

Serviço de Pós-Graduação EESC/USP

EXEMPLAR REVISADO

Data de entrada no Serviço: 26.1.05.03

Ass.: *[Assinatura]*

**USO DA ABORDAGEM BAYESIANA
PARA A ESTIMATIVA DE PARÂMETROS
SAZONAIS DOS MODELOS
AUTO-REGRESSIVOS PERIÓDICOS**



Maria Helena Rodrigues Gomes

Tese apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Doutor em Hidráulica e Saneamento.

ORIENTADOR: Prof. Dr. Fazal Hussain Chaudrhy

DEDALUS - Acervo - EESC



31100043554

São Carlos

2003

Class.	TESF-EESC
Cutt.	2503
Tombo	T 211/03
Sysno	132.6305

Ficha catalográfica preparada pela Seção de Tratamento da Informação do
Serviço de Biblioteca - EESC/USP

G633u Gomes, Maria Helena Rodrigues
Uso da abordagem bayesiana para a estimativa de parâmetros sazonais dos modelos auto-regressivos periódicos / Maria Helena Rodrigues Gomes. -- São Carlos, 2003.

Tese (Doutorado) -- Escola de Engenharia de São Carlos-Universidade de São Paulo, 2003.
Área : Hidráulica e Saneamento.
Orientador: Prof. Dr. Fazal Hussain Chaudhry.

1. Modelo PAR. 2. Séries temporais hidrológicas. 3. Abordagem bayesiana. I. Título.

FOLHA DE JULGAMENTO

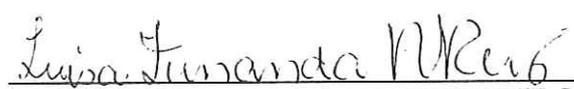
Candidata: Engenheira **MARIA HELENA RODRIGUES GOMES**

Tese defendida e julgada em 19-03-2003 perante a Comissão Julgadora:



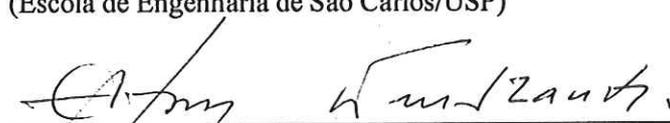
Prof. Tit. **FAZAL HUSSAIN CHAUDHRY (Orientador)**
(Escola de Engenharia de São Carlos/USP)

Aprovada



Profa. Dra. **LUISA FERNANDA RIBEIRO REIS**
(Escola de Engenharia de São Carlos/USP)

Aprovada



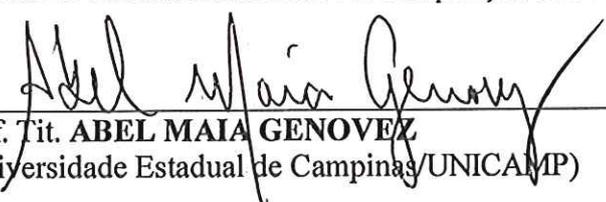
Prof. Dr. **EDSON CEZAR WENDLAND**
(Escola de Engenharia de São Carlos/USP)

Aprovada



Prof. Dr. **MARINHO GOMES DE ANDRADE FILHO**
(Instituto de Ciências Matemáticas e de Computação/USP/São Carlos)

Aprovada



Prof. Tit. **ABEL MAIA GENOVEZ**
(Universidade Estadual de Campinas/UNICAMP)

Aprovada



Profa. Associada **MARIA DO CARMO CALIJURI**
Coordenadora do Programa de Pós-Graduação em
Engenharia (Hidráulica e Saneamento) e
Presidente da Comissão de Pós-Graduação



A água é o sangue da terra.

Insubstituível.

*Nada é mais suave e, no entanto,
nada a ela resiste:*

*Aquele que conhece seus princípios
Pode agir corretamente,
tomando-a como chave e exemplo.*

*Quando a água é pura,
o coração do povo é forte.*

*Quando a água é suficiente,
o coração do povo é tranquilo.*

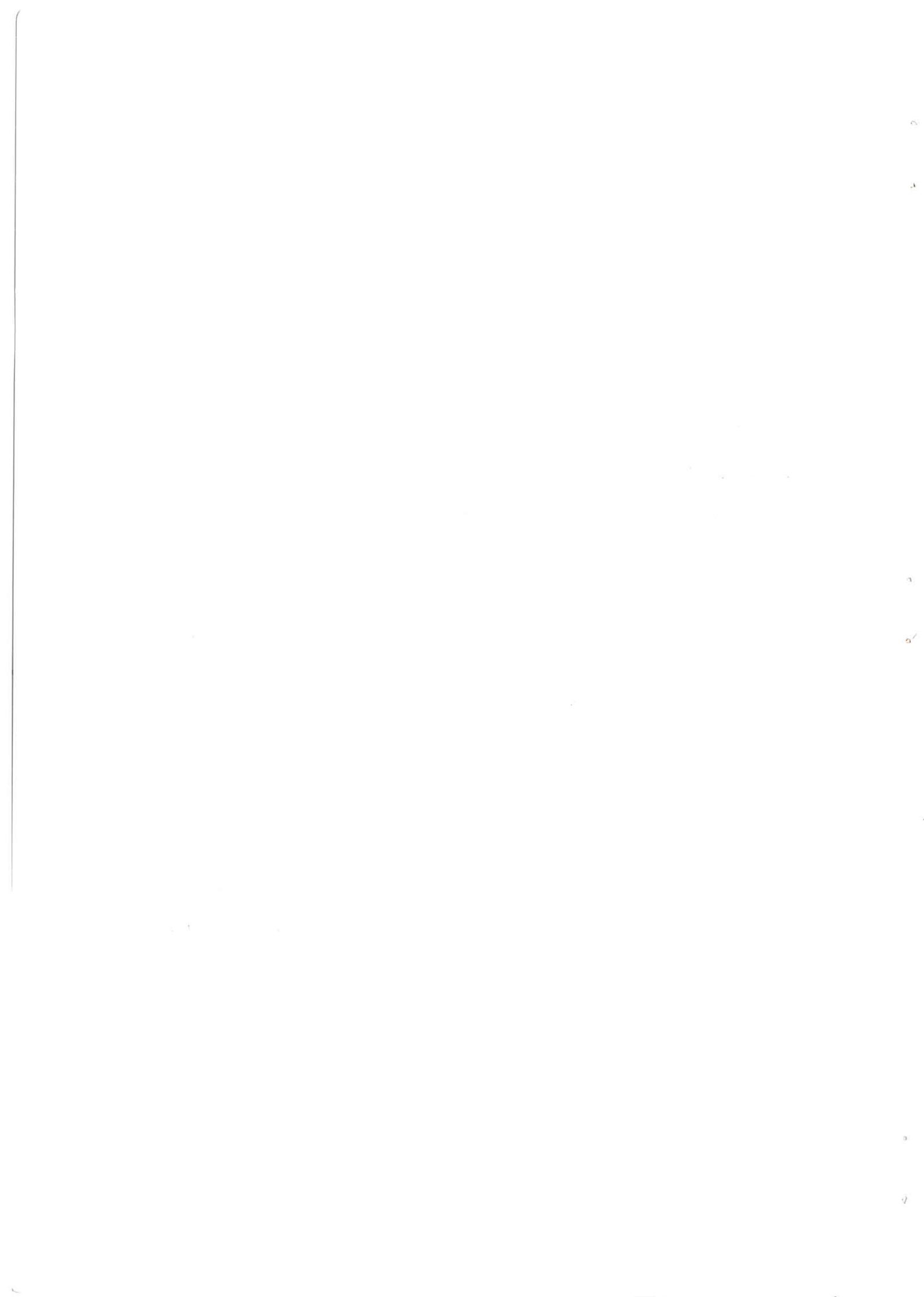
*Filósofo Chinês
Século IV – A.C.*



A Deus,

*Pai de infinita bondade e misericórdia,
que sempre esteve presente em todos os
momentos da minha vida, dando-me
forças para superar os obstáculos e
seguir a diante com fé e esperança.*

*Aos meus pais, José e Helena,
e minha irmã Mônica, por todo
amor, compreensão e apoio.*



AGRADECIMENTOS

À Escola de Engenharia de São Carlos e ao Curso de Pós-Graduação em Hidráulica e Saneamento.

Ao Professor Fazal Hussain Chaudrhy, pela orientação, amizade e atenção dispensadas durante a realização desta pesquisa.

Ao Professor Marinho Gomes de Andrade Filho, pela colaboração na realização deste trabalho.

Ao Conselho Nacional de Pesquisa (CNPq) pela bolsa concedida.

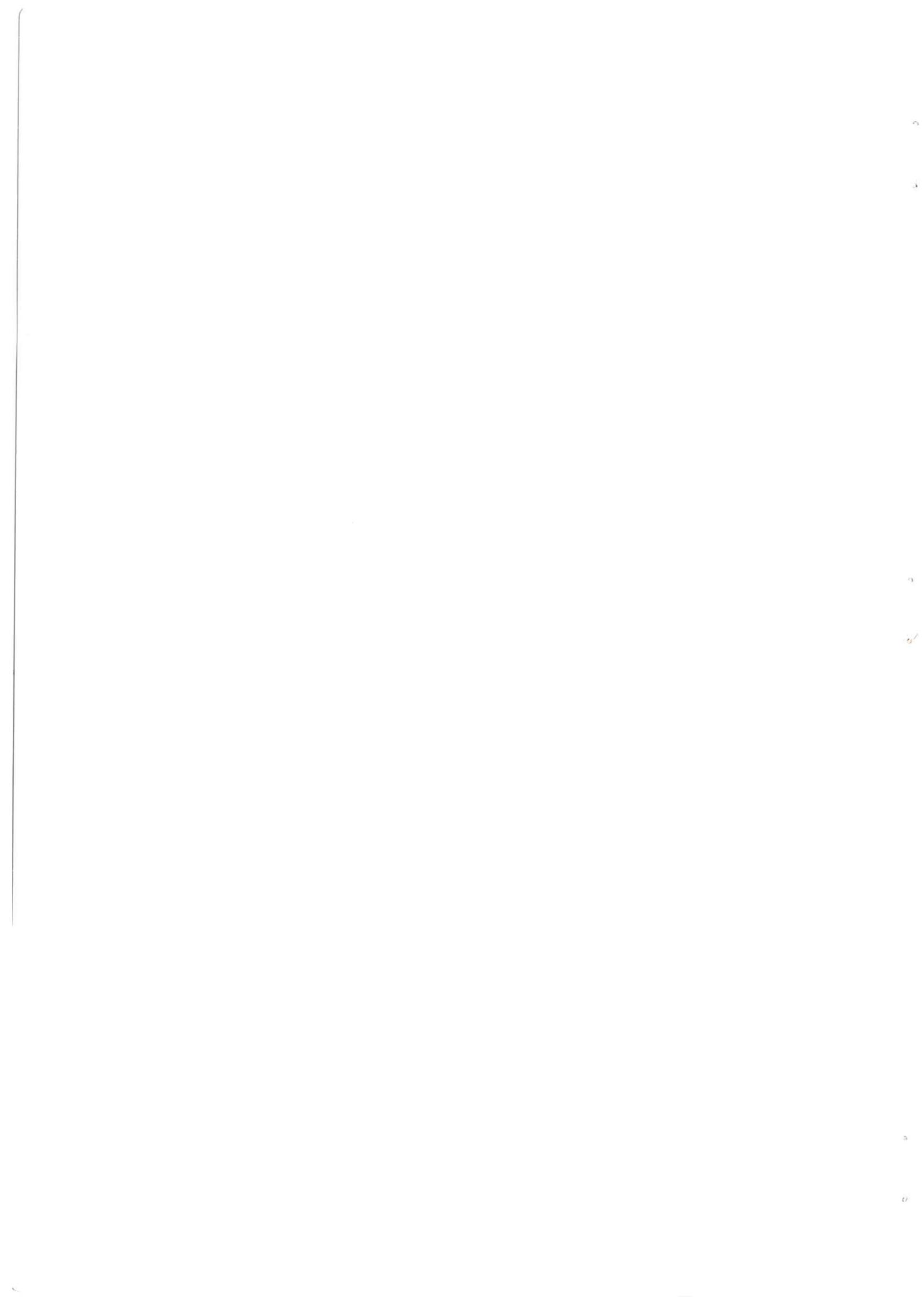
Ao Grupo de Oração Universitário Sal e Luz pelas orações e pelos momentos de fé e alegria.

Às amigas Isaura Sombra e Luciane Sarkis pela amizade e pelos momentos de alegria e apoio dispensados durante todos esses anos.

Aos amigos Victor Matos, Mario Massayuki, Eneide Baffa, Elvira Majanelli, Zezo, Lena, Sônia Luiz, Cynthia Lotti, Jefferson Pereira pelos momentos de alegre convivência.

À Eleninha, pelas correções realizadas na finalização deste trabalho.

A todos os colegas, funcionários e professores do Departamento de Hidráulica e Saneamento que contribuíram para a realização deste trabalho, meus sinceros agradecimento.



SUMÁRIO

LISTA DE FIGURAS.....	ix
LISTA DE TABELAS.....	xi
LISTA DE ABREVIATURAS E SIGLAS.....	xiii
LISTA DE SÍMBOLOS.....	xv
RESUMO.....	xix
<i>ABSTRACT</i>	xxi
1 – INTRODUÇÃO.....	1
2 – OBJETIVOS.....	5
2.1 – Objetivos específicos.....	5
3 - REVISÃO DA LITERATURA	
3.1 – Séries temporais.....	7
3.1.1 – Conceito.....	7
3.1.2 – Finalidade da análise da séries temporais.....	8
3.1.3 – Classificação das séries temporais hidrológicas.....	8
3.1.4 – Séries temporais versus previsão.....	10
3.2 – Modelos para geração de séries temporais.....	10
3.2.1 – Simulação Monte Carlo.....	11
3.2.2 – Cadeias de Markov.....	12

3.2.2.1 – Simulação via Cadeias de Markov.....	13
3.2.3 – Modelos estocásticos.....	14
3.2.3.1 – Exemplos do uso dos modelos estocásticos.....	16
3.3 – Modelos estocásticos sazonal.....	20
3.3.1 – Modelo com componente sazonal determinístico.....	21
3.3.2 – Modelo com componente sazonal estocástico.....	23
3.3.3 – Modelo com parâmetros periódicos.....	23
3.3.4 – Alguns modelos estocásticos sazonais.....	26
3.4 – Métodos “clássicos” para determinação dos parâmetros do modelo.....	30
3.4.1 – Métodos dos momentos.....	30
3.4.2 – Cadeias dos mínimos quadrados.....	31
3.4.3 – Função de verossimilhança.....	32
3.4.3.1 – Exemplos do uso do método de máxima verossimilhança na determinação dos parâmetros do modelo.....	34
3.5 – Inferência bayesiana.....	37
3.5.1 – Natureza da inferência bayesiana.....	37
3.5.2 – Análise bayesiana e a hidrologia.....	39
3.5.3 – Operação bayesiana.....	40
3.5.4 – Teorema de Bayes.....	41
3.5.5 – Seleção do modelo bayesiano.....	42
3.5.6 – Método bayesiano e técnicas de simulação.....	42
3.5.6.1 – Computação bayesiana.....	43
3.5.7 – Exemplo do uso do método bayesiano em hidrologia.....	43
3.6 – Métodos para seleção do modelo.....	49
3.6.1 – Determinação da função de autocorrelação parcial periódica – PeFACP.....	49
3.6.2 – Critério do erro de predição finito – FPE.....	51
3.6.3 – Critério de informação de Akaike – AIC.....	52
3.6.4 – Critério de Hanna-Quinn –HQ.....	52

3.6.5 – Critério de Schwarz – BIC.....	53
3.6.6 – Critério de Shibata.....	53
3.6.7 – Densidade preditiva ordenada.....	54
3.6.8 – Fator de Bayes.....	54
4 – METODOLOGIA.....	
4.1 – Modelo periódico auto-regressivo –PAR.....	57
4.1.1 – Apresentação do modelo.....	57
4.1.2 – Identificação do modelo do modelo PAR.....	57
4.1.2.1 – Determinação da função de auto-correlação periódica – PeFAC.....	58
4.2 – Seleção do modelo.....	58
4.2.1 – Critérios de informação.....	58
4.3 – Estimativa dos parâmetros do modelo.....	60
4.3.1 – Método clássico – Estimadores de máxima verossimilhança.....	60
4.3.1.1 – Determinação da função de verossimilhança.....	60
4.3.1.2 – Estimadores de máxima verossimilhança – EMV.....	63
4.3.2 – Método bayesiano.....	63
4.3.2.1 – <i>Priori</i> não informativa de Jeffreys.....	64
4.3.2.1.1 – Densidade <i>a posteriori</i> conjunta para os parâmetros Φ_m e τ_m	64
4.3.2.2 – Densidade <i>a priori</i> informativa - <i>priori</i> Normal-Gama.....	67
4.3.2.2.1 – Densidade <i>a posteriori</i>	68
4.3.2.2.2 – Determinação de P_m e μ_m	72
4.4 – Aplicação.....	74
4.4.1 – Escolha das séries hidrológicas.....	75
4.5 – Cálculo dos erros.....	76

4.6 – Desenvolvimento dos programas computacionais.....	76
5 – RESULTADOS E DISCUSSÃO.....	
5.1 – Modelos periódicos auto-regressivos – PAR.....	77
5.1.1 – Apresentação das séries hidrológicas.....	77
5.1.2 – Identificação do modelo PAR.....	79
5.2 – Seleção da ordem do modelo.....	82
5.3 – Estimativa dos parâmetros do modelo.....	86
5.3.1 – Método clássico – Estimadores de máxima verossimilhança (EMV)	86
5.4 – Método bayesiano.....	88
5.4.1 – Determinação dos parâmetros usando <i>priori</i> não-informativa de Jeffreys.....	88
5.4.2 – Determinação dos parâmetros usando <i>priori</i> informativa Normal–Gama.....	89
5.4.2.1 – Determinação de P_m e μ_m	89
5.4.2.2 – Determinação dos estimadores bayesianos exatos – EBEx...	91
5.5 – Previsão.....	92
5.5.1 – Análise dos resíduos.....	99
6 - CONCLUSÕES.....	103
ANEXO A.....	105
ANEXO B.....	107
7 - REFERÊNCIAS BIBLIOGRÁFICAS.....	109
8 - OBRAS CONSULTADAS.....	115

LISTA DE FIGURAS

FIGURA 1.1 – Incerteza das vazões futuras (POLTRONIERE SILVA, 1999).....	2
FIGURA 3.1 – Classificação geral das séries temporais segundo YEVJEVICH (1972).	9
FIGURA 3.2 – Esquema do filtro linear agindo no ruído para criar uma série temporal segundo BOX & Jenkins (1994).....	15
FIGURA 3.3 – Procedimento iterativo para construção do modelo segundo BOX & JENKINS (1994).....	15
FIGURA 3.4 – Série de vazões de Furnas de maio /1931 a abril/1990.....	20
FIGURA 3.5 – Diagrama ilustrativo da operação bayesiana PEREIRA & VIANA (1982).....	41
FIGURA 4.1 – Parte do Sistema Sudeste Brasileiro (POLTRONIERE SILVA, 1990).....	75
FIGURA 5.1 – Série hidrológica de vazões médias mensais para a Usina Hidrelétrica de Furnas no período de janeiro de 1931 a dezembro de 1990.....	78
FIGURA 5.2 – Série hidrológica de vazões médias mensais para a Usina Hidrelétrica de Emborcação no período de janeiro de 1931 a dezembro de 1990.....	78
FIGURA 5.3 – Gráficos de dispersão para a série de vazões de Furnas.....	81
FIGURA 5.4 – Gráficos de dispersão para a série de vazões de Emborcação....	76
FIGURA 5.5 – Função de autocorrelação para as séries de Furnas e Emborcação.....	82
FIGURA 5.6 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1980 usando o modelo PAR, os EMV e os EBEx para a série hidrológica de Furnas.....	93

FIGURA 5.7 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1980 usando o modelo AR(2), os EMV e os EBEx para a série hidrológica de Furnas.....	93
FIGURA 5.8 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1990 usando os modelos PAR, os EMV e os EBEx para a série de vazões de Furnas.....	94
FIGURA 5.9 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1990 usando os modelos AR(2), os EMV e os EBEx para a série de vazões de Furnas.....	95
FIGURA 5.10 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1980 usando o modelo PAR, os EMV e os EBEx para a série hidrológica de Emborcação.....	96
FIGURA 5.11 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1980 usando o modelo AR(2), os EMV e os EBEx para a série hidrológica de Emborcação.....	96
FIGURA 5.12 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1990 usando o modelo PAR, os EMV e os EBEx para a série hidrológica de Emborcação.....	97
FIGURA 5.13 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1990 usando o modelo AR(2), os EMV e os EBEx para a série hidrológica de Emborcação.....	98
FIGURA 5.14 – Análise gráfica da probabilidade, da correlação e da correlação parcial dos resíduos para a série hidrológica de Furnas.....	99
FIGURA 5.15 – Análise gráfica da probabilidade, da correlação e da correlação parcial dos resíduos para a série hidrológica de Emborcação.....	100

LISTA DE TABELAS

TABELA 1.1 – Relação entre os horizontes de planejamento e dos graus de incerteza e a descrição de cada período.....	3
TABELA 3.1 – Matriz das probabilidades de transição para o processo de Markov dos períodos indicados.....	46
TABELA 3.2 – Tabela de referência do fator de Bayes.....	55
TABELA 5.1 – Valores do critério AIC para os meses da série hidrológica de Furnas.....	83
TABELA 5.2 – Valores do critério BIC para os meses da série hidrológica de Furnas.....	83
TABELA 5.3 – Valores do critério AIC para os meses da série hidrológica de Emborcação.....	84
TABELA 5.4 – Valores do critério BIC para os meses da série hidrológica de Emborcação.....	84
TABELA 5.5 – Comparação entre os valores de AIC e BIC e para a série hidrológica de Furnas.....	85
TABELA 5.6 – Comparação entre os valores de AIC e BIC para a série hidrológica de Emborcação.....	85
TABELA 5.7 – Valores das ordens mensais inferidas para o modelo PAR das séries hidrológicas.....	86
TABELA 5.8 – Valores dos parâmetros do modelo PAR para a série de Furnas calculados pelos EMV.....	87
TABELA 5.9 – Valores dos parâmetros do modelo PAR para a série de Emborcação calculados pelos EMV.....	87

TABELA 5.10 – Valores dos parâmetros do modelo PAR para a série de Furnas calculados utilizando-se a <i>priori</i> não informativa de Jeffreys.....	88
TABELA 5.11 – Valores dos parâmetros do modelo PAR para a série de Emborcação calculados utilizando-se a <i>priori</i> não informativa de Jeffreys.....	89
TABELA 5.12 – Valores dos parâmetros do modelo AR ajustado para as ordens mensais das séries em estudo.....	90
TABELA 5.13 – Valores dos parâmetros do modelo PAR para a série de Furnas calculados pelos EBEx.....	91
TABELA 5.14 – Valores dos parâmetros do modelo PAR para a série de Emborcação calculados pelos EBEx.....	91
TABELA 5.15 – Valores obtidos a partir dos EMV e o dos EBEx para a série hidrológica de Furnas para o ano de 1980.....	94
TABELA 5.16 – Valores obtidos a partir dos EMV e o dos EBEx para a série hidrológica de Furnas para o ano de 1990.....	95
TABELA 5.17 – Valores obtidos a partir dos EMV e o dos EBEx para a série hidrológica de Emborcação para o ano de 1980.....	94
TABELA 5.18 – Valores obtidos a partir dos EMV e o dos EBEx para a série hidrológica de Emborcação para o ano de 1990.....	95

LISTA DE ABREVIATURAS E SIGLAS

AIC	- Critério de informação de Akaike
AR(p)	- Modelo auto-regressivo de ordem p
ARIMA($p+d, q$)	- Modelo auto-regressivo integrado de médias móveis de ordem $p+d$ e q
ARIMAX	- Modelo auto-regressivo integrado de médias móveis com variável exógena
ARMA(p, q)	- Modelo auto-regressivo de médias móveis de ordem p e q
BIC	- Critério de informação bayesiano
CIG	- Critério de informação generalizado
DEG	- Distribuição exponencial generalizada
EAPM	- Erro absoluto percentual médio
EBEx	- Estimador bayesiano exato
EMV	- Estimador de máxima verossimilhança
EAPMax	- Erro absoluto percentual máximo
ESS	- Soma residual dos quadrados
IC	- Intervalo de confiança
HQ	- Critério de Hanna-Quinn
MA(q)	- Modelo média móvel de ordem q
MCMC	- Simulação Monte Carlo em Cadeias Markov
MF	- Método dos fragmentos
M-H	- Algoritmo de Metropolis Hasting
MV	- Método de máxima verossimilhança
NGM	- Modelo não-gaussiano multicomponente
PAR(p_m)	- Modelo auto-regressivo periódico de ordem p_m
PAR _{LN}	- Modelo AR(1) Log-Normal
PARMA	- Modelo periódico auto-regressivo de médias móveis
PDAR	- Processo binário discreto

PeFAC	- Função de auto-correlação periódica
PeFACP	- Função de auto-correlação parcial periódica
PGAR	- Modelo periódico gama AR(1)
REQM	- Raiz do erro quadrático médio
SARIMA	- Modelo sazonal auto-regressivo integrado de médias móveis multiplicativo
SARMA(p,q)(P,Q) s	- Modelo auto-regressivo de médias móveis sazonal
SDP	- Método de séries de duração parcial
SHIBATA	- Critério de Shibata
SIC	- Critério de Schwarz
WRI	- World Resources Institute

LISTA DE SÍMBOLOS

$\Phi(z)$	- Função de distribuição de uma variável aleatória normal padrão Z
$\Psi(B), \theta(B), \Phi(B)$	- Representação do filtro linear ou sistema modificador de acordo com os modelos
$\Phi(B^S)$ e $\theta(B^S)$	- Polinômios em B^S de graus P e Q
$\Phi_p(B)$ e $\theta_q(B)$	- Polinômios em B de graus p e q
$\Phi = (\phi_1, \dots, \phi_p)^T$	- Conjunto de parâmetros a serem estimados via função de verossimilhança
Φ_{m1} e Φ_m	- Matrizes dos parâmetros c e ϕ para o mês m_1 quando $r=1$ e para o mês m quando $r \neq 1$
\mathcal{F}	- Família de distribuições
α_0, α_i e β_i	- Parâmetros do modelo sazonal com componente determinístico
α_{j-z} e β_{j-z}	- Coeficientes determinados por regressão múltipla dos dados da série histórica
ε_{j-i}	- Variável aleatória normal (0, 1)
ε_T	- Variável aleatória referente a parte AR do modelo ARIMAX
ε_{T-k}	- Variável aleatória referente a parte MA do modelo ARIMAX
ϕ_0	- Parâmetro inicialmente escolhido para iniciar a estimativa do parâmetro verdadeiro
$\hat{\phi}$	- Parâmetro estimado
$\phi_{i,m}$	- Parâmetro i do modelo PAR para o mês m e representa a PeFACP
$\hat{\gamma}_m$	- Covariâncias estimadas
η_{ij}	- Parâmetros dos modelos markovianos com variação de

	regime com aproximação bayesiana
ρ_m	- Representa a PeFAC
$\hat{\sigma}_m^2$	- Estimativa de σ_m^2
μ_m	- Média do mês m do modelo PAR
σ_m	- Desvio padrão no mês m do modelo PAR
τ_m	- Desvio padrão da variável aleatória a_t do modelo PAR
a_i, b_j e c_k	- Parâmetros do modelo ARIMAX
a_t	- Processo aleatório ou ruído branco
$a_{(r,m)}$	- Processo aleatório ou ruído branco no instante t , no mês m , no ano r
B_{10}	- Razão entre o modelo H_1 e o modelo H_0
f_{m1}^1 e f_m^r	- Função densidade de probabilidade para o ano $r=1$ e mês m_1 e função densidade de probabilidade para o ano $r \neq 1$ e mês m
$F(z, \phi)$	- Função de probabilidade ou densidade de z dado ϕ
f_i	- Harmônicos da série de Fourier
H	- Defasagem inicial conveniente à variável exógena
H_0	- Hipótese nula
H_1	- Hipótese alternativa
i, j	- Contadores
I_t	- Conjunto de informações até o instante t
$\{R_t\}$	- Variável indicadora a variação do regime ($R=0, 1$ ou 2)
k e k'	- Números inteiros
$\ell(\phi)$	- Logaritmo da função de verossimilhança de ϕ
$L(\phi)$	- Função de verossimilhança de ϕ
N	- Número de observações da série histórica
m	- Índice referente ao mês do ano ($m=1, 2, \dots, 12$)
p, q, r_H	- Ordens referentes às parcelas AR, MA e à variável exógena
p_{ii}	- Probabilidade de se fazer a transição entre i e i' ou seja entre os regimes representados por I variando entre 0, 1 e 2
p_m	- Ordem do processo PAR
P e Q	- Graus do polinômio em B^S

$P(x, y)$	- Matriz de probabilidade de transição do estado do estado x para o y
$P\{i, j\}$	- Matriz de probabilidade de transição
\bar{q}	- Tamanho da série em função do número de observações da série
r	- Índice referente ao ano da série acrescido de uma unidade a cada variação do índice m de 1 a 12
R_m	- Coeficiente de regressão múltipla
s	- Período
s_t	- Componente sazonal que pode ser estocástico ou determinístico da série de vazões y_t
\bar{s}_t	- Valor médio de s_t
\hat{s}_t	- Componente sazonal estimado para o instante t
\hat{s}_{t+h}	- Valor determinístico da função s_t no instante $t+h$
t	- Indica o instante
$t_{(r,m)}$ e $t_{(r,m)-i}$	- Instante t e $t-i$ para os mês m e o ano r
u	- Variável aleatória $U[0, 1]$
$U[0, 1]$	- Variável uniforme com média zero e desvio padrão 1
x_{m-z}	- Desvio padronizado do logaritmo do volume mensal de precipitações
x_{T-k}	- Variável exógena, no caso do modelo ARIMAX representa a chuva
x_0	- Valor inicial gerado para a simulação em cadeia markoviana
x_j	- Valor gerado para a simulação em cadeia markoviana
X	- Representa a precipitação no modelo markoviano com variação de regime com aproximação bayesiana
X_{t-1}	- Representa a precipitação no modelo markoviano com variação de regime com aproximação bayesiana no instante $t-1$
y	- Novo valor candidato gerado pela cadeia markoviana
$Y_{j,i}$	- Desvio padronizado do modelo multi-regressivo chuva-deflúvio
$Y_{m-z,i}$	- Desvio padronizado do logaritmo do volume mensal de deflúvio

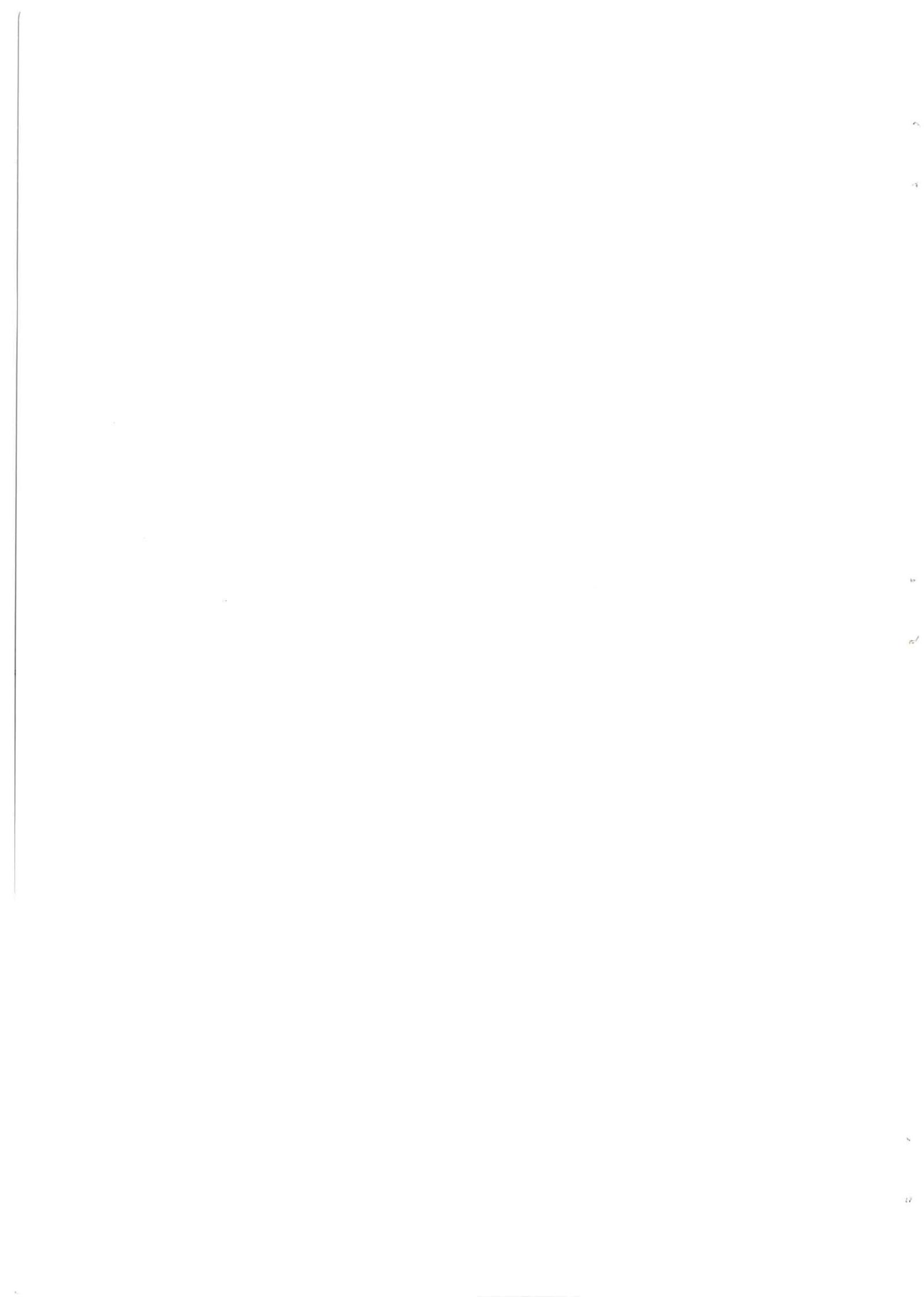
Y_t e Y_{t-1}	- Deflúvio no instante t e $t-1$
Y_T	- Variável original
y_t	- Série de vazões
y_{t+h}	- Previsão de y_t para o instante $t+h$
z	- Observação de um experimento qualquer
z_t	- Componente estacionário da série y_t de vazões
\hat{z}_{t+h}	- Previsão da componente estocástica z_t para o instante $t+h$
Z	- Representação da série histórica
Z_m e X_m	- Matrizes das séries de vazão do mês m
Z_{m1} e X_{m1}	- Matrizes das séries de vazão do mês m_1
Z_T	- Transformada da variável original Y_T
Z_t	- Série temporal
$Z_{t(r,m)}$	- Série temporal no instante t para o ano r e o mês m
$\hat{Z}_{t(r,m)}$	- Série temporal padronizada no instante t para o ano r e o mês m
W	- Valor futuro
$(\quad)'$	- Simboliza matriz transposta

RESUMO

GOMES, M. H. R. (2002). *Uso da abordagem bayesiana para a estimativa de parâmetros sazonais dos modelos auto-regressivos periódicos*. São Carlos, 2003. 118p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

O presente trabalho tem por finalidade o uso da abordagem bayesiana para a estimativa de parâmetros sazonais dos modelos periódicos auto-regressivos (PAR). Após a determinação dos estimadores bayesianos, estes são comparados com os estimadores de máxima verossimilhança. A previsão para 12 meses é realizada usando os dois estimadores e os resultados comparados por meio de gráficos, tabelas e pelos erros de previsão. Para ilustrar o problema as séries escolhidas foram as séries hidrológicas da Usinas Hidroelétricas de Furnas e Emborcação. Tais séries foram selecionadas tendo em vista a necessidade de previsões com reduzido erro já que o sistema de operação das usinas hidroelétricas depende muito da quantidade de água existente em seus reservatórios e de planejamento e gerenciamento eficazes.

