

Modelos de séries temporais para a previsão da temperatura média mensal de Lavras, MG

Renata A. Cintra^{1†}, Marcel I. P. Melo², Júlio S. S. Bueno Filho³

¹Universidade Federal de Lavras.

²Universidade Federal de Lavras. E-mail: mipmelo90@gmail.com.

³Universidade Federal de Lavras. E-mail: jssbueno@des.ufla.br.

Resumo: *Uma série temporal é caracterizada como qualquer conjunto de observações tomadas sequencialmente no tempo, cuja natureza principal é a presença de dependência nessas observações. A natureza dessa dependência é considerada um problema prático de interesse e, comumente, o objetivo de estudo da série temporal, é classificado em 4 tipos: investigar o mecanismo gerador da série, realizar previsões para valores futuros, descrever o comportamento da série e procurar periodicidade relevantes nos dados. Frequentemente, existe o interesse em modelar dados climatológicos, como por exemplo, os produtos produzidos no ramo agropecuário são influenciados pela temperatura do ar, uma vez que esta variável pode causar efeitos fisiológicos positivos ou negativos em plantas e/ou animais. Uma forma de análise para tais dados climatológicos pode ser realizada via modelos de séries temporais, visto que tais variáveis apresentam sazonalidade, correlação serial no tempo e tendem a não ser estacionárias, reunindo características ideais para o uso dessa ferramenta estatística. Portanto, este trabalho tem por objetivo estudar o comportamento futuro da série mensal da temperatura média compensada da cidade de Lavras, bem como, detectar padrões de periodicidade no comportamento da série. A seleção do modelo foi baseada naquele que apresentou menor erro quadrático médio de previsão e erro absoluto percentual médio. Destacam-se os modelos SARIMA(4,0,0)(0,1,1)₁₂ e suavização exponencial sazonal de Holt-Winters que apresentaram o menor erro quadrático médio de previsão e erro absoluto percentual médio, respectivamente. Conclui-se que ambos modelos, foram capazes de captar a periodicidade sazonal anual presente nas observações, como também, fornecem boas previsões para a série de temperatura média mensal de Lavras, sugerindo que esses modelos são adequados para a realização deste tipo de análise.*

Palavras-chave: modelo SARIMA, modelo de suavização exponencial sazonal de Holt-Winters, previsão.

Abstract: *A time series is characterized as any set of observations taken sequentially in time, whose main nature is the presence of dependence in these observations. The nature of this dependence is considered a practical problem of interest and, commonly, the objective of the study of the time series, is classified into 4 types: investigate the generating mechanism of the series, make predictions for future values, describe the behavior of the series and look for periodicity relevant in the data. There is interest in modeling climatological data, for example, products produced in the agricultural sector are influenced by air temperature, since this variable can cause positive or negative physiological effects on plants and / or animals. One way of analyzing such climatological data can be performed via time series models, since these variables present seasonality, serial correlation in time and tend to be non-stationary, bringing together ideal characteristics for the use of this statistical tool. The objective of this work is to study the future behavior of the monthly series of the compensated average temperature of the city of Lavras. We highlight the SARIMA(4,0,0)(0,1,1)₁₂ and seasonal exponential smoothing models of Holt-Winters that presented the lowest mean squared error of prediction and absolute percentage error, respectively. It is concluded that both models were able to capture the annual seasonal periodicity present in the treatments, as well as provide good predictions for the Lavras monthly average temperature series, suggesting that these models are adequate to perform this type of analysis.*

Keywords: SARIMA model; Holt-Winters exponential smoothing model; Prediction.

[†]Autora correspondente: renatapcintra@gmail.com.

Introdução

De acordo com Box e Jenkins (1970) uma série temporal é qualquer conjunto de observações tomadas sequencialmente no tempo, tendo como principal característica dependência nas observações. A natureza dessa dependência é considerada um problema prático de interesse (BOX, REINSEL, LJUNG e JENKINS, 2015). Morettin e Tolo (2006) descrevem tais interesses em 4 tipos: investigar o mecanismo gerador da série, realizar previsões para valores futuros da série, descrever o comportamento da série e procurar periodicidades relevantes nos dados.

