UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

María Sol Zelaya Arce

FATORES GENÉTICOS, BIOFÍSICOS E DE MANEJO QUE INFLUENCIAM NA CONCENTRAÇÃO DE PROTEÍNA EM LAVOURAS DE SOJA NO BRASIL

Santa Maria, RS
2024

María Sol Zelaya Arce

FATORES GENÉTICOS, BIOFÍSICOS E DE MANEJO QUE INFLUENCIAM NA CONCENTRAÇÃO DE PROTEÍNA EM LAVOURAS DE SOJA NO BRASIL

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Agrícola, Área de Concentração de Água e Solo da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Engenharia Agrícola.**

Orientador: Prof. Dr. Alencar Junior Zanon

Santa Maria, RS

2024

María Sol Zelaya Arce

FATORES GENÉTICOS, BIOFÍSICOS E DE MANEJO QUE INFLUENCIAM NA CONCENTRAÇÃO DE PROTEÍNA EM LAVOURAS DE SOJA NO BRASIL

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Agrícola, Área de Concentração de Água e Solo da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Mestre em Engenharia agrícola.**

Aprovada em 05 de Abril de 2024

gov.br Documento assinado digitalmente
ALENCAR JUNIOR ZANON
Data: 05/04/2024 16:11:50-0300
Verifique em https://validar.iti.gov.br

Alencar Junior Zanon, Dr. (UFSM)
(Presidente/Orientador)

Firmado digitalmente por:CN=GONZALO DARÍO RIZZO AYPHASSORHO, SERIALNUMBER=DNI44000749, C=UY
Fecha:08/04/2024 14:28:18 -0300

GONZALO DARÍO RIZZO AYPHASSORHO

Gonzalo Rizzo, Dr. (UDELAR) – Videoconferência

gov.br Documento assinado digitalmente
RAFAEL BATTISTI
Data: 08/04/2024 12:39:58-0300
Verifique em https://validar.iti.gov.br

Rafael Battisti, Dr. (UFG) - Videoconferência

Santa Maria, RS

2024

DEDICATÓRIA

Aos meus pais Eduardo e Gricelda.

Aos meus avós Ramón Zelaya e Carlos Arce (in memoria) que mesmo distante sempre foram minha maior inspiração.

DEDICO

AGRADECIMENTOS

A Deus pelas oportunidades que coloca na minha vida.

Aos meus pais Eduardo e Gricelda pela educação, apoio, carinho e comprenssão.

À Universidade Federal de Santa Maria e ao Programa de Pós-Graduação em Engenharia Agrícola pela oportunidade de realizar minha formação.

À CAPES pelo auxílio financeiro através de bolsa de mestrado.

Ao professor Alencar Junior Zanon, pela orientação, dedicação, ensinamentos e empenho em proporcionar uma formação completa durante o mestrado.

Ao professor Nereu Augusto Streck, pela co-orientação, exemplo e apoio para a excussão desse trabalho.

Aos membros da banca avaliadora pela disponibilidade, sugestões e contribuições para melhoria deste trabalho.

A Startup ZEIT pela parceria para a realização das análises.

Aos meus colegas e amigos pela troca de ideias e auxílio na elaboração e execução deste estudo.

Aos colegas e amigos da Equipe FieldCrops, pelo auxílio nas atividades, pelas conversas, ensinamentos e pelos diversos momentos de troca que tivemos ao longo destes anos.

Aos colegas e amigos da Crops team pela ajuda, pelas conversas, ensinamentos e pelos diversos momentos de troca que tivemos ao longo destes anos.

A todas as pessoas que não foram mencionadas, mas que incentivaram a realização desse sonho.

MUITO OBRIGADA.

EPÍGRAFE

“Suba o primeiro degrau com fé.

Não é necessário que você veja toda a escada.

Apenas dê o primeiro passo.”

(Martin Luther King)