Palavras-chave: modelo PAR; séries temporais hidrológicas; abordagem bayesiana



ABSTRACT

GOMES, M. H. R. (2002). Use of bayesian method to the estimate of sazonal parameters of periodic autoregressive models. São Carlos, 2003. 118p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

The objective of this research is to use bayesian method to estimate of sazonal parameters of periodic autoregressive models (PAR). The bayesian estimators are then compared with maximum likelihood estimators. The forecast for 12 months is made by using two estimators and comparing their results though graphs, tables and forecast error. The hydrological time series chosen were from Furnas and Emborcação Hydroelectric Power Plant. These series were chosen having in mind the necessity of series with reduced error in their forecast because system of operation in the Hydroelectric Power Plant depends on the quantity of the water in their resevoirs, efficient planning and management.

Key-Words: PAR models; hydrological time series; bayesian method



1 - INTRODUÇÃO

O Brasil possui cerca de 13,8% do deflúvio mundial, o que significa uma parcela considerável da disponibilidade hídrica do mundo. Apesar disto, o país enfrenta os mesmos problemas do restante do mundo, o que implica na necessidade de projetos de gestão, gerenciamento e planejamento do uso dos recursos hídricos para garantir o seu desenvolvimento social e econômico. Para assegurar a disponibilidade de água para os vários setores, segundo as demandas exigidas, torna-se necessário o gerenciamento de um sistema hídrico conforme mencionado por KELMAN (1987), significando a possibilidade de poder atuar no sistema, tendo como objetivo a distribuição temporal e espacial da água.

Objetivando, solucionar os problemas mencionados acima, cada vez mais surgem pesquisas em todo o mundo visando simular o comportamento futuro dos recursos hídricos (deflúvios, águas subterrâneas, etc.) a fim de se poder estabelecer metas de consumo aliadas à produção de água sem prejuízo para o homem, o meio ambiente e o desenvolvimento econômico e social das regiões.

Para alcançar este objetivo muitos métodos para geração de séries sintéticas foram criados, e são largamente utilizados. Isto ocorreu graças a popularização do uso de séries sintéticas de vazões, nos anos 60, pelo Havard Water Program, que as utilizou na resolução de problemas na operação e projeto de reservatórios de sistemas hídricos de bacias.

Os métodos empregados podem ser determinísticos ou estocásticos. Em hidrologia estocástica, os modelos empregados são erigidos sobre os conceitos de probabilidade e estatística. Vários são os métodos empregados e investigados para geração de séries sintéticas, como também muitas são as melhorias aplicadas a cada método visando a geração de séries sintéticas que se aproximem cada vez mais das séries históricas.

Os modelos matemáticos utilizados com a finalidade de gerar séries hidrológicas estão vinculados aos graus de incerteza devido aos horizontes de planejamento. A

necessidade de previsões com reduzido erro tem sido cada vez maior, tendo em vista a escassez dos recursos hídricos e a crescente necessidade de se gerenciar de forma mais eficaz os sistemas existentes, como por exemplo, o sistema de geração de energia elétrica, que no Brasil na sua maior parte, é de origem hídrica. Quando considerada a análise da aleatoriedade das vazões futuras, observa-se que a incerteza aumenta com o tempo. Analisando a Figura 1.1, pode-se observar três períodos distintos quando as características do fenômeno se diferenciam (Tabela 1.1). As faixas apresentadas no gráfico possuem larguras distintas e são dependentes das características do ciclo hidrológico. A Tabela 1.1 apresenta a relação entre os horizontes de planejamento, os graus de incerteza relativos à previsão e a descrição de cada período.

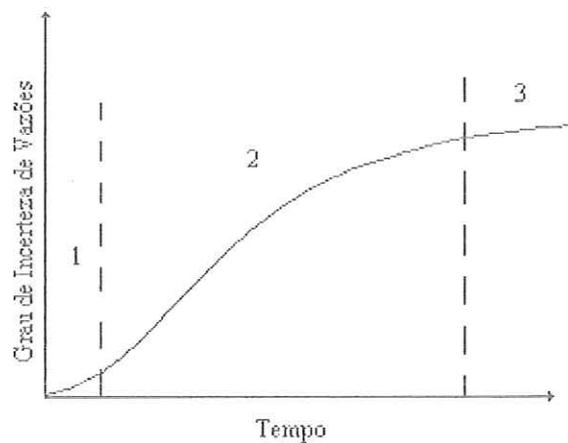


FIGURA 1.1 – Incerteza das vazões futuras (POLTRONIERE SILVA, 1999)

TABELA 1.1 – Relação entre os horizontes de planejamento e os graus de incerteza e a descrição de cada período.

Faixas	Horizonte	Grau de Incerteza
1	Curto prazo (uma semana)	A incerteza sobre as vazões futuras apresenta-se muito pequena devido à capacidade de previsão obtida pelos modelos que utilizam informação meteorológica e pelos modelos baseados na relação chuva-vazão.
2	Médio prazo (algumas semanas ou meses)	A incerteza é crescente (mas menor que a incerteza a longo prazo) e ocorre devido à correlação entre a vazão futura e a presente (e passadas). São utilizados de modelos de previsão das vazões baseadas em séries temporais.
3	Longo prazo (alguns meses ou anos)	Existe a necessidade de considerar as vazões em um futuro remoto. A incerteza cresce até atingir a saturação (isto representa a independência estocástica entre a vazão futura e a conjuntura presente). São utilizados modelos baseados em processos markovianos para as vazões.

Diversos modelos estocásticos são apresentados na literatura (BOX et al. 1995), a família de modelos aqui escolhida para investigação foi a família dos modelos auto-regressivos periódicos (PAR) cujos parâmetros variam de acordo com a sazonalidade (os parâmetros variam para cada modelo mensal). Na análise, uma etapa importante é a estimativa dos parâmetros dos modelos, uma escolha equivocada pode gerar uma representação inadequada da série. Entre os métodos para determinação dos parâmetros de modelos auto-regressivos, podem-se citar, o método dos momentos, dos mínimos quadrados e o método de máxima verossimilhança, já exaustivamente investigados. Entretanto, tem despertado interesse de alguns pesquisadores, um outro método para análise e inferência de séries temporais, inclusive de séries temporais hidrológicas, que são os métodos bayesianos. O método bayesiano se deve aos esforços do pesquisador Jacques Bernier, que foi o precursor da idéia bayesiana na hidrologia e, é devido à importância de seu trabalho que o método bayesiano tem permeado as ciências hidrológicas. O suporte para a idéia bayesiana é conceitual e prático: a aproximação bayesiana fornece uma estrutura coerente que facilita a análise de problemas de decisão sobre incertezas. (BERGER, 1985).

O uso da inferência bayesiana como um método para análise de séries temporais hidrológicas e estimativa paramétrica dos modelos representativos dessas séries tornaram-se alvo de interesse. Os métodos bayesianos foram empregadas como uma alternativa para a redução do esforço computacional no cálculo das estimativas

paramétricas dos modelos PAR e têm por finalidade a obtenção de resultados mais precisos do aqueles obtidos via análise de máxima verossimilhança, ou seja, estimativas paramétricas melhores, já que nos casos de séries temporais geralmente tem-se amostras pequenas para ajuste de modelos PAR.

Dentre os diferentes métodos para estimar os parâmetros dos modelos foi escolhido o método de máxima verossimilhança por ser um método amplamente utilizado e que fornece bons resultados. Neste método, os parâmetros de um modelo são fixados como constantes. Enquanto que na teoria da inferência bayesiana, são considerados variáveis aleatórias .

Esta pesquisa tem por finalidade avaliar e estimar os parâmetros dos modelos periódicos auto-regressivos (PAR), para séries de vazões médias mensais, via estatística bayesiana e compará-las com as estimativas paramétricas de máxima verossimilhança. A comparação das estimativas paramétricas é também realizada entre os modelos PAR e os modelos auto-regressivos (AR). Os programas utilizados foram desenvolvidos com auxílio do *software* para computação numérica MATLAB 6.0. As séries hidrológicas utilizadas são as séries de vazões médias mensais das usinas hidroelétricas de Furnas e Emborcação que fazem parte do Sistema Sudeste Brasileiro de geração de energia elétrica.

2 - OBJETIVOS

O grande crescimento da utilização de água de boa qualidade em todo o mundo tem levado a um aumento da preocupação com o uso racional deste recurso, traduzindo-se em um crescente número de pesquisas que objetivam o planejamento dos recursos hídricos. É cada vez mais necessário o gerenciamento dos sistemas hídricos de forma a garantir a distribuição da água visando o desenvolvimento social e econômico de uma região e/ou país, além de assegurar a disponibilidade de água para os vários setores, segundo as demandas exigidas. Esta pesquisa investigará o uso de modelos periódicos auto-regressivos – PAR para previsão de séries de vazões médias mensais por meio de uma abordagem bayesiana onde serão estimados os parâmetros do modelo, além de se comparar os modelos PAR com os modelos auto-regressivos (AR). O número de parâmetros de cada modelo será inferido pelos critérios de informação bayesiano (BIC) e pelo critério de informação de Akaike (AIC). As previsões realizadas com os estimadores bayesianos serão comparadas com aquelas obtidas usando-se estimadores de máxima verossimilhança (EMV).

2.1 – Objetivos específicos

- i. Estimar os parâmetros de modelos periódicos auto-regressivos – PAR por meio de uma abordagem bayesiana;
- ii. Estimar os parâmetros de modelos periódicos auto-regressivos – PAR por meio de uma abordagem clássica (método de máxima verossimilhança);
- iii. Verificação da capacidade preditiva dos modelos PAR pela abordagem bayesiana;
- iv. Comparar as previsões realizadas utilizando-se os modelos PAR tanto pelo enfoque bayesiano quanto pelo clássico;

- v. Comparar a previsão obtidas pelos modelos PAR e os modelos auto-regressivos (AR) tanto pelo enfoque bayesiano quanto pelo clássico;
- vi. Apresentar a viabilidade do uso da abordagem bayesiana como um método de estimativa paramétrica para séries hidrológicas.

3 – REVISÃO DE LITERATURA

Neste capítulo serão introduzidos alguns conceitos básicos de séries temporais, processos estocásticos e inferência bayesiana usados para modelação de séries hidrológicas. Pode-se considerar, como razão principal, para modelação de uma série de vazões, por exemplo, a necessidade de prever o comportamento futuro dessa série. Alguns exemplos de pesquisas envolvendo modelos estocásticos e bayesianos também serão apresentados como forma de ilustrar a aplicação prática desses modelos.

3.1 – Séries temporais

3.1.1 – Conceito

Uma série temporal é uma série de observações tomadas, sequencialmente, no tempo. Em um modelo de regressão padrão a ordem na qual as observações são incluídas no conjunto de dados é irrelevante, ou seja, qualquer ordenação é igualmente satisfatória. Esta propriedade de ordem é crucial para séries temporais e é ela que distingue uma série temporal de uma série não temporal. Ações tomadas em algum intervalo de tempo têm conseqüências e efeitos que são experimentados em algum tempo posterior. O próprio tempo, através dos mecanismos de casualidade, concede estrutura dentro de uma série temporal. Em aplicações típicas, sucessivas observações poderão ser igualmente espaçadas, como por exemplo: diária, semanal, mensal. Mas, tal espaçamento é somente uma conveniência e isto não é, certamente, uma restrição necessária. Qualquer seqüência de observações pode ser transformada em uma série espaçada igualmente pela adoção de uma escala temporal apropriada (POLE et al., 1999). Muitos conjuntos de dados apresentam-se como séries temporais, como por exemplo as séries hidrológicas de precipitação e vazão. Uma característica intrínseca

das séries temporais é que, tipicamente, as observações são dependentes e a natureza desta dependência é de interesse prático.

3.1.2 – Finalidade da análise das séries temporais

A análise das séries temporais, de acordo com PEREIRA et al. (1986) tem por finalidade entender o mecanismo gerador da série e prever o comportamento futuro do sistema. Para tal é necessário: (i) descrever o comportamento das séries, por meio de gráficos, histogramas, estatísticas descritivas simples das séries, verificação de tendências, ciclos, variações sazonais, etc; (ii) investigar o mecanismo gerador das séries temporais; (iii) prever valores futuros a partir de valores passados das séries, e (iv) procurar periodicidade relevante no conjunto de dados que caracterizam as séries.

3.1.3 – Classificação das séries temporais hidrológicas

A classificação de séries, particularmente, de séries temporais é adaptada por YEVJEVICH (1972) para seqüências hidrológicas. Todas as séries elementares ou os componentes de séries complexas e compostas podem ser classificados em dois grupos básicos – séries determinística e séries estocásticas – e cada uma pode, sucessivamente, ser classificada em outros grupos como mostrado na Figura (3.1). Cada grupo descrito e suas subdivisões podem ser encontrados na análise de séries hidrológicas, alguns mais freqüentes que outros.

De acordo com PARZEN (1962), as séries temporais hidrológicas determinísticas são definidas como sendo as séries expressas matematicamente por leis exatas. As seqüências determinísticas podem ser divididas em periódicas e não-periódicas. As séries periódicas podem dividir-se em:

- Séries senoidais: São aquelas que possuem um componente periódico determinístico aproximadamente senoidal, porém, nenhuma série hidrológica é uma simples função seno ou coseno;
- Séries periódicas complexas: São aquelas que têm como característica principal a repetição dos seus valores em intervalos (períodos) de tempo.

Já as séries não-periódicas dividem-se em:

- Séries quase-periódicas: São aquelas em que a relação entre duas quaisquer frequências é um número irracional, não existindo nenhum período e as frequências de funções trigonométricas são indeterminadas. Este tipo de série, dificilmente, é encontrado em hidrologia, sendo mais comum em oceanografia;
- Séries transitórias: O ambiente hidrológico pode ser afetado por mudanças repentinas produzidas pelo homem ou por transformações na natureza causando alterações em algumas características estatísticas da série como média, variância, coeficiente de assimetria e coeficiente de autocorrelação. Essas variações são lentas e podem ser lineares ou não, ou caracterizadas por saltos.

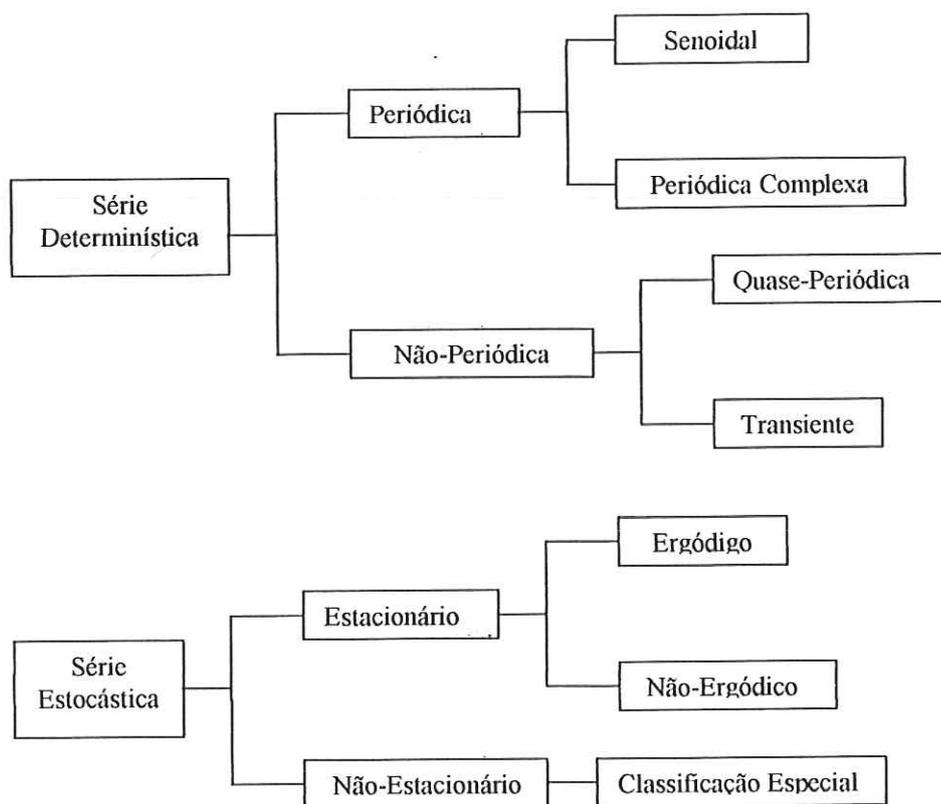


FIGURA 3.1 – Classificação geral das séries temporais segundo YEVJEVICH (1972)

PARZEN (1962) define as séries temporais hidrológicas estocásticas como as séries que seguem leis probabilísticas e podem ser divididas em estacionárias e não-estacionárias. As séries estacionárias são aquelas cuja distribuição permanece inalterada com o tempo, porque os mecanismos que as produzem não variam, ou seja, dividindo-se uma série em várias sub-séries e usando um parâmetro estatístico qualquer para caracterizar os dados em cada sub-série, quando o valor esperado desse parâmetro não se alterar em cada sub-série, a série é denominada uma série temporal estacionária. Em caso contrário, a série é denominada não-estacionária.

As séries estacionárias podem ser ergódicas ou não-ergódicas. Para que o processo estacionário seja ergódico, há a necessidade das médias amostrais, formadas de um registro da seqüência, sejam consideradas como uma aproximação da média da população, ou seja, as propriedades do processo sejam convergentes às da população. Caso contrário, o processo será não-ergódico.

3.1.4 – Séries temporais versus previsão

De acordo com POLE et al. (1999), previsão e análise de séries temporais são duas atividades distintas. A previsão de uma série é uma visão de um futuro incerto, já a análise de uma série temporal é uma descrição daquilo que já tenha acontecido. A análise pode ser feita sem realizar a previsão, igualmente, pode-se fazer uma previsão sem efetuar análise da série temporal (ou qualquer outra). Mais precisamente, a modelação e o processo de análise são concluídos uma vez que a previsão tenha sido produzida. Eventualmente, a quantidade, sendo prevista, torna-se conhecida, este valor fornece novas informações a serem incorporadas à análise. Consequentemente, em uma situação típica, uma previsão para um intervalo subsequente poderá ser produzida e o ciclo observação-análise-previsão poderá ser repetido.

3.2 – Modelos para geração de séries temporais

KELMAN (1987) considera que gerenciar um sistema hídrico significa atuar no sentido de assegurar uma distribuição temporal e espacial da água que melhor se adequam às necessidades da comunidade. Porém, não se pode garantir que o objetivo do

gerenciamento será atingido, já que inúmeras são as incertezas associadas aos problemas em questão.

Para se tratar dessas incertezas, busca-se quantificar e atribuir probabilidades a eventos futuros relacionados com a disponibilidade dos recursos hídricos. Os modelos e algoritmos empregados em hidrologia estocástica são erigidos sobre conceitos de probabilidade e estatística.

Segundo KELMAN (1987), nos processos aleatórios, a ordem dos eventos é irrelevante. Existem, porém, outros casos em que a ordem é relevante. Por exemplo, na análise do deflúvio anual de uma bacia com solos profundos e com grandes capacidades de reter água, um valor muito acima da média dificilmente será seguido de um valor muito abaixo da média. Logo, em casos como este, as observações não formam uma amostra aleatória independente, mas sim, uma *série temporal*. Neste caso, não se tem uma variável aleatória, mas um vetor de variáveis aleatórias, geralmente, dependentes entre si. Esta seqüência indexada de variáveis aleatórias é também chamada de processo estocástico. Diz-se que uma série temporal é uma realização do processo estocástico.

Os problemas envolvendo processos estocásticos são resolvidos em 4 etapas:

- i. Coleta e organização de uma série temporal do processo estocástico;
- ii. Seleção de um modelo matemático para o processo estocástico;
- iii. Estimativa dos parâmetros do modelo e teste do ajuste e adequação do modelo;
- iv. Utilização do modelo ajustado para obtenção da resposta desejada.

3.2.1 - Simulação Monte-Carlo

Ao se planejar ou operar um aproveitamento hídrico, segundo KELMAN (1987), utiliza-se uma série histórica como dado do modelo de simulação e/ou otimização que produza como saída uma grandeza relevante ao projeto. A série histórica é apenas uma das possíveis realizações, ocorrida aleatoriamente segundo leis probabilísticas, dentre as inúmeras séries do fenômeno na natureza. Uma nova ocorrência redundaria em uma série diferente, porém igualmente provável. Os modelos estocásticos tentam aproximar este comportamento fazendo, artificialmente, tantas escolhas quanto forem necessárias para o estudo do fenômeno focado. Cada escolha é associada a uma série de vazões, chamada de série sintética. Como essas séries são distintas entre si, é possível que sejam

obtidos diferentes resultados provenientes das simulações. Sendo assim, podem-se extrair, completamente, as informações contidas na série histórica e utilizá-las na execução dos projetos por meio de séries geradas, segundo suas necessidades. O método usado, neste caso, é denominado de Simulação Monte Carlo.

De acordo com MAASS et al.¹ apud BRAS & RODRIGUEZ-ITURBE (1985), a geração sintética de séries hidrológicas (simulação Monte Carlo) foi popularizada nos anos 60 através das pesquisas do Harvard Water Program. Vazões sintéticas foram usadas nos problemas de operação e projeto de reservatórios de sistemas hídricos de bacias. O objetivo das simulações hidrológicas sintéticas é produzir um conjunto de características igualmente prováveis (tão longas quanto necessário) que são estatisticamente indistinguíveis dos dados históricos.

O método de simulação Monte Carlo é muito usado para prever a operação de um sistema de recursos hídricos, quando as estruturas hidráulicas já estão dimensionadas e os armazenamentos iniciais são dados pelas condições atuais do sistema, além de ser aplicado em problemas mais complexos.

3.2.2 – Cadeias de Markov

Segundo GAMERMAN (1996), o conceito de dependência de Markov é imputado ao matemático russo Andrei Andreivich Markov que investigou, no início do século XX, a alternância entre as vogais e as consoantes existentes no poema “Onegin” de Poeshkin. Markov desenvolveu um modelo probabilístico onde os resultados sucessivos dos experimentos dependiam dos resultados antecedentes somente através de seu predecessor. Esse modelo permitiu que seu criador obtivesse uma boa estimativa da frequência com que as consoantes apareciam no poema. Quase, simultaneamente, segundo o autor, outro matemático, o francês Henri Poincaré, investigou seqüências de variáveis aleatórias que na realidade são seqüências markovianas, entretanto não houve a continuidade das investigações como fez Markov, recebendo este nome devido ao reconhecimento.

¹MAASS, A.; HUFSCHMIDT, R.; DORFMAN, R.; THOMAS, Jr. H. A.; MARGLIN, S. A. & FAIR, G. M. (1962). *Design of Water Resource Systems*. Cambridge, Mass.: Harvard Unity Press apud BRAS, R. & RODRIGUEZ-ITURBE, I. (1985). *Random Functions and Hydrology*. EUA. Addison-Wesley Publishing Company.

As cadeias de Markov são processos estocásticos que pertencem à parte dinâmica da teoria das probabilidades onde se estuda uma coleção de variáveis aleatórias sob o enfoque de sua interdependência e de seu comportamento limite.

3.2.2.1 – Simulação via cadeias de Markov

Um processo markoviano é um processo cujo valor futuro da variável aleatória depende do valor presente dessa mesma variável. Quando esse processo assume um número finito ou infinito contável de possíveis valores, então, esse processo recebe o nome de Cadeia de Markov.

Nos processos markovianos existe uma matriz de probabilidade denominada matriz de probabilidade de transição $P\{i, j\}$, ou seja, se uma cadeia de Markov está no estado i no instante t deve transitar para algum estado j no instante $t+1$ com esta probabilidade P . Normalmente, utiliza-se o algoritmo de Metropolis-Hasting (M-H), para simular uma trajetória da cadeia de Markov, apresentado a seguir:

- i) Condição inicial qualquer: x_0 ; Contador: $j \rightarrow 1$
- ii) $y \sim P(x_{j-1}, y)$ – Gerar um candidato pertencente ao núcleo P
- iii) Verificar se o candidato gerado é aceito então a transição será realizada.

Gerar um n° aleatório $u \sim U[0,1]$

Se $u \leq \alpha(x_{j-1}, y)$

$$x_{j-1} = y$$

caso contrário

$$x_j = x_{j-1}$$

fim

- iv) Atualizar o contador $j \rightarrow j+1$ e repetir os passos ii e iii.

Onde $U[0,1]$ é um número aleatório com distribuição uniforme e α a probabilidade de aceitação do candidato gerado.

Quando a cadeia alcança equilíbrio, gera-se $y \sim P(x_{j-1}, y)$ e temos $\alpha(x_{j-1}, y) = 1$. Ou seja, todo candidato é aceito, assim, o algoritmo consiste simplesmente em:

- i) Condição inicial qualquer: x_0 ; Contador: $j \rightarrow 1$;
- ii) $y \sim P(x_{j-1}, y)$ – Gerar um candidato pertencente ao núcleo P ;
- iii) $x_j = y$
 $j \rightarrow j+1$
- iv) Retornar aos passos (ii) e (iii) até a convergência.

GAMERMAN (1996) afirma que as cadeias de Markov abrem um grande leque de possibilidades em simulação para problemas complexos onde simplesmente não há como aplicar os métodos convencionais.

3.2.3 – Modelos estocásticos

Os modelos utilizados para descrever séries temporais descrevem processos estocásticos, isto é, processos controlados por leis probabilísticas. Para KELMAN (1987), a utilização do modelo estocástico de vazões deve ser encarada simplesmente como uma das maneiras de se manipular a informação existente. Nenhuma informação é criada, a não ser a adoção de hipóteses subjacentes ao modelo estocástico.

A análise de séries temporais generalizada, segundo BOX & JENKIS (1976), não parte da estrutura do modelo, mas sim do suporte (um tipo de modelo) que ajudará na identificação do modelo específico descrevendo melhor o fenômeno. Os modelos são expressos em termos de diversos componentes: tendência, periodicidade, memória, aleatoriedade, etc.

Entende-se por modelo de série o sistema modificador, ou filtro, que converte uma seqüência de variáveis aleatórias não correlacionadas (normalmente distribuídas) a_t para uma seqüência de variáveis correlacionadas Z_t . O operador representando o filtro pode ser denominado como $\psi(B)$. A seqüência de variáveis aleatórias não correlacionadas a_t pode ser denominada como ruído branco (Figura 3.2).

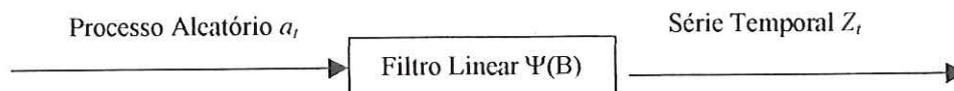


FIGURA 3.2 – Esquema do filtro linear agindo no ruído para criar uma série temporal segundo BOX & JENKINS (1994).

BOX & JENKINS (1994) propuseram um algoritmo iterativo para identificar o modelo, estimar os parâmetros e verificar o modelo (Figura 3.3).

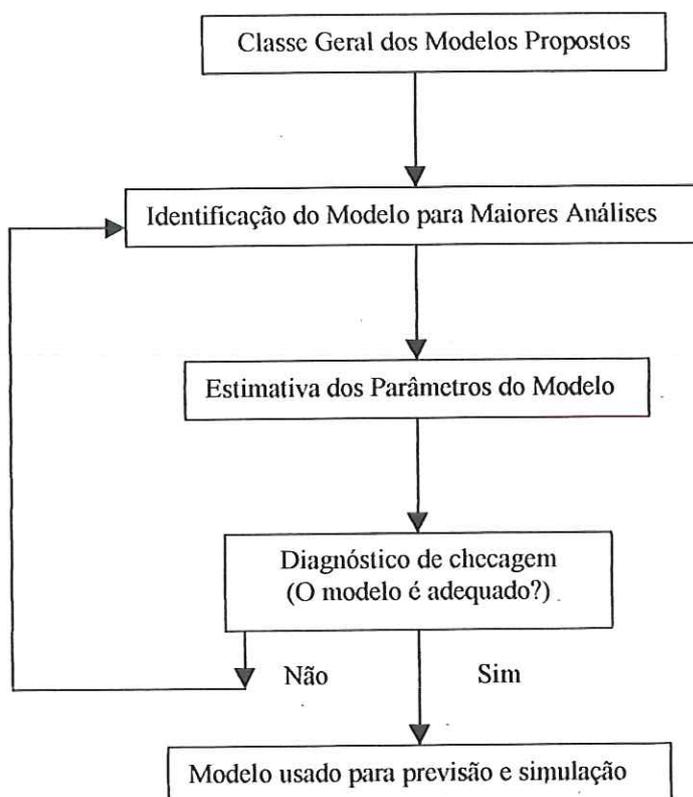


FIGURA 3.3 – Procedimento iterativo para construção do modelo segundo BOX & JENKINS (1994).

3.2.3.1 - Exemplos do uso dos modelos estocásticos

➤ Modelo para as vazões médias diárias

COIMBRA (1978) apresentou um modelo específico para vazões diárias de rios intermitentes, no qual os eventos são considerados como a soma de dois componentes. Um componente determinístico que é calculado pela equação de depleção dos aquíferos e um aleatório, baseado em matrizes de probabilidades de transição de estados mensais e diários. A ordem markoviana do processo de dias com vazão igual ou maior que zero é de fundamental importância para o modelo e é determinada por um método simples e de grande aplicação em hidrologia. Na formulação do modelo foi imposto pelo autor um processo markoviano mensal cujos estados indicam se em um determinado mês há ou não-*input*. Contudo foi necessária também uma matriz markoviana diária, pois independentemente de haver, em um mês qualquer, *input* ou não-*input*, isto não implica que este ocorra em todos os dias desse mês. As matrizes markovianas usadas mostraram os períodos com maior probabilidade de ocorrência dos eventos. A série sintética foi gerada para 100 anos e suas características comparadas com a das séries históricas. COIMBRA (1978) analisou os resultados e concluiu que, para a série de vazões do rio Jaguaribe, o modelo proposto é adequado.

➤ Modelo conceitual markoviano para previsão e geração de deflúvios

OCCHIPINTI (1979) obteve um modelo conceitual chuva-deflúvio, não-linear, auto-regressivo (AR) do tipo markoviano no qual as precipitações são consideradas insumos e os deflúvios, respostas. O modelo conceitual multi-regressivo chuva-deflúvio usado pelo autor pode ser expresso pela seguinte formulação:

$$Y_{m,r} = \sum_{z=1}^N \alpha_{m,z} Y_{m-z,r} + \sum_{z=0}^N \beta_{m,z} x_{m-z,r} + (1 - R_m^2)^{1/2} \epsilon_{m,r} \quad (3.1)$$

onde:

- z série histórica;
- $Y_{m-z,r}$ e $x_{m-z,r}$ são os desvios padronizados dos logaritmos dos volumes mensais de deflúvios e precipitações;
- R_m é o coeficiente de regressão múltipla;
- $\varepsilon_{m,r}$ é uma variável aleatória normal (0, 1);
- $\alpha_{m,z}$ e $\beta_{m,z}$ são os coeficientes determinados por regressão múltipla dos dados da série histórica;
- m é o índice referente ao mês do ano ($m = 1, 2, \dots, 12$);
- r é o índice referente ao ano da série acrescido de uma unidade a cada variação do índice m de 1 a 12.

O modelo foi sintetizado na forma de um modelo com um parâmetro único e três funções mostrando que a resposta (deflúvios) poderia ser avaliada pelo produto de duas funções; a função persistência que é a média geométrica ponderada dos eventos antecedentes chuva-deflúvio, e a função forçante que é caracterizada pela precipitação atual. De acordo com o autor, o modelo permite obter uma ótima correlação chuva-deflúvio e como consequência avalia a consistência dos dados, estende as séries de deflúvios, gera precipitações como únicas variáveis aleatórias e a partir destas gera os deflúvios. O modelo geral possui a vantagem de poder ser aplicado em qualquer bacia, em qualquer época e inclui outros modelos, como os markovianos, etc.

Segundo OCCHIPINTI (1979), a eficiência do modelo foi testada e comprovada quando comparado com modelos markovianos auto-regressivos - AR (de ordem 1 e de ordem superior) utilizando-se a mesma série histórica de chuvas e de deflúvios. O modelo também apresenta um grau de determinação superior e uniforme ao longo do ciclo anual quando comparado com outros modelos.

➤ Modelo estocástico para vazões mensais

A análise de intervenção, segundo BOX & TIAO (1975), descrevendo as mudanças no nível médio de uma seqüência devido às causas humanas ou naturais foi estudado por GENOVEZ (1982) através de um tipo de função de transferência de Box-

Jenkins em associação com o modelo da componente aleatória (modelo sazonal auto-regressivo integrado multiplicativo de médias móveis – modelo SARIMA).

