

## METODOLOGIA BOX & JENKINS PARA PREVISÃO DE TEMPERATURA MÉDIA MENSAL DA CIDADE DE BAURU (SP)

Eduardo Campana BARBOSA<sup>1</sup>  
TheIma SÁFADI<sup>2</sup>  
Moysés NASCIMENTO<sup>1</sup>  
Ana Carolina Campana NASCIMENTO<sup>1</sup>  
Carlos Henrique Osório SILVA<sup>1</sup>  
Rômulo César MANULI<sup>1</sup>

**RESUMO:** Neste trabalho aplicou-se a metodologia Box & Jenkins para obter um modelo estatístico que permitisse prever os valores futuros de temperatura média mensal da cidade de Bauru (SP). Foram utilizados dados de temperatura média de doze anos (Janeiro de 2001 a Dezembro de 2012) e por meio do Critério de Informação de Akaike (AIC) selecionou-se 6 modelos SARIMA candidatos. A escolha do modelo de previsão ocorreu segundo o indicador de Erro Quadrático Médio (EQM), calculado através da diferença quadrática entre os valores reais e previstos para o ano de 2013. Além disso, avaliou-se a significância estatística dos coeficientes do modelo e se os resíduos estimados possuíam características de um ruído branco. O modelo o SARIMA (0,0,1) x (2,1,1)<sub>12</sub> foi escolhido pois apresentou previsões mais acuradas e resultados estatísticos satisfatórios.

**PALAVRAS-CHAVE:** Modelagem; estatística; sazonalidade; SARIMA.

### 1 Introdução

Uma série temporal é definida como um conjunto de observações ordenadas no tempo, descritas por uma função matemática (série determinística) ou por distribuições de probabilidade (processo estocástico), cuja principal particularidade é a relação de dependência ou autocorrelação serial entre observações vizinhas. Logo, a análise de séries temporais é uma metodologia empregada no estudo de dados com essa natureza, pois permite a modelagem desta estrutura de autocorrelação e consequentemente o desenvolvimento de previsões futuras (Morettin & Tolo, 2006).

Neste contexto, os modelos estatísticos utilizados buscam captar o mecanismo gerador da série via dados históricos, coletados ao longo do tempo e referentes a uma variável de interesse. A adequabilidade destes é verificada em função da capacidade de ajustar as principais características de uma série temporal, denominadas como tendência e

---

1 Universidade Federal de Viçosa - UFV, Departamento de Estatística, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: [duducampana@hotmail.com](mailto:duducampana@hotmail.com); [moysesnascim@ufv.br](mailto:moysesnascim@ufv.br); [ana.oliveira8@ufv.br](mailto:ana.oliveira8@ufv.br); [chos@ufv.br](mailto:chos@ufv.br); [romulomanuli@ig.com.br](mailto:romulomanuli@ig.com.br).

2 Universidade Federal de Lavras - UFLA, Departamento de Ciências Exatas - DEX, CEP: 37200-000, Lavras, MG, Brasil. E-mail: [safadi@dex.ufl.br](mailto:safadi@dex.ufl.br).

sazonalidade (Morettin & Toloí, 2006). Os modelos ARIMA (*Auto Regressive Integrated Moving Average*) e suas parametrizações (AR, MA, ARMA, SARIMA etc.), propostos por Box & Jenkins (1976), são frequentemente empregados em estudos deste tipo. Destaca-se que a análise de séries temporais pode ser aplicada em diversas áreas do conhecimento, devido às variáveis estudadas, quase sempre, estarem relacionadas ao domínio do tempo.

Werner & Ribeiro (2003) utilizaram a metodologia Box & Jenkins para obter previsões do número de atendimentos a clientes que solicitam três diferentes tipos de serviços de assistência técnica em equipamentos de informática, em uma empresa de Porto Alegre (RS). Rezende et al. (2005) empregaram os modelos SARIMA para análise e previsão das séries temporais de preços do carvão vegetal para quatro regiões do estado de Minas Gerais (MG). Sáfadi (2004) analisou o comportamento da série de vazão de água na barragem de Furnas durante o período de 1963 a 1994, por meio de modelos SARIMA com intervenção.

Segundo Silva, Guimarães & Tavares (2008), a análise de séries temporais em dados climatológicos é de fundamental importância, visto que o descontrole ou desconhecimento de valores futuros para algumas variáveis relacionadas ao clima pode influenciar direta e negativamente a vida em sociedade, seja nas atividades e no lazer das pessoas, no planejamento das empresas (segmento industrial) e no gerenciamento da agricultura (setor rural). A importância deste controle climatológico torna-se evidente ao analisarmos os estudos de Medeiros et al. (2005) e Hubbard (2007), que demonstram como a temperatura do ar pode influenciar as atividades relacionadas a agricultura e pecuária, bem como os efeitos fisiológicos em plantas e animais. Chechi & Bayer (2012) aplicaram e compararam as previsões obtidas por modelos SARIMA e Holt-Winters para séries de temperatura mínimas e máximas da cidade de Erechim (RS). Silva, Guimarães e Tavares (2008) apresentaram um estudo semelhante, porém, utilizando apenas os modelos SARIMA em dados de temperatura média da cidade de Uberlândia (MG).

