



Inference of population effect and progeny selection via a multi-trait index in soybean breeding

Leonardo Volpato^{1*}, João Romero do Amaral Santos de Carvalho Rocha², Rodrigo Silva Alves², Willian Hytalo Ludke¹, Aluizio Borém¹ and Felipe Lopes da Silva¹

¹Departamento de Fitotecnia, Universidade Federal de Viçosa, Av. Peter Henry Rolfs, s/n, 36570-900, Viçosa, Minas Gerais, Brazil. ²Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil. *Author for correspondence. E-mail: leo.agroufv@gmail.com

ABSTRACT. The selection of superior genotypes of soybean entails a simultaneous evaluation of a number of favorable traits that provide a comparatively superior yield. Disregarding the population effect in the statistical model may compromise the estimate of variance components and the prediction of genetic values. The present study was undertaken to investigate the importance of including population effect in the statistical model and to determine the effectiveness of the index based on factor analysis and ideotype design via best linear unbiased prediction (FAI-BLUP) in the selection of erect, early, and high-yielding soybean progenies. To attain these objectives, 204 soybean progenies originating from three populations were examined for various traits of agronomic interest. The inclusion of the population effect in the statistical model was relevant in the genetic evaluation of soybean progenies. To quantify the effectiveness of the FAI-BLUP index, genetic gains were predicted and compared with those obtained by the Smith-Hazel and Additive Genetic indices. The FAI-BLUP index was effective in the selection of progenies with balanced, desirable genetic gains for all traits simultaneously. Therefore, the FAI-BLUP index is an adequate tool for the simultaneous selection of important traits in soybean breeding.

Keywords: mixed-model methodology; best linear unbiased prediction; selection index; genotype x environment interaction; factor analysis; ideotype design.

Received on September 17, 2018.
Accepted on February 21, 2019.

Introduction

Similar to most autogamous species, soybean (*Glycine max* (L.) Merrill) breeding is aimed at the development and identification of superior genotypes and the release of new cultivars. To produce better cultivars, breeding programs of autogamous plants have intensified the production of segregating populations. To this end, a large number of progenies from several crosses are typically generated each year, resulting in several populations (Bernardo, 2003; Resende et al., 2016).

In traditional methods, progenies are selected without considering the merits of the populations to which they belong. However, it is important to use statistical models that include not only the effects of progenies but also the effects of populations (Resende et al., 2016). Including the population effect in the statistical model to select the best progenies within the best populations is a measure aimed at increased selection accuracy. Duarte and Vencovsky (2001), Piepho and Williams (2006), Resende et al. (2016), and Pereira et al. (2017) highlighted the benefits of including population effect in the model.

Disregarding population effect in the statistical model may compromise the prediction of genetic values and the estimation of variance components. In this way, selection gains tend to decrease due to selection bias (Rocha & Vello, 1999; Duarte & Vencovsky, 2001; Pereira et al., 2017).

In the selection process, in addition to choosing the best statistical model to predict genetic values, plant breeders usually handle multiple traits simultaneously (Akhter & Sneller, 1996; Malek, Rafii, Shahida Sharmin Afroz, Nath, & Mondal, 2014). However, choosing high-performance soybean genotypes for multiple traits simultaneously can be a difficult task. Some selection indices have been proposed for a simultaneous selection of traits, e.g., the Smith-Hazel classical index (Smith, 1936; Hazel, 1943) and the Additive Genetic index (Resende, 2016). However, several limitations exist regarding the determination of economic weights of traits. Moreover, the Smith-Hazel index may have multicollinearity problems, undermining the effectiveness of best progeny selection.

In an effort to address the questions above, Rocha, Machado, and Carneiro (2018) proposed a multi-trait index based on factor analysis and ideotype design (FAI-BLUP index). This index considers the genetic correlation structure obtained from the data (via exploratory factor analysis according to the method description), which enables the selection of genotypes closer to those hypothesized by the breeder using the ideotype combination of desirable and undesirable factors for the selection objective.

In this scenario, the present study proposes i) to examine the importance of including population effect in the statistical model for the prediction of genetic values and ii) to evaluate the effectiveness of the FAI-BLUP index in the selection of erect, early, and high-yielding progenies of soybean.

Material and methods

Genetic material and experimental settings

Three populations belonging to the Soybean Breeding Program of the Federal University of Viçosa (UFV) were obtained from crosses between divergent inbred lines (TMG 123 RR/M7211 RR, UFVS Citrino RR/UFVS Turqueza RR, and M7908 RR/M7211 RR) for relative maturity groups in accordance with Carpentieri-Pípolo, De Almeida, De Souza Kiihl, and Rosolem (2000); we aimed to select earlier-maturing and high-yielding progenies. The F_2 plants were separately bulk-harvested and threshed to obtain a bulk sampling of seeds produced to form 204 $F_{2:3}$ progenies, which were evaluated in the field along with nine controls.

Two trials were conducted in the 2015 crop year until mid-March 2016. The first trial took place in the municipality of Viçosa, Minas Gerais State, Brazil (20°45'45" S, 42°49'27" W, and 647 m altitude), and the second was carried out in São Jose do Triunfo, Minas Gerais State, Brazil (20°45'14" S, 42°52'55" W, and 667 m altitude). The experiments were set up as randomized complete block designs with three replicates. Plots consisted of two 2.0-m rows spaced 0.5 m apart with a planting stand of 13 seeds per meter, totaling a density of 256,000 plants ha⁻¹. All management and training operations were undertaken according to the crop requirements, following recommendations of Sedyama, Felipe, and Borem (2015).

