

## Ajuste de um modelo de séries temporais para prever a precipitação pluviométrica

Adjustment of a time series model to predict rainfall

Ajuste de un modelo de serie de tiempo para predecir la lluvia

Recebido: 30/04/2021 | Revisado: 07/05/2021 | Aceito: 18/05/2021 | Publicado: 04/06/2021

**Pedro Henrique Alves Bittencourt Santos**

ORCID: <https://orcid.org/0000-0001-7485-324X>

Centro Federal de Educação Tecnológica de Minas Gerais, Brasil

E-mail: [pedro.henrique2512@hotmail.com](mailto:pedro.henrique2512@hotmail.com)

**Otávio Augusto dos Santos Delfino**

ORCID: <https://orcid.org/0000-0002-6006-0510>

Centro Federal de Educação Tecnológica de Minas Gerais, Brasil

E-mail: [otavioaugusto.2503@gmail.com](mailto:otavioaugusto.2503@gmail.com)

**Ricardo Vitor Ribeiro dos Santos**

ORCID: <https://orcid.org/0000-0002-8623-1509>

Centro Federal de Educação Tecnológica de Minas Gerais, Brasil

E-mail: [ricardoribeiro@cefetmg.br](mailto:ricardoribeiro@cefetmg.br)

**Mateus do Nascimento**

ORCID: <https://orcid.org/0000-0003-1152-8882>

Centro Federal de Educação Tecnológica de Minas Gerais, Brasil

E-mail: [mateusnascimento@cefetmg.br](mailto:mateusnascimento@cefetmg.br)

### Resumo

A precipitação pluviométrica é uma das variáveis meteorológicas mais relevantes para estudos climáticos. Conhecer sua variabilidade espacial e temporal permite planejar diversas atividades humanas, tanto do ponto de vista econômico quanto social. Tal importância deve-se às consequências que ela pode ocasionar, em excesso ou em falta, causando enchentes, inundações, secas, queda no fornecimento de energia, baixa produção de alimentos, entre outros. Este trabalho teve como objetivo estudar a série histórica de precipitação pluviométrica média mensal da cidade de Lavras/MG a fim de obter um modelo estatístico que permita realizar previsões, para isso utilizou-se 228 observações correspondente ao período de janeiro de 2000 a dezembro de 2018. Para esta análise, foi verificada a existência das componentes tendência e sazonalidade. A metodologia de Box e Jenkins foi utilizada na modelagem dos dados. Foram ajustados alguns modelos utilizando a classe SARIMA, pois a série em estudo apresentou sazonalidade estocástica. A comparação entre os modelos considerados adequados à série foi realizada através do EQM e AIC. O modelo SARIMA (0,0,0) x (0,1,1)<sub>12</sub> foi utilizado para fazer previsões de observações futuras. A série de precipitação pluviométrica média mensal da cidade de Lavras/MG apresentou uma componente sazonal com periodicidade de 12 meses. O modelo ajustado obteve um resultado muito bom, pois os intervalos de confiança a 95% contiveram os doze valores reais de precipitação pluviométrica média mensal da cidade de Lavras/MG para o ano de 2019, mesmo diante dos imprevistos e incertezas associadas a fatores climáticos. O modelo em questão pode ser utilizado na tomada de decisões para a realização de planejamentos estratégicos futuros que envolvem questões públicas associadas à cidade de Lavras. Estas previsões também podem ser utilizadas para auxiliar os gestores da usina hidrelétrica do Funil/MG a programar operações e manutenções futuras da vazão de água de forma adequada, pois esta se encontra próxima ao município de Lavras.

**Palavras-chave:** Meteorológicas; Previsões; Sazonalidade; Modelo SARIMA.

### Abstract

Precipitation is one of the most relevant meteorological variables for climate studies. Knowing its spatial and temporal variability allows planning various human activities, both from an economic and social point of view. Such importance is due to the consequences that it can cause, in excess or in lack, causing floods, floods, droughts, decrease in energy supply, low food production, among others. This study aimed to study the historical series of average monthly rainfall in the city of Lavras/MG in order to obtain a statistical model that allows predictions to be made. For this purpose, 228 observations were used corresponding to the period from January 2000 to December 2018 for this analysis, the existence of the trend and seasonality components was verified. The Box and Jenkins methodology was used to model the data. Some models were adjusted using the SARIMA class, as the series under study showed stochastic seasonality. The comparison between the models considered suitable for the series was performed using the NDE and AIC. The SARIMA (0,0,0) x (0,1,1)<sub>12</sub> model was used to make predictions of future observations. The series of monthly average rainfall in the city of Lavras/MG presented a seasonal component with a periodicity of 12 months. The adjusted model obtained a very good result, since the 95% confidence intervals contained the twelve real values of average monthly rainfall in the city of Lavras/MG for the year 2019, even in the face of unforeseen and

uncertainties associated with climatic factors. The model in question can be used in decision making to carry out future strategic plans that involve public issues associated with the city of Lavras. These forecasts can also be used to assist the managers of the Funil/MG hydroelectric plant to schedule future water flow operations and maintenance properly, as it is close to the municipality of Lavras.

