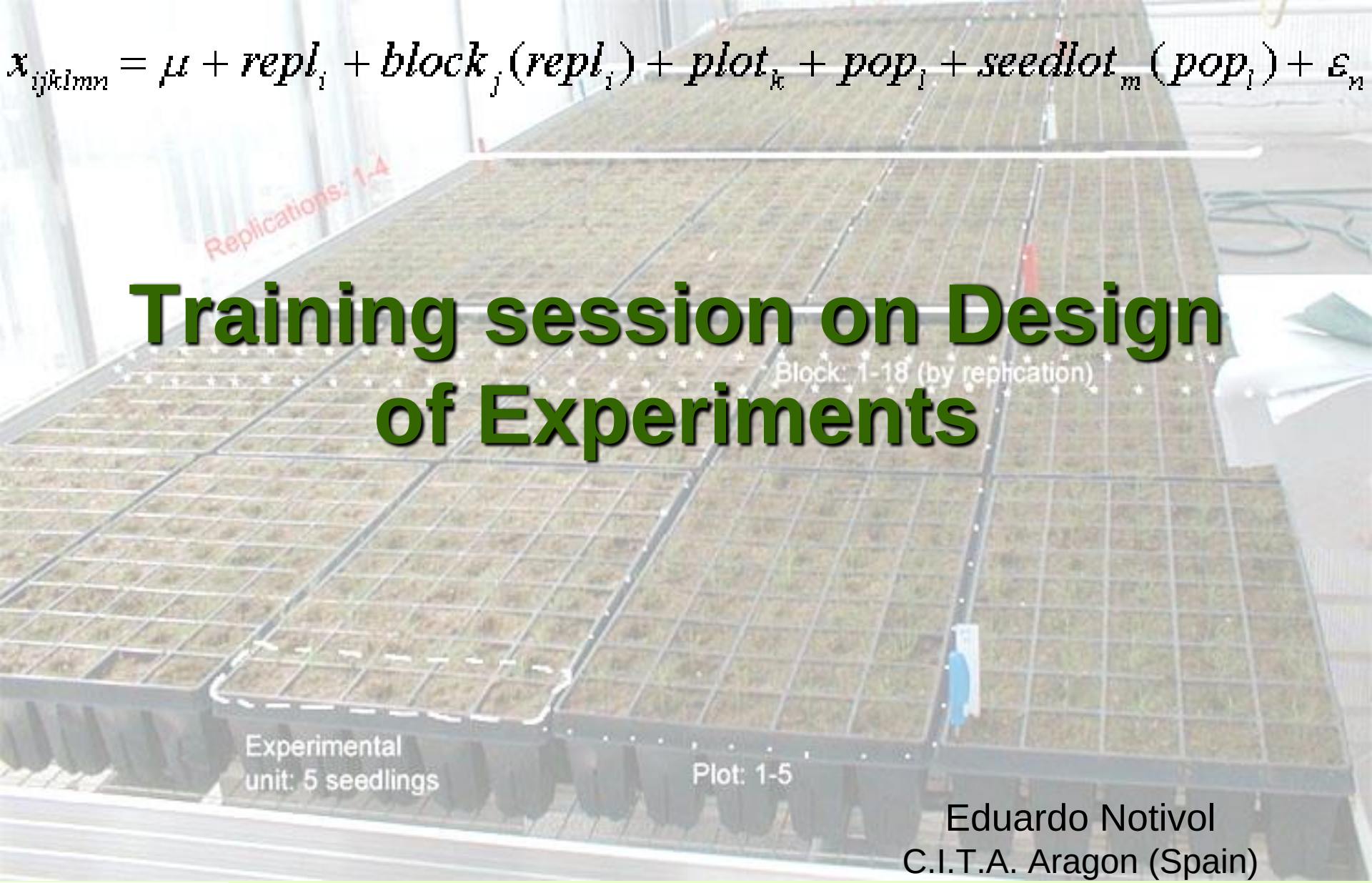
$$x_{ijklmn} = \mu + repl_i + block_j(repl_i) + plot_k + pop_l + seedlot_m(pop_l) + \varepsilon_n$$



Training session on Design of Experiments

Eduardo Notivol
C.I.T.A. Aragon (Spain)
Unidad de Recursos Forestales
enotivol@aragon.es

GOBIERNO DE ARAGON

cita
CENTRO DE INVESTIGACIÓN Y TECNOLOGÍA AGROALIMENTARIA DE ARAGÓN

Sketch out

- Introduction
- Stages
- Principles
- Designs and analysis theory
- Operational tools for design
- [Analysis of experiments]

