

Combined BLUP selection indexes with parents and F_2 populations in soybean (*Glycine max*) breeding

André Ricardo Gomes Bezerra^{1*}, Marcos Deon Vilela de Resende², Felipe Lopes da Silva², Tuneo Sediya², Cosme Damião Cruz², Francisco Charles dos Santos Silva² and Amilton Ferreira da Silva³

Abstract: This study aimed to apply selection indexes in situations in which F_2 populations and their parents are evaluated simultaneously in field trials to predict population genotypic values and selecting the best populations. Fifteen F_2 soybean populations were evaluated for number of days to flowering and to maturity and grain yield. The data was analyzed using restricted maximum likelihood/best linear unbiased prediction (REML/BLUP) to obtain genotypic effects, variance components and accuracies. Four selection methods were compared: genotypic values of the F_2 populations (g_{F_2}), combined index with the mean genotypic value of parents and F_2 population (I_o) and two indexes that add (to g_{F_2} and to I_o) and consider genetic variability within populations (I_b and I_c). The index I_o presented a result similar to that obtained with the g_{F_2} selection method. Selection indexes (I_b and I_c) based on several sources of information were more efficient than selection based on g_{F_2} values only.

Keywords: Grain yield, selection gain, early selection, days to maturity

INTRODUCTION

Soybean breeding programs obtain many segregating populations annually, mainly from converging crosses. Due to the demand for resources and time, breeders should focus on the significantly superior populations, because these have a higher probability of generating lines superior to the parents or a certain standard, which is usually the best commercial cultivar. However, the efficiency of breeding programs has been low, as parental crosses that do not result in new cultivars consume more than 99% of the resources (Witcombe et al. 2013).

In the breeding of autogamous plants, such as soybeans, selection has been performed from the evaluation of $F_{2,4}$ progenies or more advanced generations (Ramalho et al. 2012). In addition, the selection procedures have taken as reference only the last generation in which the progenies were evaluated. According to Resende et al. (2015), the selection could be made using information from previous generations (Wricke and Weber 1986, Resende 2002, Resende et al. 2016).

Resende (2002) proposed the use of mixed model methods, through the best linear unbiased prediction (BLUP) (Henderson 1975) in the sequential

Crop Breeding and Applied Biotechnology
23(4): e45412346, 2023
Brazilian Society of Plant Breeding.
Printed in Brazil
<http://dx.doi.org/10.1590/1984-70332023v23n4a41>



*Corresponding author:

E-mail: bezerra.agro@yahoo.com.br

 ORCID: 0000-0003-0029-4179

Received: 01 May 2023

Accepted: 31 July 2023

Published: 10 October 2023

¹ Limagrains Brasil S.A., Rua Rio Verde, 56, Dom Benedito, 75800-115, Jataí, GO, Brazil

² Universidade Federal de Viçosa, Avenida PH Rolfs, s/n, Campus Universitário, 36570-900, Viçosa, MG, Brazil

³ Universidade Federal de São João del-Rei, MG 424, km 47, 35701-970, Sete Lagoas, MG, Brazil

analysis of successive generations, according to the genealogical method. Later, a multigeneration index that considers all information across generations and covariance between generations, aiming to identify the best progenies, was proposed (Resende et al. 2015). This multigeneration index was successfully applied to common bean breeding by Batista et al. (2017). Another method called selection index with parents, populations, progenies and effects of generations (SIPPPG) was recommended for selection in the breeding of autogamous species (Resende et al. 2016). This index includes not only the effects of progenies in different generations, but also the effects of populations in all generations, parental data and F_1 and F_2 generations all simultaneously.

There are no reports in the literature of a specific index for selecting F_2 populations. Previous studies were carried out focusing only on population's data without consider the parents information. In this context, this study applied new selection methods that can be used in situations in which F_2 populations and their parents are simultaneously evaluated in an experiment, aiming to estimate the genetic gain by selecting the best populations for earliness and grain yield in soybean.

MATERIAL AND METHODS

Obtaining phenotypic data

Fifteen F_2 populations from biparental crosses, obtained from balanced complete diallel cross between six parents (Table 1), was performed in the soybean breeding program of the Universidade Federal de Viçosa (UFV), at Viçosa, Minas Gerais, Brazil, in the 2013/2014 crop season. In the following year (2014/2015), the progenies were assessed at the Teaching, Research, and Academic-Extension Units at UFV (lat 20° 46' 08" S, long 42° 52' 14" W, alt 663 m asl). The experiment was laid out in a randomized complete block design with three replications. Populations and parents were evaluated in plots of three 6.0 m rows, spaced 0.7 m apart. In each plot 140 seeds were distributed, resulting in a sowing density of approximately 8 seeds per linear meter. All plant management operations were undertaken in accordance with the requirements of the crop in the region (Sediyama et al. 2022).

