Crop Yield Calculations for Estimating Nutrient Application and Long-term Loads

UPDATE: 6/14/24

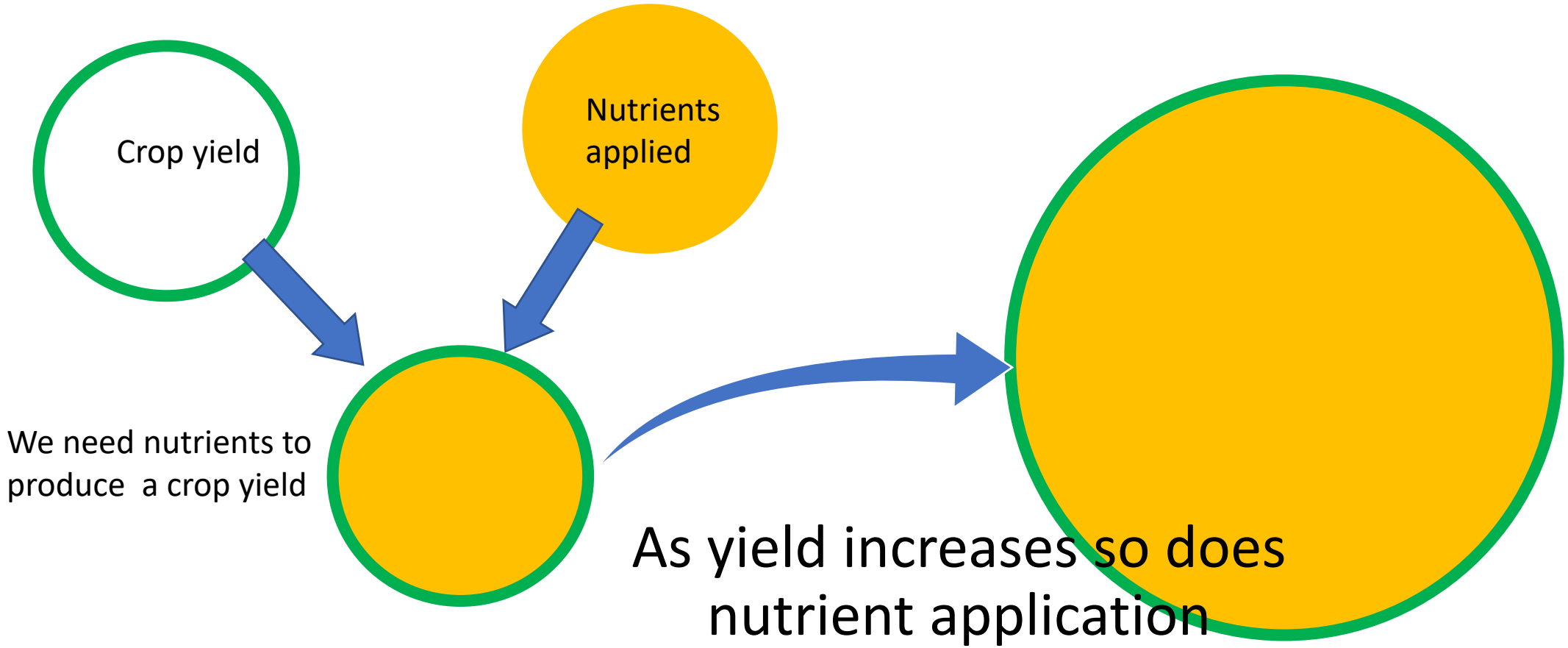
Joseph Delesantro

ORISE Fellow, CBPO Modeling Team

Photo: WILL PARSON/CHESAPEAKE BAY PROGRAM

