

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-US			
	Data de Depósito:		
	Assinatura:		

Modelos autorregressivos periódicos para previsão e geração de séries de vazões médias mensais

Ricardo Luis dos Reis

Orientador: Prof. Dr. Marinho Gomes de Andrade Filho

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências - Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*.

USP – São Carlos Abril de 2013

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi e Seção Técnica de Informática, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

Reis, Ricardo Luis dos

R347m

Modelos autorregressivos periódicos para previsão e geração de séries de vazões médias mensais / Ricardo Luis dos Reis; orientador Marinho Gomes de Andrade Filho. -- São Carlos, 2013. 108 p.

Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2013.

1. Processos periódicos. 2. Séries temporais. 3. Inferência bayesiana. 4. Hidrologia. 5. Métodos computacionais. I. Andrade Filho, Marinho Gomes de , orient. II. Título.

A Deus e a Santo Expedito, a meus pais, Luzia e José Vicente, a minha irmã, Maria Aparecida e ao meu sobrinho, Pedro. Obrigado por tudo!

Agradecimentos

Agradeço a Deus e a Santo Expedito, por sempre guiarem meus passos aonde quer que eu vá. A meus pais, Luzia e José Vicente, a minha irmã, Maria Aparecida e ao meu sobrinho, Pedro, pelo carinho, amor, alegria e incentivo em todos os momentos da minha vida. Não posso esquecer também dos outros integrantes da família animal, Peninha, Mel, Bila, Pipoca, Nina, Chiquinha e Chico. Que Deus abençoe todos vocês!

Ao professor Marinho, pela orientação, confiança, amizade, ensinamentos e, sobretudo, pelo exemplo.

Aos professores e funcionários do Instituto de Ciências Matemáticas e de Computação e ao Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional.

Aos colegas e amigos do mestrado e do doutorado do ICMC e da UFSCar pelos grandes momentos de alegrias e estudos. Vocês fazem parte dessa vitória.

A Júlia, que me acolheu em sua pensão e que morei por todo o meu doutorado. Sua força aos mais de 80 anos sempre foi uma fonte inspiradora. A Juliana e sua família, Luiz, Isaura e Luiz Felipe. Aos companheiros de pensão, em especial a Vanessa e a Teca.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo apoio com bolsas de estudo e reserva técnica.

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

A todos que, de forma direta ou indireta, me auxiliaram neste trabalho.

Este trabalho aborda o problema de modelagem de séries de vazões médias mensais visando previsão e geração de séries sintéticas. Destaca-se que a importância da previsão de valores futuros das séries mensais de vazões assim como a geração de séries sintéticas são fundamentais para o planejamento da operação de sistemas hidroelétricos brasileiros. Estas séries possuem um comportamento periódico na média, na variância e na função de autocorrelação e, portanto, considera-se para a série padronizada os modelos autorregressivos periódicos $PAR(p_m)$. Em relação à previsão clássica, a análise do erro de previsão é feita em função do horizonte de previsão. Neste estudo, os erros de previsão são calculados, na escala original da série de vazão, em função dos parâmetros dos modelos ajustados e avaliados para horizontes de previsão h variando de 1 a 12 meses. Estes erros são comparados com as estimativas das variâncias das vazões para o mês que está sendo previsto. Em relação à previsão bayesiana, adota-se os modelos Normal, Log-Normal e t-Student nos processos de estimação e, após, é realizado um estudo da perfomance destes modelos usando o erro quadrático médio, erro absoluto médio e erro percentual absoluto médio. Em relação à geração de séries sintéticas de vazões, um modelo multivariado Log-Normal com três parâmetros e um modelo Log-Normal generalizado foram desenvolvidos. As séries geradas são comparadas com as séries históricas reais utilizando o critério de Kullback-Leibler. Como resultado, tem-se uma avaliação da capacidade de previsão, em meses, dos modelos ajustados para cada mês e a escolha do modelo Log-Normal nos procedimentos de análise bayesiana. Além disso, o modelo utilizado para a geração de séries sintéticas de vazões mensais forneceu evidências que o apontam como uma alternativa ao modelo amplamente adotado no setor elétrico brasileiro para geração de séries de vazões.

This work addresses the problem of forecasting and generation series monthly average streamflows. It is noteworthy that the importance of forecasting future values of the series of monthly streamflows as well as the generation of synthetic series are fundamental for planning the operation of Brazilian hydroelectric systems. These series have a periodic behavior on average, variance and autocorrelation function and therefore it is considered for standard series periodic autoregressive models $PAR(p_m)$. At the forecast classical analysis of the prediction error is made in function of the prediction horizon. In this study, the forecasting errors are calculated in the original scale of the series of streamflow, depending on the model parameters adjusted and evaluated for forecasting horizons h ranging from 1 to 12 months. These errors are compared with estimates of the variances of the streamflows for the month is provided. Regarding the bayesian prediction, we adopt the models Normal, Log-Normal and t-Student in estimation procedures and, then, is a study of the performance of these models using the mean square error, mean absolute error and mean absolute percentage error. In relation to generation, a Log-Normal multivariate model with three parameters and a Log-Normal generalized model were developed and analyzed using the Kullback-Leibler criterion. As a result there has been an assessment of the predictive power, in months, the adjusted models for each month, the choice of the

Log-Normal model in the procedures for bayesian analysis and the model used to generate synthetic series of monthly streamflows provided evidence that point as an alternative model adopted in the Brazilian electric sector.

Sumário

1	Intr	oduça	0	15
	1.1	Consid	derações iniciais	15
	1.2	Objet	ivos	21
	1.3	Organ	ização	22
2	Mo	delo A	utorregressivo Periódico - Abordagem Clássica	23
	2.1	Introd	lução	23
	2.2	Identi	ficação da ordem	24
	2.3	Estim	ação dos parâmetros	27
	2.4	Previs	são clássica	28
		2.4.1	Introdução	28
		2.4.2	MSE da previsão para séries de vazões sazonais	30
		2.4.3	MSE com um processo $PAR(p_m)$	33
3	Mo	delo A	utorregressivo Periódico - Abordagem Bayesiana	39
	3.1	Inferê	ncia bayesiana	39
		3.1.1	Distribuição a priori não informativa	41
	3.2	Métod	lo de Monte Carlo via Cadeias de Markov	41
		3.2.1	Algoritmo de Metropolis-Hastings	42
		3.2.2	Avaliação da convergência	42
	3.3	Estim	ação dos parâmetros	43
		3.3.1	Modelo Normal	44
		3.3.2	Modelo Log-Normal	45

10 SUMÁRIO

		3.3.3 Modelo <i>t</i> -Student	45
	3.4	Seleção de modelos	47
	3.5	Previsão bayesiana	49
		3.5.1 Densidade preditiva um passo a frente	49
4	Ger	ação de Séries Sintéticas	53
	4.1	Introdução	53
	4.2	Modelo Log-Normal com três parâmetros	54
	4.3	Modelo Log-Normal com três parâmetros multivariado	57
	4.4	Modelo Log-Normal generalizado	65
		4.4.1 Análise bayesiana	65
		4.4.2 Geração univariada de série sintética	67
	4.5	Critério de validação da série gerada	67
5	\mathbf{Apl}	icação	71
	5.1	Considerações iniciais	71
	5.2	Abordagem clássica dos modelos PAR	72
	5.3	Abordagem bayesiana dos modelos PAR	81
	5.4	Geração de séries sintéticas	89
6	Con	clusões e Propostas Futuras	96
\mathbf{R}	eferê	ncias 10)3

Lista de Tabelas

Ordens dos modelos para as vazões de Furnas	73
Modelos ajustados com o método clássico para as vazões de Furnas	73
Razão RMSE/MLT (em %) com $h=1$ para as vazões de Furnas	74
Ordens dos modelos para as vazões de Emborcação	76
Modelos ajustados com o método clássico para as vazões de Emborcação.	76
Razão RMSE/MLT (em %) com $h=1$ paras as vazões de Emborcação.	77
Ordens dos modelos para as vazões de Sobradinho	78
Modelos ajustados com o método clássico para as vazões de Sobradinho.	79
Razão RMSE/MLT (em %) com $h=1$ para as vazões de Sobradinho	79
Ordem fornecida pelo critério BIC para Água Vermelha	82
Ordem fornecida pelo critério BIC para São Simão	82
Estimativas pontuais para o modelo Normal do reservatório Água Ver-	
melha	82
Estimativas pontuais para o modelo Normal do reservatório São Simão.	83
Estimativas pontuais para o modelo Log-Normal do reservatório Água	
Vermelha	83
Estimativas pontuais para o modelo Log-Normal do reservatório São	
Simão	84
Estimativas pontuais para o modelo t-Student do reservatório Água Ver-	
melha	84
Estimativas pontuais para o modelo t -Student do reservatório São Simão.	85
Métodos de seleção do modelo para Água Vermelha	85
Métodos de seleção do modelo para São Simão	85
	Modelos ajustados com o método clássico para as vazões de Furnas Razão RMSE/MLT (em %) com $h=1$ para as vazões de Furnas Ordens dos modelos para as vazões de Emborcação Modelos ajustados com o método clássico para as vazões de Emborcação. Razão RMSE/MLT (em %) com $h=1$ paras as vazões de Emborcação. Ordens dos modelos para as vazões de Sobradinho

5.20	Erros de previsão para o modelo Normal de Água Vermelha	86
5.21	Erros de previsão para o modelo Log-Normal de Água Vermelha	86
5.22	Erros de previsão para o modelo t -Student de Água Vermelha	87
5.23	Erros de previsão para o modelo Normal de São Simão	87
5.24	Erros de previsão para o modelo Log-Normal de São Simão	88
5.25	Erros de previsão para o modelo t -Student de São Simão	88
5.26	Divergência de Kullback-Leibler (KL)	94
5.27	Estimativas pontuais para os parâmetros de um modelo PAR(1)	95

Lista de Figuras

1.1	Vazões médias mensais da Usina de Agua Vermelha	16
1.2	Vazões médias mensais da Usina de São Simão	16
1.3	Gráfico de dispersão da série de Água Vermelha de janeiro e fevereiro	17
1.4	Gráfico de dispersão da série de Água Vermelha de abril e maio	17
5.1	RMSE da previsão da vazão de Furnas para o mês de janeiro	75
5.2	RMSE da previsão da vazão de Furnas para o mês de agosto	75
5.3	RMSE da previsão da vazão de Emborcação para o mês de fevereiro	77
5.4	RMSE da previsão da vazão de Emborcação para o mês de julho	78
5.5	MSE da previsão da vazão para o mês de março para o posto de Sobradinho.	80
5.6	MSE da previsão da vazão para o mês de setembro para o posto de	
	Sobradinho	80
5.7	Previsão do modelo Log-Normal de Água Vermelha (2008-2010)	89
5.8	Previsão do modelo Log-Normal de São Simão (2008-2010)	89
5.9	Média mensal para o reservatório de Itaipu	90
5.10	Desvio padrão mensal para o reservatório de Itaipu	90
5.11	Coeficiente de assimetria mensal para o reservatório de Itaipu	91
5.12	Coeficiente de curtose mensal para o reservatório de Itaipu	91
5.13	Correlação espacial entre os reservatórios de Itaipu e Xingó	92
5.14	Correlação espacial entre os reservatórios de Itaipu e Itutinga	92
5.15	Correlação espacial entre os reservatórios de Corumbá IV e Espora.	92
5.16	Correlação espacial entre os reservatórios de Corumbá I e Corumbá IV.	93
5.17	Séries históricas e sintéticas para Itaipu	94

5.18	Média Mensal para séries histórica e sintética de Furnas	95
5.19	Desvio padrão mensal para séries histórica e sintética de Furnas	96
5.20	Coeficiente de assimetria para séries histórica e sintética de Furnas	96
5.21	Coeficiente de curtose para séries histórica e sintética de Furnas	96
5.22	Série histórica e sintética de Furnas	97

Capítulo

1

Introdução

O objetivo deste capítulo é apresentar uma revisão bibliográfica sobre o estado da arte em modelos de previsão e geração de séries temporais. São apresentados os aspectos de motivação para a abordagem do problema proposto e detalhados os objetivos, destacando-se a contribuição do trabalho no contexto dos modelos para séries temporais.

1.1 Considerações iniciais

A análise estatística dos modelos de séries temporais tem aplicações em diferentes áreas da ciência, como engenharia, economia, meteorologia, entre outras. Os modelos autorregressivos e de médias móveis (ARMA) são bem conhecidos e considerados adequados para analisar séries estacionárias (Box $et\ al.,\ 2008$). No entanto, nem todas as séries temporais encontradas na natureza podem ser modeladas como processos ARMA. Em algumas situações práticas, encontram-se séries temporais que apresentam um comportamento sazonal ou periódico. Assim, diz-se que uma série temporal apresenta um comportamento sazonal com período s, quando ocorrem similaridades na série após s intervalos de tempo.

Para ilustrar, as figuras (1.1) e (1.2) apresentam as séries sazonais de vazões médias mensais (m^3/s) das usinas hidrelétricas de Água Vermelha e São Simão entre os anos de 2001 a 2010.

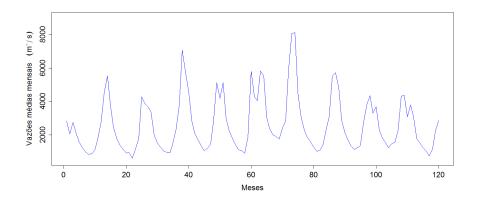


Figura 1.1: Vazões médias mensais da Usina de Água Vermelha.

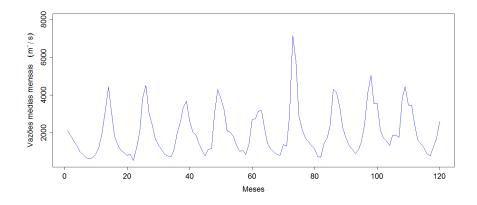


Figura 1.2: Vazões médias mensais da Usina de São Simão.

Para estas séries sazonais, muitas vezes a correlação entre observações vizinhas, Z_t e Z_{t-1} , dependem de t. Nestes casos, é mais adequado ajustar um modelo cujos parâmetros também variam de forma periódica em vez de um modelo com parâmetros constantes para todos os instantes t. Então, quando a correlação entre os meses se modifica de forma periódica, estas séries são ditas periodicamente estacionárias (McLeod, 1994). Segundo Tiao & Grupe (1980), os processos periódicos podem ser especificados erroneamente pelos modelos SARMA (autorregressivos e de médias móveis sazonais), pois nestes modelos é feita a hipótese de que as correlações entre os valores ocorridos entre meses consecutivos é constante independente do mês e um estudo com intuito de identificação da correlação periódica deve ser realizado.

Segundo Hurd & Gerr (1991) a correlação periódica pode ser analisada por meio dos gráficos de dispersão de diferentes períodos. Como ilustração, as figuras (1.3) e (1.4) apresentam as dispersões dos meses de janeiro e fevereiro e de abril e maio de Água Vermelha. Como pode ser visto no gráfico, a hipótese de existência de correlação periódica pode ser facilmente visualizada. Para casos mais extremos utiliza-se os testes propostos por Hurd & Gerr (1991) e Vecchia & Ballerini (1991).

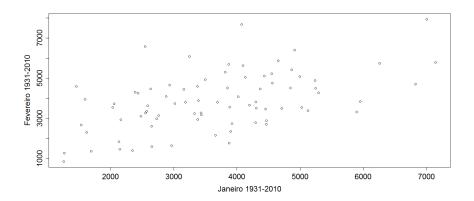


Figura 1.3: Gráfico de dispersão da série de Água Vermelha de janeiro e fevereiro.

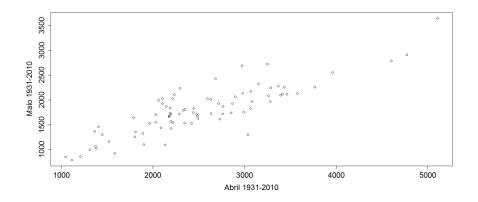


Figura 1.4: Gráfico de dispersão da série de Água Vermelha de abril e maio.

Os processos periódicos foram propostos originalmente em Gladysev (1961) e várias aplicações desses modelos nas áreas de hidrologia e econometria são apresentadas em Parzen & Pagano (1979), Salas *et al.* (1982), Vecchia (1985a,b), Salas & Obeysekera (1988), McLeod (1994) e Rasmussen *et al.* (1996).

Uma classe de processos periódicos muito pesquisada é a dos modelos autorregressivos periódicos (PAR), estudados inicialmente por Thomas & Fiering (1962) e Jones & Brelsford (1967) e amplamente analisados em Pagano (1978), onde foram apresentadas as propriedades assintóticas destes modelos assumindo normalidade do processo

e Troutman (1979) que mostra as condições necessárias e suficientes para a estacionariedade periódica.

A estacionariedade periódica dos modelos PAR conduz às função de autocorrelação e autocorrelação parcial periódicas. Nesse sentido, vale ressaltar os trabalhos de Anderson & Vecchia (1993), que desenvolveram alguns resultados assintóticos para a função de autocorrelação periódica amostral e Sakai (1982), que estudou a função de autocorrelação parcial periódica dos processos PAR. Outras propriedades dos modelos PAR e suas relações com os modelos SARMA podem ser encontradas em Tiao & Grupe (1980) e Salas et al. (1982).

A geração de energia elétrica no Brasil é predominantemente hidráulica e o setor elétrico é o responsável pelo gerenciamento da água armazenada nos reservatórios do país. Estes reservatórios, antes destinados exclusivamente à geração de energia, atualmente são alvos de interesses diversos e conflitantes. Sendo assim, a solução de problemas envolvendo projetos de gestão, gerenciamento e planejamento dos recursos hídricos é de grande importância.

Nos modelos de Planejamento da Operação de Sistemas Hidrotérmicos (POSH), uma das variáveis mais importantes é a vazão afluente aos aproveitamentos hidrelétricos, sendo que estas vazões interferem significativamente na operação destes sistemas. As vazões nos rios dependem, entre outros fatores, das condições climáticas, o que leva a atribuir às vazões um modelo estocástico. Visto que as decisões futuras, ou seja, a política de operação, depende dos cenários destas vazões, torna-se fundamental adotar um modelo estocástico adequado para calcular a política ótima de operação destes sistemas.

Assim, o processo de produção de energia de um sistema hidroelétrico depende da série de vazões afluentes, sendo esta operação realizada de forma otimizada para atender a demanda de energia. As séries de vazões naturais médias mensais se caracterizam por um comportamento periódico, ou seja, a média, a variância, a assimetria e a estrutura de autocorrelação apresentam características periódicas. Neste trabalho, os modelos PAR são aplicados na modelagem das séries de vazões médias mensais. Uma revisão detalhada de algumas características de séries temporais sazonais, tais como tendências e periodicidades, aplicadas em Hidrologia são apresentadas em Machiwal & Jha (2009).

Em McLeod (1994) encontra-se uma descrição sobre as etapas de desenvolvimento dos modelos PAR. O processo de modelagem dos modelos PAR inicia-se com a identificação da ordem p, para cada mês m, ou seja, o número de parâmetros autorregressivos em cada mês. No procedimento de identificação, Stedinger & Taylor (1982) e Ula & Smadi (2003) consideraram a identificação da ordem dos modelos PAR usando a função de autocorrelação parcial periódica. Bentarzi $et\ al.\ (2008)$ consideraram uma adapta-

ção do critério de densidade preditiva, via abordagem bayesiana, para o problema de seleção da ordem dos modelos PAR. Além destes métodos, outros critérios para identificação da ordem são o critério de informação bayesiano (Schwarz, 1978), o método de *Bootstrap* (Oliveira, 2011), entre outros.

No processo de estimação dos parâmetros do modelo PAR, a utilização de métodos clássicos é bastante popular, destacando-se o método dos momentos, da máxima verossimilhança e de mínimos quadrados. Pagano (1978) considerou a estimação pelo método dos momentos para os parâmetros do modelo PAR.

No trabalho desenvolvido por Vecchia (1985b) apresentou-se resultados úteis sobre a estrutura de correlação e estimativa de parâmetros apresentando algumas propriedades de correlação, discutindo o método dos momentos e máxima verossimilhança e mostrando uma aplicação destas técnicas. Esse autor também desenvolve uma aproximação para a verossimilhança exata para processos PAR normais e implementou um algoritmo bastante simples para a sua maximização.

Outro algoritmo, desenvolvido por Anderson et al. (1999), para a estimação dos processos periodicamente estacionário é utilizado por esses autores em vários trabalhos subsequentes. Shao & Ni (2004) estudaram as propriedades assintóticas dos estimadores de mínimos quadrados. Trabalhos mais recentes têm utilizado a representação dos modelos PAR no domínio da frequência visando o ajuste de modelos parcimoniosos (Lund et al., 2006; Tesfaye et al., 2011) e na presença de outliers aditivos (Sarnaglia et al., 2010).

Outro método que tem sido bastante empregado para inferência em séries temporais, inclusive séries hidrológicas, são os métodos bayesianos (Barreto & Andrade, 2004). Entre as vantagens da abordagem bayesiana destacam-se, principalmente, a incorporação de informações a priori sobre os parâmetros do modelo e o uso de técnicas de simulação de Monte Carlo via cadeias de Markov (MCMC) (Gelfand & Smith, 1990; Gelfand et al., 1990), que permite fazer inferência de parâmetros em modelos mais complexos. Após a fase de estimação dos modelos PAR, as próximas etapas se referem à previsão e geração de séries sintéticas.

Em relação à previsão, este estudo faz uma análise do erro de previsão em função do horizonte de previsão para modelos PAR. O uso de modelos estocásticos para previsão de valores de uma série temporal tem como objetivo fazer a previsão com o menor erro possível. Nesta tese, adota-se o erro médio quadrático da previsão, MSE (Mean Squared Error) como critério para avaliar a performance das previsões. A vantagem deste critério é que ele pode ser comparado com a variância incondicional da série.

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

Vários trabalhos retratam a previsão de séries temporais tais como Noakes et al. (1985) e Thompstone et al. (1985), que demonstraram a superioridade dos modelos PAR. Ula (1993) apresentou a previsão do erro quadrático médio mínimo e Lund & Basawa (2000), que explorou o cálculo da previsão recursiva e do erro quadrático médio. Outros modelos de previsão e a comparação do desempenho com outras técnicas, tais como redes neurais artificiais (Ballini et al., 2003) ou abordagens bayesianas como em Wang (2006), Wang et al. (2009) e Wang & Robertson (2011) podem ser encontradas na literatura.

Em relação à geração de séries sintéticas, pode-se afirmar que os modelos geradores de séries de vazões são de grande utilidade na comparação de diferentes técnicas de planejamento do uso dos recursos hídricos renováveis. Um modelo de geração de série de vazões deve reproduzir da forma mais fiel possível características estocásticas dos dados históricos de vazões. Dentre estas características destacam-se a função de autocorrelação e a distribuição de probabilidade das vazões. Além disso, um modelo de geração de séries de vazões deve permitir a geração de vazões sem exigir grande esforço computacional, pois o procedimento de geração deverá ser repetido um grande número de vezes para proporcionar um grande número de séries longas.

O gerador de série sintética univariada tem com objetivo gerar séries de vazões para estudos envolvendo uma única usina hidrelétrica ou um único subsistema. O gerador multivariado, por sua vez, considera a correlação espacial presente em sistemas com múltiplos reservatórios. Quando estes vários reservatórios não apresentam correlação espacial, o gerador de série sintética univariada pode ser empregado para gerar séries sintéticas para cada reservatório independentemente. Vale ressaltar que o uso do gerador multivariado deve apresentar resultados semelhantes nestes casos.

Um dos fatores que dificulta a construção dos modelos de geração de vazões é a presença de assimetria na distribuição de probabilidade das vazões. Para considerar modelos com assimetria e com facilidade de geração de valores, vários autores propuseram o uso da distribuição Log-Normal com três parâmetros (Stedinger & Taylor, 1982; Pereira et al., 1984). Esta distribuição é usada pelo modelo GEVAZP, que é um modelo estocástico multivariado de geração de séries sintéticas de vazões incrementais e totais, que se baseia no modelo PAR. Este modelo é amplamente utilizado no Sistema Interligado Nacional (CEPEL, 2002).

1.2 Objetivos 21

O trabalho de Tesfaye et al. (2006) propõe o uso de mais de uma distribuição, ou seja, um modelo de mistura de distribuições para modelar séries de vazões. Assim, os autores utilizam uma distribuição Log-Normal para as vazões intermediárias e uma distribuição de Pareto truncada para modelar as caudas das distribuições. Outras técnicas também têm sido utilizadas nos procedimentos de geração de séries sintéticas, como as técnicas Bootstrap (Kim et al., 2004), que, de acordo com os autores, preservam o coeficiente de assimetria dos dados históricos, modelos com desagregação (Valencia & Schaake, 1973; Mondal & Wasimi, 2007), redes neurais artificiais (Ahmed & Sarma, 2007), entre outras. Muitos trabalhos compararam estas técnicas, como, por exemplo, Lohani et al. (2012), que comparou as técnicas de redes neurais artificiais, sistemas neuro-fuzzy e modelos PAR.

