



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS DE CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
CURSO DE BACHARELADO EM ESTATÍSTICA**

MARCOS RENATO BASILIO AMORIM

**Modelagem em séries temporais aplicados a dados de
temperatura máxima na cidade de Campina Grande-PB**

CAMPINA GRANDE – PB
Dezembro/2012

MARCOS RENATO BASILIO AMORIM

Modelagem em séries temporais aplicados a dados de temperatura máxima na cidade de Campina Grande-PB

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de 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: Prof. Dr. Ricardo Alves de Olinda

CAMPINA GRANDE – PB
Dezembro/2012

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

A524m

Amorim, Marcos Renato Basilio.

Modelagem em séries temporais aplicados a dados de temperatura máxima na cidade de Campina grande – PB [manuscrito] / Marcos Renato Basílio Amorim. – 2012.

50 f. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2012.

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

1. Climatologia. 2. Modelo ARIMA. 3. Sazonalidade.
I. Título.

21. ed. CDD 551.6

Marcos Renato Basilio Amorim

Modelagem em séries temporais aplicados a dados de temperatura máxima na cidade de Campina Grande-PB

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de 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: 06 / 12 / 2012

Banca Examinadora:

Ricardo Alves de Olinda

Prof. Dr. Ricardo Alves de Olinda
Orientador

Tiago Almeida de Oliveira

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

Ana Patricia Bastos Peixoto

Prof. Msc. Ana Patricia Bastos Peixoto
Universidade Estadual da Paraíba - UEPB

DEDICATÓRIA

A todos que contribuíram para a conclusão deste trabalho, pela dedicação, companheirismo e amizade, DEDICO.

AGRADECIMENTOS

Ao professor Dr. Gustavo Henrique Esteves, coordenador do curso de Bacharelado em Estatística, por seu empenho.

Ao professor Dr. Ricardo Alves de Olinda pelas leituras sugeridas ao longo dessa orientação e pela dedicação.

A todos os professores do Curso de Bacharelado em Estatística da UEPB que contribuíram na minha formação durante o período de quatro anos.

Aos funcionários da UEPB, pela atenção no atendimento quando nos foi necessário.

Aos colegas de classe pelos momentos de amizade e apoio.

Todos vivemos dias difíceis
mas nada disso é em vão (Beco sem saída - Charlie
Brown Jr).

RESUMO

A Região nordeste apresenta acentuada variabilidade inter-anual, particularmente em relação à precipitação e a disponibilidade dos recursos hídricos, com anos extremamente secos, outros chuvosos, isto caracteriza a sua climatologia bem distinta das demais regiões, fato este, decorrente da irregularidade das chuvas e aos baixos índices pluviométricos. Desta forma se faz necessário o uso de modelos de séries temporais no estudo de variáveis climatológicas e neste estudo modela-se a temperatura máxima da cidade de Campina Grande – PB. Modelos do tipo ARIMA sazonal foram ajustados e por meio do critério BIC (Bayesian Information Criterion) foi selecionado o modelo ARIMA (1,0,0)(2,0,0) que foi o mais adequado para a previsão da temperatura máxima para a cidade de Campina Grande – PB. A análise dessa variável permitiu identificar a presença dos componentes de tendência e sazonalidade. O software permitiu o tratamento estatístico dos dados em que foi possível a obtenção de gráficos para análise. Os gráficos obtidos e analisados permitiram uma maior compreensão do que possa vir a acontecer com o comportamento da temperatura máxima da cidade nos próximos anos.

PALAVRAS-CHAVE: Variável climatológica, Modelo ARIMA, Sazonalidade.

A B S T R A C T

The Northeast Region has strong inter-annual variability, particularly in relation to rainfall and availability of water resources, with extremely dry years, other rainy, that characterizes its climatology quite distinct from other regions, this fact, due to irregular rainfall and low rainfall. Thus it is necessary the use of time series models in the study of the weather and this study is modeled maximum temperature in the city of Campina Grande - PB. Seasonal ARIMA type models were fitted using the criterion and BIC (Bayesian Information Criterion) was selected ARIMA (1,0,0) (2,0,0) that was suitable for predicting the maximum temperature for city of Campina Grande - PB. The analysis of this variable allowed us to identify the presence of components of trend and seasonality. The software allowed the statistical treatment of the data it was possible to obtain graphs for analysis. The graphs obtained and analyzed allowed a greater understanding of what might happen to the behavior of the maximum temperature of the city in the coming years.

KEYWORDS: Climatological variable, Arima model, Seasonal.

LISTA DE FIGURAS

- FIGURA - 1. Gráfico da Temperatura máxima da cidade de Campina Grande correspondente aos períodos de 2002 a 2010.....31.
- FIGURA - 2. Gráfico de Box – Plot para a temperatura máxima da cidade de Campina grande entre os meses de janeiro de 2002 a dezembro de 2010.....32.
- FIGURA - 3. Gráfico dos resíduos e p-valor utilizando-se a estatística de Ljung–Box.....33.
- FIGURA - 4. Histograma sob distribuição normal dos valores para a temperatura máxima da cidade de Campina Grande no período de janeiro de 2002 a dezembro de 2010.....34.
- FIGURA - 5. Gráfico Q - Qplot para os quantis teóricos versus quantis amostrais.....34.
- FIGURA - 6. Gráfico da Distribuição normal para os quantis teóricos versus quantis amostrais.....35.
- FIGURA -7. Gráfico de previsão do comportamento da temperatura máxima para a cidade de Campina Grande para 2011.....36.

SUMÁRIO

1	Introdução.....	12
2	Fundamentação Teórica.....	14
2.1	Noções Básicas.....	14
2.2	Fundamentos Probabilísticos.....	15
2.3	Processos Estacionários.....	16
2.3.1	Estacionariedade de Segunda Ordem.....	16
2.3.2	Estacionariedade Forte.....	19
2.4	Modelos de Séries Temporais.....	19
2.4.1	Modelo de Regressão.....	19
2.4.2	Modelos Lineares.....	20
2.4.3	Modelos Sazonais.....	22
2.4.4	Modelos Arima.....	23
2.4.5	Modelos Arma.....	24
2.4.6	Método dos Momentos.....	25
2.4.7	Método de Máxima Verossimilhança.....	26
2.4.8	Variância dos Estimadores.....	27
2.4.9	Diagnóstico dos Modelos.....	28
2.5	Estimação dos Parâmetros.....	28
2.6	Critério de Informação Bayesiano (BIC).....	29
3	Previsão.....	29
4	Aplicação.....	30
4.1	Análise dos Resíduos.....	33
4.2	Estudo da Previsão.....	35
5	Conclusão.....	37
	Referências Bibliográficas.....	38

1 - Introdução

A região Nordeste apresenta uma climatologia com características bem diferentes das demais regiões do Brasil, este fato deve-se as mudanças climáticas e, as chuvas estarem ocorrendo com menor intensidade. O aumento das temperaturas ocasiona a escassez da chuva, devido ao aquecimento global como também o aquecimento local, provocado na maioria das vezes pelo desmatamento e também pelo processo de urbanização das cidades. Walker (1928) documentou uma notável coincidência do aquecimento anômalo das águas superficiais no Pacífico Equatorial e as secas do nordeste do Brasil.

Do ponto de vista climático, a região Nordeste é considerada semi-árida por apresentar substanciais variações temporais e espaciais da precipitação pluviométrica, e elevadas temperaturas ao longo do ano (AZEVEDO; SILVA; RODRIGUES, 1998). Apesar das elevadas temperaturas registradas durante todo o ano, as amplitudes térmicas máximas são em torno de 6°C.

Segundo Oyama e Nobre (2003) o impacto das mudanças climáticas sobre a estabilidade dos biomas predominantes no Brasil indica maior vulnerabilidade do bioma Caatinga num cenário de aumento das temperaturas globais, dessa forma a Região Nordeste enfrenta diversos problemas e é preciso estar sempre alerta a respeito desses mesmos, tendo em vista que, a vulnerabilidade do bioma Caatinga apresenta efeitos consideráveis nas mudanças climáticas e ocasiona um forte fator de pressão que pode acarretar na desertificação de diversas áreas.

É necessário conhecer as condições do ambiente climático de forma que os diversos setores das atividades humanas possam se desenvolver sejam esses setores relacionados às atividades urbanas, atividades agrícolas, de planejamento ambiental, dentre outras. No estudo da climatologia é preciso fazer uma análise do clima juntamente com o espaço geográfico. Em vista disto, muitos estudiosos em climatologia e meteorologia têm dedicado seus conhecimentos em estudos voltados a superfície de um determinado lugar, porém eles também levam em consideração episódios ao longo do tempo, ou seja, estudo de séries temporais, e realizam análises voltadas a uma perspectiva regional.

Segundo Maack (1981) as condições do clima de uma determinada região não evidenciam apenas os valores meteorológicos, elas também permitem uma classificação climática por meio da obtenção de elementos importantes para análises estatísticas.

Conseqüentemente, catástrofes como veranicos, aumento de temperatura, maremotos entre outras podem ser previstas e assim as pessoas podem ser retiradas das áreas de risco. Também é importante ressaltar que pode ocorrer o desaparecimento de espécies de microrganismos, e até plantas e animais sensíveis as mudanças de temperatura, e ainda a seleção de indivíduos patogênicos mais resistentes.

