PLANT BREEDING & GENOMICS

Field Phenomics Part 2 of 2: Data Analysis

Presented by

Michael Gore
Cornell University

&

Kelly Thorp
USDA-ARS in Maricopa, Arizona

Hosted by
Shawn Yarnes
Plant Breeding and Genomics

nite extension iculture
ce

U.S. Arid Land Agricultural Research Center, Maricopa, Arizona

Geospatial data processing

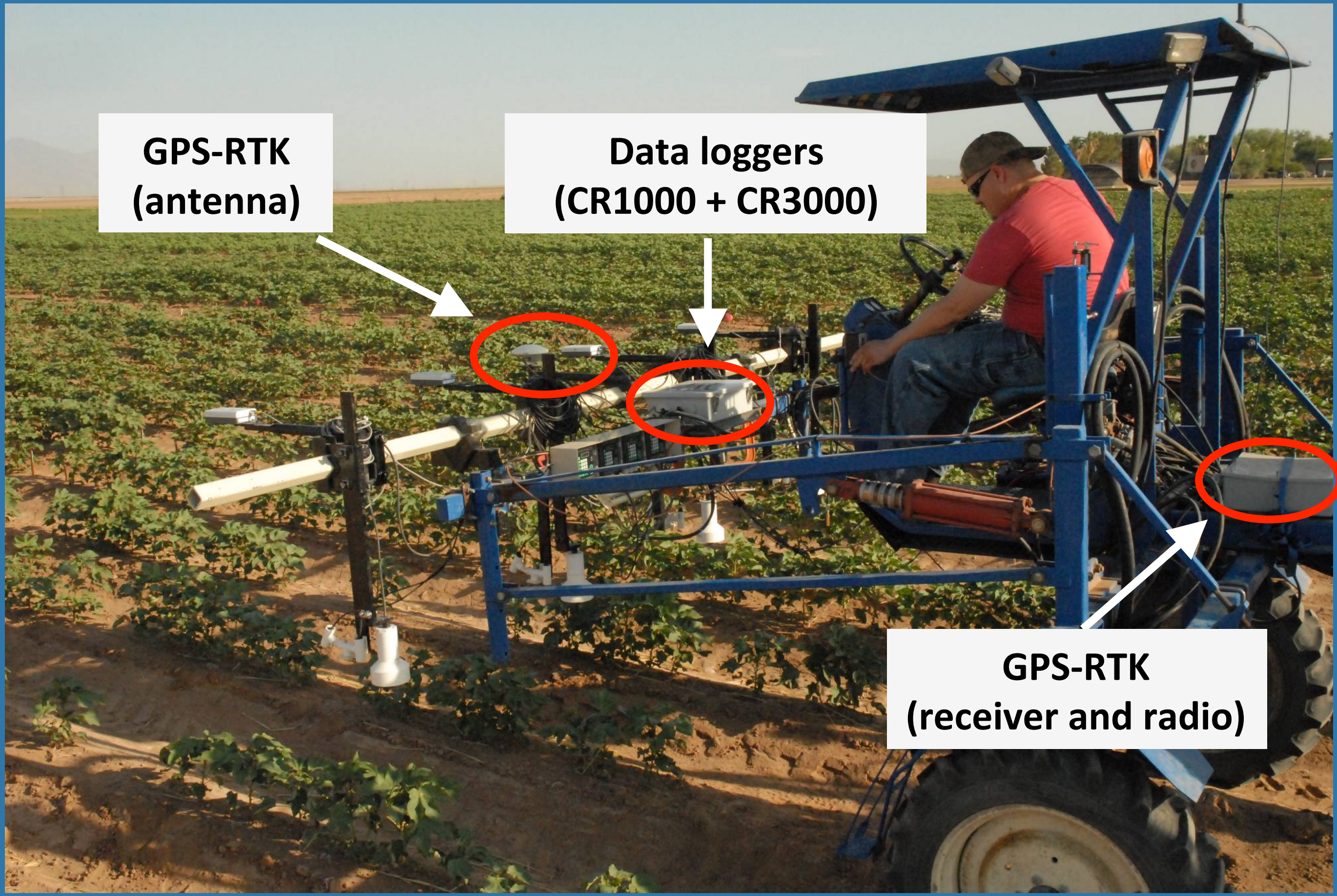
United States Department of Agriculture
Agricultural Research Service
U.S. Arid Land Agricultural Research Center, Maricopa, Arizona

HTP: Sensors, platform, and vehicle

Infrared thermometer

Ultrasonic Transducer

Multi-spectral sensor



Average rate of 0.84 ha h^{-1}

Andrade-Sanchez, P., Gore, M. A. et al. 2013 Functional Plant Biology In press

Field Map in Spreadsheet Format

Dirt Road

Range: | **Border 1 WW** | S W + E N | **Border 2 WL**

Dirt Road (west side) | 1 ROW PHY 800 borders on each side of each block | 1 ROW DP 393 | SEPERATE COTTON EXPERIMENTS

Border 1 WW

Rep3 1 Range of PHY 800 Rep4 0.85 m

Range										
11	321 Pima S-4 22	322 P70 8	323 Pima S-6 1	324 Pima S-2 12	325 St. Vin. 9	421 Mon. S.I. 2	422 Pima 32 13	423 Pima S-3 19	424 St. Vin. 9	425 Old Pima 14
10	316 89590 3	317 P73 6	318 PSI 425 21	319 Mon. S.I. 2	320 P62 24	416 Pima S-6 1	417 Pima S-5 17	418 OA360 25	419 P62 24	420 94220 4
9	311 OA360 25	312 94217 11	313 Pima S-1 10	314 Amsak 18	315 Old Pima 14	411 94217 11	412 Pima S-7 23	413 93260 7	414 P73 6	415 Pima S-2 12
8	306 Pima S-3 19	307 94220 4	308 Pima S-7 23	309 P53 15	310 8810 16	406 P53 15	407 PSI 425 21	408 P76 20	409 Pima S-4 22	410 Pima S-1 10
7	301 93260 7	302 P76 20	303 PSI 113 5	304 Pima S-5 17	305 Pima 32 13	401 8810 16	402 PSI 113 5	403 89590 3	404 P70 8	405 Amsak 18
6	1 Range of PHY 800									
5	121 Pima S-2 12	122 P62 24	123 P53 15	124 Amsak 18	125 Pima 32 13	221 PSI 113 5	222 Pima S-3 19	223 Pima S-4 22	224 P62 24	225 94217 11
4	116 Pima S-5 17	117 8810 16	118 Old Pima 14	119 P73 6	120 Pima S-4 22	216 P70 8	217 Pima 32 13	218 94220 4	219 Pima S-1 10	220 P73 6
3	111 Mon. S.I. 2	112 Pima S-1 10	113 PSI 113 5	114 Pima S-7 23	115 Pima S-6 1	211 Pima S-6 1	212 P53 15	213 Old Pima 14	214 89590 3	215 93260 7
2	106 PSI 425 21	107 93260 7	108 P70 8	109 Pima S-3 19	110 OA360 25	206 Pima S-7 23	207 PSI 425 21	208 St. Vin. 9	209 Amsak 18	210 Pima S-5 17
1	101 St. Vin. 9	102 P76 20	103 94220 4	104 89590 3	105 94217 11	201 Pima S-2 12	202 OA360 25	203 P76 20	204 8810 16	205 Mon. S.I. 2

Rep1 1 Range of PHY 800 Rep2 0.85 m

Border 2 WL

Rep3 1 Range of PHY 800 Rep4 0.85 m

Range										
11	321 8810 16	322 Pima S-1 10	323 P62 24	324 Pima S-2 12	325 Pima 32 13	421 PSI 113 5	422 Pima S-3 19	423 Amsak 18	424 Pima S-1 10	425 89590 3
10	316 OA360 25	317 PSI 113 5	318 Pima S-4 22	319 PSI 425 21	320 93260 7	416 P62 24	417 Old Pima 14	418 P73 6	419 93260 7	420 P76 20
9	311 89590 3	312 Pima S-5 17	313 P73 6	314 94220 4	315 P70 8	411 Pima S-2 12	412 Mon. S.I. 2	413 P53 15	414 OA360 25	415 P70 8
8	306 Pima S-3 19	307 St. Vin. 9	308 Mon. S.I. 2	309 94217 11	310 Old Pima 14	406 Pima S-4 22	407 8810 16	408 94217 11	409 94220 4	410 Pima S-7 23
7	301 P76 20	302 Pima S-6 1	303 P53 15	304 Pima S-7 23	305 Amsak 18	401 PSI 425 21	402 St. Vin. 9	403 Pima S-5 17	404 Pima S-6 1	405 Pima 32 13
6	1 Range of PHY 800									
5	121 OA360 25	122 Pima S-6 1	123 94220 4	124 Old Pima 14	125 Pima S-1 10	221 Mon. S.I. 2	222 P73 6	223 PSI 113 5	224 Pima S-6 1	225 8810 16
4	116 Pima S-7 23	117 Pima S-3 19	118 Pima S-2 12	119 PSI 425 21	120 P73 6	216 Pima S-1 10	217 P70 8	218 P76 20	219 94217 11	220 PSI 425 21
3	111 P76 20	112 Pima 32 13	113 Pima S-4 22	114 89590 3	115 Mon. S.I. 2	211 Old Pima 14	212 Pima S-4 22	213 Amsak 18	214 Pima S-2 12	215 Pima S-5 17
2	106 Amsak 18	107 P70 8	108 8810 16	109 93260 7	110 St. Vin. 9	206 Pima S-3 19	207 93260 7	208 P53 15	209 94220 4	210 Pima 32 13
1	101 94217 11	102 P53 15	103 Pima S-5 17	104 PSI 113 5	105 P62 24	201 89590 3	202 Pima S-7 23	203 St. Vin. 9	204 P62 24	205 OA360 25

