

## Modelagem e Previsão de Vazões Médias Mensais do Rio Potiribu Utilizando Modelos de Séries Temporais

Débora Missio Bayer; Nilza Maria dos Reis Castro

Instituto de Pesquisas Hidráulicas – UFRGS

deborabayer@gmail.com; nilza@iph.ufrgs.br

Fábio Mariano Bayer

Departamento de Estatística – UFSM

bayer@ufsm.br

Recebido: 06/07/11 - revisado: 23/10/11 - aceito: 12/12/11

---

### RESUMO

Este trabalho apresenta uma análise de séries temporais de dados de vazões médias mensais utilizando modelos estocásticos da classe ARIMA (autorregressivo integrado de médias móveis). A análise de séries temporais teve por objetivo modelar e prever as vazões da bacia hidrográfica do rio Potiribu, localizada na região noroeste do Rio Grande do Sul e afluente do rio Ijuí. Para selecionar o de melhor desempenho foram testados diversos modelos do tipo SARIMA, ou seja, modelos ARIMA que consideram a característica sazonal da série. Dentre os modelos candidatos foram selecionados aqueles que obtiveram os menores valores do critério de seleção de modelos AIC (Critério de Informação de Akaike). O modelo escolhido foi o SARIMA(3,0,0)(2,1,2)<sub>12</sub>, que obteve um bom ajuste à série de vazões observadas, com valor do coeficiente de Nash-Sutcliffe (CNS) de 0,68. O referido modelo foi utilizado para realizar previsões de vazões com antecedência de seis meses à frente. No horizonte previsto foi obtido um valor de 0,81 para o CNS.

**Palavras-chave:** ARIMA, modelos estocásticos, modelos hidrológicos, vazão.

---

### INTRODUÇÃO

Conhecer as bacias hidrográficas e conseguir descrever e prever o comportamento das variáveis envolvidas no ciclo hidrológico têm sido um desafio ao longo dos anos para muitos pesquisadores.

Em Hidrologia existem dois modos de descrever as variáveis envolvidas no ciclo hidrológico e obter previsões: (i) usar equações com base conceitual ou empírica para descrever os processos físicos, que são chamados de modelos determinísticos; (ii) considerar que os processos são controlados por leis probabilísticas, ou seja, modelos estocásticos. (Tucci, 1998)

Dentre os modelos estocásticos, os modelos de séries temporais da classe ARIMA (autorregressivo integrado de médias móveis) possuem grande destaque. Entendendo que uma série temporal é um conjunto de observações ordenadas no tempo, conclui-se que os processos que compõem o ciclo hidrológico como chuva, vazão, evaporação, infiltração, entre outros, por serem objetos de medições siste-

máticas em tempos discretos, podem ser tratados como séries temporais.

A análise de séries temporais consiste em procurar alguma relação de dependência temporal entre os dados, identificando o mecanismo gerador da série com o objetivo de extrair periodicidades relevantes nas observações; descrever o seu comportamento e fazer previsões (Bayer e Souza, 2010).

A previsão de vazões tem uma significativa importância no gerenciamento dos recursos hídricos. Essas previsões podem ser realizadas tanto a curto prazo, geralmente utilizando modelos determinísticos, quanto a longo prazo, utilizando modelos estocásticos (Colischonn et al, 2005). As previsões a longo prazo trazem como benefício direto a estimativa de afluências aos reservatórios hidrelétricos, o que permite planejar a sua melhor operação. Além disso, a previsão da disponibilidade hídrica é importante para outros usos como irrigação, abastecimento, navegação, etc. (Colischonn et al, 2005).

O objetivo deste trabalho é realizar a análise e modelagem da série temporal de vazões médias mensais do rio Potiribu. A análise considerará modelos sazonais da classe ARIMA, conhecidos como

modelos SARIMA, para modelagem e previsão da variável de interesse.

Este artigo está organizado em seis seções: a primeira seção apresenta um breve resumo de algumas aplicações da metodologia; a segunda mostra o referencial teórico dos modelos estocásticos, em particular os da classe ARIMA, e medidas de qualidade de ajuste; a terceira caracteriza a bacia em estudo; na quarta seção é apresentada a análise descritiva dos dados, seguida da modelagem dos dados de acordo com a metodologia Box e Jenkins (Box et al, 2008) e das previsões passos a frente. Na última seção encontram-se as conclusões.

**Aplicações de modelos ARIMA na hidrologia**

Modelos estatísticos de análise de séries temporais da classe ARIMA têm sido pouco utilizados na área hidrológica, apesar de seu grande poder de modelagem e de previsão. Alguns trabalhos antigos introduziram o uso de séries temporais na hidrologia, como pode ser visto em Ledolter (1976, 1977). Atualmente o seu uso tem sido mais restrito devido, principalmente, ao grande foco da área no uso de modelos determinísticos. Os conceitos e a teoria acerca da análise e da modelagem de séries temporais hidrológicas e suas aplicações na engenharia são apresentados de forma detalhada em Salas et al (1980) e Salas (1992). Modarres e Eslami-an (2006) utilizaram modelos do tipo ARIMA para modelar vazões médias mensais do rio Zayandehrud, localizado na Província de Isfahan Ocidental, no Irã. Lucas et al (2009) utilizaram dois modelos, um determinístico e outro estocástico do tipo ARIMA, para simular a vazão média mensal das sub-bacias da região hidrográfica do Xingu no Estado do Pará. Segundo os autores, o modelo estocástico ARIMA conseguiu capturar a dinâmica das séries temporais, apresentando resultados muito satisfatórios na simulação da vazão média mensal. Birinci e Akay (2010) utilizaram modelos do tipo ARIMA para a previsão de precipitações, que, por sua vez, são insumos (*inputs*) em modelos de redes neurais artificiais para previsão de vazões diárias. Já no trabalho realizado por Freitas et al (2010), são utilizados modelos de séries temporais com o objetivo de identificar tendências anuais de precipitação e de vazão na Bacia Hidrográfica do Rio São Mateus, no Estado do Espírito Santo. Outra boa aplicação de modelos do tipo ARIMA foi feita por Gemitzi e Stefanopoulos (2011) para investigar os efeitos das condições meteorológicas e das intervenções do homem sobre os aquíferos subterrâneos. Os autores obtiveram bons resultados e sugerem o uso de modelos estocásticos como

uma alternativa, especialmente devido à limitação na obtenção de dados em relação a propriedades hidráulicas e geométricas dos aquíferos.