Neste sentido, o presente trabalho teve por objetivo o estudo da série histórica de temperatura média mensal da cidade de Bauru (SP) e a obtenção de um modelo de previsão, via metodologia Box & Jenkins, que possa ser utilizado para fornecer estimativas futuras de temperatura média. O município de Bauru está localizado no noroeste do estado de São Paulo e segundo o censo de 2010, realizado pelo IBGE, o número de habitantes que residiam na cidade era de 344.039. A classificação Köppen-Geiger indica que o clima da região é tropical do tipo Aw, com inverno seco e verão quente e úmido. Em Bauru existem diversas Universidades e Faculdades, dentre estas a Universidade Estadual Paulista (UNESP), Universidade Paulista (UNIP) e Faculdade de Odontologia de Bauru (FOB). Embora o turismo seja um ponto forte, devido às atrações em áreas urbanas e rurais, destaca-se que o interesse científico de pesquisadores com relação à cidade tem aumentado constantemente, devido a gama de instituições de ensino e hospitais que se destinam a pesquisa e atendimento de algumas anomalias. A agricultura baseia-se no cultivo de abacaxi, batata-doce e cana-de-açúcar. O setor industrial também possui grande importância e representatividade na economia da cidade, sendo as industriais de metal-mecânicas e alimentos as com maior relevância.

Com a definição de um modelo de previsão e, conseqüentemente, baseado em seus resultados, seria possível auxiliar a gestão pública da cidade, contribuindo para que as principais atividades nos setores e segmentos anteriormente citados não sejam gravemente

afetadas por uma variação térmica inesperada. Além disso, a população de Bauru é composta por diversos estudantes universitários, fato que requer atenção das autoridades em relação à saúde pública nos períodos de baixas e altas temperaturas. Destaca-se ainda que não foram encontrados na literatura outros trabalhos envolvendo a análise de séries temporais aos dados de temperatura média da cidade de Bauru, portanto, devido à escassez de estudos esta é também uma justificativa para o desenvolvimento do presente trabalho.

## 2 Material e métodos

Os dados utilizados foram obtidos na página do Instituto de Pesquisas Meteorológicas-IPMet da UNESP e referem-se à série histórica de temperatura média mensal, em graus celsius (°C), da cidade de Bauru (SP). As informações foram coletadas na estação meteorológica automática do referido Instituto, que se encontra localizado sob as seguintes coordenadas geográficas: Latitude: 22,35° Sul, Longitude: 49,03° Oeste e Altitude: 620 metros. A obtenção dos valores de temperatura média mensal ocorreu segundo o critério da Organização Mundial de Meteorologia (OMM). Mais especificamente, coletam-se todos os dias medidas de temperatura a cada 30 minutos, o que em 24 horas (um dia) resulta no total de 48 observações. A temperatura média diária é o valor médio destas 48 observações e, conseqüentemente, a temperatura média mensal o valor médio das temperaturas médias diárias. Tal base de dados contém informações dos últimos 13 anos sobre a variável de estudo, totalizando 156 observações mensais, de janeiro de 2001 a dezembro de 2013. O horizonte de previsão contemplou o ano de 2013, portanto, os valores reais dos 12 últimos doze meses não foram incluídos na modelagem e sim reservados para comparações com as previsões.

O procedimento metodológico inicial consistiu em decompor graficamente a série histórica em suas componentes de tendência, sazonalidade e parte aleatória. Realizou-se também uma análise descritiva, apresentando os valores mínimo, médio e máximo de temperatura média, além de algumas medidas que quantificam a dispersão dos dados, tais como o desvio-padrão e coeficiente de variação.

Para identificar a ordem de integração da série temporal, procedeu-se o teste de raiz unitária de Dickey-Fuller Aumentado (*Argumented Dickey-Fuller* ou *ADF*) (Dickey & Fuller, 1979). Esta é uma versão generalizada do teste de Dickey-Fuller, que parte do pressuposto de que o processo gerador dos dados é um autoregressivo de ordem um, AR(1), conforme em (1).

$$Y_t = \rho Y_{t-1} + \varepsilon_t \quad (1)$$

Em que  $Y_t$  e  $Y_{t-1}$  são, respectivamente, os valores da série no tempo  $t$  e  $t-1$ ,  $\rho$  é um parâmetro e  $\varepsilon_t$  são os erros aleatórios (ruído branco), independentes e identicamente distribuídos com média zero e variância constante.

A ideia principal consiste em testar a hipótese de que  $\rho = 1$  (por isso o nome raiz unitária), o que poderia ser verificado através do teste  $t$  de Student após a estimação de  $\hat{\rho}$ , via Mínimos Quadrados Ordinários, por meio da regressão de  $Y_t$  em função de  $Y_{t-1}$ . No entanto, se a hipótese nula for verdadeira, a distribuição de  $\hat{\rho}$  não é uma  $t$  de Student e possui viés negativo (Dickey & Fuller, 1979). Por meio de métodos de simulação, tais autores obtiveram a distribuição exata e os valores críticos do teste quando  $\rho = 1$ , cujas

tabelas podem ser obtidas em Dickey & Fuller (1981). Por razões teóricas, os softwares estatísticos/econométricos realizam o teste para série em primeira diferença, e procede-se a seguinte substituição  $\delta = \rho - 1$ . Logo, testar  $\rho = 1$  é equivalente a testar  $\delta = 0$ .