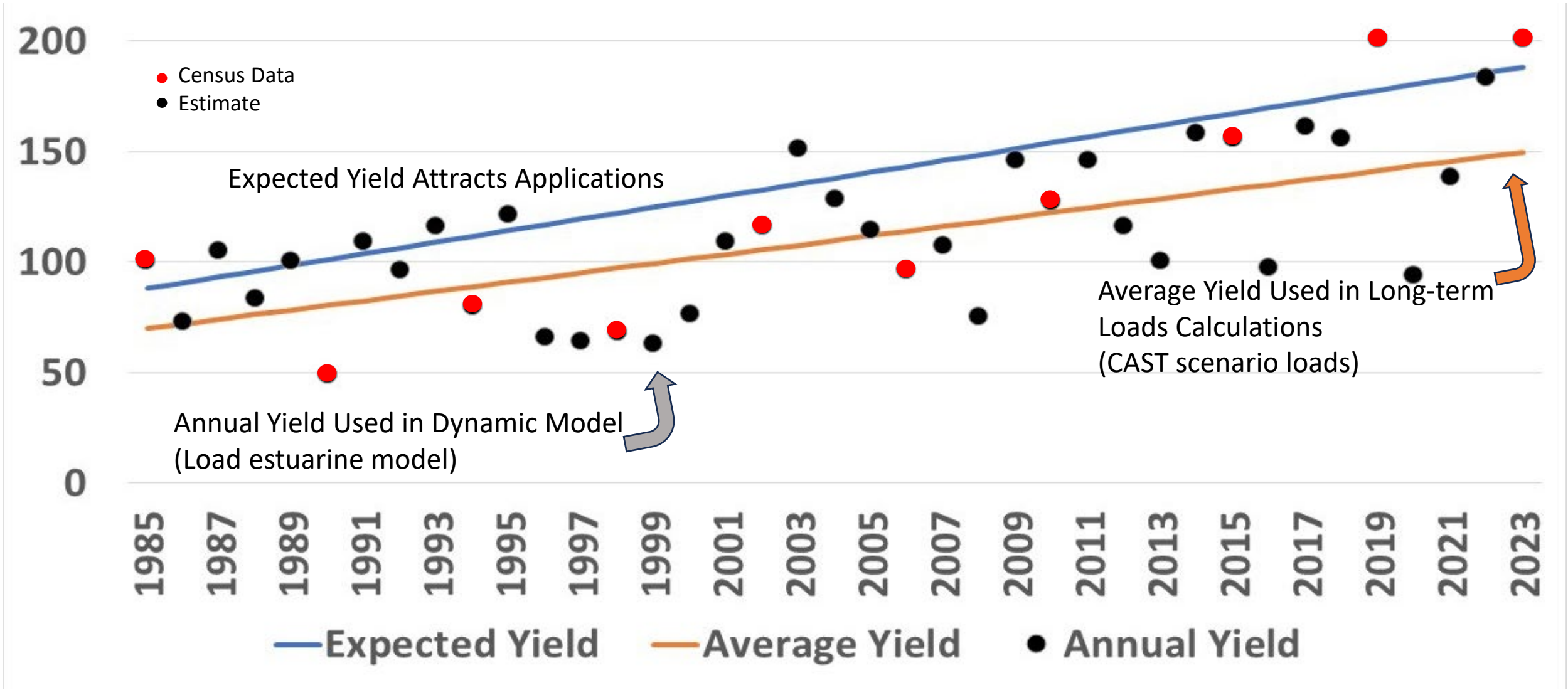
Why crop yields matter

- Yields and nutrient applications are tied together



*EXAMPLE
DATA ONLY

$$N\ applied_{(crop\ i)} = Acres_{(crop\ i)} * \underline{Expected\ Yield_{(crop\ i)}} * lbs\ N/unit\ yield_{(crop\ i)}$$



