
ANÁLISE DO ERRO DE PREVISÃO DE VAZÕES MENSIS COM DIFERENTES HORIZONTES DE PREVISÃO

Marinho G. Andrade*

marinho@icmc.usp.br

Secundino Soares†

dino@cose.fee.unicamp.br

Ricardo Luis dos Reis*

rreis18@gmail.com

Donato da Silva Filho‡

donato.filho@edpbr.com.br

*Departamento de Matemática Aplicada e Estatística, ICMC - USP, São Carlos, SP

†Departamento de Engenharia de Sistemas- FEEC - UNICAMP, Campinas, SP

‡Energias do Brasil - EDP - São Paulo, SP

ABSTRACT

Analysis of Forecast Error of Monthly Streamflow with Different Forecast Horizons

This paper addresses the problem of forecasting for monthly mean streamflow series, in which we call the forecast horizon (h), the interval of time between the last observation used in fitting the model prediction and the future value being predicted. The analysis of the forecast error is made on the basis of the forecast horizon. These series have a periodic behavior on average, variance and autocorrelation function. Therefore, we consider the widely used approach to modeling these series that initially consists of removing the periodicity in mean and variance of the streamflow series and then calculating a standardized series for which stochastic models are adjusted. In this study we consider the series to the standard periodic autoregressive models PAR (p_m). Orders p_m of the adjusted models for each month are determined from the analysis of periodic partial autocorrelation function (PePACF), using the Bayesian Information Criterion (BIC) applied to PAR models, proposed in (MecLeod, 1994) and analysis of PePACF proposed in (Stedinger, 2001). The forecast errors are calculated on the basis of parameters adjusted and evaluated for forecasting horizons h , ranging from 1 to 12 months on the

original scale of the flow series. These errors are compared with estimates of the variances of the flows for the month being predicted. As a result we have an evaluation of the predictive performance in months of the adjusted models for each month.

KEYWORDS: PAR model, streamflow forecasting, forecast error, hydroelectric systems

RESUMO

Este trabalho aborda o problema de previsão para séries de vazões médias mensais, no qual denomina-se de horizonte de previsão (h), o intervalo de tempo que separa a última observação usada no ajuste do modelo de previsão e o valor futuro a ser previsto. A análise do erro de previsão é feita em função deste horizonte de previsão. Estas séries possuem um comportamento periódico na média, na variância e na função de autocorrelação. Portanto, considera-se a abordagem amplamente usada para a modelagem destas séries que consiste inicialmente em remover a periodicidade na média e na variância das séries de vazões e em seguida calcular uma série padronizada para a qual são ajustados modelos estocásticos. Neste estudo considera-se para a série padronizada os modelos autorregressivos periódicos PAR(p_m). As ordens p_m dos modelos ajustados para cada mês são determinadas usando os seguintes critérios: a análise clássica da função de autocorrelação parcial periódica (FACPPe); usando-se o *Bayesian*

Artigo submetido em 03/05/2011 (Id.: 1332)

Revisado em 11/06/2011, 05/07/2011, 19/08/2011

Aceito sob recomendação do Editor Associado Prof. Jês de Jesus Fiais Cerqueira

Information Criterion (BIC) proposto em (MecLeod, 1994); e com a análise da FACPPE proposta em (Stedinger, 2001). Os erros de previsão são calculados, na escala original da série de vazão, em função dos parâmetros dos modelos ajustados e avaliados para horizontes de previsão h variando de 1 a 12 meses. Estes erros são comparados com as estimativas das variâncias das vazões para o mês que está sendo previsto. Como resultado tem-se uma avaliação da capacidade de previsão, em meses, dos modelos ajustados para cada mês.

PALAVRAS-CHAVE: Modelo PAR, previsão de vazão, erro de previsão, sistemas hidrelétricos.

1 INTRODUÇÃO

Nos modelos de planejamento da operação de sistemas hidrotérmicos uma das variáveis importante é a vazão afluente aos aproveitamentos hidrelétricos. As vazões nos rios dependem, entre outros fatores, das condições climáticas, o que leva a atribuir às vazões um modelo estocástico. Alguns modelos de planejamento da operação de sistemas hidrotérmicos consideram explicitamente o modelo estocástico para as vazões (ver (Stedinger et al., 1984), (Maceira et al., 2002)). Isto torna o planejamento da operação um problema de otimização de grande porte, não linear e, conseqüentemente, estocástico. Outras abordagens consideram de forma implícita o modelo estocástico para as vazões como é feito em (Martinez and Soares, 2002). Nestas abordagens, os modelos para as vazões são utilizados para obter previsões. Neste contexto, a previsão de vazão é uma importante etapa do planejamento da operação de sistemas hidrotérmicos. No entanto, esta previsão está sempre sujeita a erros. Erros de previsões podem levar a tomada de decisões incorretas, causando grandes prejuízos ao sistema. Por isso, torna-se importante ter uma boa avaliação do erro de previsão associado ao modelo usado para previsão.

O uso de modelos estocásticos para previsão de valores de uma série temporal tem como um dos objetivos fazer a previsão com o menor erro possível. As séries de vazões mensais possuem um comportamento periódico na média, na variância e na função de autocorrelação. Por isso, considera-se a abordagem amplamente usada na modelagem destas séries (Hippel and McLeod, 1994), que consiste inicialmente em remover a periodicidade na média e na variância e em seguida calcular uma série padronizada para a qual são ajustados modelos do tipo autorregressivo periódico $PAR(p_m)$. Este trabalho faz uma análise do erro de previsão para séries de vazões médias mensais em função dos parâmetros do modelo adotado para a previsão.

Neste trabalho adota-se o erro médio quadrático da previsão, MSE (*Mean Squared Error*) como critério para comparar as previsões feitas com diferentes modelos. A vantagem deste

critério é que ele pode ser comparado com a variância incondicional da série. A variância incondicional de uma série reflete a máxima incerteza ou variabilidade que se pode esperar da série, quando ignora-se toda e qualquer informação que pode ser extraída dos valores passado e presente da série para prever um valor futuro. Portanto a variância incondicional representa o maior erro que se pode cometer ao fazer uma previsão. Sendo assim, esta variância serve como referência para avaliar o MSE calculado com qualquer modelo que faz usos de algum tipo de ferramenta estatística para melhorar a previsão.

Denomina-se de horizonte de previsão (h), o intervalo de tempo que separa a última observação usada no ajuste do modelo de previsão e o valor futuro a ser previsto. Um comportamento que se deve esperar intuitivamente do erro de previsão é que este erro deve crescer com o horizonte de previsão, ou seja, quanto mais a frente se quer prever valores futuros, maior é o erro de previsão ao qual se está sujeito. Isto se deve a limitada capacidade que o passado e o presente tem para prever o crescente efeito das variáveis aleatórias imprevisíveis, que podem influenciar os valores futuros da série quando distancia-se do presente em direção ao futuro. Assim, é natural a expectativa de que, a medida que o horizonte de previsão aumenta, o erro de previsão tenda para a variância incondicional da série. Neste trabalho, os erros de previsão são avaliados na escala original das séries de vazões e usa-se como referência a variância incondicional sazonal da série, permitindo avaliar a rapidez com que estes erros de previsão tendem para a variância incondicional da série, quando diferentes modelos são considerados para fazer previsões.

Os modelos $PAR(p_m)$, considerados neste trabalho, são ajustados considerando-se a equação de Yule-Walker (Mondal and Wasimib, 2006; Box et al., 2008). Um ponto crítico no ajuste destes modelos é a escolha da ordem p_m do modelo de cada mês. As ordens p_m dos modelos ajustados para as vazões de cada mês podem ser determinadas fazendo-se uma análise da função de autocorrelação parcial periódica (FACPPE) ou usando-se *Bayesian information criterion* (BIC) proposto por (MecLeod, 1994). Estes critério podem apresentar resultados que não se sustentam com uma análise dos fenômenos naturais. Ou seja, ao ser determinado um valor para a ordem p_m , pode ocorrer que alguns parâmetros do modelo $PAR(p_m)$, para defasagens intermediárias $k < p_m$, sejam não significativos. Portanto, torna-se difícil explicar, com base nos fenômenos naturais, como a vazão de um mês m depende da vazão de um mês com defasagem p_m e não depende das vazões de meses com defasagens $k < p_m$. Para contornar esta dificuldade, foi usada a análise da FACPPE proposto por (Stedinger, 2001), a qual considera para cada mês m , a maior ordem p_m tal que todas as estimativas da FACPPE de defasagem menor que p_m sejam significativas.

