EXPERIMENTAL DESIGNS APPLIED TO HIGH-THROUGHPUT PHENOTYPING IN PLANT BREEDING

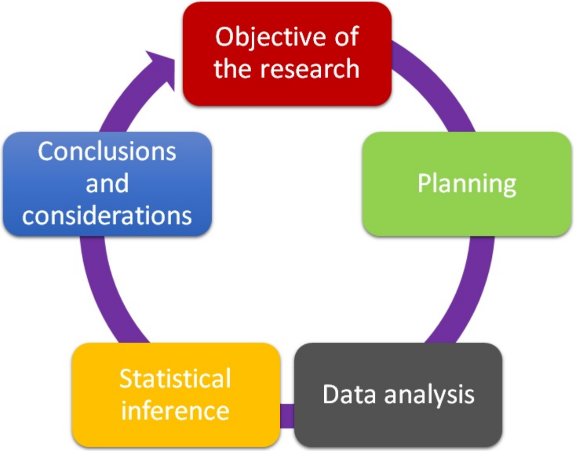
Andreza Jardelino da Silva
Ana Letycia Basso Garcia
José Hernrique Soler Guilhen

Prof. Dr. Roberto Fritsche Neto

Department of Genetics

September 12, 2017


Do you mean, not enough replicates? It is a field experiment! kkkkkkk
Are you crazy? What about the acuracy?!?!

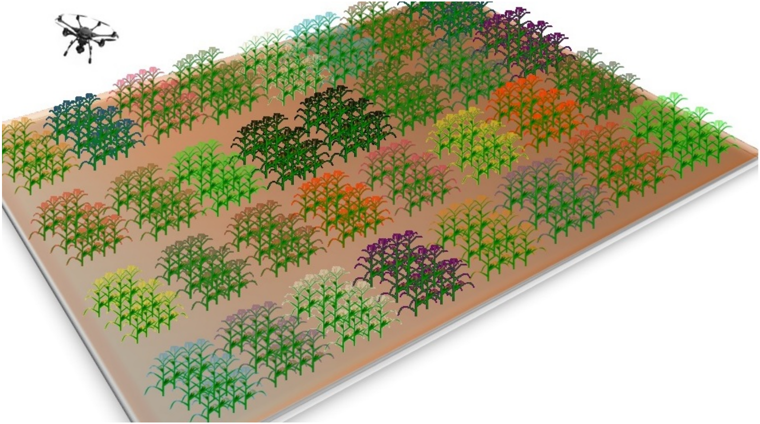

Objective of the research
Planning
Data analysis
Statistical inference
Conclusions and considerations

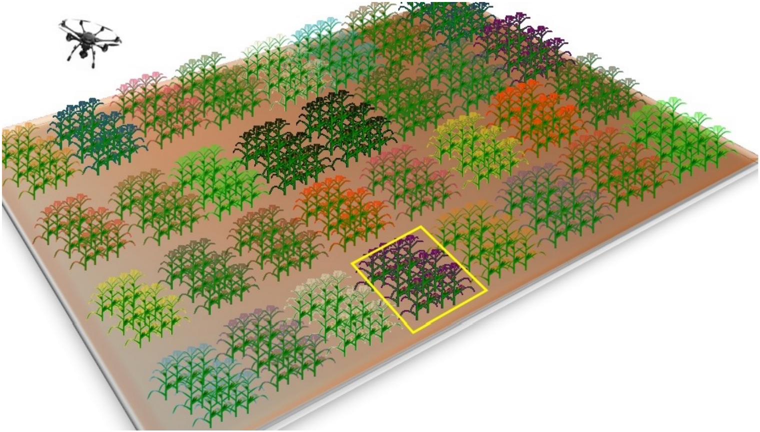
Experimental Designs

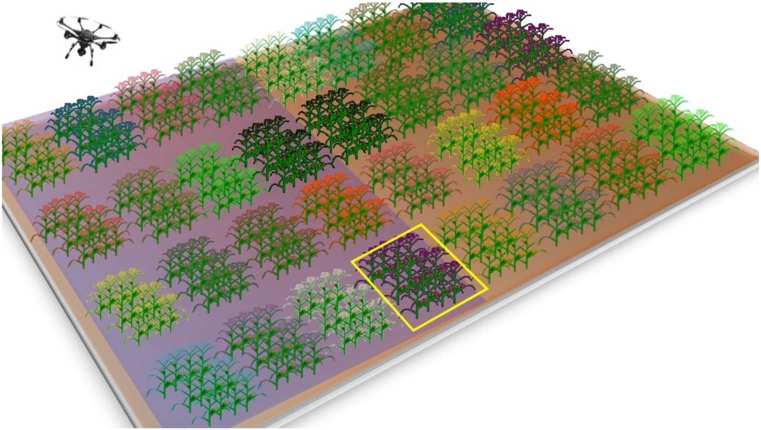
Experimental Structures

Data Collection

Statistical Issues

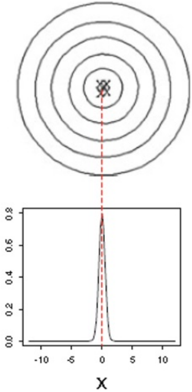






Acuracy

$$\hat{r}_{g\hat{g}} = \frac{\sigma^2_{g\hat{g}}}{\sqrt{\sigma^2_g \cdot \sigma^2_{\hat{g}}}} = \sqrt{h^2}$$

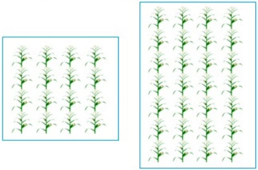


(FRITSCHE NETO & BORÉM, 2015)

Plot

<u>Size and shape</u>

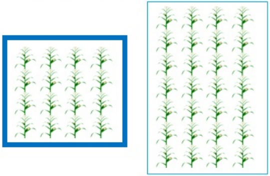
- ✓ To reduce the experimental error
- ✓ Must be uniform
- ✓ Methods to estimate the optimal plot size
- ✓ Which is the best shape?
  - •Environment effects
  - •Image-based data collection

