Formulating mixed models for experiments, including longitudinal experiments
(accepted for publication in JABES)

Chris Brien[1] & Clarice Demétrio[2]

[1]University of South Australia, [2]ESALQ, Universidade de São Paulo

Web address for Multitiered experiments site:
http://chris.brien.name/multitier

Chris.brien@unisa.edu.au

Outline

1. Preliminaries
2. A longitudinal Randomized Complete Block Design (RCBD)
3. Why randomization-based models?
4. A three-phase example
5. Concluding comments

1) Three-stage method

(motivated by Piepho et al., 2004); extension of Brien and Bailey, 2006, section 7)

I. **Intratier Random and Intratier Fixed models:**
Essentially models equivalent to a randomization model.

↓

II. **Homogeneous Random and Fixed models:**
Terms added to intratier models and others shifted between intratier random and intratier fixed models.

↓

III. **General Random and General Fixed models:**
Perhaps reparameterize terms in homogeneous models, particularly if a longitudinal experiment, and omit aliased terms from random model.

- May yield a model of convenience, not full mixed model.
- Demonstrate by example

Tiers

- Randomization-based tiers are the foundation of the method.
- View randomization as the assignment of one set of objects to another.
  - e.g. treatments to plots.
- A tier is a set of factors indexing a set of objects.
- Would not be need if all experiments were two-tiered, as only two sets of factors needed:
  - block or unit or unrandomized factors;
  - treatment or randomized factors.
- Tiers is a general term for these sets
- For two-tiered factorial RCBD:
  - the units or unrandomized tier might be {Blocks, Plots} and
  - the treatments or randomized tier {A, B}

Notation

Factor relationships

A*B factors A and B are crossed
A/B factor B is nested within A

Generalized factor

A$\wedge$B is the *ab*-level factor formed from the combinations of A with *a* levels and B with *b* levels

Symbolic mixed model

Fixed terms | random terms   e.g. (A*B | Blocks/Runs)

A*B = A + B + A$\wedge$B

Blocks/Runs = Blocks + Blocks$\wedge$Runs

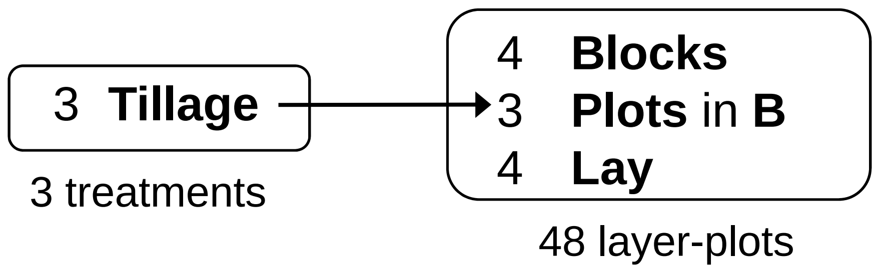
Functions on generalized factors

gf(.) generalized factor from all factors in argument.
uc(.) some, possibly structured, form of unequal correlation between levels of generalized factors.
td(.) systematic trend across levels of generalized factors.

2) A longitudinal RCBD

(Piepho et al., 2004, Example 1)

- A field experiment comparing 3 different tillage methods
- Laid out according to an RCBD with 4 blocks.
- On each plot one water collector is installed in each of 4 layers and the amount of nitrogen leaching measured.



Intratier Random and Intratier Fixed models:

- The **unrandomized tier** is {Block, Plot, Lay};
- The **randomized tier** is {Tillage}.
- The only **longitudinal factor** is Lay.

Intratier Random: (Block / Plot) * Lay
= Block + Lay + Block∧Lay + Block∧Plot + Block∧Plot∧Lay ;

Intratier Fixed: Tillage.

- So not the "Split-plot-in-Time“ analysis with

  Random: Block / Plot / Subplot;

  Fixed: Tillage * Lay.

- But, what are Subplots?

A longitudinal RCBD (cont'd)

I. **Intratier Random**: (Block / Plot) * Lay
= Block + Lay + Block∧Lay + Block∧Plot + Block∧Plot∧Lay ;
Intratier Fixed: Tillage.

- Have all possible terms given the randomization.

II. **Homogeneous Random and Fixed models:**
Terms added to intratier models and others shifted from intratier random to intratier fixed models and vice versa.

