Global Journal of HUMAN-SOCIAL SCIENCE: E
Economics
Volume 24 Issue 1 Version 1.0 Year 2024
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-460x & Print ISSN: 0975-587X

ARIMAX Model to Forecast Grain Production under Rainfall Instabilities in Brazilian Semi-Arid Region

By José de Jesus Sousa Lemos & Filomena Nádia Rodrigues Bezerra

Federal University of Ceara

Abstract- The state of Ceará has most of its area in Brazil's semi-arid region. Initially, the research segmented Ceará's annual rainfall into 6 periods: very rainy, rainy, normal-humid, normal-dry, drought and very drought. This segmentation was based on the annual rainfall in the state between 1901 and 2020. The research estimated the average rainfall and instability of both the annual rainfall in the state during the period and those estimated for the periods in which the rainfall was segmented. The research then developed forecast models for harvested areas, yields, production values and average annual grain prices between 1947 and 2020, the years in which this information is available. To make these forecasts, the research used the ARIMAX model, which is an extension of the Box-Jenkins model, with the addition of an exogenous variable. The exogenous variable included in the model was the annual rainfall observed between 1947 and 2020, assuming that this variable influences these forecasts. The results showed that the state's rainfall has a high level of instability and that the adjusted models proved to be parsimonious and robust from a statistical point of view.

Keywords: climate adversity. vulnerable ecosystems. systematic occurrence of droughts. aridity index. edaphoclimatic factors.

GJHSS-E Classification: JEL Code: O13, Q13, Q16

ARIMAX Model to Forecast Grain Production under Rainfall Instabilities in Brazilian Semi-Arid Region

José de Jesus Sousa Lemos [α] & Filomena Nádia Rodrigues Bezerra [σ]

Abstract- The state of Ceará has most of its area in Brazil's semi-arid region. Initially, the research segmented Ceará's annual rainfall into 6 periods: very rainy, rainy, normal-humid, normal-dry, drought and very drought. This segmentation was based on the annual rainfall in the state between 1901 and 2020. The research estimated the average rainfall and instability of both the annual rainfall in the state during the period and those estimated for the periods in which the rainfall was segmented. The research then developed forecast models for harvested areas, yields, production values and average annual grain prices between 1947 and 2020, the years in which this information is available. To make these forecasts, the research used the ARIMAX model, which is an extension of the Box-Jenkins model, with the addition of an exogenous variable. The exogenous variable included in the model was the annual rainfall observed between 1947 and 2020, assuming that this variable influences these forecasts. The results showed that the state's rainfall has a high level of instability and that the adjusted models proved to be parsimonious and robust from a statistical point of view.

Keywords: climate adversity. vulnerable ecosystems. systematic occurrence of droughts. aridity index. edaphoclimatic factors.

I. Introduction

The Brazilian semi-arid region is not homogeneous in terms of landscape, availability of natural resources or floral cover. The convergence that exists in the immense area that makes it up is climatic instability, reflected in the poor distribution of rainfall, both from a spatial and temporal point of view. Also common among the mosaics found in this Brazilian ecosystem are agricultural activities, especially those producing food such as rice, beans, manioc and corn, which are practiced mainly by family farmers on a rainfed basis, as well as extensive livestock farming, both of which are high-risk activities. These activities have very low yields, even when compared to those observed in the Northeast, which is not part of the Semi-Arid.

There is no doubt that the Brazilian semi-arid region is one of the most vulnerable ecosystems due to the instability of the rainfall regime, which leads to the systematic occurrence of droughts and, often, the incidence of floods (Assad & Pinto 2008; CEDEPLAR & FIOCRUZ, 2009). The adverse climate conditions in the semi-arid region turn its population into potential migrants. These migrations to other cities and/or other states tend to aggravate social problems that already exist in urban areas of cities of all sizes, especially in large cities, by the so-called "climate refugees".

Due to these aspects related to climate instability and the lack of more consistent policies for the Semi-Arid population to live with climate adversity, the social and economic indicators prevailing in the populations that survive in the municipalities located in the Semi-Arid are quite critical (Marengo et al., 2011; Lemos, 2015, 2020).

In Brazil, there are differences between what is Semi-arid from a technical point of view, which is measured only by the aridity index (AI), and that defined by the Federal Government, which uses other definition criteria in addition to the AI. It is worth noting that municipalities officially recognized by the Federal Government as belonging to the semi-arid region receive differentiated treatment in public policies.

According to the latest revision of the Semi-Arid Map, carried out by the Deliberative Council of the Northeast Development Superintendence (CONDEL/ SUDENE) at a meeting held in December 2021, the Brazilian Semi-Arid now has 1,427 municipalities. In the new delimitation, Ceará now has 171 of its 184 municipalities (97%) recognized as part of this ecosystem (SUDENE, 2017, 2021).

