

## ANÁLISE E PREVISÃO DE TEMPERATURA MÁXIMA MENSAL DA CIDADE DO RIO DE JANEIRO (RJ) PARA O ANO DE 2016

Eduardo Campana BARBOSA<sup>1</sup>  
Carlos Henrique Osório SILVA<sup>1</sup>  
Moysés NASCIMENTO<sup>1</sup>  
João Eustáquio de LIMA<sup>2</sup>

- RESUMO: Neste trabalho comparou-se o ajuste dos modelos aditivo de Holt-Winters e SARIMA à série de temperatura máxima mensal da cidade do Rio de Janeiro (RJ), com dados de Janeiro de 2003 a Dezembro de 2015. O modelo aditivo de Holt-Winters foi escolhido devido a característica sazonal da série e o modelo SARIMA  $(1,0,1) \times (1,1,1)_{12}$  via Critérios de Informação de Akaike e de Schwarz. A escolha do modelo final ocorreu pelos indicadores de Erro Quadrático Médio e Erro Percentual Médio Absoluto, calculados pelo ajuste aos dados históricos. Adicionalmente, o teste de Diebold e Mariano e as características de ruído branco dos resíduos foram avaliados. O modelo SARIMA  $(1,0,1) \times (1,1,1)_{12}$  foi o modelo que melhor se ajustou aos dados e, por isso, foi adotado para previsão de temperatura máxima mensal para o ano de 2016.
- PALAVRAS-CHAVE: Holt-Winters; SARIMA; sazonalidade; temperatura.

### 1 Introdução

Sob a abordagem estatística clássica, cada observação de uma série temporal é composta pela combinação de suas componentes não observáveis, denominadas como Tendência, Sazonalidade, Ciclos e Variações Aleatórias (BROCWELL e RICHARD, 1991). Na abordagem moderna, define-se uma série temporal como a realização de um processo estocástico, isto é, um conjunto de observações ordenadas no tempo e descritas por suas distribuições de probabilidade (SHUMWAY e STOFFER, 2010). Uma série temporal pode apresentar diferentes formas de dependência entre as observações atuais e passadas, sendo de grande importância a modelagem dessas relações, principalmente, para o desenvolvimento de previsões futuras (NAJAFI e TARAZKAR, 2006).

É neste contexto que a análise de séries temporais é empregada. Nessa metodologia, modelos estatísticos são ajustados a dados históricos, coletados no domínio do tempo e referentes a uma variável de interesse, para captar padrões e o mecanismo gerador da série. Dentre os modelos mais utilizados destacam-se os de suavização exponencial de Holt-Winters (HOLT, 1957; WINTERS, 1960) e a ampla classe de modelos propostos por

---

<sup>1</sup> Universidade Federal de Viçosa - UFV, Departamento de Estatística, CEP: 36570-977, Viçosa, MG, Brasil. E-mail: duducampana@hotmail.com; chos@ufv.br; moysesnascm@ufv.br

<sup>2</sup> Universidade Federal de Viçosa - UFV, Departamento de Economia Rural, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: jelima@ufv.br

Box e Jenkins (1976), dentre estes o SARIMA ou ARIMA sazonal. Métodos computacionais mais recentes, como o de Redes Neurais Artificiais (RUMELHART *et al.*, 1985), são também utilizados para modelagem e previsão.

A análise de séries temporais é aplicada em diversas áreas do conhecimento, pois muitas variáveis relacionam-se ao domínio do tempo. Cho (2002) verificou que as Redes Neurais Artificiais conseguiram prever melhor a demanda de viagens para Hong Kong, em relação aos modelos de Holt-Winters e ARIMA. Dooley e Lenihan (2005) utilizaram o modelo ARIMA para prever o preço de Chumbo e Zinco em Londres. Najafi e Tarazkar (2006) verificaram melhor desempenho do modelo ARIMA em relação às Redes Neurais Artificiais, para previsão de exportação de pistache no Irã. Henry *et al.* (2007) compararam Redes Neurais Artificiais com os modelos de Holt-Winters e SARIMA, para previsão de exportação mensal de arroz na Tailândia, optando pela segunda abordagem. Em um estudo recente, Huddleston, Porter e Brown (2015) aplicaram os modelos de Holt-Winters e ARIMA para previsão semanal de roubos na cidade de Pittsburgh na Pensilvânia.

Mais especificamente, para variáveis climatológicas, a análise de séries temporais pode contribuir em diversos aspectos, seja na melhoria da vida em sociedade, em atividades e no lazer das pessoas, no turismo, na análise de precipitação e de recursos hídricos; seja no ambiente empresarial, para o planejamento de recursos e investimentos; seja no gerenciamento da agricultura, etc. (BACK, 2001). Medeiros *et al.* (2005) e Hubbard (2007) corroboraram algumas dessas informações ao demonstrarem, respectivamente, a influência dos efeitos fisiológicos da temperatura do ar sobre plantas e animais, bem como nas atividades relacionadas à agricultura e pecuária. Karner (2009) ajustou modelos ARIMA para analisar séries temporais diárias de temperatura do ar, com dados de 10 estações meteorológicas localizadas na Europa e 4 na Ásia. Tularam e Ilahee (2010) ajustaram modelos ARIMA com variáveis regressoras, com dados de 4 estações meteorológicas de Queensland, na Austrália, para desenvolver previsões e concluir sobre a relação inversa entre precipitação e temperatura. No Brasil, Chechi e Bayer (2012) compararam as previsões dos modelos de Holt-Winters e SARIMA para séries de temperatura mínimas e máximas da cidade de Erechin, no Rio Grande do Sul. Silva, Guimarães e Tavares (2008) apresentaram estudos semelhantes, no entanto, comparando modelos SARIMA para previsão de temperatura média da cidade de Uberlândia, em Minas Gerais.

Neste sentido, o presente trabalho tem por objetivo o estudo da série de temperatura máxima mensal da cidade do Rio de Janeiro-RJ e, principalmente, a definição de um modelo estatístico que forneça previsões confiáveis para os 12 meses do ano de 2016. O Rio de Janeiro ou Cidade Maravilhosa (epíteto) é internacionalmente conhecido por diversos e belos pontos turísticos e culturais, como as praias de Copacabana e Ipanema, as florestas da Tijuca e da Pedra Branca, o bairro boêmio da Lapa e seus arcos, o estádio do Maracanã, o Teatro Municipal entre muitos outros (UNESCO, 2012). Devido a tantos atrativos, a cidade foi considerada a maior rota de turismo internacional do Brasil, da América Latina e do Hemisfério Sul (BUENO, 2007). Destaca-se que, em épocas específicas do ano, como em Fevereiro e Dezembro, a atividade turística é intensificada devido a eventos como o carnaval e o réveillon. Logo, conhecer a priori os valores máximos que a temperatura do ar pode atingir é uma informação de fundamental

importância no que se refere ao: *i*) turismo, devido ao clima mais frio de outras regiões ou países e *ii*) saúde pública, pois a exposição excessiva aos raios solares, em períodos mais quentes no ano, pode ocasionar problemas de queda de pressão arterial, tonturas ou queimaduras. Adicionalmente, em 2016, a importância de informações desse tipo torna-se mais necessária, pois o Rio de Janeiro sediará as Olimpíadas (de 5 a 21 de Agosto) e as paraolimpíadas (de 7 a 18 de Setembro), recebendo atletas, turistas e a atenção do mundo inteiro.

