



Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Fábio Sandro dos Santos

Modelos de Séries Temporais Aplicados a Dados de Precipitação da Região do Cariri Paraibano

Campina Grande - PB

Junho de 2015

Fábio Sandro dos Santos

Modelos de Séries Temporais Aplicados a Dados de Precipitação da Região do Cariri Paraibano

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Orientador:

Dr. Ricardo Alves de Olinda

Campina Grande - PB

Junho de 2015

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

S237m Santos, Fábio Sandro dos.
Modelos de séries temporais aplicados a dados de precipitação da região do Cariri Paraibano [manuscrito] / Fabio Sandro dos Santos. - 2015.
42 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2015.

"Orientação: Prof. Dr. Ricardo Alves de Olinda, Departamento de Estatística".

1. Variáveis climáticas. 2. Modelos ARIMA. 3. Sazonalidade. I. Título.

21. ed. CDD 519.5

Fábio Sandro dos Santos

Modelos de Séries Temporais Aplicados a Dados de Precipitação da Região do Cariri Paraibano

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Aprovado em: 30/06/2015

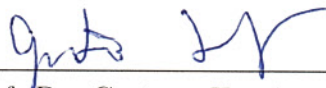
Banca Examinadora:



Prof. Dr. Ricardo Alves de Olinda
Orientador



Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba - UEPB



Prof. Dr. Gustavo Henrique Esteves
Universidade Estadual da Paraíba - UEPB

Dedicatória

Dedico este trabalho aos meus pais João e Joana. Aos meus irmãos Sandra, Alex e Simone.

Agradecimentos

Escrever os agradecimentos de um trabalho acadêmico levou-me a comparar com momentos de tarefas difíceis e árduas. Surge um misto de lembranças, esquecimentos, entrelaçados com julgamentos de injustiças, para com alguns. E mesmo diante dos riscos de ser injusto, represento no decorrer desta seção, meu reconhecimento a personagens fundamentais que cruzaram algum dia meu caminho e contribuíram de alguma forma para esta realização. Venho aqui agradecer:

A todos e a todas que compõem o curso de Bacharelado em Estatística da Universidade Estadual da Paraíba. Em especial, o meu orientador Ricardo Alves de Olinda, que desde os primeiros contatos depositou em mim suas palavras de conforto e confiança, muito obrigado pela atenção, paciência, conselhos, dicas de trabalho e pesquisa.

Aos professores João Gil de Luna, Gustavo Henrique Esteve e Tiago Almeida de Oliveira, com sensibilidade, conseguiram instigar minha curiosidade pela Estatística.

Resumo

A Região do Cariri paraibano fica situada no Nordeste brasileiro, apresentando acentuada variabilidade em relação à precipitação pluviométrica. Esta Região apresenta alguns anos extremamente secos e poucos anos chuvosos. O período chuvoso concentra-se entre os meses de janeiro a maio podendo se estender, em alguns anos, até o mês de julho. Nos demais períodos ocorrem irregularidades de chuvas e baixos índices pluviométricos. Sendo assim, se faz necessário o uso de modelos de séries temporais para a variável precipitação. Os dados constituíram-se de uma série temporal compreendida entre os anos de 1935 a 2013, nos municípios de Soledade, Taperoá, Santa Luzia e Cabaceira, Estado da Paraíba. Para o estudo dessas séries foram utilizados os testes da raiz unitária (Dickey-Fuller) e o teste de tendência (Mann-Kendall). Na sequência, ajustou-se aos dados os modelos autorregressivos integrados de médias móveis (ARIMA), adotando-se, como critério de escolha, o valor de AIC e BIC. Neste estudo foi possível identificar no modelo autorregressivo a presença de não-estacionariedade e ausência de tendência. Os gráficos de previsão, obtidos a partir do modelo ajustado, permitiram uma interpretação coerente do comportamento da variável precipitação para a próxima década nos municípios analisados.

Palavras-chave: Variáveis climáticas, modelos ARIMA, sazonalidade.

Abstract

The Region of Cariri is located in Brazilian northeastern, showing marked variability in relation to precipitation of rainfall. This region has some extremely dry years and some rainy years. The rainy season is concentrated among months from January to May and some years going to until month of July. In other periods feature irregularities of rains and low rates of rainfall. Thus, it is necessary to use time series models for rainfall variable . The data consisted of a time series between the years 1935 to 2013 in the municipalities of Soledade, Taperoá, Santa Luzia and Cabaceira, State of Paraíba . For the study of these series the unit root tests were used (Dickey-Fuller) and trend test (Mann-Kendall). then, seted autoregressive models Integrated moving-average(ARIMA), adopting as a criterion of choice, value of AIC and BIC. In this study it was possible indetificar on the models autoregressive the presence of non-stationary and absence trend. The graphics forecast, obtained from the model adjusted allowed a interpretation of the behavior rainfall variable for the next decade in the municipalities analyzed.

key-words: Climatic variables, ARIMA models, Seasonality

Sumário

Lista de Tabelas

Lista de Figuras

1	Introdução	p. 10
2	Material e Métodos	p. 12
2.1	Processos Estocásticos	p. 12
2.2	Estacionariedade	p. 13
2.3	Tendência e Sazonalidade	p. 15
2.4	Modelos para Séries Temporais	p. 18
2.4.1	Modelos Auto-Regressivo - AR (p)	p. 19
2.4.2	Modelos de Médias Móveis - MA (q)	p. 19
2.4.3	Modelos Auto-Regressivo e de médias Móveis ARMA (p, q)	p. 20
2.4.3.1	Modelo Auto-Regressivo Integrado e de Médias Móveis ARIMA (p, d, q)	p. 21
2.5	Previsão com Modelos ARIMA	p. 26
3	Resultados e Discussão	p. 27
4	Conclusão	p. 37
	Referências	p. 38
	Apêndice	p. 40

Lista de Tabelas

1	Resultados obtidos com a aplicação dos respectivos testes, Dickey-Fuller e Mann-Kendall para as localidades de Soledade, Taperoá Santa Luzia e Cabaceira.	p. 29
2	Critérios de seleção de modelos autorregressivos integrados e de médias móveis para o município de Soledade	p. 30
3	Critérios de seleção de modelos autorregressivos integrados e de médias móveis para o município de Taperoá	p. 31
4	Critérios de seleção de modelos autorregressivos e de médias móveis para o município de Santa Luzia	p. 32
5	Critérios de seleção de modelos autorregressivos integrados e de médias móveis para o município de Cabaceira	p. 33

Lista de Figuras

1	Processo estocástico interpretado como uma família de variáveis aleatórias (Fonte: Bezerra, (2006).	p. 13
2	Série não-estacionária quanto ao nível e inclinação (Fonte: Bezerra, (2006).	p. 14
3	Precipitação acumulada anual(mm) no período de 1935 a 2013 para os municípios de Soledade (a), Taperoá (b), Santa Luzia (c) e Cabaceira(d), localizados na região do Cariri paraibano.	p. 28
4	Funções de autocorrelações (FAC) e funções autocorrelações parciais (FACP) para os municípios de Soledade (a, b), Taperoá (c, d), Santa Luzia (e, f) e Cabaceira (g, h).	p. 29
5	Gráficos dos resíduos e Gráficos de Box-Pierce para os municípios de Soledade, Taperoá, Santa Luzia e Cabaceira, respectivamente.	p. 34
6	Previsão da Precipitação acumulada anual(mm) para os próximos dez anos nos municípios de Soledade (a), Taperoá (b), Santa Luzia (c) e Cabaceira (d)	p. 35

1 Introdução

De acordo com Alves (2007), o domínio ecogeográfico da Caatinga ocupa uma área de cerca de 750.000km^2 sob as latitudes sub-equatorial compreendidas entre $2^\circ 45'$ e $17^\circ 21'$, Latitude Sul. Engloba partes dos territórios pertencentes aos estados do Maranhão, Piauí, Ceará, Rio Grande do Norte, Pernambuco, Paraíba, Alagoas, Sergipe, Bahia, e parte de Minas Gerais. Sua área corresponde a 50% da Região Nordeste e a 11% do território brasileiro e constitui o chamado Polígono das Secas.

A região do Cariri paraibano têm seus municípios concentrados predominantemente em uma área localizada no centro do espaço geográfico da Paraíba. O clima o qual a região do Cariri paraibano está submetida, varia de semi-áridos a sub-áridos secos tropicais de exceção e são caracterizados por uma pluviometria que se concentra em um só período (3 a 4 meses), com médias anuais situadas entre 250 a 900 mm, irregularmente distribuídas no tempo e no espaço. As temperaturas médias anuais são relativamente elevadas, 25°C a 27°C , e a insolação média é de 2.800 horas/ano. A umidade relativa do ar é de cerca de 50% e as taxas médias de evaporação são em torno de 2.000 mm/ano (VIEIRA *et al.*, 2007).

Para Nascimento e Alves (2008), na Paraíba, os Cariris formam uma diagonal Nordeste-Sudoeste, onde a mesma é denominada diagonal seca. Ao ponto de vista climatológico se caracterizam por dois aspectos marcantes como, a extrema e generalizada fraqueza das precipitações que são marcadas pela situação-posição geográfica, exposição solar e altitude (como no sudoeste da região) e pela forte irregularidade do regime das chuvas que se manifesta por uma grande viabilidade inter-anual (ela pode alcançar 50% de um ano para outro) e anual (ao nível das datas de início e do final das estações ou ao nível das chuvas).

Levando-se em consideração as irregularidades no regime pluviométrico do Estado da Paraíba e a preocupação da sociedade no que se refere aos recursos hídricos desse Estado, se faz necessário o uso de modelos de séries temporais para que se possa quantificar e prever o comportamento da variável precipitação para os próximos anos. Com os avanços computacionais, modelos para análise de dados em séries temporais evoluíram significativamente, facilitando assim o uso em diversas áreas de aplicação.

Diante do exposto este trabalho tem como objetivos: verificar o comportamento de tendência e estacionariedade da precipitação nos municípios de Soledade, Taperoá, Santa Luzia e Cabaceira situadas na região do Cariri paraibano; ajustar modelos de séries temporais à variável precipitação, adotando-se alguns critérios de seleção de modelos; identificar qual modelo melhor se ajusta aos dados em estudo e por fim; realizar previsões na variável precipitação para próximos anos nos municípios em estudo.

