Exploration of trait relationships in mungbean using a multivariate linear mixed model

Eugenia Settecase[1,2], Cristina Cuesta[2], M. Gabriela Borgognone[1], M. Valeria Paccapelo[1]

[1]*Department of Agriculture and Fisheries, Leslie Research Facility, Toowoomba, Queensland, Australia.*
[2]*Universidad Nacional de Rosario, Rosario, Santa Fe, Argentina.*

1. Introduction

- Breeders select the best genotypes based on the analysis of multiple traits which could often be correlated [1,2].
- It is important to obtain the predictions of the traits with the best possible accuracy.
- Multi-variate analysis using a multivariate linear mixed model can increase the accuracy with which the traits are predicted, compared to separate univariate analyses.

2. Objectives

- Present a trivariate linear mixed model based on the bivariate analysis proposed by Ganesalingam et al. (2013) [3].
- Evaluate the accuracy increment of the trivariate method when applied to two scenarios (two sets of three traits each).

3. Materials

- Both scenarios are from the same replicated field trial with 25 diverse mungbean inbred genotypes.

Scenario 1: Highly heritable traits

Seed colour traits: L^*, a^*, b^*. Measured using a scanner (each observation is the average of 20 seeds – one observation per plot).
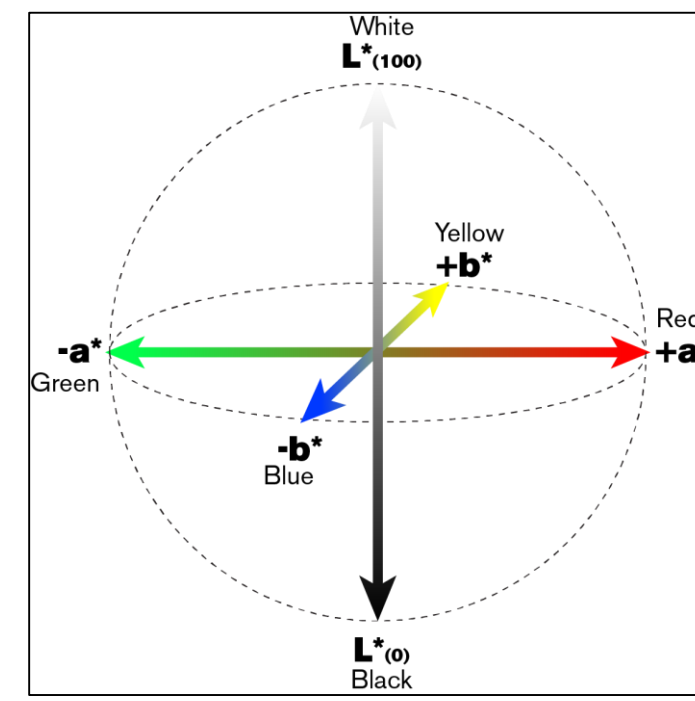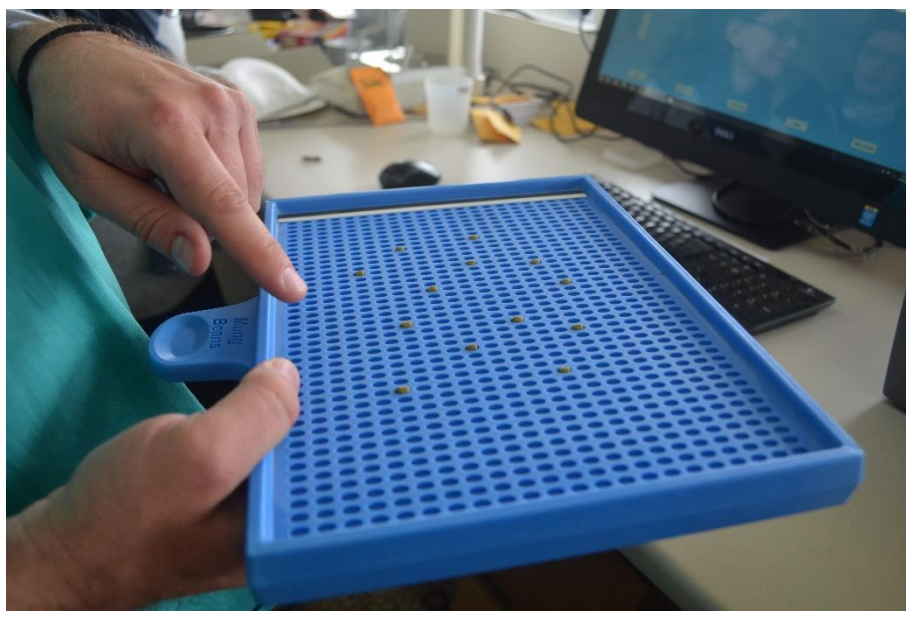
Scenario 2: Less heritable traits

