IOSR Journal Of Humanities And Social Science (IOSR-JHSS)
Volume 29, Issue 11, Series 4 (November, 2024) 38-48
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Forecast Of Rainfed Agricultural Production In The State Of Pernambuco, Brazil

Elizama Cavalcante De Paiva [1], Erika Costa Sousa[1], Jair Andrade De Araujo[2], Jose De Jesus Sousa Lemos[3], Milena Monteiro Feitosa[4]
[1](Phd Student In Agricultural Economics At Federal University Of Ceará (UFC), Brazil)
[2](Associate Professor Of The Postgraduate Program In Rural Economy. Federal University Of Ceará (UFC), Brazil)
[3](Professor In Post Graduation Program Of Agricutural Economics. Department Of Agricultural Economics, Federal University Of Ceará (UFC), Brazil)
[4](Phd Student In Program Of Developmente And Environmente At Federal University Of Ceará (UFC), Brazil)

Abstract:
This research assesses rainfed agricultural production in the state of Pernambuco, which has 137 of its 185 municipalities officially characterized as being part of the semi-arid climate regime. The study classified rainfall periods in Pernambuco between 1944 and 2022 and investigated rainfall instability and its impact on aggregate productivity, aggregate harvested area and production value per hectare of bean, cassava and maize crops. The Auto-Regressive Integrated Moving Average model with Exogenous Variable Inputs (ARIMAX) was used, considering rainfall as an exogenous variable. The results showed that rainfall has a direct effect on the variables investigated, especially the harvested area. It was proven that in periods classified as normal and rainy, productivity, harvested area and production value per hectare showed higher results than those observed in the dry periods of, as was the assumption of the research. The study also confirmed that the models developed were parsimonious and capable of predicting the variables investigated. The results showed that productivities, harvested areas and production values per hectare for beans, cassava and corn experienced greater instability during periods of drought. The research also shows that the aggregated production values per hectare for the three crops are very low. In dry years, they are even lower than the minimum wage. This result confirms that the rural areas of Pernambuco where these crops are mostly grown are very poor.

***Key Word**: Climatic region of the Brazilian Semi-arid; resilience of family farming; Edaphoclimatic factors; Long periods of drought; High temperatures.*

Date of Submission: 01-11-2024 Date of Acceptance: 11-11-2024

I. Introduction

The climatic region of the Brazilian semi-arid region is characterized by high temperatures throughout the year, with a low temperature range, low relative humidity and rainfall instability, with the systematic occurrence of periods considered to be dry, according to the different definitions of this phenomenon [1,2,3].

From a technical point of view, the climatic regions classified worldwide as semi-arid, anchored in the aridity index (AI) based on the work of Thornthwaite (1948)[4], are those in which this AI varies between 0.20 and 0.50. This means that in semi-arid areas the soil is only capable of retaining between 20% and 50% of the rainfall that has occurred, given that the AI is calculated from the rainfall/potential evapotranspiration ratio in observations of at least 30 years. Potential evapotranspiration measures the evaporation of water precipitated by rainfall, through heat, wind and the transpiration of the remaining vegetation.

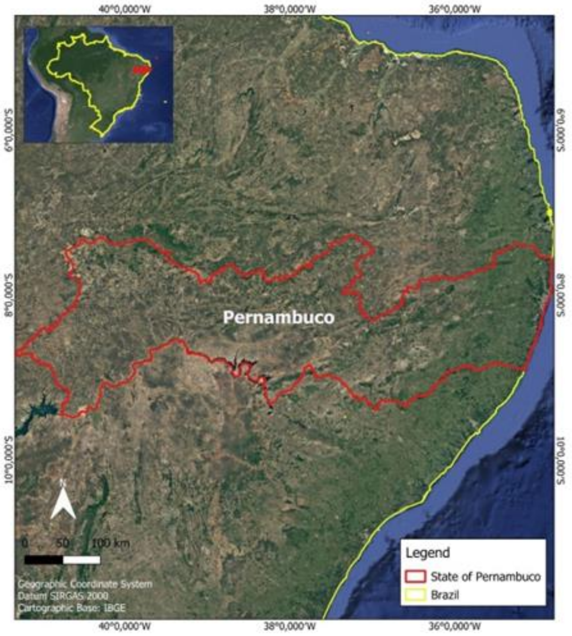
From a political point of view, the Brazilian semi-arid climate is based on the definition of the Deliberative Council of the Northeast Development Superintendence (Condel/Sudene)[5], which in December 2023 redefined the Brazilian semi-arid region to include 1477 municipalities. In this new redefinition, the state of Pernambuco now has 137 of its 185 (74%) municipalities forming part of this climate region

In the semi-arid region in general, and in the region where the 137 municipalities of Pernambuco are located in particular, rainfed crops prevail. These are activities that depend exclusively on rainfall. In general, the farmers who practise them do so without having access to technologies that could help them with the difficulties that arise when there is a shortage of water or when water availability falls short of what is required for the activities that provide them with food security and monetary income[6].

Among these activities, beans, cassava and corn are the most important in the state of Pernambuco. These crops are mostly grown by Pernambuco farmers, especially those located in family farming units[7].

Anchored in these foundations, this research seeks to answer the following questions: 1) Is it possible to classify the rainfall observed in the state of Pernambuco into dry, normal and rainy periods between the years 1944 and 2022, the period for which this research was designed?; 2) Does the rainfall instability observed in Pernambuco (Figure 1) in the period evaluated in this research (1944-2022) affect the predictive capacity of the production of these crops?.

