The Crops County Estimates Program: Developing Official Statistics Based on Available Data

Linda J Young
USDA National Agricultural Statistics Service
May 16, 2022

The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

Outline

- Crops County Estimates
- Historical Approach
- CNSTAT Review
- Models for Crops County Estimates
- Current Efforts
- Challenges

Crops County Estimates: Corn

Agricultural Survey

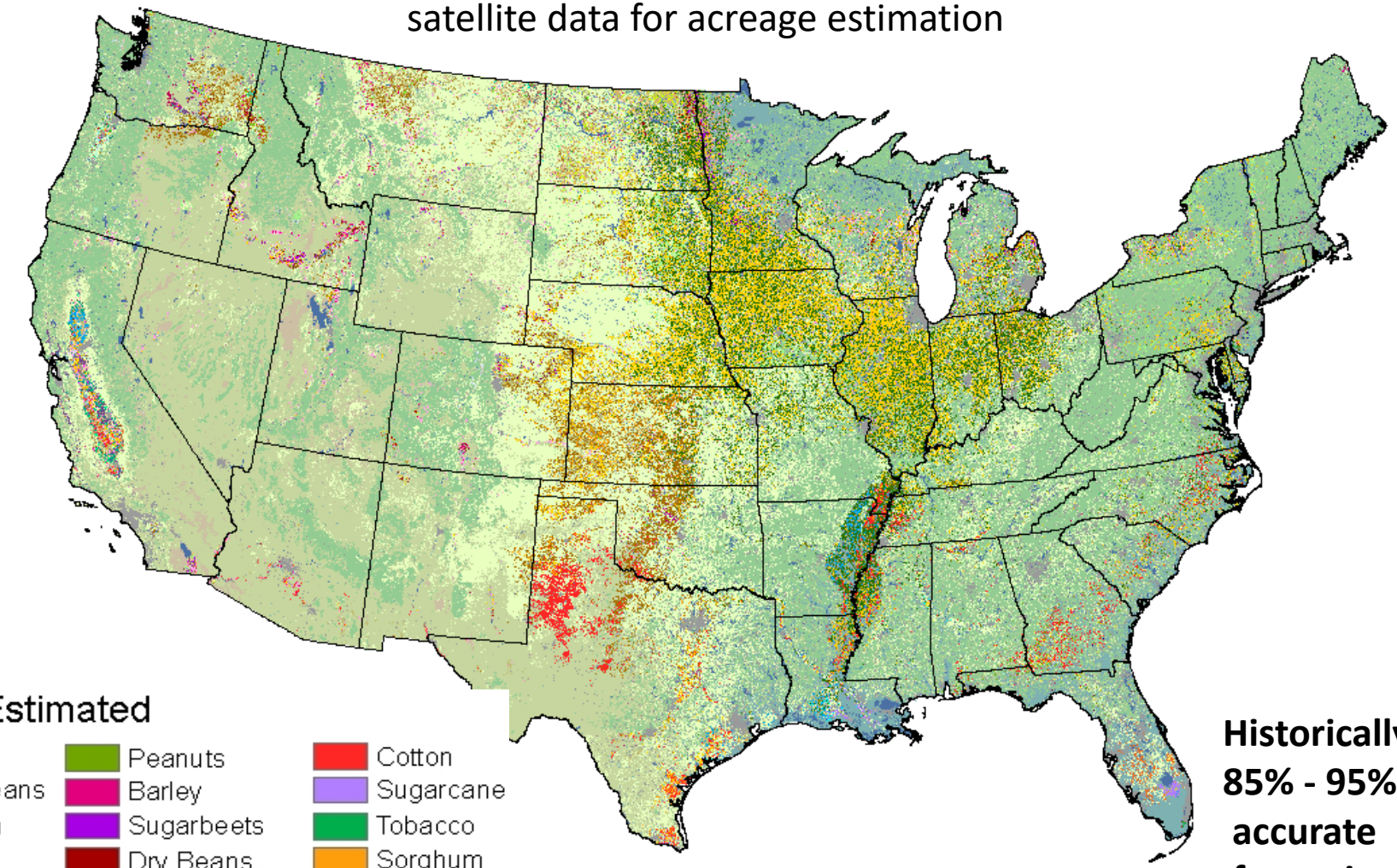
- Conducted quarterly
- National estimates (or predictions) of crop acreage, yields, and production for major crops and quantities of grain and oilseeds stored on farms
- State estimates for the states with the largest production

County Agricultural Survey

- Additional data collected in December
- December surveys provide foundation for county estimates
  - Planted acreages
  - Harvested acreages
  - Production
  - Yield