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + \varepsilon_t \\ \Delta Y_t &= (\rho - 1)Y_{t-1} + \varepsilon_t \\ \Delta Y_t &= \delta Y_{t-1} + \varepsilon_t \\ \Delta Y_t &= \alpha + \beta t + \delta Y_{t-1} + \varepsilon_t \end{aligned} \quad (2)$$

Sem perda de generalidade é possível incorporar ao modelo uma constante de nível e um termo de tendência determinística, conforme em (2).

A não rejeição da hipótese nula implica que a série possui uma raiz unitária e não é estacionária. Neste caso, procedimentos matemáticos (diferenças) devem ser aplicados para estabilizar a média e variância dos dados. No entanto, em algumas séries temporais, a representação do processo estocástico como um AR(1) não é adequada, visto que os resíduos do teste de Dickey-Fuller podem ser autocorrelacionados. Nestes casos sugere-se a incorporação de outras defasagens no modelo, o que caracteriza o teste de Dickey-Fuller Aumentado, equacionado conforme em (3).

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{i=1}^p \gamma_i \Delta Y_{t-1} + \varepsilon_t \quad (3)$$

Embora o teste de Dickey-Fuller Aumentado (ADF) tenha sido empregado neste trabalho, existem outros testes que podem ser utilizados para a detecção de tendências na série temporal, dentre estes destacamos os testes não paramétricos de Cox-Stuart (Box & Jenkins, 1976), Mann-Kendall e o teste baseado no coeficiente de correlação de Spearman (Spearman rank correlation), conforme apresentado em Sonali e Kumar (2013). Os autores agradecem aos revisores anônimos pela indicação desta referência.

A existência de sazonalidade foi verificada através da análise espectral e pelo teste G de Fisher (com 0,05 de significância). O periodograma da análise espectral consiste na decomposição da série temporal em uma série de Fourier (no domínio de frequências). No eixo das ordenadas ficam localizadas as frequências da série ( $f_i$ ) e no eixo das abscissas as respectivas intensidades de cada frequência  $I_p(f_i)$ , definidas conforme em (4):

$$I_p(f_i) = \frac{2}{\left(\frac{N}{2}\right)} \left\{ \sum_{t=1}^n \varepsilon_t \cos \left[ \frac{2\pi i}{\left(\frac{N}{2}\right)} t \right]^2 \sum_{t=1}^n \varepsilon_t \sen \left[ \frac{2\pi i}{\left(\frac{N}{2}\right)} t \right]^2 \right\} \quad (4)$$

Em que  $\varepsilon_t$  é o componente estocástico da série temporal associado ao tempo  $t$  e  $N$  é o número de observações da série. Para determinar a periodicidade sazonal ( $s$ ) basta verificar a qual frequência  $f_i$  está associado o maior valor de intensidade  $I_p(f_i)$  e então dividir 1 por este valor de frequência. Posteriormente, para testar sua significância estatística utilizou-se o teste G de Fisher. Se comprovado que o valor da estatística G é superior ao valor de Z conclui-se que a componente sazonal com periodicidade  $s = 1/f_i$  é significativa. Maiores detalhes sobre estes procedimentos podem ser encontrados em

Morettin & Toloí (2006) e Jenkins & Watts (1968). Ressalta-se que a detecção de sazonalidade nos dados pode também ser constatada por meio de métodos de regressão linear múltipla com variáveis Dummies (Morettin & Toloí, 2006), método da Análise de Variâncias e teste de Kruskal-Walis, conforme mencionado por Findlley et al.(1998).

$$G = \frac{\max[I_p(f_i)]}{\sum_{i=1}^{\binom{N}{2}} I_p(f_i)} \quad (5)$$

$$Z = 1 - \left[ \frac{\alpha}{\binom{N}{2}} \right]^{\frac{1}{\binom{N}{2} - 1}} \quad (6)$$

O modelo estatístico selecionado foi o SARIMA multiplicativo, que é adequado para modelagem de dados com autocorrelação periódica e inferior a um ano, o que caracteriza o efeito sazonal. Além dos operadores autoregressivos e de médias móveis de ordem p e q do modelo ARIMA, este consta com operadores autoregressivos e de médias móveis sazonais de ordem P e Q, conforme a Eq. (7).

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (7)$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (8)$$

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

$$\Phi_P(B^s) = 1 - \Phi_s B^s - \Phi_{2s} B^{2s} - \dots - \Phi_{Ps} B^{Ps} \quad (10)$$

$$\Theta_Q(B^s) = 1 - \Theta_s B^s - \Theta_{2s} B^{2s} - \dots - \Theta_{Qs} B^{Qs} \quad (11)$$

Em que  $\varepsilon_t$  é o ruído aleatório no tempo t,  $\phi_p(B)$  e  $\theta_q(B)$  são respectivamente os operadores auto-regressivos e de médias móveis definidos em (8) e (9) e  $\Phi_P(B^s)$  e  $\Theta_Q(B^s)$  os operadores auto-regressivos e de médias móveis sazonais, conforme em (10) e (11). O termo  $W_t$  representa os valores da série original ( $Y_t$ ) após a aplicação de d e/ou D diferenças de ordem 1 e ordem s, necessárias para tornar a série estacionária, o que segundo Makridakis et al. (1998), é uma pressuposição para aplicação da metodologia Box & Jenkins.

