https://doi.org/10.18406/2316-1817v17nunico20251956 e-ISSN 2316-1817

Management and edaphoclimatic factors that determine soybean top-performing farmers in Brazil

Victor Delino Barasuol Scarton[1], Ivan Ricardo Carvalho[2], Christiane de Fátima Colet[3], Leonardo Cesar Pradebon[4], Willyan Júnior Adorian Bandeira[5], Jaqueline Piesanti Sangiovo[6], Murilo Vieira Loro[7], José Antonio Gonzalez da Silva[8]

[1] Universidade Regional do Noroeste do Rio Grande do Sul, victorscarton@hotmail.com
[2] Universidade Regional do Noroeste do Rio Grande do Sul, carvalhoirc@gmail.com
[3] Universidade Regional do Noroeste do Rio Grande do Sul, chris_colet@gmail.com
[4] Universidade Federal de Santa Maria, pradebon_prad@gmail.com
[5] Universidade Regional do Noroeste do Rio Grande do Sul, willyan.jab@hotmail.com
[6] Universidade Federal de Santa Maria, jpsangiovo@gmail.com
[7] Universidade Federal de Santa Maria, muriloloro@gmail.com
[8] Universidade Regional do Noroeste do Rio Grande do Sul, gonzalez_silva@gmail.com

Received in: 17/09/2024
Accepted in: 05/05/2025

Abstract

The integration of meteorological, edaphic, and genetic data with robust analyses such as machine learning and factorial regression helps clarify the factors related to high soybean productivity. This study was developed in order to identify the key management and edaphoclimatic factors determining Brazil's top-performing soybean farmers. Data were collected from the Brazilian Soy Strategic Committee (CESB) website, covering 50 farmers from 36 environments between 2014 and 2023. A total of 18 top-performing cultivars were identified, with relative maturity groups ranging from 5.4 to 8.3. Grain yield was analyzed using centered means with partial least squares, followed by linear regression models and t-tests ($p < 0.05$). Reaction norm parameters were estimated via the Finlay-Wilkinson method, stratified by production region. Factorial regression included meteorological, geographic, satellite, and soil variables as predictors. A regression tree algorithm identified the most influential variables, and farmer profiles were grouped using principal component biplots and K-means clustering. Machine learning models proved superior to traditional methods for predicting productivity, offering a strategic tool for agribusiness. Key factors positively associated with yield included mean temperature (around 30°C), relative humidity, longwave and shortwave radiation, high altitude, early sowing, high plant population, elevated soil organic matter, and high cation exchange capacity. Interestingly, yields were higher in soils with magnesium and calcium contents below 13% and 27%, respectively, decreasing beyond those levels. The highest yields (>6 t ha^{-1}) were observed in Rio Grande do Sul, Paraná, São Paulo, and Minas Gerais. Future research should validate these models in low-tech environments and include socio-economic variables.

Keywords: Soybean yield. Crop management. Cultivar maturity. Sowing date.

Introduction

Soy (*Glycine max* L.) is the most cultivated oilseed in the world, with great agricultural, economic, and nutritional relevance (GOMES, 2023). Brazil currently leads global production, with 45.7 million hectares cultivated and an output of 147.6 million tons in the 2023/2024 season, distributed across all regions, predominantly in the Center-West (CONAB, 2024). Its wide adaptation to different environments and multiple uses – including oil extraction and animal feed – contributes to its large-scale cultivation (DE MELLO *et al.*, 2020).

Given its strategic relevance, soybean production must increasingly align with sustainable practices to positively contributes to food security and address the challenges posed by climate change (KANDEL *et al.*, 2024). This includes the adoption of cultivars with high yield potential, the use of efficient management strategies, and the integration of environmental knowledge to reduce resource use and ecological impacts (STRALIOTTO *et al.*, 2022). Sustainable intensification also demands consideration of genotype x environment x management interactions, which are essential for resilient and profitable systems (WINCK *et al.*, 2023).

Cultivar recommendations must account for ecophysiological factors such as temperature, precipitation, altitude, and solar radiation (OLIVOTO *et al.*, 2018). In parallel, sustainable soil use – considering fertility, organic carbon, and physical structure – supports both productivity and environmental conservation. Tools that estimate environmental and genetic deviations, as well as phenotypic stability and adaptability, are valuable for optimizing these systems (PRADEBON *et al.*, 2023a). In this context, the integration of meteorological, edaphic and genetic data with robust analyses such as machine learning and factorial regression are ways of clarifying the factors related to high production performance, which is the main focus and differential of this study.

Among environmental challenges, temperature stands out: ideal development occurs between 10 ºC and 30 ºC, while values outside this range increase physiological stress and limit yield potential (DA SILVA *et al.*, 2020; NEUMAIER *et al.*, 2020). In addition to temperature, variables such as relative humidity, precipitation, soil attributes, and solar radiation influence key physiological processes, including photosynthesis, internode elongation, stem and branch growth, leaf expansion, and biological nitrogen fixation (MARENCO, ANTEZANA-VERA, 2021). Therefore, aligning management strategies and genetic materials with environments that provide at least 840 mm well distributed throughout the crop cycle is crucial for sustaining high productivity under climate variability.

Although much is known about the environmental requirements and management practices for soybean cultivation, uncertainties persist regarding which combinations of management and edaphoclimatic factors most effectively contribute to sustainable grain yields above six tons per hectare under Brazilian diverse agricultural conditions. In this context, this work was developed in order to identify which management and edaphoclimatic factors determine the soybean top-performing farmers in Brazil.

Material and methods

The data were obtained through a public consultation carried out directly on the CESB - Brazilian Soy Strategic Committee website via the website www.cesbrasil.org.br. The 50-soybean top-performing farmers from 2014 to 2023 were located in the following Environments (Figure 1): seven municipalities in Rio Grande do Sul, five in Bahia, four in São Paulo and Goiás, three in Minas Gerais and Paraná (PR), and one in Piauí, Mato Grosso, Rondônia, Tocantins and Mato Grosso do Sul. This makes it possible to represent different geographical positions and agro-climatic characteristics.

Throughout this period, the genetic constitutions that allowed for high yields was: Pioneer 98Y30, BMX Desafio RR, M 5947 IPRO, M 5917 IPRO, NA 5909 RR, BMX ELITE IPRO, BMX ATIVA, DM53I54 RSF IPRO, M 8349 IPRO, M 8210 IPRO, BMX BÔNUS IPRO, BMX OLIMPO IPRO, BRS 284 CV, BMX FOCO IPRO, AS 3610 IPRO, TMG 7062, BMX LANÇA IPRO and BMX ZEUS IPRO (Table 1). The genotypes had relative maturity group (RMG) ranging from 5.4 to 8.3.

Data collection and yield estimation

The methodology for estimating soybean productivity was carried out according to CESB Official Rules (CESB, 2024). The audited area consists of a minimum of 2.5 hectares and a maximum of 10 hectares. From this area, five to seven representative points were chosen, where the number of plants and grains per linear meter was measured. With these attributes the grain yield (in kg per ha) was estimated with 21% correction based on the thousand grain weight of the cultivar.

Variables collected

Categorical variables: category (irrigated or rainfed), municipality, state, latitude, longitude, harvest, cultivar, system fertilization (performed or not carried out), pre-sowing fertilization, fertilizer formulation, base fertilization, top dressing fertilization, dates (sowing, emergence, flowering and harvest) and predecessor crop.