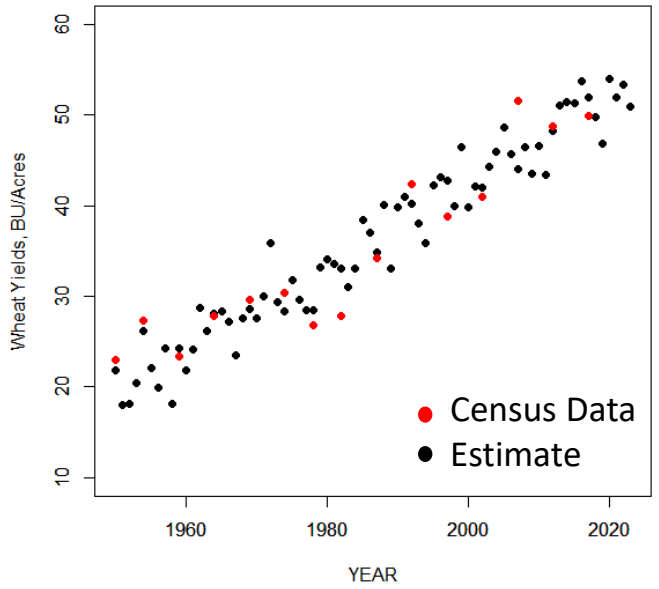
Path of investigation

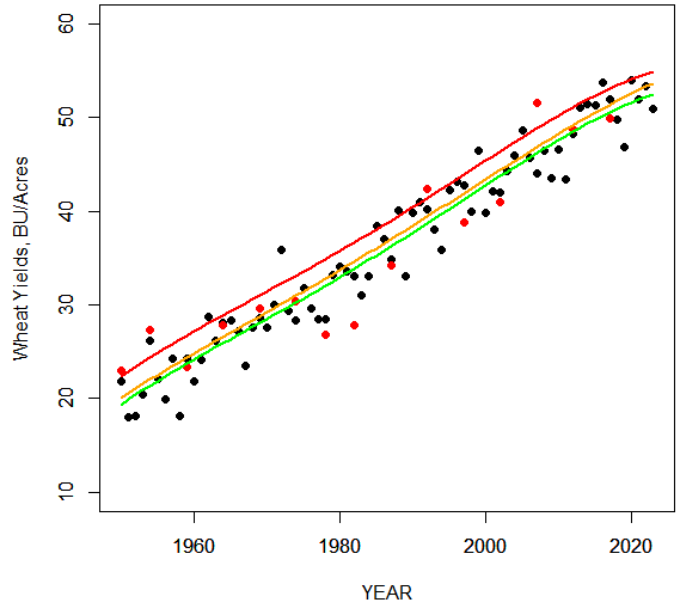
Goals:

- Estimate farmer yield expectations at the county level which drive the application of nutrients.
- Estimate various yield trends to support potential scenarios.

Collect crop yield data from 1950 to present → Estimate annual yields → Apply trend analyses

USDA Census and Survey data





Crop data collection

- 94 CAST-crops with both a potential yield and N-application
  - Excludes pasture, fallow, unmanaged or wild covers

- "Complete" data for 23 of these CAST-crops
  - Complete = data spanning >85% of period 1950-2022
  - **91% of crop land area**
  - **95% of N applied to crop land**
  - **89% of P applied to crop land**

- Partial data for an additional 40 crops
  - Partial = partial spatial range, partial time range, state-level only
  - 2.2% of crop land area, 3% of N applied to crop land

- No yield data for 31 crops
  - 6% of crop land area, 2% of N applied to crop land

corn for grain
soybeans for beans
other haylage; grass silage and greenchop
corn for silage or greenchop
alfalfa hay
wheat for grain
haylage or greenchop from alfalfa or alfalfa mixtures
wild hay
small grain hay
barley for grain
oats for grain
cotton
rye for grain
sorghum for grain
potatoes
sorghum for silage or greenchop
tobacco
buckwheat
sunflower seed - oil varieties
popcorn
other managed hay
peanuts for nuts
sunflower seed - non-oil varieties