OUTLINE

- Introduction
- Stages
- Principles
- Designs and analysis theory
- Operational tools for design
- [Analysis of experiments]

- DoE ← → AoE
- Practical aspects
- Some statistics theory
- Operational tools for design
- Data collecting
- Preprocessing (screening, outliers)
- Checking assumptions
- [Analysis of experiments]

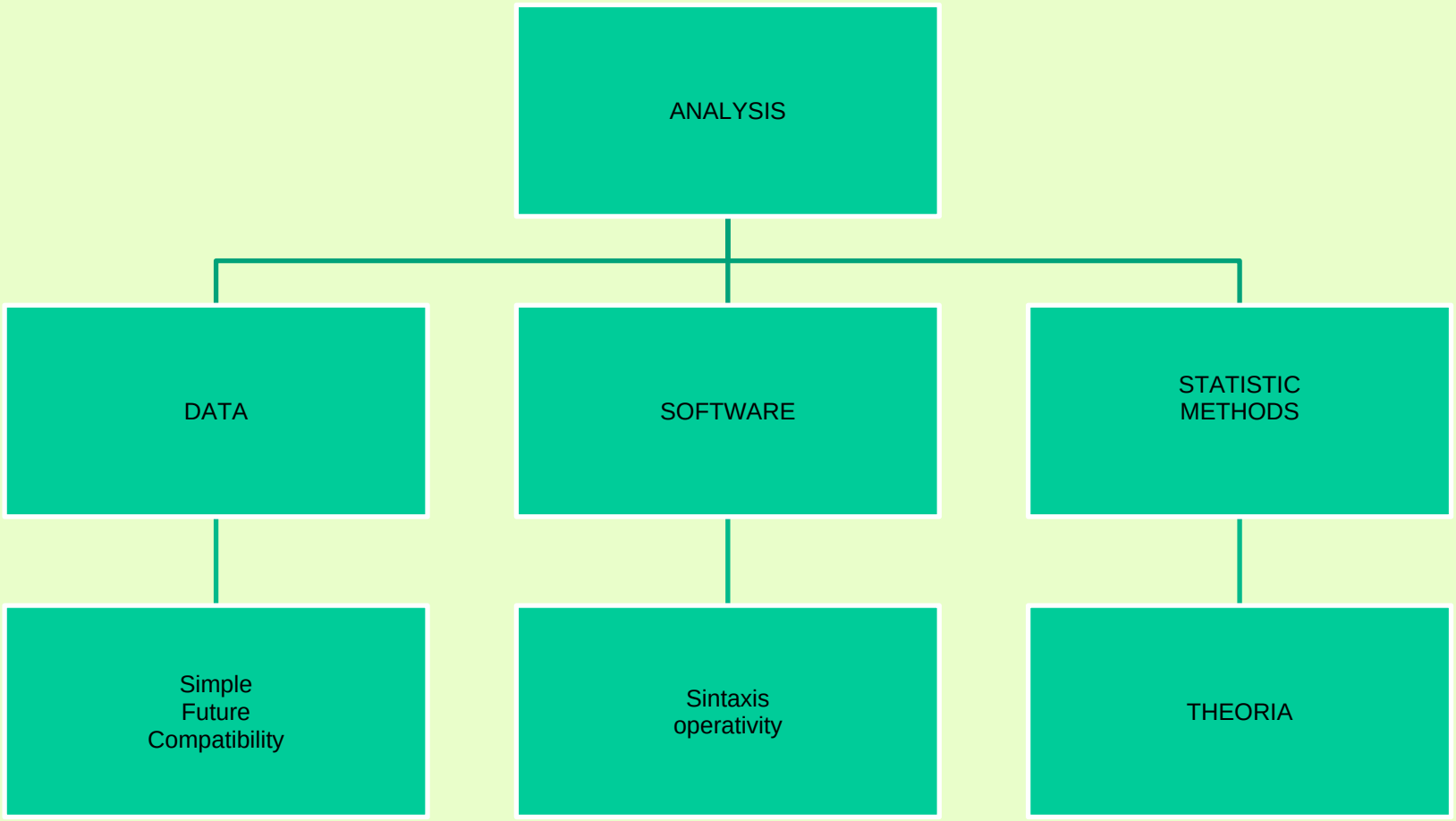
Introduction

- Goal

Tool for addressing analytical problems without fixed laws

- Variability
  - Existence
  - Dealing and understanding
  - Modeling and controlling

experimental error


ANALYSIS
DATA
SOFTWARE
STATISTIC METHODS
Simple
Future
Compatibility
Sintaxis
operativity
THEORIA

Planning of an experiment (i)