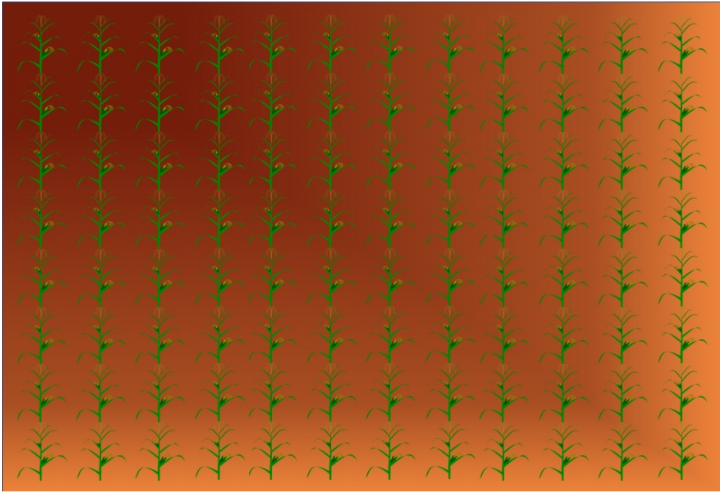


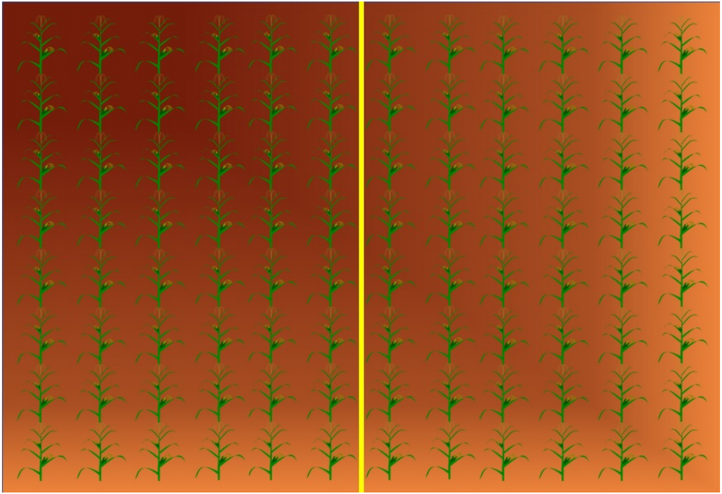
Plot

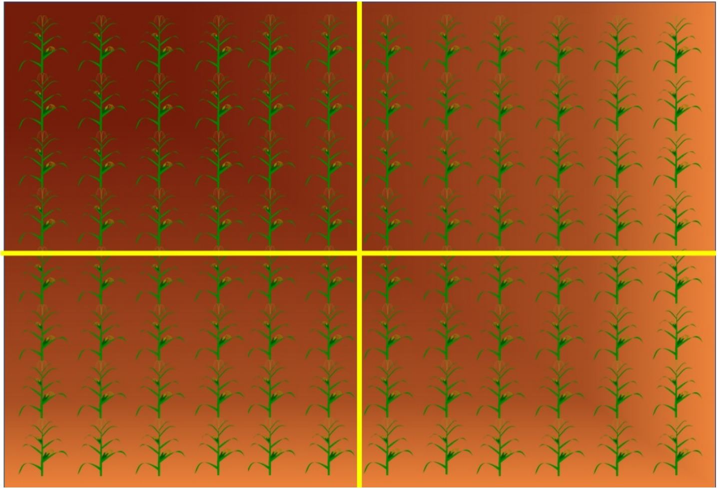
<u>Size and shape</u>

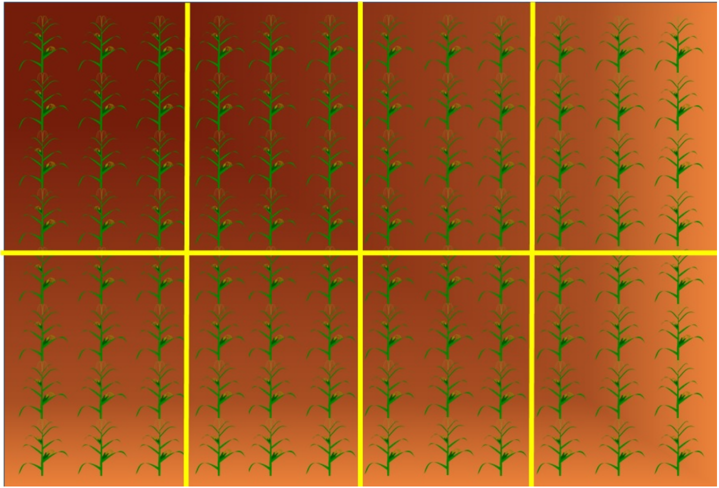
- ✓ To reduce the experimental error
- ✓ Must be uniform
- ✓ Methods to estimate the optimal plot size
- ✓ Which is the best shape?
  - •Environment effects
  - •Image-based data collection

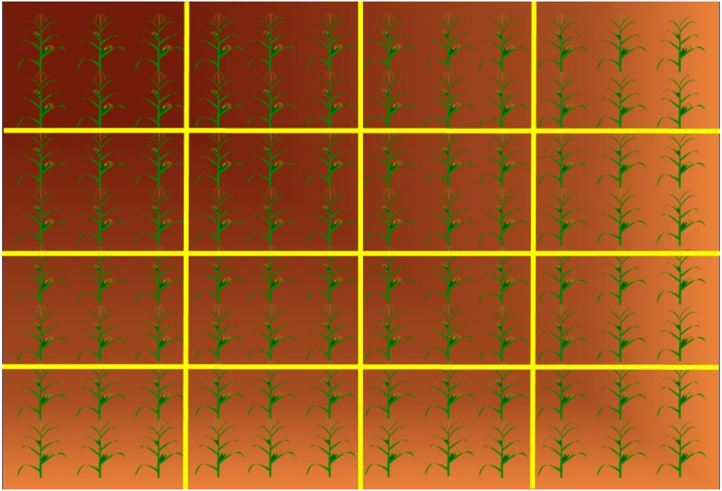


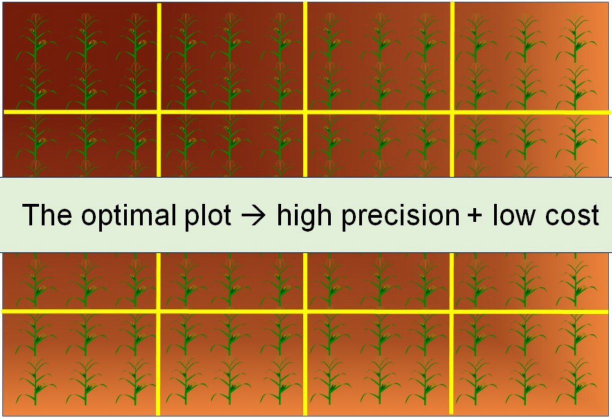




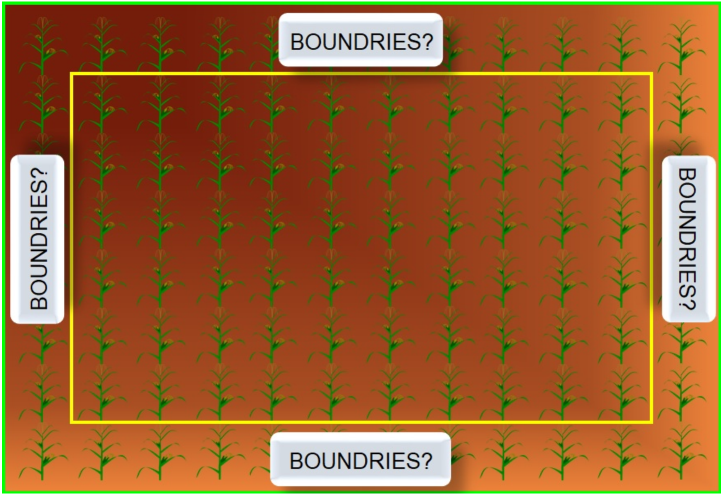








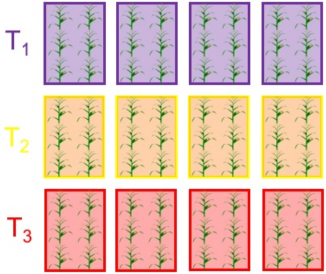
The optimal plot → high precision + low cost


BOUNDRIES?
BOUNDRIES?
BOUNDRIES?
BOUNDRIES?

