

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

**USO DE MODELOS DE MARKOV COM ESTADOS
OCULTOS E INFORMAÇÕES CLIMÁTICAS PARA A
PREVISÃO DE AFLUÊNCIAS**

LUIS CARLOS HERNÁNDEZ HERNÁNDEZ

ORIENTADOR: DIRCEU SILVEIRA REIS JUNIOR

**TESE DE DOUTORADO EM TECNOLOGIA AMBIENTAL E
RECURSOS HÍDRICOS**

BRASÍLIA/DF: ABRIL 10 – 2019

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

**USO DE MODELOS DE MARKOV COM ESTADOS OCULTOS E
INFORMAÇÕES CLIMÁTICAS PARA A PREVISÃO DE
AFLUÊNCIAS**

LUIS CARLOS HERNÁNDEZ HERNÁNDEZ

**TESE DOUTORADO 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
TÍTULO DE DOUTOR 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)**

**Prof. Wilson dos Santos Fernandes, PhD (UFMG)
(Examinador Externo)**

BRASÍLIA/DF, 10 DE ABRIL DE 2019

FICHA CATALOGRÁFICA

HERNÁNDEZ, HERNÁNDEZ LUIS CARLOS

Uso de Modelos de Markov com Estados Ocultos e Informações Climáticas para a Previsão de Afluências [Distrito Federal] 2019.

xxv, 213p., 210 x 297 mm (ENC/FT/UnB, Doutor, Tecnologia Ambiental e Recursos Hídricos, 2019).

Tese de Doutorado – Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Civil e Ambiental.

1. Previsão de Vazão

2. Índices Climáticos

3. Modelos de Markov

4. Modelos Estocásticos

I. ENC/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

HERNÁNDEZ, L. C. H. (2019). Uso de Modelos de Markov com Estados Ocultos e Informações Climáticas para a Previsão de Afluências. Tese de Doutorado em Tecnologia Ambiental e Recursos Hídricos, Publicação PTARH.TD-023/2019, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF, 213p.

CESSÃO DE DIREITOS

AUTOR: Luis Carlos Hernández Hernández.

TÍTULO: Uso de Modelos de Markov com Estados Ocultos e Informações Climáticas para a Previsão de Afluências

GRAU: Doutor

ANO: 2019

É concedida à Universidade de Brasília permissão para reproduzir cópias desta tese de doutorado 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

Aos que se foram: Elkín, Tía María, Tía Rosalía e Tía Rosaura

*“.....Passado é um pé no chão
E um sabiá
Presente é a porta aberta
E futuro é o que virá mas, e daí
Oh oh, eh eh ah
O moleque acabou de chegar
Oh oh, eh eh ah
Nessa cama é que eu quero sonhar
Oh oh, eh eh ah
Amanhã boto a perna no mundo
Oh oh, eh eh ah
É que o mundo é que é meu lugar
Oh oh, eh eh ah.....[.....]”*

Com a perna no mundo Gonzaguinha

AGRADECIMENTOS

À vida e a Deus por ter me trazido até onde estou e me dar forças para continuar, por ter colocado no meu caminho muitíssimas pessoas que me ajudaram nessa caminhada. Gratidão!!

À minha família pelo seu apoio incondicional, ela é minha força e inspiração para seguir em frente.

À Lulúzinha pela companhia, amor e ensinamentos, também agradeço à sua família pelo apoio e companhia nessa etapa.

Ao meu orientador Prof. Dirceu Reis pela sua ajuda e paciência, seus ensinamentos e questionamentos me inspiraram a realizar esta pesquisa.

Ao Prof. Carlos Lima pelas suas grandes contribuições neste trabalho, e também pela sua ajuda e disposição para atender e solucionar minhas dúvidas.

Ao Prof. Eduardo Savio Martins pelas suas sugestões e contribuições na etapa final deste estudo.

Ao Prof. Wilson dos Santos pelos seus comentários e sugestões na conclusão deste trabalho.

Aos meus colegas de turma do mestrado e outras tantas turmas, corpo docente, técnicos e funcionários do PTARH, uma família que me ajudou bastante desde a época do mestrado. Muito obrigado!!!

A cada pessoa que de alguma ou outra forma me ajudou, me deu forças, esclareceu assuntos, me corrigiu e me ensinou, é impossível citar aqui todas essas pessoas, obrigado a cada uma delas. Namastê!!

À Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME) e ao *International Research institute for Climate And Society* (IRI), pelos dados de afluências e mapas digitais.
Ao *International Research institute for Climate And Society* (IRI), pelos dados de indicadores climáticos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa concedida.

RESUMO

USO DE MODELOS DE MARKOV COM ESTADOS OCULTOS E INFORMAÇÕES CLIMÁTICAS PARA A PREVISÃO DE AFLUÊNCIAS

Sistemas probabilísticos de previsão de vazão nas escalas sazonal e interanual podem fornecer informações valiosas, ainda que incertas, para o processo de alocação de água numa bacia hidrográfica, reservatório ou conjunto de reservatórios, permitindo que se estabeleça um *trade-off* entre os benefícios que se obtém pelo uso da água disponível e os riscos, inerentes à decisão tomada, de não se ter água suficiente para atender as demandas no futuro. Tais sistemas de previsão podem e são utilizados na gestão do risco de secas, permitindo que gestores e tomadores de decisão tenham o tempo necessário para colocar em prática ações previamente planejadas no plano de secas, possibilitando uma redução das perdas econômicas, sociais e ambientais resultantes da ocorrência de uma possível seca. Esta tese explora o uso de Modelos de Markov com Estados Ocultos *Hidden Markov Models* (HMMs), com a inclusão de informações climáticas de larga escala, para realizar a previsão na escala climática de vazões médias anuais afluentes ao reservatório Orós/CE, visando contribuir para o entendimento de como essa classe de modelos se comporta na previsão de vazões. Utilizaram-se três extensões da versão básica dos HMMs, as quais incorporam informações climáticas: i) HMMs não Homogêneos (NHMMs); ii) NHMMs Autoregressivos (AR-NHMMs); e iii) HMMs com informações climáticas nas distribuições dependentes dos estados ocultos (IHMMs), resultando em um total de 132 HMMs. As diferenças entre os HMMs dependem da distribuição dependente (Log-Normal e Gama), do número de estados ocultos (2 e 3) e das 11 combinações de indicadores climáticos (NINO3 e Dipolo do Atlântico utilizados em diferentes momentos e escalas temporais). O desempenho dos diferentes modelos probabilísticos foi avaliado empregando um período de validação de 60 anos (1941-200) tendo como base o *Continuous Ranked Probability Skill Score* (CRPSS), assim como métricas de qualidade de previsões pontuais, tais como o Erro Médio Absoluto e Percentual (EMA, EMPA), Distancia Multicritério (DM) e coeficientes de Nash Sutcliffe (NS) e correlação. HMMs com 2 estados, baseados na distribuição Log-Normal se destacaram dos outros modelos, principalmente aquele que emprega o índice NINO3 de dezembro (NL) do ano anterior para realizar a previsão em janeiro do ano corrente. Em geral, os HMMs obtiveram desempenho melhor em anos mais

secos, quando comparados com a climatologia, mas tiveram dificuldade na previsão de anos mais úmidos. No entanto, IHMMs que empregam o índice do Dipolo do Atlântico são alternativas para a previsão de anos úmidos. Os resultados mostraram que o modelo NL obteve desempenho destacado, quando comparado com modelos que não incluem informações climáticas e modelos autoregressivos com variáveis exógenas (ARX). Os resultados obtidos indicam o potencial de sistemas de previsão de vazão empregando a modelagem baseada em HMMs, em conjunto com informações climáticas. Esse caminho permite uma melhor compreensão da influência climática nos regimes do rio, permitindo flexibilizar o uso de indicadores climáticos e capturar melhor a variabilidade e a ocorrência de vazões baixas.

ABSTRACT

USE OF HIDDEN MARKOV MODELS WITH CLIMATE INFORMATION FOR CLIMATE STREAMFLOW FORECASTS

Seasonal and interannual probabilistic streamflow forecasts, albeit uncertain, can provide valuable information to water allocation processes carried out in watersheds, reservoirs or system of reservoirs, allowing for the evaluation of tradeoffs between the benefits of using the current storage water and the risks of facing a shortage in the future. Such streamflow forecast systems can and are used in drought risk management, giving water managers and decision makers the time needed to take the appropriate measures, usually defined in the drought management plan, resulting in potential reduction in economic, social and environmental damages due to the occurrence of a given drought. This thesis explores the use of Hidden Markov Models (HMMs), in conjunction of large-scale climate information, to provide climate streamflow forecast of mean annual flows for the Orós reservoir/CE. The goal is to evaluate how this class of model can contribute to the understanding of how mean annual streamflow varies in the region and which climate indices modulates the process. Three extension of the basic version of HMMs that employ large-scale climate information were used: i) non-homogeneous HMMs (NHMMs); ii) auto-regressive non-homogeneous models (AR-NHMMs); e iii) HMMs with climate information modulating the dependent probabilistic distributions of HMMs (IHMMs), resulting in a total of 132 models. Differences among models are due to the probability distribution used to model annual flows (Log-normal and Gamma), number of hidden states (2 and 3), and sets of climate indices (Nino3 and Atlantic Dipole taken in different times of the year and different time scales). The quality of the different probabilistic forecast models was evaluated in a 60-year independent period (1941-2000) based on the Continuous Ranked Probability Score (CRPS), as well as on the following deterministic verification measures: mean absolute and relative errors, muticriteria distance, Nash-Sutcliff efficiency coefficient, and correlation. HMMs with two hidden states and based on the Log-normal distribution outperformed other models, in particular the one that employs the value of NINO3 measured in December of the previous year to issue in January the forecast of the current year. In general, the HMMs performed better in dry years, when compared to climatology, but had weak results for wet years. However, IHMMs that uses the Atlantic Dipole index

seems to be a reasonable alternative for wet years. Results show that the NL model obtained a distinguished performance when compared to both models that don't include climate information and auto-regressive models with exogenous variables (ARX). The results point out forecast systems HMMs based along with climate information for could potentially be used for inflow forecasting. Influence of climate on river regimes could be further understood by using this forecast modelling which allows flexibility when using climate indices and a better representation of both variability and lower flow occurrences.

SUMÁRIO

LISTA DE TABELAS.....	XV
LISTA DE FIGURAS	XVI
LISTA DE NOMENCLATURA E ABREVIACÕES	XXI
LISTA DE SÍMBOLOS.....	XXIV
1 - INTRODUÇÃO	1
2 - OBJETIVOS	8
2.1 - OBJETIVO GERAL	8
2.2 - OBJETIVOS ESPECÍFICOS	8
3 - REVISÃO DE LITERATURA.....	9
3.1 - PREVISÃO DE AFLUÊNCIAS E MODELOS ESTOCÁSTICOS DE PREVISÃO DE VAZÃO	9
3.1.1 - Previsão de afluências e seu uso no gerenciamento dos recursos hídricos	9
3.1.2 - Modelos de previsão de vazão	10
3.1.3 - Não estacionariedade em séries temporais de vazões	12
3.1.4 - Persistência hidrológica em séries temporais de vazões	13
3.1.5 - Uso de modelos estocásticos na previsão de vazões	15
3.1.6 - Considerações finais.....	17
3.2 - APLICAÇÕES DOS MODELOS DE MARKOV COM ESTADOS OCULTOS Na PREVISÃO DE VAZÃO	18
3.3 - VARIABILIDADE CLIMÁTICA E O REGIME DE VAZÕES NO ESTADO DO CEARÁ	26
3.3.1 - Características gerais do estado do Ceará.....	26
3.3.2 - Fenômenos climáticos e atmosféricos	27
3.3.3 - Variabilidade temporal e espacial do regime de chuva.....	30
3.3.4 - A influência das oscilações decenais dos oceanos Atlântico e Pacífico sobre o regime de chuvas.....	32
3.3.5 - Variabilidade climática interanual e a sua relação com o regime de vazões.....	35

3.4 - MODELOS UTILIZADOS PARA A PREVISÃO DE VAZÕES NO ESTADO DO CEARÁ	36
4 - MODELOS AUTOREGRESSIVOS COM VARIÁVEIS EXÓGENAS E MODELOS DE MARKOV COM ESTADOS OCULTOS	43
4.1 - MODELOS AUTOREGRESSIVOS COM VARIÁVEIS EXÓGENAS	43
4.1.1 - Função de autocorrelação	43
4.1.2 - Modelos autoregressivos de medias móveis	44
4.1.3 - Modelos autoregressivos com variáveis exógenas	47
4.2 - MODELOS DE MARKOV COM ESTADOS OCULTOS – <i>HIDDEN MARKOV MODELS</i>	48
4.2.1 - Modelos de mistura independente e cadeias de Markov	48
4.2.2 - Modelos de Markov com estados ocultos.....	51
4.2.3 - Estimação de parâmetros	53
4.2.4 - Previsão utilizando Modelos de Markov com estados ocultos	55
4.2.5 - Considerações finais	56
4.3 - EXTENSÕES DA FORMA BÁSICA DOS MODELOS DE MARKOV COM ESTADOS OCULTOS	57
4.3.1 - HMMs para Observações que dependem de covariáveis	57
4.3.2 - Modelos Autoregressivos de Markov com estados ocultos.....	62
4.3.3 - Outras extensões da forma básica de um HMM	63
5 - METODOLOGIA.....	64
5.1 - ESTRATÉGIA METODOLÓGICA.....	64
5.2 - ESTUDO DE CASO: RESERVATÓRIO ORÓS ESTADO DO CEARÁ.....	66
5.2.1 - Dados utilizados.....	68
5.3 - MODELOS ESTOCÁSTICOS PARA A PREVISÃO DE AFLUÊNCIAS UTILIZADOS.....	70
5.3.1 - Definição e estruturação dos modelos de previsão de vazão.....	70

5.3.2 - Critérios de seleção do número de estados ocultos em HMMs e da ordem em modelos AR	75
5.4 - MÉTRICAS DE COMPARAÇÃO DO DESEMPENHO DOS MODELOS DE PREVISÃO DE VAZÃO	76
5.4.1 - Métricas de qualidade de previsões.....	76
5.4.2 - Análise de anos secos e úmidos	80
5.5 - VALIDAÇÃO CRUZADA	80
5.5.1 - Validação cruzada <i>K-fold</i> para HMMs.....	82
6 - RESULTADOS E DISCUSSÃO.....	83
6.1 - CARACTERÍSTICAS DA SÉRIE OBSERVADA.....	83
6.2 - ANÁLISE DE CORRELAÇÃO ENTRE INDICADORES CLIMÁTICOS E A SÉRIE OBSERVADA DE AFLUÊNCIAS	89
6.3 - AVALIAÇÃO DO DESEMPENHO DOS MODELOS DE PREVISÃO DE VAZÃO	92
6.3.1 - Qualidade de previsões probabilísticas para um ano na frente.....	93
6.3.2 - Qualidade de previsões pontuais para um ano na frente	108
6.4 - VALIDAÇÃO CRUZADA	124
7 - CONCLUSÕES.....	127
REFERÊNCIAS BIBLIOGRÁFICAS	132
APÊNDICE A	156
APÊNDICE B.....	174
APÊNDICE C	195
ANEXO A.....	207

LISTA DE TABELAS

Tabela 3.1 – Desempenho dos HMMs na previsão de vazões.	24
Tabela 3.2- Coordenadas as regiões dos índices <i>El Niño</i>	29
Tabela 3.3 – Previsão de vazões no estado do Ceará utilizando informação climática.....	40
Tabela 3.4 – Problemas e recomendações dos trabalhos citados na Tabela 3.4.....	41
Tabela 3.5 – Trabalhos sobre previsão de vazões no Nordeste do Brasil	42
Tabela 3.6 – Problemas e recomendações dos trabalhos citados na Tabela 3.6.....	42
Tabela 5.1- Características do reservatório Orós (Ribeiro, 2011).....	68
Tabela 5.2 – Resumo da estruturação dos modelos de previsão	72
Tabela 6.1 – Estatísticas descritivas da série de vazões observada completa.	84
Tabela 6.2 – Parâmetros e intervalos de confiança de 95% estimados para a distribuição Normal e Gama das vazões observadas, Q, e Log-transformadas, IQ. Período 1911-2000.	88
Tabela 6.3 – Testes de aderência ao 5 % de nível de significância para a série sem transformar, Q, e Log-transformada, IQ. Período 1911-2000.	88
Tabela 6.4 – Correlação tipo Spearman (ρ_S) entre a série observada período 1911-2000, e o os indicadores climáticos NINO3 e Dipolo do Atlântico do ano anterior período 1910-1999.....	92
Tabela 6.5 – “Melhores” modelos de previsão das análises a e b. Todos os HMMs têm 2 estados ocultos.....	95
Tabela 6.6 – Modelo NL e modelos de comparação. Todos os modelos têm 2 estados ocultos.	96
Tabela 6.7 – Resumo do desempenho do modelo NL. Os intervalos de confiança de 95% foram calculados via <i>Bootstrap</i>	124
Tabela 6.8 – Resumo do desempenho do modelo NL. Os intervalos de confiança de 95% foram calculados via <i>Bootstrap</i> . Valores para os anos em que o CRPSS é maior do que o CRPSSm do modelo NL.	124
Tabela 6.9 – Resultados da validação cruzada <i>K-fold</i> para o modelo NL. K=10. Os intervalos de confiança de 95% foram calculados via <i>Bootstrap</i>	125
Tabela 6.10 – Valores do CRPSS para os períodos da validação cruzada <i>K-fold</i> . K = 10.	126

LISTA DE FIGURAS

Figura 3.1- Estrutura de um estudo de simulação do desempenho de um sistema hídrico. .10	
Figura 3.2 – Fases do ENSO: a) condições normais, b) condições <i>El Niño</i> e c) condições <i>La Niña</i> (INPE/CPTEC, 2017).28	
Figura 3.3 – Regiões <i>El Niño</i> (NOAA, 2017a)29	
Figura 3.4 – Localização das zonas que constituem o DIPOLLO do Atlântico (IRI, 2012a) 31	
Figura 3.5 – Efeitos produzidos pelo Dipolo do Atlântico (DA): a) períodos chuvosos e b) secos no Nordeste do Brasil (Ferreira e Mello, 2005)32	
Figura 3.6 – Ocorrência temporal dos modos DA indicados por: 1 se são definidos pela OMA e ATN e -1 se são definidos pela OMA e ATS. As barras cheias indicam PDA (NDA) precedidos por um EN (LN). As barras não preenchidas indicam PDA (NDA) não precedidos por um EN (LN). O índice da OMA (em °C) foi multiplicado por 3 (linha contínua) (Kayano <i>et al.</i> , 2018).35	
Figura 4.1 – Distribuição de mistura de dois componentes. A sequência das componentes ativas é: 1, 2, 1, 1, 2, 1 (esquerda). As densidades ou distribuições ativas (meio) geram as observações (direita) (Zucchini e MacDonald, 2009, traduzido).49	
Figura 4.2 – Representação gráfica de um Modelo de Markov Oculto – HMM.....51	
Figura 4.3 – Geração das observações em um HMM discreto de dois estados ocultos (Zucchini e MacDonald, 2009, traduzido)52	
Figura 4.4 – Representação gráfica de um Modelo de Markov Oculto com covariáveis nas distribuições dependentes dos estados.57	
Figura 4.5 – Representação gráfica de um Modelo de Markov Oculto Não Homogêneo, NHMM.60	
Figura 4.6 – Representação gráfica de um AR(1)–HMM (superior).e de um AR(2)–HMM (inferior)62	
Figura 5.1 – Estratégia Metodológica65	
Figura 5.2 – Localização da bacia do rio Jaguaribe e do reservatório Orós67	
Figura 5.3 – Sistemas de distribuição de água no Estado do Ceará (ANA 2010).69	
Figura 5.4 – Modelos utilizados para a previsão de afluências. Classe I (superior): HMMs com informação climática nas Matrizes de Probabilidade de Transição de estados ocultos – NHMMs e AR-NHMMs; Classe II (meio): HMMs com informação climática nas FDPs	

IHMMS; Classe III (inferior): Modelos Autoregressivos com variáveis exógenas ARX. As setas vermelhas indicam uma autocorrelação de primeira ordem.	73
Figura 5.5 – Fluxograma para o ajuste dos modelos e cálculo de previsões.....	74
Figura 5.6 – Ilustração do <i>Continuous Ranked Probability Score</i> (Wilks 2006).....	77
Figura 5.7 – Estatísticas de períodos secos e úmidos	80
Figura 5.8 – Exemplo de validação cruzada <i>K-fold</i> quando $K=5$	81
Figura 5.9 – Validação cruzada <i>Out - of -Sample</i> (Hyndman, 2019)	82
Figura 6.1 – Série de vazões afluentes ao reservatório Orós, período 1911-2000. O período de validação, 1941-2000, está marcado em cinza. Observe a frequência de vazões “baixas” para vários anos.	83
Figura 6.2 – Histograma das Afluências Observadas	84
Figura 6.3 – Diagrama de caixa das afluências observadas	85
Figura 6.4 – Função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) da série observada. Período 1911-2000.	86
Figura 6.5 – Gráfico quantil-quantil das vazões observadas. Período 1911-2000.	86
Figura 6.6 – Função de probabilidade acumulada Gama das vazões observadas. Período 1911-2000.....	87
Figura 6.7 – Gráfico quantil-quantil das vazões observadas log-transformadas. Período 1911-2000.....	88
Figura 6.8 – Ocorrência de anos secos e úmidos com relação ao percentil 33% e 66%. P_u e P_s representam os anos úmidos e secos, respectivamente.	89
Figura 6.9 – Matriz de Correlação entre a série observada e os indicadores climáticos NINO3 e DA. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Q: vazão observada; ND: NINO3 de dezembro do ano anterior (AA); DAJAS: Dipolo do Atlântico (DA) do período de julho a setembro do AA; DAA: DA de agosto do AA; DAM: média DA do AA.....	91
Figura 6.10 – Resultados das análises a, b e c. As análises seguem os critérios A, B e C....	95
Figura 6.11 – Valor do CRPSSm para o modelo NL, NG, IL e modelos sem informações climáticas.....	97
Figura 6.12 – Diagramas de caixa do CRPSS para o modelo NL, NG, IL e modelos sem informações climáticas. A linha horizontal representa o desempenho da climatologia (CRPSS = 0).....	98
Figura 6.13 – Valor do CRPSSms para o modelo NL, NG, IL e modelos sem informações climáticas.....	99

Figura 6.14 – Diagramas de caixa do CRPSS dos anos secos para o modelo NL, NG, IL e modelos sem informações climáticas. A linha horizontal representa o desempenho da climatologia (CRPSS = 0).....	100
Figura 6.15 – Valor do CRPSS _{mu} para o modelo NL, MG, IL e modelos sem informações climáticas.....	100
Figura 6.16 – Diagramas de caixa do CRPSS dos anos úmidos para o modelo NL, NG, IL e modelos sem informações climáticas. A linha horizontal representa o desempenho da climatologia (CRPSS = 0).....	101
Figura 6.17 – Variação do CRPSS produzido pelo modelo NL ao longo dos 60 anos do período de validação. As linhas azuis horizontais indicam o valor médio do CRPSS na década e a linha vermelha tracejada indica a média móvel de 10 anos. Os retângulos verticais indicam valores do CRPSS selecionados para mostrar o desempenho pontual do modelo NL.....	103
Figura 6.18 – Forma das distribuições de previsão acumuladas para os anos com melhor (a) e pior (b) desempenho. Esses valores são diferentes dos <i>outliers</i> gerados pelo modelo NL.	104
Figura 6.19 – Forma das Distribuições de previsão acumuladas para anos próximos ao valor do CRPSS _m . Anos 1945 (a) e 1992 (b).	105
Figura 6.20 – Superior: valores do CRPSS produzidos pelo modelo NL para os anos secos ao longo dos 60 anos do período de validação. Inferior: valores dos indicadores climáticos ND e da ODP para o período 1940 – 1999	106
Figura 6.21 – Superior: valores do CRPSS produzidos pelo modelo NL para os anos úmidos ao longo dos 60 anos do período de validação. Inferior: valores dos indicadores climáticos ND e da ODP para o período 1940 – 1999.....	107
Figura 6.22 – Valores do Erro Médio Absoluto (EMA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i>	110
Figura 6.23 – Valores do Erro Médio Percentual Absoluto (EMPA) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i>	111
Figura 6.24 – Valores do coeficiente de Nash-Sutcliffe (NS) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e	

NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i>	112
Figura 6.25 – Valores da Distância Multicritério (DM) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i>	112
Figura 6.26 – Valores da Correlação (Equação 5.11) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i>	113
Figura 6.27 – Matriz de correlação tipo Spearman entre as vazões observadas e as médias previstas nos 60 anos do período de validação para pelos modelos Gama, HL e NL. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Qo: vazão observada; Qp: vazão prevista; Gama: Modelo Gama; HL: HMM Log-Normal de 2 estados ocultos; NL: NHMM Log-Normal de 2 estados ocultos que inclui o indicador NINO 3 de dezembro do ano anterior.	114
Figura 6.28 – Erro Médio Absoluto (EMA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i> . Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.	115
Figura 6.29 - Erro Médio Percentual Absoluto (EMPA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i> . Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.	116
Figura 6.30 – Coeficiente de Nash-Sutcliffe (NS) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i> . Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.	117
Figura 6.31 – Distancia Multicritério (DM) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via <i>Bootstrap</i> . Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.	118
Figura 6.32 – Correlação para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os	

intervalos de confiança de 95% calculados via <i>Bootstrap</i> . Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.....	119
Figura 6.33 - Matriz de correlação tipo Spearman entre as vazões observadas e as médias previstas nos 60 anos do período de validação para pelos modelos Gama, HL e NL. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL. Qo: vazão observada; Qp: vazão prevista; Gama: modelo Gama; HL: HMM Log-Normal de 2 estados ocultos; NL: NHMM Log-Normal de 2 estados ocultos que inclui o indicador NINO3 de dezembro do ano anterior.	119
Figura 6.34 – Valores das observações e previsões produzidas pelo modelo Gama (Qp-Gama) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL	121
Figura 6.35 – Valores das observações e previsões produzidas pelo modelo AR(1) (Qp-AR(1)) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.....	122
Figura 6.36 – Valores das observações e previsões produzidas pelo modelo HL (Qp-HL) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.....	122
Figura 6.37 – Valores das observações e previsões produzidas pelo modelo HL (Qp-ARX(1)-ND) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.....	122
Figura 6.38 – Valores das observações e previsões produzidas pelo modelo NL (Qp-NL) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.....	123

LISTA DE NOMENCLATURA E ABREVIACOES

AA	Ano anterior ao ano em que acontece a vazo observada
ACF	Funo de Autocorrelao
AIC	<i>Akaike Information Criterion</i>
AR	Modelo Autoregressivo
AR(1)	Modelo Autoregressivo de Ordem 1
AR(1)-HL	Modelo de Markov Oculito Autoregressivo Log-Normal
AR(1)-NL	NHMM Autoregressivo Log-Normal
ARMA	Modelo Autoregressivo de Mdia Mvel
ARX	Modelo Autoregressivo com Variveis Exgenas
ARX(1)	ARX de Primeira Ordem
ATN	Atlntico Norte Tropical
ATS	Atlntico Sul Tropical
BIC	<i>Bayesian Information Criterion</i>
CCM	Complexos Convectivos de Mesoescala
CRPS	<i>Continuous Ranked Probability Skill</i>
CRPSS	<i>Continuous Ranked Probability Skill</i> <i>Score</i>
CRPSSm	CRPSS mdio para o perodo de validao
CRPSSms	CRPSS mdio dos anos secos do perodo de validao
CRPSSmu	CRPSS mdio dos anos midos do perodo de validao
DA	Dipolo do Atlntico
DAA	DA do ms de agosto do ano anterior
DAJAS	DA do perodo de julho-agosto-setembro do ano anterior
DAM	Media do DA do ano anterior em que

	acontece a vazão observada
DM	Distancia Multicritério
EMA	Erro Médio Absoluto
EMPA	Erro Médio Percentual Absoluto
ENSO	<i>El Niño-Southern Oscillation</i>
FAP	Função de Probabilidade Acumulada
FDP	Função de Densidade de Probabilidade
FF	Frentes Frias
FUNCEME	Fundação Cearense de Meteorologia e Recursos Hídricos
HG	HMM com Distribuições Dependentes Gama
HHMM	HMM Hierárquico
HL	HMM com distribuições Log-Normais e Log-Transformação nos dados
HMM	Modelo de Markov com Estados Ocultos
IG	IHMM com distribuições dependentes Gama
IHMM	HMM com Covariáveis nas distribuições dependentes dos estados ocultos
IL	IHMM com distribuições Log-Normais e Log-Transformação nos dados
IRI	<i>International Research Institute for Climate and Society</i>
KNN	<i>K-Nearest Neighbor</i>
LI	Linhas de Instabilidade
MA	Mesmo ano em que acontece a vazão observada
ME	Algoritmo de Maximização da Expectativa
MPT	Matriz de Probabilidades de Transição
MVS	Máxima Verossimilhança
ND	NINO3 do mês de dezembro do ano

	anterior
NG	Versão Não Homogênea do HG
NHMM	Modelo de Markov com Estados Ocultos Não Homogêneo
NINO3	Índice do ENSO na região Niño 3
NJ	NINO3 do mês de janeiro do mesmo ano
NL	Versão Não Homogênea do HL
NS	Coefficiente de Nash-Sutcliffe
OAN	Oscilação do Atlântico Norte
ODP	Oscilação Decadal do Pacífico
OL	Ondas de Leste
OMA	Oscilação Multi Decadal do Atlântico
PACF	Função de Autocorrelação Parcial
PAR	Modelo Periódico Autoregressivo
PARX	Modelo Periódico Autoregressivo com Variáveis Exógenas
SL	<i>Shifting Mean/Level Model</i>
SOI	<i>Southern Oscillation Index</i>
TSM	Temperatura da Superfície do Mar
VCANs	Vórtices Ciclônicos de Altos Níveis
ZCIT	Zona de Convergência Intertropical

LISTA DE SÍMBOLOS

A		Amplitude
AIQ		Amplitude Inter-Quartis
$AIQP$		Amplitude Inter-Percentis
η		Parâmetro de forma da distribuição Gama
α_t		Probabilidade para frente
θ		Parâmetro de escala da distribuição Gama
β_t		Probabilidade para trás
C_s		Coeficiente de Assimetria Amostral
Γ		Matriz de Probabilidades de Transição
γ_{ij}		Probabilidade de Transição do Estado i ao Estado j
$F_r(r)$		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
$Corr$		Coeficiente de correlação entre previsões e observações
IC		Intervalo de Confiança
Θ		Vetor de Parâmetros de um HMM ou um NHMM
θ_i		Vetor dos parâmetros da distribuição componentes i de um HMM
θ_q		Parâmetro de um Modelo de Média Móvel de Ordem q
k		Defasagem ou <i>Lag</i>
L_N		Verossimilhança em um HMM
$\lambda_i, \sigma_{ij}, \rho_i$		Parâmetros de Transição do Estado Oculto i ao Estado Oculto j em um NHMM
m		Número de estados ou Ordem em um HMM
μ_R		Média ou Valor esperado de R
μ_P		Média esperada ou prevista para $T+h$
μ_i		Média da Distribuição dependente i de um HMM
N		Comprimento da série
N_p		Número de Parâmetros em um HMM
n		Intervalos de Classe
δ_i		Distribuição Estacionaria do estado oculto i em um HMM

P_{33}	Percentil de 33%
P_{66}	Percentil de 66%
P_s	Ano seco segundo o Percentil de 33%
P_u	Ano úmido segundo o Percentil de 66%
P_r	Probabilidade no caso discreto
$P_r(r)$	Função acumulada de probabilidade de r no caso discreto
p	Ordem de um Modelo Autoregressivo
$p_i(r)$	FDP da distribuição dependente do estado i em um HMM no caso discreto
$p_r(r)$	Função massa de probabilidade de r no caso discreto
ρ_k	Autocorrelação na Defasagem k
ρ_s	Coefficiente de Correlação Spearman
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
R	Variável aleatória
R_t, x_t	Variável Aleatória no Tempo t
\bar{R}	Média amostral de R
$R_{1:N}$	Sequência 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
$r_{1:N}$	Sequência de possíveis observações de comprimento N , $\{r_1, r_2, \dots, r_N\}$

r_{pk}	Autocorrelação Parcial Amostral no <i>Lag</i> 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
σ_R^2	Variância de R
σ_i^2	Variância da Distribuição dependente i em um HMM
σ_Q	Desvio Padrão Populacional de Q
t	Tempo
$u_i(t)$	Probabilidade de Marginal do Estado Oculto i no tempo t
w_t	Variável Aleatória Normalmente Distribuída com $\mu_w = 0$ e $\sigma_w^2 = 1$
X_t	<i>Input</i> ou variável exógena em NHMMs, AR-NHMMs e IHMMs
$X_{1:N}$	Sequência 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 <i>White Noise</i> ou Ruído Aleatório
ϕ_p	Parâmetro de um Modelo autoregressivo de Ordem p
Ω	Matriz dos Parâmetros $\lambda_i, \sigma_{ij}, \rho_i$ de tamanho $1 \times m$ em um NHMM

1 - INTRODUÇÃO

A previsão de vazão permite, embora com incerteza, estimar a quantidade de água que afluirá em um reservatório em um horizonte de tempo futuro ou horizonte de previsão. Sistemas probabilísticos baseados em previsões de afluências são de grande valor para os tomadores de decisão no planejamento e gestão dos recursos hídricos, uma vez possibilitam criar estratégias de operação de um reservatório, ou sistemas de reservatórios, baseadas em satisfazer demandas conflitantes entre os usuários da água. Tais estratégias tratam os benefícios pelo uso da água e incluem os riscos associados ao decidir a alocação desse recurso para um período futuro. Ainda, uma operação de reservatórios que utiliza previsões de vazão permite contemplar cenários críticos, por exemplo, a ocorrência de uma seca, e, dessa forma, realizar uma adequada gestão de secas, que vise evitar ao máximo seus impactos socioeconômicos e ambientais associados.

Os modelos de previsão de vazão entregam previsões probabilísticas ou valores pontuais de vazão, as primeiras são bastante flexíveis, já que permitem o cálculo de valores pontuais de vazão, previsão por intervalos e cálculos de quantis. Essas previsões são incluídas em simulações da operação de reservatórios (veja, por exemplo, Loucks *et al.*, 2005), embora, essas simulações tenham a desvantagem de que a própria análise defina as regras de operação a ser utilizadas, e quando o sistema de reservatórios é complexo, o número de alternativas para as regras de operação aumenta, assim, a simulação do sistema torna-se mais complexa (Cardoso *et al.*, 2006). Dessa forma, simultaneamente a técnicas de simulação são empregados métodos de otimização que ajudam a atingir níveis meta de entrega de água, permitindo arcar com vários possíveis cenários de operação.

Basicamente, a previsão de vazões é realizada a partir do conhecimento das condições hidro-climáticas em um instante de tempo, que podem persistir por um período de tempo sazonal ou interanual, e que ajudam para prever a vazão em um tempo posterior a tal instante. A previsão de vazão, ainda, pode aproveitar a predição das condições futuras da atmosfera e oceanos. Existem principalmente duas linhas de modelos de previsão de vazão, uma que utiliza modelos estocásticos e outra que emprega modelos dinâmicos. Essa última estratégia, tem se popularizado nos últimos anos, devido ao avanço do entendimento das interação atmosfera-oceanos e aos avanços computacionais, e utiliza modelos numéricos

do clima global ou regional. Nessa modelagem utilizando, por exemplo, um modelo climático regional, se realizam previsões de precipitação que depois são transformadas em afluições utilizando um modelo conceitual de chuva-vazão (veja, por exemplo, Reis *et al.*, 2009). Essa metodologia de previsão, ainda, pode incluir técnicas estatísticas (*Downscaling*) para um refinamento das previsões de chuva, advindas de modelos climáticos globais ou regionais.

A previsão de vazão de base estocástica emprega modelos estatísticos que procuram encontrar uma relação empírica entre séries históricas de variáveis hidrológicas, por exemplo, afluições e padrões climáticos de larga escala, associados a fenômenos climáticos como o *El Niño – Southern Oscillation* (ENSO) e a temperaturas da superfície do mar em diversos locais dos oceanos Pacífico ou Atlântico (Uvo *et al.*, 1998). Desde a década dos anos 2000, vários estudos têm utilizado esse tipo de previsão no Brasil (veja, por exemplo Lima e Lall, 2010; Souza Filho e Lall, 2003 e Araújo *et al.*, 2015). Tais estudos tentaram desenvolver ou utilizar modelos estatísticos aproveitando as altas correlações entre vazões e indicadores climáticos, relacionados ao ENSO ou tele conexões climáticas, nas escalas intra e interanual, encontradas no oceano Atlântico Tropical, que modificam o sistema de altas pressões na zona equatorial influenciando a magnitude das chuvas, por exemplo, no Nordeste do Brasil (Hastenrath e Heller, 1977, Uvo *et al.*, 1998; Kayano *et al.*, 2018) .

Uma das regiões com maior influência de tele conexões climáticas, presentes nos oceanos Pacífico e Atlântico, no mundo é o semiárido brasileiro, localizado no Nordeste do Brasil (NEB). Essa região, que possui alta densidade populacional, tem sido afetada historicamente com severos períodos de seca impactando a produção agrícola, a produção alimentícia, e contribuindo para o êxodo massivo da população desde o século XVI (Pilz *et al.*, 2019). Por exemplo, apenas em 2017, ao redor de 30 milhões de pessoas foram afetadas por eventos de seca (ANA, 2018). Vários autores (veja, por exemplo, Hastenrath e Heller, 1977; Moura e Shukla, 1981; Uvo *et al.*, 1998; Marengo e Valverde, 2007; Nys *et al.*, 2016; Kayano *et al.*, 2018) têm demonstrado que a ocorrência de tais períodos de seca está relacionada ao ENSO e a o gradiente de temperaturas no Atlântico Tropical Dipolo do Atlântico (DA) , esse último, modifica o deslocamento Norte-Sul da Zona de convergência intertropical (ZCIT), principal modulador da quantidade de chuvas do NEB (Cabral, 2014).

Ainda no que concerne ao NEB, um estado que se insere nessa região e que sofre os efeitos das secas é o Estado do Ceará. Nesse estado, a alta variação espaço temporal do regime de chuvas está relacionada à alta variabilidade climática intranual, interanual e decenal, associada às tele conexões climáticas que afetam o NEB. Intranualmente, a média pluviométrica se concentra no primeiro semestre do ano janeiro a maio, com maior concentração entre fevereiro a maio e existe ausência de chuva no segundo. Na escala interanual, a média histórica pluviométrica é de 805 mm, com tendência a anos com média inferior, o que produz os conhecidos períodos de secas. Por outro lado, a variação espacial da precipitação nesse estado é influenciada por fenômenos atmosféricos locais que modificam as chuvas nas zonas norte, sul-sudeste e leste (Marengo e Valverde, 2007; Campos, 2014).

Em face disso, no Ceará, existe uma ausência de escoamento superficial na segunda metade do ano, o que contribui para rios com vazões com coeficientes de variação acima de 1 (um), que é relativamente alto em relação a outros rios no mundo (Reis *et al.*, 2009). Em virtude disso, uma estratégia para garantir a segurança hídrica em períodos de seca tomada pelo governo foi a construção, ao longo do tempo, de um sistema de reservatórios para atender aos períodos de secas e dessa forma armazenar água para satisfazer os setores: agrícola, industrial e humano. A estratégia engloba ainda reservatórios funcionando em períodos de cheias para a regularização de vazões. No entanto, pelas características variáveis dos rios existe a alta incerteza nas afluições produzindo problemas na operação desses reservatórios. Dessa maneira, o Ceará é uma região altamente vulnerável à variabilidade climática futura e às possíveis mudanças climáticas, já que ela pode produzir impactos severos nas suas reservas de água (Souza Filho e Lall 2003; Araújo *et al.*, 2015; Marengo *et al.*, 2016). Por exemplo, mesmo com períodos chuvas ao redor da média histórica em 2017 e 2018, os maiores reservatórios do estado apresentaram um volume útil de menos do 10%, resultado dos efeitos da seca que começou desde 2012 (Pilz *et al.*, 2019).

Nesse quadro, a alocação da água para os diferentes setores é um desafio imposto aos gestores de recursos hídricos e exige lidar continuamente com conflitos sobre as demandas de água a ser entregues de um ano para outro, normalmente solucionados por medidas conservadoras. Por exemplo, a companhia de gerenciamento de recursos hídricos do Estado do Ceará (COGERH) assume vazões iguais a zero no período seco do próximo ano,

enquanto, em períodos úmidos a incerteza da ocorrência ou não de vazões é alta (Ribeiro, 2011; Silveira, 2014; Araújo *et al.*, 2015). Dessa forma, fornecer previsões, com antecedência à quadra chuvosa na região que representem a variabilidade dos rios da região e persistência de longo prazo das suas vazões, é de grande valor para essa agência ou para os tomadores de decisão na região. Nessa linha, estudos de previsão que abordem essas características das vazões, principalmente, aqueles que tratem a variabilidade climática, são valiosos para o estado. Alguns estudos desse tipo realizados no estado tem demonstrado a aplicabilidade de vários modelos estocásticos e dinâmicos (veja, por exemplo, Souza Filho e Lall 2003; Reis *et al.*, 2009; Araújo *et al.*, 2015), porém, ainda são necessários estudos que consigam utilizar as informações climáticas disponíveis, por exemplo, indicadores climáticos, para capturar melhor as características históricas das vazões, variabilidade e persistência, para com isso, obter previsões mais acuradas e, logo, diminuir a incerteza na tomada de decisão na operação de reservatórios.

Comumente utilizados para fins de previsão de vazão, a família de modelos Autoregressivos com Médias Móveis (ARMA) apresenta dificuldades para representar a persistência de longo prazo, variabilidade ao longo do tempo e não estacionaridade no espectro (Bracken *et al.*, 2014). Esses modelos têm deficiências para representar longos períodos de vazões mais secas ou úmidas e a sua variação ao longo do tempo, essa deficiência pode produzir previsões errôneas e, portanto, influenciar para a tomada de decisões incorretas na operação de um sistema de reservatórios. Dessa forma, modelos que tratem principalmente a persistência e não estacionariedade podem ajudar a reduzir a incerteza nas previsões de vazão (Chen *et al.*, 2016, 2018; Liu *et al.*, 2018).

Um exemplo de modelos que permitem tratar a persistência de longo prazo são os modelos multifractais (veja, por exemplo, Lohre *et al.*, 2003 e Li *et al.*, 2015). Porém, esses modelos não apresentam flexibilidade para incluir variáveis externas, como informações climáticas na sua estrutura, limitando o seu uso (Bracken *et al.*, 2014). Embora, em alguns casos pela sua estrutura de autocorrelação, correlação entre membros da mesma série, os modelos ARMA representem a persistência; no caso da variabilidade, devido a essa estrutura linear nesses modelos não é possível criar uma variação entre regimes secos ou úmidos, dificultando a representação de saltos ou mudanças abruptas de vazões encontradas em séries históricas, especialmente de regiões com forte influência da variabilidade de climática, relacionada, por exemplo, a eventos como o ENSO.

Nesse sentido, um tipo de modelos estocásticos, denominados Modelos de Markov com estados Ocultos *Hidden Markov Models* (HMMs), tem sido utilizado na literatura para representar padrões de persistência e variabilidade ao longo do tempo, além de características de não estacionariedade, encontrados em séries de vazões (veja por exemplo, Jackson, 1975; Thyer e Kuczera, 2000; Akintug e Rasmussen, 2005; Whiting, 2006; Bracken *et al.*, 2014, 2016). Esses modelos têm tido bastante sucesso na representação desses padrões e, ainda, têm conseguido representar tais padrões, quando eles foram associados a fenômenos climáticos de larga escala.

Com relação à sua estrutura, um HMM é composto por um número de estados não observados. No tempo, o sistema varia aleatoriamente entre os estados ocultos que geram as observações, a partir de distribuições de probabilidade dependentes desses estados. A chance de mudar de estado dado ao atual estado do sistema é determinada pela denominada Matriz de Probabilidade de Transição (MPT). Além disso, nos HMMs variáveis externas, podem ser incluídas para que os parâmetros do modelo, por exemplo nas MPT, mudem ao longo do tempo, podendo tal variação ser relacionada a variações interanuais de padrões climáticos. Embora, essa inclusão de variáveis seja possível em modelos ARMA com variáveis exógenas (ARMAX), os ARMAX também são modelos lineares que não permitem representar mudanças ou saltos abruptos na média, o que sim é possível nos HMMs, pela presença de um mecanismo interno, uma cadeia de Markov que varia no tempo entre estados discretos, os denominados estados ocultos. Esses estados ocultos podem representar estados hidro-climáticos, e ainda, podem ser relacionados a fenômenos climáticos que influenciem a sua ocorrência, embora esses estados não sejam medíveis, daí o nome de ocultos. Essa característica diferencia os HMMs dos modelos multifractais, já que, os HMMs apresentam maior flexibilidade no uso de informações climáticas, permitindo avaliar a influência dessas informações sobre os estados ocultos. Dessa forma, os HMMs são potenciais candidatos para a previsão de vazões para a região semiárida do NEB.

Portanto, neste estudo utilizaram-se, para a previsão de vazão, extensões dos HMMs que permitem a inclusão de informações climáticas nas MPT (*Non-Homogenous HMMs* NHMMs) e nas distribuições dependentes (IHMMs), também se empregaram NHMMs que incluem modelos autoregressivos como distribuições dependentes (AR-NHMMs). Embora,

os AR-NHMMs apresentem uma estrutura mais complexa com relação aos NHMMs, esses modelos apresentam igual número de parâmetros que os NHMMs. Estudos de previsão de vazão utilizando HMMs são poucos: Fortin *et al.*, 2004; Bracken, 2011 e Liu *et al.*, (2018) e o único uso de NHMMs para fins de previsão foi realizado por Gelati *et al.*, (2010b), enquanto, AR-NHMMs não têm sido explorados na literatura. No Brasil nenhum desses modelos para a previsão de aflúências têm sido utilizados. Os IHMMs exploram a inclusão de variáveis externas nas distribuições de probabilidade de ocorrência dependentes dos estados ocultos e o seu uso, corresponde à tentativa de capturar melhor a variabilidade das observações, incorporando não estacionariedade nas probabilidades de ocorrência, e, dessa forma, obter mais acurácia nas previsões de aflúências. A modelagem de previsão utilizando IHMMs propostos neste estudo não é encontrada na literatura.

Logo, em conjunto, o uso de modelos HMMs para a previsão de aflúências, procura contribuir na aplicação de uma forma de modelagem de previsão, não encontrada na literatura no caso dos IHMMs e não encontrada no Brasil nesse caso os NHMMs e AR-NHMMs, visando entender melhor como funcionam esse tipo de modelos para a previsão de vazão em uma região na qual a variabilidade climática influencia o regime dos rios, com o atrativo de essas extensões dos HMMs flexibilizam o uso das informações climáticas disponíveis, no caso indicadores climáticos. Procura-se, ainda, ao utilizar esses modelos representar-se melhor a distribuição de frequências, variações e persistência observadas em séries históricas de vazões de tal região. Pretende-se, também obter ganhos dessa representatividade com relação a modelos que não incluem informações climáticas ou aqueles que não representam adequadamente a variação de estados hidro-climáticos.

Tomando todas as considerações mencionadas, esta tese explora a previsão de aflúências anuais, com antecedência mensal, utilizando Modelos Markov com Estados Ocultos *Hidden Markov Models* (HMMs), para a previsão de aflúências médias anuais ao reservatório Orós, utilizando uma série histórica de aflúências desse reservatório (90 anos). O Orós pereniza o rio Jaguaribe e é localizado no Estado do Ceará, Nordeste do Brasil. Esse reservatório é a segunda maior reserva hídrica do Ceará e faz parte do sistema de abastecimento que abastece a região metropolitana de Fortaleza, com 55% da população urbana do estado, além disso, abastece as bacias do médio e baixo Jaguaribe.

Considerando-se tais colocações, espera-se, ao utilizar HMMs, capturar as frequências, a variabilidade e persistência observadas em séries históricas de vazão, para obter previsões de afluições mais acuradas em uma região com forte influência da variabilidade climática Estado do Ceará, Nordeste do Brasil. Principalmente a intenção é obter algum ganho na previsão de períodos mais úmidos, mas, também melhorar a previsão de períodos secos, e, dessa forma, auxiliar a gestão dos reservatórios dessa região. Para determinar a acurácia das previsões obtidas por HMMs, compara-se sua qualidade frente à qualidade das afluições previstas por outros modelos estocásticos de previsão utilizados na região, entre eles os modelos Autoregressivos com variáveis exógenas (ARX), bastante utilizados no NEB para a previsão de vazão.

Este documento prossegue com os objetivos da tese no Capítulo 2. No Capítulo 3 é apresentada a revisão de literatura sobre: a previsão de vazões, o uso dos HMMs para fins de previsão de vazão, a variabilidade climática e a sua relação com o regime de vazões no estado do Ceará, e o desempenho dos modelos utilizados nesse estado, para a previsão de afluições. No Capítulo 4 é apresentado um embasamento teórico dos modelos ARMA, ARMAX e dos HMMs incluindo as suas variações NHMMs, AR-NHMMS e IHMMs. Seguidamente, no Capítulo 5 é apresentada a metodologia a ser empregada, que inclui os dados a serem utilizados e as metodologias de avaliação do desempenho dos modelos de previsão. O capítulo 6 traz os resultados deste estudo. Por fim, o Capítulo 7 apresenta as conclusões da pesquisa.

2 - OBJETIVOS

2.1 - OBJETIVO GERAL

Explorar o uso dos Modelos de Markov com Estados Ocultos (HMMs) que utilizam informações climáticas, para a previsão, na escala anual, de vazões médias afluentes ao reservatório Orós, Estado do Ceará, visando contribuir no entendimento de como essa modelagem se comporta na previsão de afluências em uma região semiárida do Nordeste do Brasil (NEB).

2.2 - OBJETIVOS ESPECÍFICOS

- Identificar as informações climáticas, na escala interanual, relacionadas com a temperatura da superfície do mar nos oceanos Pacífico e Atlântico equatorial, que exercem maior influência sobre a série histórica de afluências;
- Avaliar a qualidade das previsões de afluências obtidas através de Modelos de Markov com Estados Ocultos (HMMs), baseados nas distribuições Gama e Log-Normal e que utilizam informações climáticas para influenciar essas distribuições e a transição de estados ocultos.
- Comparar a qualidade das previsões de afluências produzidas por Modelos de Markov com Estados Ocultos (HMMs) e que incorporam informações climáticas com a qualidade de previsões de afluências geradas por Modelos Autoregressivos com Variáveis Exógenas (ARX), que utilizam informações climáticas como variáveis externas;
- Avaliar a capacidade de previsão dos HMMs que incluem informações climáticas e os ARX para prever anos mais secos e mais úmidos definidos por limiares históricos da série de afluências.

3 - REVISÃO DE LITERATURA

3.1 - PREVISÃO DE AFLUÊNCIAS E MODELOS ESTOCÁSTICOS DE PREVISÃO DE VAZÃO

3.1.1 - Previsão de afluições e seu uso no gerenciamento dos recursos hídricos

A previsão de afluições consiste na estimativa da vazão de entrada a um reservatório com determinada antecedência temporal (Fan e Meller, 2015). Também pode ser definida como uma estimativa de vazões futuras ou a geração de vazões para um período na frente – horizonte de previsão –, que depende do problema (Salas *et al.*, 2014; Box *et al.*, 2016). Essas estimativas são valores que não pertencem à amostra que se utiliza para ajustar o modelo de previsão (Stock e Watson, 2010).

De acordo com a antecedência, as previsões são de curto ou longo prazo. A previsão de longo prazo envolve a previsão com antecedência de meses ou anos, sendo muitas vezes chamada de sazonal (Fan e Meller, 2015). As previsões de vazão ao ser empregadas no gerenciamento dos recursos hídricos ajudam a melhorar a eficiência na operação de sistemas de abastecimento de água e hidro energia, e entregar informações uteis para sistemas de alerta de cheias, alocação de água para irrigação, e o acompanhamento da dinâmica de secas em curso. Quanto maior for o grau de acurácia das previsões de vazão, maior será o valor das mesmas para auxiliar a tomada de decisão (Salas *et al.*, 2014; Chen *et al.*, 2016, 2018).

Nesse contexto, Loucks *et al.*, (2005) expõem a estrutura de um estudo de simulação da operação de um reservatório, ou de um hidrossistema, figura 3.1. Nessa estrutura, transformam-se previsões de vazão, demandas futuras e as políticas operacionais, dentro de um modelo de simulação para produzir estatísticas do desempenho do hidrossistema. Esse tipo de simulação pode se tornar complexo dependendo dos usos conflitantes da água, do número de reservatórios e das políticas operacionais. As previsões de vazão podem ajudar na redução dos conflitos, porém elas estão condicionadas à eficiência dos modelos de previsão (Silveira, 2014).

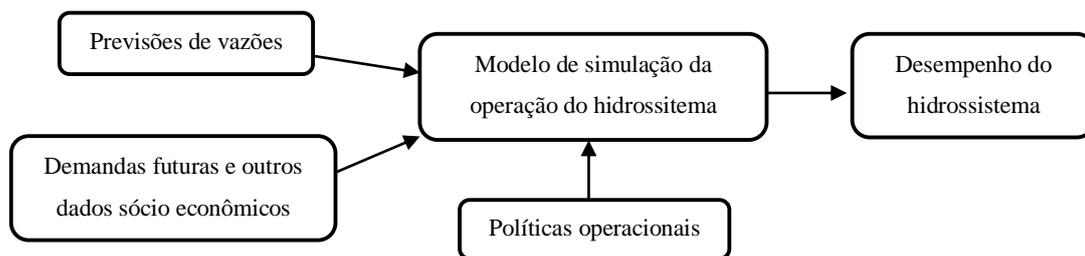


Figura 3.1- Estrutura de um estudo de simulação do desempenho de um sistema hídrico.

Um exemplo do uso de previsões de aflúências na simulação e otimização de um sistema de reservatórios no estado do Ceará é mostrado por Cardoso *et al.*, (2006). Esses autores utilizaram o método de otimização simplex Nelder-Mead, o qual entrega as alocações ótimas. O método resolve um conjunto de equações baseadas em um balanço hídrico dos reservatórios que incluem as aflúências previstas. Silveira (2014) menciona que esse tipo de estudos de simulação é empregado para o processo de alocação por parte da Companhia de Gestão de Recursos Hídricos do Ceará (COGERH), para um horizonte de planejamento de 6 a 18 meses. Nesse processo, a COGERH assume que na estação seca a vazão afluyente é zero, enquanto na estação úmida existe alta incerteza da magnitude das aflúências.

3.1.2 - Modelos de previsão de vazão

A previsão de vazões pode ser realizada utilizando modelos de base física ou probabilística. A primeira visa reproduzir os diferentes processos do ciclo hidrológico empregando conjuntamente a modelagem atmosférica e hidrológica. Os resultados desse tipo de modelos são uma representação física dos processos hidro-climáticos de uma dada bacia hidrográfica (Silveira, 2004). A segunda, se realiza a partir de modelos estocásticos baseados nas propriedades históricas da série. Tais modelos consistem em modelos matemáticos, compostos ou não de parâmetros, que tentam inferir propriedades do processo estocástico – que representa a população, a partir de uma única amostra – no caso uma série temporal. O denominado processo estocástico é um fenômeno estatístico que evolui no tempo seguindo leis probabilísticas. Portanto, uma série temporal pode ser considerada como uma realização particular desse processo, produzida por um mecanismo de probabilidades características do sistema analisado (Box *et al.*, 2016).

Em virtude disso, as diferentes classes de modelos estocásticos de previsão de vazão inferem a distribuição de probabilidade de uma observação futura, dado um registro histórico, empregando distintas formas de descrever a população, amostras, valores máximos e mínimos e valores atípicos ou *outliers* (Box *et al.*, 2016). Na literatura são citadas diferentes classes de modelos estocásticos para a previsão de vazão. No entanto devido à natureza complexa dos processos hidrológicos, à dificuldade para o tratamento de dados faltantes, ao fenômeno de Hurst (Hurst, 1957), à suposição de estacionariedade e à incerteza do modelo e dos parâmetros, há dificuldades em utilizar uma única classe de modelo generalizada, ou seja, aplicável a qualquer situação ou local (Sveinsson e Salas, 2003; Ilich e Despotovic, 2008).

A incerteza nos parâmetros e do modelo influenciam a acurácia de previsões de vazão, e deve ser considerada como parte essencial no processo de previsão, ou pelo menos descrita, já que, afetará o processo de alocação de água (Kwon *et al.*, 2011; Chen *et al.*, 2018). A incerteza do modelo está relacionada à escolha do melhor modelo que se adapte às necessidades do problema de previsão estudado. Dada a quantidade de modelos disponíveis, relativamente grande, junto com um número limitado de variáveis disponíveis, a incerteza dos modelos pode se tornar um sério problema (Montgomery e Nyhan, 2010).

Tradicionalmente, modelos estocásticos são selecionados utilizando critérios que penalizam modelos complexos, que geralmente possuem muitos parâmetros, como o Critério de Informação de Akaike (AIC) (Akaike, 1974) e o Critério de Informação Bayesiano (BIC) (Schwarz, 1978), outra alternativa é a ponderação de modelos, por exemplo, técnicas de ponderação Bayesiana de modelos (veja, por exemplo, Pokhrel *et al.*, 2013; Oliveira e Lima, 2016; Meira Neto *et al.*, 2018).

No caso do tratamento da incerteza dos parâmetros uma abordagem utilizada, por exemplo, em modelos autoregressivos de médias móveis (ARMA), é a análise assintótica das distribuições dos parâmetros, veja, por exemplo, Lee *et al.*, (2007) e Box *et al.*, 2016. Outras técnicas utilizadas são baseadas em análise Bayesiana como pode ser visto para modelos ARMA, os trabalhos de Valdés *et al.*, (1977); Stedinger e Taylor, (1982). Cabe aqui destacar que, muito embora influencie na previsão de afluências, a incerteza do

modelo e dos parâmetros, acima descrito, não foi objeto de tratamento nesta tese, portanto a sua análise não será realizada dado o caráter exploratório desta tese.

3.1.3 - Não estacionariedade em séries temporais de vazões

A maioria de modelos estocásticos de previsão de afluências é denominada de modelos estacionários. Esses modelos assumem que as séries temporais permanecem em “equilíbrio” estatístico. Portanto, a distribuição de probabilidade que descreve o processo não muda ao longo do tempo – possui média fixa e variância constante (Box *et al.*, 2016).

Atualmente existe um debate entre se a estacionariedade ainda é válida ou não no gerenciamento de recursos hídricos. Por exemplo, Milly *et al.*, (2008) mencionam que a estacionariedade não é mais válida. Para esses autores, uma das razões dessa invalidez é que as mudanças antropogênicas do clima alteram as médias e extremos de precipitação, taxas de evapotranspiração e o regime das vazões em rios. De outro lado existem vários autores que criticam que a estacionariedade seja inválida, por exemplo, Matalas, (2012); Lins e Cohn, (2011).

Lins e Cohn (2011) mencionam que a não estacionariedade existe e ela é uma característica natural das variáveis hidrológicas. Com relação ao clima, citam que não é claro como ele influencia a hidrologia. Portanto é impossível remover completamente a não estacionariedade através da modelagem determinística. Também detalham que a magnitude da não estacionariedade não é clara e pode ser influenciada por outras fontes de variabilidade que ainda se desconhecem.

Milly *et al.*, (2008) mencionam que existe a necessidade de se encontrar formas de identificar modelos probabilísticos que considerem a não-estacionariedade e utilizá-los para otimizar sistemas de recursos hídricos. Portanto, requer-se um bom uso e entendimento das ferramentas estocásticas e tecnológicas disponíveis na atualidade, para desenvolver modelos mais sofisticados que permitam entender melhor a variabilidade hidro-climática (Salas *et al.*, 2014).

Também Nayak e Villarini (2016) mencionam que o tratamento da não estacionariedade é um desafio para os hidrólogos, porque não considerá-la pode resultar em efeitos no

desenho e estruturas hidráulicas, incerteza no gerenciamento e operação de reservatórios e em estudos sobre impactos de mudanças climáticas nos recursos hídricos (Maftai *et al.*, 2016).

Outra possível causa da não estacionariedade tem sido citada na literatura nos últimos anos, já que, vários estudos mostram que as variações em séries de vazões estão relacionadas com a variabilidade climática, principalmente com fenômenos como o *El Niño – Southern Oscillation* (ENSO). Essa relação é ainda maior com fenômenos de larga escala como a Oscilação Decenal do Pacífico (ODP) e a Oscilação Multi-decenal do Atlântico (OMA), as quais produzem oscilações nas magnitudes e valores médios de vazões e na frequência de secas e cheias, (veja por exemplo, Fortin *et al.*, 2004; Sveinsson *et al.*, 2003; Robertson *et al.*, 2004; Souza Filho e Lall, 2003, Erkyihun *et al.*, 2016; Kayano *et al.*, 2018).

Nesse debate, Lins e Cohn (2011) concluem que pode ser aceito um planejamento ignorando a não estacionariedade e gerenciar os recursos hídricos sob princípios de risco e adaptação. Também mencionam que na ausência de alternativas, deve-se manter a estacionariedade válida, para garantir que o “sucesso” do gerenciamento tradicional em sistemas hídricos não perca vigor. O leitor pode encontrar mais sobre o assunto em Matalas (2012); Milly *et al.*, (2015) e Zhang *et al.*, (2019).

3.1.4 - Persistência hidrológica em séries temporais de vazões

Outra questão de interesse na previsão de vazões é a persistência hidrológica. Essa persistência representa a memória em séries temporais e se refere à dependência dos dados uns sobre os outros, ou seja, a tendência que tem valores sucessivos de lembrar seus antecedentes e de ser influenciados por eles (Klemeš *et al.*, 1981; Machiwal e Jha, 2012). Uma das razões da persistência em séries temporais de vazão é a influência de fenômenos climáticos de longa escala, por exemplo, o ENSO (Nigam *et al.*, 2011). Outra causa da persistência de vazões é o armazenamento subsuperficial na bacia (Bayazit e Bulu, 1988; Whiting, 2006).

A persistência em uma série anual de vazões é observada como longos períodos de anos seguidos com vazões maiores ou menores a um limiar. Alguns limiares comumente empregados para determinar esses períodos são: a média, a mediana e os percentis da série

observada (Sveinsson *et al.*, 2003). Esses períodos podem ser denominados como “secos” ou “úmidos”, dependendo do caso, por exemplo, anos com vazões menores que a média serão períodos secos, e maiores são definidos como úmidos. A análise dos comprimentos desses períodos e da sua variabilidade é de bastante interesse no gerenciamento e planejamento de recursos hídricos.

A investigação do comportamento de períodos secos prolongados permite identificar suas propriedades estatísticas, realizar previsões probabilísticas com antecedência à sua ocorrência, quantificar sua duração, severidade, e a sua área de influência (Saldarriaga e Yevjevich, 1970). Também é possível analisar a influência de secas sobre outras regiões e evitar prejuízos econômicos ambientais e sociais. Por exemplo, se duas ou mais regiões produzem uma cultura importante, ou abastecem complexos urbanos/industriais, então a probabilidade condicional de uma seca atingir simultaneamente essas duas regiões pode ter implicações no abastecimento de água ou na produção agrícola (Saldarriaga e Yevjevich, 1970; Marengo *et al.*, 2016).

Saldarriaga e Yevjevich, (1970) mencionam que os períodos úmidos trazem consequências parecidas às secas. Porém, períodos chuvosos que não ultrapassam os picos de cheias e de curta duração podem ser importantes para a restauração da cobertura vegetal em regiões áridas e semiáridas porque permitem restaurar as vazões naturais. Também podem ajudar à restauração de ecossistemas aquáticos e evitar a poluição produzida durante períodos de secas, por exemplo a poluição do ar devido à poeira produzida pelo ressecamento dos solos (Saldarriaga e Yevjevich 1970; Propst *et al.*, 2008; Feng *et al.*, 2011).

3.1.4.1 - Alguns estudos sobre persistência hidrológica em séries de vazões

Tradicionalmente os modelos ARMA tem sido utilizados em estudos de persistência, porém, esses modelos apresentam em alguns casos problemas para capturar persistência no longo prazo e não estacionariedade (Bracken *et al.*, 2014). Eles pertencem a classe de modelos de memória curta e apresentam dificuldades para simular valores acima de 0.7 do coeficiente de Hurst. Este coeficiente é um indicador de persistência relacionado ao fenômeno Hurst, o qual é encontrado em séries temporais geofísicas, incluindo séries de vazões (O’Connell *et al.*, 2016).

Hurst (1951,1957) definiu a persistência em termos de um parâmetro denominado coeficiente de Hurst (H), com um valor médio observado de 0,73 para séries de grande comprimento. Porém, teoricamente esse coeficiente para séries de vazões pode apresentar valor de 0,5. Quando os valores observados e teóricos do coeficiente de Hurst não correspondem, se apresenta o que é denominado como fenômeno Hurst e corresponde a uma persistência de longo prazo.

A maioria dos modelos estocásticos propostos simulam a persistência de curto prazo – relacionada com o decaimento rápido da Função de Autocorrelação (ACF), função que mede o grau de dependência temporal entre valores da mesma série. Entanto, a persistência de longo prazo – associada ao decaimento lento da ACF – é virtualmente difícil de identificar devido ao tamanho, curto, das séries de vazões comumente disponíveis (Box *et al.*, 2016; Whiting, 2006; Koscielny-Bunde *et al.*, 2006).

Dessa forma, outros modelos, diferentes aos ARMA, são propostos na literatura para tratar a persistência, além de representar não estacionariedade e mudanças abruptas ou saltos na média. São exemplos: os modelos de memória curta – de médias móveis e de regime variável (veja, por exemplo, Salas e Boes, 1980; Sveinsson *et al.*, 2003 e Fortin *et al.*, 2004) e os modelos de memória longa como os *Fractional* ARMA (FARMA) e *Fractional Gaussian Noise Models* (veja, por exemplo, Koutsoyiannis, 2002).

Destacam-se alguns estudos de persistência de vazões que analisaram as propriedades estatísticas de comprimentos de períodos secos e úmidos utilizando modelos ARMA, de memória curta e outros modelos: Saldarriaga e Yevjevich (1970); Bayazit e Bulu (1988); Sveinsson *et al.*, (2003); Whiting (2006); Prairie *et al.*, (2008); Bracken (2011), Hernandez (2013); Bracken *et al.*, (2014); Turner e Galelli, (2016); Guo *et al.*, (2018). Vários desses trabalhos utilizaram Modelos de Markov simples ou Modelos de Markov com Estados Ocultos (HMMs).

3.1.5 - Uso de modelos estocásticos na previsão de vazões

Alguns modelos estocásticos para a previsão de vazões surgiram desde a década de 1910. Nas décadas de 1950 e 1960 Hurst (1951), Thomas e Fiering (1962), Matalas (1967), e

Mandelbrot e Wallis (1969) propuseram vários tipos de modelos que deram origem a outros tantos modelos em diversas diretrizes.

Na década de 1970, Box e Jenkins criaram uma classe de modelos que é considerada clássica, a família de modelos Autoregressivos de Médias Móveis –ARMA. A partir desses modelos se criaram modelos mais sofisticados e até hoje são desenvolvidos novos tipos de modelos. Salas *et al.*, (1980) e Box *et al.*, (2016) compilam alguns desses trabalhos.

Entre os modelos estocásticos comumente utilizados para previsão, destacam-se os modelos de regressão, de regressão baseados em componentes principais, *Autoregressive Integrated Moving Average* (ARIMA), ARMA com variáveis externas (ARMAX) e *Transfer Function Noise* (TFN). Os modelos ARMAX e autoregressivos com variáveis externas (ARX) vêm sendo utilizados desde a década de 1980. Por exemplo, Hannan *et al.*, (1980) começaram o seu emprego na previsão hidrológica.

Além desses modelos, desde a década de 1990, são utilizados modelos de redes neuronais artificiais – *Artificial Neural Networks* (ANN) (Govindaraju e Ramachandra, 2000). Também desde essa década, surgiram abordagens baseadas em preditores hidrológicos, oceânicos e atmosféricos. Por exemplo, têm sido empregados indicadores das temperaturas da superfície do mar (TSMs), indicadores do ENSO, ODP, OMA e da Oscilação do Atlântico Norte (OAN), devido à influência desses fenômenos sobre variáveis atmosféricas como pressão atmosférica, direção do vento, entre outras, e hidrológicas como precipitação e vazão (Uvo *et al.*, 1998).

Vários trabalhos têm demonstrado uma melhoria na qualidade de previsões quando se incluem esses indicadores climáticos, empregando tanto abordagens paramétricas quanto não paramétricas (veja Souza Filho e Lall, 2003; Córdoba-Machado *et al.*, 2016). Por exemplo, no caso dos modelos ARMAX, Kwon (2002) realizou previsões de vazões diárias utilizando modelos ARMAX na China e Gelati *et al.*, (2010b) utilizaram modelos ARX, que variam regulados por uma cadeia de Markov com estados ocultos, para previsões de vazões mensais que sofrem a influência do ENSO no Equador.

No Brasil, Lima and Lall, (2010a), (2010b) previram vazões afluentes a aproveitamentos hidroelétricos do Sistema Interligado Nacional (SIN) no Brasil. Esses autores obtiveram

melhores previsões mensais, incluindo um horizonte de 20 meses com os modelos Periódicos Autoregressivos com variáveis externas (PARX), comparadas às obtidas com os modelos Periódicos Autoregressivos (PAR) e Autoregressivos (AR). Nesses trabalhos foi demonstrada a possibilidade de utilizar índices do ENSO e TSMs do Oceano Atlântico Tropical para o caso do Nordeste, e ventos zonais para o caso do Sudeste do país.

3.1.6 - Considerações finais

Diante o que foi exposto ao longo desta seção, nesta tese serão realizadas previsões de afluências de longo prazo, logo, estimou-se a vazão que ocorrerá no próximo ano –um ano na frente, a partir de vazões históricas e indicadores climáticos. Para tal finalidade, são utilizadas variações dos modelos estocásticos de Markov com Estados Ocultos (HMMs).

A razão da utilização dos HMMs deve-se ao fato de esse tipo de modelo permitir tratar a não estacionariedade, variabilidade e persistência observada em séries de vazões com maior robustez do que utilizando outros modelos estocásticos (Thyer e Kuczera, 2000; Akintug e Rasmussen, 2005; Bracken *et al.* 2014; Liu *et al.*, 2018). Nesse sentido, Chen *et al.*, 2016, 2018 mencionam que a não estacionariedade e persistência são quesitos importantes a serem tratados para tentar reduzir a incerteza dos valores previstos de afluências. Além disso, os HMMs possibilitam a inclusão de informações climáticas, que pelo mencionado nesta seção, e na seção 3.3, influenciam a variabilidade das vazões e persistência hidrológica (Sveinsson *et al.*, 2003; Bracken *et al.* 2014). Quanto a outros modelos estocásticos de previsão disponíveis, algumas deficiências em relação aos HMMs são apresentadas a seguir.