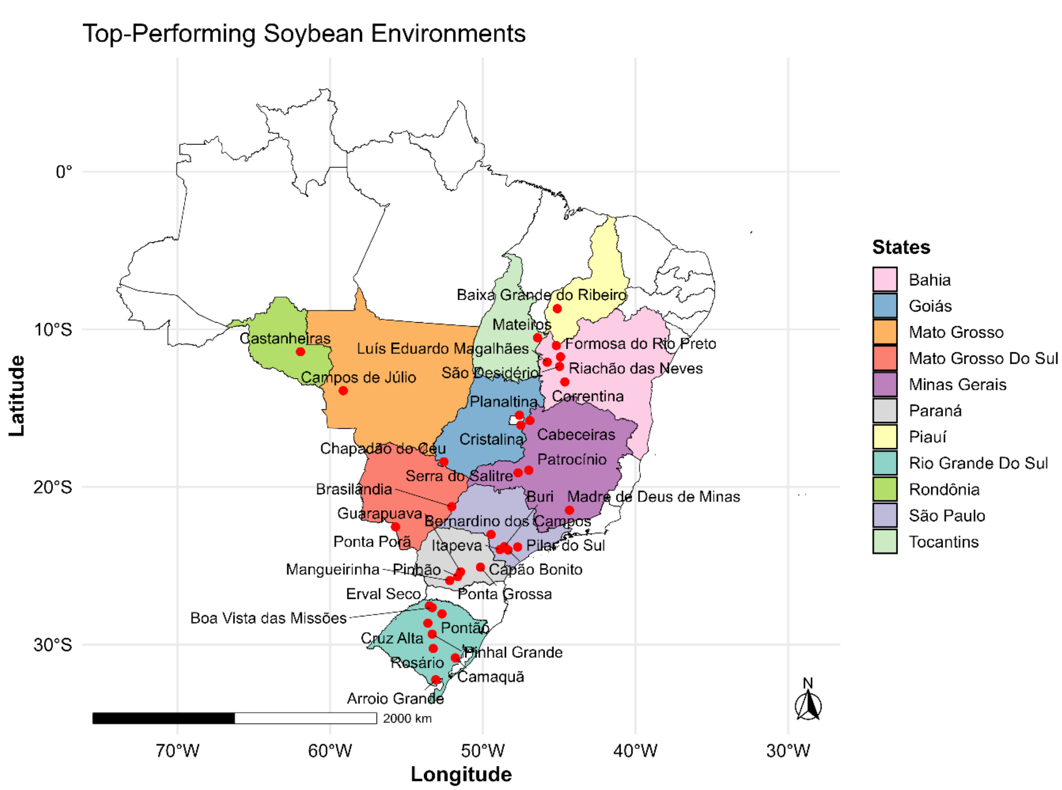
Continuous variables: cycle (days), plant stand (p mL^{-1}), row spacing (cm), sowing depth (cm), clay (%), organic matter (%), cation exchange capacity (CEC, units), pH (units) , aluminum (%), aluminum saturation (m%), base saturation (V%), phosphorus (mg kg^{-1}), potassium (mg kg^{-1}), calcium (mg kg^{-1}), magnesium (mg kg^{-1}), sulfur (mg kg^{-1}), manganese (mg kg^{-1}), iron (mg kg^{-1}), copper (mg kg^{-1}), boron (mg kg^{-1}), number of fungicide applications (units), total cost (R$), soybean price (R$), profit per hectare (R$), harvested area (ha), altitude (m) and grain yield (kg ha^{-1}).

Meteorological and satellite-derived variables: mean, minimum and maximum air temperature (ºC), precipitation (mm), wind speed (m s^{-1}), relative humidity (%), long and shortwave incident radiation (MJ m^2 day^{-1}) coming from the NASA Power platform (NASA POWER, 2024). The edaphic variables were obtained by the SoilGrids platform, and included: clay (CLA, %); cation exchange capacity (CEC, %); nitrogen (N, %); organic matter (OM, %); soil acidity (pH, scale); sand (SAND, %); silt (SILT, %) (SOILGRIDS, 2024).

Statistical analysis

The data obtained were subjected to adjustment of the averages centered by partial least squares (PLS). Subsequently, grain yield

Figure 1. Geographic positioning of production environments in the states of Rio Grande do Sul, Paraná, São Paulo, Mato Grosso do Sul, Minas Gerais, Goiás, Mato Grosso, Rondônia, Bahia and Piauí.



Source: authors (2024).

Table 1. Cultivars used, breeders, technologies, thousand seed weight, cycle and relative maturation group.

CULTIVAR	Obtainer	TEC	TSW	CYCLE	RMG	FLOWERC	VEGC	SS	INTEGC	PHYP	PUBC	PUBDEN	HCOR	PEX	GT	BRI
Pioneer 98Y30	PIONEER	RR	160	140	8.3	White	Dark grey	Spherical	Yellow	Absent	Gray	Medium	Light brown	Positive	Determined	High
BMX Desafio RR	BRASMAX	RR	170	120	7.4	White	Dark grey	Spherical	Yellow	Absent	Gray	Medium	Light brown	Positive	Undetermined	High
Monsoy 5947 IPRO	MONSOY	IPRO	140	139	5.9	Purple	Light gray	Spherical	Yellow	Present	Gray	Medium	Imperfect black	Negative	Undetermined	Low
M5917 IPRO	MONSOY	IPRO	204	150	5.9	Purple	Dark grey	Spherical	Yellow	Present	Gray	Medium	Imperfect black	Negative	Undetermined	Low
NA 5909	NIDERA	RR	150	132	5.9	Purple	Light brown	Spherical	Yellow	Present	Gray	Medium	Imperfect black	Positive	Undetermined	Medium
BMX ELITE IPRO	BRASMAX	IPRO	170	128	5.5	Purple	Dark grey	Flattened spherical	Yellow	Present	Gray	Medium	Light brown	Positive	Undetermined	Low
BMX ATIVA	BRASMAX	RR	174	126	5.6	Purple	Light gray	Spherical	Yellow	Present	Gray	Medium	Imperfect black	Negative	Determined	Low
DM 53i54 RSF IPRO	DON MARIO	IPRO	200	143	5.4	Purple	Dark brown	Flattened spherical	Yellow	Present	Light brown	Medium	Black	Positive	Undetermined	Low
M8349 IPRO	MONSOY	IPRO	190	129	8.3	Purple	Dark grey	Spherical	Yellow	Present	Gray	Medium	Light brown	Negative	Determined	Low
Monsoy 8210 IPRO	MONSOY	IPRO	177	133	8.2	White	Medium brown	Spherical	Yellow	Absent	Brown	Medium	Black	Negative	Determined	Low
BMX BÔNUS IPRO	BRASMAX	IPRO	190	120	7.9	Purple	Light gray	Flattened spherical	Yellow	Present	Gray	Medium	Imperfect black	Negative	Undetermined	Medium
BMX OLIMPO IPRO	BRASMAX	IPRO	181	119	8	White	Dark grey	Flattened spherical	Yellow	Absent	Gray	Low	Light brown	Negative	Undetermined	Medium
BRS 284CV	EMBRAPA	CV	146	123	6.3	Purple	Light gray	Spherical	Yellow	Present	Gray	Medium	Light brown	Positive	Undetermined	Low
BMX FOCO IPRO	BRASMAX	IPRO	176	108	7.2	Purple	Light gray	Flattened spherical	Yellow	Present	Gray	Medium	Light brown	Positive	Undetermined	Low
AS 3610 IPRO	AGROESTE	IPRO	153	134	6.1	Purple	Light gray	Flattened spherical	Yellow	Present	Gray	Medium	Light brown	Negative	Undetermined	Low
TMG 7062	TROPICAL M. E GENÉTICA	IPRO	200	153	6.2	White	Light gray	Spherical	Yellow	Absent	Gray	Medium	Light brown	Positive	Semi determined	Medium
BMX LANÇA IPRO	BRASMAX	IPRO	177	128	5.8	White	Light gray	Spherical	Yellow	Absent	Gray	Medium	Light brown	Positive	Undetermined	Low
BMX ZEUS IPRO	BRASMAX	IPRO	209	137	5.5	White	Dark brown	Spherical	Yellow	Absent	Light brown	Medium	Brown	Negative	Undetermined	Low

TEC - technology, TSW - thousand seed weight (grams), CYCLE - (days), RMG - relative maturation group, VEGC - vegetable color, SS - seed shape, INTEGC - integument color, PHYP- pigmentation of the hypocotyl, PUBC- pubescence color, PUBDEN - pubescence density, HCOR - hypocotyl color, PEX - peroxidase reaction, TP - growth type, BRI - seed brightness.

Source: authors (2024).

was set as a dependent character and the other quantities as independent, in this way linear models of regression and significance were constructed using the t test at 5% probability. The reaction norm parameters were estimated using the Finlay-Wilkinson method, stratified by production regions in Brazil, obtaining stability, predictability and their relationships.

Next, a factorial regression was carried out, where grain yield was fixed as the main one and all meteorological, geographical, satellite and edaphic variables as explanatory of the model, which was stratified for each soybean grower, an analysis of variance was estimated, the sums of squares and multiple angular coefficients.

To identify the most efficient management strategies and cultural practices, an artificial intelligence model was built based on the regression tree algorithm, where the determining variables for grain yield were verified. With the need to identify which producers relate to these selected variables, the biplot main components were used and the general grouping profiles of these soybean farmers were obtained using the K-means algorithm.

All analyzes were carried out in the R software, with packages ExpDes.pt, EnvRtype, agricolae, FW, metan, factoextra, rnaturalearth and neuralnet (R Core Team, 2024).

Results and discussion

The analysis of variance for the factorial regression (Tables 2 and 3) revealed significant effects at a 5% of probability using F-test, for the meteorological variables mean temperature, relative air humidity, longwave radiation, and shortwave radiation in 50 out of the 50 treatments evaluated. Studies by Malosseti *et al*. (2013) and Resende *et al*. (2021) report that the use of environmental information in study analysis provides a precise interpretation and allows predicting the behavior of genotypes through complementary analyzes such as regressions, reaction norms, among others. Corbellini *et al*. (2024) emphasize that the sensitivity of soybeans to temperature directly influences the responsiveness of the genotypes, inflating the variances. This reflects the need to thoroughly evaluate the effect of these environmental covariates on the management and cultivars used in certain regions, in order to maximize soybean yield.

