Acta Scientiarum

http://www.periodicos.uem.br/ojs/
ISSN on-line: 1807-8621
Doi: 10.4025/actasciagron.v45i1.56156

PLANT BREEDING

Methods for estimation of genetic parameters in soybeans: an alternative to adjust residual variability

Ivan Ricardo Carvalho[1*], José Antonio Gonzalez da Silva[1], Natã Balssan Moura[1], Luiz Leonardo Ferreira[2], Francine Lautenchleger[3] and Velci Queiróz de Souza[4]

[1]Departamento de Estudos Agrários, Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Avenida do Comércio, 3000, 97800-000, Ijuí, Rio Grande do Sul, Brazil. [2]Departamento de Agronomia, Centro Universitário de Mineiros, Mineiros, Goiás, Brazil. [3]Departamento de Agronomia, Universidade Estadual do Centro-Oeste, Guarapuava, Paraná, Brazil. [4]Universidade Federal do Pampa, São Gabriel, Rio Grande do Sul, Brazil. *Author for correpondence. E-mail: ivan.carvalho@unijui.edu.br

ABSTRACT. Selection practices are maximized when plant breeders have the availability of consolidated parameters, which will guide direct and indirect selection methods. This study aimed to apply a biometric alternative to minimize residual variance and maximize selection parameters by parent-progeny regression, interim controls, and mixed linear models intrinsic to breeding. The obtained data were subjected to the assumptions of the statistical model, which identified the normality and homogeneity of the residual variances and model additivity. Subsequently, two analysis scenarios were created. The first preserved all information obtained in the experiment, both from segregating families and pure-line cultivars, and was called original scenario. The other scenario preserved progeny data, but the residual variability of controls was restricted using as criterion observations contained between the interval of the first sample standard deviation. Thereby, an acceptable residue limit could be obtained. Both scenarios were submitted to three consolidated frequentist methods (genitor-progeny regression; sum of squares of augmented block design with interim controls; and mixed linear models, wherein random genetic effects are taken as weighted genetic parameters by the genealogical matrix). Restricting residual variation in parents or controls can maximize genetic parameters and genetic gains in soybean breeding. Significant heritability estimate gains were obtained in the augmented blocks with interim control approach. Mixed linear models with random genetic effects can be considered a great tool to obtain genetic parameters in experiments with a high magnitude of common and regular treatments.

Keywords: *Glycine max* (L.) Merr.; narrow-sense heritability; genitor-progeny regression; augmented blocks with interim controls; mixed linear models with random genetic effects.

Received on September 20, 2020.
Accepted on February 4, 2021.

Introduction

Soybean [*Glycine max* (L.) Merr.] is important for the Brazilian and world economy due to its diverse uses in the market, human consumption, feed manufacture, biofuels, pharmaceutical and cosmetic raw material. According to the Brazilian Institute of Geography and Statistics (IBGE), the world population has shown an upward growth in the last decades, which has also increased the demand for food. In 2050, the world population may reach 10 billion inhabitants, for which sustainability has been sought in the production chain, using highly productive and efficient areas and genotypes to maximize natural and agronomic resources (Conab, 2019; Chechi, Deuner, Forcelini, & Boller, 2020).

Since it has been hard to expand agriculture to new areas, an increased productivity is essential for soybeans (Yang et al., 2020). Currently, cultural practices available for this crop have contributed a lot. Still, the search for an increase in genetic potential is fundamental. Thus, soybean breeding programs have aimed at genetic gain in seed production, in addition to tolerance to biotic and abiotic factors (Volpato et al., 2019). Breeders have sought to select superior soybean genotypes, which could maximize qualitative and quantitative agronomic traits, what has been an arduous and costly task (Almeida, Peluzio, and Afferri, 2010; Nogueira et al., 2012).

Selection practices are maximized when plant breeders have consolidated parameters available to guide direct and indirect selection methods (Woyann et al., 2019). Among the essential parameters, heritability, additive genetic variations, and expected genetic gain in progenies stand out (Torres, Teodoro, Sagrilo, Ceccon, & Correa, 2015; Troyjack et al., 2017). However, these parameters can often be inflated by variations or fluctuations and unconsolidated effects. This is due to a residual variation imposed by lack of control of disturbing and potentially disturbing characteristics in the experimental scenario, as well as low magnitude of plants measured in experimental units that do not capitalize on minimal effects to support estimation of variance components and genetic parameters (Carvalho et al., 2017). In this context, the mathematical nature of estimation and prediction methods may be related to the success and reliability of biometric inferences (Szareski et al., 2015).

Among the models and methods available for estimates and predictions, those pertinent to frequentist statistics and based on Bayesian inference are evident (Resende, Silva, & Azevedo, 2014). In plant breeding, theoretical and practical aspects of selection are directed to frequentist methods, using linear models, among which the generalized linear models have fixed and mixed effects (Liu et al., 2019). For both biometric natures, estimates present disturbances due to residual variability expressed between experimental units composed by treatments known as pure lines, controls, or parents.

Given the prevalence of determinant events for residual variance inflation, difficulties in controlling noise involved in statistical models, and failure to meet analysis assumptions, new alternatives that minimize such effects and increase selection efficiency in breeding programs should be searched, as models will present greater explicability regardless of its mathematical or biometric nature. Due to the lack of information about this new method for obtaining genetic parameters for soybeans, this study aimed to apply a biometric alternative to minimize residual variance and maximize selection parameters obtained by parent-progeny regression, interim controls, and mixed linear models intrinsic to breeding.

