# Sorghum Field Effects: Location Correlations

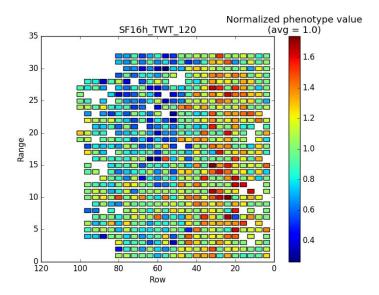
Ben Parr

bparr@cmu.edu

07/03/17

#### Motivation

- What is the correlation between a plant's location and harvest/robot features?
- Is it significant?



A seeming correlation between location and a total weight of the harvest plant. Credit: <u>Simon Heath's</u>

#### Methods

- Mantel Test using GPS locations.
  - Determined significance by randomly shuffling the GPS location of the data 10,000 times.
  - Used <u>scikit-bio implementation</u>.
- Computed GPS correlation for all available harvest and robot features.
- Computed correlations with GPS Eastings+Northings location, as well as with just Eastings, and with just Northings

$$\mathbf{A}_{R} = \begin{bmatrix} 0 & \mathbf{a}_{35} & \cdots & \mathbf{a}_{15} \\ \mathbf{a}_{35} & 0 & \cdots & \mathbf{a}_{13} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \mathbf{a}_{15} & \mathbf{a}_{13} & \cdots & 0 \end{bmatrix}.$$

Example shuffling of matrix A with the new 5, 3, ..., 1 order.

#### Results

- See Appendix for full results. Each table is sorted separately by p-value.
- All correlation coefficients are between -0.05 and 0.11
  - 2016\_08\_05-08\_vegetation\_index has largest absolute correlation (0.1080)
- Robot features:
  - Vegetation Index, Light Interception and Laser Plant Height features have low-p-values.
- Harvest features:
  - The Total Weight feature has the lowest p-value. A field effect on this feature can directly affect biofuel output.
  - Fewer harvest features have low p-values compared to robot features.

#### Discussion

- Field effects on chemical composition features?
  - Data is not currently available.
- How to correct for field effects?
  - When evaluating plants for breeding, need to normalize for field effects because the most productive plant might not have the highest biofuel output if it happened to grow in a poorer part of the field.
  - When predicting harvest features, consider synthesizing new features to improve predictive performance. Otherwise the assumption is the machine learning algorithm can take the field effects into account given the GPS locations.
- Does a max correlation of 0.11 warrant a correction?
- Investigate robot features with low p-values?

### Appendix A: Data for GPS Eastings and Northings

Feature Label	Observation of Feature	Correlation Coefficient	p-value
2016_08_05-08_vegetation_inde			
x	638	0.1080	0.0001
2016_09_light_interception	623	0.0560	0.0001
2016_07_13_laser_plant_height	745	0.0361	0.0010
SF16h_TWT_120	698	0.0388	0.0016
2016_07_light_interception	482	0.0490	0.0023
2016_08_light_interception	416	0.0402	0.0027
SF16h_WTL_120	698	0.0325	0.0091
SF16h_HGT3_120	698	0.0192	0.0663
SF16h_HGT1_120	698	0.0148	0.1556
2016_07_13-14_vegetation_inde			
x	526	0.0175	0.2150
SF16h_WTP_120	214	-0.0271	0.2325
SF16h_HGT2_120	698	0.0126	0.2385
SF16h_PAN2_120	214	0.0299	0.2500
SF16h_PAN3_120	214	0.0215	0.4085
SF16h_PAN1_120	214	0.0194	0.4472
2016_07_13_BAP_Leaf_Area	583	0.0084	0.5268

# Appendix B: Data for GPS Eastings Only

Feature Label	Observation of Feature	Correlation Coefficient	p-value
2016_07_13_laser_plant_height	745	0.0704	0.0001
2016_08_05-08_vegetation_inde x	638	0.0623	0.0003
2016_09_light_interception	623	0.0508	0.0006
2016_07_light_interception	482	0.0603	0.0013
SF16h_HGT1_120	698	0.0330	0.0091
SF16h_HGT3_120	698	0.0300	0.0156
SF16h_WTL_120	698	0.0354	0.0224
SF16h_HGT2_120	698	0.0295	0.0238
SF16h_TWT_120	698	0.0261	0.0899
SF16h_PAN3_120	214	-0.0419	0.0959
SF16h_PAN1_120	214	-0.0358	0.1504
SF16h_PAN2_120	214	-0.0228	0.3869
2016_08_light_interception	416	-0.0141	0.4239
2016_07_13-14_Leaf_Necrosis	674	-0.0098	0.4461
2016_07_13-14_vegetation_inde			
x	526	0.0132	0.4468
SF16h_WTP_120	214	0.0054	0.8118

## Appendix C: Data for GPS Northings Only

Feature Label	Observation of Feature	Correlation Coefficient	p-value
2016_08_05-08_vegetation_inde			
x	638	0.0961	0.0001
2016_08_light_interception	416	0.0477	0.0002
2016_09_light_interception	623	0.0365	0.0015
2016_07_light_interception	482	0.0313	0.0415
SF16h_TWT_120	698	0.0239	0.0589
SF16h_PAN3_120	214	0.0531	0.0743
SF16h_PAN2_120	214	0.0515	0.0887
SF16h_PAN1_120	214	0.0484	0.0963
SF16h_WTP_120	214	-0.0352	0.1802
2016_07_13_laser_plant_height	745	0.0131	0.2464
SF16h_WTL_120	698	0.0132	0.2985
2016_07_13-14_vegetation_inde			
x	526	0.0104	0.4733
SF16h_HGT2_120	698	-0.0054	0.6157
SF16h_HGT1_120	698	-0.0048	0.6435
2016_07_13-14_Leaf_Necrosis	674	0.0046	0.6630
2016_07_13_BAP_Leaf_Area	583	0.0022	0.8688