Ciência e Natura, Santa Maria, v. 37 n. 4, 2015, p. 1–11
Revista do Centro de Ciências Naturais e Exatas - UFSM
ISSN impressa: 0100-8307 ISSN on-line: 2179-460X

CIÊNCIA e NATURA 35 anos

Análise comparativa de modelos de séries temporais para modelagem e previsão de regimes de vazões médias mensais do Rio Doce, Colatina - Espírito Santo

Comparative analysis of models for times to series modeling and forecasting of scheme of average monthly streamflow of the Doce River, Colatina Espirito Santo, Brazil

Wanderson de Paula Pinto*[1], Gemael Barbosa Lima†[2] e Juliano Bras Zanetti‡[3]

[1]Doutorando em Engenharia Ambiental, PPGEA - UFES. Faculdade da Região Serrana - FARESE e Centro Universitário do Espírito Santo - UNESC, ES, Brasil
[2]Faculdade da Região Serrana - FARESE e Centro Universitário do Espírito Santo - UNESC, ES, Brasil
[3]Faculdade da Região Serrana - FARESE, ES, Brasil

Resumo

Este trabalho apresenta um estudo de séries temporais de vazões médias mensais utilizando modelos estocásticos da classe SARIMA (Autorregressivo Integrado e de Médias Móveis Sazonal). A modelagem hidrológica é uma importante ferramenta no planejamento e gerenciamento de programas de recursos hídricos de bacias hidrográficas. Objetivou-se com o presente artigo analisar, comparativamente, modelos de séries temporais, bem como realizar previsões das vazões médias mensais para o rio Doce/Colatina/ES. A metodologia aplicada neste trabalho baseou-se na proposta por Box e Jenkins. A identificação da ordem dos modelos foi feita através da análise gráfica das funções de autocorrelação (ACF) e autocorrelação parcial (PACF). Dentre os modelos candidatos foram selecionados aqueles que obtiveram os menores valores dos critérios de seleção adotados (AIC, BIC, EQM e EAP). O modelo escolhido foi o $SARIMA(1,1,1)(1,1,2)_{12}$*, que representou bem a dinâmica da série em estudo e apresentou os melhores resultados numéricos nas fases de calibração e validação. O referido modelo foi utilizado para realizar previsões para os meses de 2014 e 2015. As previsões de vazão média mensal apresentadas, obtidas com o modelo considerado, podem ser usadas para analisar o regime hidrológico do rio Doce, principalmente para os meses críticos de 2015, e com isso propor medidas preventivas para a diminuição dos impactos causados, não só ao meio ambiente, mas na economia dessa Cidade.*

Palavras-chave: *Recursos hídricos, SARIMA, sazonalidade, modelagem hidrológica, séries temporais.*

Abstract

This work shows a time series study of the average monthly flow using stochastic model of the SARIMA class (Seasonal Autoregressive Moving Average). The hydrology model is an important tool in the planning and management of water resources programs. This article aimed to analyse, comparatively, time series models, as well as to realize forecasting of average monthly flow taken Doce River as watershed, it's situated in Colatina municipality/ES. The applied methodology in this work was based in the proposal by Box and Jenkins. The identification of the order of the models was made by graphic analysis of the autocorrelation function (ACF) e partial autocorrelation (PACF). Among possible models was selected those that obtained minor value of the following parameters: AIC, BIC, EQM and EAM. The chosen model was $SARIMA(1,1,1)(1,1,2)_{12}$*, that represented well the dynamic of the series studied and showed the best numerical results in both calibration and validation stages. The chosen model was used to make forecasting to months of 2014 and 2015. The average monthly flow forecasting showed, obtained with the considered models, can be used to analyse hydrology regime of the Doce river, mainly to the critical months of 2015, as well as it can propose preventive measures to reduce both environmental and economic impacts in this municipality.*

Keywords: *Water resources, SARIMA, seasonality, hydrology modelling, times series.*

*W. P. Pinto: wandersondpp@gmail.com
†G. B. Lima: gemaelbl@yahoo.com.br
‡J. B. Zanetti: zanettijb@yahoo.com.br
Recebido: 05/03/2015 Revisado: 19/07/2015 Aceito: 05/08/2015

1 Introdução

A água é de fundamental importância para a manutenção de qualquer vida na Terra. O Brasil é um país privilegiado, pois o mesmo detém 8% das reservas hídricas mundiais, no entanto, somente a bacia do rio Amazonas corresponde corresponde a 72% das reservas brasileiras onde concentra-se apenas 5% da população nacional, por outro lado, o nordeste contém apenas 3,3% das reservas hídricas e sua população representa 27% . Logo, a água não é bem distribuída no espaço e no tempo de forma uniforme. De acordo com Setti et al. (2001) a ideia de abundância serviu durante muito tempo como suporte à cultura do desperdício da água disponível, à sua pouca valorização como recurso e ao adiamento dos investimentos necessários à otimização de seu uso.

Os problemas de escassez hídrica no Brasil decorrem, fundamentalmente, da combinação entre o crescimento exagerado das demandas localizadas e da degradação da qualidade da água. Esse cenário é devido ao crescimento desordenado da industrialização, urbanização e expansão agrícola. Nesse sentido, é de suma importância a adoção de uma gestão dos recursos hídricos eficiente para minimizar as pressões sobre os mesmos, bem como a garantia do desenvolvimento sustentável.

O estudo de vazões é fundamental para o entendimento da dinâmica do ciclo hidrológico em bacias hidrográficas, pois representa a resposta da bacia às alterações climáticas e antrópicas, como eventos extremos de precipitação e mudanças do uso e ocupação do solo, respectivamente. As vazões podem ser caracterizadas por mínimas, médias e máximas, sendo que as duas primeiras são de fundamental importância para o estudo acerca da disponibilidade de água, assim como concessões de outorga de direito de uso da água, ferramentas essas de capital importância para se ter uma gestão adequada dos recursos hídricos (Oliveira, 2013).

A previsão de vazões tem uma significativa importância no gerenciamento dos recursos hídricos. Essas previsões podem ser realizadas tanto em curto prazo, geralmente utilizando modelos determinísticos, quanto a longo prazo, utilizando modelos estocásticos. As previsões a longo prazo trazem benefício direto à estimativa de afluências aos reservatórios hidrelétricos, o que permite planejar a sua melhor operação (Collischonn et al., 2005). Já as previsões de vazão de curto prazo são usadas com frequência na operação de sistemas de recursos hídricos destinados aos usos múltiplos, a saber: controle de cheia, geração de energia elétrica, navegação, dentre outros (Barbosa e Fill, 2001). Além disso, a previsão da disponibilidade hídrica é importante para outros usos como irrigação, abastecimento, navegação, etc (Collischonn et al., 2005).

Segundo Bayer et al. (2012), dentre os modelos estocásticos, os modelos de séries temporais da classe ARIMA (autorregressivo integrado de médias móveis) possuem grande destaque. Ainda segundo o autor, como série temporal é um conjunto de observações ordenadas no tempo, conclui-se que os processos que compõem o ciclo hidrológico como a chuva, vazão, evaporação, infiltração, dentre outros, por serem objetos de medições sistemáticas em tempos discretos, podem ser tratados como séries temporais.

