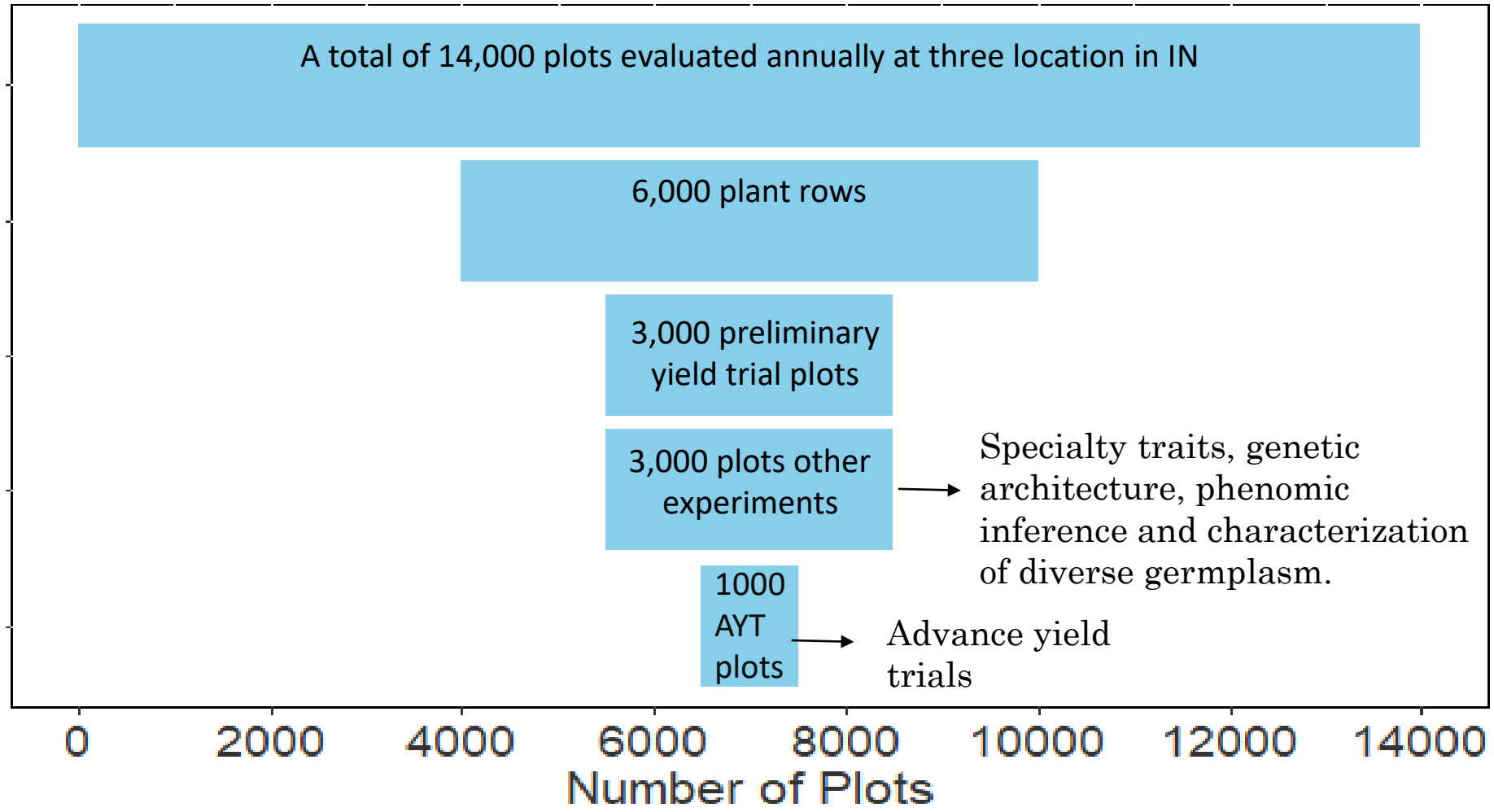
Managing UAS Imagery for Developmentally-Driven Decision Making and Genetic Analysis

2020 G2F GxE Collaborator's Meeting
Feb 24, 2020
Katy Martin Rainey

PURDUE UNIVERSITY | PHENOMICS

UAS: Unmanned Aerial Systems, aka drone.

The number of plots for multi-environment yield trials in the Rainey lab in 2019, a typical year.



PLANT ROWS: the first stage in soybean breeding that selected inbred plants are grown in rows/plots.

progeny

K.M. Rainey is co-founder of Progeny Drone, Inc.

Dr. Anthony Hearst (CEO) (Full Time)

PhD - Drone Image Processing & Software Development

Dr. Katy Rainey (CTO)

ahearst@progenydrone.com
https://www.progenydrone.com/

NSF | CORPS NSF Innovation Corps

The goals of field-based phenomics.

FOR SELECTION:

Predict yield *per se* or genetic yield potential.

- Using genetically correlated secondary traits.

Adjust your yield data with a covariate or other index.

FOR GENETICS:

Conduct growth analyses, find genes that control how the crop is growing.

Identify a new trait, or new phenotypes for a trait, or both.

FOR EITHER:

Replace ground-based phenotyping methods.

- Convert a visual score to a quantitative phenotype to obtain better precision, throughput or both.
- Two methods:
    1. Computer vision/AI/ML.
    2. An regression from growth analyses.

GENETIC CORRELATION, $|r_A|$ **:** in multivariate quantitative genetics, a genetic correlation is the proportion of variance that two traits share due to genetic causes; covariance of breeding values.

GENETIC YIELD POTENTIAL: the yield of optimal or ideal genotypes in a target environment, across major growing regions and environments

Define the Data You Need

Define your objectives.

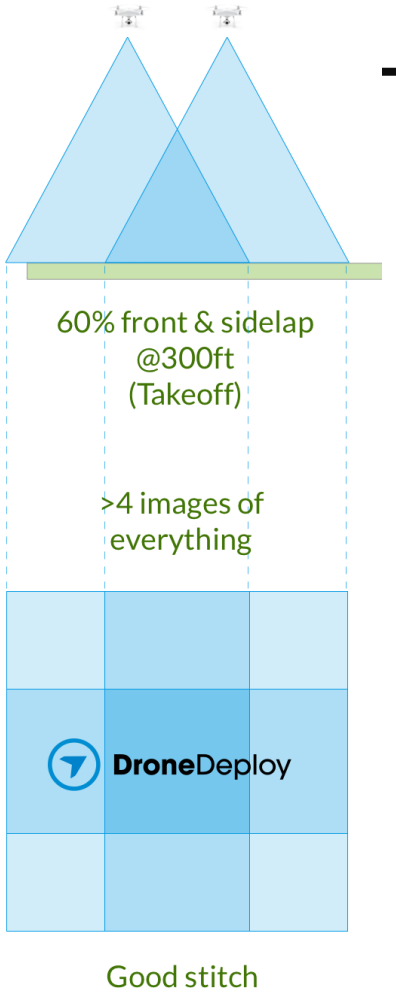
- Genetic analyses needs many observations.
- *Ergo* robust UAS platforms for row crops.

  - RGB: growth analyses, canopy features.
  - MS: color changes (i.e. senescence).
  - RGB + Thermal: yield prediction?

Define the relevant phenology.

Define <u>your</u> *minimal* spatial and temporal resolution.

Always at least 80% overlap.



RGB: red, green and blue; images from a typical camera.

MS: multispectral; sensor data quantifying reflectance of spectral bands.

PHENOLOGY: plant life cycle events influenced by seasonal and interannual variations in weather and climate.

LONGITUDINAL TRAITS: traits recorded multiple times during a season that have continuously-variable phenotypes.

*Implications for defining **your** minimal spatial and temporal resolution.*

There is a LOT of biology and genetics to be done on growth analyses of phenotypes from RGB images.

-Recognize data snobbery.

-Do Genetics Now!

Collect images and other data when developmentally relevant.

-This is *dynamic*.

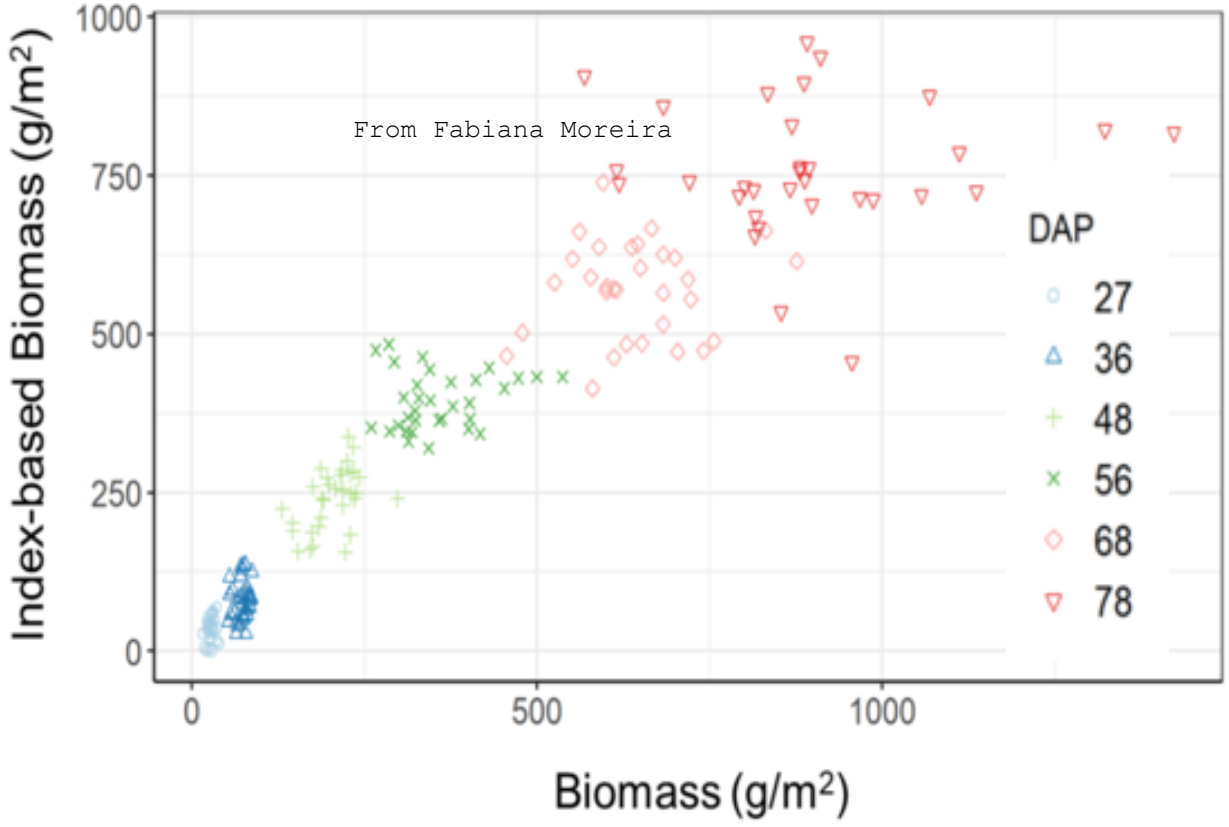
Each site and each trait has different requirements to get resolution.

-Not helpful to standardize platforms or data collection protocols.

Stand Counts: 1 cm/pixel

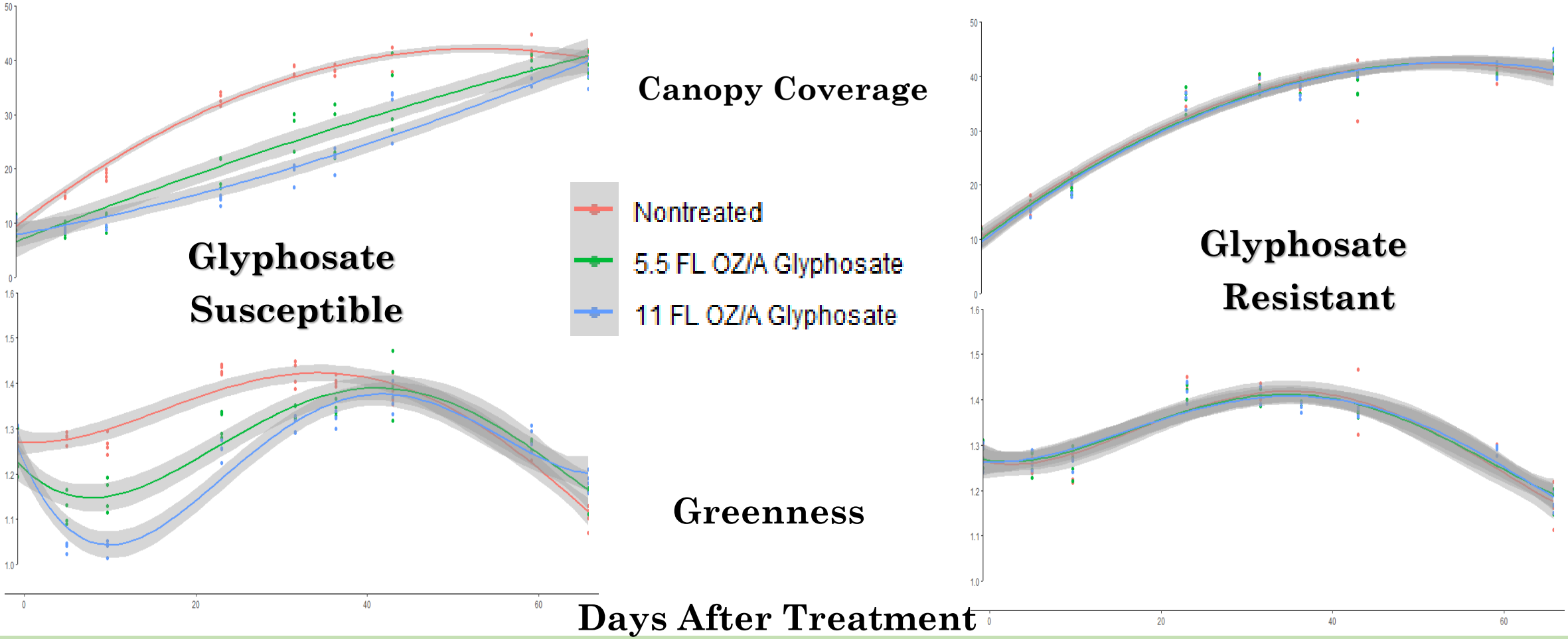
Canopy Cover: 4 cm/pixel

Biomass Accumulation: dynamic



Changes in soybean canopy coverage and greenness due to stress.