Este critério não admite valores intermediários não significativos.

2 O MODELOS PAR(P_M)

Denotando-se por $q_{t(r,m)}$, $t(r,m) = s(r-1) + m$, uma série temporal de vazões mensais em que $m = 1, \dots, s$ ($s = 12$) representa os meses e $r = 1, \dots, n$ é índice dos anos ($m = 1$ refere-se ao mês de janeiro e $m = 12$ refere-se ao mês de dezembro). Define-se a série padronizada por:

$$z_{t(r,m)} = \frac{q_{t(r,m)} - \mu_m}{\sigma_m} \quad (1)$$

em que $\mu_m = E(q_{t(r,m)})$ e $\sigma_m^2 = Var(q_{t(r,m)})$ são as médias e as variâncias mensais, respectivamente. Portanto, a série padronizada $z_{t(r,m)}$ tem $E(z_{t(r,m)}) = 0$ e $Var(z_{t(r,m)}) = 1$. A média μ_m , e variância σ_m^2 de cada mês podem ser estimadas respectivamente por:

$$\hat{\mu}_m = \frac{1}{n} \sum_{r=1}^n q_{t(r,m)}, \quad m = 1, \dots, s \quad (2)$$

$$\hat{\sigma}_m^2 = \frac{1}{n-1} \sum_{r=1}^n (q_{t(r,m)} - \hat{\mu}_m)^2, \quad m = 1, \dots, s \quad (3)$$

Uma característica das séries de vazões mensais é que a função de autocorrelação também é periódica. Assim, modelos autorregressivos periódicos (PAR) são frequentemente adotados para representar esta característica (Vechia, 1985). O modelo PAR(p_m) para a série padronizada $z_{t(r,m)} > 0$, $r = 1, \dots, n$, $m = 1, \dots, s$ é escrito como:

$$z_{t(r,m)} = \sum_{j=1}^{p_m} \phi_{m,j} z_{t(r,m)-j} + a_{t(r,m)} \quad (4)$$

Nos modelos PAR(p_m) o processo $z_{t(r,m)}$ possui estacionariedade periódica, ou seja, as raízes de cada polinômio $\phi_m(B) = 1 - \phi_{m,1}B - \dots - \phi_{m,p_m}B^{p_m} = 0$ estão situadas no plano complexo, fora do círculo de raio unitário. E os ruídos $a_{t(r,m)}$ são não-correlacionados, $E(a_{t(r,m)}a_{t(r,m)+k}) = 0$, para todo $k \neq 0$, com variância $Var(a_{t(r,m)}) = \tau_m^2$. Nestes modelos os parâmetros $\phi_{m,j}$ e a ordem do modelo p_m variam com m (neste estudo assumi-se que $p_m < s$ para todo m). Além disto, tem-se que para qualquer k inteiro $\phi_{m,j} = \phi_{m+k,s,j}$ e como consequência se tem que para qualquer $\tau = t(r,m)$:

$$E(z_\tau) = E(z_{\tau+ks}) \quad (5)$$

$$Var(z_\tau) = Var(z_{\tau+ks}) \quad (6)$$

$$Cov(z_\tau, z_{\tau+\delta}) = Cov(z_{\tau+ks}, z_{\tau+\delta+ks}) \quad (7)$$

3 IDENTIFICAÇÃO E ESTIMAÇÃO DO MODELO PAR(P_M)

A identificação de um modelo PAR(p_m) consiste em determinar as ordens p_m , mais apropriadas dos operadores autorregressivos para cada mês $m = 1, \dots, s$. Neste trabalho adota-se, para a escolha da ordem do modelo de cada mês, três procedimentos. Dois destes são baseados na análise da função de autocorrelação parcial periódica (FACPPe) do modelo PAR(p_m), o terceiro é o procedimento adotado em (MecLeod, 1994), que considera o critério BIC proposto por (Schwarz, 1978).

A função de autocorrelação periódica (FACPe), ρ_k^m entre $z_{t(r,m)}$ e $z_{t(r,m)-k}$, é dada por $\rho_k^m = E(z_{t(r,m)}z_{t(r,m)-k})$. Considerando-se o modelo PAR(p_m), tem-se que:

$$\begin{aligned} \rho_k^m &= \phi_{m,1}E(z_{t(r,m)-1}z_{t(r,m)-k}) + \\ &\phi_{m,2}E(z_{t(r,m)-2}z_{t(r,m)-k}) + \dots + \\ &\phi_{m,p_m}E(z_{t(r,m)-p_m}z_{t(r,m)-k}) + \\ &E(a_{t(r,m)}z_{t(r,m)-k}) \end{aligned} \quad (8)$$

Fixando-se o mês m e variando k de 1 até p_m em (8) encontra-se, para cada mês m , um conjunto de equações lineares, comumente denominadas de equações de Yule-Walker. Para um mês m qualquer, as equações de Yule-Walker são dadas por:

$$\begin{pmatrix} \rho_1^m \\ \rho_2^m \\ \rho_3^m \\ \vdots \\ \rho_{p_m}^m \end{pmatrix} = R \begin{pmatrix} \phi_{m,1} \\ \phi_{m,2} \\ \phi_{m,3} \\ \vdots \\ \phi_{m,p_m} \end{pmatrix} \quad (9)$$

em que R é a matriz definida por:

$$R = \begin{pmatrix} 1 & \rho_1^{m-1} & \rho_2^{m-1} & \dots & \rho_{p_m-1}^{m-1} \\ \rho_1^{m-1} & 1 & \rho_1^{m-2} & \dots & \rho_{p_m-2}^{m-2} \\ \rho_2^{m-1} & \rho_1^{m-2} & 1 & \dots & \rho_{p_m-3}^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p_m-1}^{m-1} & \rho_{p_m-2}^{m-2} & \rho_{p_m-3}^{m-3} & \dots & 1 \end{pmatrix} \quad (10)$$

Denota-se por $\varphi_{k,j}^m$ o j -ésimo parâmetro autorregressivo de um processo PAR de ordem k , em que $\varphi_{k,k}^m$ é o último parâmetro deste processo. As equações de Yule-Walker para cada mês m podem ser reescritas da seguinte forma:

$$\begin{pmatrix} \rho_1^m \\ \rho_2^m \\ \rho_3^m \\ \vdots \\ \rho_k^m \end{pmatrix} = R(k, k) \begin{pmatrix} \varphi_{k,1}^m \\ \varphi_{k,2}^m \\ \varphi_{k,3}^m \\ \vdots \\ \varphi_{k,k}^m \end{pmatrix} \quad (11)$$

em que $R(k, k)$ é a matriz de dimensão $k \times k$ formada pelas primeiras k -ésima linhas e k -ésima colunas da matriz R , dada em (10).

Os valores dos parâmetros $\{\varphi_{kk}, k = 1, 2, \dots\}$ correspondem à função autocorrelação parcial periódica do processo PAR no período m . Esta função representa também a estrutura de dependência do processo estocástico ao longo do tempo. Em um processo periódico autorregressivo de ordem p_m , a FACPPE é o valor do último parâmetro φ_{kk}^m dado em (11). Uma propriedade dos modelos PAR(p_m) é que $\{\varphi_{kk}^m\}$ é diferente de zero, para k menor ou igual a p_m , e zero para k maior que p_m . Portanto, um procedimento de identificação de um modelo PAR(p_m) consiste em determinar as ordens p_m dos modelos autorregressivos de cada mês $m = 1, \dots, s$ tal que todos os termos da FACPPE sejam nulos (ou não-significativos) para todas as defasagens (lags) k maiores que p_m . Isto pode ser feito obtendo-se estimativas de $\varphi_{kk}^m, k = 1, 2, \dots$ que são calculadas substituindo-se na equação (11) as autocorrelações ρ_i^{m-j} , para $i = 1, \dots, p_m$ e $j = 0, \dots, p_m - 1$, pelos respectivos valores amostrais e resolvendo-se esse sistema de equações para obter $\hat{\varphi}_{kk}^m$.

