

Análises *In Sample* e *Out of Sample* na
Modelagem e Previsão da Vazão de Afluentes da
Usina Hidrelétrica de Sobradinho sob as
Abordagens de Holt-Winters e de Box & Jenkins

*In Sample and Out of Sample analysis forecasting in Sobradinho
Hydroelectric plant affluent flow by Holt-Winters and Box & Jenkins
approaches*

Lorena Tereza Bezerra Gamboni*

Moisés Lima de Menezes†

Keila Mara Cassiano‡

Resumo: O Brasil é um país rico em recursos naturais e sua faixa fluviométrica é uma das mais extensas do mundo com uma vasta quantidade de rios em seu território. O bom aproveitamento deste recurso inclui, dentre outras coisas, a geração de energia hidrelétrica. Neste país, mais de 80% de toda a energia gerada é de origem hidrelétrica e para que a energia possa ser gerada, usinas são instaladas em pontos estratégicos. Para que se tenha sempre energia suficiente para a demanda alocada à usina, é necessário um planejamento adequado e, para que isso possa ser feito com eficiência, estimar a vazão futura na barragem da usina é essencial. Este artigo aborda a modelagem e previsão da série temporal de vazão de afluentes da usina hidrelétrica de Sobradinho com o modelo de amortecimento exponencial de Holt-Winters e o modelo de Box & Jenkin com o intuito de verificar qual o modelo mais adequado para este tipo série. A capacidade preditiva dos modelos é medida a partir das estatísticas de aderência *MAD*, *MAPE*, *RMSE*, *BIC* e R^2 para dados dentro (*In Sample*) da amostra e *MAD* e *MAPE* para dados fora (*Out of Sample*) da amostra. A análise das estatísticas de aderência mostra que apesar de os dois modelos

*Departamento de Estatística, Universidade Federal Fluminense, lorena_gamboni@id.uff.br

†Departamento de Estatística, Universidade Federal Fluminense, moises_lima@id.uff.br

‡Departamento de Estatística, Universidade Federal Fluminense, keilamath@hotmail.com

terem comportamentos parecidos, os resultados obtidos a partir da modelagem de Box & Jenkins são melhores para este caso.

Palavras-chave: Séries temporais. Vazão de afluentes. Holt-Winters. Box & Jenkins. Previsão.

Abstract: Brazil is a very rich country in natural resources and your fluviometric range is one of the largest in the world with a vast amount of rivers in its territory. The good use of this resource include, among other things, the generation of hydroelectric power. In Brazil, over 80% of the generated energy is hydroelectric and that energy can be generated, plants are installed at strategic points. In order to always have enough power to demand which the plant is allocated, you must have proper planning and for this to be done efficiently, knowing the amount of future flow in the plant's dam is essential. In order to verify which model is most suitable, this article discusses the modeling and forecasting of time series tributary of the hydroelectric plant flow Sobradrinho with the exponential smoothing model Holt-Winters and model Box & Jenkin. The predictive ability of the models is measured from the adherence statistics, *MAD*, *MAPE*, *RMSE*, *BIC* and R^2 for data within the sample and *MAD* and *MAPE* for data outside the sample. The adherence statistics analysis shows that although the two have similar behavior models, the results obtained from modeling Box and Jenkins are better for this case.

Keywords: Time series. Affluent flow. Holt-Winters. Box & Jenkins. Forecasting.

1 Introdução

A energia gerada a partir de usinas hidrelétricas é produzida através do potencial hidráulico da corrente de um rio. Para que isso ocorra é preciso que esse rio tenha grande volume de água e desníveis em seu fluxo. Também é preciso que a corrente gerada pelo rio tenha uma grande força e velocidade ao passar pelas turbinas da usina, para que a energia obtida vá até o gerador, se transforme em energia elétrica e depois seja conduzida pelos fios até os centros residenciais e comerciais [1].

A vazão afluente de um rio é dita como uma vazão que chega até uma certa localização, tendo como principal objetivo a produção de energia elétrica. Temos que a vazão está proporcionalmente ligada a quantidade de energia gerada por uma hidrelétrica, uma vez que define o volume de água contido na barragem e que passa pelas turbinas.

Para que o abastecimento possa atender as necessidades das demandas de energia se faz necessário ter um planejamento adequado do uso deste bem. A modelagem e previsão de séries temporais de vazão de afluentes é uma ferramenta estatística essencial para este tipo de planejamento. Existem muitos modelos de previsão de séries temporais, porém para cada fenômeno há um modelo adequado.

Por sua vez, [3] apresenta uma pesquisa com o objetivo de gerar uma simulação de vazões e níveis de médios de água para o Rio Tapajós. Foram analisados os níveis de água e vazões médias diárias nos períodos de 1976 até 2008. O método de Box & Jenkins foi utilizado para realizar os estudos destas variáveis e o modelo ajustado para este caso foi o SARIMA. O principal resultado obtido nessa pesquisa foi que o modelo utilizado alcançou de forma adequada o comportamento das séries de níveis de água e de vazão.

O artigo [7] realizou uma pesquisa com o objetivo de obter uma modelagem para previsão de demanda de energia com filtragem Singular Spectrum Analysis (SSA). Foi analisado o consumo residencial de energia elétrica de uma distribuidora de eletricidade que atende uma parcela do estado do Rio de Janeiro contendo 136 observações, sendo usadas como amostra de teste apenas as últimas 6 observações. Foram utilizadas as modelagens ARIMA e Holt-Winters associadas à filtragem SSA [4]. O principal resultado obtido nessa pesquisa foi que a filtragem de séries temporais por SSA antes de sua modelagem aponta uma melhora considerável nas previsões.

As abordagens *in sample* e *out of sample* permitem fazer uma análise de séries temporais com mais precisão em relação à sua previsão, uma vez que elas verificam o poder preditivo do modelo comparando previsões a dados históricos.

Neste artigo são utilizados, para a modelagem de uma série temporal de média mensal de vazão de afluentes da Usina Hidrelétrica Sobradinho do Operador Nacional do Sistema Elétrico (ONS) [11], os modelos de Holt-Winter e de Box & Jenkins com o objetivo de verificar qual dos dois modelos, em termos de ajuste, é o mais adequado para este caso. O poder preditivo de cada modelo é testado via estatísticas de aderência: MAD , $MAPE$, R^2 , $RMSE$, BIC através de análises *in sample* e *out of sample* de modo que a primeira parte é composta de boa parte da amostra e a modelagem é feita a partir dos dados amostrados e é também conhecida como “amostra de treinamento” e a segunda composta pelo último ano de observações em que o modelo é obtido a partir dos próprios dados previstos pelo modelo é conhecida como “amostra de teste”.