Rep1 1 Range of PHY 800 Rep2 0.85 m

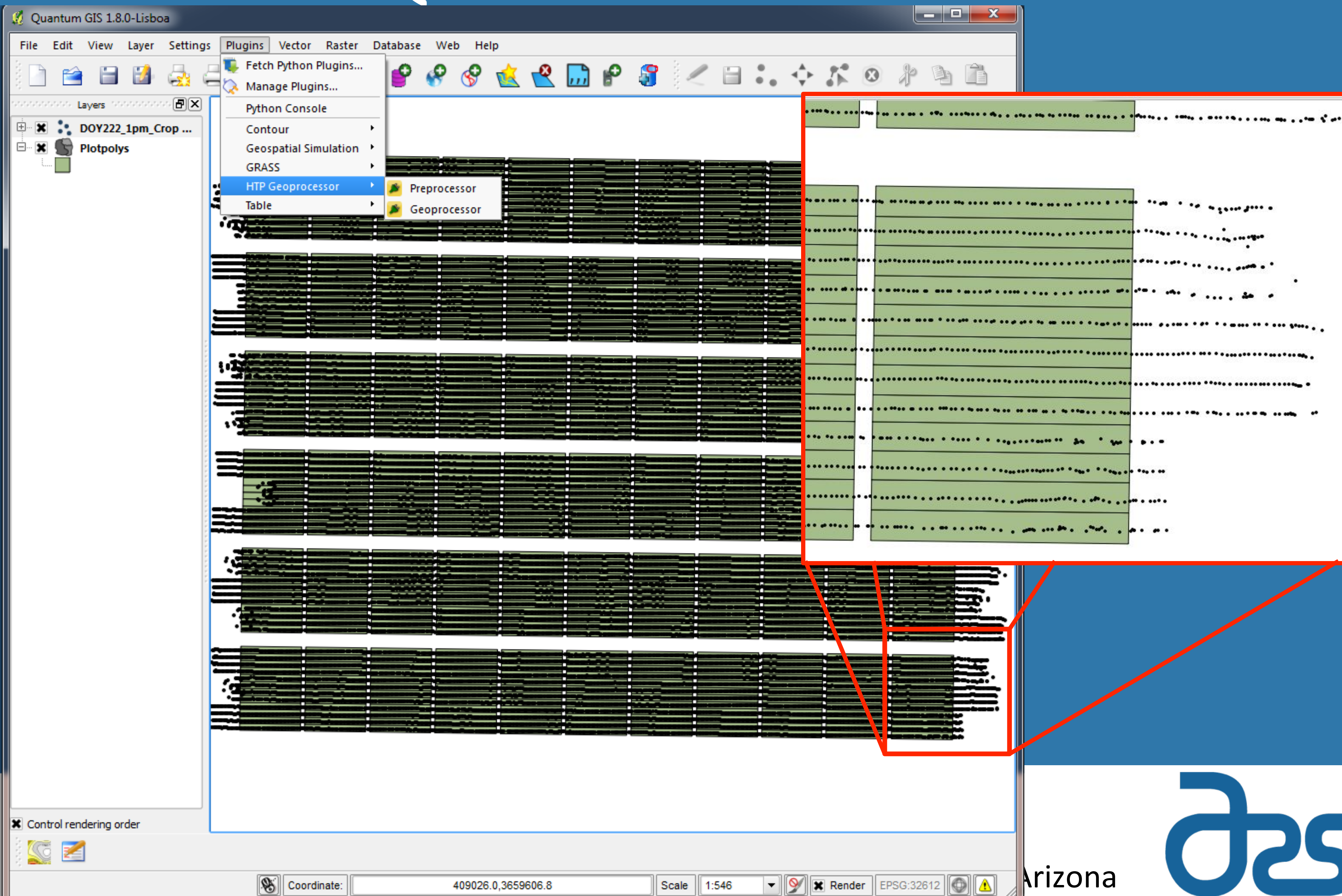
Not drawn to scale

Dry Ravine and Paved Road

Geographic Information Systems

- Appropriately consider the geospatial nature of the data
- GRASS GIS, Quantum GIS, gvSig, ESRI ArcGIS, etc.
- Quantum GIS (QGIS; www.qgis.org)
  - Open-source, desktop GIS
  - New – Version 1.0 released in January 2009
  - Lightweight – Less RAM, hard drive, CPU requirements
  - Platform independent – Linux, Unix, Mac OS, Windows
  - Easily extendable
    - Scripts or 'plug-ins' written in C++ or Python
    - Object-oriented API for accessing QGIS functionality
    - Qt or PyQt libraries for implementing GUIs
- HTP Geoprocessor
  - Open-source Quantum GIS plug-in for processing HTP data
  - Independent of data collection in the field
  - Independent of the sensor data file structure

QGIS Environment

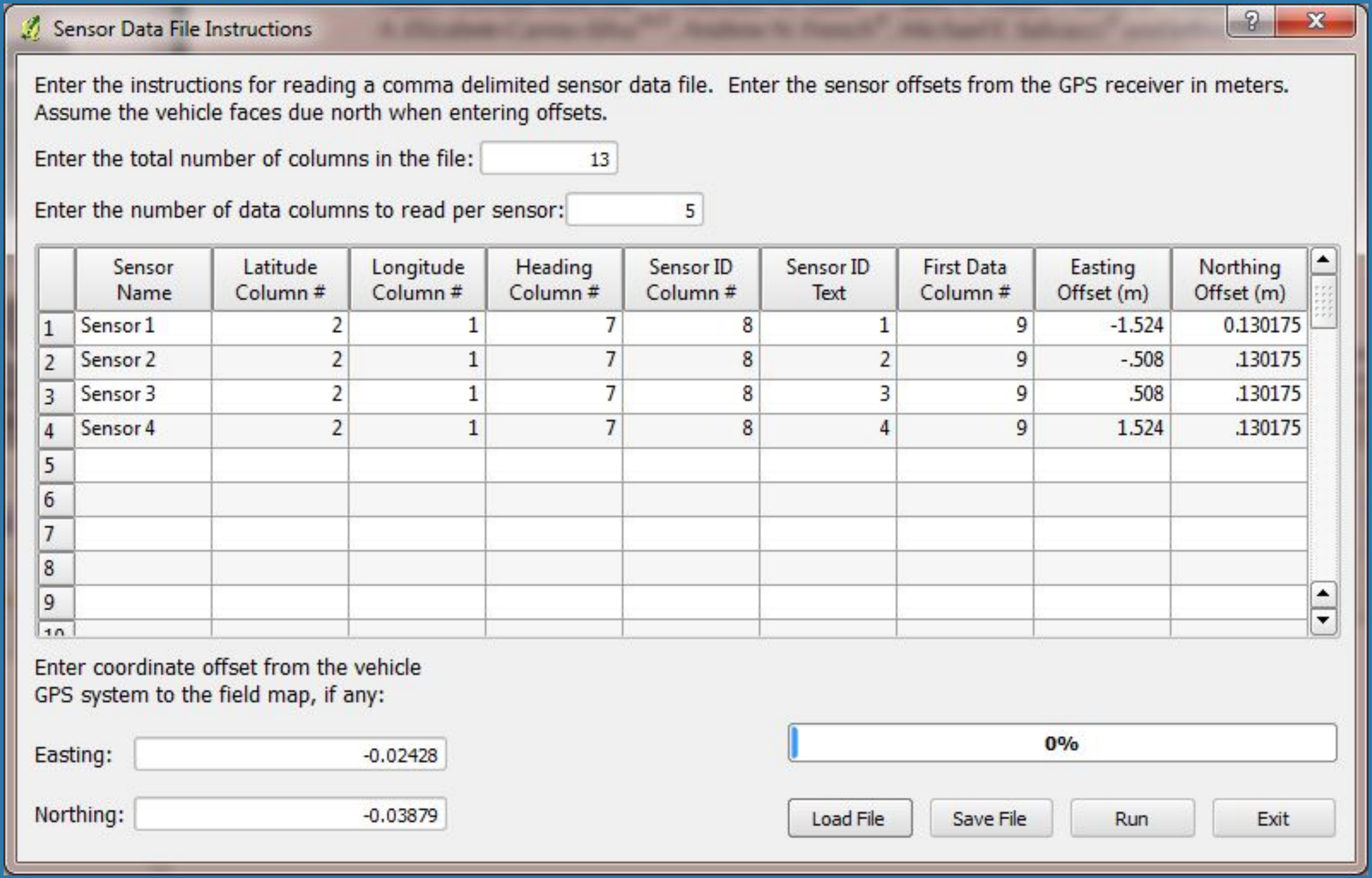


Arizona

HTP Geoprocessor Tools

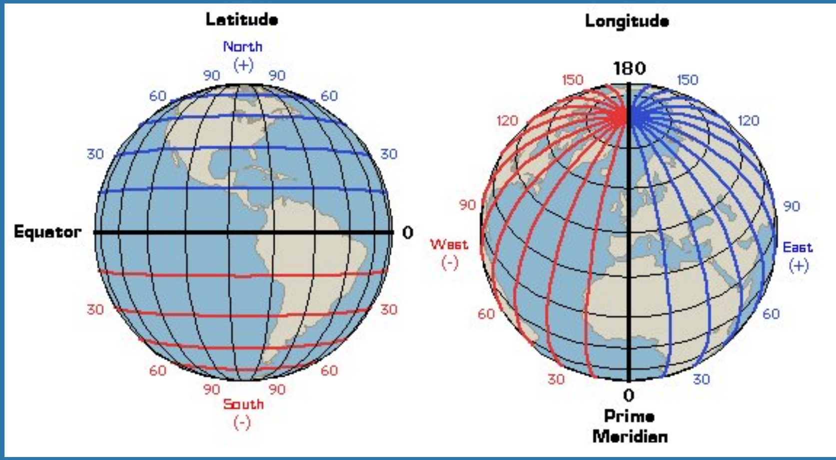
- Preprocessor
  - Read sensor data from files (independent of file structure)
  - Convert latitude/longitude to UTM coordinates
  - Calculate UTM coordinate transformations
    - From GPS receiver position to each sensor position along the boom
    - Depends on course (heading) of vehicle travel
- Geoprocessor
  - Analyze sensor data iteratively within plot boundaries
    - Calculate summary statistics for each plot
    - Assign plot ID number to each sensor data point
  - Requires accurate plot boundary map as polygon shapefile