Pod dissection traits: *pod length, seeds per pod,* and *50 seed weight* (calculated from the average weight of all seeds in the pod). The data is unbalanced (different number of samples per plot).
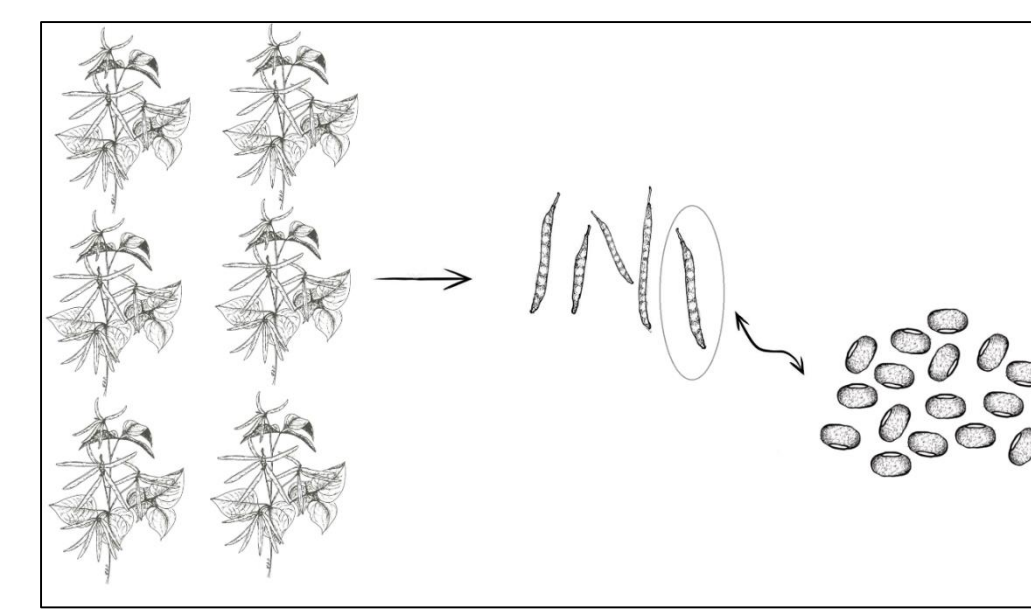
4. Methods

- Trivariate linear mixed model for each scenario: $y^{3N \times 1} = X\tau + Z_g u_g + Z_d u_d + e$

    $y_{3N \times 1}$: data for the traits stacked under each other

    τ: fixed effects u_g: genotypic effects, $u_g \sim MN(0, G_g)$

    u_d: design effects, $u_d \sim MN(0, G_d)$ e: residual effects, $e \sim MN(0, R)$

    X, Z_g, and Z_d are the associated design matrices

- Different covariance structures were proposed for the genotypic and residual effects.

    - **Unstructured** (all variances and covariances can differ)

    $$G_g^{3N} = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_{12}} & \sigma_{g_{13}} \\ & \sigma_{g_2}^2 & \sigma_{g_{23}} \\ & & \sigma_{g_3}^2 \end{bmatrix} \otimes I^N \quad R^{3N} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} & \sigma_{e_{13}} \\ & \sigma_{e_2}^2 & \sigma_{e_{23}} \\ & & \sigma_{e_3}^2 \end{bmatrix} \otimes I^N$$

    - **Diagonal**, where $\sigma_{g\,jk} = 0$ and $\sigma_{e\,jk} = 0 \; \forall j \neq k$ (equivalent to fitting three univariate models)

Accuracy measure for the genotypic predictions [4]

$$r_{i_w} = \sqrt{1 - \frac{sep_{i_w}^2}{\sigma_{g_w}^2}} \text{, where}$$

sep_{i_w}: standard error of prediction for genotype i at trait w
$\sigma_{g_w}^2$: genotypic variance component for trait w

This measure will increase when the model can account for the residual covariance between traits.

Accuracy gain (%): quantifies the gains in accuracy for each trait when using a trivariate analysis instead of three univariate analyses.

5. Results

- The genotypic effect was significant in the analyses of both scenarios and the main source of variability was the genotypic variance.
- The covariance structures of best fit for both scenarios were the unstructured at the genotypic and residual levels (*REML ratio test*).

Scenario 1: Highly heritable traits

	L*	a*	b*
Accuracy gain	0.001%	0.003%	0.008%
Heritability	0.99	0.96	0.99

Figure 1. Genotypic correlations between traits (lower diagonal). Genotypic predictions (upper diagonal).

