

Application of the BLUPe predictor in the selection of potential soybean varieties for Orinoquia

Aplicación del predictor BLUPe en la selección de variedades potenciales de soya para la Orinoquia

Rubén Alfredo Valencia Ramírez¹ and Yuli Stephani Tibocho Ardila^{1*}

ABSTRACT

The Colombian Orinoquia and the Altillanura subregion show comparative and competitive advantages for soybean production (edaphoclimatic conditions, cost-benefit ratio, potential area, and location) essential food with a high protein content (~37%) and used mostly for poultry and pig nutrition. However, this immense region has scarce varietal alternatives of high grain yield and quality that are adapted to its climatic and edaphic conditions. The current research is based on the selection of superior lines or potential varieties with high genetic merit using the restricted maximum likelihood/standardized best linear unbiased predictor (REML/BLUPe) procedure. Sixty advanced lines and four commercial varieties were tested in an 8x8 alpha lattice design. Grain yield (GY) oscillated between 1,117 and 4,431 kg ha⁻¹, the population average yield was 2,682 kg ha⁻¹, and BLUPe predictors ranged between 5.37 and -3.71. With a t-test at a significance of 5% (1.67) and a predictor comparator of $(t_1-t_2) \geq 1.67 (\sqrt{2})$, six outstanding lines were identified with superior BLUPe values compared to the mean ($P < 0.05$) and $GY > 3,500$ kg ha⁻¹. In descending order (kg ha⁻¹), the GY was: L-041 (4,431), L-019 (4,326), L-104 (3,923), L-149 (3,832), L-202 (3,536), and L-201 (3,519 kg ha⁻¹). The BLUPe standardized predictor allowed an effective selection (92%) of lines.

Key words: Altillanura, genetic breeding, mixed models, REML.

RESUMEN

La Orinoquia colombiana y en particular, la subregión de la Altillanura presenta ventajas comparativas y competitivas para la producción de soya (condiciones edafoclimáticas, relación costo-beneficio, área potencial y ubicación), alimento esencial por su contenido proteico (~37%), especialmente para aves y cerdos. Sin embargo, las alternativas varietales adaptadas y de alto rendimiento de grano son escasas para esta vasta región. Por ello, la presente investigación fue orientada hacia la selección de líneas mejoradas o variedades potenciales con alto mérito genético mediante la aplicación del mejor predictor lineal insesgado estandarizado (REML/BLUPe). Se evaluaron 60 líneas avanzadas y cuatro variedades comerciales en un diseño Alfa látice 8x8. El rendimiento de grano (RG) osciló entre 1117 y 4431 kg ha⁻¹, el promedio poblacional fue de 2682 kg ha⁻¹ y los predictores BLUPe oscilaron entre 5.37 y -3.71. Con una significancia t del 5% (1.67) y un comparador de predictores $(t_1-t_2) \geq 1.67 (\sqrt{2})$, se identificaron 6 líneas sobresalientes con BLUPe superiores a la media ($P < 0.05$) y $RG > 3500$ kg ha⁻¹. En orden descendente el RG fue: L-041 (4431), L-019 (4326), L-104 (3923), L-149 (3832), L-202 (3536) y L-201 (3519 kg ha⁻¹). El predictor estandarizado BLUPe permitió una selección efectiva (92%) de líneas.

Palabras clave: Altillanura, mejoramiento genético, modelos mixtos, REML.

Introduction

The Orinoco basin is shared by Colombia and Venezuela and has an area of 981,446 km². Thirty-five percent of this vast region is located in the equatorial zone of Colombia between 2°N and 10°N and 60°W and 75°W (Vásquez Cerón *et al.*, 2019). In the savannah complex of the Orinoco basin, the Colombian Altillanura includes 2.8 million ha that, according to the CONPES document No. 3797 (DNP, 2014), has high agricultural potential with typical oxidic soils, where soybean is projected to have comparative and competitive advantages as a first-class raw material as

balanced feed, especially for poultry, pigs, and the production of vegetable oil.

In the Altillanura only 39,793 ha are planted per year, including two sowing cycles, with a limited varietal offering and a production of close to 119,412 t of grain (FENALCE, 2021). This supply, representing less than 7% of the national demand, has motivated agribusinesses to approach national soybean production as an inclusive, economically viable, and ecologically sustainable regional agricultural development. Currently, the Corporación Colombiana de Investigación Agropecuaria - Agrosavia is carrying out a

Received for publication: March 16, 2022. Accepted for publication: May 23, 2022.

Doi: 10.15446/agron.colomb.v40n2.101137

¹ Corporación Colombiana de Investigación Agropecuaria AGROSAVIA, Centro de investigación La Libertad, Villavicencio, Meta (Colombia).

* Corresponding author: ytibocho@agrosavia.co



soybean breeding program for Oxisols from the Orinoquía region. This development aims to generate varieties adapted to the low tropics with high yield potential, vegetative cycles compatible to integrated crop rotation systems, tolerant to adverse biotic and abiotic factors. The importance of expanding the existing varietal genetic base for this vast region and the need to close the grain yield gap with varieties that exceed 3,400 kg ha⁻¹ is highlighted.

The process of genetic breeding is arduous and complex. Its initial phase implies the introduction, characterization, and evaluation of diverse germplasm, a crossing plan in search of the best genetic combinations, and later a selection of superior individuals that will become advanced lines or potential varieties. These lines, their components, and sanitary behavior are evaluated in a first cycle through preliminary yield trial (PYT). Then, the best lines are subjected to multi-environment trials (MET) to determine their level of phenotypic stability. The most outstanding ones are subjected to regional (RT) or agronomic evaluation trials (AET), regulated by the Colombian Agricultural Institute (ICA, 2020), where the superior line or lines with identity attributes, homogeneity, and stability are registered as new cultivars.