Para testar a hipótese nula $H_0 : \varphi_{kk} = 0$ para todos os valores de $k = 1, 2, \dots$, utiliza-se o resultado assintótico, o qual afirma que, se a ordem do modelo autorregressivo de um mês m qualquer é p_m , então o estimador $\hat{\varphi}_{kk}^m$ de φ_{kk}^m , para $k > p_m$ tem distribuição aproximadamente normal de média zero e variância $1/n$. Portanto não se rejeita H_0 com nível de significância $\alpha < 5\%$ se o zero está contido no intervalo $\hat{\varphi}_{kk}^m \pm 2/\sqrt{n}$. Neste caso se diz que a estimativa $\hat{\varphi}_{kk}^m$ é não significativa. Portanto o procedimento de identificação da ordem p_m , consiste em procurar, para cada mês m , a maior ordem p_m , tal que todas as estimativas $\hat{\varphi}_{kk}^m$ para $k > p_m$ não sejam mais significativas. Este procedimento será referido neste trabalho como o procedimento clássico de identificação.

Este procedimento clássico de identificação da ordem dos modelos PAR(p_m) pode levar a aceitação de processos como sendo ruídos brancos, quando, na verdade, trata-se de modelos que possuem baixos valores de ρ_k^m e φ_{kk}^m . Valores intermediários baixos de φ_{kk}^m para $k < p_m$ também podem ser considerados não-significativos e isto leva a identificação de modelos com parâmetros intermediários nulos, $\phi_{m,j} = 0$, para alguns $j < p_m$. Estas situações corroboram a ideia de que a forma clássica de identificação das ordens p_m pode não ser a mais adequada.

Por outro lado, (Stedinger, 2001) afirma que uma modelagem que considera individualmente cada mês não pode produzir o conjunto de modelos mais adequados. Sem maiores detalhes matemáticos e somente considerando a natureza dos fenômenos naturais envolvidos, ele argumenta que não

faz sentido considerar que a vazão de determinado mês m , devido a parâmetros intermediários nulos, dependa da vazão do mês $m - 2$ e não do mês $m - 1$. Para compreender melhor, suponha um modelo PAR(p_m) de ordem elevada ($p_m = 6$) que tenha somente estimativas estatisticamente significativas para os coeficientes $\phi_{m,1}$ e $\phi_{m,6}$. Portanto, o modelo tem quatro parâmetros $\phi_{m,2}, \phi_{m,3}, \phi_{m,4}$ e $\phi_{m,5}$ não-significativos. Neste caso, a proposta de (Stedinger, 2001) leva a supor que a correlação parcial φ_{06}^m seja espúria e um modelo de ordem $p_m = 1$ deve ser considerado adequado. Portanto, (Stedinger, 2001) propõe como critério para a seleção das ordens p_m , que para cada mês m , encontre-se a maior ordem p_m tal que todas as estimativas $\hat{\varphi}_{kk}$ para $k < p_m$ sejam significativas. Este critério, ao contrário da forma clássica de seleção, não admite valores intermediários de $\hat{\varphi}_{kk}$ não-significativos. Uma abordagem de identificação da ordem dos modelos PAR(p_m) usando técnicas de bootstrap foi proposta por (Oliveira et al., 2010) a qual avalia melhor os intervalos de confiança da FACP-Pe e corrobora com os valores de ordens p_m encontradas em (Stedinger, 2001).

Além dos dois procedimentos apresentados, um terceiro procedimento, utilizado em (MecLeod, 1994), considera o critério BIC proposto por (Schwarz, 1978) para selecionar as ordens p_m dos modelos autorregressivos de cada mês m . Neste caso o critério considera o cálculo de BIC_m , dado por

$$BIC_m = n \log(\tau_m^2) + p_m \log(n) \quad (12)$$

Este critério pondera entre a qualidade do ajuste, representada pela estimativa da variância do resíduo $\hat{\tau}_m^2$ e a complexidade do modelo, dado pela ordem p_m . Portanto escolhe-se as ordens p_m , tal que o critério $BIC = \sum_{m=1}^s BIC_m$ seja mínimo.

4 MSE DA PREVISÃO PARA SÉRIES DE VAZÕES SAZONAIS

Seja $z_{t(r,m)}$, $r = 1, \dots, n$, $m = 1, \dots, s$ uma série temporal padronizada observada até o mês m_0 do ano n . Denota-se por $T = t(n, m_0) = s(n - 1) + m_0$ o instante da última observação e por \hat{z}_{T+h} a previsão estimada para z_{T+h} com um passo de previsão h , considerando somente as observações conhecidas até o instante T . Então, o MSE desta previsão é dado por:

$$MSE(\hat{z}_{T+h}) = E(z_{T+h} - \hat{z}_{T+h})^2 \quad (13)$$

Mostra-se facilmente que o predictor \hat{z}_{T+h} de mínimo erro médio quadrático é o valor esperado condicional, dado por:

$$\hat{z}_{T+h} = E(z_{T+h} | z_T, z_{T-1}, \dots). \quad (14)$$

Para avaliar o MSE na escala original das vazões, supõe-se que são conhecidas as vazões históricas até um determinado

mês m_0 do ano n e se quer prever a vazão $q_{t(n+1,m)}$, no mês m do ano $(n+1)$. Considerando a relação (1), então a vazão $q_{t(n+1,m)}$ é escrita como:

$$q_{t(n+1,m)} = \mu_m + \sigma_m z_{t(n+1,m)} \quad (15)$$

Sendo $t(n+1, m) = sn + m$ e $t(n, m_0) = s(n-1) + m_0$ e considerando o passo da previsão $h = t(n+1, m) - t(n, m_0)$, (considera-se sempre que $1 \leq h \leq s$). Pode ser escrito então que $t(n+1, m) = t(n, m_0) + h$, assim, a equação (15) pode ser escrita como:

$$q_{t(n,m_0)+h} = \mu_m + \sigma_m z_{t(n,m_0)+h} \quad (16)$$

em (16) m_0 representa o mês de origem, do qual se faz a previsão e $m = m_0 + h - s$ é o mês no futuro, para o qual a vazão será prevista e h é o passo da previsão.

A previsão de mínimo erro médio quadrático para $q_{t(n+1,m)}$, feita do mês m_0 , é dada por:

$$\hat{q}_{t(n,m_0)+h} = E(q_{t(n,m_0)+h} | q_{t(n,m_0)}, q_{t(n,m_0)-1}, \dots) \quad (17)$$

o conhecimento de $\{q_{t(n,m_0)}, q_{t(n,m_0)-1}, \dots\}$ implica no conhecimento de $\{z_{t(n,m_0)}, z_{t(n,m_0)-1}, \dots\}$, então usando as equações (16) e (17) tem-se que:

$$\hat{q}_{t(n,m_0)+h} = \mu_m + \sigma_m \hat{z}_{t(n,m_0)+h} \quad (18)$$

O MSE associado a previsão $\hat{q}_{t(n,m_0)+h}$, é dado por:

$$MSE_m(\hat{q}_{t(n,m_0)+h}) = E(q_{t(n,m_0)+h} - \hat{q}_{t(n,m_0)+h})^2 = \sigma_m^2 E(z_{t(n,m_0)+h} - \hat{z}_{t(n,m_0)+h})^2 \quad (19)$$

Portanto, o MSE para a previsão da série sazonal tem um padrão sazonal com período s . Usa-se a notação $MSE_m(\hat{q}_{t(n,m_0)+h})$, referindo-se ao MSE da previsão de passo h , da vazão $q_{t(n,m_0)+h}$. O MSE, dado na equação (19), pode ser escrito como:

$$MSE_m(\hat{q}_{t(n,m_0)+h}) = \sigma_m^2 MSE(\hat{z}_{t(n,m_0)+h}) \quad (20)$$

Portanto, o MSE de uma previsão $\hat{q}_{t(n,m_0)+h}$ da série original $q_{t(n,m_0)+h}$ é proporcional ao MSE da previsão da série padronizada $\hat{z}_{t(n,m_0)+h}$ e a constante de proporcionalidade é a variância mensal σ_m^2 .