O fato de encontrar um grande número de trabalhos que adotam a distribuição Log-Normal na área de Hidrologia motivou, neste trabalho, o desenvolvimento de um gerador bayesiano univariado usando a distribuição Log-Normal generalizada, que é construída a partir de uma distribuição Normal generalizada (Nadarajah, 2005). Alguns trabalhos, na área de análise de sobrevivência, têm utilizado esta distribuição usando uma abordagem bayesiana. Nestes trabalhos, os autores utilizam distribuições a priori informativas e não-informativas (Martín & Pérez, 2009) e em Herández & Usuga (2011) são consideradas apenas não-informativas.

A literatura apresenta um grande número de modelos estocásticos multivariados para séries temporais (Pereira et al., 1984), ainda que para o caso de séries temporais sazonais esse número seja bem menor, em que se pode destacar o artigo de Salas & Pegram (1977). O principal motivo desta limitação é que estes modelos são não-parcimoniosos, com identificação complexa e com estimação de parâmetros sujeita a uma grande imprecisão (Vieira et al., 1998). Outra boa referência sobre a utilização destes modelos multivariados em Hidrologia é Camacho et al. (1987).

1.2 Objetivos

Os objetivos desta tese são:

a) Contribuir para o problema de previsão clássica de vazões médias mensais mediante estudos sobre o erro quadrático médio. Esse problema é importante para o setor elétrico brasileiro, pois ajuda no planejamento da operação ótima do sistema hidroelétrico brasileiro, que é responsável por mais de 90% da energia elétrica consumida no Brasil;

- b) Dar uma contribuição bayesiana sobre os procedimentos de identificação, estimação, seleção e previsão dos modelos PAR. Em hidrologia, a adoção da estimação bayesiana ainda é bem escassa e espera-se que os resultados demonstrem as vantagens e flexibização deste procedimento pelo uso de técnicas de simulação MCMC;
- c) Contribuir para a solução do problema de geração de vazões médias mensais. Nesse sentido, um modelo multivariado para geração de séries sintéticas é desenvolvido e comparado com o modelo GEVAZP, utilizado no SIN. Também é apresentado um modelo gerador univariado Log-Normal generalizado via abordagem bayesiana. Utilizase a divergência de Kullback-Leibler como um método de validação das séries geradas.

1.3 Organização

Esta tese está organizada da seguinte maneira:

O capítulo 2 introduz os modelos PAR, abordando os procedimentos clássicos de identificação, estimação e previsão. Um estudo assintótico sobre o erro quadrático médio de previsão também é desenvolvido.

O capítulo 3 apresenta a teoria sobre inferência bayesiana e, após, é realizada uma análise dos modelos PAR usando os procedimentos de estimação e previsão.

O capítulo 4 trata da geração de séries sintéticas. Primeiramente, descreve-se o modelo Log-Normal com três parâmetros e, após, é desenvolvido um modelo multivariado usando esta distribuição. Após, é desenvolvido, via abordagem bayesiana, um modelo Log-Normal generalizado para geração de séries sintéticas univariadas. Finalmente, um critério de validação das séries geradas, a divergência de Kullback-Leibler, é apresentado.

O capítulo 5 apresenta estudos de caso para os problemas de estimação e previsão, usando as abordagens clássicas e bayesianas, dos modelos PAR. Também é realizado estudos sobre a geração de vazões médias mensais.

O capítulo 6 destaca as principais conclusões deste trabalho, com possíveis tópicos para trabalhos futuros.

Capítulo

2

Modelo Autorregressivo Periódico - Abordagem Clássica

Neste capítulo é apresentado o modelo autorregressivo periódico. Primeiramente, são definidos os conceitos iniciais do modelo e, a seguir, são apresentadas as etapas de identificação da ordem, estimação e previsão do modelo. Finalmente, é desenvolvido um estudo assintótico do erro quadrático médio de previsão.

2.1 Introdução

Algumas séries temporais históricas, tais como as hidrológicas sazonais, exibem uma estrutura de autocorrelação que depende não somente do intervalo de tempo entre as observações, mas também do período observado (Hipel & McLeod, 1994). Na classe dos modelos periódicos, se destaca o modelo autorregressivo PAR (periodic autoregressive). O modelo PAR (p_m) ajusta para cada mês m um modelo AR (p_m) , em que p_m representa a ordem do modelo, ou seja, o número de termos autorregressivos do modelo em cada mês.

O modelo $PAR(p_m)$ é descrito matematicamente por:

$$\left(\frac{x_{t(r,m)} - \mu_m}{\sigma_m}\right) = \sum_{j=1}^{p_m} \phi_{j,m} \left(\frac{x_{t(r,m)-j} - \mu_{m-j}}{\sigma_{m-j}}\right) + a_{t(r,m)},$$
(2.1)

em que $x_{t(r,m)}$ é a vazão média mensal para o ano r e o mês m, μ_m é a média do mês m, σ_m é o desvio padrão do mês m, $\phi_{1,m}, \cdots, \phi_{p_m,m}$ são os coeficientes do modelo autorregressivo para cada mês m e $a_{t(r,m)}$ é um ruído não-correlacionado para o ano r e o mês m com $E[a_{t(r,m)}] = 0$ e $Var[a_{t(r,m)}] = \sigma_m^{(a)2}$ dada por:

$$\sigma_m^{(a)2} = \sigma_m^2 \left(1 + \sum_{j=1}^{p_m} \rho_j^m \right),$$
 (2.2)

em que ρ_j^m é a autocorrelação entre $x_{t(r,m)}$ e $x_{t(r,m)-j}$, dada por:

$$\rho_j^m = E\left[\left(\frac{x_{t(r,m)} - \mu_m}{\sigma_m}\right) \left(\frac{x_{t(r,m)-j} - \mu_{m-j}}{\sigma_{m-j}}\right)\right]. \tag{2.3}$$

Na sequência deste trabalho, usa-se a seguinte notação para representar a série temporal periódica padronizada:

$$z_{t(r,m)} = \left(\frac{x_{t(r,m)} - \mu_m}{\sigma_m}\right). \tag{2.4}$$

Esta padronização tem como objetivo obter uma série sem sazonalidade. Segundo Salas et al. (1982), esses processos estocásticos naturais são, em geral, estacionários no sentido lato, isto é, os momentos de primeira e segunda ordem da distribuição de probabilidades não são afetados por variações devido à escolha da origem dos tempos.

2.2 Identificação da ordem

Nesta tese são utilizados três critérios para identificação da ordem do modelo PAR. A identificação da ordem p, em cada mês m, para o modelo PAR (p_m) , geralmente é realizada por meio da análise da função de autocorrelação parcial (PACF). Inicialmente, define-se a função de autocorrelação (ACF) da série, ou seja, considere ρ_j^m a correlação entre $z_{t(r,m)}$ e $z_{t(r,m)-j}$ dada por:

$$\rho_j^m = E\left[\left(\frac{x_{t(r,m)} - \mu_m}{\sigma_m}\right) \left(\frac{x_{t(r,m)-j} - \mu_{m-j}}{\sigma_{m-j}}\right)\right] = E\left[z_{t(r,m)} z_{t(r,m)-j}\right]. \tag{2.5}$$

O conjunto de funções de autocorrelação ρ_j^m dos períodos $m=1,\cdots,s$, descrevem a estrutura de dependência temporal da série. Estas funções são dadas por:

$$\rho_j^m = \phi_{1,m} E\left[z_{t(r,m)-1} z_{t(r,m)-j}\right] + \dots + \phi_{p_m,m} E\left[z_{t(r,m)-p_m} z_{t(r,m)-j}\right], \quad j \ge 1, \quad (2.6)$$

em que $\rho_0^m = 1$.

Fixando-se o mês m e variando j de 1 a p_m em (2.6) encontra-se para cada mês m, um conjunto de equações, comumente denominadas de equações de Yule-Walker. Para um mês m qualquer:

$$\begin{bmatrix} 1 & r_1^{m-1} & r_2^{m-1} & \cdots & r_{p_m-1}^{m-1} \\ r_1^{m-1} & 1 & r_1^{m-2} & \cdots & r_{p_m-2}^{m-2} \\ r_2^{m-1} & r_1^{m-2} & 1 & \cdots & r_{p_m-3}^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p_m-1}^{m-1} & r_{p_m-2}^{m-2} & r_{p_m-3}^{m-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \widehat{\phi}_{1,m} \\ \widehat{\phi}_{2,m} \\ \widehat{\phi}_{3,m} \\ \vdots \\ \widehat{\phi}_{p_m,m} \end{bmatrix} = \begin{bmatrix} r_1^m \\ r_2^m \\ r_3^m \\ \vdots \\ r_{p_m}^m \end{bmatrix}$$

Para facilitar as notações, omite-se o período m e denota-se por ϕ_{kj} , o j-ésimo parâmetro autorregressivo de ordem k, em que ϕ_{kk} representa o último parâmetro deste processo. As equações de Yule-Walker, para cada período m, podem ser reescritas da seguinte forma:

$$\begin{bmatrix} 1 & r_1^{m-1} & r_2^{m-1} & \cdots & r_{p_m-1}^{m-1} \\ r_1^{m-1} & 1 & r_1^{m-2} & \cdots & r_{p_m-2}^{m-2} \\ r_2^{m-1} & r_1^{m-2} & 1 & \cdots & r_{p_m-3}^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p_m-1}^{m-1} & r_{p_m-2}^{m-2} & r_{p_m-3}^{m-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \widehat{\phi}_{k1} \\ \widehat{\phi}_{k2} \\ \widehat{\phi}_{k3} \end{bmatrix} = \begin{bmatrix} r_1^m \\ r_2^m \\ \widehat{\phi}_{k3} \\ \vdots \\ \widehat{\phi}_{kk} \end{bmatrix}$$

Os valores dos parâmetros ϕ_{kk} , k=1,2,... representam a função de autocorrelação parcial no mês m. Em um processo $PAR(p_m)$, a função de autocorrelação parcial é o valor do último parâmetro ϕ_{kk} e será diferente de zero para k menor ou igual a p_m e zero para k maior que p_m .

Portanto, a identificação de um modelo $PAR(p_m)$ consiste em determinar as ordens p_m mais apropriadas dos operadores autorregressivos de cada mês, m=1,2,...,s. Isto pode ser feito obtendo-se as estimativas de ϕ_{kk} , k=1,2,...,N/4. Para testar a hipótese nula $H_0: \phi_{kk} = 0$ para todos os valores de k=1,2,..., utiliza-se o resultado assintótico

o qual afirma que, se a ordem do operador autorregressivo de um mês m qualquer é p_m , então o estimador $\widehat{\phi}_{kk}$ de ϕ_{kk} , para $k>p_m$, tem distribuição aproximadamente Normal com média zero e variância 1/n. Portanto, não se rejeita H_0 com nível de significância $\alpha<5\%$ se o zero está contido no intervalo $\widehat{\phi}_{kk}\pm2/\sqrt{n}$. Neste caso, afirma-se que a estimativa $\widehat{\phi}_{kk}$ é não significativa. Um procedimento de identificação da ordem p_m consiste em procurar, para cada mês m, a maior ordem p_m tal que todas as estimativas $\widehat{\phi}_{kk}$ para $k>p_m$ não sejam mais significativas.

Este procedimento clássico de identificação da ordem pode permitir a aceitação de processos como sendo ruídos brancos, quando, na verdade, trata-se de modelos que possuem baixos valores de ρ_k^m e ϕ_{kk} . Valores intermediários baixos de ϕ_{jj} para $j < p_m$ também podem ser considerados não significativos e isso possibilita a identificação de modelos com parâmetros intermediários nulos $\phi_{kj} = 0$, para alguns $j < p_m$.

Stedinger (2001) afirma que uma modelagem que considera individualmente cada mês não pode produzir um conjunto de modelos mais adequado. Sem maiores detalhes matemáticos e somente considerando a natureza dos fenômenos naturais envolvidos, o autor argumenta que não faria sentido considerar que a vazão de determinado mês dependa de vários meses anteriores, ou seja, o modelo $PAR(p_m)$ tem ordem elevada $p_m = 6$ e para o mês seguinte, m + 1, a ordem seja pequena $p_{m+1} = 1$. Assim como também não é razoável supor que, devido a parâmetros intermediários nulos, a vazão de um mês m dependa da vazão do mês m - 2 e não do mês m - 1.

Estes resultados corroboram a idéia de que a forma clássica de identificação das ordens p_m pode não ser a mais adequada. Stedinger (2001) propõe como critério para a seleção das defasagens j significantes, que para cada período m, encontre-se a maior ordem j tal que todas as estimativas $\widehat{\phi}_{kk}$ para k < j sejam significativas. Este critério não admite valores intermediários de $\widehat{\phi}_{kk}$ não significativos, o que ocorre na forma clássica.

Além dos dois procedimentos apresentados, considera-se também o critério BIC (Bayesian information criterion), proposto por Schwarz (1978) para selecionar as ordens p_m dos modelos autorregressivos de cada mês m. Neste caso, o critério para o cálculo de BIC_m é dado por:

$$BIC_m = n\log(\tau_m^2) + p_m\log(n). \tag{2.7}$$

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

2.3 Estimação dos parâmetros

Após a etapa de identificação da ordem p_m , é necessário obter as estimativas dos parâmetros do modelo PAR (p_m) . Segundo Hipel & McLeod (1994), o método dos momentos é tão eficiente quanto o método de máxima verossimilhança para processos autorregressivos. Os parâmetros $\phi_{1,m}, \phi_{2,m}, ..., \phi_{p_m,m}$ são estimados usando as equações de Yule-Walker. As estimativas de $\sigma_m^{(a)2}$ são obtidas usando-se a expressão dada em (2.2).

Os parâmetros $\Phi_m = \phi_{1,m}, \phi_{2,m}, ..., \phi_{p_m,m}$ e $\tau_m = \frac{1}{\sigma_m^{(a)2}}$ também são estimados pelo método de máxima verossimilhança. Assim, a função de verossimilhança é definida por:

$$L_m(\mathbf{\Phi}_m, \tau_m | \mathbf{Z}_m) \propto \tau_m^{(n-1)/2} \exp\left\{-\frac{\tau_m}{2}[(\mathbf{Z}_m - \mathbf{X}_m \mathbf{\Phi}_m)'(\mathbf{Z}_m - \mathbf{X}_m \mathbf{\Phi}_m)]\right\},$$
 (2.8)

em que

$$oldsymbol{Z}_m = egin{pmatrix} z_{t(1,m)} \ z_{t(2,m)} \ dots \ z_{t(n,m)} \end{pmatrix} \; ; \; oldsymbol{\Phi}_m = egin{pmatrix} \phi_{1,m} \ \phi_{2,m} \ dots \ \phi_{p_m,m} \end{pmatrix}$$

$$m{X}_m = egin{pmatrix} z_{t(1,m)-1} & z_{t(1,m)-2} & \cdots & z_{t(1,m)-p_m} \ z_{t(2,m)-1} & z_{t(2,m)-2} & \cdots & z_{t(2,m)-p_m} \ dots & dots & \ddots & dots \ z_{t(n,m)-1} & z_{t(n,m)-2} & \cdots & z_{t(n,m)-p_m} \end{pmatrix},$$

com \mathbf{Z}_m sendo uma matriz $n \ge 1$, \mathbf{X}_m sendo $p_m \ge n$ e $\mathbf{\Phi}_m$ sendo $p_m \ge 1$.

Assim, os estimadores de máxima verossimilhança para os parâmetros Φ_m e τ_m são dados por:

$$\hat{\boldsymbol{\Phi}}_{m} = (\boldsymbol{X}_{m}^{'} \boldsymbol{X}_{m})^{-1} (\boldsymbol{X}_{m}^{'} \boldsymbol{Z}_{m}) \tag{2.9}$$

$$\hat{\tau}_m^{-1} = (\boldsymbol{Z}_m - \boldsymbol{X}_m \hat{\boldsymbol{\Phi}}_m)' (\boldsymbol{Z}_m - \boldsymbol{X}_m \hat{\boldsymbol{\Phi}}_m)$$
 (2.10)

2.4 Previsão clássica

Um dos objetivos em análise de séries temporais refere-se à previsão, que é realizada após os procedimentos de identificação e estimação dos parâmetros do modelo PAR. Nesta fase, a avaliação da perfomance do modelo estimado pode ser realizada com o uso do erro quadrático médio de previsão. Este critério será o tema das próximas seções, em que é realizado um estudo assintótico em relação aos modelos PAR. Maiores detalhes sobre a previsão clássica podem ser encontrados em Hipel & McLeod (1994).

2.4.1 Introdução

Considerando o modelo $PAR(p_m)$ dado em (2.1) para o ano (n+h), $h \ge 1$, em que h representa o horizonte de previsão, tem-se:

$$z_{t(n+h,m)} = \sum_{j=1}^{p_m} \phi_{j,m} z_{t(n+h,m)-j} + a_{t(n+h,m)}.$$
 (2.11)

Usando a equação (2.11), $\hat{z}_{t(n+h,m)} = E(z_{t(n+h,m)}|\mathbf{Z})$ é dado por:

$$\hat{z}_{t(n+h,m)} = \sum_{j=1}^{p_m} \phi_{j,m} \hat{z}_{t(n+h,m)-j}, \qquad (2.12)$$

em que $\hat{z}_{t(n+h,m)-j} = z_{t(n+h,m)-j}$, se $t(n+h,m)-j < sn, j=1,...,p_m$. Para encontrar uma forma recursiva de calcular as previsões $\hat{z}_{t(n+h,m)-j}$, defini-se o vetor \mathbf{Z}_{fh} com os s valores futuros da série, para um ano (n+h) e por \mathbf{a}_{fh} os respectivos ruídos, como:

$$Z_{fh} = (z_{t(n+h,1)}, z_{t(n+h,2)}, ..., z_{t(n+h,s)})',$$
 (2.13)

$$\mathbf{a}_{fh} = (a_{t(n+h,1)}, a_{t(n+h,2)}, ..., a_{t(n+h,s)})'. \tag{2.14}$$

Denotando por $\Phi_m = (\phi_{1,m}, \phi_{2,m}, ..., \phi_{s,m})$, com m = 1, ..., s, pode-se escrever (2.11) na forma matricial por:

$$\mathbf{A}(\mathbf{\Phi}_m)\mathbf{Z}_{fh} + \mathbf{B}(\mathbf{\Phi}_m)\mathbf{Z}_{fh-1} = \mathbf{a}_{fh}, \tag{2.15}$$

em que $A(\Phi_m)$ e $B(\Phi_m)$ são matrizes $s \times s$ dadas por:

2.4 Previsão clássica 29

$$\boldsymbol{A}(\boldsymbol{\Phi}_m) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{1,1} & 1 & 0 & \cdots & 0 \\ -\phi_{1,2} & -\phi_{2,2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{s-1,s} & -\phi_{s-2,s} & -\phi_{s-3,s} & \cdots & 1 \end{pmatrix},$$

$$\boldsymbol{B}(\boldsymbol{\Phi}_m) = \begin{pmatrix} -\phi_{s,1} & -\phi_{s-1,1} & -\phi_{s-2,1} & \cdots & -\phi_{1,1} \\ 0 & -\phi_{s,2} & -\phi_{s-1,2} & \cdots & -\phi_{2,2} \\ 0 & 0 & -\phi_{s,3} & \cdots & -\phi_{3,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -\phi_{s+1,s} & \cdots & -\phi_{s,s} \end{pmatrix}.$$

Assim, tem-se:

$$Z_{fh} = -A(\Phi_m)^{-1}B(\Phi_m)Z_{fh-1} + A(\Phi_m)^{-1}a_{fh}.$$
 (2.16)

Denotando por \mathbf{Z}_{f0} as observações do último ano do histórico, ou seja:

$$\mathbf{Z}_{f0} = (z_{t(n,1)}, z_{t(n,2)}, ..., z_{t(n,12)})'$$
(2.17)

e escrevendo a equação (2.16) para h=1 tem-se uma condição inicial para a equação recursiva (2.16) dada por:

$$Z_{f1} = -A(\Phi_m)^{-1}B(\Phi_m)Z_{f0} + A(\Phi_m)^{-1}a_{f1}.$$
 (2.18)

Escrevendo (2.16) recursivamente, tem-se que:

$$\boldsymbol{Z}_{fh} = -\left[\boldsymbol{A}(\boldsymbol{\Phi}_m)^{-1}\boldsymbol{B}(\boldsymbol{\Phi}_m)\right]^h \boldsymbol{Z}_{f0} + \sum_{k=1}^h D_k(\boldsymbol{\Phi}_m)\boldsymbol{a}_{fh}, \qquad (2.19)$$

com

$$D_k(\mathbf{\Phi}_m) = \left[\mathbf{A}(\mathbf{\Phi}_m)^{-1} \mathbf{B}(\mathbf{\Phi}_m) \right]^{h-k} \mathbf{A}(\mathbf{\Phi}_m)^{-1}. \tag{2.20}$$

Assumindo que \boldsymbol{a}_{fk} , k=1,...,h são vetores aleatórios independentes com distribuição $N(0,\Gamma)$, $\Gamma=diag(\gamma_1^{-1},...,\gamma_s^{-1})$, então \boldsymbol{Z}_{fh} também é um vetor aleatório normalmente distribuído com média $\boldsymbol{E}(\boldsymbol{Z}_{fh}|\boldsymbol{Z}_{f0})$ e matriz de covariância $\boldsymbol{V}(\boldsymbol{Z}_{fh}|\boldsymbol{Z}_{f0})$.

No contexto clássico, a previsão para o ano (n+h) usando as observações do último ano n, ou seja, as previsões h-passos a frente são calculadas como:

$$\boldsymbol{E}(\boldsymbol{Z}_{fh}|\boldsymbol{Z}_{f0}) = -\left[\boldsymbol{A}(\boldsymbol{\Phi}_m)^{-1}\boldsymbol{B}(\boldsymbol{\Phi}_m)\right]^h \boldsymbol{Z}_{f0}.$$
 (2.21)

A variância de \boldsymbol{Z}_{fh} condicionada a \boldsymbol{Z}_{f0} é dada por:

$$V(\mathbf{Z}_{fh}|\mathbf{Z}_{f0}) = \sum_{k=1}^{h} D_k(\mathbf{\Phi}_m) \Gamma D_k(\mathbf{\Phi}_m)'. \tag{2.22}$$

As estimativas de $E(Z_{fh}|Z_{f0})$ e $V(Z_{fh}|Z_{f0})$ são calculadas substituindo-se nas equações (2.21) e (2.22) as estimativas de $\hat{\Phi}_m$ como:

$$\hat{\boldsymbol{Z}}_{fh} = -\left[\boldsymbol{A}(\hat{\boldsymbol{\Phi}}_m)^{-1}\boldsymbol{B}(\hat{\boldsymbol{\Phi}}_m)\right]^h \boldsymbol{Z}_{f0}, \tag{2.23}$$

$$\hat{\boldsymbol{S}}_{fh} = \sum_{k=1}^{h} D_k(\hat{\boldsymbol{\Phi}}_m) \hat{\Gamma} D_k(\hat{\boldsymbol{\Phi}}_m)'. \tag{2.24}$$

As previsões dos valores futuros $\mathbf{Z}_{fh} = z_{t(n+h,m)}, m=1,...,s$, têm intervalos com $100(1-\alpha)\%$ de confiança calculados por $\hat{z}_{t(n+h,m)} \pm \epsilon(\alpha/2)\sqrt{\hat{s}_m}$, sendo $\hat{z}_{t(n+h,m)}$ a previsão de $z_{t(n+h,m)}$, calculada em (2.23) e \hat{s}_m é o m-ésimo elemento da diagonal da matriz \hat{S}_{fh} estimada em (2.24), que representa a variância da previsão $\hat{z}_{t(n+h,m)}$. A seguir, apresenta-se um estudo sobre o erro de previsão.

2.4.2 MSE da previsão para séries de vazões sazonais

O cálculo do erro quadrático médio (MSE) de previsão para séries de vazões sazonais representa a etapa final da previsão e tem como objetivo avaliar a perfomance do modelo adotado. Nesta análise, o MSE deve ser comparado com a variância incondicional, em um horizonte de previsão de h=1,...,12 passos a frente.

Denomina-se de horizonte de previsão (h), o intervalo de tempo que separa a última observação usada no ajuste do modelo de previsão e o valor futuro a ser previsto. Um comportamento que se deve esperar intuitivamente do erro de previsão é que este erro deve crescer com o horizonte de previsão, ou seja, quanto mais à frente se deseja prever valores futuros, maior é o erro de previsão. Isto se deve à limitada capacidade que o passado e o presente tem para prever o crescente efeito de variáveis aleatórias imprevisíveis que podem influenciar os valores futuros da série, quando se distancia do presente em direção ao futuro. Assim, é natural a expectativa de que, à medida que o

2.4 Previsão clássica 31

horizonte de previsão aumente, o erro de previsão tenda para a variância incondicional da série.

Neste trabalho, os erros de previsão são avaliados na escala da série original para os diferentes modelos e usa-se como referência a variância incondicional sazonal da série original. Sendo assim, tem-se como objetivo avaliar a velocidade com que estes erros de previsão tendem para a variância incondicional da série, quando diferentes modelos são considerados para o cálculo das previsões.