Em nível internacional, as previsões de vazões a partir de modelos de séries temporais têm sido largamente utilizadas. Entre os estudos destacam-se o desenvolvido por Lohani et al. (2012) que empregaram modelos autorregressivos juntamente com sistemas de interferência fuzzy na previsão de vazões mensais. Wu e Chau (2010) utilizaram modelos ARIMA e redes neurais na previsão de vazões em diferentes bacias hidrográficas na China. Ghanbarpour et al. (2010) testaram a performance dos modelos ARIMA e Dessazonalizada Autorregressivo Médias Móveis (DARMA), em que o primeiro foi melhor que o segundo na aplicação da previsão de vazão em bacias hidrográficas cársticas. Koutroumanidis et al. (2009) desenvolveram modelos para a previsão de vazão para a bacia hidrográfica do rio Nestos na Bulgária com proposta híbrida baseada em modelos ARIMA.

No Brasil, Figueredo e Blanco (2014) apresentaram análise de previsão de vazão média mensais, com antecedência de 24 meses, para a bacia do rio Tapajós, PA, utilizando modelos ARIMA e, a partir dos resultados obtidos concluíram que a metodologia adotada é uma ferramenta interessante podendo ser utilizada para analisar regime hidrológico. Meller et al. (2012) descreveram um método empírico de assimilação de dados com a modelagem hidrológica distribuída, com o intuito de prever as cheias do rio Piracicaba, os resultados apontaram que o método de assimilação tem impactos positivos nos resultados das previsões. Lucas et al. (2009) utilizaram dois modelos, um determinístico e outro estocástico tipo ARIMA, com o objetivo de simular a vazão média mensal das sub-bacias da região hidrográfica do rio Xingu/PA. Mine e Tucci (1999) utilizaram um modelo ARIMA, conjuntamente com um modelo determinístico chuva-vazão denominado IPH-II, na previsão de vazões afluentes à estação Usina Foz de Areia, no rio Iguaçu/PR. Ambos modelos obtiveram bom desempenho.

Diante do exposto, o presente artigo teve por objetivo analisar, comparativamente, modelos de séries temporais, bem como, realizar previsões para as vazões médias mensais para o rio Doce no município de Colatina/ES.

2 Materiais e métodos

2.1 Área de estudo

A bacia hidrográfica do rio Doce (Figura 1) apresenta área aproximada de 83.400 km^2, dos quais 86% pertencem ao Estado de Minas Gerais e o restante ao Estado do Espírito Santo. Abrange, total ou parcialmente, áreas de 228 municípios, sendo 202 em Minas Gerais e 26 no Espírito Santo. O rio Doce, com uma extensão de 853 km, tem como formadores os rios Piranga e Carmo, cujas nascentes estão situadas nas encostas das serras da Mantiqueira e Espinhaço, onde as altitudes atingem cerca de 1.200 m e de água no mar em Regência no município de Linhares/ES. Seus principais afluentes são: Santo Antônio, Suaçuí Grande, Casca, Matipó, Caratinga-Cuieté e Manhuaçu, em Minas Gerais e Pancas, São José e Guandu no Espírito Santo.

De acordo com a classificação de Köppen, identificam-se três tipos climáticos na bacia, a saber: o clima tropical de altitude com chuvas de verão e verões frescos, presente nas vertentes das serras da Mantiqueira e do Espinhaço e nas nascentes do rio Doce; clima tropical de altitude com chuvas de verão e verões quentes, presentes nas nascentes de seus afluentes; e clima quente com chuvas de verão, presentes nos trechos médio e baixo do rio Doce e de seus afluentes. A precipitação média anual na bacia varia de 1.500 mm, nas nascentes localizadas nas serras da Mantiqueira e do Espinhaço, a 900 mm, na região da cidade de Aimorés/MG, voltando a crescer em direção ao litoral.

O solo predominante na bacia é o Latossolo Vermelho Amarelo e Podizólico Vermelho Amarelo. Outros tipos de solo que ocorrem em menor percentagem são o latossolo húmico, solos litólicos, cambissolos e afloramentos de rochas, dentre outros.

A bacia hidrográfica do rio Doce é originalmente coberta por Mata Atlântica. A intensa devastação restringiu o revestimento florístico originário basicamente à área do Parque Estadual do rio Doce. As demais matas correspondem a uma vegetação que sofreu influência antrópica intensa, constituindo-se em vegetação secundária. Estima-se que menos de 7% da área possui hoje cobertura vegetal (C.B.H. Doce, 2014). Destas, menos de 1% encontra-se em estágio primário (C.B.H. Doce, 2014). Na zona rural encontram-se vastas áreas em estado avançado de desertificação, lagoas eutrofizadas, nascentes desprotegidas e processos erosivos. Da cobertura vegetal original, mais de 90% foi extinta. Do restante, menos de 1% encontra-se em estágio primário (C.B.H. Doce, 2014).

A economia está baseada principalmente nas seguintes atividades: (I) agricultura com destaque para a pecuária de leite e corte, suinocultura, café, cana-de-açúcar, hortifrutigranjeiros e cacau; (II) indústria: siderurgia, metalurgia, mecânica, química, alimentícia, álcool, têxtil, curtume, papel e celulose; e (III) mineração: ferro, ouro, bauxita, manganês, rochas calcáreas e pedras preciosas; e (IV) geração de energia elétrica com uma capacidade de geração de cerca de 4.055 MW.

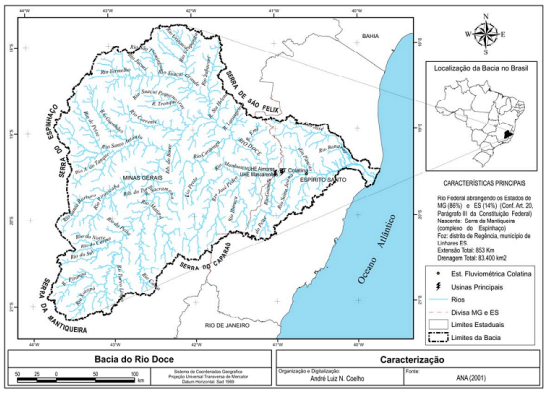

Figura 1: Localização da Bacia Hidrográfica do rio Doce e área de drenagem.
Fonte: Coelho (2007).

2.2 Dados

A série histórica de vazões médias mensais utilizadas no presente artigo foi obtida junto ao Sistema de Informações Hidrológicas (HIDROWEB) mantido pela Agência Nacional de Água (ANA), disponível no site http://hidroweb.ana.gov.br/. A estação de monitoramento escolhida foi a de Colatina, localizada no município de mesmo nome, cujo código é 56994500 e coordenadas geográficas 19º 32'00"S e 40º37'47"W, com área de drenagem de 76.400 km^2.

A estação de monitoramento de Colatina opera desde o ano de 1939 e, para a utilização dessa pesquisa, utilizou-se dados até de 2013, resultando assim em uma série histórica com dados de 74 anos. Tal série caracteriza-se por haver dados faltantes, uma vez que durante os anos de 1978 a 1982, a estação ficou inoperante. No entanto, o cálculo dessas obervações foi baseada na metodologia estudada por Pinto (2013).