4.1 MSE com um processo independente

Considerando que o processo $z_{t(r,m)}$, $r = 1, \dots, n$, $m = 1, \dots, s$ é uma sequência de variáveis aleatórias independentes, então, o processo padronizado $z_{t(r,m)}$, $r = 1, \dots, n$, $m = 0, 1, \dots, s$ dado pela equação (1) também é uma sequência de variáveis aleatórias independentes.

Para calcular o MSE da previsão $\hat{q}_{t(n,m_0)+h}$ considera-se a equação (19). Com a hipótese de independência do processo $z_{t(r,m)}$, $r = 1, \dots, n$, $m = 1, \dots, s$, tem-se que:

$$E(z_{t(n,m_0)+h} | z_{t(n,m_0)}, z_{t(n,m_0)-1}, \dots) = E(z_{t(n,m_0)+h})$$

então a previsão é dada por:

$$\hat{z}_{t(n,m_0)+h} = E(z_{t(n,m_0)+h})$$

portanto,

$$E(z_{t(n,m_0)+h} - \hat{z}_{t(n,m_0)+h})^2 = Var(z_{t(n,m_0)+h}).$$

Se o processo $z_{t(r,m)}$, $r = 1, \dots, n$, $m = 1, \dots, s$ é um processo estacionário no sentido fraco, então tem-se que $Var(z_{t(r,m)}) = Var(z_{t(r,m)+h}) = \sigma_z^2$ para qualquer $t(r, m)$. Portanto, com a suposição de independência, o MSE da previsão $\hat{q}_{t(n,m_0)+h}$ é dado por:

$$MSE_m(\hat{q}_{t(n,m_0)+h}) = \sigma_m^2 \sigma_z^2 \quad (21)$$

Nota: Teoricamente a relação (1) assegura que $Var(z_{t(r,m)}) = 1$, na prática se constata que $\sigma_z^2 \approx 1$, portanto, do ponto de vista prático a suposição de independência leva a um $MSE_m(\hat{q}_{t(n,m_0)+h}) \approx \sigma_m^2$. A suposição de independência feita para $q_{t(r,m)}$ implica em desprezar toda e qualquer informação vinda das observações $\{q_{t(n,m_0)}, q_{t(n,m_0)-1}, \dots\}$. Se o único efeito das observações é diminuir o MSE da previsão, então o MSE dado em (21) pode ser considerado como o maior erro médio quadrático da previsão que se pode cometer, ou seja, σ_m^2 (a variância mensal incondicional) pode ser considerada como um limitante superior para o erro médio quadrático da previsão para qualquer mês de origem m_0 e qualquer passo de previsão h .

4.2 MSE com um processo PAR(p_m)

O modelo PAR(p_m), dado pela equação (4) pode ser escrito como uma soma infinita dos ruídos aleatórios periódicos como se segue.

$$z_{t(r,m)} = \sum_{j=0}^{\infty} \psi_{m,j} a_{t(r,m)-j} \quad \text{com } \psi_{m,0} = 1 \quad (22)$$

neste modelo tem-se que, para qualquer k inteiro, $\psi_{m,j} = \psi_{m+ks,j}$, e portanto o valor da série $z_{t(n,m_0)+h}$, calculada h passos à frente, pode ser escrito usando-se um modelo PAR(p_m), como:

$$z_{t(n,m_0)+h} = \sum_{j=0}^{\infty} \psi_{m,j} a_{t(n,m_0)+h-j}, \quad \text{com } \psi_{m,0} = 1 \quad (23)$$

Na equação (23) tem-se que $t(n, m_0) + h = t(n+1, m)$, e por isso os coeficientes $\psi_{m,j}$, que multiplicam os ruídos

$a_{t(n,m_0)+h-j}$, tem o índice m que se refere ao mês para o qual se considera a série $z_{t(n+1,m)}$.

Denota-se por $\hat{z}_{t(n,m_0)+h}$ a previsão para $z_{t(n,m_0)+h}$ calculada com h passos de antecedência usando-se um modelo $PAR(p_m)$. Então $\hat{z}_{t(n,m_0)+h}$ deve ser escrita em função dos valores passados $z_{t(n,m_0)}, z_{t(n,m_0)-1}, \dots$. Consequentemente, usando a equação (23), $\hat{z}_{t(n,m_0)+h}$ é escrita como uma função dos ruídos passados $a_{t(n,m_0)}, a_{t(n,m_0)-1}, \dots$. Além disto, suponha que $\hat{z}_{t(n,m_0)+h}$ é uma função linear. Então, denotando-se a melhor previsão por:

$$\hat{z}_{t(n,m_0)+h} = \sum_{j=0}^{\infty} \psi_{m,h+j}^* a_{t(n,m_0)-j} \quad (24)$$

Pode-se determinar os pesos $\psi_{m,h+j}^*$ que minimizem o erro médio quadrático (MSE) da previsão. Usando-se as equações (23) e (24), o erro de previsão, $e_{t(n,m_0)+h} = z_{t(n,m_0)+h} - \hat{z}_{t(n,m_0)+h}$, do modelo $PAR(p_m)$, com passo de previsão h , é escrito em função dos ruídos aleatórios, como:

$$e_{t(n,m_0)+h} = \sum_{j=0}^{\infty} \psi_{m,j} a_{t(n,m_0)+h-j} - \sum_{j=0}^{\infty} \psi_{m,h+j}^* a_{t(n,m_0)-j} \quad (25)$$

a equação (25) que pode ser reescrita como:

$$e_{t(n,m_0)+h} = \sum_{j=0}^{h-1} \psi_{m,j} a_{t(n,m_0)+h-j} - \sum_{j=0}^{\infty} (\psi_{m,h+j} - \psi_{m,h+j}^*) a_{t(n,m_0)-j} \quad (26)$$

Considerando o $MSE(\hat{z}_{t(n,m_0)+h})$ escrito em função do erro de previsão $e_{t(n,m_0)+h}$, dado por:

$$MSE(\hat{z}_{t(n,m_0)+h}) = E(e_{t(n,m_0)+h}^2). \quad (27)$$

então, tem-se da equação (27), que o mínimo $MSE(\hat{z}_{t(n,m_0)+h})$ ocorre quando $\psi_{m,h+j}^* = \psi_{m,h+j}$, para $j = 0, 1, \dots$. Sendo os ruídos aleatórios $a_{t(n,m_0)+h-j}$, $j = 0, \dots, h-1$ não correlacionados e periodicamente estacionários, tem-se que $Var(a_{t(n,m_0)+h-j}) = Var(a_{m_0+h-j}) = \tau_{m_0+h-j}^2$, para $j = 0, \dots, h-1$. A estacionariedade periódica permite escrever que $\tau_{m_0+h-j}^2 = \tau_i^2$ para $i = 1, \dots, s$. Os índices i são associados com cada um dos valores $m_0 + h - j$ por:

$$i = \begin{cases} m_0 + h - j - \nu s & \text{se } m_0 + h - j \neq s \\ s & \text{se } m_0 + h - j = s \end{cases} \quad (28)$$

em que $\nu = \text{mod}(m_0 + h - j, s)$ representa a parte inteira da divisão entre $m_0 + h - j$ e s . Então, o $MSE(\hat{z}_{t(n,m_0)+h})$ do modelo $PAR(p_m)$ pode ser avaliado, em função dos parâmetros do modelo e das variâncias dos resíduos, como:

$$MSE(\hat{z}_{t(n,m_0)+h}) = (\tau_{m_0+h}^2 + \psi_{m,1}^2 \tau_{m_0+h-1}^2 + \dots + \psi_{m,h-1}^2 \tau_{m_0+1}^2) \quad (29)$$

em que $\tau_{m_0+h-j}^2 = \tau_i^2$, para $i = 1, \dots, s$.

