COLÓQUIO
REVISTA DO DESENVOLVIMENTO REGIONAL
FACCAT
ISSN 2318-180X

EVOLUTION OF SOYBEAN PRODUCTIVITY IN RIO GRANDE DO SUL: AN ANALYSIS OF VARIANCE AND CLUSTER ANALYSIS

Daniel Gross [1]
Nilson Luiz Costa [2]
Fouad Fabio El Beitune Said [3]
Lucas França Tanaro [4]
Geferson Gustavo Wagner Mota da Silva [5]

Abstract: This study aims to analyze the evolution of the spatial variability of soybean productivity across municipalities in the state of Rio Grande do Sul, Brazil. To achieve this objective, data from the Brazilian Institute of Geography and Statistics (IBGE) regarding soybean productivity were used, applying Analysis of Variance (ANOVA) and cluster analysis techniques. ANOVA was conducted for the period from 1974 to 2023, allowing the identification of three distinct phases in the productive dynamics of soybean cultivation, characterized by a transition from a relatively homogeneous pattern, with a standard deviation of 236.57 kg/ha, to a highly heterogeneous scenario, in which the standard deviation reached 716.95 kg/ha. Cluster analysis was performed for the period from 2004 to 2023, when the state's political-administrative division was already consolidated, with no significant municipal emancipation processes. The results enabled the identification of seven groups composed of municipalities with similar productivity levels. The observed spatial variability can be attributed to multiple factors, particularly the territorial expansion of soybean cultivation into marginally suitable areas and the influence of edaphoclimatic conditions on regional productive performance. The findings contribute to a better understanding of the spatial evolution of soybean production in Rio Grande do Sul and the productivity differences among its territories.

Keywords*: Agricultural productivity; Soybeans; Spatial variability; Rio Grande do Sul.*

EVOLUCIÓN DE LA PRODUCTIVIDAD DE LA SOJA EN RIO GRANDE DO SUL: UN ANÁLISIS DE VARIANZA Y ANÁLISIS DE CLÚSTERES

Resumen: Este estudio tiene como objetivo analizar la evolución de la variabilidad espacial de la productividad de la soja entre los municipios del estado de Rio Grande do Sul, Brasil. Para ello, se utilizaron datos del Instituto Brasileño de Geografía y Estadística (IBGE) sobre la productividad de la soja, aplicando técnicas de Análisis de Varianza (ANOVA) y análisis de clústeres. El ANOVA se realizó para el período comprendido entre 1974 y 2023, permitiendo identificar tres fases distintas en la dinámica productiva del cultivo de soja, caracterizadas por una transición desde un patrón relativamente homogéneo, con una desviación estándar de 236,57 kg/ha, hacia un escenario altamente heterogéneo, en el que la desviación estándar alcanzó los 716,95 kg/ha. El análisis de clústeres se llevó a cabo para el período 2004–2023, cuando la división político-administrativa del estado ya se encontraba consolidada, sin procesos significativos de emancipación municipal. Los resultados permitieron identificar siete grupos compuestos por municipios con niveles de productividad similares. La variabilidad espacial observada puede atribuirse a múltiples factores, en particular a la expansión territorial del cultivo de soja hacia áreas marginalmente aptas y a la influencia

---

[1] Ph.D. Student in the Graduate Program in Agribusiness, Federal University of Santa Maria (UFSM), Brazil. ORCID: https://orcid.org/0000-0001-8591-7122. Lattes: https://lattes.cnpq.br/6678901641057217. E-mail: dgross88@gmail.com.

[2] Ph.D. in Agricultural Sciences from the Federal Rural University of the Amazon (UFRA), Brazil. ORCID: https://orcid.org/0000-0003-2000-1521. Lattes: http://lattes.cnpq.br/4436596248591572. E-mail: nilson.costa@ufsm.com.br

[3] Ph.D. Student in the Graduate Program in Agribusiness, Federal University of Santa Maria (UFSM), Brazil. ORCID: https://orcid.org/0009-0002-5250-8570. Lattes: http://lattes.cnpq.br/0582060139848149. E-mail: ffbeitune@yahoo.com.br

[4] M.Sc. Student in the Graduate Program in Agribusiness, Federal University of Santa Maria (UFSM), Brazil. ORCID: https://orcid.org/0009-0009-1684-8755. Lattes: http://lattes.cnpq.br/2598083718730791. E-mail: lucas.franca004@gmail.com

[5] Ph.D. Student in the Graduate Program in Agribusiness, Federal University of Santa Maria (UFSM), Brazil. ORCID: https://orcid.org/0000-0002-2325-4666. Lattes: http://lattes.cnpq.br/9832170280673981. E-mail: geferson_gustavo@hotmail.com

de las condiciones edafoclimáticas sobre el desempeño productivo regional. Los hallazgos contribuyen a una mejor comprensión de la evolución espacial de la producción de soja en Rio Grande do Sul y de las diferencias de productividad entre sus territorios.

Palabras clave: Productividad agrícola; Soja; Variabilidad espacial; Rio Grande do Sul

1 Introduction

The soybean-cultivated area in the state of Rio Grande do Sul remained relatively stable between the mid-1970s and the early 2000s, increasing from approximately 3.1 million to 3.3 million hectares. During this period, the average cultivated area was approximately 3.32 million hectares, reaching a peak of 4.03 million hectares in 1979 and a minimum of 2.49 million hectares in 1996 (IBGE, 2025).

From 2002 onward, soybean cultivation entered a continuous expansion phase, substantially reshaping the agricultural production structure of Rio Grande do Sul. By the end of that decade, the cultivated area had increased by approximately 27.3%, a trend that intensified after 2010, when soybean acreage expanded by nearly 45% through 2020. In 2010, soybean occupied approximately 4 million hectares, increasing to nearly 6.6 million hectares by 2023, according to IBGE data (2025). As a result, soybean accounted for 63% of the total area allocated to temporary crops in the state, compared with 51.76% in 2010, representing an increase of 21.72 percentage points in the agricultural production structure of Rio Grande do Sul.

Until the early 2000s, soybean production in Rio Grande do Sul was concentrated primarily in the Mid-Plateau region and adjacent areas, which have traditionally been recognized for their high suitability for soybean cultivation (Oderich, Elias, & Waquil, 2019). However, between 2000 and 2017, soybean acreage expanded substantially, increasing from approximately 3.0 to 5.5 million hectares. During this period, the soybean production chain significantly increased its economic importance within the state, with approximately 87% of the expansion occurring in agricultural frontier regions (Costa et al., 2020). Oderich, Elias, and Waquil (2019) characterize this process as the main manifestation of the so-called “agribusiness cycle” in Rio Grande do Sul. The incorporation of new production areas contributed to increasing regional productivity disparities, making the analysis of the crop’s spatial variability an important element for understanding the recent dynamics of soybean production in the state.

The territorial expansion of soybean cultivation occurred through the incorporation of regions previously considered marginal for crop production, such as the Southeastern Highlands (Serra do Sudeste) and the Campanha region (Oderich, Elias, & Waquil, 2019). Similarly, Torres et al. (2023) identified the West-Central, Southeast, and Southwest mesoregions as the areas where soybean cultivation advanced most rapidly over recent decades. In this context, soybean cultivation, together with forestry activities, expanded over areas substantially larger than those displaced from other agricultural crops, indicating the conversion of land traditionally occupied by livestock production within the Pampa biome (Oderich, Elias, & Waquil, 2019).

Between the 2000/01 and 2014/15 growing seasons, soybean acreage within the Pampa biome increased by approximately 57%, corresponding to the conversion of nearly 1.53 million hectares—primarily natural grasslands—into agricultural land dedicated to soybean production (Silveira, González, & Fonseca, 2017). On a broader scale, Torres et al. (2023) reported that soybean acreage in Rio Grande do Sul increased by 100.5% between 2001 and 2020, expanding over areas previously allocated to rice and corn cultivation as well as livestock production. The West-Central, Southeast, and Southwest mesoregions accounted for approximately 62.4% of all land converted to soybean production in the state.