2.3 Modelos

Uma série temporal é um conjunto de observações de qualquer fenômeno aleatório ordenadas no tempo. A análise de séries temporais consiste em procurar alguma relação de dependência existente temporalmente nos dados, identificando o mecanismo gerador da série com objetivo de extrair periodicidades relevantes nas observações, descrever o seu comportamento e fazer previsões (Bayer e Souza, 2010).

Seja Z_t $(t = 1, 2, 3, ...)$ um processo que satisfaz a equação

$$\phi_p(B)Z_t = \theta_q(B)\epsilon_t, \quad (1)$$

em que ϵ_t é ruído branco (RB), definido como uma sequência de variáveis aleatórias não-correlacionadas com média zero e variância constante ao longo do tempo (Wei (2006), i. e., $\epsilon_t \sim RB(0,\sigma_\epsilon^2)$, $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ e B é o operador de defasagem definido como $B^k Z_t = Z_{t-k}$. O processo Z_t definido em (1) é chamado de processo autorregressivo e de médias móveis de ordens (p,q), abreviado por $ARMA(p,q)$.

O processo (1) é estacionário e invertível se as raízes de $\phi_p(B)$ e $\theta_q(B)$ não são comuns e encontram-se fora do círculo unitário. Para maiores detalhes consultar em Brockwel e Davis (2002), Wei (2006) e Box et al. (2008).

Segundo Bayer e Souza (2010) geralmente as realizações encontradas na prática não apresentam a característica de estacionariedade, sendo necessária a utilização de transformações para torná-la estacionária. O procedimento comumente utilizado é o processo de diferenciação da série. Se a série torna-se estacionária após d diferenças a série é dita ser integrada de ordem d. Sendo assim, o modelo $ARMA$ integrado passa a ser denominado de modelo $ARIMA$. Uma série temporal Z_t é dita um processo $ARIMA(p,d,q)$ se,

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)\epsilon_t, \quad \epsilon_t \sim RB(0,\sigma_\epsilon^2) \quad (2)$$

em que d é a ordem de integração, sendo dada pelo menor número de diferenças necessárias para se alcançar a estacionariedade. Portanto, Z_t é processo não-estacionário que depois de diferenciado $(\nabla^d Z_t (d \geq 1)$ e $\nabla = 1 - B)$ se transforma em processo estacionário e invertível $ARMA(p,q)$.

Os modelos ARIMA exploram a autocorrelação entre os valores da série em instantes sucessivos, mas, a série também pode apresentar autocorrelação para uma estação de sazonalidades. Os modelos que contemplam as séries que apresentam autocorrelação sazonal são conhecidos como SARIMA.

Modelo $SARIMA(p,d,q) \times (P,D,Q)_s$. Seja $Z_t = \{Z_t; t \in \mathbb{Z}\}$ um processo linear com representação dada por:

$$\Phi(B^s)\phi(B)\nabla^d Z_t = \Theta(B^s)\theta(B)\epsilon_t, \quad (3)$$

em que s é chamado de período sazonal do processo e $\epsilon_t \sim RB(0,\sigma_\epsilon^2)$. O operador ∇^d, em que $\mathbf{d} = (d,D)$ e d, D são números inteiros não negativos e representam o número de diferenças simples e sazonais, respectivamente, aplicadas sobre o processo Z_t , é definido por:

$$\nabla^d = (1-B)^d(1-B^s)^D. \quad (4)$$

Temos que $\Phi(z^s) = 1 - \sum_{i=1}^{p} \Phi_i z^{is}$, $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$, $\Theta(z^s) = 1 - \sum_{k=1}^{Q} \Theta_k z^{ks}$ e $\theta(z) = 1 - \sum_{l=1}^{q} \theta_l z^l$ são polinômios de ordem $P, p, Q, q \in \mathbb{N}$, respectivamente, com $z \in \mathbb{C}$, em que, $\mathbb{C}$ representa o conjunto dos números complexos e $\{\Phi_i\}, \{\phi_j\}, \{\Theta_k\}, \{\theta_l\}$ são sequências de números reais. O processo Z_t com representação dada em (4) é denominado $ARIMA$ sazonal multiplicativo $(SARIMA)$ de ordem $(p,d,q) \times (P,D,Q)_s$. Para detalhes ver, Brockwel e Davis (2002), Morettin e Toloi (2006), Wei (2006), Box et al. (2008), entre outros.

2.4 Metodologia de modelagem

No estudo de séries temporais o objetivo principal é encontrar um modelo apropriado que descreva o fenômeno gerador de cada série estudada. Nesta seção, aborda-se as etapas da metodologia de Box e Jenkins (1970) para o processo de escolha do melhor modelo que se ajusta ao conjunto de dados. Neste trabalho, a série temporal Z_t, representará as vazões médias mensais.

A metodologia de Box e Jenkins aplicada neste trabalho está dividida nas seguintes etapas: (i) Identificação; (ii) estimação; (iii) previsão. A identificação do modelo a ser ajustado aos dados é uma dadas fases mais complicadas. Para isso, serão utilizados os critérios de seleção de modelos.

O AIC *(Akaike Information Criterion)* (Akaike, 1973) é um critério de seleção comumente utilizado. Utilizando os estimadores de máxima verossimilhança para os parâmentros do modelo, em que $l(\widehat{\xi})$ é a função de log-verossimilhança maximizada e k é o número de parâmetros do modelo, o AIC é dado por:

$$AIC = -2l(\widehat{\xi}) + 2k.$$

Em uma perspectiva bayesiana, Schwarz (1978) e Akaike (1978) introduziram o BIC *(Bayesian Information Criterion)* dado por:

$$BIC = -2l(\widehat{\xi}) + k\log(n),$$

em que n é o número de observações da amostra.

Com base nesses critérios de informação, pode-se ajustar diversos modelos e escolher aquele que obtever o menor valor para o critério de informação.

Além dos critérios de seleção de modelos outras duas medidas de qualidade de ajuste foram utilizadas para auxiliar na seleção domodelo que melhor se ajusta à série temporal em estudo e, principalmente, na avaliação da qualidade do ajuste. As medidas utilizadas neste trabalho foram o erro quadrático médio (EQM) e erro absoluto médio (EAM), definidas por:

Tabela 1: Medidas descritivas da variável vazão.

Medidas descritivas	Valor
Média (m^3/s)	856,90
Mediana (m^3/s)	650,60
Desvio padrão (m^3/s)	609,81
Coeficiente de variação %	71,16
Valor máximo (m^3/s)	4237,00
Valor mínimo (m^3/s)	72,22
Assimetria	1,77
Curtose	3,95

$$EQM = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \widehat{Z}_i)^2$$

e

$$EAM = \frac{1}{N}\sum_{i=1}^{N}\left|Z_i - \widehat{Z}_i\right|$$

em que Z_i e $\widehat{Z}_i$ são, respectivamente, os valores observados e previstos no instante i.

