Revista Ciência Agronômica, v. 54, e20228567, 2023
Centro de Ciências Agrárias - Universidade Federal do Ceará, Fortaleza, CE
www.ccarevista.ufc.br

Scientific Article
ISSN 1806-6690

Performance prediction of crosses using estimated breeding values for regions of soybean production in Brazil[1]

Anderson Dallastra[2]*, Joênes Mucci Peluzio[2], Leandro de Freitas Mendonça[3], Rafael Ravaneli Chagas[4], Bruno de Almeida Soares[4], Gabriel Mendes Villela[5], Nizio Fernando Giasson[6]

ABSTRACT - The aim of this study was use the performance prediction of crosses in a group of conventional soybean genotypes to obtain the breeding value (BV), and to evaluate the correlation between the prediction and the actual productive potential of the progeny generated by this method in experimental tests for different seasons and environments, and determine whether the methodology is efficient in generating progeny of high productive potential for the soybean macro-regions (SMR) and soil and climate regions (SCR) of Brazil. A total of 481 conventional elite genotypes were selected as parents, the BV were generated, and crosses were predicted using the restricted maximum likelihood/best linear unbiased prediction mixed-model procedure (REML/BLUP). In 2019, the predicted crosses and advancement of the F_1 and F_2 segregating populations were carried and sent to the breeding programs of a private company in Passo Fundo-RS, Cambé-PR, Rio Verde-GO, Lucas do Rio Verde-MT and Porto Nacional-TO, where they were sown during the 2019/2020 crop season. During the 2020/2021 season, 1868 progeny were selected and tested in experimental trials at these locations. The progeny were again tested during the 2021/2022 season in experimental trials in 50 environments in SCR throughout Brazil. The result of the analysis showed a very weak to moderate correlation, indicating little efficiency for the prediction model used in this study. It is suggested that the prediction model be revised to include a greater number of variables, such as the kinship matrix, so that the BV of the genotypes can be more assertively estimated, especially when the aim is to select progeny in early generations with a high degree of heterozygosity.

Key words: *Glycine max*. Breeding value. Correlation. Genetic improvement. Hybridization.

DOI: 10.5935/1806-6690.20230060
Editor-in-Chief: Profa. Charline Zaratin Alves - charline.alves@ufms.br
*Author for correspondence
Received for publication 02/08/2022; approved on 22/03/2023
[1]Extracted from the doctoral thesis of the lead author, presented to the Graduate Programme in Plant Production of the Federal University of Tocantins (UFT)
[2]Graduate Programme in Plant Production, Federal University of Tocantins (UFT), Gurupi Campus, Gurupi-TO, Brazil, anderson_dallastra@hotmail.com (ORCID ID 0000-0003-2303-1451), joenesp@mail.uft.edu.br (ORCID ID 0000-0002-9336-2072)
[3]Department of Genetics, 'Luiz de Queiroz' Higher School of Agriculture of the University of São Paulo (ESALQ/USP), Piracicaba-SP, Brazil, lmendonca@gdmseeds.com (ORCID ID 0000-0003-1723-2302)
[4]Department of Agronomy, Federal University of Viçosa (UFV), Viçosa-MG, Brazil, rafaelr_chagas@hotmail.com (ORCID ID 0000-0002-7195-0009), bruno.a.soares@ufv.br (ORCID ID 0000-0002-4604-2199)
[5]Departament of Biology, Federal University of Lavras (UFLA), Lavras-MG, Brazil, gabrielmendesagro@gmail.com (ORCID ID 0000-0002-7300-8527)
[6]GDM Genética do Brasil S.A, Cambé-PR, Brazil, ngiasson@gdmseeds.com (ORCID ID 0000-0002-4497-1330)

INTRODUCTION

The principal aim of a breeding program is to develop genetic combinations between parents (RESENDE; ALVES, 2021) and then develop cultivars that are superior to current cultivars grown and selected in areas that represent the growing region of the species (FRITSCHE-NETO, 2013).

Many cultivar development programs grow a large number of crosses in the expectation that one might result in a superior genetic combination that can be released as a new cultivar. With an effective method of choosing the parents, the number of crosses could be considerably reduced (BORÉM; MIRANDA, 2013).

Since its inception, plant improvement has been based on the visual selection of individuals, i.e., selection based on phenotypic value only (ALLARD, 1999).

It is presumed that the use of mixed models can predict the genotypic value, also known as the breeding value (BV), increasing the efficiency of the selective processes in plant breeding. The REML/BLUP (Restricted Maximum Likelihood/Best Unbiased Linear Prediction) mixed-model method is currently used for studying families, allowing the genetic parameters to be estimated and the genotypic values of the families to be predicted (RESENDE, 2002). According to Pinheiro *et al*. (2013), there is a growing number of reports on the use of BLUP/REML in soybean improvement.

Heffner, Sorrells and Jannink (2009), describes how the success of a genetic improvement program depends on the accuracy of predicting the genetic value (BV) from phenotypic values, however, such predictive procedures still need to demonstrate their practical effectiveness in field tests so that they can be incorporated into routines as a tool to aid and advance genetic improvement programs. The selection of good parents is the key to success in plant breeding, with individuals to be used as future parents being originally selected based on their superior genotypic value (BERNARDO, 2020), in order that methodologies which employ this information to satisfactorily predict the performance of crosses of high genetic potential might be promising.

Furthermore, performance prediction in crosses of proven value can make plant breeding programs more efficient in terms of genetic gains, reducing the costs of selection processes, and of developing and obtaining new cultivars, thereby increasing the supply of new cultivars that can be recommended for the different agricultural regions.

The aim of this study was to predict the performance of crosses in a group of conventional soybean genotypes from a private breeding program to obtain the breeding value (BV) of the genotype combination. Also, by means of experimental testing in different seasons and environments, to evaluate the correlation between the result of the prediction and the actual productive potential of the progeny generated by this method and thereby determine whether the methodology is efficient in generating progeny of high productive potential for the soybean macro-regions (SMR) and soil and climate regions (SCR) of Brazil.

MATERIAL AND METHODS

All the genotypes in this study are conventional and come from the soybean breeding programs of GDM Genética do Brasil S.A (GDM) implemented in each SMR and SCR in Brazil, as per the third approximation (KASTER; FARIAS, 2012).

A total of 481 elite genotypes were selected as genitors for predicting the crosses, including 31 genotypes from breeding program M1, relating to SMR 1; 120 genotypes from breeding program M2, related to SMR 2; 118 genotypes from breeding program M3, relating to SMR 3; 176 genotypes from breeding program M4, related to SMR 4, and 36 genotypes from breeding program M5, relating to SMR 5.

The REML\BLUP mixed-model procedure was used with the ASReml-R statistical package (BUTLER *et al*., 2017) to predict the genetic values of the genotypes (BV) and predict the performance of the crossings employing the Shiny package of the R software (R CORE TEAM, 2016), as per the following model:

$$\mathrm{Re}\,nd_{ijkl} = \mu + G_i + E_j + (G \times E)_{ij} + T(E)_{jkl} + e_{ijkl} \quad (1)$$

where μ is the overall mean value, G_i is the fixed effect of the *i*th genótipo($i = 1,2,...,p$); *Ei* is the random effect of the *j*th environment (combination of locality + crop year + sowing date) $N(0, \sigma_E^2)$; $(G \times E)_{ij}$ is the Random effect of the interaction $N(0, \sigma^2)$; $T(E)_{ik}$ is the Random effect of the *k*th trial in the *j*th environment $\sim N(0, \sigma_T^2¿)$; $e_{i_{jk}}$ is the experimental error associated with the experimental unit of the *i*th genótipo in the *k*th trial in the *j*th environment $N(0, \sigma_{ej}^2)$ with a different variance σ_{ej}^2 for each environment *j*.