Três diferentes modelos foram investigados com diferentes graus de abrangência. Primeiro, houve a tentativa de se explicar vazão através da intervenção (urbanização), das entradas (precipitação do mês e do mês anterior) e do modelo SARIMA. Como os parâmetros estimados da componente de intervenção não foram significativos, GENOVEZ (1982) ajustou um outro modelo com entradas (precipitação do mês e do mês anterior) e o modelo SARIMA, sem a componente de intervenção. Já para o terceiro modelo foi utilizado apenas a estrutura SARIMA para explicar as vazões, ou seja, sem a intervenção e sem as entradas. Para os três modelos, o autor sugeriu a dessazonalização dos dados antes de ajustar os modelos. De acordo com o critério de informação de Akaike (AIC), o modelo com entradas é mais apropriado já que este fornece um valor menor para o critério AIC do que o modelo sem as entradas. Como a maior vantagem do tipo de análise empregada, o autor aponta a possibilidade de aproveitamento de toda extensão das séries hidrológicas observadas, mesmo quando ocorrem perturbações devido às atividades de desenvolvimento, na modelação estocástica.

➤ Modelo para previsão de vazão em tempo real

COSTA (1989) utilizou os modelos da família ARIMAX visando a previsão de vazões em tempo real, sem levar em conta os termos de médias móveis (MA), porém, dando ênfase à participação da variável auto-regressiva (AR), variáveis exógenas representadas pela letra X e diferenças de 1ª ordem representadas pela letra I . Um modelo ARIMAX além de conter variáveis aleatórias do tipo Y_t , abrange variáveis exógenas, como é o caso da chuva, em transformação chuva-vazão. A expressão matemática de um modelo ARIMAX é dada da forma:

$$Z_T = \left(\sum_{i=1}^p a_i Z_{T-i} + \varepsilon_t \right) - \sum_{j=1}^q b_j \varepsilon_{T-j} - \sum_{k=H}^{I_H} c_k X_{T-k} \quad (3.2)$$

onde:

- Z_T é uma transformada da variável original Y_t ;
- a_i , b_j e c_k são parâmetros do modelo;

- x_{T-k} é a variável exógena e,
- H é a defasagem inicial conveniente à variável exógena;
- p, q, r_H são as ordens referentes às parcelas AR, MA e da variável exógena.

Na eq.(3.2), a 1ª parcela refere-se aos termos auto-regressivos, a 2ª engloba os termos média móvel e a 3ª envolve a variável exógena x_{T-k} . Nota-se, porém, que não é necessário incluir apenas uma única variável exógena, outras podem ser incluídas.

As variáveis do tipo AR, utilizadas por COSTA (1989), são as vazões na seção de previsão, que é coincidente com a seção de jusante de um trecho de rio previamente escolhido. Já as variáveis exógenas possuem duas origens: as vazões são observadas em uma seção de montante e as alturas médias de chuva que exercem maior influência nas contribuições laterais.

COSTA (1989) aplicou os modelos ARIMAX a 3 (três) sub-bacias, 2 (duas) situadas a montante e 1 (uma) a jusante do reservatório de Americana, na bacia do rio Piracicaba, estado de São Paulo. Os coeficientes de correlação para as 3 (três) sub-bacias foram superiores a 95%, entre valores previstos e observados ao longo do ano de 1982, adotando a hipótese de conhecimento antecipado, de um intervalo de tempo, das alturas médias.

➤ **Modelo estocástico para simulação de vazões diárias em bacias de desenvolvimento elevado**

KOTTEGODA et al. (1995) utilizaram um processo markoviano para períodos secos e chuvosos na modelação de vazões diárias. O objetivo principal foi simular baixas vazões em bacias com grande desenvolvimento, sujeitas à irrigação e outras influências. A aproximação considerada foi baseada no uso de seqüências históricas de descargas diárias do rio Po, localizado na Itália, para o desenvolvimento e calibração do modelo estocástico. As vazões foram obtidas de medidas realizadas nas principais estações do rio colaborando para um detalhado estudo das baixas vazões.

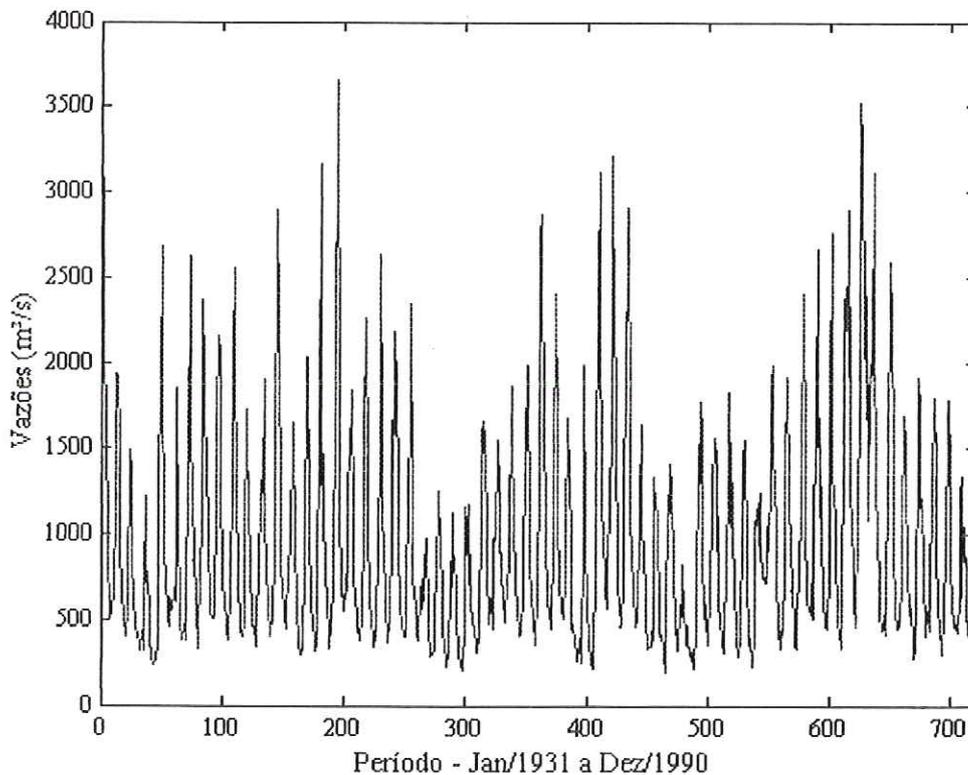
Um gerador log-normal foi usado, por KOTTEGODA et al. (1998), para as magnitudes de baixas vazões, as quais são dependentes da duração de recessão, mas possuem baixa correlação serial. Um processo ARIMA foi adotado para simular a

recessão anterior no tempo entre uma baixa e uma elevada vazão. O estudo realizado forneceu um procedimento alternativo para simulação de vazões diárias com particular referência aos eventos extremos mínimos. De acordo com os autores, essa aproximação pode ser adaptada para as condições peculiares de cada bacia.

As vantagens desse estudo, apontadas pelos autores, foram: a modelação das seqüências de períodos chuvosos e secos tão próximas quanto possíveis; modelação realística das baixas vazões; simulação do comportamento da recessão usando-se características das recessões observadas e também se impondo efeitos aleatórios causados por vários processos de entrada e saída.

3.3 – Modelos estocásticos sazonais

As séries hidrológicas, geralmente, apresentam um comportamento periódico (sazonal), como ilustra a Figura (3.4). Esse tipo de comportamento representa um caso particular de não estacionariedade da série.



FIGURAS 3.4 – Série de Vazões Médias Mensais da Usina Hidroelétrica Furnas de Janeiro/1931 a Dezembro/1990

Uma série temporal sazonal, geralmente, é representada por dois componentes, um sazonal que pode ser considerado estocástico ou determinístico s_t e outro estacionário z_t . Assim, a vazão y_t pode ser escrita como:

$$y_t = s_t + z_t \quad (3.4)$$

3.3.1 – Modelo com componente sazonal determinístico

Quando se considera a componente sazonal de uma série, esta pode ser representada por uma função determinística periódica que se ajusta aos dados utilizando-se uma série de Fourier. A componente sazonal pode então ser escrita como:

$$\hat{s}_t = \alpha_0 + \sum_{i=1}^q (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) \quad (3.5)$$

sendo:

$$\alpha_0 = \bar{s}_t, \text{ com } \bar{s}_t = \frac{1}{N} \sum_{t=1}^N s_t \quad (3.6)$$

$$\alpha_i = \frac{2}{N} \sum_{t=1}^N y_t \cos 2\pi f_i t \quad (3.7)$$

$$\beta_i = \frac{2}{N} \sum_{t=1}^N y_t \sin 2\pi f_i t \quad (3.8)$$

onde:

- N é o número de observações da série,
- $i = 1, 2, \dots, q$,
- $q = \frac{(N-1)}{2}$ quando N é ímpar ou $q = \frac{N}{2}$ quando N é par.

A identificação dos harmônicos ($f_i = i/s, i = 1, 2, \dots, q$), onde s é o período que melhor representa s_t , é a grande dificuldade desse método. Após o ajuste do componente sazonal da série, é possível isolar o componente estocástico estacionário:

$$z_t = y_t - \hat{s}_t \quad (3.9)$$

Dessa forma a série z_t pode ser modelada como um processo auto-regressivo de médias móveis ARMA(p, q): $\phi(B)z_t = \theta(B)\alpha_t$. Assim, quando supomos a existência de um componente sazonal determinístico, o modelo representativo completo da série será composto por duas equações:

$$\begin{cases} y_t = s_t + z_t \\ \phi(B)z_t = \theta(B)\alpha_t \end{cases} \quad (3.10)$$

A previsão pode então ser feita por:

$$\hat{y}_{t+h} = \hat{s}_{t+h} + \hat{z}_{t+h} \quad (3.11)$$

onde:

- \hat{y}_{t+h} é a previsão de y_t para o instante $t+h$;
- \hat{s}_{t+h} é o valor determinístico da função s_t no instante $t+h$;
- \hat{z}_{t+h} é a previsão da componente estocástica z_t para o instante $t+h$ e é dada por $\hat{z}_{t+h} = E(z_{t+h} | I_t)$, I_t : conjunto das informações até o instante t .

BOX & JENKINS (1994) mostram uma alternativa para ajustar e remover a sazonalidade de séries de vazões:

$$\bar{z}_{t(r,m)} = \frac{z_{t(r,m)} - \mu_m}{\sigma_m} \quad (3.12)$$

onde:

- μ_m e σ_m são respectivamente a média e o desvio padrão do mês m ($m=1, 2, \dots, s$),
- r é o ano ($r = 1, 2, \dots, n$),
- n é o número de anos de vazões históricas e;
- $t_{(r, m)} = s(r-1) + m$.

A previsão, neste caso, para o tempo $t+h$ é dada por:

$$\hat{y}_{t(r, m)} = \mu_m + \sigma_m \hat{z}_{t(r, m)} \quad (3.13)$$

onde: $r=n+1, n+2, \dots$; $m=1, 2, \dots, 12$ e a previsão $\hat{z}_{t(r, m)+h}$ se obtém como antes, pelo valor esperado condicional $\hat{z}_{t(r, m)+h} = E(z_{t(r, m)+h} | I_t)$.

3.3.2 – Modelo com componente sazonal estocástico

De acordo com BOX & JENKINS (1994), nas séries de vazões, a natureza sazonal se deve às condições climáticas e às estações do ano. Dessa forma, supõe-se que uma observação para um determinado mês deve estar correlacionada com as observações do mesmo mês de anos anteriores e com as vazões de meses anteriores do mesmo ano. O modelo do tipo auto-regressivo de médias móveis sazonal (SARMA(p, q)(P, Q) $_s$) pode representar essa relação:

$$\Phi(B^s)\phi_p(B)y_t = \Theta(B^s)\theta_q(B)a_{t(r, m)} \quad (3.14)$$

onde: $\Phi(B^s)$ e $\Theta(B^s)$ são polinômios em B^s de graus P e Q , respectivamente e $\phi(B)$ e $\theta(B)$ são polinômios de graus p e q , respectivamente.

3.3.3 – Modelo com parâmetros periódicos

O modelo utilizado nesta pesquisa faz parte da classe de modelos auto-regressivos (AR), porém levando-se em consideração sua componente periódica. A

família de modelos autoregressivos periódicos - PAR(p_m) foi, originalmente, introduzida por THOMAS & FIERING (1962) para modelação e simulação de séries de vazões mensais. De acordo com McLEOD (1992), o modelo PAR de ordens p_m pode ser definido como:

$$Z_{t(r,m)} = \mu_m + \sum_{i=1}^{p_m} \phi_{i,m} (Z_{t(r,m)-i} - \mu_{m-i}) + \alpha_{t(r,m)} \quad (3.15)$$

onde : $t(r,m) \geq p_m + 1$

onde:

- Z_t são as observações consecutivas de uma série temporal periódica com período s ;
- $t (=1, \dots, ns)$ é o índice de tempo que pode ser escrito como uma função do número de anos r e o número de meses m , ou seja, $t = t(r,m) = (r-1)s + m$;
- s é o período (por exemplo, se $s = 12$, a série possui período de 12 meses);
- n é o comprimento da série;
- r é o conjunto de todos os anos da série ($r = 1, \dots, n$);
- m representa o número de meses do ano ($m = 1, \dots, s$).
- p_m é a ordem do modelo;
- $\alpha_{t(r,m)}$ é o ruído branco com $N(0, \tau_m^{-1})$ onde $\tau_m^{-1} = \sigma_m^2 = \tau_m > 0$

Se $\mu_m = E(Z_{t(r,m)})$ e $\gamma_{t(r,m)(j)} = cov(Z_{t(r,m)}, Z_{t(r,m)+j})$ existem e dependem somente de $t(r,m)$ e m , a série Z_t é considerada correlacionada periodicamente ou estacionária periódica segundo GLADYSEV² apud McLEOD (1992). O autor observou que se não forem dependentes de m , a série torna-se uma série temporal estacionária com covariância ordinária. Definidas a média $\mu_{t(r,m)}$ e autocovariância $\gamma_{t(r,m)}$, o autor afirma que o processo $\{Z_{t(r,m)}, r = 1, \dots, n; m = 1, \dots, s\}$ é periodicamente estacionário com período s se para dois números j e k quaisquer:

² GLADYSEV, E. G., (1961) Periodically correlated random sequences. *Sov. Math. Dokl.* 2, 385 - 388 apud McLEOD, A. I. (1992). Diagnostic checking of periodic autoregressive models with application. *Journal of Time Series Analysis* vol. 15. n^o2, p.221 - 233.

$$\mu_{t(r,m)} = \mu_{t(r,m)+ks} \quad (3.16)$$

$$\gamma_{t(r,m)}(j) = \gamma_{t(r,m)+ks}(j) \quad (3.17)$$

onde: $\gamma_{t(r,m)}(0) = \sigma_m^2$ é a variância do processo no mês m .

Utilizando as eq.(3.16) e (3.17) escreve-se a média e a função de autocovariância em termos de m para os seguintes casos:

i) $r = 1 \rightarrow t_{(1,m)} = m$ e $t_{(1,m)+ks} = m+ks$

$$\begin{aligned} \mu_m &= \mu_{m+ks} \\ \sigma_m^2 &= \sigma_{m+ks}^2 \\ \gamma_m(j) &= \gamma_{m+ks}(j) \end{aligned} \quad (3.18)$$

ii) $r \neq 1 \rightarrow t_{(r,m)} = (r-1)s+m+ks = m+k's$ e $k' = r+k+1$

$$\begin{aligned} \mu_m &= \mu_{m+k's} \\ \sigma_m^2 &= \sigma_{m+k's}^2 \\ \gamma_m(j) &= \gamma_{m+k's}(j) \end{aligned} \quad (3.19)$$

Os parâmetros que variam de forma periódica no tempo serão indexados apenas pelo índice m , ou seja, o parâmetro se refere ao mês do ano.

O modelo PAR(p_m) pode ser apresentado na forma de um modelo médias móveis (MA):

$$Z_{t(r,m)} = \mu_m + \sum_{i=0}^{\infty} \psi_{i,m} \alpha_{t(r,m)-i} \quad (3.20)$$

Onde os ψ 's podem ser calculados recursivamente usando a expressão:

$$\psi_{i,m} = \sum_{j=0}^{\infty} \phi_{j,m} \psi_{i-j,m-j} \quad i \geq 1 \quad (3.21)$$

$\psi_{0,m}=1$ e $\psi_{i,m}=0$ se $i \leq 0$. Uma condição necessária e suficiente para a estacionariedade periódica é:

$$\sum_{i=0}^{\infty} \psi_{i,m}^2 < \infty \quad m = 1, \dots, s \quad (3.22)$$

3.3.4 – Alguns exemplos do uso dos modelos estocásticos sazonais

➤ Modelos não-gaussianos multicomponentes (NGM) sua comparação com os modelos periódicos auto-regressivos (PAR) para vazões

O modelo não-gaussiano multicomponente (NGM) para vazões de VANDEWIELE & DOM (1989) foi modificado por ESHETE & VANDEWIELE (1992) a fim de facilitar a estimativa de máxima verossimilhança. O modelo também foi generalizado de modo que uma grande variedade de vazões em uma diversidade de passos no tempo pudesse ser modelado. A simulação é útil na análise de projetos de decisão, visto que previsões foram empregadas em casos de controle real no tempo. A escolha da base temporal do modelo (vazão anual, mensal, semanal), primeiramente, depende da aplicação em estudo (prevenção de enchentes, geração de energia elétrica, suprimento de água, melhoria das baixas vazões) e na área da bacia. Como uma consequência, é útil ter uma variedade de modelos disponíveis, capazes de encontrar uma variedade correspondente de exigências. Este modelo foi aplicado a duas bacias na Bélgica e na França com diferentes áreas, ambos em escala temporal mensal e semanal.

Os autores compararam a qualidade dos resultados da previsão e simulação (especialmente simulação dos volumes de altas e baixas vazões) com os modelos auto-regressivos periódicos (PAR), os quais possuem um bom desempenho na previsão, através do erro quadrático médio das previsões. Neste caso os autores perceberam que os modelos NGM geram resultados sempre melhores que os modelos PAR, isto ocorre pelo fato dos modelos NGM, explicitamente, considerarem à presença dos componentes da vazão tais como fluxo base e recessão direta da vazão, os quais são fenômenos bem conhecidos dos hidrólogos. O modelo PAR usa, exclusivamente, a auto-correlação

comum a qual é simétrica no tempo e assim incapaz de explicar a assimetria de repentinas variações do hidrograma.

Os modelos NGM não foram investigados nesta pesquisa pela ausência de referências bibliográficas que dessem suporte a uma investigação adequada.

➤ **Processos auto-regressivos periódicos para modelação de vazões mensais intermitentes**

Diversas tentativas foram feitas no passado para modelação de processos hidrológicos tais como vazões mensais em regiões secas. De acordo com CHEBAANE et al (1995), um dos problemas cruciais na modelação deste tipo de processo é o tratamento da ausência de vazões. As vazões mensais são, geralmente, processos nos quais a média, a variância, coeficiente de assimetria e correlação mês a mês variam através do ano. Além disso, vazões mensais em regiões áridas e semi-áridas podem ser intermitentes, por exemplo, o padrão de vazão é uma seqüência alternada de valores não-zeros e zeros. O processo de vazão não-zero/zero é altamente variável com o coeficiente elevado de variação e alto coeficiente de assimetria quando comparado a vazões mensais em regiões úmidas ou temperadas em rios perenes. Os modelos desenvolvidos para regiões úmidas e temperadas não são de fácil adaptação para regiões áridas, conseqüentemente, o desenvolvimento de modelos específicos para vazões mensais, segundo os autores torna-se essencial.

Um modelo estocástico foi apresentado por CHEBAANE et al (1995), o qual reproduz a percentagem de vazões nulas em cada mês, a média e a variância mensal, e a correlação mês a mês das vazões intermitentes. O modelo considera o processo de vazão intermitente mensal como um produto um processo binário discreto PDAR(1) e um processo periódico contínuo sendo ambos modelos auto-regressivos de primeira ordem. A estimativa dos parâmetros foi desenvolvida baseando-se no método dos momentos, método de probabilidade de transição e método da máxima verossimilhança. Os processos contínuos dependentes, os quais representaram as magnitudes das ocorrências não-zeros, foram descritos por um processo periódico de passo um. Para esta etapa, dois tipos de modelos auto-regressivos foram usados: modelo AR(1) log-normal, denotado como PAR_{LN} e o modelo periódico gama AR(1) denotado PGAR.

Ambos os modelos PDAR-PAR_{LN} e PDAR-PGAR justificam, explicitamente, a variação sazonal de vazões não-zeros e auto-covariância de passo um da ocorrência/não-ocorrência de vazões, bem como a média, variância, e auto-correlação de passo um para processos intermitentes de vazão. O modelo periódico discreto-contínuo proposto pelos autores pareceu ser útil para modelação de processos hidrológicos intermitentes sazonais dependentes.

➤ **Modelo para previsão de séries com características sazonais**

Os modelos para séries temporais do tipo auto-regressivo (AR) são bem conhecidos e considerados adequados para modelar séries estacionárias. Todavia, alguns conjuntos de dados não podem ser expressos por modelos AR estacionários. Em muitas situações práticas, encontramos séries temporais com um comportamento sazonal. Tal comportamento ocorre quando há similaridades na série após s períodos ou intervalos. HUTTER (1998) cita como exemplos de séries com características sazonais, a série formada pelo número de passageiros em uma companhia aérea internacional, o consumo de energia elétrica no setor industrial, a série de valores mensais de concentração de ozônio (em partes por milhão) e a série de vazões médias mensais (m^3/s) que chegam ao reservatório da Usina Hidroelétrica de Furnas.

Para as séries sazonais, onde o valor observado no instante t está correlacionado com os valores passados no instante $t-s$ é mais adequado ajustar um modelo cujos parâmetros também variam de uma forma periódica do que um modelo com parâmetros constantes para todos os meses. Estas séries são ditas periodicamente estacionárias quando a correlação entre os meses varia de forma periódica. Dentro da classe de modelos periódicos, a autora considera os modelo auto-regressivos periódicos (PAR). Um modelo PAR pode ser representado pela equação abaixo:

$$\tilde{Z}_{t(r,m)} = \sum_{i=1}^{p_m} \phi_{i,m} \tilde{Z}_{t(r,m)-i} + \alpha_{t(r,m)} \quad (3.23)$$

onde:

– $t(r,m) \geq p_m+1$

- $\tilde{Z}_{t(r,m)} = \frac{Z_{t(r,m)} - \mu_m}{\sigma_m}$ representa a série temporal $Z_{t(r,m)}$ padronizada;
- μ_m e $\phi_{i,m}$ são as funções periódicas com período s que representam respectivamente a média e os coeficientes do modelo para cada mês;
- $\alpha_{t(r,m)}$ é o ruído gaussiano com as seguintes propriedades:

$$\begin{aligned}
 E(\alpha_{t(r,m)}) &= 0 \\
 E(\alpha_{t(r,m)} - \alpha_{t(r,m)+j}) &= \begin{cases} \tau_m^{-1}, & j = 0 \\ 0, & j \neq 0 \end{cases} \\
 \text{onde } \tau_m^{-1} &= \sigma_m^2, \tau_m > 0
 \end{aligned} \tag{3.24}$$

HUTTER (1998) estabeleceu as metodologias clássica e bayesiana para análise de séries temporais periódicas. Na análise bayesiana, a autora, considerou três alternativas de funções de distribuição de densidades *a priori* para os parâmetros. Através da densidade preditiva o melhor modelo foi escolhido e previsões um passo à frente de valores futuros de séries foram realizadas, usando-se dados da simulação Monte Carlo em Cadeia Markov (MCMC).

As técnicas bayesianas foram empregadas como uma alternativa para reduzir o esforço computacional, no cálculo das estimativas de máxima verossimilhança, e tem por finalidade a obtenção de resultados mais precisos que a análise de verossimilhança, ou seja, obtenção de estimativas melhores, já que nos casos de séries temporais geralmente temos amostras pequenas para ajuste de modelos PAR.

O método proposto por HUTTER (1998) é ilustrado por duas séries estacionárias geradas e duas séries sazonais geradas. Os resultados apresentados mostraram que técnicas de simulação em MCMC foram bastante apropriadas para a análise de dados de séries temporais, quando se trabalha com séries periódicas onde muitos parâmetros estão envolvidos e as inferências usuais baseadas na estimação de máxima verossimilhança não são viáveis. Segundo a autora, os resultados obtidos com os estimadores de máxima verossimilhança se mostraram equivalentes aos obtidos pela simulação MCMC, além disso, o uso do algoritmo do amostrador de Gibbs e Metropolis-Hasting é de simples implementação e não exigem conhecimentos computacionais sofisticados.

3.4 – Métodos “clássicos” para determinação dos parâmetros do modelo

Existem diferentes métodos para estimar os parâmetros dos modelos para séries temporais, entre os métodos considerados clássicos estão o método dos momentos, o método dos mínimos quadrados e o método de máxima verossimilhança. Uma breve exposição desses métodos será apresentada, porém, os métodos utilizados nesta pesquisa são os métodos de máxima verossimilhança e bayesiano, este último será apresentado posteriormente.

3.4.1 – Método dos momentos

De acordo com OLIVEIRA (1998), a determinação das funções de autocorrelação (FAC) e auto correlação periódica (FACP) fornecem uma estimativa inicial dos parâmetros de um modelo AR(p) através da relação dessas funções e a equação de Yule-Walker, assim sendo os estimadores obtidos são denominados estimadores por método dos momentos. Segundo o autor este procedimento, geralmente, não apresenta boa precisão nas estimativas dos parâmetros ϕ 's quando o modelo está próximo das fronteiras da região de estacionariedade no espaço dos parâmetros (BOX et al., 1994). Considerando $Z=(Z_1, \dots, Z_N)$ um conjunto de observações de uma série temporal $\{Z_t, t \geq 0\}$ supostamente estacionária. A estimativa da FAC é dada por:

$$r_k = \frac{c_k}{c_0} \quad (3.25)$$

Onde:

$$c_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (Z_t - \hat{\mu})(Z_{t-k} - \hat{\mu}) \quad (3.26)$$

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^N Z_t \quad (3.27)$$

$$c_0 = \frac{1}{N} \sum_{t=1}^n (Z_t - \hat{\mu})^2 \quad (3.28)$$

A estimativa dos parâmetros pelo método dos momentos não é mais simples, segundo SALAS et al. (1988), do que a estimativa por outros métodos, entretanto, freqüentemente, as estimativas dos momentos são usadas como uma aproximação inicial para estimativa dos parâmetros por outros métodos.

3.4.2 – Método dos mínimos quadrados

Outro método usado para determinação dos parâmetros do modelo, segundo SALAS et al. (1988), é o método dos mínimos quadrados. Os autores descrevem o método da seguinte maneira: considerando que o modelo de uma série temporal $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N$ é $\bar{x} = f(\bar{x}_1, \bar{x}_2, \dots, \phi_1, \dots, \phi_p) + a_t$, onde ϕ_1, \dots, ϕ_p são os parâmetros do modelo e a_t é o resíduo ou erro da série o qual tem média zero. O método de estimativa dos mínimos quadrados é baseado na determinação das estimativas dos parâmetros $\hat{\phi}_1, \dots, \hat{\phi}_p$ de modo que a soma das diferenças dos quadrados entre os valores observados \bar{x}_t e calculados pelo modelo $\hat{x}_t = f(\bar{x}_{t-1}, \dots, \hat{\phi}_1, \dots, \hat{\phi}_p)$, $t=1, \dots, N$, respectivamente, é minimizado. Isso é, $\hat{\phi}_1, \dots, \hat{\phi}_p$ devem ser escolhidos de modo a minimizar o residual:

$$\sum_{t=1}^N a_t^2 = \sum_{t=1}^N (\bar{x}_t - \hat{x}_t)^2 = \sum_{t=1}^N [\bar{x}_t - f(\bar{x}_{t-1}, \dots, \hat{\phi}_1, \dots, \hat{\phi}_p)]^2 \quad (3.31)$$

Para encontrar o mínimo da soma na eq.(3.31) todas as derivadas parciais da soma em relação à $\hat{\phi}_1, \dots, \hat{\phi}_p$ devem ser iguais a zero. Isto é:

$$\frac{\partial \sum_{t=1}^N (\bar{x}_t - \hat{x}_t)^2}{\partial \hat{\phi}_1} = 0, \dots, \dots, \frac{\partial \sum_{t=1}^N (\bar{x}_t - \hat{x}_t)^2}{\partial \hat{\phi}_p} = 0 \quad (3.32)$$

Estas p equações com p incógnitas devem ser resolvidas simultaneamente para obtenção das estimativas dos parâmetros $\hat{\phi}_1, \dots, \hat{\phi}_p$ do modelo. Estas estimativas são eficientes quando certas condições são satisfeitas segundo YEVJEVICH (1972a).

Considerando o modelo da série temporal $x_t = \phi_1 \bar{x}_{t-1} + a_t$ onde a média de \bar{x}_t é zero, a média de a_t é também zero e a variância σ_a^2 de a_t é considerada conhecida. Pode-se determinar a estimativa dos mínimos quadrados do parâmetro ϕ_1 , baseando-se na série amostrada $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N$. A soma dos quadrados dos erros na eq.(3.32) é dada por:

$$\sum a_t^2 = \sum_{t=1}^N (\bar{x}_t - \hat{\phi}_1 \bar{x}_{t-1})^2 \quad (3.33)$$

E a derivada parcial da soma em relação à $\hat{\phi}_1$ é:

$$\frac{\partial}{\partial \hat{\phi}_1} \sum (\bar{x}_t - \hat{\phi}_1 \bar{x}_{t-1})^2 = \sum 2(\bar{x}_t - \hat{\phi}_1 \bar{x}_{t-1}) \bar{x}_t - (-\bar{x}_{t-1}) \quad (3.34)$$

Igualando a zero e resolvendo para $\hat{\phi}_1$ tem-se:

$$\hat{\phi}_1 = \frac{\sum \bar{x}_t \bar{x}_{t-1}}{\sum \bar{x}_{t-1}^2} \quad (3.35)$$

Observa-se que nas expressões acima o somatório varia de $t=2$ a $t=N$.

3.4.3 – Função máxima de verossimilhança

Considerando a observação z como o resultado de um experimento qualquer de uma variável Z que pode ser discreta ou contínua e é caracterizada por uma distribuição de probabilidade pertencente à família \mathcal{T} de distribuições, especificada com um número finito de parâmetros $\Phi = (\phi_1, \dots, \phi_p)$, geralmente, desconhecidos. Seja $f(z, \phi)$ a função de probabilidade ou densidade de Z dado ϕ e Φ (denominado de espaço paramétrico) é o conjunto dos valores possíveis para o vetor ϕ .

A teoria da probabilidade trata do problema da distribuição da variável Z quando esta é completamente definida, ou seja, todos os parâmetros ϕ são conhecidos. Entretanto, se a distribuição da variável Z não estiver completamente definida e supondo que a sua distribuição pertença à família \mathfrak{F} , o problema passa a ser de inferência estatística. O objetivo é, então, a partir das observações de z , ter alguma informação sobre a distribuição de Z ou, equivalentemente, sobre uma hipótese expressa através de ϕ .

A função de verossimilhança pode ser definida como sendo proporcional à distribuição de probabilidade conjunta das realizações de Z , a menos de uma constante de proporcionalidade positiva dependente dos dados de z , mas não do parâmetro ϕ . De forma geral, este parâmetro é considerado igual a 1. A verossimilhança é interpretada, como função do vetor de parâmetros para um conjunto fixo de dados, e tem por finalidade medir o quanto aqueles dados suportam uma hipótese sobre ϕ .

Obviamente, escolhe-se a família \mathfrak{F} antes de se trabalhar com a verossimilhança. Esta escolha pode ser feita por intuição, por princípios genéricos ou baseada em experimentações anteriores. Neste sentido, ela informa a ordem natural, de preferência, entre as diversas possibilidades de ϕ . É necessário, porém, salientar que não é possível comparar diferentes hipóteses sobre ϕ se elas forem baseadas em conjunto de dados distintos. A função de verossimilhança pode ser escrita de forma genérica da seguinte maneira:

$$L(\phi) = f(Z_1, \dots, Z_i; \phi) \quad (3.36)$$

Verossimilhança e probabilidade são conceitos fundamentalmente distintos, já que as hipóteses sobre ϕ não podem ser tratadas como se fossem resultados de experimentos. Nos conceitos de probabilidade, os resultados da experimentação são variáveis supondo a hipótese sobre ϕ constante, enquanto que para a verossimilhança os resultados do experimento (os dados) são fixados como constantes e as possibilidades de ϕ são as variáveis.

FISHER (1921) propôs a noção de verossimilhança como um método de estimação e como critério para comparar duas hipóteses rivais a serem testadas. O método de máxima verossimilhança não contradiz os fatos (representados pelos dados)

observados e permite escolher o valor do parâmetro que fornece uma chance mais provável para que os fatos ocorram novamente. Para a estimativa de ϕ_0 (valor verdadeiro do parâmetro) é necessário escolher o vetor de parâmetros que maximiza a função de verossimilhança no domínio Φ . Logo, a estimativa de máxima verossimilhança (EMV) de ϕ é o vetor $\hat{\phi}$ que maximiza o $L(\phi)$ em Φ . A EMV de ϕ desempenha um papel central na inferência paramétrica de grandes amostras. Geralmente, no processo de maximização, trabalha-se com o logaritmo natural de $L(\phi)$ dado por $\ell(\phi) = \log L(\phi)$. Para determinação dos estimadores de máxima verossimilhança (EMV) basta calcular as derivadas parciais de $\ell(\phi)$ em relação aos parâmetros ϕ_1, \dots, ϕ_p igualadas a zero são:

$$\frac{\partial(\ell(\phi))}{\partial\phi_1} = 0, \dots, \dots, \frac{\partial(\ell(\phi))}{\partial\phi_p} = 0 \quad (3.37)$$

Então, a solução da eq.(3.37) apresenta as estimativas de máxima verossimilhança para os parâmetros ϕ_1, \dots, ϕ_p . Os estimadores de máxima verossimilhança são assintoticamente eficientes.

3.4.3.1 – Exemplos do uso do método de máxima verossimilhança na determinação dos parâmetros do modelo

➤ Estimativa da máxima verossimilhança para modelos periódicos auto-regressivos de médias móveis

Segundo VECCHIA (1985) há muitas séries temporais periódicas que não podem ser filtradas ou padronizadas para obtenção da estacionariedade de 2ª ordem, devido ao fato da estrutura de correlação dessas séries dependerem do período. A classe de modelos considerada útil e analisada, pelo autor, é a classe de modelos periódicos

auto-regressivos de média móvel (PARMA) que são uma extensão dos modelos auto-regressivos de média móvel (ARMA) levando em conta os parâmetros periódicos.