In the estimates and predictions, 50 managements from top-performing farmers producers of the Brazilian Soy Strategic Committee were used, where the grain yield variable was fixed as the dependent variable of the model and the other variables as explanatory of the model. All models presented coefficients of determination greater than 0.7 and significance at 5% probability. This allows a good degree of adjustment of the values to the trend line, allowing good inferences to be made.

It was observed that soybean grain yield had small fluctuations with increasing altitude, with increases of 0.390 kg ha^{-1} per meter of altitude increased (Figure 2). Scarton *et al*. (2023), when evaluating six soybean genotypes in seven environments in Brazil and Paraguay, observed a negative influence of altitude on the expression of soybean grain yield, disagreeing with the present study. However, Gonçalves *et al*. (2020) observed a positive response in soybean productivity in Paraná with increasing altitude. At higher altitudes, lower temperatures – particularly during the night – reduce the energy expenditure in soybean plants, thereby enhancing the availability of net photoassimilates.

The response of the relative maturity group (RMG), according to linear regression (Figure 2), there was a decrease in grain yield, with the increase in RMG, and with each increase in this degree, there was a reduction of 500 kg ha^{-1} in productivity of grains. In this way, a high amplitude is observed, since soybeans are responsive to the photoperiod and the choice of genotypes recommended for the region is essential to obtain high productivity. Mourtzinis *et al*. (2017) conclude that sowing outside the preferred date, combined with the choice of genotypes with inappropriate RMG for the region, significantly reduces soybean grain yield, due to the differential response of genotypes to the photoperiod, as well as lower accumulation of assimilates.

Another key point of high soybean grain yield is the plant population, considered one of the main components of yield (Figure 2). In this scenario, it was found that increasing plant population per hectare responds positively to

Table 2. Analysis of variance and sums of squares of factorial regression.

Management	Variable	DF	SUMS	MS	F. value	Pr (p<0.05)
P17	LON	1	4.77E-20	4.77E-20	4.385304	0.036501
P17	VV	1	1.73E-19	1.73E-19	15.88218	7.23E-05
P17	RAD_C	1	5.39E-20	5.39E-20	4.957379	0.026202
P17	Residual	1000	1.09E-17	1.09E-20		
P18	LON	1	1.64E-20	1.64E-20	4.385304	0.036501
P18	VV	1	5.93E-20	5.93E-20	15.88218	7.23E-05
P18	RAD_C	1	1.85E-20	1.85E-20	4.957379	0.026202
P18	Residual	1000	3.74E-18	3.74E-21		
P19	LON	1	1.32E-20	1.32E-20	4.385304	0.036501
P19	VV	1	4.77E-20	4.77E-20	15.88218	7.23E-05
P19	RAD_C	1	1.49E-20	1.49E-20	4.957379	0.026202
P19	Residual	1000	3.01E-18	3.01E-21		
P20	LON	1	1.07E-20	1.07E-20	4.385304	0.036501
P20	VV	1	3.88E-20	3.88E-20	15.88218	7.23E-05
P20	RAD_C	1	1.21E-20	1.21E-20	4.957379	0.026202
P20	Residual	1000	2.44E-18	2.44E-21		
P21	LON	1	1.59E-20	1.59E-20	4.385304	0.036501
P21	VV	1	5.76E-20	5.76E-20	15.88218	7.23E-05
P21	RAD_C	1	1.8E-20	1.8E-20	4.957379	0.026202
P21	Residual	1000	3.63E-18	3.63E-21		
P22	LON	1	5.39E-21	5.39E-21	4.385304	0.036501
P22	VV	1	1.95E-20	1.95E-20	15.88218	7.23E-05
P22	RAD_C	1	6.09E-21	6.09E-21	4.957379	0.026202
P22	Residual	1000	1.23E-18	1.23E-21		
P23	LON	1	2.44E-20	2.44E-20	4.385304	0.036501
P23	VV	1	8.84E-20	8.84E-20	15.88218	7.23E-05
P23	RAD_C	1	2.76E-20	2.76E-20	4.957379	0.026202
P23	Residual	1000	5.57E-18	5.57E-21		
P24	LON	1	3.69E-20	3.69E-20	4.385304	0.036501
P24	VV	1	1.33E-19	1.33E-19	15.88218	7.23E-05
P24	RAD_C	1	4.17E-20	4.17E-20	4.957379	0.026202
P24	Residual	1000	8.4E-18	8.4E-21		
P2	Tmed	1	5.12E-22	5.12E-22	6.377968	0.011691
P2	UR	1	1.62E-21	1.62E-21	20.14072	7.93E-06
P2	RAD_L	1	4.17E-22	4.17E-22	5.193294	0.022861
P2	RAD_C	1	5.37E-22	5.37E-22	6.695869	0.009788
P2	Residual	1127	9.05E-20	8.03E-23		

DF - degrees of freedom, SUMS - sum of squares, MS - mean square, F. value - f test, Pr - probability, LON - longitude, VV - wind speed (m s^{-1}), RAD_C - shortwave radiation (MJ m^2 day^{-1}), Tmed - mean temperature (°C) and UR - relative air humidity (%). P 17 (Correntina), P 18 (Rosário), P 19 (Luís Eduardo Magalhães), P 20 (São Desidério), P 21 (Riachão das Neves), P 22 (Baixa Grande do Ribeiro), P 23 (Formosa do Rio Preto), P 24 (Castanheiras) and P 2 (Planaltina).

Source: authors (2024).

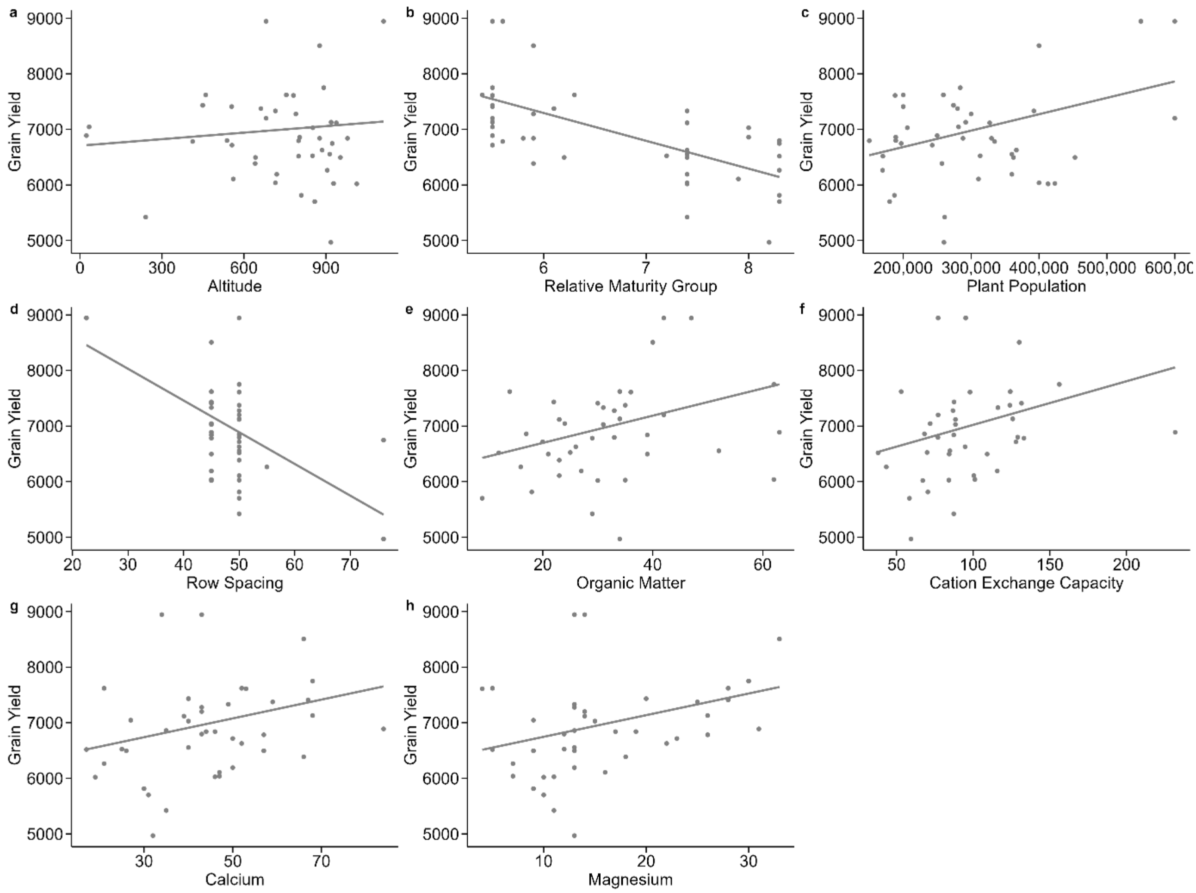
Table 3. Analysis of variance and sums of squares of factorial regression.