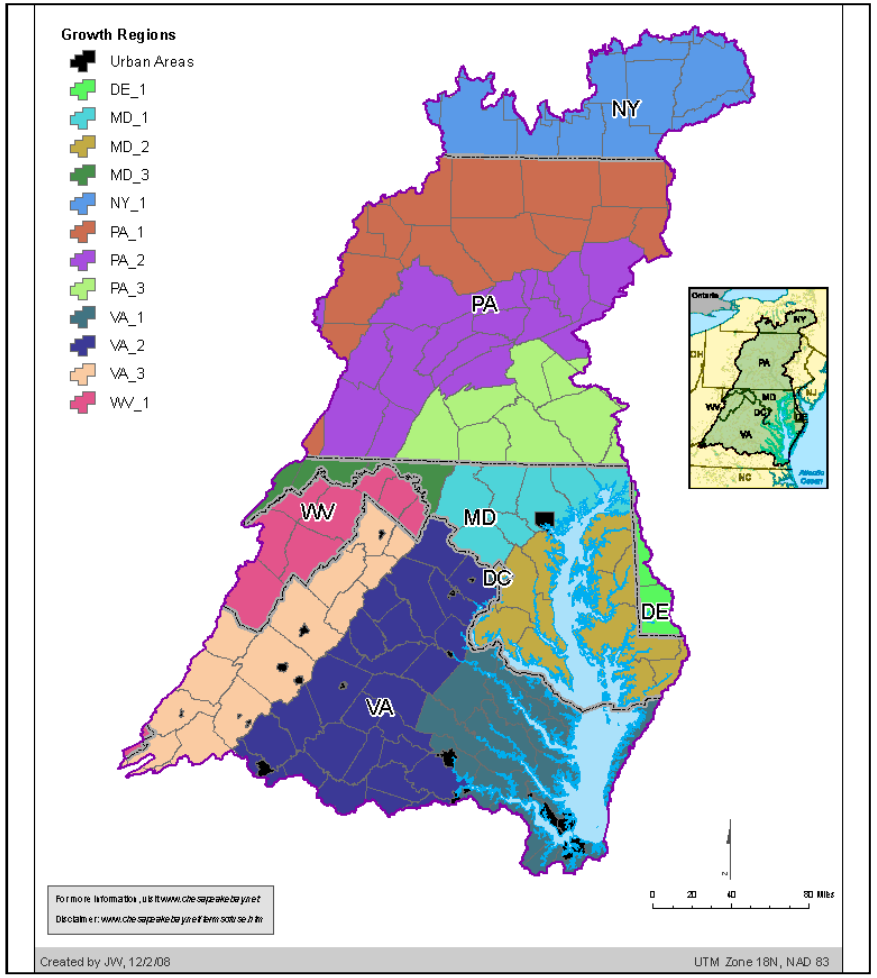
Statistical modeling method for estimating annual yields

multivariate linear models, bootstrapped (LOO) BIC and conceptual model selection

$Yield_{\text{crop i, growth region j}}$ ~ f(time, weather, climate,
Survey crop yields,
Survey economics)

*where appropriate and complete

Aggregate to growth regions for more consistent yield data



Statistical modeling method for estimating annual yields

multivariate linear models, bootstrapped (LOO) BIC and conceptual model selection

$$Yield_{\text{crop i, growth region j}} \sim f(\text{time, weather, climate, Survey crop yields, economics})$$

Weather and climate:

- Precip. – growing season
- Precip. – winter
- Wet day frequency – growing season
- Avg temp. - growing season
- Avg temp. – annual
- Growing degree day
- Heat stress – 5 consecutive days with max temp. > 86F
- Drought – 40 days Apr.-Jul. with < 2 in rainfall

Survey crop yields:

