



THE UNIVERSITY
of EDINBURGH



Biotechnology and
Biological Sciences
Research Council



Day 4

Spatial variation and genotype by environment interaction

Design of plant breeding field trials

Daniel Tolhurst, Jon Bancic, Chris Gaynor, Gregor Gorjanc



Lecture overview

1. Design of plant breeding field trials ←

- Fundamental concepts of experimental design
- Classical and model-based designs

2. Linear mixed models for plants

- Complex residual variance structures
- Spatial variation

A wide-angle photograph of a vast agricultural field, likely corn, stretching towards a horizon under a dramatic sunset or sunrise sky. The sky is filled with warm, golden, and orange hues, with darker clouds on the right side. In the far distance, there are a few small clusters of trees. The foreground is dominated by the green and yellow foliage of the crops.

Design of plant breeding field trials

Outline

- **Background**
- **Fundamental concepts of experimental design**
 - Randomisation
 - Replication
 - Blocking
- **Classical and model-based designs**

Plant breeding – a series of tests + selections

Year	Stage	Population	Env	h^2	Action
1	Crossing	$P_i \times P_j$	100 crosses		Crosses with 100 parents
2	F_1/DH	100 families			Produce 100 DHs per cross
3	HDRW	10,000	1	0.10	Grow DHs
4	PYT	500	2	0.20	Field trial
5	AYT	50	5	0.50	Field trial
6	EYT	10	20	0.70	Field trial
7	Variety		1		Release inbred variety

The diagram illustrates the plant breeding process through a series of stages:

- Year 1:** Crossing ($P_i \times P_j$) leads to 100 crosses, each producing 100 DHs (Doubled Haploid). This stage involves crosses with 100 parents.
- Year 2:** The 100 crosses produce 100 families of DHs. This stage involves producing 100 DHs per cross.
- Year 3:** The population is 10,000 individuals. This stage involves growing DHs.
- Year 4:** The population is 500 individuals. This stage involves a field trial.
- Year 5:** The population is 50 individuals. This stage involves a field trial.
- Year 6:** The population is 10 individuals. This stage involves a field trial.
- Year 7:** The final stage is the release of the inbred variety.

Environmental factors (Env) and heritability (h^2) values are provided for each stage:

- Year 1: Env 1, h^2 -
- Year 2: Env - , h^2 -
- Year 3: Env 1, h^2 0.10
- Year 4: Env 2, h^2 0.20
- Year 5: Env 5, h^2 0.50
- Year 6: Env 20, h^2 0.70
- Year 7: Env 1, h^2 -