Este artigo está organizado como segue. Na seção 2 são apresentados os materiais e métodos dando enfoque aos modelos de amortecimento exponencial de Holt-Winters e os modelos de Box & Jenkins bem como as estatísticas de aderência a serem utilizadas. Na seção 3 é apresentada a usina hidrelétrica de Sobradinho e os dados de vazão de afluentes. Também são apresentadas estatísticas descritivas e os softwares utilizados na modelagem. As análises e discussões estão na seção 4 e na seção 5 estão as conclusões.

2 Materiais e Métodos

2.1 Séries Temporais

Uma série temporal é um conjunto de observações ordenadas no tempo. As séries temporais podem ser divididas em dois grupos: contínuas e discretas. As séries do segundo grupo podem ser obtidas através de amostras retiradas em períodos de tempo semelhantes. Os objetivos da análise de séries temporais são de levantar modelos paramétricos, onde o número de parâmetros é finito, ou não-paramétricos, onde o número de parâmetros é infinito, fazendo projeções de valores futuros ou apenas relatando o comportamento da série, construindo modelos probabilísticos ou estocásticos, ou seja, controlados por leis probabilísticas [8]. Uma série temporal de comprimento T será representada por

$$Z_t = \{z_1, z_2, \dots, z_T\}. \quad (1)$$

2.1.1 Decomposição de Séries Temporais

Segundo o modelo clássico, todas as séries temporais são compostas de quatro padrões:

- **Tendência** (T_t), que é o comportamento de longo prazo da série (onde t é o instante atual);
- **Variações cíclicas ou ciclos** (C_t), que resultam em flutuações nos valores da variável com duração superior a um ano, e que se repetem com certa periodicidade;
- **Variações sazonais ou sazonalidade** (S_t), que resultam em flutuações nos valores da variável com duração inferior a um ano, e que se repetem todos os anos;
- **Variações irregulares** (I_t), que são as flutuações inexplicáveis, resultado de fatos fortuitos e inesperados como catástrofes naturais, atentados terroristas como o de 11 de setembro de 2001, decisões intempestivas de governos, etc.

A decomposição da série permitirá identificar quais componentes estão atuando naquele conjunto em particular, além de possibilitar obter índices e/ou equações para realizar previsões para períodos futuros da série.

Dessa forma, é possível decompor a série em (1) como se segue:

$$Z_t = T_t + S_t + R_t, \quad (2)$$

onde T_t e S_t representam a tendência e a sazonalidade, respectivamente, e R_t é uma componente ruidosa (aleatória que pode incluir os elementos irregulares (I_t) e alguns elementos identificados nas variações cíclicas (C_t)).

Para se estimar a tendência, deve ser usada algum tipo de suavização. As mais utilizados são as médias móveis e as medianas moveis. Também existem procedimentos para se estimar a sazonalidade, onde os mais utilizados são método de regressão e método de médias móveis, onde o método de regressão é melhor para sazonalidade determinística, enquanto o de médias móveis é melhor para sazonalidade estocástica. Pode-se verificar a existência de tendência e sazonalidade através de gráficos e testes de hipótese que podem ser feitos antes ou depois da estimação [8].

2.2 Modelos de Holt-Winters

O método de amortecimento exponencial de Holt-Winters foi sugerido por Holt (1957) e Winters (1960), que trabalharam na *School of Industrial Administration*, do *Carnegie Institute of Technology*. O método usa médias móveis ponderadas exponencialmente para atualizar as estimativas da média ajustada sazonalmente (chamada de nível), inclinação e sazonalidade. Este modelo possui duas equações de previsão: aditiva e multiplicativa. Em [6], os autores sugerem que o melhor critério de escolha entre os fatores multiplicativos ou aditivos está em calcular medidas de precisão, assim a opção recai naquela que apresentar o(s) menor(es) erro(s). Outra possível forma de indicar o modelo aditivo, reside na possibilidade de as séries apresentarem oscilações aproximadamente constantes. Entretanto, se as oscilações sazonais forem proporcionais ao nível da série, o modelo multiplicativo é mais indicado [13].

Existem dois tipos de suavização exponencial de Holt-Winters: com sazonalidade multiplicativa e a com sazonalidade aditiva. O modelo com sazonalidade multiplicativa apresenta o modelo geral

$$Z_t = (a_1 + a_2 t) \times \rho_t + \varepsilon_t, \quad (3)$$

onde a_1 e a_2 são os parâmetros do modelo a serem estimados a cada instante de tempo (nível e tendência, respectivamente), ρ_t são os fatores sazonais (onde são estimados dependendo do ciclo sazonal e da periodicidade dos dados) e ε_t é um componente de erro (uma variável aleatória *iid* com média zero e variância constante). A equação de previsão deste modelo é dada por

$$\hat{Z}_t(h) = (\hat{a}_1(t) + \hat{a}_2(t)h) \times \hat{\rho}_{m(t+h)}(h), \quad (4)$$

onde h representa o número de passos à frente da previsão.

As estimativas em (4) são obtidas a partir das equações de atualizações dos parâmetros:

$$\begin{aligned} \hat{a}_1(t) &= \alpha \left[\frac{Z_t}{\hat{\rho}_{m(t)}(t-1)} \right] + (1 - \alpha) [\hat{a}_1(t-1) + \hat{a}_2(t-1)]; \\ \hat{a}_2(t) &= \beta \left[\frac{\hat{a}_1(t)}{\hat{a}_1(t-1)} \right] + (1 - \beta) [\hat{a}_2(t-1)]; \\ \hat{\rho}_{m(t)}(t) &= \gamma \left[\frac{Z_t}{\hat{a}_1(t)} \right] + (1 - \gamma) [\hat{\rho}_{m(t)}(t-1)], \end{aligned}$$

onde α , β e γ são constantes de amortecimento do nível, da tendência e dos fatores sazonais, respectivamente.

O modelo aditivo é representado por

$$Z_t = a_1 + a_2 t + \rho_t + \varepsilon_t, \quad (5)$$

com equação de previsão

$$\hat{Z}_t(h) = \hat{a}_1(t) + \hat{a}_2(t)h + \hat{\rho}_{m(t+h)}(h) \quad (6)$$

e com atualização dos parâmetros dada pelas equações:

$$\begin{aligned} \hat{a}_1(t) &= \alpha [Z_t - \hat{\rho}_{m(t)}(t-1)] + (1 - \alpha) [\hat{a}_1(t-1) + \hat{a}_2(t-1)]; \\ \hat{a}_2(t) &= \beta [\hat{a}_1(t) - \hat{a}_1(t-1)] + (1 - \beta) [\hat{a}_2(t-1)]; \\ \hat{\rho}_{m(t)}(t) &= \gamma [Z_t - \hat{a}_1(t)] + (1 - \gamma) [\hat{\rho}_{m(t)}(t-1)]. \end{aligned}$$

2.3 Modelos de Box & Jenkins

A classe de modelos desenvolvida por Box & Jenkins [2] associa a ideia de que o modelo pode ser visto como um filtro linear onde um ruído branco ε_t (variável aleatória *iid* com média zero e variância constante) entra no filtro para a saída de um modelo Z_t . Os modelos de Box & Jenkins são conhecidos como modelos *ARIMA*

por serem compostos de uma parte autorregressiva (AR), uma parte integrada (I) e uma parte de médias móveis (MA).