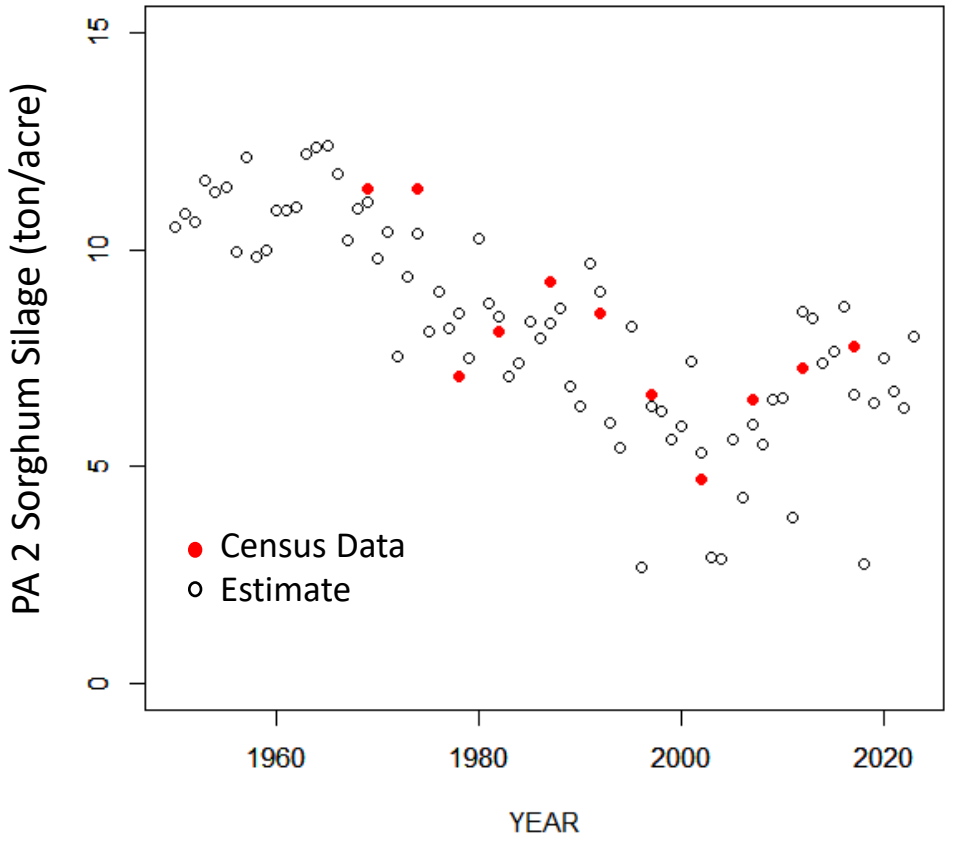
- Corn-grain
- Corn-silage
- Oats
- Wheat
- Soy

Survey price data:

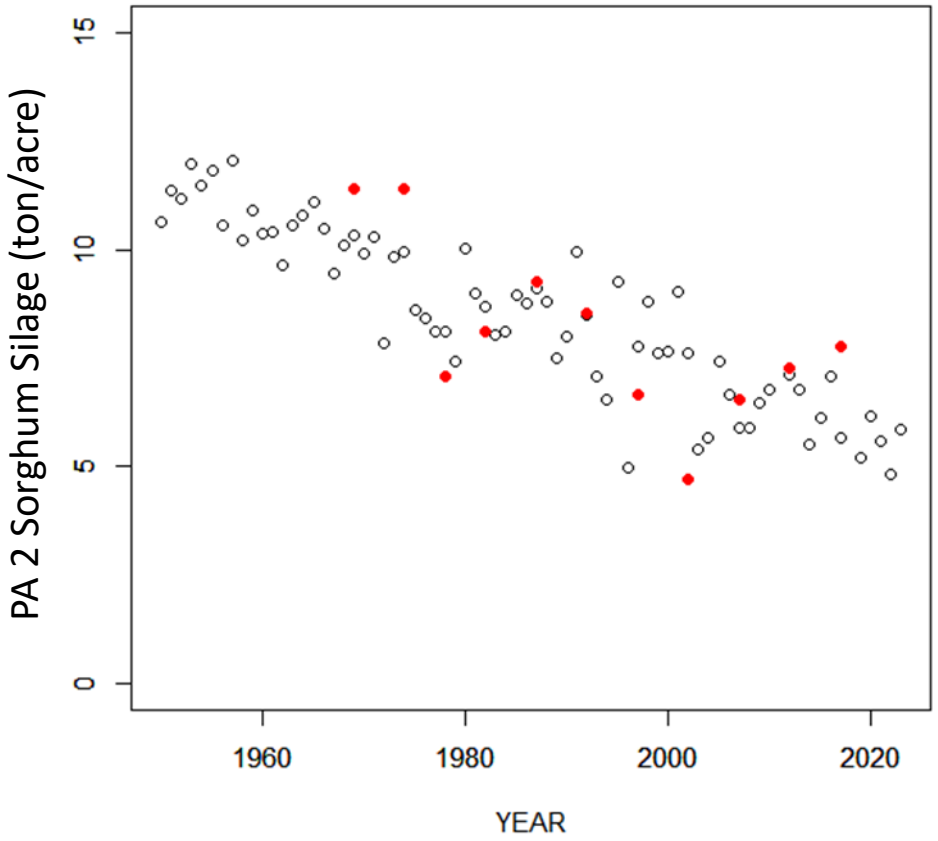
- Corn
- Sorghum
- Hay
- Wheat
- Oil/gas