Cropland Data Layer (CDL)

Annual national coverage since 2008
A raster, crop-specific, land cover data set produced using satellite data for acreage estimation



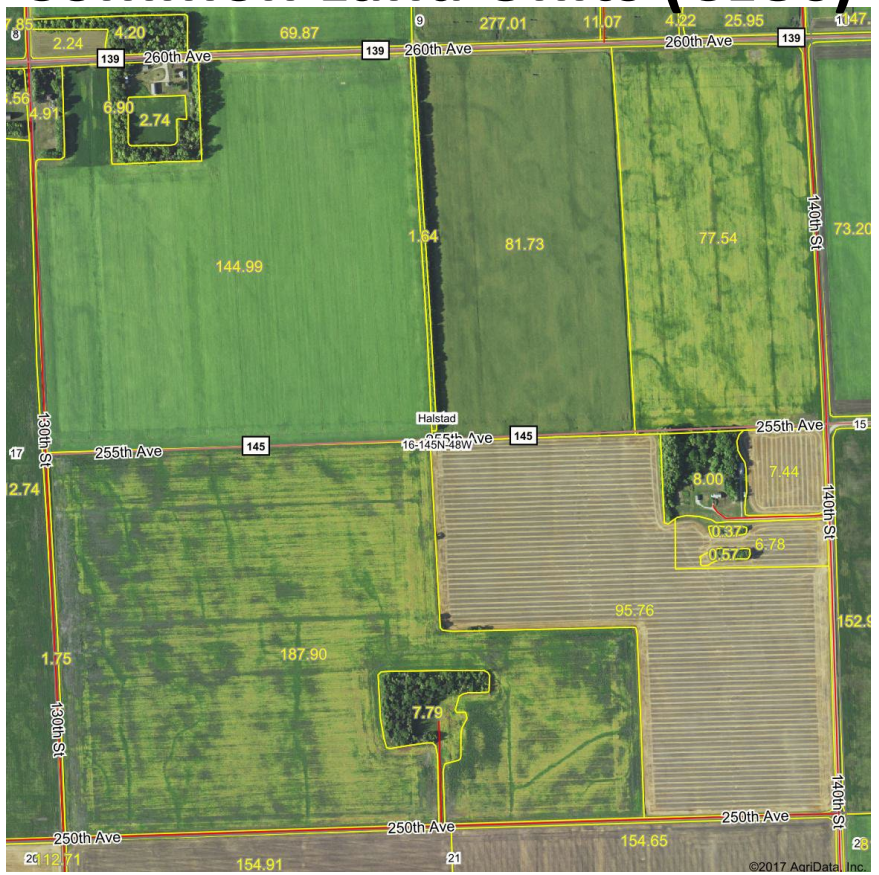
*** 9 billion pixel 30m product**

Historically, 85% - 95% accurate for major crops

Farm Service Agency (FSA) Form FSA-578

- Completed by all producers participating in a USDA program for that crop season
- Information for each Common Land Unit
  - Crops
  - Acreage
  - Irrigation
- Variable coverage for crops and states, but high in major corn states
- Provides lower bound for acreages planted to a crop within a county

Common Land Units (CLUs)



https://www.agridatainc.com/Home/Products/Mapping%20Features/Land%20Resource%20Intelligence/FSA%20Field%20Boundaries%20(CLU)

Risk Management Agency (RMA)

- Manages the USDA's crop insurance program
- Producers report adverse events that impact crops, including decisions not to harvest certain acreages
- Acreages planted and not harvested for each crop
- Other information not available prior to publishing estimates
- Provides
  - Lower bound for the number of acres planted in a county to corn
  - Lower bound for the number of acres in a county planted but not harvested

Expert Opinion to Produce Official Estimates

- Historical approach used by NASS
- State estimates set prior to county estimates
- Available information
  - Survey estimates by county
  - Estimates derived through remote sensing (CDL)
  - FSA- and RMA-reported planted acres by county
  - RMA-reported acres planted but not harvested by county
  - Recommendations from each state based on data and local factors
- Concerns
  - Lack of transparency and reproducibility
  - No valid measure of uncertainty

CNSTAT Panel

- Established through a cooperative agreement in September 2014
- Purpose: Review NASS county estimates programs and provide suggestions for how to better combine all available data to produce improved county estimates for (hopefully) more counties
- Consensus report released in October 2017