## 2 Material e métodos

### 2.1 Descrição dos dados

Os dados utilizados foram obtidos na página do Instituto Nacional de Meteorologia INMET (BDMEP – Banco de Dados Meteorológicos para Ensino e Pesquisa) e referem-se à série histórica de temperatura máxima mensal, em graus celsius (°C), da cidade do Rio de Janeiro (RJ). A base de dados contém informações dos últimos 13 anos sobre a variável em estudo, totalizando 156 observações mensais, de janeiro de 2003 a dezembro de 2015. Mais especificamente, cada observação consiste no valor máximo de temperatura referente ao dia mais quente de cada mês analisado. Essas informações foram coletadas na estação meteorológica convencional do Rio de Janeiro, código OMM 83743, localizada sob as seguintes coordenadas geográficas: Latitude: -22,89°, Longitude: -43,18° e Altitude: 11,10 metros.

### 2.2 Análise da série temporal

O procedimento inicial consistiu em análise gráfica e descritiva da série, apresentando o boxplot mensal e algumas estatísticas básicas, como os valores mínimos e máximos, a média, o desvio-padrão e o coeficiente de variação dos dados.

Posteriormente, para a análise de tendência determinística e/ou estocástica na série e, conseqüentemente, verificação de sua estacionariedade, procedeu-se o teste de raiz unitária de Dickey-Fuller aumentado ou ADF (DICKEY e FULLER, 1979) com um intercepto e um termo de tendência determinística linear, conforme o modelo de regressão em (1).

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{i=1}^p \gamma_i \Delta Y_{t-i} + \varepsilon_t \quad (1)$$

em que  $\Delta Y_t$  são os valores da série em primeira diferença,  $Y_{t-1}$  são os valores da série original defasados um período de tempo,  $\alpha$ ,  $\beta$  e  $\delta$  são parâmetros a serem estimados por Mínimos Quadrados Ordinários (MQO) e que representam, respectivamente, os efeitos do intercepto, da tendência determinística linear e da raiz unitária na série e  $\varepsilon_t$  são erros ruído branco, isto é,  $\varepsilon_t \sim RB(0, \sigma^2)$ , não autocorrelacionados, com média zero e variância constante (PFAFF, 2008). Os termos dentro do operador somatório são defasagens da variável resposta em primeira diferença, incluídas no modelo para remover a autocorrelação residual. Mais especificamente, se o valor da estatística  $t$  indicar que a estimativa de  $\beta$  não foi significativa, conclui-se sobre a ausência de tendência linear

determinística. Se o valor da estatística de Dickey-Fuller indicar que a estimativa de  $\delta$  foi significativa, conclui-se sobre a ausência de tendência estocástica.

Adicionalmente, para detecção de tendência linear determinística ou monotônica, foi empregado o teste não paramétrico de Mann-Kendall, cuja estatística  $\tau$  é baseada no coeficiente de correlação de Kendall entre  $Y_t$  e  $t$ , e a hipótese nula é que não há tendência na série (detalhes em MANN, 1945; KENDALL, 1948). Testes não paramétricos como o Mann-Kendall, Cox-Stuart e o Spearman rank correlation são frequentemente empregados para análise de tendência em séries climatológicas (SONALI e KUMAR, 2013).

A existência de sazonalidade, sob a estacionariedade da série  $Y_t$ , foi verificada via análise espectral, decompondo a série no domínio das frequências de Fourier. Foi utilizado o gráfico do periodograma, que possui no eixo das ordenadas às frequências de Fourier ( $f_i$ ) da série, em que  $0 \leq f_i = i/n \leq 1/2$ , e no eixo das abscissas as respectivas intensidade  $I(f_i)$ , definidas conforme em (2) (CRYER e CHAN, 2008):

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

A periodicidade sazonal ( $s$ ) está associada à frequência de Fourier de maior intensidade segundo a relação  $s = 1/f_i$ . Para testar a significância estatística dessa periodicidade utilizou-se o teste G de Fisher. A decisão do teste, sobre a existência de sazonalidade com periodicidade  $s = 1/f_i$ , é comprovada se o valor da estatística G (3) for superior ao valor de Z (4) (FISHER, 1929). Morettin e Tolo (2006) sugerem a aproximação em (5) para a distribuição exata da estatística G, que permite calcular o p-valor do teste.

$$G = \frac{\max[I(f_i)]}{\sum_{i=1}^{n/2} I(f_i)} \quad (3)$$

$$Z = 1 - \frac{\alpha}{\binom{n}{2}} \left[ \frac{1}{\left(\frac{n}{2}\right)^{-1}} \right] \quad (4)$$

$$P(G > g) \approx \binom{n}{2} (1 - G)^{\frac{n}{2}-1} \quad (5)$$

em que  $n$  é o número de observações da série e  $\alpha$  o nível de significância adotado (5%).

### 2.3 Modelos de Holt-Winters

Os modelos de suavização exponencial de Holt-Winters (HOLT, 1957; WINTERS, 1960) são empregados para ajustar séries temporais com tendência e sazonalidade determinísticas. Apresentam-se sob a forma aditiva (6) ou multiplicativa (7), de acordo com o comportamento sazonal exibido pela série. Segundo Chatfield e Yar (1988), se a

amplitude da variação sazonal for constante, o modelo mais apropriado é o aditivo, caso contrário, se essa amplitude for proporcional ao nível da série, o multiplicativo.

$$Y_t = N_t + T_t + S_t + \varepsilon_t \quad (6)$$

$$Y_t = N_t S_t + T_t + \varepsilon_t \quad (7)$$

em que  $N_t$  é a componente de nível,  $T_t$  a componente de tendência,  $S_t$  a componente sazonal e  $\varepsilon_t$  são erros ruído branco.

Neste trabalho, a decisão sobre qual dos dois modelos adotar ocorreu por análise gráfica da série e Regressão Linear Simples, utilizando a amplitude sazonal de cada ano como variável dependente e, o tempo (em anos), como variável independente. Se o coeficiente da variável tempo não for significativo, há evidências de que a variação da amplitude sazonal é constante e o modelo (6) é o mais indicado, caso contrário, o modelo (7). O ajuste desses modelos é realizado por três constantes de suavização, denominadas como  $\alpha$ ,  $\beta$  e  $\gamma$ , que representam, respectivamente, o efeito de nível, de tendência e de sazonalidade na série. O valor dessas constantes varia entre 0 e 1, isto é,  $0 \leq \alpha, \beta, \gamma \leq 1$  e os níveis ótimos foram estimados visando tornar mínima a soma de quadrados dos erros (GRANGER e NEWBOLD, 1977). O quadro 1 apresenta as equações de suavização e de previsão para ambos os tipos de modelo.

