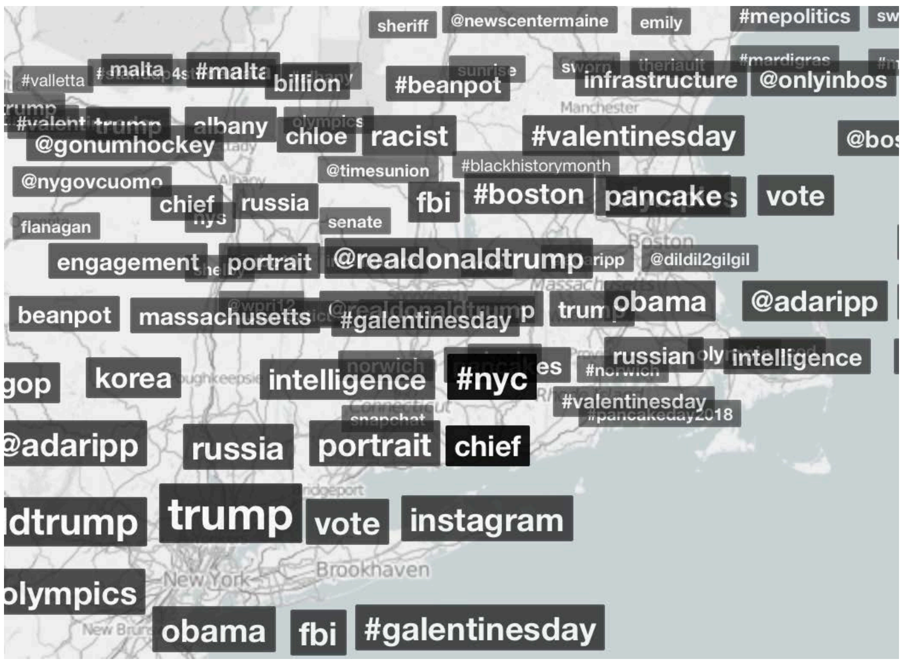
Using big data and machine learning to optimize cranberry production

Jed Colquhoun, University of Wisconsin-Madison

Sometimes, it's hard to put your finger on the problem…

- Yield variation among years
- Yield variation among beds
- Some marshes produce greater yields year after year than others
- Sometimes, we know the crop is "stressed" but we don't know why, and we don't know the real impact on yield and quality
- Some things we can control, others we can't – let's not worry about what we can't…

Big data and machine learning are nothing new



trendmaps.com

Big data:

- Unstructured data, such as trends on Twitter
- Structured data, such as medical record keeping

Machine learning:

- Email filtering
- Online medical diagnostics

In essence, there are two ways to tackle questions with research:

TRADITIONAL RESEARCH

- Hypothesize about a cause:
  - Harvest stress
  - Winter stress
  - Viruses
  - Herbicide damage
- Test the hypothesis
  - Subject cranberries to the stressor
  - Compare to cranberries without the stressor
- Lather, rinse and repeat until you're confident that the solution is "clean"

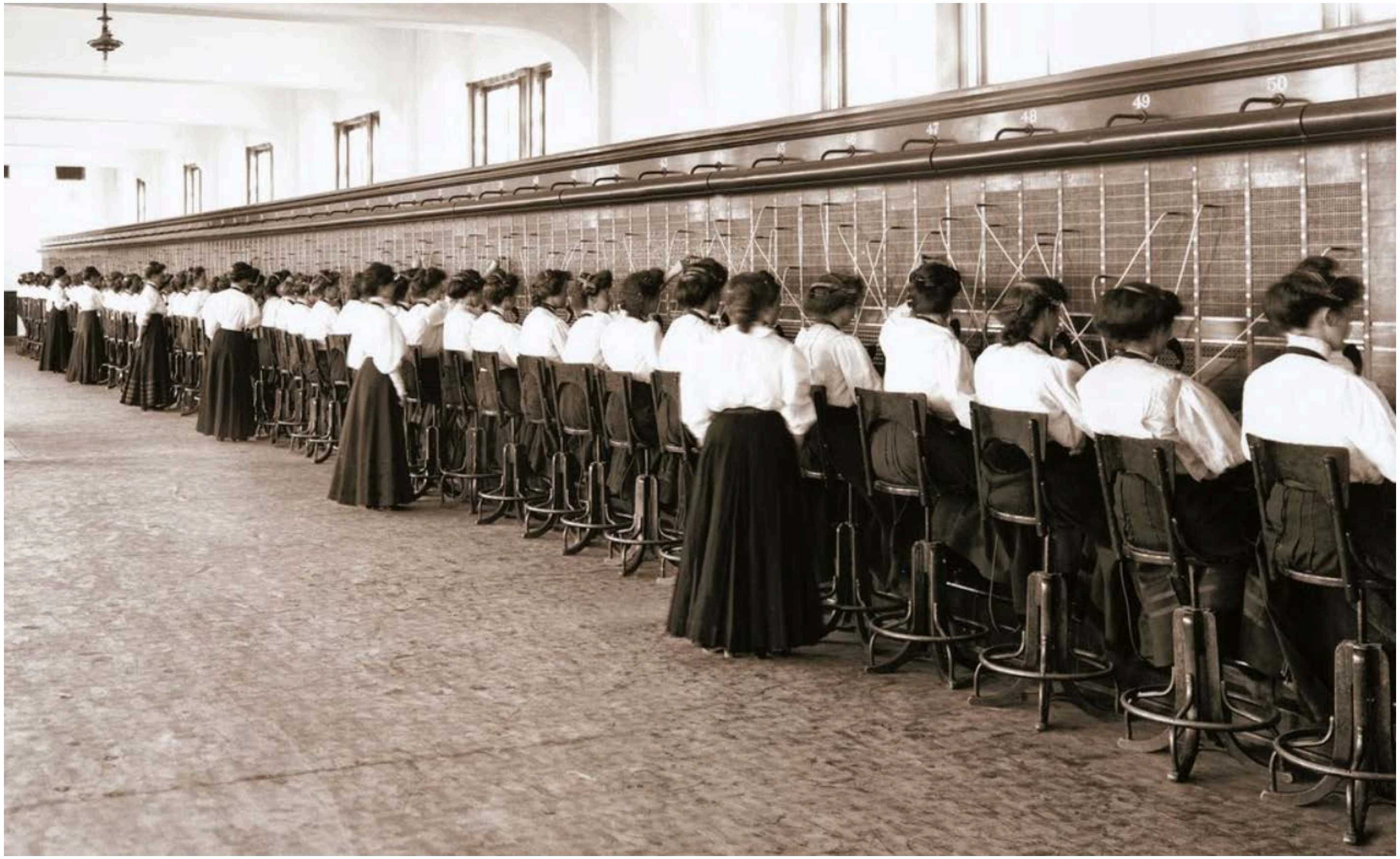
BIG DATA EXPLORATION

- Collect data on as many relevant variables as possible
- Collect as much data as possible
  - Inputs (pesticides, water, fertilizer, variety, stand age, etc.)
  - Weather
  - Outputs (yield and quality)
- Determine relationship among inputs, weather and outputs
- Add data over time to improve confidence in relationships