RESUMO

FATORES GENÉTICOS, BIOFÍSICOS E DE MANEJO QUE INFLUENCIAM NA CONCENTRAÇÃO DE PROTEÍNA EM LAVOURAS DE SOJA NO BRASIL

Autor: María Sol Zelaya Arce
Orientador: Alencar Junior Zanon

Contexto: Espera-se que a demanda por soja de alta qualidade aumente. A composição da soja pode variar devido a fatores genéticos, biofísicos e práticas de manejo. Em particular, os estudos sobre o teor de proteínas ainda estão engatinhando no país. Objetivos: Quantificar os efeitos genéticos sobre a concentração de proteína em sementes e identificar os fatores biofísicos e de manejo que influenciam a concentração de proteína em sementes de soja em diferentes sistemas de produção de soja no Brasil. Métodos: Foram coletadas amostras de soja e dados de localização e manejo da cultura por meio de levantamentos em 194 lavouras de produtores de soja em duas safras (2018/2019; 2022/2023) em onze estados do Brasil. A proteína das sementes foi determinada pelo método de Kjeldahl. Foram utilizadas árvores de regressão, regressões aleatórias florestais e comparações entre campos de alta e baixa proteína para identificar as principais causas de variação nas concentrações de proteína nas sementes de soja. Resultados: As lavouras com maior concentração de proteína foram observadas em cultivares mais velhas lançadas em (2011), com menores produtividades (3082 kg ha-1), semeadura tardia (DOY 313), temperaturas mais elevadas (25,6 °C-1) e menor coeficiente fototérmico (0,79 MJ m-2 d-1 °C-1). Por outro lado, baixa concentração de proteína foi observada em campos com maiores produtividades (4220 kg ha-1), semeadura precoce (DOY 302), menores temperaturas (24,8°C-1) e maior coeficiente fototérmico (0,84 MJ m-2 d-1 °C-1) e cultivares mais novas lançadas em (2016). A árvore de regressão e a floresta aleatória explicaram 58% da variabilidade proteica. Em relação a essa variação explicada, a cultivar (39%) foi o fator mais importante, seguida pela latitude (12%) e época de semeadura (7%). Os ambientes de sequeiro apresentaram maior concentração de proteína das sementes em relação aos campos irrigados. Conclusão: A cultivar foi o fator mais importante que afetou a concentração proteica nas sementes de soja. A época de semeadura foi a prática de manejo com maior variação na concentração de proteína. O ano de lançamento das cultivares, as empresas de melhoramento, a latitude, a temperatura, o coeficiente fototérmico e o fornecimento de água também afetaram a concentração final de proteína das sementes de soja. Significância: Nosso estudo fornece informações úteis para orientar investimentos e aplicar estratégias para produtores, tomada de decisão de processadores e exportadores de todo o complexo soja produzido no Brasil para atender a alta demanda de qualidade.

Palavras-chave: Genética. Fatores biofísicos. Práticas de manejo. Proteína

ABSTRACT

ASSESSING GENETIC, BIOPHYSICAL AND MANAGEMENT FACTORS RELATED TO SOYBEAN SEED PROTEIN VARIATION IN BRAZIL

Author: María Sol Zelaya Arce
Advisor: Dr. Alencar Junior Zanon

Context: The demand for high-quality soybeans is expected to increase. The composition of soybeans can vary due to genetics, biophysical factors, and management practices. In particular, studies on protein content are still in their infancy in the country. Objectives: Our objectives were to quantify the genetic effects on seed protein concentration and to identify the biophysical and management factors that influence seed protein concentration in the different soybean production systems in Brazil. Methods: We collected soybean samples and data of location and crop management through surveys in 194 soybean producers' fields in two growing seasons (2018/2019; 2022/2023) across eleven states in Brazil. Seed protein was determined by Kjeldahl method. We used regression trees, random forest regressions and comparisons between high and low protein fields to identify the main causes of variation in soybean seed protein concentrations. Results: 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 MJ m-2 d-1 °C-1). 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 MJ m-2 d-1 °C-1) and newer cultivars released in (2016). The regression tree and random forest explained 58% of the protein variability. Relative to this explained variation, cultivar (39%) was the most important factor, following by latitude (12%) and sowing date (7%). Rainfed environments had higher seed protein concentration compared to irrigated fields. Conclusion: Cultivar was the most important factor affecting soybean seed protein concentration. Sowing date was the management practice with higher variation in protein concentration. Cultivar release year, breeding companies, latitude, temperature, photothermal coefficient and water supply also affected the final concentration of soybean seed protein. Significance: Our study provides useful information for guiding investments and apply strategies for producers, decision-making of processor and exporters of the entire soybean complex produced in Brazil to attend high quality demand.

Keywords: Genetic. Biophysical factors. Management practices. Protein

SUMÁRIO

1. DESENVOLVIMENTO

1.1 ARTIGO - ASSESSING GENETIC, BIOPHYSICAL AND MANAGEMENT FACTORS RELATED TO SOYBEAN SEED PROTEIN VARIATION IN BRAZIL – *(será submetido a Field Crops Research)*

María Sol Zelaya Arce [a], Eduardo Lago Tagliapietra [b], José Eduardo Minussi Winck [b], Alexandre Ferigolo Alves [a], Felipe Schmidt Dalla Porta [a], Tiago Facco [a], Nereu Augusto Streck [a], Mauricio Fornalski Soares [a], Gregori E Ferrão [c], Daniel Debona [d], Claudio Martins [e], Rodrigo 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 Universitario 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, Concordia, SC 89715-899, Brazil