Material and methods

An experiment was developed in the 2019/2020 crop season in Campos Borges, Rio Grande do Sul State, Brazil. It is located at the coordinates 28°52'31'' S latitude and 53°00'55'' W longitude. The area is characterized as a humid subtropical climate, *Cfa* type according to Köppen's classification. The local soil is classified as dark red Latosol (Oxisol). The experimental design used was augmented blocks, testing 292 F3 segregating families (common treatments) and seven cultivars (Don Mario 7.0i RR, Roos Camino RR, BMX Potência RR, NS 6700 IPRO, DM5958 RSF IPRO, TMG 7166 RR, and Don Mario 5.8i RR), which corresponded to regular treatments arranged in four repetitions. The F_3 segregating families (75% endogamous level with 25% heterozygosity) were obtained through crossings carried out in 2014/2015, F_1 generation (2015/2016), F_2 generation (2016/2017). Table 1 details the genealogical information.

Table 1. Description of the genealogy for soybean F3 segregating families.

Maternal genitor		Paternal genitor		F1 (2017/2018)	F_2 Population (2018/2019)	F_3 Family (2019/2020)
G_1	DM7.0 BMX Magna RR	G_2	Fundacep 66 RR	1	IRC 001	15
G_3	DM 5.8 BMX Apolo RR	G_2	Fundacep 66 RR	2	IRC 002	4
G_1	DM7.0 BMX Magna RR	G_4	Monasca RR	3	IRC 003	12
G_5	FPS URANO RR	G_6	Mar M4	4	IRC 004	1
G_7	ROOS Camino RR	G_8	FPS Paranapanema RR	5	IRC 005	6
G_3	DM 5.8 BMX Apolo RR	G_6	Mar M4	6	IRC 006	3
G_3	DM 5.8 BMX Apolo RR	G_6	Mar M4	7	IRC 007	2
G_3	DM 5.8 BMX Apolo RR	G_5	FPS Urano RR	8	IRC 008	7
G_3	DM 5.8 BMX Apolo RR	G_8	FPS Paranapanema RR	9	IRC 010	8
G_{10}	BMX Potência RR	G_6	Mar M4	10	IRC 011	15
G_{11}	FPS Netuno RR	G_3	DM 5.8 BMX Apolo RR	11	IRC 012	18
G_1	DM 7.0 BMX Magna RR	G_3	DM 5.8 BMX Apolo RR	12	IRC 013	16
G_7	ROOS Camino RR	G_5	FPS Urano RR	13	IRC 016	3
G_{12}	FPS Júpiter RR	G_4	Monasca RR	14	IRC 017	10
G_2	Fundacep 66 RR	G_{13}	NK 7059 Vmax RR	15	IRC 019	6
G_{16}	TMG 7062 RR	G_{17}	Rota 54 IPRO RR	16	IRC 021	4
G_{17}	6700 RR	G_{18}	6700 RR	17	IRC 022	6
G_{20}	Elite Line 1	G_{20}	NS 5909 RR	18	IRC 024	8
G_{21}	Elite Line 2	G_{21}	NS 5909 RR	19	IRC 025	6
G_{22}	Elite Line 3	G_{22}	6700 RR	20	IRC 026	5
G_6	Elite Line 4	G_6	NS 5909 RR	21	IRC 027	1

Maternal genitor		Paternal genitor		F1 (2017/2018)	F_2 Population (2018/2019)	F_3 Family (2019/2020)
G_{23}	Elite Line 5	G_{23}	TMG 7166 RR	22	IRC 028	8
G_{24}	Elite Line 6	G_{24}	NS 5958 RR	23	IRC 029	6
G_{25}	BMX Força RR	G_6	Mar M4	24	IRC 030	14
G_{26}	NS 5958 RR	G_{21}	Mar M2	25	IRC 031	4
G_1	DM 7.0 BMX Magna RR	G_{27}	NS 5909 RR	26	IRC 032	22
G_6	Mar M4	G_{27}	NS 5909 RR	27	IRC 033	1
G_{26}	NS 5958 RR	G_{17}	6700 RR	28	IRC 034	6
G_{25}	BMX Força RR	G_6	Mar M4	29	IRC 035	19
G_3	DM 5.8 BMX Apolo RR	G_{28}	TMG 7166 RR	30	IRC 036	14
G_{27}	NS 5909 RR	G_6	Mar M4	31	IRC 037	11
G_{17}	6700 RR	G_{22}	Mar M3	32	IRC 038	15
G_{28}	TMG 7166 RR	G_{27}	NS 5909 RR	33	IRC 039	4
G_{29}	BMX Turbo RR	G_{28}	TMG 7166 RR	34	IRC 040	12
Number of genotypes						292

Experimental units comprised two 5-m-long sowing lines spaced 0.50 m apart. Sowing was performed manually in a no-till system in the second half of November 2019, at a population density of 14 seeds per linear meter. A basal dressing was performed with 250 kg ha^{-1} of the N-P-K in formulation 10-20-20. Preventive control measures against weeds, insect pests, and diseases were recommended to minimize biotic effects on results. Traits of agronomic interest were measured in 10 plants selected at random, within the useful area of each experimental unit. The traits measured comprised:

First pod insertion height (FPI, cm): which was the distance between the ground level and the spot where the first pod formed was inserted.

Plant height (PH, cm): the distance between the ground level and the spot of the last pod formed at the apex of the plant.

Number of pods on the main stem (NPMS, units): the counting of viable pods located on the main stem of the plant.

Number of pods on the branches (NPB, units): the counting of viable pods located on the side branches of the plants.

Number of branches (NB, units): counting of branches longer than ten centimeters.

Number of pods with one seed (NP_1, units): counting of pods containing only one formed seed.

Number of pods with two seeds (NP_2, units): counting of pods containing two formed seeds.