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

$$MSE(\hat{z}_{T+h}) = E(z_{T+h} - \hat{z}_{T+h})^2.$$
 (2.25)

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

$$\hat{z}_{T+h} = E(z_{T+h}|z_T, z_{T-1}, \dots).$$
 (2.26)

Para avaliar o MSE na escala original das vazões supõe-se que são conhecidas as vazões históricas até um determinado mês m_0 do ano n e se quer prever a vazão $x_{t(n+1,m)}$, no mês m do ano (n+1). Então, a vazão $x_{t(n+1,m)}$ é escrita como:

$$x_{t(n+1,m)} = \mu_m + \sigma_m z_{t(n+1,m)}. (2.27)$$

Sendo t(n+1,m) = sn + m e $t(n,m_0) = s(n-1) + m_0$ e considerando o passo da previsão $h = t(n+1,m) - t(n,m_0)$ (considera-se sempre que $1 \le h \le s$), pode ser escrito então que $t(n+1,m) = t(n,m_0) + h$ e, assim, a equação (2.27), pode ser escrita como:

$$x_{t(n,m_0)+h} = \mu_m + \sigma_m x_{t(n,m_0)+h}, \tag{2.28}$$

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

A previsão de mínimo erro quadrático médio para $x_{t(n+1,m)}$, feita no mês m_0 , é dada por:

$$\widehat{x}_{t(n,m_0)+h} = E(x_{t(n,m_0)+h} | x_{t(n,m_0)}, x_{t(n,m_0)-1}, \ldots), \tag{2.29}$$

em que o conhecimento de $\{x_{t(n,m_0)}, x_{t(n,m_0)-1}, \ldots\}$ implica em conhecimento a respeito de $\{z_{(n,m_0)}, z_{(n,m_0)-1}, \ldots\}$. Usando as equações (2.28) e (2.29) tem-se que:

$$\widehat{x}_{t(n,m_0)+h} = \mu_m + \sigma_m \widehat{z}_{t(n,m_0)+h}.$$
(2.30)

O MSE, associado à previsão $\widehat{x}_{t(n,m_0)+h}$, é dado por:

$$MSE_{m}(\widehat{x}_{t(n,m_{0})+h}) = E(x_{t(n,m_{0})+h} - \widehat{x}_{t(n,m_{0})+h})^{2},$$

$$= \sigma_{m}^{2} E(z_{t(n,m_{0})+h} - \widehat{z}_{t(n,m_{0})+h})^{2}.$$
(2.31)

Portanto, o MSE para a previsão da série sazonal tem um padrão sazonal com período s. Usa-se a notação $MSE_m(\widehat{x}_{t(n,m_0)+h})$, referindo-se ao MSE da previsão de passo h, da vazão $x_{t(n,m_0)+h}$. O MSE, dado na equação (2.31), pode ser escrito como:

$$MSE_m(\widehat{x}_{t(n,m_0)+h}) = \sigma_m^2 MSE(\widehat{z}_{t(n,m_0)+h}). \tag{2.32}$$

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

MSE com um processo independente

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

Para calcular o MSE da previsão $\widehat{x}_{t(n,m_0)+h}$ considera-se a equação (2.31). Com a hipótese de independência do processo $z_{t(r,m)}$, $r=1,\ldots,n$, $m=1,\ldots,s$, tem-se que:

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

e, então, a previsão é dada por:

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

e, portanto,

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

2.4 Previsão clássica 33

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

$$MSE_m(\widehat{x}_{t(n,m_0)+h}) = \sigma_m^2 \sigma_z^2. \tag{2.33}$$

Nota: Teoricamente, a relação (2.1) assegura que $Var(z_{t(r,m)}) = 1$, mas, na prática, se constata que $\sigma_z^2 \approx 1$ e, portanto, do ponto de vista prático, a suposição de independência leva a um $MSE_m(\widehat{x}_{t(n,m_0)+h}) \approx \sigma_m^2$. A suposição de independência feita para $x_{t(r,m)}$ implica em desprezar toda e qualquer informação vinda das observações $\{x_{t(n,m_0)}, x_{t(n,m_0)-1}, \ldots\}$. Se o único efeito das observações é diminuir o MSE da previsão, então o MSE pode ser considerado como o maior erro quadrático médio da previsão que se pode cometer, ou seja, σ_m^2 (a variância mensal incondicional) pode ser considerada como um limitante superior para o erro quadrático médio da previsão para qualquer mês de origem m_0 e qualquer passo de previsão h.

2.4.3 MSE com um processo PAR(p_m)

O modelo $PAR(p_m)$, dado pela equação (2.1), pode ser escrito como uma soma infinita dos ruídos aleatórios periódicos, dado por:

$$z_{t(r,m)} = \sum_{j=0}^{\infty} \psi_{j,m} a_{t(r,m)-j} \quad \text{com} \quad \psi_{0,m} = 1,$$
 (2.34)

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

$$z_{t(n,m_0)+h} = \sum_{j=0}^{\infty} \psi_{j,m} a_{t(n,m_0)+h-j}, \quad \text{com} \quad \psi_{0,m} = 1.$$
 (2.35)

Na equação (2.35) tem-se que $t(n, m_0) + h = t(n+1, m)$ e, por isso, os coeficientes $\psi_{j,m}$, que multiplicam os ruídos $a_{t(n,m_0)+h-j}$, tem o índice m, que se refere ao mês para o qual se considera a série $z_{t(n+1,m)}$.

Denota-se por $\widehat{z}_{t(n,m_0)+h}$, a previsão para $z_{t(n,m_0)+h}$ calculada com h passos de antecedência e usando-se um modelo $PAR(p_m)$. Então, $\widehat{z}_{t(n,m_0)+h}$ deve ser escrita em função dos valores passados $z_{t(n,m_0)}$, $z_{t(n,m_0)-1}$, Consequentemente, usando a equação (2.35), $\widehat{z}_{t(n,m_0)+h}$ é escrita como uma função dos ruídos passados $a_{t(n,m_0)}$, $a_{t(n,m_0)-1}$,

Além disto, suponha que $\hat{z}_{t(n,m_0)+h}$ é uma função linear. Então, denotando-se a melhor previsão por:

$$\widehat{z}_{t(n,m_0)+h} = \sum_{j=0}^{\infty} \psi_{h+j,m}^* a_{t(n,m_0)-j}, \qquad (2.36)$$

pode-se determinar os pesos $\psi_{h+j,m}^*$ que minimizem o erro quadrático médio da previsão. Usando-se as equações (2.35) e (2.36), o erro de previsão, $e_{t(n,m_0)+h} = z_{t(n,m_0)+h} - \hat{z}_{t(n,m_0)+h}$, do modelo PAR (p_m) , com passo de previsão h, é escrito em função dos ruídos aleatórios, como:

$$e_{t(n,m_0)+h} = \sum_{j=0}^{\infty} \psi_{j,m} a_{t(n,m_0)+h-j} - \sum_{j=0}^{\infty} \psi_{h+j,m}^* a_{t(n,m_0)-j},$$
(2.37)

em que a equação (2.37) pode ser reescrita como:

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

$$\sum_{j=0}^{\infty} (\psi_{h+j,m} - \psi_{h+j,m}^*) a_{t(n,m_0)-j}.$$
(2.38)

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

$$MSE(\widehat{z}_{t(n,m_0)+h}) = E(e_{t(n,m_0)+h}^2),$$
 (2.39)

então, tem-se da equação (2.39), que o mínimo $MSE(\widehat{z}_{t(n,m_0)+h})$ ocorre quando $\psi_{h+j,m}^* = \psi_{h+j,m}$, para $j=0,1,\ldots$ Sendo os ruídos aleatórios $a_{t(n,m_0)+h-j}$, $j=0,\ldots,h-1$ não correlacionados e periodicamente estacionários, tem-se que $Var(a_{t(n,m_0)+h-j}) = Var(a_{m_0+h-j}) = \tau_{m_0+h-j}^2$, para $j=0,\ldots,h-1$. A estacionariedade periódica permite escrever que $\tau_{m_0+h-j}^2 = \tau_i^2$ para $i=1,\ldots,s$. Os índices i são associados com cada um dos valores m_0+h-j por:

2.4 Previsão clássica 35

$$i = \left\{ \begin{array}{cc} m_0 + h - j - \nu s, & \text{se } m_0 + h - j \neq s \\ s, & \text{se } m_0 + h - j = s \end{array} \right\}, \tag{2.40}$$

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

$$MSE(\widehat{z}_{t(n,m_0)+h}) = (\tau_{m_0+h}^2 + \psi_{1,m}^2 \tau_{m_0+h-1}^2 + \cdots + \psi_{h-1,m}^2 \tau_{m_0+1}^2), \qquad (2.41)$$

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

Cálculo dos parâmetros $\psi_{j,m}$

Para calcular os parâmetros $\psi_{j,m}$, $j=1,\ldots,h-1$ da equação (2.41), é necessário escrever o modelo PAR (p_m) na forma da equação (2.34). Para isto, adota-se um procedimento recursivo e, para facilitar os cálculos, supõe-se, sem perda de generalidade, que os modelos PAR (p_m) têm a mesma ordem $p=\max\{p_m\}$, para todos os meses $m=1,\ldots,s$. Esta suposição pode ser feita adotando-se $\phi_{k,m}=0$ para $p_m < k \le p$, em todos os modelos com $p_m < p$.

Para iniciar o procedimento recursivo de construção da equação (2.34), primeiro considera-se o modelo $PAR(p_m)$ escrito como um modelo PAR(p), dado por:

$$z_{t(r,m)} = \phi_{1,m} z_{t(r,m)-1} + \phi_{2,m} z_{t(r,m)-2} + \cdots + \phi_{p,m} z_{t(r,m)-p} + a_{t(r,m)}.$$
(2.42)

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

$$z_{t(r,m)-1} = \phi_{1,m-1} z_{t(r,m)-2} + \phi_{2,m-1} z_{t(r,m)-3} + \cdots + \phi_{p,m-1} z_{t(r,m)-1-p} + a_{t(r,m)-1}.$$
(2.43)

Substituindo (2.43) em (2.42), obtêm-se:

$$z_{t(r,m)} = a_{t(r,m)} + \phi_{1,m} a_{t(r,m)-1} + (\phi_{1,m} \phi_{1,m-1} + \phi_{2,m}) z_{t(r,m)-2} + (2.44)$$

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

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

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

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

$$+ [(\phi_{1,m} \phi_{1,m-1} + \phi_{2,m}) \phi_{p,m-2} + (\phi_{1,m} \phi_{p,m-1} + \phi_{p,m})] z_{t(r,m)-p}.$$
(2.45)

Das equações (2.44) e (2.45), conclui-se que:

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

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

$$\psi_{j,m} = \phi_{1,m-j+1}\psi_{j-1,m} + \phi_{2,m-j+2}\psi_{j-2,m} + \cdots$$

$$\cdots + \phi_{j-1,m-1}\psi_{1,m} + \phi_{j,m}. \tag{2.46}$$

A equação (2.46) é válida para qualquer mês m, assim escreve-se, de forma geral, a equação para $\psi_{j,m}$, como:

$$\psi_{j,m} = \sum_{k=0}^{j-1} \phi_{j-k,m-k} \psi_{k,m}.$$
 (2.47)

2.4 Previsão clássica 37

Após serem ajustados os modelos $PAR(p_m)$, avalia-se o $MSE(\widehat{z}_{t(n,m_0)+h})$, dado em (2.41), calculando-se recursivamente os coeficientes $\psi_{j,m}$ com a equação (2.47). Usando-se a equação (2.32), avalia-se o $MSE_m(\widehat{x}_{t(n,m_0)+h})$ na escala original da série de vazões, como se segue:

$$MSE_{m}(\widehat{x}_{t(n,m_{0})+h}) = \sigma_{m}^{2}(\tau_{m_{0}+h}^{2} + \psi_{1,m}^{2}\tau_{m_{0}+h-1}^{2} + \cdots + \psi_{h-1,m}^{2}\tau_{m_{0}+1}^{2}).$$
(2.48)

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

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

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

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

$$\lim_{h \to \infty} MSE_m(\widehat{x}_{t(n,m_0)+h}) = \sigma_m^2 \sum_{j=0}^{\infty} \psi_{j,m}^2 \tau_j^2.$$
 (2.49)

Portanto, da equação (2.34), pode ser mostrado que a série da equação (2.49) corresponde à variância incondicional da série padronizada $z_{t(r,m)}$, que como pode ser vista em (2.1), é igual a 1. Desta forma, pode-se afirmar que, no caso dos modelos $PAR(p_m)$, a variância incondicional σ_m^2 é um limitante superior para o MSE da previsão para cada mês.

Capítulo

3

Modelo Autorregressivo Periódico - Abordagem Bayesiana

Esse capítulo tem por objetivo apresentar a metodologia de inferência bayesiana, assim como os métodos MCMC. A seguir, apresenta-se as etapas de identificação, estimação, seleção e previsão para os modelos PAR.

3.1 Inferência bayesiana

Fazer inferências é uma das principais finalidades da estatística. Na abordagem clássica, os parâmetros desconhecidos são considerados fixos e toda a análise se restringe às informações contidas na amostra dos dados. Segundo Paulino et al. (2003), esta abordagem foi adotada, de forma quase unânime, pelos estatísticos durante a primeira metade do século XX. No entanto, uma abordagem alternativa ressurge devido aos avanços computacionais, a inferência bayesiana, que considera úteis todas as informações disponíveis para reduzir as incertezas na análise.

Portanto, a inferência bayesiana trata o parâmetro desconhecido como quantidade aleatória e, consequentemente, permite incorporar algum conhecimento sobre esse, assumindo, assim, distribuições de probabilidade, antes que os dados tenham sido coletados (Box & Tiao, 1992). De maneira bem geral, a distribuição a priori é o principal fator que diferencia a estatística clássica da bayesiana. Portanto, é alvo de crítica para muitos dos estatísticos clássicos, que alegam ser a situação a priori um processo subjetivo, já que dois pesquisadores podem ter diferentes graus de incertezas sobre uma quantidade desconhecida (Paulino et al., 2003).

Por meio do Teorema de Bayes, essa distribuição a priori para o vetor de parâmetros $\boldsymbol{\theta}$ de um modelo, $\pi(\boldsymbol{\theta})$, é combinada com a informação contida nos dados amostrais, ou seja, a função de verossimilhança, $L(\boldsymbol{\theta}|x)$, induzindo a uma distribuição a posteriori, $\pi(\boldsymbol{\theta}|x)$. Portanto, toda inferência é realizada usando-se a distribuição a posteriori e esta informação pode ser resumida pela média, moda, mediana ou pelos intervalos de credibilidade (Paulino et al., 2003). Dessa forma, a inferência bayesiana oferece resultados mais completos, pois, além da verossimilhança, utiliza também a informação a priori relativa aos parâmetros.

O Teorema de Bayes é fundamental na construção da inferência bayesiana e, basicamente, é o resultado de uma probabilidade condicional. Para o caso em que o vetor de parâmetros θ é contínuo, o Teorema é dado por:

$$\pi(\boldsymbol{\theta}|x) = \frac{L(\boldsymbol{\theta}|x)\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|x)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$
(3.1)

e, no caso em que θ é discreto, tem-se:

$$\pi(\boldsymbol{\theta}|x) = \frac{L(\boldsymbol{\theta}|x)\pi(\boldsymbol{\theta})}{\sum L(\boldsymbol{\theta}|x)\pi(\boldsymbol{\theta})}.$$
 (3.2)

Como nas expressões (3.1) e (3.2) o denominador independe do parâmetro θ , este pode ser considerado como uma constante e, então, estas expressões podem ser representadas por:

$$\pi(\boldsymbol{\theta}|x) \propto L(\boldsymbol{\theta}|x)\pi(\boldsymbol{\theta}).$$
 (3.3)

Com a utilização dos métodos computacionais Monte Carlo via Cadeias de Markov (MCMC) é possível gerar amostras da densidade *a posteriori* sem conhecer a constante normalizadora. Este fato permite fazer um sumário da densidade *a posteriori* usando estimadores via simulação de Monte Carlo e, assim, a forma proporcional da densidade *a posteriori* dada em (3.3) tornou-se bastante utilizada.

3.1.1 Distribuição a priori não informativa

Existem algumas situações em que a informação disponível a respeito do vetor de parâmetros de interesse é pouco significativa e, assim, o uso de distribuições *a priori* não informativas acarreta que a informação resultante na distribuição *a posteriori* se refere apenas aos dados (Carlin & Louis, 1997).

Nestes casos, pode-se considerar uma distribuição a priori Uniforme, ou seja, $\pi(\theta) \propto constante$, pois os valores de θ são igualmente prováveis ou a densidade a priori de Jeffreys (Jeffreys, 1961), que é invariante e imprópria e se baseia na informação de Fisher.

3.2 Método de Monte Carlo via Cadeias de Markov

Segundo Paulino et~al.~(2003), para se inferir em relação a qualquer elemento de θ , a distribuição conjunta a posteriori dos parâmetros deve ser integrada em relação a todos os outros parâmetros que a constituem, ou seja, procura-se obter a distribuição marginal de cada um dos parâmetros. A integração dessa distribuição, geralmente, não é analítica, necessitando de algoritmos iterativos especializados denominados de algoritmos MCMC (Markov Chain Monte Carlo).

Uma cadeia de Markov é um processo estocástico no qual o próximo estado da cadeia, ω_{t+1} , depende somente do estado atual, ω_t e dos dados e não da história passada da cadeia (Gamerman & Lopes, 2006). Segundo este autor, as primeiras iterações são influenciadas pelo estado inicial ω_1 e são descartadas. Este período é conhecido como aquecimento da cadeia ou burn-in. Também se considera uma dependência entre as observações subsequentes da cadeia e, para se obter uma amostra independente, as observações finais devem ser obtidas a cada k iterações, sendo este valor conhecido como salto, thin ou intervalo de amostragem.

A idéia dos métodos MCMC é obter uma amostra das distribuições marginais a posteriori dos parâmetros de interesse, por meio de um processo iterativo, utilizando as distribuições condicionais completas de cada parâmetro. Por sua vez, esses valores gerados são considerados amostras aleatórias de uma determinada distribuição de probabilidade, caracterizando, assim, o método de simulação de Monte Carlo. Dessa forma, tem-se uma ação conjunta dos métodos, que resulta no processo MCMC, cujos principais algoritmos são o Metropolis-Hastings e o amostrador de Gibbs.

3.2.1 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings permite gerar uma amostra da distribuição marginal dos parâmetros a partir das distribuições condicionais completas com formas desconhecidas. Eles usam a idéia de que um valor é gerado a partir de uma distribuição auxiliar ou candidata e este é aceito com uma dada probabilidade (Metropolis *et al.*, 1953; Hastings, 1970). O algoritmo de Metropolis-Hastings é estruturado da seguinte maneira:

- I. Inicialize o contador de iterações i=0 e especifique valores iniciais $\boldsymbol{\theta}^{(0)}=(\theta_1^{(0)},...,\theta_d^{(0)})$, em que d representa o número de parâmetros;
 - II. Gere um novo valor θ^c da distribuição proposta $q(.|\theta_1)$;
- III. Calcule a probabilidade de aceitação $\alpha(\theta_1, \theta^c)$ e gere $u \sim U(0, 1)$, em que U representa a distribuição Uniforme;

$$\alpha(\theta_1, \theta^c) = \min \left\{ 1, \frac{\pi(\theta^c | \theta_2, \dots, \theta_d) q(\theta_1 | \theta^c)}{\pi(\theta_1 | \theta_2, \dots, \theta_d) q(\theta^c | \theta_1)} \right\}$$
(3.4)

- IV. Se $u \leq \alpha$, então aceite o novo valor e faça $\theta_1^{i+1} = \theta^c$. Caso contrário, rejeite e faça $\theta_1^{i+1} = \theta_1^i$;
- V. Incremente o contador de i para i+1 e volte ao passo II até atingir a convergência.

Quando as distribuições condicionais completas possuem formas conhecidas, utilizase um caso especial do algoritmo de Metropolis-Hastings, o amostrador de Gibbs, no sentido de que seja fácil amostrar de seus elementos (Gelfand & Smith, 1990; Casella & George, 1992). O amostrador de Gibbs é estruturado da seguinte maneira:

- I. Inicialize o contador de iterações i=0 e especifique os valores iniciais $\boldsymbol{\theta}^{(0)}=(\theta_1^{(0)},...,\theta_d^{(0)});$
 - II. Gere um novo valor θ^c da distribuição condicional conhecida;
- III. Incremente o contador de i para i+1 e volte ao passo II até atingir a convergência.

3.2.2 Avaliação da convergência

As técnicas MCMC constituem uma ótima ferramenta para a resolução de muitos problemas na análise bayesiana, porém, apresentam algumas limitações. Entre elas,

incluem-se, as influências dos valores iniciais da cadeia, a correlação entre os dados gerados e a falta de convergência da cadeia.

Um procedimento útil na avaliação informal da convergência é a utilização de gráficos. Segundo Gelman et al. (2004), os gráficos mais frequentes nesta análise são o gráfico do parâmetro ao longo das iterações e um gráfico da estimativa da distribuição a posteriori do parâmetro, por exemplo, um histograma ou uma densidade kernel. Outra forma é o cálculo da função de autocorrelação da cadeia com determinados valores de defasagens ou lags. Se a cadeia apresentar alta correlação, isto é um forte indicativo de falta da convergência.

Para avaliar a convergência mais detalhadamente, alguns critérios foram desenvolvidos, entre os quais se pode destacar o critério de Gelman e Rubin (Gelman & Rubin, 1992) e o critério de Geweke (Geweke, 1992). Estes critérios estão implementados no pacote BOA (Bayesian Output Analysis) que pode ser instalado no software livre R (R Development Core Team, 2010).

O critério de Gelman e Rubin é baseado na comparação entre duas ou mais cadeias paralelas. Ele analisa as variâncias dentro e entre as cadeias e utiliza esta informação para estimar o fator pelo qual o parâmetro escalar da distribuição marginal a posteriori deveria ser reduzido se a cadeia fosse repetida infinitas vezes. Este fator é expresso pelo valor \hat{R} (fator de redução de escala potencial ou fator de diagnóstico da convergência) e sugere que valores de \hat{R} próximos a 1 indicam que a convergência foi atingida para n iterações.

O critério de Geweke propõe uma avaliação de convergência que se baseia no teste de igualdade das médias da primeira e da última parte da cadeia, geralmente, dos primeiros 10% e dos últimos 50%. Segundo o autor, a cadeia converge quando a maioria dos dados está entre os limites de uma distribuição Normal padronizada.

3.3 Estimação dos parâmetros

Considerando o modelo autorregressivo periódico de ordem p_m nas análises, o processo de ajuste pode ser dividido, para cada período m, em: identificação da ordem p_m , estimação dos parâmetros $\Phi_m = (\phi_1, \phi_2, ..., \phi_{p_m})$ e τ_m , seleção do modelo e, finalmente, a previsão da série.

No procedimento de identificação da ordem p_m utiliza-se o critério BIC, já mencionado na seção (2.2). Em relação à estimação dos parâmetros, consideram-se, neste estudo, os modelos com distribuição Normal, com distribuição Log-Normal e o modelo

t-Student. A seguir, apresenta-se todas as etapas necessárias para o procedimento de estimação bayesiana dos parâmetros.

3.3.1 Modelo Normal

A abordagem bayesiana para estimação dos parâmetros Φ_m e τ_m de um modelo $PAR(p_m)$ inicia-se com a determinação da função de verossimilhança. Para o modelo com distribuição Normal, esta é dada por:

$$L_{m}(\boldsymbol{\Phi}_{m}, \tau_{m} | \boldsymbol{Z}_{m}) \propto \tau_{m}^{(n-1)/2} \exp \left\{-\frac{\tau_{m}}{2} [(\boldsymbol{Z}_{m} - \boldsymbol{X}_{m} \hat{\boldsymbol{\Phi}}_{m})' (\boldsymbol{Z}_{m} - \boldsymbol{X}_{m} \hat{\boldsymbol{\Phi}}_{m})]\right\} \exp \left\{-\frac{\tau_{m}}{2} [(\boldsymbol{\Phi}_{m} - \hat{\boldsymbol{\Phi}}_{m})' (\boldsymbol{X}'_{m} \boldsymbol{X}_{m}) (\boldsymbol{\Phi}_{m} - \hat{\boldsymbol{\Phi}}_{m})]\right\}.$$

Segundo Pagano (1978), pode-se assumir independência entre os parâmetros, o que resulta em uma função de verossimilhança para cada mês.