A previsão de passos à frente é o cálculo do valor esperado de uma futura observação condicionada a valores passados e ao valor presente da variável, ou seja,

$$\widehat{Z}_t(h) = E(Z_{(t+h)} \mid Z_t, Z_{(t-1)}, \cdots), \qquad (5)$$

em que $\widehat{Z}_t(h)$ é o valor estimado da variável Z_t no horizonte de h períodos de tempos futuros com base em t observações passadas. O valor de Z_{t+h} é calculado com o modelo que melhor se ajusta aos dados.

Um dos principais objetivos deste trabalho foi a comparação entre os modelos quanto ao seu desempenho no processo de ajustamento aos dados em estudo e uma das formas de verificar essa qualidade é o estudo de previsão.

3 Resultados e discussões

3.1 Análise descritiva dos dados

Para um entendimento preliminar da variável vazão estudada é apresentado, na Tabela 1, algumas medidas descritivas. A análise apresentada corresponde à série de médias mensais de vazões do período de janeiro de 1939 a dezembro de 2013 do rio Doce. Todas as análises foram feitas utilizando o *software* livre R (R Development Core Team, 2014).

A vazão média do período foi de 856,90 m^3/s, com desvio padrão de 609,81 m^3/s. O desvio padrão e o coeficiente de variação alto indicam uma média pouco representativa. Essa pouca representatividade pode ser explicada pela grande variabilidade intra-anual, caracterizada como sazonalidade (Bayer et al., 2012).

A menor vazão média mensal foi registrada em setembro de 2010, 72,22 m^3/s, e a maior foi de 4237,00 m^3/s em fevereiro de 1949. A distribuição da série em estudo possui assimetria positiva igual a 1,77, que pode ser visualizada na Figura 2, e curtose igual a 3,95 o que indica que a distribuição tem caudas mais pesadas do que a normal.

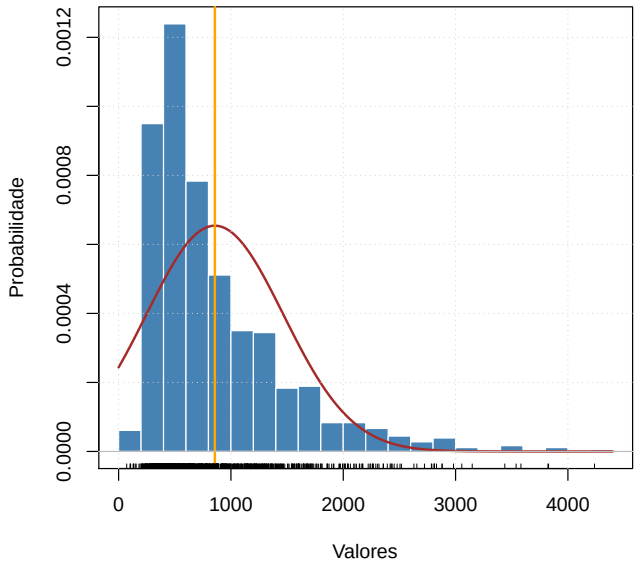

Figura 2: Histograma dos dados em estudo.

Na Figura 3 apresentam-se as séries original, com dados faltantes, e com dados imputados. Utiliza-se a série imputada para o estudo proposto neste artigo. Como pode ser observada, Figura 3, a série apresenta um padrão intra-anual de variabilidade, com períodos de cheias seguidos por períodos de estiagem, o que caracteriza a propriedade de sazonalidade.

Para confirmar a presença da sazonalidade e identificar tendências são apresentados os gráficos de sazonalidade, Figura 4, e da decomposição da série de vazões médias mensal, Figura 5. Com a análise da Figura 3 percebe-se que a série possui um período de cheias, de novembro a abril, e um período de seca, de maio a outubro, sugerindo a propriedade de sazonalidade, que foi confirmada com a aplicação do teste de Kruskal-Wallis (Morettin e Toloi, 2006). Percebe-se ainda que a série é não estacionária na média e na variância. A não estacionariedade na média foi confirmada com o uso do teste de Dickey-Fuller (Dickey e Fuller, 1981).

Já na Figura 5, nota-se a presença de valores atípicos, o que pode justificar o aumento na variabilidade dos

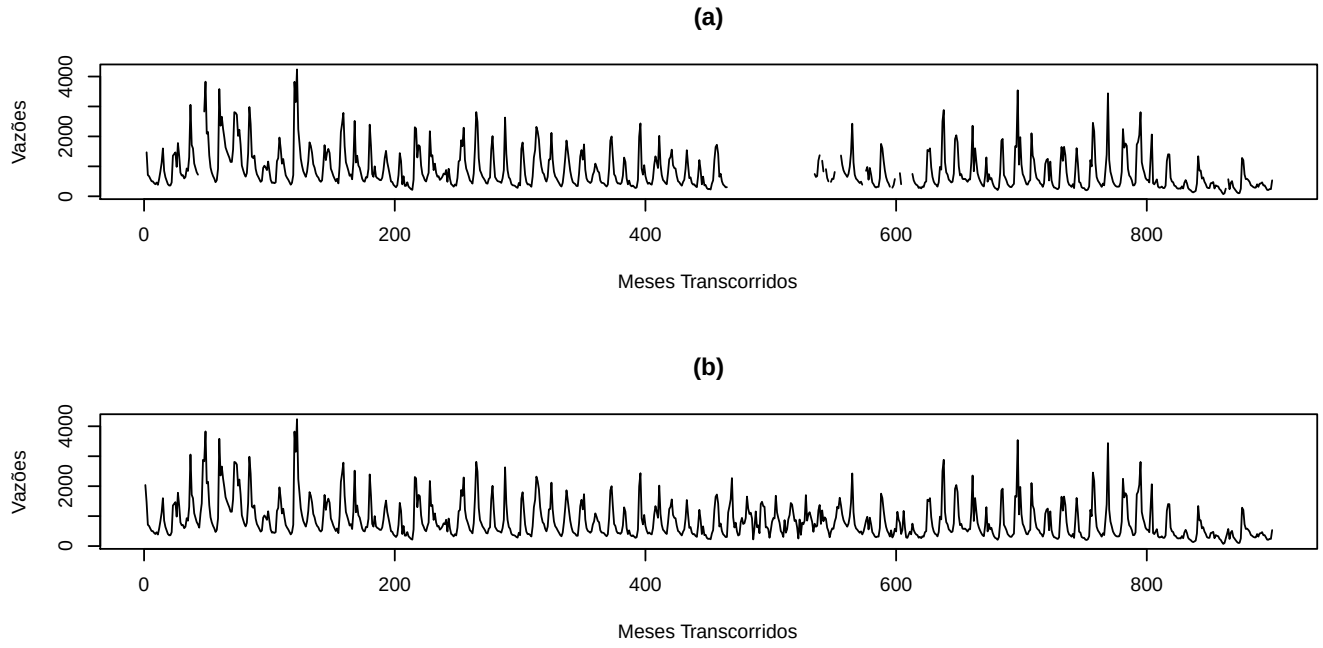

Figura 3: Gráfico da série de vazões médias mensais (a) com dados faltantes, (b) imputada.