Agricultural activities in any ecosystem depend directly on edaphoclimatic factors and therefore become more sensitive to climatic fluctuations. In the semi-arid region of Ceará, this is very evident, given that the technological advances that allow cultivation with water difficulties have not yet reached the vast majority of Ceará's farmers. For this reason, favorable climatic conditions are still considered the defining factors for obtaining good levels of productivity, production and income in agricultural activities. This is evidence of the vulnerability of the agricultural sector to the climatic instability that permeates the reality of the rural population whose main productive activity is agriculture, both crop cultivation and domestic animal husbandry (Deschênes & Greenstone, 2007; Fisher et al., 2009; Alemaw & Simalenga, 2015; Mallari, 2016).

Therefore, the exercise of trying to predict what is likely to happen in the state's grain crops, even knowing that rainfall, being a natural phenomenon, is totally unpredictable, could be useful for both farmers and those promoting public policies aimed at the production of these items in Ceará.

Author α σ: Department of Agricultural Economics, Postgraduate Program in Rural Economics, Federal University of Ceara.
e-mail: lemos@ufc.br

The general objective of this study is to assess how the instability associated with rainfall influences the forecast of grain production in the semi-arid region of Ceará, from 1947 to 2020.

Specifically, the research aims to: a) classify the rainfall in the state of Ceará between 1901 and 2020; b) assess the behavior of the expected values and coefficients of variation of the variables associated with grain production in Ceará in each of the climate definitions constructed in the research; c) estimate how grain farmers in the semi-arid region of Ceará make projections about the area harvested, land productivity, production value per hectare and average price received, from 1947 to 2020; d) estimate how rainfall affects forecasts of the area harvested, productivity, production value per hectare and grain prices in the period evaluated.

II. Methodology

The study uses rainfall information provided by the National Oceanic and Atmospheric Agency (NOAA, 2022) and Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME, 2022), for the period 1901/2020. Information on the crops studied was collected from the Sidra database and the Statistical Yearbooks of the Brazilian Institute of Geography (IBGE). The period of data availability extends from 1947 to 2020. The grains that make up the series studied are: cotton, peanuts, rice, broad beans, beans, castor beans, corn, sunflower seeds and soybeans (IBGE, 1947-2021).

In this study, three types of variables are considered: endogenous, exogenous and constructed. Endogenous variables are those over which producers have some (but not all) autonomy in defining their magnitude and can make expectations about their future behavior. For the purposes of this study, these are: harvested areas (in hectares) and yields (in kilograms per hectare). Exogenous variables are those over which farmers have no decision-making power. In this group of variables there are those in which the farmer can build expectations by gathering information, such as prices. But there is also the exogenous variable, over which farmers not only have no influence, but also cannot build expectations because it is linked to natural phenomena. This is the case of annual rainfall. The third group of variables studied are those that are constructed from exogenous and/or endogenous variables. This group includes the quantity produced and the value of grain production per hectare. The average price and value of production per hectare, used in the research as a proxy for gross income, are updated to 2020 values, using the general price index of the Getúlio Vargas Foundation (IGP-DI) as an indexer. The 2020 values in reais were converted into US dollars, using the exchange rate of R$5.1558/US$. The methodological procedures adopted are presented below, followed by the research objectives.

a) *Methodologies used to achieve objectives "a" and "b"*

The first objective of the research is to create an outline of the distribution of rainfall in Ceará between 1901 and 2020 and try to capture within it descriptive statistics associated with the variables involved in grain production, and what was the behavior of rainfall in Ceará between the years 1947 and 2020.

To help interpret the rainfall observed in the semi-arid region of Ceará, a classification table was drawn up based on the data observed in the series from 1901 to 2020. To this end, the average and standard deviation of rainfall during this period were estimated and six (6) periods were outlined for the distribution of rainfall in Ceará. These periods are outlined in Table 1.

Table 1: Classification of rainfall in the semi-arid region of Ceará taking into account the mean and standard deviation (sd) of the rainfall distribution observed between 1901 and 2020

Períods	Range
Very rainy	Rainfall > (Average + 1 sd)
Rainy	(Average + 1 sd) > Rainfall > (Average +1/2 sd)
Normal-humid	(Average +1/2 sd) > Rainfall > Average <
Normal-dry	Average > Rainfall > (Average – 1/2 sd)
Drought	(Average – 1/2 sd)> Rainfall > (Average – 1sd)
Very drought	Rainfall < (Average – 1 sd)

Source: Values to be established based on NOAA data (2022).

