

Rosendo Chagas de Albuquerque

**MODELAGEM EM SÉRIES TEMPORAIS: APLICAÇÃO
EM DADOS DE PRECIPITAÇÃO NA REGIÃO DO
SERTÃO DE PERNAMBUCO - BRASIL**

Recife

13 de janeiro de 2019



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**MODELAGEM EM SÉRIES TEMPORAIS: APLICAÇÃO EM DADOS DE
PRECIPITAÇÃO NA REGIÃO DO SERTÃO DE PERNAMBUCO - BRASIL**

Dissertação apresentada ao curso de
Mestrado em Biometria e Estatística
Aplicada, como pré-requisito para ob-
tenção do título de mestre

**Área de concentração: Biometria e Esta-
tística Aplicada**

**Orientador: Prof. Dr. Moacyr Cunha
Filho**

**Coorientador: Prof. Dr. Ricardo Alves
de Olinda**

Recife

13 de janeiro de 2019

Ficha Catalográfica

MODELAGEM EM SÉRIES TEMPORAIS: APLICAÇÃO EM DADOS DE
PRECIPITAÇÃO NA REGIÃO DO SERTÃO DE PERNAMBUCO - BRASIL/
Rosendo Chagas de Albuquerque. – Recife, 13 de janeiro de 2019-
Orientador: Prof. Dr. Moacyr Cunha Filho

– UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLI-
CADA, 13 de janeiro de 2019.

1. Precipitação. 2. Previsão. I. Moacyr Cunha Filho. II. Universidade Federal
Rural de Pernambuco. III. Modelagem em Séries Temporais: Aplicação em Dados
de Precipitação na Região do Sertão de Pernambuco - Brasil

CDU 02:141:005.7

Rosendo Chagas de Albuquerque

**MODELAGEM EM SÉRIES TEMPORAIS: APLICAÇÃO EM DADOS DE
PRECIPITAÇÃO NA REGIÃO DO SERTÃO DE PERNAMBUCO - BRASIL**

Dissertação apresentada ao curso de Mestrado
em Biometria e Estatística Aplicada, como
pré-requisito para obtenção do título de mes-
tre

COMISSÃO EXAMINADORA

Prof. Dr. Moacyr Cunha Filho
Orientador

Prof. Dr. Ricardo Alves de Olinda
Universidade Estadual de Paraíba
(UEPB)

Prof. Dr. Guilherme Rocha Moreira
Universidade Federal Rural de
Pernambuco (UFRPE)

APROVADO EM:

Dedico este trabalho, aos meus pais José de Oliveira e Maria de Fátima, a quem devo toda esta caminhada, a eles o meu reconhecimento e carinho.

A minha esposa Poliana Azevedo, pela compreensão e ajuda em todos os momentos, sua ajuda foi fundamental para conclusão desta caminhada.

A minha filha Sophie Chagas, lembro todas as vez que ela queria brincar na hora de estudar.

Aos professores, pelo estímulo, experiência e ensinamentos.

Aos meus amigos e colegas de estudos, pela compreensão e força nos momentos mais necessários.

Agradecimentos

Agradecimento especial ao professor Ricardo Alves de Olinda, por toda orientação e incentivo para realização deste trabalho, professor com quem sempre pude contar.

Ao Professor Moacyr Cunha, pela orientação segura e disponibilidade para realização deste trabalho.

Aos professores do Departamento de Estatística - UFRPE e aos colegas de aula, pessoas com quem sempre pude contar.

Aos familiares pelo incentivo e apoio, que se fizeram sempre presentes.

De forma especial, Poliana, Sophie, Leomir, Fábio, Abraão, Andre, Diego, Henrique, Anderson, Junior, Bia, Alan, Ciça, Osmar, Renan, Marcilio, Socorro, Fernando, Estela, Francisco, José e Fátima, um muito obrigado a todos.

Agradecimento a CAPES, pelo incentivo em todas as etapas da pesquisa.

Se você perdeu dinheiro, perdeu pouco.

Se perdeu a honra, perdeu muito.

Se perdeu a coragem, perdeu tudo.

(Van Gogh)

Resumo

Uma das grandes preocupações dos especialistas em ciências climáticas e meteorologistas, nos últimos anos, diz respeito às alterações climáticas e seus impactos para a humanidade. Este trabalho apresenta uma análise de séries temporais de dados de precipitação mensal, utilizando modelos da classe ARIMA (Autorregressivo Integrado de Médias Móveis). Objetivou-se modelar e prever a precipitação de três municípios (Arcoverde, Ouricuri e Petrolina), localizados na região do sertão de estado de Pernambuco, Brasil. Foram testados modelos do tipo SARIMA, ou seja, modelos ARIMA que consideram a característica sazonal da série. Os modelos selecionados foram aqueles que obtiveram os menores valores do critério de seleção de modelos Erro Quadrado Médio (EQM). Os modelos selecionados foram ARMA(3,3) município de Arcoverde, SARIMA (3,2)(1,1) município de Ouricuri e SARMA (2,3)(1,1) município de Petrolina, os quais, obtiveram um bom ajuste as séries de dados de precipitação estudadas. Os modelos selecionados e devidamente verificados, foram utilizados para realizar previsões de precipitação dos próximos meses, os resultados obtidos estão de acordo com o fato das séries serem todas estacionárias, sendo assim, sua média e variância são constantes no decorrer do tempo.

Palavras-chave: Box-Jenkins. Precipitação. Previsão.

Abstract

One of the major concerns of climate scientists and meteorologists in recent years is climate change and its impacts on mankind. This work presents an analysis of time series of monthly precipitation data, using models of the class ARIMA (Integrated Autoregressive of Moving Averages). The objective was to model and predict the precipitation of three municipalities (Arcaverde, Ouricuri and Petrolina), located in the hinterland region of Pernambuco State, Brazil. SARIMA models were tested, that is, ARIMA models that consider the seasonal feature of the series. The selected models were those that obtained the lowest values the criterion of selection of models Medium Square Error (EQM). The selected models were ARMA (3,3) municipality of Arcaverde, SARIMA (3,2) (1,1) municipality of Ouricuri and SARMA (2,3) (1,1) municipality of Petrolina, which obtained a good adjusted the series of precipitation data studied. The selected and duly verified models were used to perform precipitation forecasts of the next months, the results obtained are in agreement with the fact that the series are all stationary, so that their mean and variance are constant over time.

Key-words: Box-Jenkins. Precipitation. Forecast.

Lista de Figuras

Figura 1 – Tendências de uma série temporal (Barros, 2003)	9
Figura 2 – Precipitação acumulada mensal (mm) no período de 1970 a 2017 para os municípios de Arcoverde (A), Ouricuri (B) e Petrolina (C).	18
Figura 3 – Funções de autocorrelação (FAC)(a,c,e) e funções de autocorrelações parciais (FACP)(b,d,f), nos dados, para os municípios de Arcoverde, Ouricuri e Petrolina, respectivamente.	22
Figura 4 – Funções de autocorrelação dos resíduos(FAC)(A,C,E) e Estatística de Ljung-Box(B,D,F) para os municípios de Arcoverde, Ouricuri e Petrolina.	23
Figura 5 – Previsão da precipitação acumulada mensal(mm) para os próximo ano, nos municípios de Arcoverde, Ouricuri e Petrolina.	24

Lista de tabelas

Tabela 1 – Resultados obtidos com a aplicação dos respectivos testes, Dickey Fuller(estacionariedade) e Mann-Kendall(tendência), para as localidades de Arcoverde, Ouricuri e Petrolina.	19
Tabela 2 – Seleção dos modelos Auto-regressivos e de médias móveis ajustados para o município de Arcoverde, seguido dos respectivos critérios de avaliação.	20
Tabela 3 – Seleção dos modelos Espacial Auto-regressivos e de médias móveis ajustados para o município de Ouricuri, seguido dos respectivos critérios de avaliação.	20
Tabela 4 – Seleção dos modelos Espacial Auto-regressivos e de médias móveis ajustados para o município de Petrolina, seguido dos respectivos critérios de avaliação.	21
Tabela 5 – Estimativas dos parâmetros dos modelos selecionados.	21

Sumário

1	Introdução	1
2	Objetivos	2
2.1	Objetivo geral	2
2.2	Objetivos específicos	2
3	Revisão de Literatura	3
4	Metodologia	5
4.1	Processos Estocásticos	6
4.2	Estacionariedade	6
4.3	Tendência	9
4.4	Sazonalidade	11
4.5	Metodologia de Box-Jenkins	12
4.5.1	Auto-regressivos (AR)	12
4.5.2	Modelos de Médias Móveis (MA)	12
4.5.3	Modelos Auto-regressivos e de Médias Móveis (ARMA)	13
4.6	Modelos Auto-regressivos Integrado de Médias Móveis (ARIMA)	13
4.7	Modelos Auto-regressivos e de Médias Móveis (SARMA)	14
4.8	Etapas de Modelagem	15
5	Resultados e Discussão	18
6	Conclusão	25
	Referências Bibliográficas	26

1 Introdução

Uma série temporal é qualquer conjunto de observações ordenadas no tempo. Sendo assim, a modelagem em série temporal é uma das metodologias mais utilizadas para pesquisas nas mais variadas áreas do conhecimento, tais como: medicina, indústria, confiabilidade, economia, fenômenos meteorológicos, epidemiologia, ciências agrárias entre outras.

Os dados de precipitação têm atraído cada vez mais a atenção de diversos pesquisadores, uma vez, que a chuva interfere diretamente na economia do país. De tal modo, a modelagem em séries temporais vem sendo muito empregada nessa área de estudo seguindo a metodologia de Box e Jenkins, com intuito de ajustar um modelo que melhor represente os dados, sendo possível fazer previsões com estimativas próximas da realidade. Nesta pesquisa os dados são referentes a precipitação mensal de chuva no Sertão do Estado de Pernambuco, representado pelas cidades de Arcoverde, Ouricuri e Petrolina, sendo umas das cidades mais populosas do Sertão Pernambucano.