2 Material e Métodos

Neste capítulo serão evidenciados as metodologias que servirão de base para este trabalho. Os dados utilizados para realização deste trabalho foram obtidos na Agência Executiva de Gestão das Águas do Estado da Paraíba (AESA), relativos ao período de 1935 a 2013, observados no município de Soledade, Taperoá, Santa Luzia e Cabaceira localizados na região do Cariri paraibano.

2.1 Processos Estocásticos

Segundo Bezerra (2006), um processo estocástico é uma família $Z = Z_t$, tal que, para cada $t \in T$, Z_t é uma variável aleatória (v.a.) definida num espaço de probabilidades. Em matemática, um espaço de probabilidade é uma tripla $(\Omega, \mathbb{A}, \mathbb{P})$ formada por um conjunto ômega Ω , uma σ -álgebra \mathbb{A} em Ω e uma medida positiva \mathbb{P} nessa σ -álgebra tal que $P(\Omega) = 1$. O conjunto Ω é chamado de espaço amostral e os elementos de \mathbb{A} são chamados os eventos.

Seja o conjunto T arbitrário, sendo normalmente tomado como o conjunto dos inteiros $\mathbb{Z} = 0, \pm 1, \pm 2, \dots$ ou no conjunto dos reais \mathbb{R} . Como, para $t \in T$, Z_t é uma v.a. na realidade Z_t é uma função de dois argumentos, $Z(t, \omega)$, $t \in T$, $\omega \in \Omega$, conforme descrito na Figura 1.

O conjunto dos valores $Z_t, t \in T$ é chamado espaço dos estados do processo estocástico, e os valores de Z_t são chamados estados, onde Z_t e T podem ser discretos ou contínuos. Um processo estocástico $Z = \{Z_{(t)}, t \in T\}$ diz-se fracamente estacionário ou estacionário de segunda ordem se, e somente se:

$$i) E(Z_t) = \mu_{(t)} = \mu, \text{ é constante para todo } t \in T;$$

$$ii) E_t^2 < \infty, \text{ para todo } t \in T;$$

$$iii) \gamma_j = cov[Z_t, Z_{t-j}] = E[(Z_t - \mu)], \text{ só depende dos passos } (j) \forall t \in T \text{ e } j \in \mathbb{R}.$$

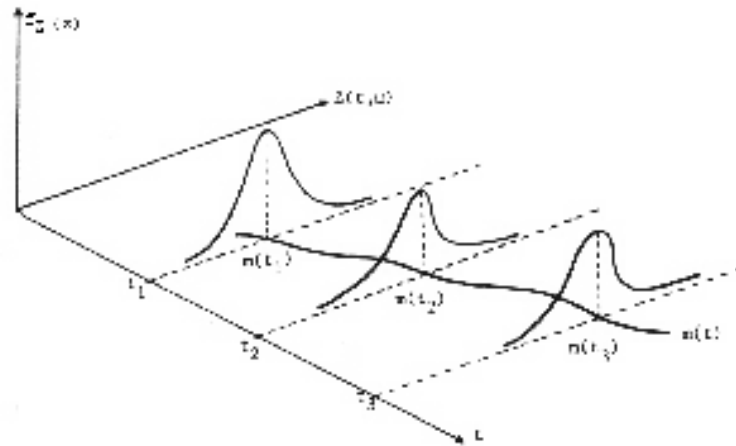


Figura 1: Processo estocástico interpretado como uma família de variáveis aleatórias (Fonte: Bezerra, (2006)).

Conforme Morettin e Tolo (2006), a função definida em (iii) é chamada de função de autocovariância (FACV).

2.2 Estacionariedade

Segundo Morettin e Tolo (2004), uma série temporal é estacionária quando ela se desenvolve aleatoriamente, no tempo, em torno de uma média constante, refletindo alguma forma de equilíbrio estável. Entretanto, a maior parte das séries que encontra-se na prática apresentam alguma forma de não estacionariedade. As séries econômicas, por exemplo, apresentam em geral tendências lineares positivas ou negativas. Pode-se ter, também, uma forma de não-estacionariedade explosiva, como o crescimento de uma colônia de bactérias.

Com o ajuste de alguns modelos é possível descrever de maneira satisfatória séries estacionárias e não-estacionárias, mas que não apresentam comportamento explosivo. A série pode ser estacionária, flutuando ao redor de um nível, por um certo tempo, depois mudar de nível e flutuar ao redor de um novo nível e assim por diante, ou então mudar de inclinação, ou ambas as coisas. A Figura 2 ilustra esta forma de não-estacionariedade. Como a maioria dos procedimentos de análise estatística de séries temporais supõem que estas sejam estacionárias, deve-se transformar os dados originais, se estes não formam uma série estacionária.

A transformação mais comum consiste em tomar diferenças sucessivas da série original,

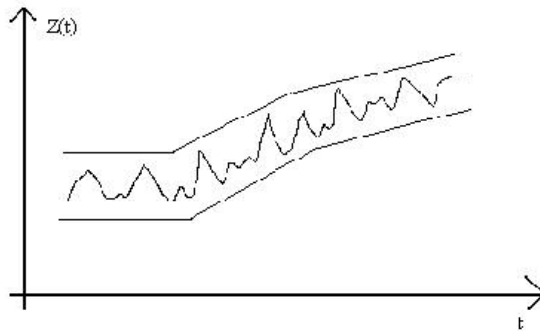


Figura 2: Série não-estacionária quanto ao nível e inclinação (Fonte: Bezerra, (2006)).

até se obter uma série estacionária. A primeira diferença de Z_t é definida pela equação de um passo.

$$\Delta Z_t = Z_t - Z_{t-1}, \quad (2.1)$$

a segunda diferença é denotada pela equação

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

De modo geral, a n -ésima diferença de Z_t é $\Delta^n Z_t = \Delta[\Delta^n - Z_t]$. Em situações normais, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária.

Segundo Morettin e Tolo (2006), uma importante classe de processos estocásticos são chamados de processos estacionários. Uma série temporal é dita estritamente estacionária se a distribuição conjunta de probabilidade Y_{t_1}, \dots, Y_{t_k} é a mesma de $Y_{t_1 + \tau}, \dots, Y_{t_k + \tau}$. Sendo τ a defasagem, ou seja, o deslocamento da origem dos tempos por uma quantidade τ não tem efeito na distribuição conjunta que, portanto depende apenas dos intervalos entre t_1, \dots, t_k . Em particular, para $k=1$ a estacionariedade escrita implica que a distribuição de Y_t é a mesma para todo t de modo que, se os dois primeiros momentos forem finitos, ou seja,

$$\mu_t = \mu \text{ e } \sigma_t^2 = \sigma^2,$$

em que μ e σ^2 são constantes que não dependem de t .

Segundo FAVA (2000), na análise das séries estacionárias é possível separar a estaci-

onariedade em dois grupos:

- i*) Estacionariedade no amplo senso: médias, variâncias e covariâncias constante no tempo;
- ii*) Estacionariedade no estrito senso: probabilidade de uma dada flutuação no processo em torno da média é a mesma em qualquer momento do processo.

Teste de Estacionariedade

Morettin e Tolo (2004) afirmam que as suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela seja estacionária, ou seja, ela se desenvolva no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. A suposição de estacionariedade pode ser verificada da seguinte forma.

Ao realizar um estudo em séries temporais é necessário verificar o pressuposto de estacionariedade, para isto, pode-se aplicar o teste de Dickey-Fuller. Este teste procede tabulando os valores críticos através da simulação de Monte Carlo, desenvolver a estatística de teste τ (*tau*) para testar formalmente o problema. Se o valor absoluto da estatística τ calculado for maior que o valor absoluto tabelado, aceita-se a hipótese nula, logo a série é não-estacionária.

Grande parte dos recursos para séries temporais foram elaborados utilizando o conceito de estacionariedade nas séries. Uma forma geral para analisar este fato é fazendo um estudo da existência de alguma raiz dos operadores de retardos dentro do círculo unitário, denominada simplesmente por raiz unitária. Portanto, utilizamos testes de hipóteses que em geral, possui as seguintes hipóteses:

H_0 : “Existe pelo menos uma raiz dentro do círculo unitário”

H_1 : “Não existem raízes dentro do círculo unitário”

2.3 Tendência e Sazonalidade

Considerando-se as observações $Z_t = 1, \dots, N$ de uma série temporal, em que um modelo de decomposição consiste em escrever Z_t como uma soma de três componentes não-observáveis, isto é,

$$Z_t = T_t + S_t + a_t,$$

em que T_t e S_t representam a tendência e sazonalidade, respectivamente, enquanto a_t é um componente aleatório, de média zero e variância constante σ_a^2 . Se a_t for um ruído branco, então $E(a_t, a_s) = 0, s \neq t$; no entanto pode-se eventualmente, relaxar esta suposição, adotando-se a_t como um processo estacionário.

O interesse principal é considerar um modelo do tipo que será o de estimador de \hat{S}_t e construir as séries livres de sazonalidade ou sazonalmente ajustadas. Isto é, se \hat{S}_t for uma estimativa de S .

$$Z_t^S = Z_t - \hat{S}_t$$

é a série sazonalmente ajustada. Há várias razões para considerar este procedimento de ajustamento sazonal. Conforme Kopecky e Pierce (1979) as componentes T_t e S_t são, em geral, bastante relacionadas e a influência de tendência sobre a componente sazonal pode ser muito forte, por duas razões:

- i) Métodos de estimação de \hat{S}_t podem ser bastante afetados se não levar em consideração a tendência;
- ii) A especificação de S_t depende da especificação de T_t .

Devido a estes fatos, não é possível isolar uma das componentes sem tentar isolar a outra. Estimando-se T_t e S_t e subtraindo-se de Z_t obtém-se uma estimativa da componente aleatória a_t .