Fitting prioritizes consistency

Least squares fit



Outlier resistant fitting
Limit predictor values to +/- 15% of observed range

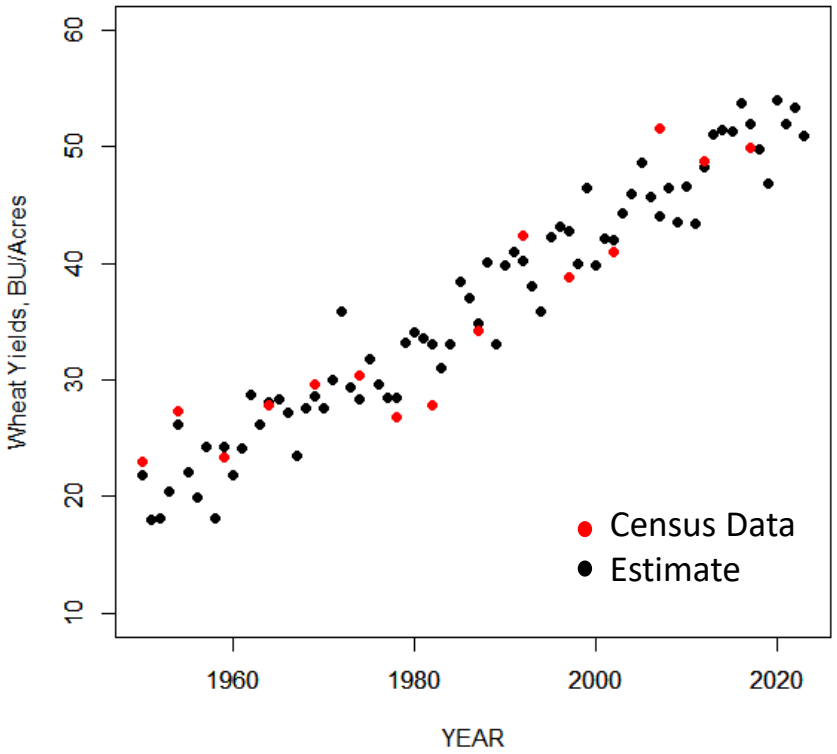


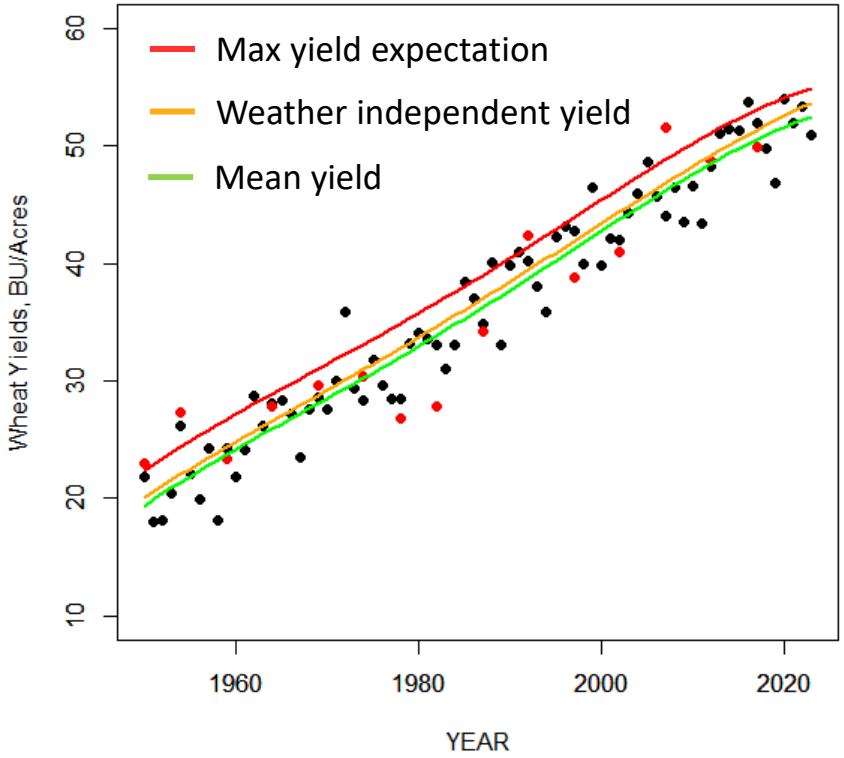
Statistical modeling method for estimating annual yields