A aplicação de tal metodologia consiste em três etapas usuais: identificação dos parâmetros do modelo, estimação e diagnóstico. Segundo Morrettin & Toloí (2006), com a série estacionária, a identificação dos parâmetros ocorre como em um processo Auto Regressivo de Médias Móveis (ARMA), via funções de Autocorrelação (ACF) e Autocorrelação Parcial (PACP). No entanto, Makridakis, Wheelwright e Hyndman(1998) afirmam que obter a ordem exata do modelo através destes gráficos é algo complexo e subjetivo, devido ao grande número de modelos que podem ser ajustados. Logo, neste trabalho esta etapa foi conduzida simulando diversos modelos SARIMA e selecionando os

seis mais parcimoniosos por meio do Critério de Informação de Akaike (AIC) (Akaike, 1974), calculado como em (12):

$$AIC = -2.(LMV) + 2.(n) \quad (12)$$

Em que LMV é o logaritmo da máxima verossimilhança e n é número de parâmetros do modelo. A estimação dos parâmetros ocorreu pela distribuição Gaussiana (normal) e método da máxima verossimilhança exata, satisfazendo as condições de invertibilidade e unicidade dos parâmetros (Rezende et al., 2005).

O diagnóstico consistiu em verificar se os resíduos possuíam características de um ruído branco, ou seja,  $\varepsilon \stackrel{iid}{\sim} N(0, \sigma^2)$ . Para avaliar a presença de autocorrelação serial utilizou-se o teste estatístico Ljung-Box (Ljung & Box, 1978) além da ACF para os resíduos. A estatística do teste Ljung-Box possui distribuição Qui-Quadrado com k - n graus de liberdade, dada em (13):

$$Q = N(N + 2) \sum_{i=1}^k (N - k)^{-1} r_k^2 \quad (13)$$

Em que k é o número de defasagens utilizadas (normalmente 15) e  $r_k$  é valor estimado do k-ésimo coeficiente de autocorrelação serial. Logo, o valor-p é dado como  $aProb [Q \geq \chi_{G.L.}^2]$ . A normalidade foi também verificada por meio do valor-p para o formal e frequentemente empregado teste Shapiro-Wilk (Shapiro & Wilk, 1965).

Após estes procedimentos, foi realizada a previsão dos valores de temperatura média mensal para os 12 meses do ano de 2013. A escolha do modelo de previsão ocorreu baseada no indicador de Erro Quadrático Médio (EQM) em (14):

$$EQM = \frac{1}{12} \sum_{t=1}^{12} e_t^2 \quad (14)$$

Em que  $e_t$  é o erro aleatório estimado para cada mês do ano de 2013. Utilizou-se o software livre R (R Development Core Team, 2012) para manipulação e análise estatística dos dados, sendo o *stats*, *forecast* e *time Series* os principais pacotes utilizados. As análises foram conduzidas considerando um nível de significância  $\alpha = 5\%$ .

### 3 Resultados e discussão

A Figura 1 apresenta o gráfico da série de temperatura média mensal e sua decomposição em componentes de tendência, sazonalidade e parte aleatória. Nota-se que a série oscila em torno de um valor médio e com um padrão periódico bem definido, o que deixa subentendido a presença de sazonalidade aditiva e a inexistência de tendência nos dados. A Tabela 1 apresenta uma análise descritiva da variável em estudo.

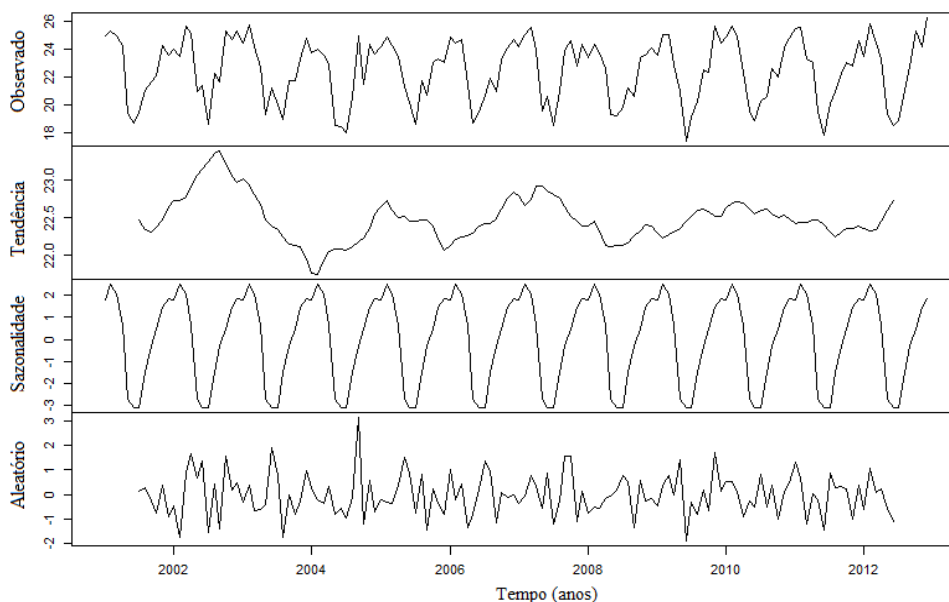


Figura 1 - Gráfico e decomposição da Série Temporal de temperatura média.