A estimativa da máxima verossimilhança para o processo PARMA foi utilizada por VECCHIA (1985). Uma aproximação computacional simples para a função de verossimilhança exata para processos PARMA foi desenvolvida e um algoritmo foi apresentado para obtenção da estimativa aproximada da máxima verossimilhança dos parâmetros do modelo.

O modelo $PARMA_s(p, q)$ pode ser representado da seguinte forma:

$$z_{t(n,m)} - \sum_{k=1}^p \phi_{k(m)} z_{t(n,m)-k} = \alpha_{t(n,m)} - \sum_{k=1}^q \theta_{k(m)} z_{t(n,m)-k} \quad (3.38)$$

onde:

- Z_t representa a série de vazões;
- n ($n = 0, \pm 1, \dots$) representa o ano da série;
- m ($m = 1, 2, \dots, s$) representa o número do intervalo dentro do período s (por exemplo, meses do ano);
- p, q representam a ordem do modelo PARMA.

O modelo PARMA (3.38), também, pode ser expresso em termos de um modelo ARMA multivariado para estimativa da máxima verossimilhança segundo HILLMER & TIAO (1979). Entretanto, as dificuldades computacionais associadas com tal aproximação são excessivas quando comparadas com a aproximação usada pelo autor, especialmente, quando s é maior que p e q .

O primeiro passo na obtenção dos estimadores de máxima verossimilhança (EMV) dos parâmetros na eq.(3.38) é escrever a função de verossimilhança na forma que é facilmente avaliada. Fazendo $Z = [z_1, z_2, \dots, z_{ns}]'$, a função de verossimilhança de Z para o modelo apresentado na eq.(3.38) pode ser expressa como:

$$L(Z) = (2\pi)^{-\frac{ns}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} Z' \Sigma^{-1} Z\right\} \quad (3.39)$$

onde: $\Sigma = cov(Z)$ é uma função dos parâmetros sazonais. A verossimilhança dada na eq.(3.39) possui duas grandes dificuldades. Primeira, o cálculo de Σ é uma difícil tarefa, e segunda, a dimensão de Σ aumenta com n , fazendo o cálculo de seu inverso e de seu determinante ter um alto custo computacional. Para solucionar estes problemas o autor considera o caso de $p=0$ para a determinação da função de máxima verossimilhança exata. Quando calculada a verossimilhança aproximada para as últimas $n-p$ observações em Z considera-se $p>0$ e observam-se as primeiras p -observações como sendo fixas, dessa maneira a máxima verossimilhança é facilmente avaliada e interpretada.

O algoritmo proposto para a estimativa da máxima verossimilhança aproximada dos parâmetros do modelo é direto e facilmente executado. As simulações mostram que um melhor desempenho ocorreu para modelo PARMA(1, 1). VECCHIA (1985) sugere que mais pesquisas precisam ser realizadas antes de serem apresentadas conclusões gerais para modelos de ordem elevada. O algoritmo foi testado e seus resultados comparados com as estimativas dos momentos. A aplicabilidade da técnica é demonstrada através de uma série de vazões do Rio Caroni na Venezuela e os resultados encontrados por VECCHIA (1985) são sensivelmente melhores.

➤ **Aplicação da simulação Monte Carlo com cadeia Markov (MCMC) na estimação da máxima verossimilhança para processos AR(p) e MA(q)**

ANDRADE FILHO & MIZOI (1998) propuseram o uso de algoritmos de simulação Monte Carlo com Cadeia de Markov (MCMC), mais especificamente, o algoritmo Amostrador de Gibbs (CASELLA & GEORGE, 1992) e o algoritmo de Metropolis-Hasting (CHIB & GREENBERG, 1995) para o cálculo das estimativas dos parâmetros dos processos AR(p) e MA(q) pelo método de máxima verossimilhança. Os parâmetros dos modelos foram interpretados como variáveis aleatórias cuja densidade de probabilidade tem como núcleo a função de verossimilhança do modelo. Os algoritmos de simulação MCMC foram usados para gerar uma amostra dos parâmetros e, a estimativa de máxima verossimilhança desses parâmetros foi tomada como sendo a moda da amostra gerada.

As séries usadas foram geradas por modelos AR(2) e MA(2). Para os modelos AR(2), os resultados foram comparados com os obtidos através de métodos numéricos (estimativa dos mínimos quadrado, aproximação das estimativas de máxima verossimilhança (BOX et al., 1994)) e pelo método ARMLE (MILLER, 1995). Para os modelos MA(2), os métodos numéricos usados na comparação foram as estimativas de máxima verossimilhança condicional e máxima verossimilhança incondicional (BOX et al., 1994). ANDRADE FILHO & MIZOI (1998) mostraram que o uso dos algoritmos MCMC, para o cálculo dos EMV em modelos do tipo AR(p) é viável, pois dispensa as simplificações necessárias quando usados métodos numéricos. Já para os modelos do tipo MA(q), o cálculo das estimativas de máxima verossimilhança apresenta a vantagem de estimar de forma mais precisa os parâmetros do modelo. Quando comparados os resultados obtidos usando os algoritmos de simulação MCMC com aqueles obtidos por métodos numéricos, os resultados mostram que as estimativas via MCMC são próximas daquelas obtidas pelos outros métodos, no caso de modelos AR(p). Para os modelos MA(q), a diferença entre as estimativas do parâmetro θ mostraram-se um pouco maiores, mas com resultados muito satisfatórios. Entretanto, para os modelos MA(q), o custo computacional torna-se elevado quando usados os algoritmos MCMC devido à necessidade de se operar com matrizes muito grandes, fazendo com que os algoritmos fiquem mais lentos que os demais. Os autores, também, citam que as vantagens dos métodos numéricos diminuem à medida que a discretização do intervalo de variação dos parâmetros nestes métodos aumenta e esta discretização também afeta a precisão desses dados.

3.5 – Inferência bayesiana

3.5.1 - Natureza da inferência bayesiana

As opiniões a respeito do valor do teorema de Bayes, de acordo com BOX & TIAO (1992), como uma base para a inferência estatística tem oscilado entre aceitação e rejeição desde sua publicação em 1763. Durante muitos anos, quando se pensava que argumentos alternativos forneciam um suporte satisfatório para a inferência bayesiana, os resultados obtidos eram vistos, muitas vezes, com condescendência, como um

interessante, mas errôneo esforço para solucionar problemas importantes. Quando, posteriormente, encontrou-se argumentos, inicialmente, despercebidos acompanhando as alternativas, o interesse foi reacendido. O modo bayesiano de pensamento, em muitas ocasiões esquecido, ergueu-se, com vigor e de acordo com os autores, três fatores adicionais podem esclarecer esse reavivamento:

- i) Os trabalhos de vários autores, embora nem sempre direcionados a esse fim, ajudaram a esclarecer e superar algumas das dificuldades filosóficas e práticas;
- ii) Enquanto outras teorias de inferência produziam soluções viáveis em casos onde hipóteses especiais, tais como normalidade e independência de erros, podiam ser realizadas; em outros casos, particularmente, onde não existiam estatísticas suficientes, as soluções eram, freqüentemente, insatisfatórias e confusas. Embora seja verdadeiro que estas hipóteses especiais cubram um número de situações de interesse científico, seria inútil fingir que o conjunto dos problemas estatísticos cuja solução tenha sido ou viria a ser necessária ao pesquisador coincida com o conjunto dos problemas e desta maneira tratado de forma conveniente. A obtenção dos dados é, muitas vezes, dispendiosa quando comparada com a análise. É razoável, então, que esses dados sejam analisados sob diferentes pontos de vista. Na seleção dos pontos de vista, os métodos bayesianos permitem que uma ênfase maior seja dada ao interesse científico e uma ênfase menor à conveniência matemática;
- iii) As soluções viáveis baseadas nas hipóteses especiais tornaram-se populares por outra razão – elas eram fáceis de computar. Esta consideração, porém, tem menor força hoje em dia, pois a calculadora de mesa não é mais considerada como um “instrumento poderoso” na execução da análise estatística. Com o surgimento dos computadores, o tempo computacional gasto na busca de cada solução foi reduzido e um número maior de análises pode ser realizado.

Segundo BOX & TIAO (1992), a teoria da inferência bayesiana os parâmetros de um modelo são considerados variáveis aleatórias, para os quais temos alguma informação *a priori*, resumida na densidade de probabilidade *a priori*, a qual expressa a

incerteza sobre estes parâmetros. Essa distribuição desempenha um importante papel na análise bayesiana.

As aplicações do teorema de Bayes, discutida por BOX & TIAO (1992) são exemplos da inferência estatística, embora esta seja somente uma etapa de projeto e análise, usada na iteração investigatória, todavia é uma importante etapa.

Entre diferentes sistemas de inferência estatística existentes que derivam do teorema de Bayes, os autores acreditam que este teorema poderá ter propriedades as quais o farão, de maneira particularmente apropriada, tornar-se um importante instrumento na investigação científica.

3.5.2 – Análise bayesiana e a hidrologia

BERGER (1995) afirma que o precursor do uso da idéia bayesiana na hidrologia foi o pesquisador Jacques Bernier e é devido à importância de seu trabalho que o método bayesiano tem permeado as ciências hidrológicas.

Segundo o autor, o suporte para a idéia bayesiana é conceitual e prático: o procedimento bayesiano fornece uma estrutura coerente que facilita a análise de problemas de decisão sobre incertezas (BERGER, 1985). As críticas relativas aos métodos bayesianos estão centradas em três aspectos:

- **Computacionais:** A implementação do suporte bayesiano muitas vezes aponta para as dificuldades dos problemas computacionais. Como uma consequência, a implementação é, algumas vezes, questionada numa tentativa de chamar atenção para modelos muito simplificados ou para empreender uma análise não bayesiana. O desenvolvimento recente dos métodos MCMC e outras ferramentas computacionais bayesianas têm ultrapassado estas críticas, levando em consideração uma modelação mais realística (e tipicamente complexa);
- **Imprecisão:** É, freqüentemente, apontado que o suporte bayesiano demanda precisão excessiva nos julgamentos do decisor, particularmente, quanto à especificação da distribuição *a priori*. O autor recorda que o suporte bayesiano é normativo. Quando aplicado, este suporte serve como um guia para agir sob a incerteza. Portanto, em estágios anteriores de uma análise, o modelo pode não ser capaz de produzir

informações precisas (*prioris*, modelo, utilidade). Para esclarecer este fato, métodos bayesianos robustos têm sido desenvolvidos. Esses métodos consistem em empreender uma família de análises bayesianas e as suas conclusões em razões comuns. Se há também muita discrepância, as ferramentas bayesianas robustas sugerem como resolvê-las, guiando-as à elucidação;

- **Descritiva:** Algumas vezes, discute-se que decisores atuais não confirmam os postulados bayesianos. ALLAIS³ apud BERGER (1998) em seu trabalho mostra que muitos estudos experimentais têm apontado que muitos decisores violam os postulados bayesianos em tarefas desassistidas o que sugere fragilidade do procedimento bayesiano como uma teoria descritiva. Alguns autores interpretam e tratam o assunto com status normativo, embora tal interpretação não carregue nenhuma força lógica. Segundo RIOS INSUA⁴ apud BERGER (1998), em muitos casos, novas teorias têm surgido tentando aperfeiçoar, de modo geral, a análise bayesiana sob um ponto de vista descritivo.

3.5.3 – Operação bayesiana

Na operação bayesiana, de acordo com PEREIRA & VIANA (1982), a incerteza é descrita probabilisticamente. Dessa forma, a receita bayesiana é simples: para o cálculo de uma quantidade desconhecida ϕ de interesse (parâmetro), a partir de valores passados H , são calculadas as probabilidades condicionais dessa quantidade ϕ , por meio de regras de probabilidade. Observa-se que se H representa o conhecimento de ϕ até determinado momento, no instante seguinte deverá conter também as informações adicionais adquiridas naquele momento. O objetivo da inferência bayesiana é estudar as formas com as quais os resultados experimentais, adicionados a H alteram a probabilidade de ϕ , dado H . Para ilustrar a operação bayesiana, os autores propuseram o esquema abaixo (Figura 3.6):

³ALLAIS, M. (1953). Le comportement de l'homme rationel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica*, 21, p. 503-546 apud BERGER, J. O. (1998). *Recent developments in Bayesian inference with applications in hydrology*. In: *Statistica and bayesian methods in hydrological sciences*. Ed IHP-V Technical Documents in Hydrology n 20, UNESCO, Paris, Cap 3. p.43-61.

⁴RIOS INSUA, D. (1994). Ambiguity, imprecision and sensitivity in Decision Theory, in Puri and Vilaplana (eds). *New Progress in Probability and Statistics*, SVP apud BERGER, J. O. (1998). *Recent developments in Bayesian inference with applications in hydrology*. In: *Statistica and bayesian methods in hydrological sciences*. Ed IHP-V Technical Documents in Hydrology n 20, UNESCO, Paris, Cap 3. p.43-61.

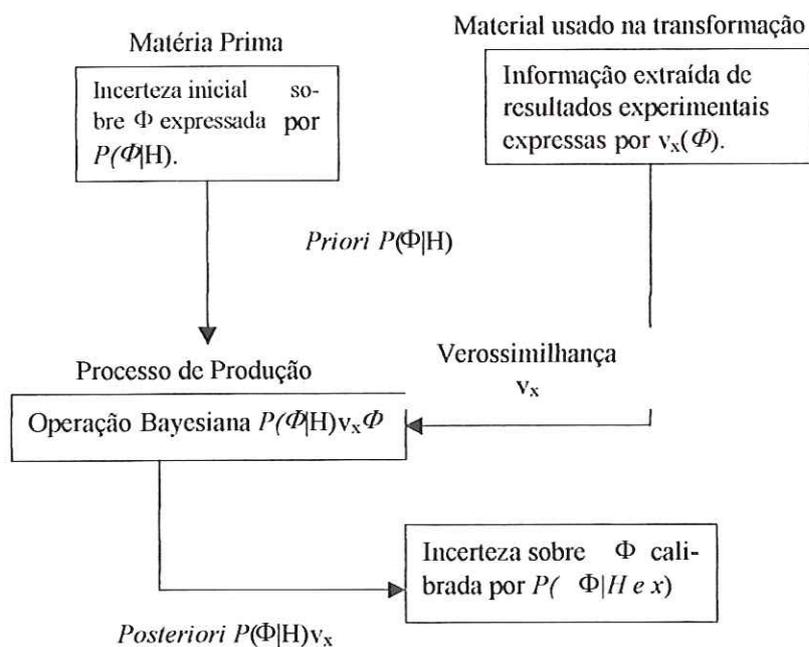


FIGURA 3.5 – Diagrama ilustrativo da operação bayesiana PEREIRA & VIANA (1982)

3.5.4 – Teorema de Bayes

É uma das relações mais utilizadas que faz uso da probabilidade condicional. Esse teorema afirma que:

$$P(E1_k|E2) = \frac{P(E2|E1_k)P(E1_k)}{\sum_{i=1}^N P(E2E1_i)P(E1_i)} \quad (3.40)$$

Onde a probabilidade do evento $E2$ pode ser calculada conhecendo-se as probabilidades condicionadas $P(E2|E1_i)$ e $P(E1_i)$ para $i=1, \dots, n$, como:

$$P(E2) = P\left(\bigcup_{i=1}^n E2 \cap E1_i\right) = \sum_{i=1}^n P(E2 \cap E1_i) \quad (3.41)$$

$$P(E2) = \sum_{i=1}^n P(E2|E1_i)P(E1_i)$$

3.5.5 – Seleção do modelo bayesiano

O procedimento bayesiano para seleção do modelo inicia-se pela designação das probabilidades para cada modelo a ser investigado, freqüentemente, são utilizadas probabilidades à *priori* iguais. É, também, necessário escolher distribuições a *priori* para os parâmetros desconhecidos de cada modelo. Tipicamente, seleciona-se o modelo ou modelos com a maior probabilidade a *posteriori*.

Segundo BERGER (1998), a mais evidente vantagem do procedimento bayesiano é a simplicidade da interpretação das respostas, até mesmo aquelas pessoas com conhecimento estatístico básico podem, facilmente, interpretar as conclusões das probabilidades a *posteriori*. A segunda vantagem é a consistência, de certo modo, quando se obtém mais e mais dados, está garantida a seleção do modelo mais apropriado ou do modelo próximo do modelo mais apropriado entre todos os escolhidos. Os métodos clássicos, tipicamente, falham neste critério mínimo, usualmente, pelos modelos selecionados que são muito complexos quando há uma grande quantidade de dados. A terceira vantagem, apresentada pelo autor, da seleção do modelo bayesiano é que se pode incluir a incerteza no modelo. Além das vantagens anteriormente apresentadas, ainda existe a possibilidade da seleção do modelo bayesiano de ser aplicado na comparação de diversos modelos e aplicado muito genericamente; estes modelos não precisam pertencer à família padrão e nem estar agrupados.

3.5.6 – Método bayesiano e técnicas de simulação

BERGER (1998) afirma que são duas as dificuldades na implementação do modelo bayesiano: a escolha da distribuição a *priori* e a computação do modelo escolhido. Porém, a escolha da distribuição a *priori* é considerada o maior problema.

Este pode, bem, ser o caso em que o conhecimento subjetivo sobre os parâmetros desconhecidos é avaliado e pode ser incorporado à subjetividade própria das densidades a *prioris* para estes parâmetros. Isto é, claramente, desejável se puder ser realizado. Freqüentemente, entretanto, os parâmetros podem, ser super dimensionados e a obtenção subjetiva de todas as *prioris* pode ser impossível.

3.5.6.1 – Computação bayesiana

Algumas ferramentas computacionais recentes têm permitido a aplicação de métodos bayesianos para modelos de alta complexidade e não padronizados. Na verdade, para modelos mais complicados, a análise bayesiana tenha, talvez, se tornado o mais simples, e freqüentemente o único, método de análise (BERGER, 1998).

O método de análise bayesiano possui algumas limitações nas aplicações quando se trata da resolução de integrais bayesianas para obtenção de distribuições marginais que são muitas vezes de difícil solução analítica e que acabam exigindo o uso de métodos de aproximação de integrais como, por exemplo, o método de Laplace. Recentemente, porém, têm sido utilizadas com sucesso as técnicas de simulação baseadas na técnica de simulação Monte Carlo em Cadeias de Markov (MCMC) como o Amostrador de Gibbs e o algoritmo de Metropolis-Hasting (M-H) (CHIB & GREENBERG, 1995). A convergência desses algoritmos pode ser verificada, informalmente, por meio de métodos gráficos, quantitativos ou qualitativos e, também, formalmente, por vários métodos propostos na literatura. O critério comumente usado é o critério proposto por GELMAN & RUBIN (1992).

3.5.7 – Exemplos do uso do método bayesiano em hidrologia

Nesta parte do trabalho, apresentam-se alguns exemplos do uso do método bayesiano aplicado à hidrologia de modo a mostrar a viabilidade de tal ferramenta nas ciências hidrológicas.

➤ Modelo bayesiano de processos de previsão de vazão

KRZYSZTOFOWICZ (1983), para encontrar decisões ótimas para previsões de vazão de rios ou para avaliar o desempenho econômico de tais previsões, desenvolveu um modelo do processo resposta-previsão de vazões. Um dos componentes desse modelo seguiu uma lei estocástica Λ que governa o estado de transição do processo de previsão de vazão. O processo Λ foi conceituado como uma cadeia markoviana discreta no tempo, finita, de duração aleatória, bivariada. As funções não-estacionárias de transição de probabilidade foram obtidas via procedimento bayesiano.

Os procedimentos de decisão bayesianos justificaram não somente a incerteza da previsão, mas em contraste com outros procedimentos, eles asseguraram um lucro na previsão temporal real não importando quão grande era a incerteza na previsão. Esta vantagem do procedimento bayesiano foi elucidada teoricamente e numericamente por KRZYSZTOFOWICZ (1983a). Do ponto de vista metodológico, o autor estende este estudo para um procedimento bayesiano de um processo markoviano com duas ramificações e embutidos em um algoritmo de programação dinâmica que soluciona um problema de decisão seqüencial. Além disso, oferece uma aproximação (também no espírito bayesiano) para funções de verossimilhança bivariada para torná-las mais amenas para avaliações práticas. Além da vantagem apresentada acima, o autor cita também como vantagens: primeiro, permite que se explore todos os registros de dados relevantes que foram tipicamente muito mais longos para o processo da inundação do que para o processo da previsão e segundo, realça a implementação do modelo para oferecer aproximações para as funções bivariadas de transição de probabilidade as quais requerem a estimativa de somente funções univariadas.

➤ **Método bayesiano para estimativa de vazões extremas em períodos de seca**

Um método para estimativa regional de vazões extremas em período de seca foi apresentado por MADSEN & ROSBJERG (1995). A modelação local foi baseada no método de séries de duração parcial (SDP) e assumiu-se que a ocorrência de secas poderia se descrita por meio de um processo de Poisson e a magnitude das secas e *déficit* de volumes seguem uma distribuição exponencial generalizada (DEG). Os autores relacionaram os parâmetros do modelo às características da bacia (fisiográficas, meteorológicas, entre outras) utilizando técnicas de regressão de mínimos quadrados que consideram os erros amostrais e dependência entre locais. As propriedades regionais dos parâmetros do modelo foram usadas como informações à *priori* no procedimento bayesiano que tornou possível a estimativa de eventos de T -anos em locais sem postos de medida. Já em locais com medição a procedimento bayesiano forneceu um esquema de atualização para combinar as informações à *priori* e as informações locais específicas.

O modelo de estimativa de vazões em períodos de seca foi aplicado a 11 (onze) conjuntos de dados de vazões do rio Danish com períodos de observação de 60 a 77

anos. MADSEN & ROSBJERG (1995) observaram que para a duração dos períodos de seca a DEG forneceu um ajuste apropriado, visto que uma ausência de ajuste é verificada na calda da distribuição para o déficit de volume em um número significativo de estações e concluíram que neste aspecto mais pesquisas precisam ser realizadas. Entretanto, algumas conclusões preliminares puderam ser esboçadas. A correlação interlocal observada foi relativamente grande implicando que um severo período de seca possui uma grande cobertura regional na Dinamarca. Para todos os parâmetros DEG, exceto o parâmetro de forma do déficit de volume, as equações de regressão regional foram formuladas. Para o parâmetro de forma do déficit de volume um modelo de médias regional que implica em um valor constante do parâmetro na região foi considerado adequado. A comparação entre os procedimentos de estimação dos eventos de T -anos revelou que o procedimento bayesiano regional foi superior à estimativa baseada somente nos dados locais. Além disso, a procedimento bayesiano fornece um procedimento de estimativas para locais sem medição os quais são muito importantes para a avaliação geral dos períodos de secas regionais.

➤ **Modelos auto-regressivos para variação de regimes de escoamento de LU & BERLINER (1999)**

LU & BERLINER (1999) consideraram uma classe de modelos dinâmicos bayesianos que envolvem variações entre regimes de escoamento. Este modelo é uma combinação de três modelos auto-regressivos, os quais são ajustados para os estados normal, de acréscimo e redução no processo de escoamento. O modelo para variação entre regimes é dado por uma cadeia de Markov em três estados, onde as probabilidades de transição são modeladas tendo por base os valores passados dos escoamentos e das séries de dados de alturas pluviométricas.

Os modelos formulados são analisados usando-se um procedimento bayesiano e o amostrador de Gibbs para a simulação Monte Carlo via Cadeia de Markov (MCMC). Neste exemplo, os autores, utilizaram três modelos auto-regressivos de ordem 1 (isto é, dado o regime, somente Y_{t-1} entra no modelo para Y_t), onde são definidas variáveis indicadoras $\{I_t\}$ para cada regime em estudo: para $I=0$ tem-se a representação para o regime normal, $I=1$ indica um sobressalto ou acréscimo no período de escoamento, e $I=2$ denota o período de restabelecimento ou diminuição do regime. Para o regime onde

há acréscimo do escoamento ($I=1$), os autores incluíram valores passados de X no modelo de regressão.

Os modelos markovianos para a variação entre os regimes de escoamento, utilizando-se de um procedimento bayesiano são dadas por:

$$p_{01}(t) = \Phi(\eta_{01} + \eta_{02}Y_{t-1} + \eta_{03}X_{t-1}) \quad (3.43)$$

$$p_{11}(t) = \Phi(\eta_{11} + \eta_{13}X_{t-1}) \quad (3.44)$$

$$p_{22}(t) = \Phi(\eta_{21} + \eta_{22}Y_{t-1}) \quad (3.45)$$

onde: $\Phi(z)$ denota a função de distribuição de uma variável aleatória normal padrão Z , isto é, $\Phi(z)$ é a probabilidade de Z ser menos que z . Especificações destes três valores de p completam o modelo devido à existência de zeros na Tabela 3.2. Os outros três valores são obtidos pela subtração, por exemplo, $p_{00}(t)=1-p_{01}(t)$.

No modelo markoviano para a variação entre regimes de escoamento, a probabilidade condicional para I_t depende somente dos valores passados de I_{t-1} . A probabilidade de se fazer a transição de um estado para outro é denotada pelos p_{ij} apresentada na Tabela 3.1. Observa-se que o padrão de zeros ($p_{02}=p_{10}=p_{21}=0$) que aparece na matriz de transição, neste exemplo, reflete as definições dos regimes realizadas por LU & BERLINER (1999).

TABELA 3.1 - Matriz das probabilidades de transição para o processo de Markov dos períodos indicados

I_{t-1}	I_t		
	0	1	2
0	p_{00}	p_{01}	0
1	0	p_{11}	p_{12}
2	p_{20}	0	p_{22}

Para os autores, uma importante aplicação de uma análise de série temporal é a previsão de valores futuros. No problema de escoamento apresentado por eles, foi desenvolvido um procedimento para incorporar novos dados de precipitação. De acordo com os autores, o modelo produziu resultados razoáveis.

➤ **Estimativa paramétrica via estatística bayesiana e simulação Monte Carlo de BARRETO & ANDRADE (2000)**

BARRETO & ANDRADE (2000) propuseram uma abordagem bayesiana e o uso de simulação Monte Carlo em Cadeias de Markov (MCMC) para a estimativa dos parâmetros de modelos estocásticos auto-regressivos de ordem p - AR(p), e compararam os resultados obtidos com aqueles calculados pelo método de máxima verossimilhança (MV). De acordo com os autores, a diferença entre os métodos é o tratamento dado ao parâmetro ϕ do modelo. No método de MV, o parâmetro é tratado como uma constante enquanto a abordagem bayesiana entende o parâmetro como uma variável aleatória que obedece a uma determinada distribuição de probabilidade. Conforme se viu nos itens anteriores, o processo de inferência bayesiana requer a especificação de uma densidade de probabilidade *a priori* que reflete o conhecimento prévio a respeito da distribuição do parâmetro em questão.

Inicialmente, foi determinada a ordem p do modelo AR que melhor representa a série temporal observada. Dentre os critérios mais utilizados para a seleção do modelo estão os critérios AIC (AKAIKE, 1974) e BIC (SCHWARZ, 1978), os métodos baseados na densidade preditiva *a posteriori* e utilizando o fator de Bayes. Os autores avaliaram os resultados por estes métodos e observaram o modelo AR(2) era o mais indicado. Em seguida, os autores utilizaram a simulação MCMC para estimar dos parâmetros do modelo, através da geração de uma amostra de onde se obteve inferências a cerca dos parâmetros, e analisar a robustez do estimador bayesiano. Os parâmetros calculados pelo método bayesiano e MV foram usados para se fazer a previsão de valores futuros, calculados 12 passos à frente.

BARRETO & ANDRADE (2000) escolheram a série de vazões médias mensais afluentes ao reservatório de Furnas no período de janeiro de 1960 a dezembro de 1990 para análise, na tentativa de melhorar a previsão e conseqüentemente, ajudar na solução

dos problemas de planejamento da operação do sistema energético visando às tomadas de decisão que resultem na operação ótima do sistema. Os resultados obtidos, pelos autores, mostraram que o desempenho da técnica bayesiana é ligeiramente melhor do que o dos métodos clássicos e que o uso da estatística bayesiana e das técnicas de simulação MCMC tornaram o processo de estimação paramétrica mais flexível e poderoso quando comparado com os métodos clássicos.

➤ **Procedimento estocástico para o problema de controle de vazão**

ANDRADE et al. (2001) propuseram a junção do modelo estocástico com conceitos de análise da confiabilidade para melhorar as estimativas para o volume de proteção que podem ser alocados no reservatório para controlar uma onda de cheia. Neste procedimento, a vazão de entrada que alcança o reservatório durante uma cheia foi considerada sendo uma carga, e a capacidade do reservatório para controlar esta cheia foi considerado como a resistência que o reservatório oferece contra a propagação da cheia. Neste estudo a carga e a resistência são modeladas como um processo de difusão estocástico, e o volume de proteção foi determinado via fórmula de Ito. Neste cenário, uma fórmula explícita para o risco de falha foi derivado.

A inferência do parâmetro foi realizada pelo procedimento bayesiano para uma versão discreta no tempo da carga, e as estimativas foram obtidas usando-se os algoritmos de simulação Monte-Carlo em Cadeia Markov (MCMC) e Amostrador de Gibbs cuja convergência foi monitorada usando-se o Critério de Convergência de Gelman-Rubin. Neste caso o número considerado de iterações foi suficiente para a convergência aproximada se a escala potencial estimada de redução for $\sqrt{\hat{R}} < 1,1$. Os estimadores de máxima verossimilhança foram usados na comparação. Os dados utilizados compreendem nove anos de vazões diárias de entrada durante os períodos de cheia (novembro a maio) que chegam à Hidroelétrica de Xavantes no sudeste do Brasil.

Quando comparados os métodos bayesiano e de máxima verossimilhança pode-se observar, segundo os autores, que mesmo sendo o volume de proteção alocado pelo método bayesiano tenha sido sensivelmente maior do que aquele alocado pela estimativa de máxima verossimilhança; o método bayesiano foi mais significativo para

o número de dias de falha, certamente devido ao excelente uso dos volumes utilizados obtidos por meio do método bayesiano.

3.6 – Métodos para seleção do modelo

A seleção da ordem do modelo é uma das etapas importantes para a análise das séries temporais, pois através dessa seleção pode-se escolher o modelo matemático mais parcimonioso para a série em estudo. A identificação da ordem do modelo mais adequado consiste na determinação do valor de p , ou seja, do número de termos a serem considerados em um modelo auto-regressivo. Caso haja a escolha de um valor pequeno de p , pode-se ter uma representação inadequada da série, por outro lado, a escolha de um valor elevado pode levar a um alto grau de complexidade do modelo, o que é desnecessário, além de que, em alguns casos, pode ocorrer uma representação inadequada para o conjunto de dados. Vários são os métodos apresentados na literatura, tanto para o enfoque clássico quanto para o bayesiano e um breve resumo destes métodos é apresentado a seguir.

3.6.1 - Determinação da função de autocorrelação parcial periódica – PeFACP

A investigação das propriedades seqüenciais de uma série pela análise da autocorrelação é uma técnica da estatística clássica. É usada para determinação da dependência linear entre valores sucessivos do processo em instantes distintos. A função de autocorrelação periódica (PeFAC) indica graficamente a presença de correlações periódicas entre os meses e pode ser apresentada da seguinte forma:

$$\rho_m(j) = \frac{\hat{\gamma}_m(j)}{(\hat{\gamma}_m(0)\hat{\gamma}_{m+j}(0))^{1/2}} \quad (3.46)$$

sendo $\hat{\gamma}_m$'s as covariâncias estimadas.

Os valores das PeFAC's podem ser apresentados em um gráfico que relaciona as PeFAC's e os passos significativos (lags).

Após detectar a presença da correlação periódica, um modelo $PAR(p_m)$ adequado pode ser selecionado pela função de autocorrelação parcial periódica (PeFACP) $\phi_{i,m}$, que determina graficamente a ordem do modelo ou por algum critério de informação.

No caso em que for utilizado a PeFACP para identificação da ordem p_m do modelo, a mesma é determinada pelo passo no qual a PeFACP amostrada é cortada. A equação geral representativa da PeFACP pode ser escrita como:

$$\rho_m(j) = \phi_{1,m}\rho_m(j-1) + \phi_{2,m}\rho_m(j-2) + \dots + \phi_{p_m,m}\rho_m(j-p_m) \quad j > 0 \quad (3.47)$$

Denota-se por $\phi_{i,m}$ o i -ésimo coeficiente de um modelo $PAR(p_m)$ para m -ésimo mês, de tal modo que $\phi_{p_m,m}$ seja o último coeficiente. A PeFACP será calculada pela eq.(3.47) fazendo-se $j=1, \dots, p_m$ para obtenção de um conjunto de equações denominado equações de Yule-Walker dadas por:

$$\begin{cases} \rho_m(1) = \phi_{1,m} + \phi_{2,m}\rho_m(1) + \dots + \phi_{p_m,m}\rho_m(p_m-1) \\ \rho_m(2) = \phi_{1,m}\rho_m(1) + \phi_{2,m} + \dots + \phi_{p_m,m}\rho_m(p_m-2) \\ \vdots \\ \rho_m(p_m) = \phi_{1,m}\rho_m(p_m-1) + \phi_{2,m}\rho_m(p_m-2) + \dots + \phi_{p_m,m} \end{cases} \quad (3.48)$$

Obtendo-se uma estimativa dos coeficientes $\phi_{i,m}$ do modelo $PAR(p_m)$ e substituindo as PeFAC $\rho_m(j)$ por suas estimativas $\gamma_m(j)$. A eq.(3.48) pode também ser escrita na forma matricial como segue:



$$\begin{bmatrix} 1 & \rho_m(1) & \cdots & \rho_m(p_m-1) \\ \rho_m(1) & 1 & \cdots & \rho_m(p_m-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_m(p_m-1) & \rho_m(p_m-2) & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{1,m} \\ \phi_{2,m} \\ \vdots \\ \phi_{p_m,m} \end{bmatrix} = \begin{bmatrix} \rho_m(1) \\ \rho_m(2) \\ \vdots \\ \rho_m(p_m) \end{bmatrix} \quad (3.49)$$

O parâmetro $\phi_{p_m,m}$ é denominado de função de autocorrelação parcial periódica. Na prática utilizam-se seus valores estimados.