4.2.1 Cálculo dos parâmetros $\psi_{m,j}$

Para calcular os parâmetros $\psi_{m,j}$, $j = 1, \dots, h-1$ da equação (29), é necessário escrever o modelo $PAR(p)$ na forma da equação (22). Para isto, adota-se um procedimento recursivo e, para facilitar os cálculos, supõe-se, sem perda de generalidade, que os modelos $PAR(p_m)$ têm a mesma ordem $p = \max\{p_m\}$, para todos os meses $m = 1, \dots, s$. Esta suposição pode ser feita adotando-se $\phi_{m,k} = 0$ para $p_m < k \leq p$, em todos os modelos com $p_m < p$.

Para iniciar o procedimento recursivo de construção da equação (22), primeiro considera-se o modelo $PAR(p_m)$ escrito como um modelo $PAR(p)$, dado por:

$$z_{t(r,m)} = \phi_{m,1} z_{t(r,m)-1} + \phi_{m,2} z_{t(r,m)-2} + \dots + \phi_{m,p} z_{t(r,m)-p} + a_{t(r,m)} \quad (30)$$

Em seguida escreve-se o modelo $PAR(p)$ para $z_{t(r,m)-1}$ como segue:

$$z_{t(r,m)-1} = \phi_{m-1,1} z_{t(r,m)-2} + \phi_{m-1,2} z_{t(r,m)-3} + \dots + \phi_{m-1,p} z_{t(r,m)-1-p} + a_{t(r,m)-1} \quad (31)$$

Substituindo (31) em (30) obtêm-se:

$$z_{t(r,m)} = a_{t(r,m)} + \phi_{m,1} a_{t(r,m)-1} + (\phi_{m,1} \phi_{m-1,1} + \phi_{m,2}) z_{t(r,m)-2} + (\phi_{m,1} \phi_{m-1,2} + \phi_{m,3}) z_{t(r,m)-3} + \dots + (\phi_{m,1} \phi_{m-1,p} + \phi_{m,p}) z_{t(r,m)-p} \quad (32)$$

e repetindo o procedimento para $z_{t(r,m)-2}$, tem-se:

$$z_{t(r,m)} = a_{t(r,m)} + \phi_{m,1} a_{t(r,m)-1} + (\phi_{m,1} \phi_{m-1,1} + \phi_{m,2}) a_{t(r,m)-2} + [(\phi_{m,1} \phi_{m-1,1} + \phi_{m,2}) \phi_{m-2,1} + (\phi_{m,1} \phi_{m-1,2} + \phi_{m,3})] z_{t(r,m)-3} + \dots + [(\phi_{m,1} \phi_{m-1,1} + \phi_{m,2}) \phi_{m-2,p} + (\phi_{m,1} \phi_{m-1,p} + \phi_{m,p})] z_{t(r,m)-p} \quad (33)$$

Das equações (32) e (33) conclui-se que:

$$\begin{aligned}\psi_{m,1} &= \phi_{m,1} \\ \psi_{m,2} &= \psi_{m,1}\phi_{m-1,1} + \phi_{m,2} \\ \psi_{m,3} &= (\phi_{m,1}\phi_{m-1,1} + \phi_{m,2})\phi_{m-2,1} + \\ &\quad \phi_{m,1}\phi_{m-1,2} + \phi_{m,3} \\ &= \psi_{m,2}\phi_{m-2,1} + \psi_{m,1}\phi_{m-1,2} + \phi_{m,3}\end{aligned}$$

e por indução chega-se a fórmula geral dos coeficientes $\psi_{m,j}$, dada por:

$$\begin{aligned}\psi_{m,j} &= \phi_{m-j+1,1}\psi_{m,j-1} + \phi_{m-j+2,2}\psi_{m,j-2} + \dots \\ &\quad \dots + \phi_{m-1,j-1}\psi_{m,1} + \phi_{m,j}\end{aligned}\quad (34)$$

A equação (34) é válida para qualquer mês m , assim escreve-se, de forma geral, a equação para $\psi_{m,j}$, como:

$$\psi_{m,j} = \sum_{k=0}^{j-1} \phi_{m-k,j-k}\psi_{m,k}\quad (35)$$

Após serem ajustados os modelos $PAR(p_m)$, avalia-se o $MSE(\hat{z}_{t(n,m_0)+h})$, dado em (29), calculando-se recursivamente os coeficientes $\psi_{m,j}$ com a equação (35). Usando-se a equação (20), avalia-se o $MSE_m(\hat{q}_{t(n,m_0)+h})$ na escala original da série de vazões, como se segue:

$$\begin{aligned}MSE_m(\hat{q}_{t(n,m_0)+h}) &= \sigma_m^2(\tau_{m_0+h}^2 + \psi_{m,1}^2\tau_{m_0+h-1}^2 + \\ &\quad \dots + \psi_{m,h-1}^2\tau_{m_0+1}^2)\end{aligned}\quad (36)$$

Portanto, o cálculo do $MSE_m(\hat{q}_{t(n,m_0)+h})$, envolve não só a estimativa dos coeficientes $\psi_{m,j}$, $j = 1, \dots, h-1$ como também as estimativas das variâncias dos ruídos τ_i^2 , para $i = 1, \dots, s$.

No limite, quando $h \rightarrow \infty$, tem-se que:

$$\lim_{h \rightarrow \infty} MSE_m(\hat{q}_{t(n,m_0)+h}) = \sigma_m^2 \lim_{h \rightarrow \infty} \sum_{j=0}^{h-1} \psi_{m,j}^2 \tau_{i(m_0,h)}^2,$$

com $\psi_{m,0} = 1$, de forma que:

$$\lim_{h \rightarrow \infty} MSE_m(\hat{q}_{t(n,m_0)+h}) = \sigma_m^2 \sum_{j=0}^{\infty} \psi_{m,j}^2 \tau_j^2.\quad (37)$$

Portanto, da equação (22), pode ser mostrado que a série da equação (37) corresponde a variância incondicional da série padronizada $z_{t(r,m)}$. A qual, como pode ser vista em (1), é igual a 1. Desta forma, pode-se afirmar que, no caso dos modelos $PAR(p_m)$, a variância incondicional σ_m^2 é um limitante superior para o MSE da previsão para cada mês.

5 ESTUDO DE CASO

Neste estudo, considera-se os modelos ajustados às séries históricas de vazões mensais nos postos de Furnas, Emborcação e Sobradinho. Foram ajustados modelos autorregressivos periódicos $PAR(p_m)$, cujas ordens p_m foram selecionadas com a análise da função de autocorrelação parcial, considerando o procedimento clássico (Hippel and McLeod, 1994), no qual a identificação da ordem p_m , consiste em procurar para cada mês m a maior ordem p , tal que todas as estimativas da FACPPE $\hat{\varphi}_{kk}^m$, para $k > p$, não sejam mais significativas. Para comparação foram considerados também o método proposto em (Stedinger, 2001) e o método proposto por (McLeod, 1994), baseado no critério BIC.

As comparações entre os modelos de previsão são avaliadas considerando-se a variação da raiz quadrada do Erro Médio Quadrático ($RMSE$) da previsão, feita com cada modelo ajustado, em função do passo de previsão. São usados como referência os desvios-padrões mensais da série original.

Outra análise realizada refere-se a capacidade de previsão dos modelos $PAR(p_m)$ ajustados. Nesta análise considera-se o coeficiente de variação (cv_m) de cada mês, o qual consiste na razão entre o desvio-padrão e a Média de Longo Termo (MLT) de cada mês (e.g. $cv_m = \sigma_m/\mu_m$). Os meses com alto cv_m apresentam maiores erros de previsão e os modelos ajustados para estes meses apresentam baixa capacidade preditiva, ou seja, para estes meses observa-se que os erros de previsão tendem rapidamente para o limitante superior dos erros, definido pelo desvio-padrão mensal.

Na Tabela 1 são apresentadas, como ilustração, as ordens dos modelos $PAR(p_m)$ selecionadas com os três métodos considerados, para as vazões de Furnas. Na Tabela 2 encontram-se os modelos mensais ajustados para as vazões de Furnas quando a ordem de cada modelo foi selecionada pelo método clássico de análise da FACPPE. Portanto, estes modelos apresentam alguns coeficientes intermediários nulos.

Tabela 1: Ordens dos modelos para as vazões de Furnas.