Tabela 1 - Estatística Descritiva para a Série de Temperatura Média de Bauru

Medidas Estatísticas	Média	Desvio Padrão	Mínimo	Máximo	Coefficiente de Variação
Valores	22,5	2,2652	17,4	26,3	10,06%

Os baixos valores de desvio padrão e coeficiente de variação corroboram as conclusões iniciais a respeito da inexistência de tendência na série, devido à pequena variabilidade dos dados em torno do valor médio de temperatura média mensal de 22,5°C. Percebe-se que ao longo de 12 anos menor valor de temperatura média mensal foi de 17,4 °C (junho de 2009) e o maior de 26,3 °C (dezembro de 2012). Estes resultados coincidem com as informações fornecidas por Ventura (1964), de que na região norte e noroeste do estado de São Paulo a temperatura média anual gira em torno de 22,6 °C, e nos meses mais frios é próxima ou superior a 18°C.

Na Tabela 2 apresentam-se os resultados para o Teste de Dickey-Fuller Aumentado (ADF), na qual verifica-se que a série em estudo é estacionária em nível, visto que o teste rejeitou a hipótese nula de raiz unitária nos dados. Foram necessárias 7 defasagens (ambas significantes à 5% de probabilidade) para que os resíduos do teste ADF se tornassem um ruído branco. Nota-se ainda que o coeficiente de tendência determinística não foi significativo, o que confirma as conclusões anteriores, de que a série não apresenta componente de tendência.

Tabela 2 - Resultados para o Teste de Dickey-Fuller Aumentado (ADF)

Parâmetros	$\alpha$	$\beta$	$\delta$
Estimativas	39,1717	-0,0012	-1,7378
Erro-Padrão	4,3010	0,0028	0,1904
Estatísticas do Teste	9,108	-0,4280	-9,1260
Valores Críticos (5%)	4,75	6,49	-3,43
valor-p	0,0000*	0,6697	0,0000*

\* Significativo a 0,05 de probabilidade

Como mencionado anteriormente parece existir um padrão periódico nos dados, portanto, para detectar sua frequência de ocorrência e eliminá-lo através de diferenças sazonais, foi realizada a decomposição espectral, conforme mostra a Figura 2.

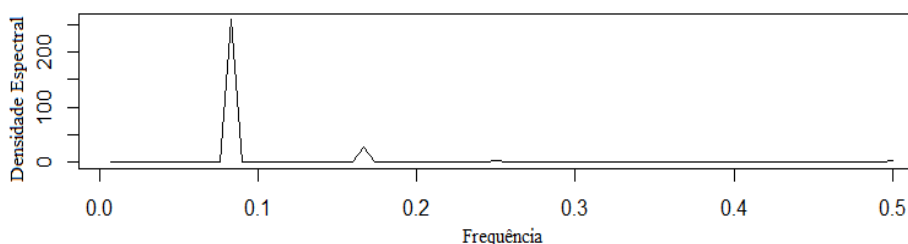


Figura 2 - Periodograma da Série de Temperatura Média.

O Periodograma demonstra que a maior densidade espectral (pico) está associada à frequência de 0,08333, o que implica em uma componente sazonal com periodicidade de  $s=12$  meses. Pelo teste G de Fisher foram obtidos os valores de  $G = 0,8706$  e  $Z = 0,0973$ . Como o valor de  $G$  é superior ao de  $Z$ , conclui-se sobre a presença de sazonalidade anual (12 em 12 meses) na série de temperatura média.

Tais resultados referentes à ausência de tendência e principalmente quanto à presença de sazonalidade já eram esperados, visto que a variável em estudo é climatológica. Em estudos como os de Silva, Guimarães & Tavares (2008), Chechi & Bayer (2012) e Cavalcanti et al. (2005), que avaliam a temperatura média do ar em determinadas regiões, observou-se que quando a componente de tendência está presente na série é de forma suave, diferentemente da sazonalidade, sempre bem definida. Neste sentido uma diferença de ordem  $s = 12$  foi aplicada para eliminar a sazonalidade presente na série de temperatura média da cidade de Bauru (SP). A Figura 3 ilustra a série sazonalmente ajustada e as respectivas ACF e PACF.