Figure 1: Location of the State of Pernambuco on the map of Brazil



Source: IBGE (2022)

In order to answer these questions, the research has the following specific objectives: a) To classify the rainfall observed in Pernambuco from 1944 to 2022, the years for which IBGE provides information for the crops studied, into three specific periods: dry, normal and rainy; b) Evaluate the instability of Pernambuco's annual rainfall between the years 1944 to 2022; c) Create models that are capable of making aggregate projections of the harvested areas, yields and value of production per hectare of beans, cassava and corn between 1944 and 2022; d) To assess how the stability/instability of annual rainfall impacts on the projections of the variables harvested areas, yields and production values per hectare of beans, cassava and corn in Pernambuco during the period under investigation.

In addition to this introductory section, the research contains four more sections. The second shows a brief review of the literature on the semi-arid northeast. The third section discusses the source of the data and the econometric methods applied. The fourth section presents a discussion of the estimated results. Finally, the fifth section is dedicated to the final considerations

II. Dryland Production In Pernambuco's Semi-Arid Region

The Brazilian semi-arid region, characterized by its unique climatic conditions, faces significant challenges for agricultural production. The region, defined by irregular rainfall, long periods of drought and high temperatures, requires specific management strategies to ensure the sustainability of agricultural activities. These climatic particularities have a direct impact on production capacity and the quality of agricultural products, making it essential to understand the dynamics involved in order to implement practices that can mitigate the adverse effects of the climate.

Lemos (2015)[8] emphasizes that the Brazilian semi-arid region is defined by a unique climatic rhythm, marked by low rainfall, dry air, irregular distribution of rainfall concentrated at certain times and long periods of drought. In general, the distribution and seasonality of rainfall in the Brazilian semi-arid region are not uniform. In the north of the semi-arid region, the rainy season is concentrated between February and May. In the center-south and west, the rainy season is from December to February. On the eastern side, the rains appear from May to August (Marengo *et. al*., 2019)[9]. Travassos *et. al.* (2013)[10] define this alternation between dry and wet periods as the drought-rain binomial, a fundamental element in the identity of the semi-arid region.

The intrinsic characteristics of the region significantly affect agricultural production. Da Silva *et al.* (2019)[11] found that the hottest months for the state of Pernambuco caused stressful conditions for dairy cattle, with losses being observed in the quality and quantity of food provided to the animals, as well as a decrease in food consumption and lactation.

Duque (1980)[12] considers that farming is a high-risk activity, marked by frequent crop losses due to the irregularity of rainfall, both in terms of volume and spatial distribution. This climate variability, coupled with the prevalence of soils with low agricultural suitability, makes production a constant challenge.

Studies carried out by Paiva *et. al.* (2022)[13] have highlighted the interconnected effects of climate and the economy on livestock farming in Ceará. The research, using the ARIMA model, proved that rainfall instability and price fluctuations play a crucial role in the decisions of rural producers in the semi-arid region of Ceará.

Lessa (2023)[14] investigated the resilience of family farming in Paraíba in the face of the high rainfall variability observed in the state over 120 years (1945-2020). The analysis revealed that, although 40 years of this period were marked by droughts, the productivity and aggregate income per hectare of rainfed crops did not show a direct relationship with rainfall variability, suggesting the existence of resilience mechanisms that allowed farmers to maintain their income even during periods of drought.

It is therefore a fact that nature has an influence on agricultural production. Edaphoclimatic factors such as temperature, rainfall, luminosity and soil quality dictate the rhythm of harvests and the health of crops. Even so, in the face of adversity, family farming remains the basis of income for a large part of the rural population of the Semi-Arid [15,16]. This activity, although limited in its diversity and productivity, guarantees the survival and livelihood of thousands of families.

Therefore, understanding all the factors surrounding production is fundamental to developing public policies and management strategies that help producers deal with climatic and economic risks, increasing the resilience of agricultural activities and promoting the region's development.

III. Material And Methods

The database that underpins this work was drawn up from information extracted from the National Oceanic and Atmospheric Agency (NOAA, 2022)[17] regarding rainfall data for the period from 1901 to 2020 and from the Pernambuco Water and Climate Agency (APAC) for the years 2021 and 2022. Information on the crops studied for the years 1944 to 1973 was collected from the Brazilian Statistical Yearbooks (IBGE, various years). Data for the years 1974 to 2022 was taken from the Municipal Agricultural Survey[18]. The period in which the data was available spans from 1944 to 2022. The grains that make up the series studied are: beans, cassava and corn.

The variables analyzed in this research are defined as follows: $_{CHt}$ corresponds to the annual rainfall in the state of Pernambuco in year t (where t ranges from 1901 to 2022). $_{ARjt}$ refers to the aggregate harvested area in hectares (ha) in the state of Pernambuco, obtained through factor analysis of bean, cassava and corn crops in year t (with t ranging from 1944 to 2022). The variable $_{PDjt}$ indicates aggregate land productivity (kg.$^{ha-1}$) or simply aggregate productivity, in the state of Pernambuco, for bean, cassava or corn crops in year t, calculated by dividing the quantity produced by the area harvested, with subsequent aggregation via factor analysis. Finally, $_{PRjt}$represents the aggregate price (R$.Kg), in values corrected for 2023, of the crops studied in year t, obtained by dividing the value of production by the quantity produced of the crop in each year.

Methodology applied to achieve objective “a”

For the classification of rainfall, the methodology used by Lemos and Bezerra (2019)[19], Lessa (2023)[14] was followed, where the oscillations of half the standard deviation (SD) around the average rainfall observed over the years analyzed are considered, as shown in Table 1:

Table no 1: Classification of rainfall (dry, normal and rainy) according to mean and standard deviation from 1944 to 2022

Period	Range of variation
Dry	Rainfall < Average over the period – 0,5 Standard deviation (SD)
Normal	Rainfall = Average over the period ± 0,5 Standard deviation (SD)
Rainy	Rainfall > Average over the period + 0,5 Standard deviation (SD)

Source: Lemos e Bezerra (2019)

Separating periods based on classification alone does not guarantee complete consistency. According to Lemos and Bezerra (2019)[19], this approach should be validated by applying the regression method, using dummy variables to check whether the rainfall averages estimated for each rainy season are statistically different.

