

**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL**

**MODELOS DE MARKOV COM ESTADOS OCULTOS NA  
MODELAGEM DE SÉRIES DE VAZÕES ANUAIS**

**LUIS CARLOS HERNÁNDEZ HERNÁNDEZ**

**ORIENTADOR: DIRCEU SILVEIRA REIS JUNIOR**

**DISSERTAÇÃO DE MESTRADO EM TECNOLOGIA AMBIENTAL E  
RECURSOS HÍDRICOS**

**BRASÍLIA/DF: NOVEMBRO – 2013**

**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL**

**MODELOS DE MARKOV COM ESTADOS OCULTOS NA  
MODELAGEM DE SÉRIES DE VAZÕES ANUAIS**

**LUIS CARLOS HERNÁNDEZ HERNÁNDEZ**

**DISSERTAÇÃO SUBMETIDA AO DEPARTAMENTO DE  
ENGENHARIA CIVIL E AMBIENTAL DA FACULDADE DE  
TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA COMO PARTE  
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU  
DE MESTRE EM TECNOLOGIA AMBIENTAL E RECURSOS  
HÍDRICOS.**

**APROVADA POR:**

---

**Prof. Dirceu Silveira Reis Junior, PhD (ENC-UnB)  
(Orientador)**

---

**Prof. Carlos Henrique Ribeiro Lima, PhD (ENC-UnB)  
(Examinador Interno)**

---

**Prof. Eduardo Sávio Passos Rodrigues Martins, PhD (UFC-FUNCEME)  
(Examinador Externo)**

**BRASÍLIA/DF, 29 DE NOVEMBRO DE 2013**

## **FICHA CATALOGRÁFICA**

HERNÁNDEZ, HERNÁNDEZ LUIS CARLOS

Modelos de Markov com Estados Ocultos na Modelagem de Séries de Vazões Anuais

[Distrito Federal] 2013.

xxiii, 145p., 210 x 297 mm (ENC/FT/UnB, Mestre, Tecnologia Ambiental e Recursos Hídricos, 2013).

Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Civil e Ambiental.

1. MODELOS DE MARKOV OCULTO

2. PERSISTÊNCIA HIDROLÓGICA

3. ÍNDICES CLIMÁTICOS

4. VAZÕES ANUAIS

I. ENC/FT/UnB

II. Título (série)

## **REFERÊNCIA BIBLIOGRÁFICA**

HERNÁNDEZ, L. C. H. (2013). Modelos de Markov com Estados Ocultos na Modelagem de Séries de Vazões Anuais. Dissertação de Mestrado em Tecnologia Ambiental e Recursos Hídricos, Publicação PTARH.DM-155/2013, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF, 145p.

## **CESSÃO DE DIREITOS**

AUTOR: Luis Carlos Hernández Hernández.

TÍTULO: Modelos de Markov com Estados Ocultos na Modelagem de Séries de Vazões Anuais

GRAU: Mestre

ANO: 2013

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

---

Luis Carlos Hernández Hernández

flecks85@gmail.com

*À minha mãe*  
*À Linita*  
*À Guesa*

***“A tu escuela llegué sin entender porque llegaba, en tus salones encuentro mil caminos y encrucijadas y aprendo mucho y no aprendo nada... [....]”***

*Maestra Vida - Rubén Blades*

## AGRADECIMENTOS

Agradeço à vida e principalmente a Deus, por me permitir desfrutar tantas experiências maravilhosas, ter me trazido ao lugar em que estou e me ajudar a cada dia.

À minha mãe, pelo apoio, confiança e amor incondicional.

Ao meu pai, pelas conversas, exemplo, apoio constante e por ser fonte de inspiração para seguir este caminho.

Ao meu orientador, Dirceu Reis, pela paciência e conhecimento transmitido, pela ajuda constante e pela confiança em meu trabalho.

Ao professor e quase co-orientador, Carlos Lima, por suas indicações, sugestões e apoio. Sem dúvida, parte fundamental no desenvolvimento deste trabalho.

Ao Professor Eduardo Martins, pelas sugestões e comentários como membro da minha banca examinadora.

À Linita, minha irmã, por ter sido minha primeira professora na vida, pelo amor e por todos os conselhos, fundamentais para enfrentar essa experiência.

À Guesa, pelo amor, constante apoio e por sempre me proporcionar momentos inesquecíveis.

Aos meus irmãos Diana, Esteban e Pablo, pelo amor e apoio, sempre me motivando a continuar meu caminho.

Aos tios Jorge, Jaime, Pedro e Bairon, meus segundos pais e à tia Laura, minha segunda mãe, pelo apoio, carinho e por sempre me aconselharem.

Às Reinas, Valen, Daniela e Lamby, pelo carinho e risadas.

Ao que foi, durante esse tempo fora de casa, meu primo, confidente e quase irmão Elkin. Não esquecerei seu carinho e apoio. À Mamita Rosalía e à tia Rosaura, pelo carinho, risadas e comidinhas.

Aos professores do PTARH, Segio Koide, Cristina Brandão, Marco Antônio, Conceição Alves, Lenora Gomes, Ariuska Amorim, Ricardo Minoti, João Bosco, Nestor Campana, Oscar Cordeiro Netto e Yovanka Perez, por todos os conhecimentos, paciência, apoio e acolhida no Programa e no Brasil.

Às Chaparro, Doña Martica, Johanna e Yuliana, pelo carinho e apoio. Vocês são, sem dúvida, minha família no Brasil.

À minha colega Carol Alvarenga, pela amizade e apoio na elaboração desta dissertação.

À Renata Araújo, pela amizade e a ajuda nesse tempo no Brasil.

À minha companheira de estudo e amiga Diana Monsalve, pela ajuda em todos os momentos.

Ao meu amigo de longe Lucho pela grande amizade, risadas e momentos. Obrigado por ser meu irmão.

Ao meu amigo Diego Ocampo, pela sua paciência, conselhos e apoio nessa experiência fora de casa.

Aos meus amigos de mestrado Lucas Achaval, Adriane Dias, Leonardo Piau, Olga Caminha, Maria Elisa Elite, pela amizade, apoio e recepção no Brasil.

Aos colegas de mestrado Patricia Bermond, Rodrigo Otsuki, Arthur Tavares, Luiz Gustavo, Felipe Sampaio, Fabiano Silva, Reuel Lopes, Mateus Amaral, Cássio Rampinelli, Ana Maria Motta, Vinicius Grossi, Sara Ferrigo, Naiara Raiol, Lucas Liporoni, Orlandina Martins, Diogo Gebrim, pelas caronas, conhecimento e ajudas nesse tempo do mestrado.

À minha professora de graduação Yamile Valencia, da Universidade Nacional da Colômbia, pela ajuda e pelos conselhos para ingressar no mestrado.

Aos meus amigos da pós-graduação em Geotecnia da Universidade de Brasília Daniel Calvo, Marcelo Llano, Robinson Zuluaga, Carolina Lopez, Carolina Ruiz, Estefania Muñoz, Julian Buritica e Maria Paz Duque e ao professor Hernan Martinez, pela amizade e ajuda nesse tempo no Brasil.

À Fundação Cearense de Meteorologia e Recursos Hídricos, FUNCEME, pelos dados de vazões necessários para elaboração dos mapas digitais.

Ao professor Julio Cañon da Universidade de Antioquia e University of Arizona, pelas dúvidas sanadas.

À engenharia sanitária Ana Maria Jaramillo, pela amizade e ajuda nos temas de Qualidade da Água, Saneamento e Gestão Ambiental.

Aos meus amigos na pós-graduação em Hidráulica e Saneamento da Universidade de São Paulo Antônio, Maria e Irene, pela ajuda durante o tempo em São Carlos.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, pela bolsa concedida.

## RESUMO

Os modelos de Markov com estados ocultos (*Hidden Markov Models* - HMMs) apresentam características propícias para simular a persistência observada em séries de vazões, especialmente pela sua estrutura de variação de estados “ocultos”, que é descrita por uma Cadeia de Markov, sendo potencialmente úteis em regiões onde a variabilidade climática produz regimes ou estados hidrológicos secos e úmidos, até porque permitem considerar explicitamente índices climáticos que influenciam a persistência.

Este trabalho avaliou o desempenho dos HMMs, e sua versão não-homogênea (NHMMs), que fazem uso dos indicadores climáticos NINO3 e Dipolo do Atlântico, na geração de séries sintéticas de vazões anuais, observando a capacidade de tais modelos em representar adequadamente diversas estatísticas da série anual de vazões afluentes ao reservatório Orós, localizado no Nordeste do Brasil. O desempenho de tais modelos foi comparado com o desempenho do clássico modelo ARMA, levando em consideração diversas estatísticas de persistência hidrológica.

Modelos HMM com dois e três estados ocultos, baseados na distribuição Gama (HMM-G) e Log-Normal (HMM-NLT), foram capazes de representar adequadamente a grande maioria das estatísticas de persistência selecionadas, tendo obtido desempenho similar ao do modelo AR(2), o mais adequado dentre os modelos ARMA para a região. Testes de hipóteses foram realizados para cada modelo e para cada estatística de persistência. Apenas o modelo HMM-G2 (dois estados ocultos) foi rejeitado para uma das métricas de persistência (valor-p = 6.8%), o volume de déficit médio de período seco. Todos os outros modelos HMMs e o AR(2) passaram nos testes de hipóteses para todas as métricas empregadas. Entretanto, observaram-se diferenças nos desempenhos de cada modelo, embora não tenha sido possível indicar um melhor modelo, pois o desempenho varia com a métrica escolhida. A inclusão de indicadores climáticos na análise traz alguns benefícios pontuais, como melhorar a representação da autocorrelação, quando se utiliza o NHMM-NLT2, e do volume médio de deficit do modelo NHMM-G2.

Os resultados mostram que os modelos HMMs e NHMMs apresentaram resultados satisfatórios ao representar as características estatísticas da série estudada, com desempenho similar aos modelos ARMA, tendo superado tais modelos na representação de vazões máximas e mínimas, assim como em algumas métricas de persistência, tais como, comprimentos máximo e médio de períodos úmidos.

## ABSTRACT

Hidden Markov models (HMMs) have specific characteristics that make them well suited to simulate observed hydrologic persistence in streamflow records, especially because of the way it models the variability of hidden states, described by a Markov chain. HMMs are potentially useful in regions where climate variability produces persistent dry and wet hydrologic states, especially because it may explicitly employ climate indicators that influence persistence.

This dissertation evaluated the performance of HMMs, and its non-homogeneous version (NHMMs), which employes NINO3 and Atlantic Dipole climate indicators, in the generation of synthetic annual inflow series onto Orós reservoir, located in Northeastern Brazil. The performance of such models was compared to the performance of classical ARMA models, focusing in their ability to reproduce several hydrologic persistence statistics.

HMMs with two and three hidden states, based upon Gamma (HMM-G) and log-Normal (HMM-NLT) distributions, were able to represent properly most of hydrologic persistent statistics and had a performance as good as the one provided by AR(2) model, the best ARMA model for the basin. Hypothesis tests were performed for each model and for each one of the persistent statistics employed in the study. Among HMMs, just HMM-G2 (two states) was statistically rejected ( $p$ -value = 6.8%) for not being able to reproduce the mean deficit volume in dry periods. All other HMMs and the AR(2) model passed the test. However, differences in performance among models were observed, but it was not possible to select the best model because no single model outperformed others in the reproduction of all statistics. The inclusion of climate indicators in the analysis provided some specific benefits in the performance of some models, for example, it improved the way NHMM-NLT2 reproduces the lag-1 autocorrelation and helped NHMM-G2 with the mean deficit volume during dry periods.

In general, results show that HMMs and NHMMs provided synthetic annual streamflow series that properly reproduce many desirable statistics observed in the record, especially those related to hydrologic persistence, had performance similar to the one achieved by ARMA models, and outperformed ARMA models in reproducing some specific statistics, such as, minimum and maximum flows, and maximum and mean length of wet periods.

## SUMARIO

LISTA DE TABELAS .....	xiii
LISTA DE FIGURAS .....	xiv
LISTA DE NOMENCLATURA E ABREVIACOES .....	xviii
LISTA DE SMBOLOS .....	xx
1 - INTRODUO .....	1
2 - OBJETIVOS .....	4
2.1 - OBJETIVO GERAL .....	4
2.2 - OBJETIVOS ESPECFICOS .....	4
3 - MODELAGEM ESTOCSTICA DE SRIES DE VAZOES .....	5
3.1 - MOTIVAO .....	5
3.2 - BREVE HISTRICO .....	7
3.3 - PERSISTNCIA EM SRIES TEMPORAIS HIDROLGICAS .....	8
4 - REVISO DE LITERATURA .....	12
4.1 - APLICAOES DOS MODELOS DE MARKOV OCULTO NA MODELAGEM HIDROLGICA .....	12
4.1.1 - Aplicaes dos Modelos Homogneos de Markov com Estados Ocultos (HMM) .....	12
4.1.2 - Aplicaes dos Modelos de Markov com Estados Ocultos No Homogneos .....	16
4.2 - MODELAGEM DE VAZOES NA REGIO DO NORDESTE DO BRASIL .....	22
5 - REFERENCIAL TERICO .....	26
5.1 - PROCESSOS ESTOCSTICOS E SRIES TEMPORAIS .....	26
5.2 - FUNO DE AUTOCORRELAO .....	31
5.3 - MODELOS AUTOREGRESSIVOS DE MDIAS MVEIS .....	33
5.4 - MODELOS DE MARKOV COM ESTADOS OCULTOS .....	35
5.4.1 - Definies e Propriedades .....	36
5.4.2 - Definio dos Modelos de Markov Oculto .....	40
5.4.3 - Modelos de Markov Oculto No Homogneo .....	53
6 - INFLUENCIA DA VARIABILIDADE CLIMTICA NO REGIME HIDROLGICO DO NORDESTE DO BRASIL .....	56
6.1 - CARACTERSTICAS DO NORDESTE DO BRASIL .....	56
6.2 - FENMENOS CLIMTICOS E SISTEMAS ATMOSFRICOS SOBRE O NORDESTE DO BRASIL .....	57

6.2.1 - <i>El Niño – Southern Oscillation</i> .....	58
6.2.2 - Sistemas de Circulação Atmosférica .....	64
6.3 - VARIABILIDADE CLIMÁTICA SOBRE O NORDESTE .....	67
6.3.1 - Variabilidade Sazonal e Espacial.....	67
6.3.2 - Variabilidade Interanual .....	68
6.3.3 - Variabilidade Decenal e de Longo Prazo .....	72
6.3.4 - Variabilidade climática e o regime das vazões no Nordeste do Brasil .....	74
7 - METODOLOGIA.....	76
7.1 - ESTRATÉGIA METODOLÓGICA .....	76
7.2.1 - Definição dos Modelos HMM propostos.....	77
7.2 - ESTATÍSTICAS PARA IDENTIFICAR A PERSISTÊNCIA.....	78
7.2.1 - Comprimentos e Volumes de Períodos Hidrológicos .....	78
7.2.2 - Número total de comprimentos .....	79
7.3 - TESTE DE ADERÊNCIA DO QUI-QUADRADO ( $\chi^2$ ) .....	79
7.4 - MVN-HMM TOOLBOX.....	80
7.5 - CRITÉRIOS DE SELEÇÃO DE MODELOS ESTOCÁSTICOS .....	81
8 - ESTUDO DE CASO.....	83
8.1 - LOCAL DE APLICAÇÃO .....	83
8.2 - DADOS UTILIZADOS.....	84
8.2.1 - Dados dos Índices Climáticos.....	84
9 - RESULTADOS E DISCUSSÃO.....	86
9.1 - ANÁLISE DA SÉRIE DE VAZÕES OBSERVADAS .....	86
9.2 - AJUSTE DOS MODELOS ARMA .....	94
9.3 - AJUSTE DOS MODELOS HMM.....	95
9.3.1 - Momentos dos HMM Ajustados.....	96
9.3.2 - Estimação das probabilidades dos estados ocultos nos HMM ao longo do tempo.....	100
9.3.3 - Séries Viterbi .....	102
9.3.4 - Distribuições de Probabilidade Marginais .....	105
9.4 - COMPARAÇÃO DAS SÉRIES SINTÉTICAS GERADAS PELOS MODELOS ARMA E HMMs.....	107
9.4.1 - Momentos.....	108
9.4.2 - Métricas de persistência .....	112
9.5 - AJUSTE DE MODELOS DE MARKOV OCULTO NÃO HOMOGÊNEO .....	117

9.5.1 - CORRELAÇÃO ENTRE AS VAZÕES OBSERVADAS E OS ÍNDICES CLIMÁTICOS NINO3 E DIPOLO DO ATLÂNTICO .....	117
9.6 - COMPARAÇÃO DAS SÉRIES SINTÉTICAS GERADAS PELOS MODELOS ARMA E NHMMs .....	119
9.6.1 - Momentos .....	120
9.6.2 - Métricas de Persistência .....	122
10 - CONCLUSÕES .....	127
REFERÊNCIAS BIBLIOGRÁFICAS .....	130
APÊNDICE A .....	140

## LISTA DE TABELAS

Tabela 4.1 - Características dos modelos estocásticos nos trabalhos relatados na literatura, que utilizaram HMM (m estados), para a modelagem de séries de vazões.....	21
Tabela 4.2 - Características dos modelos estocásticos nos trabalhos relatados na literatura, que utilizaram NHMM (m estados) .....	22
Tabela 6.1 - Regiões dos índices do ENSO, (Schöngart e Junk, 2007).....	62
Tabela 6.2 - Anos de seca no Nordeste brasileiro, coincidentes com anos de El Niño, durante os últimos quatro séculos atualizado para 1998 (Marengo, 2006). .....	73
Tabela 8.1- Características do reservatório Orós (COGERH, 2011).....	83
Tabela 9.1 - Estatísticas descritivas da série de vazões observada.....	88
Tabela 9.2 - Parâmetros e intervalos de confiança de 95% estimados para a distribuição Normal, Log-Normal e Gama das vazões observadas. ....	91
Tabela 9.3- Estatística T para a ocorrência de períodos secos e úmidos .....	93
Tabela 9.4 - Métricas de persistência com relação aos percentis P33 e P66.....	93
Tabela 9.5- Valores do AIC e BIC dos melhores modelos .....	94
Tabela 9.6 - Valores do AIC, BIC, Log-Likelihood (LLL) e numero de parâmetros (k), para os HMM de m estados ocultos. ....	96
Tabela 9.7 – Matrizes de probabilidades de transição dos HMM de 2 estados ocultos ....	101
Tabela 9.8 - Duração média dos períodos secos e úmidos em anos, nos HMM de 2 estados ocultos.....	101
Tabela 9.9 - Coeficientes de correlação Spearman ( $\rho_s$ ) e os seus respectivos valores p para a correlação entre as vazões anuais e os índices climáticos NINO3 e DIPOLo do Atlântico mensais para o período de 1910-2000.....	118

## LISTA DE FIGURAS

Figura 3.1 - Estrutura de um estudo de simulação, indicando a transformação de uma sequencia de vazões sintéticas, demandas futuras, desenho e as políticas operacionais para produzir estatísticas do desempenho do sistema, (Loucks et al. 2005, traduzido).	7
Figura 5.1 - Serie temporal de vazões afluentes ao reservatório Orós. Período 1911-200. Nível de agregação: série mensal (acima) e série anual (abaixo), (Fonte: Base de dados).	28
Figura 5.2 - Processo de uma distribuição de mistura de dois componentes. De cima para baixo, os estados são 1, 2, 1, 1, 2, 1. No meio são apresentadas as correspondentes distribuições componentes. As observações no lado direito são geradas pelas densidades componentes ativas, (Zucchini e MacDonald, 2009, traduzido).	37
Figura 5.3 - Representação gráfica de uma Cadeia de Markov de primeira ordem	38
Figura 5.4 - Representação gráfica de um Modelo de Markov Oculto, HMM	42
Figura 5.5 - Geração das observações em um HMM de dois estados ocultos. A Cadeia de Markov segue o caminho 2, 1, 1, 1, 2, 1 como se amostra na esquerda. As distribuições que são ativadas dependem dos estados ativos, apresentadas não meio. As observações são geradas dependendo da distribuição que é ativada, (Zucchini e MacDonald, 2009, traduzido)	42
Figura 5.6 - Representação gráfica de um AR-HMM.	52
Figura 5.7 – Representação gráfica de um Input-Output HMM.	53
Figura 5.8 – Representação gráfica de um HMM Fatorial.	53
Figura 5.9 - Representação gráfica de um Modelo de Markov Oculto Não Homogêneo, NHMM	54
Figura 6.1 - Estados do ENSO: a) Condições El Niño para o mês de julho de 1997 e b) condições La Niña para o mês de Dezembro de 1999, (IRI, 2012b).	59
Figura 6.2 - Mecanismos do ENSO: a) condições normais observam-se águas superficiais relativamente mais frias no Pacífico equatorial leste, junto à costa oeste da América do Sul, e relativamente mais aquecidas no Pacífico equatorial Oeste, b) condições El Niño, ocorre uma diminuição das águas mais frias que afloram próximo à costa oeste da América do Sul e c) condições La Niña, As águas adjacentes à costa oeste da América do Sul tornam-se ainda mais frias devido à intensificação do movimento de ressurgência, (INPE/CPTEC, 2012).	60

Figura 6.3 - Convencional índice da Oscilação Sul (Southern Oscillation Index-SOI) medido como a diferença entre a pressão atmosférica média ao nível do mar entre Darwin (Austrália) e Papeete (Taiti) dado em mbars (linha tracejada) e um índice da SST para a Oscilação Sul (linha solida). Nesta figura o SOI tem sido duplicado, (Storch e Zwiers, 1999, traduzido).....	61
Figura 6.4 - Efeitos globais do ENSO para os meses de dezembro janeiro e fevereiro: a) efeitos do El Niño, b) efeitos da La Niña (INPE/CPTEC, 2012) .....	63
Figura 6.5 - Localização das regiões dos índices ENSO (IRI, 2012b) .....	63
Figura 6.6 - Zona de Convergência Intertropical - ZCIT, sinalada pela elipse vermelha, (INPE/ CPTEC, 2012).....	65
Figura 6.7 - Localização das zonas das regiões que constituem o DIPOLO do Atlântico, o dipolo é definido como a diferencia aritmética das anomalias da SST do Atlântico norte tropical (TAN) e as SST do Atlântico sul tropical (TAS), (IRI, 2012 a) .....	71
Figura 6.8 – Representação dos padrões atmosféricos e SST no Nordeste produzidos pelo DIPOLO do Atlântico: a) períodos secos e b) chuvosos no Nordeste (Ferreira e Mello, 2005).....	72
Figura 7.1 - Estratégia metodologia.....	77
Figura 7.2 - Definição de Estatísticas de períodos secos e úmidos (Prairie et al. 2008, traduzido).....	79
Figura 8.1 - Localização do reservatório Orós (Autoria própria).....	83
Figura 9.1 - Série de vazões observadas afluentes no reservatório Orós. Período 1911-200. ....	87
Figura 9.2 - Histograma das vazões observadas.....	87
Figura 9.3 - Diagrama Box Plot para as vazões médias anuais afluentes no reservatório Orós. Período 1911-2000.....	88
Figura 9.4 - Função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) para a série de vazões observadas. ....	89
Figura 9.5 – Função de densidade de probabilidade normal das vazões observadas. ....	90
Figura 9.6 - Função de densidade de probabilidade log-normal das vazões observadas. ....	91
Figura 9.7 - Função de densidade de probabilidade Gama das vazões observadas.....	92
Figura 9.8 - Mediana das vazões observadas (Qmd) (superior) e ocorrência de períodos secos e úmidos com relação à Qmd (inferior) .....	93

Figura 9.9 - Percentil 33% (P33) e percentil 66% (P66) da série de vazões observada (superior) e ocorrência de períodos secos e úmidos com relação aos percentis P33 e P66 (inferior).....	94
Figura 9.10 - Média (a) e desvio padrão (b) produzidos pelos modelos HMM-Gs.....	97
Figura 9.11 - Coeficiente de assimetria (a) e r1 (b), produzidos pelos modelos HMM-Gs.	98
Figura 9.12 - Média (a) e desvio padrão (b) produzidos pelos modelos HMM-NLTs.....	99
Figura 9.13 - Coeficiente de assimetria (a) e r1 (b), produzidos pelos modelos HMM-NLTs .....	99
Figura 9.14 - Probabilidade do estado s1 e s2 ao longo do tempo no HMM-NLT (2).....	101
Figura 9.15 - Probabilidade do estado s1 e s2 ao longo do tempo no HMM-G (2) .....	102
Figura 9.16 - Sequência de estados mais prováveis no HMM-NLT (2) (superior), vazões observadas (centro) e sequência de estados mais prováveis no HMM-G (2) (inferior) .....	104
Figura 9.17 - Sequência de estados mais prováveis no HMM-G (3) (superior), vazões observadas (centro) e sequência de estados mais prováveis no HMM-NLT (6) (inferior).....	104
Figura 9.18 – Probabilidade das vazões ajustadas ao HMM-NLT (2) em “escala logarítmica” (superior) e em “escala real” (exp(Qsim)) (inferior) .....	106
Figura 9.19 - Probabilidade das vazões ajustadas ao HMM-G (2) (superior) em escala semi-logarítmica e em “escala real” (inferior).....	106
Figura 9.20 - Probabilidade das vazões ajustadas ao HMM-G (3) (superior) em escala semi-logarítmica e em “escala real” (inferior).....	107
Figura 9.21 - Média (superior) e desvio padrão (inferior) das series sintéticas geradas pelos HMM, AR(2) e GA (2).....	110
Figura 9.22 – Coeficiente de assimetria (superior) e autocorrelação de primeira ordem (inferior) das series sintéticas geradas pelos HMM, AR(2) e GA (2).....	110
Figura 9.23 – Vazão máxima (superior) e vazão mínima (inferior) das series sintéticas geradas pelos HMM, AR(2) e GA (2).....	112
Figura 9.24 - Estatística T para os períodos secos com relação ao percentil P33 (superior) e à mediana Qmd (inferior). .....	113
Figura 9.25 - Comprimento máximo dos períodos úmidos com relação ao percentil P66 (superior) e à mediana Qmd (inferior). .....	114
Figura 9.26 - Comprimento médio dos períodos secos com relação ao percentil P33 (superior) e à mediana Qmd (inferior) .....	115

Figura 9.27 - Volume máximo dos períodos úmidos com relação ao percentil P66 (superior) e à mediana Qmd (inferior) .....	116
Figura 9.28 - Volume médio dos períodos úmidos com relação ao percentil P66 (superior) e à mediana Qmd (inferior).....	116
Figura 9.29 - Correlação Spearman entre as vazões anuais e os índices climáticos NINO3 (superior) e DIPOLO do Atlântico (inferior) mensais. Período 1910-2000.....	118
Figura 9.30 - Média (superior) e desvio padrão (inferior) das series sintéticas geradas pelos NHMMs, AR(2) e o GA(2). .....	121
Figura 9.31 - Coeficiente de assimetria (superior) e autocorrelação de primeira ordem (inferior) das series sintéticas geradas pelos NHMMs, AR(2) e o GA(2).....	121
Figura 9.32 - Vazão máxima (superior) e vazão mínima (inferior) das series sintéticas geradas pelos NHMMs, AR(2) e o GA(2).....	122
Figura 9.33 - Estatística T dos períodos úmidos com relação ao percentil P66 (superior) e à mediana Qmd (inferior).....	123
Figura 9.34 - Comprimento máximo dos períodos úmidos com relação ao percentil P66 (superior) e à mediana Qmd (inferior) .....	124
Figura 9.35 - Comprimento médio dos períodos úmidos com relação ao percentil P66 (superior) e à mediana Qmd (inferior) .....	124
Figura 9.36 - Volume de déficit máximo com relação ao percentil P33 (superior) e à mediana Qmd (inferior).....	125
Figura 9.37 - Volume de déficit médio com relação ao percentil P33 (superior) e à mediana Qmd (inferior).....	125
Figura 9.36 - Volume máximo dos períodos úmidos com relação ao percentil P33 (superior) e à mediana Qmd (inferior) .....	144

## LISTA DE NOMENCLATURA E ABREVIACOES

ACF	Funo de Autocorrelao
ACP	Anlise de Componentes Principais
AGP	Altura Geopotencial
AR(1)	Modelo Autoregressivo de Ordem 1
AR-HMM	Modelo de Markov Oculto Autoregressivo
ARMA	Modelo Autoregressivo de Mdia Mvel
CCM	Complexos Convectivos de Mesoescala
CDF	Funo de Probabilidade Acumulada
CRD	Correlao entre os Resduos da Regresso
DEP	Distribuio Espacial da Precipitao
DTPO	Depresso da Temperatura do Ponto de Orvalho
EM	Algoritmo de Maximizao da Expectativa
ENSO	<i>El Nio-Southern Oscillation</i>
FF	Frentes Frias
GA (2)	Distribuio Gama de Dois Parmetros
HMM	Modelo de Markov Oculto
HMM-G	HMM com Distribuio Dependentes Gama
HMM-LN	HMM com Distribuio Dependentes Log-Normais
HMM-N	HMM com Distribuio Dependentes Normais
HMM-NLT	HMM com Distribuio Normais e Log-Transformao nos Dados
IDO	<i>Indian Dipole Index</i>
IPO	<i>Interdecadal Pacific Oscillation</i>
IRI	<i>International Research Institute for Climate and Society</i>
KNN	<i>K-Nearest Neighbor</i>
L	Verossimilhana
LI	Linhas de Instabilidade
LL	Logaritmo da Mxima Verossimilhana
LT	Log-Transformao
M	Multivariado
MA	Modelo de Mdia Mvel
MBAM	Modelo Binrio Autologstico Multivariado

MC	Cadeia de Markov
MPT	Matriz de Probabilidades de Transição
MVN-HMM	<i>MultiVariate Nonhomogeneous Hidden Markov toolbox</i>
MVS	Máxima Verossimilhança
NEB	Nordeste do Brasil
NHMM	Modelo de Markov Oculto Não Homogêneo
NHMM-G	Versão Não Homogênea do HMM-G
NHMM-NLT	Versão Não Homogênea do HMM-NLT
NINO3	Índice do ENSO na região Niño 3
NP	Não Paramétrico
OL	Ondas de Leste
P	Precipitação
PACF	Função de Autocorrelação Parcial
PAR	Modelo Periódico Autoregressivo
PARX	Modelo Periódico Autoregressivo com Variáveis Exógenas
PDF	Função de Densidade de Probabilidade
PDO	<i>Pacific Decadal Oscillation</i>
PMF	Função de Probabilidade de Massa
PMNM	Pressão Media ao Nível do Mar
PNM	Pressão no Nível do Mar
SIN	Sistema Interligado Nacional
SL	<i>Shifting Mean/Level Model</i>
SOI	<i>Southern Oscillation Index</i>
SSI	<i>State Signal Index</i>
SST	<i>Sea Surface Temperature</i>
TAN	Atlântico Norte Tropical
TAS	Atlântico Sul Tropical
TTAP	Teor Total de Água Precipitável
U	Univariado
VCANs	Vórtices Ciclônicos de Altos Níveis
ZCIT	Zona de Convergência Intertropical

## LISTA DE SÍMBOLOS

$A$	Amplitude
$AIQ$	Amplitude Inter-Quartis
$\alpha$	Parâmetro de forma da distribuição Gama
$\alpha_t$	Probabilidade para frente
$\beta$	Parâmetro de escala da distribuição Gama
$\beta_t$	Probabilidade para trás
$CM_{up}, CM_{umd}$	Comprimento Máximo dos Períodos Úmidos com relação ao $P_{66}$ e à $Q_{md}$
$CM_{sp}, CM_{smd}$	Comprimento Máximo dos Períodos Secos com relação ao $P_{33}$ e à $Q_{md}$
$CME_{up}, CME_{um}$	Comprimento Médio dos Períodos Úmidos com relação ao $P_{66}$ e à $Q_{md}$
$CME_{sp}, CME_{smc}$	Comprimento Médio dos Períodos Secos com relação ao $P_{33}$ e à $Q_{md}$
$C_s$	Coefficiente de Assimetria Amostral
$\Gamma$	Matriz de Probabilidades de Transição
$\gamma$	Coefficiente de Assimetria
$\gamma_{ij}$	Probabilidade de Transição do Estado $i$ ao Estado $j$
$E_i$	Frequência Esperada
$F_r(r)$	Função de probabilidade Acumulada de $r$ no caso contínuo
$f_i(r)$	PDF da distribuição dependente do estado $i$ em um HMM no caso contínuo
$f_r(r)$	Função de densidade de probabilidade de $r$
$H_0$	Hipótese Nula
$H_1$	Hipótese Alternativa
$IC$	Intervalo de Confiança
$\Theta$	Vetor de Parâmetros de um HMM ou um NHMM
$\theta_i$	Vetor dos parâmetros da distribuição componentes $i$ de um HMM
$\theta_q$	Parâmetro de um Modelo de Média Móvel de Ordem $q$
$k$	Defasagem ou <i>Lag</i>
$\kappa$	Coefficiente de Curtose
$L_N$	Verossimilhança em um HMM
$l$	Anos Consecutivos em um Estado Climático seco ou úmido
$\lambda_i, \sigma_{ij}, \rho_i$	Parâmetros de Transição do Estado Oculto $i$ ao Estado Oculto $j$ em um

$m$	Número de estados ou Ordem em um HMM
$\mu_R$	Média ou Valor esperado de $R$
$\mu_i$	Média da Distribuição dependente $i$ de um HMM
$\mu_Q$	Média Populacional de $Q$
$\mu_Y$	Média Populacional de $Y$
$N$	Comprimento da série
$N_p$	Número de Parâmetros em um HMM
$n$	Intervalos de Classe
$n_c(l)$	Número Total de <i>Clusters</i> com $l$ Anos Consecutivos em um Estado Climático
$O_i$	Frequência Observada
$\pi_i$	Distribuição Estacionaria do estado oculto $i$ em um HMM
$P_{33}$	Percentil de 33%
$P_{66}$	Percentil de 66%
$P$	Probabilidade no caso contínuo
$P_r$	Probabilidade no caso discreto
$P_r(r)$	Função acumulada de probabilidade de $r$ no caso discreto
$P_{sp}, P_{smd}$	Período seco com relação ao $P_{33}$ e à $Q_{md}$
$P_{up}, P_{umd}$	Período úmido com relação ao $P_{66}$ e à $Q_{md}$
$p$	Ordem de um Modelo Autoregressivo
$p_i(r)$	PDF da distribuição dependente do estado $i$ em um HMM no caso discreto
$p_r(r)$	Função massa de probabilidade de $r$ no caso discreto
$\rho_k$	Autocorrelação na Defasagem $k$
$Q_1$	Primeiro Quartil
$Q_3$	Terceiro Quartil
$Q$	Série de Vazões Observadas
$Q(t)$	Vazão do tempo $t$
$Q_{cs}$	Coefficiente de Assimetria de $Q$
$Q_m$	Média de $Q$
$Q_{max}$	Vazão Máxima

$Q_{md}$	Mediana de $Q$
$Q_{min}$	Vazão Mínima
$S_Q$	Desvio padrão de $Q$
$Q_s^2$	Variância de $Q$
$Q_{sim}$	Série Sintética ou Simulada de Vazões
$q$	Ordem de um Modelo de Média Móvel
$R$	Variável aleatória
$R_t, x_t$	Variável Aleatória no Tempo $t$
$\bar{R}$	Média amostral de $R$
$R_{1:N}$	Sequencia de observações de comprimento $N$ , $\{R_1, R_2, \dots, R_N\}$
$R_{md}$	Mediana de $R$
$R_{-t}$	Sequência de observações em tempos diferentes ao tempo $t$
$r$	Possível valor de $R$
$r_1$	Autocorrelação de ordem 1
$r_k$	Autocorrelação Amostral defasagem $k$
$r_{pk}$	Autocorrelação Parcial Amostral no Lag $k$
$r_{-t}$	Possível Valor de $R_{-t}$
$\overline{S_R^2}$	Variância amostral de $R$
$s_i$	Estado Oculto $i$
$s_t$	Estado Oculto no tempo $t$
$\sigma_R^2$	Variância de $R$
$\sigma_i^2$	Variância da Distribuição dependente $i$ em um HMM
$\sigma_Q$	Desvio Padrão Populacional de $Q$
$\sigma_Y$	Desvio Padrão Populacional de $Y$
$T$	Número total de Comprimentos com Diferentes magnitudes na Série $Q$ ou $Q_{sim}$
$T_{sp}, T_{smd}$	$T$ dos Períodos Secos com relação ao $P_{33}$ e à $Q_{md}$
$T_{up}, T_{umd}$	$T$ dos Períodos Úmidos com relação ao $P_{66}$ e à $Q_{md}$
$t$	Tempo
$u_i(t)$	Probabilidade de Marginal do Estado $i$ no tempo $t$
$VM_{up}, VM_{umd}$	Volume Máximo dos Períodos Úmidos com relação ao $P_{66}$ e à $Q_{md}$
$VM_{sp}, VM_{smd}$	Volume de Déficit Máximo com relação ao $P_{33}$ e à $Q_{md}$

$VME_{up}, ME_{umd}$	Volume Médio dos Períodos Úmidos com relação ao $P_{66}$ e à $Q_{md}$
$VME_{sp}, VME_{smc}$	Volume de Déficit Médio com relação ao $P_{33}$ e à $Q_{md}$
$w_t$	Variável Aleatória Normalmente Distribuída com $\mu_w = 0$ e $\sigma_w^2 = 1$
$X_t$	<i>Input</i> ou variável exógena em um NHMM
$X_{1:N}$	Sequencia de variáveis exógenas de comprimento $N$ , $\{X_1, X_2, \dots, X_N\}$
$Y$	Série do Logaritmo Natural das Vazões Observadas, $Y = \ln(Q)$
$z_t$	Sequencia White Noise ou Ruído Aleatório
$\phi_p$	Parâmetro de um Modelo autoregressivo de Ordem $p$
$\Omega$	Matriz dos Parâmetros $\lambda_i, \sigma_{ij}, \rho_i$ de tamanho $1 \times m$ em um NHMM

# 1 - INTRODUÇÃO

Modelos de geração de séries sintéticas de vazão vêm sendo utilizados desde 1960 para avaliar o desempenho de alternativas de sistemas hídricos, tanto em nível de planejamento quanto de gerenciamento dos recursos hídricos (Salas *et al.* 1980). Esse tipo de abordagem permite avaliar o tamanho ótimo e os benefícios da construção de um novo reservatório, da transposição de águas de uma bacia para fins de abastecimento, na avaliação de diferentes regras de operação de um sistema de reservatórios, assim como na avaliação de diferentes políticas de alocação de água (Loucks *et al.* 2005). A utilização desses modelos se baseia no fato de que a série histórica disponível não é suficiente para representar as possíveis sequências de vazão que podem ocorrer no futuro. A justificativa para empregar tais modelos é ainda mais forte quando o interesse está em avaliar o desempenho de sistemas hídricos quando da ocorrência de eventos extremos, como uma sequência de anos secos (Thyer e Kuczera, 2000; Loucks *et al.* 2005).

De uma forma geral, modelos estocásticos de vazão empregados no passado consideravam apenas variações sazonais e mensais no comportamento das vazões, e não levavam em consideração possíveis variações dos parâmetros do modelo em escala anual. Esse fato pode ser visto como uma limitação, pois observa-se em várias partes do mundo a presença de ciclos de variação de baixa frequência nas séries de vazões observadas, com sequências relativamente longas de anos secos, seguidas de sequências de anos mais úmidos (Akintug e Rasmussen, 2005). Esses ciclos de variabilidade estão, muitas vezes, relacionados com mecanismos climáticos de larga escala, geralmente associados aos padrões de temperatura da superfície dos oceanos (Uvo *et al.*, 1998). Por exemplo, na região do Nordeste do Brasil (NEB), sabe-se que esses ciclos de anos secos e úmidos estão relacionados com o fenômeno do *El Niño-Southern Oscillation*, que ocorre em uma escala interanual entre 4-7 anos, e com o sistema de altas pressões no oceano Atlântico tropical, que tende a produzir ocorrência de regimes climáticos secos e úmidos principalmente na escala intranual e interanual, produto de um gradiente das temperaturas da superfície desse oceano. (Uvo *et al.* 1998; Hastenrath, 2000; Kayano e Andreoli, 2006; Marengo, 2008). Os estudos de Souza Filho e Lall, (2003) e Lima, (2010), por exemplo, mostram que ciclos de estados úmidos e secos vêm sendo observados no NEB.

Embora não seja possível afirmar categoricamente, modelos estocásticos tradicionais de geração de séries sintéticas de vazão baseados na família de modelos ARMA (Modelos Autoregressivos e de Média Móvel) podem ter dificuldade em simular adequadamente esses ciclos mais longos, o que pode levar a conclusões errôneas sobre o desempenho de sistemas hídricos. A observação dessas longas persistências de anos secos e úmidos induziu ao desenvolvimento de outros modelos estocásticos, que tentam de maneira explícita representar essas mudanças de estado.

Modelos de Markov com estados ocultos (HMM) são fortes candidatos para a geração de séries sintéticas de vazão nas quais se observa longa persistência de períodos úmidos e secos, pois os mesmos possuem uma maneira explícita de definir o regime de vazões em que o rio se encontra. Os HMM podem, ainda, levar em consideração variáveis exógenas, tal como indicadores climáticos que sabidamente exercem influência sobre as vazões do rio, passando a ser chamados então de Modelos de Markov com Estados Ocultos Não-Homogêneos (NHMM). Os NHMM são úteis para a geração de séries sintéticas de vazão, ainda que não se conheça a série de indicadores no futuro, sendo possível realizar a simulação com as séries observadas de indicadores climáticos. O uso potencial dos NHMMs está na previsão climática de vazões, embora a mesma não seja abordada neste trabalho.

Considerando isso, esta dissertação explora o uso de HMMs e NHMMs para modelar a série de vazões anuais afluentes ao reservatório Orós, localizado no estado do Ceará, no NEB. Esta série foi estudada devido à importância e ao porte do rio Jaguaribe na região, além de ser uma série estatisticamente útil por não apresentar falhas e ter um comprimento longo (90 anos). O objetivo foi avaliar o desempenho desses modelos e compará-los com modelos do tipo ARMA. A avaliação não se restringe às métricas usualmente empregadas nesses casos, como a representação dos momentos da distribuição de probabilidades das vazões anuais e sua função de autocorrelação, mas engloba também estatísticas relacionadas à persistência das vazões em estados secos e úmidos.

Depois dessa introdução, o documento prossegue com a apresentação dos objetivos do estudo seguida de uma revisão bibliográfica de modelagem estocástica de vazões. Na sequência, o Capítulo 4 segue com uma revisão de literatura dos HMMs e NHMMs e o Capítulo 5 apresenta formulação teórica, que inclui tanto os modelos ARMA quanto os

modelos HMMs e NHMMs. Uma revisão da variabilidade climática na região do NEB é dada no Capítulo 6, seguido do Capítulo 7, que contém a metodologia do estudo e do Capítulo 8, com a descrição do estudo de caso. Os resultados e a discussão, Capítulo 9, precedem então as conclusões da pesquisa, apresentada por fim no Capítulo 10.

## **2 - OBJETIVOS**

### **2.1 - OBJETIVO GERAL**

Explorar o uso de modelos de Markov com Estados Ocultos na geração de séries sintéticas de vazões anuais, e compará-los com modelos do tipo ARMA e a distribuição Gama de dois parâmetros na tentativa de reproduzir ciclos de períodos úmidos e secos observados na série de vazão afluyente ao reservatório Orós, localizado no Estado do Ceará.

### **2.2 - OBJETIVOS ESPECÍFICOS**

- Avaliar o desempenho de modelos de Markov com Estados Ocultos (HMMs) empregando as distribuições Gama, para representar as vazões anuais na escala real, e a distribuição Normal, para representar as vazões anuais na escala logarítmica;
- Avaliar o ganho de se incorporar os indicadores climáticos na representação das vazões anuais com o emprego da versão não-homogênea dos modelos de Markov com Estados Ocultos (NHMMs);
- Comparar o desempenho dos modelos HMMs e NHMMs na geração de séries sintéticas de vazão anual com o desempenho de modelos do tipo Autoregressivos de Médias Móveis (ARMA) e a distribuição Gama de dois parâmetros, empregando métricas que expressem a questão da persistência das vazões em estados secos e úmidos.

### **3 - MODELAGEM ESTOCÁSTICA DE SÉRIES DE VAZÕES**

#### **3.1 - MOTIVAÇÃO**

Na hidrologia, existe sempre o interesse de descrever os processos hidrológicos que ocorrem em uma bacia. Para alcançar esse objetivo, faz-se uso de modelos para descrever o comportamento das variáveis que fazem parte do ciclo hidrológico, tais como precipitação, vazão, infiltração, e evapotranspiração, empregando características físicas próprias da bacia. Existe uma ampla variedade de modelos que realizam a simulação dos processos na bacia, que podem ser classificados em duas grandes classes: os determinísticos e os probabilísticos. A segunda classe de modelos, objeto desta dissertação, assume que os processos físicos podem ser modelados por variáveis aleatórias, que seguem leis probabilísticas, sendo então chamados de modelos estocásticos ou estatísticos (Bayer e Castro, 2012).

Os modelos estocásticos formam o que se chama de hidrologia estocástica, que por sua vez se divide em duas partes. A primeira parte se refere ao estudo de frequência de eventos extremos, com diversas aplicações, como por exemplo, no dimensionamento de obras de controle de cheias, enquanto que a segunda parte trata da análise de séries temporais, que pode ser empregada no planejamento e análise de sistemas de recursos hídricos cujas variáveis de entrada possuem grande variabilidade e incerteza. O estudo de séries temporais de variáveis hidrometeorológicas possui um papel importante no planejamento e operação de sistemas hídricos.

Ao longo do século passado, surgiram diversos modelos estocásticos, que tinham como objetivo principal gerar múltiplas sequências de variáveis hidrometeorológicas para que fosse possível avaliar o desempenho de sistemas hídricos via simulação, por exemplo os modelos Autoregressivos (AR). (Bras e Rodriguez-Iturbe, 1992; Loucks *et al.* 2005).

Tendo em vista a importância prática de modelos desse tipo, o campo do conhecimento referente à geração estocástica de variáveis hidrometeorológicas, em especial de vazão, é bastante extenso e inclui modelos com diversos níveis de complexidade, desde os mais simples, com propriedades estatísticas estacionárias para geração de vazões em um único

local, aos mais complexos, que incorporam alguma forma de não-estacionariedade da série, e que podem ser aplicados em situações onde é necessário gerar séries simultâneas em diversos locais, levando em consideração a estrutura de correlação espacial das vazões, e possivelmente alguma forma de relação com indicadores climáticos de larga escala, como o *El Niño*.

O principal objetivo desses modelos é gerar séries de variáveis hidrometeorológicas que sejam estatisticamente similares à série histórica, denominadas séries sintéticas. Frequentemente se diz que um modelo estocástico de geração de vazões é útil para gerar séries sintéticas de vazão se ele produz vazões com a mesma média, variância, coeficiente de assimetria, coeficiente de autocorrelação e/ou estrutura de correlação espacial das vazões observadas (Loucks *et al.* 2005). Mas essa é apenas uma interpretação possível de similaridade entre as séries geradas pelo modelo e a série observada. Outra interpretação, também válida, seria obter um modelo estocástico que reproduza de maneira fiel a distribuição marginal da vazões, no caso univariado, ou a distribuição conjunta, no caso multivariado.

A Figura 3.1 ilustra como séries sintéticas de vazão ou séries de outras variáveis hidrometeorológicas (chuva, evapotranspiração potencial e temperatura) são utilizadas, em conjunto com sequências de demandas futuras e de outras informações de caráter econômico, para avaliar o desempenho de sistemas hídricos quando simulados com diferentes políticas operativas e até mesmo com diferentes configurações de reservatórios. O emprego de séries sintéticas permite que se avalie o desempenho do sistema de forma mais ampla, possivelmente em condições extremas mais adversas do que aquelas encontradas na série histórica disponível, muito embora as múltiplas séries geradas possuam as mesmas características estatísticas da série histórica, permitindo que se obtenham sistemas hídricos e políticas operacionais mais robustos (Vogel e Shallcross, 1996).

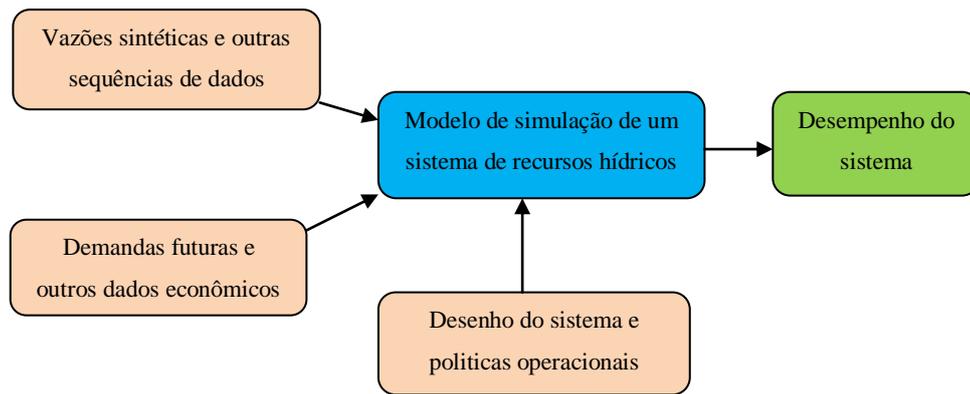


Figura 3.1 - Estrutura de um estudo de simulação, indicando a transformação de uma sequência de vazões sintéticas, demandas futuras, desenho e as políticas operacionais para produzir estatísticas do desempenho do sistema (Loucks *et al.* 2005, traduzido).