Tendência

Conforme Morettin e Tolo (2006), há vários métodos para estimar a tendência de uma série temporal. Os mais utilizados consistem em

- i) Ajustar uma função tempo, como um polinômio, uma exponencial ou outra função suave de t ;
- ii) Suavizar (ou filtrar) os valores da série ao redor de um ponto, para estimar a tendência naquele ponto;
- iii) Suavizar os valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados.

Estimando-se a tendência através de \hat{T}_t , pode-se obter a série ajustada para tendência ou livre de tendência.

Teste para Tendência

Um teste utilizado para verificar possíveis tendências numa sérietemporal é o teste de Mann-Kendall, o qual foi proposto por Sneyers (1975), considerando-se, as hipótese de estabilidade de uma série temporal, a sucessão de valores ocorre de forma independente, e a distribuição de probabilidade deve permanecer sempre a mesma. Goossens e Berger () afirmam que o teste de Mann-Kendall é o método mais apropriado para analisar mudanças climáticas, além de permitir a detecção e localização aproximada do ponto inicial de determinadas tendências.

Seja as observações y_1, y_2, \dots, y_n de uma série temporal. Podemos aplicar o teste de Mann-Kendall para tendência somente se a série for serialmente independente. Então, tem-se interesse em testar se as observações da série são independentes e identicamente distribuídas, isto é, quer-se testar as hipóteses

H_0 : “As observações da série são independentes e identicamente distribuídas”

H_1 : “As observações da série possuem tendência monótona no tempo”

Sazonalidade

Segundo Morettin e Tolo (2004), a sazonalidade numa série temporal é ajustada à uma série para o componente sazonal, ou seja, estima-se a série temporal e subtrai-se Z_t no modelo. Um procedimento de ajuste sazonal consiste em:

i) obter estimativas \hat{S}_t de S_t ;

ii) calcular:

$$Z_t^S = Z_t - \hat{S}_t.$$

Se o modelo for multiplicativo, da forma

$$Z_t = T_t \times S_t a_t,$$

a série sazonalmente ajustada será

$$Z_t^S = Z_t / \hat{S}_t.$$

O modelo geralmente é adequado para séries econômicas, que apresentam um crescimento exponencial. Estimando-se \hat{S}_t comete-se um erro de ajustamento sazonal, dado

por

$$a_t = S_t - \hat{S}_t.$$

O procedimento de ajustamento sazonal é ótimo se minimizar $E(a_t^2)$. Sem perda de generalidade, considere o caso que tem-se dados mensais e o número total de observações, N , é múltiplo de 12, isto é, $N = 12p$, p é o número de anos, de modo que os dados podem ser representados como

$$Z_{ij} = T_{ij} + S_j + a_{ij}, \quad i = 1, \dots, p; \quad j = 1, \dots, 12. \quad (2.2)$$

No modelo (2.5) a sazonalidade é considerada constante. Para sazonalidade não-constante, ou seja, o padrão sazonal muda de ano para ano, devendo-se considerar o modelo

$$Z_{ij} = T_{ij} + S_{ij} + a_{ij}, \quad i = 1, \dots, p; \quad j = 1, \dots, 12, \quad (2.3)$$

segundo-se

$$\bar{Z}_i = \frac{1}{12p} \sum_{j=1}^p \sum_{j=1}^{12} Z_{ij}$$

2.4 Modelos para Séries Temporais

A metodologia das séries temporais apresentam uma série de modelos, os quais podem ser classificados de acordo com seus respectivos parâmetros. Estes podem ser classificados como sendo paramétricos, pois se caracterizam pelo número finito de parâmetros, e tem suas análises feitas no domínio do tempo. Neste estudo, serão descritos apenas os modelos mais comuns.

Segundo Fischer (1982), no estudo dos modelos de séries temporais busca-se obter meios capazes de inferir as características de seus processos geradores, bem como buscar modelos estocásticos que sejam capazes de descrever as situações particulares que ocorrem na verdadeira realidade.

2.4.1 Modelos Auto-Regressivo - AR (p)

Em um modelo auto-regressivo, a série de dados histórica Z_t é descrita por seus valores passados regressores e pelo ruído aleatório a_t . Assim, um modelo AR(p) é dado por.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (2.4)$$

em que ϕ_i é o parâmetro que descreve como Z_t se relaciona com o valor Z_{t-i} para $i = 1, 2, \dots, p$.

O modelo AR(p) dado pela Equação (2.7) pode ser reescrito conforme apresentado na Equação (2.8), adicionando-se um argumento chamado de operador de defasagens $L = Z_t - Z_{t-1}$ e aplicando-se o operador de defasagens em Z_t tem-se

$$1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p = \phi(L) \times Z_t = a_t. \quad (2.5)$$

O modelo auto-regressivo de ordem 1 ou AR(1) é a versão mais simples dessa classe de modelos. Sua apresentação algébrica é dada pela seguinte equação

$$Z_t = \phi_1 Z_{t-1} + a_t.$$

Para o modelo ser estacionário é necessário que $|\phi_1| < 1$ (condição de estacionariedade) e que as autocovariâncias (γ_k) sejam independentes. No caso do modelo AR(1), as autocovariâncias são dadas por

$$\gamma_k = \phi_1^k \gamma_0$$

e as autocorrelações ρ_k são dadas pela equação

$$\rho_k = \frac{\rho_k}{\rho_0} = \phi_1^k \quad k = 0, 1, 2, \dots \quad (2.6)$$

A função de autocorrelação decai exponencialmente quando ϕ_1 é positivo; quando ϕ_1 é negativo, a função de autocorrelação também decai exponencialmente, mas apresenta alternância de sinais positivos e negativos.

2.4.2 Modelos de Médias Móveis - MA (q)

Segundo Morettin e Tolo (2004), o nome média móvel é utilizado porque, a cada período, a observação mais antiga é substituída pela mais recente, calculando-se uma mais recente. Um processo é classificado como sendo de médias móveis de ordem q , onde

o mesmo é denotado por $MA(q)$, quando apresentar as seguintes características, a_t é um processo aleatório com média μ e variância σ_a^2 . Com isto, o modelo de médias móveis tem a seguinte representação,

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

onde define-se a_t como ruído branco da série e θ_j é a média móvel.

2.4.3 Modelos Auto-Regressivo e de médias Móveis ARMA (p, q)

Modelos auto regressivos e de médias móveis é uma junção dos modelos AR e MA , onde esta junção é denotada por $ARMA(p, q)$. Uma das formas de realizar previsões é ajustar um modelo capaz de descrever o comportamento da série. Com isto, o modelo autoregressivo e de médias móveis é um modelo eficiente, pois é utilizado para ajustar funções com base nos seus valores passados e na média da série, seguindo uma ordem (p, q) , o modelo tem a seguinte representação

$$Z_t = \phi Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q},$$

em que os $Z_{t-1} + Z_{t-2} + \dots + Z_{t-p}$ são independentes de a_t . Os valores da série Z_t são combinações lineares dos p -valores passados mais um termo a_t , no qual incorpora argumentos na série até o tempo t que não é explicado pelos valores passados (MORETTIN; TOLOI, 2006).

Metodologia de Box-Jenkins

A metodologia de Box e Jenkins é bastante utilizada em análise de séries temporais. Esta metodologia ajusta modelos auto regressivos integrados e de médias móveis, ARIMA (p, d, q) a um conjunto de dados. Pode-se observar particularidades nos modelos ARIMA como processos AR (Autoregressivos) e MA (Médias Móveis).

Diante do enfoque de Box e Jenkins, pode-se dizer que a análise de séries temporais tem por objetivo principal a realização de previsões. Essa metodologia estabelece mecanismo no qual valores futuros de uma série possam ser previstos com base apenas em seus valores presentes e passados. As técnicas desse processo de previsão fundamentam-se na exploração da correlação temporal que pode existir entre os valores exibidos pela série.

2.4.3.1 Modelo Auto-Regressivo Integrado e de Médias Móveis ARIMA (p, d, q)

Uma das modelagens mais utilizadas em previsões de séries temporais, o modelo auto-regressivo integrado de médias móveis (ARIMA) é uma generalização do modelo auto-regressivo de médias móveis (ARMA). Desenvolvido pelos estatísticos Box e Jenkins (1976), este também é conhecido por modelo de Box-Jenkins.

Então, a representação ARIMA (p, d, q) , refere-se, à ordens de auto-regressão, integração e de média móvel. Formando a classe de processos lineares estacionários, sendo assim, p é o número de termos auto-regressivo, d é o número de diferenças, q é o número de termos da média móvel.

Mas os modelos ARIMA podem descrever uma classe de processos lineares não estacionários homogêneos, ou seja, as séries não são estacionárias em nível de inclinação. De acordo com Morettin e Tolo (1987), estas séries podem ser transformadas em séries estacionárias, aplicando um número determinado de diferenças entre seus dados, onde este procedimento é necessário ser realizado no máximo, duas vezes. A primeira diferenciação entre os dados é aplicado da seguinte forma;

$$W_t = \Delta Z_t = Z_t - Z_{t-1}$$

sendo,

- i)* y_t observação y , no período t da série y_t sem diferenciação;
- ii)* y_{t-1} observação y , no período $t - 1$ da série y_t sem diferenciação;
- iii)* $\Delta y_t = Z_t$ observação Z , no período t , pertencente a série Z_t com dados da série y_t diferenciados a primeira vez.

Se com a primeira diferenciação entre os dados não for obtida uma série estacionária, então será realizada a segunda diferenciação, para isto a aplicação é dada da seguinte forma

$$\Delta^2 Z_t = \Delta (\Delta Z_t) = \Delta (Z_t - Z_{t-1})$$

Identificação de Modelos ARIMA

Sendo considerada a etapa mais crítica ao se utilizar a modelagem ARIMA, a identificação particular de um modelo é feita principalmente com base nas autocorrelações e autocorrelações parciais estimadas, este procedimento é realizado com o intuito de deter-

minar os possíveis valores dos parâmetros p , d e q do modelo $ARIMA(p, d, q)$. De acordo com Morettin e Tolo (2004), o procedimento de identificação consiste de três etapas:

- 1^a: Verificar se a série é estacionária, por meio da análise do gráfico da série original e sua função de autocorrelação.
- 2^a: Diferenciar a série até que se torne estacionária por meio do operador de diferença, onde o número de diferenças necessárias d é obtido quando a FAC amostral $\Delta^d Z_t$ decai com rapidez para zero.
- 3^a: Conhecer o processo $ARMA(p, q)$ por meio da análise das autocorrelações e autocorrelações parciais estimadas que devem imitar os comportamentos das respectivas quantidades teóricas.