Quadro 1 – Equações de suavização e previsão para os modelos de Holt-Winters

Holt-Winters	Modelo Aditivo	Modelo Multiplicativo
Nível	$\hat{N}_t = \alpha(Y_t - \hat{S}_{t-s}) + (1 - \alpha)(\hat{N}_{t-1} + \hat{T}_{t-1})$	$\hat{N}_t = \alpha(Y_t/\hat{S}_{t-s}) + (1 - \alpha)(\hat{N}_{t-1} + \hat{T}_{t-1})$
Tendência	$\hat{T}_t = \beta(\hat{N}_t - \hat{N}_{t-1}) + (1 - \beta)\hat{T}_{t-1}$	$\hat{T}_t = \beta(\hat{N}_t - \hat{N}_{t-1}) + (1 - \beta)\hat{T}_{t-1}$
Sazonalidade	$\hat{S}_t = \gamma(Y_t - \hat{N}_t) + (1 - \gamma)\hat{S}_{t-s}$	$\hat{S}_t = \gamma(Y_t/\hat{N}_t) + (1 - \gamma)\hat{S}_{t-s}$
Previsão	$\hat{Y}_{t+h} = \hat{N}_t + h\hat{T}_t + \hat{S}_{t+h-s}$	$\hat{Y}_{t+h} = (\hat{N}_t + h\hat{T}_t)\hat{S}_{t+h-s}$

em que  $h$  é o horizonte da previsão e  $s$  é o número de vezes que a série é observada por ano ( $s = 12$ ).

## 2.4 Modelo SARIMA

O modelo SARIMA é utilizado para dados com tendência e sazonalidade, sejam essas componentes determinísticas e/ou estocásticas. Além dos operadores autoregressivos e de médias móveis de ordem “ $p$ ” e “ $q$ ” do modelo ARIMA [BOX e JENKINS (1976), BOX, JENKINS e REINSEL (2008) e MONTGOMERY, JENNINGS e KULAHCI (2008)], este modelo consta com operadores autoregressivos e de médias móveis sazonais de ordem “ $P$ ” e “ $Q$ ”, conforme a equações de (8) a (12).

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (8)$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (9)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (10)$$

$$\Phi_P(B^s) = 1 - \phi_s B^s - \phi_{2s} B^{2s} - \dots - \phi_{ps} B^{ps} \quad (11)$$

$$\Theta_Q(B^s) = 1 - \theta_s B^s - \theta_{2s} B^{2s} - \dots - \theta_{qs} B^{qs} \quad (12)$$

em que  $\varepsilon_t$  são erros ruído branco,  $\phi_p(B)$  e  $\theta_q(B)$  são, respectivamente, os operadores auto-regressivos e de médias móveis definidos em (9) e (10) e  $\Phi_P(B^s)$  e  $\Theta_Q(B^s)$  são os operadores auto-regressivos e de médias móveis sazonais, conforme em (11) e (12). Os termos  $(1 - B)^d$  e  $(1 - B^s)^D$  representam diferenças de primeira ordem e de ordem “s” aplicadas sob a série original  $Y_t$ , isto é,  $\Delta^d$  e  $\Delta_s^D$ , para torná-la estacionária, dado a presença de tendências determinísticas ou estocásticas, e para remover o efeito sazonal.

A aplicação do modelo SARIMA consiste em três etapas usuais: identificação dos parâmetros do modelo, estimação e diagnóstico.

Com a série estacionária, a identificação dos parâmetros ocorre via funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP). No entanto, Makridakis, Wheelwright e Hyndman (2008) afirmaram que obter a ordem exata dos parâmetros apenas por estes gráficos é uma tarefa complexa e subjetiva, devido ao grande número de modelos que podem ser ajustados, devido a uma mesma estrutura de covariância. Por este motivo, essa etapa foi conduzida simulando diversos modelos SARIMA e selecionando o modelo mais parcimonioso via Critérios de Informação Akaike (1974) e de Schwarz (1978), descritos, respectivamente, em (13) e (14).

$$AIC = -2 \log(L) + 2v \quad (13)$$

$$BIC = -2 \log(L) + v \ln(n) \quad (14)$$

em que  $L$  é o máximo da função de verossimilhança,  $v$  é o número de parâmetros do modelo e  $n$  o número de observações da série.

A estimação dos parâmetros ocorreu por máxima verossimilhança exata, através do algoritmo proposto por Gardner e Phillips (1980), que utiliza o filtro de Kalman (HARVEY e PHILLIPS, 1976) para obter os coeficientes e suas respectivas variâncias assintóticas, satisfazendo as condições de invertibilidade e, conseqüentemente, de unicidade do modelo (PRIESTLEY, 1981; SHUMWAY e STOFFER, 2010).

O diagnóstico consistiu em verificar se os resíduos do modelo são ruído branco. Para avaliar a presença de autocorrelação utilizou-se o teste proposto por Ljung e Box (1978), cuja hipótese nula é de que os resíduos não são autocorrelacionados. Sua estatística de teste em (15) possui distribuição assintótica de Qui-quadrado com  $k - v$  graus de liberdade.

$$Q = n(n + 2) \sum_{i=1}^k \frac{r_k^2}{(n - k)} \sim \chi^2_{k-v} \quad (15)$$

em que  $k$  é o número de defasagens utilizadas no teste ( $k = 20$ ),  $n$  é o número de observações da série,  $v$  o número de parâmetros do modelo ajustado e  $r_k$  é valor estimado para o  $k$ -ésimo coeficiente de autocorrelação.

O teste (15) foi também aplicado à série dos resíduos quadráticos, para testar sua homogeneidade de variância. Utilizou-se o histograma de probabilidade, o gráfico de probabilidade ou QQ-plot e o formal teste de normalidade de Shapiro-Wilk (SHAPIRO e WILK, 1965), no intuito de verificar a normalidade dos resíduos. Destaca-se que esses mesmos procedimentos foram realizados para os resíduos do modelo de Holt-Winters.

Com o modelo SARIMA, os valores previstos para o horizonte  $h$  foram estimados pela esperança condicional apresentada em (16), que minimiza o Erro Quadrático Médio da previsão (MORETTIN e TOLOI, 2006).

$$\hat{Y}_{t+h} = E[Y_{t+h} | Y_t, Y_{t-1}, \dots, Y_2, Y_1] \quad (16)$$

## 2.5 Indicadores de erro

Para avaliar a qualidade do ajuste aos dados históricos, foram calculados os indicadores de Erro Quadrático Médio (*Mean Squared Error* ou MSE) e Erro Percentual Médio Absoluto (*Mean Absolute Percentual Error* ou MAPE), conforme apresentado, respectivamente, em (17) e (18) (TSAY, 2005).

$$MSE = \frac{1}{h} \sum_t^{t+h} (Y_t - \hat{Y}_t)^2 \quad (17)$$

$$MAPE = \frac{1}{h} \sum_t^{t+h} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| 100\% \quad (18)$$

## 2.6 Previsão

Após as etapas de ajuste e diagnóstico, selecionou-se como modelo final aquele que obteve os menores valores para os indicadores de erro apresentados em (17) e (18). Posteriormente, este modelo foi utilizado para fornecer previsões de temperatura máxima mensal para a cidade do Rio de Janeiro (RJ), referentes aos 12 meses do ano de 2016. Adicionalmente, empregou-se o teste proposto por Diebold e Mariano (2012), que tem por finalidade comparar a acurácia das previsões obtidas por dois modelos ou métodos de previsão. Neste sentido, comparou-se a acurácia da previsão obtida pelo modelo selecionado como final com a do modelo concorrente, visando verificar se existe superioridade estatística do primeiro. Os autores agradecem aos revisores anônimos por essa valiosa contribuição.

