Using Data to Drive Crop Production Decisions

Shawn P. Conley, Ph.D.
Soybean and Wheat Extension Specialist
University of Wisconsin, Madison
Overview

• In this session we will explore three unique methods that utilize data to drive crop production decisions behind the farm gate.

• These examples will include:
  – probability risk matrix using small plot research data
  – developing prescription variable seeding rates in soybean using multiple on-farm replicated trials
  – utilizing yield gap analysis from farmer supplied big data sets to drive grower decision making
Response of Broad Spectrum and Target Specific Seed Treatments and Seeding Rate on Soybean Seed Yield, Profitability, and Economic Risk

Unpredictable Field Variability

• Traditional IPM practices and seed treatment decisions are often at odds
  – Field-scale insect and disease levels are largely unknown and unquantifiable
    before planting

• Growing conditions vary within a year due to weather, soil type, and
  topography resulting in varying and often unpredictable levels of disease, 
insect pressure and planting date
Trial Information

• RCBD 2 x 3 x 6 factorial or split plot with 4 reps
• Varieties used: AG2136, RS213NR2, AG2636
• Six seeding rates (seeds a\(^{-1}\))
  – 40,000
  – 60,000
  – 80,000
  – 100,000
  – 120,000
  – 140,000 (current WI rec.)
• Three seed treatments
  – UTC
  – CB: Evergol Energy, Allegiance, Poncho/Votivo
    – Prothioconazole (F)
    – Penflufen (F)
    – Metalaxyl (F)
    – Clothianidin (I)
    – Bacillus firmus (N)
  – ILeVO
    – CB (FIN)
    – Fluopyram (F)
Yield Potential: Locations

- We examined treatments across various yield potentials and ultimately, responsive and non-responsive environments
  - WI = 20 environments
  - IA = 2 environments
  - IN = 1 environment
  - MI = 1 environment
  - ON = 2 Environments


= Location displayed visual SDS symptomology
Economic Risk

• Uncontrollable factors during planting and early season growth

• Products and practices that are valuable:
  – Show consistent yield gains
  – Provide profit stability over a wide range of situations and environments
  – Help manage long term margins and economic risk with volatile grain markets

• Assessing economic risk at various seeding rates and how seed treatment affects risk
  – “Base case” = 140k seeds a\(^{-1}\) with no seed treatment (UTC)
  – Our trial allows us 20 comparisons to the base case.
  – The break-even probability shows us the probability that a certain seeding rate x seed trt. combination will increase profit over the base case.
    • Or essentially the risk of a certain treatment combination
## Economic Risk Table for $8\text{ bu}^{-1}$ Soybeans

<table>
<thead>
<tr>
<th>Treatment combination</th>
<th>Seed Treatment</th>
<th>Seeding Rate</th>
<th>Break-even probability</th>
<th>Avg. profit increase over the Base Case</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Seeds $a^{-1}$</td>
<td></td>
<td>$a^{-1}$</td>
<td></td>
</tr>
<tr>
<td>UTC</td>
<td>$120,000$</td>
<td>0.99</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>$100,000$</td>
<td>0.97</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$80,000$</td>
<td>0.04</td>
<td>1</td>
<td>-5</td>
</tr>
<tr>
<td></td>
<td>$60,000$</td>
<td>0.00</td>
<td>na</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td>$40,000$</td>
<td>0.00</td>
<td>na</td>
<td>-96</td>
</tr>
<tr>
<td>CB</td>
<td>$140,000$</td>
<td>0.70</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>$120,000$</td>
<td>0.93</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>$100,000$</td>
<td>0.97</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>$80,000$</td>
<td>0.76</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>$60,000$</td>
<td>0.01</td>
<td>2</td>
<td>-20</td>
</tr>
<tr>
<td></td>
<td>$40,000$</td>
<td>0.00</td>
<td>na</td>
<td>-83</td>
</tr>
<tr>
<td>ILeVO</td>
<td>$140,000$</td>
<td>0.87</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>$120,000$</td>
<td>0.98</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>$100,000$</td>
<td>0.99</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>$80,000$</td>
<td>0.95</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>$60,000$</td>
<td>0.01</td>
<td>3</td>
<td>-17</td>
</tr>
<tr>
<td></td>
<td>$40,000$</td>
<td>0.00</td>
<td>na</td>
<td>-83</td>
</tr>
<tr>
<td>EOSR</td>
<td>$110,300$</td>
<td>0.99</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>$103,700$</td>
<td>0.97</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>$103,200$</td>
<td>0.99</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>
Acknowledgements
Key Predictors for Variable Rate Soybean Seeding Prescriptions at the Landscape Scale

