Use of mixed model for evaluation and selection of Burley tobacco

Dr. Elaine BATISTA

Alliance One International

2023 CORESTA Conference
Agronomy & Leaf Integrity and Phytopathology & Genetics Conference
Cancun, 15-19 October, Mexico

Use of mixed model for evaluation and selection of Burley tobacco

Elaine Cristina Batista[1], Ezequiel de Oliveira[1], Eder Leandro Frantz[1], Adriano Teodoro Bruzi[2], Carlos Eduardo Pulcinelli[1].

[1] Alliance One International, Global Center of Research, Development & Deployment, Brazil
[2] Federal University of Lavras, Brazil

AGENDA

1   Introduction

2   Objectives

3   Material & Methods

4   Results

5   Conclusion

What is a Linear Mixed Model?

The mixed models methodology was proposed by Charles HENDERSON (1949) to be used in the genetic evaluation of cattle, and was presented for the first time in 1973 (HENDERSON, 1973)

BIOMETRICS 31, 423–447
June 1975

BEST LINEAR UNBIASED ESTIMATION AND PREDICTION
UNDER A SELECTION MODEL

C. R. HENDERSON

Department of Animal Science, Cornell University, Ithaca, N. Y. 14850, U.S.A.

SUMMARY

Mixed linear models are assumed in most animal breeding applications. Convenient methods for computing BLUE of the estimable linear functions of the fixed elements of the model and for computing best linear unbiased predictions of the random elements of the model have been available. Most data available to animal breeders, however, do not meet the usual requirements of random sampling, the problem being that the data arise either from selection experiments or from breeders' herds which are undergoing selection. Consequently, the usual methods are likely to yield biased estimates and predictions. Methods for dealing with such data are presented in this paper.
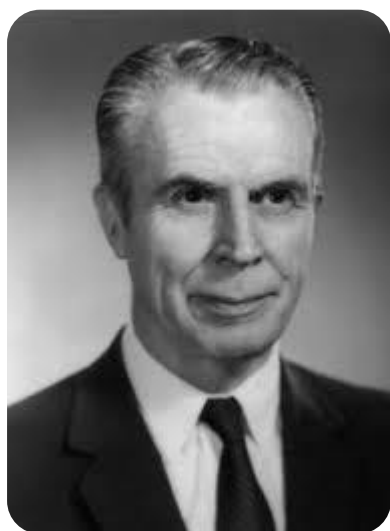
What is a Linear Mixed Model?

The Mixed Model started to be used in practice from the 1990s onwards due to the computational technological advances that allowed its use
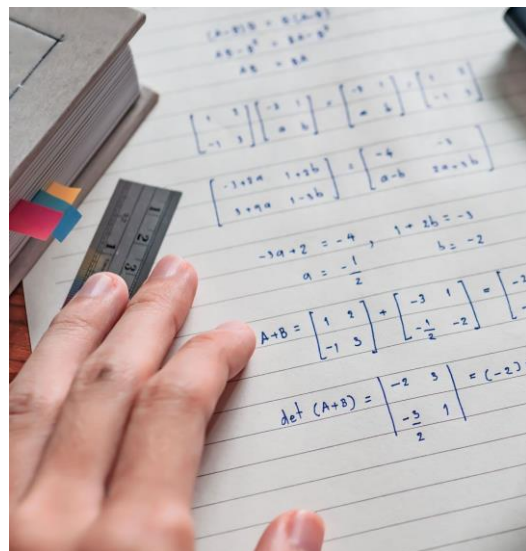
1961: First All-Electronic Desktop Calculator

What is a Linear Mixed Model?

- A statistical model has the function of explaining the observations of a dependent variable in terms of other possible causes of variation (independent variables).

- Such variables can be classified as having a fixed or random effect.

Analysis of variance (ANOVA)= fixed OR random

What is a Linear Mixed Model?

A mixed model is one that presents both fixed and random effects factors, in addition to the experimental error and the constant μ

$$y = X\beta + Z\gamma + \varepsilon$$

Where:

y: is the vector of phenotypic observations with dimension n x 1; n: number of observations

X: fixed effects incidence matrix with dimension n x p;

β: fixed effects vector with dimension p x 1; p: number of fixed effects

Z: incidence matrix of random effects with dimensions n x q;

γ: vector of random events with dimension q x 1; q: number of random effects

ε: residue vector with dimension n x 1.

What is a Linear Mixed Model?

Best Linear Unbiased Estimates **(BLUEs)** are the solutions (or estimates) associated with the fixed effects and Best Linear Unbiased Predictions **(BLUPs)** are the solutions (but identified as predictions) associated with the random effects of a model.

What can Mixed Models do in Plant Breeding?

1- Can account for multiple readings on the same subject (eg.: more than one harvest in the same plant, seed germination count – 7 and 16 days)
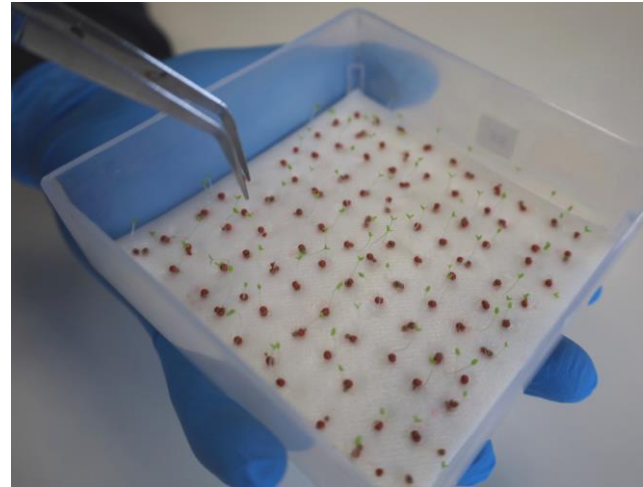
2- Account for nested structure in data. They can account for data that has a nested or hierarchical structure (eg.:harvest effect within the location, genotype within environment)

What can Mixed Models do in Plant Breeding?