- Definition of objectives
- Definition of all sources of variation
  - Treatments and their levels
  - Experimental units
  - Nuisance factors: blocking, noise & covs
- Setting up the experimental units & treatments

Planning of an experiment (ii)

- Definition of response variable, experimental process and issues foresight
- Set up the model
- Scheme of analysis steps
- Set up sampling size
- Review all foregoing points and modifying if necessary

Stages (i)

- Definition of the problem
- Definition of objectives
- Selection of treatments to test (interactions)
- Selection of the material to test
- Selection of the experimental design (simple)
- Selection of the experimental unit size and number of replications

Stages (ii)

- Control of “surroundings” effects
- Kind of data to be taken
- Selection of statistical tests
- Accomplishment of the experiment
- Analysis and interpretation of results
- Final reporting (conclusions)

Principles (i)

1.Replicatio**n**s. (experimental error basis)

Standard Error of Difference

Agronomic trails SED<1/3 diff

Material selection SED<1/6 diff

Knowing s^2 & d ==> n

$$SED = \sqrt{\frac{2\sigma^2}{n}}$$

2. Treatment (broad sense) **R**andomization

3. Local control of existing variation in trial site

(**B**locking or spatial analysis)

ANOVA / SDA / Robust methods

“typical numbers & expressions”

4 replications

25 plants per plot

4 treatments

randomization

latinization

3 sites

2 border lines

RCB

25 genotypes

5 x 5 m spacing

single tree plot

Principles (ii)

Operational Limiting factors

Number of available effectives
Site constraints (topography, surface …)
Technical limitations (machinery, ….)
Measurements
Competence, specific needs,
Future treatments, thinings,…
Spacing, density

Experimental design

Initial assumptions or constraints:

Additivity
Normality
Homocedasticity.

Different treatment errors are independient & distributed $N(0,\sigma^2)$

Statistic tests:

N: Shapiro-Wilks, graphs distrib, freq acum., res * pred
H: Barlett, Levenne, ratios variances

Transformations

No parametric methods

Elementary Designs

C.R.

Model: $y_{ij} = \mu + t_i + \varepsilon_{ij}$

	dof	SS	MS	F	EMS
Total	rt-1	a	a/rt-1		
Treat	t-1	b	b/t-1	MS_T / MS_E	$\sigma^2_e + r\,\sigma^2_t$
Error	t(r-1)	c	c/t(r-1)		σ^2_e

Model y = treat;

Elementary Designs

R.C.B

Model: $y_{ijk} = \mu + t_i + b_j + \varepsilon_{ijk}$

	dof	SS	MS	F	EMS
Total	rb-1	a	a/rb-1		
Treat	t-1	b	b/t-1	MS_T / MS_E	$\sigma^2_e + b\,\sigma^2_t$
Blq	b-1	c	c/b-1	MS_B / MS_E	$\sigma^2_e + t\,\sigma^2_b$
Error	t-1)(r-1)	d	d/t(r-1)		σ^2_e

Model y = treat blq;

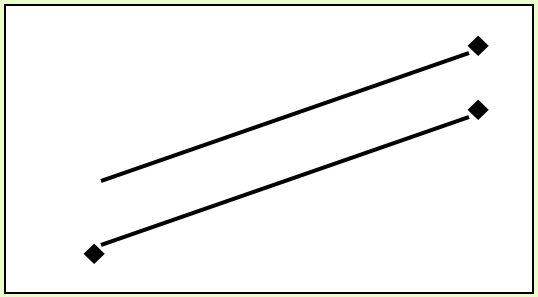
Experimental design

Possible structure of treatments

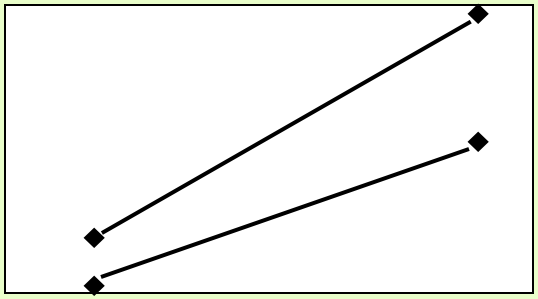
Factorial: total combination all x all

Possibility interactions study (GxE)

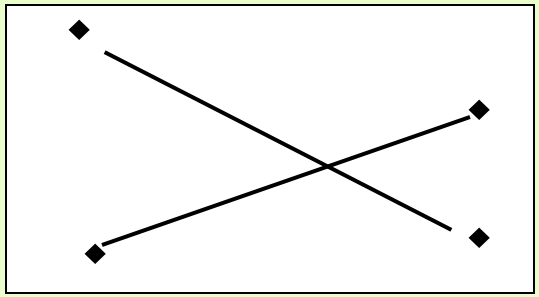
Reaction norms $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$



No interaction



quantitative interaction



qualitative interaction

Experimental design

Possible structure of treatments

Hierarquical or nested: Impossible combination

$$y_{ijk} = \mu + \alpha_i + \beta_j(\alpha_i) + \varepsilon_{ijk}$$

Model y = pop fam(pop);

Is it important the treatment structure ?

