

Modelagem de Séries Temporais Sazonais na Presença de Ooutliers Estudo de Caso da Vazão Máxima Mensal do Rio Jucu, ES, Brasil

Valdério Anselmo Reisen

DEST, NuMEs, CCE, PPGEA – UFES

valderio@cce.ufes.br

Fabio A. Fajardo Molinares

NuMes, Departamento de Estatística, Ices – UFMG

ffajardo@est.dout.ufmg.br

Edmilson Costa Teixeira

LabGest, DEA, PPGEA, CT - UFES

edmilson@npd.ufes.br

Recebido: 06/09/06 – revisado: 25/04/07 – aceito: 09/04/08

RESUMO

Na maioria dos estudos de séries temporais de dados hidrológicos, a informação da existência de observações atípicas (outliers) não é considerada como parte integrante da modelagem dos dados. A consideração dos dados atípicos nos processos de modelagem estocástica tem como objetivo melhorar a eficiência do modelo ajustado e proporcionar uma maior confiabilidade em seus resultados. Por isso, usa-se neste artigo ferramenta teórica na modelagem de séries temporais sazonais com outliers, tendo como estudo de caso a vazão máxima do Rio Jucu, entre os dois principais mananciais responsáveis pelo abastecimento da Grande Vitória, estado do Espírito Santo. Os resultados mostram que o modelo SARIMA (Auto-Regressivo Integrado e de Médias Móveis Sazonal), considerando a informação da presença de outliers, representou melhor a dinâmica da série em estudo, ao prever o ciclo anual de picos máximos de vazão com uma redução de 46% no Erro Quadrático Médio (EQM) de previsão para 1-passo à frente.

Palavras-chave: Séries temporais, sazonalidade, outliers.

INTRODUÇÃO

Uma série temporal é uma seqüência de dados observados em intervalos regulares de tempo. O estudo das séries temporais pode ser motivado pelo interesse em investigar o mecanismo gerador de um conjunto de dados observados ao longo do tempo, para descrever sua dinâmica, com o objetivo de gerar previsões acerca do seu comportamento futuro.

Para tanto, são construídos modelos probabilísticos que pertencem a um domínio temporal previamente estabelecido. Os modelos de séries temporais devem respeitar o princípio da parcimônia, ou seja, devem envolver o menor número possível de parâmetros.

A metodologia mais utilizada para a modelagem de séries temporais é a apresentada por Box e Jenkins no ano 1970 (ver Box, Jenkins e Reinsel (1994)), a qual é baseada na utilização de modelos

que consideram as partes auto-regressivas e de médias móveis. Esses modelos são conhecidos como modelos ARMA (Auto-regressivos e de Médias Móveis) e são amplamente utilizados em diversas áreas do conhecimento, como por exemplo, a hidrologia.

Na análise de dados hidrológicos, além da sazonalidade, é comum encontrar a presença de observações influenciadas por eventos externos. Essas observações são conhecidas na literatura por dados atípicos ou *outliers*. De acordo com a natureza do *outlier*, seus efeitos podem ser expressivos sobre os resultados obtidos, influenciando de forma significativa inferências realizadas sobre os modelos ajustados e, conseqüentemente, podendo levar a erros de análises de previsões, bem como na tomada de decisões em processos de planejamento e gestão de recursos hídricos. Ooms e Franses (2001) apresentam uma revisão sobre modelagem de séries temporais aplicada a dados de vazão.

Fox (1972) introduziu o conceito de *outliers* no contexto de séries temporais, considerando dois tipos: *outlier aditivo* (AO – *additive outlier*), *outlier de inovação* (IO – *innovation outlier*). Como extensão do trabalho de Fox (1972), os artigos de Chang, Tiao e Chen (1988), Chen e Liu (1993a, 1993b) e Tsay (1986) consideraram as seguintes alterações na estrutura da série: mudança de nível permanente (LS – *Level Shift*) e mudança transitória (TC – *Transitional Change*). Os mesmos autores, adotando a formulação de Fox (1972), consideraram os dados atípicos como casos particulares do modelo geral de intervenção de Box e Tiao (1975). Estes quatro tipos de *outliers*, AO, IO, LS e TC, interferem na série de modo diferente, ora causando impacto somente no instante da sua ocorrência, ora afetando toda série subjacente ao instante que ocorre a perturbação. Por exemplo, Chang, Tiao e Chen (1988) e Chen e Liu (1993a) mostraram que a presença de dados atípicos afeta os procedimentos convencionais de análise estatística, além de prejudicar as estimativas dos parâmetros e as possíveis previsões obtidas a partir dos modelos ARMA ajustados.