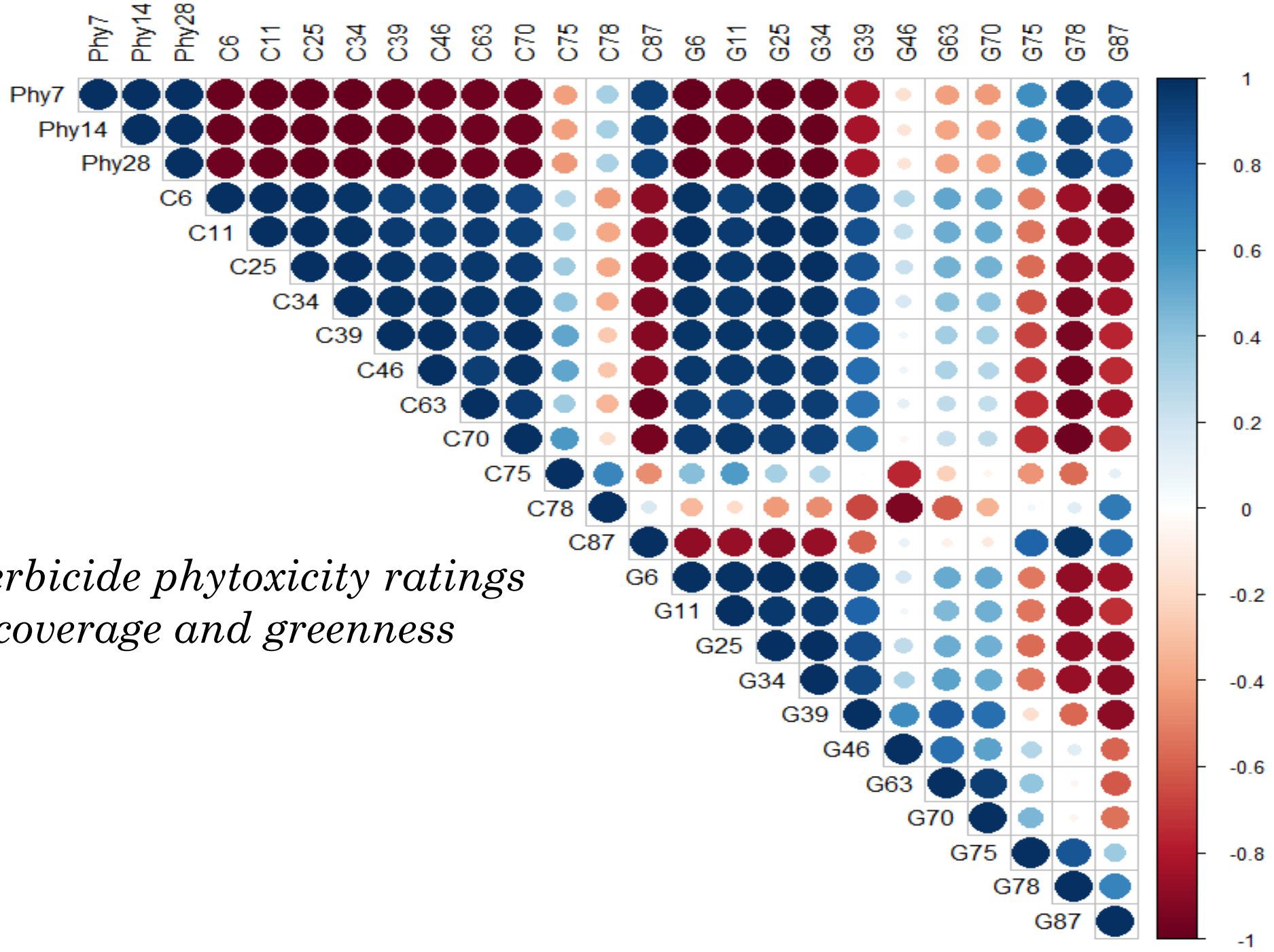
There is a LOT of biology and genetics to be done on growth analyses of phenotypes from RGB images.



CC/ACC: canopy coverage measured with UAS calculated as percentage pixels from canopy-segmented RGB plot clips/average canopy coverage is a phenotype derived from fitting CC observations to a model (logistic) and averaging the interpolated values.

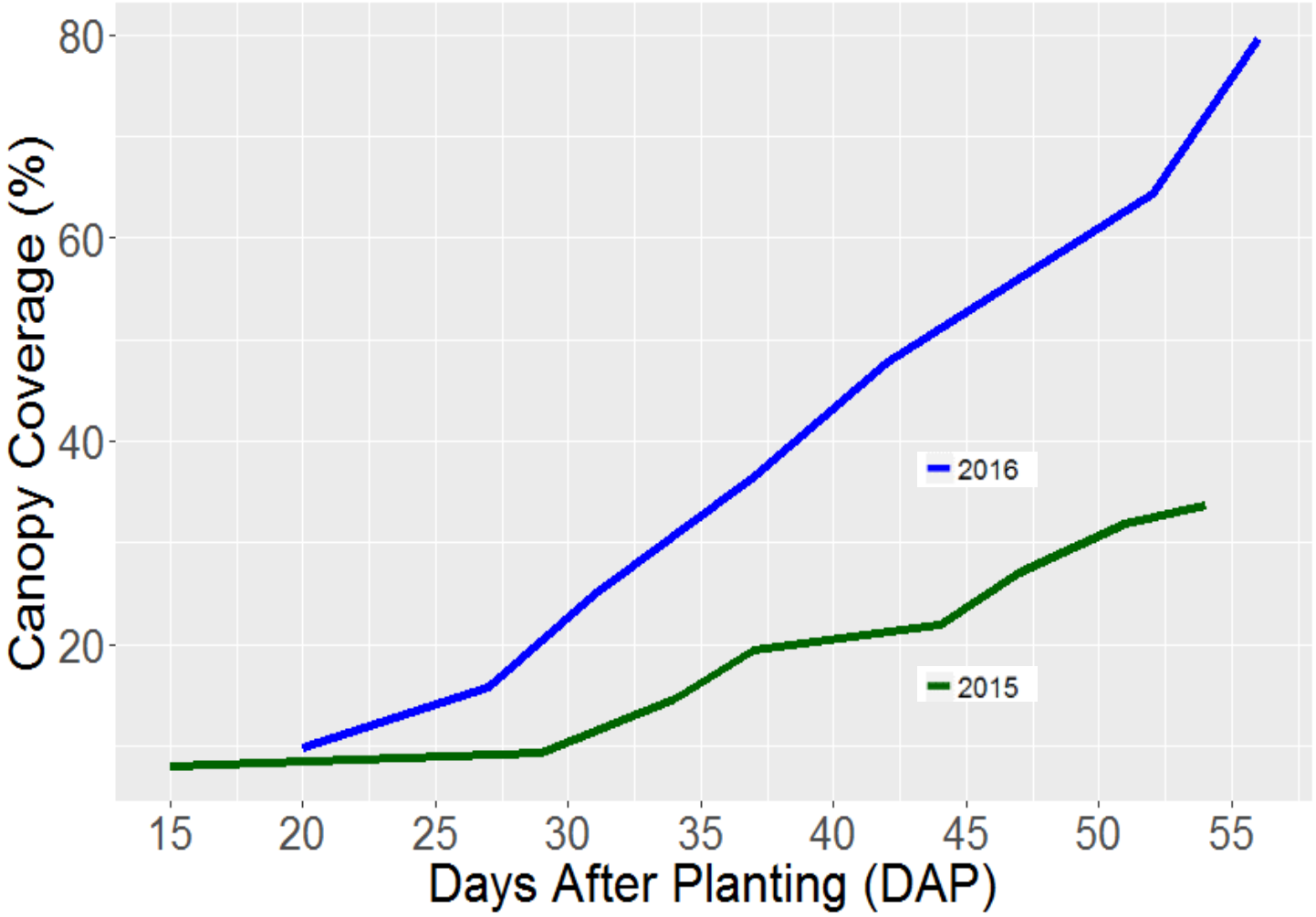
GREENNESS: green to red ratio.

Correlation of herbicide phytoxicity ratings with canopy coverage and greenness



Distribution of average canopy coverage of the checks by days after planting for progeny rows 2015 and 2016.

Collect images and other data when developmentally relevant.



How to describe/ prescribe a phenotyping protocol?

F.F Moreira, Hearst, A.A., Cherkauer, K.A. **& K.M. Rainey**. Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. *Plant Methods* **15**, 139 (2019). https://doi.org/10.1186/s13007-019-0519-4

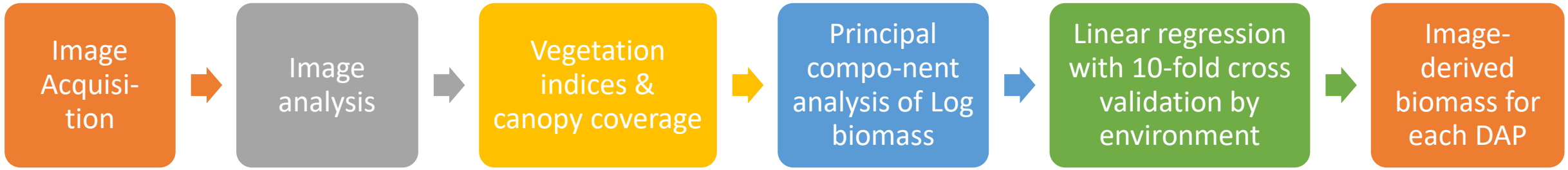
Ground Reference Data and Metadata

Do you need to develop a remote-sensing prediction equation?

- Predict ground reference phenotypes from remote data with cross-validation.
- You may need to calibrate both the spectral data and the prediction equation.

Example biomass estimation pipeline

From Fabiana Moreira

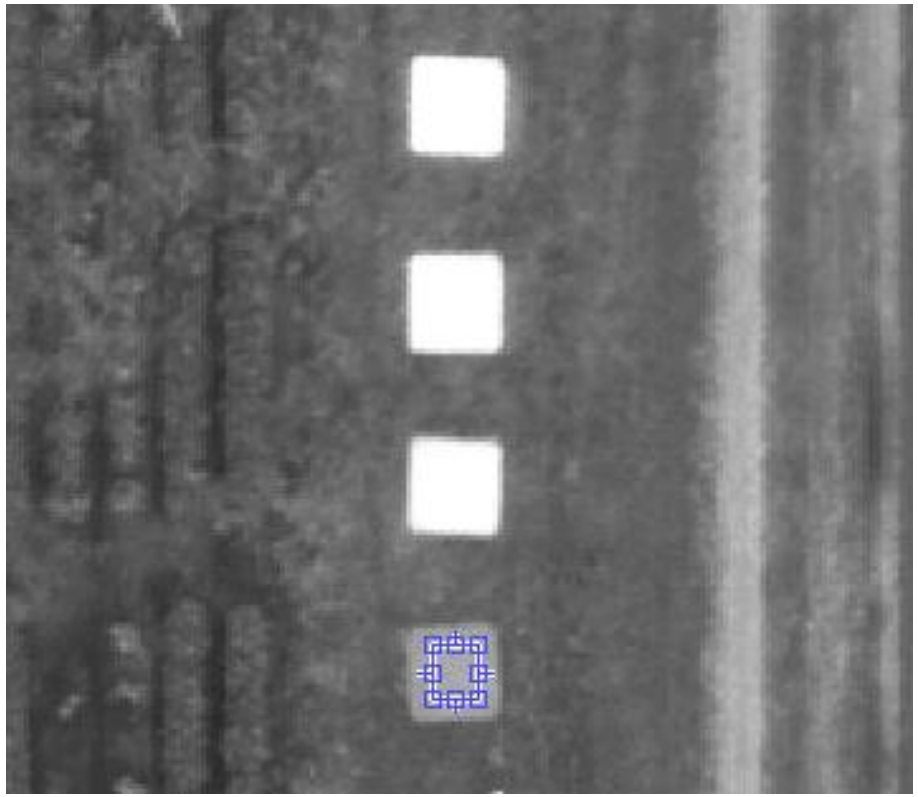


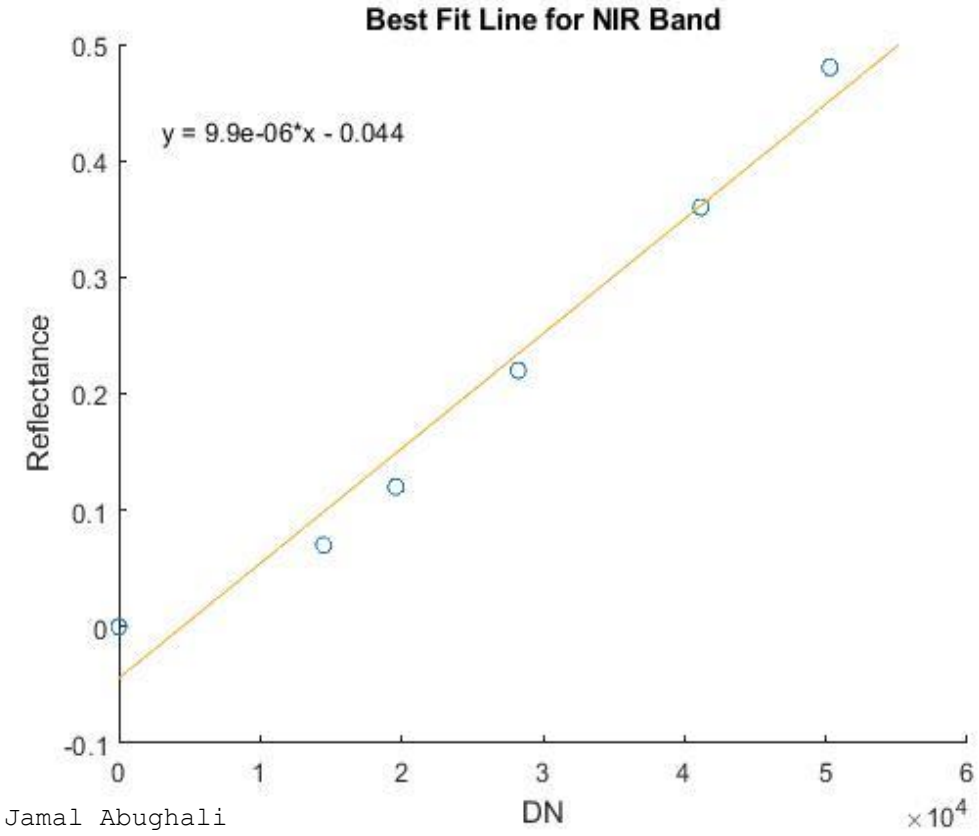
GROUND-REFERENCE DATA: ground observations of the phenotype to be estimated from remote-sensing, and metadata needed for calibration of *prediction equations*.

Data Type	Description	Format	Short-term Storage
Supplementary ground-reference data.	Includes the location of all ground control points and spectral targets. Field spectrometer + thermal target data.	Varies, but mostly ASCII text data files. Size: < 10 GB	Field computer.

Ground Reference Data and Metadata

Do you need to calibrate your spectral data for reflectance?





From Bilal Jamal Abughali

SPECTRAL REFLECTORS, OTHER TARGETS OR REFLECTORS: ground features and fixed points visible in images for calibration of reflectance, color, temperature, height.

REFLECTANCE PANELS: reflect at a specific and consistent percentage of light across the Visible and near Infra-red spectrum.