Preprocessing – Reading Data

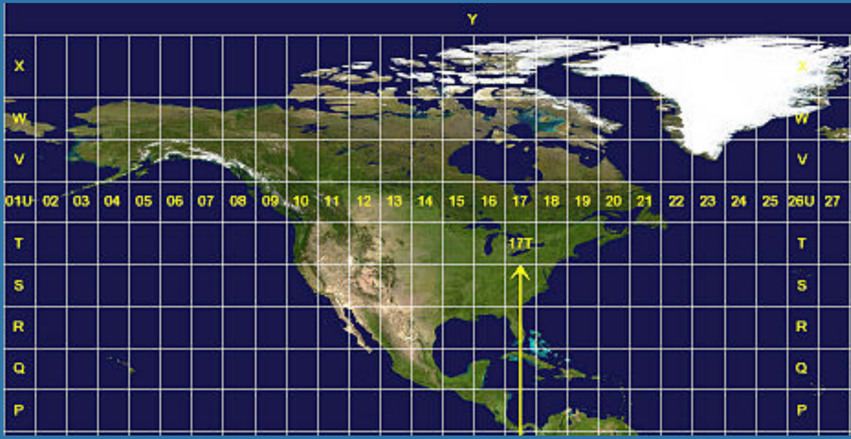


Preprocessing – LL to UTM

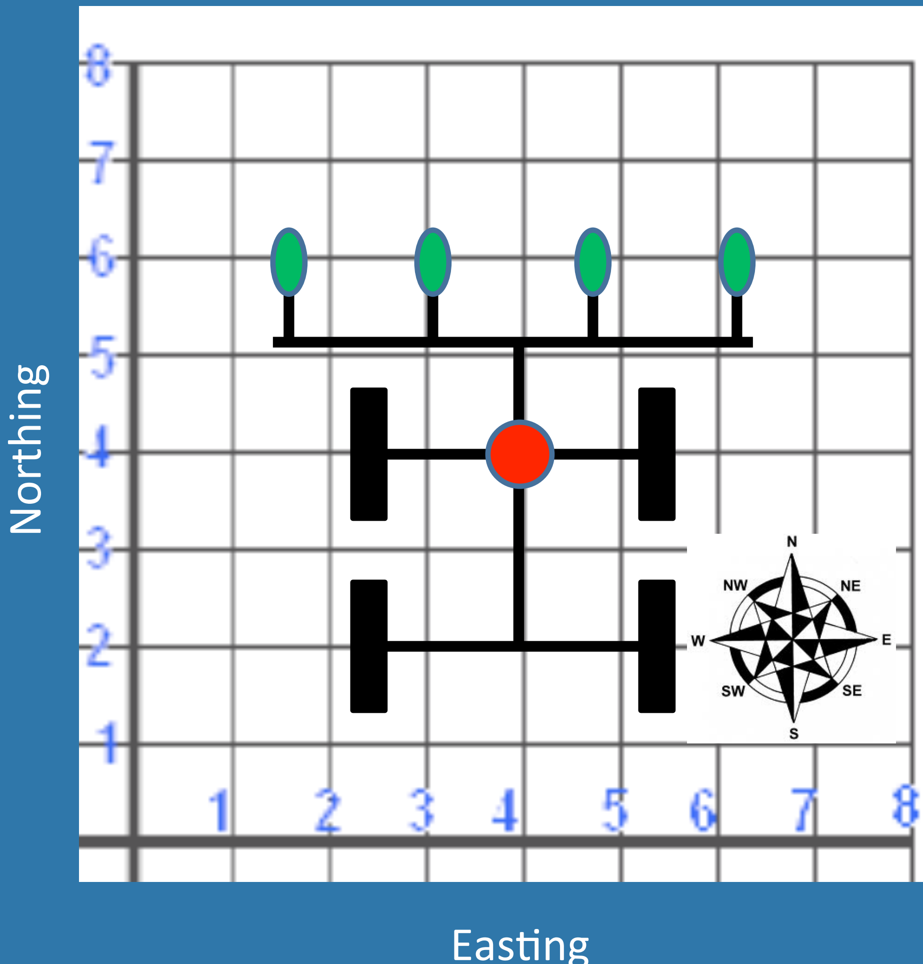
- GPS Receivers
  - Provide Lat/Long (degrees)
  - Provide elevation (meters)
  - Use the WGS 84 datum
    - Reference ellipsoid (shape of Earth)
    - Geoid (shape of ocean surface)

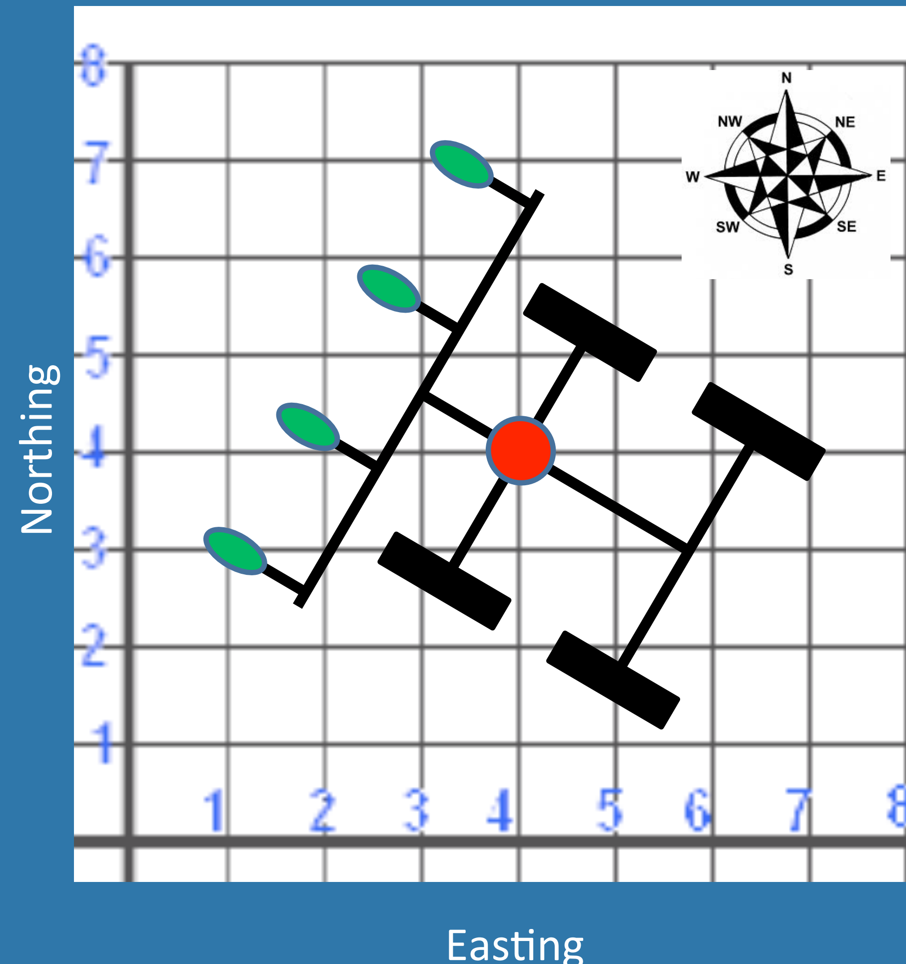


- UTM Map Projections
  - Converts Lat/Long to a Cartesian coordinate system in meters
  - Permits treatment of geospatial data by simple Euclidean distance
  - Minimal distortion at HTP scales



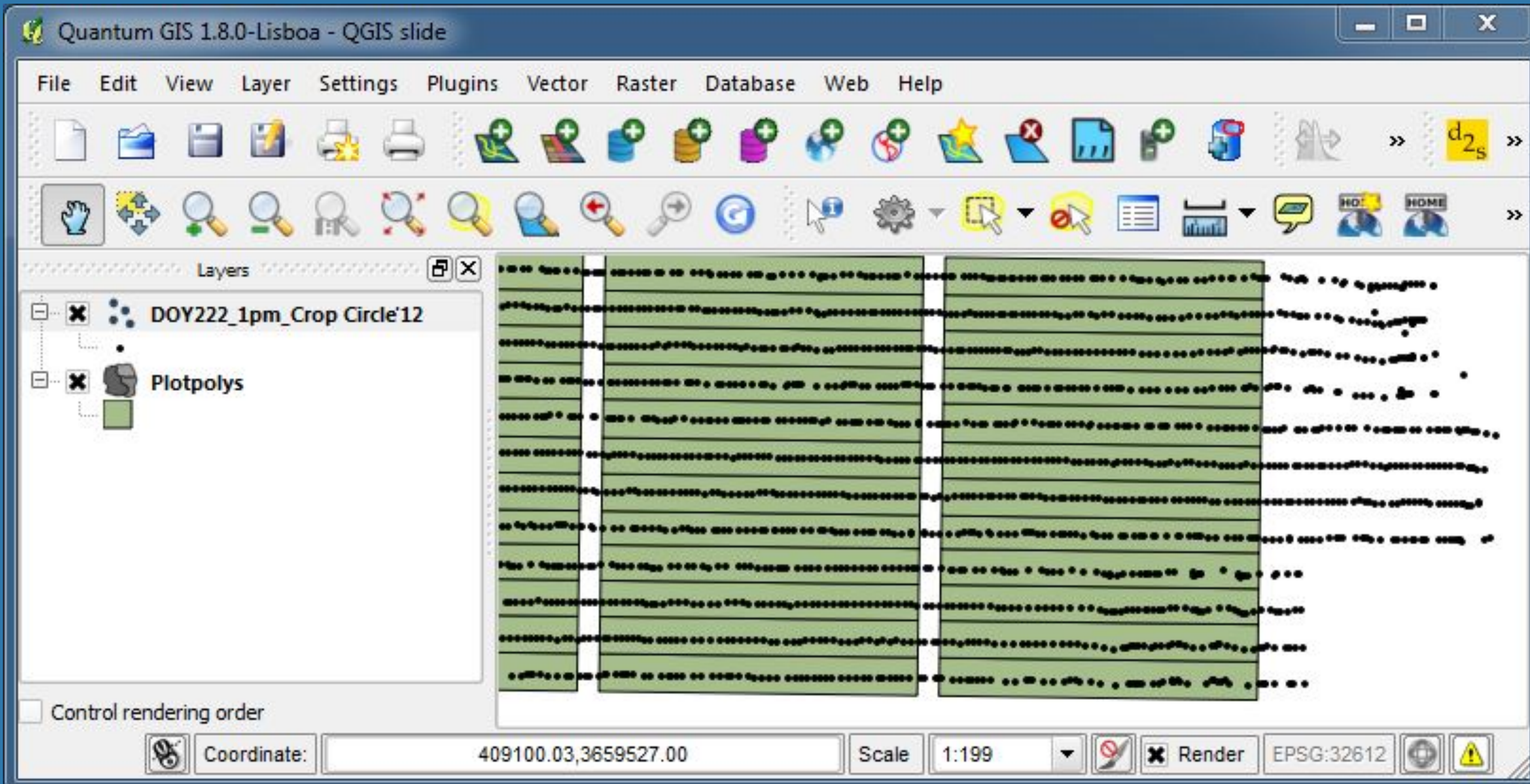
Preprocessing – Sensor Positions





United States Department of Agriculture
Agricultural Research Service
U.S. Arid Land Agricultural Research Center, Maricopa, Arizona

Geoprocessing – Load files into QGIS



Geoprocessing – Attribute Table for Sensor Data Shapefile

Attribute table - DOY222_1pm_Crop Circle'12 :: 0 / 50836 feature(s) selected

	Sensor	SensorID	Latitude	Longitude	Heading	UTMX-rec	UTMY-rec	UTMZ-rec	UTMX-sen	UTMY-sen	UTMZ-sen	Data1	Data2
0	Sensor 1	1	33.070436	-111.9736343	267.29	409117.670231	3659516.87469	12S	409117.612258	3659515.34624	12S	0.845	3.1
1	Sensor 2	2	33.070436	-111.9736343	267.29	409117.670231	3659516.87469	12S	409117.564221	3659516.36111	12S	1	[illegible]
2	Sensor 3	3	33.070436	-111.9736343	267.29	409117.670231	3659516.87469	12S	409117.516183	3659517.37597	12S	1	[illegible]
3	Sensor 4	4	33.070436	-111.9736343	267.29	409117.670231	3659516.87469	12S	409117.468146	3659518.39084	12S	-1	0.44
4	Sensor 1	1	33.070436	-111.9736362	271.74	409117.492872	3659516.87634	12S	409117.316482	3659515.35699	12S	0.777	1.83
5	Sensor 2	2	33.070436	-111.9736362	271.74	409117.492872	3659516.87634	12S	409117.347332	3659516.37253	12S	0.838	2.53
6	Sensor 3	3	33.070436	-111.9736362	271.74	409117.492872	3659516.87634	12S	409117.378182	3659517.38806	12S	-0.989	0.70
7	Sensor 4	4	33.070436	-111.9736362	271.74	409117.492872	3659516.87634	12S	409117.409032	3659518.40359	12S	0.48	0.6
8	Sensor 1	1	33.0704359	-111.9736381	267.61	409117.31541	3659516.8669	12S	409117.248901	3659515.33879	12S	0.69	1.5
9	Sensor 2	2	33.0704359	-111.9736381	267.61	409117.31541	3659516.8669	12S	409117.206532	3659516.35391	12S	0.464	0.96
10	Sensor 3	3	33.0704359	-111.9736381	267.61	409117.31541	3659516.8669	12S	409117.164164	3659517.36903	12S	0.447	1.01
11	Sensor 4	4	33.0704359	-111.9736381	267.61	409117.31541	3659516.8669	12S	409117.121795	3659518.38414	12S	0.418	0.59