Plant breeding field trials

- Formally designed comparative field experiments



Source: B. Ovenden

Plant breeding field trials



Source: Crop Circle Consulting and Research

Plant breeding field trials



Source: R.C. Gaynor

Plant breeding field trials



Source: R.C. Gaynor

Plant breeding field trials



Source: Top crop manager

Plant breeding field trials



Source: B. Ovenden

Comparative field experiments

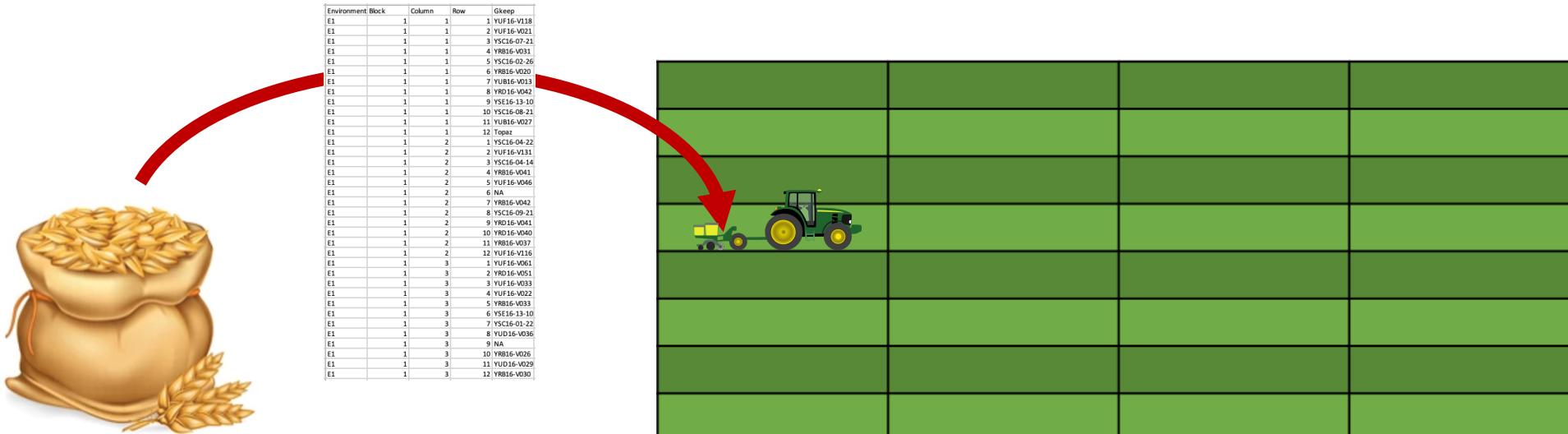
- **Field layout**
 - Field columns and rows
 - Plot lengths and widths
 - Number of rows per plot and seed rate
- **Experimental design**
 - Allocation of treatments to plots



Source: Crop Circle Consulting and Research

What is experimental design?

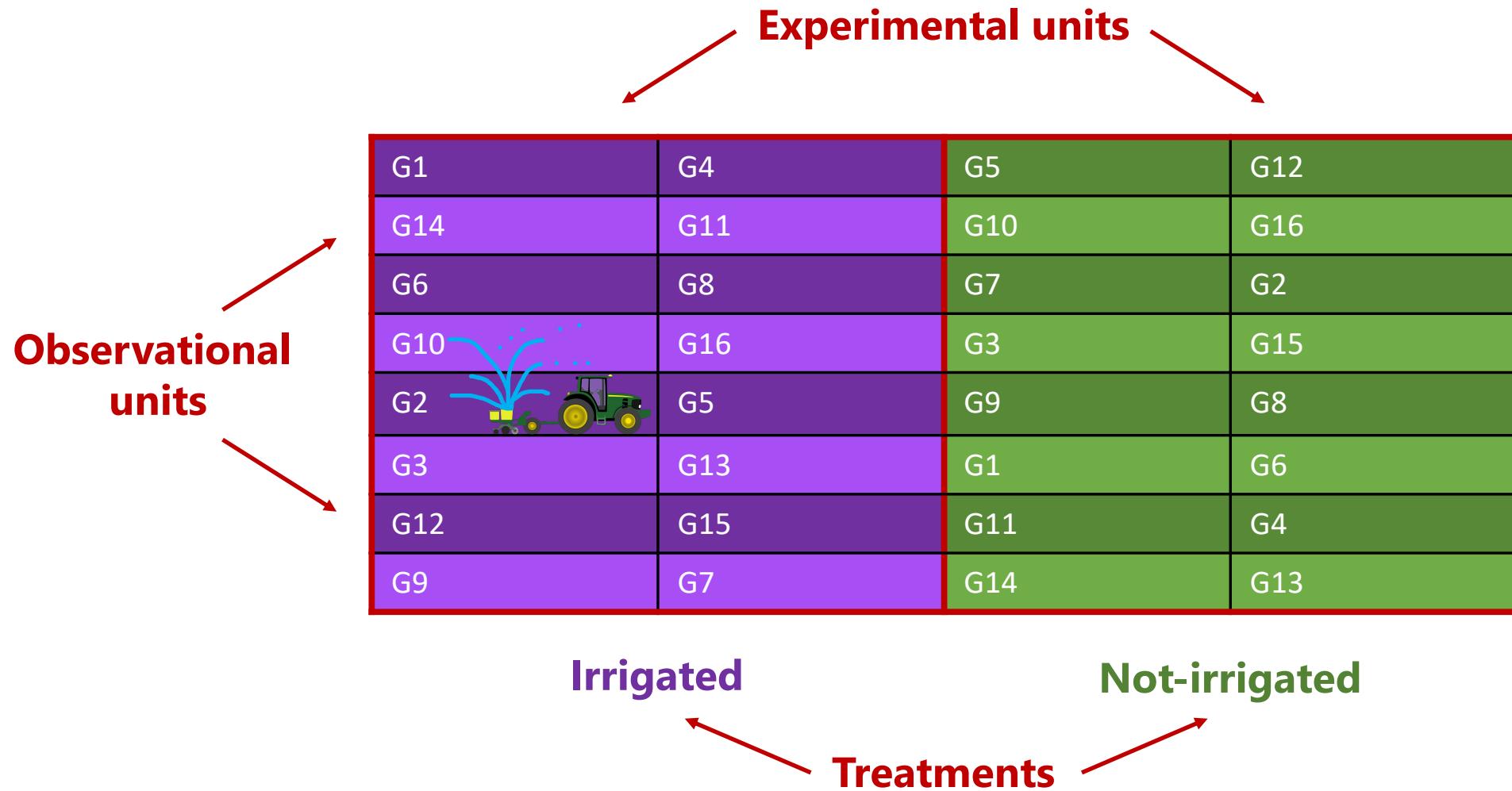
- Allocation of treatments to experimental units based on a design function, chosen to satisfy some criterion
- Requires definition of key concepts!



