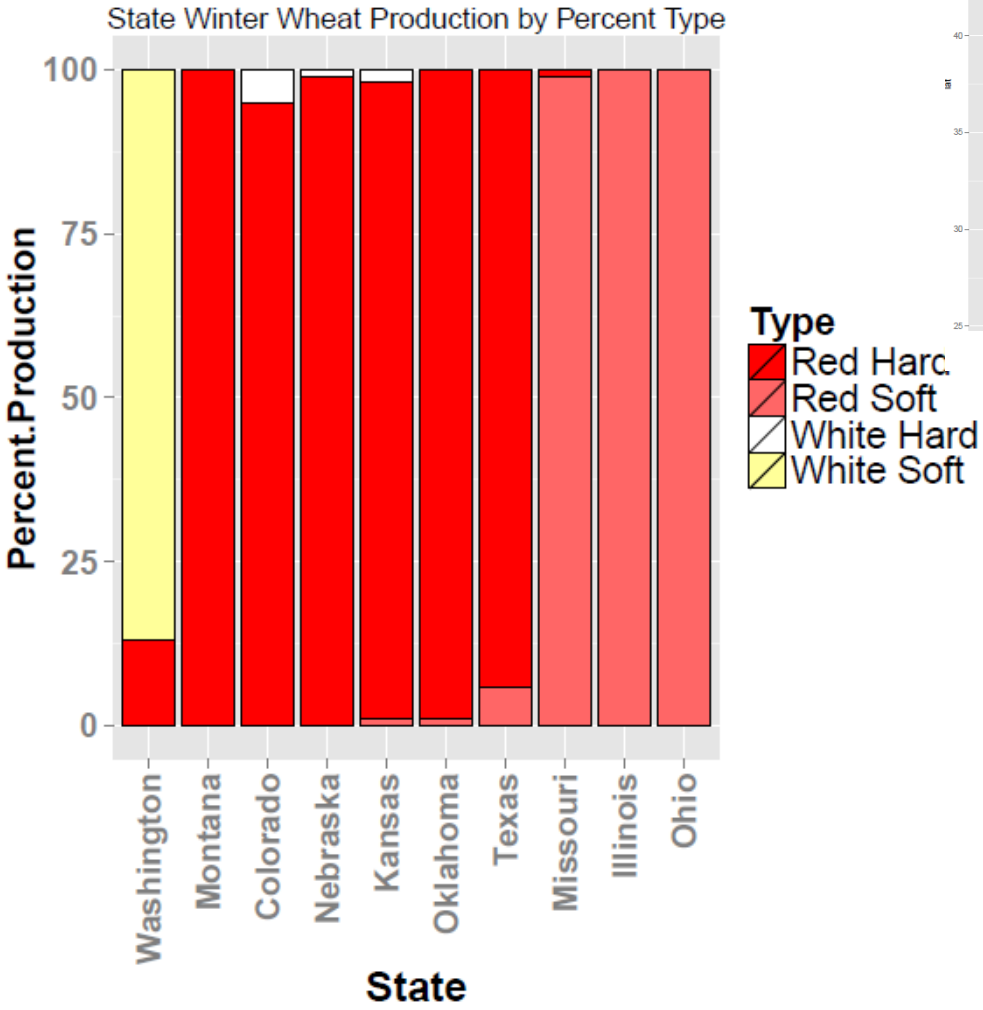
Challenges

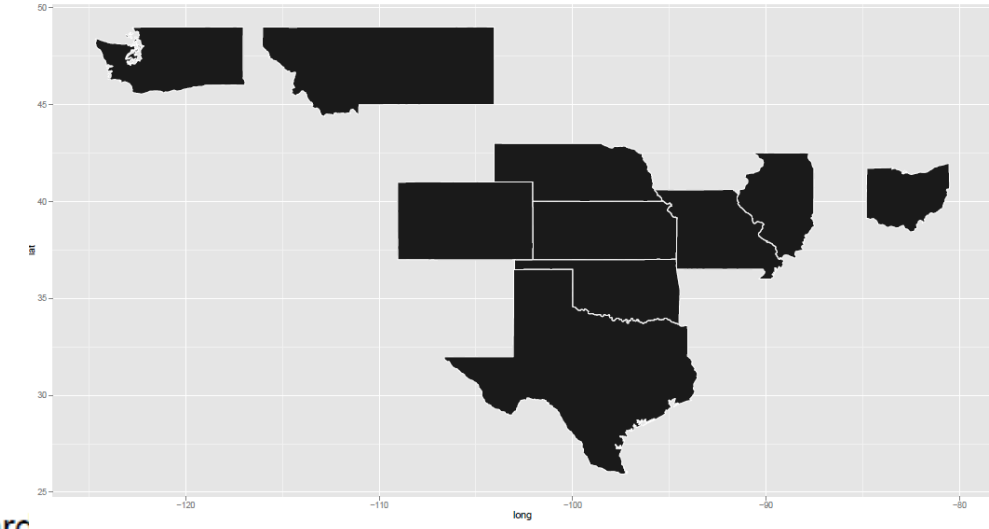
- Lacks transparency
- Lacks reproducibility

