Temporal Rainfall Variations Induce Forecast Errors in Rainfed Agriculture in the Brazilian State of Ceará, Brazil

José de Jesus Sousa Lemos[1], Filomena Nádia Rodrigues Bezerra[1], Elizama Cavalcante Paiva[1], Antonia Leudiane Mariano Ipolito[1], Erika Costa Sousa[1] & João da Costa Filho[1]

[1] Federal University of Ceara (UFC), Brazil

Correspondence: Erika Costa Sousa, Federal University of Ceara (UFC), Brazil.

Received: January 31, 2024 Accepted: March 6, 2024 Online Published: March 15, 2024

doi:10.5430/ijba.v15n1p36 URL: https://doi.org/10.5430/ijba.v15n1p36

Abstract

The research aims to: a) assess the instabilities associated with rainfall and the variables that define the production of rice, beans, cassava and corn in the state of Ceará between 1945 and 2020; b) estimate models that can be used to make projections of harvested areas, yields and prices for these crops between 1945 and 2020; c) assess the impact of rainfall on the estimated forecasting models; d) assess how rainfall affects the likelihood of farmers making forecasts of the variables that define agricultural production. Rainfall data was obtained from the National Centers for Environmental Information (NOAA). Crop yield data came from the Brazilian Institute of Geography and Statistics (IBGE). Instabilities were measured by the coefficients of variation. ARIMA models (autoregressive, integrated and moving average model) were used to make the forecasts. The hypothesis that the residuals generated by the models are influenced by annual rainfall was tested. The results showed high instabilities in annual rainfall, which spread to the variables that define crop yields. Parsimonious and robust adjustments were obtained from a statistical point of view and it was shown that the errors generated, including their magnitudes, in the models used to forecast all the variables that define bean and corn yields, harvested areas and rice yields, as well as cassava yields, are influenced by annual rainfall in Ceará between 1945 and 2020.

Keywords: occurrence of droughts, rainfall instability, Brazilian semi-arid region, synergy of events, potential evapotranspiration

Jel: Q15, Q18, C22, Q54, Q58

1. Introduction

The aggressive conditions of Brazil's semi-arid climate are mainly characterized by an average sunshine of 2,800 hours per year. Although it is relatively rainier than other semi-arid regions on the planet, with an average annual rainfall of around 800 mm, it has an average annual evaporation of approximately 2000 mm. Rainfall is concentrated in the first three to four months of the year. The temperatures are high with annual averages of 23 ℃ to 27 ℃. This rainfall is distributed over the years with a coefficient of variation of around 2%, which indicates that the temperatures in this climatic region in Brazil are always high, and relative humidity low (50%). This synergy of events leads to the occurrence of water balances that are almost always negative and become unfavorable to the development of agricultural practices, especially when carried out without the use of appropriate technologies adapted to these difficulties (Bezerra, 2022; Brasil, 2021; Lemos, 2020;; Salviano, 2021).

Another characteristic of this climate is its instability in rainfall both spatially and temporally (Carvalho, 2014; Lemos, 2020; Melo *et. al.*, 2019; Mohammed, 2019; Salviano, 2021). Rainfall instability, which is the norm in Brazil's semi-arid climate regime, leads to errors in forecasts and in the physical and economic results observed in the cultivation of the main crops, especially rainfed crops. The state of Ceará is one of the most affected by rainfall instability. In fact, research carried out by Lemos & Bezerra (2023) showed that between 1901 and 2020 (200 years) the average rainfall in Ceará was 799.5 mm with a coefficient of variation of 33.6%. The study classified rainfall into three periods: drought, normal and rainy. To count the drought period, the survey measured rainfall below the average minus half a standard deviation. Based on this criterion, 38 years of drought were counted (31.7%), therefore, with rainfall below 648.5 mm, which is the upper limit of this period.

In the most common cases of water scarcity, caused by the systematic occurrence of droughts, the use of technologies that neutralize, or at least mitigate, the impacts of rainfall irregularities are quite rare. These facts make agricultural practices under this climate regime difficult to conduct and subject to economic, social and environmental risks (Beyer *et. al.*, 2016; Costa Filho, 2019; Fischer *et. al.*, 2002; Pereira, 2018; Rosenzweig, 2005; Thornton *et. al.*, 2008).

The semi-arid region is defined internationally by the relationship between rainfall and potential evapotranspiration, measured by the aridity index (AI). The AI in semi-arid regions varies between 0.20 and 0.50

(Köppen, 1936; Köppen, 1918; Thornthwaite, 1948). In general, agricultural activities depend exclusively on rainfall in Brazil's semi-arid region. Soil and water management in these areas are the main constraints to maintaining sustainable crop yields and productivity (Rockstrom *et. al.*, 2010; Mohinder Singh *et. al.*, 2017; Wani *et. al.*, 2009).

Between the years 1949 and 2020, it was possible to classify Ceará's rainfall into dry periods in which rainfall ranged from 286.9mm to 621.7mm. Normal periods with rainfall ranging from 622mm to 959mm. Rainy period when rainfall was above 959mm (Salviano, 2021).

To make matters worse, farmers in Ceará, in general, do not use irrigation and genetically modified seeds adapted to the hostile conditions of the semi-arid regime, which makes agricultural practices in this state difficult and less profitable. The crops grown in the northeastern semi-arid region are mainly rice, beans, cassava and corn (IBGE, 2020). These crops help to provide food security, provide occupational opportunities and generate monetary income for family farmers, who constitute the vast majority of producers of these crops in the semi-arid zones of the state of Ceará (IBGE, 2017; Costa Filho, 2019; Fischer *et. al.*, 2002; Pereira, 2018; Rosenzweig *et. al.*, 2005; Salviano, 2021; Thornton *et. al.*, 2008; Wani *et. al.*, 2009).