2.3.1 Testes de Normalidade e de Estacionariedade

De acordo com [8], para se modelar $ARIMA$ é necessário que as séries atendam alguns condições como normalidade e estacionariedade. Para se detectar tais condições, alguns testes podem ser feitos como por exemplo, para normalidade:

- teste de Kolmogorov-Smirnov;
- teste de Shapiro-Wilk;
- Teste de Doornik-Hansen;
- Teste de Jarque-Bera.

Nos exemplos citados acima, a hipótese nula é de normalidade e nos casos particulares dos testes de Doornik-Hansen e Jarque-Bera, os testes são feitos a partir das medidas de assimetria e curtose enquanto os outros dois são baseados na função de densidade. Em todos os casos, os programas retornam o valor- p , que para não rejeitar a hipótese nula, precisa ser maior que o nível de significância do teste, que em geral é 0,05.

Para testar a estacionariedade dos dados, o teste mais comum a ser utilizado são os testes de Dickey-Fuller e de Phillips Perron, cujas hipóteses nulas são de existência de raiz unitária, que leva a não estacionariedade. Neste caso, deseja-se rejeitar a hipótese nula para que se verifique a estacionariedade. Assim como nos testes de normalidade, os programas retornam o valor- p , que agora tem que ser menor que o nível de significância do teste para se confirmar a estacionariedade.

Tanto para a normalidade, quanto para a estacionariedade, se os testes não confirmam as hipóteses necessárias para a modelagem $ARIMA$, então alguns procedimentos podem ser seguidos, como por exemplo, para normalidade pode-se proceder com transformações logarítmicas ou de Box-Cox e para estacionariedade, realizar sucessivas diferenças. O processo de se fazer diferenças sucessivas é uma das transformações mais utilizadas para se atingir uma série estacionária. Em geral, são suficientes duas diferenças. O operador de diferenças Δ é definido por

$$\Delta Z_t = Z_t - Z_{t-1}. \tag{7}$$

2.3.2 O Modelo $ARIMA(p, d, q)$

A equação geral dos modelos $ARIMA(p, d, q)$ ¹ é dada por:

$$\phi(B)(1 - B)^d Z_t = \theta(B)\varepsilon_t, \quad (8)$$

onde $\phi(B)$ e $\theta(B)$ são polinômios autorregressivos e de médias móveis respectivamente, p e q são os graus destes polinômios e as raízes de $\phi(B)$ e $\theta(B)$ definem os coeficientes do modelo, B é um operador de defasagem de modo que $B^k Z_t = Z_{t-k}$, $(1 - B) = \Delta$ e d é a quantidade de diferença necessárias para que a série se torne estacionária.

Dentro do modelo $ARIMA$ temos casos particulares. Quando $\phi(B) = 1$ e $d = 0$, o modelo é chamado modelo de média móvel (MA). Quando $\theta(B) = 1$ e $d = 0$, o modelo é chamado modelo autorregressivo (AR) e quando apenas $d = 0$, ou seja, a série possui estacionariedade e não é preciso fazer diferença na série original, o modelo é chamado modelo autorregressivo de média móvel ($ARMA$). A equação de previsão do modelo $ARIMA$ é descrita por

$$\hat{Z}_t(h) = E(Z_{t+h}|Z_t) = \hat{Z}_{t+h}. \quad (9)$$

Quando a série possui sazonalidade, deve-se utilizar a versão do modelo $ARIMA$ com sazonalidade: o modelo $SARIMA$ (ou $ARIMA$ sazonal, ou ainda $ARIMA$ multiplicativo). A equação do modelo $SARIMA(p, d, q) \times (P, D, Q)_s$ é apresentada por

$$\phi(B)\Phi(B^s)(1 - B)^d(B^s)^D Z_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad (10)$$

onde as ordens P e Q são os graus dos polinômios $\Phi(B^s)$ e $\Theta(B^s)$ respectivamente, s é o número de fatores sazonais e D é o número de diferenças sazonais necessárias para que a série se torne sazonalmente estacionária.

2.4 Estatísticas de Aderência

As estatísticas de aderência medem a capacidade de predição de um modelo, sendo o melhor modelo aquele que as minimizam.

Esta seção apresenta as estatísticas de aderência mais utilizadas na literatura. Todas elas são obtidas com base no erro de previsão, por isso dá-se créditos aos mode-

¹A expressão $ARIMA(p, d, q)$ é lida como segue: Modelos autorregressivos, integrados e de médias móveis de ordem autorregressiva p , ordem de diferenças d e ordem de médias móveis q . Para maiores detalhes, veja [8].

los que apresentam valores baixos. Dentre as estatísticas de aderência, destacam-se:

- **Desvio Médio Absoluto** (MAD - *Mean Absolute Deviation*), definido por

$$MAD = \sum_{i=1}^n \frac{|Z_i - \hat{Z}_i|}{n}, \quad (11)$$

onde Z_i é o valor observado no instante i e \hat{Z}_i é o valor ajustado no mesmo instante.

- **Raiz Quadrada do Erro Quadrático Médio** (RMSE - *Root Mean Squared Error*), definido por

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z_i - \hat{Z}_i)^2}{n}}, \quad (12)$$

onde Z_i é o valor observado no instante i e \hat{Z}_i é o valor ajustado no mesmo instante.

- **Erro Percentual Médio Absoluto** (MAPE - *Mean Absolute Percentage Error*) definido por

$$MAPE = \sum_{i=1}^n \left| \frac{Z_i - \hat{Z}_i}{n} \right| \times 100\%, \quad (13)$$

onde Z_i é o valor observado no instante i e \hat{Z}_i é o valor ajustado no mesmo instante.

- **Critério de Informação Bayesiana** (BIC - *Bayesian Information Criterion*) dado por

$$BIC = T \ln(\hat{\sigma}_{\varepsilon_t}^2) - (T - m) \ln \left(1 - \left(\frac{m}{T} \right) \right) + m \ln(T) + m \ln \left[\frac{\left(\frac{\sigma_{Z_t}^2}{\hat{\sigma}_{\varepsilon_t}^2} \right) - 1}{m} \right] \quad (14)$$

leva em consideração o número de parâmetros (m) da modelagem e a varância dos resíduos ($\hat{\sigma}_{\varepsilon_t}^2$). O *BIC* penaliza modelos que tenham um número elevado de parâmetros. Dessa forma, o modelo será considerado adequado se minimizar o *BIC*.