## 2.7 Aspectos computacionais

Utilizou-se o *software* R (R CORE TEAM, 2016) para a análise de dados, desde a confecção de gráficos e condução de testes de hipóteses, ao ajuste e diagnóstico dos modelos de Holt-Winters e Box e Jenkins. Dentre os principais pacotes utilizados destacam-se o “*forecast*”, “*timeseries*”, “*urca*” e “*TSA*”. Detalhes teóricos e procedimentos computacionais sobre a análise de séries temporais no *software* R podem ser consultados em Cryer e Chan (2008), Pfaff (2008) e Shumway e Stoffer (2010).

## 3 Resultados e discussão

A Figura 1 apresenta o gráfico da série de temperatura máxima mensal do Rio de Janeiro (RJ). A série oscila em torno de um valor médio constante e parece não apresentar tendência. Por se tratar de uma variável climatológica, há um padrão periódico bem definido, como verificado no boxplot da Figura 2, com temperaturas mais baixas nos meses de Maio, Junho, Julho e Agosto e mais elevadas nos meses de Dezembro, Janeiro, Fevereiro e Março. Apenas duas observações atípicas foram detectadas, uma no mês de Outubro de 2005 e outra no mês de Junho de 2012. A análise descritiva indicou que a menor temperatura máxima mensal foi de 31,2 °C (Julho de 2014) e a maior de 41,8 °C (Fevereiro de 2010), com média de 36,70 °C e desvio-padrão de 2,60 °C. O baixo coeficiente de variação,  $CV = 7,09\%$ , corrobora a ideia de que a série não apresenta tendência, devido à baixa variabilidade dos dados em torno da média.

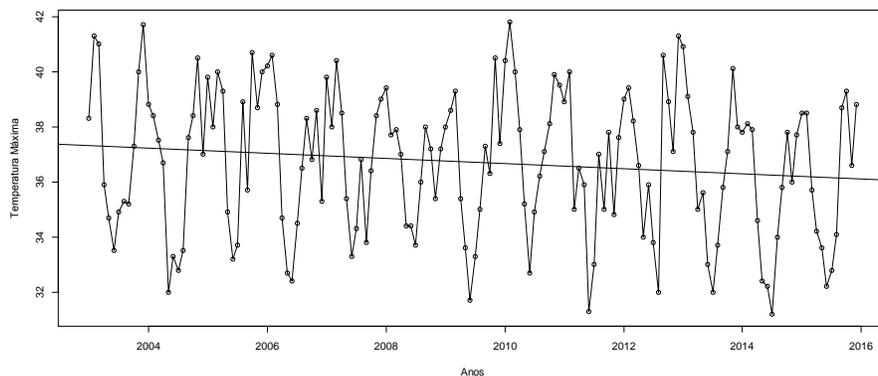


Figura 1 - Gráfico da Série de temperatura máxima mensal do Rio de Janeiro (Janeiro-2003 a Dezembro-2015).

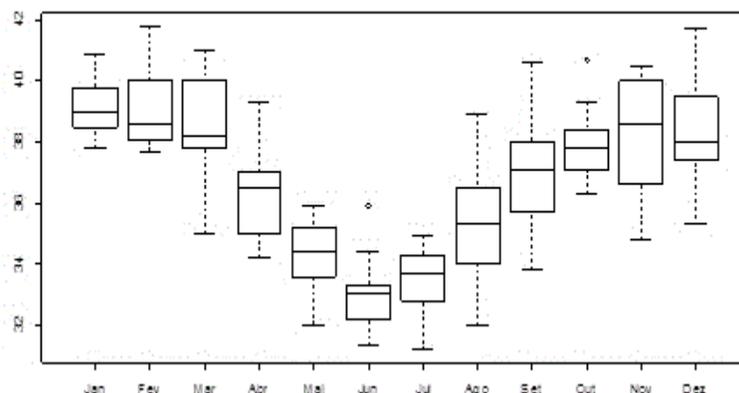


Figura 2 - Boxplots mensais da série de temperatura máxima do Rio de Janeiro (Janeiro-2003 a Dezembro-2015).

Na Tabela 1 estão apresentados os resultados do teste ADF. Verificou-se que a série em estudo é estacionária, pois a hipótese de raiz unitária foi rejeitada a 5% de probabilidade ( $p$ -valor = 0,0051). Adicionalmente, o coeficiente para a tendência determinística linear não foi significativo ( $p$ -valor = 0,1688) e o teste de Mann-Kendall não rejeitou a hipótese de ausência de tendência ( $p$ -valor = 0,1265), o que corroborou as conclusões obtidas a partir da análise gráfica e descritiva da série. Portanto, não foram detectadas tendências determinísticas ou estocásticas. O teste ADF foi realizado com 11 defasagens da variável em primeira diferença, necessárias para remover a autocorrelação residual. O teste Ljung-Box confirma esse resultado.

Tabela 1 - Resultados para o teste de Dickey-Fuller aumentado (ADF)

Parâmetros	$\alpha$	$\beta$	$\delta$
Estimativa	29,7066	- 0,0056	- 0,7985
Erro-Padrão	10,4764	0,0041	0,2803
Estatísticas do Teste	2,8360	- 1,3840	- 2,8490
p-valor	0,0053	0,1688	0,0051
Mann-Kendall	$\tau = -0,0829$	p-valor = 0,1265	
Teste de Ljung-Box	$Q_{20} = 27,88$	p-valor = 0,1122	

\* Significativo a 0,05 de probabilidade

Como mencionado anteriormente, por ser tratar de uma variável climatológica, há evidências de que a série em estudo apresente sazonalidade. Para analisar a periodicidade dessa componente, a Figura 2 apresenta o Periodograma da análise espectral.

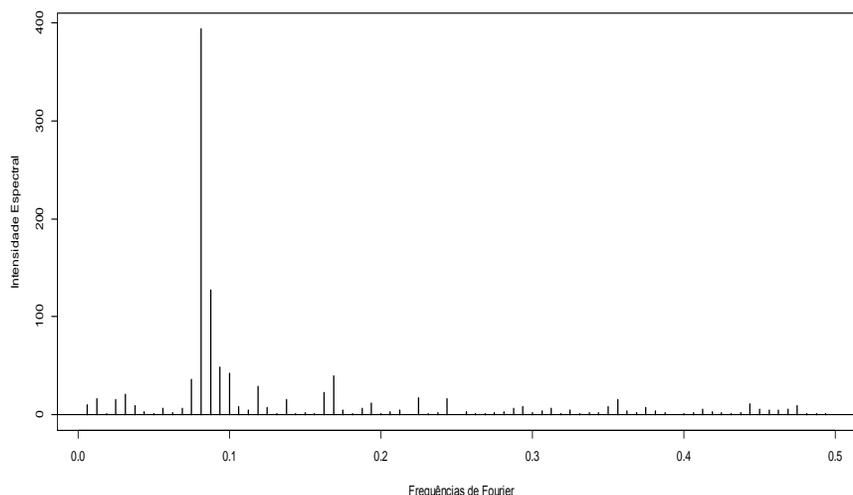


Figura 3 - Periodograma da série de temperatura máxima mensal do Rio de Janeiro (Janeiro-2003 a Dezembro-2015).