Number of pods with three seeds (NP_3, units): counting of pods containing three formed seeds.

Number of pods with four seeds (NP_4, units): counting of pods containing four formed seeds.

Seed mass per plant (MSP, grams): weighing viable seeds, husked and individually cleaned, at 13% humidity.

The obtained data were subjected to the assumptions of the statistical model, which identified the normality and homogeneity of the residual variances and model additivity. Subsequently, two analysis scenarios were created. In the first, all information obtained in the experiment was preserved, both from segregating families and pure-line cultivars (controls), this scenario was called original (scenario I - Figure 1a). In the other scenario, progeny data were preserved, but the residual variability of controls was restricted using as criterion observations contained between the interval of the first sample standard deviation (-1s to + 1s). Thereby, an acceptable residue limit could be obtained (scenario II - Figure 1b).

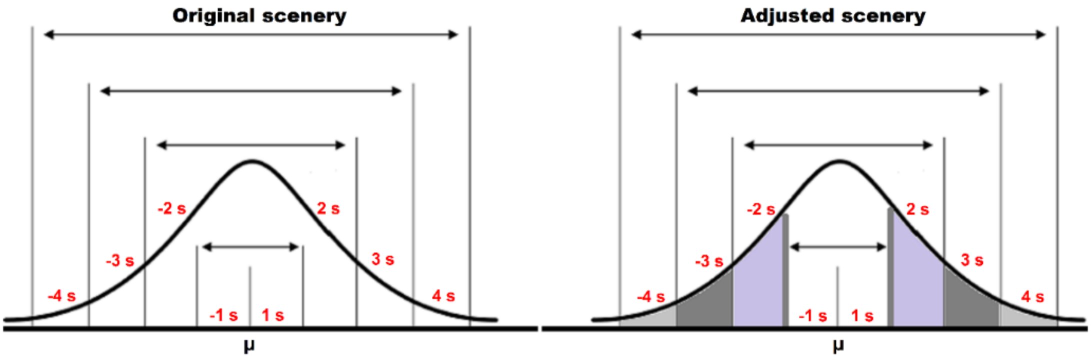

Figure 1. Illustration of scenarios (original and adjusted), standard deviations (- s and +s)

Both scenarios were submitted to three consolidated frequentist methods (approaches) to obtain components of variance and genetic parameters. The approach I is referring to genitor-progeny regression method, which is based on fixed effects (Lynch & Walsh, 1998) and wherein standardization through Z-score notations is crucial, as follows:

$$Z = \frac{Xi - \mu}{\theta}$$

where in: Xi corresponds to the value observed in the experimental unit, μ corresponds to the sample mean of trait of interest, and θ is sample standard deviation (Cruz, Carneiro, & Regazzi, 2014). Subsequently, a statistical model was used, as follows:

$$Y_i = \beta_0 + \beta_1 + \varepsilon_i$$

where in: Yi is based on the dependent character from the effects of progenies, β_0 shows the coefficient responsible for the origin of the information (intercept), β_1 represents the angular coefficient obtained between the explanatory information of the controls positioned on the x and x abscissa axis, ε_i reveals the residual effects between the adjustments of matrix X and Y.

The approach II is based on the use of the sum of squares of augmented block design with interim controls, using the model as follows:

$$\gamma_{ikj} = \mu + T_{k\prime} + T_{(j)k} + B_j + e_{kj}$$

where in: γ_{kj} is obtained in the i^{-th} experimental unit located in the j^{-th} block that housed the k^{-th} regular treatment (k') or common treatment (k), μ represents the overall mean of the experiment for the trait of interest, Tk' shows the fixed effect of regular treatments (k'= T_1, T_2, T_3, ... T_7), $T(j)k$ is the random effect of common treatments (k = P_1, P_2, ... P_{292}, r + c), r is the number of regular treatments considered, and c is the number of common treatments, $T(j)k$ ~ is the standardized normal distribution with mean centered at zero and variance equal to $T(j)k$ (NID ~ 0, σ^2), Bj represents the block effect (j = 1, 2, 3, 4), and e_{kj} represents the residual random effect.

The approach III was based on the method of mixed linear models wherein random genetic effects were considered, with genetic parameters weighted by a genealogical matrix (additive genetic variance between ½ full siblings of the genetic variance), as follows:

$$y = X\beta + Z\vartheta + e$$

where in: y corresponds to the vector of observations at the level of the experimental unit, β represents the parametric vector of the fixed effects, with incidence matrix X', ϑ shows the parametric vector of the random effects, with incidence matrix Z, and e is the vector responsible for capitalizing the residual variance (Resende et al., 2014). The estimates were based on the general equation of the mixed models, as follows:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'^{R^{-1}}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta^{\circ} \\ \vartheta \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

where in: y corresponds to the vector of observations at the experimental unit level, β represents the parametric vector of the fixed effects, with incidence matrix X', ϑ shows the parametric vector of the random effects, with incidence matrix Z, R determines the variance and covariance of the errors, and G presents the matrix of variance and covariance of the random effects.

By using these approaches, we could estimate the standard deviation of samples (s), the upper limit (ul) and lower limit (ll) of classes, the percentage of information within each class (%), arithmetic mean ($\overline{x}$), sample variance (s^2), sample covariance (COV), inter-class correlation (r), phenotypic variance (s^2P), environmental variance (s^2E), additive genetic variance (s^2AG), and narrow-sense heritability (h^2).