Posteriormente, adota-se a densidade *a priori* não informativa de Jeffreys. Supõe-se que Φ_m e τ_m sejam independentes e com distribuições *a priori* dadas, respectivamente, por $\pi_m(\Phi_m) \propto constante$ e $\pi_m(\tau_m) \propto 1/\tau_m$, para $\tau_m > 0$. Tem-se, assim, a densidade conjunta *a priori* dada por:

$$\pi_m(\mathbf{\Phi}_m, \tau_m) \propto 1/\tau_m.$$

Combinando a função de verossimilhança com a densidade conjunta *a priori* obtémse a densidade conjunta *a posteriori* dada pela expressão:

$$\pi_{m}(\boldsymbol{\Phi}_{m}, \tau_{m} | \boldsymbol{Z}_{m}) \propto \tau_{m}^{((n-1)/2)-1} \exp \left\{-\frac{\tau_{m}}{2} [(\boldsymbol{Z}_{m} - \boldsymbol{X}_{m} \hat{\boldsymbol{\Phi}}_{m})' (\boldsymbol{Z}_{m} - \boldsymbol{X}_{m} \hat{\boldsymbol{\Phi}}_{m})]\right\}$$

$$\exp \left\{-\frac{\tau_{m}}{2} [(\boldsymbol{\Phi}_{m} - \hat{\boldsymbol{\Phi}}_{m})' (\boldsymbol{X}_{m}' \boldsymbol{X}_{m}) (\boldsymbol{\Phi}_{m} - \hat{\boldsymbol{\Phi}}_{m})]\right\}.$$

Apesar de se conseguir calcular analiticamente as densidades marginais a posteriori utilizou-se aqui os procedimentos MCMC, por serem de fácil implementação computacional e úteis em análises de previsão. Sendo assim, a partir da densidade conjunta a posteriori determinam-se as densidades condicionais a posteriori para os parâmetros Φ_m e τ_m , dadas, respectivamente, por:

$$\pi_m(\mathbf{\Phi}_m|\tau_m, \mathbf{Z}_m) \sim N_M(\mathbf{\hat{\Phi}}_m, \tau_m^{-1}(\mathbf{X}_m'\mathbf{X}_m)^{-1}),$$

$$\pi_m(\tau_m|\mathbf{\Phi}_m,\mathbf{Z}_m) \sim G((n-1)/2,B_m/2),$$

em que
$$B_m = [(\mathbf{Z}_m - \mathbf{X}_m \hat{\mathbf{\Phi}}_m)'(\mathbf{Z}_m - \mathbf{X}_m \hat{\mathbf{\Phi}}_m) + (\mathbf{\Phi}_m - \hat{\mathbf{\Phi}}_m)'(\mathbf{X}_m' \mathbf{X}_m)(\mathbf{\Phi}_m - \hat{\mathbf{\Phi}}_m)].$$

Como pode ser observado, as densidades condicionais *a posteriori* dos parâmetros Φ_m e τ_m apresentam formas padronizadas dadas, respectivamente, pelas distribuições Normal Multivariada e Gama e, assim, permitem o uso do amostrador de Gibbs.

3.3.2 Modelo Log-Normal

Caso haja necessidade, pode-se usar uma transformação, com o objetivo de corrigir problemas tais como a heterocedasticidade e/ou não-normalidade nos resíduos do modelo PAR ajustado para a série. Neste trabalho, será considerada apenas a transformação logarítmica, ou seja, $\lambda=0$, sendo esta associada à suposição de um modelo Log-Normal.

Assim, a série transformada e padronizada é dada por:

$$z_{t(r,m)} = \frac{\log(\widetilde{x}_{t(r,m)}) - \mu_m^{LN}}{\sigma_m^{LN}},$$

em que μ_m^{LN} e σ_m^{LN} representam, respectivamente, a média e o desvio padrão da série transformada para cada mês.

3.3.3 Modelo t-Student

Neste trabalho, discute-se também a aplicação do modelo com caudas pesadas, ou seja, com distribuição t-Student para séries temporais periódicas. Muitas séries de vazões apresentam este comportamento e é importante na adequação destas séries a definição de um modelo que captura a característica de caudas pesadas. Tipicamente, nestes casos, a solução seria transformar a série a fim de se obter um comportamento aproximadamente Normal. Sendo assim, propõe-se um modelo que permita uma representação mais confiável de séries de vazões médias mensais sem preliminares transformações.

Em particular, pode-se escrever a distribuição t-Student como uma mistura da distribuição Normal com a distribuição Gama da seguinte forma:

$$tSt(\mu, \tau^{-1}, v) = \int_0^\infty f_N(z|\mu, [w\tau]^{-1}) f_G(w|v/2, v/2) dw,$$

em que $f_N(z|\mu,[w\tau]^{-1})$ representa uma distribuição Normal dada por:

$$f(z|w) \propto (w\tau)^{1/2} \exp\left\{\frac{-w\tau}{2}(z-\mu)^2\right\}$$

e $f_G(w|v/2,v/2)$ representa a distribuição Gama dada por:

$$f(w) \propto w^{(v/2)-1} \exp\left\{\frac{-wv}{2}\right\},$$

em que v representa o grau de liberdade e controla o peso nas caudas e w é a variável latente.

Em um primeiro momento, foram adotadas distribuições *a priori* para o parâmetro grau de liberdade *v*. Assim, definiu-se, como primeiro caso, a utilização da distribuição *a priori* não informativa de Jeffreys, dada por:

$$\pi_m(v) \propto \left\{ 2h(v) - \psi^{(2)} \left((v+1)/2 \right) + \psi^{(2)} \left(v/2 \right) - 2/v \right\}^{1/2},$$

em que $\psi^{(2)}(z)$ é a função trigama de z e $h(v) = \frac{2}{v+1} - \frac{v+2}{v(v+3)}$

Em um segundo caso, foi adotada a distribuição exponencial deslocada como distribuição a priori para v, ou seja, $\pi_m(v) = \exp\{-(v-2)\}$. Após, este parâmetro foi considerado fixo, pois não representa um parâmetro de interesse e, assim, foi adotado um valor constante nas análises, sendo escolhido, após uma análise de sensibilidade, o valor v=3. Como os resultados apresentados neste caso foram superiores aos outros casos, apresenta-se, na sequência, a análise bayesiana apenas para esse caso.

Assim, a abordagem bayesiana para estimação dos parâmetros inicia-se com a determinação da função de verossimilhança. Para o modelo com distribuicao t-Student, esta é dada por:

$$L_m(\mathbf{\Phi}_m, \tau_m, w | \mathbf{Z}_m) \propto w^{n(v-1)/2} \tau_m^{(n/2)} \exp \left\{ \frac{-w(\tau_m B_m + nv)}{2} \right\}.$$

Combinando a função de verossimilhança com a distribuição a priori não informativa conjunta $\pi_m(\Phi_m, \tau_m) \propto 1/\tau_m$, obtém-se a densidade conjunta a posteriori dada por:

$$\pi_m(\mathbf{\Phi}_m, \tau_m, w | \mathbf{Z}_m) \propto w^{n(v-1)/2} \tau_m^{(n/2)-1} \exp\left\{\frac{-w(\tau_m B_m + nv)}{2}\right\}.$$

A partir da densidade conjunta *a posteriori* determinam-se as distribuições condicionais *a posteriori* dadas por:

$$\pi_m(\mathbf{\Phi}_m|\tau_m, w, \mathbf{Z}_m) \sim N_M(\hat{\mathbf{\Phi}}_m, (w\tau_m)^{-1}(\mathbf{X}_m'\mathbf{X}_m)^{-1}),$$

$$\pi_m(\tau_m|\mathbf{\Phi}_m, w, \mathbf{Z}_m) \sim G(n/2, (wB_m)/2),$$

$$\pi_m(w|\mathbf{\Phi}_m, \tau_m, \mathbf{Z}_m) \sim G\left(\frac{n(v-1)+2}{2}, \frac{(\tau_m B_m + nv)}{2}\right).$$

Como se pode observar, as densidades condicionais a posteriori dos parâmetros Φ_m , τ_m e da variável latente w apresentam formas padronizadas que são dadas, respectivamente, pelas distribuições Normal Multivariada, Gama e Gama e, assim, permitem o uso do amostrador de Gibbs.

3.4 Seleção de modelos

A seleção de modelos pode ser realizada por meio de muitos métodos propostos na literatura. Entre estes, destacam-se o BIC, já mencionado, o DIC (*Deviance Information Criterion*), proposto por Spiegelhalter *et al.* (2002) e o critério da densidade preditiva ordenada (CPO), proposto por Gelfand & Dey (1994).

O DIC é um critério de informação em que o número efetivo de parâmetros é definido como $d_i = \bar{D}_i - D(\tilde{\Phi}_{m,i})$, em que $D(\Phi_{m,i}) = -2 \log L_i(\Phi_{m,i}|x)$ é a função desvio ou deviance, $\tilde{\Phi}_{m,i} = E[\Phi_{m,i}|x, M_i]$ e $\bar{D}_i = E[D(\Phi_{m,i}|x, M_i)]$. Assim, o critério é definido

por $DIC(M_i) = -2 \log L_i(\tilde{\Phi}_{m,i}|x) + 2d_i$, com \bar{D}_i e d_i representando, respectivamente, medidas do ajuste e da complexidade do modelo. O DIC generaliza o AIC e os menores valores de DIC determinam os melhores modelos. Computacionalmente, o DIC é bem atrativo, pois este pode ser facilmente incorporado às rotinas MCMC. Para propósitos computacionais, o DIC pode ser reescrito como $DIC(M_i) = 2\bar{D}_i - D(\tilde{\Phi}_{m,i})$.

O critério da CPO para as observações \mathbf{Z}_{n+k} dado $\mathbf{Z}_{n+k-1} = (Z_1, ..., Z_n, ..., Z_{n+k-1})$ e os parâmetros $\mathbf{\Phi}$ e τ pode ser especificado para um modelo M_p , como um $AR(p_m)$, da seguinte forma:

$$c_k^{M_p} = f(\boldsymbol{Z}_{n+k}|\boldsymbol{Z}_{n+k-1}) \propto \int_{\boldsymbol{\Phi}} \int_{\tau} \tau^{1/2} \exp\left\{\frac{-\tau}{2} \left(\boldsymbol{Z}_{n+k} - \sum_{i=1}^{p_m} \boldsymbol{\Phi}_i \boldsymbol{Z}_{n+k-i}\right)^2\right\}$$

$$\pi(\boldsymbol{\Phi}, \tau|\boldsymbol{Z}) d\boldsymbol{\Phi} d\tau,$$
(3.5)

em que $\pi(\mathbf{\Phi}, \tau | \mathbf{Z})$ é a densidade *a posteriori* conjunta para os parâmetros $\mathbf{\Phi}$ e τ e M_p é o modelo que está sendo avaliado, no caso, um $AR(p_m)$.

Usando as amostras geradas pelos algoritmos MCMC, a equação (3.5) pode ser aproximada pelo estimador de Monte Carlo dado por:

$$\hat{f}(\boldsymbol{Z}_{n+k}|\boldsymbol{Z}_{n+k-1}) = \frac{1}{T} \sum_{i=1}^{T} \tau^{1/2(j)} \exp \left\{ \frac{-\tau^{(j)}}{2} \left(\boldsymbol{Z}_{n+k} - \sum_{i=1}^{p_m} \boldsymbol{\Phi}_{i,m}^{(j)} \boldsymbol{Z}_{n+k-i} \right)^2 \right\}$$

$$\pi(\boldsymbol{\Phi}^{(j)}, \tau^{(j)}|\boldsymbol{Z}),$$
(3.6)

em que T representa o tamanho da amostra final nos processos MCMC.

A estimativa $\hat{c}_k^{M_p} = \hat{f}(\boldsymbol{Z}_{n+k}|\boldsymbol{Z}_{n+k-1})$ pode ser usada nos processos de seleção do modelo. Esta etapa pode ser ilustrada pelo gráfico de $c_k^{M_p}$ versus i (i=1,...,K) para cada um dos modelos M_p . O critério da CPO a ser usado na seleção do melhor modelo M_p está na escolha do valor máximo de $C^{M_p} = \prod_{k=1}^K c_k^{M_p}$. Costuma-se calcular $\log C^{M_p} = \sum_{k=1}^K \log c_k^{M_p}$.

A seguir, apresenta-se os procedimentos para previsão bayesiana. Uma vantagem da previsão usando a abordagem bayesiana é que esta pode ser calculada facilmente usando técnicas de simulação MCMC.

3.5 Previsão bayesiana

Em muitas ocasiões, dada a informação amostral, pode-se ter interesse em fazer inferência sobre algum valor que ainda não foi observado, isto é, fazer uma previsão para um valor futuro. Na análise bayesiana, a função densidade de probabilidade para um valor ainda não observado, dada a informação amostral, é conhecida como densidade preditiva para o valor futuro.

Assim, calcula-se a previsão para o mês m, $\hat{Z}_{t(n+1,m)}$, usando a função densidade preditiva bayesiana, condicionada às observações $\mathbf{Z}_m = (Z_{t(1,m)}, ..., Z_{t(n,m)})$, dada por:

$$f_m(Z_{t(n+1,m)}|\mathbf{Z}_m) = \int_{\mathbf{\Phi}_m} \int_{\tau_m} f_m(Z_{t(n+1,m)}|\mathbf{\Phi}_m, \tau_m, \mathbf{Z}_m)$$

$$\pi_m(\mathbf{\Phi}_m, \tau_m|\mathbf{Z}_m) d\mathbf{\Phi}_m d\tau_m,$$
(3.7)

em que $\pi_m(\boldsymbol{\Phi}_m, \tau_m | \boldsymbol{Z}_m)$ é a densidade *a posteriori* dos parâmetros $\boldsymbol{\Phi}_m$ e τ_m e a expressão $f_m(Z_{t(n+1,m)} | \boldsymbol{\Phi}_m, \tau_m, \boldsymbol{Z}_m)$ é a função densidade de $Z_{t(n+1,m)}$, condicionada aos dados \boldsymbol{Z}_m e aos parâmetros $\boldsymbol{\Phi}_m$ e τ_m . Desta forma, tem-se que:

$$f_m(Z_{t(n+1,m)}|\mathbf{Z}_m) = E[f_m(Z_{t(n+1,m)}|\mathbf{\Phi}_m, \tau_m, \mathbf{Z}_m)].$$
 (3.8)

3.5.1 Densidade preditiva um passo a frente

Conforme visto, a densidade preditiva bayesiana é a densidade de valores futuros condicionada às observações passadas. Como exemplo, considere a densidade preditiva um passo a frente, $Z_{t(n+1,1)}$, para o mês de janeiro m=1 condicionado às observações $\mathbf{Z}_1 = (Z_{t(1,1)}, ..., Z_{t(n,1)})$. Portanto, tem-se:

$$f_m(Z_{t(n+1,1)}|\mathbf{Z}_1) = \int_{\mathbf{\Phi}_1} \int_{\tau_1} f_m(Z_{t(n+1,1)}|\mathbf{\Phi}_1, \tau_1, \mathbf{Z}_1)$$

$$\pi_1(\mathbf{\Phi}_1, \tau_1|\mathbf{Z}_1) d\mathbf{\Phi}_1 d\tau_1,$$
(3.9)

isto é,

$$f_1(Z_{t(n+1,1)}|\mathbf{Z}_1) = E[f_1(Z_{t(n+1,1)}|\mathbf{\Phi}_1, \tau_1, \mathbf{Z}_1)].$$
 (3.10)

A solução da expressão (3.9) é difícil de ser obtida analiticamente e uma alternativa é utilizar os métodos de simulação MCMC. Para isso, sejam as amostras $\Phi_1^{(j)}$, $\tau_1^{(j)}$, j=1,2,...,T, geradas a partir de um algoritmo MCMC. Um estimador para (3.9) pode ser dado por:

$$E[Z_{t(n+1,1)}|\mathbf{Z}_1] = \int_{-\infty}^{+\infty} Z_{t(n+1,1)} f_1(Z_{t(n+1,1)}|\mathbf{Z}_1) dZ_{t(n+1,1)}.$$
 (3.11)

Substituindo a equação (3.9) em (3.11) e mudando a ordem de integração, a equação (3.11) pode ser reescrita da seguinte forma:

$$E[Z_{t(n+1,1)}|\mathbf{Z}_{1}] = \int_{\mathbf{\Phi}_{1}} \int_{\tau_{1}} E[(Z_{t(n+1,1)}|\mathbf{\Phi}_{1}, \tau_{1}, \mathbf{Z}_{1})]$$

$$\pi_{1}(\mathbf{\Phi}_{1}, \tau_{1}|\mathbf{Z}_{1})d\mathbf{\Phi}_{1}d\tau_{1}.$$
(3.12)

Assim, um estimador para a expressão (3.12) considerando as amostras MCMC $\Phi_1^{(j)}, \tau_1^{(j)}$ é dado por:

$$\hat{E}[Z_{t(n+1,1)}|\mathbf{Z}_1] = \frac{1}{T} \sum_{i=1}^{T} E[(Z_{t(n+1,1)}|\mathbf{\Phi}_1^{(j)}, \tau_1^{(j)}, \mathbf{Z}_1)],$$
(3.13)

em que a equação (3.13) fornece a estimativa para a previsão de $Z_{t(n+1,1)}$, denotada por $\hat{Z}_{t(n+1,1)} = \hat{E}[Z_{t(n+1,1)}|\mathbf{Z}_1]$. Nota-se também a simplicidade no cálculo da previsão com o uso de algoritmos computacionais MCMC.

Agora, usando um modelo $PAR(p_1)$, tem-se que:

$$Z_{t(n+1,1)} = \sum_{i=1}^{p_1} \phi_{i,1} Z_{t(n+1,1)-i} + a_{t(n+1,1)}, \tag{3.14}$$

em que $a_{t(n+1,1)} \sim N(0, \tau_1^{-1}).$

A função densidade de probabilidade de $Z_{t(n+1,1)}$, condicionada a todos os parâmetros do modelo e às observações passadas é dada por:

$$f_1(Z_{t(n+1,1)}|\mathbf{\Phi}_1,\tau_1) \propto \tau_1^{1/2} \left\{ \exp\left[\frac{-\tau_1}{2} \left(Z_{t(n+1,1)} - \sum_{i=1}^{p_1} \phi_{i,1} Z_{t(n+1,1)-i} \right)^2 \right] \right\}.$$
 (3.15)

A densidade preditiva bayesiana de $Z_{t(n+1,1)}$ dados Φ_1, τ_1 e \mathbf{Z}_1 tem valor esperado dado por:

$$E[Z_{t(n+1,1)}|\boldsymbol{\Phi}_1, \tau_1, \boldsymbol{Z}_1] = \sum_{i=1}^{p_1} \phi_{i,1} Z_{t(n+1,1)-i}.$$
 (3.16)

Observa-se que $Z_{t(n+1,1)}=Z_{ns+1}$ e os valores passados são dados por $Z_{t(n+1,1)-1}=Z_{ns}, Z_{t(n+1,1)-2}=Z_{ns-1},..., Z_{t(n+1,1)-p_1-1}=Z_{ns-p_1}$. Dessa forma, tem-se:

$$E[Z_{t(n+1,1)}|\boldsymbol{\Phi}_{1}^{(j)},\tau_{1}^{(j)},\boldsymbol{Z}_{1}] = \sum_{i=1}^{p_{1}} \phi_{i,1}^{(j)} Z_{ns-i}.$$
(3.17)

A substituição da equação (3.17) em (3.16) mostra que a previsão do valor $Z_{t(n+1,1)}$ pode ser facilmente calculada usando-se os métodos MCMC.

Para comparar a capacidade preditiva dos modelos propostos, via inferência bayesiana, utiliza-se os erros de previsão, que comparam os valores previstos $(\hat{Z}_{t(n+h,m)})$ com os valores observados $(Z_{t(n+h,m)})$ para a série de vazões médias mensais, em que h representa o horizonte de previsão. Estes erros são representados pelo erro quadrático médio (MSE), erro absoluto médio (MAD) e o erro percentual absoluto médio (MAPE), H passos à frente, dados, respectivamente, por:

$$MSE = \frac{1}{H} \sum_{h=1}^{H} (Z_{t(n+h,m)} - \widehat{Z}_{t(n+h,m)})^{2},$$

$$MAD = \frac{1}{H} \sum_{h=1}^{H} |Z_{t(n+h,m)} - \widehat{Z}_{t(n+h,m)}|,$$

$$MAPE = \frac{100}{H} \sum_{h=1}^{H} \left| \frac{(Z_{t(n+h,m)} - \widehat{Z}_{t(n+h,m)})}{Z_{t(n+h,m)}} \right|.$$

Como as previsões não são perfeitas, o objetivo é tentar reduzir os erros de previsão.

Capítulo

4

Geração de Séries Sintéticas

Este capítulo trata o problema da geração de séries sintéticas de vazões médias mensais. Em um primeiro momento, são abordadas algumas considerações iniciais e, posteriormente, são introduzidas a distribuição Log-Normal com três parâmetros, o desenvolvimento para a geração de séries sintéticas multivariadas, o modelo Log-Normal generalizado para geração univariada via estimação bayesiana e, por último, um critério de validação das séries geradas.

4.1 Introdução

O histórico de vazões médias mensais disponível para o ajuste dos modelos é uma série temporal que consiste em uma das possíveis realizações do processo estocástico gerador. O objetivo é, portanto, dado um modelo autorregressivo periódico ajustado à série original, aproximar o comportamento estocástico e, sinteticamente, gerar tantas novas séries temporais quanto se queira, diferentes do histórico original, mas igualmente possíveis do ponto de vista estatístico.

O problema do planejamento anual da operação energética apresenta característica estocástica, pois não se conhece antecipadamente as vazões referentes às usinas

hidrelétricas. Essas vazões interferem significativamente na operação de sistemas hidrotérmicos, visto que as decisões futuras e, consequentemente, a política de operação, dependem dos cenários de vazões. Assim, é de fundamental importância desenvolver um modelo eficiente para geração das vazões. Portanto, faz-se necessário apresentar uma metodologia para transformar a série de observações, isto é, o histórico de vazões médias mensais, em um procedimento adequado para geração das séries sintéticas.

Muitos pesquisadores consideram que os resíduos no ajuste dos modelos PAR apresentam distribuição Normal e uma possível não-normalidade pode ser corrigida por transformações, tais como a transformação de Box & Cox (1964). Entretanto, como essas séries sintéticas geradas são utilizadas em modelos que calculam as estratégias ótimas de operação de um sistema multireservatórios, baseados em programação dinâmica dual estocástica, o modelo de geração de séries sintéticas deve ser aplicado diretamente à série temporal original e deve ser capaz de lidar com resíduos que apresentem um forte coeficiente de assimetria.

Nesta tese um modelo Log-Normal com três parâmetros aos resíduos mensais foi proposto. A seguir, apresenta-se os passos para modelagem dos ruídos que são usados nos processos de geração de séries sintéticas.

4.2 Modelo Log-Normal com três parâmetros

O primeiro passo para construir um sistema que possa gerar séries sintéticas de vazões consiste em ajustar um modelo à série histórica. Neste trabalho, considera-se, para modelar a correlação temporal entre as vazões mensais, os modelos $PAR(p_m)$, dado por (2.1). Uma restrição ao uso destes modelos, para geração de séries de vazões mensais, é que o modelo deve sempre gerar uma vazão $x_{t(r,m)} > x_0$, em que x_0 representa o valor mínimo de vazão em um determinado mês. Para atender esta restrição, os ruídos $a_{t(r,m)}$ devem satisfazer a condição:

$$a_{t(r,m)} > x_0 - \left(\frac{\mu_m}{\sigma_m}\right) - \sum_{j=1}^{p_m} \phi_{j,m} \left(\frac{x_{t(r,m)-j} - \mu_{m-j}}{\sigma_{m-j}}\right).$$
 (4.1)

No entanto, para a geração de séries sintéticas é fundamental a consideração de modelos probabilísticos desta série para que se possa gerar tantas séries temporais quanto sejam necessárias. A análise de séries de vazões reais revela que, geralmente, estas séries têm uma distribuição de probabilidade assimétrica. Portanto, os modelos de geração de séries sintéticas, os quais são ajustados diretamente à série temporal original, devem ser capaz de lidarem com ruídos que apresentem um forte coeficiente de assime-

tria. Além disso, é necessário que a distribuição dos ruídos $a_{t(r,m)}$ seja adequadamente definida para que a restrição imposta pela equação (4.1) seja satisfeita.

A solução adotada, segundo Stedinger & Taylor (1982), foi ajustar uma distribuição Log-Normal com três parâmetros para os ruídos $a_{t(r,m)}$. Portanto, supõe-se que $a_{t(r,m)}$ tem uma distribuição Log-Normal com três parâmetros, tal que se pode definir um processo aleatório $\xi_{t(r,m)}$, com distribuição Normal $N(\mu_m^{(\xi)}, \sigma_m^{(\xi)2})$, que está relacionado ao processo $a_{t(r,m)}$ pela equação:

$$\xi_{t(r,m)} = \log(a_{t(r,m)} + \tau_m).$$
 (4.2)

A equação (4.2) impõe uma nova restrição para $a_{t(r,m)}$, ou seja, $a_{t(r,m)} > -\tau_m$. Portanto, comparando-se as equações (4.1) e (4.2), tem-se uma possível definição para o parâmetro τ_m dada por:

$$\tau_m = \frac{\mu_m - x_0}{\sigma_m} + \sum_{i=1}^{p_m} \phi_{j,m} \left(\frac{x_{t(r,m)-j} - \mu_{m-j}}{\sigma_{m-j}} \right). \tag{4.3}$$

Os parâmetros $\mu_m^{(\xi)}$ e $\sigma_m^{(\xi)2}$ devem ser calculados de forma a preservar os dois primeiros momentos do ruído $a_{t(r,m)}$, denotados por $E(a_{t(r,m)}) = \mu_m^{(a)}$ e $Var(a_{t(r,m)}) = \sigma_m^{(a)2}$. Isto é feito usando-se as relações entre os parâmetros da distribuição Normal $N(\mu_m^{(\xi)}, \sigma_m^{(\xi)2})$ e os da distribuição Log-Normal com três parâmetros $LN(\mu_m^{(a)}, \sigma_m^{(a)2}, \tau_m)$ dadas por:

$$\mu_m^{(\xi)} = \frac{1}{2} \log \left(\frac{\sigma_m^{(a)2}}{\theta_m^2 - \theta_m} \right), \tag{4.4}$$

$$\sigma_m^{(\xi)2} = \log(\theta_m),\tag{4.5}$$

em que θ_m é definido por:

$$\theta_m = 1 + \frac{\sigma_m^{(a)2}}{(\mu_m^{(a)} - \tau_m)^2}. (4.6)$$

Portanto, após a identificação da ordem p_m e o ajuste dos modelos $PAR(p_m)$, estimase $\sigma_m^{(a)2}$ e τ_m com as equações (2.2) e (4.3) e os parâmetros $\mu_m^{(\xi)}$ e $\sigma_m^{(\xi)2}$ são estimados com as equações (4.4) - (4.6).