Vision for 2025

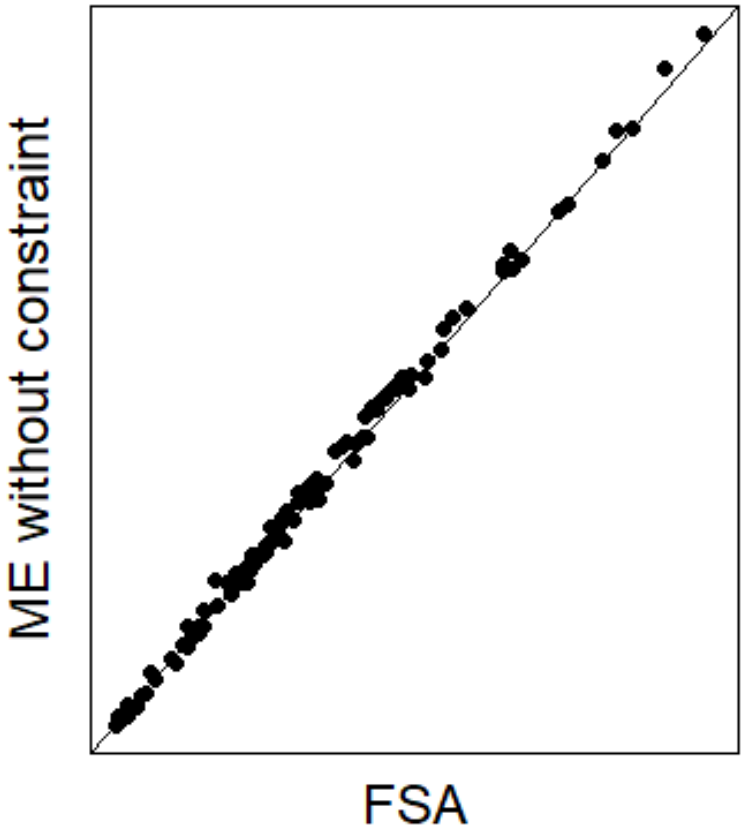
- NASS should evolve its Agricultural Statistics Board process
  - County-level estimates based on models incorporating multiple data sources with uncertainty measures
  - ASB would review predictions, macro-edit, and ensure models are continually reviewed
- NASS should acquire all georeferenced administrative and remotely sensed and ground-gathered data relevant to developing estimates and use the data to complement its survey data
  - Current data (FSA, RMA, Remotely Sensed, Weather)
  - Potential new data (Precision agriculture, in situ sensors, drones, etc.)

Crops County Estimates: Corn

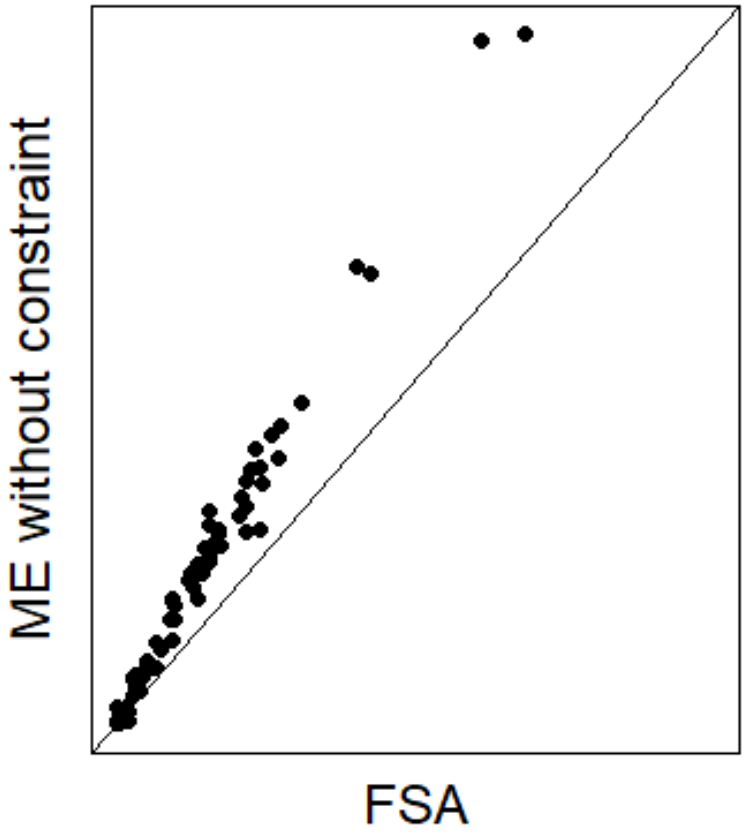
- Three Bayesian hierarchical models used to combine information at the county level
  - **Planted acreage**
  - Harvested acreage, which must be no greater than planted acreage
  - Yield—production estimated by (yield) · (harvested acreage)
- Challenges
  - County estimates must sum to state estimate
  - Honoring the bounds obtained from FSA and RMA
  - Rounding