Basic principles of experimental design

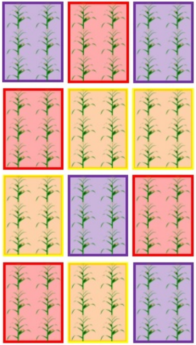
1st: Replication

- ✓ To estimate the experimental variation
- ✓ How many replications are necessary?



Basic principles of experimental design

2nd: Randomization

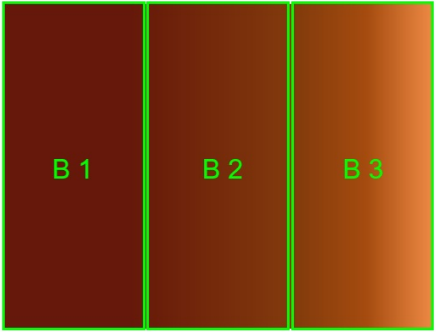


✓ All treatments have the same probability to be in any plot

Basic principles of experimental design

3rd: Local control

✓ Area is divided in homogenous blocks (*blocking*)

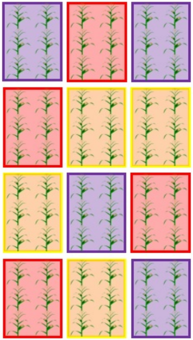


Completely Randomized Design

- ✓ Homogeneous experimental area
- ✓ Statistical model:

$$Y_{ij} = \mu + t_i + \varepsilon_{ij}$$

- ✓ Source of variation (S.V.) → treatments



Randomization

```r
n <- 10 #number of Plots
Standard.Order <- factor(1:n)
Standard.Order
Variety <- factor(rep(c("A", "B", "C"), times = c(3, 4, 3)))
Variety
Plot <- order(r<-runif(n)) # generates 12 random numbers and use the order
# of them as the number for a Plot
r
Plot
CRDPotato.Design <- data.frame(Standard.Order,Plot,Variety)
CRDPotato.Design
remove("Standard.Order", "Plot", "Variety")
CRDPotato.Design[CRDPotato.Design$"Plot",] <- CRDPotato.Design #sort according to
CRDPotato.Design #use the last two columns to give to the scientist
```

	Standard.Order	Plot	Variety
1	1	8	A
2	2	6	A
3	3	3	A
4	4	2	B
5	5	9	B
6	6	1	B
7	7	10	B
8	8	4	C
9	9	7	C
10	10	5	C

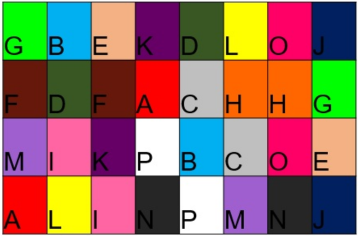
Figure 1: Before randomization.

	Standard.Order	Plot	Variety
1	6	1	B
2	4	2	B
3	3	3	A
4	8	4	C
5	10	5	C
6	2	6	A
7	9	7	C
8	1	8	A
9	5	9	B
10	7	10	B

Figure 2: After randomization.

Table 1. Output from R (package *agricolae*) for a randomization of a completely randomized design experiment, with 16 treatments and two replications.

plots	r	trat	plots	r	trat
1	1	G	17	1	M
2	1	B	18	1	I
3	1	E	19	2	K
4	1	K	20	1	P
5	1	D	21	2	B
6	1	L	22	2	C
7	1	O	23	2	O
8	1	J	24	2	E
9	1	F	25	2	A
10	2	D	26	2	L
11	2	F	27	2	I
12	1	A	28	1	N
13	1	C	29	2	P
14	1	H	30	2	M
15	2	H	31	2	N
16	2	G	32	2	J



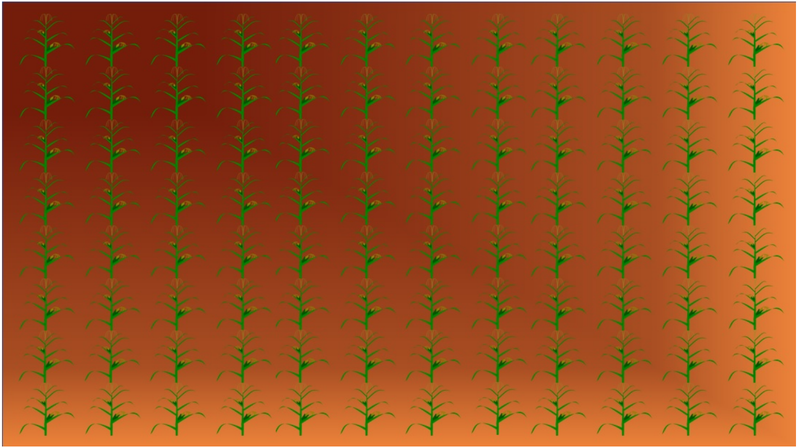
Alógamas Melhoramento ESALQ

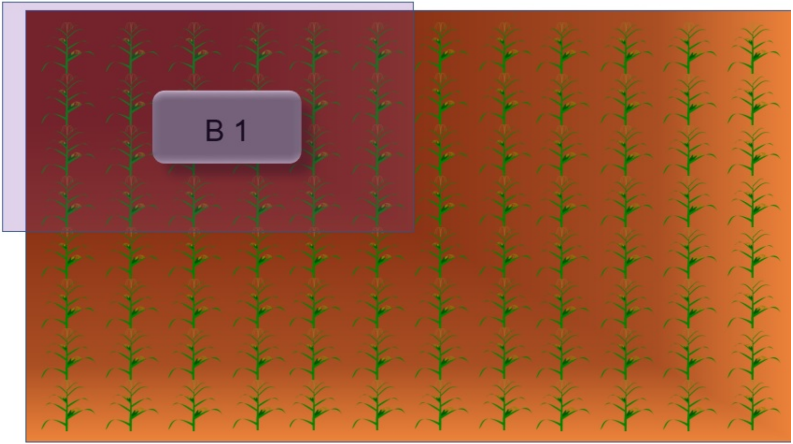
Randomized blocks design

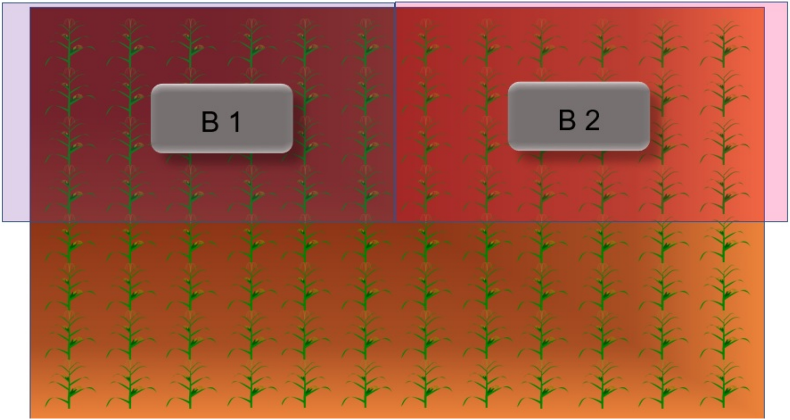
- ✓ Replication, randomization and local control are essential
- ✓ Blocks may contain the same replication of all treatments
- ✓ Randomization inside the blocks