Um modelo para geração de uma série sintética de vazões mensais pode ser obtido substituindo na equação (2.1), o valor do ruído $a_{t(r,m)}$ retirado da equação (4.2) dada por:

$$a_{t(r,m)} = \exp\{\xi_{t(r,m)}\} - \tau_m.$$
 (4.7)

Substituindo em (4.7) a expressão de τ_m dada em (4.3) e, após algumas manipulações algébricas, obtém-se uma equação final para a geração de uma vazão sintética $x_{t(r,m)}$, para o mês m do ano r. Esta equação final é dada por:

$$x_{t(r,m)} = x_0 + \sigma_m \exp\{\xi_{t(r,m)}\}. \tag{4.8}$$

Portanto, um algoritmo para gerar uma série de vazões sintéticas é constituído dos seguintes passos:

Algoritmo para geração de série sintética

1. Considere o modelo PAR (p_m) ajustado e calcule as estimativas dos ruídos, $a_{t(r,m)}$, para $r=2,\ldots,n$ e $m=1,\ldots,s$, com

$$\hat{a}_{t(r,m)} = \left(\frac{x_{t(r,m)} - \hat{\mu}_m}{\hat{\sigma}_m}\right) - \left\{\sum_{j=1}^{p_m} \hat{\phi}_{j,m} \left(\frac{x_{t(r,m)-j} - \hat{\mu}_{m-j}}{\hat{\sigma}_{m-j}}\right)\right\}.$$

2. Calcule as estimativas para $\mu_m^{(a)}$, a média do resíduo para o mês m e $\sigma_m^{(a)2}$, a variância do resíduo para o mês m, $m = 1, \ldots, s$, com

$$\hat{\mu}_{m}^{(a)} = \frac{1}{n-1} \sum_{r=2}^{n} \hat{a}_{t(r,m)},$$

$$\hat{\sigma}_{m}^{(a)2} = \frac{1}{n-2} \sum_{r=2}^{n} \left(\hat{a}_{t(r,m)} - \hat{\mu}_{m}^{(a)} \right)^{2}.$$

- 3. Faça m = 1, r = n + 1.
- 4. Com as vazões iniciais $\{x_{t(r,m)-1}, \ldots, x_{t(r,m)-p_m}\}$ conhecidas, calcule a estimativa de τ_m com

$$\hat{\tau}_m = \left(\frac{\hat{\mu}_m - x_0}{\hat{\sigma}_m}\right) + \sum_{j=1}^{p_m} \hat{\phi}_{j,m} \left(\frac{x_{t(r,m)-j} - \hat{\mu}_{m-j}}{\hat{\sigma}_{m-j}}\right).$$

5. Calcule as estimativas dos parâmetros $\theta_m, \, \mu_m^{(\xi)}$ e $\sigma_m^{(\xi)2}$ com:

$$\hat{\theta}_{m} = 1 + \frac{\hat{\sigma}_{m}^{(a)2}}{(\hat{\mu}_{m}^{(a)} - \hat{\tau}_{m})^{2}},$$

$$\hat{\sigma}_{m}^{(\xi)2} = \log(\hat{\theta}_{m}),$$

$$\hat{\mu}_{m}^{(\xi)} = \frac{1}{2}\log\left(\frac{\hat{\sigma}_{m}^{(a)2}}{\hat{\theta}_{m}^{2} - \hat{\theta}_{m}}\right).$$

- 6. Gere o processo $\xi_{t(r,m)} \sim N(\hat{\mu}_m^{(\xi)}, \hat{\sigma}_m^{(\xi)^2})$.
- 7. Gere a vazão $x_{t(r,m)}$, com

$$x_{t(r,m)} = x_0 + \hat{\sigma}_m \exp\{\xi_{t(r,m)}\}.$$

- 8. Se m < s, faça m = m + 1 e volte para o Passo 4. Caso contrário, vá para o Passo 9.
- 9. Faça m = 1, r = r + 1 e volte para o Passo 4 e atualiza o ano.

Repetindo-se este procedimento recursivamente, tem-se como resultado a geração de uma série sintética de vazões mensais, cuja correlação temporal é representada por um modelo $PAR(p_m)$ e a série tem distribuição Log-Normal com três parâmetros. Observase que a dependência entre as vazões $x_{t(r,m)}$ e $x_{t(r,m)-j}$ são consideradas no modelo por meio do parâmetro τ_m introduzido no cálculo dos parâmetros $\mu_m^{(\xi)}$ e $\sigma_m^{(\xi)2}$.

A geração de séries sintéticas univariadas usando o modelo Log-Normal com três parâmetros é amplamente destacada na literatura. Neste trabalho, é desenvolvida uma abordagem multivariada deste modelo para geração de vazões médias mensais.

4.3 Modelo Log-Normal com três parâmetros multivariado

A modelagem de séries de vazões médias mensais para um conjunto de usinas é amplamente utilizada no planejamento da operação dos sistemas de recursos hídricos de grande porte. Esta modelagem, adotada pelo Setor Elétrico Brasileiro, abrange vastas regiões, ou seja, mais de uma bacia hidrográfica e, assim, é desejável que os estudos com modelos estocásticos multivariados sobre geração de séries sintéticas preservem a correlação espacial das vazões entre os diferentes locais.

O modelo utilizado pelo Setor Elétrico Brasileiro é o GEVAZP (CEPEL, 2002). Este utiliza a abordagem paramétrica, considerando uma distribuição Log-Normal de três parâmetros, para cada mês, ajustada aos ruídos de cada posto. Na parte multivariada, este modelo considera uma multiplicação do vetor de ruídos normais independentes por uma matriz de carga, resultando em ruídos normais espacialmente correlacionados. A correlação espacial é considerada única para todos os estágios de geração, sendo a matriz de carga estimada a partir da matriz de correlações espaciais das vazões.

Nesta tese é proposto o modelo Log-Normal com três parâmetros multivariado, em que não se considera a etapa de utilização da matriz de carga. Todo o desenvolvimento deste modelo é apresentado, a seguir. Posteriormente, este modelo será comparado com o modelo GEVAZP.

Considerando n-usinas, tem-se que para cada usina i = 1, ..., n associa-se um modelo $PAR(p_m)$, dado por:

$$\left(\frac{x_{t(r,m)}^{(i)} - \mu_m^{(i)}}{\sigma_m^{(i)}}\right) = \sum_{j=1}^{p_m^{(i)}} \phi_{j,m}^{(i)} \left(\frac{x_{t(r,m)-j}^{(i)} - \mu_{m-j}^{(i)}}{\sigma_{m-j}^{(i)}}\right) + a_{t(r,m)}^{(i)}.$$
(4.9)

Considerando as restrições $x_{t(r,m)}^{(i)} \geq x_0^{(i)}$, para cada usina i=1,...,n, tem-se:

$$a_{t(r,m)}^{(i)} \ge \frac{x_0^{(i)}}{\sigma_m^{(i)}} - \frac{\mu_m^{(i)}}{\sigma_m^{(i)}} - \sum_{i=1}^{p_m^{(i)}} \phi_{j,m}^{(i)} \left(\frac{x_{t(r,m)-j}^{(i)} - \mu_{m-j}^{(i)}}{\sigma_{m-j}^{(i)}} \right). \tag{4.10}$$

Considera-se que $a_{t(r,m)}^{(i)}$ são ruídos com distribuição Log-Normal com três parâmetros, tal que existe um processo gaussiano $\xi_{t(r,m)}^{(i)}$ dado por:

$$\xi_{t(r,m)}^{(i)} = \log(a_{t(r,m)}^{(i)} + \tau_m^{(i)}), \tag{4.11}$$

com

$$\tau_m^{(i)} = \frac{-x_0^{(i)}}{\sigma_m^{(i)}} + \frac{\mu_m^{(i)}}{\sigma_m^{(i)}} + \sum_{j=1}^{p_m^{(i)}} \phi_{j,m}^{(i)} \left(\frac{x_{t(r,m)-j}^{(i)} - \mu_{m-j}^{(i)}}{\sigma_{m-j}^{(i)}} \right). \tag{4.12}$$

Denota-se, na sequência, a seguinte notação:

$$b_{t(r,m)}^{(i)} = a_{t(r,m)}^{(i)} + \tau_m^{(i)}. (4.13)$$

Desta forma, tem-se:

$$E[b_{t(r,m)}^{(i)}] = E[a_{t(r,m)}^{(i)}] + \tau_m^{(i)}, \tag{4.14}$$

$$Var[b_{t(r,m)}^{(i)}] = Var[a_{t(r,m)}^{(i)}],$$
 (4.15)

$$Cov[b_{t(r,m)}^{(i)}, b_{t(r,m)}^{(j)}] = Cov[a_{t(r,m)}^{(i)}, a_{t(r,m)}^{(j)}].$$
 (4.16)

Denota-se também:

$$\gamma_{(i,j)}^{(a)} = Cov[a_{t(r,m)}^{(i)}, a_{t(r,m)}^{(j)}]. \tag{4.17}$$

Usando $\xi_{t(r,m)}^{(i)},\,i=1,...,n,$ tem-se que:

$$\xi_{t(r,m)}^{(i)} = \log(b_{t(r,m)}^{(i)}). \tag{4.18}$$

Considerando o processo n-dimensional:

$$\boldsymbol{\xi}_{t(r,m)} \sim N(\boldsymbol{\mu}_m^{(\xi)}, \boldsymbol{\Sigma}_m^{(\xi)}), \tag{4.19}$$

em que $\boldsymbol{\mu}_m^{(\xi)}=(\mu_m^{(1)},\mu_m^{(2)},\cdots,\mu_m^{(n)})^t$ e a matriz de covariância é dada por:

$$\Sigma_{m}^{(\xi)} = \begin{pmatrix} \gamma_{m}^{(\xi)}(1,1) & \gamma_{m}^{(\xi)}(1,2) & \gamma_{m}^{(\xi)}(1,3) & \cdots & \gamma_{m}^{(\xi)}(1,n) \\ \gamma_{m}^{(\xi)}(2,1) & \gamma_{m}^{(\xi)}(2,2) & \gamma_{m}^{(\xi)}(2,3) & \cdots & \gamma_{m}^{(\xi)}(2,n) \\ \gamma_{m}^{(\xi)}(3,1) & \gamma_{m}^{(\xi)}(3,2) & \gamma_{m}^{(\xi)}(3,3) & \cdots & \gamma_{m}^{(\xi)}(3,n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{m}^{(\xi)}(n,1) & \gamma_{m}^{(\xi)}(n,2) & \gamma_{m}^{(\xi)}(n,3) & \cdots & \gamma_{m}^{(\xi)}(n,n) \end{pmatrix},$$

tal que

$$\gamma_m^{(\xi)}(i,j) = Cov[\xi_{t(r,m)}^{(i)}, \xi_{t(r,m)}^{(j)}]. \tag{4.20}$$

Vamos relacionar $\gamma_m^{(\xi)}(i,j)$ com $\gamma_m^{(a)}(i,j)$.

Tem-se que:

$$\mu_m^{(\xi)}(i) = E[\xi_{t(r,m)}^{(i)}] \qquad \gamma_m^{(\xi)}(i,i) = Var[\xi_{t(r,m)}^{(i)}],$$
(4.21)

$$\mu_m^{(b)}(i) = E[b_{t(r,m)}^{(i)}] \qquad \gamma_m^{(b)}(i,i) = Var[b_{t(r,m)}^{(i)}],$$
(4.22)

$$\xi_{t(r,m)}^{(i)} = \log(b_{t(r,m)}^{(i)}) \qquad b_{t(r,m)}^{(i)} = \exp(\xi_{t(r,m)}^{(i)}). \tag{4.23}$$

Portanto, para $i = 1, 2, \dots, n$, segue que:

$$\mu_m^{(b)}(i) = E[\exp(\xi_{t(r,m)}^{(i)})] = \exp\left[\mu_m^{(\xi)}(i) + \frac{1}{2}\gamma_m^{(\xi)}(i,i)\right],\tag{4.24}$$

$$\gamma_m^{(b)}(i,i) = Var[\exp(\xi_{t(r,m)}^{(i)})] = \exp[2\mu_m^{(\xi)}(i) + \gamma_m^{(\xi)}(i,i)][\exp(\gamma_m^{(\xi)}(i,i)) - 1]. \tag{4.25}$$

Relacionando (4.24) e (4.25), tem-se que:

$$[\mu_m^{(b)}(i)]^2 = \exp[2\mu_m^{(\xi)}(i) + \gamma_m^{(\xi)}(i,i)] \tag{4.26}$$

$$\frac{\gamma_m^{(b)}(i,i)}{[\mu_m^{(b)}(i)]^2} = \exp(\gamma_m^{(\xi)}(i,i)) - 1 \tag{4.27}$$

$$\gamma_m^{(\xi)}(i,i) = \log \left[1 + \frac{\gamma_m^{(b)}(i,i)}{[\mu_m^{(b)}(i)]^2} \right]. \tag{4.28}$$

Substituindo $\gamma_m^{(\xi)}(i,i)$ em (4.24) , tem-se que:

$$\mu_m^{(\xi)}(i) = \log \left\{ \mu_m^{(b)}(i) - \frac{1}{2} \log \left[1 + \frac{\gamma_m^{(b)}(i,i)}{[\mu_m^{(b)}(i)]^2} \right] \right\}. \tag{4.29}$$

Resumindo, tem-se:

$$\theta_m(i) = 1 + \frac{\gamma_m^{(b)}(i,i)}{[\mu_m^{(b)}(i)]^2},\tag{4.30}$$

$$\gamma_m^{(\xi)}(i,i) = \log[\theta_m(i)],\tag{4.31}$$

$$\mu_m^{(\xi)}(i) = \frac{1}{2} \log \left[\frac{\gamma_m^{(b)}(i,i)}{\theta_m^2(i) - \theta_m(i)} \right], \tag{4.32}$$

em que:

$$\mu_m^{(b)}(i) = \mu_m^{(a)}(i) + \tau_m(i), \tag{4.33}$$

$$\gamma_m^{(b)}(i,i) = \gamma_m^{(a)}(i,i). \tag{4.34}$$

O cálculo da covariância segue da seguinte maneira:

$$\gamma_m^{(b)}(i,j) = E\{[b_{t(r,m)}^{(i)} - \mu_m^{(b)}(i)][b_{t(r,m)}^{(j)} - \mu_m^{(b)}(j)]\},
= E[b_{t(r,m)}^{(i)} b_{t(r,m)}^{(j)}] - \mu_m^{(b)}(i)\mu_m^{(b)}(j).$$
(4.35)

Calculando o valor esperado cruzado:

$$E[b_{t(r,m)}^{(i)}b_{t(r,m)}^{(j)}] = E\{\exp[\xi_{t(r,m)}^{(i)}]\exp[\xi_{t(r,m)}^{(j)}]\},$$

$$= E\{\exp[\xi_{t(r,m)}^{(i)} + \xi_{t(r,m)}^{(j)}]\},$$

$$= \exp\left\{E[\xi_{t(r,m)}^{(i)}] + E[\xi_{t(r,m)}^{(j)}] + \frac{1}{2}Var[\xi_{t(r,m)}^{(i)} + \xi_{t(r,m)}^{(j)}]\right\},$$

$$= \exp\left\{\mu_m^{(\xi)}(i) + \mu_m^{(\xi)}(j) + \frac{1}{2}Var[\xi_{t(r,m)}^{(i)} + \xi_{t(r,m)}^{(j)}]\right\}.$$

$$(4.36)$$

O cálculo da variância da soma é dado por:

$$Var[\xi_{t(r,m)}^{(i)} + \xi_{t(r,m)}^{(j)}] = Var[\xi_{t(r,m)}^{(i)}] + Var[\xi_{t(r,m)}^{(j)}] + 2Cov[\xi_{t(r,m)}^{(i)}, \xi_{t(r,m)}^{(j)}],$$

$$= \gamma_m^{(\xi)}(i,i) + \gamma_m^{(\xi)}(j,j) + 2\gamma_m^{(\xi)}(i,j).$$
(4.37)

Substituindo, tem-se:

$$E[b_{t(r,m)}^{(i)}b_{t(r,m)}^{(j)}] = \exp\{\mu_m^{(\xi)}(i) + \mu_m^{(\xi)}(j) + 0.5[\gamma_m^{(\xi)}(i,i) + \gamma_m^{(\xi)}(j,j) + 2\gamma_m^{(\xi)}(i,j)]\}.$$

$$(4.38)$$

Assim, o cálculo da covariância é dado por:

$$\gamma_m^{(b)}(i,j) = \exp\{\mu_m^{(\xi)}(i) + \mu_m^{(\xi)}(j) + 0.5[\gamma_m^{(\xi)}(i,i) + \gamma_m^{(\xi)}(j,j)]\}
\exp\{\gamma_m^{(\xi)}(i,j)\} - \mu_m^{(b)}(i) - \mu_m^{(b)}(j).$$
(4.39)

Sendo,

$$\mu_m^{(b)}(i) = \exp[\mu_m^{(\xi)}(i) + 0.5\gamma_m^{(\xi)}(i,i)]. \tag{4.40}$$

e

$$\mu_m^{(b)}(j) = \exp[\mu_m^{(\xi)}(j) + 0.5\gamma_m^{(\xi)}(j,j)]. \tag{4.41}$$

Então,

$$\gamma_m^{(b)}(i,j) = \exp\{\mu_m^{(\xi)}(i) + \mu_m^{(\xi)}(j) + 0.5[\gamma_m^{(\xi)}(i,i) + \gamma_m^{(\xi)}(j,j)]\} \{\exp[\gamma_m^{(\xi)}(i,j)] - 1\}.$$
(4.42)

Considerando os termos:

$$\mu_{m}^{(\xi)}(i) + \mu_{m}^{(\xi)}(j) + 0.5[\gamma_{m}^{(\xi)}(i,i) + \gamma_{m}^{(\xi)}(j,j)] = 0.5 \log \left[\frac{\gamma_{m}^{(b)}(i,i)}{\theta_{m}^{2}(i) - \theta_{m}(i)} \right] +$$

$$+ 0.5 \log \left[\frac{\gamma_{m}^{(b)}(j,j)}{\theta_{m}^{2}(j) - \theta_{m}(j)} \right] + 0.5 \{ \log[\theta_{m}(i)] + \log[\theta_{m}(j)] \} =$$

$$= \log \left\{ \frac{\gamma_{m}^{(b)}(i,i)\gamma_{m}^{(b)}(j,j)\theta_{m}(i)\theta_{m}(j)}{[\theta_{m}^{2}(i) - \theta_{m}(i)][\theta_{m}^{2}(j) - \theta_{m}(j)]} \right\}^{0.5}.$$

$$(4.43)$$

Tem-se que:

$$\gamma_m^{(b)}(i,j) = \left[\frac{\gamma_m^{(b)}(i,i)\gamma_m^{(b)}(j,j)\theta_m(i)\theta_m(j)}{[\theta_m^2(i) - \theta_m(i)][\theta_m^2(j) - \theta_m(j)]} \right]^{0.5} \{ \exp[\gamma_m^{(\xi)}(i,j)] - 1 \}.$$
 (4.44)

Denotando por:

$$\Delta_m(i,j) = \frac{\gamma_m^{(b)}(i,j)}{\left[\frac{\gamma_m^{(b)}(i,i)\gamma_m^{(b)}(j,j)\theta_m(i)\theta_m(j)}{[\theta_m^2(i)-\theta_m(i)][\theta_m^2(j)-\theta_m(j)]}\right]^{0.5}}.$$
(4.45)

Tem-se que:

$$\exp[\gamma_m^{(\xi)}(i,j)] - 1 = \Delta_m(i,j). \tag{4.46}$$

Portanto,

$$\gamma_m^{(\xi)}(i,j) = \log[1 + \Delta_m(i,j)]. \tag{4.47}$$

Resumindo, tem-se para cada usina i = 1, 2, ..., n:

$$\theta_m(i) = 1 + \frac{\gamma_m^{(a)}(i,i)}{[\mu_m^{(a)}(i) - \tau_m(i)]^2},\tag{4.48}$$

$$\mu_m^{(\xi)}(i) = 0.5 \log \left[\frac{\gamma_m^{(a)}(i,i)}{\theta_m^2(i) - \theta_m(i)} \right],$$
(4.49)

$$\gamma_m^{(\xi)}(i) = \log[\theta_m(i)],\tag{4.50}$$

$$\Delta_m(i,j) = \frac{\gamma_m^{(a)}(i,j)}{\left[\frac{\gamma_m^{(a)}(i,i)\gamma_m^{(a)}(j,j)\theta_m(i)\theta_m(j)}{[\theta_m^2(i)-\theta_m(i)][\theta_m^2(j)-\theta_m(j)]}\right]^{0.5}},$$
(4.51)

$$\gamma_m^{(\xi)}(i,j) = \log[1 + \Delta_m(i,j)]. \tag{4.52}$$

Portanto $\boldsymbol{\xi}_{t(r,m)}^{(i)}, i=1,2,...,n,$ é uma Normal n-variada com $\boldsymbol{\xi}_{t(r,m)} \sim N[\boldsymbol{\mu}_m^{(\xi)}, \boldsymbol{\Sigma}_m^{(\xi)}],$ em que:

$$\boldsymbol{\mu}_{m}^{(\xi)} = (\mu_{m}^{(1)}, \mu_{m}^{(2)}, \cdots, \mu_{m}^{(n)})^{t} \tag{4.53}$$

e a matriz de covariância é dada por:

$$\begin{pmatrix} \gamma_m^{(\xi)}(1,1) & \gamma_m^{(\xi)}(1,2) & \gamma_m^{(\xi)}(1,3) & \cdots & \gamma_m^{(\xi)}(1,n) \\ \gamma_m^{(\xi)}(2,1) & \gamma_m^{(\xi)}(2,2) & \gamma_m^{(\xi)}(2,3) & \cdots & \gamma_m^{(\xi)}(2,n) \\ \gamma_m^{(\xi)}(3,1) & \gamma_m^{(\xi)}(3,2) & \gamma_m^{(\xi)}(3,3) & \cdots & \gamma_m^{(\xi)}(3,n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_m^{(\xi)}(n,1) & \gamma_m^{(\xi)}(n,2) & \gamma_m^{(\xi)}(n,3) & \cdots & \gamma_m^{(\xi)}(n,n) \end{pmatrix}$$

Para gerar amostras de $\boldsymbol{\xi}_{t(r,m)}$:

1) Considera-se a decomposição de Cholesky da matriz de covariância $\Sigma_m^{(\xi)}$ dada por:

$$\mathbf{V}_{m}^{(\xi)} = chol[\mathbf{\Sigma}_{m}^{(\xi)}],\tag{4.54}$$

tal que:

$$\Sigma_m^{(\xi)} = [V_m^{(\xi)}]^t [V_m^{(\xi)}]. \tag{4.55}$$

- 2) Seja $\mathbf{Z} = (Z^{(1)}, Z^{(2)}, ..., Z^{(n)})^t$ uma variável aleatória Normal com $\mathbf{Z} \sim N(0, \mathbf{I}_{nxn})$, em que \mathbf{I}_{nxn} é a matriz identidade.
 - 3) Então tem-se que:

$$\boldsymbol{\xi}_{t(r,m)} = \boldsymbol{\mu}_m^{(\xi)} + \boldsymbol{V}_m^{(\xi)t} \boldsymbol{Z}, \tag{4.56}$$

tal que $\pmb{\xi}_{t(r,m)} = (\xi_{t(r,m)}^{(1)}, \xi_{t(r,m)}^{(2)}, ..., \xi_{t(r,m)}^{(n)})$ e

$$\Sigma_{m}^{(\xi)} = E\{ [\boldsymbol{\xi}_{t(r,m)} - \boldsymbol{\mu}_{m}^{(\xi)}] [\boldsymbol{\xi}_{t(r,m)} - \boldsymbol{\mu}_{m}^{(\xi)}]^{t} \},$$

$$= E\{ [\boldsymbol{V}_{m}^{(\xi)t} \boldsymbol{Z}] [\boldsymbol{V}_{m}^{(\xi)t} \boldsymbol{Z}]^{t} \},$$

$$= [\boldsymbol{V}_{m}^{(\xi)t}] E[\boldsymbol{Z} \boldsymbol{Z}^{t}] [\boldsymbol{V}_{m}^{(\xi)}],$$

$$= [\boldsymbol{V}_{m}^{(\xi)t}] [\boldsymbol{I}_{nxn}] [\boldsymbol{V}_{m}^{(\xi)}],$$

$$= [\boldsymbol{V}_{m}^{(\xi)}]^{t} [\boldsymbol{V}_{m}^{(\xi)}].$$

$$(4.57)$$

Na sequência, é desenvolvido um modelo Log-Normal generalizado para geração de séries sintéticas univariadas usando, nos procedimentos de estimação, a abordagem bayesiana.