Um modelo de séries temporais consiste em decompor as observações Y_t como a soma de três componentes, se tratando de um modelo aditivo, como:

$$Y_t = T_t + S_t + e_t \quad (1)$$

em que T_t e S_t representam, respectivamente, a tendência e sazonalidade, enquanto e_t é uma componente aleatória, de média zero e variância constante.

Constantemente, tem-se interesse em modelar dados climatológicos, pois, de acordo com Medeiros et al. (2005) o ramo agrícola é influenciado pela temperatura do ar, uma vez que podem causar efeitos fisiológicos positivos ou negativos em plantas e/ou animais. Neste sentido, Chechi e Bayer (2012) afirmam que muitos empreendimentos agropecuários podem ir a falência devido a não considerarem o efeito da temperatura.

Adicionalmente, diversos trabalhos em diferentes regiões do país tiveram o interesse no estudo dessa variável climatológica. No Rio Grande do Sul, Cargnelutti Filho et al. (2008) estudaram o comportamento das séries de temperatura média decendial e máxima média decendial do ar para 41 municípios. Utilizando uma abordagem via modelos de regressão linear de efeitos fixos, Bardin et al. (2010) estimaram os valores de temperatura mensais e anuais na região de municípios que compõem o Polo Turístico do Circuito das Frutas do estado de São Paulo. Enquanto que, Silva et al. (2008) descreveram o comportamento da série de temperatura média mensal em Uberlândia, MG, através de modelos da classe ARIMA. Portanto, o ajuste de modelos para a temperatura é de interesse público e devem ser estudados.

Chechi e Bayer (2012) complementam que a modelagem utilizando séries temporais para variáveis climáticas é útil, uma vez que tais variáveis, geralmente, se comportam no sentido de apresentarem sazonalidade, correlação serial no tempo e tendem não serem estacionárias. Essas características são ideias para o uso dessa ferramenta estatística. Desse modo, este trabalho tem por objetivo estudar o comportamento futuro da série mensal da temperatura média compensada da cidade de Lavras/MG, bem como, detectar padrões de periodicidade no comportamento da série.

Material e Métodos

Os dados são referentes à Temperatura Compensada Média mensal da cidade de Lavras/MG no período de janeiro de 1998 a dezembro de 2018 e foram obtidos junto ao Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) do Instituto Nacional de Meteorologia (INMET). Uma vez que realizar previsão para a temperatura da cidade de Lavras é o objetivo principal deste trabalho, utilizaremos para o ajuste dos modelos das séries, a classe de modelos ARIMA e o modelo de suavização exponencial sazonal de Holt-Winters. Assim sendo, a escolha do modelo que melhor se ajusta aos dados da série será definido por aquele que apresentar menor erro quadrático médio de previsão (EQMP) e erro absoluto percentual médio (MAPE). Para que isso se faça possível, os dados referentes aos meses do ano de 2018 foram excluídos do ajuste dos modelos e utilizados para o cálculo do EQMP e MAPE.

Inicialmente, sugere-se que em uma análise de série temporal, plotar o gráfico *amplitude versus média*, com a finalidade de verificar se o modelo possui componentes aditivas e/ou multiplicativas e se apresenta ou não variância constante. Em seguida, com auxílio visual dos gráficos da série temporal, correlograma e periodograma, pode-se inferir se a série possui tendência, sazonalidade, observações atípicas e etc. Testes de Hipóteses sempre devem ser aplicados para sustentarem as inferências realizadas. Neste trabalho, foi utilizado o teste da raiz unitária de Dickey-Fuller aumentado e o teste de Friedman para verificar a presença de tendência e sazonalidade na série, respectivamente. Ambos testes estão descritos em Morettin e Tolo (2006).

Obtida uma série livre das componentes de tendência e sazonalidade, isto é, a série é um processo estacionário, ajusta-se um modelo à série. Morettin e Tolo (2006) enfatizam que a abordagem de Box e Jenkins (1970) é uma metodologia bastante utilizada na análises de modelos paramétricos de séries temporais. A metodologia consiste em ajustar modelos auto-regressivos integrados de médias móveis - $ARIMA(p, d, q)$ a um conjunto de dados. Em que, p é a ordem do polinômio auto-regressivo, d é o número de diferenças tomadas na série até que a mesma fique livre de tendência, enquanto que q é a ordem do polinômio de médias móveis.