Management	Variable	DF	SUMS	MS	F. value	Pr (p<0.05)
P35	Tmed	1	5.33E-20	5.33E-20	6.377968	0.011691
P35	UR	1	1.68E-19	1.68E-19	20.14072	7.93E-06
P35	RAD_L	1	4.34E-20	4.34E-20	5.193294	0.022861
P35	RAD_C	1	5.59E-20	5.59E-20	6.695869	0.009788
P35	Residual	1127	9.41E-18	8.35E-21		
P36	Tmed	1	1.21E-19	1.21E-19	6.377968	0.011691
P36	UR	1	3.83E-19	3.83E-19	20.14072	7.93E-06
P36	RAD_L	1	9.89E-20	9.89E-20	5.193294	0.022861
P36	RAD_C	1	1.27E-19	1.27E-19	6.695869	0.009788
P36	Residual	1127	2.15E-17	1.9E-20		
P37	Tmed	1	4.43E-20	4.43E-20	6.377968	0.011691
P37	UR	1	1.4E-19	1.4E-19	20.14072	7.93E-06
P37	RAD_L	1	3.6E-20	3.6E-20	5.193294	0.022861
P37	RAD_C	1	4.65E-20	4.65E-20	6.695869	0.009788
P37	Residual	1127	7.82E-18	6.94E-21		
P38	Tmed	1	4.05E-20	4.05E-20	6.377968	0.011691
P38	UR	1	1.28E-19	1.28E-19	20.14072	7.93E-06
P38	RAD_L	1	3.3E-20	3.3E-20	5.193294	0.022861
P38	RAD_C	1	4.26E-20	4.26E-20	6.695869	0.009788
P38	Residual	1127	7.16E-18	6.36E-21		
P39	Tmed	1	1.44E-20	1.44E-20	6.377968	0.011691
P39	UR	1	4.54E-20	4.54E-20	20.14072	7.93E-06
P39	RAD_L	1	1.17E-20	1.17E-20	5.193294	0.022861
P39	RAD_C	1	1.51E-20	1.51E-20	6.695869	0.009788
P39	Residual	1127	2.54E-18	2.25E-21		
P40	Tmed	1	5.97E-20	5.97E-20	6.377968	0.011691
P40	UR	1	1.89E-19	1.89E-19	20.14072	7.93E-06
P40	RAD_L	1	4.86E-20	4.86E-20	5.193294	0.022861
P40	RAD_C	1	6.27E-20	6.27E-20	6.695869	0.009788
P40	Residual	1127	1.06E-17	9.36E-21		
P41	Tmed	1	2.26E-21	2.26E-21	6.377968	0.011691
P41	UR	1	7.13E-21	7.13E-21	20.14072	7.93E-06
P41	RAD_L	1	1.84E-21	1.84E-21	5.193294	0.022861
P41	RAD_C	1	2.37E-21	2.37E-21	6.695869	0.009788
P41	Residual	1127	3.99E-19	3.54E-22		
P42	Tmed	1	1.78E-20	1.78E-20	6.377968	0.011691
P42	UR	1	5.63E-20	5.63E-20	20.14072	7.93E-06
P42	RAD_L	1	1.45E-20	1.45E-20	5.193294	0.022861
P42	RAD_C	1	1.87E-20	1.87E-20	6.695869	0.009788
P42	Residual	1127	3.15E-18	2.79E-21		

DF - degrees of freedom, SUMS - sum of squares, MS - mean square, F. value - f test, Pr - probability, RAD_C - shortwave radiation (MJ m^2 day^{-1}), RAD_L - longwave radiation (MJ m^2 day^{-1}), Tmed - mean temperature (°C) and UR - relative humidity (%). P 35 (Ponta Porã), P 36 (Campos de Júlio), P 37 (Campos de Júlio), P 38 (Cabeceiras), P 39 (Rio Verde), P 40 (Campos de Júlio), P 41 (Cristalina) and P 42 (Chapadão do Céu).

Source: authors (2024).

Figure 2. Regression analysis of 50 rural producers with grain yield (kg ha^{-1}) as the dependent variable and independent: graph a (GY= 6706 + 0.390x) variable altitude (m), graph b (10299 – 500x) variable (RMG (relative maturity group), graph c (GY= 6093 + 2.94x) variable plant population (plants ha^{-1}), graph d (GY= 9741 – 57.05x) variable row spacing (cm), graph e (GY= 6205 + 24.52x) variable OM (organic matter), graph f (GY= 6237 + 7.84x) variable CEC (cation exchange capacity), graph g (GY= 6231 + 16.20x) variable Ca (calcium) and graph h (GY= 6356 + 38.95x) variable Mg (magnesium).



Source: authors (2024)

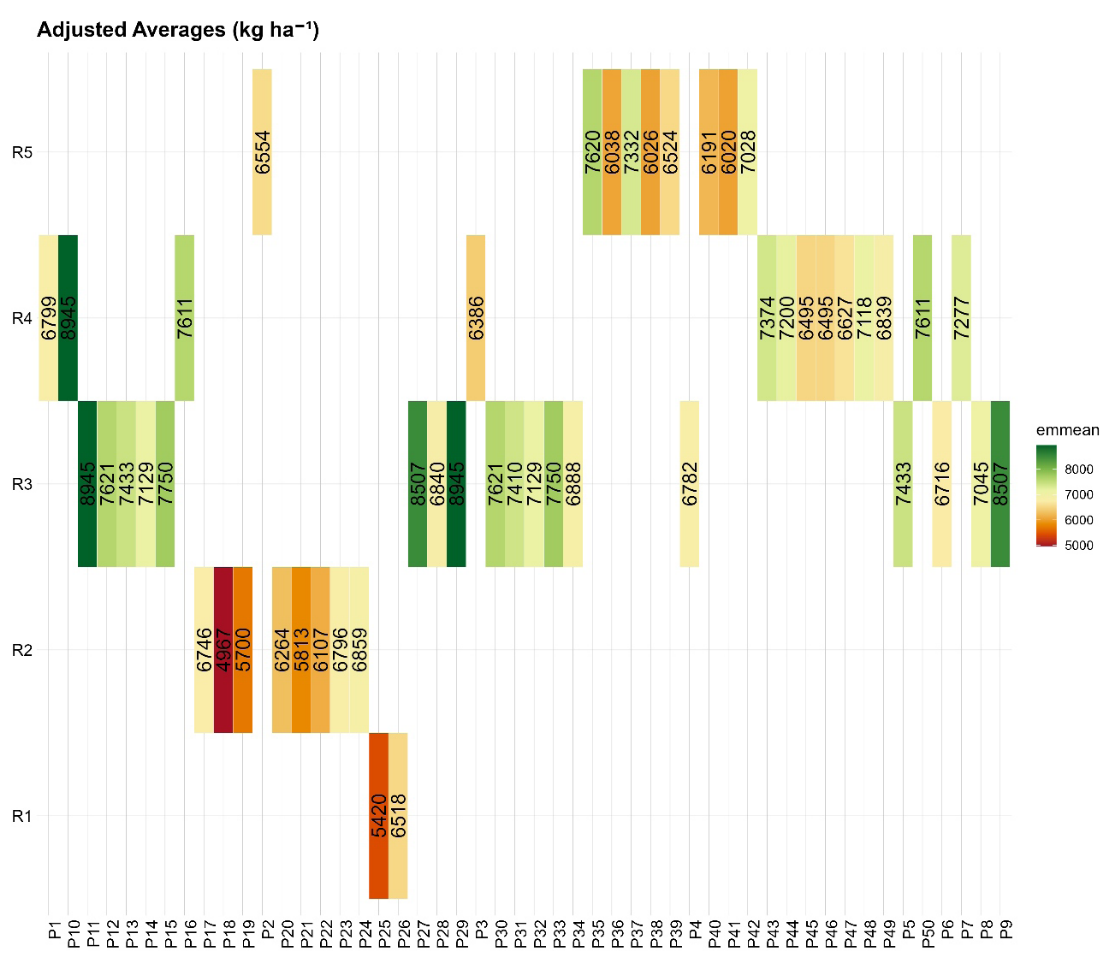
grain yield (GY=6093+2.94x). The optimal adjustment of plant density favors increased effectiveness in interception of solar radiation (FONTANA *et al*., 2020), possibly influencing the growth rate, vegetable fixation and grain yield (FRIGERI *et al*., 2019). Different results were observed by Pradebon *et al*. (2023b), in which the increase in plant density reduced grain yield, as well as the thousand soybean grains weight and the weight of grains per plant, as there is greater intraspecific competition between plants. Another factor linked to plant population is the spacing between soybean rows, it was found that larger spacings end up reducing soybean grain yield, (GY=9741–57.05x). Studies by Cervieri Filho *et al*. (2021) reports that the optimal arrangement would be 30 cm between rows, corroborating that increasing spacing reduces productivity.

