

MAURÍCIO SILVA LACERDA

ABORDAGEM ESTATÍSTICA DE VARIÁVEIS CLIMÁTICAS DE VIÇOSA - MG

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2017

**Ficha catalográfica preparada pela Biblioteca Central da
Universidade Federal de Viçosa - Câmpus Viçosa**

T

L131a Lacerda, Maurício Silva, 1991-
2017 Abordagem estatística de variáveis climáticas de Viçosa
- MG / Maurício Silva Lacerda. - Viçosa, MG, 2017.
ix, 93f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador : Paulo César Emiliano.

Dissertação (mestrado) - Universidade Federal de
Viçosa.

Inclui bibliografia.

1. Climatologia - Viçosa (MG). 2. Classificação climática.
3. Temperatura atmosférica. 4. Precipitação (Meteorologia).
I. Universidade Federal de Viçosa. Departamento de
Estatística. Programa de Pós-graduação em Estatística
Aplicada e Biometria. II. Título.

CDD 22 ed. 551.68151

MAURÍCIO SILVA LACERDA

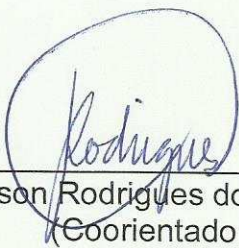
ABORDAGEM ESTATÍSTICA DE VARIÁVEIS CLIMÁTICAS DE VIÇOSA – MG.

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

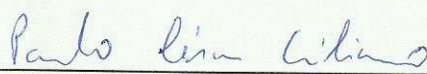
APROVADA: 15 de fevereiro de 2017.



João Marcos Louzada



Gerson Rodrigues dos Santos
(Coorientador)



Paulo César Emiliano
(Orientador)

*Dedico a Deus, pois sem
Ele nada seria possível*

AGRADECIMENTOS

Primeiramente agradeço a Deus, que concedeu-me força, saúde e determinação em todos os momentos, sempre me guiando pelo melhor caminho no decorrer deste trabalho e de toda minha vida.

A todos, que direta ou indiretamente, contribuíram para a realização deste trabalho, muito obrigado!

Meus familiares, em especial aqueles que me incentivaram e acompanharam de perto nesta caminhada, minha namorada que sempre apoiou, incentivou e cobrou meus estudos. Aos meus amigos, aqueles da infância que sempre torcem por mim, e também os amigos que tive o prazer de conquistar ao longo da graduação e mestrado, vocês foram muito importantes neste trabalho, não tem como esquecer das inúmeras horas de estudos juntos, esclarecendo minhas dúvidas e colaborando com meu aprendizado.

Ao professor e orientador Paulo César Emiliano, pela paciência, ensinamentos, força e dedicação ao longo do trabalho.

Ao amigo e professor Eduardo Campana Barbosa pelas incansáveis horas de contribuição, inclusive nos fins de semana e pelas redes sociais.

A todos os professores do Departamento de Estatística da UFV pelo empenho em ensinar, esclarecer dúvidas e incentivar. Aos servidores por estarem sempre a disposição para esclarecimentos, auxiliar nas atividades burocráticas e solucionar problemas.

Ao Departamento de Engenharia Agrícola da UFV pela disponibilidade dos dados que foram utilizados neste trabalho.

Aos membro da banca que fizeram observações relevantes, acrescentando e melhorando este trabalho.

Enfim, a todos da cidade de Viçosa que acolheu-me, ajudando no processo de adaptação à esta nova etapa da vida.

“Quanto mais um homem se aproxima de suas metas, tanto mais crescem as dificuldades.”

Johann Goethe

“Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá.”

Ayrton Senna

SUMÁRIO

	Lista de Figuras	vi
	Lista de Tabelas	vii
	RESUMO	viii
	ABSTRACT	ix
	CAPÍTULO 1	1
1	INTRODUÇÃO GERAL	1
2	REFERENCIAL TEÓRICO	4
2.1	Estacionariedade	6
2.1.1	Teste de Dickey-Fuller Aumentado	8
2.2	Sazonalidade	9
2.3	Seleção de modelos	12
2.3.1	Classes de modelos de Box e Jenkins	12
	REFERÊNCIAS	23
	CAPÍTULO 2 Estudo da média mensal das temperaturas máxima e mínima de Viçosa-MG por meio de séries temporais	25
1	INTRODUÇÃO	26
2	MATERIAL E MÉTODOS	30
3	RESULTADOS E DISCUSSÃO	33
3.1	Média mensal da temperatura máxima	33
3.2	Média mensal da temperatura mínima	42
4	CONCLUSÃO	52
	REFERÊNCIAS	54
	CAPÍTULO 3 Estudo da precipitação pluviométrica total mensal de Viçosa-MG por meio de séries temporais	56
1	INTRODUÇÃO	57
2	MATERIAL E MÉTODOS	61
3	RESULTADOS E DISCUSSÃO	65
4	CONCLUSÃO	75
	REFERÊNCIAS	76
	APÊNDICE	77

LISTA DE FIGURAS

Capítulo 2	25
Gráfico 1 Série histórica das médias mensais das temperaturas máximas da cidade de Viçosa-MG com a reta de tendência	34
Gráfico 2 A - Série da média mensal da temperatura máxima de Viçosa-MG; B - Componente de tendência da série; C - Componente sazonal da série; D - Fator aleatório da série	35
Gráfico 3 Boxplot da série com os diferentes meses da média mensal da temperatura máxima	36
Gráfico 4 Periodograma da série da média mensal da temperatura máxima de Viçosa-MG	37
Gráfico 5 A - Fac da série original; B - Facp da série original; C - Fac da série diferenciada; D - Facp da série diferenciada	38
Gráfico 6 A - Histograma dos resíduos; B - QQ-plot dos resíduos	40
Gráfico 7 A - Fac dos resíduos; B - Facp dos resíduos; C - Fac dos resíduos elevados ao quadrado; D - Facp dos resíduos elevados ao quadrado	40
Gráfico 8 Série histórica da média mensal das temperaturas mínimas da cidade de Viçosa-MG com a reta de tendência	43
Gráfico 9 A - Série da média mensal da temperatura mínima de Viçosa-MG; B - Componente de tendência da série; C - Componente sazonal da série; D - Fator aleatório da série	44
Gráfico 10 Boxplot de cada mês da série em anos diferentes	45
Gráfico 11 Periodograma da série da média mensal das temperaturas mínimas de Viçosa-MG	46
Gráfico 12 A - Fac da série original; B - Facp da série original; C - Fac da série diferenciada; D - Facp da série diferenciada	47
Gráfico 13 A - Histograma dos resíduos; B - QQ-plot dos resíduos	50
Gráfico 14 A - Fac dos resíduos; B - Facp dos resíduos; C - Fac dos resíduos elevados ao quadrado; D - Facp dos resíduos elevados ao quadrado	50
Capítulo 3	56
Gráfico 1 Série histórica da precipitação pluviométrica mensal da cidade de Viçosa-MG com a reta de tendência	66
Gráfico 2 A - Série do índice mensal da precipitação pluviométrica de Viçosa-MG; B - Componente de tendência da série; C - Componente sazonal da série; D - Fator aleatório da série	67
Gráfico 3 Boxplot de cada mês da série em anos diferentes	68
Gráfico 4 Periodograma da série da precipitação total mensal de Viçosa-MG	69
Gráfico 5 A - Fac da série original; B - Facp da série original; C - Fac da série diferenciada; D - Facp da série diferenciada	69
Gráfico 6 A - Histograma dos resíduos; B - QQ-plot dos resíduos	72
Gráfico 7 Gráfico da fac e facp dos resíduos e dos resíduos elevados ao quadrado	72

LISTA DE TABELAS

Capítulo 2	25
Tabela 1 Resultado da análise para tendência de elevação da temperatura	35
Tabela 2 Resultado do teste de Dickey-Fuller aumentado	36
Tabela 3 Resultado do teste de Dickey-Fuller aumentado para a série após aplicação de uma diferença	37
Tabela 4 Comparação de diversos modelos concorrentes para a série da média mensal da temperatura máxima de Viçosa-MG	38
Tabela 5 Estimativa dos parâmetros do modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$	39
Tabela 6 Resultado da análise residual do modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$	39
Tabela 7 Previsão para os próximos 12 meses da média mensal das temperaturas máximas	41
Tabela 8 Resultado da análise para tendência de elevação da temperatura	44
Tabela 9 Resultado do teste de Dickey-Fuller aumentado	45
Tabela 10 Resultado do teste de Dickey-Fuller aumentado para a série após aplicação de uma diferença	46
Tabela 11 Comparação de diversos modelos concorrentes para a série da média mensal das temperaturas mínimas de Viçosa-MG	48
Tabela 12 Estimativa dos parâmetros do modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$	48
Tabela 13 Estimativa dos parâmetros do modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$	48
Tabela 14 Resultado da análise residual do modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$	49
Tabela 15 Comparação de modelos concorrentes	51
Tabela 16 Estimativa dos parâmetros do modelo $ARCH(2)$	51
Tabela 17 Previsão para os próximos 12 meses da média mensal da temperatura mínima	52
Capítulo 3	56
Tabela 1 Análise da tendência para a série de precipitação pluviométrica	67
Tabela 2 Resultado do teste de Dickey-Fuller aumentado	68
Tabela 3 Comparação de diversos modelos concorrentes para a série da precipitação pluviométrica total mensal de Viçosa-MG	70
Tabela 4 Estimativa dos parâmetros do modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$	71
Tabela 5 Resultado da análise residual do modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$	71
Tabela 6 Comparação de modelos concorrentes da classe ARCH/GARCH	73
Tabela 7 Estimativa dos parâmetros do modelo $ARCH(2)$	73
Tabela 8 Previsão para os próximos 12 meses da precipitação total mensal	74

RESUMO

LACERDA, Maurício Silva, M.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Abordagem estatística de variáveis climáticas de Viçosa - MG.** Orientador: Paulo César Emiliano. Coorientadores: Moysés Nascimento e Gerson Rodrigues dos Santos.

Conhecer o comportamento de variáveis climáticas como a temperatura e o índice de chuvas de uma cidade ou região é de grande importância para que haja melhor planejamento e tomadas de decisões futuras, por este motivo estudos destas variáveis estão sempre em evidência. Desta forma, tem-se como objetivo geral realizar um estudo de algumas características do clima da cidade de Viçosa-MG, por meio de séries temporais, com enfoque nas temperaturas máximas e mínimas e no nível de precipitação pluviométrica, com objetivo específico de identificar alterações nestes fatores. Para isso analisou-se um conjunto de dados históricos acerca do clima da cidade com registros desde janeiro de 1968 até julho de 2016, iniciando o tratamento dos dados calculando a média mensal para as temperaturas e o total mensal para a precipitação, em seguida foi aplicado o teste de Dickey-Fuller aumentado para análise da tendência, e uma decomposição espectral da série nas frequências de Fourier para detectar sazonalidade. Posteriormente, escolheu-se a classe dos modelos SARIMA e as suas ordens escolhidas por meio dos critérios de informação bayesiano (BIC) e Akaike corrigido (AICc). O modelo escolhido foi verificado através de uma análise residual com os testes de Shapiro-Wilk e Ljung-Box. Para a série histórica da média mensal da temperatura máxima o modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$ se apresentou com bom ajuste. As séries da média mensal da temperatura mínima e da precipitação pluviométrica total mensal necessitaram de um modelo que ajustassem os resíduos para que os pressupostos da modelagem fossem satisfeitos, com isso a primeira teve sua correlação serial modelada pelo $SARIMA(1, 1, 2) \times (0, 1, 1)$ e seus resíduos pelo $ARCH(2)$ e a segunda série teve os modelos $SARIMA(1, 0, 0) \times (0, 1, 1)$ para ajustar a correlação serial e $ARCH(2)$ para os resíduos. Com o presente estudo conclui-se que a cidade de Viçosa-MG teve uma elevação acima de 1°C na média mensal de suas temperaturas máximas e mínimas, enquanto que o nível de precipitação não sofreu alterações ao longo do tempo estudado.

ABSTRACT

LACERDA, Maurício Silva, M.Sc., Universidade Federal de Viçosa, February, 2017. **Statistical approach of climate variables of Viçosa - MG.** Advisor: Paulo César Emiliano. Co-Advisors: Moysés Nascimento and Gerson Rodrigues dos Santos.

Knowing the behavior of climatic variables such as temperature behavior and rainfall index of a city or region is of great importance for better planning and future decision making, therefore, studies of these variables are always in evidence. In this way, our general aim was to study some climate characteristics from Viçosa-MG city using time series analyses, focusing on maximum and minimum temperatures and rainfall levels, with the specific aim to identify changes in these factors. For that, a historical data set about the climate of the city with records from January 1968 to July 2016 was analyzed. We first calculated the monthly mean of temperatures and the monthly total mean of precipitation, then, the increased Dickey-Fuller test for trend analysis, and a spectral decomposition of the series on Fourier frequencies was applied to detect the seasonality. Subsequently, class of SARIMA models and their orders were chosen by Bayesian information criteria (BIC) and corrected Akaike (AICc). The chosen model was verified through a residual analysis using Shapiro-Wilk and Ljung-Box tests. SARIMA model $(2,1,2) \times (0,1,1)$ presented good fitting for the historical series of monthly mean of maximum temperature. The series of monthly mean of minimum temperature and monthly rainfall required a model to fit the residuals so that the modeling assumptions were satisfied, with the first having its serial correlation modeled by SARIMA $(1,1,2) \times (0,1,1)$ and its residuals by ARCH (2) and the second series had the SARIMA models $(1,0,0) \times (0,1,1)$ to fit the serial correlation and also the ARCH (2) for residuals. The present study allowed us to conclude that Viçosa-MG had an increase of more than 1°C in the monthly mean of its maximum and minimum temperatures, while the precipitation level did not change during the studied time.

CAPÍTULO 1

1 INTRODUÇÃO GERAL

A Terra é o terceiro planeta mais próximo do sol dentre aqueles que se encontram no sistema solar. É o único que possui vida dentre todos os conhecidos e estudados. A presença de vida na Terra é possível devido as suas características climáticas, que são favoráveis para que isso aconteça, caracterizando-o como um planeta com temperaturas moderadas se comparado aos demais.

Segundo João (2009), o planeta Terra é dotado de dois movimentos principais que são fortemente relacionados ao clima e suas variações, são eles: o movimento de translação e o de rotação. O primeiro é efetuado ao redor do sol, sendo a fonte de luz e calor, que regula todo o processo climático da Terra. O segundo é o movimento que ela realiza em torno de seu eixo imaginário sendo responsável pelo dia e a noite. O movimento de translação é responsável também pelas estações e pelas variações climáticas que ocorrem durante o ano.

As estações do ano que caracterizam períodos mais quentes no verão e mais frios no inverno estão relacionadas com a inclinação da Terra em relação ao seu eixo imaginário. A inclinação faz com que a orientação dela mude durante o ano. Quando a Terra inclina-se, tendo o hemisfério norte voltado para o sol, tem-se o verão neste hemisfério, enquanto que no hemisfério sul ocorre o inverno, posteriormente esta inclinação inverte ocorrendo também a inversão das estações nos hemisférios norte e sul. A inclinação é também responsável pelo ângulo de incidência dos raios de sol sobre determinada região do planeta. Os raios que atingem a Terra verticalmente são mais concentrados, enquanto que, os raios com certa inclinação são mais dispersos e percorrem maiores caminhos na atmosfera até atingirem a superfície da Terra, perdendo intensidade.

João (2009) afirma que, não se deve confundir um estado climático com uma condição de clima. Um estado climático é avaliado ao longo de grandes períodos de tempo, tais como, anos, décadas ou séculos. Já algumas condições de climas ocorrem por um período bem menor, ou seja, por alguns dias. Observações do tempo, obtidas em curtos

períodos de tempo, não podem ser inferidas como mudanças climáticas. Desta forma, a ocorrência de dias seguidos muito quentes, muito frios ou chuvosos estão ligados a uma condição do clima naquele momento e não se pode concluir que está havendo uma mudança climática. Para caracterizar-se como mudança climática é preciso que a alteração climatológica repita-se por longos períodos de tempo.

Segundo Marengo (2009) estudos referentes a mudanças climáticas na América do Sul mostraram que nos últimos 50 anos as temperaturas da superfície da Terra aumentaram em $0,75^{\circ}\text{C}$ enquanto que a temperatura mínima teve um aumento um pouco maior, de 1°C .

Ao analisar as temperaturas registradas em diversas cidades do estado de Minas Gerais, Minuzzi et al. (2010) concluíram que ocorreu maior tendência de elevação da temperatura nos meses de setembro e dezembro, resultando em uma elevação entre $1,3^{\circ}\text{C}$ e $3,9^{\circ}\text{C}$ no período de 1961 a 2004.

O clima ocupa grande espaço nas pesquisas e estudos, uma vez que suas mudanças podem provocar alteração e redução no ano agrícola, redução da diversidade genética, modificação na ocorrência de chuvas e suas intensidades, entre outras alterações que podem prejudicar a economia de uma região, se o comportamento climático for muito diferente do que se espera. Baseando-se nisso, estudos apontam o interesse em fazer análises e realizar previsões e, apoiado nelas, tomar decisões para conter ou amenizar os prejuízos causados pelo aumento das temperaturas e a falta de água.

Conforme Marengo (2001) desde a década de 1980 são expressivas as evidências científicas acerca da possibilidade de mudança climática mundial, e desde então vem despertando um grande interesse do público e da comunidade acadêmica.

Os modelos de séries temporais são bastante utilizados e se mostram uma boa ferramenta para análise de dados climatológicos, levando em conta a dependência entre as observações e possíveis ciclos ou sazonalidades que geralmente ocorrem neste tipo de dados. Ao ajustar um bom modelo é possível fazer análises, previsões e interpretar os resultados obtidos de forma rápida e eficiente.

Com as informações a respeito do aquecimento global e a escassez de água que se repete anualmente na cidade de Viçosa-MG, objetiva-se aqui elaborar uma análise do comportamento das médias mensais das temperaturas máximas e mínimas e o índice pluviométrico total mensal ao longo do tempo, desde o ano de 1968 (ano em que a estação meteorológica de Viçosa começou a registrar os dados) se valendo dos modelos de séries temporais para verificar se há indícios de aquecimento ao longo do tempo e, se a escassez

de água é um problema de chuvas ou de planejamento.

Tem-se, como objetivos específicos, identificar a tendência, sazonalidade, ciclos, caso existam, para cada variável do estudo. Objetiva-se também verificar se os modelos escolhidos e analisados permitem fazer previsões de forma adequada e enfim realizar as previsões para os próximos meses, para as médias mensais das temperaturas máximas e mínimas e o índice pluviométrico. Isso permitirá que a população, os setores de abastecimento de água e energia elétrica, o setor da agricultura e agropecuária, entre outros, possam se programarem utilizando essas previsões durante o processo de planejamento.

Vale destacar que, as previsões realizadas por meio de séries temporais não constituem um fim, mas sim um meio que fornece informações e baseando-se nelas decisões podem ser tomadas com determinados objetivos.

Para isso, o trabalho está dividido em três capítulos organizados da seguinte maneira: No capítulo 1, será apresentado uma breve introdução e o referencial teórico com alguns conceitos básicos necessários para o desenvolvimento dos capítulos 2 e 3. Consistirá em uma breve introdução do tema e revisão bibliográfica acerca de séries temporais, sendo apresentados conceitos, testes para detectar características peculiares, modelos de séries temporais e os métodos para a seleção dos melhores modelos a serem utilizados neste estudo.

Nos capítulos 2 e 3, os modelos de séries temporais que melhor se ajustarem aos dados serão especificados, terão seus parâmetros estimados, verificados e com tais modelos serão realizadas previsões. Com os testes do processo de modelagem será possível extrair análises mais precisas acerca do clima de Viçosa-MG e tirar conclusões. No capítulo 2 serão estudadas as médias mensais das temperaturas mínimas e máximas enquanto que no capítulo 3 serão estudados os índices de precipitação total mensal.

2 REFERENCIAL TEÓRICO

Segundo Morettin e Tolo (2006) uma série temporal é qualquer conjunto de observações ordenadas no tempo. Em uma outra definição, uma série temporal é uma realização de um processo estocástico.

João (2009) afirma que uma série temporal pode ser classificada como determinística ou estocástica. Diz-se que a série é determinística quando os seus futuros valores podem ser estabelecidos de forma precisa por alguma função que dependa apenas do tempo. Ela é dita ser estocástica se seus valores futuros puderem ser expressos em termos probabilísticos, pois a série é descrita por meio de uma relação que não envolve apenas o tempo, mas também uma variável aleatória ou termo aleatório residual. A inclusão do termo residual aleatório, se faz necessária, quando não é possível explicar algum comportamento irregular da série utilizando-se apenas uma relação matemática.

Diversos estudos estatísticos são caracterizados pela independência entre as observações, contudo os estudos de séries temporais vão no sentido contrário, sendo caracterizados pela dependência serial, ou seja, a observação do momento em estudo depende das observações que ocorreram anteriores a ela. Esta dependência pode ser avaliada pela correlação entre as observações.

Conforme Morettin e Tolo (2006) há dois enfoques nos estudos de séries temporais, sendo no primeiro a análise feita no domínio temporal e os modelos propostos são paramétricos (número finito de parâmetros). No segundo, a análise é conduzida no domínio das frequências e os modelos propostos são não-paramétricos.

Em estudos de séries temporais são utilizados dados discretos de alguma variável ao longo do tempo, porém grande parte dos dados históricos precisam ser obtidos de séries que apresentam comportamento contínuo. Desta forma, deve-se extrair da variável contínua algum valor de interesse, que represente bem esta variável por meio de uma amostragem, obtendo uma série discreta. Por exemplo, a temperatura é uma variável contínua, assim é extraído-se a temperatura média, máxima ou mínima de cada dia para obter informações discretas e estudar suas características ao longo do tempo. Cotação de ações na bolsa também apresentam valores contínuos, desta forma seleciona-se a máxima, mínima ou média de cada dia para estudar suas características ao longo de grandes períodos.

Conforme Box et al. (2008) ao usar séries temporais para prever um valor num tempo $t + \tau$ consegue-se informações para:

- i) Planejamento econômico de negócios;
- ii) Planejamento de produção;
- iii) Controle da produção de um setor;
- iv) Controle e otimização de processos industriais.

Numa análise de séries temporais busca-se aqueles modelos que melhor represente os dados para diversos objetivos, tais como: sondar o mecanismo que gera a série, fazer previsões de valores futuros, verificar o comportamento, procurar propriedades expressivas nos dados, etc. Entretanto, nem sempre o melhor modelo para um objetivo é também o melhor para outros.

Morettin e Tolo (2006) afirmam que a classe dos modelos paramétricos são feitas no domínio do tempo e os modelos mais comuns são: autorregressivo (AR), autorregressivo de média móvel (ARMA), autorregressivo integrado de média móvel (ARIMA), sazonal autorregressivo integrado de média móvel (SARIMA). Os modelos não paramétricos são baseados no domínio das frequências e os mais utilizados são as funções de autocovariância (ou autocorrelação) e sua transformada de Fourier, o espectro. A vantagem de escrever a série no domínio das frequências está no fato de eliminar o problema de correlação serial.

Morettin e Tolo (2006) afirmam que uma observação da série temporal pode ser decomposta em uma soma de três componentes não-observáveis:

$$Z_t = T_t + S_t + a_t, \text{ com } t = 1, 2, \dots, N,$$

em que T_t é a componente de tendência que modela a variação de crescimento ou decréscimo da série, S_t é a componente de sazonalidade que modela os efeitos da série que se repetem em períodos idênticos de tempo e a_t é a componente aleatória da série, que modela a parte residual não explicada nem pela tendência e nem pela sazonalidade. Para aplicação dos estudos de séries temporais supõe-se que as variáveis aleatórias a_t não sejam correlacionadas e tenham média e variância constantes, ou seja:

$$E(a_t) = 0, \forall t; E(a_t^2) = \sigma_a^2, \forall t \text{ e } E(a_t a_s) = 0 \text{ se } t \neq s.$$

Satisfeitas as condições acima a_t é dito ser ruído branco, caso contrário tem-se indícios de que a modelagem não está boa.

Os resíduos de uma série podem apresentar um padrão que seja não linear. Isto ocorre, por exemplo, quando a variância da série é dependente do tempo. Neste caso tem-se a pressuposição de ruído branco violada. Quando isso ocorre, uma alternativa é utilizar modelos não-lineares para modelar a variância do fenômeno. Os modelos da família autorregressivo com heteroscedasticidade condicional (ARCH) ou autorregressivo com heteroscedasticidade condicional generalizado (GARCH) são exemplos de modelos que podem ser utilizados nestas situações.

2.1 Estacionariedade

Em estudos de séries temporais uma pressuposição geralmente requerida é a de que os dados em análise apresentem comportamento estacionário, ou seja, a série se desenvolve no tempo de forma aleatória ao redor de uma média constante, o que reflete uma forma de equilíbrio estável.

Segundo Morettin e Tolo (2006) uma série estacionária possui as seguintes propriedades:

- i) $E(Z_t) = \mu$;
- ii) $V(Z_t) = \sigma^2$;
- iii) $Cov(Z_t, Z_{t-s}) = \gamma$ (independente do tempo) com $t \neq s$ em que t e s são medidas do tempo.

Grande parte das séries temporais encontradas na prática não apresentam padrão estacionário. Em geral, elas apresentam uma certa tendência de crescimento ou decréscimo, violando a pressuposição dos modelos a serem utilizados.

Um procedimento bastante comum para estimar a tendência T_t é ajustar uma função do tempo aos dados. Morettin e Tolo (2006) afirmam que os tipos de funções comumente utilizados são as polinomiais, exponenciais, logarítmicas ou quaisquer outras funções suaves de t , ou seja, funções que possui derivadas de todas as ordens em relação a t . Geralmente ajustamos um polinômio de grau m dado por:

$$T_t = \beta_0 + \beta_1 t + \dots + \beta_m t^m,$$

em que β_i com $i = 0, 1, 2, \dots, m$ são os parâmetros do modelo que, em geral, são obtidos via métodos dos mínimos quadrados.

Após modelar a tendência obtém-se uma série livre de tendência dada por:

$$Y_t = Z_t - \hat{T}_t,$$

em que Y_t é a nova série livre de tendência, Z_t é a série original e \hat{T}_t é a estimativa da tendência.

Outro procedimento utilizado para eliminar a tendência é fazer alguma transformação nos dados originais, se a série não for estacionária. A transformação mais comum consiste em tomar diferenças sucessivas nos dados originais. A primeira diferença pode ser definida como:

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

em que Z_t é a observação no instante t e Z_{t-1} é a observação em um passo anterior, no instante $t - 1$. Montgomery et al. (2008) afirmam que, em geral, uma ou duas diferenças são suficientes para remover a tendência dos dados. Outra transformação que pode ser utilizada é aplicar alguma função para suavizar a série, exemplo: logaritmo ou raiz quadrada.

Um operador muito utilizado em séries temporais é o operador de tempo passado, que transforma uma observação no instante t em observações passadas. Este operador é denotado por B e definido da seguinte forma:

$$Z_{t-1} = BZ_t.$$

Desta maneira, a equação (1) pode ser reescrita em função do operador B , da seguinte forma:

$$\Delta Z = Z_t - Z_{t-1} = Z_t - BZ_t = (1 - B)Z_t.$$

Para iniciar a análise dos dados de uma série temporal é interessante a construção do seu gráfico, podendo este revelar as principais características encontradas na série, tais como: tendência, sazonalidade, variações ao longo do tempo, entre outras.

A verificação da existência de tendência na série, apenas pela análise visual não basta, pois a mesma é falha. Em geral, utiliza-se algum teste, neste trabalho utilizou-se o teste de Dickey-Fuller aumentado. Na próxima seção tem-se um estudo mais detalhado acerca deste teste.

Para construir o gráfico, o ideal é que a série não possua dados faltantes, entretanto, uma série histórica longa dificilmente conterà todos os dados, uma vez que séries

históricas são registradas por longos períodos de tempo, podendo ocorrer defeitos nos equipamentos de medidas e falhas nos registros. Para contornar este problema, a interpolação pode ser um meio bastante útil.

As ferramentas de séries temporais podem ser utilizadas na interpolação de alguns dados faltantes em meio a uma série histórica. Se os dados anteriores àquele dado faltante apresentarem padrão sazonal, leva-se em consideração uma decomposição sazonal na hora de interpolar, caso contrário, procede-se com a interpolação linear para o preenchimento dos dados.