A variância estimada para as PeFACP's é dada por:

$$\text{Var}(\hat{\phi}_{i,m}) \cong \frac{1}{n} \quad (3.50)$$

Para um valor de n grande e supondo que o processo é PAR(p_m), o parâmetro $\hat{\phi}_{i,m}$ possui distribuição aproximadamente normal com média zero e variância calculada pela eq.(4.13), de modo que o intervalo de confiança é dado por:

$$IC = \hat{\phi}_{i,m} \pm \frac{2}{\sqrt{n}} \quad (3.51)$$

3.6.2 – Critério do erro de predição finito - FPE

AKAIKE (1969) desenvolveu o Critério do Erro de Predição Finito (FPE) para seleção da ordem (identificação) de modelos auto-regressiva. Esse critério é definido como sendo a predição do erro de uma realização independente do modelo ajustado e é apresentado como:

$$FPE = \hat{\sigma}_k^2 \frac{n+k}{n-k} \quad (3.52)$$

onde $\hat{\sigma}_k^2$ é o estimador de máxima verossimilhança de σ^2 , n o tamanho da série e $k=1, 2, \dots, p$ onde k é o passo (lag) e p é a ordem do modelo. O menor valor para FPE indica a melhor ordem do modelo.

3.6.3 – Critério de informação de Akaike - AIC

Segundo McLEOD (1992), um critério bem conhecido é o Critério de Informação de Akaike (AIC) desenvolvido por AKAIKE (1974, 1977⁵). O critério AIC é dado por:

$$AIC = \ln(\hat{\sigma}^2) + \frac{2M}{n} + cte \quad (3.53)$$

onde: $M = p+1$ denota o número de parâmetros estimados no modelo incluindo o termo constante; $\hat{\sigma}^2$ denota a estimativa de σ^2 pela função de máxima verossimilhança, n é o tamanho da série.

Por este método, o modelo escolhido será aquele que minimizará a expressão (3.53).

3.6.4 – Critério de Hanna-Quinn – HQ

O critério AIC é um exemplo do critério informação generalizado (CIG) o qual minimiza a expressão $CIG = \log(\sigma^2) + Cp/T$. Para o critério AIC, $C=2$. Um estimador consistente pode ser obtido pela recolocação de C como sendo uma função de decréscimo de T tal que Cp/T tende a zero quando T tende ao infinito, onde T é o número de observações avaliadas. HANNAH & QUINN (1979) provaram que um limite menor na razão de acréscimo de Cp é $\log(T)$. O critério de Hanna-Quinn pode ser escrito como:

⁵ AKAIKE, H. (1977). In entropy maximization principle. Proceedings of the Symposium on Application of Statistics (ed. P. R. Krishnaiah). Amsterdam: North Holland apud McLEOD, A. I. (1992) Diagnostic checking of periodic autoregressive models with application. *Journal of Time Series Analysis* vol.15, nº2, p.221-233.

$$HQ = (\log T)^{\frac{2k}{T}} ESS \quad (3.54)$$

onde ESS é a soma residual dos quadrados.

3.6.5 – Critério de Schwarz

SCHWARZ (1978) forneceu um critério no qual é um procedimento para uma grande amostra para a razão de probabilidade *a posteriori* dos modelos considerados. O critério do autor é baseado na minimização de:

$$BIC = \log(\sigma^2) + \frac{(\log T)p}{T} \quad (3.55)$$

De acordo com os resultados de HANNA & QUINN (1979), este critério tem a propriedade de uma forte consistência. SNEEK⁶ apud ENGLE & BROWN (1986) verificou que o critério de Schwarz possui um desempenho melhor do que o critério AIC na seleção da ordem correta de um modelo ARMA, especialmente se houver uma grande diferença entre os modelos competidores.

3.6.6 – Critério de Shibata

SHIBATA (1981) desenvolveu um critério para a seleção das variáveis de regressão ótimas (assintoticamente) sob uma hipótese de que o número de variáveis possíveis é infinito ou aumenta com o tamanho da amostra. Esse critério pode ser escrito como:

$$SHIBATA = \left(1 + \frac{2p}{T}\right) ESS \quad (3.56)$$

⁶ SNEEK, J. M. (1984). Modelling procedures fore univariate economic time series. Free University Press, Amsterdam apud ENGLE, R. F.; BROWN, S. J. (1986). Model Selection for Forecasting. *Applied Mathematics and Computation* 20: 313-327.

3.6.7 – Densidade preditiva ordenada

A seleção do melhor modelo pode também ser realizada através do enfoque bayesiano, dentre as técnicas utilizadas está aquela que utiliza as densidades preditivas ordenadas (GELMAN et al., 1995). E consiste na determinação da densidade preditiva ordenada para uma observação Z_{n+p} dada a série $(Z_1, \dots, Z_m, \dots, Z_{n+p-1})$, o parâmetro ϕ e τ , é dada por:

$$C_k^{M_p} = f(Z_{n+k} | Z_{n+k-1}) \propto \int \int \tau^{\frac{1}{2}} \exp \left\{ -\frac{\tau}{2} \left(Z_{n+k} - \sum_{i=1}^p \phi_i Z_{n+k-1} \right)^2 \right\} f(\phi, \tau | Z) d\phi d\tau \quad (3.57)$$

onde $f(\phi, \tau | Z)$ é a densidade a *posteriori* para os parâmetros ϕ e τ , e M_p é o modelo que está sendo avaliado. Calcula-se $\hat{C}^{M_p} = \prod_{k=1}^K C_{M_p}^k$ para cada modelo M_p e seleciona-se o modelo que maximiza esse produto (ANDRADE & BARRETO, 1999).

3.6.8 – Fator de Bayes

Uma outra forma de determinação da ordem, segundo ANDRADE FILHO (2001), do modelo via teoria bayesiana é através do fator de Bayes que pode ser definido da seguinte maneira:

$$B_{ij} = \frac{p(Z | M_i)}{p(Z | M_j)} \quad (3.58)$$

sendo $p(Z | M_p)$ o fator de normalização dada por:

$$p(Z | M_p) = \int \int L(Z | \phi, \tau, M_p) p(\phi, \tau | M_p) d\phi d\tau \quad (3.59)$$

Onde $p(\phi, \tau | M_p)$ é a densidade a *priori* e $L(Z | \phi, \tau, M_p)$ é a função de verossimilhança.

Os valores para análise do fator de Bayes, segundo KASS & RAFTERY (1995), estão reunidos na Tabela 3.2, onde H_0 é a hipótese nula que afirma se um modelo é melhor que o modelo na hipótese alternativa H_1 e B_{10} , significa a razão entre o modelo H_1 e modelo H_0 (H_1/H_0).

TABELA 3.2 – Tabela de referência do fator de Bayes

Log 10 (B_{10})	B_{10}	Evidência contra H_0
0 a 0,5	1 a 3,2	Não-Significativos
0,5 a 1	3,2 a 10	Substancial
1 a 2	10 a 100	Forte
> 2	> 100	Decisivo

Nos capítulos seguintes serão apresentados a metodologia empregada nesta pesquisa e os resultados gerados por esta.

4 – METODOLOGIA

Neste capítulo é apresentada a metodologia usada na investigação dos modelos auto-regressivos periódicos (PAR) sob o enfoque clássico (método de verossimilhança) e bayesiano.

4.1 – Modelo periódico auto-regressivo – PAR

4.1.1 - Apresentação do modelo

A família de modelos autoregressivos periódicos - $PAR(p_m)$ escolhida para ser investigada foi introduzida por THOMAS & FIERING (1962) para modelação e simulação de séries de vazões mensais. A equação (3.15) representa o modelo PAR de ordem p_m .

4.1.2 – Identificação do modelo PAR

Segundo McLEOD (1991), uma forte componente sazonal bem como uma alta variabilidade sazonal dos dados são evidentes nas séries temporais, o que pode ser observado quando plotamos os dados de vazão versus período (meses). A presença de correlações periódicas nas séries temporais pode, freqüentemente, ser indicada pelo exame do gráfico de dispersão, onde os dados plotados mostram a correlação periódica. Os gráficos de dispersão serão apresentados da seguinte forma: janeiro versus fevereiro, fevereiro versus março,....., novembro versus dezembro. Dessa maneira pretende-se mostrar, graficamente, a existência de correlação periódica entre os meses.

4.1.2.1 - Determinação da função de autocorrelação – PeFAC

A investigação das propriedades seqüenciais de uma série pela análise da autocorrelação é uma técnica da estatística clássica. É usada para determinação da dependência linear entre valores sucessivos do processo em instantes distintos. A função de autocorrelação periódica (PeFAC) indica, graficamente, a presença de correlações periódicas entre os meses. Os valores das PeFAC's podem ser plotados em um gráfico que relaciona as PeFAC's e os passos significativos (lags). A análise dos gráficos das PeFAC's foi realizada para identificar o tipo de modelo que melhor representaria a série e, em seguida, selecionar o número de parâmetros que melhor se adequar a esse modelo usando os critérios de seleção do modelo. Quando observados os gráficos, pode-se notar que para os modelos do tipo auto-regressivo periódico, as PeFAC's possuem características específicas para este tipo de modelo, ou seja, possuem sua função de autocorrelação com extensão infinita, consistindo-se de quedas exponenciais e/ou ondas. Entretanto deve-se salientar que cada tipo de modelo possui a sua PeFAC com características específicas em função dos mesmos.

4.2 - Seleção do modelo

4.2.1 – Critérios de informação

A seleção de um modelo $PAR(p_m)$ adequado pode ser realizada pela função de autocorrelação parcial periódica (PeFACP) que identifica graficamente a ordem do modelo. Entretanto, optou-se por determinar a ordem do modelo pelos critérios de seleção apresentados nas equações (4.1) e (4.2), a partir daí, estimar os valores dos parâmetros pelos métodos de máxima verossimilhança e estimador bayesiano.

A identificação da ordem do modelo consiste na determinação do valor de p_m , ou seja, do número adequado de parâmetros a serem considerados em um modelo. Uma escolha inconveniente de p_m pode gerar uma representação inadequada da série ou um alto grau de complexidade do modelo, o que é desnecessário, além de ocorrer, em alguns casos, uma representação errônea do conjunto de dados. Os critérios de seleção do modelo vêm ajudar na escolha do modelo visando atender ao critério de parcimônia

de Box (BOX et al., 1994). Um modelo mais parcimonioso é aquele que tem o menor número possível de parâmetros e menor variância dos resíduos.

Além da PeFACP que determina a ordem do modelo a partir da interpretação de gráficos, uma outra maneira de selecionar o modelo mais parcimonioso é o uso dos critérios de informação, como, por exemplo, o Critério de Informação de Akaike (AIC) ou Critério de Informação Bayesiano (BIC), detalhados no Capítulo 3.

Os modelos que fornecem um menor valor para os critérios de informação são os preferidos e os valores AIC e BIC são comparados entre vários modelos como base para a seleção. Como o critério BIC impõe maior penalidade para o número de parâmetros do modelo estimado do que o AIC, o uso do menor valor de BIC para a seleção do modelo poderá resultar sempre na escolha do modelo no qual o número de parâmetros é menor que aquele escolhido pelo AIC. Uma desvantagem desta aproximação é que vários modelos deverão ser estimados pela função de máxima verossimilhança o que demanda elevado tempo computacional.

Para os modelos PAR, os critérios de seleção do modelo podem ser rescritos como:

$$AIC = \ln(\hat{\sigma}_m^2) + \frac{2M}{n_m} \quad (4.1)$$

onde: $\hat{\sigma}_m^2$ denota a estimativa de σ^2 pela função de máxima verossimilhança, $M=p_m+1$ é o número de parâmetros do modelo e n_m é o tamanho da série para cada mês.

Para os modelos PAR, o critério BIC pode se fatorado para obtenção de um critério separado para cada período. A ordem do modelo é escolhida a partir da análise desses critérios e uma ordem para a série é inferida. Assim:

$$BIC = \sum_{m=1}^s BIC_m \quad \text{onde} \quad BIC_m = -\ln \hat{\tau}_m^{-1} + M \frac{\ln(n_m)}{n_m} \quad (4.2)$$

onde: $\hat{\tau}_m$ denota a estimativa da variância, $M=p_m+1$ é o número de parâmetros do modelo, incluindo o termo constante, e n_m é o tamanho da série para cada mês.

4.3 - Estimativa dos parâmetros do modelo

Existem diferentes métodos para estimar os parâmetros dos modelos auto-regressivos, entre eles estão os métodos dos momentos, dos mínimos quadrados e de máxima verossimilhança (BOX et al., 1994). Atualmente, os métodos bayesianos tem sido investigados como uma alternativa na análise e inferência de modelos de séries temporais. Optou-se por usar, aqui, os métodos de máxima verossimilhança e bayesiano na análise e determinação dos parâmetros de séries de vazões mensais.

4.3.1 – Método clássico – Estimadores de máxima verossimilhança

4.3.1.1 – Determinação da função de verossimilhança

A família de distribuições escolhida é a distribuição normal, que é a mais importante da estatística, segundo CORDEIRO (1992), face às suas inúmeras aplicações e por constituir uma aproximação razoável para um grande número de distribuições de interesse, além de ser uma das distribuições mais fáceis de se trabalhar matematicamente.

Rescrevendo a eq.(3.15) para o modelo PAR em termos de $\alpha_{t(r,m)}$, tem-se:

$$\alpha_{t(r,m)} = Z_{t(r,m)} - \sum_{i=1}^{p_m} \phi_{i,m} Z_{t(r,m)-i}, \quad t(r,m) \geq p_m + 1 \quad (4.3)$$

onde: $Z_{t(r,m)} = \frac{z_{t(r,m)} - \mu_m}{\sigma_m^2}$
 $t = t(r, m) = (r - 1)s + m$

A função de densidade de probabilidade (fdp) conjunta dos $a_{i(r,m)}$ para cada mês é dada por:

$$\begin{aligned} f_m(a_{i(r,1)} | \Phi_1, \tau_1) &= \prod_{r=2}^n \tau_1^{1/2} \exp\left\{-\frac{\tau_1}{2} a_{i(r,1)}^2\right\} \\ f_m(a_{i(r,2)} | \Phi_2, \tau_2) &= \prod_{r=2}^n \tau_2^{1/2} \exp\left\{-\frac{\tau_2}{2} a_{i(r,2)}^2\right\} \\ &\vdots \\ f_m(a_{i(r,12)} | \Phi_{12}, \tau_{12}) &= \prod_{r=2}^n \tau_{12}^{1/2} \exp\left\{-\frac{\tau_{12}}{2} a_{i(r,12)}^2\right\} \end{aligned} \quad (4.4)$$

Pode-se escrever a eq.(4.4) de forma geral:

$$f_m(a_{i(r,m)} | \Phi_m, \tau_m) = \prod_{r=2}^n \tau_m^{1/2} \exp\left\{-\frac{\tau_m}{2} a_{i(r,m)}^2\right\} \quad (4.5)$$

A fdp pode ser escrita como uma função de Z :

$$L_m(Z_m | \Phi_m, \tau_m) = \tau_m^{(n-p_m)/2} \exp\left\{-\frac{\tau_m}{2} \left(\sum_{r=2}^n \left(Z_{i(r,m)} - \sum_{i=1}^{p_m} \phi_{i,m}(Z_{i(r,m)-i})\right)\right)^2\right\} \quad (4.6)$$

Escrevendo a eq.(4.6) na forma matricial tem-se:

$$L_m(\Phi_m, \tau_m | Z_m) \propto \left\{ \tau_m^{(n-p_m)/2} \exp\left(-\frac{\tau_m}{2} (Z_m - X_m \Phi_m)' (Z_m - X_m \Phi_m)\right) \right\} \quad (4.7)$$

sendo:

$$Z_m = \begin{bmatrix} Z_{s+m} \\ Z_{2s+m} \\ \vdots \\ Z_{(n-1)s+m} \end{bmatrix} \quad (4.8)$$

$$X_m = \begin{bmatrix} Z_{s+m-1} & Z_{s+m-2} & \cdots & Z_{s+m-1} \\ Z_{2s+m-1} & Z_{2s+m-2} & \cdots & Z_{2s+m} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{(n-1)s+m-1} & Z_{(n-1)s+m-2} & \cdots & Z_{(n-1)s+m-1} \end{bmatrix} \quad (4.9)$$

$$\Phi_m = \begin{bmatrix} \phi_{1,m} \\ \phi_{2,m} \\ \vdots \\ \phi_{p,m} \end{bmatrix} \quad (4.10)$$

Para o modelo PAR em estudo, a função de verossimilhança é o produto das funções (eq.4.7) para todos os meses e é apresentada em sua forma matricial:

$$L_m(\Phi_1, \dots, \Phi_{12}, \tau_1, \dots, \tau_{12} | Z_m) \propto \prod_{m=1}^{12} \left\{ \tau_m^{(n-p_m)/2} \exp\left(-\frac{\tau_m}{2} (Z_m - X_m \Phi_m)' (Z_m - X_m \Phi_m)\right) \right\} \quad (4.11)$$

sendo: $\Phi_1 = (\phi_{1,1}, \dots, \phi_{p_m})'$, \dots , $\Phi_{12} = (\phi_{12,1}, \dots, \phi_{p_{12},12})'$, o vetor de parâmetros associado aos parâmetros ϕ 's para cada mês.

Os cálculos serão feitos usando-se somente a eq.(4.7) pela possibilidade de assumir a independência entre parâmetros, o que resulta em se trabalhar com uma função para cada mês.

4.3.1.2 – Estimadores de máxima verossimilhança - EMV

Os valores que maximizam a eq.(4.7) são denominados de estimadores de máxima verossimilhança (EMV) e são obtidos usando-se o logaritmo natural da eq.(4.7) e as suas derivadas parciais em função dos parâmetros Φ e τ :

$$\ell_m(\Phi_m, \tau_m | Z_m) \propto \log \left(\tau_m^{(n_m - p_m)/2} \exp \left(-\frac{\tau_m}{2} (Z_m - X_m \Phi_m)' (Z_m - X_m \Phi_m) \right) \right) \quad (4.12)$$

Os EMV's $\hat{\Phi}_m$ e $\hat{\tau}_m$ podem ser determinados pelas expressões abaixo:

$$\hat{\Phi}_m = (X_m' X_m)^{-1} (X_m' Z_m) \quad (4.13)$$

$$\hat{\tau}_m = \frac{1}{n_m} (Z_m - X_m \hat{\Phi}_m)' (Z_m - X_m \hat{\Phi}_m) \quad (4.14)$$

A estimativa de Z é dada por: $\hat{Z}_m = X_m \hat{\Phi}_m$ (4.15)

4.3.2 – Método bayesiano

Nesta etapa apresenta-se a análise bayesiana para estimativa dos parâmetros dos modelos autoregressivos periódicos (PAR) para séries temporais hidrológicas. Na teoria da inferência bayesiana, os parâmetros de um modelo são considerados variáveis aleatórias, para os quais temos alguma informação *a priori*, resumida na densidade de probabilidade *a priori* que expressa a incerteza sobre estes parâmetros e a densidade *a posteriori* mostra a distribuição da variável aleatória ϕ após as variáveis terem sido observadas. A metodologia empregada para a análise bayesiana teve por base o trabalho de HUTTER (1998), entretanto algumas modificações foram necessárias para a determinação dos estimadores bayesianos.

4.3.2.1 – *Priori* não-informativa de Jeffreys

A inferência bayesiana para os modelos PAR inicia-se pela escolha da densidade *a priori* que pode ser definida como a densidade de probabilidade que permite incorporar conhecimento prévio sobre os parâmetros do modelo. Entretanto, quando não há informações sobre os parâmetros, utiliza-se uma *priori* não-informativa. Neste trabalho foi usada *a priori* não-informativa de Jeffreys (BOX & TIAO, 1992). Supondo a independência dos parâmetros, encontra-se uma densidade *a priori* conjunta dada por:

$$\pi_0(\Phi_m, \tau_m) \propto \frac{1}{\tau_m}, \tau_m > 0 \quad (4.16)$$

4.3.2.1.1 – Densidade *a posteriori* conjunta para os parâmetros ϕ_m e τ_m

A função densidade *a posteriori* $\pi(\Phi_m, \tau_m)$ é obtida pela combinação da eq.(4.6) e a eq.(4.16). A densidade *a posteriori* conjunta é dada por:

$$\pi(\Phi_m, \tau_m | Z_m) \propto \tau_m^{\left(\frac{n_m - p_m}{2}\right) - 1} \exp\left\{-\frac{\tau_m}{2} [(Z_m - X_m \Phi_m)' (Z_m - X_m \Phi_m)]\right\} \quad (4.17)$$

onde n_m é o número de anos para um determinado mês.

Integrando a eq.(4.17) em relação aos parâmetros Φ_m e τ_m obtém-se as densidades marginais e condicionais para esses parâmetros: onde $\pi_m(\Phi_m, \tau_m | Z_m)$ tem distribuição normal-gama, ou seja $\pi_P(\Phi_m | \tau_m, Z_m)$ possui distribuição Normal e $\pi(\tau_m | \hat{\Phi}_m, Z_m)$ possui distribuição Gama.

A densidade *a posteriori* para o parâmetro τ_m é dada por:

$$\pi(\tau_m | Z_m) = \int_{\phi_{1,m}} \dots \int_{\phi_{p_m,m}} \pi_{P,m}(\phi_m, \tau_m | Z_m) d\phi_{1,m} \dots d\phi_{p_m,m} \quad (4.18)$$

$$\pi(\tau_m | Z_m) \propto \tau_m^{\left(\frac{n_m - p_m}{2}\right) - 1} \int_{D(\Phi_m)} \exp\left\{-\frac{\tau_m}{2} [(Z_m - X_m \Phi_m)' (Z_m - X_m \Phi_m)]\right\} d\Phi \quad (4.19)$$

$$\begin{aligned} \pi(\tau_m | Z_m) \propto \tau_m^{\left(\frac{n_m - p_m}{2}\right) - 1} \exp\left\{-\frac{\tau_m}{2} [(Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m)]\right\} \times \\ \times \underbrace{\int_{D(\Phi_m)} \exp\left\{-\frac{\tau_m}{2} [(\Phi_m - \hat{\Phi}_m)' (X'_m - \hat{X}_m)(\Phi_m - \hat{\Phi}_m)]\right\} d\Phi_m}_{\frac{-p_m}{\tau_m^2}} \end{aligned} \quad (4.20)$$

$\frac{n_m - p_m}{2} - 1 - \frac{p_m}{2}$
 $\frac{n_m - p_m}{2} - 1$
 $\left(\frac{n_m - p_m}{2}\right) - 1$

Observa-se que o parâmetro τ_m possui uma densidade *a posterior* $Gama(n_m - p_m/2, b_m)$ onde $b_m = (Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m) / 2$.

O estimador bayesiano de τ_m é dado por:

$$\hat{\tau}_{B,m} = (n_m - p_m) [(Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m)]^{-1}, \quad n_m \gg 2p_m \quad (4.21)$$

$$\text{Var}(\tau_m | Z_m) = 2(n_m - p_m) [(Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m)]^{-2} \quad (4.22)$$

A densidade *a posteriori* para o parâmetro Φ_m é dada por:

$$\pi_m(\Phi_m | Z_m) = \int_0^{\infty} \Pi_m(\Phi_m, \tau_m | Z_m) d\tau_m \quad (4.23)$$

$$\begin{aligned} \pi_m(\Phi_m | Z_m) \propto \int_{\tau_m} \left(\frac{n_m - p_m}{2} \right)^{-1} & \left\{ \exp \left\{ -\frac{\tau_m}{2} \left[(\Phi_m - \hat{\Phi}_m)' (X_m' X_m) (\Phi_m - \hat{\Phi}_m) \right] \right\} + \right. \\ & \left. + \exp \left\{ \left[(Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m) \right] \right\} \right\} \end{aligned} \quad (4.24)$$

$$\text{onde: } \alpha_m = \frac{n_m - p_m}{2} \quad (4.25)$$

$$\begin{aligned} \beta_m = \frac{1}{2} \left[(\Phi_m - \hat{\Phi}_m)' (X_m' X_m) (\Phi_m - \hat{\Phi}_m) + (Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m) \right] \\ \left[(X_m' X_m) \left[(Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m) \right] \right]^{-1} \end{aligned} \quad (4.26)$$

$$\text{então: } \pi(\Phi_m | Z_m) \propto \left[(\Phi_m - \hat{\Phi}_m)' (\Phi_m - \hat{\Phi}_m) + (Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m) \right]^{-\alpha_m} \quad (4.27)$$

$$\pi(\Phi_m | Z_m) \propto \left[1 + \frac{(\Phi_m - \hat{\Phi}_m)' V_m (\Phi_m - \hat{\Phi}_m)}{\nu} \right]^{-\frac{1}{2}(\nu_n + p_n)} \quad (4.28)$$

$$\text{onde: } \nu_m = n_m - 2p_m \\ V_m = (n_m - 2p_m)$$

O parâmetro Φ_m possui densidade *a posteriori* marginal *t*-Student p_m -dimensional com ν_m graus de liberdade, vetor de localização e matriz de precisão V_m dados por:

$$\hat{\Phi}_{B,m} = E(\Phi_m | Z_m) = (X_m' X_m)^{-1} (X_m' Z_m) \quad (4.29)$$

$$\text{Var}(\Phi_m | Z_m) = \frac{\nu_m}{\nu_m - 2} V_m^{-1} = \frac{(Z_m - \hat{Z}_m)' (Z_m - \hat{Z}_m) (X_m' X_m)^{-1}}{n_m - p_m - 2} \quad (4.30)$$

Pode-se observar que o estimador bayesiano $\hat{\Phi}_{B,m}$ coincide com o EMV $\hat{\Phi}_m$, quando é usada uma *priori* não-informativa.

4.3.2.2 – Densidades *a priori* informativas – *priori* Normal-Gama

O grau de conhecimento a respeito dos parâmetros do modelo pode ser avaliado pelas funções de densidade *a priori* informativas. Considera-se, nesta pesquisa, duas densidades *a priori* informativa: uma densidade *a priori* Normal e densidade *a priori* Gama.

$$\pi_0(\Phi_m, \tau_m) = \pi_1(\Phi_m | \tau_m) \pi_2(\tau_m) \quad (4.31)$$

A distribuição π_1 é uma distribuição Normal:

$$\pi_1(\Phi_m | \tau_m) \propto \tau_m^{\frac{pm}{2}} \exp\left\{-\frac{\tau_m}{2} (\Phi_m - \mu_m)' P_m (\Phi_m - \mu_m)\right\} \quad (4.32)$$

sendo P_m a matriz de precisão e μ_m é o vetor de parâmetros iniciais calculado usando - se os dados da série.

A distribuição π_2 é uma distribuição Gama:

$$\pi_2(\tau_m) \propto \tau_m^{\alpha_m - 1} \exp(-B_m \tau_m) \quad (3.33)$$

Onde:

$$B_m(\Phi_m) = \left[\Phi_m - (X_m' X_m D_m)^{-1} (X_m' Z_m + P_m \mu_m) \right]^2 \times \\ \times (X_m' X_m + P_m)^{-1} (X_m' Z_m + P_m \mu_m) \quad (4.34)$$

A *priori* Normal-Gama $\pi_0(\Phi_m, \tau_m)$ pode ser rescrita da seguinte forma:

$$\pi_0(\Phi_m, \tau_m) \propto \tau_m^{\frac{pm}{2}} \exp\left\{-\frac{\tau_m}{2} (\Phi_m - \mu_m)' P_m (\Phi_m - \mu_m)\right\} \times \tau_m^{\alpha_m - 1} \exp\{B_m \tau_m\} \quad (4.35)$$

$$\pi_0(\Phi_m, \tau_m) \propto \tau_m^{\frac{pm}{2} + \alpha_m - 1} \exp\left\{-\frac{\tau_m}{2} (\Phi_m - \mu_m)' P_m (\Phi_m - \mu_m) + 2B_m \tau_m\right\} \quad (4.36)$$

4.3.2.2.1 – Densidades *a posteriori*

A densidade *a posteriori* de Φ_m e τ_m será obtida usando-se o teorema de Bayes e *a priori* Normal-Gama.

$$\begin{aligned} \pi_0(\Phi_m, \tau_m | Z_m) &\propto \tau_m^{\frac{n_m}{2}} \exp\left\{\frac{\tau_m}{2} (Z_m - X_m \Phi_m)' (Z_m - X_m \Phi_m)\right\} \times \\ &\times \tau_m^{\frac{p_m}{2}} \exp\left\{-\frac{\tau_m}{2} (\Phi_m - \mu_m)' P_m (\Phi_m - \mu_m)\right\} \times \tau_m^{\alpha_m - 1} \exp\{-\tau_m B_m\} \end{aligned} \quad (4.37)$$

$$\text{onde : } \begin{cases} \Phi_m \in \mathfrak{R}^{p_m} \\ \alpha_m > 0 \\ B_m > 0 \end{cases}$$

Resolvendo eq.(4.37) temos:

$$\begin{aligned} \pi(\Phi_m, \tau_m | Z_m) &\propto \tau_m^{\left(\frac{n_m + 2\alpha_m}{2}\right) - 1} \times \\ &\times \exp\left\{-\tau_m \left[B_m + \frac{Z_m' Z_m + \mu_m' P_m \mu_m - (X_m' Z_m + P_m \mu_m)' (X_m' Z_m P_m)^{-1} (X_m' Z_m P_m)}{2} \right]\right\} \times \\ &\times \left\{ \exp\left[-\frac{\tau_m}{2} \left(\Phi_m + X_m' X_m + \frac{(B_m + Z_m' Z_m + \mu_m' P_m \mu_m) (X_m' Z_m P_m)^{-1}}{2} \right) \right] - \right. \\ &\quad \left. - \exp\left[-\frac{\tau_m}{2} \left(\frac{(X_m' Z_m + P_m \mu_m)' (X_m' X_m P_m)^{-1} (X_m' Z_m + P_m \mu_m)}{2} \right) \right] \right\} \end{aligned} \quad (4.38)$$

Rearranjando os termos da eq.(4.37), pode-se escrever a densidade a *posteriori* da seguinte forma:

$$\pi(\Phi_m, \tau_m | Z_m) \propto \tau_m^{\left(\frac{n_m + 2\alpha_m}{2}\right) - 1} \exp\{\tau_m D_m\} \times \tau_m^{\frac{p_m}{2}} \exp\left\{-\frac{\tau_m}{2} B_m(\Phi_m)\right\} \quad (4.39)$$

onde :

$$D_m = B_m + \frac{(Z_m' Z_m + \mu_m' P_m \mu_m)}{2} - \frac{(X_m' Z_m + P_m \mu_m)(X_m' X_m P_m)^{-1}(X_m' Z_m + P_m \mu_m)}{2} \quad (4.40)$$

A densidade *a posteriori* conjunta do parâmetro τ_m é uma distribuição Gama:

$$\pi(\tau_m | Z_m) \propto \tau_m^{\left(\frac{n_m + 2\alpha_m}{2}\right) - 1} \exp\{-\tau_m D_m\} \times \tau_m^{\frac{p_m}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{\tau_m}{2} B_m(\Phi_m)\right\} d\Phi_m \quad (4.41)$$

$$\pi(\tau_m | Z_m) \propto \tau_m^{\left(\frac{n_m + 2\alpha_m}{2}\right) - 1 + \frac{p_m}{2}} \exp\{-\tau_m D_m\} \quad (4.42)$$

$$\pi(\tau_m | Z_m) \propto \text{Gama}\left(\frac{n_m + 2\alpha_m}{2}, D_m\right) \quad (4.43)$$

Integrando a eq.(4.38) em relação à τ_m , obtemos a *posteriori* marginal de Φ_m :

$$\pi(\Phi_m | Z_m) \propto \int_0^\infty \tau_m^{\left(\frac{n_m+2\alpha_m}{2}\right)-1+\frac{p_m}{2}} \exp\left\{-\frac{\tau_m}{2}(2D_m + B_m(\Phi_m))\right\} d\tau_m \quad (4.44)$$

onde :

$$\xi_m = \frac{n_m - 2\alpha_m + p_m}{2}$$

$$\eta_m(\Phi_m) = 2D_m + B_m(\Phi_m)$$

Considerando apenas os termos dependentes de Φ_m temos:

$$\pi(\Phi_m | Z_m) \propto (2D_m + B_m(\Phi_m))^{-\xi_m} \quad (4.45)$$

A densidade $\pi_m(\Phi_m | Z_m)$ é uma função *t*-Student com $(n_m+2\alpha_m)$ graus de liberdade.

O estimador bayesiano Φ_m é dado por:

$$E(\Phi_m | Z_m) = (X_m' X_m + P_m)^{-1} (X_m' Z_m + P_m \mu_m) \quad (4.46)$$

$$\hat{\Phi}_{Bm} = E(\Phi_m | Z_m) = (X_m' X_m + P_m)^{-1} (X_m' Z_m + P_m \mu_m) \quad (4.47)$$

Integrando a eq.(4.38) em relação à Φ_m , obtemos a *posteriori* marginal de τ_m :

$$\pi(\tau_m | Z_m) \propto \tau_m^{\left(\frac{n_m - p_m + 2\alpha_m}{2}\right) - 1} \exp(-\tau_m D_m) \times \int_{\phi_m} \tau^{\frac{p_m}{2}} \exp\left\{-\frac{\tau}{2}(\phi_m - \hat{\phi}_m)' V_m (\phi_m - \hat{\phi}_m)\right\} d\phi_m \quad (4.48)$$

O estimador bayesiano τ_m é dado por:

$$\hat{\tau}_m = E(\tau_m | Z_m) = (n_m - p_m + 2\alpha_m)(2D_m)^{-1} \quad (4.49)$$

$$Var(\tau_m | Z_m) = E(\tau_m | Z_m)(D_m)^{-1} \quad (4.50)$$

4.3.2.2 – Determinação de P_m e μ_m

A matriz de precisão P_m usada nesta pesquisa foi a matriz identidade, já o valor do vetor de parâmetros iniciais μ_m foi determinado ajustando-se as séries em análise segundo um modelo do tipo auto-regressivo (AR(p)), porém, considerando-se as ordens determinadas para cada mês, de acordo com McLEOD (1992).

➤ Apresentação do modelo AR(p)

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \dots + \phi_p Z_{t-p} + a_t, \quad a_t \approx N(0, \sigma^2) \quad (4.51)$$

Onde: - Z_t são as observações consecutivas de uma série temporal;

- ϕ_i são os parâmetros do modelo;

- a_t é o ruído branco $N \approx (0, \sigma^2)$.