4.4 Modelo Log-Normal generalizado

Uma das mais comuns e importantes distribuições de probabilidade usadas em Hidrologia é a distribuição Log-Normal. Porém, em algumas situações, esta distribuição não apresenta uma flexibilidade suficiente para a modelagem de vazões médias mensais. Neste trabalho, considera-se uma generalização da distribuição Log-Normal, a distribuição Log-Normal generalizada (GLN).

Esta distribuição é analisada por meio da abordagem bayesiana usando distribuições a priori não informativas de Jeffreys independentes para os parâmetros de um modelo PAR(1), ou seja, para cada mês é estimado um modelo AR de ordem fixa 1. A ordem fixa utilizada aqui é justificada pelo fato da sua utilização nos procedimentos da programação dinâmica dual estocástica. Posteriormente, é desenvolvido um gerador univariado de séries sintéticas de vazões médias mensais.

Uma variável aleatória X tem distribuição Log-Normal generalizada se a transformação $Y = \log(X)$ tem distribuição Normal generalizada. A função densidade de probabilidade para uma variável GLN com parâmetros μ , σ e s é dada por:

$$f(x|\mu,\sigma,s) = \frac{s}{2x\sigma\Gamma(\frac{1}{s})} \exp\left(-\left|\frac{\log(x) - \mu}{\sigma}\right|^{s}\right),\tag{4.58}$$

em que $x > 0, -\infty < \mu < +\infty, \sigma > 0, s \ge 1$ e Γ denota a função gama.

Uma proposição foi apresentada por Martín & Pérez (2009) e mostra as relações entre as distribuições Gama e GLN.

Proposição 1. Se U e X são variáveis aleatórias tais que $U \sim G(\alpha = 1+1/s, \gamma = 1)$ e $f(x|u) = I[\exp(\mu - \sigma u^{1/s}) < x < \exp(\mu + \sigma u^{1/s})]/(2x\sigma u^{1/s})$ então $X \sim GLN(\mu, \sigma, s)$, em que I[.] representa uma função indicadora.

4.4.1 Análise bayesiana

De acordo com a proposição 1 é possível representar a distribuição GLN em (4.58) como uma distribuição marginal dada por:

$$f(z_{t(r,m)}|z_{t(r,m)-1},\alpha_m,\beta_m,\sigma_m,s_m,u_{r,m}) = \frac{1}{2z_{t(r,m)}\sigma_m u_{r,m}^{1/s_m}}I[.],$$
(4.59)

$$f(u_{r,m}|s_m) = G(1+1/s_m, 1), \tag{4.60}$$

em que $I[.] = I[\exp\{\alpha_m + \beta_m z_{t(r,m)-1} - \sigma_m u_{r,m}^{1/s_m}\} < z_{t(r,m)} < \exp\{\alpha_m + \beta_m z_{t(r,m)-1} + \sigma_m u_{r,m}^{1/s_m}\}]$ denota a função indicadora e $u_{r,m}$ representa uma variável latente.

Nesta etapa do trabalho, utiliza-se a seguinte representação para o modelo PAR(1):

$$z_{t(r,m)} = \alpha_m + \beta_m z_{t(r,m)-1} + a_{t(r,m)}. \tag{4.61}$$

Os procedimentos da análise bayesiana iniciam com a determinação da função de verossimilhança para os parâmetros α_m , β_m , σ_m , s_m e $\mathbf{u}_m = (u_{1,m}, u_{2,m}, ..., u_{n,m})'$. Esta função é definida por:

$$L(\mathbf{z_m}|\alpha_m, \beta_m, \sigma_m, s_m, \mathbf{u_m}) = \frac{1}{(2\sigma_m)^n} \prod_{r=1}^n \frac{1}{z_{t(r,m)} u_{r,m}^{1/s_m}} I[.],$$
(4.62)

em que
$$I[.] = I[\exp{\{\alpha_m + \beta_m z_{t(r,m)-1} - \sigma_m u_{r,m}^{1/s_m}\}} < z_{t(r,m)} < \exp{\{\alpha_m + \beta_m z_{t(r,m)-1} + \sigma_m u_{r,m}^{1/s_m}\}}].$$

Martín & Pérez (2009) propuseram o uso de distribuições a priori não informativas de Jeffreyes independentes para os parâmetros da distribuição GLN. Sendo assim, temse que $\pi(\alpha_m, \beta_m) \propto constante$, $\pi(\sigma_m) \propto \frac{1}{\sigma_m}$ e $\pi(s_m) \propto \frac{1}{s_m}$.

Combinando a função de verossimilhança com as distribuições *a priori* define-se a distribuição *a posteriori* dada por:

$$\pi(\alpha_m, \beta_m, \sigma_m, s_m, \boldsymbol{u_m} | \boldsymbol{z_m}) \propto \frac{s_m^{n-1}}{\sigma_m^{n+1} \Gamma^n \left(\frac{1}{s_m}\right)} \prod_{r=1}^n \frac{\exp\{-u_{r,m}\}}{z_{t(r,m)}} I[.], \tag{4.63}$$

em que
$$I[.] = I[\exp\{\alpha_m + \beta_m z_{t(r,m)-1} - \sigma_m u_{r,m}^{1/s_m}\} < z_{t(r,m)} < \exp\{\alpha_m + \beta_m z_{t(r,m)-1} + \sigma_m u_{r,m}^{1/s_m}\}].$$

A partir da distribuição *a posteriori* define-se as distribuições condicionais, para cada um dos parâmetros, como:

$$\pi(\alpha_m|\beta_m, \sigma_m, s_m, \boldsymbol{u_m}, \boldsymbol{z_m}) \propto 1,$$
 (4.64)

$$I\left[max_r \{ \log[z_{t(r,m)}] - \sigma_m u_{r,m}^{1/s_m} - \beta_m z_{t(r,m)-1} \} < \alpha_m < min_r \{ \log[z_{t(r,m)}] + \sigma_m u_{r,m}^{1/s_m} - \beta_m z_{t(r,m)-1} \} \right],$$

$$\pi(\beta_m | \alpha_m, \sigma_m, s_m, \boldsymbol{u_m}, \boldsymbol{z_m}) \propto 1,$$
 (4.65)

em que
$$I\left[\max_{r}\left\{\frac{\log[z_{t(r,m)}] - \sigma_{m}u_{r,m}^{1/s_{m}} - \alpha_{m}}{z_{t(r,m)-1}}\right\} < \beta_{m} < \min_{r}\left\{\frac{\log[z_{t(r,m)}] + \sigma_{m}u_{r,m}^{1/s_{m}} - \alpha_{m}}{z_{t(r,m)-1}}\right\}\right],$$

$$\pi(\sigma_m | \alpha_m, \beta_m, s_m, \boldsymbol{u_m}, \boldsymbol{z_m}) \propto \frac{1}{\sigma_m^{n+1}},$$
 (4.66)

em que
$$I\left[\sigma_m > \max_r \left\{ \frac{|\alpha_m + \beta_m z_{t(r,m)-1} - \log[z_{t(r,m)}]|}{u_{r,m}^{1/s_m}} \right\} \right]$$

$$\pi(s_m | \alpha_m, \beta_m, \sigma_m, \boldsymbol{u_m}, \boldsymbol{z_m}) \propto \frac{s_m^{n-1}}{\Gamma^n \left(\frac{1}{s_m}\right)},$$
 (4.67)

em que
$$I\left[s_m > \max_r \left\{ \frac{\log(u_{r,m})}{\log(|\alpha_m + \beta_m z_{t(r,m)-1} - \log[z_{t(r,m)}]|/\sigma_m)} \right\} \right]$$

$$\pi(u_{r,m}|\alpha_m, \beta_m, \sigma_m, s_m, \boldsymbol{z_m}) \propto \exp\{-u_{r,m}\},$$
 (4.68)

em que
$$I\left[u_{r,m} > \left(\frac{|\log[z_{t(r,m)}] - \alpha_m - \beta_m z_{t(r,m)-1}|}{\sigma_m}\right)^{s_m}\right], r = 1, 2, ..., n.$$

No processo de simulação MCMC, a partir das distribuições condicionais, é impor-

No processo de simulação MCMC, a partir das distribuições condicionais, é importante identificar os núcleos característicos das distribuições. Note que as densidades dadas em (4.64), (4.65), (4.66) e (4.68) são dadas pelas distribuições Uniforme, Uniforme, Pareto e Exponencial, respectivamente. A densidade dada em (4.67) não apresenta um núcleo padrão e, neste caso, utiliza-se o algoritmo de Metropolis-Hastings.

4.4.2 Geração univariada de série sintética

Após a estimação dos parâmetros, o modelo GLN pode ser usado na geração univariada de séries sintéticas de vazões médias mensais. Neste trabalho, são gerados 2000 anos de séries sintéticas para análises. O seguinte algoritmo é implementado para facilitar o processo de geração:

- 1. Gere $U \sim G(1 + 1/s_m, 1)$;
- 2. Gere $V \sim U(-1, 1)$;
- 3. Defina $z_{t(r,m)} = \exp\{\sigma_m U^{1/s_m} V + \alpha_m + \beta_m z_{t(r,m)-1}\};$
- 4. Repita os passos (1) (3) até que o tamanho desejado das séries sintéticas seja atingido.

As séries sintéticas geradas pelos modelos desenvolvidos devem ser validadas mediante algum critério de diagnóstico. Neste trabalho, é adotada a divergência de Kullback-Leibler.

4.5 Critério de validação da série gerada

Uma série de vazão gerada de uma distribuição de probabilidade ajustada pode não representar adequadamente as características estatísticas da série real. Isto deve-se,

principalmente, as limitações dos modelos propostos e a qualidade dos ajustes destes modelos. Uma alternativa para avaliar se uma série de vazões gerada da distribuição de probabilidade ajustada representa uma série de vazão plausível é comparar a densidade da série gerada, denotada aqui por $f_{\hat{q}}$, com a densidade da série real (f_q) , com ambas densidades calculadas de forma não-paramétrica. Considera-se que as densidades são avaliadas em M pontos de forma que tem-se, para ambas, os pares $\{(\hat{q}_1, f_{\hat{q}}^1), \dots (\hat{q}_M, f_{\hat{q}}^M)\}$ e $\{(q_1, f_q^1), \dots (q_M, f_q^M)\}$. Estes pontos podem ser avaliados construindo um histograma das séries e adotando cada ponto como o ponto médio de cada classe do histograma. A comparação é feita calculando a divergência de Kullback-Leibler (Kullback & Leibler, 1951), que mede a distância entre duas densidades de probabilidade, dada por

$$KL(f_q(q), f_{\hat{q}}(\hat{q})) = \int f_q(q) \log \left(\frac{f_q(q)}{f_{\hat{q}}(\hat{q})}\right) dq.$$

$$(4.69)$$

Para se avaliar a divergência $KL(f_q(q), f_{\hat{q}}(\hat{q}))$ considerando-se os pontos avaliados, considera-se um método de integração numérica, tal como o método dos trapézios. Neste caso, tem-se que:

$$KL(f_q(q), f_{\hat{q}}(\hat{q})) = \left\{ \frac{f_q^1}{2} \log \left(\frac{f_q^1}{f_{\hat{q}}^1} \right) + \sum_{i=2}^{M-1} f_q^i \log \left(\frac{f_q^i}{f_{\hat{q}}^i} \right) + \frac{f_q^M}{2} \log \left(\frac{f_q^M}{f_{\hat{q}}^M} \right) \right\} \delta q, \tag{4.70}$$

em que δq representa a amplitude das classes do histograma.

Uma medida de calibração para a divergência $KL(f_q(q), f_{\hat{q}}(\hat{q}))$, denotada aqui por p, é obtida resolvendo-se a equação:

$$KL(f_q(q), f_{\hat{q}}(\hat{q})) = KL(B(0.5), B(p)) = -\log[4p(1-p)/2],$$
 (4.71)

em que B(p) denota a distribuição de Bernoulli com probabilidade de sucesso p. Isto significa aceitar a série gerada com a densidade de probabilidade $f_{\hat{q}}(\hat{q})$ como vinda da distribuição de probabilidade $f_q(q)$ e é equivalente a descrever um evento não observado como tendo probabilidade p, quando a probabilidade correta é 0.5. Resolvendo-se a equação para p, tem-se que:

$$p = \frac{1}{2} \left\{ 1 + \sqrt{1 - \exp[-2KL(f_q(q), f_{\hat{q}}(\hat{q}))]} \right\}.$$
 (4.72)

Isto implica que $0.5 \le p \le 1$. Portanto, quanto menor a divergência entre as densidades $f_q(q)$ e $f_{\hat{q}}(\hat{q})$, tem-se p mais próximo do valor verdadeiro $(p \approx 0, 5)$. Para $p \gg 0, 5$, tem-se que a divergência entre as densidades $f_q(q)$ e $f_{\hat{q}}(\hat{q})$ é grande e, conse-

quentemente, a série gerada, provavelmente, não tem mesma distribuição da série real. Recomenda-se a rejeição da série gerada para $p \geq 0,95$.

Capítulo 5

Aplicação

Este capítulo apresenta alguns estudos de caso sobre os processos de estimação e previsão, via abordagens clássica e bayesiana e sobre os procedimentos para geração de séries sintéticas usando uma abordagem univariada e multivariada. Algumas considerações são feitas no início deste capítulo.

5.1 Considerações iniciais

Neste trabalho, primeiramente, os modelos $PAR(p_m)$ são ajustados às séries históricas de vazões médias mensais pertencentes aos reservatórios das usinas hidroelétricas de Emborcação, Furnas e Sobradinho no período de janeiro de 1931 a dezembro de 2010 (80 anos). Estes dados são utilizados nos processos de estimação e previsão, via abordagens clássica.

Em seguida, os modelos $PAR(p_m)$ são ajustados às séries históricas de vazões médias mensais pertencentes aos reservatórios das usinas de Água Vermelha e São Simão no período de janeiro de 1931 a dezembro de 2010 (80 anos). Estes dados são utilizados nos processos de estimação e previsão, via abordagem bayesiana.

Na geração de séries sintéticas usando a abordagem multivariada foram utilizados os seguintes reservatórios principais: Cachoeira Dourada, Capim Branco 2, Corumbá 1, Corumbá 4, Emborcação, Espora, Foz do Rio Claro, Salto do Rio Verdinho, São Simão, Água Vermelha, Armando de Salles Oliveira, Candonga, Canoas 1, Fundão, Funil Grande, Governador Pedro Viriato Parigot de Souza, Guarapiranga, Ilha Pombos, Itaipu, Itutinga, Jaguari, Jauru, Lajes, Manso, Mascarenhas, Navanhandava, Porto Colômbia, Porto Estrela, Picada, Ponte Pedra, Porto Primavera, Retiro, Rosal, Sá Carvalho, Santa Branca, Sobragi, Três Marias, Xingó.

Na geração univariada de séries sintéticas, usando o modelo Log-Normal generalizado, utilizou-se, para as análises, o reservatório de Furnas. Todos estes reservatórios, utilizados para geração, apresentam séries históricas de vazões médias mensais no período de janeiro de 1931 a dezembro de 2010 (80 anos). Todos os procedimentos descritos foram implementados em linguagem de programação C e R.

5.2 Abordagem clássica dos modelos PAR

As comparações entre os modelos de previsão são avaliadas considerando-se a raiz quadrada do erro quadrático médio (RMSE) da previsão, feita com cada modelo ajustado, em função do passo de previsão. São usados como referência os desvio padrão mensal da série original.

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

Na Tabela 5.1 são apresentadas, como ilustração, as ordens dos modelos $PAR(p_m)$ selecionadas com os três métodos considerados, para as vazões de Furnas. Na Tabela 5.2 encontram-se os modelos mensais ajustados para as vazões de Furnas quando a ordem de cada modelo foi selecionada pelo método clássico de análise da função de autocorrelação parcial periódica. Portanto, estes modelos apresentam alguns coeficientes intermediários nulos.

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

Tabela 5.1: Ordens dos modelos para as vazões de Furnas.

Tabela 5.2: Modelos ajustados com o método clássico para as vazões de Furnas.

mês	$\phi_{1,m}$	$\phi_{2,m}$	$\phi_{3,m}$	$\phi_{4,m}$	$\phi_{5,m}$	$\phi_{6,m}$
1	0,52	0	0	-0, 3	0, 3	0
2	0,44	0	0	0	0	0
3	0,43	0,28	0	0	0,33	-0,29
4	0,59	0, 26	0	0	0	0
5	0,51	0	0,27	0	0	0
6	0,84	0	0	0	0	0
7	0,6	0,37	0	0	0	0
8	0,93	0	0	0	0	0
9	0, 5	0,32	0,43	-0,43	0	0
10	0,4	-0,32	0,44	0,46	0	-0,27
11	0,72	0, 25	-0,26	0,45	-0,44	0
12	0,51	0	0	0, 5	-0,56	0,35

O método de escolha da ordem de cada modelo, utilizado pelo método clássico, resulta em ordens mais elevadas do que as ordens escolhidas pelo método proposto por Stedinger (2001) ou pelo método proposto por McLeod (1994) que faz uso do critério BIC. Os procedimentos propostos por Stedinger (2001) e McLeod (1994), na maioria, dos meses coincidem na escolha das ordens dos modelos, diferenciando-se em poucos meses e em apenas uma unidade na ordem selecionada (ver Tabela 5.1). Apesar do método clássico selecionar modelos de ordem mais elevadas, vários termos intermediários são nulos, como pode ser visto nos modelos ajustados para as vazões de Furnas, apresentados na Tabela 5.2.

Para análise dos erros de previsão usa-se como referência o coeficiente de variação (cv_m) de cada mês. Na Tabela 5.3 apresenta-se os coeficientes de variação cv_m e a razão entre o RMSE (calculado com o passo de previsão h=1) e a MLT de cada mês para as vazões de Furnas. Nota-se um alto coeficiente de variação no mês de setembro,

devido às vazões ocorridas neste mês, nos anos de 1976 e 1983, que foram muito acima dos valores normalmente observados para as vazões de setembro.

Nota-se na Tabela 5.3 que os meses de janeiro, fevereiro e março apresentam maior erro relativo de previsão RMSE/MLT (em %) quando comparados com os demais meses. Meses com alto coeficiente de variação são mais difíceis de serem previstos, consequentemente apresentam maiores erros de previsão, como pode ser visto na Tabela 5.3.

Método	jan	fev	mar	abr	mai	jun
$\overline{cv_m}$	40,95	39, 39	40,72	34,79	30,46	42,09
Clássico	35, 18	35, 22	32, 17	21,91	14, 22	22,61
Stedinger (2001)	34,97	35, 22	33, 20	21,91	14, 22	22,61
BIC	34,97	33,57	33, 20	21,91	13,61	22,61
Método	jul	ago	set	out	nov	dez
$\overline{cv_m}$	30,39	30,05	54, 23	43, 16	43,00	38,76
Clássico	10, 13	10, 51	27,47	23, 25	28,91	27,84
Stedinger (2001)	10, 13	10,51	27,47	23,78	29, 29	28, 17
BIC	10, 13	10,51	27,47	23,78	29, 29	28,17

Tabela 5.3: Razão RMSE/MLT (em %) com h=1 para as vazões de Furnas.

Portanto, o aumento na ordem dos modelos, para os meses com alto coeficiente de variação, não resulta em diminuição significativa dos erros das previsões das vazões destes meses, como pode ser visto na Tabela 5.3, para os meses de janeiro, fevereiro e março, com o passo de previsão de um mês a frente (h = 1).

A Figura 5.1, ilustra os erros de previsão para as vazões afluentes em Furnas, no mês de janeiro, onde observa-se que os erros da previsão, cometidos com as três propostas de identificação consideradas neste trabalho, são praticamente iguais e crescem rapidamente, chegando a mais de 90% do limite superior nos três primeiros meses.

No entanto, para os meses com baixo coeficiente de variação, como o mês de agosto, ao se considerar modelos de ordem mais elevada obtêm-se uma diminuição nos erros de previsão a longo prazo. O gráfico apresentado na Figura 5.2 ilustra este fato para a previsão da vazão afluente à Furnas no mês de agosto. Independente do critério adotado na identificação da ordem dos modelos PAR, as ordens dos modelos ajustados para os três meses que precedem o mês de agosto são iguais. Este fato é refletido no erro de previsão para este mês, notando-se que somente para passos de previsão $h \geq 4$ surgem as diferenças nos erros de previsões que é provocada pela diferença nas ordens dos modelos ajustados.

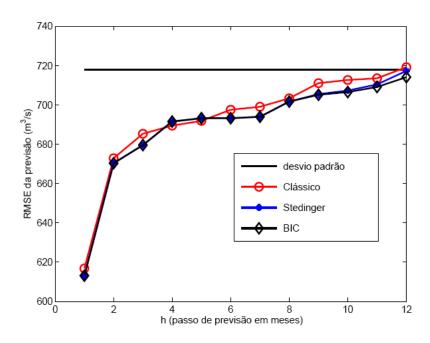


Figura 5.1: RMSE da previsão da vazão de Furnas para o mês de janeiro.

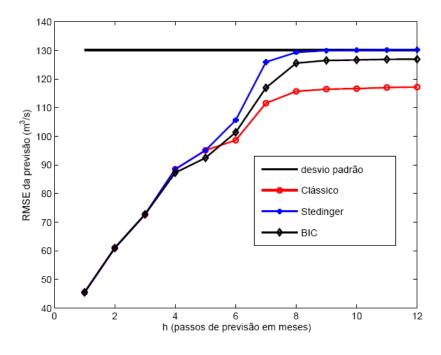


Figura 5.2: RMSE da previsão da vazão de Furnas para o mês de agosto.

dez

6

1

1

nov

4

2

2

out

6

1

1

Uma análise semelhante foi feita para as vazões afluentes à usina de Emborcação. Na Tabela 5.4 são apresentadas as ordens dos modelos selecionadas com as três propostas consideradas neste trabalho. Nota-se para estes dados, que o método clássico seleciona ordens mais elevadas que os outros dois métodos. Observando-se os modelos ajustados, apresentados na Tabela 5.5, nota-se que vários destes modelos apresentam estimativas dos parâmetros intermediários $\phi_{m,j}$, com $1 \le j \le p_m$ nulos, ressaltando a incoerência deste método de identificação da ordem dos modelos com os fenômenos naturais.

Método	jan	fev	mar	abr	mai	jun
Clássico	6	1	1	5	3	1
Stedinger (2001)	1	1	1	1	1	1
RIC	1	2	1	2	3	1

ago

3

1

3

set

4

1

1

jul

2

2

2

Método

Clássico

BIC

Stedinger (2001)

Tabela 5.4: Ordens dos modelos para as vazões de Emborcação.

Tabela 5.5:	Modelos aiustados o	com o método clássico :	para as vazões de Emborcação.

mês	$\phi_{1,m}$	$\phi_{2,m}$	$\phi_{3,m}$	$\phi_{4,m}$	$\phi_{5,m}$	$\phi_{6,m}$
1	0,37	0,24	-0,39	0	0	0,49
2	0,6	0	0	0	0	0
3	0,49	0	0	0	0	0
4	0,54	0	0	0	0, 26	0
5	0,66	0	0, 29	0	0	0
6	0,88	0	0	0	0	0
7	0,68	0,31	0	0	0	0
8	0,89	0	0,28	0	0	0
9	1,01	0,47	-0,29	-0,35	0	0
10	0,38	0	0,37	-0,34	0	0,36
11	0,73	0	-0,54	0,42	0	0
12	0,54	0	0	0,47	-1, 12	0,62

A Tabela 5.6 apresenta, para cada mês do ano, o coeficiente de variação e a razão RMSE/MLT (%) para cada método de identificação da ordem do modelo $PAR(p_m)$, para a usina de Emborcação. O mês de fevereiro apresenta o maior coeficiente de variação e, consequentemente, os maiores erros de previsão com h=1.

A Figura 5.3 mostra como estes erros de previsão tendem rapidamente para o limite superior, definido pelo desvio padrão das vazões deste mês, quando o passo de previ-

Método	jan	fev	mar	abr	mai	jun
$\overline{cv_m}$	38, 59	47,03	42, 23	37,64	30,92	29,99
Clássico	34,07	37, 33	36,77	24,32	11,52	13,91
Stedinger (2001)	35, 28	37, 33	36,77	26, 36	14, 11	13,91
BIC	35, 28	36,01	36,77	25,47	11,52	13,91
Método	jul	ago	set	out	nov	dez
$\overline{cv_m}$	29,32	30,04	32,01	34,87	47,01	46,20
Clássico	7,01	8,94	15,01	28,49	38,08	38,97
Stedinger (2001)	7,01	8,94	15,62	28,49	38, 12	38,86
BIC	7,01	8,23	15,62	28, 49	38, 12	38,86

Tabela 5.6: Razão RMSE/MLT (em %) com h=1 paras as vazões de Emborcação.

são cresce. Resultados semelhantes são encontrados para os meses de janeiro, março, novembro e dezembro. Para os meses com baixo coeficiente de variação observa-se a diminuição do erro de previsão para os modelos de ordem maiores. Isso é ilustrado na Figura 5.4, a qual apresenta os erros das previsões para as vazões de Emborcação no mês de julho.