A maior intensidade espectral (pico) foi associada à frequência de Fourier de 0,08125, que sugere uma componente sazonal com periodicidade  $s = 12,3076 \approx 12$  meses. Pelo teste  $G$  de Fisher concluiu-se sobre sua significância estatística, pois o valor da estatística  $G = 0,3652$  foi superior ao da estatística  $Z = 0,0910$ , com um p-valor  $\approx 0$ . Logo, a série de temperatura máxima mensal do Rio de Janeiro apresenta sazonalidade anual, isto é, de 12 em 12 meses.

Conclusões similares, sobre a ausência de tendências e o comportamento sazonal com periodicidade anual, foram também obtidas por Stein e Lioret (2001), Silva, Guimarães e Tavares (2008) e Chechi e Bayer (2012), em trabalhos envolvendo a temperatura do ar de outras localidades.

### 3.1 Ajuste do modelo de Holt-Winters

Pelo gráfico da série (Figura 1), não há evidências de que a amplitude sazonal varia com o tempo, isto é, a diferença entre o maior e o menor valor de temperatura, associados a cada um dos 13 anos em estudo, parece ser constante. A Tabela 2 apresenta a regressão linear simples da amplitude sazonal pelo tempo em anos.

Tabela 2 - Regressão linear simples Amplitude Anual  $\times$  Tempo (Anos)

Parâmetros	Estimativa	Erro Padrão	t	p-valor
Intercepto	7,9653	0,6526	12,21	0,0000
Tempo (anos)	0,0049	0,0822	0,06	0,9530

Como a variável Tempo não foi significativa a 5% de probabilidade (p-valor = 0,9530), a amplitude sazonal não variou com o tempo. Portanto, optou-se por ajustar o modelo aditivo de Holt-Winters. Em outros estudos de temperatura do ar, como os de Kamruzzaman, Beecham e Metcalfe (2011) e Chechi e Bayer (2012), verificou-se também esta característica do comportamento sazonal e o modelo aditivo foi adotado, o que corrobora essa escolha. A Tabela 3 apresenta a estimativa das constantes de suavização e os indicadores MSE e MAPE desse modelo. Na Figura 4 os gráficos dos seus resíduos.

Tabela 3 - Resultados para o modelo aditivo de Holt-Winters

Holt-Winters	$\alpha$	$\beta$	$\gamma$	MSE	MAPE (%)
Aditivo	0,1299	0,0191	0,2985	3,0284	3,5091

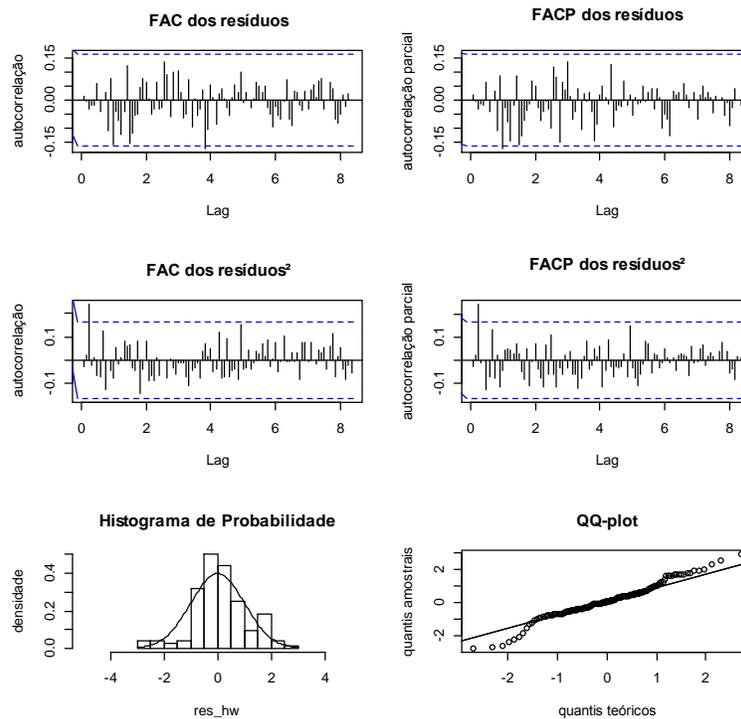


Figura 4 - Análise dos resíduos padronizados do modelo multiplicativo de Holt-Winters.

Na Tabela 3, a constante de suavização estimada para o efeito de tendência na série foi próxima de zero ( $\alpha = 0,0191$ ), devido à ausência dessa componente. Já a constante do efeito sazonal ( $\gamma = 0,2985$ ) obteve o maior valor estimado, devido à periodicidade anual bem definida. Pelos gráficos da FAC e da FACP, os resíduos do modelo aditivo não são autocorrelacionados, pois poucos coeficientes de autocorrelação e de autocorrelação parcial ultrapassaram os limites de confiança estabelecidos a 5% de probabilidade. O teste Ljung-Box corroborou essa conclusão,  $Q_{20} = 20,34$  e p-valor = 0,43. A homogeneidade de variância dos resíduos também foi constatada por este teste,  $Q_{20} = 19,90$  e p-valor = 0,46. O histograma de probabilidades e o QQ-plot indicam que os resíduos não se ajustaram a distribuição normal, possivelmente devido a valores discrepantes, que tornaram as caudas dessa distribuição mais densas. Como esperado, o teste de Shapiro-Wilk rejeitou a hipótese de normalidade, apresentando estatística  $W = 0,97$  e p-valor = 0,0047. Portanto, os resíduos do modelo aditivo de Holt-Winters são ruído branco, mas não são gaussianos.

### 3.2 Ajuste do modelo SARIMA

Para aplicação da metodologia Box & Jenkins, não foi preciso aplicar diferenças de primeira ordem à série original, devido à ausência de tendências determinísticas ou estocásticas significativas (Tabela 1). Para eliminar o efeito sazonal aplicou-se uma diferença de ordem  $s = 12$ . A Figura 5 ilustra o gráfico, a FAC e a FACP da série original e da série diferenciada.

A aplicação da diferença minimizou relativamente a estrutura de autocorrelação dos dados, no entanto, em defasagens múltiplas e próximas de  $s = 12$ , alguns coeficientes de autocorrelação e de autocorrelação parcial ainda foram significativos a 5%. Portanto, o dentre a ampla classe de modelos propostos por Box & Jenkins, o SARIMA é o mais indicado.