- Take the fixed factors to be Block, Tillage and Lay and the random factor to be Plot.
- Terms involving Block and Lay that are in the Intratier Random model are shifted to the fixed model.
- Lay#Tillage is of interest so that the fixed model includes Tillage * Lay.

Homogeneous Random:
Block∧Plot + Block∧Plot∧Lay
= (Block∧Plot) / Lay
Fixed: Block + Lay + Block∧Lay + Tillage + Tillage∧Lay
= (Block + Tillage) * Lay

A longitudinal RCBD (cont'd)

II. **Homogeneous Random**:

Block∧Plot + Block∧Plot∧Lay
= (Block∧Plot) / Lay

Fixed: Block + Lay + Block∧Lay + Tillage + Tillage∧Lay
= (Block + Tillage) * Lay

III. **General Random and General Fixed models:** Reparameterize terms in homogeneous models and omit aliased terms from random model.

- The **subject term** for Lay is Block∧Plot;
- Expected that there will be unequal correlation between observations with different levels of Lay and same levels of Block∧Plot;
- No aliased random terms.

General random: (Block∧Plot) / uc(gf(Lay))
⇒(*Block*∧*Plot*) / uc(Lay)

- Trends for Lay are of interest, but not for the qualitative factor Tillage nor for Block.

General fixed: (Block + Tilllage) * td(Lay)

- For longitudinal experiments, form **longitudinal error terms**: (*subject term*) ^ gf(longitudinal factors)
- A ***subject term*** for a longitudinal factor is a generalized factor whose levels are units on which the successive observations are taken.
- Allow unequal correlation (uc) between longitudinal factor levels and using gf allows arbitrary uc between these factors.

Mixed model:
(Block + Tilllage) * td(Lay) |
(*Block*∧*Plot*) / uc(Lay)

3) Why randomization-based models?

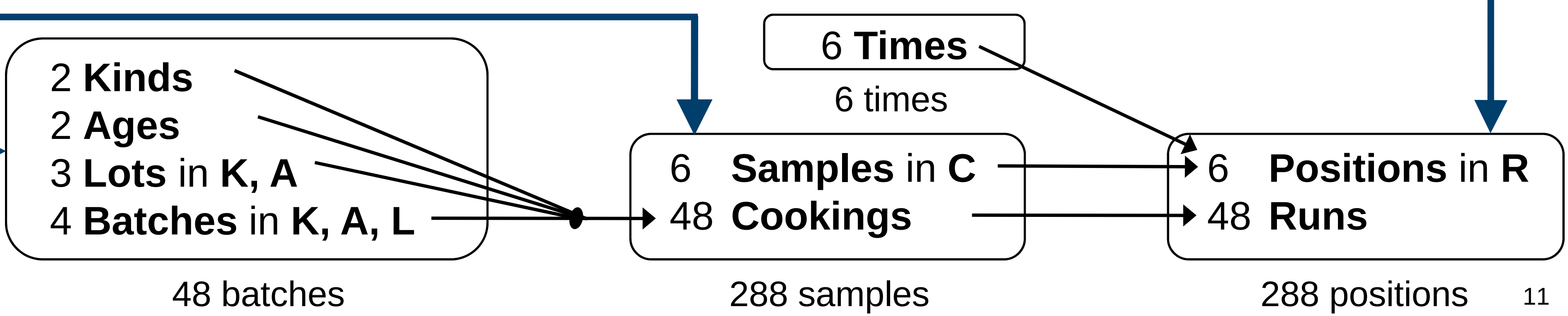
- It is common to form models by writing down a list of terms, sometimes drawing on models for related experiments, and designating each term as fixed or random.
  - e.g. Split-plot-in-Time for the longitudinal RCBD
- Here derive models from tiers: factors indexing sets of objects.
  - Ensures all the terms, taken into account in the randomization, are included in the analysis and that the incorporation of any other terms is intentional.
  - Call such models **randomization-based** in that randomization is used in determining the terms in the model.
- Strongly recommend against using Rule 5 in Piepho et al. (2003), as done by Littel et al. (2006, Sec. 4.2).
  - Rule 5 involves substituting randomized factors for unrandomized factors.
  - Leads to a misidentification of sources of variation, including the possible omission of the experimental units (EUs) as sources.