Results and discussion

Genetic basis used for inferences is founded by genetic recombination of 29 parents. These were crossed to obtain 34 breeding F2 populations, which, in turn, allowed selection of 292 F3 soybean progenies. Just as in pure lines (controls), these progenies were assessed according to the following traits: first pod insertion height (FPI), plant height (PH), number of pods on the main stem (NPMS), number of pods on the branches

(NPB), number of branches (NB), number of pods with one seed (NP1), number of pods with two seeds (NP2), number of pods with three seeds (NP3), number of pods with four seeds (NP4), and seed weight per plant (SMP). All these attributes demonstrate variability at 5% probability by the F-test (approaches I and II) and significance by the likelihood ratio test (LRT) at 5% probability by the x^2 test (approach III) for both scenarios. This provides conditions for both scenarios and approaches to estimate fundamental and reliable components of variance and genetic parameters, with a large capacity of repeating trends in subsequent studies. Martins, Unêda-Trevisoli, Môro, and Vieira (2016) defined that high genetic variability in soybeans is essential to attest biometric models applied in breeding programs. The inferences made in this study are widely used in the core of statistics, agricultural experimentation, genetic improvement, and biometrics, as they brought together wide genetic variability built by genetic complementarity of elite parents. In this way, trends, estimates, and forecasts can be directed to other studies and scenarios. Several are the difficulties during selection and management of segregating progenies, which hinders extraction of genetic parameters from tiny and unbalanced data, or due to their complex traits. In this sense, it is essential to develop effective methods to maximize genetic and selection gains, minimizing financial and labor costs and reproducing reliable information and parameters (Rezende, Cruz, Borém, & Rosado, 2021; Carvalho et al., 2020; Matta, Tomé, Salgado, Cruz, & Silva, 2015).

Approach I - genitor - progeny regression

When using approach I in parent-progeny regression, in the scenario I (Table 2), which includes original information, the controls (pure lines) expressed averages greater than progenies for all measured traits but first pod insertion height. This was expected in segregating F3 soybean generation since no direct selection was made for quantitative agronomic traits. It should be defined that compliance with an agronomic soybean ideotype will be served by maximizing the magnitude of pods, seeds, and mass of seeds per plant. Such a result would occur after several selection cycles and consolidated additive genetic gains in a progeny. The identification of phenotypic and genotypic attributes able to maximize soybean yield has contributed to genetic gains in breeding programs, both for statistical and biometric inference and for genomics (Jarquín et al., 2014). Based on the sample variance of superior progenies, the controls showed in scenario I the possibility of selection using characters related to plant architecture, such as height and height of insertion of the first pod, in addition to the number of lateral branches. One valid alternative to prove obtaining and selection of genetic variability that meets the agronomic ideotype is understanding covariation and genetic correlation between classes of individuals. These, when significant, indicate that attributes in progenies are similar to their parents and/ or cultivars assigned as control. Among these attributes are plant height and first pod insertion height, sizes of pods with three and four seeds, which express a similar sense in selection. The phenotypic inferences obtained in soybean breeding programs and the achievement of these parameters obtained by this linear regression model fit into selection assumptions that meet the soybean agronomic ideotype (Hanyu, Ferreira, Cecon, & Matsuo, 2020).

Scenario II revealed the maintenance above 64% of the original observations (scenario I) for all the traits measured. This was due to the smaller amplitude between the lower and upper limits defined by the residual variability criterion. In these conditions, the mean of the progenies was not altered and maintained the original trend. However, all traits revealed a reduction in the intrinsic mean of the controls, proving the hypothesis that the residual variability may inflate parameters extracted directly from the original information. The control variance was minimized by more than 60% if compared to the scenario I for all the traits measured. This trend maximized covariation and demonstrated an inversion of directions in inter-class correlations for plant height and number of pods, regardless of the number of seeds contained therein. Stochastic statistical models seek to quantify, control, and extract residual variance from the biological phenomenon under study. The focus is to allow maximization of parameters used as a basis for explicability and inferences about the dependent variable. In one parent regression models, the effects involved are based on the best fit to linear model able to maximize genetic covariation between classes (maternal and paternal parents, means of parents x those of sibling or half-sibling progenies), thus controlling further residual variations of observations made in each experimental unit of parents or controls. This allows inclination (b) to be maximized and express true genetic deviation and heritability (Falconer & Mackay, 1996).

Table 2. Estimation of the variance components and genetic parameters obtained by linear regression for genitors and progenies.

Approach I: linear regression for genitors and progenies

Trait	Scenario I						Scenario II					
	Mean		Variance		Covariance	Correlation	Mean		Variance		Covariance	Correlation
	Control	Progeny	Control	Progeny			Control	Progeny	Control	Progeny		
FPI[1]	11.19	12.92	11.93	53.76	0.41	0.02	10.78	12.92	4.91	53.76	1.22	0.08
PH	74.78	69.27	79.83	755.39	9.38	0.04	75.08	69.27	18.68	755.39	-9.61	-0.08
NPMS	37.35	28.39	572.83	197.77	-14.90	-0.04	36.36	28.39	165.84	197.77	-17.24	-0.10
NPB	70.95	30.69	1442.21	1054.21	-17.21	-0.01	66.37	30.69	459.19	1054.21	-4.94	-0.01
NB	6.04	3.54	4.93	8.77	-0.01	0.00	5.93	3.54	1.56	8.77	-0.47	-0.13
NP1	18.02	9.04	73.60	58.66	-0.34	-0.01	17.72	9.04	21.70	58.66	-3.97	-0.11
NP2	45.24	27.55	567.45	434.79	0.72	0.00	43.48	27.55	189.89	434.79	-2.16	-0.01
NP3	43.21	21.55	522.38	336.43	3.64	0.01	43.74	21.55	161.65	336.43	-31.93	-0.14
NP4	0.30	0.05	0.05	0.09	0.00	0.02	0.26	0.05	0.02	0.09	0.00	-0.02
SMP	32.36	18.00	244.74	178.69	-10.00	-0.05	32.11	18.00	68.24	178.69	-18.24	-0.17