Foram ajustados diversos modelos  $SARIMA(p, 0, q) \times (P, 1, Q)_{12}$ , com os valores de “ $p$ ” e “ $q$ ” variando de 0 a 3, e “ $P$ ” e “ $Q$ ” de 0 a 2. Após várias tentativas, o modelo selecionado pelos critérios de informação de AIC e BIC foi o SARIMA  $(1,0,1) \times (1,1,1)_{12}$ , que obteve todos os coeficientes significativos a 5%. Tais resultados e os indicadores de MSE e MAPE estão apresentados na Tabela 4.

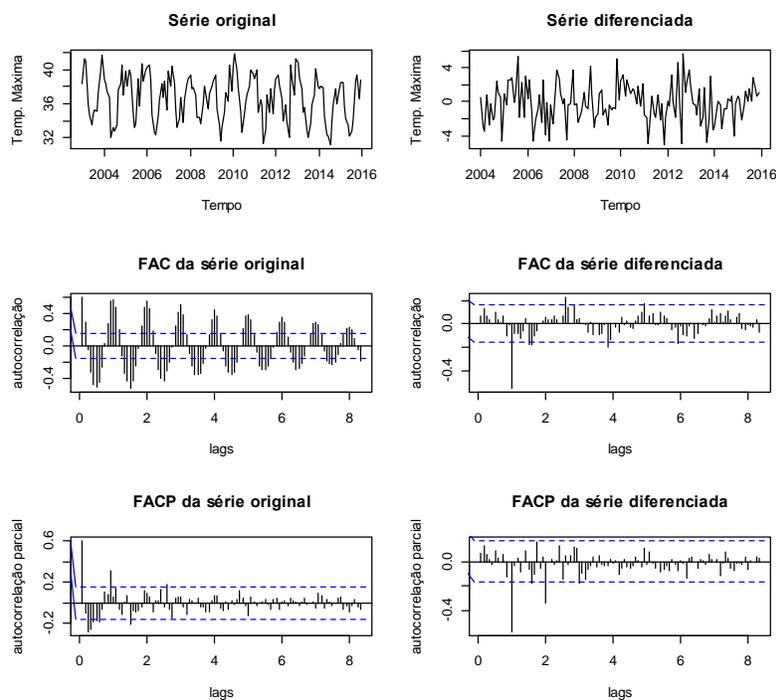


Figura 5 - Gráfico, FAC e FACP da série original e da série após diferença sazonal.

Tabela 4 - Resultados para o modelo  $SARIMA(1,0,1) \times (1,1,1)_{12}$

Parâmetros	Coefficientes	Erro-Padrão	t	p-valor
$\phi_1$	0,8949	0,1156	7,7351	0,0000*
$\theta_1$	-0,8073	0,1441	-5,5993	0,0000*
$\Phi_1$	-0,2465	0,0999	-2,4671	0,0136*
$\Theta_1$	-0,7502	0,0929	-8,0730	0,0000*
$-2 \log(L) = 551,11$		AIC = 559,11	BIC = 571,31	
MSE = 2,2457		MAPE = 3,0861%		

\* Significativo a 0,05 de probabilidade

Os indicadores MSE e MAPE do modelo SARIMA foram menores do que os do modelo aditivo de Holt-Winters (MSE = 3,0284 e MAPE = 3,5091%), o que indica que o mesmo se ajustou de forma mais precisa aos dados analisados. Tal resultado pode estar associado à característica estocástica da componente sazonal, não eliminada completamente pela diferença aplicada, como verificado nos gráficos da FAC e da FACP (Figura 5). No entanto, o modelo SARIMA conseguiu captar essa dinâmica com seus termos auto-regressivos e de médias móveis sazonais, que o modelo de Holt-Winters,

apenas com algoritmos de suavização, não conseguiu (BOX, JENKINS e REINSEL, 2008). Chatfield e Yar (2012) corrobora afirmando que os modelos de Holt-Winters, embora precisos, simples de estimar e de baixo custo computacional, possuem como desvantagem a ausência de conceitos probabilísticos em seu desenvolvimento. O melhor desempenho do modelo SARIMA em relação aos modelos de Holt-Winters, para estudos com variáveis climatológicas, foi também verificado por Chechi e Bayer (2012), na previsão de temperatura média do ar da cidade de Erechim (RS). Resultados similares também foram verificados por Newaz (2008) e Assis *et al.* (2010), no entanto, para previsão de taxa de câmbio e preços de feijão e de cacau, respectivamente.

A Figura 6 apresenta os gráficos da FAC e FACP dos resíduos, FAC e FACP dos resíduos quadráticos, o histograma de probabilidade e o QQ-Plot para os resíduos do SARIMA  $(1,0,1) \times (1,1,1)_{12}$ .

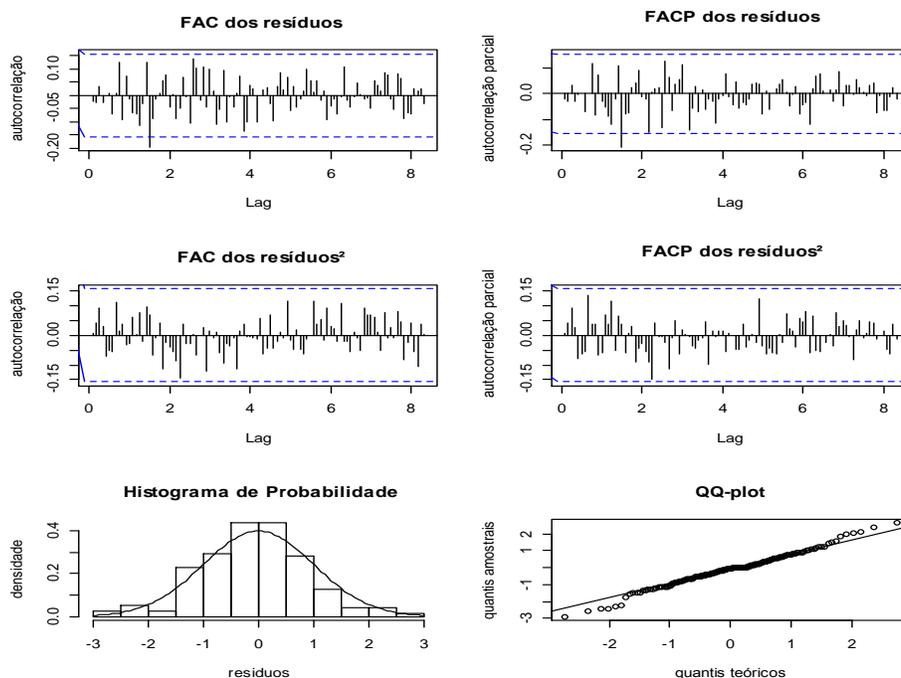


Figura 6 - Comprovação das características de ruído branco.