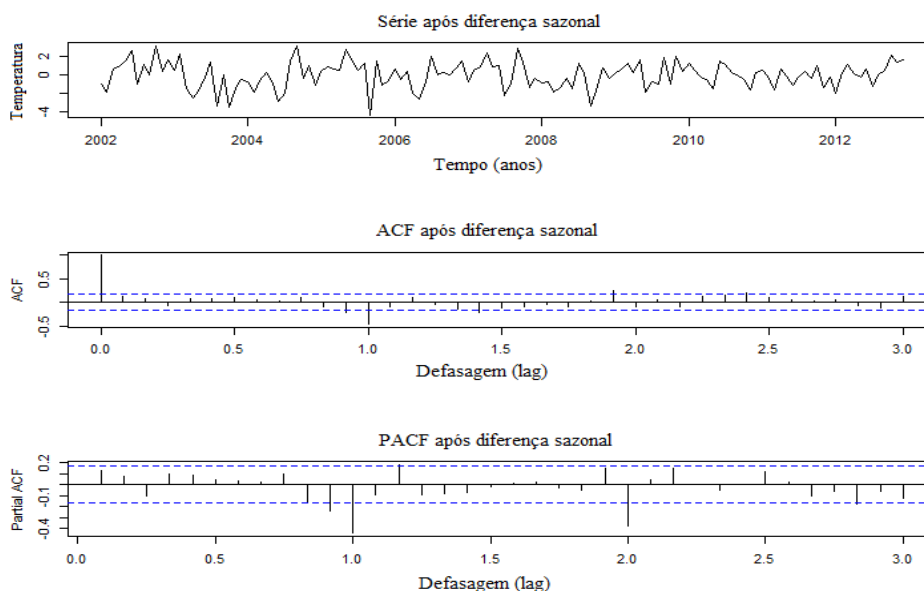


Figura 3 - Série Temporal, ACF e PACF após aplicação da diferença sazonal.

Nota-se que alguns lags múltiplos de 12 na ACF e PACF ainda foram significativos, o que segundo Morettin e Tolo (2006) indica a presença de sazonalidade estocástica, que deve ser eliminada ajustando-se um modelo que a considere, além da determinística, retirada por diferença.

Ajustou-se diversos modelos SARIMA  $(p, d, q) \times (P, D, Q)_s$ , com os valores de  $p$  e  $q$  variando de 0 a 3, e  $P$  e  $Q$  de 0 a 2. Como a série não apresenta tendência, não houve necessidade de aplicar diferenças de ordem 1, então, automaticamente tem-se que  $d=0$ . Como uma diferença sazonal (de ordem  $s=12$ ) foi aplicada para eliminar a sazonalidade, tem-se  $D = 1$ . Os resultados podem ser visualizados na Tabela 3.

Tabela 3 - Comparação entre modelos SARIMA

Modelo	SARIMA (s=12)	AIC	Ljung-Box (valor-p)	Shapiro-Wilk (valor-p)	EQM
1	(1,0,0) x (0,1,2)	401,88	0,3383	0,3790	0,6626
2	(0,0,2) x (0,1,2)	399,76	0,5679	0,2582	0,6869
3	(1,0,3) x (1,1,0)	440,73	0,1656	0,0928	1,1992
4	(0,0,1) x (2,1,1)	396,27	0,4137	0,4123	0,6636
5	(1,0,1) x (0,1,2)	403,19	0,3057	0,3903	0,7228
6	(0,0,2) x (2,1,2)	398,42	0,6516	0,4256	0,6991

Estes foram os 6 modelos selecionados segundo o AIC. Todos eles estimam resíduos que são ruído branco, visto que o valor-p para os testes Ljung-Box e Shapiro-Wilk foram superiores a 0,05, o que implica em resíduos independentes e normalmente distribuídos. O modelo SARIMA (0,0,1) x (2,1,1)<sub>12</sub> e o SARIMA (1,0,0) x (0,1,2)<sub>12</sub> apresentaram os menores valores de EQM, no entanto, o primeiro foi selecionado por possuir menor AIC e ainda ter apresentado todos os coeficientes significativos de acordo com o teste t de Student, conforme apresentado na Tabela 4.

Tabela 4 - Estatísticas para o Modelo SARIMA (0,0,1) x (2,1,1)<sub>12</sub>

Parâmetros	Coefficientes	Erro Padrão	t de Student	Valor-p
$\theta_1$	0,1757	0,0893	2,0456	0,0408*
$\Phi_1$	-0,3005	0,1235	-2,4317	0,0150*
$\Phi_2$	-0,2566	0,1109	-2,3141	0,2060*
$\theta_1$	-0,7064	0,1214	-5,8169	0,0000*

\* Significativo a 0,05 de probabilidade

Destaca-se que todos os modelos foram estimados sem o termo intercepto devido à aplicação de uma diferença de ordem 12 à série original, visto que os novos valores passam a oscilar em torno do valor zero, implicando que tal termo seja não significativo.

Na Figura 4, além da ACF dos resíduos padronizados é possível verificar a homogeneidade dos mesmos, o que corrobora as características de ruído branco.

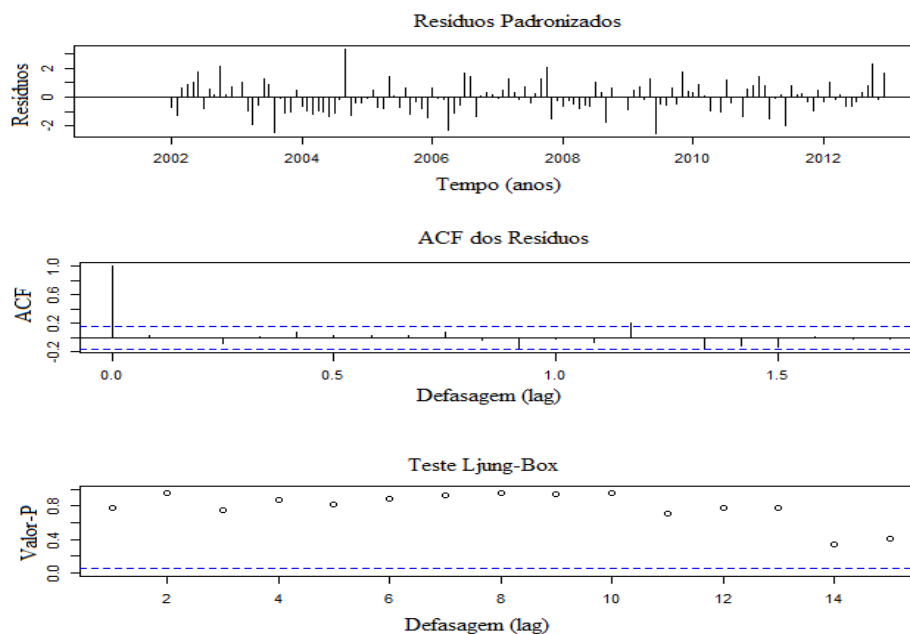


Figura 4 - Comprovação das características de ruído branco.