The traits evaluated at individual level were grain yield (g plant^{-1}), number of days to flowering and days to maturity. Days to flowering is the period between the emergence of the seedling and the appearance of the first flower on the main stem; days to maturity is the number of days between the emergence of the seedling and full maturity (plants with 95% of pods that have reached the ripe pod color). For the number of days to flowering and maturity, selection was carried out in both directions, that is, for selection of early and late plants.

Statistical analyses

Data were analyzed using restricted maximum likelihood/best linear unbiased prediction (REML/BLUP). REML developed by Patterson and Thompson (1971) was used to estimate variance components and BLUP (Henderson 1975) was used to predict genotypic values. The analysis strategy adjusted simultaneously one model for crosses and another for parents (that were included in the experiments as controls) in the same analysis (Resende et al. 2014, Resende and Alves 2020). In this case, a special encoding in the column "Fixed Effect", separating crosses and parents, became necessary and was possible using Selegen-REML/BLUP software (Resende 2016), model 189. Thus, we have the implicit model for data of crosses:

Table 1. Description of segregating soybean populations, with their respective parents and number of F_2 individuals evaluated for grain yield, days to flowering and days to maturity

Population	Parents	Number of F_2 individuals
1	MSOY6001 x RSF6563IPRO	236
2	MSOY6001 x TMG123RR	265
3	MSOY6001 X SYN9078RR	197
4	MSOY6001 X TMG801	211
5	MSOY6001 X MSOY9144RR	262
6	RSF6563IPRO X TMG123RR	187
7	RSF6563IPRO X SYN9078RR	235
8	RSF6563IPRO X TMG801	226
9	RSF6563IPRO X MSOY9144RR	211
10	TMG123RR X SYN9078RR	209
11	TMG123RR X TMG801	240
12	TMG123RR X MSOY9144RR	221
13	SYN9078RR x TMG801	213
14	SYN9078RR X MSOY9144RR	188
15	TMG801 X MSOY144 RR	222
Total		3323

$$y = Xc + Zm + Wf + Sd + Tp + e \tag{1}$$

where y is the vector of phenotypic data; c is the fixed effect of the combination of the average of the cross in each block; m is the vector of the effects of general combining abilities of the male parents (assumed as random), in which $m \sim N(0, I\sigma_m^2)$; f is the vector of the effects of the general combining abilities of the female parents (assumed as random), in which $f \sim N(0, I\sigma_f^2)$; d is the vector of the effects of the specific combining abilities of the crosses (assumed as random), in which $d \sim N(0, I\sigma_d^2)$; p is the vector of the effects of plots (assumed as random), in which $p \sim N(0, I\sigma_p^2)$; e is the vector of the effects of the error (assumed as 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 c, m, f, d and p effects, respectively. The total genotypic effect of each F₂ population was given by:

$$\hat{g}_{F2ij} = \hat{m}_i + \hat{f}_j + \hat{d}_{ij} \tag{2}$$

Implicit model for parents' data:

$$y = Xu + Hg + Tp + e \tag{3}$$

where y is the vector of phenotypic data; u is the fixed effect of the overall mean of the parents in each block; g is the vector of the additive genetic effects of the parents (assumed as random), in which $g \sim N(0, I\sigma_g^2)$; p is the vector of the effects of plots (assumed as random), in which $p \sim N(0, I\sigma_p^2)$; e is the vector of the effects of error, in which $e \sim N(0, I\sigma_e^2)$. The capital letters (X, H and T) represent the incidence matrices for u, g and p effects, respectively.

The BLUP index I_a , which considers the effects of populations and parents, has the following form (Resende 2015):

$$I_a = \hat{b}_1 \left[\frac{(g_1 + g_2)}{2} \right] + \hat{b}_2 g_{F2} \tag{4}$$

In which g_1 and g_2 refer to the predicted additive genetic values of the parents 1 and 2, respectively; g_{F2} is equivalent to the predicted additive genetic value of the F₂ population. The weights are given by: $\hat{b} = P^{-1}C$, where:

$$P = \begin{bmatrix} r_{aaPar}^2 & r_{aaPar}^2 r_{aaF2}^2 & r_{aaPar}^2 r_{ParF2}^2 \\ r_{aaF2}^2 & r_{aaF2}^2 & r_{aaF2}^2 r_{ParF2}^2 \\ r_{aaF2}^2 r_{aaPar}^2 & r_{aaF2}^2 r_{aaF2}^2 & r_{aaF2}^2 r_{ParF2}^2 \end{bmatrix} \tag{5}$$

