Importance of Data Quality for Data Aggregation

Elizabeth M. Hawkins
The Ohio State University
July 24, 2019
St. Louis, Missouri
Growing Amount of Data in Ag

http://websoilsurvey.sc.egov.usda.gov/
http://www.ent.iastate.edu/pest/cornborer/tasseling-corn-hybrids
https://news.illinois.edu/blog/view/6367/204576
Data is a tool. When used correctly, data can provide insights that improve farm management.

Using data to improve decision-making can help optimize farm management whether that be through:

• Reduced risks
• Increased profits
• Reduced inputs
Turning Data into Decisions

Data utilization is the process of creating usable information or knowledge from your data.
Value of Sharing and Aggregating Data

• More robust datasets for analysis and mining
Potential for Better Answers

Optimum Management Practices?
Digital Tools

Many different tools are available to make using your data easier.

- Data Warehousing
- In-Season Monitoring
- Production Benchmarking
- Crop Modeling
- Production Analysis
- Recommendations
eFields Program

- Standardized research protocols and data collection

2018
- 95+ trials
- 25 counties
- 5,600+ acres

2017
- 39 trials
- 13 counties
- 3,000+ acres
2018 Soybean Seeding Rate Results

Target Seeding Rate (1000 seeds/acre)

Relative Yield

- Crawford Co.
- Darke Co. A
- Darke Co. B
- Defiance Co.
- Fayette Co. A
- Fayette Co. B
- Fayette Co. C
- Fulton Co.
- Hardin Co.
- Knox Co. A
- Knox Co. B
- Miami Co.
- Pickaway Co.
- Ross Co.
- Sandusky Co. A
- Sandusky Co. B
- Tuscarawas Co.
- Union Co. A
- Union Co. B
- Williams Co.
Good Decisions Require High Quality Data

- Quality data is important to avoid making the wrong conclusions
- Yield data is especially prone to problems that affect quality
Collecting High Quality Data takes Work
Collecting High Quality Data
Data Quality Impacts

Corn Yield (bu/ac)

Accuracy – Calibrating Yield Monitors

12 lb/sec
How Does a Yield Monitor Work?

- 12 lb/sec
- Test Weight
- Distance
- BU/AC
Why is Calibration Important?

A

150 bu/ac

B

180 bu/ac
Why **Proper** Calibration is Important

Source: http://msue.anr.msu.edu/news/yield_monitor_calibration_procedure
Check the Accuracy of Your Yield Monitor

- Accuracy of your yield monitor changes as harvest conditions change
  - Moisture
  - Test weight
  - Temperature
  - Adjustments made to YM components
- Check your calibration regularly by weighing a load and comparing to the yield monitor estimate
  - Recalibrate if the difference is greater than 5%
## Check the Accuracy of Your Yield Monitor

<table>
<thead>
<tr>
<th>Year</th>
<th>Scale Weight (lbs)</th>
<th>Yield Monitor Weight (lbs)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>156,510</td>
<td>158,363</td>
<td>1.2</td>
</tr>
<tr>
<td>2009</td>
<td>155,690</td>
<td>169,569</td>
<td>8.9</td>
</tr>
<tr>
<td>2011</td>
<td>116,690</td>
<td>110,757</td>
<td>-5.1</td>
</tr>
<tr>
<td>2013</td>
<td>117,740</td>
<td>116,976</td>
<td>-0.7</td>
</tr>
<tr>
<td>2015</td>
<td>126,480</td>
<td>124,247</td>
<td>-1.8</td>
</tr>
</tbody>
</table>
Precision – Spatial Integrity
Precision – Flow Delay

0 second flow delay setting
Precision – Flow Delay

0 second flow delay setting  12 second flow delay setting
Data Cleaning – Yield Monitor Errors

https://www.youtube.com/watch?v=CM_PIYJc9o
Remove Errors and Artifacts in Data

- Areas with sudden speed changes
- Swath width errors
- Header height errors

Picture Source: Joe Luck, UNL, 2015
Using Historical Yield Data

- Data cleaning is critical to remove errors from areas that may impact decisions
- Multiple years of data will help mask errors over time
Data Cleaning – Persistent Errors

• Combine operation causes errors that will occur in the same areas over time
  • Headlands
• Accurate field boundaries
• Scout and document underlying yield causes
Other Issues

- Post-processing
  - Double check settings
    - Which crop setting was used?
  - Moisture
  - Weight to volume
- Lots of opportunities for errors!
Missing and Incorrect Information
Making the Correct Decisions

Correct interpretation will require more information than yield.
What Information Do We Need?