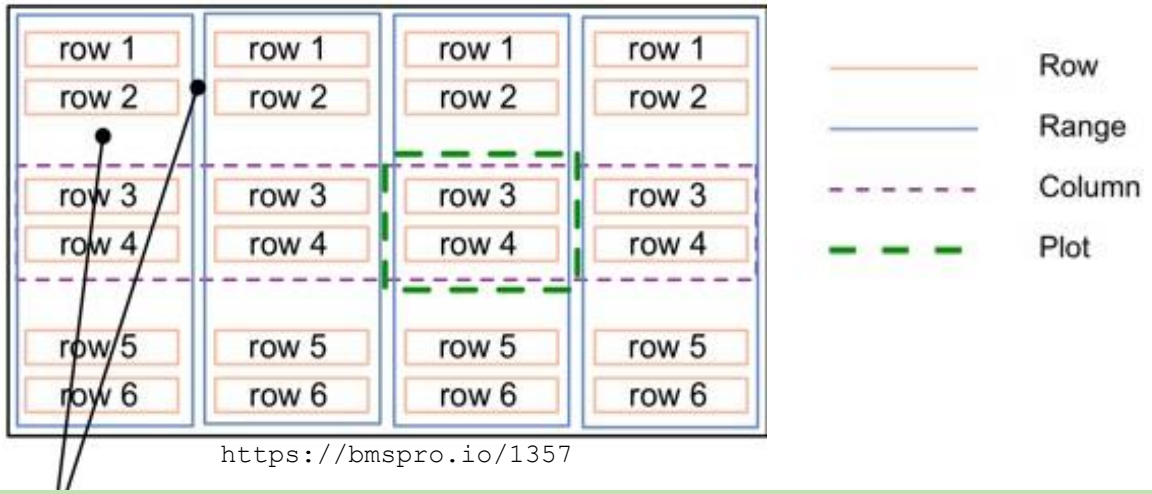
Ground Reference Data and Metadata

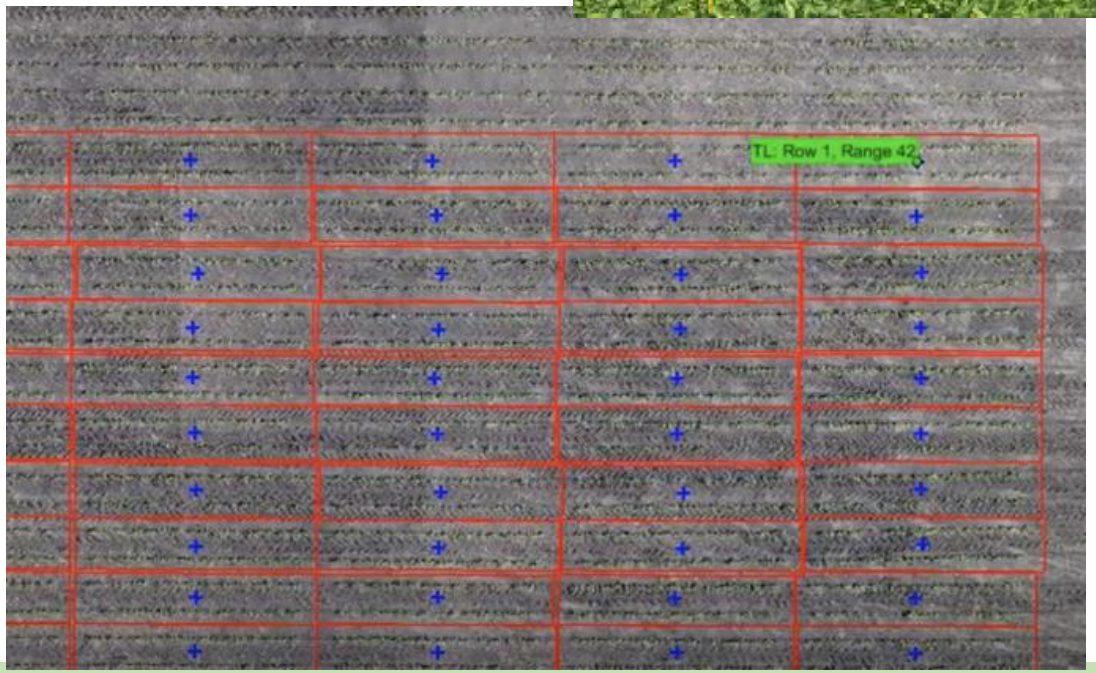
Do you need georeferencing?

- **Probably not**. Just use range, row coordinates (needed for GPS planters).
- Experiments blocked in field by UAS objective to minimize flight number.
- Merge all data from range, row coordinates. Include borders.

Minimum features and metadata:

- GPS coordinates of field corners and/or blocks (arranged by plot size).
- Visible flagging or feature in first and last plot.





GROUND CONTROL POINTS (GCPS): ground features visible in images that provide fixed points for geo-referencing.

Flight Naming Convention

Example 1: Flight

- DateofFlight_PlatformCameraLocationExperiment
  190709_eBSODAACRESoy

Example 2: Pre-processing

- DateofFlight_ PlatformCameraLocnExp_PreprocessingDate_Initials
  190709_eBSODAACRESoy_191013_xx

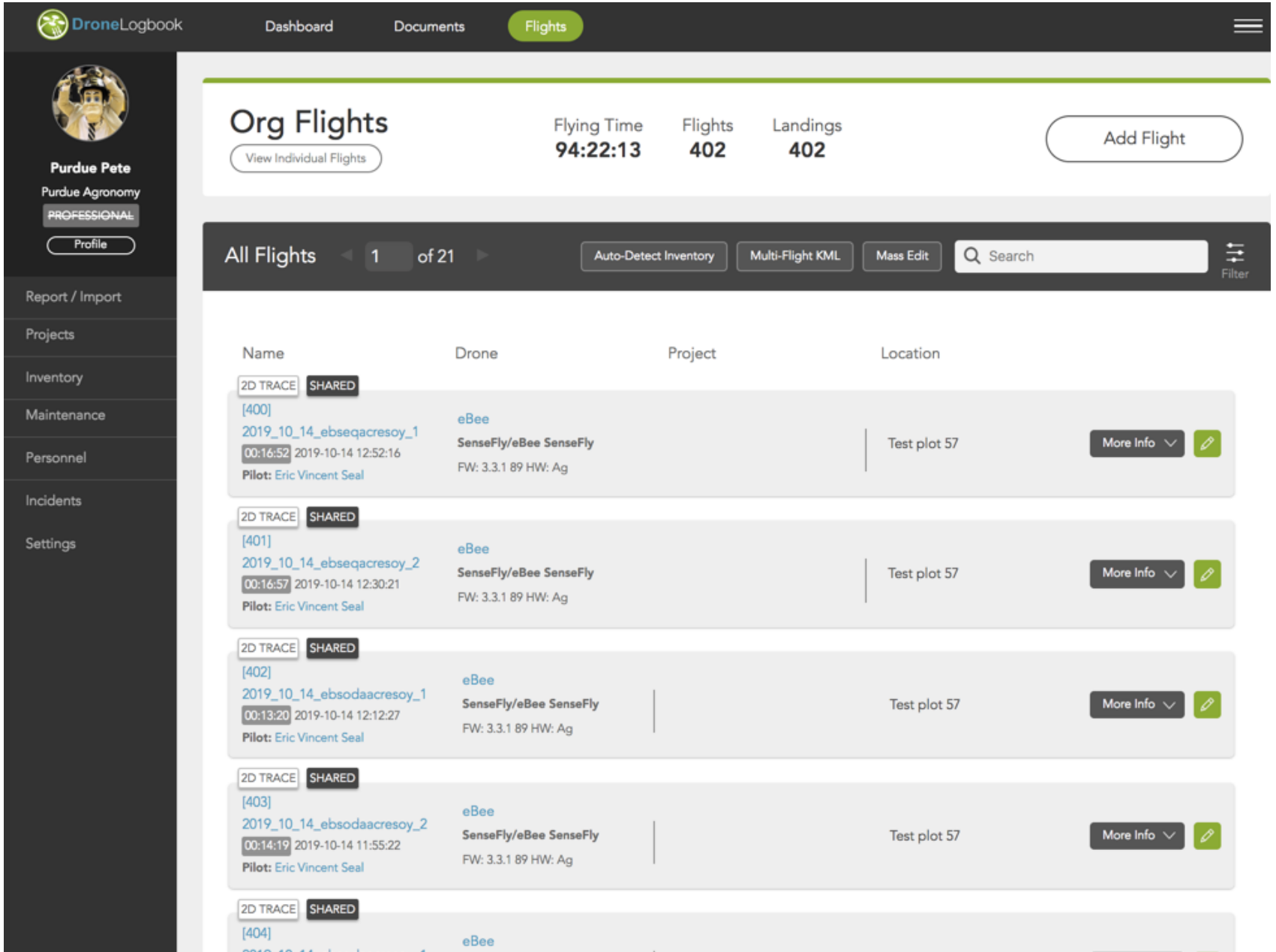
Pre-processing is a recursive process where multiple products may be generated to produce the most useful product for use downstream. Having a timestamp of preprocessing is helpful to compare products generated on different dates.

PLATFORM: UAS make and model.

CAMERA: Camera make and model.

FRAME CAMERA: A given area is instantaneously imaged.

Free Flight Data Management with Drone Logbook



Free Flight Data Management with Drone Logbook

Date	Flight name	Drone	Duration	Location
2019-10-14 12:52:16 16:52:16 UTC	**2019_10_14_ebseqacresoy_1**	eBee SenseFly/eBee SenseFly (**FW:**3.3.1 89 **HW:**Ag)	00:16:52	Test plot 57 (40.48023154518763, -86.99382681436384)

Landing Time:17:09:08 UTC **Flight Type:**

Operation Type: VLOS **Sunset / Sunrise:** Day
Personnel: Eric Vincent Seal [Pilot] **Pilot info:**
Equipment onboard: Sequoia
Nb landing: 1 **Distance:** 11505 m **Max altitude:** 48 m
Conditions:
Cloud cover: 0 % **Temperature:** 52 F **Wind:** 8.2 miles/hour (259°) **Humidity:** 50 %
Notes:

IGC File | KML File

Date	Flight name	Drone	Duration	Location
2019-10-14 12:30:21 16:30:21 UTC	**2019_10_14_ebseqacresoy_2**	eBee SenseFly/eBee SenseFly (**FW:**3.3.1 89 **HW:**Ag)	00:16:57	Test plot 57 (40.48023154518763, -86.99382681436384)

Landing Time:16:47:18 UTC **Flight Type:**

Operation Type: VLOS **Sunset / Sunrise:** Day
Personnel: Eric Vincent Seal [Pilot] **Pilot info:**
Equipment onboard: Sequoia
Nb landing: 1 **Distance:** 11916 m **Max altitude:** 49 m
Conditions:
Cloud cover: 0 % **Temperature:** 52 F **Wind:** 7.24 miles/hour (288°) **Humidity:** 51 %
Notes:

IGC File | KML File

Date	Flight name	Drone	Duration	Location
2019-10-14 12:12:27 16:12:27 UTC	**2019_10_14_ebsodaacresoy_1**	eBee SenseFly/eBee SenseFly (**FW:**3.3.1 89 **HW:**Ag)	00:13:20	Test plot 57 (40.48023154518763, -86.99382681436384)

Landing Time:16:25:47 UTC **Flight Type:**

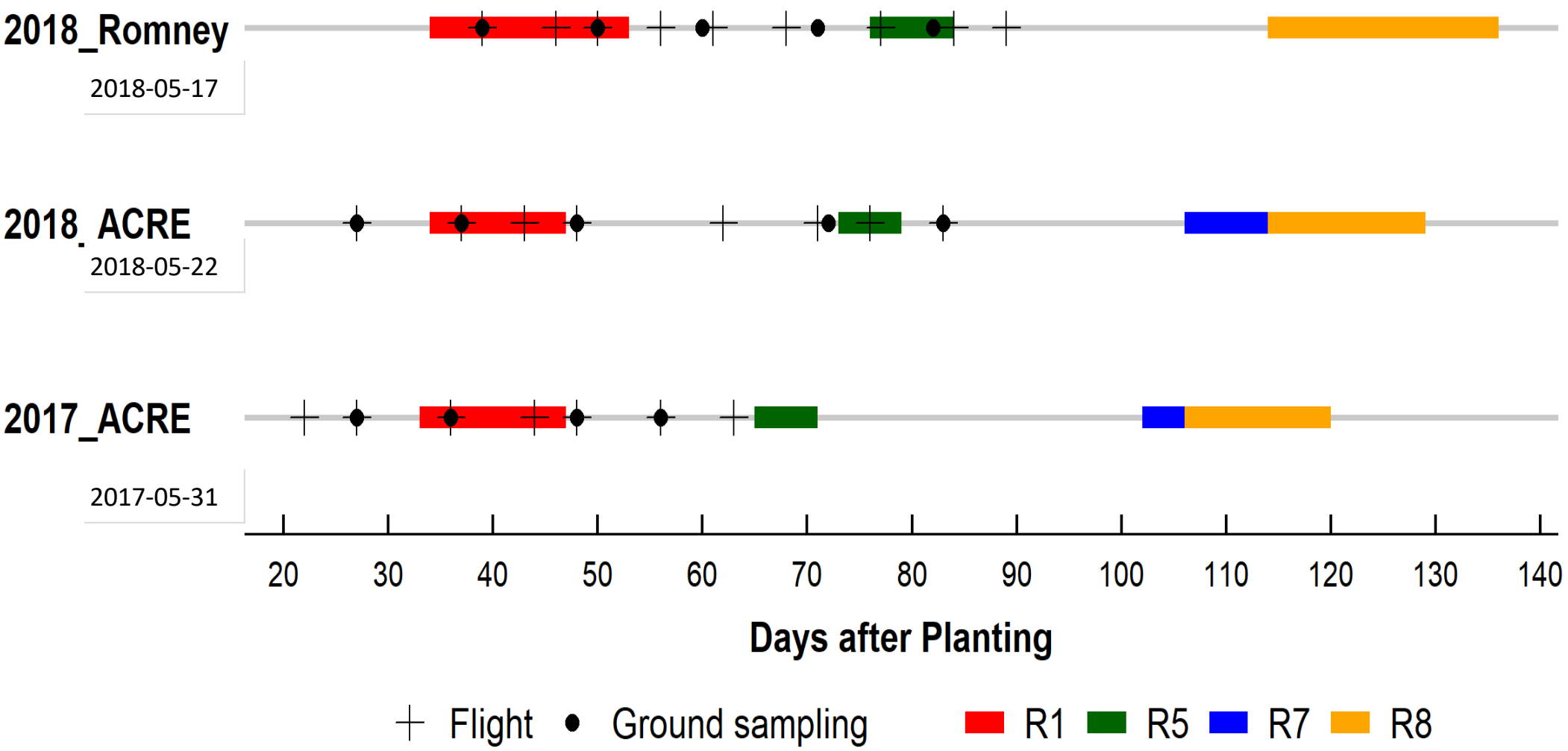
Operation Type: VLOS **Sunset / Sunrise:** Day
Personnel: Eric Vincent Seal [Pilot] **Pilot info:**
Equipment onboard: S.O.D.A.
Nb landing: 1 **Distance:** 9236 m **Max altitude:** 45 m

GPS

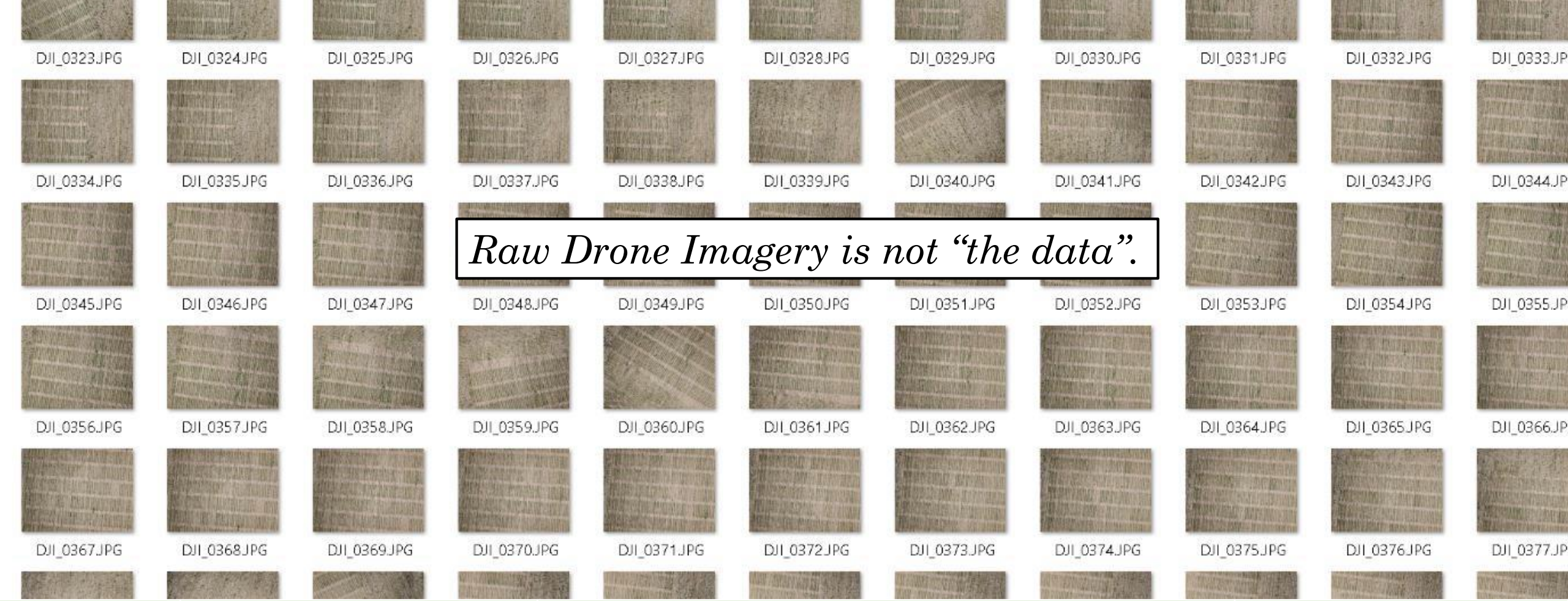
Google Earth

Data Type	Description	Format	Short-term Storage
UAS Flight Info	Flight parameter inputs, and actual flight positions, including GPS locations of all images.	Varies depending on data stream, but primarily ASCII text files. Size: < 1 GB	UAS data card to field computer.