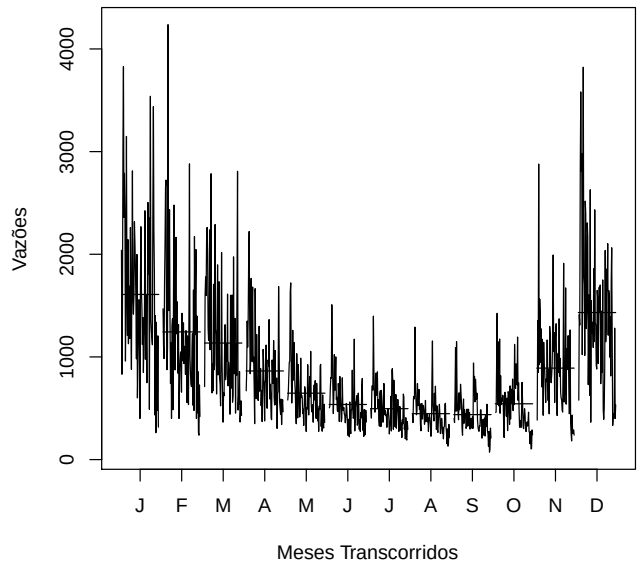

Figura 4: Análise gráfica da sazonalidade da série de vazões médias mensais.

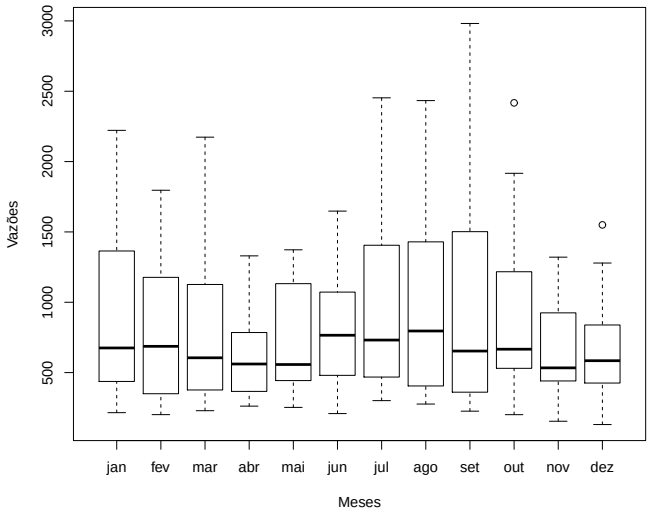

Figura 5: Box plot dos dados.

dados de vazão, porém o tratamento desse tipo de dados não é o objetivo deste trabalho. Nesse contexto, as metodologias apresentadas por Reisen et al. (2008) podem ser consideradas para futuros trabalhos de análise de séries temporais de vazões médias mensais.

Da Figura 6, pode-se concluir que a série possui sazonalidade e uma pequena tendência, confirmada através do teste do sinal, verificando que além do comportamento intra-anual há uma pequena evidência de decréscimo da variável ao longo do tempo.

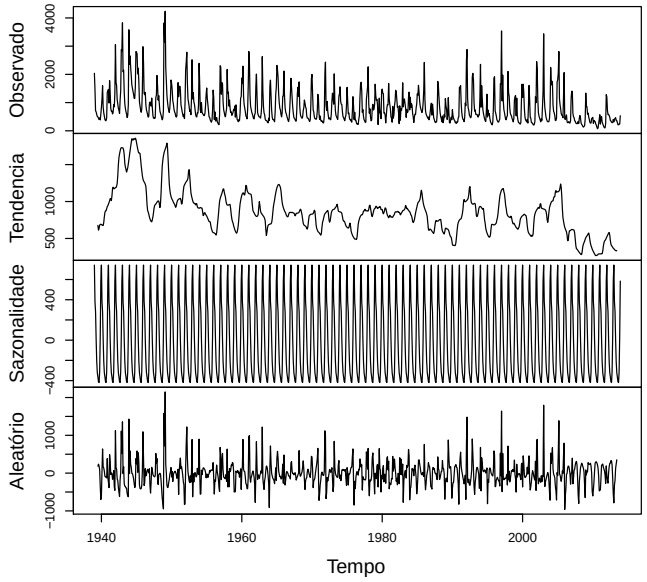

Figura 6: Decomposição da série temporal em componentes de sazonalidade, de tendência e de aleatoriedade.

3.2 Modelagem e previsão

3.2.1 Ajuste dos modelos

Com objetivo de realizar a avaliação das previsões, o conjunto de dados foi dividido em dois subconjuntos. O primeiro formado pelos dados de vazões médias mensais de 1939 até 2012, utilizados no processo de modelagem e o segundo, formado pelos dados de 2013, que foi utilizado para o cálculo das estatísticas de avaliação das previsões um passo à frente.

Conforme verificado a seção 3.1 a série apresenta o comportamento sazonal. Observou-se também que a série não é estacionária na variância. Dessa forma, para estabilizar o aumento da variância das observações realizam-se testes estatísticos para examinar a aplicação de uma transformação. No caso em estudo, considerou-se a transformação sugerida por (Box e Cox, 1964), no qual se estima o λ, baseadas nesta estimativa define-se o tipo de tranformação. Segundo Reisen et al. (2008) frequentemente, a transformação não só estabiliza a variância, mas também melhora na aproximação da distribuição dos dados para a distribuição normal.

A série transformada é dada por

$$T(Z_t) = \begin{cases} Z_t^{\lambda} & \text{se} \quad \lambda \neq 0 \\ ln(Z_t) & \text{se} \quad \lambda = 0 \end{cases} \tag{6}$$

Em que λ é o parâmetro de transformação e escolhido no intervalo $[-1, +1]$ e os valores mais correntes correspondem às transformações propostas por (Box e Cox, 1964).

Tabela 2: Modelos SARIMA ajustados e seus valores de AIC, BIC e das medidas de qualidade.

(p,d,q)(P,D,Q)	AIC	BIC	EQM	EAM
$(3,0,0)(0,1,2)_{12}$	547,45	571,32	0,1041	0,2297
$(3,0,0)(1,1,2)_{12}$	549,68	583,11	0,1038	0,2298
$(0,0,1)(2,1,2)_{12}$	728,43	757,08	0,1285	0,2715
$(1,0,2)(0,1,3)_{12}$	542,75	576,18	0,1027	0,2288
$(1,0,1)(2,1,3)_{12}$	545,25	583,45	0,1029	0,2312
$(1,1,1)(2,0,2)_{12}$	544,24	577,75	0,1020	0,2311
$(1,1,1)(1,1,2)_{12}$	**541,84**	**570,48**	**0,0994**	**0,2252**
$(0,0,3)(2,0,2)_{12}$	602,22	645,32	0,1111	0,2476

Dessa forma, como o λ obtido foi $-0{,}0789$, ou seja, próximo de zero, o logaritmo natural foi a transformação adequada para estabilizar a variância.

Para o procedimento de modelagem, considerou-se a série transformada. A identificação do modelo para o cálculo de previsões baseou-se na metodologia sugerida por Box e Jenkins (seção 2.4), para detalhes ver Brockwel e Davis (2002) e Box et al. (2008).