The following traits were evaluated: number of days to flowering (FL, days); number of days to maturity (MT, days); seed-filling period (FP, days); hypocotyl diameter, measured just above the hypocotyl node (HD, mm); lodging angle (LA, degrees); 100-seed weight (SW, g); average seed yield per plant (SYPL, g/plant), corrected 'for the stand' (plant survival rate); and average seed yield per plot (SY, g). The FL and MT variables correspond to the number of days from seedling emergence until more than 50% of the plants in the plot reached the R2 stage and the number of days before 95% of the pods were mature, as indicated by their color, respectively (Fehr & Caviness, 1977). The seed-filling period was determined as the difference between FL and MT (Panthee et al., 2004). Lodging angle was measured at maturity, ranging from 81° (all plants erect) to 9° (all plants prostrate). To determine the SW, SYPL, and SY traits, the seeds were dried until reaching 13% moisture before analysis.

Statistical analyses

The Restricted Maximum Likelihood/Best Linear Unbiased Prediction (REML/BLUP) procedure was adopted for statistical analyses, following Patterson and Thompson (1971) and Henderson (1975). The statistical model associated with the evaluation of genotypes in a randomized complete block design, with one observation per plot in more than one environment and with population data, as shown below:

$$y = X_r + Z_f + W_i + S_p + T_j + e$$

where: y is the vector of phenotypic data; r is the vector of fixed effects (controls, replicate, and location) added to the overall mean; f is the effect of $F_{2:3}$ progeny among progenies within population (assumed random), in which $f \sim N(0, I\sigma_f^2)$; i is the vector of progeny x location interaction effects (random), in which, $i \sim N(0, I\sigma_i^2)$; p is the vector of among-population effects (assumed random), in which, $p \sim N(0, I\sigma_p^2)$; j is the vector of population x location interaction effects (random), in which, $j \sim N(0, I\sigma_j^2)$; and e is the error vector (random), in which $e \sim N(0, I\sigma_e^2)$. The capital letters (X , Z , W , S , and T) represent the incidence matrices for the r , f , i , p , and j effects, respectively. For the model without among-population effects, we removed the among-population and among population x environment interaction effects from the above model.

Evaluation of population effect and selection of progenies for multiple traits

The fit of the models with and without the among-population effects was compared by the Akaike information criterion (AIC) (Akaike, 1974) and the likelihood ratio test (LRT), following Wilks (1938) and using chi-square statistics with one degree of freedom.

After the best-fitting model was chosen, it was used to predict the genetic values (BLUPs) of the progenies; these were calculated as the sum of significant among-progeny within-population and among-population effects. Otherwise, the values were calculated only as the effect that was significant. These genetic values were used in three different selection indices: i) the classical Smith-Hazel index (Smith, 1936; Hazel, 1943); ii) the Additive Genetic index (AGI) (Resende, 2016); and, lastly, iii) the FAI-BLUP index proposed by Rocha et al. (2018). For all indices, selection was aimed at reducing FL and MT and increasing the other traits. For i), the multicollinearity diagnosis was carried out in the phenotypic correlation matrix, as recommended by Montgomery and Peck (1992), and the variables that provided a condition number higher than 100 were discarded to solve multicollinearity problems. The genetic coefficients of variation of the progenies were used as the relative economic weight for indices i) and ii) (Bhering et al., 2012). For iii), after ideotypes were determined, the distances from each genotype according to ideotypes (genotype-ideotype distance) were estimated and converted into spatial probability, enabling the genotype ranking (Rocha et al., 2018). Oblique criterion rotation was used (Coan, 1959) for analytic rotation and the factor scores were calculated using weighted least squares method according to Bartlett (1938).

In oblique rotations, the assumptions of independent factors are relaxed, and the new axes are free to take any position in the factor space. However, the degree of correlation allowed among factors is generally small because two highly correlated factors are better interpreted as only one factor. Oblique rotations relax the orthogonality constraint to gain simplicity in the interpretation. Oblique rotation produces new axes that are not required to be orthogonal (Bernaards & Jennrich, 2005).

Index comparisons

Comparisons among the SH-BLUP, AGI, and FAI-BLUP indices were carried out based on the predicted genetic gains. For a more valid comparison, predicted genetic gains were calculated using the genotypes indicated by the classical Smith-Hazel index based on the genetic values (SH-BLUP) and using the genotypes indicated by the AGI and FAI-BLUP indices. Lastly, the 24 best progenies (12% selection intensity, approximately) were selected according to each index.

Software

R software (R.Core Team, 2017) was used for deviance analysis, prediction of genetic values using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015), and adjustment of the FAI-BLUP index with a protocol provided by Rocha et al. (2018). Selegen-REML/BLUP software (Resende, 2016) was used for genetic variance-covariance and genetic correlation analyses and to adjust the AGI index. The SH-BLUP index was run in the GENES software (Cruz, 2013).