According to Valle and Rebelo (2002)[20], the significant impact of including dummy variables in econometric analysis is evident. These variables are used to assess the relationship between a dependent variable and an independent variable, and are a common statistical technique in regression modeling. As discussed by Gujarati and Porter (2011)[21], a model that incorporates dummy variables, also known as binary variables, can be described as follows:

$$Y_t = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \mu_t \qquad (1)$$

Such that,
D1 and D2 are binary variables (*dummies*) established as follows:
D1 = 0 in the dry and rainy years;
D1 = 1 in years of normality;
D2 = 0 in the dry and normality years;
D2 = 1 in the rainy years ;
D1 = D2 = 0 in the dry years.

In this context, when applying these criteria, the expectation is that the average rainfall during the rainy years will be higher, both numerically and statistically, than the years with normal rainfall, which in turn should be higher than the rainfall during the periods of rainfall scarcity in the state of Pernambuco, from 1944 to 2022.

Methodology applied to achieve objective "b"

The level of stability/instability of the rainfall averages for each defined period will be analyzed using the coefficient of variation (CV). According to Gomes (1985)[22] and Garcia (1989)[23], the higher the CV, the more heterogeneous or unstable the variable studied, while the lower the CV, the more homogeneous or stable the rainfall during the period under investigation. Gomes (1985)[22] classified the CV according to Table 2:

Table no 2: CV classification according to its amplitude

Classificação do CV	Amplitude do CV
Low	CV < 10%
Middle	10% ≤ CV < 20%
High	20% ≤ CV < 30%
Very high	CV ≥ 30%

Source: Gomes, 1985.

Thus, choosing to use the coefficient of variance in this evaluation method instead of other measures of variability makes it possible to compare disparities or instabilities between variables measured in different units 24,23,25,26,27,28,29,30.

Methodology applied to achieve objective "c"

To estimate aggregate projections for bean, cassava and maize crops in Pernambuco in terms of harvested areas, productivity and production value per hectare between 1944 and 2022.

The same technique as Lessa (2023)[14] will be used to aggregate the different crops in terms of harvested area and productivity: Factor Analysis (FA). Using FA, a weighted average of the aggregate harvested area and aggregate productivity will be constructed, the weights of which will be defined using expression (2):

$$M\text{Aij}=(P1\times\text{APF}j) + (P2\times\text{APM}j) + (P3\times\text{APML}j) \qquad (2)$$

Where:
MAij consists of the aggregate (weighted) average of harvested area/productivities (i) in year j;
P1, P2 and P3 are the weights associated with the harvested area/productivity of the crops;
APFj is the weighted aggregate of the harvested area/productivity of beans in year j;
APMj is the weighted aggregate of the harvested area/productivity of cassava in year j;
APMIj is the weighted aggregate of the harvested area/productivity of corn in year j.

For the aggregate of crop incomes, which consists of the value of production per hectare of harvested area, it will be obtained through the linear sum of the production values for each of the crops. Therefore, in order to assess the trajectories of the three variables, as well as the effects of rainfall on the projections, it is necessary to understand time series.

Methodology applied to achieve objective "d"

According to Enders (2009)[31], a time series is a collection of values, ordered in time {Xt}n t−1 = {X1, X2,..., Xn}, which represent the change of a random variable over regular intervals. The main objective of time series analysis is to make forecasts. Box, Jenkins and Reinsel (1994)[32] emphasize that this methodology establishes mechanisms in which future values of a series can be predicted based solely on its present and past values.

Some fundamental pillars are needed to understand and build accurate forecasting models. Among these foundations is the random or stochastic process, defined as a set of observations of random variables ordered in time. According to Gujarati and Porter (2011)[21] and Wooldridge (2013)[33], a stochastic process is considered stationary when its mean, variance and autocorrelation structure do not change over time, i.e. they remain constant over time. This fundamental property ensures that the statistical characteristics of the process do not change over time, allowing for a more robust and reliable analysis. So our first step is to analyze the stationarity of the series.

Unit root test

The unit root test has become popular and widespread in recent years as a way of detecting the stationarity of a series. The starting point for analysis is the (stochastic) unit root process:

$$Y_t = \rho Y_{t-1} + \xi_t \qquad (3)$$

Where Y_t is the time series analyzed, e is the value lagged by one period, and t is the error term. Knowing that it varies from -1 to 1. By subtracting from both sides, the following is obtained :

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + \xi_t \\ Y_t - Y_{t-1} &= (\rho - 1) Y_{t-1} + \xi_t \\ \Delta Y_{it} &= \rho_i Y_{it-1} + z'_{it} Y + \xi_{it} \end{aligned} \qquad (4)$$

Where, is the first price difference; is the lagged price, when for each individual and periods, is the deterministic component, which could be zero, one, the fixed effect, or the fixed time trend effect. In simple series models, vs. is tested. For = 0, it is assumed that = 1. Therefore, the hypotheses are:

$$\begin{aligned} &H_0: \rho_i = 0 \text{ vs.} \\ &H_1: \rho_i < 0, \text{for all i } = 1,2,\ldots,\text{N.} \end{aligned} \qquad (5)$$

The purpose of the Augmented Dickey-Fuller (ADF) test, equation 04, is to determine the order of integration of the series. If the integrated series is of order zero, no modification is needed to apply the regression. However, if the null hypothesis is rejected, differentiation will have to be made in order to obtain the stationarity process, given the criterion and the number of lags (Enders, 2009)[31].

$$\Delta y_{it} = \mu_i + \rho y_{i,t-1} + \sum_{j=1}^{k} \theta_{ik} \Delta y_{it-k} + \beta_i t + \xi_{it} \qquad (6)$$

Where is the first differences operator; is the lag length; is a unit specific fixed effect; is a deterministic trend. The subscript indexes the cross-sectional units. Under the null hypothesis, there is a unit root, while the alternative hypothesis indicates that there is no unit root [21,34].