Segundo Box et al. (2008) para fazer a identificação dos modelos é essencial estudar o comportamento da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP). A Figura 7 mostra a FAC (a) e FACP (b) da série e a Figura 8 mostra a FAC (a) e FACP (b) da série diferenciada de ordem 12.

Através da análise da FAC e FACP, Figura 7, percebe-se a autocorrelação significativamente diferente de zero nas defasagens 1 e 2, o que evidencia que existe correlação serial. Além disso, percebe-se que a FAC tem comportamento senoidal, indicando um processo autorregressivo e sazonal. Na Figura 8, pode-se perceber características de um modelo estacionário, autorregressivo de ordem 1, pela análise da FAC, e de médias móveis de ordem 2, pela análise da FACP. Essa análise preliminar indica que o modelo a ser escolhido dever considerar uma diferença sazonal.

Com a série diferenciada sazonalmente, realizam-se os procedimentos de identificação e estimação dos parâmetros do modelo. Com os valores escolhidos pela indicações dos correlograma, apresentados nas Figuras 7 e 8 testaram-se alguns modelos com estas combinações. Feito isso, a identificação das ordens autorregressivas e de médias móveis, do melhor modelo para representar os dados, baseou-se nos critérios de informação de Akaike (AIC), Bayesiano (BIC) e nas medidas de qualidade do erro quadrático médio (EQM) e erro absoluto médio (EAM). Os modelos ajustados com seus valores de AIC, BIC e das medidas de qualidade são apresentados na Tabela 2

Observou-se, Tabela 2, que os menores valores ob-

tidos para os critérios de informação e medidas de qualidade foram evidenciados para um modelo $SARIMA(1,1,1)(1,1,2)_{12}$. Sendo assim, esse torna-se o modelo de interesse.

Para verificar a adequação do modelo, a teoria estatística estabelece suposições básicas das propriedades do mesmo, tais como, não correlação e normalidade dos erros (Wei, 2006). A Figura 9 mostra o histograma dos resíduos, os valores da FAC amostral e FACP dos resíduos do modelo ajustado. Observa-se que não há evidência contra a hipótese de ausência de autocorrelação dos resíduos, como também contra a hipótese de normalidade. Desta forma, conclui-se que o modelo descreve adequadamente os dados, validando o seu uso para traçar previsões.

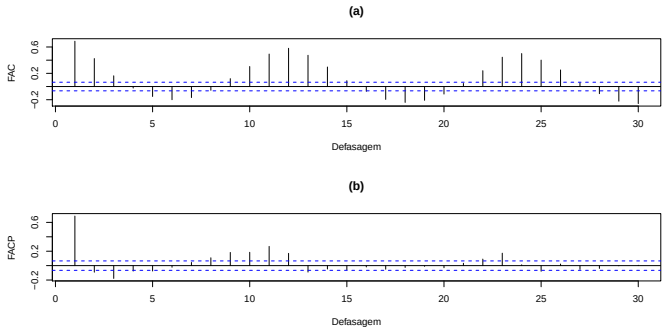

Figura 7: Funções amostrais de autocorrelação (a) e autocorrelação parcial (b).

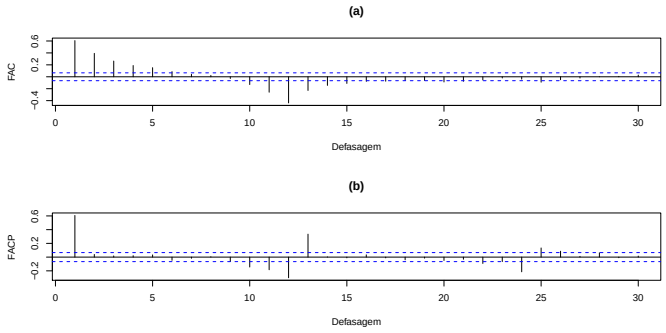

Figura 8: Funções amostrais de autocorrelação (a) e autocorrelação parcial (b) da série diferenciada de ordem 12.

As estimativas dos parâmetros do modelo selecionado $SARIMA(1,1,1)(1,1,2)_{12}$, calculadas por máxima verossimilhança, e o erro padrão de cada parâmetro, são apresentados da Tabela 3.

Para efeitos de comparação de previsão e para mostrar a importância de modelar a sazonalidade, o modelo $ARIMA(1,1,1)$ foi ajustado, na série transformada, desconsiderando a informação de sazonalidade. A série ajustada com os modelos é apresentado na Figura 10.

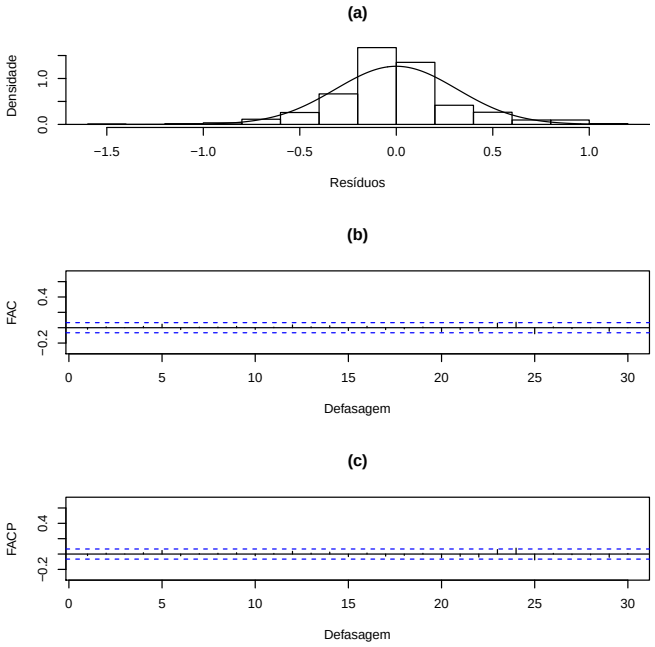

Figura 9: Histograma dos resíduos (a), FAC amostral (b) e FACP dos resíduos (c).

Tabela 3: Estimativas dos parâmetros do modelo ajustado $SARIMA(1,1,1)(1,1,2)_{12}$.

Parâmetro	Estimativa	Erro padrão
ϕ_1	0,5947	0,0391
θ_1	-0,9559	0,0196
Φ_1	-0,8855	0,1117
Θ_2	-0,0891	0,1227
Θ_1	-0,9105	0,1190

3.2.2 Estudo de previsões

Para fins de validação e, principalmente, comparação de desempenho entre modelos, podem ser utilizadas medidas de previsão. Nesse contexto, foram calculadas as previsões mensais para o ano de 2013 para ambos os modelos ajustados. A análise baseou-se nas medidas de qualidade de previsão, EQM e EAM, que mostram a significância estatística das diferenças dos ajustes, e os resultados estão na Tabela 4.

Observa-se que, baseado no EQM, o modelo SARIMA apresentou um aumento de 45,6% na precisão das previsões no horizonte $(h = 0)$ e 54,25% no horizonte $(h = 1)$ em relação ao modelo ARIMA que desconsidera a influência da sazonalidade. Da mesma forma, para a medida EAM o aumento é significativo. Esses resultados empíricos mostram a importância de se considerar o efeito da propriedade sazonalidade na modelagem de séries observadas no tempo.