To confirm the consistency of the classification outlined in the paper, a statistical test is carried out to assess whether the rainfall averages estimated for each of the periods are statistically different. If they are, it can be assumed that the classification adopted is of practical use. The test used was to estimate the following regression:

$$C_t = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D3 + \beta_4 D_4 + \beta_5 D_{5} + \eta_t \qquad (1)$$

In equation (1), C_t is the annual rainfall; D_1, D_2, D_3, D_4 and D_5 are binary variables, defined as follows: $D_1 = 1$ in the drought period; $D_1 = 0$ in other periods. $D_2 = 1$ in the normal-dry period; $D_2 = 0$ in the other periods. $D_3 = 1$ in the normal-humid period; $D_3 = 0$ in the other periods; $D_4 = 1$ in the rainy period; D4 = 0 in the other periods. $D_5 = 1$ in the very rainy period; $D_5 = 0$ in the other periods. When $D_1 = D_2 = D_3 = D_4 = D_5 = 0$, the very drought period is defined. The linear coefficient β_0 being statistically different from zero will be the average rainfall for the years of the very drought period. Since the other β_r coefficients (r = 1, 2, 3, 4, 5) are statistically different from zero, this means that the rainfall associated with the period β_r in question is different from the other periods. The random term η_t, by hypothesis, meets the assumptions of the classic linear model and so the parameters of equation (1) can be estimated using the ordinary least squares (OLS) method (Wooldridge, 2019). The hypothesis of this stage is that rainfall in the semi-arid region can be ranked as follows:

Very rainy period > Rainy period > Normal-humid period > Normal-dry period > Drought period > Very drought period.

b) *Measuring the instabilities/stabilities of the variables studied (objective "c")*

In order to assess the instabilities or stabilities associated with rainfall, as well as the variables used to study grain production in Ceará, the coefficients of variation (CV) estimated for each variable were used. By definition, the CV measures the percentage relationship between the standard deviation and the arithmetic mean of a random variable. The CV is useful for measuring the heterogeneity or homogeneity observed in the distribution of the values of a random variable around its mean. CV can be used as a measure of inequality and/or to measure the accuracy of experimental results (Gomes, 1985; Garcia, 1989).

The advantage of using CV in this evaluation model over other measures of variability is that it is independent of the units in which the variables are measured. Thus, it allows the comparison of homogeneities/heterogeneities or stabilities/instabilities between variables measured in different units of measurement (Allison, 1978; Garcia, 1989; Lemos & Bezerra, 2019; Garcia, 1989; O'Reilly et al., 1989; Punt, 2003; Wiersema & Bantel, 1993).

The lower the CV, the more homogeneous the distribution of observations around the mean. In order to use CV as a measure of homogeneity/ heterogeneity or stability/instability in a distribution, it is necessary to define its critical values. Gomes (1985) established limits for classifying CVs in agricultural experimentation (Table 2).

Table 2: Classification of the coefficient of variation (CV) according to its amplitude

Classification of CV range	Range CV (%)
Low	CV < 10
Medium	10 ≤ CV < 20
High	20 ≤ CV < 30
Very high	CV ≥ 30

Source: Gomes (1985).

c) *Methodological strategies for achieving objectives "d" and "e"*

To achieve the fourth specific objective, the research uses the definitions and procedures discussed below.

i. *Definition of the model used*

The study is based on the equation for defining the quantity of grain produced (Qt), which is the result of multiplying the harvested area (At) by the productivity of the land (Rt) defined by equation (2):

$$Q_t = A_t.R_t \quad (2)$$

Taking the logarithm of equation (2), and calculating the derivative with respect to time (T), we obtain:

$$d[\ln(Q_t)]/dT = d[\ln(A_t)]/dT + d[\ln(R_t)]/dT \quad (3)$$

Assuming that the derivative of the logarithm of a variable in relation to time measures the growth rate of that variable over time, equation (3) shows that the growth rate of grain production will depend on the addition of the growth rates of harvested areas and yields.

The value of production per hectare of grain (Vt), which can be understood as a proxy for the gross income per hectare associated with grain production, is in turn defined as shown in equation (4):

$$V_t = Q_t.P_t/A_t \quad (4)$$

Substituting into equation (5) the value of Qt shown in equation (3), taking the natural logarithm, and making the total differential in relation to the time of the result, we obtain the result shown in equation (5).

$$d[\ln(V_t)]/dT = d[\ln(R_t)]/dT + d[\ln(P_t)]/dT \quad (5)$$

Equation (5) shows that the growth rate over time of the production value per hectare of grain is the sum of the growth rates of yields and grain prices.

ii. *Predicted value of a random variable*

Given a random variable (Y_t), its predicted value (Y_P) will differ from its observed value if there are information shocks caused by unforeseen situations at the time expectations were formed. This can be represented by the error term ξ_t shown in equation (6):

$$Y_t = E(Y_t) + \xi_t = Y_P + \xi_t \quad (6)$$

If the Y_t series, which in this study can take the values of harvested areas, yields, production values per hectare or average grain prices, is stationary, ξ_t is endogenously white noise, Although it is endogenously white noise, in this study we assume the hypothesis that ξ_t is affected by the exogenous variable, rainfall (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 grain production decision-makers. This can be seen in equation (7):

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

Substituting equation (7) into equation (6) gives equation (8), defined as follows:

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

In this research, the empirical framework designed to estimate each of the predicted values (Y_P) for harvested area, land productivity, production value per hectare and average grain prices is anchored in the ARIMAX methodology, which is an expansion of the Autoregressive Integrated Moving Average (ARIMA) models, with the insertion of exogenous variables (Box & Tiao, 1975; Camelo et al., 2018).

iii. *Box-Jenkins ARIMA models*

The formulations proposed by Box and Jenkins (1978), ARIMA (Auto Regressive Integrated Moving Average), are mathematical frameworks that aim to capture the behavior of a random variable that has values distributed in realizations in the form of time series. The time series Y_t can be represented as follows:

$$Y_t = \mu + \Sigma\psi_{k.}u_{(t-k)} = \mu + \psi(B).u_t \qquad (9)$$

Where ψ, defined as the linear filter, is represented by:

$$\psi(B) = \theta(B) / \phi(B) \qquad (10)$$

The terms in equation (10) are defined by the following polynomials:

$$\theta(B) = 1 – \theta_1B - \theta_2B^2 -.... - \theta_qB^q ;\ e\ \ \phi(B) = 1 – \phi_1B – \phi_2B_2 - ... - \phi_pB^p$$

Defining $\tilde{Y}_t = Y_t – \mu_t$, it will be possible to obtain the following transformation:

$$\phi(B)\ \tilde{Y}_t = \theta(B).u_t \qquad (11)$$

In equation (11), u_t is a generally Gaussian "white noise". To do this, it must meet the following conditions: i) $E(u_t) = 0$; ii) $E(u_t^2) = \sigma_u^2 < \infty$; e iii) $E(u_t,u_{t+k}) = 0$, para $k = \pm1, \pm2, ...$ (Cochrane, 1997).

According to Box and Jenkins (1978) equation (12) is called ARMA(p,q) and can be rewritten as follows:

$$\tilde{Y}_t = \theta(B)\ \phi^{-1}.(B).u_t \qquad (12)$$

Types of Box and Jenkins models when the series is stationary:

i) *Autoregressive (AR) models:* These are 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 (t - 1, t - 2, ...);

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

iii) Models which are both autoregressive and moving average (ARMA) are those which are made up of two parts: one part (AR) another part MA, and have the notation ARMA(p.q).

iv) If the series is not stationary, converting it to assume this condition will require differentiation (d) of the dependent variable (Y_t). In general, with up to three differentiations, a non-stationary series will become stationary. In this case, it is said to be an ARIMA (p.d.q) model, where "d" is the number of differences needed to make the original series stationary.

III. Arimax Model

The ARIMAX model is a generalization of the ARIMA method with the inclusion of exogenous variables. In this study, the exogenous variable used is the annual rainfall observed for the state of Ceará between 1947 and 2020.

The ARIMAX model is considered multi-variate 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 (Bennet, 2014; Box & Jenkins, 1978; Box et al., 2015; Camelo et al., 2018).

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 the state of Ceará between 1947 and 2020, assuming the hypothesis that temporal instabilities in rainfall affect the forecasting capacity of the variables that define annual grain production in the state between 1947 and 2020. A simplified way of mathematically representing this model in a generalized way is described in equation (13), which is the application of equation (8) to this research.

$$Y_t = [\rho + \Sigma\beta_jY_{(t-i)} + \Sigma\theta_jЄ_{(t-j)}] + [\Sigma\omega_jX_j + \varepsilon_t] \qquad (13)$$

According to the model proposed in equation (8), the predicted value (Y_P) is defined in expression (13) as follows:

$$[\rho + \Sigma\beta_jY_{(t-i)} + \Sigma\theta_jЄ_{(t-j)}] \qquad (13a)$$

And the term $f(X_t)$ is represented by:

$$[\Sigma\omega_j X_j + \varepsilon_t] \quad (13b)$$

In this research, the variables Y_t to be predicted as Y_P are: harvested area; productivity; production value per hectare and average grain price in the semi-arid region of Ceará between the years 1947 and 2020. In this research, the exogenous variable (Xj) is annual rainfall.

a) Stationarity in the Box and Jenkins model

A stochastic process $Y_t = \psi(B)u_t$ is stationary if

$$\psi(B) = \Sigma_{k=0}^{\infty} \psi_k(B)^k \text{ converge to } |B| < 1$$

It is assumed that most statistical analysis procedures for time series are stationary. If they are not, it is necessary to transform the non-stationary ones into stationary ones. The autocorrelation function between the residuals is estimated. If the autocorrelation function stabilizes at the first lag, then the series can be assumed to be stationary. If this is not the case, the second, third or more lags are used to find stationarity. In general, series do not need more than three lags to become stationary (Makridakis et al., 1998; Morettin & Toloi, 2006).

b) Steps to follow to achieve the best fit in ARIMA models

In general, three (3) phases are followed to define the best fit for the Box-Jenkins model so that forecasts can be made. The first phase is model identification, which consists of two stages: data preparation and model selection. At this stage we check whether the series is stationary. Stationarity is verified using the ADF (Augmented Dickey-Fuller) unit root test. The autocorrelation function (FAC) and the partial autocorrelation function (FACP) of the series are then calculated, in addition to the graphical analysis which allowed the ARIMA model (p, d, q) to be selected.