RCBD and rule 5

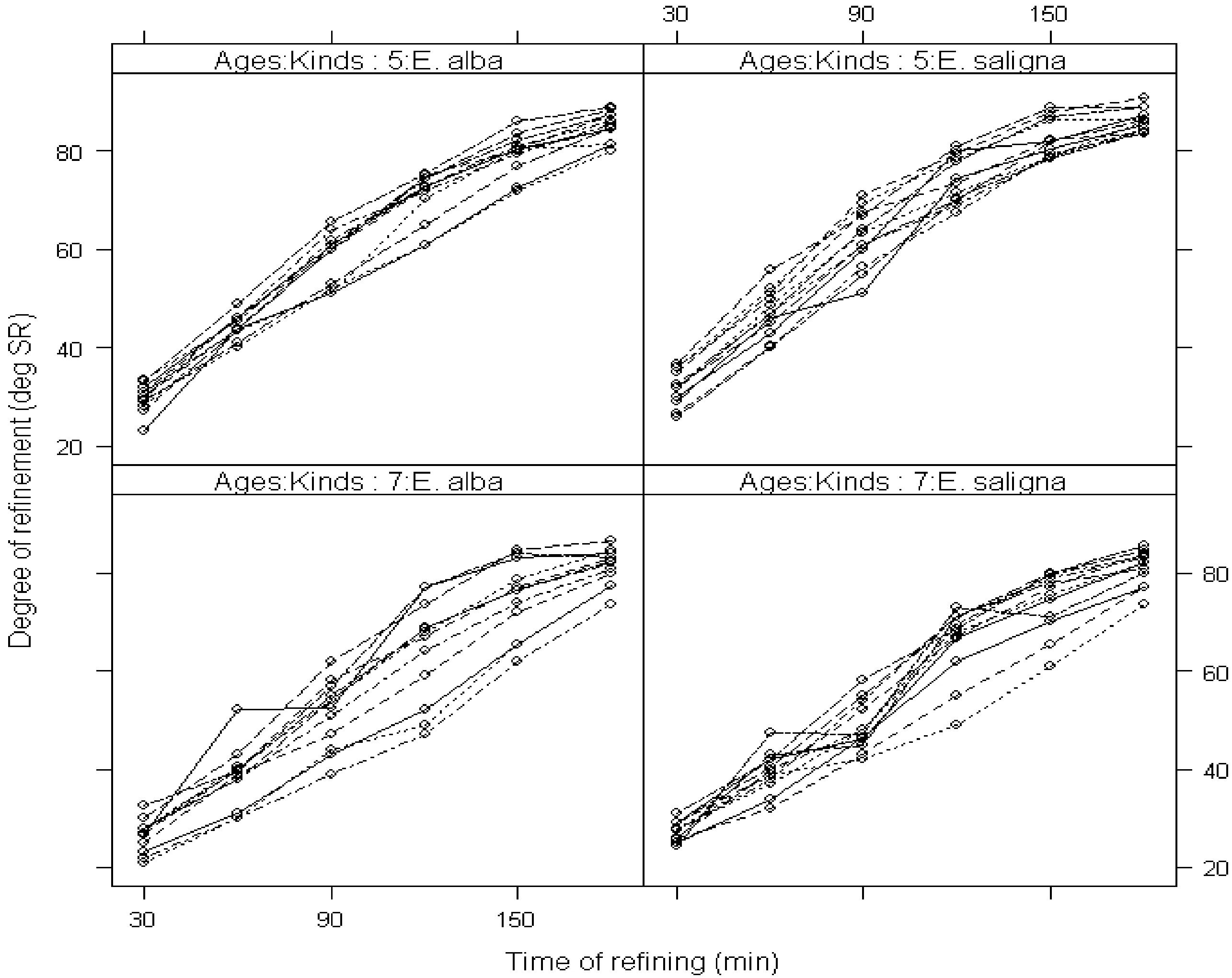
- Mixed model, equivalent to randomization model for an RCBD: Treatments | Blocks + Blocks$\wedge$Plots.
- Rule 5 modifies this to Treatments | Blocks + Blocks$\wedge$Treatments.
- Of course, latter more economical as Plots no longer needed.
- However, latter does not include Blocks$\wedge$Plots, whose levels are the EUs.
- Clearly, levels of Blocks$\wedge$Treatments are not EUs, as Treatments not applied to it levels.
- Blocks$\wedge$Plots and Blocks$\wedge$Treatments are two different sources of variablity: inherent variability vs block-treatment interaction.
- This "trick" is confusing, unnecessary and not always possible.