### 3.2 - BREVE HISTÓRICO

Segundo Salas *et al.* (1980), a modelagem estocástica de vazões se inicia com os primeiros estudos de Hazen (1914) e Sudler (1927) que mostraram como utilizar a análise estatística para analisar sequências de vazões de um rio. Mais tarde, Hurst (1951) investigou a longa série existente de vazões do rio Nilo e de outras séries de variáveis geofísicas para a elaboração do projeto da barragem de Aswan. Esse estudo produziu um impacto teórico e prático para analisar séries temporais de fenômenos físicos e hidrológicos. Em seguida, Barnes (1954) estendeu os estudos empíricos de Hazen e Sudler, e introduziu a ideia de geração sintética de vazão utilizando uma tabela de números aleatórios Normais.

No entanto, o anterior não foi útil até o começo da década dos 60's com os aportes de Box-Jenkins, ali foi que o desenvolvimento formal da modelagem estocástica começou com a introdução e aplicação dos modelos AR para vazões anuais e sazonais com Thomas e Fiering (1962) e Yevjevich (1963). Nesse mesmo período, Matalas (1967) com o intuito de expandir o modelo de Thomas e Fiering para regiões onde as vazões possuíam assimetria positiva, introduziu modificações do tipo logarítmico para obter os estimadores estocásticos do modelo original de Thomas e Fiering (Farias, 2003).

Entretanto, apenas na década de 1960 deu-se de fato o início formal da modelagem estocástica. Assim, muitos modelos com as anteriores considerações têm sido desenvolvidos para modelar séries temporais e aplicar-se na modelagem estocástica de vazões. Por conseguinte, existem os denominados Modelos Autoregressivos

(*Autoregressive Models-AR*) e os modelos Autoregressivos de Médias Móveis (*Autoregressive Moving Average - ARMA*), modelos que são mencionados como clássicos ou tradicionais (Thyer e Kuczera, 2000). E a partir desses modelos surgiram uma ampla variedade de modelos, como os trabalhos de Salas *et al.* (1980); Salas (1993); Bras e Rodriguez-Iturbe (1992); Marco *et al.* (1989), que fazem uma recopilação dos trabalhos e conceitos desses modelos aplicados na hidrologia e mencionam que o desenvolvimento histórico da modelagem de séries temporais, começa com os trabalhos na década dos 60's e 70's de Box-Jenkins.

Os modelos propostos por Box- Jenkins (1970) partem da ideia de que cada valor da série temporal pode ser explicado por valores prévios, a partir do uso da estrutura de correlação temporal que geralmente há entre os valores da série (Werner e Ribeiro, 2003).

A partir daí, surgiu o interesse de desenvolver novos modelos a partir do modelo básico AR que fossem justificados fisicamente, com o objetivo de serem utilizados para avaliar impactos nos recursos hídricos. Dessa forma, foram desenvolvidos muitos modelos e com eles surgiram muitas críticas sobre a não representação de dependências curtas e longas, dificuldade para calcular os parâmetros e incerteza neles e também falta de justificação física e a suposição de estacionariedade que nem sempre é válida (Salas *et al.*, 1980; Loucks *et al.*, 2005). Os detalhes dos modelos AR e ARMA, assim como outros modelos derivados deles, que têm sido utilizados na modelagem hidrológica, são tratados nas seguintes seções.

### **3.3 - PERSISTÊNCIA EM SÉRIES TEMPORAIS HIDROLÓGICAS**

Em muitas ocasiões, as séries temporais de observações hidrológicas apresentam variações entre períodos ou estados climáticos. Variações que em alguns casos ocorriam ao redor de uma estatística (média, mediana, etc.). Assim, dado um limiar (*threshold*) específico, é possível determinar em uma série histórica períodos ou estados climáticos, por exemplo, em uma série de vazões observadas, tomando como limiar à média, pode-se definir como períodos secos à sequência consecutiva de observações abaixo da média. Os períodos úmidos serão então definidos como os períodos onde as observações se encontrem acima da média (Sveinsson *et al.* 2003).

Contabilizar o tamanho dos comprimentos desses períodos é de interesse em varias áreas. Na hidrologia, por exemplo, é importante analisar essa persistência, a qual se denomina: persistência hidrológica. Sendo útil a sua análise para ajustar modelos estocásticos e gerar dados que apresentem a variação observada entre os estados climáticos (períodos secos e úmidos) ou a persistência desses estados. Segundo Whiting (2006), a persistência hidrológica em séries de precipitação e vazões é o resultado da influencia de variáveis atmosféricas ou armazenamento sub-superficial na bacia. Persistência que repercute principalmente no desenho e no gerenciamento de infraestrutura de abastecimento de água. Nesse aspecto, Douglas *et al.* (2002) mencionam que uma longa vida útil de um sistema de abastecimento de água é possível de se obter quando a persistência das vazões é levada em consideração na etapa de desenho. Com tudo isso, a análise da persistência hidrológica ajuda a desenvolver modelos estatísticos que representem as observações históricas para se aproximar melhor à realidade que sirvam de suporte à gestão dos recursos hídricos (Whiting, 2006).

Vários modelos estocásticos têm sido propostos na literatura para simular a persistência hidrológica, entre eles estão os modelos tradicionais ARMA, *Fractional ARMA* (FARMA), *Fractional Gaussian Noise* e *Broken Line* que são úteis para representar variabilidade de longo prazo e podem produzir mudanças aparentes de estados em séries climáticas e hidro-climáticas (Salas, 1993). Em adição a esses modelos estão os de Média Móvel (*Shifting Mean Models*), propostos por Boes e Salas (1978) e Salas e Boes (1980) e também utilizados por Sveinsson *et al.* (2003). Os Modelos de Markov Oculto (*Hidden Markov Models-HMM*), capazes de gerar padrões de variabilidade (Sveinsson *et al.*, 2003), foram utilizados por Zuchinni e Guttorp (1991), Thyer e Kuczera (2000), Robertson *et al.* (2004) e Lima (2010).

Existem vários trabalhos relatados na literatura que buscam estudar a persistência hidrológica com o objetivo de produzir simulações de comprimentos de períodos climáticos utilizando alguns desses modelos estocásticos. Entre os primeiros trabalhos está o realizado por Saldarriaga e Yevjevich (1970). Eles analisaram as mudanças repentinas nas vazões anuais no Rio Colorado nos Estados Unidos, no exutório da parte superior da bacia do Rio Colorado denominado Lees Ferry, um local no Rio Colorado que recebe as águas da parte superior da bacia desse rio, investigando várias propriedades dos comprimentos dos períodos climáticos com a hipótese que esses comprimentos são

propriedades das series hidrológicas. Com o objetivo de calcular períodos secos e úmidos utilizando modelos AR(1), eles demonstraram a utilidade desse tipo de modelos para essas análises.

Do mesmo modo, Sveinsson *et al.* (2003) analisaram séries climáticas e hidro-climáticas que apresentam mudanças abruptas ou repentinas entre estados. Propuseram modelos de medias mudáveis (*Shifting Mean Models*) com a intenção de capturar essa variabilidade entre estados e tiveram sucesso em representar a dinâmica dessas variações aplicando os modelos propostos a uma série do índice climático de Oscilação Decenal do Pacífico (*Pacific Decadal Oscillation- PDO*) e a série de vazões médias anuais do Rio Níger, na África. Obtiveram uma boa representação para essas séries onde a média observada mudava ao longo do tempo. Eles compararam seus modelos com o modelo ARMA analisando a representação das estatísticas observada das simulações de séries geradas por esses modelos e utilizando estatísticas que definiam períodos secos e úmidos propostos a partir de um limiar definido pela média observada e níveis de demanda. Também analisaram a capacidade de armazenamento e o período de retorno de secas severas e demonstraram a aplicabilidade desses modelos para a geração de cenários realísticos de períodos secos e úmidos, com menor incerteza e variabilidade no valor de estatísticas e métricas de desempenho dos modelos frente aos modelos ARMA.

Prairie *et al.* (2008) tentaram propor uma metodologia para analisar esse tipo de persistência analisando a série de vazões anuais do Rio Colorado, no ponto Lees Ferry. A metodologia incluía juntar os pontos fortes de dois métodos utilizados na análise de séries temporais, eles são a Paleo reconstrução de dados por meio de anéis das árvores e séries observadas. Formando assim uma metodologia que combina a longa informação das vazões paleo reconstruídas com a curta, mas confiável, informação das vazões observadas. Os autores utilizaram a informação paleo construída para simular os estados hidrológicos por meio de um modelo que incluía uma Cadeia de Markov Não Homogênea, e logo as magnitudes das vazões simuladas foram calculadas utilizando uma série temporal (*K-Nearest Neighbor (K-NN) bootstrap*) e condicionadas ao estado hidrológico prévio e à vazão previa. Assim, obtiveram que essa metodologia apresentou uma grande habilidade para uma ampla variedade de sequências úmidas e secas. Utilizaram também, várias estatísticas a partir de um limiar que definiu os períodos climáticos (foi utilizada a mediana da série observada para tal definição), comparando essas estatísticas definidas com as

mesmas estatísticas observadas. Além disso, também foi analisada a série de vazões para avaliar a capacidade de armazenamento num reservatório frente a vários níveis de demanda.

Por último, informação detalhada sobre conceitos, estudos e modelos utilizados na análise de persistência hidrológica e em especial a análise de secas é oferecida por Mishra e Singh (2011). Existem diversos estudos além dos já citados, mas, em especial, aqueles que se utilizam dos modelos de Markov oculto são destacados no capítulo quatro.

## 4 - REVISÃO DE LITERATURA

### 4.1 - APLICAÇÕES DOS MODELOS DE MARKOV OCULTO NA MODELAGEM HIDROLÓGICA

#### 4.1.1 - Aplicações dos Modelos Homogêneos de Markov com Estados Ocultos (HMM)

Uma das primeiras aplicações de modelos HMM para simular séries sintéticas de vazões anuais foi realizada por Jackson (1975). O foco do estudo foi obter um modelo, então denominado “*Markov Mixture Model*”, que fosse capaz de simular longos períodos de anos secos. Um modelo com dois estados foi analisado, um para vazões normais e outros para vazões baixas, aplicando-se a distribuição Normal. O método foi capaz de simular sequências de anos seco mais longas do que o tradicional modelo de Thomas-Fiering (1962).

Mais tarde, Thyer e Kuczera (2000) utilizaram um HMM para simular a persistência de períodos secos e úmidos, porém eles aplicaram o modelo tanto para séries de vazões anuais quanto para precipitação anual. As análises foram realizadas em quatro regiões da Austrália. No caso da vazão, as séries foram empregadas na escala logarítmica para que fosse possível empregar a distribuição Normal para os dois estados analisados (seco e úmido). Eles assumiram que as probabilidades de transição de estados eram estacionárias (HMM homogêneo) e trabalharam sob a hipótese de que os eventos secos e úmidos estão vinculados a fenômenos climáticos globais como o ENSO, a *Interdecadal Pacific Oscillation* (IPO) e o Vórtice Circumpolar, ou eram influenciados por sistemas climáticos do pacífico tropical (White, 2006). A estimativa dos parâmetros do modelo foi realizada sob uma abordagem Bayesiana empregando o amostrados Gibbs. Os autores concluíram que os HMMs obtiveram sucesso na representação da persistência dos estados climáticos quando comparado com os resultados obtidos pelo modelo AR(1), que era utilizado na época para estudos desempenho de sistemas hídricos.

Thyer e Kuczera (2003a e b) estendem o trabalho realizado anteriormente focando no ajuste de modelos HMM para simulação de séries de precipitação em múltiplos locais empregando análise Bayesiana. Foi dada ênfase aos aspectos de estimativa de parâmetros

utilizando uma análise Bayesiana, com foco na determinação das distribuições a priori e de como lidar com valores faltantes no histórico. A metodologia foi aplicada a três bacias hidrográficas localizadas na costa leste Australiana, que servem como manancial de abastecimento. Os resultados mostraram que, para duas das três regiões, o modelo com dois estados é mais provável do que o de apenas um estado sem qualquer persistência. Os resultados para a terceira região foram inconclusivos, pois se obteve uma distribuição a posterior bimodal para o tempo médio de residência de estados e os autores chamam a atenção sobre a necessidade de entender melhor como identificar o grau de persistência (homogênea) numa região. De fato, as bacias estudadas eram melhor representadas por modelos com dois estados.

Thyer e Kuczera (2003b, 2006), então, estendem o trabalho anterior testando o comprimento necessário para um melhor ajuste dos HMM, o cálculo dos parâmetros desses modelos e como empregar dados provenientes de múltiplas locais de medição. Um dos resultados desses trabalhos é que para um maior comprimento das séries analisadas tem-se um melhor ajuste do modelo e uma maior representação da persistência de estados climáticos.

Um trabalho parecido ao de Thyer e Kuczera (2000) foi realizado por Akintug e Rasmussen (2005), utilizando um HMM de  $m$  estados ao que chamaram “*Markov Switching Model*” de ordem  $m$ . Eles analisaram uma série anual de vazões, e ajustaram modelos HMM, para depois compará-los com vários modelos ARMA conseguindo bons resultados na hora de simular métricas persistência de períodos secos, da capacidade de armazenamento necessária num reservatório, onde, obtiveram métricas que se aproximaram a valores observados. Eles, também, derivaram uma expressão para a função de autocorrelação (ACF) dos HMM, tendo por conta o trabalho de Timmermann (2000). Mencionaram o que também foi ressaltado por Zucchini e MacDonald (2009) a respeito da ACF dos HMM, citando que essa função tem a mesma estrutura da ACF dos modelos ARMA ( $m-1, m-1$ ). Logo, pode-se encontrar o modelo ARMA que tenha exatamente a mesma ACF de um modelo HMM de  $m$  estados e que os HMM são úteis quando os dados observados tem uma forte autocorrelação, inclusive para valores de defasagem (*Lags*) altos comparados com os modelos ARMA. Mas a deficiência dos HMM está na hora do ajuste da distribuição marginal dos dados observados e da incerteza nos parâmetros, como

quando se tem uma série curta e o número de parâmetros é alto, visto que, para um aumento de estados do modelo, tem-se um aumento nos parâmetros que se devem calcular.

A mesma linha de análises de persistência utilizando HMM seguiu Whiting (2006), que identificou a persistência e variabilidade de períodos secos e úmidos. O autor analisou a magnitude desses períodos climáticos em séries hidrológicas de vazões e precipitação na escala anual e mensal empregando séries de precipitação de múltiplas estações pluviométricas ao longo da Austrália e séries de vazões de bacias australianas com forte variabilidade climática, como as regiões áridas. A persistência foi então relacionada com fenômenos climáticos como o ENSO e o DIPOLO do Oceano Índico (*Indian Dipole Index-IDO*). Devido às fortes correlações de índices climáticos pertencentes a esses fenômenos com as series de precipitação e vazão notou-se que a maior influencia sobre a persistência sobre as vazões a exerce o ENSO. Portanto, esse trabalho demonstrou que os modelos HMM são uma forte ferramenta para a análise da persistência de estados climáticos por representarem a variabilidade e a persistência de estados secos e úmidos, inclusive por acima dos modelos ARMA; e simularem séries com períodos correspondentes aos registros históricos.

Whiting, (2006) também mencionou que a deficiência dos HMM está ao assumir as distribuições de probabilidade dos dados, ele assumiu que os dados distribuam-se normalmente para os dados anuais como nos trabalhos de Thyer e Kuczera (2000) e Akintug e Rasmussen (2005) e na escala mensal ele utilizou um HMM com distribuições Log-Normais. Suposições estatísticas que deram problemas na hora de ajustar a distribuição de mistura. O mesmo problema de ajuste mencionado por Akintug e Rasmussen (2005) e que se torna mais relevante quando os dados apresentam assimetria positiva significativa. Dessa forma, o autor propôs um Modelo Não Paramétrico de Markov Oculto (*Non Parametric Hidden Markov Model - NP HMM*) para suprir as deficiências dos HMM básicos. Mas, todos esses problemas os HMM utilizados demonstram um claro sucesso na hora de representar a persistência obtendo-se melhores simulações (comprimentos de períodos mais longos) que os modelos ARMA.

No caso da simulação e previsão de vazões, Fortin *et al.* (2004) reformularam o “*Shifting Level Model*” (*SL Model*) de Salas e Boes (1980) configurando um HMM que utilizaram

para simular a persistência e fazer previsões de vazões anuais. O modelo “*SL Bayesiano*”, nome dado ao modelo proposto, foi comparado aos modelos ARMA. O *SL Model* não apresentou o sucesso esperado na hora das previsões pontuais, com maiores erros na previsão (*Bias*) em relação aos modelos ARMA, pois além do *SL Model* ser um modelo não linear, a série observada não era estacionária. Logo, os modelos lineares ARMA apresentaram um melhor desempenho para a previsão em concordância ao mencionado por Ramsey (1996) e Dacco e Satchell (1999) em relação aos *SL Models*. No entanto, deve-se informar que o *SL Model* apresentou melhores previsões condicionais, não descartando que esse modelo possa ser útil para prever eventos extremos e que consiga simular as variações na média da série observada.

Do mesmo modo, Bracken (2011) utilizou modelos HMM para simular a variabilidade de estados na série de vazões anuais do Rio Colorado, em Lees Ferry. Uma boa representação da série analisada foi obtida através desses modelos comparado ao desempenho do modelo AR(1), demonstrando novamente que os HMM são úteis para representar a ocorrência de longos períodos de estados climáticos e simular bem as densidades de probabilidade desses períodos e o histórico dos comprimentos. Os melhores resultados obtidos aconteceram quando empregou-se um modelo com distribuições dependentes Gama. Além disso, os HMM foram capazes de simular comprimentos de maior magnitude que os modelos AR(1), embora tenha apresentado deficiência na hora de capturar a autocorrelação. Isso não foi impedimento para recomendar o uso dos HMM, porque a série utilizada tinha uma autocorrelação muito baixa, dada principalmente pela variação de regimes climáticos.

Esse autor também relata o uso dos HMM para fazer previsões, sendo melhor, nesse caso, um HMM com distribuições dependentes Normais, frente ao outro HMM e, sobretudo, frente ao modelo AR(1). É mencionado inclusive que os HMM prevêm melhor os períodos úmidos e os modelos AR(1) os períodos secos. Esse trabalho também realizou a decodificação global que mostrou que é possível representar o sistema (série de vazões anuais), sugerindo que não só é suficiente capturar a ACF, como é preciso representar a variação de estados, além da persistência.

A partir dessas considerações, conclui-se que os HMM têm sido utilizados em vários estudos com grande sucesso. Mas, apresentam várias deficiências especificamente na

suposição das distribuições dependentes o que leva ao fato de que essas distribuições não se ajustam adequadamente a uma distribuição de mistura e ainda existem deficiências na hora de representar a autocorrelação dos dados observados. Pelo que se propõem modelos diferentes aos HMM básicos, para suprir as deficiências, como os HMM não paramétricos.

#### **4.1.2 - Aplicações dos Modelos de Markov com Estados Ocultos Não Homogêneos**

Muitos estudos sobre as interações entre oceano e a atmosfera têm mostrado que os processos hidro-climáticos exibem uma forte variação de padrões ou estados climáticos, (Yonetani e Gordon, 2001; Rial *et al.*, 2004; Sveinsson *et al.*, 2003; Schwing *et al.*, 2003; Fortin *et al.*, 2004). Na região do Nordeste Brasileiro (NEB) há diversos autores que trazem trabalhos sobre a influência dos padrões espaço-temporais dessas interações oceano-atmosfera com a precipitação e as vazões, assim como mencionado anteriormente e que é aprofundado no capítulo 6.

Entre as recomendações sugeridas por muitos trabalhos que utilizaram os modelos HMM homogêneos para a geração de séries sintéticas de vazões, existe a sugestão de incluir, na análise dessas séries, variáveis que influenciam o ciclo hidrológico, como as variáveis atmosféricas. A influência de padrões atmosféricos globais na escala hidrológica local foi recomendada por Thyer e Kuczera (2000) que sugeriram utilizar índices climáticos na modelagem da persistência para que as probabilidades de transição sejam função desses índices e a persistência seja função de alguma variável climática persistente como as temperaturas da superfície do mar (*Surface Sea Temperatures-SSTs*). Whiting (2006) recomendou utilizar índices climáticos que representam mudanças na persistência de padrões climáticos e relacionar essa persistência com variáveis hidrológicas. De fato, ele encontrou altas correlações entre o NINO3 e as vazões na Austrália. Considerando todas as informações acima, observa-se que na literatura existem vários trabalhos que utilizaram NHMM aplicados à hidrologia, relacionando os modelos HMM com variáveis climatológicas.

Os NHMM foram inicialmente utilizados Hughes e Guttorp (1994) para modelar a ocorrência de chuva. Nesse trabalho, um NHMM permitiu relacionar variáveis atmosféricas à ocorrência da chuva em uma densa rede de estações de medição. Os autores mencionaram que o principal atrativo do NHMM é que os estados ocultos corresponderam

a estados climáticos influenciados por variáveis atmosféricas. Esse trabalho foi similar ao de Hughes et al. (1999), que ajustaram um NHMM a uma parte dos dados observados e depois avaliaram o desempenho do modelo com os dados restantes, apresentando, novamente o sucesso do NHMM na modelação da ocorrência de chuva, assim como no trabalho de Hughes e Guttorp (1994). Charles *et al.* (1999) continuaram o trabalho de Hughes *et al.* (1999) utilizando um NHMM como ferramenta de redução de escala estatística (*Statistical Downscaling- SD*) e considerando mudança climática encontrando a mesma utilidade do modelo no campo do SD.

Na região do NEB, Robertson *et al.* (2004), baseado no trabalho de Robertson e Smyth (2003), simularam a ocorrência da precipitação diária em 10 estações, ajustando vários HMM. Escolheram o melhor número de estados por meio de Validação Cruzada e *Bayes Information Criterion* (BIC). Com o modelo escolhido, realizaram simulação da persistência e a correlação espacial observada entre estações e interpretaram fisicamente os estados ocultos, de forma que os estados mais bem definidos (um estado úmido e um seco) foram relacionados com padrões atmosféricos da região como a ZCIT (o seu deslocamento norte-sul no Atlântico) e o SAMS (*South American Monsoon System*) (expansões e contrações na região central da Amazônia). Os autores encontraram relação com os padrões de ventos do reanalises NCEP–NCAR e a interpolada *outgoing longwave radiation* (OLR) e que os estados úmido e seco estão mais relacionados a teleconexões com o ENSO e oscilação do Atlântico tropical. Os autores encontraram que os anos com períodos úmidos estão mais relacionados estatisticamente com *La Niña* e os anos com períodos secos com *El Niño*, sendo esses fenômenos grandes forçantes na escala interanual. Mencionaram que os estados com um comportamento menos marcado encontraram-se mais relacionados a os fatores do OLR e os padrões dos ventos no trópico e modulam a ocorrência da precipitação de norte a sul na região. Com tudo isso, a partir das simulações realizadas, foi demonstrada a utilidade do HMM para simular a precipitação diária em cada estação e a sua variabilidade intrasazonal e interanual, representando com bastante sucesso os comprimentos de períodos secos e úmidos, em especial os períodos secos de 10 ou mais dias sem chuva. Mas possui uma tendência a subestimar esses comprimentos, sendo recomendado um modelo AR-HMM.

Depois de realizar a análise de estados, Robertson *et al.* (2004) ajustaram um NHMM, para fazer *Statistical Downscaling - SD* utilizando simulações de vários modelos de circulação geral (*general circulation model - GCM*), como preditores (*inputs*). A escolha desses *inputs* foi realizada fazendo uma Análise de Correlação Canônica (*Canonical Correlation Analyses - CCA*) e conservaram-se os estados ocultos do HMM, pois, uma análise (*k-means clustering*) paralela; demonstrou que a existência desses estados era inerente aos dados. O NHMM representou, novamente, consistentemente a variabilidade da chuva, inclusive obtendo correlação espacial entre varias estações, demonstrando ser uma ferramenta útil para o SD. Os autores também recomendaram, para um melhor ajuste, o uso de modelos mais sofisticados, como o de Hughes Guttorp (1994) os modelos *Chow-Li* relatados por (Kirshner, 2005a).

Continuando com a aplicação dos NHMM, Robertson *et al.* (2005) analisaram a variabilidade da precipitação diária durante o verão de uma rede estações na Austrália. Visando relacionar essa variabilidade com a circulação atmosférica nas escalas diária e decenal, por meio da análise da variabilidade da Monção australiana. Eles ajustaram primeiramente um HMM, relacionando cada estado oculto com padrões de circulação atmosféricos, em que a variação dos estados representou a variação da Monção, nas diferentes escalas. Na escala interanual, encontrou-se forte relação dessa variação com o ENSO e SSTs do oceano Índico. Depois, foi ajustado um NHMM utilizando preditores obtidos de várias fontes climáticas como o Reanalises, GCM e SSTs históricas. Conseguindo um sucesso na simulação (SD), da intensidade de chuva diária, na quantidade e ocorrência junto com as variações interanuais. Nesse trabalho, não foram simuladas séries para períodos futuros. No entanto, recomendou-se o emprego dos NHMM para um análise do clima futuro, trabalho que foi realizado por Greene *et al.* (2011).

Mehrotra e Sharma (2005), fizeram uma análise da ocorrência da precipitação diária em várias estações de medição na Austrália. Utilizando um NHMM não paramétrico como ferramenta de SD. Esse modelo foi baseado em estados climáticos ocultos contínuos e utilizou informação atmosférica. Diferente dos modelos paramétricos que utilizam estados discretos, esse modelo utilizou estados definidos por 3 variáveis de estado baseadas na distribuição espacial da precipitação e as coordenadas das estações de medição a respeito de uma origem imposta. O modelo foi baseado no modelo de Mehrotra *et al.* (2004) que propôs um método de SD que inclui uma reamostragem *K-Nearest-Neighbor* (KNN).

Portanto, o NHMM incluiu esses dois aspectos: a reamostragem e os estados contínuos. O modelo obteve um notável sucesso frente ao paramétrico NHMM e ao modelo de Mehrotra *et al.* (2004). Ressaltando a importância de incluir estados contínuos. Uma estatística significativa desse trabalho, que representou bem, foi o comprimento máximo de períodos secos e úmidos, sendo nesse trabalho, os comprimentos, o número de dias de ocorrência de chuva consecutivos em um período.

Igualmente, Lima (2010) analisou a série anual de precipitação de Fortaleza, ajustando um NHMM para simular a ocorrência de: clusters de períodos secos (precipitação abaixo do percentil 33%), chuvosos (precipitação acima do percentil 66%) e normais com valores de precipitação dentre os anteriores períodos. O autor analisou a relação de índices climáticos com os períodos secos, períodos sinalados de maior interesse no planejamento e gerenciamento dos recursos hídricos (Thyer e Kuczera, 2000), e demonstrou a forte influência dos índices climáticos DIPOLO do atlântico e NINO3, na ocorrência de eventos secos e úmidos. Principalmente o DIPOLO do Atlântico apresentou forte influência sobre os eventos secos. Demonstrou, também, que utilizando um NHMM, obtêm-se simulações coerentes com os valores históricos de períodos secos, ressaltando-se que a maior ocorrência desses períodos ocorre com valores positivos do DIPOLO do atlântico.

Greene *et al.* (2008) analisaram a variabilidade da precipitação de várias estações de medição na época de Monção na Índia. Demonstrou-se o sucesso dos modelos HMM para representar a variação espaço-temporal monçônica. Comprovado por método de Validação Cruzada entre as observações e previsões realizadas por um HMM. Baseado nesse trabalho e seguindo a recomendação de utilizar o HMM para SD de projeções advindas de GCM, Greene *et al.* (2011), realizaram um trabalho similar ao de Robertson *et al.* (2005). Onde, propuseram um NHMM, para analisar as mudanças potenciais na precipitação diária monçônica para o período 2070-2099. Utilizando cenários de desenvolvimento futuro e projeções de GCM. O preditor, no NHMM foi uma média de um índice dinâmico de circulação atmosférica na Moção, produto de um conjunto de projeções de GCM e escolheu-se pela forte correlação com a precipitação monçônica. Assim, obteve-se um sucesso da representação da variação espaço-temporal da precipitação em várias escalas, de acordo as observações. Além disso, mostrou-se que uma diminuição na circulação oferecida pelo preditor, levará a um aumento de: os dias secos, o comprimento médio de

períodos secos, a intensidade média de chuva diária e a excedência na frequência de quase todos os percentis diários de chuva, em muitas localidades para o período de análises.

Um estudo recente, da ocorrência da precipitação diária para múltiplas estações, Fu *et al.* (2012), semelhante a Greene *et al.* (2011), analisaram mudanças da precipitação utilizando projeções futuras do clima na Austrália. encontrando resultados parecidos a Robertson *et al.* (2005) e obtendo uma diminuição na chuva anual e no período do inverno. Aumento de: dias secos e máximos dias secos consecutivos, diminuição de dias úmidos e máximos dias úmidos consecutivos. Ressalta-se que esses efeitos não são desejáveis em uma bacia por diminuir a disponibilidade dos recursos hídricos.

Em relação a simulações de vazões utilizando NHMM, Whiting *et al.* (2004) ajustaram um HMM, um HMM hierárquico (HHMM), e um NHMM para representar a variabilidade e persistência de séries de precipitação e vazões observadas na Austrália, nas escalas anual e mensal. Utilizaram também uma forma não paramétrica, desses modelos, em vista das recomendações de Lambert *et al.* (2003). Encontraram que o modelo HHMM, um modelo que condiciona estados climáticos anuais e mensais, representou o comportamento dos estados climáticos observados, por valores altos do índice SSI (State Sinal Index) de Thyer e Kuczera(2000), inclusive em bacias de grande área e onde a persistência joga um papel importante. No entanto, mencionaram que o ajuste dos modelos NHMM e HHMM requer muitos parâmetros, comparados com o HMM. Além disso, o HMM representa um bom ajuste às séries de vazões e precipitação.

Diante do que foi exposto, existem um notável número de trabalhos relacionados à análise da ocorrência da precipitação, que apresentam sucessos do uso dos NHMM, assim como em versões não paramétricas desse modelo, como o modelo de Mehrotra e Sharma (2005). Porém, os trabalhos utilizando o NHMM para a modelagem de vazões são escassos, notando que o trabalho Whiting *et al.* (2004) relacionado com essa modelagem apresentou resultados não tão satisfatórios aos esperados, menor sucesso na representação de períodos seco, onde por resultar sobre parametrizado recomendou-se um HMM básico.

Como resumo do tratado nessa seção 4.1, a Tabela 4.1 e Tabela 4.2 apresentam, respectivamente, informações adicionais sobre a forma dos modelos estocásticos, dos trabalhos relatados na literatura que incluíram HMM e NHMM. Nessas tabelas, as

seguintes siglas representam as distribuições utilizadas: N = Normal, L = Log-Normal, G = Gama, Ex = exponencial, Ge = Geométrica, D = Delta, MBAM = Modelo binário autolístico multivariado. A sigla LT significa Log-transformação, que é o tipo de transformação nos dados e NP significa Não Paramétrico, com respeito à forma do modelo; M de Multivariado e U para Univariado. PMNM trata da pressão média ao nível do mar, AGP é a altura geopotencial, DTPO é a depressão da temperatura do ponto de orvalho e TTAP é o teor total de água precipitável, que representam as variáveis externas. E, por ultimo, P é a Precipitação; KNN é o *K-nearest-neighbor*; DEP é a distribuição espacial da precipitação e HHMM é o HMM hierárquico. Ainda, na Tabela 4.2, os trabalhos marcados com (\*\*) empregaram, para ajustar e rodar as simulações dos modelos utilizados, o *software MultiVariate Nonhomogeneous Hidden Markov Model toolbox (MVN-HMM toolbox)*. Observa-se a ampla utilização desse código, o qual é utilizado nesta dissertação. Para finalizar, vale ressaltar que os trabalhos de Robertson *et al.* (2005) e Fu *et al.* (2012) encontram-se em detalhe no trabalho de Kirshner (2005). Além disso, em Kirshner (2005) se encontram as formulações e desenvolvimento dos HMM em todas suas vertentes e as descrições dos algoritmos utilizados no MVN-HMM *toolbox*.

Tabela 4.1 - Características dos modelos estocásticos nos trabalhos relatados na literatura que utilizaram HMM (*m estados*) para a modelagem de séries de vazões.

Trabalho	Local	Modelos utilizados	Distribuições dependentes nos HMM	Transformação dos dados	Modelos sugeridos
Jackson, (1975)	Rio Quaboag (Estados Unidos)	HMM (2), AR(1)	N	LT	Nenhum
Thyer e Kuczera, (2000)	Rio Burdekin (Austrália)	HMM (2), AR(1)	N	LT (HMM) e Box-Cox (AR(1))	HMM com informação climática
Fortin <i>et al.</i> , (2004)	Rio Senegal (África)	SL (2 e 3 média variável), AR(2) e ARMA(1,1)	N (variância constante e média variável)	Nenhuma	Modelos SL NP
Akintug e Rasmussen, (2005)	Rio Niágara (América do Norte)	HMM (1, 2 e 3); ARMA(1,1), ARMA(2,2)	N	Box-Cox e escalamento (ARMA(2,2)), Box -Cox (HMM e ARMA(1,1))	HMM com L e AR-HMM
Whiting, (2006)	Várias bacias na Austrália	HMM (2), AR(3), NP HMM (2)	N e L	Nenhuma	NP HMM com informação climática e AR-HMM
Bracken, (2011)	Rio Colorado (Estados Unidos)	HMM (2), AR(1)	N e G	LT (HMM com L)	Nenhum

Tabela 4.2 - Características dos modelos estocásticos nos trabalhos relatados na literatura que utilizaram NHMM (*m estados*)

Trabalho	Local	Modelos	Distribuições dependentes nos NHMM	Variáveis externas	Modelos sugeridos
Hughes e Guttorp, (1994)	Oeste de Washington (Estados Unidos)	NHMM (5) M	MBAM	PMNM, AGP	NHMM com N ou G
Hughes <i>et al.</i> , (1999)	Sudoeste da Austrália	NHMM (6) M	MBAM	PMNM, AGP	NHMM com maior número de estados e variáveis atmosféricas
Charles <i>et al.</i> , (1999)	Sudoeste da Austrália	NHMM (6) M	MBAM	PMNM, AGP, DTPO	Ampliar a escala do modelo
Robertson <i>et al.</i> , (2004)**	Ceará (Brasil)	HMM (4) e NHMM (4) M	Assumida como Geométrica	Simulações de GCM	NHMM sazonal, com SST, AR-HMM
Robertson <i>et al.</i> , (2005)**	Norte de Queensland (Austrália)	HMM (5) NHMM (5) M	D (P = 0) Mistura de 2 Ex (P > 0)	Simulações de GCM	AR-HMM, NHMM sazonal
Mehrotra e Sharma, (2005)	Sidney (Austrália)	KNN, NHMM (3) NP baseado no KNN, NHMM (6)	DEP no NHMM NP, MBAM no NHMM	PMNM, AGP, TTAP	NHMM NP maior número de estados Melhoras no KNN
Lima, (2010)**	Fortaleza (Brasil)	HMM (2), NHMM (4) U	Bernoulli	DIPOLo do Atlântico e NINO3	NHMM para modelar totais anuais e NHMM M
Greene <i>et al.</i> , (2011)**	Centro oeste da Índia	NHMM (4) M	D se P = 0 Mistura de duas Ex se P > 0	Western Shear Index (WSI11)	Projeções futuras escala local
Fu <i>et al.</i> , (2012)**	Sudeste da Austrália	NHMM (5) M	MBAM	PMNM, AGP, DTPO	Maior número de cenários e projeções futuras para avaliar uma potencial mudança climática
Whiting <i>et al.</i> , (2004) (Vazões)	Bacia Murray-Darling (Austrália)	HMM/NP (2), NHMM/NP (2), HHMM/NP (2) U	N para os modelos paramétricos	Nenhuma	NHMM sazonal

#### 4.2 - MODELAGEM DE VAZÕES NA REGIÃO DO NORDESTE DO BRASIL

Diversos trabalhos foram realizados para modelar séries de vazões na região do Nordeste do Brasil (NEB). Por exemplo, Sarmiento (1989) realizou um estudo de diversos modelos de geração de vazões anuais e os aplicou a dois rios intermitentes no Estado do Ceará, o rio

Acaraú e o rio Bastiões. O autor gerou várias séries sintéticas e fez a comparação dessas séries geradas com a série histórica mediante o emprego dos parâmetros estatísticos: média, desvio padrão, coeficiente de variação, coeficiente de assimetria, coeficiente de correlação e coeficiente de autocorrelação. Encontrou-se uma boa representação dessas estatísticas, inclusive ao se fazer uma desagregação para a escala mensal. O único problema encontrado foi a não boa preservação da autocorrelação (Farias, 2003).

Campos (1996) fez a geração sintética de vazões anuais para alguns rios intermitentes do NEB que apresentam uma longa estação de vazão nula seguida de uma estação úmida. Com essas características, Campos supôs que as vazões eram serialmente independentes, e então, propôs que as séries de vazões anuais podem ser obtidas pela função de densidade de probabilidade Gama de dois parâmetros, gerando números aleatórios com essa distribuição, tomando como base o que foi feito por Araújo (1991), que assegurou que esse modelo era útil para gerar vazões anuais e mensais na região do semiárido nordestino. Campos (1996) também considerou útil a distribuição Log-Normal de dois parâmetros para essa modelagem (Santos, 2009).

Essa metodologia, usada para gerar séries de vazões utilizando a distribuição gama de dois parâmetros, é bastante utilizada na região nordestina, pois trabalhos como os de Cysne (2007); Filho (2007); Santos (2009) a utilizaram como ferramenta para gerar vazões de entrada para a simulação de reservatórios. Cysne (2007) utilizou essa simulação para avaliar impactos de mudanças climáticas; Filho (2007) analisou a sustentabilidade de vários reservatórios calculando índices de sustentabilidade baseados no desempenho do sistema hídrico e Santos (2009) utilizou essas simulações para dimensionar reservatórios.

Um trabalho incluindo duas variações do modelo de Thomas-Fiering (1962) foi realizado por Farias (2003), no qual se utilizaram esses modelos e a geração de números aleatórios com as distribuições Gama com dois parâmetros. Concluindo que a geração utilizando a distribuição Gama forneceu melhores resultados e que nem sempre a escolha da metodologia com maior sofisticação matemática significa a melhor reprodução da realidade dos dados estudados.

Recentemente, alguns trabalhos procuraram incorporar informação climática na modelagem das vazões na região. Souza Filho e Lall (2003) realizaram previsão de vazões

utilizando uma abordagem semi-paramétrica em vários locais de medição em uma rede de rios (incluindo vazões afluentes ao reservatório Orós, Ceará) condicionados a preditores climáticos na região. A estratégia considerou previsões estatísticas de vazões anuais e/ou sazonais em todos os locais e sua desagregação mensal usando um k-vizinho mais próximo para reamostragem da aproximação (*K-Nearest Neighbor Resampling*), que mantém a consistência de espaço e tempo em diferentes locais e subperíodos. Essa metodologia forneceu uma boa capacidade de reamostragem condicionada aos registros históricos, considerando o estado dos principais preditores climáticos.

Nesse caso, foram utilizados os índices climáticos relacionados ao fenômeno *El Niño-Southern Oscillation* (ENSO) e anomalias das temperaturas da superfície do mar (*Surface Sea Temperatures-SSTs*) no Oceano Atlântico para avaliar a correlação espacial entre as observações dos locais e a correlação das observações com esses índices climáticos. Os resultados das previsões apresentaram uma boa correlação entre as previsões e as observações (correlação de 0.9). Concluiu-se que a metodologia é útil para auxiliar a tomada de decisões no gerenciamento de recursos hídricos e tem a vantagem de incluir informação climática.

Ribeiro (2011) utilizou informação climática para a previsão de vazões sazonais/interanuais no estado do Ceará e em vários reservatórios do NEB, incluindo aqueles reservatórios que estarão envolvidos involucrados na transposição do rio São Francisco. O estudo avaliou separado o reservatório Orós e o sistema de vários reservatórios que abastecem a região metropolitana de Fortaleza, os outros reservatórios foram analisados em conjunto. A metodologia teve vários modelos baseados no modelo de Souza Filho e Lall (2003), variando os tipos de regressão (*linear, stepwise, pool*) e como preditores os ventos meridionais e altura geopotencial de 1000mb, por correlacionar-se com as vazões e incluíram os índices do trabalho de Souza Filho e Lall (2003). Esse trabalho, além de fazer previsões, realizou a simulação dos reservatórios estudados, comprovando que a utilização da informação climática em modelos de previsão conseguem representar bem a variabilidade das vazões da região, apoiando a aplicabilidade desses modelos. Ademais, ressaltou-se que é importante fazer previsões sazonais porque dão uma melhor ideia do desempenho dos reservatórios em períodos de anos.

Além dos dois anteriores trabalhos, Alexandre (2012) utilizou modelos autoregressivos com a inclusão de variáveis exógenas analisando as vazões afluentes aos reservatórios que conformam a o Sistema Interligado Nacional (SIN), entre os quais encontra-se o reservatório Sobradinho, na Bahia. Considerou-se no estudo índices climáticos correspondentes a SSTs no oceano Atlântico; índices correspondentes ao ENSO e ventos zonais. O trabalho comparou os modelos Periódicos Autoregressivos (PAR), utilizados pelo SIN, com os modelos Periódicos Autoregressivos com Variáveis Exógenas (PARX) na previsão simultânea de vazões na escala mensal e sazonal, no qual o melhor desempenho foi o dos modelos PARX, que utilizaram correlação entre os ruídos da regressão (CRD) e a análise de componentes principais (ACP) para os períodos secos na maioria das regiões de estudo, incluindo o reservatório de Sobradinho. Desses modelos, aqueles com ACP apresentaram melhor desempenho devido aos altos valores das correlações que apresentaram com as observações. Assim, novamente demonstrou-se a aplicabilidade de modelos que utilizam informação climática.

Através dos estudos apresentados, pode-se apreciar que existem vários modelos que vem sendo utilizados na modelagem de vazões na região do NEB, dentre os quais, a geração por meio da distribuição Gama de dois parâmetros pode ser considerada como uma metodologia tradicional e aceita na região. Mas, apresentam-se recentemente correntes que introduzem novos modelos que empregam informação climática. Os trabalhos desse tipo mencionados anteriormente (Souza Filho e Lall, 2003; Ribeiro, 2011; Alexandre, 2012) demonstraram que é possível criar modelos utilizando as características climáticas próprias da região. Esses trabalhos também ressaltaram que o campo da modelagem incluindo informação climática na região apenas está começando, especialmente no estado do Ceará. Outros modelos desse tipo serão tratados no capítulo seguinte.

## 5 - REFERENCIAL TEÓRICO

### 5.1 - PROCESSOS ESTOCÁSTICOS E SÉRIES TEMPORAIS

Muitas variáveis aleatórias importantes em sistemas de recursos hídricos são funções cujos valores mudam com o tempo. Registros históricos de chuva ou vazão em um local particular constituem uma sequência de observações denominadas séries temporais cujas as observações são ordenadas pelo tempo. Geralmente, o valor observado de uma variável aleatória em um tempo influencia a avaliação da distribuição de probabilidade dessa variável aleatória em tempos posteriores, o que significa que as observações não são independentes (Loucks *et al.*, 2005).

Então, a variável cujo valor muda através do tempo de acordo a leis probabilísticas é denominada um processo estocástico. Dessa forma, uma série temporal observada é considerada uma realização de um processo estocástico, sendo só uma única observação de uma variável aleatória e um possível valor que essa variável pode assumir. Isso está de acordo à definição probabilística de *ensemble* (conjunto), no qual uma série temporal é parte de um conjunto de observações que definem um processo estocástico (Bras e Rodriguez-Iturbe, 1992; Loucks *et al.* 2005).

A área da hidrologia denominada hidrologia estocástica trata da análise de séries temporais, cujo principal interesse da análise é a magnitude e a sequência de eventos hidroclimáticos, por exemplo, vazões médias mensais, precipitação diária ou mensal, entre outros (Salas, 1993). Especificamente, os objetivos da análise de séries temporais são (Steffen, 2010):

- Descrição: análise das propriedades da série como, por exemplo, o padrão de tendência, a existência de alterações estruturais, valores atípicos ou *outliers*, sazonalidade, ruído aleatório, persistência de eventos ou *clustering*, etc.
- Explicação: construir modelos que permitam explicar o comportamento da série no período observado. Por exemplo, modelos baseados em sistemas lineares para explicar a relação entre duas variáveis.
- Controle de Processos: por exemplo, controle estatístico de qualidade.

- Previsão: prever valores futuros de uma variável com base em valores passados.

Usualmente, os hidrólogos tratam séries temporais de variáveis com valores discretos e contínuos, o que define dois tipos de series temporais: séries discretas e contínuas. Por exemplo, uma série temporal de vazões cujos valores  $Q(t)$  são contínuos no tempo representa uma serie continua. Essa série pode-se discretizar em valores de tempos discretos específicos ou fixos. A definição desses tempos define a escala temporal da série ou o nível de agregação. Os tempos definidos dependem dos objetivos da análise e das medições realizadas sobre a variável de interesse. O nível de agregação ocasiona um impacto não só na representação gráfica da série, mas também nas suas propriedades estatísticas (Salas, *et al.*, 1980; Steffen, 2010).

Outra classificação das séries temporais depende do seu número de variáveis. Dessa maneira, as séries temporais univariadas são as que possuem observações em um único local, como as observações da precipitação produzidas por uma única estação de medição em uma bacia. Já as observações provenientes de várias estações em diferentes locais de uma bacia produzem uma série temporal multivariada (Salas, 1993). A Figura 5.1 apresenta duas séries de vazões para um período de observações de 90 anos. Essas séries representam séries univariadas contínuas e discretas no tempo, mostrada em diferentes níveis de agregação, nível mensal e anual. Níveis são comumente usados em estudos hidrológicos que utilizam vazões (algumas vezes é de interesse o nível diário). Por exemplo, a escala mensal é utilizada nas vazões de entrada de um sistema hídrico. Nesta dissertação, serão analisadas séries temporais univariadas anuais.

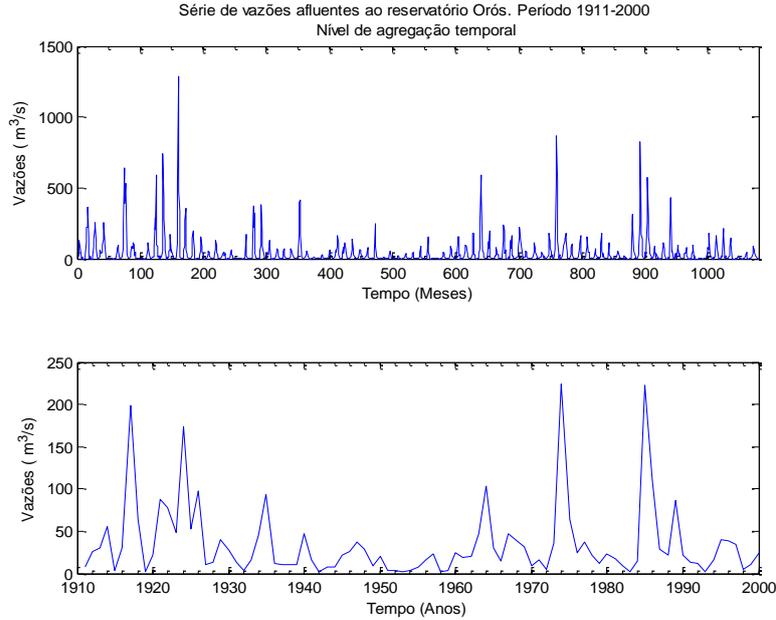


Figura 5.1 - Serie temporal de vazões afluentes ao reservatório Orós. Período 1911-200. Nível de agregação: série mensal (acima) e série anual (abaixo), (Fonte: Base de dados).

Considerando as informações acima, define-se um processo estocástico como a série temporal definida pela sequência de variáveis aleatórias  $R_{1:N} = \{R_1, \dots, R_t, \dots, R_N\}$ , ordenadas e discretas no tempo, onde  $N$  é o comprimento da série. As propriedades desse processo estocástico são geralmente determinadas a partir de uma única série temporal. Para fazer essa determinação, várias suposições são usualmente realizadas, como assumir que o processo é estacionário. Isso diz que a distribuição de probabilidade do processo não muda no tempo. Mas, se um processo é estritamente estacionário, então a distribuição conjunta das variáveis aleatórias  $R_1, \dots, R_t, \dots, R_N$  é idêntica à distribuição  $R_{1+t}, \dots, R_{t+t}, \dots, R_{N+t}$ , para qualquer  $t = 1, 2, \dots, N$  e essa ultima distribuição depende somente das diferenças  $t_i - t_j$  entre os tempos da ocorrência dos eventos (Loucks *et al.* 2005).

As estatísticas comumente utilizadas para descrever um processo estocástico estacionário são: a média, a variância e a autocorrelação. A média e a variância são definidas como (Loucks *et al.* 2005):

$$\mu_R = E[R_t] \tag{5.1}$$

$$\sigma_R^2 = Var[R_t] \quad (5.2)$$

E a autocorrelação, que representa a correlação entre as variáveis da própria série define-se para qualquer defasagem (*Lag*) positiva  $k$  como:

$$\rho_R(k) = \frac{Cov[R_t, R_{t+k}]}{\sigma_R^2} \quad (5.3)$$

Segundo a definição de série temporal como parte de um processo estocástico, uma série temporal só representa uma amostra. Portanto, essas estatísticas precisam ser calculadas por meio de estimadores amostrais, assim:

$$\hat{\mu}_R = \bar{R} = \frac{1}{N} \sum_{t=1}^N R_t \quad (5.4)$$

$$\hat{\sigma}_R^2 = \bar{S}_R = \frac{1}{N} \sum_{t=1}^N (R_t - \bar{R})^2 \quad (5.5)$$

E,

$$\hat{\rho}_R(k) = r_k = \frac{\sum_{t=1}^{N-k} (R_t - \bar{R})(R_{t+k} - \bar{R})}{\sum_{t=1}^N (R_t - \bar{R})^2} \quad (5.6)$$

Os valores desses estimadores possuem variabilidade que depende da estrutura de autocorrelação da série e se as observações são ou não independentes. Também, os estimadores são válidos sob a suposição de estacionariedade. Mas, os processos estocásticos não são sempre estacionários, como, por exemplo, as mudanças nas propriedades de uma bacia produzidas pelo desenvolvimento urbano e agrícola, além do gerenciamento dos seus recursos hídricos, que produzem variabilidade e mudanças dos padrões espaço-temporais de chuvas, vazões, concentrações de poluentes, entre outras variáveis. O que fará com que os processos não sejam mais estacionários no tempo, e assim, os estimadores anteriores já não serão válidos. Quando um processo deixa de ser estacionário, se torna mais difícil de modelar (Loucks *et al.* 2005). A respeito da autocorrelação, maiores informações serão tratadas na seção 5.2.