The predicted crossings and generation advancement of the F_1 and F_2 populations were carried out in 2019 in Porto Nacional-TO. After harvesting, these were sent to the GDM breeding programs in Passo Fundo-RS (M1), Cambé-PR (M2), Rio Verde-GO (M3), Lucas do Rio Verde-MT (M4), and Porto Nacional-TO (M5), based on the crossing guidelines for each SMR and the aims of each breeder. In the 2019/2020 season, these segregating F_2 populations were sown in experimental trials (POP trials) that included controls, to characterize the relative maturity group (MGP) of each progeny based on comparative phenotypic observations.

In all, 1868 F_3 progeny were selected, as follows: 206 progeny from 26 M1 pedigrees with an MGP ranging from 5.0 to 6.8; 202 progeny from 106 M2 pedigrees with an MGP ranging from 6.0 to 7.0; 679 progeny from 120 M3 pedigrees with an MGP ranging from 6.5 to 7.6; 732 progeny from 180 M4 pedigrees with an MGP ranging from 7.4 to 8.5, and 49 progeny from 28 M5 pedigrees with an MGP ranging from 8.0 to 8.7.

In the 2020/2021 season, the F_3 progeny were sown in the respective breeding programs to evaluate phenotype and yield (YLD) in trials (MROW trials) consisting of 50 treatments containing progeny with no repetitions, and 10 treatments containing controls repeated between each trial. The plot layout comprised two rows of five meters at a spacing of half a meter between rows and half a meter between plots (plant corridor) in an augmented block design (ABD), as described by Federer (1956). To analyze the data, the R software (R CORE TEAM, 2016) was used employing mixed-model methodology and considering random effects, as per the formula:

$$\mathrm{Re}\,nd_{ijk} = \mu + G_i + B_k + e_{ik} \quad (2)$$

where μ is the overall mean; G_i is the fixed effect of the ith genotype $(i = 1,2,\ldots, p)$; B_k is the random effect of the kth block $N(0, \sigma_B^2)$; e_{ik} is the experimental error associated with the experimental unit of the ith genotype in the kth block $N(0, \sigma_e^2)$.

For the 2021/2022 season, the F_4 progeny were sown in experimental trials (RETEST trials) in 50 environments throughout Brazil, corresponding to the SCR in each SMR under the responsibility of the breeding programs, as shown in Table 1, allowing information on the genotype x environment interaction (GxE) to be captured.

Table 1 - Environments, SMR, SCR and number of progeny sown in the 2021/2022 season

Environment	SMR	SCR	Program	Number of progeny
Cachoeira do Sul-RS	SMR 1	101	M1	206
Restinga Seca-RS	SMR 1	101	M1	206
Abelardo Luz-SC	SMR 1	102	M1	206
Condor-RS	SMR 1	102	M1	206
Passo Fundo-RS	SMR 1	102	M1	206
Giruá-RS	SMR 1	102	M1	206
São Luiz Gonzaga-RS	SMR 1	102	M1	206
Itapeva-SP	SMR 1	103	M1	206
Muitos Capões-RS	SMR 1	103	M1	206
Ponta Grossa-PR	SMR 1	103	M1	206
Cascavel-PR	SMR 2	201	M2	202
Itaipulândia-PR	SMR 2	201	M2	202
Rolândia-PR	SMR 2	201	M2	202
Batayporã-MS	SMR 2	202	M2	202
Caarapó-MS	SMR 2	202	M2	202
Francisco Alves-PR	SMR 2	202	M2	202
Naviraí-MS	SMR 2	202	M2	202
Dourados-MS	SMR 2	204	M2	202
Maracaju-MS	SMR 2	204	M2	202
Ponta Porã-MS	SMR 2	204	M2	202
Rio Brilhante-MS	SMR 2	204	M2	202
Sidrolândia-MS	SMR 2	204	M2	202
Jataí -GO	SMR 3	301	M3	679
Rio Verde -GO	SMR 3	301	M3	679
Santa Helena-GO	SMR 3	302	M3	679

Continuation Table 1

Turvelândia -GO	SMR 3	302	M3	679
Uberlândia -MG	SMR 3	303	M3	679
São Miguel do Passa Quatro-GO	SMR 3	304	M3	679
Montividiu-GO	SMR 4	401	M3	679
Paraúna -GO	SMR 4	401	M3	679
Campo Verde-MT	SMR 4	401	M4	732
Primavera do Leste -MT	SMR 4	401	M4	732
Campo Novo do Parecis-MT	SMR 4	402	M4	732
Campos de Julio-MT	SMR 4	402	M4	732
Lucas do Rio Verde-MT	SMR 4	402	M4	732
Nova Mutum-MT	SMR 4	402	M4	732
Santa Rita do Trivelato-MT	SMR 4	402	M4	732
Sinop-MT	SMR 4	402	M4	732
Sorriso-MT	SMR 4	402	M4	732
Santo Antonio do Leste-MT	SMR 4	403	M4	732
Santa Rosa do Tocantins-TO	SMR 4	404	M5	49
Barreiras-BA	SMR 4	405	M5	49
Correntina-BA	SMR 4	405	M5	49
Baixa Grande do Ribeiro-PI	SMR 5	501	M5	49
Campos Lindos-TO	SMR 5	501	M5	49
Caseara-TO	SMR 5	501	M5	49
Porto Nacional-TO	SMR 5	501	M5	49
São Domingos do Azeitão-MA	SMR 5	501	M5	49
Tasso Fragoso-MA	SMR 5	501	M5	49
Uruçuí-PI	SMR 5	501	M5	49

The RETEST trials comprised 50 treatments containing progeny with no replications, and 10 treatments containing controls that were repeated between each trial, using the ABD experimental scheme in plots of four rows of five meters, at a spacing of half a meter between rows and half a meter between plots. The data were analyzed using the R software (R CORE TEAM, 2016) to obtain the final yield of each progeny, considering the dataset from all the environments, respectively, using mixed-model methodology and including random effects, as per the formula:

$$\mathrm{Re}\,nd_{ijk} = \mu + G_i + E_j + B(E)_{jk} + e_{ijk} \quad (3)$$

where μ is the overall mean; G_i is the fixed effect of the *i*th genotype $(i = 1,2,..., p)$; E_i is the Random effect of the *j*th environment $\sim N(0, \sigma_E^2$; $B(E)_{jk})$ is the Random effect of the *k*th block in the *j*th environment $\sim N(0), \sigma_B^2$; e_{ijk} is the experimental error associated with the experimental unit of the *i*th genotype in the *k*th block of the do *j*th environment $N(0, \sigma_{ej}^2)$ with a different variance σ_{ej}^2 for each environment *j*.

To compare the effectiveness of the predictions (YLD BV) in relation to the actual potential of the progeny in the trials (YLD MROW and YLD RETEST), correlation analysis was used, which, as described by Henriques (2011), studies the relationship between a dependent variable and other independent variables, expressed by one equation that associates them all. According to Martins (2014), the coefficient of determination (R^2) gives the percentage variability of the independent variable that can be explained as a function of the variability of the dependent variable. The square root of the coefficient of determination corresponds to the correlation coefficient (r) whose value should vary between 0 and 1. A value of zero means that there is no linear relationship between the variables (HENRIQUES, 2011). When interpreting the correlations, three aspects should be considered: magnitude, direction, and significance (NOGUEIRA *et al.* 2012). The linear regression model is represented by:

$$Y = \beta_0 + \beta_1 X + e \quad (4)$$

where Y = Dependent variable; β_0 = Coefficient of intersection (Value of Y for X = 0); β1 = Inclination of the line (can be positive, negative or zero); x = Independent variable; e = Error due to random effects.