Função de Autocovariância - FAC

Seja $\{Z_t, t \in T\}$ um processo estacionário real discreto, de média zero e $\gamma_\tau = E\{Z_t \times Z_{t+\tau}\}$. A função de autocovariância (FAC) de um processo estocástico satisfaz às seguintes propriedades:

- i) $\gamma_{-\tau} = \gamma_\tau$;
- ii) $|\gamma| \leq 0$;
- iii) γ_τ é positiva definida, no sentido que,

a função de autocorrelação do processo é definida por,

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad \tau \in T$$

e têm as propriedades de γ_τ , exceto para $\rho_\tau = 1$. Sendo Z_1, \dots, Z_n , a FAC ρ_j é estimada por,

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, n-1, \quad (2.7)$$

em que c_j é a estimativa da função de auto-covariância γ_τ .

Função de Autocorrelação Parcial - FACP

Outra forma de realizar a identificação do modelo, a função de autocorrelação parcial é uma medida que corresponde a correlação entre Z_t e Z_{t-1} . Denotado por ϕ_{ρ_i} o i -ésimo coeficiente de um modelo $AR(\rho)$, tem-se que ϕ_{ρ_i} é o último coeficiente.

Segundo Morettin e Toloi (2004), um método geral para encontrar a FACP para um processo estacionário com *FAC* ρ_i consiste em utilizar as equações de Yule-Walker, com isto,

$$\rho_j = \phi_{\rho}\rho_{i-1} + \phi_{\rho 2}\rho_1 + \dots + \phi_{\rho p}\rho_{i-1}, \quad i = 1, 2, \dots, \rho.$$

Desenvolvendo as equações,

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_3 + \rho_2^2\rho_1 + \rho_1^3 - 2\rho_1\rho_2 - \rho_1^2\rho_3}{1 - 2\rho_1^2 - \rho_2^2}.$$

Para $\hat{\rho} = 1, 2, \dots$, obtém-se $\phi_{\rho\rho}$ da seguinte maneira,

$$\phi_{11} = \rho_1.$$

De modo geral, escreve-se

$$\phi_{\rho\rho} = \frac{|\mathbf{P}_\rho^*|}{|\mathbf{P}_\rho|}.$$

Sendo, \mathbf{P}_ρ a matriz de autocorrelações, \mathbf{P}_ρ^* é a matriz \mathbf{P}_ρ com a última coluna substituída pelo vetor de autocorrelações. Nos processos *AR*, *MA* e *ARMA* seguem com as seguintes FACP. Um processo *AR*(p) tem FACP $\phi_{kk} \neq 0$ para $k \leq p$ e $\phi_{kk} = 0$, $k > p$. Um processo *MA*(q) tem FACP similar a FAC de um *AR*(p). Um processo *ARMA*(p, q) tem FACP que se comporta como a FACP de um processo *MA*.

De acordo com os fatores acima, vale ressaltar que a FACP é útil para identificar modelos *AR* puros, não sendo útil para identificar modelos *MA* e *ARMA*. Obtém-se a

FACP substituindo-se as FA ϕ_j por suas estimativas nas equações de Yule-Walker,

$$r_j = \hat{\phi}_{k1}r_{j-1} + \dots + \hat{\phi}_{kk}r_{j-k}, \quad j = 1, \dots, k, \quad (2.8)$$

e resolver estas equações para $k = 1, 2, \dots$

Estimativas dos Modelos ARIMA

De acordo com Morettin e Toloi (2004), no modelo autorregressivo pode-se definir facilmente os resíduos. Considerando-se em particular um modelo $AR(2)$ com um termo constante:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$$

tendo estimado ϕ_1, ϕ_2 , os resíduos são definidos como

$$\hat{a}_t = \hat{Z}_t - \hat{\phi}_1 Z_{t-1} - \hat{\phi}_2 Z_{t-2}.$$

Os resíduos iniciais \hat{a}_1 e \hat{a}_2 podem ser obtidos do procedimento de estimação utilizando valores passados para Z_0 e Z_{t-1} . Para o modelo geral $ARMA$, é necessário colocar na forma autorregressiva de ordem infinita, para definir os resíduos, tem-se então a seguinte equação

$$Z_t = \sum_{j=1}^{\infty} \pi_j Z_{t-j} + a_t \quad (2.9)$$

e os resíduos serão definidos da seguinte expressão

$$\hat{a}_t = Z_t - \sum_{j=1}^{\infty} \hat{\pi}_j Z_{t-j}.$$

Assim sendo, os π_j não são estimadores diretamente da expressão acima denotada, mas de funções implícitas, de ϕ 's e θ 's, utilizando-se a equação $|\hat{\phi}_{jj}| > \frac{2}{N}$, $j > p$ tem-se que

$$\hat{Z}_t = \sum_{j=1}^{\infty} \hat{\pi}_j Z_{t-j}$$

em que a melhor previsão de Z_t se baseia em $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$. Além disso a equação

$$\rho_1 = \frac{-\theta}{1 + \theta^2}$$

O Teste de Box-Pierce

Box e Pierce (1970), propuseram um teste para as autocorrelações dos resíduos estimados, em que, apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos. Se o modelo for adequado, a estatística Q aproxima-se de uma distribuição qui-quadrado (χ^2).

$$Q = n(n + 2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{(n - k)}$$

em que n é o número de observações da série temporal e k é o número de lags da FAC amostral.

Sua distribuição amostral também é aproximadamente qui-quadrado com $m-p-q$ graus de liberdade. Na prática o número m de autocorrelações amostrais é tipicamente escolhido, com distribuição qui-quadrado, χ^2 , com (k, p, q) graus de liberdade (LJUNG; BOX, 1978). A hipótese de ruído branco para os resíduos é rejeitada para valores grandes de Q . Em geral basta tomar as 20 ou 25 primeiras $\hat{\rho}_k$.

Crítérios de Seleção de Modelos

Crítério de Informação Akaike - AIC

O critério de informação de Akaike (AIC) foi desenvolvido a partir da distância de Kullback-Leibler, a qual é uma distância entre o modelo verdadeiro, que geralmente é uma abstração, e o modelo candidato. Burnham e Anderson (2002) recomendam a utilização do AIC apenas quando $\frac{n}{p} \geq 40$. O critério de AIC é dado por

$$AIC = n \ln L + 2(p + 1),$$

em que L é a função de máxima verossimilhança do modelo; p é o número de parâmetros a serem estimados e n é o número de observações da amostra.

Deve-se atentar que o critério AIC compara apenas modelos e, portanto, não deve-se utilizá-lo como ferramenta isolada para a conclusão de um bom modelo. Neste estudo, além de se utilizar este critério, foi verificada a qualidade das previsões dos modelos por meio do Erro Quadrático Médio (EQM), que é denotado pela equação abaixo:

$$EQM = \frac{1}{n} \sum_{h=1}^n [Z_{t-h} - \hat{Z}_{t-h}]^2,$$

em que h representa o período de previsão além do tempo t e n é a quantidade de previsões realizadas.

Crítério de Informação Bayesiano - BIC

O critério de informação bayesiano é baseado no *AIC*, onde o mesmo é apresentado da seguinte forma:

$$BIC = -2 \ln L + p \ln n,$$

em que L é a função de máxima verossimilhança do modelo; p é o número de parâmetros a serem estimados e n é o número de observações da amostra.

De forma equivalente ao *AIC*, o modelo que apresentar o menor valor da estatística *BIC*, será o mais adequado. Vale ressaltar, que este critério de informação penaliza bem mais a inclusão de parâmetros do que o *AIC*, portanto tende a selecionar modelos mais parcimoniosos (EHLERS, 2007).

2.5 Previsão com Modelos ARIMA

Segundo Box e Jenkins (1976), para um modelo auto-regressivo de primeira ordem, o preditor h passo a frente é dado por:

$$\hat{Z}_t(h) = \phi_1^h Z_t + \mu - \mu \phi_1^h,$$

em que: $\hat{Z}_t(h)$ é a previsão h passos a frente com base no tempo t ; ϕ_1 é o coeficiente da variável em estudo defasado um período; e μ a média do processo.

Segundo Amorim (2014) na previsão de valores futuros com base em valores passados, é necessário que se disponha de uma memória histórica de dados ocorridos anteriormente. Assim sendo, o conjunto de dados, por si só, não permite a previsão dos valores futuros. Para isso, é necessário a utilização de algoritmos, técnicas ou métodos de previsão de séries temporais.

Mas com tudo isso, existem diversos métodos para auxiliar na previsão de série temporal. A metodologia de previsão é geralmente compreendida como um conjunto de abordagens, métodos e ferramentas para coleta de dados da série temporal para serem utilizados em previsão de valores futuros, baseando sempre em valores passados (PALIT; POPOVIC, 2005).

3 Resultados e Discussão

Darci início nesta seção com o estudo descritivo das precipitações climáticas dos municípios de Soledade, Taperoá, Cabaceira e Santa Luzia, localizados na Região do Cariri paraibano, ao longo dos anos de 1935 a 2013. Na avaliação da precipitação foram considerados os totais anuais. Os dados utilizados para realização deste trabalho foram obtidos na Agência Executiva de Gestão das Águas do Estado da Paraíba (AESAs).

Na sequência foram verificados os principais pressupostos para análise de séries temporais com objetivo de ajustar, de forma parcimoniosa, alguns modelos para a variável em estudo. Sendo assim, pode-se identificar o modelo que mais se adequa a caracterização da precipitação climática dos municípios em estudo. As análises foram realizadas com o auxílio do *software* estatístico R versão 3.2.0.¹.