Notwithstanding the technical definition of climatic regimes, the selection of Brazilian municipalities that are officially part of the semi-arid region is defined by the following criteria: 1 - Thornthwaite Aridity Index less than or equal to 0.50; 2 - Average annual rainfall equal to or less than 800mm and; and 3 - Daily Water Deficit Percentage equal to or greater than 60%, considering all days of the year. The municipality will be officially recognized by the Federal Government if it has at least one of these characteristics. These criteria are not set to prevail indefinitely. They are defined by the Sudene Deliberative Council (CONDEL/SUDENE) which, from time to time, may or may not review these criteria for defining the political composition of the Brazilian semi-arid region, and may include new municipalities, or remove others that were part of previous definitions (Brasil, 2021).

In view of these criteria, the new official definition of the Brazilian Semi-Arid affects all nine states in the Northeast and municipalities in the states of Minas Gerais and Espírito Santo located in the Southeast region of Brazil. According to the latest CONDEL/SUDENE definition, which took place in December 2021, the Brazilian semi-arid region now has 1,427 municipalities. Of the 184 municipalities in the state of Ceará, 175 are currently recognized by the Federal Government as being in this climate regime. As a result, they benefit from the resulting effects such as differentiated access to public policies, like the Northeast Constitutional Fund (FNE), for example. Ceará is the Brazilian state with the largest relative area and population within this climate regime, in which it is difficult to carry out agricultural activities, both in terms of plant and animal production (Brasil, 2021).

Rainfall instability, which is the norm in Brazil's semi-arid climate regime, leads to errors in forecasts in the physical and economic results observed in the cultivation of the main crops, especially rainfed crops. Based on this, this research sought to understand how this rainfall instability, which is an exogenous variable to farmers' decisions, affects the ability to forecast and, consequently, plan for the future cultivation of these crops in the state of Ceará. To this end, this study has the following objectives: a) to assess the instabilities associated with rainfall and the variables that define the production of rice, beans, cassava and maize in the state of Ceará between 1945 and 2020; b) to estimate models that can be used to draw up projections of harvested areas, yields and average prices for the selected crops between 1945 and 2020; c) evaluate the impact of rainfall on the forecasting models estimated for the production of the agricultural crops being analysed; d) evaluate how the distribution of rainfall affects the likelihood of farmers overestimating or underestimating the expected forecasts of the variables that define agricultural production (Bezerra, 2022; Lemos & Bezerra, 2019; Lessa, 2023; Sousa, 2023).

2. Method

The state of Ceará is located in the Northeast region of Brazil and is one of the poorest states in the country[2]. Its population in 2020 was 9187886 (IBGE, 2020) inhabitants. Ceará is almost completely politically recognized as being part of the semi-arid region. In fact, 175 of the state's 184 municipalities, as defined in the latest resolution of SUDENE's Deliberative Council. In physical terms, Ceará has approximately 95% of its territory officially recognized as being part of the semi-arid region of the Northeast (IBGE, 2020). Figure 1 shows the position of the state of Ceará in Brazil.

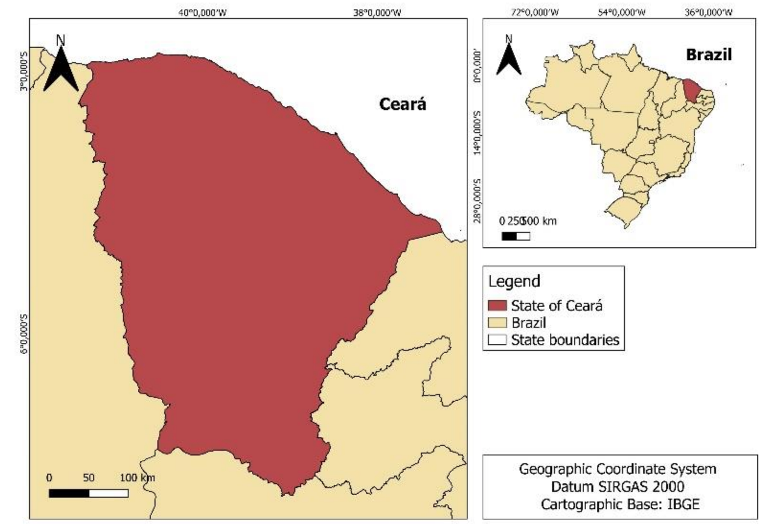

Figure 1. Geographic location map of the State of Ceará, Brazil

Source: IBGE (2022).

2.1 Database for the Research

The research used rainfall data collected from the National Oceanic and Atmospheric Agency (NOAA). Information on harvested areas, productivity and crop prices was collected from the Brazilian Institute of Geography and Statistics (IBGE), which provides data on Municipal Agricultural Production (IBGE, 2020). The period of analysis corresponds to the period between 1945 and 2020. Food crops such as rice, beans, cassava and corn were selected because they are grown by the majority of farmers in Ceará, as can be seen from the data provided by the 2017 Agricultural Census (IBGE, 2017). Prices were corrected to 2020 values, using the General Price Index (IGP-DI) as an indexer. The values were converted into US dollars based on the average exchange rate observed in 2020. The period of analysis extends from 1945 to 2020. The data for rice, beans, cassava and corn, which are mostly grown by farmers in Ceará, were collected from the IBGE (IBGE, 2020). Rainfall for the same period was obtained from the National Oceanic and Atmospheric Agency (NOAA) (NOAA, 2022).