The degree of correlation between the variables (YLD BV, YLD MROW and YLD RETEST) was analyzed as per Devore (2006) and is shown in Table 2.

The correlations were analyzed using the R v4.2.1 software (R CORE TEAM, 2016), independently in the following different scenarios: Scenario M1, Scenario M2, Scenario M3, Scenario M4, Scenario M5, and jointly in a general scenario, to study the methodology applied in each SMR and breeding program.

Table 2 - Reference correlation coefficient (DEVORE, 2006)

Value of r	Definition
0.00 to 0.19	Very weak correlation
0.20 to 0.39	Weak correlation
0.40 to 0.69	Moderate correlation
0.7 to 0.89	Strong correlation
0.90 to 1.00	Very strong correlation

RESULTS AND DISCUSSION

In Scenario M1, the result of 0.3278 for r shows that there is a weak correlation between YLD BV and YLD MROW, i.e., around 32.7% of the values of YLD BV explain the result of YLD MROW. The result of 0.2206 for r shows that there is a weak correlation between YLD BV and YLD RETEST, around 22% of the values of YLD BV explaining the result of YLD RETEST, as shown in Table 3.

The results showed a weak correlation between YLD BV and YLD MROW (0.088). The values of the variables are highly dispersed and distanced in relation to the line, which makes this correlation effectively weak. The correlation between YLD BV and YLD RETEST is also weak and only minimally positive (0.086), with highly dispersed data, as shown in Figure 1.

Table 3 - Correlation analysis between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M1

	Dependent variable: YLD BV	
	(1)	(2)
YLD MROW	0.088***	
	(0.018)	
YLD RETEST		0.086***
		(0.026)
Constant	4,427.646***	4,422.639***
	(75.072)	(116.579)
Observations	206	206
R^2	0.108	0.049
r	0.3278	0.2206
Adjusted R^2	0.103	0.044
Residual Std. Error (df = 204)	83.690	86.406
F Statistic (df = 1; 204)	24.580***	10.434***

Note: *p < 0.1; **p < 0.05; ***p < 0.01

Figure 1 - Correlation between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M1

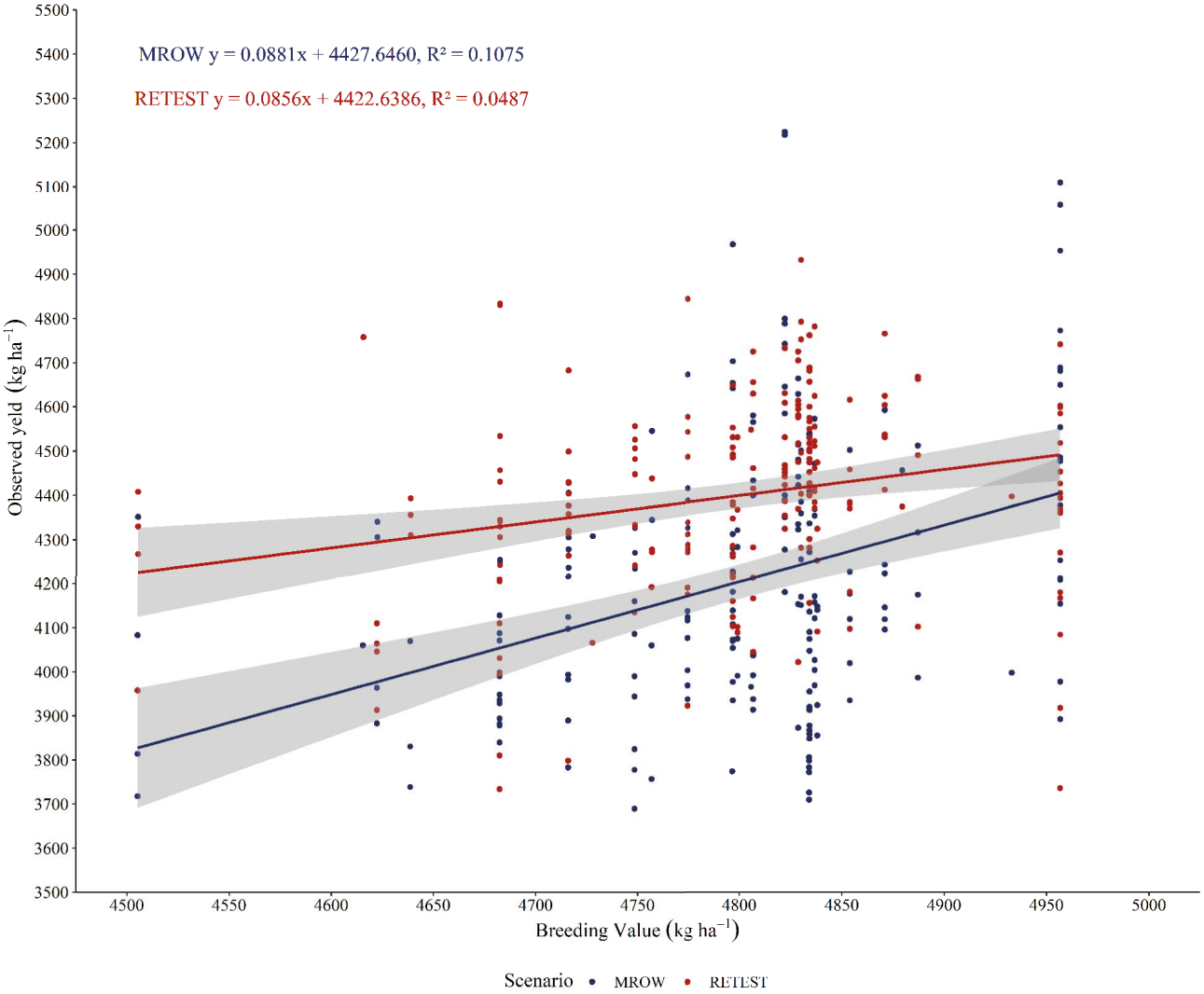


Table 4 - Correlation analysis between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M2

	Dependent variable: YLD BV	
	(1)	(2)
YLD MROW	-0.070**	
	(0.034)	
YLD RETEST		0.022
		(0.052)
Constant	4,594.055***	4,029.229***
	(226.818)	(242.660)
Observations	202	202
R^2	0.020	0.001
r	0.1428	0.030
Adjusted R^2	0.016	-0.004
Residual Std. Error (df = 200)	252.781	255.286
F Statistic (df = 1; 200)	4.169**	0.182

Note: *p < 0.1; **p < 0.05; ***p < 0.01

In Scenario M2, the result of 0.1428 for r shows that there is a very weak correlation between YLD BV and YLD MROW, i.e., 14.2% of the values for YLD BV explain the result of YLD MROW. The result of 0.030 for r shows that there is a very weak correlation between YLD BV and YLD RETEST, only 3% of the values for YLD BV explaining the result of YLD RETEST, as shown in Table 4.

The results also showed a very weak correlation between YLD BV and YLD MROW (-0.070) with highly dispersed data. The correlation between YLD BV and YLD RETEST is also weak (0.022) with highly dispersed data, as shown in Figure 2.

In Scenario M3, the result of 0.0360 for r shows that there is a very weak correlation between YLD BV and YLD MROW, where only 3.6% of the YLD BV values explain the result of YLD MROW. The result of 0.010 for r shows that there is a very weak correlation between YLD BV and YLD RETEST, only 1% of the YLD BV values explaining the result of YLD RETEST, as shown in Table 5.