The more data we have, the more confident we become…

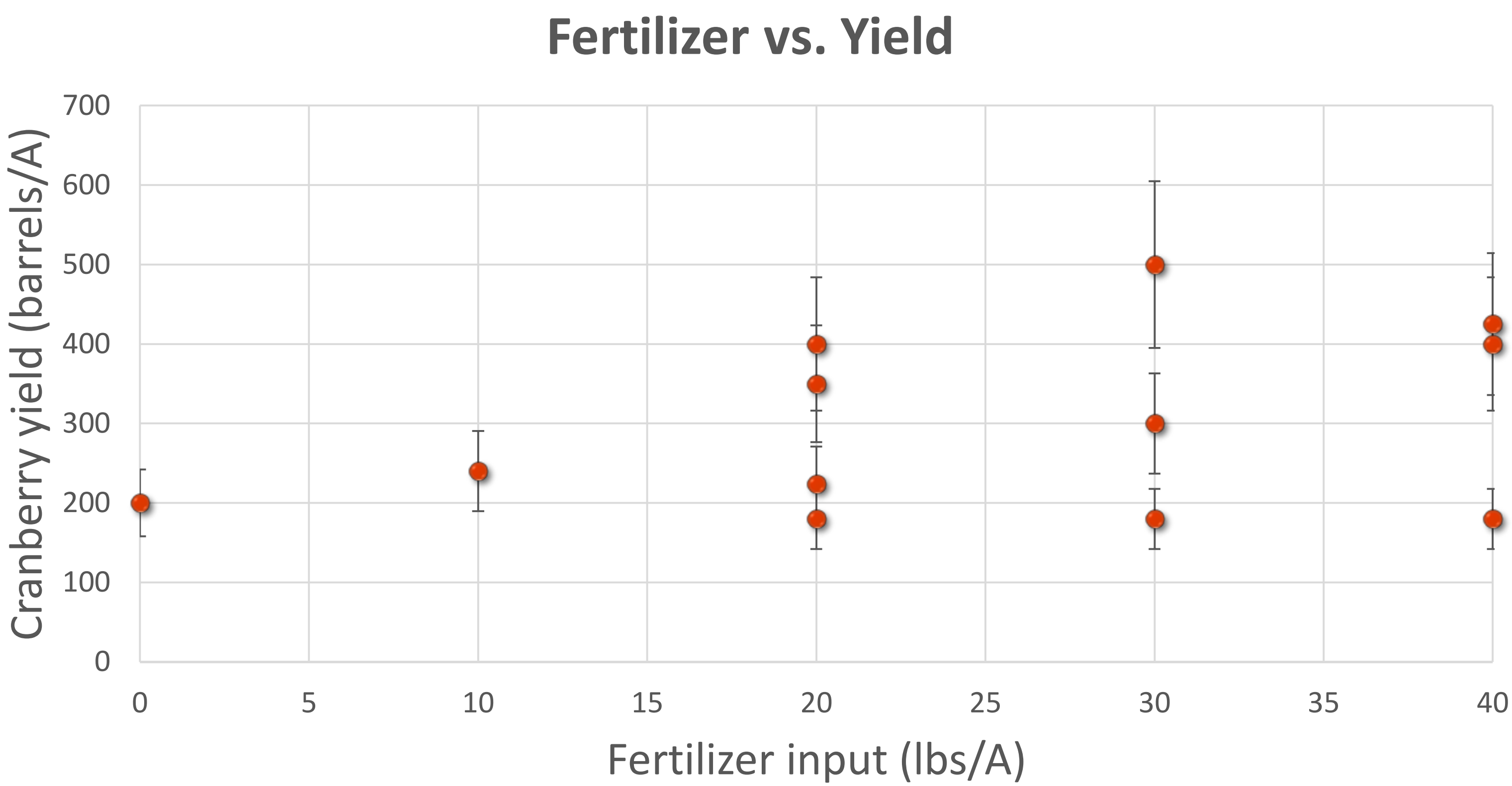
thenation.com

The more data we have, the more confident we become…

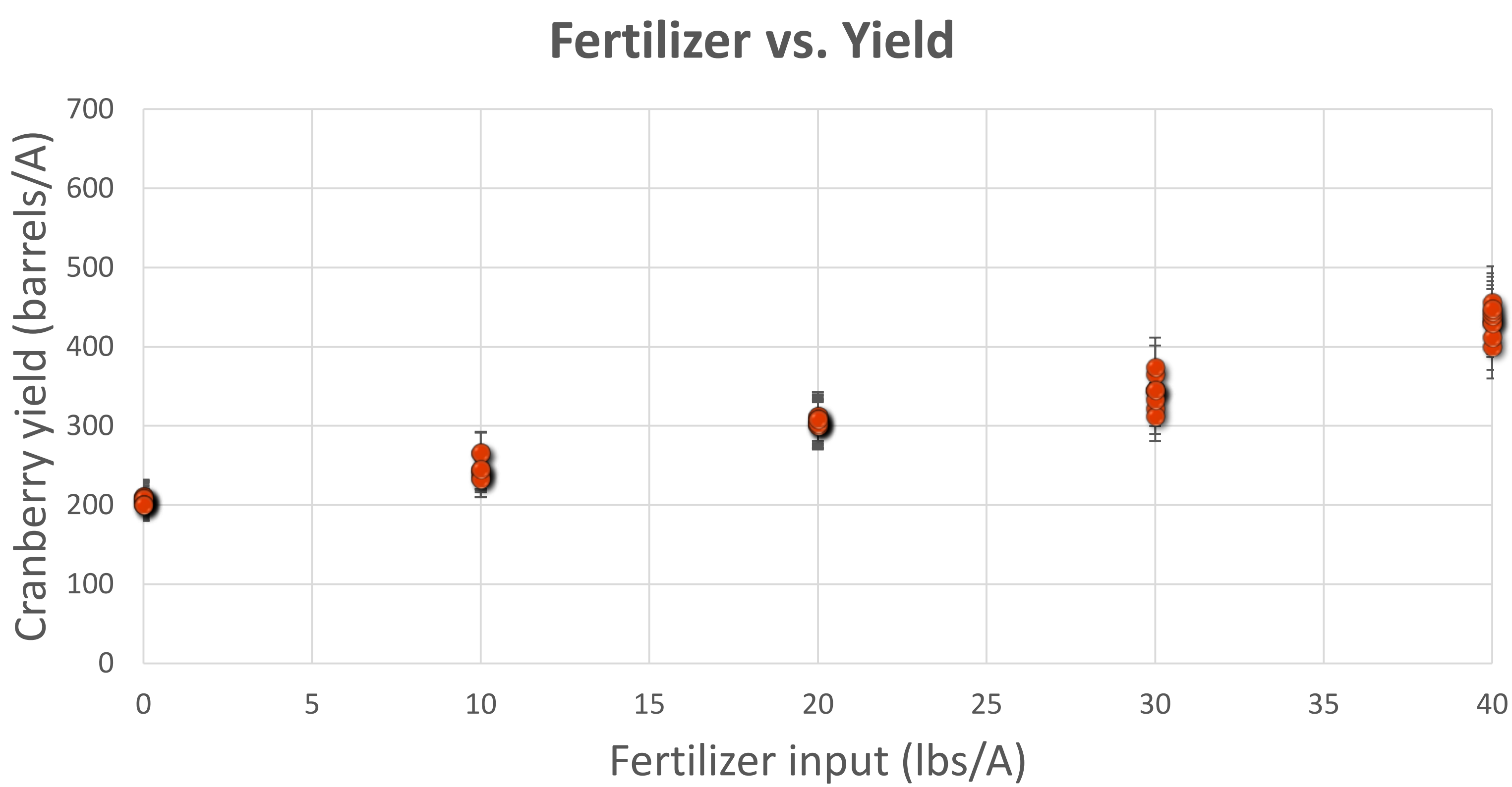


vintage.es