ABSTRACT

Context: The demand for high-quality soybeans is expected to increase. The composition of soybeans can vary due to genetics, biophysical factors, and management practices. In particular, studies on protein content are still in their infancy in the country. Objectives: Our objectives were to quantify the genetic effects on seed protein concentration and to identify the biophysical and management factors that influence seed protein concentration in the different soybean production systems in Brazil. Methods: We collected soybean samples and data of location and crop management through surveys in 194 soybean producers' fields in two growing seasons (2018/2019; 2022/2023) across eleven states in Brazil. Seed protein was determined by Kjeldahl method. We used regression trees, random forest regressions and comparisons between high and low protein fields to identify the main causes of variation in soybean seed protein concentrations. Results: 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 MJ m-2 d-1 °C-1). 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 MJ m-2 d-1 °C-1) and newer cultivars released in (2016). The regression tree and random forest explained 58% of the protein variability. Relative to this explained variation, cultivar (39%) was the most important factor, following by latitude (12%) and sowing date (7%). Rainfed environments had higher seed protein concentration compared to irrigated fields. Conclusion: Cultivar was the most important factor affecting soybean seed protein concentration. Sowing date was the management practice with higher variation in protein concentration. Cultivar release year, breeding companies, latitude, temperature, photothermal coefficient and water supply also affected the final concentration of soybean seed protein. Significance: Our study provides useful information for guiding investments and apply strategies for producers, decision-making of processor and exporters of the entire soybean complex produced in Brazil to attend high quality demand.

Keywords: genetic, biophysical factors, management practices, protein

1.1.1 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; Smárason et al., 2019; Parisi et al., 2020; Wajid et al., 2020). The economic value of soybean seeds depends on their protein and oil content (Hurburgh et al., 1990; Hurburgh, 1994). In Brazil, soybeans are currently the most important agricultural product. With a production of 154.6 million tons in the 2022/2023 harvest on 44 million hectares and an average yield of 3.5 Mg ha $^{-1}$ (CONAB, 2024), Brazil is the world's largest producer and exporter of this grain.

In 2050, China will account for 46% of the global soybean trade and the import volume is expected to reach 126 Mt. The main trading partners are Brazil and the United States, and the patterns of bilateral trade are projected to change in the future (53% import from Brazil and 37% import from the United States), which would differ from those of 2010, when both countries exported similar volumes of soybeans to China. The share of imports in total demand for oil crops (mainly soybeans) is expected to increase from 54% to 70% between 2010 and 2050. This corresponds to a projected import of 66 Mt of soybeans from Brazil in 2050, which would account for 40% of Brazil’s current soybean production (Zhao et al., 2021; FAO 2023; OECD, 2020).

On the other hand, demand for high-quality soybeans is expected to increase in the coming years, precisely 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).

In this context, there is a need to significantly increase the quantity and quality of oilseed production. At the same time, current international market regulations warn of a change in soybean quality requirements towards a higher protein concentration in the chemical composition of the seeds, which would have a strong impact on the processes of the entire Brazilian soybean chain. The low protein content in the seeds is frequently observed in different regions, making it difficult to produce a high-protein flour demanded by international markets, and studies on protein content are still in their infancy in the country (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), genetics × environment, and to a lesser extent by the agronomic management decisions (management) of producers and their interactions with genetics and environment (Grassini et al., 2021). In central Argentina, for example, genetics, environment, and management were accounted for 70%, 27% and 3% of the variation in protein concentration in soybeans in production fields, respectively (Bozas et al., 2019). However, the causes of these fluctuations are poorly understood in Brazil. Therefore, there is a need for a quantitative understanding of the factors that influence soybean seed protein concentration. 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 Brazilian crops.

Therefore, the objectives of 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 different soybean production systems in Brazil. To this end, we collected soybean samples from producers' fields with specific information of crop management and characteristics reported during two growing seasons in different regions of Brazil, including 11 states.

1.1.2 Material and Methods

1.1.2.1 *Study region*

This study was conducted for the 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 the south, southeast, central-west, northeast and north with extending over a wide latitudinal range from 3° south in the state of Maranhão to 33° south in the state of Rio

Grande do Sul (Fig. 1). Further details on the climatic characteristics of each site can be found in (Alvares et al., 2013) in relation to the Koppen´s classification.