A vegetação do sertão de Pernambuco é composta pela Caatinga, com árvores de médio porte, arbustos e estepe, também apresenta o menor índice pluviométrico em relação as outras regiões do Estado.

Assim sendo, este trabalho consiste em ajustar modelos de séries temporais (ARMA e SARMA) aos totais mensais de precipitação dos municípios de Arcoverde, Ouricuri e Petrolina localizados na região do sertão do Estado de Pernambuco - Brasil, no período de 1970 a 2017, com intenção de encontrar o modelo que melhor representa seu caráter preditivo, testes estatísticos executados nesta etapa envolvem verificações das condições de estacionariedade e tendência na série. Bem como, fazer previsões de valores futuros da série, estas podem ser a curto prazo (precipitação diária, semanal) ou a longo prazo (precipitação mensal, anual).

A dissertação encontra-se organizada em três momentos: o primeiro, em que situamos o embasamento teórico da modelagem em séries temporal e sua relação aos estudos referentes a dados de precipitação; o segundo em que apresentamos dados sobre a precipitação de chuva dos municípios de Arcoverde, Ouricuri e Petrolina, e analisamos tendo como finalidade verificar a condição de estacionariedade, tendência e ciclos sazonais para realizar previsões; o terceiro momento em que apresentamos os resultados e discussões da metodologia de Box e Jenkins aplicada aos dados da presente pesquisa.

2 Objetivos

2.1 Objetivo geral

O objetivo deste trabalho consiste em aplicar modelos de séries temporais (ARMA e SARMA) aos totais mensais de precipitação nas cidades de Arcoverde, Ouricuri e Petrolina localizadas na região do sertão do estado de Pernambuco, Brasil, ao período de 1970 a 2017 com intenção de encontrar o modelo que melhor representa seu caráter preditivo.

2.2 Objetivos específicos

Descrever o comportamento da série, neste caso, a construção do gráfico da série histórica, a verificação de existência de tendência, ciclos e variações sazonais;

Identificar o modelo que melhor representa os dados em estudo, testes estatísticos, envolvendo verificações das condições de estacionariedade e tendência na série;

Realizar previsões de valores futuros da série, estas podem ser a curto prazo (precipitação diária, semanal) ou a longo prazo (precipitação mensal, anual);

3 Revisão de Literatura

O objetivo do trabalho apresentado por [Silva, Kayano e Sousa \(2010\)](#) foi usar o índice simples como o Índice de Anomalia de Chuva (IAC) e através da Análise de de Ondaleta (AO) estudar a variabilidade pluviométrica da bacia hidrográfica do rio Mundaú, examinando aspectos da variabilidade sazonal e interanual, bem como determinando as escalas temporais dominantes. Esta bacia localiza-se nos Estados de Alagoas e Pernambuco. As diferenças e similaridades entre as regiões também foram analisadas.

[Fisch e Valério \(2008\)](#) investigaram a variabilidade anual da precipitação observada no Vale do Paraíba e sua relação com fenômenos de teleconexão global, como é o caso dos eventos El Niño ou La Niña. Não foi observada uma conexão forte (do tipo causa/efeito) entre os eventos El Niño/La Niña e as chuvas anuais ou mesmo sazonais no Vale do Paraíba.

[FERRAZ \(1999\)](#) utilizaram modelos de séries temporais na previsão da série de precipitações pluviais mensais da cidade de Lavras, MG. Dentre os modelos ajustados, modelo de regressão, modelo de alisamento exponencial de HoltWinters e modelo SARIMA determinístico. Este último apresentou melhores valores de previsão, segundo os resultados obtidos pelos autores.

Um levantamento sobre a aptidão agroecológica do Estado de Pernambuco ([SOLOS, 2001](#)) mostrou que nas regiões do Agreste e do Sertão, uma fração muito reduzida das áreas propícias para atividade voltada à agricultura, que é onde se situam as principais bacias leiteiras do Estado, cuja aptidão para pastagens nativas e cultivadas mostra relativamente elevada. Sua vegetação é composta pela Caatinga, com árvores de médio porte e arbustos que perdem as folhas nas estações secas. Alguns animais da Caatinga são os lagartos, as serpentes, o carcará, a capivara, o macaco-prego, o veado-catingueiro e aves.

Para um melhor entendimento no estudo da climatologia é necessário sempre estar estudando os dados para conhecer as características das chuvas e sua sazonalidade que ocorrem em função da temperatura, umidade relativa do ar e direção do vento: O objetivo de modelar séries temporais está associado ao entendimento do seu padrão de comportamento, descrição e previsão desses fenômenos naturais ([PADILHA et al., 2013](#)).

[Marengo e Alves \(2005\)](#) estudaram a ocorrência de tendência hidrológica na Bacia do Rio Paraíba do Sul, usando longas séries temporais de dados de precipitação e vazão fluvial para várias localidades e concluíram que, embora tenham observado vazões com tendências negativas, não há como associar tal redução a uma possível diminuição da precipitação, mas sim à forma de gerenciamento das águas dos reservatórios da região

(principalmente o reservatório de Paraibuna).

Já no trabalho realizado por [Freitas et al. \(2010\)](#), são utilizados modelos de séries temporais com o objetivo de identificar tendências anuais de precipitação e de vazão na Bacia Hidrográfica do Rio São Mateus, no Estado do Espírito Santo.

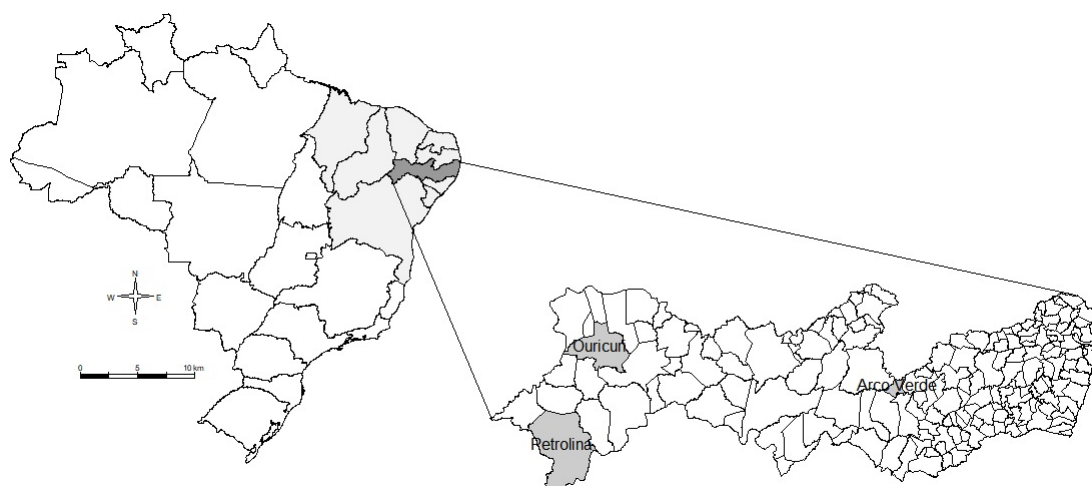
Uma aplicação de modelos do tipo ARIMA foi realizada por [Gemitzi e Stefanopoulos \(2011\)](#) para investigar os efeitos das condições meteorológicas e das intervenções do homem nas águas subterrâneas. Os autores obtiveram bons resultados e sugerem o uso de modelos estocásticos como uma alternativa, especialmente devido à limitação na obtenção de dados em relação a propriedades hidráulicas e geométricas dos aquíferos.

4 Metodologia

A iníco serão apresentados os resultados das séries históricas das cidades de Arcoverde, Ouricuri e Petrolina, localizadas no Estado de Pernambuco, os dados foram obtidos na estação meteorológica digital, Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP), relativos no período de 1970 a 2017. Na avaliação da precipitação foram considerados os totais mensais das cidades estudadas.

No primeiro momento verificou-se os gráficos da série histórica das cidades de Arcoverde, Ouricuri e Petrolina, trata-se de um procedimento simples e esclarecedor. No qual permite ter uma ideia do comportamento da série a respeito de sua estacionariedade, tendência e sazonalidade, na sequência da análise testes estatísticos foram aplicados aos dados com finalidade de comprovar as observações da primeira parte, prosseguindo logo após conhecer as características da série, inicia-se a seleção dos modelos que melhor representa os dados nesta etapa é de fundamental importância fixar os critérios de seleção adotado na pesquisa. Logo após a seleção dos modelos devidamente selecionados e verificados, a ultima etapa a ser realizada é a previsão da série, neste estudo a previsão foi apresentada de forma gráfica, uma vez que as séries estudadas sejam estacionarias a previsão também vai ter sua média e variância constante no tempo.

A seguir sera apresentado um gráfico com a localização dos municípios, Arcoverde, Ouricuri e Petrolina.



Todas as análises foram realizadas com o auxílio do software ([RCORETEAM](http://www.rcoreteam.com), 2018), encontrado de forma livre no site (www.r-projet.org).

4.1 Processos Estocásticos

Seja T um conjunto arbitrário, um processo estocástico é uma família $Z = \{Z_{(t)}, t \in T\}$, tal que para cada $t \in T$, $Z_{(t)}$ é uma variável aleatória.

O conjunto T é, normalmente, considerado como o conjunto dos inteiros \mathbb{Z} , ou conjunto dos números reais \mathbb{R} , e o conjunto de variáveis aleatórias, distribuídas equiespaçadamente no tempo, definidas num mesmo espaço de probabilidades (Ω, A, P) , (MORETTIN; TOLOI, 2004), em que Ω , é o espaço amostral; A são os eventos associados ao espaço amostral e P é uma medida de probabilidade.