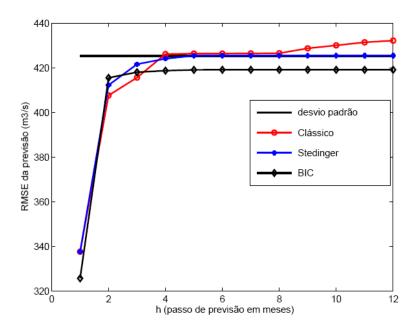


Figura 5.3: RMSE da previsão da vazão de Emborcação para o mês de fevereiro.

Na Tabela 5.7 são apresentadas as ordens dos modelos selecionadas, com os três critérios considerados neste trabalho, para as vazões de Sobradinho. Nota-se que todos os métodos resultam na mesma escolha da ordem dos modelos em quase todos os meses,

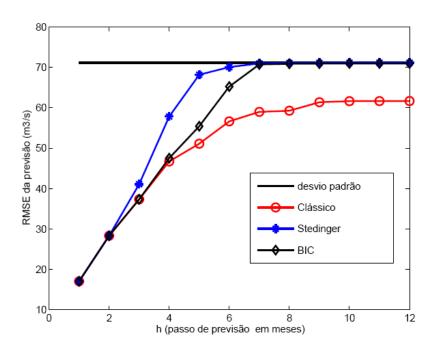


Figura 5.4: RMSE da previsão da vazão de Emborcação para o mês de julho.

exceto para os meses de fevereiro e dezembro, nos quais o critério clássico selecionou modelos de ordem mais elevada e nos meses de junho e julho, no qual o critério BIC selecionou modelos de ordem mais elevada.

Tabela 5.7:	Ordens	dos	modelos	nara	ลร	vazões	de	Sobradinho
Tabula 0.1.	Oraciis	uos	modelos	para	ab	vazocs	uc	bobiadiiiio.

Método	jan	fev	mar	abr	mai	jun
Clássico	1	4	1	1	1	2
Stedinger (2001)	1	1	1	1	1	2
BIC	1	1	1	1	1	5
N/4+-1-	• 1					-
Método	jul	ago	set	out	nov	dez
Clássico	1 jui	ago 1	$\frac{\text{set}}{1}$	out 1	nov 1	dez 5
	1 1	ago 1 1 1	$\frac{\text{set}}{1}$	1 1	nov 1 1	

A Tabela 5.8 apresenta as estimativas dos parâmetros dos modelos ajustados para a vazão de Sobradinho, em cada mês, quando foi adotado o critério clássico de seleção das ordens dos modelos. Nota-se nesta tabela, que os modelos de ordem elevada, apresentam estimativas nulas para vários parâmetros intermediários, ressaltando-se a incoerência entre o método utilizado e o fenômeno natural.

mês	$\phi_{1,m}$	$\phi_{2,m}$	$\phi_{3,m}$	$\phi_{4,m}$	$\phi_{5,m}$	$\phi_{6,m}$
1	0,57	0	0	0	0	0
2	0,42	0	0	0,33	0	0
3	0,74	0	0	0	0	0
4	0,61	0	0	0	0	0
5	0,79	0	0	0	0	0
6	0,67	0,31	0	0	0	0
7	0,97	0	0	0	0	0
8	0,97	0	0	0	0	0
9	0,94	0	0	0	0	0
10	0,79	0	0	0	0	0
11	0,63	0	0	0	0	0
12	0,56	0	0	-0.92	0.82	0

Tabela 5.8: Modelos ajustados com o método clássico para as vazões de Sobradinho.

A Tabela 5.9 apresenta, para cada mês do ano, o coeficiente de variação e a razão RMSE/MLT (%) para cada método de identificação da ordem do modelo $PAR(p_m)$, na usina de Sobradinho. As vazões afluentes ao reservatório de Sobradinho apresentam o maior coeficiente de variação no mês de março e o menor se verifica no mês de setembro. A variação do erro de previsão em função do passo de previsão para estes meses são apresentados nos gráficos das Figura 5.5 e 5.6.

Tabela 5.9: Razã	5 RMSE/MLT	(em %)	com h = 1	l para as va	azões de	Sobradinho
------------------	------------	--------	-----------	--------------	----------	------------

Método	jan	fev	mar	abr	mai	jun
$\overline{cv_m}$	29,40	40,68	53, 17	44, 78	48, 45	32, 24
Clássico	24,06	36,64	35, 31	35, 43	29, 21	10, 32
Stedinger (2001)	24,06	37,40	35, 31	35, 43	29, 21	10, 32
BIC	24,06	37,40	35, 31	35, 43	29,21	8,22
Método	jul	ago	set	out	nov	dez
cv_m	27,41	25,07	24,98	25,76	37, 30	34,54
Clássico	6,61	5, 19	8, 13	15, 56	28,89	27,44
Stedinger (2001)	6,61	5, 19	8, 13	15, 56	28,89	27,44
BIC	6,38	5, 19	8, 13	15, 56	28,89	27,44

Observa-se na Figura 5.5, que o erro de previsão para o mês de março tende rapidamente para o limite superior definido pelo desvio padrão das vazões deste mês. Nota-se, nestes gráficos, que estes erros praticamente coincidem para os modelos selecionados com os três critérios adotados. Na Figura 5.6, observa-se que a evolução do erro corre

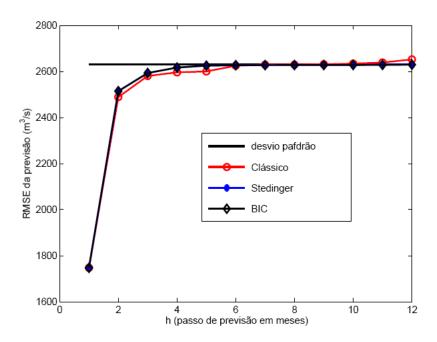


Figura 5.5: MSE da previsão da vazão para o mês de março para o posto de Sobradinho.

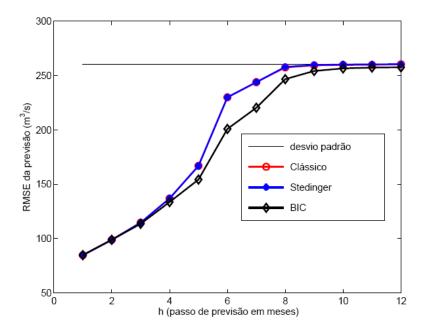


Figura 5.6: MSE da previsão da vazão para o mês de setembro para o posto de Sobradinho.

de forma mais lenta, destacando-se a capacidade preditiva dos modelos para horizontes maiores.

Quando se utiliza o critério BIC não se considera testes de hipóteses para os parâmetros dos modelos, portanto, com este critério, não se admite valores intermediários nulos. Pode-se observar que nos três estudos de caso considerados, em algumas situações, o critério BIC selecionou ordem superior ao método proposto em Stedinger (2001) e eventualmente ao método clássico, como é visto na Tabela 5.1 para Furnas, nos meses de fevereiro e maio, na Tabela 5.4 para Emborcação, no mês de fevereiro e na Tabela 5.7 para Sobradinho, nos meses de junho e julho.

Embora tais parâmetros possam ser considerados nulos, segundo algum teste de hipótese, uma diminuição na razão RMSE/MLT (h=1) é verificada nestes casos, como se pode notar na Tabela 5.3, para os meses de fevereiro e maio; Tabela 5.6, para o mês de fevereiro e Tabela 5.9, para os meses de junho e julho.

Estes resultados indicam que tais testes de hipóteses, os quais são fundamentados em propriedades assintóticas, são válidos somente para grandes amostras e podem ser imprecisos para o ajuste dos modelos $PAR(p_m)$, uma vez que, nestes ajustes o tamanho da amostra corresponde ao número de anos disponíveis no histórico de vazões. Portanto, tem-se amostras de tamanho reduzido do ponto de vista da teoria assintótica.

Recomenda-se que os critérios propostos por Stedinger (2001) ou o BIC sejam utilizados quando opta-se pelo critério da parcimônia. O uso do critério clássico, seleciona sempre ordens mais elevadas que produzem uma pequena diminuição na razão RMSE/MLT (h=1). No entanto, esta pequena redução no erro de previsão não justifica a complexidade proveniente do uso de modelos de ordens elevadas.

5.3 Abordagem bayesiana dos modelos PAR

Nos processos MCMC foram realizadas 50.000 iterações, descartadas 10000 para o aquecimento da cadeia e usou-se um salto de 10 iterações. Sendo assim, a amostra final totalizou 4000 iterações. Nestas análises, a convergência foi verificada usando-se o critério de Gelman e Rubin e o critério de Geweke.

Para avaliar o desempenho dos modelos Normal, Log-Normal e t-Student foram utilizados os erros de previsão, MSE, MAD e MAPE. Nesta avaliação, foram selecionados os últimos 3 anos (36 meses) do histórico de vazões, compreendendo o período de 2008 a 2010. Os modelos $PAR(p_m)$ foram ajustados utilizando o período de janeiro de 1931 a dezembro de 2007.

As ordens selecionadas considerando-se o critério BIC para cada modelo Normal, Log-Normal e t-Student e para os reservatórios de Água Vermelha e São Simão, são apresentadas, respectivamente, nas Tabelas 5.10 e 5.11.

Tabela 5.10: Ordem fornecida pelo critério BIC para Água Vermelha.

Mês	1	2	3	4	5	6	7	8	9	10	11	12
Normal	1	1	1	2	4	1	3	1	1	3	1	1
Log-Normal	1	1	1	2	3	1	3	1	1	2	1	1
$t ext{-Student}$	2	1	2	2	4	1	4	1	3	2	2	1

Tabela 5.11: Ordem fornecida pelo critério BIC para São Simão.

Mês	1	2	3	4	5	6	7	8	9	10	11	12
Normal	1	1	1	1	3	1	2	2	1	1	3	1
Log-Normal	1	1	1	1	3	1	2	1	1	1	1	1
$t ext{-Student}$	2	1	2	2	3	1	2	2	1	1	2	1

Pode-se notar que, para os meses secos (abril a outubro) a ordem dos modelos selecionados, em geral, foi maior que para os meses de cheia (novembro a março) indicando que durante o período seco a correlação com os meses anteriores é mais forte que para os meses úmidos. Observa-se, também, que o modelo Log-Normal foi o que apresentou o menor número de parâmetros e que a maior ordem selecionada foi 4.

As Tabelas 5.12, 5.13, 5.14, 5.15, 5.16 e 5.17 apresentam as estimativas dos parâmetros ϕ_m e τ_m para os modelos Normal, Log-Normal e t-Student e para os reservatórios de Água Vermelha e São Simão.

Tabela 5.12: Estimativas pontuais para o modelo Normal do reservatório Água

	7	<u>Vermelh</u>	ıa.		
Mês	$\hat{\phi}_{1,m}$	$\hat{\phi}_{2,m}$	$\hat{\phi}_{3,m}$	$\hat{\phi}_{4,m}$	$\hat{ au}_m$
Janeiro	0,61	-	-	-	1,58
Fevereiro	0,5	-	-	-	1,49
Março	0,62	-	-	-	1,67
Abril	$0,\!55$	0,31	-	-	2,82
Maio	$0,\!56$	0,18	0,17	0,12	4,86
Junho	0,84	-	-	-	$3,\!27$
Julho	0,7	0,11	0,2	-	9,6
Agosto	0,92	-	-	-	6,97
Setembro	0,82	-	-	-	3,13
Outubro	0,4	-0,08	$0,\!53$	-	2,58
Novembro	0,77	-	-	-	2,49
Dezembro	0,65	-	-	-	1,73

Tabela 5.13: Estimativas pontuais para o modelo Normal do reservatório São Simão.

Mês	$\hat{\phi}_{1,m}$	$\hat{\phi}_{2,m}$	$\hat{\phi}_{3,m}$	$\hat{\phi}_{4,m}$	$\hat{ au}_m$
Janeiro	0,51	-	-	-	1,33
Fevereiro	0,64	-	-	-	1,74
Março	$0,\!57$	-	-	-	$1,\!57$
Abril	0,74	-	-	-	2,34
Maio	0,72	0,07	$0,\!26$	_	8,16
$_{ m Junho}$	0,88	_	_	_	4,07
Julho	0,65	0,32	_	_	13,3
Agosto	1,18	-0,25	-	-	8,69
Setembro	0,88	_	_	_	4,72
Outubro	0,75	-	-	-	2,2
Novembro	0,94	0,03	-0,44	_	1,85
Dezembro	0,60	-	-	-	1,56

Tabela 5.14: Estimativas pontuais para o modelo Log-Normal do reservatório Água Vermelha.

	\	<u>/ermell</u>	ıa.		
Mês	$\hat{\phi}_{1,m}$	$\hat{\phi}_{2,m}$	$\hat{\phi}_{3,m}$	$\hat{\phi}_{4,m}$	$\hat{ au}_m$
Janeiro	0,64	-	-	-	1,69
Fevereiro	$0,\!54$	-	-	-	$1,\!47$
Março	0,68	-	-	-	1,92
Abril	0,6	$0,\!25$	-	-	2,99
Maio	$0,\!56$	0,2	0,2	-	4,69
Junho	$0,\!89$	-	-	-	$4,\!87$
Julho	0,74	0,03	$0,\!22$	-	9,64
Agosto	0,92	-	-	-	$6,\!65$
Setembro	0,8	-	-	-	2,83
Outubro	$0,\!35$	$0,\!38$	-	-	1,89
Novembro	0,71	-	-	-	1,97
Dezembro	$0,\!59$	-	-	-	1,51

 ${\bf Tabela~5.15:}$ Estimativas pontuais para o modelo Log-Normal do reservatório São

		Simão	Э.		
Mês	$\hat{\phi}_{1,m}$	$\hat{\phi}_{2,m}$	$\hat{\phi}_{3,m}$	$\hat{\phi}_{4,m}$	$\hat{ au}_m$
Janeiro	0,49	-	-	-	1,29
Fevereiro	0,65	-	-	-	1,74
Março	0,62	-	-	-	1,66
Abril	0,77	-	-	-	2,6
Maio	0,69	0,12	0,21	-	7,14
Junho	0,90	-	-	-	4,85
Julho	0,69	$0,\!28$	-	-	14,65
Agosto	0,94	-	-	-	9,45
Setembro	0,89	-	-	-	4,99
Outubro	0,7	-	-	-	1,94
Novembro	$0,\!58$	_	_	_	1,47
Dezembro	0,63	-	-	-	1,66

Tabela 5.16: Estimativas pontuais para o modelo t-Student do reservatório Água Vermelha.

	$\hat{\phi}_{1,m}$	$\hat{\phi}_{2,m}$	$\hat{\phi}_{3,m}$	$\hat{\phi}_{4,m}$	$\hat{ au}_m$
Janeiro	0,61	-	-	-	4,68
Fevereiro	$0,\!51$	-	-	-	4,02
Março	0,64	-0,30	-	-	5,04
Abril	$0,\!56$	$0,\!32$	-	-	8,54
Maio	$0,\!56$	0,18	$0,\!15$	0,12	14,6
Junho	0,83	-	-	-	9,79
Julho	0,7	0,07	0,16	0,05	28,91
Agosto	0,92	-	-	-	21,08
Setembro	$0,\!57$	0,09	$0,\!19$	-	10,34
Outubro	$0,\!46$	$0,\!34$	-	-	$7,\!25$
Novembro	0,67	$0,\!14$	-	-	7,66
Dezembro	0,65	-	-	-	$5,\!24$

Simão.							
Mês	$\hat{\phi}_{1,m}$	$\hat{\phi}_{2,m}$	$\hat{\phi}_{3,m}$	$\hat{\phi}_{4,m}$	$\hat{ au}_m$		
Janeiro	0,51	-	-	-	3,95		
Fevereiro	0,64	-	-	-	5,11		
Março	0,5	0,14	-	-	4,67		
Abril	0,67	0,14	-	-	$7,\!25$		
Maio	0,71	0,06	0,26	-	24,91		
Junho	0,86	-	-	_	12,03		
Julho	0,65	$0,\!34$	_	_	39,52		
Agosto	1,2	-0,28	-	_	26,4		
Setembro	0,88	_	-	_	14,26		
Outubro	0,74	-	-	_	6,67		
Novembro	0,87	-0,32	-	_	5,34		
Dezembro	0,6	_	-	_	4,68		

Tabela 5.17: Estimativas pontuais para o modelo t-Student do reservatório São

Vale ressaltar que todas as estimativas apresentadas foram significativas. Os métodos de análise de convergência de Gelman e Rubin e Geweke indicaram convergência das cadeias para todos os parâmetros.

O próximo passo se refere a seleção dos modelos. As Tabelas 5.18 e 5.19 apresentam os resultados para os métodos de seleção BIC e DIC para os modelos Normal, Log-Normal e t-Student e para os reservatórios de Água Vermelha e São Simão. Pode-se notar que para todos os critérios e reservatórios analisados seleciona-se o modelo Log-Normal.

Tabela 5.18: Métodos de seleção do modelo para Água Vermelha.

		,		
Método	Normal	Log-Normal	t-Stı	ident
BIC	-303,54	- 408,76	-32	3,67
DIC	-387,87	- 456,89	-398	8,76

Tabela 5.19: Métodos de seleção do modelo para São Simão.

Método	Normal	Log-Normal	t-Student
BIC	-244,39	- 290,55	-261,9
DIC	-269,12	- 305,7	-277,66

As Tabelas 5.20, 5.21, 5.22, 5.23, 5.24 e 5.25 mostram os erros de previsão, RMSE, MAD e MAPE, da aplicação dos modelos Normal, Log-Normal e *t*-Student para os reservatórios de Água Vermelha e São Simão, respectivamente.

 ${\bf Tabela~5.20:~Erros~de~previsão~para~o~modelo~Normal~de~\'Agua~Vermelha}.$

	1		
Mês	RMSE	MAD	MAPE
Janeiro	782,33	757,12	23,6
Fevereiro	807,92	732,79	16,4
Março	$429,\!63$	277,16	6,85
Abril	808,54	$667,\!25$	19,29
Maio	379,78	348,42	16,59
Junho	$234,\!59$	210,87	12,34
Julho	211,91	$159,\!28$	10,7
Agosto	219,84	203,75	$17,\!56$
Setembro	$515,\!35$	388,1	$27,\!54$
Outubro	$419,\!54$	$282,\!57$	16,23
Novembro	160,83	137,29	8,01
Dezembro	783,98	500,96	13,72

Tabela 5.21: Erros de previsão para o modelo Log-Normal de Água Vermelha.

Mês	RMSE	MAD	MAPE
Janeiro	906,02	844,65	27,44
Fevereiro	766,01	675,78	15,09
Março	382,09	$247,\!52$	6,12
Abril	784,72	$646,\!22$	18,67
Maio	354,96	$330,\!17$	15,79
Junho	214,75	$195,\!16$	11,45
Julho	$205,\!51$	148,71	9,91
Agosto	214,38	196,12	16,91
Setembro	$515,\!69$	$396,\!37$	$28,\!51$
Outubro	418,94	292,38	17,16
Novembro	$158,\!58$	126,72	$7,\!33$
Dezembro	782,87	$509,\!26$	14,08

Tabela 5.22: Erros de previsão para o modelo t-Student de Água Vermelha.

Mês	RMSE	MAD	\overline{MAPE}
Janeiro	884,77	875,14	25,5
Fevereiro	988,49	812,24	$17,\!37$
Março	$560,\!6$	$461,\!66$	11,88
Abril	902,2	764,98	22,36
Maio	$415,\!57$	363,26	16,9
Junho	$272,\!81$	223,74	12,8
Julho	239,44	187,79	12,74
Agosto	233,81	210,36	17,74
Setembro	538,81	$373,\!52$	24,58
Outubro	469,03	331,19	19,43
Novembro	235,39	217,07	12,81
Dezembro	872,38	615,6	17,65

Tabela 5.23: Erros de previsão para o modelo Normal de São Simão.

Mês	RMSE	MAD	MAPE
Janeiro	419,02	394,24	11,33
Fevereiro	1146,59	$945,\!96$	23,05
Março	1138,95	1011,46	22,9
Abril	$1007,\!87$	$752,\!66$	17,12
Maio	$532,\!14$	$451,\!21$	18,88
Junho	334,61	284,01	14,73
Julho	246,77	219,98	$14,\!29$
Agosto	179,24	155,94	12,48
Setembro	319,48	294,18	26,97
Outubro	$222,\!12$	$162,\!27$	$11,\!26$
Novembro	$439,\!54$	438,93	$23,\!55$
Dezembro	782,73	574,39	15

Ties de previ	sao para c	, modelo	208 110111
Mês	RMSE	MAD	MAPE
Janeiro	649,18	546,32	16,29
Fevereiro	994,81	826,43	20,87
Março	1054,19	954,05	21,92
Abril	954,11	719,85	16,43
Maio	$493,\!54$	$425,\!95$	18
Junho	$308,\!23$	267,09	13,93
Julho	227,69	206,81	13,49
Agosto	163,5	$145,\!32$	11,68
Setembro	314,36	284,03	26,07
Outubro	$219,\!55$	152,9	10,48
Novembro	$443,\!81$	443,49	23,89
Dezembro	784,19	579,5	15,19

Tabela 5.24: Erros de previsão para o modelo Log-Normal de São Simão.

Tabela 5.25: Erros de previsão para o modelo t-Student de São Simão.

Mês	RMSE	MAD	MAPE
Janeiro	413,4	357,53	9,85
Fevereiro	1193,2	$1035,\!53$	$23,\!57$
Março	1178	$910,\!24$	19,17
Abril	1120,94	841,19	19,1
Maio	$559,\!38$	472,07	$19,\!26$
Junho	367,09	300,77	$15,\!28$
Julho	272,89	232,05	14,79
Agosto	$204,\!45$	167,22	13,13
Setembro	340,14	307,97	27,12
Outubro	259,24	187,75	12,96
Novembro	484,28	470,21	23,98
Dezembro	887,93	537,97	12,77

Para ambos os reservatórios, o modelo Log-Normal, na maioria dos meses, apresentou os menores erros de previsão em relação aos modelos Normal e t-Student, reforçando o resultado apresentado no processo de seleção de modelos. Pode-se notar também que os modelos Normal e t-Student apresentaram resultados que ora demonstraram superioridade ora inferioridade de um modelo em relação ao outro. As Figuras 5.7 e 5.8 mostram o gráfico das previsões para o modelo Log-Normal para os reservatórios de Água Vermelha e São Simão, respectivamente.

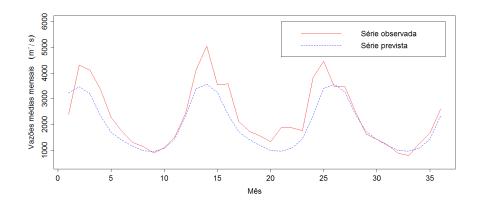


Figura 5.7: Previsão do modelo Log-Normal de Água Vermelha (2008-2010).

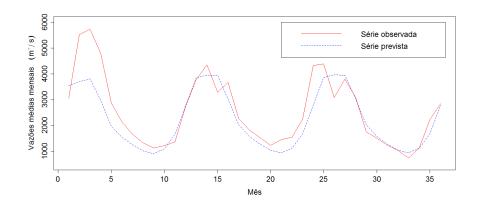


Figura 5.8: Previsão do modelo Log-Normal de São Simão (2008-2010).

5.4 Geração de séries sintéticas

O modelo PAR multivariado gerou um total de 2000 anos de séries sintéticas de vazões médias mensais, para cada reservatório. Após este procedimento, estas séries foram comparadas com os 80 anos (1931-2010) de séries históricas e com os 2000 anos das séries sintéticas geradas pelo modelo Gevazp.

O modelo PAR multivariado para geração de vazões médias mensais foi primeiramente avaliado quanto à capacidade em reproduzir os momentos amostrais, como a média, o desvio padrão, o coeficiente de assimetria e o coeficente de curtose. Posteriormente, foi realizada uma análise referente à capacidade de reproduzir a correlação espacial entre dois reservatórios. Nesta análise, foi considerada a correlação mensal

entre os dados históricos, os dados sintéticos gerados pelo modelo PAR multivariado e os dados sintéticos gerados pelo modelo Gevazp.

Uma avaliação mais criteriosa também é analisada referente ao cálculo da divergência de Kullback-Leibler. Nesta análise, procura-se verificar o quanto a distribuição da série sintética se distancia da série histórica e, avalia, se essa distância é aceitável ou não.

As Figuras 5.9, 5.10, 5.11 e 5.12 apresentam as estatísticas média, desvio padrão, coeficiente de assimetria e coeficiente de curtose, em cada mês, para o reservatório de Itaipu.

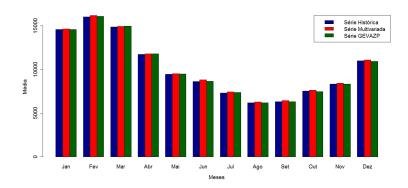


Figura 5.9: Média mensal para o reservatório de Itaipu.

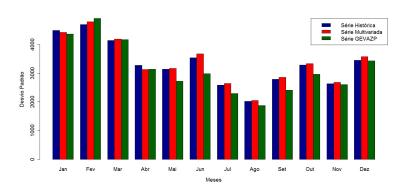


Figura 5.10: Desvio padrão mensal para o reservatório de Itaipu.