In which r_{aaPar}^2 is the reliability of the parents information and r_{aaF2}^2 is the reliability of the F₂ populations information. However, under a completely additive genetic model, the correlation (r_{ParF2}^2) between the standardized predicted F₂ population genotypic values and the standardized predicted mean of parents genotypic values is equal to 1 (F₂ population and mean of parents seeks for the same quantity, i.e., the parametric genotypic mean of a population coming from crossing of the pair of parents), then we have $r_{ParF2}^2 = 1$, and consequently:

$$P = \begin{bmatrix} r_{aaPar}^2 & r_{aaPar}^2 r_{aaF2}^2 \\ r_{aaF2}^2 & r_{aaF2}^2 \end{bmatrix} \tag{6}$$

$$\text{and } C = \begin{bmatrix} r_{aaPar}^2 & r_{aaPar}^2 r_{ParF2}^2 \\ r_{aaF2}^2 & 1 \end{bmatrix} = \begin{bmatrix} r_{aaPar}^2 \\ r_{aaF2}^2 \end{bmatrix} \tag{7}$$

The weights of the index are:

$$\hat{b}_1 = \frac{(1 - r_{aaF2}^2)}{1 - r_{aaPar}^2 r_{aaF2}^2} \tag{8}$$

$$\hat{b}_2 = \frac{(1 - r_{aaPar}^2)}{1 - r_{aaPar}^2 r_{aaF2}^2} \tag{9}$$

The accuracy of the index is given by $r_{aaI}^2 = \sqrt{1 - \frac{(1 - r_{aaPar}^2)(1 - r_{aaF2}^2)}{1 - r_{aaPar}^2 r_{aaF2}^2}}$.

where, according to Resende (2017), $r_{aaParents}^2 = (1/2)(r_{aaMother}^2 + r_{aaFather}^2)$ is the square of accuracy of the genetic value predicted from the mean of the two parents, based on the performance of the parents in the experiment; $r_{aaMother}^2$: square of accuracy of the genetic value predicted from the parent used as a mother; $r_{aaFather}^2$: square of accuracy of the genetic value predicted from the parent used as a father; r_{aaF2}^2 : square of the accuracy of the genetic value predicted from the F₂ population, based on data from the populations in the experiment. This accuracy is given by: $r_{aaF2}^2 = 2F_{ST}^2 \sigma_a^2 / V_{phenot/F2}$ in which F_{ST}^2 : square of the inbreeding coefficient due to the differentiation between populations; σ_a^2 : additive genetic variance; and $V_{phenot/F2}$: total phenotypic variance; and $F_{ST} = \text{Var}_{g(F2)} / \text{Var}_{g(Par)}$ where $\text{Var}_{g(F2)}$ and $\text{Var}_{g(Par)}$ are the genotypic variances among F₂ and among parents, respectively. The definitions of the squared accuracy presented are valid at generation F ∞ .

To obtain the accuracies $r_{\hat{a}aPar}^2$ and $r_{\hat{a}aF_2}^2$ we divided the accuracies from the output file with extension ‘fam’ of the Selegen-REML/BLUP software (Resende 2016) by 4 and $1/F_{ST}$, respectively. The second index (I_b) for population selection, which considers the additive genetic value of the population and genetic variability within the population, as described by Resende (2015, 2017) is:

$$I_b = F_{2pop_i} + 3.09 \sqrt{(1 - F_{ST})2\sigma_{a0}^2 \frac{Var_{(Residual_i)}}{Var_{(Residual)}}} \quad (10)$$

where F_{2pop_i} is the BLUP of the population i effect, and the expression within the root refers to the magnitude of genetic variability within population i; σ_{a0}^2 represents the original additive genetic variance of parent’s population and $Var_{(Residual_i)}$ is the phenotypic variability within population i and $Var_{(Residual)}$ corresponds to the residual variance of the experiment considering the population data. Finally, the value of the index I_b is an estimate of the genetic value of the best line to be selected at the end of the selection process and the constant 3.09 refers to the number of standard deviations in the normal distribution curve, associated with the selection of one in thousand lines. The sum of three standard deviations represents 99.7% of the data of a normal distribution. The accuracy of the index corresponds to the estimation accuracy of the genetic effects of the F_2 population, as discussed earlier.

The third index (I_c) was presented by Resende (2015) how an improved index and the same purpose of I_b , it uses the combined selection of F_2 parents and variability among individuals within the F_2 population. The index can be obtained as follows:

$$I_c = \hat{b}_1 \left[\frac{(g_1 + g_2)}{2} \right] + \hat{b}_2 g_{F_2} + 3.09 \sqrt{(1 - F_{ST})2\sigma_{a0}^2 \frac{Var_{(Residual_i)}}{Var_{(Residual)}}} \quad (11)$$

The definitions of the terms are the same as presented above and the accuracy of the index is equal the accuracy of the index I_a . Statistical analyses were processed using the Selegen-REML/BLUP software (Resende 2016). Spearman’s rank correlation coefficients were calculated among the ranking given by the F_2 genotypic values, I_a , I_b and I_c . The analyses were performed in R software (R Core Team 2023) and the figures created with package ggplot2 (Wickham 2016).