Para dizer que um processo estocástico está especificado, necessita-se de certas condições, isto é, pressupostos. Sejam t_1, t_2, \dots, t_n elementos quaisquer de T e considerando-se

$$F(z_{(1)}, \dots, z_{(n)}; t_1, \dots, t_n) = P[Z_{(t_1)} \leq z_1, \dots, Z_{(t_n)} \leq z_n],$$

então, o processo estocástico $Z = \{Z_{(t)}, t \in T\}$ estará especificado se conhecermos as distribuições para todo $n > 1$.

4.2 Estacionariedade

Uma sériel temporal é estacionária quando suas características estatísticas como média, variância, autocorrelação, são constantes ao longo do tempo. É uma série que se desenvolve aleatoriamente no tempo, em torno de uma média constante (SATO, 2013).

A maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, portanto, será necessário transformar os dados originais se estes não formam uma série estacionária. A transformação mais comum consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária. A primeira diferença $Z_{(t)}$ é definida por

$$\Delta Z_{(t)} = Z_{(t)} - Z_{(t-1)}$$

então, a segunda diferença é

$$\begin{aligned} \Delta^2 Z_{(t)} &= \Delta[\Delta Z_{(t)}] = \Delta[Z_{(t)} - Z_{(t-1)}]; \\ &= \Delta^2 Z_{(t)} = Z_{(t)} - 2Z_{(t-1)} + Z_{(t-2)}. \end{aligned}$$

De modo geral, a n -ésima diferença de $Z_{(t)}$ é dada por

$$\Delta^n Z_{(t)} = \Delta[\Delta^{n-1} Z_{(t-1)}].$$

Normalmente, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária.

Segundo MORETTIN e TOLOI (2004), processos estacionários caracterizam-se pelo comportamento não alterado ao longo do tempo, o processo se desenvolve no tempo em torno de uma média constante ao longo de t . Logo um processo estocástico $Z = Z_{(t)}, t \in T$ é dito estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja, $F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n)$, para quaisquer $t_1, \dots, t_n, \tau \in T$. Isto significa que todas as distribuições unidimensionais são invariantes sob translações ao longo do tempo, logo a média e a variância são constantes.

Testes de Estacionariedade

Segundo Arêdes e Pereira (2008), o teste mais simples para analisar a estacionariedade em séries temporais é dado pela obtenção dos coeficientes de autocorrelação e autocorrelação parcial, a partir dos quais são construídos os respectivos correlogramas: FAC (Função de Autocorrelação) e FACP (Função de Autocorrelação Parcial), que por sua vez representam as inspeções gráficas das defasagens. Este correlograma delimita um intervalo de confiança para os coeficientes no qual as estatísticas da FAC e FACP devem variar entre:

$$IC(95\%) = 0 \pm 1,96 \times \left(\frac{1}{\sqrt{n}}\right), \quad (4.1)$$

em que n é o tamanho da amostra, 1,96 é o quantil (0,025) da normal padrão.

O teste de Dickey-Fuller considera o modelo:

$$Y_t = \rho Y_{t-1} + \mu_t \quad -1 \leq \rho \leq 1 \quad t = 1, 2, \dots, N,$$

em que Y_t é a série a ser modelada, ρ o coeficiente de Y_{t-1} e μ_t é um termo de ruído branco, com distribuição normal.

Para a realização do teste a equação $Y_t = \rho Y_{t-1} + \mu_t$ é trabalhada na forma de diferenças, conforme:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \mu_t = (\rho - 1)Y_{t-1} + \mu_t$$

que pode ser escrita conforme a equação:

$$\Delta Y_t = \delta Y_{t-1} + \mu_t$$

onde $\delta = \rho - 1$ e Δ é o primeiro operador de diferença.

Em [Dickey e Fuller \(1981\)](#) é demonstrado que, sob a hipótese nula de que $\delta = 0$, o valor estimado do coeficiente de Y_{t-1} na equação $\Delta Y_t = \delta Y_{t-1} + \mu_t$, segue a estatística δ (delta). Mediante a utilização da simulação de Monte Carlo, os autores computaram os valores fundamentais dessa estatística.

O teste de Dickey-Fuller foi criado para verificar se um modelo autorregressivo tem ou não raiz unitária. Para que a série temporal z_t seja estacionária tem-se que obter ϕ que atenda a restrição $|\phi_1, \dots, \phi_n| < 1$. Logo, as hipóteses do teste são

$$H_0 : \phi = 1, z_t \text{ não é estacionária};$$

$$H_1 : |\phi| < 1, z_t \text{ é estacionária}$$

Nesse caso, a hipótese nula H_0 é de que exista pelo menos uma raiz unitária, logo a variável não é estacionária. Por sua vez, a hipótese alternativa H_1 é que a variável seja francamente estacionária, nesse caso não há nenhuma raiz unitária.

O teste de Kwiatkowski-Philips-Schmidt-Shinn (KPSS), proposto em [Kwiatkowski et al. \(1992\)](#), avalia a estacionariedade considerando um modelo com tendência, passeio aleatório e erro, conforme a equação:

$$Y_t = \xi t - r_t + \varepsilon_t,$$

em que ξ é uma constante que representa o coeficiente de t , ε_t é um erro estacionário e um passeio aleatório $r_t = r_{t-1} + \mu_t$, onde μ_t é independente e identicamente distribuído (iid) com distribuição normal, média zero e variância σ^2 ; e Y_t a variável estudada.

Considerando que ε_t representa os resíduos de uma regressão em Y_t , explicado pelas componentes de tendência, passeio aleatório e intercepto, um estimador para a variância dos erros dessa regressão, ainda de acordo com [Kwiatkowski et al. \(1992\)](#), é dado pela equação:

$$\hat{\sigma}_\varepsilon^2 = \frac{SQE}{N},$$

em que SQE é a Soma dos Quadrados dos Erros e N é o número de observações.

Define-se a soma parcial dos resíduos, S_t , na seguinte equação:

$$S_t = \sum_{i=1}^t e_i, \quad t = 1, 2, \dots, N$$

de modo que a estatística LM é dada pela equação:

$$LM = \sum_{t=1}^N \frac{S_t^2}{N^2 \hat{\sigma}_\varepsilon^2},$$

em que $\hat{\sigma}_\varepsilon^2$ é a estimativa da variância de longo prazo de Y_t . Essa estatística é utilizada para testar as hipóteses:

$H_0 : \sigma_\mu^2 = 0$ (o modelo não possui raiz unitária, a série é estacionária);

$H_1 : \sigma_\mu^2 > 0$ (o modelo possui raiz unitária, a série não é estacionária).

Os valores da estatística LM são comparados com valores tabelados para dar suporte à tomada de decisão.

4.3 Tendência

A tendência de uma série indica o seu comportamento de longo prazo, isto é, uma série com tendência caracteriza-se por revelar, ao longo do tempo, um comportamento que pode ser linear, não-linear, crescente, decrescente ou constante. Pode-se representar esses tipos de séries da seguinte forma:

$Y_t = \mu_t + \alpha_t$; $t = 1, 2, \dots, n$ em que o nível $\mu_t = \mu_{t-1} + \tau_{t-1}$, (em que τ_{t-1} é a taxa de crescimento da série no instante t) varia de acordo com a tendência e α_t são observações de um ruído branco.

Nos casos mais comuns trabalha-se com tendência constantes, linear ou quadrática, como ilustrado a seguir:

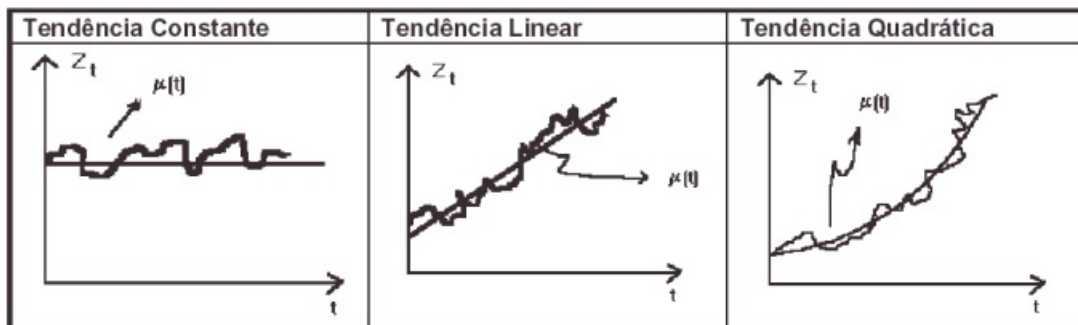


Figura 1 – Tendências de uma série temporal (Barros, 2003)

Testes de Tendência

O teste sequencial de Mann-Kendal (SNEYERS, 1975), é uma análise estatística não paramétrica que considera a hipótese de estabilidade de uma série temporal cuja sucessão de valores ocorre de forma independente e a distribuição de probabilidade deve permanecer sempre a mesma (série aleatória simples) Pinheiro, Graciano e Severo (2013).

O teste considera uma série temporal Y_i de N termos a ser analisada, sendo: $1 \leq i \leq N$. O procedimento consiste em realizar a soma $t_n = \sum_{i=1}^N m_i$ do número de termos m_i da série, relativo ao valor Y_i cujos termos precedentes ($j < i$) são inferiores ao mesmo

($Y_j < Y_i$). Para séries com grande número de termos (N), sob a hipótese nula de ausência de tendência H_0 , t_n apresentará uma distribuição normal com média e variância dada por:

$$E(t_n) = \frac{N(N-1)}{4} \quad \text{var}(t_n) = \frac{N(N-1)(2N+5)}{72}$$

Testando a significância estatística de t_n para a hipótese nula, usando um teste bilateral, esta pode ser rejeitada para grandes valores da estatística $U(t_n)$ fornecida pela equação:

$$U(t_n) = \frac{t_n - E(t_n)}{\sqrt{\text{var}(t_n)}}$$

O valor da probabilidade α_1 é calculado por meio de uma tabela e distribuição normal padronizada, tal que: $\alpha_1 = \text{prob}(|U| > |U(t_n)|)$. Sendo α_0 o nível de significância do teste (geralmente $\alpha_0 = 0,05$), a hipótese nula (H_0) é aceitar se $\alpha_1 > \alpha_0$. Caso a hipótese seja rejeitada, implicará a existência de tendência significativa, sendo que sinal de $U(t_n)$ indica se na tendência é crescente $U(t_n) > 0$ ou decrescente $U(t_n) < 0$.