**Keywords:** Meteorological; Forecasts; Seasonality; Model SARIMA.

### Resumen

La precipitación es una de las variables meteorológicas más relevantes para el estudio del clima. Conocer su variabilidad espacial y temporal permite planificar diversas actividades humanas, tanto desde el punto de vista económico como social. Dicha importancia se debe a las consecuencias que puede ocasionar, por exceso o por defecto, provocando inundaciones, riadas, sequías, disminución del suministro energético, baja producción de alimentos, entre otras. Este estudio tuvo como objetivo estudiar la serie histórica de la precipitación media mensual en la ciudad de Lavras/MG con el fin de obtener un modelo estadístico que permita realizar predicciones. Para ello, se utilizaron 228 observaciones correspondientes al período comprendido entre enero de 2000 y diciembre de 2018. Para este análisis, se verificó la existencia de los componentes de tendencia y estacionalidad. Se utilizó la metodología de Box y Jenkins para modelar los datos. Algunos modelos se ajustaron mediante la clase SARIMA, ya que la serie en estudio presentaba estacionalidad estocástica. La comparación entre los modelos considerados adecuados para la serie se realizó mediante el NDE y el AIC. Se utilizó el modelo SARIMA  $(0,0,0) \times (0,1,1)_{12}$  para realizar las predicciones de las observaciones futuras. La serie de precipitaciones medias mensuales en la ciudad de Lavras/MG presentó una componente estacional con una periodicidad de 12 meses. El modelo ajustado obtuvo un muy buen resultado, ya que los intervalos de confianza del 95% contenían los doce valores reales de la precipitación media mensual en la ciudad de Lavras/MG para el año 2019, incluso ante imprevistos e incertidumbres asociadas a factores climáticos. El modelo en cuestión puede ser utilizado en la toma de decisiones para llevar a cabo futuros planes estratégicos que implican cuestiones públicas asociadas a la ciudad de Lavras. Estas previsiones también pueden utilizarse para ayudar a los gestores de la central hidroeléctrica de Funil/MG a programar adecuadamente las futuras operaciones de flujo de agua y su mantenimiento, ya que está cerca del municipio de Lavras.

**Palabras clave:** Meteorológico; Previsión; Estacionalidad; Modelo SARIMA.

## 1. Introdução

A precipitação pluviométrica é uma das variáveis meteorológicas mais relevantes para estudos climáticos. Conhecer sua variabilidade espacial e temporal permite planejar diversas atividades humanas, tanto do ponto de vista econômico quanto social. Tal importância deve-se às consequências que ela pode ocasionar, em excesso ou em falta, causando enchentes, inundações, secas, queda no fornecimento de energia, baixa produção de alimentos, entre outros. Precipitação é toda água oriunda da atmosfera que atinge a superfície terrestre, nas formas de chuva, geada, neve, granizo, neblina e orvalho (Santos e Aquino, 2017).

Ao longo dos anos, observações de vários fenômenos que ocorrem na natureza, dentre eles, a precipitação, produziram séries históricas que servem de base para a elaboração de modelos. Estes têm sido utilizados por pesquisadores para estudar a variabilidade das precipitações pluviométricas, vislumbrando identificar anos secos e chuvosos. Entre esses estudos, alguns serão destacados a seguir.

Araújo, Silva, Neto e Souza (2007) analisaram a variabilidade espaço-temporal da precipitação na bacia do Rio Paraíba usando os índices de anomalias de chuva. Mello e Oliveira (2016) realizaram um estudo com análise estatística e geoestatística na precipitação média no município de Joinville/SC. Moares e Costa (2005) estudaram a variação espacial e temporal da precipitação no estado do Pará. Passos, Raposo e Mendes (2017) estudaram a precipitação pluviométrica mensal e anual para o município de São Mateus/ES. Almeida, Oliveira, Sanches e Páscoa (2020) analisaram o comportamento e a previsão do nível do Rio Cuiabá por meio de modelos SETAR.

Neste contexto, por se tratar de dados observados ao longo do tempo, a análise de séries temporais é uma metodologia de previsão amplamente utilizada em estudos de precipitação pluviométrica. Uma série temporal consiste em uma coleção de observações feitas sequencialmente ao longo do tempo, caracterizadas por uma função matemática ou por distribuições de probabilidade. Uma característica muito importante é a relação de dependência ou autocorrelação serial entre as observações

vizinhas. A análise de séries temporais é uma metodologia aplicada no estudo de dados com essa natureza que possibilita modelar este padrão de autocorrelação e realizar previsões (Cintra, Melo e Bueno Filho, 2019).

Os modelos estatísticos usados, neste contexto, procuram assimilar o mecanismo que gera a série através de dados históricos, acumulados ao longo do tempo referentes a uma variável de interesse qualquer, tais como: finanças, economia, seguros, demografia, ciências sociais, meteorologia, energia, epidemiologia, etc. A adequabilidade destes é verificada através do ajuste das principais características de uma série temporal, denominadas de tendência e sazonalidade. Para modelar uma série temporal, a metodologia mais utilizada foi a proposta por Box e Jenkins, que se baseia na suposição de que uma observação, e/ou seu erro, em um determinado momento é influenciada pelo que ocorreu em um passado recente. Esses modelos são conhecidos, na literatura, como ARIMA (Autoregressive Integrated Moving Average) e suas decomposições (AR, MA, ARMA e SARIMA) (Barbosa et al., 2015).

Neste sentido, prever como o clima muda de um ano para o outro, possibilita um melhor gerenciamento da gestão pública, fazendo com que as principais atividades econômicas e sociais não sejam gravemente afetadas por períodos extremamente chuvosos ou secos. Destaca-se ainda que o modelo de previsão pode ser utilizado para auxiliar na tomada de decisão e na realização de planejamentos estratégicos futuros para aferir a vazão de água em usinas hidrelétricas próximas.

O presente trabalho tem por objetivo estudar a série histórica de precipitação pluviométrica média mensal da cidade de Lavras/MG a fim de obter um modelo estatístico que permita realizar previsões.

