



Assessing genetics, biophysical, and management factors related to soybean seed protein variation in Brazil

María Sol Zelaya Arce^{a,*}, Eduardo Lago Lago Tagliapietra^a, José Eduardo Minussi Winck^b, Alexandre Ferigolo Alves^a, Felipe Schmidt Dalla Porta^a, Tiago Broilo Facco^a, Nereu Augusto Streck^a, Mauricio Fornalski Soares^a, Gregori Da Encarnação Ferrão^c, Daniel Debona^d, Claudio Hideo Martins da Costa^e, Rodrigo Merighi Bega^f, Elizandro Fochesatto^g, Everton Luis Krabbe^h, Alencar Junior Zanon^a

^a Departamento de Fitotecnia, Universidade Federal de Santa Maria, Roraima Avenue1000, Hall 77, Santa Maria, Rio Grande do Sul, Brazil

^b Crops Team – Consulting, Research and Development, Roraima, Avenue1000, Hall 61H, Santa Maria, Rio Grande do Sul, Brazil

^c Centro de Ciências Agrárias e ambientais, Universidade Federal de Maranhão, Chapadinha, Maranhão, Brazil

^d Laboratório de Fitossanidade, Universidade Tecnológica Federal do Paraná - Campus Santa Helena, Santa Helena, Paraná, Brazil

^e Centro de Ciências Agrárias, Universidade Federal de Jataí, Jataí, Goiás, Brazil

^f Centro de Ciências Agrárias, Centro Universitário do Rio Preto, São José do Rio Preto, São Paulo, Brazil

^g Universidade de Alto Vale do Rio do Peixe, Caçador, Santa Catarina, Brazil

^h Embrapa Suínos e Aves, Concórdia, Santa Catarina 89715-899, Brazil

ARTICLE INFO

Keywords:

Genetic
Biophysical factors
Management practices
Seed protein

ABSTRACT

The demand for high-quality soybeans is increasing. The composition of soybean grain can vary with genetics, biophysical, and management factors. In particular, studies on protein concentration are increasing worldwide. The objectives in this study were: (i) to quantify the genetic effects on seed protein concentration and (ii) to identify the biophysical and management factors affecting seed protein concentration in soybean production systems in Brazil. We collected soybean samples and crop management data through surveys in 194 soybean farms in two growing seasons (2018/2019; 2022/2023) across eleven states in Brazil. Seed protein was determined by the Kjeldahl method. Random forest regressions and comparisons between high and low protein fields to identify the main causes of variation in soybean protein concentration were used. Fields with highest protein concentration were observed in older cultivars released in (2011), at lower yields (3082 kg ha^{-1}), late sowing (DOY 313), higher temperatures ($25.6 \text{ }^{\circ}\text{C}^{-1}$) and a lower photothermal coefficient ($0.79 \text{ MJ m}^{-2} \text{ d}^{-1} \text{ }^{\circ}\text{C}^{-1}$). Conversely, low protein concentration was observed in fields with higher yields (4220 kg ha^{-1}), early sowing (DOY 313), lower temperatures ($24.8^{\circ}\text{C}^{-1}$) and a higher photothermal coefficient ($0.84 \text{ MJ m}^{-2} \text{ d}^{-1} \text{ }^{\circ}\text{C}^{-1}$) and newer cultivars released in (2016). The regression tree and random forest explained 58 % of the protein variability, including cultivar (39 %), latitude (12 %) and sowing date (7 %). Cultivar was the most important factor affecting soybean protein concentration, followed by sowing date. The year of cultivar release, breeding company, latitude, temperature, photothermal coefficient and water supply also affected the final concentration of soybean seed protein. The results emphasize the need for breeding programs to evaluate protein concentration in new soybean varieties. Additionally, we now have clear biophysical and management indicators to help achieve higher protein concentrations in soybean crops.

1. Introduction

Soybean [*Glycine max* (L.) Merr.] production provides a base for

global food security as it is the main source of protein used in many food and feed products (Beta and Isaak, 2016; Smáráson et al., 2019; Parisi et al., 2020; Wajid et al., 2020). The economic value of soybean seeds

* Corresponding author.

E-mail address: solzelaya1997@gmail.com (M.S. Zelaya Arce).

¹ ORCID number: 0000-0002-5546-7578

<https://doi.org/10.1016/j.eja.2025.127541>

Received 25 September 2024; Received in revised form 4 February 2025; Accepted 8 February 2025

Available online 15 February 2025

1161-0301/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

depends on their protein and oil content (Hurburgh et al., 1990; Hurburgh, 1994). Brazil is the world largest producer and exporter of soybean, with a production of 143.2 million tons and an average yield of 3.2 Mg ha⁻¹ in 44 million hectares (CONAB, 2025). Predictions indicate that in 2050, China will account for 46 % of the global soybean trade, and its import volume is expected to reach 126 (Mt), being 53 % of its soybean from Brazil and 37 % from the United States. This corresponds to a projected import of 66 Mt, which would account for 40 % of Brazil's current soybean production (Zhao et al., 2021).

Demand for high-quality soybeans is expected to increase in the coming years because of the wide range of products and by-products that will emerge due to population growth and the positive relationship between rising incomes and protein (especially animal protein) intake (Bheemanahalli, 2022; Messina, 2022). It has been projected that protein production will need to increase by 78 % to meet the needs of an expected population of 9.6 billion people in 2050 if everyone consumes the current maximum protein intake (estimated at 103 g/d) (Henchion et al., 2017).

Furthermore, some processing companies in Brazil already offer bonuses for soybeans with protein levels above 34 % (William et al., 2019). Likewise, international export regulations establish minimum requirements for protein and oil content. For example, Brazilian soybeans exported to China may be penalized if they fail to meet these quality standards (Hertsgaard et al., 2019). This suggests that, gradually, the value of soybeans will increasingly be determined by their composition, significantly impacting the processes across the entire soybean supply chain. In this context, there is a need to significantly increase the quantity and quality of oilseed production.

The low protein content in the seeds is frequently observed in

different regions, making it difficult to produce a high-protein meal demanded by international markets, and therefore more studies on protein content in soybean grain need to be conducted worldwide (Mertz-Henning et al., 2017).

Variation in the main components of the seed (i.e. protein, oil) can be influenced by genetic factors (genetics), environmental conditions during the growing season (environment), the interaction genetics × environment, and to a lesser extent by agronomic management practices (management) (Grassini et al., 2021). In Central Argentina, for example, genetics, environment, and management accounted for 70 %, 27 % and 3 % of the variation in protein concentration in soybeans in production fields, respectively (Bosaz et al., 2019). However, the causes of these fluctuations are poorly understood in Brazil where genetics, environment and management practices are different, which constituted the rationale for this study. Therefore, understanding the interaction between genotype, biophysical factors, and management practices is essential to design management strategies or alternatives aimed at increasing protein concentration in soybean seeds grown in Brazil and that can impact global trading of this commodity.

The objectives in this study were: (i) to quantify the genetic effects on seed protein concentration and (ii) to identify the biophysical and management factors affecting seed protein concentration in soybean production systems in Brazil. To achieve these objectives, we collected soybean samples from producers fields with specific information of crop management and characteristics reported during two growing seasons in 11 states of Brazil.