O teste *Run* é um teste não-paramétrico apresentado por [Thom \(1966\)](#), que é usado para avaliar se uma série ocorre aleatoriamente. Consiste em realizar a contagem do número de oscilações dos valores acima e abaixo da mediana, numa série de dados naturalmente ordenada. O número de oscilações é chamado de *Run*, e deve-se testar se o valor observado está dentro da faixa de distribuição considerada normal. Um valor alto de *Run* indica muitas oscilações, e valores baixos indicam um desvio em relação à mediana durante o período de registros.

Se a sequência contém N_1 símbolos de um tipo, e N_2 símbolos de outro tipo, a distribuição amostral do números de *Runs* total pode ser aproximada pela distribuição normal com média:

$$E(\mu) = \frac{2N_1N_2}{N_1 + N_2} + 1$$

e a variância da distribuição pode ser estimada por

$$\text{Var}(\mu) = \frac{2N_1N_2(2N_1N_2) - N_1 - N_2}{(N_1 + N_2)^2(N_1 + N_2 - 1)},$$

em que u representa o número de *Runs*. Por isso, a hipótese nula (H_0) de que as observações são iid, pode ser testada com base na estatística: $Z = \frac{\mu - E(\mu)}{\sqrt{\text{var}(\mu)}}$. Este valor calculado pode ser comparados com valores de Z para distribuição normal. Para o nível de significância de 5%, Z deve estar entre $-1,96$ e $1,96$. Caso Z calculado seja maior que o valor tabelado, deve-se rejeitar a hipótese de nulidade, portando as observações da série possuem tendência monótona ao longo do tempo.

4.4 Sazonalidade

As variações sazonais podem ser aferidas por medidas diárias, semanais ou mensais. É preciso notar que há alguma irregularidade na forma de registro do tempo, nas avaliações de sazonalidade.

[Figueredo \(2008\)](#), define sazonalidade como o conjunto dos movimentos ou flutuações com período igual ou inferior a um ano, "sistemático, mas não necessariamente regulares", que ocorrem em uma série temporal. Para [Figueredo \(2008\)](#), tais movimentos decorrem normalmente das variações climáticas relacionadas às estações do ano, ou a fatores culturais relacionados a efeitos de calendário.

Também é possível utilizar métodos para determinar sazonalidade determinísticas, sendo necessário eliminar antes a tendência da série se ela existir. Considera-se que não existe sazonalidade determinística na série temporal se a hipótese nula de que todos os coeficientes dos termos sazonais são nulos não for rejeitada.

É importante reconhecer a presença do componente sazonal, estimá-lo e removê-lo, para então analisar o comportamento da tendência de longo prazo da série. De forma geral, os estudos tradicionais sobre sazonalidade tentam decompor o fenômeno em três partes, através de: $Y_t = \text{tendência} + \text{sazonalidade} + \text{irregularidade}$. Na qual o componente de tendência representa o movimento de longo prazo. O movimento irregular reflete movimentos não-sistemáticos e residuais. Isolando-se os efeitos da tendência e dos erros não-sistemáticos, pode-se então captar os efeitos sazonais ([ANTUNES; CARDOSO, 2015](#)).

Testes para Sazonalidade

O teste de Kruskal-Wallis [Marôco \(2011\)](#) supõe uma amostra de uma população, subdividida em $K > 4$ conjuntos de amostras com tamanho n não necessariamente iguais entre si. Onde K é o número de anos e n observações da série.

A hipótese nula toma em H_0 a não existência de sazonalidade, que é testada através da seguinte estatística

$$T_1 = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1),$$

Rejeita-se H_0 quando essa estatística for maior ou igual ao valor crítico T_{1c} , onde T_{1c} é tal que $P_H(T_1 > T_{1c}) = \alpha$, α é o nível de significância do teste.

Outro teste que pode ser utilizado é o teste F de Snedecor a uma análise de variância. O modelo subjacente é

$$Y_{ij} = S_j + e_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, k,$$

supondo-se que $e_{ij} \sim N(0, \sigma^2)$, independentes. Sob a hipótese nula $H_0 : S_1 = S_2 \dots = S_k$, a estatística

$$T = \frac{N - k}{k - 1} \frac{\sum_{j=1}^k n_j (\hat{Y}_j - \hat{Y})^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \hat{Y}_j)^2}$$

tem distribuição F e graus de liberdades $(k - 1, N - k)$, descrito da seguinte forma: $F_{(k-1, N-k)}$.

4.5 Metodologia de Box-Jenkins

Tal metodologia consiste em propor e ajustar modelos lineares estacionários ou não-estacionários a uma série de tempo observada. A estratégia para a construção do modelo será baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseado nos próprios dados

4.5.1 Auto-regressivos (AR)

Segundo [Padilha et al. \(2013\)](#), O modelo Auto-Regressivos (AR) foram criados com a ideia de que a presente observação de série Z_t pode ser explicada como uma função das p observações passadas, $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$, em que p determina o número de passos entre as observações passadas e a previsão da próxima observação. A estrutura auto-regressiva geral é expressa por:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t,$$

em que ϕ_i são os parâmetros da estrutura auto-regressiva, $i = 1, \dots, p$ (ordem da estrutura); a_t é o ruído branco com média zero e variância σ_a^2 .

4.5.2 Modelos de Médias Móveis (MA)

Segundo [Souza et al. \(2006\)](#), os modelos de médias móveis são formados por combinações lineares de ruído branco, a_t , ocorridos no período corrente e nos períodos passados. A estrutura de médias móveis geral é expressa por

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \dots - \theta_q a_{t-q},$$

em que, θ_i são os parâmetros da estrutura médias móveis, $i = 1, \dots, q$ (a ordem da estrutura); a_t é o ruído branco com média zero e variância σ_a^2 .

4.5.3 Modelos Auto-regressivos e de Médias Móveis (ARMA)

Box e Jenkins (1994) formalizam a teoria da utilização de componentes auto-regressivos e de médias móveis na modelagem de séries temporais utilizando-se de duas ideias básicas na criação de sua metodologia de construção de modelos: a parcimônia que é a utilização do menor número possível de parâmetros para se obter uma representação adequada do fenômeno em estudo e a interatividade, ou seja, a informação empírica é analisada teoricamente e o resultado deste estágio é confrontado com a prática sucessivas vezes, até que o modelo obtido seja satisfatório.

Segundo Figueredo (2008) o modelo ARMA é formado por uma combinação do modelo AR com o modelo MA, onde Z_t é descrito por seus valores passados e pelo ruído branco corrente e passados.

A estrutura geral do modelo ARMA(p,q) é expressa por:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

em que, ϕ_i são os parâmetros da estrutura auto-regressiva, $i = 1, \dots, p$; θ_i são os parâmetros da estrutura médias móveis, $i = 1, \dots, q$; a_t é o ruído branco.

Segundo MORETTIN e TOLOI (2004), modelos com essa característica são bastante utilizados em algumas áreas, como em Economia, onde é natural o valor de alguma variável no instante t como função de valores defasados da mesma variável. Em outras áreas, como em ciências, física e geografia, o interesse em modelos auto-regressivos reside em outro aspecto que não seja o da previsão: deseja-se estimar o espectro do processo e os estimadores auto-regressivos são utilizados para tal fim.

4.6 Modelos Auto-regressivos Integrado de Médias Móveis (ARIMA)

De acordo com BAYER (2010) a classe de modelos tradicionalmente mais utilizada em análise de séries temporais é a dos modelos ARIMA. Esta metodologia consiste em ajustar modelos autorregressivos integrados de média móveis, a um conjunto de dados.

Se Z_t segue um processo autorregressivo e de média móveis de ordens p, q , ARMA(p,q), estão

$$\phi(B)Z_t = c + \theta(B)\varepsilon_t, \quad (4.2)$$

em que ε_t é ruído branco com média zero e variância σ^2 , c é um parâmetro que permite ao processo ter média diferente de zero, o polinômio $B^k Z_t = Z_{t-k}$ é o operador retroativo, $\phi(B)$ e $\theta(B)$ são o polinômio autorregressivo e o polinômio de médias móveis, respectivamente,

apresentados nas equações abaixo:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (4.3)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (4.4)$$

Uma série que não apresenta característica de estacionariedade pode torna-se estacionária pela utilização de um processo de diferenciação. Sendo assim, se uma série se torna estacionária após d diferenças, a série é dita integrada (I) de ordem d . Com isso, o modelo ARMA integrado passa a ser denominado de modelo ARIMA.

O modelo ARIMA(p,d,q) tem a seguinte forma:

$$\phi(B)(1 - B)^d Z_t = \theta(B)\varepsilon_t, \quad (4.5)$$

em que d é a ordem de integração, ou seja, o número de diferença necessárias para torna a série estacionária (FAVA, 2000).