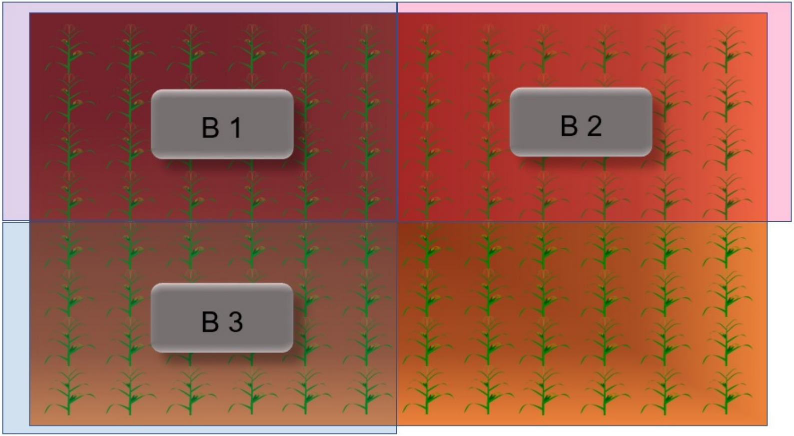
Statistical model:

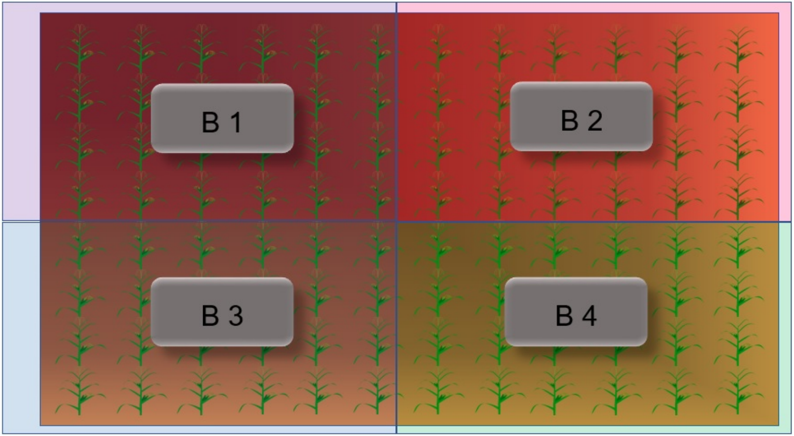
$$Y_{ij} = \mu + \underbrace{t_i + b_j}_{\text{S.V.}} + \varepsilon_{ij}$$




B 1


B 1
B 2


B 1
B 2
B 3


B 1
B 2
B 3
B 4

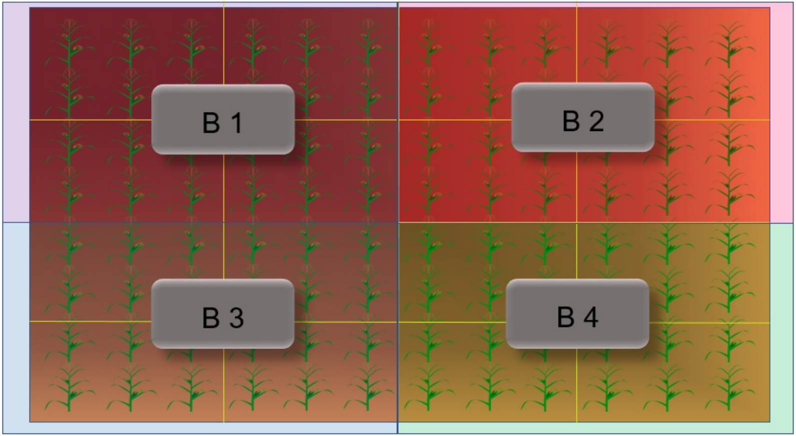

B 1
B 2
B 3
B 4

Table 2. Output from R (package *agricolae*) for a randomization of a RCBD, with 16 treatments and two replications.

plots	block	trat	plots	block	trat
101	1	H	201	2	O
102	1	L	202	2	K
103	1	D	203	2	M
104	1	P	204	2	I
105	1	B	205	2	G
106	1	G	206	2	L
107	1	N	207	2	B
108	1	E	208	2	F
109	1	M	209	2	J
110	1	I	210	2	H
111	1	J	211	2	E
112	1	O	212	2	N
113	1	A	213	2	C
114	1	F	214	2	A
115	1	C	215	2	P
116	1	K	216	2	D

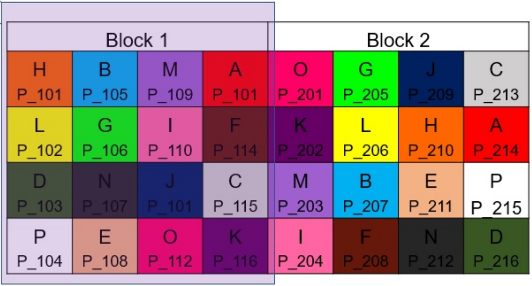
Table 2. Output from R (package *agricolae*) for a randomization of a RCBD, with 16 treatments and two replications.

plots	block	trat	plots	block	trat
101	1	H	201	2	O
102	1	L	202	2	K
103	1	D	203	2	M
104	1	P	204	2	I
105	1	B	205	2	G
106	1	G	206	2	L
107	1	N	207	2	B
108	1	E	208	2	F
109	1	M	209	2	J
110	1	I	210	2	H
111	1	J	211	2	E
112	1	O	212	2	N
113	1	A	213	2	C
114	1	F	214	2	A
115	1	C	215	2	P
116	1	K	216	2	D

Block 1				Block 2			
H P_101	B P_105	M P_109	A P_101	O P_201	G P_205	J P_209	C P_213
L P_102	G P_106	I P_110	F P_114	K P_202	L P_206	H P_210	A P_214
D P_103	N P_107	J P_101	C P_115	M P_203	B P_207	E P_211	P P_215
P P_104	E P_108	O P_112	K P_116	I P_204	F P_208	N P_212	D P_216

Table 2. Output from R (package *agricolae*) for a randomization of a RCBD, with 16 treatments and two replications.