The more data we have, the more confident we become…

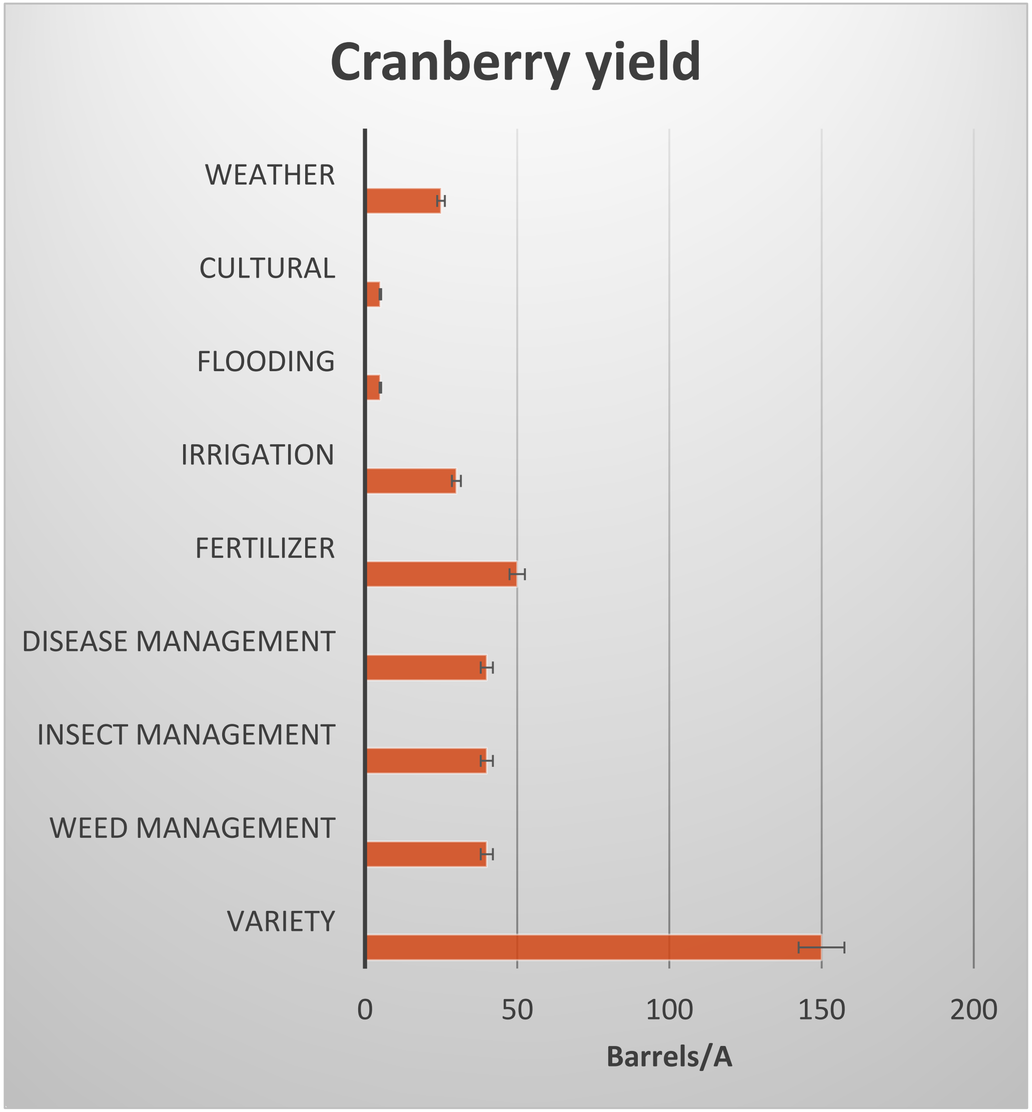


The more data we have, the more confident we become…



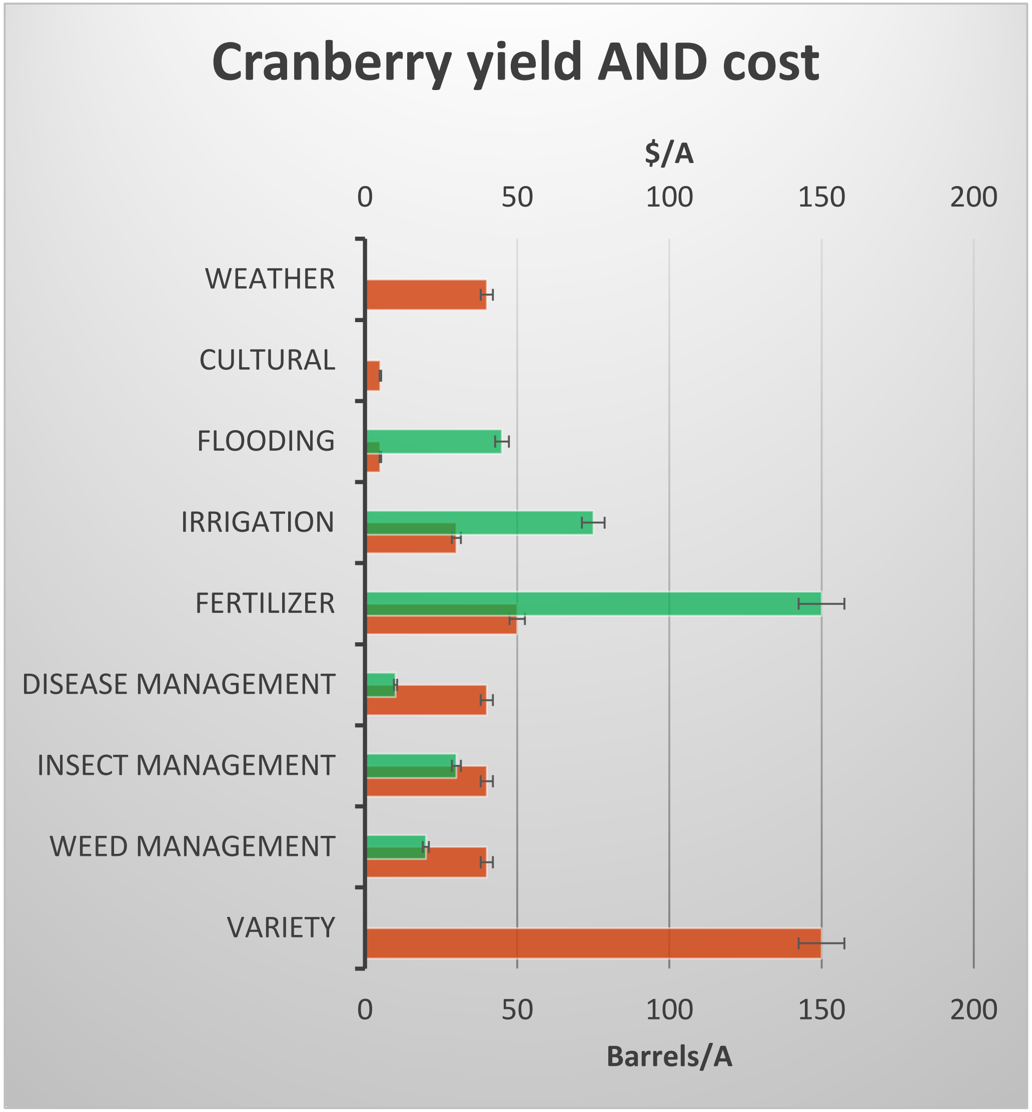
400 barrel/A projection

What contributes to yield?