Some definitions

- **Experimental unit (EU)**
 - Smallest unit to which a treatment is applied (e.g. blocks or plots)
- **Observational unit (OU)**
 - Smallest unit to which a measurement is taken (e.g. plants within plots or plots)
- **Treatment**
 - Entire description of what is applied to the experimental units (e.g. irrigation/fertiliser level or genotype)

EUs, OUs and treatment



Some concepts

- **Randomisation**
 - Required to obtain unbiased estimates of error variance
 - Sources of bias: systematic, selection, accidental and cheating
- **Replication**
 - Required to estimate sources of error (including error variance)
 - Achieved through true replication of the EUs
- **Blocking**
 - Historically used in field trials to mitigate spatial variability
 - Balanced, complete, resolvable → orthogonality

Systematic bias

- e.g. allocating genotypes 1:16 to plots 1:16 in all trials within a breeding programme

G1	G9	G5	G12
G2	G10	G10	G16
G3	G11	G7	G2
G4	G12	G3	G15
G5	G13	G9	G8
G6	G14	G1	G6
G7	G15	G11	G4
G8	G16	G14	G13

Selection bias

- e.g. placing all genotypes from a preferred family into areas with higher fertility

High fertility
↑
Low fertility

G1	G3	G2	G4
G2	G4	G3	G1
G14	G8	G7	G5
G11	G16	G12	G10
G5	G6	G9	G8
G10	G13	G15	G6
G12	G15	G11	G16
G9	G7	G14	G13

Accidental bias

- e.g. placing family groups of genotypes together, based on delivery of seed packets

G1	G2	G3	G4
G14	G11	G10	G16
G6	G8	G7	G12
G10	G16	G5	G15
G12	G5	G9	G8
G7	G13	G14	G6
G12	G15	G11	G13
G4	G3	G2	G1

Cheating

- **Cheating** by the experimenter
 - May involve some combination of the previous examples