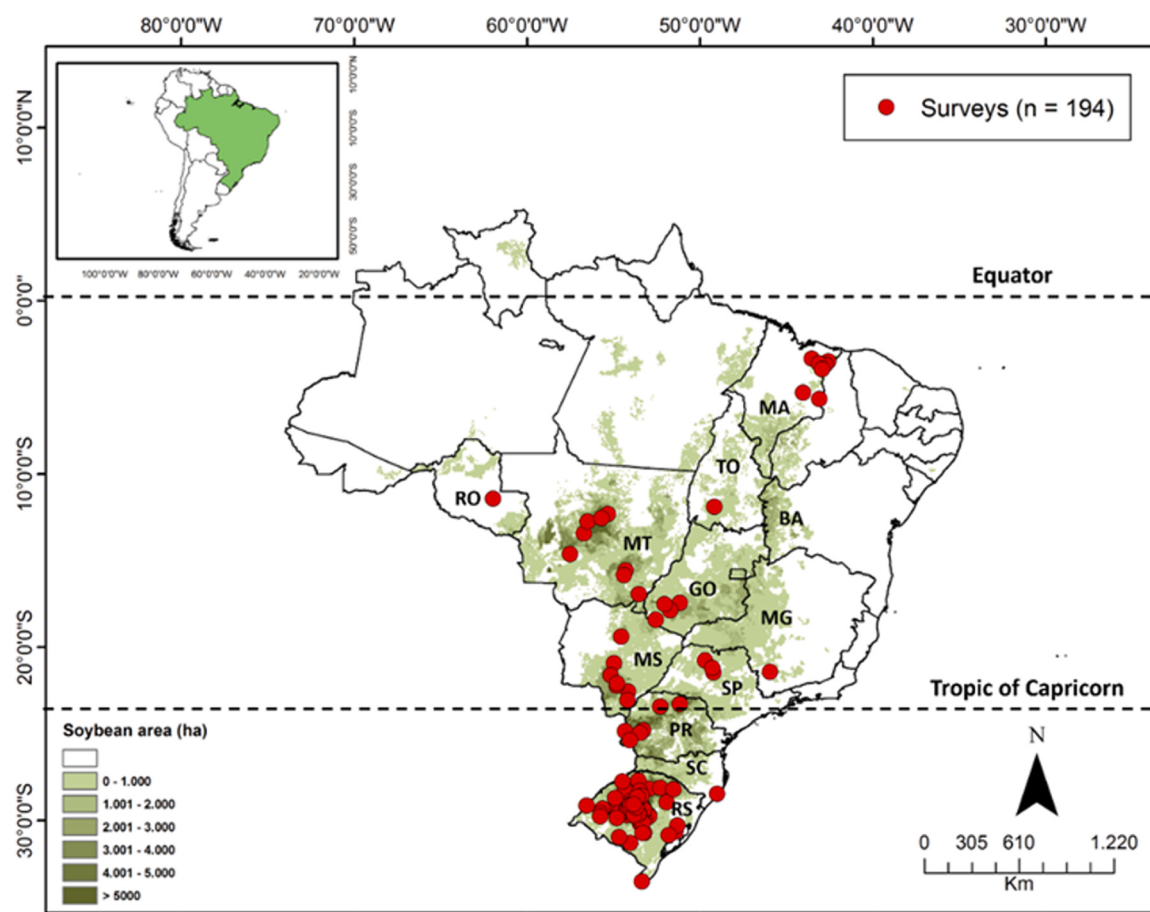


Fig. 1. Location of the surveyed fields (red circles) with soybean harvest area distribution in Brazil (green shaded area). Source: Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística), (2023).

2. Material and methods

2.1. Study region

This study was conducted in the principal soybean-growing regions of Brazil, including 11 states (Rio Grande do Sul, Santa Catarina, Paraná, São Paulo, Mato Grosso do Sul, Mato Grosso, Rondônia, Goiás, Minas Gerais, Tocantins, and Maranhão) and covering a wide range of latitude from 3° S to 33° S (Fig. 1). Further details on climate of each site can be found elsewhere (Alvares et al., 2013).

2.2. Data collection of genetic, biophysical, and management factors

Data of location, crop management practices, and other factors were collected in two growing seasons (2018/2019 and 2022/2023) from each farmer field participating in the Soybean Money Maker Project led by the FieldCrops Team from the Federal University of Santa Maria: sowing date, cultivar name, phosphorus (P) and potassium (K) fertilizer rates, lime application, soil chemical properties, water supply, yield, latitude, longitude, altitude, and others. It was not used in the study of those fields that were severely affected by unexpected adversities such as flooding, extreme drought, etc. The survey data were input into a digital database and incorrect or very incomplete data entries were removed. After quality control, a database containing data from a total of 194 fields (96 % of the total fields surveyed) were used. From a total of 194 surveys applied the sampling size (n) was lower for some variables analyzed due to incomplete responses in the surveys.

For each producer field, daily maximum and minimum temperature, and incoming solar radiation were obtained from the nearest meteorological station of the Brazilian National Institute of Meteorology. Field-specific average mean daily air temperature (°C), average mean daily incoming solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$), and photothermal coefficient ($\text{MJ m}^{-2} \text{d}^{-1} \text{°C}^{-1}$) were calculated during the seed filling phase (from R5-R7) (Fehr and Caviness, 1977). Genetic, biophysical, and management variables used or analyzed in the study are in Table 1.

2.3. Sampling and compositional analysis

The observational unit corresponded to an individual field, each with its crop management data, within a specific growing season and site

Table 1
Variables and management practices of the soybean fields used in the study.

Variables	Type	Units	Explored range
Latitude	Quantitative	Degrees	−3.3 to −33.5
Longitude	Quantitative	Degrees	−42.5 to −61.9
Altitude	Quantitative	m	9–889
Cultivar	Qualitative	Name	
Sowing date	Quantitative	Day of year	272–362
P fertilizer	Quantitative	kg ha ^{−1}	23–198
K fertilizer	Quantitative	kg ha ^{−1}	10–207
Total rainfall	Quantitative	mm	228–1700
Lime application	Quantitative	kg ha ^{−1}	1000–7000
Psoil	Quantitative	Mg dm ^{−3}	1–39
Ksoil	Quantitative	Mg dm ^{−3}	39–366
Ssoil	Quantitative	Mg dm ^{−3}	3.9 – 87.5
pH soil	Quantitative	-	4.4 – 6.6
Yield	Quantitative	kg ha ^{−1}	720–6230
Temperature at R5-R7	Quantitative	°C	22.1 – 27.7
Solar Radiation at R5-R7	Quantitative	MJ m ^{−2} d ^{−1}	16.4 – 25.4
Photothermal coefficient at R5-R7	Quantitative	MJ m ^{−2} d ^{−1} °C ^{−1}	0.65 – 0–97
Water supply	Qualitative	Irrigated/ Rainfed	
Cultivar release year	Quantitative	Year	2008–2023
Breeding company	Qualitative	Name	Brasmax, Don Mario, Monsoy, Nidera, Pionner ^a

^a These breeding companies were the most frequently observed.

(survey). A sample of 1 kg of seed was collected during harvest time from each of the reported fields for subsequent moisture and protein analysis. To do this, it was applied a protocol with three subsamples per field that were collected when roughly 30 %, 55 %, and 80 % of the field was harvested. The soybean samples were cleaned of external matter. Seed protein was determined using the Kjeldahl method (Lynch and Barbano, 1999), in which the percentage of N is converted to a percentage of protein by a factor of 6.25. The concentrations were determined at 13 % seed moisture content.

2.4. Data analysis

In a first step to identify the factors, the protein concentration was divided into an upper tercile (high protein – HP) and a lower tercile (low protein – LP) according to the protein values in the seeds using the Infostat Analysis software (Grassini et al., 2015). The relationship between HP and LP for each quantitative factor such as cultivar release year, sowing date, or P fertilizer was analyzed using the *t*-test or a Wilcoxon test if the distribution of observed values deviated from normality. Outlier data of protein concentrations were not used in the analysis. Variables that were found to be statistically significant in terms of their influence on seed protein when comparing HP and LP fields were further analyzed.

A regression tree analysis was performed to determine the genetic, biophysical and management factors causing the variation in soybean seed protein in fields. The package “caret” in R was used for this purpose. The caret package in R was used to create data balance splits with random sampling and data were split into 70 % training dataset and a 30 % test dataset (default) for model development and evaluation, respectively, for the regression tree analysis. Regression tree analysis is a nonparametric approach that classifies data into successively smaller groups with binary subdivisions based on a single continuous predictor variable (Breiman et al., 1984). Regression tree analysis produces as output a tree diagram with branches determined by the splitting rules and a series of endpoints containing the average response and the number of observations in each end node (Tagliapietra et al., 2021). The method first generates the maximum number of branches in the tree and then applies a cross-validation approach to prune the tree to an ideal size (Therneau and Atkinson, 1997).