É necessário à determinação de índices de tendência climática para a compreensão das variáveis objetivando-se realizar um estudo em uma determinada região ou um local específico. De acordo com as características de cada ambiente, torna-se importante o monitoramento da temperatura em cada um dos locais que se deseja estudar, que pode ser uma região, estado ou cidade.

Por se localizar dentro da faixa equatorial, o Estado da Paraíba é submetido à alta radiação solar durante um longo período. Essa condição determina para este estado um clima quente, como também uma temperatura média anual de 26°C, além de pouca variação intra-anual da temperatura altamente influenciada pelo relevo. Utilizando-se o software R, pôde-se fazer o tratamento dos dados da variável em estudo por meio de testes estatísticos e também especificar qual o modelo mais adequado à situação abordada. Tendo em vista estes fatos, este trabalho tem por objetivo investigar o mecanismo gerador de séries temporais por meio de dados de temperatura máxima do município de Campina Grande; descrever o comportamento da série e verificar a existência de tendência e variações sazonais, aplicar a modelagem de séries temporais para a variável temperatura máxima da cidade de Campina Grande entre os anos de 2002 a 2010; verificar qual modelo é o mais adequado para a solução do problema abordado; descrever um modelo de série temporal que melhor se ajusta a variável temperatura máxima e prever valores de temperatura máxima para a cidade de Campina Grande referente ao ano de 2011.

2 - Fundamentação Teórica

Aborda-se nesse trabalho um estudo de séries temporais, utilizando-se de técnicas e testes estatísticos, para a descrição e entendimento do mecanismo de série e aplicação a dados de temperatura máxima para a cidade em estudo.

2.1 - Noções Básicas

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo (EHLERS, 2007). A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e o interesse é analisar e modelar esta dependência. Enquanto em modelos de regressão, por exemplo, a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é crucial. Vale notar também que o tempo pode ser substituído por outra variável como espaço, profundidade, etc.

Como a maior parte dos procedimentos estatísticos foram desenvolvidos para analisar observações independentes, o estudo de séries temporais requer o uso de técnicas específicas. Segundo Costa (2009) dados de séries temporais surgem em vários campos do conhecimento como Economia (preços diários de ações, taxa mensal de desemprego, produção industrial), Medicina (eletrocardiograma, eletroencefalograma), Epidemiologia (número mensal de novos casos de meningite), Meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento).

A classe de fenômenos cujo processo observacional e consequente quantificação numérica geram uma sequência de dados distribuídos no tempo é denominada *série temporal* (SOUZA, 1989). A natureza de uma série temporal e a estrutura de seu mecanismo gerador está relacionada com o intervalo de ocorrência das observações no tempo (ANDERSON, 1971). Caso o levantamento das observações da série possa ser feito a qualquer momento do tempo, a série temporal é dita *contínua*, sendo denotada por $X(t)$ (GRANGER, 1977). Entretanto, de acordo com Granger, (1977) e Nelson, (1973), na maioria das séries, as observações são tomadas em intervalos de tempo discretos e equidistantes. Uma série temporal *discreta* pode ser representada por $(X_T \in \mathbb{R} / t = 0, 1, \dots, n)$, sendo que cada observação discreta x_t está associada a um instante de

tempo distinto, existindo uma relação de dependência serial entre essas observações (SOUZA, 1989).

A base do tratamento requerido pelas séries de tempo está assentada na possibilidade de serem tiradas conclusões sobre o comportamento passado da variável e que poderão ser úteis para proporcionar informações sobre o seu comportamento futuro provável (FISCHER, 1982). Portanto, a análise de um modelo de séries temporais significa um estudo que busca obter as características comportamentais sistemáticas da série, capazes de propiciar a construção de um modelo que descreva os movimentos passados de uma variável, com o que pode-se prever os futuros movimentos da mesma (FISCHER, 1982). De uma maneira formal, pode-se dizer que uma série temporal é uma realização de um processo estocástico (BARROS, 2003).

No estudo de uma série temporal existe interesse em três aspectos:

- i) Análise e modelagem da série temporal – descrever a série, verificar suas características mais relevantes e suas possíveis relações com outras séries;
- ii) A previsão da série no instante $T+k$ será denotada por Z_{t+k} . O número de instantes à frente para o qual é feita a previsão (neste caso, k) é chamado de horizonte de previsão. Por exemplo, a previsão de Z_{t+1} é denotada por Z_{t+k} (BARROS, 2003).
- iii) Previsão da série temporal – a partir de valores passados da série (e também de outras séries) encontrar boas previsões (de curto prazo);

2.2 - Fundamentos probabilísticos

Definição: Seja T um conjunto arbitrário. Um processo estocástico é uma família $\{Y(t), t \in T\}$ tal que, $\forall t \in T$, $Y(t)$ é uma variável aleatória. Uma série temporal é uma particular realização do processo estocástico. O conjunto de valores $\{Y(t), t \in T\}$ é chamado de espaço de estados e os valores $Y(t)$ são chamados de estados no qual podem representar valores discretos ou contínuos.

2.3 - Processos estacionários

Uma importante classe de processos estocásticos são os chamados processos estacionários. Uma série temporal é dita estritamente estacionária se a distribuição de probabilidade conjunta de $X(t_1), \dots, X(t_k)$ é a mesma de $X(t_1 + \tau), \dots, X(t_k + \tau)$. 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 estrita implica que a distribuição de $X(t)$ é a mesma para todo t de modo que, se os dois primeiros momentos forem finitos, tem-se que,

$$\mu(t) = \mu \text{ e } \sigma^2(t) = \sigma^2,$$

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

Como a quantidade de parâmetros é usualmente maior que o número de observações, são necessárias hipóteses simplificadoras. A mais comum em séries temporais é a de estacionariedade. Basicamente, isso significa que o comportamento da série não se altera com o passar do tempo, ou seja, a média não muda ao caminhar no tempo.

Segundo Fava (2000) a análise das estatísticas básicas das séries estacionárias permite separar a estacionariedade em dois grupos:

- i) estacionariedade no amplo senso: médias, variâncias e covariâncias constantes 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.

2.3.1 – Estacionariedade de segunda ordem

Segundo Gujarati (2000) um processo é dito fracamente estacionário (ou estacionário de segunda ordem) se as condições a seguir são satisfeitas (média constante, variância constante e covariância que só depende do “lag” ou retardamento). Essas condições referem-se apenas aos dois primeiros momentos da distribuição de probabilidade dos Y_t 's, o que explica a terminologia “processo estacionário de segunda

ordem”. A definição de estacionariedade mais geral envolve momentos de todas as ordens e é muito mais complicada de verificar que a estacionariedade de segunda ordem. Se os Y_t 's são conjuntamente gaussianos, as duas condições (estacionariedade estrita e estacionariedade de segunda ordem) são equivalentes. Na prática, aceita-se que as séries observadas sejam séries fracamente estacionárias, situações nas quais garante-se apenas médias e variâncias invariantes no tempo.

Segundo Jenkins (1968), Wilks (1995) e Chatfield (1996) uma série estacionária é aquela que está em *equilíbrio estatístico*, no sentido que contém *nenhuma tendência*, enquanto que uma série não-estacionária é aquela cujas propriedades mudam com o tempo. Na prática, as séries são usualmente de 3 tipos: aquelas que *exibem propriedades de estacionariedade em longo período*, como, por exemplo, análises de estacionariedade das séries de vazões médias anuais naturais de 32 usinas hidrelétricas da Bacia do Rio Paraná, as séries utilizadas são provenientes do banco de dados de vazões naturais utilizados pelo setor elétrico. Aquelas que possuem uma razoável *estacionariedade em períodos curtos*, por exemplo, medidas de turbulência na atmosfera; e séries que são *obviamente não estacionárias*, no sentido que suas propriedades estão continuamente mudando com o tempo. Exemplos óbvios de não-estacionariedade: temperatura em altas e médias latitudes, ventos (apresentam ciclos diurnos e anuais).

Segundo Carvalho e Sedyama (2005) o termo estacionariedade é interpretado como *fraca estacionariedade* ou *estacionariedade da covariância*. Neste sentido, estacionariedade implica que a média e a função de auto-correlação de uma série de dados não muda com o tempo. Diferentes pedaços de uma série de dados estacionária (por exemplo, os dados observados hoje e no futuro) podem ser considerados como *tendo uma mesma média e variância*. Além disso, uma correlação entre variáveis em uma série estacionária é determinada apenas pela sua separação no tempo (ou seja, pelo seu “lag”) e não pela sua absoluta posição no tempo. Isso significa que valores individuais em distintas porções da série podem ser diferentes embora essas duas porções da série se pareçam *a estacionariedade de covariância* é uma suposição menos restritiva que estacionariedade restrita, a qual implica que a distribuição total das variáveis na série não muda com o tempo.

A maior parte dos métodos que trata com não-estacionariedade de séries temporais está baseada em técnicas para remover ou filtrar a parte não-estacionária, deixando apenas a parte que pode ser tratada como estacionária. Em climatologia, utiliza-se esse tipo de técnica quando deseja-se conhecer o comportamento das anomalias de uma

determinada variável. Existem duas aproximações para tratar-se com séries não-estacionárias. Ambas objetivam processar os dados de forma que permitam que uma subsequente estacionaridade seja assumida. Por exemplo: subtração de uma função periódica média a partir dos dados sujeitos a um ciclo anual produziria uma nova série transformada com média constante igual à zero. A fim de produzir uma série com média e variância constante, seria necessário transformar essas anomalias em anomalias normalizadas, definidas por:

$$Z = \frac{x - \bar{x}}{s_x}, \quad (1)$$

em que Z é a anomalia padronizada, calculada simplesmente pela subtração da média da amostra (que no caso seria igual a zero após remoção do ciclo anual) e dividindo pelo respectivo desvio padrão s_x , o qual varia, denota-se x como sendo um valor amostral e \bar{x} a média. Por exemplo, não apenas as temperaturas tendem a ser mais frias durante o inverno, mas sua variabilidade tende a ser mais alta em regiões de latitudes médias.