2.2 Methodology for Achieving the First Objective: To Access Instabilities

The stability/instability of the selected variables was measured by the Coefficient of Variation (CV), which measures the percentage relationship between the standard deviation and the mean of a random variable. In practice, the CV measures the degree of homogeneity or heterogeneity of the distribution of the values of a random variable around its mean. It can be assumed that the CV measures the instability/stability of the way in which the observations of a random variable are distributed around its expected value. The greater the magnitude of the CV, the more unstable or heterogeneous the distribution of the observed values of the random variable around its mean. Thus, CV can also be interpreted as a measure of risk and has the added advantage of comparing variables measured in different units of measurement (Markowitz, 1952; Garcia, 1989; Gomes, 1985; Santos *et. al.*, 2021).

It is worth clarifying that in order to use CV as a measure of the stability or instability of a distribution, it is necessary to define amplitudes for its critical values. With this in mind, Gomes (1985) created the limits for classifying CVs, which are established as described in Table 1.

Table 1. Classification of the Coefficients of Variation (CV) according to their ranges

CV ranking	CV ranges
Low	CV < 10%
Medium	10% ≤ CV < 20%
High	20% ≤ CV < 30%
Very high	CV ≥ 30%

Source: Gomes (1985).

2.3 Methodology for Achieving the Objective «b»: Fitted Models to Make Projections of Decision Variables in Rainfed Crop Production in the State of Ceará

When analyzing time series, some concepts are relevant to understanding them and creating forecasting models. In this concept, it is worth noting that a random or stochastic process is configured as a set of observations of random variables ordered in time, which exhibit serial dependence and is known as a time series (Gujarati, 2011; Wooldridge, 2013). For these authors, in general, a stochastic process will be called stationary if its mean and variance are constant over time and the covariance between the variables does not depend on time.

Thus, considering the stationary time series represented by the random variable (Y_t) its predicted value (Y_P)will differ from the observed value due to the occurrence of random factors (ξt) along its trajectory. This information can be summarised in equation (1).

$$Y_t - Y_P = \xi_t\text{; or, similarly, } Y_t = Y_P + \xi_t \quad (1)$$

This study aims to assess the impact of annual rainfall on the forecasting capacity of the following variables: harvested areas, yields and prices of rice, beans, cassava and corn in Ceará from 1945 to 2020. It is assumed that these impacts will occur in the noise generated in the forecasting model, moving the observed values away from or closer to the projected value. Thus, it is assumed that the random term ξ_t can be represented as shown in equation (2):

$$\xi_t = f(C_t) \quad (2)$$

Substituting this value of ξ_t into equation (1) produces the result that will be tested in this research:

$$Y_t = Y_P + f(C_t) \quad (3)$$

The choice made in this study to find the projected values for Y_t was to adopt the Autoregressive Integrated Moving Average (ARIMA) model, developed by Box, Jenkins (1976). The following is a brief explanation of the ARIMA method as it applies to this study.

ARIMA model as it applies to this study

This model aims to capture the behavior of a random variable that has values distributed in the form of a time series. This model is suitable for stationary time series, or variables whose means, variances and auto-covariances are constant over time (Gujarati, 2011; Wooldridge, 2013). It is assumed that the time series Yt can be represented by equation (4):

$$Y_t = \mu + \Sigma\psi k u_{(t-k)} = \mu + \psi(B)u_t \quad (4)$$

In equation (4) the definition of the linear filter (ψ) is represented as follows:

$$\psi(B) = \theta(B)/\phi(B) \quad (5)$$

Equation (5) can be seen from the definition of the polynomials described below:

$$\theta(B)=1-\theta_1B_1- \theta_2B_2 -....- \theta qBq \quad e \quad \phi(B)= 1- \phi_1B_1- \phi_2B_2-...- \phi pBp \quad (5A)$$

In Equation (5) Box et al. (2016) clarify that: 1) $\phi(B)$ will be called an autoregressive operator and considered stationary if the roots of $\phi(B) = 0$ are outside the unit circle; 2) $\psi(B)$ will be called a generalized autoregressive operator, i.e. a non-stationary operator with "d" roots if $\varphi(B) = 0$ is equal to unity; and 3) $\theta(B)$ will be called the moving average operator. It is assumed that it can be inversion passive and that the roots of $(B) = 0$ are outside the unit circle.

If we define ($\tilde{Y}t$) as the expected value of Yt, then we can write that:

$$\tilde{Y}_t = Y_t - \mu_t$$

You can obtain its transformation in the time series as shown in equation (6):

$$\phi(B)\tilde{Y}_t=\theta(B)u_t \tag{6}$$

For the definition of equation (6), the random term "u_t" must be "white noise" and, for this, it must observe the following characteristics, according to Cochrane (1997):

$E(u_t)=0$; $E(u_t^2)=\sigma_u^2< \infty$; $E(u_t,u_{t+k})= 0$, quando $k = \pm1, \pm2, ...$

Based on this set of information, equation (6) can be rewritten as the Box and Jenkins (1976) model and is known as the autoregressive average of order p and the moving average of order q, or ARMA(p,q). This result is shown in equation (7):

$$\tilde{Y}_t= \theta(B)\phi^{-1}(B)u_t \tag{7}$$

As has already been discussed in this text, in order to use the Box-Jenkins methodology, the time series must be stationary. When this is not the case, stationarity must be sought by differentiating the time series. In general, with one, and up to a maximum of three differentiations, it is possible to transform a series that has been shown not to be stationary into a stationary series (Gujarati, 2011; Wooldridge, 2013; Paiva *et. al.*, 2022).

The steps to forecast the values of a time series using the Box-Jenkins model are as follows: i) Examine the series for stationarity. This step can be examined by measuring the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) or by performing a unit root analysis. ii) If the time series is not stationary, it will be necessary to differentiate one or more times until stationarity is reached (Box; Jenkins, 1976; Gujarati, 2011; Greene, 2012).