Bayesian Hierarchical Models of Yield

- Provide model-based yield to NASS Agricultural Statistics Board: corn and soybean (2011) and winter wheat (2015)
- Development of models to combine data
  - Include factors considered by the experts
    - Survey data
    - Remotely sensed data
    - Weather information
    - Farming practices
  - Potential additional sources
    - Additional remotely sensed data
    - Use of drones?
    - On-farm information

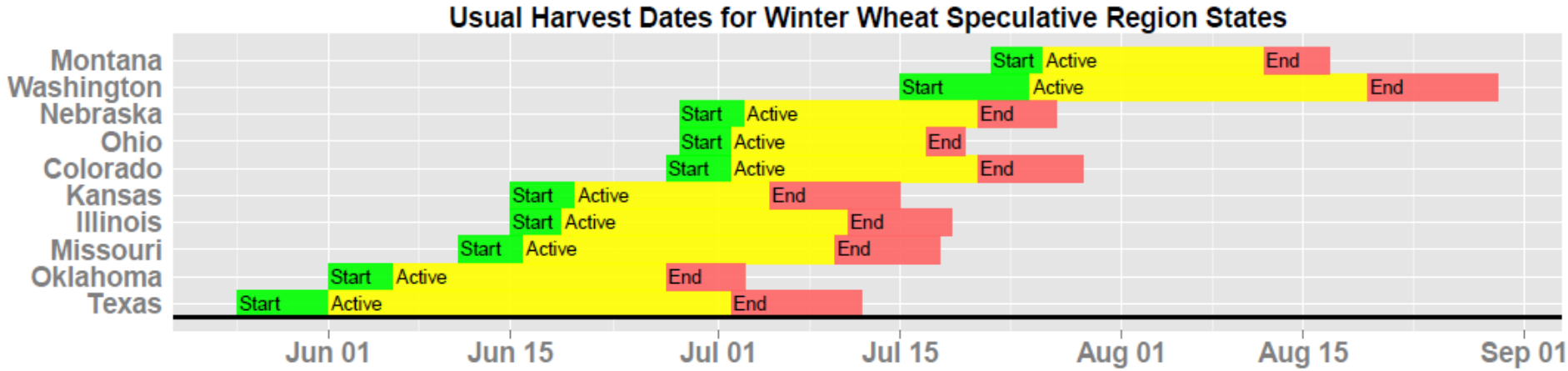
Winter Wheat Speculative Region: Types of Wheat