The factors underlying this reconfiguration of the agricultural production structure are primarily associated with the appreciation of soybean prices in international markets beginning in the early 2000s (Silveira, González, & Fonseca, 2017). Furthermore, Torres et al. (2023) emphasize that the adoption of new agricultural technologies and the introduction of genetically modified cultivars contributed significantly to this expansion process. Consequently, the importance of soybean production within local economies increased across the interior regions of Rio Grande do Sul, making many municipalities highly dependent on this crop (Oderich, Elias, & Waquil, 2019). This process also enhanced the relevance of agribusiness to the state economy, as the soybean production chain accounted for at least 52% of Rio Grande do Sul’s exports over the past 21 years (Costa et al., 2020). Therefore, fluctuations in soybean productivity have the potential to affect not only the economic performance of producing municipalities but also the competitiveness of the state’s agribusiness sector in national and international markets.

The expansion of soybean cultivation throughout the state has also contributed to increasing heterogeneity in agricultural productivity across producing regions. This process is closely linked to climatic challenges, particularly drought events, which substantially affect agricultural productivity in Rio Grande do Sul. The recurrence of water deficits represents one of the main factors reducing crop yields, as most agricultural production is conducted under rainfed conditions, resulting in significant economic and social losses (Lazzari, 2006; Sentelhas et al., 2015; Leivas et al., 2014). This dynamic helps explain the spatial variability in productivity observed across different regions of the state (Arsego et al., 2019).

In this regard, Torres et al. (2023) observed that, between 2011 and 2020, the Southeast and Southwest regions—characterized by the most intensive expansion of soybean cultivation into previously non-agricultural areas—recorded the lowest average yields per hectare in the state. In contrast, the Northwest and East-Central regions achieved productivity levels above the state average, in a context where soybean expansion occurred predominantly through the replacement of areas previously occupied by corn cultivation. These findings suggest that land-use history exerts a direct influence on agricultural performance.

Given this context, analyzing the area, production, and productivity of the state's main agricultural crop is of central importance for regional development studies. Excessive productive specialization may increase the economic vulnerability of territories, particularly in regions exposed to recurring climatic risks. Identifying spatial productivity patterns not only contributes to understanding regional differences in agricultural performance but also provides valuable insights for the formulation of public policies aimed at mitigating climatic risks, promoting productive diversification, reducing regional inequalities, and strengthening the economic and social development of municipalities.

Against this backdrop, in which soybean cultivation advances into marginal lands, reshapes land-use patterns, and faces recurrent climatic instability, it becomes essential to understand how these processes are reflected in agricultural productivity levels throughout Rio Grande do Sul. Accordingly, the following research question emerges: Is there statistically significant evidence of spatial and temporal heterogeneity in soybean productivity among municipalities in Rio Grande do Sul that allows the identification of territories exhibiting persistently higher productive suitability over time?

Based on this guiding question, the objective of this study is to analyze the spatial variability and temporal evolution of soybean productivity across municipalities in Rio Grande do Sul through the application of Analysis of Variance (ANOVA) and Cluster Analysis (CA).

2 Methodological Framework

This study adopts a quantitative approach. The methodological procedures consisted of applying Analysis of Variance (ANOVA) to soybean productivity in the state of Rio Grande do Sul over a fifty-year period (1974–2023), as well as grouping the state's 497 municipalities into clusters based on a temporal analysis of soybean productivity during the most recent twenty-year period (2004–2023).

Additionally, a bibliographic and documentary review was conducted to contextualize the evolution of soybean production in the state, particularly regarding the territorial expansion of the crop, the climatic factors affecting productivity, and the transformations in the regional production structure. This stage enabled the interpretation of quantitative results in light of the specialized literature, contributing to a more comprehensive understanding of the productive differences identified among municipal clusters.

2.1 Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a statistical technique that allows multiple group comparisons within a single test using data obtained from samples drawn from the same population. According to Hair et al. (2009), ANOVA makes it possible to determine whether differences among group means are statistically significant or merely attributable to random variation and sampling error.

Fávero (2017) describes ANOVA as a statistical procedure designed to compare the means of three or more populations based on samples extracted from each population, with the objective of verifying whether significant differences exist among the sample means.

In the present study, ANOVA was employed to analyze soybean productivity across municipalities in Rio Grande do Sul between 1974 and 2023 using data from the Municipal Agricultural Production Survey (PAM) conducted by the Brazilian Institute of Geography and Statistics (IBGE, 2025), extracted in May 2025. For analytical purposes, municipalities were grouped according to their respective Regional Development Councils (COREDEs). The adopted classification comprises 28 regional councils established under State Decree No. 54,572 of April 14, 2019. Municipalities are allocated to each COREDE by public authorities based on shared socioeconomic and territorial characteristics.

The ANOVA procedure is based on the assumption that two independent estimates of variance can be obtained: one associated with variation within groups and another associated with variation between groups. The technique relies on comparing these two sources of variance, which is why it is referred to as Analysis of Variance (ANOVA). The procedure involves calculating variation between groups $(\bar{Y}_i - \bar{Y})$ and variation within groups $(Y_{ij} - \bar{Y}_i)$. The Sum of Squares Within Groups (SSW) is calculated as follows (Fávero, 2017):

$$SQU = \sum_{i=1}^{k} \sum_{j=1}^{n_j} (Y_{ij} - \bar{Y}_i)^2$$

The Sum of Squares Between Groups (SSB), also referred to as the factor sum of squares, is computed as follows (Fávero, 2017):

$$SQF = \sum_{i=1}^{k} n_i \,.\, (\bar{Y}_i - \bar{Y})^2$$

The Total Sum of Squares (SST) is obtained by:

$$SQT = SQU + SQF = \sum_{i=1}^{k} \sum_{j=1}^{n\ i} (Y_{ij} - \bar{Y})^2$$

The ANOVA F-statistic is calculated as the ratio between the variance explained by the factor (SSB divided by (k – 1) degrees of freedom) and the residual variance (SSW divided by (N-k) degrees of freedom), as follows (Fávero, 2017):

$$F_{cal} = \frac{\frac{SQF}{k-1}}{\frac{SQU}{N-k}} = \frac{QMF}{QME}$$

ANOVA procedures were performed using IBM® SPSS® Statistics software.

2.2 Cluster Analysis

This study employed Cluster Analysis, also referred to as clustering or grouping analysis. Cluster analysis is an exploratory statistical technique whose primary objective is to identify similarities among the elements of a sample. The method enables the formation of groups composed of observations that share internal similarities while remaining distinct from other groups based on the variables under investigation (Fávero, 2017).

Unlike other multivariate techniques designed to estimate statistical relationships among variables, cluster analysis seeks to classify observations into relatively homogeneous groups according to characteristics previously selected by the researcher (Hair et al., 2009).

For this purpose, data obtained from the Brazilian Institute of Geography and Statistics (IBGE, 2025), extracted in May 2025, were used. The dataset consisted of productivity information from temporary crops, with particular emphasis on soybean production, derived from the Municipal Agricultural Production Survey (PAM). The objective was to classify municipalities in Rio Grande do Sul according to their average soybean productivity during the period from 2004 to 2023.

The selection of this time frame was motivated by changes in the state's administrative structure. In 1974, Rio Grande do Sul comprised 235 municipalities. Beginning in 1987, the state experienced an intense process of municipal emancipation, characterized by three successive waves of municipal creation (1987–1988, 1992, and 1995–1996), which substantially expanded the state's political-administrative division. The most recently established municipality was Pinto Bandeira in 2001. However, it was dissolved in 2003 and reincorporated into Bento Gonçalves before being re-established in 2013. As a result, Rio Grande do Sul currently consists of 497 municipalities (Rio Grande do Sul, 2018). Consequently, restricting the analysis to the period from 2004 to 2023 minimizes potential distortions arising from territorial changes and ensures greater consistency when comparing spatial productivity patterns among municipalities.

Given the substantial increase in the number of municipalities over time, conducting cluster analysis for earlier periods could compromise the results, as traditionally productive agricultural areas might be allocated to different groups solely due to administrative boundary changes. In such cases, the resulting clusters would reflect the process of territorial emancipation rather than actual soybean productivity patterns.