Grain yield and other agronomic traits in genetic breeding programs have traditionally been valued using the classical generalized linear model (GLM). However, these models do not satisfy the statistical assumptions, particularly with unbalanced databases (Resende, 2007) nor through the predominant application of fixed factors usually analyzed through mean comparison tests such as DHS Tukey, Scheffé, and LSD, etc. There are statistical procedures, such as restricted maximum likelihood/best linear unbiased prediction (REML/BLUP), generically called a mixed model methodology (Sturion & Resende, 2010; Bandera-Fernández & Pérez-Pelea, 2018) that improves the estimation and prediction of genetic parameters and the effectiveness of selection to reduce these biases. Mixed models have provided valuable information for selecting lines by genetic merit. According to Piepho *et al.* (2008), BLUP is better in genetic breeding and variety evaluation precision than the best linear unbiased estimator (BLUE). The BLUP predictor has advantages over BLUE by reducing the response due to environmental effects and efficiently discriminating genotypes with high varietal potential (Pacheco *et al.*, 2020). The mean associated with fixed-effects in BLUE is an average performed over all the effect levels in the population, while BLUP is a regression towards the general mean based on the variance and covariance

components associated with the random-effects model (Casanoves & Balzarini, 2002). BLUP has a shrinkage property, so the prediction values tend to be less separated from the mean than the original values, increasing the precision of the analysis (Baselga & Blasco, 2008). In this sense, BLUP is a standard method for random effects and requires the estimation of the genetic variance (σ^2_g) and residual (σ^2) components, preferably through REML (Mora & Arnhold, 2006). The latter compensates for the loss of degrees of freedom that results from the estimation of the fixed effects and produces less biased estimates (Patterson & Thompson, 1971). Overall, for authors such as Searle *et al.* (1992), Robinson (1991), Piepho and Möhring (2006), Resende (2016), and Volpato *et al.* (2019), REML/BLUP has become an effective procedure for estimating parameters and predicting genetic values to optimize selection.

Resende and Duarte (2007) suggest subjecting experiments with more than 10 cultivars or genotypes to a BLUP analysis as a random source of the mixed model. The genotype ranking could be similar for phenotypic selection and the one performed by BLUP when data are balanced. Although it is essential to link pedigree information in the genetic merit analysis, for Piepho *et al.* (2008), it is easy to exploit the information via BLUP through a simple mixed model without explicit reference to the pedigree. BLUP maximizes the correlation between the real and predicted genotypic values, *i.e.*, the main objective of the breeder. In this regard, Panter and Allen (1995), evaluating highly related individuals, found that the kinship matrix link in the BLUP analysis is not justified. However, if the data of the historical parents are available in crosses that are not closely related, pedigree data should be included because it can improve the prediction accuracy of progeny performance.

Accordingly, the current research was carried out to select advanced soybean lines or potential varieties with high genetic merit and grain yield >3,500 kg ha⁻¹ in a preliminary yield trial for Oxisols of the Colombian Orinoquean region, using the REML/BLUP methodology and its standardized BLUPe value.

Materials and methods

Study area

A preliminary yield trial (PYT) of 60 advanced soybean lines of the Agrosavia genetic breeding program and four commercial varieties (C. Superior 6, C. Achagua 8, C. Primavera 11, and Soyica P-34) was carried out during the second period of 2020 at the research center La Libertad,

Villavicencio, Meta (Colombia). The study site is located at 4°22' N and 72°13' W, in Oxisols with a pH of 4.9, 5.9 mg L⁻¹ of phosphorus, 2.4% of organic matter, and a base saturation of 57.3%.

Plant material, trials, and variables

The soybean lines and varieties were planted in plots with four rows of 5 m long, 0.45 m apart, with a distance between plants of 0.07 m. Mineral and biological fertilization with *Bradyrhizobium japonicum* was uniform according to soil analysis and crop requirements. The experimental units were distributed in an alpha lattice design proposed by Patterson and Williams (1976) of 8x8 with two replicates. Grain yield was evaluated as the response variable of interest for the BLUP analysis, and the following phenotypic variables that characterize the genotypes under study were registered.

Grain yield (GY) included grain weight in kg ha⁻¹ per experimental unit and grain moisture contents of 14%; days to flowering (DF) was the number of days between emergence and flowering in the upper nodes of the plant; days to physiological maturity (DM) was the number of days from emergence to reaching 95% of mature pods in each plot, at which point the average height (H) and the number of nodes (NN) per plant were also recorded.

Additionally, the qualitative description of the lines was made using flower color (FC), growth habit (GH), and pubescence color (PC).

To determine water excess and deficits during the crop cycle, the hydric balance model (Allen *et al.*, 2006) and the database of the meteorological station La Libertad ascribed to IDEAM were used. Real evapotranspiration was obtained using the tank coefficient (Kp) (Cruz Valderama, 2015) and the FAO cultivation coefficients (Kc) (FAO, 2000).