O *BIC* não é uma estatística de aderência baseada no erro de previsão, mas é um dos principais critérios utilizados para escolha do melhor modelo.

- **Coefficiente de Determinação de Pearson R^2** definido por

$$R^2 = 1 - \frac{\sum_{i=1}^n (Z_i - \hat{Z}_i)^2}{\sum_{i=1}^n (Z_i - \bar{Z})^2}, \quad (15)$$

onde \bar{Z} é a média dos dados observados na série.

O coeficiente de Determinação de Pearson não é uma estatística de aderência que precisa ser minimizada. Ele verifica o quanto o modelo explica os dados, de modo que quanto maior for o seu valor, melhor será o modelo.

3 Estudo de Caso

A Usina Hidrelétrica de Sobradinho foi construída na década de 1970 no interior da Bahia no rio São Francisco. Para que Sobradinho mantenha sua capacidade hídrica é necessário que haja chuva na nascente do rio São Francisco, localizada no norte de Minas Gerais. De acordo com a EBC (2016), nos últimos anos houve frequente falta de chuva nessa região, o que fez com que o volume de água retido na hidrelétrica de Sobradinho chegasse a apenas 1,11% de sua capacidade original (28 bilhões de metros cúbicos).

A falta de chuva e a diminuição do volume de água é um problema que vem atingindo com frequência os reservatórios das usinas hidrelétricas. Diante disso, o estudo da média mensal de vazão de afluentes da usina hidrelétrica Sobradinho através de análise de séries temporais é importante para a projeção de como estará o nível de água em um determinado prazo e se o volume contido nesses reservatórios é aceitavelmente viável para a produção de energia suficiente para suprir a necessidade da população.

Os dados utilizados para o estudo do comportamento da quantidade de água na Usina Hidrelétrica de Sobradinho são uma série temporal de vazões médias mensais medidas em m^3/s obtidos do ONS [11]. Os dados observados variam de janeiro de 1931 até dezembro de 2007, totalizando 924 observações, das quais 900 serão utilizadas para amostra de treinamento (*in sample*), onde os dados ajustados são obtidos a partir dos dados observados e as últimas 24 observações são utilizadas como amostra de teste (*out of sample*), onde os dados ajustados são obtidos a partir dos próprios dados já ajustados. A Figura 1 apresenta o comportamento da série original de médias mensais de vazão de afluentes da usina hidrelétrica de Sobradinho

enquanto que a Figura 2 apresenta as séries mensais de modo que para cada mês há uma série temporal e também é exibida sua média.

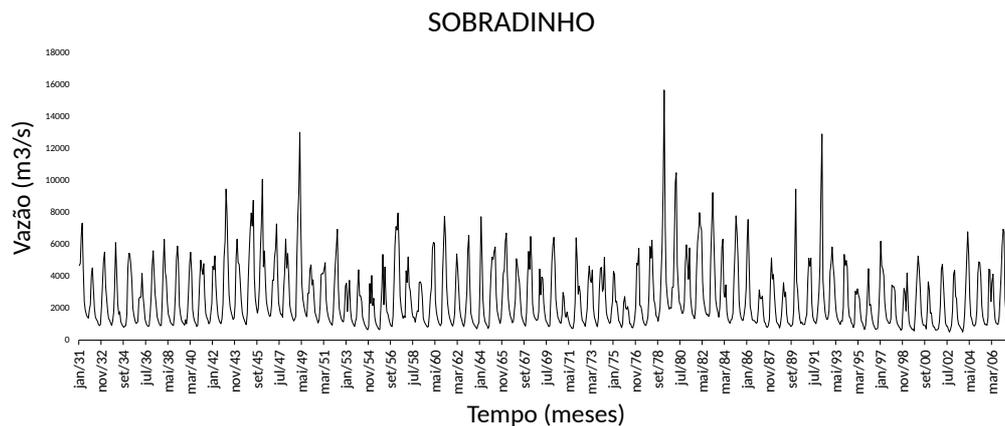


Figura 1: Série temporal original de vazão média mensal de afluentes da Usina Hidrelétrica de Sobradinho. 1931 - 2007.

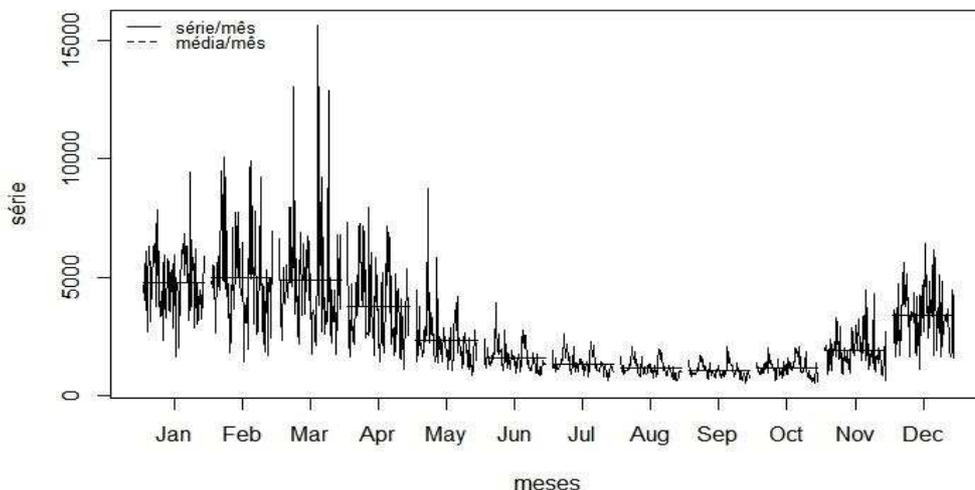


Figura 2: Séries mensais e suas respectivas médias da série de vazão da Usina Hidrelétrica de Sobradinho. 1931-2007.

Pode-se perceber na Figura 1 que a série apresenta alguns picos em algumas épocas isoladas. Já a Figura 2 mostra que a quantidade de vazão de afluentes oscila entre os meses de cada ano, onde os meses de janeiro, fevereiro e março

apresentam maiores médias e os meses de menor vazão são os meses de junho a outubro, considerados como época seca. A Tabela 1 mostra as estatísticas descritivas da série original.

Tabela 1: Estatísticas descritivas da série temporal de vazão média mensal de afluentes da usina hidrelétrica de Sobradinho 1931-2007.

Média	Desvio Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
2.687,02	1.975,84	506	1.196,25	1.925,5	3.784	15.676

A partir da Tabela 1, pode-se perceber uma diferença bem drástica entre a menor e a maior vazão média apresentada nessa série. Também pode-se perceber que a média total é bem menor que a maior vazão média. Logo, a quantidade de vazões baixas é maior do que de vazões altas. Isso também pode ser verificado no 3º Quartil. Porém, percebe-se que esta vazão máxima é fruto de um dos picos que a série dá, sendo este valor um dos casos excepcionais.