- Management details
  - Hybrid/variety
  - Population
  - In-season notes
    - Disease/Pest issues
    - Applications
- Record weather summaries
  - GDDs
  - Rainfall

http://www.ent.iastate.edu/pest/cornborer/tasseling-corn-hybrids
https://feww.wordpress.com/category/us-weather-map/
Standardized Protocols

- Randomized complete block design
- Minimum 3 reps, 4 recommended
- 3-5 treatments
- Current protocols are available at go.osu.edu/eFields
Mobile Apps

- OSU PLOTS
  - iOS App Store and Google Play
- FREE!
- Easy to use
- Share information from app
  - CSV
  - JPEG of Results Summary page
eFields Data Collection

• 22 additional variables that describe the operation and management
Seeding Rate Trials

STUDY INFORMATION

- Planting Date: 5/8/2018
- Harvest Date: 9/19/2018
- Variety: SC9238R
- Population: Treatments
- Acres: 19
- Treatments: 4
- Reps: 3
- Treatment Width: 40 ft.
- Tillage: Vertical
- Herbicide: 2,4-D, Gramoxone, Prefix, Select, Durango
- Previous Crop: Corn
- Row Width: 15 in.
- Soil Type: Ockley silt loam, 66% Crane silt loam, 17% Fox Gravelly loam, 14% Wooster silt loam, 3%

Weather Summary

- Total Total
- APR 6.49 3.93 3.38 Cumulative GDDs 133 723 1,363
- MAY
- JUN
- JUL
- AUG
- Total 18.05

PROJECT CONTACT

For inquiries about this project, contact John Barker, Extension Educator Agriculture (barker44@osu.edu).

Treatments (acs/acre)

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Avg. Emergence (plants/acre)</th>
<th>Moisture (%)</th>
<th>Yield (bu/acre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125,000</td>
<td>107,867</td>
<td>13.4</td>
<td>79 a</td>
</tr>
<tr>
<td>145,000</td>
<td>115,697</td>
<td>13.5</td>
<td>78 a</td>
</tr>
<tr>
<td>155,000</td>
<td>124,000</td>
<td>13.6</td>
<td>78 b</td>
</tr>
<tr>
<td>185,000</td>
<td>126,167</td>
<td>13.4</td>
<td>79 a</td>
</tr>
</tbody>
</table>

Treatment Means with the same letter are not significantly different according to Fisher's Protected Least Significant Difference (LSD) test at alpha = 0.1.

LSD: Not significant
CV: 2.25%

Soil Type

- Ockley silt loam, 66%
- Crane silt loam, 17%
- Fox Gravelly loam, 14%
- Wooster silt loam, 3%

Soybean

eFields Collaborating Farm
OSU Extension
Knox County - B

148 | Ohio State Digital Ag Program

2018 eFields Report | 147
Corn Yield Response to Seeding Rate
Corn Yield Response to Planting Date
# Corn Yield Response to Planting Date

## Weather Summary

<table>
<thead>
<tr>
<th></th>
<th>APR</th>
<th>MAY</th>
<th>JUN</th>
<th>JUL</th>
<th>AUG</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precip (in)</td>
<td>3.45</td>
<td>8.08</td>
<td>3.49</td>
<td>6.91</td>
<td>1.46</td>
<td>23.39</td>
</tr>
<tr>
<td>Cumulative GDDs</td>
<td>265.0</td>
<td>595.0</td>
<td>1237.0</td>
<td>1978.0</td>
<td>2590.0</td>
<td>2590.0</td>
</tr>
</tbody>
</table>

---

*The Ohio State University*
Seeding Rate Response by Soil Type

![Graph showing the response of yield to seeding rate for Canfield and Fitchville soil types. The x-axis represents the target seeding rate in thousand seeds/acre, ranging from 22 to 38. The y-axis represents the yield in bushels per acre, ranging from 170 to 230. Different symbols indicate the two soil types, with Canfield represented by grey circles and Fitchville by red squares.]
Challenge: Capturing Accurate Information

• Most data is collected during critical times of the season
  • Planting
  • Sidedress
  • Harvest
• Mistakes will happen
As-Applied Data for Treatment Verification

- Confirm treatments executed as planned
- Great record of applications
Imagery to Document Field Variability

- Targeted scouting
  - Plant health and vigor
  - Weed and pest

- Documentation of field conditions spatially to help explain yield variability
Mobile Apps

- Useful for capturing in-season records
- Many allow pictures to be added and geotagged
- 2 OSU Fact Sheets that list and describe apps created for scouting
  - [go.osu.edu/CropNutritionApps](go.osu.edu/CropNutritionApps)
  - [go.osu.edu/CropProtectionApps](go.osu.edu/CropProtectionApps)
## Digital Data Capture Success

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Number of Locations Submitted</th>
<th>Success Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot Yield Averages</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>Spatial Yield Data</td>
<td>26</td>
<td><strong>50</strong></td>
</tr>
<tr>
<td>Stand Counts</td>
<td>43</td>
<td><strong>83</strong></td>
</tr>
<tr>
<td>Planting and Harvest Dates</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>Hybrid/ Variety</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>Soil Type</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>Field Management History</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>Pesticide Applications</td>
<td>36</td>
<td><strong>69</strong></td>
</tr>
</tbody>
</table>
Collecting High Quality Data takes Work
How Do We Encourage High Quality Data Collection?

- Increased education and outreach about data in agriculture
  - Importance of data quality: calibration, flow shift and data processing
- Enable farmers to use their data and create value

CFAES
Final Thoughts

• Vast potential but big challenges still exist
• Yield data quality, overall, is likely not as good as we believe
• Interpreting aggregated yield data accurately requires information that we currently are not typically capturing
Questions?

Elizabeth Hawkins
hawkins.301@osu.edu
@Farm3rsDaught3r