Além das estatísticas já descritas, também são utilizadas outras estatísticas descritivas como a mediana e o coeficientes de assimetria. Os seus estimadores amostrais são respetivamente os seguintes (Naghattini e Pinto, 2007):

Mediana:

$$R_{md} = R_{\left(\frac{N+1}{2}\right)} \quad \text{Se } N \text{ for ímpar}$$

$$R_{md} = \frac{R_{\left(\frac{N}{2}\right)} + R_{\left(\frac{N}{2}+1\right)}}{2} \quad \text{Se } N \text{ for par} \quad (5.7)$$

Coefficiente de Assimetria:

$$C_s = \frac{N}{(N-1)(N-2)} \times \frac{\sum_{t=1}^N (R_t - \bar{R})^3}{\bar{S}_R^3} \quad (5.8)$$

Além de conhecer essas estatísticas, é necessário conhecer a distribuição de probabilidade do processo e testar a aderência do processo a essa distribuição. É importante para a análise dos processos estocásticos, inclusive, conhecer a suas distribuições de probabilidade conjunta, marginal e condicional. Por exemplo, tem-se duas variáveis aleatórias  $R$  e  $S$ , a PDF conjunta  $f_{r,s}(r, s)$  representa o comportamento conjunto dessas variáveis. Logo, a PDF marginal é a aquela que só representa o comportamento de uma variável ignorando a outra, por exemplo,  $f_r(r)$ . Por ultimo, se o comportamento de  $R$  condiciona-se a  $S$ , diz-se que a PDF é condicional, assim essa função é  $f_{r|s}(r, s) = \frac{f_{r,s}(r,s)}{f_s(s)}$  (Loucks *et al.* 2005).

A partir dessas estatísticas, procura-se encontrar um modelo matemático que represente uma série temporal de observações e, muitas vezes, essa modelagem inclui o ajuste da série a distribuições de probabilidade. Nessa modelagem, existem dois grupos de modelos: Paramétricos e Não-Paramétricos. Na modelam paramétrica, os passos básicos são (Salas *et al.* 1980): (1) seleção do tipo de modelo, (2) identificação da ordem do modelo, (3) estimação dos parâmetros, (4) provas de ajuste e (5) geração de sequências de dados com o modelo selecionado e/ou previsão. Passos que são conhecidos como a metodologia de Box-Jenkins (1970). No calculo dos parâmetros, é importante compreender a natureza dos parâmetros visto que serão calculados parâmetros amostrais o que produz incerteza nos

resultados, nos quais a qualidade dos dados e o comprimento da série possuem um papel importante na redução ou aumento dessa incerteza (Salas *et al.*, 1980; Loucks *et al.*, 2005).

## 5.2 - FUNÇÃO DE AUTOCORRELAÇÃO

A autocorrelação em séries temporais indica a correlação de cada observação  $R_t$  com seu passado e seu futuro. Em outras palavras, a autocorrelação é um valor da correlação entre os membros da própria série. As séries temporais geofísicas frequentemente possuem autocorrelação pela inércia ou transição do sistema físico (Meko, 2005). Um exemplo são as séries de vazões que apresentam uma forte autocorrelação devido à persistência de um processo, como a precipitação, ou a persistência de estado, como climático seco ou úmido (Wang e Salas, 1991). As séries não correlacionadas são chamadas séries independentes (Salas, 1993). Para alguns fins, o fato de que uma série seja autocorrelacionada dificulta a aplicação de testes estatísticos, pela redução do tamanho efetivo da amostra. Também, são encontrados problemas para as análises de regressão e dificuldade para o cálculo de covariância ou correlação entre séries temporais (por exemplo, precipitação com uma série de anéis de árvores) (Meko, 2005).

Geralmente, séries que possuem alta correlação positiva tendem a apresentar persistência, pois valores de observações altos tendem a seguir altos e o mesmo ocorre com os valores baixos. Caso contrário, uma correlação negativa tende a variar de valores altos a baixos e de valores baixos para altos de um ano para outro com relação à média (Box e Jenkins, 1994). Essa característica pode ser observada nos gráficos de séries temporais, de forma que para o caso de autocorrelação positiva a série apresentará longos ou curtos períodos, acima ou abaixo da média. Para valores negativos de autocorrelação o gráfico apresentará uma variabilidade dos valores acima ou abaixo da média, apresentando ausência de períodos longos ou períodos curtos (Bras e Rodriguez-Iturbe, 1992).

A partir dessas considerações, define-se a função de autocorrelação (ACF) como a função que entrega os coeficientes ou valores de autocorrelação entre observações ( $R_t$ ) em diferentes tempos. Esses tempos são denominados defasagens ou *Lags*  $k$ . Logo, para vários valores de  $k$ , a função de autocovariância para uma série é definida como (Wang e Salas, 1991):

$$\gamma_k = Cov(R_{t+k}, R_t) = E[(R_{t+k} - \mu_R)(R_t - \mu_R)] \quad (5.9)$$

Então, a correlação entre as observações separadas por  $k$  anos ou a ACF é:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (5.10)$$

Para uma amostra, essa função se converte na equação 5.6. Nessa equação,  $r_k$  é o coeficiente de autocorrelação no *Lag*  $k$  para a série observada de comprimento  $N$  e geralmente assume-se que  $N \gg k$ . O gráfico da ACF como uma função de  $k$ , é denominado Correlograma e o número de *Lags*  $k$  pode ser aproximado ao inteiro mais próximo de  $N/10$  ou  $N/4$  (Box e Jenkins, 1994).

A ACF permite dar uma ideia da persistência do processo estocástico em estudo. Por exemplo, a autocorrelação  $r_1$ , onde  $k = 1$  (*Lag-one autocorrelation*), é a correlação da observação  $r_t$  com a observação  $r_{t+1}$ . Essa correlação é útil e bastante utilizada na hidrologia como ferramenta para avaliar séries sintéticas geradas por modelos estocásticos, como o Modelo Autoregressivo de ordem 1 (*Lag-one Autoregressive Model -AR(1)*). Nesse modelo, o objetivo é preservar a autocorrelação de primeira e segunda ordem da série observada. Objetivo que também é sugerido para fins de simulações e previsões de curto prazo para a maioria de modelos estocásticos (Bras e Rodriguez-Iturbe, 1992).

Além disso, por meio do Correlograma é possível visualmente ter informação sobre a série analisada como a aleatoriedade das observações. No caso de resíduos de um modelo, essa informação é oferecida por  $r_1$ . Também o Correlograma é útil para analisar o melhor Modelo Autoregressivo de Média Móvel de ordem  $p, q$  (*Autoregressive Moving Average Model- ARMA(p,q)*) que se ajusta às observações, onde os valores de  $q$  são os valores de  $k$  que apresentam correlações significativas no Correlograma. E os valores de  $p$  são extraídos similarmente do gráfico da Função de Autocorrelação Parcial (PACF). A PACF mede a correlação remanescente entre  $x_t$  e  $x_{t+1}$ , depois de se eliminar a influencia de  $x_{t+1}, \dots, x_{t-k-1}$  (Bayer & Castro, 2012). Maiores detalhes desses procedimentos e o calculo da PACF encontram-se na metodologia de Box e Jenkins (1976) que pode ser encontrada em Bras e Rodriguez-Iturbe (1992). Outras informações que podem ser

extraídas do Correlograma são a correlação de curto prazo, não estacionariedade da série, flutuações sazonais e *outliers* (Chatfield, 2003).

### 5.3 - MODELOS AUTOREGRESSIVOS DE MÉDIAS MÓVEIS

Como foi mencionado anteriormente, os modelos estocásticos procuram preservar um conjunto de estatísticas da série observada e incorporar isso dentro de modelos lineares ou não lineares estacionários. São denominados modelos de memória curta (ou dependência curta), pois as suas funções de autocorrelação decaem rapidamente como o aumento dos *Lags* (Salas e Boes, 1980). Se essas correlações decaem lentamente, os modelos são então chamados modelos de longa memória. Esse tipo de modelo pertence a família de modelos lineares estacionários AR, Média Móvel (*Moving Average* - MA) e ARMA, que são modelos atrativos pela relativamente fácil aplicabilidade e alcance em diferentes campos (Whiting, 2006). As suas definições e formulação matemática foram investigadas por Box e Jenkins (1970). Nesta seção são descritos os modelos que serão utilizados no presente estudo.

Os modelos AR com especificamente autocorrelação de ordem 1 são os modelos de ordem 1, AR(1) (*Lag- One Auto-Regressive Models-AR(1)*). Utilizam-se amplamente na hidrologia para simulação de séries temporais e previsão (Bras e Rodriguez-Iturbe, 1992). Para o modelo AR(1), são comumente usados outros nomes, como *Modelo Auto Regressivo de Markov* ou *Thomas Fiering*. Esse modelo pode ser representado em termos de um processo com média zero, como:

$$R_t = \phi_1 R_{t-1} + z_t \quad (5.11)$$

Onde  $z_t$  é uma sequência *white noise* com média zero e variância  $\sigma_z^2$ . A estacionariedade se garante por meio da condição que  $|\phi_1| < 1$ . A autocorrelação do processo  $\{R_t\}$  no *Lag*  $k$  é dada pela ACF  $r_k = \phi_1^k$ . Além disso, a variância do processo é relacionada com ambas a autocorrelação *Lag*-1 e variância da sequência *white noise* por:

$$\sigma_y^2 = \frac{\sigma_z^2}{(1 - r_1^2)} \quad (5.12)$$

Frequentemente esse processo é expresso em termos dos momentos centrais do processo aleatório (Bras e Rodriguez-Iturbe, 1992), como:

$$(x_t - \mu) = r_1(x_{t-1} - \mu) + \sigma_x(1 - r_1^2)^{0.5}w_t \quad (5.12)$$

Onde  $\mu$  é a média estacionária do processo  $\{x_t\}$ , e  $(x_t - \mu)$  é  $R_t$  da equação 5.11. Nesse caso, o termo aleatório  $z_t$  é substituído por  $w_t$ , que representa uma variável normalmente distribuída com média zero e variância igual a 1. E onde  $\sigma_x(1 - r_1^2)^{0.5}$  representa o desvio padrão de  $\{z_t\}$ . As características do modelo AR(1) são refletidas no seu espectro em que processos governados com esse tipo de modelo e com parâmetros positivos são dominados por flutuações de baixa frequência. Entanto, processos com parâmetros negativos apresentam um domínio de altas frequências. O processo  $\{R_t\}$  é denominado um processo Auto Regressivo de ordem  $p$  AR( $p$ ), se (Whiting, 2006):

$$R_t = \phi_1 R_{t-1} + \dots + \phi_p R_{t-p} + z_t \quad (5.13)$$

O processo linear estacionário, denominado processo de média móvel, pode ser descrito mediante os termos *white noise*. Logo, dado que  $\{z_t\}$  é um processo puramente aleatório com média zero e variância  $\sigma_z^2$ , o processo  $\{R_t\}$  é chamado um processo de média móvel (MA) de ordem  $q$ , MA( $q$ ), se:

$$R_t = \theta_1 z_{t-1} + \dots + \theta_q z_{t-q} + z_t \quad (5.14)$$

O processo MA( $q$ ) assume estacionariedade indiferentemente do valor dos parâmetros  $\theta$ , porque  $R_t$  é formada a partir de uma finita combinação linear de termos  $z_t$ . Porém, pela condição imposta  $|\theta_q| < 1$  para todo  $q$ , se garante que o processo seja invertível.

Um fator chave na modelagem de séries temporais é ter poucos parâmetros. Portanto, quando se ajustam modelos lineares estacionários, pode ser útil combinar processos AR e MA para formar um modelo ARMA. A importância desses processos está em que séries temporais estacionárias são frequentemente ajustadas a um modelo ARMA que envolve alguns poucos parâmetros de um processo MA ou AR para realizar o modelo mistura. Isso é expresso como (Whiting, 2006):

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + z_t + \theta_1 z_{t-1} + \dots + \theta_q z_{t-q} \quad (5.15)$$

Note-se que um ARMA(p,0) é um modelo AR(p) é um processo MA(q) corresponde a um ARMA(0,q).

#### 5.4 - MODELOS DE MARKOV COM ESTADOS OCULTOS

Os Modelos de Markov Oculto (*Hidden Markov Models* - HMM) são modelos estatísticos cuja distribuição de probabilidade que gera uma observação, depende de um estado pertencente a um processo de Markov não observado ou oculto. Os HMM apresentam vários atrativos como o fácil tratamento matemático, especialmente a computação ao calcular os parâmetros do modelo. O que produz uma ampla flexibilidade desses modelos para múltiplas aplicações em séries temporais univariadas e multivariadas (Zucchini e MacDonald, 2009).

Os HMM tem sido utilizados inicialmente no processamento de sinais, especialmente no campo do reconhecimento de voz (Juang e Rabiner, 1991). Mas, o interesse de outros campos pelo uso desses modelos aumentou nos últimos anos. As aplicações incluem todos os campos do reconhecimento (facial, de escrita, de assinatura); bioinformática (análises de sequências biológicas); finanças (séries de retornos diários); biofísica (modelagem de canais iônicos); meio ambiente (direção do vento, precipitação, vazões, terremotos) (Whiting, 2006; Zucchini e MacDonald, 2009).

O interesse especial que tem surgido na hidrologia por esses modelos se deve a sua capacidade de representar a persistência de regimes climáticos ou persistência hidrológica, possibilitando representar a persistência de períodos secos ou úmidos, com aplicação na simulação e previsão da precipitação (Thyer e Kuczera, 2000; Mallya *et al.*, 2011); Whiting, 2006; Lambert *et al.*, 2003), na geração de séries vazões anuais (Thyer e Kuczera, 2000; Akintug e Rasmussen, 2005; Whiting, 2006) e também na previsão de vazões anuais (Fortin *et al.*, 2004; Bracken, 2011). São utilizados, inclusive, em casos onde as variáveis modeladas tem influencia de fatores climáticos de longa escala e forte autocorrelação de ano para ano (Thyer e Kuczera, 2000; Whiting, 2006).

## 5.4.1 - Definições e Propriedades

### 5.4.1.1 - Modelos de Mistura Independente e Cadeias de Markov

Os Modelos de Markov oculto surgem a partir de dois processos estocásticos: um processo de mistura dependente e uma Cadeia de Markov. Nesta seção, explicam-se esses processos e as suas propriedades que servirão para logo chegar à definição dos HMM. Existem varias definições, mas se utilizaram a seguir as definições de Zucchini e MacDonald (2009) e as notações utilizadas no trabalho de Kirshner (2005).

### 5.4.1.2 - Modelos de Mistura Independente

Quando se tem um conjunto de  $N$  observações  $R_{1:N} = \{R_1, \dots, R_t, \dots, R_N\}$ , com uma distribuição de probabilidade multimodal, uma distribuição que têm duas ou mais médias, o que acontece quando é que a população pode consistir de grupos não observados, tendo cada grupo uma diferente distribuição para a variável observada. Nesse caso, um método que permite ajustar heterogeneidade não observada na população é utilizar um modelo de mistura.

Esse modelo de mistura consiste em ajustar uma distribuição mistura às observações, distribuição que é composta e selecionada a partir de  $m$  distribuições independentes definidas para a variável observada. Para  $i = 1, \dots, m$ , cada distribuição independente conta com uma probabilidade  $p_i(r)$ . No caso contínuo,  $p_i(r)$  corresponde à função de densidade de probabilidade (*Probability Density Function- PDF*) ou  $f_i(r)$ . Para poder especificar a distribuição componente que gera a observação, se define então a variável aleatória discreta  $S$ , que realiza a mistura, controlando que distribuição é ativada, portanto  $S$  tem uma probabilidade de ocorrência  $\pi_i$ . Assim, para uma probabilidade  $\pi_i$  se tem um valor de  $S$  e, portanto, uma distribuição será ativada utilizando a sua  $p_i(r)$  para gerar a variável aleatória  $R_t$ .

A Figura 5.2 exemplifica o caso de uma distribuição de dois componentes ( $p_1(r)$  e  $p_2(r)$ ), logo, quando é dada  $\pi_1$  então  $S = 1$  ativa a  $p_1(r)$ ,  $S = 2$  ativa a  $p_2(r)$ , no caso contrário. Note-se também que  $\pi_2 = 1 - \pi_1$ , o que cumpre com sua definição de probabilidade. Por

enquanto, não se sabe o valor de  $S$ , só se tem que as  $p_1$  e  $p_2$  são ativadas quando as observações são geradas.

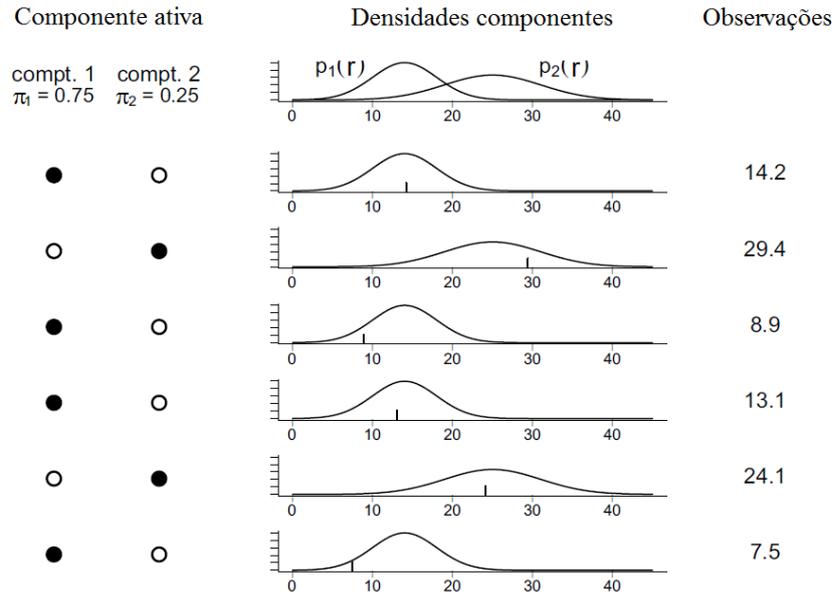


Figura 5.2 - Processo de uma distribuição de mistura de dois componentes. De cima para baixo, os estados são 1, 2, 1, 1, 2, 1. No meio são apresentadas as correspondentes distribuições componentes. As observações no lado direito são geradas pelas densidades componentes ativas, (Zucchini e MacDonald, 2009, traduzido).

A partir das considerações anteriores pode ser demonstrado que a probabilidade de  $R_t$  é dada pela equação 5.16 e também que a PDF de  $R_t$  é representada pela equação 5.17, assim:

$$p(x) = \sum_{i=1}^m p_i(x) \quad (5.16)$$

$$P_r(R = r) = \sum_{i=1}^m (P_r(R = r|S = i)P_r(S = i)) \quad (5.17)$$

Para a estimação dos parâmetros de um modelo de mistura é frequente utilizar o método de Máxima Verossimilhança (MVS) (*Maximum Likelihood-ML*), logo, a Verossimilhança ( $L$ ) para o caso discreto e contínuo é:

$$L(\theta_i, \pi_i | r_j) = \prod_{j=1}^N \sum_{i=1}^m \pi_i p_i(r_j, \theta_i) \quad (5.18)$$

Onde:

$\theta_i$ , é o vetor dos parâmetros da distribuição componente  $i$ ;  
 $\pi_i$ , as probabilidades da mistura, sendo  $\sum_{i=1}^m \pi_i = 1$  e;  
 $r_j$ , as observações.

Nota-se que os símbolos que representam vetores são ressaltados em negrito. Nesta dissertação, continuará utilizando essa notação para os vetores e matrizes.

#### 5.4.1.3 - Cadeias de Markov

A sequência de variáveis aleatórias discretas  $S_t = \{S_1, \dots, S_t\}$ , representa uma Cadeia de Markov (*Markov Chain*- MC), se para todo  $t$ , se satisfaz a propriedade:

$$P_r(S_{t+1}|S_{1:N}) = P_r(S_{t+1}|S_t) \quad (5.19)$$

Onde  $S_{1:N} = \{S_1, S_2, \dots, S_N\}$ , e se condiciona  $S_{t+1}$  ao evento observado  $S_t$ , sendo  $S_t$  o valor mais recente, portanto a variável  $S_t$  só depende de seu presente como se apresenta na seguinte figura:

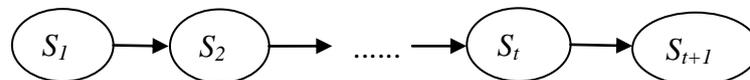


Figura 5.3 - Representação gráfica de uma Cadeia de Markov de primeira ordem

As probabilidades de transição de estados para essa MC são definidas como:

$$P_r(S_{x+t} = i | S_x = j) \quad (5.20)$$

Essa equação representa, na MC, a probabilidade no tempo  $t$  de ir ao estado  $i$ , sendo o anterior estado o estado  $j$ . Claramente  $i, j = 1, 2, \dots, m$ . Quando a MC é homogênea, essas probabilidades de transição não dependem de  $X$  (variável externa). Caso contrário, a MC será não homogênea. Por enquanto, será tratado o caso homogêneo para fins hidrológicos. Portanto, as probabilidades de transição da MC, são:

$$\gamma_{i,j}(t) = P_r(S_{x+t} = j | S_x = i) \quad (5.21)$$

Nota-se que equação 5.21 não depende de  $X$ . Logo, pode-se definir a matriz de probabilidades de transição (MPT)  $\Gamma$  contendo os  $(i, j)$  elementos  $\gamma_{i,j}(t)$ , como:

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \cdots & \ddots & \cdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}$$

Onde as filas de  $\Gamma$  somam 1, ou seja,  $\sum_{j=1}^m \gamma_{i,j} = 1$

De interesse nas MC, são também as probabilidades incondicionais (marginais),  $P_r(S_t = j)$ , logo o vetor  $\mathbf{u}(t)$ , é definido como:

$$\mathbf{u}(t) = (P_r(S_t = 1), P_r(S_t = 2), \dots, P_r(S_t = m)) \quad (5.22)$$

Onde  $\mathbf{u}(1)$  é a distribuição inicial da MC e a distribuição no tempo  $t + 1$  é:

$$\mathbf{u}(t + 1) = \mathbf{u}(t)\Gamma \quad (5.23)$$

Para essa MC, também pode ser definida a distribuição estacionária  $\boldsymbol{\pi}$  (vetor de elementos positivos) se:

$$\boldsymbol{\pi}\Gamma = \boldsymbol{\pi} \quad (5.24)$$

$$\boldsymbol{\pi}\mathbf{1}' = \boldsymbol{\pi} \quad (5.25)$$

Onde  $\mathbf{1}'$  é o vetor coluna de valores próprios de  $\Gamma$  e corresponde ao valor próprio 1. A equação 5.24 expressa estacionariedade e a equação 5.25 indica que  $\boldsymbol{\pi}$  seja de fato uma distribuição de probabilidade. Devido à equação 5.23, a Cadeia de Markov começa desde a distribuição estacionária e contínua nela pra todos os tempos  $t$ , portanto, o processo é chamado Cadeia de Markov estacionária. Como a propriedade de homogeneidade não é suficiente para que esse processo seja estacionário, então adicionalmente da homogeneidade será necessário que a cadeia de Markov tenha a propriedade de que  $\mathbf{u}(1)$  seja uma distribuição estacionária.

Ainda, tem-se que o vetor  $\boldsymbol{\pi}$  é uma distribuição estacionária se e somente se:

$$\boldsymbol{\pi}(\mathbf{I}_m - \Gamma + \mathbf{U}) = \mathbf{1} \quad (5.26)$$

Onde:

$\mathbf{1}$ , é o vetor coluna de uns;

$\mathbf{I}_m$ , é a matriz identidade de  $m \times m$  ;

$\mathbf{U}$  é a matriz de  $m \times m$  que contem todos os seus elementos iguais a 1.

Logo, a equação 5.26 constitui uma forma simples para calcular a distribuição estacionária  $\boldsymbol{\pi}$ . Até aqui, foram apresentadas as propriedades das MC de primeiro ordem. Uma análise semelhante pode ser feita para MC de ordem superior (de 2 ou mais ordens). Para maiores informações sobre o assunto, o leitor pode ir a Zucchini e MacDonald (2009) ou Bishop (2006).

#### 5.4.2 - Definição dos Modelos de Markov Oculto

Múltiplos nomes tem sido utilizados para os HMM entre eles estão: “*Hidden Markov Process*” (Ephraim e Merhav, 2002), “*Markov-Dependent Mixture*” (Leuroux e Puterman, 1992), “*Markov Switching Model*”, (Akintug e Rasmussen, 2005), “*Models Subject to Markov Regime*”, “*Markov Mixture Model*”, (Zucchini e MacDonald, 2009). Todos esses nomes em relação à origem dos HMM, e dado que as séries de dados utilizados tem uma forte autocorrelação (como as séries de vazões). O modelo surge então da união de dois processos: um processo de mistura dependente e uma Cadeia de Markov que permite fazer a dependência entre as observações de uma maneira matematicamente conveniente, pois, por definição, um processo de mistura independente não permite fazer essa dependência.

Segundo Whiting (2006), os Modelos de Markov oculto (HMM) discretos no tempo, ou seja, para séries de valores discretos; são descritos em termos de um par de processos  $\{S_t, R_t\}$ . Tendo uma serie temporal de observações  $R_{1:N} = \{R_1, \dots, R_t, \dots, R_N\}$  com uma distribuição de probabilidade determinada pelo estado  $s_t$  de um estado não observado  $m$  pertencente a uma MC. No caso  $m = 1$ , o HMM degenera uma série de variáveis aleatórias mutuamente independentes. Em aplicações para hidrologia,  $\{R_t\}$  pode representar valores totais de precipitação ou vazão em intervalos discretos de tempo (por exemplo, totais mensais ou anuais), com  $\{S_t\}$  caracterizando um conjunto de estados climáticos que influenciam o ciclo hidrológico. Isso oferece um método fácil para modelar a interação entre persistência de regimes climáticos e resposta hidrológica.

Supondo que  $S_t$  pode ser representada em qualquer tempo  $t$  ( $t = 1, 2, \dots, N$ ) como sendo parte de um conjunto de diferentes estados  $\{s_1, s_2, \dots, s_m\}$ , como resultado  $\{s_t, t = 1, 2, \dots, N\}$  é a série de estados, e no caso de um HMM de 2 estados ocultos, o valor de  $s_1$  pode representar um predominante estado úmido e o valor de  $s_2$  um estado seco (Thyer e Kuczera, 2000).

Da mesma forma, Zucchini e MacDonald (2009) descrevem primeiramente um HMM para o caso discreto, mencionando que similarmente pode ser formulado um HMM para o caso contínuo, dessa forma, um HMM de  $N$  observações  $R_{1:N} = \{R_1, \dots, R_t, \dots, R_N\}$ , é uma forma particular de uma mistura dependente; com  $R_t: \{t = 1, 2, \dots\}$  e  $S_t: \{t = 1, 2, \dots\}$  representando os históricos (ou séries), do tempo 1 até o tempo  $t$ , assim o modelo pode ser então descrito como:

$$P_r(S_t | \mathbf{S}_{1:t-1}) = P_r(S_t | S_{t-1}), \quad t = 1, 2, \dots, N \quad (5.27)$$

$$P_r(\mathbf{R}_t | \mathbf{R}_{1:t-1}, \mathbf{S}_{1:t}) = P_r(R_t | S_t), \quad t \in \mathbb{N} \quad (5.28)$$

E, portanto, o modelo consiste em dois processos:

1. Um processo (*parameter process*) não observado ou oculto (*Hidden*)  $S_t: \{t = 1, 2, \dots\}$  (equação 5.27). Processo que consiste em uma Cadeia de Markov, satisfazendo a propriedade de Markov (equação 5.19).
2. Um processo estocástico dependente dos estados  $S_t$ , (*state-dependent process*)  $R_t: \{t = 1, 2, \dots\}$ , tal que quando o estado  $S_t$  é conhecido, a distribuição de  $R_t$  depende só do estado atual  $S_t$  e não a estados prévios (equação 5.28). De acordo com isso, a Figura 5.4 representa graficamente um HMM e a Figura 5.5 representa o caso de um HMM de 2 estados ocultos.

Se a Cadeia de Markov  $\{S_t\}$  tem  $m$  estados, então  $\{R_t\}$  será um HMM de  $m$  estados ocultos tendo uma distribuição estacionária  $\boldsymbol{\pi}$  com elementos  $\pi_i = P_r(S_1 = i)$  e MPT  $\boldsymbol{\Gamma}$ , de elementos  $\gamma_{ij} = P_r(S_t = i | S_{t-1} = j)$  regulando a transição entre os estados  $\{S_t\}$  da Cadeia de Markov.

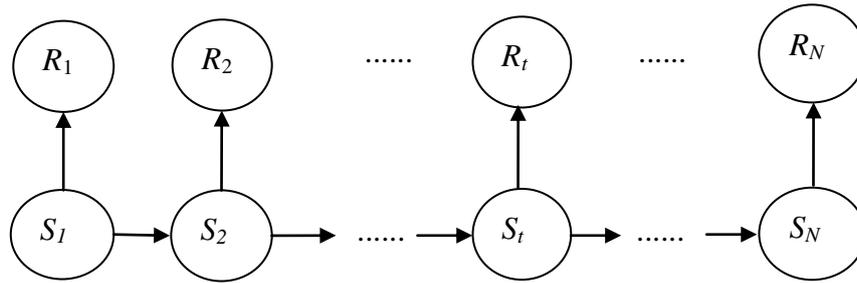


Figura 5.4 - Representação gráfica de um Modelo de Markov Oculto, HMM

Define-se também para o conjunto de observações discretas,  $i = 1, 2, \dots, m$ :

$$p_i(r) = P_r(R_t = r | S_t = i) \quad (5.29)$$

Onde  $p_i(r)$  é a função massa de probabilidade de  $R_t$  se a cadeia de Markov está no estado  $i$  no tempo  $t$ . Analogamente, se define para o caso contínuo  $p_i(r)$  como a PDF de  $R_t$ , se a cadeia de Markov está no estado  $i$  no tempo  $t$ . Assim, as  $m$  distribuições  $p_i(r)$  serão as distribuições dependentes dos estados (*state-dependent distributions*).

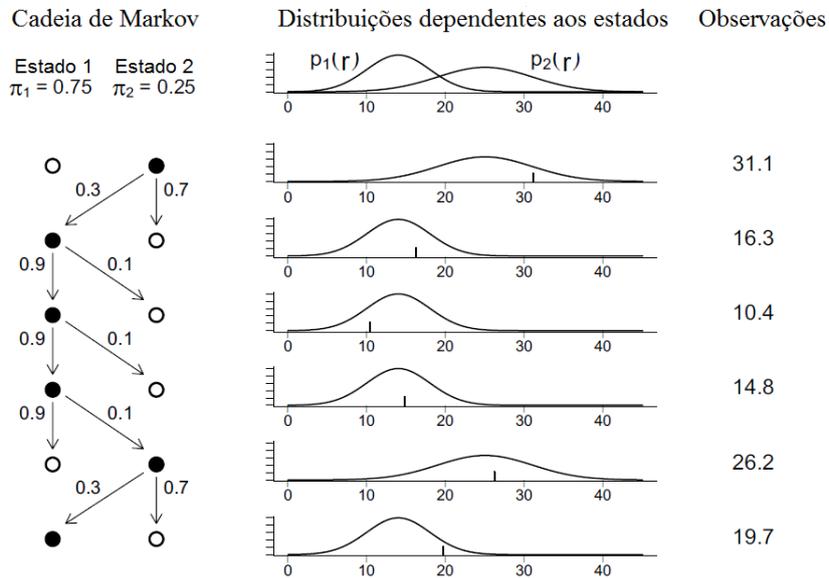


Figura 5.5 - Geração das observações em um HMM de dois estados ocultos. A Cadeia de Markov segue o caminho 2, 1, 1, 2, 1 como indicado na esquerda. As distribuições que são ativadas dependem dos estados ativos, apresentadas no meio. As observações são geradas dependendo da distribuição que é ativada (Zucchini e MacDonald, 2009, traduzido)

#### 5.4.2.1. - Distribuições marginais de um HMM

Para o caso univariado de observações discretas  $R_t$ , se redefina a equação 5.22 como:

$$u_i(t) = P_r(S_t = i) , \text{ para } t = 1, 2, \dots, N \quad (5.30)$$

E se tem:

$$P_r(R_t = r) = \sum_{i=1}^m P_r(S_t = i)P_r(R_t = r|S_t = i) = \sum_{i=1}^m u_i(t)p_i(r) \quad (5.31)$$

Matricialmente 5.16 pode ser escrita como:

$$P_r(R_t = r) = \mathbf{u}(t)\mathbf{P}(r)\mathbf{1}' \quad (5.32)$$

Onde,  $\mathbf{P}(r)$ , é a matriz diagonal com o  $i$  – ésimo elemento da diagonal igual a  $p_i(r)$ . Da equação 5.23 pode-se demonstrar que para uma MC homogênea e estacionaria que:

$$P_r(R_t = r) = \boldsymbol{\pi}\mathbf{P}(r)\mathbf{1}' \quad (5.33)$$

#### 5.4.2.2. - Momentos de um HMM

Para um conjunto de distribuições  $p_i(r), i = 1, 2, \dots, m$ , com parâmetros  $\mu_i$  e  $\sigma_i^2$ , representando à média e à variância, respectivamente, sendo elas distribuições componentes de um HMM homogêneo com distribuição estacionaria  $\boldsymbol{\pi}$ , tem-se que, segundo Akintug e Rasmussen (2005) e Zucchini e MacDonald (2009), para o conjunto de observações  $\{R_t\}$ , o valor esperado ( $\mu$ ), variância ( $\sigma^2$ ), coeficiente de assimetria ( $C_k$ ) e autocorrelação ( $\rho(k)$ ) para uma defasagem  $k \in \mathbb{N}$ , como segue:

$$\mu = E[R_t] = \sum_{i=1}^m \pi_i \mu_i = \boldsymbol{\pi}\boldsymbol{\mu}' \quad (5.33)$$

$$\sigma^2 = E[(R_t - \mu)^2] = \sum_{i=1}^m \pi_i (\sigma_i^2 + u_i^2) - \mu^2 \quad (5.34)$$

$$C_s = \frac{E[(X_t - \mu)^3]}{\sigma^3} = \sigma^3 \sum_{i=1}^m \pi_i [(\mu_i - \mu)^3 + 3\sigma_i^2(\mu_i - \mu)] \quad (5.35)$$

$$\rho(k) = \text{Corr}(X_t, X_{t+k}) = \frac{\boldsymbol{\pi} \mathbf{M} \boldsymbol{\Gamma}^k \boldsymbol{\mu}' - (\boldsymbol{\pi} \boldsymbol{\mu}')^2}{\sigma^2} \quad (5.36)$$

Onde:

$\boldsymbol{\mu}$  é o vetor de médias  $(\mu_1, \mu_2, \dots, \mu_m)$

$\mathbf{M}$  a matriz diagonal de  $\boldsymbol{\mu}$ ,  $\mathbf{M} = \text{diag}(\boldsymbol{\mu})$

Os detalhes sobre a derivação dos momentos para os HMM encontram-se no trabalho de Timmermann (2000). E os detalhes da derivação da ACF encontram-se no trabalho de Akintug e Rasmussen (2005).

#### 5.4.2.3. - Estimação dos parâmetros de um HMM

Para estimar os parâmetros de um HMM, é utilizado o método da Máxima Verossimilhança (MVS). A Verossimilhança denotada como  $L_N$ , é o resultado da soma de  $m^N$  termos onde cada termo é o produto de  $2N$  fatores (Zucchini e MacDonald, 2009), se o modelo têm as características antes descritas, e assumindo estacionariedade (mas, não é necessário essa suposição), de modo geral se tem:

$$L_N = \boldsymbol{\pi} \mathbf{P}(r_1) \boldsymbol{\Gamma} \mathbf{P}(r_2) \boldsymbol{\Gamma} \mathbf{P}(r_3) \dots \boldsymbol{\Gamma} \mathbf{P}(r_N) \mathbf{1}' \quad (5.37)$$

Onde  $\mathbf{P}(r_t)$  representa a  $\text{diag}(p_1(r_t), p_2(r_t), \dots, p_m(r_t))$ . Logo, define-se a  $\alpha_t$  para  $t = 1, 2, \dots, N$ , como:

$$\alpha_t = \boldsymbol{\pi} \mathbf{P}(r_1) \boldsymbol{\Gamma} \mathbf{P}(r_2) \boldsymbol{\Gamma} \mathbf{P}(r_3) \dots \boldsymbol{\Gamma} \mathbf{P}(r_N) = \boldsymbol{\pi} \mathbf{P}(r_1) \prod_{s=2}^t \boldsymbol{\Gamma} \mathbf{P}(r_s) \quad (5.38)$$

E pode-se provar que:

$$L_N = \alpha_N \mathbf{1}' \quad (5.39)$$

E,

$$\alpha_t = \alpha_{t-1} \mathbf{\Gamma P}(r_t), \quad t \geq 2$$

Assim, pode-se computar a  $L_N$  utilizando a equação 5.38 e considerando que:

$$\alpha_1 = \pi \mathbf{P}(r_1)$$

$$\alpha_t = \alpha_{t-1} \mathbf{\Gamma P}(r_t), \quad t = 2, 3, \dots, N$$

E, para o caso estacionário:

$$\alpha_0 = \pi$$

$$\alpha_t = \alpha_{t-1} \mathbf{\Gamma P}(r_t), \quad t = 1, 2, \dots, N$$

Depois, calcula-se  $L_N$  com base na equação 5.39.

A definida  $\alpha_t$  (equação 5.38) representa as probabilidades para frente (*forward probabilities*). Pode também definir-se para  $t = 1, 2, \dots, N$ , as probabilidade para trás  $\beta_t$  (*backward probabilities*) como:

$$\beta'_t = \mathbf{\Gamma P}(r_{t+1}) \mathbf{\Gamma P}(r_{t+2}) \dots \mathbf{\Gamma P}(r_N) \mathbf{1}' = \left( \prod_{s=t+1}^N \mathbf{\Gamma P}(r_s) \right) \mathbf{1}' \quad (5.40)$$

Especificamente,  $\alpha_t$  representa a probabilidade de observar a sequência parcial  $r_1, r_2, \dots, r_t$ , terminando no estado  $i$  no tempo  $t$ , e  $\beta'_t$  é a probabilidade de observar a sequência restante  $r_{t+1}, r_{t+2}, \dots, r_N$ , dado o estado  $i$  no tempo  $t$  (Akintug e Rasmussen, 2005).

Existem vários algoritmos utilizados em estatística para encontrar os parâmetros de modelos, maximizando  $L_N$ . Entre eles, um algoritmo comumente utilizado é o algoritmo

Maximização da Expectativa (*Expectation-Maximization-EM*) útil para encontrar a MVS quando existem variáveis ocultas ou dados faltantes. De maneira geral, o algoritmo opera mediante dois passos a Expectativa (E) e Maximização (M). No passo E, se computa as expectativas dos parâmetros condicionadas, ou seja, as expectativas dos parâmetros faltantes condicionadas aos dados observados, utilizando um valor inicial dos parâmetros dos dados não observados. O passo (M) maximiza o logaritmo da verossimilhança dos dados faltantes e observados, ou seja, maximiza o logaritmo da verossimilhança completa, utilizando as esperanças achadas no passo E e as esperanças dos parâmetros dos dados observados. Logo, os valores achados são utilizados em outro passo E, e algoritmo repete os passos até um critério de convergência imposto (Zucchini e MacDonald, 2009).

No contexto dos HMM, o algoritmo EM é conhecido como o algoritmo Baum-Welch e serve para estimar os parâmetros de um HMM, dado que, no caso dos HMM, existem funções ou distribuições com parâmetros desconhecidos. Portanto, o algoritmo serve para estimar o vetor de parâmetros  $\theta = \{\mu, \sigma, \Gamma, \pi\}$  com os parâmetros das distribuições dependentes dos estados, a matriz de probabilidades de transição  $\Gamma$ , e  $\pi$  a distribuição inicial (no caso estacionário é a distribuição estacionária) (Zucchini e MacDonald, 2009). Esse algoritmo utiliza as probabilidades para frente ( $\alpha_t$ ) e probabilidades para trás ( $\beta_t$ ). Maiores informações sobre o algoritmo podem ser encontradas em Zucchini e MacDonald (2009) e para modelos multivariados em Kirshner (2005).

#### 5.4.2.4. - Previsão utilizando HMM

Para fazer previsões utilizando um HMM, é necessário definir a uma distribuição de previsão, que surge da formulação da distribuição para a observação  $R_t$  condicionada a todas as outras observações no HMM. Dessa forma, a seguir define-se essa distribuição condicional, para definir, em seguida, a distribuição de previsão desejada. Seja  $R_{-t} = R_1, \dots, R_{t-1}, R_{t+1}, \dots, R_N$  as observações em todos os tempos diferentes ao tempo  $t$ , de forma similar se define a  $r_{-t}$ . Das equações 5.38 e 5.40 pode-se definir:

$$\alpha_t(j) = P_r(R_{1:t} = r_{1:t}, S_t = j) \quad (5.41)$$

Para  $t = 1, 2, \dots, N$  e  $j = 1, 2, \dots, m$

$$\beta_t(i) = P_r(R_{t+1} = r_{t+1}, R_{t+2} = r_{t+2}, \dots, R_N = r_N | R_t = i) \quad (5.42)$$

Para  $t = 1, 2, \dots, N - 1$  e  $i = 1, 2, \dots, m$  condição que  $P_r(S_t = i) > 0$

Ou

$$\beta_t(i) = P_r(R_{t+1}^N = r_{t+1}^N | S_t = i)$$

E pode-se demonstrar que:

$$\alpha_t(i)\beta_t(i) = P_r(R_N = r_N, S_t = i) \quad (5.43)$$

Para  $t = 1, 2, \dots, N$  e  $i = 1, 2, \dots, m$

$$\alpha_t \beta'_t = P_r(\mathbf{R}_N = \mathbf{r}_N) = L_N; \text{ Para cada } t \quad (5.44)$$

Onde,  $\mathbf{R}_N = R_{1:N} = \{R_1, \dots, R_N\}$  e  $\mathbf{r}_N$  são definidos de maneira similar. Agora, utilizando esses resultados e as equações 5.38 e 5.40, pode-se derivar a distribuição condicional de  $R_t$  como:

$$P_r(R_t = r | \mathbf{R}_{-t} = \mathbf{r}_{-t}) = \sum_{i=1}^m \omega_i(t) p_i(t) \quad (5.45)$$

Para  $t = 2, \dots, N$  e  $i = 1, 2, \dots, m$

Onde,  $\omega_i(t) = d_i(t) / \sum_{j=1}^m d_j(t)$  são as probabilidades de mistura, sendo elas funções das observações  $\mathbf{r}_{-t}$  e os parâmetros do modelo. E  $d_i(t)$  representa o produto da  $i$  – ésima entrada do vetor  $\alpha_{t-1} \mathbf{F}$ . No caso em que  $t = 1$ ,  $d_i(t)$  representa o produto da  $i$  – ésima entrada do vetor  $\boldsymbol{\pi}$  e a  $i$  – ésima entrada do vetor  $\boldsymbol{\beta}_1$ .

Uma vez definida a distribuição condicional para  $R_t$ , define-se a distribuição de previsão, que é um tipo dessa distribuição. Ela é uma distribuição condicionada a  $R_{N+h}$  dado  $\mathbf{R}_N = \mathbf{r}_N$ , e  $h$  é o horizonte de previsão. Assim, define-se a distribuição de previsão  $P_r(R_{N+h} = r | \mathbf{R}_N = \mathbf{r}_N)$  de maneira similar à equação 5.45 como:

$$P_r(R_{N+h} = r | \mathbf{R}_N = \mathbf{r}_N) = \frac{\alpha_N \Gamma^h \mathbf{P}(x) \mathbf{1}'}{\alpha_N \mathbf{1}'} \quad (5.46)$$

Essa equação pode ser escrita como uma mistura de distribuições de probabilidade dependentes de estados, assim:

$$P_r(R_{N+h} = r | \mathbf{R}_N = \mathbf{r}_N) = \sum_{i=1}^m \xi_i(h) p_i(x) \quad (5.47)$$

Onde  $\xi_i(h)$  é a  $i$ -ésima entrada do vetor  $\alpha_N \Gamma^h / \alpha_N \mathbf{1}'$ . Utilizando a equação 5.47 é possível fazer previsões pontuais e também previsões em intervalos. Quando  $h$  incrementa, a distribuição de previsão converge à distribuição marginal do HMM estacionário, assim:

$$\lim_{h \rightarrow \infty} P_r(R_{N+h} = r | \mathbf{R}_N = \mathbf{r}_N) = \boldsymbol{\pi}^* P(r) \mathbf{1}'$$

Onde,  $\boldsymbol{\pi}^*$  denota temporalmente a distribuição estacionária da MC diferente à distribuição inicial  $\boldsymbol{\pi}$ .

#### 5.4.2.5. - Sequência de Estados mais Prováveis

Em muitas aplicações é de interesse determinar os estados da MC que são mais prováveis (sob o modelo ajustado) para ter dado origem a sequência de observações, isso é conhecido como o problema de decodificação (*the decoding problem*) (Juang e Rabiner, 1991). O problema consiste em encontrar a decodificação local (*local decoding*) do estado  $i$  no tempo  $t$ , ou seja, qual estado é mais provável num dado tempo  $t$ . O problema em conjunto é denominado decodificação global (*global decoding*) e consiste em determinar a sequência de estados mais provável. A seguir serão apresentadas expressões para encontrar a probabilidade de um estado no tempo; e posteriormente expressões para encontrar a decodificação local e global.

Tomando as equações 5.38, 5.40 e 5.43, é possível chegar à distribuição condicional de  $S_t$  dadas as observações que podem ser obtidas, para  $i = 1, 2, \dots, m$ ; como:

$$P_r(S_t = i | \mathbf{R}_N = \mathbf{r}_N) = \frac{\alpha_t(i) \beta_t(i)}{L_N} \quad (5.48)$$

Equação que é necessária resolver para encontrar a decodificação local. Essa equação pode ser calculada por um método de dimensionamento, dado que o cálculo de  $\alpha_t(i)\beta_t(i)$  é numericamente instável. Seguindo o procedimento descrito por Rabiner (1989) e utilizado por Akintug e Rasmussen (2005), pode-se redimensionar a  $\alpha_t(i)$  e  $\beta_t(i)$ , normalizando-os para todos os estados. O procedimento começa calculando  $\alpha_1(i)$  e depois dimensionando-o como segue:

$$\hat{\alpha}_1(i) = \frac{\alpha_t(i)}{\sum_{j=1}^m \alpha_t(j)} \quad (5.49)$$

Onde,  $\hat{\alpha}_1(i)$  é o valor dimensionado. Para  $t = 2, 3, \dots, N$ ,  $\alpha_t(i)$  pode ser calculado recursivamente baseado nos valores dimensionados como:

$$\alpha_t(i) = \left( \sum_{j=1}^m \hat{\alpha}_{(t-1)}(j) \gamma_{ij} \right) p_i(r) \quad (5.50)$$

$$\hat{\alpha}_t(i) = \frac{\alpha_t(i)}{\sum_{j=1}^m \alpha_t(j)} \quad (5.51)$$

O mesmo procedimento é utilizado para redimensionar  $\beta_t(i)$ . Para  $t = N$ , se tem o conjunto  $\beta_N(i) = 1$ , e calcula-se o conjunto de coeficientes dimensionados  $\hat{\beta}_N(i)$  como:

$$\hat{\beta}_N(i) = \frac{\beta_t(i)}{\sum_{j=1}^m \beta_t(j)} \quad (5.52)$$

Então, para  $t = N - 1, N - 2, \dots, 1$ , calcula-se  $\hat{\beta}_t(i)$  como:

$$\beta_t(i) = \sum_{j=1}^m \gamma_{ij} p_i(r) \hat{\beta}_{t+1}(j) \quad (5.53)$$

$$\hat{\beta}_t(i) = \frac{\beta_t(i)}{\sum_{j=1}^m \beta_t(j)} \quad (5.54)$$

Visto que o cálculo da equação 5.48 envolve rádios (intervalos) de  $\alpha_t(i)$  e  $\beta_t(i)$ , os valores redimensionados cancelam-se. Assim,  $\hat{\alpha}_t(i)$  e  $\hat{\beta}_t(i)$  podem ser utilizados nas equações 5.51 e 5.52. Akintug e Rasmussen (2005) encontraram que esse procedimento de redimensionamento é necessário na maioria dos casos e que ele elimina completamente qualquer instabilidade numérica.

Outra abordagem para resolver a equação 5.48 é descrita por Zucchini e MacDonald (2009), segundo eles, para cada  $t \in \{1, \dots, N\}$ , o estado mais provável  $i_t^*$  dadas as  $N$  observações, é definido como:

$$i_t^* = \operatorname{argmax}_{i=1, \dots, m} P_r(S_t = i | \mathbf{R}_N = \mathbf{r}_N) \quad (5.55)$$

Essa equação determina o estado mais provável separadamente para cada  $t$  pela maximização da probabilidade condicional  $P_r(S_t = i | \mathbf{R}_N = \mathbf{r}_N)$  e, portanto, é chamada decodificação local. Na decodificação global em vez de maximizar essa probabilidade, busca-se a sequencia  $S_{1:N} = \{S_1, S_2, \dots, S_N\}$  que maximiza a probabilidade condicional:

$$P_r(\mathbf{S}_N = \mathbf{s}_N | \mathbf{R}_N = \mathbf{r}_N) \quad (5.56)$$

Ou maximizar a probabilidade conjunta:

$$P_r(\mathbf{R}_N, \mathbf{S}_N) = \pi_{s_1} \prod_{t=2}^N \gamma_{S_{t-1}, S_t} \prod_{t=1}^N p_{S_t}(r_t) \quad (5.57)$$

Algumas vezes, a decodificação local e similar à decodificação global, mas nunca serão idênticas.