4 Resultados e Discussões

Para realizar as modelagens de Holt-Winters e de Box & Jenkins foi utilizada a validação *out of sample*, uma vez que para uma comparação de modelos, esse método é essencial. A validação *out of sample* consiste em reter parte da amostra no processo de identificação e estimação do modelo e então fazer uma previsão [9]. Nessa análise foram utilizados os dois últimos anos dos dados históricos como amostra de teste *out of sample*.

As modelagens foram feitas a partir do software Forecast Pro for Windows (FPW) e as análises e testes de normalidade, estacionariedade e de heterocedasticidade foram feitas a partir do software R. As estatísticas de aderência foram calculadas a partir do FPW, bem como as análises dos resíduos.

4.1 Modelagem Holt-Winters

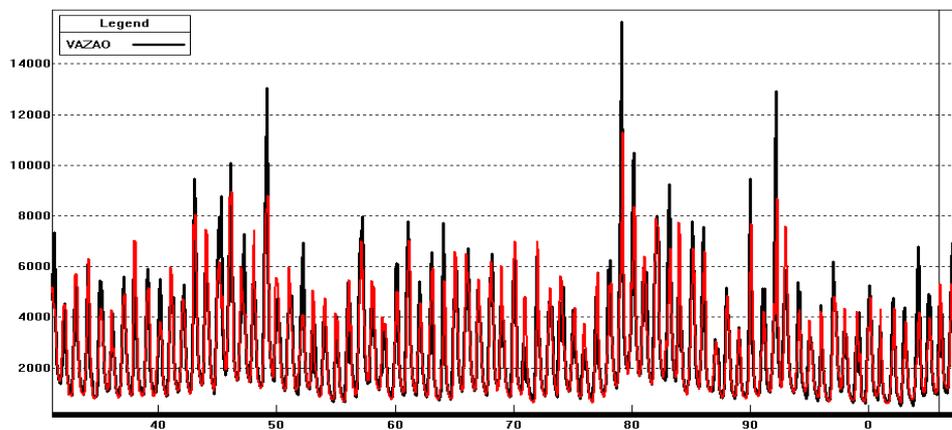
Ao utilizar o modo otimizado do FPW, que escolhe o melhor modelo com base naquele que minimiza a soma dos quadrados dos erros de previsão, o modelo de Amortecimento Exponencial de Holt-Winters sem tendência e com sazonalidade multiplicativa foi o modelo escolhido. Os parâmetros do modelo estão na Tabela 2.

Utilizando as primeiras 900 observações como amostra de treinamento (*in sample*) e as demais 24 como amostra de teste (*out of sample*), a série ajustada e os dados

Tabela 2: Parâmetros estimados na modelagem de Holt-Winters.

		Fatores Sazonais					
Nível	Sazonalidade	Jan	Fev	Mar	Abr	Mai	Jun
0,57563	0,14757	2,32504	2,04641	1,91141	1,40169	0,85929	0,66303
		Jul	Ago	Set	Out	Nov	Dez
		0,60053	0,53702	0,51096	0,54425	0,91205	0,91205

previstos foram obtidos a partir do modelo de amortecimento exponencial de Holt-Winters com os parâmetros da Tabela 2. A comparação gráfica da série observada (original) com a série modela (dados ajustados) via Holt-Winters está na Figura 3, na qual pode-se perceber que o modelo de Holt-Winters utilizado reproduziu bem o comportamento da série original deixando apenas os picos, que são fenômenos isolados fora de sua trajetória temporal. Pode-se perceber uma linha vertical no extremo direito da Figura que divide as análises *in sample* e *out of sample*.


Figura 3: Série original (preto) e série ajustada via Holt-Winters (vermelho).

A Figura 4 apresenta uma aproximação (zoom) das últimas 36 observações (os últimos 3 anos) para que se tenha uma visualização mais adequada desta região em relação aos dados modelados na amostra de treinamento e de teste.

Com a aproximação da Figura 4, pode-se perceber que, mesmo quando os dados são obtidos a partir dos dados ajustados, o modelo tem a capacidade de acompanhar os dados observados, comprovando ser este um bom modelo.

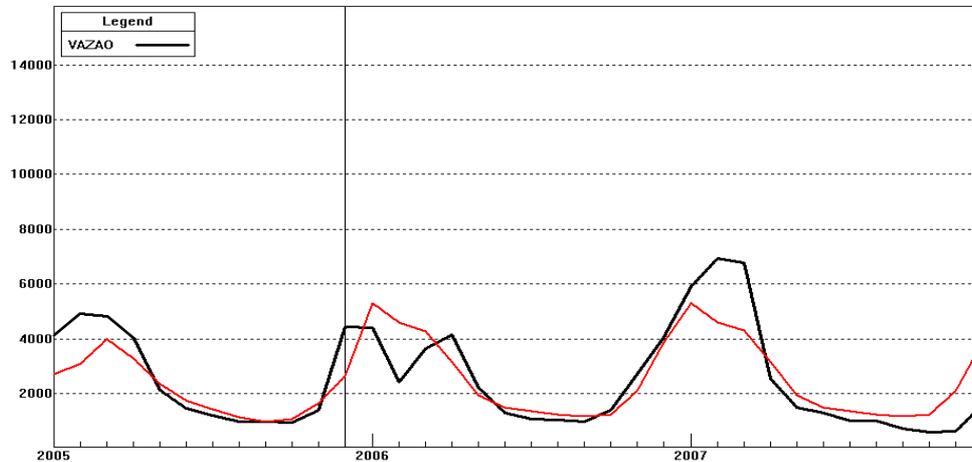


Figura 4: Aproximação (zoom) da Figura 3 na região de previsão (*out of sample*) do modelo ajustado via Holt-Winters.

Após a modelagem, uma análise dos resíduos foi feita a partir da análise da função de autocorrelação (FAC) dos mesmos. Para que o modelo seja adequado, os resíduos não devem ser serialmente correlacionados e para que isso ocorra, espera-se que os valores da FAC não ultrapasse um limite de $\pm 1,96/\sqrt{T} = \pm 1,96/\sqrt{924} = \pm 0,06$. A Figura 5 apresenta a FAC dos resíduos para este modelo.