Results

Population effect in the models

According to AIC, the model including population effects ($Pop(+)$) showed the best fit (lowest AIC value) for all traits (Table 1); thus, it was used for the prediction of genetic values. Table 1 also presents the results of deviance analysis for the $Pop(+)$ model and the model without the effect of populations ($Pop(-)$). In both analyzed models, a significant effect was observed for the estimates of variance associated with the effect of progeny within populations or among progenies for all traits. However, for the $Pop(+)$ model, significant progeny x environment interaction effects ($p \leq 0.05$) were only observed for the FL, FP, LA, and SW traits. The $Pop(-)$ model also showed a significant effect of the progeny x environment interaction for the same traits, but the SYPL and SY variables were also significant. However, for the among-population effects, variability was present ($p \leq 0.05$) for the FL, MT, FP, DH, and SW traits. For the population x environment interaction effects, a significant effect was only detected for the LA, SYPL, and SY traits.

Table 1. Deviance (Dev.) and likelihood ratio test (LRT) for number of days to flowering (FL), number of days to maturity (MT), seed-filling period (FP), hypocotyl diameter (HD), lodging angle (LA), 100-seed weight (SW), average seed yield per plant (SYPL), and average seed yield per plot (SY) evaluated in 204 soybean $F_{2:3}$ progenies (G, among progenies within population or among progenies) derived from three populations (P, among populations), in two environments (E).

Effect	FL				MT			
	Pop(+)		Pop(-)		Pop(+)		Pop(-)	
	Dev.	LRT [#]	Dev.	LRT	Dev.	LRT	Dev.	LRT
Progeny (G)	5522.3	117.8*	5684.4	24.4*	7202.7	78.1*	7387.1	159.2*
G×E interaction	5426.6	22.1*	5996.2	336.2*	7127.9	3.3 ^{NS}	7230.6	2.7 ^{NS}
Population (P)	5414.5	10*	-	-	7132.8	8.2*	-	-
P×E interaction	5405.1	0.6 ^{NS}	-	-	7124.7	0.10 ^{NS}	-	-
Full model	5404.5	-	5660	-	7124.6	-	7227.9	-
AIC	5462.484		5714.019		7182.630		7281.870	
Effect	Pop(+)		Pop(-)		Pop(+)		Pop(-)	
	Dev.	LRT	Dev.	LRT	Dev.	LRT	Dev.	LRT
	Progeny (G)	7146.1	26.8*	7189.8	40.8*	3284.1	17.2*	3365.7
G×E interaction	7124.8	5.5*	7153.8	4.8*	3266.9	0 ^{NS}	3328.1	0 ^{NS}
Population (P)	7123.6	4.3*	-	-	3274	7.1*	-	-
P×E interaction	7119.6	0.3 ^{NS}	-	-	3266.9	0 ^{NS}	-	-
Full model	7119.3	-	7149	-	3266.9	-	3328.1	-
AIC	7177.316		7202.983		3324.941		3382.065	
Effect	Pop(+)		Pop(-)		Pop(+)		Pop(-)	
	Dev.	LRT	Dev.	LRT	Dev.	LRT	Dev.	LRT
	Progeny (G)	9229.2	39.8*	9245.1	33.8*	4568.9	28.5*	4831.6
G×E interaction	9203.4	14*	9235.1	23.8*	4553.3	12.9*	4725.8	11.6*
Population (P)	9189.4	0 ^{NS}	-	-	4548.1	7.7*	-	-
P×E interaction	9208.5	19.1*	-	-	4541.2	0.8 ^{NS}	-	-
Full model	9189.4	-	9211.3	-	4540.4	-	4714.2	-
AIC	9247.401		9265.291		4598.434		4768.21	
Effect	Pop(+)		Pop(-)		Pop(+)		Pop(-)	
	Dev.	LRT	Dev.	LRT	Dev.	LRT	Dev.	LRT
	Progeny (G)	7326.9	8.6*	7474.9	28.6*	13338	7.0*	13519.3
G×E interaction	7320.3	2 ^{NS}	7458.1	11.8*	13333.5	2.5 ^{NS}	13536.6	31.3*
Population (P)	7319.5	1.2 ^{NS}	-	-	13331.3	0.3 ^{NS}	-	-
P×E interaction	7347.6	29.3*	-	-	13400.8	69.8*	-	-
Full model	7318.3	-	7446.3	-	13331	-	13505.3	-
AIC	7376.329		7500.337		13388.980		13559.320	

*,^{NS} Significant and not significant at 5% probability level by the chi-square test considering 1 degree of freedom, respectively. [#] Likelihood ratio test performed with the difference between the full model and model effects.

Exploratory factor analysis - FAI-BLUP index

The first four principal components had eigenvalues higher than one (Kaiser, 1958). Thus, the data may be condensed (dimensional reduction) into four factors that explain 90% of total variability. After oblimin rotation (Table 2), high genetic correlation for the first factor was observed among the DH, FP, and SW traits; the first factor was thus named the ‘accumulation factor’. For the second factor, high genetic correlation was observed between SYPL and SY; therefore, it was named the ‘yield factor’. The third factor was named ‘time factor’, since high genetic correlation was observed between MT and FL. The fourth and last factor was termed ‘lodging’ because LA was the only variable to have a high load on this factor.

Table 2. Factor loadings after oblimin rotation and communalities.