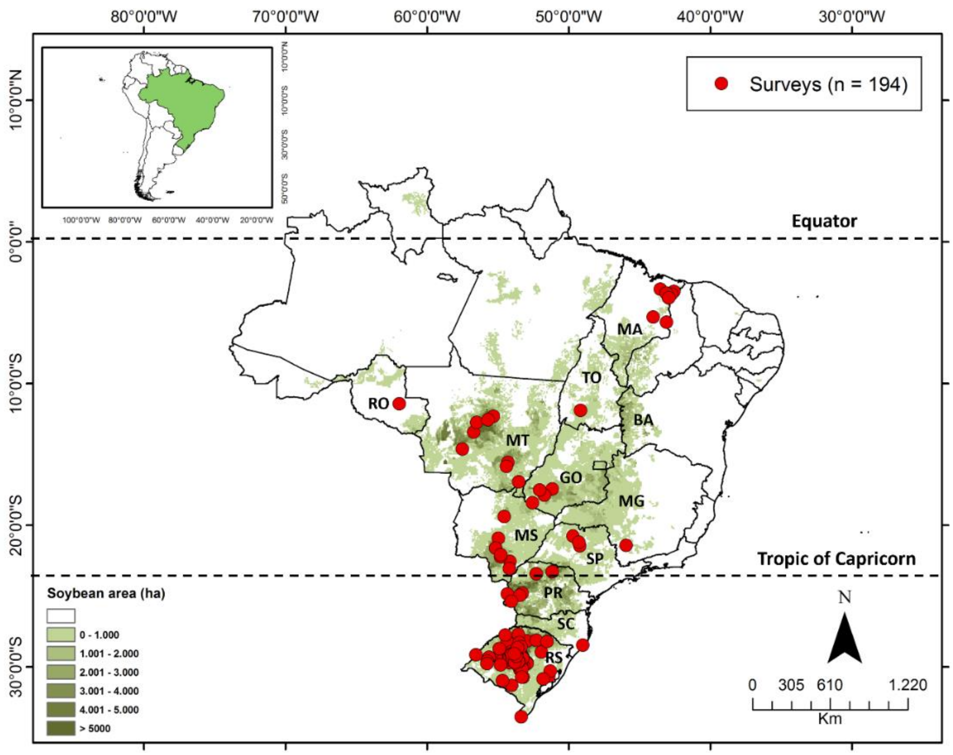

Fig. 1. Location of the surveyed fields (red circles) with soybean harvest area distribution in Brazil (green). Data source: Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística, 2023).

1.1.2.2 Data collection of genetic, biophysical and management factors

The fields of soybean producers in 2 growing seasons (2018/2019 and 2022/2023) were recorded in a database. Soybean producers provided data through surveys by members of the FieldCrops Research Group at the Federal University of Santa Maria and a database of “Efficiency and Sustainability Contest: Soybean Money Maker” was also used. Data of location, crop management practices and other factors were collected from each farmer's field: Sowing date, cultivar name, phosphorus (P) and potassium (K) fertilizer rates, lime application, soil chemical properties, water supply, yield, latitude, longitude, altitude, and others. The survey data was entered into a digital database and incorrect or very incomplete data entries were removed. After quality control, the database contained data from a total of 194 fields (96% of the total fields surveyed).

For each producer field were obtained the maximum and minimum temperature, and incident solar radiation from the meteorological stations of the National Institute of Meteorology. Field-specific average mean daily temperature (°C), average mean daily incident solar radiation (MJ m^{-2} d^{-1}) and photothermal coefficient (MJ m^{-2} d^{-1} $°C^{-1}$) were calculated for seed filling period (from R5-R7) (Fehr and Caviness, 1977). Table 1 describes all analyzed genetic, biophysical and management variables.

Table 1. Measured variables recorded for each analyzed soybean field. Management information corresponds to farm productions fields of Brazil.

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] These breeding companies were the most frequently observed.

1.1.2.3 Obtaining samples and compositional analysis

Each sample of soybean seeds corresponded to a specific field with independent cultivation in a specific growing season and site (survey). A sample of 1 kg of seed was collected from the truck at the harvest auditory and we also requested producers and sponsors to collect samples from each of the reported fields for subsequent moisture and protein analyses. 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.

1.1.2.4 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. Database mean outliers in protein concentrations were not considered 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. Regression tree analysis is a nonparametric method 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).

The caret package in R was used to create data balance splits with random sampling and to divide the data set into 70% and 30% for calibration and validation of the regression tree analysis. For this analysis, we included all qualitative and quantitative variables collected *via* our 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 importance of each variable. Random forest regression analysis was performed using random Forest package in R (Liaw and Wiener, 2002) to quantify the importance of each independent variable, which was based on how much the mean accuracy of the prediction decreases when a variable is excluded. Random forest is included in nonparametric methods, so they do not require distribution assumptions.

Linear regression analysis was used to determine the relationships between protein concentration and the reported quantitative variables previously observed to be significantly different in the other analyses. Significance was determined at p= 0.01, p= 0.05 and p= 0.1 using the GraphPad Prism statistical package (GraphPad Software Inc.).

Boxplots were used to summarize the protein variations for the qualitative management factors, and the t-test (parametric data), Wilcoxon test (non-parametric data) and Tukey test for multiple comparison of means, were used to assess significance.

1.1.3 Results

1.1.3.1 Relationship between 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% to 40.2%. A consistent trade-off correlation between protein concentration and yield was evident across production systems in different regions of Brazil.