If the series Y_t, which in this study can take the values of harvested areas, productivity (yield), production values per hectare, is stationary, ξ_t is endogenous white noise. Although it is endogenous white noise, this study assumes that ξ_t is affected by the exogenous variable, precipitation (X_t). This is because it is assumed that the unstable temporal distribution of rainfall could cause the residual (ξ_t), which makes the predicted value Y_P different from the observed value of Y_t, to be affected by this exogenous variable (X_t), which is unpredictable for cereal production decisions. This can be seen in equation (7):

$$\xi_t = f(X_t) \qquad (7)$$

Thus, the Y_t series can be defined according to expression 06, where Y_t is the observed series, Y_p the predicted series and $f(X_t)$ the impact of the rainfall.

$$Y_t = Y_P + f(X_t) \qquad (8)$$

Types of Box and Jenkins Models:

i) Moving average (MA) models: these are those in which $\phi(B) = 1$ and are said to be MA(q);

ii) Auto regressive (AR) models: those in which $\theta(B) = 1$ and are said to be AR(p). These models are so called because Y_t, at time t, is a function of the values of this variable at times prior to t;

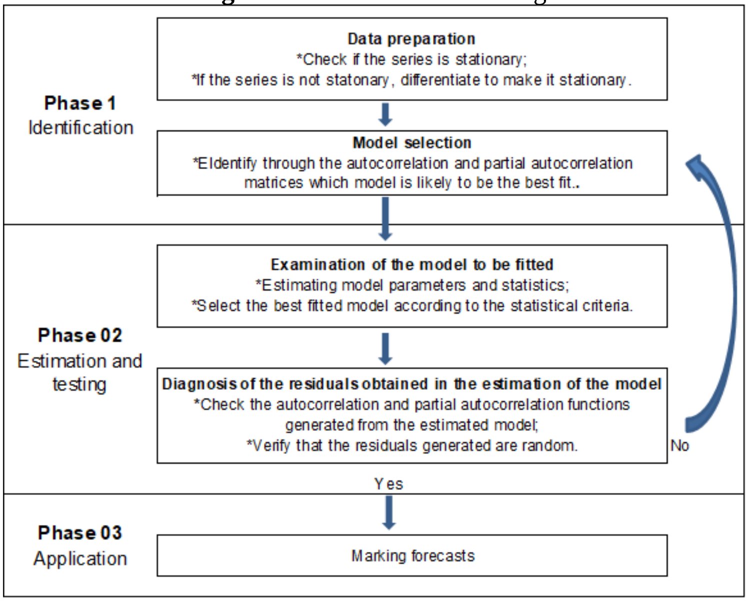
iii) Auto regressive-moving average (ARMA) models are those that have an AR part with an MA part, and will have the notation ARMA(p,q);

iv) Auto-Regressive Integrated Moving Average (ARIMA) models are those known as integrated ARMA models or ARIMA is an extension of the ARMA model class to include the "d" ARIMA differentiation (p,d,q).

It should be noted that if it is necessary to differentiate a time series "d" times in order to make it stationary and apply the ARMA (p, q) model to it, the original time series is said to be ARIMA (p,d,q), as shown in item IV. Therefore, the integrated autoregressive time series of moving averages, where "p" denotes the numbers of autoregressive terms, "d" the number of times the series must be differentiated before becoming stationary and "q" the number of moving average terms [35,21,34].

To build the ARIMA model that best suits the study, it is necessary to follow the stages of the Box-Jenkins methodology (Makridakis *et. al*, 1998)[36] shown in figure 01. There are three phases, the first of which is known as the identification phase, in which the model is selected by analyzing the data. The second, in which the model parameters are estimated and the acceptance criteria checked. The third and final stage, called application, is when the model is used in forecasting tasks (Figure 2).

Figure 2 - Time series modeling.



Source: Prepared by the authors based on Gujarati and Porter (2011).

In the data preparation stage, it will be checked whether the series is stationary; if it is not, a transformation will be made to make it stationary using the successive differences procedure. Stationarity is then checked using the*Augmented Dickey-Fuller*(ADF) unit root test. Next, the autocorrelation (FAC) and partial autocorrelation (FACP) functions of the series are calculated, in addition to the graphical analysis which will allow the ARIMA model (p, d, q) to be selected.

Once the identification stage is complete, the model parameters are estimated. The d parameter refers to the number of times the difference between the elements of the series was taken until it became stationary. Calculating the autoregressive parameters p and moving average q involves analyzing the FAC and FACP functions respectively, as the FAC function will show the peaks that identify p, while the FACP function will show the peaks that identify the value of q. Finally, the result of the ARIMA (p,d,q) model is obtained.

ARIMAX model

The ARIMAX model is a generalization of the ARIMA method with the inclusion of exogenous variables (Box and Tiao, 1975)[37]. In this study, the exogenous variable used is the annual rainfall observed for the semi-arid regions of the state analyzed.

The ARIMAX model is considered multivariate because it adds a linear component to the observations of exogenous variables. The main difference between this model and ARIMA is that ARIMAX has, in addition to the autoregressive and moving average parameters, the input of exogenous and linear variables [3,35,38,39].

The ARIMAX model can be understood as a combination of the Auto-Regressive AR(p), Integrated (d), Moving Average MA(q) and Exogenous X(r) models, and can therefore be symbolized as ARIMAX (p,d,q,r). In this study, the exogenous variable to be inserted is the annual rainfall observed in Pernambuco, assuming the hypothesis that temporal instabilities in rainfall affect the forecasting capacity of the variables that define the annual production of beans, cassava and corn. A simplified way of mathematically representing this model in a generalized form is described in equation (9), which is the application of equation (8) to this research.

$$Y_t = [\rho + \Sigma\beta_j Y_{(t-i)} + \Sigma\theta_j \,Є(_{(t-i)})\,] + [\Sigma\omega_j X_{(t-i)} + \xi_t] \qquad (9)$$

In this research, the variables Y_t to be predicted as Y_p are: harvested area, productivity and production value per hectare of grains in the state of Pernambuco between the years 1944 and 2022. The exogenous variable (X_j) is annual rainfall.