Look for [ ] in Sensor — Search

Show selected only — Search selected only — Case sensitive — Advanced search — ? — Close

Geoprocessing – Attribute Table for Plot Boundary Shapefile

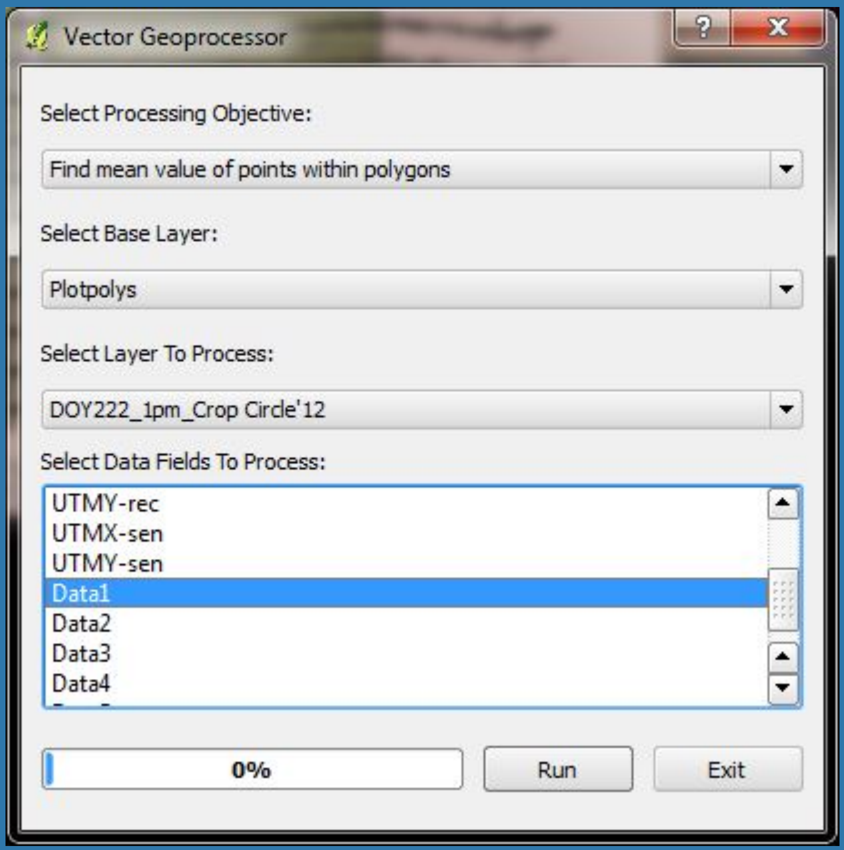
Attribute table - Plotpolys :: 0 / 792 feature(s) selected

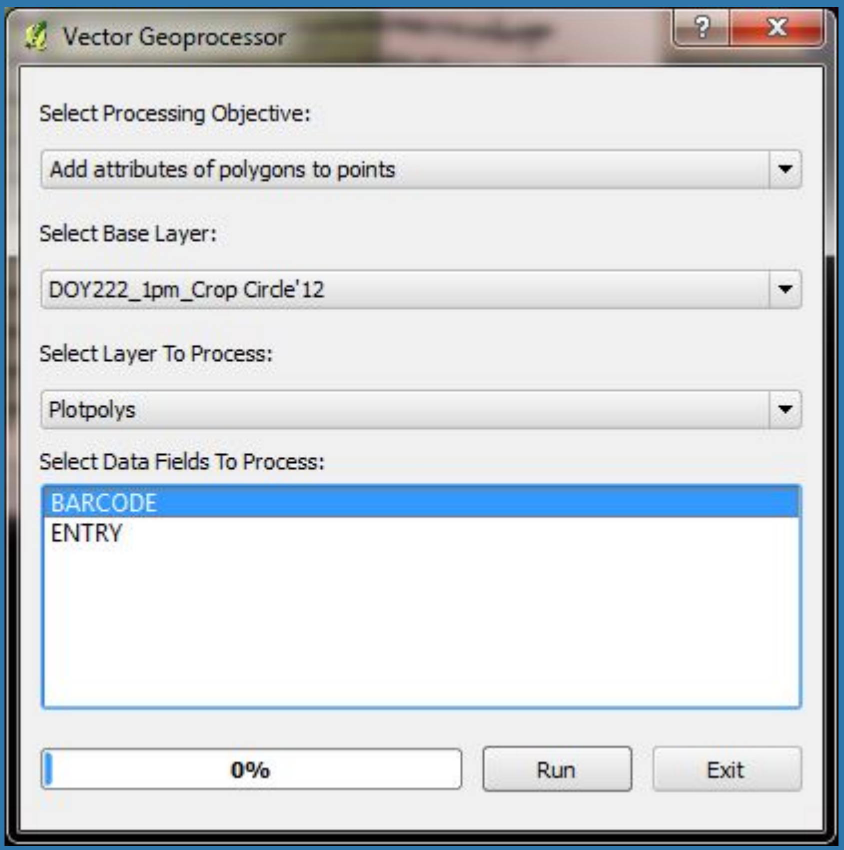
	ID	MEAST	MNORTH	ROW	RANGE	BORDER	PLOT	BLOCK	TRT	REP	BARCODE	ENTRY
0	1	409111.914453	3659515.459046	1	1	1	0	0	0	0	NULL	NULL
1	2	409111.926272	3659516.505072	2	1	1	1	1	1	1	11W101	S. I. St. Vin. V-135
2	3	409111.933851	3659517.518477	3	1	1	6	2	2	1	11W102	P76
3	4	409111.954499	3659518.512428	4	1	1	11	3	4	1	11W103	94220
4	5	409111.957906	3659519.469435	5	1	1	16	4	8	1	11W104	89590
5	6	409111.98368	3659520.447233	6	1	1	21	5	17	1	11W105	94217
6	7	409111.995141	3659521.465521	7	1	1	26	6	24	2	11W201	Pima S-2
7	8	409112.014021	3659522.503686	8	1	1	31	7	9	2	11W202	OA360
8	9	409112.023058	3659523.532953	9	1	1	36	8	6	2	11W203	P76
9	10	409112.044764	3659524.544605	10	1	1	41	9	22	2	11W204	8810
10	11	409112.054217	3659525.561085	11	1	1	46	10	20	2	11W205	Monseratt S. I.
11	12	409112.072056	3659526.579257	12	1	1	0	0	0	0	NULL	NULL
12	13	409102.466491	3659515.579444	1	2	1	0	0	0	0	NULL	NULL
13	14	409102.478249	3659516.619466	2	2	1	2	1	12	1	11W106	PSI 425
14	15	409102.485943	3659517.634676	3	2	1	7	2	15	1	11W107	93260
15	16	409102.506658	3659518.63113	4	2	1	12	3	23	1	11W108	P70
16	17	409102.510312	3659519.593202	5	2	1	17	4	22	1	11W109	Pima S-3

Look for [ ] in ID Search

Show selected only   Search selected only   Case sensitive   Advanced search   ?   Close

Geoprocessing – Vector Shapefile Geoprocessor





Geoprocessing – Final Attribute Table for Sensor Data Shapefile

Attribute table - DOY222_1pm_Crop Circle'12 :: 0 / 50836 feature(s) selected

	UTMX-sen	UTMY-sen	UTMZ-sen	Data1	Data2	Data3	Data4	Data5	BARCODE
222	409107.248684	3659517.51563	12S	0.487	1.272	0.1023	0.2969	0.1307	NULL
223	409107.201178	3659518.53052	12S	0.756	1.563	0.0634	0.4573	0.1785	NULL
224	409107.036947	3659515.49596	12S	0.685	1.372	0.0854	0.4562	0.1923	Fill
225	409107.063188	3659516.51162	12S	0.482	2.05	0.1111	0.3182	0.1043	12D1011
226	409107.089429	3659517.52728	12S	0.378	0.802	0.1202	0.2666	0.1479	12D1012
227	409107.115671	3659518.54294	12S	0.693	1.34	0.0744	0.411	0.1756	12D1013
228	409106.932969	3659515.49032	12S	0.761	1.771	0.0647	0.4761	0.1718	Fill
229	409106.897866	3659516.50571	12S	0.473	2.236	0.1059	0.2963	0.0916	12D1011
230	409106.862763	3659517.5211	12S	0.249	0.531	0.1486	0.2471	0.1614	12D1012
231	409106.827659	3659518.5365	12S	0.42	0.717	0.1149	0.281	0.1637	12D1013
232	409106.624476	3659515.50201	12S	0.713	2.033	0.0813	0.4855	0.1601	Fill
233	409106.664364	3659516.51723	12S	0.424	1.323	0.1126	0.2784	0.1198	12D1011
234	409106.704252	3659517.53244	12S	0.475	0.798	0.1047	0.2942	0.1636	12D1012
235	409106.74414	3659518.54766	12S	0.408	0.636	0.1111	0.264	0.1614	12D1013
236	409106.408417	3659515.5077	12S	0.685	1.858	0.0915	0.4889	0.1711	Fill
237	409106.467964	3659516.52196	12S	0.308	1.007	0.1462	0.2765	0.1377	12D1011

Look for [ ] in Sensor Search

Show selected only Search selected only Case sensitive Advanced search ? Close

Statistical data processing and analysis

- Format the geoprocessed raw data for statistical analysis with SAS or R software packages