In order to find the best adjustments, this study used a number of criteria. One of these is the parsimony criterion with regard to the number of estimated parameters. The smaller this number, the better is the model fit. In addition, the following criteria were used: coefficient of determination (R^2); percentage of mean absolute error (MAPE); Ljung-Box Q statistic (tests whether the residuals generated in the adjusted model are "white noise") and Pearson's correlation coefficient between the observed series and the series projected by the estimated models. This coefficient must be high (close to one) and statistically different from zero (Box; Jenkins, 1976; Box *et al*., 2016; Greene, 2012; Wooldridge, 2013).

2.4 Methodology for Achieving the Objective "c": Relationship Between Forecasting Models and Rainfall in the State of Ceará

This research assumes that, in the scenario for forecasting harvested areas, yields and prices of selected crops (rice, beans, cassava and corn) in the semi-arid region, whose characteristics, as discussed, are rainfall instability, the shocks ξ_t can be affected by an exogenous variable: annual rainfall (C_t).

Thus, the noise ξ_t associated with the projections of harvested areas, yields and prices of rice, beans, cassava and corn between 1945 and 2020, as shown in Equation (2), is affected by annual rainfall, as shown in equation (8):

$$\xi_t= \lambda_0 + \lambda_1C_t+ \upsilon_t \tag{8}$$

In equation (8), the coefficient λ_0 represents the linear parameter; λ_1 is the angular coefficient which, if it is not statistically different from zero, implies that the residuals associated with the projections are not affected by rainfall. This means that the variables can be projected using the model as estimated. If λ_1 is statistically different from zero (regardless of the sign), this means that the model estimated to make the projections of the variables must take this fact into account. It is worth noting that if the errors induced by rainfall change the values projected by the farmers for the better, they will be welcome. On the other hand, if the errors induced by rainfall are negative, which means that the farmers' results will be less than they expected, these errors will be disastrous. This is because expecting to harvest a certain area, having productivity expectations or prices that are lower than those projected, due to the lack of rainfall, will be frustrating for them. The random term υ_t, by hypothesis, is also endogenously "white noise". If this hypothesis is true, the coefficients λ_0 and λ_1 can be estimated using the Ordinary Least Squares (OLS) method (Wooldridge, 2013).

2.5 Methodology for Achieving the Objective "d":

To achieve this goal, forecasted error (Є) is estimated by subtracting the forecasted value from the observed value (Є) = (Observed value – Forecasted value)] in the adjusted models to predict the decision variables chosen for the research (harvested area, productivity, prices of rice, beans, cassava and corn).

The occurrences of negative errors (-Є) and positive errors (+Є) are checked and aggregated. The sum of these two plots must be equal to (N-d), where "N" is the number of observations that make up the series, and "d" is the number of differentiations made to make the series stationary, if it is not. When the series is originally stationary d = 0. Thus, the following result is obtained

$$[\Sigma(+ \epsilon)] + [\Sigma(- \epsilon)] = (N\text{-}d)$$

Then the probabilities of the sum of the positive errors {P[Σ(+Є)]} and the probability of the sum of the negative errors {P[Σ(-Є)]} are estimated, as follows:

$$P[\Sigma(+\epsilon)] = [\Sigma(+\epsilon)] / (N\text{-}d); \text{ and } P[\Sigma(-\epsilon)] = [\Sigma(-\epsilon)] / (N\text{-}d).$$

It is clear that when the forecast errors are negatives; this implies that the farmers' expectations of the results found were higher than the results that actually occurred. This could be a measure of risk associated with rainfall instability, for example, affecting the variable for which the forecast was made. On the other hand, when the forecast errors are positive, this means that the farmers obtained results that exceeded their expectations.

3. Results and Discussion

The results, as well as the methodologies used, will be presented and discussed in accordance with the specific objectives of the research.

3.1 Results Found for Objective "a": Rainfall and Areas, Produvities, and Prices Instabilities

Initially, the minimum, maximum and average values were estimated, as well as the coefficients of variation (CV) of the harvested areas, yields and prices of rice, beans, cassava and corn in Ceará for the period from 1945 to 2020. These results are shown in Table 2.

Table 2. Minimum, maximum, average and Coefficients of Variation (CV) of the variables studied

Variáveis	**Minimum**	**Maximum**	**Average**	**CV (%)**
Annual rainfal (mm)	286,90	1773,40	777,76	33,56
Harvested area with rice (ha)	5250,00	79993,00	40532,01	45,34
Rice productivity (kg.ha^{-1})	409.68	3130.82	1939.54	33.87
Average price of rice (USD.kg^{-1})	0.93	6.25	0.52	43.94
Harvested area bean (ha)	74775.00	765654.00	381713.85	42.14
Bean productivity (kg.ha^{-1})	116.85	608.22	340.80	40.44
Average price of beans (USD.kg^{-1})	2.25	15.96	0.93	54.43
Harvested area with cassava (ha)	32283.00	176000.00	89810.80	39.75
Cassava productivity (kg.ha^{-1})	3356.92	16905.08	10178.85	32.25
Average price of cassava (USD.kg^{-1})	0.27	1.26	0.08	51.15
Harvested area with corn (ha)	78460.00	726777.00	433491.71	39.21
Corn productivity (kg.ha^{-1})	120.00	1254.14	646.66	42.30
Average corn price (USD.kg^{-1})	0.55	3.68	0.32	41.53

Source: PAM 2020 (IBGE, 2020); NOAA (2022).