**MODELOS ESTOCÁSTICOS**

Um processo estocástico é uma família de variáveis aleatórias que se supõem definidas em um mesmo espaço de probabilidades. Formalmente, conforme Morettin e Toloi (2006), sendo  $T$  um conjunto arbitrário, um processo estocástico é uma família  $Z = \{Z_t, t \in T\}$ , tal que, para cada  $t \in T$ ,  $Z_t$  é uma variável aleatória. Sendo assim, uma série temporal é, na verdade, uma realização de um processo estocástico.

Um processo é dito estacionário se ele oscila ao redor de uma média ( $\mu$ ) constante e com uma variância ( $\sigma^2$ ) também constante. Em uma definição mais formal, segundo Morettin e Toloi (2006), um processo estocástico  $Z = \{Z_t, t \in T\}$  é dito estacionário se, e somente se,

$$E[Z_t] = \mu_T = \mu, \text{ para todo } t \in T, \tag{1}$$

$$E[Z_t^2] < \infty, \text{ para todo } t \in T, \tag{2}$$

$$\gamma_h = Cov[Z_{t_1}, Z_{t_2}] \text{ é função de } h = |t_1 - t_2|, \tag{3}$$

em que  $E[Z]$  indica o valor esperado da variável aleatória  $Z$  e a covariância  $Cov[\cdot]$  é definida neste artigo na Equação 5.

A introdução do conceito de estacionariedade é importante para o estudo de séries temporais, pois esta é a suposição mais frequente. No entanto, uma grande parte das séries encontradas na prática são não-estacionárias, podendo ser estacionarizadas por meio do processo de diferenças sucessivas, o qual é apresentado mais a frente.

A correlação serial é chamada de auto-correlação, pois se refere a medidas de associação entre instantes de tempo de uma mesma série temporal. A função de auto-correlação é dada por

$$\rho_h = Cor[Z_t, Z_{t+h}] = \frac{Cov[Z_t, Z_{t+h}]}{\sqrt{\sigma_t^2 \sigma_{t+h}^2}}, \tag{4}$$

em que  $h$  é um número inteiro e

$$\gamma_h = Cov[Z_t, Z_{t+h}] = E[(Z_t - \mu_t)(Z_{t+h} - \mu_{t+h})] \tag{5}$$

é a covariância. A função de auto-correlação parcial mede a correlação remanescente entre  $Z_t$  e  $Z_{t+h}$ , depois de eliminada a influência de  $Z_{t+1}, \dots, Z_{t+h-1}$ . Para maiores detalhes ver Morettin e Toloi (2006) ou Box et al(2008).

As definições acima se referem às populações. Na prática, utilizam-se estimadores desses parâmetros, ou seja, considerando as  $n$  observações  $z_1, \dots, z_n$ , a função de auto-correlação  $\rho_h$  é estimada por

$$\hat{\rho}_h = r_h = \frac{c_h}{c_0}, \quad h=0, 1, \dots, (n-1), \quad (6)$$

em que  $\hat{\gamma}_h$  é a estimativa da covariância, dada por

$$\hat{\gamma}_h = \frac{1}{n} \sum_{t=1}^{(n-h)} [(z_t - \bar{z})(z_{t+h} - \bar{z})], \quad (7)$$

com  $h=0, 1, \dots, (n-1)$  e

$$\bar{z} = \frac{1}{n} \sum_{t=1}^n z_t. \quad (8)$$

As estimativas da função de auto-correlação parcial podem ser obtidas estimando sucessivamente modelos autorregressivos de ordens  $p = 1, 2, 3, \dots$  por mínimos quadrados e tomando as estimativas dos parâmetros de maior defasagem (Morettin e Toloi, 2006).

Outra definição importante acerca dos processos estocásticos é a de ruído branco, em que  $\{\varepsilon_t, t \in T\}$  é dito ser um ruído branco se as variáveis aleatórias  $\varepsilon_t$  são não-correlacionadas, isto é,  $Cov[\varepsilon_t, \varepsilon_s] = 0, t \neq s$  (Morettin e Toloi, 2006).

Na classe dos modelos paramétricos os mais frequentemente utilizados são: os modelos de regressão, os modelos autorregressivos integrados de médias móveis (ARIMA), os modelos sazonais autorregressivos integrados de médias móveis (SARIMA), os modelos de memória longa (ARFIMA) e os modelos não-lineares.

**Modelos da classe ARIMA**

De acordo com Bayer e Souza (2010) a classe de modelos tradicionalmente mais utilizada em análise de séries temporais é a dos modelos ARIMA.

Se  $Z_t$  segue um processo autorregressivo e de médias móveis de ordens  $p$  e  $q$ , ARMA( $p, q$ ), então

$$\phi(B)Z_t = c + \theta(B)\varepsilon_t, \quad (9)$$

em que  $\varepsilon_t$  é ruído branco com média zero e variância  $\sigma^2$ ,  $c$  é um parâmetro que permite ao processo ter média diferente de zero, o polinômio  $B^k Z_t = Z_{t-k}$  é o operador retroativo,  $\phi(B)$  e  $\theta(B)$  são o polinômio autorregressivo e o polinômio de médias móveis, respectivamente, apresentados nas Equações 10 e 11:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad (10)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \quad (11)$$

Como foi citado anteriormente, uma série que não apresenta característica de estacionariedade pode tornar-se estacionária pela utilização de um processo de diferenciação. Sendo assim, se uma série se torna estacionária após  $d$  diferenças, a série é dita integrada (I) de ordem  $d$ . Com isso, o modelo ARMA integrado passa a ser denominado de modelo ARIMA.

O modelo ARIMA( $p, d, q$ ) tem a seguinte forma:

$$\phi(B)(1 - B)^d Z_t = \theta(B)\varepsilon_t, \quad (12)$$

em que  $d$  é a ordem de integração, ou seja, o número de diferenças necessárias para tornar a série estacionária.

Contudo, em estudos hidrológicos é comum encontrar sazonalidade nas séries. Isso nos remete à necessidade de utilizar modelos que incorporem essa importante característica temporal. Neste contexto, ampliam-se os modelos ARIMA para os modelos SARIMA.