multivariate linear models, bootstrapped (LOO) BIC and conceptual model selection

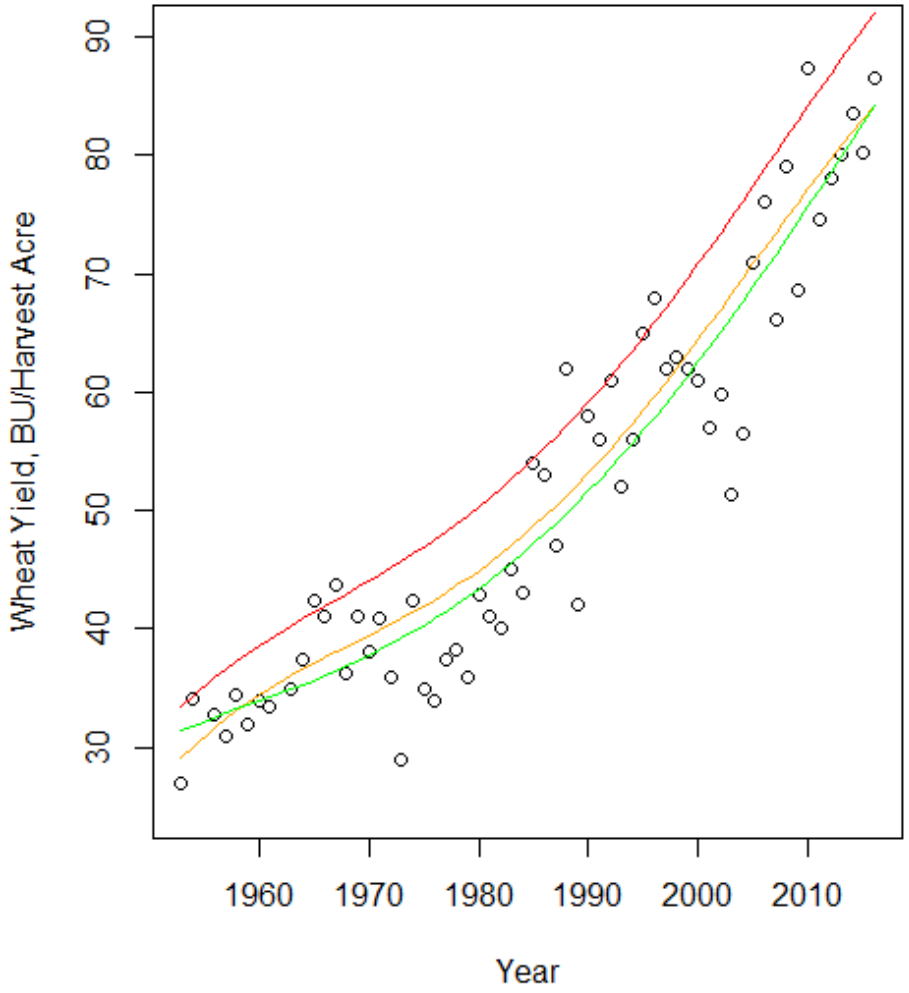
$$Yield_{\text{crop i, growth region j}} \sim f(\text{time, weather, climate, Survey crop yields, economics})$$