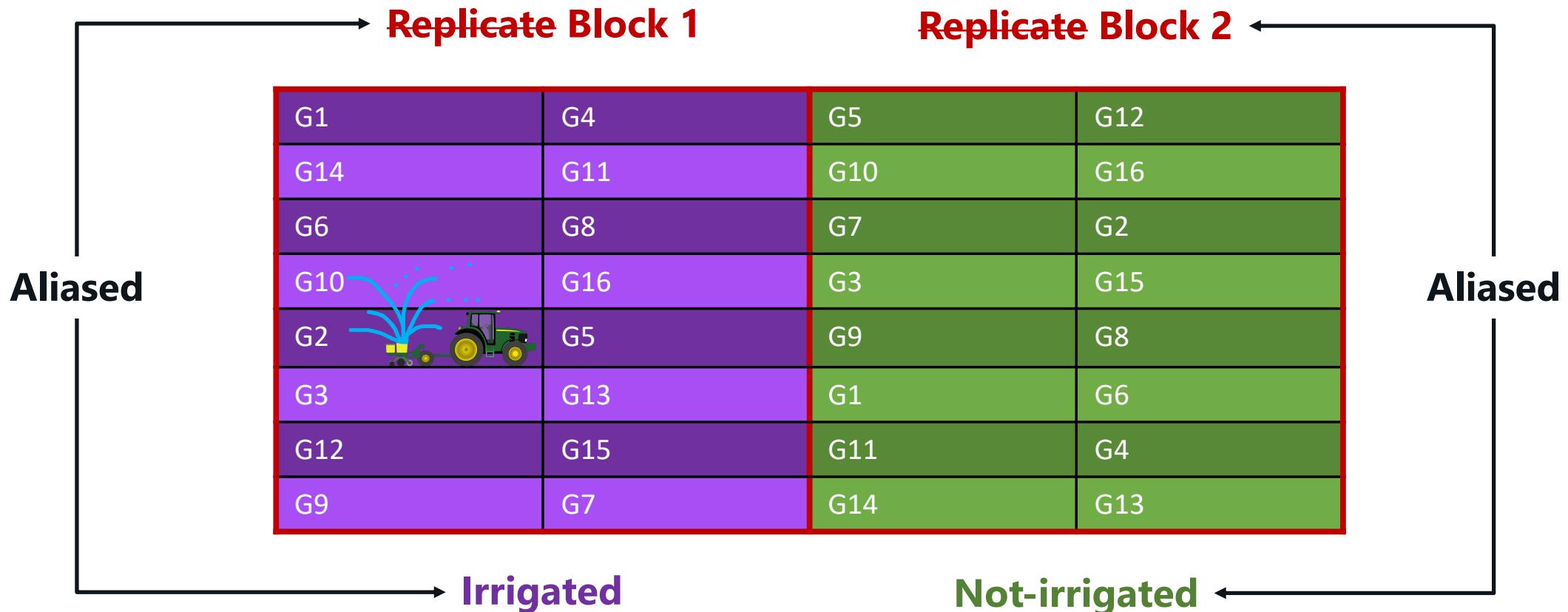


Randomisation, replication and blocking

Replicate Block 1		Replicate Block 2	
G1	G4	G5	G12
G14	G11	G10	G16
G6	G8	G7	G2
G10 	G16	G3	G15
G2 	G5	G9	G8
G3	G13	G1	G6
G12	G15	G11	G4
G9	G7	G14	G13

Irrigated **Not-irrigated**

Pseudo/false-replication



Cannot disentangle treatment and block effects → cannot perform significance testing

Classical experimental designs

Randomised complete block (RCB) design

$$y_{ij} = \mu + g_i + b_j + e_{ij}$$

phenotype mean genotype block residual

- Simple to construct
- Balanced, complete and resolvable
- Genotypes and blocks are orthogonal
- **But**, assumes blocks are homogeneous

Block 1		Block 2	
G1	G4	G5	G12
G14	G11	G10	G16
G6	G8	G7	G2
G10	G16	G3	G15
G2	G5	G9	G8
G3	G13	G1	G6
G12	G15	G11	G4
G9	G7	G14	G13

Resolvable row-column design

$$y_{ijklm} = \mu + g_i + b_{c_j} + c_k + b_{r_l} + r_m + e_{ijklm}$$

phenotype mean genotype ColBlock Column RowBlock Row residual

- Unbalanced (or balanced, see latinised designs)
- Column / row blocks are complete and resolvable
- Genotypes and column / row blocks are orthogonal

	ColBlock 1			ColBlock 2		
	1	2	3	4	5	6
Row	1	2	3	4	5	6
RowBlock 1	G1	G4	G5	G2		
	G14	G11	G16	G9		
	G6	G8	G7	G12		
	G10	G15	G13	G3		
RowBlock 2	G2	G5	G4	G10		
	G3	G13	G11	G14		
	G12	G16	G6	G8		
	G9	G7	G15	G1		
Column	1	2	3	4		