Quando há na série a presença de sazonalidade estocástica, necessita-se ajustar um modelo ARIMA sazonal, ou simplesmente, $SARIMA(p, d, q)(P, D, Q)_s$, de modo que P , D e Q possuem as mesmas características de interpretação presentes em p , d e q , no entanto, estão em um contexto sazonal.

Segundo, Box e Jenkins (1970), um modelo $SARIMA(p, d, q)(P, D, Q)_s$, de período s , pode ser representado por:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)Y_t = \theta(B)\Theta(B^s)e_t \quad (2)$$

em que:

- $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é o operador autorregressivo de ordem p ;
- $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ é o operador de médias móveis de ordem q ;
- $(1-B)^d$ é o operador diferença, sendo d o número de diferenças simples aplicadas à série;
- $\Phi(B^s) = 1 - \Phi_1 B - \dots - \phi_P B^P$ é o operador autorregressivo sazonal de ordem P ;
- $\Theta(B^s) = 1 - \Theta_1 B - \dots - \Theta_Q B^Q$ é o operador de médias móveis sazonal de ordem Q ;
- $(1-B)^D$ é o operador diferença sazonal, sendo D o número de diferenças sazonais aplicadas à série;

Nicolau (2012) propõe três etapas para seguir a modelagem de Box e Jenkins (1970). Identificação, Estimação e Diagnóstico. Por identificação, entende-se que ao estacionarizar a série, identifica-se a ordem de q e p via função de autocorrelação (FAC) e função de autocorrelação parcial (FACP), respectivamente. Em estimação, uma vez obtida a ordem dos parâmetros desconhecidos, é necessário estimá-los. Neste trabalho, o método de estimação utilizado foi o de máxima verossimilhança. Em avaliação do diagnóstico, o autor ressalta que deve-se verificar os seguintes aspectos: significância estatística dos parâmetros; invertibilidade e estacionariedade; análise da matriz de correlação entre os estimadores; redundância entre as estimativas dos parâmetros e resíduo seguindo ruído branco. O fluxograma apresentado na Figura 1 ilustra o procedimento.

Note que após a realização dos procedimentos de identificação, estimação e avaliação de diagnóstico, caso o modelo não seja satisfatório, deve-se voltar ao processo inicial da metodologia de Box e Jenkins. Por outro lado, se o modelo for satisfatório, significa que o modelo pode ser utilizado para fazer inferência sobre a série, no entanto, este modelo só deve ser usado se for o melhor modelo satisfatório obtido, sob algum critério de seleção.

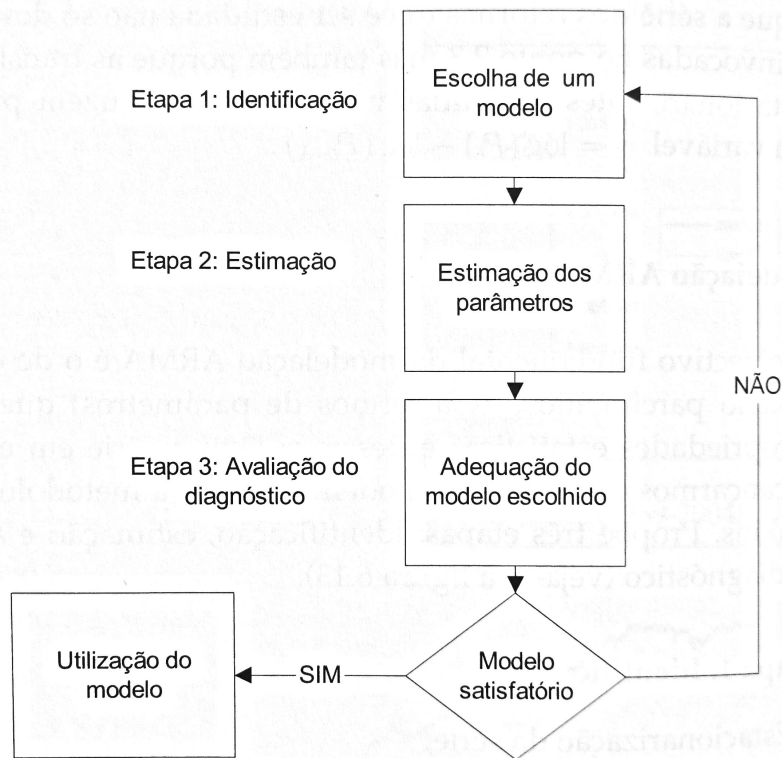


Figura 1: Fluxograma da metodologia de Box e Jenkins.
Fonte: Nicolau (2012).