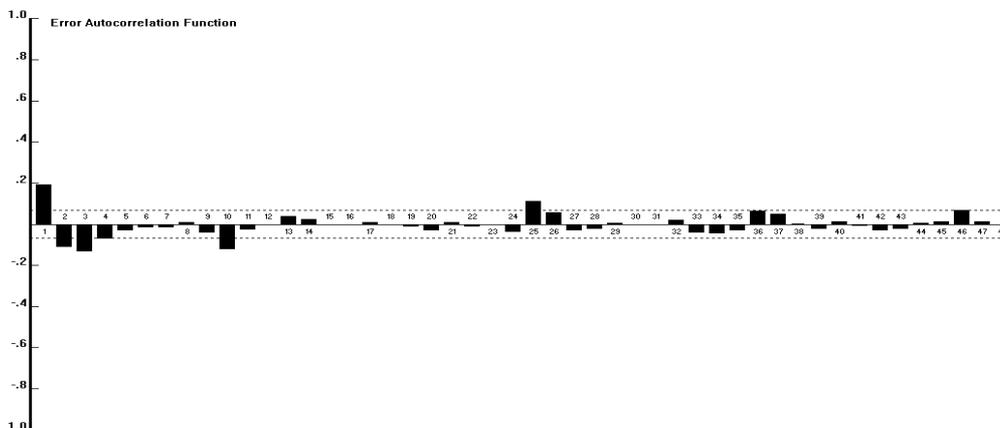


Figura 5: FAC dos resíduos para o modelo de Holt-Winters.

Conforme pode ser observado na Figura 5, os valores da Função de Autocorrelação dos resíduos estão, em sua maioria, dentro do limite esperado, o que pode confirmar que o modelo é adequado.

4.2 Modelagem de Box & Jenkins

Para a modelagem de Box & Jenkins é necessário fazer testes preliminares para que as condições de normalidade e estacionariedade sejam satisfeitas.

O teste de normalidade utilizado foi o de Shapiro-Wilk, implementado no Gretl e no R. Na ocasião, o teste retornou um valor- p igual a 0,0023, que é menor que o nível de significância de 0,05, portanto a série não possui uma distribuição Normal de probabilidade. Para obter a normalidade, uma transformação logarítmica foi aplicada aos dados originais e a série obtida foi testada novamente via Shapiro-Wilk. Para esta nova série, o teste retornou um valor- p igual a 0,4256, maior que o nível de significância, portanto confirmado a normalidade. A Figura 6 apresenta a série após a transformação logarítmica.

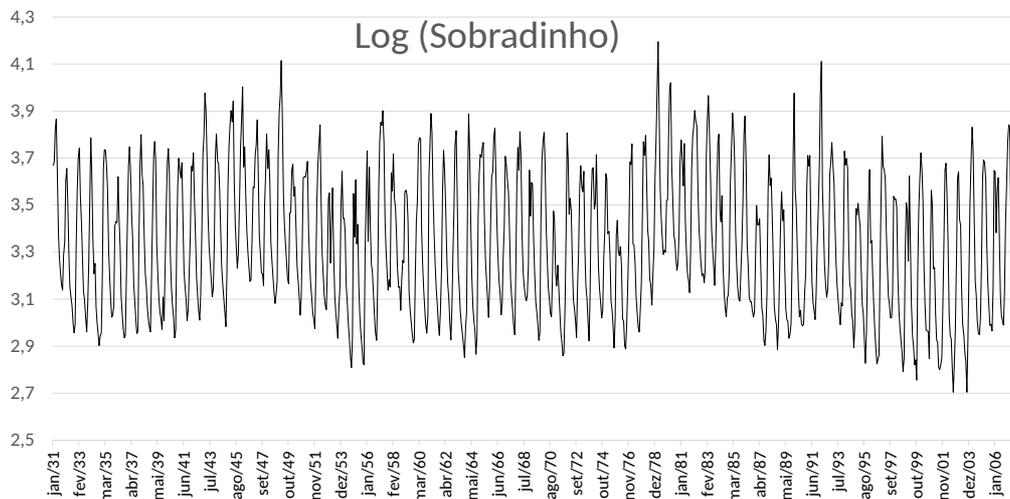


Figura 6: Série temporal da vazão média mensal de afluentes da hidrelétrica de Sobradinho com transformação logarítmica 1931-2007.

Ao realizar o teste de estacionariedade de Philips Perron, o valor- p obtido foi igual a 0,01, menor que o nível de significância de 0,05 indicando que a hipótese nula de existência de raiz unitária é rejeitada. Logo, verifica-se que não existe raiz unitária na série, concluindo que a série é estacionária não havendo então necessidade de fazer diferenças na série.

A figura 7 apresenta as funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) da série original de vazão de afluentes da Usina Hidrelétrica de Sobradinho. O comportamento destas funções sugerem um modelo autorregressivo.

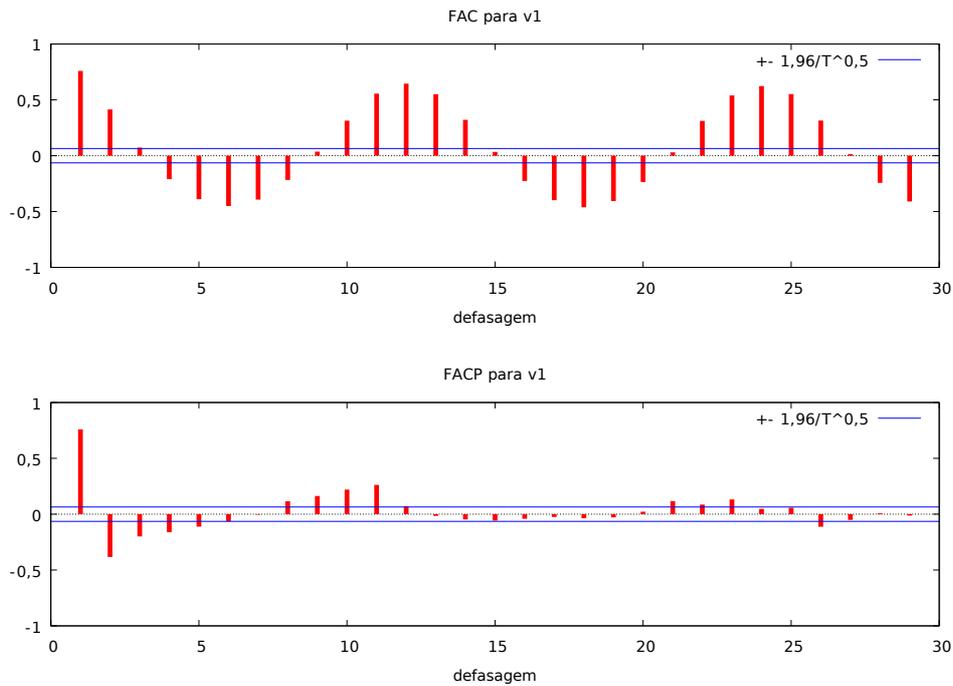


Figura 7: FAC e FACP da série original de vazão de aflunetes da Usina Hidrelétrica de Sobradinho

Depois de verificadas as condições necessárias e de fazer a transformação logarítmica, foi aplicado o modelo $SARIMA(1,0,0) \times (1,0,1)_{12}$ a partir do modo otimizado do FPW, que ajusta o modelo que minimiza a soma dos quadrados dos erros de previsão. Assim como, no modelo de Holt-Winters, o modelo de Box & Jenkins foi ajustado utilizando as 900 primeiras observações com amostra de treinamento *in sample* e as demais 24 como amostra de teste *out of sample*. A Figura 8 apresenta a comparação dos dados observados com os dados modelados via Box & Jenkins.