2.1.1 Teste de Dickey-Fuller Aumentado

O teste de Dickey-Fuller Aumentado (ADF) (Dickey e Fuller, 1979) requer um estudo da seguinte regressão:

$$\Delta Z_t = \beta_1 + \beta_2 t + \pi Z_{t-1} + \sum_{i=1}^m \alpha_i \Delta Z_{t-i} + a_t, \quad (2)$$

em que ΔZ_t é a primeira diferença da série; ΔZ_{t-i} são os valores defasados incluídos no teste para eliminar a autocorrelação serial dos resíduos; o valor de m pode ser minimizado pelo critério de informação de Akaike - AIC (Akaike, 1974) ou pelo critério de informação bayesiano - BIC (Schwarz, 1978) ou ainda ser feito à mão por tentativa e erro, que consiste em tentar o menor valor inteiro para m tal que os resíduos da equação (2) se tornem ruído branco; β_1 representa o intercepto; β_2 é o coeficiente de tendência determinística; π é o coeficiente de tendência estocástica e a_t é ruído branco.

Caso existam, as componentes de tendência e sazonalidade na série geralmente possuem relação. Desta forma, é preciso extrair a correlação para depois aplicar algum teste a série, com resíduos sendo ruído branco. A aplicação do teste ADF é bastante útil neste sentido, visto que, o teste incorpora uma componente que elimina a correlação.

Segundo Dickey e Fuller (1979) inicia-se a análise verificando a hipótese $H_0 : \rho = 1$ (A série possui raiz unitária, portanto possui tendência estocástica) que é equivalente a testar $H_0 : \pi = 0$ (A série possui tendência estocástica), pois a regressão é estimada com uma diferença (ΔZ_t) já aplicada a série. Se H_0 for rejeitada conclui-se que a série não possui tendência estocástica e pode-se seguir com o estudo. O segundo passo é verificar a presença de tendência determinística pela hipótese $H_0 : \beta_2 = 0$ (A série não possui tendência determinística). Neste ponto se H_0 não for rejeitada, conclui-se que, a série é livre de tendência determinística e prossegue-se com os estudos. Caso a tendência

determinística ou estocástica for detectada, a remoção de uma delas ou as duas juntas, pode ser conseguida aplicando diferenças sucessivas nos dados originais até obter uma série livre de tendência para prosseguir com a análise.

O número de diferenças aplicadas na série original para obter uma nova série livre de tendência determinística e/ou estocástica indicam a ordem d do componente de diferenças que o modelo utilizado para representar a série deve ter. Estas diferenças são denotadas por:

$$W_t = \Delta^d Z_t = (1 - B - B^2 - \dots - B^d) Z_t = \Delta^d(B) Z_t, \quad (3)$$

em que W_t é a série livre de tendência determinística, Δ^d é o operador com d diferenças e Z_t é a série original.

Outros testes que podem ser aplicados para verificar a presença de tendência na série são: teste de Wald-Wolfowitz, teste de Cox-Stuart também conhecido como teste do sinal e o teste baseado no coeficiente de correlação de Sperman. Informações mais detalhadas acerca destes testes podem ser encontradas em Morettin e Tolo (2006).

2.2 Sazonalidade

Diz-se que uma série temporal tem comportamento sazonal quando os fenômenos observados se repetem a cada período idêntico de tempo. Estes fenômenos podem ocorrer a cada hora, dia, mês ou ano. Efeitos que se repetem em tempos maiores que um ano são chamados de ciclo. Bons exemplos de sazonalidade anual são as vendas de passagens aéreas, que tem alta em todos os fins de ano e a produção de leite do Brasil, que aumenta de forma considerável em períodos que boa parte do país está farto de água, nos meses entre novembro e fevereiro.

Uma observação Z_t de uma série temporal com comportamento sazonal é caracterizada não apenas pela dependência das observações imediatamente anteriores Z_{t-1}, Z_{t-2}, \dots , mas também pela dependência com as observações de *lags* ou passos igualmente espaçados de tamanho s , representadas por Z_{t-s}, Z_{t-2s}, \dots .

Segundo Box et al. (2008) a função de autocorrelação dos resíduos não é um indicador sensível de periodicidade, pois os efeitos de sazonalidade se diluem entre várias autocorrelações, que expressam o grau de dependência entre as observações. Desta forma a ocorrência de sazonalidade e a ordem do *lag* sazonal da série livre de tendência será verificada via análise espectral, decompondo a série no domínio das frequências de Fourier

(f_i). Cryer e Chan (2008) definem a intensidade das frequências de Fourier como:

$$I(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n Z_t \cos(2\pi f_i t) \right)^2 + \left(\sum_{t=1}^n Z_t \sin(2\pi f_i t) \right)^2 \right]. \quad (4)$$

O gráfico do periodograma foi devidamente criado para a detecção de fatores periódicos, sendo que seu gráfico possui no eixo das abscissas as frequências (f_i) e no eixo das ordenadas as intensidades delas $I(f_i)$, ou seja, o periodograma representa o gráfico da equação (4).

Com base na função dada pela equação (4) e na construção de seu gráfico, obtém-se o período sazonal s por meio da fórmula $s = 1/f_c$ sendo f_c a frequência de Fourier crítica, ou seja, aquela frequência responsável pela maior intensidade $I(f_c)$.

Segundo Morettin e Tolo (2006) mesmo que o periodograma apresente diversos picos, não é possível concluir, a priori, que cada pico corresponde a uma componente periódica na série Z_t . Desta forma a significância estatística do período sazonal s encontrado deve ser avaliada, o que pode ser feito pelo teste G de Fisher (Fisher, 1929) que possui a seguinte estatística de teste:

$$G_{calc} = \frac{\max(I(f_i))}{\sum_{i=1}^{n/2} I(f_i)}, \quad (5)$$

em que $I(f_i)$ é definido na equação (4).

Morettin e Tolo (2006) afirmam que uma boa aproximação do valor p para o teste G de Fisher é dado por:

$$P(G > G_{calc}) \approx n(1 - G_{calc})^{n-1},$$

em que n é o número de observação da série e G_{calc} é dado pela expressão (5).

Supondo Z_t gaussiano, Fisher (1929) apresenta uma forma para testar a hipótese

$$H_0 : R_i \equiv 0 \text{ (Não existe periodicidade)}$$

baseado na estatística G_{calc} dada na equação (5). Caso seja detectado sazonalidade de lag s , para prosseguir com o estudo é preciso aplicar diferenças sazonais de tamanho s até obter estacionariedade.

A diferença sazonal é bastante utilizada para remover a sazonalidade de uma série temporal. Esta diferença é caracterizada por possuir um lag ou espaço entre as observa-

ções. Pode-se defini-la da seguinte maneira:

$$W_t = Z_t - Z_{t-s} = Z_t - B^s Z_t = (1 - B^s) Z_t,$$

em que W_t é a nova série com uma diferença sazonal de tamanho s aplicada, Z_t é a observação no tempo t , Z_{t-s} é a observação no tempo $t - s$ e B é o operador de tempo passado.

Conforme Morettin e Toloi (2006) os modelos de regressão são ótimos para séries que apresentam sazonalidade determinística. Assim a sazonalidade pode ser estimada por:

$$\hat{S}_t = \sum_{j=1}^{12} \alpha_j d_{jt},$$

em que α_j são os parâmetros de estrutura do modelo, d_{jt} são variáveis periódicas (senos, cossenos, ou variáveis sazonais indicadoras).

Após estimar as componentes T_t e S_t obtém-se a componente aleatória subtraindo estas componentes da série original por meio da expressão:

$$Y_t = Z_t - \hat{T}_t - \hat{S}_t,$$

em que Y_t é a série livre de tendência e sazonalidade, contendo apenas a variável aleatória a_t .

Se a tendência e a sazonalidade forem eliminadas da série por meio das diferenças simples e sazonal obtém-se uma série estacionária contendo apenas a componente aleatória.

Segundo João (2009) a componente aleatória ou resíduo, refere-se não somente aos componentes esporádicos resultantes de eventos aleatórios imprevisíveis, mas também todos os movimentos da série que não foram passíveis de identificação pelos componentes de tendência e sazonalidade, pois não obedecem a nenhuma lei comportamental capaz de ser descrita de forma determinística por meio de relações exclusivamente matemáticas.

Outros testes que podem ser aplicados para verificar a presença de sazonalidade na série são: teste de Kruskal-Wallis, teste de Friedman ou análise de variância. Mais informações acerca destes testes podem ser obtidas em Morettin e Toloi (2006).

2.3 Seleção de modelos

Uma metodologia comumente utilizada em análise de séries temporais é a metodologia de Box e Jenkins (Box et al., 2008). Tal metodologia é fundamentada em ajustar modelos a um conjunto de dados, considerando que a construção do modelo deve ser baseada num ciclo iterativo de quatro etapas:

- Especificação: em que considera-se uma classe geral de modelos que caracterize os dados;
- Identificação: com base na função de autocorrelação (fac) e autocorrelação parcial (facp) e alguns critérios;
- Estimação: os parâmetros do modelo identificado anteriormente são estimados por métodos de máxima verossimilhança condicional, não-condicional e exata, para mais detalhes ver Morettin e Tolo (2006);
- Verificação: é realizada uma análise de resíduos do modelo para verificar sua adequabilidade.

2.3.1 Classes de modelos de Box e Jenkins

Para análise de séries temporais tem-se diversos modelos. A seguir estão definidos os mais utilizados na metodologia de Box e Jenkins.

Os modelos autorregressivos (*AR*) foram criados com base na ideia de que as observações presentes em uma série Z_t podem ser representadas se valendo apenas das p observações anteriores. Box et al. (2008) definem o modelo $AR(p)$ como:

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t, \quad (6)$$

em que μ é o intercepto do modelo, ϕ_i são os parâmetros de estrutura do modelo com $i = 1, 2, \dots, p$ e a_t é ruído branco.

É possível reescrever a equação (6) usando o operador de tempo passado B . Desta forma tem-se:

$$Z_t = \mu + \phi_1 B Z_t + \phi_2 B^2 Z_t + \cdots + \phi_p B^p Z_t + a_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Z_t = \mu + a_t$$

$$\phi_p(B) Z_t = \mu + a_t$$

em que

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (7)$$

é o polinômio autorregressivo de ordem p .

O processo linear de médias móveis de ordem q denotado por $MA(q)$ é uma combinação linear do ruído branco. O modelo de médias móveis é definido segundo Morettin e Toloi (2006) por:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad (8)$$

em que μ é o intercepto do modelo, θ_i são os parâmetros de estrutura do modelo com $i = 1, 2, \dots, q$ e a_t é ruído branco.

É possível reescrever a equação (8) usando o operador B , assim tem-se:

$$Z_t = \mu + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t$$

$$Z_t = \mu + \theta_q(B) a_t,$$

em que

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (9)$$

é o operador de médias móveis de ordem q .

O modelo $ARMA(p, q)$ definido em Box et al. (2008) como a junção do $AR(p)$ e $MA(q)$. Para uma série que não apresenta tendência nem sazonalidade, este modelo se mostra uma boa opção no processo de seleção dos modelos. O modelo $ARMA(p, q)$ é dado por:

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

assim

$$Z_t - \phi_1 B Z_t - \phi_2 B^2 Z_t - \dots - \phi_p B^p Z_t = \mu + a_t - \theta_1 B a_t - \theta_2 B^2 a_t - \dots - \theta_q B^q a_t$$

e finalmente

$$\phi_p(B) Z_t = \mu + \theta_q(B) a_t, \quad (10)$$

em que μ é o intercepto do modelo, B é um operador de retardo no tempo, $\phi_p(B)$ definido

na equação (7) é o polinômio autorregressivo de ordem p , $\theta_q(B)$ definido na equação (9) é o polinômio de médias móveis de ordem q e a_t é o ruído branco.

Segundo Box et al. (2008) o modelo autorregressivo integrado de médias móveis $ARIMA(p, d, q)$ é apropriado para séries temporais em que o processo estocástico apresenta tendência. Assim a série deve passar por um número d de diferenças em seu processo de ajuste, até eliminar a tendência.

Dada uma série Z_t que possui apenas a componente de tendência (T_t) e ruído branco (a_t), pode-se obter uma nova série W_t com comportamento estacionário conforme definido anteriormente na equação (3). A nova série não possui tendência, podendo ser modelada por uma classe $ARMA(p, q)$ já definida anteriormente, na equação (10), porém o intercepto do modelo é zero ($\mu = 0$) devido a diferença aplicada. Desta maneira, Z_t é uma série integrada de W_t , daí diz-se que Z_t é definida por um modelo autorregressivo integrado de médias móveis $ARIMA(p, d, q)$ dado por:

$$\Delta^d(B)Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

e reescrevendo-a utilizando o operador B tem-se:

$$\phi_p(B) \Delta^d(B)Z_t = \theta_q(B) a_t,$$

em que $\phi_p(B)$ dado pela equação (7) é o polinômio autorregressivo de ordem p , $\theta_q(B)$ dado pela equação (9) é o polinômio de médias móveis de ordem q , a_t é o ruído branco e $\Delta^d(B) = 1 - B - B^2 - \dots - B^d$, é o polinômio de diferenças, em que d representa o número de diferenças que tornou a série estacionária.

O modelo $SARIMA(p, d, q) \times (P, D, Q)$ autorregressivo integrado de médias móveis sazonal, também conhecido como ARIMA sazonal, é bastante utilizado quando tem-se tendência e sazonalidade presentes na mesma série, características bastantes comuns em séries climatológicas e econômicas. Além disso, o modelo ainda leva em conta os operadores autorregressivos e de médias móveis de ordens p e q , respectivamente, os operadores autorregressivos e médias móveis sazonal tem ordem P e Q , respectivamente e as diferenças simples e sazonal ordem d e D , respectivamente.

Box e Jenkins realizaram uma generalização no modelo ARIMA para trabalhar com o fator sazonal, dando origem ao modelo SARIMA definido em Box et al. (2008)

como:

$$\phi_p(B)\Phi_P(B)\Delta^d(B)\Delta_s^D Z_t(B) = \theta_q(B)\Theta_Q(B)a_t,$$

em que $\phi_p(B)$ é o polinômio autorregressivo, $\theta_q(B)$ é o polinômio de médias móveis, $\Phi_P(B)$ é o polinômio autorregressivo sazonal, $\Theta_Q(B)$ é o polinômio de médias móveis sazonal, $\Delta^d(B)$ é o operador de diferenças simples, $\Delta_s^D(B)$ é o operador de diferenças sazonais de lag com tamanho s e a_t é o ruído branco.

Após aplicar os testes para detectar a presença de tendência e/ou sazonalidade será possível especificar a classe de modelos a ser considerada. Com a série livre de tendência e sazonalidade a identificação da ordem do modelo pode ser obtida analisando os picos significativos dos gráficos da função de autocorrelação (fac) e autocorrelação parcial (facp). Porém, segundo Makridakis et al. (2008) obter a ordem exata dos parâmetros é uma tarefa difícil e subjetiva, pois há um número grande de modelos que podem ser ajustados. Se a ordem de um modelo for obtida com base nos gráficos da fac e facp pode haver ordens diferentes para mesma série, dependendo do rigor utilizado por cada autor.

A função de autocorrelação (fac) é definida segundo Box et al. (2008) por:

$$\rho_\tau = \frac{\text{Cov}(Z_t, Z_{t+\tau})}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_{t+\tau})}} = \frac{\text{Cov}(Z_t, Z_{t+\tau})}{\text{Var}(Z_t)} = \frac{\gamma_\tau}{\gamma_0},$$

em que τ é um número inteiro.

A função de autocorrelação pode ser estimada por:

$$\hat{\rho}_K = \frac{\sum_{t=K+1}^N (Z_t - \bar{Z})(Z_{t-K} - \bar{Z})}{\sum_{t=1}^N (Z_t - \bar{Z})^2},$$

em que $\bar{Z} = \frac{1}{N} \sum_{i=1}^N Z_i$ e K é o número de defasagens utilizados para realizar a estimação.

Box et al. (2008) definem a facp como sendo a medida da correlação entre duas observações Z_t e Z_{t+k} , eliminando os termos intermediários. A autocorrelação parcial é definida por:

$$\phi_{kk} = \text{Corr}(Z_t, Z_{t+k} | Z_{t+1}, \dots, Z_{t+k-1}).$$

Uma maneira de estimar a função de autocorrelação parcial é resolvendo as equações do sistema de Yule-Walker, que é um método trabalhoso, envolvendo operações de matrizes. Para mais detalhes acerca do método veja Morettin e Tolo (2006) e Box et al. (2008).

A fac e $facp$ medem a dependência entre as observações da série, indicando em seus gráficos os maiores picos para observações com maiores dependências entre si, dando indícios das ordens dos modelos.

Devido a complexidade de identificar modelos com base na fac e $facp$, a seleção pode ser realizada por meio dos critérios de informação, tais como, o critério de Akaike (AIC) (Akaike, 1974), o Akaike corrigido (AICc) (Sugiura, 1978) e o critério de informação Bayesiano (BIC) (Schwarz, 1978). Estes critérios podem ser explicados como uma forma de ponderar um modelo baseando na sua adequabilidade e na sua ordem. O modelo escolhido será aquele que apresentar menores valores para os critérios.

Akaike (1974) propôs seu critério da seguinte forma:

$$AIC = -2 \ln L(\hat{\theta}) + 2p, \quad (11)$$

em que $L(\hat{\theta})$ é a função de verossimilhança avaliada no estimador de máxima verossimilhança ($\hat{\theta}$) e p é o número de parâmetros.

Sugiura (1978) propôs uma correção para a equação (11) da seguinte forma:

$$AICc = AIC + \frac{2p(p+1)}{n-p-1},$$

em que n é o número de observações da série.

Schwarz (1978) propôs o BIC dado por:

$$BIC = -2 \ln L(\hat{\theta}) + p \ln(n).$$

Emiliano et al. (2014) avaliaram os critérios AIC, AICc e BIC por meio de simulações e concluíram que o critério BIC apresenta melhores resultados na comparação de modelos de séries temporais, tendo atingido estabilidade com índice de acerto de 96% contra 85% do AIC e AICc. Desta forma utilizaremos o critério BIC para seleção de modelos. De maneira complementar utilizaremos o AICc, ao invés do AIC por este critério penalizar com maior rigor os modelos que envolvem maior número de parâmetros.

O último passo do processo iterativo de Box e Jenkins consiste na verificação. Segundo Tsay (2002) um modelo ajustado deve ser cuidadosamente examinado para verificar se há alguma inadequação.

Segundo Sáfadi (2004) para um bom ajuste de um modelo de Box e Jenkins é necessário se valer das suas técnicas de tal forma que a estrutura residual seja um ruído

branco, ou seja, o resíduo seja uma variável aleatória independente e identicamente distribuída.

Qualquer insuficiência detectada neste estágio do estudo pode sugerir um novo modelo como sendo o adequado. O teste de de Ljung e Box (Ljung e Box, 1978) é bastante utilizado na verificação do modelo. O teste de Shapiro-Wilk (Shapiro e Wilk, 1965) pode ser aplicado para verificar se os resíduos da série provém de uma distribuição normal.

Aplica-se o teste de Shapiro-Wilk (Shapiro e Wilk, 1965) para verificar a normalidade dos resíduos. O teste possui as hipóteses:

$$\begin{cases} H_0 : & \text{A amostra provém de uma população normal.} \\ H_1 : & \text{A amostra não provém de uma população normal.} \end{cases}$$

O fato do resíduo seguir ou não uma distribuição normal não é indicativo de ajuste do modelo, o teste é utilizado para realizar a construção do intervalo de confiança.

Para o caso em que H_0 não for rejeitada constrói-se um intervalo de confiança para as previsões com base na distribuição normal. Em caso de rejeição de H_0 pode-se prosseguir com os estudos, porém a construção de intervalos de confiança se torna mais complicada, utilizando intervalos de confiança aproximados ou baseados em métodos não-paramétricos, tal como, *bootstrap*.

O método de reamostragem intitulado de *bootstrap* foi proposto por Efron (1979). Segundo Cunha e Colosimo (2003) este método é útil para inferir acerca das distribuições de estatísticas de interesse, o que muitas vezes é tarefa bastante difícil de ser obtido por métodos tradicionais como os métodos exatos e assintóticos.

O *bootstrap* é um método que consiste em gerar novas amostras de mesmo tamanho da amostra original por meio de um sorteio aleatório com reposição. Este método baseia-se no princípio de usar a informação da amostra para inferir sobre a população, sendo possível a construção de intervalos de confiança e estimativas de erros.

Para verificar a adequabilidade do modelo, Box e Pierce (1970) propuseram um teste para a autocorrelação dos resíduos estimados. Este teste possui a seguinte estatística:

$$Q(K) = n \sum_{j=1}^K \hat{r}_j^2, \quad (12)$$

em que K é o número de defasagens que foram utilizadas para estimar a função de autocorrelação, n é o número de observação da série e \hat{r}_j^2 é o valor estimado para o j -ésimo

coeficiente de autocorrelação. Caso o modelo que está sendo verificado seja adequado a equação (12) tem distribuição de qui-quadrado com K graus de liberdade.

Ljung e Box (1978) defendem a ideia de que a distribuição de qui-quadrado não fornece uma aproximação suficientemente rigorosa para a distribuição da estatística Q sob a hipótese nula, e propuseram uma generalização do teste de Box-Pierce com um fator de correção. O teste de Ljung e Box possui as mesmas hipóteses e interpretações do teste de Box-Pierce, tendo alterada apenas a sua estatística que é dada por:

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)},$$

em que K é o número de defasagens que foram utilizadas para estimar a função de autocorrelação, n é o número de observação da série e \hat{r}_j^2 é o valor estimado para o j -ésimo coeficiente de autocorrelação. Caso o modelo seja apropriado esta estatística terá distribuição qui-quadrado com $K - \nu$ graus de liberdade, em que ν é o número de parâmetros usados no modelo ajustado. Segundo Morettin e Tolo (2006), em geral, basta testar as 10 ou 15 primeiras autocorrelações \hat{r}_j^2 .

O teste de Box-Pierce ou o teste de Ljung-Box são aplicados duas vezes para verificar a adequabilidade do modelo, primeiramente nos resíduos do modelo e depois nos resíduos elevados ao quadrado. Aplicado aos resíduos o teste é usado para testar a correlação residual sob as seguintes hipóteses:

$$\begin{cases} H_0 : & \text{Os resíduos são não autocorrelacionados.} \\ H_1 : & \text{Os resíduos são autocorrelacionados.} \end{cases}$$

Espera-se com este teste não rejeitar H_0 para garantir um bom ajuste do modelo, mas se a hipótese de nulidade for rejeitada tem-se o indício de que o modelo não está bem ajustado. Sendo assim, um novo modelo alternativo deve ser proposto, ajustado e verificado, até obter o resultado desejado.

O mesmo teste aplicado aos resíduos elevados ao quadrado é utilizado para testar a homocedasticidade das variâncias com as seguintes hipóteses:

$$\begin{cases} H_0 : & \text{O modelo possui variâncias homocedásticas.} \\ H_1 : & \text{O modelo não possui variâncias homocedásticas.} \end{cases}$$

Segundo Tsay (2002) o incremento proposto por Ljung e Box torna o teste mais poderoso, desta forma ele será utilizado neste estudo.

A não rejeição da hipótese de homoscedasticidade da variância indica adequação do modelo, visto que tal condição é um pré-requisito para as classes de modelos de Box e Jenkins apresentadas neste trabalho. Entretanto, se H_0 for rejeitada será preciso modelar os resíduos de modo que a sua heteroscedasticidade seja controlada para não interferir nos resultados obtidos pelo modelo.

Os modelos da família ARCH ou autorregressivo com heteroscedasticidade condicional foram propostos por Engle (1982) com o objetivo de estimar a variância da inflação no Reino Unido. A sua ideia fundamental é que a observação a_t da série formada pelos resíduos do modelo ajustado no instante t é não-correlacionada, mas a variância condicional possui dependência serial com as observações anteriores.

A construção de um modelo $ARCH(r)$ é realizada para modelar a heteroscedasticidade de um modelo. Tal modelagem deve ser iniciada com a construção de um modelo de Box e Jenkins ($AR, MA, ARMA, ARIMA, SARIMA, \dots$) aplicado aos dados para modelar a correlação da série e, posteriormente modelar sua variação (série dos resíduos) por meio de um $ARCH(r)$. Por exemplo, considerando um modelo $ARMA(p, q)$ que seja suficiente para remover a correlação serial, mas que possua heteroscedasticidade na variância, tem-se:

$$\phi(B)Z_t = \mu + \theta(B)a_t,$$

com $a_t \sim ARCH(r)$.

Segundo Morettin e Toloi (2006) as estimativas dos parâmetros do modelo $ARCH(r)$ são obtidos pelo método da máxima verossimilhança condicional, pois os resíduos não são independentes.

Um modelo $ARCH(r)$ é definido por Morettin e Toloi (2006) como:

$$a_t = \sqrt{h_t}\varepsilon_t,$$

com

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_r a_{t-r}^2,$$

assim

$$a_t = \sqrt{(\alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_r a_{t-r}^2)}\varepsilon_t,$$

em que a_t é a observação do resíduo no instante t , α_i com $i = 0, 1, 2, \dots, r$ são os parâmetros do modelo a serem estimados, ε_t é o resíduo que possui distribuição normal, t de Student ou ainda GVE (distribuição generalizada de valores extremos). As restrições

$\alpha_0 > 0$ e $\alpha_i \geq 0$ para $i > 0$, são necessárias para satisfazer a condição da variância não-negativa. A ordem r do modelo pode ser identificada por meio da função de auto-correlação parcial (facp) dos quadrados dos resíduos. Da mesma forma que acontece na construção de modelos para os dados, esta tarefa é difícil e subjetiva, podendo indicar ordens diferentes para uma mesma série residual. Desta maneira a obtenção da ordem r pode ser feita por comparação de modelos utilizando critérios de informação, AIC, AICc ou BIC. Assim, a variância condicional é representada por r defasagens do quadrado das observações dos resíduos passados.

Geralmente supõe-se que $\varepsilon_t \sim N(0, 1)$ ou $\varepsilon_t \sim t_\nu$, distribuição t de Student com ν graus de liberdade, sendo ν o número de parâmetros utilizados no modelo ajustado ou $\varepsilon_t \sim GEV$ em que GEV é uma distribuição generalizada de valores extremos.

Morettin e Tolo (2006) afirmam que as previsões da variância dos resíduos por meio de um modelo $ARCH(r)$ são realizadas pela expressão:

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_r a_{t-r}^2,$$

em que $\hat{h}_t(1)$ é a previsão de h_{t+1} com origem fixada no instante t .

Os modelos $ARCH(r)$ possuem uma generalização, que são os modelos da classe $GARCH(r, s)$ proposta por Bollerslev (1986). Embora esta família de modelos sejam uma generalização, eles podem ser usados para descrever a variância residual de uma série utilizando menos parâmetros que um modelo $ARCH(r)$. Segundo Lima et al. (2010) a variância condicional da série pode depender do quadrado de observações passadas de tempos muito distantes, ou seja, o r pode ser elevado, desta forma o uso do modelo $ARCH(r)$ resultaria numa estimação de um número grande de parâmetros.

O modelo $GARCH(r, s)$ é definido segundo Morettin e Tolo (2006) por:

$$a_t = \sqrt{h_t} \varepsilon_t,$$

em que

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j},$$

ou seja,

$$a_t = \sqrt{\left(\alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j} \right)} \varepsilon_t,$$

a_t é a observação do resíduo no instante t , α_i com $i = 0, 1, 2, \dots, r$ e β_j com $j =$

$1, 2, \dots, s$ são os parâmetros do modelo a serem estimados; ε_t é o resíduo que possui distribuição normal, t de Student ou ainda GVE (distribuição generalizada de valores extremos). As restrições $\alpha_0 > 0$, $\alpha_i \geq 0$ para $i > 0$, $\beta_i \geq 0$ para $j > 0$, são necessárias para satisfazer a condição da variância não-negativa. E ainda $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$, $q = \max(r, s)$.

O ajuste de um modelo da família $GARCH(r, s)$ é realizada da mesma forma que foi detalhado para o ajuste de um modelo $ARCH(r)$.