Na descrição dos resultados, primeiramente procede-se com uma análise gráfica, conforme apresentado nas Figuras 3a, 3b, 3c e 3d, nos quais é possível identificar a necessidade de verificar alguns pressupostos exigidos na análise de séries temporais, tais como, estacionariedade e tendência.

De acordo com FAVA (2000), ao realizar a análise de séries temporais é necessário verificar a estacionariedade, para isto, aplicou-se o teste Dickey-Fuller (DF). Este teste procede-se tabulando-se os valores críticos através da simulação de Monte Carlo.

Se o valor absoluto da estatística τ for maior que o valor absoluto tabelado por DF, aceita-se a hipótese nula, logo a série é não estacionária. Procede-se também com a análise de tendência, sendo analisada com o teste de Mann-Kendall (SNEYERS, 1991), considerando-se que na hipótese de estabilidade de uma série temporal, a sucessão de valores ocorrem de forma independente e a distribuição de probabilidade deve permanecer sempre a mesma. Goossens e Berger (), afirmam que o teste de Mann-Kendall é o método mais apropriado para analisar mudanças climáticas, além de permitir a detecção e localização aproximada do ponto inicial de determinada tendência.

¹www.r-project.org

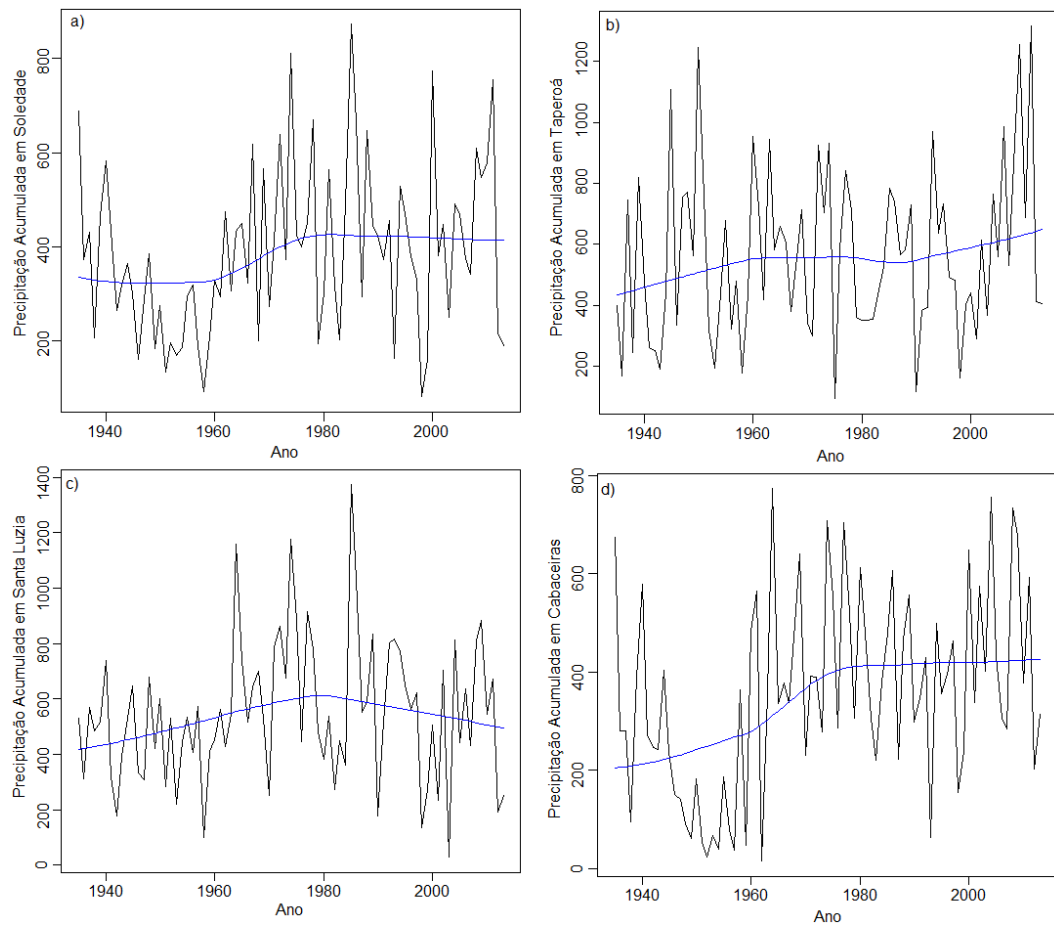


Figura 3: Precipitação acumulada anual(mm) no período de 1935 a 2013 para os municípios de Soledade (a), Taperoá (b), Santa Luzia (c) e Cabaceira(d), localizados na região do Cariri paraibano.

Dando sequência às análises, pode-se observar, por meio da Tabela 1, os resultados do teste de Mann-Kendall. Estes resultados corroboram para uma tendência monótona ao longo do tempo para os municípios de Soledade e Cabaceira (p -valor $< 0,05$) e ausência de tendência para os municípios de Taperoá e Santa Luzia, concluindo assim que as observações das séries temporais desses dois municípios são independentes e identicamente distribuídas.

Em contrapartida, pode-se observar pelo teste da raiz unitária de Dick-Fuller (Tabela 1), que a hipótese de nulidade H_0 não foi rejeitada (p -valor $< 0,05$) para as localidades de Soledade, Taperoá e Cabaceira. É importante observar que a série temporal do município de Taperoá é independente ao longo do tempo. Sendo assim, admite-se, para estas três localidades, que as séries temporais são não-estacionárias. É importante ressaltar que para o município de Santa Luzia, o teste de Dickey-Fuller corroborou para a estacionariedade da série temporal.

Tabela 1: Resultados obtidos com a aplicação dos respectivos testes, Dickey-Fuller e Mann-Kendall para as localidades de Soledade, Taperoá Santa Luzia e Cabaceira.

Cidades	<i>Dickey – Fuller</i>	<i>p – valor</i>	<i>Mann – Kendall</i>	<i>p – valor</i>
Soledade	-2,81	0,246	0,164	0,033
Taperoá	-3,46	0,052	0,105	0,170
Santa Luzia	-3,693	0,031	0,101	0,191
Cabaceira	-2,058	0,552	0,221	0,004

O próximo passo é analisar as funções de autocorrelações (FAC) e de autocorrelações parciais (FACP) da série precipitação nos municípios em estudo. O comportamento dessas funções poderão auxiliar na verificação da estacionariedade e na proposição do modelo, conforme observa-se na Figura 4.

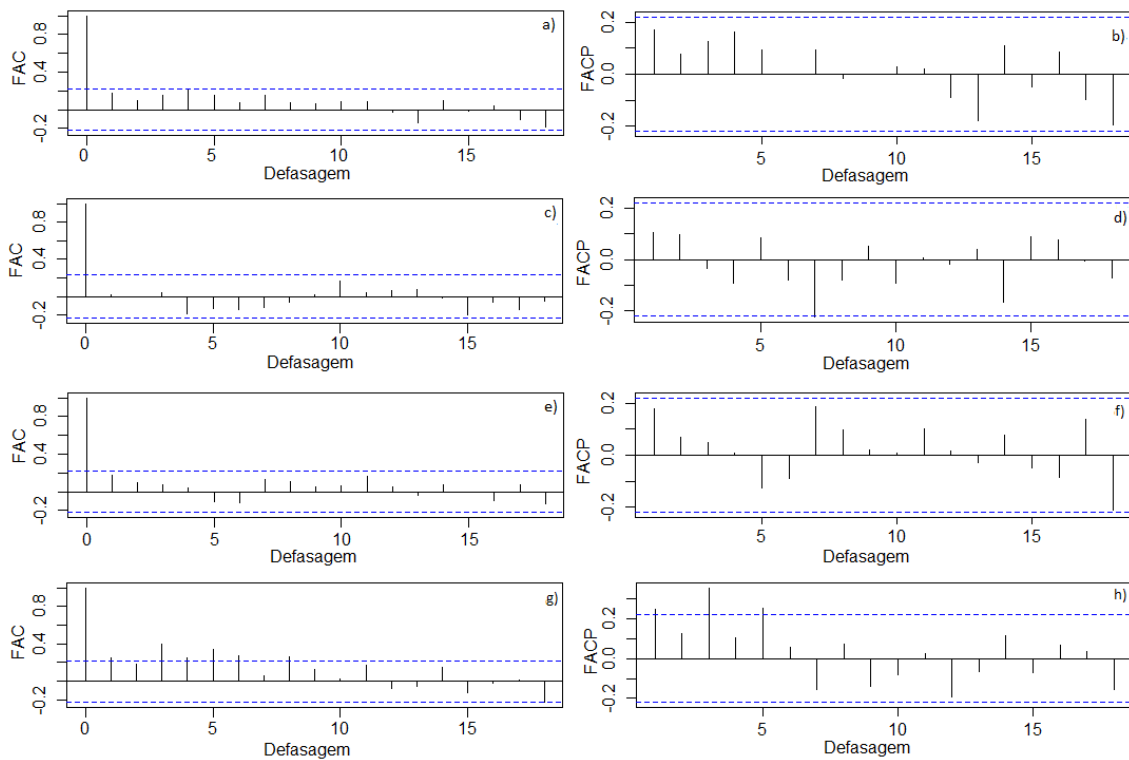


Figura 4: Funções de autocorrelações (FAC) e funções autocorrelações parciais (FACP) para os municípios de Soledade (a, b), Taperoá (c, d), Santa Luzia (e, f) e Cabaceira (g, h).

Com base nas FAC's (Figuras 4a, 4c, 4e e 4g) e nas FACP's (Figuras 4b, 4d, 4f e 4h), iniciou-se o processo de identificação da ordem dos modelos a serem ajustados aos dados de precipitação para os municípios de Soledade, Taperoá, Santa Luzia e Cabaceira,

com objetivo de encontrar um modelo que mais se adequa aos dados em análise. Desta forma, foram analisados alguns modelos com diferentes estruturas para as séries temporais de precipitação nos municípios descritos anteriormente. Comparou-se os resultados e a significância dos parâmetros adotando-se alguns critérios de adequabilidade de modelos.