Maximizar a equação 5.57 não é fácil e, portanto, para solucionar esse problema, comumente é utilizado um algoritmo eficiente de programação dinâmica, denominado o algoritmo *Viterbi*. Esse algoritmo precisa das seguintes definições:

$$\xi_{1i} = P_r(S_1 = i, R_1 = r_1) = \pi_i p_i(r_1) \quad (5.58)$$

E para  $t = 2, 3, \dots, N$ , se tem:

$$\xi_{ti} = \max_{s_1, s_2, \dots, s_{t-1}} P_r(\mathbf{S}_{t-1} = \mathbf{s}_{t-1}, S_t = i, \mathbf{R}_N = \mathbf{r}_N) \quad (5.59)$$

E pode-se provar a seguinte recursividade para  $t = 2, 3, \dots, N$  e  $i = 1, 2, \dots, m$ :

$$\xi_{tj} = (\max_i (\xi_{t-1,i} \gamma_{i,j})) p_j(r_t) \quad (5.60)$$

O que é útil para calcular a equação 5.59, e a maximização das sequencias  $i_1, i_2, \dots, i_N$ , pode ser determinada recursivamente por:

$$i_N = \operatorname{argmax}_{i=1, \dots, m} (\xi_{Ni}) \quad (5.61)$$

E para  $t = N - 1, N - 2, \dots, 1$

$$i_t = \operatorname{argmax}_{i=1, \dots, m} (\xi_{ti} \gamma_{i,i+1}) \quad (5.62)$$

Esse algoritmo pode ser reescrito em términos dos logaritmos das probabilidades, e pode ser aplicado nos casos em que a MC é estacionária e não estacionária, assim, não é necessário assumir que a distribuição inicial é a distribuição estacionária. A decodificação global é o principal objetivo em muitas aplicações, especialmente quando se tem interpretações de fundo para os estados (caso de variação de estados climáticos, seco e úmido). É, então, de interesse investigar o desempenho da decodificação global em identificar os estados corretos. Isso pode ser feito pela simulação das séries a partir de um HMM, aplicando diretamente o algoritmo, a fim de decodificar as observações simuladas, e, então comparar o caminho Viterbi (sequência Viterbi) com a (conhecida) série de estados simulados (Zucchini e MacDonald, 2009).

Isso permite fazer uma checagem da simulação, pois o caminho é uma matriz de probabilidades (a sequência é similar a essa matriz) em que as filas e colunas devem somar 1. E pode-se achar a probabilidade de que um estado inferido ( $j$ ) seja na realidade outro estado ( $i$ ), ou seja, as probabilidades  $p(i/j)$  e  $p(j/i)$ . Isto ajuda a quantificar a precisão do caminho Viterbi e é recomendado quando a interpretação desse caminho é um importante objetivo de análises (Zucchini e MacDonald, 2009). Detalhes sobre esse algoritmo podem ser encontrados em Kirshner (2005).

#### 5.4.2.6. - Extensões dos HMM

O HMM básico, conforme estudado, incluindo a sua estimação de parâmetros, pode ser estendido em numerosas vias para satisfazer os requerimentos de aplicações particulares (Bishop, 2006). A continuação apresenta representações esquemáticas de alguns casos importantes, detalhes desses modelos podem ser encontrados em Kirshner (2005) e Bishop (2006).

A Figura 5.6 representa um HMM Autoregressivo (AR-HMM), no qual as observações  $R_t$  dependem dos estados ocultos  $S_t$ , da observação  $R_{t-1}$  e também de  $k$  observações anteriores. No caso específico da Figura 5.6,  $k = 1$ .

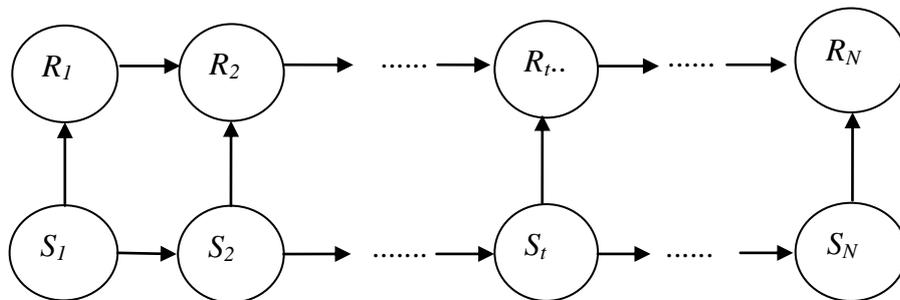


Figura 5.6 - Representação gráfica de um AR-HMM.

Quando em um HMM as observações  $R_t$  dependem de  $S_t$  e de variáveis externas  $X_t$ , diz-se que o modelo é um *Input-Output* HMM. A Figura 5.7 representa esse tipo de modelo, onde as setas tracejadas representam a dependência de  $R_t$  a  $X_t$ .

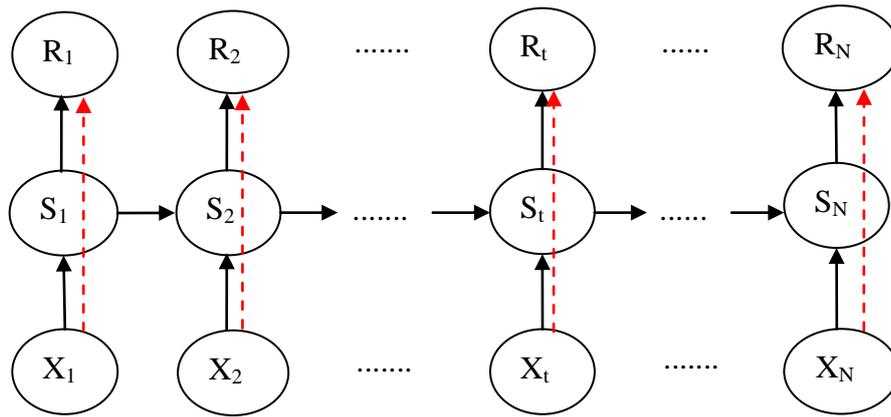


Figura 5.7 – Representação gráfica de um Input-Output HMM.

Outro caso ocorre quando se tem várias MC. Nesse caso, o modelo é denominado um HMM Fatorial. A Figura 5.8 representa esse tipo de modelo que possui dois MC ( $S1_t$  e  $S2_t$ ). As setas tracejadas representam a dependência de  $R_t$  à segunda MC ( $S2_t$ ).

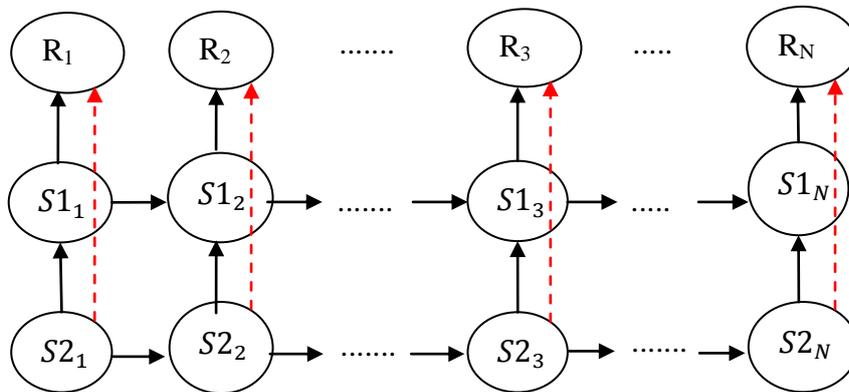


Figura 5.8 – Representação gráfica de um HMM Fatorial

O caso no qual as probabilidades de transição dependem de outras variáveis externas foi mencionado na seção 4.1.1.2. Nesse caso, o HMM converte-se num modelo não-homogêneo. Esse modelo será descrito na próxima seção.

### 5.4.3 - Modelos de Markov Oculto Não Homogêneo

No caso em que algumas variáveis ( $X$ ) influenciam as observações  $R_t$ , por exemplo, a chuva pode ser influenciada por fatores atmosféricos como a temperatura, velocidade do vento; entre outras variáveis ou influenciada por processos climáticos de larga escala resultantes de anomalias nas temperaturas no oceano, essas variáveis  $X$  podem ser entradas (*inputs*) num HMM homogêneo e fazer que esse processo não seja mais homogêneo, influenciando as probabilidades de transição de estados, e dessa forma, influenciar as

distribuições de probabilidade que geram as observações (Kirshner, 2005a). Os modelos que trata o descrito anteriormente são chamados Modelos de Markov Oculto Não Homogêneos (*Non-homogeneous Hidden Markov Models - NHMM*). Kirshner (2005a) os descreve como segue.

Dado um conjunto  $D$ -dimensional de variáveis de entrada para o modelo  $X_{1:N} = \{X_1, \dots, X_t, \dots, X_N\}$ , que convertem o HMM em não homogêneo, ao fazer que a probabilidade do estado oculto  $S_t$  dependa do estado  $S_{t-1}$  e também dependa do valor de  $X_t$ , redefine-se a equação 6.12 e, portanto, tem-se que:

$$P(S_t | S_{1:t-1}, X_{1:N}) = P(S_t | S_{1:t-1}, X_{1:t}), \quad t = 1, 2, \dots, N \quad (5.63)$$

E a Figura 5.9, representa graficamente a estrutura desse modelo.

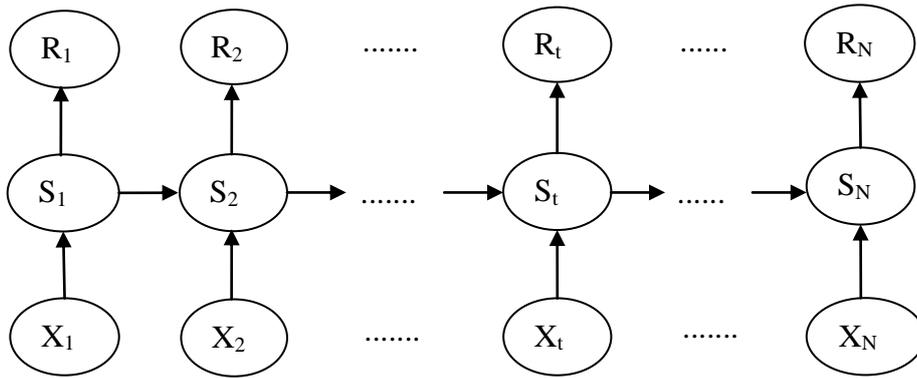


Figura 5.9 - Representação gráfica de um Modelo de Markov Oculto Não Homogêneo, NHMM.

Além disso, as probabilidades de transição  $\gamma_{ij}(r)$  e as distribuições iniciais  $\pi_i(r)$  dependem da variável  $X_t$ , assim:

$$\pi_i(r) = P(S_1 = i | X_1 = x) \quad (5.64)$$

$$\gamma_{ij}(r) = P(S_t = i | S_{t-1} = j, X_t = x) \quad (5.65)$$

E para modelar a transição de estados utiliza-se uma regressão multinomial logística, assim:

$$P(S_t = i | S_{t-1} = j, \mathbf{X}_t = \mathbf{x}) = \frac{\exp(\sigma_{ji} + \rho_i x^t)}{\sum_{i=1}^m \exp(\sigma_{ji} + \rho_i x^t)} \quad \text{para } t = 2, \dots, N \quad (5.66)$$

$$P(S_1 = i | \mathbf{X}_1 = \mathbf{x}) = \frac{\exp(\lambda_i + \rho_i x^t)}{\sum_{i=1}^m \exp(\lambda_i + \rho_i x^t)} \quad \text{para } t = 1 \quad (5.67)$$

Onde,  $m$  representa o número de total de estados e  $\lambda_i, \sigma_{ij} \in \mathbb{R}$  e  $\rho_i \in \mathbb{R}^D$ , logo, se define a  $\Theta = (\Omega, \Gamma)$  como o conjunto total de parâmetros do modelo NHMM, onde  $\Omega$ , representa a matriz dos parâmetros  $\lambda_i, \sigma_{ij}$  e  $\rho_i$  de tamanho  $1 \times m$  e  $\Gamma$  é a matriz de transição de probabilidades (MPT).

Pode-se notar que o HMM é um caso especial de um NHMM com:

$$\pi_i = \frac{\exp(\lambda_i)}{\sum_{i=1}^m \exp(\lambda_i)}$$

E:

$$\gamma_{ij} = \frac{\exp(\sigma_{ji})}{\sum_{i=1}^m \exp(\sigma_{ji})}$$

Com  $i, j = 1, 2, \dots, m$  ou equivalentemente  $\rho_i = 0, i = 1, 2, \dots, m$ .

Essa conclusão permite que as considerações feitas para os HMM sejam válidas para os NHMM. Então, é possível, a partir da extensão das equações dos HMM, definir as propriedades para os NHMM. E também, por meio do algoritmo *Baum-Welch*, estimar o conjunto de parâmetros  $\Theta$  e utilizar o algoritmo Viterbi para a decodificação global e realizar a decodificação local. Detalhes sobre esses algoritmos para NHMM podem ser encontrados em Kirshner (2005).

## **6 - INFLUENCIA DA VARIABILIDADE CLIMÁTICA NO REGIME HIDROLÓGICO DO NORDESTE DO BRASIL**

As relações entre fatores climáticos e o regime hidrológico é um assunto de interesse em regiões que apresentam baixa disponibilidade hídrica, como a região do Nordeste do Brasil (NEB) (Souza Filho e Lall, 2003). Porém, como apresentado na seção 4.2, os modelos estocásticos que incluem informação climática na modelagem de séries sintéticas de vazões ainda estão em surgimento. Portanto, neste capítulo, será tratada a relação de fenômenos climáticos globais de baixa frequência com o regime climático do NEB, para contextualizar a influência de fatores climáticos na variabilidade e persistência de variáveis hidrológicas como a precipitação e vazões.

### **6.1 - CARACTERÍSTICAS DO NORDESTE DO BRASIL**

O NEB é principalmente uma região semiárida, que ocupa grande parte do território nacional. O semiárido abrange os Estados de Alagoas, Bahia, Ceará, Minas Gerais, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe, ocupando uma área de 977 mil km<sup>2</sup>. Compreende 1.133 municípios e abriga uma população de 20 milhões de habitantes, correspondente a 12% da população brasileira. Desses 20 milhões, aproximadamente 56% encontram-se na área urbana, enquanto 44% na zona rural. Especificamente, o Estado do Ceará possui 86,8% da sua área inserida no semiárido (Ceará, 2008; ANA, 2010a).

O semiárido Cearense apresenta um conjunto de características representativas, entre as principais estão as seguintes (Ceará, 2008):

- Precipitação pluviométrica média anual inferior a 800 mm;
- Índice de aridez de até 0,5 calculado pelo balanço hídrico que relaciona as precipitações e a evapotranspiração potencial, no período entre 1961 e 1990; e
- Risco de seca maior que 60%, tomando-se por base o período entre 1970 e 1990.

Essas são algumas das características do ponto de vista científico, mas do ponto de vista cultural, agrícola e industrial, outros critérios podem descrever o semiárido, por exemplo,

comumente a região é chamada como a região onde incidem as secas prolongadas. A ideia de seca, por sua vez, vai desde a falta de precipitação, deficiência de umidade no solo agrícola, quebra de produção agropecuária, até impactos sociais e econômicos negativos em geral, ou seja, identificação de áreas de secas, em função de causas e efeitos, com diversos níveis de abrangência. Além disso, existem outras características marcantes da região, elas são (Ceará, 2008; Silveira, *et al.*, 2011):

- Cobertura vegetal de caatinga;
- Embasamento cristalino predominante;
- Solos geralmente rasos, pouco permeáveis e sujeitos à erosão;
- Rios, em sua maioria, intermitentes;
- Eventos hidrológicos extremos frequentes: secas e cheias;
- Escoamento específico reduzido: 4 L/s/km<sup>2</sup> ou 1.260 m<sup>3</sup> 21 /ha./ano;
- Temperaturas altas entre 24.1°C e 27.1°C, com insolação intensa entre 2.417 h/ano e 2.983 h/ano o que leva a taxas elevadas de evapotranspiração de 1.469 mm/ano a 2.904 mm/ano.

Em vista disso, na região frequentemente apresenta um balanço hídrico desfavorável, com déficit hídrico em quase todo estado. Isso repercute na diminuição dos recursos hídricos para os diferentes usos, em particular o abastecimento humano. Trata-se, portanto, de um território vulnerável, em que a irregularidade interanual das chuvas pode chegar a condições extremas representadas por frequentes e longos períodos de estiagem, que são o principal impacto da variabilidade climática na região. Esses períodos críticos têm sido os maiores responsáveis pelo histórico êxodo de grande parte da sua população (Marengo, 2008; ANA, 2010a, 2010b).

## **6.2 - FENÔMENOS CLIMÁTICOS E SISTEMAS ATMOSFÉRICOS SOBRE O NORDESTE DO BRASIL**

Segundo Hastenrath e Heller (1977), os mecanismos que governam o regime de chuva no NEB são os eventos relacionados ao fenômeno do *El Niño- Southern Oscillation* (ENSO); a dinâmica dos ventos tropicais e os sistemas de circulação atmosférica como: a Zona de Convergência Intertropical (ZCIT) sobre o oceano Atlântico; Frentes Frias (FF); Vórtices Ciclônicos de Altos Níveis (VCANs); linhas de Instabilidade (LI); Complexos Convectivos

de Mesoescala (CCM), e o efeito das brisas marítima e terrestre na precipitação. Esses mecanismos descrevem-se a seguir.

### **6.2.1 - *El Niño – Southern Oscillation***

*El Niño – Southern Oscillation* (ENSO) é um sistema/fenômeno de interações entre o Oceano Pacífico equatorial e a atmosfera. Os estados desse sistema possuem variabilidade interanual. Uma das principais formas de observar as variações entre estados do ENSO é através das flutuações das temperaturas da superfície do mar (*Surface Sea Temperatures-SSTs*) do oceano Pacífico equatorial. A variabilidade do ENSO afeta o clima estacional de muitas regiões do mundo incluindo o NEB. O ENSO apresenta duas fases ou estados opostos, eles são: *El Niño* e *La Niña*. *El Niño* descreve os eventos nos que o oceano Pacífico equatorial encontra-se mais quente que a média e *La Niña* representa eventos contrários deixando o Pacífico equatorial mais frio que a média (IRI, 2012a). A Figura 6.1 apresenta o ENSO nas suas fases *El Niño* e *La Niña*, determinadas por anomalias da SST a respeito da média. Pode-se apreciar as elevadas anomalias no caso *El Niño* (Figura 6.1 a) e baixas anomalias no caso *La Niña* (Figura 6.1 b).

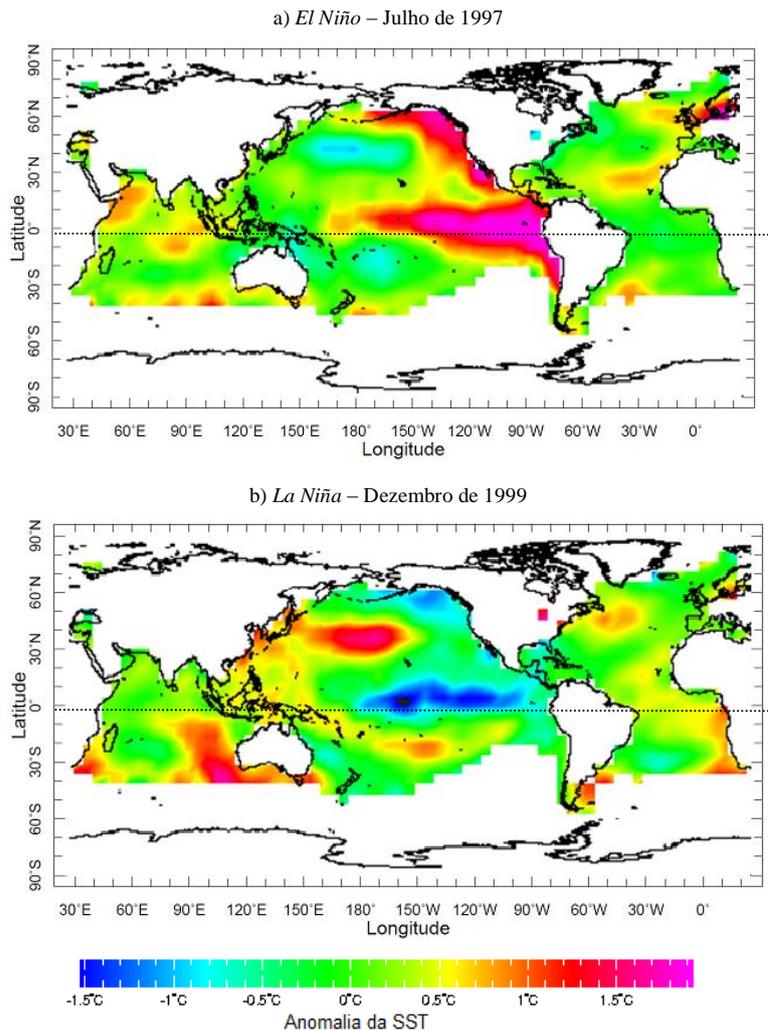


Figura 6.1 - Estados do ENSO: a) Condições *El Niño* para o mês de julho de 1997 e b) condições *La Niña* para o mês de Dezembro de 1999, (IRI, 2012b).

Além disso, os estados do ENSO estão associados com mudanças nos ventos dos trópicos, a pressão atmosférica e os padrões das chuvas. Logo, nessa zona, as chuvas ocorrem quando as SSTs estão mais quentes, portanto, as tendências das chuvas podem ser previstas utilizando as condições do ENSO. Em **condições normais** do ENSO, a água mais quente encontra-se no Oceano Pacífico Ocidental, e assim, ocorrem maiores precipitações nessa zona. Os ventos perto da superfície viajam do leste a oeste através do Pacífico (ventos do este), ver Figura 6.2a. Enquanto, que nas **condições *El Niño*** (Figura 6.2b), os ventos do este se debilitam, as SSTs estão mais quentes que a média para grande parte do Pacífico tropical central do leste e a região de chuvas mais fortes movimenta-se para o leste. Por último, nas **condições *La Niña***, apresentam-se as condições opostas a *El Niño*. Nessa fase (Figura 6.2 c), os ventos do este se fortalecem, a água mais fria do oceano estende-se para o oeste até o Pacífico central. As SSTs são mais quentes que a média no Pacífico ocidental

e estão acompanhadas por precipitações maiores que o habitual na Austrália, no NEB e na região do sudeste da Ásia (IRI, 2012a).

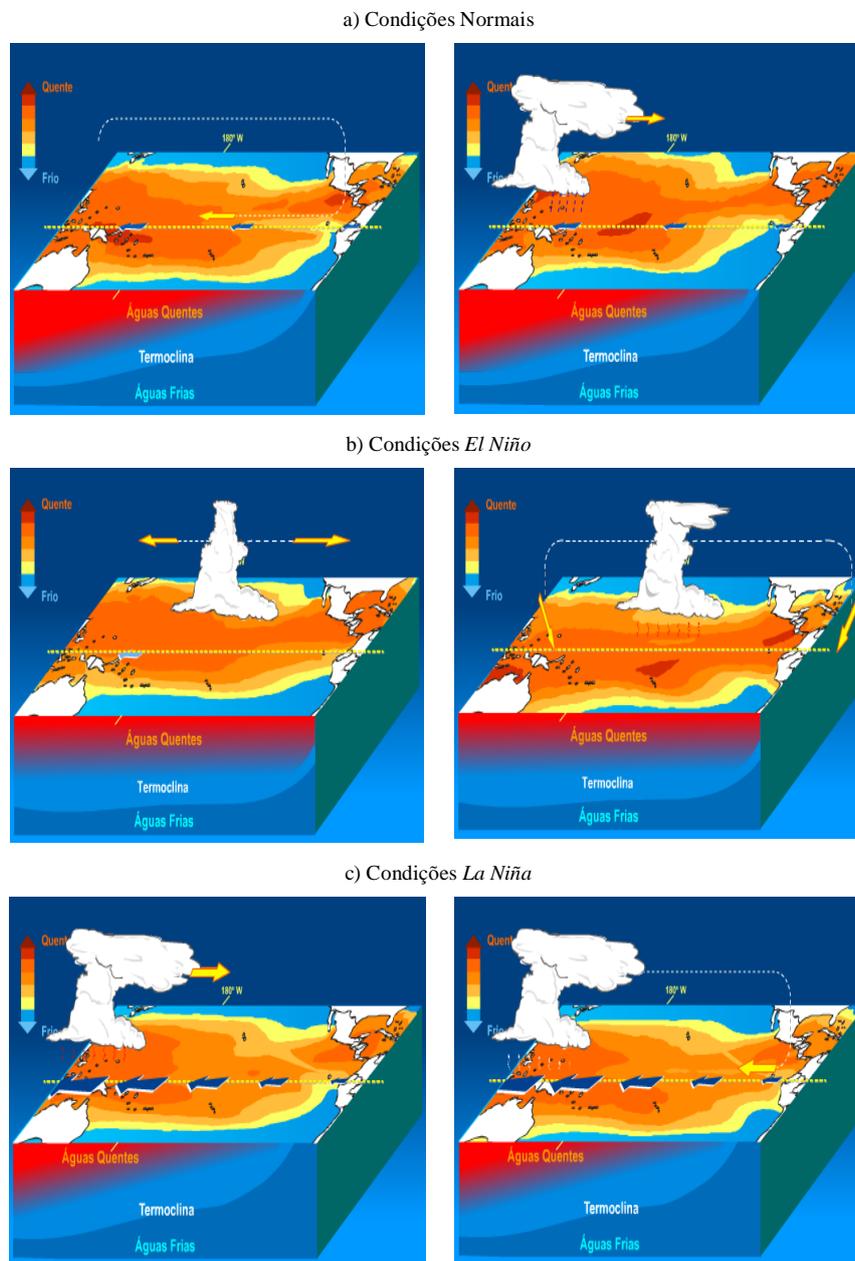


Figura 6.2 - Mecanismos do ENSO: a) condições normais observam-se águas superficiais relativamente mais frias no Pacífico equatorial leste, junto à costa oeste da América do Sul, e relativamente mais aquecidas no Pacífico equatorial Oeste, b) condições *El Niño*, ocorre uma diminuição das águas mais frias que afloram próximo à costa oeste da América do Sul e c) condições *La Niña*, As águas adjacentes à costa oeste da América do Sul tornam-se ainda mais frias devido à intensificação do movimento de ressurgência, (INPE/CPTEC, 2012).

Enquanto que o Pacífico tropical afeta a atmosfera, também a atmosfera tem uma influência sobre o oceano. De fato, a interação entre a atmosfera e oceano é uma parte

essencial do *El Niño* e *La Niña*, pois, durante *El Niño*, a pressão ao nível do mar tende a ser menor no Pacífico oriental e superior no Pacífico ocidental. Um fenômeno contrário tende a ocorrer em um episódio *La Niña*. Essas variações das pressões atmosféricas se denomina a Oscilação Sul (*Southern Oscillation*). Como o *El Niño* e a Oscilação Sul estão relacionados, comumente utiliza-se a frase *El Niño – Southern Oscillation* (ENSO), para referir-se ao fenômeno completo (IRI, 2012a). Para ilustrar essa relação, a Figura 6.3 apresenta a relação entre as anomalias da SST no Pacífico tropical e as variações das pressões entre a Austrália e Taiti (*Southern Oscillation Index-SOI*), para correlação gráfica de 0.67.

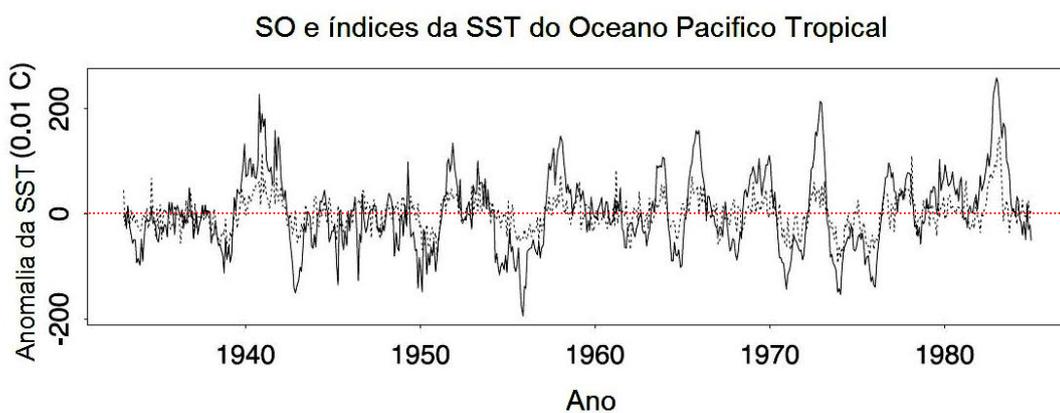


Figura 6.3 - Convencional índice da Oscilação Sul (*Southern Oscillation Index-SOI*) medido como a diferença entre a pressão atmosférica média ao nível do mar entre Darwin (Austrália) e Papeete (Taiti) dado em mbars (linha tracejada) e um índice da SST para a Oscilação Sul (linha sólida). Nesta figura o SOI tem sido duplicado (Storch e Zwiers, 1999, traduzido)

Além dos efeitos climáticos no oceano Pacífico tropical, o ENSO também afeta muitas regiões, chegando a produzir variabilidade sazonal. A Figura 6.4 apresenta os efeitos do ENSO em diversas partes do mundo, para seus dois estados. Esses efeitos são produto das diferentes interações dos sistemas atmosféricos globais, que entram em relação com o ENSO. Nota-se, na Figura 6.4, efeitos contrários no NEB, pois se tem condições secas e chuvosas para cada estado. Além desse fenômeno, outros fenômenos atmosféricos e climáticos somados a os fatores característicos de uma determinada região também produzem variabilidade climática, mas, o ENSO é considerado a principal influência para o clima mundial (Whiting, 2006; IRI, 2012a).

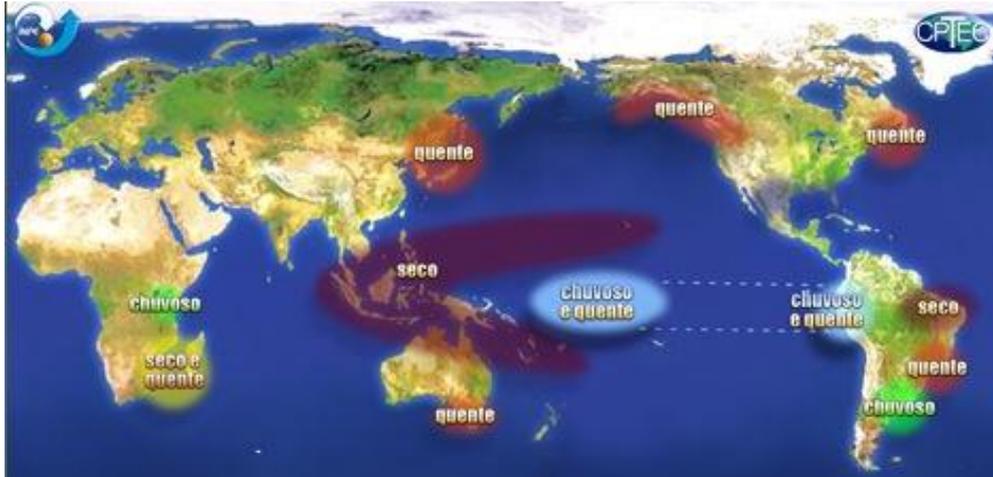
Em consequência da importância do ENSO, mudanças desse fenômeno precisam ser monitoradas. Assim, as interações atmosféricas do Pacífico com o ENSO são monitoradas através de variáveis que tenham relação a esses fenômenos. Uma medida comumente utilizada é o índice SOI, o qual consiste em uma diferença de pressões. Utilizando esse índice tem-se que os eventos *El Niño* são caracterizados por períodos extensos de valores do SOI negativos e *La Niña* ocorre no caso contrário. Dessa forma, diz-se que o ENSO é um fenômeno climático persistente. Como um monitoramento atmosférico se sujeita a outros fenômenos atmosféricos de pequena escala que podem intervir nessa diferença de pressões, um método alternativo é monitorar as SSTs no leste do Oceano Pacífico equatorial, para identificar anomalias dessas SSTs e, assim, identificar a variabilidade do ENSO. Deste modo, tem sido propostos índices climáticos para quantificar as anomalias relacionadas ao fenômeno, medindo diretamente no oceano. Eles são definidos como anomalias da SST em quatro regiões do Pacífico tropical (Whiting, 2006).

A Tabela 6.1 apresenta as regiões desses índices e a nomenclatura utilizada, enquanto que a Figura 6.5 apresenta a sua localização geográfica no oceano Pacífico. Essas regiões apresentam as seguintes características: a região Niño 1+2 é a região que tipicamente esquenta primeiro quando surge um evento *El Niño*; a região Niño 3 apresenta maior variabilidade na SST nas diferentes escalas do *El Niño*; da mesma forma a região Niño 3.4 é similar à Niño 3, mas, as variações são mais claras, pois a região é onde essas variações produzem precipitações locais (Figura 6.2 b), além de ser a região mais afastada dos continentes no Pacífico ocidental. A Niño 4 é a região onde valores elevados de mudanças da SST (ao redor de 27.5°C) ajudam à produção de precipitação na zona (IRI, 2012c).

Tabela 6.1 - Regiões dos índices do ENSO, (Schöngart e Junk, 2007)

<b>Região</b>	<b>Índice ENSO</b>	<b>Latitude</b>	<b>Longitude</b>
Niño 1+2	NINO12	0°-10°S	90°W-80°W
Niño 3	NINO3	5°N-5°S	150°W-90°W
Niño 3.4	NINO34	5°N-5°S	170°W-120°W
Niño 4	NINO4	5°N-5°S	160°E-150°W

a) *El Niño*



b) *La Niña*

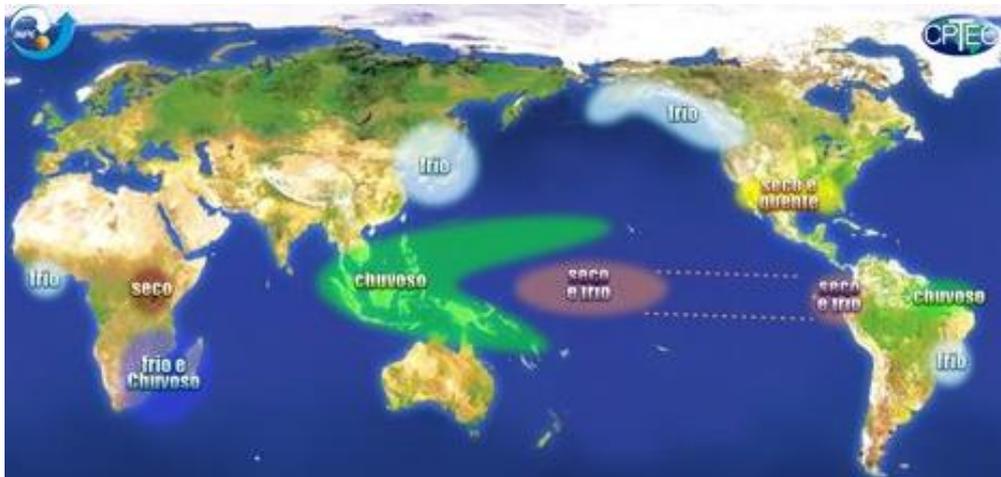


Figura 6.4 - Efeitos globais do ENSO para os meses de dezembro janeiro e fevereiro: a) efeitos do *El Niño*, b) efeitos da *La Niña* (INPE/CPTEC, 2012).

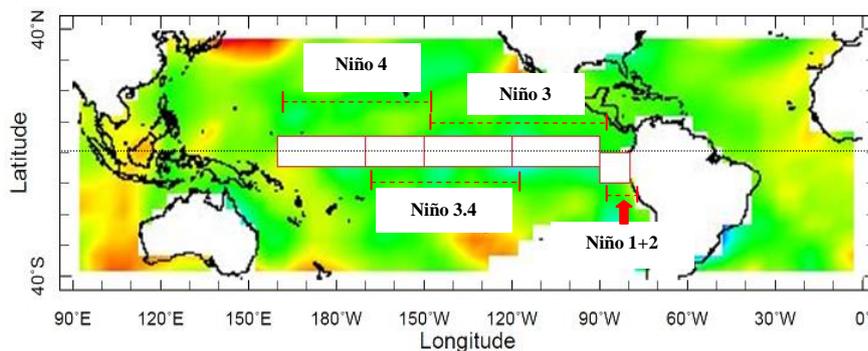


Figura 6.5 - Localização das regiões dos índices ENSO (IRI, 2012b).

Frequentemente climatologistas australianos sinalam que valores positivos desses índices acima de  $+0.8^{\circ}\text{C}$  por vários meses são associados a condições *El Niño*, e valores abaixo de  $-0.8^{\circ}\text{C}$  são condições de *La Niña*, esses valores dos índices são aproximadamente um desvio padrão de uma média de longo prazo (ABOM, 2012). Um dos índices comumente

utilizados é o NINO3, embora pela sua importância o NINO3.4 também é muito utilizado. Para o índice NINO3, toma-se uma média por 5 meses e períodos no mínimo de 6 meses com valores acima de  $+0.4^{\circ}\text{C}$  são associados a *El Niño* e valores com a mesma duração abaixo de  $-0.4^{\circ}\text{C}$  são relacionados a eventos *La Niña* (Whiting, 2006; IRI, 2012c).

## 6.2.2 - Sistemas de Circulação Atmosférica

A dinâmica da circulação atmosférica sobre a região tropical é fortemente modulada e modificada pelos padrões termodinâmicos dos oceanos Pacífico e Atlântico tropicais. Em anos nos quais se verificam anomalias positivas ou negativas das SSTs nesses oceanos, a célula de Hadley, que atua no sentido meridional (ramo ascendente sobre os trópicos e ramos descendentes nas latitudes subtropicais), e a célula de Walker, que atua no sentido zonal (ramo ascendente no Pacífico oeste e ramo descendente no Pacífico leste) são perturbadas, causando fortes anomalias na circulação atmosférica sobre os trópicos, visto que essas células são deslocadas de suas posições climatológicas. Consequentemente, as intensidades e duração do período chuvoso dessa região também são afetadas (Hastenrath e Heller, 1977). Além disso, vários sistemas atmosféricos apresentam-se na região do NEB, esses sistemas definem-se a seguir (FUNCEME, 2012):

### 6.2.2.1. - Zona de Convergência Intertropical

A Zona de Convergência Intertropical (ZCIT) é uma banda de nuvens que circunda a faixa equatorial do globo terrestre (Figura 6.6), formada principalmente pela confluência dos ventos alísios do hemisfério norte com os ventos alísios do hemisfério sul, em baixos níveis (o choque entre eles faz com que o ar quente e úmido ascenda e provoque a formação das nuvens), baixas pressões, altas SST, intensa atividade convectiva e precipitação. Esse mecanismo é o fator mais importante do regime de chuvas no norte do NEB. Normalmente a ZCIT migra sazonalmente de sua posição mais ao norte, aproximadamente  $14^{\circ}\text{N}$  em agosto-outubro, para posições mais ao sul, aproximadamente  $2^{\circ}\text{S}$  a  $4^{\circ}\text{S}$  entre fevereiro a abril. Esse deslocamento da ZCIT está relacionado aos padrões da SST sobre o Atlântico tropical, como foi demonstrado por Uvo *et al.* (1998). A ZCIT é mais significativa sobre os oceanos, e por isso, a SST é um dos fatores determinantes na sua posição e intensidade.

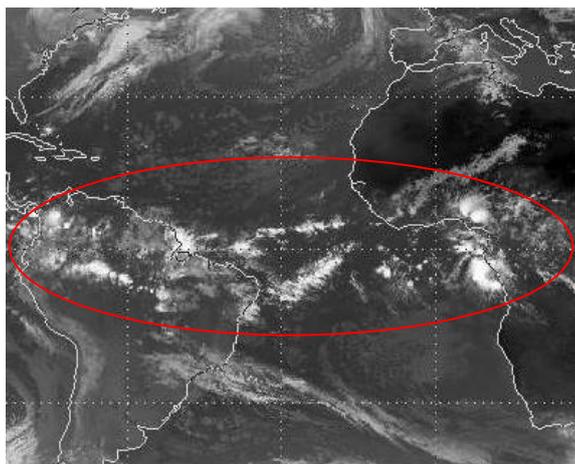


Figura 6.6 - Zona de Convergência Intertropical - ZCIT, sinalada pela elipse vermelha, (INPE/ CPTEC, 2012)

#### 6.2.2.2. - Frentes Frias

As Frentes Frias (FF) são outro fator importante para a produção de chuva no NEB, esse fator está relacionado à penetração de frentes frias até às latitudes tropicais entre os meses de novembro e janeiro. As FF são bandas de nuvens organizadas que se formam na região de confluência entre uma massa de ar frio (mais densa) com uma massa de ar quente (menos densa). A massa de ar frio penetra por baixo da quente, e faz com que o ar quente e úmido suba, formem-se as nuvens e apresentem-se chuvas.

#### 6.2.2.3. - Vórtices Ciclônicos de Altos Níveis

Os Vórtices Ciclônicos de Altos Níveis (VCANs) são vórtices ciclônicos ar superior que penetram no NEB e formam-se no Atlântico, principalmente entre os meses de novembro e março, e sua trajetória normalmente é de leste para oeste, com maior frequência entre os meses de janeiro e fevereiro. O tempo de vida desses sistemas varia em média, entre 7 a 10 dias. Os VCANs são um conjunto de nuvens que tem a forma aproximada de um círculo girando no sentido horário. Na sua periferia há formação de nuvens causadoras de chuva e no centro há movimentos de ar de cima para baixo (subsidiência), aumentando a pressão e inibindo a formação de nuvens.

#### 6.2.2.4. - Linhas de Instabilidade

As Linhas de Instabilidade (LI) são bandas de nuvens causadoras de chuva, normalmente do tipo cúmulos, organizadas em forma de linha, daí o seu nome. Sua formação se dá basicamente pelo fato de que, com a grande quantidade de radiação solar incidente sobre a região tropical ocorre o desenvolvimento das nuvens cúmulos, que atingem um número maior à tarde e início da noite, quando a convecção é máxima, com consequentes chuvas. Outro fator que contribui para o incremento das LI, principalmente nos meses de fevereiro e março, é a proximidade da ZCIT.

#### 6.2.2.5. - Complexos Convectivos de Mesoescala

Os Complexos Convectivos de Mesoescala (CCMs) são aglomerados de nuvens que se formam devido às condições locais favoráveis como temperatura, relevo, pressão, etc., e provocam chuvas fortes e de curta duração, normalmente acompanhadas de fortes rajadas de vento. Os CCMs, na região subtropical, ocorrem preferencialmente durante os meses de primavera e de verão no hemisfério sul, formando-se no período noturno com um ciclo de vida entre 10 e 20 horas. Normalmente as chuvas associadas aos CCMs ocorrem de forma isolada.

#### 6.2.2.6. - Ondas de Leste

As Ondas de Leste (OL) são ondas que se formam no campo de pressão atmosférica, na faixa tropical, na área de influência dos ventos alísios, e se deslocam de oeste para leste, ou seja, desde a costa da África até o litoral leste do Brasil. Provocam chuvas principalmente na Zona da Mata que se estende desde o Recôncavo Baiano até o litoral do Rio Grande do Norte, mas, quando as condições oceânicas e atmosféricas estão favoráveis, as OL também provocam chuvas no estado do Ceará nos meses de junho, julho e agosto, principalmente na parte centro-norte do estado.

#### 6.2.2.7. - Brisas Marítimas e Terrestres

Segundo Ferreira e Mello (2005), outros mecanismos são as brisas marítimas e terrestres. Essas brisas resultam do aquecimento e resfriamento diferenciais que se estabelecem entre a terra e a água. Durante o dia, o continente se aquece mais rapidamente que o oceano adjacente, fazendo com que a pressão sobre o continente seja mais baixa que sobre o

oceano. Isto faz com que o vento à superfície sopra do oceano para o continente, vento que é denominado brisa marítima. Essa brisa pode chegar a penetrar até 100 km para dentro do continente. No período da noite o continente perde calor mais rapidamente que o oceano, produzindo com que esse fique com temperaturas mais elevadas se comparadas às do continente. Dessa forma, a pressão fica maior sobre o continente, e o vento sopra do litoral para o oceano, esse vento chama-se brisa terrestre. A brisa terrestre também afeta até 100 km para dentro do mar. As brisas terrestre e marítima nem sempre são percebidas. No NEB, por exemplo, onde os ventos alísios são persistentes e intensos durante todo o ano, quase sempre as brisas apenas contribuem para mudar um pouco a direção e a velocidade dos ventos. Dependendo da orientação da costa, a velocidade do vento, resultante da superposição alísio-brisa, pode ser maior ou menor que a do alísio (Varejão-Silva, 2001).

### **6.3 - VARIABILIDADE CLIMÁTICA SOBRE O NORDESTE**

O clima de qualquer região é determinado em grande parte pela circulação geral da atmosfera, resultante do aquecimento diferencial do globo pela radiação solar, da distribuição assimétrica de oceanos e continentes e também das características topográficas dos continentes. Padrões de circulação atmosférica redistribuem calor, umidade e momento (quantidade de movimento) por todo o globo. No entanto, essa redistribuição não é homogênea, agindo algumas vezes no sentido de diminuir as variações regionais dos elementos climáticos, tais como, temperatura e precipitação, as quais tem enorme influência nas atividades humanas (Ferreira e Mello, 2005).

Logo, esse quadro e as características do semiárido nordestino fazem com que, na região, predomine um clima semiárido com intensa **variabilidade temporal** (intra, interanual, decenal) e **variabilidade espacial** de chuvas (Marengo, 2006), devido fundamentalmente à atuação simultânea dos sistemas atmosféricos antes mencionados e à dinâmica das suas interações. Esses sistemas podem variar com as características fisiográficas da região, assim como pelos padrões anômalos de grande escala de circulação atmosférica global associada aos Oceanos Pacífico e Atlântico Tropical, o que modifica a frequência, distribuição espacial e intensidade desses sistemas (Lázaro *et al.* 2011).

#### **6.3.1 - Variabilidade Sazonal e Espacial**

No que se refere à **variação intrasazonal**, o NEB possui um ciclo anual bem definido, onde predominam duas estações distintas: a estação de chuvosa e a estação seca. A estação de chuva concentra-se entre dezembro e subdivide-se em pré-estação chuvosa (dezembro e janeiro). Nessa estação, os principais sistemas causadores de chuva são: a proximidade das FF e os VCANs. A quadra chuvosa compreende os meses de fevereiro, março, abril e maio e tem à ZCIT como principal sistema causador de chuva seguido de sistemas secundários como as LI e os CCMs e efeitos de brisas. Os meses de junho e julho são considerados pós-estação chuvosa e os sistemas causadores de chuva são as OL e os CCMs. No segundo semestre a ocorrência de chuva é pequena, há um predomínio de altas pressões atmosféricas e uma quase total ausência de fenômenos atmosféricos causadores de chuva, por isso chama-se de estação seca (Silveira *et al.* 2011). Essas variações interanuais também estão ligadas a **variações espaciais** em toda a região nordestina, segundo Marengo (2008), a distribuição espacial das chuvas é a seguinte:

- No sul-sudoeste, o principal período chuvoso é de outubro a fevereiro. As chuvas são ocasionadas principalmente pela passagem de FF que vêm do sul do país. Além disso, pode-se observar chuvas em forma de pancadas isoladas, que normalmente ocorrem no final da tarde e início da noite, devido ao aquecimento durante o dia.
- No norte, abrangendo a maior parte do semiárido, o principal período chuvoso ocorre entre os meses de fevereiro e maio. Essa região é onde ocorrem as maiores secas. O mais importante sistema causador de chuvas nessa sub-região é a ZCIT.
- No Leste ou Zona da Mata, o principal período chuvoso ocorre de abril a agosto e as chuvas são ocasionadas principalmente pelos contrastes entre a temperatura do mar junto à costa e a temperatura do continente. Os ventos, ao soprar continente adentro, carregam a umidade do oceano que condensa e precipita na faixa litorânea e na região da Zona da Mata.

### 6.3.2 - Variabilidade Interanual

O fenômeno ENSO sobre o Pacífico equatorial e anomalias da SST sobre o Atlântico tropical modulam conjuntamente uma grande parte da variabilidade interanual do clima sobre o NEB. A combinação das circulações atmosféricas anômalas induzidas pelas distribuições espaciais de SST sobre os oceanos Pacífico equatorial e Atlântico tropical afetam o posicionamento latitudinal da ZCIT sobre o Atlântico, influenciando a

distribuição da pluviometria sobre o NEB. Apesar da variabilidade interanual das SSTs e ventos sobre o Atlântico tropical serem significativamente menores das variações observadas no Pacífico equatorial, essa variabilidade exerce profunda influência na variabilidade climática sobre o NEB (Uvo *et al.*, 1998; Marengo, 2006).

Como o mencionado na seção 5.2.1, o NEB apresenta sinais consistentes de *El Niño*, pois, na região, existe uma tendência para as secas que ocorre devido ao aquecimento das águas superficiais no Pacífico Central, que interfere no regime de ventos sobre toda a região equatorial do Pacífico. As nuvens, que normalmente produzem chuvas abundantes na parte oeste do oceano Pacífico, nas vizinhanças da Indonésia, deslocam-se para leste, para o Pacífico central e, posteriormente, para a costa oeste da América do Sul, trazendo as chuvas ao deserto peruano. A mudança de posição das chuvas no Pacífico provoca alterações nas condições climáticas, devido à grande quantidade de energia envolvida no processo de formação da chuva, portanto, produzem-se grandes secas no NEB. Durante *La Niña*, ocorrem anomalias climáticas inversas ao *El Niño*, embora o fenômeno ENSO não seja estritamente linear. Nas condições *La Niña*, os Ventos Alísios no Pacífico são mais intensos que a média climatológica. Apresenta-se uma intensificação da pressão no Pacífico Central e Oriental, em relação à pressão no Pacífico Ocidental, o que produz aumento da produção de chuvas no NEB. Em geral, o episódio começa a se desenvolver em meados do ano, atinge sua intensidade máxima no final daquele mesmo ano e se dissipa cerca de seis meses depois (Marengo, 2006).