Pelos gráficos da FAC e da FACP, poucos coeficientes de autocorrelação e de autocorrelação parcial da série de resíduos e de resíduos quadráticos foram estatisticamente significativos ao nível de 5% de probabilidade, o que indica a ausência de autocorrelação ( $Q_{20} = 21,55$  e  $p\text{-valor} = 0,36$ ) e homogeneidade de variâncias ( $Q_{20} = 11,17$  e  $p\text{-valor} = 0,94$ ). O histograma de probabilidade e o QQ-plot demonstram que os resíduos se ajustaram a distribuição normal, o que foi confirmado pelo teste Shapiro-Wilk,

que apresentou estatística  $W = 0,98$  e  $p\text{-valor} = 0,20$ . Este é um resultado importante, pois sob normalidade, a ausência de autocorrelação implica em independência, portanto, os resíduos do modelo SARIMA são ruído branco e independentes (PFAFF, 2008). Adicionalmente, os intervalos de confiança para as previsões são mais precisos, pois não possuem sua cobertura afetada, como ocorre se tal condição não for verificada (GUAN *et al.*, 2010).

### 3.3 Previsão

Adotou-se o SARIMA  $(1,0,1) \times (1,1,1)_{12}$  como o modelo de previsão, devido sua maior eficiência e melhores características residuais. A Figura 7 ilustra o seu ajuste aos dados históricos de temperatura máxima mensal da cidade do Rio de Janeiro (RJ) e as previsões fornecidas. A linha vermelha indica os valores previstos pelo modelo, a linha preta os valores reais e as linhas em azul indicam, respectivamente, o limite inferior e o limite superior do intervalo com 95% confiança dos valores previstos para os 12 meses do ano de 2016.

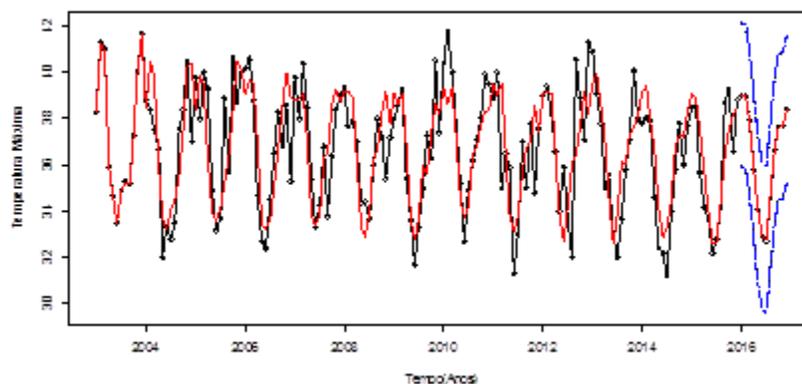


Figura 7 - Ajuste e Previsão com o modelo SARIMA  $(1,0,1) \times (1,1,1)_{12}$

O modelo conseguiu captar a dinâmica temporal da série, detectando principalmente o padrão periódico. Apenas nos anos de 2010 e 2013, em meses mais quentes ou próximos do verão, os valores ajustados foram subestimados em relação aos valores reais.

Na Tabela 5 estão indicados os valores previstos de temperatura máxima mensal para os 12 meses do ano de 2016 e os valores para os respectivos intervalos com 95% de confiança.

Como esperado, as previsões mais elevadas referem-se aos meses de Janeiro (Férias Escolares), Fevereiro (Carnaval) e Dezembro (Natal e Réveillon). Os intervalos demonstram, com 95% de confiança, o quanto essas estimativas podem variar. Portanto, os turistas e a população do Rio de Janeiro podem se prevenir e, conseqüentemente, aproveitar melhor o período de férias e festas. Adicionalmente, nos períodos em que a olimpíada (Agosto) e a paraolímpico (Setembro) serão realizadas, essas informações podem também ser utilizadas.

Embora o modelo SARIMA (1,0,1) x (1,1,1)<sub>12</sub> tenha se ajustado melhor aos dados históricos e por isso tenha sido selecionado como modelo final, o teste estatístico de Diebold e Mariano não rejeitou a hipótese nula (DM = 0, p-valor = 0,5) de que os dois modelos possuem a mesma acurácia de previsão. Portanto, as previsões obtidas pelo modelo aditivo de Holt-Winter para o ano de 2016 são tão acuradas quanto às obtidas pelo modelo SARIMA e também poderiam ser adotadas. A Figura 8 corrobora tal conclusão.

Tabela 5 - Estimativas e Intervalos de Confiança para as previsões de Temperatura máxima mensal em 2016

Mês	Limite Inferior	Previsto	Limite Superior
Janeiro	35,9016	39,0215	42,1414
Fevereiro	35,8289	38,9607	42,0926
Março	34,7980	37,9393	41,0807
Abril	32,6375	35,7865	38,9355
Mai	30,9380	34,0931	37,2482
Junho	29,7490	32,9090	36,0689
Julho	29,5655	32,7293	35,8931
Agosto	31,3706	34,5375	37,7045
Setembro	33,4909	36,6604	39,8298
Outubro	34,4995	37,6709	40,8423
Novembro	34,5135	37,6865	40,8595
Dezembro	35,2205	38,3948	41,5691

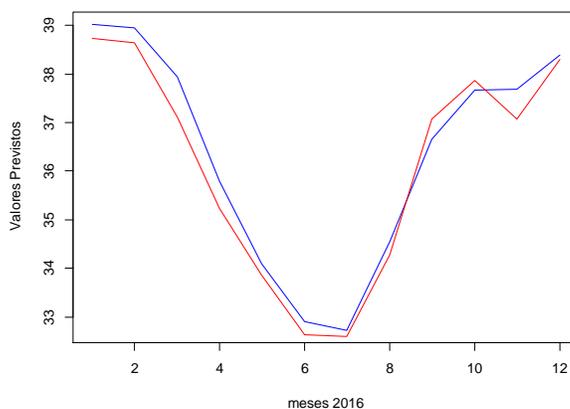


Figura 8 - Previsão para os meses de 2016 com o modelo SARIMA (linha azul) e Holt-Winters aditivo (linha vermelha).

## 4 Conclusão

A série de temperatura máxima mensal do Rio de Janeiro (RJ) não apresentou tendência e sim uma componente sazonal com periodicidade de 12 meses. Foram comparados o modelo aditivo de Holt-Winters e o modelo  $SARIMA (1,0,1) \times (1,1,1)_{12}$ , sendo que o segundo se ajustou de forma mais precisa aos dados históricos e apresentou menores valores para os indicadores MSE e MAPE, além de resíduos ruído branco, normais e independentes.

O modelo  $SARIMA (1,0,1) \times (1,1,1)_{12}$  forneceu previsões de temperatura máxima para os 12 meses de 2016. Adicionalmente, os intervalos de confiança com 95% de probabilidade foram apresentados. Com estes resultados, informações podem ser disponibilizadas à população do Rio e aos turistas, para que possam se preparar para o clima da cidade, minimizando riscos e aproveitando suas várias atrações, principalmente, no período de férias, carnaval e réveillon.

A abordagem proposta mostrou-se adequada para a modelagem de dados de temperatura máxima mensal. Para futuros trabalhos sugere-se a comparação com Redes Neurais Artificiais e modelos SARIMA com intervenções, para estudar os *outliers* apresentados no boxplot (Figura 2) e avaliar a melhoria das previsões, no que se refere à superioridade estatística e acurácia.