To perform the clustering procedure, two complementary techniques were employed: hierarchical clustering and the k-means clustering method.

Hierarchical clustering is characterized by a sequential process of combining observations based on their similarities through clustering algorithms. This procedure progressively groups observations until a tree-like structure, known as a dendrogram, is formed. The dendrogram allows researchers to visualize alternative grouping structures and assists in determining the most appropriate number of clusters (Hair et al., 2009).

Subsequently, the k-means clustering method was applied. This non-hierarchical procedure allocates observations to clusters according to their proximity to predefined cluster centroids, thereby optimizing within-group homogeneity (Fávero, 2017).

The combined use of hierarchical and non-hierarchical clustering techniques is justified by their complementary strengths. Hierarchical clustering provides a visual representation of potential grouping structures and supports the determination of the optimal number of clusters. In contrast, non-hierarchical methods such as k-means require the number of clusters to be specified a priori and subsequently allocate observations to clusters based on their similarity to cluster centroids (Fávero, 2017).

Accordingly, the analysis was conducted using IBM® SPSS® Statistics software. Initially, hierarchical clustering was performed, and the resulting dendrogram indicated the formation of seven clusters. Subsequently, the k-means procedure was applied to classify the 497 municipalities of Rio Grande do Sul into the seven previously defined groups.

To present the results, graphs were generated using Microsoft Excel, illustrating the mean productivity values and standard deviations associated with each cluster. The spatial distribution of the clusters was mapped using QGIS version 3.40.6 (Bratislava), allowing for the geographic visualization of the productivity patterns identified through the clustering analysis.

3 Results and Discussion

This section presents the main findings of the study. First, the results of the Analysis of Variance (ANOVA) conducted on soybean productivity in Rio Grande do Sul between 1974 and 2023 are discussed. Subsequently, the findings from the Cluster Analysis are presented, which grouped the state's 497 municipalities according to their soybean productivity patterns during the period from 2004 to 2023.

3.1 Analysis of Variance (ANOVA)

The analysis of soybean productivity across municipalities in Rio Grande do Sul, based on the complete historical series available from the Brazilian Institute of Geography and Statistics (IBGE, 2025) covering the period from 1974 to 2023, provides insights not only into the evolution of soybean production and changes in productivity levels over time but also into the territorial expansion of soybean cultivation into new producing regions. These dynamics are evidenced by the data presented in Table 1

Table 1. ANOVA results for soybean productivity in Rio Grande do Sul, Brazil, 1974–2023.

ANOVA														
		Sum of	df	Mean	F	Sig.			Sum of	df	Mean	F	Sig.	
1974	Between	19069937,134	27	706293,968	1,413	,084		1999	Between	111563933,036	27	4131997,520	11,136	,000
	Within	234497571,563	469	499994,822					Within	174014917,229	469	371033,939		
	Total	253567508,696	496						Total	285578850,266	496			
1975	Between	24525999,760	27	908370,361	1,693	,017		2000	Between	159516833,024	27	5908030,853	14,002	,000
	Within	251649110,683	469	536565,268					Within	197892814,859	469	421946,300		
	Total	276175110,443	496						Total	357409647,883	496			
1976	Between	26852427,023	27	994534,334	1,777	,010		2001	Between	311344856,331	27	11531290,975	27,877	,000
	Within	262463615,524	469	559623,914					Within	194004821,935	469	413656,337		
	Total	289316042,547	496						Total	505349678,266	496			
1977	Between	25534176,644	27	945710,246	1,511	,050		2002	Between	170305075,590	27	6307595,392	15,831	,000
	Within	293500982,587	469	625801,669					Within	186863626,551	469	398429,907		
	Total	319035159,231	496						Total	357168702,141	496			
1978	Between	13712494,426	27	507870,164	1,426	,078		2003	Between	387000605,893	27	14333355,774	32,636	,000
	Within	167024908,556	469	356129,869					Within	205977444,663	469	439184,317		
	Total	180737402,982	496						Total	592978050,555	496			
1979	Between	7801485,764	27	288943,917	1,210	,217		2004	Between	122344938,158	27	4531294,006	18,782	,000
	Within	111965754,445	469	238732,952					Within	113150805,641	469	241259,714		
	Total	119767240,209	496						Total	235495743,799	496			
1980	Between	18126677,959	27	671358,443	1,215	,212		2005	Between	24536980,735	27	908777,064	7,580	,000
	Within	259126113,108	469	552507,704					Within	56227795,164	469	119888,689		
	Total	277252791,066	496						Total	80764775,899	496			
1981	Between	22984772,972	27	851287,888	1,353	,113		2006	Between	249644328,970	27	9246086,258	24,156	,000
	Within	295161010,883	469	629341,175					Within	179520458,394	469	382772,832		
	Total	318145783,855	496						Total	429164787,364	496			
1982	Between	14255529,825	27	527982,586	1,448	,070		2007	Between	370551305,199	27	13724122,415	29,574	,000
	Within	170982141,555	469	364567,466					Within	217641378,315	469	464054,112		
	Total	185237671,380	496						Total	588192683,513	496			
1983	Between	33027256,094	27	1223231,707	2,424	,000		2008	Between	252662965,076	27	9357887,595	24,201	,000
	Within	236661873,588	469	504609,539					Within	181350864,558	469	386675,617		
	Total	269689129,682	496						Total	434013829,634	496			
1984	Between	26968670,290	27	998839,640	1,977	,003		2009	Between	244208703,588	27	9044766,800	20,139	,000
	Within	236985290,635	469	505299,127					Within	210631933,764	469	449108,601		
	Total	263953960,926	496						Total	454840637,352	496			
1985	Between	30407316,971	27	1126196,925	2,023	,002		2010	Between	383966164,397	27	14220969,052	29,080	,000
	Within	261071521,057	469	556655,695					Within	229356635,736	469	489033,338		
	Total	291478838,028	496						Total	613322800,133	496			
1986	Between	16170876,801	27	598921,363	2,449	,000		2011	Between	481162194,509	27	17820822,019	34,133	,000
	Within	114697631,368	469	244557,849					Within	244865537,909	469	522101,360		
	Total	130868508,169	496						Total	726027732,419	496			
1987	Between	36239018,785	27	1342185,881	2,610	,000		2012	Between	180948901,477	27	6701811,166	14,955	,000
	Within	241169451,014	469	514220,578					Within	210166846,914	469	448116,944		
	Total	277408469,799	496						Total	391115748,390	496			
1988	Between	14076379,979	27	521347,407	1,949	,003		2013	Between	356240610,383	27	13194096,681	22,150	,000
	Within	125450180,637	469	267484,394					Within	279368069,754	469	595667,526		
	Total	139526560,616	496						Total	635608680,137	496			
1989	Between	68894088,130	27	2551632,894	4,231	,000		2014	Between	329960736,270	27	12220768,010	21,334	,000
	Within	282854382,417	469	603101,029					Within	268662415,363	469	572840,971		
	Total	351748470,547	496						Total	598623151,634	496			
1990	Between	62265269,117	27	2306121,078	3,722	,000		2015	Between	440222917,150	27	16304552,487	25,245	,000
	Within	290598667,993	469	619613,365					Within	302907942,367	469	645859,152		
	Total	352863937,111	496						Total	743130859,517	496			
1991	Between	11387445,438	27	421757,238	2,669	,000		2016	Between	470727901,849	27	17434366,735	27,933	,000
	Within	74112408,047	469	158022,192					Within	292721325,978	469	624139,288		
	Total	85499853,485	496						Total	763449227,827	496			
1992	Between	90778518,534	27	3362167,353	4,876	,000		2017	Between	489688977,019	27	18136628,778	26,731	,000
	Within	323402351,731	469	689557,253					Within	318213181,443	469	678492,924		
	Total	414180870,266	496						Total	807902158,463	496			
1993	Between	145566095,351	27	5391336,865	10,590	,000		2018	Between	454922698,916	27	16848988,849	24,464	,000
	Within	238772602,641	469	509110,027					Within	323011815,329	469	688724,553		
	Total	384338697,992	496						Total	777934514,245	496			
1994	Between	117567188,982	27	4354340,333	9,734	,000		2019	Between	450512881,901	27	16685662,293	25,565	,000
	Within	209801366,559	469	447337,669					Within	306104670,856	469	652675,204		
	Total	327368555,541	496						Total	756617552,757	496			
1995	Between	143418355,571	27	5311790,947	10,002	,000		2020	Between	229050506,883	27	8483352,107	18,338	,000
	Within	249071889,749	469	531070,127					Within	216966912,469	469	462616,018		
	Total	392490245,320	496						Total	446017419,352	496			
1996	Between	99194520,351	27	3673871,124	8,225	,000		2021	Between	433142146,425	27	16042301,719	22,719	,000