Marengo (2006) menciona que o fenômeno *El Niño* pode ser caracterizado como cíclico, mas não possui um período estritamente regular, ele é caracterizado por uma variabilidade interanual aparecendo em uma escala geralmente de ao longo de 2 a 7 anos. Entretanto, podem existir períodos nos quais o *El Niño* ocorre com fraca intensidade, como na década de 1920 -1930 e em meados do século XVIII. Ou eventos anômalos como *El Niño* de 1982/83. Nesse evento, o aquecimento da SST ocorreu inicialmente no Pacífico Central e depois se estendeu para a costa da América do Sul. O episódio de 1986/1987 também foi similar, pois, apesar de ter-se monitorado a Oscilação Sul e as condições das SST, pareciam indicar um episódio *El Niño* durante 1986, as condições só intensificaram-se depois de maio de 1987.

Em comparação com casos anteriores, o evento 1997/98 apresenta uma característica diferente do padrão normal e extremamente importante: a taxa de crescimento da anomalia de SST foi muito superior ao normal e ao prognosticado nas previsões realizadas pelos principais centros meteorológicos mundiais no final do ano 1996. Em adição, o episódio começou um pouco mais cedo que o normal, comparando-se apenas ao caso de 1972/73. *El Niño* também produziu as secas extremas dos anos 1932 e 1983 e o evento 1997/98 gerou a grande seca de 1998. Nos últimos 15 anos, foram apenas três ocasiões em que o *El Niño* foi sucedido pela *La Niña*. O episódio intenso de *El Niño* em 1982/83 foi seguido de um evento fraco de *La Niña* em 1984/85, e um *El Niño* menos intenso, ocorrido em 1986/87, um *El Niño* relativamente longo entre 1990-94 e o evento muito intenso em 1997/98. O evento em 1986/87 foi seguido de um forte *La Niña* em 1988/89. O *El Niño* longo, mas pouco intenso do 1990-94 foi seguido de um episódio fraco *La Niña* em 1995/96.

O oceano Atlântico tropical tem também um papel fundamental na variabilidade climática interanual do NEB. Uma série de trabalhos sobre esse assunto é apresentada por Marengo (2006). Muitos desses trabalhos apresentam fortes evidências observacionais, teóricas e resultados de modelos de circulação geral da atmosfera, de que as condições oceânicas e atmosféricas do Atlântico tropical influem fortemente na variabilidade interanual na região (Moura e Shukla, 1981; Hastenrath, 1994, 2000, 2006; Souza Filho e Lall, 2003; Robertson *et al.*, 2004; Lima, 2010).

Portanto, o padrão espacial predominante do ciclo anual e da variabilidade interanual das SSTs é que os ventos à superfície sobre o Atlântico apresentam uma estrutura norte-sul mais pronunciada do que a estrutura leste-oeste. A estrutura norte-sul das anomalias das SSTs observadas no Atlântico é conhecida como o DIPOLO do Atlântico tropical (Uvo *et al.*, 1998). Esse dipolo é definido como a diferença aritmética entre as anomalias da SST do oceano Atlântico norte tropical (TAN) na região de coordenadas geográficas 5° N - 20° N, 60° W - 30° W e anomalias da SST do oceano Atlântico sul tropical (TAS) nas coordenadas 0° - 20°S, 30° W - 10°, a Figura 6.7 apresenta essas duas regiões no Atlântico tropical e a Figura 6.8 mostra um modelo conceitual desse dipolo, observa-se que o dipolo no Atlântico tropical propicia a ocorrência de gradientes meridionais de anomalias de SST, os quais afetam fortemente na posição latitudinal da ZCIT, modulando a distribuição sazonal de precipitação pluviométrica sobre o Atlântico equatorial, na parte norte do NEB, (Uvo *et al.*, 1998; Marengo, 2006).

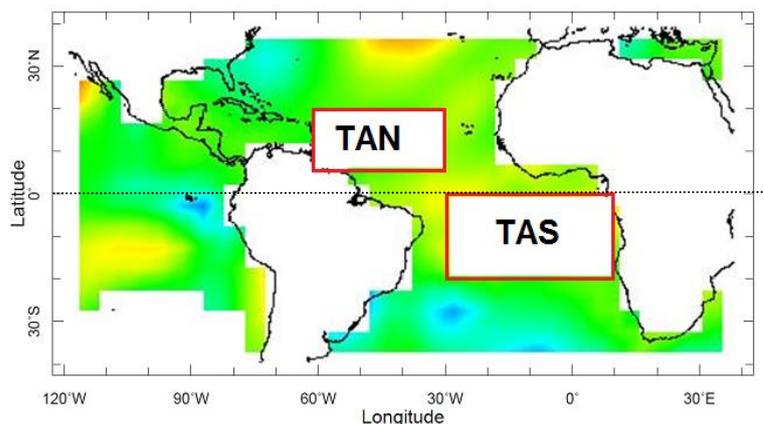


Figura 6.7 - Localização das zonas das regiões que constituem o DIPOLO do Atlântico, o dipolo é definido como a diferença aritmética das anomalias da SST do Atlântico norte tropical (TAN) e as SST do Atlântico sul tropical (TAS), (IRI, 2012 a)

Em anos nos quais a SST sobre o TAS (entre a linha do Equador e 15°S) encontra-se mais alta que a média de longo período, durante março-abril-maio, e o TAN (entre 12°N e 20°N) está menos aquecido do que a média, há a formação de um gradiente meridional de anomalias de SST no sentido de norte para sul. Nessa situação observa-se, simultaneamente, uma pressão no nível do mar (PNM) mais baixa do que a média sobre o Atlântico sul e mais alta do que a média sobre o Atlântico norte (alta pressão no Atlântico norte, AAN na Figura 6.8 a). Os alísios de sudeste mais fracos do que a média e os alísios de nordeste mais intensos do que a média. O eixo de baixa pressão à superfície e confluência dos ventos alísios deslocando-se mais para sul, relativamente ao seu posicionamento médio, e totais pluviométricos acima da média sobre o norte do NEB (Marengo, 2006).

O Atlântico pode, em algumas ocasiões, apresentar variações de SST em um modo que se assemelha à variabilidade do ENSO em vários aspectos, devido ao fato das variações anuais do Pacífico e do Atlântico serem similares. A variabilidade interanual no Pacífico e Atlântico relaciona-se com deslocamentos da ZCIT. A oscilação sul possui um componente adicional que o Atlântico não apresenta: o deslocamento leste-oeste da convergência sobre o Pacífico tropical do oeste. Para explicar o que acontece com a variabilidade no Atlântico tropical, devem ser consideradas as mudanças na circulação global da atmosfera. Fatores similares devem afetar o Pacífico, e esse é um possível motivo da frequência de episódios *El Niño* oscilarem na escala de tempo decenal (Marengo, 2006).

A relação entre o ENSO e o oceano Atlântico também é mencionada por Alves *et al.* (1997), eles sinalam que o *El Niño*, quando acontece conjuntamente com um valor positivo do DIPOLO do Atlântico, tem anos secos ou muito secos no NEB. A relação inversa ocorre com o fenômeno *La Niña* associado a um valor DIPOLO do Atlântico negativo, é normalmente responsável por anos considerados normais, chuvosos ou muito chuvosos na região. Por fim, quando as águas do Pacífico estão em condições normais e o dipolo do Atlântico negativo (positivo), a probabilidade de se ter um ano normal ou chuvoso (seco) também é alta. Essa consideração também é apresentada por Lima (2010), que chegou a essa conclusão ao analisar a precipitação anual em Fortaleza, Ceará.

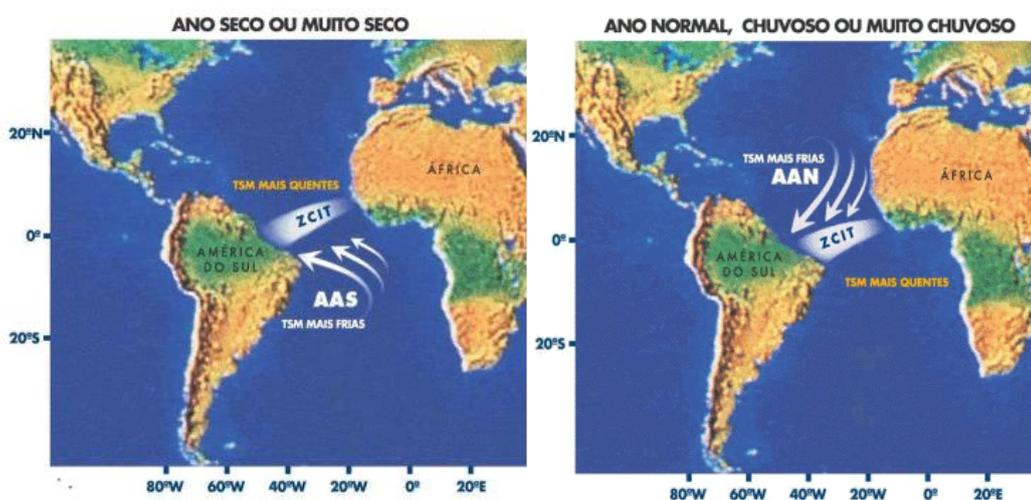


Figura 6.8 – Padrões atmosféricos e SST no Nordeste produzidos pelo DIPOLO do Atlântico: a) períodos secos e b) chuvosos no Nordeste (Ferreira e Mello, 2005)

### 6.3.3 - Variabilidade Decenal e de Longo Prazo

Historicamente, a região do NEB sempre foi afetada por grandes secas ou grandes cheias, que estatisticamente acontecem de 18 a 20 anos a cada 100 anos. Segundo Kane (1989), para a região no período 1849-1985 (137 anos) os 29 anos de *El Niño*, apenas 12 foram associados às secas na região. Wagner (1996) estudou no Atlântico tropical e no NEB, tendências decadais nos mecanismos que controlam o gradiente meridional da SST nessa região. Posteriormente, Hastenrath (2000) identificou tendência de longo prazo na chuva do Nordeste e do setor do Atlântico tropical adjacente, caracterizado por um deslocamento da ZCIT e banda de chuvas mais para o sul da sua posição climatológica. Isso poderia explicar as tendências positivas de chuva no NEB. Esses trabalhos detectaram um aquecimento sistemático do Atlântico tropical sul, observado principalmente nos meses de

verão (fevereiro/março), enquanto que a SST no Atlântico norte aumenta no inverno (agosto/ setembro) e diminui no verão. Como consequência do incremento do gradiente meridional de SST no verão, a ZCIT se desloca mais para o Sul e a chuva no norte do NEB tende a incrementar no período 1951-90 (Marengo e Valverde, 2007)

A partir da década de 1970, o volume de chuvas tem sido menor em relação a outros anteriores e especificamente ao ano de 1985, que foi muito úmido. Esta variabilidade também tem sido observada nas vazões do rio São Francisco em Sobradinho na Bahia, onde a tendência relativamente positiva desde 1931 contrasta com a tendência negativa observada a partir de 1979. Ainda que essa queda de vazões possa estar associada em parte a essa variabilidade de chuva, também poderia estar associada ao uso de água para irrigação e outras formas de uso (Marengo e Valverde, 2007). Para sumarizar a variabilidade no longo prazo, a Tabela 6.2 apresenta a ocorrência de secas nos últimos quatro séculos, além disso, Marengo (2008) menciona que secas menores ocorreram nos anos 2003 e 2005.

Tabela 6.2 - Anos de seca no Nordeste brasileiro, coincidentes com anos de *El Niño*, durante os últimos quatro séculos atualizado para 1998 (Marengo, 2006).

Século XVII	Século XVIII	Século XIX	Século XX
1603	1711	1804	1900
1614	1721	1809	1902
1692	1723/24	1810	1907
-	1736/37	1816/17	1915
-	1754	1824/25	1919
-	1760	1827	1932/33
-	1772	1830/33	1936
-	1776/7	1845	1941/44
-	1784	1877/79	1951
-	1790/94	1888/89	1953
-	-	1891	1958
-	-	1898	1970
-	-	-	1979/80
-	-	-	1981
-	-	-	1982/83
-	-	-	1986/87
-	-	-	1991/92/97/98

#### **6.3.4 - Variabilidade climática e o regime das vazões no Nordeste do Brasil**

Conforme visto anteriormente, a variabilidade climática na região do NEB influi fortemente na variabilidade da precipitação. Vários trabalhos tem tratado esse assunto (Hastenrath e Heller, 1977; Moura e Shukla, 1981; Hastenrath, 1994, 2000, 2006; Wagner, 1996; Uvo *et al.*, 1998; Souza Filho e Lall, 2003; Robertson *et al.*, 2004; Marengo, 2006; Marengo e Valverde, 2007; Andreoli e Kayano, 2007; Kayano e Andreoli, (2006); Marengo, 2008; Lima, 2010; Ito e Coelho, 2012) e demonstraram a influência da variabilidade climática na precipitação.

Em seus trabalhos, Hastenrath (1994, 2000) fez uma análise da dinâmica dos padrões atmosféricos na região. Hastenrath (2006), em referencia aos seus trabalhos anteriores, analisou a ocorrência a longo prazo das secas, determinando que as variações da ZCIT controlam essa ocorrência. Uvo *et al.* (1998) realizou uma análise da precipitação pelo comportamento da ZCIT com as SST dos oceanos Pacífico e Atlântico. Eles encontraram uma alta correlação do período chuvoso com índices do ENSO propostos nesse trabalho. Além disso, encontraram que esse período também esteve correlacionado com o DIPOLo do Atlântico. Marengo (2006); Marengo e Valverde (2007); Marengo (2008) recopilam trabalhos e informações da variabilidade da precipitação, além de indicar a variabilidade associada a tendências de mudanças climáticas e fazem um análise da vulnerabilidade da região. Eles mencionam que para o NEB existem altas correlações positivas entre a precipitação e o SOI, o que implica menores precipitações na região. Essas correlações tendem ser menores que com a chuva nos caso das vazões, mas, continuam sendo significativas.

Os trabalhos de Andreoli e Kayano (2007); Kayano e Andreoli (2006) realizaram as relações dos oceanos pacífico e atlântico com a estação chuvosa e a ocorrência de secas. Lima (2010) modelou a formação e persistência de períodos secos e não secos de precipitação em Fortaleza. Ito e Coelho (2012) analisaram a padrões atmosféricos de grande escala junto com as relações do ENSO nas variações da precipitação no Nordeste, especificamente eles analisaram as secas de 2010 e 2012 e os períodos chuvosos de 2009 e 2011. Eles encontraram que os eventos secos estiveram relacionados com o ENSO e os períodos úmidos com as anomalias das SST no oceano Atlântico, mas também a seca de 2012 esteve relacionada com a SST do Atlântico tropical.

Reconhecida a variabilidade da precipitação, as vazões, por relacionar-se à precipitação, também exibem variações na região. As relações com o ENSO e as SST do oceano Atlântico tropical, também aplicam para as vazões (Souza Filho e Lall, 2003, Marengo, 2006; Whiting, 2006). Por exemplo, Souza Filho e Lall, (2003) utilizaram os resultados de Uvo *et al.* (1998) para analisar a relação entre os índices do ENSO e o DIPOLO do Atlântico com vazões de vários locais, entre eles as vazões de entrada do reservatório Orós, no Ceará. Eles encontraram significativas correlações entre o DIPOLO e o NINO3 com as vazões. Utilizaram o NINO3 tomando como referência os resultados da correlação desse índice com o período chuvoso no NEB, encontradas no trabalho de Chiang *et al.* (2000), que também mencionou a influencia das SST do Atlântico. Lima (2010) também utilizou esses índices, o mesmo que Alexandre (2012) e Ribeiro (2011). Ribeiro (2011) encontrou uma correlação semelhante à de Souza Filho e Lall (2003).

Em vista das considerações tratadas neste capítulo, existe uma variabilidade climática importante no NEB. Portanto, é importante, na modelagem de séries de vazões, incluir informação climática que represente essa variabilidade. Logo, neste estudo serão tratados modelos que representam variabilidade de estados climáticos e incluem informação climática, como os vistos no capítulo 5.

## 7 - METODOLOGIA

Neste capítulo, apresenta-se a estratégia metodológica que foi utilizada para a modelagem de uma série histórica de vazões anuais, assim como também são apresentados os métodos e materiais que foram utilizados.

### 7.1 - ESTRATÉGIA METODOLÓGICA

A estratégia metodológica que foi proposta para a geração de séries de vazões anuais consistiu de cinco etapas:

- 1) Análise da série observada de vazões (série  $Q$ );
- 2) Ajuste de modelos ARMA ( $p, q$ ) e HMM com até 6 estados ocultos;
- 3) Comparação das séries geradas através da distribuição Gama e as simuladas pelos modelos ARMA e HMM;
- 4) Ajuste de modelos NHMM;
- 5) Comparação das séries sintéticas obtidas por meio da distribuição Gama e as geradas pelos modelos ARMA e NHMM.

A Figura 7.1 apresenta o fluxograma da metodologia. Pode-se observar que cada etapa está composta por várias sub-etapas que explicam melhor os procedimentos realizados.

Assim, o estudo consistiu em ajustar vários modelos estocásticos à série de  $Q$  para gerar séries sintéticas que simularam adequadamente a persistência observada nela. Assim, primeiramente analisou-se essa série. Depois, ajustaram-se vários modelos ARMA, para então escolher o melhor modelo ARMA. Seguidamente ajustaram-se vários HMM e compararam-se com os modelos ARMA e a distribuição Gama para analisar a utilidade dos HMM em representar a  $Q$ . Logo, realizou-se uma análise de correlação entre índices e a série observada. Com isso, ajustaram-se vários NHMM, com o intuito de obter modelos que simulem a persistência observada utilizando informação climática a partir do resultado da análise de correlação. Finalmente, foram comparadas as séries geradas pelos NHMM com os modelos ARMA e as séries produzidas pela distribuição Gama. Com relação aos tipos de modelos HMM que foram utilizados, eles definem-se a seguir.

### 7.2.1 - Definição dos Modelos HMM propostos

Foram ajustados duas classes de modelos HMM à série  $Q$ , os quais são:

- 1) O modelo HMM-NLT, que consistiu em um modelo o qual utilizou uma Log-transformação (LT) nas vazões, portanto, utilizou a série  $Y = \ln(Q)$  e teve distribuições Normais dependentes dos estados ocultos; e
- 2) O modelo HMM-G que utilizou diretamente a série  $Q$ , com distribuições Gama dependentes dos estados ocultos.

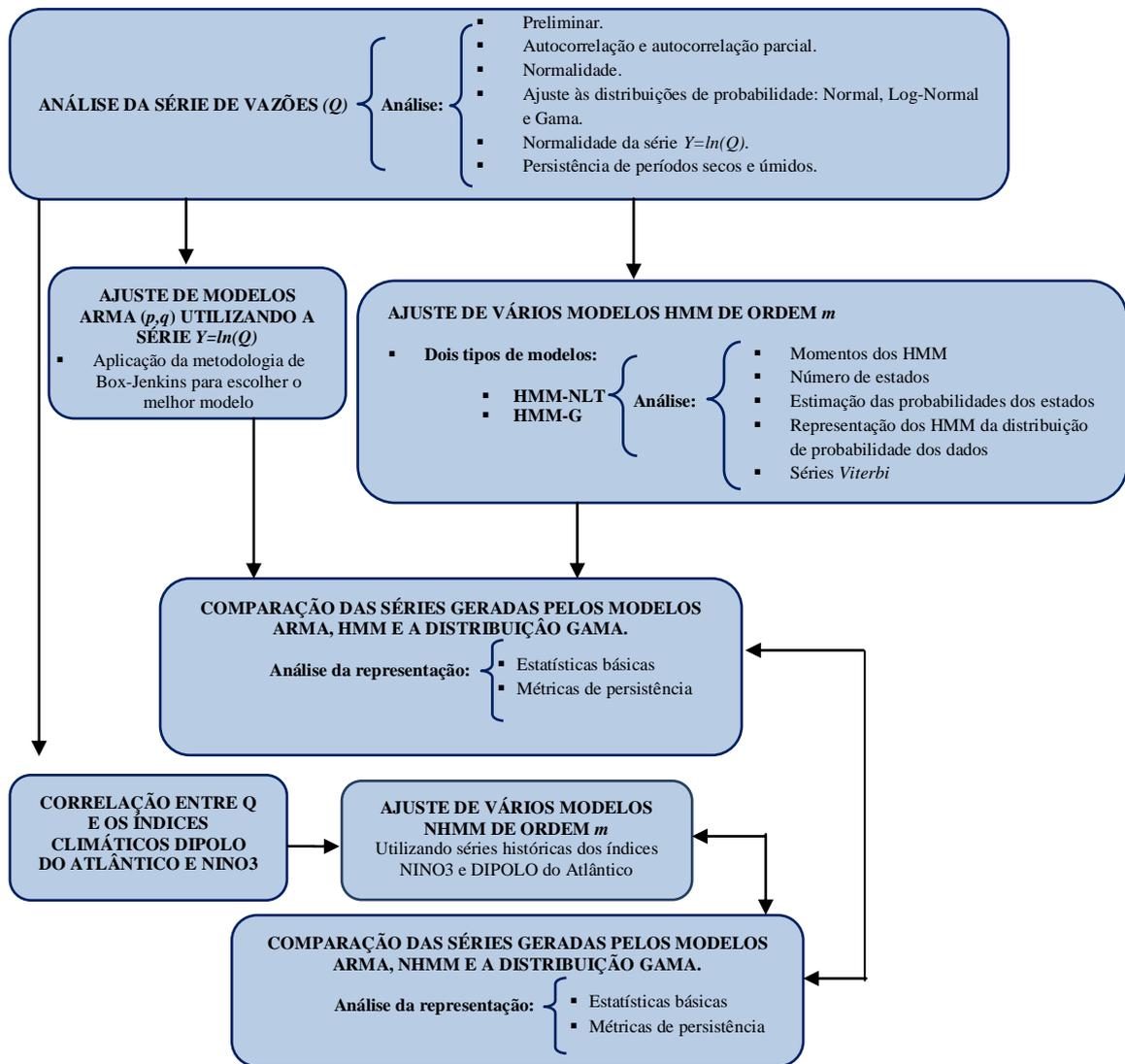


Figura 7.1 - Estratégia metodologia

## 7.2 - ESTATÍSTICAS PARA IDENTIFICAR A PERSISTÊNCIA

A seguir são definidas as estatísticas que foram utilizadas nesta dissertação para a análise de persistência hidrológica. Persistência que foi definida como a análise dos comprimentos de períodos úmidos e secos, definidos pelos percentis 33% e 66% utilizados no trabalho de Lima (2010); e a mediana, utilizada por Prairie *et al.* (2008) e Whiting (2006). Esse último autor menciona que a escolha da mediana reduz a influência de dados assimétricos sobre o comprimento e magnitude dos períodos hidrológicos.

### 7.2.1 - Comprimentos e Volumes de Períodos Hidrológicos

Nos trabalhos de Sveinsson *et al.* (2003) e Prairie *et al.* (2008) foram definidas várias métricas para analisar os períodos úmidos e secos na série observada, para testar o desempenho de modelos estocásticos para simular essas estatísticas, para obter séries com variabilidade entre estados e persistência hidrológica. A Figura 7.2 apresenta as definições feitas por Prairie *et al.* (2008) para analisar a persistência e variabilidade em séries de vazões. Na figura, os períodos secos e úmidos são definidos por um limiar dado pela mediana. Logo, um período seco define-se como um período onde as vazões encontram-se abaixo da mediana e o período úmido o contrário, sendo definido pelas vazões que excedem a mediana. Portanto, um comprimento seco é a longitude onde as vazões permanecem abaixo da mediana, conformando um período seco. Esse comprimento é dado em anos para o caso de vazões anuais. Do mesmo modo, define-se para os períodos úmidos.

A Figura 7.2 também apresenta a definição dos volumes desses períodos. É claro que o volume úmido corresponde ao volume de excedência que pertence a um período úmido. No caso do volume seco, esse volume não corresponde ao volume das vazões que conformam o período seco, em vez disso, corresponde ao volume de déficit, como o mostrado na figura, sendo então, o volume necessário para que o período alcance à mediana. As análises dos trabalhos de Sveinsson *et al.* (2003) e Prairie *et al.* (2008) consistiram em avaliar os valores máximos, mínimos, médios e a somatória dos máximos valores dessas métricas, além de realizar a análise das distribuições dos comprimentos. Uma análise similar foi realizada nesta dissertação.

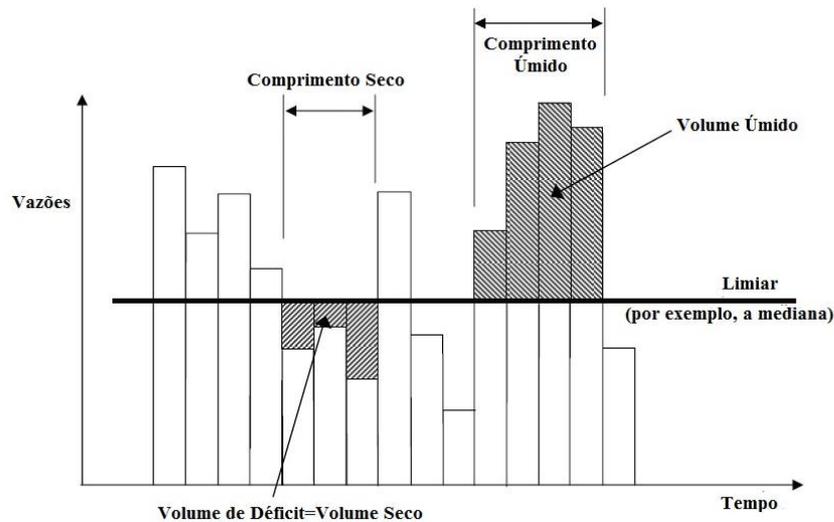


Figura 7.2 - Definição de Estatísticas de períodos secos e úmidos (Prairie *et al.* 2008, traduzido)

### 7.2.2 - Número total de comprimentos

Lima (2010) definiu um estatística para contabilizar o número total de comprimentos de diferentes magnitudes em uma série, permitindo medir os períodos secos ou úmidos de uma série observada. Matematicamente essa estatística é definida como:

$$T = \sum_{l=2}^6 n_c(l) \quad (7.1)$$

Onde,  $n_c(l)$  indica o número total de clusters com  $l$  anos consecutivos no estado seco. Por exemplo, se existe uma série com quatro anos consecutivos de eventos seco, o valor de  $T$  será:

$$T = n_c(2) + n_c(3) + n_c(4) + n_c(5) + n_c(6) = 3 + 2 + 1 + 0 + 0 = 6$$

### 7.3 - TESTE DE ADERÊNCIA DO QUI-QUADRADO ( $\chi^2$ )

Comumente na análise de séries temporais, o um teste de aderência é utilizado em modelos estocásticos univariados no qual as variáveis sejam Normalmente distribuídas, embora, sejam utilizadas outras distribuições. Para determinar que distribuição ajusta-se a uma série

histórica, pode-se realizar uma análise gráfica. No que refere à distribuição Normal, analisa-se a forma do histograma de frequências ou diagrama de caixa, o que além de dar uma tentativa da distribuição a ajustar, oferecem uma representação gráfica dos dados. Outra forma é plotar os valores da probabilidade normal das observações e avaliar a sua linearidade. Além disso, existem testes de ajuste ou aderência que são válidos para qualquer distribuição de probabilidade, entre esses testes estão o Qui-quadrado e o Kolmogorov – Smirnov. O teste do Qui-Quadrado é um dos testes mais utilizados e, por isso, foi o teste aplicado nessa dissertação. Ele consiste em comparar a frequência observada dos dados,  $O_i$ , com a frequência esperada,  $E_i$ , segundo a distribuição de probabilidade escolhida mediante ordenação dos dados em  $n$  intervalos de classes. Assim, avaliam-se as hipóteses  $H_0$ : os dados seguem a distribuição de probabilidade escolhida.  $H_1$ : os dados não seguem a distribuição de probabilidade escolhida. Utilizando o valor de uma estatística calculada  $\chi_c^2$ , denominada Qui-Quadrado definida como (Farias, 2003):

$$\chi_c^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (7.2)$$

Logo, essa estatística compara-se com a estatística Qui-Quadrado  $\chi_{1-\alpha, \nu}^2$  tabelada, onde  $\alpha$  é o nível de confiança que pode ser o 95 % e  $\nu = n - k - 1$ , que representa os graus de liberdade, sendo  $k$  o número de parâmetros calculados da distribuição a ser ajustada. Assim, se  $\chi_c^2 < \chi_{1-\alpha, \nu}^2$ , aceita-se a hipótese  $H_0$ , caso contrário, aceita-se a hipótese  $H_1$ . Os procedimentos para calcular esse teste foram tomados de IME USP (2012).

#### 7.4 - MVN-HMM TOOLBOX

Como foi mencionado no capítulo 4, para o ajuste dos modelos HMM utilizou-se o *software MultiVariate Nonhomogeneous Hidden Markov Model toolbox (MVN-HMM toolbox)*, desenvolvido por Kirshner (2005) e encontra-se disponível no site: <http://www.stat.purdue.edu/~skirshne/>. O *software* é uma coleção de algoritmos implementados na linguagem C++, permitindo realizar várias opções para o ajuste e análise de dados. As três ações principais são: *Learn*, *Viterbi* e *Simulation*, ações que foram utilizadas neste trabalho. A ação *Learn* corresponde à estimação dos parâmetros necessários do modelo empregando o algoritmo Baum-Welch. Na ação *Viterbi* estima-se a

sequência de estados mais provável (decodificação global) e a última ação, *Simulation*, realiza as simulações de dados utilizando o HMM ajustado. Maiores informações e detalhes sobre o funcionamento do MVN-HMM *toolbox* se podem encontrar em Kirshner (2005a) e Kirshner (2005b).

## 7.5 - CRITÉRIOS DE SELEÇÃO DE MODELOS ESTOCÁSTICOS

Várias técnicas existem para a escolha do modelo adequado ou modelo que melhor represente as observações. No caso dos modelos ARMA, é comumente utilizada a metodologia de Box de Box e Jenkins (1976) (ver Bras e Rodriguez-Iturbe, 1992). Para identificar os modelos ARMA, escolhem-se os modelos candidatos que melhor ajustam-se aos dados e, posteriormente, escolhe-se o melhor modelo. Uma abordagem incluída nessa metodologia consiste em utilizar vários de critérios como: o *Akaike Information Criterion* (AIC) (Akaike, 1974) e o *Bayesian Information Criterion* (BIC) (Schwarz, 1978). Ambos os critérios são baseados na função de máxima verossimilhança e a diferença entre esses critérios está na forma de penalizar os modelos. Essa penalização baseia-se no número de parâmetros sendo mais forte o BIC, que penaliza severamente a modelos com muitos parâmetros (Bellone *et al.* 2000; Storch e Zwiers, 1999). Logo, o melhor modelo é aquele que apresenta os menores valores de AIC e BIC. Esses critérios são definidos como:

$$AIC = -2 \ln(L) + 2k \quad (7.3)$$

$$BIC = -2 \ln(L) + k \ln(N) \quad (7.4)$$

Onde:

$k$  = Numero de parâmetros do modelo. Nos HMM,  $k$  é o numero de parâmetros independentes, entregados pelo MVN-HMM *toolbox*;

$L$  = Logaritmo do valor obtido de maximizar a função de máxima verossimilhança;

$N$  = Número de observações.

Para os modelos HMM; o interesse está em saber o número de estados  $m$  adequados. Nesse caso, uma abordagem semelhante aos modelos ARMA e os critérios AIC e BIC são utilizados. Existem várias preferências sobre um critério ou outro, por exemplo, para Prairie *et al.* (2008) é mais adequado o AIC para modelos HMM. Entanto que, para

Akintug e Rasmussen (2005), o BIC é mais adequado para essa classe de modelos. Geralmente, ambos os critérios são utilizados, variando a preferência entre autores, notando a importância do peso que tem o BIC, visto que um HMM com muitos estados apresenta demasiados parâmetros. No caso de Bracken (2011), utilizou os dois critérios para escolher os estados dos HMM que utilizou.

Zucchini e MacDonald (2009) recomendam o uso de ambos os critérios. Mencionam também que em algumas aplicações o BIC tende a não ser aplicado. Eles indicam a importância da análise da ACF que produzem os HMM de vários estados comparada com a ACF das observações. Também ressaltam a importância da análise da representação dos estados observados para cada HMM e ajuste das distribuições à distribuição conjunta das observações, apreciação que também fazem Akintug and Rasmussen (2005). Mas, com tudo isso, é importante analisar se os estados escolhidos fisicamente são possíveis ou se tem alguma interpretação física (Greene *et al.*, 2008). No caso de Robertson *et al.* (2004), a escolha de quatro estados ocultos foi analisada adequadamente, concluindo que cada estado representa uma característica específica para a ocorrência de chuva, sendo cada estado realmente possível.

## 8 - ESTUDO DE CASO

### 8.1 - LOCAL DE APLICAÇÃO

O estudo objetivou analisar a série histórica das vazões de entrada ao reservatório Orós, localizado na bacia do Rio Jaguaribe, no Estado do Ceará, no Nordeste do Brasil (NEB). A bacia ocupa 51.9 % da área total do estado e localiza-se entre as coordenadas 4°3'00'' S - 7°45'00''S e 37°30'00''W- 41°00'00''W. O reservatório está localizado dentre os municípios Quixelô, Iguatu e Orós, a aproximadamente 270 km de Fortaleza (em linha reta) (Figura 8.1). O reservatório Orós foi construído em 1961 e seu objetivo principal é o armazenamento de água para o período de secas, mas também faz parte do Sistema Metropolitano-Jaguaribe que serve de abastecimento para a Região Metropolitana de Fortaleza (RMF) (Gatto, 1999; COGERH, 2011). A Tabela 8.1 apresenta algumas das características principais do reservatório.

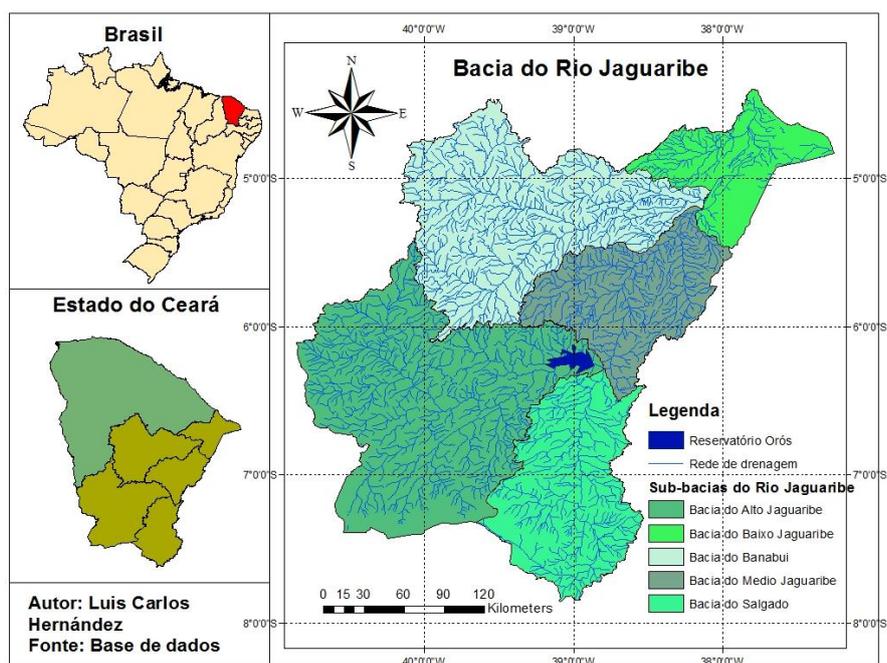


Figura 8.1 - Localização do reservatório Orós (Fonte: Base de dados)

Tabela 8.1- Características do reservatório Orós (COGERH, 2011)

Bacia	Capacidade (m <sup>3</sup> )	Vazão Regularizada (m <sup>3</sup> /s)	Área de drenagem (km <sup>2</sup> )
Alto Jaguaribe	1940000000	20.40	24538

## 8.2 - DADOS UTILIZADOS

As vazões médias anuais utilizadas nesta dissertação para a geração de séries sintéticas de vazões anuais foram obtidas das vazões médias mensais afluentes no reservatório Orós para o período 1911-2000. Os dados dos valores mensais das séries dos índices climáticos NINO3 e DIPOLO do Atlântico para o período 1910-2000, utilizados como *inputs* dos modelos NHMM foram obtidos no site do *International Research Institute for Climate and Society* (IRI), dados que correspondem a uma versão estendida dos trabalhos de Reynolds e Smith (1994) e Kaplan *et al.* (1998). O site do IRI entrega valores de anomalias da temperatura da superfície do mar (*Surface Sea Temperature-SST*), espaçados geograficamente 5° longitudinal e latitudinalmente. Logo, cada série climática foi conformada utilizando anomalias da SST, dependendo das coordenadas geográficas que definem o índice. Então, para umas coordenadas específicas se teve uma grid com vários valores de SST e o valor mensal foi obtido como uma média desses valores, com exceção da série do índice NINO3, que o site entrega diretamente os valores mensais. A seguir são apresentadas as considerações que foram realizadas para obter cada uma dessas séries climáticas.

### 8.2.1 - Dados dos Índices Climáticos

As anomalias da SST para construir a série mensal do DIPOLO do Atlântico foram tomadas do site do IRI, disponíveis no seguinte link <http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/>. Primeiramente, para a série de anomalias da SST do Oceano Atlântico norte tropical (TAN) foi utilizada uma grid formada pelas coordenadas 2.5°N - 27.5N, 67.5°W - 27.5°W que incluem as coordenadas que definem a região TAN. A grid esteve composta por nove pontos longitudinalmente e seis pontos latitudinalmente para um total de 54 pontos para cada mês, cada ponto com um valor de anomalia. Para a série de anomalias da SST do Oceano Atlântico sul tropical (TAS) foi utilizada a uma grid formada pelas coordenadas 2.5° N - 27.5°S, 37.5°W - 17.5°E, que incluem as coordenadas que definem a zona TAS. A grid é composta por 12 pontos longitudinalmente e 7 pontos latitudinalmente para um total de 84 valores de SST para cada mês. Logo, o DIPOLO foi obtido como a diferença aritmética entre essas duas anomalias, (Souza Filho e Lall, 2003; Brabo Alves *et al.*, 2009). Por ultimo, os valores médios mensais do índice climático NINO3 foram obtidos do site do IRI

disponíveis pelo seguinte endereço:  
[http://iridl.ldeo.columbia.edu/SOURCES/.Indices/.nino/.EXTENDED/.NINO3/.](http://iridl.ldeo.columbia.edu/SOURCES/.Indices/.nino/.EXTENDED/.NINO3/)

## 9 - RESULTADOS E DISCUSSÃO

Muitos dos resultados foram obtidos utilizando o *software MATLAB*, calculados por meio das funções próprias desse *software* ou criando um código (*script*) específico, para se obter um determinado resultado. No caso dos Modelos de Markov Oculto (HMMs), como mencionado anteriormente, foi utilizado o *software MVN-HMM toolbox* para calcular: os parâmetros, a sequência *Viterbi* e fazer as simulações. Os modelos Autoregressivos de Médias Móveis (ARMA) foram ajustados por meio do *MATLAB*. Assim, os resultados obtidos do estudo são apresentados nas seguintes seções, assim como a sua discussão.

### 9.1 - ANÁLISE DA SÉRIE DE VAZÕES OBSERVADAS

Esta seção apresenta uma descrição da série de vazões anuais observadas, incluindo um conjunto de estatísticas descritivas, tais como média, desvio-padrão e coeficiente de assimetria, assim como a própria função de autocorrelação amostral. Além disso, uma descrição da persistência de períodos úmidos e secos também é apresentada e que servirá de base para a análise de desempenho dos modelos avaliados.

A Figura 9.1 apresenta a série de vazões anuais afluentes ao reservatório Orós no período entre 1911 e 2000. Observam-se anos seguidos com vazões baixas, como na década de 1950, e anos com vazões mais elevadas, como na década de 1980.

O histograma das vazões apresentado na Figura 9.2 mostra que a maioria dos valores observados de vazão são inferiores a  $25 \text{ m}^3/\text{s}$ , porém com a ocorrência de valores acima de  $150 \text{ m}^3/\text{s}$ , algo típico de regiões semiáridas, onde o coeficiente de variação costuma ser bastante elevado.

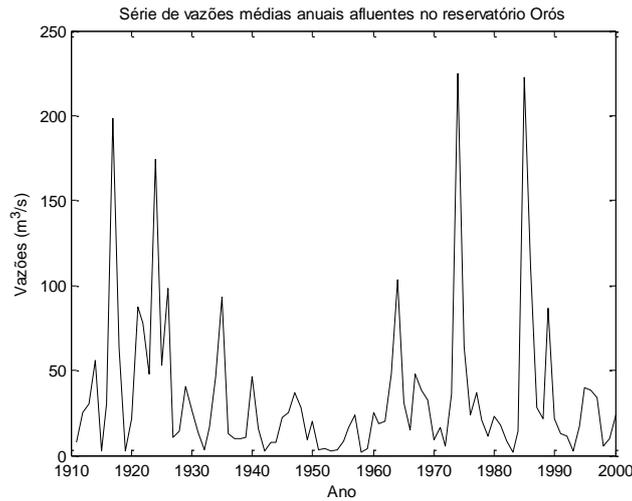


Figura 9.1 - Série de vazões observadas afluentes no reservatório Orós. Período 1911-2000.

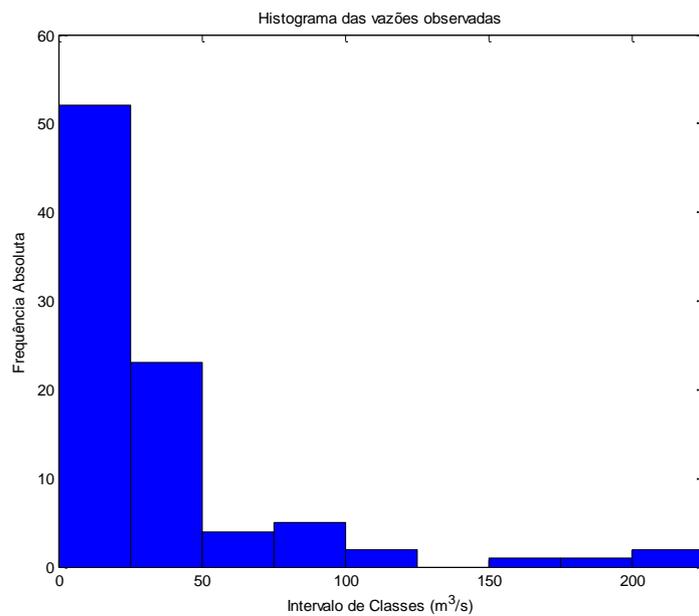


Figura 9.2 - Histograma das vazões observadas.

A Tabela 9.1 fornece um resumo das principais estatísticas descritivas da série anual afluente ao reservatório Orós. Com média amostral de aproximadamente  $35 \text{ m}^3/\text{s}$ , e desvio-padrão de quase  $45 \text{ m}^3/\text{s}$ , o coeficiente de variação do rio Jaguaribe neste local é de aproximadamente 1.27, um valor extremamente elevado, muito maior do que em outros locais do mundo. O bloco de caixa, apresentado na Figura 9.3, permite visualizar o elevado grau de assimetria positiva da distribuição empírica de vazões, assim como o elevado número de anos em que as vazões são consideradas “outliers”.

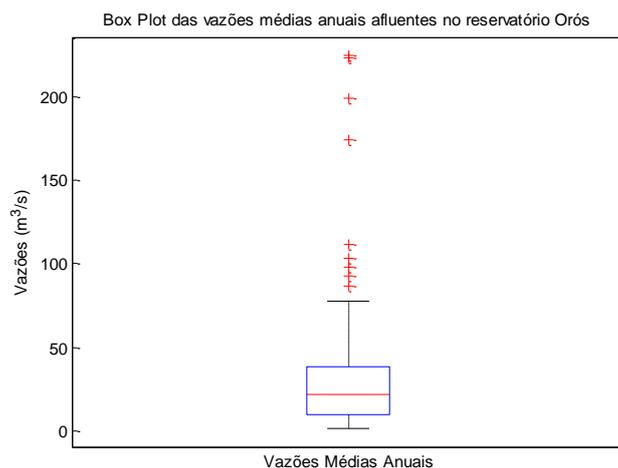


Figura 9.3 - Diagrama *Box Plot* para as vazões médias anuais afluentes no reservatório Orós. Período 1911-2000.

Tabela 9.1 - Estatísticas descritivas da série de vazões observada

<b>Estatísticas descritivas das vazões médias anuais afluentes ao reservatório Orós. Período 1911-2000.</b>				
<b>Estatística Amostral</b>	<b>Notação</b>	<b>Valor</b>	<b>Unidades</b>	<b>Cálculo</b>
Média	$Q_m$	35.059	$m^3/s$	Equação (5.4)
Mediana	$Q_{md}$	21.625	$m^3/s$	Equação (5.7)
Vazão Máxima	$Q_{max}$	224.550	$m^3/s$	Valor máximo
Vazão Mínima	$Q_{min}$	1.810	$m^3/s$	Valor mínimo
Amplitude	$A$	222.740	$m^3/s$	$Q_{max} - Q_{min}$
Primeiro Quartil	$Q_1$	10.020	$m^3/s$	Equação (5.7)
Terceiro Quartil	$Q_3$	38.700	$m^3/s$	Equação (5.7)
Amplitude Inter - Quartis	$AIQ$	28.680	$m^3/s$	$Q_3 - Q_1$
Variância	$S_Q^2$	1983.400	$(m^3/s)^2$	Equação (5.5)
Desvio padrão	$S_Q$	44.535	$m^3/s$	$\sqrt{S_Q^2}$
Coefficiente de Assimetria	$Q_{Cs}$	2.729	Adimensional	Equação (5.8)

A análise da autocorrelação de  $Q$  foi realizada mediante o cálculo das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), ambas apresentadas na Figura 9.4. Essas figuras também mostram uma baixa dependência de longo prazo, talvez pode estar relacionada a um valor baixo do coeficiente de Hurst (Hurst, 1951). Mas, aqui esse valor não foi calculado.

Pode-se observar que a autocorrelação com “lags” 1 e 11 são estatisticamente diferentes de zero. O mesmo ocorre com a autocorrelação parcial. Esses valores da ACF e a PACF, foram utilizados como valores  $p$  e  $q$  para identificar o grau dos modelos ARMA, dado que a ACF entrega valores  $q$  para identificar modelos MA e, similarmente, a PACF entrega valores de  $p$  para identificar modelos AR. Com isso, é possível ajustar os modelos ARMA (Bras e Rodriguez-Iturbe, 1992). Assim, foram testadas diferentes combinações de valores  $p$  e  $q$  dos modelos ARMA.

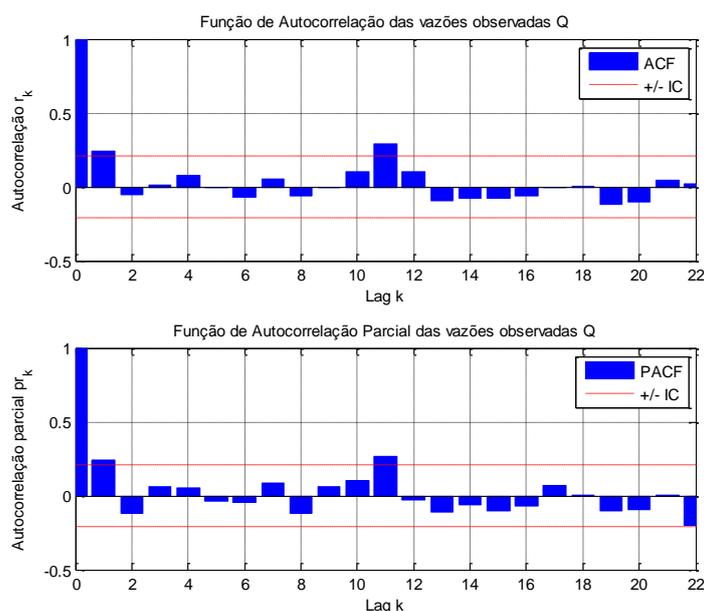


Figura 9.4 - Função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) para a série de vazões observadas.

Procurou-se avaliar qual seria a distribuição teórica de probabilidades que poderia ser empregada para modelar a série em questão. Num primeiro momento, decidiu-se verificar a adequabilidade da distribuição Normal, desconfiando-se, de antemão, com base nas estatísticas descritivas, que a mesma não seria adequada. De todo modo, foi realizado um teste de hipóteses, baseado no teste de aderência Qui-Quadrado. O resultado mostrou que há evidências contundentes (valor  $p = 3.42 \times 10^{-4}$ ) de que a hipótese nula é falsa, concluindo-se que a distribuição Normal não é indicada para modelar a série de vazões anuais afluentes ao reservatório Orós.