Segundo Jenkins (1968), Wilks (1995) e Chatfield (1996) uma aproximação possível para transformar séries de temperaturas mensais em uma série (aproximadamente) estacionária seria calcular as 12 médias mensais e os 12 desvios-padrão e então aplicar a equação a equação 1, usando diferentes médias e desvios-padrão para o mês do calendário apropriado. Uma alternativa seria a estratificação dos dados. Isto é, poder-se-ia conduzir análises separadas de subconjuntos dos dados que são curtas o suficiente para serem consideradas aproximadamente estacionárias. Por exemplo, poder-se-ia analisar observações diárias para todos os dados disponíveis de janeiro para uma dada localização, assumindo-se que cada conjunto de 31 dias de dados é uma amostra que sofreu os mesmos processos físicos. Não necessariamente os processos seriam os mesmos para julho, ou fevereiro.

Assim, o objetivo básico da análise estatística das séries temporais é buscar, a partir da realização amostral do processo (isto é, a série $Z_1, Z_2, Z_3, \dots, Z_t$), descrever as características de sua aleatoriedade, com o objetivo de proporcionar os instrumentos para a inferência sobre as probabilidades associadas com o conjunto de valores futuros alternativos da série. Ao se conseguir especificar numericamente como é a função de probabilidade da série, torna-se viável inferir a probabilidade de um ou outro futuro valor ocorrer (FISCHER, 1982).

Para Fischer (1982) estudar modelos de séries temporais significa buscar obter meios capazes de inferir as características de seu processo gerador, bem como buscar

modelos estocásticos que sejam capazes de descrever as situações particulares que ocorrem na realidade.

De acordo com Barros (2003) um processo estocástico é dito ergódico se uma única realização do processo é o suficiente para caracterizá-lo. Na análise de séries temporais existe apenas uma realização do processo disponível e, portanto é preciso supor que o processo subjacente é ergódico, pois irá usar apenas uma de suas realizações para caracterizá-lo.

2.3.2- Estacionariedade forte

Na estacionariedade forte ou estrita a forma da distribuição conjunta do processo permanece sem variação mediante uma translação temporal. Como na prática é muito difícil especificar a distribuição conjunta de um processo estocástico, há a versão mais fraca, na qual somente alguns momentos do processo permanecem inalterados no tempo (MEDEIROS, 2005).

2.4- Modelos de séries temporais

Segundo Morettin e Toloí, (2004) os modelos podem ser divididos em duas classes :

i) paramétricos – número finito de parâmetros. análise é feita no domínio do tempo.

ii) não paramétricos – número infinito de parâmetros. análise é feita no domínio da frequência.

2.4.1- Modelos de regressão

Conforme Demétrio e Zocchi (2006) a teoria de regressão teve origem no século XIX com Francis Galton. Em um de seus trabalhos ele estudou a relação entre a altura dos pais e dos filhos, procurando-se saber como a altura do pai influenciava a altura do filho. Galton observou que se o pai fosse muito alto ou muito baixo, o filho teria uma altura tendendo à média. Por isso, ele chamou de regressão, ou seja, existe uma tendência de os dados regredirem à média. Segundo Demétrio e Zocchi (2006), a utilização de modelos de regressão pode ter por objetivos:

i) Predição: Uma vez que se espera que uma parte (que se deseja que seja a maior) da variação de Y é explicada pelas variáveis X , então, pode-se utilizar o modelo para obter valores de Y correspondentes a valores de X que não estavam entre os dados. Esse processo denomina-se predição e, em geral, são usados valores X que não estão dentro do intervalo de variação estudado. A utilização de valores fora desse intervalo recebe o nome de extrapolação e, deve ser usada com muito cuidado, pois o modelo adotado pode não ser correto fora do intervalo estudado. Este, talvez, seja o uso mais comum dos modelos de regressão.

ii) Seleção de variáveis: Frequentemente, não se tem ideia de quais são as variáveis que afetam significativamente a variação de Y . Para responder a esse tipo de questão, conduzem-se estudos em que está presente um grande número de variáveis. A análise de regressão pode auxiliar no processo de seleção de variáveis, eliminando aquelas cuja contribuição não seja importante.

iii) Estimação de parâmetros: Dado um modelo e um conjunto de dados (amostra) referente às variáveis respostas e preditoras, estimar parâmetros, ou ainda, ajustar o modelo aos dados significa obter valores (estimativas) para os parâmetros, por algum processo, tendo por base o modelo e os dados observados. Em alguns casos, o valor do coeficiente tem valor por si só. Como exemplo, pode-se citar o estudo de estabilidade de variedades. Em outros casos, o interesse está em uma função dos parâmetros. Como por exemplo, pode-se citar o cálculo de doses letais.

iv) Inferência: O ajuste de um modelo de regressão tem em geral, por objetivos básicos, além de estimar os parâmetros, realizar inferências sobre eles, tais como testes de hipóteses e intervalos de confiança.

2.4.2- Modelos lineares

Conforme Costa (2003) o modelo linear clássico utilizado na análise é definido por,

$$Y = X\beta + \varepsilon \quad (2)$$

em que, Y representa o vetor de dimensões $n \times 1$, de dados observados; X , de dimensões $n \times p$, é a matriz de delineamento; β , de dimensões $p \times 1$, é um vetor de parâmetros

desconhecidos de efeitos fixos e ε é o vetor de dimensões $n \times 1$, erros aleatórios. O objetivo do modelo linear clássico é modelar a média de Y , usando-se o vetor de parâmetros de efeitos fixos β . Os componentes do vetor ε são variáveis aleatórias independentes e identicamente distribuídas com média 0 e variância σ^2 .

A Equação (2) é denominada componente sistemático do modelo. Considerando-se o modelo de regressão simples, β_0 e β_1 são parâmetros desconhecidos a serem estimados. Além da componente sistemática, o modelo é composto de uma parte não determinística, devido a fatores não observáveis, denominada componente aleatória e denotada por uma variável aleatória. Dessa forma, o modelo completo fica estabelecido pela seguinte equação,

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (3)$$

Considera-se inicialmente o caso mais simples em que têm-se uma variável resposta Y e uma única variável independente X , admitindo-se que a relação entre a média condicional de Y dado X , ou seja, $\mu(x) = E[Y | X = x]$, seja da forma $\mu = \mu(x) = \beta_0 + \beta_1 x$.

Segundo Júnior (2005) o modelo de regressão normal linear simples é obtido a partir das seguintes suposições assumidas para a componente aleatória:

$$E(\varepsilon) = 0$$

$$\text{Var}(\varepsilon) = \sigma^2$$

$$\varepsilon \sim N(0, \sigma^2)$$

No modelo de regressão normal linear múltiplo, no qual utiliza-se mais de uma variável independente, considera-se que a média condicional da variável resposta Y pode ser expressa em função das variáveis explicativas X_1, X_2, \dots, X_p por meio da relação linear.

$$\mu = E[Y | X_1 = x_1, \dots, X_p = x_p]$$

$$= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

O componente aleatório do modelo também é representado por uma variável aleatória ε , que segue uma distribuição normal com média zero e variância σ^2 . Assim, para cada observação do tipo $(y_i, x_{1i}, x_2, \dots, x_{pi})$, tem-se o seguinte modelo $y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i$; para $i = 1, 2, \dots, n$; $p < n$, em que $\varepsilon_i \sim N(0; \sigma^2)$, para todo i .

A seguir mostra-se um tipo de modelo adequado a situação em estudo, diferente do que é citado acima, esse modelo permite um estudo mais característico da situação abordada neste trabalho, explorando o comportamento da temperatura máxima por meio

de séries temporais no período de 2002 a 2010 visando à elaboração de previsões sazonais, analisando as séries e os mecanismos (processos) que a geram, propondo um modelo estocástico de previsão que combine resultados da variável meteorológica utilizada.

2.4.3 - Modelos sazonais

Segundo Ehlers (2007) uma forma bastante simples de eliminar o efeito sazonal é simplesmente tomar médias sazonais. Por exemplo, em dados mensais com sazonalidade anual, as médias anuais estarão livres do efeito sazonal. Embora este procedimento esteja correto, muitos dados serão perdidos e ao invés disto podem-se recorrer mais uma vez às médias móveis. Um comportamento periódico ou cíclico pode ser encontrado em várias séries temporais. É importante que se consiga descrever o padrão sazonal da série por meio de quantidades que possam ser estimadas incluindo-se assim este padrão na função de previsão.

Segundo Montgomery, Fohson e Gardiner (1990) grande parte das séries temporais, principalmente na indústria, apresenta variações sazonais. Isto ocorre quando a série exibe uma característica periódica que se repete a cada s intervalos de tempo. Por exemplo, em séries compostas por observações mensais e sazonalidade anual, s é igual a 12.