	Within	209499796,743	469	446694,663				Within	331168173,116	469	706115,508		
	Total	308694317,095	496					Total	764310319,541	496			
1997	Between	127780193,043	27	4732599,742	13,088	,000	2022	Between	194075483,702	27	7187980,878	13,945	,000
	Within	169594270,349	469	361608,252				Within	241745014,237	469	515447,792		
	Total	297374463,392	496					Total	435820497,940	496			
1998	Between	205773738,717	27	7621249,582	15,816	,000	2023	Between	215794747,464	27	7992398,054	12,664	,000
	Within	225999085,557	469	481874,383				Within	295994250,810	469	631117,806		
	Total	431772824,274	496					Total	511788998,274	496			

Source: Prepared by the authors based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

The Analysis of Variance (ANOVA) revealed a clear evolutionary trajectory in soybean productivity across Rio Grande do Sul over the past five decades. The results indicate the existence of three distinct phases in the development of soybean production within the state, as presented in Table 1.

Table 1. Evolution of the spatial heterogeneity of soybean productivity in Rio Grande do Sul, Brazil.

Phase	Period	Mean Productivity	Mean Standard Deviation
Phase 1	1974–1982	1,380.22 kg ha^{-1}	236.57 kg ha^{-1}
Phase 2	1983–1999	1,581.00 kg ha^{-1}	363.60 kg ha^{-1}
Phase 3	2000–2023	2,312.12 kg ha^{-1}	716.95 kg ha^{-1}

Source: Authors' elaboration based on IBGE (2025).

Phase 1 encompasses the period from 1974 to 1982. During this interval, differences in soybean productivity among municipalities were relatively small, resulting in a fairly homogeneous production pattern throughout the state. Average soybean productivity reached 1,380.22 kg ha^{-1}, while the mean standard deviation was 236.57 kg ha^{-1}. The highest municipal productivity recorded during this phase was observed in Bagé, which achieved 2,690 kg ha^{-1} in 1981.

Phase 2 covers the period from 1983 to 1999. During this stage, the initial signs of productive differentiation among municipalities became evident. This trend is reflected in the 53.7% increase in the average standard deviation, which rose to 363.60 kg ha^{-1}, while mean state productivity increased by 14.6%, reaching 1,581 kg ha^{-1}. The highest municipal productivity during this period was recorded in Colorado, which achieved 3,044 kg ha^{-1} in 1992.

Finally, Phase 3, spanning the period from 2000 to 2023, was characterized by pronounced heterogeneity in municipal productivity levels. Average state productivity increased by 46.2% relative to the previous period, reaching 2,312.12 kg ha^{-1}. However, the mean standard deviation rose to 716.95 kg ha^{-1}, representing increases of 97.2% and 203.1% compared with Phases 2 and 1, respectively. These results reveal the substantial spatial variability identified by the ANOVA. During this period, the municipality of Nova Bassano recorded the highest productivity observed in the dataset, reaching 4,800 kg ha^{-1} in 2015, whereas Herval recorded only 1,200 kg ha^{-1} in the same year, corresponding to a fourfold difference between the two municipalities.

The findings observed for Phase 3 are consistent with those reported by Torres et al. (2023), who identified a substantial increase in soybean productivity in Rio Grande do Sul during the period of rapid crop expansion. According to these authors, average state productivity increased from 1,936 kg ha^{-1} during 2001–2010 to 2,828 kg ha^{-1} during 2011–2020. This pattern is consistent with the results obtained in the present study, which likewise identified significant growth in average soybean productivity during the third phase. Nevertheless, the ANOVA results demonstrate that these productivity gains were unevenly distributed across municipalities, resulting in an even greater increase in spatial heterogeneity.

Overall, the ANOVA results indicate that soybean productivity in Rio Grande do Sul has undergone a continuous process of spatial differentiation over the last five decades. Initially concentrated in traditional production areas characterized by relatively homogeneous production conditions, soybean cultivation progressively expanded into regions with distinct edaphoclimatic characteristics and varying levels of technological adoption. As a consequence, productivity disparities among municipalities became increasingly pronounced.

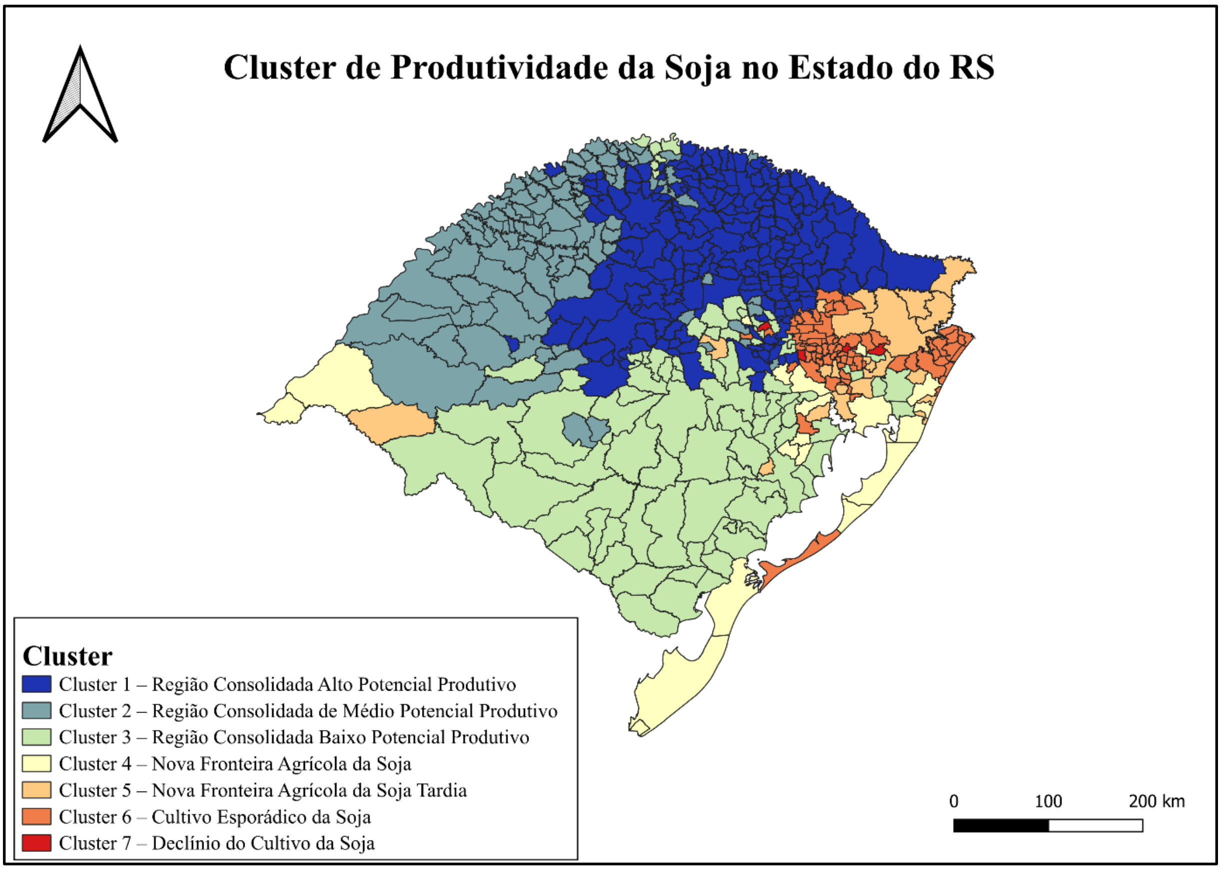
In this context, the spatial evolution of soybean production in Rio Grande do Sul cannot be explained solely by increases in cultivated area or total output. Rather, it also reflects a growing differentiation among producing territories, resulting in the emergence of regions with distinct productive potentials, which are subsequently identified through cluster analysis.