O modelo SARIMA( $p, d, q$ )( $P, D, Q$ ) é dado por:

$$\phi(B)\Phi(B)(1 - B)^d(1 - B^s)^D Z_t = \theta(B)\Theta(B)\varepsilon_t, \quad (13)$$

em que ( $p, d, q$ ) são as ordens do modelo referentes à dinâmica ordinal já mencionados e ( $P, D, Q$ ) são as ordens da parte sazonal. Os parâmetros  $\Phi_1, \dots, \Phi_P$  e  $\Theta_1, \dots, \Theta_Q$  são os parâmetros autorregressivos sazonais e de médias móveis sazonais, respectivamente. Para observações mensais, com sazonalidade de 12 meses, temos que  $s = 12$ .

Para estabelecer um modelo da classe ARIMA é utilizado a metodologia Box & Jenkins (Box et al, 2008), que é definida pelas seguintes etapas: (i) identificação do modelo; (ii) estimação dos parâmetros e (iii) diagnóstico do modelo ajustado.

Na fase de identificação são determinadas as ordens de integração  $d$  e  $D$  e as ordens  $p$ ,  $q$ ,  $P$  e  $Q$  do modelo. Essa identificação pode ser feita através da análise dos correlogramas. Contudo, um método bastante utilizado para identificar as ordens do modelo se dá na utilização dos critérios de seleção de modelos, também conhecidos como critérios de informação. A utilização desses critérios faz com que as fases de identificação e estimação do modelo ocorram simultaneamente. A estimação dos parâmetros do modelo é realizada pelo método da máxima verossimilhança. Para maiores detalhes consultar Morettin e Toloi (2006).

Dentre os critérios de seleção de modelo o AIC (*Akaike Information Criterion*), proposto por Akaike (1973), tem sido muito utilizado. Esse critério é dado por:

$$AIC = -2\ln[L(\hat{\xi})] + 2k, \tag{14}$$

em que  $L(\cdot)$  é a função de verossimilhança,  $\hat{\xi}$  são as estimativas de máxima verossimilhança do vetor paramétrico  $\xi$  e  $k$  é o número de parâmetros do modelo. No AIC o termo  $-2\ln [L(\hat{\xi})]$  é uma medida de qualidade do ajuste e  $2k$  é um termo penalizador, uma vez que o mesmo aumenta com a inclusão de parâmetros no modelo. Dessa forma, o AIC busca selecionar modelos parcimoniosos, equilibrando explicabilidade e complexidade do modelo.

A fase de diagnóstico tem por objetivo verificar se o modelo, com seus parâmetros já estimados, representa adequadamente os dados. Em geral os testes de adequação do modelo são baseados nas auto-correlações estimadas a partir dos resíduos. Se o modelo estiver bem ajustado os resíduos estimados devem estar próximos dos erros verdadeiros e, portanto, não-correlacionados. Isso pode ser verificado no correlograma residual.

O ajuste do modelo também pode ser verificado pelo teste de Ljung-Box (Ljung e Box, 1978). Nesse teste estatístico a hipótese nula a ser testada é a de que os erros são um processo ruído branco. Portanto, ao não rejeitar a hipótese nula, ou seja, quando o  $p$ -valor decorrente do teste estatístico é maior do que o nível de significância  $\alpha$ , conclui-se que o modelo está adequadamente ajustado.

**Medidas de qualidade**

Além do critério de seleção de modelos (AIC), comentado anteriormente, algumas medidas de qualidade podem auxiliar na escolha do melhor modelo para a série temporal e, principalmente, na

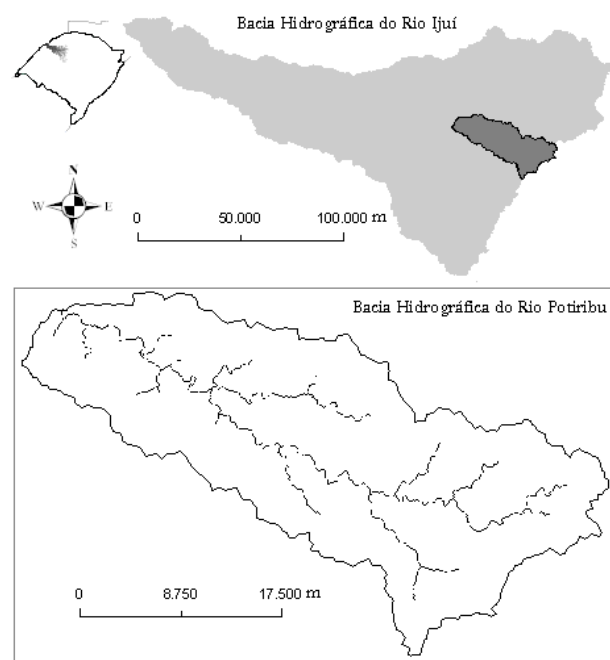
avaliação da qualidade do ajuste. As medidas utilizadas neste trabalho estão definidas na Tabela 1, em que  $z_i$  e  $\hat{z}_i$  são os valores observados e previstos no instante  $i$ , respectivamente, e  $\bar{Z}$  é a média aritmética da série observada.

**Tabela 1 - Medidas de qualidade das previsões.**

Critério	Expressão
Erro quadrático médio - EQM	$\frac{1}{h} \sum_{i=1}^h (z_i - \hat{z}_i)^2$
Erro agregado percentual - EAP	$\frac{\sum_{i=1}^h z_i - \sum_{i=1}^h \hat{z}_i}{\sum_{i=1}^h z_i}$
Coefficiente de Nash-Sutcliffe - CNS	$1 - \frac{\sum_{i=1}^h (z_i - \hat{z}_i)^2}{\sum_{i=1}^h (z_i - \bar{Z})^2}$

**BACIA HIDROGRÁFICA DO RIO POTIRIBU**

A bacia hidrográfica do rio Potiribu está localizada na região noroeste do estado Rio Grande do Sul (Figura 1). O rio Potiribu é um afluente da margem esquerda do rio Ijuí que, por sua vez, é afluente do rio Uruguai.