Criterion	FPI	PH	NPMS	LPB	NB	NP1	NP2	NP3	NP4	SMP
ANOVA (5% F)	*	*	*	*	*	*	*	*	*	*
Overall mean	11.41	75.18	37.74	70.78	5.95	17.90	45.64	42.80	0.31	32.92
Standard deviation	3.87	9.17	23.26	38.60	2.28	8.92	24.67	23.30	0.23	16.37
Lower limit (-1DP)	7.54	66.02	14.47	32.19	3.66	8.98	20.97	19.50	0.08	16.55
Upper limit (+1DP)	15.28	84.35	61.00	109.38	8.23	26.82	70.31	66.11	0.54	49.29
Observations maintained (%)	75.10	69.40	64.00	65.50	67.60	66.80	65.50	64.20	68.70	69.70

[1]first pod insertion height (FPI); plant height (PH); number of pods on the main stem (NPMS); number of pods on the branches (NPB); number of branches (NB); number of pods with one seed (NP1); number of pods with two seeds (NP2); number of pods with three seeds (NP3); number of pods with four seeds (NP4) and seed mass per plant (SMP).

Approach II - augmented blocks with interim controls

In this approach, total variance was partitioned into variance components, which are attributed to a statistical model derived from augmented block design. These components are intended to improve the explicability of genetic inferences. These were divided into phenotypic, genetic, and environmental variations (Table 3).

Table 3. Estimation of the variance components and genetic parameters obtained by the model based on analysis of variance with augmented blocks and interim controls.

Approach II: model based on analysis of variance with augmented blocks and interim controls

Trait	Scenario I			Scenario II		
	Phenotypic	Environmental	Genetic	Phenotypic	Environmental	Genetic
FPI[1]	22.739	11.089	11.650	22.739	2.198	20.541
PH	268.402	49.186	219.216	268.402	14.189	254.214
NPMS	57.487	467.565	-410.078	57.487	151.491	-94.004
NPB	365.865	514.472	-148.607	365.865	145.404	220.461
NB	2.993	1.578	1.415	2.993	0.495	2.499
NP1	23.977	38.243	-14.266	23.977	15.693	8.283
NP2	171.866	355.895	-184.029	171.866	94.060	77.806
NP3	105.625	71.122	34.504	105.625	49.554	56.071
NP4	0.022	0.026	-0.004	0.022	0.008	0.014
SMP	60.931	71.070	-10.139	60.941	25.135	35.805

Criterion	FPI	PH	NPMS	LPB	NB	NP1	NP2	NP3	NP4	SMP
ANOVA (5% F)	*	*	*	*	*	*	*	*	*	*
Overall mean	11.05	74.07	40.01	79.93	6.36	19.22	46.71	37.91	0.27	31.33
Standard deviation	3.68	7.22	24.05	29.06	1.58	7.76	20.57	13.27	0.23	11.74
Lower limit (-1DP)	7.36	66.85	15.95	50.87	4.77	11.46	26.14	24.63	0.04	19.59
Upper limit (+1DP)	14.73	81.28	64.06	108.99	7.94	26.98	67.28	51.18	0.5	43.08
Observations maintained (%)	80	70	62.5	67.5	67.5	70	71.25	68.75	76.25	77.5

[1]first pod insertion height (FPI); plant height (PH); number of pods on the main stem (NPMS); number of pods on the branches (NPB); number of branches (NB); number of pods with one seed (NP1); number of pods with two seeds (NP2); number of pods with three seeds (NP3); number of pods with four seeds (NP4) and seed mass per plant (SMP).

Under these conditions, a direct comparison was directed to estimates obtained in the scenarios. We verified that the elaboration of scenario II resulted in the maintenance of more than 62% of original information. Phenotypic variance changes were minimal before the elaboration of scenarios. However, high distortions were obtained through environmental and genetic components. This is because their nature of estimation requires weightings and partitions previously established through the kinship matrix, as well as homozygosity coefficient. The component responsible for harboring environmental variance capitalized on a 20% reduction in variability, due to the use of scenario II, will certainly provide gains in genetic estimates for some traits. The use of augmented block design allows partitioning of variances assertively since they meet the additivity assumptions of statistical models, residue independence, as well as residual variance homogeneity and normality. Altogether, it allows including repeated and randomized parents and controls. Moreover, when segregating progenies are intercalated, major phenotypic and genetic parameters can be maximized, fitting large tests in smaller areas, and obtaining the best accuracy of biometric estimates and predictions (Peternelli, Souza, Barbosa, & Carvalho, 2009; Coimbra et al., 2006).

Basic assumptions are expressed in components of variance, and negative estimates were not considered since they are expressions of variance inflation and thus should be discarded. In this sense, new studies should be directed so that number of plants per experimental unit is increased at the family level. This can be justified by number of pods in the main stem, which, regardless of the improvement in the quality of the experimental data, did not reflect in reliable estimates. In this context, biometric gains from the new approach are listed, wherein genetic effects were maximized in scenario II. This is because, by restricting residual variance of pure control lines, satisfactory genetic estimates to the proposed genetic-statistical model were obtained. This is because the traits number of pods in branches; number of pods with one, two, and four seeds; and seed mass per plant are of extreme importance for a soybean breeding program aiming at genetic gains in productivity.

Approach III - mixed linear models with random genetic effects

When considering random genetic effects for the 292 F3 segregating families, scenario-independent phenotypic variance was increased when compared to the approaches I and II, which maintained fixed genotypic effects (Table 4).

Table 4. Estimation of the variance components and genetic parameters obtained by restricted maximum likelihood (REML).