In this Scenario there was a very weak correlation between YLD BV and YLD MROW (0.006) with highly dispersed data. The correlation between YLD BV and YLD RETEST is also very weak (-0.003), again showing highly dispersed data, as shown in Table 3.

In Scenario M4, the result of 0.0574 for r showed that there is a very weak correlation between YLD BV and YLD MROW, i.e., 5.7% of the values for YLD BV explain the result of YLD MROW. The result of 0.0223 for r shows that there is a very weak correlation between YLD BV and YLD RETEST, with only 2.2% of the values for YLD BV explaining the result of YLD RETEST, as shown in Table 6.

Figure 2 - Correlation between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M2

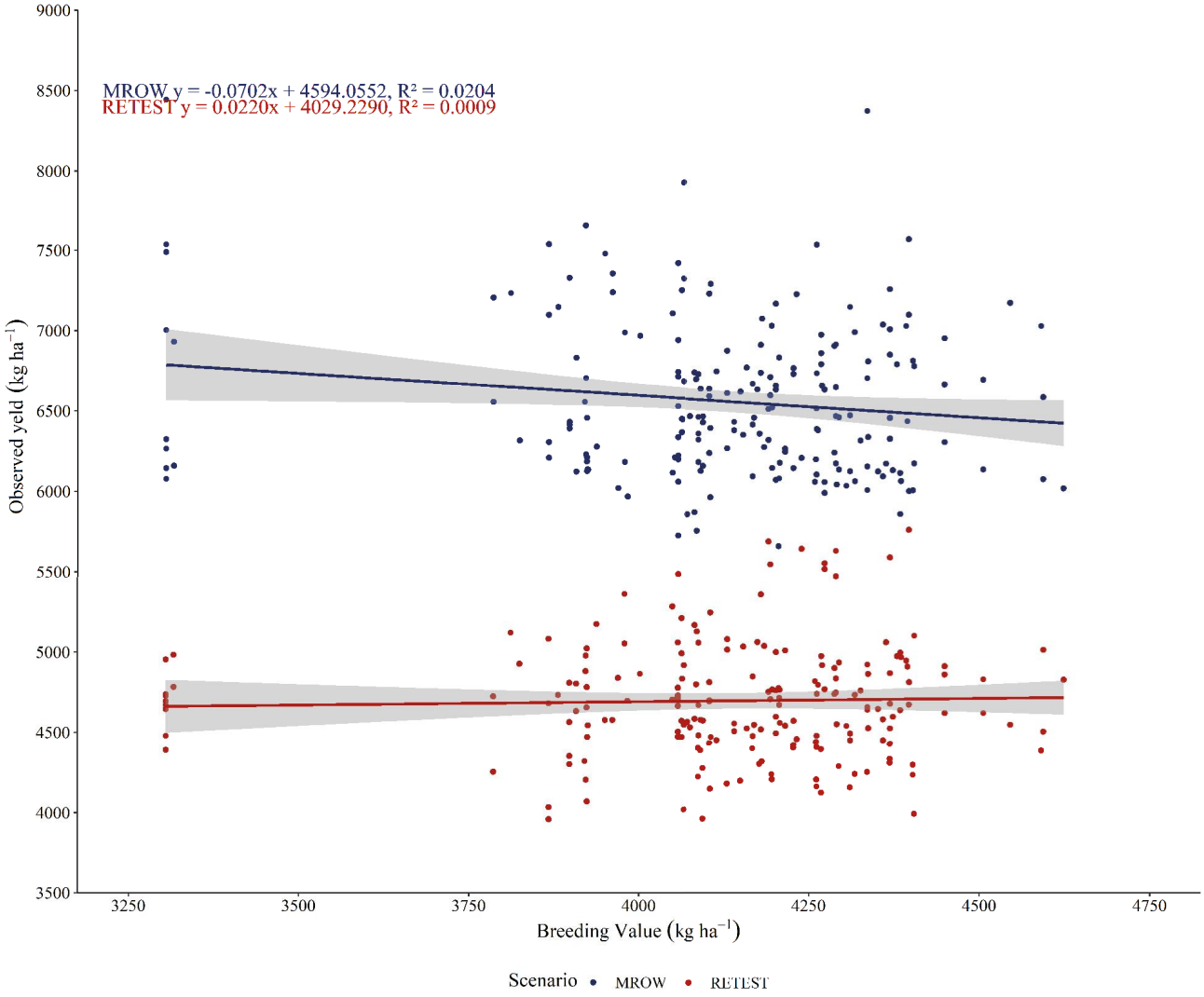


Table 5 - Correlation analysis between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M3

	Dependent variable: YLD BV	
	(1)	(2)
YLD MROW	0.006	
	(0.007)	
YLD RETEST		-0.003
		(0.015)
Constant	4,118.097***	4,164.108***
	(36.323)	(64.027)
Observations	679	679
R^2	0.001	0.0001
r	0.0360	0.010
Adjusted R^2	-0.0002	-0.001
Residual Std. Error (df = 677)	109.798	109.864
F Statistic (df = 1; 677)	0.856	0.039

Note: *p < 0.1; **p < 0.05; ***p < 0.01

Figure 3 - Correlation between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M3