Ethan Smidt, Jun Zhu, and Shawn Conley
University of Wisconsin - Madison
Reasoning

- Growers are collecting data at all times

- GPS and equipment advances have allowed for increased use of variable rate technology (VRT)

- Growers are unsure which data layer(s) to use when creating these prescriptions
Objective

• Find key predictor variable(s) influencing soybean yield so future prescriptions can be created more accurately

Materials and Methods

- Created prescriptions with high, medium, and low seeding rates
- Rates confirmed by as-planted data and stand counts
- Data layers also collected from soil samples, yield monitors, and soil surveys
Materials and Methods

• Soil Sampling:
  – Soil pH, phosphorus, potassium, organic matter

• Planting/Yield Monitors:
  – Elevation, seeding rate

• Soil Survey:
  – Available water supply (0-100 cm, 0-150 cm), depth upper soil horizon, slope, soil symbol
  – NCCPI (v2.0)
Materials and Methods (continued)

- Range of locations, conditions, and soils
- Multiple varieties
- 15, 20, and 30 in row widths

Field size range:
2013/14: 25-200 a

Seeding rates: 100,000-200,000 seeds/a
Materials and Methods (continued)

- Soybean yield data was “cleaned” to discard outliers and incorrect data points
- Elevation data was obtained from planter tractors and harvesters
Statistics

- Decision Tree Analysis
  - Robust
  - Gives yield estimates
  - Tendency to over-fit without user input
Example Decision Tree

Key:

- **Percent of total dataset**
- **Average Yield (bu/ac)**
- **Number of data points**
- **Separation of most important parameter to this specific node of data**
Statistics

- Random Forest Process (500 individual trees)
- Cross-validation
Random Forest Results (Pooled Data)

Most important predictors – 2013 Pooled Data

Most important predictors – 2014 Pooled Data

Node Purity (measure of importance)
Random Forest - Single Field Analyses

<table>
<thead>
<tr>
<th>2013 Variable Rankings</th>
<th>2014 Variable Rankings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Elevation (1.55)</td>
<td>1. Elevation (2.00)</td>
</tr>
<tr>
<td>2. Soil Organic Matter (3.18)</td>
<td>2. Soil pH (3.09)</td>
</tr>
<tr>
<td>3. Soil Potassium (3.36)</td>
<td>3. Soil Potassium (3.27)</td>
</tr>
<tr>
<td>5. Soil pH (4.09)</td>
<td>5. Soil Phosphorus (3.82)</td>
</tr>
</tbody>
</table>

• Very different story
• Local knowledge is still very important
• Next closest is Soil Symbol (>6.0) in 2013 and 2014 (>6.0)
Results (continued)
Statistics

• Quantile Regression
  – Analyze responses at different yield levels
  – Compare results to SLR
  – “quantreg” (Koenker, 2015)
Quantile Regression Results
Quantile Regression Results

2014 - Richland 2

Seeding Rate

0.00
0.05
0.10
0.15

0.2
0.4
0.6
0.8

Quantile


Acknowledgements

• Thank you to Bob Berkevich, Arnie Imholte, and all of DuPont Pioneer for organizational assistance and funding, as well as John Gaska with UW-Madison for software and mapping assistance.