Organic matter is considered one of the main attributes resulting from the conservation and improvement of the soil profile. Cultivation environments with higher levels of organic matter in the soil, revealed increases in soybean grain yield at the level of Brazil (GY= 6205 + 24.52x) (Figure 2). Although it represents a small part of the soil weight, soil organic matter controls its quality, as it strongly influences plant productivity, conditions various physical, chemical and physical-water properties of the soil, buffers acidity and is a substrate for biota (SILVA *et al*., 2023).

Another key element in soil quality is the cation exchange capacity. It was inferred that environments with high CEC resulted in productivity greater than six tons of soybean grains (Figure 2), and the increase of each CEC unit results in a gradual increase of 7.84 kg of soybeans (GY=6237.192+7.844x). Environments with high levels of Ca and Mg in the soil show gains in soybean grain yield; these elements are directly related to growth and energy production processes (MALAVOLTA *et al*., 1997), which justifies their importance for high productivity is achieved.

In order to explore the behavior of data and environments, the heatmap was used to highlight the performance of soybeans according to regions and producers' management. It was evident that the region with the lowest soybean grain yield was the northeast region (R2), with productivity between 4,967 and 6,859 kg ha^{-1} of grains (Figure 3). On the other hand, higher yields were observed in the southern region (R3) with values between 6,782 and 8,945 kg ha^{-1} of grains.

Figure 3. Heatmap corresponding to the average grain yield (kg ha^{-1}) of the 50 rural producers evaluated in 36 environments. Each color of the grid represents an average magnitude expression, with red representing low yield, yellow representing average grain yield and green representing high average grain yield. R1- North, R2- Northeast, R3- South, R4- Southeast and R5- Midwest.



Source: authors (2024).

Lobell *et al*. (2009) and Van Ittersum *et al*. (2013), when quantifying the gaps in soybean productivity in Brazil, found that management and water deficit were responsible for approximately 74 and 26% of the difference in soybean grain yield in Brazil. Due to the great climate variability between Brazilian producing regions, as well as sowing times, they lead to a limited understanding of these determining factors in the expression of productivity, both climate and crop management (SENTELHAS *et al*., 2015; EDREIRA *et al*., 2017).

The responsiveness of producer management to the environment indicates which producers performed better than others in each region (Figure 4); which the responsiveness of producers was positive for the southern (3.93), northeast (1.06) and southeast (0.84) regions. Producers P 20 and P 46 obtained the best engagement in the environments, located in São Desidério-BA and Serra do Salitre-MG. Both data refer to the 2017/18 harvest, with sowing in November and a productivity above 6,000 kg ha^{-1}.

The regions with a negative deviation were the north (-1.87) and central-west (-1.26), a condition in which the producers with the lowest indices are located in Brasilândia-MG, Buri-SP, Pontão-RS, Formosa do Rio Preto- BA, Campos de Júlio-MT and Chapadão do Céu-GO, corresponding to producers P 1, P 10, P 12, P 23, P 28, P 40, P 42, P 44 and P 9. These results demonstrate and corroborate the previous ones that climatic conditions are the main limiting factor for yield gaps, and soybean yield is negatively affected in all crop management conditions (BATTISTI *et al*., 2018).

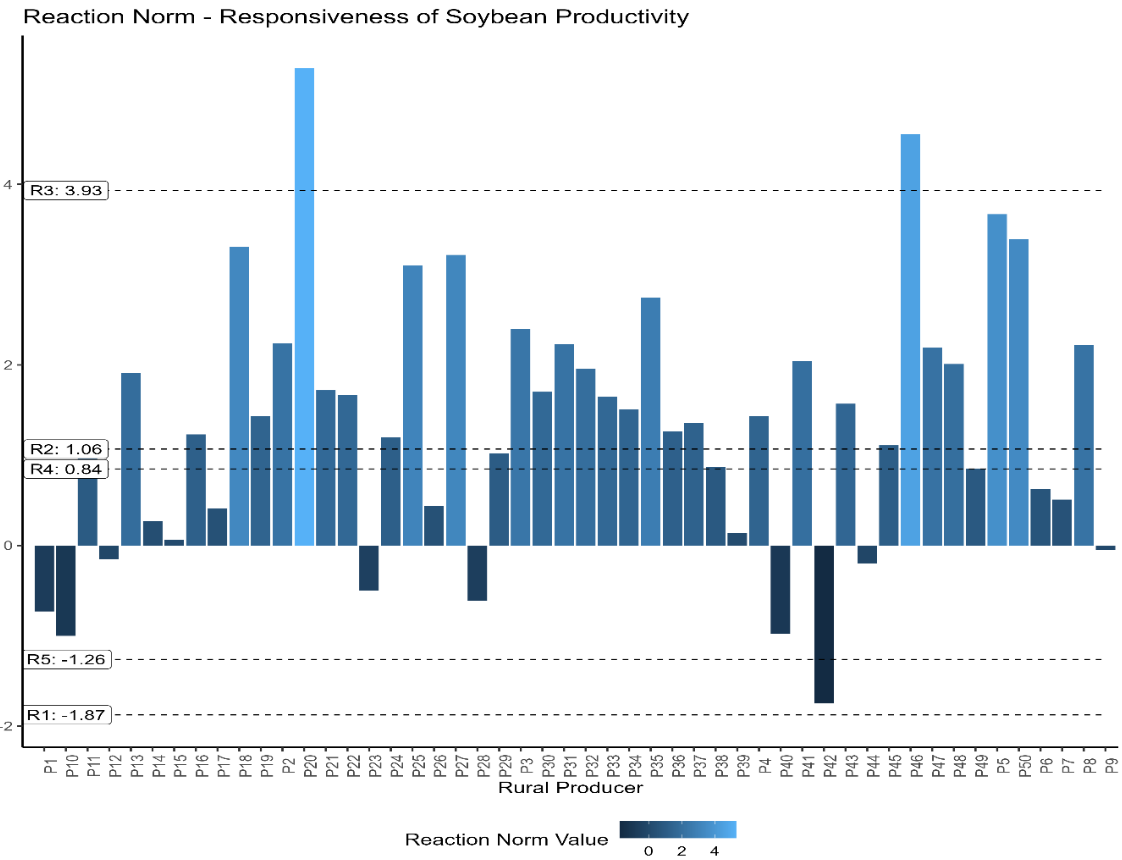
A pronounced effect of longitude, wind speed and shortwave radiation was observed on the first group, formed by producers P 17, P 18, P 19, P 20, P 21, P 22, P 23 and P 24 (Figure 5). These are part of Region 2: Correntina-BA, Rosário-BA, Luís Eduardo Magalhães-BA, São Desidério-BA, Riachão das Neves-BA, Baixa Grande do Ribeiro-PI and Formosa do Rio Preto-BA. The variables analyzed oscillated as follows, longitude had an amplitude of -84.77 O to -13.20 O, wind speed (0.66 m s^{-1} and 12.57 m s^{-1}) and shortwave radiation (3.35 MJ m^{-2} day^{-1} and 33.1 MJ m^{-2} day^{-1}).

The second group had a great influence on mean temperature, relative humidity, longwave radiation and shortwave radiation. The group was composed of producers P 2, P 35, P 36, P 37, P 38, P 39, P 40, P 41 and P 42 from Region 5: Planaltina-GO, Ponta Porã-MS, Campos de Júlio-MT, Cabeceiras-GO, Cristalina-GO, Chapadão do Céu-GO. The meteorological variables varied as follows, mean temperature (20.48°C to 32.54°C), relative humidity (24.5% to 97.12%), longwave radiation (336.29 MJ m^{-2} day^{-1} to 450.35 MJ m^{-2} day^{-1}) and shortwave radiation (3.65 MJ m^{-2} day^{-1} to 31.72 MJ m^{-2} day^{-1}). For Silva *et al*. (2020), temperatures below 10°C and above 40°C can cause losses in soybean productivity, with the ideal mean temperature being around 30°C.

The decomposition of variance by mean squares (Figure 5) enables the identification of the variable with the greatest influence on grain yield for each specific producer. Genotypes located below the abscissa represent native regression, that is, greater sensitivity to the environmental variable (LORO *et al*., 2022). On the other hand, when management was found to be above the abscissa, it indicates positive slope coefficients, that is, the management response tended to be higher than the average, as there is an increase in the environmental variable. In this logic, management that promotes low negative sensitivity to environmental variables is sought.