RESULTS AND DISCUSSION

The predicted genotypic values, phenotypic variances within population and selection for grain yield are presented in Table 2. According to the genotypic value, the best populations were 8, 6, 7, 9, 13, 11, 10, 1 and 15, because they showed values above the overall mean of the experiment (57.32 g). The phenotypic variance within the population ranged from 808.76 (population 2) to 1896.78 (population 8).

Table 2. Predicted genotypic value (Gv), phenotypic variance within population ($Var_{(Res)}$), accuracy ($r_{\hat{a}a}$) and selection indexes applied to fifteen F_2 soybean populations for grain yield

F_2 average			Index I_a with Parents and F_2		Index I_b with F_2 and variance of F_2		Index I_c with Parents, F_2 and variance of F_2	
Pop	Gv	$Var_{(Res)}$	Pop	I_a	Pop	I_b	Pop	I_c
8	64.26	1896.78	8	68.30	8	75.54	8	79.59
6	62.45	1250.04	6	64.85	6	71.61	6	74.01
7	60.57	868.01	7	63.41	9	68.62	7	71.04
9	59.24	1311.37	9	60.59	7	68.21	9	69.97
13	59.04	1085.99	13	60.55	15	68.09	13	69.09
11	58.25	1336.56	11	59.36	11	67.72	11	68.84
10	58.19	1074.60	1	58.06	13	67.58	15	68.11
1	57.96	1145.63	10	58.02	4	67.05	1	66.84
15	57.88	1551.35	15	57.90	1	66.73	10	66.52
4	56.31	1716.06	4	55.09	10	66.69	4	65.83
12	55.51	1249.42	12	53.88	12	64.67	12	63.04
14	54.92	1097.43	14	53.71	3	63.59	14	62.30
3	54.76	1161.45	3	52.28	14	63.51	3	61.11
2	50.39	808.76	2	47.62	5	58.50	2	54.99
5	50.12	1046.14	5	46.25	2	57.76	5	54.64
$r_{\hat{a}a}$	0.54			0.55		0.54		0.55

The index based on the additive genetic value of parents and F₂ populations (I_o) showed a result similar to that obtained with the selection on the genotypic value of population. In this case, populations 11, 10 and 1 are considered in this order of superiority, while by index I_o , population 10 reversed position with population 1. In addition, the accuracy of the index I_o (0.55) was higher than the accuracy of the selection based on the genotypic value of the F₂ population (0.54), that is, the use of the index provided 1% more genetic gain in grain yield, due to the inclusion of information from the parents. In terms of magnitudes, these accuracies are classified as moderate (Resende and Duarte 2007, Resende and Alves 2022).

For grain yield the selection by genotypic values and index I_o has provided genetic gain similar to those obtained annually by breeding programs. Felipe et al. (2016) reported that the genetic gains obtained for soybean in Argentina in the period from 1980 to 2015 were on the order of 1.1%, which is also similar to the gains observed in the United States. In Brazil, Toledo et al. (1990) evaluated the efficiency of a soybean breeding program in the state of Paraná in the period from 1981 to 1986 and found that genetic gains were 1.8% in the early group and 1.3% for genotypes of the semi-early group. Recently a new study revealed an average rate of yield gain of 45.9 kg ha⁻¹ yr⁻¹ (2.1% ha⁻¹ yr⁻¹) over the past 50 years in southern Brazil (Umburanas et al. 2022).

The selection based on the index with additive values of population and phenotypic variance within population (I_b) classified the best populations similarly to the other indexes. However, through this index it is possible to obtain gains, when performing the selection, in all the studied populations, because all the values obtained were above the overall mean of the experiment. The value of the index I_b refers to an estimate of the genetic value of the best line to be selected in generation F_∞. Taking as an example the best population, in this case, the population 8, when comparing with the overall mean, we have 75.54/57.32 = 1.32 and, therefore, the genetic gain will be 32%. When comparing with the mean of population 8, we have 75.54/64.26 = 1.18, so the genetic gain with selection within population 8 will be 18%. Thus, through this index it is possible to obtain gains of up to 32% with the selection between populations and ranging from 13 to 19% by selecting within populations (Table 3).

For grain yield, the genetic gains obtained with index I_b are considered satisfactory for the genetic improvement of soybeans, given that in the literature gains between 3 and 32.8% are reported in different generations of inbreeding (Reis et al. 2004, Costa et al. 2008, Bárbaro et al. 2013). The great advantage of the index I_b is the combination of genetic information from populations and variance within populations. According to Resende (2015), the best indexes include both high mean and wide genetic variability. As a disadvantage of the index, the cost of phenotyping of all individuals to estimate variance within the population can be cited.