$R^2 \sim 0.74$
Crop area weighted





Trend analysis



- Regression, weighted residuals
- 1/3 smoothed moving max and averages
- Weather normalized estimate

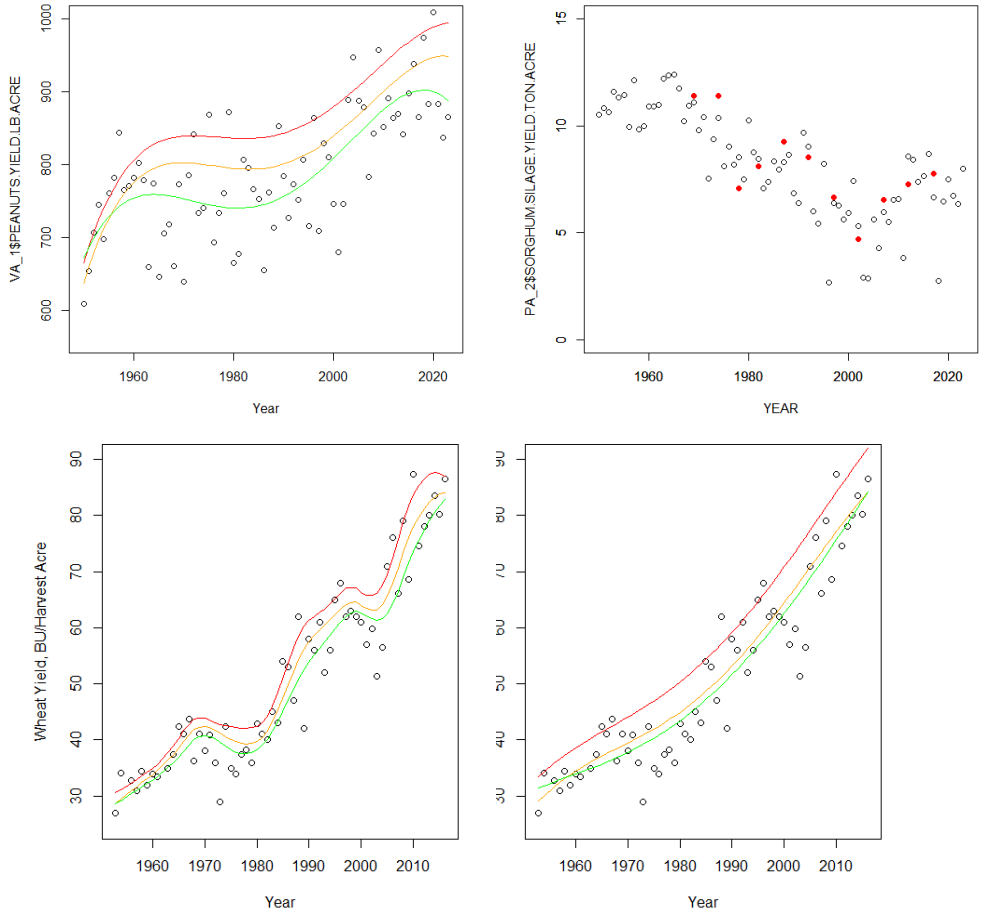
Yield expectation attracting application

Weather independent yield

Mean yield

Suite of metrics to describe the yield estimates and trend analyses

- Generating thousands of these plots
- We need a way to iterate without visually inspecting all of the generated data
- A suite of metrics to quickly assess the changes to numerical methods and flag issue to bring to the working group
  - “Smoothness”
  - Change metrics
  - Fit



Assessment of the method

- Does not provide good results for 0.9% of N applied, 1.0% of P applied (as estimated by CAST ‘23 2016-2020)
- Applied to 89% of P application, 95% of N application and performs well across 99% of that application