Primeiro, no caso da representação da não estacionariedade, os modelos ARMA apresentam dificuldades para capturar características não estacionárias no espectro, pois embora, ao adicionar uma estrutura de onduletas – modelos *Wavelet* ARMA – consigam representar tais características, a estrutura da ACF dificulta a representação de distribuições bimodais. Nessa linha, Bracken *et al.*, (2014) mostraram que os HMMs conseguem capturar não estacionariedade no espectro com melhor desempenho de HMMs com informações climáticas .

Segundo, no caso da persistência, conforme descrito na seção 3.1.2, os modelos ARMA apresentam dificuldades para capturar a persistência de longo prazo especificamente para representar o coeficiente de Hurst. Embora, em alguns casos, esses modelos pela sua estrutura de autocorrelação logrem capturar a persistência observada. No caso de modelos multifractais (veja por exemplo, Lohre et al 2003 e Li *et al.*, 2015), tais modelos conseguem representar o coeficiente de Hurst, embora, podem não capturar características espectrais. Por outro lado, a flexibilidade na inclusão de informações climáticas em modelos multifractais limita a sua utilização (Bracken et al., 2014).

Por fim, com relação à variabilidade, os modelos ARMA utilizam uma estrutura linear, baseada na autocorrelação da série, para relacionar vazões em defasagens anteriores com a vazão em um tempo determinado, porém, tal estrutura não permite uma variação de regimes de vazões – conforme pode ser visto na seção 4.1.

Os HMMs, por sua vez, permitem ajustar relações lineares, ou não, a subconjuntos de dados pertencentes a uma amostra, ainda, os HMMs possuem um mecanismo baseado em uma cadeia de Markov que possibilita variar de uma relação à outra, com isso é possível representar a variação de regimes de vazões. Logo, os HMMs são úteis na análise de séries temporais que apresentam mudanças ou saltos abruptos na média no tempo, sendo conhecidos como modelos de regime variável ou *switching models*, tais regimes ou estados são denominados não observados, ou ocultos, porque não são medíveis.

Além disso, nos HMMs os estados ocultos podem ser associados, embora, subjetivamente, a estados hidro-climáticos, que por sua vez podem ser influenciados por variáveis climáticas de larga escala. Essas variáveis podem ser incluídas nos HMMs igual aos modelos ARMAX, porém, esses últimos modelos também não permitem representar subconjuntos amostrais nem a variação entre os mesmos. E ainda, a flexibilidade no uso de informações climáticas nos HMMs pode ajudar identificar o grau de influência dessas informações sobre os subconjuntos ou estados ocultos.

3.2 - APLICAÇÕES DOS MODELOS DE MARKOV COM ESTADOS OCULTOS NA PREVISÃO DE VAZÃO

Ao longo do tempo têm sido propostos alguns modelos para reproduzir mudanças súbitas em séries temporais juntamente com a memória do processo ou persistência, de longo prazo, um tipo desses modelos são os Modelos de Markov com Estados Ocultos – *Hidden Markov Models* (HMMs).

De forma geral, na literatura a utilização dos HMMs na simulação de precipitação e vazão tem tido bastante sucesso. Vários estudos relatam a aplicabilidade desses modelos para a geração de vazões anuais, por exemplo, Jackson, (1975); Thyer e Kuczera, (2000); Fortin *et al.*, (2004); Akintug e Rasmussen, (2005); Whiting, (2006) e Bracken, (2011); Bracken *et al.*, (2014); Bracken *et al.*, (2016). Esses trabalhos utilizaram series históricas de vazões pertencentes a regiões com forte influência da variabilidade climática como a América do Norte, a África e a Austrália.

Nesses trabalhos, os HMMs mostraram bastante sucesso na geração de sequências de vazões estatisticamente similares às observadas, comparadas com as geradas com os tradicionais modelos Autoregressivos de Médias Móveis –ARMA. Os HMMs tiveram também bom desempenho na representação gráfica da variabilidade das séries estudadas. Isso acompanhado de um sucesso satisfatório na reprodução dos comprimentos de períodos hidro-climáticos, secos e úmidos, demonstrado a partir da análise das propriedades estatísticas desses comprimentos. Em virtude disso, os HMMs conseguiram reproduzir a persistência hidrológica observada. No caso da previsão de vazões anuais Fortin *et al.*, (2004) e Bracken, (2011) utilizaram HMMs logrando também melhor desempenho, de algumas métricas relacionadas com a qualidade de previsões, que os modelos ARMA.

Uma extensão dos HMMs, a qual permite a inclusão de variáveis exógenas, para regular a transição de estados ocultos, denominada Modelos de Markov com Estados Ocultos Não Homogêneos – *Non Homogeneous Hidden Markov Models* (NHMMs) vem sendo também utilizada na hidrologia. Geralmente, os NHMMs foram utilizados para a geração e previsão de precipitação (veja, por exemplo, Hughes *et al.*, 1999; Charles *et al.*, 1999; Robertson *et al.*, 2005; Greene *et al.*, 2011; Fu *et al.*, 2012; Mares *et al.*, 2014), empregando variáveis climáticas como variáveis externas.

A utilização de NHMMs para a geração de series sintéticas de vazões, em escala mensal, foi realizada por Whiting *et al.*, (2004). Esses autores a partir de NHMMs geraram series

sintéticas para várias bacias da Austrália, embora, a não homogeneidade dos modelos não se relacionou à informação climática. Nesse estudo, utilizou-se uma hierarquização temporal, desde vazões anuais, passando por vazões sazonais até mensais, que produziu modelos NHMMs, logo, não foram empregadas variáveis externas.

A aplicabilidade dos HMMs e NHMMs para geração de vazões anuais no Nordeste do Brasil foi realizada por Hernández, (2013). Nesse estudo, demonstrou-se um bom desempenho dos desses modelos na reprodução de estatísticas de persistência de períodos secos e úmidos – comprimentos de períodos úmidos, volume de déficit em períodos secos. Encontrou-se também melhor autocorrelação em comparação com: o modelo Autoregressivo de segunda ordem (AR(2)) e com o a geração de séries sintéticas a partir da distribuição Gama independente. Entretanto, o uso dos HMMs e NHMMs para a previsão de vazões no Brasil não tem sido registrado.

Outra versão de HMMs que tem sido utilizada na hidrologia são os HMM hierárquicos (HHMMs). Por exemplo, Whiting et al. (2004) e Whiting (2006), utilizaram esse tipo de modelos para geração de series sintéticas de precipitação e vazão. Nesses trabalhos, os autores utilizaram um HHMM com duas sequencias ocultas que simulavam a persistência de estados secos e úmidos em escala anual e mensal, respectivamente. Esses trabalhos mostraram a diferença do desempenho de HHMMs com relação ao HMMs básicos, isso medido por métricas relacionadas à persistência dos estados ocultos. Detalhes desses trabalhos podem ser encontrados no Apêndice A.

No apêndice A apresentam-se informações detalhadas de vários trabalhos que utilizaram HMMs. As tabelas A.1 e A.2 trazem estudos aplicados à precipitação e a vazão, respectivamente. Enquanto que, as tabelas A.3 e A.4 mostram os problemas e recomendações dos trabalhos citados nas tabelas A.1 e A.2, respectivamente. Nessas tabelas, nota-se que vários problemas e limitações têm sido solucionados mediante o desenvolvimento de novas técnicas e metodologias. Porém, ainda se faz necessário aprofundar em várias questões, como:

- Realizar previsões de vazões futuras;
- Tamanho da série de observações;
- Incerteza nos parâmetros e nos dados gerados;

- Seleção do melhor modelo e do número de estados ocultos;
- Interpretação física dos estados ocultos;
- Seleção e ajuste das distribuições dependentes;
- Representação da distribuição de probabilidade das observações e do coeficiente de assimetria;
- Seleção dos preditores ou variáveis externas, tratamento estatístico dessa informação.
- Validação do desempenho do modelo;
- Representação da Autocorrelação;
- Geração de valores atípicos ou *outliers*;
- Escala temporal para realizar as análises;
- Métodos de estimação de parâmetros;
- Técnicas de transformação das observações;
- Representação da correlação espacial e temporal dos dados gerados;
- Criação de modelos multivariados, regiões com estados climáticos homogêneos;
- Utilizar outras versões dos HMMs.

Este trabalho não procura analisar todas essas questões, mas sim pretende tratar a previsão de vazões médias anuais utilizando HMMs. Em relação a isso, nos parágrafos seguintes se discutem alguns detalhes dos trabalhos de Fortin *et al.*, (2004); Gelati *et al.*, (2010b); Bracken, (2011) e Liu *et al.*, (2018) que utilizaram HMMs para a previsão de vazões. Sugere-se ao leitor complementar essas informações com os detalhes do desempenho dos HMMs nesses trabalhos, mostrados na tabela 3.1.

A previsão de vazão utilizando HMMs foi inicialmente realizada por Fortin *et al.*, (2004). Esses autores utilizaram o *Shifting Level Model* SL, uma versão melhorada do modelo de Salas e Boes (1980), com estrutura similar à de um HMM. Esse trabalho procurou demonstrar a aplicabilidade do modelo SL para realizar previsões pontuais e probabilísticas de vazões anuais para o rio Senegal na África. Esses autores estimaram as distribuições de densidade de probabilidade (FDPs) de previsão para um ano na frente utilizando amostrados de Gibbs e *Kernel smoothing*. Essas distribuições foram condicionadas modelo SL condicionado, ou não, modelo SL não condicionado ou linear a que existia uma persistência da média das vazões, a partir do ano em que se iniciou a previsão. Com isso, os autores prosseguiram a realizar uma comparação das FDPs de previsão e as observações.

No caso da previsão pontual Fortin *et al.*, (2004) compararam o desempenho do modelo SL frente aos modelos ARMA(1,1), AR(2), um ruído aleatório AR(0), e um modelo em que a previsão da vazão do próximo ano foi a vazão do ano anterior, esse último modelo foi denominado como modelo *Naive* - nome utilizado na economia. Vale ressaltar que, as previsões com os modelos ARMA (1,1), AR(2), AR(0) e *Naive*, não foram realizadas de forma Bayesiana, como foram realizadas pelo modelo SL, elas foram calculadas pelo método de Box e Jenkins (1994) para ter uma comparação inicial, já que estimar previsões via análise Bayesiana saia do escopo do trabalho.

Seguidamente, Gelati *et al.*, (2010b) empregaram um HMM com dois estados ocultos e com informação climática, indicadores do ENSO, para gerar séries sintéticas e previsões para as anomalias trimestrais das afluições ao reservatório Daule Peripa no Equador. Nesse estudo, as anomalias foram padronizadas e log-transformadas para eliminar assimetria positiva na série. O HMM utilizado foi similar aos modelos de Hughes e Guttorp, (1994a); Hughes *et al.*, (1999) e Robertson *et al.*, (2004), com distribuições de densidade de probabilidade iguais a um modelo autoregressivo com variáveis exógenas (ARX) condicionado aos estados ocultos do HMM, razão pela qual esse modelo foi denominado *Markov-Modulated Autoregressive Model with Exogenous Input* (MARX). Na verificação do desempenho das previsões, esses autores realizaram uma análise visual de previsões, calculadas para um trimestre à frente, e observações, testando a correlação entre ambas. Nessa análise encontraram que o HMM se desempenhou melhor para anomalias de vazões altas que para baixas, já que as vazões baixas foram sob estimadas.

Mais tarde, Bracken, (2011) realizou previsões pontuais de vazões anuais - ano hidrológico - para uma bacia do alto Colorado nos Estados Unidos, embora, sem o uso de informações climáticas. Esse autor utilizou um HMMs, com dois e três estados ocultos, não estacionários, com distribuições de probabilidade para as observações Log-Normal - com previa log-transformação das vazões - e Gama. Além disso, o autor utilizou o modelo Autoregressivo de ordem 1 – AR(1). A verificação das previsões baseou-se no *Ranked Probability Skill Score* (RPSS). Nesse caso, um valor igual a 1 indicava uma previsão perfeita, um valor igual a 0 mostrava que não existe diferença entre a previsão climatológica (aquela baseada em valores históricos) e valores negativos indicavam uma previsão pior com respeito à climatológica. Para os HMMs os resultados mostraram que a

mediana, para o período de previsão, do RPSS foi ao redor de 0,2, enquanto, para o AR(1) foi igual a -0,03. Os resultados também indicaram que os HMMs foram melhores para prever anos secos porque em 50% do período das previsões o RPSS foi superior à climatologia, RPSS = 0,26 (HMMs Gama) e RPSS = 0,20 (HMMs Normal), enquanto, para o AR(1) foi próxima da climatologia (RPSS = 0,07). A verificação das previsões foi complementada com análises visuais entre distribuições previstas e valores observados que mostraram também bom desempenho dos HMMs Logo, o autor concluiu que os HMMs são melhores que o AR(1), sendo que são preferíveis modelos com 2 estados ocultos e distribuições Log-Normais, sendo melhores para prever anos secos. Vale a pena ressaltar que, Bracken (2011) utilizou um processo de simulação sem utilizar a distribuições probabilísticas de previsão, para gerar as previsões pontuais utilizando HMMs, sinalizando que essa forma é levemente similar a utilizar distribuições de previsão.

Mais recentemente, Liu *et al.*, (2018) utilizaram um HMM de dois estados ocultos para realizar a previsão de vazões mensais na bacia do rio Yangtzé na China. A distribuição das observações seguiu uma distribuição Normal e utilizaram para gerar distribuições probabilísticas e pontuais uma regressão que mistura distribuições Normais. Na verificação das previsões probabilísticas esses autores utilizaram a versão continua do RPSS, o CRPSS – *Continuous Probability Skill Score*, o qual descreve a qualidade das previsões igual que o RPSS, e para a avaliação visual a métrica PIT - *Probability Integral Transform*, a qual é baseada na distribuição uniforme e mostra se as previsões são uniformes e seguem a distribuição das observações. Para verificar as previsões pontuais utilizaram o coeficiente de Nash Sutcliffe (NS). Os resultados mostraram que as previsões geradas pelo HMM são foram melhores que a climatologia (CRPSS>0,10), também que os HMMs entregam previsões pontuais similares às produzidas por um VSM *Vector Support Machine*, já que, para ambos os modelos o NS foi em torno de 0,2. A partir das análises da métrica PIT, os autores sinalaram que os HMMs entregam previsões confiáveis, sem viés quando existiram viés eles não apresentaram uma tendência significativa . Dessa forma, esses autores concluíram que os HMM entregam previsões de vazões médias mensais satisfatórias.

Tabela 3.1 – Desempenho dos HMMs na previsão de vazões.

Trabalho	Indicadores de Desempenho	Desempenho
Fortin <i>et al.</i> , (2004)	<p>Inspeção visual das FDPs observadas e previstas do SL condicionado e do SL não condicionado.</p> <p>Cálculo do volume total futuro para um período de 7 anos.</p> <p>Comparação da moda da FDP de previsão</p> <p>Comparação utilizando as métricas Viés; Erro médio Absoluto (EMA); Raiz do erro médio absoluto (REMA); versões ponderadas do Viés EMA e REMA. Em que o fator de ponderação era o número de observações mínimo para ajustar o modelo como fator de ponderação.</p>	<p>A FDP de previsão do modelo SL condicionado se aproximou melhor á FDP observada</p> <p>Melhor aproximação do volume total futuro calculado com o SL condicionado. Porém, esse modelo subestimou a probabilidade de excedência desse volume.</p> <p>Os dos tipos de modelo SL conseguiram um bom desempenho em estimar a moda observada comparado com o desempenho dos modelos ARMA(1,1), AR(2), AR(0) e <i>Naive</i></p> <p>O modelo <i>Naive</i> obteve menor valor do viés</p> <p>Os modelos SL apresentaram menores valores dos EMA, REMA e suas versões ponderadas variando entre 2 e 7 % dependendo da métrica empregada.</p> <p>Os modelos SL e ARMA (1,1) apresentaram um comportamento similar nas métricas de comparação, seguido dos modelos AR(2) e AR(0).</p> <p>O modelo <i>Naive</i> igualou aos modelos AR(2) e AR(0) em termos do ponderado EMA</p> <p>Todos os modelos apresentam um aumento da vazão final utilizada para realizar as previsões, valor final, de 6 a 9 vezes. No entanto, o modelo SL tem um desempenho levemente melhor</p>
Gelati <i>et al.</i> , (2010b)	<p>Inspeção visual</p> <p>Correlação entre previsões e observações</p>	<p>O MARX se desempenhou melhor para prever vazões altas, porém, sob estimação as vazões baixas. Na representação da autocorrelação a autocorrelação é mais alta no período de validação, 0,68, que no de calibração, 0,46.</p> <p>As FDPs no período de previsão seguem a tendência da FDP observada.</p> <p>As curvas de probabilidade de excedência no período de previsão mostraram uma maior persistência de vazões baixas.</p>

Tabela 3.1 (Continuação) – Desempenho dos HMMs na previsão de vazões

<p>Bracken (2011)</p>	<p><i>Ranked Probability Skill Score (RPSS)</i> (Wilks, 2006) Correlação entre a média da FDP de previsão e os dados observados Inspeção das FDPs previstas e FDPs observadas</p>	<p>De forma geral os HMMs Normais apresentaram melhor desempenho que os HMMs Gama, reproduzindo variações de vazões adequadamente. Isso foi refletido nas métricas da qualidade de previsões. Especificamente o HMM Normal foi melhor que o AR(1) com alto RPSS e alta correlação (embora não significativa e similar ao modelo HMM Gama) entre a mediana da previsão e observações. O AR(1) teve um desempenho deficiente aproximadamente em um 50 % do período das previsões segundo a métrica RPSS. Os HMMs se comportaram melhor para prever períodos secos e com deficiências para períodos úmidos, isso foi refletido nos valores do RPSS correspondentes a esses períodos. O AR(1) é levemente diferente à climatologia em períodos secos e realmente ruim (RPSS negativo) para períodos úmidos Um modelo HMMs com três estados não entregou melhores resultados, só se destacou na previsão de uma seca intensa observada ao redor de 2004. As FDPs de previsão apresentaram um deslocamento com relação à FDP da climatologia, com formas não Normais, e com valores próximos de zero (0,09).</p>
<p>Liu <i>et al.</i>, (2018)</p>	<p><i>Continuous Ranked Probability Skill Score (CRPSS)</i> (Wilks, 2006) Coeficiente de Nash-Sutcliffe (NS) Inspeção Visual a partir do <i>probability integral transform</i> (PIT) – Métrica baseada na distribuição uniforme</p>	<p>O HMM utilizado entregou valores do CRPSS > 0,1, logo, as previsões obtidas a partir do HMMs foram melhores que a climatologia. O valor do NS foi superior a 0,2. A partir de várias análises baseadas no PIT e de comparações entre a média climatológica e a média das previsões, encontrou-se que as previsões seguem as observações e não existe um viés significativo, nem existe uma tendência de viés. Também pelo PIT o HMM gera previsões que seguem a forma da distribuição observada, já que, as previsões seguem a distribuição uniforme sem notáveis afastamentos dela.</p>

3.3 - VARIABILIDADE CLIMÁTICA E O REGIME DE VAZÕES NO ESTADO DO CEARÁ

A previsão de aflúncias no estado do Ceará é um problema complexo que inclui múltiplas variáveis. Ele está diretamente influenciado por variáveis climáticas, portanto, faz-se necessário o entendimento de como essas variáveis se comportam e dos principais fatores que modificam o seu comportamento (Araújo *et al.*, 2015). Dessa forma, esta seção traz a influência do clima sobre as vazões da região.

3.3.1 - Características gerais do estado do Ceará

O estado do Ceará está localizado no Nordeste do Brasil e tem a maior parte do seu território, 86,8%, inserida dentro de uma região sujeita historicamente a repetidas estiagens, produzidas principalmente pelo déficit de precipitação, abaixo dos 805 mm. No Ceará existe uma alta variabilidade espacial e temporal do regime de chuva, já que, há regiões com índices de precipitação variáveis entre regiões próximas e com concentração do período chuvoso em poucos meses do ano. Somado a isso, o estado tem elevadas temperaturas durante todo o ano, entre 24,1 e 27,1°C, com insolação intensa, entre 2.417 e 2.983 h/ano, que ocasionam elevadas taxas de evapotranspiração, entre 1.469 a 2.904 mm/ano (Ceará, 2008).

A vegetação característica no Ceará é principalmente a caatinga rala e plantas xerófilas de pequeno porte. Os solos apresentam predominantemente embasamento cristalino com pequenas profundidades e grande pedregosidade. Logo, os lençóis freáticos são raros e pouco volumosos, o que dificulta a recarga de aquíferos. Além disso, as águas superficiais e subterrâneas são bastante mineralizadas (Silveira *et al.* 2011b).

Dessa maneira, no Ceará os rios são altamente intermitentes, já que, atingem rapidamente seu ponto de esgotamento durante as estiagens e produzem cheias violentas durante o período de chuvas. Essa intermitência ocasiona uma alta variabilidade dos volumes dos reservatórios, convertendo o Ceará em uma zona altamente vulnerável às condições climáticas (Silveira *et al.*, 2011b).

3.3.2 - Fenômenos climáticos e atmosféricos

Com relação ao clima, o Ceará apresenta um clima semiárido com intensa variabilidade temporal e espacial de chuvas devido fundamentalmente à atuação simultânea de diversos sistemas atmosféricos de várias escalas e à dinâmica de suas interações. Esses sistemas podem variar com as características fisiográficas da região, assim como pelos padrões anômalos de grande escala de circulação atmosférica global associados à ocorrência do *El Niño–Southern Oscillation* (ENSO) e a variações nas temperaturas da superfície do oceano Atlântico tropical (Hastenrath e Heller, 1977; Silveira *et al.*, 2011b).

Com relação ao ENSO, ele é um fenômeno de interações entre a atmosfera e o oceano Pacífico tropical. Possui duas fases principais: *El Niño* (EN) e *La Niña* (LN), embora, exista uma fase neutra ou normal. O desenvolvimento de cada uma dessas fases está relacionado às flutuações das temperaturas da superfície do mar (TSMs) no Oceano Pacífico Tropical e a sua ocorrência é irregular, de 2 a 7 anos, com persistência de 9 a 12 meses, embora, essa persistência em ocasiões seja de até 2 anos (IRI, 2017).

Cada uma das fases do ENSO é apresentada na figura 3.2. Nessa figura observa-se que nas condições normais, as TSMs no Pacífico equatorial leste se encontram relativamente mais frias, junto à costa oeste da América do Sul, e relativamente mais aquecidas no Pacífico equatorial Oeste. Essa figura também mostra que na fase EN ocorre uma diminuição das águas mais frias que afloram próximas à costa oeste da América do Sul e que em condições LN, as águas adjacentes a essa costa tornam-se ainda mais frias devido à intensificação do movimento de ressurgência (INPE/CPTEC, 2017).

As condições atmosféricas no Nordeste do Brasil também são modificadas pelo ENSO. Por exemplo, na fase EN, na maioria das vezes, apresenta-se um deslocamento das células convectivas atmosféricas da Indonésia para o Pacífico Central e Leste, o que acarreta o aumento do ramo descendente sobre o leste da região da Amazônia e a região Nordeste do Brasil. Com isso, inibe-se o processo de formação de nebulosidade local, a Zona de Convergência Intertropical (ZCIT) se desloca para o norte, ocasionando seca na região Nordeste e Leste da Amazônia. Já a ocorrência da fase LN, provoca-se o efeito contrário, favorecendo o aumento da precipitação na região Nordeste e leste da região da Amazônia (Araújo *et al.*, 2015).

A variação das pressões ao nível do mar entre a Austrália e Taiti está também relacionada com o ENSO, do monitoramento dessas pressões resulta o *Southern Oscillation Index* (SOI), o qual ajuda a identificar o desenvolvimento das fases do ENSO. Os eventos EN resultam em prolongados valores negativos do SOI e episódios LN acontecem no caso contrário (IRI, 2017).

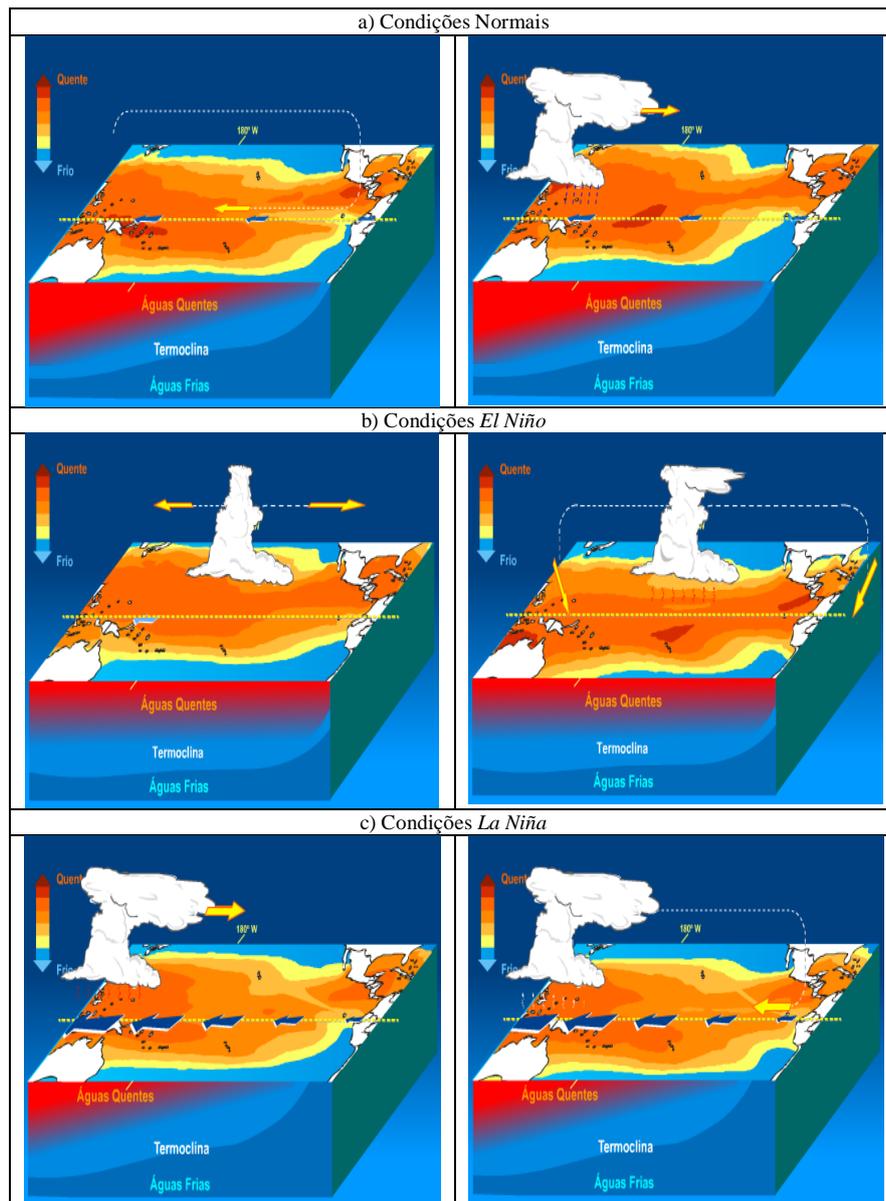


Figura 3.2 – Fases do ENSO: a) condições normais, b) condições *El Niño* e c) condições *La Niña* (INPE/CPTEC, 2017).

Como a variação de pressões está sujeita a fatores atmosféricos de pequena escala, uma forma alternativa de monitorar o surgimento de eventos EN ou LN é através do monitoramento das TSMs no Oceano Pacífico Tropical. Existem quatro regiões principais

de monitoramento, que definem os quatro indicadores climáticos do ENSO. A tabela 3.2 mostra as coordenadas geográficas dessas regiões e a figura 3.3 a sua localização.

Tabela 3.2- Coordenadas as regiões dos índices *El Niño*

Índice	Latitude	Longitude
NINO1+2	0°-10°S	90°W-80°W
NINO3	5°N-5°S	150°W-90°W
NINO3.4	5°N-5°S	170°W-120°W
NINO4	5°N-5°S	160°E-150°W

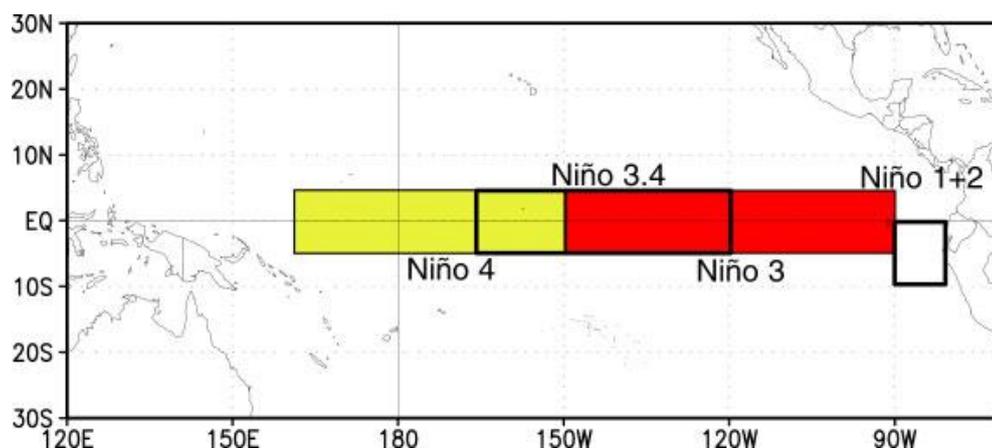


Figura 3.3 – Regiões *El Niño* (NOAA, 2017a)

Especificamente, o NINO1+2 é a zona que primeiro se esquentava durante um episódio EN, a região NINO3 apresenta a maior variação de temperaturas e as medições do NINO4 estão fortemente relacionadas com as condições de precipitação da Indonésia. Porém, quem informa a maioria dos eventos do ENSO e como eles afetam a variabilidade do clima global, e as mudanças nos padrões de precipitação, é a região NINO3.4, já que essa região captura a variabilidade das TSMs e mudanças fortes da precipitação local (IRI, 2017). Mais detalhes desses indicadores podem ser encontrados em NOAA (2017b).

As variações dos indicadores do ENSO acompanhadas com mudanças nos padrões das TSMs do oceano Atlântico tropical modulam e modificam a dinâmica da circulação atmosférica equatorial. Quando as anomalias dos indicadores do ENSO são negativas ou positivas as células de Hadley e Walker se deslocam das suas posições climatológicas e produzem alterações na duração e intensidade das chuvas na região dos trópicos (Hastenrath e Heller, 1977).

Além do ENSO, existem outros sistemas de circulação atmosférica no Estado do Ceará que modificam a intensidade, temporal e espacial, das chuvas, eles são: a Zona de Convergência Intertropical (ZCIT), as Frentes Frias (FF), os Vórtices Ciclônicos de Altos Níveis (VCANs), as Linhas de Instabilidade (LI), os Complexos Convectivos de Mesoescala (CCM), as Ondas de Leste (OL) e as Brisas Marítimas e Terrestres (BMT). O leitor é sugerido a encontrar mais informações desses sistemas em Uvo, *et al.*, (1998); Ferreira e Mello (2005) e FUNCEME, (2017).

Desses sistemas é interessante ressaltar os efeitos da ZCIT sobre a climatologia do NEB. A ZCIT corresponde a uma faixa de nuvens ao redor da zona equatorial terrestre, originada principalmente pela confluência dos ventos alísios dos hemisférios Norte e Sul, essa convergência dos ventos faz com que o ar, quente e úmido ascenda, carregando umidade do oceano para os altos níveis da atmosfera contribuindo à formação das nuvens (Cabral, 2014). A ZCIT determina a magnitude das precipitações na região, já que migra sazonalmente entre os 14°N, em agosto a outubro, para até os 2 a 4°S, entre fevereiro a abril. Uvo, *et al.*, (1998) encontraram que essa variação latitudinal e a intensidade da ZCIT está relacionada às variações das TSMs no oceano Atlântico Tropical (AT), já que, ela é mais significativa sobre os oceanos (Ferreira e Mello, 2005).

3.3.3 - Variabilidade temporal e espacial do regime de chuva

Hastenrath e Heller (1977) e Kousky (1979) mencionam que no Ceará existem duas estações: uma chuvosa e uma seca. A primeira estação, entre dezembro e julho, divide-se em: i) pré-chuvosa, entre dezembro e janeiro, com as FF e os VCANs atuando como principais sistemas causadores de chuva; ii) chuvosa, de fevereiro a maio, na qual a ZCIT é principalmente o sistema causador de precipitações, seguido dos sistemas secundários LI, CCMs e BMTs; e iii) pós-estação, de junho a julho, em que a chuva é produzida pelas OE e os CCMs. A segunda estação, após de julho, se caracteriza por altas pressões atmosféricas e uma quase total ausência de fenômenos atmosféricos causadores de chuva

Espacialmente, Marengo *et al.*, (2008) mencionam que a distribuição de chuvas no NEB, apresenta-se da seguinte forma: i) no Sul-Sudoeste, as chuvas são principalmente de outubro a fevereiro, devido às passagens de FF, procedentes do Sul do país, nesses meses também existem pancadas isoladas de chuvas no final e início, da tarde e noite,

respectivamente; ii) no Norte, cobrindo quase toda a extensão do semiárido, ocorre o principal período chuvoso entre fevereiro e maio - com forte influência da ZCIT; e iii) no Leste, as chuvas são maioritariamente de abril a agosto - ocasionadas em grande parte pelas BMTs. Essa variabilidade espacial do regime de chuvas, pelo comentado no parágrafo anterior, relaciona-se com a variabilidade intranual.

Como comentado ao longo desta seção, o ENSO, principalmente, e o gradiente meridional de anomalias das TSMs sobre o AT influenciam a variabilidade interanual da chuva no Ceará. Embora, a variabilidade interanual das TSMs no AT seja menor do que aquela observada produzida em eventos do ENSO, essa variabilidade influência consideravelmente na variação climática sobre o Ceará (Nobre e Shukla 1996; Fernandes, 2012). Para monitorar essa variabilidade é utilizado o indicador climático Dipolo do Atlântico (DA): diferença aritmética entre as anomalias das TSMs no AT Norte (ATN) coordenadas $5^{\circ}\text{N} - 20^{\circ}\text{N}$, $60^{\circ}\text{W}-30^{\circ}\text{W}$ e Sul (ATS) coordenadas $0^{\circ}\text{S}- 20^{\circ}\text{S}$, $30^{\circ}\text{W} - 10^{\circ}\text{E}$, figura 3.4. De acordo ao sinal do DA, e em conjunto com os padrões do ENSO, A ZCIT muda de posição, produzindo efeitos sobre a precipitação da bacia do Atlântico equatorial e norte da América do Sul incluindo o Norte do NEB e a parte central da Amazônia (Moura e Shukla, 1981; Marengo e Hastenrath, 1993; Nobre e Shukla, 1996).

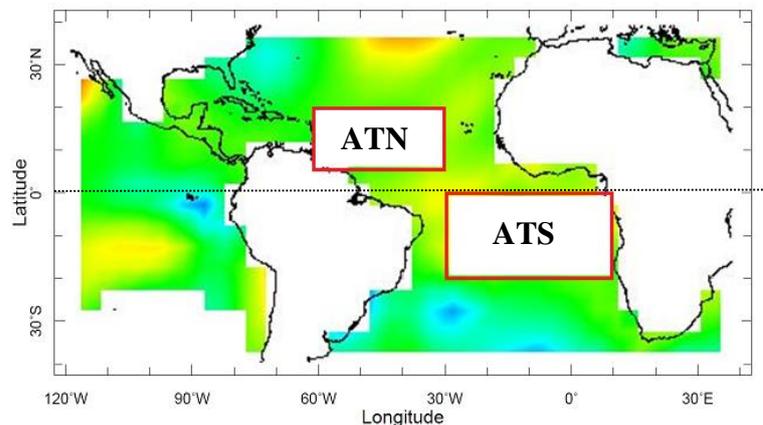


Figura 3.4 – Localização das zonas que constituem o DIPOLO do Atlântico (IRI, 2012a)

A figura 3.5a mostra os efeitos interanuais produzidos pelo DA quando as anomalias das TSMs sobre o ATS estão mais altas, no período de março a maio, e no ATN mais frias. Nessa situação, as anomalias: de TSMs formam um gradiente meridional no sentido Norte-Sul; de pressão ao nível do mar são negativas sobre o ATS e positivas no ATN; dos ventos alísios de sudeste são negativas e positivas para os ventos alísios de nordeste. Observa-se

também, sob essas condições, o eixo de baixa pressão superficial e a confluência dos ventos alísios, com relação às suas posições médias, mais para o Sul. Portanto, para esse padrão do DA, resultam anomalias positivas da precipitação no NEB. Nos períodos secos, padrão oposto do DA, ocorrem efeitos contrários, figura 3.5b (Marengo, 2006).

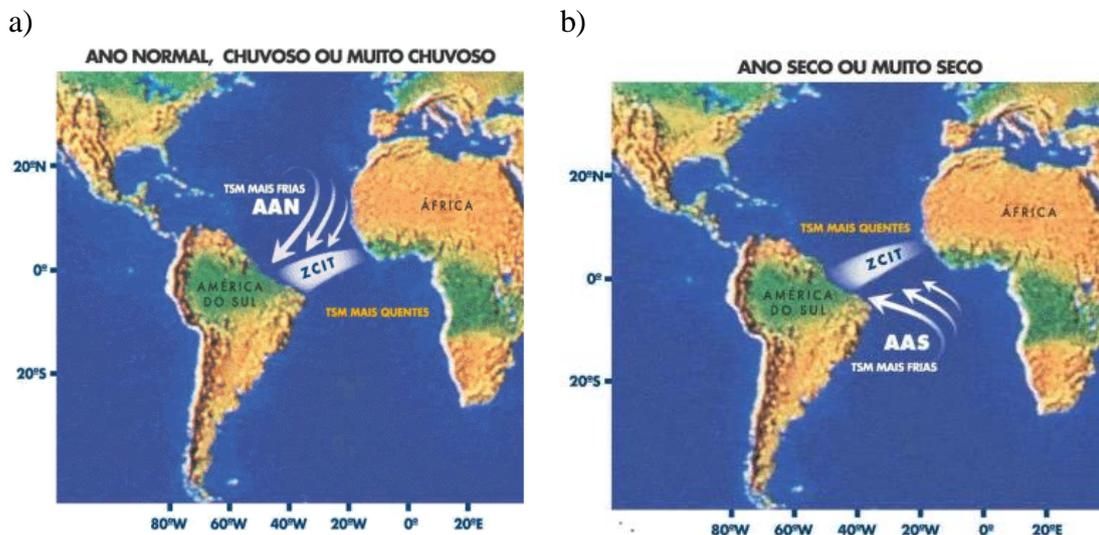


Figura 3.5 – Efeitos produzidos pelo Dipolo do Atlântico (DA): a) períodos chuvosos e b) secos no Nordeste do Brasil (Ferreira e Mello, 2005)

Alves *et al.* (1997) e Lima (2010) mencionam que quando existe um evento EN e um valor positivo do DA (PDA), tem-se anos secos ou muito secos no NEB. A relação é inversa quando ocorre um LN e um valor negativo do DA (NDA), já que, acontecem anos normais, chuvosos ou muito chuvosos nessa região. Quando o ENSO está em condições normais e um NDA (PDA), um ano normal ou chuvoso (seco) é altamente provável.

3.3.4 - A influência das oscilações decenais dos oceanos Atlântico e Pacífico sobre o regime de chuvas

Várias oscilações em diferentes escalas temporais e locais, nos oceanos Pacífico e Atlântico, além das relacionadas ao ENSO e a variações no oceano Atlântico Tropical (AT), influenciam na variabilidade climática em várias regiões do planeta incluindo o NEB. A continuação algumas generalidades dessas oscilações e uma breve revisão sobre a relação com a climatologia do NEB é apresentada.

Um padrão de variabilidade climática, semelhante ao ENSO, descrito na literatura desde finais dos anos 1990 é a Oscilação Decadal do Pacífico (ODP), a qual apresenta uma variação decenal e não interanual como o ENSO. A ODP é caracterizada por anomalias de TSM de sinais opostos no Pacífico: no Centro-Norte e Norte – ao longo da costa da América do Norte e a parte Norte da América do Sul - e no Leste perto das Ilhas Aleutas e do Golfo do Alasca. Essa oscilação afeta principalmente os padrões climáticos e as TSMs no Noroeste do Pacífico, no Alasca e nas ilhas do norte do Pacífico (NOAA, 2018a).

Duas características principais distinguem a ODP do ENSO: primeiro, os “eventos” ODP do século XX persistiram por 20 a 30 anos - enquanto os eventos típicos do ENSO persistiram por 6 a 18 meses -; segundo: as impressões digitais climáticas da ODP são mais visíveis no setor Norte Pacífico e Norte-Americano, enquanto assinaturas secundárias existem nos trópicos – o oposto, visibilidade dessas assinaturas, é verdadeiro para o ENSO. Vários estudos encontraram evidências de apenas dois ciclos completos da ODP no século passado: a fase “fria” de 1890 a 1924 e novamente de 1947 a 1976, enquanto uma fase “quente” de 1925 a 1946 e de 1977 até meados da década de 1990. As causas para a ODP não são conhecidas atualmente, assim como a sua potencial previsibilidade (NOAA, 2018a).

Knigh *et al.*, (2006) menciona que ao longo do século XX têm sido registradas variações multidecadais nas TSMs do Atlântico, as quais modificaram particularmente as condições de precipitação no NEB e no deserto do Sahel (África), a ocorrência de furacões no Atlântico e as condições climáticas no verão norte-americano e europeu. Essas variações das TSMs correspondem à Oscilação Multidecadal do Atlântico (OMA) e são parte de uma variação de temperatura em grande parte do Hemisfério Norte. Fases “frias” da OMA ocorreram nos anos 1900-1920 e 1960-1980, enquanto uma fase “quente” ocorreu entre 1930-1950. Esses períodos coincidem com anomalias na precipitação no NEB: de 1930 a 1950 existiu uma diminuição dos fluxos de chuvas do NEB, enquanto, de 1960 a 1980 ocorreu alta precipitação.

Outra variação climática encontrada no oceano Atlântico é a Oscilação do Atlântico Norte (OAN), uma variação em grande escala na pressão atmosférica entre o sistema subtropical de alta pressão, localizado perto dos Açores, e o sistema de baixa pressão subpolar, perto da Islândia. Essa variação de pressões leva ventos superficiais e tempestades de inverno do

oeste para o leste através do Atlântico Norte, afetando o clima da Nova Inglaterra à Europa Ocidental até o leste - como a Sibéria central e o leste do Mediterrâneo - e ao sul até a África Ocidental. As anomalias na pressão ao nível do mar entre os sistemas de baixa e alta pressão que relacionados com a OAN, quantificam-se pelo índice OAN. Quando o OAN é positivo o nordeste dos Estados Unidos apresenta um aumento na temperatura e uma diminuição nos dias de neve; e no centro dos EUA aumenta a precipitação, o Mar do Norte tem um aumento nas tempestades; e a Noruega, juntamente com o norte da Europa, tem temperaturas mais altas e aumento da precipitação. Quando o OAN é negativo no Atlântico Tropical e na costa do Golfo dos Estados Unidos o número de fortes furacões aumenta; o norte da Europa é mais seco, e na Turquia, junto com outros países do Mediterrâneo, aumenta a precipitação (NOAA, 2018a).

Lucena, *et al.*, (2011) analisaram as relações da chuva do norte do NEB com o ENSO e os modos do DA durante o período de 1948-1997 no contexto das mudanças de fase de longo prazo da ODP e a OAN. Nesse período, esses autores encontraram maior número, e intensidade, de eventos EN, os quais produziram secas nessa região, e eventos LN relacionados a um NDA, ambos contribuindo com anos mais úmidos na região. Antes de 1970 - ano de referência - esses autores encontraram que um LN e um PDA reduziram a magnitude das anomalias de chuva na região. Esses autores concluíram que as mudanças entre as relações do ENSO, o PDA e NDA, e as chuvas no NEB estavam associadas às mudanças de fase do ODP e OAN da metade da década de 1970. No entanto, mencionaram que a influência do Atlântico equatorial sobre o ENSO poderia existir antes de 1970.

Kayano *et al.*, (2018) comentaram que a distribuição temporal do DA durante o período 1901-2012 mostra que eventos PDA são igualmente distribuídos para ambas as fases da OMA, e, que, eventos NDA ocorrem preferivelmente durante a fase fria da OMA, figura 3.8. O alto número de NDAs durante a fase fria da OMA é consistente com os encontrados por Giannini *et al.*, (2004) e confirmam os resultados de Knight *et al.*, (2006). Giannini *et al.*, (2004) encontraram mais frequência do NDA no AT durante o período de 1950-1994, período no qual sobressai a fase fria da OMA. Enquanto, Knight *et al.*, (2006) encontrou uma anomalia da posição da ZCIT durante a fase fria da OMA que produziu condições mais chuvosas sobre o NEB. Outros trabalhos que analisam a influência das oscilações ODP e OMA sobre a quadra chuvosa no estado do Ceará são Knight *et al.*, (2006); Capistrano e Kayano (2012); Kayano *et al.*, (2016) e Costa e Silva (2017).

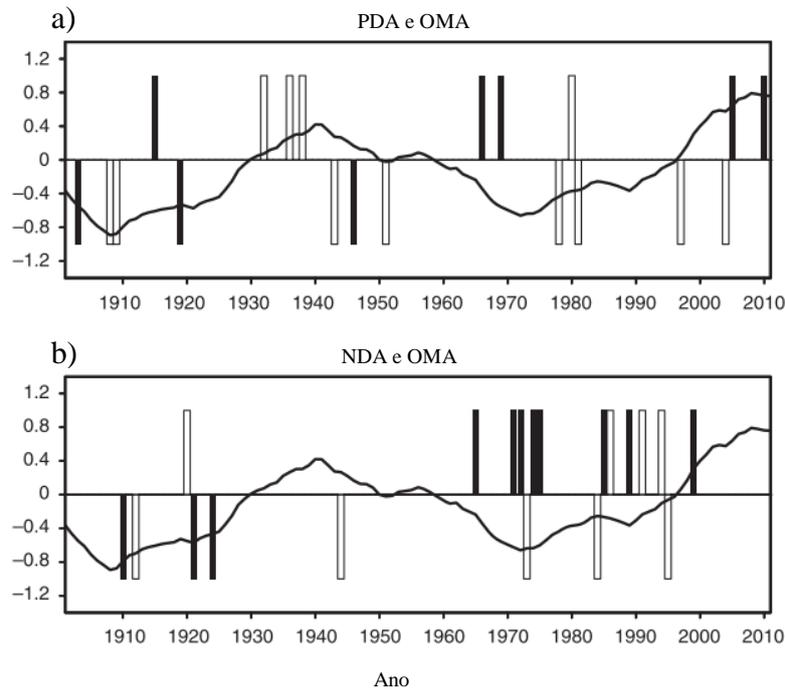


Figura 3.6 – Ocorrência temporal dos modos DA indicados por: 1 se são definidos pela OMA e ATN e -1 se são definidos pela OMA e ATS. As barras cheias indicam PDA (NDA) precedidos por um EN (LN). As barras não preenchidas indicam PDA (NDA) não precedidos por um EN (LN). O índice da OMA (em °C) foi multiplicado por 3 (linha contínua) (Kayano *et al.*, 2018).

3.3.5 - Variabilidade climática interanual e a sua relação com o regime de vazões

Na literatura existem vários trabalhos que mencionam a influência da variabilidade climática no regime de chuva no Nordeste do Brasil e pelo mencionado nas seções anteriores também existe no estado do Ceará. Alguns trabalhos são: 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, Marengo *et al.*, 2015, Nys *et al.*, 2016; Kayano *et al.*, 2018.

No caso das vazões, a relação entre regime das vazões e a variabilidade climática também é encontrada. Por exemplo, Souza Filho e Lall, (2003) a partir do trabalho de Uvo *et al.*, (1998) analisaram a relação entre os índices NINO3 e DA e as afluências de vários reservatórios no Ceará. Para as afluências ao açude Orós esses autores encontraram

“significativas” correlações: -0,21 para o NINO3, período outubro a dezembro, e 0,51 e 0,47 para o DA, períodos de outubro a dezembro e de julho a setembro, respectivamente. Vale mencionar que esses períodos de meses corresponderam a períodos do ano anterior em que aconteceu a afluência observada. Ribeiro (2011) também encontrou correlações semelhantes. Outros trabalhos que tratam esse assunto no Nordeste são Lima (2010) e Alexandre (2012). Esses estudos e outros aplicados especificamente a reservatórios no Ceará são apresentados na seguinte seção.

3.4 - MODELOS UTILIZADOS PARA A PREVISÃO DE VAZÕES NO ESTADO DO CEARÁ

Como se comentou nas seções anteriores as vazões no Ceará apresentam características particulares e são influenciadas por condições climáticas globais e regionais. Logo, esta seção relata alguns detalhes sobre a previsão de afluências no estado do Ceará, e no NEB, para ilustrar como se tem tratado a complexidade do problema da previsão e o uso de informações climáticas.

O estado do Ceará está localizado na região semiárida do Brasil. Os rios dessa região se caracterizam por apresentarem vazões médias anuais intermitentes, com coeficientes de variação maiores que um (1) e distribuições de densidade de probabilidade com assimetria positiva. Em vista disso, na região tem se assumido que as vazões são serialmente independentes e, portanto, utiliza-se uma modelagem baseada nas distribuições de probabilidade Gama e Log-Normal, sendo a primeira a mais utilizada.

Essa metodologia é bastante aceita na região, já que vários trabalhos utilizaram a distribuição Gama ou Log-Normal para gerar números aleatórios e dessa forma gerar vazões sintéticas (Santos, 2009). Devido à performance dessa metodologia, no estado também se realizam previsões de vazões empregando a distribuição Gama.

Como foi mostrado na seção 3.3.5 existem correlações entre afluências e fenômenos climáticos no estado do Ceará. Dessa forma, desde o ano 2000 vem se realizando vários estudos visando aproveitar essas correlações em modelos estatísticos de previsão de vazões. A tabela 3.4 apresenta alguns desses trabalhos e a tabela 3.5 menciona alguns

problemas e recomendações. O leitor é indicado a acompanhar a leitura dessas tabelas com os detalhes adicionais desses trabalhos que se comentam a continuação.

Souza Filho e Lall (2003) utilizaram um método semi-paramétrico: o *K-Nearest Neighbor Resampling* (KNN), para manter a correlação espacial de diferentes locais no Ceará, e também conservar correlação temporal de subperíodos intranuais, de aflúncias observadas. Esses autores empregaram os indicadores NINO3 e DA como variáveis externas que influenciavam as vazões e obtiveram um desempenho satisfatório do KNN, o qual foi mostrado pela alta correlação entre observações e previsões para o período de validação das previsões (1993-2000).

Empregando uma abordagem similar Cardoso *et al.*, (2006,2009) e Ribeiro (2011) também encontraram resultados adequados. Esse último autor incluiu modificações no modelo de Souza Filho e Lall (2003) ao utilizar modelos com três tipos de regressões: linear, *stepwise* e *pool*. Antes de ser incluídas em modelos de previsão, esses trabalhos normalizaram as vazões analisadas. Porém, Souza Filho e Lall (2003) e Cardoso *et al.*, (2006, 2009) aplicaram uma transformação de raiz cubica previamente.

Ribeiro (2011), também simulou a operação dos reservatórios analisados através do *software* AcquaNet empregando as previsões de aflúncias obtidas. Os resultados dessas simulações mostraram que é possível obter cenários com volume alocável diferente de zero. Também encontrou volumes de espera variáveis, que permitiram cenários diferentes no controle de cheias.

Dois trabalhos que não utilizaram modelos estatísticos, mencionados na tabela 3.4, são os estudos de Reis *et al.*, (2009) e Araújo *et al.*, (2015). O primeiro trabalho foi incluído porque atualmente a Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME) realiza previsões sazonais e mensais empregando essa abordagem, vale mencionar que a FUNCEME também a utiliza as abordagens de Souza Filho e Lall (2003) e Cardoso *et al.*, (2006, 2009). O segundo estudo, Araújo *et al.*, (2015), foi citado na tabela porque é um estudo recente e empregou uma abordagem que, embora, seja baseada em Redes Neurais Artificiais (RNA), incluiu informações climáticas.

Dessa forma, Reis *et al.*, (2009) realizaram uma previsão dinâmica de vazões, a partir de previsão climática de chuva e o uso de modelos chuva vazão. As previsões de chuva foram obtidas por conjunto e utilizaram os modelos climáticos regionais: *Regional Atmospheric Modeling System* (RAMS) e *Regional Spectral Model* (RSM). O modelo chuva-vazão utilizado foi *Soil Moisture Accounting Procedure* (SMAP). Esse estudo entregou um bom desempenho, avaliado a partir de uma análise visual, das previsões obtidas comparadas com as observações.

Araújo *et al.*, (2015) utilizaram RNA incluindo informações climáticas, com previa Normalização das observações. O estudo mostrou um bom desempenho dessa abordagem, apesar de não ter realizado a comparação do desempenho frente a outros modelos de previsão, mostrou que as RNA são úteis para a previsão de afluências em reservatórios do Ceará.

Alguns comentários de dois trabalhos que realizaram previsões de afluências para os reservatórios Três Marias, Sobradinho e Itaparica, na bacia do Rio São Francisco no NEB são apresentados na tabela 3.6 e na tabela 3.7, são citados alguns problemas e recomendações desses trabalhos. Tais estudos obtiveram um bom desempenho das previsões obtidas a partir de modelos Periódicos Autoregressivos com Variáveis Exógenas (PARX). Outros trabalhos que realizaram previsões para vazões afluentes em reservatórios no Brasil também com modelos PARX e que não são mencionados na tabela 3.6 são estudos de Silveira, (2014) e Oliveira e Lima, (2016).

Silveira, (2014), além dos modelos PARX, utilizou o KNN, de Souza Filho e Lall (2003), na escala mensal e sazonal e também realizou previsões dinâmicas de vazões. Esse estudo analisou várias escalas temporais de previsão e uma ampla gama de modelos procurando um planejamento integrado dos reservatórios das usinas hidrelétricas que fazem parte do Sistema Interligado Nacional (SIN), incluindo também o Sistema Jaguaribe–Metropolitano.

Oliveira e Lima, (2016) analisaram os reservatórios do Sistema Interligado Nacional (SIN), incluindo também os reservatórios da bacia do Rio São Francisco, propondo a integração de previsões semanais e mensais a partir da técnica de Ponderação Bayesiana de Modelos – BMA. Nesse trabalho as previsões semanais foram obtidas por modelos próprios do

Operador Nacional do Sistema Elétrico –ONS. Esse trabalho e o estudo de Silveira, (2014) também apresentaram um sucesso ao empregar os modelos PARX para realizar previsões de afluências.

Tabela 3.3 – Previsão de vazões no estado do Ceará utilizando informação climática

Trabalho	Local	Escala	Modelos utilizados	Indicadores de Desempenho	Desempenho	Variáveis externas
Souza Filho e Lall (2003)	Estado do Ceará Sistema Jaguaribe-Metropolitano	Anual Sazonal	KNN	Inspeção gráfica entre percentis de previsões e observações	Correlação de 0.9 entre a mediana da previsão e as vazões observadas	NINO3 e Dipolo do Atlântico
Cardoso <i>et al.</i> , (2006, 2009)	Estado do Ceará Sistema Jaguaribe-Metropolitano e Reservatório Orós	Anual	KNN	Inspeção gráfica entre percentis de previsões e observações Erro Quadrático Médio (EQM).	As vazões observadas seguem a tendência da mediana das previsões com correlações maiores a 0.6. Nos anos 1995, 1996, 1998, 2000 e 2001 o EQM da previsão foi menor que o da climatologia. Nos anos 1994, 1997 e 1999 o apresentam EQM da previsão foi maior que o da climatologia.	NINO3 e Dipolo do Atlântico
Reis <i>et al.</i> , (2009)	Estado do Ceará 8 reservatórios incluindo o Sistema Jaguaribe-Metropolitano	Sazonal Mensal	RSM, RAMS e SMAP	Inspeção gráfica entre percentis de previsões, observações e valores observados pontuais. Erro Quadrático Médio (EQM)	A faixa de variação das previsões foi visivelmente menor que a faixa de variação das observações. As previsões seguem a tendência das observações.	Variáveis climáticas da previsão de chuva
Ribeiro (2011)	Nordeste do Brasil, incluindo o Sistema Jaguaribe-Metropolitano	Sazonal Interanual	KNN	Comparação das verossimilhanças das previsões e climatologia com distribuições de densidade de probabilidade <i>Kernel</i> e Gama. Inspeção gráfica entre os percentis das previsões, climatologia e observações	Reservatório Orós: As previsões seguem a tendência de valores altos e baixos. Anos de menor Q obtiveram máximas verossimilhanças. O modelo com regressão linear e que inclui todos os preditores teve melhor desempenho ao comparar as verossimilhanças Sistema-Jaguaribe Metropolitano: Resultados similares ao do reservatório Orós. O desempenho das previsões no Orós melhorou ao ser incluído no sistema todo diferente aos outros reservatórios. A verossimilhança da distribuição Gama foi superior que a <i>Kernel</i> .	TSM, ventos meridionais, altura geopotencial de 1000mb, NINO3 e Dipolo do Atlântico
Araújo <i>et al.</i> , (2015)	Estado do Ceará Reservatório Orós	Sazonal Interanual	RNA Usando <i>Perceptons</i>	Coefficiente de determinação (R^2), Nash-Sutcliffe (NS). Inspeção Visual das distribuições de densidade de probabilidade previstas e observadas	Na calibração: NS=0.8 e $R^2=0.8$ Na validação: Previsões adequadas de vazões pico (cheias), NS=0.3. Para anos secos NS=0.65. A partir das distribuições de frequência encontrou-se sobre estimativa de vazões baixas e melhor previsão de vazões altas	NINO3 TSM do Atlântico Norte e Sul Tropical

Tabela 3.4 – Problemas e recomendações dos trabalhos citados na Tabela 3.4

Trabalho	Problemas	Recomendações
Souza Filho e Lall (2003)	Vieses problemáticos em anos secos	Tratar a incerteza Empregar os resultados para simular a operação de reservatórios. Utilizar diferentes métodos de previsão na estrutura interna do modelo. Refinar o modelo para entregar resultados práticos para a tomada de decisão.
Cardoso <i>et al.</i> , (2006,2009)	Resultados das correlações entre previsões e observações sensíveis à inclusão de novos anos no período de validação, devido aos poucos anos desse período.	Continuar realizando este tipo de estudos. Validação cruzada para avaliar a performance.
Reis <i>et al.</i> , (2009)	Falta de dados para a calibração do modelo hidrológico.	Continuar realizando este tipo de estudos. Analisar melhor a qualidade das previsões. Aumentar e variar o horizonte de previsão.
Ribeiro (2011)	Analisar separadamente e em conjunto os reservatórios se reflete no desempenho dos modelos	Utilizar mais preditores climáticos ou procurar outros novos. Utilizar séries mais extensas. Utilizar outras métricas de desempenho, por exemplo, <i>skill score</i> . Empregar esse tipo de abordagens na previsão de vazões
Araújo <i>et al.</i> , (2015)	Desempenho pobre na validação. Porém as previsões ficaram dentro os intervalos de confiança de 95%	Ter em conta que as anomalias das TSMs do oceano Atlântico influenciaram em um 63 % a previsão das vazões. Comparar as previsões com as obtidas com outros modelos Empregar o modelo como ferramenta de suporte à decisão na operação de reservatórios

Tabela 3.5 – Trabalhos sobre previsão de vazões no Nordeste do Brasil

Trabalho	Local	Escala	Modelos utilizados	Indicadores de Desempenho	Desempenho	Variáveis externas
Lima (2008)	Estado da Bahia Reservatório Sobradinho	Mensal	PARX PAR AR	Inspeção gráfica comparando previsões e observações. Coeficiente de determinação, R^2 Erro Absoluto Médio –EAM Viés Definição de vazões altas e baixas a partir dos percentis 33% e 66%.	Bom desempenho dos modelos PARX em todas as métricas analisadas para previsões de até 4 meses de antecedência. Com até 1 mês de antecedência melhor reprodução de vazões altas, $R^2=0.61$. Com até dois meses de antecedência boa representação de vazões altas e baixas. Com até três meses de antecedência, $R^2= 0.11$. Para vazões baixas com até 5 meses de antecedência, $R^2=0.61$. O modelo AR teve o pior desempenho de todos os modelos.	Vento zonal, NINO3, Temperaturas da superfície dos Oceanos Atlântico Tropical e Pacífico
Alexandre (2012)	Estado da Bahia Reservatórios Tremarias Sobradinho e Itaparica	Mensal	PARX PAR Correlação Espacial modelada com: Correlação dos Resíduos Preditivos (CRP), correlação entre outros locais e Análise de Componentes Principais (ACP).	Na calibração: Erro Médio Absoluto (EMA) Coeficiente de Correlação Na Validação: Erro Médio Percentual Absoluto (EMPA) Coeficiente de Nash-Sutcliffe (NS) Distância Multicritério (DM)	Na correlação espacial os resultados foram variáveis para cada reservatório. Mas, de forma geral, os modelos PAR resultaram ter melhor desempenho ao analisar a série toda. Os PARX obtiveram melhor desempenho em vazões sazonais e para períodos secos. Desses modelos também foram melhores os quais tiveram CRP e muito melhores foram os que tiveram ACP. De forma geral: NS acima de 0.7, DM ao redor de 0.34 e EMPA acima de 0.22.	27 índices climáticos que incluíram: ventos zonais, índices do ENSO, temperaturas da superfície do Oceano Atlântico

Tabela 3.6 – Problemas e recomendações dos trabalhos citados na Tabela 3.6

Trabalho	Problemas	Recomendações
Lima (2008)	Tendência das previsões a sub estimar os valores observados devido a viés negativos.	Aplicar o modelo PARX a outros reservatórios Procurar outros índices climáticos
Alexandre (2012)	Não se estabeleceu um melhor modelo. Alguns modelos apresentam EMPA de quase de 100% para o período de dezembro janeiro e fevereiro.	Aperfeiçoar os modelos com CRP e mais ainda aqueles com ACP.

4 - MODELOS AUTOREGRESSIVOS COM VARIÁVEIS EXÓGENAS E MODELOS DE MARKOV COM ESTADOS OCULTOS

A seguir são apresentados conceitos e definições dos modelos estocásticos utilizados neste estudo. A primeira parte deste capítulo trata sobre os ARX, começando por uma breve revisão dos modelos autoregressivos de médias móveis (ARMA), e a segunda discute os HMMs. Vale ressaltar que nas expressões matemáticas mostradas a seguir as nomenclaturas marcadas em negrito representam matrizes ou vetores.

4.1 - MODELOS AUTOREGRESSIVOS COM VARIÁVEIS EXÓGENAS

Os modelos Autoregressivos são considerados clássicos e bastante utilizados na hidrologia. O leitor pode encontrar informação detalhada sobre o processo de ajuste, seleção, validação e previsão de valores utilizando modelos AR e ARX em textos clássicos como Box e Jenkins (1994) ou Bras e Rodriguez-Iturbe (1992).

4.1.1 - Função de autocorrelação

Os modelos Autoregressivos surgem do conceito de autocorrelação encontrada em séries temporais. A autocorrelação de forma geral é definida como a relação de uma observação com observações anteriores. Na ausência de autocorrelação a série é considerada independente (Salas, 1993; Meko, 2005). A função de autocorrelação (ACF) quantifica essa correlação, a qual expressa o grau de dependência temporal entre as observações. A ACF é um processo de auto-comparação e expressa a correlação linear entre as observações, de uma mesma série, R_t , e R_{t-k} , em que k representa uma defasagem ou *lag* específica. A ACF serial ou amostral pode ser calculada como:

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

em que $\bar{R} = \frac{1}{N} \sum_{t=1}^N R_t$ é a média amostral. Para $k = 0$ se tem que $r_0 = 1$ e na prática é comum que $k = \frac{N}{10}$ ou $k = \frac{N}{4}$, sendo N o tamanho da série.

A ACF oferece informação detalhada da estrutura interna da série temporal através do correlograma, no qual se plotam os valores de r_k em função de k . Se os valores de r_k caem dentro dos valores críticos máximos, $r_{k_{max}}$, e mínimos, $r_{k_{min}}$ comumente assumidos como $\pm 2/\sqrt{N}$, respectivamente, a hipótese nula, $r_k = 0$, é rejeitada. Isso indica que a série não é puramente aleatória e apresenta algum tipo de persistência (Machiwal *et al.*, 2002; Guimarães e Santos, 2011; Box *et al.*, 2016). Outras informações que podem ser extraídas do correlograma são a correlação de curto prazo, não estacionariedade da série, flutuações sazonais e *outliers* (Chatfield, 2003).

Em alguns casos, se a série é auto correlacionada dificulta a aplicação de testes de hipótese, por exemplo, Mann Kendall, pela redução do tamanho efetivo da amostra. A autocorrelação, em alguns casos dificulta análises de regressão, cálculo de covariância e análises de correlação com outras séries temporais, por exemplo, séries de vazões em diferentes locais (Meko, 2005).

Com relação às séries de vazões, elas apresentam forte autocorrelação devido à inércia ou transição do sistema físico, efeitos de armazenamento subterrâneo, tipo de solos, ou devido à persistência de um estado climático. Uma alta autocorrelação positiva indica uma tendência de que valores altos continuem altos, igualmente para valores baixos. Se a autocorrelação é negativa indica que existe variabilidade na série (Box e Jenkins, 1994). Graficamente, uma autocorrelação positiva mostra uma persistência na série, com longos ou curtos períodos, acima ou abaixo de um limiar, por exemplo, a mediana. No caso de autocorrelação negativa, existirá uma variação dos valores, em relação a um limiar, e ausência de períodos longos ou períodos curtos (Bras e Rodriguez-Iturbe, 1992).

4.1.2 - Modelos autoregressivos de medias móveis

Quando a ACF decai rapidamente com o aumento de k se podem definir uma classe de modelos denominados de memória ou dependência curta (Salas e Boes, 1980). No caso contrário, se a ACF decai lentamente, definem-se os modelos de memória longa. À primeira classe pertencem a família de modelos lineares e estacionários Autoregressivos de Médias Móveis (ARMA), a qual inclui os modelos Autoregressivos (AR) e de medias moveis (MA). Os modelos ARMA têm como principais atrativos a sua fácil aplicabilidade e alcance para tratar diferentes tipos de dados, desde séries biológicas até de dados

financeiros, entre outros (Whiting, 2006). A formulação desse tipo de modelos surgiu na década dos anos 70 com os trabalhos de Box e Jenkins.

Um modelo bastante utilizado para previsão de vazões é o modelo Autoregressivo de primeira ordem $AR(1)$, o qual utiliza a autocorrelação no *lag* 1. A sua definição em termos de um processo com média igual a zero é:

$$R_t = \phi_1 R_{t-1} + z_t \quad (4.1)$$

em que z_t é um conjunto de variáveis aleatórias independentes e identicamente distribuídas com média zero, $\mu_z = 0$, e variância σ_z^2 denominado sequência *White noise* ou ruído aleatório. Essa sequência é o principal exemplo de um processo aleatório estacionário. Na série z_t a condição de estacionariedade implica que essa sequência seja não auto correlacionada, por tanto, a sua auto covariância é:

$$\gamma_k = Cov(z_{t+k}, z_t) = E[(z_{t+k} - \mu_z)(z_t - \mu_z)] = \begin{cases} \sigma_z^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

Essa condição de estacionariedade para z_t é equivalente a que essa sequência seja normalmente distribuída com $\mu_z = 0$ e σ_z^2 . No caso do modelo $AR(1)$, a estacionariedade implica que a componente endógena, ϕ_1 , satisfaça que $|\phi_1| < 1$. Nesse modelo, a ACF é dada por $r_k = \phi_1^k$ e a variância é definida por:

$$\sigma_y^2 = \frac{\sigma_z^2}{(1 - r_1^2)} \quad (4.2)$$

No modelo $AR(1)$ as suas características são refletidas no seu espectro, dessa forma, processos governados por esse modelo são dominados por variações de baixa frequência. No caso contrário, quando $\phi_1 < 0$, o processo apresenta variações de altas frequências (Whiting, 2006).

Para uma ordem p o modelo Autoregressivo $AR(p)$ assume a seguinte forma:

$$R_t = \phi_1 R_{t-1} + \dots + \phi_p R_{t-p} + z_t \quad (4.3)$$

Box e Jenkins (1994) mencionam que além da dependência serial, os valores de x_t também podem ser afetados por um erro aleatório passado que não é contemplado no modelo AR, tal erro é levado em conta por um processo de Média Móvel MA. O modelo de média móvel de ordem 1 MA(1) é definido como:

$$R_t = \mu + z_t - \theta_1 z_{t-1} \quad (4.4)$$

em que μ é a média populacional; θ_1 , são os parâmetros de médias móveis; z_t e z_{t-1} são as componentes dos erros aleatórios nos tempos t e $t - 1$, respectivamente. Logo, cada valor de x_t é produzido por um erro aleatório e uma combinação linear de erros aleatórios de valores prévios. Similarmente aos modelos autoregressivos existem modelos MA de ordem q MA(q), definidos como:

$$R_t = \mu + z_t - \theta_1 z_{t-1} - \dots - \theta_q z_{t-q} \quad (4.5)$$

Whiting (2006) mencionam que para garantir estacionariedade do processo MA(q) a condição de invertibilidade deve ser cumprida, isso é equivalente a que $|\theta_q| < 1$ para todo q .

Na análise de algumas séries temporais resulta útil combinar os processos AR e MA, visando aproveitar as vantagens de ambos os dois e procurando parcimônia, logo, reúnem-se modelos AR e MA com poucos parâmetros, nos denominados modelos Autoregressivos de Médias Móveis ARMA(p, q) que seguem a seguinte forma:

$$R_t = \mu + \phi_1 R_{t-1} + \phi_2 R_{t-2} + \dots + \phi_p R_{t-p} + z_t - \theta_1 z_{t-1} - \theta_2 z_{t-2} - \dots - \theta_q z_{t-q} \quad (4.6)$$

Dessa forma, os processos AR(p) e MA(q) são um caso especiais de um processo ARMA, em que um ARMA($p, 0$) é um processo AR(p) e um ARMA($0, q$) corresponde a um modelo MA(q).