EDUARDO, C. B.; SILVA, C. H. O.; NASCIMENTO, M.; LIMA, J. E. Analysis and forecast for monthly maximum temperature of Rio de Janeiro city for the year 2016. *Rev. Bras. Biom.* Lavras, v.35, n.1, p.174-193, 2017.

- **ABSTRACT:** *In this paper we have compared the fit of additive models of Holt-Winters and SARIMA in time series of monthly maximum temperature in the city of Rio de Janeiro (RJ) using data from January 2003 to December 2015. The Holt-Winters additive model was chosen because the seasonal behavior of the time series. The SARIMA (1,0,1)×(1,1,1)<sub>12</sub> was chosen because provided the best Akaike Information and Schwarz Criteria. The final model was chosen by using mean square error and mean absolute percentual error. These measures were calculated using the fitted model to the historical data. Additionally, a Diebold-Mariano test for predictive accuracy was applied and white noise characteristics of the residues were evaluated. The SARIMA (1,0,1)×(1,1,1)<sub>12</sub> was the best model fitted and so it was adopted for maximum monthly temperature forecast for the year 2016.*
- **KEYWORDS:** *Holt-Winters; SARIMA; seasonality; temperature.*

## Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v.19, n.6, p.716-723, 1974.
- ASSIS, K.; AMRAN, A.; REMALI, Y.; AFFENDY, H.; ASSIS, K. A comparison of univariate time series methods for forecasting cocoa bean prices. *Trends in Agricultural Economics*, 2010.
- BACK, A. J. Application of statistical analysis for identification of climatic tendency. *Pesquisa Agropecuária Brasileira*, v.36, n.5, p.717-726, 2001.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. *Time series analysis: forecasting and control*. John Wiley & Sons, 2008.

- Box, G. E. P.; Jenkins, G. M. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1976.
- BROCKWELL, P. J.; RICHARD A, D. *Time series: Theory and methods*. Springer Series in Statistics. Springer-Verlag, 1991.
- BUENO, J. C. C. Polos de desenvolvimento do Estado do Rio de Janeiro. *Revista de Estudos e Pesquisas*, n.180, 2007.
- CHATFIELD, C.; YAR, M. Holt-Winters forecasting: some practical issues. *The Statistician*, p.129-140, 1988.
- CHECI, L.; BAYER, F. M. Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.16, n.12, p.1321-1329, 2012.
- CHO, V. A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, v.24, n.3, p.323-330, 2002.
- Cryer, J. D.; Chan, K. S. *Time series analysis: With R applications*. Springer, 2008.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, v. 74, n. 366a, p. 427-431, 1979.
- DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, v.20, n.1, p.134-144, 2012.
- DOOLEY, G.; LENIHAN, H. An assessment of time series methods in metal price forecasting. *Resources Policy*, v.30, n.3, p.208-217, 2005.
- FISHER, Ronald Aylmer. Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, v.125, n.796, p.54-59, 1929.
- GARDNER, G.; HARVEY, A. C.; PHILLIPS, G. D. A. Algorithm AS 154: An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v.29, n.3, p.311-322, 1980.
- Granger, C. W. J.; Newbold, P. *Forecasting economic time series*. Economic Theory and Mathematical Economics (USA), 1977.
- GUAN, C.; LUH, P. B.; MICHEL, L. D.; COOLBETH, M.; FRIEDLAND, P. B. (2010). Hybrid Kalman algorithms for very short-term load forecasting and confidence interval estimation. In: *Power and Energy Society General Meeting, 2010 IEEE*. IEEE, 2010. p. 1-8.
- CO, H. C.; BOOSARAWONGSE, R. Forecasting Thailand's rice export: Statistical techniques vs. artificial neural networks. *Computers & industrial engineering*, v.53, n.4, p.610-627, 2007.
- HOLT, C. C. Forecasting trends and seasonals by exponentially weighted moving averages. *Management Science*, v.6, p.324-342, 1957.

- HUBBARD, K. G. Agricultural climatology. *Journal of Service Climatology*, v.1, p.1-9, 2007.
- HUDDLESTON, S. H.; PORTER, J. H.; BROWN, D. E. Improving forecasts for noisy geographic time series. *Journal of Business Research*, v.68, n.8, p.1810-1818, 2015.
- KAMRUZZAMAN, M.; BEECHAM, S.; METCALFE, A. V. Non-stationarity in rainfall and temperature in the Murray Darling Basin. *Hydrological Processes*, v.25, n.10, p.1659-1675, 2011.
- KARNER, O. ARIMA representation for daily solar irradiance and surface air temperature time series. *Journal of Atmospheric and Solar-Terrestrial Physics*, v.71, n.8, p.841-847, 2009.
- KENDALL, M. G. *Rank correlation methods*. London: Charless Griffin, 1948.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack in time series models. *Biometrika*, London, v.65, p.297-303, 1978.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. *Forecasting methods and applications*. John Wiley & Sons, 2008.
- MANN, H. B. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, v.13, p.245-259, 1945.
- MEDEIROS, S. D. S.; CECÍLIO, R. A.; MELO JÚNIOR, J. D.; SILVA JUNIOR, J. D. 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.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2008.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. 2.ed. São Paulo: Edgard Blücher, 2006.
- NAJAFI N. R., TARAZKAR M. Forecasting of Iran pistachio export: application of artificial neural network. *Journal of Trading Bulletin*, v.29, p.191-214, 2006.
- NEWAZ, M. K. Comparing the performance of time series models for forecasting exchange rate, *BRAC University Journal*, v.2, p.55-65, 2008.
- PFÄFF, B. *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media, 2008.
- PRIESTLEY, M. B. *Spectral Analysis and Time Series*. Academic Press, London, 1981.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning internal representations by backpropagation errors. *Nature*, London, v.323, p.533-536, 1985.
- SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v.6, n.2, p.461-464, 1978.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, p.591-611, 1965.

SHUMWAY, R. H.; STOFFER, D. S. *Time series analysis and its applications with R examples*. Springer Science & Business Media, 2010.

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.

SONALI, P.; KUMAR, D. N. Review of trend detection methods and their application to detect temperature changes in India. *Journal of Hydrology*, v.476, p.212-227, 2013.

STEIN, M.; LLORET, J. Forecasting of air and water temperatures for fishery purposes with selected examples from Northwest Atlantic. *Journal of Northwest Atlantic Fishery Science*, v.29, p.23-30, 2001.

TSAY, R. S. *Analysis of financial time series*. John Wiley & Sons, 2005.

TULARAM, G. A.; ILAHEE, M. Time series analysis of rainfall and temperature interactions in coastal catchments. *Journal of Mathematics and Statistics*, v.6, n.3, p.372-380, 2010.

UNESCO. *Dossiê: Carioca Landscapes between the Mountain and the Sea*. 2012

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. *Management Science*, v.6, n.3, p.324-342, 1960.

Recebido em 29.03.2016

Aprovado após revisão em 15.07.2016