Identification of the model

Testing the quality of the adjustments

Still in phase 2 of estimation and testing, after defining the appropriate values, it is necessary to analyze the diagnosis of the residuals (ξ_t), which must be white noise. When choosing statistically appropriate models, one of the criteria is to look for more parsimonious models, with the view that the smaller the number of regressors, the better the adjustment model.

In addition to this, other criteria are used to assess the suitability of the adjusted models for the research objectives. These are coefficient of determination (R²), which assesses the percentage of variation in the analyzed variable that is explained by the structured model; Pearson's correlation coefficient, which is able to verify the proximity between the observed values and the values predicted by the model; percentage of mean absolute error (MAPE) which considers the relative error of each forecast, with the aim of comparing the values predicted by the model with the observed values of the series, characterizing the forecasting capacity of the model adopted; Schwartz's Bayesian criterion (BIC) and the Ljung-Box Q statistic which must be non-significant at a significance level of at least 10% [35,33]. Once these steps have been completed, the forecasting model will be used and explored. The estimates will be made using the Statistical Package for the Social Sciences (SPSS) software version 20.0 [36,33,40,35,34,21].

IV. Result And Discussion

Results found for the first and second objectives

The average rainfall for the period between 1944 and 2022 was 898.75mm. The minimum rainfall (486.10mm) occurred in 1993. According to IPEA (2009), one of the biggest water rationings in history took place in Pernambuco at that time, when the metropolitan region only received piped water once a week. The maximum rainfall was recorded in 1985 (1,506.68 mm). The CV was 22.06%, suggesting that the rainfall observed for the state of Pernambuco during the period evaluated by this study was highly unstable, on the scale defined by Gomes (1985)[22]. Table 3 shows a summary of these results.

Table no 3: Descriptive statistics for the defined rainfall periods

Period	Range of variation (mm)	N° of years	Average	CV (%)
Total	-	79	898,75	22,06
Dry	Rainfall < 796,27	27	707,72	12,24
Normal	796,27 > Rainfall < 983,12	31	891,83	6,27
Rainy	Rainfall > 983,12	21	1.154,57	12,48

Source: Survey results (2024).

In the 79-year analyzed series, it was possible to identify 27 years that were considered “dry”, with an average annual rainfall of 707.72 mm and a CV of 12.24%, which is considered midle The “rainy” period was identified with a volume of rain greater than 983.12 mm, counting 21 years, with an average annual rainfall of 1,154.57 mm and a CV characterized as midle (12.48%). The largest number of years (31) was characterized as “normal” with an annual average of 891.83 mm and a low coefficient of variation (6.27%).

Table 4 shows the results of the tests carried out to assess whether the averages of the rainy periods created for the research are statistically different. These results confirm the assumptions that guided this research, that the average rainfall observed in Pernambuco during the period analyzed can be classified into dry, normal and rainy periods.

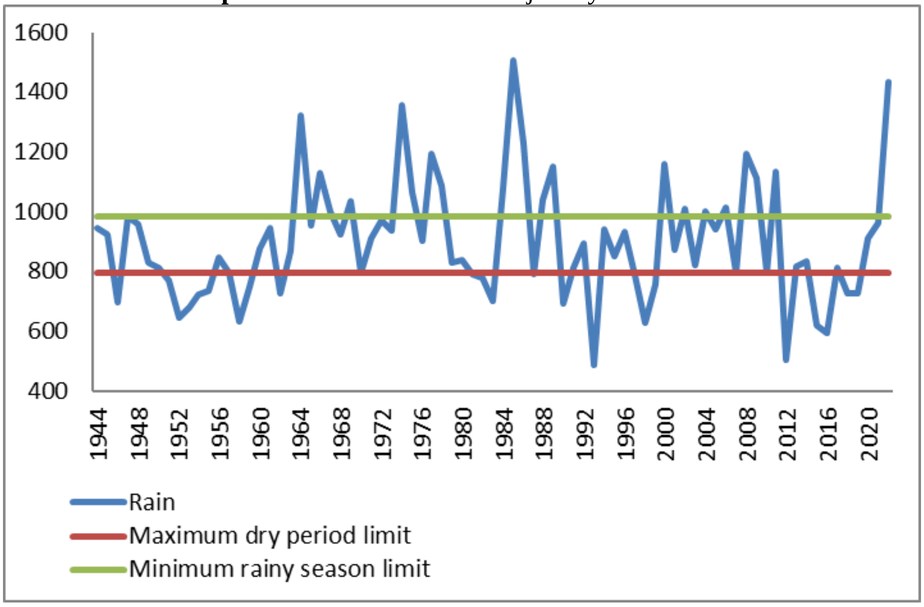
Table no 4: Result of the comparison between the average rainfall of the periods defined in the survey

Varible	Coefficients	T Estatístics	sign.	adjusted R²
Constant	707,72	38,178	0,000*	0,764
D1	184,10	7,261	0,000*	
D2	446,85	15,944	0,000*	

Source: Research results (2024) .*significant up to 1% significance level

The evidence presented in Table 4 shows that the average rainfall observed for the years that make up the dry period was 707.72 mm. For the years in which rainfall is classified as normal, the average was 891.82 mm. In the years classified as rainy, the average rainfall was 1,393.12 mm. Graph 1 shows the observed annual rainfall for the state of Pernambuco between 1944 and 2022.

Graph 1 - Pernambuco Rain Trajectory 1944 to 2022



Source: Survey results (2024).