Nos trabalhos recentes de Fajardo, Reisen e Cribari-Neto (2008) e de Haldrup e Nielsen (2007) são apresentados resultados teóricos e empíricos sobre os efeitos causados pelos dados atípicos para um caso mais geral dos processos ARMA integrados: os processos ARIMA fracionários (ver, por exemplo, Reisen (1994), Baillie e Chung (2002), entre outros).

Os efeitos das observações atípicas são, no entanto, muitas vezes omitidos pela falta de conhecimento dos métodos para detectá-las e acomodá-las ao processo de modelagem da série temporal. Diferentes procedimentos de detecção e correção de dados atípicos têm sido propostos na literatura, por exemplo, Battaglia e Orfei (2005), Burridge e Taylor (2006), Chen e Liu (1993b), Tsay (1988) entre outros.

Analistas e pesquisadores têm adotado diferentes métodos para a detecção de *outliers* em dados de vazão. Entretanto, até hoje não existe um método eficiente para identificação e tratamento dos dados atípicos, tornando-se uma linha de pesquisa de muito interesse entre os pesquisadores da área.

Entre as metodologias propostas podem-se citar as desenvolvidas pelos pesquisadores do US-WRC (United States Water Resources Council) e do US-EPA (US Environmental Protection Agency). Na metodologia proposta no US-WRC, a análise estatística é realizada considerando uma distribuição de probabilidade log-Pearson tipo III, ou Gamma,

para avaliar o viés positivo causado pelos dados atípicos. No entanto, os pesquisadores do US-EPA (US Environmental Protection Agency) desenvolveram uma metodologia baseada em análise multivariada utilizando a distância de Mahalanobis. Diferentes métodos utilizando análise de séries temporais no domínio da frequência, para realizar a identificação e tratamento de *outliers*, também têm sido sugeridos na literatura (ver Ng, Panu e Lennox (2007)). Outros métodos alternativos são apresentados em Kottegodda (1984).

Neste artigo, analisa-se a série mensal de vazão máxima do rio Jacu, ES, Brasil, um dos maiores responsáveis pelo abastecimento de água na Grande Vitória, estado do Espírito Santo. No estudo, realiza-se a modelagem da série de observações, através do modelo SARIMA (Auto-regressivo Integrado e de Médias Móveis Sazonal), e com o intuito de validar o modelo ajustado, a análise de previsões é considerada para verificação.

A série em estudo possui observações atípicas e faltantes, sendo características interessantes de pesquisa nas áreas de séries temporais e hidrologia. Nesse contexto, o artigo apresenta um estudo invocando ferramentas sofisticadas de séries temporais, através da metodologia de Box e Jenkins aplicada ao conjunto de dados em estudo.

Na análise da série utilizou-se a metodologia desenvolvida por Gómez e Maravall (1994), implementada no pacote estatístico TRAMO (Time Series Regression with Arima Noise, Missing Observations and Outliers), para identificar os possíveis *outliers* e estimar os valores faltantes. O software e algumas das suas aplicações com dados reais se encontram disponíveis no site do Banco da Espanha (<http://www.bde.es>).

A metodologia utilizada no TRAMO para detecção e tratamento de *outliers* é uma modificação melhorada da proposta de Chen e Liu (1993a), que funciona de forma satisfatória e requer um procedimento mais simples de cálculo. De forma semelhante, a análise de dados faltantes é baseada na teoria de filtros de Kalman (para detalhes ver Gómez e Maravall (1994)).

Este artigo está organizado da seguinte forma: na seção 2 o modelo sazonal ARIMA é sumarizado. A seção 3 trata de metodologias encontradas para a detecção de outliers; na seção 4 é apresentada uma breve descrição da metodologia empregada; na seção 5 são exibidos os resultados; e na seção 6 são apresentadas conclusões.

MODELOS DE SÉRIES TEMPORAIS

Peiris e Singh (1996) e Hassler (1994) propuseram um modelo probabilístico que avalia a sazonalidade do conjunto de observações dentro da equação do modelo, considerando os parâmetros de diferenciação como valores reais, i.e., sendo $d, D \in R$.

Seja $\{Z_t\}$, $t=1,2,\dots,n$, o processo que satisfaz a equação em diferenças dada por

$$\Phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D(Z_t - \mu) = \Theta(B)\Theta(B^s)a_t, \quad (1)$$

onde B é o operador de defasagem definido por

$$B^m Z_t = Z_{t-m}, \quad \Phi(x) = 1 - \sum_{i=1}^p \phi_i x^i \text{ e } \Theta(x) = 1 + \sum_{i=1}^q \theta_i x^i,$$

$\{a_t\}$ é chamado de processo de ruído branco com média 0 e variância constante e μ representa a média do processo $\{Z_t\}$. O filtro de diferenciação

$(1-B)^d$ é definido pela seguinte expansão binomial

$$(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j,$$

onde $\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$, $j=0,1,\dots$, e $\Gamma(\cdot)$ é a

função Gamma. Os polinômios $\Phi(x)$ e $\Theta(x)$ devem ter todas as raízes fora do círculo unitário para cumprir as condições de estacionariedade e invertibilidade, necessárias para o cálculo das estimativas dos parâmetros desconhecidos (para detalhes ver Brockwell e Davis, 2006).