After the identification stage, the model's parameters were estimated (second stage). The "d" parameter refers to the number of times the difference between the elements of the series was taken until it became stationary. In this study, only the first difference had to be taken to make the series stationary ($d = 1$). Calculating the autoregressive parameters p and the moving average q involves analyzing the FAC and FACP functions, respectively.

Still in phase 2 of estimation and testing, once the appropriate values have been defined, the analysis of the residuals (ξ) is carried out, which must be white noise. For this purpose, the Ljung-Box (LB) statistic is used, which must be non-significant at a significance level of at least 10%. Next, the mean absolute percentage error (MAPE) is analysed, which considers the relative error of each forecast, in order to compare the values predicted by the model with the observed values of the series, characterizing the forecasting capacity of the model adopted.

To assess the suitability of the adjusted model, in addition to the statistical significance of the parameters, it is considered that the model will fit better if the number of estimated parameters is as small as possible. This is the parsimony criterion associated with the number of regressors to be estimated. The following criteria are also used: the magnitude of the coefficient of determination (R^2), which assesses the percentage of variation in the variable analyzed that is explained by the structured model; Pearson's correlation coefficient, to assess the level of adherence of the values predicted by the adjusted model to the data observed in the series under analysis.

In the verification phase, the analysis of the model consists of checking for the lowest values for the AIC and BIC criteria (Akaike Information Criterion and Bayesian Information Criterion, respectively), since these criteria aim to indicate the most parsimonious model, i.e. with the fewest parameters, since they are built based on the estimated variance (σ) and sample size (n). The model with the lowest AIC and BIC values will be the one that best fits the data (Brockwell; Davis, 1991). Once the most suitable model has been fitted, the predictions obtained are assessed for their accuracy using the Mean Absolute Percentage Error (MAPE) value. The best model should have the MAPE value (Camelo et al., 2018; Box et al., 2015; Wooldridge, 2019).

IV. Results and Discussion

The results of the research are presented in the order in which the specific objectives are distributed in the initial section of the paper.

a) Results obtained for the first and second objectives

The results found for annual rainfall in the state of Ceará between 1901 and 2020 showed that the lowest rainfall was 250.90 mm in 1919. The highest was 1773.40 mm in 1985. These values gravitated around an average of 798.82 mm, with very high instability, according to the classification of (Gomes, 1985) measured by CV=33.56%). In the period from 1947 to 2020, the state's average rainfall was 768.76 mm, showing very high instability, as measured by CV= 33.64%. The highest and lowest rainfall occurred in 1985 (1773.40) and 1958 (286.90 mm), respectively.

Table 3: Years of occurrence, averages and coefficients of variation (CV) of rainfall classified into periods between the years 1901/2020 and the period 1947/2020 in Ceará.

	Period from 1901 to 2020				Period from 1947 to 2020			
	Years of occurrence		Average	CV	Years of occurrence		Average	CV
Periods	Totals	%	(mm)	(%)	Totals	%	(mm)	(%)
Very rainy	18	15.00	1250.03^{A}	15.32	7	9.46	1302.83	18.21
Rainy	17	14.17	989.28^{B}	3.73	12	16.22	988.47	3.29
Normal humid	20	16.67	860.86^{C}	4,32	10	13.51	852.45	4.98
Normal dry	28	23.33	716.61^{D}	6.26	18	24.32	726.39	6.51
Drought	23	19.17	589.37^{E}	6.80	17	22.97	591.94	7.05
Very drought	14	11.67	407.37^{F}	20.19	10	13.51	424.44	18.29
Ceará	120	100.00	798.82	33.56	74	100.00	768.76	33.64

Source: Estimated values based on NOAA data (2022). Observation: the indices superimposed on the rainfall averages observed in each period have the following meaning: A>B>C>D>E>F.

Applying the test to check whether the averages of the periods defined in Table 1 are statistically different or equal, the results shown in Table 4 were found.

Table 4: Tests to assess whether the rainfall periods defined in the research are statiscally different.

Varibles	Estimated Coefficients	Student statistics test (t)	Signifcance
(Constant)	407.372	17.661	0.000
D_1	181.994	6.221	0.000
D_2	309.233	10.946	0.000
D_3	453.485	15.079	0.000
D_4	581.903	18.682	0.000
D_5	842.662	27.400	0.000

Source: NOAA (2022). Obs: Adjusted R^2 = 0.896.

The results shown in Table 4 suggest that all 6 periods tested are statistically different, with practically zero probability of error. The adjusted coefficient of determination (R^2 = 0.896) corroborates the information that the adjustment achieved was robust from a statistical point of view. Based on these results, the hierarchy assumed in the research definition is confirmed: rainfall averages: Very rainy> Rany> Normal humid > Normal dry> Drought> Very drought.