Portanto, pode-se concluir que o modelo $SARIMA$-

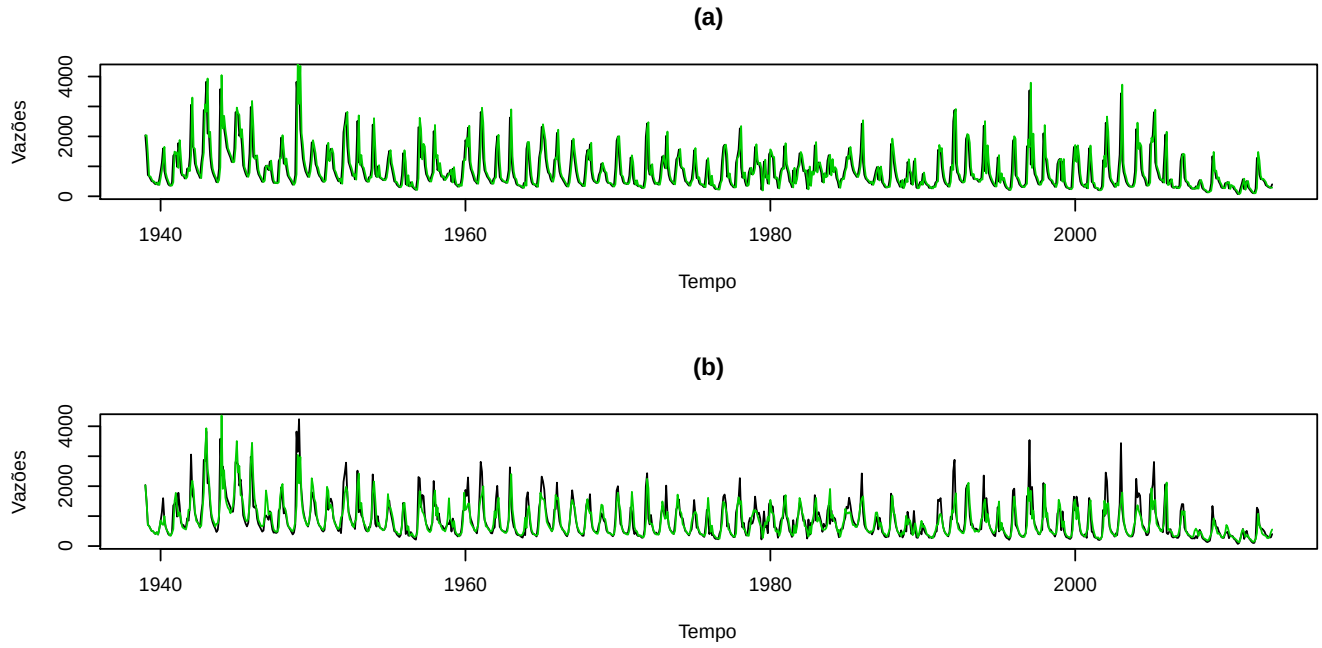

Figura 10: Valores observados da vazão média mensal e valores da vazão média simulados pelos modelos $SARIMA(1,1,1)(2,1,2)_{12}$ (a) e $ARMA(1,1,1)$ (b).

Tabela 5: Previsão de vazões médias mensais (m^3/s) para os meses de 2014 e 2015.

Ano	Meses											
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2014	649,78	499,94	482,05	374,89	289,44	241,78	223,27	197,87	192,53	236,28	381,19	611,97
2015	664,23	513,90	471,95	372,93	282,25	232,22	216,99	197,17	188,67	229,08	370,37	584,03

Tabela 4: Avaliação da precisão das previsões obtidas a partir dos modelos ajustados.

Horizonte	Medida	Modelos	
		SARIMA	ARIMA
h=0	EQM	0,0991	0,1825
	EAM	0,2256	0,3124
h=1	EQM	0,1001	0,1845
	EAM	0,2285	0,3166

$(1,1,1)(2,1,2)_{12}$ é o mais, entre os modelos testados, adequado para prever as vazões médias para os meses de 2014 e 2015. As previsões para os meses de 2014 e 2015 são apresentadas na Tabela 5.

De acordo com as previsões, Tabela 5, a vazão média anual para 2014 será de 365,1 m^3/s e para 2015 360,3 m^3/s. Assim, segundo o modelo SARIMA, deverá haver um aumento nas médias anuais, quando comparado com o ano de 2013, no qual a média anual foi de 341,0 m^3/s. Porém, quando observado os valores de mínimos, que ocorrerão em setembro de 2014 e 2015, 192,5 m^3/s e 188,7 m^3/s, respectivamente, conclui-se que haverá um período de seca mais intensa, quando comparado com o ano de 2013, cujo o mínimo registrado em setembro foi de 210,22 m^3/s.

A maior vazão registrada em 2013 ocorreu no mês de dezembro, 533,87 m^3/s, período de cheias do rio Doce, segundo as previsões será maior em 2014, 649,78 m^3/s, e 2015, de 664,23 m^3/s.

De maneira geral, os resultados obtidos com a modelagem dos dados de vazão média mensal com o modelo SARIMA, tanto no ajuste quanto no estudo de previsão, podem ser considerados bons, uma vez que tais resultados estão condizentes com os encontrados nos trabalhos de Reisen et al. (2008), Bayer et al. (2012), Meller et al. (2012) e Figueredo e Blanco (2014). Nestes trabalhos os autores aplicaram a metodologia de séries temporais para modelagem e previsão de vazões.

Segundo Pruski et al. (2006), a vazão máxima está associada às condições de inundação do local e é uti-

lizada no dimensionamento de obras hidráulicas. A vazão média de longa duração permite quantificar a disponibilidade de água na bacia e calcular o volume de regularização. Já a vazão mínima é utilizada na avaliação da disponibilidade hídrica, na elaboração de projetos de irrigação e de energia elétrica e concessão de uso da água para uma dada finalidade. Neste contexto, e em acordo com as previsões apresentadas, neste trabalho, sugere-se que o sucesso das estratégias de investimentos, passará pela inclusão de medidas adicionais que promovam a cooperação entre vários setores que são importantes utilizadores e consumidores de água, nomeadamente os associados à agricultura, produção de energia, turismo, indústria e habitação.

4 Conclusões

Neste artigo avaliaram-se a qualidade do ajuste e a capacidade de previsão de modelos que representam a dinâmica das vazões médias mensais do rio Doce, Colatina, ES. Comparou-se modelos considerando ou não a presença da sazonalidade no conjunto de observações. Baseando-se em indicadores de desempenho de modelagem (EQM e EAM), verificou-se que o modelo que considera a informação de sazonalidade ($SARIMA(1,1,1)(2,1,2)_{12}$) fornece previsões mais precisas. Após a etapa de validação do modelo foram feitas previsões para os anos de 2014 e 2015. Desta forma, a adoção de modelos confiáveis e precisos pode favorecer a tomada de decisões no planejamento e gestão de recursos hídricos.