A Figura 5 ilustra o ajuste de tal modelo aos dados históricos de temperatura média mensal da cidade de Bauru (SP). Nota-se que os valores ajustados estão bem próximos dos reais, o que significa que o modelo conseguiu captar a dinâmica temporal da série.

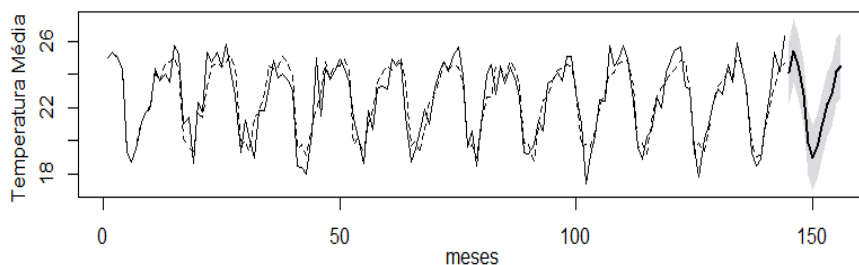


Figura 5 - Valores Reais (Linha Contínua) e Valores Previstos (Tracejada).

Na Tabela 5 são apresentadas algumas informações sobre a previsão para os 12 meses de 2013, desde os valores previstos, reais, os resíduos e os respectivos intervalos de confiança de 95%.

Tabela 5 - Estimativas de Temperatura Média mensal da Cidade de Bauru (SP) para o ano de 2013

Meses	Limite Inferior (5%)	Previsto	Limite Superior (95%)	Valor Real	Resíduos
Janeiro	22,3296	24,2454	26,1612	23,9	-0,345
Fevereiro	23,442	25,358	27,274	24,7	-0,658
Março	22,748	24,693	26,638	23,7	-0,993
Abril	20,878	22,823	24,768	21,9	-0,923
Maio	17,892	19,837	21,782	20,5	0,662
Junho	17,046	18,991	20,936	19,6	0,608
Julho	17,646	19,591	21,536	18,2	-1,391
Agosto	18,884	20,829	22,774	19,7	-1,129
Setembro	20,285	22,231	24,176	21,5	-0,73
Outubro	20,817	22,763	24,708	22,6	-0,163
Novembro	22,184	24,129	26,074	23,9	-0,229
Dezembro	22,486	24,432	26,377	25,4	0,967

Destaca-se que no trabalho de Silva, Guimarães e Tavares (2008) os modelos SARIMA foram também aplicados para previsão de temperatura média da cidade de Uberlândia (MG) e apresentaram resultados muito satisfatórios, com baixos indicadores

de erro e resíduos com características de ruído branco. Resultados similares foram encontrados por Chechi e Bayer (2012), que aplicaram estes modelos para previsão de temperatura mínima e máxima para a cidade de Erechim (RS). Ainda no trabalho de Chechi e Bayer (2012) verifica-se que os modelos SARIMA obtiveram melhor desempenho do que os modelos de Alisamento Exponencial ou Holt-Winters, apresentando menores indicadores de erro para as previsões.

Para complementar os resultados, apresentamos na Figura 6 as previsões com o modelo ajustado para os próximos 36 meses, ou seja, para os meses dos anos de 2014, 2015 e 2016. No entanto, destaca-se que previsões mais precisas podem ser obtidas se houver uma atualização periódica dos dados (semestralmente ou anualmente), isto é, se após o conhecimento de tais valores reais, estes forem incorporados ao conjunto de dados e a análise refeita, no intuito de agregar histórico de observações e captar novos padrões temporais.

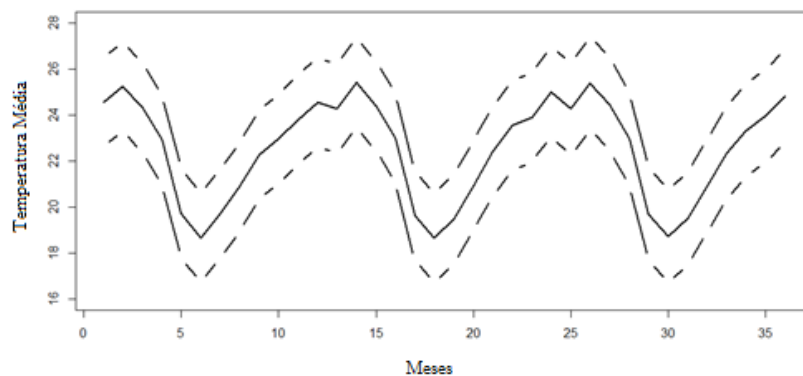


Figura 6 - Previsão (Linha Contínua) e Limites de Confiança de 95% (Linha Tracejada).