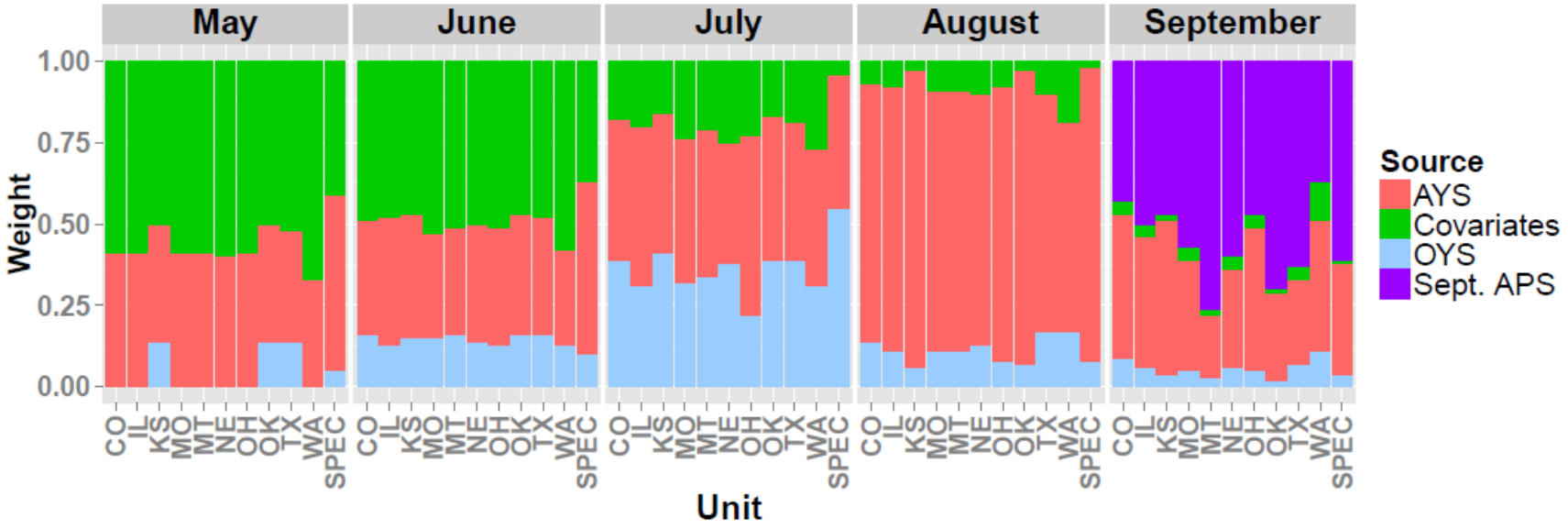
- States `specialize'
- Soft varieties associated with higher yield
- Washington, Missouri, Illinois, Ohio have higher yields
- Confounding with state

Winter Wheat in Speculative Region: Differential Harvest



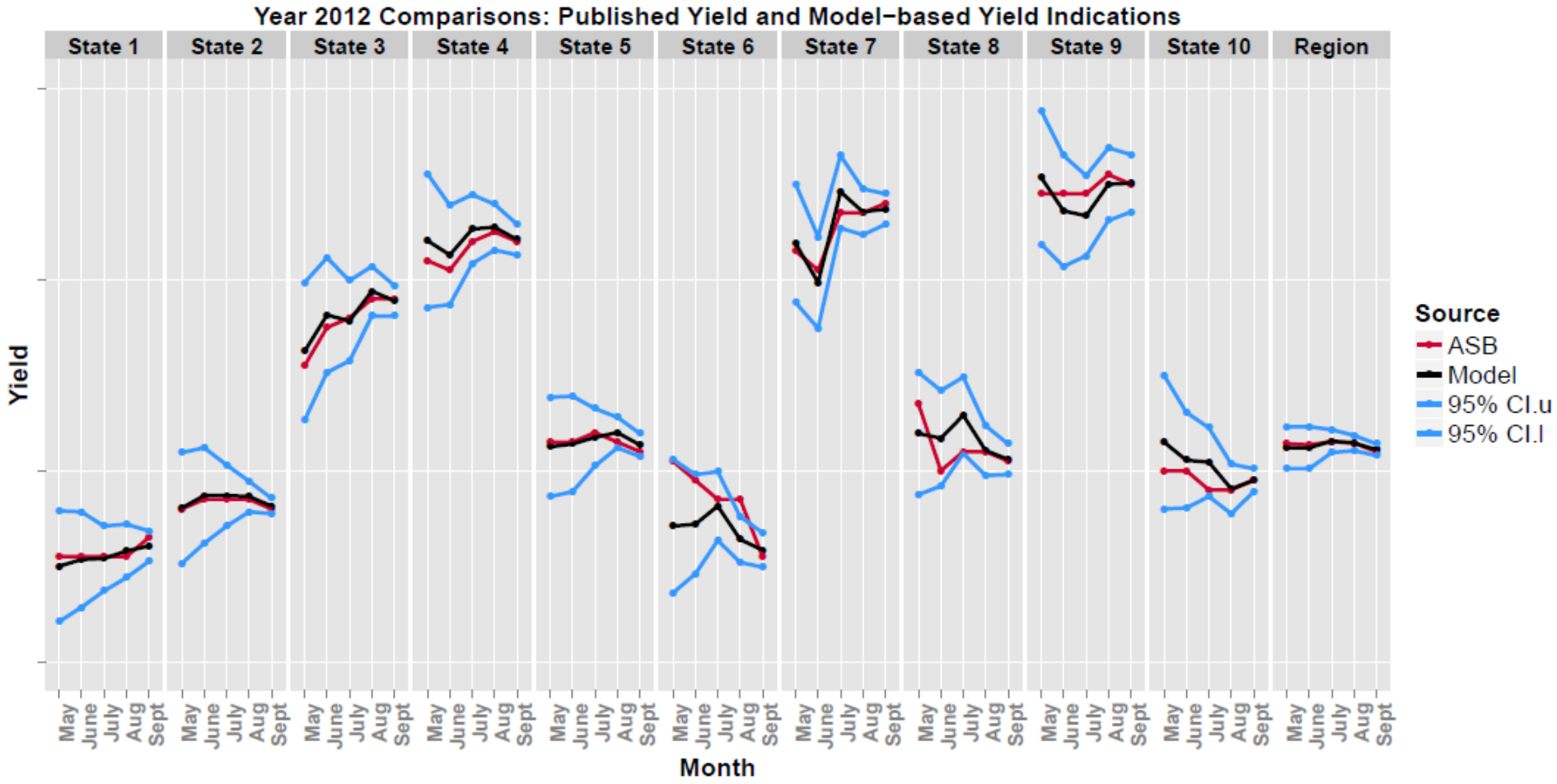
- May OYS: only TX, OK, KS
- Southern states complete harvest before northern states begin
- Timing of covariates
- Deriving covariates for the region