UAS and Ground Data Collection & Phenology Visualization



Traveling phenomics teams will be ineffective.



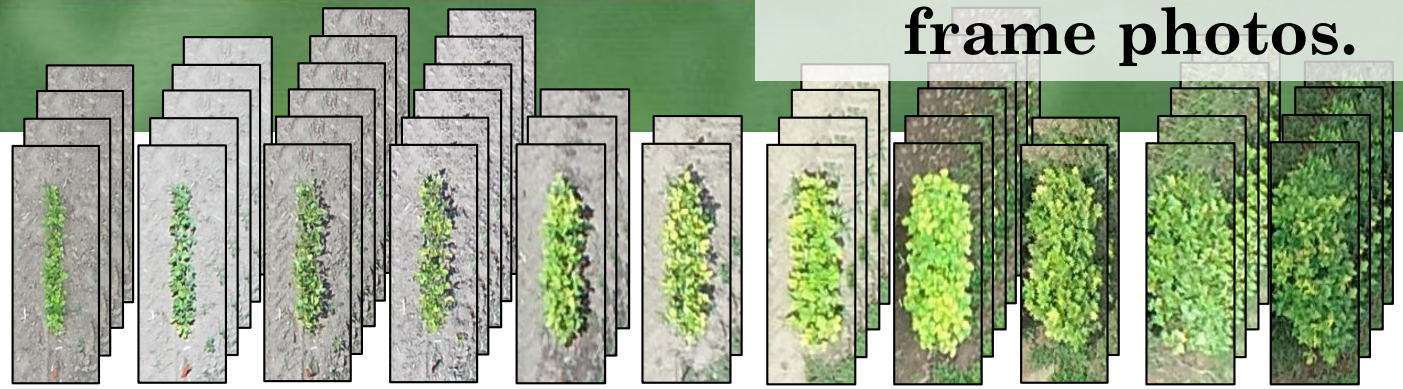
Raw Drone Imagery is not "the data".

Data Type	Description	Format	Short-term Storage
Raw UAS Imagery	Original imagery collected by the UAS cameras, includes RGB, multispectral and TIR.	RAW or other uncompressed image file format. Actual format will be camera dependent. Size: 100s of GB per day, size varies cameras used. Cumulative 5-7 TB per season.	Camera memory card to field computer.

Don't quantify phenotypes from orthomosaic "pretty pictures".

Extract multiple replicate images of individual plots from the raw, overlapping frame photos.

Multi-layer Mosaics



PLOT CLIPS: Extracted smaller images of plots from raw photos. Orthorectified and labeled images of individual plots of *uniform pixel dimensions*. Could be calibrated, segmented, binary, etc.

ORTHORECTIFIED: effects of image perspective (tilt) and relief (terrain) are removed from the image data.

MLM: Multi Layer Mosaic, analysis technique using all raw UAS images

Data Type	Description	Format	Short-term Storage
Processed UAS Imagery	Intermediate labeled image products, potentially atmospherically corrected or segmented.	Various image file formats. Size: Variable from 1 to 100s of GB.	Local computer hard drive.

Advantages of replicate labeled plot clips of uniform pixel dimensions.

Preserve raw image colors & sharpness.
- Avoids mosaicking errors & color distortion.

Trace to original raw photo, which provides position of UAS when photo was taken (cardinal orientation, altitude, pitch, roll, yaw).

Can be filtered, discarded and subset for quality control.
- QC: blurriness and location of a plot within the photo.
- Photo perspective (cardinal) relative to plot, i.e. can correct for glare.

Can be batch processed.

Provide standard deviation of phenotype.
- Enable statistical analyses & quality control.
- Quantify precision & improve accuracy.

Avoids HPC and cloud computing.

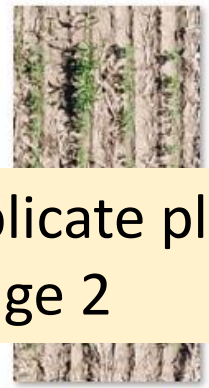
AAAS
Plant Phenomics
Volume 2019, Article ID 2591849, 9 pages
https://doi.org/10.34133/2019/2591849

Plant Phenomics
A SCIENCE PARTNER JOURNAL

Research Article

Easy MPE: Extraction of Quality Microplot Images for UAV-Based High-Throughput Field Phenotyping

Léa Tresch,[1,2] Yue Mu,[3] Atsushi Itoh,[4] Akito Kaga,[5] Kazunori Taguchi,[4] Masayuki Hirafuji,[1] Seishi Ninomiya,[1,3] and Wei Guo[1]



Replicate plot image 2

180606_djifc220_locn_exp_row_1_range_2_rep_2.tif
180606_djifc220_locn_exp_row_1_range_2_rep_3.tif
180606_djifc220_locn_exp_row_1_range_2_rep_4.tif
180606_djifc220_locn_exp_row_1_range_2_rep_5.tif
180606_djifc220_locn_exp_row_1_range_2_rep_6.tif
180606_djifc220_locn_exp_row_1_r_1.tif

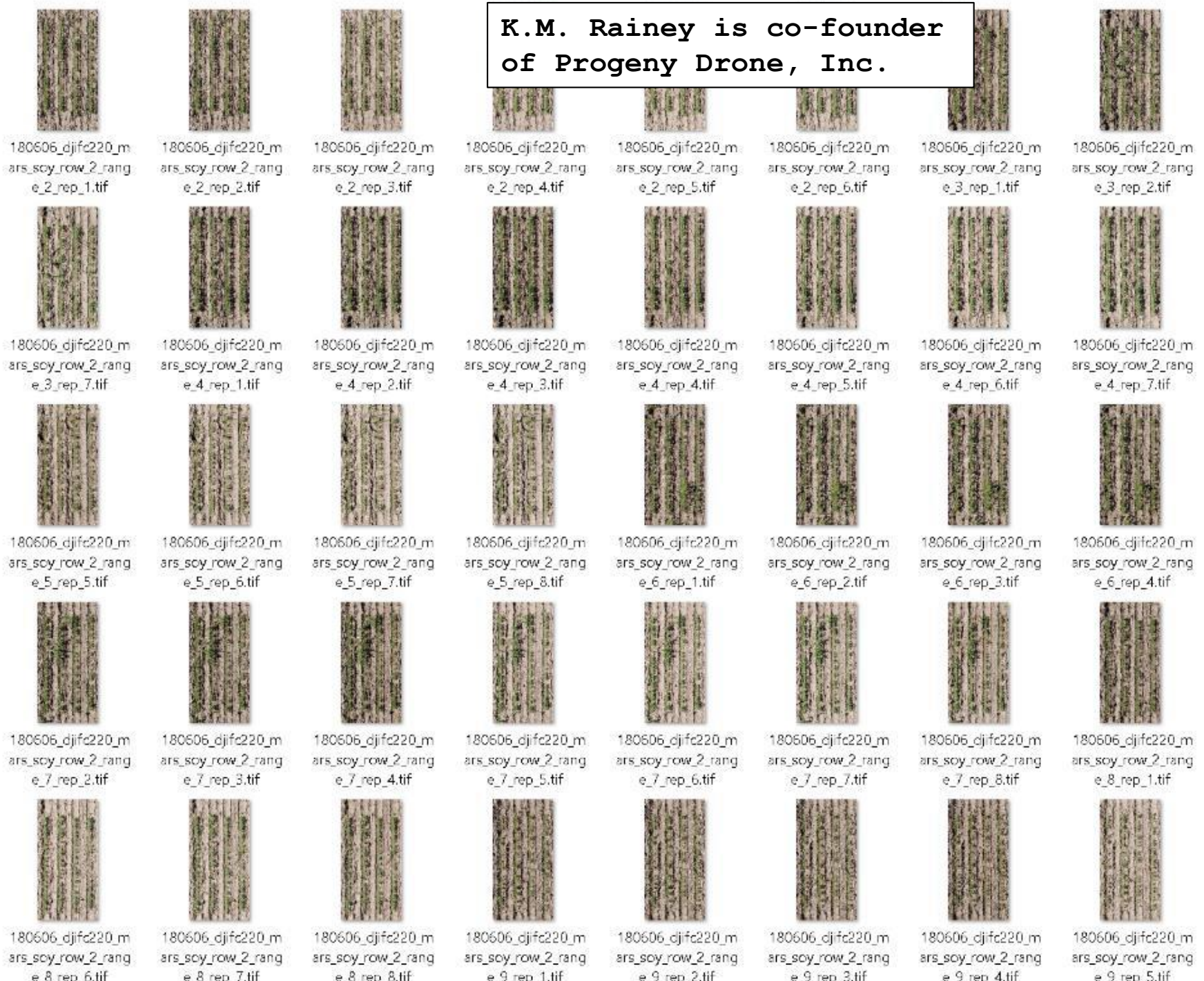
progeny

Software Outputs

- ✕ No RTK-GPS
- ✕ No ground control points
- ✕ No digital terrain models
- ✕ No polygons or shape files
- ✕ No orthomosaics
- ✕ No cloud-computing
- ✕ No high-performance computing

- ✓ Automated
- ✓ 10-20 min
- ✓ At fields edge

- ✓ Imagery QC at site
- ✓ Phenotype QC at site
- ✓ Decision-making at site



K.M. Rainey is co-founder of Progeny Drone, Inc.

progeny

Software Outputs

The replicate phenotypes quantified from plot clips.

K.M. Rainey is co-founder of Progeny Drone, Inc.

	A	B	C	D	E	F	G	H	I	J
1	location	experiment	flight_date	row	range	n reps	median crop cover (m^2)	reference crop cover (m^2)	median G/R ratio	reference G/R ratio
2	laffayatte	soy	180606	2	2	6	2.0493135	2.134933	1.091056	1.09808
3	laffayatte	soy	180606	2	3	7	2.348427	2.348427	1.105727	1.117293
4	laffayatte	soy	180606	2	4	7	2.754763	2.834982	1.128892	1.114134
5	laffayatte	soy	180606	2	5	8	0.9860565	0.972634	1.0683715	1.065476
6	laffayatte	soy	180606	2	6	7	2.260424	2.260424	1.09015	1.091426
7	laffayatte	soy	180606	2	7	8	1.983153	2.105546	1.1090265	1.135468
8	laffayatte	soy	180606	2	8	8	1.672126	1.638688	1.084102	1.077887
9	laffayatte	soy	180606	2	9	8	0.7385695	0.767401	1.056775	1.055729
10	laffayatte	soy	180606	2	10	8	1.4466395	1.520187	1.0675035	1.069511
11	laffayatte	soy	180606	2	11	8	2.0401	2.006424	1.0882015	1.077705
12	laffayatte	soy	180606	2	12	8	2.0730615	2.221983	1.078626	1.074737
13	laffayatte	soy	180606	2	13	8	0.926011	0.883678	1.061797	1.0576
14	laffayatte	soy	180606	2	14	7	0.790751	0.790751	1.067205	1.067205
15	laffayatte	soy	180606	2	15	8	0.6095045	0.598385	1.044362	1.042522
16	laffayatte	soy	180606	2	16	8	1.8588535	1.890147	1.0719365	1.077407
17	laffayatte	soy	180606	2	17	8	1.200662	1.207254	1.08168	1.076523
18	laffayatte	soy	180606	2	18	8	1.723514	1.738922	1.0753835	1.071648
19	laffayatte	soy	180606	2	19	9	2.843719	2.843719	1.114134	1.119063
20	laffayatte	soy	180606	3	2	7	2.718704	2.797335	1.120491	1.127401
21	laffayatte	soy	180606	3	3	7	2.617517	2.55779	1.104525	1.103564
22	laffayatte	soy	180606	3	4	7	3.303905	3.422565	1.122695	1.134504

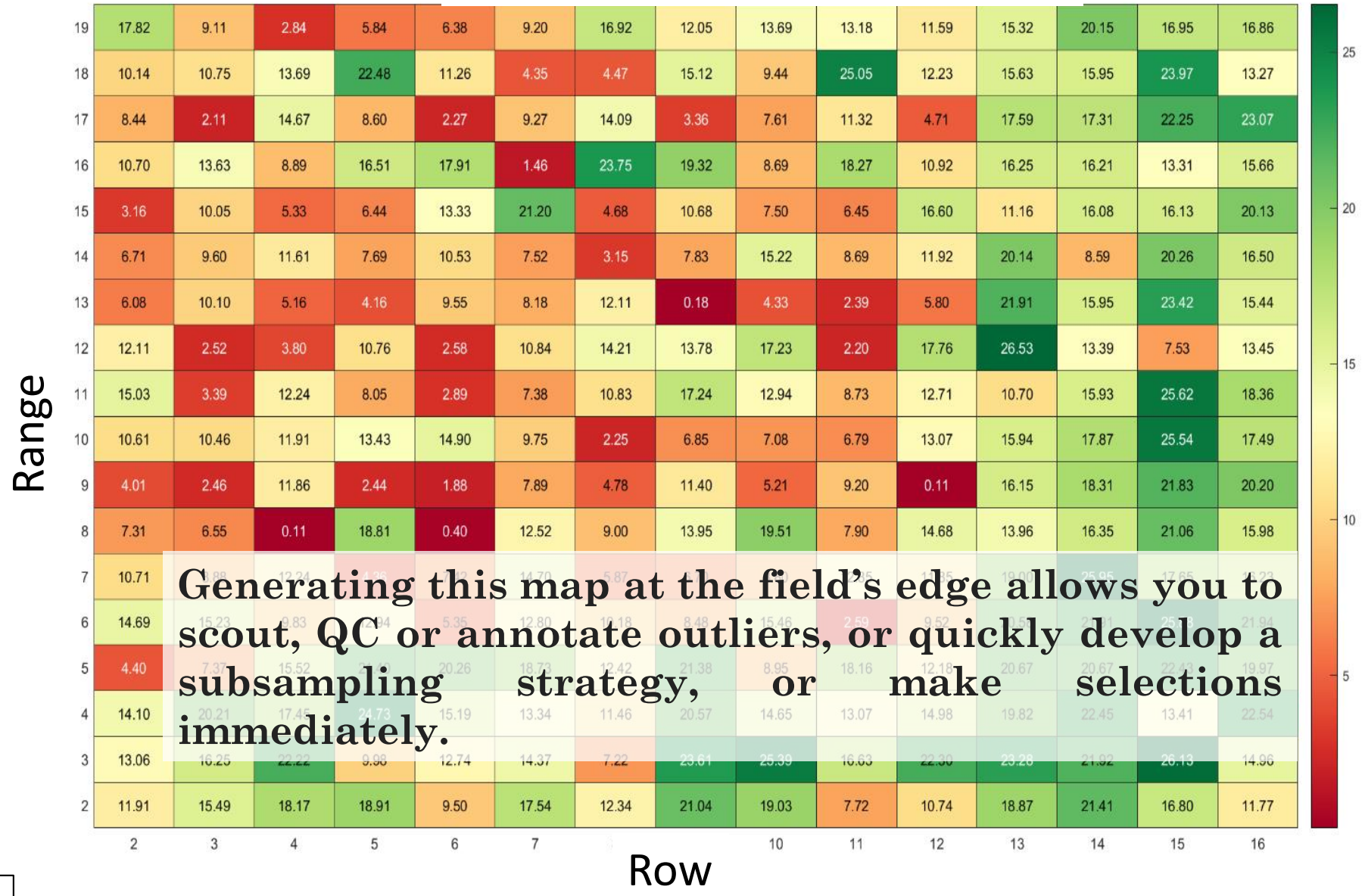
REFERENCE IMAGE/ REFERENCE VALUE: the plot clip taken from the photo that was most nadir/ the phenotype quantified from the reference image.