Trait [#]	Factor				Communality
	Accumulation	Yield	Time	Lodging	
SW	-0.9253	-0.0249	-0.122	-0.0213	0.8721
FP	-0.8581	0.0425	-0.0218	0.1265	0.7547
HD	-0.85	-0.0209	0.1845	-0.1415	0.7769
SYPL	0.0106	0.9683	0.0236	0.0344	0.9395
SY	-0.0037	0.9671	-0.0241	-0.0435	0.9378
FL	0.1942	-0.0101	0.9963	-0.0306	1.0313
MT	-0.3315	0.0196	0.8609	0.0633	0.8554
LA	-0.0176	0.009	-0.0122	-0.9894	0.9795
Mean	-	-	-	-	0.8934

[#]FL = days to flowering; MT = days to maturity; FP = seed-filling period; HD = hypocotyl diameter; LA = lodging angle; SW = 100-seed weight; SYPL = average seed yield per plant; and SY = average seed yield per plot.

Index comparisons

Table 3 describes the comparisons among the indices based on predicted genetic gains. According to the SH-BLUP index, the selection and recombination of the 24 best progenies would lead to undesirable gains for FP, whereas AGI indicated undesirable gains for MT. On the other hand, the gains predicted by the FAI-BLUP index were as desired for all traits.

Direct selection provides the maximum predicted gain when considering one trait at a time. The SH-BLUP, AGI, and FAI-BLUP indices revealed respective gains of 6.32, 48.20, and 48.32% in direct selection for traits whose values are desired to be increased; the indices revealed gains of 58.95, 11.66, and 24.10% for traits whose values are desired to be reduced (Table 3). These results indicate greater balance in the gains obtained via FAI-BLUP. The 24 best progenies were selected using the SH-BLUP, AGI, and FAI-BLUP indices. Coincidences between the progenies selected by the FAI-BLUP and AGI, FAI-BLUP and SH-BLUP, and AGI and SH-BLUP indices were 75, 25, and 16.16%, respectively.

Table 3. Predicted genetic gains based on direct selection by genetic values, SH-BLUP index, AGI-BLUP, and FAI-BLUP index.

Factor	Trait [#]	Genetic value	GCV (%) [†]	Predicted genetic gain ^{##} (%)			
				Direct selection	SH-BLUP	AGI-BLUP	FAI-BLUP
1	FP [‡]	6.05	3.93	6.73	-0.56	2.25	2.93
	HD	0.88	4.11	9.06	0.42	3.68	5.24
	SW	3.48	5.81	19.11	1.66	7.05	10.64
2	SYPL	2.33	10.38	9.57	0.95	4.57	3.41
	SY	37.35	10.51	9.22	1.65	4.98	3.77
3	FL	-7.81	5.61	-15.53	-10.22	-3.28	-5.85
	MT	-12.37	3.64	-8.83	-4.14	0.44	-0.02
4	LA	12.94	19.89	23.84	2.44	14.84	11.47
Total desirable (positive) gain (%) [§]				-	6.32	48.2	48.32
Total desirable (negative) gain (%) [¶]				-	58.95	11.66	24.1
Genotypes selected by the SH index - 159, 359, 343, 155, 160, 157, 162, 229, 337, 364, 167, 370, 335, 147, 152, 161, 310, 150, 110, 136, 327, 163, 258, 332							
Genotypes selected by the AGI index - 322, 355, 335, 138, 370, 368, 324, 33, 360, 233, 248, 328, 134, 113, 12, 316, 332, 371, 329, 167, 314, 39, 125, 169							
Genotypes selected by the FAI-BLUP index - 335, 355, 322, 370, 328, 33, 39, 324, 368, 316, 332, 317, 342, 326, 314, 371, 343, 125, 138, 364, 167, 360, 143, 134							

[#]FL = days to flowering (days); MT = days to maturity (days); FP = seed-filling period (days); HD = hypocotyl diameter (mm); LA = lodging angle (degrees); SW = 100-seed weight (g); SYPL = average seed yield (g/plant); and SY = average seed yield (g/plot). ^{##}12% selection intensity (24 genotypes selected) [†]GCV = genetic coefficient of variation (%). [‡]Traits excluded in SH-BLUP due to multicollinearity problems. [§]Proportion of gain predicted for traits whose values are desired to be increased (FP, HD, LA, SW, SYPL, and SY) in relation to the gain predicted through direct selection. [¶]Proportion of gain predicted for traits whose values are desired to be reduced (FL and MT) in relation to the gain predicted through direct selection.

Figure 1 shows the ranking of the 204 genotypes according to the FAI-BLUP index and its associated spatial probability (color gradient). The results allowed for a simpler, easier, and more objective genotype selection process. The first 24 progenies selected according to the FAI-BLUP index present potential to generate lines of plants that are simultaneously erect, early, and high-yielding.

The AIC is an estimator of the relative quality of statistical models for a given set of data. In a given set of candidate models for the data, the preferred model is that with the minimum AIC value. Thus, AIC rewards goodness of fit but also includes a penalty that is an increasing function of the number of estimated parameters (parsimony criterion). The penalty discourages overfitting because increasing the number of parameters in the model almost always improves the goodness of fit (Akaike, 1974).

The significance of among-progeny within-population effects and among-population effects indicated genetic variability and the possibility of obtaining gains with selection. The results found in this study reveal that the model fitted by including population effects is the most suitable for all evaluated traits (Table 1).