Com o intuito de encontrar a melhor distribuição que se ajustasse às observações, procedeu-se ao ajuste e teste de aderência das distribuições Log-Normal e Gama, o que

daria uma ideia melhor sobre as distribuições candidatas a serem empregadas no ajuste do modelo HMM. O ajuste da distribuição Log-Normal e Gama foi realizado pelo Método dos Momentos. Os resultados do teste do Qui-Quadrado indicaram que não existem evidências suficientes para rejeitar a hipótese nula de que a série anual de vazões afluentes ao reservatório Orós possa ser modelada pelas distribuições Log-Normal (valor  $p = 0.976$ ) e Gama (valor  $p = 0.484$ ). Os resultados dos parâmetros estimados e seus intervalos de confiança para o ajuste dessas distribuições apresentam-se na Tabela 9.2. Utilizando esses valores foi construída a função de densidade de probabilidade (PDF) Normal para a série de vazões observadas (Figura 9.5). Pode-se notar claramente a não adequabilidade da distribuição Normal para a série em questão, com probabilidade alta de se obter vazões negativas. A Figura 9.6 e a Figura 9.7 apresentam, respectivamente, as funções densidade de probabilidade ajustadas da distribuição Log-Normal e Gama aos dados.

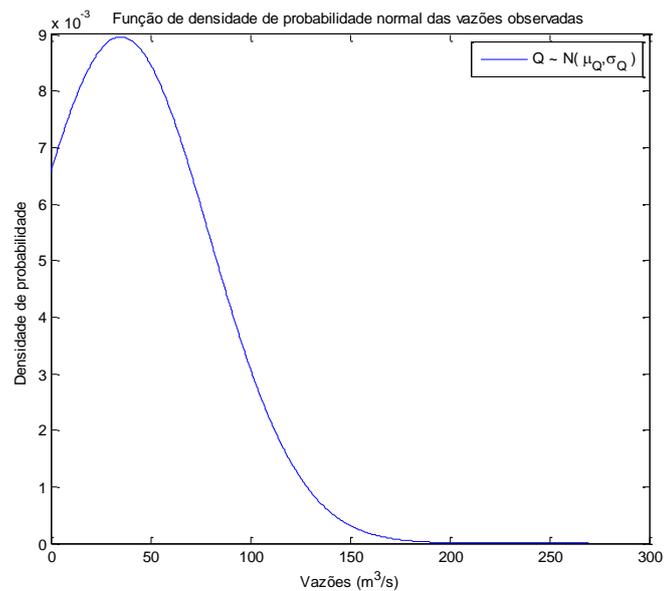


Figura 9.5 – Função de densidade de probabilidade normal das vazões observadas.

Tabela 9.2 - Parâmetros e intervalos de confiança de 95% estimados para a distribuição Normal, Log-Normal e Gama das vazões observadas.

Parâmetro	Distribuição	Valor	Intervalos de confiança de 95 %		Unidades
			Limite Inferior	Limite Superior	
Média de $Q$ ( $\mu_Q$ )	Normal	35.059	25.731	44.387	m <sup>3</sup> /s
Desvio padrão de $Q$ ( $\sigma_Q$ )	Normal	44.535	38.844	52.196	m <sup>3</sup> /s
Média de $Y$ ( $\mu_Y$ )	Log-Normal	2.969	2.733	3.205	m <sup>3</sup> /s
Desvio padrão de $Y$ ( $\sigma_Y$ )	Log-Normal	1.126	0.982	1.320	m <sup>3</sup> /s
Parâmetro de forma ( $\alpha$ )	Gama	0.983	0.760	1.271	Adimensional
Parâmetro de taxa ( $\beta$ )	Gama	35.663	25.620	49.644	m <sup>3</sup> /s

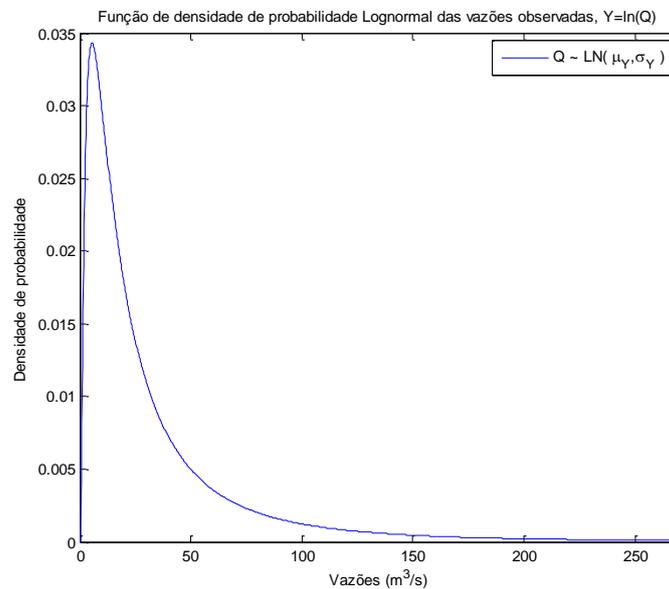


Figura 9.6 - Função de densidade de probabilidade log-normal das vazões observadas.

Além de fornecer as estatísticas descritivas da série e identificar distribuições teóricas de probabilidade que são candidatas à sua modelagem, é imprescindível descrever o grau de persistência da série observada. De acordo com o que foi mencionado na seção 7.2, decidiu-se por utilizar três limiares para análise de persistência da série, o percentil de 33% ( $P_{33}$ ), o percentil de 66% ( $P_{66}$ ) e a mediana ( $Q_{md}$ ). Sendo assim, os períodos úmidos ( $P_u$  ou  $P_{umd}$ ) foram definidos como sendo aqueles com vazões superiores a  $P_{66}$  ou  $Q_{md}$ , enquanto que os períodos secos ( $P_s$  ou  $P_{smd}$ ) são definidos como sendo aqueles em que as vazões são inferiores a  $P_{33}$  ou  $Q_{md}$ .

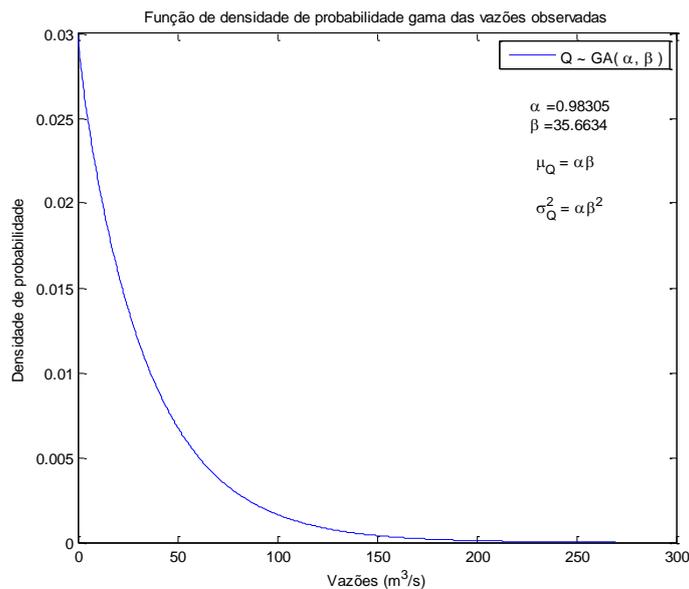


Figura 9.7 - Função de densidade de probabilidade Gama das vazões observadas.

A Figura 9.9 apresenta, em sua parte superior, os percentis  $P_{33}$  (linha vermelha) e  $P_{66}$  (linha azul) junto com as séries de vazões, enquanto que a parte inferior indica a ocorrência de anos úmidos e secos. A Figura 9.8 é similar à Figura 9.9, sendo que a única diferença são os limiares para determinação do que é um ano seco ou úmido. Em ambas as figuras, pode-se observar a ocorrência de aglomerados de anos secos nas décadas de 1940, 1950, 1980 e 1990, e de anos úmidos nas décadas de 1920, 1960, 1970.

Na tentativa de tornar essa análise a mais objetiva possível, procedeu-se ao cálculo da estatística  $T$ , métrica de persistência definida na secção 7.2. Os resultados são apresentados na Tabela 9.3. Observa-se que os valores de  $T$ , quando baseados na mediana das vazões, são maiores que aqueles baseados nos percentis de 33% e 66%. Isso é esperado porque não há um estado climático intermediário entre o úmido e o seco.

Além da estatística  $T$ , foram calculados os comprimentos médio e máximo dos períodos úmido e seco, assim como os volumes afluentes correspondentes. Essas informações são apresentadas na Tabela 9.4. Essas estatísticas relacionadas à persistência das vazões anuais serão utilizadas para avaliar o desempenho dos modelos HMM.

Tabela 9.3- Estatística  $T$  para a ocorrência de períodos secos e úmidos

Estatística $T$	Notação	Valor
Períodos secos com relação a $P_{33}$	$T_{sp}$	25
Períodos úmidos com relação a $P_{66}$	$T_{up}$	32
Períodos secos com relação à $Q_{md}$	$T_{smd}$	60
Períodos úmidos com relação à $Q_{md}$	$T_{umd}$	57

Tabela 9.4 - Métricas de persistência com relação aos percentis  $P_{33}$  e  $P_{66}$

Métrica/Período	Déficit		Úmido	
	Máximo	Médio	Máximo	Médio
Comprimento com relação a $P_{33}$ (anos)	5	1.9	6	2.2
Comprimento com relação à $Q_{md}$ (anos)	8	2.6	7	2.6
Volume com relação a $P_{33}$ ( $m^3$ )	7026.22	1008.21	67376.41	10566.10
Volume com relação à $Q_{md}$ ( $m^3$ )	26752.62	3772.31	90112.07	13214.65

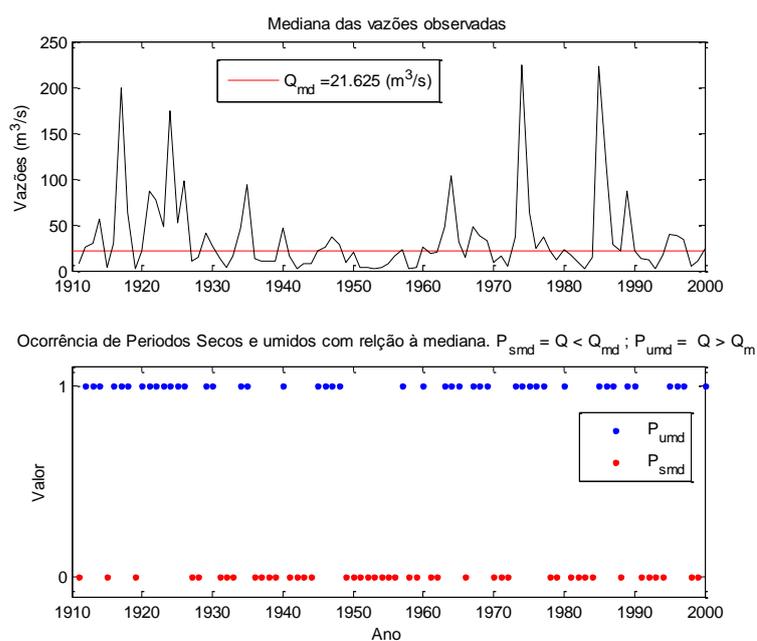


Figura 9.8 - Mediana das vazões observadas ( $Q_{md}$ ) (superior) e ocorrência de períodos secos e úmidos com relação à  $Q_{md}$  (inferior)

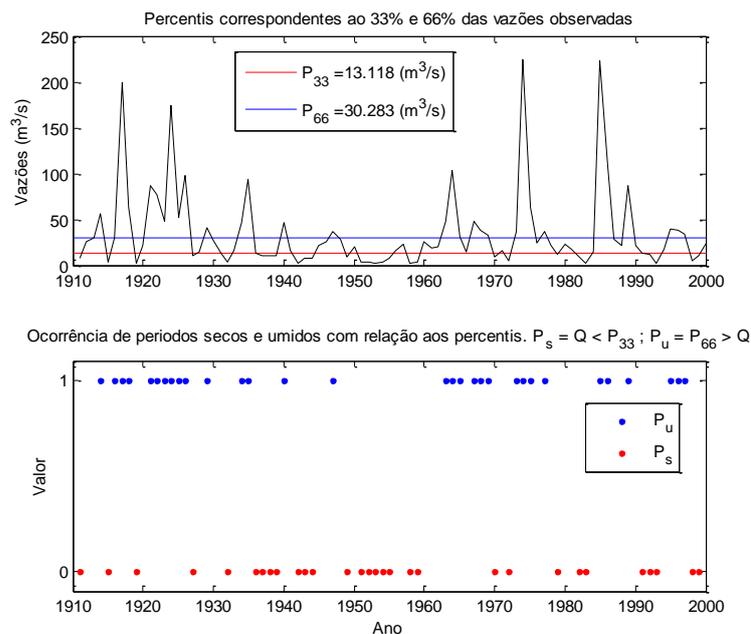


Figura 9.9 - Percentil 33% ( $P_{33}$ ) e percentil 66% ( $P_{66}$ ) da série de vazões observada (superior) e ocorrência de períodos secos e úmidos com relação aos percentis  $P_{33}$  e  $P_{66}$  (inferior)

## 9.2 - AJUSTE DOS MODELOS ARMA

O ajuste de vários modelos ARMA foi realizado para a série transformada de vazões,  $Y = \ln(Q)$ , empregando o *software MATLAB* (função *garchset*), e seguindo a metodologia de Box e Jenkins (1976). A escolha do modelo ARMA( $p, q$ ) mais adequado à série foi baseada nos critérios de *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC), tendo sido eliminados aqueles modelos cujos parâmetros não eram estatisticamente diferentes de zero ou cujos resíduos não estivessem de acordo com as premissas do modelo; razões que foram levadas em consideração para eliminar o AR(1)). Os resultados dos três melhores modelos ARMA, incluindo o AR(1), são apresentados na Tabela 9.5.

Tabela 9.5- Valores do AIC e BIC dos melhores modelos

Modelo	AIC	BIC	Log-likelihood	Número de parâmetros
ARMA(1,0)	267,952	275,452	-130,976	3
ARMA(2,0)	264,591	274,590	-128,296	4
ARMA(2,2)	265,020	280,019	-126,510	6
ARMA(4,2)	265,693	285,692	-124,847	8

Os resultados mostram que o modelo ARMA(2,0), ou AR(2), é o mais indicado dentre aqueles da família ARMA para modelar a série de vazão anual afluyente ao reservatório Orós, pois possui os melhores valores de AIC e BIC.

### 9.3 - AJUSTE DOS MODELOS HMM

Como foi mencionado na metodologia e levando em consideração a análise da série de vazões observadas, foram ajustadas duas classes do modelo HMM (HMM-NLT e HMM-G), uma para cada tipo de distribuição dependente, quais sejam, distribuição dependente Normal (HMM-NLT), e distribuição dependente Gama (HMM-G).

O ajuste desses modelos depende do número de estados ocultos,  $m$ , que deve ser definido *a priori*. Para cada uma das duas classes de modelos listadas acima, foram ajustados e avaliados modelos com 2 a 6 estados, totalizando 10 modelos. O ajuste propriamente dito foi realizado com o MVN-HMM *Toolbox*, desenvolvido por Kirshner (2005).

A

Tabela 9.6 apresenta os resultados do ajuste das duas classes de modelos HMM empregando diferentes números de estados ocultos. Os valores de AIC, BIC e o logaritmo da função de máxima verossimilhança (*Log-likelihood* – LLL) são apresentados. Note-se como um aumento no número de estados incrementa o número de parâmetros.

Para a classe HMM-NLT, o melhor modelo é aquele que emprega apenas 2 estados ocultos. Nesse caso, diferentemente das outras classes de HMM estudadas, quando  $m = 2$ , AIC e BIC apresentam os menores valores, não deixando dúvidas na escolha do modelo mais adequado.

O modelo HMM-G mais adequado depende do critério a ser utilizado. Se o critério AIC é utilizado, o melhor modelo contém 3 estados ocultos, mas se o critério BIC é empregado, o modelo mais adequado passa a ser o que contém apenas 2 estados. Esse tipo de situação é relativamente comum na literatura, e acontece pelo fato de que o critério BIC penalizar mais a complexidade do modelo do que o critério AIC.

Tabela 9.6 - Valores do AIC, BIC, Log-Likelihood (LLL) e número de parâmetros ( $k$ ), para os HMM de  $m$  estados ocultos.

Modelo	$m$	$k$	AIC	BIC	LLL
HMM-NLT	2	7	<b>277,518</b>	<b>295,017</b>	<b>-131,759</b>
	3	14	278,289	313,286	-125,144
	4	23	281,359	338,854	-117,679
	5	34	279,032	364,025	-105,516
	6	47	292,885	410,376	-99,443
HMM-G	2	7	814,773	<b>832,271</b>	<b>-400,386</b>
	3	14	<b>813,406</b>	848,403	-392,703
	4	23	814,096	871,592	-384,048
	5	34	816,560	901,553	-374,280
	6	47	828,874	946,366	-367,437

### 9.3.1 - Momentos dos HMM Ajustados

Esta seção compara os momentos dos modelos HMM ajustados. Eles são apresentados junto com os valores dos momentos observados, apresentados na Tabela 9.1 e representados pela linha tracejada preta, nas figuras seguintes. Assim, mostra-se: a média, desvio padrão, coeficiente de assimetria e a autocorrelação de ordem 1 ( $r_1$ ), que foram calculadas utilizando as equações da seção 5.4.2.2. A diante, a notação dos HMM será acompanhada do nome específico da classe e o número de estados dentre parêntesis, por exemplo, um HMM-NLT (2) representa um modelo HMM-NLT com 2 estados ocultos.

A Figura 9.10 a apresenta o valor da média para os HMM-Gs. Nessa figura, observa-se que a média produzida pelos diferentes modelos cai dentre os intervalos de confiança (IC) apresentados na Tabela 9.2, portanto, pode-se dizer que ela é representada apropriadamente por esses modelos. Os valores mais pertos a  $Q_m$  são produzidos pelo modelo HMM-G (3) ( $35.260 \text{ m}^3/\text{s}$ ). O desvio padrão também é bem representado (Figura 9.10 b) pelos HMM-Gs, tomando valores incluídos nos IC; se destacando o HMM-G (2) ( $44.379 \text{ m}^3/\text{s}$ ) por ter valores próximos à  $S_Q$ .

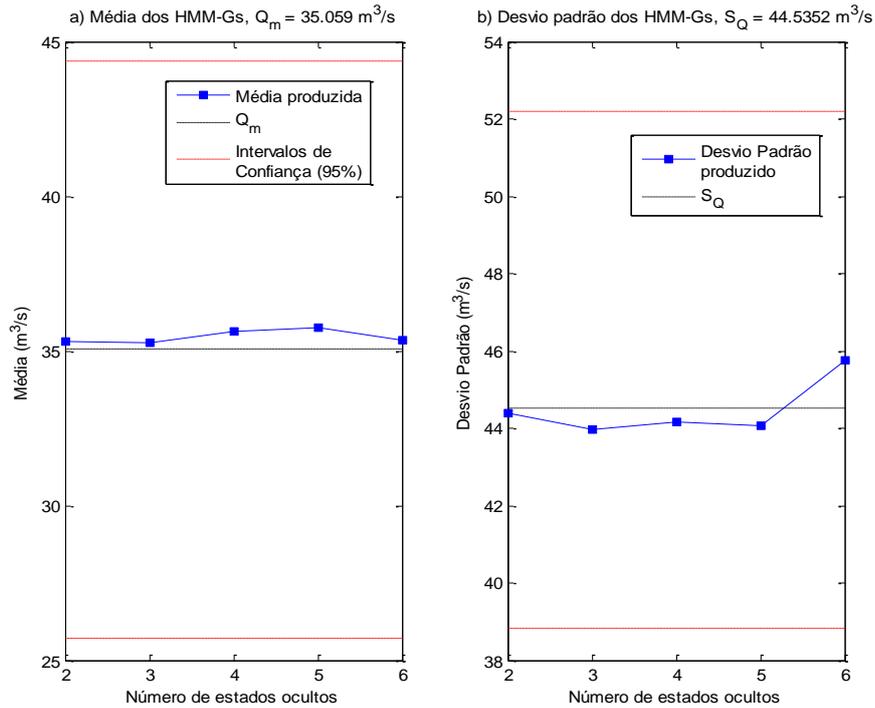


Figura 9.10 - Média (a) e desvio padrão (b) produzidos pelos modelos HMM-Gs.

Para o coeficiente de assimetria, a Figura 9.11a apresenta que nos HMM-G são obtidos valores menores ao  $Q_{Cs}$  observado, tendo-se um valor próximo só no modelo com 6 estados ocultos. Embora menores os coeficientes obtidos estejam dentre os IC (calculados via *Bootstrap*), somente no HMM-G (2) é apresentado um coeficiente que sai do intervalo de confiança inferior.

Para a autocorrelação de primeira ordem, pode-se apreciar na Figura 9.11.b que os HMM-Gs conseguem representar o  $r_{1\text{obs}}$  e os valores estão incluídos nos intervalos de confiança, ressaltando-se o valor mais próximo de 0.237 para o HMM-G (4).

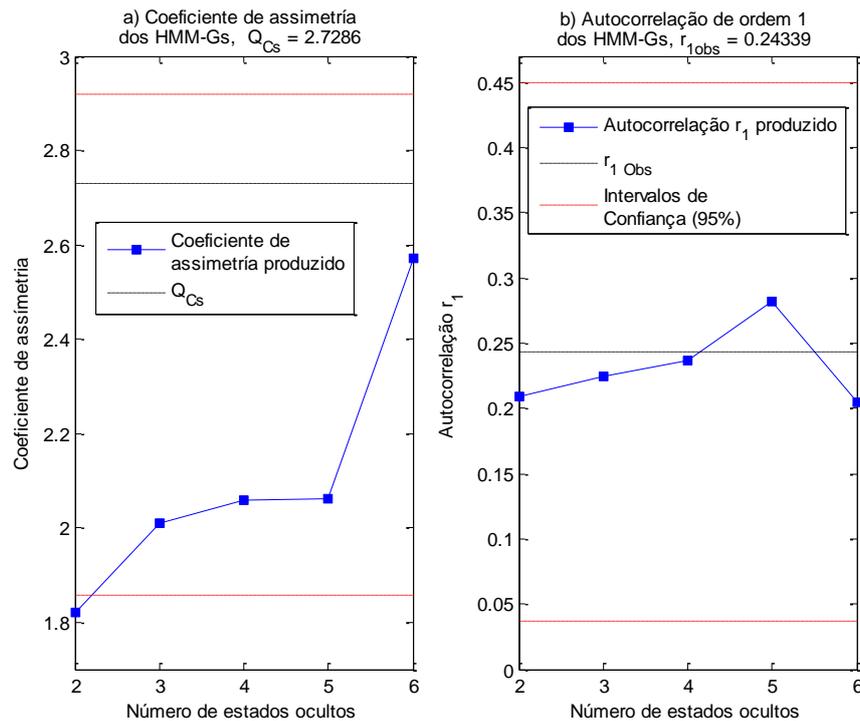


Figura 9.11 - Coeficiente de assimetria (a) e  $r_1$  (b), produzidos pelos modelos HMM-Gs

A mesma análise sobre a média e o desvio padrão dos modelos anteriores, pode ser realizada sobre os modelos HMM-NLT. Na Figura 9.12a e Figura 9.12b, observa-se que tanto os valores da média como o desvio padrão permanecem dentro dos IC dados na Tabela 9.2 e as variações desses valores são bem pequenas ao redor dos valores observados. No entanto, o HMM-NLT (3) apresenta valores mais pertos dos observados, mesmo para o coeficiente de assimetria (Figura 9.13 a). No caso da autocorrelação  $r_1$ , a Figura 9.13 b apresenta que os valores produzidos são menores do que o valor observado, onde o valor mais próximo é o produzido pelo HMM-NLT (5) e a autocorrelação do HMM-NLT (2) está fora dos IC.

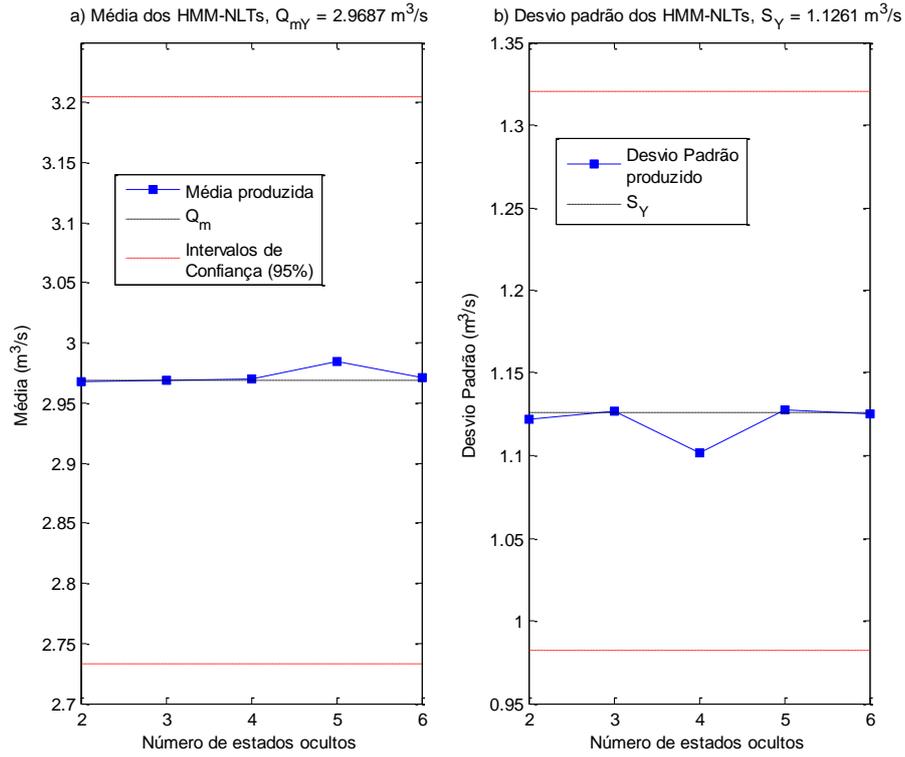


Figura 9.12 - Média (a) e desvio padrão (b) produzidos pelos modelos HMM-NLTs

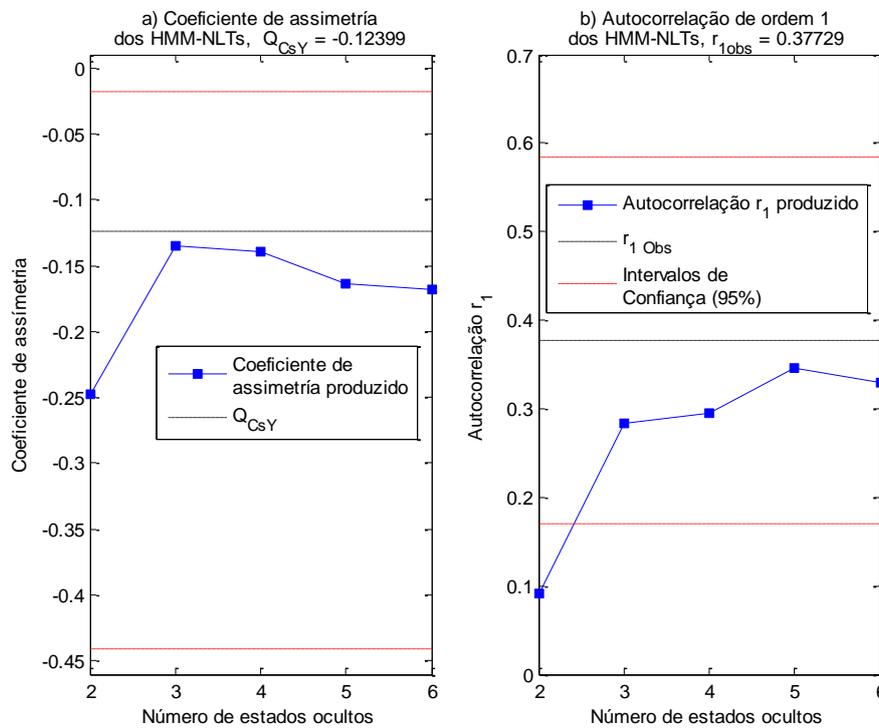


Figura 9.13 - Coeficiente de assimetria (a) e  $r_1$  (b), produzidos pelos modelos HMM-NLTs

### 9.3.2 - Estimação das probabilidades dos estados ocultos nos HMM ao longo do tempo

Estimar a probabilidade dos estados ocultos num modelo HMM ao longo do tempo possui dois propósitos, que são avaliar o quão confiante o modelo está em determinar um dado estado das vazões, por exemplo, seco ou úmido, no caso de apenas dois estados, e de permitir que se identifique e entenda se há alguma causa física que justifique as mudanças de estado ao longo do tempo.

O procedimento empregado aqui foi o mesmo mencionado na seção 5.4.2.5, utilizando a equação 5.48 e o procedimento utilizado por Akintug e Rasmussen (2005), descrito também nessa seção.

Em tese, seria possível apresentar os resultados para todos os modelos ajustados, o que incluiria modelos com até 6 estados  $\{s_1, s_2, \dots, s_6\}$ , sendo  $s_1$  o estado mais úmido e  $s_6$  o estado mais seco. Entretanto, a visualização e análise para mais de 4 estados ficam bastante complexas e de difícil conclusão sobre a dinâmica dos sistemas. Sendo assim, a seção corrente apresenta os resultados apenas para os modelos HMM com dois estados ocultos.

A Tabela 9.7 apresenta as matrizes de probabilidades de transição para os modelos HMM-NLT, e HMM-G, todos com apenas 2 estados ocultos, enquanto a Tabela 9.8 apresenta o valor médio da persistência no tempo de cada um dos estados ocultos.

A Figura 9.14 apresenta as probabilidades dos estados  $S_1$  (úmido) e  $S_2$  (seco) ao longo do tempo para o modelo HMM-NLT (2), assim como as vazões anuais observadas no tempo, com as respectivas médias de cada estado oculto. Observa-se que as probabilidades, de uma forma geral, oscilam entre 0 e 1, indicando que os estados foram claramente identificados pelo modelo, com exceção de períodos curtos, como no ano de 1952 e 1959. Pode-se observar também a elevada persistência no tempo do estado úmido, com pequenas alternâncias para o estado seco, voltando rapidamente ao estado úmido, constatação que está de acordo com o apresentado na Tabela 9.8.

Tabela 9.7 – Matrizes de probabilidades de transição dos HMM de 2 estados ocultos

Modelo		HMM-NLT		HMM-G	
Para o estado		1	2	1	2
Desde o estado	1	0.909	0.091	0.646	0.354
	2	0.686	0.314	0.129	0.871

Tabela 9.8 - Duração média dos períodos secos e úmidos em anos, nos HMM de 2 estados ocultos

Período	Modelo	
	HMM-NLT	HMM-G
Úmido	11	2.8
Seco	1.5	7.8

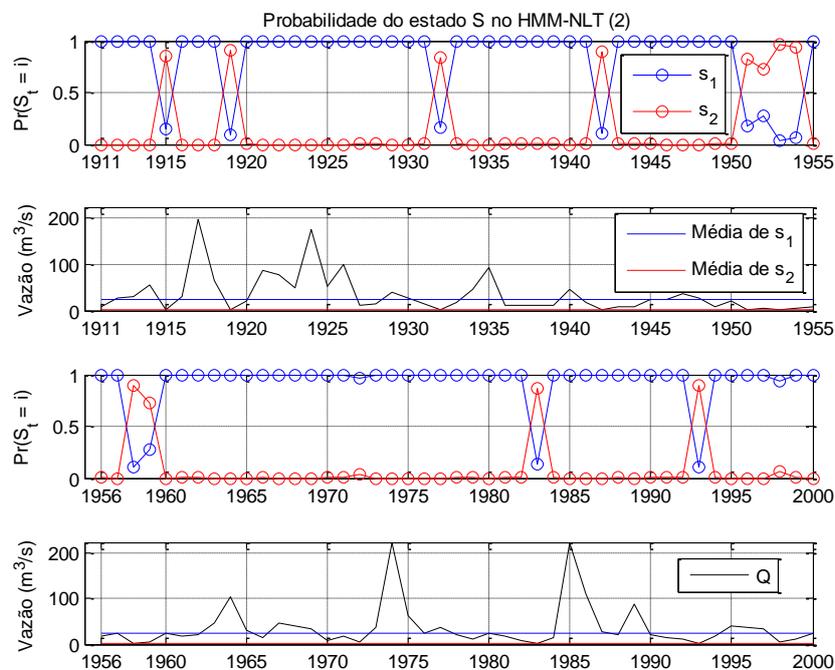


Figura 9.14 - Probabilidade do estado  $s_1$  e  $s_2$  ao longo do tempo no HMM-NLT (2)

Os resultados do modelo HMM-G (2) apresentam um comportamento um pouco diferente no que tange as probabilidades dos estados úmido e seco. Embora na maior parte dos anos os estados tenham sido bem definidos, ou seja, com probabilidades próximas a 0 ou 1, observa-se um número maior de anos, quando comparados com os HMM-NLTs, em que a definição do estado não é tão clara, como nos períodos 1912-1914, 1963-1964, 1966-1969,

1987, 1989, 1996-1997 (Figura 9.15). A relação entre as probabilidades dos estados úmido e seco parece razoável em relação ao valor das vazões anuais.

De uma forma geral, pode-se dizer que os HMMs com dois estados ocultos identificaram com um certo grau de clareza os estados úmidos e secos, e que esses estados estão bem relacionados com as vazões observadas. Vale lembrar que as denominações úmido e seco são relativas, pois dependem das distribuições dependentes aos estados ocultos. Ressalta-se que o modelo HMM-G (2) possui períodos secos de comprimento maior (7.8 anos, segundo a Tabela 9.8) em comparação ao HMM-NLT (2).

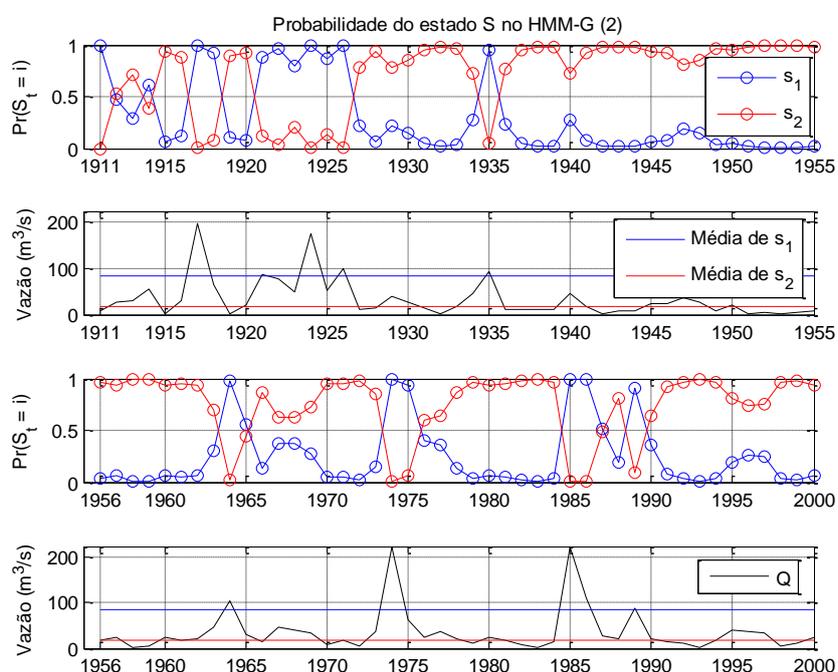


Figura 9.15 - Probabilidade do estado  $s_1$  e  $s_2$  ao longo do tempo no HMM-G (2)

### 9.3.3 - Séries Viterbi

Esta seção apresenta a série de estados ocultos mais prováveis, estimada pelo procedimento Viterbi. Como dito anteriormente, modelos com muitos estados apresentam problemas no ajuste dos parâmetros das distribuições. Além disso, modelos HMM com mais de 3 estados são muito complexos e de difícil interpretação. Sendo assim, os

resultados apresentados aqui limitam-se aos modelos HMM-NLT (2) e HMM-G (2) e (3). Apenas a título de ilustração, é apresentada também a série Viterbi do HMM-NLT (6).

A Figura 9.16 (superior) apresenta as séries Viterbi dos modelos HMM-NLT (2) e HMM-G (2), além da série de vazões observadas. Observa-se uma persistência elevada do estado úmido, (ou de vazões mais altas) na série Viterbi referente ao modelo HMM-NLT (2), e também uma persistência elevada do estado seco (vazões mais baixas) referente ao modelo HMM-G (2). Vale de novo destacar que os estados úmido e seco não possuem qualquer relação quando comparados entre modelos. A sequência de estados úmido e seco, observadas nas séries Viterbi dos dois modelos, é um indício, ou pelo menos uma possibilidades, de que esses modelos possam representar adequadamente as estatísticas de persistências observadas na série histórica.

No modelo HMM-G (3) (Figura 9.17, superior), pode-se observar também uma alternância entre os 3 estados, mas sempre com uma certa persistência em cada estado, porém com uma persistência maior no estado intermediário. Esse tipo de comportamento pode ser visto como um indício de que esse modelo possa vir a reproduzir adequadamente as métricas de persistência observadas na série histórica.

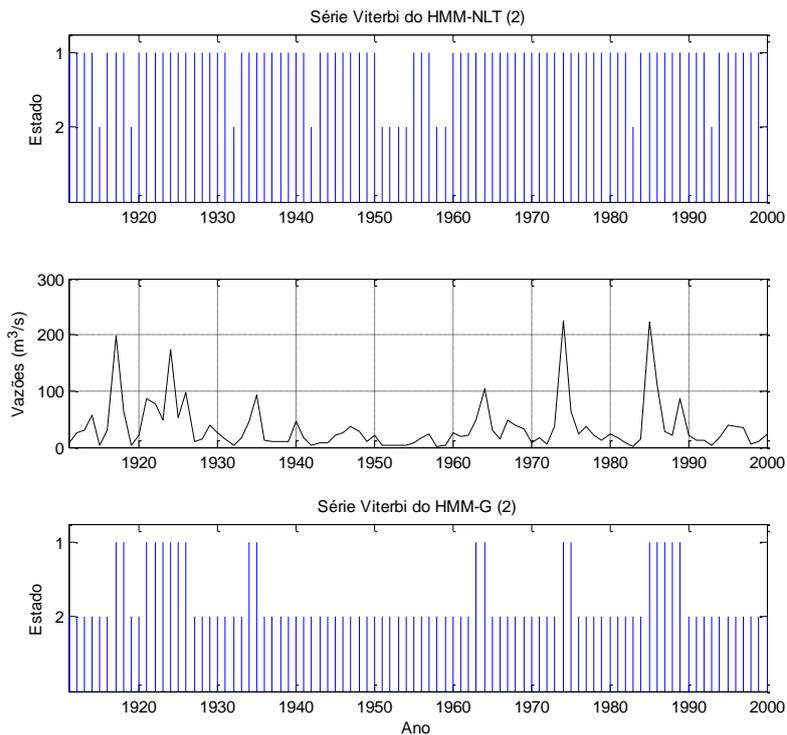


Figura 9.16 - Sequência de estados mais prováveis no HMM-NLT (2) (superior), vazões observadas (centro) e sequência de estados mais prováveis no HMM-G (2) (inferior)

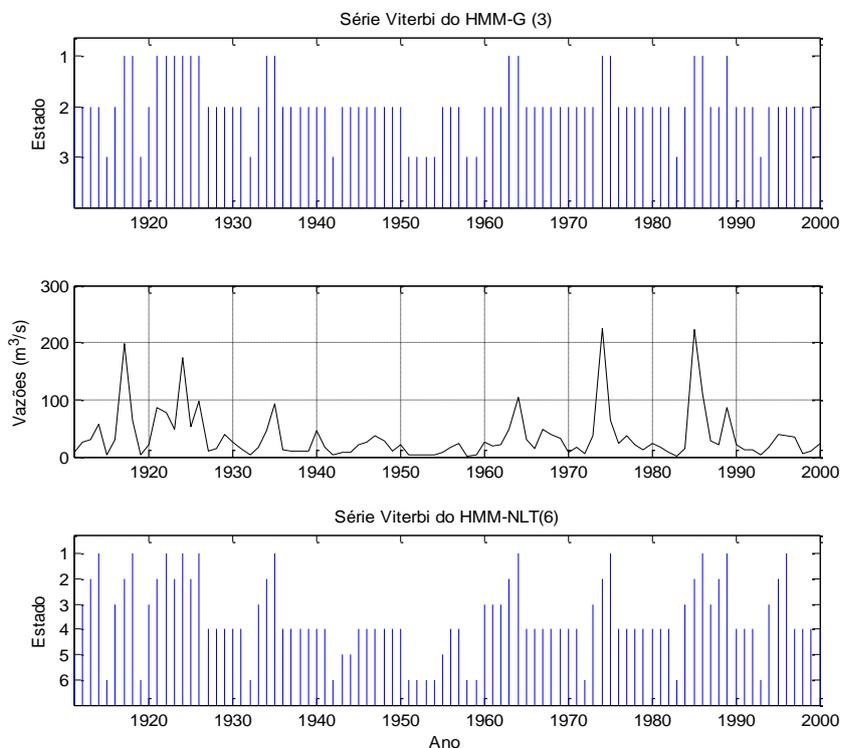


Figura 9.17 - Sequência de estados mais prováveis no HMM-G (3) (superior), vazões observadas (centro) e sequência de estados mais prováveis no HMM-NLT (6) (inferior)

### 9.3.4 - Distribuições de Probabilidade Marginais

Os modelos HMM devem ser capazes de representar adequadamente a distribuição marginal das vazões anuais observadas. Esse é mais um critério que deve ser empregado para determinar a adequabilidade dos modelos na modelagem de séries anuais de vazão.

Esta seção possui o objetivo de apresentar as curvas de frequência de vazões anuais dos modelos HMM-NLT e HMM-G, com 2 e 3 estados, já que esses foram os modelos utilizados para avaliar a capacidade dos modelos HMM em reproduzir as estatísticas de persistência da série observada, o que será apresentado na próxima seção.

As curvas de frequência apresentadas abaixo foram determinadas conforme descrição na seção 5.4.2.1. Não foi realizado nenhum teste estatístico para determinar a aderência da distribuição teórica produzida por um HMM aos dados observados, porque no caso do teste Qui-Quadrado, não se tinham as frequências observadas dos estados ocultos, apenas se tem as frequências esperadas. Análises apresentadas na sequência são baseadas em inspeção visual.

A Figura 9.18 apresenta a curva de frequência cumulativa de vazões anuais do modelo HMM-NLT com 2 estados ocultos. Pode-se observar visualmente a boa aderência dos valores observados com a curva do modelo. O mesmo pode ser dito para o restante dos modelos HMM-Gs de 2 e 3 estados ocultos, cujas curvas de frequência são apresentadas na sequência, Figura 9.19 e Figura 9.20.

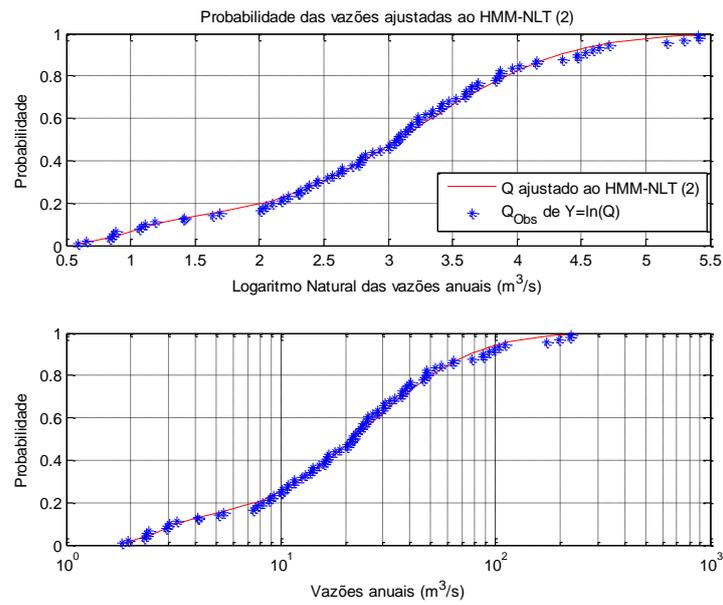


Figura 9.18 – Probabilidade das vazões ajustadas ao HMM-NLT (2) em “escala logarítmica” (superior) e em “escala real” ( $exp(Q_{sim})$ ) (inferior)

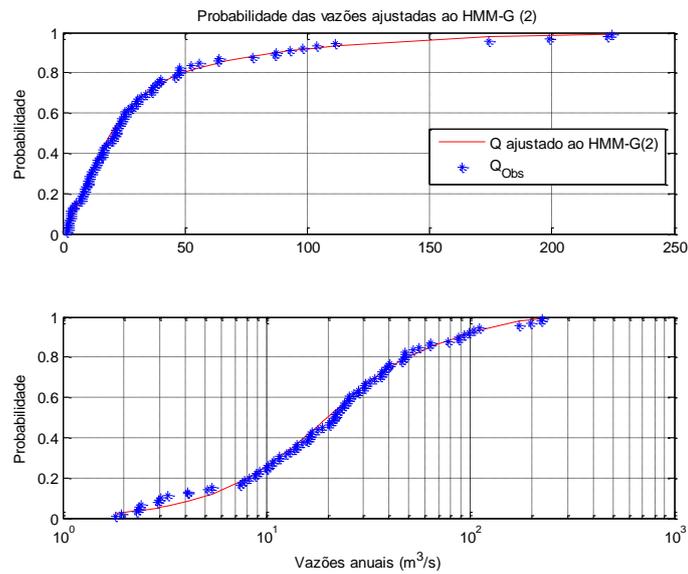


Figura 9.19 - Probabilidade das vazões ajustadas ao HMM-G (2) (superior) em escala semi-logarítmica e em “escala real” (inferior).

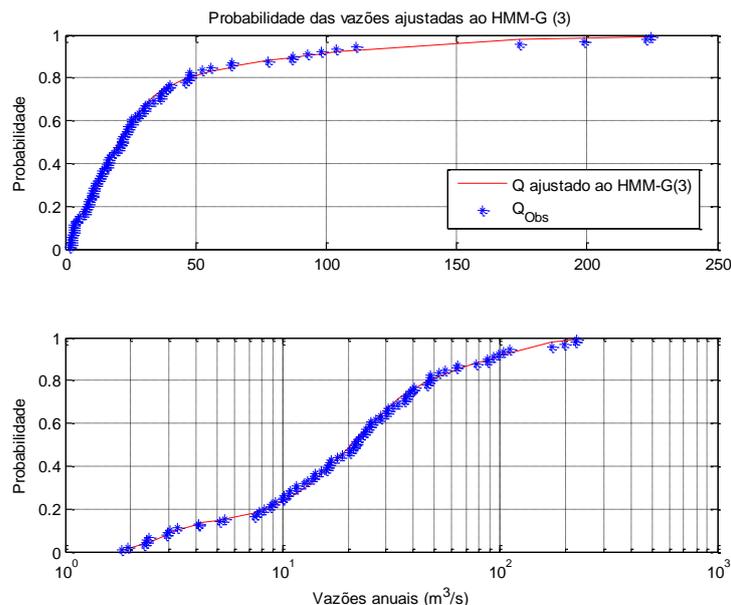


Figura 9.20 - Probabilidade das vazões ajustadas ao HMM-G (3) (superior) em escala semi-logarítmica e em “escala real” (inferior).

#### 9.4 - COMPARAÇÃO DAS SÉRIES SINTÉTICAS GERADAS PELOS MODELOS ARMA E HMMs

Esta seção apresenta uma comparação entre os diversos modelos HMM ajustados à série de vazões anuais afluentes ao reservatório Orós com o modelo AR(2), escolhido para representar a classe de modelos do tipo ARMA. Além disso, as séries simuladas por esses modelos são comparadas com as produzidas pela distribuição Gama de 2 parâmetros (GA(2)). A comparação é feita com base em diversas métricas, incluindo estatísticas descritivas, tais como média, desvio-padrão, coeficiente de assimetria, e autocorrelação (lag 1), vazões máximas e mínimas, e métricas de persistência definidas anteriormente, que incluem: a estatística  $T$ , comprimentos máximos e médios de períodos úmidos e secos, definidos de duas maneiras diferentes, além de volumes máximos e médios de períodos úmidos e secos.

Para que essa comparação fosse possível, cada modelo HMM, GA(2) e o próprio modelo AR (2), foi utilizado para gerar 1000 séries sintéticas, com 180 anos cada, tendo sido utilizado os últimos 90 anos de cada série para a determinação das métricas elencadas acima.

Apesar de terem sido geradas séries para modelos com até 6 estados, serão apresentados apenas os resultados dos modelos com até 3 estados ocultos. No caso dos modelos AR(2) e HMM-NLT, ajustados com base no logarítmico das vazões observadas, as vazões simuladas foram transformadas de volta para a escala real do problema.

Dessa forma, as análises apresentadas nesta seção referem-se aos modelos AR(2) e aos modelos: HMM-NLT (3) e HMM-G com 2 e 3 estados ocultos, que tiveram sucesso em representar melhor a série observada. A análise baseia-se numa comparação visual entre a métrica observada na série histórica e os diferentes diagramas de caixa de cada modelo, que são utilizados como uma descrição estatística das métricas obtidas por cada modelo para cada uma das 1000 séries geradas. Essa é a mesma abordagem empregada, por exemplo, em Sveisson *et al.* (2003), Akitung e Rasmussen (2005), e Bracken (2011) para avaliar a adequabilidade de modelos similares na representação de algumas estatísticas observadas em séries observadas.

A abordagem empregada aqui assume que se os valores observados das métricas ficarem entre o primeiro e o terceiro quartil, ou seja, dentro do intervalo interquartil (dentro da caixa), o modelo é capaz de representar bem as características da série observada. Vale notar que se a métrica baseada na série observada cair fora da caixa que corresponde ao intervalo interquartil, não é possível afirmar categoricamente que o modelo não é apropriado.

Então, para auxiliar que modelos podem ser úteis para a modelagem da série  $Q$ , foi calculado um valor  $p$  “empírico” que representa a frequência com que cada modelo calcula métricas acima do valor observado. Assim, para uma significância de 5%, o modelo pode ser considerado de “aceito”. Essa abordagem é similar à realizada por Stedinger e Taylor (1982). Assim, o valor  $p$  correspondente a cada modelo é colocado nas figuras na sequência, como percentagem ao lado do diagrama de caixa.