• Thank you to Jun Zhu, Francisco Arriaga, and Shawn Conley for help throughout the process.
Benchmarking soybean production system in the North-central USA

**Core team:** Patricio Grassini, Shawn P. Conley, Juan I. Rattalino Edreira, Spyridon Mourtzinis and Adam Roth

**Regional collaborators:** Shaun Casteel, Ignacio A. Ciampitti, Mark Licht, Hans Kandel, Laura Lindsey, Darren Muller, Seth Naeve, Emerson Nafziger, Michael Staton

rattalino@unl.edu
Main goal

To benchmark current yield and management practices in producer fields across the North-Central US region in order to identify KEY management factors that can be used by individual producers to (i) increase soybean yield on their farms, and (ii) improve input-use efficiency.

We argue that having a database with well-contextualized farmers data might be equivalent to run hundreds field experiments.
Process Diagram

1 Collection
- Extension Educators
- Crop Production Clinics
- State growers boards
- Natural Res. Districts
- Crop consultants
- Newsletters
- Press releases

2 Assimilation
- Data entry
- Quality control
- Field location and mapping
- Data contextualization (soil & weather)

3 Analyzes
- Frontier analysis
- Machine learning
- Crop simulation models

4 Dissemination
- Scientific papers
- Extension newsletters
- Field days

User’s feedback

Strategic farm visits
1. Data collection

Farmer data from 10 states

Core team
- Patricio Grassini (Univ. of Nebraska)
- Shawn Conley (Univ. of Wisconsin)
- Juani Rattalino (Univ. of Nebraska)
- Adam Roth & Spyros Mourtzinis (Univ. of Wisconsin)

Regional collaborators
- Shaun Casteel (Purdue Univ.)
- Ignacio Ciampitti (Kansas State Univ.)
- Mark Licht (Iowa State Univ.)
- Hans Kandel (N. Dakota Univ.)
- Laura Lindsey (Ohio State Univ.)
- Daran Mueller (Iowa State Univ.)
- Seth Naeve (Univ. of Minnesota)
- Emerson Nafziger (Univ. of Illinois)
- Michael Staton (Michigan State Univ.)
1. Data collection

- ND
- MN
- NE
- IA
- KS
- MI
- OH
- WI
- IN
- IL

- 1700 per season
- 1500 per season
# 1. Data collection

<table>
<thead>
<tr>
<th>Field location</th>
<th>Field boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irrig. system</td>
<td>Irrig. amount</td>
</tr>
<tr>
<td>Drainage type</td>
<td>Average yield</td>
</tr>
<tr>
<td>Yield variation</td>
<td>Planting date</td>
</tr>
<tr>
<td>Variety</td>
<td>Seeding rate</td>
</tr>
<tr>
<td>Seed treat.</td>
<td>Tillage</td>
</tr>
<tr>
<td>Fertilizers</td>
<td>Lime, Manure</td>
</tr>
<tr>
<td>Starters</td>
<td>Herbicide</td>
</tr>
<tr>
<td>Fungicide</td>
<td>Insecticide</td>
</tr>
<tr>
<td>SCN, IDC, others</td>
<td></td>
</tr>
</tbody>
</table>

Survey form

Contact info and logos were customized for each state.
1. Data collection

Location of surveyed 2014-2016 soybean fields

501,838 acres
1. Data collection

- Data entry
  - >8000 fields by 2018
- Quality control
  - After applying our protocol <1% data is discarded
- Data contextualization
  - Each field has soil and weather data

2. Data assimilation

- Field delimitation
  - Field boundaries are used to retrieve soil data and other attributes

3. Data analysis

- Data entry
  - 1700 per season
- Data contextualization
  - 1500 per season
Data description

% of no-tilled fields

Proportion of tillage system
USDA-ERS (2010)
No-till/strip-till 46%

Comparison of current database versus USDA-NASS county yield data

Full report available in http://cropwatch.unl.edu/2016-soybean-survey
Materials and Methods cont.

Results - Yields

The chart shows the yield (Mg ha\(^{-1}\)) of different TED (Tillage and Equipment Design) treatments. The data is presented in box plots for each TED, with the median, interquartile range, and outliers marked. The yield values range from approximately 2 to 6 Mg ha\(^{-1}\).
3. Data analysis

Data were contextualized based on the Global Yield Gap Atlas spatial framework (www.yield.gap.org)

25% of US-soy area

- Delays in planting reduces yield potential up to 0.5 bu/A per day after late April.
- In this TED, soybean yield in can be increased, on average, 9 bu/A by planting earlier.
Statistical Analysis

- We used conditional inference trees to identify management and field variables influencing soybean yields in farmer fields within each TED.