<table>
<tr><th colspan="11">Approach III: mixed models with random genetic effects</th></tr>
<tr><th rowspan="2">Trait</th><th colspan="3">Scenario I</th><th colspan="7">Scenario II</th></tr>
<tr><th>Phenotypic</th><th>Environmental</th><th>Genetic</th><th colspan="3">Phenotypic</th><th colspan="2">Environmental</th><th colspan="2">Genetic</th></tr>
<tr><td>FPI[1]</td><td>48.541</td><td>2.238</td><td>16.068</td><td colspan="3">47.477</td><td colspan="2">1.406</td><td colspan="2">16.365</td></tr>
<tr><td>PH</td><td>583.476</td><td>4.183</td><td>155.917</td><td colspan="3">581.217</td><td colspan="2">1.608</td><td colspan="2">158.176</td></tr>
<tr><td>NPMS</td><td>187.471</td><td>6.850</td><td>17.176</td><td colspan="3">162.992</td><td colspan="2">0.923</td><td colspan="2">21.510</td></tr>
<tr><td>NPB</td><td>1062.327</td><td>20.676</td><td>203.277</td><td colspan="3">1020.302</td><td colspan="2">17.852</td><td colspan="2">212.374</td></tr>
<tr><td>NB</td><td>7.749</td><td>0.148</td><td>1.094</td><td colspan="3">7.639</td><td colspan="2">0.136</td><td colspan="2">1.124</td></tr>
<tr><td>NP1</td><td>58.812</td><td>1.838</td><td>14.706</td><td colspan="3">55.839</td><td colspan="2">0.591</td><td colspan="2">15.632</td></tr>
<tr><td>NP2</td><td>416.725</td><td>17.225</td><td>83.289</td><td colspan="3">396.618</td><td colspan="2">15.394</td><td colspan="2">85.389</td></tr>
<tr><td>NP3</td><td>322.437</td><td>2.104</td><td>54.271</td><td colspan="3">315.080</td><td colspan="2">0.746</td><td colspan="2">56.831</td></tr>
<tr><td>NP4</td><td>0.440</td><td>0.007</td><td>0.248</td><td colspan="3">0.425</td><td colspan="2">0.000</td><td colspan="2">0.281</td></tr>
<tr><td>SMP</td><td>173.898</td><td>8.711</td><td>29.071</td><td colspan="3">165.539</td><td colspan="2">3.917</td><td colspan="2">30.810</td></tr>
<tr><th>Criterion</th><th>FPI</th><th>PH</th><th>NPMS</th><th>LPB</th><th>NB</th><th>NP1</th><th>NP2</th><th>NP3</th><th>NP4</th><th>SMP</th></tr>
<tr><td>Deviance/LRT (5% X^2)</td><td>273,2*</td><td>142.89*</td><td>71.81*</td><td>192.52*</td><td>97.22*</td><td>245.47*</td><td>182.6*</td><td>102.86*</td><td>17.11*</td><td>167.12*</td></tr>
<tr><td>Overall mean</td><td>11.05</td><td>74.07</td><td>40.01</td><td>79.93</td><td>6.36</td><td>19.22</td><td>46.71</td><td>37.91</td><td>0.27</td><td>31.33</td></tr>
<tr><td>Standard deviation</td><td>3.68</td><td>7.22</td><td>24.05</td><td>29.06</td><td>1.58</td><td>7.76</td><td>20.57</td><td>13.27</td><td>0.23</td><td>11.74</td></tr>
<tr><td>Lower limit (-1DP)</td><td>7.36</td><td>66.85</td><td>15.95</td><td>50.87</td><td>4.77</td><td>11.46</td><td>26.14</td><td>24.63</td><td>0.04</td><td>19.59</td></tr>
<tr><td>Upper limit (+1DP)</td><td>14.73</td><td>81.28</td><td>64.06</td><td>108.99</td><td>7.94</td><td>26.98</td><td>67.28</td><td>51.18</td><td>0.5</td><td>43.08</td></tr>
<tr><td>Observations maintained (%)</td><td>80</td><td>70</td><td>62.5</td><td>67.5</td><td>67.5</td><td>70</td><td>71.25</td><td>68.75</td><td>76.25</td><td>77.5</td></tr>
</table>

[1] first pod insertion height (FPI); plant height (PH); number of pods on the main stem (NPMS); number of pods on the branches (NPB); number of branches (NB); number of pods with one seed (NP1); number of pods with two seeds (NP2); number of pods with three seeds (NP3); number of pods with four seeds (NP4) and seed mass per plant (SMP).

In this approach, we obtained parameter estimates through maximum restricted likelihood, which presents greater flexibility in terms of residual variance, mainly for normality and homogeneity (Volpato et al., 2018). In this context, the comparison among scenarios showed that the scenario with residual variance restriction forced a gradual reduction in phenotypic component of estimates. This fact was not evident in approach II, which is based on sum of squares.

In approach I, variance in the environmental component was extremely minimized, especially in scenario II. This did not present negative estimates for the genetic component given the nature of mathematical estimates. On the other hand, when reconciling mixed model method, residual variability restriction in controls showed an abrupt maximization of genetic components, regardless of the target trait. This is seen as a great gain for genetic improvement, especially in early generations where the size of plants does not always reach an optimum number to minimize waste and maximize parameters (Del Conte, Carneiro, Resende, Silva, & Peternelli, 2020).