**Figura 1 - Localização da Bacia Hidrográfica do Rio Potiribu.**

Essa bacia possui atividade econômica essencialmente agrícola e tem sido monitorada desde 1989 por um grupo de pesquisadores do Instituto de Pesquisas Hidráulicas (IPH) da Universidade Federal do Rio Grande do Sul (UFRGS) por meio do Projeto Potiribu (Agra et al, 2000). Esse monitoramento é devido à sua representatividade da região central de um derrame basáltico, com características físico-climáticas homogêneas (Bordas et al., 1990).

A série de vazões médias mensais utilizada neste estudo foi obtida no Sistema de Informações Hidrológica (HidroWeb) mantido pela Agência Nacional de Águas (ANA). A estação escolhida foi a Ponte Nova do Potiribu – Jusante localizada no município de Ijuí na coordenada geográfica 28°22'37,99"S e 53°52'45.12"W, com uma área de drenagem de 629 km<sup>2</sup>.

### ANÁLISE DESCRITIVA DOS DADOS

Em todas as análises desenvolvidas neste trabalho foi utilizado o software livre R (R Development Core Team, 2010).

Para um melhor entendimento da variável vazão estudada são apresentadas, na Tabela 2, algumas medidas descritivas. A análise apresentada corresponde à série de médias mensais de vazões do período de julho de 2002 a setembro de 2009 do rio Potiribu.

Tabela 2 - Medidas descritivas da variável vazão em estudo.

Medida descritiva	Valor
Média (m <sup>3</sup> /s)	17,8
Mediana (m <sup>3</sup> /s)	12,5
Desvio padrão (m <sup>3</sup> /s)	14,5
Coefficiente de variação (%)	81,4
Valor máximo (m <sup>3</sup> /s)	64,6
Valor mínimo (m <sup>3</sup> /s)	2,5
Assimetria	1,5

A média das vazões médias mensais do período foi de 17,8 m<sup>3</sup>/s, com desvio padrão de 14,5 m<sup>3</sup>/s. Os valores altos do desvio padrão e do coeficiente de variação indicam uma média pouco representativa. Essa pouca representatividade pode ser explicada pela grande variabilidade intra-anual, caracterizada como sazonalidade.

A vazão média mensal máxima foi registrada em outubro de 2002, com 64,6 m<sup>3</sup>/s, enquanto que a mínima foi de 2,5 m<sup>3</sup>/s em março de 2005. A distribuição da série de vazões médias mensais possui assimetria positiva igual a 1,5.

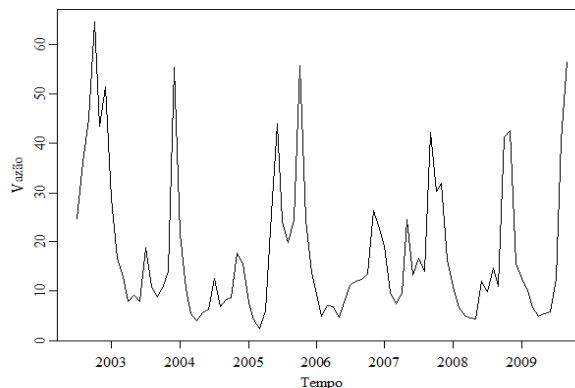


Figura 2 - Gráfico da série de vazões médias mensais de julho de 2002 a setembro de 2009.

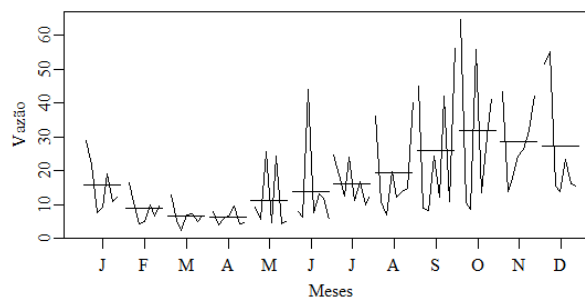


Figura 3 - Análise gráfica da sazonalidade da série de vazões médias mensais.

O gráfico de linha com a série temporal mensal de vazões médias é apresentada na Figura 2. Como pode ser observada, a série de vazões possui um padrão de variabilidade intra-anual, com períodos de cheias seguidos por períodos de estiagem, caracterizando a presença de sazonalidade.

Para identificar componentes da série como sazonalidade e tendência são apresentados os gráficos da sazonalidade, na Figura 3, e da decomposição da série, na Figura 4.

Percebe-se na Figura 3 que a série de vazões possui um período de cheias, de agosto a janeiro, e um período de seca, de fevereiro a julho, demonstrando a existência de sazonalidade na série em estudo. Já na Figura 4, nota-se que a série possui sazonalidade e uma ausência de tendência da mes-

ma, verificando que além do comportamento intra-anual não há evidências de acréscimo ou decréscimo sistemático da variável ao longo do tempo.

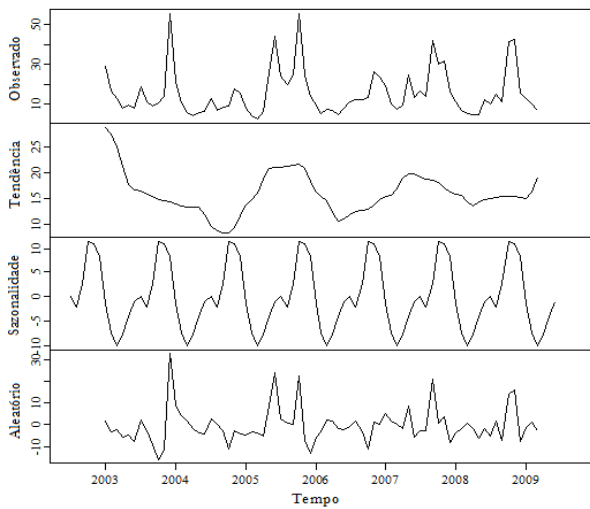


Figura 4 - Decomposição da série temporal em componentes de sazonalidade, de tendência e de aleatoriedade.

## MODELAGEM E PREVISÃO

### Identificação e estimação do modelo

Em uma investigação inicial dos dados são verificadas as funções de auto-correlação amostral (FAC) e de auto-correlação parcial amostral (FACP) da série em nível, Figura 5, da série diferenciada de ordem 1, Figura 6, e da série diferenciada de ordem 12, Figura 7.