4.7 Modelos Auto-regressivos e de Médias Móveis (SARMA)

Os modelos espacial auto-regressivos e de médias móveis SARMA(P,Q)s podem ser combinados em uma especificação seguindo a ideia nos modelos auto-regressivos e médias móveis (ARMA). Os modelos SARMA têm uma especificação da forma:

$$y = \rho \mathbf{W}_1 y + X\beta + \mu,$$

na qual os resíduos da equação observada possuem uma estrutura autorregressiva, da forma:

$$\mu = \lambda \mathbf{W}_2 \varepsilon + \varepsilon.$$

As matrizes \mathbf{W}_1 e \mathbf{W}_2 são matrizes de contiguidade não necessariamente iguais. De fato, quando $\mathbf{W}_1 = \mathbf{W}_2$, o modelo é não identificado, e as estimativas para os coeficientes λ e ρ podem resultar bastante instáveis (ESPINOSA; PRADO; GHELLERE, 2010). Uma das críticas em relação à utilização dos modelos SARMA é justamente o fato de eles exigirem, em alguns casos, a especificação de duas matrizes de contiguidade diferentes. Em geral, a escolha de uma matriz de contiguidade é arbitrária, a escolha de duas matrizes diferentes implica em um grau de arbitrariedade ainda mais criticável.

Espinosa, Prado e Ghellere (2010), a estimação dos parâmetros do modelo SARMA pode ser feita via máxima verossimilhança. A partir das expressões acima, pode-se escrever $(I - \rho \mathbf{W}_1)y = X\beta + (I - \lambda \mathbf{W}_2)^{-1}\varepsilon \Rightarrow y = (I - \rho \mathbf{W}_1)^{-1}X\beta + (I - \rho \mathbf{W}_1)^{-1}(I - \rho \mathbf{W}_2)^{-1}\varepsilon$.

Assumido-se que ε possui distribuição normal multivariada, com média zero e covariância $\sigma^2 I$, conclui-se que o vetor de observações para variável resposta y possui distribuição normal multivariada com média condicional

$$E[Y|X = x] = (I - \rho \mathbf{W}_1)^{-1} X \beta,$$

e matriz de variância condicional

$$\Sigma_{y|X} = \sigma^2 (I - \rho \mathbf{W}_1)^{-1} (I - \rho \mathbf{W}_2)^{-1} [(I - \rho \mathbf{W}_1)^{-1} (I - \rho \mathbf{W}_2)^{-1}]^T$$

Utilizando-se a fórmula para a distribuição normal multivariada, pode-se chegar à função de log-verossimilhança $\log L(\lambda, \rho, \beta, \sigma^2)$, como função dos parâmetros desconhecidos do modelo.

4.8 Etapas de Modelagem

A aplicação dos modelos para fins e previsão de série temporais segue as seguintes etapas: identificação, estimação, verificação e previsão.

Na primeira fase o que se deseja é identificar o processo aleatório que gerou os dados, para em seguida estimar os parâmetros que o caracterizam e verificar se as hipóteses do modelo foram cumpridas. Caso negativo, uma nova fase de identificação deve ser considerada até que a verificação das hipóteses seja finalmente positiva, permitindo então a realização da previsão.

Identificação do Modelo

Na etapa de identificação de um modelo empregam-se procedimento que possam identificar a estrutura do modelo, isto é, permitam conhecer os valores dos parâmetros do modelo que caracterizam o processo estocástico.

Os procedimentos de identificação consistem em duas partes: 1) inicialmente diferencia-se a série temporal original tantas vezes quantas necessárias para obtenção de uma série estacionária, de modo a possibilitar a análise do processo com o modelo ARMA(p,q); o número de diferenciações d é aquele necessário para que a função de autocorrelação amostral (ACF) da série transformada decresça rapidamente para zero; 2) a identificação de um processo AR(p), MA(q) ou ARMA(p,q) é feita através da análise das funções de autocorrelação simples (ACF) e da autocorrelação parcial (PACF), com determinação dos valores dos parâmetros p , q (BOX, 1994); (MORETTIN; TOLOI, 2004).

Um ponto importante na identificação do modelo ARIMA é a parcimônia, ou seja, deve-se buscar um modelo com melhor ajuste e menor número possível de parâmetros. Em

outras palavras, dados dois modelos com ajustes igualmente bons, escolhe-se aquele com menor número de parâmetros.

Estimativas dos Parâmetros

Após a identificação dos valores d , p , q passa-se à estimativa dos parâmetros do modelo. Esta etapa é geralmente executada através de "software" específico para análise de série temporais como programa computacional (RCORETEAM, 2018), utilizado no desenvolvimento desta pesquisa. Existem diferentes métodos para a estimativa dos parâmetros do modelo, e todos produzem resultados semelhantes. De maneira geral, durante a etapa de estimativa usa-se algum procedimento iterativo de estimação de mínimos quadrados não linear ou através dos métodos de máxima verossimilhança. (MORETTIN; TOLOI, 2004)

Verificação através de testes estatísticos

Ao ajustarmos uns modelos AR, MA, ARMA, ARIMA, SARIMA, SARMA, entre outros, estamos buscando capturar toda a estrutura de dependência serial presente na série. Logo, os resíduos do modelo não devem apresentar qualquer tipo de dependência no tempo. Num modelo bem ajustado, os erros de previsão um passo à frente (resíduos) devem ser uma sequência de ruídos brancos, ou seja, não devem exibir autocorrelações (e autocorrelações parciais) significativas. Outros testes estatísticos executados nesta etapa envolvem verificações da condição de estacionariedade, análise de significância dos parâmetros estimados e critérios de seleção do modelo (AIC, BIC, EQM). (BOX et al., 2015).

Para validar os modelos ajustados aos dados estudados no presente trabalho foi aplicado o teste de Ljung-Box aos resíduos. O teste examina m autocorrelações dos resíduos. Em geral o teste é definido como:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{r}_k^2}{n-k}, \quad (4.6)$$

em que \hat{r}_k^2 é a autocorrelação estimada da série no lag k , e m é o número de lags que está sendo testado.

Como o teste é aplicado aos resíduos, os graus de liberdade devem considerar os parâmetros do modelo estimado de modo que $h = m - p - q$, onde p e q indicam o número de parâmetros do modelo ARMA(p,q) ajustado aos dados. (BOX et al., 2015)

Crítérios de seleção de modelos

Ao selecionarmos modelos estamos em busca do modelo que mais aproxima-se da realidade. Deste modo, tem que ser feita uma seleção do melhor modelo, dentre aqueles ajustados, para explicar tal fenômeno.

Critério de Informação de Akaike - AIC

[Akaike \(1974\)](#), mostrou que o viés é dado assintoticamente por p , em que p é o número de parâmetros a serem estimados no modelo, e definiu seu critério de informação como:

$$AIC = -2\log L(\hat{\theta}) + 2(p) \quad (4.7)$$

Critério de informação Bayesiano - BIC

o Critério de informação Bayesiano(BIC), proposto por [Schwarz et al. \(1978\)](#) é dado por:

$$BIC = -2\log f(x_n | \theta) + p\log n, \quad (4.8)$$

em que $f(x_n | \theta)$ é o modelo escolhido, p é o número de parâmetros a serem estimados e n é o número de observações da amostra.

Erro Quadrado Médio - EQM

[Schmittlein, Kim e Morrison \(1990\)](#) constroem um modelo que associa a minimização do EQM à variância dos erros das previsões e suas correlações, como forma de otimizar a seleção de combinações quando estes parâmetros têm que ser estimados, definido por:

$$EQM = \frac{1}{T} \sum_{t=1}^T u_{t+k,t}^2 \quad (4.9)$$

Previsão de valores

A última etapa a ser executada na modelagem em séries temporais, consiste em realizar a previsão. Ou seja, testa-se o potencial do modelo em prever os potenciais valores futuros para a série estudada. A previsão a ser realizada pode assumir dois aspectos. Num momento ela pode ser usada para prever valores futuros, que ainda não existem. Em outro momento, pode haver também a realização de previsões acerca dos valores já existentes dentro da série estudada. O modelo selecionado na etapa de identificação dos parâmetros for analisada, com vista à obtenção das previsões ([MORAES; SALES; COSTA, 2011](#)).

5 Resultados e Discussão

O presente capítulo tem por finalidade apresentar os resultados e discussão acerca da precipitação mensal (mm) para os municípios de Arcoverde, Ouricuri e Petrolina localizados no sertão do estado de Pernambuco, relativos ao período entre 1970 a 2017.

Na Figura 2 observa-se a precipitação em (mm) dos três municípios em estudo. As tendências de crescimento visualizadas em Arcoverde (Figura 2A) e Petrolina (Figura 2C), não necessariamente farão parte da composição da previsão para períodos próximos, pois se torna importante a análise das séries temporais observando a sucessão de períodos de acréscimo/decrécimo e tendência de valores próximos à média.

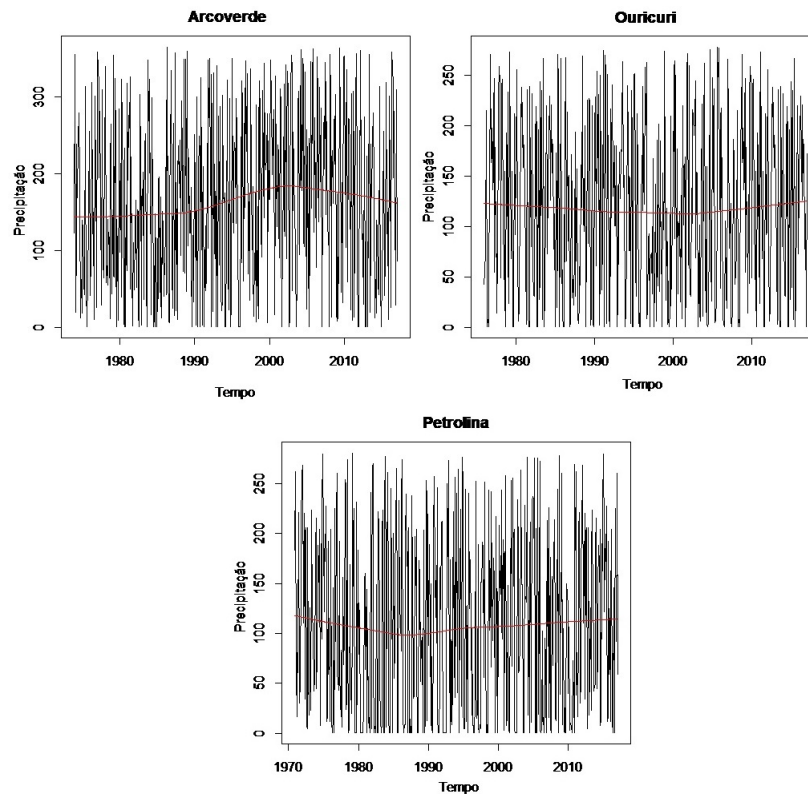


Figura 2 – Precipitação acumulada mensal (mm) no período de 1970 a 2017 para os municípios de Arcoverde (A), Ouricuri (B) e Petrolina (C).