Método	jan	fev	mar	abr	mai	jun
Clássico	5	1	6	2	3	1
(Stedinger, 2001)	1	1	1	2	3	1
BIC	1	2	1	2	4	1
Método	jul	ago	set	out	nov	dez
Clássico	2	1	4	6	5	6
(Stedinger, 2001)	2	1	4	4	1	2
BIC	2	1	4	4	1	2

O método de escolha da ordem de cada modelo, utilizado pelo método clássico, resulta em ordens mais elevadas do que as ordens escolhidas pelo método proposto por (Stedinger, 2001) ou pelo método proposto por (MecLeod, 1994) que faz uso do critério BIC. Os procedimentos propostos por (Stedinger, 2001) e (MecLeod, 1994) geralmente coincidem na escolha das ordens dos modelos, diferenciando-se em poucos meses e em apenas uma unidade na ordem selecionada (ver Tabela 1). Apesar do método clássico selecionar modelos de ordem mais elevadas, vários termos intermediários são nulos, como pode ser visto nos modelos ajustados para as vazões de Furnas, apresentados na Tabela 2.

Tabela 2: Modelos ajustados com o método clássico para as vazões de Furnas.

mês	$\phi_{m,1}$	$\phi_{m,2}$	$\phi_{m,3}$	$\phi_{m,4}$	$\phi_{m,5}$	$\phi_{m,6}$
1	0.5213	0	0	-0.3047	0.3022	0
2	0.4476	0	0	0	0	0
3	0.4323	0.2898	0	0	0.3324	-0.2905
4	0.5972	0.2641	0	0	0	0
5	0.5124	0	0.2727	0	0	0
6	0.8434	0	0	0	0	0
7	0.6051	0.3756	0	0	0	0
8	0.9368	0	0	0	0	0
9	0.5053	0.3296	0.4353	-0.4379	0	0
10	0.4019	-0.3266	0.4488	0.4665	0	-0.2799
11	0.7246	0.2518	-0.2603	0.4542	-0.4417	0
12	0.5170	0	0	0.5086	-0.5699	0.3503

Para análise dos erros de previsão usa-se como referência o coeficiente de variação (cv_m) de cada mês. Na Tabela 3 apresenta-se os coeficientes de variação cv_m e a razão entre o $RMSE$ (calculado com o passo de previsão $h = 1$) e a MLT de cada mês para as vazões de Furnas. Nota-se um alto coeficiente de variação no mês de setembro, devido as vazões ocorridas neste mês, nos anos de 1976 e 1983, que foram muito acima dos valores normalmente observados para as vazões de setembro.

Nota-se na Tabela 3 que os meses de janeiro, fevereiro e março apresentam maior erro relativo de previsão $RMSE/MLT$ (em %) quando comparados com os demais meses. Meses com alto coeficiente de variação são mais difíceis de serem previstos, consequentemente apresentam maiores erros de previsão, como pode ser visto na Tabela 3.

Tabela 3: Razão $RMSE/MLT$ (em %) com $h = 1$ para as vazões de Furnas.

Método	jan	fev	mar	abr	mai	jun
cv_m	40.95	39.39	40.72	34.79	30.46	42.09
Clássico	35.18	35.22	32.17	21.91	14.22	22.61
(Stedinger, 2001)	34.97	35.22	33.20	21.91	14.22	22.61
BIC	34.97	33.57	33.20	21.91	13.61	22.61
Método	jul	ago	set	out	nov	dez
cv_m	30.39	30.05	54.23	43.16	43.00	38.76
Clássico	10.13	10.51	27.47	23.25	28.91	27.84
(Stedinger, 2001)	10.13	10.51	27.47	23.78	29.29	28.17
BIC	10.13	10.51	27.47	23.78	29.29	28.17

Portanto, o aumento na ordem dos modelos, para os meses com alto coeficiente de variação, não resulta em diminuição significativa dos erros das previsões das vazões destes meses, como pode ser visto na Tabela 3, para os meses de janeiro, fevereiro e março, com o passo de previsão de um mês a frente ($h = 1$).

A Figura 1, ilustra os erros de previsão para as vazões afluentes em Furnas, no mês de janeiro, onde observa-se que os erros da previsão, cometidos com as três propostas de identificação consideradas neste trabalho, são praticamente iguais e crescem rapidamente, chegando a mais de 90% do limite superior nos três primeiros meses.

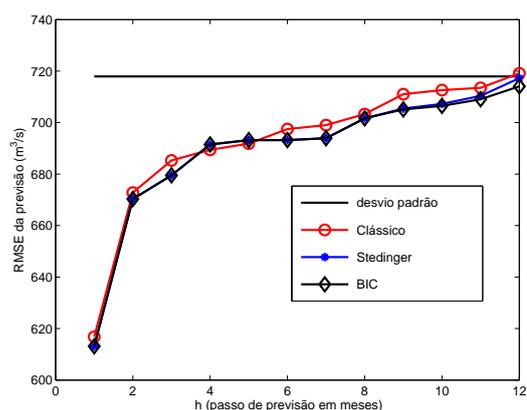


Figura 1: $RMSE$ da previsão da vazão de Furnas para o mês de janeiro.

No entanto, para os meses com baixo coeficiente de variação, como o mês de agosto, ao se considerar modelos de ordem mais elevada obtêm-se uma diminuição nos erros de previsão a longo prazo. Os gráficos apresentados na Figura 2 ilustram este fato para a previsão da vazão afluente à Furnas no mês de agosto. Independente do critério adotado na identificação da ordem dos modelos PAR, as ordens dos modelos ajustados para os três meses que precedem o mês de agosto são iguais. Este fato é refletido no erro de previsão para este mês, notando-se que somente para passos de previsão $h \geq 4$ surgem as diferenças nos erros de previsões que é provocada pela diferença nas ordens dos modelos ajustados.

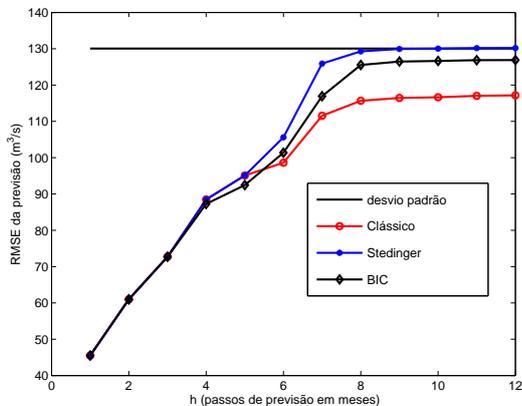


Figura 2: *RMSE* da previsão da vazão de Furnas para o mês de agosto.

Uma análise semelhante foi feita para as vazões afluentes à usina de Emborcação. Na Tabela 4 são apresentadas as ordens dos modelos selecionadas com as três propostas consideradas neste trabalho. Nota-se para estes dados, que o método clássico seleciona ordens mais elevadas que os outros dois métodos. Observando-se os modelos ajustados, apresentados na Tabela 5, nota-se que vários destes modelos apresentam estimativas dos parâmetros intermediários $\phi_{m,j}$, com $1 \leq j \leq p_m$ nulos, ressaltando a incoerência deste método de identificação da ordem dos modelos com os fenômenos naturais.

Tabela 4: Ordens dos modelos para as vazões de Emborcação.