Split-plot design

$$y_{hijk} = \mu + t_h + g_i + b_j + m_k + e_{hijk}$$

phenotype mean irrigation genotype block MainPlot residual

- Irrigation to main plots
Genotypes to sub plots
- Blocks are complete and resolvable wrt irrigation
- Irrigation and blocks are orthogonal

	MainPlot 1	Block 1	Block 2	MainPlot 3	MainPlot 4
MainPlot 1	G1	G5	G8	G3	
MainPlot 2	G4	G2	G6	G5	
MainPlot 3	G6	G8	G7	G2	
MainPlot 4	G3	G7	G1	G4	
MainPlot 1	G2	G4	G3	G6	
MainPlot 2	G7	G1	G5	G8	
MainPlot 3	G5	G3	G2	G1	
MainPlot 4	G8	G6	G4	G7	
	Irrigated			Not-irrigated	

Split-plot design

$$y_{hijk} = \mu + t_h + g_i + b_j + m_k + e_{hijk}$$

phenotype **mean irrigation genotype** **block** **MainPlot** **residual**

- Irrigation to main plots.
 - **Genotypes to sub plots**
 - Main plots are complete and resolvable wrt genotypes
 - Genotypes and main plots are orthogonal

	MainPlot 1	Block 1	Block 2	MainPlot 3	MainPlot 4
MainPlot 1	G1	G5	G8	G3	
MainPlot 2	G4	G2	G6	G5	
MainPlot 3	G6	G8	G7	G2	
MainPlot 4	G3	G7	G1	G4	
MainPlot 5	G2	G4	G3	G6	
MainPlot 6	G7	G1	G5	G8	
MainPlot 7	G5	G3	G2	G1	
MainPlot 8	G8	G6	G4	G7	

Other classical designs

- Completely randomised (CR) design
- Balanced and unbalanced incomplete block (IB) designs
- Augmented grid-plot design (Federer, 1956)
- Strip plot design (Cochran and Cox, 1957)
- Latin squares (sudoku)
- Latinized rectangular lattice design (Harshbarger and Davis, 1952)
- α lattice design (Patterson and Williams, 1976)
- Latinized α lattice design (Williams, 1986)
- t -latinized design (John and Williams, 2002)

A photograph of a field of corn plants. The plants are tall with long, narrow green leaves. At the top of each plant is a yellow tassel. The background is slightly blurred, showing more of the field extending into the distance.

Model-based designs

Partially replicated (p-rep) design

- Motivated by augmented designs (Federer, 1956)
 - Add replicated checks to unreplicated trials
 - Comprise grid plots with checks inside each grid
 - 20% of plots used for checks 8/40 plots
 - Can be wasteful
- *p*-rep designs
 - Efficient
 - Remove checks and replicate a percentage of genotypes 32, 8 spare
 - $p = 25\%$ replication level 8/32 genos

Partially replicated (prep) design

$$y_{ij} = \mu + g_i + b_j + c_k + r_k + e_{ij}$$

phenotype mean genotype block column row residual

- Unbalanced and incomplete
- **Replicated** genotypes occur once in each block
- **Unreplicated** genotypes are spread across blocks
- Optimality criterion used to allocate genos to plots

		Block 1	Block 2	
		G1	G2	
Row	1	G20	G23	
2	G10	G14	G25	G3
3	G16	G8	G28	G32
4	G4	G11	G7	G6
5	G19	G5	G21	G27
6	G6	G13	G29	G31
7	G12	G2	G8	G24
8	G9	G18	G22	G1
9	G3	G15	G5	G30
10	G17	G7	G26	G4
		1	2	3
		Column		4

Other model-based designs

- Augmented p -rep designs (Williams et al., 2011)
- Multi-phase experiments (Smith et al., 2006)
- Sparse testing (Jarquin et al., 2020)
- Designs using genetic relatedness (Cullis et al., 2020)

Lecture overview

1. Design of plant breeding field trials ←

- Fundamental concepts of experimental design ✓
- Classical and model-based designs ✓

2. Linear mixed models for plants

- Complex residual variance structures
- Spatial variation