F E.M.S.

other structure + important.....

¿Fixed o Random?

F E.M.S.

1. Critic decision
2. Not well documented on texts
3. Usually based on subjective statistic agreements

Fixed: Levels of factor clearly targeted or selected
Results & conclusions from anova are for these levels
Main aim: Mean estimation of the variable for each level
(BLUE)

Random: Levels are a random sample from all posible.
Results & conclusions from anova can be extrapolated + levels
Main aim: Variability estimation of the variable or factor
or perhaps prediction at a given level
(BLUP)

	dof	MS	A y B fixed	A y B rand	A:fix B:rand
Total	abr-1				
A	a-1	MS_A	MS_A / MS_E	MS_A / MS_{AB}	MS_A / MS_{AB}
B	b-1	MS_B	MS_B / MS_E	MS_B / MS_{AB}	MS_B / MS_E
AxB	(a-1)(b-1)	MS_{AB}	MS_{AB} / MS_E	MS_{AB} / MS_E	MS_{AB} / MS_E
Error	ab(r-1)	MS_E			

$\sigma^2_e + c_1 \Phi_\alpha$ | $\sigma^2_e + n \sigma^2_{ab} + nb \sigma^2_a$

¿Fixed o Random?

How to asses?

A PRIORI

Scientific Criteria :

1) is it possible to repeat the factor levels in other site or year?

2) has it meaning this replication?

Yes + Yes = Fixed

Statistic Criteria :

“Random” few levels (3-5) =>weak variance estimation, Better setting as fixed and use the results only at these levels

“Fixed” with many levels (>10) without structure, better setting as random and estimating means by BLUPs

E.M.S. Numeric difficulty

¿Fixed o Random?

GLM

Model y =loc blq(loc) var var*loc;
Random loc blq(loc) var*loc /test;

1º Calculation as fixed
2º Calculation EMS
3º Repeat F-tests with proper denominators

MIXED MODELS

Incomplete Blocks

Evaluation : high nº genotypes
limited material

'Many genotypes' means huge blocks # no control

I.B. Not all treat by block, so several blocks are needed for a complete replication

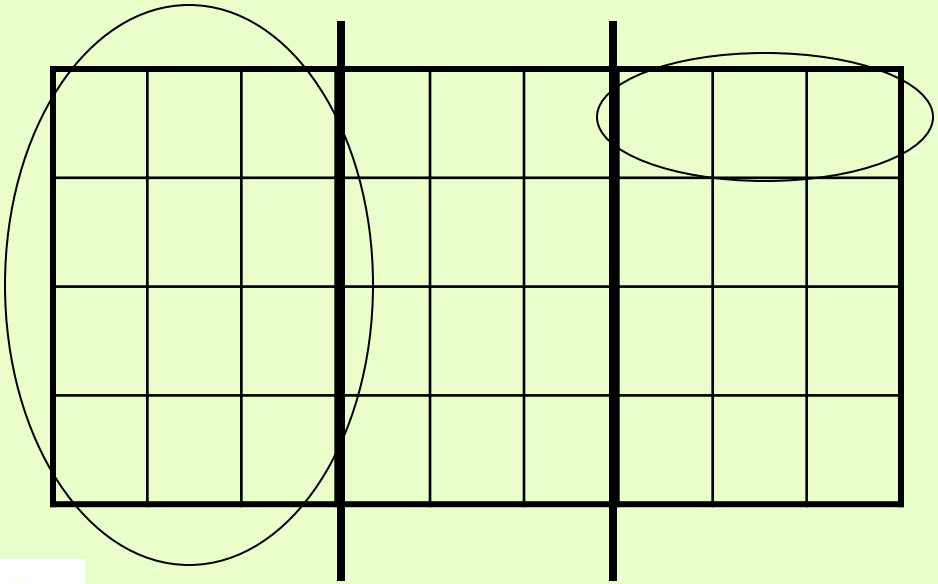
.	.	B	A	.	*1*
.	.	A	C	.	*2*
B	.	.	.	C	*3*