Analisando-se os correlogramas apresentados na Figura 5 pode-se perceber a auto-correlação significativamente diferente de zero nas defasagens 1 e 2, ou seja, os valores da FAC amostral estão fora do intervalo de confiança, delimitado pelas linhas tracejadas, evidenciando que existe correlação serial a ser modelada. Além disso, percebe-se que a FAC amostral é infinita em extensão e tem comportamento senoidal, indicando um possível processo do tipo autorregressivo, não estacionário e sazonal.

Analisando-se a Figura 6 pode-se observar que a série com uma defasagem (*lag* 1) não é auto-correlacionada, indicando que o processo gerador da série não é integrado de ordem um. Na Figura 7, em que são apresentados os correlogramas da série com uma diferença de ordem doze, ou seja, com uma diferença sazonal, percebe-se características de um modelo estacionário, autorregressivo de ordem

4, pela análise da FAC, e de médias móveis de ordem 1, pela análise da FACP. Essa análise dos correlogramas indica que o modelo a ser escolhido deve considerar diferenças na parte sazonal. Já para a parte ordinal o modelo não deve possuir diferenças e deve-se considerar um modelo autorregressivo.

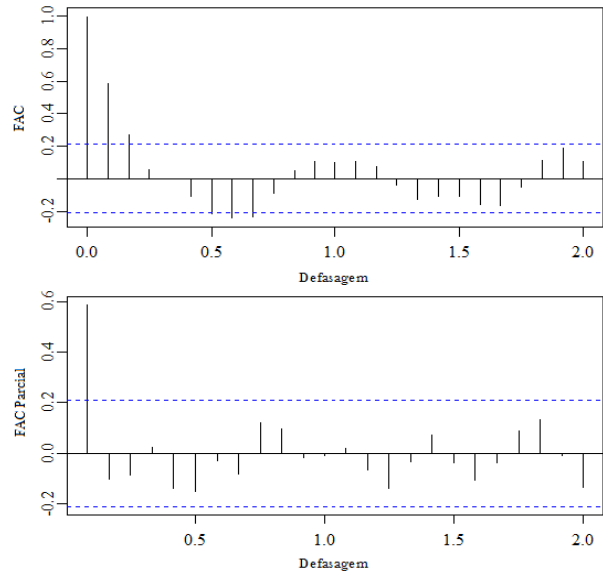


Figura 5 - Função de auto-correlação amostral (FAC) e função de auto-correlação parcial amostral (FACP) da série.

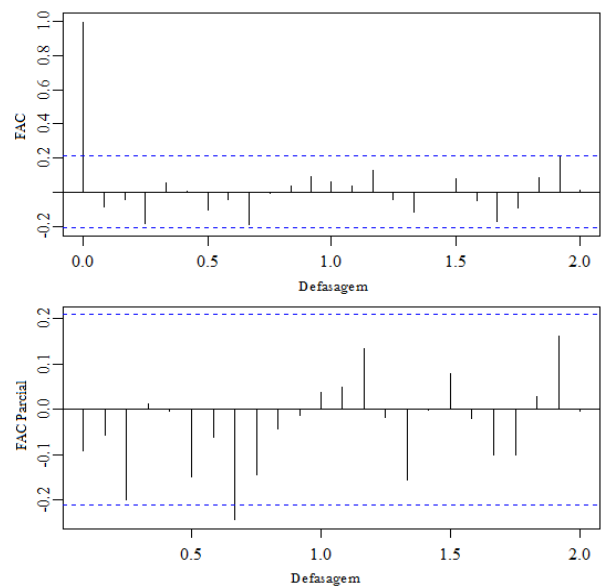


Figura 6 - Função de auto-correlação amostral (FAC) e função de auto-correlação parcial amostral (FACP) da série diferenciada de ordem 1.

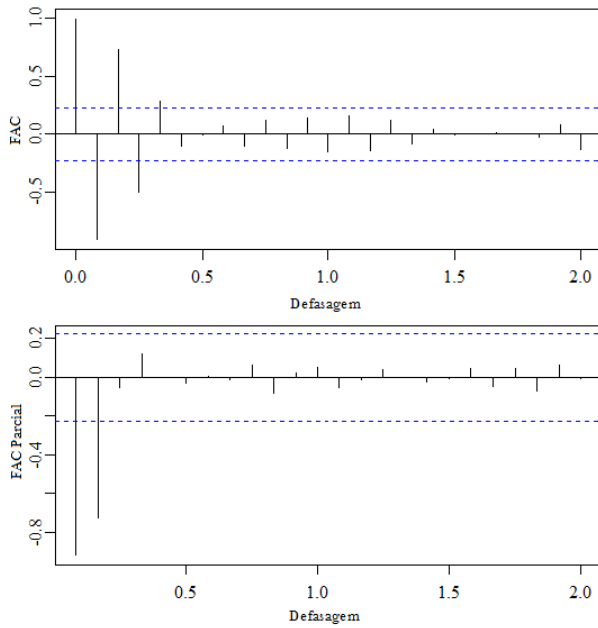


Figura 7 - Função de auto-correlação amostral (FAC) e função de auto-correlação parcial amostral (FACP) da série diferenciada de ordem 12.

Sendo assim, conclui-se que a vazão do Rio Potiribu possui uma sazonalidade bem definida, apresentando-se claramente nos correlogramas, em que a mesma pode ser descrita em grande parte pelas vazões dos mesmos meses dos anos anteriores. Além disso, a auto-correlação da série evidencia que a vazão do período atual também pode ser descrita pela vazão dos últimos períodos.

Contudo, para determinar o modelo mais adequado lançamos mão do critério de seleção de modelos AIC, seguindo a metodologia apresentada a seguir. Inicialmente testaram-se todos os possíveis modelos tais que: a ordem autorregressiva ( $p$  e  $P$ ) e de médias móveis ( $q$  e  $Q$ ) fossem inferiores a 3 e as diferenças ( $d$  e  $D$ ) menores que 2. Esses valores foram escolhidos pelas indicações dos correlogramas, apresentados nas Figuras 5, 6 e 7, e por se procurar um modelo que equilibre a qualidade do ajuste com a sua complexidade. Feito isso, foram selecionados seis modelos concorrentes que apresentaram os menores valores para o AIC e, para esses, calculadas as medidas de qualidade do erro quadrático médio (EQM), do erro agregado percentual (EAP) e do coeficiente de Nash-Sutcliffe (CNS). Os modelos concorrentes com seus valores de AIC e das medidas de qualidade são apresentados na Tabela 3.