➤ **Identificação de um modelo AR(p)**

A identificação de um modelo é feita pela análise de sua função de autocorrelação (FAC). Em um modelo AR tem-se que sua função de autocorrelação tem por características: extensão infinita consistindo-se de quedas exponenciais e/ou ondas; já a função de autocorrelação parcial (FACP) possui extensão finita, com picos no passo (lag) um e sendo cortada no passo p .

➤ **Determinação da ordem do modelo**

A ordem do modelo foi escolhida usando-se os critérios AIC e BIC, apresentados no Capítulo 3, seus valores foram calculados para ordem 1 até ordem 10 e o modelo AR(p) ajustado para cada uma das ordens.

➤ **Determinação dos parâmetros do modelo AR usando EMV**

Os parâmetros do modelo AR foram calculados pelo estimador de máxima verossimilhança ajustando cada série para cada uma das ordens mensais determinadas anteriormente. Estes parâmetros foram usados como o conjunto inicial de parâmetros μ_m para a determinação dos estimadores bayesianos exatos – EBEx do modelo PAR(p_m). Para o modelo AR, a função de verossimilhança $L(Z|\phi)$ onde Z representa a série padronizada e $\Phi=(\phi_1, \phi_2, \dots, \phi_p, \tau)$ é o vetor de parâmetros desconhecidos do modelo, é dada pela seguinte equação:

$$L(Z|\phi, \tau) \propto \tau^{\frac{n-p}{2}} \exp\left\{-\frac{\tau}{2}\left[(\phi - \hat{\phi})X'X(\phi - \hat{\phi}) + (Z - \hat{Z})(Z - \hat{Z})\right]\right\} \quad (4.52)$$

Os estimadores de máxima verossimilhança (EMV) são obtidos pelas derivadas parciais do logaritmo natural da eq.(4.52):

Os EMV's $\hat{\phi}$ e $\hat{\tau}$ podem ser determinados pelas expressões abaixo:

$$\hat{\phi} = (X'X)^{-1}(X'Z) \quad (4.53)$$

$$\hat{\tau} = \frac{1}{n}(Z - X\hat{\phi})(Z - X\hat{\phi}) \quad (4.54)$$

A estimativa de Z é dada por: $\hat{Z} = X\hat{\phi}$ (4.55)

Onde: $Z = \begin{pmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_N \end{pmatrix}$ e $X = \begin{pmatrix} Z_1 & \dots & Z_p \\ \vdots & \vdots & \vdots \\ Z_{N-p} & \dots & Z_N \end{pmatrix}$ (4.56)

4.4 – Aplicação

A verificação do uso de estimadores bayesianos para determinação dos parâmetros do modelo foi realizada comparando-se a geração de dados da série para 12 meses para estes estimadores e os EMV. Para tal foram usadas séries hidrológicas de vazões médias mensais dos reservatórios das Usinas Hidrelétricas de Furnas e Emborcação que fazem parte do Sistema Sudeste Brasileiro de geração de energia elétrica. Os resultados foram comparados como aqueles encontrados usando-se os modelos AR com abordagem bayesiana e exata.

4.4.1 – Escolha das séries hidrológicas

Para se chegar aos objetivos propostos no Capítulo 2, foi necessária a escolha das séries hidrológicas tendo em vista o desafio de se obter previsões de boa qualidade em vários setores. As séries escolhidas foram as séries hidrológicas de vazões médias mensais das usinas hidroelétricas para produção de energia devido à necessidade de um planejamento da operação do sistema hidrotérmico de geração de energia elétrica com a finalidade de se determinar uma política de operação para as usinas hidroelétricas e termoelétricas de maneira que a demanda requerida de energia seja atingida de forma econômica e confiável. O problema da operação do sistema consiste na quantificação da disponibilidade de água para geração hidroelétrica (de custo relativamente baixo) na tentativa de se reduzir ao máximo a geração termoelétrica (de custo operacional elevado) (POLTRONIERE SILVA, 1999).

As séries hidrológicas escolhidas são as séries de vazões médias mensais para um período de 60 anos das Usinas Hidrelétricas de Furnas e Emborcação localizadas no Rio Grande e no Rio Paranaíba, respectivamente (Figura 4.1).

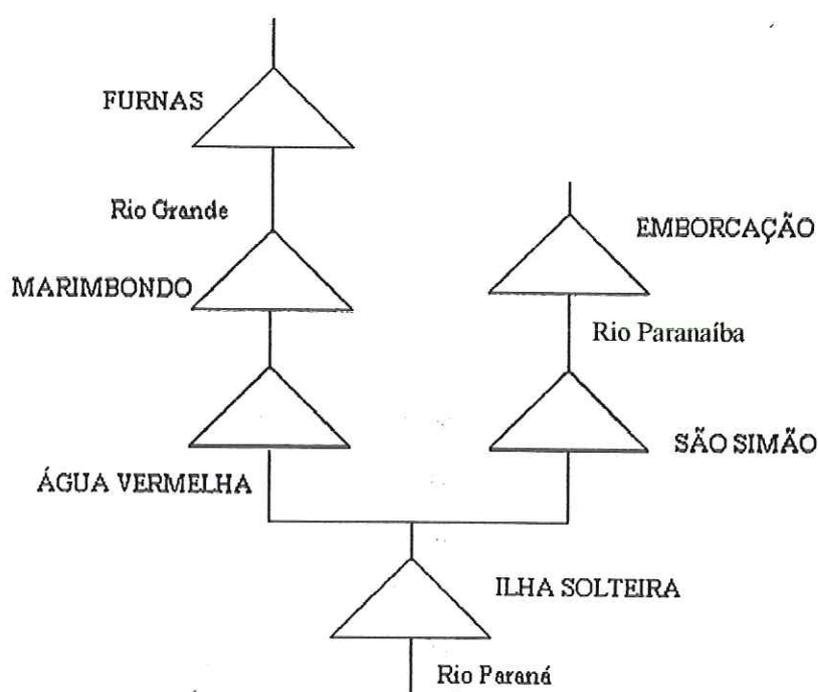


FIGURA 4.1 – Parte do Sistema Sudeste Brasileiro (POLTRONIERE SILVA, 1999).

4.5 – Cálculo dos erros

Após a determinação dos parâmetros do modelo pelos EMV e bayesianos dos modelos PAR, a geração de dados para 12 meses foi realizada e comparada por meio de uma avaliação dos erros. As medidas de avaliação dos erros, de acordo com ANDRADE FILHO (2001), mais usadas são: Raiz do Erro Quadrático Médio (REQM), Erro Absoluto Percentual Médio (EAPM) e Erro Absoluto Percentual Máximo (EAPMax). Os erros foram calculados usando-se as equações abaixo:

$$REQM = \sqrt{\frac{1}{N} \sum_{k=1}^N [(Z_{obs(t+k)} - Z_{prev(t+k)})]^2} \quad (4.57)$$

$$EAPM(\%) = \frac{100}{N} \sum_{k=1}^N \left| \frac{Z_{obs(t+k)} - Z_{prev(t+k)}}{Z_{obs(t+k)}} \right| \quad (4.58)$$

$$EAPMax(\%) = 100 \max_{k=1, \dots, N} \left\{ \left| \frac{Z_{obs(t+k)} - Z_{prev(t+k)}}{Z_{obs(t+k)}} \right| \right\} \quad (4.59)$$

4.6 – Desenvolvimento dos programas computacionais

Os programas computacionais para determinação paramétrica (EMV e EBEx), geração de dados e cálculo dos erros foram desenvolvidos com o auxílio do software para simulação numérica MATLAB 6.0.

5 – RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os resultados obtidos pelo método de verossimilhança e bayesiano para determinação dos parâmetros, além da previsão para 12 meses usando-se os valores históricos das séries hidrológicas das Usinas Hidroelétricas de Furnas e Emborcação.

5.1 – Modelo periódico auto-regressivo – PAR

5.1.1 – Apresentação das séries hidrológicas

Os conjuntos de dados utilizados para ilustrar o uso dos métodos aqui propostos foram os de vazões médias mensais das Usinas Hidroelétricas de Furnas e Emborcação, que compõe o Sistema Sudeste Brasileiro de Geração de Energia, tendo em vista o desafio de se obter previsões de boa qualidade em vários setores, inclusive no setor de geração de energia. As Figuras 5.1 e 5.2 mostram as séries históricas de Furnas e Emborcação, respectivamente, para um período de 60 anos.

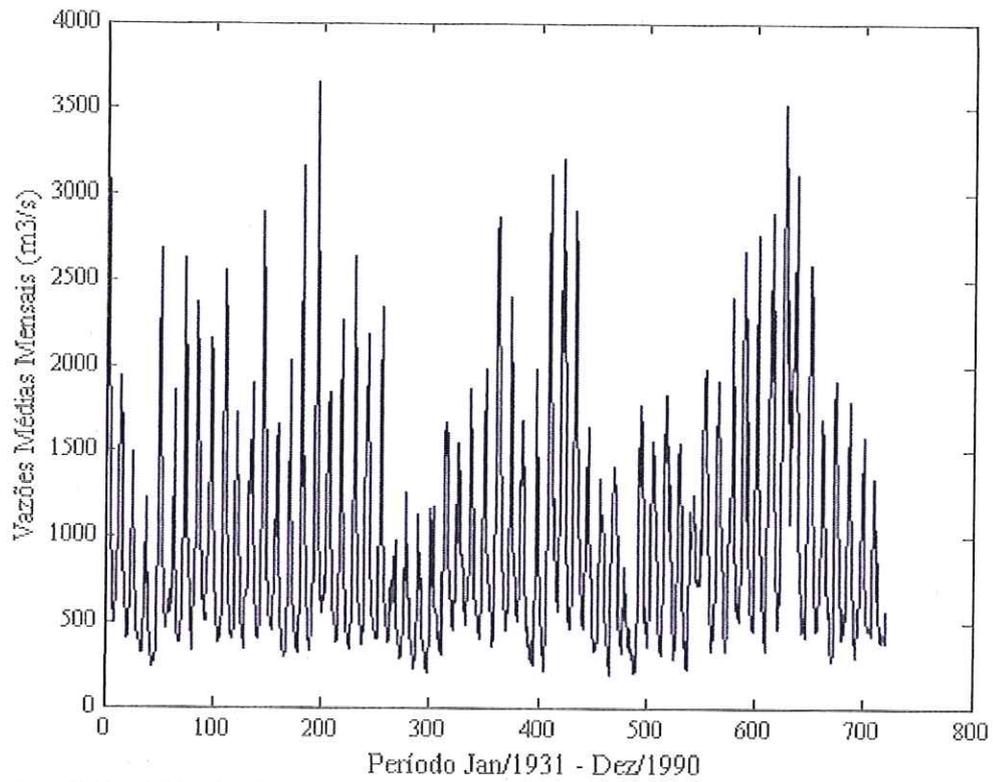


FIGURA 5.1: Série hidrológica de vazões médias mensais para a Usina Hidrelétrica de Furnas no período de janeiro de 1931 a dezembro de 1990.

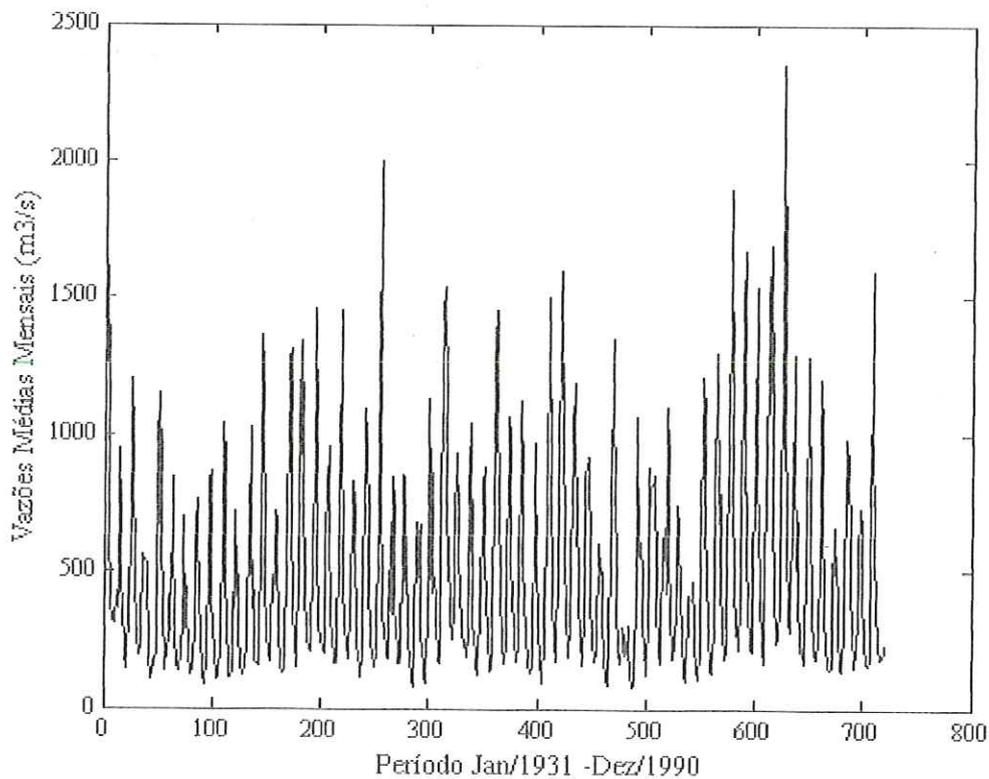


FIGURA 5.2: Série hidrológica de vazões médias mensais para a Usina Hidrelétrica de Emborcação no período de janeiro de 1931 a dezembro de 1990.

5.1.2 – Identificação do modelo PAR

De acordo com McLEOD (1991), a presença de correlação periódica pode ser indicada pelo exame de dispersão apresentado em gráficos que mostram a correlação periódica entre os meses. As Figuras 5.3 e 5.4 apresentam os gráficos de dispersão para as séries em estudo e pode-se observar a correlação periódica entre os meses do ano para cada série. Observando as figuras pode-se ver que existe uma maior correlação periódica entre os meses de estiagem, enquanto que para os meses de chuva, apesar de também existir uma correlação, ela não é tão explícita. Uma outra forma de identificação da correlação periódica foi realizada pela análise da função de autocorrelação periódica (PeFAC). A investigação das propriedades seqüenciais de uma série pela análise da autocorrelação é uma técnica da estatística clássica e é usada para determinar a dependência linear entre os valores sucessivos do processo em instantes distintos. Pela análise da PeFAC, pode-se verificar, também, o tipo de modelo que melhor se ajusta ao conjunto de dados. Nas séries aqui utilizadas, o modelo mais adequado é um modelo do tipo auto-regressivo (AR). A Figura (5.5) apresenta os gráficos das FACP's para as séries dessazonalizadas em estudo. Observando os gráficos apresentados percebe-se que as PeFAC's para as séries investigadas apresentam as características próprias de um modelo AR, ou seja, possuem sua função de autocorrelação com extensão infinita, consistindo-se de quedas exponenciais e/ou ondas.

Sabendo-se que as séries em análise são adequadas a um modelo do tipo AR, adotou-se este tipo de modelo para cada mês das séries investigadas. Sendo assim, tem-se um modelo auto-regressivo periódico (PAR) e o número de parâmetros determinado para cada mês da série.

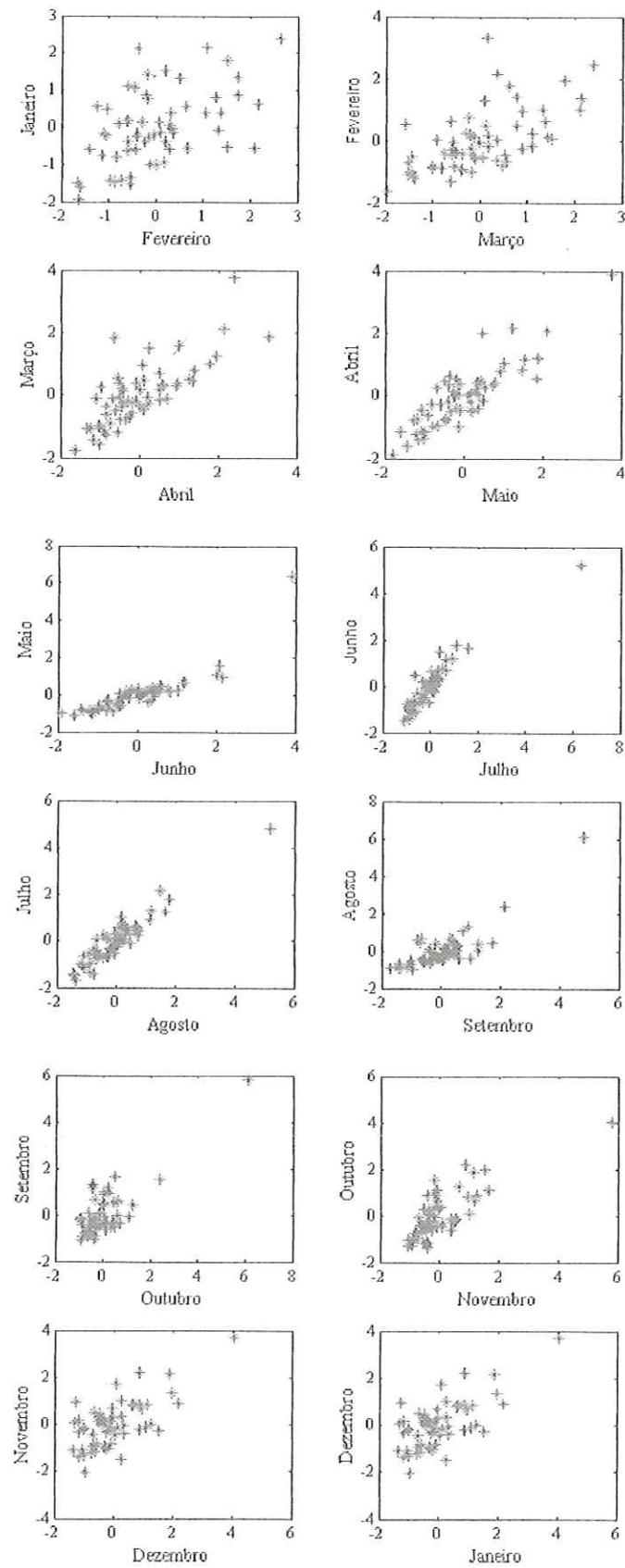


FIGURA 5.3: Gráficos da dispersão para a série de Furnas

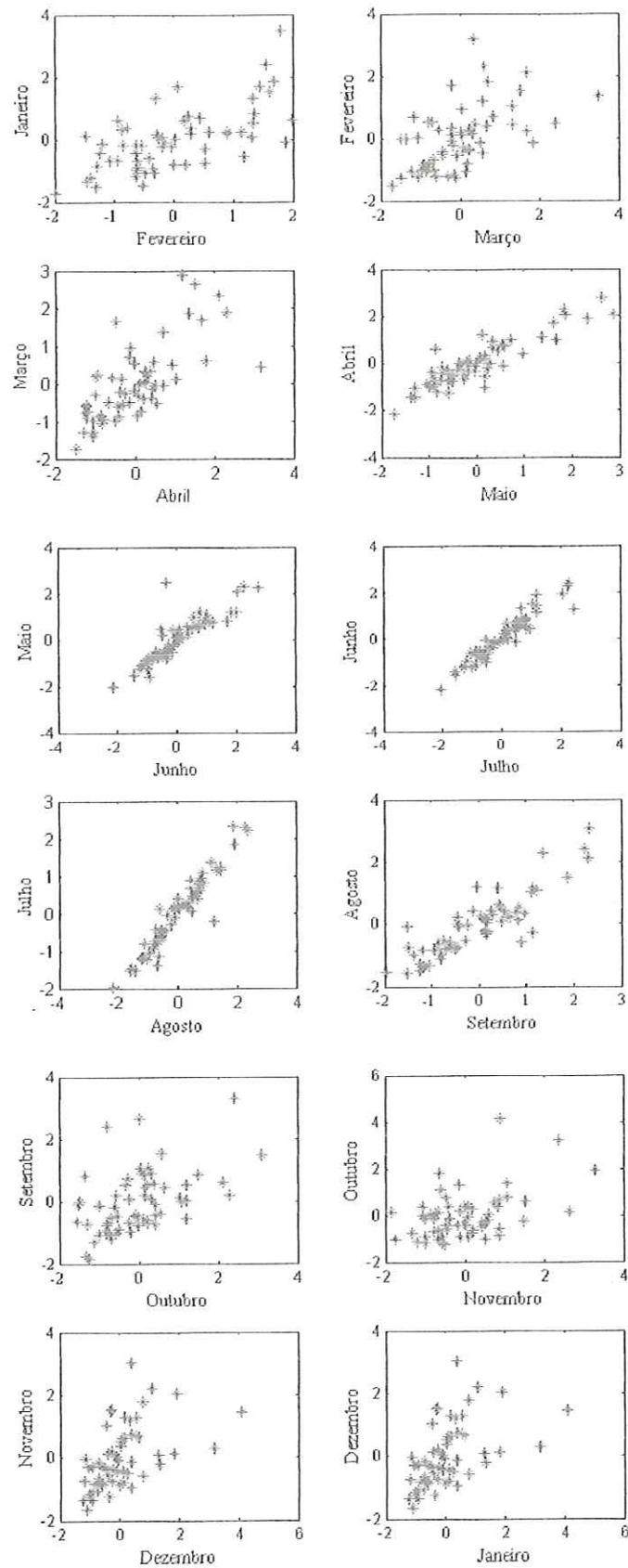
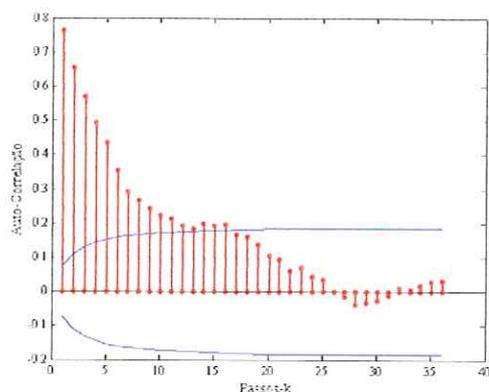
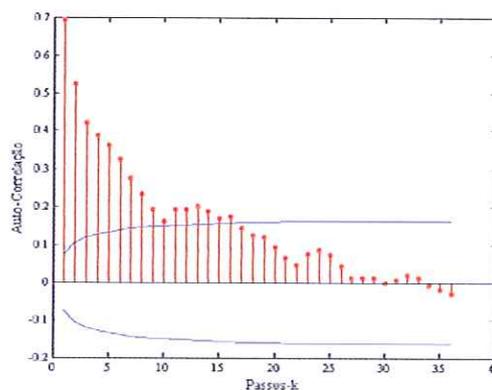


FIGURA 5.4: Gráficos da dispersão da série de Emborcação



a) Função de autocorrelação para a série de Furnas



b) Função de autocorrelação para a série de Emborcação

FIGURA 5.5: Função de autocorrelação para as séries dessazonalizadas de Furnas e Emborcação

5.2 - Seleção da ordem do modelo

Para seleção da ordem do modelo foram escolhidos os Critérios de Seleção AIC e BIC, conforme mencionados no Capítulo 4. Os valores dos critérios AIC e BIC para cada mês de cada série hidrológica foram determinados e estão apresentadas nas Tabelas 5.1 a 5.4. Os critérios foram calculados para valores de p_m de 1 a 10 com a finalidade de se determinar o número de parâmetros mais adequado para cada mês de cada série, sempre levando em consideração o critério de parcimônia de Box (BOX et al., 1994). Na avaliação dos critérios de informação, os modelos escolhidos foram aqueles que forneceram um menor valor para esses critérios (Tabelas 5.5 e 5.6). De acordo com a literatura, o critério BIC impõe maior penalidade para o número de parâmetros do modelo estimado do que o critério AIC. O uso do menor valor de BIC para a seleção do modelo resultou na escolha do modelo no qual o número de parâmetros é menor que aquele escolhido pelo AIC. Entretanto, em alguns meses, o critério AIC forneceu um menor valor que o critério BIC, mas a ordem estimada por ambos era a mesma. Sendo assim, as ordens escolhidas para cada mês foram aquelas determinadas em sua maioria pelo critério BIC. Tendo sido inferidas as ordens mensais para os modelos $PAR(p_m)$ de cada série hidrológica aqui pesquisada, o passo seguinte foi a determinação dos parâmetros mensais pelos métodos escolhidos.

TABELA 5.1 – Valores dos critérios AIC os meses da série hidrológica de Furnas

Critério AIC para os 12 meses da Série de Furnas												
p	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
1	-0,3637	-0,2777	-0,4025	-0,8377	-1,2941	-1,3311	-1,8999	-2,1448	-1,1255	-0,9613	-0,9008	-0,5173
2	-0,3344	-0,4087	-0,4336	-0,9077	-1,3932	-1,3386	-2,1526	-2,1089	-1,1498	-0,983	-0,8673	-0,5968
3	-0,3286	-0,3936	-0,417	-0,9793	-1,6167	-1,3254	-2,1175	-2,0916	-1,2453	-1,0543	-0,8325	-0,5611
4	-0,2958	-0,358	-0,378	-1,0104	-1,6308	-1,2987	-2,1041	-2,144	-1,4789	-1,1188	-0,8021	-0,5389
5	-0,3708	-0,3339	-0,3603	-0,9822	-1,6903	-1,2586	-2,0707	-2,1036	-1,4598	-1,1063	-0,8013	-0,5186
6	-0,3582	-0,3111	-0,3876	-0,956	-1,6535	-1,2318	-2,0405	-2,1214	-1,4339	-1,2021	-0,7603	-0,4946
7	-0,3326	-0,2801	-0,3886	-0,9448	-1,6107	-1,197	-2,0114	-2,081	-1,3998	-1,1592	-0,7398	-0,4513
8	-0,3395	-0,2367	-0,3462	-0,9429	-1,5741	-1,1522	-1,9944	-2,0424	-1,3817	-1,1142	-0,6948	-0,4071
9	-0,3136	-0,2139	-0,2991	-0,896	-1,5368	-1,1289	-2,0372	-2,0128	-1,3482	-1,1464	-0,6834	-0,4077
10	-0,2646	-0,2112	-0,2526	-0,8638	-1,5358	-1,0803	-2,0243	-1,9759	-1,3609	-1,1244	-0,6639	-0,4081

TABELA 5.2– Valores dos critérios BIC os meses da série hidrológica de Furnas

Critério BIC para os 12 meses da Série de Furnas												
p	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
1	-0,2926	-0,2067	-0,3315	-0,7667	-1,223	-1,2601	-1,8288	-2,0737	-1,0544	-0,8902	-0,8297	-0,4463
2	-0,2269	-0,3012	-0,3261	-0,8001	-1,2856	-1,2311	-2,0451	-2,0014	-1,0422	-0,8754	-0,7598	-0,4892
3	-0,1839	-0,249	-0,2723	-0,8346	-1,472	-1,1808	-1,9728	-1,947	-1,1006	-0,9096	-0,6879	-0,4165
4	-0,1134	-0,1755	-0,1956	-0,8279	-1,4483	-1,1162	-1,9217	-1,9615	-1,2964	-0,9364	-0,6196	-0,3564
5	-0,1498	-0,1129	-0,1393	-0,7612	-1,4693	-1,0376	-1,8497	-1,8826	-1,2388	-0,8853	-0,5803	-0,2976
6	-0,098	-0,0508	-0,1274	-0,6958	-1,3933	-0,9716	-1,7802	-1,8612	-1,1736	-0,9419	-0,5000	-0,2344
7	-0,0324	0,0201	-0,0884	-0,6446	-1,3105	-0,8968	-1,7112	-1,7808	-1,0996	-0,859	-0,4396	-0,1511
8	0,0015	0,1042	-0,0053	-0,602	-1,2331	-0,8113	-1,6535	-1,7015	-1,0408	-0,7732	-0,3539	-0,0661
9	0,0688	0,1685	0,0833	-0,5136	-1,1544	-0,7465	-1,6548	-1,6304	-0,9658	-0,764	-0,301	-0,0253
10	0,1601	0,2135	0,172	-0,4391	-1,1111	-0,6556	-1,5996	-1,5512	-0,9362	-0,6997	-0,2393	0,0166

TABELA 5.3 - Valores dos critérios AIC os meses da série hidrológica de Emborcação

Critério AIC para os 12 meses da Série de Emborcação												
p	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
1	-0,1293	-0,5383	-0,2715	-0,6583	-1,6452	-1,4008	-2,4343	-2,3154	-1,3649	-0,2016	-0,172	-0,2658
2	-0,1076	-0,622	-0,2353	-0,6905	-1,7425	-1,3672	-2,7387	-2,3375	-1,3286	-0,1948	-0,2363	-0,2333
3	-0,1149	-0,5948	-0,2357	-0,6568	-2,0029	-1,4012	-2,7095	-2,5344	-1,3172	-0,1581	-0,1989	-0,1978
4	-0,1847	-0,5796	-0,2388	-0,6537	-2,0318	-1,3736	-2,7445	-2,4986	-1,3634	-0,1428	-0,2070	-0,1591
5	-0,1634	-0,5467	-0,2046	-0,7118	-2,0689	-1,3564	-2,8346	-2,460	-1,3234	-0,1061	-0,1677	-0,1447
6	-0,1714	-0,5132	-0,1627	-0,6701	-2,0538	-1,3471	-2,8112	-2,435	-1,2833	-0,1350	-0,1259	-0,1414
7	-0,1321	-0,4885	-0,2072	-0,6497	-2,0600	-1,3221	-2,8391	-2,4715	-1,2492	-0,0983	-0,1787	-0,1001
8	-0,0965	-0,4885	-0,1621	-0,6047	-2,0708	-1,2773	-2,7980	-2,5060	-1,2125	-0,0775	-0,1415	-0,0576
9	-0,0535	-0,465	-0,1151	-0,562	-2,0281	-1,2399	-2,7681	-2,4990	-1,3251	-0,0347	-0,0944	-0,0112
10	-0,0395	-0,4225	-0,0681	-0,5147	-1,9961	-1,2156	-2,8248	-2,5046	-1,2900	0,0043	-0,0473	0,0224

TABELA 5.4 - Valores dos critérios BIC os meses da série hidrológica de Emborcação

Critério BIC para os 12 meses da Série de Emborcação												
p	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
1	-0,0583	-0,4673	-0,2005	-0,5873	-1,5741	-1,3297	-2,3633	-2,2443	-1,2938	-0,1305	-0,1010	-0,1948
2	-0,0001	-0,5145	-0,1277	-0,5830	-1,6349	-1,2597	-2,6312	-2,2300	-1,2211	-0,0872	-0,1288	-0,1257
3	0,0298	-0,4502	-0,0910	-0,5122	-1,8582	-1,2566	-2,5649	-2,3898	-1,1725	-0,0134	-0,0542	-0,0531
4	-0,0022	-0,3972	-0,0563	-0,4713	-1,8493	-1,1911	-2,5620	-2,3161	-1,1809	0,0396	-0,0245	0,0234
5	0,0576	-0,3257	0,0163	-0,4908	-1,8479	-1,1354	-2,6136	-2,2390	-1,1024	0,1149	0,0533	0,0763
6	0,0888	-0,2530	0,0975	-0,4098	-1,7936	-1,0868	-2,5510	-2,1748	-1,0231	0,1253	0,1343	0,1188
7	0,1681	-0,1883	0,0930	-0,3495	-1,7598	-1,0219	-2,5389	-2,1713	-0,9490	0,2019	0,1215	0,2001
8	0,2444	-0,1476	0,1789	-0,2638	-1,7299	-0,9363	-2,4571	-2,1651	-0,8716	0,2634	0,1994	0,2833
9	0,3289	-0,0826	0,2674	-0,1796	-1,6457	-0,8575	-2,3857	-2,1166	-0,9427	0,3477	0,2880	0,3712
10	0,3852	0,0022	0,3566	-0,0900	-1,5714	-0,7909	-2,4001	-2,0799	-0,8653	0,4290	0,3774	0,4471

TABELA 5.5 – Comparação entre os valores de BIC e AIC para a série hidrológica de Furnas

	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
AIC	-0,3708	-0,4087	-0,4336	-1,0104	-1,6903	-1,3386	-2,1526	-2,1448	-1,4789	-1,2021	-0,9008	-0,5968
BIC	-0,2926	-0,3012	-0,3315	-0,8346	-1,472	-1,2601	-2,0451	-2,0737	-1,2964	-0,9419	-0,8297	-0,4892
menor	-0,3708	-0,4087	-0,4336	-1,0104	-1,6903	-1,3386	-2,1526	-2,1448	-1,4789	-1,2021	-0,9008	-0,5968

TABELA 5.6 – Comparação entre os valores de BIC e AIC para a série hidrológica de Emborcação

	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
AIC	-0,1847	-0,622	-0,2715	-0,7118	-2,0708	-1,4012	-2,8391	-2,5344	-1,3699	-0,2016	-0,2363	-0,2658
BIC	-0,0583	-0,5145	-0,2005	-0,5873	-1,8582	-1,3297	-2,6312	-2,3898	-1,2938	-0,1305	-0,1288	-0,1948
menor	-0,1847	-0,622	-0,2715	-0,7118	-2,0708	-1,4012	-2,8391	-2,5344	-1,3649	-0,2016	-0,2363	-0,2658

A Tabela 5.7 apresenta os valores das ordens mensais para cada série. Estas ordens inferidas foram usadas para ajustar um modelo AR na determinação do vetor de parâmetros iniciais μ_m para o cálculo dos estimadores bayesianos exatos e serão apresentados posteriormente. Pode-se observar que os meses de setembro e outubro da série hidrológica de Furnas possui ordens elevadas (4 e 6, respectivamente) quando comparadas com os outros meses. Tal fato ocorre devido às altas correlações entre os meses secos levando a estimativa de ordens elevadas, para o modelo AR, no período correspondente aos meses secos. Pode-se, também, atribuir às ordens elevadas do modelo AR a não normalidade da série.