O processo $\{Z_t\}$ definido em (1) é chamado de processo Auto-Regressivo Integrado e de Médias Móveis Fracionário Sazonal, SARFIMA (ver Reisen *et al.* (2006, a,b)).

Os parâmetros d e D representam o número de diferenças simples e sazonais, respectivamente, aplicadas sobre o processo $\{Z_t\}$, com o período sazonal s . Estudos recentes envolvendo um caso mais geral do modelo (1) com diferenças fracionárias sazonais múltiplas são apresentados nos trabalhos de Reisen *et al.* (2006, a,b). Os modelos SARFIMA são apresentados como uma proposta para a modelagem de séries caracterizadas pelo fenômeno de Hurst (ou memória longa), propriedade comumente encontrada nas séries hidrológicas e geofísicas, e inicialmente estudada pelo hidrólogo H. E. Hurst na década de 50.

Neste artigo considera-se um caso particular desses modelos, onde os parâmetros de diferenciação admitem valores inteiros, i.e., os modelos SARIMA.

Em estudos ambientais, os modelos SARIMA são amplamente utilizados. Por exemplo, Tong (2002) analisa a variabilidade do clima em Townsville, Austrália, levando em consideração a transmissão de certo vírus. Nobre (2001) realiza uma comparação entre os modelos SARIMA e linear dinâmico em dados epidemiológicos, através do desempenho em previsões futuras; entre outros. Portanto, a utilização dos modelos SARIMA em estudos ambientais tem apresentado resultados consistentes na modelagem e previsão de séries temporais, mostrando assim ser uma ferramenta útil e motivando o desenvolvimento de novos trabalhos na área.

DETECÇÃO E MODELAGEM DE OUTLIERS EM SÉRIES TEMPORAIS

Quando *outliers* estão presentes na série observada, o efeito desses combinados ao processo sem contaminação pode ser representado pelo seguinte:

$$Y_t = \sum_{j=1}^m \xi_j(B) \omega_j I_t^{(\tau_j)} + Z_t, \quad (2)$$

onde $\{Y_t\}$ é o processo contaminado, $\{Z_t\}$ é o processo dado em (1), ω_j representa a magnitude do i -ésimo *outlier* no tempo τ_j , $I_t^{(\tau_j)}$ é uma variável indicadora que assume o valor 1 se $t = \tau_j$ e zero em caso contrário, $\xi_j(B)$ determina a dinâmica do *outlier* no tempo $t = \tau_j$. O tipo de *outlier* é definido de acordo com as seguintes equações:

$$AO : \xi_j(B) = 1, \quad (3)$$

$$LS : \xi_j(B) = 1/(1-B), \quad (4)$$

$$TC : \xi_j(B) = 1/(1-\delta B), \quad 0 < \delta < 1, \quad (5)$$

$$IO: \xi_j(B) = \theta(B) / \phi(B), \quad (6)$$

Esses tipos de *outliers* afetam a série de diferentes maneiras. Os efeitos gerados pelos tipos AO, LS e TC não dependem do modelo SARIMA ajustado, o que não ocorre com o tipo IO, já que esse utiliza a representação em médias móveis do modelo SARIMA ajustado para gerar o efeito sobre a série de dados. AO e TC geram um efeito transitório sobre a série, enquanto que LS tem efeito permanente. Todavia, os efeitos gerados por AO, TC e LS são limitados, diferentemente do que acontece com IO. Chen e Liu (1993a) sugerem uma aproximação para detecção e correção de outliers que conduz a procedimentos contáveis e eficientes. Esses procedimentos foram implementados no software estatístico TRAMO desenvolvido por Gómez e Maravall (1994).



Figura 1 - Localização da Bacia do Rio Jacu e área de drenagem dos postos fluviométricos.

METODOLOGIA

A bacia hidrográfica do rio Jacu tem uma área de 2.032 km², com cotas altimétricas variando entre 0 m e 1.800 m, e perímetro de 340 km. Na bacia do rio Jacu existem dois postos fluviométricos, denominados de córrego do Galo e Fazenda Jucuruaba (Figura 1). Para o desenvolvimento do presente trabalho, adotou-se o ponto fluviométrico Fazenda Jucuruaba, por essa compreender a maior área de drenagem da bacia (1687 Km²).

Os dados de vazão máxima mensal referente aos anos de 1972 até 2001 foram obtidos através do Sistema de Informações Hidrológicas (HIDROWEB), disponíveis no <http://hidroweb.ana.gov.br/>.