3.2 Cluster Analysis

Cluster analysis enabled the identification of similar spatial patterns of soybean productivity among municipalities in the state of Rio Grande do Sul. The analysis was based on average soybean productivity levels observed over the most recent twenty-year period (2004–2023), using data obtained from the Brazilian Institute of Geography and Statistics (IBGE, 2025). The selection of this temporal framework was supported by two main considerations. First, the ANOVA results revealed that the period after 2000 was characterized by greater spatial heterogeneity in soybean productivity, making it particularly suitable for identifying distinct productivity patterns among municipalities. Second, this period was marked by greater territorial stability, as no significant municipal emancipations occurred within the state. Consequently, the selected time frame provided a consistent basis for comparing productivity patterns across municipalities and minimizing distortions associated with changes in administrative boundaries.

Based on these criteria, seven clusters were identified, grouping municipalities according to similar soybean productivity patterns over the twenty-year period. The spatial distribution of these clusters is presented in figure 1.

Figure 1. Soybean productivity clusters in Rio Grande do Sul, Brazil, based on average municipal productivity during the period 2004–2023



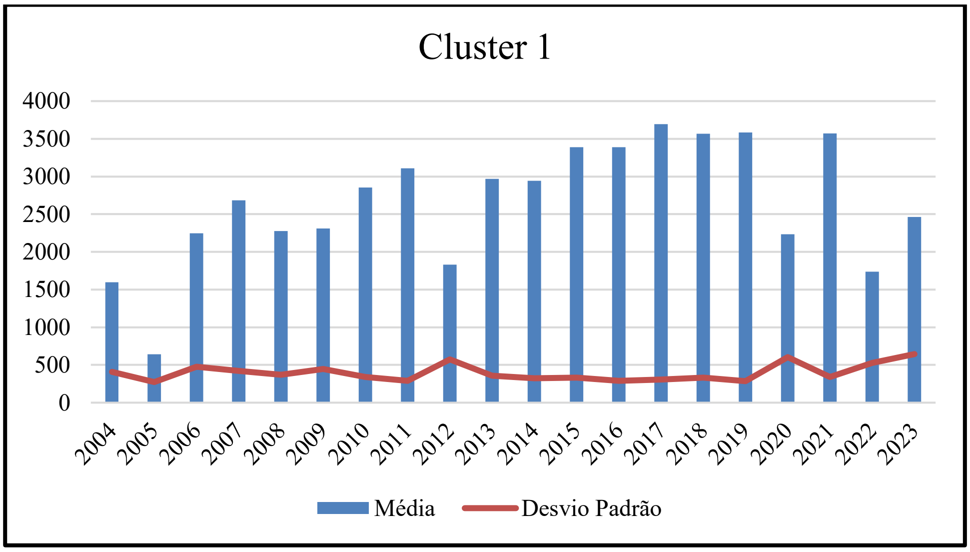
Fonte: Elaborado a partir de dados do Instituto Brasileiro de Geografia e Estatística (IBGE, 2025).

3.2.1 Cluster 1 – Consolidated High-Productivity Region

Cluster 1 comprises 196 municipalities and represents the group with the highest average productivity levels and the greatest productive potential in the state. Between 2004 and 2023, municipalities within this cluster recorded an average soybean productivity of 2,655.1 kg ha^{-1}, which is 12.5% higher than the state average of 2,359.6 kg ha^{-1}. The mean standard deviation was 809.42 kg ha^{-1}, indicating moderate variability among municipalities within the cluster.

The analysis of Figure 2, which presents the annual mean productivity and standard deviation values, reveals that this is the only cluster that consistently exceeded the threshold of 3,500 kg ha^{-1} during several years of the study period. This performance highlights the superior productive suitability of the municipalities belonging to Cluster 1, establishing this region as the most productive soybean-growing area in Rio Grande do Sul.

Figure 2. Mean soybean productivity and standard deviation (kg ha⁻¹) for Cluster 1, Rio Grande do Sul, Brazil, 2004–2023.



Source: Authors' elaboration based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

The results observed for Cluster 1 are geographically consistent with the municipalities comprising the region designated as PRA in the study by Melo, Fontana, and Berlato (2004). According to these authors, this region exhibits the highest soybean productivity levels in Rio Grande do Sul, a pattern attributed to favorable climatic conditions, soil characteristics, and crop management practices. Although the authors identify thermal limitations during specific stages of the soybean growth cycle as the primary constraint on yield potential, municipalities within this cluster consistently outperform those in the other groups analyzed, highlighting the region's high agricultural suitability for soybean cultivation.

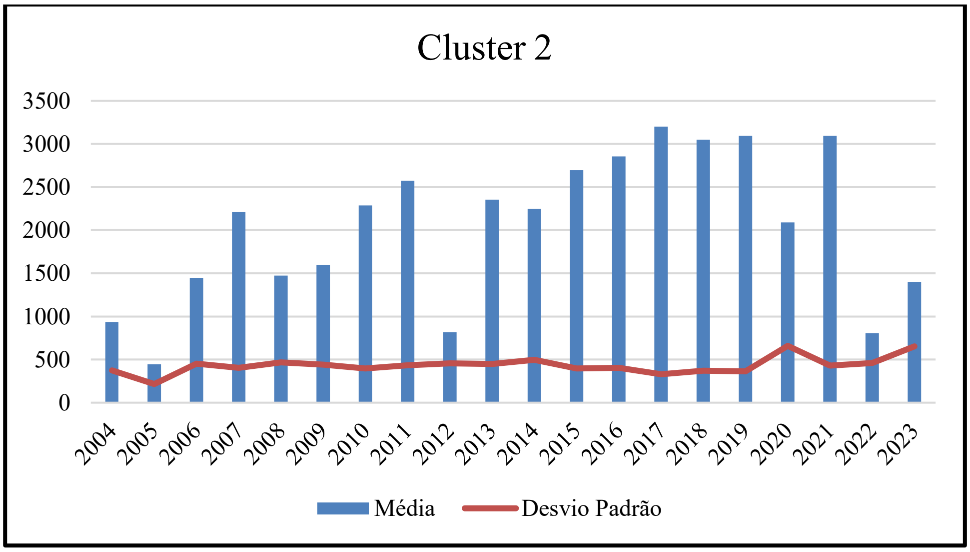
3.2.2 Cluster 2 – Consolidated Medium-Productivity Region

Cluster 2 comprises 103 municipalities and exhibits substantial variability in soybean productivity throughout the period from 2004 to 2023. The mean standard deviation of 866.63 kg ha⁻¹ is among the highest observed across all clusters, indicating considerable differences in productivity levels among municipalities within the group.

Although this cluster displays the second-highest productive potential in the state, with productivity levels exceeding 3,000 kg ha⁻¹ in several years, as illustrated in Figure 3, its average productivity over the entire study period was 2,033.6 kg ha⁻¹, representing a value 13.8% below the state average of 2,359.6 kg ha⁻¹. Consequently, despite its high yield potential, Cluster 2 ranked third among the seven clusters in terms of average productivity.

While Cluster 2 demonstrates high productivity potential under favorable conditions, its municipalities appear to be more susceptible to adverse climatic events, particularly droughts, than those belonging to other clusters. This greater climatic vulnerability is reflected in recurrent yield losses that alternate with periods of exceptionally high productivity. As a result, production stability is reduced, contributing to the pronounced variability observed in the cluster's productivity indicators.

Figure 3. Mean soybean productivity and standard deviation (kg ha^{-1}) for Cluster 2, Rio Grande do Sul, Brazil, 2004–2023.