Com uma abordagem simples e didática, Wonnacott e Wonnacott (1990) discutem os conceitos relacionados à sazonalidade. Segundo os autores, a análise da série dessazonalizada é fundamental, uma vez que a aplicação de uma regressão simples da variável dependente sobre o tempo acusa uma substancial tendenciosidade. Para tanto, deve-se incluir em um modelo de regressão tanto a tendência como o efeito sazonal, a fim de estimar seus efeitos em separado. A tendência permite uma aproximação real aos dados, o efeito sazonal é interpretado como um movimento regular da série dentro de um ano.

2.4.4 - Modelos Arima (Autoregressivos integrados e de médias móveis)

Ao longo dos anos diversas ferramentas para modelagem e previsão de séries temporais têm sido desenvolvidas, mas, no entanto, a maioria destes métodos baseia-se em hipóteses fundamentais que são: a série adapta-se a um modelo linear;

estacionariedade ou redução (por meio de diferenciação) para a estacionariedade, homocedasticidade e gaussianidade.

Os modelos ARIMA ou Box e Jenkins são excelentes modelos de previsão de curto prazo (GRANGER; NEWBOLD, 1977). Resultados de análises com esses modelos mostram que os melhores resultados (previsões) são obtidos com informação entre 5 a 10 anos, particularmente na presença de sazonalidade. Como já visto, a importância do processo observado ser estacionário é a possibilidade de fixar parâmetros do modelo válido para previsão do futuro a partir do passado. Assim, como primeiro passo para essa modelagem são realizados procedimentos para a remoção da não-estacionariedade.

A metodologia de Box e Jenkins é bastante utilizada em análise de modelos paramétricos por meio da abordagem de Box e Jenkins (1970), basicamente, esta metodologia ajusta modelos auto regressivos integrados de médias móveis, ARIMA, (p,d,q) a um conjunto de dados. Como casos particulares dos modelos ARIMA têm-se os processos AR (Autoregressivo) e MA (Médias móveis ou moving average). Existem também modelos ARIMA multivariados, geralmente denominados modelos de função de transferência (BARROS, 2003).

A metodologia de Box e Jenkins para a previsão se baseia no ajuste de modelos tentativos denominados ARIMA a séries temporais de valores observados de forma que a diferença entre os valores gerados pelos modelos e os valores observados resulte em séries de resíduos de comportamento aleatório em torno de zero (BOWERMAN; O'CONNELL, 1987).

Segundo Dickey e Fuller, (1979) os modelos ARIMA são capazes de descrever os processos de geração de uma variedade de séries temporais para os previsores (que correspondem aos filtros) sem precisar levar em conta as relações econômicas, por exemplo, que geraram as séries.

Segundo a sistemática da metodologia de Box e Jenkins os modelos ARIMA descrevem tanto o comportamento estacionário como o não estacionário. Dessa forma, pode-se afirmar que é uma metodologia de modelagem flexível em que as previsões com base nesses modelos são feitas a partir dos valores correntes e passados dessas séries (PANKRATZ, 1983).

Segundo Dickey e Fuller (1979), Dickey e Fuller (1981), Fuller (1976) e Enders (2004) a construção dos modelos ARIMA (p,d,q) baseia-se em um ciclo iterativo no qual a escolha da estrutura do modelo é baseada nos próprios dados. Para a construção de modelos ARIMA é necessário passar pelos estágios do ciclo iterativo, especificação de

uma classe geral de modelos, identificação de um modelo, com base na análise de auto correlações, auto correlações parciais e outros critérios, estimação do modelo, na qual os parâmetros do modelo identificado são estimados e por último, diagnóstico do modelo ajustado, por meio da análise de resíduo, entretanto, se o modelo não for adequado volta-se a fase de identificação.

2.4.5- Modelos Arma (Auto regressivos e de médias móveis)

Existe uma equivalência entre os modelos ARIMA e os modelos ARMA. Esses últimos são ajustados a séries já estacionárias transformadas pelo método das diferenças de ordem, ou seja, cujas séries originais são séries não-estacionárias homogêneas (assim denominadas por ter sido possível obter a estacionariedade com um número finito de diferenciações).

Os modelos ARIMA ainda podem conter uma componente sazonal. Segundo Morettin (2002), se a correlação for significativa em *lag's* (defasagem) de baixa ordem, tem-se uma sazonalidade pura. A equação abaixo representa um modelo auto regressivo de ordem p com uma sazonalidade no período S , isto é, um AR(P)S (modelo autoregressivo) ou um SAR(P)S (modelo sarima).

$$x_t = \phi_1 X_{t-1S} + \phi_2 X_{t-2S} + \dots + \phi_p X_{t-pS} + e_t \quad (4)$$

Esse modelo também pode ser expresso abreviadamente por,

$$\Phi(B^S)X_t + e_t \quad (5)$$

Quando existe correlação significativa em *lag's* sazonais, existe uma componente sazonal multiplicativa.

A equação abaixo representa um modelo auto regressivo de ordem p com uma sazonalidade no período S , isto é, um AR(p)x(P)S.

$$\phi(B)\Phi(B^S)x_t = e_t \quad (6)$$

Segundo Pindyck, Rubinfel (1991) a diferença básica entre a regressão clássica e os modelos de séries temporais é que nos modelos de séries temporais ARIMA não se pode assumir independência entre observações. Ao contrário, os modelos auto regressivos e de médias móveis vão modelar o grau de auto correlação entre desvios e

observações defasadas. Em suma, de forma geral, quando faz-se referência a modelos ARIMA esses modelos estão sendo ajustados à série original. Já ao fazer-se referência a modelos ARMA, considera-se que a série é uma série diferenciada. Supondo Y_t a série já diferenciada, os modelos ARMA em sua forma geral se escrevem da seguinte forma:

$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + L + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, \quad (7)$$

em que, ϕ_1, \dots, ϕ_p e $\theta_1, \dots, \theta_q$ são parâmetros regressores.

2.4.6- Método dos momentos

Conforme Silva et al. (2007) o método dos momentos é um dos métodos de estimação mais simples e antigo. Este método consiste em substituir, nas equações que relacionam as auto correlações e os parâmetros do modelo, os momentos teóricos (média, variância e auto correlação) pelos respectivos momentos amostrais e, em seguida, resolver as equações resultantes.

Para o modelo AR(p), o estimador ϕ do vetor parâmetro pelo método dos momentos é definido por

$$\hat{\phi}_{MM} = (\hat{\phi}_{1,MM}, \dots, \hat{\phi}_{p,MM})^t = R_p^{-1} r_p, \quad (8)$$

em que,

$$R_p = \begin{pmatrix} 1 & r_1 & r_2 & \dots & r_{p-1} \\ r_1 & 1 & r_1 & \dots & r_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \dots & 1 \end{pmatrix} \text{ e } r_p = (r_1, r_2, \dots, r_p)^t.$$

Utilizando-se $\hat{\phi}_{MM}$ pode-se também estimar $\hat{\sigma}_{MM}^2$, por meio do método dos momentos, obtendo-se,

$$\begin{aligned} \hat{\sigma}_{MM}^2 &= c_0(1 - \hat{\phi}_{1,MMr_1} - \dots - \hat{\phi}_{1,MMr_p}) \\ &= c_0(1 - \mathbf{r}_p^t \hat{\phi}_{MM}) = c_0(1 - \mathbf{r}_p^t - R_p^{-1} r_p). \end{aligned}$$

Em particular, para $p = 1$, tem-se que,

$$\hat{\phi}_{MM} = \hat{\phi}_{1,MM} = r_1 \text{ e } \hat{\sigma}_{MM}^2 = c_0(1 - r_1^2). \quad (9)$$

Para o modelo MA(q), o estimador de θ , utilizando-se o método dos momentos é obtido resolvendo-se as equações,

$$r_k = \frac{-\hat{\theta}_{k,MM} + \hat{\theta}_{1,MM}\hat{\theta}_{k+1,MM} + \dots + \hat{\theta}_{q-k,MM}\hat{\theta}_{q,MM}}{1 + \hat{\theta}_{1,MM}^2 + \hat{\theta}_{2,MM}^2 + \dots + \hat{\theta}_{q,MM}^2} \quad k = 1, 2, \dots, q.$$

A variância residual estimada por meio do método dos momentos é,

$$\hat{\sigma}_{MM}^2 = c_0 / (1 + \hat{\theta}_{1,MM}^2 + \hat{\theta}_{2,MM}^2 + \dots + \hat{\theta}_{q,MM}^2).$$

Em particular, para $q = 1$, tem-se que,

$$r_1 = \frac{-\hat{\theta}_{1,MM}}{1 + \hat{\theta}_{1,MM}^2} \quad \text{e} \quad r_{MM}^2 = \frac{c_0}{1 + \hat{\theta}_{1,MM}^2}$$

Segundo Areli (2007) o modelo ARMA (p,q) , o vetor de parâmetros $\hat{\phi}$ e $\hat{\theta}$, são estimados, por meio do método dos momentos, em duas etapas:

i) Estimação de ϕ , por meio da solução $\hat{\phi}_{MM} = (\phi_{1,MM}, \dots, \phi_{p,MM})$ da seguinte equação,

$$r_k = \phi_{1,MM}r_{k-1} + \dots + \phi_{p,MM}r_{k-p} \quad k = q+1, \dots, q+p.$$

ii) Estimação de θ , por meio da solução $\hat{\theta}_{MM} = (\theta_{1,MM}, \dots, \theta_{q,MM})$, utilizando-se as auto covariâncias amostrais c_k e os estimadores $\hat{\phi}_{MM}$ obtidos na etapa anterior.