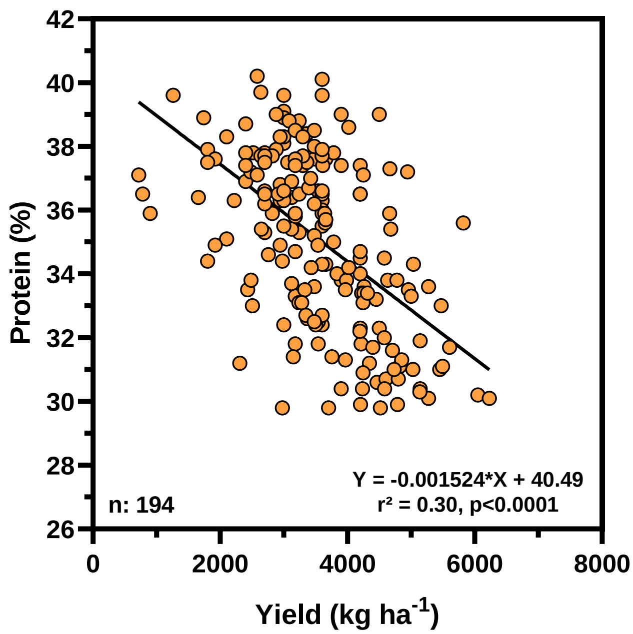

Fig. 2. Relationship between soybean seed protein and yield. Each point represents a producer field. Black line represents significant linear regression. P value and coefficient of determination (r^2) are also shown.

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

The protein concentrations in the fields with high protein concentration reached an average of 38.0%, while they averaged 31.8% in the fields with low protein concentration (Table 2). Fields with high protein concentration had 6.2% higher protein than fields with low protein concentration. 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 (cultivar release year, 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 MJ m^{-2} d^{-1} $°C^{-1}$). 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 MJ m^{-2} d^{-1} $°C^{-1}$) 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 genetic 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 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 cultivar release year) and biophysical and management factors (latitude, sowing date, and others) were responsible for 73% and 27%.

Table 2. Comparison of genetics, biophysical and management factors between the protein highest tercile (HP) and the lowest tercile (LP) in soybean fields in Brazil. The values indicate the mean differences (HP – LP) between the upper and lower protein tercile. 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	**Unit**	**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	0.8	0.8	-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_2O_5 fertilization	$kg\ ha^{-1}$	121	69	73	-4	ns	0.3941
Total K_2O 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

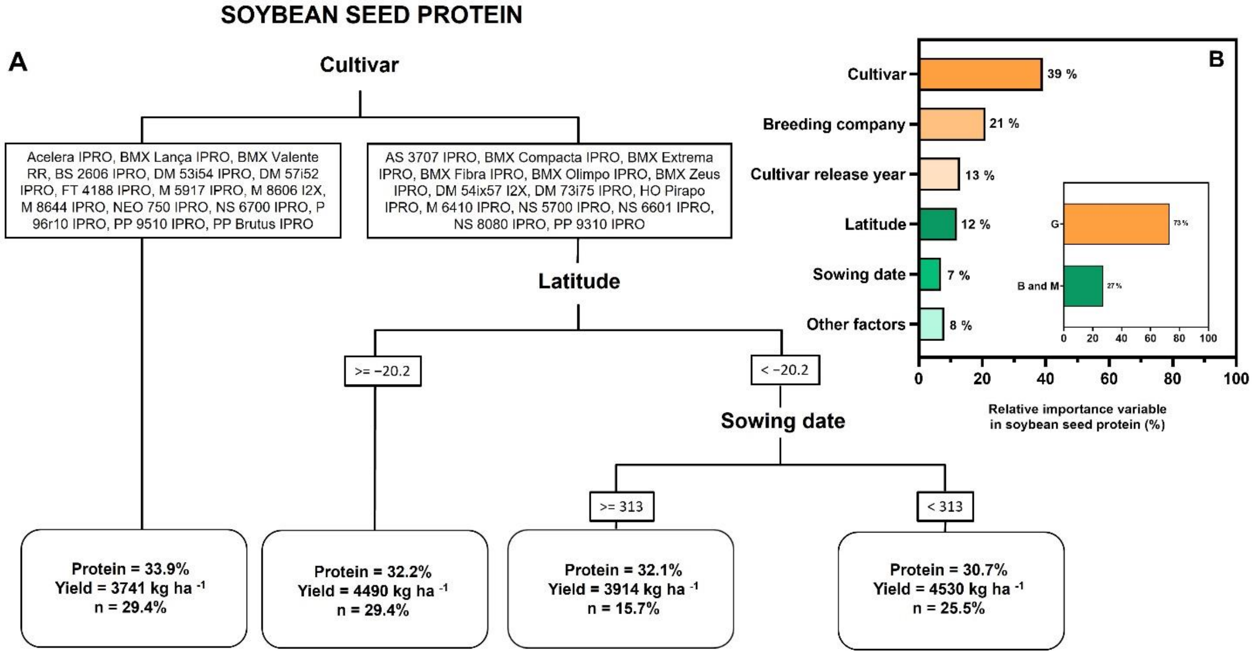

Fig. 3. Regression tree showing sources of variation in soybean seed protein due to genetic, biophysical and management factors in 2022/2023 growing season in Brazil. The boxes are splitting nodes, with bottom boxes representing terminal nodes. Values within each terminal node indicate average of seed 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).