Estimation of narrow sense heritabilities

Regardless of crop, segregating generation, and selection place, trait heritability is by far the most important and indispensable parameter in breeding programs. This parameter defines the ability of an individual or progeny to determine phenotypic manifestation of a trait through genetic effects (Falconer & Mackay, 1996). Among the types of heritability, the narrow-sense is vital for breeders to succeed in breeding programs. The magnitude of this parameter is estimated as the ratio between additive genetic variance and total phenotypic variance. Here, we sought to minimize distortions intrinsic to total phenotypic variance and to maximize additive genetic variance, hence maximizing the extracted parameter. Moreover, this ratio ranges between 0 and 1 and does not assume negative behavior (Mather & Jinks, 1984). Therefore, many traits are difficult to estimate because breeders are often unable to cover adversities in selection fields, which can distort general variation, and hence phenotypic variation (Lynch & Walsh, 1998).

Based on these precepts, approaches based on parent-progeny regression are robust and not quite plastic in their experimental nature. Therefore, regardless of the scenario, heritability estimates were not proven in strict sense (Table 5). For approach II, where assumptions of augmented blocks are applied with provisional controls, it is evident that scenario I has no basis for number of pods on the main stem; number of pods on branches; number of pods with one, two, and four seeds; and seed mass per plant. In this approach, the restriction of residual variability in controls is justified since 90% of the measured variables had reliable estimates. This demonstrates the specificity of restricting variability in augmented designs with interim controls with similar traits.

Table 5. Narrow-sense heritability for soybean attributes of agronomic interest.

	Approach I		Approach II		Approach III	
Trait	C.I	C.II	C.I	C.II	C.I	C.II
FPI[1]	0.034	0.246	0.510	0.900	0.330	0.340
PH	0.116	-0.510	0.810	0.980	0.260	0.270
NPMS	-0.026	-0.103	-7.130	-1.630	0.090	0.130
NPB	-0.012	-0.011	-0.400	0.600	0.190	0.200
NB	-0.002	-0.297	0.470	0.830	0.140	0.140
NP1	-0.005	-0.181	-0.590	0.340	0.250	0.270
NP2	0.001	-0.011	-1.070	0.850	0.190	0.210
NP3	0.007	-0.196	0.320	0.530	0.160	0.180
NP4	0.023	-0.056	-0.200	0.620	0.560	0.650
SMP	-0.040	-0.265	-0.160	0.580	0.160	0.180

[1] first pod insertion height (FPI); plant height (PH); number of pods on the main stem (NPMS); number of pods on the branches (NPB); number of branches (NB); number of pods with one seed (NP1); number of pods with two seeds (NP2); number of pods with three seeds (NP3); number of pods with four seeds (NP4) and seed mass per plant (SMP).

Regarding approach III, all estimates were reliable regardless of the experimental scenario used. It arises from the nature of the component estimation since the maximum likelihood assumes that the model is true and that the initial and final parameters are reliable. When comparing the use of the new residual variance restriction approach in pure lines, controls, or parents, all parameters assessed in this study showed an increase in the genetic component, and hence narrow-sense heritability. This fact demonstrates its applicability in plant breeding to maximize genetic gains during segregating generations and selection strategies.

Conclusion

Restricting residual variation in parents or controls can maximize genetic parameters and gains in soybean breeding. Relevant gains in heritability estimates are obtained in the augmented blocks with interim control approach. Mixed linear models with random genetic effects can be considered a great tool to obtain genetic parameters in experiments with a high magnitude of common and regular treatments.

References

Almeida, R. D., Peluzio, J. M., & Afferri, F. S. (2010). Correlações fenotípicas, genotípicas e ambientais em soja cultivada sob condições várzea irrigada, sul do Tocantins. *Bioscience Journal, 26*(1), 95-99.

Carvalho, I. R., Nardino, M., Demari, G., Pelegrin, A. J., Ferrari, M., Szareski, V. J., ... Maia, L. (2017). Components of variance and inter-relation of important traits for maize (Zea mays) breeding. *Australian Journal of Crop Science*, *11*(8), 982-988. DOI: https://doi.org/10.21475 / ajcs.17.11.08.pne474

Carvalho, I. R., Silva, J. A. G., Ferreira, L. L., Szareski, V. J., Demari, G., Facchinello, P. H. K., ... Souza, V. Q. (2020). Relative contribution of expected sum of squares values for soybean genotypes × growing environments interaction. *Australian Journal of Crop Science, 14*(3), 382-390. DOI: https://doi.org/10.21475/ajcs.20.14.03.p1515

Chechi, A., Deuner, C. C., Forcelini, C. A., & Boller, W. (2020). Asian soybean rust control in response to rainfall simulation after fungicide application. *Acta Scientiarum. Agronomy*, *43*(1), 1-9. DOI: https://doi.org/10.4025/actasciagron.v43i1.45689

Cruz, C. D., Carneiro, P. C. S., & Regazzi, A. J. (2014). *Modelos biométricos aplicados ao melhoramento genético* (5. ed.). Viçosa, MG: Editora UFV.

Companhia Nacional de Abastecimento [CONAB]. (2019). *Acompanhamento da safra brasileira de grãos: 9º levantamento grãos safra 2016/17*. Retrieved on july 10, 2020 from http://www.conab.gov.br/OlalaCMS/uploads/arquivos/17_06_08_09_02_48_boletim_graos_junho_2017.pdf

Coimbra, J. L. M., Souza, V. Q. D., Kopp, M. M., Silva, J. G. C. D., Oliveira, A. C. D., & Carvalho, F. I. F. D. (2006). Esperanças matemáticas dos quadrados médios: uma análise essencial. *Ciência Rural*, *36*(6), 1730-1738. DOI: http://dx.doi.org/10.1590/S0103-84782006000600010

Del Conte, M. V., Carneiro, P. C. S., Resende, M. D., Silva, F. L., & Peternelli, L. A. (2020). Overcoming collinearity in path analysis of soybean [*Glycine max* (L.) Merr.] grain oil content. *PLoS ONE*, *15*(5), 1-15. DOI: https://doi.org/10.1371/journal.pone.0233290

Hanyu, J., Ferreira, S. C., Cecon, P. R., & Matsuo, E. (2020). Genetic parameters estimate and characters analysis in phenotypic phase of soybean during two evaluation periods. *Agronomy Science and Biotechnology*, *6*, 1-12. DOI: https://doi.org/10.33158 / ASB.r104.v6.2020

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. (4. ed.). London, UK: Longman.