Tanto modelos da família $ARCH(r)$ quanto $GARCH(r, s)$ após serem ajustados podem ser verificados por meio de análise residual. Se uma destas duas classes de modelos for ajustada a uma série temporal a_t , o intuito é que a série seja não-correlacionada, porém haja dependência serial. Conforme Morettin e Tolo (2006) seus resíduos padronizados são dados por:

$$\tilde{a}_t = \frac{a_t}{\sqrt{h_t}},$$

formando uma sequência independente e identicamente distribuída. Desta forma a adequação do modelo pode ser verificada por algum teste de normalidade, por exemplo o teste de Shapiro-Wilk já mencionado na página 17 e, também um teste de aleatoriedade, tais como o teste de Ljung-Box, aplicado aos resíduos elevados ao quadrado, para concluir acerca da adequabilidade do modelo.

As medidas de erro ou de acurácia de um modelo são estatísticas interessantes de serem calculadas, pois são bastante úteis na comparação de modelos, principalmente por expressarem o quão distante estão as previsões da realidade. Existem diversas medidas de erro para verificar a qualidade de ajuste dos modelos aos dados, neste trabalho foram utilizados o indicativo de erro quadrático médio (Mean Squared Error - MSE) e o indicativo de erro percentual médio absoluto (Mean Absolute Percentual Error - MAPE).

As estatísticas das medidas de acurácia são definidas em Tsay (2002) por:

$$MSE = \frac{1}{h} \sum_{j=t}^{t+h} (Z_j - \hat{Z}_j)^2;$$

$$MAPE = \frac{1}{h} \sum_{j=t}^{t+h} \left| \frac{Z_j - \hat{Z}_j}{Z_j} \right| \times 100\%,$$

em que Z_t são as observações, \hat{Z}_t são as estimativas das observações e h é o número de observações que foram previstas. Note que para um modelo perfeito as medidas de acurácia resultam em zero, pois \hat{Z}_t é igual a Z_t , desta forma quanto menores forem os

valores de MSE e MAPE tem-se indicativo de maior proximidade dos valores estimados aos observados, resultando em um bom ajuste do modelo.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Notre Dame, v. 19, n. 6, p. 717-723, 1974.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, North Holland, v. 31, n. 3, p. 307-327, 1986.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis forecasting and control**. 4. ed. Hoboken: John Wiley & Sons, 2008. 746 p.

BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelation in autoregressive-integrated moving average Time Series models. **Journal of the American Statistical Association**, Alexandria, v. 65, n. 332, p. 1509-1526, 1970.

CRYER, J. D.; CHAN, K. S. **Time series analysis with applications in R**. 2. ed. Nova York: Springer, 2008. 492 p.

CUNHA, W. J.; COLOSIMO, E. A. Intervalos de confiança bootstrap para modelos de regressão com erros de medida. **Revista de Matemática e Estatística**, Jaboticabal, v. 21, n. 2, p. 25-41, 2003.

DICKEY, D.; FULLER, W. A. Distribution of the estimates for autoregressive Time Series with a unit root. **Journal of the American Statistical Association**, Alexandria, v. 74, n. 366, p. 427-431, 1979.

EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, v. 7, n. 1, p. 1-26, 1979.

EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S. Information criteria: How do they behave in different models?. **Computational Statistics and Data Analysis**, v. 69, n. 1, p. 141-153, 2014.

ENGLE, F. R. Autorregressive conditional heteroscedasticity with estimates of the variance United Kingdom inflation. **Econometrica**, v. 50, n. 4, p. 987-1008, 1982.

FISHER, R. A. Tests of significance in harmonic analysis. **Proceeding of the Royal Societ**, Series A, v. 125, n. 796, p. 54-59, 1929.

JOÃO, M. M. **Análise do comportamento das temperaturas máximas e mínimas médias mensais para o estado do Rio Grande do Sul**. 2009. 137 p. Dissertação (Mestrado em Meteorologia) - Programa de pós-graduação da Faculdade de Meteorologia, Universidade Federal de Pelotas, Pelotas. 2009.

LIMA, F. G.; KIMURA, H.; NETO, A. A.; PERERA, L. C. J. Previsão de preços de commodities com modelos ARIMA-GARCH e redes neurais com ondaletas: velhas tecnologias-novos resultados. **Revista de Administração**, São Paulo, v. 45, n. 2, p. 188-202, 2010.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, Oxford, v. 65, n. 2, p. 297-303, 1978.

MAKRIDAKIS, S. G.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: methods and applications**. 3. ed. Nova Jersey: Wiley, 2008. 656 p.

MARENCO, J. A. Mudanças climáticas globais e regionais: avaliação do clima atual do Brasil e projeções de cenários climáticos do futuro. **Revista brasileira de meteorologia**, São José dos Campos, v. 16, n. 1, p. 1-18, 2001.

_____. Impactos de extremos relacionados com o tempo e o clima - Impactos sociais e econômicos. **Boletim do Grupo de Pesquisas em Mudanças Climáticas - GPMC**, Edição Especial, n. 8, Mai. 2009.

MINUZZI, R. B.; VIANELLO, R. L.; SEDIYAMA, G. C. Oscilações climáticas em Minas Gerais. **Revista Brasileira de Meteorologia**, São José dos Campos, v. 25, n. 2, p. 227-236, 2010.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. Hoboken: John Wiley & Sons, 2008. 456 p.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: Edgard Blücher, 2006. 538 p.

SÁFADI, T. Uso de séries temporais na análise de vazão de água na represa de Furnas. **Ciência e agrotecnologia**, Lavras, v. 28, n. 1, p. 142-148, 2004.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, Hayward, v. 6, n. 2, p. 461-464, 1978.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Oxford, v. 52, n. 3/4, p. 591-611, 1965.

SUGIURA, N. Further analysts of the data by akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, Ontario, v. 7, n. 1, p. 13-26, 1978.

TSAY, R. S. **Analysis of financial time series**. Hoboken: Jonh Wiley & Sons, 2002. 448 p.

CAPÍTULO 2

Estudo da média mensal das temperaturas máxima e mínima de Viçosa-MG por meio de séries temporais

RESUMO

Com este trabalho objetivou-se principalmente realizar uma análise da série histórica da média mensal das temperaturas máximas e mínimas da cidade de Viçosa-MG, valendo-se da análise de séries temporais. Verificou-se o efeito da tendência e sazonalidade contidas nos dados históricos que são registrados desde janeiro de 1968, além de identificar possíveis alterações climáticas e realizar previsões para os próximos 12 meses. Para análise e modelagem da série, foram escolhidos diversos modelos da classe SARIMA e o que melhor se ajustou, dentre os modelos concorrentes, foi selecionado por meio do critério de informação bayesiano (BIC) e confirmado pelo critério de informação de Akaike corrigido (AICc). O modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$ apresentou melhor ajuste para a série da temperatura máxima, enquanto que a temperatura mínima teve sua correlação serial modelada pelo modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$ e seus resíduos pelo modelo $ARCH(2)$, pois o resíduo da série das temperaturas mínimas não é ruído branco. Com estes modelos foi possível realizar previsão das médias mensais das temperaturas máximas e mínimas para os próximos 12 meses, além de identificar uma elevação significativa na temperatura da cidade de Viçosa desde o início dos registros.

Palavras-chave: Aquecimento. Clima. SARIMA. ARCH. Previsão.

1 INTRODUÇÃO

O estudo climatológico, em especial das temperaturas vem ganhando cada vez mais espaço entre os pesquisadores. Este fenômeno vem causando inquietações entre os estudiosos devido aos indícios de aquecimento global e suas consequências que ainda são, em parte, incertas.

Não se deve confundir dois termos importantes: “variabilidade climática” e “mudança climática”, segundo Steinke et al. (2005) o termo “variabilidade climática” é utilizado para descrever variações do clima em função de condições naturais. Já o termo “mudança climática” é utilizado para referenciar as alterações da variabilidade natural do clima devido as atividade humanas na Terra. Assim em um estudo acerca do clima, é necessário verificar se estão ocorrendo alterações, além da variação natural, para afirmar que estão ocorrendo mudanças climáticas.

É sabido que, o planeta Terra passa por constantes variações de temperaturas, sejam elas oriundas de efeitos que ocorrem com certa frequência e são passíveis de previsão, ou de maneira aleatória tornando difícil suas previsões. Não é fácil e nem pode-se concluir que certas alterações climatológicas ocorrem por influências da humanidade, visto que, as temperaturas de um determinado ano nunca são idênticas a de anos anteriores. Para conclusões neste sentido é preciso averiguar uma mudança que se concretize ao longo de vários anos e não apenas em ocorrências esporádicas.

Segundo Silva et al. (2008) é inegável que o sucesso ou insucesso de determinado empreendimento está diretamente ligado às condições da temperatura em uma região, país ou planeta.

Do ponto de vista da agricultura as temperaturas extremas, tanto baixas, quanto altas, são preocupantes, uma vez que a população de uma determinada localidade está habituada a cultivar plantas e criar animais que são adaptados para temperaturas que geralmente ocorrem naquela região. Uma variação extrema na temperatura certamente irá prejudicar a produção. Por exemplo, a “geada negra” que atingiu o norte do estado do Paraná, em julho de 1975, causou grande estrago na plantação de café daquele estado, que era até então o maior produtor do país. Segundo Priori et al. (2012) depois de verem seus cafezais condenados pela geada, grande parte dos cafeicultores paranaenses optaram pela produção diversificada como a soja e o milho.

Do ponto de vista da produção e reprodução de animais, a preocupação com o clima não é diferente. Quando os assuntos são as temperaturas máximas e mínimas o

cuidado se volta para o bem estar animal. Desta forma, a temperatura muito elevada ou muito baixa se torna um problema. Com as altas temperaturas o produtor precisa investir cada vez mais para manter o conforto dos animais, tornando o custo de produção mais elevado. Nos casos em que não há o investimento no conforto do animal, a quantidade e qualidade da produção sofrem uma notável queda. Salgado e Nääs (2010) em um estudo acerca do risco da produção de aves de corte no estado de São Paulo, em função da temperatura ambiente, afirmam que a perda na produção, motivada pelas temperaturas, são potencialmente grandes, principalmente ocasionadas pelas altas temperaturas, que podem causar a morte de alguns animais.

Do ponto de vista do setor de fornecimento de energia, as temperaturas também são de grande importância, uma vez que, em dias ou meses com temperaturas mais elevadas o consumo de energia elétrica aumenta, sendo esta alta motivada, principalmente, pelo uso de aparelhos de ar condicionado e maior exigências de aparelhos utilizados para resfriamento e/ou congelamento. Portanto, é de grande importância para o setor energético um estudo acerca da elevação das temperaturas para que se possa ter um controle, planejamento e previsão no fornecimento de energia.

Séries climatológicas, em geral, apresentam dois aspectos bem comuns: a dependência serial entre as observações, ou seja, a temperatura de hoje é influenciada em boa parte pela temperatura dos dias anteriores e, o fator sazonal que é caracterizada pelo comportamento parecido das temperaturas, nas mesmas épocas, de anos diferentes. Desta forma, um estudo por meio de séries temporais pode ser de grande valia, haja vista que esta área do conhecimento possui boas ferramentas para modelar tais aspectos.

Nota-se diversos estudos acerca das mudanças climáticas por meio de séries temporais. Em vários artigos publicados, tais como Marengo (2001), Minuzzi et al. (2010), Gandra et al. (2014) dentre outros, é possível notar elevação significativa na temperatura para o período analisado. Entretanto, é preciso tomar cuidado ao tentar explicar tal elevação na temperatura creditando o fenômeno ao aquecimento global, uma vez que, estudos realizados em algumas cidades, ou regiões isoladas, não servem como base para inferir acerca de fenômenos globais. Para verificar características globais é necessário a utilização de dados gerais, de diversas regiões do planeta.

Marengo (2009) afirma que estudos realizados na América do Sul seguem a tendência mundial, indicando pequeno acréscimo nas temperaturas ao longo dos anos. Estudos no Brasil também vão no mesmo sentido, segundo Marengo (2001) alguns estudos detectaram aumentos sistemáticos de temperatura do ar, em diversas regiões brasileiras.

Este aquecimento foi associado, pelo autor, ao fenômeno de urbanização que o país vem passando e também as causas naturais, por exemplo, o aquecimento das águas do Oceano Atlântico Sul.

Minuzzi et al. (2010) em um estudo climatológico para diversas cidades de diferentes regiões de Minas Gerais, afirmam que o estado também segue a tendência global, indicando acréscimo de temperaturas em quase todas as cidades que foram estudadas.

Gandra et al. (2014) utilizaram a metodologia de Box e Jenkins para modelar as séries temporais das médias anuais das temperaturas mínimas, médias e máximas da região de Pelotas-RS no período de 1931 até 2011 detectando elevação de 1,12 °C nas médias anuais das temperaturas mínimas a partir do ano de 1970.

Steinke et al. (2005) em um estudo realizado com dados obtidos em diversas estações meteorológicas da região do Distrito Federal afirmam que as mudanças na temperatura, nem sempre, estão ligadas ao aquecimento global, visto que a urbanização, a redução dos dias chuvosos, entre outros fenômenos podem influenciar na alta das temperaturas. Os autores confirmam esta afirmação baseando-se no resultado obtido, em que detectaram elevação das temperaturas em algumas estações, enquanto que em outras não houve tal elevação, logo não foi possível associar o aumento da temperatura a eventos globais e sim a fenômenos ocorridos em algumas microrregiões.

Blain et al. (2009) realizaram um estudo das temperaturas mínimas de algumas cidades do estado de São Paulo, com objetivo de verificar se, ambas sofriam mudanças parecidas e se elas ocorriam nas mesmas épocas, para que pudessem confirmar a presença de um efeito global sobre o estado. Entretanto os autores concluíram que foi detectado aumento de temperatura em algumas cidades nas mesmas épocas, mas isso não foi geral. Eles acreditam que esta elevação não tenha sido causada pelo efeito do aquecimento global e sim por algumas ondas de calor sobre uma certa região, provavelmente influenciadas por efeitos de urbanização.

Blain (2010) em um estudo realizado para as séries temporais de temperaturas máximas de oito cidades de São Paulo detectou elevação de temperatura significativa em quatro delas, enquanto que as outras tiveram temperaturas constantes ou até mesmo pequena tendência de queda na temperatura. Confirmando com isto que, não houve, uma mudança de temperatura significativa para todas as regiões, concluindo que, não é possível afirmar que tal elevação de temperatura foi influenciada por algum fenômeno global e sim microrregional.

Neste sentido, objetiva-se estudar, de forma mais específica, e restrita, as médias

mensais das temperaturas máximas e mínimas da cidade de Viçosa-MG com o intuito de identificar possíveis alterações, verificando se a cidade também segue a tendência de aquecimento vista em algumas regiões. Objetiva-se ainda fazer uma análise de cada mês, em anos diferentes, para detectar quais meses sofrem mais com a variação climatológica e finalmente realizar previsões dos próximos 12 meses das médias mensais das temperaturas máximas e mínimas da cidade de Viçosa-MG. Este estudo, juntamente com as previsões, podem ser utilizados como um meio para tomadas de decisões de setores que dependem das temperaturas direta e indiretamente. Quanto a população de Viçosa e região é interessante que se tenha melhor compreensão do comportamento do clima da cidade.

2 MATERIAL E MÉTODOS

As séries temporais sob estudo referem-se aos dados diários acerca da temperatura mínima e máxima da cidade de Viçosa-MG. Estas duas características foram extraídas do banco de dados do clima da cidade que é registrado diariamente na estação meteorológica da Universidade Federal de Viçosa no departamento de Engenharia Agrícola (UFV, 2016).

Foi utilizada a interpolação para alguns dados não registrados, que ocorreram por defeito ou manutenção do equipamento de medida. Para as temperaturas foi calculada a média mensal, pois uma série com 365 observações anuais se torna mais complicada de ser modelada, acarretando muitos erros. Calculando a média mensal a série passa a ter 12 observações anuais, o que torna mais fácil a manipulação dos dados com menor erro de modelagem e previsão.

Após transformar os dados em observações mensais, realizou-se a construção do gráfico da série objetivando perceber as suas principais características, tais como tendência, sazonalidade e efeitos aleatórios. Construiu-se ainda um gráfico de cada componente da série, com intuito de analisar o comportamento de cada uma separadamente. Com estes gráficos, realizou-se uma análise visual e, a partir dela, foi possível ter uma indicação de quais testes seriam viáveis para verificar algumas características da série e ainda ter uma visão prévia de qual classe dos modelos de Box e Jenkins seria a mais adequada para os dados.

Para obter um entendimento mais detalhado acerca das temperaturas, foi realizada uma análise de cada mês em diferentes anos, com intuito de verificar em quais deles, ou em quais estações do ano, ocorreram maiores alterações nas temperaturas.

A construção do box-plot de cada mês da série para anos diferentes foi realizado para identificar a variação presente em cada mês, verificando aqueles que apresentam maior discrepância e os que apresentam temperaturas concentradas em intervalos menores. Com este gráfico foi possível notar alguns candidatos a *outliers*.

Os gráficos da função de autocorrelação (fac) e autocorrelação parcial (facp) da série foram construídos para identificar, por meio da análise visual, uma dentre várias classes de modelos de Box e Jenkins que pode ser utilizada para modelar a série e quais as ordens de cada parâmetro. Pode-se verificar também se a série apresenta componente sazonal e, em caso positivo, detectar qual o tamanho do *lag* sazonal. A tarefa de identificação de modelos e, suas ordens, por meio dos gráficos da fac e facp não é fácil, mas se torna possível ter uma ideia de quais modelos se ajustam melhor após analisá-los.

Com o intuito de verificar a presença de tendência determinística e/ou estocástica procedeu-se o teste de Dickey-Fuller aumentado (ADF) estimando a regressão do teste aplicado aos dados. Para o caso em que este teste detectou tendência, efetuou-se uma transformação nos dados aplicando diferenças sucessivas até obter uma série livre de tendência. A série transformada e livre de tendência foi confirmada pelo mesmo teste ADF e ilustrada por meio de um gráfico em que foi possível ver a série oscilando em torno de uma reta horizontal, caracterizando comportamento estacionário.

De posse da série diferenciada e livre de tendência, analisou-se a presença de sazonalidade por meio de uma análise espectral, decompondo a série no domínio das frequências de Fourier. Para tal, foi utilizado o gráfico do periodograma que contém as frequências de Fourier dispostas no eixo das abscissas e suas intensidades representadas no eixo das ordenadas, com este gráfico foi possível identificar a sazonalidade e o período sazonal da série. O teste G de Fisher foi aplicado para testar a hipótese de significância do período sazonal encontrado.

Após identificar sazonalidade anual, componente muito comum em dados climatológicos, aplicou-se diferenças com *lag* sazonal de tamanho 12 para eliminar também este fator, obtendo enfim uma série estacionária, característica necessária para seguir com a metodologia de Box e Jenkins.

A classe de modelos $SARIMA(p, d, q) \times (P, D, Q)$ é uma generalização dos modelos de Box e Jenkins, que é apropriada para descrever séries temporais que apresentam a componente de sazonalidade, sendo capaz de modelar também a componente de tendência, caso exista.

O início do processo de modelagem da série se deu considerando o valor de d como sendo o número de diferenças sucessivas que foram necessárias para eliminar a tendência da série e, D sendo o número de diferenças sucessivas de *lag* doze que foram necessárias para eliminar a sazonalidade, tornando a série estacionária. Os demais parâmetros foram identificados criando modelos concorrentes, variando e combinando valores para os parâmetros, na busca de um modelo parcimonioso que apresentou menor valor para o critério de informação bayesiano (BIC).

O critério de informação de Akaike corrigido (AICc) foi utilizado para confirmação do melhor modelo para casos em que o BIC não apresentou grandes diferenças. As medidas de erros MSE e MAPE foram utilizadas na busca de um bom modelo para realizar previsões. No momento em que foram criados os modelos concorrentes para comparação, eles tiveram seus parâmetros estimados pelo método da máxima verossimilhança

condicional, com o auxílio do software R (2016) por meio do pacote *stats*.

Após especificar, identificar o modelo e estimar os parâmetros, chega o instante de verificar a adequabilidade do mesmo com base na análise residual. Iniciou-se a análise dos resíduos por meio do teste de Shapiro-Wilk para verificar se os mesmos são normalmente distribuídos. Enquanto que o teste de Ljung-Box foi aplicado duas vezes, a primeira nos resíduos do modelo, para verificar a correlação serial e, a segunda nos resíduos elevados ao quadrado para verificar a homocedasticidade da variância do modelo.

No teste de Ljung-Box aplicado aos resíduos elevados ao quadrado, avalia-se a hipótese de variância homogênea. Caso a hipótese seja rejeitada, há indícios de que a variância dos resíduos não possui um padrão constante, apresentando oscilações ao longo do tempo, desta forma, deve-se modelar tal variação, o que pode ser feito por meio de um modelo $ARCH(r)$.

Um bom modelo da classe ARCH foi obtido criando vários modelos concorrentes e escolhendo aquele que apresentou menor valor para o critério de informação de Akaike (AIC), visto que os critérios de seleção BIC e AICc apresentaram resultados semelhantes, indicando um bom ajuste. Modelando a variância condicional por meio de um $ARCH(r)$, os resíduos passaram a se comportar como ruído branco. Assim a série teve sua correlação serial descrita por um modelo da classe SARIMA. Já a variância condicional foi descrita por um modelo da classe ARCH que foi implementada ao modelo para realizar as previsões, tornando-as mais precisas. Este procedimento resultou em melhor ajuste e adequabilidade do modelo, conservando a pressuposição de ruído branco; condição necessária para modelagem de séries temporais.

Depois de passar pela fase de verificação da adequabilidade, os modelos ajustados para cada série temporal foram utilizados para realizar as previsões para os próximos 12 meses das médias mensais das temperaturas máximas e mínimas.

Após o ajuste dos modelos adequados, realizou-se previsões para os 12 meses seguintes, de agosto de 2016 até julho de 2017. Os dados reais de agosto de 2016 até janeiro de 2017 foram coletados e reservados para comparação com os dados preditos para analisar o quão bom ficou a previsão elaborada por meio dos modelos selecionados.

3 RESULTADOS E DISCUSSÃO

Na seção 3.1 serão apresentados os resultados obtidos do estudos de séries temporais para os dados da média mensal da temperatura máxima da cidade de Viçosa-MG, enquanto que na seção 3.2 serão apresentados os resultados para a média mensal da temperatura mínima.

3.1 Média mensal da temperatura máxima

Nesta seção serão apresentados os resultados dos estudos acerca das características da média mensal da temperatura máxima da cidade de Viçosa-MG com registros desde janeiro de 1968 a julho de 2016.

No período estudado a temperatura máxima de Viçosa apresentou uma média geral de 26,62 °C com desvio padrão de 2,10 °C. O mês que apresentou maior valor para a média mensal da temperatura máxima foi fevereiro de 2002, atingindo 32,42 °C, já o mês em que a média mensal obteve o menor valor foi em julho de 1976 em que registrou-se 22,21 °C.

No Gráfico 1 tem-se a representação da série temporal das médias mensais das temperaturas máximas da cidade de Viçosa-MG juntamente com a reta de tendência, indicando uma elevação da média mensal das temperaturas ao longo do tempo. Além disso nota-se no gráfico os picos, com frequência anual, que são indícios de sazonalidade, fator comum em dados climatológicos.

No Gráfico 2 pode-se observar a série decomposta em suas três componentes não observáveis: tendência, sazonalidade e fator aleatório. Por meio de uma análise visual verifica-se a presença de tendência positiva e um fator sazonal bem definido.

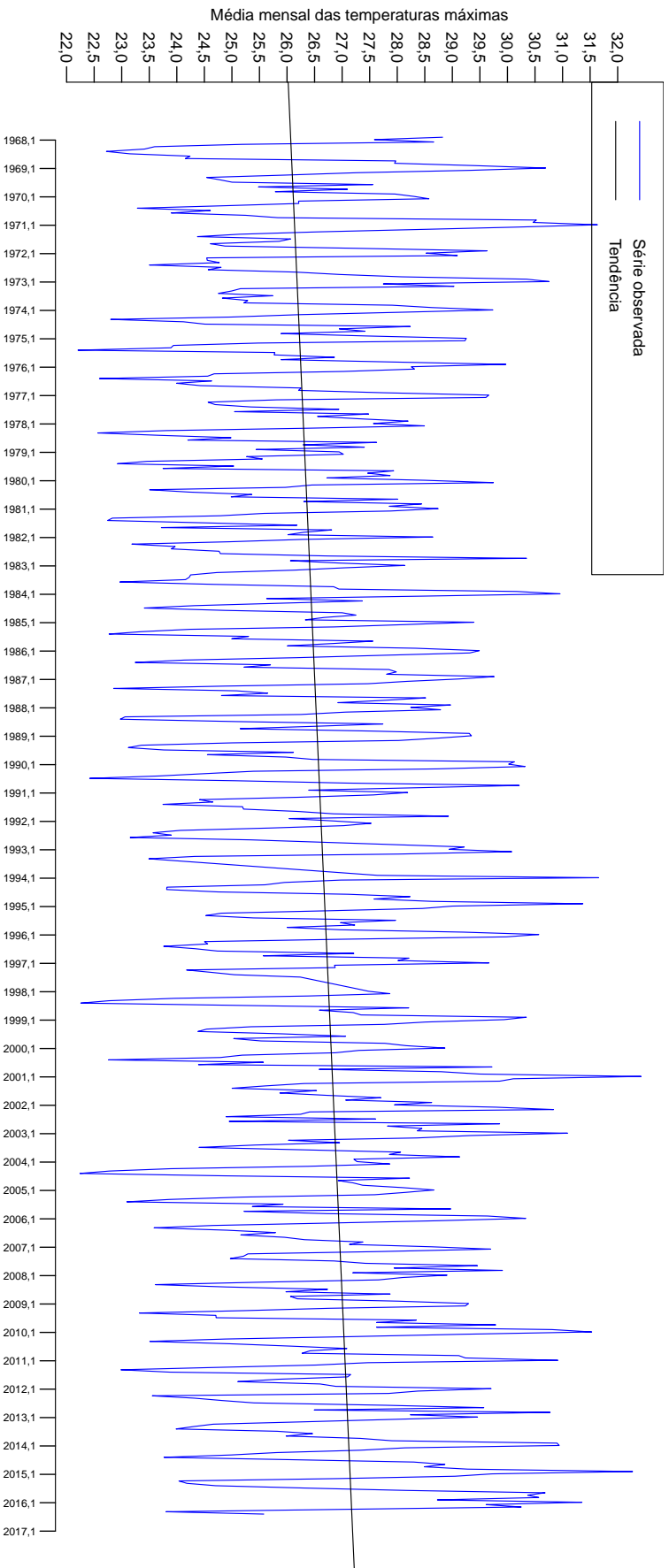


Gráfico 1 Série histórica das médias mensais das temperaturas máximas da cidade de Vigosa-MG com a reta de tendência

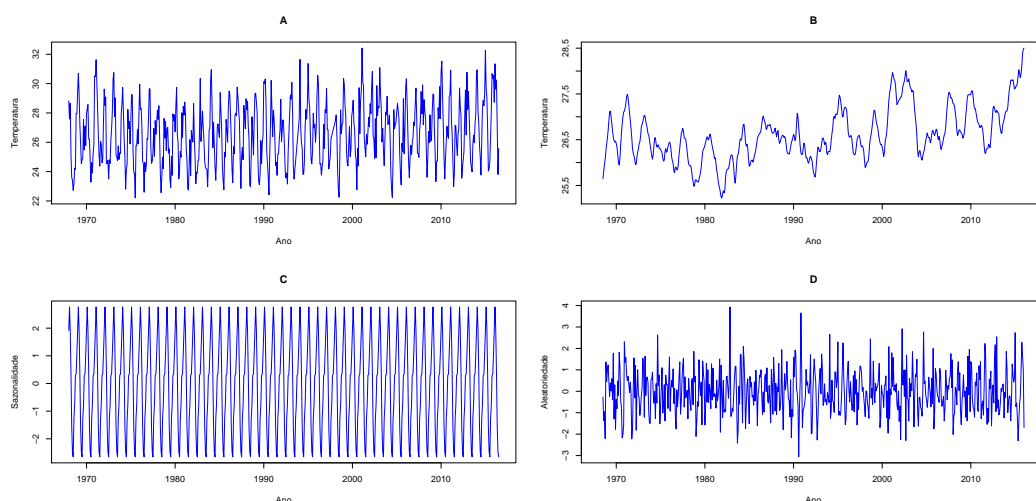


Gráfico 2 A - Série da média mensal da temperatura máxima de Viçosa-MG; B - Componente de tendência da série; C - Componente sazonal da série; D - Fator aleatório da série

Na Tabela 1 tem-se o resultado para a análise da tendência de elevação da temperatura, em que verifica-se uma alta significativa de $0,0229\text{ }^{\circ}\text{C}$ por ano na média mensal da temperatura máxima da cidade de Viçosa, resultando em uma elevação de mais de $1\text{ }^{\circ}\text{C}$ grau na média ao longo dos 48 anos e 7 meses em estudo.