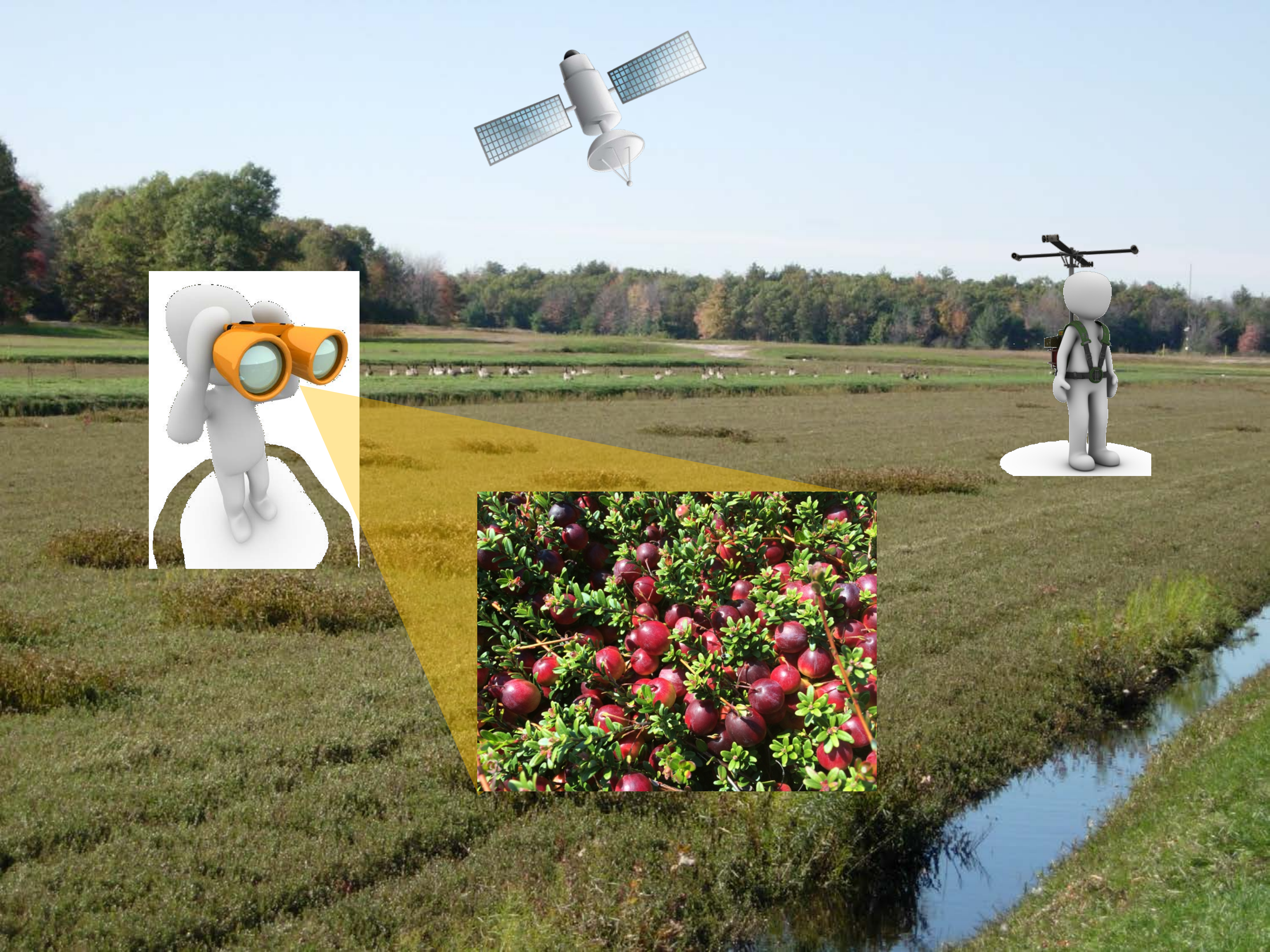
400 barrel/A projection

What contributes to yield?