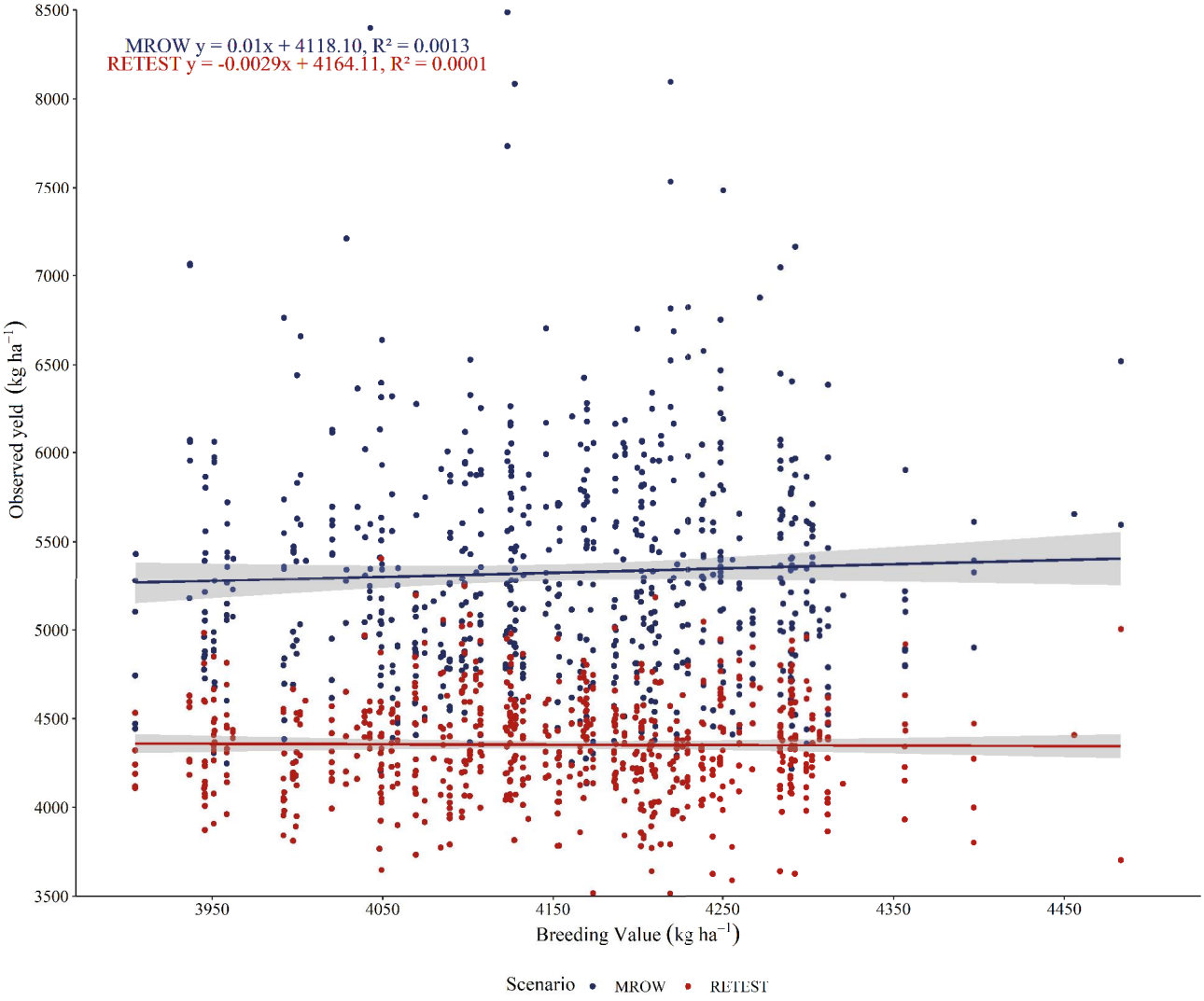
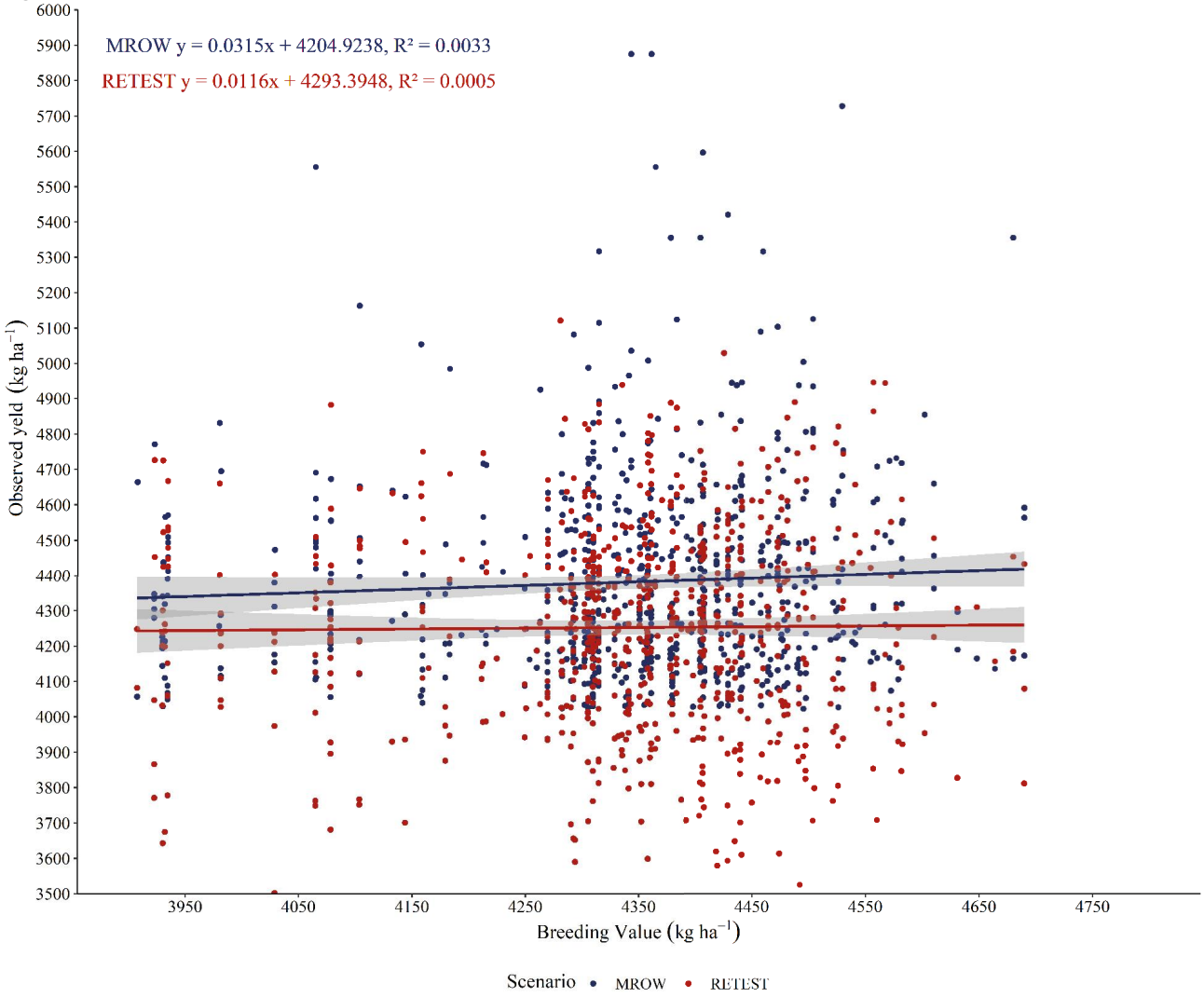


Table 6 - Correlation analysis between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M4

	Dependent variable: YLD BV	
	(1)	(2)
YLD MROW	0.031	
	(0.020)	
YLD RETEST		0.012
		(0.020)
Constant	4,204.924***	4,293.395***
	(89.043)	(83.968)
Observations	732	732
R^2	0.003	0.0005
r	0.0574	0.0223
Adjusted R^2	0.002	-0.001
Residual Std. Error (df = 730)	149.704	149.915
F Statistic (df = 1; 730)	2.408	0.348

Note: *p < 0.1; **p < 0.05; ***p < 0.01

Figure 4 - Correlation between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M4



There was also a very weak correlation between YLD BV and YLD MROW (0.031) showing high data dispersion. The correlation between YLD BV and YLD RETEST is also characterized as very weak (0.012) and with high data dispersion, as shown in Figure 4.

For Scenario M5, the result of 0.0141 for r shows that there was a very weak correlation between YLD BV and YLD MROW, i.e., only 1.4% of the values for YLD BV explain the result of YLD MROW. The result of 0.0812 for r shows that there is a very weak correlation between YLD BV and YLD RETEST, approximately 8.1% of the values for YLD BV explaining the result of YLD RETEST, as per Table 7.

The correlation between YLD BV and YLD MROW is very weak (0.004) and shows high data dispersion. The correlation between YLD BV and YLD RETEST is also very weak (-0.039) with high data dispersion (Figure 5).

The general scenario is shown in Table 8. The result of 0.4800 for r shows that there is a moderate correlation between YLD BV and YLD MROW, where around 48% of the values for YLD BV explain the result of YLD MROW. The result of 0.0447 for r shows that there is a very weak correlation between YLD BV and YLD RETEST, with approximately 4.4% of the values for YLD BV explaining the result of YLD RETEST.

The correlation between YLD BV and YLD MROW is classified as moderate and negative (-0.138) and is shown in Figure 6. The correlation between YLD BV and YLD RETEST was also very weak (-0.035) and with highly dispersed data.

As the environmental effect exerts a great influence on the behavior of the germplasm, it was certainly one of the causes of the differences in YLD between and in each SMR.