Devido a problemas técnicos de coleta da informação, um número significativo das observações dos anos 1990, 1991 e 1993 é faltante. Ainda, a medição do mês de dezembro do ano 1992 não foi realizada. Assim, com um total de 17 observações faltantes o cálculo dessas observações foi baseado na metodologia de Gómez e Maravall (1994) utilizando filtros de Kalman para a avaliação da máxima verossimilhança no cálculo das estimativas desses valores.

Na Figura 2 apresentam-se as séries original e com dados interpolados. Utiliza-se a série interpolada para o estudo proposto neste artigo.

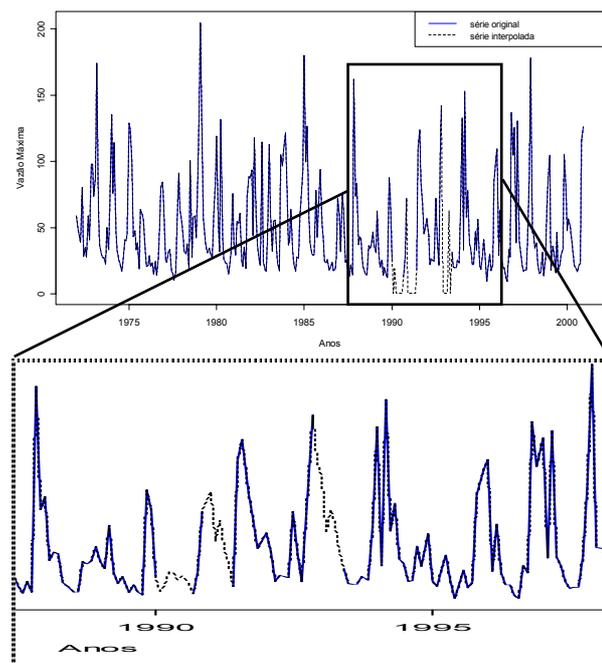


Figura 2 - Série original (contínua) e interpolada (pontilhada) da vazão máxima, período 01/1972-12/2001.

Com o objetivo de realizar a avaliação das previsões, o conjunto de dados foi dividido em dois subconjuntos: O primeiro formado pelos dados de 1972 até 2000, utilizados na modelagem; e o segundo, formado pelos dados de 2001, utilizados para o cálculo das estatísticas de avaliação das

previsões um passo à frente, i.e., cada valor de vazão estimado é inserido automaticamente na série, para estimar o novo valor de vazão (maiores detalhes da teoria de previsão em modelos ARIMA podem ser obtidos em Wei (2006)).

Como mencionado anteriormente, o apoio computacional utilizado foi o pacote estatístico TRAMO. A utilização desse pacote alicerça-se ao fato do mesmo possuir uma grande flexibilidade no que tange a estimação, previsão e interpolação de modelos de regressão. Muitas instituições internacionais, bancos, universidades, centros de pesquisa e companhias privadas fazem uso desse ferramental estatístico.

O procedimento de detecção de *outliers* de TRAMO é baseado na seguinte estatística

$$\lambda_j = \frac{\hat{\omega}_j}{\sqrt{\text{var}(\hat{\omega}_j)}}, \quad (7)$$

onde $\hat{\omega}_j$ é o estimador da magnitude do *outlier* no tempo $t = \tau_j$, ou seja, a j -ésima componente do vetor $\hat{\omega} = (\hat{X}'\hat{X})^{-1}\hat{X}'r^*$, r^* são chamados de resíduos estimados e \hat{X} é o estimador da matriz X que depende do tipo de *outliers* (para detalhes ver Gómez e Maravall (1994)). A estatística do teste é dada por

$$\eta(\tau) = \max_j \{|\lambda_j|\}. \quad (8)$$

Se $\eta(\tau) > C$, é possível que a série observada esteja sobre influência de um tipo qualquer de *outliers* (AO, LS, TS e IO) no instante $t = \tau$. A constante C é o nível crítico pré-determinado, que depende do número de observações. Para a escolha do valor C , Chang, Tiao e Chen (1988) sugerem a seguinte regra:

- I. Se o número de observações (N) for menor ou igual a 50, $C = 3$.
- II. Se $0 \leq N \leq 450$, $C = 3 + 0,0025 \cdot (N - 50)$, caso contrário $C = 4$.

RESULTADOS

A Figura 2 sugere que a série de vazão máxima do rio Jucu apresenta comportamento

sazonal ao redor de um valor médio estimado de $49,58\text{m}^3/\text{h}$.