Statistical analysis

In the analysis of variance, the statistical model described by Singh and Bhatia (2017) for incomplete blocks was applied (Equation 1):

$$Y_{ijk} = \mu + G_i + \gamma_j + \rho_{k(j)} + \varepsilon_{ijk} \quad (1)$$

Where Y_{ijk} is the phenotype of i^{th} genotype in j^{th} replication and k block; G_i is the genotype effect $i=1,2,\dots,k$; γ_j is the replicate effect $j=1,2,\dots,r$; $\rho_{k(j)}$ is the block within replicate effect $k=1,2,\dots,S$; ε_{ijk} is the random error. For the

statistical analysis, the matrix structure of the mixed linear model (Piepho *et al.*, 2008) was applied using the SAS 9.4 software (SAS Institute, 2014). The variance components were estimated with the PROC MIXED/REML procedure of the SAS System (Bueno Filho & Vencovsky, 2000). This analysis considers lines and the block effects within the repetition as random effects. The comparison of means was carried out using the BLUP predictors (best linear unbiased prediction) that represent the predicted value for each genotype with respect to the general mean (Biasutti, 2012). The comparison of the BLUPs between lines/varieties was performed with the t statistic (Yan & Rajcan, 2002). This statistical comparator called BLUPe (standardized BLUP) was obtained from the relationship between the predicted empirical BLUP value and the associated prediction error. The BLUPe (standardized BLUP) was obtained from the BLUP predictor or predicted value for each genotype concerning the general mean (Yan & Rajcan, 2002). For the comparison analysis of the standardized predicted values, the genotypes are considered different if the BLUPe values met the following condition, with a t-test significance of 5%: $(t_1-t_2) \geq 1.67 (\sqrt{2})$ (Yan *et al.*, 2002). Additionally, the probabilistic values generated by SAS/Mixed for each BLUPe were used to determine the superiority ($P \leq 0.05$) or inferiority ($P > 0.05$) of the lines/varieties concerning the general mean (Casanoves & Balzarini, 2002). The BLUPe pair GY were ordered in descending order to identify the genotypes or superior lines. This methodology allowed comparing free genetic values of environmental effects and not the phenotypic means to improve genetic gain in the subsequent selection cycle.

The elevated kinship of the parents that gave rise to the group of lines makes using the coefficient of coancestry in the BLUP analysis unnecessary since it does not represent a change in the results. Furthermore, when the purpose is to estimate the total genotypic value, it is reasonable not to use the coefficient of coancestry (Piepho *et al.*, 2008).

The level of experimental precision in selecting superior lines was measured based on the square root of broad-sense heritability, calculated on a mean plot basis using the following equation (Hacker & Cuany, 1997):

$$H^2 = \frac{Vg}{\left(Vg + \frac{Ve}{r} \right)} \quad (2)$$

Where Vg represents genetic variance, Ve is the environmental variance, and r is the repetitions.

Results and discussion

The descriptive analysis of the variables DF, DM, H, NN, and GY (Tab. 1) allowed inferring a high variability as a genetic source for varietal improvement. The wide ranges of DF and DM (22 and 29 d, respectively) underlined the differential behavior of the genotypes in precociousness, together with the high variation of H, NN, and GY. In general, the lines and varieties with indeterminate GH showed higher H than those with semi-determined GH, and these, in turn, were higher than those with determined GH. Plants with indeterminate GH continue to grow for a long time after flowering, while those with determinate GH finish stem growth when or shortly after flowering begins, often resulting in fewer nodes than the indeterminate GH (Fehr & Caviness, 1977). These differences in H and GH do not necessarily represent lower yields in those with determined GH since agronomic-importance traits such as GY are quantitative (Volpato *et al.*, 2019), where the environmental effect represents a large part of the variation. In the analysis of variance, highly significant differences ($P < 0.01$) between lines and varieties were ratified for the case of GY by BLUPe predictor values ranging between 5.37 and -3.71. The GY range was between 1,117 and 4,431 kg ha⁻¹. It is noteworthy that, in the edaphoclimatic conditions of the current study, there were no phytosanitary problems that affected the GY response variable.

A highly significant positive Pearson correlation ($P < 0.01$) was found between GY and DM ($r = 0.32$) that means that the higher the DM, the higher the GY under the agroclimatic conditions of the preliminary yield trial. In contrast, a shortening of the filling period is frequently associated with decreasing grain weight (Kantolic *et al.*, 2004).

BLUPe vs. grain yield

The BLUP and BLUPe predictors and the significance of the random effects for GY are presented in Table 2. The BLUPe ($t_1 - t_2 \geq 1.67 (\sqrt{2})$) comparator allowed an effective

differentiation of the genotypes; by ordering them in descending order, a range with a maximum of 5.37 and a minimum of -3.71 for lines L-019 and L-078 was reached. Genotypes with positive BLUPe for GY were classified as higher, and those with a negative value were considered lower with respect to the general mean of 2,682 kg ha⁻¹. The higher the BLUPe, the greater the probability of success in selecting superior lines. The standardized BLUP (BLUPe) is more discriminant than the empirical BLUP when selecting outstanding lines because the prediction error adjusts it (Casanoves & Balzarini, 2002). Therefore, a higher mean does not necessarily represent a higher BLUPe, as occurs between lines L-041 and L-104. The lines with $GY \geq 3,519$ kg ha⁻¹ showed statistical differences ($P > |t|$) concerning the general mean and positive BLUPe ≥ 2.52 (Fig. 1), and they constituted promising lines or potential varieties. In descending order, the GY (kg ha⁻¹) were: L-041 (4,431), L-019 (4,326), L-104 (3,923), L-201 (3,832), L-149 (3,536), and L-202 (3,519), with higher GY for the Colombian Orinoco region.

The genetic base of these lines includes the varieties that were developed for the Colombian Orinoco, such as Ori-noquía 3, C. Libertad 4, and C. Taluma 5, sources of adaptation to the low tropics, with prominent differences in precocity and high yield potential that were crossed with elite materials from the EMBRAPA variety bank (Brazil).