TABELA 5.7: Valores das ordens para cada mês inferidas para as séries hidrológicas para o modelo PAR

Série/Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Furnas	1	2	1	3	3	1	2	1	4	6	1	2
Emborcação	1	2	1	1	3	1	2	3	1	1	2	1

5.3 - Estimativa dos parâmetros do modelo

5.3.1 - Método clássico – Estimadores de máxima verossimilhança (EMV)

Para a determinação da função de máxima verossimilhança, escolheu-se a distribuição normal, por ser a mais importante da estatística, segundo CORDEIRO (1992) face às suas inúmeras aplicações e por constituir uma aproximação razoável para um grande número de distribuições de interesse. A função de probabilidade conjunta para cada mês foi determinada em função dos a_t e escrita em função dos Z_m 's. Os cálculos foram feitos usando-se somente a eq.(4.20) pela possibilidade de assumir a independência entre parâmetros, o que resultou em se trabalhar com uma função para cada mês. Os valores que maximizaram a eq.(4.20) são os estimadores de máxima verossimilhança (EMV) e foram obtidos usando-se o seu logaritmo natural e as suas derivadas parciais em função dos parâmetros Φ e τ (eq. 4.12 e 4.13). Para a determinação da matriz Z_m (eq. 4.7) foram usados os valores das vazões médias mensais das séries iniciando-se do segundo ano de dados ($n_m = 2, \dots, 60$, onde n_m indica o ano).

Os parâmetros para cada mês foram calculados pelos estimadores de máxima verossimilhança (EMV) considerando as ordens da Tabela 5.7. Para cálculo dos EMV foram usadas as eq.(4.16) e (4.17) e os resultados foram reunidos nas Tabelas 5.8 e 5.9.

TABELA 5.8: Valores dos parâmetros do modelo PAR da série de Furnas calculados pelos EMV

Mês	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	τ_c
Jan	0,5969	-	-	-	-	-	1,5311
Fev	0,6386	-0,1870	-	-	-	-	1,5796
Mar	0,6238	-	-	-	-	-	1,6044
Abr	0,5446	0,1373	0,2617	-	-	-	3,0610
Mai	0,5363	0,1833	0,3125	-	-	-	5,6354
Jun	0,8741	-	-	-	-	-	4,0658
Jul	0,5976	0,3823	-	-	-	-	9,4282
Ago	0,9454	-	-	-	-	-	9,1079
Set	0,6048	0,1884	0,6420	-0,5889	-	-	4,9299
Out	0,3376	-0,2654	0,4473	0,6277	0,0680	-0,4085	3,9917
Nov	0,7850	-	-	-	-	-	2,5913
Dez	0,3565	0,3891	-	-	-	-	1,9494

TABELA 5.9: Valores dos parâmetros do modelo PAR da série de Emborcação calculados pelos EMV

Mês	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	τ_c
Jan	0,4172	-	-	-	-	-	1,2030
Fev	0,7343	-0,1925	-	-	-	-	2,0058
Mar	0,4984	-	-	-	-	-	1,4517
Abr	0,6853	-	-	-	-	-	2,1160
Mai	0,6864	0,1436	0,2476	-	-	-	8,1353
Jun	0,8915	-	-	-	-	-	4,4008
Jul	0,6900	0,2924	-	-	-	-	17,3241
Ago	0,8363	-0,2047	0,3281	-	-	-	14,5241
Set	0,8239	-	-	-	-	-	4,5060
Out	0,4858	-	-	-	-	-	1,3078
Nov	0,6043	-0,3118	-	-	-	-	1,3665
Dez	0,5187	-	-	-	-	-	1,3850

5.4 – Método bayesiano

5.4.1 – Determinação dos parâmetros usando *priori* não-informativa de Jeffreys

A inferência bayesiana para os modelos PAR iniciou-se pela escolha da densidade *a priori* que pode ser definida como a densidade de probabilidade que permite incorporar conhecimento prévio sobre os parâmetros do modelo. Entretanto, quando são insuficientes ou não há informações sobre os parâmetros utiliza-se uma *priori* não-informativa. Utilizou-se, aqui, densidade de probabilidade *a priori* não-informativa de Jeffreys (eq. 4.19). Ao utilizarmos a *priori* não-informativa para determinação dos estimadores bayesianos, foi possível observar que os valores encontrados coincidem com os valores encontrados pelos EMV, o que é corroborado pela teoria bayesiana.

TABELA 5.10: Valores dos parâmetros do modelo PAR da série de Furnas calculados utilizando *priori* não-informativa de Jeffreys

Mês	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	τ_c
Jan	0,5969	-	-	-	-	-	1,5311
Fev	0,6386	-0,1870	-	-	-	-	1,5796
Mar	0,6238	-	-	-	-	-	1,6044
Abr	0,5446	0,1373	0,2617	-	-	-	3,0610
Mai	0,5363	0,1833	0,3125	-	-	-	5,6354
Jun	0,8741	-	-	-	-	-	4,0658
Jul	0,5976	0,3823	-	-	-	-	9,4282
Ago	0,9454	-	-	-	-	-	9,1079
Set	0,6048	0,1884	0,6420	-0,5889	-	-	4,9299
Out	0,3376	-0,2654	0,4473	0,6277	0,0680	-0,4085	3,9917
Nov	0,7850	-	-	-	-	-	2,5913
Dez	0,3565	0,3891	-	-	-	-	1,9494

TABELA 5.11: Valores dos parâmetros do modelo PAR da série de Emborcação calculados utilizando *priori* não-informativa de Jeffreys

Mês	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	τ_c
Jan	0,4172	-	-	-	-	-	1,2030
Fev	0,7343	-0,1925	-	-	-	-	2,0058
Mar	0,4984	-	-	-	-	-	1,4517
Abr	0,6853	-	-	-	-	-	2,1160
Mai	0,6864	0,1436	0,2476	-	-	-	8,1353
Jun	0,8915	-	-	-	-	-	4,4008
Jul	0,6900	0,2924	-	-	-	-	17,3241
Ago	0,8363	-0,2047	0,3281	-	-	-	14,5241
Set	0,8239	-	-	-	-	-	4,5060
Out	0,4858	-	-	-	-	-	1,3078
Nov	0,6043	-0,3118	-	-	-	-	1,3665
Dez	0,5187	-	-	-	-	-	1,3850

5.4.2 - Determinação dos parâmetros usando *a priori* informativa Normal-Gama

As *prioris* informativas são aquelas que representam o grau de conhecimento sobre os parâmetros do modelo. Optou-se por utilizar uma densidade *a priori* conjugada Normal-Gama (eq. 4.34) para os parâmetros Φ e τ . Sendo uma densidade *a priori* Normal para o parâmetro Φ e para o parâmetro τ , uma distribuição *a priori* Gama. As *prioris* informativas foram apresentadas nas equações (4.31) e (4.32). Ao se trabalhar as equações, duas novas variáveis são necessárias para determinação dos estimadores bayesianos. São elas: P_m e μ_m .

5.4.2.1- Determinação de P_m e μ_m

A matriz P_m usada é a matriz identidade e será ajustada em função de cada ordem mensal. Já o vetor de parâmetros μ_m foi estimado a partir dos parâmetros do modelo auto-regressivo periódico (AR) ajustado para a ordem para cada mês das séries investigadas. A identificação do tipo de modelo foi realizada pela análise de sua função de autocorrelação (FAC). Observando o gráfico das FAC's (Figura 5.5) das séries

nota-se uma extensão finita consistindo de quedas exponenciais e/ou ondas o que indica um modelo do tipo AR

➤ **Determinação da ordem do modelo**

Após a identificação do tipo de modelo a ser utilizado foram determinadas as ordens desses modelos AR usando-se os critérios AIC e BIC e as ordens pré-determinadas na Tabela 5.7. As séries foram ajustadas para iniciar o processo de estimação paramétrica do modelo PAR via equações (4.47) e (4.48), os resultados apresentam-se na Tabela 5.12.

➤ **Determinação dos parâmetros do modelo AR usando EMV**

Os parâmetros ϕ do modelo AR foram calculados pelo estimador de máxima verossimilhança (EMV) considerando-se as ordens mensais determinadas anteriormente (eq. 4.5). Estes valores foram usados como valores do vetor de parâmetros iniciais μ_m no cálculo dos estimadores bayesianos exatos utilizando-se de uma *priori* Normal-Gama.

TABELA 5.12: Valores dos parâmetros do modelo AR ajustado para as ordens mensais das séries em estudo

	Ordem	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
Furnas	1	0,7268	-	-	-	-	-
	2	0,6729	0,0732	-	-	-	-
	3	0,6616	0,0251	0,0769	-	-	-
	4	0,6536	0,0201	0,0330	0,0710	-	-
	5	0,6493	0,0186	0,0312	0,0385	0,0503	-
	6	0,6480	0,0182	0,0309	0,0387	0,0325	0,0257
Emborcação	1	0,7027	-	-	-	-	-
	2	0,6491	0,0763	-	-	-	-
	3	0,6398	0,0435	0,0558	-	-	-

5.4.2.2- Determinação dos estimadores bayesianos exatos (EBEx)

Os estimadores bayesianos exatos (EBEx) foram determinados pelas equações (4.46) e (4.47) e estão reunidos nas Tabelas 5.13 e 5.14.

TABELA 5.13: Valores dos parâmetros do modelo PAR da série de Furnas calculados pelos EBEx

Mês	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	τ_c
Jan	0,5994	-	-	-	-	-	1,5311
Fev	0,6337	-0,1786	-	-	-	-	1,5796
Mar	0,626	-	-	-	-	-	1,6044
Abr	0,548	0,1381	0,2565	-	-	-	3,061
Mai	0,5436	0,1815	0,3049	-	-	-	5,6354
Jun	0,8724	-	-	-	-	-	4,0658
Jul	0,6099	0,3685	-	-	-	-	9,4282
Ago	0,9427	-	-	-	-	-	9,1079
Set	0,6033	0,1955	0,5763	-0,5280	-	-	4,9299
Out	0,3698	-0,1869	0,3548	0,5682	0,0762	-0,3652	3,9917
Nov	0,7846	-	-	-	-	-	2,5913
Dez	0,3731	0,3730	-	-	-	-	1,9494

TABELA 5.14: Valores dos parâmetros do modelo PAR da série de Emborcação calculados pelos EBEx

Mês	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	τ_c
Jan	0,4216	-	-	-	-	-	1,203
Fev	0,7343	-0,1925	-	-	-	-	2,0058
Mar	0,5016	-	-	-	-	-	1,4517
Abr	0,6856	-	-	-	-	-	2,116
Mai	0,688	0,1425	0,2443	-	-	-	8,1353
Jun	0,8885	-	-	-	-	-	4,4008
Jul	0,6989	0,2811	-	-	-	-	17,3241
Ago	0,7838	-0,1406	0,3144	0	0	-	14,5241
Set	0,8219	-	-	-	-	-	4,506
Out	0,4896	-	-	-	-	-	1,3078
Nov	0,6016	-0,3035	-	-	-	-	1,3665
Dez	0,5215	-	-	-	-	-	1,385

5.5 – Previsão

Tanto para os estimadores de máxima verossimilhança (EMV) quanto para os estimadores bayesianos exatos (EBEx) foram usados 12 meses na previsão de valores das séries de vazões médias mensais de Furnas e Emborcação. O procedimento, nesta etapa, foi simples: das séries hidrológicas foram retirados os últimos doze meses (1 ano de dados), e estes valores foram calculados ora, utilizando-se os EMV ora, utilizando-se os estimadores bayesianos exatos (EBEx). Para este procedimento, considerou-se em um primeiro momento apenas 50 anos da série (1931 a 1980) e o último ano (1980) foi retirado e previsto, usando-se os EMV e os EBEx, e comparado-o com o valor histórico. Posteriormente, o mesmo procedimento foi realizado considerando-se a série de vazões com 60 anos (1931 a 1990) e, o ano de 1990, retirado para ser simulado utilizando-se dos parâmetros estimados. Esses resultados foram comparados com resultados obtidos usando-se os modelos $AR(p)$, também analisado sob os dois enfoques. Os valores encontrados foram apresentados em forma de gráficos e tabelas. Nas tabelas foram inseridos também os erros de previsão calculados.

Observando-se a Figuras 5.6 verifica-se que tanto para os EMV quanto para os EBEx, utilizando-se dos modelos $PAR(p_m)$, os valores previstos para 12 meses (ano 1980) acompanham o comportamento dos valores históricos para a série hidrológica de Furnas, o que vem mostrar a viabilidade do uso dos EBEx para previsão de valores. A Tabela 5.15 reúne os resultados obtidos para estes valores gerados e os erros calculados. Verifica-se que para os EBEx os resultados são levemente melhores do que aqueles calculados pelos EMV.

Tal afirmação pode ser percebida, com mais nitidez, ao compararmos os erros de previsão calculados entre os valores previstos pelos EMV e EBEx. Apesar dessa diferença ser bastante sutil, mesmo em si tratando de modelos que venham a gerar valores futuros e até mesmo investigados para previsão em hidrologia, toda e qualquer melhora no cálculo desses estimadores e, conseqüentemente, dos resultados obtidos torna-se útil.

A comparação entre os valores histórico e previsto usando um modelo $AR(2)$, tanto para a análise clássica quanto para a bayesiana, estão apresentados na Figura (5.7) para a série de Furnas para os anos de 1980. Pode-se observar que neste caso também os

EBEx geram valores melhores que os EMV. Entretanto, quando comparados os valores de um modelo AR(2) com aqueles gerados pelo modelo PAR(p_m), usando os EBEx, é possível perceber que este último oferece melhores resultados (Tabela 5.15).

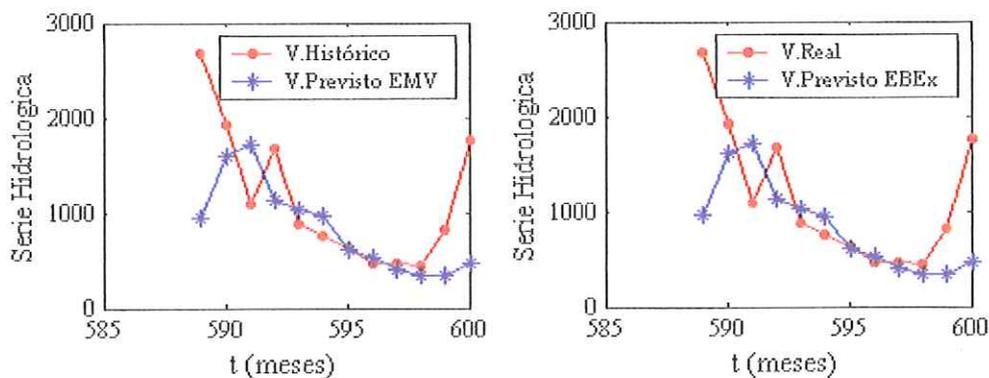


FIGURA 5.6: Comparação gráfica entre os valores históricos e valores previstos para o ano de 1980 usando os modelos PAR, os EMV e os EBEx para a série hidrológica de Furnas.

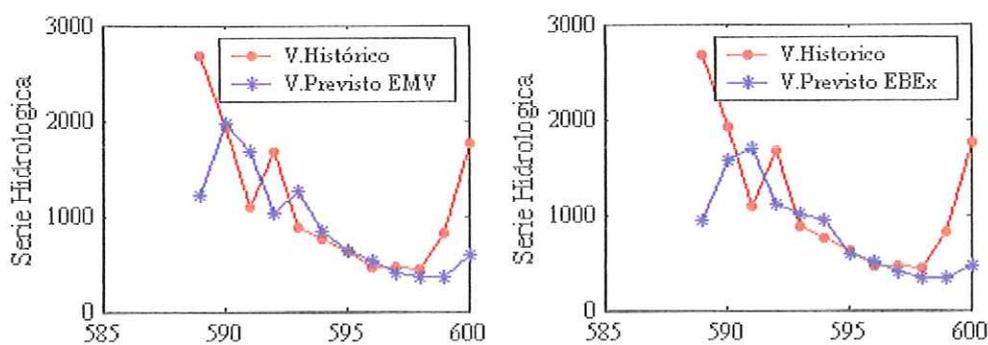


FIGURA 5.7: Comparação gráfica entre os valores históricos e valores previstos para o ano de 1980 usando os modelos AR(2), os EMV e os EBEx para a série hidrológica de Furnas.

TABELA 5.15: Valores obtidos a partir dos os EMV e os EBEx para a série hidrológica de Furnas para de 1980.

Valores Históricos (Ano:1980)	Valores Previstos			
	EMV		EBEx	
	PAR	AR(2)	PAR	AR(2)
2674	956,6	1209,9	968,8	942,2
1918	1596,9	1978,1	1614,6	1572,4
1103	1724,2	1672,5	1713,0	1692,1
1682	1139,5	1026,7	1126,9	1117,3
878	1028,8	1255,0	1039,0	1012,8
758	967,5	843,7	955,0	948,3
639	611,9	630,5	610,3	600,9
462	522,7	534,6	521,1	513,3
466	413,3	402,0	410,7	405,6
445	345,9	374,4	346,2	339,9
823	340,0	361,7	339,8	334,0
1753	466,5	597,9	473,0	459,6
REQM	689,49	619,94	685,16	693,39
EAPM	66,24%	52,71%	65,55%	67,87%
EAPMax.	275,74%	193,19%	270,58%	281,39%
Perc(90%)	208,40%	204,37%	204,37%	204,37%

As Figuras (5.8) e (5.9) representam as previsões realizadas para a série de Furnas, com 60 anos de dados históricos, utilizando os EMV e os EBEx para os modelos $PAR(p_m)$ e $AR(2)$. Observa-se, neste caso, que os EBEx oferecem melhores resultados do que os EMV. Pode-se notar, também que os modelos $PAR(p_m)$ proporcionaram melhores resultados do que os modelos $AR(2)$. Estes resultados vem corroborar aqueles, anteriormente, apresentados e mostram que o uso dos EBEx é válido para qualquer período que se deseja prever.

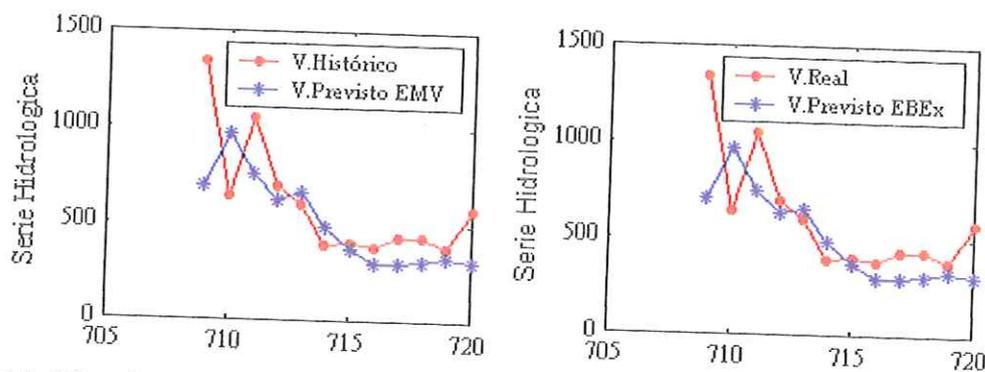


FIGURA 5.8 – Comparação gráfica entre os valores históricos e valores previstos para ano 1990 usando os modelos PAR, os EMV e os EBEx para a série hidrológica de Furnas.

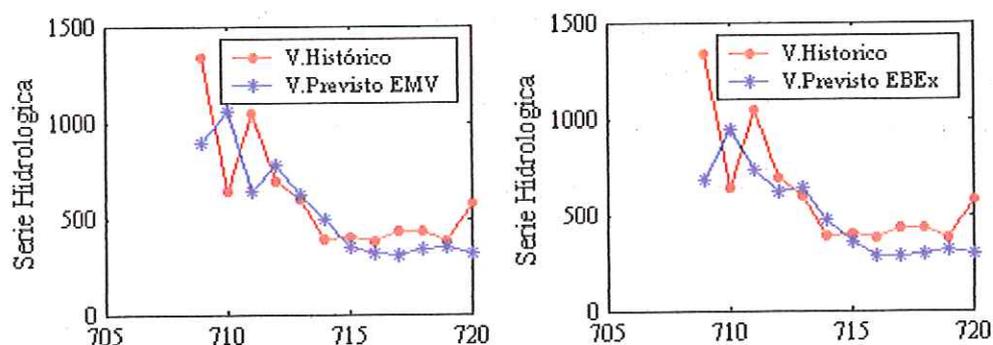


FIGURA 5.9 – Comparação gráfica entre os valores previstos e valores históricos para o ano 1990 usando o modelo AR(2), os EMV e os EBEx para a série hidrológica de Furnas.

Os resultados para os valores previstos pelos estimadores e os valores históricos, bem como os erros de previsão, para o ano de 1990 da série hidrológica de Furnas, estão reunidos na Tabela 5.16.

TABELA 5.16: Valores obtidos a partir dos os EMV e os EBEx para a série hidrológica de Furnas para o ano de 1990.

Valores Históricos (Ano:1990)	Valores Previstos			
	EMV		EBEx	
	PAR	AR(2)	PAR	AR(2)
1343	696,3	895,1	706,2	686,1
643	958,9	1062,3	961,4	942,7
1051	751,8	636,8	740,8	736,6
695	624,9	776,3	632,0	615,3
595	656,7	620	651,3	644,3
391	482,5	495,9	481,2	473,9
403	370,9	349,5	367,8	363,9
379	295,8	322,3	296,2	290,7
435	291,9	309,1	291,7	286,8
432	302,5	340,5	303,7	297,5
378	323,3	348,2	323,4	317,7
576	302,8	313,4	302,2	297,4
REQM	249,57	235,67	248,56	252,7
EAPM	36,74%	31,92%	36,58%	38,25%
EAPMax.	92,87%	83,78%	90,62%	95,76%
Perc(90%)	91,00%	90,31%	90,31%	90,31%

As Figuras 5.10 e 5.11 apresentam os gráficos obtidos para as previsões realizadas para a série hidrológica de Emborcação e as Tabelas 5.17 e 5.18 reúnem os valores obtidos para estas previsões e os erros calculados. Neste caso, também, verifica-se uma previsão melhor para os EBEx, conforme já mencionado anteriormente.

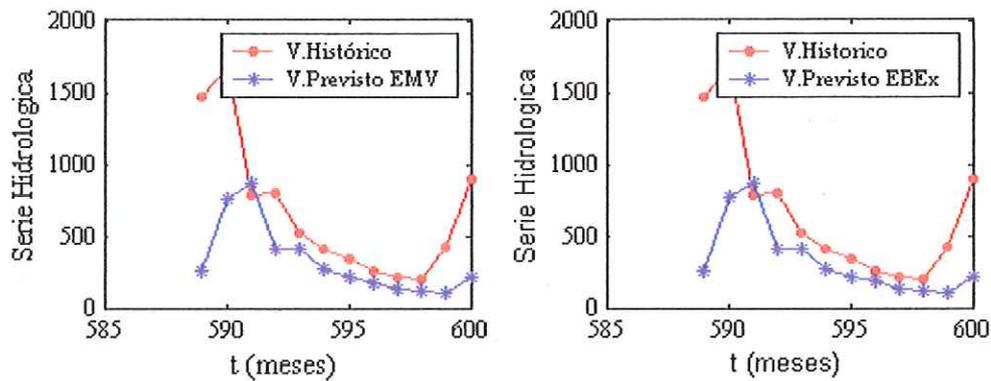


FIGURA 10: Comparação gráfica entre os valores históricos e valores previstos usando os modelos PAR, os EMV e os EBEx para a série hidrológica de Emborcação para o ano de 1980

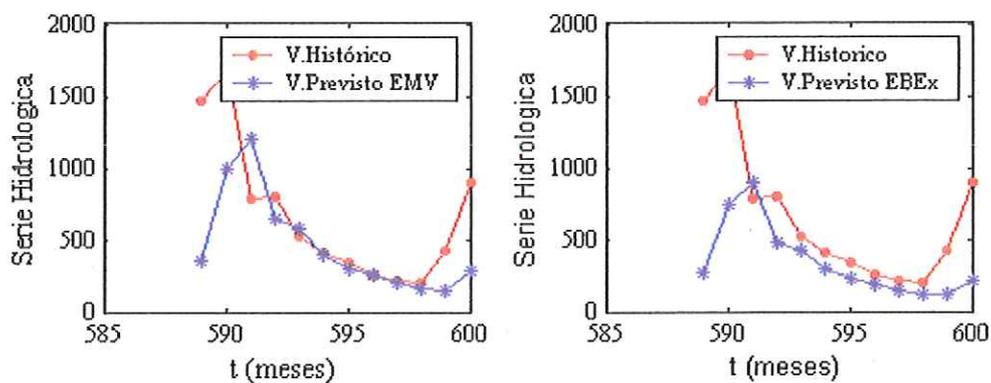


FIGURA 11: Comparação gráfica entre os valores históricos e valores previstos para o ano de 1980 usando os modelos AR(2), os EMV e os EBEx para a série hidrológica de Emborcação.

TABELA 5.17: Valores obtidos a partir dos os EMV e os EBEx para a série hidrológica de Emborcação para os valores históricos do ano de 1980.

Valores Históricos (Ano:1980)	Valores Previstos			
	EMV		EBEx	
	PAR	AR(2)	PAR	AR(2)
1470	260,4	356,2	261,8	265,7
1670	762,5	992,5	766,6	740,4
788	866,2	1196,2	870,9	892,3
793	408,8	638,9	410,9	476,5
527	411,3	574,9	413,5	428,8
407	273,4	402,6	274,8	300,3
347	211,1	304,4	212,3	227,1
259	180,0	256,3	181,0	191,2
215	134,3	194,6	135,1	145,1
205	111,5	159,3	112,1	118,8
417	106,3	149,5	106,9	111,5
890	216,3	286,3	217,5	213,6
REQM	505,15	440,88	503,88	502,38
EAPM	134,95%	74,39%	133,78%	126,25%
EAPMax.	464,54%	312,67%	461,51%	453,20%
Perc(90%)	357,39%	354,94%	355,84%	354,90%

Analisando as figuras e tabelas pode-se observar que no caso da série de Emborcação assim como na série de Furnas, ao se trabalhar com diferentes períodos para simulação da série, os EBEx oferecem melhores resultados do que os EMV, tanto no caso dos modelos $PAR(p_m)$ quanto dos modelos $AR(2)$. Sendo, porém, os valores previstos utilizando os modelos $PAR(p_m)$ e os estimadores EBEx, sensivelmente melhores.

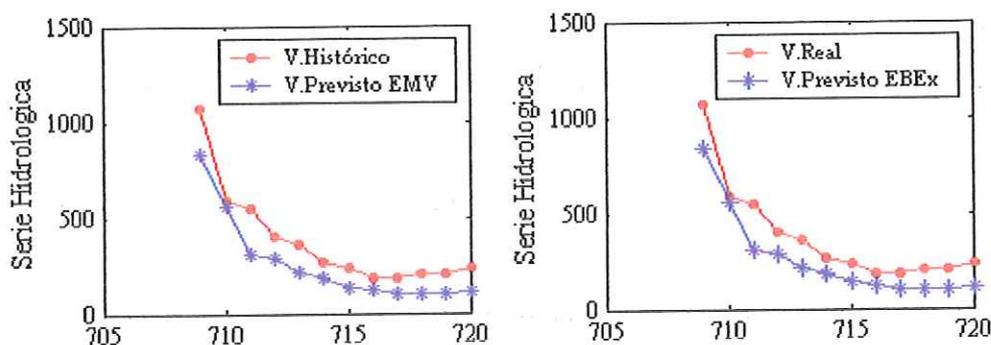


FIGURA 5.12 – Comparação gráfica entre os valores históricos e valores previstos para o ano de 1990 usando os modelos PAR, os EMV e os EBEx para a série hidrológica de Emborcação.

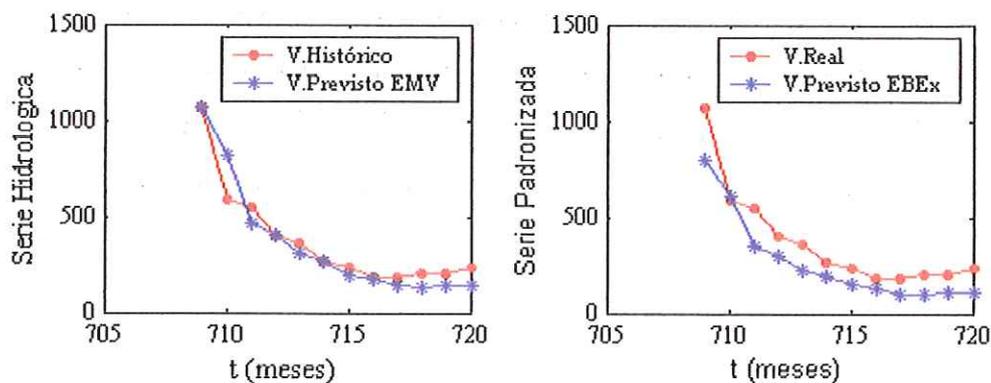


FIGURA 5.13 – Comparação gráfica entre os valores previstos e valores históricos para o ano de 1990 usando o modelo AR(2), os EMV e os EBEx para a série hidrológica a de Emborcação.

TABELA 5.18: Valores obtidos a partir dos os EMV e os EBEx para a série hidrológica Emborcação para o ano de 1990.

Valores Históricos (ano 1990)	Valores Previstos			
	EMV		EBEx	
	PAR	AR(2)	PAR	AR(2)
1070	829,9	1070,6	834,4	798,7
586	555,0	816,6	558,0	609,1
548	303,9	462,0	305,6	344,6
404	284,3	400,4	285,8	298,7
355	209,5	304,0	210,7	226,8
261	184,1	261,3	185,1	194,9
238	135,4	196,5	136,1	146,6
182	123,5	174,4	124,1	130,1
182	94,4	136,3	94,9	101,7
199	94,4	132,0	94,9	98,5
201	103,2	143,1	103,8	106,7
236	104,3	145,7	104,8	108,7
REQM	135,11	83,20	133,67	129,14
EAPM	67,99%	23,08%	23,07%	59,29%
EAPMax.	126,36%	62,03%	125,14%	117,22%
Perc(90%)	115,47%	113,30%	114,31%	114,31%

Os erros calculados usando-se os modelos AR(2) e PAR(p_m) foram inseridos na Tabela 5.15 e 5.16 para a série hidrológica de Furnas e nas Tabelas 5.17 e 5.18 para a série hidrológica de Emborcação. Analisando estes valores observa-se que os erros calculados para o modelo AR(2) quando usado o EMV é maior do que o erro calculado quando usado o EBEx. A mesma observação pode ser feita quando são avaliados os erros obtidos para o modelo PAR(p_m) considerando a análise clássica (EMV) e a

bayesiana. O que corrobora as análises realizadas anteriormente, ou seja, que o uso da inferência bayesiana na estimativa paramétrica oferece resultados melhores que os EMV. Pode-se, também, comparar o modelo periódico com o não-periódico evidenciando que o uso dos modelos periódico e a estimativa dos parâmetros considerando as características mensais da série proporciona melhores resultados.

5.5.1 - Análise dos Resíduos

Outra análise realizada foi a análise dos ruídos das séries hidrológicas em estudo. As Figuras 5.10 e 5.11 apresentam na parte superior o histograma dos ruídos (lado esquerdo) e a curva da probabilidade normal (lado direito) e, na parte inferior a correlação e a correlação parcial quando usados os EMV.

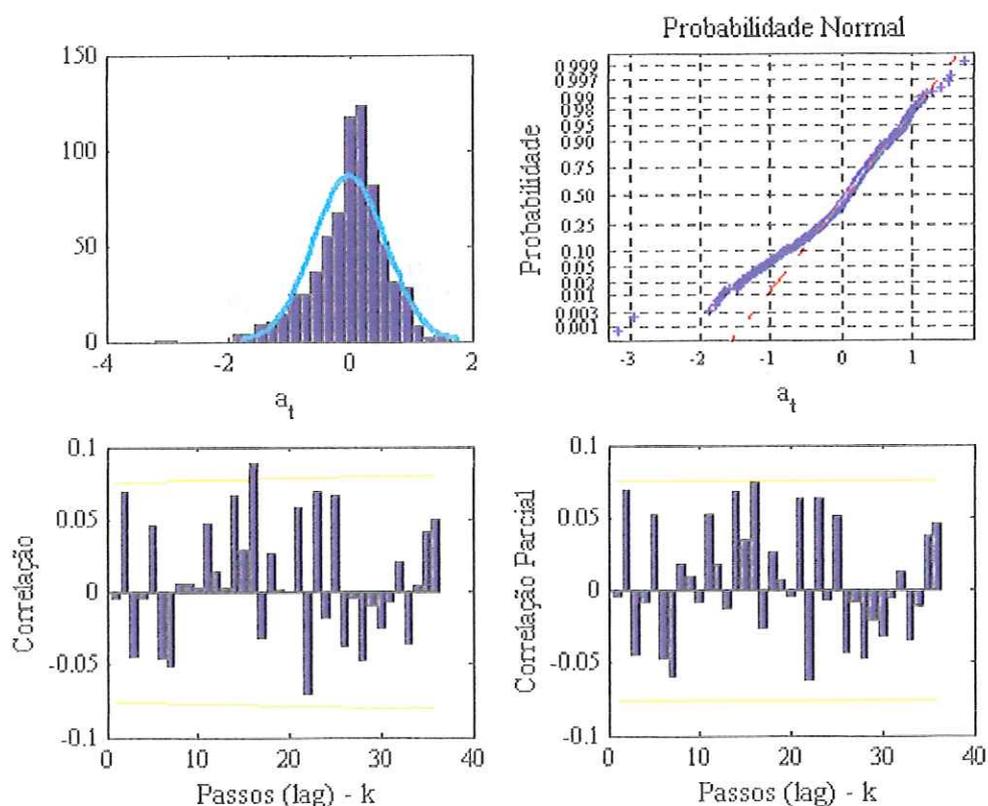


FIGURA 5.14: Análise gráfica da probabilidade, da correlação e da correlação parcial dos resíduos para a série hidrológica de Furnas.

Observando o histograma dos ruídos a_t pode-se notar que estes possuem uma distribuição aproximadamente normal. Avaliando o gráfico da probabilidade normal, verifica-se que na parte inferior os pontos se afastam da curva normal enquanto que no centro é possível notar que as estimativas oferecem uma menor sensibilidade às variações. Quanto aos gráficos de correlação e correlação parcial, verifica-se que as duas funções estão confinadas dentro de uma faixa aceitável. Considerando-se, portanto, que o modelo ajustado representa bem as correlações da série original.