1.1.3.3 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 correlation with the year of cultivar release, decreasing by 0.40% yr^{-1} (Fig. 4a), and yield showed a positive correlation with the year of cultivar release, increasing by 91 kg ha^{-1} 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 kg ha^{-1} yr^{-1} (Fig. 4c).

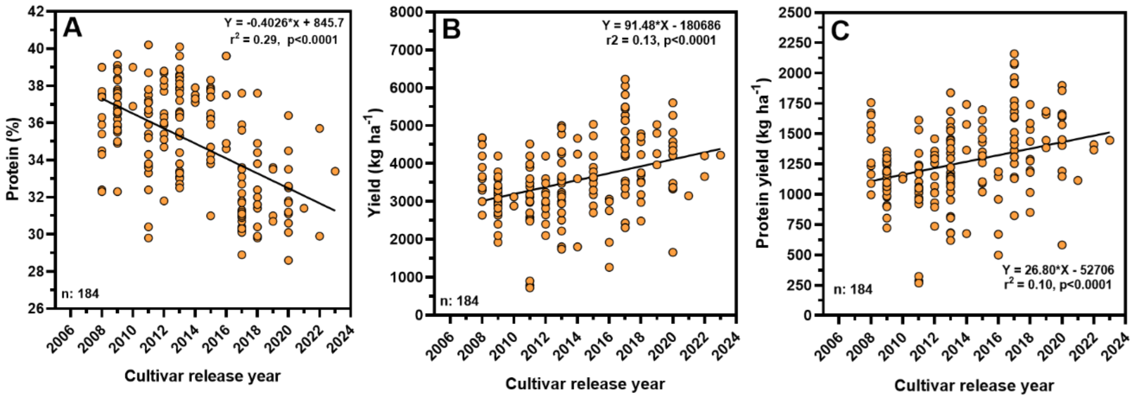

Fig. 4. Changes in soybean seed protein concentration in Brazil during (2008-2023) based on database of cultivars release year (a), genetic gain estimated by the yields of the database and cultivars release year (b), relationship between protein yield and cultivars release year (c). Each point represents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

1.1.3.4 Impact of management practices

The relationships between the protein concentration of soybean seeds show a positive correlation with the sowing date, which increases by 0.06% 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.

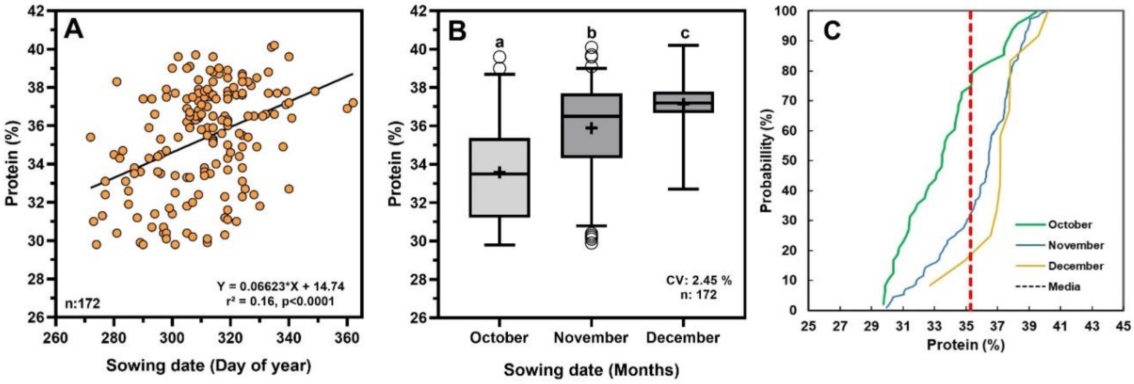

Fig. 5. Relationship between protein and sowing date (a), soybean seed protein sowing in October, November, and December (b), probability analysis for soybean seed protein of 35.3 % (dashed red line) as a function of sowing date (c) in Brazil. Black line represents significant linear regression. Boxes 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 difference with months of sowing (evaluated using Tukey test) are also shown.

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).