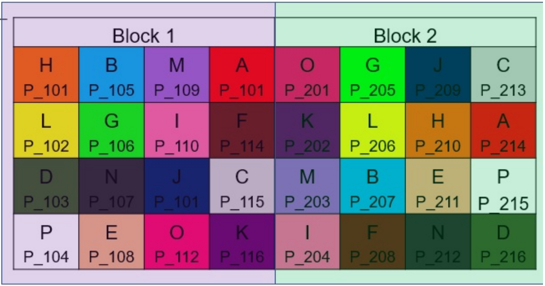
plots	block	trat	plots	block	trat
101	1	H	201	2	O
102	1	L	202	2	K
103	1	D	203	2	M
104	1	P	204	2	I
105	1	B	205	2	G
106	1	G	206	2	L
107	1	N	207	2	B
108	1	E	208	2	F
109	1	M	209	2	J
110	1	I	210	2	H
111	1	J	211	2	E
112	1	O	212	2	N
113	1	A	213	2	C
114	1	F	214	2	A
115	1	C	215	2	P
116	1	K	216	2	D



Alógamas Melhoramento ESALQ

Table 2. Output from R (package *agricolae*) for a randomization of a RCBD, with 16 treatments and two replications.

plots	block	trat	plots	block	trat
101	1	H	201	2	O
102	1	L	202	2	K
103	1	D	203	2	M
104	1	P	204	2	I
105	1	B	205	2	G
106	1	G	206	2	L
107	1	N	207	2	B
108	1	E	208	2	F
109	1	M	209	2	J
110	1	I	210	2	H
111	1	J	211	2	E
112	1	O	212	2	N
113	1	A	213	2	C
114	1	F	214	2	A
115	1	C	215	2	P
116	1	K	216	2	D



Double local control designs - Latin Square Design (LSD)

- ✓ Blocks are build in rows and columns (complete blocks)

- ✓ Replication number = treatment number

$$Y_{ijk} = \mu + \underbrace{c_i + l_j + t_k}_{\text{s.v.}} + \varepsilon_{ij}$$

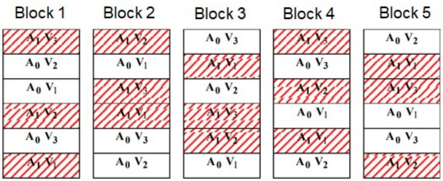
- ✓ e.g.: t = 5 ; r = 5 (LSD) →

	col_1	col_2	col_3	col_4	col_5
row1	E	B	C	A	D
row2	C	E	A	D	B
row3	D	A	B	E	C
row4	A	C	D	B	E
row5	B	D	E	C	A

	CRD	RBD	LSD
Main characteristics	– Replication – Randomization	– Replication – Randomization – Simple local control	– Replication – Randomization – Double local control
Utilization	Homogenous environment (ex.: green house)	Field experiments (heterogeneous area)	Very heterogeneous area
Causes of variation	Treatments	Treatments and blocks	Treatments, rows and columns
Advantages	Highest D.F. for residuals	Precise estimation of residual variation	Residuals are better estimated
Disadvantages	Requires homogeneous area	Plot homogeneity is required within each block	Number of plots increases exponentially as we increase number of treatments
Model	$Y_{ij} = \mu + t_i + \varepsilon_{ij}$	$Y_{ij} = \mu + t_i + b_j + \varepsilon_{ij}$	$Y_{ijk} = \mu + c_i + l_j + t_k + \varepsilon_{ij}$

Experimental Structures- Differences

Factorial Design in Blocks



Experimental Structures- Differences

Factorial Design in Blocks

Block 1	Block 2	Block 3	Block 4	Block 5
$A_1 V_3$	$A_1 V_2$	$A_0 V_3$	$A_1 V_3$	$A_0 V_2$
$A_0 V_2$	$A_0 V_1$	$A_1 V_1$	$A_0 V_3$	$A_1 V_1$
$A_0 V_1$	$A_1 V_3$	$A_0 V_2$	$A_1 V_2$	$A_1 V_3$
$A_1 V_2$	$A_1 V_1$	$A_1 V_3$	$A_0 V_1$	$A_0 V_1$
$A_0 V_3$	$A_0 V_3$	$A_1 V_2$	$A_1 V_1$	$A_0 V_3$
$A_1 V_1$	$A_0 V_2$	$A_0 V_1$	$A_0 V_2$	$A_1 V_2$

Split Plots in Blocks

Block 1	Block 2	Block 3	Block 4	Block 5
$A_1 V_2$	$A_1 V_1$	$A_0 V_2$	$A_1 V_1$	$A_0 V_3$
$A_1 V_1$	$A_1 V_3$	$A_0 V_3$	$A_1 V_2$	$A_0 V_1$
$A_1 V_3$	$A_1 V_2$	$A_0 V_1$	$A_1 V_3$	$A_0 V_2$
$A_0 V_2$	$A_0 V_3$	$A_1 V_3$	$A_0 V_2$	$A_1 V_2$
$A_0 V_3$	$A_0 V_1$	$A_1 V_2$	$A_0 V_3$	$A_1 V_1$
$A_0 V_1$	$A_0 V_2$	$A_1 V_1$	$A_0 V_1$	$A_1 V_3$

ANOVA

Table: Factorial Design in Blocks

Source	df
Block	r-1
Factor A	a-1
Factor B	b-1
$(A \times B)$	(a-1)(b-1)
Residual	(ab-1)(r-1)

ANOVA

Table: Factorial Design in Blocks

Source	df
Block	r-1
Factor A	a-1
Factor B	b-1
(A×B)	(a-1)(b-1)
Residual	(ab-1)(r-1)

Table: Split Plots in Blocks

Source	df
Block	r-1
Factor A	a-1
Residual (a)	(r-1)(a-1)
Factor B	b-1
(A×B)	(a-1)(b-1)
Residual (b)	a(b-1)(r-1)

α − *Design*

Principles

- ✓ Randomization

α – Design

Principles

- ✓ Randomization
- ✓ Local control

α − Design

Principles

- ✓ Randomization
- ✓ Local control
- ✓ Replications

α – Design

Principles

- ✓ Randomization
- ✓ Local control
- ✓ Replications

Example

Replication-I				Replication-II				Replication-III			
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
1	2	3	4	1	2	3	4	1	2	3	4
5	6	7	8	7	8	5	6	8	5	6	7
9	10	11	12	12	9	10	11	10	11	12	9

Balanced incomplete block design (BIB)

where

$$\lambda(t-1) = r(k-1)$$

Balanced incomplete block design (BIB)

where

$$\lambda(t-1) = r(k-1)$$

✓ λ is pairs, k is plot, t is treatment, r is replication

Balanced incomplete block design (BIB)

where

$$\lambda(t-1) = r(k-1)$$

- ✓ λ is pairs, k is plot, t is treatment, r is replication

Example

B1	A	C	B
B2	B	D	A
B3	C	A	D
B4	B	D	C

Augmented Block Designs

P_1	T_1	Block 1 = set 1 = small block 1
P_2	T_2	
$\vdots$	$\vdots$	
P_{k1}	T_c	
P_{16}	T_1	Block 2 = set 2 = small block 2
P_{17}	T_2	
$\vdots$	$\vdots$	
P_{k2}	T_c	
$\vdots$		$\vdots$
P_{41}	T_1	Block 3 = set 3 = small block 3
P_{42}	T_2	
$\vdots$	$\vdots$	
P_{kb}	T_c	