The only control variety that exceeded the general average was C. Primavera 11, although with a BLUPe significantly lower than L-019. The other controls did not differ statistically from the mean, where the C Achagua 8 variety had a positive BLUPe (0.55), while in the Soyica P-34 and C. Superior 6 varieties, the predictor was negative (-0.03 and -1.71). The upper lines, representing a selection pressure of 10% with an estimated genetic gain of 37.2% concerning the average, were subjected to multi-environment trials to assess their behavior in different agroclimatic conditions and establish their phenotypic stability.

TABLE 1. Descriptive analysis of agronomic and yield characteristics of lines evaluated.

Statistics	DF	DM	H (cm)	NN	GY (kg ha ⁻¹)
N	127	124	123	124	116
Minimum	31	80	43	11	1,117
Maximum	53	109	112	25	4,431
Mean	40	93	77	17	2,682
SD	5	8	18	3	690
CV (%)	13	8	14	16	26

DF: days to flowering; DM: days to physiological maturity; H: average plant height at maturity; NN: average number of nodes; GY: grain yield; N: data number; SD: standard deviation; CV: coefficient of variation.

TABLE 2. Grain yield (GY, kg ha⁻¹), BLUP and BLUPe predictors, and significance level of soybean lines and varieties in the preliminary yield trial (PYT), Research Center La Libertad, Orinoquía, 2020.

Line/Variety	GY	BLUP	BLUPe	Pr> t	Line/Variety	GY	BLUP	BLUPe	P> t
L-019	4,326	1,377.76	5.37	<.0001	L-052	2,632	-30.02	-0.12	0.9076
L-104	3,923	1,042.97	4.06	0.0002	L-003	2,620	-40.06	-0.16	0.8769
L-041	4,431	1,253.49	3.81	0.0005	L-051	2,615	-43.94	-0.17	0.8651
C.Primavera11	3,556	737.77	2.87	0.0067	L-139	2,588	-66.87	-0.26	0.7960
L-149	3,536	721.54	2.81	0.0079	L-171	2,515	-127.22	-0.5	0.6232
L-202	3,519	706.98	2.75	0.0091	L-123	2,502	-137.68	-0.54	0.5951
L-201	3,832	827.67	2.52	0.0163	L-029	2,459	-173.67	-0.68	0.5031
L-058	3,371	584.31	2.28	0.0288	L-006	2,416	-209.55	-0.82	0.4197
L-085	3,292	518.88	2.02	0.0506	L-154	2,406	-217.57	-0.85	0.4023
L-102	3,288	514.99	2.01	0.0523	L-071	2,277	-278.12	-0.85	0.4031
L-061	3,281	509.10	1.98	0.0549	L-143	2,396	-226.27	-0.88	0.3839
L-145	3,276	505.16	1.97	0.0567	L-174	2,231	-311.09	-0.95	0.3502
L-153	3,267	498.20	1.94	0.0600	L-022	2,193	-337.79	-1.03	0.3109
L-049	3,208	448.79	1.75	0.0888	L-124	2,034	-450.82	-1.37	0.1786
L-103	3,205	446.59	1.74	0.0903	L-193	2,236	-359.24	-1.4	0.1702
L-136	3,178	424.16	1.65	0.1071	L-055	2,228	-365.60	-1.42	0.1629
L-090	3,150	400.47	1.56	0.1274	L-050	2,203	-386.95	-1.51	0.1404
L-140	3,090	350.37	1.36	0.1807	L-007	2,187	-399.55	-1.56	0.1282
L-068	3,076	339.00	1.32	0.1949	L-166	2,173	-411.66	-1.6	0.1174
L-016	2,971	251.41	0.98	0.3339	C.Superior 6	2,141	-438.31	-1.71	0.0962
L-115	2,946	230.62	0.9	0.3750	L-183	2,076	-492.40	-1.92	0.0629
L-167	2,924	212.71	0.83	0.4128	L-189	2,028	-531.79	-2.07	0.0454
L-204	3,032	259.02	0.79	0.4359	L-175	2,026	-533.44	-2.08	0.0448
L-048	2,910	201.18	0.78	0.4384	L-047	2,011	-546.06	-2.13	0.0402
C. Achagua 8	2,837	140.08	0.55	0.5887	L-106	1,985	-567.57	-2.21	0.0334
L-217	2,875	147.07	0.45	0.6573	L-180	1,377	-918.42	-2.79	0.0082
L-150	2,789	100.82	0.39	0.6969	L-229	1,380	-916.17	-2.79	0.0084
L-215	2,811	101.94	0.31	0.7583	L-107	1,803	-719.40	-2.8	0.0080
L-082	2,741	60.20	0.23	0.8160	L-216	1,762	-753.15	-2.93	0.0057
L-056	2,695	22.11	0.09	0.9319	L-182	1,637	-856.78	-3.34	0.0019
Soyica P-34	2,658	-8.48	-0.03	0.9738	L-002	1,117	-1102.91	-3.35	0.0018
L-147	2,638	-24.94	-0.1	0.9232	L-078	1,523	-951.85	-3.71	0.0007

BLUP: best linear unbiased predictor; Standardized BLUP: BLUPe=BLUP/prediction error; P>|t|: t-value probability; GY: grain yield (kg ha⁻¹).

The correlation between grain yield (GY) and BLUP predictors was 0.997, and the BLUPe value was 0.985, meaning that phenotypic selection and predicted values are equally effective for selection in the target population. However, there is no doubt that BLUPe is very useful when deciding about the upper lines, mainly when the data are unbalanced, while Tukey's (5%) mean comparison test did not allow the precise discrimination of the lines. Several researchers have shown significant and positive correlations

between predicted genetic and observed yield values (Casaroves & Balzarini, 2002; Milla-Lewis & Isleib, 2005), a highly dependent situation on trait heritability.