Tabela 1 Resultado da análise para tendência de elevação da temperatura

	Estimativa	Desvio padrão	Estatística t	Valor p
Intercepto	-18,9880	12,2157	-1,5540	0,1206
Tempo (x)	0,0229	0,0061	3,7340	0,0002

No Gráfico 3 encontra-se o boxplot de cada mês, ao longo dos anos, da série em estudo, que ilustra o esperado, indicando que as temperaturas mais elevadas ocorrem nos meses de janeiro, fevereiro e março, enquanto que as menores ocorrem em maio, junho e julho. As maiores amplitudes térmicas podem ser vistas em janeiro, fevereiro, abril e outubro. Neste mesmo gráfico identifica-se alguns meses em certos anos que são possíveis *outliers*, entretanto não foram identificadas observações com valores extremos que fossem de fato candidatos a *outliers*, desta forma, não foram necessárias intervenções.

Na Tabela 2 encontram-se os resultados do teste de Dickey-Fuller aumentado (ADF). Verificou-se que a série em estudos não apresenta raiz unitária, que equivale afirmar que a série não possui tendência estocástica, pois a regressão é estimada com uma diferença já aplicada aos dados. A hipótese de raiz unitária foi rejeitada a 5% de probabilidade (valor $p < 0,0001$). Porém o coeficiente para tendência determinística foi significativo com valor $p = 0,0067$, ou seja, rejeitou-se a hipótese de nulidade que afirma não possuir tendência determinística na série. Desta forma a tendência determinística da

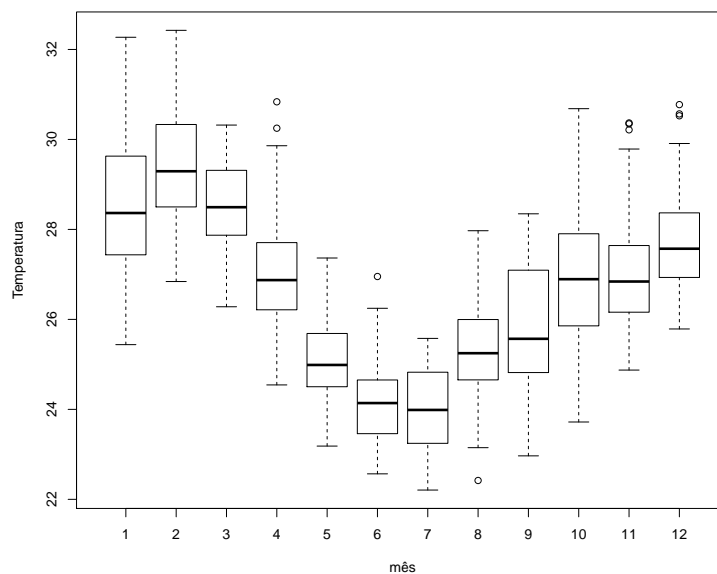


Gráfico 3 Boxplot da série com os diferentes meses da média mensal da temperatura máxima

série foi eliminada após a aplicação de uma diferença nos dados. Realizou-se o teste ADF com 13 defasagens da variável em primeira diferença, tais defasagens foram necessárias para eliminar a autocorrelação residual, que pode influenciar o teste. O teste de Ljung-Box apresentou valor $p = 0,0828$ confirmando a 5% de probabilidade que a autocorrelação residual foi eliminada, conservando a hipótese nula que, afirma que os resíduos são não autocorrelacionados.

Tabela 2 Resultado do teste de Dickey-Fuller aumentado

Parâmetros	β_1	β_2	π
Estimativa	12,6908	0,0010	-0,4874
Erro-padrão	2,8954	0,0003	0,1109
Estatística do teste	4,3830	2,7190	-4,3940
Valor p	< 0,0001	0,0067	< 0,0001
Teste Ljung-Box	$Q = 169,99$	valor $p = 0,0828$	

Após aplicação de uma diferença simples nos dados da série, procedeu-se novamente com o teste ADF, e os resultados estão apresentados na Tabela 3 em que verificou-se a série livre de tendência estocástica e determinística. Na série transformada, a aplicação do teste ADF necessitou de 15 defasagens para eliminar a autocorrelação residual.

Conforme dito anteriormente, é provável que a série em estudo apresente sazonalidade, pois se trata de uma série de dados climatológicos. Para verificar a periodicidade desta componente foi realizado o periodograma da análise espectral apresentado no Gráfico 4.

Tabela 3 Resultado do teste de Dickey-Fuller aumentado para a série após aplicação de uma diferença

Parâmetros	β_1	β_2	π
Estimativa	0,0399	0,0002	-6,1187
Erro-padrão	0,1118	0,0003	0,6779
Estatística do teste	-0,3570	0,6300	-9,0260
Valor p	0,7211	0,5287	< 0,0001
Teste Ljung-Box	$Q = 170,9100$ valor $p = 0,0757$		

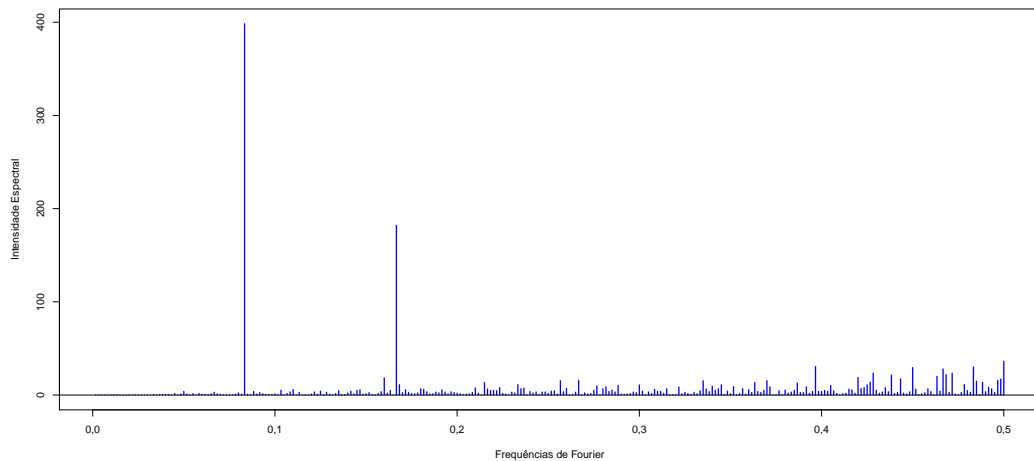


Gráfico 4 Periodograma da série da média mensal da temperatura máxima de Viçosa-MG

Com auxílio do Gráfico 4 tem-se que a maior intensidade espectral está associada a frequência de Fourier $f = 0,0833$ que fornece uma periodicidade $s = 12$. Este valor calculado foi testado pelo teste G de Fisher de onde conclui-se, baseado na estatística $G = 0,2173$, que se apresentou superior a estatística $Z = 0,0294$ com valor $p \leq 0,0001$. Portanto a série histórica das médias mensais das temperaturas de Viçosa-MG possui periodicidade de 12 meses, ou seja, sazonalidade anual. Para eliminar a sazonalidade foi aplicada uma diferença sazonal de *lag* 12, obtendo uma série transformada, que se mostrou livre de tendência e sazonalidade.

Após o tratamento dos dados foi possível aplicar a metodologia de Box e Jenkins na modelagem. Analisando o Gráfico 5, torna-se difícil ajustar um modelo baseando-se apenas nas funções de autocorrelação (fac) e na função de autocorrelação parcial (facp) das séries originais e após aplicação das diferenças. Nota-se que após aplicação das diferenças, a estrutura de autocorrelação dos dados foi relativamente minimizada, porém ainda é visível que nas defasagens múltiplas e próximas de $s = 12$ alguns coeficientes de autocorrelação e autocorrelação parcial se apresentam significativos a 5% de probabilidade, indicando que um modelo mais indicado para tal é da classe $SARIMA(p, d, q) \times (P, D, Q)$.

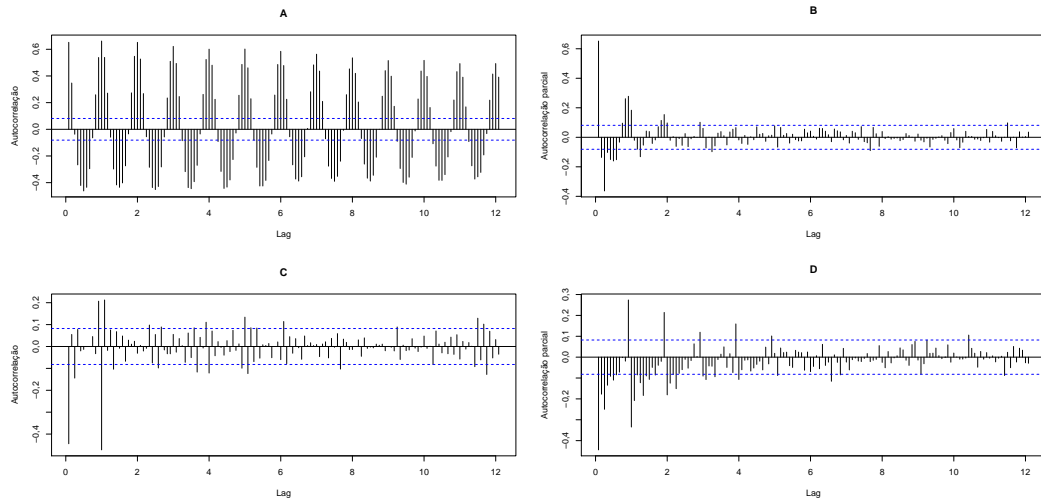


Gráfico 5 A - Fac da série original; B - Fapc da série original; C - Fac da série diferenciada; D - Fapc da série diferenciada

Como foi aplicado uma diferença simples e uma sazonal para tratamento prévio dos dados, tem-se que um bom modelo para representar a série histórica é da forma $SARIMA(p, 1, q) \times (P, 1, Q)$. Na Tabela 4 estão apresentados diversos modelos concorrentes que foram ajustados e selecionados por meio do critério de informação bayesiano (BIC) e o critério de informação de Akaike corrigido (AICc).

Tabela 4 Comparação de diversos modelos concorrentes para a série da média mensal da temperatura máxima de Viçosa-MG

Modelo	BIC	AICc	MSE	MAPE
SARIMA(0,1,0)x(0,1,0)	2398,6830	2394,3450	0,8443	1,6507
SARIMA(1,1,0)x(0,1,0)	2279,4890	2271,8190	0,4617	1,2017
SARIMA(2,1,0)x(0,1,0)	2267,8590	2254,8650	0,4078	1,0823
SARIMA(2,1,1)x(0,1,0)	2160,7950	2143,4840	0,3546	1,0768
SARIMA(2,1,2)x(0,1,0)	2164,4300	2142,8080	0,3765	1,1356
SARIMA(2,1,2)x(0,1,1)	1848,9290	1823,0050	0,2047	0,8440
SARIMA(2,1,2)x(0,1,2)	1855,2750	1825,0540	0,2046	0,8438
SARIMA(2,1,2)x(1,1,1)	1855,2750	1825,0540	0,2046	0,8438
SARIMA(2,1,2)x(2,1,1)	1861,6090	1827,1010	0,2045	0,8436
SARIMA(3,1,2)x(0,1,1)	1855,1720	1824,9520	0,2058	0,8488
SARIMA(2,1,3)x(0,1,1)	1855,0240	1824,8030	0,2072	0,8549

O modelo escolhido por meio do BIC e confirmado pelo AICc para modelar a série foi o $SARIMA(2, 1, 2) \times (0, 1, 1)$ que teve seus parâmetros estimados pelo método da máxima verossimilhança condicional por meio do pacote *stats* do software R. Na Tabela 5 encontram-se as estimativas dos parâmetros deste modelo, juntamente com a significância de cada um, baseando-se no valor p .

O modelo escolhido para ajustar os dados nem sempre são bons também para realizar previsões de dados futuros. Assim, o modelo para ajuste dos dados foi escolhido pelos

Tabela 5 Estimativa dos parâmetros do modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$

Coefficiente	Estimativa	Valor p
ϕ_1	-0,2226	0,2497
ϕ_2	0,2579	< 0,0001
θ_1	-0,5036	0,0107
θ_2	-0,4733	0,0131
Θ_1	-0,9333	< 0,0001

critérios de seleção BIC e AICc, enquanto que o modelo para realização das previsões foi escolhido por meio das medidas de erro quadrático médio (MSE) e erro percentual médio absoluto (MAPE).

Como os valores para o MSE e MAPE apresentaram resultados muito próximos para alguns modelos, com diferença apenas na quarta casa decimal, o modelo escolhido para realização das previsões será o mesmo escolhido para modelar a série, ou seja, o $SARIMA(2, 1, 2) \times (0, 1, 1)$.

Após a escolha do modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$ a análise residual acerca da adequabilidade do mesmo foi realizada e os resultados se encontram na Tabela 6, na qual é possível notar que os resíduos não seguem distribuição normal, pois a hipótese de nulidade foi rejeitada pelo teste de Shapiro-Wilk, com isso o intervalo de confiança para previsão não pode ser construído com base na distribuição normal. O intervalo de confiança foi construído por meio de reamostragem, em que utilizou-se o método de *bootstrap*.

Quanto ao teste de Ljung-Box, aplicado aos resíduos, conclui-se que os resíduos do modelo são não autocorrelacionados. Enquanto que ao aplicar o mesmo teste aos resíduos elevados ao quadrado conclui-se que o modelo apresenta variâncias homogêneas. De posse destes resultados tem-se que o modelo está bem ajustado pela metodologia de Box e Jenkins.

Tabela 6 Resultado da análise residual do modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$

Teste	Estatística	Valor p
Shapiro-Wilk	0,9919	0,0028
Ljung-Box para resíduos	124,6700	0,8962
Ljung-Box para resíduos elevados ao quadrado	20,0000	0,1652

O resultado do teste de Shapiro-Wilk está ratificado no Gráfico 6 em que é possível notar os resíduos observados desviando-se da normalidade. No Gráfico 7 tem-se a ratificação para o resultado do teste de Ljung-Box aplicado aos resíduos e aos resíduos elevados ao quadrado, em que é possível notar um número pequeno de picos de autocorrelação que se mostram significativos.

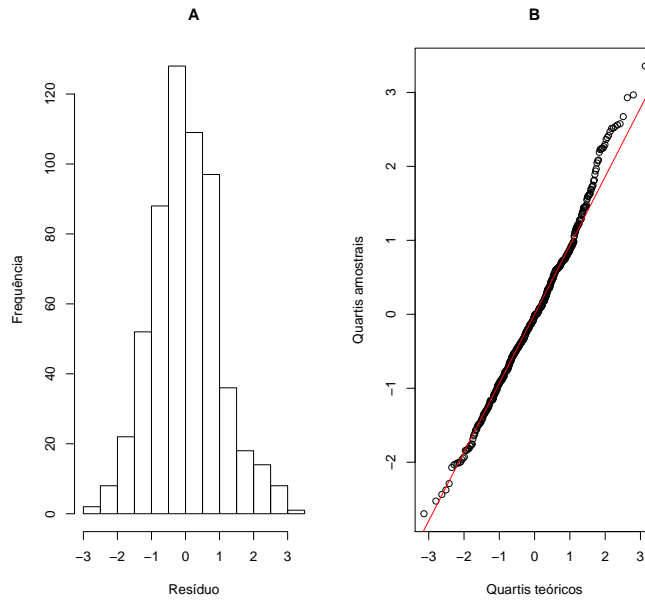


Gráfico 6 A - Histograma dos resíduos; B - QQ-plot dos resíduos

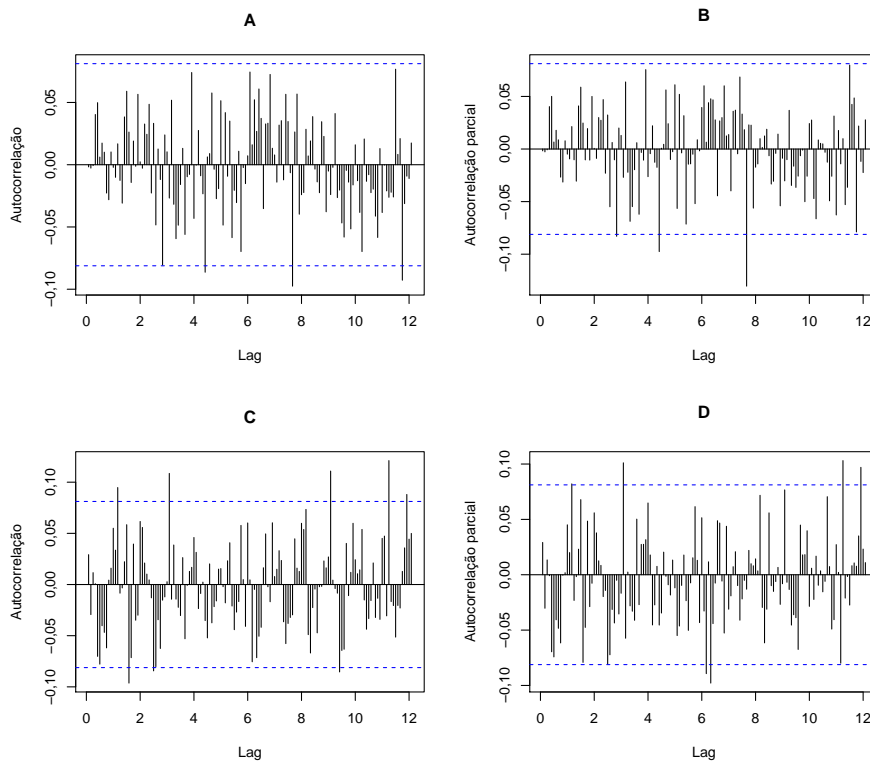


Gráfico 7 A - Fac dos resíduos; B - Facp dos resíduos; C - Fac dos resíduos elevados ao quadrado; D - Facp dos resíduos elevados ao quadrado

Na Tabela 7 encontra-se as previsões realizadas por meio do modelo, escolhido e verificado anteriormente, $SARIMA(2, 1, 2) \times (0, 1, 1)$. Como os dados em estudo não provém de uma população normal a construção do intervalo de confiança é menos precisa, uma vez que foi realizada por meio de reamostragem.

Nota-se nas previsões apresentadas na Tabela 7 que a média mensal das tempera-

Tabela 7 Previsão para os próximos 12 meses da média mensal das temperaturas máximas

Mês	Previsão	Intervalo de confiança	
		LI	LS
ago/16	26,3556	24,2603	28,9743
set/16	27,3171	25,0867	29,9964
out/16	28,0877	25,8277	30,7714
nov/16	27,8542	25,5919	30,5387
dez/16	28,7128	26,4384	31,3800
jan/17	29,4247	27,1725	32,0805
fev/17	30,2954	28,0189	32,9736
mar/17	29,1996	26,9389	31,8453
abr/17	27,9821	25,7009	30,7010
mai/17	25,6919	23,4170	28,3615
jun/17	24,7507	22,4784	27,4483
jul/17	24,8593	22,5893	27,5203

turas máximas estão bem altas, com previsão acima dos 30 °C para o mês de fevereiro, e para janeiro e março próxima dos 30 °C, destaca-se ainda que nenhuma previsão ficou abaixo dos 24 °C.

3.2 Média mensal da temperatura mínima

Nesta seção serão apresentados os resultados dos estudos acerca das características da média mensal das temperaturas mínimas da cidade de Viçosa-MG com registros desde janeiro de 1968 a julho de 2016.

No período estudado a temperatura mínima de Viçosa apresentou uma média geral de 15,43 °C com desvio padrão de 3,0285 °C. O mês que apresentou maior valor de temperatura média mensal foi janeiro de 2003, atingindo 20,15 °C, já o mês em que a média mensal obteve o menor valor foi em junho de 1968 em que registrou-se uma média mensal de 7,283 °C.

No Gráfico 8 tem-se a representação da série temporal das médias mensais das temperaturas mínimas da cidade de Viçosa-MG juntamente com a reta de tendência, indicando uma elevação da média mensal das temperaturas ao longo do tempo. Além disso nota-se no gráfico os picos frequentes da série, que são indícios de sazonalidade, fator comum em dados climatológicos.

No Gráfico 9 pode-se observar a série decomposta em suas três componentes não observáveis: tendência, sazonalidade e fator aleatório, sendo possível notar, por meio de uma análise visual, a presença de tendência crescente e um fator sazonal bem definido.

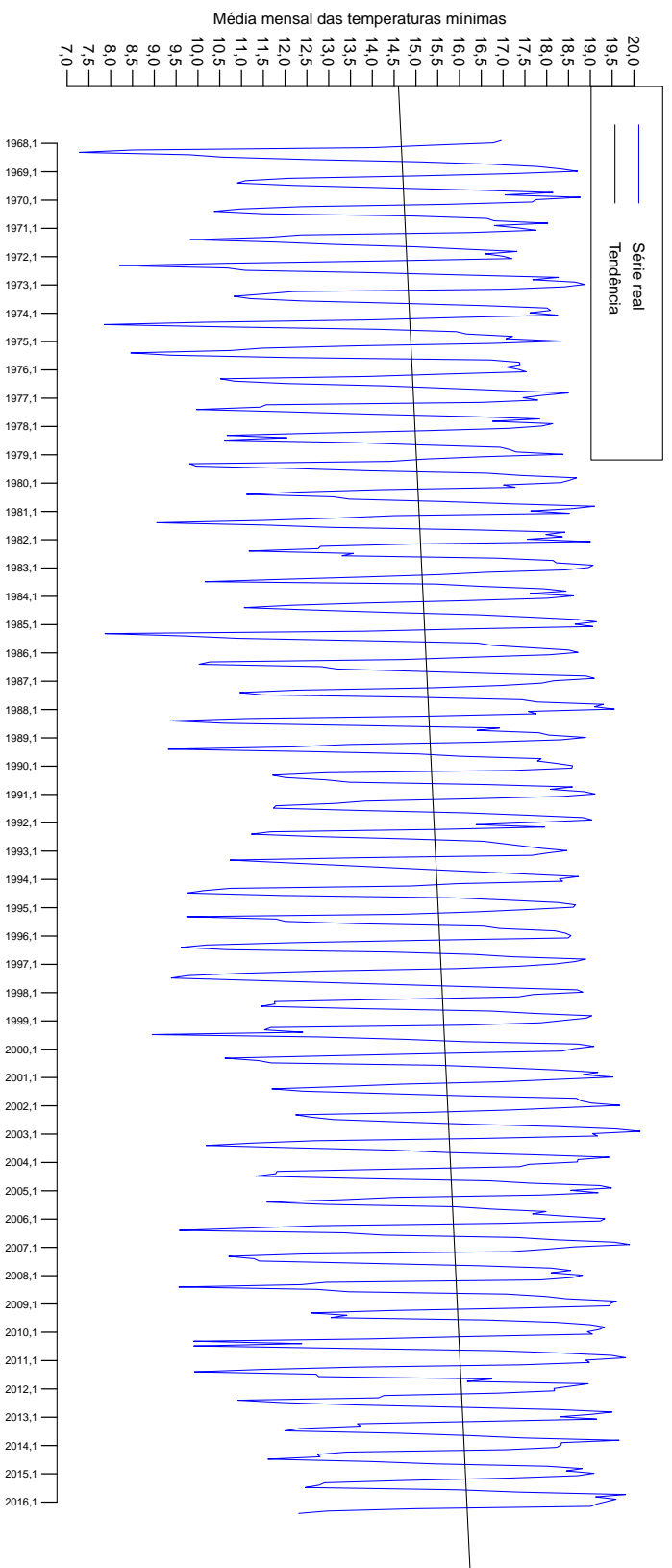


Gráfico 8 Série histórica da média mensal das temperaturas mínimas da cidade de Viçosa-MG com a reta de tendência

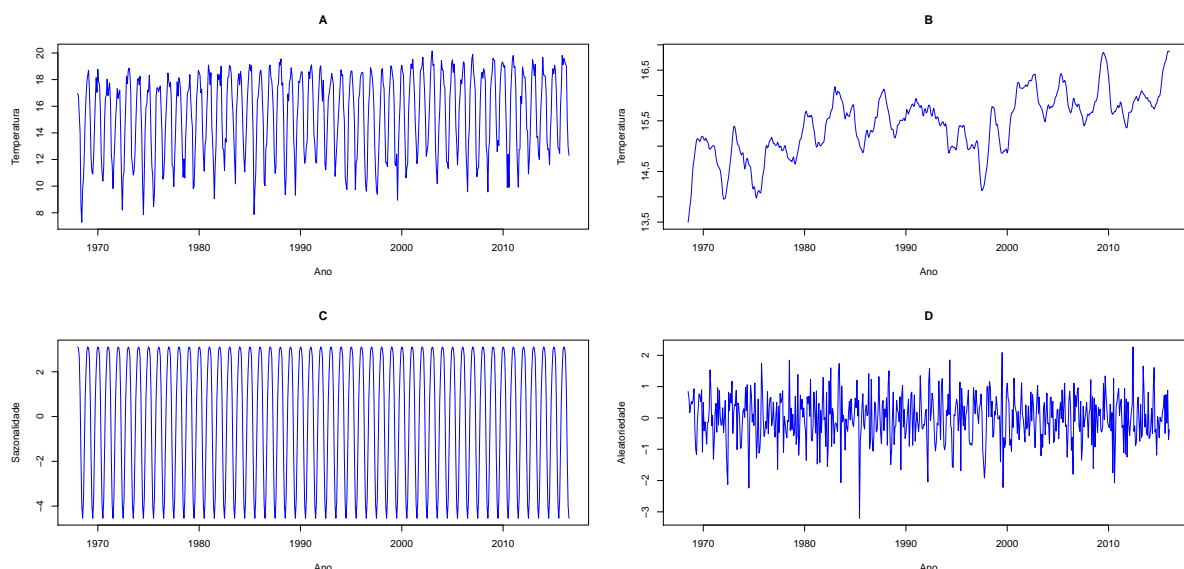


Gráfico 9 A - Série da média mensal da temperatura mínima de Viçosa-MG; B - Componente de tendência da série; C - Componente sazonal da série; D - Fator aleatório da série

O resultado para a análise da tendência de acréscimo da temperatura está apresentado na Tabela 8, em que é possível notar uma elevação significativa de 0,0313 °C, por ano, na média mensal das temperaturas mínimas da cidade de Viçosa, resultando em uma elevação um pouco acima de 1,5 °C, na média, ao longo dos 48 anos e 7 meses estudados.

Tabela 8 Resultado da análise para tendência de elevação da temperatura

	Estimativa	Desvio padrão	Estatística t	Valor p
Intercepto	-47,0114	17,6438	-2,6640	0,0079
Tempo (x)	0,0313	0,0089	3,5390	0,0004

No Gráfico 10 encontra-se o boxplot de cada mês ao longo dos anos da série em estudo, indicando que as temperaturas mais elevadas ocorrem nos meses de dezembro, janeiro e fevereiro enquanto que as menores ocorrem em junho, julho e agosto. As maiores amplitudes térmicas podem ser vistas em abril, maio, julho e novembro. Neste mesmo gráfico identifica-se alguns meses em certos anos que são possíveis *outliers*, entretanto não foi identificada observação com valor extremo de tal modo que fosse necessário intervir.