In soybean breeding, the use of population effect to estimate genetic and phenotypic components is often neglected. However, in forest breeding (Furtini, Ramalho, Abad, & Aguiar, 2012; Cappa et al., 2013) and animal breeding (Daetwyler, Kemper, Van Der Werf, & Hayes, 2012; Li et al., 2016), this information improves the accuracy of estimates and consequently contributes to the selection of superior genotypes. Duarte and Vencovsky (2001) and Pereira et al. (2017) pointed out that the inclusion of population information can change the ranking of progenies even under data balancing and orthogonality conditions and thus alter the genetic and phenotypic components.

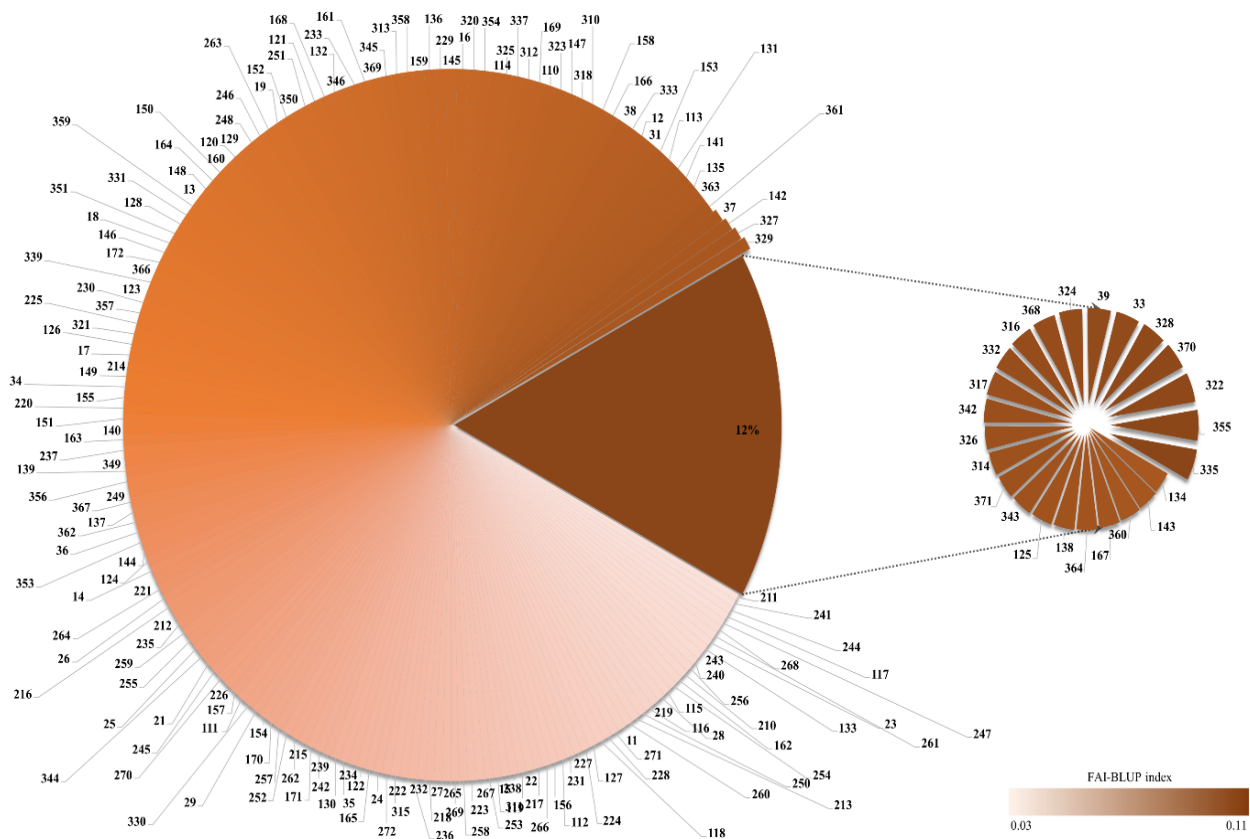


Figure 1. Ranking of genotypes and genotypes selected using the FAI-BLUP index.

Discussion

Population effects and genetic variability

Soybean ideotype

At present, in soybean breeding, the ideal plant would be erect and have a short production cycle and high grain yields, among other characteristics (Silva, Borém, Sediya, & Ludke, 2017). On this basis, selecting genotypes with desirable earliness- and yield-related traits in different environments under biotic and abiotic stress conditions will be fundamental for obtaining future gains in soybean breeding programs (Kyei-Boahen & Zhang, 2006; Abrahão & Costa, 2018). This requirement allows farmers from different regions of the world to maximize their growing areas and accelerate the soybean cycle. As a result, a second harvest (with other crops) may be implemented and production losses may be minimized due to the shorter time of exposure to stress factors in the field (Marcos-Filho, Chamma, Casagrande, Marcos, & Regitano-d'Arce, 1994; Diniz et al., 2013).