## 2. Metodologia

A série histórica de precipitação pluviométrica média mensal, em milímetros (mm), da cidade de Lavras/MG, foi utilizada para a presente análise estando disponível na página do Instituto Nacional de Meteorologia (INMET). Deste modo, fizeram parte da base de dados para modelagem de séries temporais 228 observações. Os valores de precipitação pluviométrica foram coletados diariamente e a precipitação pluviométrica média mensal foi o valor médio das precipitações diárias. Tal base de dados contém 240 observações mensais sobre a variável em estudo, de janeiro de 2000 a dezembro de 2019. O horizonte de previsão contemplou o ano de 2019, portanto, os valores reais dos 12 últimos meses não foram incluídos na modelagem e sim reservados para comparações com as previsões.

O procedimento inicial consiste em decompor graficamente a série temporal em suas componentes de tendência, sazonalidade e parte aleatória.

Para detectar a existência da componente de tendência pode-se utilizar um dos seguintes testes: Cox-Stuart, Wald-Wolfowitz ou Mann-Kendall. Neste trabalho aplicou-se o teste de Cox-Stuart (ou teste do sinal), o qual é baseado no coeficiente de relação de Spearman e no teste F para análise de variância. De acordo com Batista (2009), são necessários cinco passos para aplicar o teste:

- (i) agrupam-se as observações em pares  $(Z_1, Z_{c+1}), (Z_2, Z_{c+2}), (Z_{N-c}, Z_N)$ , em que  $c = \frac{n}{2}$  se  $N$ , números de observações, for par e  $c = \frac{N+1}{2}$  se  $N$  for ímpar;
- (ii) para cada par  $(Z_i, Z_{c+i})$ , associa-se um sinal. Este sinal é positivo (+) se  $Z_i < Z_{c+i}$  e negativo (-) se  $Z_i > Z_{c+i}$ . Descarta-se a comparação se  $Z_i = Z_{c+i}$ ;
- (iii) testa-se a hipótese de nulidade.  $H_0$ : Não existe tendência. A estatística do teste a ser utilizada é  $T$ , sendo que  $T$  é igual ao número de sinais positivos atribuídos no final da comparação entre  $Z_i$  e  $Z_{c+i}$ ;
- (iv) calcula-se o valor de  $t$ , que é dado por:  $t = \frac{1}{2} (n + \frac{\omega_\alpha}{2} \sqrt{n})$  em que  $n$  é o número de vezes que não houve empate na comparação entre  $Z_i$  e  $Z_{c+i}$ , e  $\frac{\omega_\alpha}{2}$  é o valor tabelado ao nível de significância  $\alpha$ . Caso  $n \leq 20$ , o valor tabelado de  $\frac{\omega_\alpha}{2}$  é dado

pela distribuição binomial com  $p = 0,5$  e, caso  $n > 20$ , utiliza-se a distribuição normal e

(v) se  $T \leq t$  ou  $T \geq n - t$  rejeita  $H_0$ , a série possui tendência.

Para verificar a presença da componente de sazonalidade pode-se aplicar o teste de Kruskal-Wallis ou o teste de Friedman. Aqui, utilizou-se o teste de Kruskal-Wallis, sendo este, análogo ao teste F utilizado na ANOVA de um fator, exceto pelo fato de que ele não coloca nenhuma restrição sobre a comparação, enquanto na análise de variância as populações em confronto são independentes e normalmente distribuídas. De acordo com Morettin e Tolo (2006), seis passos são necessários para realização do teste:

- (i) estabelecem-se as hipóteses:  $H_0 : t_1 = t_2 = \dots = t_k$  e  $H_1 : t_1, t_2, \dots, t_k$  não são todos iguais;
- (ii) ordena-se de forma crescente de magnitude os valores deste novo conjunto de dados e associa-se a cada valor seu posto correspondente tendo cada posto o mesmo sinal do valor que este representa;
- (iii) calcula-se o valor da estatística  $H$ . Em seguida, fixa-se o nível de significância  $\alpha$ ;
- (iv) encontram-se os valores críticos referentes ao nível de significância fixado. Neste caso, calcula-se o valor  $Q_\alpha$  (percentil da distribuição Qui-Quadrado referente ao grau de liberdade e ao nível de significância) de modo que  $P [H > Q_\alpha] = \alpha$  (sob  $H_0$ );
- (v) se  $H_{obs} > Q_\alpha$ , rejeita-se a hipótese nula de que as amostras provêm de populações igualmente distribuídas e
- (vi) o p-valor é calculado da seguinte forma  $P_{valor} = P[\chi_{k-1}^2 \geq H | H_0]$ .

A presença de sazonalidade também foi verificada através do gráfico do periodograma. Suponha que  $\epsilon_t$  sejam  $n$  observações de um processo estocástico da série temporal vinculadas ao tempo  $t$ . Um estimador de espectro  $I_p(f_i)$  do processo é dado por:

$$I_p(f_i) = \frac{2}{\left(\frac{n}{2}\right)} \left( \sum_{t=1}^n \epsilon_t \cos\left[\frac{2\pi i}{\left(\frac{n}{2}\right)} t\right]^2 - \sum_{t=1}^n \epsilon_t \sin\left[\frac{2\pi i}{\left(\frac{n}{2}\right)} t\right]^2 \right) \quad (1)$$

em que:

$0 < f_i < \frac{1}{2}$ . Este estimador tem a finalidade de detectar periodicidades nos dados. Um pico na frequência  $f_i = \frac{i}{n}$  indica uma periodicidade de período  $\frac{1}{f_i}$  (Morettin e Tolo, 2006).