Encontra-se, na Tabela 9, os resultados do teste de Dickey-Fuller aumentado (ADF), no qual verificou-se que a série em estudo não apresenta raiz unitária, o que equivale a afirmar que a série não possui tendência estocástica, pois a regressão é estimada com uma diferença já aplicada aos dados. A hipótese de raiz unitária foi rejeitada a 5% de probabilidade (valor $p = 0,0004$). Porém o coeficiente para tendência determinística possui valor

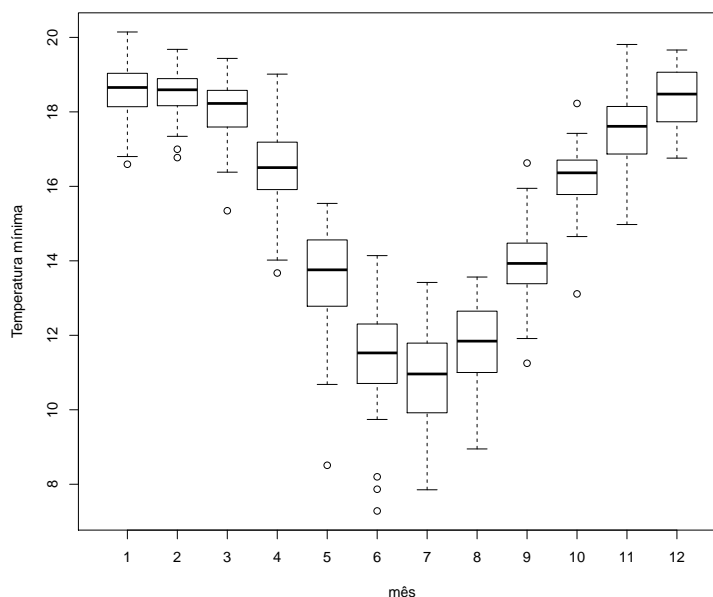


Gráfico 10 Boxplot de cada mês da série em anos diferentes

significativo com valor $p < 0,0001$, ou seja, rejeitou-se a hipótese de nulidade que afirma não possuir tendência determinística na série. Desta forma eliminou-se a tendência determinística após a aplicação de uma diferença nos dados. Realizou-se o teste ADF com 21 defasagens da variável em primeira diferença, tais defasagens foram necessárias para eliminar a autocorrelação residual, que pode influenciar no teste. O teste de Ljung-Box apresentou valor $p = 0,1587$ confirmando a 5% de probabilidade que a autocorrelação residual foi eliminada, não rejeitando a hipótese de nulidade (H_0 : Os resíduos são não autocorrelacionados).

Tabela 9 Resultado do teste de Dickey-Fuller aumentado

Parâmetros	β_1	β_2	π
Estimativa	7,5911	0,0013	-0,5160
Erro-padrão	1,6995	0,0004	0,1153
Estatística do teste	4,4670	3,5330	-4,4740
Valor p	< 0,0001	0,0004	< 0,0001
Teste Ljung-Box	$Q = 162,78$ valor $p = 0,1587$		

Após aplicação de uma diferença simples nos dados da série, aplicou-se novamente o teste ADF e os resultados estão apresentados na Tabela 10, em que verificou-se a série livre tendência. Na série transformada, a aplicação do teste ADF necessitou de 26 defasagens para eliminar a autocorrelação residual.

É provável que a série em estudo apresente sazonalidade, pois se trata de uma série de dados climatológicos. Para auxiliar na verificação da periodicidade desta componente,

Tabela 10 Resultado do teste de Dickey-Fuller aumentado para a série após aplicação de uma diferença

Parâmetros	β_1	β_2	π
Estimativa	-0,0201	0,0001	-7,0703
Erro-padrão	0,0854	0,0002	1,1758
Estatística do teste	-0,2360	0,5290	-6,0130
Valor p	0,8137	0,5971	< 0,0001
Teste Ljung-Box	$Q = 171,7500$ valor $p = 0,0695$		

utilizou-se o Gráfico 11, que representa o periodograma da análise espectral da série livre de tendência.

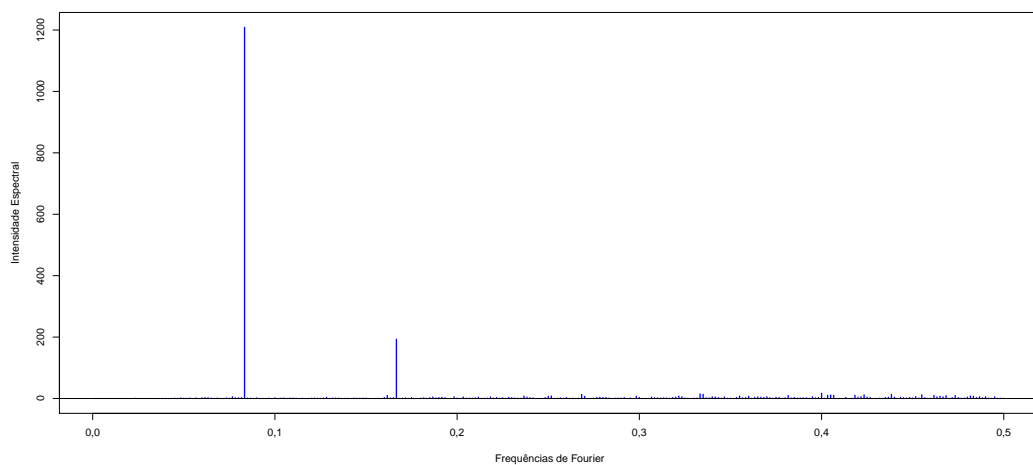


Gráfico 11 Periodograma da série da média mensal das temperaturas mínimas de Viçosa-MG

Ao analisar o gráfico, nota-se, que a maior intensidade espectral está associada a frequência de Fourier $f = 0,0833$ que oferece uma periodicidade $s = 12$. Este valor encontrado para a periodicidade foi testado pelo teste G de Fisher que apresentou estatística $G = 0,5667$, sendo superior estatisticamente ao valor de comparação $Z = 0,02941$, com valor $p < 0,0001$, o que confirma que G é estatisticamente maior que Z . Portanto, a série histórica da média mensal das temperaturas mínimas de Viçosa-MG possui periodicidade de 12 meses, ou seja, sazonalidade anual. Para eliminar tal fator foi aplicada uma diferença sazonal de tamanho 12, obtendo uma série histórica transformada e livre de sazonalidade.

Após o tratamento dos dados foi possível aplicar a metodologia de Box e Jenkins na modelagem. O processo de modelagem iniciou-se com base no Gráfico 12, no qual encontra-se a função de autocorrelação (fac) e autocorrelação parcial (facp) da série original e da série após aplicação das diferenças (com comportamento estacionário).

Porém, é difícil ajustar completamente um modelo de Box e Jenkins baseando-se apenas nos gráficos da *fac* e *facp* da série original e da série diferenciada, mas pode-se notar que após aplicação das diferenças, a estrutura de autocorrelação dos dados foi relativamente minimizada, entretanto é visível que nas defasagens múltiplas e próximas de $s = 12$ alguns coeficientes de autocorrelação e autocorrelação parcial se apresentam significativos a 5% de probabilidade, indicando que o modelo mais indicado para tal é da classe $SARIMA(p, d, q) \times (P, D, Q)$.

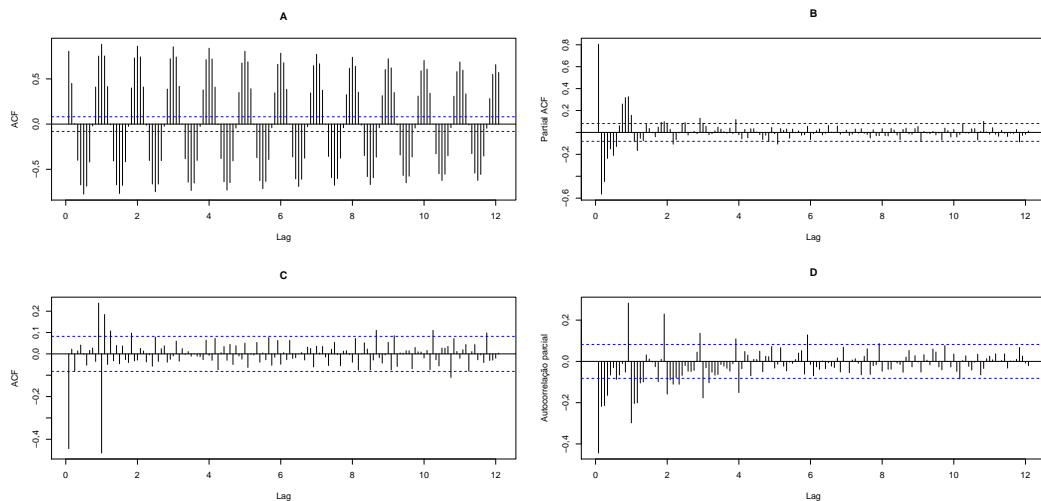


Gráfico 12 A - *Fac* da série original; B - *Facp* da série original; C - *Fac* da série diferenciada; D - *Facp* da série diferenciada

Como foi aplicada uma diferença simples e uma sazonal para tratamento prévio dos dados, tem-se que um bom modelo para representar os dados é da forma $SARIMA(p, 1, q) \times (P, 1, Q)$. Na Tabela 11 estão apresentados diversos modelos concorrentes que foram ajustados e selecionados por meio dos critérios de informação bayesiano (BIC) e de Akaike corrigido (AICc). Enquanto que a medida do erro quadrático médio (MSE) e erro percentual médio absoluto (MAPE) foram utilizadas para escolher um bom modelo para realizar previsões, pois um modelo que se mostra bom para ajuste dos dados pode não ser bom para realizar previsões.

O modelo escolhido para o ajuste dos dados foi o $SARIMA(1, 1, 2) \times (0, 1, 1)$, enquanto que o $SARIMA(2, 1, 2) \times (0, 1, 1)$ se mostrou melhor para realização de previsões, pois apresentou menores valores para as medidas de erro MSE e MAPE.

O modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$ escolhido por meio do BIC e AICc teve seus parâmetros estimados pelo método da máxima verossimilhanças condicional por meio pacote *stats* do software R.

Encontra-se na Tabela 12 as estimativas dos parâmetros do modelo juntamente com a significância de cada valor baseando-se no valor p . O mesmo foi feito para o

Tabela 11 Comparação de diversos modelos concorrentes para a série da média mensal das temperaturas mínimas de Viçosa-MG

Modelo	BIC	AICc	MSE	MAPE
$SARIMA(0, 1, 0) \times (0, 1, 0)$	2069,7020	2065,3630	0,7822	3,1423
$SARIMA(1, 1, 0) \times (0, 1, 0)$	1951,1040	1942,4340	0,7607	3,5538
$SARIMA(2, 1, 0) \times (0, 1, 0)$	1929,8880	1916,8930	0,6158	3,3061
$SARIMA(2, 1, 1) \times (0, 1, 0)$	1846,0410	1828,7290	0,5085	2,9090
$SARIMA(2, 1, 2) \times (0, 1, 0)$	1850,7770	1829,1560	0,489	2,8641
$SARIMA(2, 1, 2) \times (0, 1, 1)$	1540,2790	1514,3540	0,2594	2,0877
$SARIMA(2, 1, 2) \times (0, 1, 2)$	1546,5420	1516,3210	0,2622	2,0988
$SARIMA(2, 1, 2) \times (1, 1, 1)$	1546,5480	1516,3280	0,2620	2,0979
$SARIMA(1, 1, 2) \times (0, 1, 1)$	1534,0370	1512,4160	0,2617	2,1046
$SARIMA(1, 1, 2) \times (1, 1, 1)$	1540,2980	1514,3730	0,2644	2,1147
$SARIMA(1, 1, 2) \times (0, 1, 2)$	1540,2900	1514,3660	0,2646	2,1156

modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$ selecionado pelas medidas de erro MSE e MAPE e os resultados se encontram na Tabela 13, em que é possível notar que o valor estimado para ϕ_2 se apresentou como não significativo.

O modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$ com o termo autorregressivo de segunda ordem não significativo resulta no modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$. Portanto o modelo escolhido tanto para ajustar os dados históricos quanto para realizar previsões será o $SARIMA(1, 1, 2) \times (0, 1, 1)$.

Tabela 12 Estimativa dos parâmetros do modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$

Coefficiente	Estimativa	Valor p
ϕ_1	0,7690	< 0,0001
θ_1	-1,4875	< 0,0001
θ_2	0,4943	< 0,0001
Θ_1	-0,9412	< 0,0001

Tabela 13 Estimativa dos parâmetros do modelo $SARIMA(2, 1, 2) \times (0, 1, 1)$

Coefficiente	Estimativa	Valor p
ϕ_1	0,8222	< 0,0001
ϕ_2	-0,0262	0,7511
θ_1	-1,5357	< 0,0001
θ_2	0,5414	0,0014
Θ_1	-0,9412	< 0,0001

Caso os modelos apresentassem todos os parâmetros significativos seriam escolhidos os dois, utilizando um para o ajuste dos dados e outro para a previsão dos valores futuros.

Após a escolha do modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$ a análise residual acerca

da adequabilidade do mesmo foi realizada e os resultados se encontram na Tabela 14, na qual nota-se que os resíduos não seguem distribuição normal, pois a hipótese de nulidade foi rejeitada pelo teste de Shapiro-Wilk, com isso o intervalo de confiança para a previsão não pode ser construído com base na distribuição normal, sendo contruído por meio de reamostragem, pelo método *bootstrap*.

Quanto ao teste de Ljung-Box aplicado aos resíduos do modelo, conclui-se que, os resíduos não são autocorrelacionados, indicando bom ajuste. Enquanto que, para o mesmo teste, aplicado aos resíduos elevados ao quadrado conclui-se que o modelo não apresenta variâncias homogêneas. Desta forma o modelo em questão não atende os pressupostos para o ajuste dos dados por meio da metodologia de Box e Jenkins.

Tabela 14 Resultado da análise residual do modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$

Teste	Estatística	Valor p
Shapiro-Wilk	9884	0,0001
Ljung-Box para resíduos	7,7604	0,9330
Ljung-Box para resíduos elevados ao quadrado	30,3730	0,0107

O resultado do teste de Shapiro-Wilk está ratificado no Gráfico 13 em que é possível ver os resíduos observados desviando-se da reta de normalidade. No Gráfico 14 tem-se a ratificação para o resultado do teste de Ljung-Box aplicado aos resíduos, em que é possível notar um número baixo de picos significativos de autocorrelação e com ocorrência aleatória dos mesmos, indicando que os resíduos não são autocorrelacionados. Enquanto que pelo mesmo teste, aplicado aos resíduos elevados ao quadrado, é possível notar um número relativamente alto de picos significativos de autocorrelação e a ocorrência deles de forma periódica, indicando que as variâncias não são homogêneas.

O modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$ escolhido pelos critérios de seleção BIC e AICc se mostrou inapropriado após a realização da análise residual, pois foi detectado que o modelo não possui variâncias homogêneas, violando uma pressuposição do ruído branco. Sendo assim, utilizou-se um modelo não-linear da classe ARCH/GARCH para modelar a variância.

O ajuste do modelo da classe ARCH/GARCH foi escolhido com base no critério de informação de Akaike (AIC), pois os critérios de seleção BIC e AICc apresentaram resultados semelhantes. O resultado encontra-se na Tabela 15, em que diversos modelos concorrentes foram criados e, o escolhido, foi aquele que apresentou menor valor de AIC. Na Tabela 16 tem-se as estimativas dos parâmetros do modelo escolhido bem como a significância deles baseada no valor p . Assim, o modelo escolhido para mo-

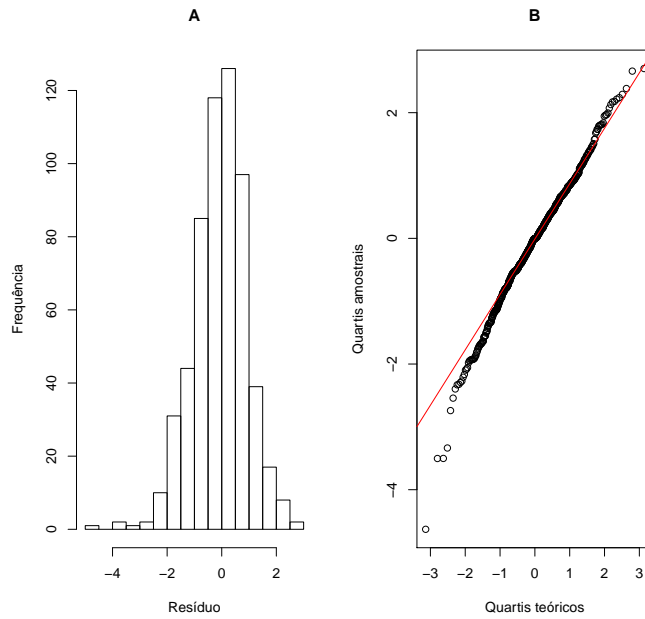


Gráfico 13 A - Histograma dos resíduos; B - QQ-plot dos resíduos

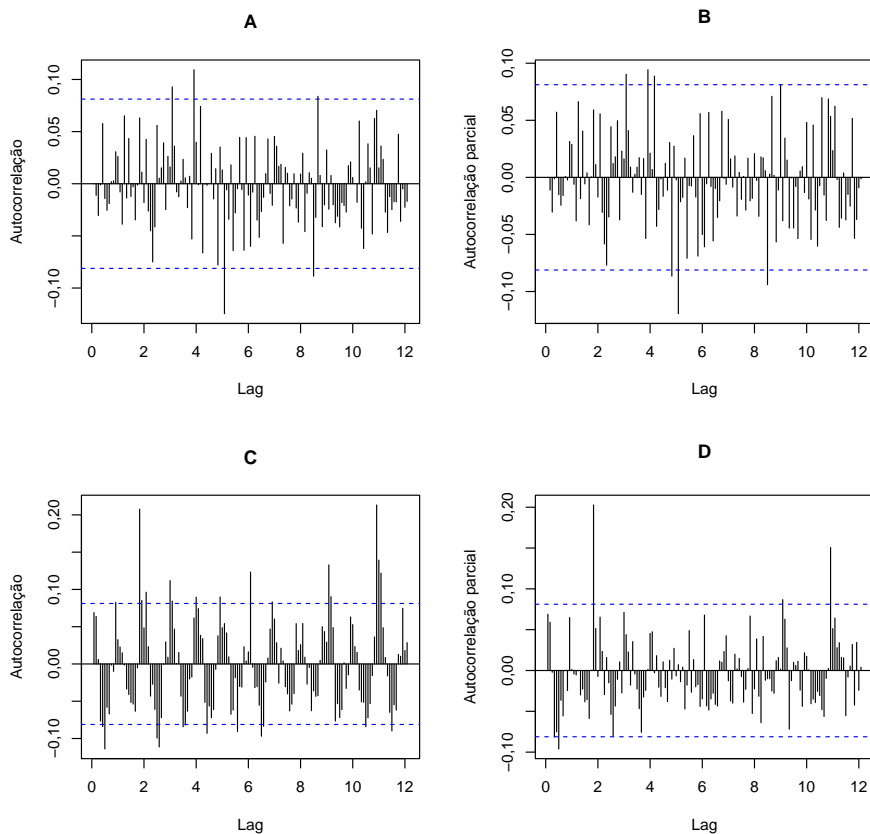


Gráfico 14 A - Fac dos resíduos; B - Fac dos resíduos; C - Fac dos resíduos elevados ao quadrado; D - Fac dos resíduos elevados ao quadrado

delar os resíduos é o $GARCH(2, 0)$, ou seja, $ARCH(2)$. Portanto tem-se o modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$ para ajustar a correlação serial e o modelo $ARCH(2)$ para ajustar os resíduos.

Modelando a série desta maneira os resíduos do modelo passaram a se comportar

Tabela 15 Comparação de modelos concorrentes

Modelo	AIC
GARCH(1,0)	1488,647
GARCH(2,0)	1483,789
GARCH(3,0)	1484,908
GARCH(2,1)	1485,792
GARCH(2,2)	1488,099
GARCH(0,1)	1493,678
GARCH(0,2)	1494,117

Tabela 16 Estimativa dos parâmetros do modelo $ARCH(2)$

Coefficiente	Estimativa	Valor
a_0	0,6072	$< 0,0001$
a_1	0,0839	0,0757
a_2	0,1223	0,0104

como ruído branco. O teste de Ljung Box aplicado aos resíduos não rejeitou a hipótese de nulidade (valor $p = 0,9232$) confirmando que os resíduos do modelo não são auto-correlacionados. O mesmo teste aplicado aos resíduos elevados ao quadrado também não rejeitou a hipótese de nulidade (valor $p = 0,1609$) indicando que as variâncias do modelo são homogêneas.

Após obter a modelagem adequada para série temporal da média mensal das temperaturas mínimas da cidade de Viçosa-MG foram realizadas as previsões para os meses de agosto de 2016 até julho de 2017. Na Tabela 17 encontra-se a previsão realizada por meio do modelo $SARIMA(1, 1, 2) \times (0, 1, 1)$ somada com a previsão para os resíduos realizadas com o modelo $ARCH(2)$. Os dois modelos juntos apresentaram melhor ajuste, além de atender aos pressupostos de uma modelagem de séries temporais por meio da classe de modelos de Box e Jenkins.

Nota-se nas previsões apresentadas na Tabela 17 que a média mensal das temperaturas mínimas estão bem altas, se comparadas com as observações de 40 anos atrás, com previsão acima dos 11°C mesmo nos meses mais frios da cidade, que já apresentou médias abaixo de 8°C , e para os meses mais quentes a previsão apresentou a média da temperatura mínima acima dos 19°C .

Tabela 17 Previsão para os próximos 12 meses da média mensal da temperatura mínima

Mês	Previsão	Intervalo de confiança	
		LI	LS
ago/16	11,4097	11,0139	14,4611
set/16	14,0669	13,1357	16,7170
out/16	16,8001	15,2965	18,9596
nov/16	18,0070	16,6045	20,3191
dez/16	19,0737	17,5346	21,2823
jan/17	18,6623	17,5909	21,3605
fev/17	18,7801	17,4334	21,2179
mar/17	19,1671	17,1262	20,9211
abr/17	18,6035	15,5680	19,3705
mai/17	14,9948	12,4675	16,2757
jun/17	12,0780	10,5780	14,3916
jul/17	11,1088	10,0090	13,8261

4 CONCLUSÃO

Ao estudar a série temporal da média mensal das temperaturas mínimas e máximas da cidade de Viçosa-MG, conclui-se que, estas características do clima desta cidade caminha na mesma direção de outras regiões que também apresentaram uma elevação significativa na temperatura, entretanto, não é possível atribuir que as elevações detectadas neste estudo estão relacionadas com efeito global, visto que o estudo se valeu de dados climáticos apenas da cidade em estudo. Para Viçosa-MG, pode-se afirmar que as temperaturas máximas e mínimas vem sofrendo elevação constante ao longo do tempo, o que resultou em um aumento de mais de 1 °C na média das temperaturas nos últimos anos.

O aquecimento que vem atingindo a cidade de Viçosa-MG pode ser atribuído a fenômenos locais, tais como obras de urbanização, aumento no fluxo de veículos ou ainda uma redução nos dias chuvosos e ondas de calor sobre a região. Entretanto os fenômenos globais não podem ser totalmente excluídos, podendo também ser responsáveis por parte do aquecimento da cidade.

Neste contexto os meses que apresentaram maiores altas nas médias mensais das temperaturas mínimas foram: abril, junho e julho, enquanto que, para as máximas os meses de abril, setembro e outubro apresentaram maiores elevações.

Estes resultados são de grande valia para a população de Viçosa-MG, em especial para os setores que dependem diretamente das temperaturas, tais como, o setor da agricultura, da produção animal e do fornecimento de água e energia.

As análises e as previsões realizadas por meio de modelos bem ajustados e veri-

ficados devem ser vistas por toda a população viçosense e região, para que possam obter uma informação específica acerca do clima da cidade e, que decisões possam ser tomadas agora para conter o aquecimento e, contribuir para o cuidado da região em que se encontra a cidade de Viçosa.

REFERÊNCIAS

- BLAIN, G. C. Séries anuais de temperatura máxima média do ar no estado de São Paulo: Variações e tendências climáticas. **Revista Brasileira de Meteorologia**, São José dos Campos, v. 25, n. 1, p. 114-124, 2010.
- BLAIN, G. C.; PICOLI, M. C. A.; LULU, J. Análise estatística das tendências de elevação nas séries anuais de temperatura mínima do ar no estado de São Paulo. **Bragantia**, Campinas, v. 68, n. 3, p. 807-815, 2009.
- GANDRA, C. F. A. T.; DAMÉ, R. C. F.; SIMONETE, M. A.; BACELAR, L. C. S.; DISCONZI, P. B.; SANTOS, J. P. Modelagem estocástica: previsão das temperaturas para a localidade de Pelotas/RS/Brasil. **Revista agro@mbiente online**, Boa Vista, v. 8, n. 1, p. 18-28, 2014.
- MARENCO, J. A. Mudanças climáticas globais e regionais: avaliação do clima atual do Brasil e projeções de cenários climáticos do futuro. **Revista brasileira de meteorologia**, São José dos Campos, v. 16, n. 1, p. 1-18, 2001.
- _____. Impactos de extremos relacionados com o tempo e o clima - Impactos sociais e econômicos. **Boletim do Grupo de Pesquisas em Mudanças Climáticas - GPMC**, Edição Especial, n. 8, Mai. 2009.
- MINUZZI, R. B.; VIANELLO, R. L.; SEDIYAMA, G. C. Oscilações climáticas em Minas Gerais. **Revista Brasileira de Meteorologia**, São José dos Campos, v. 25, n. 2, p. 227-236, 2010.
- PRIORI, A.; POMARI, L. R.; AMÂNCIO, S. M.; IPÓLITO, V. K. **História do Paraná (Séculos XIX e XX)**. Maringá: Eduem, 2012. 234 p.
- R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016. Disponível em: <http://www.r-project.org>. Acesso em: 13 julho 2016.
- SALGADO, D. D.; NÄÄS, I. A. Avaliação de risco à produção de frngo de corte do estado de São Paulo em função da temperatura ambiente. **Revista Engenharia Agrícola**, Jaboticabal, v. 30, n. 3, p. 367-376, 2010.
- SILVA, M. I. S.; GUIMARÃES, E. C.; TAVARES, M. Previsão da temperatura média mensal de Uberlândia, MG, com modelos de séries temporais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 12, n. 5, p. 480-485, 2008.
- STEINKE, E. T.; SOUZA, G. A.; SAITO, C. H. Análise da variabilidade da temperatura do ar e da precipitação no Distrito Federal no período de 1965/2003 e sua relação com uma possível alteração climática. **Revista Brasileira de Climatologia**, v. 1, n. 1, p. 131-145, 2005.

UNIVERSIDADE FEDERAL DE VIÇOSA - UFV. Departamento de Engenharia Agrícola. Estação Climatológica Principal de Viçosa-MG. **Boletim meteorológico 2016**. Viçosa, 2016. Disponível em http://www.posmet.ufv.br/?page_id=1253. Acesso em 7/6/2016.

CAPÍTULO 3

Estudo da precipitação pluviométrica total mensal de Viçosa-MG por meio de séries temporais

RESUMO

Com este trabalho objetivou-se principalmente realizar uma análise da série histórica da precipitação pluviométrica total mensal da cidade de Viçosa-MG, se valendo da análise de séries temporais e verificando o efeito da tendência e sazonalidade contidas nos dados históricos que são registrados desde janeiro de 1968, além de identificar possíveis alterações no período e na quantidade das chuvas ao longo do tempo, e realizar previsões para os próximos 12 meses. Para análise e modelagem da série foram escolhidos diversos modelos da classe SARIMA e o que melhor se ajustou dentre os modelos concorrentes foi selecionado por meio do critério de informação bayesiano (BIC) e confirmado pelo critério de informação de Akaike corrigido (AICc). O modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$ apresentou melhor ajuste para modelar a correlação serial, e o modelo $ARCH(2)$ apresentou melhor ajuste para os resíduos do modelo, tal ajuste residual foi necessário pois os resíduos da série da precipitação total mensal não se comportam como ruído branco. Com este modelo foi possível realizar previsão para os próximos 12 meses para a precipitação total mensal, além de verificar que as precipitações não sofreram alteração significativa ao longo do tempo, sendo detectado apenas anos mais e outros menos chuvosos, porém, de forma aleatória.

Palavras-chave: Alterações das chuvas. Previsão. SARIMA. ARCH.

1 INTRODUÇÃO

A água é um bem essencial para a existência da vida, desta forma, estudos acerca da hidrologia se fazem cada vez mais presentes no meio acadêmico e entre os analistas do clima.