Table 3. Relative performance compared to the overall mean and population F₂ mean, obtained from the index I_b , which considers the additive genetic value and genetic variability within the population, for the traits grain yield, days to flowering and days to maturity

Population	Grain yield			Days to flowering			Days to maturity		
	F ₂ mean/Overall mean	I_b /Overall mean	I_b /F ₂ mean	F ₂ mean/Overall mean	I_b /Overall mean	I_b /F ₂ mean	F ₂ mean/Overall mean	I_b /Overall mean	I_b /F ₂ mean
1	1.01	1.16	1.15	0.88	0.59	0.67	0.94	0.82	0.88
2	0.88	1.01	1.15	0.92	0.71	0.77	0.93	0.79	0.85
3	0.96	1.11	1.16	0.99	0.53	0.54	1.01	0.80	0.80
4	0.98	1.17	1.19	1.01	0.53	0.52	1.01	0.69	0.69
5	0.87	1.02	1.17	1.01	0.52	0.52	1.00	0.69	0.69
6	1.09	1.25	1.15	0.92	0.57	0.62	0.99	0.81	0.82
7	1.06	1.19	1.13	0.87	0.64	0.73	0.94	0.82	0.87
8	1.12	1.32	1.18	0.96	0.52	0.54	1.00	0.68	0.69
9	1.03	1.20	1.16	0.97	0.45	0.47	1.00	0.72	0.72
10	1.02	1.16	1.15	1.03	0.62	0.60	0.99	0.72	0.73
11	1.02	1.18	1.16	1.03	0.55	0.53	0.99	0.70	0.70
12	0.97	1.13	1.17	1.02	0.54	0.53	1.00	0.71	0.71
13	1.03	1.18	1.14	1.09	0.64	0.58	1.04	0.72	0.70
14	0.96	1.11	1.16	1.08	0.52	0.48	1.03	0.76	0.73
15	1.01	1.19	1.18	1.22	0.98	0.81	1.14	1.00	0.88

The index I_c is a projection of I_a in the normal distribution curve, in which the index is the genetic value of the best line to be selected in generation F_∞ , and the interpretation is the same made for I_b . However, the information from individuals within populations included in I_c promoted minor changes in the order of the best populations, in relation to I_a , as can be observed for populations 1, 10 and 15 (Table 2).

For days to flowering, the best populations classified, based on genotypic value, in descending order, were 7, 1, 6, 2, 8, 9 and 3. The other populations showed means higher than the overall mean of the experiment (51.16 days), so the probability of obtaining gains when selecting these populations is lower. The variance within the population for the trait in question ranged from 12.22 to 85.06 (Table 4). The ranking made by the index with parents and F_2 populations (I_b) changed the order of the two best populations (1 and 7) and raised population 10 to three positions above the classification made by genotypic values. The populations that, notably, can provide gains in reducing the time to flowering are the same identified by genotypic population values. Through the application of the index, the possibility of a reduction from 2 to 65% in the time to flowering was found (Table 3). In addition, the use of the index I_a can provide 6% more genetic gains compared to selection based on genotypic values, due to its greater selective accuracy (0.38).

The inclusion of variances between plants in the population, in the index I_b , promoted considerable change in the ranking of populations (Table 4). Populations 1 and 7, for example, which occupied the first two positions in the order based on genotypic values and index with parents and F_2 (I_a), were reallocated in the 10th and 13th positions by including the variance of F_2 in the index (I_b). Due to the low variance between plants in populations 1 (23.01) and 7 (14.71), these were penalized by the index. The same projection of gains made previously for the grain yield from the I_b , can be applied to the number of days to flowering. The results showed that it is possible to obtain gains by performing selection in all populations, contrary to what was observed when the selection was performed from the genotypic value of the population and by the index I_a .

As verified for grain yield, the index I_b was also the one that had the highest selective accuracy for the number of days to flowering. In Table 3 we can see that the relative performance of the index I_b was higher compared to the overall mean of the experiment and compared to the mean of the F_2 population, that is, in the selection between and within populations, respectively.

The index I_c was the one that showed the greatest expectation of reducing the time to flowering, because it had a higher probability of selection of plants with greater potential to originate lines in F_∞ with less time to flowering and,

Table 4. Predicted genotypic value (Gv), phenotypic variance within population ($Var_{(Res)}$), accuracy (r_{dal}) and selection indexes applied in fifteen F_2 soybean populations for days to flowering

F ₂ average			Index I _a with Parents and F ₂		Index I _b with F ₂ and variance of F ₂		Index I _c with Parents, F ₂ and variance of F ₂	
Pop	Gv	Var _(Res)	Pop	I _a	Pop	I _b	Pop	I _c
7	44.58	14.71	1	42.13	9	23.09	9	22.71
1	45.07	23.01	7	43.03	14	26.66	8	26.34
6	47.20	33.11	6	44.74	8	26.70	3	26.66
2	47.31	12.22	2	46.13	5	26.73	6	26.90
8	49.15	52.45	8	48.79	4	26.89	1	27.26
9	49.39	71.98	9	49.01	3	27.21	5	27.51
3	50.80	57.93	3	50.26	12	27.82	4	27.71
4	51.67	63.87	10	52.46	11	27.98	14	28.67
5	51.85	65.71	4	52.48	6	29.36	12	29.17
12	52.34	62.59	5	52.64	1	30.20	11	29.33
10	52.50	45.03	12	53.69	10	31.70	7	31.15
11	52.51	62.61	11	53.86	13	32.52	10	31.65
14	55.25	85.07	14	57.26	7	32.69	13	34.52
13	55.69	55.86	13	57.69	2	36.48	2	35.30
15	62.22	15.31	15	65.31	15	50.09	15	53.18
(r_{dal})	0.32			0.38		0.32		0.38