BARCODE/Plot ID	ENTRY	ROW	RANGE	BLOCK	REP	TREATMENT	Canopy Tm °C	Day Year	Time MST
11W101	S. I. St. Vin. V-135	2	1	1	1	W	43.64	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	41.58	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	36.85	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	39.74	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	44.23	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	40.99	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	43.69	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	44.38	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	37.96	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	47.3	202	13
11W101	S. I. St. Vin. V-135	2	1	1	1	W	48.49	202	13
11D101	94217	2	1	1	1	D	36.61	202	13
11D101	94217	2	1	1	1	D	38.29	202	13
11D101	94217	2	1	1	1	D	38.51	202	13
11D101	94217	2	1	1	1	D	39.97	202	13
11D101	94217	2	1	1	1	D	40.01	202	13
11D101	94217	2	1	1	1	D	40.28	202	13
11D101	94217	2	1	1	1	D	40.37	202	13
11D101	94217	2	1	1	1	D	41.71	202	13
11W106	PSI 425	2	2	1	1	W	42.97	202	13
11W106	PSI 425	2	2	1	1	W	40.64	202	13
11W106	PSI 425	2	2	1	1	W	49.59	202	13
11W106	PSI 425	2	2	1	1	W	40.91	202	13
11W106	PSI 425	2	2	1	1	W	39.82	202	13
11W106	PSI 425	2	2	1	1	W	39.4	202	13
11W106	PSI 425	2	2	1	1	W	41.59	202	13
11D106	Amsak	2	2	1	1	D	40.89	202	13
11D106	Amsak	2	2	1	1	D	40.81	202	13
11D106	Amsak	2	2	1	1	D	41.61	202	13
11D106	Amsak	2	2	1	1	D	43.23	202	13
11D106	Amsak	2	2	1	1	D	40.37	202	13
11D106	Amsak	2	2	1	1	D	40.25	202	13
11D106	Amsak	2	2	1	1	D	39.26	202	13

Exploratory data analysis: know your data

- Histogram
- Box-plot (Box and Whisker plot)
- Scatter plot
- Quantile-Quantile plot – graphical method for comparing two probability distributions to assess goodness-of-fit
- Summary Statistics
  - Mean, standard deviation, and variance with confidence intervals
  - Tests for normality
  - Tests for outliers
  - Quantiles and related confidence intervals
  - Frequency counts for observations
  - Missing values

Phenotypic Outliers

"Not all outliers are illegitimate contaminants, and not all illegitimate scores show up as outliers." (Barnett & Lewis, 1994)

- Outliers are "unusual" data points that substantially deviate from the mean and strongly influence parameter estimates
- Should always check for outliers in your data sets
- Do not ignore outliers if detected
- Outliers can result in several negative consequences
  - Increase error variance
  - Reduce the power of statistical tests
  - Distort estimates
  - Decrease normality if non-randomly distributed
  - **Inferences made from the fitted model may not be representative of the population under study**

Barnett V, Lewis T (1994). Outliers in Statistical Data, 3rd edition, John Wiley, New York, NY, USA.

Potential Causes of Phenotypic Outliers

- Human errors (manual or code) in data collection, recording, or entry
- Technical errors from faulty or non-calibrated equipment used for measurement of plant phenotypes
- Sampling error such as underrepresentation of a subpopulation
- Incorrect assumption about the distribution (e.g., temporal trend not accounted for in experimental design)
- Real biological outlier – 1% chance of getting an outlier that is 3 standard deviations from the mean
- Large gaps in vegetation (i.e., soil exposed)

Osborne JW, Overbay A (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, 9(6).

Detection of phenotypic outliers

- Fit an initial linear model of fixed effects for each trait
  - Possible fixed effects to include in model: Cultivar (Line), Treatment, Environment, Replicate, Day of Year, Time of Day, and their interactions

- Detection of outliers with one of several possible methods:
  - Cook's distance measures influence of a data point. Data points that substantially change effect estimates.
  - Studentized deleted residuals measures leverage of a data point. Data points that affect least squares fit.
  - DFFITS criterion measures the influence of each observation on predicted values.

- Repeat model fitting until no statistical outliers are detected.

Belsley DA, Kuh E, Welsch RE (2004). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley-Interscience, Hoboken, NJ, USA.
Cook RD, and Weisberg S (1982) Residuals and Influence in Regression, Chapman & Hall, New York, NY, USA.
Kutner MH, Nachtsheim CJ, Neter J, Li W (2004). Applied Linear Statistical Models, 4th ed. McGraw-Hill, Boston, MA, USA.

Analysis of Non-Normal Trait Data

- When fitting linear models, two very important assumptions are that the error terms follow a normal distribution and that there is a constant variance.
- When data are non-normal, these two assumptions in particular could be violated.
- Generalized linear mixed models can be used to analyze non-normal data.
- The Box-Cox procedure can be used to find the most appropriate transformation that corrects for non-normality of the error terms and unequal variances.

McCullagh P and Nelder JA (1989). Generalized Linear Models, Second Edition. Chapman & Hall, New York, NY, USA.
Box GEP, Cox DR (1964). An analysis of Transformations. J. Roy. Stat. Soc. B. Met. 26: 211–252.

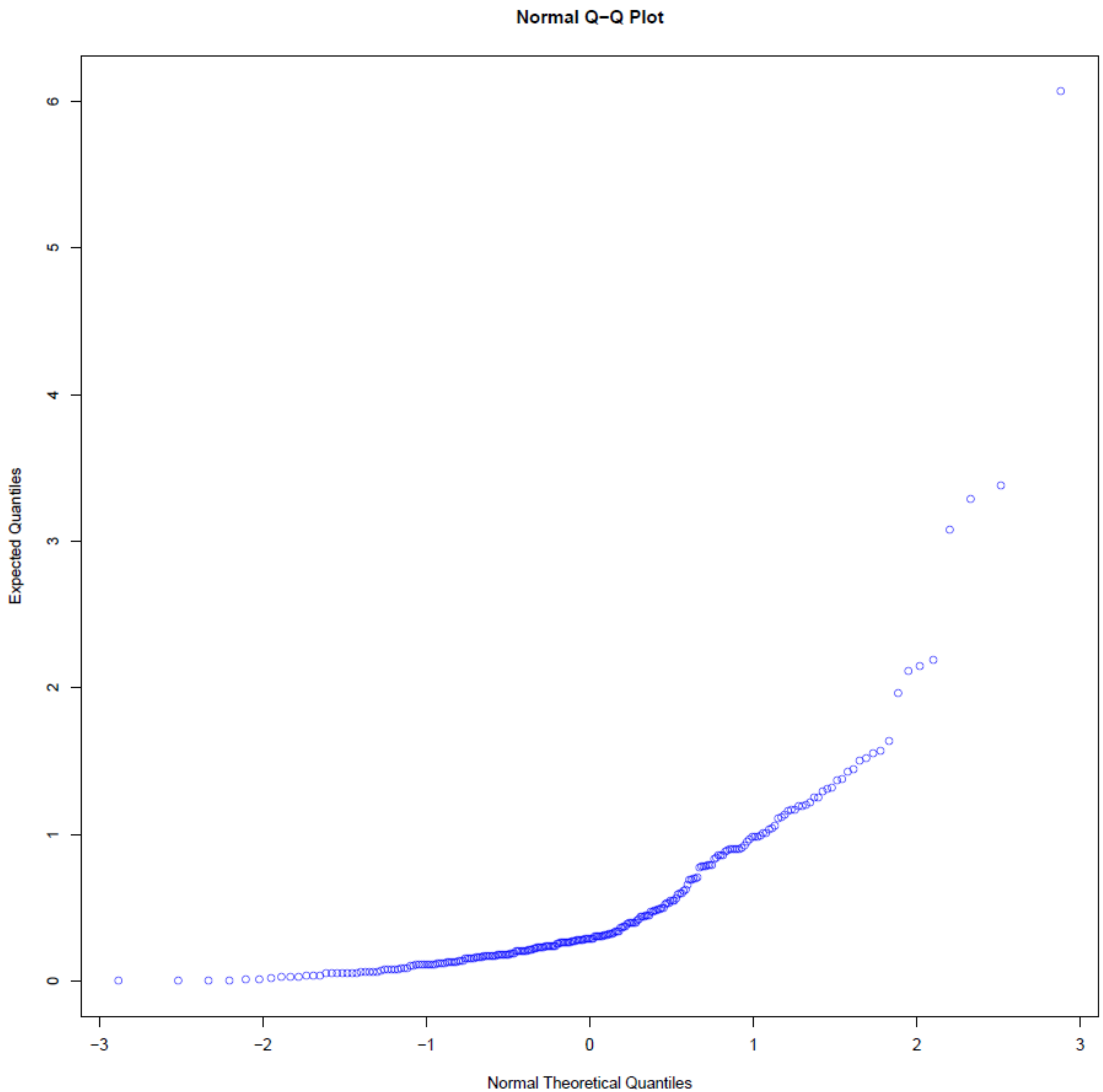
Box-Cox Transformation

Common Box-Cox Transformations

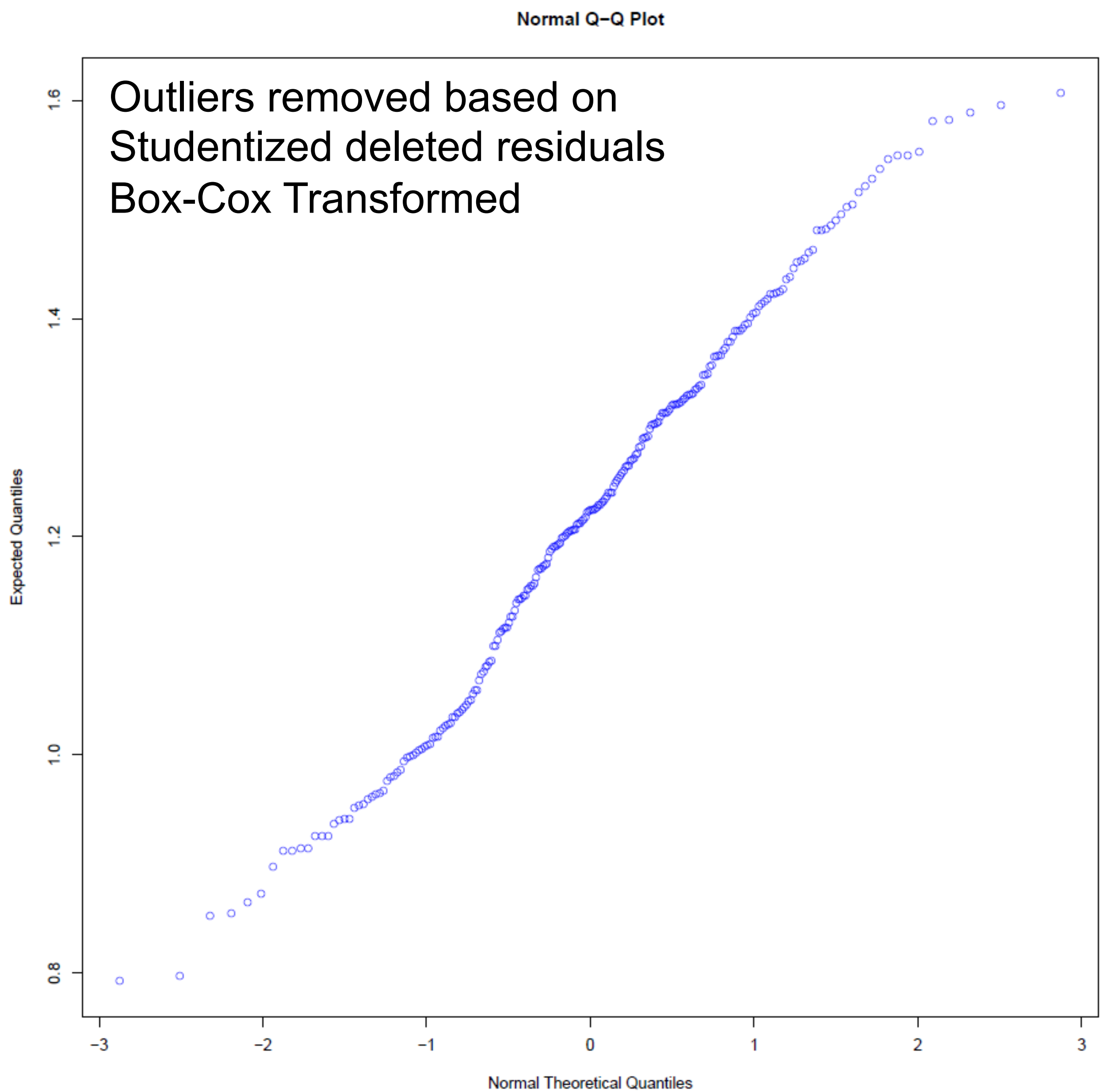
l	Y'
-2	$Y^{-2} = 1/Y^2$
-1	$Y^{-1} = 1/Y^1$
-0.5	$Y^{-0.5} = 1/(Sqrt(Y))$
0	$log(Y)$
0.5	$Y^{0.5} = Sqrt(Y)$
1	$Y^1 = Y$
2	Y^2

http://www.isixsigma.com/tools-templates/normality/making-data-normal-using-box-cox-power-transformation/