Grain yield showed a genetic variance of 381,356 ($P<0.01$), an error variance of 137,426 ($P<0.01$), a heritability value of 84.7%, and 92% efficiency in selecting the best lines, consistent with Falconer (1981), where the selection of lines for their phenotypic values is effective when the heritability is high

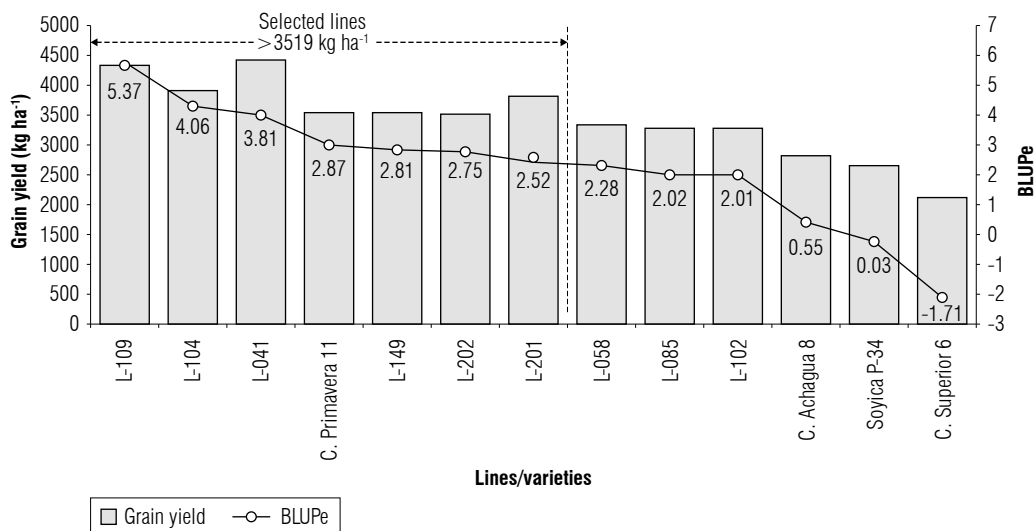


FIGURE 1. BLUPe vs. grain yield of various soybean lines/varieties, Orinoco region, 2020.

In this sense, Souza *et al.* (2000) finds that when heritability is low or very low, the correlation between genetic values predicted by BLUP and phenotypic values is also low. If heritability is moderate, the genetic values obtained by BLUP will allow a better classification of genotypes than the phenotypic values for a more efficient selection.

Generally, the genotypic variance in its optimal environment in soybean is higher than in stress environments. However, Ceccarelli (1989) states that even when the heritability of a specific trait in a stressful environment is always lower than in the optimal environment, this is not clear evidence that selection should be conducted only in optimal environments. Therefore, following the criteria stated by Allen *et al.* (1978) concerning heritability and how to conduct selection, promising lines should be subjected to multi-environment trials to identify the more stable and profitable ones for farmers in the Colombian Orinoco region.

Hydric balance vs. grain yield

During the experimental crop cycle, precipitation was variable in frequency and intensity, reaching a total volume of 618 mm from sowing to harvest, close to the water requirement of the crop in the Orinoco of 350 mm and 550 mm (Almansa, 2006). The hydric balance (Fig. 2) diagram elaborated for soybean in the study site showed a marked variability when excesses and deficits occurred. Although the excesses in the Colombian Orinoco are more noticeable in the first semester of the year and the deficits in the second semester, it is frequent to observe marked variations in the same semester. These variations in water resource availability affect plants according to the moment

of occurrence of the stress and the genotype. In general, the lines similar to the early control C. Superior 6 with DM between 81 and 84 and with a determined or semi-determined GH were highly affected by a water deficit of 13.87 mm between September 15 and 21, and very marked on September 17 and 18 (3.9 and 4.7 mm, respectively), with temperatures above 31°C, coinciding with the reproductive phase of the beginning of pod formation (R3). Almansa (2006) determines that soybean cultivation consumes 4.5 mm of water per day, and an absence of rain for four days makes irrigation necessary. According to Giménez (2014), severe water deficiencies can produce very substantial yield losses (40% or more), mainly when they occur in the critical period of pod formation and filling (Fehr & Caviness, 1977; Sawchik *et al.*, 2013). If the water deficit occurs between R3-R5, it significantly affects the number of grains, and if it does so later, it affects the weight of the grains. Additionally, high temperatures with water deficit during flowering and grain filling cause physiological changes such as stomatal closure that in turn causes premature leaf and flower drop, embryonic abortions, pod drop, and reduced grain yield.

Water excesses in the filling phase and physiological maturity were also decisive when selecting promising lines. In this sense, the control variety C Primavera 11 with a semi-determined GH and an intermediate cycle (102 DM), although it reached a GY of 3,556 kg ha⁻¹, also showed high foliar retention and a non-uniform population that negatively affected harvest and grain quality. Carvajal *et al.* (2017) find that alternate periods of wetting and drying the grains inside the pods result in a marked grain or future seed deterioration. In contrast, the selected lines used in this research had uniform maturity and drying.