From the evidence presented in Table 2, we can infer the high instability observed in annual rainfall in Ceará between 1945 and 2020, captured by the CV of 33.56%, classified as "very high" on the scale designed by Gomes (1985). It can be seen that this rainfall instability was synergistically transmitted to all the variables studied, all of which were classified as having very high levels of instability. In fact, the CVs of the variables that define rice, beans, cassava and corn production in Ceará between 1945 and 2020 ranged from 32.25% for annual cassava productivity to 54.43% for the average price of beans.

3.2 Results Found to Objectve «b»: ARIMA Models Fitted for Forecasting

The results found in the parameter estimates and statistical characteristics that measure the robustness of the estimated forecasting models are shown in Table 3. From this evidence it can be seen that all the series studied were not stationary and that they needed a difference (d =1) to achieve this characteristic. It can be seen that all the models are parsimonious in terms of the number of parameters estimated.

Table 3. Models fitted to the forecasts of harvested areas, yields and prices for rice, beans, cassava and corn in the state of Ceará between 1945 and 2020

		Variables	Area	Productivity	Price
Rice	Fitted models		ARIMA (0,1,1)	ARIMA (0,1,1)	ARIMA (2,1,2)
	AR	Lag1	0.000	0.000	0.598 *
		Lag2			-0.622*
	MA	Lag1	0.455*	0.643*	0.997*
		Lag2			-0.578*
	R^2		0.638	0.579	0.771
	Ljung Box		14.857 ^{NS}	10.765^{NS}	20.049^{NS}
	MAPE		27.059	24.334	15.476
	R Pearson		0.803*	0.764*	0.880*
	Variables		**Area**	**Productivity**	**Price**
Bean		**Fitted models**	**ARIMA (0.1.1)**	**ARIMA (0.1.1)**	**ARIMA (2.1.2)**
	AR	Lag1	0.000	-0.261**	0.670*
		Lag2			-0.433*
	MA	Lag1	0.725	0.678*	1.295
		Lag2			-0.679
	R^2		0.481	0.342	0.424
	Ljung-Box		8.149^{NS}	16.278^{NS}	13.836^{NS}
	MAPE		32.943	35.309	32.304
	R Pearson		0.701*	0.597*	0.671*
	Variables		**Area**	**Productivity**	**Price**
Cassava		**Fitted models**	**ARIMA (0.1.1)**	**ARIMA (0.1.1)**	**ARIMA (2.1.2)**
	AR	Lag1	0.847*	0.000	0.347**
		Lag2			
	MA	Lag1	0.000	0.439*	0.741
		Lag2			
	R^2		0.728	0.700	0.433
	Ljung Box		17.298^{NS}	8.323^{NS}	24.103^{NS}
	MAPE		16.040	15.785	101.726
	R Pearson		0.853*	0.842*	0.601*
	Variables		**Area**	**Productivity**	**Price**
Corn	**Fitted models**		**ARIMA (0.1.1)**	**ARIMA (2.1.0)**	**ARIMA (0.1.1)**
	AR	Lag1	0.000	-0.789*	0.000
		Lag2		-0.438*	
	MA	Lag1	0.661*	0.000	0.517*
		Lag2			
	R^2		0.475	0.074	0.620
	Ljung Box		10.817^{NS}	22.067^{NS}	0.662^{NS}
	MAPE		33.342	48.769	17.040
	R Pearson		0.698*	0.388*	0.792*

Sources: Values estimated from IBGE data (2020). Note: *Significant at 1%; NS = not significant at at least 15% error.

The Ljung Box statistics were statistically non-significant, at least at 10% error. This suggests that the noise generated in all the adjusted models is endogenously random. In all the estimated models, it was observed that the linear parameters were not statistically different from zero. From the evidence presented in Table 3, it can also be seen that, with the exception of the model estimated for predicting the price of cassava (MAPE = 101.726%), the percentages of the average absolute errors of the other models were relatively low. In the other cases, the MAPE ranged from 15.476% in the model estimated to predict bean prices to 48.769% in the model built to predict corn yields.

Although the adjustment estimated for cassava prices was the one whose results are shown in Table 3, caution should be exercised when using these estimates to make forecasts in this case in view of the high MAPE estimated. The study also estimated the correlation coefficients between the values observed in the original series and those predicted in the models generated. All these correlations were statistically different from zero, with a margin of error of less than 1%.

3.3 Results Found to Objectve «c»: Relationship Between the Residuals Generated in the Models and Rainfall

Table 5 shows the results found in the estimates of the relationship between rainfall and the residuals estimated in each of the models created for the forecasts of the variables studied. It can be seen that in the case of the estimated residuals for the forecasts of harvested areas, cassava prices and rice prices, annual rainfall did not significantly influence the residuals generated in the models, with at least a 10% error. For the other variables, in all the crops studied, the expectations made when the research was carried out were confirmed, with rainfall having an impact on the residuals generated in the models' forecasts, with error probabilities of less than 10% (Table 4).

Table 4. Results found to assess the relationship between the residuals estimated in the adjusted models and annual rainfall between 1945 and 2020 in Ceará

Crops	Variables (Resíduo)	Constant Coef.	Constant Sig.	Regression coefficient Coef.	Regression coefficient Sig.	Adjusted R^2
Rice	Harvested area	-12205.024	0.002	15.669	0.001	0.123
	Productivity	-482.001	0.001	0.672	0.000	0.156
	Average price	0.022	0.917	-7.577E-005	0.768	0.001
Bean	Harvested area	-140620.973	0.001	183.044	0.000	0.155
	Productivity	-96.334	0.018	0.119	0.017	0.063
	Average price	2.056	0.004	-0.003	0.002	0.124
Casava	Harvested area	-3282.076	0.630	4.105	0.621	0.030
	Productivity	-1052.990	0.110	1.367	0.091	0.039
	Average price	0.025	0.692	-3.264E-005	0.669	0.030
Corn	Harvested area	-177134.947	0.000	232.864	0.000	0.239
	Productivity	-330.204	0.000	0.457	0.000	0.184
	Average price	0.469	0.003	-0.001	0.003	0.114

Sources: Values estimated from IBGE data (2020) and NOAA (2022).