4) A three-phase example (Pereira, 1969)

- Experiment to investigate differences between pulp produced from different Eucalypt trees.
- Chip phase:
  - 3 lots of 5 trees from each of 4 areas were processed into wood chips.
  - Each area differed in i) kinds of trees (2 species) and ii) age (5 and 7 years).
  - For each of 12 lots, chips from 5 trees were combined and 4 batches selected.
- Pulp phase:
  - Batches were cooked to produce pulp & 6 samples obtained from each cooking.
- Measurement phase:
  - Each batch processed in one of 48 Runs of a laboratory refiner with its 6 samples randomly placed on 6 positions in a pan in the refiner.
  - For each run, 6 times of refinement (30, 60, 90, 120, 150 and 180 minutes) were randomized to the 6 positions in the pan.
  - After allotted time, a sample taken from a pan and its degree of refinement measured.



Profile plot of data



Shows: a) curvature in the trend over time; b) some trend variability; c) variance heterogeneity, in particular between the Ages

Formulated and fitted mixed models

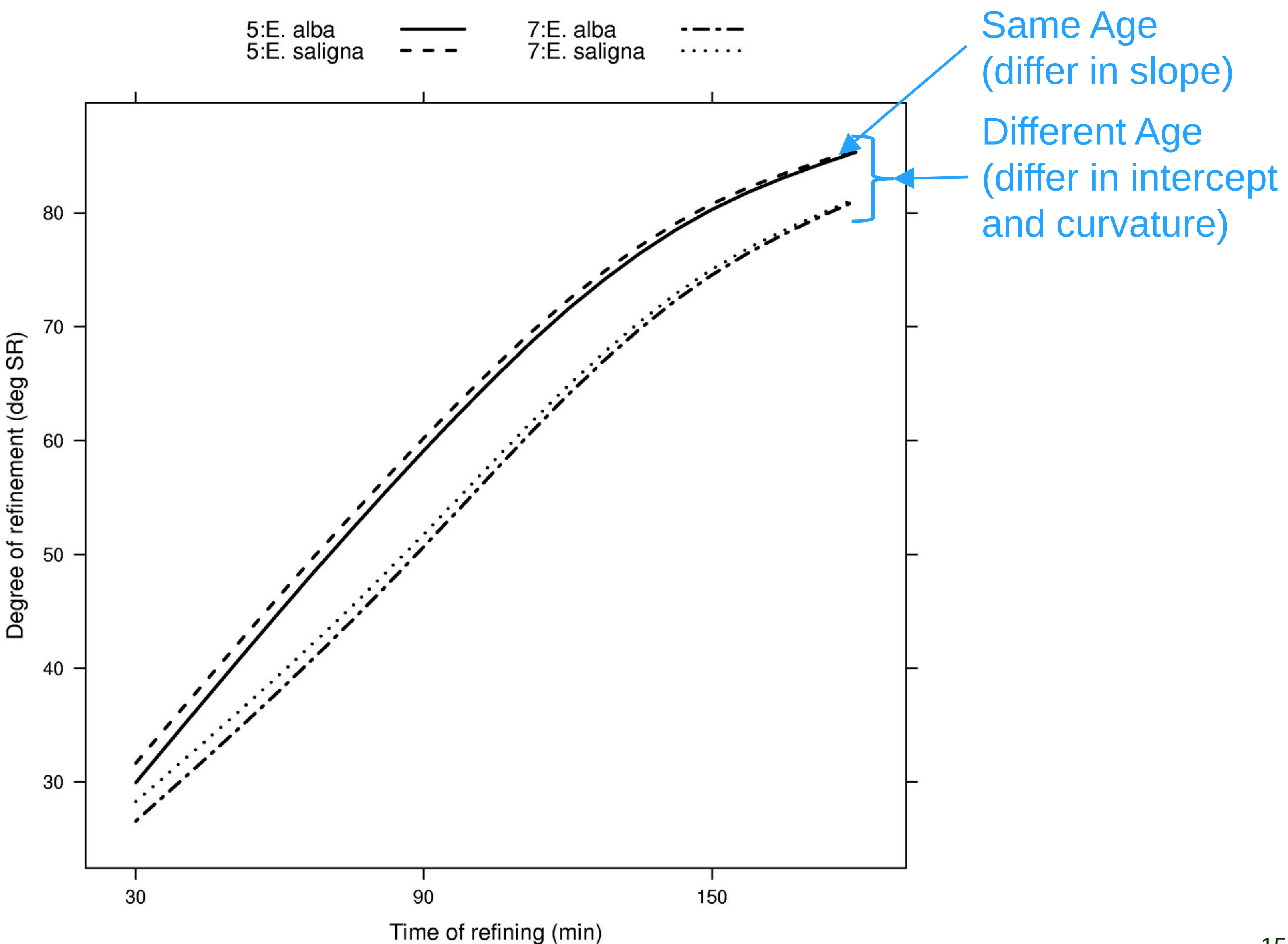
(details in Brien & Demétrio, 2009.)