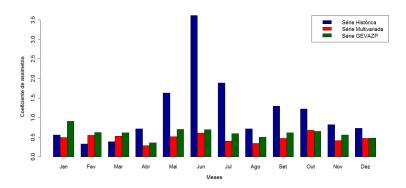


Figura 5.11: Coeficiente de assimetria mensal para o reservatório de Itaipu.

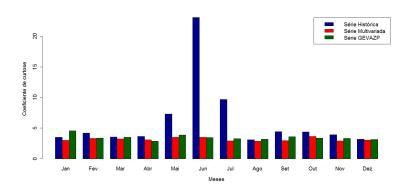


Figura 5.12: Coeficiente de curtose mensal para o reservatório de Itaipu.

O procedimento de geração usando o modelo PAR multivariado preserva bem os dois primeiros momentos, ou seja, a média e o desvio padrão. Em relação aos outros momentos, o que se observa é que, para alguns meses, essa capacidade de reprodução não é mantida. Em geral, a grande dificuldade encontrada foi a reprodução do coeficiente de assimetria e coeficiente de curtose.

As Figuras 5.13, 5.14, 5.15 e 5.16 exibem a estrutura de correlação mensal para dados históricos, dados sintéticos gerados pelo modelo PAR multivariado e dados sintéticos gerados pelo modelo GEVAZP entre os reservatórios de Itaipu e Itutinga, Itaipu e Xingó, Corumbá IV e Espora e Corumbá I e Corumbá IV, respectivamente.

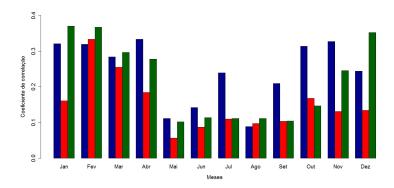


Figura 5.13: Correlação espacial entre os reservatórios de Itaipu e Xingó.

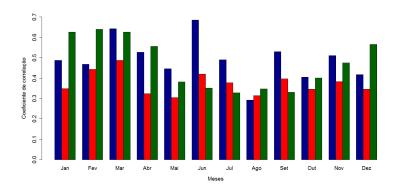


Figura 5.14: Correlação espacial entre os reservatórios de Itaipu e Itutinga.

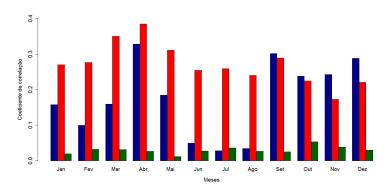


Figura 5.15: Correlação espacial entre os reservatórios de Corumbá IV e Espora.

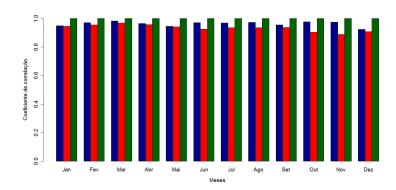


Figura 5.16: Correlação espacial entre os reservatórios de Corumbá I e Corumbá IV.

Nota-se que, por exemplo, entre os reservatórios de Itaipu e Xingó, a correlação nos meses de julho e setembro foram bem próximas paras as séries sintéticas geradas pelo modelo PAR multivariado e GEVAZP. Observa-se também que, para alguns meses, as séries sintéticas geradas pelo modelo PAR multivariado apresentam valores mais próximos ao dados históricos. Para o caso de Corumbá IV e Espora, os resultados possibilitam afirmar que, para alguns meses, as séries sintéticas geradas pelo modelo PAR multivariado apresentaram valores superiores as séries sintéticas geradas pelo modelo GEVAZP e, que o inverso, também ocorre, para outros meses. Já para o caso de Corumbá I e Corumbá IV, pode-se atestar que os dois modelos conseguiram reproduzir bem a correlação espacial.

A Tabela 5.26 apresenta os valores da divergência de Kullback-Leibler (KL) para todos os reservatórios estudados. Pode-se afirmar que, para a maioria dos reservatórios, a série gerada foi validada. Apenas os reservatórios de Guarapiranga, Lajes e Ponte Pedra não tiveram suas séries validadas.

A Figura 5.17 apresenta a trajetória para os 10 últimos anos da série histórica de Itaipu, com os 10 primeiros anos para a série sintética gerada pelo modelo PAR multivariado e com os 10 primeiros anos para a série sintética gerada pelo modelo GEVAZP. Observa-se que o comportamento sazonal da série histórica é mantido pelos modelos geradores de séries sintéticas.

Tabela 5.26: Divergencia de Kullback-Leibler (KL).							
Reservatório	KL	Reservatório	KL				
Água Vermelha	0,63	Jauru	0,91				
A.S.Oliveira	$0,\!54$	Lajes	0,98				
Cachoeira Dourada	$0,\!58$	Manso	0,62				
Candonga	0,61	Mascarenhas	$0,\!55$				
Canoas I	$0,\!68$	Navanhandava	0,64				
Capim Branco II	$0,\!58$	Porto Colômbia	0,63				
Corumbá I	$0,\!55$	Porto Estrela	0,61				
Corumbá IV	0,6	Picada	0,81				
Emborcação	$0,\!57$	Ponte Pedra	0,98				
Espora	0,88	Porto Primavera	0,81				
Foz Rio Claro	0,6	Retiro	0,6				
Fundão	0,85	Rosal	$0,\!66$				
Funil Grande	0,61	Sá Carvalho	0,62				
G.P. Souza	0,89	Salto Rio Verdinho	0,72				
Guarapiranga	0,96	Santa Branca	0,64				
Ilha Pombos	0,61	São Simão	0,62				
Itaipu	0,78	Sobragi	$0,\!67$				
Itutinga	0,61	Três Marias	$0,\!55$				
Jaguari	0,93	Xingó	0,69				

Tabela 5.26: Divergência de Kullback-Leibler (KL)

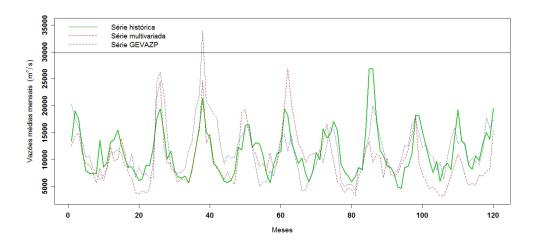


Figura 5.17: Séries históricas e sintéticas para Itaipu.

O modelo PAR(1) univariado usando a distribuição Log-Normal generalizada gerou um total de 2000 anos de séries sintéticas de vazões médias mensais para o reservatório de Furnas. Neste procedimento, foram utilizadas séries históricas de Furnas correspondentes ao período de janeiro de 1931 a dezembro de 2010.

Vale ressaltar que, no processo de estimação dos parâmetros via abordagem bayesiana, foram simuladas um total de 20.000 iterações, sendo descartadas as 10.000 primeiras iterações para o aquecimento da cadeia e foi considerado um salto de tamanho 10 nas iterações restantes. Por fim, a amostra final totalizou um tamanho de 1000 iterações, que foram utilizadas nas análises posteriores. Destaca-se também que a cadeia convergiu, para todos os parâmetros, por meio das análises dos critérios de Gelman e Rubin e Geweke.

A Tabela 5.27 apresenta as estimativas pontuais para os parâmetros do modelo PAR(1), em que todas estas estimativas foram significativas.

Tabela 3.27. Estimativas polituais para os parametros de um modelo i Art(1).												
Parâmetros	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
α	3,13	1,69	3,51	1,99	2,56	1,49	1,52	1,77	2,44	3,41	1,97	4,5
β	0,6	0,76	0,5	0,67	$0,\!58$	0,74	0,73	0,68	$0,\!59$	0,46	0,73	$0,\!39$
σ	$0,\!48$	$0,\!51$	0,4	$0,\!27$	$0,\!23$	0,19	$0,\!15$	0,18	$0,\!37$	0,39	0,4	$0,\!41$
s	1,9	1,87	1,85	1,9	1,88	1,88	1,92	1,88	1,89	1,88	1,87	1,88

Tabela 5.27: Estimativas pontuais para os parâmetros de um modelo PAR(1)

Novamente, as estatísticas média, desvio padrão, coeficiente de assimetria e coeficente de curtose são utilizadas na avaliação da capacidade de reprodução dos momentos do modelo de geração univariada usando a distribuição Log-Normal generalizada. As Figuras 5.18 e 5.19 mostram as estatísticas média e desvio padrão desse modelo para o reservatório de Furnas. Como pode-se observar, este modelo também consegue reproduzir satisfatoriamente estes momentos.

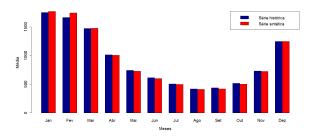


Figura 5.18: Média Mensal para séries histórica e sintética de Furnas.



Figura 5.19: Desvio padrão mensal para séries histórica e sintética de Furnas.

As Figuras 5.20 e 5.21 mostram os coeficientes de assimetria e curtose para o reservatório de Furnas. Mais uma vez nota-se a insastifatória capacidade de reproduzir estes momentos.

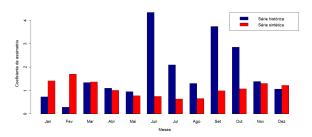


Figura 5.20: Coeficiente de assimetria para séries histórica e sintética de Furnas.

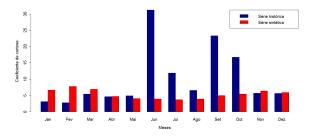


Figura 5.21: Coeficiente de curtose para séries histórica e sintética de Furnas.

As figuras aqui apresentadas mostram a necessidade de considerar uma distribuição marginal adequada para a geração de séries sintéticas que sejam estatisticamente semelhantes às séries históricas. Embora a falta de conhecimento sobre a distribuição teórica ou uma hipótese inadequada aparentemente não afetem significativamente as estimativas da média e do desvio padrão, a reprodução do coeficiente de assimetria e do coeficiente de curtose não são satisfatórias. Este é um ponto que merece destaque em pesquisas futuras.

O valor da calibração da divergência de Kullback-Leibler foi 0,89, indicando que o modelo Log-Normal generalizado desenvolvido neste estudo pode ser considerado como uma potencial alternativa para geração univariada de séries sintéticas. A Figura 5.22 mostra os gráficos da série observada e da série sintética para o reservatório de Furnas. Como observa-se, o comportamento sazonal é mais deslocado pela série sintética.

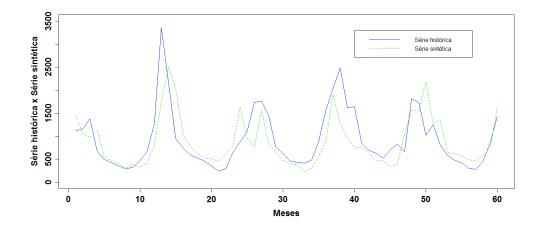


Figura 5.22: Série histórica e sintética de Furnas.

Capítulo

Conclusões e Propostas Futuras

Em Hidrologia, os modelos autorregressivos periódicos apresentam papel importante nos procedimentos de previsão e geração de vazões médias mensais. Antes destas análises e não menos importante, deve-se realizar a estimação dos parâmetros, que neste caso, foi realizado usando as abordagens clássica e bayesiana.

Em um primeiro momento, foi utilizada a estimação clássica usando as equações de Yule-Walker e, posteriormente, foram calculados os erros quadráticos médios das previsões de vazões em função do passo de previsão para modelos $PAR(p_m)$. Foram considerados três métodos de identificação da ordem dos modelos, em que se observa que o uso do método clássico de análise da função de autocorrelação parcial leva à escolha de ordens elevadas para alguns meses. No entanto, os modelos, para estes meses, apresentam estimativas nulas para alguns parâmetros intermediários. Considerar parâmetros nulos entre o primeiro e o último parâmetro de um modelo $PAR(p_m)$ ressalta uma incoerência entre o modelo e o fenômeno natural. Portanto, o método clássico de análise da função de autocorrelação parcial não é adequado para identificar a ordem dos modelos $PAR(p_m)$, sendo que os outros dois métodos considerados identificam as ordens de forma mais coerente.

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

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

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

Em um segundo momento, foi utilizada a estimação bayesiana e, após, os parâmetros estimados foram utilizados nos procedimentos de previsão. Assim, este trabalho abordou a aplicação da inferência bayesiana e dos métodos de Monte Carlo via Cadeias de Markov para a previsão de vazões médias mensais. Os modelos autorregressivos periódicos foram ajustados para os reservatórios de Emborcação, Furnas e Sobradinho e, posteriormente, foram estudadas as propriedades geradas pelas previsões. Em particular, os algoritmos MCMC têm desempenhado um papel essencial quando as densidades a posteriori não apresentam uma forma padrão e, assim, estes são utilizados para a obtenção de amostras dos parâmetros.

O procedimento de previsão é de grande importância para os estudos de planejamento e operação de um sistema hidrelétrico e são os principais resultados práticos deste trabalho. Em particular, os estudos de previsão podem ser usados juntos como um critério final para selecionar um método de estimação para o parâmetro. Podese concluir que a metodologia obteve resultados significativos na previsão de vazões afluentes naturais médias mensais h passos a frente para os reservatórios estudados. Conclui-se, também, que os modelos apresentaram erros de previsão muito próximos apesar do modelo Log-Normal apresentar menor MAPE e MSE. Recomenda-se o uso do modelo Log-Normal por terem ordens bastante reduzidas o que facilita a utilização do modelo, no contexto de otimização do planejamento e operação de sistemas hidroelétricos.

Vale ressaltar que as análises bayesianas juntamente com algoritmos MCMC deram mais flexibilidade ao processo de modelagem de séries temporais periódicas e durante todas as etapas, tais como a estimativa de parâmetros, seleção de modelos e previsão.

E, por último, este trabalho apresentou os procedimentos de geração de séries sintéticas usando o modelo Log-Normal com três parâmetros multivariado, modelo Log-Normal generalizado e avaliação das séries geradas. O modelo multivariado foi aplicado a um conjunto de usinas do Sistema Elétrico Brasileiro. Os resultados para esses estudos, quando comparados com o modelo GEVAZP, amplamente utilizado pelo SEB, indicaram que esse modelo representa uma potencial abordagem para a geração de séries sintéticas.

Vale destacar que, em geral, os valores médios obtidos das séries sintéticas quando comparados aos valores reais das séries de vazões foram similares para todos os modelos e todos os meses. Por outro lado, os modelos apresentaram dificuldades em replicar os coeficientes de assimetria e curtose para alguns meses das séries de vazões. Os resultados forneceram evidências que apontam que os modelos utilizados são uma alternativa para a geração de séries sintéticas de vazões mensais e que reproduzem as principais características estatísticas das séries observadas. Em relação ao modelo Log-Normal generalizado desenvolvido neste estudo pode-se concluir que este apresenta uma alternativa potencial para os procedimentos de geração univariada de séries sintéticas de vazões médias mensais.

Como sugestões para trabalhos futuros, pode-se destacar: a utilização dos modelos autorregressivos e de médias móveis periódicos (PARMA) em todos os procedimentos de estimação, geração e previsão; comparação da abordagem proposta para geração e previsão com outras técnicas comuns, tais como redes neurais artificiais e os algoritmos neuro-fuzzy; em relação à abordagem bayesiana, pode-se ressaltar a utilização de diferentes modelos e distribuições a priori; Desenvolvimento de novos critérios para validação de séries sintéticas; Utilização de outros modelos para geração, além do modelo Log-Normal com três parâmetros, que captem a assimetria dos dados de vazões

médias mensais; Adoção de distribuições a~priori informativas na estimação bayesiana do modelo Log-Normal generalizado; Ampliação da abordagem do modelo Log-Normal generalizado, utilizando um modelo PAR (p_m) e desenvolvendo um modelo multivariado.

Referências

- Ahmed, J. & Sarma, A. (2007). Artificial neural network model for synthetic streamflow generation. Water resources management, **21**(6), 1015–1029.
- Anderson, P. & Vecchia, A. (1993). Asymptotic results for periodic autoregressive moving-average process. *Journal of Time Series Analysis*, **14**(1), 1–18.
- Anderson, P., Meerschaert, M. & Vecchia, A. (1999). Innovations algorithm for periodically stationary time series. *Stochastic processes and their applications*, **83**(1), 149–169.
- Ballini, R., Soares, S. & Andrade, M. (2003). Previsão de vazões médias mensais usando redes neurais nebulosas. Sba: Controle & Automação Sociedade Brasileira de Automatica, 14(3), 680–693.
- Barreto, G. & Andrade, M. (2004). Robust bayesian approach for ar (p) models applied to stream flow forecasting. *Journal of Applied Statistical Science*, **12**, 269–292.
- Bentarzi, M., Guerbyenne, H. & Hemis, R. (2008). Predictive density order selection of periodic ar models. *Communications in Statistics: Simulation and Computation*(R), **37**(6), 1167–1182.
- Box, G. & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252.
- Box, G. & Tiao, G. (1992). Bayesian inference in statistical analysis. Wiley Online Library.

Box, G., Jenkins, G. & Reinsel, G. (2008). Time Series Analysis: Forecasting and Control. John Wiley, Hoboken, NJ.

- Camacho, F., McLeod, A. & Hipel, K. (1987). Multivariate contemporaneous arma model with hydrological applications. *Stochastic Hydrology and Hydraulics*, **2**(1), 141–154.
- Carlin, B. & Louis, T. (1997). Bayes and empirical bayes methods for data analysis. Statistics and Computing, 7(2), 153–154.
- Casella, G. & George, E. (1992). Explaining the gibbs sampler. *American Statistician*, pages 167–174.
- CEPEL (2002). Manual de Referência do Modelo de Geração de Séries Sintéticas de Energias e Vazões, o Modelo GEVAZP. Centro de Pesquisa de Energia Elétrica.
- Gamerman, D. & Lopes, H. (2006). Markov chain Monte Carlo: stochastic simulation for Bayesian inference, volume 68. Chapman & Hall/CRC.
- Gelfand, A. & Dey, D. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 501–514.
- Gelfand, A. & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, pages 398–409.
- Gelfand, A., Hills, S., Racine-Poon, A. & Smith, A. (1990). Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, pages 972–985.
- Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, **7**(4), 457–472.
- Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004). *Bayesian data analysis*. CRC press.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. *Bayesian Statistics*, pages 625–631.
- Gladysev, E. (1961). Periodically random sequences. Soviet Mathematics, 2, 385–388.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, **57**(1), 97–109.

Herández, F. & Usuga, O. (2011). Bayesian analysis for the generalized lognormal distribution applied to failure time analysis. *Revista Colombiana de Estadística*, **34**(1), 95–109.

- Hipel, K. & McLeod, A. (1994). Time series modelling of water resources and environmental systems. Elsevier Science B.V., Amsterdam.
- Hurd, H. & Gerr, N. (1991). Graphical methods for determining the presence of periodic correlation. *Journal of Time Series Analysis*, **12**(4), 337–350.
- Jeffreys, H. (1961). Theory of probability. Clarendon. Oxford. GB.
- Jones, R. & Brelsford, W. (1967). Time series with periodic structure. *Biometrika*, **54**(3-4), 403–408.
- Kim, B., Kim, H. & Seoh, B. (2004). Streamflow simulation and skewness preservation based on the bootstrapped stochastic models. *Stochastic Environmental Research and Risk Assessment*, **18**(6), 386–400.
- Kullback, S. & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**(1), 79–86.
- Lohani, A., Kumar, R. & Singh, R. (2012). Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology*, pages 129–148.
- Lund, R. & Basawa, I. (2000). Recursive prediction and likelihood evaluation for periodic arma models. *Journal of Time Series Analysis*, **21**(1), 75–93.
- Lund, R., Shao, Q. & Basawa, I. (2006). Parsimonious periodic time series modeling. Australian & New Zealand Journal of Statistics, 48(1), 33–47.
- Machiwal, D. & Jha, M. (2009). Time series analysis of hydrologic data for water resources planning and management: A review. *Journal of Hydrology and Hydrome-chanics*, **54**(3), 237–257.
- Martín, J. & Pérez, C. (2009). Bayesian analysis of a generalized lognormal distribution. Computational Statistics & Data Analysis, 53(4), 1377–1387.
- McLeod, A. (1994). Diagnostic checking of periodic autoregression models with application. *Journal of Time Series Analysis*, **15**(2), 221–233.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, **21**, 1087.

- Mondal, M. & Wasimi, S. (2007). Disaggregation model for synthetic stream-flow generation.
- Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, **32**(7), 685–694.
- Noakes, D., McLeod, A. & Hipel, K. (1985). Forecasting monthly riverflow time series. *International Journal of Forecasting*, **1**(2), 179–190.
- Oliveira, F.L.C.; Souza, R. C. (2011). A new approach to identify the structural order of par (p) models. *Pesquisa Operacional (Impresso)*, **31**, 487–498.
- Pagano, M. (1978). On periodic and multiple autoregressions. *The Annals of Statistics*, pages 1310–1317.
- Parzen, E. & Pagano, M. (1979). An approach to modeling seasonally stationary time series. *Journal of Econometrics*, **9**, 137–153.
- Paulino, C., Turkman, M. & Murteira, B. (2003). *Estatística bayesiana*. Fundação Calouste Gulbenkian.
- Pereira, M., Oliveira, G., Costa, C. & Kelman, J. (1984). Stochastic streamflow models for hydroelectric systems. Water Resources Research, 20(3), 379–390.
- R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rasmussen, P., Salas, J., Fagherazzi, L., Rassam, J. & Bobée, B. (1996). Estimation and validation of contemporaneous parma models for streamflow simulation. *Water resources research*, **32**(10), 3151–3160.
- Sakai, H. (1982). Circular lattice filtering using pagano's method. Acoustics, Speech and Signal Processing, IEEE Transactions on, **30**(2), 279–287.
- Salas, J. & Obeysekera, J. (1988). Multivariate periodic arma (1, 1) processes. Water Resources Research, 24(8), 1237–1246.

Salas, J. & Pegram, G. (1977). A seasonal multivariate multilag autoregressive model in hydrology. *Modeling Hydrologic Processes*, pages 45–60.

- Salas, J., Boes, D. & Smith, R. (1982). Estimation of arma models with seasonal parameters. Water Resources Research, 18(4), 1006–1010.
- Sarnaglia, A., Reisen, V. & Lévy-Leduc, C. (2010). Robust estimation of periodic autoregressive processes in the presence of additive outliers. *Journal of multivariate analysis*, **101**(9), 2168–2183.
- Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, **6**(2), 461–464.
- Shao, Q. & Ni, P. (2004). Least-squares estimation and anova for periodic autoregressive time series. Statistics & probability letters, **69**(3), 287–297.
- Spiegelhalter, D., Best, N., Carlin, B. & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* (Statistical Methodology), **64**(4), 583–639.
- Stedinger, J. (2001). Report on the evaluation of CEPEL PAR models. Cornell University, School of Civil and Environmental Engineering, Ithaca, New York.
- Stedinger, J. & Taylor, M. (1982). Synthetic streamflow generation: 1. model verification and validation. *Water Resources Research*, **18**(4), 909–918.
- Tesfaye, Y., Meerschaert, M. & Anderson, P. (2006). Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resources Research*, **42**(1), W01419.
- Tesfaye, Y., Anderson, P. & Meerschaert, M. (2011). Asymptotic results for fourier-parma time series. *Journal of Time Series Analysis*, **32**(2), 157–174.
- Thomas, H. & Fiering, M. (1962). Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation in Design of Water Resources Systems. *Cambridge*, Mass, 1, 459–493.
- Thompstone, R., Hipel, K. & McLeod, A. (1985). Forecasting quarter-monthly riverflow. *JAWRA Journal of the American Water Resources Association*, **21**(5), 731–741.
- Tiao, G. & Grupe, M. (1980). Hidden periodic autoregressiye-moving average models in time series data. *Biometrika*, **67**(2), 365–373.

Troutman, B. (1979). Some results in periodic autoregression. *Biometrika*, **66**(2), 219–228.

- Ula, T. (1993). Forecasting of multivariate periodic autoregressive moving-average process. *Journal of Time Series Analysis*, **14**(6), 645–657.
- Ula, T. & Smadi, A. (2003). Identification of periodic moving-average models. *Communications in Statistics-Theory and Methods*, **32**(12), 2465–2475.
- Valencia, D. & Schaake, J. (1973). Disaggregation processes in stochastic hydrology. Water Resour. Res, 9(3), 211–219.
- Vecchia, A. (1985a). Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics*, **27**, 375–384.
- Vecchia, A. (1985b). Periodic autoregressive-moving average (parma) modeling with applications to water resources. *JAWRA Journal of the American Water Resources Association*, **21**(5), 721–730.
- Vecchia, A. & Ballerini, R. (1991). Testing for periodic autocorrelations in seasonal time series data. *Biometrika*, **78**(1), 53–63.
- Vieira, A., Pereira, B. & Sales, P. (1998). Estimação conjunta dos parâmetros de um modelo estocástico multivariado contemporâneo periódico auto-regressivo par(p). Revista Brasileira de Recursos Hídricos, 3(4), 5–17.
- Wang, Q. & Robertson, D. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, **47**(2), W02546.
- Wang, Q., Robertson, D. & Chiew, F. (2009). A bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, **45**(5), W05407.
- Wang, W. (2006). Stochasticity, nonlinearity and forecasting of streamflow processes.

 Delft University Press.