Source: Authors' elaboration based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

The results observed for Cluster 2 are consistent with the findings reported by Melo, Fontana, and Berlato (2004). In their study, the authors identified that municipalities located in the western portion of the northern region of the state, corresponding to the area classified as PRc, exhibit geographic characteristics similar to those observed in Cluster 2. According to the authors, the primary constraint on soybean productivity in this region is associated with water deficits, which significantly reduce the crop's productive potential. This condition may help explain the high productivity variability observed within the cluster, characterized by alternating periods of high yields and years marked by substantial productivity losses.

3.2.3 Cluster 3 – Consolidated Low-Productivity Region

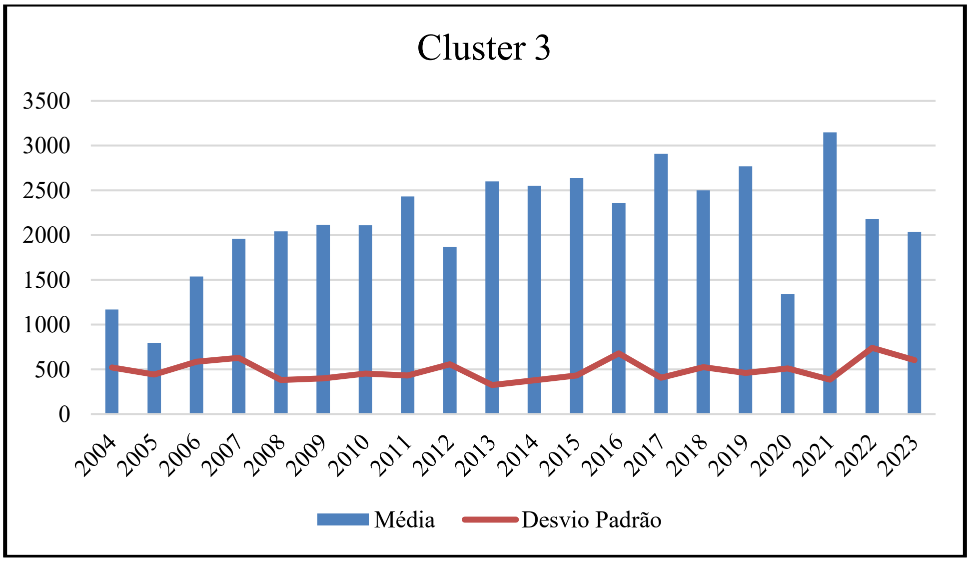
Cluster 3 comprises 85 municipalities and is distinguished by exhibiting the greatest productivity homogeneity among the major clusters, with a mean standard deviation of 596.94 kg ha^{-1}, the lowest value among all clusters analyzed. Average soybean productivity within the group was 2,152.3 kg ha^{-1}, which is 5.9% higher than that observed for Cluster 2 (2,033.6 kg ha^{-1}), although still 8.8% lower than the state average of 2,359.6 kg ha^{-1} during the same period.

Despite presenting a higher average productivity than Cluster 2, the yield potential of

Cluster 3 appears to be somewhat lower. While municipalities belonging to Cluster 2 frequently achieved average productivity levels exceeding 3,000 kg ha^{-1} during years with favorable climatic conditions, municipalities within Cluster 3 reached this threshold only once, in 2021. During years characterized by favorable growing conditions, municipalities in this cluster typically achieved average productivity levels close to 2,500 kg ha^{-1}, indicating a more stable but less pronounced production performance in terms of maximum yield potential.

The lower yield ceiling observed in Cluster 3 is consistent with the findings reported by Torres et al. (2023), who identified the Southeast and Southwest mesoregions as the areas experiencing the greatest expansion of soybean cultivation in recent decades, particularly through the conversion of land previously used for livestock production. Complementarily, Oderich, Elias, and Waquil (2019) emphasize that regions historically considered unsuitable for soybean cultivation, such as the Southeastern Highlands (Serra do Sudeste) and the Campanha region, have increasingly incorporated soybean production as an important economic activity following the recent expansion of the agricultural frontier. Although this process has increased the crop's economic relevance within these regions, the results suggest that these areas still face productive limitations when compared with the state's traditional soybean-producing regions, which is reflected in the comparatively lower yield potential observed for this cluster.

Figure 4. Mean soybean productivity and standard deviation (kg ha^{-1}) for Cluster 3, Rio Grande do Sul, Brazil, 2004–2023



Source: Authors' elaboration based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

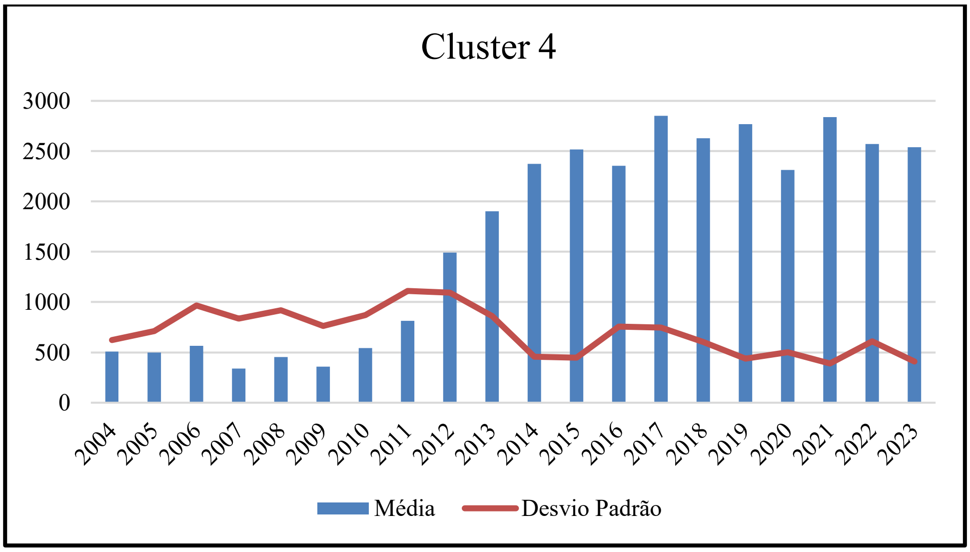
3.2.4 Cluster 4 – Emerging Soybean Agricultural Frontier

Cluster 4 represents a region that can be characterized as an emerging agricultural frontier for soybean cultivation in Rio Grande do Sul. Of the 19 municipalities included in this cluster, only nine reported soybean production in 2004. Between 2007 and 2008, soybean cultivation was recorded in only three municipalities. Beginning in 2010, however, soybean production resumed and expanded substantially, becoming established in all 19 municipalities by 2014 and remaining present throughout the subsequent years of the study period. The cluster recorded an average productivity of 1,661.65 kg ha^{-1} between 2004 and 2023, with a mean standard deviation of 1,014.31 kg ha^{-1}.

The results observed for Cluster 4 highlight the incorporation of municipalities traditionally associated with irrigated rice production, such as Uruguaiana and Santa Vitória do Palmar, as well as municipalities located along the state's coastal zone, into the ongoing process of soybean expansion. These regions, historically regarded as marginal for soybean cultivation, have increasingly become part of the territorial expansion process previously identified in the West-Central, Southeast, and Southwest mesoregions of Rio Grande do Sul by Oderich, Elias, and Waquil (2019), Costa et al. (2020), and Torres et al. (2023).

The expansion of soybean cultivation into these municipalities has contributed to the reconfiguration of the regional production structure, reflecting the growing economic attractiveness of soybean production in both domestic and international markets, as highlighted by Silveira, González, and Fonseca (2017). The relatively low average productivity combined with high variability suggests that this cluster is currently undergoing a process of productive consolidation, characteristic of regions recently incorporated into the soybean production system.

Figure 5. Mean soybean productivity and standard deviation (kg ha^{-1}) for Cluster 4, Rio Grande do Sul, Brazil, 2004–2023.