From the evidence shown in Table 3, it can be seen that in the 74 years evaluated, 17 years (23%) were considered drought and 10 years (13.5%) were considered very dry. Thus, it can be concluded that in 36.5% of the years studied in grain production, rainfall fell into the drought and very drought periods. The average rainfall for these two periods was 529.90 mm.

The very high instability estimated for Ceará's rainfall between 1947 and 2020 influenced those associated with the variables used to study grain production, all of which were classified as having very high instability. In fact, the estimated coefficient of variation for harvested areas was 41.02%; the estimated CV for yields was 58.68%. The CV calculated for production values per hectare was 42.18% and that estimated for observed prices was 61.24%.

The averages, as well as the coefficients of variation (CV), estimated to gauge the levels of instability associated with the variables harvested areas, yields, production value per hectare and average grain prices are shown in Table 5. The results shown in this table suggest that the instability observed in rainfall in the state of Ceará between 1947 and 2020 is also evident in all the variables studied.

In fact, only the harvested areas showed instability classified as medium (CV = 19.33%) in the very rany period. The value of production per hectare showed high instability in the very rany period (CV = 27.73%). In all the other periods, the variables studied showed instabilities classified as very high according to Gomes (1985). This confirms the assumptions underlying this study that rainfall instability is transmitted to the variables studied for grain production in the semi-arid region of the state of Ceará (Table 5).

Table 5: Averages and coefficients of variation (CV) of the defining variables of grain production in the semi-arid region of Ceará according to the rainfall season, between 1947 and 2020.

	Variables							
	Area (1000 ha)		Yield (kg.ha^{-1})		Value (USD.ha^{-1})		Price (USD.kg^{-1})	
Períods	Average	CV	Average	CV	Average	CV	Average	CV
Very rany	1729.92	19.33	376.86	27.73	329.62	37.39	0.95	42.95
Rany	1587.82	33.14	619.05	53.78	298.40	32.29	0.60	58.98
Normal humid	1363.42	40.72	731.10	51.96	321.35	31.22	0.55	54.86
Normal dry	1239.26	45.66	588.43	53.33	368.40	39.21	0.81	61.78
Drought	1303.81	38.37	418.24	52.66	246.33	53.44	0.75	71.91
Very drought	884.95	47.47	308.85	53.15	270.86	47.43	0.90	54.24

Source: Estimated values based on original data from NOAA (2022) and IBGE (1947-2021).

b) *Results achieved to meet the third and fourth objectives*

Having presented the descriptive statistics of the decision variables in grain production in the semi-arid region of Ceará, we will now present and discuss the results of the adjustments to foster expectations associated with the study variables. The stationarity tests showed that the series of harvested areas, yields, production value per hectare and average grain prices were not stationary, but that it was possible to reverse this situation by making just one difference.

Table 3 summarizes these results. Generally speaking, it can be seen that the best adjustments were made using ARMAX (0.1.1.1) models for the variables harvested areas, yields and production values per hectare. For the stationary price series, the best fit was an ARIMAX (1.1.0.1).

Table 6 also shows that the Ljung-Box statistics, which test the hypothesis that the residuals are all not significantly different from zero, at least at the 35% probability level, suggest that the resulting residuals are white noise. The adjustments obtained showed coefficients of determination (R^2) ranging from 0.613 in the model estimated to forecast yields to 0.721 in the model estimated to forecast prices.

The MAPE ranged from 18.848 for predicting harvested areas to a maximum value of 31.396 in the model estimated for predicting yields. The BIC statistics ranged from 0.593 for the model created to estimate prices to 25.637 for the model created to estimate harvested areas. The estimated correlation coefficients between the observed values and the values estimated using the models ranged from 0.890 for the model created to forecast yields to 0.974 for the model created to forecast prices. In this way, all the results obtained for estimating the four variables can be considered parsimonious and robust from a statistical point of view.

The estimated coefficients associated with the regressions, including the exogenous variable (rainfall), were all statistically different from zero at least at an error level of less than 2%. These results consolidate the assumption that the exogenous variable used in the research does affect the projections of the variables used to forecast harvested areas, yields, production values per hectare and average grain prices in Ceará between 1947 and 2020.

Table 6: Adjustments obtained for the ARIMAX models for the forecasts of harvested areas (ha), yield per hectare (ton/ha), production value per hectare (US$/ha) and grain prices (US$/kg) in the semi-arid region of Ceará from 1914 to 2020.