For this analysis, we included all qualitative and quantitative variables collected in the farmer survey and we used the 2022/2023 growing season database (n = 70). We did not use all the entire database because of the large number of observations concentrated in South Region (n = 124) in 2018/2019 growing season. After the regression tree analysis, a randomized forest analysis was performed to determine the influence of the genetics (cultivar, breeding company, year of cultivar release) and non-genetic (biophysical and management) factors through the quantification of the relative importance of each factor in the variation of soybean seed protein concentration. Random forest regression analysis was performed using the Random Forest package in R (Liaw and Wiener, 2002), which was based on how much the mean accuracy of the prediction decreases when a variable is excluded. Random Forest is also a nonparametric approach, so there is no requirement of assumptions on probability density distribution.

Linear regression analysis was used to determine the relationships between seed protein concentration and quantitative explanatory variables previously observed to be significantly different in the other analyses. Significance was assumed at *p* = 0.01, *p* = 0.05 and *p* = 0.1 using the GraphPad Prism statistical package (GraphPad Software Inc.). Box-plots were used to summarize the seed protein variations for the qualitative management factors, and the *t*-test (for parametric data), the Wilcoxon test (for non-parametric data) and the Tukey test for multiple comparison of means were used to assess significance.

3. Results

3.1. Relationship between seed protein and yield

Seed protein concentration was negatively correlated with yield (Fig. 2) and showed a decrease in protein concentration of 1.5 % for each Mg ha^{-1} of soybean. Seed yield ranged between 720 kg ha^{-1} to 6230 kg ha^{-1} and protein concentration between 29.8 % and 40.2 %. A consistent trade-off correlation between seed protein concentration and yield was observed across production systems in different regions of Brazil.

3.2. Influence of genetic, biophysical, and management factors on soybean seed protein

The two-sided *t*-test comparison of the field classes with the highest and lowest protein concentration showed a significant influence of genetics and biophysical factors (year of cultivar release, mean air temperature, and photothermal coefficient at stage R5-R7), management practices (sowing date) and yield. In general, fields with the highest protein concentration were observed in older cultivars released in 2011, at lower yields (3082 kg ha^{-1}), late sowing (DOY 313), higher temperatures (25.6°C^{-1}) and a lower photothermal coefficient ($0.79 \text{ MJ m}^{-2} \text{ d}^{-1} \text{ }^\circ\text{C}^{-1}$) (Table 2). Conversely, low protein concentration was observed in fields with higher yields (4220 kg ha^{-1}), early sowing (DOY 302), lower temperatures (24.8°C^{-1}) and a higher photothermal coefficient ($0.84 \text{ MJ m}^{-2} \text{ d}^{-1} \text{ }^\circ\text{C}^{-1}$) and newer cultivars released in (2016).

The regression tree for soybean seed protein showed 4 terminal nodes with a protein concentration between 30.7 % and 33.9 % (Fig. 3a). Of all the genetics and biophysical factors analyzed, as well as management practices, cultivar was the most important variable for seed protein concentration. The field observations were divided into two groups: cultivars with a protein concentration of 33.9 % (right group) and others with a protein concentration lower than 33.9 % (left group). The analysis also identified other factors that explain soybean seed protein variability, such as latitude and sowing date. In general, the higher protein concentrations were observed at higher latitude (≥ -20.2) and late sowing date (≥ 313), similar to what we observed in the first analysis.

The random forest analysis showed that among the levels of relative

importance for soybean seed protein, cultivar accounted for 39 %, breeding company 21 %, cultivar release year 13 %, latitude for 12 %, sowing date for 7 %, and other factors for 8 % (Fig. 3b).

In summary, genetics (cultivar, breeding company, and year of cultivar release), and biophysical and management factors (latitude, sowing date, and others) were responsible for 73 % and 27 %, respectively.

3.3. Changes in soybean protein concentration in cultivars released from 2008 to 2023

A linear regression was estimated between the year of release of the cultivars and the protein, yield, and protein yield of the fields based on information from 184 farmers, 2 growing seasons and 64 cultivars registered between 2008 and 2023. Seed protein concentration showed a negative relationship with the year of cultivar release, decreasing by $0.40 \text{ \% yr}^{-1}$ (Fig. 4a), and yield showed a positive correlation with the year of cultivar release, increasing by $91 \text{ kg ha}^{-1} \text{ yr}^{-1}$ (Fig. 4b). Based on the database for protein concentration and yield of the cultivar in the year of release, we observed an increase in protein yield of $26 \text{ kg ha}^{-1} \text{ yr}^{-1}$ (Fig. 4c).

3.4. Impact of management practices

Protein concentration of soybean seeds showed a positive relationship with the sowing date, with an increase of $0.06 \text{ \% day}^{-1}$ (Fig. 5a). The comparison of fields with different sowing months (October, November, and December) showed a significant effect on soybean seed protein in a decreasing sequence: December (37.1 %) > November (35.9 %) > October (33.6 %) (Fig. 5b). A cumulative protein probability function was performed for the three sowing dates (October, November, December) (Fig. 5c). The probabilities of reaching a protein concentration of 35.3 % are indicated by a vertical line (dashed red). The probability analysis shows that the probability of achieving a protein concentration of 35.3 % is 25 % for sowing in October, while the probability for sowing in November and December is 70 % and 85 % respectively.

The analysis of 70 fields indicated that rainfed fields had on average a 1.07 % higher seed protein concentration compared to irrigated fields (Fig. 6a), while irrigated fields obtained an average protein yield 318 kg ha^{-1} higher than rainfed fields (Fig. 6b).

4. Discussion

In this study, the use of on-farm data allows the identification of the genetic, biophysical and management factors that influence the variation in protein concentration of soybean seeds in Brazil, as well as possible ways to increase protein concentration in soybean fields. A negative correlation ($p < 0.0001$) was found between seed protein concentration and yield across Brazil (Fig. 2). This relationship has already been observed in other countries such as Argentina (Bosaz et al., 2019) and can be explained by the remobilization of C and N from the plant parts into the seeds. With higher yields, the demand for these substances in the seeds increases (Rotundo and Westgate, 2009), which leads to a dilution effect and reduces the protein concentration in the seeds. Also, it was demonstrated that for every one percent increase in protein concentration, the yield will decrease 1 % with the same given amount of photosynthetic energy input (Boote and Tollenaar, 1994).

Studies in Nebraska-USA reported the lack of a trade-off between yield and soybean protein concentration in irrigated environments, indicating that this can also be attainable in other regions and environments (Carciochi et al., 2023). This response is unusual and can be explained by some specific management factors and soil properties unique to that study region. For example, $\text{NO}_3^- \text{-N}$ is commonly present in Nebraska groundwater, and thus the irrigated crops would be

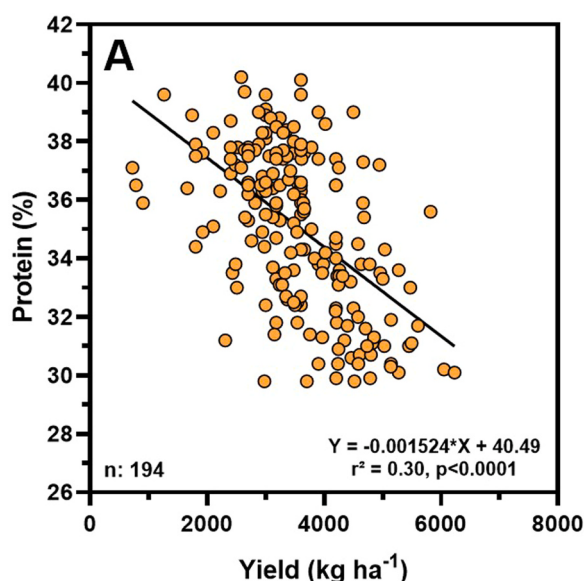


Fig. 2. Relationship between soybean seed protein and yield in Brazil. Each point represents a producer field. Black line represents significant linear regression, and p-value and coefficient of determination (r^2) are also shown.

Table 2
Comparison of genetics, biophysical, and management factors between the protein highest tertile (HP) and the lowest tertile (LP) in soybean fields in Brazil. “n” is the number of observations for each variable analyzed. The values Δ indicate the mean differences (HP – LP) between the upper and lower protein tertile. Means are indicated for each variable in the HP and LP field. Asterisks indicate significance at $p < 0.01$ ***, $p < 0.05$ ** and $p < 0.1$ *.