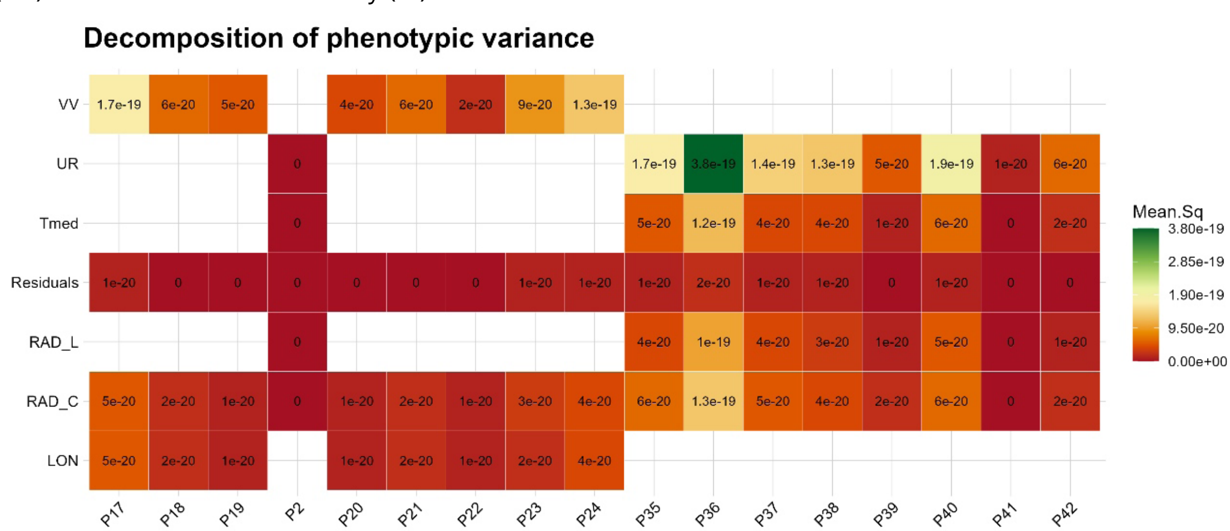
For example, the longitude variable (Figure 6), it was observed that the P 35 management presented high responsiveness, that is, a positive slope trend, with the increase in longitude, relative air humidity, longwave radiation and shortwave radiation, would be

Figure 4. Reaction norm of average genetic values of grain yield (kg ha^{-1}) and slope of responsiveness genetics for 50 rural producers evaluated in 5 regions. R1- region 1, R2- region 2, R3- region 3, R4- region 4, R5- region 5 and PR- producers.



Source: authors (2024).

Figure 5. Decomposition of phenotypic variance by mean squares. LON- longitude, VV - wind speed (m s^{-1}), RAD_C- shortwave radiation (MJ m^2 day^{-1}), RAD_L- longwave radiation (MJ m^2 day^{-1}), Tmed- mean temperature (° C) and UR- relative air humidity (%).



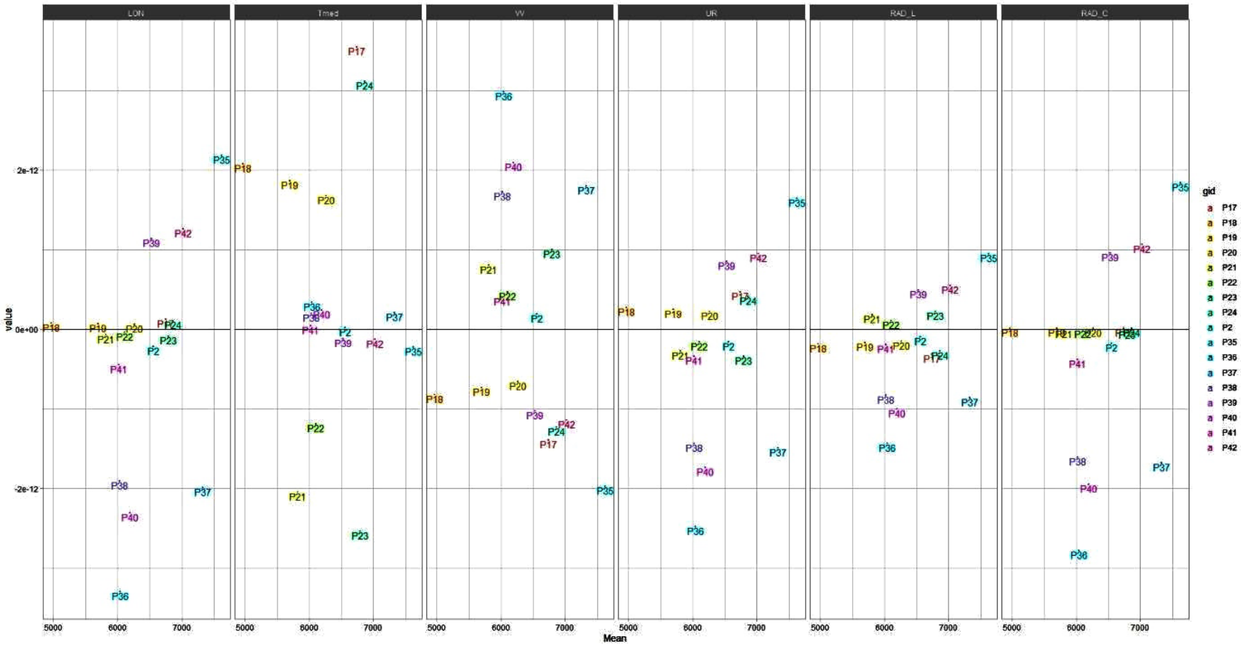
Source: authors (2024).

beneficial for increasing productivity for this management. However, at shorter lengths P 37 would be more responsive. In relation to mean air temperature, P 37 management is positively influenced by higher mean temperatures and wind speed. Studies by Scarton *et al*. (2023), when positioning organic soybean genotypes in Brazil and Paraguay, found a different response from the genotypes depending on temperature variation. These results corroborate those of the present study, in which the genotypes end up responding differently to edaphoclimatic fluctuations between the producing regions.

In this way, it is possible to determine specific cultivars and management for the regions. These results lead to a consistent recommendation for management and positioning of cultivars within the most diverse producing regions. This is of interest to producers, as well as genetic improvement programs, as it allows the identification of specific genotype adaptations in the target environmental population.

In the regression tree of edaphoclimatic elements on grain yield (Figure 7), decision-making starts from organic matter, with more than 39% showing grain yield greater than 7,820 kg ha^{-1}. In this scenario, if the color of the soybean flower is white, the productivity is 6,995 kg ha^{-1}, and if the color is purple, the productivity is 8,508 kg ha^{-1}. For organic matter contents lower than 39%, decision-making is based on the cycle that forms another branch, genotypes with cycles longer than 133 days produce approximately 7,147 kg ha^{-1} of grains, on the other hand, if the soil has a lower OM equal to 23% productivity drops to 6,830 kg ha^{-1} of grains. However, if it is higher than 23% OM, and associated with a CEC higher than 97.9%, it tends to produce 7,471 kg ha^{-1} of grains, now if it is lower than this value, productivity drops to 7,204 kg ha^{-1} of grains. Ramos et al. (2018) also reported a positive relationship between OM and CEC, on both poor and fertile soils. This explains its collinearity and relationship with higher grain yields.

Figure 6. Response capacity stratified by meteorological variables: LON - longitude, VV - wind speed (m s^{-1}), RAD_C - shortwave radiation (MJ m^{-2} day^{-1}), RAD_L - longwave radiation (MJ m^{-2} day^{-1}), Tmed - mean temperature (°C) and UR - relative humidity (%) for 50 rural producers evaluated in 36 growing environments.



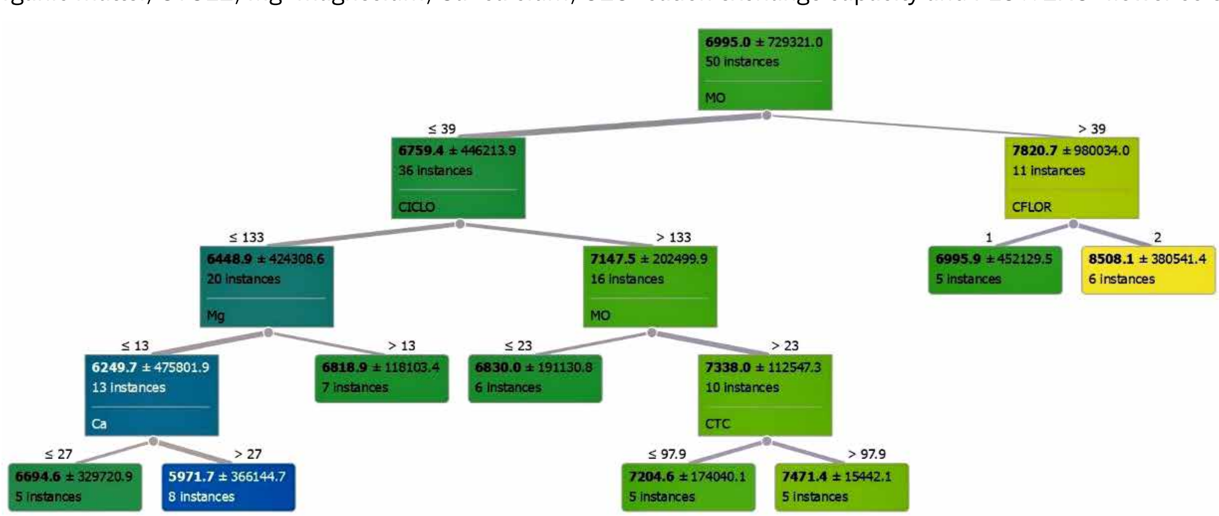
Source: authors (2024).