Na sequência, procede-se com o ajuste dos modelos aos dados de precipitação desses municípios. Conforme se observa nas Tabelas 2, 3, 4 e 5, para as localidades de Soledade, Taperoá, Santa Luzia e Cabaceira, conclui-se que o potencial modelo ARIMA, para os municípios identificados é o AR, I e MA, ou seja, com o parâmetro p , d e q , em que p representa o número de termos autoregressivos, d é o número de diferenças e q o número de termos da média móvel.

Ainda de acordo com as Tabelas 2, 3, 4 e 5, pode-se afirmar que o melhores modelos para representar as séries de precipitação são os que apresentam menores valores de acordo com os critérios de seleção adotados respectivamente, como se segue, equivalentemente, pois eles apresentam menores valores de AIC , menores valores de BIC e menores variabilidades estimadas entre os modelos concorrentes nos municípios em estudo.

Tabela 2: Critérios de seleção de modelos autorregressivos integrados e de médias móveis para o município de Soledade

MODELOS	$\hat{\sigma}^2$	EQM	AIC	BIC
ARIMA (1, 1, 1)	29065	170,29	1032,84	1039,91
ARIMA (1, 1, 2)	30068	172,30	1034,84	1044,26
ARIMA (2, 1, 1)	29790	171,50	1034,10	1043,52
ARIMA (2, 1, 2)	29726	171,32	1035,92	1047,71

Conforme resultados apresentados na Tabela 2, o modelo autoregressivo integrado e de médias móveis, ARIMA(1,1,1), é o melhor modelo para representar as previsões de precipitação no município de Soledade, tendo em vista que esse modelo possui menores valores de AIC , BIC , EQM e $\hat{\sigma}^2$, ou seja, o processo que descreve a série temporal é um ARIMA(1,1,1) com as seguintes estimativas dos parâmetros

$$\begin{aligned}\hat{Z}_t &= \hat{\phi} \times Z_{t-1} - \hat{\theta} \times a_{t-1} + a_t \\ \hat{Z}_t &= 0,0066 \times (Z_{t-1}) + 0,8406 \times (a_{t-1}) + a_t\end{aligned}$$

Estes resultados corroboram com Chechi e Sanches (2013). Os referidos autores anali-

saram o comportamento das séries temporais de temperatura e precipitação no município de Nossa Senhora da Glória, Estado do Sergipe. Utilizando-se do modelo ARIMA os autores verificaram que a variação, tanto da temperatura quanto da precipitação, não são regulares ao longo do tempo, demonstrando que o trimestre janeiro, fevereiro e março possuem maiores variações. Os resultados mostraram ainda que as temperaturas no mês de março apresentaram os maiores índices de variação.

É importante observar que a precipitação no Estado da Paraíba é afetada por vários sistemas meteorológicos, tais como: Zona de Convergência Intertropical (ZCIT), Vórtices Ciclônicos de Altos Níveis (VCAN), linhas de instabilidade, brisas marítimas e terrestres, perturbações de frentes frias dentre outros que atuam em diferentes meses do ano, com características diferentes de um ano para outro (PEDROZA, 2009). A média anual de precipitação pluvial é inferior aos $800mm$ na região do Cariri paraibano e pode superar os $1500mm$ na região litorânea (VICENTE *et al.*, 2003). Essa variabilidade de chuva tem causado bastante interesse por parte dos pesquisadores na busca de explicar tais fenômenos ao longo do tempo.

Tabela 3: Critérios de seleção de modelos autorregressivos integrados e de médias móveis para o município de Taperoá

MODELOS	$\hat{\sigma}^2$	<i>EQM</i>	<i>AIC</i>	<i>BIC</i>
ARIMA (1, 1, 1)	73736	269,82	1106,73	1116,80
ARIMA (1, 1, 2)	74604	271,40	1107,80	1117,23
ARIMA (2, 1, 1)	72943	268,34	1106,71	1116,14
ARIMA (2, 1, 2)	72986	268,36	1108,68	1120,47

Continuando-se com a discussão e a descrição dos resultados pode-se observar, por meio da Tabela 3, que o modelo autorregressivo integrado e de médias móveis com parâmetros $p=2$, $d=1$ e $q=1$, é o melhor modelo para representar as previsões de precipitação no município de Taperoá, ou seja, o processo que descreve a série temporal é um ARIMA(2, 1, 1) com as seguintes estimativas

$$\begin{aligned}\hat{Z}_t &= \hat{\phi}_1 \times Z_{t-1} - \hat{\phi}_2 \times Z_{t-2} + \hat{\theta} \times a_{t-1} + a_t \\ \hat{Z}_t &= 0,1100 \times Z_{t-1} - 0,1156 \times Z_{t-2} - 1,0000 \times a_{t-1} + a_t\end{aligned}$$

Tabela 4: Critérios de seleção de modelos autorregressivos e de médias móveis para o município de Santa Luzia

MODELOS	$\hat{\sigma}^2$	<i>EQM</i>	<i>AIC</i>	<i>BIC</i>
ARIMA (1, 0, 1)	60436	245,84	1101,98	1111,45
ARIMA (1, 0, 0)	60327	245,63	1100,47	1107,58
ARIMA (0, 0, 1)	61112	247,21	1100,83	1107,94
ARIMA (1, 0, 2)	60400	245,76	1103,93	1115,78
ARIMA (2, 0, 1)	60412	245,78	1103,95	1115,79
ARIMA (2, 0, 2)	60430	245,83	1105,97	1120,19

De acordo com os modelos autorregressivos testados para a variável precipitação no município de Santa Luzia (Tabela 4), o modelo que melhor se ajustou, em termos de significância estatística, foi o ARIMA (1,0,0), ou seja, o modelo AR(1) sem diferenciação e sem média móvel, em que o AR(1), indica o número de termos autorregressivos.

A representação da escolha desse modelo para a variável precipitação no município de Santa Luzia, de acordo com os critérios de seleção, AIC, BIC, EQM e $\hat{\sigma}^2$, ficaram muito próximos dos demais modelos testados. Percebe-se que nos demais modelos, a inclusão de um termo de médias móveis não altera significativamente os resultados dos critérios de seleção, a mudança acontece apenas nas casas decimais. O processo que descreve a série temporal é um AR(1) com a seguinte estimativa

$$\begin{aligned}\hat{Z}_t &= \hat{\phi}_1 \times \hat{Z}_{t-1} + \hat{a}_t \\ \hat{Z}_t &= 0,1809 \times Z_{t-1} + a_t.\end{aligned}$$

Sendo, $\hat{\phi} = 0,1809$ a função de autocorrelação que decai exponencialmente, com valores todos positivos. Sendo-se assim, a função de autocorrelação decai exponencialmente para zero.

Tabela 5: Critérios de seleção de modelos autorregressivos integrados e de médias móveis para o município de Cabaceira

MODELOS	$\hat{\sigma}^2$	EQM	AIC	BIC
ARIMA (1, 1, 1)	34200	183,75	1042,69	1049,76
ARIMA (1, 1, 2)	32717	179,72	1041,43	1050,85
ARIMA (2, 1, 1)	31432	176,00	1039,04	1048,46
ARIMA (2, 1, 2)	31495	176,34	1040,46	1052,24

De acordo com a Tabela 5, o modelo que melhor se ajustou a variável precipitação na localidade de Cabaceira foi um ARIMA (2, 1, 1). Tais composições tomam como base o critério de análise dos índices penalizadores AIC e BIC, EQM e $\hat{\sigma}^2$, procurando-se os menores valores de tais índices. Observa-se no modelo selecionado (Tabela 5), uma redução considerável na variância estimada ($\hat{\sigma}^2$), corroborando assim para uma redução na estimativa do Erro Quadrático Médio (EQM), com as seguintes estimativas dos parâmetros

$$\begin{aligned}\hat{Z}_t &= \hat{\phi} \times Z_{(t-1)} + \hat{\phi} \times Z_{(t-1+1)} - \hat{\theta} \times a_t + a_t \\ \hat{Z}_t &= 0,4629 \times Z_{t-1} - 1,4296 \times Z_{t-1} - 0,6117 \times a_t + a_t\end{aligned}$$

Michael *et al.* (2006) ao estudarem a regionalização e distribuição espacial e temporal da precipitação na bacia hidrográfica do Xingu, Estado do Pará, observaram que a precipitação nesta região variava, aproximadamente, entre 1.500 a 2.500mm por ano, sendo que nas partes norte e sudoeste da bacia, ocorreram as maiores magnitudes de chuvas, principalmente causadas pela atuação dos sistemas atmosféricos conhecidos como Zona de Convergência Intertropical (ZCIT) e Zona de Convergência do Atlântico Sul (ZCAS), respectivamente.

Após a escolha dos modelos é necessário realizar a análise de resíduos (observações *versus* valores ajustados). Se o modelo for adequado aos dados espera-se que os resíduos se distribuam aleatoriamente em torno de zero com variância aproximadamente constante. Também espera-se que os resíduos sejam independentes e possuam, aproximadamente, uma distribuição normal. Tal ajuste está diretamente associado à capacidade de explicação dos modelos, no que se refere à variabilidade das precipitações para cada localidade utilizada neste estudo.

Com base nos modelos selecionados, pode-se observar que os pressupostos associados aos resíduos não foram violados. Este fato foi confirmado pelo teste de normalidade de

Shapiro-Wilk, em que a estatística calculada para os municípios de Soledade $W=0,9769$ (p -valor= $0,1628$), Taperoá $W=0,9661$ (p -valor= $0,05365$), Santa Luzia $W=0,9756$ (p -valor = $0,135$) e Cabaceira $W=0,9776$ (p -valor= $0,1781$), permaneceram num patamar desejável. Sendo assim, não rejeitou-se (p -valor $> 0,05$) a hipótese nula de que os dados provinham de uma distribuição normal.

Analisando-se as funções de autocorrelações residuais (Figura 5), pode-se afirmar que os modelos ajustados aos dados mostraram eficientes. Estas evidências também foram confirmadas pelo teste de Ljung-Box, em que o valor calculado da estatística para os municípios de Soledade $Q=0,0249$ (p -valor= $0,8747$), Taperoá $Q=0,0454$ e (p -valor= $0,8313$), Santa Luzia $Q=0,0096$ e (p -valor= $0,9221$) e $Q=0,5578$ e (p -valor= $0,4552$), permaneceram num patamar aceitável. E, portanto, não rejeitando-se a hipótese nula da não existência de autocorrelação residual.