3.4 Results Found to Objectve «d»: How the Distribution of Rainfall Affects the Likelihood of Farmers Overestimating or Underestimating the Expected Forecasts of the Variables That Define Agricultural Production

The fact that rainfall has no statistical influence on the residuals generated in the forecasting models means that these estimated models can be used for this purpose, in the form in which they were estimated, given that the errors made in the forecast are not influenced by the observed rainfall. However, in order to make these projections, when rainfall induces forecasting errors, the model generated needs to incorporate this information. Table 5 shows the average values of the cumulative errors induced by rainfall, both positive and negative, as well as their respective probabilities of occurrence during the period analyzed. Table 5 also shows the average rainfall observed in the periods in which the positive and negative forecast errors were influenced by rainfall accumulated.

Table 5. Forecast probalities errors in harvested areas, yields and prices, induced by the temporal instability of rainfall in Cear ábetween 1945 and 2020

	Cumulative negative errors			Cumulative positive errors		
Forecast Errors in Variables	Prob. (%)	Average Error	Rainfall, Average	Prob. (%)	Average Error	Rainfall, average
Harvested area rice (ha)	40.8	2201.52	638.43	59.2	4212.04	1.047.74
Produtivit of rice ($kg.ha^{-1}$)	30.3	82.22	594.91	69.7	188.15	997.25
Harvested area of beans (ha)	40.8	23760.60	638.43	59.2	51162.03	1.047.74
Productivity of beans ($kg.ha^{-1}$)	43.4	19.20	648.16	56.6	29.72	1.059.31
Price of beans ($US\$. kg^{-}$)	76.3	0.88	977.31	23.7	0.35	569.75
Productivity cassava ($kg.ha^{-1}$)	42.1	173.27	643.54	57.9	386.90	1.053.32
Harvested area corn (ha)	42.1	27277.36	643.54	57.9	68146.31	1.053.32
Productivity of corn ($kg.ha^{-1}$)	30.3	58.3303	594.91	69.7	125.54	997.25
Price of corn ($US\$.kg^{-1}$)	96.1	0.43	901.46	3.9	0.09	377.75

Sources: Values estimated from IBGE data (2020) and NOAA (2022).

Taking as a reference the historical trajectory of expectations of the projected values of each of the variables based on the estimated models, the results shown in Table 5 are presented.

As can be seen from the evidence shown in Table 5, the probabilities of negative cumulative errors (with the exception of those observed for bean and corn prices) ranged from 30.3% (rice and corn yields) to 43.4% (bean yields). It can be seen that in the years in which there were negative forecast errors, the average rainfall observed in the state was always very close to the upper limit of the drought period, as defined in the research by Lemos & Bezerra (2019) and Lemos & Bezerra (2023). These results confirm the assumptions that motivated this research.

On the other hand, when rainfall is more abundant, the observed values tend to be higher than those predicted in the historical series of adjusted forecasts. This means that in periods of abundant rainfall, farmers who grow rainfed crops run less risk. For this reason, the probabilities of accumulated positive errors ranged from 56.6% (bean yields) to 69.7% (rice and corn yields).

With regard to the trajectories of the bean and maize price series (the cassava price projections were not affected by the rains), the observed price values will be further away from the trajectory (negative errors). In the case of the prices of these crops, when the rains are more abundant, as was the case for bean prices (977.31 mm) and maize prices (901.46 mm), production is higher and, as a result, supply increases in relation to demand, causing the observed prices to be lower than the projected ones (Table 5).

4. Conclusions

The research showed that the distribution of rainfall in Cear ábetween 1945 and 2020 is quite unstable, according to the scale defined by Gomes (1985). In fact, the coefficient of variation of the distribution of rainfall showed "very high" values 33,56%), reinforcing the result that confirms the instability of rainfall in Cear á during the period analyzed. This was evident in all the variables used to evaluate the production of rice, beans, cassava and corn in Cear á: harvested areas, yields and prices of rice, beans, cassava and corn in that period.

The study also generated robust models, from an econometric point of view, that can be used to project these variables between 1945 and 2020. The hypothesis that the errors generated in the forecasts were influenced by annual rainfall was only rejected in the cases of rice prices, harvested areas and cassava prices.

Based on the results of the research, it can be inferred that in the case of these decision variables, whose residuals were not influenced by rainfall, the estimated models can be used, in the form in which they were created, to make the forecasts, although caution should be exercised in relation to the forecasts for cassava prices, given that the best adjusted model in this case presented one of its statistics (MAPE) with a very high value, which is not desirable.

In the cases in which rainfall was found to have influenced the formation of the forecast error, it can be concluded that it is difficult to make forecasts for agricultural production in the semi-arid region of Cear á, given that induced errors that turn out to be negative will inevitably cause frustration for farmers, because their expectations for the variables in which this happens are impacted to lower values. And this happened with very high probabilities of between 30.3% and 43.4%. This evidence confirms that farmers in Ceará who grow these crops (the majority in the state) always run the risk of seeing the observed values of the variables that define the production of these crops lower than those they projected.

In general terms, the main conclusion of this study is that the temporal rainfall instability observed in the state of Ceará between 1945 and 2020 played an important role in the behavior of the projections of the variables that define the production of rice, beans, cassava and corn, crops mostly grown by family farmers who have the characteristic of being rainfed: they depend exclusively on rainfall to develop at all stages.