Na Figura 8 pode-se perceber que o modelo de Box & Jenkins utilizado reproduziu bem o comportamento da série original deixando apenas os picos, que são fenômenos isolados, fora de sua trajetória temporal, assim como no ajuste do modelo de Holt-Winters.

A Figura 9 apresenta uma aproximação (zoom) das últimas 36 observações (os últimos 3 anos) para que se tenha uma visualização mais adequada da região que

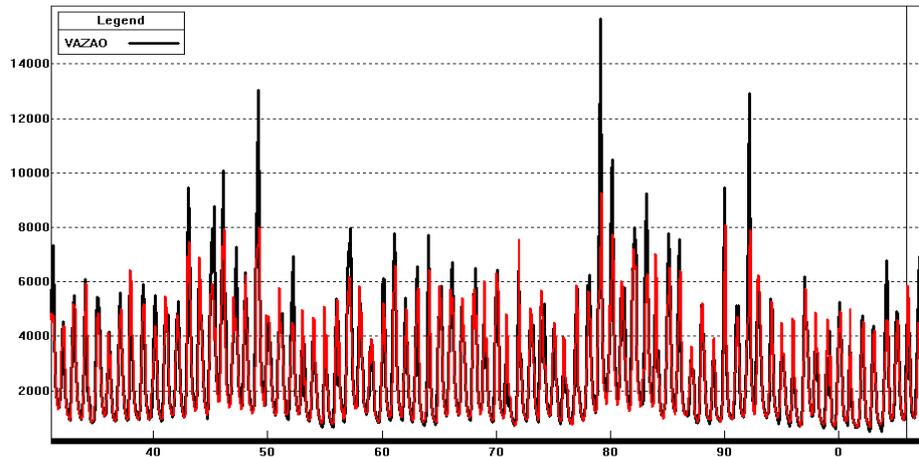


Figura 8: Série original (preto) e série ajustada via Box & Jenkins (vermelho)

envolve a análise *out of sample* em relação aos dados modelados na amostra de treinamento e de teste.

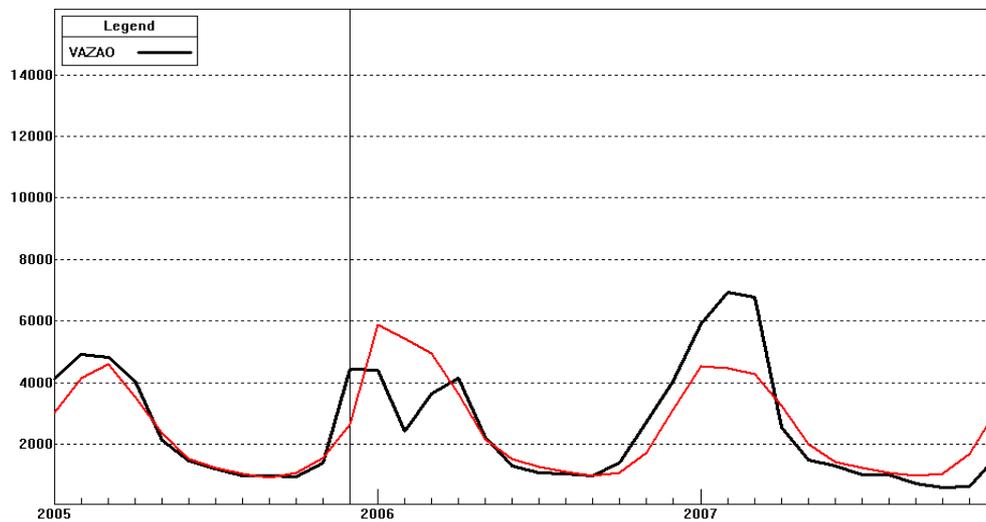


Figura 9: Aproximação (zoom) da Figura 8 na região de previsão (*out of sample*) do modelo ajustado via Box & Jenkins.

Da mesma forma como foi visto com o modelo de Holt-Winters, a aproximação da Figura 9 mostra que, mesmo quando os dados são obtidos a partir dos dados ajustados, o modelo tem a capacidade de acompanhar os dados observados, comprovando ser este também um bom modelo a ser utilizado para prever a vazão média

mensal da usina hidrelétrica de Sobradinho.

A análise dos resíduos também foi feita a partir da FAC dos resíduos de modo análogo ao modelos de Holt-Winters e o correlograma mostrado na Figura 10.

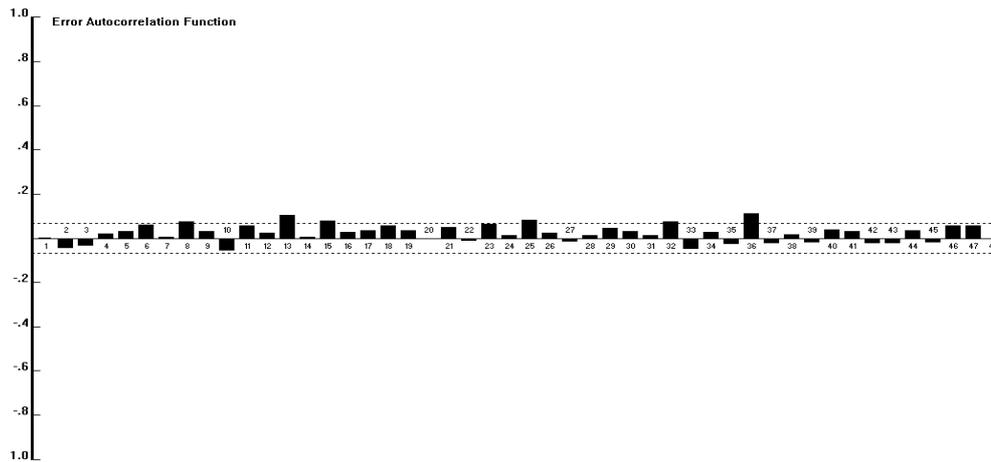


Figura 10: FAC dos resíduos para o modelo Box & Jenkins.

Assim como foi observado com o modelo de Holt-Winters e conforme pode ser observado na Figura 10, os valores da Função de Autocorrelação dos resíduos estão dentro do limite esperado, confirmando que o modelo de Box & Jenkins também é adequado.