at the same time, it had selective accuracy superior to the genotypic value of population and index I_b , and equivalent to the index I_a .

The genotypic values, variances of each population and indexes obtained for number of days to maturity are presented in Table 5. It was found that the ordering of the populations most favorable to reduction of the characteristic was similar when the selection based on genotypic value of population and index with Parents and F₂ (I_a) was used. Regarding the index, the results showed that, although the change in the ranking of the best populations was modest, compared to genotypic values, the inclusion of information from the parents contributed to obtaining expected gains with the selection (Table 3). Considering the reduction of the number of days to maturity, through the application of the index I_a , it was possible to obtain genetic gains of up to 32% compared to the overall mean of the experiment and 31% compared to the population mean.

The ordering and genetic gain with the selection of the best populations based on the index I_c were similar to those obtained with the index I_b (Table 5 and Figure 1) and much higher than those obtained with the selection by genotypic

Table 5. Predicted genotypic value (Gv), residual variance ($Var_{(Res)}$), accuracy ($r_{\hat{a}ai}$) and selection indexes applied in fifteen F₂ soybean populations for days to maturity

F ₂ average			Index I_a with Parents and F ₂		Index I_b with F ₂ and variance of F ₂		Index I_c with Parents, F ₂ and variance of F ₂	
Pop	Gv	$Var_{(Res)}$	Pop	I_a	Pop	I_b	Pop	I_c
2	121.09	24.77	2	119.23	8	88.95	8	89.91
1	122.23	18.19	1	119.85	5	90.01	5	91.20
7	122.43	18.97	7	121.66	4	90.36	4	91.43
6	128.43	40.03	6	126.13	11	90.50	11	91.95
10	128.83	96.02	10	128.30	12	92.16	10	93.34
11	129.21	117.75	3	129.88	10	93.88	12	93.64
8	129.46	128.93	8	130.41	9	93.99	9	94.98
5	129.79	124.31	11	130.66	13	94.19	13	96.72
9	130.44	104.45	5	130.97	14	98.37	14	100.99
12	130.46	115.24	9	131.43	2	103.33	2	101.48
4	130.80	128.52	4	131.87	3	104.20	3	103.27
3	130.81	55.66	12	131.93	6	105.86	6	103.56
14	134.42	102.13	14	137.04	7	106.90	1	104.64
13	134.89	130.19	13	137.42	1	107.02	7	106.13
15	147.62	22.78	15	151.73	15	130.59	15	134.70
$r_{\hat{a}ai}$	0.30			0.35		0.30		0.35

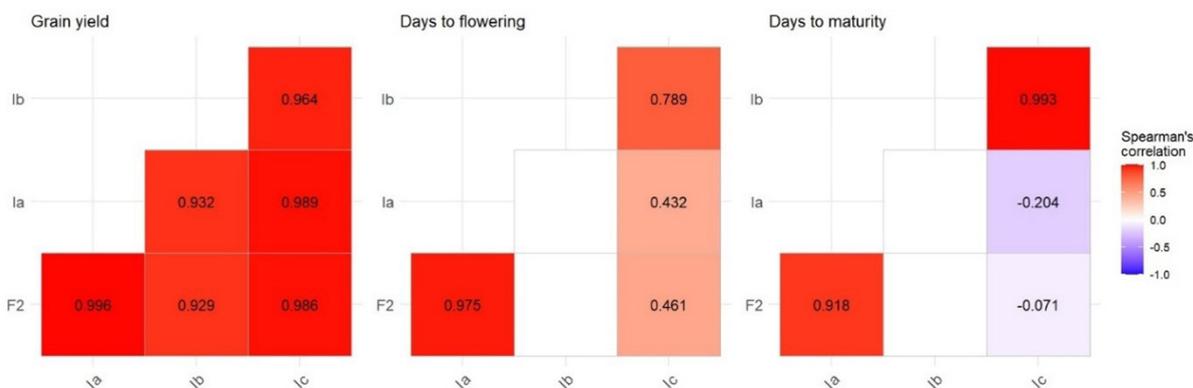


Figure 1. Spearman's rank correlations among ranking by F₂ genotypic values and selection indexes (I_a , I_b and I_c) applied in fifteen F₂ soybean populations for days to flowering, days to maturity and grain yield. * Only significant values ($p < 0.05$) are presented.

value and index I_{σ} . The indexes that contemplate the variance between individuals within populations (I_b and I_c) provided greater gains compared to the selection based on the genotypic value of the population and the index that includes information of parents and F_2 (I_{σ}).