Merten e Minella (2002) afirmam que a água doce é um recurso natural finito, cuja qualidade vem piorando devido ao aumento da população e a falta de políticas públicas para a sua preservação. Estima-se que a cada ano 12 milhões de mortes que ocorrem no mundo estão relacionadas com problemas na qualidade da água. Neste sentido não se deve ater apenas a quantidade de água doce que o Planeta Terra contém, mas também para a qualidade e o seu estado de conservação.

Para que a população mundial tenha acesso a água de qualidade será necessário maior cuidado com os recursos hídricos do planeta Terra, visto que a população está em constante crescimento.

Para analisar o comportamento hidrológico de uma região por meio de séries temporais é preciso utilizar algum conjunto de dados históricos, para isso é preciso levar em conta uma característica mensurável. É comum estudos por meio de séries temporais em que utiliza-se dados obtidos via mensuração da vazão de um rio ou de uma comporta de alguma barragem da região estudada, sendo visto também muitos estudos que utilizam dados da precipitação pluviométrica registrada diariamente por meio de estações climatológicas. Estas formas de captação de dados são correlacionadas, visto que a vazão de um rio ou barragem sofre alterações quando o índice de chuvas nas regiões próximas dele também sofre alterações em dias anteriores.

Segundo Silva et al. (2002) a variável precipitação possui grande responsabilidade no desenvolvimento e no comportamento do clima de uma região. Sendo assim, estudos acerca da precipitação pluviométrica se tornam de grande valia no planejamento econômico de uma localidade. A decisão de iniciar uma atividade agrícola, ou qualquer outro empreendimento que depende do clima, está diretamente ligada ao índice de precipitação pluviométrica no lugar em que será desenvolvida tal atividade.

Conforme Ayoade (2006) os parâmetros do clima exercem influências sobre todos os estágios da cadeia de produção, começando pela preparação da terra, semeadura, crescimento dos cultivos, colheita, armazenamento, transporte e comercialização.

A cidade de Viçosa - MG tem na agricultura uma grande atividade local, sendo caracterizada por pequenos produtores e agricultura familiar, que dependem ainda mais

de chuvas regulares, pois pequenos produtores geralmente não possuem recursos econômicos para investir em sistemas de irrigação avançados, limitando sua produção apenas no período em que as chuvas são mais frequentes. Este é um dos fatores que motivaram o presente estudo, pois com a análise do comportamento das precipitações da cidade de Viçosa será possível perceber de que forma as chuvas vem se comportando, tornando possível destacar se as chuvas estão ocorrendo em quantidades diferentes ao longo do tempo ou se elas estão ocorrendo em espaços de tempos mais longos ou mais curtos.

Ayoade (2006) afirma que o clima é o fator da produção agrícola mais difícil de ser controlado e de maior impacto quando se deseja a produção máxima, pois é o fator que apresenta maior imprevisibilidade. Desta forma, um estudo detalhado acerca do clima se valendo de uma série de dados históricos com quase 50 anos ajuda na compreensão desta variável, sendo possível a detecção de algumas alterações no seu padrão. Dentre os fatores do clima que apresentam grande imprevisibilidade destaca-se a precipitação, pois é um elemento influenciado por diversos outros.

Do ponto de vista do setor de abastecimento de água tratada para a população da cidade de Viçosa - MG, o índice de precipitação pluviométrica total e a regularidade das chuvas causam grande preocupação. Visto que, todos os anos a cidade passa por uma escassez muito grande de água.

A cidade em questão contém reservatórios de água para ser tratada que são abastecidos por rio de pequeno porte (São Bartolomeu), que se torna muitas vezes incapaz de fornecer água para reabastecer o reservatório na mesma proporção que é retirada nos períodos de estiagens prolongadas, acarretando em forte economia ou até mesmo na falta de água tratada para toda a população em algumas épocas do ano, geralmente no final do inverno nos meses de agosto e setembro. Este fato também influenciou o presente estudo, que busca verificar se a falta de água está relacionada com crescimento populacional apenas ou com a redução de chuvas.

Não é raro que a falta de chuvas regulares cause prejuízos em todas as partes. Além de causarem escassez ou falta de água para consumo da população de uma região, grandes períodos de secas ainda prejudicam muito na produção agrícola. Se as chuvas não ocorrerem de forma regular em todas as etapas de uma atividade agrícola dificilmente a produção será máxima, o que pode gerar baixa oferta de alimentos no mercado e alta nos preços, visto que a demanda por alimentos vem aumentando com o crescimento da população.

Silva et al. (2011) afirmam que variações extremas na precipitação pluviométrica

influenciam na vida humana e nas atividades agrícolas em diversas regiões do planeta. Os autores afirmam ainda que o conhecimento do regime de precipitação pluviométrica com relação à duração e ao final da estação chuvosa, e a possibilidade de conhecer períodos mais suscetíveis a estiagens (veranico) no decorrer desta estação são fundamentais para a elaboração de um calendário agrícola e a implementação de projetos agrícolas.

Os procedimentos e características de um estudo de séries temporais se apresentam como boas ferramentas para analisar séries históricas de precipitações pluviométricas, pois os dados desta natureza geralmente apresentam dependência serial entre as observações e sazonalidade.

Silva et al. (2002) utilizaram modelos de séries temporais para modelar e fazer previsões da precipitação pluviométrica total mensal do município de Uberaba - MG, verificando um bom ajuste e confirmando que os modelos da classe de Box e Jenkins podem ser utilizados para este fim.

Sáfadi (2004) em um estudo com dados acerca da vazão da Barragem de Furnas se valeu de técnicas de séries temporais para ajustar aos dados um modelo de Box e Jenkins com e sem intervenção, confirmando que a análise por meio de séries temporais aplicadas em dados desta natureza é uma boa ferramenta.

Bayer e Castro (2012) realizaram um estudo acerca da vazão do rio Potiribu na região noroeste do estado do Rio Grande do Sul, para este estudo ajustaram modelos de Box e Jenkins aos dados da série, concluindo que o modelo SARIMA da classe ARIMA que incorpora a sazonalidade proporcionou um bom ajuste, tornando possível a previsão para a vazão do rio Potiribu para os próximos meses.

Pereira et al. (2015) utilizaram modelos de Box e Jenkins para modelar, analisar o comportamento e realizar previsões dos dados mensais da série temporal de temperatura e precipitação pluviométrica da cidade de Areia - PB. Os autores afirmam que tais modelos se mostraram satisfatórios para o procedimento.

Batista (2009) em seus estudos acerca da vazão do rio Grande, afluente a um reservatório da Usina Hidrelétrica de Camargos/CEMIG, utilizou modelos de Box e Jenkins e redes neurais artificiais para analisar os dados e realizar previsões. As duas técnicas utilizadas se mostraram eficientes na modelagem e previsão dos dados, porém com algumas vantagens e desvantagens para cada, concluindo que um modelo da classe SARIMA apresentou um bom ajuste e uma boa previsão, apresentando um erro elevado apenas em um dos 4 meses que foram feitas as previsões.

De modo análogo aos estudos citados, a análise dos dados da precipitação pluvio-

métrica total mensal da cidade de Viçosa-MG serão avaliados, objetivando verificar possíveis alterações no padrão das precipitações mensais que ocorreram ao longo do tempo.

Com o presente estudo objetiva-se confirmar se a população, que depende diretamente de chuvas regulares para realizar seus trabalhos, estão conseguindo desempenhar suas atividades de forma normal e se há tendência de redução no volume das chuvas em Viçosa.

Objetiva-se ainda verificar se o índice de precipitação pluviométrica total mensal da cidade de Viçosa-MG vem sofrendo alterações ao longo do tempo estudado, para verificar se a falta de água que a população da cidade tem passado nos últimos anos se deve a um problema relacionado a redução da precipitação pluviométrica mensal, ou se o problema está no planejamento e em políticas públicas sobre a preservação e uso consciente da água, visto que a população da cidade vem crescendo de tal forma que a estrutura atual de fornecimento não se mostra suficiente.

Outro objetivo deste mesmo trabalho será realizar previsão da precipitação pluviométrica mensal da cidade de Viçosa se valendo de modelos de séries temporais ajustados aos dados. A previsão da precipitação pode não ser tão próxima da realidade, pois tal fator possui uma variância bastante alta.

2 MATERIAL E MÉTODOS

A série temporal sob estudo refere-se aos dados diários acerca da precipitação pluviométrica da cidade de Viçosa-MG. Esta característica foi extraída do banco de dados do clima da cidade que é registrado diariamente na estação meteorológica da Universidade Federal de Viçosa no departamento de Engenharia Agrícola (UFV, 2016).

Para os índices de precipitação com registros diários foi calculado uma soma mensal, para trabalhar com os dados sem perder informações dos anos anteriores, pois uma série histórica contendo 365 observações por ano se torna mais difícil de ser tratada, acarretando maior erro que uma série com 12 observações por ano, dado que cada observação apresenta um erro ao ser modelada. Além disso foi utilizado interpolação para alguns dados não registrados, que geralmente ocorrem por defeito ou manutenção do equipamento de medida. Uma vez que dados faltantes poderiam acarretar problemas no tratamento dos dados.

Calculando a precipitação total mensal, a série passa a ter 12 observações por ano, o que torna mais fácil o tratamento estatístico dos dados com menor erro de modelagem e previsão. A esperança de chuvas para um certo dia do ano não pode ser detectada com boa confiabilidade baseando-se apenas em informações anteriores, pois a chuva em um determinado dia depende de diversos fatores, enquanto que a precipitação pluviométrica para um mês pode ser obtida com melhor confiabilidade baseando-se nos registros dos meses que o antecedem, e também, nas observações deste mesmo mês nos anos anteriores, visto que observações de meses anteriores são bons indicativos da precipitação total do mês seguinte.

Após transformar os dados em observações mensais, originando a série da precipitação total mensal da cidade de Viçosa, foi realizada a construção do gráfico desta série para analisar as suas principais características, tais como tendência, sazonalidade e efeitos aleatórios. Foi construído ainda um gráfico de cada componente da série, em que foi possível avaliar o comportamento de cada uma separadamente. Com base nestes gráficos realizou-se uma análise visual e a partir dela foi possível ter uma ideia de quais testes seriam viáveis para verificar a presença de algumas características da série e ainda ter uma visão a priori de qual classe dentre os modelos de Box e Jenkins seria a mais adequada para modelar os dados.

Para obter uma visão detalhada acerca da precipitação pluviométrica mensal de Viçosa efetuou-se uma análise de cada mês em diferentes anos, com intuito de verificar

em quais meses ou estações ocorreram maiores alterações no índice de precipitação ao longo do tempo.

O boxplot de cada mês da série para anos diferentes foi efetuado a fim de ilustrar a variação presente em cada período, em que verificou-se os meses ou estações que apresentam índices de precipitação pluviométrica concentrada ao redor da mediana dos dados com pouca variação, e os meses que apresentam maior dispersão em relação a mediana, oscilando em intervalos maiores. Com este gráfico foi possível notar diversos candidatos a *outliers*, entretanto, não foi observado dado com valor extremo que fosse necessário aplicar intervenção no processo de modelagem.

Os gráficos da função de autocorrelação (fac) e autocorrelação parcial (facp) da série foram construídos com intuito de identificar, por meio de uma análise visual, uma dentre várias classes dos modelos de Box e Jenkins que pode ser utilizada para modelar a série e quais as ordens de cada parâmetro, além de verificar se a série apresenta componente sazonal. Porém, a tarefa de identificação de modelos e suas ordens por meio destes gráficos não é fácil, tornando-se possível ter apenas uma ideia de quais modelos podem se ajustar a série.

A fim de verificar a presença de tendência determinística e/ou estocástica realizou-se o teste de Dickey-Fuller aumentado (ADF) aplicado aos dados da precipitação total mensal. Com o gráfico da série foi possível analisar a sua oscilação além de confirmar e ilustrar o resultado do teste.

Como a série em estudo não apresentou tendência, a presença de sazonalidade foi analisada por meio de uma análise espectral, decompondo a série no domínio das frequências de Fourier. Para tal, foi realizado o gráfico do periodograma contendo as frequências de Fourier dispostas no eixo das abscissas e suas respectivas intensidades representadas no eixo das ordenadas. De posse deste gráfico identificou-se a sazonalidade e o seu período sazonal. O valor encontrado para o período foi analisado pelo teste G de Fisher para confirmar a significância do valor encontrado.

A sazonalidade, componente muito comum em dados climatológicos, foi identificada e, para eliminá-la, efetuou-se uma diferença sazonal de tamanho 12, resultando em uma série estacionária, característica necessária para seguir com a metodologia de Box e Jenkins.

A classe de modelos $SARIMA(p, d, q) \times (P, D, Q)$ é uma generalização dos modelos $ARIMA$ de Box e Jenkins, sendo apropriada para descrever séries temporais que apresentam a componente de sazonalidade além de modelar a componente de tendência,

caso exista. Como foi detectado sazonalidade na série em estudo, esta classe se mostra uma boa alternativa para a modelagem.

Após a escolha da classe do modelo, iniciou-se a identificação das ordens dos parâmetros, considerando os valores de d como sendo o número de diferenças aplicadas para eliminar a tendência e D para eliminar a sazonalidade. Os demais parâmetros foram identificados criando modelos concorrentes, variando e combinando valores para os parâmetros, na busca de um modelo parcimonioso e que apresentou menor valor para o critério de informação bayesiano (BIC); utilizou-se ainda o critério de informação de Akaike corrigido (AICc) para confirmação do melhor modelo. As medidas de erros MSE e MAPE foram utilizadas na busca de um modelo bom para realizar previsões.

Os modelos concorrentes e suas comparações foram realizadas com o auxílio do software R (R, 2016) por meio do pacote *stats*, que ao criar os modelos concorrentes estimou seus parâmetros pelo método da máxima verossimilhança condicional, realizando a etapa de escolha e estimação dos modelos num mesmo processo.

Após especificar, identificar e ajustar o modelo, iniciou-se a verificação com base na análise residual. A análise dos resíduos foi iniciada por meio do teste de Shapiro-Wilk verificando se os resíduos seguem distribuição normal. Enquanto que o teste de Ljung-Box foi aplicado duas vezes, a primeira nos resíduos do modelo para verificar a correlação serial e a segunda nos resíduos elevados ao quadrado para verificar a homocedasticidade da variância do modelo.

Se o modelo ajustado para a série temporal tem a hipótese de variância homogênea rejeitada pelo teste de Ljung-Box, aplicado aos resíduos do modelo elevados ao quadrado, tem-se que a variância do modelo ajustado não possui um padrão constante, apresentando oscilações significativas ao longo do tempo, desta forma, deve-se modelar a variação dos resíduos, o que pode ser feito por meio de um modelo $GARCH(r)$.

Um bom modelo da classe $GARCH(r)$ pode ser obtido criando vários modelos concorrentes e ajustando aquele que apresentar menor valor para algum critério, neste caso utilizou-se o critério de informação de Akaike (AIC), que se mostrou um bom indicador de ajuste para tal classe. Modelando a variância condicional utilizando-se este modelo, faz com que os resíduos geralmente se comportem como ruído branco. Assim a série passa a ter sua correlação serial descrita por um modelo da classe SARIMA e a variância condicional descrita por um modelo da classe GARCH. Resultando em melhor ajuste e adequabilidade do modelo, conservando todas as condições necessárias para o uso de modelos de séries temporais.

Depois de passar pela fase de verificação da adequabilidade, os modelos ajustados para a série temporal foram utilizados para realizar as previsões para os próximos 12 meses da precipitação pluviométrica total mensal da cidade de Viçosa-MG.

O presente estudo contém os dados de janeiro de 1968 até julho de 2016 e após verificar o ajuste de modelo realizou-se previsões para os próximos 12 meses. Os dados reais de agosto de 2016 até janeiro de 2017 foram coletados ao longo do estudo e reservados para realizar comparação com os valores preditos pelo modelo, verificando a sua precisão.

3 RESULTADOS E DISCUSSÃO

No período estudado a série temporal da precipitação pluviométrica total mensal da cidade de Viçosa-MG apresentou uma média de 106,5 mm de chuva por mês, com um desvio padrão bastante alto de 108,92 mm. O desvio padrão elevado se deve ao fato de que em alguns meses no período de verão chove bastante e em outros meses chove pouco ou nada. O mês que apresentou maior quantidade de chuvas desde que os dados passaram a ser registrados foi dezembro de 2008 com 626 mm de chuvas.

Realizando uma análise visual no Gráfico 1 observa-se a série temporal da precipitação total mensal da cidade de Viçosa-MG incluindo ainda a reta de tendência no gráfico, que indica um valor constante para a precipitação pluviométrica, indicando que o índice de chuvas pode sofrer alterações de um ano para outro, porém não caracteriza-se nem com elevação nem com redução na quantidade de chuvas da cidade em estudo desde o ano de 1968.

No Gráfico 2 tem-se a série histórica juntamente com a sua decomposição em três componentes não observáveis: tendência, sazonalidade e efeito aleatório. Analisando cada componente separadamente nota-se por meio de uma análise visual que a série aparentemente não apresenta tendência, entretanto a sazonalidade anual fica bastante clara e bem definida, quando se analisa apenas esta componente separada das demais.

A análise da tendência com seu respectivo valor p para verificar a significância está apresentada na Tabela 1, em que é possível notar que não foi detectada tendência na série da precipitação pluviométrica mensal, contrariando o senso comum de que atualmente chove menos que anos atrás, o que muitas pessoas acreditam ser verdade.

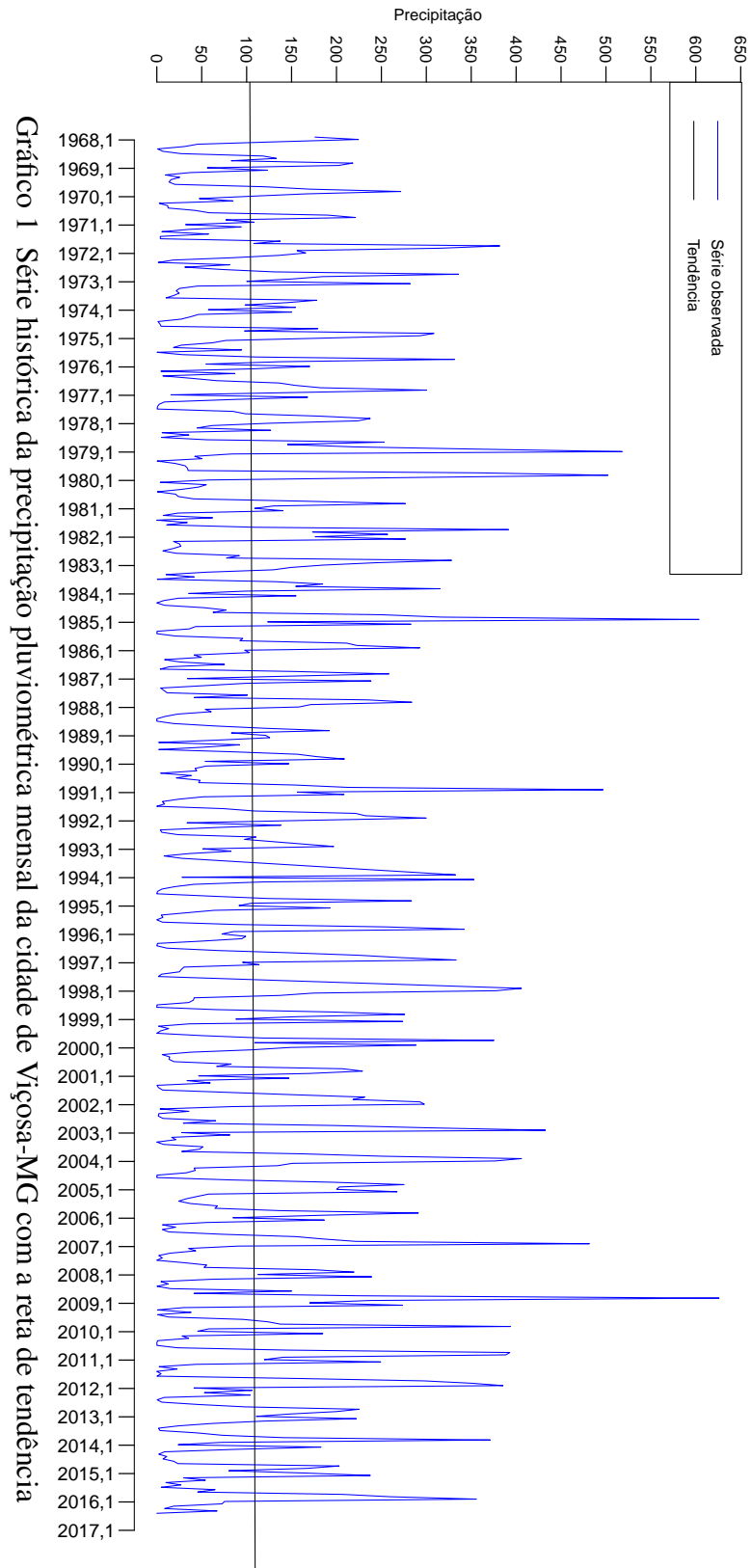


Gráfico 1 Série histórica da precipitação pluviométrica mensal da cidade de Viçosa-MG com a reta de tendência

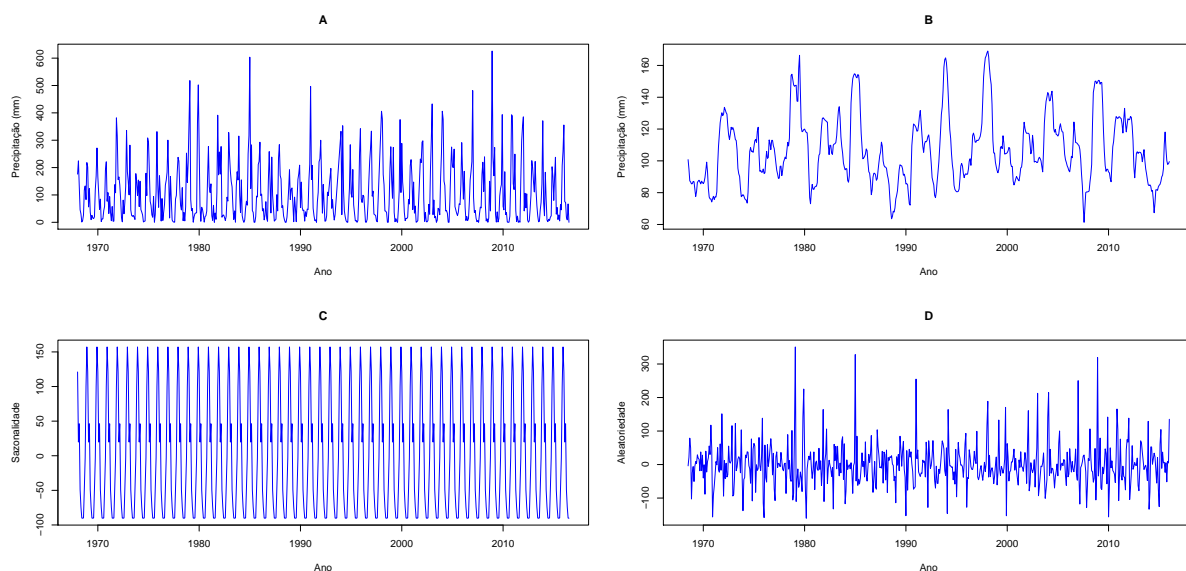


Gráfico 2 A - Série do índice mensal da precipitação pluviométrica de Viçosa-MG; B - Componente de tendência da série; C - Componente sazonal da série; D - Fator aleatório da série

Tabela 1 Análise da tendência para a série de precipitação pluviométrica

	Estimativa	Desvio padrão	Estatística t	Valor p
Intercepto	-118,3561	641,2812	-0,1850	0,8540
Tempo (x)	0,1129	0,3219	0,3510	0,7260

O boxplot de cada mês ao longo dos anos da série estudada encontra-se no Gráfico 3 e representa o esperado, com maiores índices de chuvas nos meses do verão e menores nos do inverno. Os meses que apresentam maiores intervalos de variação na precipitação, ou seja, que em alguns anos chove mais e em outros chove menos são janeiro e março, enquanto que os meses de junho, julho e agosto tem praticamente a mesma quantidade de chuvas todo ano, oscilando em um intervalo bem pequeno (amplitude baixa).

Os resultados do teste de Dickey-Fuller aumentado (ADF) encontram-se na Tabela 2, na qual é possível notar que a série em estudo não apresenta raiz unitária, o que equivale afirmar que a série não possui tendência estocástica, uma vez que a regressão do teste é estimada já aplicando uma diferença nos dados. A hipótese de raiz unitária ou tendência estocástica (H_0 : A série possui raiz unitária ou H_0 : A série possui tendência estocástica) foi rejeitada a 5% de probabilidade (valor $p < 0,0001$). O valor para a tendência determinística não rejeita a hipótese de nulidade (H_0 : A série não possui tendência determinística) a 5% com valor $p = 0,7257$ indicando que a série não apresenta tendência. Realizou-se o teste ADF com 22 defasagens da variável em primeira diferença, tais defasagens foram necessárias para eliminar a autocorrelação residual, que pode influenciar no teste. Com

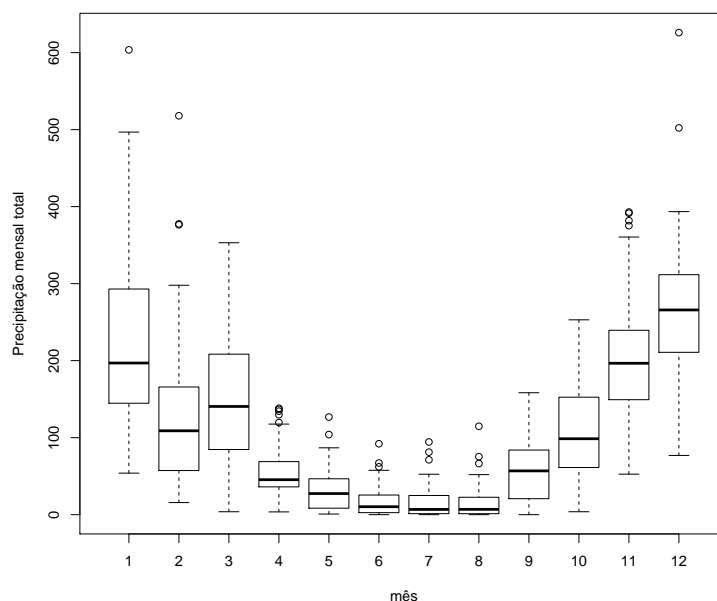


Gráfico 3 Boxplot de cada mês da série em anos diferentes

tais defasagens o teste de Ljung-Box apresentou valor $p = 0,1435$ confirmando a 5% de probabilidade que a correlação residual foi eliminada, não rejeitando a hipótese H_0 : Os resíduos são não autocorrelacionados.

Tabela 2 Resultado do teste de Dickey-Fuller aumentado

Parâmetros	β_1	β_2	π
Estimativa	127,3279	0,0071	-1,2075
Erro-padrão	24,2183	0,0203	0,2243
Estatística do teste	5,2570	0,3510	-5,3830
Valor p	< 0,0001	0,7257	< 0,0001
Teste Ljung-Box	Q=163,9800	valor $p = 0,1435$	

Para verificar a presença de sazonalidade, fator comum em dados climatológicos, foi utilizado o periodograma, apresentado no Gráfico 4. Note que não há necessidade de realizar transformações nos dados da série histórica, pois ela não apresenta o fator de tendência, visto pelo teste de Dickey-Fuller aumentado.

No gráfico do periodograma detecta-se que a maior intensidade espectral está associada a frequência de Fourier $f = 0,0841$ indicando uma periodicidade $s = 11,8906$, encontrada pela relação $s = 1/f$. O valor encontrado para a periodicidade da série foi testado pelo teste G de Fisher que apresentou estatística $G = 0,2745$, superior ao valor de comparação $Z = 0,0294$, tem-se ainda o valor $p < 0,0001$ confirmando que G é estatisticamente maior que Z. Portanto a série em estudo apresenta sazonalidade de tamanho aproximadamente 12 (sazonalidade anual), que pode ser eliminada aplicando uma

diferença sazonal de tamanho 12 na série, tornando a série livre de sazonalidade.