4.1.3 - Modelos autoregressivos com variáveis exógenas

A partir das definições da seção 4.1.2, um modelo $ARX(p)$ é um modelo $AR(p)$ no qual se incluem variáveis exógenas, a sua forma geral é a seguinte:

$$R_t = \sum_{i=1}^p \phi_i R_{t-i} + \sum_{j=0}^r \omega_j X_{t-1} + z_t \quad (4.7)$$

em que R_t é a variável resposta no tempo t ; R_{t-i} é a variável resposta na defasagem i ; ϕ_i são as componentes endógenas autoregressivas; X_{t-1} são as componentes exógenas no instante $t-1$; e ω_j são parâmetros da regressão com as variáveis externas; z_t é definido igual que na equação 4.1; e p e r são a ordem no modelo AR e o número de variáveis exógenas, respectivamente. Esses modelos e a sua versão ARMAX pertencem à classe de modelos lineares de função de transferência, em que uma variável externa ou *input* entra a influenciar o processo estocástico, tal impulso gera respostas diferentes em um sistema dinâmico (Box *et al.*, 2016).

Algumas aplicações na previsão de vazões utilizando esse tipo de modelos foram apresentadas na seção 3.3.3, além desses estudos estão os trabalhos de: Haltiner e Salas (1988); Awwad e Valdes (1992); Awwad *et al.*, (1994); Kelman *et al.*, (2000), Castellano-Mendez *et al.*, (2004), Kwon (2002); Sveinsson *et al.*, (2008); Bogner e Kalas (2008), que realizaram previsão de vazões no curto prazo, desde vazões diárias a vazões mensais, em diferentes locais.

Todos esses trabalhos conseguiram um bom desempenho ao incluir variáveis de base física no processo como: derretimento de neve, precipitação, temperaturas da superfície do mar e outras variáveis atmosféricas. Outras aplicações dos modelos ARMAX e sua versão não linear podem ser encontradas em Chang (2009) e aplicações em outras áreas em Ochoa-Riviera *et al.*, (2002). Os estudos específicos empregando modelos ARMAX e a sua versão periódica PARMAX para previsão de vazões no estado do ceara, e no Nordeste, foram apresentados na seção 3.4.

4.2 - MODELOS DE MARKOV COM ESTADOS OCULTOS – *HIDDEN MARKOV MODELS*

Nesta seção algumas definições e propriedades dos Modelos de Markov com Estados Ocultos –*Hidden Markov Models* (HMMs) são tratadas. Começa-se por sua as bases do modelo e sua definição formal, seguidamente se apresenta a forma de estimação dos parâmetros em HMMs, a forma de realizar previsões e por último se apresentam extensões da forma básica dos HMMs. As definições apresentadas nesta seção são baseadas nas apresentadas em Zucchini e MacDonald (2009) e já que, não existe uma nomenclatura padronizada, as notações e símbolos matemáticos utilizados são os utilizados por Zucchini e MacDonald (2009) e Kirshner (2005). O leitor pode encontrar mais informações sobre os HMMs nesses trabalhos e em Bishop (2008).

4.2.1 - Modelos de mistura independente e cadeias de Markov

Os HMMs consistem em dois processos estocásticos: i) uma mistura independente e ii) um processo de Markov – o qual permite a dependência no processo de mistura. Dessa forma, a seguir se apresentam as definições desses dois processos.

Um modelo de mistura independente permite representar uma população que tem grupos não observados, ou seja, quando as observações seguem uma distribuição multimodal distribuição com uma ou mais médias. Por exemplo, se o conjunto de N observações independentes $R_{1:N} = \{R_1, \dots, R_t, \dots, R_N\}$ de variáveis contínuas possui um Função de Densidade de Probabilidade (FDP) multimodal, um modelo de mistura independente ajusta uma distribuição de mistura a essas observações, tal distribuição, é composta por $f_i(r)$, $i = 1, \dots, m$, FDPs independentes. A mistura das $f_i(r)$ é realizada através de uma variável aleatória discreta, S , com probabilidade de ocorrência δ_i . Logo, o modelo opera da seguinte maneira: uma variável aleatória R_t é gerada mediante uma $f_i(r)$, a qual se ativa pelo valor de S que é definido pela probabilidade δ_i . Dessa forma, a FDP de mistura para R_t é definida como:

$$f(r) = \sum_{i=1}^m \delta_i f_i(r) \quad (4.8)$$

A Figura 4.1 mostra uma mistura de dois componentes ($f_1(r)$ e $f_2(r)$). Nesse caso, quando é definida δ_1 então $S = 1$ e se ativa a $f_1(r)$, no caso contrario, $S = 2$ ativa a $f_2(r)$, quando é dada δ_2 . Pela sua definição de probabilidades, deve-se cumprir que $\delta_2 = 1 - \delta_1$. Por enquanto, não se sabe muito sobre o processo da variável S , simplesmente se sabe que quando ativadas f_1 e f_2 , geram-se as observações.

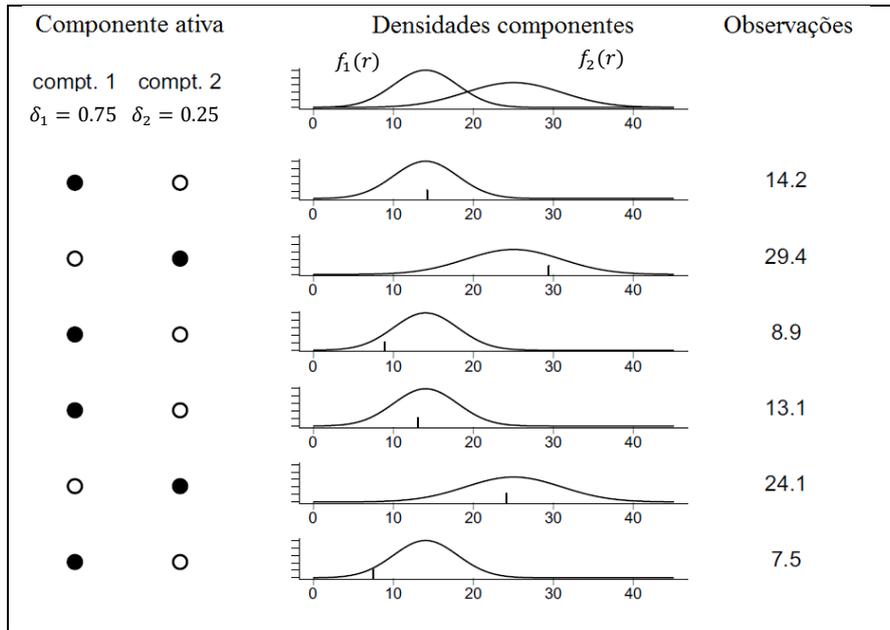


Figura 4.1 – Distribuição de mistura de dois componentes. A sequência das componentes ativas é: 1, 2, 1, 1, 2, 1 (esquerda). As densidades ou distribuições ativas (meio) geram as observações (direita) (Zucchini e MacDonald, 2009, traduzido).

O outro processo incluído em um HMM é uma Cadeia de Markov de primeira ordem, a qual é definida pelo conjunto de variáveis aleatórias discretas $S_{1:N} = \{S_1, \dots, S_t, \dots, S_N\}$ que satisfazem, para todo t , a seguinte propriedade:

$$P_r(S_{t+1}|S_{1:N}) = P_r(S_{t+1}|S_t) \quad (4.9)$$

logo, S_{t+1} é condicionado ao valor mais recente S_t , por sua vez, S_t é condicionado a S_{t-1} . Se $i, j = 1, 2, \dots, m$ são os estados da Cadeia de Markov, a probabilidade de S_t ser igual i quando S_{t-1} é igual a j é denominada probabilidade de transição e define a transição entre os valores da sequência $S_{1:N}$. Dessa forma, essas probabilidades de transição definem-se como:

$$\gamma_{i,j}(t) = P_r(S_{t+1} = i | S_t = j) \quad (4.10)$$

a equação 4.10 define as probabilidades de transição, $\gamma_{i,j}(t)$, só em função do tempo, portanto, a Cadeia de Markov é homogênea. Se uma variável externa influencia essas probabilidades de transição a Cadeia de Markov será então não homogênea. Comumente, as $\gamma_{i,j}(t)$ são reunidas em forma matricial como:

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

$\mathbf{\Gamma}$ é uma matriz de $m \times m$ estados denominada matriz de probabilidades de transição (MPT) e contém os elementos i, j de $\gamma_{i,j}(t)$, tal que $\sum_{j=1}^m \gamma_{i,j} = 1$. Também, para essa Cadeia de Markov se pode definir a:

$$\mathbf{u}(t) = (P_r(S_t = 1), P_r(S_t = 2), \dots, P_r(S_t = m)) \quad (4.11)$$

Em que $P_r(S_t = j)$ representa a probabilidade marginal de acontecer o estado j no tempo t . Para $t = 1$, $\mathbf{u}(1)$ descreve a distribuição inicial da Cadeia de Markov e se pode demonstrar que $\mathbf{u}(t + 1) = \mathbf{u}(t)\mathbf{\Gamma}$.

Outra definição importante para as Cadeias de Markov é a distribuição estacionária $\boldsymbol{\delta}$, a qual satisfaz:

$$\boldsymbol{\delta}\mathbf{\Gamma} = \boldsymbol{\delta} \quad (4.12)$$

$$\boldsymbol{\delta}\mathbf{1}' = \boldsymbol{\delta} \quad (4.13)$$

em que $\mathbf{1}'$ é autovetor direito de $\mathbf{\Gamma}$. As equações 4.12 e 4.13 representam, respectivamente, estacionariedade e que $\boldsymbol{\delta}$ seja realmente uma distribuição de probabilidade.

Com isso, uma Cadeia de Markov é estacionária se começa em $\boldsymbol{\delta}$ e continua em ela para todos os tempos t subsequentes, isso devido ao fato de que $\mathbf{u}(t + 1) = \mathbf{u}(t)\mathbf{\Gamma}$. Se a MPT é homogênea, isso não é suficiente para que a Cadeia de Markov seja estacionária, portanto, torna-se necessário que $\mathbf{u}(1)$ seja uma distribuição estacionária. Em vista disso, para que $\boldsymbol{\delta}$, seja uma distribuição estacionária se deve cumprir que:

$$\boldsymbol{\delta}(\mathbf{I}_m - \mathbf{\Gamma} + \mathbf{U}) = \mathbf{1} \quad (4.14)$$

Em que \mathbf{I} , é um vetor coluna de uns; \mathbf{I}_m , é uma matriz identidade de $m \times m$ e \mathbf{U} é uma matriz unitária de $m \times m$, que contém todos os seus elementos iguais a 1.

A equação 4.14 é suficiente e necessária para garantir estacionariedade e permite calcular facilmente δ . Para Cadeias de Markov de ordem superior, segunda ordem ou maior, uma análise similar pode ser encontrada em Zucchini e MacDonald (2009) e Bishop (2006).

4.2.2 - Modelos de Markov com estados ocultos

Um HMM é um tipo de modelo de mistura dependente, no qual é adicionada uma Cadeia de Markov a um modelo de mistura independente para realizar a dependência entre as observações, já que, um processo de mistura independente não permite realizar essa dependência. Logo, para as séries $R_{1:N}$ e $S_{1:N}$, um HMM se define como:

$$P_r(S_t | \mathbf{S}_{1:t-1}) = P_r(S_t | S_{t-1}), \quad t = 2, 3 \dots, N \quad (4.15)$$

$$P(\mathbf{R}_t | \mathbf{R}_{1:t-1}, \mathbf{S}_{1:t}) = P(R_t | S_t), \quad t \in \mathbb{N} \quad (4.16)$$

A equação 4.15 representa um processo paramétrico não observado ou oculto que consiste em uma Cadeia de Markov, de primeira ordem, e satisfaz a equação 4.9. Já a equação 4.16 equivale a um processo estocástico dependente dos estados S_t . Portanto, quando o estado S_t é conhecido, a distribuição de R_t depende só do estado atual S_t e não a estados anteriores. Assim, um HMM consiste em dois processos, os quais são mostrados na figura 4.2. Para visualizar melhor a geração das observações em um HMM, a figura 4.3, análoga à figura 4.1, ilustra um HMM discreto com dois estados ocultos. Nessa figura, a Cadeia de Markov segue o caminho 2, 1, 1, 1, 2, 1, portanto, ativam-se as distribuições de probabilidade de cada um desses estados, as quais geram cada uma das observações.

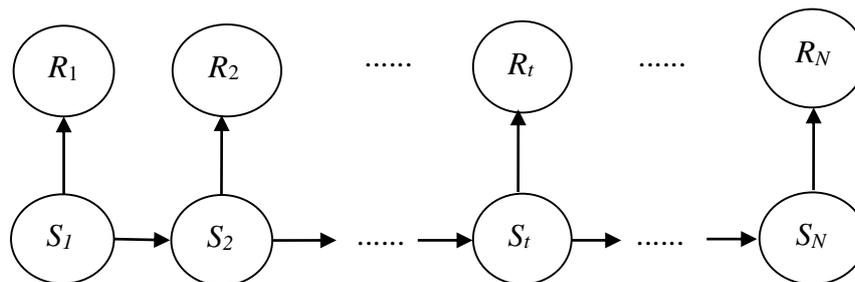


Figura 4.2 – Representação gráfica de um Modelo de Markov Oculto – HMM

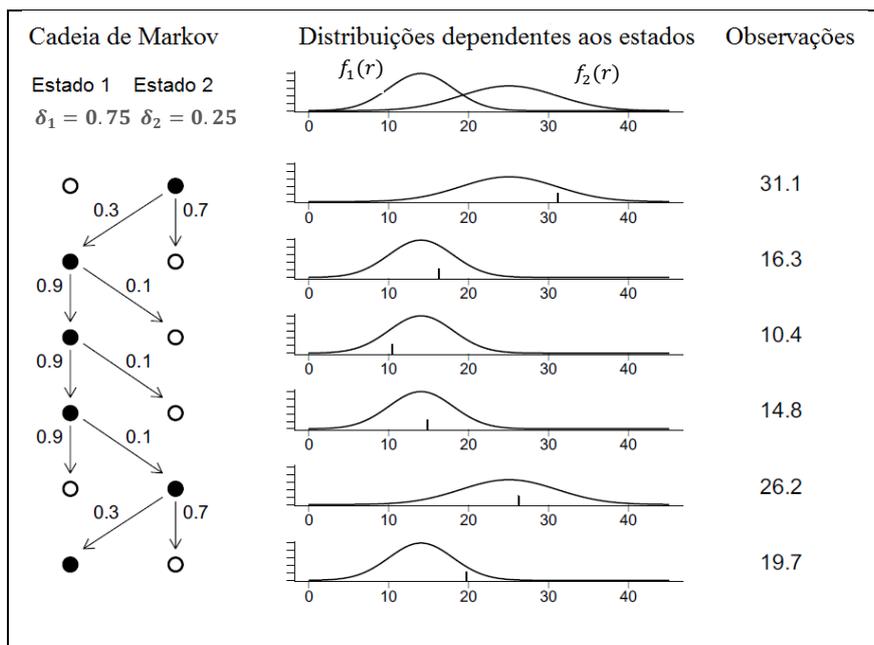


Figura 4.3 – Geração das observações em um HMM discreto de dois estados ocultos (Zucchini e MacDonald, 2009, traduzido)

Vale mencionar que, o número de estados ocultos m em um HMM corresponde ao número de estados discretos da Cadeia de Markov S_t . Portanto, a série R_t , de variáveis contínuas, poderá ser modelada por um HMM de m estados ocultos com: i) uma distribuição estacionária, δ , de tal forma que $\delta_i = P_r(S_1 = i)$; ii) uma MPT, Γ , com elementos $\gamma_{ij} = P_r(S_t = i | S_{t-1} = j)$ os quais regulam a transição entre os estados da Cadeia de Markov e iii) m FDPs, $f_i(r)$, dependentes dos estados ocultos, tal que se R_t está no estado i , a probabilidade $p_i(r)$ corresponde a:

$$p_i(r) = P(a < r \leq b | S_t = i) = \int_a^b f_i(r) dr \quad i = 1, 2, \dots, m \quad (4.17)$$

Um HMM com $m = 1$ gera uma série de variáveis aleatórias mutuamente independentes ou ruído aleatório, *White Noise*, com distribuição $f(r)$ e sem variação de estados (Akintug e Rasmussen, 2005).

Neste trabalho R_t , representa uma série de aflúncias médias anuais e S_t um conjunto de estados hidro-climáticos que influenciam essas vazões, tal que $\{s_1, s_2, \dots, s_m\}$ para $\{s_t, t = 1, 2, \dots, N\}$. Isso oferece uma alternativa para modelar a interação entre persistência de regimes climáticos e resposta hidrológica. Por exemplo, em um HMM de 2 estados

ocultos, o valor s_1 pode ser um estado “úmido” e o valor de s_2 um estado “seco” (Thyer e Kuczera, 2000). Vale a pena ressaltar aqui, que a definição de um estado ser úmido ou seco é subjetiva e depende dos propósitos do estudo.

Com relação às distribuições marginais, a equação 4.11 pode ser rescrita da seguinte forma:

$$u_i(t) = P_r(S_t = i) , \text{ para } t = 1, 2, \dots, N \quad (4.18)$$

a partir dessa equação, tem-se que:

$$p(r) = \sum_{i=1}^m P_r(S_t = i)P(a < r \leq b|S_t = i) = \sum_{i=1}^m u_i(t)p_i(r) \quad (4.19)$$

em forma matricial a equação 4.19, resulta em:

$$p(r) = \mathbf{u}(t)\mathbf{P}(r)\mathbf{1}' \quad (4.20)$$

em que $\mathbf{P}(r)$ representa a $diag(p_1(r_t), p_2(r_t), \dots, p_m(r_t))$ para $t = 1, 2, \dots, N$.

Por outro lado, para que a Cadeia de Markov seja homogênea e estacionaria se deve satisfazer que:

$$P(a < r \leq b) = \delta\mathbf{P}(r)\mathbf{1}' \quad (4.21)$$

4.2.3 - Estimação de parâmetros

Comumente para estimar os parâmetros de um HMM é empregado o Método da Máxima Verossimilhança (MVS), em que a verossimilhança, L_N , é descrita como:

$$L_N = \delta\mathbf{P}(r_1)\mathbf{\Gamma}\mathbf{P}(r_2)\mathbf{\Gamma}\mathbf{P}(r_3) \dots \mathbf{\Gamma}\mathbf{P}(r_N)\mathbf{1}' \quad (4.22)$$

e sob a suposição de estacionariedade:

$$L_N = \boldsymbol{\delta} \boldsymbol{\Gamma} \mathbf{P}(r_1) \boldsymbol{\Gamma} \mathbf{P}(r_2) \boldsymbol{\Gamma} \mathbf{P}(r_3) \dots \boldsymbol{\Gamma} \mathbf{P}(r_N) \mathbf{1}' \quad (4.23)$$

Dois resultados bastante utilizados no contexto dos HMMs, obtidos a partir da equação 4.22, são as probabilidades para frente e para trás $\boldsymbol{\alpha}_t, \boldsymbol{\beta}_t$, definidas como:

$$\boldsymbol{\alpha}_t = \boldsymbol{\delta} \mathbf{P}(r_1) \boldsymbol{\Gamma} \mathbf{P}(r_2) \boldsymbol{\Gamma} \mathbf{P}(r_3) \dots \boldsymbol{\Gamma} \mathbf{P}(r_N) = \boldsymbol{\delta} \mathbf{P}(r_1) \prod_{c=2}^t \boldsymbol{\Gamma} \mathbf{P}(r_c) \quad (4.24)$$

e:

$$\boldsymbol{\beta}'_t = \boldsymbol{\Gamma} \mathbf{P}(r_{t+1}) \boldsymbol{\Gamma} \mathbf{P}(r_{t+2}) \dots \boldsymbol{\Gamma} \mathbf{P}(r_N) \mathbf{1}' = \left(\prod_{c=t+1}^N \boldsymbol{\Gamma} \mathbf{P}(r_c) \right) \mathbf{1}' \quad (4.25)$$

em que $\boldsymbol{\alpha}_t$ é a probabilidade de observar a sequência parcial r_1, r_2, \dots, r_t , terminando no estado i no tempo t . Similarmente, $\boldsymbol{\beta}'_t$ 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).

Com esses resultados, a verossimilhança pode ser calculada recursivamente a partir da equação 4.24 como:

$$L_N = \boldsymbol{\alpha}_N \mathbf{1}' \quad (4.26)$$

em que $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \mathbf{P}(r_t)$, $t \geq 2$, e $\boldsymbol{\alpha}_1 = \boldsymbol{\delta} \mathbf{P}(r_1)$. No caso estacionário $\boldsymbol{\alpha}_0 = \boldsymbol{\delta}$ e $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \mathbf{P}(r_t)$, $t = 1, 2, \dots, N$.

Maximizar a equação 4.22 não é uma tarefa simples porque os parâmetros $\boldsymbol{\delta}$ e $\boldsymbol{\Gamma}$ estão sujeitos a restrições, já que são matrizes que contêm probabilidades. Igualmente, dependendo do tipo de FDPs dependentes os seus parâmetros possuem restrições. Em vista disso, vários métodos têm sido propostos na literatura. Dois métodos comumente empregados são: a Maximização Direta da Verossimilhança e o Algoritmo de Maximização da Expectativa, o leitor pode encontrar detalhes desses procedimentos no Anexo A.

4.2.4 - Previsão utilizando Modelos de Markov com estados ocultos

Nos HMMs para realizar previsões é necessário definir uma distribuição de previsão, a qual se define a partir da distribuição da observação R_t condicionada a todas as outras observações. Dessa forma, se $\mathbf{R}_{-t} = R_1, \dots, R_{t-1}, R_{t+1}, \dots, R_N$, e $\mathbf{r}_{-t} = r_1, \dots, r_{t-1}, r_{t+1}, \dots, r_N$, são as observações em todos os tempos diferentes ao tempo t e a partir dos resultados e as propriedades de $\boldsymbol{\alpha}_t$ e $\boldsymbol{\beta}_t$, se tem, no caso discreto:

$$P_r(R_t = r | \mathbf{R}_{-t} = \mathbf{r}_{-t}) = \sum_{i=1}^m \omega_i(t) p_i(r) \quad (4.27)$$

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

em que $\omega_i(t) = d_i(t) / \sum_{j=1}^m d_j(t)$ 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 $\boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma}$. No caso em que $t = 1$, $d_i(t)$ representa o produto da i -ésima entrada do vetor $\boldsymbol{\delta}$ e a i -ésima entrada do vetor $\boldsymbol{\beta}_1$.

Com essa distribuição condicional é possível definir a distribuição de previsão, também uma distribuição condicional, a qual é a distribuição de R_{N+h} condicionada a $\mathbf{R}_N = \mathbf{r}_N$, com h , sendo o horizonte de previsão. Assim, define-se a distribuição de previsão $P_r(R_{N+h} = r | \mathbf{R}_N = \mathbf{r}_N)$ como:

$$P_r(R_{N+h} = r | \mathbf{R}_N = \mathbf{r}_N) = \frac{\boldsymbol{\alpha}_N \boldsymbol{\Gamma}^h \mathbf{P}(x) \mathbf{1}'}{\boldsymbol{\alpha}_N \mathbf{1}'} \quad (4.28)$$

em que $\mathbf{R}_N = R_{1:N} = \{R_1, \dots, R_N\}$ e $\mathbf{r}_N = r_{1:N} = \{r_1, \dots, r_N\}$. A equação 4.28 se pode escrever como uma mistura de FDPs dependentes dos estados ocultos, da seguinte forma:

$$P_r(R_{N+h} = r | \mathbf{R}_N = \mathbf{r}_N) = \sum_{i=1}^m \xi_i(h) p_i(r) \quad (4.29)$$

em que $\xi_i(h)$ é a i -ésima entrada do vetor $\alpha_N \Gamma^h / \alpha_N \mathbf{1}'$. Através da equação 4.29 é possível realizar previsões pontuais e também previsões em intervalos. No caso contínuo a equação 4.29 também é válida, só se deve lembrar que são utilizadas FDPs.

4.2.5 - Considerações finais

Diante o que foi exposto, nos HMMs uma série de observações que possua subgrupos não observados é representada a partir de uma série de variáveis discretas não observadas, ou latentes, que seguem um processo de Markov. Logo, cada observação está condicionada ao estado da variável latente e é gerada por uma distribuição dependente. Dessa forma, um HMM corresponde a uma extensão dos modelos de mistura independente, em que a escolha da distribuição geradora não é selecionada de forma independente, já que essa seleção depende da escolha da distribuição dependente previa (Bishop, 2008).

Os HMM foram inicialmente propostos para variáveis discretas, porém seu uso tem-se estendido para problemas que tratam variáveis contínuas. Bishop (2008) menciona que os HMMs têm sido amplamente utilizados em reconhecimento de voz e de assinatura, veja, por exemplo, Jelinek, 1997; Rabiner e Juang, 1993; Nag *et al.*, 1986, e análises de sequências biológicas como: proteínas e Ácido desoxirribonucleico (ADN) como nos trabalhos de Krogh *et al.*, 1994; Durbin *et al.*, 1998 e Baldi e Brunak, 2001. Outras várias aplicações em áreas como: bioestatística, geofísica, biologia e econometria são mostradas em Zucchini e MacDonald (2009).

Na hidrologia várias aplicações dos HMMs já foram apresentadas na seção 3.2. Como foi mencionado, em tal seção, os HMMs têm sido empregados para representar a variabilidade de estados hidrológicos ou climatológicos, por exemplo, estados secos e úmidos, ou outros estados fisicamente possíveis. Além disso, nesses trabalhos tem se analisado a influência de variáveis climáticas sobre esses estados hidro-climáticos. Ainda, esses estudos, demonstraram a capacidade dos HMMs para capturar a persistência de longo prazo observada em séries temporais de precipitação e vazão.

4.3 - EXTENSÕES DA FORMA BÁSICA DOS MODELOS DE MARKOV COM ESTADOS OCULTOS

O HMM mostrado na seção 4.2 consiste na forma básica desse modelo. Tal forma pode se estender a múltiplas formas, todas elas para satisfazer os requerimentos do problema a ser tratado. Vale mencionar que as análises realizadas para o HMM básico são válidas também para as outras formas (Bishop, 2006). Nesse sentido, a seguir se apresentam alguns casos bastante aplicados na literatura, alguns deles são aplicados em este trabalho.

4.3.1 - HMMs para Observações que dependem de covariáveis

Quando variáveis como temperatura, velocidade do vento, temperaturas da superfície do mar (TSMs) ou outras variáveis influenciam as afluências, duas formas, embora, outras formas são possíveis, de incluir em um HMM covariáveis X_t que influenciam as observações R_t . A primeira forma é incluí-las nas FDPs dependentes dos estados ocultos e a segunda é incorporá-las nas MPTs. Essas duas classes de HMMs são discutidas a seguir.

4.3.1.1 - HMMs com covariáveis nas distribuições de densidade de probabilidade

Com relação à primeira classe de HMMs, a figura 4.4 apresenta um HMM com covariáveis nas FDPs.

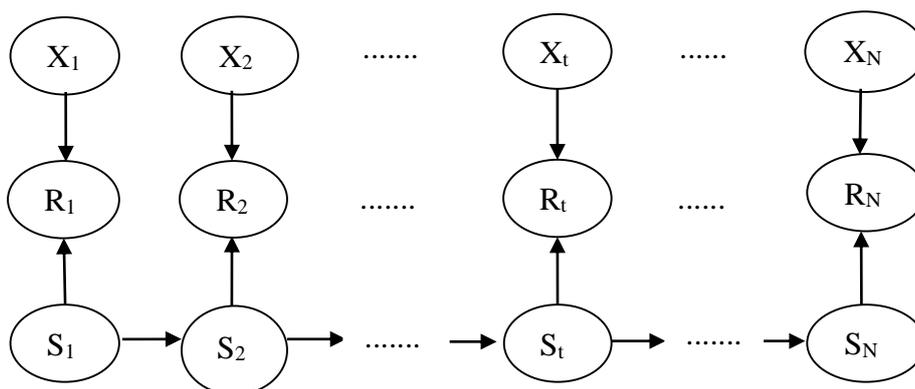


Figura 4.4 – Representação gráfica de um Modelo de Markov Oculto com covariáveis nas distribuições dependentes dos estados.

Nesse modelo a verossimilhança é:

$$L_N = \delta \mathbf{P}(r_1, x_1) \mathbf{\Gamma} \mathbf{P}(r_2, x_2) \mathbf{\Gamma} \mathbf{P}(r_3, x_3) \dots \mathbf{\Gamma} \mathbf{P}(r_N, x_N) \mathbf{1}' \quad (4.30)$$

em que $\mathbf{P}(r, x)$ representa a $\text{diag}(p_1(r_t, x_t), p_2(r_t, x_t), \dots, p_m(r_t, x_t))$ para $t = 1, 2, \dots, N$.

Em esta versão de HMMs as probabilidades de transição dos ocultos seguem a forma da equação 4.15, assume-se, então, que a Cadeia de Markov é homogênea e estacionária (Zucchini et al. 2016). Já, equação 4.16 se transforma em:

$$P(R_t | R_{1:t-1}, S_{1:t}, X_{1:t-1}) = P(R_t | S_t, X_t = x) \quad (4.31)$$

dessa forma, as probabilidades das observações são condicionadas a variáveis externas, o que permite a utilização de diversas funções de ligação, por exemplo, aquelas pertencentes a Modelos Lineares Generalizados (MLGs), o que o torna a esse tipo de HMMs bastante flexíveis (Zucchini e MacDonald, 2009). Dois MLGs úteis para modelar séries de vazões, variáveis contínuas e positivas, são os MLGs Normais e Gama. Mesmo que a série de vazões não apresente normalidade, uma transformação logarítmica pode ser utilizada e empregar MLGs do tipo Normal.

Dobson (2002) menciona que em um MLG Normal:

$$E(R_t) = \mu_t = x_t' \beta; \quad R_t \sim N(\mu_t, \sigma^2) \quad (4.32)$$

em que $R_t = \{R_1, \dots, R_t, \dots, R_N\}$ é o conjunto de variáveis aleatórias independentes. Logo, em este modelo a função de relação é $g(\mu_t) = \mu_t$ e a função de variância é $V(\mu_t) = 1$ (McCullagh e Nelder, 1989). Matricialmente a equação 4.32 pode ser escrita como:

$$\mathbf{R} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (4.33)$$

em que:

$$\mathbf{R} = \begin{bmatrix} R_1 \\ \vdots \\ R_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_1' \\ \vdots \\ x_N' \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$$

com $e_t = \{e_1, \dots, e_t, \dots, e_N\}$ representando um conjunto de variáveis aleatórias independentes identicamente distribuídas, logo $e_t \sim N(0, \sigma^2)$. Os estimadores de máxima verossimilhança para β e σ^2 são:

$$\hat{\beta} = (X'X)^{-1}X'R \quad (4.34)$$

$$\hat{\sigma}^2 = \frac{1}{N - p - 1} (R - X\hat{\beta})'(R - X\hat{\beta}) \quad (4.35)$$

em que $X'X$ deve ser uma matriz invertível.

No caso do MLG Gama, Altman (2009) comenta que existem três funções de relação comumente utilizadas, elas são:

$$1) \text{ relação inversa: } g(\mu_t) = \frac{1}{\mu_t}; \quad (4.36)$$

$$2) \text{ relação logarítmica: } g(\mu_t) = \log \mu_t; \quad (4.37)$$

$$3) \text{ relação identidade: } g(\mu_t) = \mu_t. \quad (4.38)$$

em que μ_t é definida pela equação 4.32 e a equação 4.38 é a mesma função de relação do MLG Normal. Para a modelagem de séries de afluições, a função de relação logarítmica resulta útil, já que evita valores negativos da média. De outro lado a função de variância é $V(\mu_t) = \mu_t^2$ (McCullagh e Nelder, 1989).

Logo, se $R_t \sim \text{Gama}(\theta_t, \eta_t)$ e ao assumir a função de relação logarítmica (Davidian 2008),

$$E(R_t) = \exp(\mu_t) = \exp(x_t'\beta) \quad (4.39)$$

$$\text{Var}(R_t) = \sigma_t^2 = \sigma^2 E(R_t)^2 = \sigma^2 [\exp(x_t'\beta)]^2 \quad (4.40)$$

e, da mesma forma que na distribuição Gama, os parâmetros θ_t e η_t , são:

$$\theta_t = \frac{\sigma_t^2}{E(R_t)} \quad (4.41)$$

$$\eta_t = \frac{E(R_t)}{\theta_t} \quad (4.42)$$

4.3.1.2 - HMMs com covariáveis nas probabilidades de transição de estados ocultos

A segunda classe de HMMs é comumente conhecida como HMMs não homogêneos – *Non Homogeneous Hidden Markov Models* (NHMMs). Nesses modelos, a cadeia de Markov não é mais homogênea devido a que as MPT têm a influência de covariáveis. A figura 4.4, representa graficamente a estrutura desse modelo.

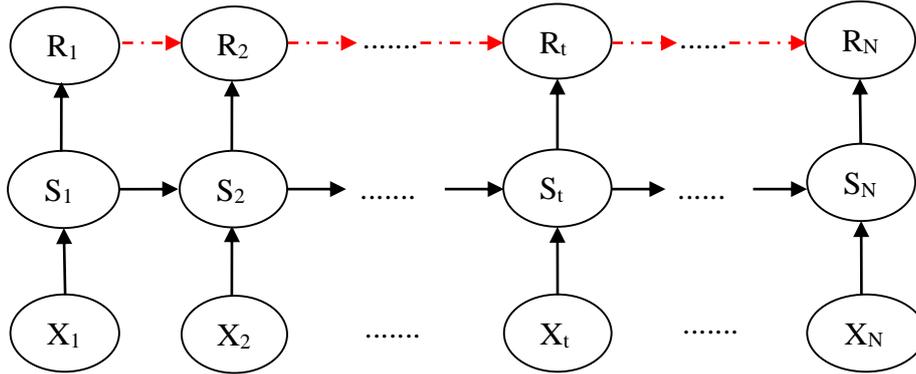


Figura 4.5 – Representação gráfica de um Modelo de Markov Oculto Não Homogêneo, NHMM.

Dessa forma, em um NHMM a verossimilhança é:

$$L_N = \delta(x_1)\mathbf{P}(r_1)\Gamma(x_2)\mathbf{P}(r_2)\Gamma(x_3)\mathbf{P}(r_3) \dots \Gamma(x_N)\mathbf{P}(r_N)\mathbf{1}' \quad (4.43)$$

em que as probabilidades de transição de estados são:

$$P(S_t | \mathbf{S}_{1:t-1}, \mathbf{X}_{1:N}) = P(S_t | \mathbf{S}_{1:t-1}, \mathbf{X}_{1:t}), \quad t = 1, 2, \dots, N \quad (4.44)$$

portanto:

$$\delta_i(\mathbf{x}) = P(S_1 = i | \mathbf{X}_1 = \mathbf{x}) \quad (4.45)$$

$$\gamma_{ij}(\mathbf{x}) = P(S_t = i | S_{t-1} = j, \mathbf{X}_t = \mathbf{x}) \quad (4.46)$$

Na literatura vários autores utilizaram diferentes tipos de modelos logísticos para modelar essas probabilidades de transição. Por exemplo, Hughes *et al.* (1999) utilizaram um modelo autolístico multivariado para observações binárias, esse modelo segue a seguinte forma:

$$\Pr(S_t = j | S_{t-1} = i, \mathbf{X}_t) \propto \Pr(S_t = j | S_{t-1} = i) P(\mathbf{X}_t | S_{t-1} = i, S_t = j) = \quad (4.47)$$

$$\gamma_{ij} \exp \left\{ -\frac{1}{2} (\mathbf{X}_t - \mu_{ij}) \mathbf{V}^{-1} (\mathbf{X}_t - \mu_{ij})' \right\}$$

em que μ_{ij} é a média de \mathbf{X}_t condicionada a S_{t-1} e S_t , \mathbf{V} é a matriz de covariância de \mathbf{X}_t . Esses autores também mencionam que para encontrar os parâmetros, desse modelo autolístico, deve-se cumprir que $\sum_{j=1}^m \gamma_{ij} = 1$ e $\sum_{j=1}^m u_{ij} = 0$. Esse modelo é bastante utilizado na literatura, por exemplo, Hughes e Guttorp, (1994); Bellone *et al.*, (2000) e Gelati *et al.*, (2010b) utilizaram ele.

Um modelo multilístico equivalente ao de Hughes *et al.* (1999) é o mencionado por Robertson e Smyth, (2003); Robertson *et al.*, (2004); Kirshner (2005) e Gelati *et al.*, (2010b), o qual segue a seguinte estrutura:

$$P(S_t = i | S_{t-1} = j, \mathbf{X}_t = \mathbf{x}) = \frac{\exp(\sigma_{ji} + \rho_i' \mathbf{x})}{\sum_{i=1}^m \exp(\sigma_{ji} + \rho_i' \mathbf{x})} \quad \text{para } t = 2, \dots, N \quad (4.48)$$

$$P(S_1 = i | \mathbf{X}_1 = \mathbf{x}) = \frac{\exp(\lambda_i + \rho_i' \mathbf{x})}{\sum_{i=1}^m \exp(\lambda_i + \rho_i' \mathbf{x})} \quad \text{para } t = 1 \quad (4.49)$$

em que os parâmetros $\lambda_i, \sigma_{ij} \in \mathbb{R}$ e $\rho_i \in \mathbb{R}^D$ é um vetor com D sendo o número de covariáveis incluídas no problema. Para identificar os parâmetros desse modelo é necessário que $\lambda_1 = 0$, $\sigma_{j1} = 0$ e $\rho_1 = 0$. Hughes *et al.*, (1999) e Kirshner (2005) mencionam que a partir desses modelos logísticos é possível chegar ao HMM básico.

Gelati *et al.*, 2010b, no seu trabalho calcularam para $t = 1$ a distribuição inicial como a distribuição estacionária e depois condicionaram essa distribuição aos índices climáticos para $t = 1$, empregando o procedimento de Hughes e Guttorp, (1994a) e Hughes *et al.*, (1999).

Para um HMM de dois estados ocultos Zucchini e MacDonald, (2009) e Bracken *et al.*, (2014), mencionam que pode ser utilizada uma regressão binomial logística da forma:

$$\eta_t = \log \frac{\pi_t}{1 - \pi_t} = \beta_0 + \beta \mathbf{x}_t \quad (4.50)$$

em que é η_t a função de relação logística, β_0 é o interceto e $\beta \in \mathbb{R}^D$ é um vetor dos parâmetros da regressão logística. Dessa forma:

$$P(S_t = i | S_{t-1} = j, \mathbf{X}_t = \mathbf{x}) = \frac{\exp(\beta_0 + \beta \mathbf{x}_t)}{1 + \exp(\beta_0 + \beta \mathbf{x}_t)} \quad (4.51)$$

a função de relação logística toma só valores entre zero e um que são interpretados como probabilidades. Hastie *et al.*, (2008) mostram um procedimento para encontrar os parâmetros dessa regressão, assim como para regressões de ordem superior.

4.3.2 - Modelos Autoregressivos de Markov com estados ocultos

Quando a observação R_t depende do estado S_t , e da observação R_{t-p} , sendo p a defasagem, tal processo pode ser modelado por um HMM Autoregressivo AR(p)–HMM. A figura 4.6, apresenta a estrutura de um AR(1) –HMM e um AR(2) –HMM.

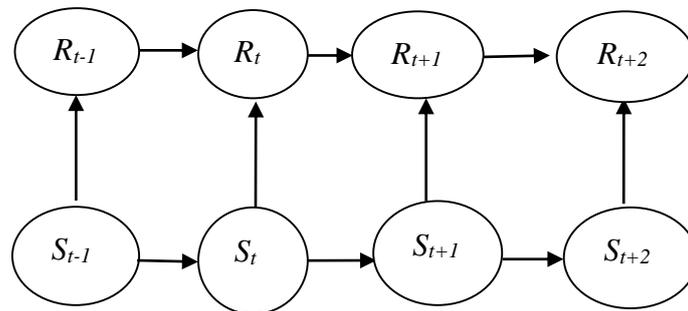


Figura 4.6 – Representação gráfica de um AR(1)–HMM (superior).e de um AR(2)–HMM (inferior)

Quando $p = 1$ se tem o modelo AR(1)–HMM com verossimilhança igual a:

$$L_N = \delta \mathbf{P}(r_1, r_0) \mathbf{\Gamma} \mathbf{P}(r_2, r_1) \mathbf{\Gamma} \mathbf{P}(r_3, r_2) \dots \mathbf{\Gamma} \mathbf{P}(r_N, r_{N-1}) \mathbf{1}' \quad (4.52)$$

Também pode existir um AR(1) –NHMM, nesse caso a verossimilhança é dada por:

$$L_N = \delta(x_1) \mathbf{P}(r_1, r_0) \mathbf{\Gamma}(x_2) \mathbf{P}(r_2, r_1) \mathbf{\Gamma}(x_3) \mathbf{P}(r_3, r_2) \dots \mathbf{\Gamma}(x_N) \mathbf{P}(r_N, r_{N-1}) \mathbf{1}' \quad (4.53)$$

A figura 4.5 apresenta um AR(1) –NHMM, em que as setas vermelhas representam a dependência entre observações.

Outras formas de modelos Autoregressivos que podem ser mencionados são as combinações com os modelos da seção 4.3.1. Dessa forma, surgem os modelos ARX(p)-HMMs modelos autoregressivos de Markov com estados ocultos com variáveis externas e ARX(p)-NHMMs modelos autoregressivos de Markov com estados ocultos não homogêneos com variáveis externas, tal versão foi empregada por Gelati *et al.*, (2010b).

4.3.3 - Outras extensões da forma básica de um HMM

Outras extensões da forma básica de um HMM, além das mostradas nesta seção são, por exemplo, utilizadas na biologia e econometria, entre outras áreas. Uma dessas versões é quando as observações R_t dependem simultaneamente de S_t e de variáveis externas X_t , o modelo é denominado *Input-Output* HMM. Outra versão de HMMs surge quando se tem várias Cadeias de Markov, esse modelo é conhecido como um HMM Fatorial. Também existe uma versão de HMMs em que as observações dependem quanto de S_t tanto do estado anterior S_{t-1} . O leitor é sugerido a encontrar mais informações sobre esses tipos de HMMs em Bishop (2006) e Zucchini *et al.*, (2016)

Uma outra classe de HMMs a qual foi utilizada por Whiting et al. (2004) e Whiting (2006), é a classe Hierárquica – HMMs Hierárquicos (HHMMs). Esse tipo de HMMs foi introduzido por Fine *et al.*, (1998) mostrando algumas aplicações relacionadas ao reconhecimento de voz e de escrita, similares às aplicações iniciais dos HMMs realizadas por Rabiner (1989). Os HHMMs são processos estocásticos estruturados em camadas multinível, que generalizam os HMMs básicos, tornando cada um dos estados ocultos em um modelo probabilístico “autônomo”, ou seja, cada estado também é um HHMM. Portanto, os estados ocultos de um HHMM emitem sequências, em vez de observações, por uma ativação recursiva de um dos sub estados. A produção de sequencias ocultas segue a mesma estrutura dos HMMs básicos. Nos HHMMs, uma sequência oculta de ultimo nível, nível inferior, da hierarquia é denominada de produção, essa sequência produz as observações que são geradas através de FDPs dependentes. Os estados ocultos que não emitem observações diretamente são chamados de estados internos.(Fine *et al.*, 1998).

5 - METODOLOGIA

A seguir se apresenta uma descrição dos métodos e técnicas que foram utilizadas no presente estudo. Este capítulo está dividido em várias seções que abrangem principalmente: a estratégia metodológica, os dados utilizados, características dos modelos de previsão de vazão empregados e a sua estrutura, tipos de variáveis utilizadas, a estratégia de previsão e as métricas de avaliação da qualidade do desempenho dos modelos.

5.1 - ESTRATÉGIA METODOLÓGICA

A estratégia metodológica utilizada neste estudo é apresentada na figura 5.1. A partir dessa estratégia foram exploradas diferentes metodologias baseadas em Modelos de Markov com Estados Ocultos (HMMs) que utilizam informações climáticas para a previsão de afluências médias anuais.

Dessa forma, utilizaram-se três classes de modelos: I) HMMs com informação climática nas Matrizes de Probabilidade de Transição (MPTs) – Modelos de Markov com estados ocultos Não Homogêneos (NHMMs), II) HMMs com informação climática nas Funções de Densidade de Probabilidade (FDPs) dependentes dos estados ocultos (IHMMs) e III) Modelos Autoregressivos com variáveis exógenas (ARXs) – comumente utilizados para a previsão de afluências no Nordeste do Brasil. Cada tipo de modelo inclui várias subclasses que se comentam a seguir.

1. **HMMs com informação climática nas MPTs (NHMMs).** Possuem duas subclasses:
 - i) Autoregressiva e
 - ii) Não Autoregressiva, em que cada uma delas possui duas subclasses: Log-Normal e Gama. A diferença entre um modelo ser Log-Normal ou Gama não é unicamente a FDP dependente utilizada no modelo, também se dá pelo tipo de variável tratada, já que, modelos “Log-Normais” empregam as vazões log-transformadas e modelos “Gama” utilizam as afluências em escala real ou sem transformação.

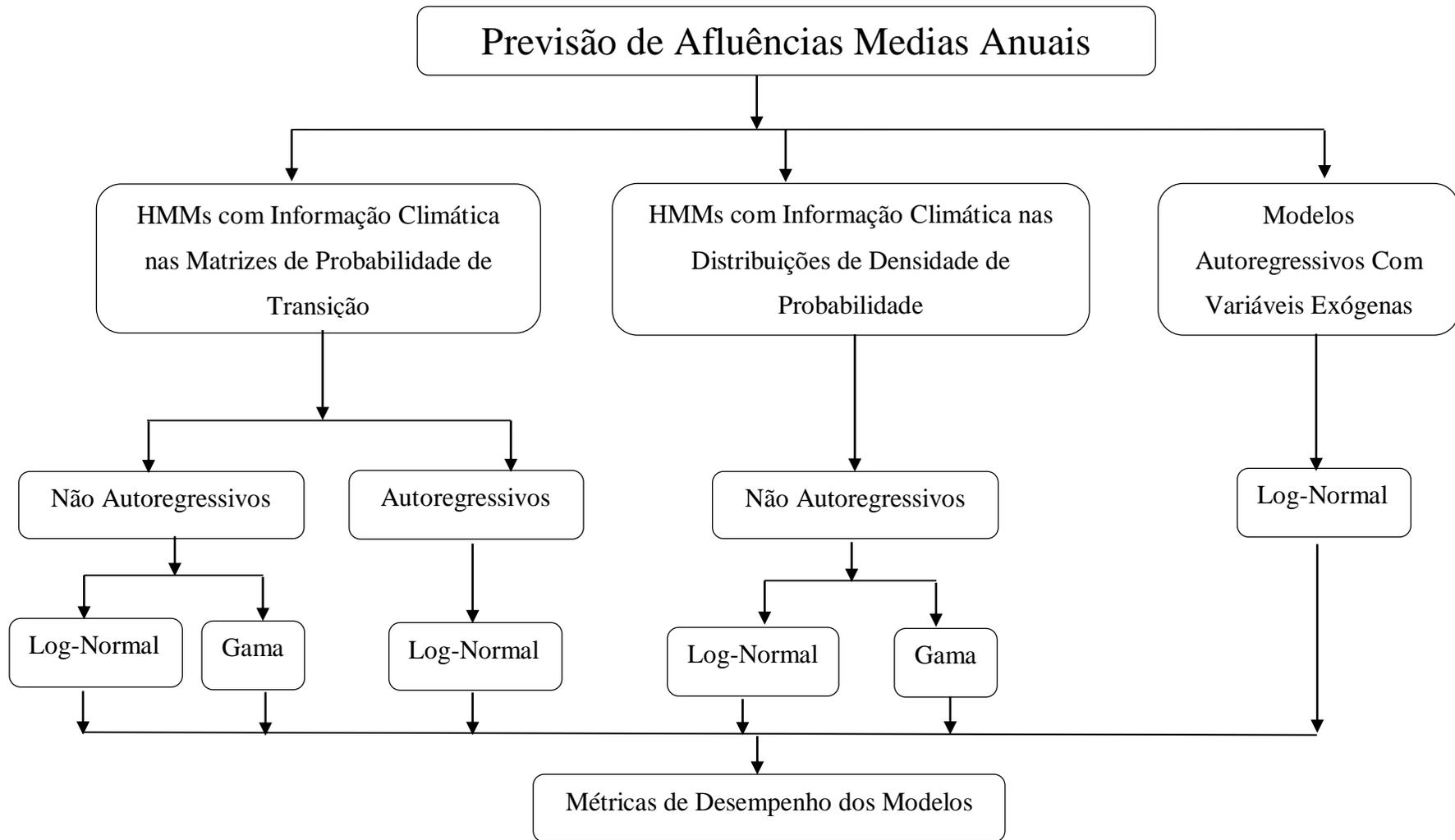


Figura 5.1 – Estratégia Metodológica

2. **HMMs com informação climática nas FDPs (IHMMs).** Esses modelos possuem somente uma classe: i) Não Autoregressiva, nesta têm-se as subclasses Log-Normal e Gama, definidas de igual forma que no item anterior.
3. **ARXs.** São modelos Autoregressivos com informação climática, em que as variáveis dependentes seguem uma distribuição Log-Normal com parâmetros estabelecidos a partir de uma estrutura de autocorrelação e coeficientes de regressão linear que dependem de variáveis exógenas.

Maiores detalhes sobre a estrutura dos modelos de previsão são apresentados na seção 5.3

5.2 - ESTUDO DE CASO: RESERVATÓRIO ORÓS ESTADO DO CEARÁ

Nesta tese foram realizadas previsões de afluências médias anuais, com antecedência mensal, ao reservatório Orós. A ideia foi prever a vazão média anual para esse reservatório antes da ocorrência da quadra chuvosa no Ceará – fevereiro a maio. O Orós está localizado na bacia do rio Jaguaribe no estado do Ceará, Nordeste do Brasil; figura 5.2. Escolheu-se esse reservatório pela sua importância para a região e porque faz parte do sistema de abastecimento de água da Região Metropolitana de Fortaleza (RMF). A RMF é composta por 15 municípios e possui uma população correspondente a 55% da população urbana do Estado de Ceará (ANA, 2010).

A bacia do rio Jaguaribe ocupa aproximadamente o 51.9 %, 6415,10 km², da área total do estado e está localizada entre as coordenadas geográficas 4°3'00''- 7°45'00''S e 37°30'00''- 41°00'00''W. Essa bacia é composta por três sub-bacias: alto, médio e baixo Jaguaribe, mas, envolve também as bacias dos rios Salgado e Banabuiú. O rio Jaguaribe tem uma extensão de aproximadamente 633 km e os seus afluentes mais importantes são: pela margem direita os rios Cariús, Salgado e Figueiredo; e pela margem esquerda o riacho do Sangue e os rios Banabuiú e Palhano, (SRHC, 2005).

O sistema de abastecimento da RMF inclui a transposição do rio Jaguaribe e é denominado sistema Jaguaribe-Metropolitano. A transposição é realizada por meio do Eixo De Integração Castanhão/Região Metropolitana de Fortaleza (EIC/RMF) e o Canal do Trabalhador. De forma geral, a configuração do sistema é a seguinte: o reservatório Orós é

ligado ao reservatório Castanhão e nessa parte começa o EIC/RMF, o qual recebe as águas do reservatório Banabuiú, essa configuração pode ser vista na figura 5.3.

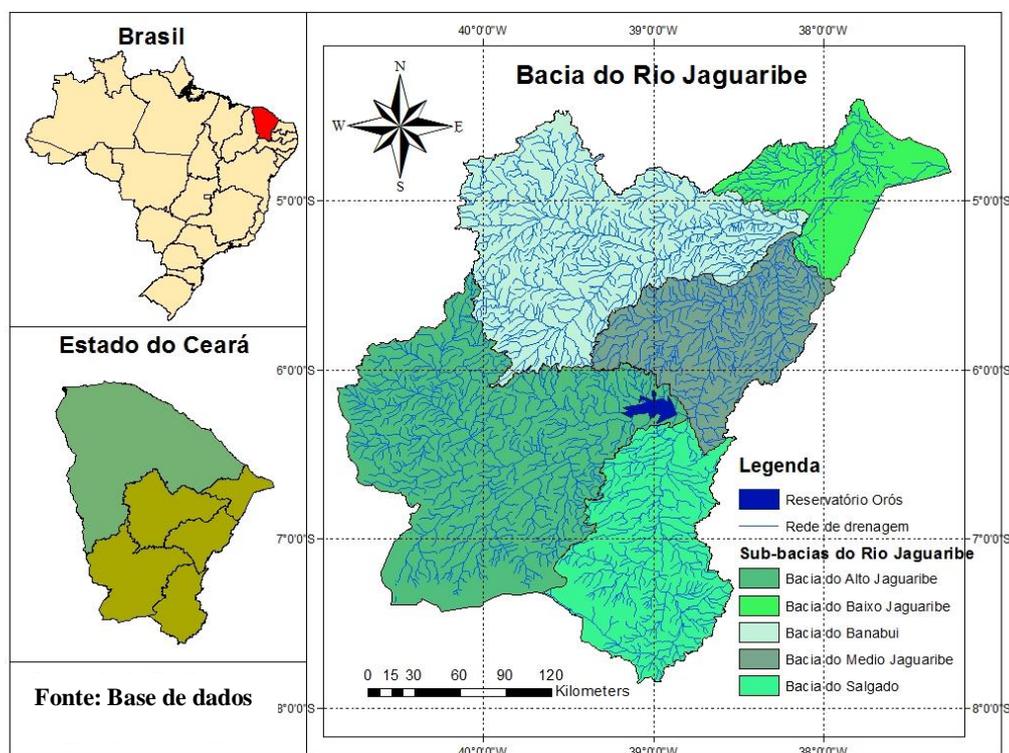


Figura 5.2 – Localização da bacia do rio Jaguaribe e do reservatório Orós

Especificamente, o reservatório Orós está localizado entre os municípios Quixelô, Iguatu e Orós, a aproximadamente 270 km de Fortaleza, em linha reta. O reservatório foi construído em 1961 com o objetivo principal de armazenar água para o período de secas. No entanto, faz parte do Sistema Jaguaribe-Metropolitano (Gatto, 1999; COGERH, 2011).

O açude Orós é o segundo maior reservatório do Estado do Ceará, perdendo apenas para o açude Castanhão. O Orós se localiza próximo ao exutório da bacia do Alto Jaguaribe, pereniza o rio Jaguaribe até sua foz (308 km), tornando-o o maior rio perenizado artificialmente do mundo (SRHC, 2005).

Com essa perenização, também são beneficiadas as bacias do Médio e Baixo Jaguaribe, além da RMF, a qual, através de transposição de bacias, utiliza suas águas desde a construção do Canal do Trabalhador, em 1994. Há duas obras de transposição das águas do açude Orós: a primeira e a mais antiga é a transferência de água para o açude Lima

Campos; a segunda, é a transposição de água da bacia do Jaguaribe para o sistema Pacoti/Riachão/Gavião que abastece a RMF, com vazão máxima aduzida de 5 m³/s, (SRHC, 2005). Algumas características do reservatório Orós são apresentadas na Tabela 5.1.

Tabela 5.1- Características do reservatório Orós (Ribeiro, 2011)

Bacia	Capacidade (hm ³)	Área de drenagem (km ²)
Alto Jaguaribe	1940	24538

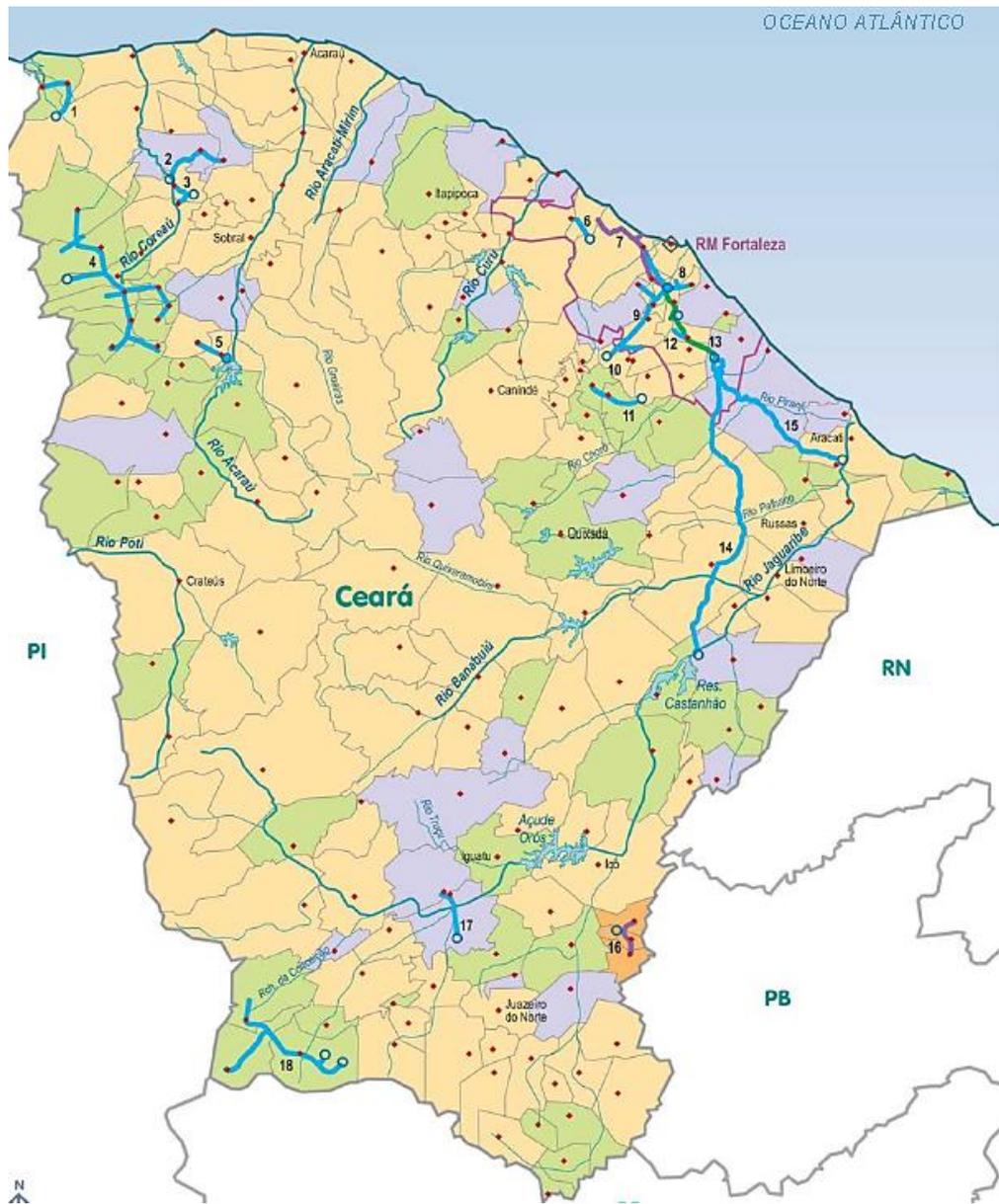
5.2.1 - Dados utilizados

5.2.1.1 - Série de afluições

Para realizar previsões de afluições médias anuais ao reservatório Orós foi utilizada a série de afluições médias anuais, obtida a partir da série de afluições médias mensais correspondentes ao período 1911-2000, a qual foi disponibilizada pela Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME). A série de vazões médias mensais foi obtida a partir do modelo chuva-vazão SMAP *Soil Moisture Accounting Procedure*, com dados observados de chuva do posto pluviométrico Iguatu/CE. Escolheu-se essa série de afluições, pela importância do reservatório Orós e por ser uma série de dados consistentes, sem falhas, e extensa (90 anos).

5.2.1.2 - Indicadores climáticos

Este estudo incluiu informações climáticas em modelos estocásticos para a previsão de afluições, informações que pelo comentado na seção 3.3.1 influenciam a variabilidade e persistência do regime de vazões no Ceará. De tal forma, os índices climáticos foram conformados a partir dos dados disponíveis no site do *International Research Institute for Climate and Society* (IRI) – <http://iri.columbia.edu/resources/>, o qual disponibiliza séries históricas de TSMs, na escala mensal. Os indicadores climáticos utilizados foram o índice NINO3, relacionado ao *El Niño –Southern Oscillation* (ENSO) e Dipolo do Atlântico (DA).



- 1 SIN Chaval-Barroquinha
- 2 SIN Senador Sá-Uruoca
- 3 SIN Coreaú-Moraújo
- 4 SIN Jaburu-Ibiapaba
- 5 SIN Reriutaba-Varijota
- 6 SIS São Gonçalo do Amarante
- 7 SIN Gavião Pecém
- 8 SIN Gavião
- 9 SIN Acarape do Meio
- 10 SIN Acarape-Barreira-Redenção
- 11 SIN Araújoaba-Baturité
- 12 SIN Pacajus-Horizonte-Chorozinho
- 13 Eixo Castanhão - RM Fortaleza
- 14 Eixo Castanhão - RM Fortaleza
- 15 Canal do Trabalhador
- 16 SIN Baixo-Ipimirim-Umari
- 17 SIN Cariús-Jucás
- 18 SIN Araripe-Campos Sales-Salitre

OBRAS PREVISTAS - Cenário 2025

- Conexão a sistema integrado
- Adoção de novo manancial
- Adequação do sistema existente
- Abastecimento satisfatório
- Sistema existente
- Sistema em obras
- Sistema planejado
- Principais aglomerados urbanos



Figura 5.3 – Sistemas de distribuição de água no Estado do Ceará (ANA 2010).

5.3 - MODELOS ESTOCÁSTICOS PARA A PREVISÃO DE AFLUÊNCIAS UTILIZADOS

Os modelos estocásticos utilizados descreveram-se brevemente na estratégia metodológica. No entanto, na seção 5.3.1 se apresenta uma definição mais detalhada desses modelos e na seção 5.3.2 se mencionam alguns critérios úteis para a seleção de modelos de previsão.

5.3.1 - Definição e estruturação dos modelos de previsão de vazão

A ordem (m) dos HMMs que foram utilizados neste trabalho se limitou em dois e três estados ocultos, valores geralmente encontrados na literatura para a modelagem de séries de vazões (veja, por exemplo, as ordens dos HMMs nas tabelas A.1 e A.2), embora, tenha-se ajustado modelos de até 4 estados ocultos. Essa limitante de estados está baseada no fato de modelos de quatro ou mais estados apresentarem um grande número de parâmetros, o que aumenta a incerteza nas previsões, e também no fato de um maior número de estados ocultos dificultar uma interpretação física dos mesmos. Além disso, modelos de mais de quatro estados requerem maior gasto computacional e o número de informações para a análises aumenta. Logo, tomando essas considerações e devido ao foco exploratório desta tese foi pertinente reduzir as análises a modelos de dois e três estados ocultos. As distribuições dependentes dos estados ocultos foram a Log-Normal e Gama, a seleção dessas distribuições foi baseada na literatura (veja as tabelas A.1 e A.2 no apêndice A).

Com relação à ordem (p) dos modelos AR, foi utilizada a metodologia de Box e Jenkins (1976) para determinar a melhor ordem desses modelos.

A configuração interna dos modelos de previsão de afluências que se utilizou foi a seguinte:

- I. HMMs com informação climática nas MPTs.
 - i. HMMs Autoregressivos e Não Homogêneos ($AR(p)$ -NHMMs (m)): são NHMMs de ordem m , em que um modelo Autoregressivo é quem realiza o papel de FDP, adicionando dependência entre as observações. Nesta classe, os processos AR são modelados na forma Log-Normal ($AR(p)$ -NL (m)) e

utilizam as afluências observadas log-transformadas. Esse tipo de modelos seguem as propriedades mostradas na seção 4.3.2.

- ii. HMMs Não autoregressivos e Não Homogêneos (NHMMs (m)): são HMMs de ordem m , os quais correspondem a modelos NHMMs, segunda classe de HMMs apresentada na seção 4.3.1.2, e incluem duas formas determinadas pelas FDPs dependentes: Log-Normal (NL (m)) e Gama (NG (m)). Sendo que os NL utilizam as afluências observadas log-transformadas e os NG as empregam sem transformar.

II. HMMs com informação climática nas FDPs:

- i. HMMs Não Autoregressivos (IHMMs (m)): esses modelos são HMMs nos quais as FDPs são influenciadas por indicadores climáticos e tem a forma Log-Normal (IL (m)) e Gama (IG (m)), em que as FDPs seguem MLGs Log-Normais e Gama, respectivamente.

III. Modelos ARX

- i. Modelos Autoregressivos com variáveis exógenas (ARX (p)): são modelos ARX da forma Log-Normal.

Uma ferramenta comumente utilizada em estudos de previsão de variáveis hidrometeorológicas, precipitação, vazão, temperatura, entre outras, que permite avaliar o desempenho de modelos de previsão é a climatologia. Essa ferramenta não é propriamente um modelo de previsão, portanto, não foi incluída na estratégia metodológica. A climatologia corresponde a um modelo no qual a variável de interesse segue uma distribuição empírica de probabilidades de ocorrência, ou seja, é um histórico que representa as condições médias da variável de interesse, tal histórico pode contemplar um período fixo ou variável de tempo, comumente um período de 30 anos. Nesta tese a climatologia utilizada foi um período de 30 anos, fixo no início da série observada, período 1911-1940.

Um resumo da nomenclatura e estrutura desses modelos de previsão é apresentado na tabela 5.2. e na figura 5.4.

Tabela 5.2 – Resumo da estruturação dos modelos de previsão

Modelo	Subclasse		Abreviação	Equação da Verossimilhança	Variáveis Utilizadas	Informação Climática
I	Autoregressivos	Log-Normal	AR (p)-NL (m)	4.56	Série de afluências Q_t e série de indicadores climáticos X_t	NINO3, Dipolo do Atlântico
	Não Autoregressivos	Log-Norma	NL (m)	4.43		
Gama		NG (m)				
II	Não Autoregressivos	Log-Normal	IL (m)	4.30		
		Gama	IG (m)			
III	Log-Normal		ARX (p)	Ver Kwon (2002), página 51		
Climatologia*-				Ver Naghettini e Pinto, (2007), página 24	Série de afluências Q_t	Não utiliza

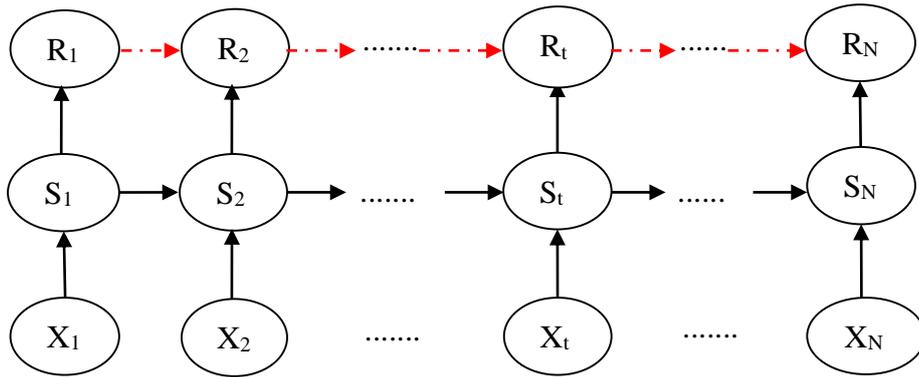
*Não é um modelo de previsão. É utilizada para efeitos de comparação

5.3.1.1 - Ajuste dos modelos de previsão e procedimento de previsão

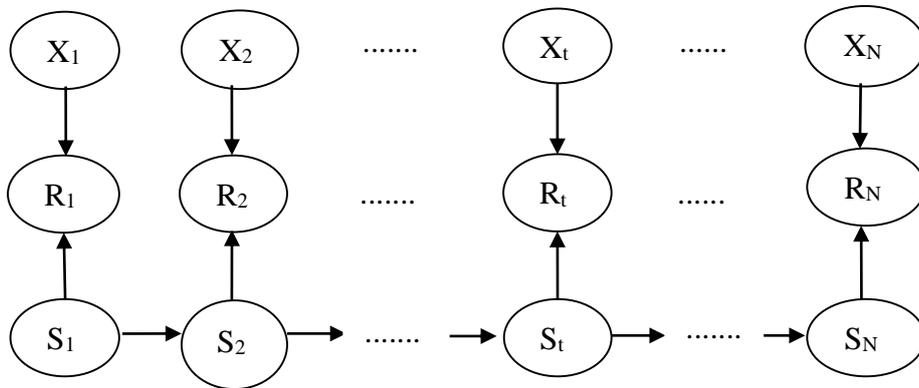
Para realizar previsões é necessário definir um período de calibração e outro de validação. Logo, primeiramente foi selecionado um período de calibração de 30 anos. Com essa série inicial se ajustaram os modelos estocásticos de previsão e realizadas previsões para os 60 anos restantes da série. Os passos do ajuste do modelo até a previsão seguem o fluxo da figura 5.5. Como pode se observar nesse fluxograma o processo de previsão se assemelha a um sistema de previsão real, o que o torna um sistema com potencial uso.

Na figura 5.5, observa-se que com os dados de entrada, séries de afluências Q_t e de indicadores climáticos X_t , e já definido o período de calibração, se têm dois caminhos a seguir dependendo da variável a ser analisada. Em cada caminho, forma-se um conjunto de trabalho que inclui a nova série R_t de variáveis independentes e variáveis exógenas independentes X_t . Na sequência se faz uma análise de estatísticas amostrais de R_t , para os períodos de calibração e validação, e se avalia a correlação entre a série R_t , e a série X_t . Depois disso, passa-se a escolher o tipo de modelo, para logo realizar o ajuste e, na parte final, realizam-se as previsões para um horizonte de previsão que nesta tese foi igual a 1 (um). Por fim, avalia-se o desempenho dos modelos.

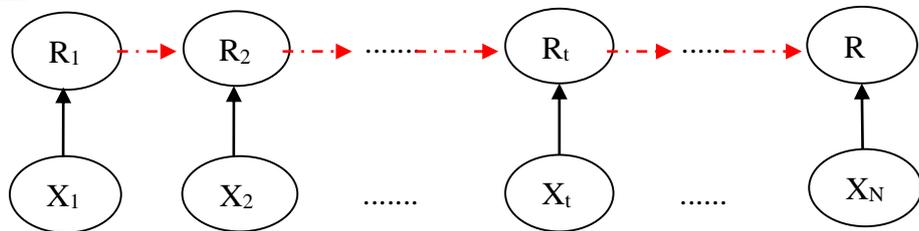
Classe I



Classe II



Classe III



Legenda:

R_t : Série de afluências observadas, que seguem FDPs dependentes dos estados ocultos Log-Normal ou Gama.

S_t : Série de estados ocultos, os estados ocultos tomam valores de 1 até o número de estados ocultos (m).

X_t : Série de variáveis exógenas, série de indicadores climáticos que influenciam as afluências.

Figura 5.4 – Modelos utilizados para a previsão de afluências. Classe I (superior): HMMs com informação climática nas Matrizes de Probabilidade de Transição de estados ocultos – NHMMs e AR-NHMMs; Classe II (meio): HMMs com informação climática nas FDPs IHMMS; Classe III (inferior): Modelos Autoregressivos com variáveis exógenas ARX. As setas vermelhas indicam uma autocorrelação de primeira ordem.