#### **9.4.1 - Momentos**

A Figura 9.21 apresenta a média (superior) e o desvio-padrão (inferior) gerados por cada modelo HMM. De uma forma geral, todos os modelos HMM representam bem os dois primeiros momentos. Isto já havia sido mostrado na seção 9.3.1. Vale notar que os modelos

HMM-G com 2 e 3 estados apresentam menor número de *outliers* do desvio-padrão, quando comparados aos modelos HMM-NLT (3) e AR(2), esse último com *outliers* extremamente elevados de desvio-padrão, o que explica, de certa maneira, os valores extremamente elevados de vazão máxima, como será discutido mais adiante. Entanto, que o “modelo” Gama (GA (2), nas figuras) apresenta deficiências ao representar o desvio padrão.

O coeficiente de assimetria (Figura 9.22, superior) também foi bem representado por todos os modelos, excetuando o GA (2). Com o valor observado bem próximo da mediana dos valores gerados pelos modelos HMMs, tendo-se resultados bastante similares nos modelos HMM-Gs.

No caso da autocorrelação  $r_1$  (Figura 9.22, inferior), observa-se um desempenho satisfatório de todos os modelos HMM e do AR(2), com exceção dos modelos HMM-NLT (3) e GA (2). Embora, teoricamente, os HMMs, não sejam concebidos para capturar a autocorrelação (Zucchini e MacDonald, 2009).

No modelo Gama, a incapacidade de reproduzir  $r_1$  é claramente visualizada na Figura 9.22. Para a representação dos momentos, esse modelo não consegue representar adequadamente os valores observados, exceto a média. Nota-se que para o desvio padrão e a autocorrelação o valor-p não chega aos 5%, o modelo atende esse valor ao calcular o coeficiente de assimetria. Por enquanto, sob essa observação, estatisticamente não existem argumentos suficientes para aceitar essa classe de modelo para modelar a série histórica.

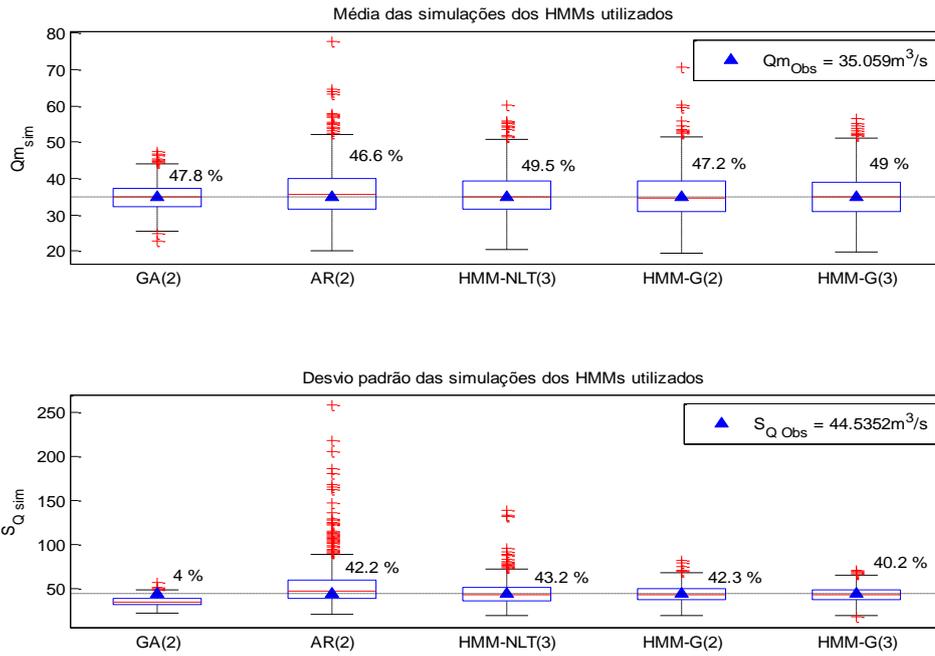


Figura 9.21 - Média (superior) e desvio padrão (inferior) das series sintéticas geradas pelos HMM, AR(2) e GA (2).

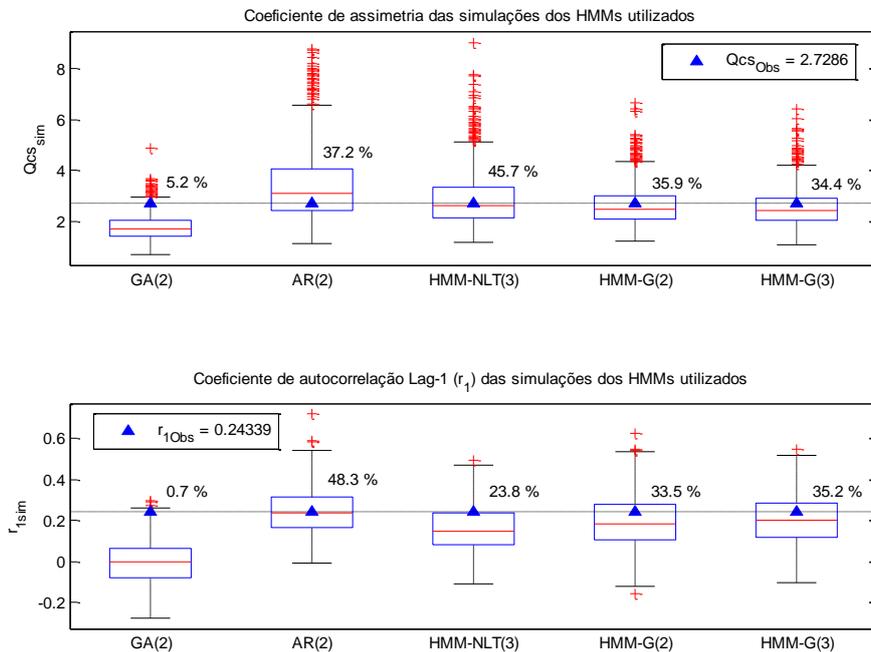


Figura 9.22 – Coeficiente de assimetria (superior) e autocorrelação de primeira ordem (inferior) das series sintéticas geradas pelos HMM, AR(2) e GA (2).

Em relação às vazões máximas (Figura 9.23, superior), observa-se que os HMMs desempenham-se melhor do que o AR(2), que além de obter um diagrama de caixa relativamente elevado, com o primeiro quartil próximo da vazão máxima observada, acabou gerando um número relativamente elevado de vazões extremamente altas, atingindo valores de mais de 2000 m<sup>3</sup>/s, enquanto que a vazão máxima observada foi de 224.6 m<sup>3</sup>/s. Nesse aspecto, os modelos HMM-Gs, independente do número de estados, geraram vazões máximas de no máximo 500 m<sup>3</sup>/s, o que parece ser mais próximo da realidade, embora não tenha sido feita nenhuma análise em relação ao tempo de recorrências de tais valores de vazão.

Em relação às vazões mínimas (Figura 9.23, inferior), o modelo HMM-NLT (3), assim como os modelos HMM-Gs, foi capaz de representar essas vazões mínimas. O modelo AR(2), por sua vez, teve um pouco dificuldade em representar as vazões mínimas, assim como o HMM-G (2) (sendo melhor o HMM-G (3) ) e o GA (2), que obteve o pior desempenho.

De uma forma geral, os modelos HMM representaram relativamente bem os momentos da distribuição, assim como as vazões máximas e mínimas, embora alguns deles tenham tido dificuldade de representar as mínimas e outras tenham tido problemas em representar a autocorrelação (lag 1). De novo, o modelo GA (2), apresenta indícios de não ser o correto para representar a série  $Q$ , por subestimar as vazões, além de ter valores  $p$  baixos. O modelo AR (2) foi capaz de representar bem os momentos, porém gerou valores muito elevados de desvio-padrão, o que acabou acarretando em valores muito elevados de vazões máximas; tal vez resultado da transformação nas vazões.

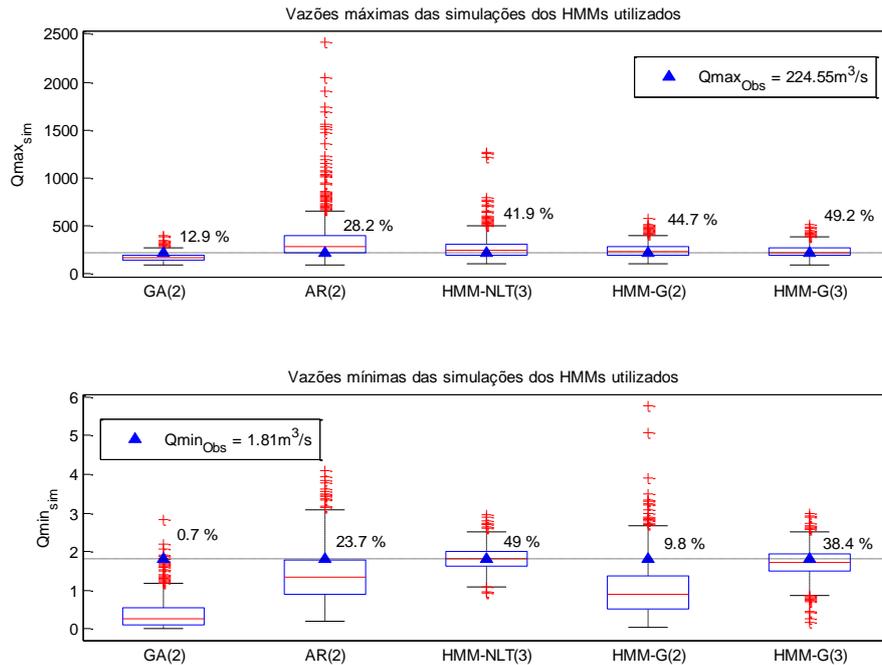


Figura 9.23 – Vazão máxima (superior) e vazão mínima (inferior) das series sintéticas geradas pelos HMM, AR(2) e GA (2).

#### 9.4.2 - Métricas de persistência

A Figura 9.24 apresenta os resultados obtidos para a estatística  $T$  para os períodos secos, sendo o gráfico superior relativo ao percentil de  $P_{33}$ , enquanto o gráfico inferior refere-se à mediana. Pode-se observar que, quando o período seco é determinado por um limiar mais severo, os modelos HMM possuem um pouco de dificuldade em representar adequadamente a estatística  $T$ . Nesse caso, o modelo AR (2) apresenta um desempenho superior, embora os modelos HMM-NLT (3) e HMM-G (2, 3) não sejam necessariamente ruins. Quando a mediana é empregada para definir o período seco, os modelos HMM, apresentam desempenho superior ao AR (2), com valor observado muito próximo da mediana dos valores gerados sinteticamente.

No caso da estatística  $T$  para os períodos úmidos, o que se observa é um desempenho satisfatório dos modelos HMM e superior ao modelo AR (2), quando o percentil  $P_{66}$  é empregado na definição do período úmido (ver Figura 9.33, superior). Quando a mediana é empregada na definição do período, todos eles passam a ter desempenho satisfatório, inclusive o AR (2), como pode ser observado na Figura 9.33, inferior. Entretanto, para os

períodos úmidos, quanto os secos o modelo GA(2), se mostra bem deficiente ao representar a estatística  $T$ .

Pode-se dizer, então, que em relação à estatística  $T$ , que o desempenho relativo dos modelos HMM e AR(2) depende do fato do período ser úmido ou seco, e do próprio limiar que define os períodos úmido e seco. No caso do período seco, o AR(2) possui desempenho melhor na definição mais crítica, enquanto que os modelos HMM possuem desempenho melhor do que o AR(2) quando a mediana é empregada, embora o AR(2) possua desempenho aceitável nesse caso. No caso dos períodos úmidos, os modelos HMM possuem desempenho melhor, especialmente se o limiar for o de  $P_{66}$ .

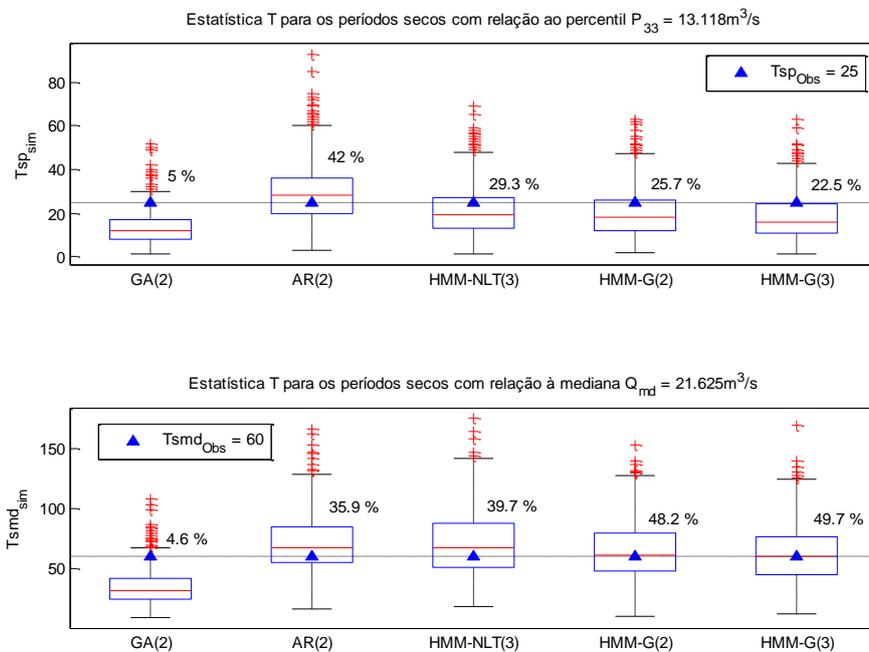


Figura 9.24 - Estatística  $T$  para os períodos secos com relação ao percentil  $P_{33}$  (superior) e à mediana  $Q_{md}$  (inferior).

Em relação ao comprimento máximo dos períodos gerados pelos HMMs, obteve-se para os períodos secos um bom desempenho do AR(2) com relação a  $P_{33}$  frente aos HMMs, porém, a representação dos HMMs não foi tão deficiente (por exemplo o HMM-NLT (3) e o HMM-G (3)) O comportamento melhorou em relação à mediana, com melhor desempenho do HMM-NLT (3), que não subestimou essa métrica. Deficiências na representação desse

comprimento foram encontradas para o GA (2), sem importar o limiar o modelo subestima os valores observados.

Para a produção de comprimentos máximos de períodos úmidos (Figura 9.25), existe uma tendência dos HMMs a produzir comprimentos mais longos. Para o  $P_{66}$ , todos os modelos apresentam a tendência a subestimar essa estatística. O comportamento de novo melhora para a mediana, destacando-se o HMM-NLT (3), ainda os HMM-G não sejam ruins. No caso do GA (2), o modelo se comporta similar ao AR(2), aqui esse modelo não apresenta valores  $p$  tão baixos comparados aos obtidos nas métricas anteriores e especialmente frente ao AR(2).

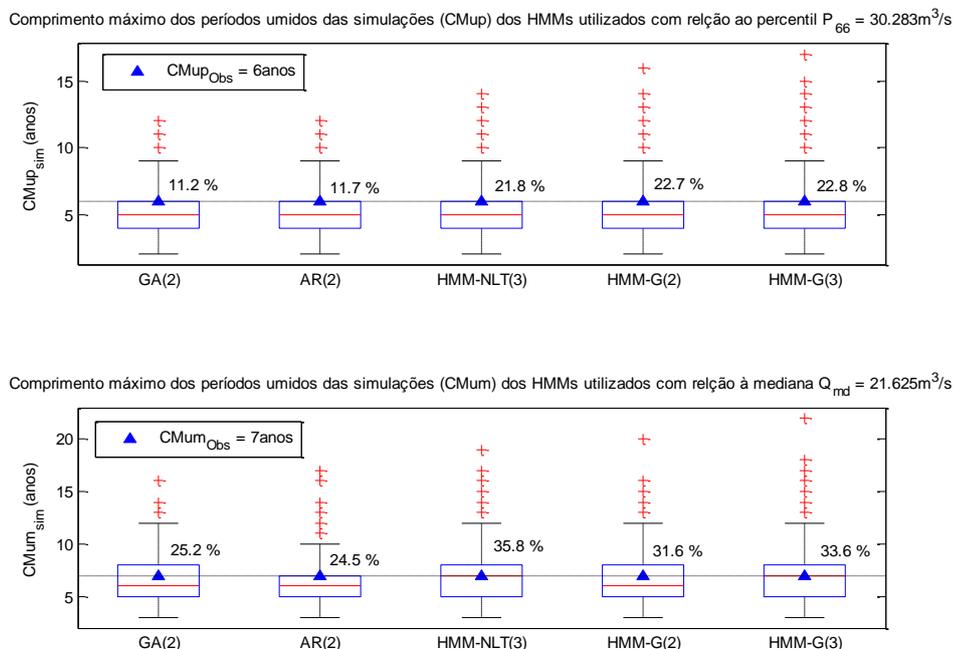


Figura 9.25 - Comprimento máximo dos períodos úmidos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior).

No caso dos comprimentos médios, a Figura 9.26 apresenta os resultados para os estados secos. Para esta métrica, foi encontrado um desempenho similar aos comprimentos máximos de períodos secos e úmidos dos modelos estocásticos. O comportamento dos HMMs não foi tão satisfatório para os períodos úmidos (Figura 9.34). Porém, segundo os valores da Tabela 9.4, existe uma simetria para comprimentos médios observados de períodos secos e úmidos, esperava-se uma reprodução adequada. Dessa representação,

ressaltasse o desempenho do AR(2) e o HMM-NLT (3). Claramente, o GA (2) fica por fora de um bom desempenho, sobretudo na representação de comprimentos médios de períodos secos.

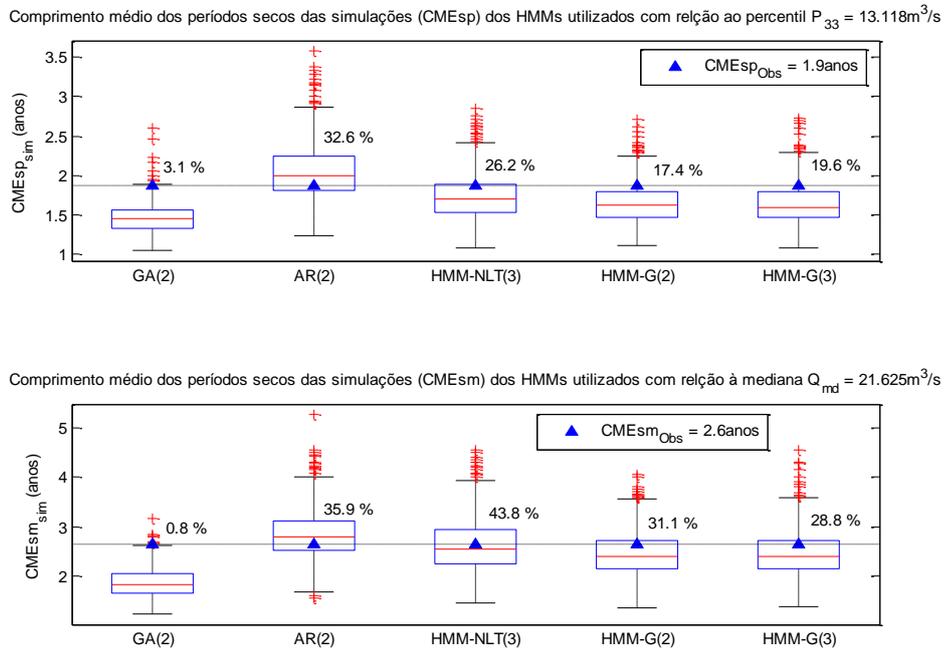


Figura 9.26 - Comprimento médio dos períodos secos com relação ao percentil  $P_{33}$  (superior) e à mediana  $Q_{md}$  (inferior)

Na Figura 9.27, aprecia-se como, em geral, os modelos utilizados representam com sucesso o volume máximo úmido, excluindo o GA (2). Especialmente o desempenho melhora com respeito à  $Q_m$ . No caso dos volumes de déficit, na Figura 9.36, os HMM-Gs subestimam essa métrica e os modelos AR(2) e HMM-NLT (3) conseguem ser melhores, o comportamento melhora ainda mais ao utilizar a mediana como limiar e o GA (2) continua tendo dificuldade em representar bem esse volume.

No caso dos volumes médios úmidos, os modelos apresentam resultados menos adequados, em relação aos produzidos para os volumes máximos úmidos, visto que subestimam essa métrica, exceto o HMM-G (3). Os resultados dos HMMs para os estados úmidos são melhores que para os estados secos (Figura 9.37). O HMM-NLT (3) e o AR(2), de novo resultam melhores para ambos os limiares e volumes de déficit. Entanto, para os úmidos, observa-se na Figura 9.28 melhor reprodução do HMM-G (3) e AR(2), sobretudo com relação à mediana. O GA (2) novamente demonstra incapacidade em simular essa métrica.

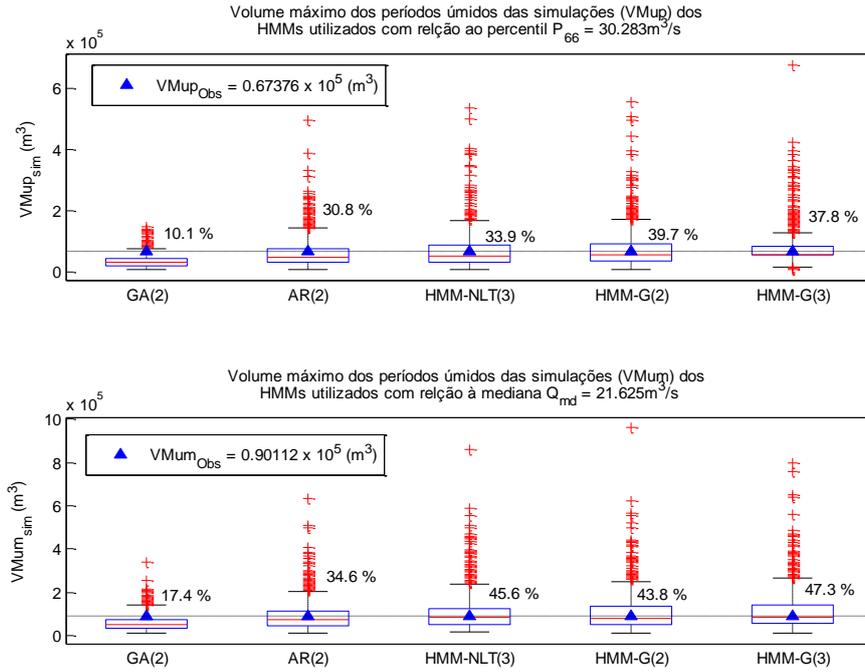


Figura 9.27 - Volume máximo dos períodos úmidos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

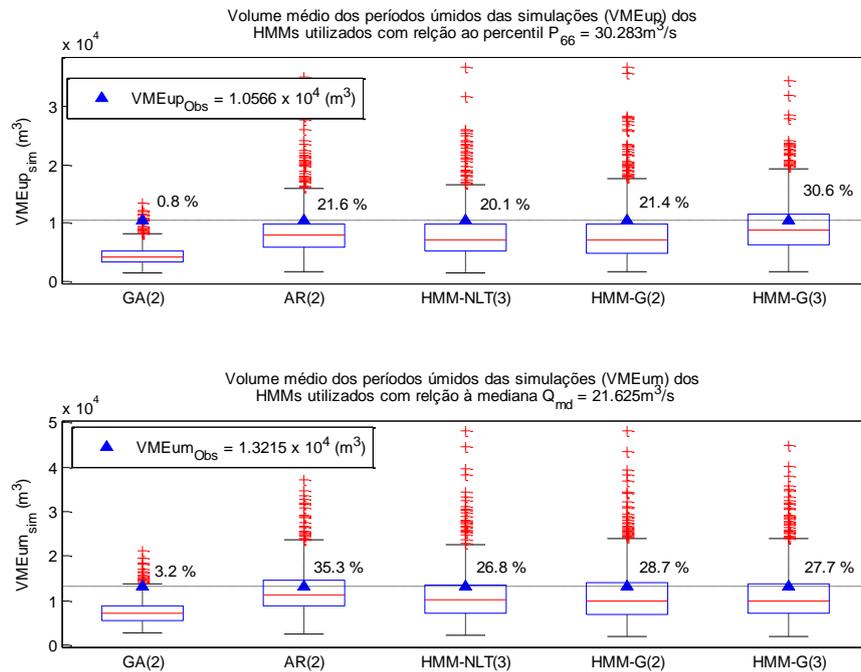


Figura 9.28 - Volume médio dos períodos úmidos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

## 9.5 - AJUSTE DE MODELOS DE MARKOV OCULTO NÃO HOMOGÊNEO

### 9.5.1 - CORRELAÇÃO ENTRE AS VAZÕES OBSERVADAS E OS ÍNDICES CLIMÁTICOS NINO3 E DIPOLO DO ATLÂNTICO

Continuando com o mencionado na metodologia, foi analisada a correlação entre as vazões anuais observadas e os índices climáticos NINO3 e DIPOLO do Atlântico (gradiente de temperaturas da superfície do mar no oceano Atlântico Tropical), dos meses do período 1910-2000. Essa correlação foi calculada com o método não paramétrico de Spearman, que a diferença da técnica de Pearson não assume uma relação linear entre as variáveis (para mais detalhes ver Storch e Zwiers, 1999).

Assim, levou-se em conta que se o mês era menor que 7 (Janeiro até Junho), utiliza-se os índices mensais do mesmo ano que a vazão, caso contrário, a correlação calcula-se com os índices do ano anterior à vazão. Por exemplo, para a vazão de 1911 foram tomados os índices dos meses de Janeiro até Junho de 1911 e os índices mensais de Julho até Dezembro de 1910. Assim, conjuntamente tinham duas séries (1910-1999 e 1911-2000) de índices climáticos para essa análise.

Dessa maneira, a Figura 9.29 apresenta graficamente os valores encontrados para o coeficiente de correlação Spearman ( $\rho_s$ ), entre as vazões e os índices NINO3 e DIPOLO do Atlântico na parte superior e inferior dessa figura, respectivamente. Igualmente, a Tabela 9.9 apresenta os valores das correlações encontradas.

Observa-se tanto na figura como na tabela que as vazões encontram-se mais correlacionadas com o índice NINO3 do mês de fevereiro ( $\rho_s = -0.281$ ), enquanto que, para o DIPOLO existem duas correlações significativas, elas apresentam-se nos meses de Julho ( $\rho_s = -0.483$ ) e Agosto ( $\rho_s = 0.489$ ), sendo “maior” para esse último mês. Além disso, da Tabela 9.9 observa-se, para a maioria das correlações como o NINO3, que os valores  $p$  são menores a 0.05, o que indica que as correlações são significativas e no caso do DIPOLO, igualmente estão abaixo desse limiar. Também, nota-se que o DIPOLO é mais bem correlacionado que o NINO3, devido aos altos valores de  $\rho_s$ .

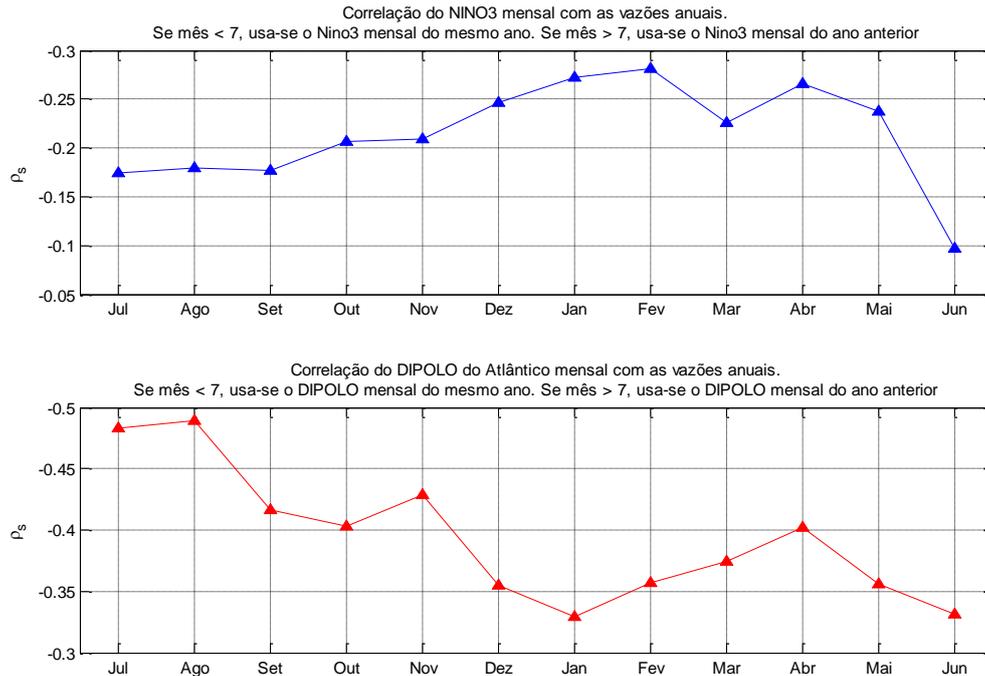


Figura 9.29 - Correlação Spearman entre as vazões anuais e os índices climáticos NINO3 (superior) e DIPOLO do Atlântico (inferior) mensais. Período 1910-2000

Tabela 9.9 - Coeficientes de correlação Spearman ( $\rho_s$ ) e os seus respectivos valores p para a correlação entre as vazões anuais e os índices climáticos NINO3 e DIPOLO do Atlântico mensais para o período de 1910-2000

Mês	Correlação com o NINO3		Correlação com o DIPOLO do Atlântico	
	$\rho_s$	Valor $p$	$\rho_s$	Valor $p$
Janeiro	-0.272	0.010	-0.329	0.002
Fevereiro	<b>-0.281</b>	0.007	-0.357	0.001
Março	-0.225	0.033	-0.374	2.80E-04
Abril	-0.266	0.011	-0.402	8.41E-05
Maio	-0.237	0.024	-0.355	0.001
Junho	-0.097	0.365	-0.331	0.001
Julho	-0.174	0.101	<b>-0.483</b>	1.42E-06
Agosto	-0.179	0.091	<b>-0.489</b>	9.86E-07
Setembro	-0.177	0.096	-0.416	4.44E-05
Outubro	-0.206	0.051	-0.403	8.10E-05
Novembro	-0.210	0.047	-0.429	2.51E-05
Dezembro	-0.247	0.019	-0.355	0.001

Com os resultados encontrados das correlações e os índices climáticos NINO3 e DIPOLo do Atlântico, ajustaram-se vários modelos HMMs não homogêneos (NHMM), com as séries desses índices como variáveis exógenas (*inputs*). Assim, foi empregada a série do índice NINO3 dos meses de fevereiro do período 1911-2000 e as séries do índice DIPOLo dos meses de julho e agosto do período 1910-1999. De acordo aos resultados das seções 9.3 e 9.4, continua-se a análise utilizando os mesmos HMMs, neste caso não homogêneos: NHMM-NLT (3) e NHMM-G (2 e 3).

Ajustaram-se NHMMs empregando, separadamente, cada índice climático e a combinação dos índices. Portanto, foram obtidos 5 grupos de NHMM: (1) NHMMs com o NINO3 do mês de Fevereiro, (2) NHMMs com o DIPOLo do mês de Agosto, (3) NHMMs com o DIPOLo do mês de Julho, (4) NHMMs com o NINO3 do mês de Fevereiro e DIPOLo de Agosto, (5) NHMMs com o NINO3 do mês de Fevereiro e DIPOLo de Julho. Dessas classes, foi encontrado pelos valores do BIC, que as classes 5 e 6 eram melhores que a combinação DIPOLo e NINO3. Igualmente para o ajuste variou-se a opção *num\_restarts*, tendo-se uma considerável quantidade de modelos.

## **9.6 - COMPARAÇÃO DAS SÉRIES SINTÉTICAS GERADAS PELOS MODELOS ARMA E NHMMs**

Depois de analisar os momentos produzidos pelas 5 classes de NHMM com diferentes *inputs*, mencionadas na seção anterior. Foi encontrado que somente alguns NHMMs com o preditor NINO3 geravam séries com média aproximada à observada, os outros modelos não conseguiram representar essa estatística básica, visto que quando fizeram os diagramas de caixa, as caixas ficaram bem distantes da média observada e com essa deficiente representação, os modelos apresentam problemas para representar as outras estatísticas, portanto, mais deficientes em representar a autocorrelação de primeira ordem e a persistência.

No entanto, tiveram-se modelos que, embora, representaram mais ou menos “aceitavelmente” a média depois a variavam muito na representação das outras estatísticas ( $r_1$ ) e as métricas de persistência. O que justificou mais só tomar os resultados dos NHMMs com o índice NINO3. Sendo assim, as séries geradas que se compararam a seguir, foram as produzidas pelo AR(2) e o GA(2) e os modelos: NHMM-NLT (3),

NHMM-G (2 e 3). Portanto, uma análise similar à realizada na seção 9.4 é apresentada a continuação. Assim, nas seguintes figuras, apresentam-se os resultados desses modelos e os produzidos pelos HMMs homogêneos, sendo eles os diagramas de caixa, do lado direito do seu respectivo NHMM.

### 9.6.1 - Momentos

A Figura 9.30 (superior) mostra a média obtida pelos NHMMs. Observa-se como o desempenho deles diminui um pouco, ao observar os valores  $p$ . Para o desvio padrão os NHMMs mostram-se aceitáveis e as diferenças entre eles e os HMMs são mínimas. Para esse momento, continua o bom desempenho dos HMMs frente ao AR(2) e, principalmente, ao GA(2), porque entregam valores mais baixos e bem comportados com relação a esse modelo.

Também, existe uma diferença mínima nos resultados para o coeficiente de assimetria (Figura 9.31, superior). Os modelos continuam com um comportamento parecido aos HMMs, com um bom desempenho. Uma melhor representação da autocorrelação  $r_1$ , foi encontrada para o NHMM-NLT (3) e o NHMM-G (2), como se observa na Figura 9.31 (inferior). Mas, o comportamento do AR(2) continua sendo melhor. Enquanto que, o modelo GA(2) apresenta problemas ao representar essa autocorrelação e o coeficiente de assimetria.

O comportamento dos NHMMs para as vazões máximas continua sendo similar aos HMMs, já que existe uma mínima variação entre os resultados (Figura 9.32, superior). Nota-se como a diferença de desempenho dos NHMMs continua frente ao AR(2) e sobretudo a respeito do GA(2), que subestima as vazões máximas e mínimas.

Já na Figura 9.32 (inferior), observa-se que os HMMs tem melhor representação frente ao AR(2) e os NHMMs melhoram a representação, omitindo o NHMM-NLT (3). O NHMM-G (2), melhora a representação frente a sua versão homogênea. Portanto, existe um ganho ao utilizar HMMs não homogêneos, principalmente para representar a autocorrelação de primeira ordem.

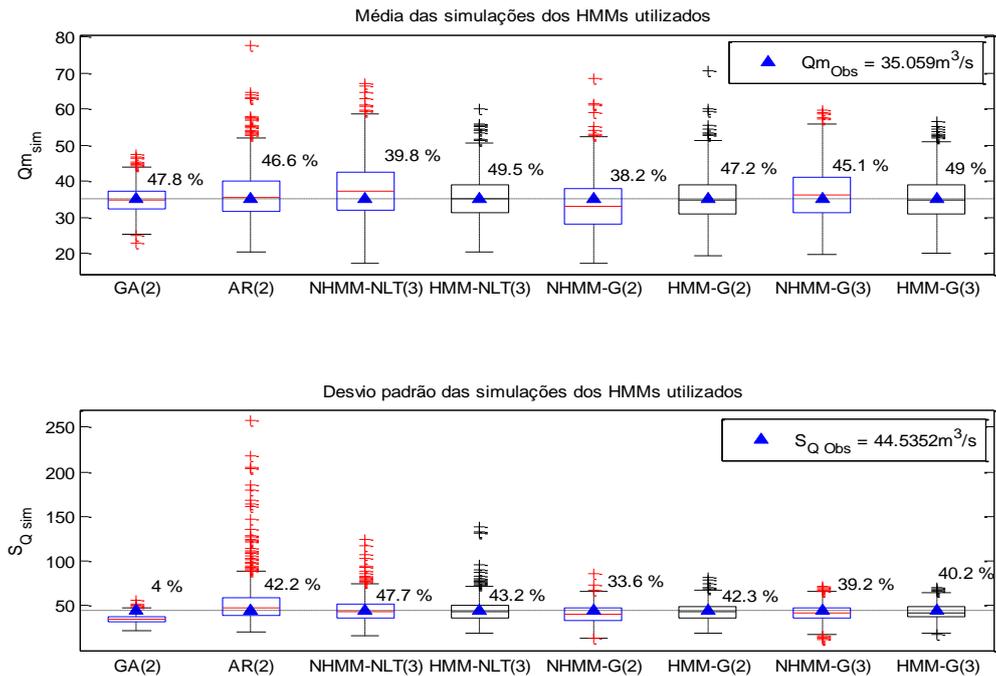


Figura 9.30 - Média (superior) e desvio padrão (inferior) das séries sintéticas geradas pelos NHMMs, AR(2) e o GA(2).

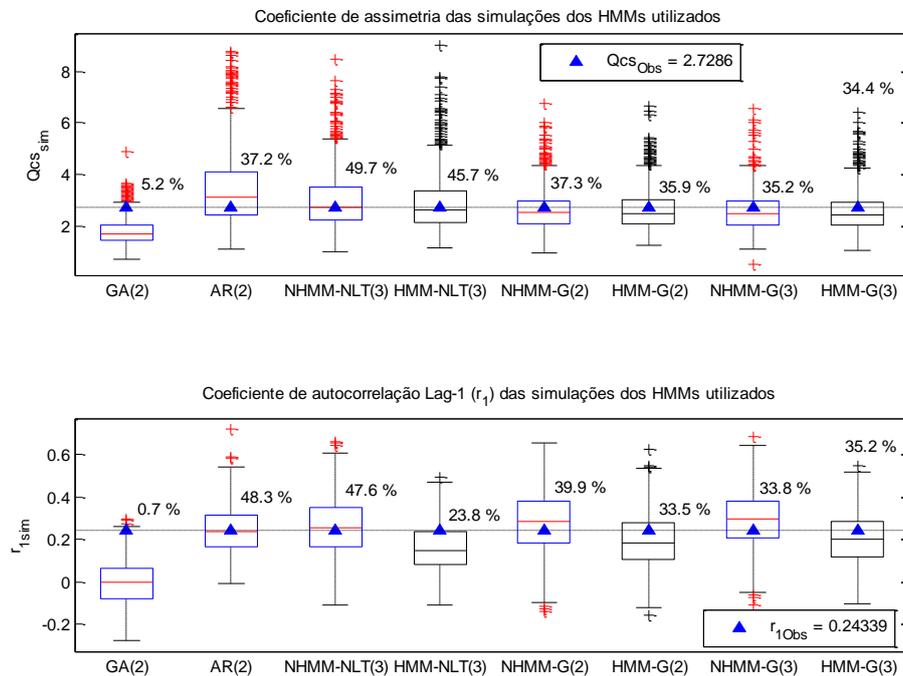


Figura 9.31 - Coeficiente de assimetria (superior) e autocorrelação de primeira ordem (inferior) das séries sintéticas geradas pelos NHMMs, AR(2) e o GA(2).

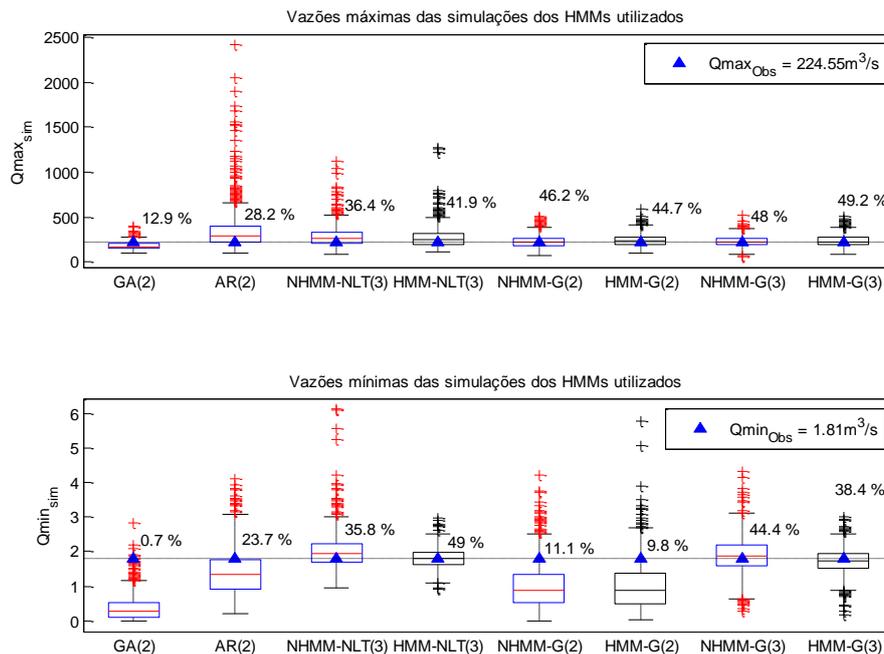


Figura 9.32 - Vazão máxima (superior) e vazão mínima (inferior) das séries sintéticas geradas pelos NHMMs, AR(2) e o GA(2).

### 9.6.2 - Métricas de Persistência

Para várias das métricas calculadas, encontrou-se que os NHMMs apresentam resultados melhores que os HMMs para os períodos úmidos. Porque, os valores  $p$  aumentam e a amplitude das caixas é maior, aumentando a tendência (ou variabilidade) de que um parâmetro seja bem simulado (Stedinger e Taylor, 1982).

No caso do valor de da estatística  $T$  para os períodos secos, encontrou-se melhor representação pelos modelos homogêneos, utilizando a  $P_{33}$ , somente o modelo NHMM-G(2) melhora a simulação dessa métrica. Para a mediana, o desempenho consegue ser mais adequado para o NHMM-NLT (3). Nos NHMM-Gs, o comportamento tende a ser pior, pois os valores  $p$  diminuem. No entanto, para os estados úmidos, obteve-se a existência de um grande aumento da variabilidade para os dos limiares utilizando NHMMs, ressaltando-se o desempenho do NHMM- G (3) (Figura 9.33). Nessa métrica o GA(2), tem um desempenho similar ao AR(2) e tendem a subestimar a  $T$ . Essa mesma análise pode ser estendida para os comprimentos máximos e médios tanto de períodos secos e úmidos

(Figura 9.34 e Figura 9.35), com a diferença que o GA(2) muda de comportamento com respeito ao AR(2) quando se calculam comprimentos médios dos períodos úmidos além de resultar em um modelo mais ruim no cálculo de comprimentos máximos e médios para períodos secos.

Tomando ambos os limiares para os volumes de déficit máximos e médios, o AR(2) representa bem essa métrica. Porém, apresenta problemas quando os volumes são úmidos (Figura 9.27 e Figura 9.28), já que tende a subestimar os valores. Entretanto, entre os NHMMs somente melhora em desempenho o NHMM-G(2), e caso inverso ocorre quando os períodos são úmidos, pois melhoram o NHMM-NLT (3) e NHMM-G(3); sem importar o limiar que se tenha, ressaltando que eles melhoram de acordo com a mediana. Por outro lado, o modelo GA(2) sempre apresenta um ruim desempenho.

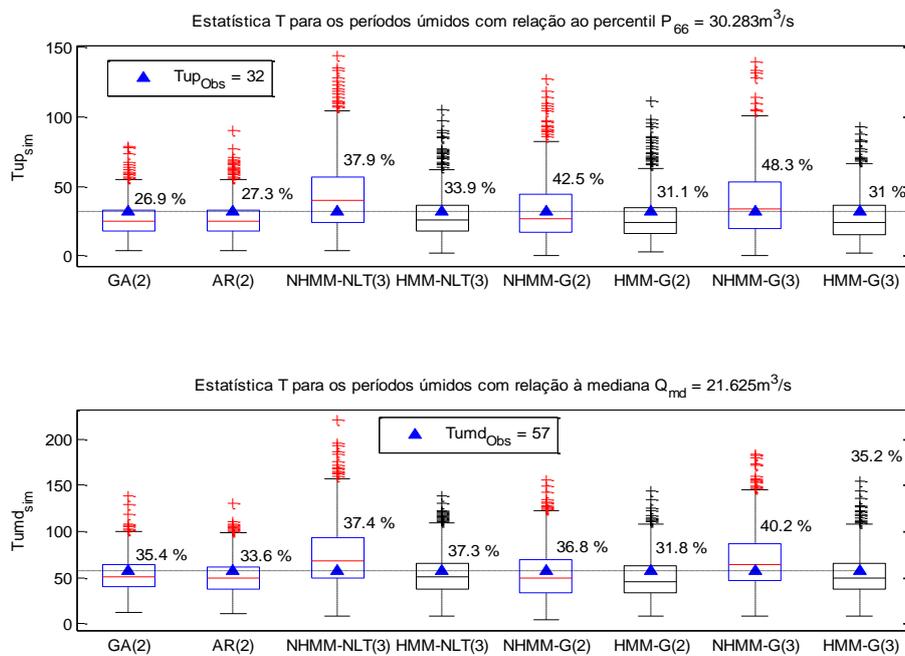
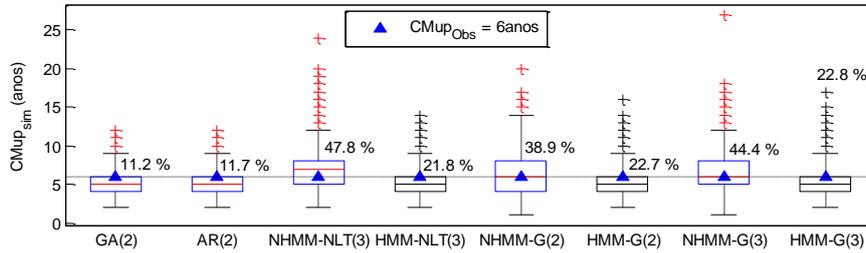


Figura 9.33 - Estatística T dos períodos úmidos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

Comprimento máximo dos períodos úmidos das simulações (CMup) dos HMMs utilizados com relação ao percentil  $P_{66} = 30.283m^3/s$



Comprimento máximo dos períodos úmidos das simulações (CMum) dos HMMs utilizados com relação à mediana  $Q_{md} = 21.625m^3/s$

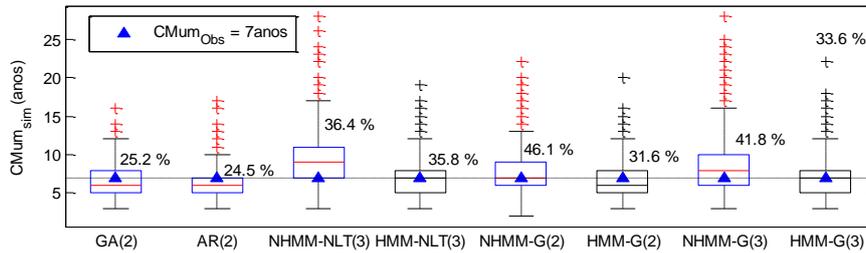
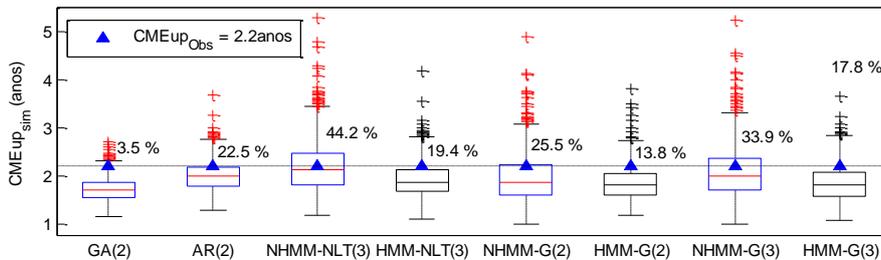


Figura 9.34 - Comprimento máximo dos períodos úmidos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

Comprimento médio dos períodos úmidos das simulações (CMEup) dos HMMs utilizados com relação ao percentil  $P_{66} = 30.283m^3/s$



Comprimento médio dos períodos úmidos das simulações (CMEum) dos HMMs utilizados com relação à média  $Q_{md} = 21.625m^3/s$

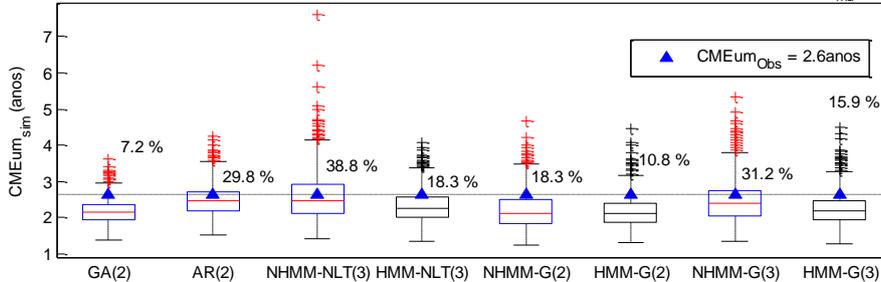


Figura 9.35 - Comprimento médio dos períodos úmidos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

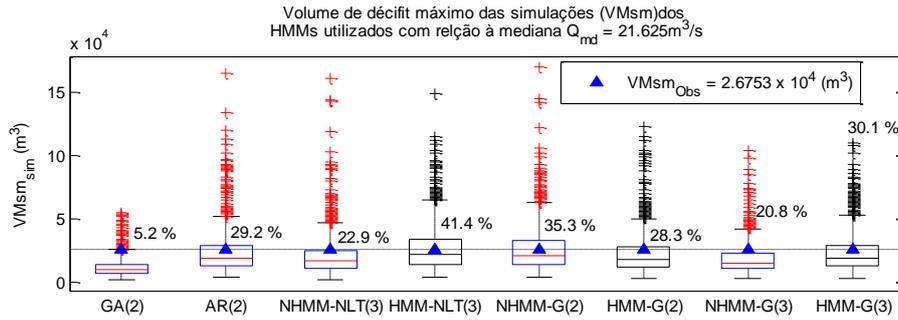
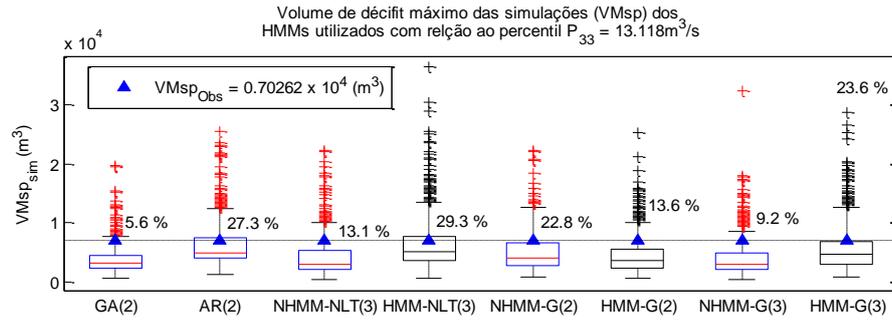


Figura 9.36 - Volume de déficit máximo com relação ao percentil  $P_{33}$  (superior) e à mediana  $Q_{md}$  (inferior)

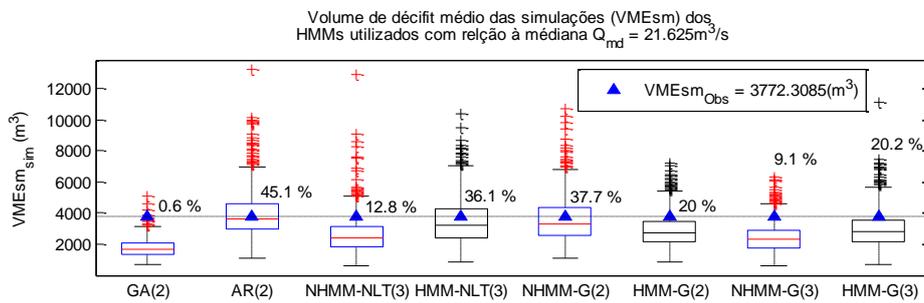
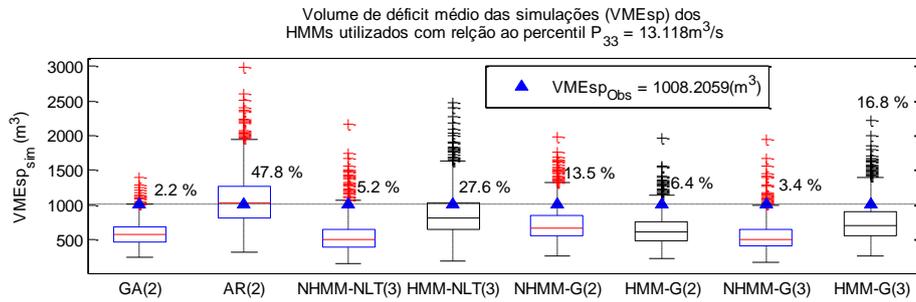


Figura 9.37 - Volume de déficit médio com relação ao percentil  $P_{33}$  (superior) e à mediana  $Q_{md}$  (inferior).