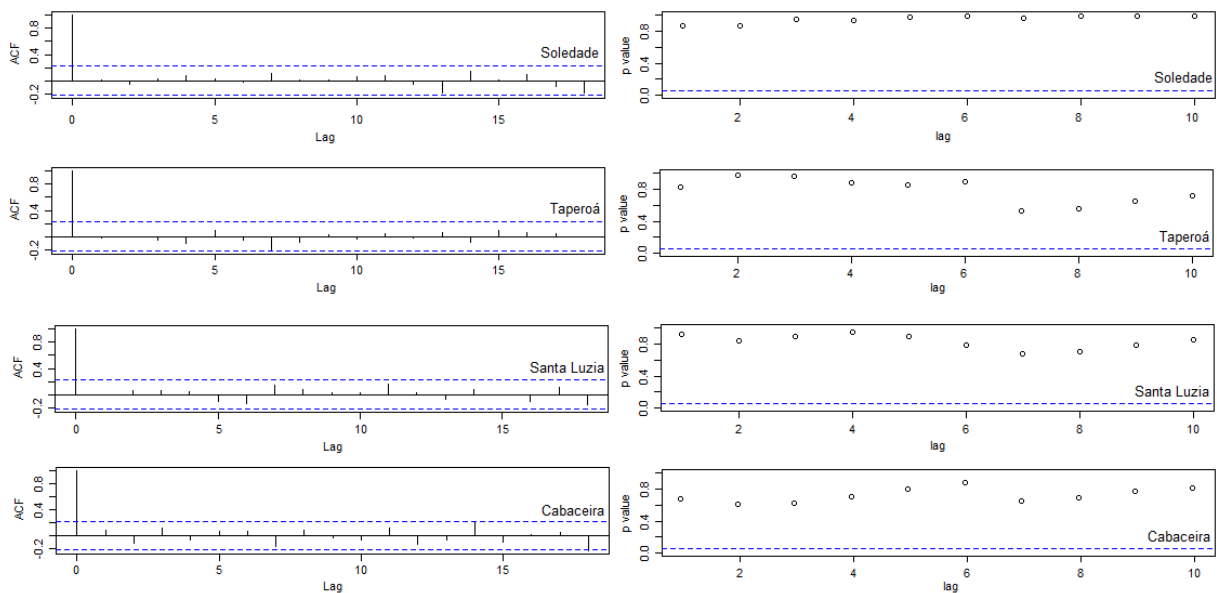


Figura 5: Gráficos dos resíduos e Gráficos de Box-Pierce para os municípios de Soledade, Taperoá, Santa Luzia e Cabaceira, respectivamente.

Analisando-se os gráficos de resíduos, pode-se observar um comportamento de ruído branco, com isto, é possível afirmar que o modelo encontra-se adequadamente especificado, ou seja, as autocorrelações das variâncias residuais quadráticas são não significativas. Este fato é corroborado pelo teste de Ljung-Box, reforçando assim, esta afirmativa. Pode-se observar que todas as observações possuem um p -valor consideravelmente alto, indicando que os resíduos são independentes.

Segundo Amorim (2014), a partir do modelo ajustado é possível prever os valores

futuros com base em valores passados. A partir da Figura 6, observa-se a série original com os valores preditos para os próximos dez anos da variável precipitação nas localidades de Soledade, Taperoá, Santa Luzia e Cabaceira. As composições dos valores preditos estão dentro do intervalo de confiança na série original, entretanto o comportamento da série apresenta uma previsão futura em torno da média acumulada com um pequeno e suave decréscimo. Estes valores de certa forma também podem ser vistos como uma média das observações passadas, mas dentro de uma tendência de estabilidade da série.

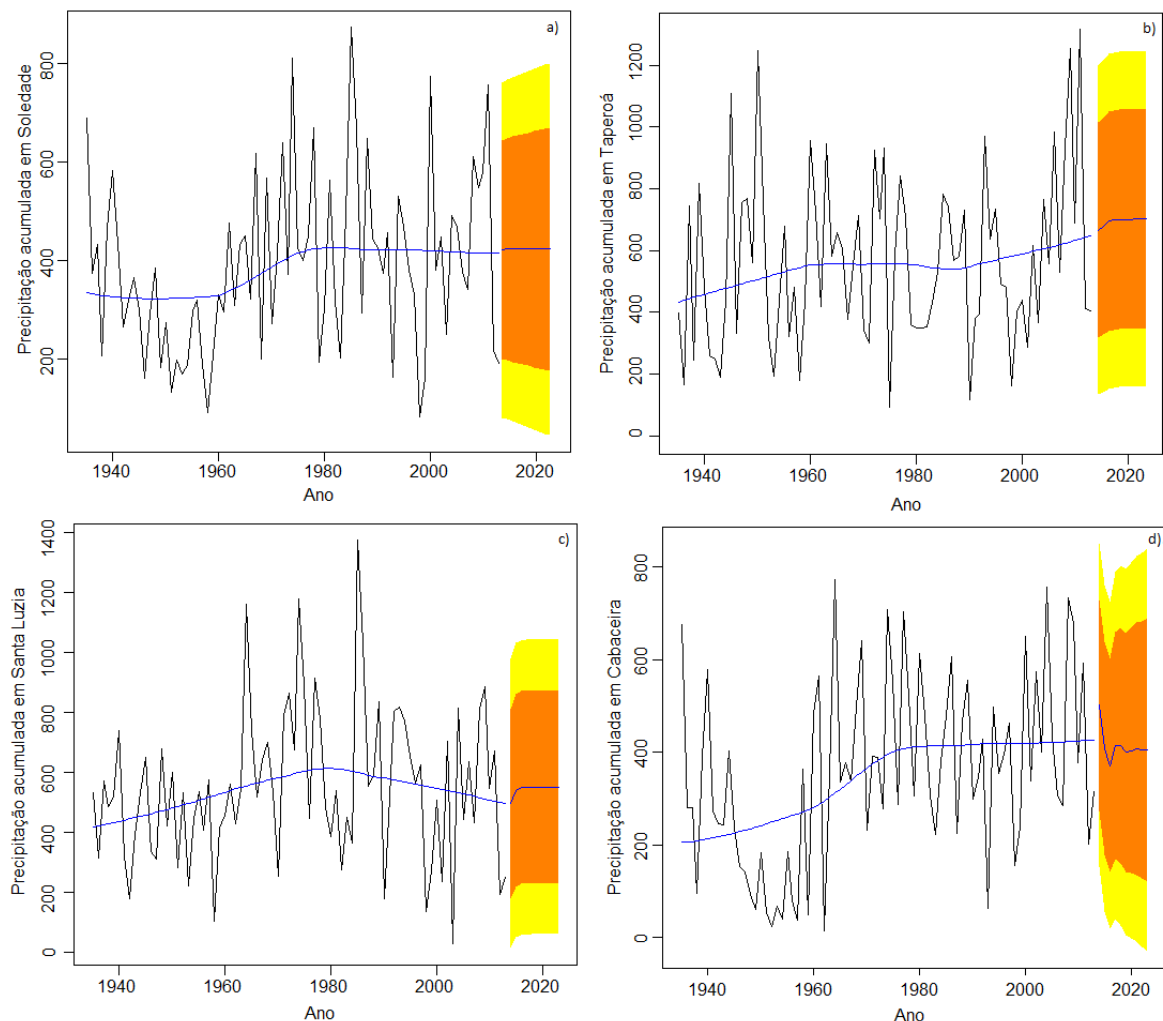


Figura 6: Previsão da Precipitação acumulada anual(mm) para os próximos dez anos nos municípios de Soledade (a), Taperoá (b), Santa Luzia (c) e Cabaceira (d)

Os valores previstos apresentam tendência significativa de aumento na precipitação acumulada para o município de Taperoá (Figura 6b), ainda é notório que o comportamento da previsão é bem melhor descrita quando comparado às demais localidades, justificado pela menor amplitude dos desvios. A aleatoriedade é extremamente associada

a precipitação acumulada dos municípios de Soledade (Figura 6a), Santa Luzia (Figura 6c) e Cabaceira (Figura 6d), fazendo com que o modelo de previsão seja suavizado em torno da média geral. Tal variabilidade na série temporal resulta em muitas possibilidades para o modelo fazendo com que sua interpretação limite-se a valores próximos a média dos valores passados.

Aparentemente, este fato parece ser bem particular da Região do Cariri paraibano, pois num estudo sob séries temporais com as variáveis precipitação e temperatura para a cidade de Rio Claro no estado de São Paulo, (MOURA, 2009) observou que há tendência de aumento nos índices extremos de precipitação, salientando que os valores não são significativos estatisticamente, no entanto, na análise de temperatura onde quase todos os índices extremos são significativos.

Entretanto, analisando índices extremos no estado da Paraíba, Nóbrega *et al.* (2014) observaram que a quantidade dos dias consecutivos secos estão aumentando a cada ano, na mesma proporção em que os dias consecutivos úmidos estão diminuindo, assim indicando que a distribuição da precipitação vem sendo modificada ao longo das décadas principalmente em termos de valores totais anuais, corroborando nas tendências dos valores abaixo da média climatológica anual, conjecturando uma possível causa a essas alterações na precipitação, ao mal uso do solo e devido a intensificação do uso da vegetação sem planejamento, degradando os recursos naturais dos ecossistemas com o desmatamento.

4 Conclusão

Diante do exposto pode-se concluir que entre os anos 1935 a 2013 houve pequenas alterações no período chuvoso, tanto na cidade de Soledade, Taperoá, Santa Luzia, quanto na cidade de Cabaceira . De acordo com os critérios de seleção AIC, BIC, EQM e $\hat{\sigma}^2$ o modelo autorregressivo integrado e de médias móveis foi o mais adequado para representar a precipitação dos municípios de Soledade, Taperoá e Cabaceira. No entanto o município de Santa Luzia o modelo que mais se Adequou aos foi o autorregressivo para representa onde através dos mesmo, constatou-se que a precipitação para os próximos anos permanecerá em torno da sua média.