Tabela 3 - Modelos concorrentes e seus valores de AIC e das medidas de qualidade.

SARIMA (p,d,q) (P,D,Q) <sub>s</sub>	Critério AIC	Medidas		
		EQM	EAP	CNS
(3,0,0) (2,1,1) <sub>12</sub>	596	101,1	-0,03	0,52
(3,0,0) (2,1,2) <sub>12</sub>	595	<b>66,4</b>	<b>-0,02</b>	<b>0,68</b>
(2,0,0) (2,1,0) <sub>12</sub>	593	104,5	-0,04	0,50
(1,1,1) (2,1,2) <sub>12</sub>	<b>590</b>	68,2	0,07	0,67
(1,1,1) (0,0,3) <sub>12</sub>	664	75,0	-0,03	0,64
(3,1,1) (0,0,3) <sub>12</sub>	675	83,8	<b>0,02</b>	0,60

Todos os modelos selecionados possuem valores de AIC muito próximos, com número de parâmetros a serem estimados variando de 5 a 8, sendo que o menor valor foi obtido no modelo SARIMA(1,1,1) (2,1,2)<sub>12</sub>. No entanto, o modelo que apresentou um bom valor de AIC, os melhores valores para as medidas de qualidade e que está de acordo com a análise dos correlogramas foi o modelo SARIMA(3,0,0) (2,1,2)<sub>12</sub>. Sendo assim, esse torna-se o modelo de interesse.

#### Diagnóstico do modelo

Após ter sido determinado o modelo candidato e estimados os parâmetros foi realizada a análise de diagnóstico, com o objetivo de validar o modelo para traçar previsões. Foram traçados os gráficos, apresentados na Figura 8, dos resíduos padronizados, ou seja, resíduos com uma distribuição próxima da normal padrão, o correlograma dos resíduos do modelo e uma interpretação gráfica do teste de Ljung-Box.

Analisando-se a Figura 8 percebe-se que todos os resíduos padronizados encontram-se entre os valores de -3 e 3. Em relação ao correlograma dos resíduos do modelo, podemos observar que o mesmo não apresenta nenhuma auto-correlação significativamente diferente de zero em nenhuma defasagem, indicando o bom ajuste do modelo e que o mesmo conseguiu captar a auto-correlação serial. Já em relação aos  $p$ -valores do teste de Ljung-Box apresentados graficamente para  $m$  diferentes defasagens, todos os valores estão acima de 5%, indicando que não se rejeita a hipótese nula de que as  $m$  primeiras auto-correlações dos erros são iguais a zero. Desta maneira, conclui-se que o modelo descreve adequadamente os dados, validando o seu uso para traçar previsões.

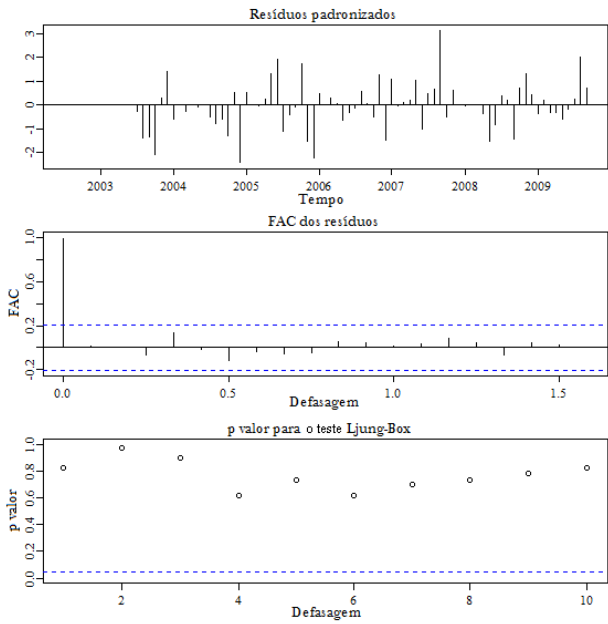


Figura 8 - Análise de diagnóstico do modelo SARIMA.

**Modelo selecionado**

As estimativas de máxima verossimilhança dos parâmetros do modelo selecionado, SARI-MA(3,0,0)(2,1,2), estão apresentadas na Tabela 4. O modelo escolhido, com  $p = 3$ , indica que a variável vazão  $Q_t$  depende dos valores de  $Q_{t-1}, Q_{t-2}$  e  $Q_{t-3}$ . Os parâmetros autorregressivos estimados ( $\hat{\phi}_1, \hat{\phi}_2$  e  $\hat{\phi}_3$ ) mostram que  $Q_t$  possui uma relação positiva com o primeiro e o terceiro mês e negativa com o segundo mês e, é mais fortemente descrito por  $Q_{t-1}$ . Em relação à parte sazonal temos a ordem  $P = 2$  ( $\hat{\Phi}_1$  e  $\hat{\Phi}_2$ ), ou seja,  $Q_t$  é função de  $Q_{t-12}$  e de  $Q_{t-24}$ , pois a sazonalidade é de 12 meses.

Tabela 4 - Estimativas dos parâmetros do modelo selecionado.

Modelo	Parâmetros estimados
SARIMA(3,0,0)(2,1,2)	$\hat{\phi}_1 = 0,6461$
	$\hat{\phi}_2 = -0,0399$
	$\hat{\phi}_3 = 0,0286$
	$\hat{\Phi}_1 = -0,4452$
	$\hat{\Phi}_2 = -0,7185$
	$\hat{\Theta}_1 = -0,3443$
	$\hat{\Theta}_2 = 0,9978$

Os parâmetros dos polinômios de médias móveis só aparecem na parte sazonal do modelo, com  $Q = 2$  ( $\hat{\Theta}_1$  e  $\hat{\Theta}_2$ ). O mesmo ocorre em relação às diferenças, pois temos  $d = 0$  e  $D = 1$  indicando que é tomada apenas uma diferença sazonal na série.

A série ajustada com o modelo estimado é apresentada na Figura 9. As medidas de qualidade são as mesmas da Tabela 2 para o referido modelo.