De acordo com o apresentado acima, os NHMMs tendem a variar seu desempenho para representar as métricas de persistência e, em alguns casos, são deficientes em relação aos HMMs, principalmente ao representar métricas de estados secos. Em geral, nota-se como os modelos não homogêneos apresentam melhor desempenho para as métricas de períodos úmidos e quando se toma por limiar a mediana. Desta análise, não se descarta que os modelos não homogêneos brindem um ganho na representação da série de vazões afluentes ao reservatório Orós. Entretanto, se visualizou mais a deficiências das séries calculadas por meio da distribuição Gama de dois parâmetros. O leitor pode conferir alguns dos resultados não apresentados aqui no Apêndice A.

## 10 - CONCLUSÕES

Os modelos estocásticos para a geração de séries sintéticas de vazões são ferramentas importantes que permitem avaliar o desempenho de sistemas de recursos hídricos, criar políticas operacionais ou gerar cenários de risco. Dessa forma, vários modelos têm sido desenvolvidos historicamente, incluindo aqueles que permitem representar a variabilidade e a persistência de períodos secos e úmidos que são comumente observados em séries de vazões.

Nesta dissertação, procurou-se avaliar o desempenho de modelos de Markov com Estados Ocultos (HMMs), e de sua versão não-homogênea (NHMMs), que faz uso de indicadores climáticos, na representação das vazões anuais afluentes ao reservatório Orós, no Nordeste do Brasil. O desempenho desses modelos foram comparados com o desempenho dos clássicos modelos ARMA e de um modelo simples, que considera que as vazões anuais são independentes e distribuídas conforme a distribuição Gama.

De uma forma geral, baseando-se nos critérios de AIC e BIC, pode-se dizer que os modelos HMMs com dois e três estados são adequados para modelar as vazões anuais afluentes ao reservatório Orós, embora possa haver divergência, dependendo do critério escolhido. O critério AIC indicou a possibilidade de utilizar modelos com 2 e 3 estados, enquanto o critério BIC sugeriu utilizar modelos com apenas 2 estados, concordante com o critério de parcimônia, visto que um aumento do número de estados, aumenta o número de parâmetros. Quando se observa a variação das probabilidades dos estados ocultos no tempo, é possível notar que modelos com apenas 2 estados representam adequadamente as variações de vazões e a persistência observada, enquanto que modelos HMMs de ordem superior não apresentam ganho significativo, devido a pouca representação física que eles trazem para análise. Entretanto, alguns modelos com 3 estados ocultos tiveram desempenho superior na representação de estatísticas relacionadas à persistência hidrológica.

Os modelos HMM-Gs, baseados na distribuição Gama, foram capazes de representar adequadamente os momentos observados, embora o referido modelo com 2 estados tenha encontrado dificuldade de representar o coeficiente de assimetria da série histórica de

vazões. Os HMM-NLTs, baseados na distribuição Normal ajustada ao logarítmico das vazões, também conseguiram representar adequadamente os momentos da série observada, sendo que o modelo com 2 estados apresentou deficiências no cálculo da autocorrelação de primeira ordem.

Os HMMs representam bem as distribuições marginais, principalmente os HMM-NLTs, o que é uma qualidade importante de modelos estocásticos (Bishop, 2006). Em relação à autocorrelação, os modelos HMMs tiveram desempenho satisfatório, tendo em vista que não foi possível, por meio de um teste de hipóteses formal, rejeitar a hipótese nula de que a série observada tenha sido gerada por esses modelos.

Em relação à capacidade desses modelos representarem a persistência hidrológica, pode-se afirmar, com base em testes formais de hipóteses, que o modelo que admite que as vazões anuais são independentes no tempo, e distribuídas de acordo com distribuição Gama, não é adequado para representar as séries anuais de vazões afluentes ao reservatório Orós. De uma forma geral, tendo como base o valor-p de um teste formal de hipóteses realizado para cada modelo, e para cada uma das métricas escolhidas para representar a persistência hidrológica, pode-se afirmar que tanto o modelo AR(2), quanto os modelos HMMs (HMM-NLT2, HMM-G2 e HMM-G3) são capazes de representar de forma adequada a persistência hidrológica, embora não seja possível indicar um único melhor modelo, tendo em vista que o desempenho relativo varia dependendo da métrica de persistência escolhida. Com base nos resultados desses testes de hipóteses, só foi possível rejeitar a hipótese de que o modelo HMM-G2 não é adequado para representar a vazão mínima (valor-p = 9.8%) e o volume de déficit médio, baseado nas vazões anuais inferiores ao percentil 33% (valor-p = 6.4%). Em todas as outras métricas de persistência, não foi possível rejeitar a hipótese nula de que os modelos HMMs e ao modelo AR(2) são adequados. Entretanto, vale a pena chamar a atenção para o fato de que o modelo AR(2), embora tenha passado no teste de hipóteses de representação das vazões máximas, gerou valores sintéticos excessivamente altos, não condizente com a realidade, e teve dificuldades na representação dos comprimentos médio e máximo de períodos úmidos, acima do percentil de 66%, com valores-p da ordem de 11%, próximos da região de rejeição.

Além do modelo homogêneos (HMMs), foram ajustados vários HMMs não homogêneos (NHMM) empregando os indicadores climáticos Dipolo do Atlântico e o NINO3. Pode-se

observar que o emprego de indicadores climáticos resultou numa melhoria significativa na capacidade do modelo HMM-NLT3 em representar a autocorrelação temporal, embora o referido modelo tenha passado no teste de hipóteses para representar a autocorrelação mesmo sem o uso dos indicadores. A introdução dos indicadores climáticos também melhorou o desempenho dos modelos HMM-G2 e HMM-G3 em representar o comprimento médio de períodos úmidos, com valores-p aumentando de 10 e 15%, no caso homogêneo, para 18 e 31% no caso não-homogêneo. Por fim, a dificuldade que o HMM-G2 tinha de representar o volume de deficit médio (valor-p = 6.4%) foi aliviada, de modo que o valor-p do modelo NHMM-G2 passou para 13.5%.

Pelos resultados desse estudo, pode-se dizer que os modelos de Markov com Estados Ocultos, tanto homogêneos quanto não homogêneos, apresentaram resultados satisfatórios ao representar as características estatísticas da série estudada, com desempenho similar aos modelos ARMA, tendo superado tais modelos na representação de vazões máximas e mínimas, assim como em algumas métricas de persistência, tais como, comprimentos máximo e médio de períodos úmidos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Ajami, N. K., Hornberger, G. M., & Sunding, D. L. (2008). "Sustainable water resource management under hydrological uncertainty". *Water Resources Research*, **44**(11), 1–10p.
- Akintug, B., & Rasmussen, P. F. (2005). "A Markov switching model for annual hydrologic time series". *Water Resources Research*, **41**(9), 1–10p.
- Aksoy, H. (2000). "Use of Gamma Distribution in Hydrological Analysis". *Turk J Engin Environ Sci*, **24**, 419–428.
- Alexandre, A. M. B. (2012). *Previsão De Vazões Mensais Para O Sistema Interligado Nacional Utilizando Informações Climáticas*. Tese de Doutorado, Universidade Federal Do Ceará, Departamento De Engenharia Hidráulica E Ambiental, 293p.
- Alves, J. M. B., Souza, E. B. De, Repelli, C. A., Vitorino, M. I., e Ferreira, N. S. (1997). "Episódios de La Nina na Bacia Do Oceano Pacifico Equatorial e a distribuição Sazonal e Intrasazonal das Chuvas no Setor Norte do Nordeste Brasileiro". *Revista Brasileira de Meteorologia*, **12**(1), 63–76p.
- ANA (2010a): Agência Nacional de Águas (Brasil). *Atlas Brasil: abastecimento urbano de água: panorama nacional / Agência Nacional de Águas*. Engecorps/Cobrape, Brasília, Vol 1, 72p.
- ANA (2010b). Agência Nacional de Águas (Brasil). *Atlas Brasil: abastecimento urbano de água: resultados por estado / Agência Nacional de Águas*. Engecorps/Cobrape, Brasília, Vol 2, 92p.
- Andreoli, R. V., e Kayano, M. T. (2007). "A Importância Relativa Do Atlântico Tropical Sul E Pacífico Leste Na Variabilidade De Precipitação Do Nordeste Do Brasil". *Revista Brasileira de Meteorologia*, **22**(1), 63–74p.
- Araújo, J. K. (1991). *Método dos Fragmentos Aplicado a Rios Intermitentes: Avaliação dos Erros Introduzidos no Cálculo da Disponibilidade de Reservatórios*. Dissertação Mestrado, Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza.
- AsBOM, (2012). Australian Bureau of Meteorology, consultado em Outubro de 2012, disponível em: <http://www.bom.gov.au/climate/enso/indices/about.shtml>
- Barnes, F. B. (1954). "Storage required for a city water supply". *J. Inst. Eng.* , 26, 198-203p.

- Bayer, D. M., e Castro, N. M. dos R. (2012). Modelagem e Previsão de Vazões Médias Mensais do Rio Potiribu Utilizando Modelos de Séries Temporais. *Revista Brasileira de Recursos Hídricos*, **17**(2), 229–239p.
- Bellone, E., Hughes, J. P., e Guttorp, P. (2000). "A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts". *Climate Research*, **15**, 1–12p.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer, 738 p.
- Box, G. E. P.; Jenkins, G. M. (1976). *Time series analysis forecasting and control*. San Francisco: Holden- Day, E.U.A, Edição revisada, 785p.
- Box, G. E. P.; Jenkins, G. M. e Reinsel, G. C., (1994). *Time Series Analysis: forecasting and control*, 3.rd ed., Prentice Hall, New Jersey, 587p
- Brabo Alves, J., Servain, J., e Campos, J. (2009). "Relationship between ocean climatic variability and rain-fed agriculture in northeast Brazil". *Climate Research*, **38**(May), 225–236p.
- Bracken, C. W. (2011). *Seasonal to Inter-Annual Streamflow Simulation and Forecasting on the Upper Colorado River Basin and Implications for Water Resources Management*. Msc Thesis, University of Colorado, 98p.
- Campos, J. N. B. (1996). *Dimensionamento de Reservatórios: o método do diagrama triangular de regularização* , Fortaleza, Edições UFC, 71p.
- Ceará, (2008). *Cenário Atual Dos Recursos Hídricos Do Ceará*. Fortaleza - Ceará: Conselho de Altos Estudos e Assuntos Estratégicos, 88 p.
- Charles, S. P., Bates, B. C., Whetton, P. H., e Hughes, J. P. (1999). "Validation of downscaling models for changed climate conditions : case study of southwestern Australia". *Climate Research*, **12**, 1–14p.
- Chatfield, C., (1975). *The analysis of time series: Theory and practice*, Chapman and Hall, London, 263p.
- Chiang, J. C. H., Kushnir, Y., e Zebiak, S. E. (2000). "Interdecadal changes in eastern Pacific ITCZ variability and its influence on the Atlantic ITCZ". *Geophysical Research Letters*, **27**(22), 3687–3690p.
- COGERH, (2011). *Inventário Ambiental Açude Orós*. Fortaleza – Ceará, 126p.
- Cysne, A. P. C. P. (2007). *Vulnerabilidade De Reservatórios Em Rios De Alta Variabilidade Em Um Cenário De Mudanças Climáticas*. Dissertação de Mestrado, Universidade Federal Do Ceará, Departamento De Engenharia Hidráulica E Ambiental, 96p.

- Douglas, E. M., Vogel, R. M., e Kroll, C. N. (2002). "Impact of Streamflow Persistence on Hydrologic Design". *Journal Of Hydrologic Engineering*, **7**(June), 220–227p.
- Farias, J. A. M. (2003). *Métodos De Geração De Vazões Mensais E Suas Influências Sobre A Curva De Garantia Versus Vazão Regularizada Em Reservatórios Do Estado Do Ceará*. Dissertação de Mestrado, Universidade Federal Do Ceará, Departamento De Engenharia Hidráulica E Ambiental, 131p.
- Ferreira, A. G., e Mello, N. G. da S. (2005). "Principais Sistemas Atmosféricos Atuantes Sobre A Região Nordeste Do Brasil E A Influência Dos Oceanos Pacífico E Atlântico No Clima Da Região". *Revista Brasileira de Climatologia*, **1**(1), 15–28p.
- Filho, E. C. C. (2007). *Regularização De Vazões Em Reservatórios Através Dos Modelos Mensal E Bi-Sazonal: A Duração Da Estação Seca E A Proposta Da Evaporação Equivalente*. Dissertação de Mestrado, Universidade Federal Do Ceará, Departamento De Engenharia Hidráulica E Ambiental, 88p.
- Fortin, V., Perreault, L., e Salas, J. D. (2004). "Retrospective analysis and forecasting of streamflows using a shifting level model". *Journal of Hydrology*, **296**(1-4), 135–163p.
- Fu, G., Charles, S. P., e Kirshner, S. (2012). "Daily rainfall projections from general circulation models with a downscaling nonhomogeneous hidden Markov model (NHMM) for south-eastern Australia". *Hydrological Processes*, 11p.
- FUNCEME, (2012). Sistemas Atmosféricos Atuantes Sobre o Nordeste, consultado em Novembro 15, disponível em: <http://www.funceme.br/index.php/areas/tempo/saiba-mais/sistemas-atmosfericos-atuantes-sobre-o-nordeste>
- Gatto, L. C. S. (1999). *Diagnóstico Ambiental da Bacia do Rio Jaguaribe*. Salvador, Bahia, 77p.
- Grayson, R. B., Argent, R. M., Nathan, R. J., McMahon, T. A., & Mein, R. G. (1996). *Hydrological Recipes: Estimation Techniques in Australian Hydrology*. Melbourne, Victoria, Australia: Coop. Res. Cent. for Catchment Hydrol, 125p.
- Greene, A M, Robertson, A. W., e Kirshner, S. (2008). "Analysis of Indian monsoon daily rainfall on subseasonal to multidecadal time-scales using a hidden Markov model". *Quarterly Journal of the Royal Meteorological Society*, **134**, 875–887p.
- Greene, Arthur M., Robertson, A. W., Smyth, P., e Triglia, S. (2011). "Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model". *Quarterly Journal of the Royal Meteorological Society*, **137**(655), 347–359p.

- Hashimoto, T., Stedinger, J. R., & Loucks, D. P. (1982). Reliability, Resiliency, and Vulnerability Criteria For Water Resource System Performance Evaluation. *Water Resources Research*, **18**(1), 14–20p.
- Hastenrath, S. (1984). "Interannual variability and annual cycle: mechanisms of circulation and climate in the tropical Atlantic". *Mon. Wea. Rev.*, **112**, 1097-1107p.
- Hastenrath, S. (2000). "Interannual and longer-term variability of upper air circulation in the Northeast Brazil-tropical Atlantic sector". *Journal Of Geophysical Research*, **105**, 7327–7335p.
- Hastenrath, S. (2006). "Circulation and teleconnection mechanisms of Northeast Brazil droughts". *Progress in Oceanography*, **70**(2-4), 407–415p.
- Hastenrath, S., Heller, L. 1977. "Dynamics of climatic hazards in Northeast Brazil". *Quartely Journal of the Royal Meteorological Society*, **103**, 77-92p.
- Hazen, A. (1914). "Storage to be provided in impounding reservoirs for municipal water supply". *Transactions American Society of Civil Engineers*, **77**, 1539-1640, p.
- Hirsch, R. M., Helsel, D. R., Cohn, T. A. e Gilroy, E. J. (1993). "Statistical analysis of hydrologic data". In: *Handbook of Hydrology*. D. R. Maidment) (eds), Chapter 17., McGraw-Hill, New York, USA, 17.1–17.55p.
- Hughes, James P., e Guttorp, P. (1994). "Incorporating spatial dependence and atmospheric data in a model of precipitation". *Journal of Applied Meteorology*, **33**(12), 1503–1515p.
- Hughes, J. P., Guttorp, P., e Charles, S. P. (1999). "A non-homogeneous hidden Markov model for precipitation occurrence". *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **48**(1), 15–30p.
- Hurst, H. E. (1951). "Long-term storage capacity of reservoirs". *Transaction of American Society of Civil Engineering*, **166**, 770-799 p.
- IME USP, (2012). "Normalidade dos dados e dos resíduos1. Teste de aderência (qui-quadrado)". *Métodos Estatísticos*. Consultado em Janeiro, disponível em: [http://www.ime.usp.br/~yambar/MI404-Metodos Estatisticos/Aula 6 Distribuicao empirica Testes Aderencia/internet-Qui\\_Quadrado.pdf](http://www.ime.usp.br/~yambar/MI404-Metodos_Estatisticos/Aula_6_Distribuicao_empirica_Testes_Aderencia/internet-Qui_Quadrado.pdf)
- INPE/CPTEC, (2012). El Niño e La Niña. Consultado em Novembro 17, 2012, disponível em: <http://enos.cptec.inpe.br/>
- IRI, (2012a). Overview of the ENSO System. Consultado em Dezembro 15, 2012, disponível em: <http://iri.columbia.edu/climate/ENSO/background/basics.html>

- IRI, (2012b). Kaplan Extended V2. Consultado em Dezembro 20, 2012, disponível <http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/>
- IRI, (2012c). Overview of the ENSO System, Consultado em Dezembro 15, 2012, disponível em: <http://iri.columbia.edu/climate/ENSO/background/monitoring.html>
- Ito, E. R. K., e Coelho, C. A. dos S. (2012). "Impacto da Variabilidade Climática do Fenômeno ENOS na Estação Chuvosa da Região Nordeste do Brasil Nos Últimos Anos". *Anais do XVII Congresso Brasileiro de Meteorologia*.
- Jackson, B. B. (1975). "Markov Mixture Models for Drought Lengths". *Water Resources Research*, **11**(I), 64–74p.
- Juang, B. H., e Rabiner, L. R. (1991). "Hidden Markov Models for Speech Recognition". *Technometrics*, **33**(3), 251–272p.
- Kane, R. P. (1989). "Relationship between the southern oscillation/El Niño and rainfall in some tropical and midlatitude regions". *Proc. Indian Acad. Sci. Earth Planet Sci.*, **3**, 223-235p.
- Kaplan, A., Cane, M. A., Kushni, Y., Clement, A. C., Blumenthal, M. B., Rajagopalan, B., e Bortoraley, C. (1998). "Analyses of global sea surface temperature 1856-1991". *Journal of Geophysical Research*, **103**(C9), 18,567–18,589p.
- Kayano, M. T., e Andreoli, R. V. (2006). "Relationships between rainfall anomalies over northeastern Brazil and the El Niño–Southern Oscillation". *Journal of Geophysical Research*, **111**(D13).
- Kirshner, S. (2005a). *Modeling of Multivariate Time Series Using Hidden Markov Models*. PhD Thesis, University Of California, Irvine, 223p.
- Kirshner, S. (2005b). *Quick Start Manual for the MVN-HMM Toolbox*. Donald Bren School of Information and Computer Science, University of California, Irvine.
- Lambert, M. F., Whiting, P., Metcalfe, V., Whiting, J. P., e Metcalfe, A. V. (2003). "A non-parametric hidden Markov model for climate state identification". *Hydrology and Earth System Sciences*, **7**(5), 652–667p.
- Lázaro, Y. M. C., Silveira, C. da S., & Filho, F. de A. de S. (2011). Avaliação De Desempenho Dos Modelos Do Ippc-Ar4 No Nordeste Setentrional Do Brasil Quanto À Variabilidade Plurianual Da Precipitação No Século XX. *Anais do XIX Simpósio Brasileiro de Recursos Hídricos*, Maceió.
- Lima, C. H. R. (2010). "Análise E Modelagem Da Série Histórica De Fortaleza Por Meio De Um Modelo De Markov Escondido Não-Homogêneo". *X Simpósio de Recursos Hídricos do Nordeste*, 1–15p.

- Loucks, D. P., Beek, E. van, Stedinger, J. R., Dijkman, J. P. M., & Villars, M. T. (2005). Concepts in Probability, Statistics and Stochastic Modelling. In: *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications* (Studies an., pp. 168–229). WL - Delft Hydraulics, The Netherlands: UNESCO Publishing.
- Mallya, G., Tripathi, S., e Govindaraju, R. S. (2011). "Hidden Markov Model Based Probabilistic Assesment Of Droughts". *World Environmental and Water Resources Congress 2011: Bearing Knowledge for Sustainability ASCE 2011*, 1282–1291p.
- Marco, J.B.; Harboe, R. e Salas, J.D. (1989). *Stochastic hydrology and its use in water resources systems simulation and optimization*. NATO ASI Series, Dordrecht, Kluwer Academic, 487p.
- Marengo, José A. (2006). "Mudanças Climáticas Globais e seus Efeitos sobre a Biodiversidade". Em: *Biodiversidade*, Brasília, Ministério Do Meio Ambiente Secretaria De Biodiversidade E Florestas, 201p.
- Marengo, Jose A, e Valverde, M. C. (2007). "Caracterização do clima no Século XX e Cenário de Mudanças de clima para o Brasil no Século XXI usando os modelos do IPCC-AR4". *Revista Multiciência*, **8**, 5–28p.
- Marengo, Jose A. (2008). "Vulnerabilidade, impactos e adaptação à mudança do clima no semi-árido do Brasil". *Parcerias Estratégicas*, **27**, 149-176p.
- Matalas, N.C. (1967). "Mathematical assessment of synthetic hydrology". *Water Resources Research*, **3**(4), 937–45p.
- Matalas, N. C. (1997): "Stochastic hydrology in the context of climate change". *Climatic Change*, **37**, 89–101p.
- Mehrotra, R., A. Sharma, e I. Cordery (2004), Comparison of two approaches for downscaling synoptic atmospheric patterns to multisite precipitation occurrence, *J. Geophys. Res.*, **109**(D14)
- Mehrotra, R., e Sharma, A. (2005). "A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences". *Journal of Geophysical Research*, **110**(D16), 1–13p.
- Meko, D. M. (2005). GEOS 585A, *Applied Time Series Analysis. Lesson 3*. University of Arizona, 1-6p.
- Mishra, A. K., e Singh, V. P. (2011). "Drought modeling – A review". *Journal of Hydrology*, **403**(1-2), 157–175p.

- Moura, A. D., Shukla, J. (1981). "On the dynamics of the droughts in Northeast Brazil: observations, theory and numerical experiments with a general circulation model". *J. Atmospheric Science*, **38**, 2653-2673p.
- Naghettini, M., e Pinto, É. J. de A. (2007). *Hidrologia Estatística*. Belo Horizonte: CPRM - Serviço Geológico do Brasil, 552 p.
- Prairie, J., Nowak, K., Rajagopalan, B., Lall, U., e Fulp, T. (2008). "A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data". *Water Resources Research*, **44**(6), 1–11p.
- Rabiner, L. R. (1989). "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". *Proceedings of the IEEE*, **77**(2), 257–286p.
- Rajagopalan, B., Lall, U., e Tarboton, D. G. (1996). "Nonhomogeneous markov model for daily precipitation". *Journal of Hydrologic Engineering*, **1**(1), 33–40p.
- Reynolds, R. W., e Smith, T. M. (1994). "Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation". *Journal of Climate*, **7**, 929–948p.
- Ribeiro, D. D. C. M. (2011). *Modelo de Previsão de Vazão Aplicado ao Nordeste Brasileiro Utilizando a Informação Climática Para Auxiliar a Operação de Hidrossistemas*. Dissertação de Mestrado, Universidade Federal Do Ceará, Departamento De Engenharia Hidráulica E Ambiental, 188p.
- Robertson, A. W., Kirshner, S., e Smyth, P. (2004). "Downscaling of Daily Rainfall Occurrence over Northeast Brazil Using a Hidden Markov Model". *Journal Of Climate*, **17**, 4407–4424p.
- Robertson, A. W., Kirshner, S., Smyth, P., CHARLES, S. P., & BATES, B. C. (2005). Subseasonal-to-Interdecadal Variability of the Australian Monsoon Over North Queensland. *Q. J. R. Meteorol. Soc.*, **131**, 1–26. doi:10.1256/qj.yy.n
- Robertson, A. W., & Smyth, S. K. P. (2003). *Hidden Markov models for modeling daily rainfall occurrence over Brazil*. Information and Computer Science University of California, Irvine, 38p.
- Salas, J.D.; Delleur, J.W.; Yejevich, V. e Lane, W.L. (1980). *Applied modelling of hydrological time series*. Water Resources Press Publications, Littleton, Colorado, E.U.A., 485p.
- Salas, J.D., Boes, D.C., (1980). "Shifting level modeling of hydrologic series". *Advances in Water Resources*, **3**, 59–63p.
- Salas, J.D. (1993). "Analysis and modelling of hydrological time series". In: D. Maidment (ed.), *Handbook of hydrology*, Chapter 17. New York, McGraw-Hill.

- Saldarriaga, J., e Yevjevich, V. (1970). "Application of run-lengths to hydrologic series". *Hydrology Papers Colorado State University*, 40-57p.
- Santos, M. B. Dos. (2009). *Indicadores De Sustentabilidade De Reservatórios No Semi-Árido Brasileiro: Uma Proposta Para A Bacia Do Rio Jaguaribe - Ceará*. Dissertação de Mestrado, Universidade Federal do Ceará, Centro de Tecnologia, Fortaleza, 128p.
- Sarmiento, F. J. (1989). *Aplicabilidade de Modelos de Geração de Vazão no Semi-Árido do Nordeste do Brasil*. Dissertação de Mestrado, Universidade Federal do Ceará, Departamento de Engenharia Hidráulica e Ambiental, 142p.
- Schöngart, J., e Junk, W. J. (2007). "Forecasting the flood-pulse in Central Amazonia by ENSO-indices". *Journal of Hydrology*, **335**(1-2), 124–132p.
- Silveira, C. da S., Alexandre, A. M. B., Souza Filho, F. de A. de, e Fernandes, W. de S. (2011a). Avaliação Das Projeções Para As Vazões Dos Modelos Do IPCC-Ar4 Para O Cenário A1b Para As Bacias Do Estado Do Ceará, *Anais do XIX Simpósio Brasileiro de Recursos Hídricos, Maceió, Brasil*, 1–18p.
- Silveira, C. da S., Filho, F. de A. de S., Alves, B. C. C., e Lázaro, Y. M. C. (2011b). Análise De Tendência Da Precipitação Dos Modelos Do Ipccl-Ar4 No Nordeste Setentrional Brasileiro Para O Cenário A1b No Século XXI, *Anais do XIX Simpósio Brasileiro de Recursos Hídricos, Maceió, Brasil*, 1–17p.
- Souza Filho, F. A., & Lall, U. (2003). "Seasonal to interannual ensemble streamflow forecasts for Ceara , Brazil: Applications of a multivariate , semiparametric algorithm". *Water Resources Research*, **39**(11), 1–13.
- Stedinger, J. R., e Taylor, M. R. (1982). Synthetic streamflow generation: 2. Effect of parameter uncertainty. *Water Resources Research*, **18**(4), 919–924.
- Steffen, D. (2010). *Intervalo "Bootstrap" Para Previsões De Séries Temporais Obtidas Pelo Método Theta*. Dissertação de Mestrado, Universidade Federal do Paraná, 116p.
- Storch, H. Von, e Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, New York,USA, 484 p.
- Sudler, C. E. (1927). "Storage Required for the Regulation of Stream Flow". *Trans. A.S.C.E.*, **91**, 622-660p.
- Sveinsson, O. G. B., Salas, J. D., Boes, D. C., & Pielke Sr., R. A. (2003). "Modeling the Dynamics of Long-Term Variability of Hydroclimatic Processes". *Journal Of Hydrometeorology*, **4**, 489–505p.

- Thomas, H. A.; Fiering, M. B. (1962). "Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation". In: MAAS, A. et al. *Design of water resources*. Harvard University press, Cambridge, MA, E.U.A, 459-463p.
- Thyer, M., e Kuczera, G. (2000). "Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model". *Water Resources Research*, **36**(11), 3301–3310.
- Thyer, M., e Kuczera, G. (2003a). "A hidden Markov model for modelling long-term persistence in multi-site rainfall time series" .1. Model calibration using a Bayesian approach. *Journal of Hydrology*, **275**(1-2), 12–26p.
- Thyer, M., & Kuczera, G. (2003b). "A hidden Markov model for modelling long-term persistence in multi-site rainfall time series". 2. Real data analysis. *Journal of Hydrology*, **275**(1-2), 27–48p.
- Thyer, M., Frost, A. J., e Kuczera, G. (2006). "Parameter estimation and model identification for stochastic models of annual hydrological data: Is the observed record long enough?". *Journal of Hydrology*, **330**(1-2), 313–328p.
- Timmermann, A. (2000). "Moments of Markov switching models". *Journal of Econometrics*, **96**, 75–111p.
- Uvo, C. B., Repelli, C. A., Zebiak, S. E., e Kushnir, Y. (1998). "The Relationships between Tropical Pacific and Atlantic SST and Northeast Brazil Monthly Precipitation". *Journal Of Climate*, **11**, 551–562p.
- Vogel, R. M., & Shallcross, A. L. (1996). The moving blocks bootstrap versus parametric time series models. *Water Resources Research*, **32**(6), 1875–1882.
- Wagner, R. G. (1996). "Decadal-scale trend in mechanisms controlling meridional sea surface temperature gradients in the tropical Atlantic". *Journal Of Geophysical Research*, **101**(C7), 16,683–16,694p.
- Wang, D. C., e Salas, J. D. (1991). *Forecasting Streamflow For Colorado River Systems*. Fort Collins, Colorado, USA, 245p.
- Werner, L., e Ribeiro, J. L. D. (2003). "Previsão De Demanda: Uma Aplicação Dos Modelos Box- Jenkins Na Área De Assistência Técnica De Computadores Pessoais". *Gestão & Produção*, **10**(1), 47–67.
- Whiting, J., Lambert, M., Metcalfe, A., e Kuczera, G. (2004). "Development of Non-homogeneous and Hierarchical Hidden Markov Models for Modelling Monthly Rainfall and Streamflow Time Series". *World Water Congress, Critical Transitions in Water and Environmental Resources Management World. A. S. C. E.*, 1–9p.

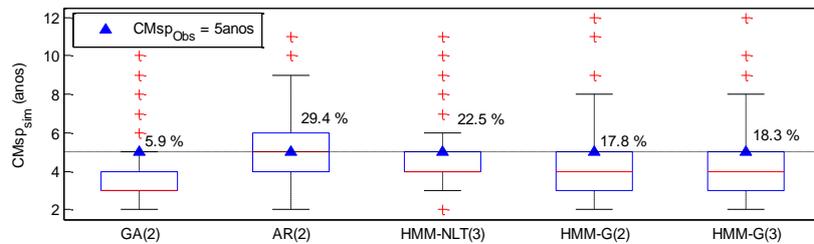
- Whiting, J. P. (2006). *Identification And Modelling Of Hydrological Persistence With Hidden Markov Models*. PhD Thesis, University Of Adelaide, Australia, 270p.
- Yevjevich, V.M. (1963). "Flutuations of wet and dry years. Part 1. Research data assembly and mathematical models". In: *Textit hydrology paper 1*, Colorado State University, Fort Collins, Colorado, E.U.A., 55p
- Zucchini, W. e Guttorp, P. (1991). "A hidden Markov model for space-time precipitation". *Water Resources Reseach*, **27**, 1917–1923p.
- Zucchini, W., e MacDonald, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction using R*. South African Actuarial Journal (Chapman e Hall/CRC., Vol. 10, p. 265). Boca Raton, FL, USA, 265p.

## **APÊNDICE A**

# APÊNDICE A – FIGURAS AUXILIARES DA COMPARAÇÃO DOS MODELOS ESTOCÁSTICOS

## A.1 – COMPARAÇÃO DOS MODELOS GA(2), AR(2) E HMMs

Comprimento máximo dos períodos secos das simulações (CMsp) dos HMMs utilizados com relação ao percentil  $P_{33} = 13.118m^3/s$



Comprimento máximo dos períodos secos das simulações (CMsm) dos HMMs utilizados com relação à mediana  $Q_{md} = 21.625m^3/s$

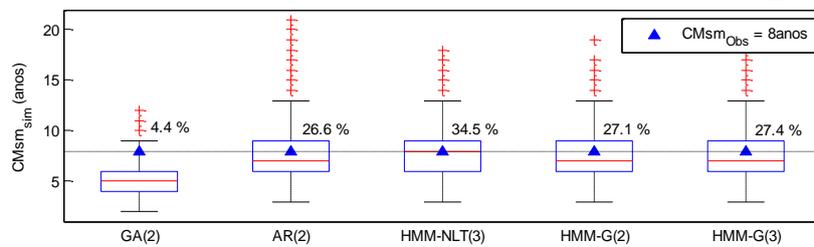
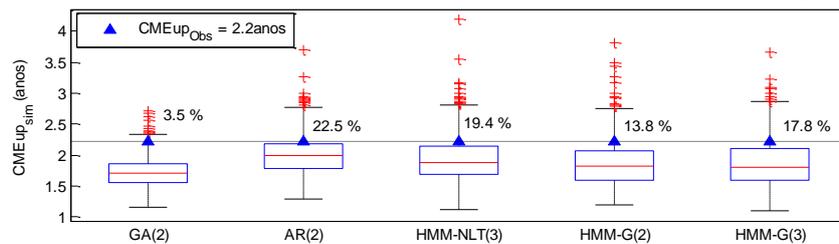


Figura A.1 - Comprimento máximo dos períodos secos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior).

Comprimento médio dos períodos úmidos das simulações (CMEup) dos HMMs utilizados com relação ao percentil  $P_{66} = 30.283m^3/s$



Comprimento médio dos períodos úmidos das simulações (CMEum) dos HMMs utilizados com relação à média  $Q_{md} = 21.625m^3/s$

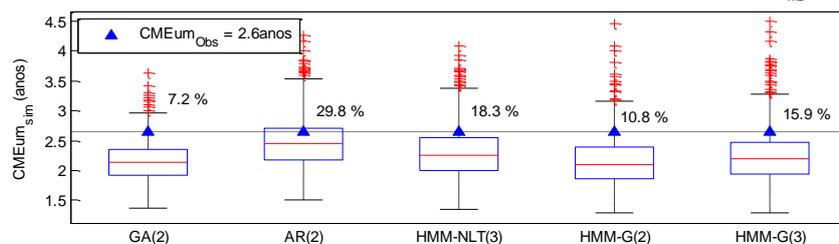


Figura A.2 - Comprimento médio dos períodos úmidos com relação ao percentil  $P_{33}$  (superior) e à mediana  $Q_{md}$  (inferior)

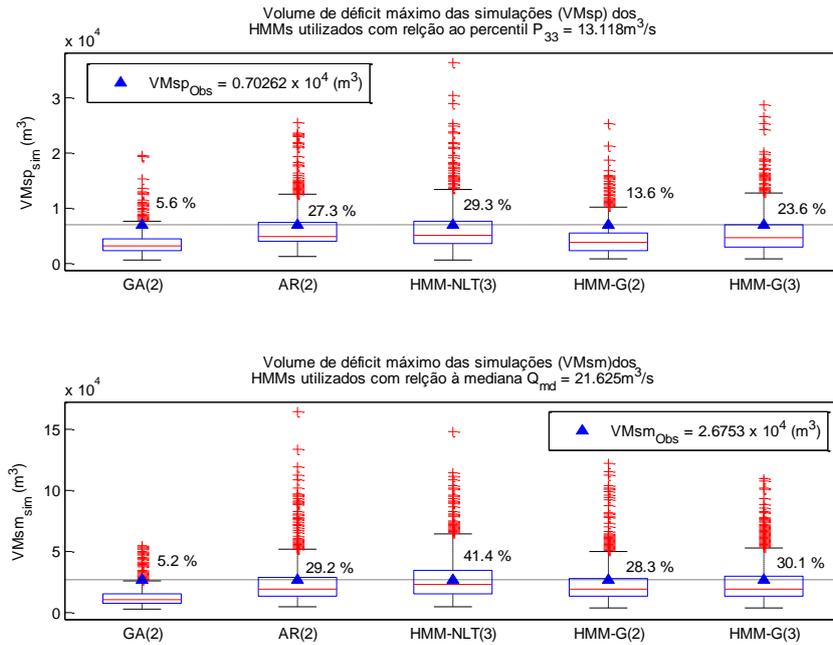


Figura A.3 - Volume de déficit máximo com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

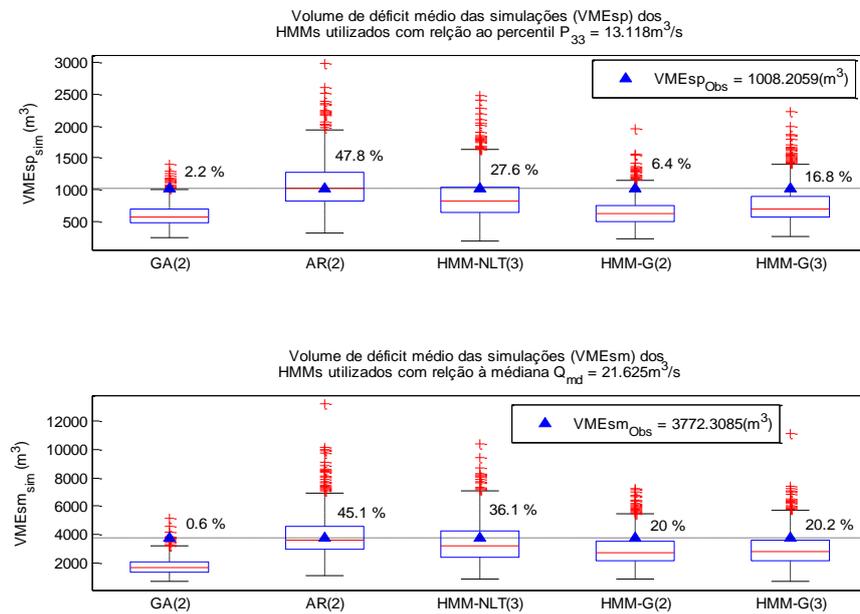


Figura A.4 - Volume de déficit médio com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

## A.2 – COMPARACAO DOS MODELOS GA(2), AR(2) E NHMMs

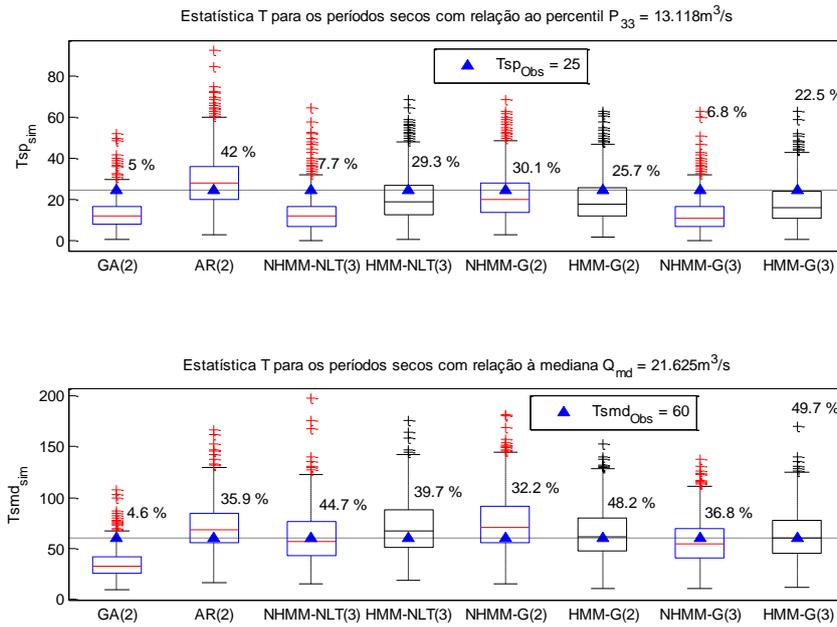


Figura A.5 - Estatística T dos períodos secos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

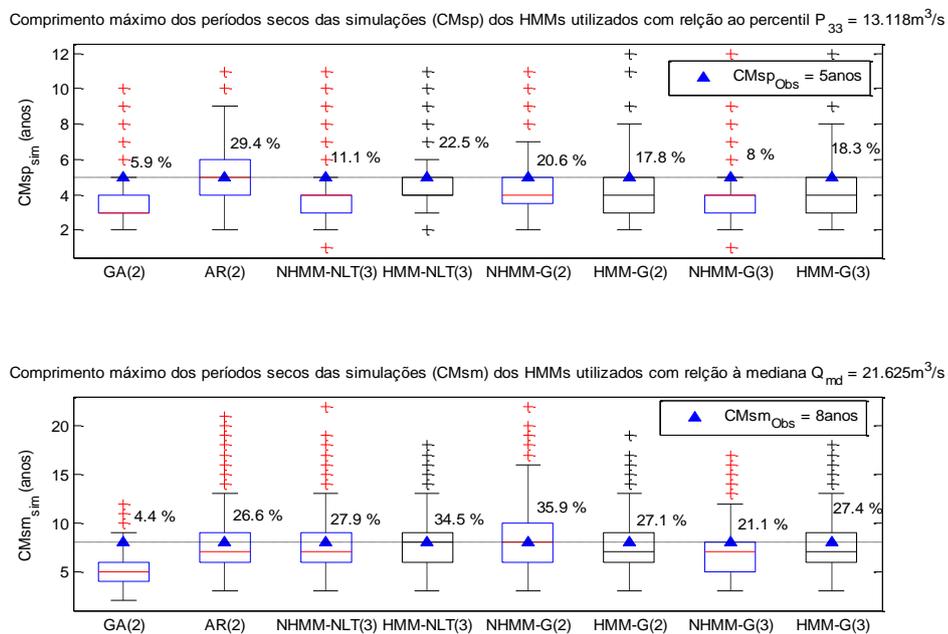
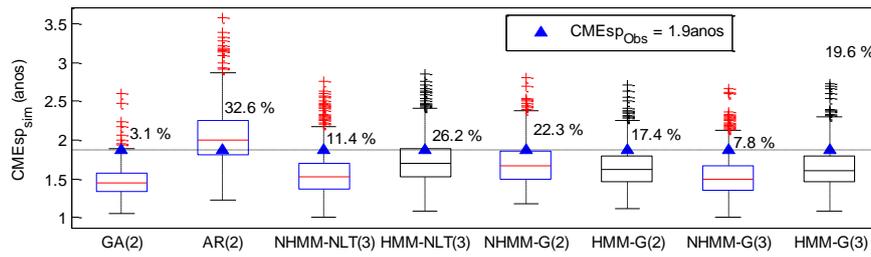


Figura A.6 - Comprimento máximo dos períodos secos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

Comprimento médio dos períodos secos das simulações (CMEsp) dos HMMs utilizados com relação ao percentil  $P_{33} = 13.118m^3/s$



Comprimento médio dos períodos secos das simulações (CMEsm) dos HMMs utilizados com relação à mediana  $Q_{md} = 21.625m^3/s$

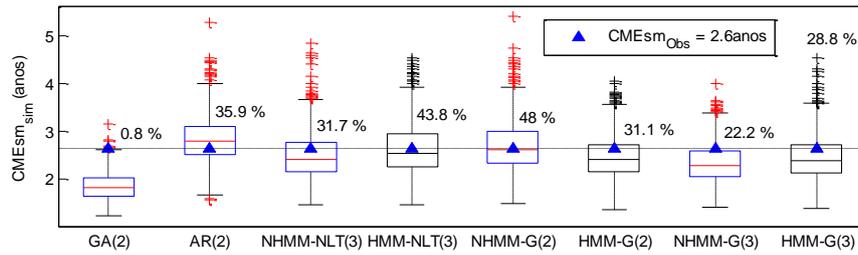


Figura A.7 - Comprimento médio dos períodos secos com relação ao percentil  $P_{66}$  (superior) e à mediana  $Q_{md}$  (inferior)

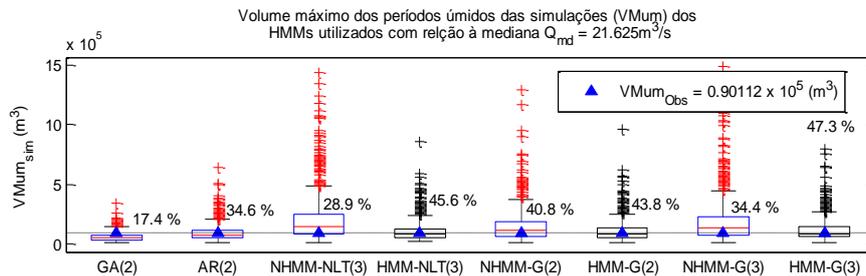
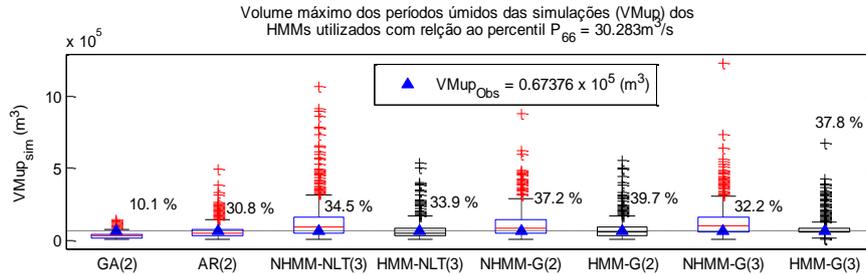


Figura A.8 - Volume máximo dos períodos úmidos com relação ao percentil  $P_{33}$  (superior) e à mediana  $Q_{md}$  (inferior)

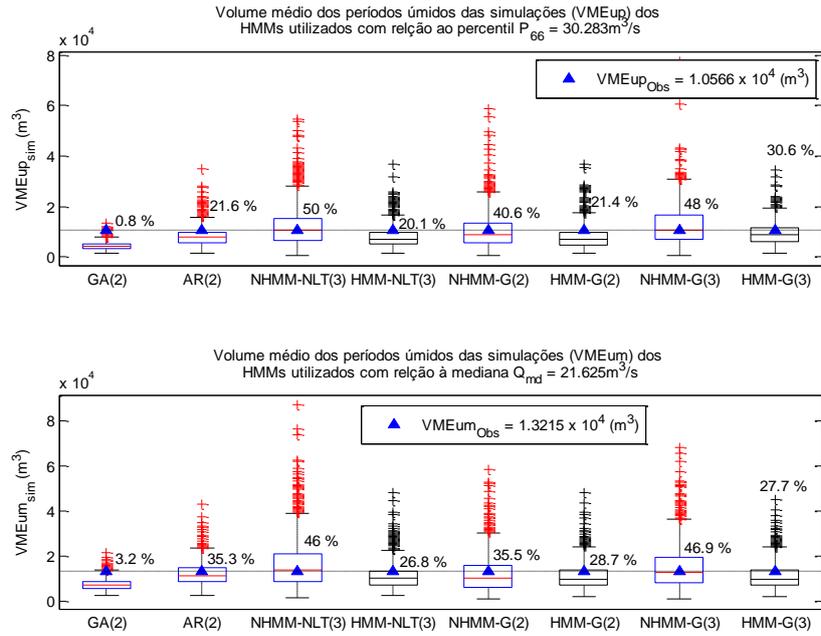


Figura A.9 - Volume médio dos períodos úmidos com relação ao percentil P<sub>33</sub> (superior) e à mediana Q<sub>md</sub> (inferior)