Método	jan	fev	mar	abr	mai	jun
Clássico	6	1	1	5	3	1
(Stedinger, 2001)	1	1	1	1	1	1
BIC	1	2	1	2	3	1
Método	jul	ago	set	out	nov	dez
Clássico	2	3	4	6	4	6
(Stedinger, 2001)	2	1	1	1	2	1
BIC	2	3	1	1	2	1

Tabela 5: Modelos ajustados com o método clássico para as vazões de Emborcação.

mês	$\phi_{m,1}$	$\phi_{m,2}$	$\phi_{m,3}$	$\phi_{m,4}$	$\phi_{m,5}$	$\phi_{m,6}$
1	0.3717	0.2484	-0.3966	0	0	0.4942
2	0.6083	0	0	0	0	0
3	0.4916	0	0	0	0	0
4	0.5483	0	0	0	0.2616	0
5	0.6656	0	0.2909	0	0	0
6	0.8858	0	0	0	0	0
7	0.6854	0.3107	0	0	0	0
8	0.8930	0	0.2860	0	0	0
9	1.0056	0.4773	-0.2937	-0.3516	0	0
10	0.3834	0	0.3716	-0.3475	0	0.3689
11	0.7310	0	-0.5455	0.4214	0	0
12	0.5493	0	0	0.4746	-1.1215	0.6238

A Tabela 6 apresenta, para cada mês do ano, o coeficiente de variação e a razão *RMSE/MLT* (%) para cada método de identificação da ordem do modelo *PAR*(*pm*), para a usina de Emborcação. O mês de fevereiro apresenta o maior coeficiente de variação e consequentemente os maiores erros de previsão com $h = 1$. A Figura 3 mostra como estes erros de previsão tendem rapidamente para o limite superior, definido pelo desvio-padrão das vazões deste mês, quando o passo de previsão cresce. Resultados semelhantes são encontrados para os meses de janeiro, março, novembro e dezembro (ver (Projeto estratégico de P & D da ANEEL: ODIN - Otimização do Despacho Interligado Nacional, 2010)).

Tabela 6: Razão *RMSE/MLT* (em %) com $h = 1$ para as vazões de Emborcação.

Método	jan	fev	mar	abr	mai	jun
cv_m	38.59	47.03	42.23	37.64	30.92	29.99
Clássico	34.07	37.33	36.77	24.32	11.52	13.91
(Stedinger, 2001)	35.28	37.33	36.77	26.36	14.11	13.91
BIC	35.28	36.01	36.77	25.47	11.52	13.91
Método	jul	ago	set	out	nov	dez
cv_m	29.32	30.04	32.01	34.87	47.01	46.20
Clássico	7.01	8.94	15.01	28.49	38.08	38.97
(Stedinger, 2001)	7.01	8.94	15.62	28.49	38.12	38.86
BIC	7.01	8.23	15.62	28.49	38.12	38.86

Para os meses com baixo coeficiente de variação observa-se a diminuição do erro de previsão para os modelos de ordem maiores, isso é ilustrado na Figura 4, a qual apresenta os erros das previsões para as vazões de Emborcação no mês de julho.

Na Tabela 7 são apresentadas as ordens dos modelos selecionadas, com os três critérios considerados neste trabalho, para as vazões de Sobradinho. Nota-se que todos os métodos resultam na mesma escolha da ordem dos modelos em quase todos os meses, exceto para os meses de fevereiro e dezembro, nos quais o critério clássico selecionou modelos de ordem mais elevada e nos meses de junho e julho, no qual o critério BIC selecionou modelos de ordem mais elevada.

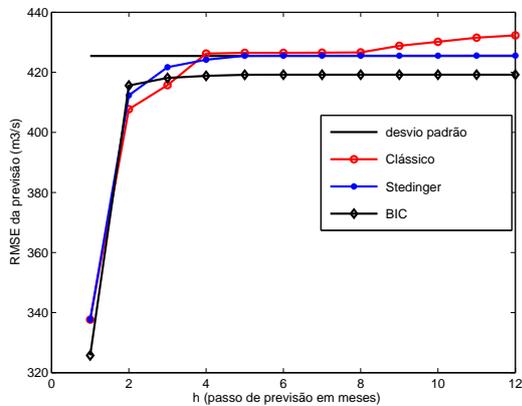


Figura 3: *RMSE* da previsão da vazão de Emborcação para o mês de fevereiro.

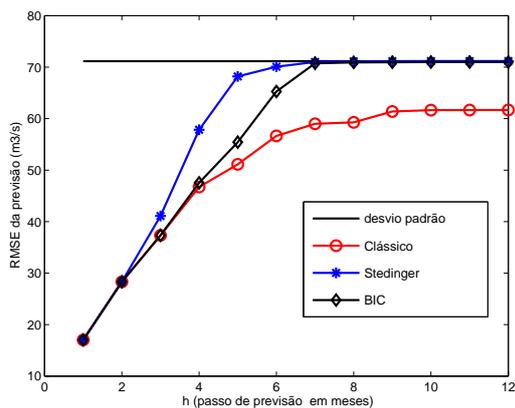


Figura 4: *RMSE* da previsão da vazão de Emborcação para o mês de julho.

Tabela 7: Ordens dos modelos para as vazões de Sobradinho.

Método	jan	fev	mar	abr	mai	jun
Clássico	1	4	1	1	1	2
(Stedinger, 2001)	1	1	1	1	1	2
BIC	1	1	1	1	1	5
Método	jul	ago	set	out	nov	dez
Clássico	1	1	1	1	1	5
(Stedinger, 2001)	1	1	1	1	1	1
BIC	2	1	1	1	1	1

A Tabela 8 apresenta as estimativas dos parâmetros dos modelos ajustados para a vazão de Sobradinho, em cada mês,

quando foi adotado o critério clássico de seleção das ordens dos modelos. Nota-se nesta tabela, que os modelos de ordem elevada, apresentam estimativas nulas para vários parâmetros intermediários, ressaltando-se a incoerência entre o método utilizado e o fenômeno natural.

Tabela 8: Modelos ajustados com o método clássico para as vazões de Sobradinho.

mês	$\phi_{m,1}$	$\phi_{m,2}$	$\phi_{m,3}$	$\phi_{m,4}$	$\phi_{m,5}$	$\phi_{m,6}$
1	0.5744	0	0	0	0	0
2	0.4232	0	0	0.3384	0	0
3	0.7476	0	0	0	0	0
4	0.6118	0	0	0	0	0
5	0.7979	0	0	0	0	0
6	0.6778	0.3148	0	0	0	0
7	0.9705	0	0	0	0	0
8	0.9783	0	0	0	0	0
9	0.9455	0	0	0	0	0
10	0.7969	0	0	0	0	0
11	0.6323	0	0	0	0	0
12	0.5630	0	0	-0.9262	0.8277	0

A Tabela 9 apresenta, para cada mês do ano, o coeficiente de variação e a razão *RMSE/MLT* (%) para cada método de identificação da ordem do modelo $PAR(p_m)$, na usina de Sobradinho. As vazões afluentes ao reservatório de Sobradinho apresentam o maior coeficiente de variação no mês de março e o menor se verifica no mês de setembro. A variação do erro de previsão em função do passo de previsão para estes meses são apresentados nos gráficos das Figura 5 e 6.

Tabela 9: Razão *RMSE/MLT* (em %) com $h = 1$ para as vazões de Sobradinho.

Método	jan	fev	mar	abr	mai	jun
cv_m	29.40	40.68	53.17	44.78	48.45	32.24
Clássico	24.06	36.64	35.31	35.43	29.21	10.32
(Stedinger, 2001)	24.06	37.40	35.31	35.43	29.21	10.32
BIC	24.06	37.40	35.31	35.43	29.21	8.22
Método	jul	ago	set	out	nov	dez
cv_m	27.41	25.07	24.98	25.76	37.30	34.54
Clássico	6.61	5.19	8.13	15.56	28.89	27.44
(Stedinger, 2001)	6.61	5.19	8.13	15.56	28.89	27.44
BIC	6.38	5.19	8.13	15.56	28.89	27.44

Observa-se na Figura 5, que o erro de previsão para o mês de março tende rapidamente para o limite superior definido pelo desvio-padrão das vazões deste mês. Nota-se, nestes gráficos, que estes erros praticamente coincidem para os modelos selecionados com os três critérios adotados. Na Figura 6, observa-se que a evolução do erro corre de forma mais lenta, destacando-se a capacidade preditiva dos modelos para horizontes maiores.

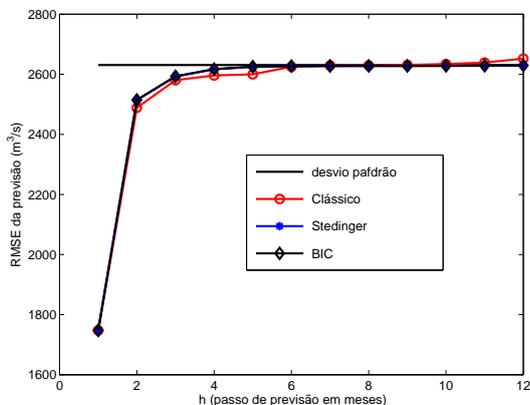


Figura 5: MSE da previsão da vazão para o mês de março para o posto de Sobradinho.