Source: Authors' elaboration based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

3.2.5 Cluster 5 – Late Soybean Agricultural Frontier

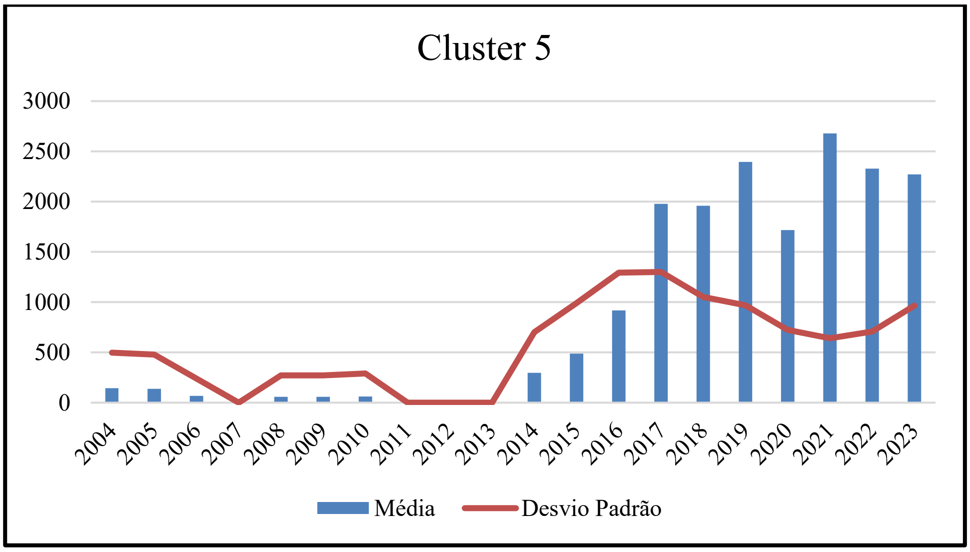
Similar to Cluster 4, Cluster 5 also represents an emerging agricultural frontier for soybean cultivation in Rio Grande do Sul. However, the incorporation of soybean production occurred even later in this group. Of the 21 municipalities included in the cluster, only two reported soybean production in 2004. By 2007, soybean cultivation had disappeared entirely from these municipalities. Between 2008 and 2010, a new attempt to establish soybean production was observed, but cultivation was once again discontinued, with no soybean production recorded between 2011 and 2013.

The definitive expansion of soybean cultivation began in 2014, when only three municipalities resumed production. It was not until 2021 that soybean cultivation became established in all 21 municipalities belonging to the cluster. During the study period, Cluster 5 recorded an average productivity of 877.72 kg ha^{-1} and a mean standard deviation of 1,026.02 kg ha^{-1}, reflecting the instability that characterized the introduction and subsequent consolidation of soybean production within the region.

The results observed for Cluster 5 further support the territorial expansion process described by Oderich, Elias, and Waquil (2019), Costa et al. (2020), and Torres et al. (2023). These authors emphasize that the increasing economic attractiveness of soybean production, combined with technological advances, enabled the incorporation of new production areas that had previously been considered marginal for soybean cultivation.

In this context, Cluster 5 represents a later stage of this expansion process. Whereas municipalities included in Cluster 4 experienced a more consistent consolidation of soybean production beginning around 2010, municipalities belonging to Cluster 5 only became effectively integrated into the soybean production chain after 2014, achieving greater productive stability only in the most recent years. These findings indicate that soybean expansion in Rio Grande do Sul occurred gradually and unevenly across space, reflecting different rates and stages of territorial incorporation into the soybean production system.

Figure 6. Mean soybean productivity and standard deviation (kg ha⁻¹) for Cluster 5, Rio Grande do Sul, Brazil, 2004–2023.



Source: Authors' elaboration based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

3.2.6 Cluster 6 – Sporadic Soybean Cultivation

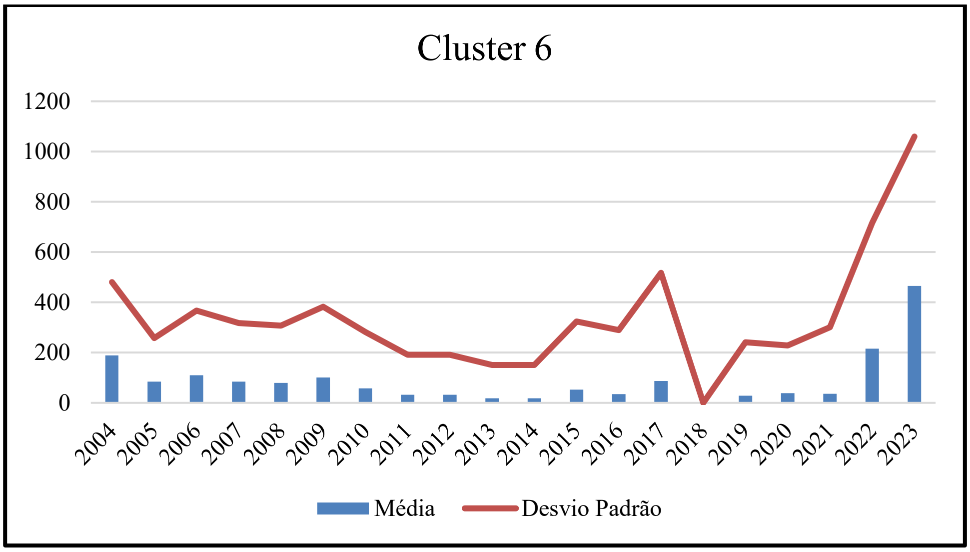
Unlike Clusters 4 and 5, which represent emerging agricultural frontiers at different stages of soybean production consolidation, Cluster 6 has not reached a similar level of development. Among the municipalities belonging to this group, soybean cultivation has failed to establish itself as a continuous or economically significant activity throughout the study period, occurring only sporadically and discontinuously.

Cluster 6 comprises 69 municipalities. In 2004, only 10 municipalities reported soybean cultivation; this number declined to five in 2009, one in 2013, and no soybean production was recorded in 2018. Beginning in 2019, however, a gradual resurgence of the activity can be observed, reaching 12 municipalities with soybean cultivation by 2023. Notably, 47 municipalities within this cluster did not report soybean production in any year of the period analyzed.

The municipalities comprising Cluster 6 are predominantly located in the Northeastern Highlands and the Northern Coastal regions of Rio Grande do Sul. These areas are characterized by relatively high population densities and a limited economic dependence on soybean production (Oderich, Elias, & Waquil, 2019). Furthermore, the region's topographical characteristics, marked by more rugged terrain and smaller farm sizes, constrain the expansion of mechanized soybean cultivation. These structural limitations contribute to the sporadic and discontinuous occurrence of soybean production observed throughout the study period.

The results suggest that, unlike the municipalities included in the emerging agricultural frontiers represented by Clusters 4 and 5, the territories belonging to Cluster 6 have not undergone a consistent process of integration into the soybean production system. Instead, soybean cultivation remains a marginal agricultural activity, with limited territorial expansion and low productive relevance within the local economic structure.

Figure 7. Mean soybean productivity and standard deviation (kg ha⁻¹) for Cluster 6, Rio Grande do Sul, Brazil, 2004–2023.

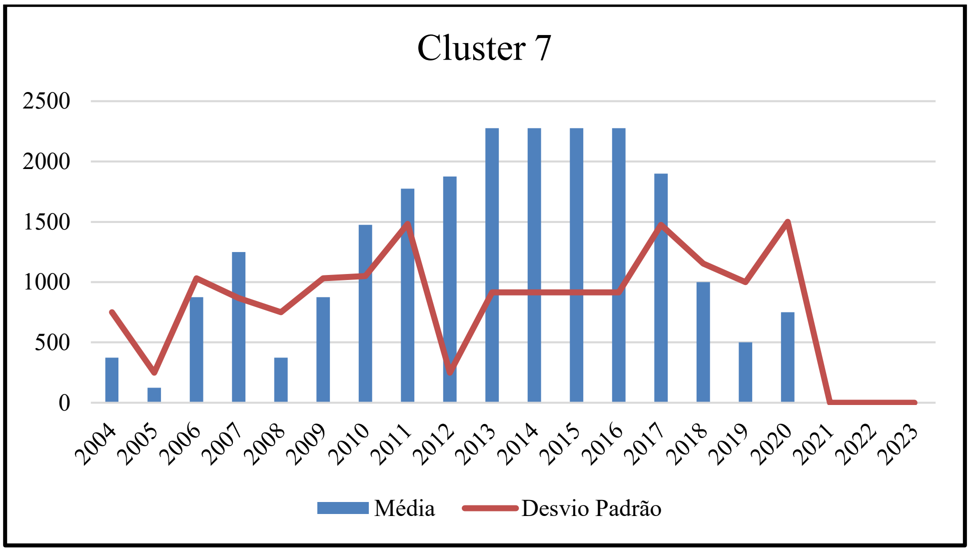


Source: Authors' elaboration based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

3.2.7 Cluster 7 – Declining Soybean Cultivation Region

Cluster 7 comprises only four municipalities and exhibits a pattern that contrasts sharply with those observed in the other clusters. Whereas the remaining groups indicate the consolidation, expansion, or initial adoption of soybean cultivation, this cluster is characterized by a process of contraction. During the early years of the study period, soybean cultivation was intermittent and unstable. However, between 2012 and 2016, all four municipalities consistently cultivated soybean, suggesting a temporary phase of consolidation. Beginning in 2017, a continuous decline in soybean cultivation became evident, ultimately culminating in the complete cessation of production across all four municipalities by 2021.