Em particular, para $p = q = 1$, obtêm-se

$$r_2 = \hat{\phi}_{1,MM} r_1$$

$$r_1 = \frac{c_1}{c_0} = (1 - \hat{\phi}_{1,MM}\hat{\theta}_{1,MM})(\hat{\phi}_{1,MM} - \hat{\theta}_{1,MM}) / (1 + \hat{\theta}_{1,MM}^2 - 2\hat{\phi}_{1,MM}\hat{\theta}_{1,MM}).$$

2.4.7 - Método de máxima verossimilhança

Denota-se por $f(\mathbf{z}|\mathfrak{G})$ a *função densidade (ou de probabilidade) conjunta* de $Z = (Z_1, Z_2, \dots, Z_N)$. Fixado \mathfrak{G} , a função $f(\mathbf{z}|\mathfrak{G})$ associa um determinado valor a cada conjunto de observações \mathbf{z} . Na sequência, quando fixa-se \mathbf{z} e varia-se \mathfrak{G} , obtem-se a *função de verossimilhança*, denotada por $f_X(x_n|\mathfrak{G})$. Segundo Box e Jenkins (1970) essa função é de

grande importância na teoria de estimação, devido aos “princípios de verossimilhança”, que diz o seguinte: dado que o modelo adotado é correto, toda a informação sobre \mathcal{G} presente na amostra está contida na função de verossimilhança; os outros aspectos dos dados são irrelevantes (BOX; JENKINS, 1976). Em geral, é conveniente trabalhar com o logaritmo natural de $f_X(x_n|\mathcal{G})$ que é denominado de *função de logaritmo de verossimilhança*.

Os valores dos parâmetros que maximizam a função de verossimilhança (ou equivalentemente, a função de logaritmo de verossimilhança) são chamados *estimadores de máxima verossimilhança* (EMV) (SARNAGLIA, 2010).

Em termos formais, a função de verossimilhança é dada por,

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^N f_y(y_i; \theta_1, \theta_2, \dots, \theta_k) \quad (10)$$

2.4.8 - Variância dos estimadores

A precisão dos estimadores encontrados deve ser avaliada por meio da construção de intervalos de confiança para os parâmetros. Considerando-se o vetor de parâmetros $\eta = (\phi, \theta)$, cuja ordem é $p + q$. Supondo n suficientemente grande, os estimadores de máxima verossimilhança têm uma distribuição assintótica normal, isto é,

$$\eta \xrightarrow{d} N_{p+q}(\eta, V)$$

$$V = 2\sigma_a^2 \begin{pmatrix} \frac{\partial^2 s_\eta}{\partial \eta_1^2} & \dots & \frac{\partial^2 s_\eta}{\partial \eta_1 \partial \eta_k} \\ \frac{\partial^2 s_\eta}{\partial \eta_k \partial \eta_1} & \dots & \frac{\partial^2 s_\eta}{\partial \eta_1^2} \end{pmatrix}.$$

Além disso, o estimador de máxima verossimilhança de σ_1^2 é definido por

$$\hat{\sigma}_1^2 = \frac{S^{(\hat{\eta})}}{\eta}.$$

e, para n suficientemente grande, $\hat{\sigma}_a^2$ e $\hat{\eta}$ são não correlacionados (MORETTIN; TOLOI, 2004).

As estimativas das variâncias dos estimadores e covariância entre os estimadores são obtidas substituindo-se σ_a^2 na matriz anterior por $\hat{\sigma}_a^2$ e calculando-se as derivadas parciais,

$$\frac{\partial^2 s_\eta}{\partial \eta_i \partial \eta_j},$$

numericamente. Utilizando-se as estimativas das variâncias, pode-se obter intervalos de confiança para os parâmetros $\eta_{i,j} = 1, 2, \dots, p + q$.

2.4.9 - Diagnóstico dos modelos

Conforme Silva (2007) após identificar o modelo e estimar seus parâmetros, devemos verificar se ele representa, satisfatoriamente, os dados observados. Esta verificação pode ser feita por meio de uma técnica chamada *superajustamento*, a qual consiste em estimar um modelo com parâmetros extras e examinar, primeiramente, se eles são significativos e, em seguida, se a inclusão dos mesmos diminui significativamente a variância residual. Para tanto, é necessário analisar os resíduos do modelo ajustado. Seja o modelo ajustado $\phi(B)W_t = \theta(B)a_t$, com $W_t = \Delta^d Z_t$, for verdadeiro, então os “erros verdadeiros” $a_t = \theta^{-1}(B)\phi(B)W_t$, serão um ruído branco (MORETTIN; TOLOI, 2004).

2.5 - Estimação dos parâmetros

Após identificar um modelo provisório a ser ajustado à série temporal, deve-se obter estimativas eficientes para os seus parâmetros. Denota-se por $\mathfrak{P} = (\phi, \theta, \sigma_a^2)$ o vetor com os $p+q+1$ parâmetros de um modelo ARIMA (p,d,q) , em que $\phi = (\phi_1, \dots, \phi_p)$ e $\theta = (\theta_1, \dots, \theta_q)$. A seguinte notação também será útil $\eta = (\phi, \theta)$. Suponha que a série original $\mathbf{Z} = (Z_1, Z_2, \dots, Z_t)$ tenha sido gerada por um processo ARIMA (p,d,q) . A partir daí, considerando-se d diferenças pode-se gerar uma série W_t estacionária: $W = (W_1, W_2, \dots, W_n)$, em que $W_t = \Delta^d Z_t$ e $n = N-d$. Com isso, o problema de estimar os parâmetros do modelo ARIMA é equivalente a estimar os parâmetros do modelo ARMA (p,q)

estacionário e invertível, representado por: $\mathbf{a}_t = \tilde{\mathbf{w}}_t - \phi_1 \tilde{\mathbf{w}}_{t-1} - \phi_2 \tilde{\mathbf{w}}_{t-2} - \dots - \phi_p \tilde{\mathbf{w}}_{t-p} + \theta_1 \mathbf{a}_{t-1} + \theta_2 \mathbf{a}_{t-2} + \dots + \theta_q \mathbf{a}_{t-q}$, em que $\mathbf{w}_t = \Delta^d z_t$, $\tilde{\mathbf{w}}_t = \mathbf{w}_t - \mu_w$ e $\mu_w = E[W_t]$.

Quando $d > 0$ é conveniente considerar $\mu_w = 0$. Caso contrário, μ_w será mais um parâmetro a ser estimado. A seguir, pode-se observar alguns métodos que possibilitam a obtenção de estimadores para os parâmetros do modelo identificado.

2.6 - Critério de informação bayesiana (BIC)

O BIC (Bayesian Information Criterion), critério de informação bayesiano, também chamado de Critério de Schwarz é um critério de seleção de modelo proposto por Schwarz (1978) e modificado por Carlin e Louis (2000) para ser aplicado considerando-se a densidade a *posteriori* dos parâmetros do modelo ajustado. Esse critério pondera entre a função do logaritmo verossimilhança maximizado e o número de parâmetros do modelo. O melhor modelo é aquele que apresenta o menor valor do BIC definido por,

$$\text{BIC} = -2 \log f_X(x_n | \mathcal{G}) + p \log n$$

em que, $f_X(x_n | \mathcal{G})$ é o modelo escolhido, p é o número de parâmetros a serem estimados e n é o número de observações da amostra. Ele é um critério de avaliação de modelos definido em termos da probabilidade a *posteriori*, sendo assim chamado porque Schwarz deu um argumento Bayesiano para prová-lo.

3 - Previsão

É muito comum se observar nos livros textos sobre séries temporais a palavra *previsão* ao se referir às estimativas futuras feitas a partir do estudo de uma série temporal, como se o modelo utilizado tivesse a capacidade de fornecer ao analista uma previsão (ver mesmo antes de o fato acontecer), o que é absolutamente falso (GARBER, 1995).

Segundo Ribeiro, Goldschmidt e Choren (2009) para que seja possível prever os valores futuros com base em valores passados, é necessário que se disponha de uma memória histórica de dados ocorridos anteriormente. Todavia, 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, que podem envolver

cálculos simples ou procedimentos complexos. Existem diversos métodos para auxiliar na tarefa de previsão de séries temporais, como por exemplo: modelos de Suavização Exponencial, modelos auto regressivos (AR), de médias móveis (MA) e Modelos ARIMA. Tecnologias de inteligência computacional, tais como redes neurais, lógica nebulosa e algoritmos genéticos, proporcionam a criação de metodologias avançadas de previsão. A metodologia de previsão é geralmente compreendida como um conjunto de abordagens, métodos e ferramentas para coleta de dados de séries temporais para serem usadas em previsão de valores futuros, baseados em valores passados (PALIT; POPOVIC, 2005).

4 - Aplicação

No presente trabalho o estudo das tendências de temperatura máxima para a cidade de Campina Grande, no período de janeiro de 2002 a dezembro de 2010 foram ajustados, fazendo-se o uso da modelagem estatística com a utilização de séries temporais, aplicando-se alguns modelos, com o objetivo de identificar qual modelo é mais adequado para a caracterização da previsão da temperatura máxima da cidade em estudo. Foi utilizado o Software R versão 2.15.1.