Nicolau (2012) ressalta que se existir mais de um modelo que cumpra as condições citadas acima, é necessário selecionar o melhor modelo à luz de determinado critério. Dessa forma, como o intuito é realizar previsão para a série da temperatura da cidade de Lavras/MG para os anos de 1998 a 2018, o critério de seleção do melhor modelo será o que apresentar o menor EQMP e MAPE.

Uma outra grande classe de métodos de previsão são os modelos de suavização exponencial. O intuito de tais modelos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e utilizá-lo para fazer previsões futuras na série (MORETTIN; TOLOI, 2006). O modelo de suavização exponencial sazonal de Holt-Winters é utilizado para fazer previsão de valores futuros para séries temporais que apresentam padrão de comportamento mais complexo. O procedimento se baseia em três equações com distintas constantes de suavização que estão associadas com as componentes nível, tendência e sazonalidade de uma série temporal. As equações para o modelo multiplicativo são:

$$Y_t = N_t S_t + T_t + e_t, \quad t = 1, \dots, N.$$

em que as três equações de suavização são dadas por:

$$\begin{aligned} \hat{S}_t &= \gamma \left(\frac{Y_t}{\hat{N}_t} \right) + (1 - \gamma) \hat{S}_{t-s}, \quad 0 < \gamma < 1, t = s + 1, \dots, N, \\ \hat{N}_t &= \alpha \left(\frac{Y_t}{\hat{S}_{t-s}} \right) + (1 - \alpha) (\hat{Y}_{t-1} + \hat{T}_{t-1}), \quad 0 < \alpha < 1, t = s + 1, \dots, N, \\ \hat{T}_t &= \beta (\hat{N}_t - \hat{N}_{t-1}) + (1 - \beta) \hat{T}_{t-1}, \quad 0 < \beta < 1, t = s + 1, \dots, N, \end{aligned}$$

e representam as estimativas da componente sazonalidade, nível e tendência, respectivamente. Enquanto que as constantes de suavização são α , β e γ e são determinadas de modo que a soma de quadrados dos erros se torne mínima. Morettin e Toloi (2006) destacam que embora o modelo seja de fácil entendimento e com implementação de baixa complexidade, medidas estatísticas, tais como a média, variância de previsão e, conseqüentemente, intervalos de confiança não são obtidas.

Para verificar se o modelo proposto era apropriado, foi aplicado o Teste de Ljung e Box (1978) que testa a hipótese de que resíduos seguem ruído branco. Em caso afirmativo, realizou-se a previsão para os meses do ano de 2018 e comparou-se com as observações reais, através do cálculo de EQMP e MAPE. A análise da série foi toda construída utilizando o programa R (R CORE TEAM, 2019), mais precisamente com os auxílios dos pacotes `stats` (R CORE TEAM, 2019) e `tseries` (TRAPLETTI; HORNIK, 2018) para o ajuste e diagnóstico dos modelos e `tidyverse` (WICKHAM, 2017) para a construção dos gráficos.

Resultados e Discussão

Na Figura 2, apresentam-se os gráficos da série temporal da temperatura compensada média de Lavras, entre o período de 1998 a 2018 e o de amplitude versus média com ajuste de uma reta de regressão linear.