Genetics, biophysical, and management factors	account for 40 % of Brazil's current soybean	n	High Protein fields (HP)	Low Protein fields (LP)	Δ	P value
Genetics						
Cultivar release year	Year	121	2011	2016	−5	*** < 0.0001
Biophysical						
Latitude	Degrees	46	−18.3	−22.4	4.1	ns 0.1354
Longitude	Degrees	46	−50.6	−52.0	1.4	ns 0.2955
Altitude	m	43	346	390	−44	ns 0.5662
Temperature at R5-R7	°C	43	25.6	24.8	0.8	*** 0.0181
Solar Radiation at R5-R7	MJ m ^{−2} d ^{−1}	43	20.2	20.9	−0.7	ns 0.2692
Photothermal coefficient at R5-R7	MJ m ^{−2} d ^{−1} °C ^{−1}	43	0.79	0.84	−0.1	** 0.0477
Total rainfall	mm	32	831	830	1	ns 0.9996
Lime application	kg ha ^{−1}	32	2661	2624	37	ns 0.9295
Psoil	Mg dm ^{−3}	34	13.8	14.6	−0.8	ns 0.8088
Ksoil	Mg dm ^{−3}	34	113.2	128.1	−14.9	ns 0.5277
Ssoil	Mg dm ^{−3}	34	24.7	24.6	0.1	ns 0.9763
pH soil	-	52	5.5	5.4	0.1	ns 0.3101
Management						
Sowing date	Day of year	115	317	302	15	*** < 0.0001
Total P ₂ O ₅ fertilization	kg ha ^{−1}	121	69	73	−4	ns 0.3941
Total K ₂ O fertilization	kg ha ^{−1}	118	82	89	−7	ns 0.3764
Integrated						
Yield	kg ha ^{−1}	130	3082	4220	−1138	*** < 0.0001
Seed Protein	%	130	38.0	31.8	6.2	*** < 0.0001

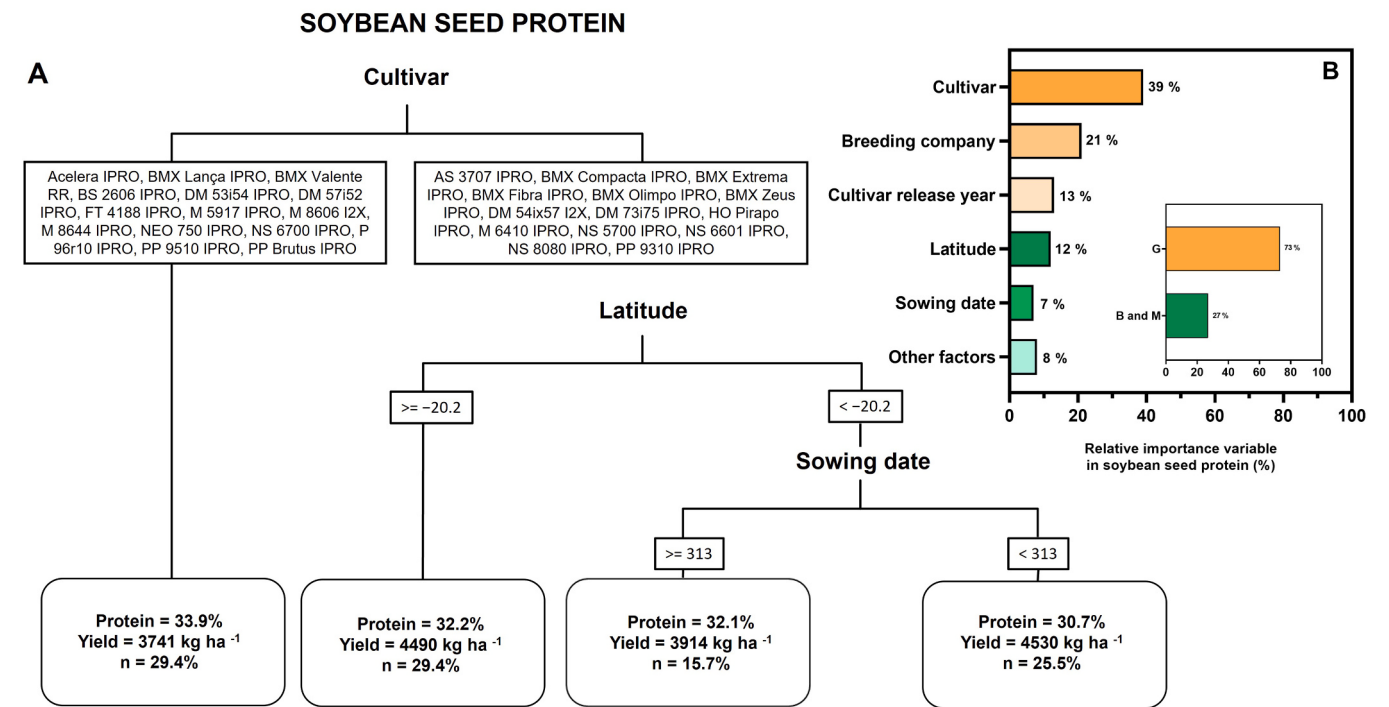


Fig. 3. Regression tree showing sources of variation in soybean seed protein due to genetics, biophysical, and management factors in the 2022/2023 growing season in Brazil. The boxes are splitting nodes, with bottom boxes representing terminal nodes. Values within each terminal node indicate average of seed protein (%) and the percentage of observations in each terminal node (a), relative importance variable ranking for the influence of all factors on variation of soybean seed protein, as determined using random forest regression (b).

supplied with N applied via irrigation water. Also, it was documented that irrigated fields had better soils due to higher soil organic matter concentrations than in rainfed environments.

Attaining high protein concentration and high yields, the total N intake must be increased (Bosaz et al., 2019). Studies have shown that increasing nitrogen supply (> 100 kg N ha^{−1}) increases protein concentration by 2 % and 3 % between planting-V6 and R1- R4, respectively (Rotundo and Westgate, 2009). Similarly, experiments comparing

full-N (to ensure non-N limiting conditions in the entire cycle by temporal additions of N fertilizer: averaged 613 kg N ha^{−1} splitted in V2, V4, R1, R3, and R5 stages) versus zero-N (inorganic soil N at sowing, in-season N mineralization, and N fixation) treatments showed that protein concentration decreased from 41 % to 38 % with increasing yield in zero-N treatments, while the high seed protein concentration and yield was maintained in full-N treatments (Cafaro La Menza et al., 2017). These observed patterns indicate that high yields can be achieved

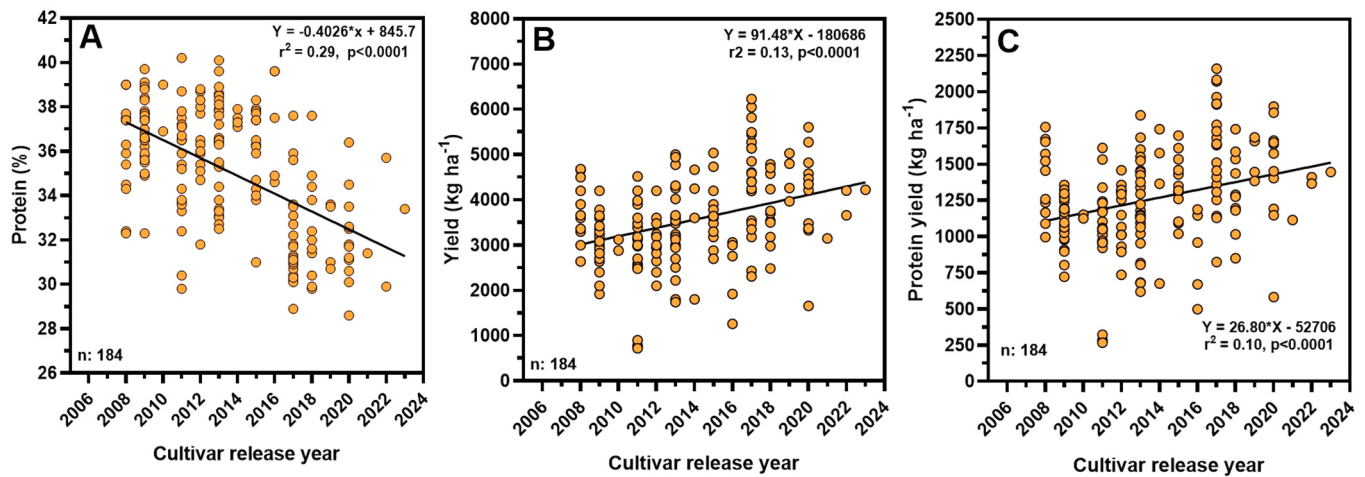


Fig. 4. Changes in soybean seed protein concentration in Brazil during the 2008–2023 period based on database of cultivars release year (a), genetic gain estimated by the yields of the database and cultivars release year (b), relationship between protein yield and cultivars release year (c). Each point represents a producer field. Black lines represent significant linear regression, and p-value and coefficient of determination (r^2) are also shown in each panel.