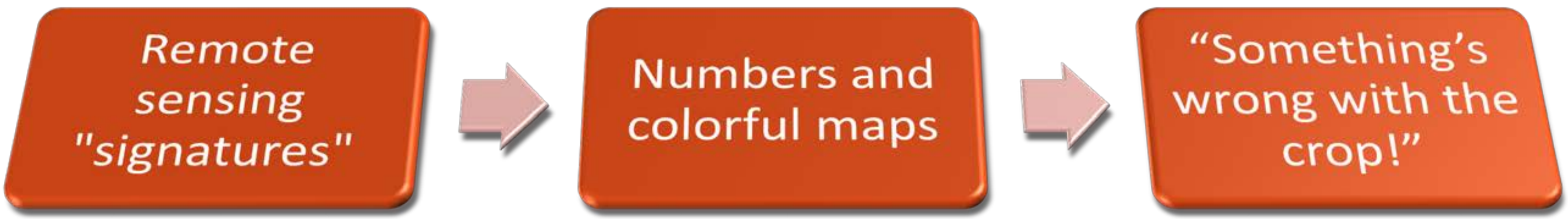


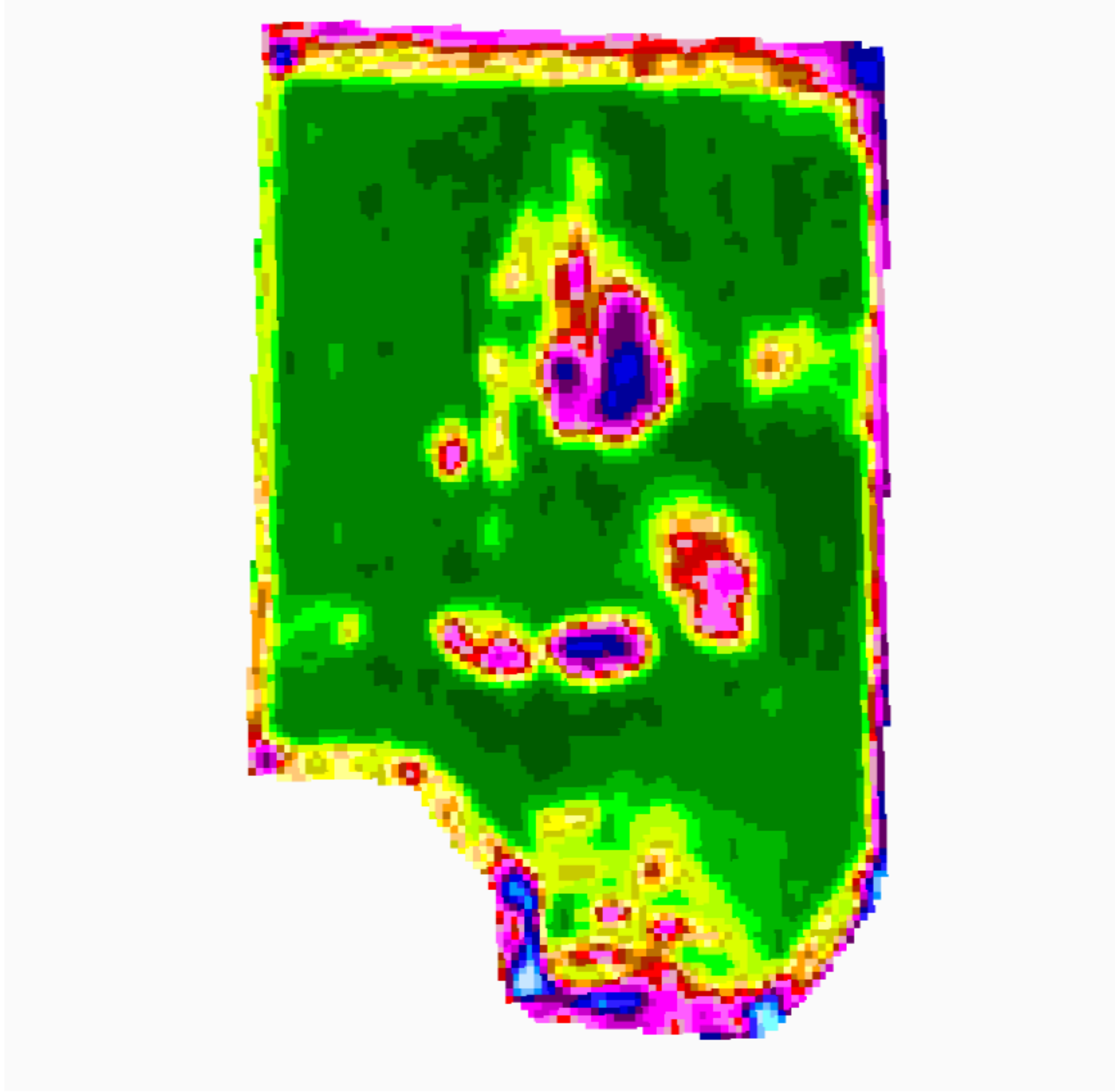
The value proposition

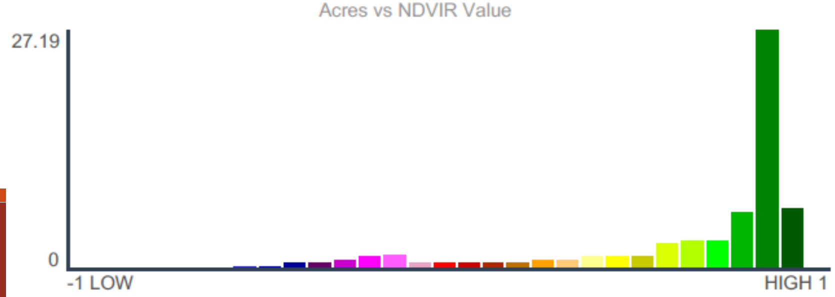
- Until we take a systems approach, it isn't likely that we'll solve multi-faceted production issues
  - Single variable research has value for single variable issues, i.e. herbicides control weeds
  - However, overall consistent production and quality aren't driven by single variables, but by a production SYSTEM
- This can be done anonymously and using existing data collection tools
- We already have much of the information, we just need to explore relationships in depth
- Get some ROI out of the information that growers already have to report, whether it be water withdrawals, pesticide use, financials, etc.