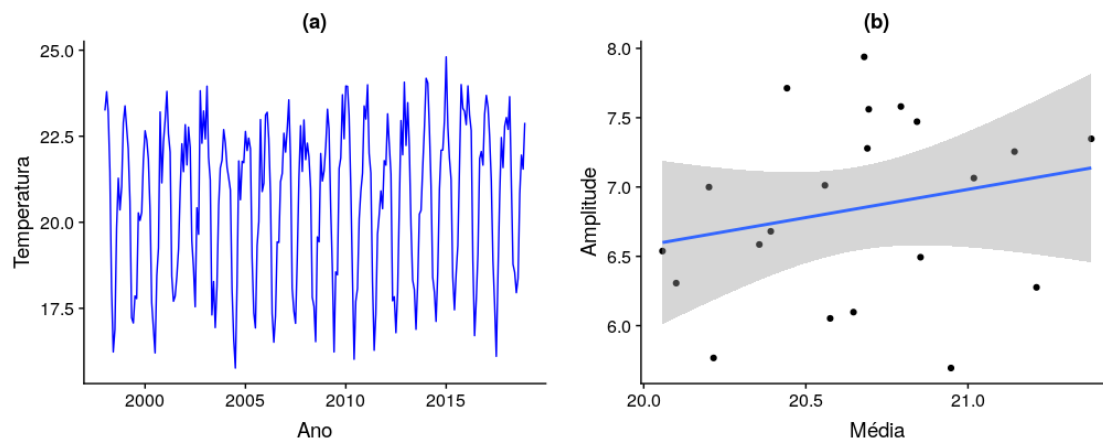


Figura 2: (a) Temperatura compensada média de Lavras de 01/1998 a 12/2018. (b) Reta de regressão linear ajustada ao gráfico de Amplitude *versus* Média

Pode-se observar através da Figura 2b que não há a necessidade de se aplicar uma transformação na escala dos dados, uma vez que os pontos não apresentam nenhum comportamento linear. O coeficiente de inclinação estimado da reta foi 0,4067 e via teste t , o valor-p foi de 0,33. Desse modo, com 5% de significância, não há razões para rejeitar a hipótese de que o coeficiente de inclinação da reta é igual a zero e, conseqüentemente, não há razão para aplicar uma transformação na escala dos dados.

Não há indícios de que a série possui tendência (Fig 2a), no entanto, percebe-se que a mesma possui sazonalidade anual, uma vez que a série apresenta comportamentos semelhantes a cada período de um ano. Essa visualização fica ainda mais evidente quando observamos o correlograma da série (Figura 3).

Observa-se que pela função de autocorrelação a série apresenta períodos sazonais de um ano. Tal ocorrência sugere que se realize uma diferença sazonal. Para uma avaliação formal da sazonalidade, foi aplicado o Teste de Friedman, cuja hipótese H_0 é de que a série temporal não

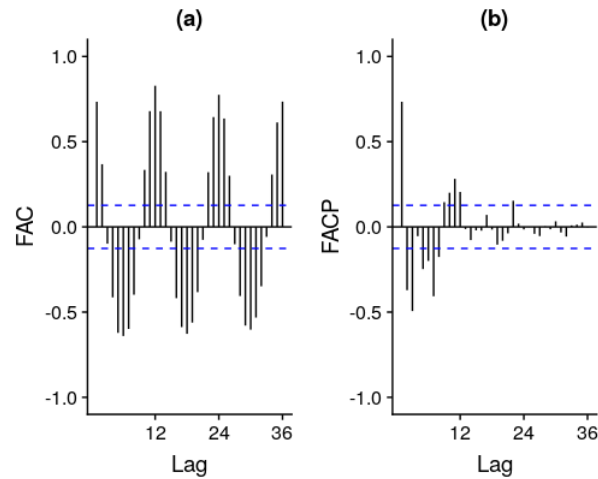


Figura 3: Gráfico das Funções de autocorrelação (a) e autocorrelação parcial (b) da Temperatura compensada média de Lavras para o período de 01/1998 a 12/2018.

possui sazonalidade. O valor da estatística do teste foi $T^2 = 195,12 > \chi_{11,0.95}^2 = 19,68$. Portanto, rejeita-se a hipótese H_0 de que não há sazonalidade na série de temperatura compensada média de Lavras 1998-2018. Uma maneira de tratar esse problema, é retirando a sazonalidade por meio de diferença sazonal. A Figura 4 apresenta o gráfico da série com uma diferença sazonal.