Para tornar uma série estacionária, um procedimento comum é tomar diferenças sucessivas da série original. Para eliminar a componente de tendência, tome a primeira diferença, definida por  $\Delta Z_t = Z_t - Z_{t-1}$  com  $t = 1, 2, \dots, n$ . Geralmente, duas diferenças são suficientes. A sazonalidade determinística pode ser eliminada por meio de uma diferença realizada na série através do método de médias móveis, levando-se em conta a ordem do período  $s$ .

Depois de eliminadas as componentes de tendência e/ou sazonalidade, a série se encontra estacionária, parte-se então para a identificação do modelo. Para sugerir um modelo é necessário analisar as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) da série diferenciada. As correlações significativas da ACF indicam a ordem do modelo de médias móveis (MA) e as da PACF indicam a ordem do modelo autorregressivo (AR). Caso haja correlações significativas nos lags múltiplos de  $s$  (período sazonal), deve-se ajustar a parte sazonal do modelo.

Box, Jenkins, Reinsel e Ljung (2015) propõem o modelo sazonal multiplicativo de período  $s$ , denotado por SARIMA  $(p, d, q) \times (P, D, Q)_s$  ou ARIMA Sazonal, representado por:

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D W_t = \theta(B)\Theta(B^s)a_t. \quad (2)$$

Em que:

$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  é o polinômio autorregressivo de ordem  $p$ ;

$\Phi(B^s) = 1 - \Phi B^s - \dots - \Phi_p B^{ps}$  é o polinômio autorregressivo sazonal de ordem  $P$ ;

$\Delta^d = (1 - B)^d$  é o operador diferença e  $d$  é o número de diferenças necessárias para retirar a tendência da série;

$\Delta_s^D = (1 - B^s)^D$  é o operador diferença generalizado, quando duas observações estão distantes entre si de  $s$  intervalos de

tempos que apresentam alguma semelhança, e  $D$  é o número de diferenças de *lags*  $s$  necessárias para retirar a sazonalidade da série;

$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  é o polinômio médias móveis de ordem  $q$ ;

$\Theta(B^s) = 1 - \theta_1 B^s - \dots - \theta_Q B^{Qs}$  é o polinômio médias moveis sazonal de ordem  $Q$ ;

$W_t$  representa os valores da série original após a aplicação de diferenças de ordem 1 e ordem  $s$ , necessárias para tornar a série estacionária e

$a_t$  é o ruído aleatório no tempo  $t$ .

Após a fase de estimação, o próximo passo é fazer o diagnóstico da adequação do modelo ajustado, isto é, verificar se os resíduos são independentes e normalmente distribuídos, podendo então ser denominados de ruído branco. Para avaliar a presença de autocorrelação serial utilizou-se o teste estatístico Ljung-Box (Ljung e Box, 1978). Se o modelo ajustado para um processo SARIMA  $(p, d, q) \times (P, D, Q)_s$  é apropriado, então

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)} \sim \chi_{K-p-q-P-Q}^2. \quad (3)$$

A hipótese de ruído branco é rejeitada para  $Q(K) > \chi_{K-p-q-P-Q}^2$ .

Dos modelos ajustados, seleciona-se aqueles que o resíduo apresenta um ruído branco. Destes, utiliza-se o Critério de Informação de Akaike (AIC) para compará-los. Akaike (1973) sugere escolher como melhor modelo, aquele cujas ordens  $p$  e  $q$  minimizam o critério de Akaike (AIC) normalizado. Na comparação de diversos modelos, com  $N$  fixo, o critério de Akaike pode ser expresso por:

$$AIC(k, l) = N \ln \hat{\sigma}_a^2 + 2(k + l + 2). \quad (4)$$

Em que:

$\hat{\sigma}_a^2$  é o estimado de máxima verossimilhança de  $\hat{\sigma}_a^2$ ,  $0 \leq k \leq p$  e  $0 \leq l \leq q$ .

$k$  e  $l$  correspondem ao número de parâmetros que minimizam o critério, independente do modelo ser completo ou incompleto.

Entre os vários modelos propostos, será escolhido como o melhor modelo aquele que tiver o menor valor de AIC. Além do AIC, existem outros critérios de informação, tais como: Bayesiano, Hanna-Quinn e Schwarz (Paiva, Heval e Sáfiadi, 2019).

Com o modelo selecionado e ajustado, pode-se utilizá-lo para prever valores futuros. Morettin e Tolo (2006) salientam que a previsão não constitui um fim em si, mas apenas um meio de fornecer informações para uma consequente tomada de decisões, visando a determinar um objetivo. A previsão  $Z_{t+h}$  denotada por  $\hat{Z}_t(h)$  é a esperança matemática condicional de  $Z_{t+h}$ , dado  $Z_t, Z_{t-1}, \dots$ , isto é,

$$\hat{Z}_t(h) = [Z_{t+h}] = E[Z_{t+h} | Z_t, Z_{t-1}, \dots]. \quad (5)$$

Se o propósito do trabalho for realizar previsões, pode-se basear no indicador de erro quadrático médio (EQM) que é uma boa medida para comparar previsões entre modelos. As estimativas de EQM são dadas pela média dos quadrados das diferenças entre valores observados e previstos. Logo, EQM com origem em  $t$  é dado por:

$$EQM_t = \frac{1}{n} \sum_{h=1}^n [Z_{t+h} - \hat{Z}_t(h)]^2. \quad (6)$$

O modelo escolhido é o que possuir menor erro quadrático médio.