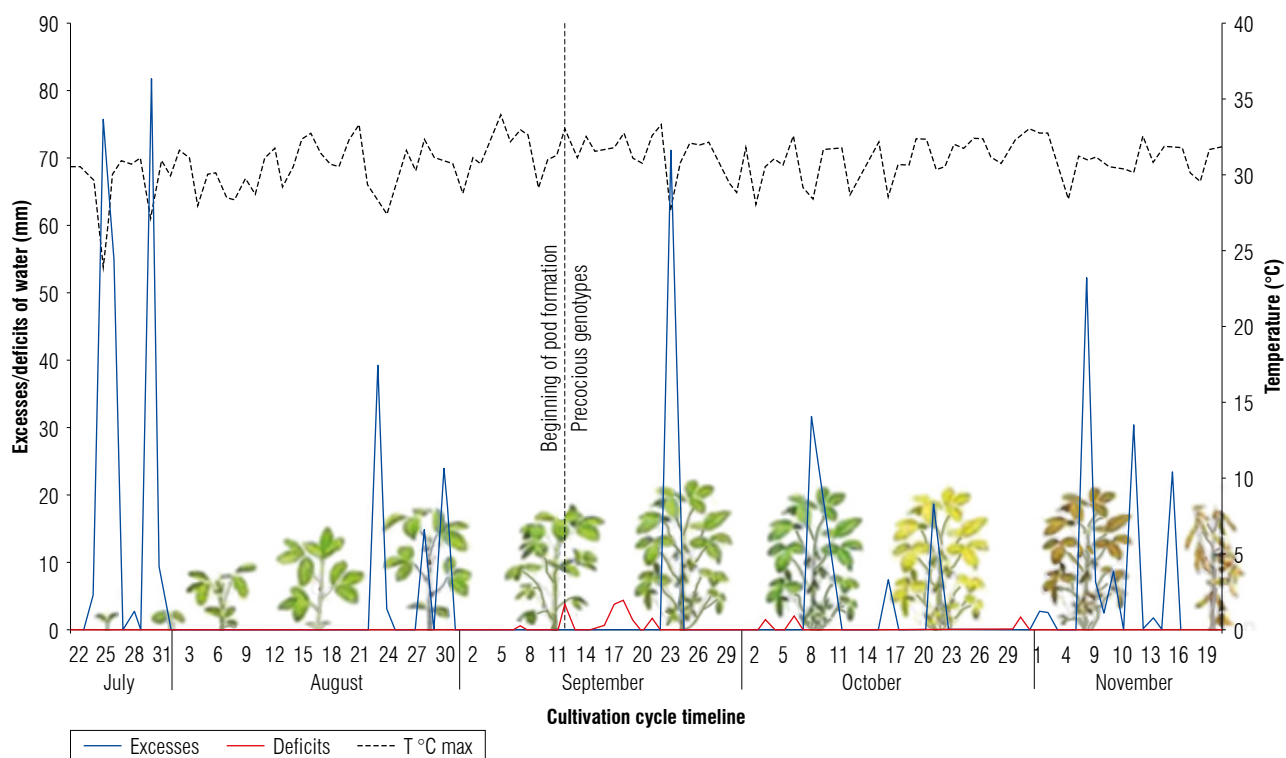


FIGURE 2. Water excesses and deficits, and average temperature during the soybean cultivation cycle. Research Center La Libertad, Orinoquía, 2020.

It is indisputable that the significant differences in GY between lines and varieties are due to the genetics and physiology of the plants in response to local environmental conditions, particularly associated with water and thermal availability. These occur during plant development and differentially influence cultivars according to their early, intermediate, or late cycle, also dependent on the moment of occurrence and intensity of the stress, the genotype, and its photoperiodic sensitivity (Sawchik *et al.*, 2013). Therefore, a timely sowing date, adequate plant population, and water and nutrient availability in the critical phases (R3-R6) will maximize grain production.

It is essential to highlight that in the edaphoclimatic conditions of Orinoquía the selection of very precocious

genotypes can reduce plant growth and final grain yield. In contrast, in the very late genotypes, the maturity of the pods is ostensibly delayed, favoring the incidence of diseases, pests, weed competition, and poor seed quality due to the variable occurrence of rainfall after maturity. The selected lines have as a comparative advantage an intermediate ripening period (88-105 d after emergence), uniform drying, and good grain quality. Some morphoagronomic traits of the promising lines are presented in Table 3.

The subsequent evaluation of these lines in MET should guarantee the genotype-by-environment interaction (GxE) assessment to select the future phenotypically stable variety or varieties with a higher genetic potential to be cultivated in Oxisols of the Orinoquía region. However, although

TABLE 3. Morphoagronomic traits of selected soybean lines from the Research Center La Libertad, Orinoquía, 2020.

Line/Variety	FC	PC	GH	DF	DM	H (cm)	NN
L-019	W	B	SD	53	105	64	16
L-041	P	B	I	36	88	103	18
L-104	W	G	D	39	89	60	13
L-149	P	B	I	38	88	92	18
L-201	P	B	I	37	97	109	22
L-202	W	G	I	37	97	100	20

FC: flower color [(white (W) and purple (P)]; PC: pubescence color [(brown (B) and gray (G)]; GH: growth habit: [determinate (D), semi-determinate (SD), and indeterminate (I)]; DF: days to flowering; DM: days to physiological maturity; H: average plant height at maturity; NN: average number of nodes.

there was broad variability in the population of soybean lines/varieties under study, the genetic base of the genetic breeding program remains narrow. Access to foreign germplasm and the implementation of moderate genotyping and high-performance phenotyping tools are necessary to accelerate the genetic gain of future cultivars for current and potential areas of Colombia through genomic selection.