Significant rank correlations were found between F_2 genotypic values and indexes, especially for grain yield (Figure 1). There was no correlation among index I_b and F_2 genotypic values and among I_b and I_{σ} for days to flowering and days to maturity ($p > 0.05$). For all studied traits, the index I_{σ} was highly correlated with F_2 genotypic values (0.918 to 0.996) and I_c with I_b . The weak correlations between I_c and F_2 genotypic values and between I_c and I_{σ} for days to maturity confirm that the parents information has changed the population's raking. For traits with a greater contribution of additive effect, the F_2 average reflects the parental average, especially for traits controlled by few genes. In this way, indexes containing only information from parents and F_2 are sufficient for selecting individuals.

The traits number of days to flowering and days to maturity are controlled by several genes. To date, ten genes involved in the genetic control of these traits have been described (Zhao et al. 2016). So, from divergent crosses it is possible to obtain various allelic combinations, which result in different phenotypes in F_2 . In addition, the occurrence of transgressive segregants for the traits in question is common, especially when the crosses involve divergent parents. The presence of transgressive individuals may increase the variance within the population, as verified in the populations 4, 5, 8, 9, 10, 11, 12, 13 and 14 originated from the cross of contrasting parents as to the number of days to maturity. Transgressive segregants are commonly observed in soybeans (Carpentieri-Pípolo et al. 2000, Tasma et al. 2001, Carpentieri-Pípolo et al. 2002). The occurrence of transgressive segregation is attributed to dispersion of favorable alleles between parents (Mackay et al. 2021).

We observed in this work that parent's information was more important than the information of the means of F_2 populations, depending on the F_{ST} . The F_{ST} is the proportion of total variability that is distributed among populations and greater efficiency is expected in F_2 when there is large total variance, the greater the efficiency of F_2 . According to Resende (2015) combining the two sources of information in an index is the optimal procedure (via BLUP) to obtain high selection accuracy. For grain yield, Table 6 shows that the accuracy of the index I_{σ} (0.55) was higher than the accuracy of F_2 (0.54) and of the parents (0.18). According to Resende (2015), these results are valid under model with completely additive inheritance.

The great advantage of the index I_{σ} compared to the other methods presented is its ease of obtaining, because it requires only the genetic value of parents and populations. Such an index dispenses with the evaluation at individual level, which makes the breeding program slow and costly. Besides, the index provided 1%, 6% and 5% more genetic gain, compared to selection by population genetic value, for the characteristics grain yield, number of days to flowering and number of days to maturity, respectively.

Indexes based on BLUP joint analyses for parents, populations and individuals in a single experiment as reported in this work provide all necessary information to obtain the BLUP of the indexes combining parents and population BLUPs. This is the most precise and efficient approach to obtain an index, which takes into account the reliability of each information as well as the correlation between them. This is also in line with the standard procedure of obtaining a selection index based on multiple traits via multivariate BLUP. The referred joint model is also crucial in computing the reliabilities of the different indexes. A BLUP index combining predicted (BLUP) genetic values of several information sources in this way is as exact as a multivariate BLUP (Resende 2015, Resende et al. 2016). The resulting BLUP is produced as an estimable function of several BLUPs.

In summary, selection using indexes based on various sources of information was more efficient than selection based on genotypic population values and the inclusion of parental information and variance within populations increased the expectation of gains for all the traits studied. The index I_{σ} , which includes the genotypic values of parents and F_2 populations, was the best strategy to increase the gains with selection.

Table 6. Inbreeding coefficient between populations (F_{ST}), accuracies of parents ($r^2_{\sigma\sigma\text{Parents}}$), F_2 ($r^2_{\sigma\sigma F_2}$) and selection index with parents and F_2 ($r^2_{\sigma\sigma I_{\sigma}}$) for grain yield, days to flowering and maturity

Trait	$r^2_{\sigma\sigma\text{Parents}}$	$r^2_{\sigma\sigma F_2}$	F_{ST}	$r^2_{\sigma\sigma I_{\sigma}}$
Grain yield	0.18	0.54	0.75	0.55
Days to flowering	0.23	0.32	0.34	0.38
Days to maturity	0.21	0.30	0.31	0.35

ACKNOWLEDGEMENTS

We wish to thank Marcos Deon Vilela de Resende for implementing the routines free of charge for the entire scientific community through SELEGEN-REML\BLUP software.