NADIR VIEW: the camera is directly above the plot.

progeny

Software Outputs

Heat map of plots with real-time UAS phenotypes



Generating this map at the field's edge allows you to scout, QC or annotate outliers, or quickly develop a subsampling strategy, or make selections immediately.

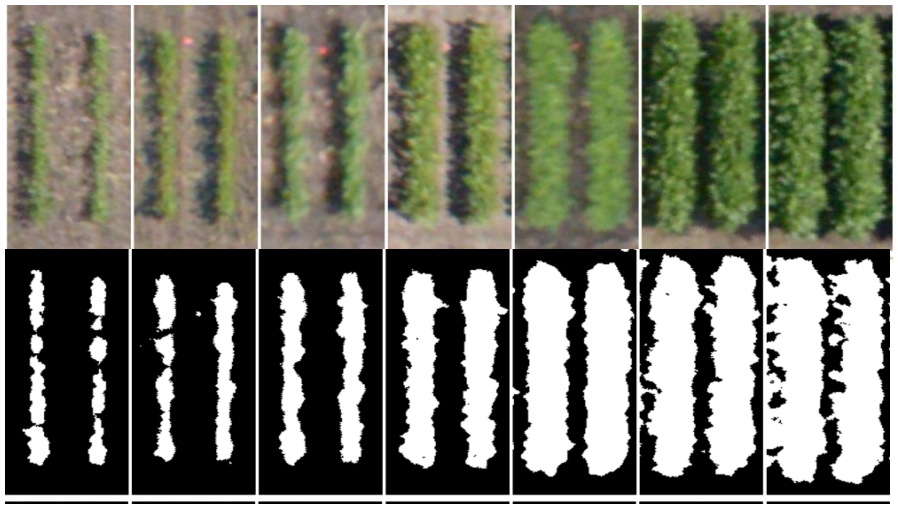
K.M. Rainey is co-founder of Progeny Drone, Inc.

Genetic Architecture of Soybean Canopy Coverage In SoyNAM

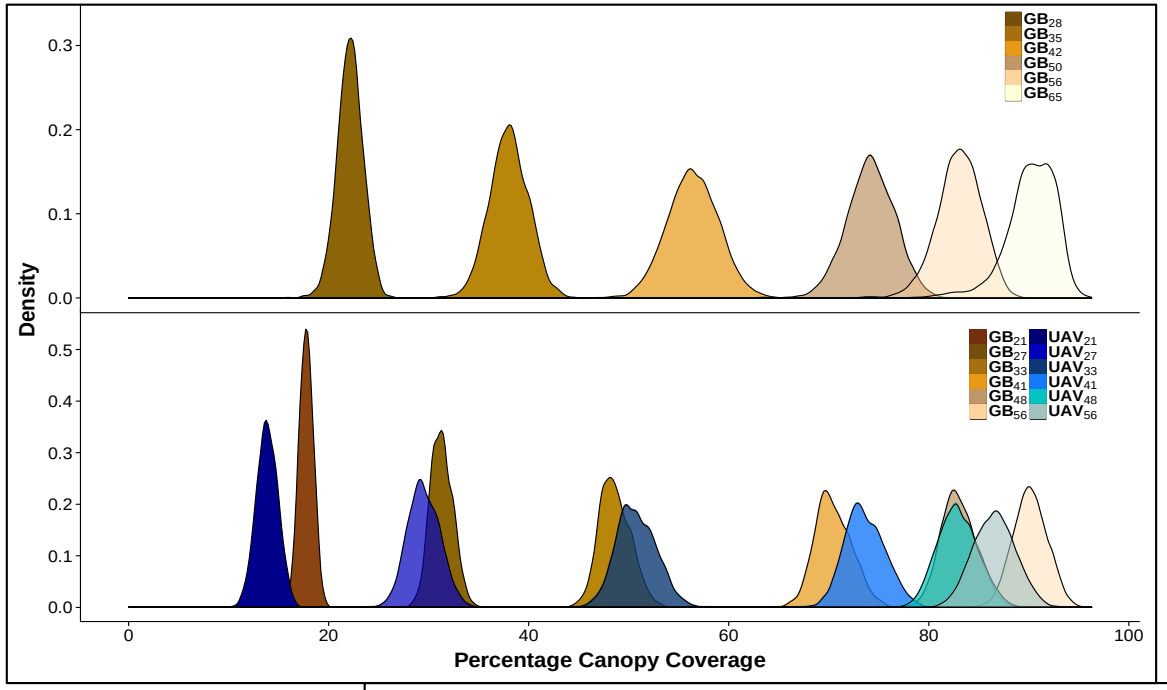
Source Material: SoyNAM

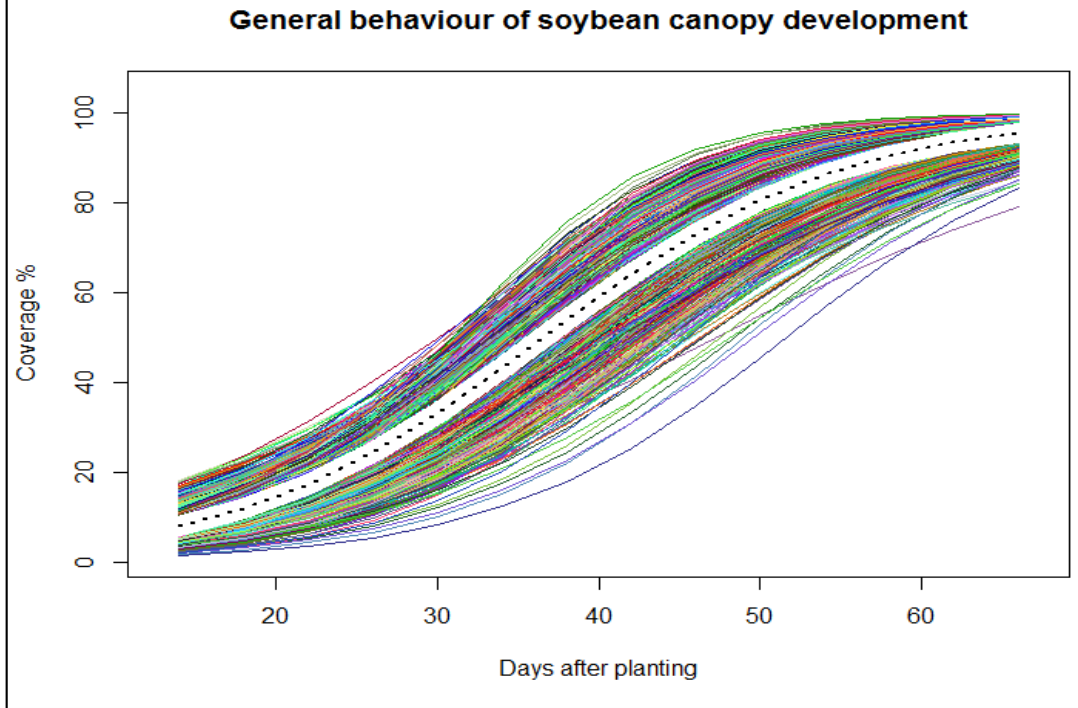
Genetic Architecture of Soybean Yield and Agronomic Traits (2018) Brian W. Diers, *et al.*
G3: GENES, GENOMES, GENETICS October 1, 2018 vol. 8 no. 10 3367-3375

25 28 32 38 43 50 53



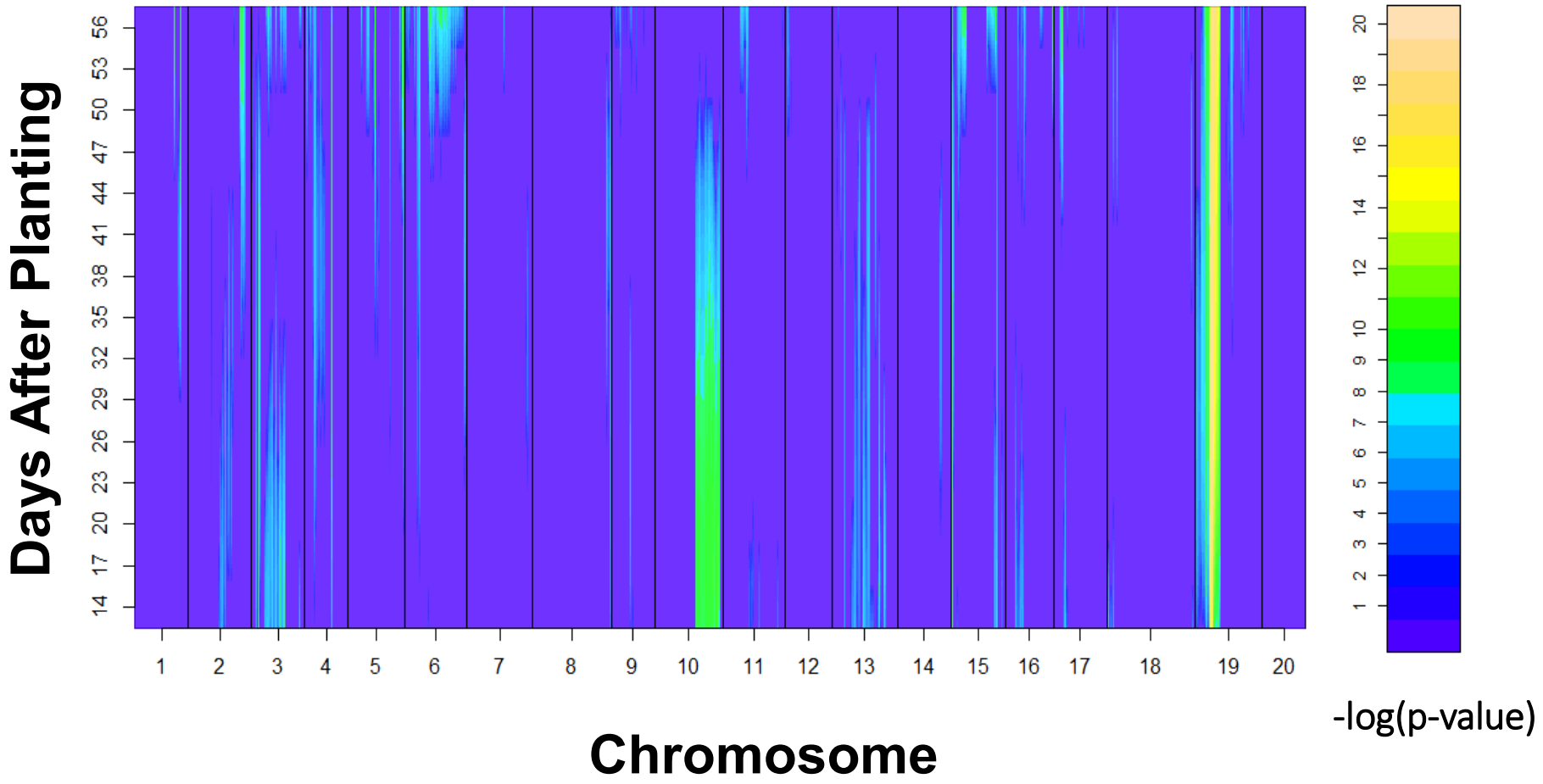
Dr. Alencar Xavier
Corteva AgriSciences
Purdue Agronomy





Alencar Xavier, **B. Hall**, A.A. Hearst, K.A. Cherkauer, and **K.M. Rainey (**2017)
Genetic architecture of phenomic-enabled canopy coverage in Glycine max.
Genetics 206(2): 1081–1089.

GWAS for canopy coverage. Genomic regions significantly associated with early-season soybean canopy coverage observed or estimated for each day from 14 to 56 days after planting.

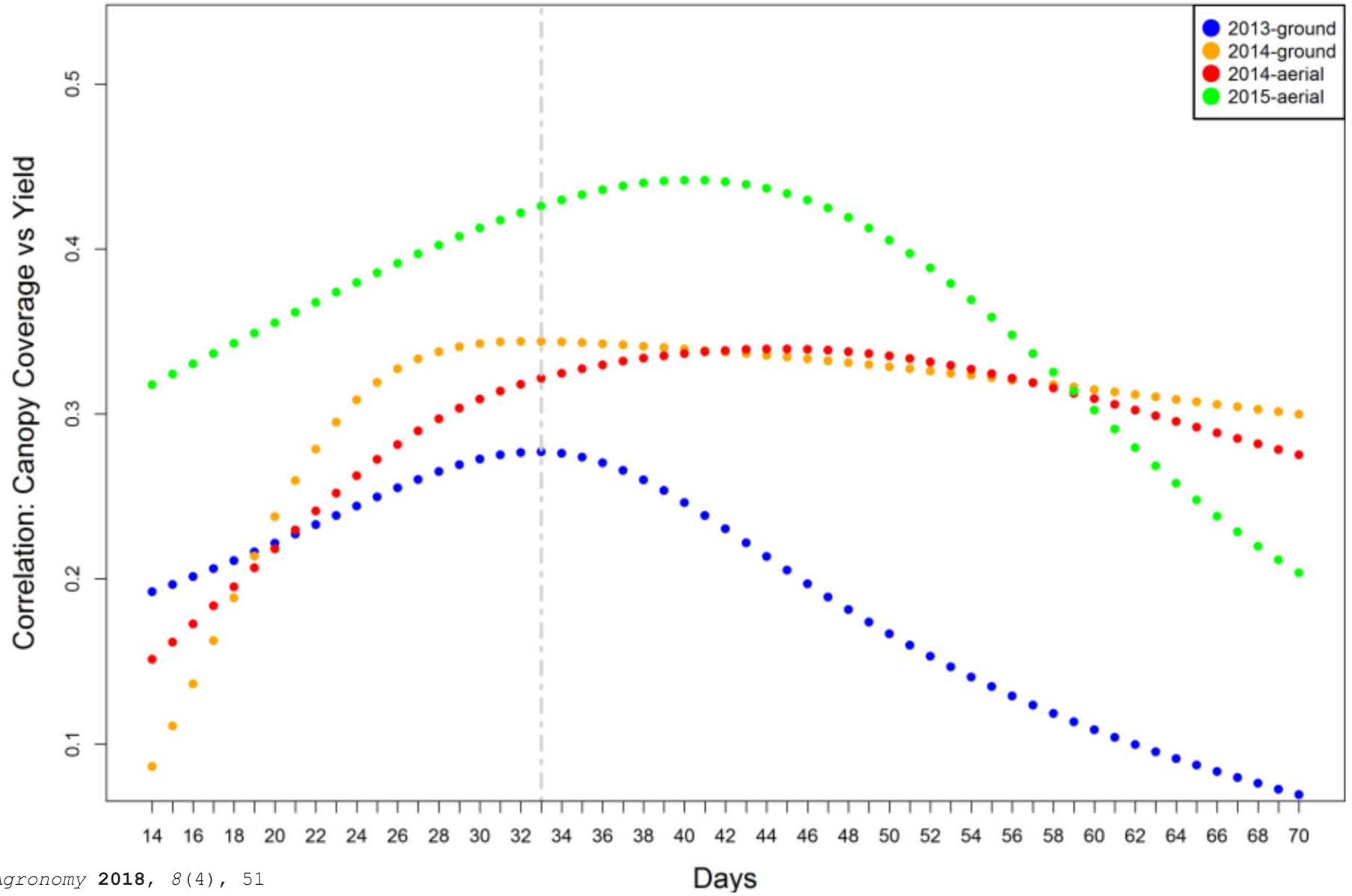


Alencar Xavier, **B. Hall**, A.A. Hearst, K.A. Cherkauer, and **K.M. Rainey (**2017)
Genetic architecture of phenomic-enabled canopy coverage in Glycine max.
Genetics 206(2): 1081–1089.