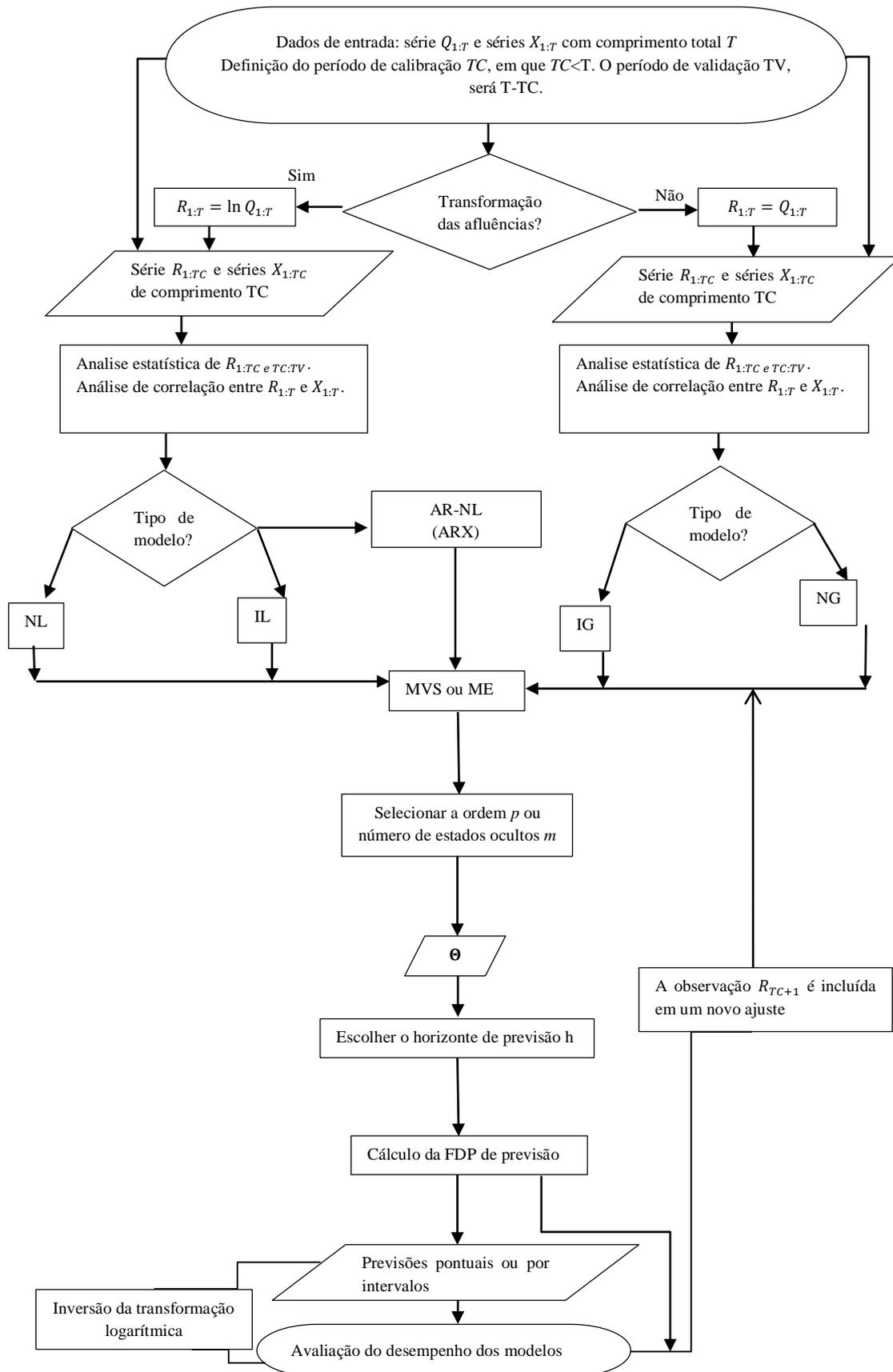


Figura 5.5 – Fluxograma para o ajuste dos modelos e cálculo de previsões

Ainda na figura 5.5, MVS corresponde ao método da máxima verossimilhança e ME o método da Maximização da Expectativa, os quais são os métodos de estimação do vetor de parâmetros $\Theta = (\delta, \lambda, \Gamma, \sigma, \rho, \theta = [\mu, \sigma, ou, \alpha, \beta])$, discutidos no Anexo A. Nesta tese a estratégia para modelar as probabilidades de transição entre estados ocultos foi a utilizada por Robertson *et al.*, (2004) e Kirshner (2005).

No caso dos IHMMs os MLGs seguem as equações da seção 4.3.1.1, sendo que no caso do MLG Gama foi utilizada a relação logarítmica.

5.3.2 - Critérios de seleção do número de estados ocultos em HMMs e da ordem em modelos AR

No Caso dos modelos da família ARMA, a seleção do modelo se refere à escolha das ordens p e q e comumente se emprega a metodologia de Box e Jenkins (1976) para tal fim. Nos HMMs, a escolha do modelo concerne à seleção do número ótimo de estados ocultos m e à eleição das FDPs dependentes desses estados.

A metodologia de Box e Jenkins (1976) inclui o *Akaike Information Criterion* (AIC) (Akaike, 1974) e o *Bayesian Information Criterion* (BIC) (Schwarz, 1978) como critérios para selecionar a ordem em modelos do tipo ARMA, esses critérios também são utilizados para os HMMs (Zucchini e MacDonald, 2009). O BIC é mais severo que o AIC, já que penaliza a modelos com muitos parâmetros, por esse motivo apresenta maior preferência na literatura. O melhor modelo é aquele que apresenta os menores valores de AIC e BIC, esses critérios se definem como:

$$AIC = -2 \ln(L_N) + 2k \quad (5.1)$$

$$BIC = -2 \ln(L_N) + k \ln(N) \quad (5.2)$$

em que:

k = Número de parâmetros do modelo. Nos HMM, k é o número de parâmetros independentes;

L_N = Valor obtido de maximizar a função de máxima verossimilhança;

N = Número de observações.

5.4 - MÉTRICAS DE COMPARAÇÃO DO DESEMPENHO DOS MODELOS DE PREVISÃO DE VAZÃO

Nesta seção são apresentadas as métricas que permitem analisar a qualidade de previsões de afluições. Também, são apresentadas algumas definições que permitem analisar a persistência de anos seguidos de vazões altas ou baixas, segundo um limiar de referência. Embora, uma análise detalhada de persistência não foi realizada neste trabalho.

5.4.1 - Métricas de qualidade de previsões

5.4.1.1 - Avaliação da qualidade de previsões probabilísticas

O *Rank probability skill score* (RPSS) é uma métrica que mede a qualidade de previsões probabilísticas no caso discreto. O RPSS compara o valor da métrica *Ranked Probability Score* (RPS) do modelo de previsão com o RPS da climatologia. No caso contínuo é utilizado o *Continuous Ranked Probability Score* (CRPS) (Wilks, 2006).

O CRPS resume qualidade de uma previsão probabilística contínua a um único valor. Em outras palavras, ela mede a integral do quadrado das diferenças entre a distribuição acumulada de probabilidade (FAP) das probabilidades previstas e a FAP das probabilidades observadas. O CRPS é definido como:

$$\text{CRPS} = \int_{-\infty}^{\infty} [F(r) - F_o(r)]^2 dr \quad (5.3)$$

$$F_o(r) = \begin{cases} 0, & r < \text{valor observado} \\ 1, & r > \text{valor observado} \end{cases} \quad (5.4)$$

em que $F_o(r)$ é a FAP das probabilidades observadas, a qual é igual a zero para valores abaixo da vazão observada e um (1) para valores iguais ou maiores à observação e $F(r)$ representa a FAP das probabilidades previstas.

A figura 5.6a mostra um exemplo de três FDPs de previsão em relação a um valor observado. Enquanto que, a figura 5.6b mostra as FAPs de previsão junto com a função de

passo, $F_o(r)$, linha em negrito. Nessa figura, a distribuição 1 produz um valor baixo, aceitável, do CRPS, porque sua FAP é uma razoável aproximação da função de passo. A distribuição 2 concentra a probabilidade longe da observação e a distribuição 3 é rejeitada por falta de forma, mesmo esteja centralizada na observação. O CRPS pode ser calculado como um valor médio para todos os anos do período de validação como:

$$CRPS = \frac{1}{n} \sum_{i=1}^n CRPS_i \quad (5.5)$$

em que n é o número de anos do período validação e $CRPS_i$ é o valor da métrica no ano i .

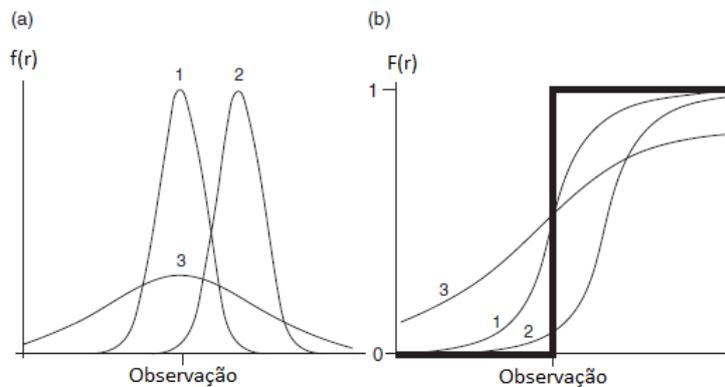


Figura 5.6 – Ilustração do *Continuous Ranked Probability Score* (Wilks 2006)

Por fim, a métrica utilizada nesta tese foi o *Continuous Probability Ranked Skill Score* (CRPSS) – que equivale ao CRPS com relação ao CRPS da climatologia:

$$CRPSS = 1 - \frac{CRPS_{Previsão}}{CRPS_{climatologia}} \quad (5.6)$$

se o valor do CRPSS é igual a 1 o modelo de previsão realiza uma previsão perfeita, um valor igual a 0 indica que não existe diferença entre a previsão realizada pelo modelo e a climatologia. Por fim, um valor negativo indica que a previsão realizada pelo modelo é pior que a calculada pela climatologia. Similarmente ao CRPS, pode-se calcular um CRPSS médio para o período de validação utilizando uma equação similar à 5.5.

5.4.1.2 - Avaliação da qualidade de previsões pontuais

Outra comparação do desempenho dos modelos de previsão que se realizou nesta tese foi a partir de várias métricas que quantificam a qualidade de previsões pontuais. Essas métricas são utilizadas pelo Operador Nacional do Sistema Elétrico (ONS) e empregaram-se nos trabalhos de Alexandre (2012), Ribeiro (2011) e (Oliveira e Lima. 2016). Tais métricas são:

Erro Médio Absoluto

$$EMA = \frac{1}{nt} \sum_{i=1}^{nt} |QO_t - QP_t| \quad (5.7)$$

Erro Médio Percentual Absoluto

$$EMPA = \frac{1}{nt} \sum_{i=1}^{nt} \frac{|QO_t - QP_t|}{QO_t} \times 100 \quad (5.8)$$

Coefficiente de Eficiência de Nash-Sutcliffe – NS

$$NS = 1 - \frac{\sum_{i=1}^{nt} (QP_t - QO_t)^2}{\sum_{i=1}^{nt} (QO_t - \overline{QO_t})^2} \quad (5.9)$$

Distância Multicritério

$$DM = \sqrt{(1 - NS)^2 + (EMPA)^2} \quad (5.10)$$

Coefficiente de Correlação

$$CORREL = \frac{Cov(QO_t, QP_t)}{\sigma_{QO_t} \cdot \sigma_{QP_t}} \quad (5.11)$$

em que:

EMA = erro médio absoluto;

$EMPA$ = erro médio percentual absoluto;

NS = coeficiente de eficiência de Nash-Sutcliffe;

DM = distância multicritério;

$CORREL$ = coeficiente de correlação entre as vazões observadas e previstas;

$Cov(QO_t, QP_t)$ = covariância entre as vazões observadas e previstas no intervalo de tempo t ;

QP_t = vazão prevista no intervalo de tempo t ;

QO_t = vazão observada no intervalo de tempo t ;

nt = número de intervalos de tempo;

σ_{QO_t} = desvio padrão das vazões observadas no intervalo de tempo t ;

σ_{QP_t} = desvio padrão das vazões previstas no intervalo de tempo t

$\overline{QO_t}$ = média da vazão observada no intervalo de tempo t

O EMA é a média aritmética dos valores absolutos da diferença entre observações e previsões. Quando um EMA é igual a zero indica previsões perfeitas. O valor do EMA incrementa quando as discrepâncias entre observações e previsões se tornam maiores. Logo, o EMA pode ser interpretado como a magnitude do erro de previsão em um conjunto de previsões referentes a um período de verificação das mesmas (Wilks, 2006)

O EMAP corresponde à média dos erros relativos das previsões, em relação às observações para um período de verificação. O EMAP é uma estatística não viesada que mede a capacidade preditiva de um modelo expressada em percentagem (Wang *et al.*, 2009). Tsai (2016) menciona que segundo o EMAP a acurácia das previsões realizadas pelo modelo de previsão pode ser: alta ($EMAP \leq 10$); boa ($10 \leq EMAP \leq 20$); viável ou factível ($20 \leq EMAP \leq 50$); e baixa ($EMAP \geq 50$).

O coeficiente NS corresponde a uma estatística normalizada que determina a magnitude relativa da variância residual em comparação com a variância dos dados observados. Os valores do NS variam entre $-\infty$ e 1, em que: $NS = 1$ corresponde a previsões perfeitas; $NS=0$ mostra que as previsões são iguais ao valor médio observado; e $NS < 0$ mostra um desempenho inaceitável do modelo utilizado, já que, o valor médio observado é um melhor modelo de previsão (Wang *et al.*, 2009).

Por fim, a *DM* foi proposta para evitar interpretações conflitantes e representa a distância euclidiana ao ponto ótimo, no caso (0,0), tal ponto indica previsões perfeitas (Oliveira e Lima, 2016).

5.4.2 - Análise de anos secos e úmidos

Neste estudo, analisou-se como os modelos de previsão conseguem prever anos “mais” secos ou “mais” úmidos com relação a limiares de referência. Para esse fim, utilizaram-se algumas definições, figura 5.7, empregadas na literatura para analisar a persistência de períodos hidro-climáticos (veja, por exemplo, Sveinsson *et al.*, 2003 e Prairie *et al.*, 2008). Nessa Figura, períodos secos e úmidos são definidos quando as vazões são inferiores ou superiores à mediana histórica, respectivamente.

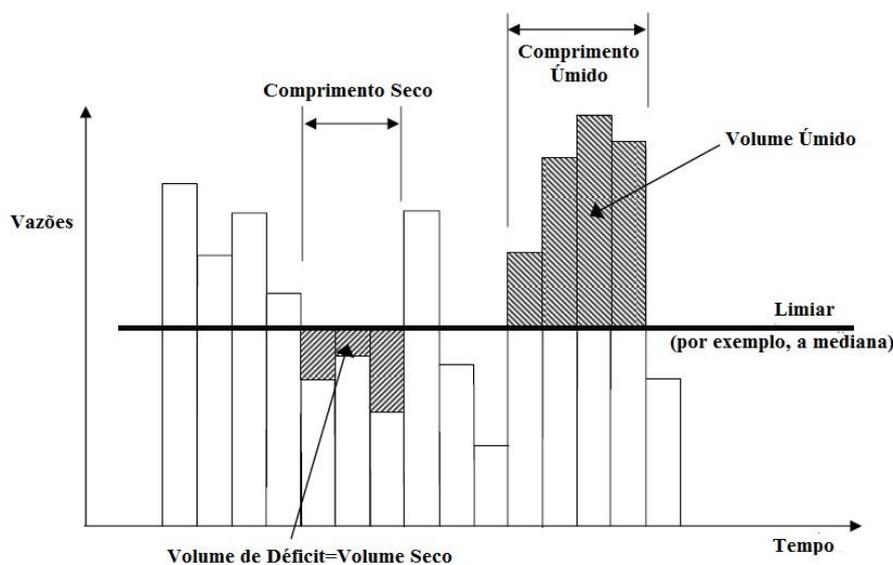


Figura 5.7 – Estatísticas de períodos secos e úmidos

5.5 - VALIDAÇÃO CRUZADA

Geralmente, deseja-se estimar a acurácia do modelo de previsão antes de ser aplicado. Porém, as estimativas de desempenho tendem a ser muito otimistas, porque no ajuste do modelo se escolhem parâmetros que “adaptam” o modelo aos dados da melhor forma possível (Davis, 1976). Esse fenômeno é chamado habilidade artificial e preocupa quando se têm amostras pequenas. Uma forma de evitar esse fenômeno é dividir os dados em

conjuntos de “ajuste” e “validação”, que servem para ajustar e validar (testar) o modelo, respectivamente (Storch e Zwiers, 1999).

Idealmente, quando existem dados suficientes, retira-se da amostra um conjunto de validação, de tamanho suficientemente grande e igual ao de ajuste, o qual é utilizado para avaliar o desempenho do modelo de previsão. Na prática, os dados costumam ser escassos, logo, essa separação geralmente não é possível. Para resolver esse problema, a validação cruzada *K-fold* usa parte dos dados disponíveis para ajustar o modelo e uma parte diferente para testá-lo (Hastie et al. 2009). Assim, dividem-se os dados em K partes aproximadamente iguais; por exemplo, quando $K = 5$, tem-se a configuração da figura 5.8. Portanto, para as outras $k-1$ partes dos dados se ajusta o modelo e para a k -ésima parte (parte – vermelha na figura 5.8) se realizam previsões, logo, calcula-se alguma métrica de desempenho das previsões realizadas pelo modelo ajustado. O procedimento se realiza para $k = 1, 2, \dots, K$, no final uma média das K métricas de desempenho é calculada.



Figura 5.8 – Exemplo de validação cruzada *K-fold* quando $K=5$

No entanto, Bergmeir *et al.* (2018) mencionam que quando se trata de previsão em séries temporais, o método *K-fold* parece dar a sensação de que se utilizam dados futuros para prever o passado. Além disso, a correlação serial nos dados, juntamente com possíveis não estacionaridades, faz com que esse método pareça problemático, uma vez que não leva em conta essas questões. Em geral, em séries temporais, se utiliza o método de validação cruzada *out-of-sample*, figura 5.9, em que uma parte final da série é retida para avaliação. Desta forma, apenas uma avaliação é realizada, no entanto, para um conjunto de dados pequeno essa avaliação torna-se problemática. Neste estudo, os métodos *K-fold* e *out-of-sample* foram utilizados.

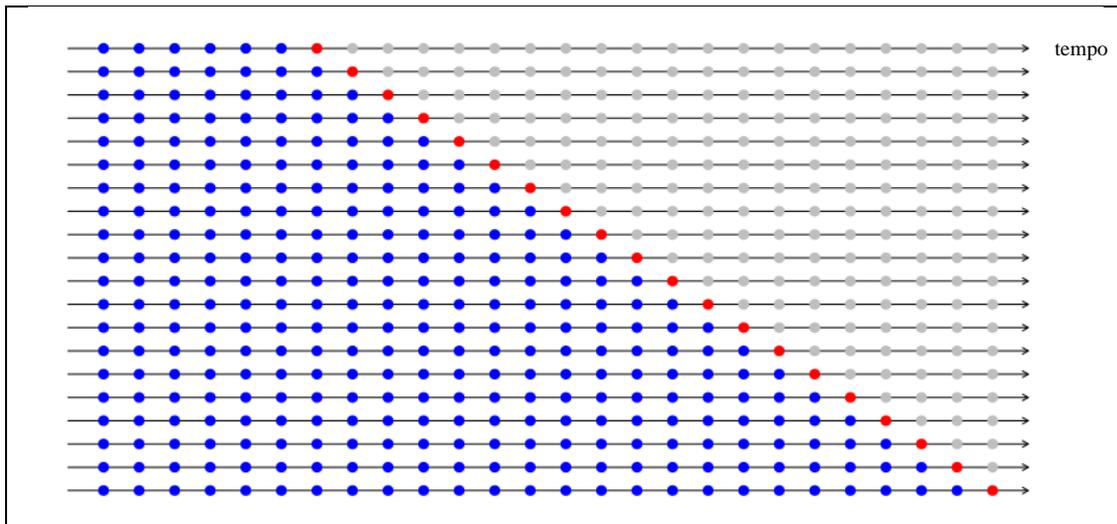


Figura 5.9 – Validação cruzada *Out - of - Sample* (Hyndman, 2019)

5.5.1 - Validação cruzada *K-fold* para HMMs

Para utilizar o método *K-fold* para HMMs, aparentemente pela sua estrutura de relação entre observações podem existir problemas no ajuste quando uma parcela de dados é retirada. No entanto, pode ser utilizado o procedimento de ajuste quando existem dados faltantes na série o qual é descrito por Zucchini and MacDonald (2009). Nesse caso, as probabilidades $p_i(r_t)$, da observação r_t faltante no tempo t , dependentes do i -ésimo estado oculto são substituídas por um (1) para todos $i = 1, 2, \dots, m$ estados ocultos. No caso de NHMMs nos quais as MPT são influenciadas por variáveis externas, tais variáveis existem e podem ser utilizadas no processo de ajuste desses modelos.

Por outro lado, segundo a equação 4.29, precisa-se do conjunto de observações passadas para realizar as previsões utilizando HMMs. Logo, para não empregar dados futuros no processo de previsão quando utilizado o método *K-fold*, optou-se por tomar só os dados anteriores ao k -ésimo conjunto extraído para validação. Logo, segundo a figura 5.8 as observações utilizadas nessa equação sempre chegam até o tempo anterior ao tempo em que começa o conjunto k .

6 - RESULTADOS E DISCUSSÃO

6.1 - CARACTERÍSTICAS DA SÉRIE OBSERVADA.

A seguir são apresentados alguns gráficos, estatísticas descritivas, autocorrelação, ajuste a distribuições da série de afluições médias anuais ao reservatório Orós completa, período 1911-2000. Como mencionado na seção 5.2 essa série foi obtida a partir das vazões médias mensais, resultantes de um modelo chuva-vazão. A série completa é mostrada na figura 6.1 e algumas estatísticas descritivas são apresentadas na tabela 6.1. Observa-se, nessa figura, vários períodos com vazões inferiores “baixas” com relação, por exemplo, à média, também se observa variabilidade das vazões ao longo do tempo.

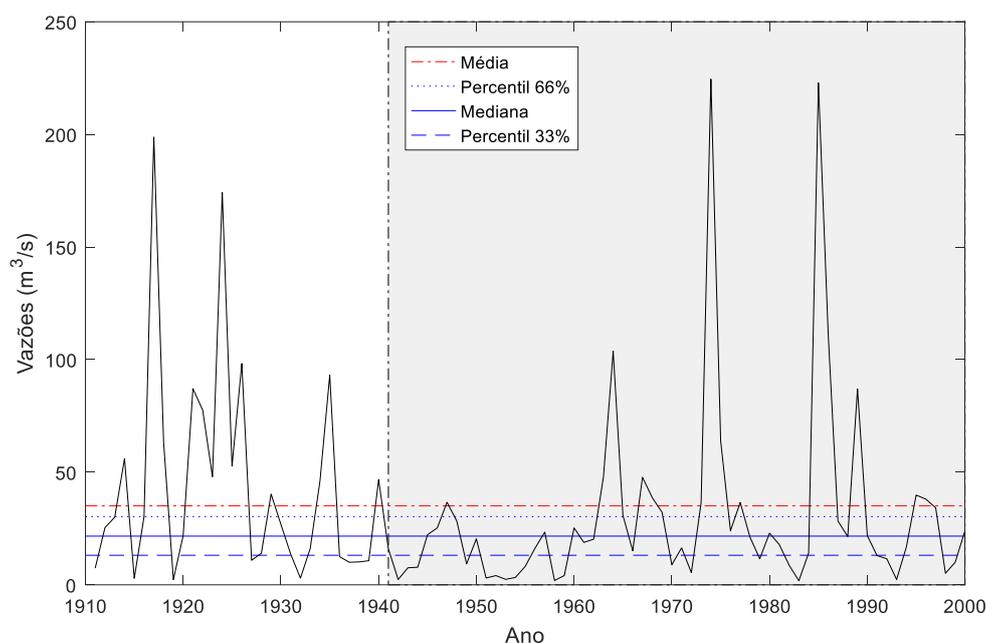


Figura 6.1 – Série de vazões afluentes ao reservatório Orós, período 1911-2000. O período de validação, 1941-2000, está marcado em cinza. Observe a frequência de vazões “baixas” para vários anos.

As figuras 6.2 e 6.3 apresentam respectivamente o histograma e diagrama de caixa da série. Observa-se uma alta frequência de vazões baixas, menores que 25 m³/s, e assimetria positiva. A série também apresenta um alto coeficiente de variação, tabela 6.1. Essas características são comuns em séries de vazões dos rios da região.

Tabela 6.1 – Estatísticas descritivas da série de vazões observada completa.

Estatística Amostral	Período Total (1911-2000)	
	Notação	Valor
Média (m ³ /s)	Q _m	35,059
Mediana (m ³ /s)	Q _{md}	21,625
Vazão Máxima (m ³ /s)	Q _{max}	224,550
Vazão Mínima (m ³ /s)	Q _{min}	1,810
Amplitude (m ³ /s)	A	222,740
Primeiro Quartil (m ³ /s)	Q ₁	10,020
Terceiro Quartil (m ³ /s)	Q ₃	38,700
Percentil do 33 % (m ³ /s)	P ₃₃	13,118
Percentil do 66 % (m ³ /s)	P ₆₆	30,283
Amplitude Inter – Quartis (m ³ /s)	AIQ	28,680
Amplitude Inter – Percentis (m ³ /s)	AIQP	17,165
Variância (m ³ /s) ²	S _Q ²	1983,400
Desvio padrão (m ³ /s)	S _Q	44,535
Coefficiente de Assimetria	Q _{Cs}	2,729
Coefficiente de Variação	CVQ	1,270
Coefficiente de autocorrelação no lag 1	r ₁	0,243

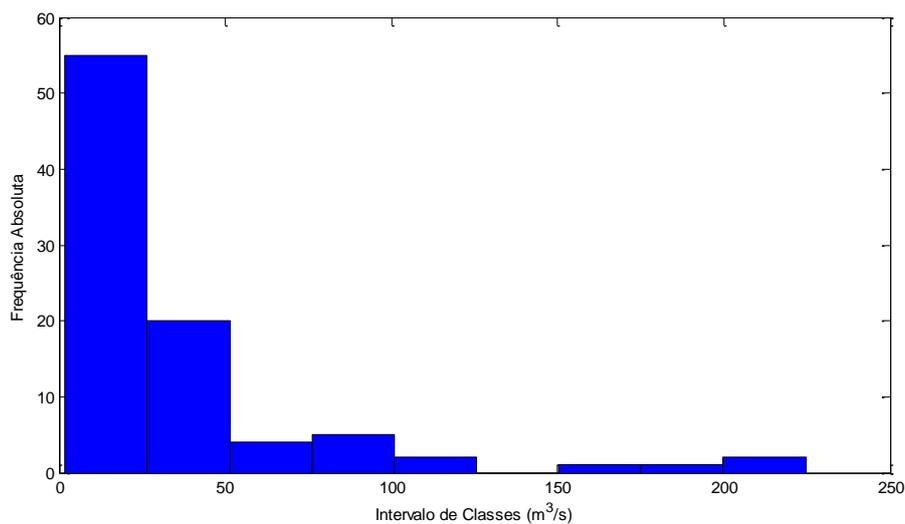


Figura 6.2 – Histograma das Afluências Observadas

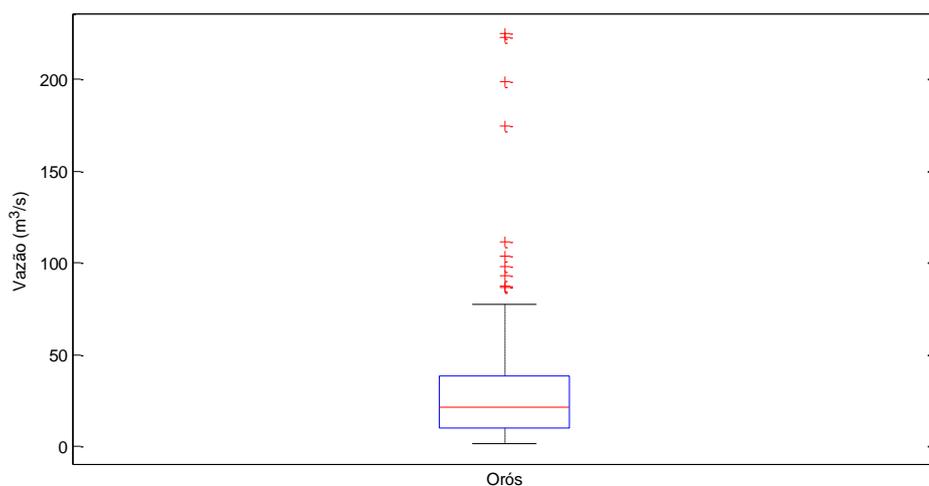


Figura 6.3 – Diagrama de caixa das afluições observadas

Com relação à autocorrelação da série existe uma autocorrelação significativa de primeira ordem, figura 6.4, superior. Existe também uma alta autocorrelação parcial na defasagem 1, mostrando que a série pode ser modelada por um modelo AR(1), figura 6.4 inferior. Embora, a figura 6.4 seja para a série observada completa, dado o processo de previsão, foi verificado para cada conjunto de ajuste começando pelo período 1911-1940, e aumentando em um ano, que pode ser utilizada a ordem 1, já que, sempre foi significativa. Em alguns casos, apareceu a ordem 2, porém ao verificar os valores dos critérios AIC e BIC os valores entre ordens 1 e 2 não eram muito diferentes, logo, por parcimônia foi escolhida a ordem 1.

Como as FDPs dependentes dos estados ocultos escolhidas foram Log-Normal e Gama, a tabela 6.2 mostra os parâmetros obtidos do ajuste da série a essas distribuições. Para a distribuição Normal, figura 6.5, um diagrama quantil-quantil mostra visualmente que a série não se ajusta adequadamente à distribuição Normal. Vários testes de hipótese de ajuste foram realizados, tabela 6.3, para verificar o ajuste a essas distribuições.

No caso da distribuição Gama, a figura 6.6 compara a FAP empírica e a FAP da distribuição Gama, nota-se que existe um bom ajuste da série observada à distribuição Gama, em concordância com os resultados dos testes de aderência da tabela 6.3. Igualmente ao caso da ACF e PACF, foi verificado esse ajuste para os períodos variáveis de calibração, nesse caso os mesmos resultados da figura 6.6 foram encontrados. Vale ressaltar que, embora as figuras desta seção apresentem os resultados para a série completa, esses resultados foram verificados

para todos os conjuntos de ajuste, os quais aumentam ano a ano dentro do período de validação, veja figura 5.5.

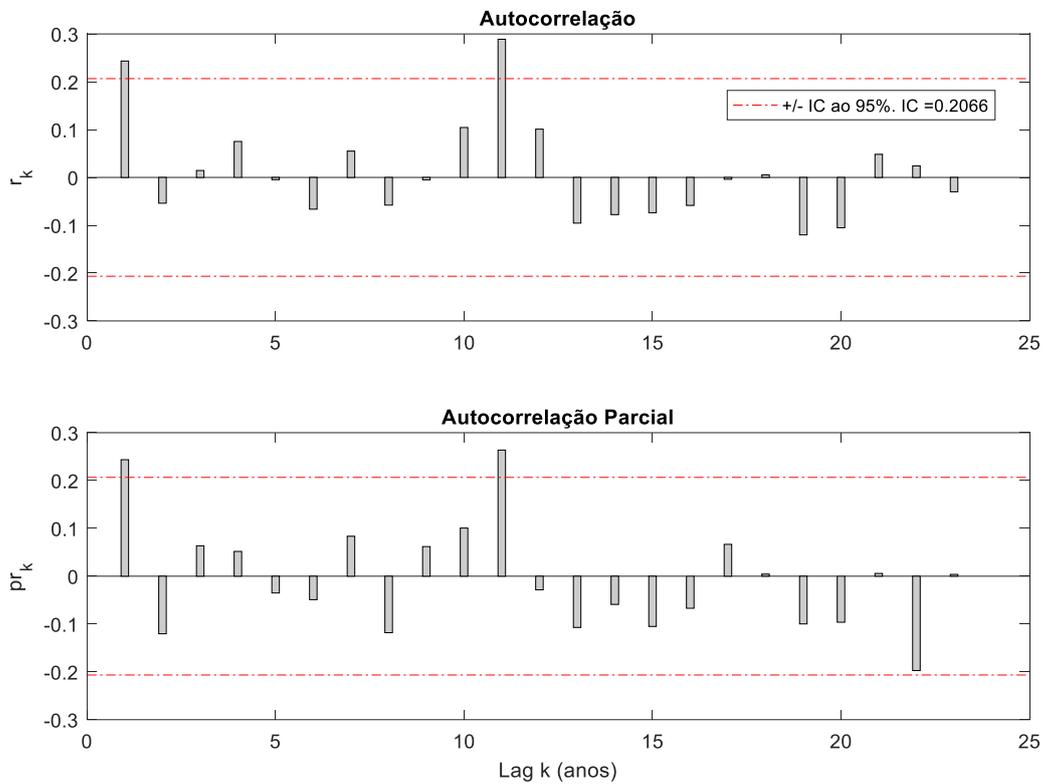


Figura 6.4 – Função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) da série observada. Período 1911-2000.

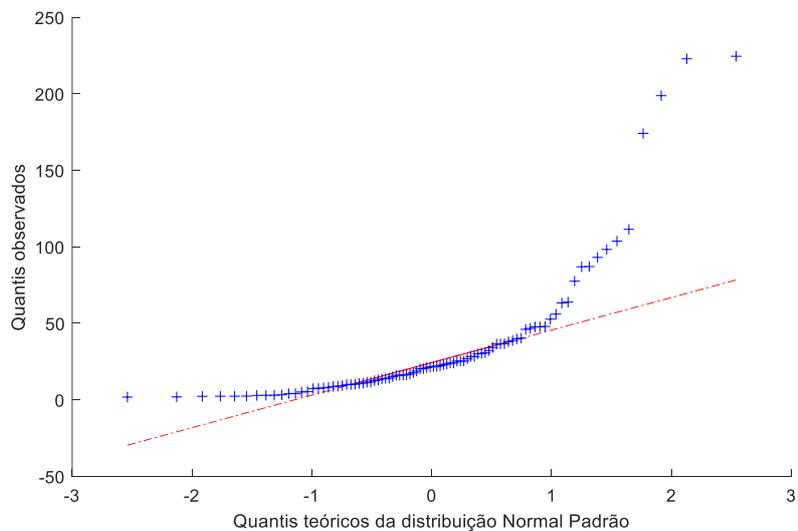


Figura 6.5 – Gráfico quantil-quantil das vazões observadas. Período 1911-2000.

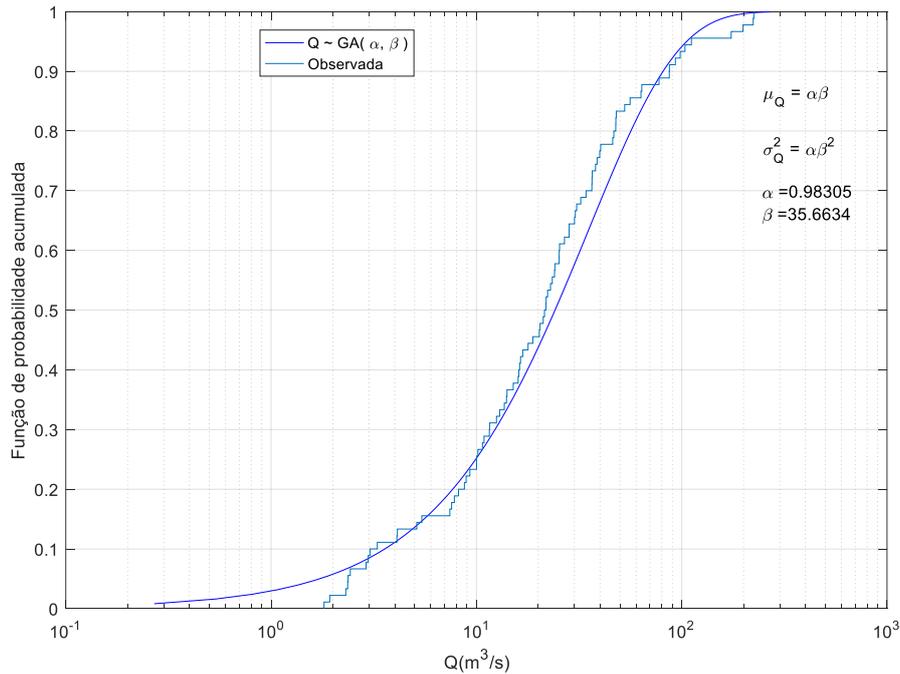


Figura 6.6 – Função de probabilidade acumulada Gama das vazões observadas. Período 1911-2000.

Devido a não distribuição normal da série, resultados da tabela 6.3, avaliou-se a Normalidade da série Log-transformada, série IQ. Os resultados apresentados na figura 6.7 e tabelas 6.3 mostram que existe a aderência da série IQ à distribuição Normal, quando os testes de Shapiro Wilk e Anderson-Darling foram empregados, valores p iguais a 0,206 e 0,280, respectivamente.

Quando o teste Kolmogorov-Smirnov foi aplicado essa aderência não foi confirmada, valor p 1,43 E-54, talvez porque há valores que não se ajustam adequadamente ao começo da linha reta da figura 6.7, já que, esse teste tem como base a diferença máxima entre as funções de probabilidades acumuladas, empírica e teórica (Naghetini e Pinto 2007). Embora isso tenha acontecido, e haja discussões sobre o assunto, tabelas A.3 e A.4, decidiu-se assumir que a série IQ distribui normalmente devido aos resultados dos outros três testes e porque na literatura comumente isso é assumido.

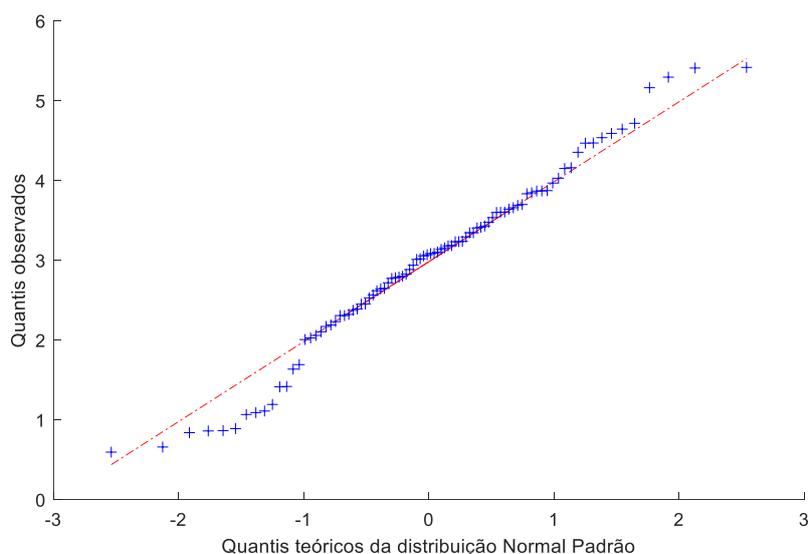


Figura 6.7 – Gráfico quantil-quantil das vazões observadas log-transformadas. Período 1911-2000.

Tabela 6.2 – Parâmetros e intervalos de confiança de 95% estimados para a distribuição Normal e Gama das vazões observadas, Q, e Log-transformadas, lQ. Período 1911-2000.

Parâmetro	Distribuição	Valor	Intervalos de Confiança de 95%		Unidades
			Limite Inferior	Limite Superior	
Média de Q (μ_Q)	Normal	35,059	25,731	44,387	m ³ /s
Desvio padrão de Q (σ_Q)	Normal	44,535	38,844	52,196	m ³ /s
Parâmetro de forma (α)	Gama	0,983	0,760	1,271	Adimensional
Parâmetro de taxa (β)	Gama	35,663	25,620	49,644	m ³ /s
Média de lQ (μ_{lQ})	Normal	2,969	2,733	3,204	m ³ /s
Desvio padrão de lQ (σ_{lQ})	Normal	1,126	0,982	1,320	m ³ /s

Tabela 6.3 – Testes de aderência ao 5 % de nível de significância para a série sem transformar, Q, e Log-transformada, lQ. Período 1911-2000.

Teste	Distribuição	Série	Resultado	Valor p
Qui-Quadrado	Normal	Q	Não se ajusta	3,418 E-4
Qui-Quadrado	Gama	Q	Se ajusta	0,484
Qui-Quadrado	Normal	lQ	Se ajusta	0,103
Kolmogorov-Smirnov	Normal	Q	Não se ajusta	2,00 E-75
Kolmogorov-Smirnov	Gama	Q	Se ajusta	3,80 E-01
Kolmogorov-Smirnov	Normal	lQ	Não se ajusta	1,43 E-54
Shapiro-Wilk	Normal	Q	Não se ajusta	1,968 E-11
Shapiro-Wilk	Normal	lQ	Se ajusta	0,206
Anderson-Darling	Normal	Q	Não se ajusta	5,000 E-04
Anderson-Darling	Normal	lQ	Se ajusta	0,280

A tabela 6.1, também mostra os valores dos limiares, percentis de 33% e 66%, escolhidos para analisar a ocorrência de anos mais secos e úmidos no tempo. Também, a figura 6.8 visualmente mostra a persistência de períodos secos e úmidos segundo esses limiares. Nessa figura é ressaltado o período de validação de 60 anos, 1941-2000. O período de validação, 1941-2000, foi escolhido devido ao fato de na literatura comumente serem utilizados períodos de calibração de 30 anos e para se ter um período de validação extenso.

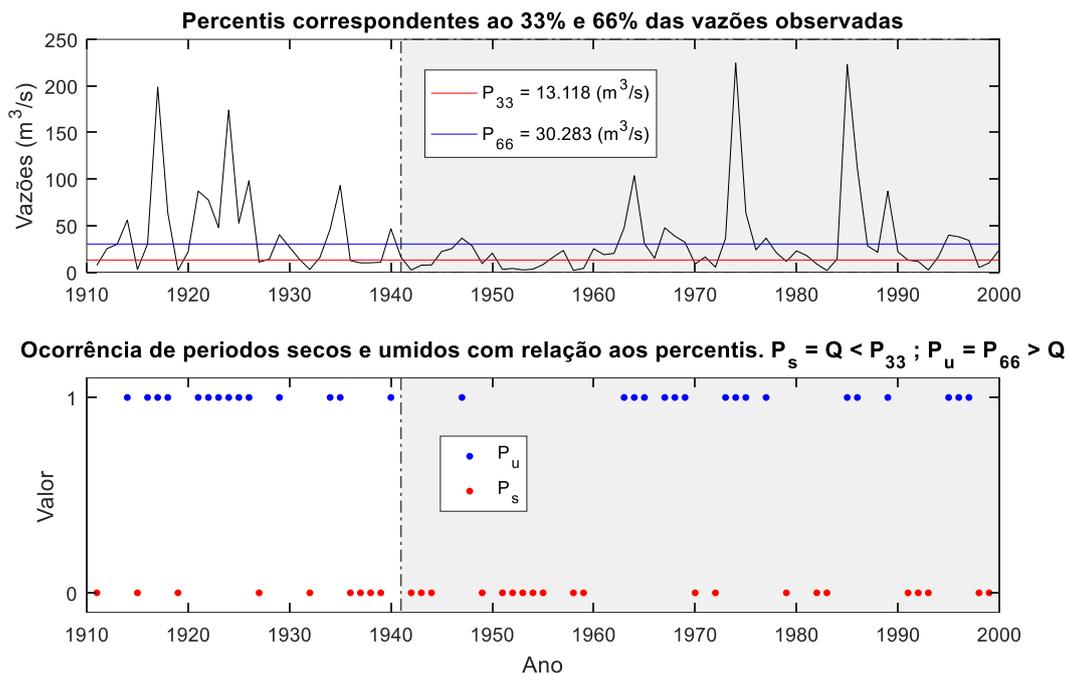


Figura 6.8 – Ocorrência de anos secos e úmidos com relação ao percentil 33% e 66%. P_u e P_s representam os anos úmidos e secos, respectivamente.

6.2 - ANÁLISE DE CORRELAÇÃO ENTRE INDICADORES CLIMÁTICOS E A SÉRIE OBSERVADA DE AFLUÊNCIAS

Antes de ajustar qualquer modelo de previsão foi necessário realizar um diagnóstico sobre a influência dos indicadores climáticos na série observada. Dessa forma, foi realizada uma análise de correlação do tipo Spearman, entre as observações, período total 1911-2000, e os índices NINO3 e Dipolo do Atlântico (DA). Para essa análise, testaram-se diferentes períodos desses indicadores, períodos baseados nos trabalhos de Souza Filho e Lall (2003); Hernandez (2013), Araújo et al., (2015) e no período chuvoso da região.

Primeiramente, para a série de anomalias da TSM do Oceano Atlântico tropical norte (ATN) foi utilizada uma grid formada pelas coordenadas $2,5^{\circ}\text{N} - 27,5^{\circ}\text{N}$, $67,5^{\circ}\text{W} - 27,5^{\circ}\text{W}$ que incluem as coordenadas que definem a região ATN. 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 TSM do Oceano Atlântico tropical sul (ATS) foi utilizada uma grid formada pelas coordenadas $2,5^{\circ}\text{N} - 27,5^{\circ}\text{S}$, $37,5^{\circ}\text{W} - 17,5^{\circ}\text{E}$, que incluem as coordenadas que definem a zona ATS. A grid foi composta por 12 pontos longitudinalmente e 7 pontos latitudinalmente para um total de 84 valores de TSM para cada mês. Logo, o DA foi obtido como a diferença aritmética entre essas duas anomalias (Souza Filho e Lall, 2003; Brabo Alves et al., 2009). Vale mencionar que os valores dos índices climáticos foram um valor médio do período de interesse. Muito embora, os indicadores climáticos OMA, OAN e ODP exerçam influência sobre o regime de chuvas no NEB, esses indicadores não foram utilizados como variáveis externas neste estudo.

A tabela 6.4 e a figura 6.9 apresentam um resumo dos resultados da análise de correlação. Na tabela 6.4, AA representa o período, mês, trimestre ou ano, do ano anterior em que acontece a vazão observada, de forma similar, MA indica o período do mesmo ano em que acontece afluência observada. Todas as correlações encontradas apresentam valores p abaixo do limiar de aceitação, 5 %, portanto, são correlações significativamente diferentes de zero.

No caso do indicador NINO3, encontraram-se altas correlações para os meses de abril e maio do AA e fevereiro do MA, porém, decidiu-se descartar esses indicadores. No caso do indicador de maio AA ele não é comumente utilizado na literatura, já o indicador de fevereiro do MA não possui utilidade prática, devido à quadra chuvosa no Ceará começar nesse mês e geralmente os indicadores climáticos são disponibilizados no começo do mês seguinte da sua ocorrência, no caso teria que se esperar até o mês de março para obter esse indicador.

Dessa forma, o indicador NINO3 utilizado foi o de dezembro do AA, um indicador comumente utilizado em estudos de previsão no Ceará e que corresponde ao mês em que começa a quadra pré-chuvosa. Para fins de comparação em vez do indicador de fevereiro do MA foi utilizado o indicador de janeiro do MA, embora, com uso limitado também pela sua disponibilidade. Com relação ao indicador DA, os indicadores utilizados foram os indicadores de agosto, julho-agosto-setembro e a média anual do AA..

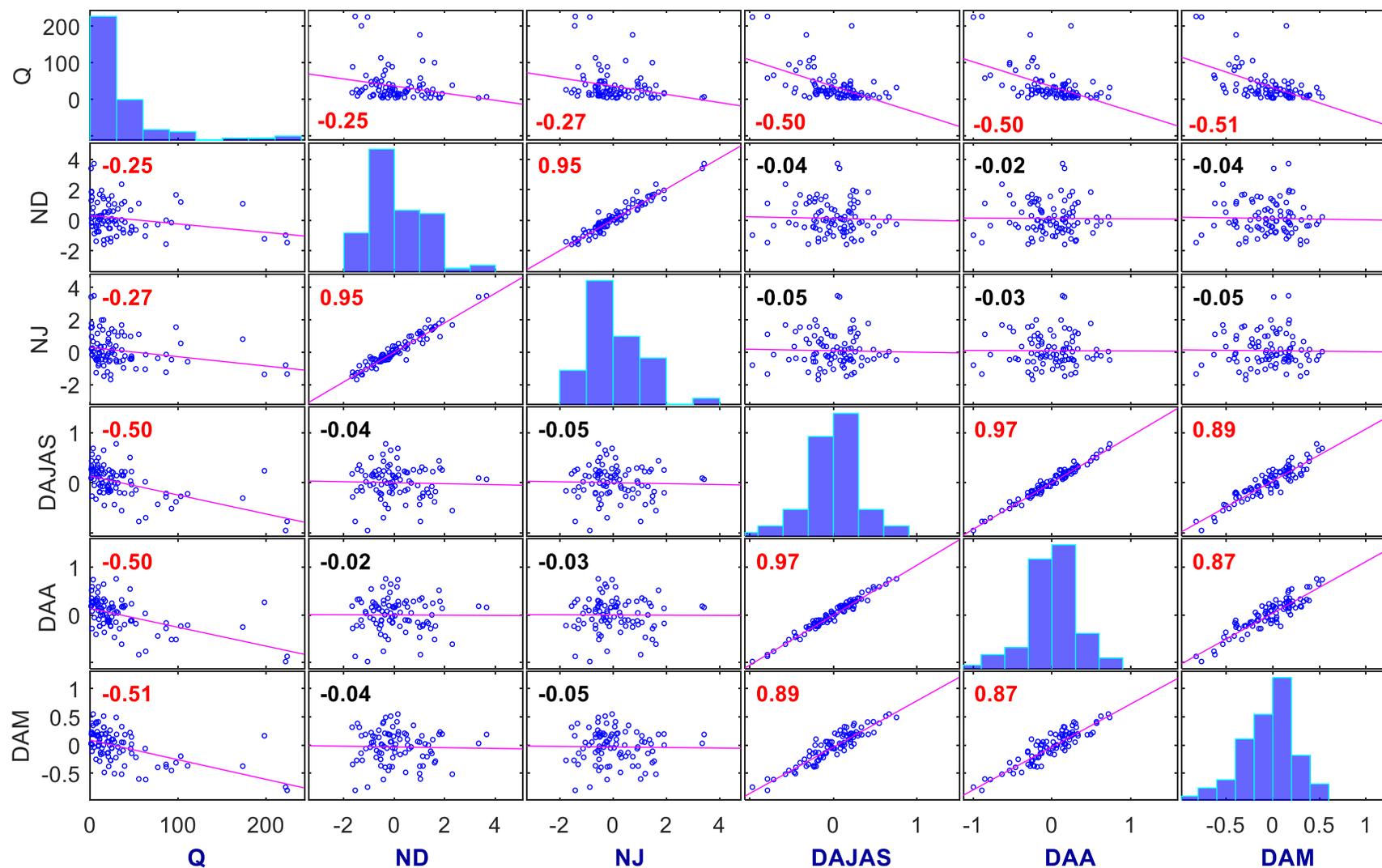


Figura 6.9 – Matriz de Correlação entre a série observada e os indicadores climáticos NINO3 e DA. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Q: vazão observada; ND: NINO3 de dezembro do ano anterior (AA); DAJAS: Dipolo do Atlântico (DA) do período de julho a setembro do AA; DAA: DA de agosto do AA; DAM: média DA do AA.

Tabela 6.4 – Correlação tipo Spearman (ρ_s) entre a série observada período 1911-2000, e os indicadores climáticos NINO3 e Dipolo do Atlântico do ano anterior período 1910-1999.

Indicador	Período/Mês	Notação	ρ_s	Valor p
NINO3	Dezembro AA	ND	-0,25	1,91E-02
	Janeiro MA	NJ	-0,27	9,50E-03
DA	Julho – Agosto – Setembro AA	DAJAS	-0,50	4,00E-04
	Agosto AA	DAA	-0,49	1,00E-03
	Média Anual AA	DAM	-0,51	2,39 E-04

6.3 - AVALIAÇÃO DO DESEMPENHO DOS MODELOS DE PREVISÃO DE VAZÃO

Foram ajustados 132 modelos HMMS – 44 NHMMs, 44 AR-NHMMs e 44 IHMMs – e 11 ARXs. Os indicadores climáticos foram incluídos nesses modelos de forma individual, ND e NJ – 2 tipos de modelos – DAJAS, DAA e DAM – 3 tipos de modelos, e combinada, combinações ND/NJ+DAJAS, ND/NJ+DAA, ND/NJ+DAM – 6 tipos de modelos. Logo, 11 indicadores foram utilizados. Por exemplo, no caso dos NHMMs, tais modelos possuem 2 subclasses, cada classe com 2 e 3 estados ocultos, portanto o número de NHMMs ajustados foi: $2 \times 2 \times 11 = 44$.

Para efeitos de comparação foram ajustados 8 modelos sem informação climática: modelo Gama, Autoregressivo de primeira ordem AR(1) e HMMS Homogêneos (Figura 4.2) com 2 e 3 estados ocultos, 2 HMMS Log-Normal –HL, 2 HMMS Gama – HG e 2 AR-HMMS Log-Normal –ARHL. 151 modelos de previsão foram ajustados.

O modelo Gama é um modelo comumente utilizado no NEB e corresponde a ajustar a série observada, até o tempo T, à distribuição Gama e realizar previsões para o tempo T+1. Logo, nesse modelo, a FDP de previsão para o tempo T+1 corresponde à PDF do ano anterior, FDP do ano T. Esse modelo corresponde também a um modelo de previsão que usa a média histórica até o tempo T para realizar a previsão de vazão do tempo T+1.

O cálculo dos parâmetros dos HMMS foi realizado a partir da maximização numérica direta da verossimilhança, detalhes do algoritmo utilizado encontram-se no Anexo A.1. Para tal maximização foram desenvolvidos vários códigos de programação no software Matlab 2016b. Os códigos produzidos permitem calcular, além dos parâmetros, as FDPs de previsão de modelos HMMS, NHMMs, AR-NHMMs e IHMMs.

Esses códigos foram baseados nos códigos disponibilizados por Zucchini e MacDonald (2009) e Zucchini *et al.*, (2016). Além disso, alguns detalhes para a convergência no ajuste dos HMMs foram tomados de Kirshner (2005a) e Kirshner (2005b). Os códigos testaram-se para HMMs e NHMMs por meio de uma comparação com os resultados obtidos com o código de Kirshner (2005b), exemplos apresentados em Zucchini e MacDonald (2009) e a partir da geração de series sintéticas utilizando a série de afluências observada.

No caso do ajuste e previsão utilizando os modelos ARX e AR, foram utilizadas as funções *arima*, *estimate*, *aibic* e *forecast* do mesmo software. A função *forecast* entrega as medias e desvios padrão previstos, com esses valores foi utilizada a função *normpdf* para gerar as FDPs de previsão desses modelos.

O procedimento do cálculo das FDPs, em base é similar nos HMMs e ARXs. Nesse procedimento, além dos parâmetros das FDPs, precisa-se de uma amostra de valores que cobram o intervalo de vazões da série observada. Além disso, a amostra deve ter observações minimamente espaceadas e ordenadas. Isso, para obter uma melhor integração numérica, devido a que, a FAP corresponde a integrar os valores da FDP utilizando a amostra de valores.

Quando a integração numérica foi realizada com amostras pequenas, foi encontrado que tal integração não era adequada, já que, em alguns casos no cálculo da métrica CRPSS altos valores eram obtidos, por exemplo, valores de CRPSS = 0,9, o que deixava dúvidas sobre o real desempenho dos modelos de previsão. O leitor pode encontrar mais detalhes desse procedimento nos códigos que foram produzidos, disponíveis no link: <https://luiskhh.wordpress.com/2019/04/28/inflow-forecasting-using-hidden-markov-models>.

Nesse link, disponibilizam-se todos os códigos produzidos, tanto para HMMs quanto para ARXs e ARs.

6.3.1 - Qualidade de previsões probabilísticas para um ano na frente

A análise da qualidade das previsões probabilísticas para um ano na frente – FDPs de previsão – foi realizada a partir do valor médio para os 60 anos do período de previsão da métrica CRPSS (CRPSSm). Nesse caso, três critérios foram empregados para avaliar o

desempenho dos modelos de previsão: A) capacidade de previsão geral – CRPSSm; B) capacidade de previsão de anos “secos” – CRPSSm de anos secos definidos pelo percentil de 33% da série observada (CRPSSms); e C) capacidade de previsão de anos “úmidos” – CRPSSm de anos úmidos definidos pelo percentil de 66% da série observada (CRPSSmu).

Essa avaliação baseou-se só em modelos que foram melhores que a climatologia (CRPSS>0), já que, alguns modelos não foram melhores que a climatologia. Esse foi um primeiro filtro que reduziu a grande quantidade de modelos. Logo dessa primeira filtragem, foi encontrado que o desempenho dos HMMs era variável, porque alguns modelos foram melhores que outros dependendo dos critérios de avaliação A, B e C.

Por outro lado, modelos com distribuições Gama saíram das análises mostradas a seguir, porque não apresentaram muitos ganhos no valor do CRPSSm quando comparados a os modelos Log-Normais. Também, HMMs com 3 estados ocultos não foram melhores que HMMs com 2 estados ocultos. Simplesmente, um HMM do tipo Gama, com 2 estados ocultos, foi incluído nas análises seguintes para efeitos de comparação, devido ao seu valor do CRPSSms. No Apêndice B, o leitor pode consultar o desempenho dos modelos não incluídos nas análises mostradas a continuação.

Dessa forma, a seguir mostra-se o desempenho do modelo com comportamento menos variável, o modelo NLND – NHMM (2) Log-Normal com o indicador ND, escolhido a partir de várias análises prévias que compararam os HMMs e os ARXs. Essas análises basearam-se no critério A e foram as seguintes: a) comparação do desempenho dos NHMMs com relação aos ARXs; b) comparação do desempenho dos NHMMs e AR-NHMMs com relação ao ARXs; e c) comparação do desempenho dos NHMMs, AR-NHMMs e IHMMS com relação ao ARXs. A comparação a) trouxe como “melhor” modelo o NLNJ – NHMM (2) Log-Normal com indicador NJ, seguido do NLND – NHMM (2) Log-Normal com o indicador ND, esses modelos serviram como referência na comparação b). A comparação b) entregou como “melhor” modelo o AR(1)-NLNJ – AR(1)-NHMM (2) Log-Normal com indicador NJ, o qual apresentou um desempenho levemente superior aos modelos NLNJ e NLND, o valor do CRPSSm desses três modelos é apresentado na tabela 6.5.

A tabela 6.5 também mostra os modelos que resultaram ser melhores nas comparações a e b para os critérios B e C, modelos NGND – NHMM (2) Gama com indicador ND, com melhor valor do CRPSSms e o NLNJ com melhor CRPSSmu. Tomando todos os modelos da tabela 6.6 como referência se realizou a comparação c, que avaliou o desempenho dos IHMMs frente aos demais modelos, nesse caso obteve-se um “melhor” modelo para o CRPSSmu, o ILDAJAS – IHMM (2) Log-Normal com indicador DAJAS. Por fim, o fluxograma das análises a, b e c é apresentado na figura 6.10.

Tabela 6.5 – “Melhores” modelos de previsão das análises a e b. Todos os HMMs têm 2 estados ocultos.

Modelo	Critério	Valor	Intervalos de Confiança de 95%	
			Inferior	Superior
AR(1)-NLNJ	Maior valor do CRPSSm	0,22	0,15	0,28
NLNJ	Segundo maior valor do CRPSSm	0,21	0,14	0,27
NLND	Terceiro maior valor do CRPSSm	0,18	0,12	0,24
NGND	Melhor valor do CRPSSms	0,41	0,29	0,53
NLNJ	Valor menos negativo do CRPSSmu	-0,01	-0,11	0,09

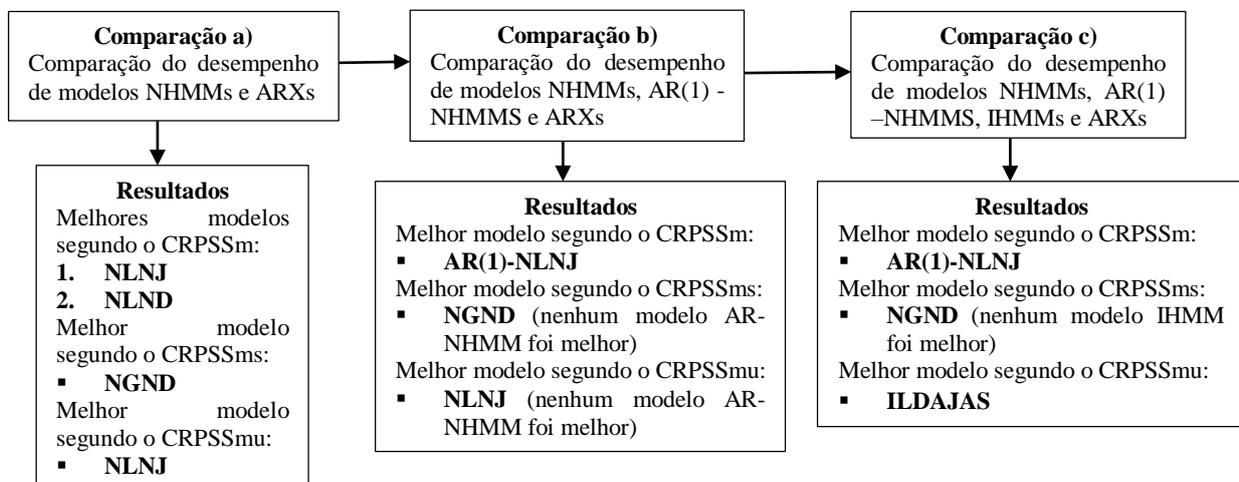


Figura 6.10 – Resultados das análise a, b e c. As análises seguem os critérios A, B e C.

A tabela 6.5 também mostra os intervalos de confiança (IC) de 95% para o CRPSSm, CRPSSms e CRPSSmu. Tais intervalos foram calculados via *Bootstrap* – função *bootstrp* do software Matlab 2016b. O procedimento de cálculo foi o seguinte: i) utilizando a série dos 60 valores do CRPSS, gerou-se uma amostra, via *bootstrapping*, de 1000 valores da média dessa série e ii) dessa série, tomaram-se os percentis de 2,5% e 97,5%, que representam o IC inferior e superior, respectivamente. Embora, esse procedimento, seja uma estimativa simples

e grosseira dos IC, serve de base para a comparação dos modelos e em geral é utilizada na literatura para gerar IC (veja por exemplo, Gilleland, 2010). Esses IC permitem construir as barras de erro que se apresentam como as linhas vermelhas nos gráficos mostrados nesta seção. Segundo a amplitude dessas barras pode ser dizer ou não se a diferença entre modelos é ou não estatisticamente significativa. Logo, barras mais estreitas indicam maior significância e barras mais amplas pouca significância.

Dos modelos da tabela 6.5, aqueles com o indicador NJ operacionalmente tem uso limitado, só ficaram até a comparação c para efeitos de comparação. No caso do modelo NGND, ele apresentou um comportamento variável e algumas limitações que serão detalhadas mais adiante, ele só foi usado para comparar o desempenho do modelo NLND. O ILDAJAS foi utilizado também para essa comparação, ele teve uma performance menos variável, embora, na métrica CRPSSm foi levemente inferior ao NLND, veja a figura B.13 no Apêndice B. A tabela 6.6 mostra os modelos considerados para a comparação do modelo NLND. Essa tabela também mostra os valores dos critérios AIC e BIC desses modelos, que indicam que o modelo de previsão mais adequado seria o IL valores mínimos desses critérios, embora as análises a seguir mostram várias deficiências desse modelo.

Tabela 6.6 – Modelo NL e modelos de comparação. Todos os modelos têm 2 estados ocultos.

Modelo	Notação	Critério de Seleção	AIC	BIC
NLND	NL	Desempenho menos variável	268	288
NGND	NG	Melhor valor do CRPSSms	802	822
ILDAJAS	IL	Melhor valor do CRPSSmu	255	275

A figura 6.11 mostra a comparação, baseada no valor do CRPSSm, do modelo NL (CRPSSm = 0,18) com relação aos modelos NG e IL e modelos sem informações climáticas. Para essa análise, os modelos HG e ARHL de dois estados ocultos tiveram valores inferiores do CRPSSm com relação ao HL (veja as figuras B.21 e B.22 no Apêndice B), logo, saíram dessa análise. Modelos HMMs sem informação climática e com 3 estados ocultos também saíram da análise porque foram inferiores aos HMMs de 2 estados ocultos.

Da figura 6.11, observa-se que o modelo NG tem desempenho inferior a todos os modelos (CRPSSm=0,1), uma das razões porque esse modelo não foi escolhido como modelo final de previsão. Embora, essa diferença não seja estatisticamente significativa. Em geral, os modelos

com distribuições Gama sempre tiveram um desempenho inferior aos modelos Log-Normais. Após do modelo NG, o desempenho do modelo Gama é estatisticamente inferior a todos os modelos. A figura 6.11 mostra também que não existe diferença estatisticamente significativa entre os modelos AR(1), HL, ARX(1) e ND, já que, suas barras de erro são similares.

Nota-se também, segundo a figura 6.11, que o desempenho do IL (CRPSSm=0,16) é levemente inferior ao NL, porém seus IC apresentam uma amplitude maior indicando maior variabilidade do CRPSSm, isso pode ser observado no diagrama de caixa do IL mostrado na figura 6.12. Esses diagramas de caixa mostram a distribuição de valores do CRPSS ao longo do período de validação, permitindo entender o grau de variabilidade do CRPSS e alguns detalhes da previsão dos modelos.

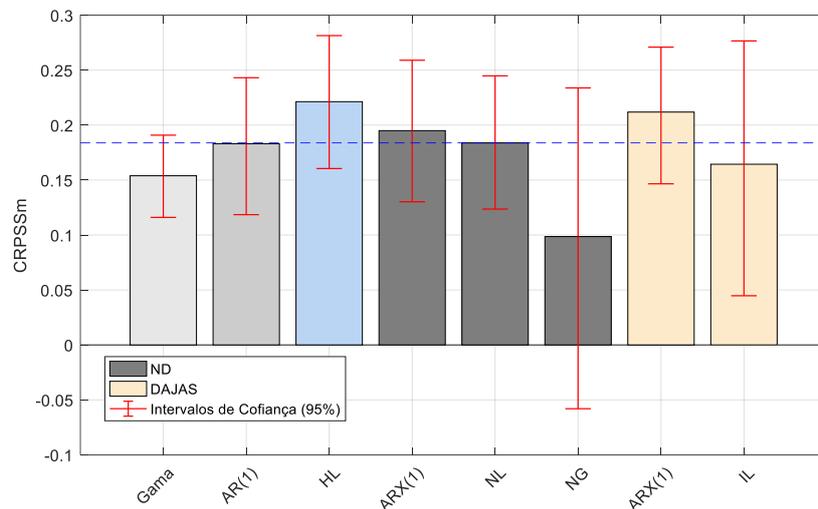


Figura 6.11 – Valor do CRPSSm para o modelo NL, NG, IL e modelos sem informações climáticas.

Observe na figura 6.12 que o modelo Gama e NL são menos variáveis, uma característica desejada em modelos de previsão. O modelo NG apresenta mais variabilidade inclusive gerando valores piores que a climatologia (CRPSS<0), uma característica não desejada em um modelo de previsão de vazão e razão pela qual essa classe de HMMs foi inferior aos HMMs Log-Normais, quando analisado o valor do CRPSSm.

Os HMMs Gama também geram valores atípicos muito piores que a climatologia, uma característica tampouco desejada em modelos de previsão, tornando-os não práticos para seu uso no processo de alocação de água. O modelo IL gera previsões com CRPSSs superiores a

0,5 em aproximadamente 15 anos do período de validação, similar ao NG, no entanto, apresenta o mesmo problema do NG, pois, gera valores negativos e atípicos não desejados. Essa é a razão pela qual seu CRPSSm é inferior ao CRPSSm do NL.

Por fim, note como a mediana do modelo Gama é inferior as medianas de todos os outros modelos. Enquanto, os modelos NL e IL possuem medianas similares, elas são ultrapassadas pelos modelos ARX, AR, HL e NG.

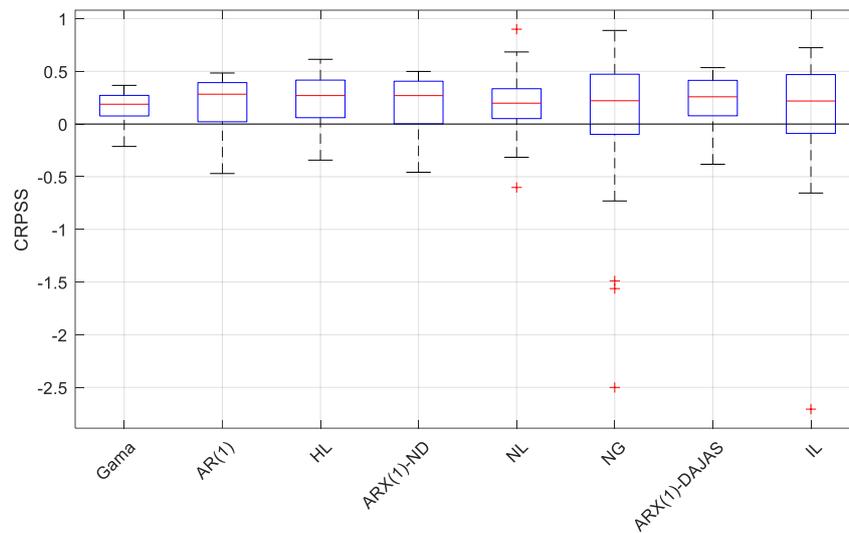


Figura 6.12 – Diagramas de caixa do CRPSS para o modelo NL, NG, IL e modelos sem informações climáticas. A linha horizontal representa o desempenho da climatologia (CRPSS = 0)

As figuras 6.13 e 6.15 mostram os valores das métricas CRPSSms e CRPSSmu. Essas métricas são baseadas numa classificação de anos secos e úmidos, utilizando os percentis de 33% e 66%, veja a figura 5.7. Nesse caso, um ano seco acontece quando a vazão é inferior ao percentil de 33% ($Q < Q_{0,33}$), enquanto, se a vazão supera o percentil de 66% ($Q > Q_{0,66}$) ocorre um ano úmido.