REFERENCES

- Bárbaro IM, Centurion MAPC, Di Mauro AO, Unêda-Trevisoli SH and Costa MM (2007) Comparação de estratégias de seleção no melhoramento de populações F₃ de soja. **Revista Ceres** **54**: 251-262.
- Batista LG, Anjos RSR, Poersch NL, Nalin R, Carneiro PCS, Carneiro JES and Resende MDV (2017) Multigeneration index in the selection of common bean inbred families. **Crop Science** **57**: 2354-2360.
- Carpentieri-Pípulo V, Almeida LA and Kiihl RAS (2002) Inheritance of a long juvenile period under short-day conditions in soybean. **Genetics and Molecular Biology** **25**: 463-469.
- Carpentieri-Pípulo V, Almeida LA, Kiihl RAS and Rosolem CA (2000) Inheritance of long juvenile period under short day conditions for the BR80-6778 soybean (*Glycine max* (L.) Merrill) line. **Euphytica** **112**: 203-209.
- Costa MM, Di Mauro AO, Unêda-Trevisoli SH, Arriel NHC, Bárbaro IM, Silveira GD and Muniz FRS (2008) Analysis of direct and indirect selection and indexes in soybean segregating populations. **Crop Breeding and Applied Biotechnology** **8**: 47-55.
- Felipe M, Gerde JA and Rotundo JL (2016) Soybean genetic gain in maturity groups III to V in Argentina from 1980 to 2015. **Crop Science** **56**: 3066-3077.
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. **Biometrics** **31**: 423-449.
- Mackay IJ, Cockram J, Howell P and Powell W (2021) Understanding the classics: the unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. **Plant Biotechnology Journal** **19**: 26-34.
- Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika** **58**: 545-554.
- R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <<https://www.r-project.org/>>. Accessed on April 30, 2023.
- Ramalho MAP, Abreu AFB, Santos JB and Nunes JAR (2012) **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. UFLA, Lavras, 522p.
- Reis EF, Reis MS, Cruz CD and Sediyaama T (2004) Comparação de procedimentos de seleção para produção de grãos em populações de soja. **Ciência Rural** **34**: 685-692.
- Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes**. Embrapa Informação Tecnológica, Brasília, 975p.
- Resende MDV (2015) **Genética quantitativa e de populações**. Suprema, Viçosa, 463p.
- Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. **Crop Breeding and Applied Biotechnology** **16**: 330-339.
- Resende MDV (2017) **Selegen Reml/Blup – Sistema estatístico e seleção genética computadorizada: manual complementar do Selegen-Reml/Blup 2017**. UFV, Viçosa, 29p.
- Resende MDV and Alves RS (2020) Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant breeding. **Functional Plant Breeding Journal** **3**: 121-132
- Resende MDV and Alves RS (2022) Statistical significance, selection accuracy and experimental precision in plant breeding. **Crop Breeding and Applied Biotechnology** **22**: 1-19.
- Resende MDV and Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical** **37**: 182-194.
- Resende MDV, Ramalho MAP, Carneiro PCS, Carneiro JES, Batista LG and Gois IB (2016) Selection index with parents, populations, progenies, and generations effects in autogamous plant breeding. **Crop Science** **56**: 530-546.
- Resende MDV, Ramalho MAP, Guilherme SR and Abreu AFB (2015) Multigeneration index in the within-progenies bulk method for breeding of self-pollinated plants. **Crop Science** **55**: 1202-1211.
- Resende MDV, Silva FF and Azevedo CF (2014) **Estatística matemática, biométrica e computacional**. Suprema, Visconde do Rio Branco, 882p.
- Sediyaama T, Silva F, Borém A and Camara G (2022) **Soja: Do plantio à colheita**. Oficina de textos, São Paulo, 312p.
- Tasma IM, Lorenzen LL, Green DE and Shoemaker RC (2001) Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. **Molecular Breeding** **8**: 25-35.
- Toledo JFF, Almeida LA, Kiihl RAS and Menosso MG (1990) Ganho genético em soja no Estado do Paraná, via melhoramento. **Pesquisa Agropecuária Brasileira** **25**: 89-94.
- Umburanas RC, Kawakami J, Ainsworth EA, Favarin JL, Anderle LZ, Dourado-Neto D and Reichardt K (2022) Changes in soybean cultivars released over the past 50 years in southern Brazil. **Scientific Reports** **12**: 1-14.
- Wickham H (2016) **ggplot2: elegant graphics for data analysis**. Springer-Verlag, New York, 213p.

Witcombe JR, Gyawali S, Subedi M, Virk DS and Joshi KD (2013) Plant breeding can be made more efficient by having fewer, better crosses. **BMC Plant Biology** **13**: 1-12.

Wricke G and Weber E (1986) **Quantitative genetics and selection in plant breeding**. De Gruyter, Berlin, New York, 406p.

Zhao C, Takeshima R, Zhu J, Xu M, Sato M, Watanabe S, Kanazawa A, Liu B, Kong F, Yamada T and Abe J (2016) A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. **BMC Plant Biology** **16**: 1-15.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.