Augmented Block Designs

✓ Recommended number of controls: 3 - 4

Augmented Block Designs

✓ Recommended number of controls: 3 - 4

Example

B1	G2	G3	G1	T2	G7	T1	G6	T3	G4
B2	G12	G10	T2	G11	G8	T1	T3	G13	G9
B3	G18	T3	G16	T1	G14	G19	G17	T2	G15
B4	G23	G25	G20	T1	T2	G24	G22	T3	G21
B5	G27	G26	G31	T2	G28	T3	G30	T1	G29

Data collection tool - Drone

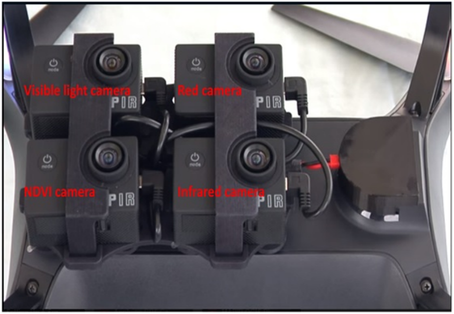


Figure: Drone and its main parts

Data collection tool - Drone

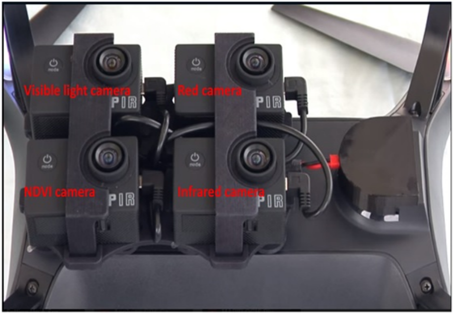


Figure: Drone and its main parts

Data collection tool - Drone

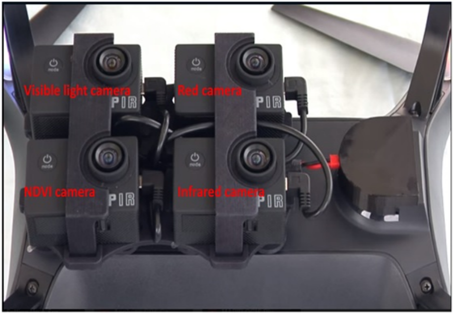


Figure: Drone and its main parts

✓ Visible light camera: RGB

Data collection tool - Drone

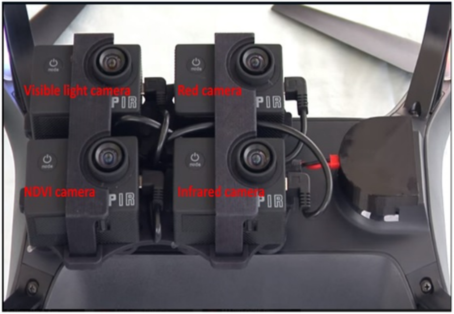


Figure: Drone and its main parts

- ✓ Visible light camera: RGB
- ✓ NDVI camera: difference between plants

Data collection tool - Drone

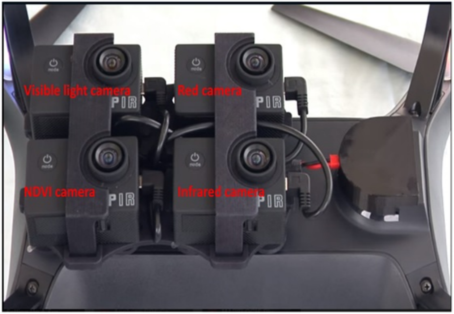


Figure: Drone and its main parts

- ✓ Visible light camera: RGB
- ✓ NDVI camera: difference between plants
- ✓ Infrared camera: adjustment of soil noise

Data collection tool - Drone

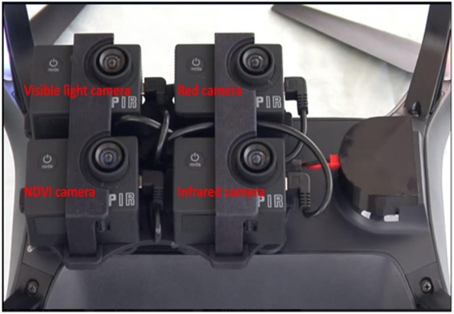


Figure: Drone and its main parts

- ✓ Visible light camera: RGB
- ✓ NDVI camera: difference between plants
- ✓ Infrared camera: adjustment of soil noise
- ✓ Red camera: infrared

Ground control points

Figure: reference point in the field

Ground control points

Figure: reference point in the field

- ✓ Points of known coordinates in the area

Ground control points

Figure: reference point in the field

- ✓ Points of known coordinates in the area
- ✓ Helps the image assembly

Ground control points

Figure: reference point in the field

- ✓ Points of known coordinates in the area
- ✓ Helps the image assembly
- ✓ Error correction (10 meters)

Ground control points

Figure: reference point in the field

- ✓ Points of known coordinates in the area
- ✓ Helps the image assembly
- ✓ Error correction (10 meters)
- ✓ 8 - 12 points

Ground control points

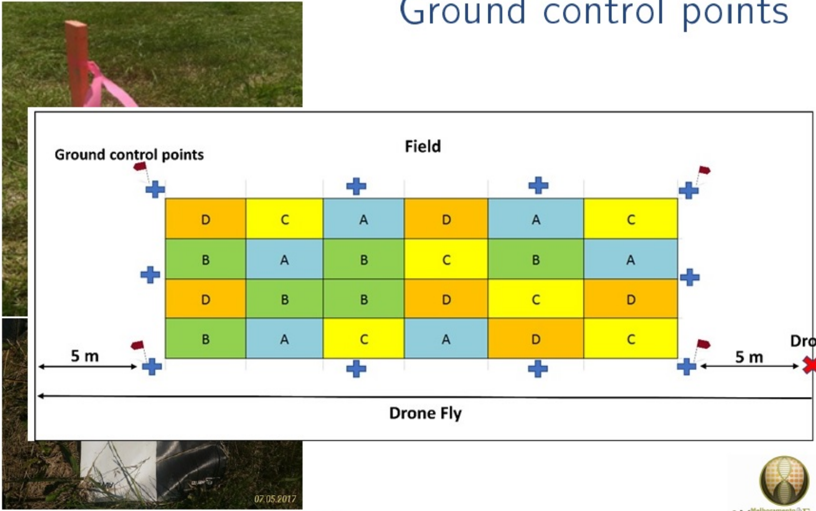


Figure: reference point in the field

Ground control points

Figure: reference point in the field

- ✓ Points of known coordinates in the area
- ✓ Helps the image assembly
- ✓ Error correction (10 meters)
- ✓ 8 - 12 points