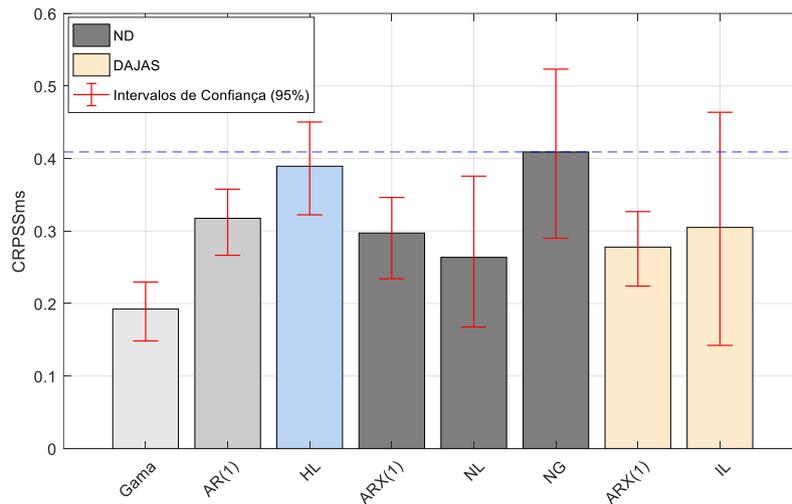


Figura 6.13 – Valor do CRPSSms para o modelo NL, NG, IL e modelos sem informações climáticas

Para os anos secos, figura 6.13, o NG foi superior a todos os modelos ($CRPSS = 0,41$), embora, os NHMMs não sejam estatisticamente diferentes comparados ao AR, ARXs, HL e Gama. Seguido desse modelo está o HL, que supera, com significância estatística, aos outros modelos com e sem informações climáticas. Enquanto, o modelo Gama apresentou pior desempenho. Em geral, veja o Apêndice B, os HMMs tendem a fornecer melhores previsões para os anos secos, em alguns casos eles superam os modelos ARXs, como o caso do IL mostrado na figura 6.13.

O NL não apresentou o melhor desempenho ($CRPSSms = 0,26$), porém, consegue ser melhor que o modelo Gama. Embora, o NL é superado pelo IL ($CRPSSms = 0,30$), o NL é menos variável que o IL ao longo do período de validação, isso pode ser observado na figura 6.14. Veja também, nessa figura, como o IL gera valores do CRPSS negativos para anos secos.

Na figura 6.14, também se observa que o modelo Gama é o menos variável de todos os modelos, com mediana superando a mediana do NL, embora, o NL gera valores do CRPSS de anos secos, muito acima dos gerados pelo o modelo Gama – aproximadamente em 15 anos do período de validação. Note também que o NG apesar de ser melhor, em média, do que os outros modelos, é mais variável que os modelos AR, ARX e HL, restringindo a sua utilização para prever esse tipo de anos.

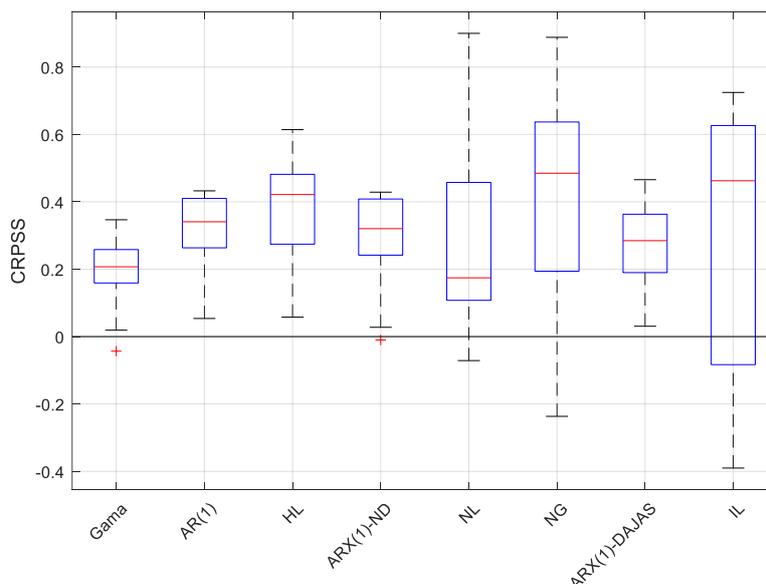


Figura 6.14 – Diagramas de caixa do CRPSS dos anos secos para o modelo NL, NG, IL e modelos sem informações climáticas. A linha horizontal representa o desempenho da climatologia (CRPSS = 0)

No caso de previsões para os anos úmidos, figura 6.15, em média, exceto o IL e o modelo Gama, todos os modelos são piores que a climatologia (CRPSS<0), embora, suas diferenças não sejam estatisticamente significativas. Nesse caso, o NL foi o menos pior entre todos os outros modelos, com um valor do CRPSSmu de -0,03, muito próximo de zero. O modelo Gama (CRPSSmu=0,10) supera o IL (CRPSSmu=0,03), porém, o modelo Gama resulta ser mais variável que o IL figura 6.16.

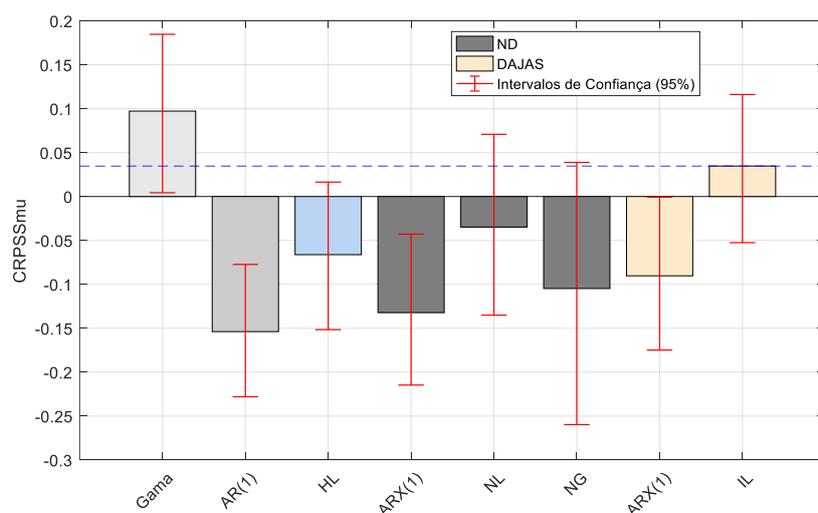


Figura 6.15 – Valor do CRPSSmu para o modelo NL, MG, IL e modelos sem informações climáticas

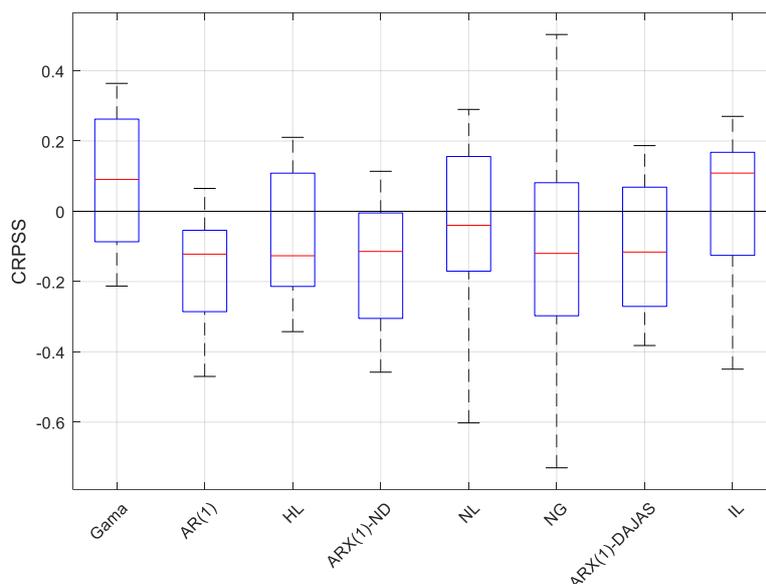


Figura 6.16 – Diagramas de caixa do CRPSS dos anos úmidos para o modelo NL, NG, IL e modelos sem informações climáticas. A linha horizontal representa o desempenho da climatologia (CRPSS = 0)

As figuras 6.15 e 6.16 mostram também que o AR(1) quem apresentou um desempenho superior para anos secos quando comparado com os modelos ARX(1), dessa vez resulta ser bem pior. Nesse caso, existe uma sequência de melhoria no desempenho desde o AR(1) até o ARX(1)-DAJAS, passando pelo ARX(1)-ND, mostrando a vantagem do uso de indicadores DA. Embora, esses modelos autoregressivos gerem valores negativos do CRPSS.

Ainda com relação à figura 6.16, o HL apresenta variabilidade similar ao NL, porém esse último modelo o supera no valor da mediana e na geração de previsões melhores que a climatologia (CRPSS>0). O modelo IL, embora seja inferior, em média, ao modelo Gama, em termos da mediana ele consegue ser levemente superior. Em geral, foi encontrado que os IHMMs melhoram o seu desempenho na previsão de anos úmidos com a inclusão de índices DA, similar ao caso dos modelos autoregressivos. Mostrando o ganho de utilizar esse tipo de indicador e indicando que indicadores climáticos relacionados ao Oceano Atlântico devem ser mais explorados para a previsão de vazão no estado do Ceará.

Para analisar o comportamento do modelo NL ao longo do período de validação, a figura 6.17 mostra a variação do CRPSS produzido por esse modelo nos 60 anos do período de validação. Pode-se observar nessa figura que existe uma tendência do NL em produzir previsões melhores que a climatologia (CRPSS>0), embora, se apresentem alguns anos com previsões

ruis ($CRPSS < 0$) e próximas à climatologia ($CRPSS = 0$). Essa tendência é conferida pelos valores do $CRPSS_m$ nas seis décadas do período de validação (linhas horizontais azuis na figura 6.17) e também pela média móvel de 10 anos representada na figura 6.17 pela linha tracejada vermelha. Tal média móvel tem um valor médio de 0.18, o que mostra que existe uma tendência decenal do modelo NL a gerar valores, a cada 10 anos, iguais ao valor do $CRPSS_m$.

A figura 6.17, ainda apresenta alguns anos (retângulos cinza verticais) selecionados na intenção de mostrar o comportamento do modelo NL nesses anos. Os anos de 1955 e 1989 representam o máximo e mínimo valor do $CRPSS$, respectivamente, depois de tirar os *outliers* dessa métrica *outliers* que aparecem na figura 6.12. Enquanto, os anos de 1945 e 1992, representam anos com $CRPSS$ próximo ao valor do $CRPSS_m$, que mostram o desempenho geral do modelo.

Em detalhe a 6.18 mostra as FAPs de previsão dos anos 1955 e 1989, figuras a e b, respectivamente, em escala logarítmica e real. Essas figuras apresentam também as FAPs observadas e da climatologia. Observe como na figura 6.18a, ano 1955, o NL gera uma FAP que simula a forma da FAP observada, embora a FAP prevista esteja deslocada do valor observado, figura 6.18a inferior. No caso do ano 1989, a FAP prevista está mais centrada à observação, no entanto, não segue a forma da FAP observada.

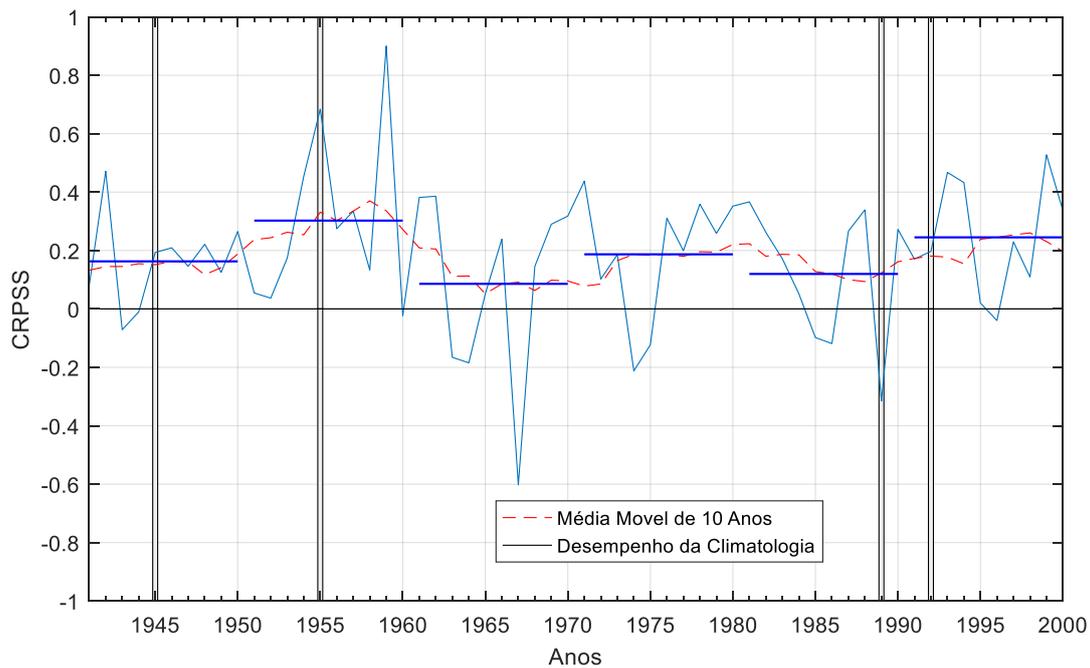
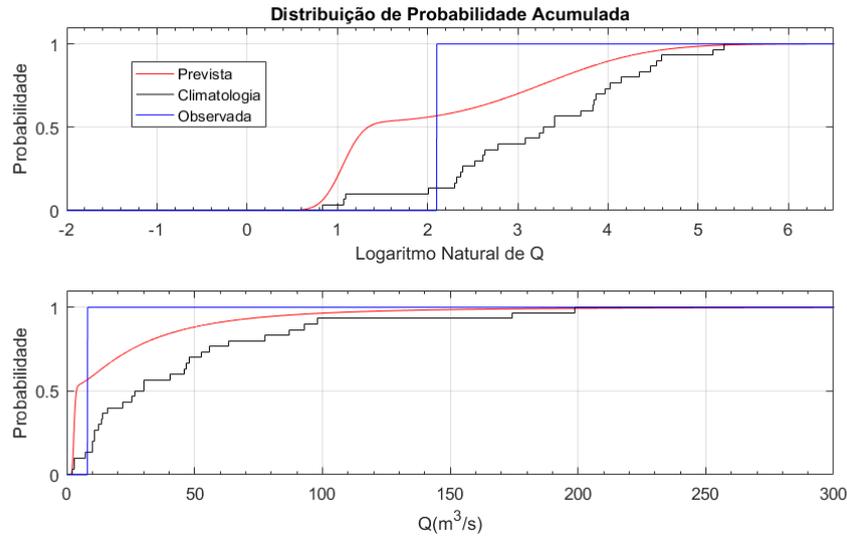


Figura 6.17 – Variação do CRPSS produzido pelo modelo NL ao longo dos 60 anos do período de validação. As linhas azuis horizontais indicam o valor médio do CRPSS na década e a linha vermelha tracejada indica a média móvel de 10 anos. Os retângulos verticais indicam valores do CRPSS selecionados para mostrar o desempenho pontual do modelo NL.

A figura 6.19 é similar à figura 6.18 e mostra os anos 1945 e 1992. Observe como o modelo NL gera FAPs previstas que seguem a forma da FAP observada, embora não se centram no valor observado, sempre deslocadas desse valor. O modelo NL nesses casos gera valores acima da observação, isso é melhor observado na análise das previsões pontuais. Esse comportamento se apresentou também para alguns anos secos. Por exemplo, o ano 1993, relacionado com o *El Niño* 1992 e o ano 1998, relacionado com *El Niño* de 1997, veja a figura B.23 no Apêndice B, nesses anos comportamento é variável, $CRPSS=0,4674$ (ano 1993) e $CRPSS=0,1094$ (ano 1998).

a) Ano 1955 – CRPSS = 0,6841



b) 1989 CRPSS = -0,3155

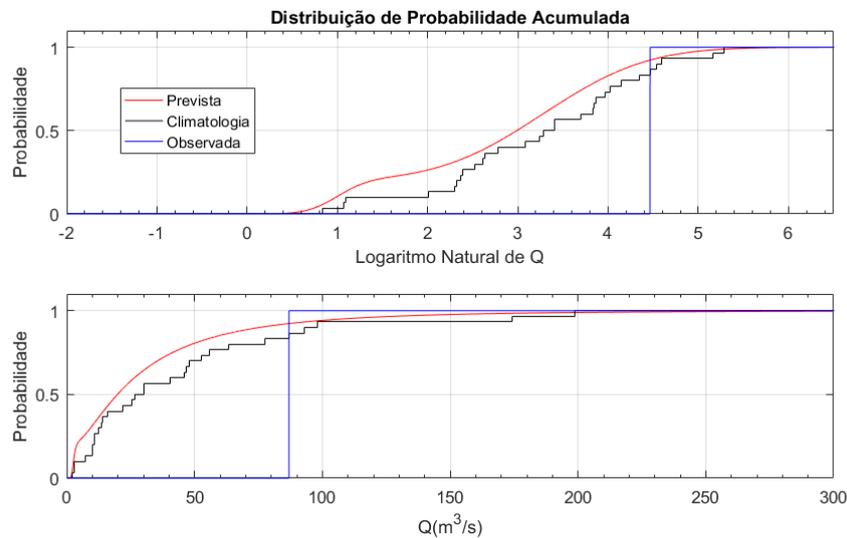
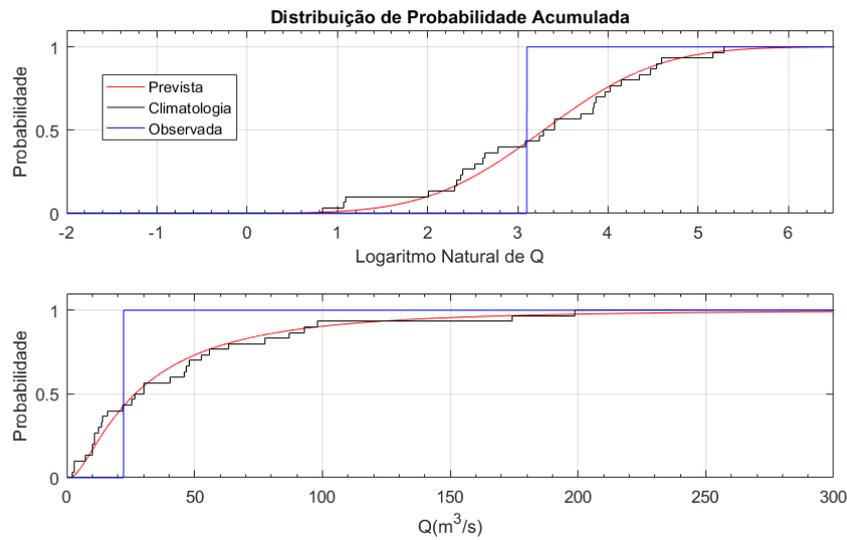


Figura 6.18 – Forma das distribuições de previsão acumuladas para os anos com melhor (a) e pior (b) desempenho. Esses valores são diferentes dos *outliers* gerados pelo modelo NL.

Outra ilustração do desempenho do modelo NL é apresentada nas figuras 6.20 e 6.21. Essas figuras mostram o valor do CRPSS para os anos secos e úmidos definidos pelo percentil de 33% e de 66%, respectivamente. Essas figuras também trazem a variação do indicador climático ND utilizado para realizar as previsões ND do período 1940-1999. Ainda, essas figuras apresentam a variação decenal, média móvel de 10 anos, do índice da ODP anual esse índice anual foi obtido a partir dos valores mensais do período 1940-1999. Os valores mensais do índice da ODP são fornecidos pelo *Joint Institute for the Study of the Atmosphere and Ocean (JISAO)*, mais informações sobre esses valores e respeito à construção desse indicador podem ser encontradas em JISAO (2019).

Ano 1945 – CRPSS = 0,1932



Ano 1992 – CRPSS = 0,1971

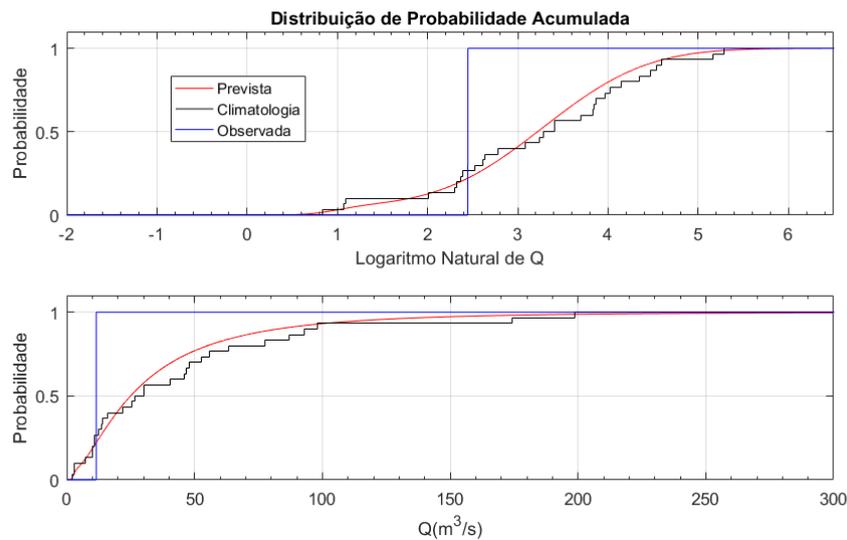


Figura 6.19 – Forma das Distribuições de previsão acumuladas para anos próximos ao valor do CRPSSm. Anos 1945 (a) e 1992 (b).

A figura 6.20 (superior) mostra que os anos secos (14 anos) sempre estão na maioria das vezes relacionados a valores positivos do indicador ND – como era de se esperar pela fase quente do ENSO. Somente em alguns anos isso não acontece, por exemplo, o ano 1942 tem um valor negativo do ND associado a um valor negativo do CRPSS. O comportamento da métrica CRPSS em anos com valores do ND negativos (figura 6.20 – inferior) não é muito claro, embora, sempre que se tem uma tendência de aumento desse indicador encontra-se um ano seco, veja, por exemplo, os anos 1944 e 1990.

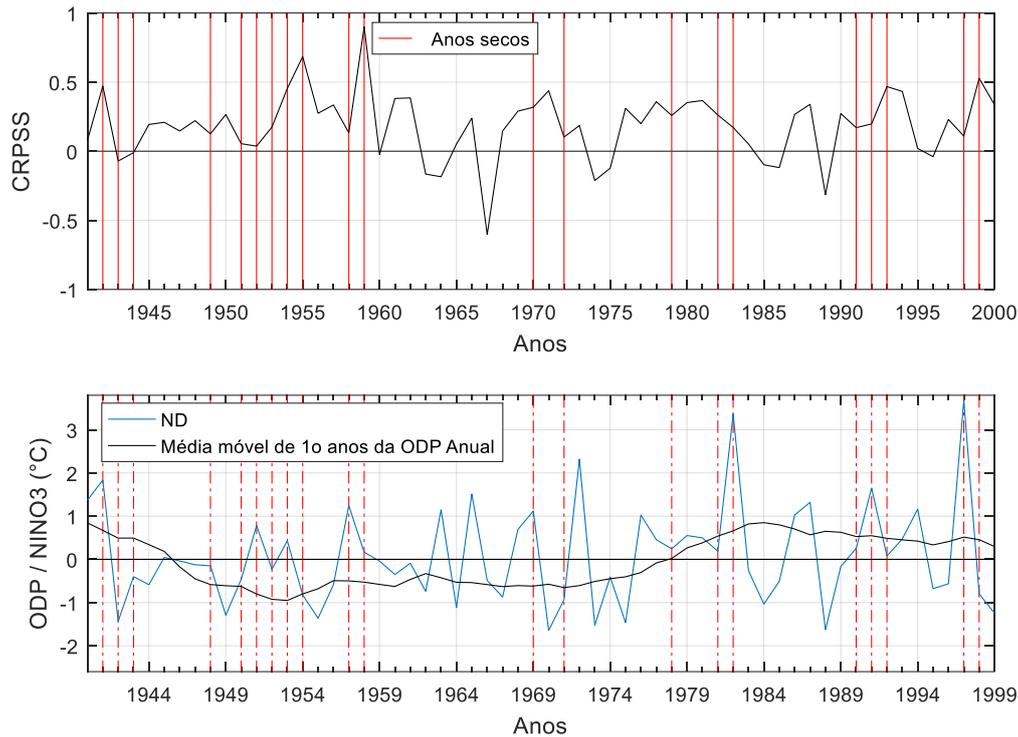


Figura 6.20 – Superior: valores do CRPSS produzidos pelo modelo NL para os anos secos ao longo dos 60 anos do período de validação. Inferior: valores dos indicadores climáticos ND e da ODP para o período 1940 – 1999

A figura 6.21 (superior) é similar à figura 6.20 (superior), nesse caso existem 11 anos úmidos relacionados a valores negativos do ND – como era de se esperar na fase fria do ENSO. A maioria de anos úmidos encontra-se no período de 1963 a 1975, época na qual existe uma variabilidade de valores positivos às negativos do ND. Embora, exista uma relação entre anos úmidos e valores negativos do ND, não existiu um adequado desempenho do modelo nesses anos, já que, existem valores negativos do CRPSS para a maioria desses anos.

As análises das figuras 6.20 e 6.21 induzem a pensar na necessidade de incluir outras modulações climáticas que possam auxiliar na previsão utilizando HMMs. Por exemplo, a figura 6.20 (inferior) e figura 6.21 (inferior) mostram praticamente as três oscilações decenais da ODP, especificamente o período frio de 1947–1976, um período quente finalizando ao redor de 1946 e outro período quente começando ao redor de 1977 e indo até 1999. Pode-se observar como nesse último período existiram mais valores positivos do ND, enquanto, no período frio houve oscilação nos valores desse índice, variando de positivos às negativos.

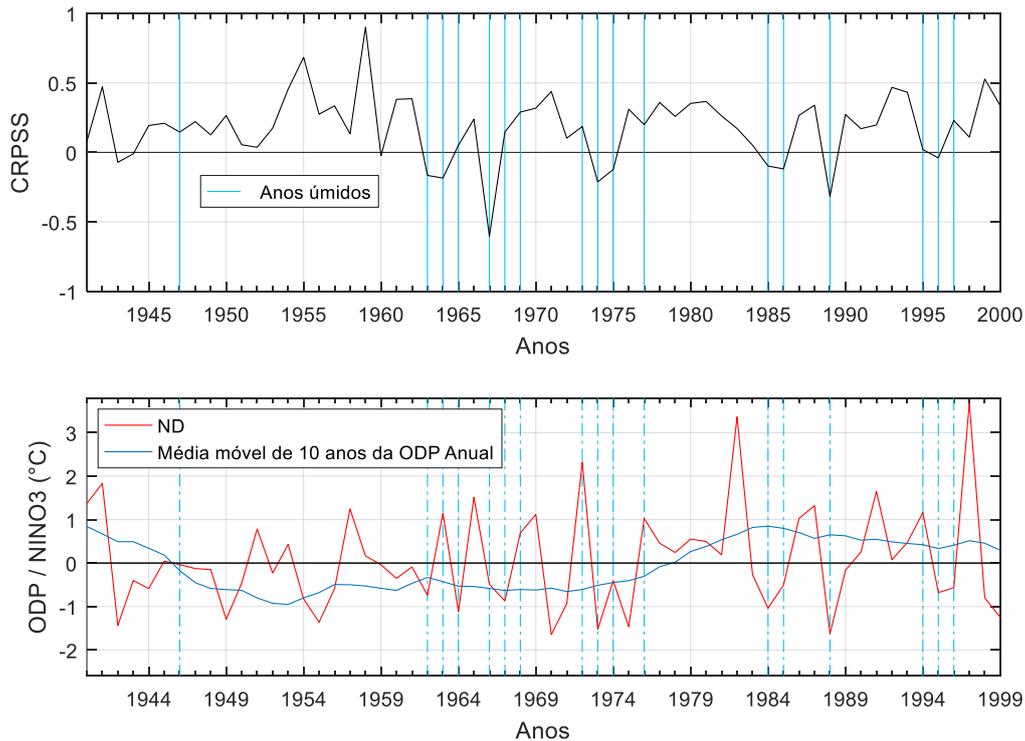


Figura 6.21 – Superior: valores do CRPSS produzidos pelo modelo NL para os anos úmidos ao longo dos 60 anos do período de validação. Inferior: valores dos indicadores climáticos ND e da ODP para o período 1940 – 1999

A ODP, poderia ser incluída para representar uma forçante climática decenal e com isso auxiliar o processo de previsão, por exemplo, ajudar a melhorar a previsão para os anos entre 1963 a 1975. Dessa forma, junto com o indicador ND, se poderia capturar a variabilidade e persistência em diferentes escalas: decenal e interanual. Isso pode ser realizado utilizando HMMs Hierárquicos (HHMMs).

No caso de previsão de aflúências realizado neste trabalho, um possível HHMM poderia ter uma cadeia de Markov superior, influenciada por indicadores climáticos decenais, como o indicador da ODP, que modularia uma cadeia de Markov, por sua vez influenciada por indicadores interanuais como o ND, em que essa última cadeia produziria as observações. Uma análise similar à realizada com a ODP, poderia ser realizada com a OMA e a OAN, nesse caso, analisando o DA, tal e como foi realizado no trabalho de Kayano et al., (2018), figura 3.6.

Outro indicador climático que poderia ser utilizado, na escala interanual, para auxiliar na previsão utilizando HHMMs é o indicador Atlântico 3 (ATL3). Esse indicador é definido

como as anomalias da TSM do Atlântico Tropical na região centro-leste – coordenadas: 3°S – 3°N; 20°W–0°E. Nessa região existe uma variação de temperaturas similar às encontradas quando se apresentam fenômenos do ENSO, produzindo variações das pressões e das direções dos ventos na região tropical modificando a ocorrência de chuvas nessa região incluindo no NEB. Alguns autores denominam tais eventos como Niños (Niñas) do Atlântico, logo, poderia ser avaliada a sua relação com as vazões utilizadas nesta tese. Maiores informações sobre o ATL3 e a sua relação com o ENSO e a variabilidade de chuvas no Atlântico equatorial podem ser encontradas em Rodriguez-Fonseca *et al.* (2009).

Outro indicador na escala interanual de possível utilização é o *Multivariate ENSO Index* (MEI), logo, também deverá ser avaliada a correlação desse indicador com a quadra chuvosa do NEB. O MEI tenta monitorar o ENSO utilizando seis variáveis observadas sobre o Pacífico tropical: pressão ao nível do mar, componentes zonal e meridional dos ventos superficiais, TSM, temperatura superficial do ar e fração total de nuvens do céu, maiores informações desse indicador podem ser consultadas em Wolter e Timlin, (1993, 1998).

Finalmente, com relação à eficiência do modelo NL, esse modelo gera previsões melhores que a climatologia (CRPSS>0) em 80% dos anos do período de validação (48 anos) e em 55% (33 anos) dos anos desse período o modelo gera valores do CRPSS maiores que a sua média (CRPSS> CRPSSm). Em 12 anos do período de verificação (20%) o modelo NL gera valores do CRPSS negativos, sendo que desses anos 9 são anos úmidos, anos em que o NL não apresentou um bom desempenho. Com isso o modelo NL apresenta uma eficiência aceitável como modelo de previsão de vazão. Embora, seja necessário verificar os valores das previsões pontuais, isso é verificado na próxima seção.

6.3.2 - Qualidade de previsões pontuais para um ano na frente

As FDPs previstas pelos HMMs podem ser transformadas em previsões pontuais fazendo uma integração numérica. No caso da média prevista ela pode ser calculada como: $\mu_p = E(R_{T+h}) = \int_{-\infty}^{\infty} rf(r) dr$. Em que $f(r)$ e r representam a FDP prevista e a amostra valores – descrita no começo desta seção, respectivamente. No caso dos modelos autoregressivos, a função *forecast* do software Matlab 2016b entrega as médias previstas. Para essas médias

previstas, calcularam-se as métricas de avaliação da qualidade de previsões pontuais, descritas na seção 5.4.1.2.

Nesta seção, as médias previstas para o período de validação de 60 anos produzidas pelo modelo NL foram comparadas com as médias previstas pelos modelos sem informações climáticas – modelo Gama, AR(1) e HL – e as produzidas pelo modelo ARX(1) com o indicador ND ARX(1) –ND. No caso do modelo Gama a média prevista corresponde a: $\mu_p = \alpha \times \beta$, em que α e β correspondem aos parâmetros de forma e de escala da FDP Gama ajustada tal e como explicado também no início desta seção.

Dessa forma, a figura 6.22 apresenta o valor do Erro Médio Absoluto (EMA) das médias previstas pelos modelos ARX(1) –ND, NL e os modelos sem informações climáticas. Essa figura também apresenta os Intervalos de confiança (IC) de 95% calculados via *Bootstrap* de forma similar aos calculados para os valores médios da métrica CRPSS.

Observe, na figura 6.22, como o modelo Gama apresenta o valor mais alto do EMA (EMA=27,0), enquanto que o NL apresentou um valor do EMA = 21,2. Os modelos autoregressivos – AR(1) e ARX(1) – apresentam valores do EMA similares ao do modelo NL. Tendo em conta que um valor do EMA = 0 indica previsões perfeitas, os modelos diferentes do Gama realizam previsões levemente melhores, sendo que o HL apresenta o EMA mais baixo igual a 20.8. Embora, todos os modelos pelas suas barras de erros apresentam não sejam diferentes, o modelo Gama realiza previsões com mais erro, já que, existem mais diferenças entre suas previsões e observações no período de validação.

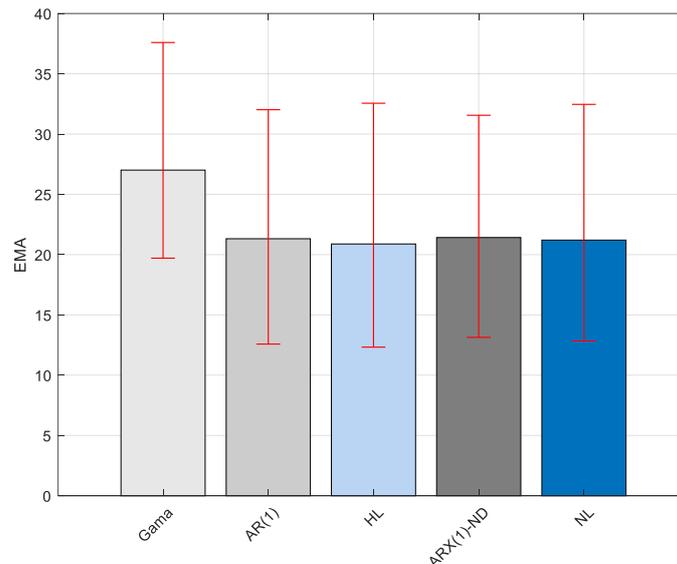


Figura 6.22 – Valores do Erro Médio Absoluto (EMA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*.

A figura 6.23 mostra os valores do Erro Médio Percentual Absoluto (EMPA) para as médias previstas. Todos os modelos entregam previsões de baixa acurácia ($EMPA \geq 50$). Observa-se como o modelo Gama com $EMPA = 297,6\%$ realiza previsões realmente ruins das médias previstas. O NL apresenta um valor de $156,2\%$ mais baixo embora, sem diferença estatística significativa, em comparação ao modelo Gama, e similar aos modelos autoregressivos. O HL com valor mais baixo, igual a $130,6\%$, e com diferença estatisticamente significativa, se mostra como o modelo com melhores valores para as médias previstas.

No caso do coeficiente de Nash-Sutcliffe (NS), figura 6.24, todos os modelos entregam valores negativos, que indica que não existe ganho ao utilizar os modelos de previsão e que é melhor um modelo de previsão igual à média do período de validação igual a $30,4 \text{ (m}^3/\text{s)}$. O modelo Gama apresenta o valor menos negativo ($NS = -0,05$), porém, os IC mostram que esse modelo é mais variável, com tendência a gerar valores do NS bastante negativos, indicando uma diferença menos significativa comparado com os outros modelos. Os modelos HL e NL, com valores do NS de $-0,07$ e $-0,08$, respectivamente, tendem a variar para valores do NS positivos, diferente dos modelos autoregressivos, que com valores do NS similares tendem sempre a gerar valores negativos dessa métrica. Embora, pela sua significância esses modelos não sejam estatisticamente diferentes.

O comportamento inadequado do modelo Gama, para as métricas EMA, EMPA e NS, é conferido de novo para os valores da Distância Multicritério (DM) mostrados na figura 6.25. Embora, todos os modelos se afastem do ponto ótimo (0,0), se ressalta que o modelo Gama apresenta maior valor dessa métrica (DM = 3,16), com tendência de gerar a valores mais altos da DM. O valor mais baixo da DM é do modelo HL (DM = 1,70), levemente diferente do modelo NL com DM = 1,90, esses valores são similares aos dos modelos autoregressivos. A maior significância de novo é para o modelo HL.

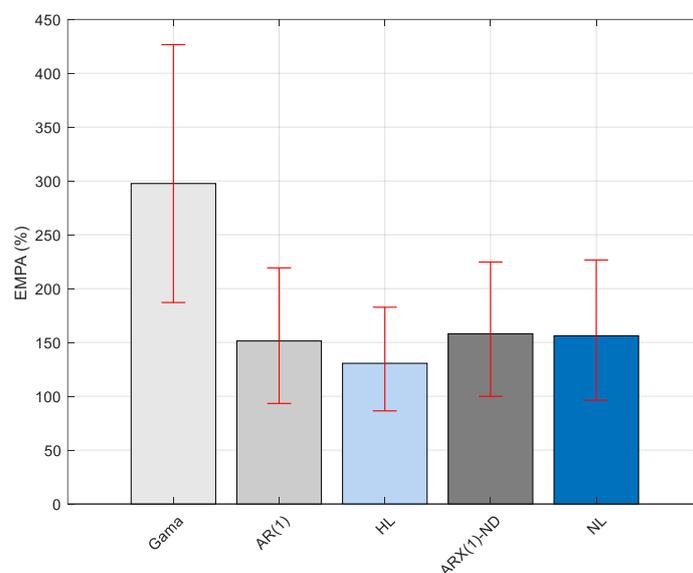


Figura 6.23 – Valores do Erro Médio Percentual Absoluto (EMPA) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*.

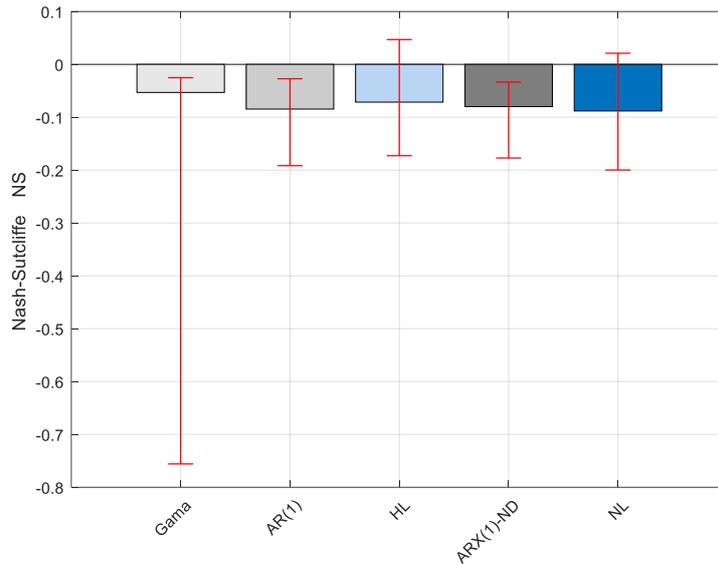


Figura 6.24 – Valores do coeficiente de Nash-Sutcliffe (NS) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*.

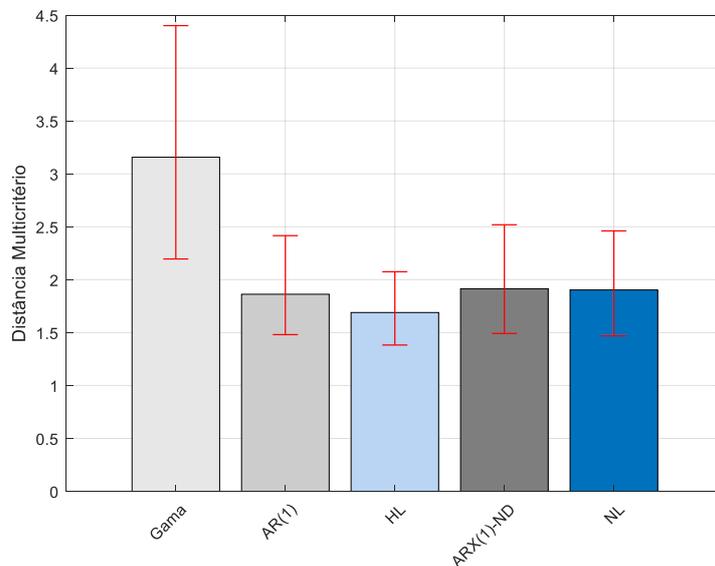


Figura 6.25 – Valores da Distância Multicritério (DM) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*.

No caso da correlação, calculada utilizando a equação 5.11, a figura 6.26 apresenta os valores dessa métrica produzida pelos modelos em questão. Observe como o modelo Gama apresenta correlação negativa mais alta (-0,25), o NL tem a correlação menos negativa igual a -0,05 e o HL apresenta uma correlação positiva igual a 0,04. Esses resultados indicam que o modelo

Gama apresenta melhores previsões, porém, contradiz as análises anteriores realizadas para as métricas EMA, EMPA, NS e DM. No caso dos modelos autoregressivos eles são estatisticamente iguais, e não existe diferença entre esses modelos e os modelos HL e NL.

Logo, para entender melhor o resultado da correlação, uma correlação do tipo *Spearman* foi realizada cujos resultados são apresentados na matriz de correlação da figura 6.27. Nessa figura, pode-se observar que a correlação para o modelo Gama ($\rho_s = -0,22$) não é significativa (valor $p = 0,09$) igual para o NL com $\rho_s = 0,07$ e valor $p = 0,60$. O modelo HL apresentou uma correlação positiva e significativa ($\rho_s = 0,28$ e valor $p = 0,03$). Dessa forma, pelos valores das métricas de qualidade pontuais o modelo HL parece entregar melhores previsões. Embora o modelo NL, apresente métricas similares às desse modelo.

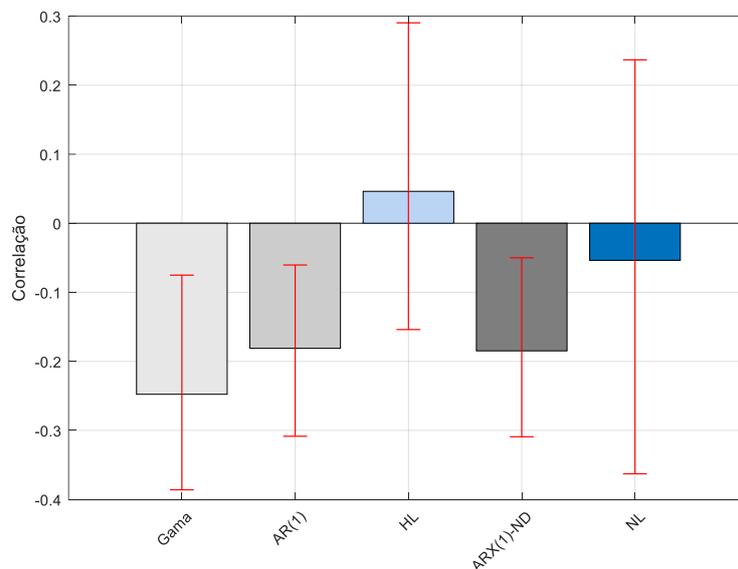


Figura 6.26 – Valores da Correlação (Equação 5.11) para os modelos sem informação climática – Gama, AR(1) e HL – e modelos com o indicador ND – ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*.

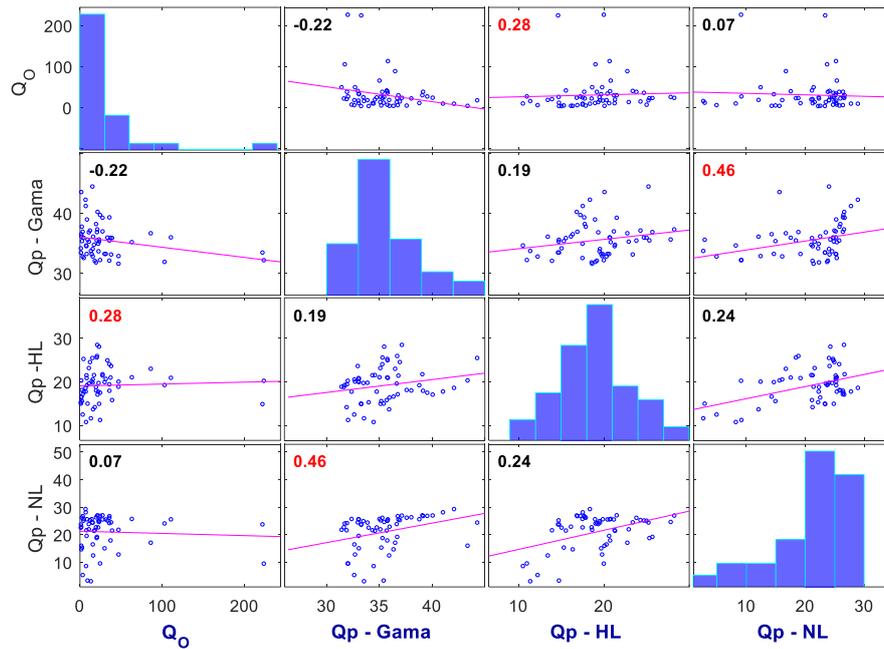


Figura 6.27 – Matriz de correlação tipo Spearman entre as vazões observadas e as médias previstas nos 60 anos do período de validação para pelos modelos Gama, HL e NL. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Qo: vazão observada; Qp: vazão prevista; Gama: Modelo Gama; HL: HMM Log-Normal de 2 estados ocultos; NL: NHMM Log-Normal de 2 estados ocultos que inclui o indicador NINO 3 de dezembro do ano anterior.

Analisando o conjunto de métricas pontuais para o período de verificação, todos os modelos são bastante ruins, sendo mais ruim o modelo Gama. Logo, foram analisadas essas métricas quando se tomam só os anos secos, definidos pelo percentil histórico de 33%. Porém resultados similares dessas métricas pontuais foram encontrados (veja o apêndice C). A mediana e a média histórica também foram tomadas para definir os anos secos e de novo os resultados das métricas pontuais são similares a quando se toma esse percentil.

Logo, tomando os resultados do CRPSS, decidiu-se avaliar o desempenho das previsões pontuais para os anos em que o modelo NL entrega valores do CRPSS maiores ou iguais ao seu valor médio ($CRPSS_m = 0,18$), os resultados para as métricas pontuais foram melhores, tendo em conta que esse modelo entrega previsões maiores ou iguais ao $CRPSS_m$ em 55% (33 anos) dos anos do período de verificação. Embora, essa forma de avaliação seja enviesada para o modelo NL, ao utilizar como referência o $CRPSS_m$ do modelo Gama resultados similares aos mostrados a continuação foram encontrados (veja o apêndice C). Assim, nas figuras 6.28 até 6.38 se apresentam os resultados dessa análise de desempenho.

Primeiramente para a métrica EMA, na figura 6.28 similar à figura 6.22 , observa-se como existe mais significância entre as diferenças dos modelos, o NL resulta ser bem melhor que os outros modelos com um valor mais próximo de zero (EMA=5,8), o modelo Gama é o pior modelo com EMA=17,1, os modelos autoregressivos e o HL são similares EMA com valores ao redor de 7,5. A diferença mais significativa está entre o HL e NL com barras de erro mais estreitas.

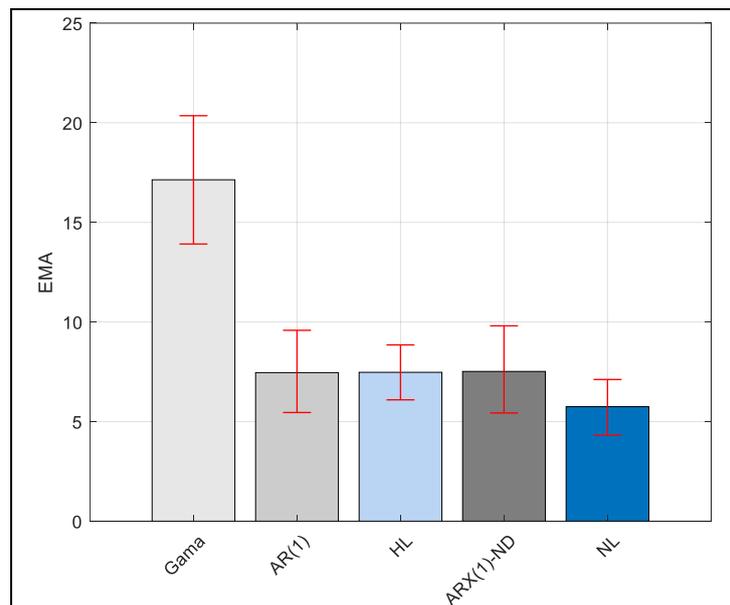


Figura 6.28 – Erro Médio Absoluto (EMA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.

Com relação ao EMPA, na figura 6.29 similar à 6.23, o NL apresenta o menor valor (EMPA=73,0%) comparado com os outros modelos e com maior significância, embora, a acurácia das previsões geradas por esse modelo seja baixa $EMPA \geq 50\%$. O modelo Gama (EMPA=235,7%) é o modelo mais ruim, porém, sua diferença não é muito significativa. Os modelos autoregressivos apresentaram comportamento similar EMPA=105,7% e EMPA=111,1% para o AR(1) e ARX(1), respectivamente e são piores que o modelo HL (EMPA=91,3).

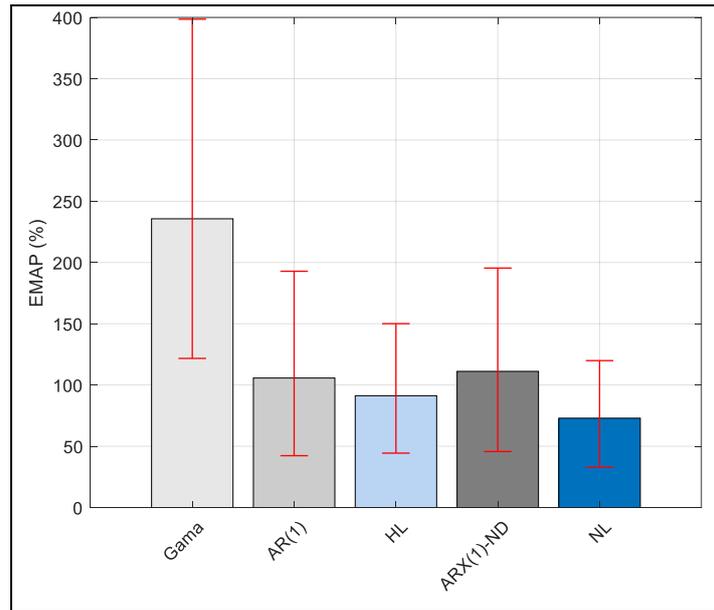


Figura 6.29 - Erro Médio Percentual Absoluto (EMPA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.

No caso do NS, a figura 6.30, nota-se que, diferente ao demonstrado na figura 6.24, o modelo NL apresenta um valor positivo ($NS=0,4$) junto com o HL ($NS=0,2$). Os outros modelos apresentam valores do NS negativos sendo mais ruim o modelo Gama ($NS=-3,3$). Enquanto, os modelos AR(1) e ARX(1), com diferencia mais significativa, apresentam valores de $NS=-0,03$ e $NS=-0,08$, respectivamente, isso indica que nesse caso ao invés de utilizar esses modelos autoregressivos é melhor utilizar a média do período de validação.

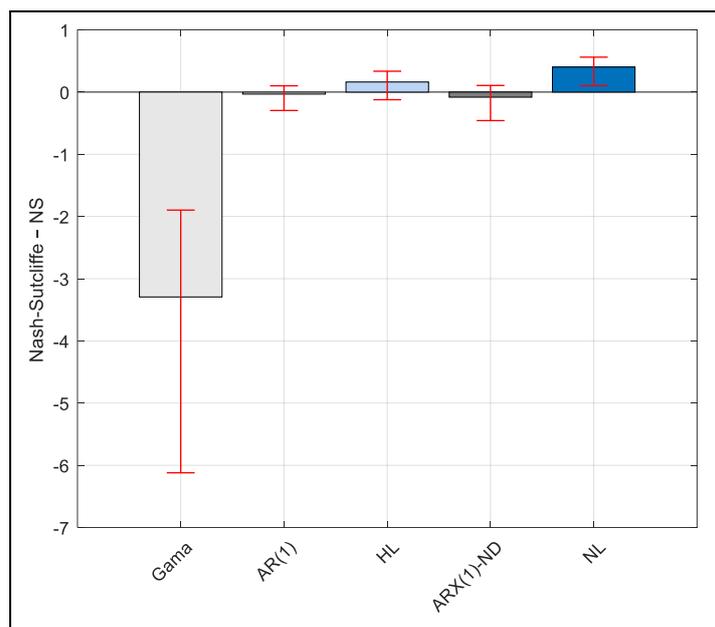


Figura 6.30 – Coeficiente de Nash-Sutcliffe (NS) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.

Para a métrica DM, na figura 6.31 similar à figura 6.25, o modelo NL tem o valor mínimo, com significância estatística, em relação aos outros modelos ($DM=0,9$), esse valor é bem próximo do ponto ótimo, indicando melhores previsões, isso se deve ao baixo EMPA e o valor positivo do NS. O modelo Gama ($DM=4,9$) resulta ser mais ruim, com tendência a gerar valores mais altos da DM, porém, com menos diferença significativa comparado aos outros modelos. Os modelos autoregressivos tem um valor da DM similar em torno a 1,5 e o modelo HL é levemente inferior a esses modelos com $DM=1,2$.

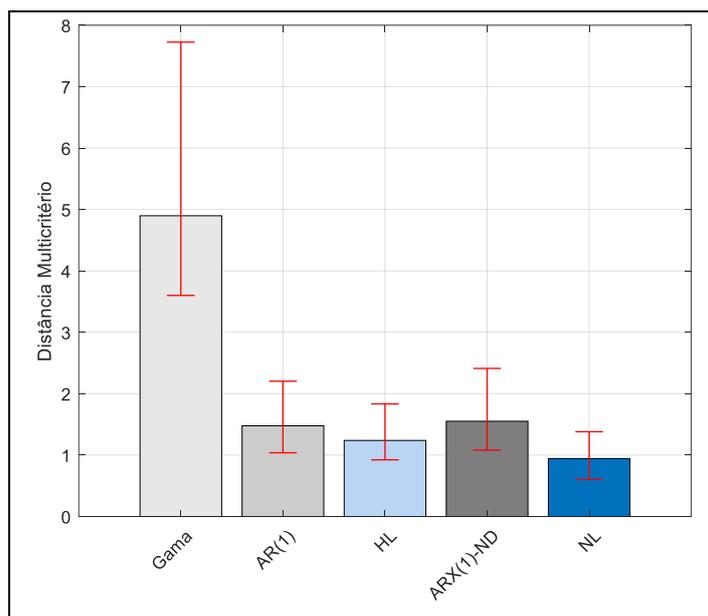


Figura 6.31 – Distância Multicritério (DM) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.

Por fim, na correlação, na figura 6.32 análoga à figura 6.26, o modelo NL apresenta alta correlação em relação aos outros modelos (Correlação=0,7) bastante significativa, barras de erros mais estreitas e superior à do modelo HL (Correlação=0,4). O modelo Gama e ARX(1) apresentam correlações negativas com valores de -0,06 e -0,03, respectivamente, já o AR(1) apresenta um valor positivo próximo de zero igual a 0,01, embora, esses modelos não apresentem uma diferença significativa.

Ao realizar uma correlação do tipo Spearman, na figura 6.33 similar à figura 6.27, pode-se observar que a correlação entregue pelo NL é alta e significativa ($\rho_s = 0,7$ valor $p = 5,0 \times 10^{-6}$), o HL também entrega uma correlação significativa ($\rho_s = 0,4$ valor $p = 0,04$). O modelo Gama com correlação $\rho_s = 0,1$ e valor $p = 0,5$ não apresenta correlação alta nem significativa. Os modelos autoregressivos apresentaram valores de ρ_s com valores p também ao redor de 0,5, similares à correlação do modelo Gama.

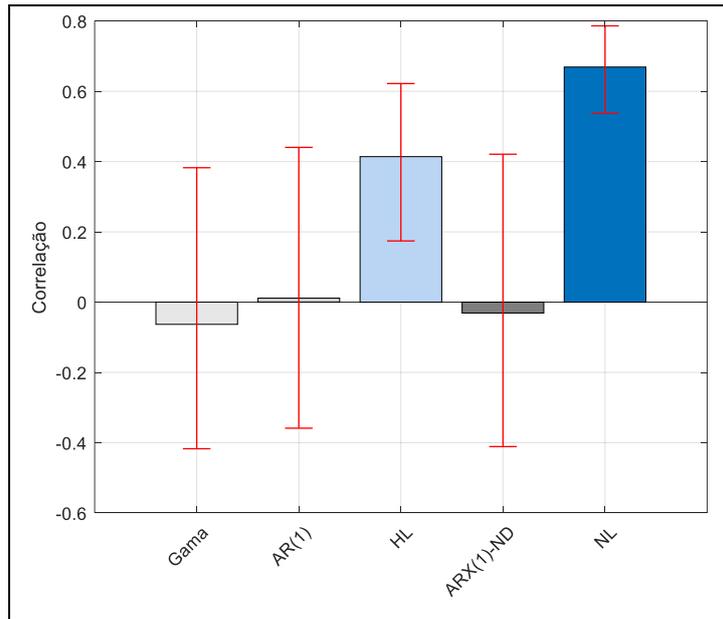


Figura 6.32 – Correlação para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via *Bootstrap*. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL.

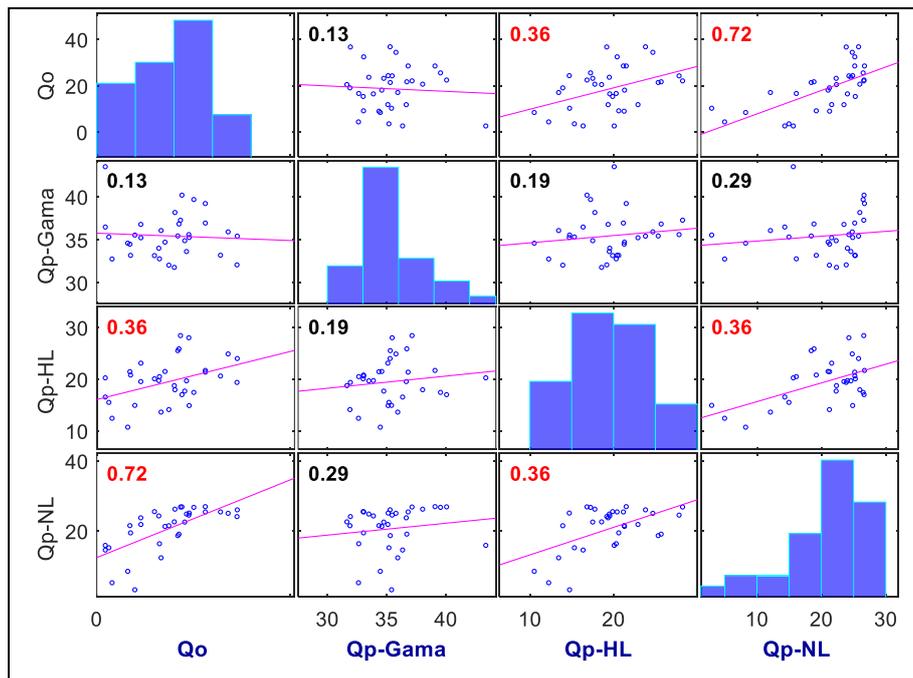


Figura 6.33 - Matriz de correlação tipo Spearman entre as vazões observadas e as médias previstas nos 60 anos do período de validação para pelos modelos Gama, HL e NL. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL. Qo: vazão observada; Qp: vazão prevista; Gama: modelo Gama; HL: HMM Log-Normal de 2 estados ocultos; NL: NHMM Log-Normal de 2 estados ocultos que inclui o indicador NINO3 de dezembro do ano anterior.

Para visualizar o comportamento dessas previsões pontuais, respectivamente, as figuras 6.34 a 6.38 apresentam as observações e as médias previstas dos anos em que o $CRPSS \geq 0,18$ CRPSSm do modelo NL para os modelos Gama, AR(1), HL, ARX(1)-ND e NL. Essas figuras mostram também os IC de 95% para essas previsões. A construção desses intervalos foi diferente para cada um dos modelos de previsão de vazão e se comenta a seguir.

No caso do modelo Gama no ajuste a essa distribuição o software Matlab 2016 entrega os IC de 95% para os parâmetros da distribuição, logo os IC foram calculados como $IC_{I/S} = \mu_p \pm (\alpha_{I/S} \times \beta_{I/S})$, em que $\alpha_{I/S}$ e $\beta_{I/S}$ dependem do IC calculado ser inferior ou superior. Já para o AR(1) e ARX(1) a função *forecast* desse mesmo software apresenta como resultado os valores da variância dos erros das previsões (σ_e^2), logo, os IC foram calculados como $IC_{I/S} = \mu_p \pm (1,96 \times \sqrt{\sigma_e^2})$ esse procedimento de cálculo é indicado por Matlab (2019). Por fim, para os HMMs os IC foram calculados tomando o procedimento que Hyndman e Athanasopoulos (2019) mencionam, assim, os IC inferior e superior, respectivamente, correspondem a $IC_{I/S} = \mu_p \pm (1,96 \times \sqrt{\sigma^2})$, em que σ^2 é a variância das previsões e se calcula com base nas FDPs previstas, essa variância foi calculada a partir das propriedades do valor esperado veja, Naghettini e Pinto, (2007), pag.75.

Dessa maneira, quando se comparam observações e previsões pontuais, observe na figura 6.34 como o modelo Gama gera previsões bastante acima das geradas pelos outros modelos. Nesse modelo os IC começam bastante amplos e diminuem na medida que a série tem mais dados para o ajuste, por exemplo, nos anos da década de 1990. Esse modelo também gera previsões menos variáveis e mais constantes, quando comparadas às geradas pelos modelos HL e NL (figuras 6.36 e 6.38). Os modelos autoregressivos (figuras 6.35 e 6.37) produzem médias inferiores às geradas pelo modelo Gama, porém, mais variáveis que as produzidas por esse modelo.

Note-se, ainda nas figuras 6.36 e 6.38, que as médias previstas pelos modelos HL e NL apresentam valores próximos às observações valores próximos às médias previstas produzidas pelos modelos autoregressivos. Observa-se também que o modelo NL consegue entregar médias bastante precisas que simulam o comportamento das vazões observadas, por exemplo, os anos 1955 e 1971. Apesar disso, os modelos HL e NL têm dificuldades em

prever valores mais altos, por exemplo os anos 1973 e 1977. Veja também que para os modelos HL e NL os IC são mais estreitos e isso é mais notável no NL, isso mostra o efeito do uso de informação climática nesse modelo.

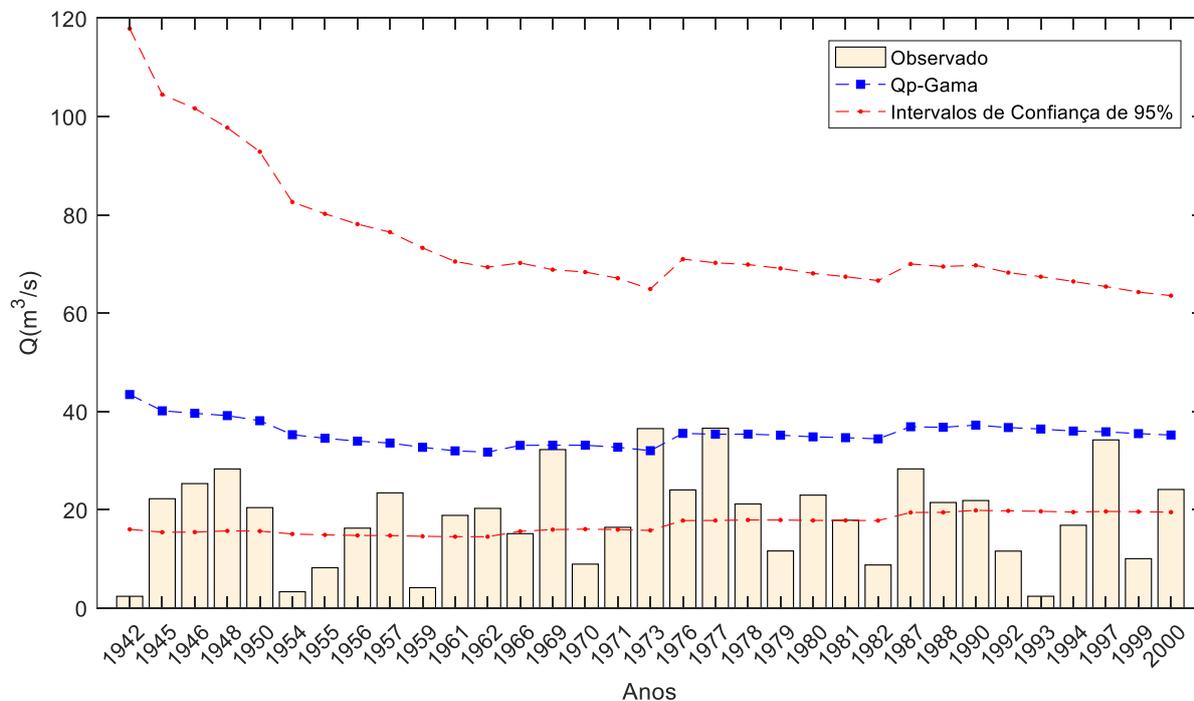


Figura 6.34 – Valores das observações e previsões produzidas pelo modelo Gama (Qp-Gama) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL

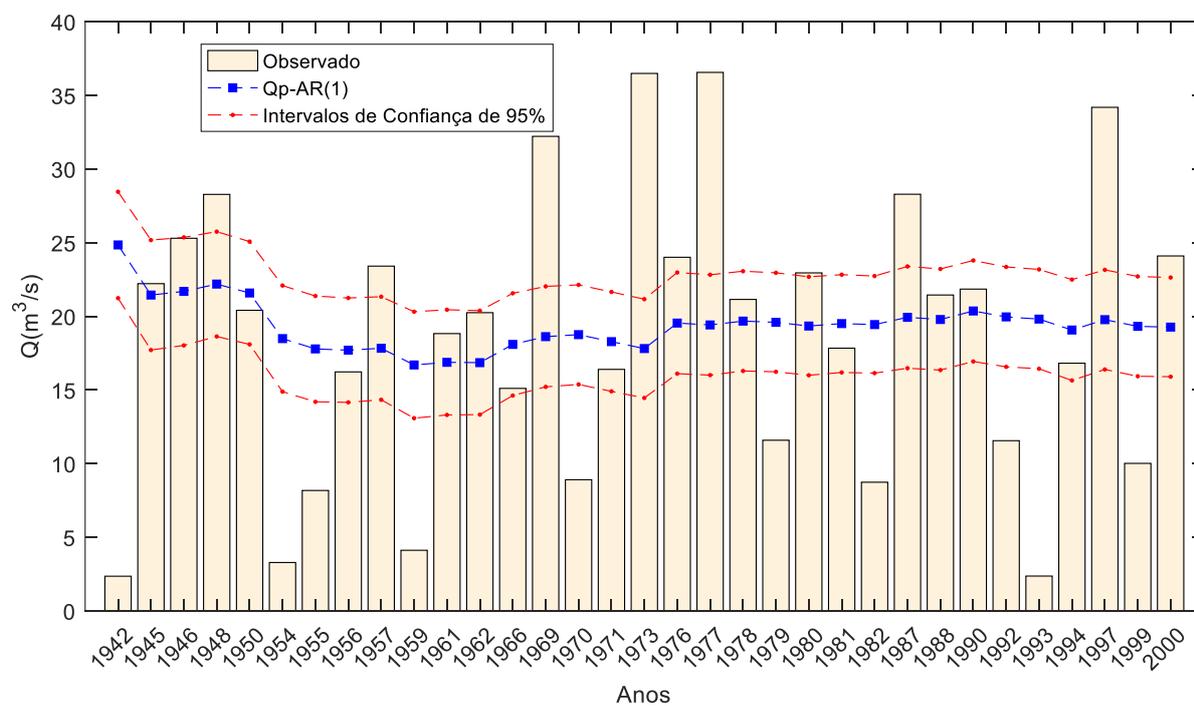


Figura 6.35 – Valores das observações e previsões produzidas pelo modelo AR(1) (Qp-AR(1)) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL

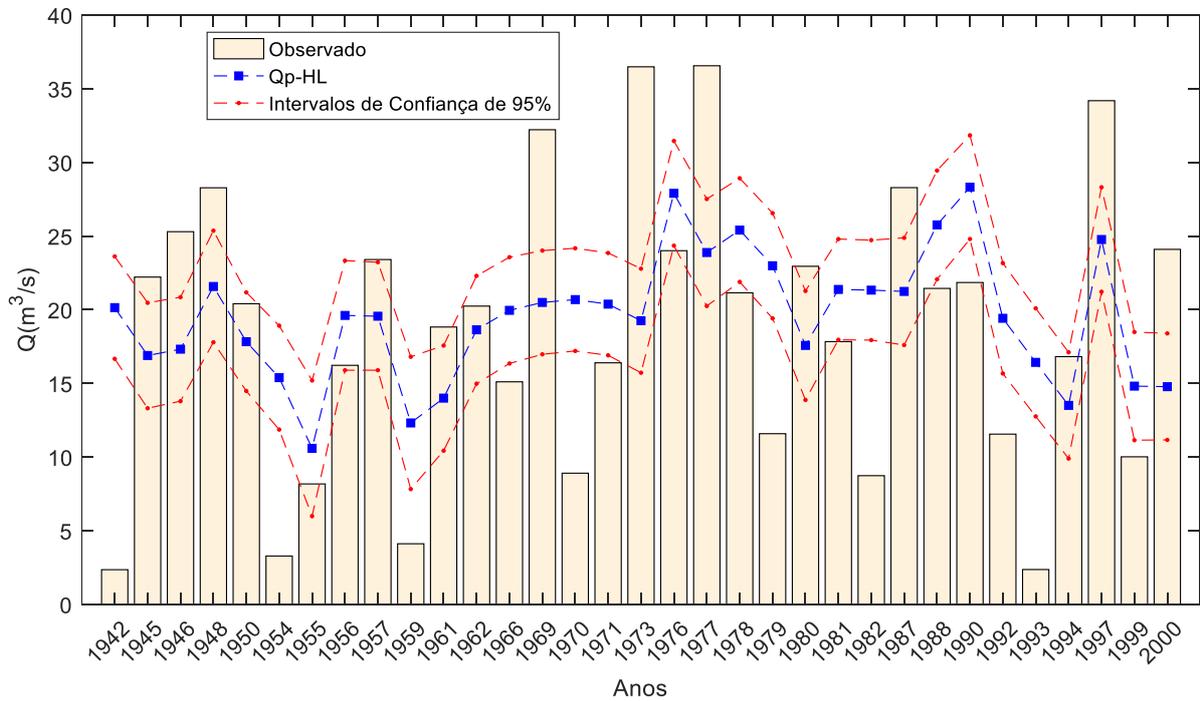


Figura 6.36 – Valores das observações e previsões produzidas pelo modelo HL (Qp-HL) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL

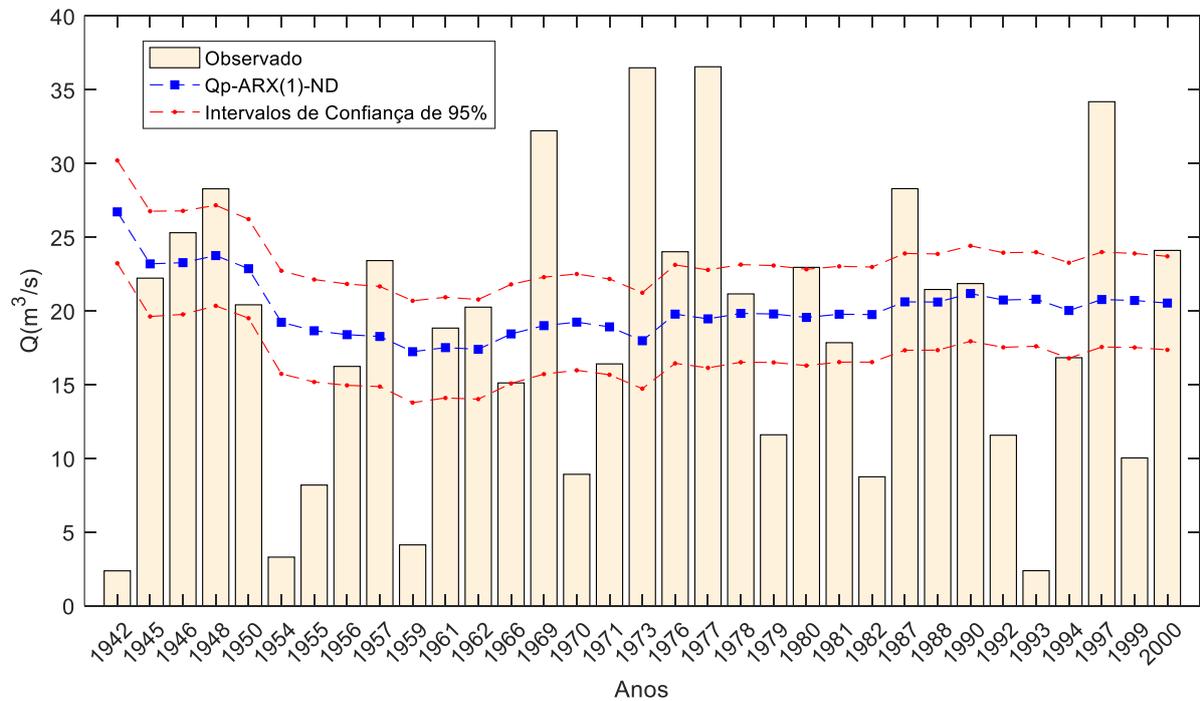


Figura 6.37 – Valores das observações e previsões produzidas pelo modelo HL (Qp-ARX(1)-ND) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL

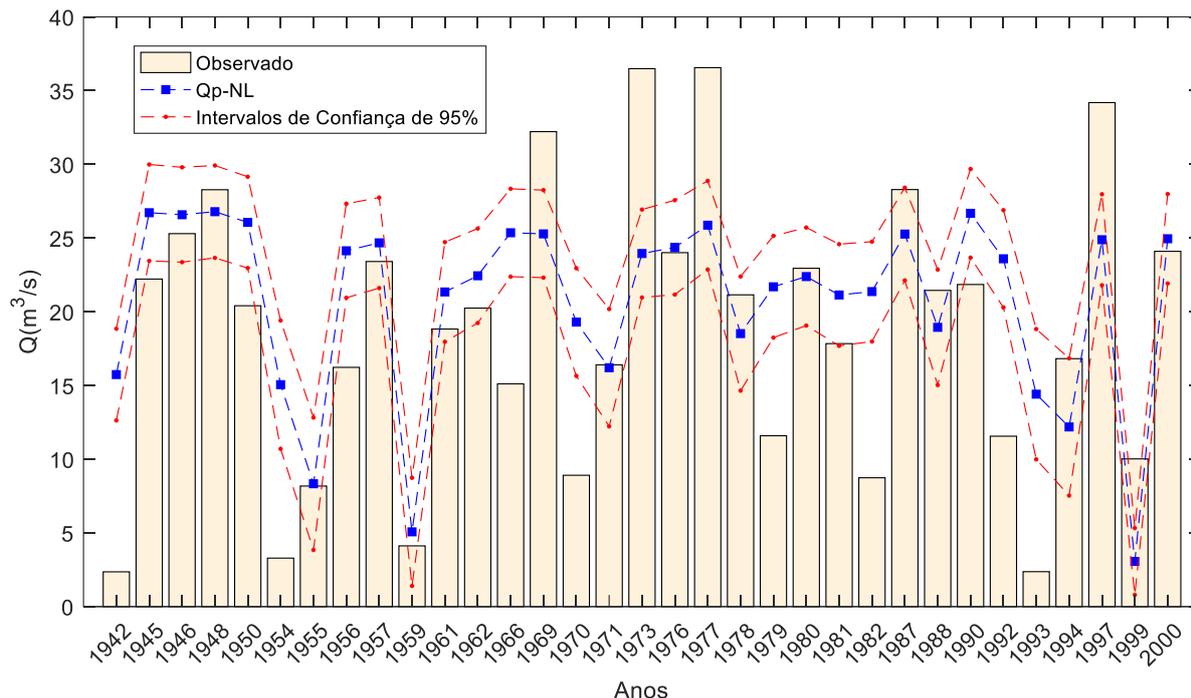


Figura 6.38 – Valores das observações e previsões produzidas pelo modelo NL (Qp-NL) para os anos em que $CRPSS \geq CRPSS_m$ do modelo NL

Dessa forma, a pesar de algumas deficiências, o modelo NL entrega médias previstas bastante próximas às observações, portanto, esse modelo se apresenta como uma alternativa para gerar valores médios das aflúências. Por outro lado, utilizar o modelo Gama é uma medida mais conservadora e inadequada, no sentido que não representa variabilidade e garante uma aflúência constante ao redor de $40 \text{ m}^3/\text{s}$, o que resulta inadequado. Modelos autoregressivos apresentam-se como uma alternativa menos conservadora, porém, não entregam a variabilidade dos modelos HL e NL, sendo esse último modelo mais adaptado à variabilidade das observações.