Weights in Wheat Forecast



- Early season emphasis on covariates
- Increasing emphasis on OYS in July
- Heavy emphasis on last AYS in August
- Heavy emphasis on quarterly survey in September

Comparing ASB Forecasts and Model Outputs



Remaining Challenges

- Model provides
  - speculative region predictions/estimates
  - predictions/estimates for states within the region
- Lacks
  - National estimates
  - State estimates outside the speculative region
- No OYS for areas outside the speculative region
- On-farm data may become available
  - Will these lead to improved predictions?
  - Can they be incorporated in the analysis on the production timeline (3 to 4 days from data to publication)

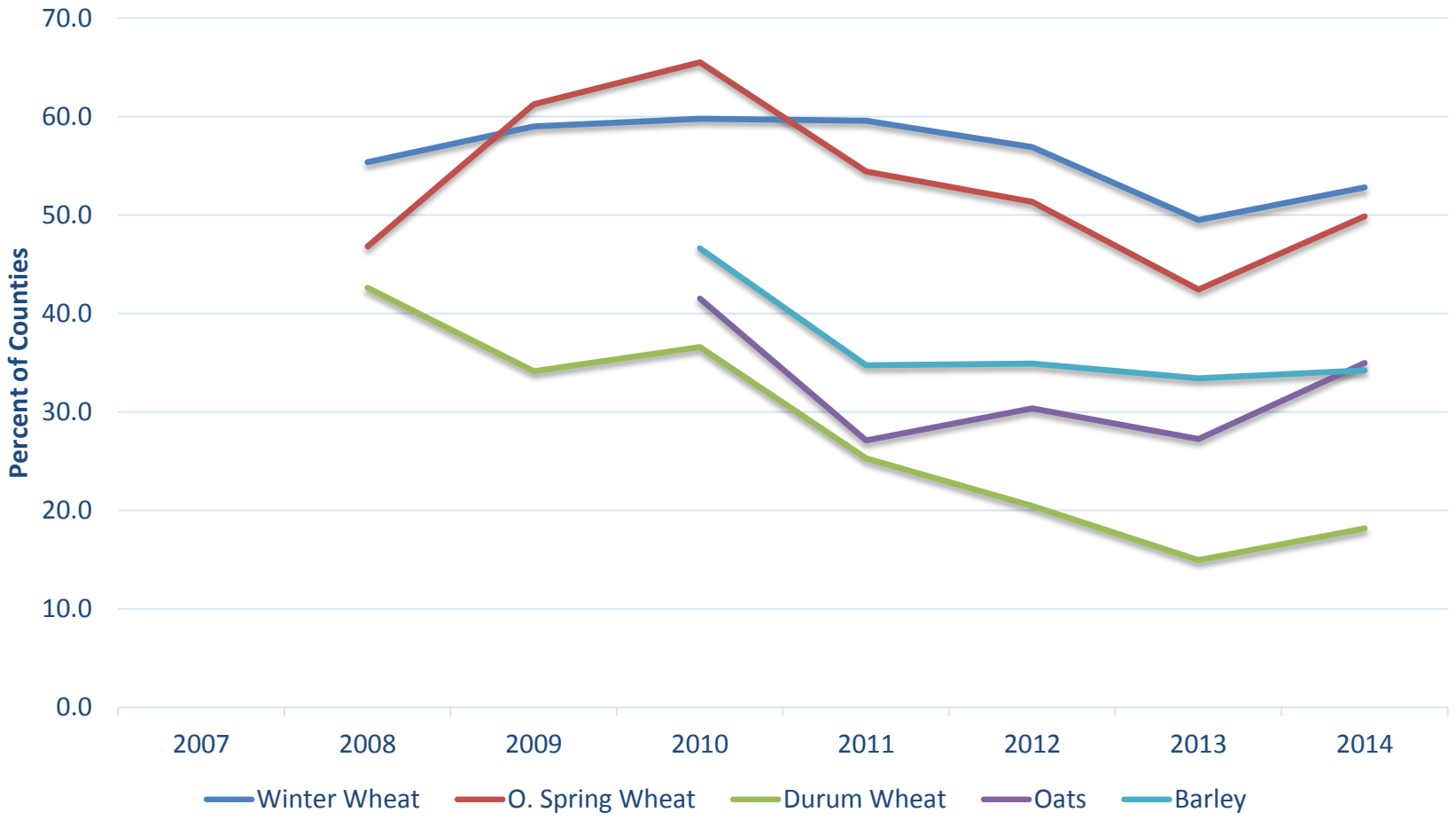
Small Area Estimation

- NASS has two programs that produce county estimates
  - County yield
  - County cash rents
- Both publish direct estimates from survey data
  - Must meet publication standards

Agriculture Loss Coverage County (ARC-CO) Program

- Initiated in 2014 Farm Bill
- Administered by USDA's Farm Services Agency
- Payments are issued when the actual county crop revenue of a covered commodity is less than the ARC-CO guarantee for the covered commodity
- Hierarchy of estimates
  - National Agricultural Statistics Service county yield surveys if available
  - FSA data-mined yields from the Risk Management Agency
  - Crop insurance; crop reporting district; and a yield set by the FSA state committee, using neighboring counties with similar production

Small Grain County Estimates
Percent of Estimated Counties Published



ARC-CO Payments

- Payments totaling $5.7 billion made for 2014 crops
- Small change in county yield can result in large change in ARC-CO payment

WATCH: Scrooged by the survey: Farmers miss out on ARC payments

Lawsuit Against USDA on ARC CO Program Problems

Challenges and Opportunities

- To produce more county estimates, NASS must turn to small area estimation