Observações significativamente acima da média evidenciam a presença de valores atípicos na série, o que indica aumento na variância dos dados, e assim alterações significativas na estrutura de correlação (ver Fajardo, Reisen e Cribari-Neto (2008)). Para estabilizar o aumento na variância das observações realizam-se testes estatísticos para examinar a aplicação de uma transformação no conjunto de dados. Neste caso, considerou-se a família de transformações sugeridas por Box e Cox (1964) baseadas na estimativa do parâmetro β que define o tipo de transformação. Frequentemente, a transformação não só estabiliza a variância, mas também melhora na aproximação da distribuição dos dados para a distribuição normal.

O processo transformado é dado por

$$X_t = \begin{cases} Y_t^\beta - 1, & \text{se } \beta \neq 0, \\ \ln Y_t, & \text{se } \beta = 0. \end{cases} \quad (9)$$

Desse modo, os dados indicaram que um valor adequado para β é zero, i.e., o logaritmo natural foi a transformação adequada para estabilizar a variância.

Ajuste dos modelos

Para o procedimento de modelagem, considerou-se a série transformada com logaritmo natural. A identificação do modelo probabilístico para o cálculo de previsões baseou-se na metodologia sugerida por Box e Jenkins (para detalhes ver Box, Jenkins e Reinsel (1994)).

A propriedade de estacionariedade confirmou-se através das funções amostrais de autocorrelação (FAC) e autocorrelação parcial (FACP), apresentadas na Figura 3.

As FAC e FACP sugerem que os dados da vazão máxima do rio Jucu apresentam comportamento sazonal, com período sazonal de 12 meses, i.e., o valor de s considerado no modelo (1) é 12.

Com a série diferenciada sazonalmente, realizam-se os procedimentos de identificação e estimação dos parâmetros do modelo. Neste caso, a identificação das ordens auto-regressivas e de médias móveis baseou-se nos critérios de informação de Akaike (AIC) e Bayesiano (BIC).

Observou-se que os menores valores obtidos para os critérios de informação foram evidenciados para um modelo SARIMA(1,0,0)(0,1,1)₁₂. Por outro lado, TRAMO detectou dois tipos de dados atípicos: *outliers* do tipo aditivo (AO) e de mudança transitória (TC). O nível crítico C, utilizado no procedimento de identificação dos *outliers*, foi calculado a partir do número de observações, obtendo um valor de 3,75.

As estimativas dos parâmetros do modelo, calculadas por máxima verossimilhança exata, e a correspondente estatística t (teste t-student) para significância de cada parâmetro, são apresentadas na Tabela 1.

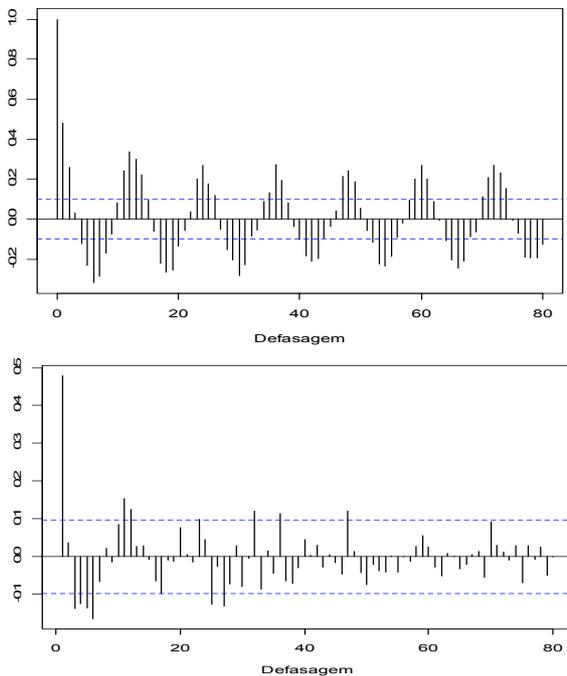


Figura 3 - Funções amostrais de autocorrelação (a) e autocorrelação parcial (b).

Tabela 1 - Estimativas dos parâmetros do modelo ajustado.

Parâmetro	Estimativa	Estatística t
ω_1 (AO)	1,7375	4,13
ω_2 (TC)	1,7386	4,13
ϕ_1	-0,4093	-8,17
θ_{12}	-0,9111	-40,51

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 4

mostra os valores da FAC amostral e probabilidade normal 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.

O modelo ajustado para os dados transformados, com a informação de *outliers*, é da seguinte forma:

$$Y_t = \omega_1 I_t^{(128)} + \frac{\omega_2}{1-0.7B} I_t^{(235)} + Z_t^* \tag{10}$$

onde o processo $\{Z_t^*\} = \{\ln Z_t\}$ é dado por

$$(1 - \phi_1 B)(Z_t^* - 3,69) = (1 + \theta_{12} B^{12}) a_t .$$

Para efeitos de comparação de previsão, o modelo SARIMA(1,0,0)(0,1,1)₁₂ foi ajustado, na série transformada, desconsiderando a informação de *outliers*.