Os resultados mostrados nesta seção e na seção 6.31, indicam que o modelo NL, escolhido como produto final deste estudo, possui um bom desempenho quando analisadas previsões probabilísticas, embora bastante ruim quando se analisam previsões pontuais para o período de verificação. Embora, esse desempenho nas previsões pontuais é bastante bom quando se selecionam os anos em que esse modelo teve bom desempenho para previsões probabilísticas. Dessa forma, o modelo NL consegue capturar aceitavelmente a variabilidade observada e consegue prever anos secos. Embora, esse modelo apresenta dificuldades para a previsão de anos úmidos, resulta ser melhor que os modelos autoregressivos para esse tipo de anos. Logo,

o modelo NL pode ser utilizado no mês de janeiro para realizar a previsão da afluência média anual do próximo ano, tendo em consideração a disponibilidade do indicador ND – indicador NINO3 do mês de dezembro do ano anterior à realização da previsão. As tabelas 6.7 e 6.8 resumem os resultados das métricas de desempenho desse modelo.

Tabela 6.7 – Resumo do desempenho do modelo NL. Os intervalos de confiança de 95% foram calculados via *Bootstrap*.

Métrica	Valor	Intervalos de Confiança de 95%	
		Inferior	Superior
CRPSSm	0.18	0.12	0.24
CRPSSms	0.26	0.17	0.37
CRPSSmu	-0.03	-0.15	0.07
EMA	21.2	12.8	31.8
EMPA (%)	156.16	95.91	224.47
NS	-0.09	-0.20	0.02
DM	1.90	1.46	2.50
Correlação Spearman	0,07 (valor p = 0,60)	-0,18	0,32

Tabela 6.8 – Resumo do desempenho do modelo NL. Os intervalos de confiança de 95% foram calculados via *Bootstrap*. Valores para os anos em que o CRPSS é maior do que o CRPSSm do modelo NL.

Métrica	Valor	Intervalos de Confiança de 95%	
		Inferior	Superior
EMA	5,75	4,32	7,11
EMPA (%)	72,96	32,90	119,93
NS	0,40	0,10	0,56
DM	0,94	0,61	1,38
Correlação Spearman	0,72 (valor p=2,4E-06)	0,51	0,83

6.4 - VALIDAÇÃO CRUZADA

Essa seção apresenta os resultados da validação cruzada para o modelo NL. Para a abordagem *K-fold* utilizou-se $K=10$, seguindo a metodologia da seção 5.5.1. Com relação ao método *out-of-sample*, o tipo de previsão realizado neste estudo (figura 5.5) no fundo consiste na aplicação desse método, já que, inicialmente um conjunto de 30 anos é utilizado para ajustar o modelo e um conjunto de um ano é utilizado para a validação. Dessa forma, para o período de validação 60 vezes é realizado o método *out-of-sample* e as métricas de desempenho para esse tipo de validação cruzada são os valores já apresentados nas seções 6.3.1 e 6.3.2. Dessa

forma, para a validação cruzada *K-fold* a tabela 6.9 apresenta os valores das métricas de desempenho de previsões probabilísticas e pontuais do modelo NL, esses valores representam o valor médio para os 10 períodos de validação.

O valor do CRPSSm para a validação cruzada é levemente superior ao valor mostrado anteriormente (CRPSSm = 0,18 tabela 6.7), muito provavelmente porque sempre para o ajuste do modelo tinham-se sempre 80 anos, o que melhora o ajuste. O comportamento do modelo NL, gera, em média, valores similares ao CRPSSm da última década (década de 1990) da figura 6.16 (CRPSSm = 0,25), quando se tem 80 ou mais anos para o ajuste.

O valor do CRPSSms foi similar ao encontrado anteriormente (CRPSSms = 0,26 tabela 6.7). Surpreende é o valor do CRPSSmu = 0,25, o qual é diferente do valor negativo encontrado nas análises da seção 6.3.1 (CRPSSmu = -0,03 tabela 6.7). Tal resultado, pode estar também relacionado ao número de dados empregados no ajuste e também ao número de anos úmidos, nesse caso para os 90 anos da série completa existem 31 anos úmidos.

Tabela 6.9 – Resultados da validação cruzada *K-fold* para o modelo NL. K=10. Os intervalos de confiança de 95% foram calculados via *Bootstrap*.

Métrica	Valor	Intervalos de Confiança de 95%	
		Inferior	Superior
CRPSSm	0,26	0,19	0,32
CRPSSms	0,31	0,20	0,44
CRPSSmu	0,25	0,18	0,32
EMA	25,33	17,26	33,92
EMPA (%)	148,35	83,91	238,85
NS	-0,56	-1,35	-0,06
DM	2,21	1,52	3,41
Correlação Spearman	0,30 (valor p=0,44)	-0,01	0,62

Em geral, veja a tabela 6.10, na validação cruzada nunca se tiveram valores negativos da métrica CRPSS. O pior desempenho foi quando k=5 (década de 1950), nesse ano o CRPSSm foi de 0,07, observe na figura 6.16 que nesse período o desempenho médio é maior do que 0,07 o CRPSSm para essa década é 0,16. Enquanto, sem anos úmidos na década de 1950 e baixos valores para as décadas de 1930 (CRPSS = 0,07) e 1940 (CRPSS = 0,09), ajudou para o alto valor do CRPSSmu, já que, a média tomou só 8 décadas, com 6 décadas com valores do CRPSSmu acima de 0,28. Isso, mostra que o NL não é tão bom para anos úmidos como tinha

sido mostrado na seção 6.3.1. Além disso, provavelmente a informação climática e a série utilizada na previsão (série cortada) não ajudam muito na previsão nas décadas de 1920 e 1930 e em geral na década de 1950.

Tabela 6.10 – Valores do CRPSS para os períodos da validação cruzada *K-fold*. $K = 10$.

<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
Década	<i>1910</i>	<i>1920</i>	<i>1930</i>	<i>1940</i>	<i>1950</i>	<i>1960</i>	<i>1970</i>	<i>1980</i>	<i>1990</i>
<i>1</i>	0,19	0,41	0,15	0,17	0,03	0,17	0,38	0,26	0,07
<i>2</i>	0,21	0,27	0,75	0,44	0,04	0,15	0,06	0,30	0,15
<i>3</i>	0,67	0,24	0,07	0,69	0,12	0,20	0,10	0,24	0,65
<i>4</i>	0,26	0,24	0,10	0,01	0,04	0,15	0,87	0,97	0,19
<i>5</i>	0,45	0,53	0,04	0,07	0,08	0,53	0,09	0,28	0,27
<i>6</i>	0,66	0,22	0,07	0,06	0,03	0,11	0,13	0,16	0,53
<i>7</i>	0,23	0,71	0,11	0,14	0,03	0,67	0,08	0,17	0,07
<i>8</i>	0,14	0,28	0,12	0,13	0,03	0,16	0,47	0,47	0,05
<i>9</i>	0,11	0,31	0,05	0,11	0,21	0,12	0,30	0,61	0,98
<i>10</i>	0,69	0,25	0,04	0,11	0,07	0,35	0,25	0,11	0,08
CRPSSm	<i>0,36</i>	<i>0,35</i>	<i>0,15</i>	<i>0,19</i>	<i>0,07</i>	<i>0,26</i>	<i>0,27</i>	<i>0,36</i>	<i>0,30</i>
CRPSSms	<i>0,25</i>	<i>0,71</i>	<i>0,22</i>	<i>0,31</i>	<i>0,08</i>	-	<i>0,24</i>	<i>0,27</i>	<i>0,38</i>
CRPSSmu	<i>0,32</i>	<i>0,32</i>	<i>0,07</i>	<i>0,09</i>	-	<i>0,30</i>	<i>0,28</i>	<i>0,35</i>	<i>0,29</i>

Para as métricas EMA e EMPA, os valores mostrados na tabela 6.9 são similares aos encontrados na tabela 6.7. Para o coeficiente de NS, DM e a correlação existe um aumento dos valores com relação aos encontrados na tabela 6.7. Ressalta-se, a diferença mais negativa do coeficiente de NS. Embora pelos IC podem existir valores que sejam similares aos encontrados na tabela 6.7.

Os resultados da validação cruzada tipo *K-fold*, mostram que o modelo consegue se comportar, no quesito desempenho, de forma similar e levemente superior quando se utiliza a estratégia de previsão empregada nesta tese. Isso, indica que o modelo pode ser generalizado e aplicado na previsão de aflúncias para o reservatório Orós, no estado do Ceará. Logo, o modelo NL, pode ser utilizado para realizar previsões em janeiro, assim que o indicador climático ND estiver disponível. Vale ressaltar que, deve-se ter um especial cuidado com a acurácia do modelo NL para a previsão de anos úmidos.

7 - CONCLUSÕES

Esta tese explorou a modelagem de previsão utilizando modelos de Markov com estados ocultos (*Hidden Markov Models* – HMMs) para a previsão, com antecedência mensal, de afluências, médias anuais, empregando indicadores climáticos, associados com as oscilações interanuais dos oceanos Pacífico e Atlântico Tropical. Utilizaram-se os indicadores NINO3 e o Dipolo do Atlântico (DA).

Os HMMs são modelos estocásticos que permitem representar a variação de estados hidroclimáticos. Tais estados, são não observados, ou ocultos, e variam no tempo utilizando uma cadeia de Markov. Essa representação é útil quando se analisam rios que apresentam uma variação entre regimes de vazões mais secas, variando para regimes de vazões mais úmidas. Uma versão dos HMMs, ainda, permite que os estados possam estar associados a flutuações climáticas.

Dessa forma, utilizaram-se três classes de modelos HMMs para a previsão de afluências: i) HMMs em que a variação entre estados ocultos foi influenciada por variáveis externas – NHMMs, nesses modelos a variação de estados não é estacionária no tempo e varia seguindo a influência de indicadores climáticos; ii) HMMs baseados nos NHMMs, embora, as distribuições associadas aos estados ocultos seguem uma estrutura de autocorrelação – AR-NHMMs; e iii) HMMs nos quais a cadeia de Markov é homogênea ou invariável no tempo, porém, as distribuições de probabilidade que dependem dos estados ocultos não são estacionárias, já que, são influenciadas por variáveis climáticas IHMMs. A modelagem de previsão de vazões utilizando NHMMs e AR-NHMMs tem sido pouco explorada na literatura, e no Brasil não tem sido utilizada. Enquanto que o uso de IHMMs para a previsão de afluências não tem sido registrado na literatura.

Este estudo, utilizou como estudo de caso o reservatório Orós, Estado do Ceará, região Nordeste do Brasil. Nessa região são observados rios que apresentam alternância de vazões mais secas para vazões mais úmidas. Logo, a previsão de vazões na região tem grande importância, já que, existe uma forte influência de tele-conexões climáticas sobre essa variação dos rios da região, o que contribui à ocorrência de fortes períodos de seca. Nesse panorama, o uso de modelos sofisticados que permitam incluir a influência climática nas

vazões observadas pode ajudar a melhorar a representatividade desses rios e, dessa forma, produzir previsões de aflúências mais acuradas, que permitam auxiliar a gestão dos recursos hídricos, evitando tomar medidas conservadoras na alocação da água.

Uma análise de correlação entre indicadores climáticos e a série histórica de aflúências médias anuais ao reservatório Orós (período 1911-200) entregou os preditores climáticos, indicadores NINO3 e DA, a ser incluídos nos HMMs como variáveis externas. Nessa análise foram explorados vários períodos mensais, trimestrais e anuais, desses indicadores. Os resultados indicaram como preditores climáticos os índices NINO3 do mês de dezembro do ano anterior à ocorrência de uma observação (ND) e os indicadores do DA do mês de agosto, período julho a setembro, e a média anual do ano anterior à ocorrência de uma observação (indicadores DAA, DAJAS e DAM, respectivamente). Para efeitos de comparação, empregou-se também o indicador de janeiro do mesmo ano em que ocorre a aflúência observada (NJ) que resultou estar correlacionado com a série observada.

Com esses indicadores climáticos ajustaram-se 132 modelos HMMs, que correspondem a modelos que utilizaram de forma individual e combinada os indicadores climáticos ND, NJ, DAA, DAJAS e DAM. Esses modelos têm diferenças também no tipo de distribuições dependentes dos estados ocultos Log-Normal e Gama, nos NHMMs e IHMMs; Log-Normal nos AR-NHMMs, e no número de estados ocultos HMMs com 2 e 3 estados ocultos. Essa restrição de número de estados obedece à ideia de oferecer modelos parcimoniosos com poucos parâmetros, já que, um modelo com muitos parâmetros aumenta a incerteza das previsões. Além disso, HMMs com 4 ou mais estados entregam maior número de informações a serem analisadas, o que foi inconveniente para o caráter exploratório desta tese.

Para o valor do CRPSS médio, durante o período de validação de 60 anos ($CRPSS_m$), os HMMs foram melhores que a climatologia ($CRPSS_m > 0$). Os resultados mostraram que HMMs com 3 estados ocultos, e com distribuições Gama, não trazem maior ganho para o valor do $CRPSS_m$, quando comparados a HMMs com 2 estados ocultos e Log-Normais. Valores do CRPSS mostraram também que os HMMs apresentam um desempenho variável, logo, que entregaram valores variáveis das métricas $CRPSS_m$, $CRPSS_{m33}$ e $CRPSS_{m66}$ de anos secos definidos pelo percentil de 33% e $CRPSS_{m66}$ e $CRPSS_{m33}$ de anos úmidos definidos pelo percentil de 66%.

Os valores do CRPSSms mostraram que os HMMs resultaram serem melhores para prever anos secos ($CRPSSms > 0,2$), esses resultados são contrários aos encontrados por Gelati et al., (2010b), que mostrou que os NHMMs com FDPs, similares às dos IHMMs, são melhores para prever vazões altas para uma bacia do Equador. A razão desse desempenho para valores de vazões mais altas, não foi mencionada por esses autores, mas pode estar relacionada à relação do indicador climático utilizado (NINO 1+2) ser mais correlacionado com esse tipo de vazões. Enquanto, para anos mais úmidos os HMMs apresentam deficiências ($CRPSSmu < 0$). Nesse caso os IHMMs apresentaram melhor performance ($CRPSSmu > 0$), que os NHMMs e AR-NHMMs, quando se incluíram indicadores DA (modelo IL), isso mostra o potencial uso dos IHMMs para a previsão desse tipo de anos especialmente quando o DA é utilizado.

O HMM com comportamento menos variável, valores mais constantes do CRPSSm, CRPSSms e CRPSSmu, foi o modelo NL – NHMM com indicador ND com 2 estados ocultos. Esse modelo teve desempenho bastante eficiente para o período de validação, mostrando adequada qualidade de previsões probabilísticas. Porém, esse modelo teve um desempenho inadequado segundo as métricas de previsões pontuais EMA, EMPA, NS, DM e Correlação, pois, seu desempenho no período de verificação das previsões foi bastante baixo. No entanto, esse desempenho ruim foi similar ao desempenho de modelos sem informações climáticas: HL HMM Log-Normal, AR(1) e o modelo Gama, e à performance do ARX(1) com o indicador ND. Dessa forma, uma seleção de anos baseada no desempenho do modelo NL anos com $CRPSS \geq 0,18$ mostrou que o NL entrega previsões pontuais bastante adequadas, uma vez que os valores das métricas pontuais melhoram consideravelmente.

Na validação cruzada o NL apresentou resultados satisfatórios, com métricas de desempenho probabilístico e pontual, similares às obtidas nas para previsões que seguem a estratégia de previsão usada nesta tese, embora, com as deficiências para prever anos úmidos e valores não tão adequados de métricas pontuais. Não obstante, o modelo NL pode ser utilizado para realizar a previsão de vazão afluente ao reservatório Orós, no mês de janeiro do ano corrente, assim que ficar disponível o indicador climático ND.

Este estudo não procurou escolher o melhor HMM, mas, sim explorar e ilustrar a aplicabilidade dos HMMs na previsão de vazões para um ano na frente, uma modelagem de previsão que não tem sido utilizada no Brasil. Os resultados aqui apresentados, são

contribuições para a utilização de HMMs na previsão de afluências, principalmente em regiões que apresentam correlações com indicadores climáticos de larga escala, como indicadores do ENSO ou os relacionados às TSMs do oceano Atlântico Tropical.

A modelagem de previsão abordada nesta tese mostrou maior flexibilidade no uso de informações climáticas, e, dessa forma, é um caminho para entender melhor as interações entre o clima e os rios de regiões com forte influência da variabilidade climática, logo, que tal modelagem conseguiu, com algumas deficiências, gerar previsões que capturam melhor a variabilidade encontrada nas observações e podem representar sua persistência, embora, analisar a persistência não ter sido o escopo desta tese, principalmente obteve-se melhor previsão de anos secos. Contudo, um sistema baseado no NL é potencialmente útil para realizar previsões de afluências ao reservatório Orós e, portanto, ajudar aos tomadores de decisão do Estado do Ceará quanto ao processo de alocação de água, evitando tomar decisões conservadoras.

Para aumentar o desempenho geral, e para prever melhor os anos úmidos, pode ser interessante explorar outras formas de HMMs. Por exemplo, utilizar uma estrutura de estados ocultos multinível a qual é possível empregando HMMS Hierárquicos (HHMMs). Logo, poderia se empregar um HHMM com uma cadeia de Markov superior que geraria uma HMM com uma cadeia de Markov de ordem inferior que produziria as observações, essas cadeias teriam a influência de indicadores decenais e interanuais, respectivamente. Na escala decenal poderiam se incluir indicadores climáticos relacionados com oscilações ODP, OMA e OAN, que segundo a literatura influenciam o regime de chuvas do NEB. Enquanto, na escala interanual, outros indicadores, que também influenciam a quadra chuvosa o NEB, que poderiam ser utilizados seriam o indicador da região Atlântico 3 (ATL3), velocidades do vento, variáveis relacionadas à produção de precipitação ou indicadores multivariados como o MEI.

Por fim, seria interessante aplicar a modelagem de previsão aqui ilustrada a outros locais, em regiões com similares características climáticas, inclusive pode-se testar uma estrutura de correlação entre reservatórios de uma mesma região. Também outros métodos de validação cruzada podem ser explorados, por exemplo, uma validação cruzada por janelas móveis, na qual um conjunto de dados móvel no tempo é dividido em subconjunto de ajuste e validação.

REFERÊNCIAS BIBLIOGRÁFICAS

- Akaike, H. (1974). "A new look at the statistical model identification, *IEEE Trans. Automat. Control*" 716–72.
- Akintug, B., and Rasmussen, P. F. (2005). "A Markov switching model for annual hydrologic time séries." *Water Resources Research*, 41(9), 1–10.
- Alexandre, A. M. B. (2012). "*Previsão De Vazões Mensais Para O Sistema Interligado Nacional Utilizando Informações Climáticas*". Tese de Doutorado em engenharia civil, Universidade Federal do Ceará.
- Altman, R. M. (2009). "Lecture 26 : Models for Gamma Data." *Generalized Linear Model Notes*, 1–4
- Alves, J. M. B., Souza, E., Repelli, C. A., Vitorino, M. Ferreira, N. (1997): "Episódios de La Niña na bacia de oceano Pacífico Equatorial e a distribuição sazonal e intra-sazonal das chuvas no setor norte do Nordeste Brasileiro". *Revista Brasileira de Meteorologia*, 12, 63-76.
- ANA. (2010). *Atlas Brasil: abastecimento urbano de água: resultados por estado*. Agência Nacional de Águas (Brasil). Consórcio Engecorps/Cobrape III. Superintendência de Planejamento de Recursos Hídricos - SPR, Brasília - DF.
- ANA. (2018). *Conjuntura dos recursos hídricos no Brasil 2018: informe anual*. Agência Nacional de Águas. Brasília:
- Andreoli, R. V., and 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), 63–74.
- 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.
- Araújo, C. B. C. de, Neto, S. A. D., and Filho, F. de A. S. (2015). "Previsão de Vazões para o Açude Orós/Ce a partir de dados Hidrometeorológicos utilizando Perceptrons." *Revista Brasileira de Meteorologia*, 30(1), 37–46.
- Awwad, H.M.; Valdes, J.B. (1992). "Adaptive parameter estimation for multisite hydrologic forecasting", *Journal of Hydraulic Engineering*. 118(9), pp.1201-1221.
- Awwad, H.M.; Valdes, J.B.; Restrepo, P.J. (1994). "Streamflow forecasting for Han River

- basin, Korea", *Journal of Water Resources Planning and Management*, 120(5), pp. 651- 673.
- Bai, Y., Chen, Z., Xie, J., and Li, C. (2016). "Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models." *Journal of Hydrology*, 532, 193–206.
- Baldi, P. and S. Brunak (2001). *Bioinformatics: The Machine Learning Approach*. Second ed.). MIT Press.
- Bayazit, M., and Bulu, A. (1988). "Complex Markov Models to Simulate Persistent Streamflows." *Journal of Hydrology*, 103, 199–207.
- Bellone, E., Hughes, J. P., and Guttorp, P. (2000). "A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts." *Climate Research*, 15, 1–12.
- Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). "A note on the validity of cross-validation for evaluating autoregressive time series prediction." *Computational Statistics and Data Analysis*, Elsevier B.V., 120, 70–83.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Science+Business.
- Bogner, K.; Kalas, M. (2008). "Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment". *Atmospheric Science Letters*. V.9 p.95–102.
- Box, G. E. P.; Jenkins, G. M. e Reinsel, G. C., (1994). *Time Séries Analysis: forecasting and control*, 3.rd ed., Prentice Hall, New Jersey, 587p
- Box, G. E. P., Reinsel, G. M. J. G. C., and Ljung, G. M. (2016). *Time Séries Analysis Forecasting and Control*. *Wiley Séries In Probability And Statistics*.
- Bracken, C., Rajagopalan, B., and Woodhouse, C. (2016). "A Bayesian hierarchical nonhomogeneous hidden Markov model for multisite streamflow reconstructions." *Water Resources Research*, 52(10), 7837–7850.
- Bracken, C., Rajagopalan, B., and Zagona, E. (2014). "A hidden Markov model combined with climate indices for multidecadal streamflow simulation." *Water Resources Research*, 50, 1–11.
- Bracken, C. W. (2011). "Seasonal to Inter-Annual Streamflow Simulation and Forecasting on the Upper Colorado River Basin and Implications for Water Resources Management." University of Colorado.
- Bras, R. L.; Rodriguez-Iturbe, I. (1992). *Random Functions and Hydrology*. Addison-

Wesley Publishing Company.

- Brabo Alves J. M. Recentes variações climáticas no Nordeste do Brasil com ênfase a precipitação: relações com ENOS, Dipolo de temperatura da superfície do Mar no Atlântico Tropical e a Oscilação decadal do Pacífico. *Rev. Geogr.* 2012; 29(3): 147-166.
- Cabral, S. L. (2014). “*Intervalo De Tempo De Máxima Previsibilidade No Acoplamento De Modelos Climáticos E Hidrológico Para O Gerenciamento De Reservatório.*” Programa de Pós-Graduação em Engenharia Civil. Área de concentração: Recursos Hídricos. Tese de doutorado. Universidade Federal do Ceará.
- Campos, J. N. B. (1996). *Dimensionamento de Reservatórios: o método do diagrama triangular de regularização*, Fortaleza, Edições UFC, 71p.
- Campos, J. N. B. (2014). “Secas e políticas públicas no semiárido : ideias , pensadores e períodos.” *Estudos Avançados*, 28(82), 65–88.
- Capistrano, V. B., and Kayano, M. T. (2012). “Influências Da Oscilação Multidecenal Do Atlântico Na Precipitação Do Norte Da Região Nordeste Do Brasil.” *Anais do XVII Congresso Brasileiro de Meteorologia*
- Cardoso, G. B. B., Martins, E. S. P. R., and Filho, F. de A. de S. (2006). “Uso de Otimização/Simulação e Previsão de Afluências na Operação Tática dos Reservatórios do Sistema Jaguaribe-Metropolitano - CE.” *Revista Brasileira de Recursos Hídricos*, 11(4), 175–186.
- Cardoso, G. B. B., Martins, E. S. P. R., and Filho, F. de A. de S. (2009). “Previsão de Afluências dos Reservatórios do Sistema Jaguaribe-Metropolitano - CE.” em: *Informacoes Climaticas em Recursos Hídricos*, 1 Edicao, FUNCEME, 160p.
- Castellano-Méndez, M., González-Manteiga, W., Febrero-Bande, M., Manuel Prada-Sánchez, J., and Lozano-Calderón, R. (2004). “Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box–Jenkins and neural networks methods.” *Journal of Hydrology*, 296(1–4), 38–58.
- Ceará, A. L. do E. do. (2008). *Cenário Atual Dos Recursos Hídricos Do Ceará*. Conselho de Altos Estudos e Assuntos Estratégicos, Fortaleza - Ceará.
- Chang, C. (2009). “A non-linear ARMAX for short term load forecasting.” *Journal of Statistics and Management Systems*, 37–41.
- Chatfield, C. (2003) *The Analysis of Time Séries: An Introduction*, 6th edn. Boca Raton, FL: Chapman and Hall/CRC.
- Chen, S., Shin, J. Y., and Kim, T.-W. (2016). “Probabilistic forecasting of drought: a

- hidden Markov model aggregated with the RCP 8.5 precipitation projection.” *Stochastic Environmental Research and Risk Assessment*, Springer Berlin Heidelberg, 1–16.
- Chen, L., Singh, V. P., Lu, W., Zhang, J., Zhou, J., & Guo, S. (2016). "Streamflow forecast uncertainty evolution and its effect on real-time reservoir operation". *Journal of Hydrology*, 540, 712–726.
- COGERH, C. de G. dos R. H. (2011). *Inventário Ambiental Açude Orós*. República Federativa do Brasil Governo do Estado do Ceará Secretaria dos Recursos Hídricos – SRH Companhia de Gestão dos Recursos Hídricos - COGERH, Fortaleza - Ceará.
- Córdoba-machado, S., Palomino-lemus, R., Gámiz-, S. R., Castro-díez, Y., and Esteban-parra, M. J. (2016). “Seasonal streamflow prediction in Colombia using atmospheric and oceanic patterns.” *Journal of Hydrology*, Elsevier B.V.
- Costa, J. A., and Silva, D. F. Da. (2017). “Distribuição espaço-temporal do Índice de anomalia de chuva para o Estado do Ceará.” *Revista brasileira de Geografia*, 50(4), 9–57.
- 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.
- Damberg, L., and AghaKouchak (2013), Global trends and patterns of drought from space, *Theor. Appl. Climatol.*, 117, 441–448.
- Davidian, M. (2008). “Chapter 11: Generalized linear models for nonnormal response.” *Notes of Applied Longitudinal Data Analysis Course*
- Davis, R. E. (1976). "Predictability of Sea Surface Temperature and Sea Level Pressure Anomalies over the North Pacific Ocean". *Journal of Physical Oceanography*, 6(3), 249–266.
- Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC.
- Durbin, R., S. Eddy, A. Krogh, and G. Mitchison (1998). *Biological Sequence Analysis*. Cambridge University Press.
- Erkyihun, S. T., Rajagopalan, B., Zagona, E., Lall, U., & Nowak, K. (2016). "Wavelet-based time series bootstrap model for multidecadal streamflow simulation using climate indicators". *Water Resources Research*, 52(5), 4061–4077.
- Fan, F. M., Ramos, M.-H., and Collischonn, W. (2015). “Sobre o uso de previsões

- hidrológicas probabilísticas para tomada de decisão.” *Revista Brasileira de Recursos Hídricos*, 20(4), 914–926.
- Feng, J., Chen, W., Tam, C.-Y., and Zhou, W. (2011). “Different impacts of El Niño and El Niño Modoki on China rainfall in the decaying phases.” *International Journal of Climatology*, 31(14), 2091–2101.
- Feng, S., Hu, Q., and Oglesby, R. J. (2011). “Influence of Atlantic sea surface temperatures on persistent drought in North America.” *Climate Dynamics*, 37(3), 569–586.
- 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.
- Fernandes, W. de S. (2012). *Avaliação do impacto das Mudanças Climáticas na Oferta Hídrica da Bacia Hidrográfica do Reservatório Óros usando os Modelos de Mudanças Climáticas do Ipcc-AR4, levando em Consideração as diversas incertezas associadas.* Dissertação de Mestrado – Universidade Federal do Ceará, Centro de Tecnologia, Departamento de Engenharia Hidráulica e Ambiental, Programa de Pós-Graduação em Engenharia Civil.
- Ferreira, A. G., and 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), 15–28.
- 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.
- Fine, S., Singer, Y., and Tishby, N. (1998). “The hierarchical hidden Markov model: Analysis and applications.” *Machine Learning*, 32(1), 41–62.
- Fioreze, A. P., Pinto, A. P., Viana, L., Júnior, J. L., Burte, Y. de B. G. J., Bursztyn, L. M. C. da S. L. G. T. de A. M., Coimbra, M. R. S. C., Nobre, P., e Paulino, (2012). *A Questão da Água no Nordeste*. (Centro de Gestão e Estudos Estratégicos e ANA, eds.), Centro de Gestão e Estudos Estratégicos (Brasil), Brasília - DF.
- Fortin, V., Perreault, L., and Salas, J. D. (2004). “Retrospective analysis and forecasting of streamflows using a shifting level model.” *Journal of Hydrology*, 296(1–4), 135–163.
- Freitas, M. A. de S. , (1996). *Geração de Vazão em Rios de Regiões Semiáridas*. Grupo de Pesquisas em Recursos Hídricos, Meio Ambiente e Computação da Universidade de

- Fortaleza. Disponível em: www.ivdialogo.com/%5Ctrabalhos%5Cdial079.pdf. acesso junho de 2017
- Fu, G., Charles, S. P., and Kirshner, S. (2012). "Daily rainfall projections from general circulation models with a downscaling nonhomogeneous hidden Markov model (NHMM) for south-eastern Australia." *Hydrological Processes*, n/a-n/a.
- FUNCEME.(2017).Sitio web da Fundacao Cearense de Meteorologia e Recursos Hidricos. Disponível em: [ttp://www.funceme.br/index.php/comunicacao/noticias/551-sistemas-atmosf%C3%A9ricos-atuantes-sobre-o-nordeste](http://www.funceme.br/index.php/comunicacao/noticias/551-sistemas-atmosf%C3%A9ricos-atuantes-sobre-o-nordeste), acesso em 28/06/2017.
- Gatto, L. C. S. (1999). *Diagnóstico Ambiental da Bacia do Rio Jáguaibe*. Salvador, Bahia.
- Gelati, E., Christensen, O. B., Rasmussen, P. F., and Rosbjerg, D. (2010a). "Downscaling atmospheric patterns to multi-site precipitation amounts in southern Scandinavia." *Hydrology Research*, 41(3–4), 193.
- Gelati, E., Madsen, H., and Rosbjerg, D. (2010b). "Markov-switching model for nonstationary runoff conditioned on El Niño information." *Water Resources Research*, 46(2), 1–13.
- Giannini, A., Saravanan, R., & Chang, P. (2004). "The preconditioning role of tropical Atlantic variability in the development of the ENSO teleconnection: Implications for the prediction of Nordeste rainfall". *Climate Dynamics*, 22, 839–855.
- Gilleland, E. (2010). "Confidence Intervals for Forecast Verification." National Center For Atmospheric Research.
- Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nuñez, and W. M. Gray (2001), The recent increase in Atlantic hurricane activity: Causes and implications, *Science*, 293, 474–479.
- Gottschalck, J.; Meet the MJO. (2008). Intermountain West Climate Summary, p. 03-24.
- Gottschalck, J. (2018). "What is the MJO, and why do we care?" Disponível em: <https://www.climate.gov/news-features/blogs/enso/what-mjo-and-why-do-we-care> acesso em 27/09/2018
- Govindaraju, R. S., and Ramachandra, A. (2000). *Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology*, Water Science and Technology Library, (R. S. Govindaraju and A. R. Rao, eds.), Springer Netherlands, Dordrecht.
- Graham, N. E. (1994). Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: Observations and model results. *Climate Dynamics*, 10, 135–162

- Greene, A. M., Robertson, A. W., and 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(May), 875–887.
- Greene, A. M., Robertson, A. W., Smyth, P., and 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), 347–359.
- Gregorio, M. de A. (2015). "Monitoramento Da Oscilação Madden Julian Utilizando O Método RMM." Dissertação de Mestrado em Metereologia. Instituto Nacional de Pesquisas Espaciais
- Guimarães, R. C., and Santos, E. G. (2011). "Principles of Stochastic Generation of Hydrologic Time Séries for Reservoir Planning and Design: Case Study." *Journal of Hydrologic Engineering*, 16(11), 891–898.
- Guo, L., Jiang, Z., Ding, M., Chen, W., & Li, L. (2018). "Downscaling and projection of summer rainfall in Eastern China using a nonhomogeneous hidden Markov model". *International Journal of Climatology*.
- Haltiner, J.P.; Salas, J.D. (1998). "Short-Term Forecasting of Snowmelt Runoff Using ARMAX Models". *Journal of the American Water Resources Association – JAWRA*. V.24, Issue 5, p.1083–1089.
- Handoh IC, Bigg GR, Matthews AJ, Stevens DP. (2006b). Interannual variability of the tropical Atlantic independent of and associated with ENSO: Part II. The South tropical Atlantic. *Int. J. Climatol.* 26: 1957–1976
- Handoh, I. C., Matthews, A. J., Bigg, G. R., & Stevens, D. P. (2006a). Interannual variability of the tropical Atlantic independent of and associated with ENSO: Part I. The north tropical Atlantic. *International Journal of Climatology*, 26, 1937–1956.
- Hannan, E. J., Dunsmuir, W. T. M., and Deistler, M. (1980). "Estimation of vector ARMAX models." *Journal of Multivariate Analysis*, 10(3), 275–295.
- Hare, S. R., & Mantua, N. J. (2000). Empirical evidence of North Pacific regime shifts in 1977 and 1989. *Progress in Oceanography*, 47, 103–146
- Hastenrath, S., and Heller, L. (1977). "Dynamics of climate hazards in Northeast Brazil." *Quart. J. Roy. Meteor. Soc.*, 103, 77–92.
- Hastenrath, S. (1984). "Interannual variability and annual cycle: mechanisms of circulation and climate in the tropical Atlantic". *Monthly Weather Review*, **112**, 1097-1107p.
- Hastenrath, S., & Greischar, L. (1993). Circulation mechanisms related to Northeast Brazil rainfall anomalies. *Journal of Geophysical Research*, 98, 5093–5102

- 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.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York, New York, NY.
- Hejazi, M. I., and Markus, M. (2009). "Impacts of Urbanization and Climate Variability on Floods in Northeastern Illinois." *Journal of Hydrologic Engineering*, 14(6), 606–616.
- Hendon, H. H., and M. L. Salby. (1994). The life cycle of the Madden–Julian oscillation. *J. Atmos. Sci.*, 51, 2225–2237
- Hendon, H. H., B. Liebmann, and J. D. Glick. (1998). Oceanic Kelvin waves and the Madden–Julian oscillation. *J. Atmos. Sci.*, 55, 88–101.
- 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, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF.
- Higgins; J.-K. E. Schemm, W. Shi, and A. Leetmaa. (2000). Extreme precipitation events in the western United States related to tropical forcing. *J. Climate*, 13, 793–820.
- Hounsou-Gbo GA, Araujo M, Boulès B, Veleda D, Servain J. (2015). Tropical Atlantic contributions to strong rainfall variability along the northeast Brazilian coast. *Adv. Meteorol.* 2015: 13, doi: 10.1155/2015/902084.
- Hughes, J. P., and Guttorp, P. (1994a). "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena." *Water Resources Research*, 30(5), 1535–1546.
- Huang, B., & Shukla, J. (1997). Characteristics of the interannual and decadal variability in a general circulation model of the tropical Atlantic Ocean. *Journal of Physical Oceanography*, 27, 1693–1712.
- Huang, H.-P., Robertson, A. W., & Kushnir, Y. (2005). Atlantic SST gradient and the influence of ENSO. *Geophysical Research Letters*, 32, L20706.
- Hughes, J. P., and Guttorp, P. (1994b). "Incorporating spatial dependence and atmospheric data in a model of precipitation." *Journal of Applied Meteorology*, 33(12), 1503–1515.
- Hughes, J. P., Guttorp, P., and Charles, S. P. (1999). "A non-homogeneous hidden Markov

- model for precipitation occurrence.” *Journal of the Royal Statistical Society: Séries C (Applied Statistics)*, 48(1), 15–30.
- Hurrell, J. W. (1995). Decadal trends in the North Atlantic oscillation and relationships to regional temperature and precipitation. *Science*, 269, 676–679.
- Hurst, H.E. (1951). "Long Term Storage Capacities of Reservoirs". *Transactions of the American Society of Civil Engineers*, 116:776-808.
- Hurst, H. E., 1957. "A suggested statistical model of some time series which occur in nature". *Nature*, 180, 494.
- Hyndman, R., J. (2019). "*Cross-validation for time series*". Acesso em 20 de fevereiro de 2019. Disponível em: <https://robjhyndman.com/hyndsight/tscv/>
- Hyndman, R., J. Athanasopoulos, G. (2019). Prediction intervals. Em: *Forecasting: Principles and Practice*. Acesso em 20 de fevereiro de 2019. Disponível em: <https://robjhyndman.com/hyndsight/tscv/>
- Ilich, N.; Despotovic, J. (2008). "A simple method for effective multi-site generation of stochastic hydrologic time series". *Stoch Environ Res Risk Assess* 22, páginas 265–279.
- INPE/CPTEC, (2017). El Niño e La Niña. Consultado em Novembro 17, 2017, disponível em: <http://enos.cptec.inpe.br/>, acesso em 24/05/2017
- IRI, (2012a). Overview of the ENSO System. Acesso em Dezembro 15, 2012, disponível em: <http://iri.columbia.edu/climate/ENSO/background/basics.html>
- IRI (2017). About ENSO. Disponível em: <http://iri.columbia.edu/our-expertise/climate/enso/enso-essentials/>. Acesso em 25/06/2017
- Ito, E. R. K., and 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), 64–74.
- Jelinek, F. (1997). *Statistical Methods for Speech Recognition*. MIT Press
- JISAO. (2019). "PDO Index". Joint Institute for the Study of the Atmosphere and Ocean. Acesso em 20 de fevereiro de 2019. Disponível em: <http://research.jisao.washington.edu/pdo/PDO.latest>

- Jones C, Waliser DE, Lau KM, Stern W. (2004a). Global occurrences of extreme precipitation and the Madden–Julian oscillation: observations and predictability. *J. Clim.* 17: 4575–4589, doi: 10.1175/3238.1.
- Jones, C. (2000). Occurrence of extreme precipitation events in California and relationships with the Madden–Julian oscillation. *J. Climate*, 13, 3576–3587
- Jones, C., Lau K. M., e Stern W. (2004b). Global occurrences of extreme precipitation and the Madden-Julian oscillation: Observations and predictability. *J. Climate*, 17, 4575–4589
- Jones, C. (2009). “A homogeneous stochastic model of the Madden-Julian oscillation.” *Journal of Climate*, 22(12), 3270–3288.
- Jong, B. T., Ting, M., Seager, R., Henderson, N., and Lee, D. E. (2018). “Role of equatorial Pacific SST forecast error in the late winter California precipitation forecast for the 2015/16 El Niño.” *Journal of Climate*, 31(2), 839–852.
- Kam, J., Sheffield, J., and Wood, E. F. (2014a). “Changes in drought risk over the contiguous United States (1901–2012): The influence of the Pacific and Atlantic Oceans.” *Geophysical Research Letters*, 1–7
- Kam, J., J. Sheffield, X. Yuan, and E. F. Wood (2013), The influence of Atlantic Tropical Cyclones on drought over the eastern United States (1980–2007), *J. Clim.*, 26, 3067–3086
- Kam, J., J. Sheffield, X. Yuan, and E. F. Wood (2014b), Did a skillful prediction of sea surface temperatures help or hinder forecasting of the 2012 Midwestern US drought?, *Environ. Res. Lett.*, 9, 034005
- Kayano MT, Kousky VE. (1999). Intraseasonal (30–60 day) variability in the global tropics: principal modes and their evolution. *Tellus A* 51: 373–386, doi: 10.1034/j.1600-0870.1999.t01-3-00003.x
- Kayano, M. T.; Jones, C.; Silva Dias, P. L.; (2009). Variabilidade Intrassazonal. In: Cavalcanti, I.D.A.; Ferreira, N. J.; Silva, M.G.A.J.; Silva Dias, M.A.F. (Orgs.). *Tempo e clima no Brasil*. São Paulo: Ofic. Textos. p. 338-351
- Kayano, M. T., and Andreoli, R. V. (2006). “Relationships between rainfall anomalies over northeastern Brazil and the El Niño–Southern Oscillation.” *Journal of Geophysical Research*, 111(D13), D13101.
- Kayano, M. T., Andreoli, R. V., Garcia, S. R., and de Souza, R. A. F. (2018). “How the two nodes of the tropical Atlantic sea surface temperature dipole relate the climate of the surrounding regions during austral autumn.” *International Journal of Climatology*,

1–15.

- Kayano, M. T., Capistrano, V. B., Andreoli, R. V., and de Souza, R. A. F. (2016). “A further analysis of the tropical Atlantic SST modes and their relations to north-eastern Brazil rainfall during different phases of Atlantic Multidecadal Oscillation.” *International Journal of Climatology*, 36(12), 4006–4018.
- Kelman, J.; Vieira, A.M.; Rodriguez-Amaya, J.E. (2000). "El Niño influence on streamflow forecasting". *Stochastic Environmental Research and Risk Assessment*, **14**, p123–138.
- Kessler, W. S., M. J. McPhaden, and K. M. Weickmann, (1995): Forcing of intraseasonal Kelvin waves in the equatorial Pacific. *J. Geophys. Res.*, 100, 10 613–10 631.
- Khalil, A. F., Kwon, H.-H., Lall, U., and Kaheil, Y. H. (2010). “Predictive downscaling based on non-homogeneous hidden Markov models.” *Hydrological Sciences Journal*, 55(3), 333–350.
- Kiladis GN, Weickmann KM. (1992). Circulation anomalies associated with tropical convection during Northern Winter. *Weather Rev.* 120: 1900–1923
- Kiladis, G. N., K. H. Straub, and P. T. Haertel, (2005): Zonal and vertical structure of the Madden–Julian oscillation. *J. Atmos. Sci.*, 62, 2790–2809
- Kirshner, S. (2005a). “Modeling of multivariate time series using hidden Markov models.” University Of California, Irvine.
- Kirshner, S. (2005b). “Quick Start Manual for the MVN-HMM Toolbox.” *Donald Bren School of Information and Computer Science*, University of California, Irvine.
- Klemeš, V., Srikanthan, R., and McMahon, T. A. (1981). “Long - memory flow models in reservoir analysis: What is their practical value?” *Water Resources Research*, 17(3), 737–751.
- Knight, J. R., Folland, C. K., and Scaife, A. A. (2006). “Climate impacts of the Atlantic multidecadal oscillation.” *Geophysical Research Letters*, 33(17), 2–5.
- Koscielny-Bunde, E., Kantelhardt, J. W., Braun, P., Bunde, A., & Havlin, S. (2006). "Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies". *Journal of Hydrology*, 322(1-4), 120–137.
- Koutsoyiannis, D. (2002). “The Hurst phenomenon and fractional Gaussian noise made easy.” *Hydrological Sciences Journal*, 47(4), 573–595.
- Kousky, V.E. (1979). “Frontal Influences on Northeast Brazil”, *Monthly Weather Review*, 107, pp. 1140–1153.
- Krogh A., Brown, M. Mian I. S., Sjolander K., and Haussler D. (1994). "Hidden Markov

- models in computational biology – applications to protein modeling". *Journal of Molecular Biology*, 235(5):1501–1531.
- Kumar, A., M. Chen, M. Hoerling, and J. Eischeid (2013), Do extreme climate events require extreme forcings?, *Geophys. Res. Lett.*, 40, 3440–3445.
- Kwon, J.-H. (2002). "Streamflow Forecasting for Geum River Basin by Using ARMAX Model.". Master Thesis, Yonsei University.
- Kwon, H.-H., de Assis de Souza Filho, F., Block, P., Sun, L., Lall, U., & Reis, D. S. (2011). "Uncertainty assessment of hydrologic and climate forecast models in Northeastern Brazil". *Hydrological Processes*, 26(25), 3875–3885
- Lambert, M. F., Whiting, P., Metcalfe, V., Whiting, J. P., and Metcalfe, A. V. (2003). "A non-parametric hidden Markov model for climate state identification." *Hydrology and Earth System Sciences*, 7(5), 652–667.
- Lau, W. K. M., (2005): El Niño–Southern Oscillation connection. In- *traseasonal Variability in the Atmosphere–Ocean Climate System*, Eds., Springer, 271–305
- Lau, W. K. M., D. E. Waliser, (2005): *Intraseasonal Variability in the Atmosphere–Ocean Climate System*. Springer, 436 pp.
- Lee, D.-J., Salas, J. D., & Boes, D. C. (2007). "Uncertainty Analysis for Synthetic Streamflow Generation". *World Environmental and Water Resources Congress 2007*.
- Lee, S. K., Lopez, H., Chung, E. S., DiNezio, P., Yeh, S. W., and Wittenberg, A. T. Lima, E. de A., Filho, M. F. G., and Da Silva, A. O. (2012). "Influência Do Oceano Pacífico Na Atuação Dos Vórtices Ciclônicos De Altos Níveis Que Atuaram No Nordeste Do Brasil Durante A Década De 1980." *Anais do XVII Congresso Brasileiro de Meteorologia*.
- Leroux, B. G. and Puterman, M. L. (1992) "Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models". *Biometrics* 48, 545—58
- Li, E., Mu, X., Zhao, G., and Gao, P. (2015). "Multifractal detrended fluctuation analysis of streamflow in the yellow river basin, China." *Water (Switzerland)*, 7(4), 1670–1686.
- 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*, 1–15.
- Lima, C. H. R., and Lall, U. (2008). "Análise e Previsões de Afluências ao Reservatório Sobradinho utilizando um Modelo Periódico Autoregressivo Exógeno (PARX)." *XVII Simpósio Brasileiro de Recursos Hídricos*, São Paulo, 1–26.

- Lima, C. H. R., e U. Lall (2010a), Climate informed long term seasonal forecasts of hydroenergy inflow for the Brazilian hydropower system, *Journal of Hydrology*, 381(1-2), 65–75.
- Lima, C. H. R., and U. Lall (2010b)." Climate informed monthly streamflow forecasts for the Brazilian hydropower network using a periodic ridge regression model", *Journal of Hydrology.*, 380(3-4).
- Lins, H. F., e Cohn, T. A. (2011). “Stationarity: Wanted dead or alive?” *Journal of the American Water Resources Association*.
- Liu, Z., Zhang, Q., & Wu, L. (2004). Remote impact on tropical Atlantic climate variability: Statistical assessment and dynamic assessment. *Journal of Climate*, 17, 1529–1549.
- Liu, Y., Ye, L., Qin, H., Hong, X., Ye, J., Yin, X. (2018). "Streamflow forecasting based on hidden Markov model and Gaussian Mixture Regression", *Journal of hydrology*, 561, 146–159.
- Lo, F, Hendon, H. H. Empirical extended-range prediction of the Madden-Julian oscillation. *Monthly weather review*, v. 128, n. 7, p. 2528-2543, 2000.
- Lohre, M., P. Sibbertsen, and T. K€onning (2003). "Modeling water flow of the Rhine River using seasonal long memory", *Water Resources. Research.*, 39(5), 1132,
- Loucks, D. P., Beek, E. van, Stedinger, J. R., Dijkman, J. P. M., and Villars, M. T. (2005). “Concepts in Probability , Statistics and Stochastic Modelling.” *Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications*, UNESCO Publishing, WL - Delft Hydraulics, The Netherlands, 168–229.
- Lu, Z. Q., and Berliner, L. M. (1999). “Markov switching time s€eries models with application to a daily runoff s€eries.” *Water Resources Research*, 35(2), 523–534.
- Lucena, D. B., Servain, J., & Gomes Filho, M. F. (2011). Rainfall response in northeast Brazil from ocean climate variability during the second half of the twentieth century. *Journal of Climate*, 24, 6174–6184.
- Luo, L., Apps, D., Arcand, S., Xu, H., Pan, M., and Hoerling, M. (2017). “Contribution of temperature and precipitation anomalies to the California drought during 2012–2015.” *Geophysical Research Letters*, 44(7), 3184–3192.
- Machiwal, D., and Jha, M. K. (2012). *Hydrologic Time S€eries Analysis: Theory and Practice*. Springer Netherlands, Dordrecht.
- Maft€ei, C., Barbulescu, A., and Carsteanu, A. A. (2016). “Long-range dependence in the time s€eries of Tai€ta River discharges.” *Hydrological Sciences Journal*, 61(9), 1740–

1747.

- Mandelbrot, B.B. and Wallis, J.R. (1969). "Some long-run properties of geophysical records", *Water Resources Research.*, 5: 321– 340
- Mann, M. E., and J. Park (1999), Oscillatory spatiotemporal signal detection in climate studies: A multiple-taper spectral domain approach, *Adv. Geophys.*, 41, 1–131
- Mantua, N. J., & Hare, S. R. (2002). The Pacific decadal oscillation. *Journal of Oceanography*, 58,35–44. <https://doi.org/10.1023/A:1015820616384>
- Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, 78, 1069–1079
- Marengo, J., Hastenrath, S. (1993). "Case studies of extreme climatic events in the Amazon basin". *Journal of Climate*, 6, 617-617.
- Marengo, J. A. (2006). *Mudanças Climáticas Globais e seus Efeitos sobre a Biodiversidade*. Ministério Do Meio Ambiente Secretaria De Biodiversidade E Florestas, BRASÍLIA – DF.
- Marengo, J. A. (2008). “Vulnerabilidade , impactos e adaptação à mudança do clima no semi-árido do Brasil.” *Parcerias Estratégicas*, 27.
- Marengo, J. A., Cunha, A. P., Alves, L. M., Ciências, C. De, Instituto, T., and Pesquisas, N. De. (2016). “A seca de 2012-15 no semiárido do Nordeste do Brasil no contexto histórico.” *Revista Climanálise*, Edição Com(November), 1–6.
- Marengo, J. A., Nobre, C. A., Seluchi, M. E., Cuartas, A., Alves, L. M., Mendiondo, E. M., Obregón, G., and Sampaio, G. (2015). “A seca e a crise hídrica de 2014-2015 em São Paulo.” *Revista USP*, (106), 31–44.
- Marengo, J. A., and 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), 5–28.
- Mares, C., Mares, I., Huebener, H., Mihailescu, M., Cubasch, U., and Stanciu, P. (2014). “A hidden Markov model applied to the daily spring precipitation over the Danube basin.” *Advances in Meteorology*, 2014.
- Matalas, N. C. (1967). "Mathematical Assessment of Synthetic Hydrology". *Water Resources Research*, v.3, p.937-945.
- Matthews AJ, Pickup G, Peatman SC, Clews P, Martin J. (2013). The effect of the Madden-Julian oscillation on station rainfall and river level in the Fly River system, Papua New Guinea. *J. Geophys. Res. Atmos.* 118: 10,926–10,935, doi: 10.1002/jgrd.50865.

- Matalas, N. C. (2012). "Comment on the Announced Death of Stationarity." *Journal of Water Resources Planning and Management*, 138(4), 311–312.
- McCabe, G. J., and M. D. Dettinger (1999), Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States, *Int. J. Climatol.*, 19, 1399–1410
- McCabe, G. J., M. A. Palecki, and J. L. Betancourt (2004), Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States, *Proc. Natl. Acad. Sci. U. S. A.*, 101, 4136–4141
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC.
- McPhaden, M.J. (2004): Evolution of the 2002/03 El Niño. *Bull. Amer. Meteor. Soc.*, 85, 677–695
- McPhaden, M. J., 1999: Genesis and evolution of the 1997-98 El Niño. *Science*, 283, 950–954.
- McPhaden, M. J., and B. A. Taft, 1988: Dynamics of seasonal and intraseasonal variability in the eastern equatorial Pacific. *J. Phys. Oceanogr.*, 18, 1713–1732.
- Mehrotra, R., and Sharma, A. (2005). "A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences." *Journal of Geophysical Research*, 110(D16), 1–13.
- Meira Neto, A. A., Oliveira, P. T. S., Rodrigues, D. B. B., & Wendland, E. (2018). "Improving Streamflow Prediction Using Uncertainty Analysis and Bayesian Model Averaging". *Journal of Hydrologic Engineering*, 23(5),.
- Meko, D. M. (2005). GEOS 585A, *Applied Time Series Analysis. Lesson 3*. University of Arizona, 1-6p.
- Melice, J. L., & Servain, J. (2003). The tropical Atlantic meridional SST gradient index and its relationships with the SOI, NAO and Southern Ocean. *Climate Dynamics*, 20, 447–464
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J. (2008). "Climate change. Stationarity is dead: whither water management?" *Science (New York, N.Y.)*, 319(5863), 573–574.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., Stouffer, R. J., Dettinger, M. D., and Krysanova, V. (2015). "On Critiques of 'stationarity is Dead: Whither Water Management?'" *Water Resources Research*, 51(9), 7785–7789.

- Mo K.C, Jones C, Nogués-Paegle J. (2012). Pan America. In *Intraseasonal Variability in the Atmosphere–Ocean Climate System*, Lau KM, Waliser DE (eds). Springer-Verlag: Berlin, Heidelberg, Germany.
- Moura, A. D., and Shukla, J. (1981). "On the dynamics of droughts in northeast Brazil: Observation, theory and numerical experiments with a general circulation model", *J. Atmos. Sci.*, 38,2653–2675
- Montgomery, J.; Nyhan, B. (2010). "Bayesian Model Averaging: Theoretical developments and practical applications". *Forthcoming, Political Analysis*,.
- Nag, R., K. Wong, and F. Fallside (1986). Script recognition using hidden markov models. In *ICASSP86*, pp. 2071–2074. IEEE
- Naghattini, M., and Pinto, É. J. de A. (2007). *Hidrologia Estatística*. CPRM - Serviço Geológico do Brasil, Belo Horizonte.
- Nayak, M. A., and Villarini, G. (2016). "Evaluation of the capability of the Lombard test in detecting abrupt changes in variance." *Journal of Hydrology*, Elsevier B.V., 534, 451–465.
- Nigam, S., Guan, B., and Ruiz-Barradas, A. (2011). "Key role of the Atlantic Multidecadal Oscillation in 20th century drought and wet periods over the Great Plains." *Geophysical Research Letters*, 38(16), 1–6.
- NOAA(2017a).Disponivelem:http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/nino_regions.shtml, acesso em 27/06/2017
- NOAA (2017b). Disponível em: <http://stateoftheocean.osmc.noaa.gov/sur/pac/>, acesso em 24/05/2017
- NOAA (2018a). Disponível em: <http://www.crh.noaa.gov/glossary>, acesso em 26/07/2018.
- Nobre, P.; Shukla, J.(1996). "Variations of sea surface temperatures, wind stress, and rainfall over the tropical over the tropical Atlantic and South America". *J. Climate*., v.9, n.10, p.2464-2479. 1996.
- Nóbrega R.S, Santiago G.A.C.F. (2014). Tendência de temperature na superfície do mar nos oceanos Atlântico e Pacífico e variabilidade de precipitação em Pernambuco. *Mercator*. 13(1): 107-118.
- Nys, E. de, Engle, N. L., and Magalhães, A. R. (2016). *Secas no Brasil : Política e gestão proativas*. Centro de Gestão e Estudos Estratégicos- CGEE; Banco Mundial, Brasília, DF.
- O’Connell, P. E., Koutsoyiannis, D., Lins, H. F., Markonis, Y., Montanari, A., and Cohn, T. (2016). "The scientific legacy of Harold Edwin Hurst (1880–1978)." *Hydrological*

- Sciences Journal*, 61(9), 1571–1590.
- Ochoa-Rivera, J.C.; García-Bartual, R.; Andreu, J. (2002). "Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks". *Hydrology and Earth System Sciences*, 6(4), p.641–654.
- Okumura, Y., Xie, S.-P., Numaguti, A., & Tanimoto, Y. (2001). Tropical Atlantic air–sea interaction and its influence on the NAO. *Geophysical Research Letters*, 28, 1507–1510
- Oliveira, V. G. De. (2015). “*Atualização das Previsões de Curto Prazo de Afluências ao Sistema Hidroelétrico Brasileiro a Partir da Técnica de Ponderação Bayesiana e de Previsões Mensais de Afluência com Uso de Informação Climática.*” Dissertação de Mestrado em Tecnologia Ambiental e Recursos Hídricos, Universidade de Brasília.
- Oliveira, V. G. De, and Lima, C. H. R. (2016). “Previsões multiescala de vazões para o sistema hidrelétrico brasileiro utilizando ponderação bayesiana de modelos (BMA).” *Revista Brasileira de Recursos Hídricos*, 21(3), 618–635.
- ONS (2017). Disponível em: http://www.ons.org.br/operacao/vazoes_naturais.aspx, acesso em 23/04/2017
- Petrova, D., Koopman, S. J., Ballester, J., and Rodó, X. (2017). “Improving the long-lead predictability of El Niño using a novel forecasting scheme based on a dynamic components model.” *Climate Dynamics*, Springer Berlin Heidelberg, 48(3–4)
- Pezzi, L. P., & Cavalcanti, I. F. A. (2001). The relative importance of ENSO and tropical Atlantic sea surface temperature anomalies for seasonal precipitation over South America: A numerical study. *Climate Dynamics*, 17, 205–212
- Pilz, T., Delgado, J. M., Voss, S., Vormoor, K., Francke, T., Cunha Costa, A., Martins, E., and Bronstert, A. (2019). “Seasonal drought prediction for semiarid northeast Brazil: What is the added value of a process-based hydrological model?” *Hydrology and Earth System Sciences*, 23(4), 1951–1971.
- Pokhrel, P., Wang, Q. J., & Robertson, D. E. (2013). "The value of model averaging and dynamical climate model predictions for improving statistical seasonal streamflow forecasts over Australia". *Water Resources Research*, 49(10), 6671–6687.
- Prairie, J., Nowak, K., Rajagopalan, B., Lall, U., and Fulp, T. (2008). “A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data.” *Water Resources Research*, 44(6), 1–11.
- Propst, D. L., Gido, K. B., and Stefferud, J. A. (2008). “Natural Flow Regimes, Nonnative Fishes, and Native Fish Persistence in Arid-Land River Systems.” *Ecological*

- Applications*, 18(5), 1236–1252.
- Quan, X.-W., Hoerling, M., Smith, L., Perlwitz, J., Zhang, T., Hoell, A., Wolter, K., and Eischeid, J. (2018). “Extreme California Rains During Winter 2015/16: A Change in El Niño Teleconnection?” *Bulletin of the American Meteorological Society*, 99(1), S49–S53.
- Rabiner, L. and B. H. Juang (1993). *Fundamentals of Speech Recognition*. Prentice Hall.
- Reis, D. S., Martins, E. S. P. R., Nascimento, L. S. V., Costa, A. A. e Alexandre, A. M., (2009). “Previsão de Afluências no estado do Ceará com modelos Climáticos e Hidrológicos” em: *Informações Climáticas em Recursos Hídricos*, 1 Edição, FUNCEME, 160p.
- Repelli CA, Souza EB, Quadro MFL, Alves JMB, Sakamoto MS. (1998). O episódio de chuvas intensas no Nordeste Brasileiro no final de março/1997: influências da oscilação de 30–60 dias. *Rev. Bras. Meteorol.* 13: 9–18.
- 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.
- Richman, M. B., and Leslie, L. M. (2015). “Uniqueness and Causes of the California Drought.” *Procedia Computer Science*, Elsevier Masson SAS, 61(December), 428–435.
- Robertson, A. W., Kirshner, S., and Smyth, P. (2004). “Downscaling of Daily Rainfall Occurrence over Northeast Brazil Using a Hidden Markov Model.” *Journal Of Climate*, 17, 4407–4424.
- Robertson, A. W., Kirshner, S., Smyth, P., CHARLES, S. P., and BATES, B. C. (2005). “Subseasonal-to-Interdecadal Variability of the Australian Monsoon Over North Queensland.” *Q. J. R. Meteorol. Soc.*, 131, 1–26.
- Robertson, A. W., and Smyth, S. K. P. (2003). *Hidden Markov models for modeling daily rainfall occurrence over Brazil*.
- Rodríguez-Fonseca, B., I. Polo, J. García-Serrano, T. Losada, E. Mohino, C. R. Mechoso, and F. Kucharski, (2009): Are Atlantic Niños enhancing Pacific ENSO events in recent decades? *Geophys. Res. Lett.*, 36, L20705,
- Rowell, D. P. (2003), The impact of Mediterranean SSTs on the Sahelian rainfall season, *J. Clim.*, 16, 849–862.
- Rowell, D. P., C. K. Folland, K. Maskell, and M. N. Ward (1995), Variability of summer rainfall over tropical North-Africa (1906–92) observations and modelling, *Q. J. R.*

- Meteorol. Soc., 121, 669–704
- Sagarika, S., Kalra, A., and Ahmad, S. (2015). “Interconnections between oceanic-atmospheric indices and variability in the U.S. streamflow.” *Journal of Hydrology*, Elsevier B.V., 525, 724–736.
- Salas, J.D., Boes, D.C., (1980). "Shifting level modeling of hydrologic séries". *Advances in Water Resources*, 3, 59–63p.
- Salas, J.D.; Delleur, J.W.; Yevjevich, V. e Lane, W.L. (1980). *Applied modelling of hydrological time séries*. Water Resources Press Publications, Littleton, Colorado, E.U.A., 485p.
- Salas, J.D. (1993). "Analysis and modelling of hydrological time séries". In: D. Maidment (ed.), *Handbook of hydrology*, Chapter 17. New York, McGraw-Hill.
- Salas, J. D., Govindaraju, R. S., Anderson, M., Arabi, M., Francés, F., Suarez, W., Lavado-Casimiro, W. S., and Green, T. R. (2014). “7. Risk and Uncertainty Analyses in Hydrology.” *Handbook of Environmental Engineering: Modern Water Resources Engineering*, L. K. Wang and C. T. Yang, eds., Humana Press, Totowa, NJ, 879.
- Saldarriaga, J. ., and Yevjevich, V. (1970). “Application of run-lengths to hydrologic séries.” *Hydrology Papers Colorado State University*, 40, 57.
- Santos, M. B. DOS. (2009). “*Indicadores de sustentabilidade de reservatórios no Semiárido Brasileiro: uma proposta para a Bacia do Rio Jaguaribe - Ceará.*” Dissertação de Mestrado – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Civil.
- Saravanan, R., & Chang, P. (2000). Interaction between tropical Atlantic variability and El Niño-Southern Oscillation. *Journal of Climate*, 13,2177–2194.
- Sarmiento, F. J. (1989). *Aplicabilidade de Modelos de Geração de Vazão no Semiárido do Nordeste do Brasil*. Dissertação de Mestrado, Universidade Federal do Ceará, Departamento de Engenharia Hidráulica e Ambiental, 142p.
- Schwarz, G. E. (1978), "Estimating the dimension of a model", *Annals of Statistics*, 6 (2): 461–464.
- Seager, R., Kushnir, Y., Visbeck, M., Naik, N., Miller, J., Krahnemann, G., & Culle, H. (2000). Causes of Atlantic Ocean climate variability between 1958 and 1998. *Journal of Climate*, 13, 2845–2862.
- Seager, R., L. Goddard, J. Nakamura, N. Henderson, and D. E. Lee (2014), Dynamical causes of the 2010/11 Texas–Northern Mexico drought, *J. Hydrometeorol.*, 15,39–68.
- Servain, J. (1991). Simple climatic indices for the tropical Atlantic Ocean and some

- applications. *Journal of Geophysical Research*, 96, 15137–15146.
- Servain, J. (1991). Simple climatic indices for the tropical Atlantic Ocean and some applications. *Journal of Geophysical Research*, 96, 15137–15146
- Servain, J., Wainer, I., McCreary, J. R., & Dessier, A. (1999). Relationship between the equatorial and meridional modes of climatic variability in the tropical Atlantic. *Geophysical Research Letters*, 26, 485–488.
- Sheffield, J., E. F. Wood, and M. L. Roderick (2012), Little change in global drought over the past 60 years, *Nature*, 491, 435–438.
- Silveira, C. D. S., Alexandre, A. M. B., Filho, F. D. A. D. S., and Fernandes, W. D. S. (2011a). “Avaliação das projeções para as vazões dos Modelos do IPCC-AR4 para o cenário A1B para as Bacias do Estado do Ceará.” *XIX Simpósio de Brasileiro de Recursos Hídricos*, Maceio, AL, 1–18.
- Silveira, C. da S. (2014). “*Modelagem Integrada De Meteorologia E Recursos Hídricos Em Múltiplas Escalas Temporais E Espaciais: Aplicação No Ceará E No Setor Hidroelétrico Brasileiro.*” Tese de Doutorado em Engenharia Civil, Universidade Federal do Ceará.
- Silveira, C. da S., Filho, F. de A. de S., Lázaro, Y. M. C., and Fernandes, W. de S. (2011b). “Avaliação de desempenho dos modelos de mudança climática do IPCC-AR4 quanto a sazonalidade e os padrões de variabilidade interanual da precipitação sobre a América do Sul.” *XIX Simpósio de Brasileiro de Recursos Hídricos*, Maceio, AL, 1–20.
- Souza E.B, Ambrizzi T. (2006). Modulation of the intraseasonal rainfall over tropical Brazil by the Madden–Julian oscillation. *Int. J. Climatol.* 26: 1759–1776.
- Souza E.B, Kayano MT, Ambrizzi T. (2005). Intraseasonal and submonthly variability over the Eastern Amazon and Northeast Brazil during the autumn rainy season. *Theor. Appl. Climatol.* 81: 177–191.
- Souza Filho, F. A., and 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.
- SRHC (2005). Secretaria de Recursos Hídricos. Estado do Ceará. Consolidação da política e dos programas de recursos hídricos do estado do Ceará. Disponível em:<http://www.srh.ce.gov.br/images/stories/PLANERH.rar>, acesso Abril de 2017.
- Stedinger, J. R., and Taylor, M. R. (1982). “Synthetic streamflow generation: 2. Effect of parameter uncertainty.” *Water Resources Research*, 18(4), 919–924.