Results found for the third and fourth objectives

The data relating to aggregate productivity and aggregate harvested area were defined based on the weighted average of bean, cassava and corn crops and their respective weights, obtained from the factor analysis scores shown in Table 5.

Table no 5: Factor scores and weights estimated and used to define aggregate productivity by weighted average

Variables	Productivity (kg^{-ha})		Harvested áarea (ha)	
	Component scores	Weights	Component scores	Weights
Beans	0.394	0.345	0.408	0.337
Cassava	0.350	0.306	0.327	0.270
Corn	0.398	0.348	0.475	0.393
Total	1.142	1.000	1.206	1.000

Source: Survey results (2024).

The aggregate income per hectare comes from the linear sum of the production values per hectare for each crop. After this stage, the ARIMAX model was used, considering rainfall as an exogenous variable.

Table 6 shows the models used to forecast the agregated productivies, aggregate harvested area and agregated value of production per hectare of beans, cassava and corn corn in Pernambuco between 1944 and 2022. The models used are ARIMAX with different parameter configurations. The results are organized into three categories: Aggregate Productivity, Aggregate Harvested Area and Aggregate Value/hectare.

The evidence shown in this table 6 shows that only the series of aggregate harveste areas of the crops studied was originally non-stationary. The series was stationarized using differentiation.

It can be seen that all the statistics that measure the quality of the adjusted models make it possible to use them to make projections of the variables studied. The coefficients of determination ranged from 0.504 to 0.716. The Ljung Box statistics showed that the residuals associated with the estimated models are white noise. The MAPE statistics ranged from 7.967 to 19.741, which means they were low. The BIC statistic ranged from 2.558 to 20.977, also low, indicating that the models were also parsimonious. The ratio between the observed and forecast series ranged from 0.711 to 0.868, which indicates high correlations between these two series.

The results also confirmed the assumption guiding this study that the rainfall instabilities observed between 1944 and 2022 interfered with aggregate yields, aggregate harvested areas and aggregate production values per hectare of beans, cassava and maize.

Table no 6: Adjusted models to forecast productivity, harvested area and value of production per hectare (aggregate) for bean, cassava and corn crops in Pernambuco between 1944 and 2022

	Agregated Productivity	**Agregated area**	**Agregated Value/hectare**
	ARIMAX (1.0.0.1) **(no transformation)**	**ARIMAX** (1.1.0.1) **(no transformation)**	**ARIMAX** (1,0,0,3) **(Natural log transformation)**
Model (p, d, q, r)	1.0.0.1	1.1.0.1	1.0.0.3
Constant	2755.129*	0.000	0.000
AR (1)	0.686*	-0.344*	0.708*
MA (1)	0.000	0.000	0.000
Rainfall	0.737*	92.521*	0.363*
Delay	0	0	3
Diference	0	1	0
R²	0.508	0.716	0.504
Ljung Box Statistics	11.541^{NS*}	20.977^{NS**}	23.53^{NS***}
MAPE Statistics	7.967	15.009	19.741
Normalized BIC	11.896	20.977	2.558
Pearson Correlation (r)	0.713*	0.851*	0.711*

Source: Survey results (2024). Notes: *significance level up to 1%; NS*:not significant at least to to 58% of significance: NS**: not significance at least to 22% of significance; NS*** (not significant level at least to 15%

A Table 7 shows the estimated averages values to aggregated productivity, aggregated harvested area and aggregated value per hectare (USD of december 2022) of bean, cassava and corn in different periods defined in this study (dry, normal and rainy), as well as the coefficients of variation (CV) for each variable and period analyzed.

Table 7 also shows the average equivalents of production values per hectare, valued in 2022 annual minimum wages. As you know, the production value per hectare can be understood as an approximation of the gross income per hectare generated by the three activities studied.

Table no 7: Effects of annual rainfall fluctuations on average in USD of December 2022 and annual minimum wage (MW) from 2022 of agregated productivity, agregated harvested area and agregated value per hectare in Pernabuco between 1974/2022

	Aggregated productivity		**Aggregated harvested area**		**Aggregated Value**		
Period	Average (kg^{ha-1})	CV (%)	Average (ha)	CV (%)	Average (USD/ha^{-1})	Average (MW)	**CV (%)**
Dry	3197.88	17.42	164499.60	24.98	2271,74	0.81	31.88
Normal	3511.15	13.00	176556.31	36.97	2456,95	1.67	35.57
Rainy	3574.90	6.54	243299.91	23.23	2503,43	1.81	46.06

Source: Survey results (2024).

Confirming the assumptions on which this research was based, the average aggregate productivity tends to be higher in the rainy (3574.90 kg^{ha-1}) and normal (3511.15 kg^{ha-1}) periods compared to the dry period (3197.88 kg^{ha-1}). It can be seen that the coefficient of variation is higher in the dry period (17.42%), indicating greater variability in productivity during this period. It is therefore suggested that productivity is more sensitive to drought confirming the supositions made to constructo this study (Table 7).

The average aggregate harvested area is lower in the dry period (164499.60 hectare) compared to the normal (176556.31 hectare) and rainy (243299.91) periods. The estimated CV for the areas harvested in normal rainfall years was classified as very high, according to Gomes (1985). For the rainy and drought periods, the CVs were classified as high, according to that author. These results suggest that the areas harvested, which are generally smaller than those planted, were very sensitive to rainfall instability in the state of Pernambuco during the period evaluated. (Table 7).

The production values per hectare estimated for the three periods are quite low. It can be seen that in the rainy years, the aggregate annual value of production per hectare (USD2503,43) was the highest observed, and represented only 1.81 of the annual minimum wage measured in 2022 values. In drought years, this variable represented only 81% of the annualized minimum wage, with a value of just USD2271.74. (Table 7).