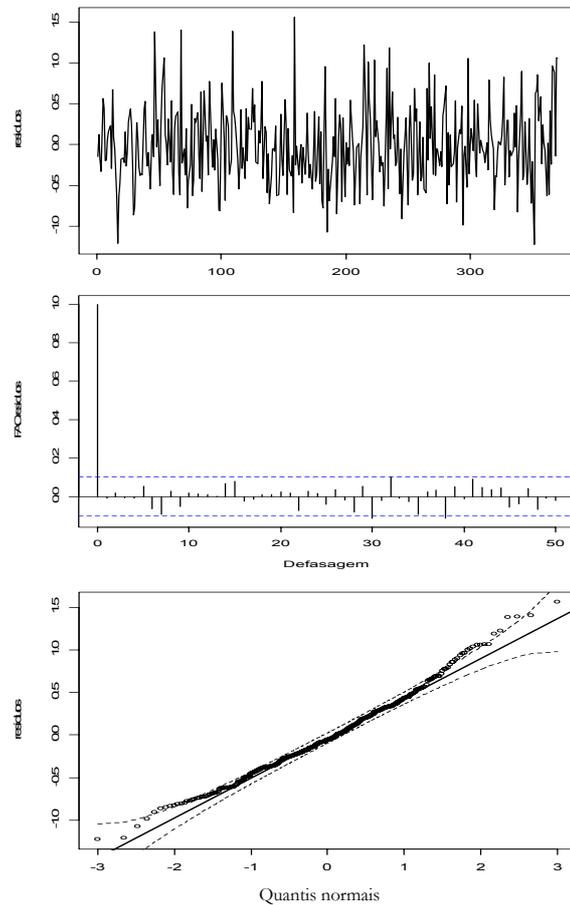


Figura 4 - Resíduos do modelo com dados transformados (a), FAC amostral dos resíduos (b) e gráfico de quantis (c).

O modelo ajustado não apresentou evidências claras de ajuste inadequado, i.e., os resíduos não mostraram evidências que violam os pressupostos estatísticos de adequação do modelo.

Estudo de previsões

Medidas de previsão podem ser também utilizadas para fins de validação e, principalmente, comparação de desempenho entre modelos. Nesse contexto, foram calculadas as previsões mensais para o ano 2001 para ambos modelos ajustados, ARIMA sazonal com e sem outlier. A análise baseou-se no cálculo de estatísticas para medir o grau de precisão das previsões e estabelecer a comparação com os valores reais da série.

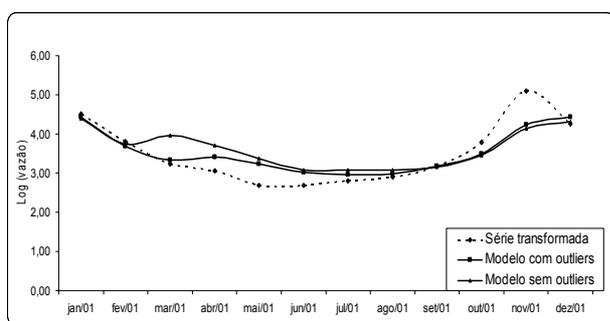


Figura 5 - Comparação das previsões e valores originais para o ano de 2001 do logaritmo da vazão máxima da Fazenda Jucuruaba.

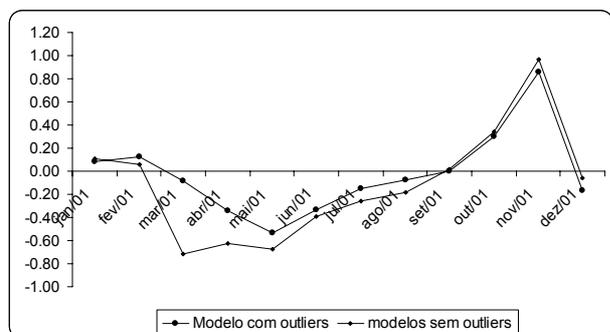


Figura 6 - Comparativos dos erros de previsão para os modelos com outliers e sem outliers.

Na Figura 5 apresenta-se comparativo dos valores originais e as previsões a 1-passo à frente utilizando o modelo dado em (10) e modelo ARIMA sazonal sem considerar a informação de outliers. A Figura 6 compara os erros de previsão obtidos a partir dos modelos analisados, evidenciando a vantagem de considerar a maior quantidade de

informação no modelo ajustado para o cálculo das previsões.

Uma análise mais precisa obtém-se através dos cálculos de medidas de qualidade de previsão, o Erro Percentual Absoluto Médio (EPAM) e o Erro Quadrático Médio (EQM), que mostram a significância das diferenças dos ajustes, e os resultados estão na Tabela 3.

Tabela 3 - Avaliação da precisão das previsões obtidas a partir dos modelos ajustados.