2019 Modeled County Estimates Without Constraint vs FSA Reported of Planted Acreage

Illinois

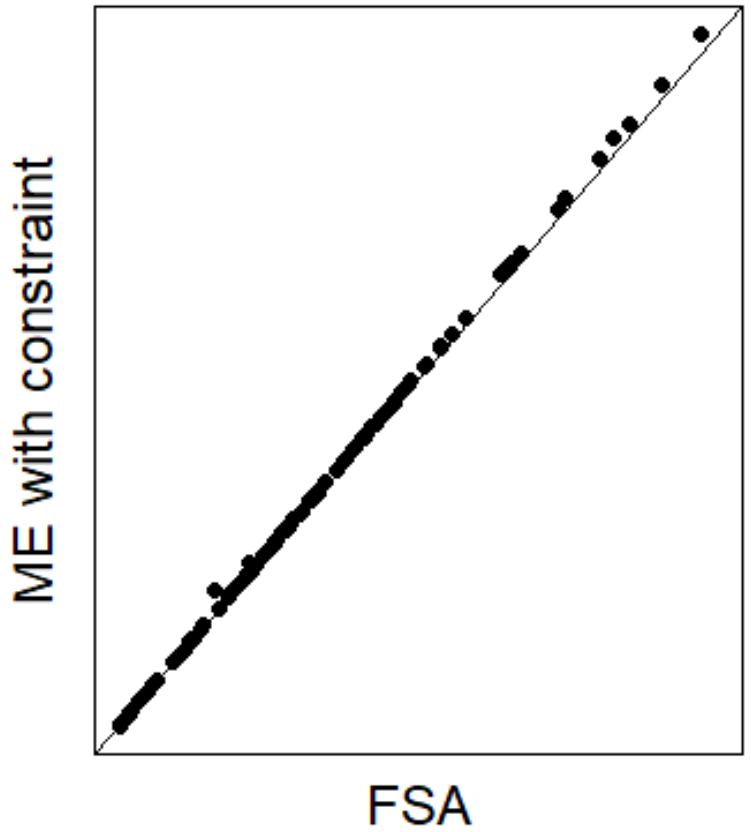


Pennsylvania

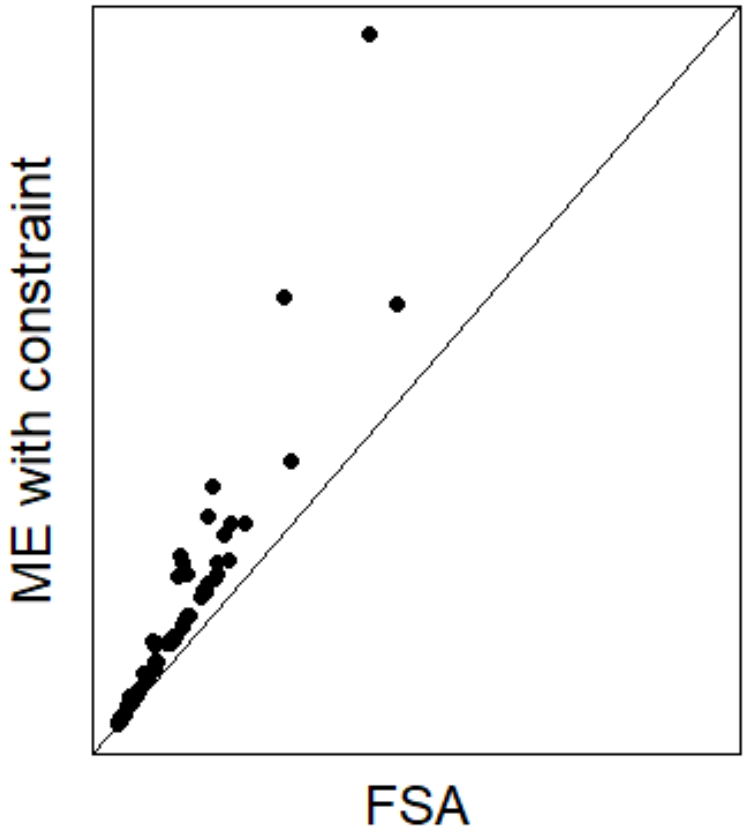


2019 Modeled County Estimates With Constraint vs FSA Reported of Planted Acreage

Illinois



Pennsylvania



Published Crops County Estimates

- Moved into production for 2020 crop year
  - Publish for 13 crops
  - Transparent and reproducible: methodology published
  - Valid measures of uncertainty
  - Rounding for publication was automated
  - Estimates reviewed by staff in State Field Offices
  - Saved substantial hours of staff time within each state