Table 3 Phenotypic Pearson (P), phenotypic Spearman (S), genetic (G), and environmental (E) correlations among GY, ACC and days to maturity (R8)

	GY-ACC	R8-ACC	GY-R8
Cor(P)	0.63	0.31	0.42
Cor(S)	0.70	0.38	0.46
Cor(G)	0.88	0.77	0.72
Cor(E)	0.18	−0.06	0.23

Correlation between interpolated daily canopy coverage values and yield in SoyNAM.



Phenomic Inference Approach

Aim:
Combine **quantitative genetics** and **physiological growth analysis** to make inferences **for new remote-sensing capabilities.**

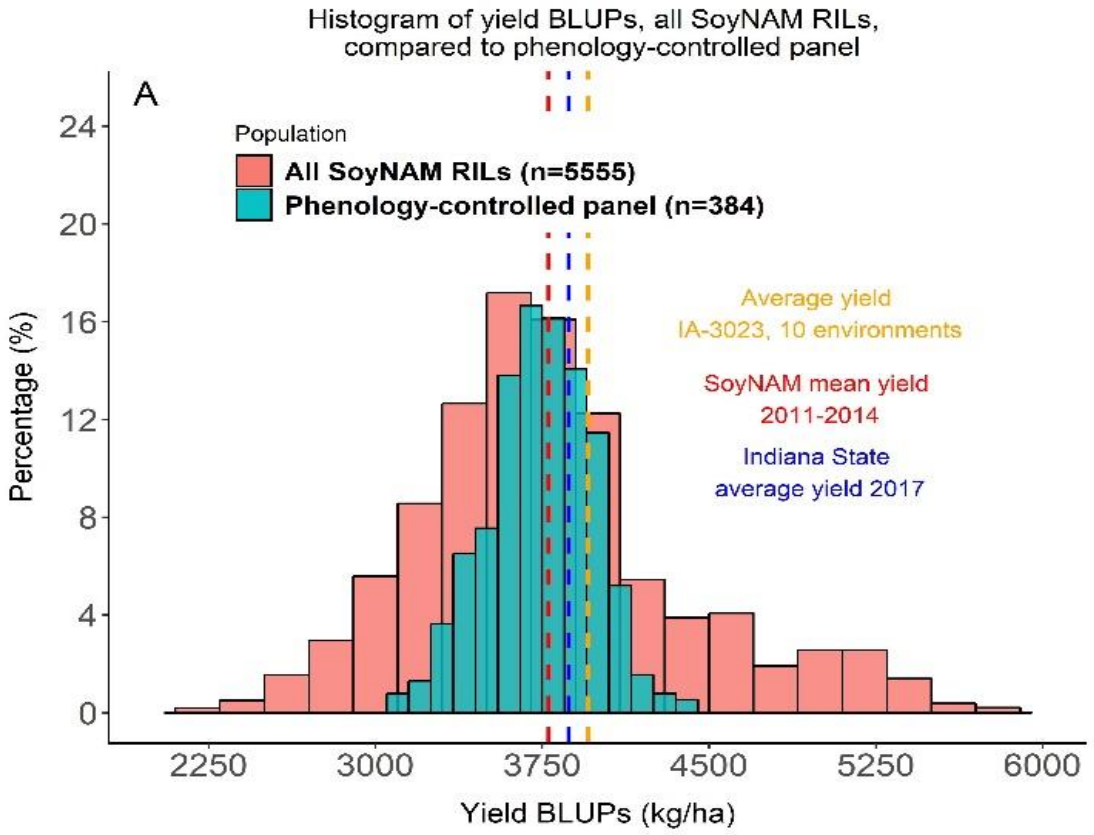
Given:
-temporal quantitative variation in phenotypes, assessed from the ground and remotely,
-a calibration panel
-and quantitative variation for grain yield, plus measures of development,
-and control of factors that confound interpretation,
-and high-density genome-wide markers to estimate kinship among lines,
-and weather data,

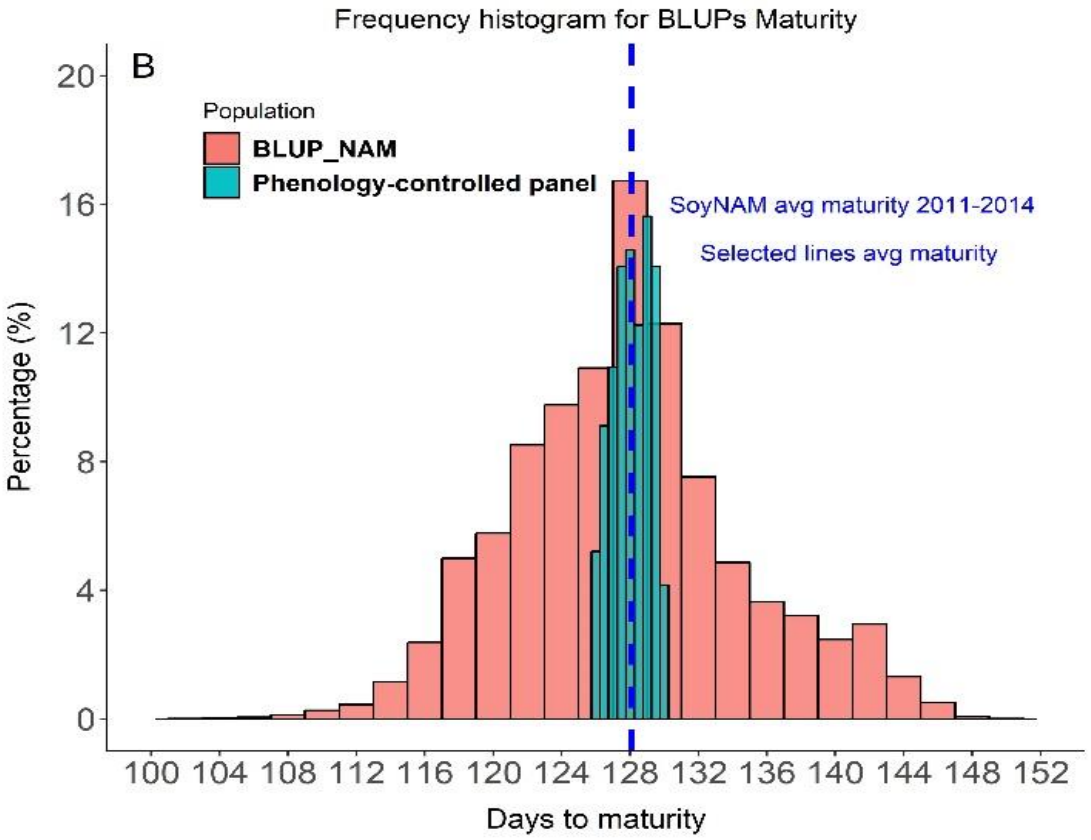
Note: subsample a NAM

calculate:
-summary parameters describing development via multivariate mixed models,
-and genetic & environmental variances for summary parameters via variance decomposition

to determine:
-GWAS outputs, including QTL, genetic architecture, functional mapping, and associations of genomic regions throughout development
-quantitative properties for application to breeding including genetic correlations among traits, other measures of association to yield, and narrow-sense heritabilities (h^2)
-remote-sensing prediction equations and methodology via training/validation approaches.

Biomass & RUE in a phenology-controlled panel (SoyNAM)

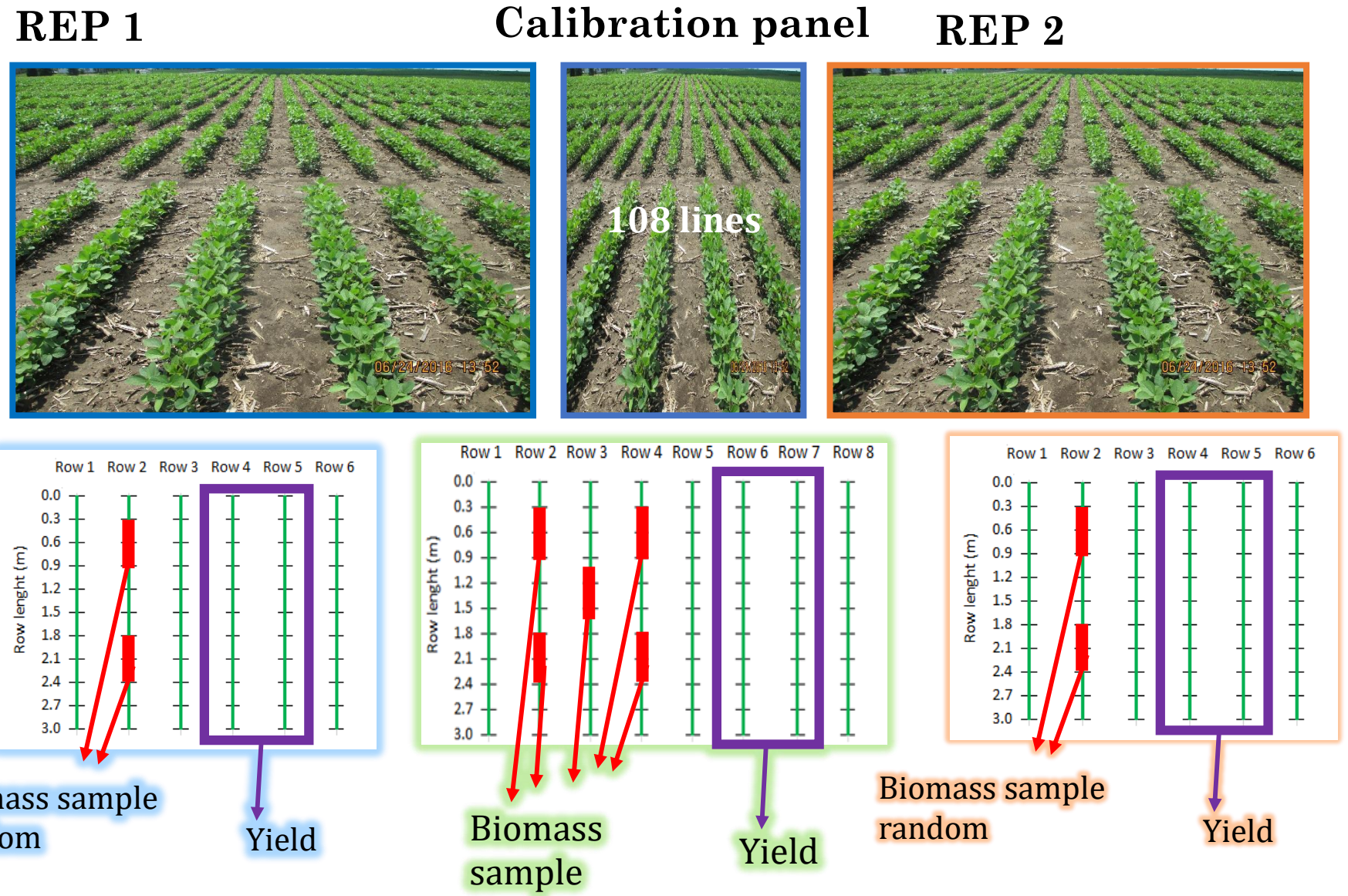




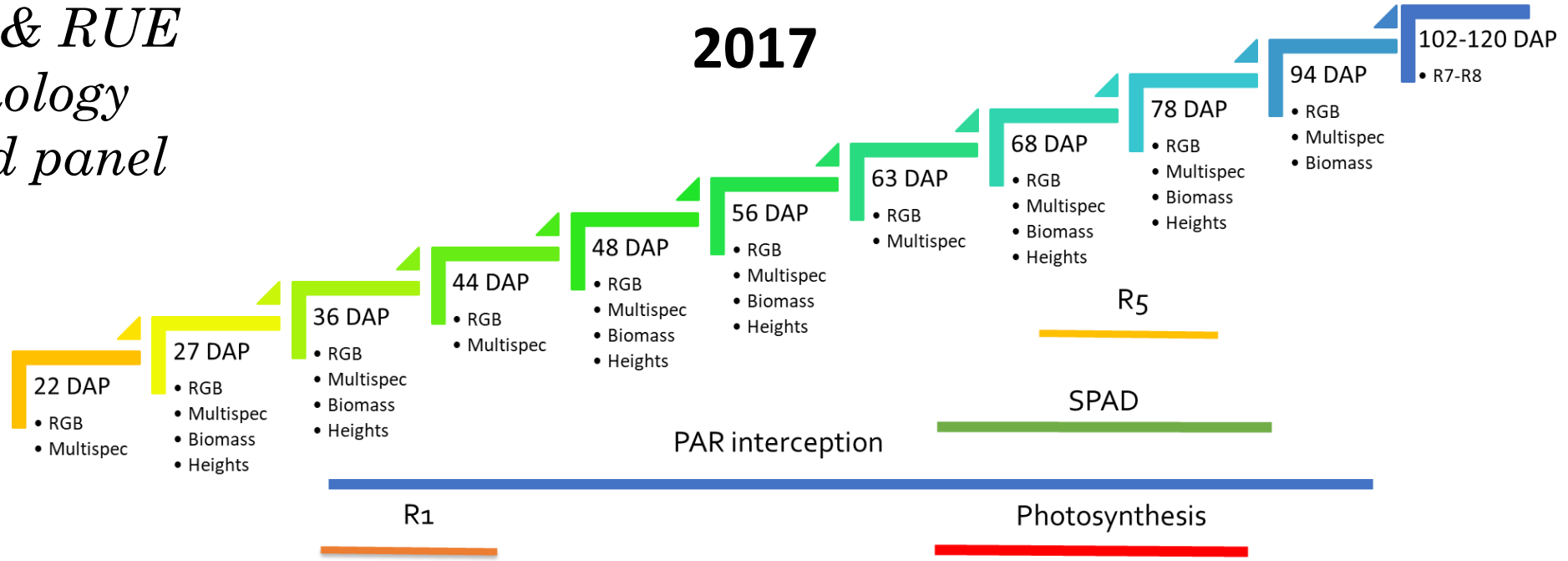
Miguel Lopez (2019). Developing The Yield Equation For Plant Breeding Purposes In Soybean (*Glycine max* L. Merr). PhD Dissertation, Purdue University, Department of Agronomy

Dr. Miguel Lopez
Director of Agronomy at Cenicaña

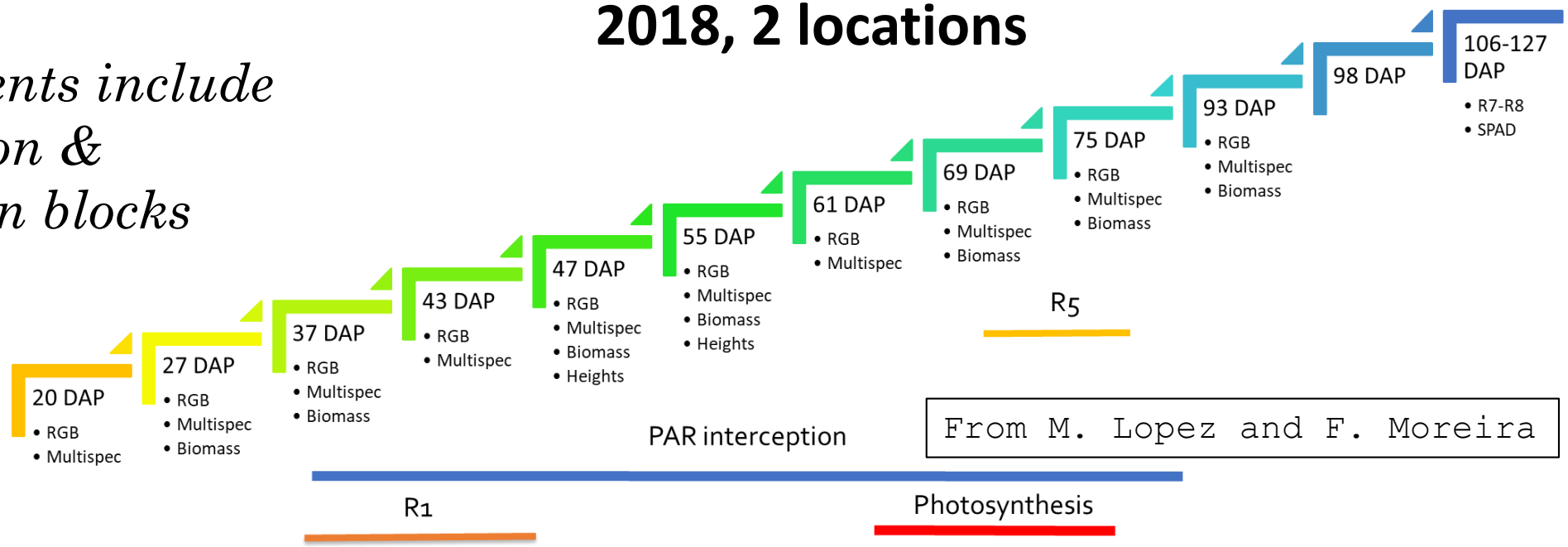
Experimental Design: model training and validation



Biomass & RUE in a phenology controlled panel



Experiments include calibration & validation blocks



Analyses of temporal genetic control of soybean biomass.