Medida	Modelo sem outliers	Modelo com outliers
EQM	0,22	0,12
EPAM	0,11	0,07

Note-se que, baseado no EQM, o modelo SARIMA com outliers apresentou uma redução de 46% na precisão das previsões em relação ao modelo que desconsidera a influencia dos outliers na série. Da mesma forma, para a medida EPAM a diminuição resultante é próxima de 33%. Esses resultados empíricos mostram a importância de se considerar o efeito de outliers na modelagem de séries observadas no tempo.

CONCLUSÕES

Na análise da série de vazão máxima do rio Jucu utilizou-se a metodologia desenvolvida por Gómez e Maravall (1994), implementada no pacote estatístico TRAMO (Time Series Regression with Arima Noise, Missing Observations and Outliers), para identificar os possíveis outliers e estimar os dados faltantes.

No presente estudo comparou-se modelos considerando ou não a presença de outliers no conjunto de observações. Baseando-se em indicadores de desempenho de modelagem (EQM e EPAM), verificou-se que o modelo que considera a informação da presença de outliers fornece previsões estatisticamente mais precisas. Certamente, a adoção de modelos confiáveis e precisos favorecerá processos de tomada de decisões no planejamento e gestão de recursos hídricos.

A investigação empírica foi feita para o logaritmo da série de vazão. A escolha da função de transformação foi baseada na metodologia Box-Cox. Estudos similares apresentados no artigo também

foram conduzidos para a série original (sem transformação). A transformação para estabilizar a variância forneceu um modelo mais parcimonioso, i.e., com menor número de parâmetros. Os resultados referentes ao modelo não transformado encontram-se disponíveis com os autores.

AGRADECIMENTOS

Os autores Valdério Anselmo Reisen e Edmilson Costa Teixeira agradecem o apoio financeiro concedido pelo CNPq em relação às bolsas de produtividade em pesquisa e aos avaliadores anônimos deste trabalho, pelas valiosas sugestões que provocaram uma mudança significativa no manuscrito inicial. O primeiro autor agradece aos seus ex-alunos André Quintão de Almeida e Daniela Bertolini Depizzol, por participarem da etapa inicial deste trabalho.

REFERÊNCIAS

- BAILLIE, R. T. AND CHUNG, S. (2002), MODELING AND FORECASTING FROM TREND-STATIONARY LONG MEMORY MODELS WITH APPLICATIONS TO CLIMATOLOGY, INTERNATIONAL JOURNAL OF FORECASTING, 18, 215-226.
- BATTAGLIA F. AND ORFEI L. (2005) OUTLIERS DETECTION AND ESTIMATION IN NONLINEAR TIME SERIES, 26, 107-121.
- BOX, G. AND COX D. (1964), AN ANALYSIS OF TRANSFORMATIONS, JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, 26, 211-252.
- BOX, G., JENKINS, C. AND REISEL, G.. (1994), TIME SERIES ANALYSIS: FORECASTING AND CONTROL, PRENTICE HALL
- BOX, G. AND TIAO, G. (1975) INTERVENTION ANALYSIS WITH APPLICATIONS TO ENVIRONMENTAL AND ECONOMIC PROBLEMS, JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 70, 71-79.
- BROCKWELL, R. AND DAVIS A. (2006) TIME SERIES: THEORY AND METHODS, SPRINGER-VERLAG.
- BURRIDGE P. AND TAYLOR R. (2006) ADDITIVE OUTLIER DETECTION VIA EXTREME-VALUE THEORY, JOURNAL OF TIME SERIES ANALYSIS, 27, 687-701.
- CHANG, I., TIAO, G. C. AND CHEN, C. (1988), "ESTIMATION OF TIME SERIES PARAMETERS IN THE PRESENCE OF OUTLIERS," TECHNOMETRICS 30, 193-204.
- CHEN, C. AND LIU, L. (1993a) JOINT ESTIMATION OF MODEL PARAMETERS AND OUTLIER EFFECTS IN TIME SERIES, JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 88, 284-297.
- CHEN, C. AND LIU, L. (1993b) FORECASTING TIME SERIES WITH OUTLIER, JOURNAL OF FORECASTING, 12, 13-35.
- FAJARDO F., REISEN V. AND CRIBARI-NETO F. (2008) ROBUST ESTIMATION IN LONG-MEMORY PROCESSES UNDER ADDITIVE OUTLIERS, JOURNAL OF TIME SERIES ANALYSIS, IN PRESS.
- FOX, A. J., (1972) OUTLIERS IN TIME SERIES, JOURNAL OF THE ROYAL STATISTICAL SOCIETY, B, 34, 350-363.
- GOMEZ V. AND MARAVALL A. (1995) ESTIMATION, PREDICTION AND INTERPOLATION FOR NONSTATIONARY SERIES WITH THE KALMAN FILTER, 89, 611-624.
- HALDRUP, N. AND NIELSEN, M. (2007). ESTIMATION OF FRACTIONAL INTEGRATION IN THE PRESENCE OF DATA NOISE, COMPUTATIONAL STATISTICS AND DATA ANALYSIS, 51, 3100-3114.
- HASSLER, U., 1994. (MIS)SPECIFICATION OF LONG MEMORY IN SEASONAL TIME SERIES. JOURNAL OF TIME SERIES ANALYSIS, 15, 19-30.
- KOTTEGODA (1984) INVESTIGATION OF OUTLIERS IN ANNUAL MAXIMUM FLOW SERIES, JOURNAL OF HYDROLOGY, 72, 105-137.
- NG W., PANU U. AND LENNOX W. (2007) CHAOS BASED ANALYTICAL TECHNIQUES FOR DAILY EXTREME HYDROLOGICAL OBSERVATIONS, 342, 17-41.
- NOBRE, F. F., MONTEIRO, A. B. S, TELLES, P. R., WILLIAMSON, G. D, (2001) DYNAMIC LINEAR MODEL AND SARIMA: A COMPARISON OF THEIR FORECASTING PERFORMANCE IN EPIDEMIOLOGY, STATISTICS IN MEDICINE, 20, 3051-3069.
- OOMS, M., FRANCES, P.H. (2001) A SEASONAL PERIODIC LONG MEMORY MODEL FOR MONTHLY RIVER FLOWS, ENVIRONMENTAL MODELLING AND SOFTWARE, 16, 559-569.
- PEIRIS, M.S., SINGH, N., 1996. PREDICTORS FOR SEASONAL AND NONSEASONAL FRACTIONALLY INTEGRATED ARIMA MODELS, BIOMETRICS, 38, 741-752.
- REISEN V. A. (1994). ESTIMATION OF THE FRACTIONAL DIFFERENCE PARAMETER IN THE ARIMA(P,D,Q) MODEL USING THE SMOOTHED PERIODOGRAM, JOURNAL OF TIME SERIES ANALYSIS, 15, 335-350.
- REISEN, V. A., RODRIGUES, A. L., PALMA, W. (2006a), ESTIMATION OF SEASONAL FRACTIONALLY