## Conclusões

Verificou-se que a série de temperatura média mensal da cidade de Bauru (SP) não apresentava tendência de crescimento e sim uma componente sazonal com periodicidade de 12 meses. Neste sentido, empregou-se a metodologia Box & Jenkins para desenvolver previsões e obter estimativas de temperatura média para os 12 meses de 2013.

Foram propostos 6 modelos da classe SARIMA e pôde-se concluir que o modelo SARIMA  $(0,0,1) \times (2,1,1)_{12}$  obteve melhor ajuste aos dados de temperatura média mensal da cidade de Bauru (SP), apresentado menor Erro Quadrático Médio (EQM), respectivamente, 0,6636, além de resíduos com características de ruído branco, o que corrobora a qualidade do ajuste. Além disso, as previsões para o ano de 2013 foram bem próximas dos valores reais. Isto significa que o modelo conseguiu captar a dinâmica temporal da série, reproduzindo principalmente a sazonalidade presente nos dados.

Observa-se também que os doze valores reais de temperatura média mensal da cidade de Bauru (SP) para o ano de 2013 situam-se dentro dos intervalos de confiança de 95% fornecidos pelo modelo ajustado. Isso é um ótimo resultado, principalmente devido às incertezas e imprevistos associados a fatores climáticos. Para questões práticas, como o

modelo ajustado mostrou-se adequado no que se refere à previsão de temperatura média mensal da cidade de Bauru, os resultados oferecidos na Figura 6 poderiam ser utilizados no intuito de auxiliar possíveis tomadas de decisões e a realização de planejamentos estratégicos futuros que envolvem questões públicas associadas à cidade de Bauru.

Como verificado, a metodologia Box & Jenkins proposta mostrou-se adequada para a modelagem de dados climatológicos. Neste sentido, para futuros trabalhos, sugere-se sua aplicação em dados de temperaturas mínimas e máximas para esta e outras regiões Brasileiras. Uma explicação para este tipo de estudo é que a temperatura média pode mascarar a variabilidade dos dados, oferecendo resultados menos precisos. Além disso, conforme verificado em Minuzzi (2011), conclui-se que a temperatura mínima de determinadas localidades do Brasil tem aumentado consideravelmente, portanto, resultados que comprovem estas teorias poderiam ser muito interessantes.

BARBOSA, E. C.; SÁFADI, T.; NASCIMENTO, M.; NASCIMENTO, A. C. C.; SILVA, C. H. O.; MANULI, R. C. Box & Jenkins methodology to forecasting of average monthly temperature of Bauru city (SP). *Rev. Bras. Biom.*, São Paulo, v.33, n.1, p.104-117, 2015.

▪ **ABSTRACT:** *In this paper was applied the Box & Jenkins methodology to obtain a statistical model which allowed forecasting the future values of monthly average temperature of Bauru city in São Paulo state. The data set utilized include information of twelve years of average temperature (January 2001 to December 2012). By the Akaike Information Criterion (AIC) we selected 6 SARIMA models and the choice of best model occurred with indicator of Mean Square Error (MSE), calculated by the quadratic sum of difference between the real values and the forecasting values obtained for the temperature values of monthly of 2013. Furthermore, we analyzed the statistical significance of model coefficients and residuals white noise. The model chosen was the SARIMA (0,0,1) x(2,1,1)<sub>12</sub>, due the accurate forecasts and satisfactory statistical results.*

▪ **Key words:** *Statistical; modeling; seasonality; SARIMA.*

## Referências

BOX, G.; JENKINS, G. *Time series analysis, forecasting and control*. San Francisco: Holden-Day, 1976. 575 p.

CAVALCANTI, E. P.; SILVA, V. P. R.; SOUSA, F. A. S. Programa computacional para a estimativa da temperatura do ar para a Região Nordeste do Brasil. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.10, p.140-147, 2005.

CHECHI, L.; BAYER, F. M. Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.16, p.1321-1329, 2012.

DICKEY, D. A.; FULLER, W. A. Distributions of the Estimators For Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, v.75, p.427-431, 1979.

DICKEY, D. A.; FULLER, W. A. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, v.49, p.1057-1072, 1981.

IBGE - Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2010. <<http://censo2010.ibge.gov.br/>>.

IPMET - Instituto de Pesquisas Meteorológicas da UNESP. Estação Meteorológica. <<http://www.ipmet.unesp.br/>>

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

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for statistical computing. <<http://www.Rproject.org>>

REZENDE, J. L. P.; COELHO JUNIOR, L. M.; OLIVEIRA, A. D.; SÁFADI, T. Análise dos preços de carvão vegetal em quatro regiões no estado de Minas Gerais. *CERNE*, v.11, p.237-252, 2005.

SÁFADI, T. Uso de séries temporais na análise de vazão de água na Represa de Furnas. *Ciência e Agrotecnologia*. v.28, p.142-148, 2004.

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

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

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

VENTURA, A. Problemas técnicos da silvicultura paulista. *Silvic*. S. Paulo, v.3, p.61-80, 1964.

WERNER, L.; RIBEIRO, J. L. D. Uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. *Gestão e Produção*, v.10, p.47-67, 2003.

Recebido em 09.09.2014

Aprovado após revisão em 10.12.2014