3- Can handle missing measurements and different numbers of measurements per subject. They will still work in situations where you have different numbers of measurements for each subject or when measurements are taken at different time intervals.

Why to use Linear Mided Models in Tobacco Breeding?

1- The trials have a large number of genotypes that, during selection cycles, are discarded, which generates unbalanced data

2- The experiments are conducted in several locations and years (multi-environment trial - MET), which aims to increase selection efficiency, deviating the environmental effect on the genotypes.

Why to use Linear Mided Models in Tobacco Breeding?

3- Due to the large number of genotypes, locations and years, there may be loss of plots in experiments

Hail rain

Objective

Evaluate the performance of 24 Burley tobacco hybrids, developed by Alliance One International, using a Mixed Model approach.

Material & Methods

Genotypes

- 24 Burley tobacco hybrids (Preliminary variety test)

- 5 checks: DBH 455
          HB 4180P
          HB 4192P
          HB 4196P
          TN 90

Imbalanced data (genotypes) !

Material & Methods

Evaluated traits

Yield

Nicotine

Quality Index

Material & Methods

Field design

Randomized Complete Block Design (RCBD) with 3 replications

Plot= 2 rows with 10 plants each

Material & Methods

Statistical analysis

• Linear Mixed Models approach (REML/BLUP)

Software:

2 steps: 1ˢᵗ Fit the appropriate mixed models to the phenotypic data of each trait.

2ⁿᵈ Consists of mixed model analysis. To obtain predictions of the genotypic values of the hybrids, the effects of nested hybrids within environments (combinations between locations and crop seasons) were stored in the requested part of the model

Material & Methods

Statistical analysis- The general phenotypic model

$$y_{ijkr} = \mu + b_r + L_j + S_k + LS_{jk} + G_{ijk} + \varepsilon_{ijkr}$$

- y_{ijkr} is te value of the observation referring to the genotype i (com $i = 1, \dots, n$), in the block r (com $r = 1,2, R = 3$), in the location j (com $j = 1,2, J = 7$), in the season k (com $k = 1, \dots, K = 2$);
- μ is a constant;
- b_r is the block effect;
- L_j is the location effect;
- S_k is the season effect;
- LS_{jk} is the effect btween locations and seasons;
- G_{ijk} is the effect of the genotype i, on location j, on season k;
- ε_{ijkr} is an experimental error.

Material & Methods

Statistical analysis

- For each character, only significant fixed effects (a = 0.05) were maintained in the phenotypic model, according to the Wald test and the random effects whose variance estimates differed from zero according to the Likelihood Ratio Test – LRT

- Obtaining predictions of the genotypic values of hybrids within environments and in the average of environments **(BLUP)**

Results

Table 1. Estimates of variance components by Wald test for Yield (kg.ha^{-1}) Nicotine (%) and Quality Index Grade (%) traits referring to the evaluation of 24 Burley hybrids in 7 different environments in the southern region of Brazil.

Effects	DF	Yield	Nicotine	Quality Index Grade
Genotype	28	8318.0146*	0.1552264**	15.711221**
Environment: Genotype	6	12673.313**	0.0058498^{ns}	5.7057516^{ns}
Residual	6	42386.767	0.1226974	57.148849

^{ns}, **, * Wald test not significant and significant at 1% and 5% respectively, by Chi Test Square

Genotype	BLUP		
	Yield	Nicotine	Quality Index
510	3038	4.71	74.70
600	3003	4.29	76.89
545	3000	4.27	77.09
597	2987	3.97	75.91
509	2964	4.70	81.62
544	2964	4.44	81.26
486	2961	4.30	77.02
505	2959	4.14	76.22
512	2959	4.32	75.34
543	2949	4.13	71.70
542	2949	4.53	69.26
594	2924	4.63	68.83
599	2916	4.98	80.22
HB 4192P	2915	4.50	78.79
550	2889	4.14	74.73
HB 4180P	2883	4.39	74.35
HB 4196P	2883	3.73	74.19
547	2880	4.04	72.39
595	2874	4.97	78.47
546	2874	4.77	74.28
TN 90	2848	4.88	77.57
549	2848	3.92	72.18
506	2845	4.55	77.81
596	2836	4.47	75.45
548	2821	3.56	76.74
504	2820	4.56	79.09
DBH 455	2808	4.72	80.17
598	2804	4.52	74.11
507	2770	5.21	81.93

Table 2. Predicted average values (BLUPs) for the Yield (kg.ha^{-1}), Nicotine (%) and Quality Index Grade (%) traits referring to the evaluation of 24 Burley hybrids in 7 different environments in the southern region of Brazil.

Conclusion

Did the use of the statistical approach of Linear Mixed Models was effective?	It was effective with regard to the selection of superior genotypes, given the nature of the data set, which had an imbalance in terms of the evaluated genotypes
Was there a difference between the genotypes evaluated and between the characters evaluated?	There were significant differences for the effect of genotype for the characters, which indicates that for these characters there is genetic variability.
Was it possible to select among the hybrids evaluated?	The behavior of the hybrids within the environments was verified, allowing the selection of hybrids according to the environments.

Thank you!

Dr. Elaine Batista
Alliance One International
e-mail: ebatista@aointl.com
Phone: +55 (51) 3793-1589