In the big picture: current state

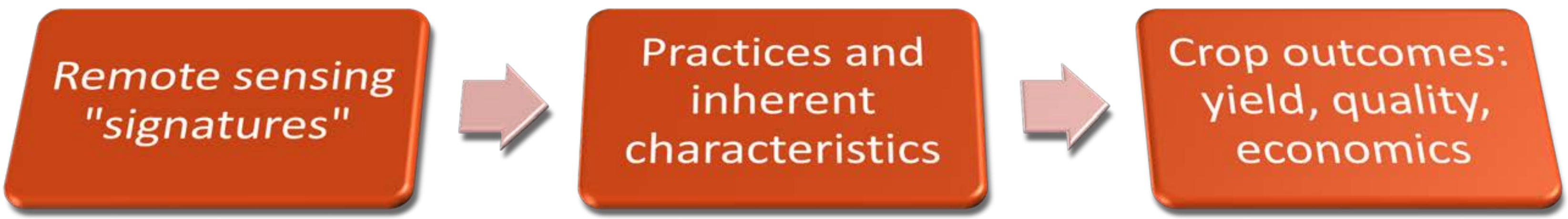


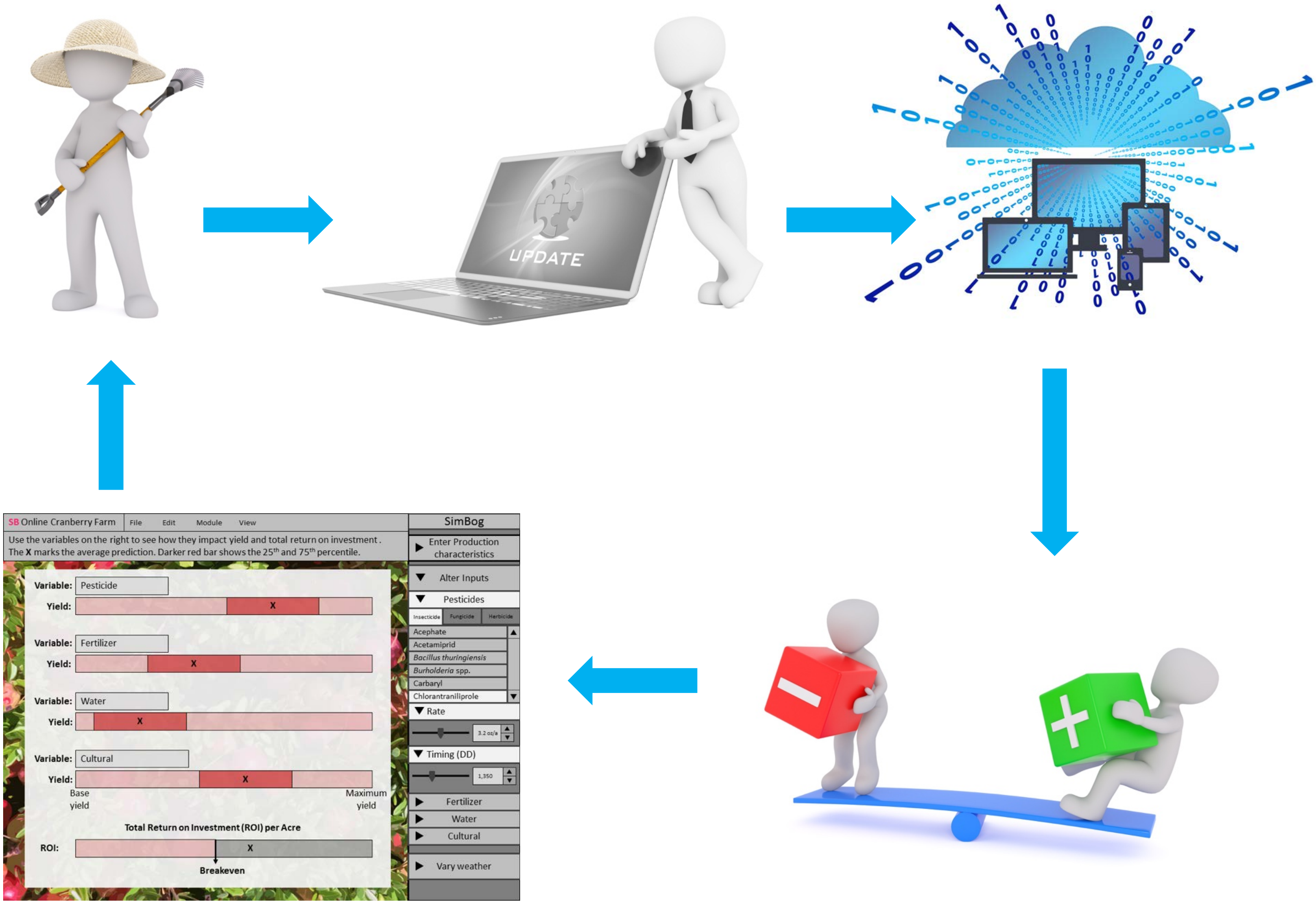




Acres vs NDVIR Value
27.19
0
-1 LOW
HIGH 1

In the big picture: future state




UPDATE
SB Online Cranberry Farm
File
Edit
Module
View
SimBog
Use the variables on the right to see how they impact yield and total return on investment.
The X marks the average prediction. Darker red bar shows the 25th and 75th percentile.
Enter Production characteristics
Alter Inputs
Pesticides
Insecticide
Fungicide
Herbicide
Acephate
Acetamiprid
Bacillus thuringiensis
Burholderia spp.
Carbaryl
Chlorantraniliprole
Rate
3.2 oz/a
Timing (DD)
1,350
Fertilizer
Water
Cultural
Vary weather
Variable: Pesticide
Yield: X
Variable: Fertilizer
Yield: X
Variable: Water
Yield: X
Variable: Cultural
Yield: X
Base yield
Maximum yield
Total Return on Investment (ROI) per Acre
ROI: X
Breakeven

Pilot study: methods

1. Established a work group of cranberry growers that created a list of data variables – inputs and outputs
2. Created an Excel-based data collection portal to collect crop production information
3. Used 2016 growing season pilot grower data to test relationships between production parameters and cranberry yield and quality

16 participating growers, over 500 beds of input

41 data variables

- Inputs:
  - Water: irrigation amount, flooding frequency, frost protection frequency
  - Pesticides: product, rate and timing
  - Cultural: flooding for insect control, crop scouting, nutrient testing, etc.
  - Fertilizer: product and rate, number and date of applications
- Production characteristics:
  - Stand age
  - Variety
  - Soil type and pH
- Weather:
  - Precipitation, GDD, frost-free growing season, sunlight
- Outputs:
  - Berry yield and quality (firmness, color, brix, size, useable, rot)

Pilot study: methods

1

THANK YOU FOR PILOTING THE CRANBERRY PRODUCTION MANAGER!

Your assistance will help build a powerful tool to optimize your cranberry production and save money by adding just the right amount of inputs. The more data that's entered, the more confident the answers become!

All data will be added to a single master spreadsheet for analysis without marsh identifiers, along with locally-available weather data. After the pilot, we'll work to streamline the final data collection parameters with currently-available electronic record keeping tools.

USE INSTRUCTIONS:

1. At the top of the sheet, please choose your county. This will allow us to add nearby weather data and growing degree days.
2. Enter your data for the first bed in the third column, labeled "Bed 1".
   If an arrow appears next to the column where you're entering data, click on the arrow and choose an answer from the drop-down menu.
3. If the information you entered for a data row for Bed 1 applies to all beds on your marsh, please select "yes" for that question from the drop-down choices in the second column. Your entry for that data row will automatically be added for the rest of the beds.
   Question: Within a data row, what if almost all the beds are the same as the first bed, but just a couple differ?
   Answer: Enter "yes" for the "Answer applies to all beds?" question in the second column, and then change the pre-populated answer for the few beds that differ from the first entry.