Os dados deste estudo corroboram com o estudo de Paz et al. (2016) que ao analisarem a precipitação dos municípios de Recife-PE e Itambé-PE no período de 1963 a 2014, identificaram que entre os anos de 1993 e 2003, as cidades apresentaram menor precipitação que quando comparadas à média climatológica.

Silva et al. (2010) também identificaram em seu estudo a alta variação de precipitação no estado de Pernambuco, obtendo valores máximos nos anos de 1978, 1985 e 2000. E valores mínimos nos anos de 1992, 1998 e 1999.

Marengo e Alves (2005) estudaram a ocorrência de tendência hidrológica, usando longas séries temporais de dados de precipitação e vazão fluvial para várias localidades e concluíram que, embora tenham observado vazões com tendências negativas, não há como associar tal redução a uma possível diminuição da precipitação, mas sim à forma de gerenciamento das águas dos reservatórios da região (principalmente o reservatório de Paraibuna).

A Tabela 1 apresenta os resultados das análises de tendência das séries temporais das precipitações para as localidades de Arcoverde, Ouricuri e Petrolina utilizando-se o teste de Mann-Kendall com nível de significância adotado de 5%, bem como os resultados de estacionariedade das séries temporais obtidos pelo teste de Dickey-Fuller. A partir dos resultados do teste de Mann-Kendall pode-se verificar que a única localidade que apresenta tendência significativa é Arcoverde (p -valor $> 0,05$). Entretanto, pelo teste da raiz unitária de Dickey-Fuller, que verifica a hipótese da série temporal de precipitação ser não estacionária (H_0), observa-se que a hipótese foi rejeitada para todas as localidades. Sendo assim, admite-se que todas as séries temporais são estacionárias, possuem média e variância constantes ao longo do tempo.

Tabela 1 – Resultados obtidos com a aplicação dos respectivos testes, Dickey Fuller(estacionariedade) e Mann-Kendall(tendência), para as localidades de Arcoverde, Ouricuri e Petrolina.

Cidades	<i>Dickey – Fuller</i>	<i>p – valor</i>	<i>Mann – Kendall</i>	<i>p – valor</i>
Arcoverde	-7,225	0,010	0,067	0,022
Ouricuri	-10,180	0,010	-0,003	0,931
Petrolina	-8,491	0,010	0,002	0,938

Dando sequência as análises, segue-se com o ajuste dos modelos aos dados para os três municípios, os quais pode-se observar nas Tabelas 2, 3 e 4. Podendo concluir que os modelos ARMA (3,3) e SARMA (3,2)(1,1) e SARMA (2,3)(1,1), foram os que melhor se ajustaram aos dados de precipitação, sendo assim, foram utilizados na última etapa da metodologia aplicada no presente estudo, a previsão.

Tais composições tomam como base o critério de análise dos índices penalizadores Critério de Informação Akaike (AIC), Critério de Informação Bayesiana (BIC) e Erro Quadrático Médio (EQM), o (EQM) foi adotado como o critério de seleção nos modelos

selecionados, sendo assim, procuram os menores valores de tal índice para escolha do modelo.

Outra característica dos modelos ARMA(p,q) é a presença do intercepto ($\hat{\beta}_0$), que representa o deslocamento da origem do eixo Y, tanto para efeitos de tendência positiva quanto negativa, os valores do intercepto de cada município em estudo são os seguintes; Arcoverde (163,349), Ouricuri (116,918) e Petrolina (109,652). Observa-se, que, o município de Arcoverde (Tabela 2), apresentou um ARMA (3,3) com menor valor do critério de seleção EQM = 110,921. No município de Ouricuri (Tabela 3) apresentou um SARMA (3,2)(1,1) com menor valor do critério de seleção EQM = 85,4899. Para o município de Petrolina (Tabela 4), apresentou um SARMA (2,3)(1,1) com menor valor do critério de seleção EQM = 87,2683.

Uma aplicação de modelos do tipo ARIMA foi feita por [Gemitzi e Stefanopoulos \(2011\)](#) para investigar os efeitos das condições meteorológicas e das intervenções do homem sobre os aquífero subterrâneos. Os autores obtiveram bons resultados e sugerem o uso de modelos ARIMA como uma alternativa, especialmente devido à limitação na obtenção de dados em relação a propriedades hidráulicas.

Tabela 2 – Seleção dos modelos Auto-regressivos e de médias móveis ajustados para o município de Arcoverde, seguido dos respectivos critérios de avaliação.

Modelo	$\hat{\sigma}^2$	EQM	AIC	BIC
ARMA(1, 1)	12445	111,558	6374,72	6391,74
ARMA(1, 2)	12455	111,602	6377,00	6398,26
ARMA(3, 1)	12453	111,594	6378,92	6404,43
ARMA(2, 2)	12437	111,520	6378,25	6403,76
ARMA(3, 3)	12304	110,921	6376,83	6410,84

Tabela 3 – Seleção dos modelos Espacial Auto-regressivos e de médias móveis ajustados para o município de Ouricuri, seguido dos respectivos critérios de avaliação.

Modelo	$\hat{\sigma}^2$	EQM	AIC	BIC
ARMA(1, 1)	12445	111,558	6374,72	6391,74
ARMA(1, 2)	12455	111,602	6377,00	6398,26
ARMA(3, 1)	12453	111,594	6378,92	6404,43
ARMA(2, 2)	12437	111,520	6378,25	6403,76
SARMA(3, 2)(1, 1)	7309	85,489	5829,40	5867,00

Tabela 4 – Seleção dos modelos Espacial Auto-regressivos e de médias móveis ajustados para o município de Petrolina, seguido dos respectivos critérios de avaliação.

Modelo	$\hat{\sigma}^2$	EQM	AIC	BIC
ARMA(1, 1)	7943	89,12	6567,00	6584,28
ARMA(1, 2)	7947	89,14	6569,27	6590,86
ARMA(3, 1)	7877	88,75	6568,51	6598,74
SARMA(2, 3)(1, 1)	7616	87,27	6557,60	6596,50
SARMA(1, 1)(1, 1)	7681	87,64	6556,40	6582,30

Na tabela 5 são apresentadas as estimativas dos parâmetros ϕ , θ e α , mais o intercepto dos modelos selecionados para as cidades de ArcoVerde, Ouricuri e Petrolina.

Tabela 5 – Estimativas dos parâmetros dos modelos selecionados.

Cidades	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\lambda}_1$	$\hat{\rho}_1$	$\hat{\beta}_0$
Arcoverde	-0,54	0,86	0,64	0,55	-0,79	-0,66	-	-	163,35
Ouricuri	0,72	0,34	-0,11	-0,71	-0,28	-	0,71	-0,58	116,92
Petrolina	-0,52	0,34	-	0,73	-0,32	-0,09	0,98	-0,95	109,65

O modelo autorregressivo e de médias móveis de ordem (p,q) Arcoverde (3,3) é representado por:

$$X_t = -0,54X_{t-1} + 0,86X_{t-2} + 0,64X_{t-3} + a_t - 0,55a_{t-1} + 0,79a_{t-2} + 0,66a_{t-3}$$

Os modelos selecionados de Ouricuri SARMA(3,2)(1,1) e Petrolina SARMA(2,3)(1,1) são representados respectivamente por:

$$Y_t = 0,72Y_{t-S} + 0,34Y_{t-2S} - 0,11Y_{t-3S} + \epsilon_t - 0,71\epsilon_{t-S} - 0,28\epsilon_{t-2S},$$

$$Y_t = -0,52Y_{t-S} + 0,34Y_{t-2S} + \epsilon_t - 0,73\epsilon_{t-S} + 0,32\epsilon_{t-2S} - 0,09\epsilon_{t-3S}.$$

Observando os gráficos FAC's E FACP's, pode-se concluir que as séries são estacionárias, uma vez que seus valores tendem a zero com a FAC e a FACP sugerindo modelo da classe ARMA com seus respectivos parâmetros. No município de Arcoverde o modelo ARMA(3,3) foi o selecionado, em seguida o modelo SARMA(3,2)(1,1) e SARMA(2,3)(1,1) selecionados para as localidades de Ouricuri e Petrolina respectivamente.