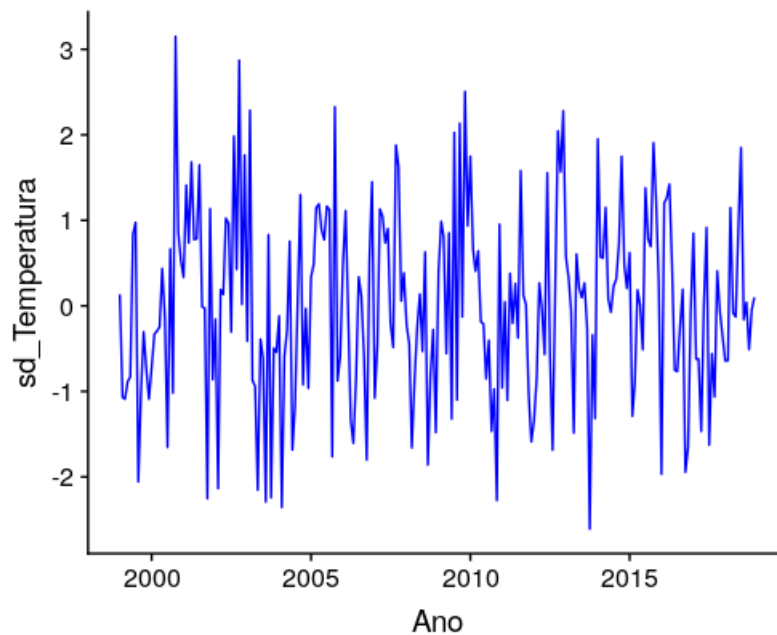


Figura 4: Temperatura compensada média de Lavras 1998-2018 com uma diferença sazonal

Espera-se que após realizar uma diferença sazonal, a série de temperatura compensada média de Lavras 1998-2018 fique estacionária. Para tanto, realizou-se o teste de raiz unitária de Dickey-Fuller Aumentado, cuja hipótese alternativa é de estacionariedade. A estatística do teste foi de -4.96 para $k = 6$, o que forneceu um valor-p menor que 0,01. Sendo assim, o resultado do teste indica que a série com uma diferença sazonal é estacionária. O correlograma da série,

apresentado na Figura 5, corrobora com o resultado do teste de Dickey-Fuller Aumentado, no que diz respeito à estacionariedade da série.

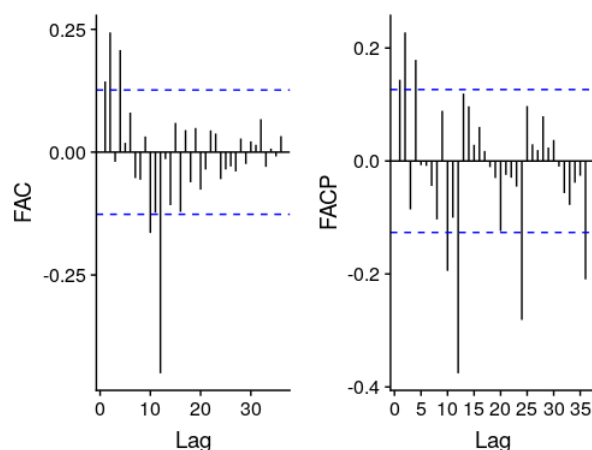


Figura 5: Correlograma da temperatura compensada média de Lavras com uma diferença sazonal

Pode-se observar que o correlogramana apresentado na Figura 5 é de uma série estacionária, no entanto, percebe-se, através da FAC, que os erros são correlacionados. Desse modo, como há a presença de autocorrelações sazonais e por ter sido realizada uma diferença sazonal na série, os modelos da classe SARIMA e o modelo de suavização exponencial Holt-Winters são os mais indicados para o ajuste do erro. Foi possível a identificação de alguns modelos que produziram ruído branco. Tais modelos estão descritos na Tabela 1

Tabela 1: Modelos de séries temporais ajustados para a temperatura compensada média de Lavras-MG no período de 1998 a 2018.