3. Don't worry if you're missing a few entry points - enter as much as you possibly can. Remember: the more data entered, the more powerful the results!

4. Save to the file often! Please also keep a copy for your own records.

QUESTIONS?

Please contact Jed Colquhoun: colquhoun@wisc.edu

Instructions | Cranberry Data | +

Pilot study: methods

Cranberry Production Manager							
County:							
	Answer applies to all beds?	Bed 1			Bed 2		
Bed identification number/name		YR01			YR02		
Production characteristics							
Cranberry variety		Ben Lear			Ben Lear		
Planting year (2016-year)		28			9		
Soil pH		5					
Tile drained bed?		0			0		
If bed is tile drained, spacing between tile lines (feet)							
If bed is tile drained, depth to tile lines (feet)							
Water management							
Irrigation							
Irrigation water pH		6.2					
Total irrigation amount (inches/season; all uses)							
Date of first irrigation frost watch (mm/dd) (days since Jan 1, 1990)		42512			42512		
Initial frost set point temperature (F)		33			33		
Fall frost/freeze protection							
# of flood protection events before ice formation		0			0		
# of ice formation floods		2			2		
Maximum ice thickness (inches)		12			12		
Ice off date (no ice on bed; mm/dd) (days since Jan 1, 1990)							
Monitor dissolved oxygen during flood events?		0			0		
Spring flooding							
Flood water pH							
Flooding for frost protection (days/season; enter '0' if none)		0			0		
Flooding for insect control (days/season; enter '0' if none)		0			0		
Flooding to remove trash (days/season; enter '0' if none)		5			5		
Average flood depth (inches of water above vines)		12			12		
Pest management							
Weed management							
Herbicide application 1		Casoron 4G	35	units	Casoron 4G	35	units
Herbicide application 2		Gly Star Plus	8:01	units	Gly Star Plus	0.33	units
Herbicide application 3							
Herbicide application 4							
Herbicide application 5							

Instructions | Cranberry Data | +

Pilot study: methods

- Machine learning: pattern recognition to make predictions (think email filtering)

- Mean
- LASSO regression to select important variables and regularize data
- Marginal effect to determine variable relationships (each pound of N is worth X barrels of cranberries)
- Next: combine marginal effects with economics (each pound of N increases net return by $X)

Pilot study: results

Variable	Mean	Minimum	Maximum	LASSO	Marginal Effect
Yield – barrels/a	354	26	695		
Nitrogen - lb	44	0	74	2.03	0.15
Phosphorus - lb	34	0	106	0	0
Potassium - lb	98	0	221	0	0
# of applications	6.5	0	10.0	11.7	6.8

Pilot study: results

Variable	Mean	Minimum	Maximum	LASSO	Marginal Effect
IRRIGATION					
Irrigation - in	27	0	241	0	0
First frost irrigation	4/28	4/19	5/22	0	0
Frost set point - F	30	28	36	0	0
FALL FREEZE					
Frost flood events	0.21	0	1	0	0
Ice floods	4.7	1.0	11.0	-2.44	-0.79
Maximum ice - in	12.0	9.5	18.0	-2.17	-0.86
Ice off date	3/14	3/2	3/31	0	0
Dissolved 0_2?	20%			0	0

Pilot study: results

Variable	Mean	Minimum	Maximum	LASSO	Marginal Effect
SPRING FLOODING					
Irrigation water pH	6.7	4.5	7.2	0	0
Flood water pH	6.9	4.5	8.5	0	0
Frost flood	1.7	0	10	0	0
Insect flood	0.3	0	2	0	0
Trash flood	2.0	0	10	-0.05	-0.02
Flood depth - in	11.3	1.0	24.0	0.14	0.03

In 2016, what mattered most?