Var.	Model	AR Lag 1	MA Lag 1	X Lag 1	MAPE	BIC	Ljung-Box Sign.	R^2	R Pearson
Area	ARIMAX (0.1.1.1)	0.000	0.510*	476.886*	18.848	25.267	0.275^{NS}	0.717	0.910
Yield	ARIMAX (0.1.1.1)	0.000	0.506*	0.342*	31.396	10.637	0,470^{NS}	0.621	0.890
Value/ha	ARIMAX (0.1.1.1)	0.000	0.723*	0.548*	21.817	11.998	0.679^{NS}	0.681	0.907
Price	ARIMAX (1.1.0.1)	0.290**	0.000	-0.001**	20.855	0.593	0.206^{NS}	0.721	0.974

*Sources: IBGE (1947-2021); NOAA (2022). Note: *significant to less than 1% error; **significant to less than 2% error; NS: not significant to at least 20.6% error.*

c) *Results obtained to achieve objective "e"*

The results shown in Table 7 suggest that the intensity and instability of rainfall interferes with the predictive capacity of the variables used to define grain production in semi-arid Ceará between 1947 and 2020. These results also show that, in general, the greatest difficulties in predicting the variables used to define grain production in the semi-arid region of Ceará between 1947 and 2020 occurred mainly during the very drought periods, especially in the case of predicting productivity (CV=40.45%), harvested áreas (CV = 31.68%) and production values per hectare (31.38%).

Table 6: Estimated mean of absolute forecasted erros (%) for each of the rainfall periods defined in the survey.

	Períods (%)					
	Very rany	Rany	Normal humid	Normal dry	Drought	Very drought
Areas	11.59	13.54	14.52	11.17	12.99	31.68
Produtivity	26.24	17.31	8.72	17.35	33.50	40.45
Value	7.51	14.71	5.80	11.36	24.50	31.38
Price	17.20	16.25	11.46	7.66	9.87	10.50

Source: IBGE (1947-2021); NOAA (2022).

To estimated mean of absolute forecasted erros of prices was was higher during the very rany period (17.20%). The results also showed that in very rainy periods, such as those defined in this study, the produtivity predictions made by the estimated models had high average absolute errors in the drought period (33.50%) and in the very rany period (26.24%). This evidence shows that high rainfall intensities can also create problems for grain production in the semi-arid region, causing high instability in its produtivity.

V. Conclusions

The methodological procedures adopted in the research made it possible to answer two questions: 1 - does the rainfall observed in Ceará, one of the Brazilian states with the highest relative number of municipalities located in the semi-arid region, influence the forecasts of the variables that define grain production in the state? 2 - Do the forecast errors of the variables that define grain production differ between the periods in which the classification of rainfall in the state of Ceará is defined?

In fact, the results found in the research showed that between the years 1901 and 2020 the average rainfall in the state of Ceará was 798.82 mm, with CV= 33.64%, classified as very high.

The evidence found in the study also shows that the objective of segmenting the rainfall observed in the semi-arid region of Ceará between 1901 and 2020 into six periods was achieved: very rainy, rainy, normal humid, normal dry, drought and very drought. It was shown that rainfall instability in Ceará between 1947 and 2017 spilled over into the variables associated with grain production in the state's semi-arid region: harvested area, productivity, production value per hectare and prices. All of these variables show statistically different values within the rainfall ranges created in the research, which made it possible to rank these periods according to the magnitudes of the average rainfall, as follows: very rainy > rainy > greater than normal humid > normal dry > drought > very drought.

Average grain prices showed the greatest heterogeneity over the period studied. As expected, when rainfall was lowest (drought), the average price was highest. However, the average values of these prices were not statistically different in the three periods in which rainfall was characterized in this study.

The adjustments made to forecast the variables studied were all satisfactory from an econometric point of view and also from the point of view of the assumptions that guided the research. These results showed that all the variables associated with grain production in Ceará showed high instabilities induced by the different rainfall regimes that the research was able to map.

Thus, the general conclusion reached in the work is that the rainfall observed in the state of Ceará over a 120-year historical series (1901/2020) can be segmented into periods according to the intensity and instability in which they occurred, and that for this reason it also interferes in the forecasts of the variables that define grain production in different ways in terms of forecast errors.

References Références Referencias

1. Alemaw, B. & Simalenga, T. (2015). Climate Change Impacts and Adaptation in Rainfed Farming Systems: A Modeling Framework for Scaling-Out Climate Smart Agriculture in Sub-Saharan Africa. *American Journal of Climate Change*, 4, 313-329.
2. Allison, P. D. (1978). Measures of inequality. *American sociological review.* p. 865-880.
3. Assad, E. & Pinto, H. S. (2008). *Aquecimento Global e a Nova Geografia da produção agrícola no Brasil.* EMBRAPA-CEPAGRI, São Paulo, 82 p.
4. Bennett, C., Stewart, R.A., & Lu, J. (2014). Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies*, 7(5), 2938-2960.
5. Box, G.E., & Tiao, G.C. (1975). Intervention analysis with application to economic and enviromental problems. *J. Am. Stat. Assoc.*, 70, 70-79.
6. Camelo et al. (2018). Proposta para Previsão de Velocidade do Vento Através de Modelagem Híbrida Elaborada a Partir dos Modelos ARIMAX e

RNA. *Revista Brasileira de Meteorologia*, 33(1), 115-129. http://dx.doi.org/10.1590/0102-7786331005.