Modelo	Coefficientes	Desvio padrão	EQMP	MAPE
m_1 : SARIMA(4,0,0)(0,1,1) ₁₂	ϕ_1 : 0.156	0.064	0.393	0.024
	ϕ_2 : 0.158	0.065		
	ϕ_4 : 0.194	0.065		
	Θ_1 : -0.999	0.093		
m_2 : SARIMA(2,0,2)(2,1,0) ₁₂	ϕ_2 : 0.677	0.114	0.567	0.031
	θ_2 : -0.450	0.130		
	Φ_1 : -0.647	0.064		
	Φ_2 : -0.386	0.064		
m_3 : SARIMA(2,0,2)(0,1,1) ₁₂	ϕ_2 : 0.721	0.113	0.402	0.025
	θ_2 : -0.506	0.134		
	Φ_1 : -0.999	0.108		
m_4 : Holt-Winters	α : 0.153	—	0.394	0.022
	β : 0.017	—		
	γ : -1.000	—		

Observa-se na Tabela 1, que dentre os modelos ajustados, o modelo que fornece menor EQMP é dado pelo modelo m_1 , enquanto que o modelo que fornece o menor MAPE é dado pelo modelo m_4 . O correlograma do resíduo do modelo m_1 é apresentado a seguir.

Perceba que os resíduos do modelo m_1 são distribuídos como ruído branco, uma vez que a função de autocorrelação é estatisticamente igual a zero. O Teste de Ljung-Box indica a não

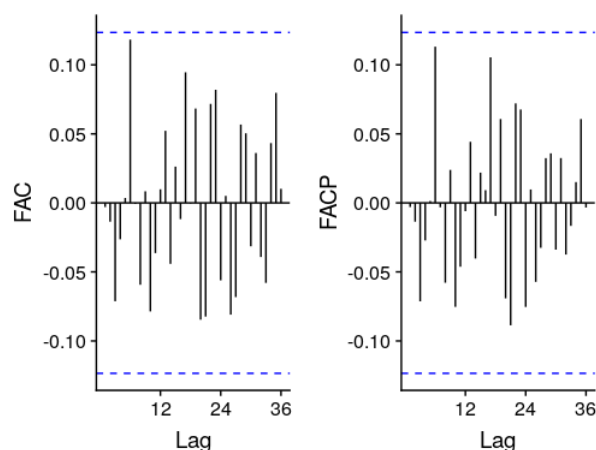


Figura 6: Correlograma do resíduo do modelo m_1

existência de autocorrelação, uma vez que o teste forneceu um valor-p de $P(\chi_{24}^2 > 21,28) = 0,6224$, corroborando a interpretação do correlograma.

Na Tabela 2 é apresentada a previsão para o ano de 2018 da temperatura, com intuito de medir a qualidade preditiva dos modelos m_1 e m_4 .

Tabela 2: Valores reais e valores preditos para a temperatura compensada média de Lavras-MG no ano de 2018.

Mês	Observado	Holt-Winters	SARIMA
Janeiro	23.04	23.12	22.98
Fevereiro	22.70	23.26	23.28
Março	23.65	22.38	22.45
Abril	21.12	21.22	21.33
Maio	18.78	18.37	18.26
Junho	18.55	17.22	17.28
Julho	17.95	17.13	17.20
Agosto	18.37	18.74	18.85
Setembro	20.83	20.85	20.52
Outubro	21.95	22.05	21.89
Novembro	21.55	21.76	21.69
Dezembro	22.89	22.76	22.56

Pela Tabela 2, nota-se que os valores preditos de ambos modelos estão próximos dos valores reais, explicando os pequenos valores de EQMP e MAPE descritos na Tabela 1, sugerindo que a utilização de modelos de séries temporais como SARIMA e o de Suavização Exponencial Sazonal de Holt-Winters é adequada para o ajuste de dados de temperatura compensada média da cidade de Lavras/MG. A Figura 7 refere-se ao ajuste e previsão dos modelos.