ADDED BARRELS

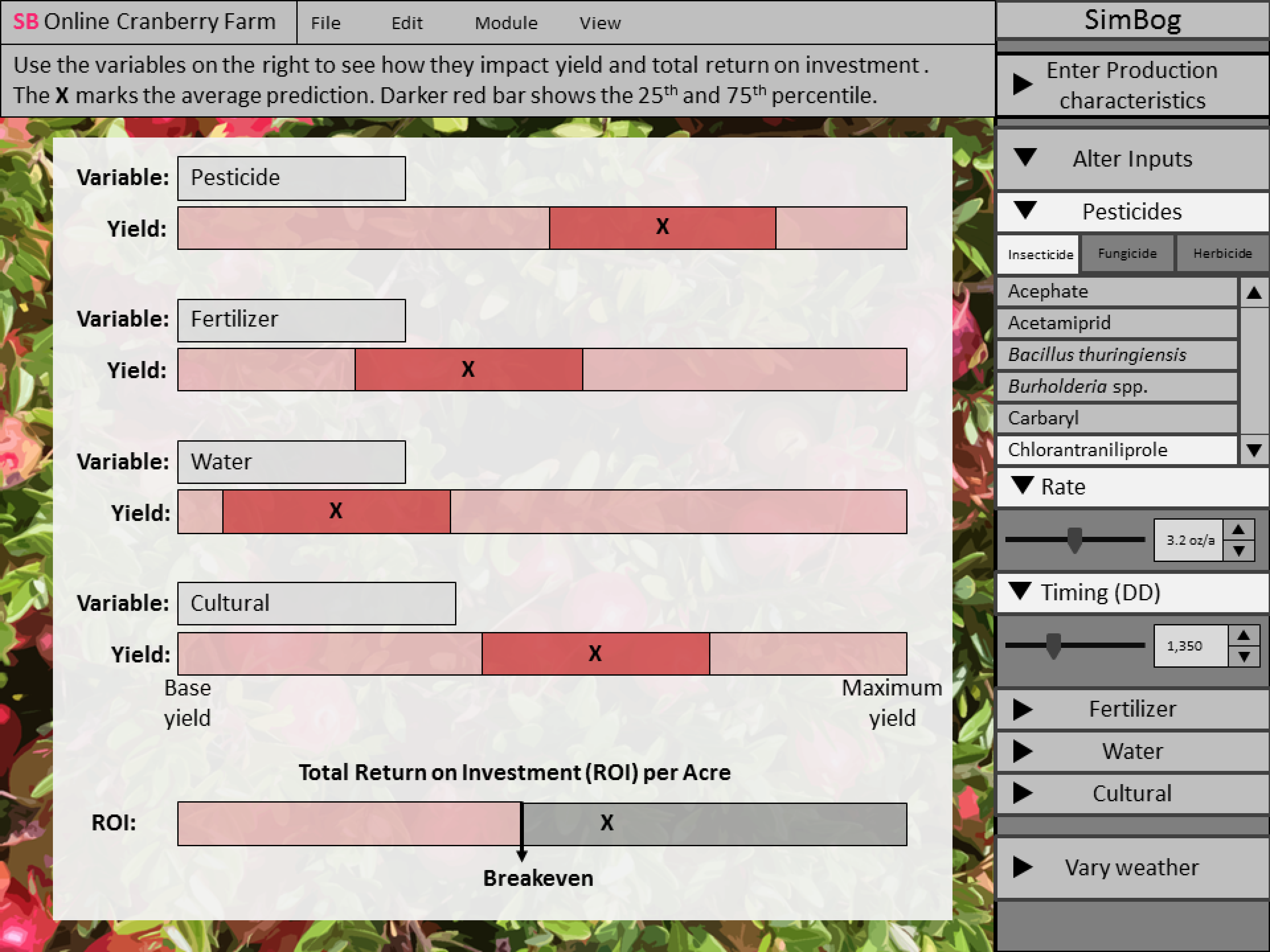
1. Variety
2. Insecticides and fungicides
3. # of fertilizer applications
4. Most recent sanding year

REDUCED BARRELS

1. Variety
2. Thick ice
3. # ice formation floods
4. Some herbicides (minor injury in absence of weeds has slight yield impact????)

Summary

- There is tremendous power here:
  - Growers already collect the data
  - We're developing a machine learning tool to efficiently predict outcomes from enormous data sets
  - After we have the data, we can ask any question that you want
- Next steps:
  - Add economics – turn the focus from more yield to quality and net return
  - Optimize scouting - connect remote sensing signatures to outcomes
  - Create SimBog – a grower-friendly interface, similar to gaming but backed by real data. Get a feeling for farm risk without turning a wheel…
- Multi-State Specialty Crop Block Grant submitted and in review


SB Online Cranberry Farm
File
Edit
Module
View
Use the variables on the right to see how they impact yield and total return on investment .
The X marks the average prediction. Darker red bar shows the 25th and 75th percentile.
Variable: Pesticide
Yield: X
Variable: Fertilizer
Yield: X
Variable: Water
Yield: X
Variable: Cultural
Yield: X
Base yield
Maximum yield
Total Return on Investment (ROI) per Acre
ROI: X
Breakeven
SimBog
Enter Production characteristics
Alter Inputs
Pesticides
Insecticide
Fungicide
Herbicide
Acephate
Acetamiprid
Bacillus thuringiensis
Burholderia spp.
Carbaryl
Chlorantraniliprole
Rate
3.2 oz/a
Timing (DD)
1,350
Fertilizer
Water
Cultural
Vary weather

This and title photo by Sevie Kenyon, UW-Madison