Conclusions

The application of mixed models using the REML/BLUP procedure allowed the generation of a standardized predictor (BLUPe) useful for the effective classification and selection of potential lines or varieties due to the high genetic merit obtained. With a heritability of 84.7% for GY and a selection pressure of 10%, six lines with positive BLUPe values ≥ 2.52 and grain yields higher than 3,519 kg ha⁻¹ were identified: L-019, L-041, L-104, L-149, L-201, and L-202. These lines have an intermediate period to physiological maturation of 88-105 d after emergence, uniform drying, and good grain quality. These lines will be subjected to multi-environment yield trials to determine their phenotypic stability and select the line or lines with the highest genetic potential for ICA registration as a variety or varieties for the Colombian Orinoco.

Acknowledgments

The authors would like to thank the Ministerio de Agricultura y Desarrollo Rural de Colombia (MADR) for financing this research work in the framework of the project “Genetic breeding of soybean for Orinoquía” (Code 1001244) developed by the Corporación Colombiana de Investigación Agropecuaria - AGROSAVIA.

Conflict of interest statement

The authors declare that there is no conflict of interest regarding the publication of this article.

Author's contributions

RAVR, leader of the Soybean Orinoquía Project, designed the experiments, performed statistical analysis and interpretation of the results, and contributed to the writing of the manuscript; YSTA carried out planting activities, data collection in the field, and contributed to the writing of the manuscript. All authors have read and approved the final version of the manuscript.

Literature cited

Allen, F. L., Comstock, R. E., & Rasmusson, D. C. (1978). Optimal environments for yield testing. *Crop Science*, 18(5), 747–751. <https://doi.org/10.2135/cropsci1978.0011183X001800050013x>

- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (2006). Evapotranspiración del cultivo: guías para la determinación de los requerimientos de agua de los cultivos. In FAO, *Estudio FAO Riego y Drenaje* (pp. 298). FAO. <http://documentacion.ideam.gov.co/openbiblio/bvirtual/021367/Evapotranspiraciondelcultivo.pdf>
- Almansa, E. F. M. (2006). Manejo de recurso hídrico para el cultivo de la soya en la Orinoquía Colombiana (No. Doc. 21393). <http://hdl.handle.net/20.500.12324/1661>
- Bandera-Fernández, E., & Pérez-Pelea, L. (2018). Los modelos lineales generalizados mixtos. Su aplicación en el mejoramiento de plantas. *Cultivos Tropicales*, 39(1), 127–133.
- Baselga, M., & Blasco, A. (2008). El BLUP y los métodos bayesianos. Aplicación a la mejora genética vegetal. In Junta de Andalucía (Ed.), *La adaptación al ambiente y los estreses abióticos en la mejora vegetal* (pp. 55–81). Lumen Gráfica. https://www.juntadeandalucia.es/export/drupaljda/la%20adaptaci%C3%B3n_baja.pdf
- Biasutti, C. A., & Balzarini, M. (2012). Estimación del comportamiento de híbridos de maíz mediante modelos mixtos. *Agriscientia*, 29(2), 59–68.
- Bueno Filho, J. S. S., & Vencovsky, R. (2000). Alternativas de análise de ensaios em látice no melhoramento vegetal. *Pesquisa Agropecuária Brasileira*, 35(2), 259–296. <https://doi.org/10.1590/S0100-204X2000000200004>
- Carvajal, C., Márquez, M., Gutiérrez, B., González-Vera, A., Arellano, J., & Ávila, M. (2017). Aspectos de fisiología, deterioro y calidad en semilla de soya. *Revista Alcance*, 73, 76–92.
- Casanoves, F., & Balzarini, M. G. (2002). Predictores lineales insesgados (BLUP) en ensayos comparativos de rendimiento. *Journal of Basic & Applied Genetics*, 14(3), 55–62.
- Ceccarelli, S. (1989). Wide adaptation: How wide? *Euphytica*, 40, 197–205. <https://doi.org/10.1007/BF00024512>
- Cruz Valderrama, J. R. (2015). Manejo eficiente del riego en el cultivo de la caña de azúcar en el valle geográfico del río Cauca (No. 631.7 C955m). Centro de Investigación de la Caña de Azúcar de Colombia (CENICAÑA).
- DNP (Departamento Nacional de Planeación). (2014). Documento CONPES No. 3797. Política para el desarrollo integral de la Orinoquía: Altillanura – Fase I. Bogotá, D.C. Departamento Nacional de Planeación. <https://ceo.uniandes.edu.co/images/Documentos/Conpes%20Altillanura%202014.pdf>
- Falconer, D. S. (1981). *Introducción a la genética cuantitativa*. Compañía Editorial Continental, SA. <https://www.iberlibro.com/buscar-libro/titulo/introduccion-genetica-cuantitativa/autor/falconer/>
- FAO. (2000). Manual de captación y aprovechamiento del agua de lluvia, experiencias en América Latina (Serie: Zonas Áridas y Semiáridas No. 13). Organización de las Naciones Unidas para la Agricultura y la Alimentación - FAO, Oficina Regional para América Latina y el Caribe. <https://www.fao.org/3/ai128s/ai128s00.pdf>
- Fehr, W. R., & Caviness, C. E. (1977). *Stages of soybean development*. Cooperative Extension Service, Agriculture and Home Economics Experiment Station. <https://dr.lib.iastate.edu/handle/20.500.12876/90239>