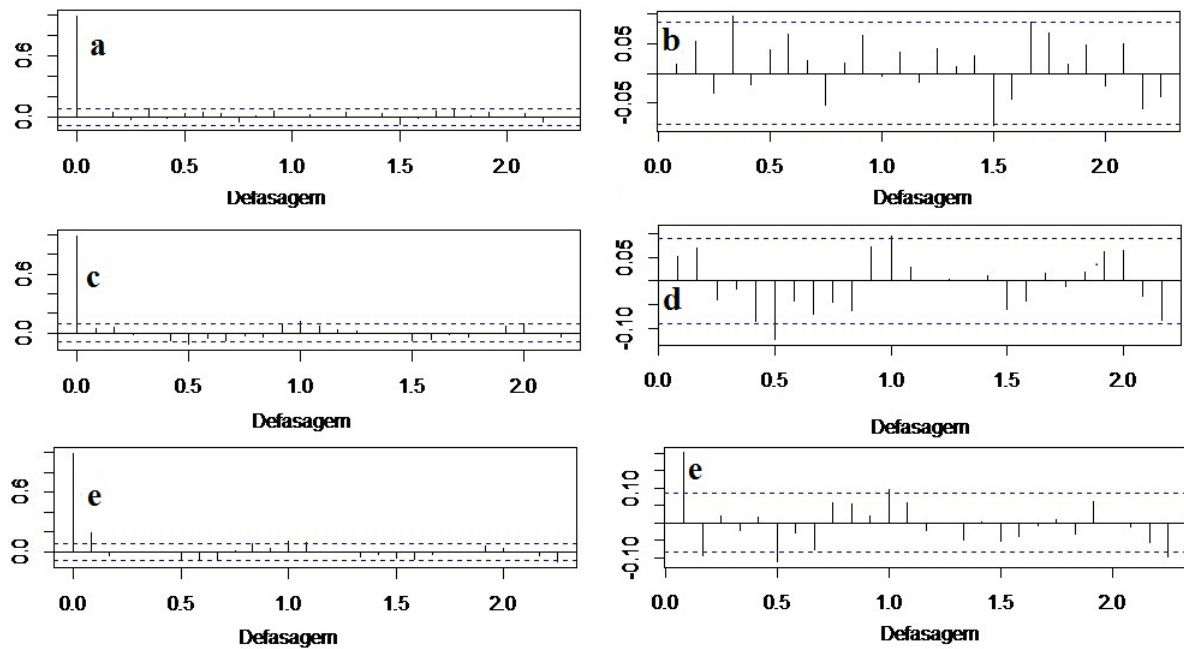


Figura 3 – Funções de autocorrelação (FAC)(a,c,e) e funções de autocorrelações parciais (FACP)(b,d,f), nos dados, para os municípios de Arcoverde, Ouricuri e Petrolina, respectivamente.

Para análise de diagnóstico do modelo, utilizou-se o teste de Ljung-Box, apresentados nas Figuras 4B, 4D e 4F para as localidades de Arcoverde, Ouricuri e Petrolina, respectivamente. O teste Ljung-Box verifica a falha do ajuste de um modelo de série temporal. O teste é aplicado aos resíduos de uma série temporal após o ajustamento de um modelo aos dados. O teste examina as autocorrelações dos resíduos. Se as autocorrelações forem muito pequenas, conclui-se que o modelo não exibe falha significativa de ajuste.

Analisando-se os resultados apresentados na Figura 4 percebe-se que os resíduos padronizados encontram-se entre os valores -3 e 3 . Em relação ao correlograma dos resíduos do modelo, pode-se observar que o mesmo não apresenta auto-correlação significativamente diferente de zero em nenhuma defasagem, indicando um bom ajuste do modelo e que o mesmo conseguiu captar a auto-correlação serial. Já em relação aos p -valores do teste de Ljung-Box apresentados graficamente para m diferentes defasagens, todos os valores estão próximos de 5%, indicando que não rejeita-se a hipótese nula de que as m primeiras auto-correlações dos erros são iguais a zero. Desta maneira, conclui-se que o modelo descreve adequadamente os dados, validando o seu uso para traçar previsões.

Diferente do que [Chechi e Sanches \(2013\)](#) encontraram em seu estudo sobre uma série temporal de precipitação de Erechim (RS) no período de 2003 a 2011, em que após

a realização do teste de Ljung-Box observou-se que nenhum valor ficou acima da linha tracejada, indicando a não rejeição da hipótese nula de não estacionariedade. O atual estudo aponta que todos os valores ficaram acima da linha tracejada, indicando que os modelos selecionados são os que mais representa os dados em estudo.

Box et al. (2015) utilizam o teste de Ljung-Box aplicado aos resíduos para validar os modelos dos dados selecionados no estudo em questão, o mesmo mostra que tal procedimento é eficaz na seleção do modelo que mais aproxima-se da realidade dos dados, uma vez que não existe modelo perfeito para cada caso, e sim uma aproximação de cada situação em estudo.

Werner e Ribeiro (2003) no trabalho em questão, usa o teste de Ljung-Box para reforça a escolha dos modelos que melhor representa os dados em estudo. Esse teste compara o valor da estatística de teste com os valores tabelados da distribuição Qui-quadrado. Em seu estudo, pela análise do teste Ljung-Box, os resíduos podem ser considerados ruído branco, isto é, suas autocorrelações devem ser não significantes.

As Figuras 4A, 4C e 4E, apresentam a Função de Auto-correlação residual dos modelos ajustados aos dados de Precipitação das localidades de Arcoverde, Ouricuri e Petrolina, respectivamente. Estes resultados evidenciam o comportamento de um ruído branco, pois os coeficientes de auto-correlação ficaram dentro dos limites de confiabilidade.

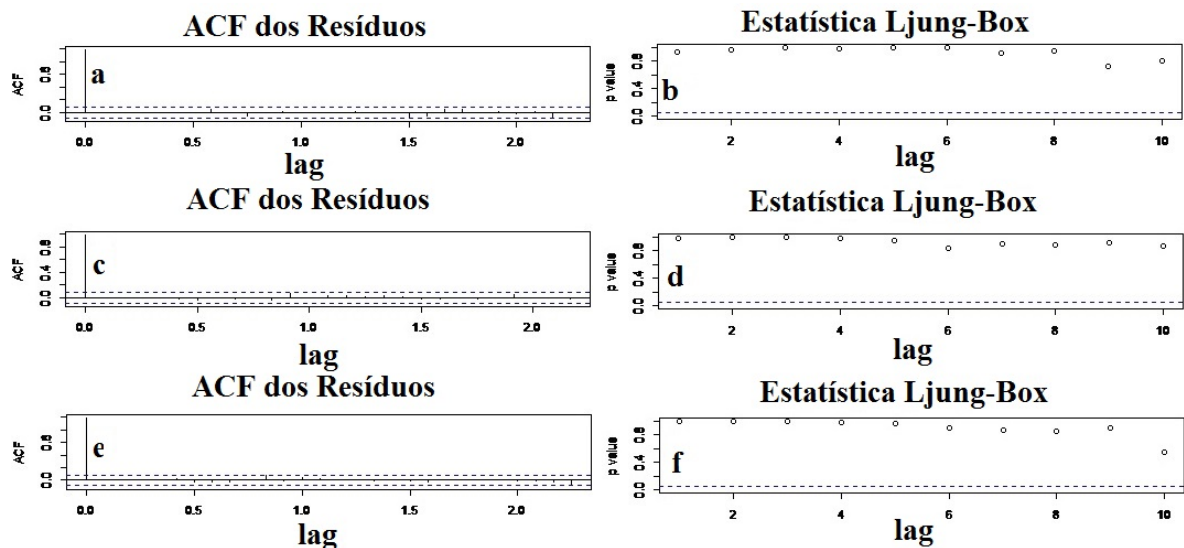


Figura 4 – Funções de autocorrelação dos resíduos(FAC)(A,C,E) e Estatística de Ljung-Box(B,D,F) para os municípios de Arcoverde, Ouricuri e Petrolina.

Partindo-se da interpretação da Figura 4, em relação aos p-valores do teste de Ljung-Box apresentados graficamente para m diferentes defasagens, todos os valores estão

acima de 5%, indicando que não se rejeita a hipótese nula de que as m primeiras autocorrelações dos erros são iguais a zero. Desta maneira, conclui-se que os modelos descrevem adequadamente os dados, validando o seu uso para traçar previsões.

A previsão é uma primeira e óbvia aplicação da análise de séries temporais. De fato, essa análise propicia a previsão dos valores futuros da série. Contudo, a incerteza sobre estimativas futuras é tanto maior quanto mais se distancia do presente, e intercorrências inesperadas são mais ou menos frequentes nas atividades humanas, o que implica risco adicional de erro não previsto.

Analisando-se a Figura 5 é possível afirmar que a média da precipitação acumulada mensal(mm) para o próximo ano, apresenta-se em torno da média dos dados da série em questão, nos três municípios em estudo, este fato era esperado, uma vez que as três séries em estudo são estacionaria, logo sua média e variância são constante ao longo do tempo.