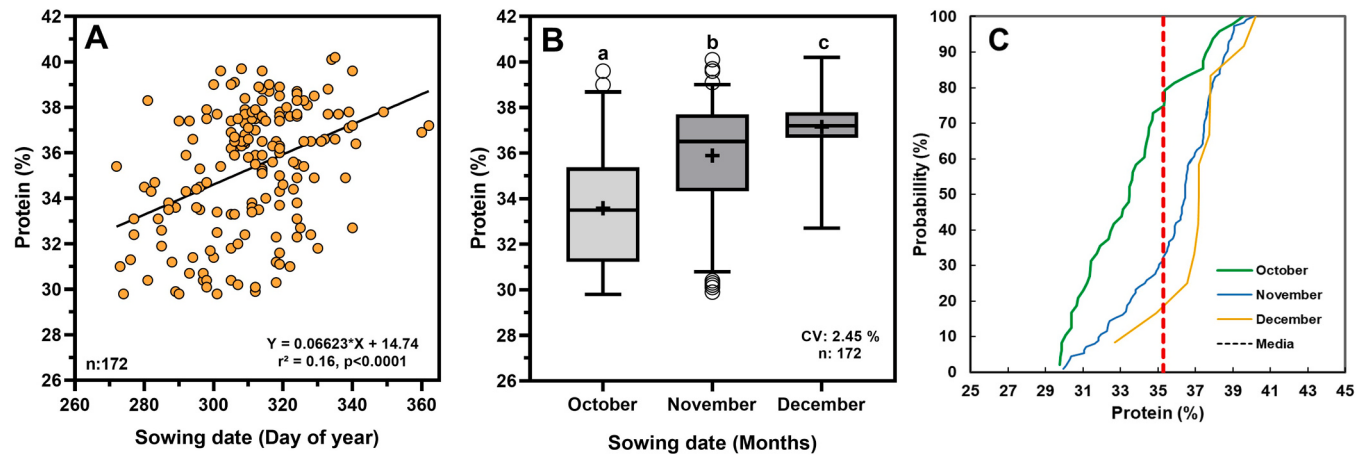


Fig. 5. Relationship between protein and sowing date (a), soybean seed protein in sowing in October, November, and December (b), probability analysis for soybean seed protein of 35.3 % (dashed red line) as a function of sowing date (c) in Brazil. Black line represents significant linear regression, and p-value and coefficient of determination (r^2) are also shown in panel (a). Boxes delimit the 5th and 95th percentiles of the distribution, horizontal line within boxes represents the median and dots are outlier in panel (b). Statistical significance for the protein difference with months of sowing (evaluated using Tukey test) are also shown in panel (b).

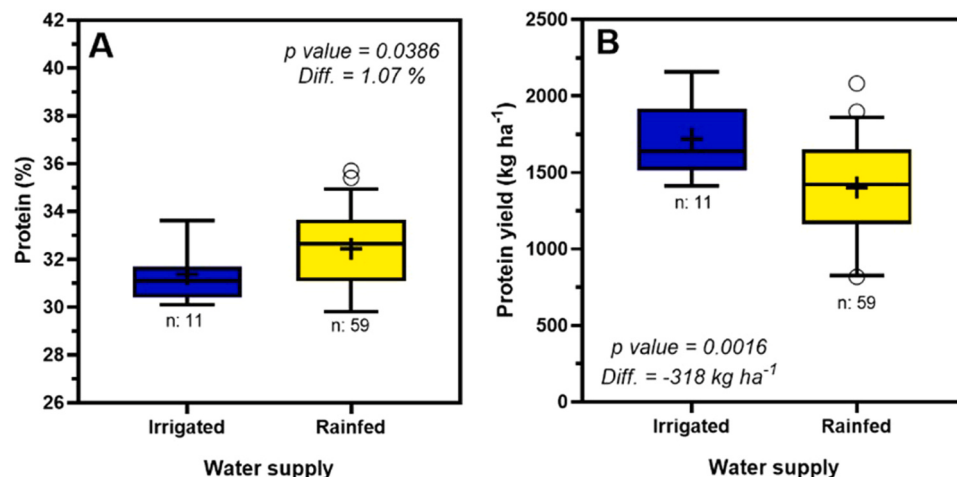


Fig. 6. Soybean seed protein concentration in fields with irrigated and rainfed conditions (a), and protein yield (kg ha^{-1}) in fields with irrigated and rainfed conditions (b). Boxes delimit the 5th and 95th percentiles of the distribution. Horizontal line within boxes represents the median. Dots in panels are outlier. Statistical significance for the protein concentration and protein yield difference (Diff.) between irrigated and rainfed conditions (evaluated using t test) are shown.

when increasing nitrogen supply without decreasing protein concentration. Therefore, the results of this study suggest that there is a gap in nitrogen

uptake in the analyzed soybean crops in Brazil to achieve high yields and maintain high seed protein concentrations in the mature seed.

Other factors that could explain this relationship between seed protein concentration and yield are the traits of each cultivar and the year of release. We have observed that in recent years genotypes express higher yield potential and consequently a decline in protein, which advises plant breeders to maintain both traits in future genotypes.

The regression tree (Fig. 3a) explained 58 % of the variability in protein concentration of soybean seeds. Cultivar was the most important factor explaining 39 % of this variability (Fig. 3b). Similar results were found in the central region of Argentina, where the cultivar accounted for 71.5 % of the variability in protein concentration in soybean seeds (Bosaz et al., 2019). This can be explained by the fact that protein and oil content, although determined by other biophysical or management factors, are complex quantitative traits controlled by multiple genes. Studies in China have shown that although numerous QTLs (Quantitative Trait Locus) controlling quality aspects such as protein accumulation have been identified through mapping and GWAS (Genome Wide Association Studies) analyses, only a few have been isolated and functionally validated in genetic improvement, which is a challenge for quality improvement (Duan et al., 2023).

Another genetic factor analyzed was the year of release of the cultivars, where a negative linear correlation was observed between protein concentration and the year of release of the cultivars released in the period 2008 – 2023 (Fig. 4a), which is strongly associated with higher yield (Fig. 4b). The same trend was observed in the United States for 13 genotypes released between 1980 and 2014 (De Borja Reis et al., 2020). According to Winck et al. (2023), the yield increase in the last 17 years in southern Brazil is due to the genetic gain from the introduction of new cultivars with higher genetic potential (42 %), environmental conditions such as the increase in CO₂ concentration and temperature associated with climate change, which subsequently led to an increase in the yield potential of the soybean (12 %) and better technologies in the management practices of producers (44 %). On the other hand, studies using the same approach in maize in Nebraska reported 48 % yield increases associated with a decadal climate trend, 39 % with agronomic improvements such as seeding rate or increasing N fertilizer applications, and 13 % due to an improvement in genetic yield potential (Rizzo et al., 2022). Part of this increase in genetic gain in southern Brazil is largely due to emerging biotechnologies, where the goal of companies breeding programs was to develop cultivars with desirable traits, such as: resistance to insects, diseases, herbicides, drought tolerance, yield, and not chemical composition, which could explain the decrease in protein concentration in cultivars released in recent years. However, the increase in yield was higher than the decrease in protein concentration, which compensated for protein per production area with the cultivar released in recent years (Fig. 4c). So, considering that the main factor was genetic dependent for the variation in protein concentrations, the next “easy” step to increase protein concentration would be the producers to sow soybean cultivars with the highest protein content.