4.3 Estatísticas de Aderência

A partir das análises gráficas e dos resíduos, pôde-se verificar que os modelos avaliados são adequados e apresentam eficiência na modelagem e previsão. No entanto, medidas estatísticas apropriadas podem verificar qual das modelagens é mais adequada ao caso particular. A Tabela 3 apresenta as estatísticas de aderência MAD , $MAPE$, $RMSE$, BIC e R^2 medidas *in sample*.

Tabela 3: Estatísticas de aderência dos modelos Holt-Winters e Box & Jenkins (*in sample*).

Abordagem	MAD	$MAPE$	$RMSE$	BIC	R^2
Holt-Winters	620,9	0,2013	1.064,0	1.077,0	0,7099
Box & Jenkins	223,2	0,1778	0,978	517,9	0,8738

De acordo com os resultados apresentados na Tabela 3, na análise *in sample* pode-se perceber que as estatísticas de aderência MAD , $MAPE$, $RMSE$ e BIC são minimizadas pelo modelo de Box & Jenkins enquanto o coeficiente de determinação R^2 é maximizado. Isso mostra que o desempenho deste modelo é melhor que o do modelo de Holt-Winters quando a análise é feita na amostra de treinamento, onde os valores ajustados são obtidos a partir dos dados amostrais.

A Tabela 4 apresenta as estatísticas MAD e $MAPE$ medidas *out of sample*.

Tabela 4: Estatísticas de aderência dos modelos Holt-Winters e Box & Jenkins (*out of sample*).

Abordagem	MAD	$MAPE$
Holt-Winters	599,7	0,269
Box & Jenkins	588,7	0,252

Assim como na Tabela 3, os resultados apresentados na Tabela 4 mostram que as estatísticas de aderência são minimizadas quando os dados são ajustados pelo modelo de Box & Jenkins nas análises *out of sample*, confirmando o desempenho deste modelo melhor que o de Holt-Winters também na amostra de teste, quando os valores ajustados são obtidos a partir dos próprios valores já ajustados.

Nesta segunda abordagem, o estudo indica o melhor modelo para se fazer previsão, uma vez que na amostra de teste ele faz uma previsão com comparação. Dessa forma, o modelo $SARIMA(1, 0, 0) \times (1, 0, 1)_{12}$ com transformação logarítmica é o modelo final estimado para esta série.

5 Conclusões

Neste artigo foram feitas modelagens de Holt-Winters e Box & Jenkins em uma série temporal de médias mensais de vazão de afluentes da Usina Hidrelétrica de Sobradinho. Na ocasião, os dados ajustados pelos modelos obtidos descreveram bem o comportamento da série original indicando que os modelos conseguiram verificar o mecanismo gerador da série. Analisando e comparando as estatísticas de aderência geradas em cada modelo, pode-se concluir que a modelagem feita através da metodologia de Box & Jenkins apresenta resultados mais significativos que a modelagem feita por Holt-Winters por conter melhor aderência aos dados observados tanto na análise *in sample* quanto na análise *out of sample*. Este resultado também mostra que uma análise adequada a partir dos testes de normalidade e de estacionariedade

podem levar a uma melhora na qualidade do ajuste, como foi feito no modelo de Box & Jenkins.

Referências

- [1] Agência Nacional de Energia Elétrica (Brasil). Atlas de energia elétrica do Brasil / Agência Nacional de Energia Elétrica. Brasília : ANEEL, 2002.
- [2] BOX, G. and Jenkins, G. **Time Series Analysis: Forecasting and Control**. 1 ed. San Francisco: Holden-Day, 1960.
- [3] FIGUEIREDO, N. e Blanco, C. Simulação de Vazões e Níveis de Água Médios Mensais para o Rio Tapajós Usando Modelos ARIMA. **Revista Brasileira de Recursos Hídricos**, Porto Alegre, v.19, n.3, p. 111-126. 2015.
- [4] GOLYANDINA, N., Nekrutkin, V., and Zhigljavsky, A. **Analysis of Time Series Structure: SSA and Related Techniques**. 1 ed. London: CHAPMAN & HALL/CRC, 2001.
- [5] HOLT, C. C. Forecasting Trends and Seasonal by Exponentially Weighted Averages. **Office of Naval Research Memorandum**, v.52. Reprinted in Holt, C. C. Forecasting Trends and Seasonal by Exponentially Weighted Averages. **International Journal of Forecasting**, v.20, n.1, p. 5-10. 1957.
- [6] LÚCIO, P. S. *et al.* Um Modelo Estocástico Combinado de Previsão Sazonal para a Precipitação no Brasil. **Revista Brasileira de Meteorologia**, São Paulo, v.25, n.1, p. 70-87. 2010
- [7] MENEZES, M. L., Cassiano, K. M., Souza, R. M., Junior, L. A. T., Pessanha, J. F. M. e Souza, R. C. Modelagem e Previsão de Demanda de Energia com Filtragem SSA. **Revista da Estatística da Universidade Federal de Ouro Preto**, Ouro Preto, v.3, n.2, p. 170-187. 2014.
- [8] MORETTIN, P. A. e Toloi, C. M. **Análise de Séries Temporais**. 1 ed. São Paulo: Edgard Blücher, 2006.
- [9] NAU, R. **Statistical Forecasting: Notes on Regression and Time Series Analysis**. Disponível em: <<http://people.duke.edu/~rnau/three.htm>>. Acesso em 23 jul. 2016.



- [10] NOGUEIRA, E. **Sobradinho**: De Volta ao Sertão. Disponível em: <<http://www.ebc.com.br/sobradinho>>. Acesso em 14 jul. 2016.
- [11] ONS - Operador Nacional do Sistema Elétrico. Histórico da Operação. Dados Hidrológicos/ Vazão. Disponível em: <http://ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_vazoes.aspx>. Acesso em 20 de jul. 2016.
- [12] SANTOS, P. G e Oliveira, T. F. Modelagem e Previsão de Vazão Afluente Média Mensal no Rio Tocantins, Usina Hidrelétrica Tucuruí-Pará, Amazônia, Brasil. **Biota Amazônia Open Jornal System**, Macapá, v.6, n.2, p. 09-16. 2015.
- [13] VASCONCELOS, A. S. e Costa, J. H. F. **Análise de Modelos de Séries temporárias para a previsão Mensal do ICMS do Estado do Maranhão**. 1 ed. São Luís: IMESC, 2008.
- [14] VERDE, P. **Tudo Sobre a Crise Energética Brasileira e Mundial**. Disponível em: <<http://www.pensamentoverde.com.br/meio-ambiente/tudo-sobre-crise-energetica-brasileira-e-mundial/>>. Acesso em 05 jul. 2016.
- [15] WINTERS, P. R. Forecasting Sales by Exponentially Weighted Moving Averages. **Management Science**, Catonsville, v.6, n.3, p.324-342. 1960.

Submetido em 5 fev. 2018

Aceito em 20 jun. 2018