One probable explanation for this low and negative correlation may be connected to the prediction model used in the study, which may not be adjusted or suitable for estimating the correct values for BV. Another possible explanation could be related to the number of data environments of each parent used in predicting the crosses, since the data from one environment might be used against the data from up to 133 different environments, depending on the phase of each parent in the breeding program.

Despite being a viable alternative, because of the low availability of seeds during the initial stages of plant breeding programs, as described by Duarte, Vencovsky and Dias (2001), the ABD experimental model may have influenced the results, since the experimental error associated with the lack of repetitions can be significant (Silva and Silva, 1999). One way to reduce the experimental error might be to change to a layout that reduces the border effect, since, as described in Silva, Souza and Montenegro (1991), this effect can result in low experimental precision. However, these improvements are subject to the volume of seeds available in the initial generations of the breeding programs.

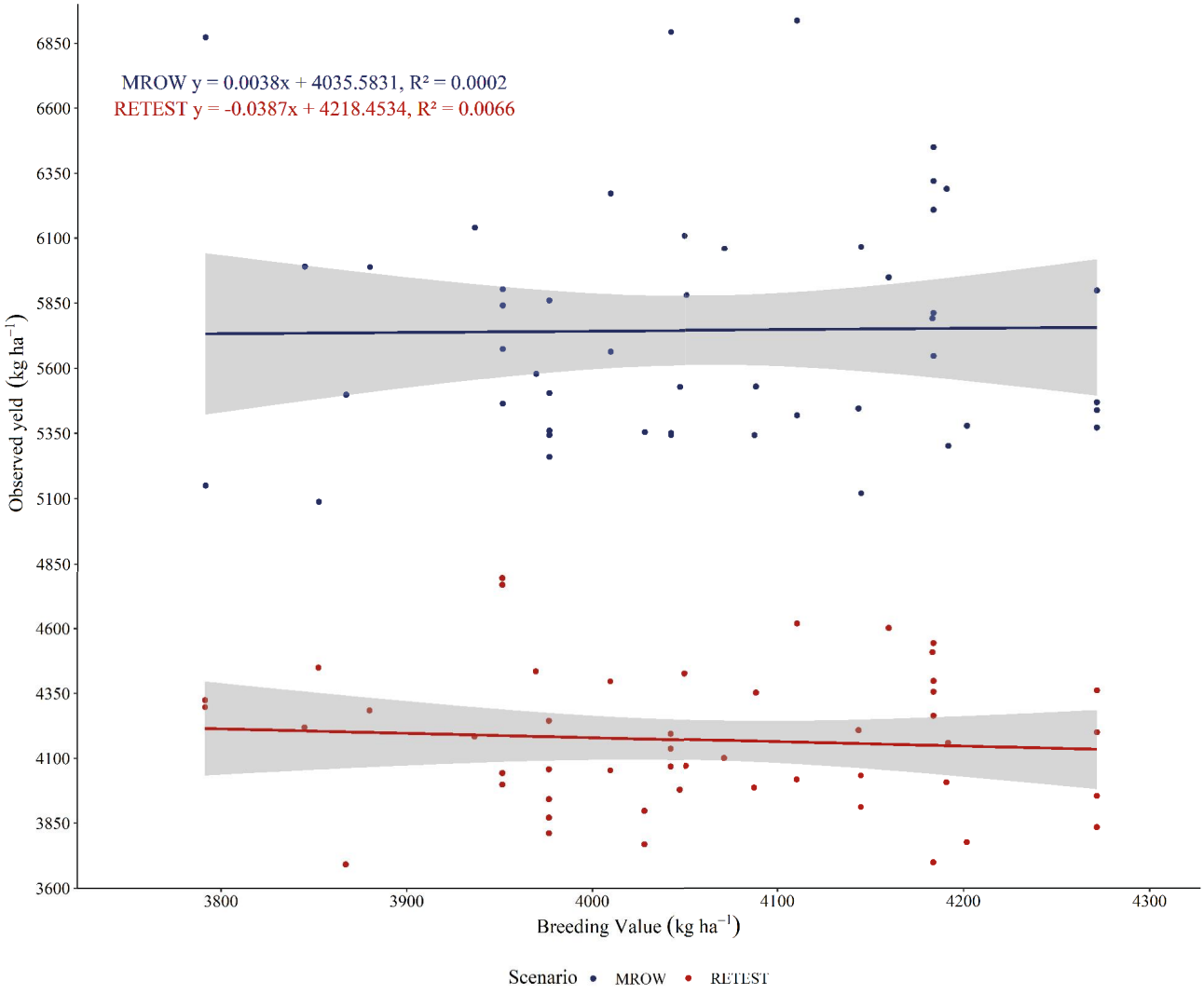
Another causal factor of the very weak to moderate correlation between variables may be related to the size and number of the different environments under analysis. Lima *et al*. (2008) describe how the genotype x environment (GxE) interaction is one of the

Table 7 - Correlation analysis between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M5

	Dependent variable: YLD BV	
	(1)	(2)
YLD MROW	0.004	
	(0.041)	
YLD RETEST		-0.039
		(0.069)
Constant	4,035.583***	4,218.453***
	(233.874)	(288.572)
Observations	49	49
R^2	0.0002	0.007
r	0.0141	0.0812
Adjusted R^2	-0.021	-0.015
Residual Std. Error (df = 47)	128.821	128.405
F Statistic (df = 1; 47)	0.009	0.314

Note: *p < 0.1; **p < 0.05; ***p < 0.01

Figure 5 - Correlation between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for Scenario M5



main complicating factors in the work of breeders. To reduce the effect of the GxE interaction, it is necessary to conduct experiments in a greater number of locations, evaluating the strength of the interaction and its possible impact when selecting and recommending genotypes.

Most agronomic characteristics are controlled by several genes, where the environment affects expression of phenotypic traits, especially quantitative traits such as YLD (LEITE *et al*., 2016). The prediction accuracy for any one characteristic across various environments can differ due to the GxE interaction (WANG *et al*., 2018).

Furthermore, most characteristics of agronomic importance have low heritability (BORÉM; MIRANDA, 2013). Because of this, one alternative might be the joint analysis of multiple traits that, according to Alimi *et al*. (2013), can improve prediction accuracy when using highly correlated characteristics, especially for some of the characteristics of low heritability. Heritability is positively related to prediction accuracy.

As these are progeny of generations F_3 and F_4 with a high degree of heterozygosity, the variation in YLD response is expected since the alleles are still undergoing genetic recombination. As stated by Mendonça *et al*. (2020), it is difficult to select for quantitative characteristics during the initial stages of breeding due to the high level of heterozygosity and the large number of new progeny. The correlation between the analyzed variables may show an increase in magnitude and a change in direction to the point where the degree of heterozygosity is reduced.

To accelerate the homozygous process, methods of rapid generation advancement can be efficiently adopted, advancing two, three or even four generations in the same

Table 8 - Correlation analysis between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for the general scenario and breeding programs

	Dependent variable: YLD BV	
	(1)	(2)
YLD MROW	-0.138***	
	(0.006)	
YLD RETEST		-0.035*
		(0.018)
Constant	4,978.505***	4,446.684***
	(29.431)	(79.666)
Observations	1,868	1,868
R^2	0.230	0.002
r	0.4800	0.0447
Adjusted R^2	0.230	0.001
Residual Std. Error (df = 1866)	218.778	249.133
F Statistic (df = 1; 1866)	558.559***	3.726*

Note: *p < 0.1; **p < 0.05; ***p < 0.01

year, depending on the cycle of the progeny and the region to be cultivated, and on such techniques and supplementary tools as greenhouses, heating and ventilation systems, supplementation or suppression of the light to change the photoperiod, the use of hormones, the early harvesting of seeds, and changes in the CO_2 concentration in controlled environments, among others. In this case, as the interest of the breeder is only to advance generations without carrying out any direct selection process where the number of seeds does not become a limiting factor, this rapid generation advancement can be achieved outside the original SMR of the breeding program, for example, in regions of low latitude. From the point of view of genetic gain, combining strategies that sum the selection of superior genotypes in the early generations of a breeding program, reducing the time and cost of obtaining a new cultivar, is undoubtedly the best strategy for the breeder to adopt.