The evidence found in the research also allows us to conclude that having access to information, especially on how rainfall behaves historically, can help producers to devise better planning and decision-making strategies, trying to circumvent the impacts caused by rainfall instability, which is the norm and part of their daily lives. With this in mind, the discussion of the results found in this research is a contribution to the literature related to rainfed crop production in the semi-arid region, where irregular rainfall occurs throughout the years. This means that farmers always need to be prepared for this fact. To do so, they need to develop adaptive capacities and devise strategies to at least minimize their difficulties in living with this reality.

On the other hand, it is up to the state, in the broad sense (federal, state and municipal) to provide public policies for technical assistance, development, extension, rural insurance and to design, together with these individuals, non-agricultural alternatives for living with this reality. Otherwise, these farmers are potential migrants, which could happen whenever water scarcity problems arise.

Some of the difficulties encountered in the research relate to the estimated MAPE values obtained in six of the estimated models. This error is generally expected to be low. It can be seen that in six of the nine models estimated, the MAPE ranged from 32.304% for the model adjusted to estimate bean price forecasts to 101.726% for the model adjusted to forecast cassava prices. In addition, the models fitted to predict the harvested areas, yields and prices of beans, cassava prices, harvested areas and corn yields had coefficients of determination below 50%. Therefore, based on these difficulties, the forecasts to be made for these variables based on the adjusted models (which were the best achieved) need to be made with due care.

Acknowledgements

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for providing scholarships for all the PhD students who took part in this research, as well as the Federal University of Ceará for providing the salaries of the professor who is the first author of this article.

Authors' contributions

All the authors, with the exception of the first, are students on the Graduate Programme in Agricultural Economics at the Federal University of Ceará. They all contributed, along with the first author, to all stages of the research, from its conception, rationale, bibliographical research, data collection and processing, as well as writing the article in Portuguese, which was later translated into English by the first author, who is a professor of all of them.

Translated with DeepL.com (free version)

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of Sciedu Press.

The journal and publisher adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Beyer, M., Wallner, M., Bahlmann, L., Thiemig, V., Dietrich, J., & Billib, M. (2016). Rainfall characteristics and their implications for rain-fed agriculture: a case study in the Upper Zambezi River Basin. *Hydrological Sciences Journal, 61*(2), 321-343.

Bezerra, Filomena Nádia Rodrigues. (2022). Avaliação da agricultura de baixa emissão de carbono e inteligente ao clima no Brasil. 2022. 263 f. Tese (Doutorado em Economia Rural) - Centro de Ciências Agrárias, Universidade Federal do Ceará, Fortaleza.

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis. Forecasting and Control.* San Francisco: Holden-Day.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis***:** *forecasting and control* (5 ed.). John Wiley & Sons. Hoboken. New Jersey. Retrieved from http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_05_Box_Time-Series-Analysis-Forecasting-and-Control-2015

BRASIL. Conselho Deliberativo da Superintendência do Desenvolvimento do Nordeste. (2021). *Delimitação do semiárido-2021*. Relatório Final (Versão Preliminar). Recife. PE. Retrieved from https://www.gov.br/sudene/pt-br/centrais-de-conteudo/02semiaridorelatorionv.pdf

Carvalho, O. (2014). Perspectivas do desenvolvimento regional no Brasil. *Revista Política e Planejamento Regional, Rio de Janeiro, 1*(2), julho/dezembro, 295, a 310. Retrieved from http://www.revistappr.com.br/artigos/extra/55257a8c74a1f.pdf

Cochrane, J. H. (1997). *Time series for macroeconomic and finances*. Graduate School of Business. Time Series. Retrieved from http://www.bseu.by/russian/faculty5/stat/docs/4/Cochran

Costa Filho, J. (2019). Efeitos da instabilidade pluviométrica sobre a previsão da produção de lavouras de sequeiro em áreas sujeitas a desertificação (ASD) no semiárido do estado do Ceará: casos de Irauçuba e Tauá. 100f. *Dissertação (Mestrado)* – Universidade Federal do Ceará. Centro de Ciências Agrárias. Programa de Pós-Graduação em Economia Rural. Fortaleza.

Fischer, G., Shah, M., & Van Velthuizen, H. (2002). *Climate change and agricultural vulnerability*. Johannesburg: International Institute for Applied Systems Analysis to World Summit on Sustainable Development. Special Report.

Garcia, C. H. (1989). *Tabelas para classificação do coeficiente de variação*. Piracicaba: IPEF. Circular técnica – p. 171.12.

Gomes, F. P. (1985). *Curso de Estatística Experimental* (13th ed.). ESALQ/USP.

Greene, W. H. (2012). *Econometric Analysis* (71st ed.). New York: Stern School of Business. New York University.

Gujarati, D. N., & Porter, D. C. (2011). *Econometria Básica* (5th ed.). AMGH Editora Ltda. São Paulo. SP.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2017). *Censo Agropecuário 2017* **-**Resultados Definitivos. Rio de Janeiro. Retrieved from https://sidra.ibge.gov.br/pesquisa/censo-agropecuario/censo-agropecuario-2017

Instituto Brasileiro de Geografia e Estatística (IBGE). (2020). *Pesquisa Agrícola Municipal – PAM 2020*. Sistema IBGE de Recuperação Automática – SIDRA. Rio de Janeiro. Retrieved from https://sidra.ibge.gov.br/pesquisa/pam/tabelas

Köppen, W. (1918). Klassifikation der Klimatenach Temperatur. Niederschlag und Jahresablauf. In *Petermanns Geographische Mitteilungen* (September/ October, pp. 193-203). Mitt. Hamburg.