- Stock, J. H., and Watson, M. W. (2010). *Introduction to Econometrics*. Pearson Education, Inc., Boston, MA.
- Storch, H. Von, and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, New York, USA.
- Sturdart, T. M. de C. (2000). Análises de Incertezas na Determinação de Vazões Regularizadas em Climas Semi-Áridos. Tese de Doutorado em Engenharia Civil - Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza.
- Sun, X., Cook, K. H., and Vizy, E. K. (2017). “The South Atlantic subtropical high: Climatology and interannual variability.” *Journal of Climate*, 30(9), 3279–3296.
- Sutton, R. T., and D. L. R. Hodson (2005), Atlantic Ocean forcing of North American and European summer climate, *Science*, 309, 115–118.
- Sutton, R. T., Jewson, S. P., & Rowell, D. P. (2000). The elements of climate variability in the tropical Atlantic region. *Journal of Climate*, 13, 3261–3284.
- Sutton, R. T., Norton, W. A., & Jewson, S. P. (2001). The North Atlantic Oscillation—What role for the ocean? *Atmospheric Science Letters*, 1, 89–100.
- Sveinsson, O. G. B., Salas, J. D., Boes, D. C., and Pielke Sr., R. A. (2003). “Modeling the Dynamics of Long-Term Variability of Hydroclimatic Processes.” *Journal of Hydrometeorology*, 4, 489–505.
- Sveinsson, O.G.B.; Lall, U.; Fortin, V. Perrault, L. Gaudet, J.; Zebiak, S.; Kushnir, Y. (2008). "Forecasting Spring Reservoir Inflows in Churchill Falls Basin in Québec Canada". *Journal of Hydrologic Engineering*. Vol.13, No.6, p.426-437,
- Thyer, M., Frost, A. J., and 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), 313–328.
- Thyer, M., and 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., and 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), 12–26.
- Thyer, M., and 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), 27–48.

- Thomas, H. A. and Fiering, M. B., (1962). "Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation." In : A. Mass et al. eds., Design Of Water Resource Systems, 459-493, Massachusetts : Harvard University Press.
- Turner, S. W. D., and Galelli, S. (2016). "Regime-shifting streamflow processes: Implications for water supply reservoir operations." *Water Resources Research*, 52(5), 3984–4002.
- Uvo, C. B., Repelli, C. A., Zebiak, S. E., and Kushnir, Y. (1998). "The Relationships between Tropical Pacific and Atlantic SST and Northeast Brazil Monthly Precipitation." *Journal Of Climate*, 11, 551–562.
- Valadão, C. E. A., Carvalho, L. M. V., Lucio, P. S., and Chaves, R. R. (2017). "Impacts of the Madden-Julian oscillation on intraseasonal precipitation over Northeast Brazil." *International Journal of Climatology*, 37(4), 1859–1884.
- Valdés, J. B., Rodríguez-Iturbe, I., and Vicens, G. J. (1977). "Bayesian generation of synthetic streamflows: 2. The multivariate case." *Water Resources Research*, 13(2), 291–295.
- Varejão-Silva, M. A. (2001). *Meteorologia e Climatologia*. PAX Gráfica e Editora Ltda, 2ª Edição, 532 pág.
- Yu, S., Wang, S., Yoon, J., Gillies, R. R., and Hsu, H. H. (2016). "The California Drought: Trends and Impacts." *Climate Extremes: Patterns and Mechanisms*, Publisher: AGU (Wiley), 223–235.
- Wagner RG, da Silva AM. (1994). Surface conditions associated with anomalous rainfall in the Guinea coastal region. *Int. J. Climatol.* 14: 179–199.
- 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), 16,683-16,694.
- Wang C. (2002). Atlantic climate variability and its associated atmospheric circulation cells. *J. Clim.* 15: 1516–1536.
- Wang, W.-C., Chau, K.-W., Cheng, C.-T., & Qiu, L. (2009). "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series". *Journal of Hydrology*, 374(3-4), 294–306.
- Wang, S.-Y., L. Hipps, R. R. Gillies, and J.-H. Yoon (2014), Probable causes of the abnormal ridge accompanying the 2013–2014 California drought: ENSO precursor and anthropogenic warming footprint, *Geophys. Res. Lett.*, 41, 3220–3226,

doi:10.1002/2014GL059748.

- Weickmann KM, Lussky GR, Kutzbach JE. (1985). Intraseasonal (30–60 day) fluctuations of outgoing longwave radiation and 250 mb stream- function during Northern winter. *Mon. Weather Rev.* 113: 941–961.
- Weickmann KM. (1983). Intraseasonal circulation and outgoing longwave radiation modes during Northern Hemisphere winter. *Mon. Weather Rev.* 111: 1838–1858,
- Whiting, J., Lambert, M., Metcalfe, A., and Kuczera, G. (2004). “Development of Non-homogeneous and Hierarchical Hidden Markov Models for Modelling Monthly Rainfall and Streamflow Time Séries.” *World Water Congress, Critical Transitions in Water and Environmental Resources Management World. American Society of Civil Engineers*, 1–9.
- Whiting, J. P. (2006). “Identification and Modelling of Hydrological Persistence with Hidden Markov Models.” University of Adelaide, Australia.
- Whiting, J. P., Lambert, M. F., and Metcalfe, A. V. (2003). “Modelling Persistence in Annual Australian Point Rainfall.” *Hydrology and Earth System Sciences*, 7(2), 197–211.
- Wilks, D., S (2006) "*Statistical Methods In The Atmospheric Sciences*". Second Edition. Elsevier, Academic Press, London UK.
- Wolter, K., and M.S. Timlin, (1993): Monitoring ENSO in COADS with a seasonally adjusted principal component index. Proc. of the 17th Climate Diagnostics Workshop, Norman, OK, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, 52-57.
- Wolter, K., and M. S. Timlin, (1998): Measuring the strength of ENSO events - how does 1997/98 rank? *Weather*, 53, 315-324..
- Zaiss, J.; Stott, L. D.; Buening, N. H. (2018). “Increased Internal Atmospheric Variability Contributed to the Persistence of the California Drought Through the 2015/16 El Niño.” American Geophysical Union, Fall Meeting 2017, 2017–2018.
- Zavala-Garay, J., A. M. Moore, C. L. Perez, and R. Kleeman, (2003). The response of a coupled model of ENSO to observed estimates of stochastic forcing. *J. Climate*, 16, 2827–2842.
- Zavala-Garay, J., C. Zhang, A. M. Moore, and R. Kleeman, (2005). The linear response of ENSO to the Madden–Julian oscillation. *J. Climate*, 18, 2441–2459.
- Zebiak SE. (1993). Air-sea interaction in the equatorial Atlantic region. *J. Clim.* 6: 1567–1586

- Zhang, C. D., (2005): "Madden-Julian oscillation". *Rev. Geophys.*, 43, 1–36
- Zhang, T., Hoerling, M. P., Wolter, K., Eischeid, J., Cheng, L., Hoell, A., Perlwitz, J., Quan, X. W., and Barsugli, J. (2018). "Predictability and prediction of Southern California rains during strong El Niño events: A focus on the failed 2016 winter rains." *Journal of Climate*, 31(2), 555–574.
- Zhang, X., Dong, Q., Costa, V., & Wang, X. (2019). "A hierarchical Bayesian model for decomposing the impacts of human activities and climate change on water resources in China". *Science of The Total Environment*, 665, 836–847.
- Zucchini, W., and Guttorp, P. (1991). "A Hidden Markov Model for Space-Time Precipitation." *Water Resources Research*, 27(8), 1917–1923.
- Zucchini, W., and MacDonald, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction using R. Monographs on statistics and applied probability; 110*, Chapman & Hall/CRC Taylor & Francis Group, Boca Raton, FL, USA.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R. C&H/CRC Monographs on Statistics & Applied Probability*, Chapman and Hall/CRC, Boca Raton, FL, USA.

APÊNDICE A

**APÊNDICE A – TABELAS SOBRE OS TRABALHOS QUE UTILIZARAM HMMs
PARA MODELAGEM HIDROLÓGICA**

**LISTA DE NOMENCLATURA E ABREVIACÕES UTILIZADA NAS TABELAS A.1
E A.2**

AIC	<i>Akaike Information Criterion</i>
AGP	Altura geopotencial
AMO	<i>Atlantic Multidecadal Oscillation</i>
AR(p)	Modelo Autoregressivo de ordem p
AR-HMM	HMM Autoregressivo
ARMA (p,q):	Modelo Autoregressivo de Médias Móveis de ordem p,q
BC	Box Cox
BIC	<i>Bayes Information Criterion</i>
BW	<i>Baum-Welch Algorithm</i>
DEP	Distribuição espacial da precipitação
D	Delta
DTPO	Depressão da temperatura do ponto de orvalho
DS	Dessazonalização
EM	<i>Expectation Maximization Algorithm</i>
ENSO	<i>El Niño–Southern Oscillation</i>
EOF	<i>Empirical Orthogonal Functions</i>
ERPL	<i>Estimation Restoration Pseudo Likelihood</i>
Ex	Exponencial
FB	<i>Forward-Backward Algorithm</i>
G	Gama
GCM	<i>Global Climate Models</i>
GS	<i>Gibbs Sampling</i>
HMM (m):	Modelo de Markov com m Estados Ocultos
HHMM (m)	HMM hierárquico
HSMM	Semi HMM não homogêneos
IC	Índices Climáticos

KNN	<i>K-Nearest-Neighbor</i>
LN	Log-Normal
LT	Log-transformação
M	Multivariado
MA	<i>Metropolis Algorithm</i>
MARX	<i>Markov Modulated AR Model</i>
MBAM	Modelo binário auto logístico multivariado
MCMC	<i>Monte Carlo Markov Chain</i>
MCML	<i>Monte Carlo Maximum Likelihood</i>
MVS	Maxima Vesrossimilhança
MM	Modelo Markoviano
MOEF	<i>Multivariate Empirical Orthogonal Functions</i>
MRL	Modelos de Regressão Linear
N	Normal
NAO	<i>North Atlantic Oscillation</i>
NHMM	HMM não homogêneo
NINO3, NINO1+2, Trans- Niño	Índices do ENSO
NP HMM	HMM não paramétricos
NP NHMM	HMM não homogêneos e não paramétricos
NP	Não Paramétrico
OLR	<i>Long Wave Radiation</i>
P	Precipitação
PDO	<i>Pacific Decadal Oscillation</i>
PL	<i>Pseudo Likelihood</i>
PNM	Pressão no nível do mar
PMNM	Pressão média ao nível do mar
Q	Vazão
SCE	<i>Shufflet Complex Evolution</i>
SL	<i>Shifting Level Model</i>
SST	<i>Surface Sea Temperature</i>
SVD	<i>Singular Value Decomposition</i>
TTAP	Teor total de água precipitável

UR

U

VA

Umidade Relativa

Univariado

Variáveis atmosféricas”

Tabela A.1 – Trabalhos que utilizaram HMMs para a modelagem da precipitação

Trabalho	Local	Tipo de Trabalho Variável	Modelos utilizados	Método de ajuste e estimação de parâmetros	Distribuições dependentes nos HMM	Transformação dos dados	Variáveis externas
Zucchini e Guttorp, (1991)	Washington (Estados Unidos)	Geração P diária	HMM (2,4) M	MVS	Frequência observada	Transformação a série binária	Nenhuma
Hughes e Guttorp, (1994 ^a)	Washington (Estados Unidos)	Geração P diária	HMM, NHMM (2,3,4) M	MVS e FB	MBAM	Nenhuma	PMNM
Hughes e Guttorp, (1994b)	Oeste de Washington (Estados Unidos)	Geração P diária	HMM, NHMM (2,3,4,5,6), M	ERPL, Dependência de estados a VA com <i>Viterbi</i>	Condicional a <i>m</i> e <i>t</i> , MBAM com correlação espacial.	Nenhuma	PMNM, AGP
Hughes <i>et al.</i> , (1999)	Sudoeste da Austrália	Geração P diária	NHMM (6,7) M	EM/MCMVS, FB, GS	MBAM	Nenhuma	PMNM, AGP
Charles <i>et al.</i> , (1999)	Sudoeste da Austrália	Geração P diária futura	NHMM (6) M	EM/MCMVS, FB, GS	MBAM	Nenhuma	GCM: PMNM, AGP, DTPO
Bellone <i>et al.</i> , (2000)	Washington (Estados Unidos)	Geração P diária	NHMM (6) M	EM/MCMVS, FB, GS	G	VA: SVD	AGP 1000mb, UR 850mb
Thyer e Kuczera, (2000)	Sidney, Brisbane, Melbourne, (Austrália)	Geração P Anual	HMM (2), AR(1) U	GS	N	BC no AR(1)	Nenhuma
Lambert <i>et al.</i> , (2003)	Costa Pacífica Australiana	Geração P Mensal	NP HMM (2)	MA	Nenhuma	Nenhuma	Nenhuma
Thyer e Kuczera, (2003 ^a)	Dados sintéticos de MCMC	Geração P Anual	HMM (2) M	GS	N	Nenhuma	Nenhuma

Tabela A.1 Continuação – Trabalhos que utilizaram HMMs para a modelagem da precipitação

Trabalho	Local	Tipo de Trabalho Variável	Modelos utilizados	Método de ajuste e estimação de parâmetros	Distribuições dependentes nos HMM	Transformação dos dados	Variáveis externas
Thyer e Kuczera, (2003b).	Bacias: Warragamba, Costa Central e do Rio Williams, (Austrália)	Geração P Diária	HMM (2), U e HMM (4), M	GS	N	Nenhuma	Nenhuma
Whiting, et al., (2003)	Sidney (Austrália)	Geração P Anual, PDO, SOI	HMM (2) P, PDO, SOI FARIMA (0,0.446,0) MRL	MCMC GS	N	LT nos MRL	Nenhuma
Robertson <i>et al.</i> , (2004)	Ceará (Brasil)	Geração P diária	HMM, NHMM (4) M	EM/MCMVS, FB, GS, BW	Geométrica	Nenhuma	Simulações de GCM
Whiting, <i>et al.</i> , (2004)	Sidney, Bourke (Austrália)	Geração P Mensal	HMM, NHMM, HHMM (2) e versões NP de todos. U	MA	LN para os modelos paramétricos	Nenhuma	Nenhuma
Mehrotra e Sharma, (2005)	Sidney (Austrália)	Geração P diária	KNN, NHMM (3) NP baseado no KNN, NHMM (6) M	AM	DEP no NHMM NP, MBAM no NHMM	Nenhuma	PMNM, AGP, TTAP
Robertson <i>et al.</i> , (2005)	Norte de Queensland (Austrália)	Geração P diária	HMM, NHMM (5) discretos-contínuos M	EM/MCMVS, FB, GS, BW	Delta se $P = 0$, se $P > 0$ Mistura de 2 Ex	Nenhuma	Reanalises e de GCM, TSMs
Thyer <i>et al.</i> , (2006).	Dados sintéticos de Thyer e Kuczera, 2000, 2003b)	Geração P Anual	HMM (2), AR(1) M	MA	N	BC no AR(1)	Nenhuma
Whiting, (2006)	Várias bacias na Austrália	Geração P Anual, mensal e diária.	P anual: HMM (2), P mensal: HMM (2, 3), HSMM (2), AR-HMM (2) até 3 Lags, HHMM (2), AR(3), NP HMM (2,3), P diária: HMM (2,3) U	EM via SCE Adaptive MA	P Anual: N P Mensal: LN, G	Escalamento, LN nos AR(p)	Nenhuma
Greene <i>et al.</i> , (2008)	Centro oeste da Índia	Geração P diária	HMM (4) M	Iterative EM, Software: Kirshner, (2005)	D se $P = 0$ se $P > 0$ Mistura de 2 Ex	Nenhuma	Nenhuma

Tabela A.1 Continuação – Trabalhos que utilizaram HMMs para a modelagem da precipitação

Trabalho	Local	Tipo de Trabalho Variável	Modelos utilizados	Método de ajuste e estimação de parâmetros	Distribuições dependentes nos HMM	Transformação dos dados	Variáveis externas
Gelati <i>et al.</i> , (2010 ^a)	Dinamarca e Sul da Suécia	Geração P diária	NHMM (8) M	EM/MCMVS, FB, GS, BW	G	VA: EOF, MEOF, SVD	AGP 1000hPa UR 850hPa
Khalil <i>et al.</i> , (2010)	Sul da Florida (Estados Unidos)	Geração e Previsão P diária	NHMM (4,5) M	EM/MCMVS, FB, GS, BW	Geométrica	Nenhuma	OLR, de GCMs NINO3.4, NAO, TSMs
Lima, (2010)	Fortaleza (Brasil)	Geração Clusters de P Anual	MM (2), NHMM (4) U	<i>Software:</i> Kirshner, (2005)	Bernoulli	IC: Filtragem de baixa frequência P: Transformação a série binária	DIPOLO do Atlântico e NINO3
Greene <i>et al.</i> , (2011)	Centro Oeste da Índia	Geração P diária futura	NHMM (4) M	Iterative EM, <i>Software:</i> Kirshner, (2005)	D se P = 0 se P > 0 Mistura de 2 Ex	Nenhuma	<i>Western Shear Index</i> (WSI11)
Fu <i>et al.</i> , (2012)	Sudeste da Austrália	Geração P diária futura	NHMM (5) M	EM/MCMVS, FB, GS	MBAM	Nenhuma	PMNM, AGP, DTPO
Mares <i>et al.</i> , (2014)	Bacia do Danúbio (Romênia)	Geração P diária	HMM, NHMM (7) M	<i>Iterative EM</i>	D se P = 0 se P > 0 Mistura de 2 Ex	IC: EOF, MEOF	PNM, AGP 850hPa

Tabela A.2 – Trabalhos que utilizaram HMMs para a modelagem de vazões

Trabalho	Local	Tipo de Trabalho Variável	Modelos utilizados	Método de ajuste e estimação de parâmetros	Distribuições dependentes nos HMM	Transformação dos dados	Variáveis externas
Jackson, (1975)	Rio Quaboag (Estados Unidos)	Geração Q Anual	HMM (2), AR(1) U	Nenhum: Estimação a partir das observações	N	LT	Nenhuma
Lu e Berliner, (1999)	Lago Taupo (Nova Zelândia)	Geração Q Diária	HMM (3) o modelo varia entre 3 AR(1) U	MCMC, GS	N	Nenhuma	P
Thyer e Kuczera, (2000)	Rio Burdekin (Austrália)/Chuva e Vazões anuais	Geração Q Anual	HMM (2), AR(1)	GS	N	LT, BC no AR(1)	Nenhuma
Lambert <i>et al.</i> , (2003)	Rios Murray, Darling, Cooper Creek (Austrália) e Nilo (África)	Geração Q Anual (Nilo), mensal	NP HMM (2)	MA	Nenhuma	Nenhuma	Nenhuma
Fortin <i>et al.</i> , (2004)	Rio Senegal (África)	Previsão Q Anual	SL (2, 3), AR(2), ARMA(1,1)	GS na parametrização e previsão	N (σ^2 constante μ e variável)	Nenhuma	Nenhuma
Whiting <i>et al.</i> , (2004)	Rios Cooper Creek e Murray (Austrália)	Geração Q Mensal	HMM, NHMM, HHMM (2) e versões NP de todos. U	MA	LN para os modelos paramétricos	Nenhuma	Nenhuma
Akintug e Rasmussen, (2005)	Rio Niágara (América do Norte)	Geração Q Anual	HMM (1,2,3), ARMA(1,1), ARMA(2,2)	MVS, BW	N	BC e escalamento ARMA(2,2), BC ARMA(1,1)	Nenhuma

Tabela A.2 – Continuação – Trabalhos que utilizara HMMs para a modelagem de vazões².

Trabalho	Local	Tipo de Trabalho Variável	Modelos utilizados	Método de ajuste e estimação de parâmetros	Distribuições dependentes nos HMM	Transformação dos dados	Variáveis externas
Whiting, (2006)	Rios Murray e Burdekin (Austrália)	Geração Q Mensal	AR(1,3), HHMM (2), HSMM (2,3) AR-HMM (até 5) até 3 Lags U	EM via SCE <i>Adaptive MA</i>	LN	Escalamento, LT nos AR(p)	Nenhuma
Gelati <i>et al.</i> , (2010b)	Reservatório Daule Peripa (Equador)	Geração e Previsão de anomalias Q Mensal	MARX(2)	SCE	N	IC e Q: DS Q: agregadas em três meses, LT	NINO 1+2 Trans-Niño
Bracken, (2011)	Rio Colorado (Estados Unidos)	Geração e Previsão Q Anual	HMM (2,3), AR(1) Não estacionários	EM, BW	N e G	LT	Nenhuma
Hernandez, (2013)	Ceará (Brasil)	Geração Q Anual	HMM, NHMM (até 6), AR(2) U	EM, BW	G e N	LT para as N	DIPOLo do Atlântico, NINO3
Bracken, <i>et al.</i> , (2014)	Rio Colorado (Estados Unidos)	Geração Q Anual	HMM, NHMM (2) Não estacionários	EM, BW	G	Nenhuma	AMO, PDO, NINO3
Turner e Galelli, (2016)	Vários reservatórios na Austrália	Geração Q Mensal	HMM (2,3) U	EM	N	LT, DS	Nenhuma
Bracken, <i>et al.</i> , (2016)	Várias estações na bacia do Rio Colorado (Estados Unidos)	Reconstrução (Geração) Q Anual (ano hidrológico)	NHMM(5) M	MVS Utilizando uma abordagem Bayesiana	LN	Escalamento: Cada Vazão se dividiu pelo desvio padrão da série	Variáveis de anéis de arvores
Liu <i>et al.</i> , (2018)	3 estações fluviométricas na bacia do Rio Yangtze (China)	Previsão Q Mensal	HMM (2) U	BW com a técnica <i>kernelized K-medoids Clustering</i> para inicialização	N Na previsão uma mistura de N	Nenhuma	Nenhuma

Tabela A.3 - Problemas e recomendações dos trabalhos citados na Tabela 1

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Zucchini e Guttorp, (1991)	A independência condicional entre estações de medição de P.	Incluir a correlação espacial. Avaliar o uso da dependência em estações próximas. Ampliar a escala do modelo. Incluir informação climática. Auxiliar estudos de GCMs. Usar PL e GS.
Hughes e Guttorp, (1994a)	Pequenos problemas no ajuste das distribuições marginais. A estrutura de correlação deu-se pelas variáveis atmosféricas, por padrões atmosféricos, e não pela correlação espacial.	Utilizar um NHMM para eventos extremos, mudanças climáticas; usando resultados de GCM. Empregar um AR-NHMM ou colocar uma estrutura de condicionante temporal, um modelo condicionado à chuva previa e ao estado climático atual, como em Hughes <i>et al.</i> , (1993), para eliminar complicações. Modificar o modelo M para estações próximas entre si, já que foi tratada independência espacial. Analisar quantidade de P. Usar distribuições Gama ou modelos transformados para usar a distribuição normal. Investigar o tamanho da amostra e as características dos dados. Usar resultados como inputs em modelos de simulação hidrológica, agrícola ou outros.
Hughes e Guttorp, (1994b)	Comparar via BIC os modelos independente e auto logístico. O modelo independente tem problemas nas correlações da probabilidade de ocorrência de P entre as estações, entregou valores altos. Viés nas simulações pelo ERPL. Problemas no ajuste resultam em problemas para representar a persistência, exemplo, duração da P em cada estação.	Utilizar um NHMM com N ou G. Analisar mudanças climáticas, ter em conta não estacionariedade. Realizar previsões no curto prazo. Utilizar resultados como <i>Inputs</i> em outros modelos. Empregar quantidades de P. Utilizar distribuições dependentes Gama ou transformar os dados e usar distribuições Normais, por exemplo, M Normal. Estender o modelo a escalas continental ou subcontinental, para criar um modelo multivariado em que cada estado seja uma região meteorológica.
Hughes <i>et al.</i> , (1999)	Não se conseguiu a simulação da variabilidade na série estudada pela possível estacionariedade climática, devido à suposição de cadeia de Markov, inviabilizando estudos de mudanças climáticas. O método de estimação de parâmetros requer um gasto computacional alto.	Utilizar modelos com menos estados e mais informação climática. Não aplicar em áreas onde a origem principal das chuvas é a convecção de ar. Definir o um adequado número de estados, e que eles tenham significado físico. Modificar o NHMM de tal forma de incluir quantidades de chuva e a correlação espacial entre estações. Colocar isso junto com a correlação de ocorrência de chuva para formar um modelo misto: Discreto-Contínuo. Criar um NHMM que permita uma interpolação a de resultados a estações que não estão na rede estudada.

Tabela A.3 Continuação - Problemas e recomendações dos trabalhos citados na Tabela 1

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Charles <i>et al.</i> , (1999)	Superestimação de probabilidade de ocorrência de P pelo uso de preditores baseados na umidade absoluta. Não foi validada uma mudança climática.	Ampliar a aplicação do modelo a regiões com chuva originada por processos convectivos. Utilizar preditores baseados na umidade relativa. Empregar dados contínuos e discretos ou ambos. Analisar melhor as simulações de eventos extremos. Simular quantidades P. Utilizar variáveis da temperatura do ar na superfície. Empregar em conjunto variáveis de mudanças climáticas, de circulação atmosférica e conteúdo de umidade.
Bellone <i>et al.</i> , (2000)	Encontrou-se um máximo local e não global na EM/MCML.	Utilizar modelos com estrutura de correlação espacial mais realística e reduzir as parametrizações. Para a dependência espacial empregar MBAM (Hughes <i>et al.</i> , 1999). Empregar para quantidades P uma distribuição Gama M ou distribuições exponenciais. Reduzir parâmetros inspecionando parâmetros comuns em regiões ou sub-regiões no NHMM. Realizar <i>Downscaling</i> e analisar mudanças climáticas. Utilizar projeções de GCMs
Thyer e Kuczera, (2000)	Analisar a estacionariedade das probabilidades de transição, porque na época não existia modelos para efetuar isso. O comprimento das séries observadas dificultou validar o modelo. Definição dos anos hidrológicos.	Utilizar modelos HMM (3), NHMM (3) com IC e outros modelos AR. Empregar outra forma de definir os anos hidrológicos. Verificar a diferença entre as distribuições dos <i>m</i> estados secos ou úmidos. Realizar previsões. Utilizar um HMM M para identificar regiões com estados climáticos iguais. Utilizar séries de maior comprimento.
Lambert <i>et al.</i> , 2003	Pouca eficiência computacional do NP HMM, porque é mais complexo computacionalmente que o HMM. Dificuldade em simular valores extremos. Encontrou-se problemas iguais aos de Whiting, (2006).	Utilizar o NP HMM já que pode competir com o HMM paramétrico. Considerar o comprimento da série. Aplicar a dados contínuos e discretos.

Tabela A.3 Continuação - Problemas e recomendações dos trabalhos citados na Tabela 1

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Thyer e Kuczera, (2003a)	<p>Tratar com as correlações espaciais altas. Quando se incluem mais locais na análise se eles têm alta correlação não dá grandes ganhos, já que, aumenta o número de parâmetros e a dificuldade de achar os parâmetros certos.</p> <p>Escolher as distribuições a priori dos parâmetros para cada estado no GS quando se tem: dados faltantes e incluem-se mais locais. Viés nos parâmetros.</p>	<p>Determinar as correlações espaciais antes de encontrar os parâmetros. Ter cuidado com altas correlações. Portanto, usar outros métodos para limitar a covariância, por exemplo, em Hughes et al., (1999) a correlação e tratada em função da distância entre locais.</p> <p>Ao tratar com dados faltantes, sempre deve existir um local com dados completos. Porque, aumenta a incerteza pelo surgimento de mais parâmetros, McDonald e Zucchini, (1991) mostram a ML para dados faltantes. Nesse caso ver a aplicabilidade do MA.</p> <p>Considerar se os locais têm os mesmos estados e a analisar a separação desses estados.</p> <p>Analisar melhor o GS para reduzir a reparametrização.</p>
Thyer e Kuczera, (2003b).	<p>Seleção do ano hidrológico.</p> <p>Não foi clara a determinação de regiões de persistência homogênea. Devido à quantidade de dados faltantes surgiram muitos parâmetros extras.</p> <p>Diferença quando se preveem as distribuições posteriores dos parâmetros de um ano para outro, já que, depende do tipo de estado utilizado (seco ou úmido).</p> <p>Encontrou-se que uma sequência de anos com valores baixos não necessariamente é uma seca.</p> <p>Não se logrou entender totalmente o HMM para múltiplos locais.</p> <p>Não convergência no GS quando a série não amostra dois estados claramente, ou não é possível identificar as distribuições da média do estado seco e media do estado úmido, ou se elas são diferentes.</p> <p>Viés nos parâmetros.</p>	<p>HMM(3), ou tratar com estados diferentes aos simples seco e úmido.</p> <p>HMM com variáveis externas e com variação espacial dos estados climáticos. Desenvolver uma metodologia para identificar regiões de persistência homogênea. Usar dados sintéticos para entender melhor o ganho de dados de múltiplos locais e entender melhor a influência de dados faltantes. Testar qual tipo de modelo se ajusta melhor a um sitio, como Hughes <i>et al.</i>, (1999) e Gelman <i>et al.</i>, (1995). Analisar modelos em único local e depois generalizar.</p> <p>Considerar que a análise de dados faltantes só é um ganho se os locais de uma região apresentam os mesmos estados climatológicos.</p>
Whiting, et al., (2003)	Ajuste das distribuições N.	<p>Atenção à definição de ano hidrológico.</p> <p>Criar modelos de regressão para simular IC.</p> <p>Utilizar modelos de regressão para precipitação com índices climáticos como preditores.</p>
Robertson <i>et al.</i> , (2004)	<p>Dados limitados para fazer uma melhor validação cruzada.</p> <p>Superestimação dos períodos de dias com e sem chuva.</p> <p>O EM/MCML encontrou o máximo local.</p>	<p>Utilizar um NHMM sazonal.</p> <p>Empregar SSTs.</p> <p>Realizar previsões.</p> <p>Analisar intensidade de chuva versus ocorrência.</p> <p>Utilizar os resultados como <i>inputs</i> para modelos hidrológicos ou outros modelos.</p> <p>Analisar quantidades de P como recomendou Charles <i>et al.</i>, (1999b).</p> <p>Empregar um AR-HMM.</p> <p>Utilizar uma área maior para definir melhor os estados climáticos e representar melhor a correlação espacial.</p> <p>Aplicar modelos mais sofisticados, modelos Chow-Liu ou os modelos de Hughes e Guttorp (1994b).</p>

Tabela A.3 Continuação - Problemas e recomendações dos trabalhos citados na Tabela 1

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Whiting, et al., (2004)	Definição do ano hidrológico. O HMM tende a degenerar as distribuições de mistura, portanto, não identificou a persistência interanual.	Utilizar um NHMM sazonal. Aplicar um HMM paramétrico com distribuições Log Normais. Empregar modelos HHMM.
Mehrotra e Sharma, (2005)	Problemas relacionados ao KNN.	Empregar um NP NHMM com maior número de estados. Utilizar estados climáticos contínuos o que permite tratar não estacionariedade e eventos extremos. Empregar um NP NHMM para interpretar os estados climáticos e os resultados, já que é útil para analisar mudanças climáticas. Analisar quantidades de P e compará-las com as ocorrências.
Robertson et al., (2005)	Moderadas correlações entre as simulações e observações quando se usam índices relacionados ao ENSO. Escolher os preditores, a escolha é subjetiva. Problemas com os GMC e as SSTs. Capturar a variabilidade interanual da chuva média, por problemas na previsão sazonal da intensidade de chuva..	Aplicar um AR-HMM ou um NHMM sazonal. Investigar se outros padrões atmosféricos criam o mesmo padrão de chuvas. Realizar previsão de valores futuros e <i>Downscaling</i> .
Thyer et al., (2006).	Incerteza nos parâmetros pelas séries comumente usadas (50-100 anos), principalmente em: parâmetros que controlam a persistência, probabilidades de transição nos HMMs e r_1 no AR(1). Séries com separação de estados alta reduziram o tamanho da amostra para calcular os parâmetros no HMM. Maior incerteza nos HMMs.	Utilizar séries de mais de 100 anos, preferivelmente 500 anos. Empregar dados que justifiquem a variabilidade climática como IC ou dados Paleoclimáticos, esses últimos dados devem ser utilizados seguindo as considerações de Haslett et al., 2006. Aplicar um HMM M a vários locais, considerando os problemas e recomendações de Thyer e Kuczera (2003a,b).
Whiting, (2006)	P Anual: o ajuste das distribuições Normais degenera a distribuição de mistura e dá problemas para identificar a persistência e identificar a persistência pelos curtos comprimentos das séries anuais. P Mensal: tratar a não estacionariedade e assumir previamente as distribuições gera problemas na simulação. P em todas as escalas: dificuldade para encontrar o efeito de Hurst pela série ser curta. Relacionar os regimes hidrológicos com as fases do ENSO, algumas características dos índices climáticos não cobrem completamente o fenômeno. Foi tratada a persistência no curto prazo com ARMA, por isso não se encontrou o efeito Hurst, porque esse fenômeno é um efeito de persistência de longo prazo.	P Mensal: Analisar conjuntamente resultados da seleção do modelo com os da calibração e utilizar um HMM (3) com distribuições dependentes LN. P Diária: Analisar valores extremos que aconteceram no mês. P em todas as escalas: Utilizar a escala mensal, porque se identifica melhor a persistência Empregar NP HMMs para identificar as distribuições dependentes antes de usar um HMM. Utilizar <i>Bayes Factor</i> para selecionar o melhor HMM. Empregar HSMM, HHMM. Utilizar HMMs para estudo de risco de secas. Empregar uma análise de persistência através de comprimentos. Melhorar a calibração dos NP HMMs na transição de estados. No NP HMM melhorar a calibração das distribuições condicionais. Utilizar métodos não paramétricos, por exemplo, técnicas <i>Kernel</i> . Realizar previsões e identificar o melhor modelo através de análise bayesiana. NP HMM com informação climática ou utilizar um AR-HMM.

Tabela A.3 Continuação - Problemas e recomendações dos trabalhos citados na Tabela 1

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Greene <i>et al.</i> , (2008)	Representar algumas variações de baixa frequência da Monção Indiana.	Utilizar HMM para outros fenômenos ou P dominada por monções. Realizar <i>Downscaling</i> . Analisar mudanças climáticas. Utilizar validação cruzada empregando simulações e observações.
Gelati <i>et al.</i> , (2010a)	Inferir resultados de mudança climática	Utilizar um modelo com Chow-Li para ocorrências de P
Khalil <i>et al.</i> , (2010)	Alto número de parâmetros nos NHMMs. Subestimação das variâncias das P sazonais e mensais.	Utilizar os resultados como inputs de modelos hidrológicos ou de cultivo. Aplicar para diminuir a incerteza das projeções futuras de GCMs. Cuidado na escolha dos preditores. Analisar a variabilidade espaço temporal. Utilizar validação cruzada empregando vários preditores para analisar o desempenho do modelo.
Lima, (2010)	Nenhum	Utilizar um NHMM para modelar totais anuais e um NHMM M para analisar a variabilidade espacial. Avaliar o número de estados ocultos.
Greene <i>et al.</i> , (2011)	Incerteza nas projeções devido às simulações de GCM.	Incorporar os resultados a modelos hidrológicos ou agrícolas. Utilizar projeções futuras. Aplicar em escala local ou ampliar a escala a ser analisada. Aplicar a outras regiões. Considerar a incerteza de GCM. Analisar a suposição de estacionariedade.
Fu <i>et al.</i> , (2012)	Incerteza pelo conjunto reduzido de projeções de GCMs e porque empregou-se só um método de <i>Downscaling</i> ; e pela natureza das projeções utilizadas. Representação da assimetria por parte dos HMM com distribuições Gama.	Utilizar um Maior número de cenários e projeções futuras para avaliar uma potencial mudança climática. Tratar não estacionariedade. Empregar várias simulações para avaliar a incerteza. Utilizar outros métodos de <i>Downscaling</i> e comparar o desempenho desses métodos e o HMM.
Mares <i>et al.</i> , (2014)	Quando se incluiu o modelo Chow-Li não melhorou o desempenho dos HMMs. Dificuldade de interpretar fisicamente os estados.	Analisar mudanças climáticas. Utilizar mais tele conexões climáticas. Realizar <i>Downscaling</i> . Utilizar um maior número de estados e empregar alguns estados para eventos extremos.

Tabela A.4 -Problemas e recomendações dos trabalhos citados na Tabela 2

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Jackson, (1975)	<p>Determinar os parâmetros do modelo. Na época os métodos dos momentos e de máxima verossimilhança não eram possíveis de realizar pelo gasto computacional alto e o tamanho da amostra ser pequeno.</p> <p>Definir os estados ocultos (1-seco 2-úmido).</p> <p>Focou-se excessivamente na probabilidade de o estado ser seco.</p> <p>Definir as vazões como baixas ou normais.</p>	<p>Utilizar HMMs.</p>
Lu e Berliner, (1999)	<p>Convergência no GS.</p> <p>Um dos modelos AR gerou altas vazões.</p>	<p>Aumentar o número de estados e de <i>Lags</i>. Criar dependência das vazões passadas.</p> <p>Modificar os preditores. Permitir o que um estado surja a partir de um determinado valor do preditor. Melhorar o GS, realizar uma atualização nele que não seja diária.</p>
Thyer e Kuczera, (2000)	<p>Analisar a estacionariedade das probabilidades de transição, na época não existia modelos para efetuar isso.</p> <p>Comprimento das séries observadas dificultou validar o modelo.</p> <p>Definição dos anos hidrológicos.</p>	<p>Utilizar um HMM (3), NHMM (3) com IC e outros AR.</p> <p>Empregar outra forma de definir os anos hidrológicos.</p> <p>Verificar a diferença entre as distribuições dos <i>m</i> estados secos ou úmidos.</p> <p>Realizar previsões.</p> <p>Utilizar um HMM M para identificar regiões com estados climáticas iguais e séries longas.</p>
Lambert et al, 2003	<p>Pouca eficiência computacional do NP HMM, porque é mais complexo computacionalmente que o HMM.</p> <p>Dificuldade em simular valores extremos.</p> <p>Encontrou-se problemas iguais aos de Whiting, (2006).</p>	<p>Utilizar o NP HMM já que pode competir com o HMM paramétrico.</p> <p>Considerar o Comprimento da série.</p> <p>Aplicar a dados contínuos e discretos.</p>
Fortin <i>et al.</i> , (2004)	<p>Escolher as distribuições a priori no GS.</p> <p>Lograr a convergência à distribuição estacionaria no GS.</p> <p>Precisão das previsões pontuais comparado com o ARMA pelo fato da série não ser estacionaria.</p>	<p>Realizar com o SL previsões de valores extremos e previsões probabilísticas.</p> <p>Utilizar modelos SL NP. Aplicar o modelo SL a outros locais, já que permite simultaneamente identificar variações da média e fazer previsões.</p> <p>Empregar um SL com autocorrelação dos resíduos.</p> <p>Aplicar modelos M e permitir a influência da variabilidade climática e o armazenamento na bacia. Aplicar a múltiplos locais, regiões com variações no mesmo tempo e espaço da média e os parâmetros que regem essas variações ao longo prazo.</p> <p>Incluir previsões de precipitação ou índices climáticos.</p> <p>Melhorar o desempenho em previsões comparando o SL e o ARMA aplicado em diferentes locais. Utilizar <i>Pseudo Bayes Factor</i> para classificar as previsões probabilísticas.</p> <p>Aperfeiçoar o GS para reduzir o tempo computacional.</p> <p>Tratar a não estacionariedade: usar simultaneamente, um modelo estacionário e não estacionário sob análises Bayesiana. Com isso testar a estacionariedade da série com um teste de hipóteses. Depois calcular as distribuições posteriores e fazer previsões tratando a não estacionariedade da série com uma mistura dos dois modelos.</p>

Tabela A.4 Continuação - Problemas e recomendações dos trabalhos citados na Tabela 2

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Whiting, et al., (2004)	Definição do ano hidrológico. O HMM tende a degenerar as distribuições de mistura, portanto, não identificou a persistência interanual.	Utilizar um NHMM sazonal. Aplicar um HMM paramétrico com distribuições Log Normais. Empregar modelos HHMM.
Akintug e Rasmussen, (2005)	As distribuições marginais não se ajustaram à distribuição N. Obtive-se sucesso desse ajuste só em alguns casos específicos. O desempenho dos modelos ARMA e HMMs foi refletida nas estatísticas de seca. Dependendo do modelo se tive ou não problemas em representar essas estatísticas. O HMM não simulou as autocorreções em altos <i>Lags</i> .	Utilizar HMM com LN. Aplicar um AR-HMM, mas, é um modelo mais complexo. Cuidado ao interpretar os resultados porque as estatísticas observadas não são uma referência absoluta para fazer comparações das séries sintéticas produzidas por diferentes modelos. Melhorar o ajuste das distribuições Normais, porém mais parâmetros não sempre geram um bom ajuste. Maior foco na incerteza dos parâmetros, já que geralmente se utilizam séries curtas de dados e o número de parâmetros é alto. Cuidado ao escolher HMMs em estudos de secas, já que, devido aos problemas com <i>Lags</i> altos seria uma escolha conservadora.
Whiting, (2006)	Usar distribuições Norma dá problemas em identificar a persistência porque se degenera a distribuição de mistura. Identificar a persistência pelos curtos comprimentos das séries anuais. Tratar a não estacionariedade. Assumir previamente as distribuições gera problemas na simulação. Dificuldade para encontrar o efeito de Hurst pela série ser curta. Relacionar os regimes hidrológicos com as fases do ENSO, algumas características dos índices climáticos não cobrem completamente o fenômeno. Tratou-se a persistência no curto prazo com ARMA, por isso não se tem efeito Hurst porque ele é um efeito de persistência de longo prazo.	Usar NP HMMs para identificar as distribuições dependentes antes de usar um HMM. Analisar conjuntamente resultados da seleção do modelo com os da calibração. HMM (3) com distribuições dependentes LN. Melhor usar a escala mensal, se identifica melhor a persistência. <i>Bayes Factor</i> para selecionar o melhor HMM. Usar HSMM, HHMM. Usar HMMs para estudo de risco de secas. Usar uma análise de persistência através de comprimentos de anos secos ou úmidos. Melhorar nos NP HMMs as calibrações na parte da representação da transição de estados e as distribuições condicionais. Usar métodos não paramétricos, exemplo, técnicas <i>Kernel</i> . Simular extremos de vazões com NP HMMs. Realizar previsões e identificar o melhor modelo através de análise bayesiana. NP HMM com informação climática. Usar AR-HMM
Gelati et al., (2010b)	Por baixas correlações entre os IC e as vazões não foi possível ter 3 estados ocultos em correspondência às fases do ENSO. Problemas para previsões de anomalias negativas de vazões. Encontrou-se que La Niña não influencia as anomalias das vazões. Leves subestimações na persistência e autocorrelação.	Utilizar as previsões para estudos de operação de reservatórios. Ampliar a escala. Analisar múltiplos locais e a correlação espacial. Identificar a forçante climática para as anomalias negativas em vazões. Desenvolver um EM mais confiável para refinar os parâmetros.

Tabela A.4 Continuação - Problemas e recomendações dos trabalhos citados na Tabela 2

Trabalho	Problemas	Modelos Sugeridos / Recomendações
Bracken, (2011)	<p>Valores negativos nas simulações, especialmente nos mínimos.</p> <p>Sobre estimação da assimetria e geração de valores altos por parte do HMM com distribuições Gama.</p> <p>Os HMM sobrestimam o r1.</p> <p>Calculo da ACF.</p> <p>Desempenho modesto dos HMMs com relação às simulações na previsão, embora, superiores ao AR(1).</p> <p>Nas previsões os HMM são piores comparados com o AR(1), quando o estado desde o qual se faz a previsão é seco, dessa forma o AR(1), é melhor. Para melhorar isso foi ajustado um HMM (3).</p>	<p>HMM com distribuições Gama para simulações, já que não geram valores negativos.</p> <p>Em previsões melhor um HMM com distribuições Normais.</p> <p>Realizar a decodificação global.</p> <p>Realizar decodificação global e relacioná-la a processos atmosféricos.</p> <p>Incorporar previsões de vazões a modelos de tomada de decisão.</p> <p>Analisar características do espectro da série.</p> <p>Decodificação global para melhor entendimento do sistema.</p>
Hernandez (2013)	<p>Selecionar o número de estados.</p> <p>Utilizar um teste formal de hipóteses de ajuste das distribuições dependentes. Encontrou-se que as simulações com os índices de maior correlação não representaram as estatísticas básica como a média.</p> <p>O critério de avaliação do desempenho entre modelos não foi claro, muitos modelos classificam.</p> <p>Divergência entre os critérios BIC e AIC.</p> <p>Representação deficiente da autocorrelação de primeira ordem.</p> <p>O AR(2) gerou vazões muito altas.</p> <p>Teve-se dificuldade com métricas de períodos úmidos.</p> <p>Não representação da assimetria dos HMMs com distribuições Gama.</p>	<p>Validar o modelo.</p> <p>Testar o ajuste das distribuições dependentes.</p> <p>Utilizar HMMs e NHMMs.</p> <p>Utilizar informação climática.</p> <p>Realizar previsões futuras.</p> <p>Utilizar distribuições Gama.</p>
Bracken, <i>et al.</i> , (2014)	<p>Encontrou-se que os mínimos e máximos tiveram valores fora do intervalo de dados disponível, o que gera condições não observadas não úteis para gerenciamento.</p> <p>Superestimação da assimetria, provavelmente pela imposição de um limiar em zero na distribuição Gama.</p> <p>O HMM não capturou as características espectrais.</p>	<p>Utilizar HMM multivariados.</p> <p>Aplicar um HMM com posterior desagregação, exemplo, Hughes <i>et al.</i>, (1999).</p> <p>Empregar um HMM ou NHMM para modelar índices climáticos conjuntamente com vazões e criar modelos M, formando uma análise Hierárquica Bayesiana, como em Cooley <i>et al.</i>, (2012). Realizar previsões.</p> <p>Utilizar um HMM U com posterior desagregação.</p> <p>Realizar análise espectral.</p> <p>Aplicar a outras bacias.</p> <p>Realizar Previsões sazonais.</p>

Tabela A.4 Continuação - Problemas e recomendações dos trabalhos citados na Tabela 2

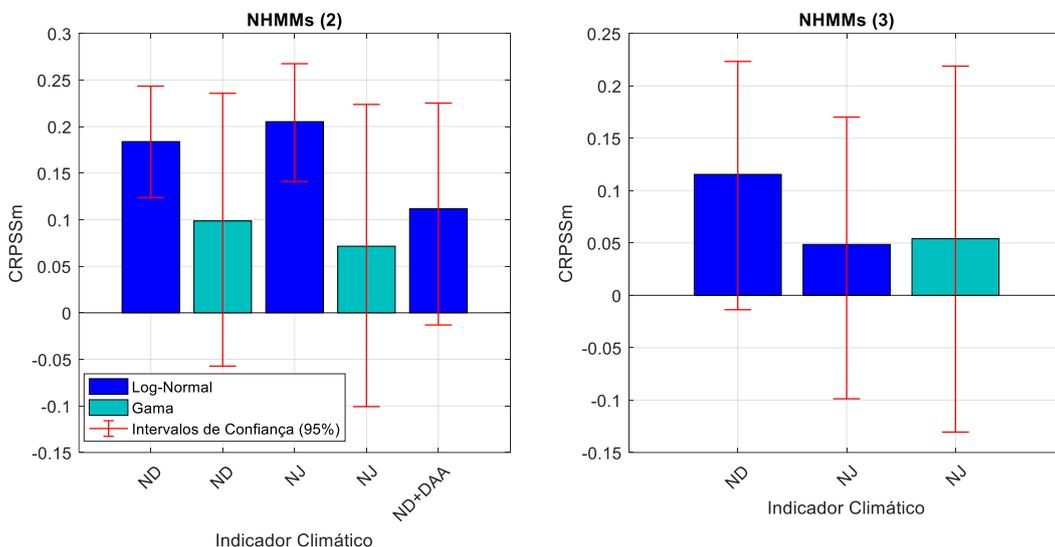
Trabalho	Problemas	Modelos Sugeridos / Recomendações
----------	-----------	-----------------------------------

Turner e Galelli, (2016)	Comprimento da série	<p>Utilizar séries de maior comprimento.</p> <p>Aplicar técnicas de reconstrução Paleoclimática de dados.</p> <p>Analisar eventos de secas extremas em outras regiões.</p> <p>Incluir HMMs para analisar operações de reservatórios.</p> <p>Utilizar a metodologia de operação proposta em tempo real.</p> <p>Condicionar o modelo a tele conexões climáticas.</p> <p>Atualizar o modelo com nova informação.</p> <p>Utilizar outra abordagem, exemplo: modelos de base física.</p> <p>Considerar modelos estacionários, já que, ainda são validos para o atual século.</p> <p>Aplicar a um sistema de múltiplos reservatórios.</p>
Bracken, et al., (2016)	<p>O NHMM comparado com modelos com outros modelos de reconstrução tende a produzir altos valores de máximos e mínimos. Os autores indicam que pode ser pelo uso de distribuições dependentes Lognormais.</p> <p>O NHMM também sub estima a correlação espacial entre as estações.</p>	<p>Geração de séries sintéticas com NHMMs.</p> <p>Usar em redes de reservatórios ou redes de estações de medição de vazões.</p> <p>Usar outro tipo de distribuições dependentes.</p> <p>Usar outro tipo de copula.</p> <p>Adicionar informação física da bacia.</p> <p>Outros locais com diferentes condições hidro climáticas.</p> <p>Outras covariáveis hidro meteorológicas, como, umidade do solo ou temperaturas.</p> <p>Usar uma estrutura hierárquica no modelo para melhorar o tempo de computação.</p>
Liu et al., (2018)	Sobre ajuste quando se utilizaram mais de dois estados ocultos, talvez pela pouca quantidade de dados observados.	<p>Utilizar outros procedimentos para selecionar os preditores, visando reduzir o sobre ajuste do modelo.</p> <p>Utilizar estados ocultos contínuos.</p> <p>Aplicar sistemas de reservatórios e que os reservatórios sejam de grande porte.</p>

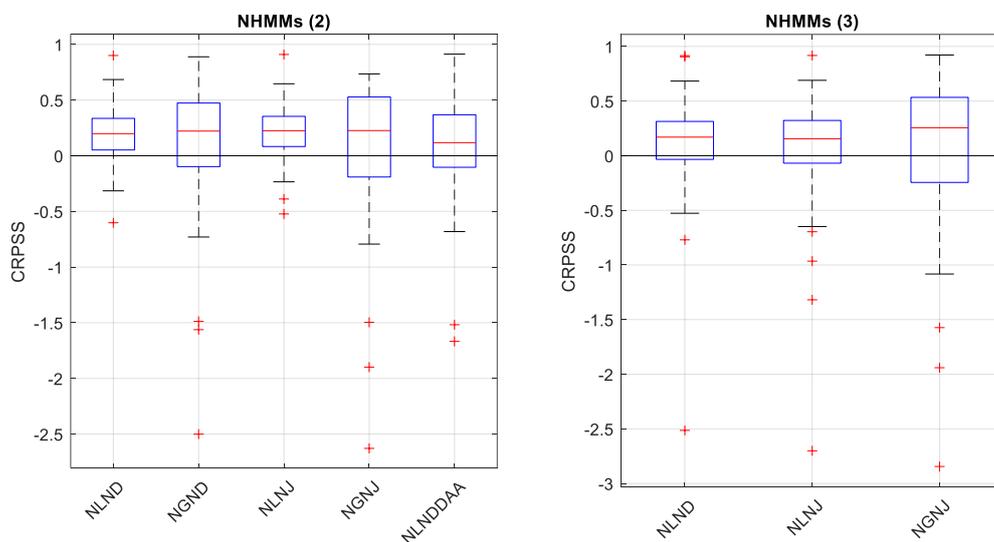
APÊNDICE B

APÊNDICE B – DESEMPENHO DE NHMMs, AR-NHMMs, IHMMs e ARXs e outros modelos

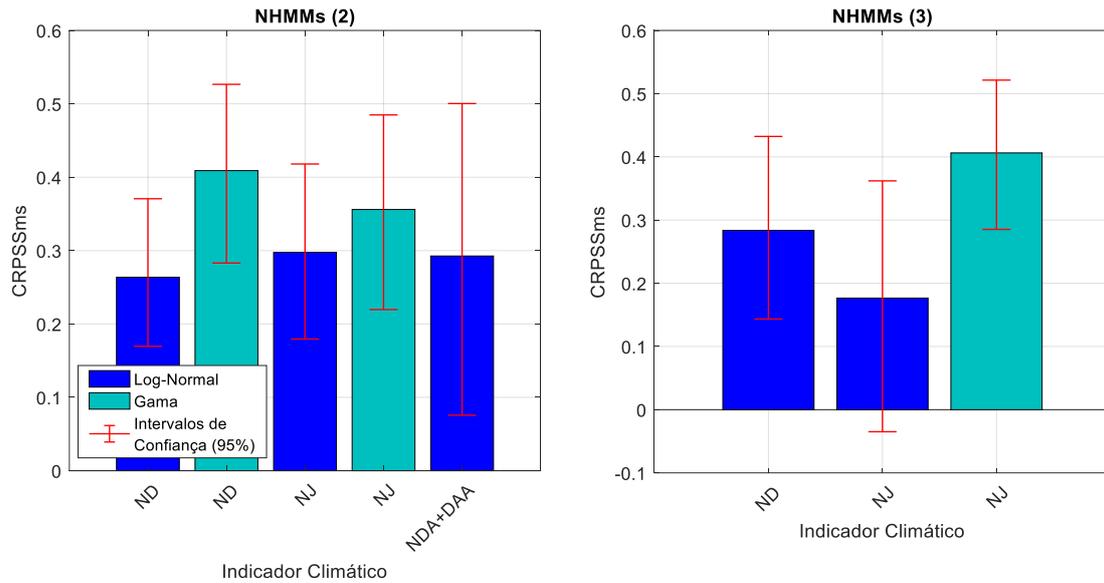
B.1 – DESEMPENHO DE NHMMs



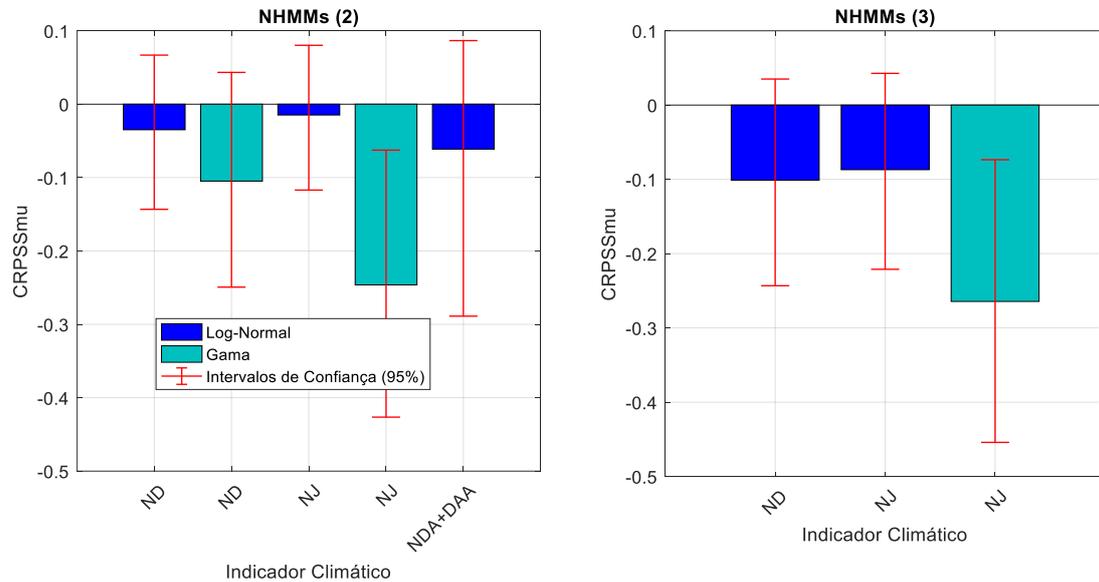
B. 1- Valores do CRPSS médio (CRPSSm) dos 60 anos do período de validação dos NHMMs de 2 (esquerda) e 3 (direita) estados ocultos. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*.



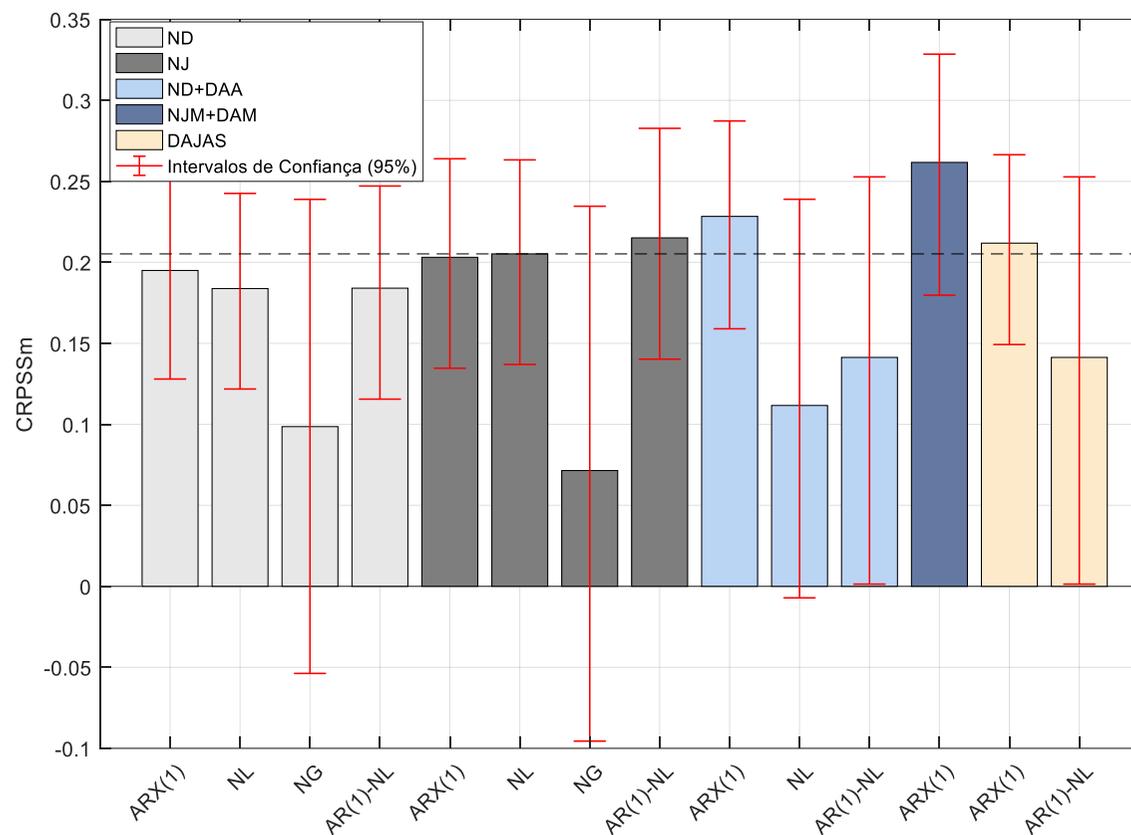
B. 2- Valores do CRPSS para os 60 anos do período de validação dos NHMMs de 2 (esquerda) e 3 (direita) estados ocultos. NL: NHMM Log-Normal, NG: NHMM Gama, ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*. A linha horizontal representa o desempenho da climatologia.



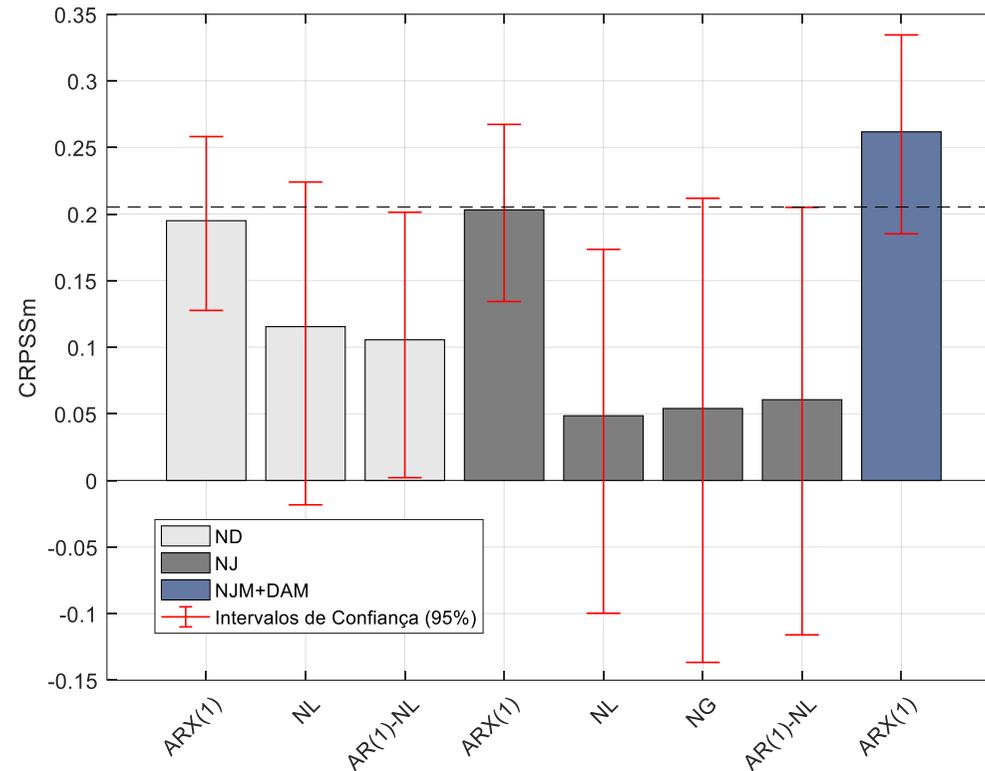
B. 3- Valores do CRPSS médio para os anos secos (CRPSSms) dos 60 anos do período de validação dos NHMMs de 2 (esquerda) e 3 (direita) estados ocultos. Os anos secos são definidos quando a vazão é inferior ao percentil de 33%. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*.



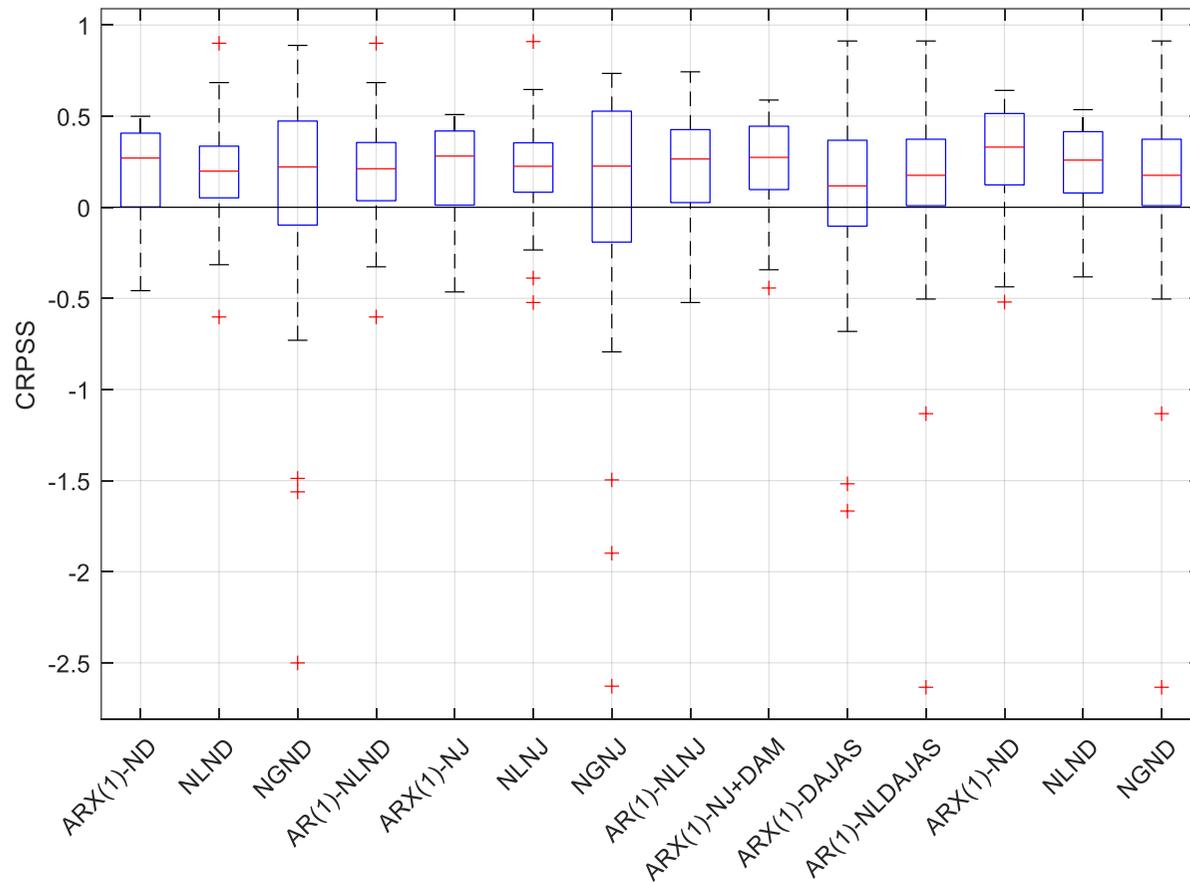
B. 4- Valores do CRPSS médio para os anos úmidos (CRPSSmu) dos 60 anos do período de validação dos NHMMs de 2 (esquerda) e 3 (direita) estados ocultos. Os anos úmidos são definidos quando a vazão é superior ao percentil de 66%. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*.



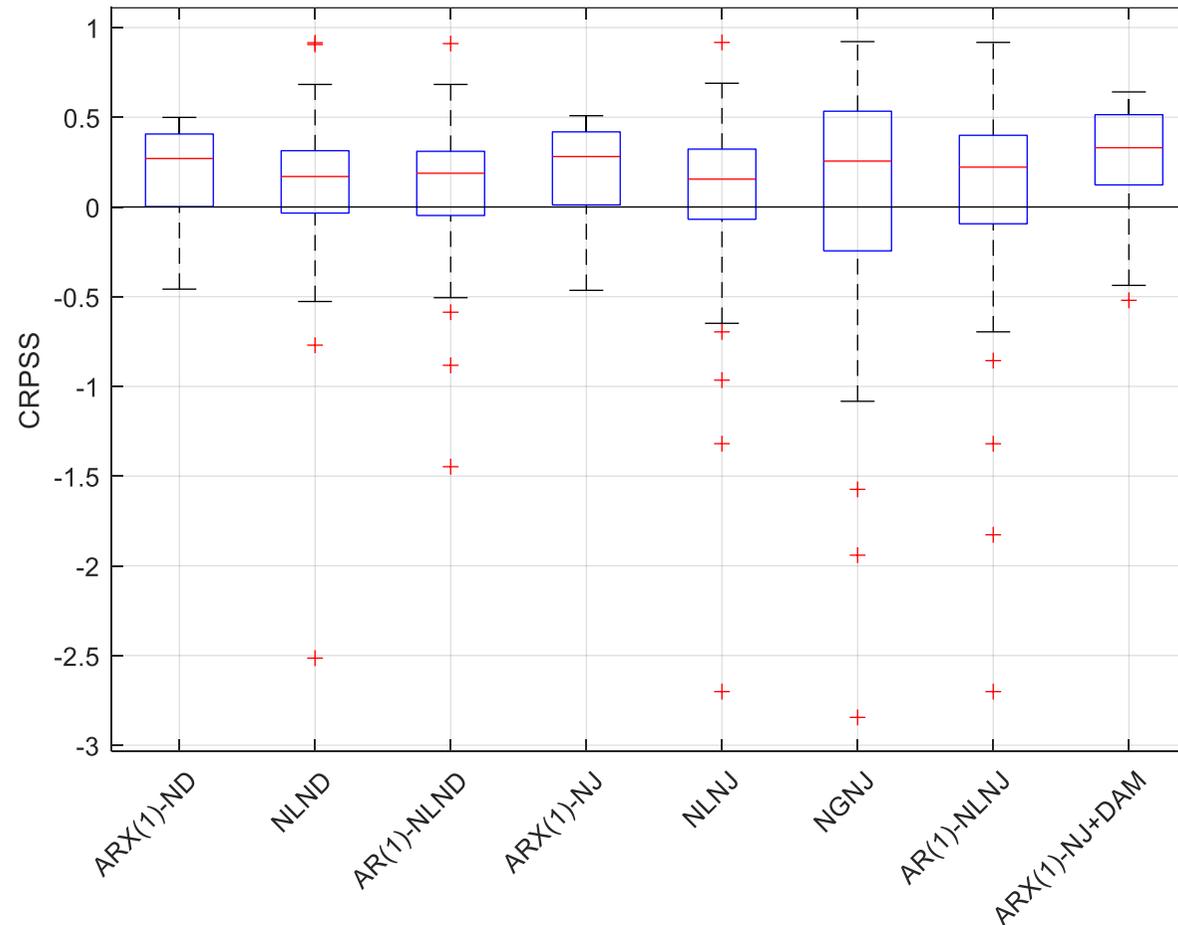
B. 5- Valores do CRPSS médio (CRPSSm) dos 60 anos do período de validação dos NHMMs e AR-NHMMs de 2 estados ocultos. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. NL: NHMM Log-Normal. NG: NHMM Gama. AR(1) e ARX(1): Modelos autoregressivos de primeira ordem sem e com variáveis exógenas, respectivamente. Os intervalos de confiança foram calculados via *Bootstrap*. A linha tracejada representa o desempenho do NLNJ.