- This method can handle categorical and continuous explanatory variables without statistical distribution assumptions, it is robust to outliers, multicollinearity, and heteroskedasticity, there is no bias and overfitting issues such as in regression trees, and can reveal variable interactions.

- The result of this procedure is a graph that looks like a tree.
Results TED=4R
Results TED=4R

TED=4R

SD
\[ p < 0.001 \]

\[ R^2 = 0.31 \]
\[ \text{RMSE} = 0.57 \text{ Mg ha}^{-1} \]

Foliar Fungicide
\[ p = 0.02 \]

Row space
\[ p = 0.007 \]

On-farm soybean yield (Mg ha\(^{-1}\))

\begin{align*}
\text{No} & : 4.1 \text{ Mg ha}^{-1} \\
\text{Yes} & : 4.4 \text{ Mg ha}^{-1} \\
\text{Narrow, Medium} & : 3.4 \text{ Mg ha}^{-1} \\
\text{Wide} & : 3.7 \text{ Mg ha}^{-1}
\end{align*}
Results TED=5R
Results TED=5R

TED=5R

R²=0.24
RMSE=0.54 Mg ha⁻¹

SD
p < 0.001

≤ 132
p = 0.013

pH (30-150)
≤ 6.5
n = 41
On-farm soybean yield
(Mg ha⁻¹)
4.3 Mg ha⁻¹

> 6.5
n = 23
3.9 Mg ha⁻¹

> 132
Foliar Fungicide
p = 0.013
No
n = 27
≤ 140
3.7 Mg ha⁻¹
Yes
n = 39
> 140
3.4 Mg ha⁻¹

n = 62
3.8 Mg ha⁻¹
Results TED=6R
Results TED=6R

TED=6R

SD $p < 0.001$

$R^2=0.25$
RMSE=0.49 Mg ha$^{-1}$

$\leq 141$

ST Insecticide $p = 0.01$

No

Yes

$> 141$

Nematodes $p = 0.024$

No

Yes, Unknown

Tillage $p = 0.034$

No-till, Reduced

Conventional

On-farm soybean yield (Mg ha$^{-1}$)

n = 19

4 Mg ha$^{-1}$

n = 17

4.6 Mg ha$^{-1}$

n = 40

4.2 Mg ha$^{-1}$

n = 30

4.4 Mg ha$^{-1}$

n = 47

3.7 Mg ha$^{-1}$
Results TED=8I
Results TED=8I

- TED=8I
  - SD p < 0.001
    - Foliar Insecticide p = 0.002
      - No
      - Yes
        - SD p = 0.009
          - ≤ 124
          - > 124
  - TWI p = 0.01
    - ≤ 10
    - > 10

On-farm soybean yield (Mg ha⁻¹)

- n = 50
  - 5 Mg ha⁻¹
- n = 18
  - 4.5 Mg ha⁻¹
- n = 38
  - 4.8 Mg ha⁻¹
- n = 22
  - 5.2 Mg ha⁻¹
- n = 24
  - 4.4 Mg ha⁻¹

R² = 0.26
RMSE = 0.44 Mg ha⁻¹
Conclusions

• It is important to highlight that the management practices that were deemed significant should not be generalized and interpreted as the only, or the most important management decisions for soybean production.

• These are the management practices that a large portion of farmers applied at a suboptimum level/rate and their optimization can increase soybean yields.
Conclusions *cont.*

- The analysis framework we used (conditional inference trees) is simple and effective in understanding and describing the effect of management practices (alone and interactive) on crop yield.

- Hence, analysis of farmer’s data should be considered as a complement to traditional randomized field experiments.

- Given the growing pressure for increasing food production, while minimizing the environmental footprint, analysis of farmer data using conditional inference trees is an effective approach to identify opportunities for increasing yield and input-use efficiency.
www.coolbean.info

@badgerbean

thesoyreport.blogspot.com