Köppen, W. (1936). Das geographische System der Klimate. In Köppen, W., & Geiger, R. (Eds.), *Handbuch der Klimatologie*. Berlin: Verlag von Gebrüder Borntraeger.

Lemos, J. J. S. (2020). *Vulnerabilidades induzidas no semiárido*. E-book. Fortaleza: Imprensa Universitária. Retrieved from http://www.repositorio.ufc.br/handle/riufc/54842

Lemos, J. J. S., & Bezerra, F. N. R. (2019). Interference of pluviometricinstability in the forecast of grain production in the semiárido of Ceará, Brazil. *Braz. J. of Develop., Curitiba, 5*(9), 15632-15652 sep. Retrieved from https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/3294/3157

Lemos, J. J. S., & Bezerra, F. N. R. (2023). Climate Resilience Agriculture (CRA) In The Brazilian Semi-Arid Region. *IOSR Journal of Humanities and Social Science, 28*(8), Series 8 (August, 2023) 01-10. Retrieved from https://www.iosrjournals.org/iosr-jhss/papers/Vol.28-Issue8/Ser-8/A2808080110.pdf

Lessa, Laura Cunha Rebouças. (2023). Resiliência e sustentabilidade da agricultura de sequeiro sob instabilidade pluviométrica na Paraíba. 2023. 96 f. Dissertação (Mestrado em Economia Rural) - Centro de Ciências Agrárias, Universidade Federal do Ceará, Fortaleza.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77-91.

Melo, R. F. de, Simoes, W. L., Pereira, L. A., Brito, L. T. de L., Ferreira, E. P., Barros, L. C. De, & Ribeiro, P. E. de A. (2019). Água para o fortalecimento dos sistemas agrícolas dependentes de chuva. In MELO, R. F., & Voltolini, T. F. (Eds.), *Agricultura familiar dependente de chuva no semiárido* (pp. 188-228). EMBRAPA Semiárido. Retrieved from https://ainfo.cnptia.embrapa.br/digital/bitstream/item/208457/1/Agua-para-o-fortalecimento-2019.pdf

Mohammed, R., & Scholz, M. (2019). Climate change and water resources in arid regions: uncertainty of the baseline time period. *Journal Theoretical and Applied Climatology*, *137*(1-2), 1365-1376. Retrieved from https://portal.research.lu.se/en/publications/climate-change-and-water-resources-in-arid-regions-uncertainty-of

Mohinder Singh, N. K. T., Naveen Kumar, K. R. D., & Dehinwal, A. K. (2017, October). Dry and Rainfed Agriculture- Characteristics and Issues to Enhance the Prosperity of Indian Farming Community. *BEPLS***,** *6*(10). Retrieved from http://bepls.com/OCT_2017/6.pdf

National Oceanic and Atmospheric Agency (NOAA). (2022, October). Global Historical-Monthly Climatology Network (GHCN-M). Washington. DC. Retrieved from https://www.globalclimatemonitor.org/#

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*, *60*(spe), e252091.

Pereira, G. R. (2018). *Correlação entre as secas e as perdas na agricultura de sequeiro no semiárido nordestino*. Instituto Nacional de Pesquisas Espaciais. Retrieved from https://editorarealize.com.br/editora/anais/conadis/2018/TRABALHO_EV116_MD1_SA23_ID185_191120181 14546.pdf

Rockstrom, J., *et al.*. (2010). Managing water in rainfed agriculture—The need for a paradigm shift. *Agricultural Water Management*, *97,* 543-55.

Rosenzweig, C., & Hillel, D. (2005). Climate change. agriculture and sustainability. In LAL, R. N., UPHOFF, B. A., Stewart., & Hansen. D. O. (Eds.), *Climate change and global food security* (pp. 243-268). Taylor & Francis. London. UK.

Salviano, J. I. A. (2021). *Relações entre instabilidades das chuvas e indicadores de produção de lavouras de sequeiro no semiárido cearense.* 130f. Dissertação (Mestrado) – Universidade Federal do Ceará. Centro de Ciências Agrárias. Programa de Pós-Graduação em Economia Rural. Fortaleza. Retrieved from http://repositorio.ufc.br/bitstream/riufc/59607/1/2021_dis_jiasalviano.pdf

Santos, C., & Dias, C. (2021). Note on the coeficiente of variation properties. *Brazilian Electronic Journal of Mathematics*, *2*(4), jul/dez.

Sousa, Erika Costa. (2023). Regionalização climática e projeções da produção per capita de lavouras alimentares de sequeiro no estado do Maranhão. 2023. Dissertação (Mestrado em Economia Rural) – Centro de Ciências Agrárias, Universidade Federal do Ceará, Fortaleza.

Thornthwaite, C. W. (1948). An Approach toward a rational classification of climate. *Geographical Review*, *38*(1), 55-94.

Thornton, P., Jones, P., Owiyo, T., Kruska, R., Herrero, M., Orindi, V., ... Omolo, A. (2008). Climate change and poverty in Africa: Mapping hotspots of vulnerability. *African Journal of Agricultural and Resource Economics*, *2*.

Wani, S. P., Sreedevi, T. K., Rockström, J., & Ramakrishna, Y. S. (2009). Rainfed agriculture: past trends and future perspectives. In *Rainfed Agriculture: Unlocking the Potential* (p. 1, 35). Retrieved from http://www.iwmi.cgiar.org/Publications/CABI_Publications/CA_CABI_Series/Rainfed_Agriculture/Protected/Rainfed_Agriculture_Unlocking_the_Potential.pdf

Wooldridge, J. M. (2013). *Introductory econometrics*: *A modern approach*. South Western Educational Publishing, p. 881.