In the last few years, soybean breeding programs have reported an increase in yields (Van Roekel, Purcell, & Salmerón, 2015), which have depended on the complex understanding of the genotype \times environment interaction and increases in selection accuracy (Kang & Gauch, 1996; Gauch, 2013; Van Eeuwijk, Bustos-Korts, & Malosetti, 2016). According to Van Roekel et al. (2015), the low increases in soybean yield obtained from selection using the FP and SW traits can be explained by the high complexity and elevated number of alleles that contribute to these traits; the low increased in yields can also be explained by the evaluation of these traits under nonideal growing conditions, which result in a high genotype \times environmental interaction and low heritability values. Therefore, new methodologies to improve the strategy of selecting the best genotypes should be investigated to minimize these effects and enable the selection of genotypes that meet production demands.

Panthee, Pantalone, Saxton, West, and Sams (2007) stated that the quantitative traits of seed-filling period and lodging are related to the seed yield, which has been corroborated by our results from the AGI and FAI-BLUP indices. Lodging tolerance is an important trait for high yields and combine-harvesting effectiveness in soybean (Yamaguchi et al., 2014). Numerous studies have investigated the effect of lodging

on yield (Cooper, 1971; Mancuso et al., 1991; Yamaguchi et al., 2014). In the present study, the LA variable had the highest gains in the positive direction of selection, which led to a successful recombination of the selected progenies and to a lower lodging angle in the genotypes that will be obtained; these gains occurred mainly when the AGI and FAI-BLUP indices were used.

Maximizing the yield of soybean under environmental stress (abiotic) conditions is not an easy task (Egli, Orf, & Pfeiffer, 1984; Kantolic, Peralta, & Slafer, 2013). Extending the period from flowering to the date of maturity (i.e., increasing the FP) usually allows for increasing final yields of genotypes (Cooper, 2003; Kantolic, Peralta, & Slafer, 2013). In this sense, selecting materials with shorter vegetative periods (Ve to Vn, according to Fehr & Caviness, 1977) and longer reproductive periods (R1 to R3) may provide longer seed-filling periods and consequently increase yield (Rowntree et al., 2013; 2014).

Selection of high-yielding, early soybean progenies

Several traits have been evaluated in soybean breeding, some of which have commercial relevance for the intrinsic characterization of each cultivar or line (Lersten & Carlson, 2004; Silva et al., 2017). The morphological characterization of the soy plant is of paramount importance in tests of adaptability and stability of the crop (Lin & Binns 1994; Chaves et al., 2017). Therefore, the selection of different soybean genotypes is based on a simultaneous evaluation of many traits of interest in different biotic and abiotic stress conditions. Effective methodologies are thus necessary for the selection of superior genotypes; in this regard, selection indices are a viable alternative.

In the SH-BLUP and AGI indices, the genetic coefficient of variation was considered an economic weight. Thus, the greatest weight was assigned to the LA, SY, and SYPL traits (Table 3). Higher genetic coefficients of variation for those traits are expected to result in higher gains, as predicted by the respective indices (Rocha et al., 2018); however, this was only true with AGI.

The FAI-BLUP index led to the selection of soybean progenies that, after selection and recombination, provided balanced and desirable gains for all traits; FAI-BLUP did not require assigning economic weights, unlike the SH-BLUP and AGI indices. Rocha et al. (2018) emphasized that FAI-BLUP is able to deal with several colinear traits-i.e., they need not be excluded-in addition to using those traits as auxiliary components.

Conclusion

To obtain truly superior genotypes, one must select those which contain a number of traits that provide comparatively higher yields and meet the consumer-market demands. Therefore, improving the phenotypic expression of several traits for which segregating populations present a continuous distribution depends on the environmental effect and on the presence of various genes involved in the genetic control of those traits. In this scenario, the FAI-BLUP index optimized genetic gains by more effectively ranking the soybean progenies that are earlier, more erect, and have a higher grain yield potential. As such, the FAI-BLUP index contributes to increasing the success in soybean breeding programs.

Acknowledgements

The authors appreciate and acknowledge the financial support from the Brazilian Government offered by the National Council for Scientific and Technological Development (CNPq) and by the Coordination for the Improvement of Higher Education Personnel (CAPES). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

- Abrahão, G. M., & Costa, M. H. (2018). Evolution of rain and photoperiod limitations on the soybean growing season in Brazil: The rise (and possible fall) of double-cropping systems. *Agricultural and Forest Meteorology*, 256-257(February), 32-45. DOI: 10.1016/j.agrformet.2018.02.031
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. DOI: 10.1109/TAC.1974.1100705
- Akhter, M., & Sneller, C. H. (1996). Yield and yield components of early maturing soybean genotypes in the mid-south. *Crop Science*, 36(4), 877-882. DOI: 10.2135/cropsci1996.0011183X0036000400010x