In addition to the genetic influence on the soybean protein concentration, an important influence of the geographical location was also determined. The results obtained in this study show that latitude is the second factor responsible for the variability across Brazil (Fig. 3a). This shows that higher protein concentrations can be achieved in latitudes close to the equator (0°), such as in the northeastern region of Brazil. Protein concentration can vary depending on the geographical location (Rotundo et al., 2016), so that locations with different environmental conditions can lead to great variability in soybean quality. These results could be related to the maturity group (MG) of cultivars that varies from the extreme south of Brazil to the northeast in a range of 5.5–9, as already reported in other studies (Assefa et al., 2019). Also, possible differences in temperatures during grain filling can be observed in the

different production systems, which have a direct reflection on the protein and oil concentration (Piper and Boote, 1999), for example, colder temperatures can result in less protein and oil synthesis because the conversion of photoassimilates is slower.

Studies comparing meal quality in soybean-producing countries show that protein concentration is higher in Brazil than in the USA (Thakur and Hurburgh, 2007). Differences can also be detected within the same country, as shown by a study at regional scale (Hurburgh, 1994b; Hurburgh et al., 1990). The protein concentration in soybean in the southern regions of the USA is higher than in the center-north region (Rotundo et al., 2016) and in Brazil it was found that the protein concentration in the southern region (Cruz Alta and Tupanciretã - RS) was lower than in the centre-north region (Sorriso - MT) (Figueiredo Moura da Silva et al., 2023).

Environmental resources or regulators are the factors determining seed composition for a given cultivar (Rotundo et al., 2016). Factors such as average temperature, average precipitation, soil fertility, soil types, production systems (e.g., soybean-cover crops; soybean-wheat; soybean-maize) can vary in different regions of Brazil. These results provide an initial step in understanding and predicting soybean seed composition at a regional scale. However, more analyzes are needed to determine the factors of variation in latitudes across states and regions in Brazil.

In addition to other biophysical factors analyzed, some significant relationships with protein concentration in seeds were also found. The mean air temperature during seed filling period was the main environmental variable associated with an increase in protein content (Table 2). At higher temperatures (> 26 °C), the composition of all grain

components are negatively affected. The accumulation of oil and residues (carbohydrates and others) is negatively affected with the increase of temperature in a high temperature range (> 26°C) during grain filling (Rotundo and Westgate, 2009); (Thomas et al., 2003). Those studies indicated that protein concentration increased with temperature (> 25°C), and this may also result from drought stress, which is often associated with higher temperatures in field situations (Piper and Boote, 1999). So, in this temperature range the oil concentration is declining rapidly with temperature, and the protein fraction may increase merely by the summation of components (i.e., the well-known negative relationship between protein and oil). Similarly, higher temperatures increase the rate of accumulation of grain constituents, and the duration of accumulation is reduced (Bhullar and Jenner, 1985; Jenner et al., 1991). In soybean, with the increase in the rate of N remobilization due to accelerated leaf senescence, an increase in the rate of protein accumulation in the seeds at higher temperatures can be expected (Egli and Wardlaw, 1980).

The photothermal coefficient also had a significant influence on the protein concentration while solar radiation did not (Table 2). The photothermal coefficient is the ratio between the incident solar radiation and the mean air temperature and integrates the effects of solar radiation on photosynthesis and temperature on cellular respiration during the critical phases of the culture (Fischer, 1985). These results indicate that temperature is the meteorological element that has a greater influence on protein concentration compared to solar radiation. In this sense, the factors that determine yield potential are genetic, temperature, solar radiation, and CO₂ (Lobell, 2009). In Brazil, the highest values of yield potential are in the more southerly latitudes (Rio Grande do Sul) (Sentelhas et al., 2015; Marin et al., 2022) due to the higher photothermal coefficient (Zanon et al., 2016), when compared to the Brazilian Cerrado (Mato Grosso), for instance. Therefore, we can assume that the protein concentration will be higher in the regions where the yield potential is lower.

When analyzing management practices, sowing date was the most important ($p < 0.0001$) for the variation in protein concentration, with an increase of 0.06 % day⁻¹ (Fig. 5a). Sowings in December had a significantly higher protein concentration than October and November, with an 85 % chance of reaching 35.3 % protein in soybean (Fig. 5b).

This result is related to the reduction in yield due to the delay in sowing date (Zanon et al., 2016; Tagliapietra et al., 2021), i.e. at earlier sowing dates the yield potential is higher, resulting in a higher N demand, which ends up impacting negatively protein synthesis in the seeds. This finding was consistent with the results obtained in Brazil (Umburanas et al., 2018), Argentina (Bozas et al., 2019), and in the midwestern United States (Rowntree et al., 2013). This management information can be used by producers to adjust their sowing date to achieve better seed quality or by buyers to better define the destination of soybean, whether for oil or meal processing industries.

The other management practice that influenced protein concentration in this study besides sowing date was irrigation. The protein concentration was higher in rainfed crops than in irrigated environments (Fig. 6a). This can be explained primarily by the higher yield response in irrigated environments and dilution effect. Secondly, a higher percentage of rainfed areas had water deficit during the 2022/2023 growing season due to the La Niña phenomenon, especially in the southern region of Brazil, which has a direct relationship with the final chemical composition of the seeds. The water deficit during grain filling (R5-R7) reduces oil synthesis (35 %) and other residual compositions (carbohydrates) (20 %) much more than protein concentration (Rotundo and Westgate, 2010; Mertz-Henning et al., 2017), with protein eventually being more concentrated due to the dilution effect. Under current climate conditions, with scarce rainfall throughout the crop cycle in the dry years, lower yields occur mainly in the south of the region (0 – 1400 kg ha⁻¹) (Battisti et al., 2016) and consequently higher protein concentration can be expected. Third, it has also been documented that soybean under conditions of water stress accelerates the rate of nitrogen remobilization from leaves (Brevedan and Egli, 2003; De Souza et al., 1997), leading to an increase in amino-N availability, resulting in higher protein concentration. However, our findings differ from other studies in which protein concentration was higher in irrigated compared to rainfed environments (Carciochi et al., 2023; Rotundo and Westgate, 2010).

However, the protein yield per area ends up being higher (Fig. 6b), which indicates that more protein is being delivered to the system but in a lower concentration. This could compensate for quantity but not quality, as soybean processors are primarily interested in the quality of protein and oil in the seeds (Cafaro La Menza et al., 2017). This information can also be used by processors to select soybean from specific regions and years with above-average levels of one or both quality constituents.

This study identified the factors that influence the variation in seed protein concentration in soybean crops in Brazil, a country projected to supply a large quantity of high-quality soybeans by 2050. The practical application of the findings may enable the implementation of strategies and management policies for producers, as well as for decision-making by processors and exporters across Brazilian soybean complex, to meet the current demands of the international market and address the protein needs of a growing global population. In this way, the processes across the entire Brazilian soybean supply chain and, consequently, global trade can be strongly impacted. Because of the large area of soybean production and different systems where soybean is a major player, further studies and the continuity of data collection over the next few years in the different production systems in Brazil are needed.

5. Conclusions

The genetics, and biophysical and management factors were responsible for 73 % and 27 % of variation in soybean seed protein concentrations in Brazil, respectively. Factors such as year of cultivar release, breeding company, latitude, temperature, photothermal coefficient, and water supply also affected the final concentration of soybean seed protein. Sowing date was the management practice that promoted the higher variation in protein concentration.

CRedit authorship contribution statement

María Sol Zelaya Arce: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Eduardo Lago Tagliapietra:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing. **José Eduardo Minussi Winck:** Investigation, Methodology, Validation, Visualization. **Alexandre Ferigolo Alves:** Investigation, Methodology. **Felipe Schmidt Dalla Porta:** Investigation, Methodology. **Tiago Broilo Facco:** Investigation, Methodology. **Nereu Augusto Streck:** Conceptualization, Investigation, Methodology, Project administration, Supervision. **Maurício Fornalski Soares:** Investigation, Methodology, Supervision. **Gregori Da Encarnação Ferrão:** Investigation, Methodology. **Daniel Debona:** Investigation, Methodology. **Claudio Hideo Martins Da Costa:** Investigation, Methodology. **Rodrigo Bega:** Investigation, Methodology. **Elizandro Fochesatto:** Investigation, Methodology. **Everton Luis Krabbe:** Methodology. **Alencar Junior Zanon:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) of the Ministry of Education of Brazil. We thank to the private companies and the public institutions of the Soybean Money Maker project for support: Universidade Federal de Maranhão (UFMA), Universidade Federal de Jataí (UFJ), Universidade Federal de Tocantins (UFT), Universidade Tecnológica Federal do Paraná (UTFPR), Universidade do Alto Vale do Rio do Peixe (UNIARP), Centro Universitário do Rio Preto (UNIRP), Fundação Mato Grosso (Fundação MT), Embrapa Aves e suínos. We also thanks Fapergs (22/2551-0000837-2) for supporting data collection and the students of the fieldcrops research team at the Universidade Federal de Santa maria (UFSM) for collecting samples and information of producers. Finally, we are grateful to the farmers of the Soybean Money Maker for sharing information in surveys and allows us to collect samples. We sincerely thank Dr. Nicolas Cafaro La Menza for his valuable suggestions and insightful ideas for data analysis.