Jarquín, D., Kocak, K., & Posadas, L. Hima, K., Jedlicka, J., Graef, G., & Lorenz, A. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics*,*15*(740), 1-10. DOI: https://doi.org/10.1186/1471-2164-15-740.

Liu, S., Xue, H., Zhang, K., Wang, P., Su, D., Li, W., & Li, X. (2019). Mapping QTL affecting the vertical distribution and seed set of soybean [*Glycine max* (L.) Merr.] pods. *The Crop Journal*, *7*(5), 694-706. DOI: https://doi.org/10.1016/j.cj.2019.04.004

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. (1. ed.). Sunderland, UK: Sinauer.

Martins, C. C., Unêda-Trevisoli, S. H., Môro, G. V., & Vieira, R. D. (2016). Metodologia para seleção de linhagens de soja visando germinação, vigor e emergência em campo. *Revista Ciência Agronômica, 47*(3), 455-461.

Mather, K., & Jinks, J. L. (1984). *Introdução à genética biométrica*. (2. ed.). Wantage, UK: Sociedade Brasileira de Genética.

Matta, L. B., Tomé, L. G. O., Salgado, C. C., Cruz, C. D., & Silva, L. F. (2015). Hierarchical genetic clusters for phenotypic analysis. *Acta Scientiarum. Agronomy*, *37*(4), 447-456. DOI: https://doi.org/10.4025/actasciagron.v37i4.19746

Nogueira, A. P. O., Sediyama, T., Sousa, L. B., Hamawaki, O. T., Cruz, C. D., & Pereira, D. G. (2012). Análise de trilha e correlações entre caracteres em soja cultivada em duas épocas de semeadura. *Bioscience Journal*, *28*(6), 877-888.

Peternelli, L. A., Souza, E. F. M. D., Barbosa, M. H. P., & Carvalho, M. P. D. (2009). Delineamentos aumentados no melhoramento de plantas em condições de restrições de recursos. *Ciência Rural*, *39*(9), 2425-2430. DOI: https://doi.org/10.1590/S0103-84782009005000209

Resende, M. D. V., Silva, F. F., & Azevedo, C. F. (2014). *Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão, aleatória, seleção genômica, QTL, GWAS, estatística espacial e temporal, competição, sobrevivência*. (1. ed.). Visconde do Rio Branco, MG: Suprema.

Rezende, W. S., Cruz, C. D., Borém, A., & Rosado, R. D. S. (2021). Half a century of studying adaptability and stability in maize and soybean in Brazil. *Scientia Agricola*, *78*(3), 1-9. DOI: https://doi.org/10.1590/1678-992x-2019-0197

Szareski, V. J., Souza, V. Q., Carvalho, I. R., Nardino, M., Follmann, D. N., Demari, G. H., ... Olivoto, T. (2015). Growing environment and its effects on morphological characters and dietetic soy. *Revista Brasileira de Agropecuária Sustentável*, *5*(2), 79-88. DOI: https://doi.org/10.21206/rbas.v5i2.247

Torres, F. E., Teodoro, P. E., Sagrilo, E., Ceccon, G., & Correa, A. M. (2015). Interação genótipo x ambiente em genótipos de feijão-caupi semiprostrado via modelos mistos. *Bragantia*, *74*(3), 255-260. DOI: https://doi.org/10.1590/1678-4499.0099

Troyjack, C., Dubal, Í. T. P., Koch, F., Szareski, V. J., Pimentel, J. R., Carvalho, I. R., ... Pedó, T. (2017). Attributes of growth, physiological quality and isoenzymatic expression of common bean seeds produced under the effect of gibberellic acid. *Australian Journal of Crop Science*, *11*(9), 1116-1122. DOI: https://doi.org/10.21475/ajcs.17.11.09.pne531

Volpato, L., Alves, R. S., Teodoro, P. E., Resende, M. D. V., Nascimento, M., Nascimento, A. C. C., ... Borém, A. (2019). Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. *PLoS ONE*, *14*(4), 1-22. DOI: https://doi.org/10.1371/journal.pone.0215315

Volpato, L., Simiqueli, G. F., Alves, R. S., Rocha, J. R. A. S. C., Del Conte, M. V., Resende, M. D. V., ... Silva, F. L. (2018). Selection of inbred soybean progeny (*Glycine max*): an approach with population effect. *Plant Breeding*, *137*(6), 865-872. DOI: https://doi.org/10.1111/pbr.12648

Yang, M. H., Mohamed, Z. Z. J., He, J., Dom, R., Hofmann, R., Siddique, K. H. M., & Li, F. M. (2020). "Effect of traditional soybean breeding on water use strategy in arid and semi-arid areas. *European Journal of Agronomy*, *120*. DOI: https://doi.org/10.1016/j.eja.2020.126128

Woyann, L. G., Meira, D., Zdziarski, A. D., Metei, G., Milioli, A. S., Rosa, A. C., ... Benin, G. (2019). Multiple-trait selection of soybean for biodiesel production in Brazil. *Industrial Crops and Products*, *140*, 111721. DOI: https://doi.org/10.1016/j.indcrop.2019.111721