Combining different methods can improve assertiveness in predicting crossings and selecting better individuals and/or better pedigrees, such as BLUP at the individual level (BLUPI) (RESENDE, 2002), simulated individual BLUP (BLUPIS) (RESENDE; BARBOSA, 2006) and modified simulated individual BLUP (BLUPISM – BLUP) (CASTRO *et al*., 2016). Furthermore, as stated by Bauer and León (2008), considering information of the parents or pedigree tends to be more efficient than only considering information of the progeny, as is the case when using a kinship matrix, as reported by Resende and Alves (2021) and Clark *et al*. (2012).

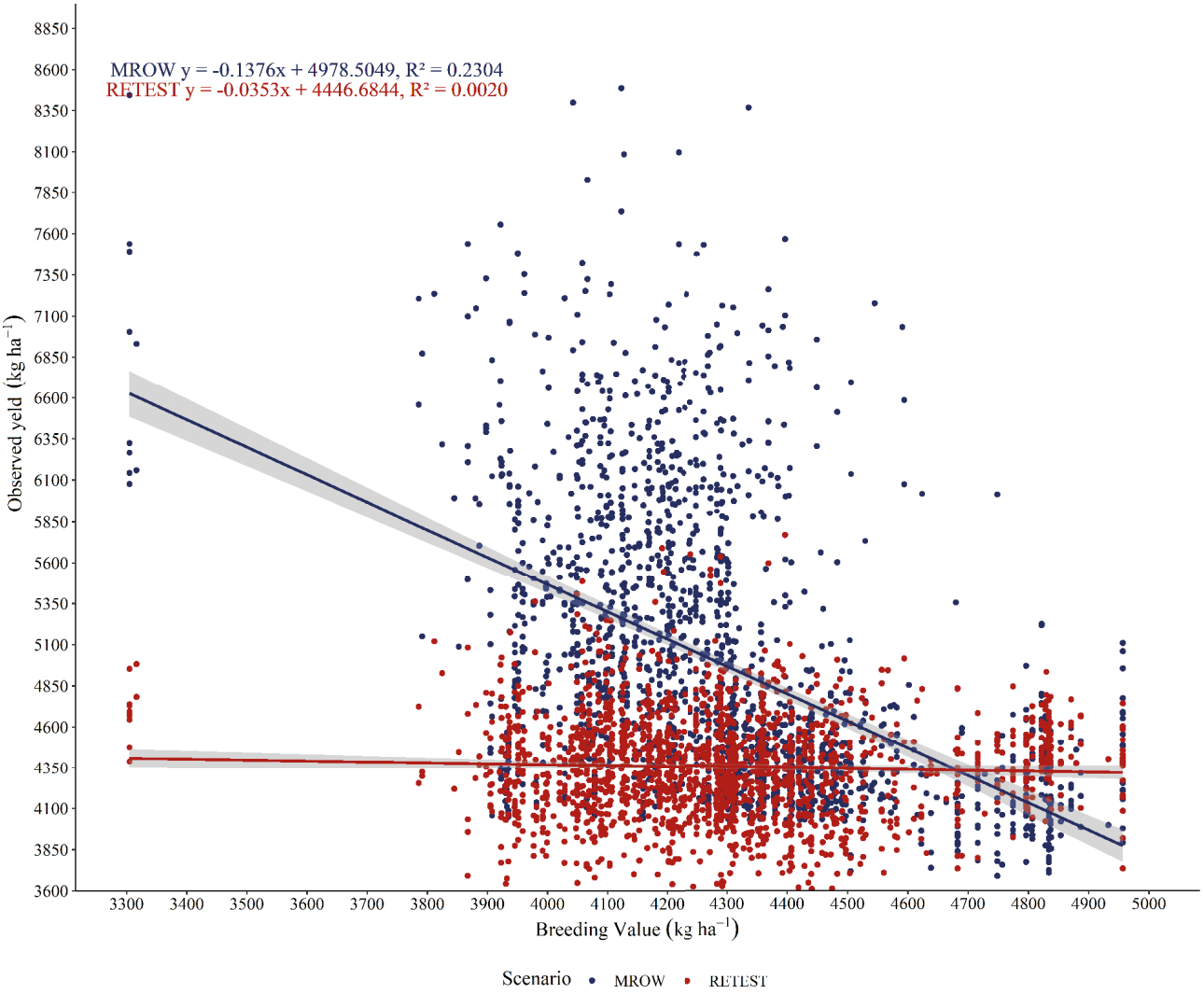
When using parents that are genetically highly dissimilar in crosses to increase genetic variability, the process of progeny selection in early generations becomes more complex and challenging. As such, the response of these correlations may be associated with divergent crosses; there is a need for a better understanding of the genetic dissimilarity between parents and the direct and indirect effects on the progeny.

In the correlations in each of the presented scenarios the data was highly dispersed, and culminated in low correlation between the variables, indicating that the regression model may not fit the data.

From the genetic point of view, another factor to be considered in the results is related the germplasm used in this study, which is essentially conventional. According to the International Service Acquisition of Agri-biotech Applications (2018), around 94% of the soybean planted in Brazil is genetically modified (GMO). The wide adoption of new biotechnologies has made practically all the companies carrying out soybean breeding, whether public or, particularly, private, migrate their research to GMOs, causing a drastic reduction in the commercial availability of conventional cultivars, and severely limiting and narrowing the genetic base.

In the germplasm of a breeding program that includes superior genotypes, it cannot necessarily be considered that these will inevitably generate agronomically superior progeny when used as parents, albeit recurrent selection proves to be efficient in most cases.

Figure 6 - Correlation between the results of YLD BV x YLD MROW and YLD BV x YLD RETEST for the general scenario



The correlation between YLD BV and YLD MROW in the general scenario, with an r value of 0.4800, was moderate, showing that, for this scenario, predicting crosses can be used as one way of identifying crosses of greater potential, or even of discarding the worst crosses, providing it is used cautiously. As a comparison, Mendonça *et al*. (2020) achieved models that reached predictive abilities of between 0.40 and 0.56, thereby allowing low-intensity selection to be applied in F_2. As a result, half of the progeny could be discarded without major losses, showing that with the use of genomic prediction, it is possible to select for quantitative characteristics during the initial stages of breeding.

Although the use of phenotypic data in this study to predict the performance of crosses and identify better crosses that might generate progeny of agronomically superior characteristics has not proven to be highly efficient to the point of being widely applied, several other studies show extremely positive results when using this methodology, such as Xu, Zhu and Zhang (2014) with rice, Daetwyler *et al.* (2014) with wheat, and Mendonça *et al*. (2020) working with segregating populations and soybean progeny.

Most of the reported results derive from the study of homozygous populations (DUHNEN *et al*., 2017; JARQUÍN *et al*., 2014; ZHANG *et al*., 2016), and consider different crops and the GxE interaction. However, this does not include the complete set of situations for which prediction can be used. As such, little information on performance is available during the early stages of breeding to predict segregating progeny or populations (MENDONÇA *et al*., 2020).