Pode-se observar por meio da Figura 1 os valores de temperatura máxima para a cidade, valores estes, que variam entre 24° C e 31° C, ao longo dos anos de 2002 a 2010. Dessa forma, observa-se que o menor valor da temperatura máxima está compreendido entre os anos de 2004 e 2006, da mesma forma verifica-se que o maior valor da temperatura máxima está quase que exatamente no ano de 2008, fazendo-se uma análise mais detalhada, os dados se mostram bem distribuídos, ou seja, aparentemente não observa-se tendência nos dados. Os pontos representam as observações obtidas para a variável temperatura máxima para a cidade de Campina Grande.

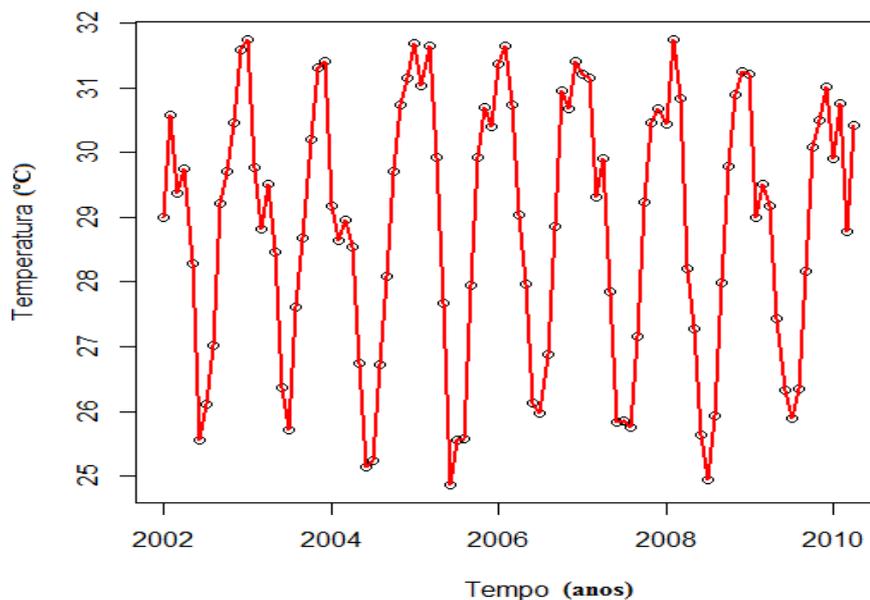


Figura - 1. Temperatura máxima para a cidade de Campina Grande correspondente aos período de 2002 a 2010.

Dando seqüência a análise, pode-se observar por meio da Figura 2 que a temperatura máxima no período de Janeiro de 2002 a Dezembro 2010 apresenta como valor mínimo aproximadamente 25°C e valor máximo 32°C. Para o primeiro quartil tem-se o valor da temperatura em torno de 27°C, a mediana corresponde a um valor de temperatura um pouco maior que 29°C e o terceiro quartil corresponde a um valor de temperatura menor que 30°C. Visualmente verifica-se uma leve assimetria negativa, pois, a distribuição dos dados esta à esquerda, de forma que a linha que representa a mediana está mais próxima do terceiro quartil do que do primeiro quartil.

Ao analisar a Tabela 1, observa-se que o modelo com menor valor do BIC é o modelo ARIMA(1,0,0)(2,0,0), isto é, dentre os dez modelos ARIMAS descritos na Tabela 1 este foi o que melhor se ajustou aos dados analisados. Este fato é ser corroborado pelo valor do Erro Quadrado Médio (EQM), apresentado na Tabela 1, corresponde ao menor valor para o EQM. Portanto, pode-se concluir que o EQM também especifica o melhor modelo ajustado, especificando-se o modelo ARIMA(1,0,0)(2,0,0) o mais adequado para o ajuste da temperatura máxima na cidade de Campina Grande.

Observam-se alguns artigos que tem por objetivo analisar uma série de temperatura média mensal de alguma cidade ou região, pode-se citar como exemplo, o artigo de Silva, Guimarães, Tavares (2008) referente a previsão da temperatura média mensal de Uberlândia-MG, com modelos de séries temporais, em que descreve-se os componentes e é feita uma previsão para períodos subsequentes por meio de modelos ajustados para a série temporal. Os referidos autores identificaram a presença dos

componentes de tendência e sazonalidade, entre outros exemplos tem-se, o artigo de Chechi e Bayer (2011) modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS, na classe de modelos ARIMA foram selecionados, utilizando-se critérios de informação. Modelos do tipo SARIMA foram ajustados e, por meio do critério BIC (Bayesian Information Criterion) e erro quadrático médio (EQM) foi selecionado o modelo SARIMA (3,1,0)(0,1,1) para fins de previsão.

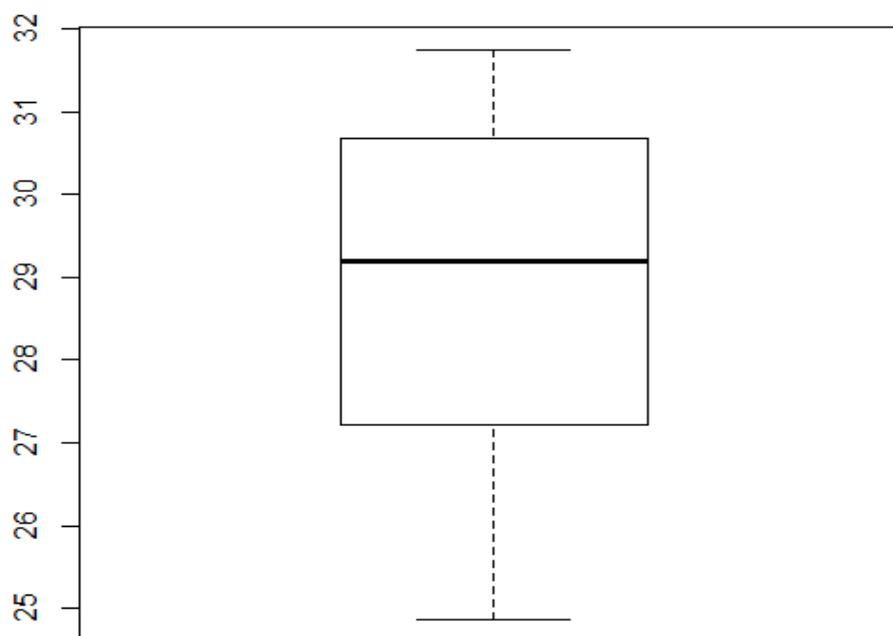


Figura - 2. Box - Plot para a temperatura máxima da cidade de Campina Grande para o período de 2002 a 2010.

Tabela 1 – Modelos Arima utilizados, valores dos BIC'S, EQM'S e Logaritmo verossimilhança dos modelos de séries temporais para as temperaturas máximas de Campina Grande.

<i>MODELO</i>	<i>LOGARITMO VEROSSIMILHANÇA</i>	<i>BIC</i>	<i>EQM</i>
ARIMA(1,0,0)(2,0,0) (M ₁)	-131,86	286,75	0, 8324
ARIMA(1,0,1)(2,0,0) (M ₂)	-131,84	291,32	0, 8385
ARIMA(1,1,0)(2,0,0) (M ₃)	-141,89	302,16	0, 9417
ARIMA(1,1,1)(2,0,0) (M ₄)	-142,27	290,37	0, 8360
ARIMA(0,1,1)(2,0,0) (M ₅)	-140,72	299,82	0, 9387
ARIMA(0,1,2)(2,0,0) (M ₆)	-133,98	290,93	0, 8499
ARIMA(1,1,2)(2,0,0) (M ₇)	-131,40	290,37	0, 8469
ARIMA(2,1,0)(2,0,0) (M ₈)	-141,52	306,02	0, 9326
ARIMA(2,0,1)(2,0,0) (M ₉)	-131,16	294,55	0, 8370
ARIMA(2,1,2)(2,0,0) (M ₁₀)	-130,68	293,52	0, 8395

4.1 Análises dos Resíduos

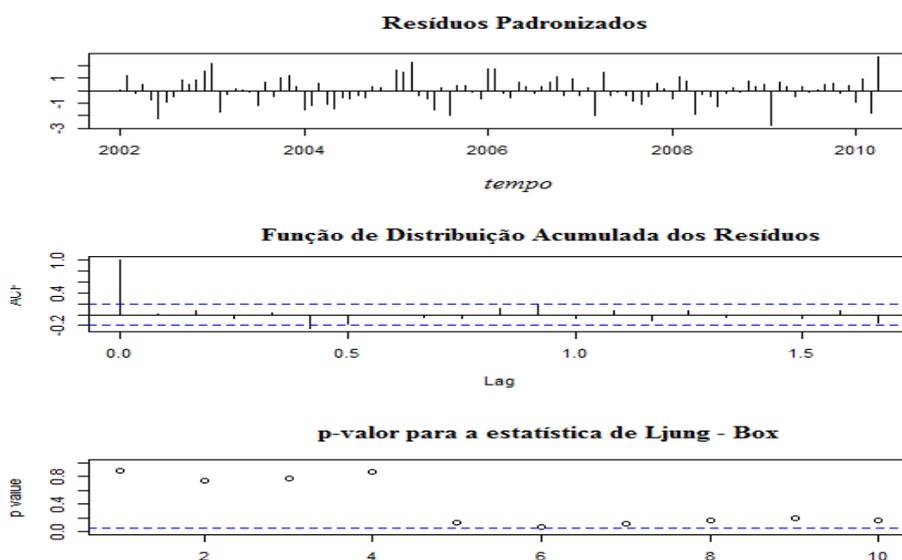


Figura - 3. Análise de Resíduos do Modelo selecionado.