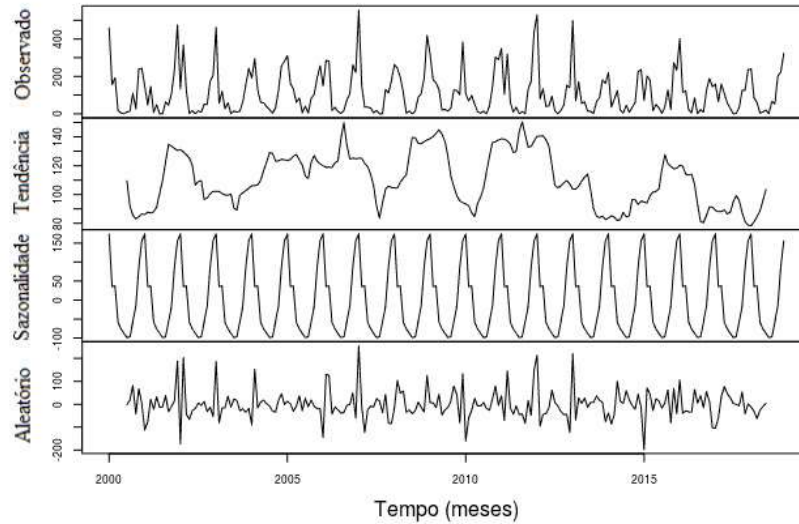
Utilizou-se o *software* livre R (R Core Team, 2021) para manipulação e análise estatística dos dados, sendo o *stats*, *forecast* e *timeSeries* os principais pacotes utilizados. As análises foram realizadas considerando um nível de significância  $\alpha = 5\%$ .

### 3. Resultados e Discussão

A Figura 1 mostra o gráfico da série temporal de precipitação pluviométrica média mensal do município de

Lavras/MG decomposta em componentes de tendência, sazonalidade e parte aleatória. A análise gráfica não indica a presença da componente tendência na série temporal e sugere a existência de sazonalidade aditiva, mas é necessário confirmar tais informações através de testes matemático-estatísticos.

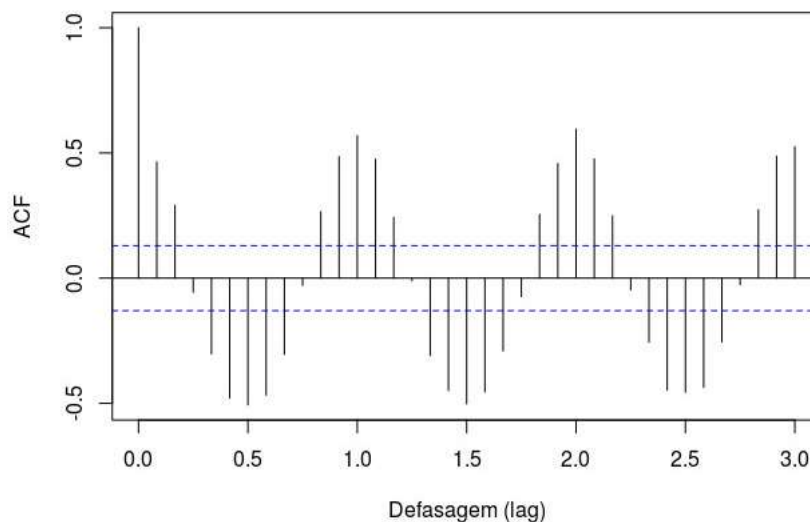
**Figura 1:** Série temporal de precipitação pluviométrica média mensal (em milímetros) decomposta em componentes de tendência, sazonalidade e parte aleatória, durante o período de janeiro de 2000 a dezembro de 2018.



Fonte: Autores.

Na Figura 2 tem-se o gráfico da função de autocorrelação (ACF) para a precipitação pluviométrica média. O comportamento senoidal da ACF sugere que a série possui uma periodicidade sazonal.

**Figura 2:** Gráfico da função de autocorrelação (ACF) da série de precipitação pluviométrica média.



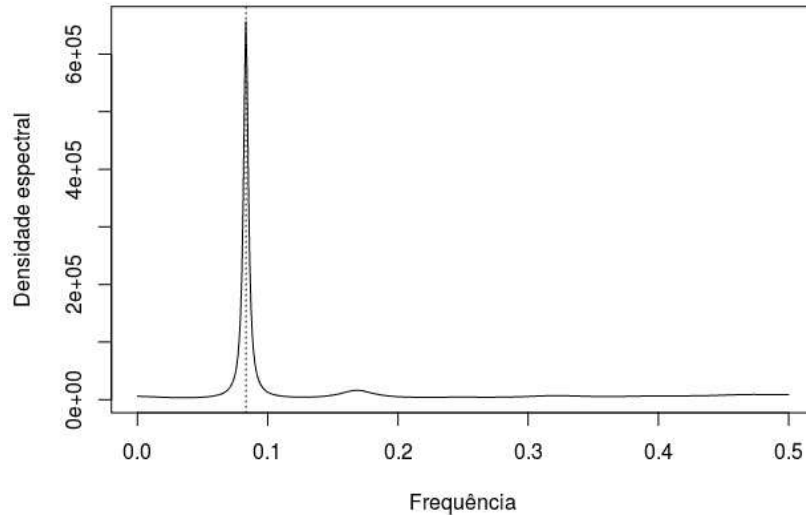
Fonte: Autores.

Através do teste de Cox-Stuart, verificou-se a ausência da componente tendência determinística com um p-valor de 0,7789, não rejeitando, ao nível de 5% de significância, a hipótese nula de existência de tendência. Sendo assim, não foi necessário realizar uma diferença de ordem 1 na série a fim de remover o efeito da tendência.