Certainly, with advances in the processes of genetic improvement, the adoption of new methodologies and equipment, new biotechnological tools, statistical models, more improved predictive models, high-throughput phenotyping, and reductions in the costs of genotyping and data analysis, the time and resources spent on obtaining progeny of high productive potential and with superior agronomic characteristics tends to be reduced, thereby increasing genetic gain per year of breeding. As such, the genetic improvement of plants will continue to make a considerable contribution to increasing productivity.

CONCLUSION

The model for predicting the performance of crosses using estimates of the breeding value was not very efficient in initially identifying crosses with a high potential for generating agronomically superior soybean progeny for the soybean regions of Brazil. The correlations between YLD BV (estimates of the breeding value) x YLD MROW (F_3 progeny) and YLD BV x YLD RETEST (F_3 progeny) were classed as very weak to moderate.

REFERENCES

ALIMI, N. A. *et al*. Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. **Theoretical and Applied Genetics**, v. 126, n. 10, p. 2597-2625, 2013.

ALLARD, R. W. **Principles of plant breeding**. New York: John Wiley & Sons, 1999.

BAUER, A. M.; LÉON, J. Multiple-trait breeding values for parental selection in selfpollinating crops. **Theoretical and Applied Genetics**, v. 116, n. 2, p. 235-242, Jan. 2008.

BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125**,** p. 375-385, 2020. DOI: https://doi.org/10.1038/s41437-020-0312-1.

BORÉM, A.; MIRANDA, G. V. **Melhoramento de plantas**. Viçosa, MG: UFV, 2013. 523 p.

BUTLER, D. G. *et al*. **ASReml-R reference manual, Version 4**. Hemel Hempstead, UK: VSN International, 2017.

CASTRO, R. D. *et al*. Selection between and within full-sib sugarcane families using the modified BLUPIS method (BLUPISM). **Genetics and Molecular Research**, v. 15, n. 1, 2016.

CLARK, S. A. *et al*. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. **Genetics Selection Evolution**, v. 44, n. 1, p. 1-9, 2012.

DAETWYLER, H. D. *et al*. Genomic prediction for rust resistance in diverse wheat landraces. **Theoretical and Applied Genetics**, v. 127, n. 8, p. 1795-1803, 2014.

DEVORE, J. L. **Probabilidade e estatística**: para engenharia e ciências. São Paulo: Thomson Pioneira, 2006. 706 p.

DUARTE, J. B.; VENCOVSKY, R.; DIAS, C. T. D. S. Estimadores de componentes de variância em delineamento de blocos aumentados com tratamentos novos de uma ou mais populações. **Pesquisa Agropecuária Brasileira**, v. 36, p. 1155-1167, 2001.

DUHNEN, A. *et al*. Genomic selection for yield and seed protein content in soybean: a study of breeding program data and assessment of prediction accuracy. **Crop Science**, v. 57, p. 1325-1337, 2017. DOI: https://doi.org/10.2135/cropsci2016.06.0496.

FEDERER, W. T. Augmented (hoonuiaku) designs. **Hawaian Planters' Record**, v. 55, p. 191-208, 1956.

FRITSCHE-NETO, R. **Técnicas experimentais e suas relações com a Lei de Proteção de Cultivares**. Piracicaba: ESALQ/USP. Departamento de Genética, 2013. Disponível em: https://edisciplinas.usp.br/pluginfile.php/978151/mod_resource/content/0/Aula%204.pdf. Acesso em: 17 abr. 2022.

HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. Genomic selection for crop breeding. **Crop Science**, v. 49, n. 1, p. 1-12, 2009.

HENRIQUES, C. **Análise de regressão linear simples e múltipla**. Viseu: Escola Superior de Tecnologia de Viseu. Departamento de Matemática, 2011.

INTERNATIONAL SERVICE FOR THE ACQUISITION OF AGRI-BIOTECH APPLICATIONS. **Global status of commercialized biotech/GM crops in 2018**: biotech crops continue to help meet the challenges of increased population and climate change. Ithaca: ISAAA, 2018.

JARQUÍN, D. *et al*. Genotyping by sequencing for genomic prediction in a soybean breeding population. **Bmc Genomics**, v. 15, n. 740, 2014. DOI: https://doi.org/10.1186/1471-2164-15-740.

KASTER, M.; FARIAS, J. R. B. **Regionalização dos testes de valor de cultivo e uso e da indicação de cultivares de soja**: terceira aproximação. Londrina: Embrapa Soja, 2012. 69 p.

LEITE, W. S. *et al*. Estimativas de parâmetros genéticos, correlações e índices de seleção para seis caracteres agronômicos em linhagens F8 de soja. **Comunicata Scientiae**, v. 7, n. 3, p. 302-310, 2016.

LIMA, W. F. *et al*. Interação genótipo-ambiente de soja convencional e transgênica resistente a glifosato, no Estado do Paraná. **Pesquisa Agropecuária Brasileira**, v. 43, p. 729-736, 2008.

MARTINS, M. E. G. Coeficiente de correlação amostral. **Revista de Ciência Elementar**, v. 2, n. 2, p. 34-36, 2014.

MENDONÇA, L. *et al*. Genomic prediction enables early but low-intensity selection in soybean segregating progenies. **Crop Science**, v. 60, n. 3, p. 1346-1361, 2020.

NOGUEIRA, A. P. O. *et al*. Análise de trilha e correlações entre caracteres em soja cultivada em duas épocas de semeadura. **Bioscience Journal**, v. 28, p. 877-888, 2012.

PINHEIRO, L. C. de M. *et al*. Parentesco na seleção para produtividade e teores de óleo e proteína de soja via modelos mistos. **Pesquisa Agropecuária Brasileira**, v. 48, p. 1246- 1253, 2013.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2016. Disponível em: http://www.R-project.org. Acesso em: 15 jul. 2021.

RESENDE, M. D. V. de; BARBOSA, M. H. P. Selection via simulated individual BLUP based on family genotypic effects in sugarcane. **Pesquisa Agropecuária Brasileira**, v. 41, n. 3, p. 421-429, mar. 2006

RESENDE, M. D. V. de. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília, DF: Embrapa Informação Tecnológica; Colombo: Embrapa Florestas, 2002.

RESENDE, M. D. V.; ALVES, R. S. Genética: estratégias de melhoramento e métodos de seleção. *In*: OLIVEIRA, E. B.; PINTO JÚNIOR, J. E. (org.). **O eucalipto e a Embrapa**: quatro décadas de pesquisa e desenvolvimento. Brasília, DF: Embrapa, 2021. p. 171-202.

SILVA, I. P.; SILVA, J. A. A. **Métodos estatísticos aplicados à pesquisa científica**: uma abordagem para profissionais da pesquisa agropecuária. Recife: UFRPE, 1999. 305 p.

SILVA, P. S. L.; SOUZA, P. G.; MONTENEGRO, E. E. Efeito de bordadura nas extremidades de parcelas de milho irrigado. **Ceres**, v. 38, n. 216, p. 101-107, 1991.

WANG, X. *et al*. Genomic selection methods for crop breeding: current status and prospects. **The Crop Journal**, v. 6, n. 4, p. 330-340, 2018.

XU, S.; ZHU, D.; ZHANG, Q. Predicting hybrid performance in rice using genomic best linear unbiased prediction. **Proceedings of the National Academy of Sciences**, v. 111, n. 34, p. 12456-12461, 2014.

ZHANG, J. *et al*. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). **Theoretical and Applied Genetics,** v. 129, p. 117-130, 2016. DOI: https://doi.org/10.1007/s00122-015-2614-x. Acesso em: 19 abr. 2022.