**Figure 2.** 3D interactive plot of genotypic predictions.

Figure 1.
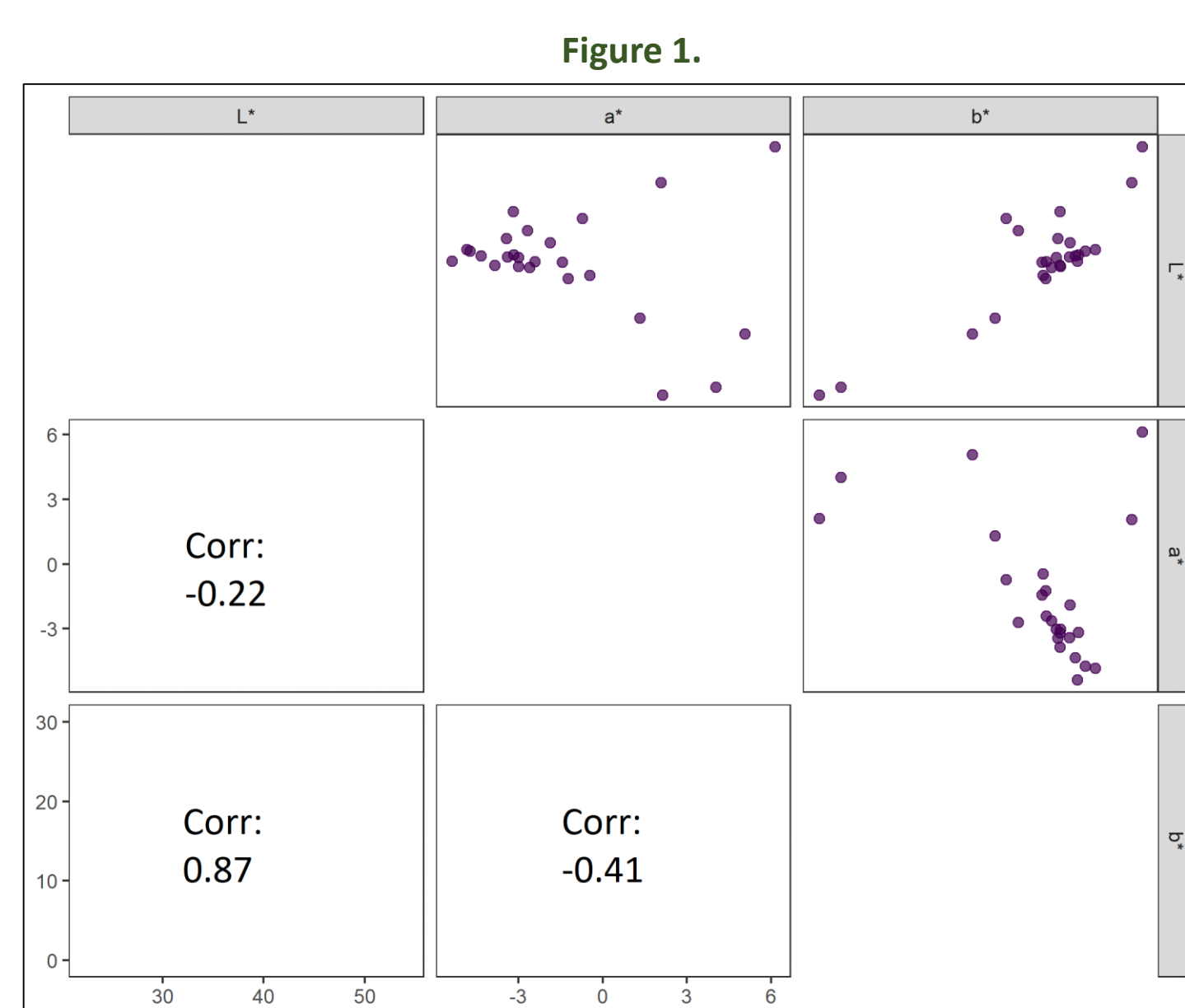
Although the trivariate model provided a better fit, the accuracy gains were negligible due to the highly heritable nature of the traits.

Scenario 2: Less heritable traits

	Pod length	Seeds per pod	50 seed weight
Accuracy gain	3.3%	6.5%	0.1%
Heritability	0.89	0.71	0.98

Figure 2. Genotypic correlations between traits (lower diagonal). Genotypic predictions (upper diagonal).

Figure 4. 3D interactive plot of genotypic predictions.