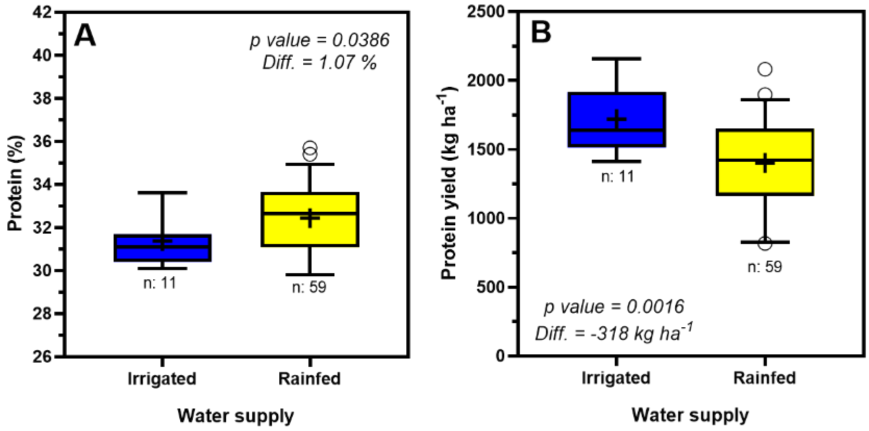

Fig. 6. Soybean seed protein in fields with irrigated and rainfed conditions (a), protein yield (kg ha $^{-1}$) in fields with irrigated and rainfed conditions (b). Boxes delimit the 5th and 95th percentiles of the distribution. Horizontal line within boxes represents the median. Dots in panels are outlier.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

1.1.4 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 crops.

A negative correlation ($p < 0.0001$) was found between protein concentration and yield in the crops studied in the different regions of Brazil (Fig. 2). This relationship has already been observed in other countries such as Central Argentina (Bozas 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 Westage, 2009), which leads to a dilution effect and reduces the protein concentration in the seeds. In Nebraska-USA, an inverse behavior was observed in the positive linear relationship between protein concentration and yield in irrigated environments, showing that this can also be attainable in other regions and environments (Carciochi et al., 2023).

Attaining high protein concentration and high yields, the total N intake must be increased (Bozas et al., 2019). A meta-analysis shows that increasing nitrogen supply increases protein concentration by ~20% (Rotundo and Westage, 2009). Similarly, experiments comparing full N versus zero N treatments showed that protein concentration decreased from 41% to 38% with increasing yield in zero N treatments, while the same trend in 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 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 in the same trend as protein concentrations in the mature seed.

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 (Bozas 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 (Minussi 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 CO_2 concentration and temperature associated with climate change, which subsequently led to an increase in the yield potential of the soybean crop (12%) and better technologies in the management practices of producers (44%). On the other hand, studies using the same approach in maize crops 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 (2008 – 2023). 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).

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 closer 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 (Assefa et al., 2019).

Studies comparing meal quality in soybean-producing countries shows 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 soybeans in the southern regions of the USA is higher than in the centre-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 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 the protein concentration in the 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 is negatively affected. According to a meta-analysis by (Rotundo and Westage, 2009), 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, while the accumulation of protein is not significantly affected. In other words, the decrease in oil and residues due to the increase in temperature leads to a net increase in protein concentration due to the dilution effect. 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 (Triboi and Triboi-Blondel, 2002; Egli and Wardlaw, 1980).

On the other hand, the photothermal coefficient also showed a significant influence with the protein concentration, conversely the solar radiation does not show to be relevant (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). This observed behavior shows 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 CO_2 (Lobell, 2009). In Brazil, the highest values of yield potential are in the lower 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 example. Therefore, we can assume that the protein concentration will be higher in the regions where the yield potential is lower.

When analyzing the management practices, the sowing date was the most important ($p < 0.0001$) for the variation in protein concentration, which showed an increase of 0.06 % day $^{-1}$ (Fig. 5a). The Tukey test showed a significant difference between sowing months (Fig. 5b), indicating that December sowings had a significantly higher protein concentration than October and November sowings, with an 85% chance of reaching 35.3% protein in soybean crops. This result is related to the reduction in yield due to the delay in sowing date (Zanon et al., 2016; Tagliapietra et al., 2021), with a dilution effect on yield. 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), but is not comparable to the results obtained by (Bellaloui et al., 2011) in Mississippi in the United States. 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 soybeans, 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’s response in irrigated environments and dilution effect mentioned above. Secondly, a higher percentage of rainfed areas were exposed to water deficit during the 2022/2023 harvest 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 of 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 than in rainfed environments (Carciochi et al., 2023; Rotundo and Westgate, 2010).

However, as can be seen in the figure (Fig. 6b), the protein yield produced per area ends up being higher, 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 soybeans from specific regions and years with above-average levels of one or both quality constituents.

Variables of the biophysical environment such as soil chemical characteristics (P soil, K soil, S soil, pH soil), longitude, altitude and solar radiation as well as management practices such as phosphorus and potassium fertilization and lime application were not significantly relevant for the variation in soybean seed protein concentration (Table 2).

This study identified the factors that influence the increase or decrease of seed protein concentration in soybean crops in Brazil, a country that will supply a large quantity of high-quality soybeans by 2050. With this, it will be possible to implement strategies and management policies for producers, as well as the decision-making of processors and exporters of the entire soybean complex produced in Brazil, to meet the current demands of the international market, which tend to sustain the current standards that deal with the quality to higher concentrations of protein. In this way, the processes of the entire Brazilian soybean chain and, consequently, trade at a global level can be strongly impacted. However, our analysis is a generalist approach for a large region, requiring continuity of data collection over the next few years in the different production systems in Brazil.