- FENALCE. (2021). *Indicadores cerealistas 2020*. Federación Nacional de Cultivadores de Cereales, Leguminosas y Soya. <https://fenalce.co/estadisticas>
- Giménez, L. (2014). Efecto de las deficiencias hídricas en diferentes etapas de desarrollo sobre el rendimiento de soja. *Agrociencia Uruguay*, 18(1), 53–64.
- Hacker, J. B., & Cuany, R. L. (1997). Genetic variation in seed production and its components in four cultivars of the pasture grass *Setaria sphacelata*. *Euphytica*, 93(3), 271–282. <https://doi.org/10.1023/A:1002932432121>
- Kantolic, A. G., Giménez, P. I., & Fuente, E. B. (2004). Ciclo ontogénico, dinámica del desarrollo y generación del rendimiento y la calidad en soja. In E. H. Satorre, R. L. Benech Arnold, G. A. Slafer, E. B. Fuente, D. J. Miralles, M. E. Otegui, & R. Savin (Eds.), *Producción de granos: Bases funcionales para su manejo* (pp. 167–201). Editorial Facultad de Agronomía.
- Milla-Lewis, S. R., & Isleib, T. G. (2005). Best linear unbiased prediction of breeding values for Tomato Spotted Wilt Virus (TSWV) incidence in Virginia-type peanuts. *Peanut Science*, 32(1), 57–67. [https://doi.org/10.3146/0095-3679\(2005\)32\[57:BLUPOB\]2.0.CO;2](https://doi.org/10.3146/0095-3679(2005)32[57:BLUPOB]2.0.CO;2)
- Mora, F., & Arnhold, E. (2006). Inferencia Bayesiana y metodología de modelos lineales mixtos aplicados al mejoramiento del maíz. *Ciencia e Investigación Agraria*, 33(3), 185–190. <https://doi.org/10.7764/rcia.v33i3.348>
- Pacheco, J. E., Urquijo, J. S., Darghan, A. E., & Rodríguez, L. E. (2020). BLUP (Best Linear Unbiased Predictors) analysis for the selection of superior yellow diploid potato genotypes (*Solanum tuberosum* Group Phureja). *Revista Colombiana de Ciencias Hortícolas*, 14(1), 125–134. <https://doi.org/10.17584/rrech.2020v14i1.8138>
- Panter, D. M., & Allen, F. L. (1995). Using best linear unbiased predictions to enhance breeding for yield in soybean: I. Choosing parents. *Crop Science*, 35(2), 397–405. <https://doi.org/10.2135/cropsci1995.0011183X003500020020x>
- Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3), 545–554. <https://doi.org/10.1093/biomet/58.3.545>
- Patterson, H. D., & Williams, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika*, 63(1), 83–92. <https://doi.org/10.1093/biomet/63.1.83>
- Piepho, H. P., & Möhring, J. (2006). Selection in cultivar trials – Is it ignorable? *Crop Science*, 46(1), 192–201. <https://doi.org/10.2135/cropsci2005.04-0038>
- Piepho, H. P., Möhring, J., Melchinger, A. E., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161, 209–228. <https://doi.org/10.1007/s10681-007-9449-8>
- Resende, M. D. V. (2007). *Matemática e estatística na análise de experimentos e no melhoramento genético*. Embrapa Florestas, Colombo.
- Resende, M. D. V., & Duarte, J. B. (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical, Goiânia*, 37(3), 182–194.
- Resende, M. D. V. (2016). Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology*, 16(4), 330–339. <https://doi.org/10.1590/1984-70332016v16n4a49>
- Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, 6(1), 15–32. <https://doi.org/10.1214/ss/1177011926>
- SAS Institute. (2014). SAS 9.4 Output delivery system: Procedure guide (3th ed.). SAS Institute. https://go.documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/odsproc/titlepage.htm
- Sawchik, J., Gaso, D., & Núñez, A. (2013). Soja: algunos aspectos a considerar para el logro de altos rendimientos. *Revista INIA*, 34, 34–37.
- Searle, S. R., Cassella, G., & McCullouch, C. E. (1992). *Variance components*. John Wiley & Sons. <https://doi.org/10.1002/9780470316856>
- Singh, P., & Bhatia, D. (2017). Incomplete block designs for plant breeding experiments. *Agricultural Research Journal*, 54(4), Article 607. <http://doi.org/10.5958/2395-146X.2017.00119.3>
- Souza, V. A. B., Byrne, D. H., & Taylor, J. F. (2000). Predicted breeding values for nine plant and fruit characteristics of 28 peach genotypes. *Journal of the American Society for Horticultural Science*, 125(4), 460–465. <http://doi.org/10.21273/JASHS.125.4.460>
- Sturion, J. A., & Resende, M. D. V. (2010). Avaliação genética e análise de deviance em um teste desbalanceado de procedência e progênie de *Ilex paraguariensis*. *Pesquisa Florestal Brasileira*, 30(62), 157–160. <https://doi.org/10.4336/2010.pfb.30.62.157>
- Vásquez, A. C., Bustamante, C. Z., Buitrago, A., Baptiste, B., & Echeverri, J. (2019). *El gran libro de la Orinoquia colombiana*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt & GIZ. <http://hdl.handle.net/20.500.11761/35408>
- Volpato, L., Alves, R. S., Teodoro, P. E., Resende, M. D. V., Nascimento, M., Nascimento, A. C. C., Ludke, W. H., Silva, F. L., & Borém, A. (2019). Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. *PLoS ONE*, 14(4), Article e0215315. <https://doi.org/10.1371/journal.pone.0215315>
- Yan, W., & Rajcan, I. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Science*, 42(1), 11–20. <https://doi.org/10.2135/cropsci2002.1100>
- Yan, W., Hunt, L. A., Johnson, P., Stewart, G., & Lu, X. (2002). On-farm strip trials vs. replicated performance trials for cultivar evaluation. *Crop Science*, 42(2), 385–392. <https://doi.org/10.2135/cropsci2002.3850>