Observou-se por meio da Figura 3 os resíduos padronizados, a função de distribuição acumulada dos resíduos e o teste de Ljung-Box, que por sua vez apresenta níveis superiores a 0,05, conseqüentemente, pode-se concluir que não há indícios para se

rejeitar a hipótese nula, isto é, os resíduos não são correlacionados, portanto os mesmos parecem estar distribuídos aleatoriamente. Segundo Moretin e Toloi (2004), basta utilizar as dez ou quinze primeiras autocorrelações para a verificação da estatística de Ljung-Box.

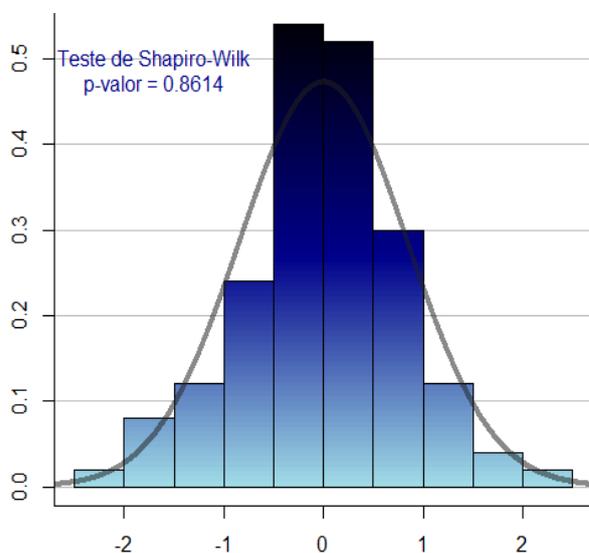


Figura - 4. Histograma dos resíduos para os valores de temperatura máxima para a cidade de Campina Grande no período de 2002 a 2010.

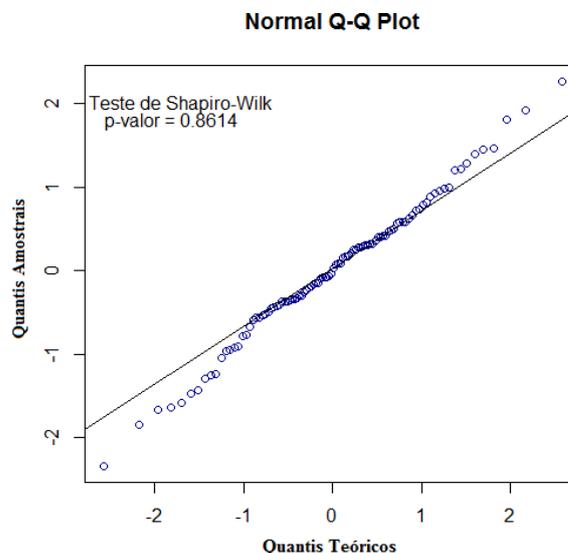


Figura - 5. Gráfico QQ-Plot que especifica a relação existente entre os quantis teóricos versus os quantis amostrais.

Pode-se observar que o histograma apresenta claramente a normalização dos dados para a temperatura máxima, fato este comprovado pelo teste de Shapiro-Wilk, que por sua vez apresenta o valor p superior ao nível de significância de 0,05. Sendo assim, há indícios para não rejeitar a hipótese nula de que os resíduos seguem uma distribuição normal (Figura – 4).

Dando sequência às análises sobre a temperatura pode-se observar por meio da Figura 6 os quantis teóricos e os quantis amostrais referentes aos dados de temperatura máxima. Verifica-se também a presença de normalidade, que por sua vez está evidenciada pelos envelopes de simulação, permitindo-se ter uma ideia do comportamento da normalidade para os dados de temperatura máxima da cidade de Campina Grande entre os anos de 2002 a 2010. Ressalta-se que os pontos representam as observações, as linhas contínuas as curvas ajustadas e as tracejadas seus intervalos de confiança com 95% de confiabilidade.

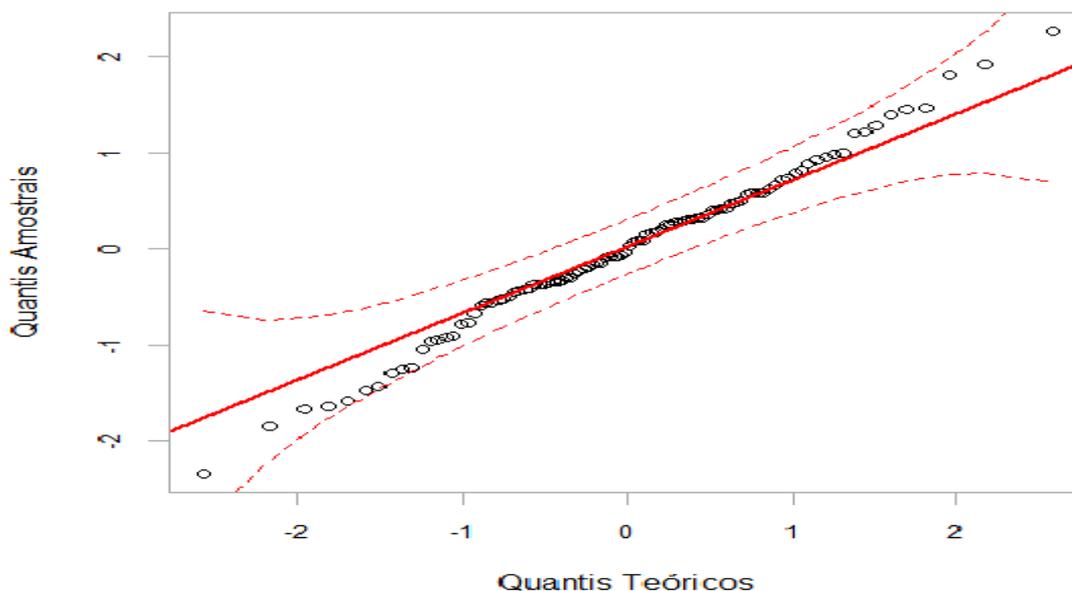


Figura - 6. Quantis amostrais esperados sob normalidade do QQ-Plot. Os pontos representam as observações, as linhas contínuas as curvas ajustadas e as tracejadas os intervalos de confiança a 95% de probabilidade

4.2 Estudo de Previsão

Ao analisar a Tabela 2, verifica-se por meio dos valores de temperatura máxima obtidos para a cidade de Campina Grande nos meses de maio, junho, julho, agosto, setembro e outubro que o mês com a temperatura máxima mais elevada é o mês de outubro, conseqüentemente o mês com temperatura máxima mais amena é o mês julho. Observando-se também os limites ilustrados na Tabela 2, pode-se dizer que os valores obtidos estão ajustados corretamente ao modelo ARIMA (1,0,0)(2,0,0) utilizado, isto é, nenhum valor esperado encontra-se fora dos limites compreendidos, caso isso acontecesse seria caracterizado como um valor extremo, ou seja, um valor fora dos limites especificados. Desta forma, o modelo utilizado foi adequado ao estudo aqui apresentado.

Tabela 2. Previsões e Limites Inferiores (LI) e Limites Superiores (LS) dos intervalos de previsões para a temperatura máxima da cidade de Campina Grande para 2011 ao nível α de 95% de confiança.

MÊS	VALORES ESPERADOS	LIMITE INFERIOR	LIMITE SUPERIOR
MAIO	28, 34598	26, 70065	29, 99131
JUNHO	26, 73148	24, 92649	28, 53648
JULHO	26, 06048	24, 22469	27, 89627
AGOSTO	26, 59601	24, 75402	28, 43800
SETEMBRO	28, 21800	26, 37474	30, 06125
OUTUBRO	29, 78160	27, 93809	31, 62511

Pode-se observar que as variações de temperatura máxima no gráfico mostra um menor valor de aproximadamente 25°C ocorrendo provavelmente no ano de 2005, já o valor máximo atingido pela temperatura máxima é de aproximadamente 32°C , mas fica bastante evidente que o ano que apresenta maior temperatura está normalizado, ou seja, semelhante aos anos anteriores, porém para o caso da menor temperatura máxima o que provavelmente acontecerá será um ano que apresente um valor mínimo para a temperatura máxima bem menor que os obtidos anteriormente.

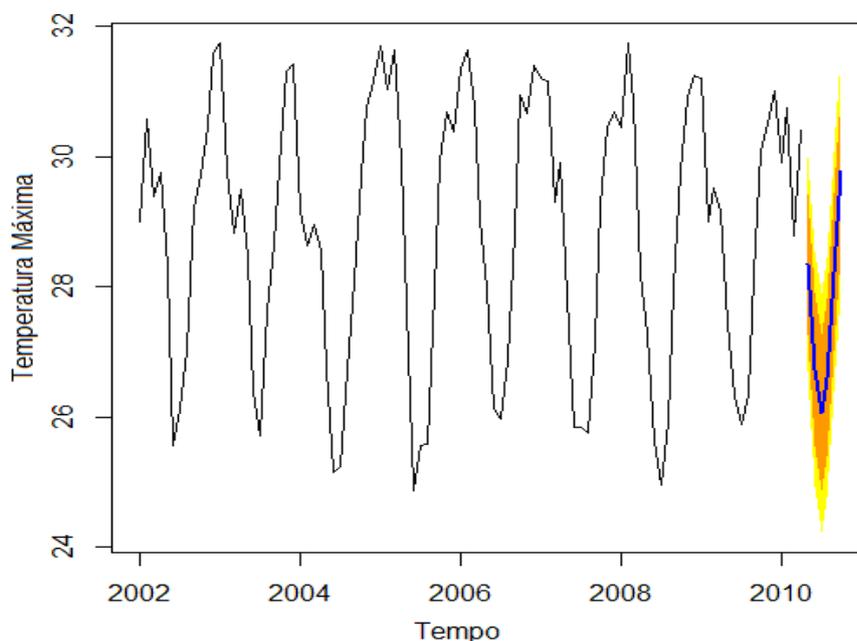


Figura -7. Gráfico que se refere à previsão para os dados de temperatura a partir do modelo selecionado.