The evidences from Table 7 also show that the aggregated value of production per hectare of beans, cassava and corn in Pernambuco in the period from 1944 to 2022 were quite unstable, in all three periods in which rainfall was classified for the state. In fact, all the estimated CVs, were classified as very high according to the metric created by Pimentel Gomes (1985), ranged from 31.88% in dry years to 46.06% in rainy years (Table 7).

The variability, as indicated by the coefficients of variation, suggests that producers face significant challenges, especially during the dry season, when the the agregated value of production/ha and harvested area are more unpredictable (Table 7).

V. Conclusion

By analyzing rainfall data between 1944 and 2022, it was possible to classify rainfall into three distinct periods: dry, normal and rainy, each of which, given its specific characteristics, influences agricultural production and the socio-economic well-being of the region.

It was also possible to verify that the rainfall instability observed in Pernambuco during the period evaluated affects the predictive capacity of the production of these crops. Therefore, the study of rainfall variability and its impact on agricultural crops in Pernambuco revealed crucial insights for managing and adapting to climate change in the semi-arid region.

The ARIMAX models used proved to be effective in predicting the variability and impact of rainfall on agricultural variables used in this research. The exogenous variable showed an influence in all of the variables studied.

Therefore, the results confirmed that rainfall variability has a significant impact on productivity, harvested area and production value per hectare of bean, cassava and corn crops. During rainy periods, conditions are more favorable for increasing productivity, harvested área and value of production/ha.

The general conclusion of the research is that these activities, which occupy the majority of farmers in the state of Pernambuco, have very low values per hectare, which confirms the high level of poverty in which farmers in this northeastern Brazilian state survive.

The information obtained is valuable for producers, policymakers and water resource managers, providing a solid basis for developing adaptation and mitigation strategies. The implementation of rainfall resilient agricultural practices and the adoption of water management policies can minimize the negative impacts of rainfall variations, ensuring greater stability in production, food security and income for the State of Pernambuco.

Therefore, the research underlines the importance of continuous analysis of climatic variations and the adaptation of agricultural practices to climate variability, promoting sustainable development and mitigating the adverse effects of climate change.

References

[1]. Assad, E., Pinto, H. S. (2008). Aquecimento Global E A Nova Geografia Da Produção Agrícola No Brasil. Embrapa-Cepagri, São Paulo, Agosto, 82 P.

[2]. Cedeplar E Fiocruz. (2008). Mudanças Climáticas, Migrações E Saúde: Cenários Para O Nordeste Brasileiro, 2000-2050. Belo Horizonte: Cedeplar/Fiocruz, Relatório De Pesquisa, Julho).

[3]. Nunes, E. M. Et Al. (2014). Políticas Agrárias E Agrícolas No Contexto Do Desenvolvimento Do Nordeste: Evolução, Desafios E Perspectivas. Planejamento E Políticas Públicas, Brasília, N. 43.

[4]. Thornthwaite, C.W. (1948). An Approach Toward A Rational Classification Of Climate. Geográfical 38 (1), 55-94.

[5]. Sudene, Superintendência Do Desenvolvimento Do Nordeste. (2021). Delimitação Do Semiárido – 2021. Recife (Versão Preliminar). Available At: Https://Www.Gov.Br/Sudene/Pt-Br/Centrais-De-Conteudo/02semiaridorelatorionv.Pdf.

[6]. Bezerra, F. N. R. (2022). Avaliação Da Agricultura De Baixo Carbono E Inteligente Ao Clima No Brasil. Fortaleza, Ce. Programa De Pós-Graduação Em Economia Rural. Ufc. (Tese De Doutorado).

[7]. Instituto Brasileiro De Geografia E Estatística (Ibge). (2019). Censo Agropecuário 2017: Resultados Definitivos. Rio De Janeiro: Ibge. Disponível Em: Https://Sidra.Ibge.Gov.Br/Pesquisa/Censoagropecuario/Censoagropecuario-2017. Acesso Em: 05 Maio 2024.

[8]. Lemos, J.J.S. (2015). Efeitos Da Expansão Da Soja Na Resiliência Da Agricultura Familiar No Maranhão. Revista De Política Agrícola. N. 2, P. 17-25.

[9]. Marengo, J. A.; Cunha, A. P.; Soares, W. R.; Torres, R. R.; Alves, L. M.; Brito, S. S. B.; Cuartas, L. A.; Leal, K. D.; Ribeiro Neto, G.; Alvalá, R. C. S.; Magalhães, A. R. (2019). Increase Risk Of Drought In The Semiarid Lands Of Northeast Brazil Due To Regional Warming Above 4 ºc. In: Nobre, C. A.; Marengo, J. A.; Soares, W. R. (Ed.). Climate Change Risks In Brazil. New York: Springer International. P. 181-200.

[10]. Travassos, D. P., Silva, V. B., Araujo, M., Nobre, P., Rocha, H. (2013). Quadros Climáticos Extremos No Semiárido Brasileiro: Padrões E Variabilidade. Ciência Hoje, 42(253), 54-61.

[11]. Da Silva, T. G. F. Et Al. (2009). Impactos Das Mudanças Climáticas Na Produção Leiteira Do Estado De Pernambuco: Análise Para Os Cenários B2 E A2 Do Ipcc. Revista Brasileira De Meteorologia, Rio De Janeiro, V. 24, N. 4, P. 489-501.

[12]. Duque, J. G. (1973). Solo E Água No Polígono Das Secas. Salvador: Abc Gráfica.

[13]. Paiva, E, C.; Lemos, J. J. S.; Campos, R. T. (2022). Previsões Para A Produção De Leite Sob Instabilidade Pluviométrica No Ceará No Período De 1974 A 2019, Revista De Economia E Sociologia Rural, V.60 (Spe), E252091. Lessa, L.C.R. (2023). Resiliência E Sustentabilidade Da Agricultura De Sequeiro Sob Instabilidade Pluviométrica Na Paraíba. 96 F. Dissertação (Mestrado Em Economia Rural) - Centro De Ciências Agrárias, Universidade Federal Do Ceará, Fortaleza.