Q-Q-plot (normal)



Q-Q-plot (normal)



Mixed Linear Models

$$\mathbf{y} = \boldsymbol{X\beta} + \boldsymbol{Z}\mathbf{u} + \mathbf{e}$$

- **y** is a vector of phenotypic observations, with mean E(**y**) = $\boldsymbol{X\beta}$
- **β** is a vector of fixed effects
- **u** is a vector of random effects with mean vector E (**u**) = 0 and variance-covariance matrix var(**u**) = **G**
- **e** is a vector of error terms with mean vector E (**e**) = 0 and variance var (**e**) = I * σ^2 (I = identity matrix; σ^2 = residual variance)
- **X** and **Z** are incidence matrices of 1s and 0s relating **y** to **β** and **u**, respectively.

Henderson CR (2004). Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, Ontario, Canada.

Mixed Linear Models

- Calculate Best linear unbiased estimators (BLUEs) of fixed effects and Best linear unbiased predictors (BLUPs) of random effects with mixed models in SAS, R, or ASReml

e.g., Mixed effects model to predict BLUPs for each line:
$Y_{ijklmno} = \mu + env_i + field(env)_{ij} + rep(field{*}env)_{ijk} + block(rep{*}field{*}env)_{ijkl} + row(field{*}env)_{ijm} + column(field{*}env)_{ijn} + line_o + env{*}line_{io} + \varepsilon_{ijklmno}$

e.g., Mixed effects model to predict BLUEs for fixed effects:
$Y_{ijklmn} = \mu + line_i + \mathrm{treatment}_j + line{*}treatment_{ij} + rep(treatment)_{jk} + block(rep{*}treatment)_{jkl} + row_m + column_n + \varepsilon_{ijklmn}$

Least square means are obtained for fixed effects: line, treatment, and line*treatment

Littell RC, Milliken GA, Stroup WW, Wolfinger R (1996). SAS System for Mixed Models. SAS Publishing, Cary, NC, USA

Mixed Linear Models

- Outlier removed, transformed data should be averaged. Then, plot mean values analyzed with a weighted mixed model, with the number observations per plot for weighting.
- Fit a separate spatially autoregressive correlations in the row and column directions among plot residuals (AR1×AR1 error structure) for each environment.
- Model the correlation structure of repeated measurements of a trait on the same experimental plots over a period of time with a variance-covariance matrix.
- Conduct likelihood ratio tests to remove all terms from the model that were not significant at $\alpha = 0.05$.
- Final model is used to estimate BLUPs for each line and to estimate variance components

Hung H-Y, Browne CJ, Guill KE, Coles N, Eller M, Garcia A et al (2011). The relationship between parental genetic or phenotypic divergence and progeny variation in the maize Nested Association Mapping population. Heredity 108:490-499.

Estimate Heritabilities and Correlations

e.g., Mixed effects model to predict BLUPs for each line:
$Y_{ijklmno} = \mu + env_i + field(env)_{ij} + rep(field*env)_{ijk} + block(rep*field*env)_{ijkl} + row(field*env)_{ijm} + column(field*env)_{ijn} + line_o + env*line_{io} + \varepsilon_{ijklmno}$

- Variance components are used to estimate broad-sense heritability on an entry-mean basis (H^2) or repeatability. – proportion of the phenotypic variability attributed to genetics
- Estimate phenotypic and genotypic correlations between BLUPs of traits with multivariate mixed linear models - important for selection in breeding populations
- Standard errors of heritabilities and correlations can be derived with the delta method

Holland JB, Nyguist WE, Cervantes-Martınez CT (2003). Estimating and interpreting heritability for plant breeding: an update. In Janick J (ed) Plant Breeding Reviews. John Wiley and Sons: Hoboken, New Jersey, USA

Statistical Genetics

- Quantitative trait locus (QTL) analysis: univariate and multivariate (e.g., ICIM and QTL cartographer software)
- Genome-wide association study (GWAS): univariate and multivariate (e.g., GAPIT, MLMM, MTMM software)
- Genomic Selection (GS): univariate and multivariate (e.g., rrBLUP, BayesCπ software)
- Consider integrating plant growth models for QTL mapping, GWAS, and GS

Papers on these topics for further reading:

Endelman JB (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Gen 4: 250-255.

Jia Y, Jannink J-L (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192: 1513-1522.

Jiang C, Zeng ZB (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140: 1111-1127.

Korte A, Vilhjalmsson BJ, Segura V, Platt A, Long Q, Nordborg M (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet 44: 1066-1071.

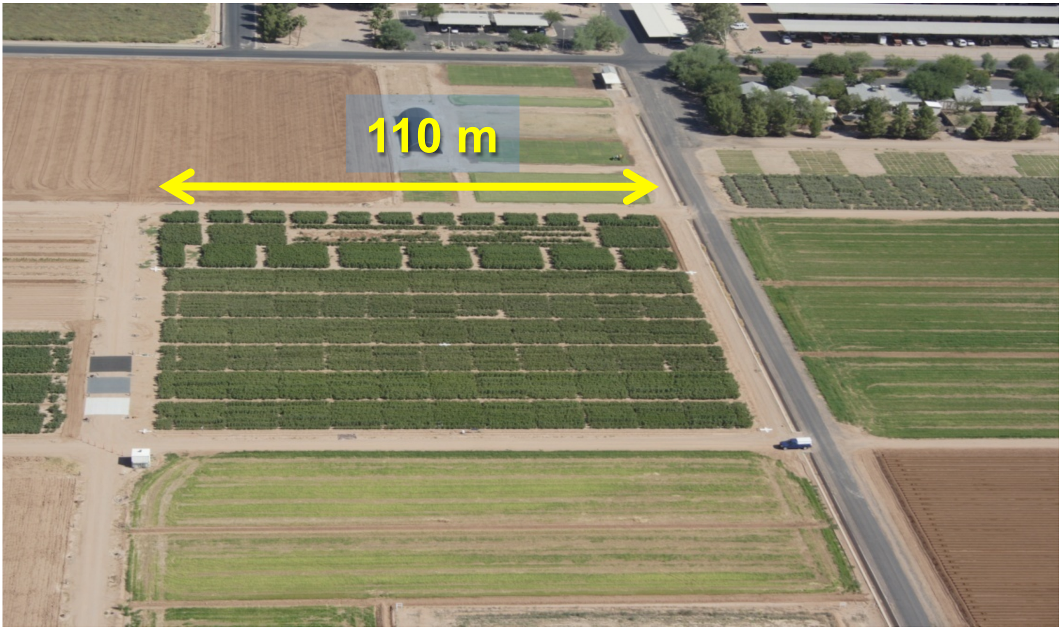
Li H, Ye G, Wang J (2007). A modified algorithm for the improvement of composite interval mapping. Genetics 175: 361-374.

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ et al (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics 28: 2397-2399.

Segura V, Vilhjalmsson BJ, Platt A, Korte A, Seren U, Long Q et al (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44: 825-830.

Wu W-R, Li W-M, Tang D-Z, Lu H-R, Worland AJ (1999). Time-related mapping of quantitative trait loci underlying tiller number in rice. Genetics 151: 297-303.