Data availability

Data will be made available on request.

References

- Alvares, C.A., Stape, J.L., Sentelhas, P.C., Gonçalves, J.L.M., Sparovek, G., 2013. Köppen's climate classification map for Brazil. *Meteorol. Z.* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- Assefa, Y., Purcell, L.C., Salmeron, M., Naeve, S., Casteel, S.N., Kovács, P., Archontoulis, S., Licht, M., Below, F., Kandel, H., Lindsey, L.E., Gaska, J., Conley, S., Shapiro, C., Orlowski, J.M., Golden, B.R., Kaur, G., Singh, M., Thelen, K., Ciampitti, I.A., 2019. Assessing variation in us soybean seed composition (protein and oil). *Front. Plant Sci.* 10 (March). <https://doi.org/10.3389/fpls.2019.00298>.
- Battisti, R., Sentelhas, P.C., Boote, K.J., Câmara, G.M.S., Farias, J.R.B., Basso, C.J., 2016. Assessment of soybean yield with altered water-related genetic improvement traits under climate change in Southern Brazil. *Eur. J. Agron.* 83, 1–14. <https://doi.org/10.1016/j.eja.2016.11.004>.
- Beta, T., Isaak, C., 2016. Grain production and consumption: overview. In: Wrigley, C., Corke, H., Seetharaman, K., Faubion, J. (Eds.), *Encyclopedia of Food Grains - the*