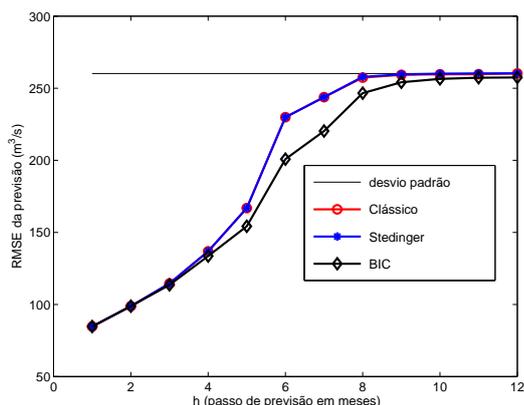


Figura 6: MSE da previsão da vazão para o mês de setembro para o posto de Sobradinho.

Quando se utiliza o critério BIC não considera-se testes de hipóteses para os parâmetros dos modelos, portanto, com este critério, não se admite valores intermediários nulos. Pode-se observar que nos três estudos de caso considerados, em algumas situações o critério BIC selecionou ordem superior ao método proposto em (Stedinger, 2001) e eventualmente ao método clássico, como é visto na Tabela 1 para Furnas, nos meses de fevereiro e maio, na Tabela 4 para Emborcação, no mês de fevereiro e na Tabela 7 para Sobradinho, nos meses de junho e julho. Embora tais parâmetros possam ser considerados nulos, segundo algum teste de hipótese, uma diminuição na razão RMSE/MLT ($h = 1$) é verificada nestes casos, como se pode notar na Tabela 3, para os meses de fevereiro e maio; Tabela 6, para o mês de fevereiro e Tabela 9, para os meses de junho e julho. Estes resultados indicam que tais testes de hipóteses, os quais são fundamentados em propri-

edades assintóticas, válidas somente para grandes amostras, podem ser imprecisos para o ajuste dos modelos $PAR(p_m)$, uma vez que, nestes ajustes o tamanho da amostra corresponde ao número de anos disponíveis no histórico de vazões, portanto, tem-se amostras de tamanho pequeno do ponto de vista da teoria assintótica. Recomenda-se que os critérios propostos por (Stedinger, 2001) ou o BIC sejam utilizados quando opta-se pelo critério da parcimônia. O uso do critério clássico, seleciona sempre ordens mais elevadas que produzem uma pequena diminuição na razão RMSE/MLT ($h = 1$). No entanto, esta pequena redução no erro de previsão não justifica a complexidade proveniente do uso de modelos de ordens elevadas.

6 CONCLUSÃO

Neste trabalho foram calculados os erros médios quadráticos das previsões de vazões em função do passo de previsão, para modelos $PAR(p_m)$. Foram considerados três métodos de identificação da ordem dos modelos. Observa-se que o uso do método clássico de análise da FACPPe leva a escolha de ordens elevadas para alguns meses. No entanto, os modelos para estes meses apresentam estimativas nulas para alguns parâmetros intermediários. Considerar parâmetros nulos entre o primeiro e o último parâmetro de um modelo $PAR(p_m)$, ressalta uma incoerência entre o modelo e o fenômeno natural. Portanto, o método clássico de análise da FACPPe não é adequado para identificar a ordem dos modelos $PAR(p_m)$. Os outros dois métodos considerados, identificam as ordens de forma coerente.

A variação dos erros de previsão em função do passo de previsão são apresentadas na forma de gráficos. Nota-se, ao analisar estes gráficos, para os meses cujos coeficientes de variação são mais elevados, que o erro de previsão tende rapidamente para o limite superior. Este fato indica que os modelos ajustados para estes meses apresentam uma capacidade preditiva limitada a um horizonte de previsão máximo de três meses ($h = 3$). Para os meses com coeficiente de variação menores, a capacidade preditiva dos modelos pode chegar até $h = 6$ meses. Para horizontes de previsão superiores a seis meses, os erros de previsão dos modelos tornam-se aproximadamente iguais ao desvio-padrão incondicional da série. Ou seja, para $h \geq 6$, os modelos perdem quase toda capacidade preditiva e comportam-se como se as vazões não apresentassem correlação temporal.

Apesar do reduzido erro de previsão para o passo de previsão de um mês ($h = 1$) quando o valor absoluto deste erro é comparado com a MLT do mês para a qual a vazão está sendo prevista, nota-se que a redução no erro de previsão só ocorre de forma mais acentuada nos meses considerados secos, ou seja entre junho e setembro. Nos meses úmidos,

novembro a abril, em geral as relações entre o *RMSE* e a *MLT* do mês são muito próximas.

Os resultados apresentados neste trabalho dão uma indicação de que é possível que os modelos de planejamento da operação de sistemas hidrelétricos, que utilizam os modelos estocásticos das vazões de forma explícita, não consigam se beneficiar destes modelos devido a baixa capacidade preditiva dos mesmos, quando comparados ao horizonte de planejamento (geralmente maior que um ano). Desta forma, é possível que os modelos de planejamento da operação que se utilizam das previsões, ou mesmo consideram as vazões de forma determinísticas, possam apresentar resultados muito próximos aos modelos estocásticos, uma vez que para horizontes superiores a seis meses ($h \geq 6$), os erros de previsão tornam-se grandes e consequentemente a *MLT* fica muito próxima das vazões previstas.

AGRADECIMENTOS

Este trabalho foi financiado pelo projeto estratégico de P&D da ANEEL denominado ODIN - Otimização do Despacho Interligado Nacional - com apoio de 18 concessionárias lideradas pela Bandeirante Energia S.A. do grupo EDP, e pelas agências de financiamento à pesquisa CNPq e FAPESP.

REFERÊNCIAS

- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008). *Time series analysis: forecasting and control*, 4th. edition edn, Holden-Day.
- Hippel, K. W. and McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*, Elsevier Science B.V., Amsterdam, Netherlands.
- Maceira, M. E. P., Terry, L., Costa, F. S. Damzio, J. M. and Melo, A. C. G. (2002). Chain of optimization models for setting the energy dispatch and spot price in the Brazilian system, *14th Power Systems Computation Conference*, Sevilla, Spanish.
- Martinez, L. and Soares, S. (2002). Comparison between closed-loop and partial open-loop feedback control policies in long term hydrothermal scheduling, *IEEE Trans. on Power Syst.* **17**(2): 330–336.
- McLeod, A. I. (1994). Diagnostic checking of periodic autoregression models with application, *J. of Time Series Analysis* **15**(2): 221–233.
- Mondal, M. S. and Wasimib, S. A. (2006). Generating and forecasting monthly flows of the Ganges river with par model, *Journal of Hydrology* **323**: 41–56.
- Oliveira, F. L. C., Miranda, C. V. C., Ferreira, P. G. C., Souza, R. C. and Calili, R. F. (2010). Critérios de identificação da ordem do modelo autorregressivo periódico - par(p), in SBPO (ed.), *XLII Simpósio Brasileiro de Pesquisa Operacional*, Vol. XLII SBPO, Bento Gonçalves, RS., pp. 753–763.
- Projeto estratégico de P & D da ANEEL: ODIN - Otimização do Despacho Interligado Nacional (2010). Análise do erro de previsão em funo do horizonte de previsão para séries de vazões médias mensais: Parte ii, *Rel. Tc.: ANEEL PE-0391-0108/2009 008/2010*, Faculdade de Engenharia Elétrica e Computação - Unicamp, Campinas / SP.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics* **6**(12): 461–464.
- Stedinger, J. R. (2001). Report on the evaluation of CEPTEL's par models, *Technical report*, School of Civil and Environmental Engineering - Cornell University, Ithaca, New York.
- Stedinger, J. R., Sule, B. F. and Loucks, D. P. (1984). Stochastic dynamic programming models for reservoir operation optimization, *Water Resources Research* **20**(11): 1499–1505.
- Vechia, A. V. (1985). Periodic autoregressive moving average (parma) modeling with application to water resources, *Water Resources Bull.* **21**: 721–730.