7. Cochran, W.G. (1977). *Sampling techniques*. 3. Ed. New York: John Wiley & Sons, 428 p.
8. Box, G. E. P., & Jenkins, G. M. (1978). *Time series analysis forecasting and control.* San Francisco: Holden-Day.
9. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
10. CEDEPLAR & FIOCRUZ. (2008). *Mudanças Climáticas, Migrações e Saúde: Cenários para o Nordeste Brasileiro, 2000-2050.* Belo Horizonte: CEDEPLAR/FICRUZ, Relatório de Pesquisa (Research Report).
11. Deschênes, O., & Greenstone, M. (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *The American Economic Review*, 97(1), 354-385.
12. Fisher, A., Hanemann, W. M., Roberts, M. J., & Schlenker, W. (2009). *Climate change and agriculture reconsidered.* University of California, Berkeley Department of Agricultural & Resource Economics. CUDARE Working Papers. https://escholarship.org/uc/item/33v2d7vc.pdf
13. FUNCEME, Fundação Cearense de Meteorologia e Recursos Hídricos. (2022). *Evolução da pluviometria no Ceará entre 1947 e 2021*. Fortaleza. http://www.funceme.br/index.php/areas/23-monitoramento/meteorol%C3%B3gico/406-chuvas-di%C3%A1rias#toddospelaagua2.
14. Garcia, C.H. (1989). *Tabelas para classificação do coeficiente de variação.* Piracicaba: IPEF. 12p. (Circular técnica, 171).
15. Gomes, F.P. (1985). *Curso de estatística experimental*. 13.ed. São Paulo: ESALQ/USP, 467p.
16. IBGE, Instituto Brasileiro de Geografia e Estatística. *Produção Agrícola Municipal*. (1947-2021). Banco SIDRA. Rio de Janeiro. https://sidra.ibge.gov.br/acervo#/S/PA/A/Q.
17. Lemos, J. J.S. (2020). *Vulnerabilidades induzidas no semiárido*. Fortaleza: Imprensa Universitária. 170 p.
18. Lemos, J. J. S. (2015). *Pobreza e Vulnerabilidades Induzidas no Nordeste e no Semiárido Brasileiro*. Tese submetida como parte dos requisitos para o concurso destinado à promoção da classe Professor Titular da Universidade Federal do Ceará-UFC, Fortaleza, Ceará.
19. Lemos, J. J. S., & Bezerra, F. N. R. (2019). Interferência da instabilidade pluviométrica na previsão da produção de grãos no semiárido do Ceará, Brasil. *Brazilian Journal of Development*, 5(9), 15632-15652.
20. Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1998). *Forecasting methods and applications.* 3. Ed. New York: John Wiley & Sons.
21. Mallari, A. E. C. (2016). Climate Change Vulnerability Assessment in the Agriculture Sector: Typhoon Santi Experience. *Procedia - Social and Behavioral Sciences*, 216, 440 – 451.
22. Marengo, J. A., Alves, L.M., Beserra, E. A., & Lacerda, F. F. (2011). Variabilidade e mudanças climáticas no semiárido brasileiro. In: Medeiros, S. S., Gheyi, H. R., Galvão, C. O., & Paz, V. P. S. (Org.). *Recursos Hídricos e, Regiões Áridas e Semiáridas*. Campina Grande, PB: INSA, 383-416.
23. Morettin, P. A., & Toloi, C. M. C. (2006). *Análise de Séries Temporais*. 2. São Paulo, SP. ABE - Projeto Fisher, Edgard Blücher.
24. NOAA, National Oceanic and Atmospheric Administration. (2022). National centers for environmental information (NCEI, EUA). *Climate monitoring*. www.ncei.noaa.gov/access/monitoring/products/
25. O'Reilly III, C. A., Caldwell, D. F., & Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative science quarterly*, 21-37.
26. Punt, C. (2003). *Measures of Poverty and Inequality: A Reference Paper.* Provide Technical Paper. http://ageconsearch.umn.edu/bitstream/15623/1/tp030004.pdf
27. SUDENE, Superintendência do Desenvolvimento do Nordeste. (2017). *Resolução nº 115, de 23 de novembro de 2017***.** Diário Oficial da União. Ministério da Integração Nacional. Conselho Deliberativo (CONDEL/SUDENE). http://sudene.gov.br/images/arquivos/semiarido/arquivos/resolucao115-23112017-delimitacaodosemiarido-DOU.pdf
28. SUDENE, Superintendência do Desenvolvimento do Nordeste. (2021). *Delimitação do Semiárido – 2021*. Relatório final. Ministério do Desenvolvimento Regional, Recife. https://www.gov.br/sudene/pt-br/centrais-de-conteudo/02semiaridorelatorionv.pdf
29. 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.
30. Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach*. 7. Ed. Cengage Learning, 816.