Image sharpness

Image sharpness

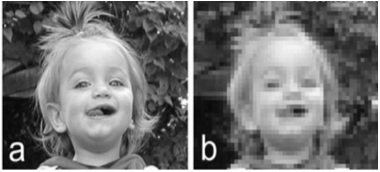


Image sharpness

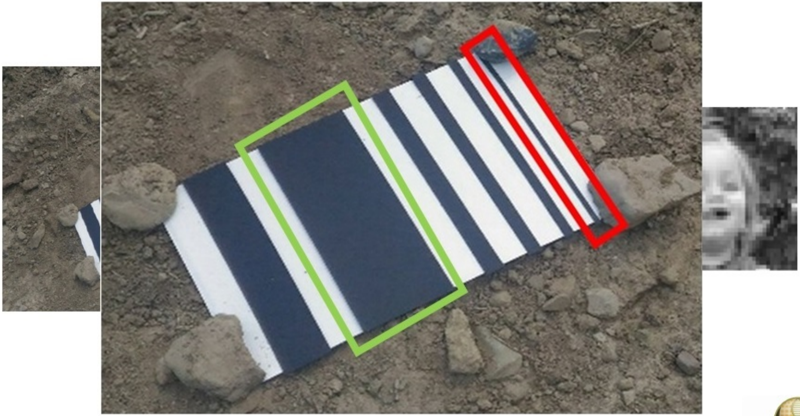
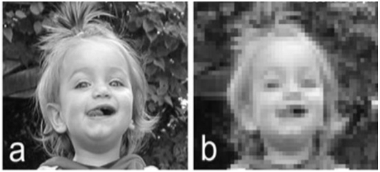
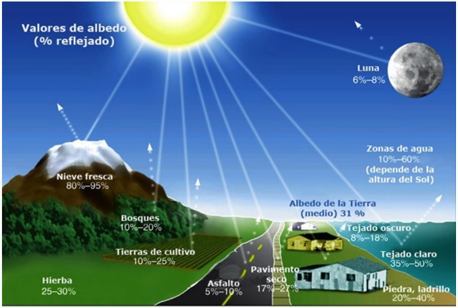


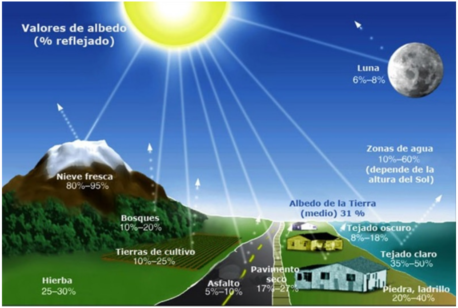
Image sharpness



Reflectance

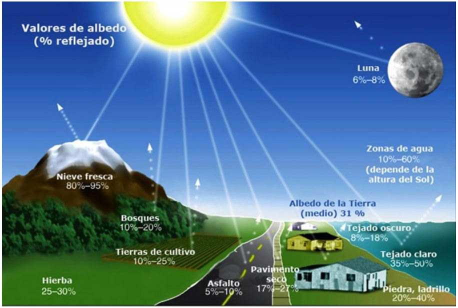


Reflectance



- ✓ Sun position must be 90 graus

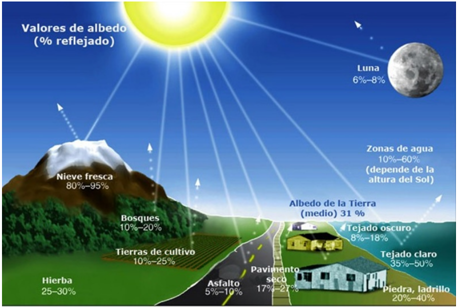
Reflectance



- ✓ Sun position must be 90 graus
- ✓ Albedo is the percentage of reflectance

Reflectance

Reflectance



- ✓ Sun position must be 90 graus
- ✓ Albedo is the percentage of reflectance

Weather station

Figure: Weather station

Weather station

Figure: Weather station

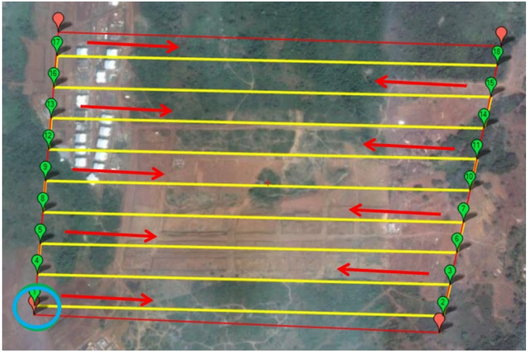
✓ Improves the aquisition of environmental variables

Weather station

Figure: Weather station

- ✓ Improves the aquisition of environmental variables
- ✓ Helpfull to correct the enviorenmental variables

Flight course start and path



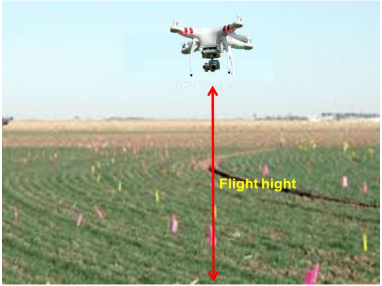
→ Flight path

○ Start of the flight course

Flight hight

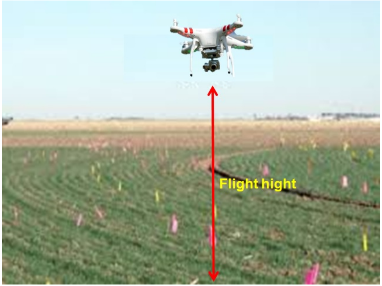
Flight hight

- ✓ Image quality and quantity are influenced by flight hight



Flight hight

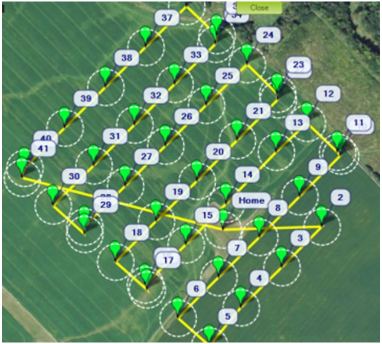
- ✓ Image quality and quantity are influenced by flight hight
- ✓ 10 - 12 meters: more accurate image



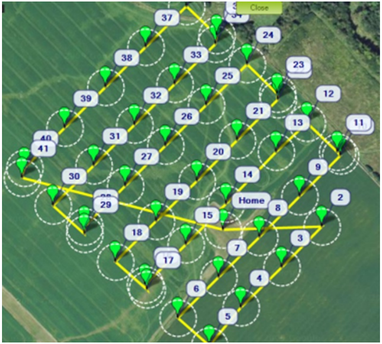
Flight hight

- ✓ Image quality and quantity are influenced by flight hight
- ✓ 10 - 12 meters: more accurate image
- ✓ 12 - 20 meters: PLH, NDVI, CT

Factors that influence flight range

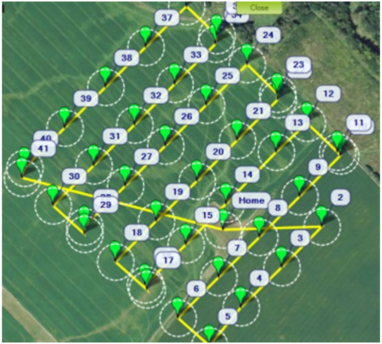


Factors that influence flight range



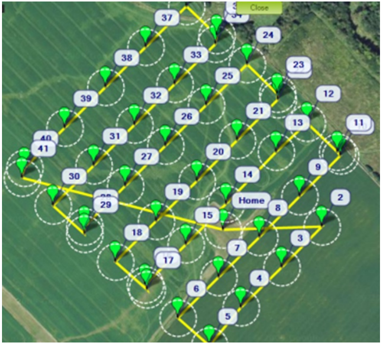
✓ Flight hight

Factors that influence flight range



- ✓ Flight hight
- ✓ Batery charge (15 - 20 minutes)