Figure 3.
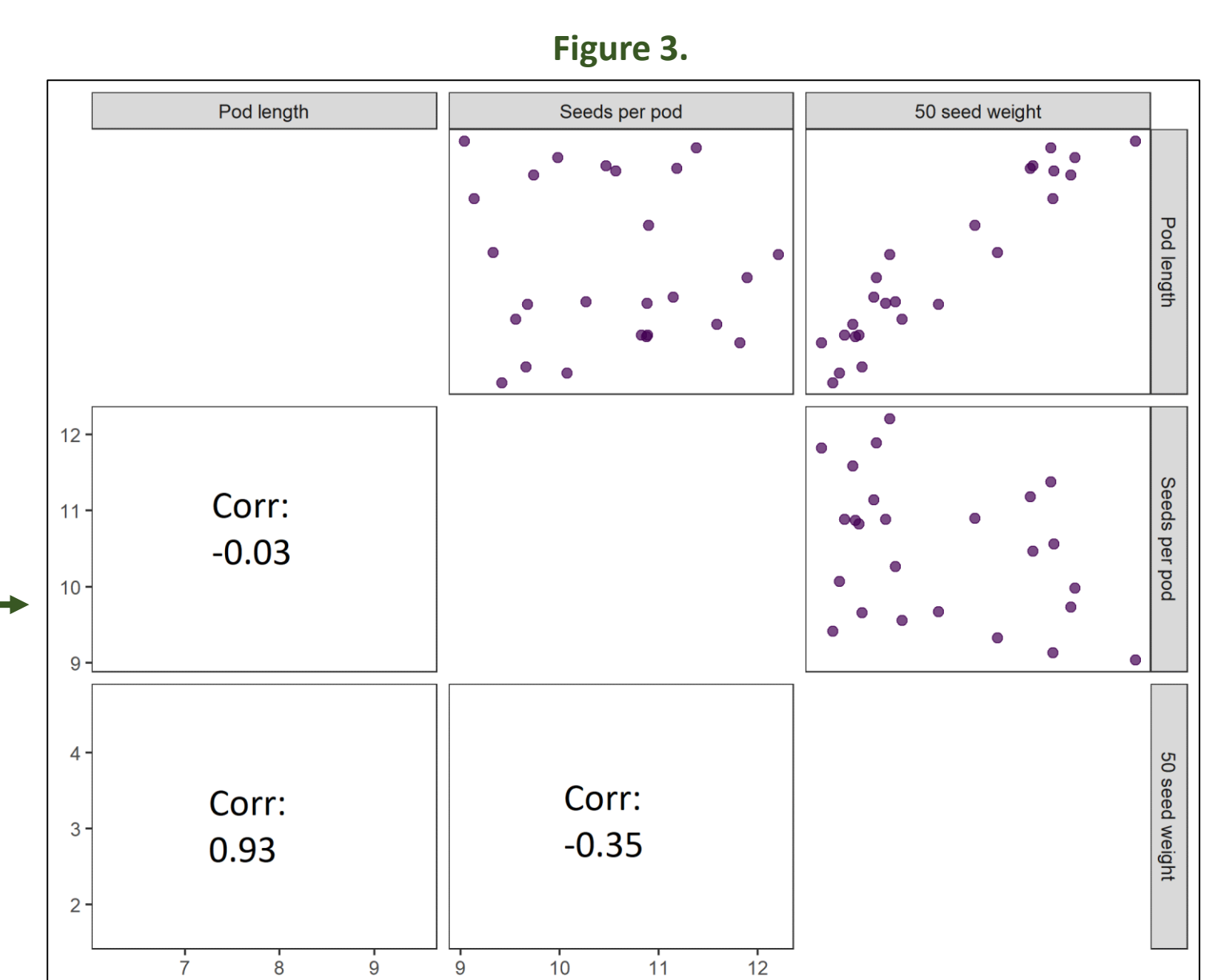
The accuracy gains for the trivariate model were higher for the two less heritable traits, which had more residual variability.

6. Discussion

- The trivariate model allowed for the examination of the genotypic effects individually for each trait while accounting for the correlations between the traits through the use of a more appropriate correlation structure (for both the genotypic and residual effects).
- The main gains in prediction accuracy occurred for the less heritable traits when the residual variability was modelled appropriately.
- Future work will aim to enable the multivariate analysis of more than three traits, as well as the incorporation of spatial variability within each trial.

References: [1] Tao, Y., Mace, E., George-Jaeggli, B., Hunt, C., Cruickshank, A., Henzell, R., Jordan, D. (2018). The Plant Genome 11 Art.89. [2] Noble, T. J., Tao, Y., Mace, E. S., Williams, B., Jordan, D. R., Douglas, C. A., Mundree, S. G. (2018). Frontiers in Plant Science, 8 Art..2102. [3] Ganesalingam, A., Smith, A. B., Thompson, R., Cullis, B. R. (2013). 190 371-383. [4] Mrode R., Thompson R. (2005). *Linear models for the prediction of animal breeding values, 2nd edn.* CABI Publishing, Wallingford.