  - Good progress on acreage model
  - Joint model with yield is more challenging
- Models produce estimates, which have good properties on average, but any one estimate may be poor
- For this and many other applications, the focus is only on the point estimates; standard errors are not used
- Can the human review of the estimates ever realistically be removed?

References

- Abreu, Denise A., Pam Arroway, Andrea C. Lamas, Kenneth K. Lopiano, and Linda J. Young. 2010. Using the Census of Agriculture list frame to assess misclassification in the June Area Survey. *2010 Proceedings of the Joint Statistical Meetings.* Pp. 2339-2350.
- Abreu, D.A., J.S. McCarthy, L.A. Colburn (2010). Impact of the Screening Procedures of the June Area Survey on the Number of Farms Estimates. Research and Development Division. RDD Research Report Number RDD-1003. Washington, DC: USDA, National Agricultural Statistics Service.
- Adrian, D. 2012. A model-based approach to forecasting corn and soybean yields. Fourth International Conference on Establishment Surveys.
- Bernard, R. 2007. Under lock and key. http://www.nass.usda.gov/Newsroom/2007/UnderLockKey_reprint.pdf. [Online, reprinted from Farm Journal, accessed May 2, 2016]
- Cruze, N.B. 2015. Integrating survey data with auxiliary sources of information to estimate crop yields. In *JSM Proceedings, Survey Research Methods Section*. Alexandria, VA: American Statistical Association.
- Nandram, B, E. Berg, and W. Barboza. 2014. A hierarchical Bayesian model for forecasting state-level corn yield. *Environmental and Ecological Statistics* 21(3): 507-530.
- Wang, J.C., S.H. Holan, B. Nandram, W. Barboza, C. Toto, and E. Anderson. 2012. A Bayesian approach to estimating agricultural yield base on multiple repeated surveys. *Journal of Agricultural, Biological and Environmental Statistics* 17(1): 84-106.

Thank you!

Linda.Young@nass.usda.gov