For cycles shorter than 133 days, the grain yield is 6,448.9 kg ha^{-1}. In soils with Mg content greater than 13%, the yield tends to be 6,818.9 kg ha^{-1}. However, for Mg values higher than 13%, the potential grain yield decreases to 6,249.7 kg ha^{-1}. In soils with Ca levels lower than 27%, the estimated grain yield is 6,694.6 kg ha^{-1}, while for values higher than 27%, the estimated yield drops to 5,971.7 kg ha^{-1}.

The Biplot principal component analysis (Figure 8) demonstrates a total of 75.6% of the graphical analysis and determines the relationship between producers and soil characteristics and cultivars used and their influence on grain yield. The choice of cultivar was decisive for grain yield in the case of producer P 7, which is the cultivar M 5947 IPRO with purple flower color. The relationship between cycle and grain yield was notable for producers P 4, P 9, P 12, P 27, P 30 and P 43 (southern and southeastern regions), with total cycles of 130, 132, 143, 132, 143 and 134 days, respectively. Regarding soil characteristics, CEC and Ca content were directly related to grain yield for producers 15, 31 and 33% (southern region), with values of 156.2, 131.5 and 156.2 for CEC and 68, 67 and 68% for Ca content, respectively.

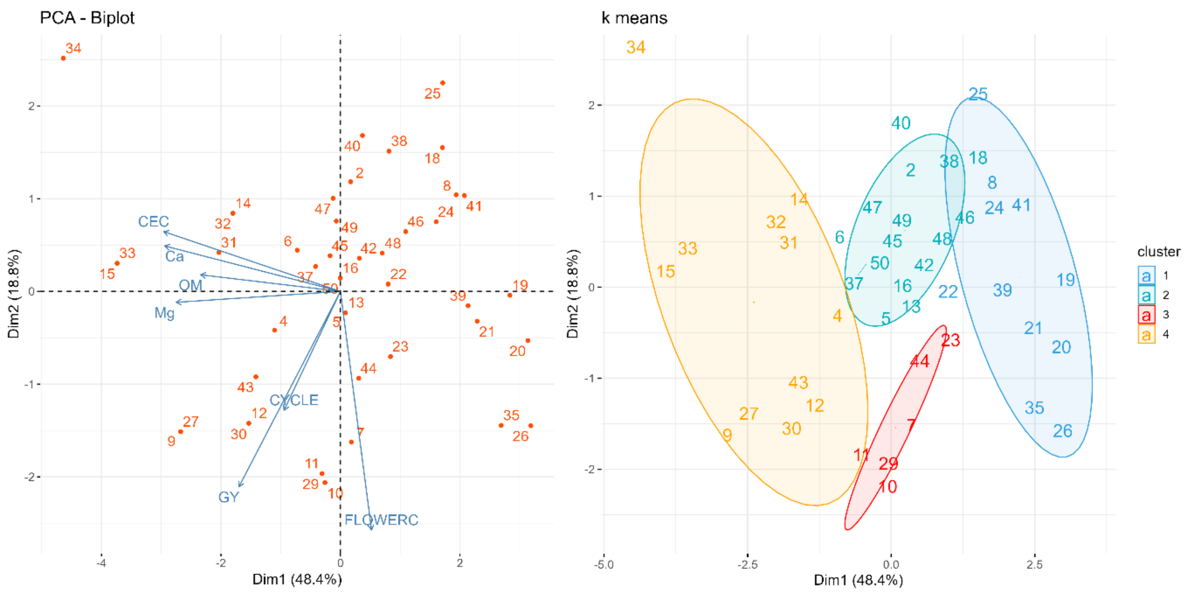
K-means clustering represents a total of 69.8% of the graphical analysis and allowed the organization of producers into four distinct groups (Figure 8). Group 1 was composed of producers P 25, P 18, P 8, P 24, P 41, P 39, P 19, P 21, P 20, P 35, and P 26, the vast majority are in the north and northeast of Brazil, using INTACTA RR2 cultivars and without the use of irrigation. Group 2 was composed of producers P 2, P 5, P 6, P 13, P 16, P 37, P 38, P 40, P 42, P 45 and P 46, all of whom achieved grain yields above 6,000 ha^{-1} and used cultivars with white flowers, with a large proportion having medium pubescence density and an indeterminate cycle. Group 3 was composed of producers P 7, P 10, P 11, P 23, P 29 and P 44, the majority of producers are located in the southeast and southern regions of Brazil, with grain yield of 6,780 kg ha^{-1}, purple flower color, gray pubescence color and low seed luster. Group 4 was composed of producers P 4, P 9, P 12, P 14, P 15, P 27, P 30, P 31, P 32, P 33 and P 43, with the exception of producer P 43, located in southeastern Brazil, all are producers from the southern region, where the majority produces rainfed crops and uses cultivars with an indeterminate cycle.

Figure 7. Decision-making tree, where grain yield (kg ha^{-1}) is the dependent variable that represents the tree trunk, indicating the beginning of decisions. The independent variables OM, CYCLE, Mg, Ca, CEC and FLOWERC represent the leaves of the tree, these are responsible for the decision tendency, as they directly influence this action. OM- organic matter, CYCLE, Mg- magnesium, Ca- calcium, CEC- cation exchange capacity and FLOWERC- flower color.



Source: authors (2024).

Figure 8. PCA and K means Biplot analysis for grain yield (kg ha^{-1}) for 50 rural producers evaluated in 50 growing environments. The numbers from 1 to 50 represent the producer's case, FLOWERC- flower color, CYCLE- days, Mg- Magnesium, Ca- calcium and CEC- cation exchange capacity.



Source: authors (2024).

Considering all the variables analyzed to achieve high productivity, it is essential to take into account geographic aspects, soybean-producing regions of the country, the genotypes used, soil conditions, regional climate, as well as plant arrangement and management practices.

Conclusions

1. The approach with machine learning models has proven to be superior to traditional methods in predicting productivity, offering a strategic tool for agribusiness.

2. Productivity greater than six tons of grains per hectare are obtained at higher altitudes, with cultivars with a lower relative maturity group, sown in high population densities, environments with high soil organic matter and cation exchange capacity.

3. Sowing in October and November, with high concentrations of calcium and magnesium in the soil, combined with mean air temperatures close to 30°C, enhances soybean productivity in Brazil, especially in Rio Grande do Sul, Paraná, São Paulo and Minas Gerais.

References

BATTISTI, R.; SENTELHAS, P. C.; PASCOALINO, J. A. L.; SAKO, H.; DANTAS, J. P. S.; MORAES, M. F. Soybean yield gap in the areas of yield contest in Brazil. **International Journal of Plant Production**, v. 12, p. 159-168, 2018.

CERVIEIRI FILHO, E.; MENEGHELLO, G. E.; VILLELA, F. A.; CARVALHO, I. R.; LAUTENCHLEGER, F.; MOURA, N. B. Soybean performance and canonic inter-relations towards spacing between sowing lines. **Brazillian Journal of Agriculture**, v. 96, n. 1, p. 314-323, 2021.

CESB. **Oficial rules**. 2024. Available in: https://www.cesbrasil.org.br/desafio-soja/regulamento/. Access on: Apr. 30th, 2025.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). **Grain bulletin – April, 2024**. Available in: https://www.conab.gov.br/info-agro/safras/graos. Access on: May 22th, 2024.

CORBELLINI, M.; BOBEK, D. V.; TOLEDO, J. F. F.; FERREIRA, L. U.; SANTANA, D. C.; GILIO, T. A. S.; TEODORO, L. P. R.; TEODOR, P. E.; TARDIN, F. D. Geographical adaptability for optimizing the recommendation of soybean cultivars in the Brazilian Cerrado. **Scientific Reports**, v. 14, n. 13076, p. 1-12, 2024.

DA SILVA, V. T.; GAVA, R.; COTRIM, M. F.; WASSOLOWSKI, C. R.; TEODORO, P. E.; SNYDER, R. L. Irrigation management in soybean cultivation in direct seeding system, on cultural remains of *Brachiaria ruziziensis*. **Research, Society and Development**, v. 9, n. 6, e64963430, 2020.