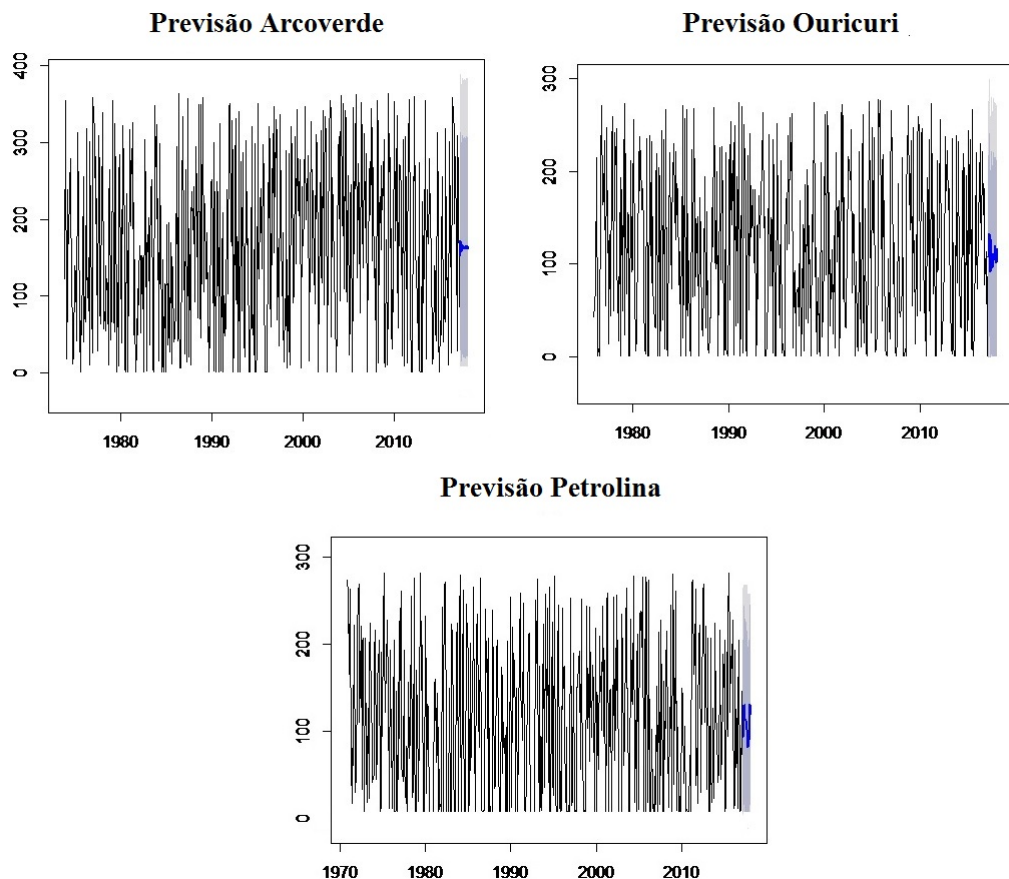


Figura 5 – Previsão da precipitação acumulada mensal(mm) para os próximo ano, nos municípios de Arcoverde, Ouricuri e Petrolina.

6 Conclusão

O presente trabalho teve o intuito de aplicar a metodologia de Box & Jenkins fundamentada em quatro passos, Identificação, Estimação, Verificação do modelo e Previsão, no qual foram analisados dados de chuva de 3 estações pluviométricas pertencentes à uma estação meteorológica digital, Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP), relativos ao período de 1970 a 2017. Na avaliação da precipitação foram considerados os totais mensais das cidades ArcoVerde, Ouricuri e Petrolina localizadas na região do sertão do estado de Pernambuco.

Quanto a escolha do melhores modelos selecionados é possível concluir que os modelos, ARMA (3,3), SARIMA (3,2)(1,1) e SARIMA(2,3)(1,1), são os modelos que melhor representa os dados das cidades Arcoverde, Ouricuri e Petrolina respectivamente. Esses modelos se mostraram satisfatórios para análise e previsão dos dados de precipitação, contornando o problema de estacionariedade e tendência, ajudando na leitura a respeito da precipitação mensal dos próximos mês.

Referências Bibliográficas

- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, Ieee, v. 19, n. 6, p. 716–723, 1974.
- ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. **Epidemiologia e Serviços de Saúde**, SciELO Public Health, v. 24, p. 565–576, 2015.
- ARÊDES, A. d.; PEREIRA, M. W. G. Potencialidade da utilização de modelos de séries temporais na previsão do preço do trigo no Estado do Paraná. **Revista. de Economia Agrícola, São Paulo**, v. 55, n. 1, p. 63–76, 2008.
- BAYER, F. M. Wavelets e modelos tradicionais de previsão: Um estudo comparativo. **Revista Brasileira Biometria**, v. 28, n. 2, p. 40–61, 2010.
- BOX, G. G. m. jenkins, 1994. **Time Series Analysis: Forecasting and Control**, 1994.
- BOX, G. E. et al. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BOX, J.; JENKINS, G. M. Reinsel. time series analysis, forecasting and control. **Tercera. NJ: Prentice Hall, Englewood Cliffs, NJ, USA**, 1994.
- CHECHI, L.; SANCHES, F. de O. Análise de uma série temporal de precipitação para erechim (rs) e um possível método de previsão climática analysis of a series of precipitation for erechim (rs) and a method of possible climate prediction. **AMBIÊNCIA**, v. 9, n. 1, p. 43–55, 2013.
- DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica: Journal of the Econometric Society**, JSTOR, p. 1057–1072, 1981.
- ESPINOSA, M. M.; PRADO, S. M.; GHELLERE, M. Uso do modelo sarima na previsão do número de focos de calor para os meses de junho a outubro no estado de mato grosso. **Ciência e Natura**, v. 32, n. 2, p. 07–21, 2010.
- FAVA, V. L. Metodologia de Box-Jenkins para modelos univariados. **VASCONCELLOS, MAS; ALVES, D. Manual de Econometria. São Paulo: Ed. Atlas**, p. 205–231, 2000.
- FERRAZ, M. I. F. **Uso de modelos de séries temporais na previsão da série de precipitações pluviiais mensais no município de Larvas-MG. 1999. 97 p.** Tese (Doutorado) — Dissertação (Mestrado em Agronomia). Universidade Federal de Lavras, Lavras, 1999.

- FIGUEREDO, C. J. Previsão de séries temporais utilizando a metodologia Box e Jenkins e redes neurais para inicialização de planejamento e controle de produção. 2008.
- FISCH, G.; VALÉRIO, M. C. Variabilidade intra e interanual da precipitação em taubaté-sp associado aos eventos el niño e la niña. **Revista Biociências**, v. 11, 2008.
- FREITAS, R. A. P. de et al. Análise de séries temporais de vazão e precipitação na bacia hidrográfica do rio são mateus. 2010.
- GEMITZI, A.; STEFANOPOULOS, K. Evaluation of the effects of climate and man intervention on ground waters and their dependent ecosystems using time series analysis. **Journal of hydrology**, Elsevier, v. 403, n. 1-2, p. 130–140, 2011.
- KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? **Journal of econometrics**, Elsevier, v. 54, n. 1-3, p. 159–178, 1992.
- MARENGO, J. A.; ALVES, L. M. Tendências hidrológicas da bacia do rio paraíba do sul. **Revista Brasileira de Meteorologia**, v. 20, n. 2, p. 215–226, 2005.
- MARÔCO, J. **Análise estatística com o SPSS Statistics**. [S.l.]: ReportNumber, Lda, 2011.
- MORAES, R. F. de; SALES, P. R. de H.; COSTA, J. F. S. Utilização do software estatístico r na previsão de séries temporais: série prêmio de resseguro. **Cadernos do IME-Série Estatística**, v. 30, n. 1, p. 1, 2011.
- MORETTIN, P. A.; TOLOI, C. M. de castro. **Análise de Séries Temporais**. São Paulo, ABE, 2004.
- PADILHA, J. B. J. et al. Modelagem box-jenkins (arima) para gestão de risco de preço na pecuária de corte do estado do paran . In: **X Congresso Virtual Brasileiro de Administra o–CONVIBRA**. [S.l.: s.n.], 2013. p. 12.
- PAZ, Y. M. et al. Temporal variability of rainfall and environmental vulnerability of the northern part of the zona da mata and metropolitan region of pernambuco. **Journal of Hyperspectral Remote Sensing**, v. 6, n. 1, p. 22–35, 2016.
- PINHEIRO, A.; GRACIANO, R. L. G.; SEVERO, D. L. An lise de tend ncia das s ries temporais de precipita o da regi o sul do brasil. **Revista Brasileira de Meteorologia**, SciELO Brasil, v. 28, n. 3, 2013.
- RCORETEAM. **R: A language and environment for statistical computing**. **R Foundation for Statistical Computing, Vienna, Austria**. 2018. [S.l.]: ISBN 3-900051-07-0, 2018.
- SATO, R. C. Disease management with arima model in time series. **Einstein (Sao Paulo)**, SciELO Brasil, v. 11, n. 1, p. 128–131, 2013.
- SCHMITTLEIN, D. C.; KIM, J.; MORRISON, D. G. Combining forecasts: Operational adjustments to theoretically optimal rules. **Management Science**, INFORMS, v. 36, n. 9, p. 1044–1056, 1990.

SCHWARZ, G. et al. Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

SILVA, D. F. D.; KAYANO, M. T.; SOUSA, F. d. A. S. de. Escalas temporais da variabilidade pluviométrica da bacia hidrográfica do rio mundaú (AL e PE). **Revista Brasileira de Meteorologia**, SciELO Brasil, v. 25, n. 3, 2010.

SILVA, R. Marques da et al. Análise da variabilidade espaço-temporal e identificação do padrão da precipitação na bacia do rio tapacurá, pernambuco. **Sociedade & Natureza**, Universidade Federal de Uberlândia, v. 22, n. 2, 2010.

SNEYERS, R. **Sur l'analyse statistique des séries d'observations**. [S.l.]: Secrétariat de l'Organisation Météorologique Mondiale, 1975.

SOLOS, E. Unidade de execução de pesquisa e desenvolvimento do recife. **Zoneamento agroecológico: Pernambuco crescendo por inteiro**. Recife: Embrapa Solos-UEP Recife: Governo do Estado de Pernambuco, Secretaria de Produção Rural e Reforma Agrária, v. 1, 2001.

SOUZA, V. et al. Utilização da metodologia de box & jenkins na previsão do preço futuro pago as exportações paranaenses de madeira serrada. **Universidade Federal do Paraná**, 2006.

THOM, H. Phenomenological analysis of $k + \lambda$ photoproduction. **Physical Review**, APS, v. 151, n. 4, p. 1322, 1966.

WERNER, L.; RIBEIRO, J. L. D. Previsão de demanda: uma aplicação dos modelos box-jenkins na área de assistência técnica de computadores pessoais. **Gestão e produção**. São Carlos, SP. Vol. 10, no. 1 (abr. 2003), p. 47-67, SciELO Brasil, 2003.