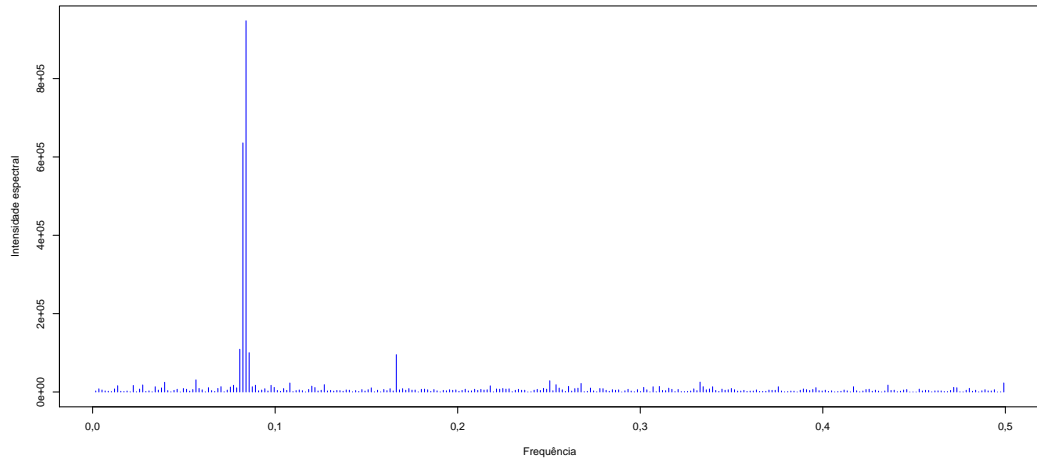


Gráfico 4 Periodograma da série da precipitação total mensal de Viçosa-MG

Após aplicar uma diferença de tamanho 12 nos dados, obteve-se uma série estacionária tornando possível a aplicação da metodologia de Box e Jenkins no processo de modelagem. No Gráfico 5 encontra-se a fac e facp da série da precipitação pluviométrica total mensal da cidade de Viçosa, que são utilizadas na escolha de um modelo para ajuste da série, no entanto, se torna difícil identificar uma classe e as ordens de um modelo baseando-se apenas na fac e facp da série original e estacionária.

Com base no Gráfico 5, nota-se que, após a aplicação da diferença a estrutura de autocorrelação dos dados foram em grande parte eliminadas, no entanto, nas defasagens próximas dos múltiplos de $s = 12$ alguns coeficientes da fac e facp que se apresentam significativos a 5% de probabilidade, indicando que um modelo da classe $SARIMA(p, d, q) \times (P, D, Q)$ se mostra apropriado para estes dados.

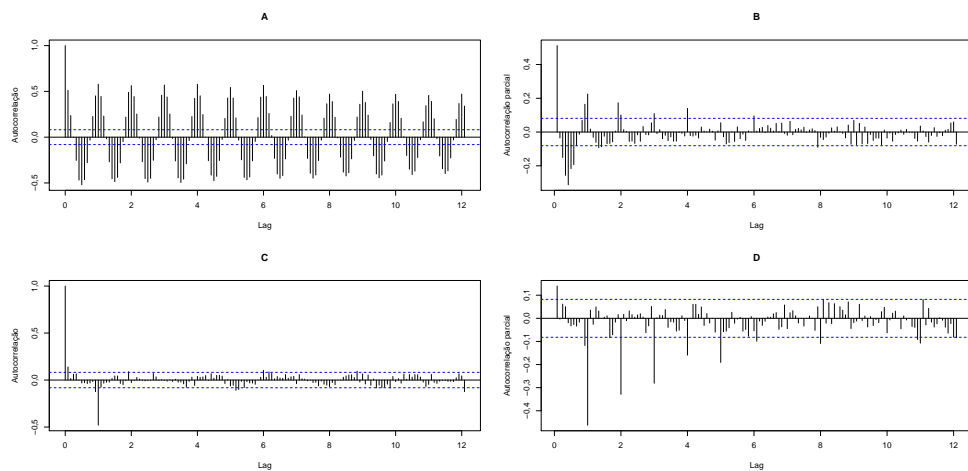


Gráfico 5 A - Fac da série original; B - Facp da série original; C - Fac da série diferenciada; D - Facp da série diferenciada

Ao realizar o tratamento dos dados a fim de obter uma série estacionária verificou-se que a mesma não possui tendência, enquanto que identificou-se um fator sazonal de tamanho 12 que foi eliminado por meio de uma diferença de mesmo tamanho. Assim o modelo a ser ajustado aos dados é da forma $SARIMA(p, d, q) \times (P, D, Q)$, em que $d = 0$ (não necessitou aplicar diferenças pra remover a tendência) e $D = 1$ (foi aplicada uma diferença de tamanho 12 para remover o fator sazonal). Desta forma, estão apresentados na Tabela 3 vários modelos concorrentes que foram criados para seleção variando a ordem dos demais fatores. Aquele que apresentou menor valor para o critério de informação bayesiano (BIC) foi o escolhido para modelar a série histórica; o critério de Akaike corrigido (AICc) foi utilizado para confirmação do modelo, enquanto que o indicativo de Erro Quadrático Médio (Mean Squared Error - MSE) e o indicativo de Erro Percentual Médio Absoluto (Mean Absolute Percentual Error - MAPE) foram calculados para obter um bom modelo para realizar previsões.

Tabela 3 Comparação de diversos modelos concorrentes para a série da precipitação pluviométrica total mensal de Viçosa-MG

Modelo	BIC	AICc	MSE	MAPE
$SARIMA(0, 0, 0) \times (0, 1, 0)$	6875,5100	6871,1690	1522,7160	40,6612
$SARIMA(0, 0, 0) \times (0, 1, 1)$	6518,0060	6509,3330	780,4648	29,1273
$SARIMA(0, 0, 0) \times (0, 1, 2)$	6523,9810	6510,9810	763,0412	28,8009
$SARIMA(1, 0, 0) \times (0, 1, 1)$	6516,6530	6503,6540	806,5600	28,8372
$SARIMA(2, 0, 0) \times (0, 1, 1)$	6522,4940	6505,1750	806,7800	29,0667
$SARIMA(1, 0, 1) \times (0, 1, 1)$	6521,8490	6504,5300	803,7225	27,9538
$SARIMA(1, 0, 0) \times (1, 1, 1)$	6522,8200	6505,5010	794,6761	28,6265

Baseando-se nos valores da Tabela 3 o modelo escolhido para um bom ajuste aos dados é o $SARIMA(1, 0, 0) \times (0, 1, 1)$, enquanto que para realização das previsões outros modelos apresentaram menores valores nos indicativos de erros, entretanto ao realizar uma análise nos coeficientes do modelo $SARIMA(0, 0, 0) \times (0, 1, 2)$ o coeficiente de ordem 2 de médias móveis sazonal apresentou-se não significativo (valor $p = 0,5402$) e o mesmo ocorreu com o modelo $SARIMA(1, 0, 1) \times (0, 1, 1)$ em que o coeficiente de médias móveis de ordem 1 apresentou-se não significativo (valor $p = 0,1498$). Portanto o modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$ foi o escolhido para representar os dados e realizar as previsões, visto que este modelo apresenta todas as ordens significativas.

O modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$ foi escolhido pelos critérios de seleção e teve seus parâmetros estimados pelo método da máxima verossimilhança condicional por meio do pacote *stats* do software R. As estimativas seguidas do valor p para o teste de significância encontram-se na Tabela 4.

Tabela 4 Estimativa dos parâmetros do modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$

Coeficiente	Estimativa	Valor p
ϕ_1	0,1157	0,0054
Θ_1	-0,9772	< 0,0001

Após especificar, identificar e estimar um modelo que melhor se ajustou aos dados, realizou-se a análise residual para verificar a qualidade do ajuste. Na Tabela 5 encontram-se os resultados dos testes aplicados aos resíduos do modelo.

Tabela 5 Resultado da análise residual do modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$

Teste	Estatística	Valor p
Shapiro-Wilk	0,8931	< 0,0001
Ljung-Box para resíduos	122,8100	0,9165
Ljung-Box para resíduos elevados ao quadrado	45,9330	< 0,0001

Pelo teste de Shapiro-Wilk verificou-se que os resíduos do modelo não seguem distribuição normal. Com isso o intervalo de confiança para a previsão não pode ser construído com base na distribuição normal, diante disso, utilizou-se o método de reamostragem *bootstrap*.

Quanto ao teste de Ljung-Box aplicado aos resíduos do modelo conclui-se que eles não são autocorrelacionados, enquanto que para o mesmo teste aplicado aos resíduos elevados ao quadrado verificou-se que o modelo não apresenta variâncias homogêneas. Portanto conclui-se que o modelo em questão não atende aos pressupostos para o ajuste dos dados por meio da metodologia de Box e Jenkins.

A ratificação do teste de Shapiro-Wilk se encontra no Gráfico 6 em que nota-se os resíduos observados desviando-se do padrão de normalidade. No Gráfico 7 tem-se a confirmação do teste de Ljung-Box aplicado aos resíduos em que nota-se os picos significativos de autocorrelação ocorrendo de forma aleatória indicando que os resíduos são não correlacionados. Para o mesmo teste aplicado aos resíduos elevados ao quadrado nota-se um número relativamente alto de autocorrelações significativas e a ocorrência deles de forma periódica, indicando a não homogeneidade da variância do modelo.

O modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$ escolhido por apresentar melhor ajuste que os demais se mostrou inadequado, apresentando heteroscedasticidade da variância residual. Desta forma, uma alternativa para contornar tal situação é modelar a variância por meio de um modelo não-linear da classe ARCH/GARCH.

O modelo não-linear da classe ARCH/GARCH foi escolhido por meio de comparação, em que foram criados diversos modelos da classe e aquele que apresentou me-

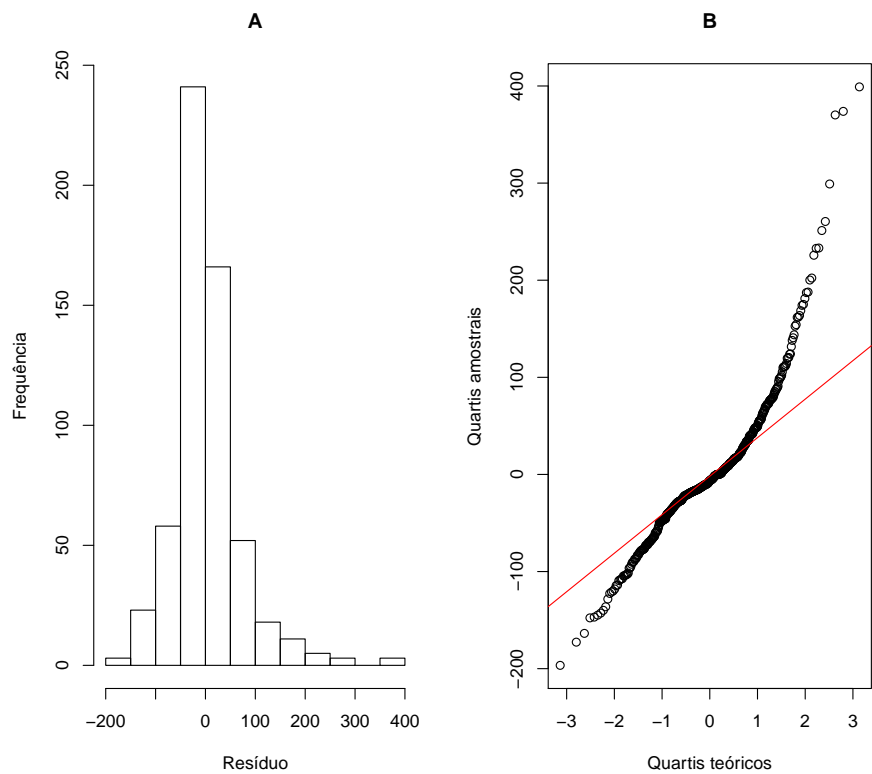


Gráfico 6 A - Histograma dos resíduos; B - QQ-plot dos resíduos

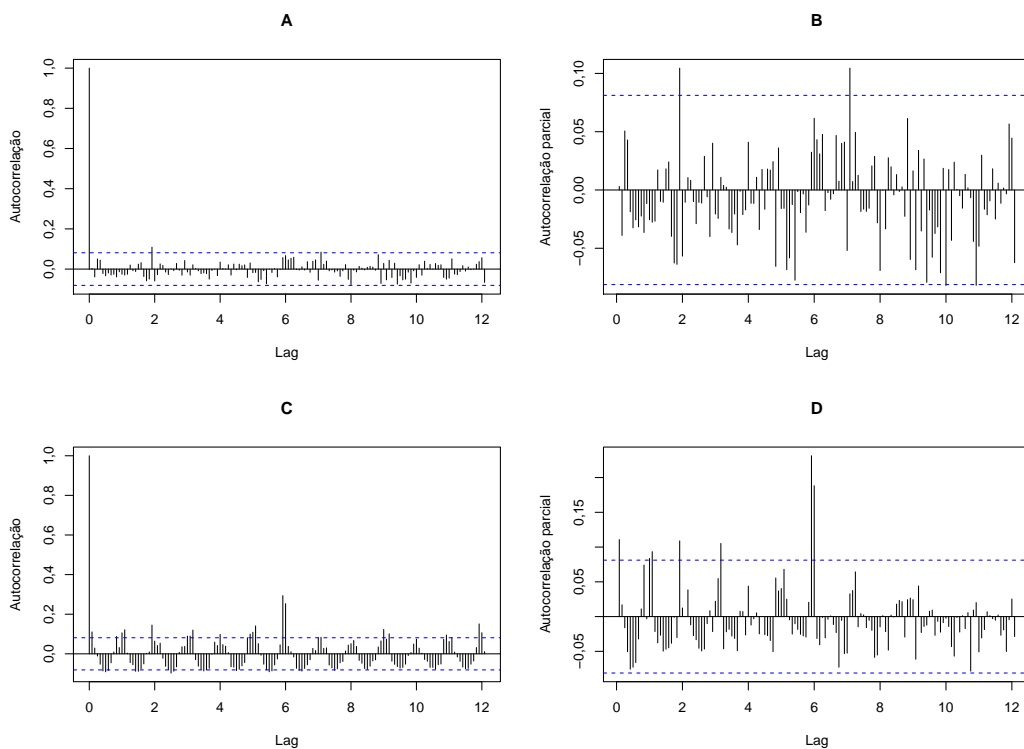


Gráfico 7 Gráfico da fac e facp dos resíduos e dos resíduos elevados ao quadrado

nor valor para o critério de informação de Akaike (AIC) foi selecionado para modelar a variância condicional, pois os critérios de seleção BIC e AICc apresentaram resultados semelhantes. Encontra-se na Tabela 6 os modelos que estão em comparação seguidos de

seus valores de AIC.

Tabela 6 Comparação de modelos concorrentes da classe ARCH/GARCH

Modelo	AIC
GARCH(1,0)	6516,2430
GARCH(1,1)	6538,8650
GARCH(2,0)	6506,3250
GARCH(3,0)	6517,8230
GARCH(4,0)	6519,4400
GARCH(2,1)	6526,5730
GARCH(0,1)	6579,1170
GARCH(0,2)	6570,8180

Com base na Tabela 6 tem-se que o melhor modelo para ajustar a variância dos resíduos do modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$ é o $GARCH(2, 0) = ARCH(2)$. Na Tabela 7 encontra-se a estimativa dos parâmetros do modelo $ARCH(2)$ juntamente com o valor p que indica a significância das estimativas para os parâmetros do modelo. Note que o termo de ordem dois do modelo $ARCH(2)$ apresenta-se como não significativo, sendo passível de retirada do modelo, porém sua presença interfere muito pouco nos resultados, devido ao fato de seu valor ser pequeno.

Tabela 7 Estimativa dos parâmetros do modelo $ARCH(2)$

Coefficiente	Estimativa	Valor p
a_0	2736,0000	< 0,0001
a_1	0,5279	< 0,0001
a_2	0,0446	0,2490

Após modelar a série combinando o modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$ para ajustar a correlação serial e o $ARCH(2)$ para modelar a variância dos resíduos obteve-se um modelo bem ajustado e com resíduos se comportando como um ruído branco. O teste de Ljung-Box aplicado aos resíduos elevados ao quadrado não rejeitou a hipótese de nulidade a 5% de probabilidade, apresentando valor $p = 0,6215$ indicando que tais modelos apresentam variâncias homogêneas. O mesmo teste aplicado aos resíduos do modelo apresentou valor $p = 0,6597$ indicando que os resíduos não são autocorrelacionados.

Com a modelagem adequada para série temporal em estudo realizou-se a previsão para os meses de agosto de 2016 até julho de 2017. Na Tabela 8 encontra-se a previsão do modelo $SARIMA(1, 0, 0) \times (0, 1, 1)$ acrescentada com a previsão dos resíduos por meio do modelo $ARCH(2)$. A junção dos modelos apresentou melhor ajuste, se comparado a apenas um modelo, além de atender aos pressupostos da modelagem de uma série temporal por meio das classes de Box e Jenkins.

Tabela 8 Previsão para os próximos 12 meses da precipitação total mensal

Mês	Previsão	Intervalo de confiança	
		LI	LS
ago/16	85,3172	0,0000	150,9159
set/16	89,2179	0,0000	195,2651
out/16	113,7109	0,0000	238,6442
nov/16	265,1579	69,6856	344,1724
dez/16	309,4285	130,4515	404,9381
jan/17	287,9018	91,0807	365,5297
fev/17	160,0424	0,0000	263,7759
mar/17	162,6419	17,0039	291,4530
abr/17	97,3947	0,0000	193,8772
mai/17	70,3734	0,0000	169,3776
jun/17	21,4735	0,0000	155,2054
jul/17	56,6052	0,0000	152,6015

A previsão para precipitação pluviométrica total mensal é bastante difícil de ser realizada, pois ela depende de diversos fatores, e não apenas de observações anteriores, por isso com base nos modelos de séries temporais que caracterizam-se por dependência entre as observações os erros podem ser elevados para alguns meses. Destaca-se na previsão a esperança de bastantes chuvas para os meses de outubro, novembro, dezembro, janeiro, fevereiro e março e poucas chuvas para os demais meses.

4 CONCLUSÃO

Ao estudar a série temporal da precipitação pluviométrica total mensal da cidade de Viçosa-MG conclui-se que o nível de precipitação segue sem alterações desde o início do registro dos dados em janeiro de 1968. Detectou-se que em alguns anos choveu mais e outros menos, porém esta variação ocorreu de forma aleatória, não podendo afirmar que houveram alterações neste fator do clima.

Conclui-se ainda que a população que depende diretamente de chuvas regulares segue conseguindo desempenhar suas atividades de forma normal, salvo alguns anos de maior estiagem ou de excesso de chuvas em determinadas épocas do ano, mas estes efeitos são observados desde o início dos dados.

A falta de água que afeta a cidade de Viçosa-MG quase todos os anos não pode ser explicada pela redução das chuvas e sim ao aumento da população, pela falta de cuidado com o meio ambiente de modo geral, e a falta de melhor planejamento das empresas que gerem os recursos hídricos da cidade.

As previsões realizadas para o índice de precipitação total mensal para a cidade de Viçosa-MG apresentaram bons resultados, porém uma variância bastante alta, devido ao fato de que as chuvas de uma região dependem de diversos fatores e os modelos de séries temporais levam em consideração apenas acontecimentos passados. As previsões para a precipitação podem apresentar grandes erros para determinados meses. Entretanto conclui-se que os modelos $SARIMA(1, 0, 0) \times (0, 1, 1)$ para ajustar a correlação serial e o $ARCH(2)$ modelando os resíduos se mostraram bons e eficientes para o ajuste proposto.

REFERÊNCIAS

AYOADE, J. O. **Introdução à climatologia para os tópicos**. 10. ed. Rio de Janeiro: Bertrand Brasil, 2006. 332 p.

BATISTA, A. L. F. **Modelos de séries temporais e redes neurais artificiais na previsão de vazão**. 2009. 89 p. Dissertação (Mestrado em Modelagem de Sistemas Biológicos) - Programa de pós-graduação em Engenharia de Sistemas, Universidade Federal de Lavras, Lavras. 2009.

BAYER, D. M.; CASTRO, N. M. R. Modelagem e previsão de vazões médias mensais do Rio Potitibu utilizando modelos de séries temporais. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v. 17, n. 2, p. 229-239, 2012.

MERTEN, G. H.; MINELLA, J. P. Qualidade da água em bacias hidrográficas rurais: um desafio atual para a sobrevivência futura. **Revista Agroecologia e Desenvolvimento Rural Sustentável**, Porto Alegre, v. 3, n. 4, p. 33-38, 2002.

PEREIRA, A. R.; COSTA, A. S.; OLIVEIRA, V. G.; BORGES, P. F.; FILHO, A. I.; Análise do comportamento das médias anuais da precipitação pluvial e temperatura da cidade de Areia, Paraíba. **Gaia Scientia**, João Pessoa, v. 9, n. 1, p. 67-73, 2015.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2016. Disponível em: <http://www.r-project.org>. Acesso em: 13 julho 2016.

SÁFADI, T. Uso de séries temporais na análise de vazão de água na represa de Furnas. **Ciência e agrotecnologia**, Lavras, v. 28, n. 1, p. 142-148, 2004.

SILVA, J. W.; BEIJO, L. A.; SÁFADI, T. Uso de modelos de séries temporais na previsão de precipitações pluviométricas mensais no município de Uberaba-MG. In: CONGRESSO DE PÓS-GRADUAÇÃO DA UFLA / ESTATÍSTICA E EXPERIMENTAÇÃO. 11., 2002, Lavras. **Anais...** Lavras, 2002. p. 75-79.

SILVA, V. P. R.; PEREIRA, E. R. R.; AZEVEDO, P. V.; SOUSA, F. A. S.; SOUSA, I. F. Análise da pluviometria e dias chuvosos na região Nordeste do Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 15, n. 2, p. 131-138, 2011.

UNIVERSIDADE FEDERAL DE VIÇOSA - UFV. Departamento de Engenharia Agrícola. Estação Climatológica Principal de Viçosa-MG. **Boletim meteorológico 2016**. Viçosa, 2016. Disponível em http://www.posmet.ufv.br/?page_id=1253. Acesso em 7/6/2016.

APÊNDICE

	Página
APÊNDICE A: Rotina para análise da temperatura máxima.	78
APÊNDICE B: Rotina para análise da temperatura mínima.	84
APÊNDICE C: Rotina para análise da precipitação pluviométrica.	89

APÊNDICE A: Rotina para análise da temperatura máxima.

```
#install.packages("forecast","timeseries","urca","stats","GeneCycle","lmtest","TSA","sme")
library("forecast")
library("tseries")
library("urca")
library("stats")
library("GeneCycle")
library("lmtest") # Obter os p-values
library("TSA")
library("sme")

#Carregando os dados
setwd("C:/Users/Paulo César/Desktop/Mauricio/Dados")
dados<-read.table('C:/Users/Paulo César/Desktop/Mauricio
/Dados/dadosatualizados.txt',h=T, sep = "")
dados
attach(dados)

#Calcular a média mensal da temperatura máxima de Viçosa

options(OutDec=",")
ano=c(31,28,31,30,31,30,31,31,30,31,30,31)
bis=c(31,29,31,30,31,30,31,31,30,31,30,31)
inibis=cbind(bis,ano,ano,ano)
diasmeses=cbind(rep(inibis,13))
#a=rnorm(sum(diasmeses),0,1)
a=dados$Tmax
mensa=NULL
tmaxmensal1=NULL
t=1
for(k in 1:length(diasmeses)){
tmaxmensal1<-rbind(tmaxmensal1,mean(a[t:(t+diasmeses[k]-1)]))
t=sum(diasmeses[1:k])+1
}
tmaxmensal1
tmaxmensal=tmaxmensal1[1:583]

#Interpolar dados da temperatura máxima

yo=ts(tmaxmensal[1:length(tmaxmensal)],start=c(1968,1),
frequency=365)
ytmax=ts(na.interp(yo, lambda = NULL),start=c(1968,1),
frequency=365)
ytmax # temperatura máxima
summary(ytmax)
var(ytmax)

#Série e a decomposição de suas componentes

x=ts(ytmax, start = c(1968, 1), end = c(2016, 7), frequency = 12)
y=ts(x[1:583], start=c(1968,1), frequency=12)
ts.plot(y,main="Série das médias mensais das temperaturas máximas
de Viçosa-MG", col=c("blue"), lwd=1, gpars=list(xaxt="n",axes=F),
xlab="", axes=F, ylab="Média mensal das temperaturas máximas",
ylim=c(min(y),max(y)))
axis(1, at=1968.1:2017.1,labels=1968.1:2017.1,
tcl= -.8,lwd.ticks=1.2,las=2, cex.axis=.7)
lim=seq(22,32,by=.5)
axis(2, at=lim, tcl= -.8,lwd.ticks=1.2,las=2)
trend=lm(x-time(x))
summary(trend) # Estimativa de time(x)
```

```

me fornece a inclinação da tendência anual
abline(trend, col="black")
legend("topleft", c("Série real","Tendência"),
col=c(4,1), lty=c(1,1))

m <- decompose(x)
par(mfrow=c(2,2))
plot(m$x, col="blue", xlab="Ano",ylab="",
main="Série observada")
plot(m$trend, col="blue", xlab="Ano",ylab="",
main="Componente de tendência da série")
plot(m$seasonal, col="blue", xlab="Ano",ylab="",
main="Componente sazonal da série")
plot(m$random, col="blue", xlab="Ano",ylab="",
main="Fator aleatório da série")

#Fazendo uma análise mês a mês

mes=cycle(x)

#Janeiro (mes==1) Fevereiro (mes==2)... Dezembro (mes==12)
jan=subset(x,mes==7)
xjan=ts(jan, start = c(1968), end = c(2016), frequency = 1)
xjan
trendjan=lm(xjan~time(xjan))
summary(trendjan) # Estimativa de time(xjan) me fornece a tendência
para o mês de janeiro (verificar qual é significativa pelo teste t)

ts.plot(xjan,main="Série da média do mês de janeiro da temperatura
máxima de Viçosa", col=c("blue"), lwd=1,
gpars=list(xaxt="n",axes=F),xlab="", axes=F,
ylab="Média mensal da temperatura máxima",
ylim=c(min(xjan),max(xjan)))
axis(1, at=1968.1:2016.12,labels=1968:2016, tcl= -.8,lwd.ticks=1.2,
las=2, cex.axis=.7)
lim=seq(20,32.5,by=.5)
axis(2, at=lim, tcl= -.8,lwd.ticks=1.2, las=2)

# Boxplot da série
mes1=rep(seq(1:12),49)
mes=mes1[1:583]
boxplot(x~mes,main="Boxplot da série")

# Verificando ACF e PACF

n=length(x) #tamanho da amostra
par(mfrow=c(2,1))
acf(x,n/4, main="FAC da Série de Média Mensal da
Temperatura Máxima de Viçosa-MG")
pacf(x,n/4, main="FACP da Série de Média Mensal
da Temperatura Máxima de Viçosa-MG")

# Teste Dickey-Fuller

ADF<-ur.df(x, type="trend", lags=13) #Foram necessárias
13 defasagens para que o resíduo não fosse correlacionados
summary(ADF)
regres=ADF@res
Box.test(regres, n/4, type = c("Ljung")) #Para verificar
se os resíduos são ruído branco

```

```

plot(ADF)

# Série , ACF e PACF após diferença

dif_1<-diff(x,1)
par(mfrow=c(1,1))
plot.ts(dif_1,main="Série de Média Mensal da Temperatura Máxima
de Viçosa-MG após uma diferença", col="blue")
trenddif=lm(dif_1~time(dif_1))
abline(trenddif, col="black")
par(mfrow=c(2,1))
acf(dif_1,n/4, main="FAC da Série de Média Mensal da Temperatura
Máxima de Viçosa-MG após uma diferença")
pacf(dif_1,n/4, main="FACP da Série de Média Mensal da Temperatura
Máxima de Viçosa-MG após uma diferença")

# Conferindo se a série se tornou estacionária após a diferença

ADF_1<-ur.df(dif_1, type=c("trend"),lags=15)
summary(ADF_1)
regres1=ADF_1@res
Box.test(regres1, n/4, type = c("Ljung"))
plot(ADF_1)

# Rejeitar H0 para z.lag.1 indica estacionaridade
(sem trend estocástico , Ro < 1)
# Não Rejeitar H0 para tt indica que a trend
determinística foi eliminada

#MannKendall(dif_1) #H0 não há tendência

# Rejeitar H0 para z.lag.1 indica estacionaridade
(sem trend estocástico , Ro < 1)
# Não Rejeitar H0 para tt indica que a trend
determinística foi eliminada

#Periodograma
par(mfrow=c(1,1))
p=periodogram(dif_1,main="Periodograma", ylab="Intensidade
Espectral",xlab="Frequências de Fourier", col="blue");abline(h=0)
f<-p$freq
ds<-p$spec
plot(f, ds, type="l", col="blue", ylab="Intensidade Espectral",
xlab="Frequência", main="Periodograma") # Olhar a posição do maior,
s = posição/número de observações
periodicidade<-1/f[ds==max(ds)]
periodicidade

# Teste G de Fisher #H0 Não existe periodicidade

n=length(x)
alpha<-0.05
N<-n/2
(G<-max(ds)/sum(ds)) # sum(ds) de 1 ate n/2
(Z<-(1-((alpha/N)^(1/(N-1))))))
(pvalor=N*((1-G)^(N-1))) # p-valor para o teste G
(Moretin e Toloi 2006, pg. 458)

# Diferença Sazonal de lag s = 12 aplicada a série já com
1 diferença de lag 1