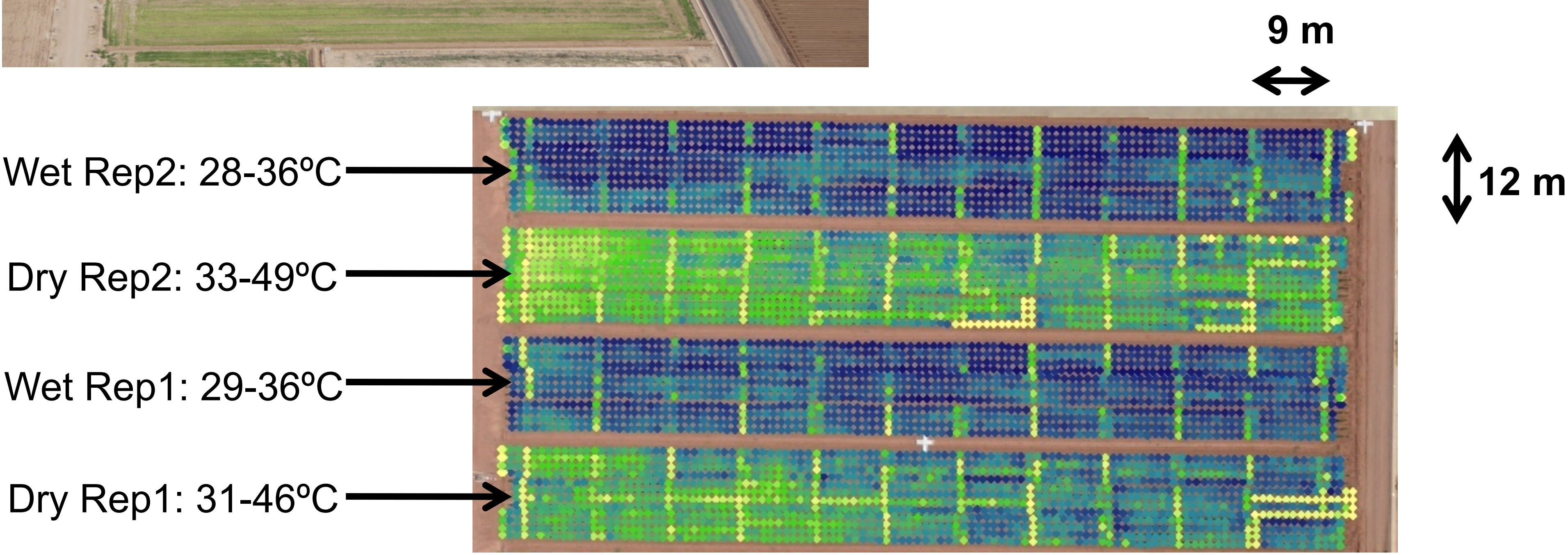
HTP: Canopy temperature



Central Arizona: clear skies, very limited rain, high temperatures

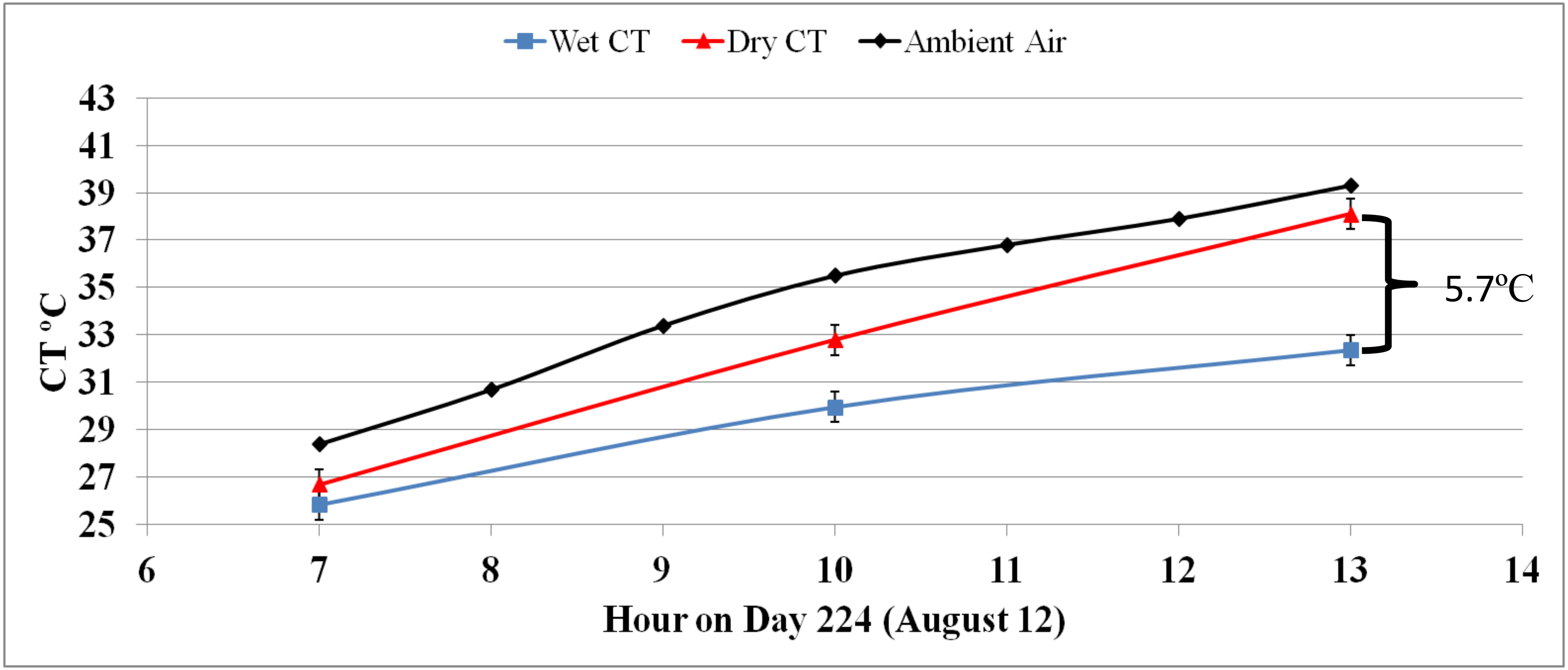
TM-1×NM24016 population: 94 RILs (*Gossypium hirsutum*)

Treatments: 100 and 50% ET (2 reps) by drip irrigation



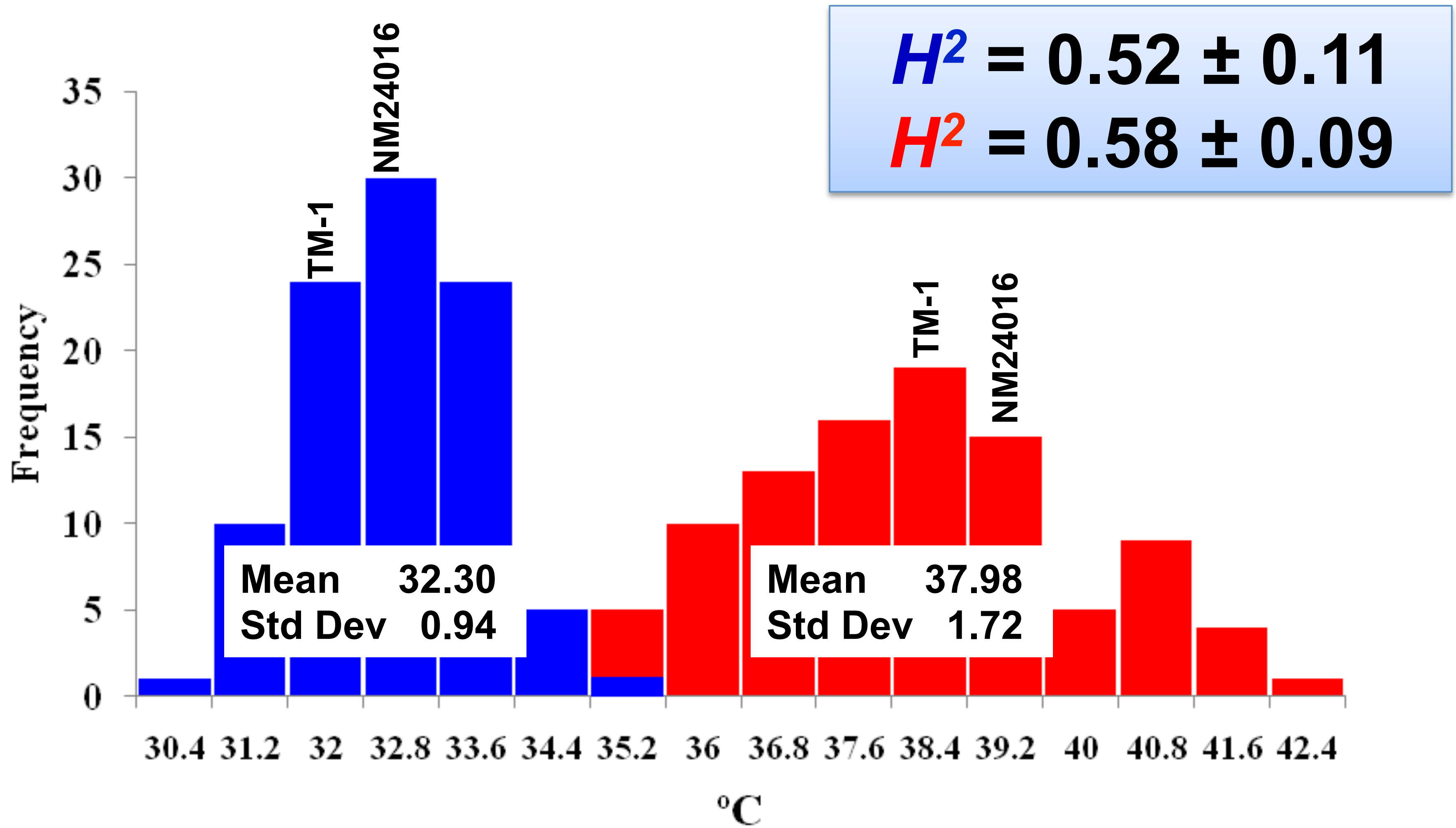
1pm MST on day 224 (12-Aug)

Significant time-by-treatment interaction for canopy temperature



Treatment	*P*<0.05
Time	*P*<0.0001
Treatment*Time	*P*<0.0001

Phenotypic variation: Canopy Tm



Wet and Dry Plots at 1 pm on Day 224

GBS-QTL Analysis: Dry/Wet Canopy Tm

Chr_LG_cM	c10_38_5	c12_51_18	c17_68_14	c20_83_18
Day 217 7 am	-0.006	0.003	-0.005	-0.001
Day 217 1 pm	**-0.019**	0.008	**-0.016**	**0.015**
Day 224 7 am	-0.004	0.005	-0.004	0.002
Day 224 10 am	-0.011	0.010	-0.009	0.009
Day 224 1 pm	**-0.020**	**0.018**	-0.010	**0.020**

Negative – decrease Tm   **Positive** – increase Tm

All statistically significant allelic effects are indicted by a black rectangle border

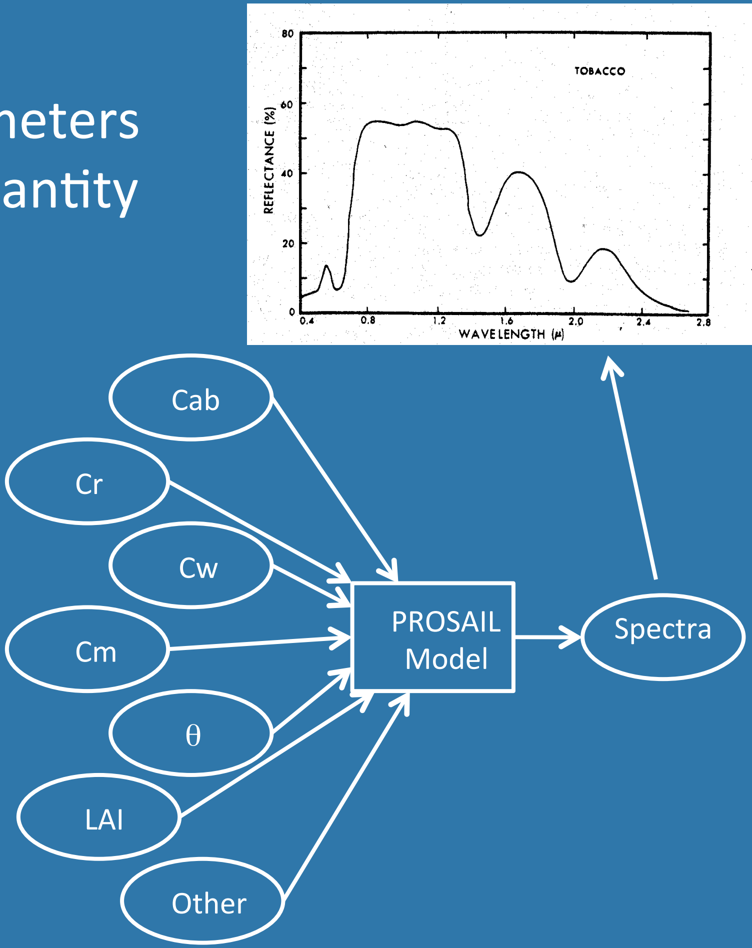
Considered identical QTL if support intervals overlapped

Inverse Modeling Introduction

- Remote sensing tools for crop phenotyping at USDA-ALARC
  - Radiometers - reflectance of radiation from crop canopy
  - Infrared thermometers - temperature of crop canopy
  - SONAR - distance from crop canopy to the sensor
- Purpose of Inverse Modeling
  - Can we increase or extend the usefulness the phenotypic data?
  - Is there a role for physical or ecophysiological models?
  - Can these models be <u>inverted</u> to estimate additional phenotypes?