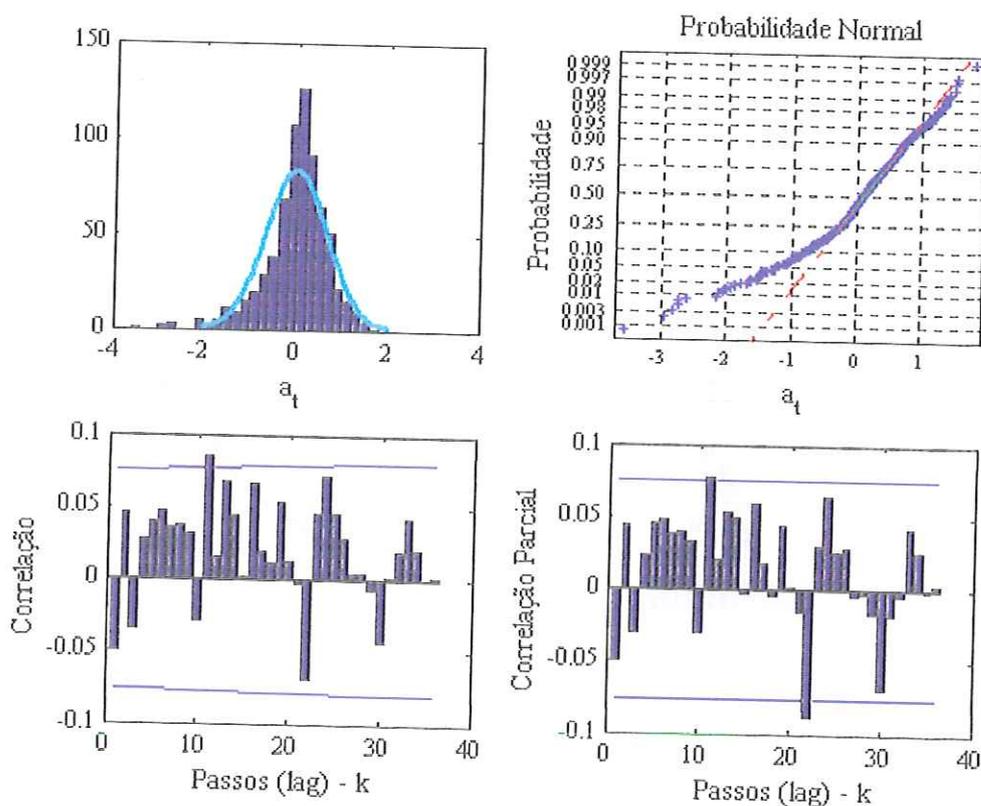


FIGURA 5.15: Análise gráfica da probabilidade, da correlação e da correlação parcial dos resíduos para a série hidrológica de Emborcação.

As considerações feitas acima podem ser atribuídas às duas séries em análise, apesar de serem diferentes, foi possível observar nos dois casos comportamentos semelhantes quando usado os EMV.

A abordagem bayesiana possui um desempenho semelhante, portanto não houve a necessidade de repetir a apresentação da análise dos resíduos

Analisando o estudo realizado pôde-se confirmar que a utilização dos EMV para estimativa dos parâmetros do modelo e para previsão de valores fornece bons resultados, entretanto pôde-se verificar, também, que o uso da abordagem bayesiana como uma ferramenta para análise de séries temporais é válida e que a estimativa de parâmetros por este método proporciona a obtenção de parâmetros que levam a resultados sensivelmente melhores na previsão de valores futuros.

6 – CONCLUSÕES

Para a obtenção dos resultados apresentados no Capítulo 5 foram escolhidos dois métodos para a estimativa de parâmetros de modelos estocásticos: o método de máxima verossimilhança e o método bayesiano. Na tentativa de alcançar os objetivos propostos, no início, desta pesquisa, uma metodologia adequada foi investigada e para avaliar o desempenho desta dois conjuntos de dados históricos de vazões médias mensais foram escolhidos. As séries hidrológicas usadas foram séries de vazões médias mensais das Usinas Hidrelétricas de Furnas e Emborcação.

Vários foram os métodos apresentados na literatura para determinação dos parâmetros de modelos estocásticos (Capítulo 3). Dentre os métodos clássicos apresentados, escolheu-se o método de máxima verossimilhança por ser um dos mais usados em hidrologia, conforme mencionado na literatura. Esse método foi escolhido para ser comparado com o método de estimativa paramétrica bayesiano, sendo este o alvo principal desta pesquisa. Algumas conclusões foram obtidas a partir deste estudo:

1. O uso do método de máxima verossimilhança para determinação paramétrica de modelos auto-regressivos e auto-regressivos periódicos proporciona bons resultados o que pode ser confirmado quando comparados os resultados obtidos para as séries em análise;
2. A comparação entre os modelos auto-regressivos periódicos - PAR e os modelos auto-regressivos - AR com parâmetros estimados pelo método de máxima verossimilhança, mostra que os resultados obtidos para a previsão para 12 meses, utilizando os modelos PAR, são melhores do que aqueles obtidos pelo modelo auto-regressivo;
3. Para a análise bayesiana determinou-se uma priori não informativa de Jeffreys que utilizada para estimativa dos parâmetros os estimadores bayesianos forneceu

estimadores bayesianos exatamente iguais aos estimadores de máxima verossimilhança, tanto para os modelos auto-regressivos quanto para os modelos auto-regressivos periódicos (PAR);

4. Na determinação dos parâmetros bayesianos exatos - EBEx foi utilizada uma priori normal-gama, conforme a metodologia desenvolvida por HUTTER (1998) com alterações na determinação da matriz μ_m conduzindo erros menores nas previsões para 12 meses quando da utilização desses parâmetros;
5. Analisando as previsões para 12 meses realizadas utilizando-se o modelo auto-regressivo periódico PAR(p_m), os estimadores de máxima verossimilhança e os estimadores bayesianos exatos, pôde-se verificar que os resultados obtidos acompanham as características dos dados históricos;
6. Comparando-se os gráficos para as previsões e os resultados tabelados, verificou-se que os EBEx proporcionam melhores resultados que os estimadores de máxima verossimilhança, mesmo sendo essa diferença seja sutil;
7. Verificou-se que para períodos diferentes, os resultados encontrados apontam como melhores previsões aquelas determinadas através do uso dos EBEx, o que pode ser confirmado comparando-se os erros de previsão tabelados;
8. As comparações dos resultados para as previsões para 12 meses das séries em estudo e dos erros mostraram a viabilidade da abordagem bayesiana na análise de séries hidrológicas e na estimativas do parâmetros dos modelos estocásticos representativos destas séries.

Entretanto, mais pesquisas envolvendo a abordagem bayesiana e as séries temporais hidrológicas podem ser desenvolvidas, sendo assim sugere-se como proposta para futuras investigações:

1. Considerar modelos não gaussianos;
2. Fazer a análise multivariada de postos de vazões;
3. Estudar o grau de dependência entre variáveis.

ANEXO A

Vazões Médias Mensais da Usina Hidroelétrica de Furnas



Valores das vazões médias mensais para o período de janeiro de 1931 a dezembro de 1990 da Usina Hidrelétrica de Furnas

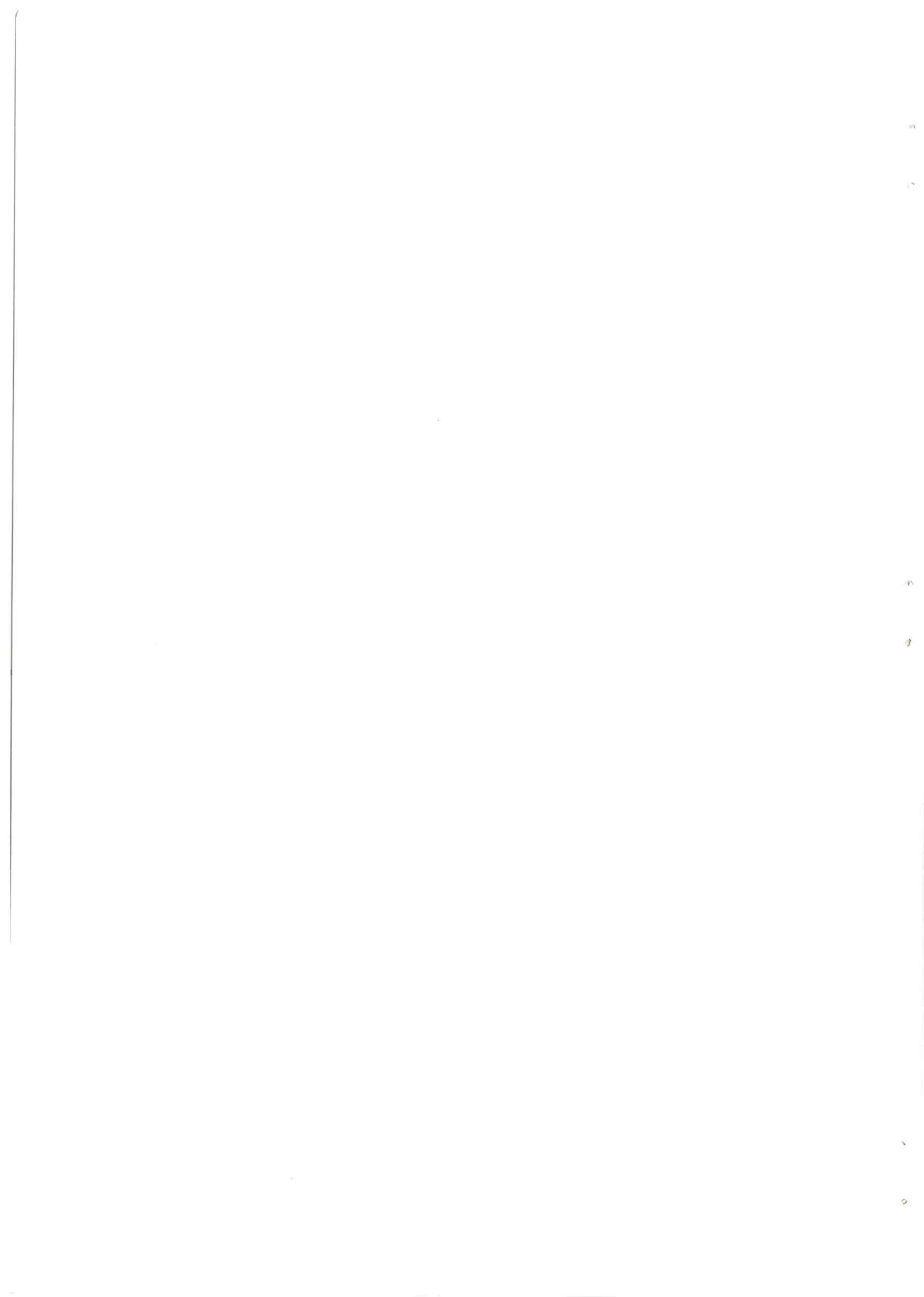
Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1931	1470	3080	2160	1580	1020	806	629	493	563	654	690	1140
1932	1930	1920	1530	884	696	655	497	399	414	546	597	1490
1933	1350	747	717	632	467	384	359	356	326	437	324	713
1934	1220	720	773	505	363	310	275	242	277	305	331	1300
1935	1850	2690	1570	1360	930	698	576	493	462	628	550	680
1936	614	602	1850	959	648	461	392	380	469	376	625	1340
1937	2630	1610	1130	851	853	645	480	392	339	842	1020	2370
1938	1950	1650	1470	955	869	661	531	505	507	812	944	1690
1939	2160	2040	1080	973	752	581	507	424	378	426	638	1400
1940	2070	2550	2150	1160	843	661	536	434	409	488	1050	1610
1941	1730	1000	972	794	514	465	464	342	610	669	704	1310
1942	1310	1250	1900	1130	767	639	551	426	398	545	855	1230
1943	2900	2260	2120	1130	798	683	560	502	449	651	659	1090
1944	1000	1530	1650	912	622	502	430	354	300	328	513	671
1945	864	2030	1200	991	635	558	505	357	321	356	724	1590
1946	3160	1290	1470	1150	755	630	553	438	339	492	692	750
1947	1760	1770	3650	1690	1030	803	700	553	765	630	708	1220
1948	1520	1750	1840	1080	768	612	520	449	377	393	530	1050
1949	1570	2260	1330	947	745	632	490	388	347	418	564	849
1950	1590	2634	1590	1180	857	685	530	423	367	495	1020	1670
1951	1590	2180	1840	1280	828	683	571	490	398	423	408	724
1952	1180	1730	2340	1200	708	693	531	414	377	387	668	745
1953	571	653	806	979	510	403	337	292	314	336	518	812
1954	745	1260	668	643	634	458	335	279	225	306	426	596
1955	1130	694	867	665	422	383	285	235	206	276	422	1160
1956	1050	704	1170	599	566	534	405	437	357	315	423	1180
1957	1620	1480	1670	1560	945	680	588	470	621	446	782	1190
1958	960	1550	1240	891	889	754	639	483	589	660	711	848
1959	1860	1410	1240	1090	692	584	504	471	408	432	668	778
1960	1450	1512	1977	971	839	639	624	513	354	421	587	1447
1961	2750	2875	2780	1463	1257	882	708	602	539	455	652	789
1962	1312	2403	1660	993	832	725	564	507	519	760	997	1675
1963	1306	1405	875	452	468	357	368	296	258	353	447	252
1964	1020	1983	982	567	562	390	373	249	213	504	608	1255
1965	2449	3114	2387	1185	1224	921	808	668	566	927	1095	1688
1966	3215	2086	2670	1385	994	697	556	462	458	743	1431	1699
1967	2904	2584	1897	1139	848	708	627	509	459	512	1084	1203
1968	1640	988	941	590	446	384	336	346	350	444	379	1338
1969	1150	1130	942	616	405	414	318	297	196	485	1230	1131
1970	1415	1270	1305	738	506	419	407	326	592	521	826	531
1971	606	355	463	379	288	344	289	200	218	464	545	1529
1972	1308	1776	1600	864	565	437	594	416	360	684	1142	1280
1973	1565	1419	903	1111	626	477	405	354	315	483	726	1396
1974	1832	1047	1537	1081	637	632	483	381	290	406	380	1093
1975	1442	1555	851	671	473	351	390	237	224	437	1020	1154
1976	944	1152	1246	929	749	721	760	721	1048	904	1366	1918
1977	1978	1583	1181	1160	645	561	441	338	565	407	833	1423
1978	1914	1267	1203	734	675	668	471	345	332	485	859	1084

cont. - Valores das vazões médias mensais para o período de janeiro de 1931 a dezembro de 1990 da Usina Hidrelétrica de Furnas

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1979	1396	2400	1399	935	772	608	551	530	715	510	943	1654
1980	2674	1918	1103	1682	878	758	639	462	466	445	823	1753
1981	2768	1327	1252	1064	749	671	475	440	331	807	1330	2338
1982	2445	1922	2900	1782	1244	1066	784	599	461	771	777	2118
1983	3529	3264	3085	2367	1679	2392	1368	1082	1989	1934	2018	3117
1984	2182	1301	972	887	828	543	440	461	528	410	629	1318
1985	2596	2194	2416	1308	843	648	540	449	486	457	758	1040
1986	1690	1538	1349	777	807	529	545	567	351	279	345	1732
1987	1907	1720	1166	1208	807	671	519	409	551	444	586	1334
1988	1289	1790	1404	810	680	612	404	355	296	555	648	799
1989	1783	1585	1590	864	561	522	460	451	470	432	659	1234
1990	1343	643	1051	695	595	391	403	379	435	432	378	576

ANEXO B

Vazões Médias Mensais da Usina Hidroelétrica de Emborcação



Valores das vazões médias mensais para o período de janeiro de 1931 a dezembro de 1990 da Usina Hidrelétrica de Emborcação

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1931	911	1610	1610	1180	636	420	377	323	316	308	314	426
1932	364	944	728	491	342	326	268	198	145	241	312	716
1933	1200	990	716	658	427	330	243	213	193	237	278	557
1934	676	531	529	414	290	162	132	105	126	172	151	237
1935	944	1150	672	1010	618	384	274	241	133	221	247	416
1936	431	284	843	581	360	233	203	154	134	138	205	322
1937	700	290	382	445	314	262	185	136	120	140	336	666
1938	766	527	529	427	253	211	197	123	95	87	184	561
1939	831	863	380	320	272	227	164	144	106	117	229	324
1940	581	1040	905	451	348	260	195	110	112	134	418	364
1941	720	473	404	487	245	193	161	122	121	154	237	449
1942	511	620	1030	614	360	537	332	177	161	155	338	519
1943	1360	1250	1100	614	388	348	278	192	174	280	483	473
1944	449	720	682	479	366	243	203	156	125	134	356	386
1945	559	1150	1280	1310	656	418	346	255	148	256	447	1060
1946	1340	927	1190	748	497	370	336	255	221	207	288	491
1947	666	825	1460	1020	525	372	304	253	213	213	203	571
1948	869	819	954	668	336	354	235	198	168	168	163	644
1949	781	1450	1010	587	416	354	295	234	181	246	321	546
1950	586	829	755	561	323	241	191	144	112	198	573	685
1951	1090	1000	950	694	400	316	244	199	150	161	188	293
1952	612	1050	2000	731	471	356	280	216	189	181	337	674
1953	372	343	845	756	388	281	216	164	162	387	384	849
1954	465	846	412	452	381	261	171	119	99	80	379	528
1955	674	630	478	667	275	214	157	119	88	206	309	1130
1956	772	470	859	423	476	397	274	228	173	166	537	1340
1957	1450	1540	1400	1250	746	517	405	322	268	249	361	773
1958	816	935	707	575	407	326	290	212	221	241	186	266
1959	951	575	1040	500	317	254	200	160	123	161	377	416
1960	696	824	879	550	393	289	230	173	136	147	361	826
1961	1350	1450	1220	652	555	380	298	224	172	161	255	405
1962	984	1060	1000	535	406	303	235	197	170	236	305	1120
1963	987	979	543	373	293	242	203	163	126	125	171	146
1964	789	966	457	391	278	192	156	122	94	263	417	621
1965	970	1200	1500	765	496	358	296	237	169	265	462	859
1966	1390	1600	932	703	522	387	306	237	190	311	450	870
1967	1040	1190	987	706	486	368	287	220	183	156	368	854
1968	890	920	918	533	371	290	234	196	169	219	236	601
1969	436	551	461	296	225	165	130	103	84	164	657	700
1970	1350	1120	793	504	334	254	207	163	172	242	295	266
1971	193	200	300	213	134	119	88	77	86	200	386	1060
1972	559	626	602	504	307	228	203	149	122	369	883	745
1973	814	819	805	851	451	330	267	200	165	279	582	599
1974	661	422	1100	951	542	383	300	245	178	267	216	405
1975	740	662	398	406	282	200	177	129	97	157	374	415
1976	401	404	463	320	223	165	142	103	158	174	482	1210
1977	1070	777	401	483	309	247	185	140	131	195	289	701
1978	1300	679	946	629	416	354	269	206	177	212	415	1030

Cont.- Valores das vazões médias mensais para o período de janeiro de 1931 a dezembro de 1990 da Usina Hidrelétrica de Emborcação

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1979	1430	1900	1020	760	502	417	324	267	275	219	398	502
1980	1470	1670	788	793	527	407	347	259	215	205	417	890
1981	1540	870	637	664	446	362	279	217	166	268	1040	1110
1982	1580	1160	1690	1070	660	501	380	296	236	265	258	390
1983	1510	2350	1340	1060	686	521	410	317	282	431	671	1290
1984	886	571	511	680	373	273	220	187	222	168	205	566
1985	1280	1010	926	639	429	318	256	204	183	176	259	567
1986	1200	1010	695	417	327	244	201	197	151	139	156	420
1987	662	504	481	560	344	247	187	147	135	175	285	982
1988	797	930	886	624	405	324	248	196	149	212	347	631
1989	664	728	659	396	285	232	190	166	158	178	420	1600
1990	1070	586	548	404	355	261	238	182	182	199	201	236

7 – REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, v21, n.2, p.243-247.
- _____. (1974). A New look at the statistical model identification. *IEEE Transactions Automation Control*, n.19, p.716-723.
- _____. (1977). An Objective use of bayesian models. *Annals of the Institute of Statistical Mathematics*, v.29, part A, n.1, p.9-20.
- _____. (1978). On The Likelihood of a time series models. *The statistician*, v.27, n.3/4, p.217-2235.
- ANDRADE, M.G.; MIZOI, M.F. (1998). Aplicação de MCMC na estimação de máxima verossimilhança para processos AR(p) e MA(q). São Carlos, ICMC/USP. p.14. /Notas n.52/
- ANDRADE, M.G.; BARRETO, G. A. (1999). Estimação e análise de robustez na abordagem bayesiana de modelos AR(p): estudo de caso em previsão de vazão. São Carlos, ICMC/USP. p.52. /Relatório Técnico/
- ANDRADE FILHO, M.G. (2001). Métodos estocásticos para previsão de carga em sistemas de energia elétrica. São Carlos, ICMC/USP. /Notas Didáticas/
- ANDRADE, M.G.; FRAGOSO, M.D.; CARNEIRO, A.A.F.M. (2001). A Stochastic approach to the flood control problem. *Applied Mathematical Modelling*, n.25, p.499-511.

- ANDERSON, P.L.; VECCHIA, A.V. (1993). Assyntotic results for periodic autoregressive moveing-avarege processs. *Journal Time Series Analysis*, v.1, p.1-18,
- BARRETO, G.A.; ANDRADE, M.G. (2000). Estimação paramétrica de modelos autoregressivos via estatística bayesiana e simulação de Monte-Carlo. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 13., Florianópolis, SC, 11 a 14 de setembro. *Anais do XIII Congresso Brasileiro de Automática – CBA 2000*. p.92-97
- BERGER, J.O. (1995). *Statistical decision theory and Bayesian analysis* (2nd edition). Springer-Verlag, N.Y.
- BERGER, J.O. (1998). Recent developments in Bayesian inference with applications in hydrology. In: *Statistical and Bayesian methods in hydrological sciences – International Hydrological Programme*. Paris, UNESCO. p.43-61. (IHP/Technical documents in hydrology, 20).
- BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. (1994). *Time series analysis – Forecasting and control*. 3.ed. New Jersey, Prentice Hall.
- BOX, G.E.P.; TIAO, G.C. (1975). Intervention analysis with application to economic and environmental problems. *Journal of the American Statistical Association*, v.70, n.349, p.70-79.
- _____. (1992). *Bayesian inference in statistical analysis*. New York, Wiley Interscience/ John Wiley.
- BRAS, R.L.; RODRIGUEZ-ITURBE, I. (1985). *Random functions and hydrology*. California, Addison-Wesley.
- CASELLA, G.; GEORGE, E.I. (1992). Explaining the gibbs sampler. *The American Statistician*, v.46, n.3, p.167-174.
- CHEEBANE, M.; SALAS, J.D.; BOES, D.C. (1995). Product periodic autoregressive processes for modeling intermittent monthly streamflows. *Water Resources Research*, v.31, n.6, p.1513-1518.

- CHIB.S; GREENBERG, E. (1980). Understanding the metropolis-hasting algorithm. *The American Statistician*, v.49, n.4, p.327-335.
- COIMBRA, A.R.S.R. (1978). *Adequação e teste de um modelo estocástico para as vazões médias diárias do Rio Jaguaribe*. São Carlos. 118p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.
- COSTA, A.R. (1989). *Modelo de previsão de vazão em tempo real aplicado à bacia do Rio Piracicaba*. São Carlos. 201p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.
- ENGLE, R.F.; BROWN, J.B. (1986). Model selection for forecasting. *Applied Mathematics and Computation*, v.20, n.(3-4), p.313-327.
- ESHETE, Z.; VANDEWIELE, G.L. (1992). Comparison of non-gaussian multicomponent and periodic autoregressive models for river flow. *Stochastic Hydrology and Hydraulics*, v.6, p.223-238.
- FISHER, R.A. (1921). On The “Probable error” of a coefficient of correlation deduced from a small sample. *Metron*, v.1
- FREITAS, M A.V., org. (1999). *O Estado da água no Brasil – 1999 – perspectivas de gestão e informação de recursos hídricos*. Brasília, Ministério de Minas e Energia. Ministério do Meio Ambiente/Secretária de Recursos Hídricos. Organização Meteorológica Mundial, Programa das Nações Unidas para o Desenvolvimento.
- GAMERMAN, D. (1996). *Simulação estocástica via cadeias de Markov*. São Paulo, Associação Brasileira de Estatística.
- GELMAN, A.E.; RUBIN, D. (1992). *Inference from iterative simulation using multiple sequences*. *Statistical Science*, v.7, p.457-472.
- GELMAN, R.; STERN, H.S.; RUBIN, D.B. (1995). *Bayesian data analysis*. London. Chapman & Hall.
- GENOVEZ, A.M. (1982). *Modelo estocástico das vazões mensais do Rio Tietê em Cumbica*. São Carlos. 121p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

- HANNAH, E.; QUINN, B. (1979). The Determination of the order of an autoregression. *Journal Royal. Statistical. Society.*, Serie B, v.41, n.2, p.190-195.
- HILLMER, S.C.; TIAO, G.C. (1979). Likelihood function of stationary mutiple autoregressive moving avarage models. *Journal of The American Statistical Association*, v.74, p.652-660.
- HUTTER, C.F.F. (1998). *Uma Abordagem bayesiana para modelos auto-regressivos periódicos –PAR*. São Carlos. 131p. Dissertação (Mestrado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.
- KASS, R.E. & RAFTERY, A.E. (1995). Bayes Factors. Joournal or the American Statistical Association, v.90, n.430, p.773-795.
- KELMAN, J. (1987). *Modelos estocásticos no gerenciamento dos recursos hídricos – Modelos para gerenciamento de recursos hídricos*. Rio de Janeiro, Nobel/ ABRH. v.1, p.327-424. (Coleção ABRH de Recursos Hídricos).
- KOTTEGODA, N.T. et al. (1998). A Stochastic model for simulating daily flows in a highly developed basin. In: *Statistical and Bayesian methods in hydrological sciences* – International Hydrological Programme. Paris, UNESCO. p.235-250. (IHP/Technical documents in hydrology, 20).
- KRZYSZTOFOWICZ, R. (1983). A Bayesian markov model of the flood forecast process. *Water Resources Research*, v.19, n.6, p.1455-1465.
- LU, Z.-Q.; BERLINER L.M. (1999). Markov switching time series models with application to a daily runoff series. *Water Resources Research*, v.35, n.2, p.523–534.
- MADSEN, H; ROSBJERG, H. (1995). A Regional bayesian method for estimation of extreme streamflow droughts/ Une méthode bayesienne d'estimation régionale pour le calcul des débits extreme en période de sécheresse. In: *Statistical and Bayesian methods in hydrological sciences* – International Hydrological Programme. Paris, UNESCO. p.327-340. (IHP/Technical documents in hydrology, 20).

- McLEOD, A.I. (1992). Diagnostic checking of periodic autoregressive models with application. *Journal of Time Series Analysis*, v.15, n.2, p.221-233.
- MILLER, J.W. (1995). Exact maximum likelihood estimation in autoregressive process. *Journal of Time Series Analysis*, v.16, n.6, p.607-615.
- OCCHIPINTI, A.G. (1979). *Aperfeiçoamento dos modelos markovianos para previsão e geração de deflúvios*. São Carlos. 91p. Tese (Livre Docência) – Escola de Engenharia de São Carlos, Universidade de São Paulo.
- OLIVEIRA, J.R.T. (1998). *Uso do amostrador de gibbs e metropolis-hasting em análise bayesiana de modelos AR(p)*. São Carlos. 75 p. Dissertação (Mestrado) - Instituto de Ciências Matemáticas e Computação de São Carlos, Universidade de São Paulo
- PARZEN, E. (1962). *Stochastic process*. San Francisco, Holden-Day.
- PEREIRA, B.B.; PAIS, M.B.Z.; SALES, P.R.H. (1986). *Análise espectral de séries temporais – uma introdução para engenharia, economia e estatística*. Rio de Janeiro, Arte Final Leasing.
- PEREIRA, C.A.B.; VIANA, M.A.G. (1982). *Elementos de inferência bayesiana*. Rio de Janeiro, Editora da UFRJ.
- POLE, A.; WEST, M.; HARRISON, J. (1999). *Applied bayesian forecasting and time series analysis – texts in statistical science*. New York, Chapman & Hall.
- SALAS, J.D. et al. (1988). Applied modeling of hydrologic time series. Fort Collins, *Water Resource*.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *The Annals of Statistical.*, n.6, p.461-464.
- SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, v.63, n.1, p.117-126.
- SHIBATA, R. (1976). An Optimasl selection of regression variables. *Biometrika*, v.68, p.45-54.

- THOMAS, H.A.; FIERING, M.B. (1962). Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation. In: MAASS, A. et al. *Design of water resources*. Cambridge, Harvard University.
- TROUTMAN, B.M. (1979). Some results in periodic autoregression. *Biometrika*, v.66, n.2, p.219-228.
- VANDEWIELE, G.L.; DOM, A. (1989). A Non-gaussian multicomponent model for river flow. *Water Resource Research*. v.25, p.397-404.
- VECCHIA, A.V. (1985). Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics*, v.27, n.4, p.375-384.
- YEVJEVICH, V. (1972). Stochastic process in hydrology. Fort Collins, *Water Resources*. p.2

8 – OBRAS CONSULTADAS

- AKAIKE, H. (1978). A New look at the bayes procedure. *Biometrika*, v.65, p.53-59.
- ANDRADE FILHO, A.G. (1994). *Uma Investigação sobre método vizinhos mais próximos na previsão de chuva-vazão*. São Carlos. 249p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.
- ANDRADE, M.G.; HUTTER, C.F.F. (1998). *Teste de sazonalidade para função de auto-correlação de processos auto-regressivos periódicos – PAR(p_m)*. São Carlos, ICMC/USP. p.12. (Notas n.46).
- ANDRADE, M.G.; BARRETO, G.A. (1999). *Estimação e análise de robustez na abordagem bayesiana de modelos AR(p)*: estudo de caso em previsão de vazão. São Carlos, ICMC/USP. p.32. (Notas n.97).
- ARAÚJO, J.K. (1991). *Método dos fragmentos aplicado a rios intermitentes: avaliação dos erros introduzidos no cálculo da disponibilidade de reservatórios*. Fortaleza. 124p. Dissertação (Mestrado) – Universidade Federal do Ceará.
- BOSHNAKOV, G. (1996). Recursive computation of the parameters of periodic autoregressive moving-average processes. *Journal of Time Series Analysis*, v.17, n.4, p.333-349.
- CORDEIRO, G.M. (1992). Introdução à teoria de verossimilhança. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 10., Rio de Janeiro, *Anais do X Simpósio Nacional de Probabilidade e Estatística* Rio de Janeiro, Universidade Federal do Rio de Janeiro/ Associação Brasileira de Estatística/ Instituto Brasileiro de Geografia e Estatística/ Laboratório Nacional de Computação Científica.
- CÓSTA NETO, P.L.O. (1977). *Estatística*. São Paulo, Edgar Blücher.

- CURRY, K.; BRAS, R.L. (1980). Multivariate seasonal time series forecast with application to adaptive control. Cambridge, Ralph M. Laboratory for Water Resources and Hydrodynamics/ M.I.T. /Technical Report, 253/.
- _____. (1995). Understanding the metropolis-hasting algorithm. *The Annals. Statistical*, v.49, p.327-335.
- CHANG, T.J.; KAVVAS, M.L.; DELLEUR, J.W. (1984). Daily precipitation modeling by discrete autoregressive moving average processes. *Water Resources Research*, v.20, n.5, p.565-580.
- CLARKE, A.B.; DISNEY, R.L. (1979). *Probabilidade e processos estocásticos*. Rio de Janeiro, Livros Técnicos e Científicos.
- DEUTSCH, C.V. (1994). Algorithmically-defined random function models. In: DIMITRAKOPULOS, K. *Geostatistics for the next century*. Norwell, Kluwer Academy. p.422-435.
- DOOB, J.L. (1953). *Stochastic process*. New York, John Wiley.
- FOFOULA-GEOGIU, E.; LETTENMAIER, D.P. (1987). A Markov renewal model for rainfall occurrences. *Water Resources Research*, v.23, n.5, p.875-884.
- HESTIR, K. et al. (1998). Inverse hydrologic modeling using stochastic growth algorithms. *Water Resources Research*, v.24, n.12, p.3335-3347.
- HUGHES, J.P.; GUTTORP, P. (1994). A Class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*, v.30, n.5, p.1535-1546.
- JONES, R.H.; BRELSFORD, W.M. (1967). Time series with periodic structure. *Biometrika*, v.54, n.3/4, p.403-408.
- KITE, G.W. (1991). *Hydrologic applications: computer programs for water resources engineering*. Fort Collins, Water Resources.

- LINSLEY, R.K.; FRANZINI, J.B. (1978). *Engenharia de recursos hídricos*. Trad.por Luiz Americo Pastorino. São Paulo, EDUSP/ McGraw-Hill do Brasil.
- MONTANARI, A.; ROSSO, R.; TAQQU, M.S. (1997). Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation. *Water Resources Research*, v.33, n.5, p.1035-1044.
- OZAKI, T. (1985). Statistical identification of storage models with application to stochastic hydrology. *Water Resources Bulletin - American Water Resources Association*, v.21, n.4, p.663-675.
- PAPOULIS, A. (1965). *Probability, random variables and stochastic process*. New York, MacGraw Hill. p.583.
- PEREIRA, M.V.F. (1985). *Optimal scheduling of hydrothermal systems – an overview*. Rio de Janeiro, IFAC Electric Energy Systems. p.1-9.
- POLTRONIERE SILVA, S.C. (1999). *Comparação entre as soluções de malha fechada e semi-aberta para otimização da operação de sistemas de produção de energia elétrica*. São Carlos. 102 p. Dissertação (Mestrado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo
- SALAS, J.D.; BOES, D.C.; SMITH, R.A. (1982). Estimation ARMA model with seasonal parameters. *Water Resources Research*, v.18, n.4, p.1006-1010.
- SALAS, J.D.; OBEYSEKERA, J.T.B. (1982). ARMA model identification of hydrologic time series. *Water Resources Research*, v.18, n.4 , p.1011-1021.
- SHIBATA, R. (1976). Selection of the order of an autoregressive model by akaike's information criterion. *Biometrika*, v.63, n.1, p.117-126.
- SOARES, S. (1987). Planejamento da operação de sistemas hidrotérmicos. *Revista SBA: controle e automação*, v.1, n.2, p.122-127.
- WILKS, D.S. (1989). Conditioning sthochastic daily precipitation models on total monthly precipitation. *Water Resources Research*, v.25, n.6, p.1429-1439.

WOOLHISER, D.A.; ROLDÁN, J. (1982). Stochastic daily precipitation models 2: a comparison of distributions of amounts. *Water Resources Research*, v.18, n.5, p.1461-1468.