- Using 3-stage process, following model of convenience is formulated from the 4 tiers:

  General random: Runs / Positions + (*Kinds*∧*Ages*∧*Lots*) / uc(Times) + (*Kinds*∧*Ages*∧*Lots*∧*Batches*) / uc(Times);

  General fixed: Kinds * Ages * td(Times)
- This model:
  - Does not contain Cooking/Samples because of aliasing.
  - Has variance components for Runs, Position, Lots, Batches.
  - Allows for some form of unequal correlation between Times.
  - Includes trends over Times.
- The **full fitted model**, obtained using ASReml-R (Butler et al., 2007), has:
  - For variance,
    - a) unstructured, heterogeneous covariance between Times arising from Runs, Batches and Cookings and that differs for Ages and
    - b) a component for Lots variability.
  - For time, trend whose intercepts and curvature (characterized by cubic smoothing splines (Verbyla, 1999)) differ for Ages and whose slopes differ for Kinds.

Predicted degree of refinement



5) Concluding comments

- Formulate a randomization-based mixed model:
  - to ensure that all terms appropriate, given the randomization, are included;
  - and makes explicit where model deviates from a randomization model.
- Based on dividing the factors in an experiment into tiers.
- To obtain fit, a model of convenience is often used:
  - When aliased random sources, terms for all but one are omitted to obtain fit;
  - But re-included in fitted model if retained term is in fitted model.
- All 11 examples from Piepho et al. (2004) are in Brien and Demétrio (2009).

References

- Brien, C.J., and Bailey, R.A. (2006) Multiple randomizations (with discussion). *J. Roy. Statist. Soc., Ser. B*, **68**, 571–609.
- Brien, C.J. and Demétrio, C.G.B. (2009) Formulating mixed models for experiments, including longitudinal experiments. Accepted for publication in JABES.
- Butler, D., Cullis, B.R., Gilmour, A.R. and Gogel, B.J. (2007) *Analysis of mixed models for S language environments: ASReml-R reference manual*. DPI Publications, Brisbane.
- Littel, R., Milliken, G., Stroup, W., Wolfinger, R. and Schabenberger, O. (2006) *SAS for Mixed Models*. 2nd edn. SAS Press, Cary.
- Pereira, R.A.G. (1969) *Estudo Comparativo das Propriedades Físico-Mecânicas da Celulose Sulfato de Madeira de Eucalyptus saligna* Smith*, Eucalyptus alba* Reinw *e Eucalyptus grandis Hill ex Maiden*. Escola Superior de Agricultura `Luiz de Queiroz', University of São Paulo, Piracicaba, Brasil.
- Piepho, H.P., Büchse, A. and Emrich, K. (2003) A hitchhiker's guide to mixed models for randomized experiments. *Journal of Agronomy and Crop Science*, **189**, 310–322.
- Piepho, H.P., Büchse, A. and Richter, C. (2004) A mixed modelling approach for randomized experiments with repeated measures. *Journal of Agronomy and Crop Science*, **190**, 230–247.
- Verbyla, A.P., Cullis, B.R., Kenward, M.G. and Welham, S.J. (1999) The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Applied Statistics*, **48**, 269–311.