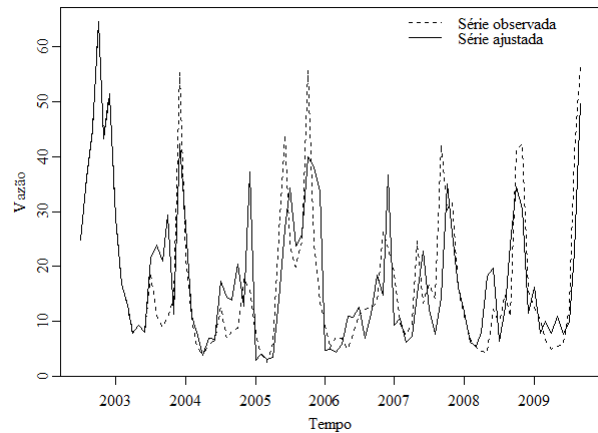


Figura 9 - Valores observados da vazão média mensal e valores da vazão média mensal ajustados (simulados) pelo modelo SARIMA (3,0,0)(2,1,2)<sub>12</sub>.

**Previsão de vazões**

Nesta etapa do trabalho foram traçadas previsões de vazões médias mensais para um horizonte de 6 meses à frente, ou seja, são meses que não foram considerados no ajuste do modelo.

No gráfico da Figura 10 são apresentados os valores previstos e os limites de previsão com 95% e 80% de confiança. Os valores obtidos na previsão podem ser vistos numericamente na Tabela 5, assim como os valores reais para o período.

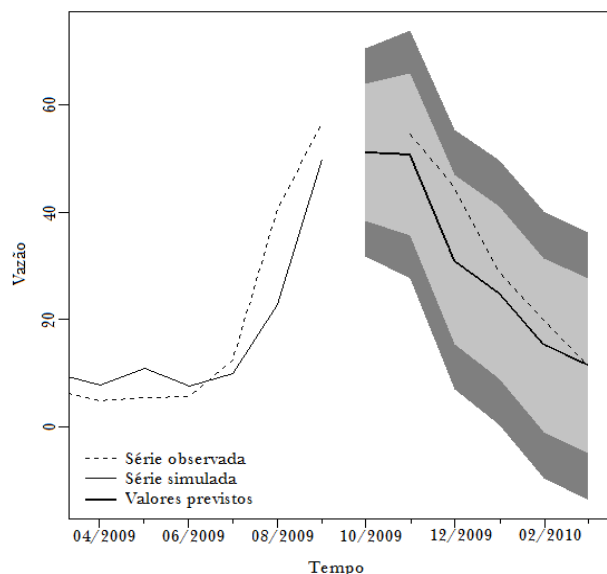
Como pode ser observado na Tabela 5 existe uma falha nos dados observados no mês de outubro de 2009, o que justifica o ajuste do modelo ter sido realizado até setembro de 2009 e previsão das vazões para os seis meses seguintes.

As medidas de qualidade obtidas para o período previsto são de EAP = 0,16, EQM = 46,63 e CNS = 0,81.

Os valores de CNS obtidos tanto no ajuste quanto na previsão podem ser considerados bons, uma vez que trabalhos utilizando modelos determinísticos, em que variáveis explicativas são consideradas, obtiveram valores próximos. Isso pode ser visto



em Gottschalk et al (2001), que utilizando um modelo físico distribuído, com variáveis explicativas como dados de clima, de solo e de vegetação, encontraram valores do CNS variando de 0,60 a 0,87 em nove bacias estudadas. Nóbrega et al (2008) obtiveram um valor para o CNS de 0,82 ao simular a vazão na bacia do rio Jaguari para dados diários em um modelo semidistribuído, com variáveis de entrada como temperatura do ar, umidade relativa, temperatura do ponto de orvalho, dados de cobertura do solo e dados fluviométricos.



**Figura 10 - Valores reais e ajustados da série e previsão das vazões médias mensais para um período de 6 meses à frente.**

**Tabela 5 - Previsão de vazão (m³/s) para os próximos 8 meses.**

Período	Valores Reais	Valores Previstos
Out/2009	-	51,2
Nov/2009	54,5	50,8
Dez/2009	44,7	31,1
Jan/2010	29,0	25,1
Fev/2010	19,8	15,3
Mar/2010	11,3	11,4

## CONCLUSÃO

Os modelos estocásticos são historicamente utilizados de maneira satisfatória para descrever o comportamento de variáveis hidrológicas. Contudo,

atualmente eles têm sido pouco explorados para tal fim, em especial pelos pesquisadores brasileiros. Neste trabalho foi realizado o ajuste de um modelo estocástico da classe ARIMA para uma série de dados de vazões mensais do rio Potiribu.

O modelo escolhido foi o SARIMA(3,0,0)(2,1,2)<sub>12</sub>, que é um modelo da classe ARIMA que incorpora a sazonalidade da série, característica temporal importante. Obteve-se um bom ajuste do modelo, com coeficiente de Nash-Sutcliffe de 0,68 no período de estimação, valor esse que pode ser considerado muito bom, uma vez que a série não possui nenhuma outra variável explicativa, sendo modelada somente pela sua correlação serial, levando em consideração seus componentes sazonais, autorregressivos e de médias móveis. Esses modelos podem ser significativamente melhorados ao considerar variáveis explicativas como precipitação, evapotranspiração, entre outras, estendendo o modelo SARIMA para os modelos que incorporam covariáveis, conhecidos como modelos SARIMAX.

Silva (2005) ressalta que modelos estocásticos, que consideram correlações temporais e sazonais, produzem bons resultados quando o escoamento na bacia se dá de forma lenta, onde predomina o armazenamento subterrâneo ou a declividade seja baixa. Essa não é uma característica da bacia em estudo, indicando que a metodologia utilizada teria resultados ainda melhores se aplicada a rios pertencentes a grandes bacias.

De acordo com a fase de diagnóstico, o modelo ajustou-se bem aos dados, o que qualificou o mesmo para realizar previsões. As previsões foram feitas em um horizonte de 6 meses. Os valores previstos acompanharam o comportamento da série e obtiveram medidas de qualidade excelentes, como o coeficiente de Nash-Sutcliffe de 0,81.