[14]. Lessa, L.C.R. (2023). Resiliência E Sustentabilidade Da Agricultura De Sequeiro Sob Instabilidade Pluviométrica Na Paraíba. 96 F. Dissertação (Mestrado Em Economia Rural) - Centro De Ciências Agrárias, Universidade Federal Do Ceará, Fortaleza

[15]. Lemos, J. J. S.; Bezerra, F. N. R. (2024). Arimax Model To Forecast Grain Production Under Rainfall Instabilities In Brazilian Semi-Arid Region. Global Journal Of Human-Social Science: Economics. V. 24, N. 1, Version 1.0.

[16]. Salviano, J.I.A. (2021). Relações Entre Instabilidades Das Chuvas E Indicadores De Produção De Lavouras De Sequeiro No Semiárido Cearense. Fortaleza, Ceará. Dissertação De Mestrado Em Economia Rural.

[17]. National Oceanic And Atmospheric Agency (Noaa, Eua). (2022). Global Historical-Monthly Climatology Network (Ghcn-M). Washington, Dc. Available At: Https://Www.Globalclimatemonitor.Org/#.

[18]. Instituto Brasileiro De Geografia E Estatística (Ibge). (2021). Produto Interno Bruto – Pib 2021, Rio De Janeiro. Available At: Https://Www.Ibge.Gov.Br/Geociencias/Cartas-E-Mapas/Mapas-Regionais/15974-Semiarido-Brasileiro.Html.

[19]. Lemos, J.J.S.; Bezerra, F.N.R. (2019). Interference Of Pluviometric Instability In The Forecast Of Grain Production In The Semiárido Of Ceará, Brazil. Braz. J. Of Develop., Curitiba, V. 5, N. 9, P. 15632-15652. Sep.

[20]. Valle, P. O., & Rebelo, E. (2002). Análise De Variância E Análise De Regressao Com Variáveis Dummy: Mais Semelhanças Do Que Diferenças. Revista De Estatistica, 1, 47-86.

[21]. Gujarati, D.N.; Porter, D. C. (2011). Econometria Básica. New York. Mcgraw Hill.

[22]. Gomes, F. P. Curso De Estatísticas Experimental. 13.Ed, São Paulo: Esalq/Usp, P.467, 1985.

[23]. Garcia, C. H. (1989). Tabelas Para Classificação Do Coeficiente De Variação. Piracicaba: Ipef. P.12. (Circular Técnica, 171).

[24]. Allison, P. D. Measures Of Inequality. (1978). American Sociological Review, New York, V. 43, N.6, P. 865-880, Dec.

[25]. O'reilly, C. A.; Caldwell, D. F; Barnett W. P. (1989). Work Group Demography, Social Integration, And Turn-Over. Administrative Science Quarterly, New York, V. 34, P. 21-37.

[26]. Wiersema, M. F., & Bantel, K. A. (1993). Top Management Team Turnover As An Adaptation Mechanism: The Role Of The Environment. Strategic Management Journal, 14(7), 485-504.

[27]. Sorensen, A. T. (2000). Equilibrium Price Dispersion In Retail Markets For Prescription Drugs. Journal Of Political Economy, Chicago, V. 108, N. 4, P. 833-850.

[28]. Pauw, K. (2003). Measures Of Poverty And Inequality: A Reference Paper. Provide Technical Paper, 4, South Africa. Disponível Em: Http://Ageconsearch.Umn.Edu/Bitstream/ 15623/1/Tp03 0004.Pdf. Acesso Em: 25 Mai. 2024.

[29]. Food And Agriculture Organization Of The United Nations (Fao). (2014). Refinements To The Fao Methodology To Restimation The Prevalence Of Under Nourishment Indicator. Working Paper, Rome, Italie, N. 14, 5. Available At: Http://Www.Fao.Org/3/A-I4046e.Pdf.

[30]. Food And Agriculture Organization Of The United Nations (Fao). (2006). Policy Impacts On Inequality Simple Inequality Measures. Available At: Http://Www.Fao.Org/Docs/Up/Easypol/448/Simple_Inequality_Mesures_080en.Pdf.

[31]. Enders, W. (2009). Applied Econometric Time Series. 3. Ed. New York.

[32]. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. (1994). Time Series Analysis: Forecasting And Control. 3thed. New York: Prentice Hall.

[33]. Wooldridge, J.M. (2015). Introductory Econometrics: A Modern Approach. South Western Educational Publishing, 881 P.

[34]. Greene, W. H. (2017). Econometric Analysis. 7. Ed. New York: Stern School Of Business. 1176p.

[35]. Box, G. E. P.; Jenkins, G. M. (1976). Time Series Analysis Forecasting And Control.San Francisco: Holden-Day. Edição Revisada.

[36]. Makridakis, S., Wheelwright, S. & Hyndman, R. J. (1998). Forecasting Methods And Applications. 3. Ed. New York: John Wiley & Sons.

[37]. Box, G. E., & Tiao, G. C. (1975). Intervention Analysis With Applications To Economic And Environmental Problems. Journal Of The American Statistical Association, 70(349), 70-79.

[38]. Box, G.E. Et Al. Time Series Analysis Forecasting And Control. New Jersey,Wiley.Fifth Edition. 2015.

[39]. Camelo Et Al. (2018). Proposta Para Previsão De Velocidade Do Vento Através De Modelagem Híbrida Elaborada A Partir Dos Modelos Arimax E Arn. Revista Brasileira De Meteorologia, 33(1), 115-129. Http://Dx.Doi.Org/10.1590/0102-7786331005.

[40]. Clement, E. P. (2014). Using Normalized Bayesian Information Criterion (Bic) To Improve Box - Jenkins Model Building. American Journal Of Mathematics And Statistics, V. 4, N. 5, Pp. 214-221.