A figura 7 corresponde aos valores previstos para a temperatura máxima para a cidade em estudo, no qual pode-se perceber que os valores obtidos estão dentro do esperado para a previsão do ano de 2011. Por meio deste gráfico ainda, percebe-se que os valores previstos são temperaturas máximas, com valores baixos ao relacionar com os valores obtidos no período de 2002 a 2010, isto quer dizer que para o ano de 2011, obteve-se um valor mínimo para a temperatura máxima em torno de 24°C aproximadamente e um valor máximo menor que o limite superior.

5 – Conclusão

Utilizou-se a modelagem estatística por meio de séries temporais para dados de temperatura máxima para a Cidade de Campina Grande referente ao período de 2002 a 2010, com a finalidade de obter uma previsão para esta variável.

Após especificar o modelo que melhor se ajustou aos dados, pôde-se verificar por meio da previsão que os valores obtidos encontram-se num patamar esperado para o ano de 2011.

Portanto, a previsão por meio da utilização de séries temporais permitiu-se verificar fatores que influenciam na variável temperatura máxima, sendo assim, a modelagem estatística via séries temporais garante uma maior confiabilidade no estudo da temperatura máxima da cidade de Campina Grande.

Referências Bibliográficas

ANDERSON, T. W. **The Statistical Analysis of Time Series**. New York: John Wiley & Sons, 1971, 540p.

AZEVEDO, P. V.; SILVA ; B. B. da.; RODRIGUES, M. F. G. Previsão Estatística das Chuvas de Outono no Estado do Ceará. **Revista Brasileira de Meteorologia**, v. 13, p19-30, 1998.

BARROS, M. **Séries Temporais e Modelagem Estatística**. In: site da M. Barros Consultoria Ltda, 2003. 150p.

BOWERMAN; O'CONNELL. **Times Series Forecasting: Unified Concepts and Computer Implementation**. Duxburg Press, 1987, 730p.

BOX, G. E. P.; JENKINS, G.M. **Time Séries Analysis, forecasting and control**, San Francisco, 1970, 450p.

CARLIN, B. P.; LOUIS, T. A. Bayes and Empirical Bayes methods for data analysis, **Chapman and Hall**, v.20, p.45-53, 2000.

CARVALHO, L.G; SEDYAMA, G.C; CECON, P.R.; ALVES, H.M.R. Aplicação da análise harmônica por séries de Fourier para a previsão de produtividade da cultura do café no Estado de Minas Gerais. **Engenharia Agrícola**. Jaboticabal, v.25, p.732-741, 2005.

CHATFIELD C. **The Analysis of Time Series: An introduction**. Chapman & Hall, fifth edition, NEW YORK, 1996, 283p.

CHECHI, L.; Bayer, F. M. Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS **Revista Brasileira de Engenharia Agrícola e Ambiental**. Campina Grande, PB.,v.16, n.12, p.1321–1329, 2012.

COSTA, S. C. da. **Modelos Lineares Generalizados Mistos para Dados Longitudinais**. São Paulo, 2003, 250p.

COSTA, K. C. de O. **Análise de DFA e de Agrupamento do perfil de densidade de poços de petróleo**. Rio Grande do Niorte, 2009, 300p.

DEMÉTRIO, C. G. B.; ZOCCHI, S. S. **Modelos de Regressão**, 2006, 258p.

DICKEY, D.A.; FULLER, W.A. Distribution of the estimator for auto-regressive time series with a unit root. **Journal of the American Statistical Association**. Alexandria, v. 74, p.427-431. 1979.

DICKEY, P.A.; FULLER, W.A. Distribution of Estimators for Autoregressive Time Series With a Unit Root, **Journal of the American Statistical Association**, v.74, p.427-431. 1979.

DICKEY, D.A.; FULLER, W.A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica**, Princeton, v. 49, p.1057-1072. 1981.

EHLERS, R. S. **Análise de séries temporais**. Curitiba: Universidade Federal do Paraná, 2007, 116p.

ENDERS, W. **Applied econometric time series**. New York: John Wiley & Sons, 2004, 433p.

FAVA, V. L. **Metodologia de Box-Jenkins para modelos univariados**. Manual de econometria:nível intermediário, São Paulo: Atlas, 2000, 308p.

FERREIRA, A. G.; MELLO, N. G. DA S. Principais Sistemas Atmosféricos atuantes sobre a Região Nordeste do Brasil e a Influência dos Oceanos Pacífico e Atlântico no Clima da Região. **Revista Brasileira de Climatologia**, v.1. p.20-31, 2004.

FISCHER, S. **Séries Univariantes de Tempo- Metodologia de Box & Jenkins**. Porto Alegre: FEE, 1982, 450p.

- FULLER, W.A. **Introduction to statistical time series**. New York: John Wiley & Sons, 1976, 424p.
- GARBER, R. **Análise de Séries Temporais**. São Paulo. 1995, 350p.
- GRANGER; NEWBOLD. **Forecasting Economic Time Series**, **Academic Press**. New York, v.1. p56-62, 1977.
- GUJARATI, D. N. **Econometria Básica**. 5ª Ed. Editora Bokman, 2000. .
- JENKINS, G. M.; D. G. WATTS. **Spectral Analysis and its Applications**. Holden-day, 1968. 525p.
- JÚNIOR, D. L. **Influência Local em Modelos de Regressão** Campina Grande - PB, 2005.
- MAACK, R. **Geografia Física do Estado do Paraná**. Curitiba: CODEPAR, 1981. 329p.
- MEDEIROS. **Análise de Séries Temporais**. Rio de Janeiro, 2005. 59p.
- MONTGOMERY, D. C.; FOHNSON, L. A.; GARDINER, J. S. **Forecasting and time seáes analysis**, 2 ed., New York. 1990 540p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Edgard Blücher, 2004. 535p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**, 2006. 400p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Previsões de séries temporais**, 2ª ed. Atual, São Paulo, 1987. 400p.
- MORETTIN, P. A. **Séries Temporais em Finaças**. São Paulo: Instituto de Matemática e Estatística – Universidade de São Paulo, 2002. 150p.
- NELSON, C. R. **Applied Time Series Analysis**. San Francisco, 1973. 450p.

PALIT, A. K., POPOVIC, D. **Computational Intelligence in Time Series Forecasting**. Londres: Springer-Verlag, 2005. 372p.

PANKRATZ, A. **Forecasting with univariate Box-Jenkins Models: Concepts and Cases**. Wiley: New York. 1983 350p.

OYAMA, M. D.; C. A. NOBRE. A new climate-vegetation equilibrium state for Tropical South America. **Geophysical Research Letter**. v.3, p.56-64, 2003.

PINDYCK, R.S.; RUBINFELD, P.L. **Econometric Models & Economic Forecasts**, McGraw-Hill International Editions. 1991 345p.

RIBEIRO, C. V.; GOLDSCHMIDT, R., CHOREN, R. Métodos para Previsão de Séries Temporais e suas Tendências de Desenvolvimento, **Instituto Militar de Engenharia Rio de Janeiro**. Rio de Janeiro, v.4. p.35-48, 2009.

SARNAGLIA, A. J. Q. **Estimação de Processos Periódicos Autorregressivos: Uma Abordagem no Domínio da Frequência**. Belo Horizonte-MG, 2010. 150p.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics** v.6. p.461–464, 1978.

SILVA, A. M. da. Estudo de Modelos ARIMA com Variáveis Angulares para Utilização na Perfuração de Poços Petrolíferos. Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática. Campina Grande – PB, 2007, p72.

SILVA, M. I. S.; GUIMARÃES, E. C.; TAVARES, M. Previsão da temperatura média mensal de Uberlândia, MG, com modelos de séries temporais. Campina Grande, PB. **Revista Brasileira de Engenharia Agrícola e Ambiental**. v.12. p.480–485, 2008.

SILVA, P. O. M. P., GOLDSCHMIDT, R. R., SOARES, J. A., FERLIN, C., **Previsão de Séries Temporais Utilizando Lógica Nebulosa**. 4º CONTECSI - Universidade de São Paulo, 2007. 200p.

SOUZA, R. C. **Modelos estruturais para previsão de séries temporais: Abordagens clássica e bayesiana**. São Paulo, 1989. 315p.

WALKER, GT. Ceará (Brazil) famines and the general air movement. **Beitrag zur Physik der freien Atmosphere**, v.14. p.88-93, 1928.

WILKS, D. S. **Statistical methods in the Atmospheric Sciences**. Academic Press, New York, 1995. 468 p.

WONNACOTT, T. H.; WONNACOTT, R. J. **Introductory Statistics for Business and Economics**. 4. ed. New York: John Wiley, 1990. 730p.