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. DOI: 10.18637/jss.v067.i01
- Bartlett, M.S. (1938). Methods of estimating mental factors. *Nature*, 141, 609-610. DOI: 10.1038/141246a0
- Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65(5), 770-790. DOI: 10.1177/0013164404272507
- Bernardo, R. (2003). Parental selection, number of breeding populations, and size of each population in inbred development. *Theoretical and Applied Genetics*, 107(7), 1252-1256. DOI: 10.1007/s00122-003-1375-0
- Bhering, L. L., Laviola, B. G., Salgado, C. C., Sanchez, C. F. B., Rosado, T. B., & Alves, A. A. (2012). Genetic gains in physic nut using selection indexes. *Pesquisa Agropecuaria Brasileira*, 47(3), 402-408. DOI: 10.1590/S0100-204X2012000300012
- Cappa, E. P., El-Kassaby, Y. A., Garcia, M. N., Acúna, C., Borralho, N. M., Grattapaglia, D., & Poltri, S. N. M. (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in eucalyptus globulus. *PLoS ONE*, 8(11), e81267. DOI: 10.1371/journal.pone.0081267
- Carpentieri-Pípulo, V., De Almeida, L. A., De Souza Kiihl, R. A., & Rosolem, C. A. (2000). Inheritance of long juvenile period under short day conditions for the BR80-6778 soybean (*Glycine max* (L.) Merrill) line. *Euphytica*, 112(2), 203-209. DOI: 10.1023/A:1003927817278
- Chaves, M. V. A., Silva, N. S., Silva, R. H. O., Jorge, G. L., Silveira, I. C., Medeiros, L. A., ... Hamawaki, C. D. L. (2017). Genotype x environment interaction and stability of soybean cultivars for vegetative-stage characters. *Genetics and Molecular Research*, 16(3), 1-10. DOI: 10.4238/gmr16039795
- Coan, R. (1959). A comparison of oblique and orthogonal factor solutions. *The Journal of Experimental Education*, 27(3), 151-166.
- Cooper, R. L. (2003). A delayed flowering barrier to higher soybean yields. *Field Crops Research*, 82(1), 27-35. DOI: 10.1016/S0378-4290(03)00003-0
- Cooper, R.L. (1971). Influence of soybean production practices on lodging environments and seed yield in highly product. *Agronomy Journal*, 63(3), 490-493. DOI: 10.2134/agronj1971.00021962006300030043x
- Cruz, C. D. (2013). GENES - Software para análise de dados em estatística experimental e em genética quantitativa. *Acta Scientiarum. Agronomy*, 35(3), 271-276. DOI: 10.4025/actasciagron.v35i3.21251
- Daetwyler, H. D., Kemper, K. E., Van Der Werf, J. H. J., & Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *Journal of Animal Science*, 90(10), 3375-3384. DOI: 10.2527/jas.2011-4557
- Diniz, F. O., Reis, M. S., Dias, L. A. dos S., Araújo, E. F., Sedyama, T., & Sedyama, C. A. (2013). Physiological quality of soybean seeds of cultivars submitted to harvesting delay and its association with seedling emergence in the field. *Journal of Seed Science*, 35(2), 147-152. DOI: 10.1590/S2317-15372013000200002
- Duarte, J. B., & Vencovsky, R. (2001). Estimação e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. *Scientia Agricola*, 58(1), 109-117. DOI: 10.1590/S0103-90162001000100017
- Egli, D. B., Orf, J. H., & Pfeiffer, T. W. (1984). Genotypic variation for duration of seedfill in soybean1. *Crop Science*, 24(3), 587-592. DOI: 10.2135/cropsci1984.0011183X002400030037x
- Fehr, W. R., & Caviness, C. E. (1977). *Stages of soybean development*. Ames, US: Iowa State University. (Special Report, 87). Retrieved on July 14, 2018 from <http://lib.dr.iastate.edu/specialreports/87>
- Furtini, I. V., Ramalho, M. A. P., Abad, J. I. M., & Aguiar, A. M. (2012). Effect of different progeny test strategies in the performance of eucalypt clones. *Silvae Genetica*, 61(1-6), 116-120. DOI: 10.1515/sg-2012-0014
- Gauch, H. G. (2013). A simple protocol for AMMI analysis of yield trials. *Crop Science*, 53(5), 1860-1869. DOI: 10.2135/cropsci2013.04.0241
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, 28(6), 476-490.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model published by: international biometric society stable. *Biometrics*, 31(2), 423-447.
- Kang, M. S., & Gauch, H. G. (1996). *Genotype -by- environment interaction*. Boca Raton, FL: CRC Press.