Ao aplicar o teste não paramétrico de Kruskal-Wallis o resultado do p-valor foi menor que 0,01, rejeitando, ao nível

de 5% de significância, a hipótese nula de ausência de sazonalidade determinística, assim, confirmando a presença da componente sazonal e corroborando com a inspeção visual da Figura 2, onde os *lags* apresentaram comportamento senoidal.

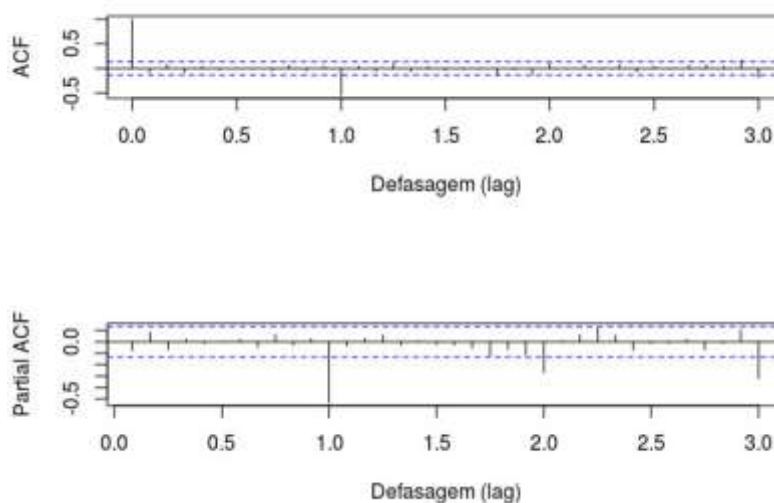
**Figura 3:** Periodograma da série de precipitação pluviométrica média.



Fonte: Autores.

Para descobrir o padrão periódico da série e detectar sua frequência de ocorrência, foi confeccionado o gráfico do periodograma (ver Figura 3). Observa-se que a maior densidade espectral (pico) está associada à frequência  $f_i = 0,083$  aproximadamente, o que indica uma componente sazonal com período de  $\frac{1}{f_i} = \frac{1}{0,083} = 12$  meses. Segundo Barbosa et al. (2015), a inexistência da componente tendência e a presença de sazonalidade já eram esperadas, uma vez que a variável em estudo é meteorológica. Assim, uma diferença de ordem 12 foi usada para retirar a componente sazonal presente na série de precipitação pluviométrica média da cidade de Lavras/MG.

**Figura 4:** Gráfico das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) após aplicação da diferença sazonal.



Fonte: Autores.

Constata-se, na Figura 4, que ainda existem *lags* significativos múltiplos de 12 nas ACF e PACF, mesmo depois de aplicada a diferença sazonal, o que de acordo com Morettin e Tolo (2006) assinala a presença de sazonalidade estocástica a qual deve ser removida ajustando-se um modelo que a represente, além da determinística, eliminada por diferença.

A partir dos correlogramas da Figura 4 foi possível propor alguns modelos SARIMA  $(p,d,q) \times (P,D,Q)_s$  para a série diferenciada. A identificação do modelo é a fase mais crítica durante o ciclo iterativo, pois esta escolha é baseada, principalmente, nas funções de autocorrelação e autocorrelação parcial estimadas, nesse sentido, é importante salientar que pode ser identificado mais de um modelo para a mesma série. Depois de identificado o modelo, o passo seguinte é estimar os parâmetros através do método de máxima verossimilhança.

Segundo Barbosa et al. (2015), alcançar a ordem precisa do modelo através destes correlogramas é algo difícil e subjetivo, devido a enorme quantidade de modelos que podem ser ajustados. Como o teste para tendência não foi significativo, não foi necessário realizar uma diferença de ordem 1, portanto  $d = 0$ . No entanto, o teste para sazonalidade determinística foi significativo, fazendo com que uma diferença sazonal de ordem  $s = 12$  fosse efetuada para remover a sazonalidade, dessa forma, tomou-se  $D = 1$ . Dentre os modelos ajustados, foram selecionados aqueles que tiveram menor AIC e p-valor superior a 5% (nível de significância) para o teste de Ljung-Box, o que implica em resíduos independentes e normalmente distribuídos, ou seja, ruído branco (ver Tabela 1).

**Tabela 1:** Critério de informação de Akaike (AIC), p-valor para o teste de Ljung-Box e raiz quadrada do erro quadrático médio ( $\sqrt{EQM}$ ) para os modelos SARIMA selecionados.

Modelo	SARIMA (s = 12)	AIC	Ljung-Box (p-valor)	$\sqrt{EQM}$
1	(0,0,0) x (0,1,1)	775,42	0,9938	63,36
2	(0,0,0) x (2,1,0)	790,87	0,4330	70,10
3	(0,0,2) x (1,1,0)	813,84	0,0675	74,45
4	(0,0,1) x (2,1,0)	792,79	0,3943	69,97
5	(0,0,2) x (2,1,0)	791,09	0,7340	68,92
6	(0,0,0) x (4,1,1)	776,39	0,9964	63,60
7	(0,0,0) x (6,1,0)	777,60	0,9967	64,59

Fonte: Autores.