O modelo escolhido foi capaz de incorporar características importantes da série, como sazonalidades e correlação serial. O ajuste apresentou um coeficiente de Nash-Sutcliffe de 0,68 no período de calibração e uma precisão ainda maior ao realizar previsões passadas a frente, demonstrado pelo valor de Nash-Sutcliffe de 0,81 no período previsto.

## REFERÊNCIAS

AGRA, S. G. ; SOUZA, V. C. B. ; CASTRO, N. M. R. ; SILVEIRA, A. L. L. (2000) . Medição de vazão e descarga sólida na bacia do rio Potiribu: Arroio Turcato, Pejuçara-RS.. In *Anais do IV Encontro Nacional de Engenharia de Sedimentos*, 2000, Santa Maria-RS. IV

- Encontro de Engenharia de Sedimentos. Porto Alegre : Associação Brasileira de Recursos Hídricos - ABRH, 2000. v. 01.
- AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proc. 2nd Inter. Symposium on Information Theory*, pp. 267-281, Budapest.
- BAYER, F.M.; SOUZA, A.M.. (2010). Wavelets e modelos tradicionais de previsão: um estudo comparativo. *Revista Brasileira de Biometria*. v.28, n. 2, pp. 40-61.
- BIRINCI, V.; AKAY, O. (2010). A Study on Modeling Daily Mean Flow with MLR, ARIMA and RBFNN. In *Anais do BALWOIS: Water observation and information system for decision support*. Ohrid, República da Macedônia. 25-29 de maio de 2010.
- BORDAS, M.P.; LANNA, A.E.L.; DE LUCA, S.J. (1990). *Projeto Potiribu: avaliação e controle dos impactos da agricultura sobre os recursos hídricos do derrame balsático sul-americano*. Porto Alegre: UFRGS-IPH. 2p.
- BOX, G.E.P.; JENKINS, G. M.; REINSEL, G.C. (2008) *Time Series Analysis: Forecasting and Control*, 4ª ed. Editora Wiley. 784 p.
- 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*. v.10, n.4. pp. 61-72.
- FREITAS, R.A.P.; DOUGLAS LINDEMANN, D.; SOUZA, L.S.; FARIA, H.; SANTOSM.; ELESBON, A.; CASAGRANDE, F. (2010). Análise de Séries Temporais de Vazão e Precipitação na Bacia Hidrográfica do Rio São Mateus. In *Anais do XVI Congresso Brasileiro de Meteorologia*, Belém, PA, 2010.
- GEMITZI, A.; STEFANOPOULOS, K. (2011). Evaluation of the effects of climate and man intervention on ground waters and their dependent ecosystems using time series analysis. *Journal of Hydrology*. v. 403. pp. 130-140.
- GOTTSCHALK, L.; BELDRING,S.; ENGELAND, K.; TALLAKSEN, L.; SAELTHUN, N.R.; KOLBERG, S.; MOTOVILOV, Y. (2001). Regional/macroscale hydrological modelling: a Scandinavian experience. *Hydrological Sciences-Journal-des Sciences Hydrologiques*. v.46, n.6. pp. 963-982.
- LEDOLTER, J..(1976). ARIMA Models and their Use in Modeling Hydrologic Sequences. *IASA Research Memorandum RM-76-069*. 45 p.
- LEDOLTER, J..(1977). The Analysis of Multivariate Time Series with a View to Applications in Hydrology. *IASA Research Memorandum RM-77-011*. 33 p.
- LJUNG, G.M.; BOX, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*. v.65, n.2, pp. 297-303.
- LUCAS, E.W.M.; SOUSA, F.A.S.; SILVA, F.D.S; LUCIO, P.S. (2009). Modelagem hidrológica determinística e estocástica aplicada à região hidrográfica do xingu – Pará. *Revista Brasileira de Meteorologia*, v.24, n.3, 308-322.
- MODARRES, R.; ESLAMIAN, S. S. (2006). Streamflow time series modeling of Zayandehrud river. *Iranian Journal of Science & Technology*, Transaction B, Engineering, Vol. 30, n.. B4. Printed in The Islamic Republic of Iran.
- MORETTIN, P.A.; TOLOI, C.M.C. (2006). *Análise de séries temporais*. 2ª ed. São Paulo: Egard Blucher.
- NOBREGA, R.S.; SOUZA, E.P.; SOUSA, F.A.S. (2008). Análise da utilização de dados do satélite TRMM em um modelo hidrológico semidistribuído na bacia do rio Jamari (RO). *Revista Brasileira de Geografia Física*. v. 01, n. 01, Recife-PE., pp.47-61
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- SALAS, J.D.; DELLEUR, J.W.; YEVJEVICH, V.; LANE, W.L.. (1980). *Applied Modeling of hydrologic time series*. Water Resources publications, Littleton, Colorado.484p.
- SALAS, J.D. (1992) Analysis and modeling of hydrologic time series. *Handbook of Hydrology*. Maidment, D.R. (editor).McGraw-Hill. Capítulo 19. 72p.
- SILVA, B.C. (2005). *Previsão hidroclimática de vazão para a bacia do rio São Francisco*. Tese submetida ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul. Instituto de Pesquisas Hidráulicas. Porto Alegre. 272 p.
- TUCCI, C.E.M. (1998). *Modelos hidrológicos*. ABRH Editora da UFRGS. Porto Alegre. 669 p.

### **Modeling and Forecasting Mean Monthly Streamflows Using Time Series Models**

#### **ABSTRACT**

*This paper considers analyzing and modeling a time series of mean monthly streamflows using the stochastic model ARIMA (autoregressive integrated moving average). This analysis aims to model and forecast the monthly streamflow in the Potiribu river basin, located in the northwest region of Rio Grande do Sul state, Brazil. We tested several models of the SARIMA type, or ARIMA models which take the seasonal characteristics of the data into account. Among all tested models we selected some competing ones that had the lowest Akaike information criteria*

(AIC). Finally, we chose the SARIMA model  $(3,0,0)$   $(2,1,2)_{12}$  which presented the best fit to the series of flows observed, with the Nash-Sutcliffe coefficient (NSC) equal to 0.68, as training data. The model was used to make forecasts for the six months forward, resulting in NSC = 0.81 as validating data.

**Key-words:** ARIMA, stochastic models, hydrological models, streamflow.