Based on
Aditivity: $B\text{-}C = (B\text{-}C)_3$
$= (B\text{-}A)_1 - (C\text{-}A)_2$
Experimental error independent of treatment

Incomplete Blocks

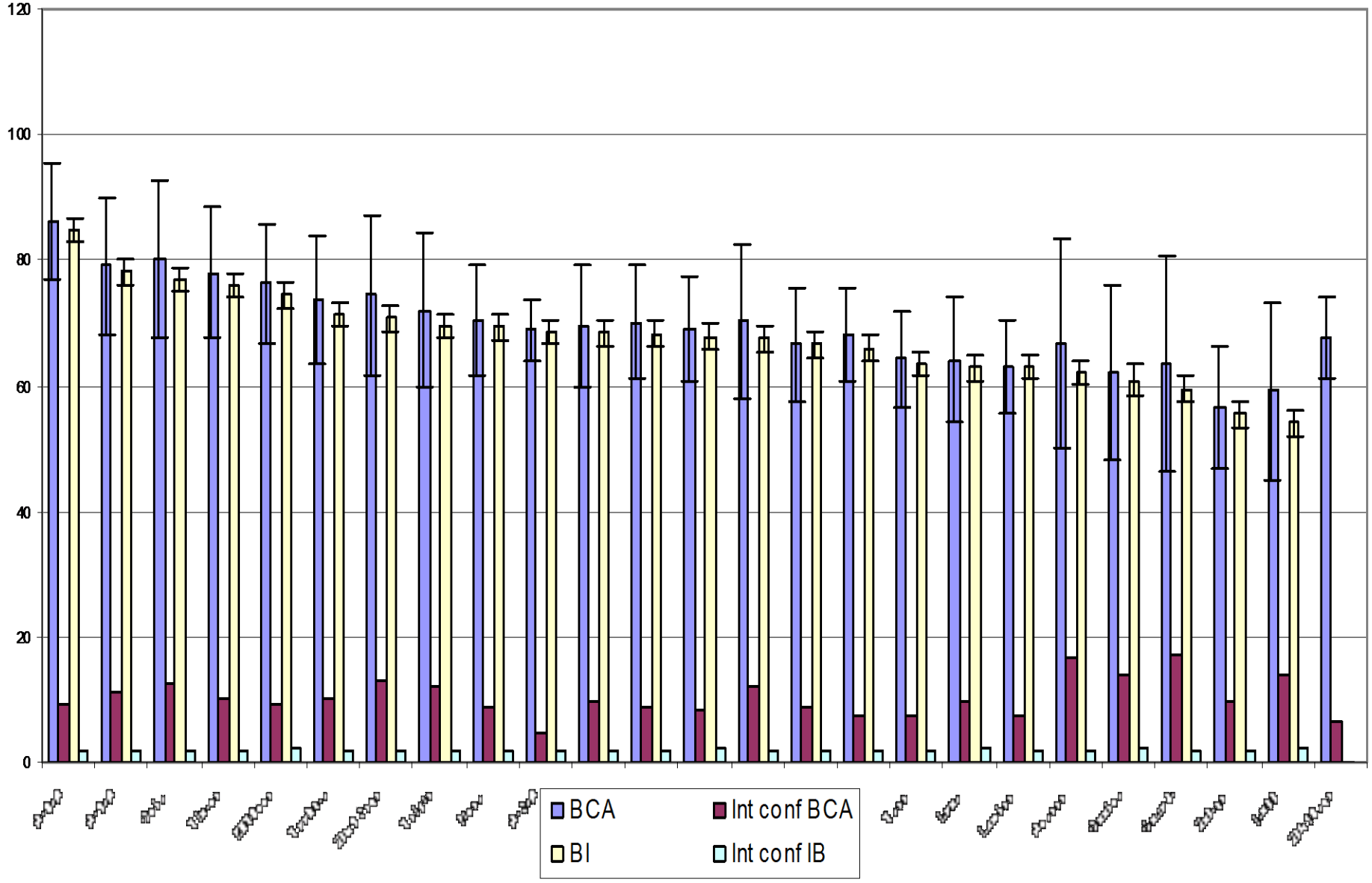
- Coexist direct & indirect comparisons
- Lost of accuracy on indirect comparisons
  but experimental error reduction

Resolvable designs
i.e.: g=k bi
α-latice, latinized, row-columns,...



Complex
specific Software

interblock info


120
100
80
60
40
20
0
BCA
Int conf BCA
BI
Int conf IB

$$x_{ijklmn} = \mu + repl_i + block_j(repl_i) + plot_k + pop_l + seedlot_m(pop_l) + \varepsilon_n$$

Mixed model REML BLUP & BLUE