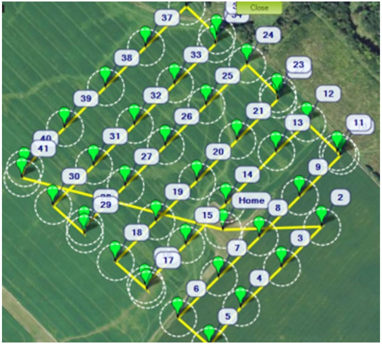
Factors that influence flight range



- ✓ Flight hight
- ✓ Batery charge (15 - 20 minutes)
- ✓ 70 - 75% of image overlap

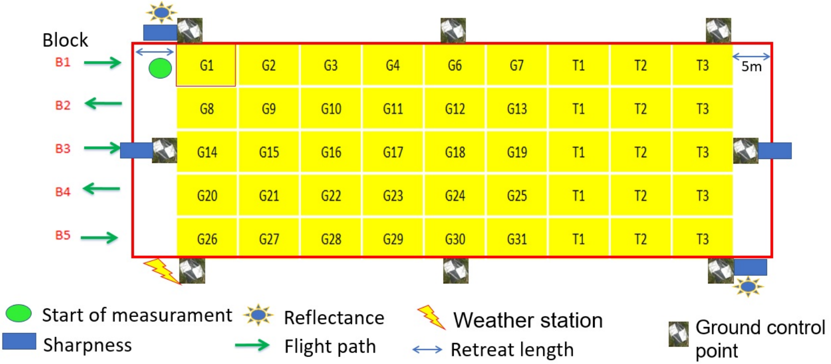
Alógamas Melhoramento ESALQ

Factors that influence flight range



- ✓ Flight hight
- ✓ Batery charge (15 - 20 minutes)
- ✓ 70 - 75% of image overlap
- ✓ 1 - 3 seconds per image

Augmented Blocks



Statistical Model

$Y = \mu + Block + Treat +$  $+$  $+$  $+ \varepsilon$

- ✓ μ: constant
- ✓ Block: effect inside the statistcal model
- ✓ Treat: treatment effect ($G_i + T_i$) inside the statistical model
- ✓ : Sharpness covariable (gradient information)
- ✓ : Reflectance covariable (enviorenment information)
- ✓ : Weather station covariable (soil and climate information)

Statistical Model

- ✓ μ: constant
- ✓ Block: effect inside the statistcal model
- ✓ Treat: treatment effect ($G_i + T_i$) inside the statistical model
- ✓ : Sharpness covariable (gradient information)
- ✓ : Reflectance covariable (enviorenment information)
- ✓ : Weather station covariable (soil and climate information)

Melhoramento Alógamas ESALQ

Statistical Model

$Y = \mu + \text{Block} + \text{Treat} +$  $+$  $+$  $+ \varepsilon$

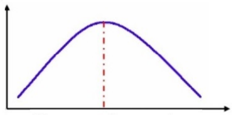
- ✓ μ: constant
- ✓ Block: effect inside the statistcal model
- ✓ Treat: treatment effect ($G_i + T_i$) inside the statistical model
- ✓ : Sharpness covariable (gradient information)
- ✓ : Reflectance covariable (envriorenment information)
- ✓ : Weather station covariable (soil and climate information)

Position measures

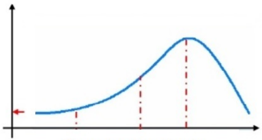
- ❑ Mean
- ❑ Median
- ❑ Mode

- Normal
- t-Student

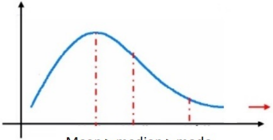
Symetric distribution



Asymmetric left-skewed distribution (or negative asymmetry)



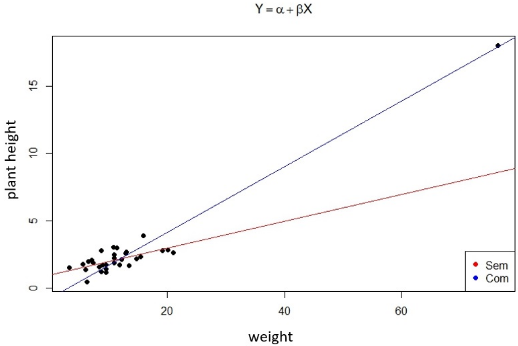
Asymmetric right-skewed distribution (or positive asymmetry)



- Asymmetric distributions: Qui-quadrado; F-Snedecor; Weibull; Poisson; Lognormal

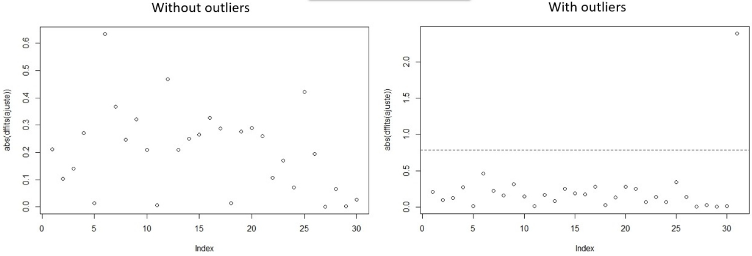
What happens when we have an experiment with distilliers values?

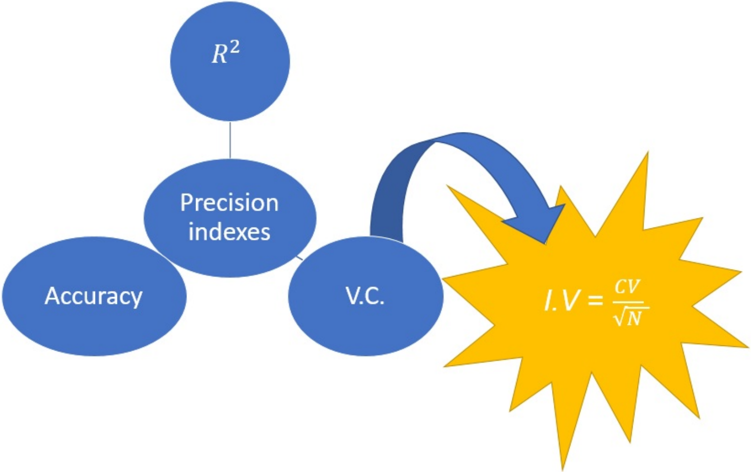
```
$media_x_sem
[1] 10.985

$media_x_com
[1] 13.69097

$media_y_sem
[1] 2.066333

$media_y_com
[1] 2.580323
```



What are the outliers effects in true value?

DFFits




R^2
Precision indexes
Accuracy
V.C.
$I.V = \frac{CV}{\sqrt{N}}$

The End

THANK YOU FOR YOUR ATTENTION!