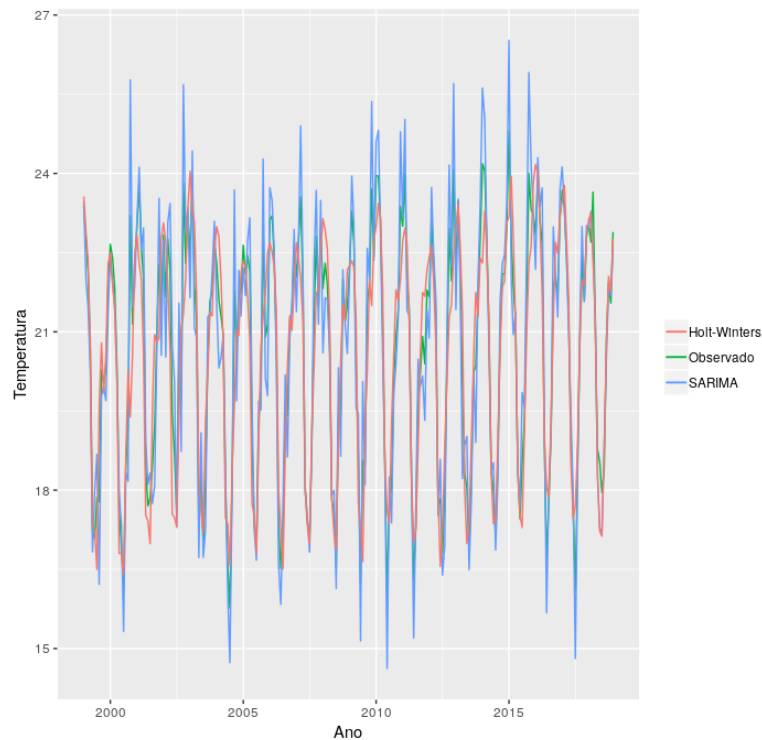


Figura 7: Valores observados, estimados e preditos para temperatura compensada média de Lavras no período de 1998 a 2018

Conclusões

A série mensal de temperatura média compensada média de Lavras no período de janeiro de 1998 a dezembro de 2018 apresentou um comportamento sazonal bem definido com periodicidade anual. Os modelos de séries temporais utilizados neste trabalho se ajustaram bem a série, resultando em boas previsões. Destaques para os modelos SARIMA m_1 e suavização exponencial sazonal Holt-Winters m_4 que obtiveram o menor EQMP e MAPE, respectivamente. Portanto, tais modelos são úteis para descrever e realizar previsões para a temperatura compensada média de Lavras.

Agradecimentos

Agradecemos à CAPES e à FAPEMIG, pelo apoio financeiro.

Referencias Bibliográficas

BARDIN, L.; PEDRO JÚNIOR, M. J.; MORAES, J. F. L. Estimativas das temperaturas máximas e mínimas do ar para a região do Circuito das Frutas, SP. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.14, n.6, p.618-624,2010.

BOX, G. E. P.; JENKINS, G. M. *Time series analysis: forecasting and control*. San Francisco: Holden-Day, 1970.

Sigmae, Alfenas, v.8, n,2, p. 596-605, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

- BOX, G. E. P. et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- CARGNELUTTI FILHO, A.; MALUF, J. R. T.; MATZENAUER, R. Coordenadas geográficas na estimativa das temperaturas máxima e média decendiais do ar no Estado do Rio Grande do Sul. *Ciência Rural*. Santa Maria. v.38, n.9, p.2448-2456, 2008.
- CHECHI, L.; BAYER, F. M. Modelos univariados de séries temporais para a previsão das temperaturas médias mensais de Erechim, RS. *Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi*, v.16, n.12, 2012.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, v.65, n.2, p.297-303, 1978.
- MEDEIROS, S. de S. et al. Estimativa e espacialização das temperaturas do ar mínimas, médias e máximas na Região Nordeste do Brasil. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.9, n.2, p.247-255, 2005.
- MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. 2ª edição ver. ampl. - São Paulo: Blucher, 2006.
- NICOLAU, J. *Modelação de séries temporais financeiras*. Coimbra: Edições Almedina, SA, 2012.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2019. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- SILVA, M. I.; 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*, v.12, n.5, p.480-485, 2008.
- TRAPLETTI, A.; HORNIK, K. *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-46, 2018. URL <https://CRAN.R-project.org/package=tseries>
- WICKHAM, H. *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1., 2017. URL <https://CRAN.R-project.org/package=tidyverse>