- World of Food Grains, 1. Oxford, UK: Elsevier, pp. 349–428. <https://doi.org/10.1016/B978-0-12-394437-5.00051-6>.
- Bheemanahalli, R., Poudel, S., Alsajri, F.A., Reddy, K.R., 2022. Phenotyping of Southern United States soybean cultivars for potential seed weight and seed quality compositions. *Agronomy* 12, 839. <https://doi.org/10.3390/agronomy12040839>.
- Bhullar, S.S., Jenner, C.F., 1985. Differential responses to high temperatures of starch and nitrogen accumulation in the grain of 4 cultivars of wheat. *Aust. J. Plant Physiol.* 12, 363–375. <https://doi.org/10.1071/PP9850363>.
- Boote, K.J., Tollenaar, M., 1994. Modeling genetic yield potential. In: Boote, K.J., Bennett, J.M., Sinclair, T.R., Paulsen, G.M. (Eds.), *Physiology and Determination of Crop Yield*. ASA-CSSA-SSSA, Madison, WI, pp. 533–565.
- Bosaz, L.B., Gerde, J.A., Borrás, L., Cipriotti, P.A., Ascheri, L., Campos, M., Gallo, S., Rotundo, J.L., 2019. Management and environmental factors explaining soybean seed protein variability in central Argentina. *Field Crops Res* 240 (May), 34–43. <https://doi.org/10.1016/j.fcr.2019.05.007>.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. **Classification and regression trees**. Wadsworth, Belmont, CA. Clark, L.A., Pregibon, D. **Tree-based models**. In: Chambers, J.M., Hastie, T.J. (Eds.), *Statistical models in S*. Wadsworth.
- Brevedan, R.E., Egli, D.B., 2003. Short periods of water stress during seed filling, leaf senescence, and yield of soybean. *Crop Sci.* 43, 2083–2088. <https://doi.org/10.2135/cropsci2003.2083>.
- Cafaro La Menza, N., Monzon, J.P., Specht, J.E., Grassini, P., 2017. Is soybean yield limited by nitrogen supply? (<https://doi.org/>). *Field Crops Res* 213 (August), 204–212. <https://doi.org/10.1016/j.fcr.2017.08.009>.
- Carciochi, W.D., Grassini, P., Naeve, S., Specht, J.E., Mamo, M., Seymour, R., Nygren, A., Mueller, N., Sivits, S., Proctor, C., Rees, J., Whitney, T., La Menza, N.C., 2023. Irrigation increases on-farm soybean yields in water-limited environments without a trade-off in seed protein concentration (Article). *Field Crops Res* 304, 109163. <https://doi.org/10.1016/j.fcr.2023.109163>.
- CONAB, 2025. Companhia nacional de abastecimento. Brasília <http://www.conab.gov.br/>, (Accessed 24 February 2024).
- De Borja Reis, A.F., Tamagno, S., Moro Rosso, L.H., Ortez, O.A., Naeve, S., Ciampitti, I.A., 2020. Historical trend on seed amino acid concentration does not follow protein changes in soybeans. *Sci. Rep.* 10, 17707. <https://doi.org/10.1038/s41598-020-74734-1>.
- De Souza, P.I., Egli, D.B., Bruening, W.P., 1997. Water stress during seed filling and leaf senescence in soybean. *Agron. J.* 89, 807–812. <https://doi.org/10.2134/agronj1997.00021962008900050015x>.
- Duan, Z., Li, Q., Wang, H., He, X., Zhang, M., 2023. Genetic regulatory networks of soybean seed size, oil and protein contents. *Front. Plant Sci.* 14, 1160418. <https://doi.org/10.3389/fpls.2023.1160418>.
- Egli, D.B., Wardlaw, I.F., 1980. Temperature response of seed growth-characteristics of soybeans. *Agron. J.* 72, 560–564. <https://doi.org/10.2134/agronj1980.00021962007200030036x>.
- Fehr, W.R., Caviness, C.E., 1977. **Stages of soybean development**. Spec. Rep. 80, 1–12.
- Figueiredo Moura da Silva, E.H., Cafaro La Menza, N., Munareto, G.G., Zanon, A.J., Santos Carvalho, K., Marin, F.R., 2023. Soybean seed protein concentration is limited by nitrogen supply in tropical and subtropical environments in Brazil. *J. Agric. Sci.* (2), 1–12. <https://doi.org/10.1017/S0021859623000199>.
- Fischer, R.A., 1985. Number of kernels in wheat crops and the influence of solar radiation and temperature. *J. Agric. Sci.* 105, 447–461. <https://doi.org/10.1017/S0021859600056495>.
- Grassini, P., Torrión, J.A., Yang, H.S., Rees, J., Andersen, D., Cassman, K.G., Specht, J.E., 2015. Soybean yield gaps and water productivity in the western U.S. corn belt. *Field Crops Res* 179, 150–163. <https://doi.org/10.1016/j.fcr.2015.04.015>.
- Grassini, P., Cafaro La Menza, N., Rattalino Edreira, J.J., Monzon, J.P., Tenorio, F.A., Specht, J.E., 2021. **Soybean**. In: Sadras, V.O., Calderini, D.F. (Eds.), **Crop Physiology. Case Histories for Major Crops**. Elsevier, pp. 283–308.
- Henchion, M., Hayes, M., Mullen, A.M., Fenelon, M., Tiwari, B., 2017. Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. *Foods* 6. <https://doi.org/10.3390/Foods6070053>.
- Hertsgaard, D.J., Wilson, W.W., Dahl, B., 2019. Costs and risks of testing and blending for essential amino acids in soybeans. *Agribusiness* 35, 265–280. <https://doi.org/10.1002/agr.21576>.
- Hurburgh, C.R., 1994. Long-term soybean composition patterns and their effect on processing. *J. Am. Oil Chem. Soc.* 71, 1425–1427. <https://doi.org/10.1007/BF02541367>.
- Hurburgh, C.R., Brumm, T.J., Guinn, J.M., Hartwig, R.A., 1990. Protein and oil patterns in united states and world soybean markets. *J. Am. Oil Chem. Soc.* 67, 966–973. <https://doi.org/10.1007/BF02541859>.
- Instituto Brasileiro de Geografia e Estatística, 2023. **Área plantada, área colhida, quantidade produzida e rendimento médio de soja**. Bras. Iia: IBGE.
- Jenner, C.F., Ugalde, T.D., Aspinall, D., 1991. The physiology of starch and protein deposition in the endosperm of wheat. *Aust. J. Plant Physiol.* 18, 211–226. <https://doi.org/10.1071/PP9910211>.
- Liaw, A., Wiener, M., 2002. **Classification and regression by RandomForest**. R. N. 2 (3), 18–22.
- Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop yield gaps: their importance, magnitudes, and causes. *Annu. Rev. Environ. Resour.* 34, 179–204. <https://doi.org/10.1146/annurev.enviro.041008.093740>.
- Lynch, J.M., Barbano, D.M., 1999. Kjeldahl nitrogen analysis as a reference method for protein determination in dairy products. *J. AOAC Int.* 82 (6), 1389–1398. <https://doi.org/10.1093/jaoac/82.6.1389>.
- Marin, F.R., Zanon, A.J., Monzon, J.P., Andrade, J.F., Silva, E.H.F.M., Richter, G.L., Antolin, L.A.S., Ribeiro, B.S.M.R., Ribas, G.R., Battisti, R., Heinemann, A.B., Grassini, P., 2022. Protecting the Amazon forest and reducing global warming via agricultural intensification. *Nat. Sustain* 5, 1018–1026. <https://doi.org/10.1038/s41893-022-00968-8>.
- Mertz-Henning, L.M., Ferreira, L.C., Henning, F.A., Mandarino, J.M.G., Santos, E.D., Oliveira, M.C.N.D., Nepomuceno, A.L., Farias, J.R.B., Neumaier, N., 2017. Effect of water deficit-induced at vegetative and reproductive stages on protein and oil content in soybean grains. *Agronomy* 8, 3. <https://doi.org/10.3390/agronomy8010003>.
- Messina, M., 2022. Perspective: soybeans can help address the caloric and protein needs of a growing global population. *Front. Nutr.* 9, 909464. <https://doi.org/10.3389/fnut.2022.909464>.
- Parisi, G., Tulli, F., Fortina, R., Marino, R., Bani, P., Dalle, Z.A., Angelis, A., Piccolo, G., Pinotti, L., Schiavone, A., Terova, G., Prandini, A., Gasco, L., Roncarati, A., Danieli, P.P., 2020. Protein hunger of the feed sector: the alternatives offered by the plant world. *Ital. J. Anim. Sci.* 19, 1204–1225. <https://doi.org/10.1080/1828051X.2020.1827993>.
- Piper, E.L., Boote, K.J., 1999. Temperature and cultivar effects on soybean seed oil and protein concentrations. *J. Am. Oil Chem. Soc.* 76, 1233–1241. <https://doi.org/10.1007/s11746-999-0099-y>.
- Rizzo, G., Monzon, J.P., Tenorio, F.A., Howard, R., Cassman, K.G., Grassini, P., 2022. Climate and agronomy, not genetics, underpin recent maize yield gains in favorable environments. *Proc. Natl. Acad. Sci.* 119 (4). <https://doi.org/10.1073/pnas.2113629119>.
- Rotundo, J.L., Westgate, M.E., 2009. Meta-analysis of environmental effects on soybean seed composition. *Field Crops Res* 110 (2), 147–156. <https://doi.org/10.1016/j.fcr.2008.07.012>.
- Rotundo, J.L., Westgate, M.E., 2010. Rate and duration of seed component accumulation in water-stressed soybean. *Crop Sci.* 50 (2), 676–684. <https://doi.org/10.2135/cropsci2009.05.0240>.
- Rotundo, J.L., Miller-Garvin, J.E., Naeve, S.L., 2016. Regional and temporal variation in soybean seed protein and oil across the United States. *Crop Sci.* 56, 797–808. <https://doi.org/10.2135/cropsci2015.06.0394>.
- Rowntree, S.C., Suhre, J.J., Weidenbenner, N.H., Wilson, E.W., Davis, V.M., Naeve, S.L., Casteel, S.N., Diers, B.W., Esker, P.D., Specht, J.E., Conley, Shawn P., S.P., 2013. Genetic gain × management interactions in soybean: I. Planting date. *Crops Sci.* <https://doi.org/10.2135/cropsci2012.03.0157>.
- Sentelhas, P.C., Battisti, R., Câmara, G.M.S., Farias, J.R.B., Hampf, A.C., Nendel, C., 2015. The soybean yield gap in Brazil – magnitude, causes and possible solutions for sustainable production. *J. Agric. Sci.* 153, 1394–1411. <https://doi.org/10.1017/S0021859615000313>.
- Smáráson, B., Alriksson, B., Jóhannsson, R., 2019. Safe and sustainable protein sources from the forest industry – the case of fish feed. *Trends Food Sci. Technol.* 84, 12–14. <https://doi.org/10.1016/j.tifs.2018.03.005>.
- Tagliapietra, E.L., Zanon, A.J., Streck, N.A., Balest, D.S., Rosa, S.L., Bexaira, K.P., Richter, G.L., Ribas, G.G., Silva, M.R., 2021. Biophysical and management factors causing yield gap in soybean in the subtropics of Brazil. *Agron. J.* 13, 1882–1894. <https://doi.org/10.1002/agj2.20586>.
- Thakur, M., Hurburgh, C.R., 2007. Quality of US soybean meal compared to the quality of soybean meal from other origins. *J. Am. Oil Chem. Soc.* 84, 835–843. <https://doi.org/10.1007/s11746-007-1107-8>.
- Therneau, T.M., Atkinson, E.J., 1997. **An introduction to recursive partitioning using the RPART routine (Technical Report 61)**. Mayo Clinic, Section of Statistics. Mayo Foundation, Rochester, United States.
- Thomas, J.M.G., Boote, K.J.L.H., Allen, J., Gallo-Meagher, M., Davis, J.M., 2003. Elevated temperature and carbon dioxide effects on soybean seed composition and transcript abundance. *Crop Sci.* 43, 1548–1557. <https://doi.org/10.2135/cropsci2003.1548>.
- Umburanas, R.C., Yokoyama, A.H., Balena, L., Lenhani, G.C., Teixeira, A.M., Kruger, R. L., Reichardt, K., Kawakami, J., 2018. 2018. Sowing Dates and Seeding Rates Affect Soybean Grain Composition. *Int. J. Plant Prod.* 12, 181–189. <https://doi.org/10.1007/s42106-018-0018-y>.
- Wajid, A., Ahmad, M.M.F., Qureshi, M., Ceyhan, A., 2020. **Nutritive potentials of soybean and its significance for humans' health and animal production: a review**. *Eurasia J. Food Sci. Technol.* 4, 41–53.
- William, W., Dahl, B., Hertsgaard, D.J., 2019. **Soybean quality differentials, blending, testing and spatial arbitrage**. *J. Commod. Mark.* 18, 100095. <https://doi.org/10.1016/j.jcomm.2019.100095>.
- Winck, J.E.M., Tagliapietra, E.L., Schneider, R.A., Inklman, V.B., Nora, M.D., Savegnago, C., Paula, L.S., Silva, M.R.D., Zanon, A.J., Streck, N.A., 2023. Decomposition of yield gap of soybean in environment × genetics × management in Southern Brazil. *Eur. J. Agron.* 145, 126795. <https://doi.org/10.1016/j.eja.2023.126795>.
- Zanon, A.J., Streck, N.A., Grassini, P., 2016. Climate and management factors influence Soybean yield potential in a subtropical environment. *Agron. J.* 108 (4), 1447–1454. <https://doi.org/10.2134/agronj2015.0535>.
- Zhao, H., Chang, J., Havlík, P., Van Dijk, M., Valin, H., Janssens, C., Ma, L., Herrero, M., Smith, P., Obersteiner, M., 2021. China's future food demand and its implications for trade and environment. *Nat. Sustain.* 4, 1042–1051. <https://doi.org/10.1038/s41893-021-00784-6>.