```

```

dif_s<-diff(dif_1 , lag = 12)
par(mfrow=c(3,1))
plot.ts(dif_s , main="Série após aplicação das diferenças",
ylab="Série após diferenças", xlab="Tempo", col="blue")
acf(dif_s ,n/4 , main="Função de autocorrelação após aplicação
das diferenças",ylab="Autocorrelação", xlab="Lag")
pacf(dif_s ,n/4 , main="Função de autocorrelação parcial após
aplicação das diferenças",ylab="Autocorrelação parcial", xlab="Lag")

# Ajustando o Modelo SARIMA
#Alguns gráficos interessantes antes de ajustar um modelo
par(mfrow=c(2,2))
acf(x,n/4 , main="FAC da série original")
pacf(x,n/4 , main="FACP da série original")
acf(dif_s ,n/4 , main="FAC da série diferenciada",
ylab="Autocorrelação", xlab="Lag")
pacf(dif_s ,n/4 , main="FACP da série diferenciada",
ylab="Autocorrelação parcial", xlab="Lag")

M1<-arima(x, order = c(2,1,3),
seasonal = list(order = c(0, 1, 1), period = 12))
M1
AICc(M1)
BIC(M1)
coefstest(M1)
fit<-fitted(M1)
acur=accuracy(fit ,ytmx) # Medida de acurácia do modelo
acur # RMSE fornece a raiz do erro quadrático médio

# Iniciando a Análise Residual

# Teste de Normalidade de Shapiro–Wilk (Normalidade)
Res<-rstandard(M1) # resíduos padronizados
a<-shapiro.test(Res)
a

#Teste de Box–Pierce e Ljung–Box (independencia)
Box.test(Res, n/4,type = c("Box–Pierce"))
Box.test(Res, n/4,type = c("Ljung"))

# ACF e PACF dos resíduos estimados

par(mfrow=c(2,1))
acf(Res,n/4 , main="FAC da série dos resíduos do modelo ajustado")
pacf(Res,n/4 , main="FACP da série dos resíduos do modelo ajustado")

# Homocedasticidade para os resíduos (Tomei a série dos Res^2)

par(mfrow = c(2,1))
acf(Res^2,n/4 , main="FAC da série dos resíduos
elevados ao quadrado do modelo ajustado")
pacf(Res^2,n/4 , main="FACP da série dos
resíduos elevados ao quadrado do modelo ajustado")
Box.test(Res^2, 15, type = c("Ljung"))
# Não Rejeitar H0 indica variância Constante

#Gráficos interessantes
par(mfrow=c(2,2))
acf(Res,n/4 , ylab="Autocorrelação",

```

```

main="FAC dos resíduos")
pacf(Res, n/4, ylab="Autocorrelação parcial",
main="FACP dos resíduos")
acf(Res^2, n/4, ylab="Autocorrelação",
main="FAC dos resíduos elevados ao quadrado")
pacf(Res^2, n/4, ylab="Autocorrelação parcial",
main="FACP dos resíduos elevados ao quadrado")
par(mfrow=c(1,2))
hist(Res, ylab="Frequência", xlab="Resíduo",
main="Histograma dos resíduos")
qqnorm(Res, ylab="Quartis amostrais",
xlab="Quartis teóricos"); qqline(Res, col=2)

#Testando um modelo SARIMA
npred=12
y=ts(ytmax[1:583], start=c(1968,1), frequency=12)
yP=ts(ytmax[1:(583 - npred)], start=c(1968,1), frequency=12)
sarimaP=arima(yP, order = c(2,1,2),
seasonal = list(order = c(0, 1, 1), period = 12))
sarimaP
predi2015_2016<-forecast(yP,12, level=c(95))
summary(predi2015_2016)

#Plotando a série dos valores observados e os valores ajustados
par(mfrow = c(1,1))
zest2=yP-sarimaP$residuals
plot.ts(yP, type="l", col="red")
lines(zest2)
legend("topleft", legend=c("valores observados",
"valores ajustados"), lty=c(1,1), col=c(2,1),
pch=c(NA,NA), bty="n")

#Predição com o modelo ajustado retirando os últimos 12 meses para
confirmação
predict(sarimaP, n.ahead = npred)$pred

#Valor predito subtraído do observado
y[(length(y)-npred+1):length(y)]
y[(length(y)-npred+1):length(y)]-predict(sarimaP,
n.ahead = npred)$pred

#Gráfico da previsão
fore <- predict(sarimaP, n.ahead=npred)
#U <- fore$pred+2*fore$se
#L <- fore$pred-2*fore$se
ts.plot(y, zest2, fore$pred, gpars=list(xlab="Data",
ylab="temperatura", lty=c(1,1,1,1,1), lwd=c(1,1,1,1,1),
col=c(1,2,3,4,4)))
legend("topleft", c("Verdadeiro", "Ajustado", "Predito"),
col=c(1,2,3), lty=c(1,1,1))

#Gráfico da previsão dando zoom na parte final
yat=48
vmax=length(y)
Diaini=vmax-yat
fy=ts(y[(vmax-yat):vmax], frequency=12, start=c(2012,7))
fyP=ts(zest2[(vmax-yat):(vmax-npred)], frequency=12, start=c(2012,7))
#U <- fore$pred+2*fore$se
#L <- fore$pred-2*fore$se

```

```

ts.plot(fy,fyP, fore$pred, main="Gráfico da previsão da média mensal
da temperatura máxima ", gpars=list(xlab="Data", ylab="Temperatura",
lty=c(1,1,1,2,2),lwd=c(2,2,2,2,2),col=c(1,2,3,4,4))
legend("topleft", c("Verdadeiro","Ajustado", "Predito"),
      col=c(1,2,3), lty=c(1,1,1))

# Previsão com o Modelo
z<-forecast(M1,12, level=c(95,95), bootstrap=T, npaths=50000)
#para calcular IC(95%) por reamostragem
fit<-fitted(M1)
plot(z, main="Previsão a partir do modelo SARIMA(2,1,2)(0,1,1)[12]",
col = "black", xlab = "meses", ylab = "Temperatura máxima")
lines(fit, col="red", lwd=1)
summary(z) # Fornece previsão, intervalo de
confiança e indicativo de erro MAPE, MAE, RMSE...

#Fim

```

APÊNDICE B: Rotina para análise da temperatura mínima.

```
#install.packages("forecast","timeseries","urca","stats","GeneCycle","lmtest","TSA","sme")
library("forecast")
library("timeseries")
library("urca")
library("stats")
library("GeneCycle")
library("lmtest") # Obter os p-values
library("TSA")
library("sme")

#Carregando os dados
setwd("C:/Users/Paulo César/Desktop/Mauricio/Dados")
dados<-read.table('C:/Users/Paulo César/Desktop/Mauricio/
Dados/dadosatualizados.txt',h=T, sep = "")
dados
attach(dados)

#Calcular a média da temperatura mínima de Viçosa

options(OutDec=",")
ano=c(31,28,31,30,31,30,31,31,30,31,30,31)
bis=c(31,29,31,30,31,30,31,31,30,31,30,31)
inibis=cbind(bis,ano,ano,ano)
diasmeses=cbind(rep(inibis,13))
#a=rnorm(sum(diasmeses),0,1)
a=dados$Tmin
mensa=NULL
tminmensal1=NULL
t=1
for(k in 1:length(diasmeses)){
tminmensal1<-rbind(tminmensal1,mean(a[t:(t+diasmeses[k]-1)]))
t=sum(diasmeses[1:k])+1
}
tminmensal1
tminmensal = tminmensal1[1:583]

#Interpolar dados da temperatura mínima
yo=ts(tminmensal[1:length(tminmensal)],start=c(1968,1),
frequency=365)
ytmin=ts(na.interp(yo, lambda = NULL),start=c(1968,1),
frequency=365)
ytmin # temperatura mínima
summary(ytmin)
var(ytmin)

#Série e a decomposição de suas componentes
x<-ts(ytmin, start = c(1968, 1), end = c(2016, 7), frequency = 12)
y=ts(x[1:583],start=c(1968,1), frequency=12)
ts.plot(y, main="Série das médias mensais das temperaturas mínimas
de Viçosa-MG", col=c("blue"), lwd=1, gpars=list(xaxt="n", axes=F),
xlab="", ylab="Média mensal das temperaturas mínimas",ylim=c(min(y),
max(y)))
axis(1, at=1968.1:2016.7, labels=1968.1:2016.7, tcl= -.8,
lwd.ticks=1.2, las=2, cex.axis=.7)
lim=seq(7,20,by=.5)
axis(2, at=lim, tcl= -.8,lwd.ticks=1.2, las=2)
trend=lm(x-time(x))
summary(trend) # Estimativa de time(x) me fornece a inclinação da
tendência anual
```

```

abline(trend)
legend("topleft", c("Série real","Tendência"), col=c(4,1),
lty=c(1,1))

m <- decompose(x)
par(mfrow=c(2,2))
plot(m$x, col="blue", xlab="Ano",ylab="", main="Série observada")
plot(m$trend, col="blue", xlab="Ano",ylab="",
main="Componente de tendência da série")
plot(m$seasonal, col="blue", xlab="Ano",ylab="",
main="Componente sazonal da série")
plot(m$random, col="blue", xlab="Ano",ylab="",
main="Fator aleatório da série")

#Fazendo uma análise mês a mês

mes=cycle(x)

#Janeiro (mes==1) Fevereiro (mes==2)... Dezembro (mes==12)
par(mfrow=c(1,1))
jan=subset(x,mes==1)
xjan=ts(jan, start = c(1968), end = c(2016), frequency = 1)
xjan
trendjan=lm(xjan~time(xjan))
summary(trendjan) # Estimativa de time(xjan) me fornece a tendência
anual para o mês de janeiro (verificar qual é significativa pelo
teste t)

ts.plot(xjan,main="Série da média do mês de janeiro da temperatura
mínima de Viçosa", col=c("blue"), lwd=1,
gpars=list(xaxt="n",axes=F), xlab="", axes=F, ylab="",
ylim=c(min(xjan),max(xjan)))
axis(1, at=1968.1:2015.12, labels=1968:2015, tcl= -.8,lwd.ticks=1.2,
las=2, cex.axis=.7)
lim=seq(1,20.5,by=.5)
axis(2, at=lim, tcl= -.8,lwd.ticks=1.2, las=2)

# Boxplot da série
mes1=rep(seq(1:12),49)
mes=mes1[1:583]
boxplot(x~mes, main="Boxplot da série")

# Verificando ACF e PACF
n=length(x) #tamanho da amostra
par(mfrow=c(2,1))
acf(x,n/4, ylab="Autocorrelação", main="FAC da Série de Média
Mensal da Temperatura Mínima de Viçosa-MG")
pacf(x,n/4, ylab="Autocorrelação parcial", main="FACP da Série
de Média Mensal da Temperatura Mínima de Viçosa-MG")

# Teste Dickey-Fuller
ADF<-ur.df(x, type="trend", lags=21) #Foram necessárias 13
defasagens para que o resíduo não fosse correlacionados
summary(ADF) # Rejeitar H0 indica estacionaridade
regres=ADF@res
Box.test(regres, n/4, type = c("Ljung")) #Para verificar se os
resíduos são ruído branco
plot(ADF)

```



```

# Série, ACF e PACF após diferença (se necessário diferenciar)
dif_1<-diff(x,1)
par(mfrow=c(1,1))
plot.ts(dif_1, main="Série de Média Mensal da Temperatura Mínima de
Viçosa-MG após uma diferença",col="blue")
par(mfrow=c(2,1))
acf(dif_1,n/4)
pacf(dif_1,n/4)

# Conferindo se a série ficou estacionária após a diferença

ADF_1<-ur.df(dif_1, type=c("trend"),lags=26)
summary(ADF_1)
regres1=ADF_1@res
Box.test(regres1, n/4, type = c("Ljung"))
plot(ADF_1)

#MannKendall(dif_1) #H0 Não há tendência

# Rejeitar H0 para z.lag.1 indica estacionaridade
(sem trend estocástico, Ro < 1)
# Não Rejeitar H0 para tt indica que a trend
determinística foi eliminada

#Periodograma
par(mfrow=c(1,1))
p=periodogram(dif_1,main="Periodograma",
ylab="Intensidade Espectral",xlab="Frequências de Fourier",
col="blue");abline(h=0)
f<-p$freq
ds<-p$spec
plot(f, ds, type="l", col="blue", ylab="Densidade Espectral",
xlab="Frequência", main="Periodograma") # Olhar a posição do maior,
s = posição/número de observações
periodicidade <-1/f[ds==max(ds)]
periodicidade

# Teste G de Fisher #H0 Não existe periodicidade
n=length(x)
alpha<-0.05
N<-n/2
(G<-max(ds)/sum(ds)) # sum(ds) de 1 ate n/2
(Z<-(1-((alpha/N)^(1/(N-1))))))
(p.valor=N*((1-G)^(N-1)))
# p-valor para o teste G (Morettin e Toloï 2006, pg. 458)

# Diferença Sazonal de lag s = 12 aplicada a série já com 1
diferença de lag 1

dif_s<-diff(dif_1, lag = 12)
par(mfrow=c(1,1))
plot.ts(dif_s, main="Série após a aplicação das
diferenças simples e sazonal", col="blue")
par(mfrow=c(2,1))
acf(dif_s,n/4, ylab="Autocorrelação",
main="Série após a aplicação das diferenças")
pacf(dif_s,n/4, ylab="Autocorrelação parcial",
main="Série após a aplicação das diferenças")

# Ajustando o Modelo SARIMA

```

```

#Alguns gráficos interessantes antes de ajustar um modelo
par(mfrow=c(2,2))
acf(x,n/4, main="FAC da série original")
pacf(x,n/4, main="FACP da série original")
acf(dif_s,n/4, main="FAC da série diferenciada",
ylab="Autocorrelação", xlab="Lag")
pacf(dif_s,n/4, main="FACP da série diferenciada",
ylab="Autocorrelação parcial", xlab="Lag")

M1<-arima(x, order = c(1,1,2),
seasonal = list(order = c(0,1,1), period = 12))
M1
AIC(M1)
AICc(M1)
BIC(M1)
coefstest(M1)
fit<-fitted(M1)
acur=accuracy(fit,ytmin) # Medida de acurácia do modelo
acur # RMSE fornece a raiz do erro quadrático médio

# Iniciando a Análise Residual

# Teste de Normalidade de Shapiro–Wilk (Normalidade)
Res<-rstandard(M1) # resíduos padronizados
a<-shapiro.test(Res)
a

#Teste de Box–Pierce e Ljung–Box (independencia)
Box.test(Res, 15,type = c("Ljung"))

# ACF e PACF dos resíduos estimados
par(mfrow=c(2,1))
acf(Res,n/4, ylab="Autocorrelação",
main="FAC da série dos resíduos do modelo ajustado")
pacf(Res,n/4, ylab="Autocorrelação parcial",
main="FACP da série dos resíduos do modelo ajustado")

# Homocedasticidade para os resíduos (Tomei a série dos Res^2)
par(mfrow = c(2,1))
acf(Res^2,n/4,ylab="Autocorrelação", main="FAC da série dos
resíduos elevados ao quadrado do modelo ajustado")
pacf(Res^2,n/4, ylab="Autocorrelação parcial",main="FACP da
série dos resíduos elevados ao quadrado do modelo ajustado")
Box.test(Res^2, 15, type = c("Ljung"))
# Não Rejeitar H0 indica variância Constante

#Gráficos interessantes
par(mfrow=c(2,2))
acf(Res,n/4, ylab="Autocorrelação", main="FAC dos resíduos")
pacf(Res,n/4, ylab="Autocorrelação parcial",
main="FACP dos resíduos")
acf(Res^2,n/4,ylab="Autocorrelação", main="FAC dos resíduos elevados
ao quadrado")
pacf(Res^2,n/4, ylab="Autocorrelação parcial",
main="FACP dos resíduos elevados ao quadrado")
par(mfrow=c(1,2))
hist(Res, ylab="Frequência", xlab="Resíduo",
main="Histograma dos resíduos")
qqnorm(Res, ylab="Quartis amostrais",

```

```

xlab="Quartis teóricos");qqline(Res, col=2)

#Modelando os resíduos
M1<-arima(x, order = c(1,1,2),seasonal = list(order = c(0, 1, 1),
period = 12))
M1
predi_M1 = forecast(M1,12)
res.M1=M1$res
Box.test(res.M1, 15, type = c("Ljung")) # Não rejeitar H0
indica que os resíduos não são autocorrelacionados
Box.test(res.M1^2, 15, type = c("Ljung")) # Não Rejeitar H0
indica variância Constante

#Modelo GARCH
Gar=garch(res.M1, order=c(0,2), trace=F)
AIC(Gar)
summary(Gar)
ResGar=Gar$res
Box.test(ResGar, 15, type = c("Ljung")) # Não rejeitar H0
indica que os resíduos não são autocorrelacionados
Box.test(ResGar^2, 15, type = c("Ljung")) # Não Rejeitar H0
indica variância Constante
ht.Gar=Gar$fit[,1]^2 # Usar primeira coluna de ajuste
plot(ht.Gar, main='Variância condicional')
predi_Gar=forecast(ht.Gar,12)
predi_Gar
predi_Gar$mean
e<-rnorm(12,0,1)
e
pred_res=sqrt(predi_Gar$mean)*e
pred_res

# Previsão SARIMA+ARCH
Previ_final= predi_M1$mean+pred_res
Previ_final

#Fim

```

APÊNDICE C: Rotina para análise da precipitação pluviométrica

```
#install.packages("forecast","timeseries","urca","stats","GeneCycle","lmtest","TSA","sme")
library("forecast")
library("tseries")
library("urca")
library("stats")
library("GeneCycle")
library("lmtest") # Obter os p-values
library("TSA")
library("sme")

#Carregando os dados
setwd("C:/Users/Paulo César/Desktop/Mauricio/Dados")
dados<-read.table('C:/Users/Paulo César/Desktop/Mauricio/Dados/
dadosatualizados.txt',h=T, sep = "")
dados
attach(dados)

#Calcular a soma mensal da precipitação pluviométrica de Viçosa

options(OutDec=",")
ano=c(31,28,31,30,31,30,31,31,30,31,30,31)
bis=c(31,29,31,30,31,30,31,31,30,31,30,31)
inibis=cbind(bis,ano,ano,ano)
diasmeses=cbind(rep(inibis,13))
#a=rnorm(sum(diasmeses),0,1)
a=dados$PP
mensa=NULL
ppmensal1=NULL
t=1
for(k in 1:length(diasmeses)){
ppmensal1<-rbind(ppmensal1,sum(a[t:(t+diasmeses[k]-1)]))
t=sum(diasmeses[1:k])+1
}
ppmensal1
ppmensal<-ppmensal1[1:583]

#Interpolar dados da precipitação

yo=ts(ppmensal[1:length(ppmensal)],start=c(1968,1), frequency=365)
ypp=ts(na.interp(yo, lambda = NULL),start=c(1968,1), frequency=365)
ypp # precipitação total mensal
summary(ypp)
var(ypp)

#Série e a decomposição de suas componentes

x<-ts(ypp, start = c(1968, 1), end = c(2016, 7), frequency = 12)
y=ts(x[1:583],start=c(1968,1), frequency=12)
ts.plot(y,main="Série da precipitação pluviométrica mensal
de Viçosa", col=c("blue"), lwd=1, gpars=list(xaxt="n",axes=F),
xlab="",ylab="Precipitação",ylim=c(min(y),max(y)))
axis(1, at=1968.1:2017.1,labels=1968.1:2017.1,
tc1= -.8,lwd.ticks=1.2,las=2, cex.axis=.7)
lim=seq(0,650,by=50)
axis(2, at=lim, tc1= -.8,lwd.ticks=1.2,las=2)
trend=lm(x-time(x))
summary(trend)
abline(trend, col="black")
legend("topleft", c("Série real","Tendência"), col=c(4,1),
```

```

lty=c(1,1))

m <- decompose(x)
par(mfrow=c(2,2))
plot(m$x, col="blue", xlab="Ano", ylab="", main="Série observada")
plot(m$trend, col="blue", xlab="Ano", ylab="",
main="Componente de tendência da série")
plot(m$seasonal, col="blue", xlab="Ano", ylab="",
main="Componente sazonal da série")
plot(m$random, col="blue", xlab="Ano", ylab="",
main="Fator aleatório da série")

#Fazendo uma análise mês a mês

mes=cycle(x)
par(mfrow=c(1,1))
#Janeiro (mes==1) Fevereiro (mes==2)... Dezembro (mes==12)
jan=subset(x,mes==12)
xjan=ts(jan, start = c(1968), end = c(2016), frequency = 1)
xjan
trendjan=lm(xjan~time(xjan))
summary(trendjan) # Estimativa de time(xjan)
me fornece a tendência anual para o mês de janeiro
(verificar qual é significativa pelo teste t)

ts.plot(xjan,main="Série da média do mês de janeiro da temperatura
máxima de Viçosa", col=c("blue"), lwd=1,
gpars=list(xaxt="n", axes=F), xlab="", axes=F,
ylab="Média mensal da temperatura máxima", ylim=c(min(xjan),
max(xjan)))
axis(1, at=1968.1:2016.12, labels=1968:2016, tcl= -.8,lwd.ticks=1.2,
las=2, cex.axis=.7)
lim=seq(50,650.5,by=25)
axis(2, at=lim, tcl= -.8,lwd.ticks=1.2, las=2)

# Boxplot da série
mes1=rep(seq(1:12),49)
mes=mes1[1:583]
boxplot(x~mes,main="Boxplot da série")

# Verificando ACF e PACF

n=length(x) #tamanho da amostra
par(mfrow=c(2,1))
acf(x,n/4, ylab="Autocorrelação", main="FAC da Série da Precipitação
Total Mensal de Viçosa-MG")
pacf(x,n/4, ylab="Autocorrelação parcial", main="FACP da Série da
Precipitação Total Mensal de Viçosa-MG")

# Teste Dickey-Fuller aumentado

ADF<-ur.df(x, type="trend", lags=22) #Foram necessárias 22
defasagens para que o resíduo não fosse correlacionados
summary(ADF) # Rejeitar H0 indica estacionaridade. Não rejeitar
H0 indica tendência estocástica para "z.lag.1" e tendência
determinística para "tt"
regres=ADF@res
Box.test(regres, n/4, type = c("Ljung")) #Para verificar se os
resíduos são ruído branco
plot(ADF)

```

```

# Rejeitar H0 para z.lag.1 indica estacionaridade (sem trend
estocástico , Ro < 1)
# Não Rejeitar H0 para tt indica que a trend determinística foi
eliminada

#Periodograma

par(mfrow=c(1,1))
t<-periodogram(x, method = "builtin")
# Para verificar a periodicidade tomei apenas a componente sazonal
da série
f<-t$freq
ds<-t$spec
plot(f, ds, type="h", col="blue", ylab="Intensidade Espectral",
xlab="Frequência", main="Periodograma")
# Olhar a posição do maior, s = posição/número de observações
#axis(1, 0:10/10, gsub(".", "", format(0:10/10), fixed=TRUE),
las=1)
periodicidade <- 1/f[ds==max(ds)]
periodicidade

# Teste G de Fisher

n=length(x)
alpha<-0.05
N<-n/2
(G<-max(ds)/sum(ds)) # sum(ds) de 1 ate n/2
(Z<-(1-((alpha/N)^(1/(N-1))))))
(pvalor=N*((1-G)^(N-1)))
# p-valor para o teste G (Moretton e Toloi 2006, pg. 458)

# Diferença Sazonal de lag s = 12 aplicada a série já com 1
diferença de lag 1

dif_s<-diff(x, lag = 12)
par(mfrow=c(1,1))
plot.ts(dif_s, main="Série após a aplicação das diferenças")
par(mfrow=c(2,1))
acf(dif_s, n/4, main="Série após a aplicação da diferença")
pacf(dif_s, n/4, main="Série após a aplicação da diferença")

# Ajustando o Modelo SARIMA

#Alguns gráficos interessantes antes de ajustar um modelo
par(mfrow=c(2,2))
acf(x, n/4, main="FAC da série original")
axis(2, -0.5:1.0, gsub(".", "", format(0:10/10), fixed=TRUE),
las=1)
pacf(x, n/4, main="FACP da série original")
acf(dif_s, n/4, main="FAC da série diferenciada",
ylab="Autocorrelação", xlab="Lag")
pacf(dif_s, n/4, main="FACP da série diferenciada",
ylab="Autocorrelação parcial", xlab="Lag")

Ml<-arima(x, order = c(1,0,0), seasonal = list(order = c(0, 1, 1),
period = 12))
Ml
AIC(Ml)
AICc(Ml)

```

```

BIC(M1)
coefstest(M1)
fit<-fitted(M1)
acur=accuracy(fit, ypp) # Medida de acurácia do modelo
acur # RMSE fornece a raiz do erro quadrático médio

# Iniciando a Análise Residual

# Teste de Normalidade de Shapiro–Wilk (Normalidade)
Res<-M1$residuals
a<-shapiro.test(Res)
a

# Teste Ljung–Box com k = n/4 (Independência)
Box.test(Res, n/4, type = c("Ljung"))

# ACF e PACF dos resíduos estimados

par(mfrow=c(2,1))
acf(Res, n/4)
pacf(Res, n/4)

# Homocedasticidade para os resíduos (Tomei a série dos Res^2)

par(mfrow = c(2,1))
acf(Res^2, n/4)
pacf(Res^2, n/4)
Box.test(Res^2, 15, type = c("Ljung"))
# Não Rejeitar H0 indica variância Constante

# Gráficos interessantes
par(mfrow=c(2,2))
acf(Res, n/4, ylab="Autocorrelação", main="FAC dos resíduos")
pacf(Res, n/4, ylab="Autocorrelação parcial",
main="FACP dos resíduos")
acf(Res^2, n/4, ylab="Autocorrelação",
main="FAC dos resíduos elevados ao quadrado")
pacf(Res^2, n/4, ylab="Autocorrelação parcial",
main="FACP dos resíduos elevados ao quadrado")
par(mfrow=c(1,2))
hist(Res, ylab="Frequência", xlab="Resíduo",
main="Histograma dos resíduos")
qqnorm(Res, ylab="Quartis amostrais",
xlab="Quartis teóricos"); qqline(Res, col=2)

# Modelando a variância condicional

M1<-arima(x, order = c(1,0,0), seasonal = list(order = c(0, 1, 1),
period = 12))
M1
predi_M1 = forecast(M1, 12)
res.M1=M1$res
res_quadrado=res.M1^2

# GARCH/ARCH
arch2=garch(res.M1, order=c(0,2), trace=F)
AIC(arch2)
summary(arch2)
ResGar=arch2$res

```

```

Box.test(ResGar, 15, type = c("Ljung")) # Não rejeitar H0 indica que
os residuos não são autocorrelacionados
predi_M1= forecast(M1,12,level=95)
ht.arch2=arch2$fit[,1]^2 # Usar primeira coluna de ajuste
plot(ht.arch2,main='Variância condicional')
predi_arch2=forecast(ht.arch2,12)
predi_arch2
predi_arch2$mean
e<-rnorm(12,0,1)
e
pred_res=sqrt(predi_arch2$mean)*e
pred_res

# Previsão SARIMA+ARCH
Previ_final= predi_M1$mean+abs(pred_res)
Previ_final

#Fim

```