Os modelos, SARIMA  $(0,0,0) \times (0,1,1)_{12}$  e SARIMA  $(0,0,0) \times (4,1,1)_{12}$ , apresentaram os menores valores de AIC, no entanto, o primeiro foi escolhido por possuir menor EQM, pois o intuito deste trabalho é encontrar um modelo que forneça melhores previsões para a precipitação pluviométrica da cidade de Lavras/MG. Os coeficientes estimados do modelo SARIMA  $(0,0,0) \times (0,1,1)_{12}$  estão na Tabela 2.

**Tabela 2:** Estimativas para o modelo SARIMA  $(0,0,0) \times (0,1,1)_{12}$ .

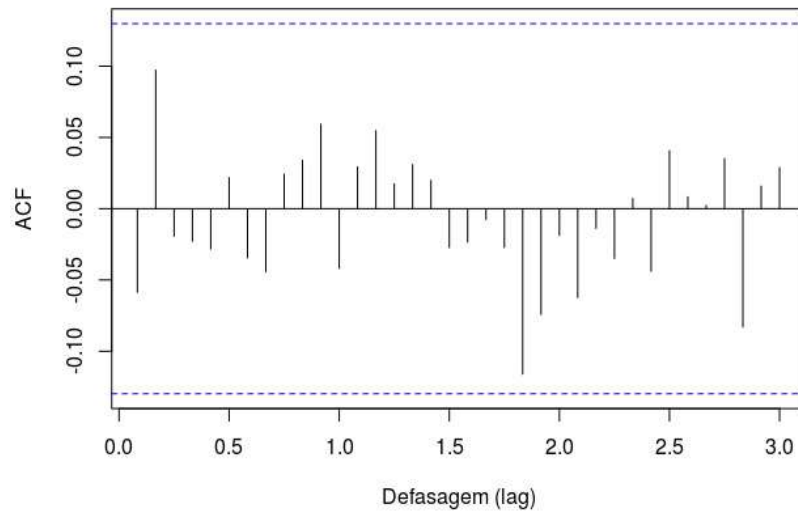
Parâmetro	Coefficiente	Erro padrão	p-valor
$\theta_1$	- 0,7593	0,0552	0,0000*

\*Significativo a 0,05 de probabilidade. Fonte: Autores.

A Figura 5 mostra o gráfico da função de autocorrelação (ACF) para os resíduos do modelo SARIMA  $(0,0,0) \times (0,1,1)_{12}$ , onde se observa que os lags estão contidos no intervalo de confiança de 95%, indicando que os resíduos são padronizados e representam ruído branco.



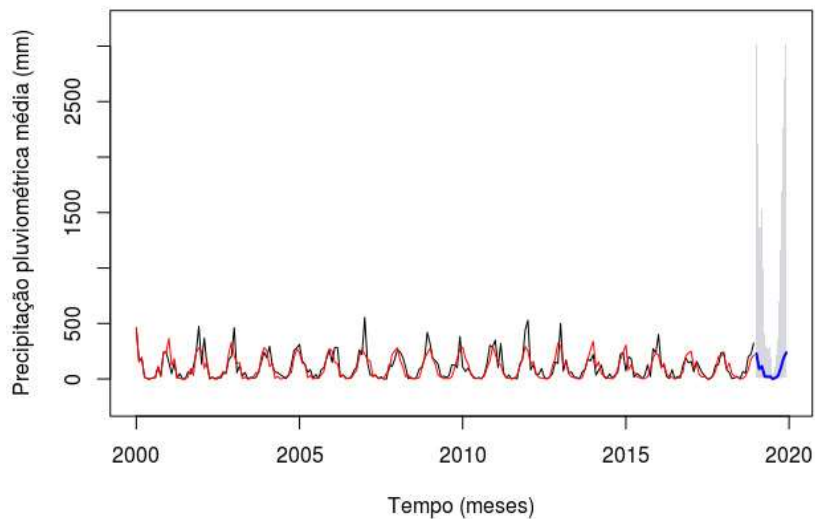
**Figura 5:** Gráfico da função de autocorrelação (ACF) residual do modelo SARIMA (0,0,0) x (0,1,1)<sub>12</sub>.



Fonte: Autores.

A Figura 6 ilustra a série histórica de precipitação pluviométrica média mensal no município de Lavras/MG (em preto), o ajuste do modelo aos dados (em vermelho), a previsão para os 12 meses de 2019 (azul escuro) e os intervalos de confiança a 95% para o ano de 2019 (azul claro). Nota-se que os valores ajustados ficaram próximos dos reais.

**Figura 6:** Valores de precipitação pluviométrica média reais (preto), ajustados (vermelho) e previstos (azul escuro).



Fonte: Autores.

A Tabela 3 apresenta os valores reais, os previstos e os respectivos intervalos de confiança para a precipitação pluviométrica média mensal da cidade de Lavras/MG para os 12 meses de 2019. Observa-se que os intervalos de confiança a 95% gerados através do modelo ajustado contiveram os doze valores reais de precipitação pluviométrica média mensal para o ano de 2019.

**Tabela 3:** Estimativas de precipitação pluviométrica média mensal da cidade de Lavras/MG para o ano de 2019.

Meses	Limite inferior (2,5%)	Limite superior (97,5%)	Previsto (mm)	Valor real (mm)
Janeiro	16,9	3053,9	227,7	143,6
Fevereiro	6,7	1211,8	90,3	195,7
Março	8,5	1544,8	115,2	221,6
Abril	1,8	335,1	24,9	77,3
Mai	1,4	261,9	19,5	24,3
Junho	1,6	297,8	22,2	2,6
Julho	0,0	9,2	0,6	9,2
Agosto	0,6	125,7	9,3	10,4
Setembro	2,3	414,4	30,9	71,3
Outubro	7,6	1369,0	102,0	82,7
Novembro	14,1	2543,7	189,6	187,9
Dezembro	17,6	3182,2	237,2	120,0

Fonte: Autores.