With Luiz Brito, Hinayah Oliveira, Fabiana Moreira

- Longitudinal traits: single time, multitrait, random regression…
- Random regression models (RRMs) were used to model biomass across 27 to 83 DAPs. Six different models were fitted: third- and fourth-order Legendre polynomials, linear and quadratic B-splines with one (at 55 DAP) and two knots (at 44 and 66 DAP). All models had the following general notation:

Presentation(s):

•437 - UAS estimation and genetic architecture field-based soybean biomass

•Wednesday, February 26

3:10 PM – 4:10 PM

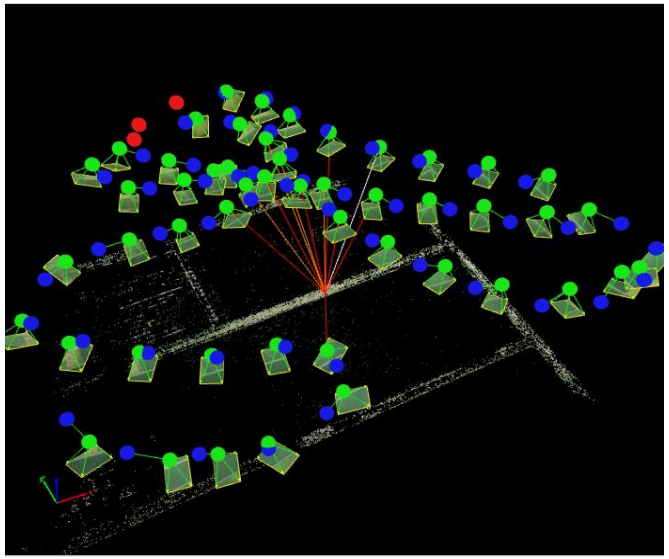
•UAS estimation and genetic architecture field-based soybean biomass

•Thursday, February 27

3:00 PM – 3:15 PM

$$y_{ijk} = Env_k + \sum_{m=1}^{m} b_m \emptyset_m(t_{ij}) + \sum_{m=1}^{m} a_{im} \emptyset_m(t_{ij}) + e_{ijk}, \qquad var\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} \otimes \boldsymbol{G}_0 & 0 \\ 0 & \boldsymbol{I} \otimes \boldsymbol{R} \end{bmatrix}$$

- where $\mathbf{G_0}$ is the (co)variance matrix of the genomic random regression coefficients, G is a genomic relationship matrix, $\mathbf{I}$ is an identity matrix, $\mathbf{R}$ is a diagonal matrix of heterogenous residual variance, and $\otimes$ is the Kronecker product between matrices. The residual variances were allowed to be heterogenous with a unique variance for each DAP with records. Assuming that the phenotype will vary continuously with time between DAP with records, the residual variances were grouped in 18 classes in the following form: 27-33, 34-36, 37, 38-41, 42-43, 44-45, 46, 47-49, 50-53, 54-58, 59-61, 62, 63-65, 66-71, 72-74, 75-76, 77-80, 81-82, and 83.

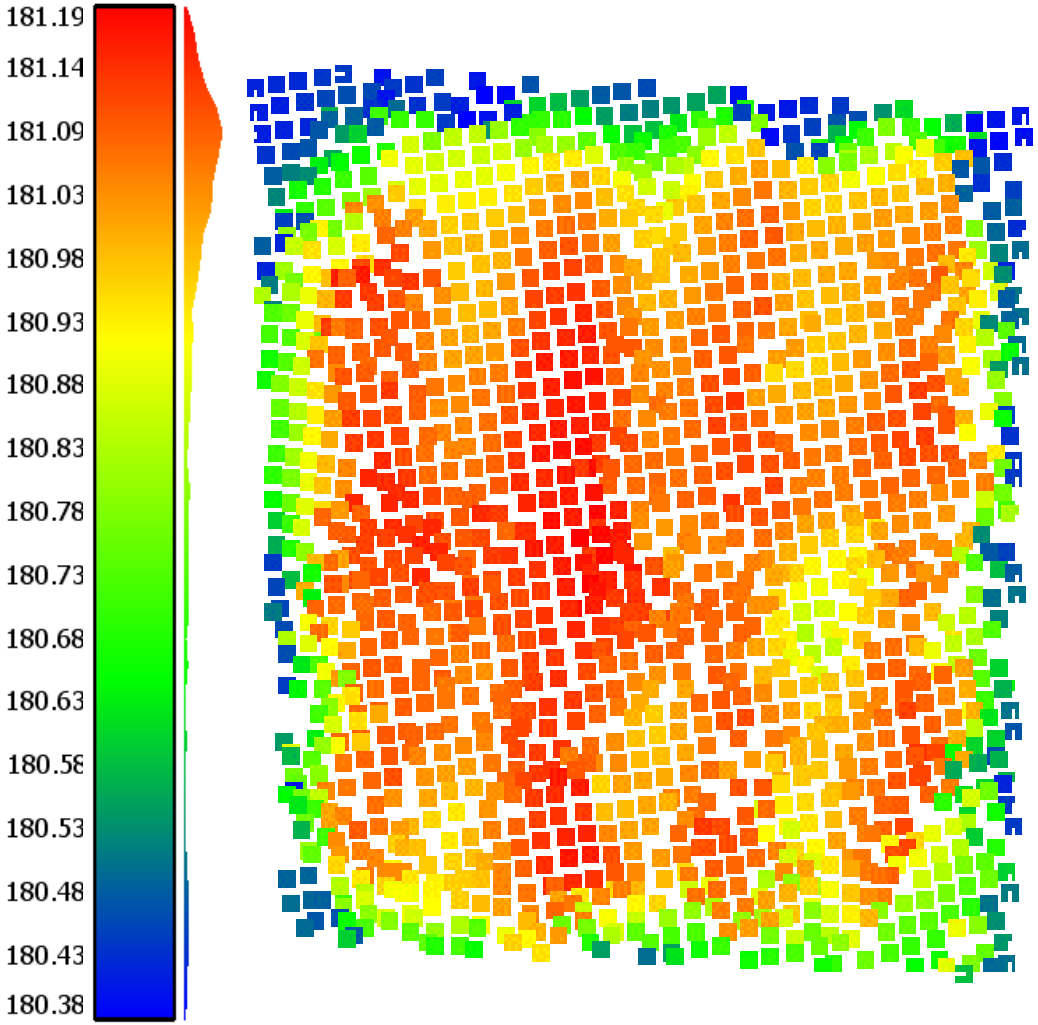


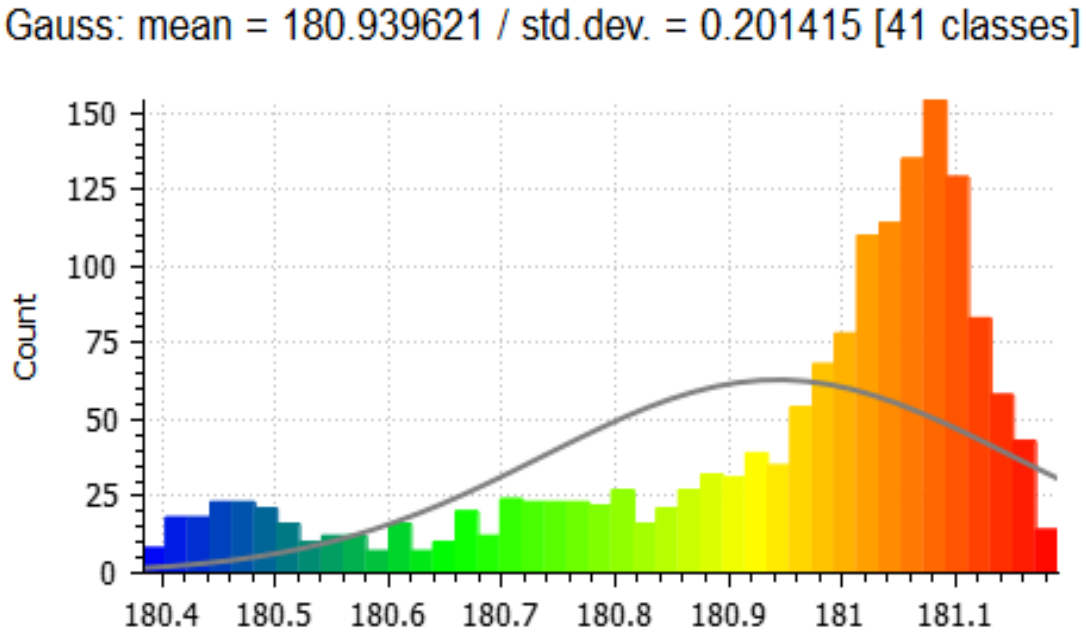
Mesh from Point Cloud by Delaunay triangulation

Dr. Monica Herrero-Huerta
Rainey Lab

Monica Herrero-Huerta, and K. M. Rainey. 'High Throughput Phenotyping of Physiological Growth Dynamics from UAS-Based 3D Modeling in Soybean'.(2019) ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4213

3D Processing: Height Point Clouds





Presentation(s):

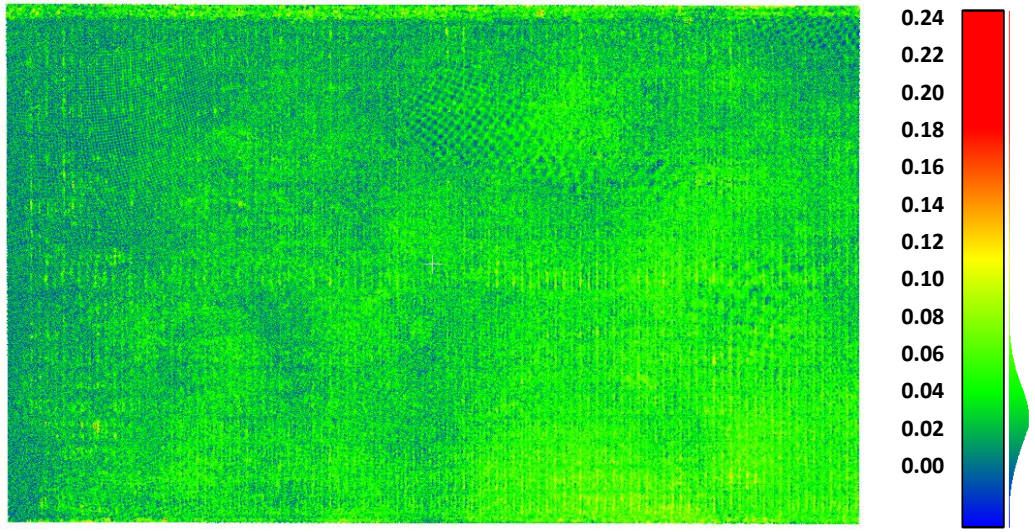
•520 - BIOMASS ESTIMATION FROM UAS POINT CLOUD-BASED STRUCTURE FROM MOTION (SfM) IN SOYBEAN

•Tuesday, February 25
3:15 PM – 4:15 PM

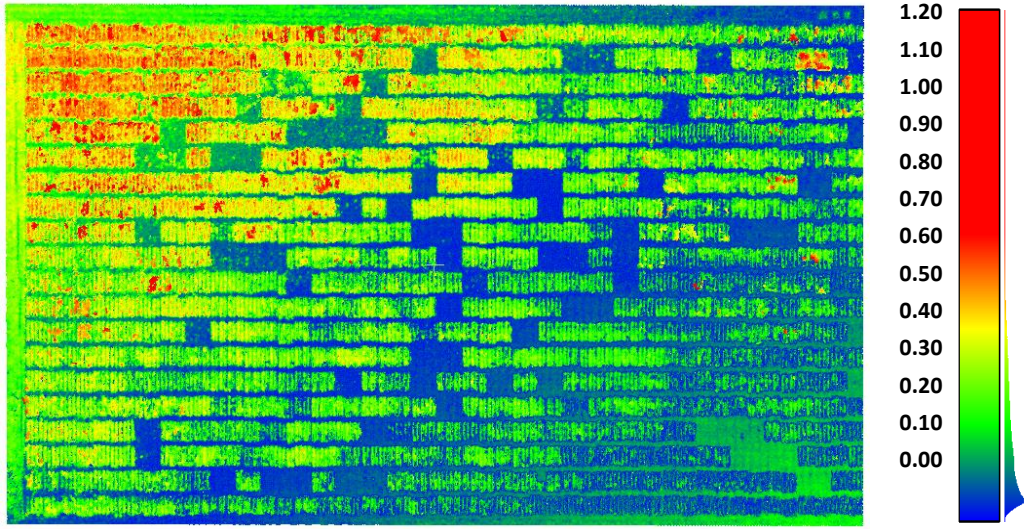
•BIOMASS ESTIMATION FROM UAS POINT CLOUD-BASED STRUCTURE FROM MOTION (SfM) IN SOYBEAN

•Thursday, February 27
9:15 AM – 9:30 AM

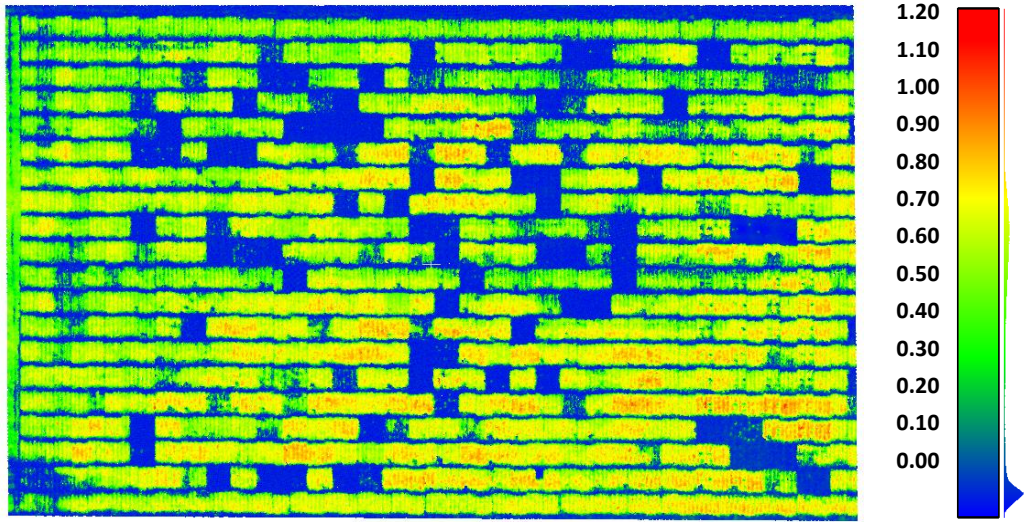
DoD (DEM of Difference)
as reference DAP 14