1.1.5 Conclusions

This study provides a comprehensive analysis of genetic, biophysical and management factors influencing the soybean seed protein concentration across Brazil. Due to the geographic coverage and number of observations, the database compiled in the present study can be used for future research related to improvement in quality of soybean.

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

1.1.6 Acknowledgments

This study was supported by the Coordenaçao de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) of the Ministry of Education in Brazil. We thank the private and public sector from a Brazilian Soybean Networking: 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 Universitario do Rio Preto (UNIRP), Fundacão Mato Grosso (Fundacao MT), Embrapa Aves e suinos. We also thank the students of the FieldCrops Research Group at the Universidade Federal de Santa Maria (UFSM) for collecting, organizing, and collected samples and information of producers. Finally, we are grateful to the Soybean Money Maker producers for sharing information for surveys and allows us to collect samples.

1.1.7 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., Kov´acs, 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. http://dx.doi.org/10.1016/j.eja.2016.11.004.

Bellaloui, N., Reddy, K.N., Bruns, H.A., Gillen, A.M., Mengistu, A., Zobiole, L.H.S., Fisher, D.K., Abbas, H.K., Zablotowicz, R.M., Kremer, R.J., 2011. Soybean seed composition and quality: interactions of environment, genotype, and management practices. In: Maxwell, J.E. (Ed.), Soybeans: Cultivation, Uses and Nutrition. Nova Science Publishers, Inc, pp. 1–42.

Beta, T., Isaak., C., 2016. Grain production and consumption: overview. In Wrigley C, Corke H, Seetharaman K and Faubion J (eds), Encyclopedia of Food Grains - the World of Food Grains, Vol. 1. Oxford, UK: Elsevier, 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. Australian Journal of Plant Physiology. 12, 363–375. https://doi.org/10.1071/PP9850363.

Bosaz, L.B., Gerde, J.A., Borr´as, 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 J. M. Chambers & T. J. Hastie (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 Science. 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? 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. Field Crops Res. 304, Article 109163. https://doi.org/10.1016/j.fcr.2023.109163.

CONAB, 2024. Companhia Nacional de Abastecimento, Brasilia. 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. Agronomy Journal 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. Agronomy Journal. 72, 560–564. https://doi.org/10.2134/agronj1980.00021962007200030036x.

FAOSTAT: Food and Agriculture Data (FAO, 2023). http://www.fao.org/faostat/en/#data.

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., Cafaro La Menza, N., Rattalino Edreira, J.I., 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.

Grassini, P., Torrion, 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.

Henchion, M., Hayes, M., Mullen, Am., 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.

Hurburgh, C.R., 1994. Long-term soybean composition patterns and their effect on processing. Journal of the American Oil Chemists Society. 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. Journal of the American Oil Chemists Society. 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. Brasília: IBGE.

Jenner, C.F., Ugalde, T.D., Aspinall, D., 1991. The physiology of starch and protein deposition in the endosperm of wheat. Australian Journal of Plant Physiology. 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.environ.041008.093740.

Lynch, J.M., Barbano, D.M., 1999. Kjeldahl Nitrogen Analysis as a Reference Method for Protein Determination in Dairy Products. Journal of AOAC INTERNATIONAL. 82, Issue 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., 2018. 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.

Minussi Winck, J.E., 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.

OECD–FAO Agricultural Outlook 2020–2029 (OECD, 2020). https://doi.org/10.1787/1112c23b-en/ #data.

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. Italian Journal of Animal Science. 19, 1204–1225. https://doi.org/10.1080/1828051X.2020.1827993.

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., 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.

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.

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., Shawn P. Conley. 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 posible solutions for sustainable production. Journal of Agricultural Science. 153, 1394–1411. https://doi:10.1017/S0021859615000313.

Smárason. B., Alriksson, B., Jóhannsson, R., 2019. Safe and sustainable protein sources from the forest industry – the case of fish feed. Trends in Food Science & Technology. 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 subtropicas 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. Journal of the American Oil Chemists' Society. 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. Rochester, United States: Mayo Foundation.

Triboi, E., Triboi-Blondel, A.M., 2002. Productivity and grain or seed composition: a new approach to an old problem—invited paper. European Journal of Agronomy. 16, 163–186. https://doi.org/10.1016/S1161-0301(01)00146-0.

Umburanas, R.C., Yokoyama, A.H., Balena, L, Lenhani, G.C., Teixeira, A.M., Kruger, R.L., Reichardt, K., Kawakami, J., 2018. Sowing Dates and Seeding Rates Affect Soybean Grain Composition. Int. J. Plant Prod. 12, 181–189 (2018). 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. Eurasian Journal of Food Science and Technology. 4, 41–53.

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. Nature Sustainability. Vol 4,1042–1051. https://doi.org/10.1038/s41893-021-00784-6.