Após a seleção dos modelos que melhor se ajustaram aos dados, pode-se verificar por meio da previsão que os valores obtidos encontrou-se num patamar esperado para os próximos dez anos. No entanto, a previsão por meio da utilização de séries temporais permitiu-se verificar fatores que influenciam na variável precipitação, sendo assim, a modelagem estatística via séries temporais garante uma maior confiabilidade no estudo da precipitação nos municípios de Soledade, Taperoá, Santa Luzia e Cabaceira, situadas na região do Cariri paraibano.

Referências

- ALVES, J. J. Geoecologia da Caatinga no Semi-árido do Nordeste Brasileiro. *CLIMEP-Climatologia e Estudos da Paisagem*, v. 2, n. 1, p. 11–15, 2007.
- AMORIM, M. R. B. *Modelagem em Séries Temporais Aplicados a Dados de Temperatura Máxima na Cidade de Campina Grande–PB*. 44 p. Monografia (Centro de Ciência e Tecnologia) — Universidade Estadual da Paraíba, 2014.
- BEZERRA, M. I. S. Apostila de Análise de Séries Temporais. *Curso de Estatística, available in <http://people.ufpr.br/lucambio/CE017/1S2010/5515941-Apostila-Series-Temporais.pdf>*, 2006.
- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: forecasting and control*, revised ed. Holden-Day, p. 450, 1976.
- BOX, G. E. P.; PIERCE, D. A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American statistical Association*, Taylor e Francis Group, v. 65, n. 332, p. 1509–1526, 1970.
- BURNHAM, K. P.; ANDERSON, D. R. *Model Selection and Multimodel Inference: a practical information-theoretic approach*. Springer Science & Business Media, p. 185, 2002.
- CHECHI, L.; OLIVEIRA SANCHES, F. 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.
- EHLERS, R. S. Análise de Séries Temporais. *Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná*, p. 90, 2007.
- FAVA, V. L. *Manual de Econometria*. Vasconcelos, MAS; Alves, D. São Paulo: Editora Atlas, p. 308p, 2000.
- FISCHER, S. Séries Univariantes de Tempo-Metodologia de box & jenkins. *Fundação de Economia e Estatística*, p. 450, 1982.
- GOOSSENS, C.; BERGER, A. Annual and Seasonal Climatic Variations Over The Northern Hemisphere and Europe During The Last Century. In: *Annales Geophysicae*. [S.l.: s.n.]. v. 4, n. 4, p. 385–400, 1986.
- KOPECKY, K. J.; PIERCE, D. A. Efficiency Of Smooth Goodness-Of-Fit Tests. *Journal of the American Statistical Association*, Taylor & Francis, v. 74, n. 366a, p. 393–397, 1979.

- LJUNG, G. M.; BOX, G. E. P. On a Measure of Lack of fit in Time Series Models. *Biometrika*, Biometrika Trust, v. 65, n. 2, p. 297–303, 1978.
- MICHAEL, D. G.; ADAMSON, P.; ALEXOPOULOS, T.; ALLISON, W. W. M.; ALNER, G. J.; ANDERSON, K.; ANDREOPOULOS, C.; ANDREWS, M.; ANDREWS, R.; ARMS, K. E. Observation of Muon Neutrino Disappearance With The Minos Detectors in The Numi Neutrino Beam. *Physical Review Letters*, APS, v. 97, n. 19, p. 191–801, 2006.
- MORETTIN, P. A.; TOLOI, C. *Análise de Séries Temporais*. [S.l.]: Blucher, 2004. 535 p.
- MORETTIN, P. A.; TOLOI, C. C. M. *Modelos Para Previsão de Séries Temporais*. [S.l.]: Instituto de matematica pura e aplicada, 2006. 356 p.
- MORETTIN, P. A.; TOLOI, C. M. C. *Previsão de Séries Temporais*. [S.l.: s.n.], 1987. 400 p.
- MOURA, M. A. C. de. A Urbanização em Campina Grande e Suas Relações com a Incidência de Doenças Respiratórias no Município e o Clima Local. p. 166, 2009.
- NASCIMENTO, S. S.; ALVES, J. J. A. Ecoclimatologia do Cariri Paraibano Ecoclimatology Of The Cariri Paraibano. *Revista Geográfica Acadêmica Volume 2 No. 3 (xii. 2008)*, v. 2, n. 3, p. 28, 2008.
- NÓBREGA, J. N.; SANTOS, C. A. C.; GOMES, O. M.; BEZERRA, B. G.; BRITO, J. I. B. Eventos Extremos de Precipitação nas Mesorregiões da Paraíba e Suas Relações Com a TSM dos Oceanos Tropicais. *Revista Brasileira de Meteorologia*, SciELO Brasil, v. 29, n. 2, p. 197–208, 2014.
- PALIT, A. K.; POPOVIC, D. Transparent Fuzzy/Neuro-Fuzzy Modelling. *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*, Springer, p. 275–303, 2005.
- PEDROZA, I. C. B. *Estimativa da Climatologia Diária da Precipitação e Investigação de Possíveis Influências das Fases da Lua nas Chuvas no Estado da Paraíba*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, Campina Grande, 2009.
- SNEYERS, R. *Sur L'analyse Statistique des Séries d'Observations*. [S.l.]: Secrétariat de l'Org Météorologique Mondiale, 1975. 415 p.
- SNEYERS, R. On The Statistical Analysis of Series of Observations. n. 194, p. 9–12, 1991.
- VICENTE, P. R. S.; ENILSON, P. C.; MARÍLIA, G. N.; JOÃO, H. B. C. C. Análises da precipitação pluvial no estado da paraíba com base na teoria da entropia. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 7, n. 2, p. 269–274, 2003.
- SILVA VIEIRA, W. L.; ARZABE, C.; SANTANA, G. G. Composição e Distribuição Espaço-Temporal de Anuros no Cariri Paraibano, Nordeste do Brasil. *Oecologia Brasiliensis*, Universidade Federal do Rio de Janeiro (PPGENF-UNIRIO), v. 11, n. 3, p. 383–396, 2007.

Apêndice

```

#Análise de dados de precipitação séries temporais para o município Soledade
#Carregando pacotes necessários para análise require(MASS)

require(epicalc)

require(forecast)

require(tseries)

require(urca)

require(stats)

require(GeneCycle)

require(Kendall) require(adehabitatLT)

# Análise gráfica da série temporal

Soledade=read.table("Precip_Soledade.txt",header=TRUE) head(Soledade) attach(Soledade)

precipsj-ts(Soledade$precip, start = c(1935, 1), end = c(2013, 1), frequency = 1) plot.ts(precips,xlab="T
lines(lowess(time(precips),precips),lwd=1, col=2)

# Aplicando as funções de Autocorrelação (Amostrai e Parcial)

#Função de Autocorrelação acf(precips,xlab="Defasagem",ylab="FAC")

#Obtendo apenas os valores da FAC fac=acf(precips,xlab="Defasagem",ylab="FAC",plot=FALSE);fac

#Função de Autocorrelação Parcial pacf(precips,xlab="Defasagem",ylab="FACP")

#Obtendo apenas os valores da FACP pacf=pacf(precips,xlab="Defasagem",ylab="FACP",plot=FALSE)

#Testes para não-estacionariedade: Dickey-Fuller e Phillips-Perron

#Teste de Dickey-Fuller

adf.test(precips) #Pelo teste de Dickey-Fuller, rejeita-se a hipótese nula #de não-estacionariedade,
ou seja, a série é estacionária

#Teste de Mann-Kendall para verificar tendência na série temporal, sob a #hipótese:

MannKendall(precips) plot(precips) lines(lowess(time(precips),precips),lwd=1, col=2)

```

```

# Série, ACF e PACF após diferença (se necessário diferenciar)

#Ordem da diferença 1 dif_1=diff(precips,lag=1,differences=1) par(mfrow=c(3,1)) plot.ts(dif_1)
acf(dif_1) pacf(dif_1)

# Ajuste do modelo ARMA, tendo em vista que a série é estacionária
modelo1= arima(precips,order=c(1,1,1));modelo1 summary(modelo1)

modelo2=arima(precips,order=c(1,1,2));modelo2 summary(modelo2) modelo3=arima(precips,order=c(2,1,1));
summary(modelo3)

modelo4=arima(precips,order=c(3,1,3));modelo4 summary(modelo4)

modelo5=arima(precips,order=c(3,1,2));modelo5) summary(modelo5)

shapiro.test(modelo5$residuals)

acf(modelo1$resid,ylab="Função de Autocorrelação",main=) acf(modelo1$resid,type="p",main=
)

#através da função auto.arima(), identificamos o modelo que será utilizado # Ter esta
função como uma "saida" caso tenha algum problema

m=auto.arima(precips,ic="aic");m accuracy(m)

S = shapiro.test(m$residuals);S

prev=forecast.Arima(modelo1,h=10);prev

plot.forecast(prev,xlab="Ano",ylab="Precipitação acumulada em Soledade",main=) lines(lowess(time(precips),
col=2)

Box.test(modelo1$residuals,lag=1) #: Análise da estatística de Box-Pierce (e Ljung-Box)

#Analise residual do modelo ajustado aos dados

#Lista do p-valor para teste Ljung-Box para o best model

length(precips)

fun=function (modelo5) n=79 w = matrix(0, nrow =79, ncol = 1) ii = 0 for (i in 1:n) ii = ii +
1 b = try(Box.test(modelo5$residuals,i,type="Ljung-Box")$p.value) w[ii,1]=b dimnames(w) =
list(NULL, c("nível descritivo")) xxx = which(w[, 1] == min(w[, 1]))[1] plot(w,main="Nível des-
critivo para estatística Ljung-Box",xlab=" defasagem",ylab="Nível descritivo",ylim = c (0.00,1.00),col="red")
return(list(res = w, min = w[xxx, ]))

pval=fun(modelo5);pval round(pval$res,2) tdiag(modelo5)

```