DE MELLO, E. S.; BRUM, A. L. The soybean production chain and some impacts on the regional development of Rio Grande do Sul. **Brazilian Journal of Development**, v. 6, n. 10, p. 74734-74750, 2020.

EDREIRA, J. I. R.; MOURTZINIS, S.; CONLEY, S. P.; ROTH, A. C.; CIAMPITTI, I. A.; LICHT, M. A.; KANDEL, H.; KYVERYGA, P. M.; LINDSEY, L. E.; MUELLER, D. S.; NAEVE, S. L.; NAFZIGER, E.; SPECHT, J. E.; STANLEY, J.; STATON, M. J.; GASSINI, P. Assessing causes of yield gaps in agricultural areas with diversity in climate and soils. **Agricultural and Forest Meteorology**, v. 247, p. 170–180, 2017.

FONTANA, D. C.; DALMAGO, G. A.; SCHIRMBECK, J.; SCHIRMBECK, L. W.; FERNANDES, J. M. C. Modificações na quantidade e qualidade da radiação solar ao atravessar a atmosfera e interagir com plantas de soja. **Agrometeors,** v. 27, n. 1, p. 101-110, 2020.

FRIGERI, A. R.; LAZARINI, E.; ORIOLI JÚNIOR, V.; BERNARDES, J. V. S. Sowing times and plant population for three soybean cultivars. **Acta Iguazu**, v. 8, n. 4, p. 41-52, 2019.

GLOBAL GRIDDED SOIL INFORMATION (SOILGRIDS). Available in: . Access on: May 28th, 2024.

GOMES, M. R. Evolution and perspectives of economic performance and soy production in the Brazilian and Paraná contexts. **Journal (Re) Definitions of Borders,** v. 1, n. 2, p. 349-360, 2023.

GONÇALVES, S. L.; OLIVEIRA, M. C. N.; FARIAS, J. R. B.; SIBALDELLI, R. N. R. Mathematical equations representing the impacts of climatic factors on soybean productivity in the 2018/2019 crop season in the Parana State, Brazil. **Agrometeoros**, v. 28, n. 026748, p. 1-10, 2020.

KANDEL, G. P.; BAVOROVA, M.; ULLAH, A.; PRADHAN, P. Food security and sustainability trough adaptation to climate change: lessons learned from Nepal. **International Journal of Disaster Risk Reduction**, v. 101, e104279, 2024.

LOBELL, D. B.; CASSMAN, K. G.; FIELD, C. B. Crop yield gaps: their importance, magnitudes, and causes. **Annual Review of Environment and Resources**, v. 34, p. 179–204, 2009.

LORO, M. V.; CARVALHO, I. R.; SILVA, J. A. G.; SFALCIN, I. C.; PRADEBON, L. C. Decomposition of white oat phenotypic variability by environmental covariates. **Brazilian Journal of Agriculture**, v. 97, n. 3, p. 279-302, 2022.

MALAVOLTA, E.; VITTI, G. C.; OLIVEIRA, S. A. **Avaliação do estado nutricional das plantas: princípios e aplicações.** 2 ed. Piracicaba: POTAFOS, 1997. 319p.

MARENCO, R. A.; ANTEZANA-VERA, S. A. Stem growth of multipurpose tree species: net effect of micrometeorological variability assessed by principal component regression. **Acta Amazonica**, v. 51, n. 3, p. 213-222, 2021.

MOURTZINIS, S.; GASPAR, A. P.; NAEVE, S. L.; CONLEY, S. P. Planting date, maturity, and temperature effects on soybean seed yield and composition. **Agronomy Journal**, v. 109, n. 5, p. 2040-2049, 2017.

National aeronautics and space administration-prediction of worldwide energy resources (NASA POWER). Available in: https://power.larc.nasa.gov/. Access on: May 28th, 2024.

NEUMAIER, N.; FARIAS, J. R. B.; NEPOMUCENO, A. L.; MERTZ-HENNING, L. M.; FOLONI, J. S. S.; MORAES, L. A. C.; GONÇALVES, S. L. Ecofisiologia da soja. In: SEIXAS, C. S.; NEUMAIER, N.; BALBINOT JUNIOR, A. A.; KRYZANOWSKI, F. C.; LEITE, R. M. V. B. C. **Tecnologias de produção de soja**. Londrina: Embrapa Soja, 2020. p. 33-54.

OLIVOTO, T.; CARVALHO, I. R.; NARDINO, M.; FERRARI, M.; DE PELEGRIN, A. J.; SZARESKI, V. J.; DE SOUZA, V. Q. Morphological characters and grain yield of simple corn hybrids in different environments in different environments. **Journal of Agroveterinary Sciences**, v. 17, n. 4, p. 462-471, 2018.

PRADEBON, L. C., CARVALHO, I. R., LORO, M. V., PORT, E. D., BONFADA, B., SFALCIN, I. C., e CHALLIOL, M. A. Soybean adaptability and stability analyzes to the organic system through AMMI, GGE Biplot and mixed models methodologies. **Ciência Rural**, v. 53, e20220262, 2023a.

PRADEBON, L. C.; CARVALHO, I. R.; SANGIOVO, J. P.; LORO, M. V.; SCARTON, V. D. B.; PORT, E. D.; MALLMANN, G.; STASIAK, G.; MACIEL, D. G.; LOPES, P. F.; CARIOLI, G. Management tendencies and needs: a joint proposal to maximize soybean grain yield. **Agronomy Science and Biotechnology**, v. 9, p. 1-11, 2023b.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing, 2024.** Available in: https://www.Rproject.org. Access on: May 28th, 2024.

RAMOS, F. T.; DORES, E. F. G. C.; WEBER, O. L. S.; BEBER, D. C.; CAMPELO JÚNIOR, J. H.; MAIA, J. C. S. Soil organic matter doubles the cation exchange capacity of tropical soil under no-till farming in Brazil. **Journal of the Science of Food and Agriculture**, v. 98, n. 9, p. 3595-3602, 2018.

RESENDE, R. T.; PIEPHO, H. P.; ROSA, G. J. M.; SILVA JUNIOR, O. B.; SILVA, F. F.; RESENDE, M. V. D.; GATTAPAGLIA, D. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. **Theoretical and Applied Genetics**, v. 134, p. 95–112, 2021.

SCARTON, V. D. B.; CARVALHO, I. R.; PRADEBON, L. C.; LORO, M. V.; ALBAN, A. A.; CHALLIOL, M. A.; SAUSEN, N. H.; BRAGA, P. M. F.; SFALCIN, I. C. Influence of meteorological variables and geographic factors in the selection of soybean lines. **Revista de Agricultura Neotropical**, v. 10, n. 3, e7246, 2023.

SENTELHAS, P. C.; BATTISTI, R.; CÂMARA, G. M. S.; FARIAS, J. R. B.; HAMPF, A.; NENDEL, C. The soybean yield gap in Brazil-magnitude, causes and possible solutions for a sustainable production. **Journal of Agricultural Science**, v. 153, n. 8, p. 1394–1411. 2015.

SILVA, C. A.; CERRI, C. E. P.; DE ANDRADE, C. A.; MARTIN NETO, L.; BETTIOL, W. Matéria orgânica do solo: ciclo, compartimentos e funções. In: BETTIOL, W.; SILVA, C. A.; CERRI, C. E. P.; MARTIN-NETO, L.; ANDRADE, C. A. **Entendendo a matéria orgânica do solo em ambientes tropical e subtropical.** Brasília, DF: Embrapa, 2023, p. 17-48.

STRALIOTTO, R.; PRADO, R. B.; FERRAZ, R. P. D.; SIMÕES, M. G.; FONTANA, A.; DENARDIN, J. E.; GIONGO, V.; AMARAL, A. J. do; BIANCHI, S. R.; BEDENDO, G. C. Intensificação da agricultura com sustentabilidade. In: TORRES, L. A.; CAMPOS, S. K. **Megatendências da Ciência do Solo 2030**. Brasília, DF: Embrapa, 2022, p. 97-117.

VAN ITTERSUM, M. K.; CASSMAN, K. G.; GRASSINI, P.; WOLF, J.; TITTONELL, P.; HOCHMAN, Z. Yield gap analysis with local to global relevance - A review. **Field Crops Research**, v. 143, p. 4–17. 2013.

WINCK, J. E. M.; TAGLIAPIETRA, E. L.; SCHNEIDER, R. A.; INKLMAN, V. B.; DALLA NORA, M.; SAVEGNAGO, C., PAULA, L. S.; SILVA, M.R.; ZANON, A.J.; STRECK, N. A. Decomposition of the soybean yield gap into environment × genetics × management in southern Brazil. **European Journal of Agronomy**, v. 145, n. 126795, p. 1-10, 2023.