Figure 8. Mean soybean productivity and standard deviation (kg ha⁻¹) for Cluster 7, Rio Grande do Sul, Brazil, 2004–2023.



Source: Authors' elaboration based on data from the Brazilian Institute of Geography and Statistics (IBGE, 2025).

4 Conclusions

The analyses conducted in this study revealed important changes in the spatial structure of soybean productivity across Rio Grande do Sul over the last five decades. The ANOVA results indicate that soybean productivity was relatively homogeneous among municipalities during the 1970s, when the average standard deviation was 236.57 kg ha⁻¹. In contrast, recent decades have been characterized by a substantial increase in productive heterogeneity, with the average standard deviation reaching 716.95 kg ha⁻¹.

This growing disparity in productivity can be partly explained by the territorial expansion of soybean cultivation into regions with marginal suitability for the crop, as highlighted by Oderich, Elias, and Waquil (2019), as well as by the edaphoclimatic variability that imposes distinct production constraints across different regions of the state (Arsego et al., 2019).

Within this context, cluster analysis provided a spatially explicit representation of the productivity differences previously reported in the literature. The methodology enabled the identification of seven clusters with distinct productivity profiles. Three of these clusters (Clusters 1, 2, and 3) represent consolidated soybean-producing regions, although they differ substantially in terms of productivity levels and yield potential. Notably, the highest-performing cluster exhibited an average productivity level 30.4% greater than that observed in the lowest-performing consolidated cluster over the last twenty years.

The findings are consistent with those reported by Torres et al. (2023), who identified the Southeast and Southwest regions of Rio Grande do Sul—areas characterized by the most intensive soybean expansion into previously non-agricultural land—as exhibiting the lowest average soybean yields in the state. In contrast, the Northwest and East-Central regions achieved productivity levels above the state average. Similarly, Melo, Fontana, and Berlato (2004) documented substantial productivity differences among soybean-producing regions, with yield gaps of approximately 800 kg ha^{-1} between the most and least productive areas. Together, these findings demonstrate that spatial heterogeneity in soybean productivity is a persistent characteristic of the agricultural landscape of Rio Grande do Sul. The results of the present study further suggest that these disparities have not only persisted over time but have intensified as a consequence of soybean expansion and the increasing differentiation of regional production systems.

The remaining clusters reflect distinct stages of territorial incorporation into the soybean production system. Clusters 4 and 5 represent expanding agricultural frontiers, where soybean cultivation has increased due to favorable economic conditions and production opportunities. In contrast, Clusters 6 and 7 correspond to regions where soybean cultivation remains sporadic or is undergoing contraction, reflecting local agronomic, territorial, and economic constraints that limit the long-term viability of the crop.

From a practical perspective, the findings contribute to a better understanding of the spatial dynamics of soybean productivity in Rio Grande do Sul. The results provide valuable insights for regional agricultural planning, the development of public policies aimed at mitigating climate-related risks, and the formulation of territorial development strategies tailored to the productive characteristics of different regions.

This study has some limitations. The analysis was conducted exclusively using agricultural productivity data and did not directly incorporate climatic, edaphic, technological, or socioeconomic variables. Future research should therefore investigate the determinants of the productivity heterogeneity identified in this study by integrating these factors into explanatory models capable of identifying the underlying drivers of regional productivity differences. Such approaches may contribute to a more comprehensive understanding of the interactions among environmental conditions, technological adoption, and agricultural performance in soybean-producing regions.

5 References

Arsego, D. A., Ferraz, S. E. T., Streck, N. A., Cardoso, A., & Zanon, A. J. (2019). Indicadores climáticos e a produtividade de soja no Rio Grande do Sul. *Revista Brasileira de Meteorologia, 34*(2), 191–200. https://doi.org/10.1590/0102-77863340024

Costa, N. L., Santana, A. C., Coronel, D. A., Brum, A. L., & Mattos, C. A. C. (2020). Aspectos da importância do complexo soja no Brasil e no Rio Grande do Sul: 1997–2017. *Redes, 25*(4), 1840–1863. https://doi.org/10.17058/redes.v25i4.12735

Fávero, L. P. (2017). *Manual de análise de dados: Estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. Grupo GEN.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Multivariate data analysis* (6th ed.). Bookman.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2024). *Produção Agrícola Municipal (PAM)*. Retrieved May 25, 2025, from https://sidra.ibge.gov.br/tabela/1613

Lazzari, M. R. (2006). Agropecuária gaúcha em 2005: estiagem, câmbio e crise. *Indicadores Econômicos FEE, 33*(4), 41–50.

Leivas, J. F., Andrade, R. G., Victoria, D. C., Torresan, F. E., Vicente, L. E., Teixeira, A. H. C., Bolfe, E. L., & Barros, T. R. (2014). Avaliação do índice de vegetação padronizado no monitoramento indicativo de estiagens em períodos críticos da soja no Sul do Brasil. *Revista Brasileira de Cartografia, 66*(5), 1145–1155.

Melo, R. W., Fontana, D. C., & Berlato, M. A. (2004). Indicadores de produção de soja no Rio Grande do Sul comparados ao zoneamento agrícola. *Pesquisa Agropecuária Brasileira, 39*(12), 1167–1175. https://doi.org/10.1590/S0100-204X2004001200002

Oderich, E. H., Elias, L. P., & Waquil, P. D. (2019). Expansão do agronegócio no Brasil: Diferentes discursos e dinâmicas socioeconômicas no Rio Grande do Sul. *Eutopía: Revista de Desarrollo Económico Territorial, 16*, 95–114. https://doi.org/10.17141/eutopia.16.2019.4120

Oderich, E. H., & Waquil, P. D. (2019). *Municípios dependentes da soja no Rio Grande do Sul: Aspectos demográficos e socioeconômicos* (Texto para Discussão No. 33). Secretaria do Planejamento, Mobilidade e Desenvolvimento Regional do Rio Grande do Sul.

Rio Grande do Sul. (2018). *Genealogia dos municípios do Rio Grande do Sul*. Secretaria de Planejamento, Governança e Gestão.

Rio Grande do Sul. (2019). *Decree No. 54,572 of April 14, 2019*. Official State Gazette of Rio Grande do Sul.

Sentelhas, P. C., Battisti, R., Câmara, G. 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. *The Journal of Agricultural Science, 153*(8), 1394–1411. https://doi.org/10.1017/S0021859615000313

Silveira, V. C. P., González, J. A., & Fonseca, E. L. (2017). Mudanças no uso da terra depois do período de aumento de preço das commodities no Rio Grande do Sul, Brasil. *Ciência Rural, 47*(4). https://doi.org/10.1590/0103-8478cr20160647

Torres, R., Zorzi, A. L., Ozaki, V. A., & Martins, G. (2023). Expansão do cultivo de soja no Rio Grande do Sul: Seu efeito sobre a gestão de riscos e composição agrícola. *Desenvolvimento em Questão, 21*(59). https://doi.org/10.21527/2237-6453.2023.59.14086

APPENDICES

Figure 2. Cluster Analysis Dendrogram