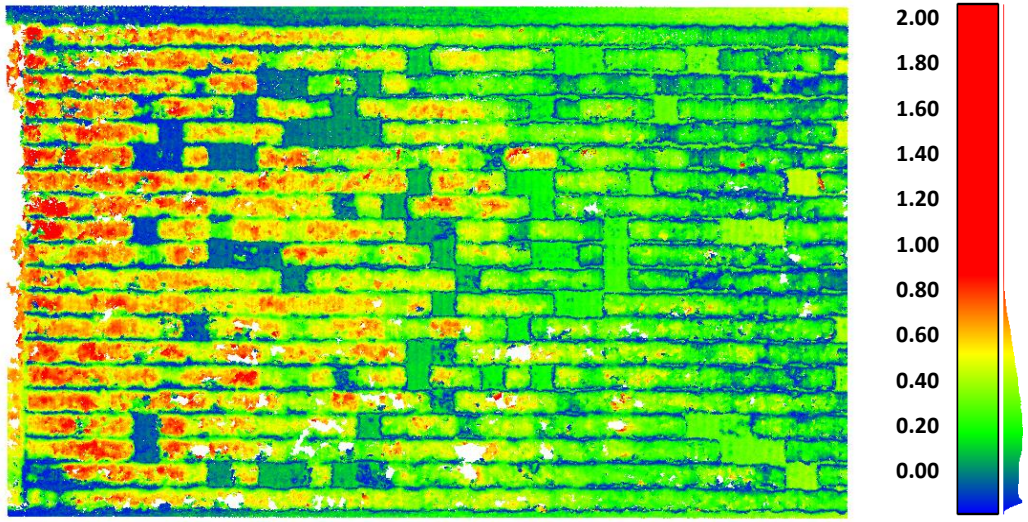
DAP 27



DAP 44



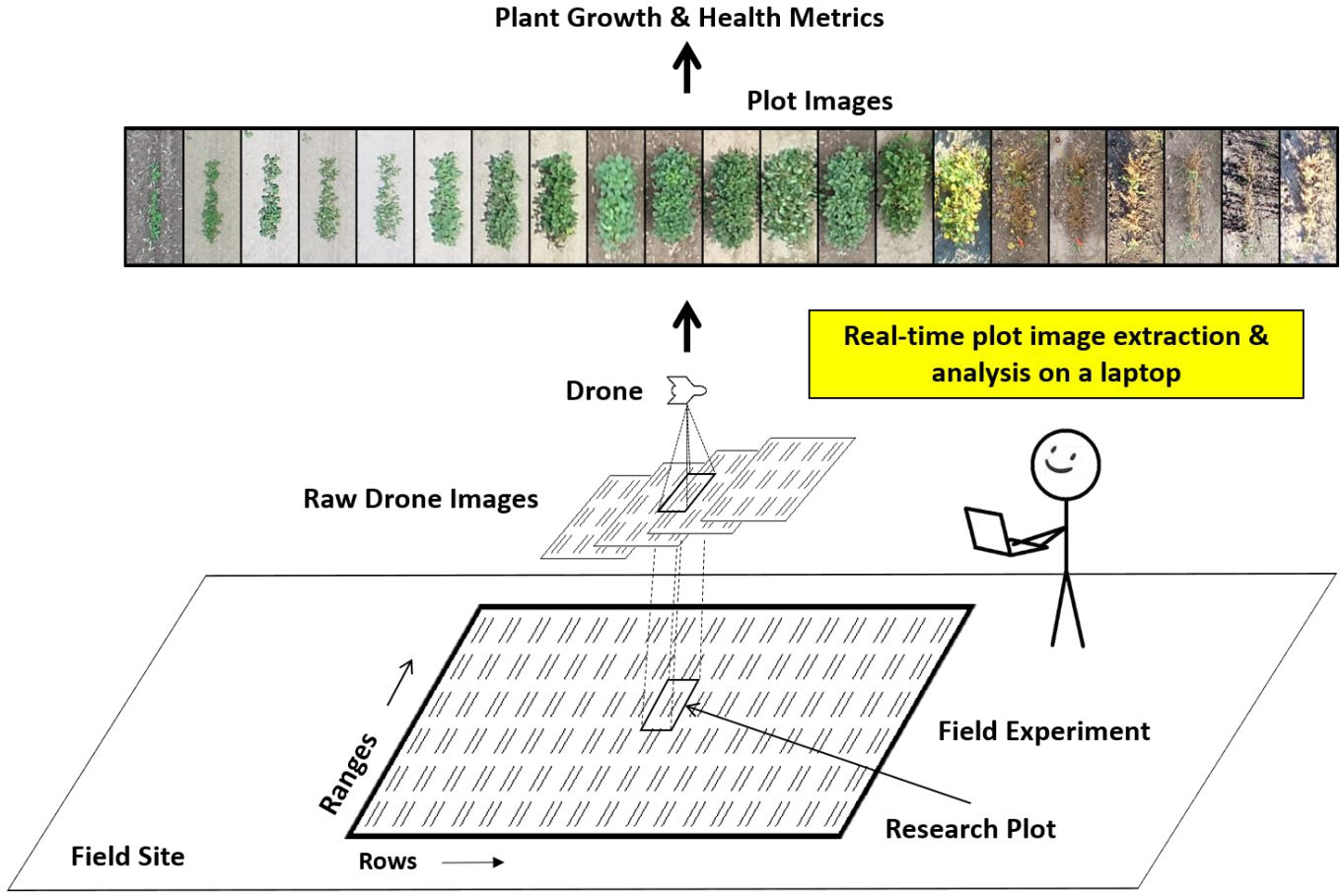
DAP 56



DAP 79

Do Genetics Now!

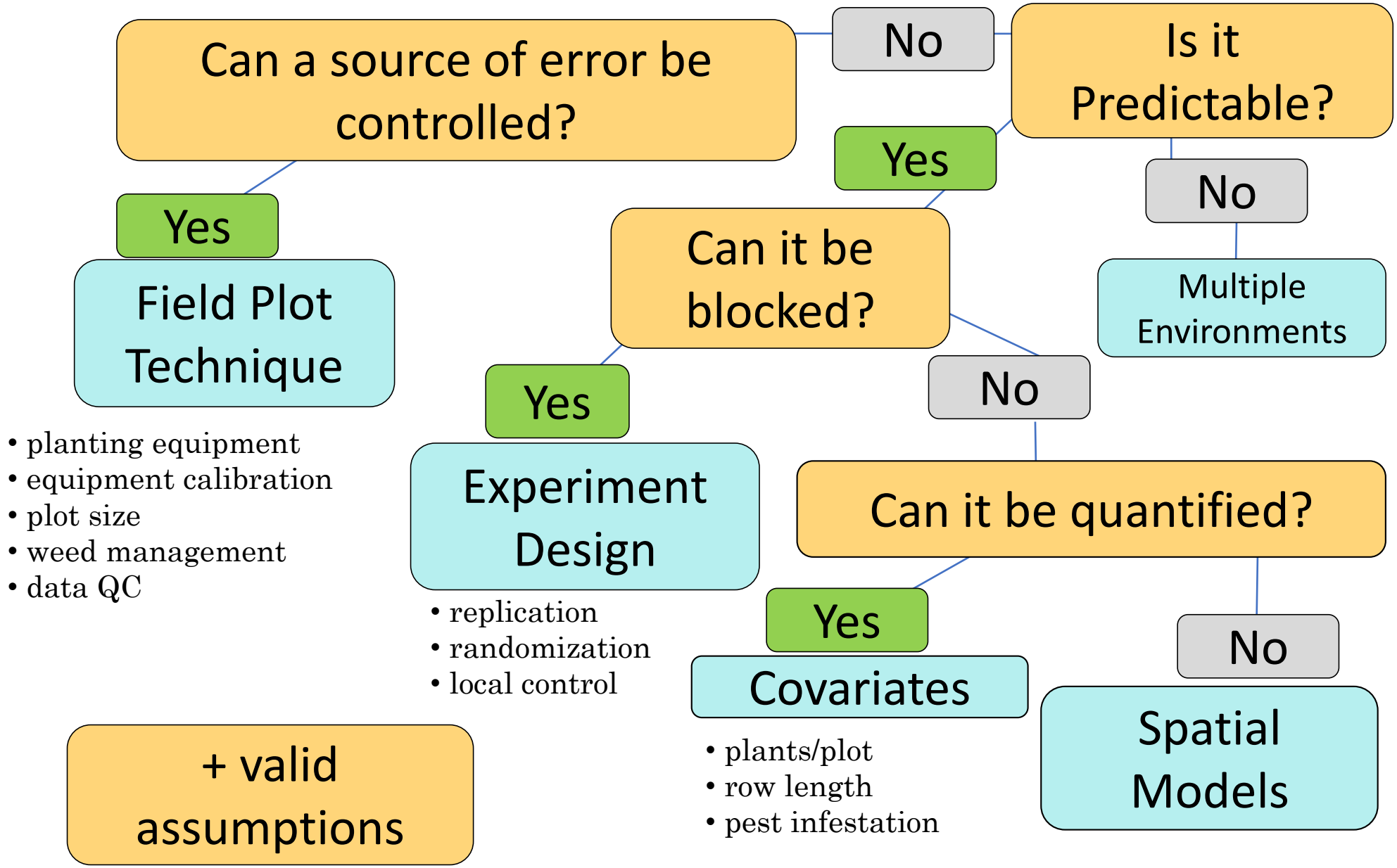
- Our approach is to use robust and inexpensive RGB data collected over multiple environments and sampling dates. We observe critical time points that quickly change.
- Crop scientists and agronomists themselves should progress with applications of these new and valuable data.
- Emerging technologies will follow in the wake of that progress.



Thank you!

Talk to me about:
Senescence
&
Targeted Management

ucing Experimental Error (Noise)



Biomass & RUE : Rationale

Understanding yield formation through processes

$$Y_p = 0.487 \sum_1^n SR \cdot \varepsilon_i \cdot RUE \cdot HI$$

where

$\boldsymbol{Y_p}$ is potential yield in grams per square meter (g m^{-2}) at zero grain moisture,

$\boldsymbol{n}$ is the duration of crop growth in days,

$\boldsymbol{\sum_1^n SR}$ is the cumulative intercepted solar radiation given in megajoules per square meter (MJ m^{-2}),

$\boldsymbol{\varepsilon_i}$ **is the crop radiation interception efficiency also known as fraction of light intercepted,**

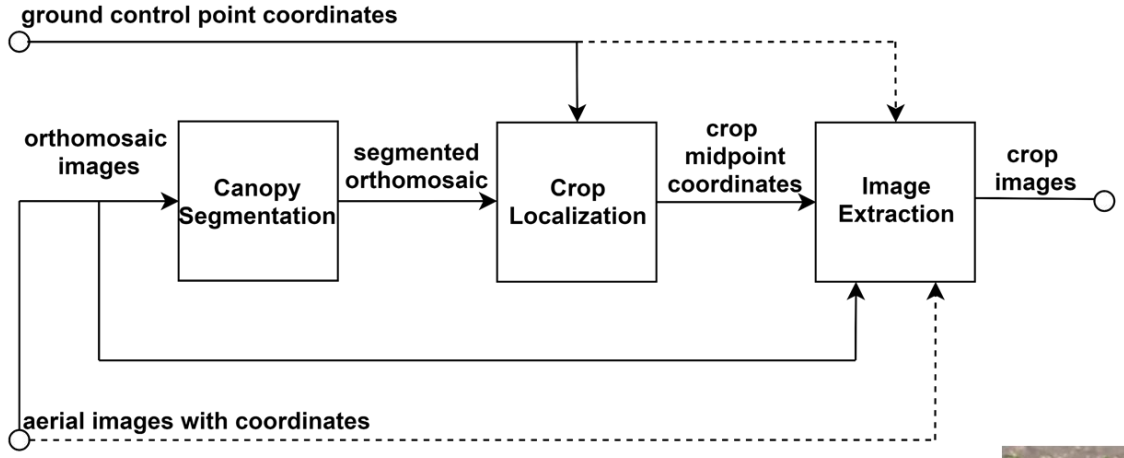
$\boldsymbol{RUE}$ **is radiation use efficiency given in grams of dry matter per megajoule of PAR intercepted (g MJ^{-1}) and**

$\boldsymbol{HI}$ is the harvest index

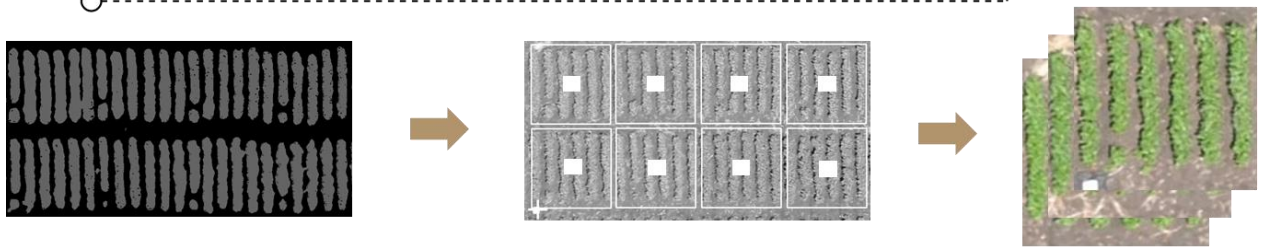
Monteith, J.L. 1972. Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9: 747–766.

CIE + VID
MLM for MS

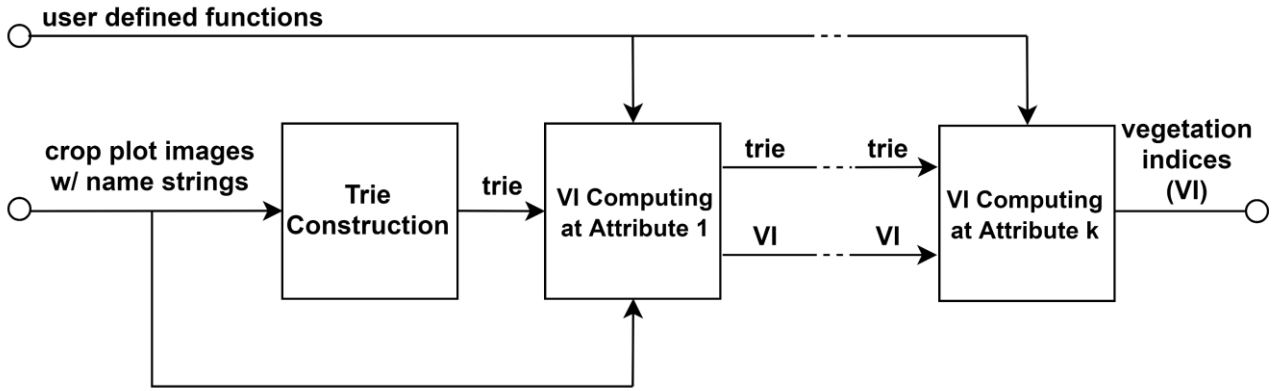
In development in the Cherkauer lab



The work flow of the first module, Crop Image Extraction (CIE)



The work flow of the second module, Vegetation Indices Derivation. A fast image indexing and customized *Vegetation Indices* (VI) derivation.



MLM: Multi Layer Mosaic, analysis technique using all raw UAS images

Software Inputs

K.M. Rainey is co-founder of Progeny Drone, Inc.

- A text file describing the basic layout of the plots

locn_exp_plot_layout.txt

```
1  rows per plot: 4
2  measurement rows: 2,3
3  vertical stagger (ft): 0.000000
4  starting row and range number: 2,2
5  total number of rows and ranges of plots to be processed: 18,15
6  bottom left corner latitude and longitude from Google Earth (decimal degrees): 46.705477,-118.998372
7  top left corner latitude and longitude from Google Earth (decimal degrees): 46.706557,-118.997504
8  top right corner latitude and longitude from Google Earth (decimal degrees): 46.706481,-118.997310
9  bottom right corner latitude and longitude from Google Earth (decimal degrees): 46.705402,-118.998174
```

- A text file describing basic parameters of the camera on the drone

djifc220.txt

```
1  total number of color bands captured by your camera: 3
2  red band number: 1
3  green band number: 2
4  blue band number: 3
5  nir band number: 0
6  sensor width that corresponds to the image pixel width (mm): 6.3
```