Published Model Estimates

- Models based on aggregated data
  - Crops county estimates
  - Cash rents county estimates
  - Farm labor

Integrating Survey and Non-survey Data

- FSA and NASS have different definitions of a farm
- NASS list frame is not fully geo-referenced
- Surveys
  - Generally, not designed to provide estimates lower than a state
  - Information at farm level does not provide field-level data
- Integration into existing production process
  - Flow of survey and non-survey data
  - Analysis methods
  - Review processes

Other Technologies

- Drones
  - USDA's Advisory Committee on Agricultural Statistics
  - Potential limited use with producer permission
- In-situ sensors
  - Establishing a nationwide system costly
  - Potential to collaborate within proposed AI Institute—Climate Resiliency in Agriculture and Forests (AI-CRAFT)

Precision Agriculture

- Immense potential
  - Reduce respondent burden
  - Improve data quality
- Major challenges
  - Access
  - Coverage
  - Data standards
  - Data quality
- Who pays?

Final Thoughts

- NASS conducts more than 100 surveys and produces over 400 reports each year
- Challenges to integrating survey and non-survey data
  - Record level versus higher level of geography
  - Survey design
- Major effort underway to modernize processes

Acknowledgments

Thank you to the numerous NASS staff upon whose work I built this presentation. A special thanks goes to Lu Chen who incorporated the constraints within the Bayesian hierarchical models and developed the graphs and most of the reference list presented here.

Selected References

Boryan C, Yang Z, Mueller R, Craig M. (2011) Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int*. 26(5):341–58.

Chen, L., Nandram, B., Cruze, N.B. (2020) Hierarchical Bayesian model with inequality constraints for county estimates. In *JSM Proceedings, Section on Government Statistics*. American Statistical Association, Alexandria, VA. 1510-1527.

Chen, L., Nandram, B., Cruze, N.B. (2021a) Hierarchical Bayesian model with inequality constraints for US county estimates. (*Journal of Official Statistics* accepted.)

Chen, L., Cruze, N.B., Nandram, B. (2021b) Preserving acreage relationships in small area agricultural models. (In preparation.)

Cruze, N.B., Chen, L., Guindin, N. and Nandram, B. (2020). Dancing distributions: developing a better understanding of county-level crop yield from posterior summaries. In *JSM Proceedings, Section on Government Statistics*. American Statistical Association, Alexandria, VA. 2262-2272.

Cruze, N.B., Erciulescu, A.L., Nandram, B., Barboza, W.J., Young, L.J. (2019) Producing official county-Level agricultural estimates in the United States: needs and challenges. *Statistical Science*. 34(2), 301-316. https://doi.org/10.1214/18-STS687

Erciulescu, A.L., Cruze, N.B., Nandram, B. (2018). Benchmarking a triplet of official statistics. *Environmental and Ecological Statistics*, 25(4), 523-547. https://doi.org/10.1007/s10651-018-0416-4

Erciulescu, A.L., Cruze, N.B., Nandram, B. (2019). Model-based county-level crop estimates incorporating auxiliary sources of information. *Journal of the Royal Statistical Society, Series A*, 182, 283-303. https://doi.org/10.1111/rssa.12390

Erciulescu, A.L., Cruze, N.B., Nandram, B. (2020) Statistical challenges in combining survey and auxiliary data to produce official estimates. *Journal of Official Statistics*. 36(1), 63-88. https://doi.org/10.2478/jos-2020-0004

Nandram, B., Cruze, N.B., Erciulescu, A.L. (2021) Bayesian benchmarking under inequality constraints and double shrinkage. (In preparation.)

Nandram, B., Erciulescu, A.L. and Cruze, N.B. (2019). Bayesian benchmarking of the Fay-Herriot model using random deletion. *Survey Methodology*, Statistics Canada, Catalogue No. 12-001-X, Vol. 45, No. 2. https://www150.statcan.gc.ca/n1/pub/12-001-x/2019002/article/00004-eng.htm

National Academies of Sciences, Engineering, and Medicine (2017). *Improving Crop Estimates by Incorporating Multiple Data Sources*. National Academies Press. https://doi.org/10.17226/24892

Thank you!

Linda.J.Young@usda.gov