- Kantolic, A. G., Peralta, G. E., & Slafer, G. A. (2013). Seed number responses to extended photoperiod and shading during reproductive stages in indeterminate soybean. *European Journal of Agronomy*, *51*, 91-100. DOI: 10.1016/J.EJA.2013.07.006
- Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187-200. DOI: 10.1007/BF02289233
- Kyei-Boahen, S., & Zhang, L. (2006). Early-Maturing Soybean in a Wheat-Soybean Double-Crop System. *Agronomy Journal*, *98*(2), 295. DOI: 10.2134/agronj2005.0198
- Lersten, N.R., & Carlson, J.B. (2004). Vegetative morphology. In: H. R. Boerma, & J. E. Specht (Eds.), *Soybeans: Improvement, production and uses* (p. 15-57). Madison: American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.
- Li, X., Lund, M. S., Zhang, Q., Costa, C. N., Ducrocq, V., & Su, G. (2016). Improving accuracy of predicting breeding values in Brazilian Holstein population by adding data from Nordic and French Holstein populations. *Journal of Dairy Science*, *99*(6), 4574-4579. DOI: 10.3168/jds.2015-10609
- Lin, C. S., & Binns, M. R. (1994). Concepts and methods of analyzing regional trial data for cultivar and location selection. *Plant Breeding*, *12*, 271-297. DOI: 10.1002/9780470650493.ch10
- Malek, M. A., Rafii, M. Y., Shahida Sharmin Afroz, M., Nath, U. K., & Mondal, M. M. A. (2014). Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. *Scientific World Journal*, *2014*, 1-12. DOI: /10.1155/2014/968796
- Mancuso, N., & Caviness, C. E. (1991). Association of selected plant traits with lodging of four determinate soybean cultivars. *Crop Science*, *31*(4) 911-914. DOI: 10.2135/cropsci1991.0011183X003100040014x
- Marcos-Filho, J., Chamma, H. M. C. P., Casagrande, J. R. R., Marcos, E. A., & Regitano-d'arce, M. A. B. (1994). Effect of harvesting time on seed physiological quality, chemical composition and storability of soybeans. *Scientia Agricola*, *51*(2), 298-304. DOI: 10.1590/S0103-90161994000200016
- Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis*. New York, US: Wiley.
- Panthee, D. R., Pantalone, V. R., Sams, C. E., Saxton, A. M., West, D. R., & Rayford, W. E. (2004). Genomic regions governing soybean seed nitrogen accumulation. *Journal of the American Oil Chemists' Society*, *81*(1), 77-81. DOI: 10.1007/s11746-004-0860-4
- Panthee, D. R., Pantalone, V. R., Saxton, A. M., West, D. R., & Sams, C. E. (2007). Quantitative trait loci for agronomic traits in soybean. *Plant Breeding*, *126*(1), 51-57. DOI: 10.1111/j.1439-0523.2006.01305.x
- Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*(3), 545-554. DOI: 10.1093/biomet/58.3.545
- Pereira, F., Bruzi, A. T., de Matos, J. W., Rezende, B. A., Prado, L. C., & Nunes, J. A. R. (2017). Implications of the population effect in the selection of soybean progeny. *Plant Breeding*, *136*(5), 679-687. DOI: 10.1111/pbr.12512
- Piepho, H. P., & Williams, E. R. (2006). A comparison of experimental designs for selection in breeding trials with nested treatment structure. *Theoretical and Applied Genetics*, *113*(8), 1505-1513. DOI: 10.1007/s00122-006-0398-8
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, AU: R Foundation for Statistical Computing. Retrieved on July 14, 2018 from <http://www.R-project.org>.
- Resende, M. D. V. de. (2016). Software Selegen-REML / BLUP : a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology*, *16*(4), 330-339. DOI: 10.1590/1984-70332016v16n4a49
- Resende, M. D. V. d., Ramalho, M. A. P., Carneiro, P. C. S., Carneiro, J. E. S., Batista, L. G., & Gois, I. B. (2016). Selection index with parents, populations, progenies, and generations effects in autogamous plant breeding. *Crop Science*, *56*(2), 530-546. DOI: 10.2135/cropsci2015.05.0303
- Rocha, M. D. M., & Vello, N. A. (1999). Interação genótipos e locais para rendimento de grãos de linhagens de soja com diferentes ciclos de maturação. *Bragantia*, *58*(1), 69-81. DOI: 10.1590/S0006-87051999000100009
- Rocha, J. R. do A. S. de C., Machado, J. C., & Carneiro, P. C. S. (2018). Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy*, *10*(1), 52-60. DOI: 10.1111/gcbb.12443

- Rowntree, S. C., Suhre, J. J., Weidenbenner, N. H., Wilson, E. W., Davis, V. M., Naeve, S. L., ... Conley, S. P. (2013). Genetic gain x management interactions in soybean: I. Planting date. *Crop Science*, *53*(3), 804-816. DOI: 10.2135/cropsci2012.03.0157
- Rowntree, S. C., Suhre, J. J., Weidenbenner, N. H., Wilson, E. W., Davis, V. M., Naeve, S. L., ... Conley, S. P. (2014). Physiological and phenological responses of historical soybean cultivar releases to earlier planting. *Crop Science*, *54*(2), 804-816. DOI: 10.2135/cropsci2013.06.0428
- Sediyama, T., Felipe, L. S., & Borem, A. (2015). *Soja: do plantio à colheita*. Viçosa, MG: Editora UFV.
- Silva, F. L., Borém, A., Sediyama, T., & Ludke, W. H. (2017). *Soybean breeding*. *Soybean Breeding*. Gewerbestrasse, SW: Springer. DOI: 10.1007/978-3-319-57433-2
- Smith, H. F. (1936). A discriminant function for plant selection. *Annals of Eugenics*, *7*(3), 240-250. DOI: 10.1111/j.1469-1809.1936.tb02143.x
- Van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype x Environment interactions? *Crop Science*, *56*(5), 2119-2140. DOI: 10.2135/cropsci2015.06.0375
- Van Roekel, R. J., Purcell, L. C., & Salmerón, M. (2015). Physiological and management factors contributing to soybean potential yield. *Field Crops Research*, *182*, 86-97. DOI: 10.1016/j.fcr.2015.05.018
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60-62. DOI: 10.1214/aoms/1177732360
- Yamaguchi, N., Sayama, T., Yamazaki, H., Miyoshi, T., Ishimoto, M., & Funatsuki, H. (2014). Quantitative trait loci associated with lodging tolerance in soybean cultivar 'Toyoharuka.' *Breeding Science*, *64*(4), 300-308. DOI: 10.1270/jsbbs.64.300