Inverse Modeling Example

- Typical model design
  - Provide system attributes as input parameters
  - Model will simulate some observable quantity
  - PROSAIL example
- Forward modeling
  - System attributes (inputs) are known
  - Use them to calculate model outputs
  - ‘Plug and chug’
- Inverse modeling
  - Model outputs are known
  - Use to estimate system attributes
  - Mathematically invert the model?
  - If not, use optimization algorithm



Optimization Procedural Questions

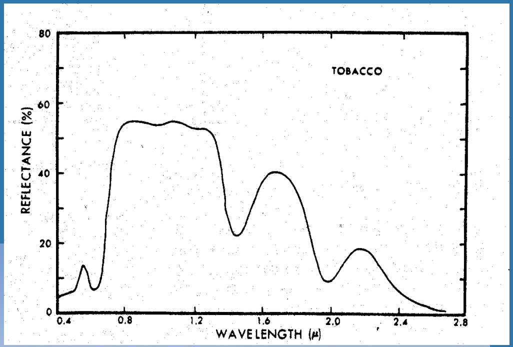
- What model parameters to adjust?
- What bounds for adjusted parameters?
- What model parameters to constrain? How?
- What spectral wavelengths to include?
- What observation dates to use?
- What objective function to minimize or maximize?
- All of these will affect optimization outcome

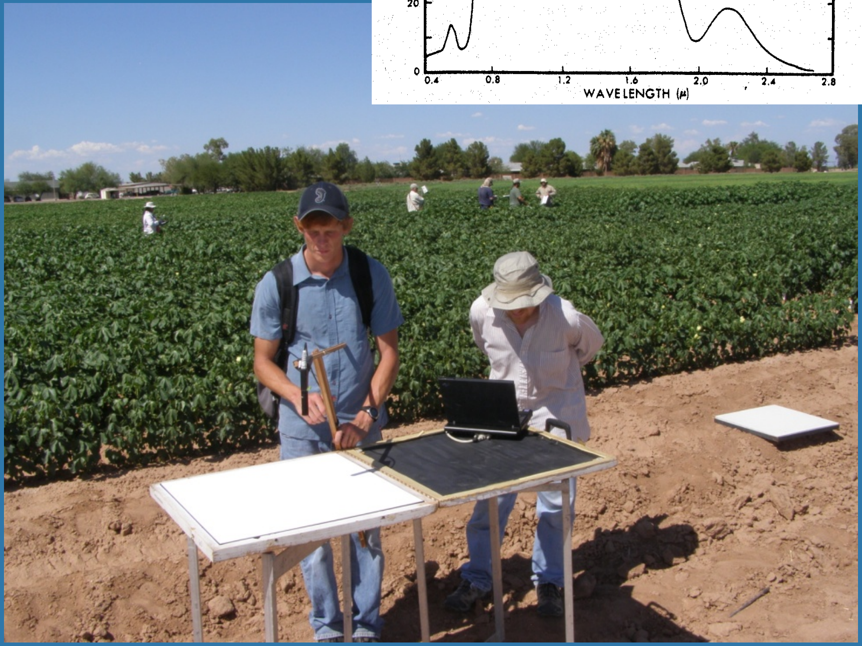
Basic Methodology

- Conduct field experiments
  - Cotton: Cultivar by Water (3 years)
  - Collect hyperspectral crop canopy reflectance information
- Build computational framework
  - Implement a radiative transfer model (PROSAIL)
  - Implement an optimization algorithm (PEST)
  - Implement control scripts (Python)
- Conduct inverse modeling (optimization) experiments
  - Evaluate alternative procedural strategies
  - Compare measured versus simulated (optimized) results
  - Test statistical significance of optimized model parameters (traits)

Cotton Phenotyping Experiments

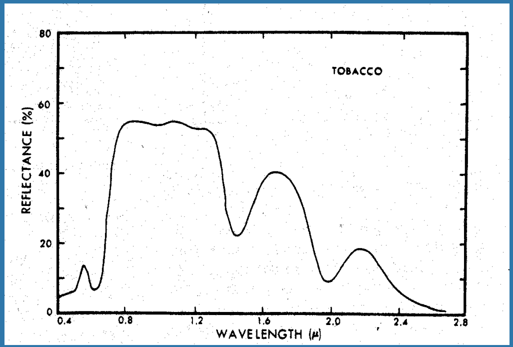
- Experimental Design
  - Maricopa, Arizona in 2010, 2011, and 2012
  - 25 extra-long staple cotton cultivars
  - Well-watered and water-limited irrigation
  - 4 replications
- Observations
  - August 4, 2010 AM
  - August 4, 2010 PM
  - August 18, 2011 AM
  - Hyperspectral remote sensing
    - ASD FieldSpec Pro
    - 350 to 2500 nm @ 1 nm bandwidth
  - Leaf punches
    - Leaf water content
    - Specific leaf area
    - Chlorophyll a & b content





PROSAIL Model

- PROSAIL
  - PROSPECT leaf optical properties model
  - SAIL canopy bidirectional reflectance model
- 14 input parameters
  - Leaf pigment content: C_{ab}, C_{cr}, C_{bp}
  - Leaf water content: C_w
  - Canopy architecture: C_m, N, LAI, Θ_l
  - Solar geometry: solar zenith and azimuth, viewer zenith
  - Soil background
  - Hot spot size
- Output
  - Canopy bidirectional reflectance
  - 400 to 2500 nm in 1 nm increments

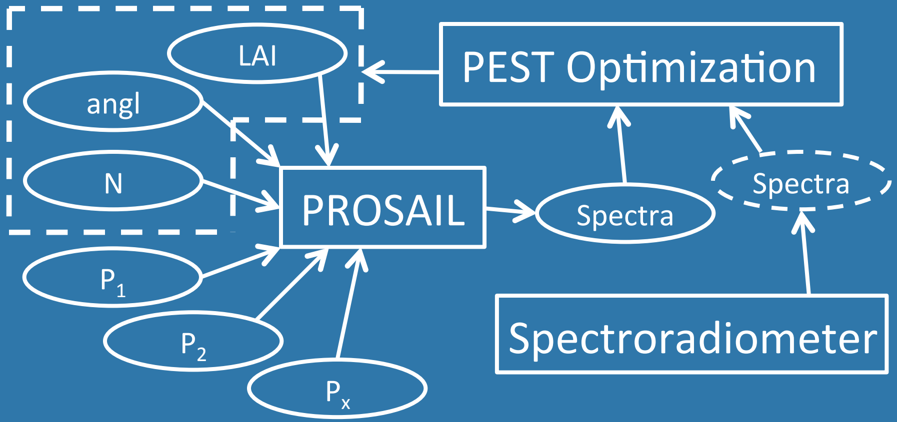


PROSAIL Model Parameterization

- Parameters
  - Cab – Chlorophyll a & b (μg cm^{-2}) – observed
  - Car – Carotenoids (μg cm^{-2}) – fixed @ 5 μg cm^{-2}
  - Cbrown – Brown pigment (unitless) – fixed at 0.0
  - Cw – Leaf water thickness (cm) – observed
  - Cm – Specific leaf area (g cm^{-2}) - observed
  - N – Structural parameter (unitless) – optimized
  - LAI – Leaf area index (unitless) – optimized
  - angl - Leaf inclination angle (degrees) – optimized
  - psoil – Soil coefficient (unitless) – fixed at 1.0
  - skyl – diffuse light (%) – observed
  - hspot – hot spot size (unitless) – fixed at 1.0
  - tts – Solar zenith angle (degrees) – computed from time stamp
  - tto – Observer zenith angle (degrees) – fixed at 0°
  - psi – Observer azimuth angle (degrees) – fixed at 0°

PROSAIL Model Inversion

- Optimization protocol
  - Parameter ESTimation (PEST)
  - Adjust
    - N structural coefficient
    - Leaf area index (LAI)
    - Leaf inclination angle
  - Minimize error between measured and simulated canopy spectral reflectance
    - 1703 wavebands
    - 400 to 1350 nm
    - 1450 to 1770 nm
    - 1970 to 2400 nm



August 4, 2010 AM Results

- ANOVA & Tukey test
- p = 0.05
- Main Treatment
  - LAI
- Cultivar
  - Structural parameter
  - LAI

	N	LAI	angl
Treatment			
Well Watered	1.37 a	**3.99 a**	56.98 a
Water Limited	1.40 a	**4.71 b**	53.83 a
Cultivar			
Pima S7	**1.07 a**	**4.94 a**	52.56 a
Pima S6	**1.24 a**	**4.85 a**	52.64 a
Pima 32	**1.58 b**	**4.29 ab**	59.14 a
Monseratt	**1.51 b**	**4.19 b**	58.33 a

August 4, 2010 PM Results

- ANOVA & Tukey test
- p = 0.05
- Main Treatment
  - LAI
  - Inclination angle
- Cultivar
  - Structural parameter
  - Inclination angle

	N	LAI	angl
Treatment			
Well Watered	1.50 a	**6.48 a**	**63.08 a**
Water Limited	1.48 a	**6.04 b**	**65.11 b**
Cultivar			
Pima S7	**1.36 a**	6.29 a	**61.76 a**
Pima S6	**1.49 ab**	6.25 a	**62.67 a**
Pima 32	**1.56 b**	6.05 a	**65.23 b**
Monseratt	**1.49 ab**	6.33 a	**66.03 b**

August 18, 2011 AM Results

- ANOVA & Tukey test
- p = 0.05
- Main Treatment
  - Inclination angle
- Cultivar
  - LAI

	N	LAI	Angl
Treatment			
Well Watered	1.55 a	5.14 a	**34.84 a**
Water Limited	1.60 a	5.46 a	**43.24 b**
Cultivar			
Pima S6	1.57 a	**5.30 ab**	41.97 a
PSI 425	1.61 a	**4.88 ab**	39.47 a
P62	1.62 a	**4.17 a**	34.70 a
Monseratt	1.44 a	**6.60 b**	35.79 a
89590	1.62 a	**5.62 ab**	43.40 a

Inverse Modeling Conclusion

- Statistically significant trait estimates
- Results dependent on optimization settings
- Crop models may help constrain PROSAIL

Plant Breeding & Genomics

Please fill out the survey evaluation.

You will be contacted via email.

Today's Presentation and Links Available
http://www.extension.org/pages/68270

Sign up for PBG News
http://pbgworks.org

Sign up for Future Webinars and View Archive
http://www.extension.org/pages/60426

eXtension