Salienta-se que no trabalho de Pereira, Costa, Oliveira, Borges e Filho (2015) os modelos SARIMA também foram aplicados para previsão de precipitação pluvial e temperatura na cidade de Areia/PB, onde apresentaram resultados satisfatórios com baixos indicadores de erro e resíduos com características de ruído branco.

#### 4. Conclusão

A série de precipitação pluviométrica média mensal da cidade de Lavras/MG apresentou uma componente sazonal com periodicidade de 12 meses.

O modelo ajustado obteve um resultado muito bom, pois os intervalos de confiança a 95% contiveram os doze valores reais de precipitação pluviométrica média mensal da cidade de Lavras/MG para o ano de 2019, mesmo diante dos imprevistos e incertezas associadas a fatores climáticos. O modelo em questão pode ser utilizado na tomada de decisões para a realização de planejamentos estratégicos futuros que envolvem questões públicas associadas à cidade de Lavras.

Estas previsões também podem ser utilizadas para auxiliar os gestores da usina hidrelétrica do Funil/MG a programar operações e manutenções futuras da vazão de água de forma adequada, pois esta se encontra próxima ao município de Lavras.

A metodologia de séries temporais proposta revelou-se apropriada para a modelagem de dados meteorológicos.

E para futuros trabalhos, sugere-se a inclusão da análise de intervenção ao modelo SARIMA ajustado.

#### Agradecimentos

Os autores Pedro Henrique Alves Bittencourt Santos e Otávio Augusto dos Santos Delfino agradecem ao Programa Institucional de Iniciação Científica Voluntária – PICV do CEFET-MG.

#### Referências

- Almeida, E. M., Oliveira, A. C. S., Sanches, L., & Pascoa, M. A. R. (2020). Comportamento e previsão do nível do rio Cuiabá por meio de modelos SETAR. *Sigma*, 9(2), 1-13.
- Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 21, 243-247.
- Araújo, L. E., Silva, D. F., Neto, J. M. M., & Souza, F. A. S. (2007). *Análise da variabilidade espaço-temporal da precipitação na bacia do Rio Paraíba usando IAC*. *Revista de Geografia*, 22(2), 23-26.

- Barbosa, E. C., Sáfadi, T., Nascimento, M., Nascimento, A. C. C., Silva, C. H. O., & Manuli, R. C. (2015). Metodologia Box & Jenkins para previsão de temperatura média mensal da cidade de Bauru (SP). *Revista Brasileira de Biometria*, 33(1), 104-117.
- Batista, A. L. F. (2009). *Modelos de séries temporais e redes neurais artificiais na previsão de vazão*. Dissertação de mestrado, Universidade Federal de Lavras, Lavras, MG, Brasil.
- Box, G. E. P., Jenkins, G. M., Reinsel, G., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control (4a ed.)*. Wiley.
- Cintra, R. A., Melo, M. I. P., & Bueno Filho, J. S. S. (2019). Modelos de séries temporais para a previsão da temperatura média mensal de Lavras, MG. *Sigmae*, 8(2), 596-605.
- Ferreira, R. A., Miranda, V. F. L., Santos, P. M., & Sáfadi, T. (2020). *Um estudo sobre a evolução de óbitos por câncer de mama no Brasil usando modelos de séries temporais*. *Research, Society and Development*, 9(12), 1-20.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2), 190-195.
- Instituto Nacional de Meteorologia. (2020). Recuperado em 15 de outubro de 2020, de <https://portal.inmet.gov.br/dadoshistoricos#>.
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 97-303.
- Mello, Y. R., & Oliveira, T. M. N. (2016). Análise estatística e geoestatística da precipitação média para o município de Joinville/SC. *Revista Brasileira de Meteorologia*, 31(2), 229-239.
- Miranda, T. P. (2016). *Previsão da precipitação mensal do município de Ouro Branco - MG, por meio de modelos de séries temporais*. Dissertação de mestrado, Universidade Federal de São João Del Rei, São João Del Rei, MG, Brasil.
- Moraes, B. C., Costa, J. M. N., Costa, A. C. L., & Costa, M. H. (2005). Variação espacial e temporal da precipitação no estado do Pará. *Acta Amazônica*, 35(2), 207-214.
- Morettin, P. A., & Toloi, C. M. C. (2006). *Análise de Séries Temporais (2a ed.)*. E. Blucher.
- Paiva, D. A., Herval, A. C. F., & Sáfadi, T. (2019). Metodologia de Séries Temporais como ferramenta de análise na produção de frangos no Brasil. *Sigmae*, 8(2), 227-237.
- Passos, M. L. V., Raposo, A. B., & Mendes, T. J. (2017). Precipitação pluviométrica mensal e anual provável para o município de São Mateus /ES. *Agropecuária Científica no Semiárido*, 13(2), 162-168.
- Pereira, A. R., Costa, A. S., Oliveira, V. G., Borges, P. F., & Filho, A. I. (2015). Análise do comportamento das médias anuais da precipitação pluvial e temperatura da cidade de Areia, Paraíba. *Gaia Scientia*, 9(1), 67-73.
- R Development Core Team. (2021). *R: a language and environment for statistical computing*. Vienna: R Foundation for statistical computing. <http://www.Rproject.org>.
- Santos, F. A., & Aquino, C. M. S. (2017). Análise da precipitação pluviométrica no município de Castelo do Piauí, Nordeste do Brasil. *Geosp*, 21(2), 619-633.