



PRECISION OF EARLY SELECTION USING MIXED MODELS IN INIA'S RICE BREEDING PROGRAM

S. Scheffel^{1*}; F. Molina¹; F. Perez de Vida¹; J. E. Rosas¹¹INIA Uruguay; *sscheffel@inia.org.uy

Introduction

Selecting genotypes in their first field evaluation (E1) has a great impact in breeding program's efficiency, as its precision determines the suitable selection intensity. It is also challenging, due to a poor representation of environments that hinders the estimation of environment and genotype by environment effects and variances. Joint analysis of multiple trials and years and the use of mixed models can lead to higher selection precision.

Objective

The objective of this work is to compare the precision of estimation or predictions at different early selection scenarios with increasing availability of yield trial data.

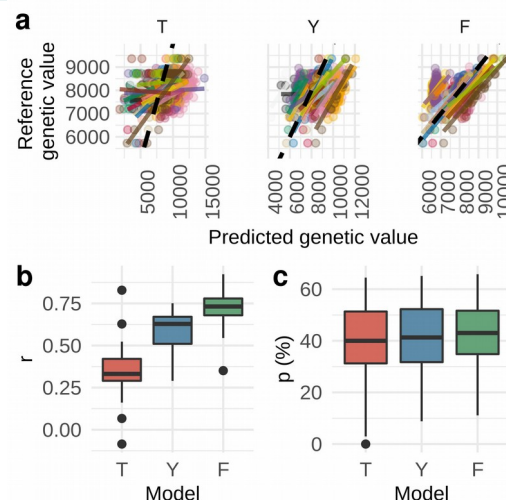
Materials and methods

Data consisted of 23 years of field evaluation available in INIA's Rice Breeding Program (IRBP) database. Different data availability scenarios (DAS) for GV estimation at E1 selection time were recreated using partial datasets: (T) one trial per dataset, (Y) all trials of the same year, (F) all trials in the previous four and in the year when evaluation occurs, and (C) complete dataset from all 23 available years. A model accounting for every source of variation available in each dataset was fitted for each DAS: for models in T genotype and block were modelled as fixed effects; for model Y genotype, block (nested in trial), and trial were modelled as random effects; and for models F and A genotype, block (nested in trial), trial and year were modelled as random effects. Reference genetic value (GV) was defined as the BLUP

obtained with model A for those genotypes evaluated in at least 3 years. Models for each DAS were: (T) $y_{ij} = \mu + \underline{g_i} + \underline{b_j} + \varepsilon_{ij}$; (Y) $y_{ijkl} = \mu + \underline{g_i} + \underline{b_{j(k)}} + \underline{t_k} + \underline{l_l} + (\underline{gt})_{ik} + \varepsilon_{ijkl}$; (F) and (C) $y_{ijklm} = \mu + \underline{g_i} + \underline{b_{j(k)}} + \underline{t_{k(m)}} + \underline{l_l} + \underline{w_m} + (\underline{gt})_{ik} + (\underline{gw})_{im} + \varepsilon_{ijklm}$. Where y_{ijklm} is yield, μ the general mean, g_i the genotypic effect, $b_{j(k)}$ block effect, nested within trial, $t_{k(m)}$ trial effect, nested within year, l_l location effect, w_m year effect, $(gt)_{ik}$ genotype by trial effect, $(gw)_{im}$ la genotype by year effect, and ε_{ijklm} is the model effect. Fixed effects are underlined, random effects are assumed $\sim N(0, \sigma^2)$, variances are assumed independent and identically distributed. Indicators for the accuracy of the GV prediction were: r, the Pearson's correlation between reference GV and the GV estimated or predicted for each DAS; and p, the percentage of genotypes selected based on each DAS that match those with best reference GV.

Results and discussion

Figure 1. (a) Reference genetic value (GV) by GV predicted for different data availability scenarios (DAS): one trial (T), all trials in one year (Y) and all trials in the previous four and in the evaluated year (F). Colors are evaluated years, solid lines are the regression for each year, dashed lines are 1:1 ratio. **(b)** Correlation (r) between reference GV and predicted GV in each DAS. **(c)** Percentage (p) of genotypes selected based on each DAS that match those with best reference GV.



Results show that as models accounts for more sources of variation, r and p values increase (Figure 1).

The variability attributable to environmental effects such as trial and year accounted for a standard deviation of 2 ton/ha, and to genotype by environment effects was around 1 ton/ha (Figure 2).

Thus, the improvement in the precision of GV estimation achieved corresponds to a better modeling of these environment and genotype by environment effects.

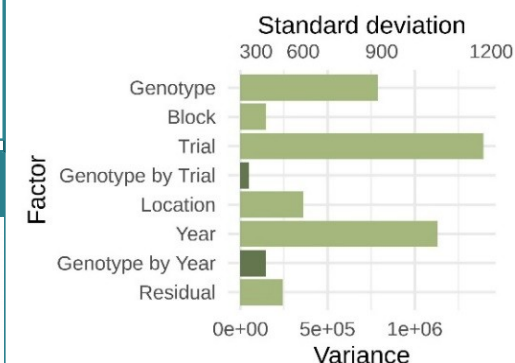


Figure 2. Variance and standard deviation distribution for a model using the complete dataset (23 years, 1997-2020)

Conclusion

The precision of GV estimates of genotypes in early evaluation stage was higher as more environment and genotype by environment effects were modeled. Forthcoming use of models allowing variance and covariance structures among genotypes and environments is anticipated to enable even higher GV prediction accuracies.