Deve-se destacar que os resultados obtidos neste trabalho podem ser usados como um indicativo do comportamento da vazão do rio Doce, principalmente as previsões para o ano de 2015, no município de Colatina, para os meses críticos. Tais previsões são importantes, pois permitem medidas preventivas para a diminuição dos impactos causados, não só ao meio ambiente, mas na economia dessa Cidade.

Como parte de um estudo posterior, sugere-se aprimorar os modelos da classe ARIMA com a inclusão de variáveis meteorológicas adicionais, que permitam melhorar as análises e dessa forma explicar significativamente melhor o comportamento das vazões médias mensais do rio Doce.

Agradecimentos

Os autores agradecem a Faculdade da Região Serrana (FARESE) e ao Centro Universitário do Espírito Santo (UNESC) pelo apoio institucional e aos revisores anônimos deste trabalho, pelas valiosas sugestões.

Referências

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *In Proceedings Intitutional Symposium on Information Theory Amenia*, ..., 267–281.

Akaike, H. (1978). A bayesian analysis of the minimum AIC procedure. *Ann Inst Stat Math*, *30*(1), 9–14.

Barbosa, A. M., Fill, H. D. (2001). Modelo de previsão de vazão baseado no hidrograma unitário. *Revista Brasileira de Recursos Hídricos*, *6*(4), 165–174.

Bayer, D. M., Castro, M. R., Bayer, F. M. (2012). Modelagem e previsão de vazões médias mensais do rio Potiribu utilizando modelos de séries temporais. *Revista Brasileira de Recursos Hídricos*, *17*(2), 229–239.

Bayer, F. M., Souza, A. M. (2010). Wavelets e modelos tradicionais de previsão: um estudo comparativo. *Rev Bras Biom*, *28*(2), 40–61.

Box, G., Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*(2), 211–252.

Box, G., Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*, revised edn. San Francisco: Holden-Dayl.

Box, G., Jenkins, G., Reinsel, G. (2008). *Time Series Analysis: Forecasting and Control*, 4º edn. Prentice Hall.

Brockwel, P., Davis, R. (2002). *Introduction to Time Series and Forecasting*, 2º edn. New York: Springer Verlag.

C.B.H. Doce (2014). *A bacia do rio Doce: Caracterização da bacia*. Comitê da bacia hidrográfica do rio Doce, Governador Valadares, Minas Gerais, URL `http://www.riodoce.cbh.gov.br/bacia_caractericao.asp`.

Coelho, A. L. N. (2007). *Alteração hidrogeomorfológica no médio-baixo rio Doce, ES. Tese (Doutorado em Geografia)*. Programa de Pós-Graduação em Geografia, Universidade Federal Fluminense. Niteroi, RJ.

Collischonn, W., Tucci, C. E. M., Clarke, R. T., Dias, P. L. S., Oliveira, G. S. (2005). Previsão sazonal de vazão na bacia do rio Uruguai 2: previsão climática-hidrológica. *Revista Brasileira de Recursos Hídricos*, *10*(4), 61–72.

Dickey, D., Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a uni root. *Econometricas*, *49*(4), 1057–1073.

Figueredo, N. M., Blanco, C. J. C. (2014). Simulação de vazão e níveis de água médio mensais para o rio Tapajó usando modelos ARIMA. *Revista Brasileira de Recursos Hídricos*, *19*(3), 111–126.

Ghanbarpour, M. R., Abbaspour, K. C., Jalalvand, G., Moghaddam, G. A. (2010). Stochastic modeling of surface stream flow at different time scales: Sangsoorakh karst basin, Iran. *Journal of Cave and Karst Studies*, *72*(1), 1–10.

Koutroumanidis, T., Sylaios, G., Zafeiriou, E., Tsihrintzis, V. A. (2009). Genetic modeling for the optimal forecasting of hydrologic time-series: Application in Nestos River. *Journal of Hydrology*, *368*, 156–164.

Lohani, A. H., Kumar, R., Singh, R. D. (2012). Hydrological time series modeling: A comparison between adaptive neurofuzzy, neural network and autoregressive techniques. *Journal of Hydrology*, *442-443*(6), 23–35.

Lucas, E. W. M., Sousa, F. d. A. S. d., Silva, F. D. d. S., Lucio, P. S. (2009). Modelagem hidrológica determinística e estocástica aplicada à Região Hidrográfica do Xingu – Pará. *Revista Brasileira de Meteorologia*, *24*(3), 208–322.

Meller, A., Bravo, J. M., Collischonn, W. (2012). Análise de dados de vazão na previsão de cheias em tempo-real com o modelo hidrológico MGH-IPH. *Revista Brasileira de Recursos Hídricos*, *17*(3), 209–224.

Mine, M. R., Tucci, C. E. (1999). Previsão em tempo real de vazões afluentes a reservatório de usinas hidrelétricas. *Revista Brasileira de Recursos Hídricos*, *9*(2), 73–99.

Morettin, P., Toloi, C. (2006). *Análise de Séries Temporais*. São Paulo: ABE - Projeto Fisher - Blucher.

Oliveira, V. A. d. (2013). *Regionalização de vazões nas regiões das unidades de planejamento e gestão de recursos hídricos GD1 e GD2, Minas Gerais. 99 f. Dissertação (Mestrado em Recursos Hídricos e Sistemas Agrícolas)*. Programa de Pós-Graduação em Recursos Hídricos e Sistemas Agrícolas, Universidade Federal de Lavras. Lavras, MG.

Pinto, W. d. P. (2013). *O uso da metodologia de dados faltantes em séries temporais com aplicações a dados de concentração (PM10) observados na Região da Grande Vitória. 85 f. Dissertação (Mestrado em Engenharia Ambiental)*. Programa de Pós-Graduação em Engenharia Ambiental, Centro Tecnológico, Universidade Federal do Espírito Santo. Vitória, ES.

Pruski, F., Silva, D., Teixeira, A., Cecílio, R., Silva, J., Griebeler, N. (2006). *Hidros: Dimensionamento de Sistemas Hidroagrícolas*. Viçosa: Editora UFV.

R Development Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`, ISBN 3-900051-07-0.

Reisen, V. A., Molinares, F. A. F., Teixeira, E. C. (2008). Modelagem de séries temporais sazonais na presença de outliers: estudo de caso da vazão máxima mensal do rio Jucu, ES, Brasil. *Revista Brasileira de Recursos Hídricos*, *13*(2), 45–53.

Schwarz, G. (1978). Estimating the dimensional of a model. *Annals of Statistics, Hayward*, *6*(2), 461–464.

Setti, A. A., Lima, J. E. F. W., Chaves, A. G. d. M., Pereira, I. d. C. (2001). *Introdução ao Gerenciamento de Recursos Hídricos*, 2º edn. Brasília: Agência Nacional de Energia Elétrica, Superintendência de Estudos e Informações hidrológicas.

Wei, W. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. New York: Addison Wesley.

Wu, C. L., Chau, K. W. (2010). Data-driven models for monthly streamflow time series prediction. *Engineering Applications of Artificial Intelligence*, *23*, 1350–1367.