- INTEGRATED PROCESSES, COMPUTATIONAL STATISTICS & DATA ANALYSIS, 50, 568 - 582.
- REISEN, V. A., RODRIGUES, A. L., PALMA, W. (2006b), ESTIMATION SEASONAL LONG-MEMORY PROCESSES: MONTE CARLO STUDY, JOURNAL OF A STATISTICAL COMPUTATION AND SIMULATION, 50, 305 - 316.
- THORNTHWAITE, C.W., MATHER, J.R. (1955) THE WATER BALANCE, NEW JERSEY : DREXEL INSTITUTE OF TECHNOLOGY, (PUBLICATIONS IN CLIMATOLOGY).
- TSAY, R. S. (1986), TIME SERIES MODEL SPECIFICATION IN THE PRESENCE OF OUTLIERS, JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 81, 132-141.
- TSAY, R. S. (1988) OUTLIERS, LEVEL SHIFTS, AND VARIANCE CHANGES IN TIME SERIES, JOURNAL OF FORECASTING, 7, 1-20.
- TONG, S. L., HU, W. B. (2002), CLIMATE VARIABILITY AND ROSS RIVER VIRUS TRANSMISSION IN TOWNSVILLE, AUSTRALIA: A SARIMA MODEL, AMERICAN JOURNAL OF EPIDEMIOLOGY, 155.
- WEI, W. (2006) TIME SERIES ANALYSIS: UNIVARIATE AND MULTIVARIATE METHODS. ADDISON-WESLEY.

in a variety of water resources problems such as floods and droughts.

Keywords: *outliers, time series analysis, seasonal ARIMA model.*

Modelling Seasonal Time Series in the Presence of Outliers. The Case Study of the Maximum Monthly Discharge of Jucu River, ES, Brazil

ABSTRACT

In most hydrological time series data studies atypical observations (outliers) are not usually considered in the data modelling process. As a result, the performance of stochastic modelling, in terms of liability and accuracy, has not been very good. This paper presents a comparative study of stochastic modelling of the monthly maximum flow rate of River Jucu (one of the two main water suppliers to Great Vitória, State of Espírito Santo, Southeastern Region of Brazil), by considering or not outliers as part of the flow rate time series. Based on the modelling performance indicators MSE (mean square error) and MAPE (mean absolute percentile error), the version of the SARIMA model taking outliers into account presented a much better performance in terms of prediction reliability and accuracy than the SARIMA model in its usual form (without considering outliers). Thus, the case study presented in this paper calls attention to the importance of taking into account the presence of outliers when modelling time series