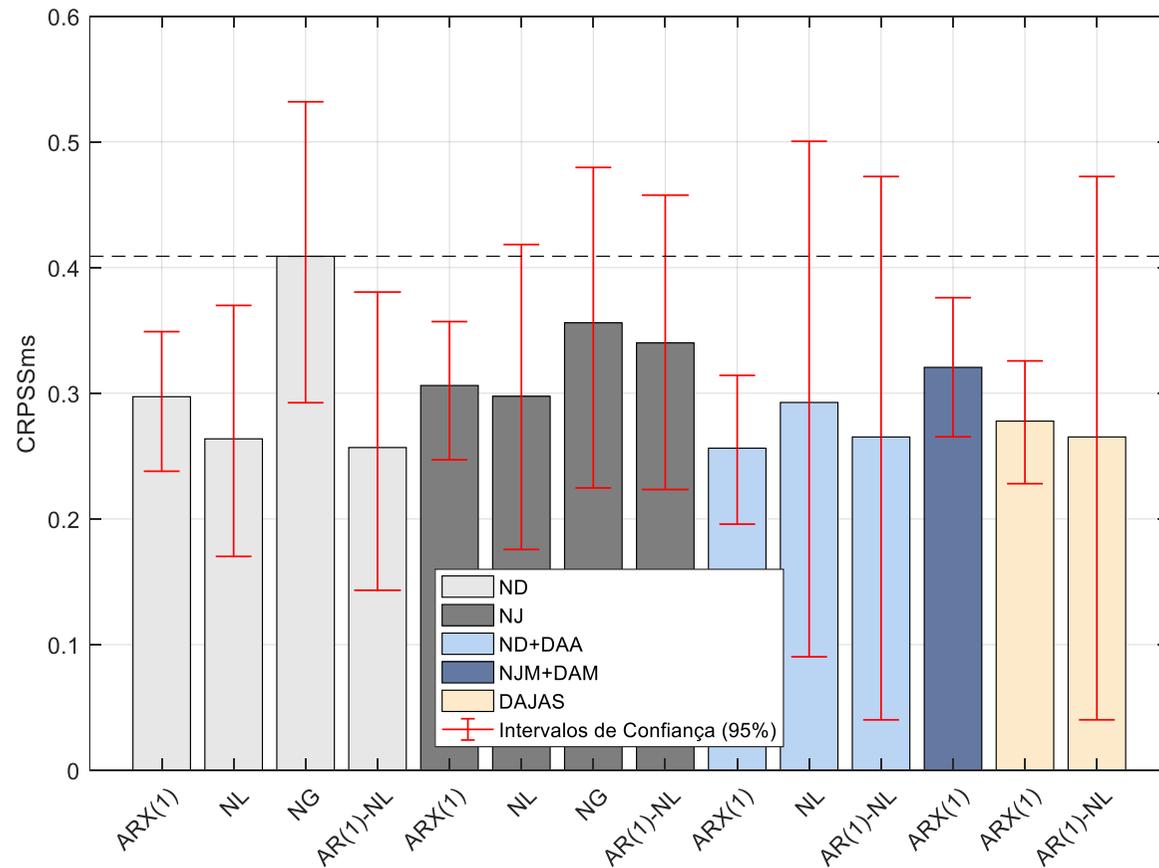
B. 6- Valores do CRPSS médio (CRPSSm) dos 60 anos do período de validação dos NHMMs e AR-NHMMs de 2 estados ocultos. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. NL: NHMM Log-Normal. NG: NHMM Gama. AR(1) e ARX(1): Modelos autoregressivos de primeira ordem sem e com variáveis exógenas, respectivamente. Os intervalos de confiança foram calculados via *Bootstrap*. A linha tracejada representa o desempenho do NLNJ com 2 estados ocultos.



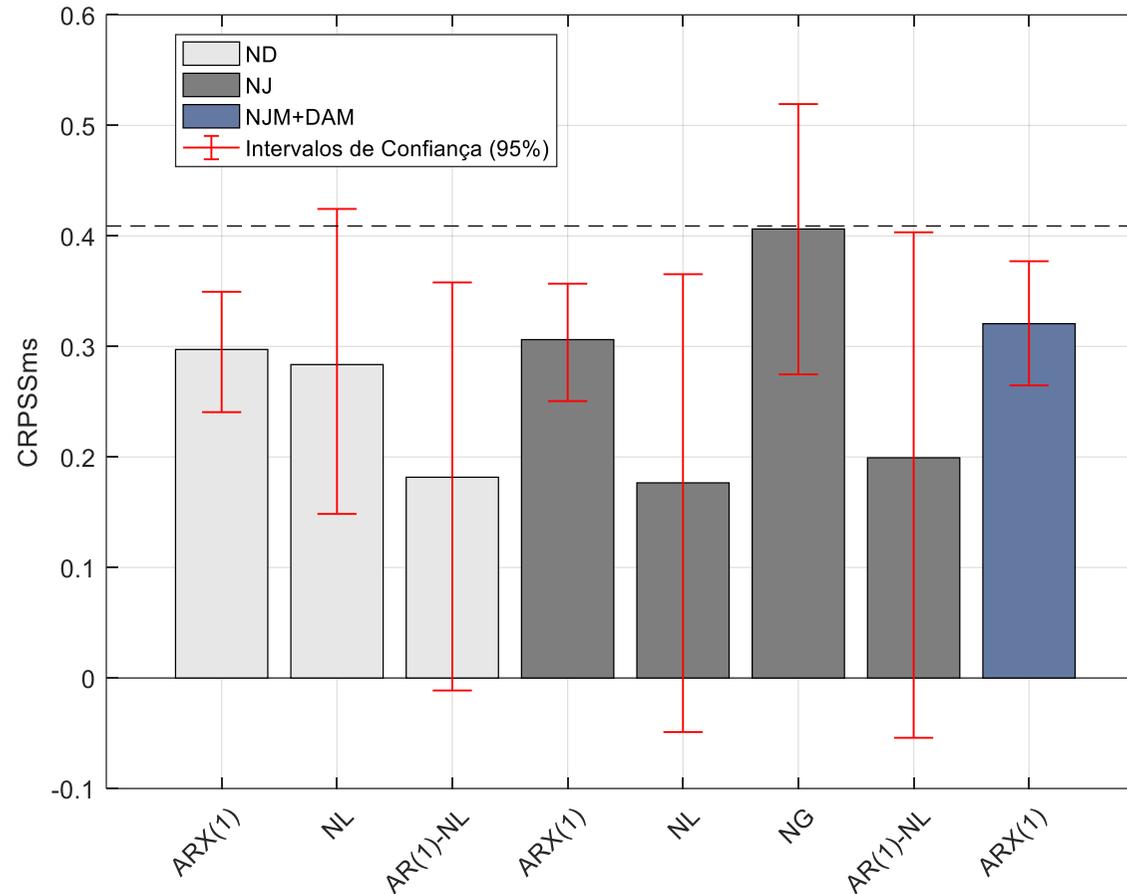
B. 7- Valores do CRPSS para os 60 anos do período de validação dos NHMMs e AR-NHMMs de 2 estados ocultos. NL: NHMM Log-Normal, NG: NHMM Gama, ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. A linha horizontal representa o desempenho da climatologia.



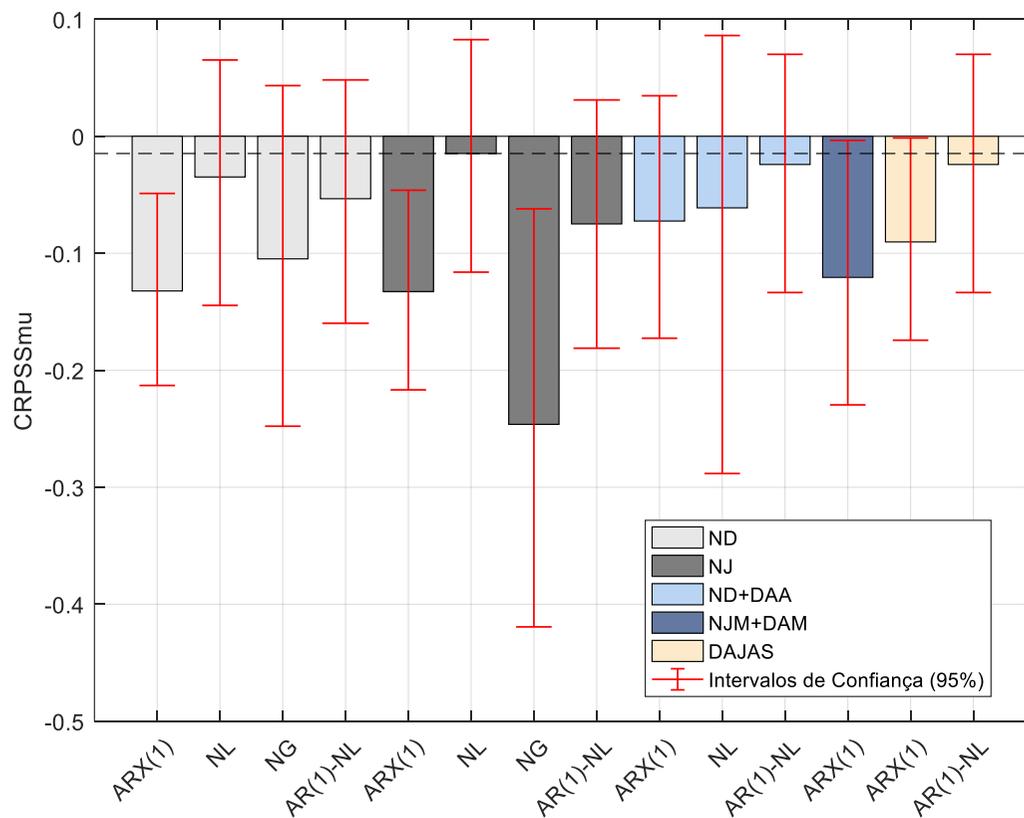
B. 8- Valores do CRPSS para os 60 anos do período de validação dos NHMMs e AR-NHMMs de 3 estados ocultos. NL: NHMM Log-Normal, NG: NHMM Gama, ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. A linha horizontal representa o desempenho da climatologia.



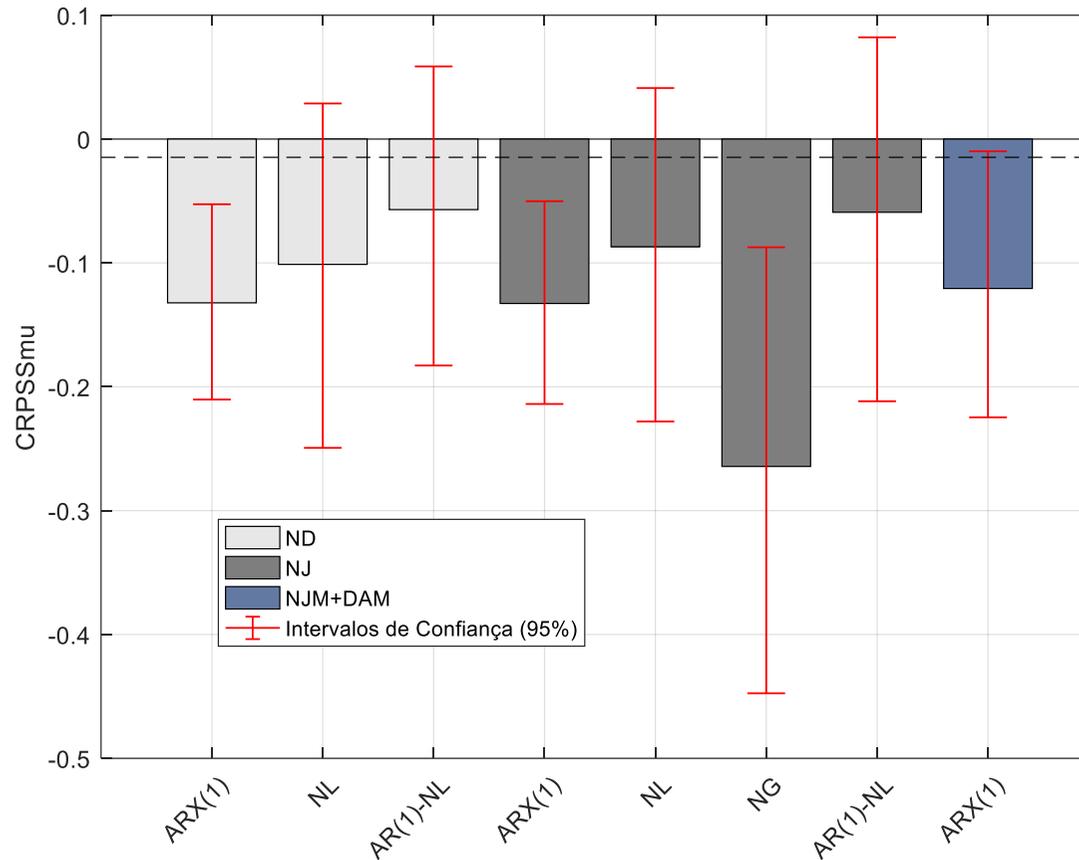
B. 9- Valores do CRPSS médio para os anos secos (CRPSSms) dos 60 anos do período de validação dos NHMMs e AR-NHMMs de 2 estados ocultos. Os anos secos são definidos quando a vazão é inferior ao percentil de 33%. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*. A linha horizontal tracejada representa o desempenho do NGND.



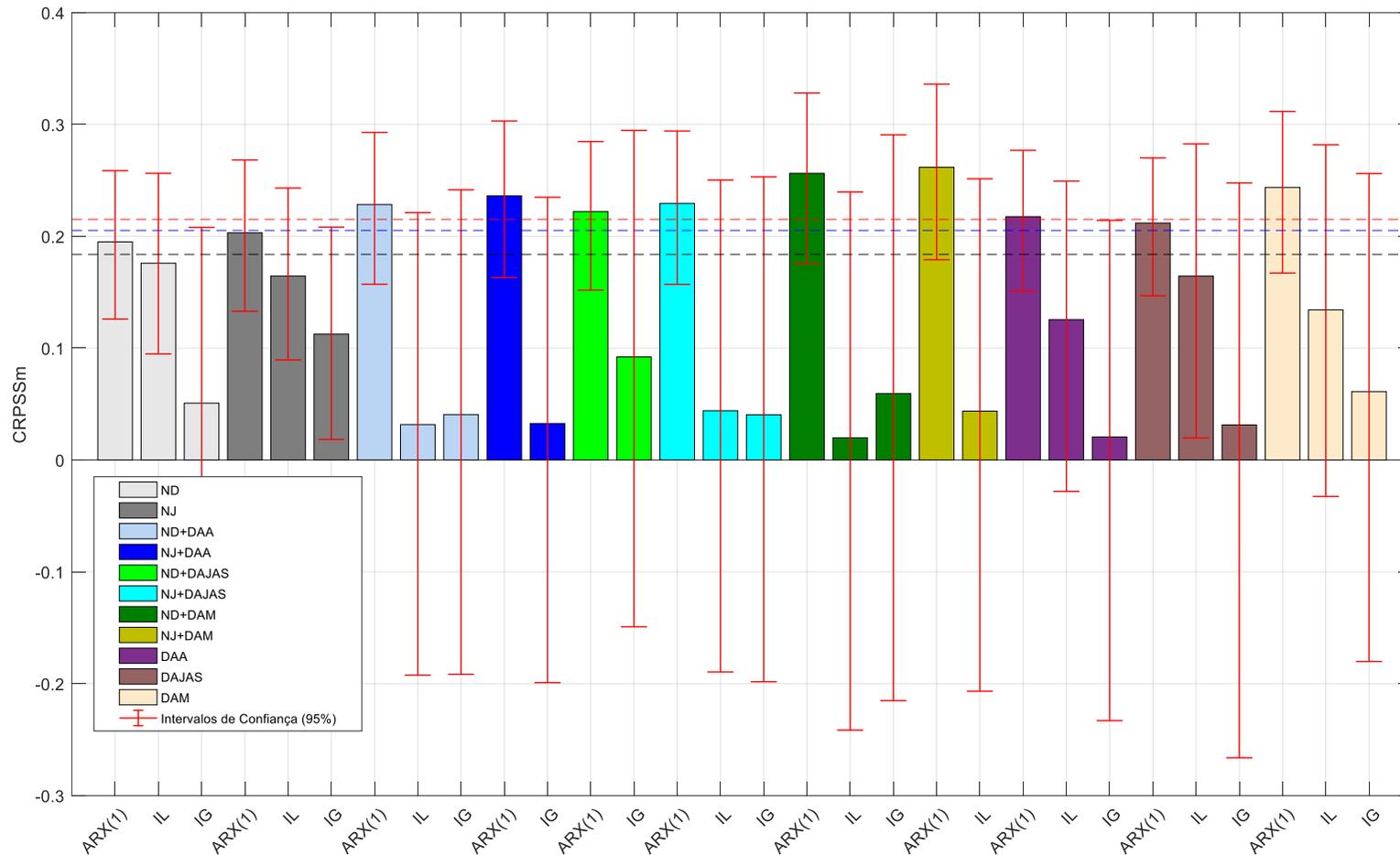
B. 10- Valores do CRPSS médio para os anos secos (CRPSSms) dos 60 anos do período de validação dos NHMMs e AR-NHMMs de 3 estados ocultos. Os anos secos são definidos quando a vazão é inferior ao percentil de 33%. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*. A linha horizontal tracejada representa o desempenho do NGND com 2 estados ocultos.



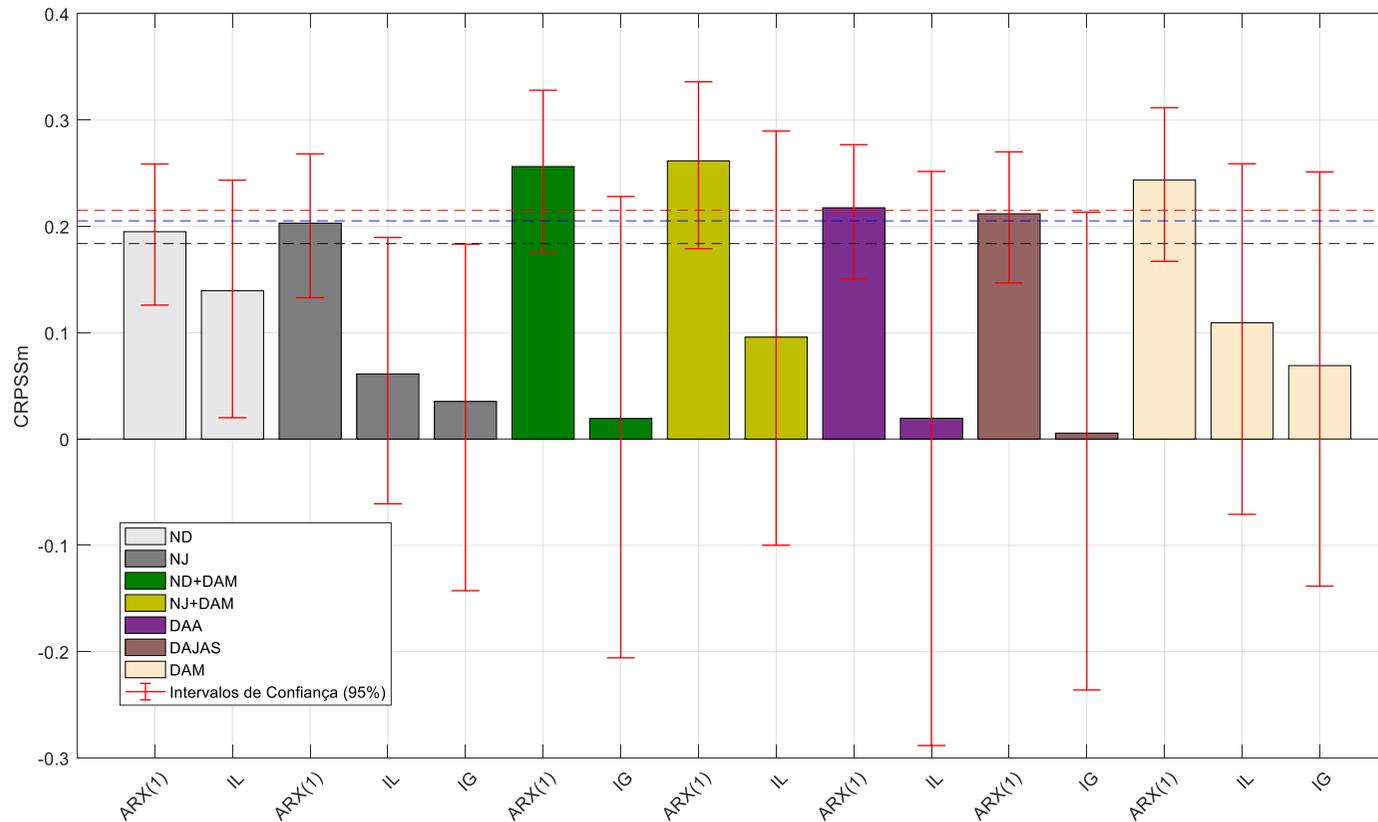
B. 11- Valores do CRPSS médio para os anos úmidos (CRPSSmu) dos 60 anos do período de validação dos NHMMs e AR-NHMMs de 2 estados ocultos. Os anos úmidos são definidos quando a vazão é superior ao percentil de 66%. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*. A linha horizontal tracejada representa o desempenho do NLNJ.



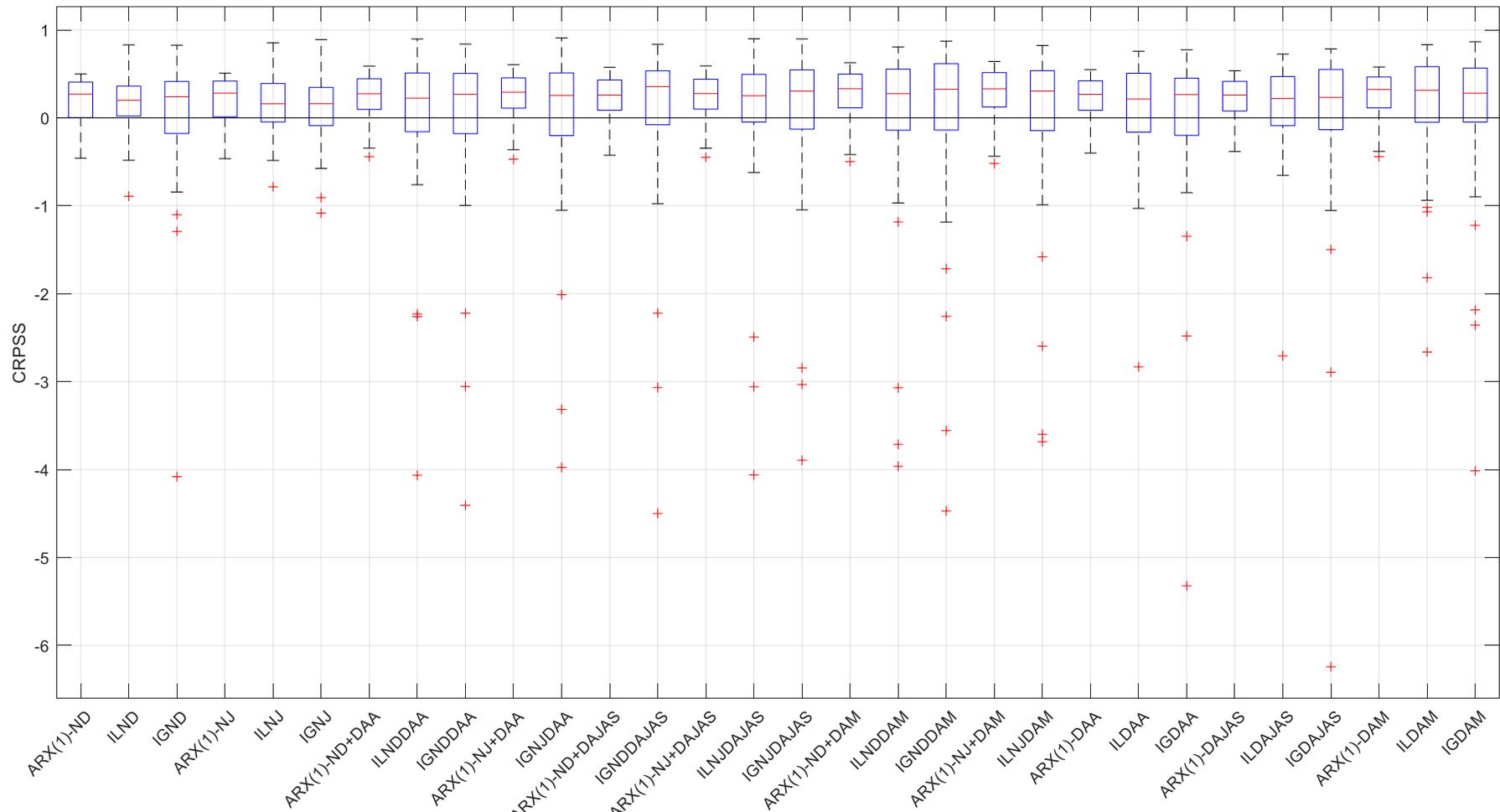
B. 12- Valores do CRPSS médio para os anos úmidos (CRPSSmu) dos 60 anos do período de validação dos NHMMs e AR-NHMMs de 3 estados ocultos. Os anos úmidos são definidos quando a vazão é superior ao percentil de 66%. ND: Indicador NINO3 do mês de dezembro do ano anterior em que acontece a afluência observada, NJ: Indicador NINO3 do mês de janeiro do mesmo ano em que acontece a afluência observada, DAA: Indicador Dipolo do Atlântico do mês de agosto do ano anterior em que acontece a afluência observada, DAJAS: Indicador Dipolo do Atlântico do período julho-agosto-setembro do ano anterior em que acontece a afluência observada, DAM: Média do indicador Dipolo do Atlântico do ano anterior em que acontece a afluência observada. Os intervalos de confiança foram calculados via *Bootstrap*. A linha horizontal tracejada representa o desempenho do NLNJ com 2 estados ocultos.



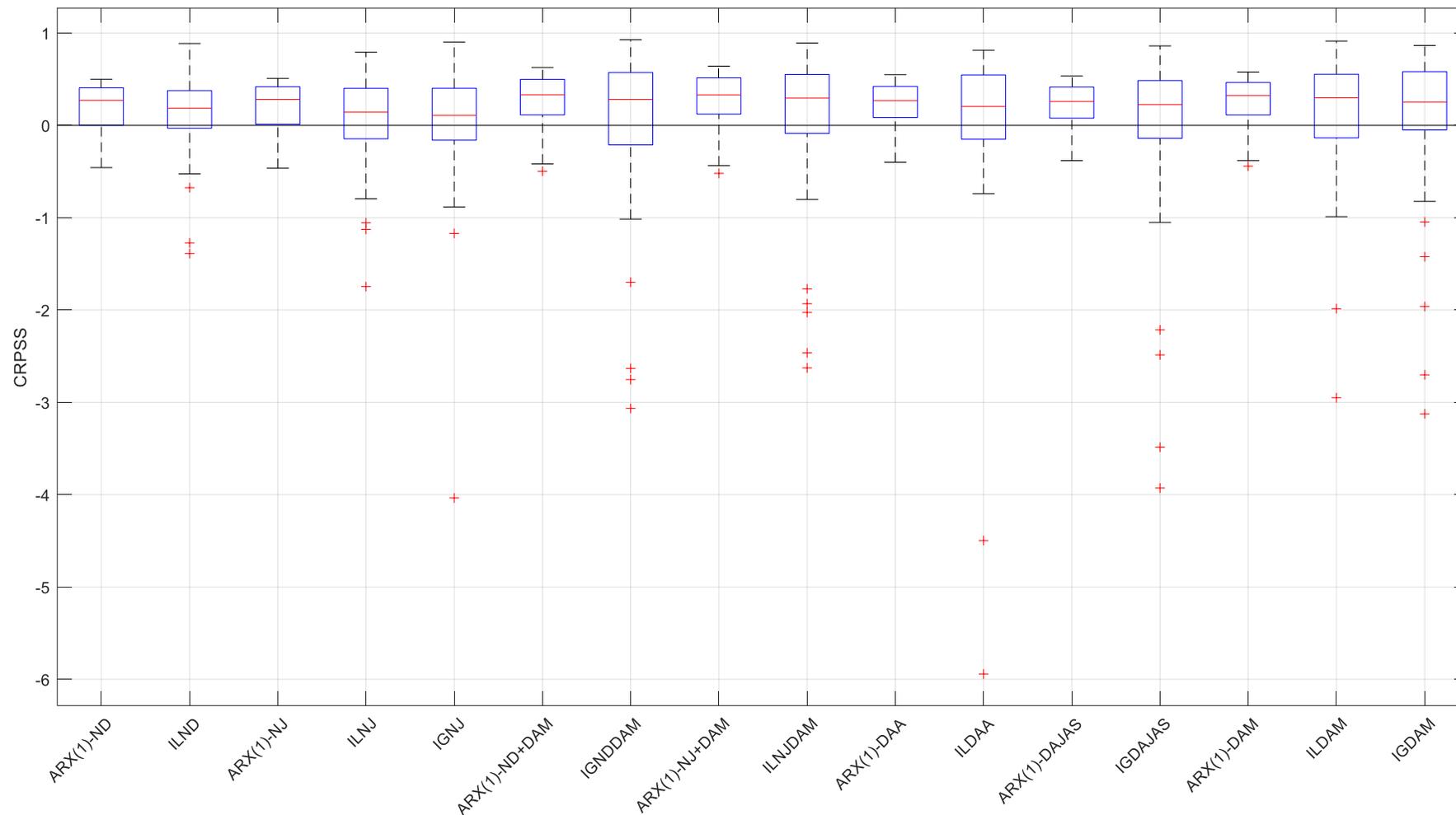
B. 13 - Valores do CRPSS médio (CRPSSm) dos 60 anos do período de validação dos IHMMs 2 estados ocultos. Os indicadores climáticos correspondem a aqueles mostrados na tabela 6.5. Os intervalos de confiança de 95% foram calculados via *Bootstrap*. A linha horizontal vermelha representa o desempenho do AR(1)-NLNJ com 2 estados ocultos. A linha horizontal azul representa o desempenho do NLNJ com 2 estados ocultos. A linha horizontal preta representa o desempenho do NLND com 2 estados ocultos. IL: IHMM Log-Normal. IG: IHMM Gama. ARX(1): Modelo Autoregressivo de primeira ordem com variáveis exógenas



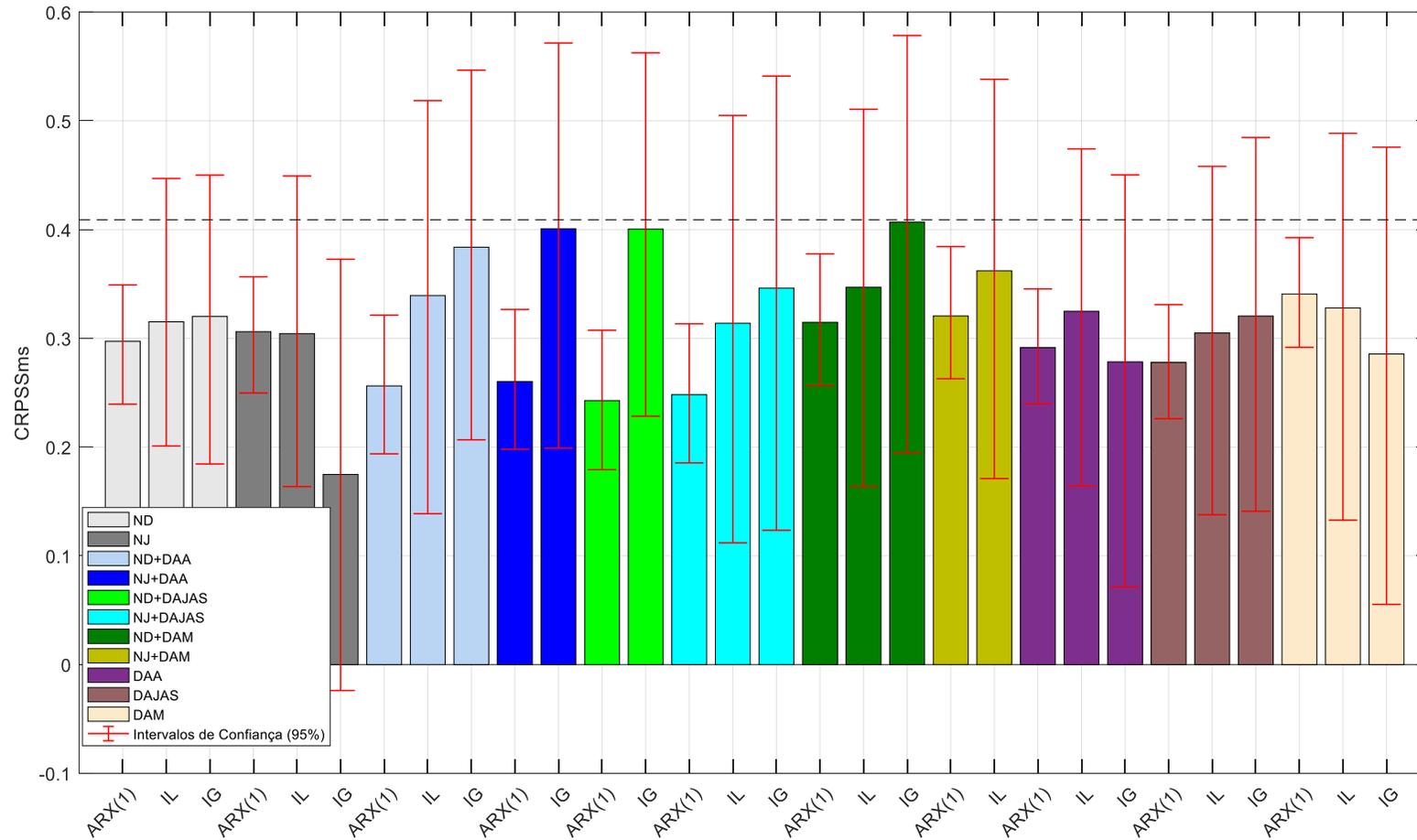
B. 14 - Valores do CRPSS médio (CRPSSm) dos 60 anos do período de validação dos IHMMs 3 estados ocultos. Os indicadores climáticos correspondem a aqueles mostrados na tabela 6.5. Os intervalos de confiança de 95% foram calculados via *Bootstrap*. A linha horizontal vermelha representa o desempenho do AR(1)-NLNJ com 2 estados ocultos. A linha horizontal azul representa o desempenho do NLNJ com 2 estados ocultos. A linha horizontal preta representa o desempenho do NLND com 2 estados ocultos. IL: IHMM Log-Normal. IG: IHMM Gama. ARX(1): Modelo Autoregressivo de primeira ordem com variáveis exógenas



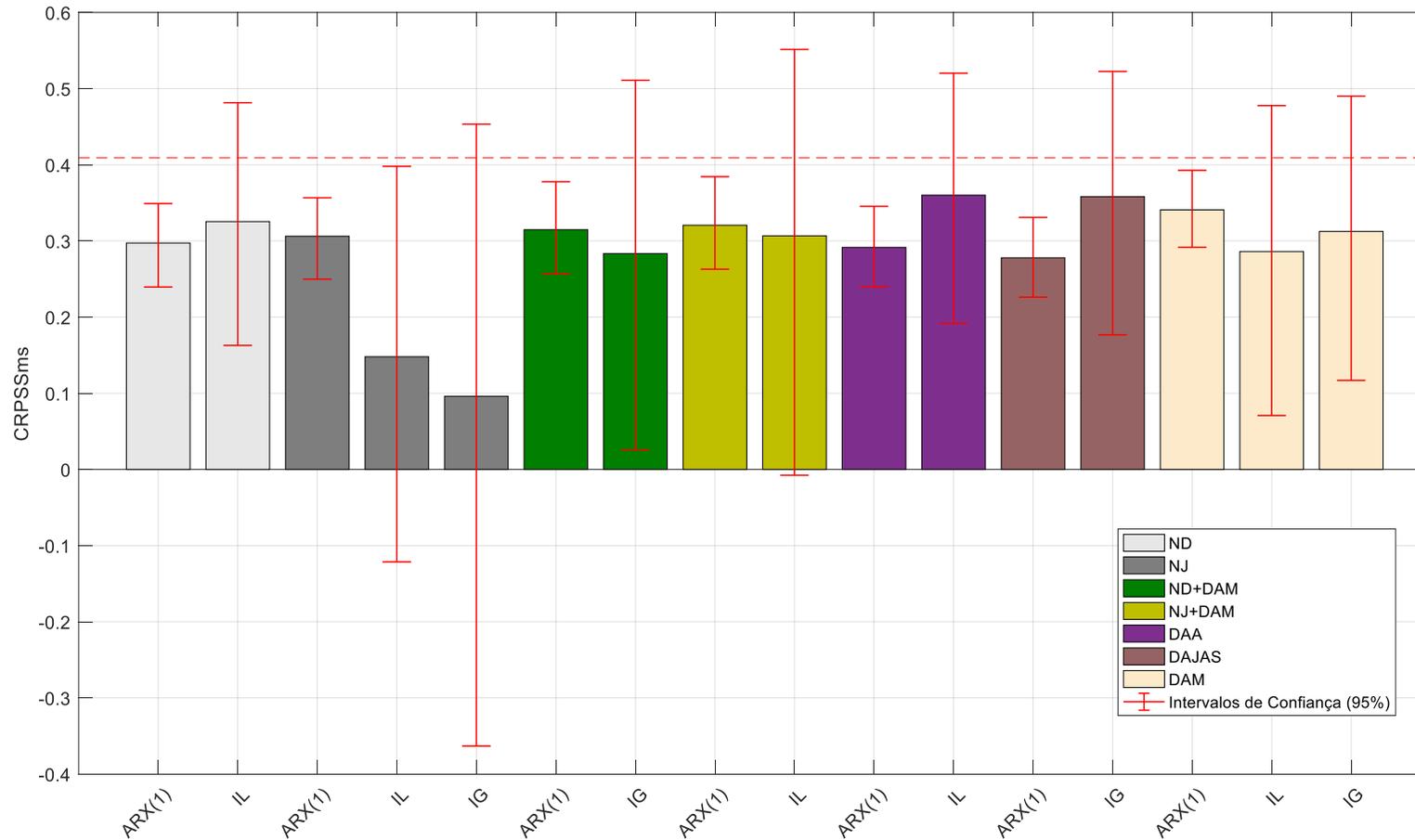
B. 15 - Valores do CRPSS dos 60 anos do período de validação dos IHMMs 2 estados ocultos. Os indicadores climáticos correspondem a aqueles mostrados na tabela 6.5. A linha horizontal representa o desempenho da climatologia



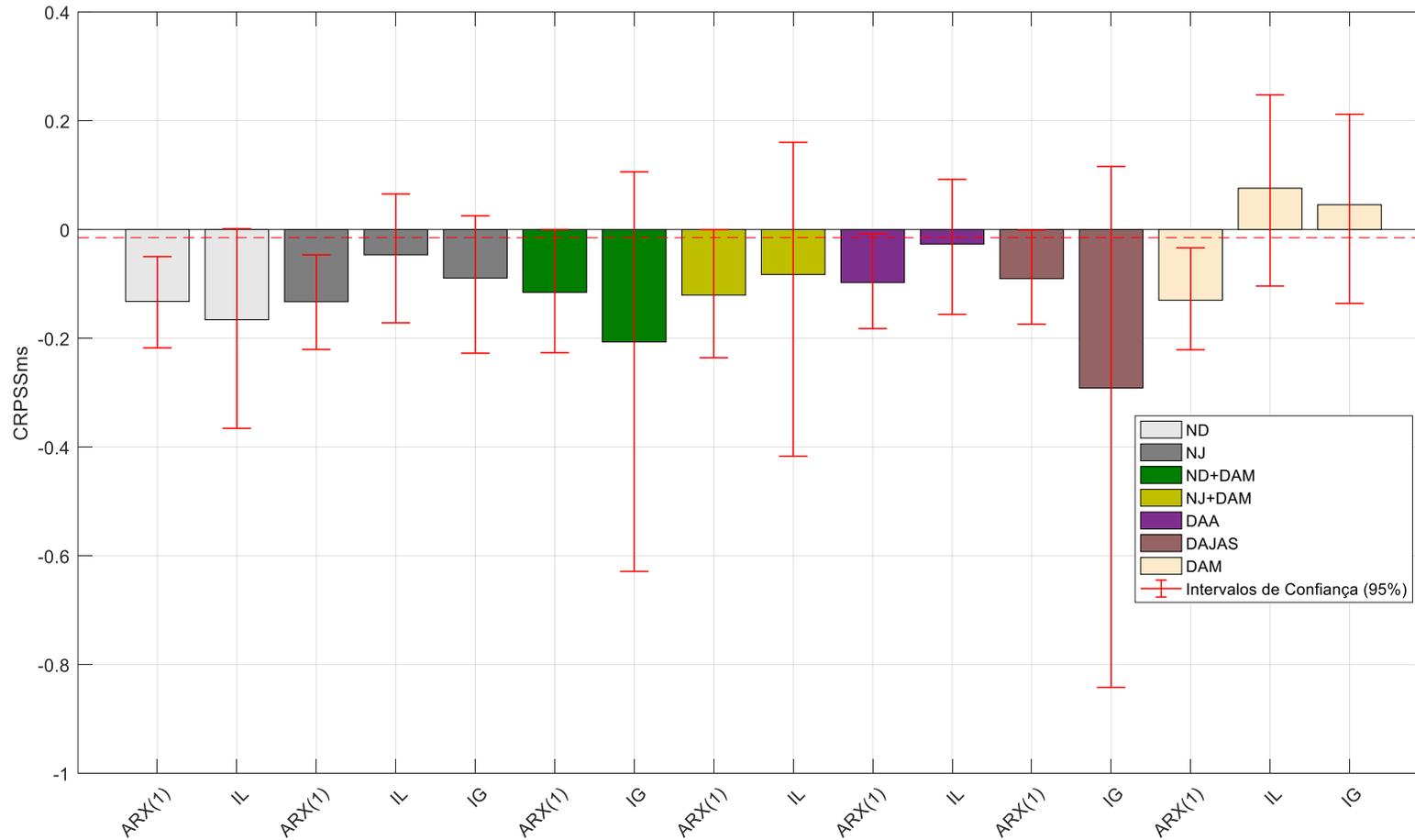
B. 16 - Valores do CRPSS dos 60 anos do período de validação dos IHMMs 3 estados ocultos. Os indicadores climáticos correspondem a aqueles mostrados na tabela 6.5. A linha horizontal representa o desempenho da climatologia



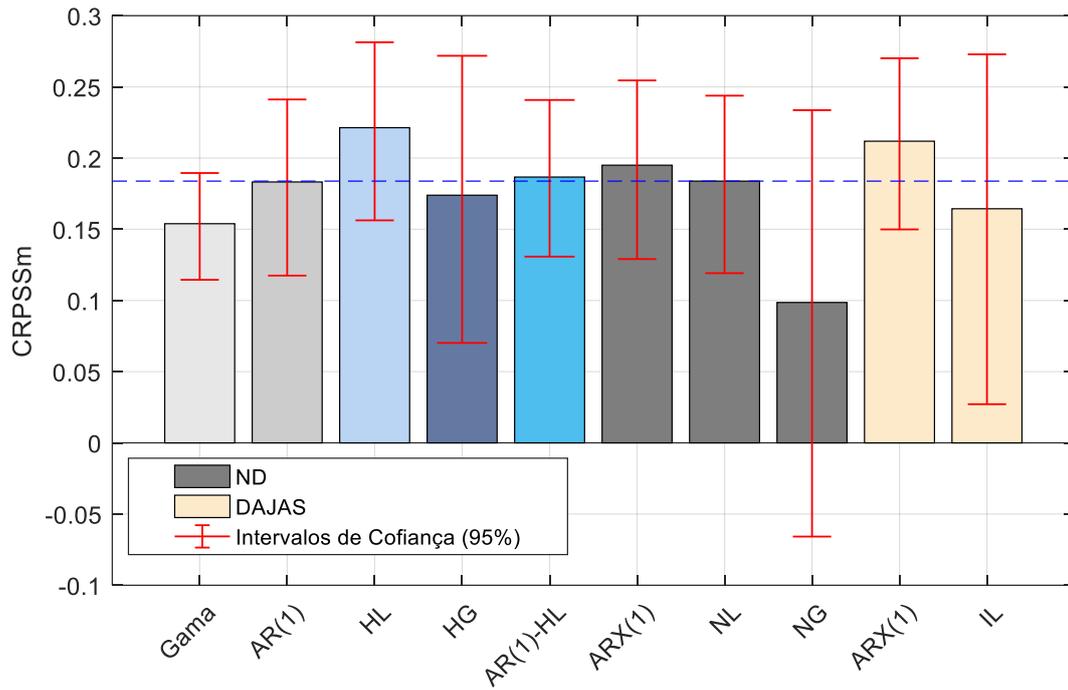
B. 17 - Valores do CRPSS médio dos anos secos (CRPSSms) dos 60 anos do período de validação dos IHMMs 2 estados ocultos. Os anos secos são definidos como os anos em que a vazão é inferior ao percentil de 33%. Os indicadores climáticos correspondem a aqueles mostrados na tabela 6.5. Os intervalos de confiança de 95% foram calculados via *Bootstrap*. A linha horizontal preta representa o desempenho do NGND com 2 estados ocultos. IL: IHMM Log-Normal. IG: IHMM Gama. ARX(1): Modelo Autoregressivo de primeira ordem com variáveis exógenas



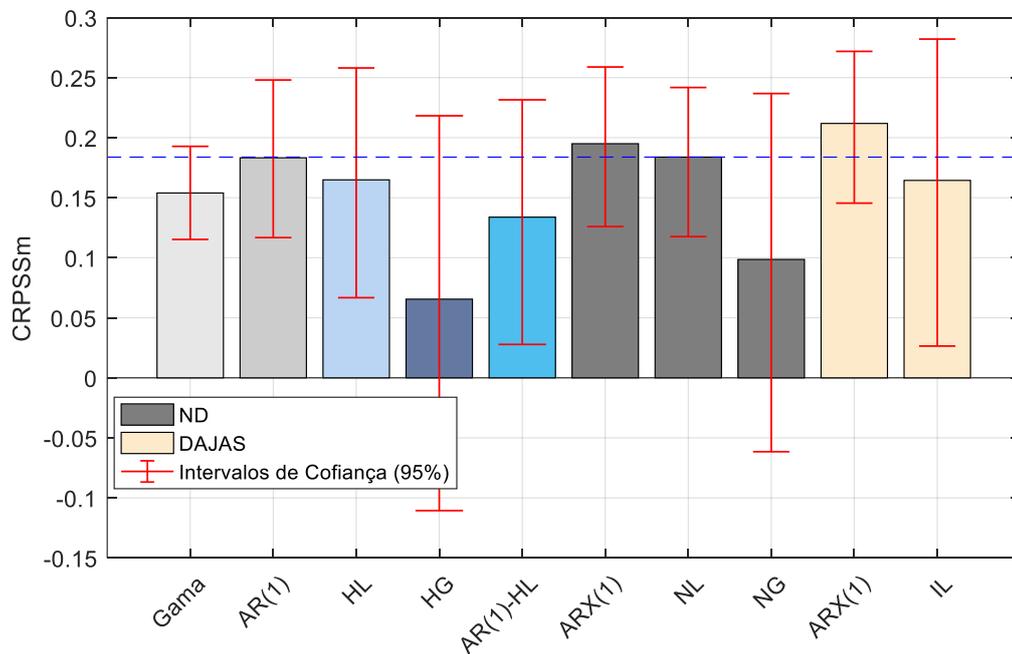
B. 18 - Valores do CRPSS médio dos anos secos (CRPSSms) dos 60 anos do período de validação dos IHMMs 3 estados ocultos. Os anos secos são definidos como os anos em que a vazão é inferior ao percentil de 33%. Os indicadores climáticos correspondem a aqueles mostrados na tabela 6.5. Os intervalos de confiança de 95% foram calculados via *Bootstrap*. A linha horizontal vermelha representa o desempenho do NGND com 2 estados ocultos. IL: IHMM Log-Normal. IG: IHMM Gama. ARX(1): Modelo Autoregressivo de primeira ordem com variáveis exógenas



B. 20 - Valores do CRPSS médio dos anos úmidos (CRPSS_{mu}) dos 60 anos do período de validação dos IHMMs 3 estados ocultos. Os anos úmidos são definidos como os anos em que a vazão é superior ao percentil de 66%. Os indicadores climáticos correspondem a aqueles mostrados na tabela 6.5. Os intervalos de confiança de 95% foram calculados via *Bootstrap*. A linha horizontal vermelha representa o desempenho do NLNJ com 2 estados ocultos. IL: IHMM Log-Normal. IG: IHMM Gama. ARX(1): Modelo Autoregressivo de primeira ordem com variáveis exógenas

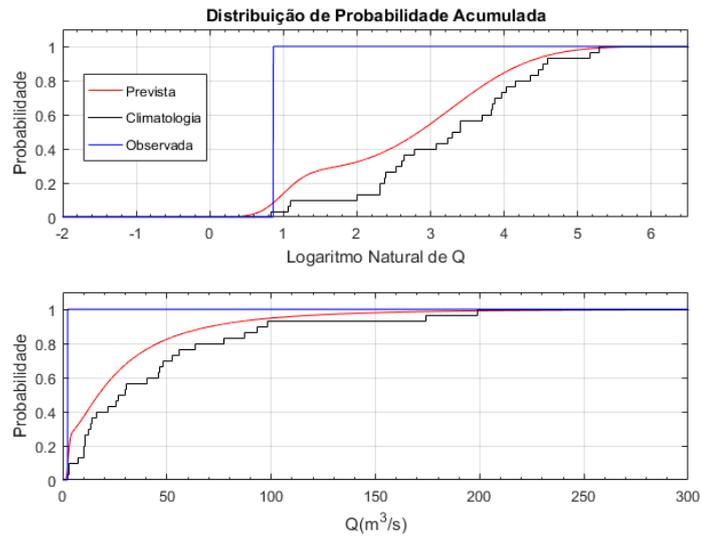


B. 21 - Similar à figura 6.11 inclui os modelos HG: HMM Gama e AR(1)-HL: HMM Autoregressivo Log-Normal. Todos os HMMs têm 2 estados ocultos.

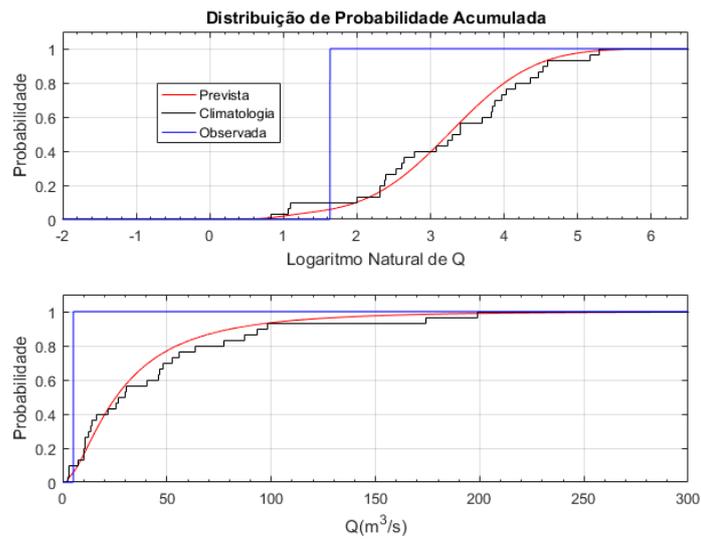


B. 22 - Similar à figura 6.11 inclui os modelos HG: HMM Gama e AR(1)-HL: HMM Autoregressivo Log-Normal. Todos os HMMs têm 3 estados ocultos.

Ano 1993 – CRPSS = 0.4674



Ano 1998 – CRPSS = 0.1094



B. 23 - Forma das Distribuições de previsão acumuladas para anos secos: anos 1993 (a) e 1998 (b).

APÊNDICE C

APÊNDICE C. QUALIDADE DE PREVISÕES PONTUAIS PARA UM ANO NA FRENTE

C.1. – ANOS SECOS DEFINIDOS PELO PERCENTIL HISTÓRICO DE 33%

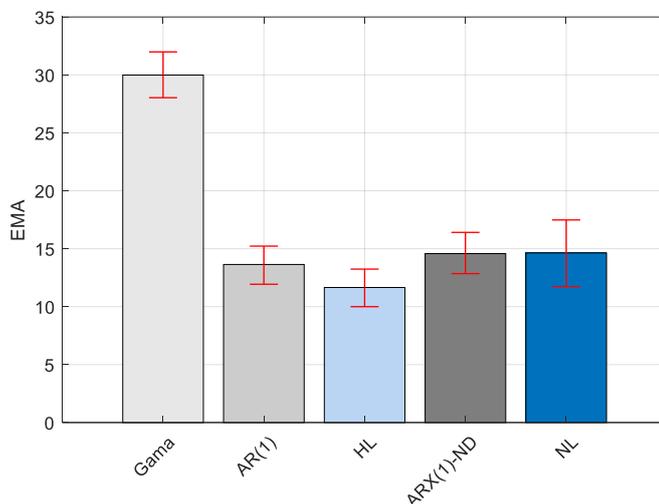


Figura C. 1 Erro Médio Absoluto (EMA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos secos definidos pelo percentil histórico de 33%.

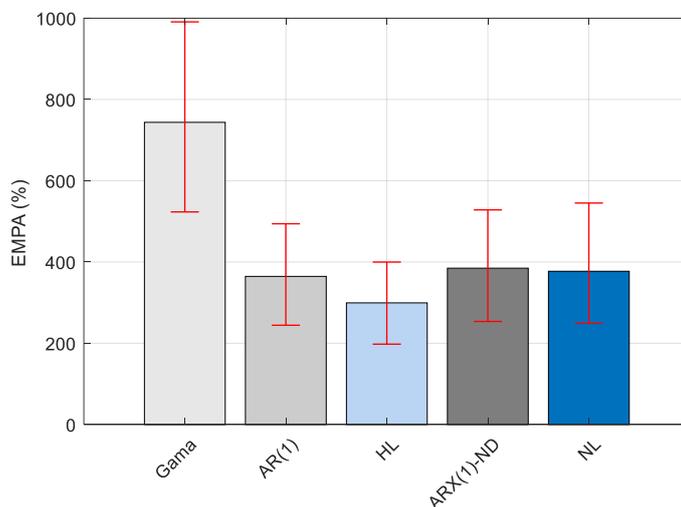


Figura C. 2 Erro Médio Percentual Absoluto (EMPA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos secos definidos pelo percentil histórico de 33%.

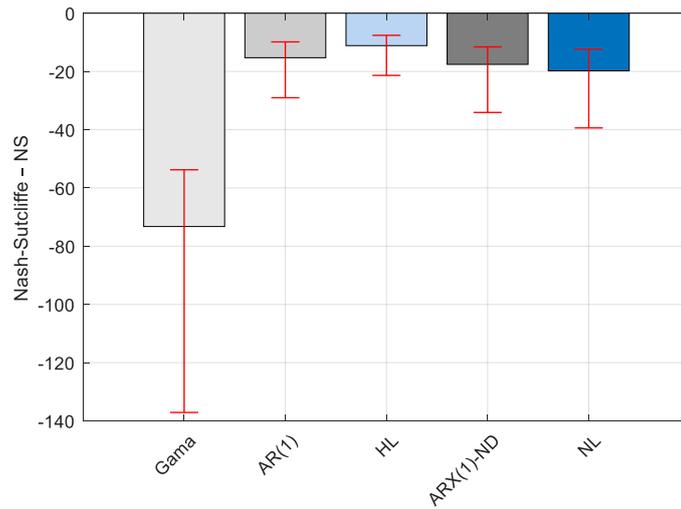


Figura C. 3 Coeficiente de Nash-Sutcliffe (NS) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos secos definidos pelo percentil histórico de 33%.

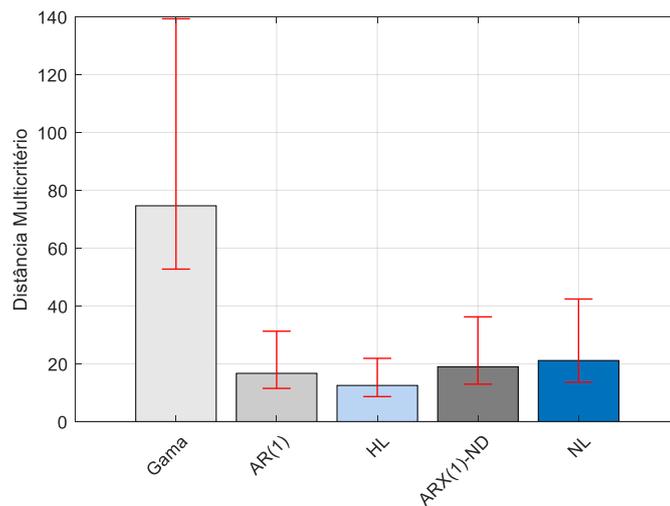


Figura C. 4 Distância Multicritério (DM) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos secos definidos pelo percentil histórico de 33%.

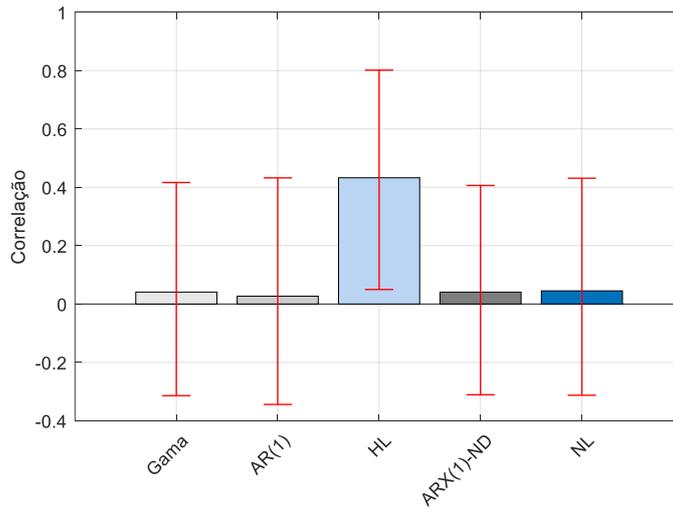


Figura C. 5 Correlação para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos secos definidos pelo percentil histórico de 33%.

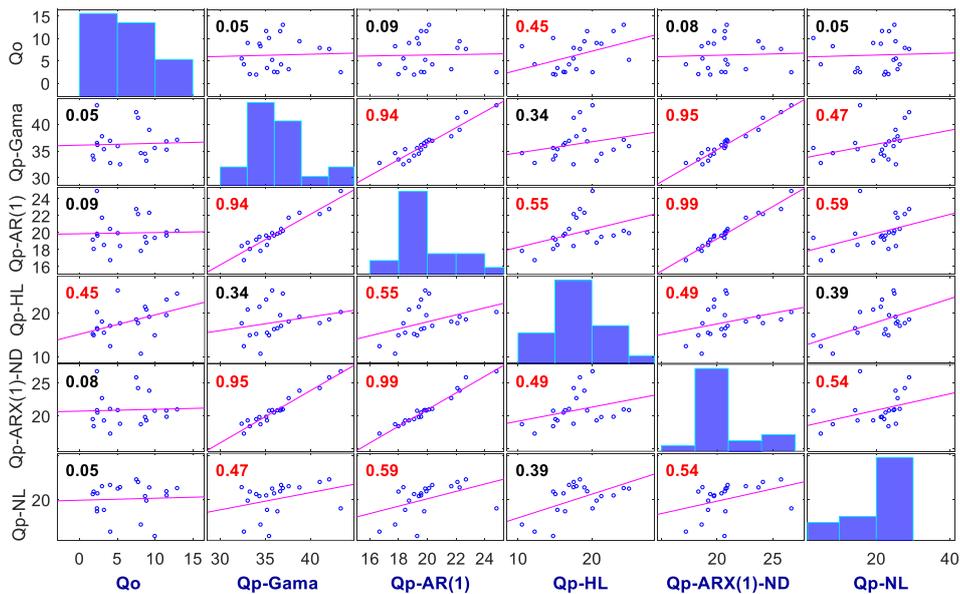


Figura C. 6 - Matriz de correlação tipo Spearman entre as vazões observadas e as médias previstas nos 60 anos do período de validação para pelos modelos Gama, HL e NL e modelos autoregressivos. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Valores para os anos em secos definidos pelo percentil histórico de 33%. Qo: vazão observada; Qp: vazão prevista; Gama: modelo Gama; HL: HMM Log-Normal de 2 estados ocultos; NL: NHMM Log-Normal de 2 estados ocultos que inclui o indicador NINO3 de dezembro do ano anterior.

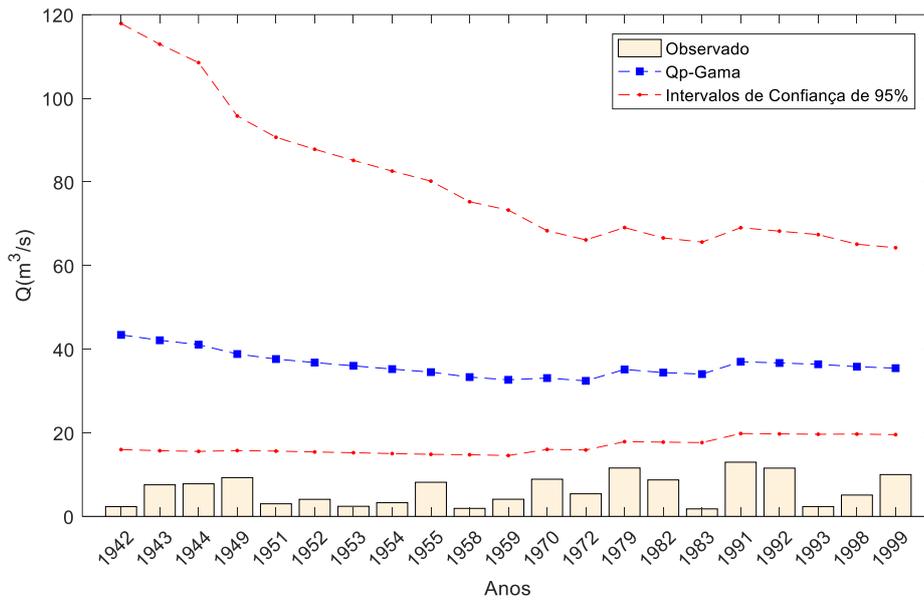


Figura C. 7 - Valores das observações e previsões produzidas pelo modelo Gama (Qp-Gama) para os anos secos definidos pelo percentil histórico de 33%.

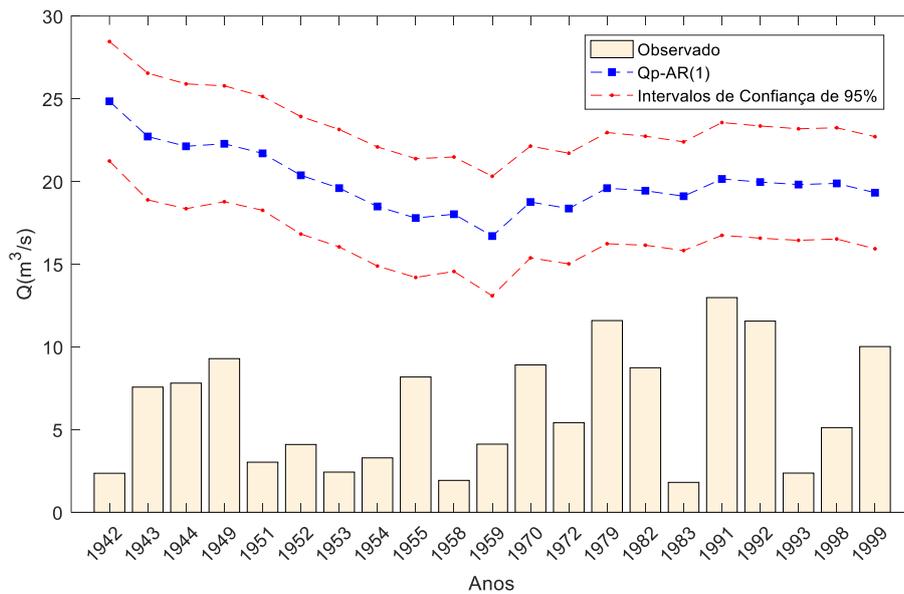


Figura C. 8 - Valores das observações e previsões produzidas pelo modelo AR(1) (Qp-AR(1)) para os anos secos definidos pelo percentil histórico de 33%.

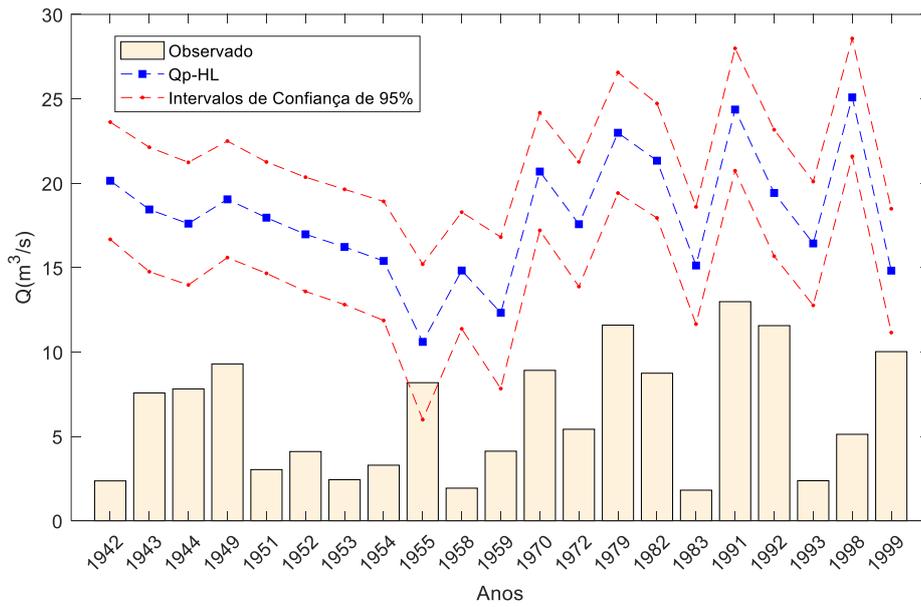


Figura C. 9 - Valores das observações e previsões produzidas pelo modelo HL (Qp-HL) para os anos secos definidos pelo percentil histórico de 33%.

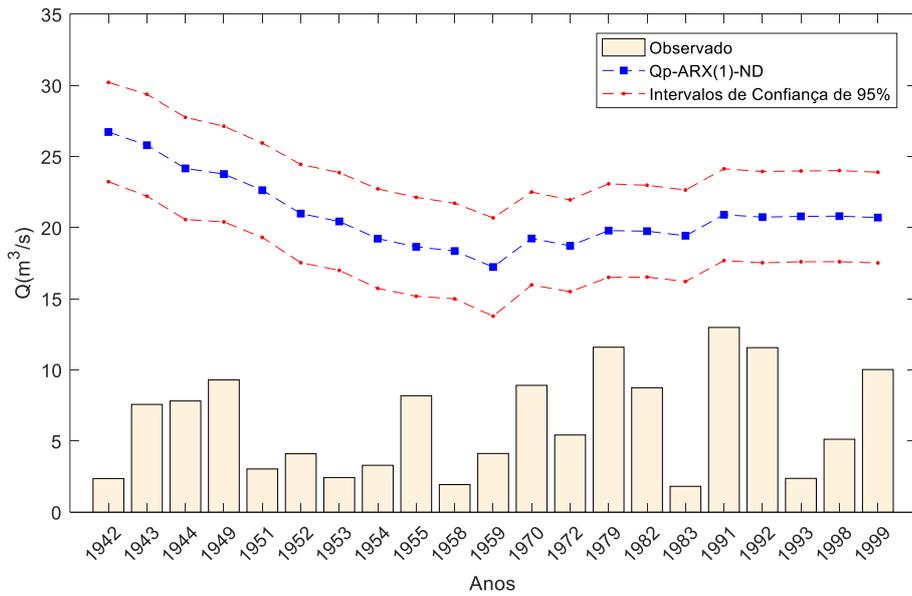


Figura C. 10 - Valores das observações e previsões produzidas pelo modelo ARX(1)-ND (Qp-ARX(1)-ND) para os anos secos definidos pelo percentil histórico de 33%.

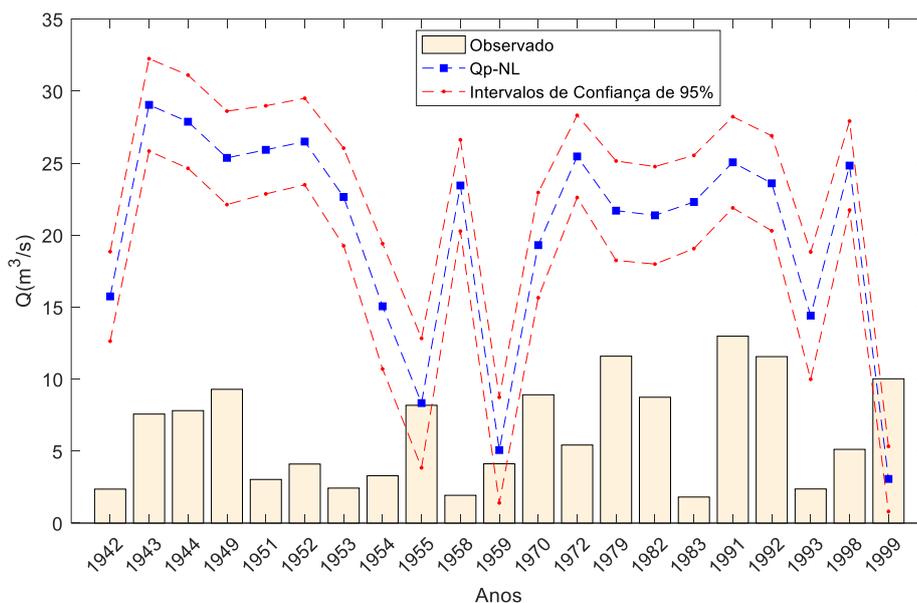


Figura C. 11 - Valores das observações e previsões produzidas pelo modelo NL (Qp-NL) para os anos secos definidos pelo percentil histórico de 33%.

C.1. – ANOS EM QUE o CRPSS é maior ou igual ao CRPSSm do modelo Gama

Anos em que $CRPSS \geq 0,15$

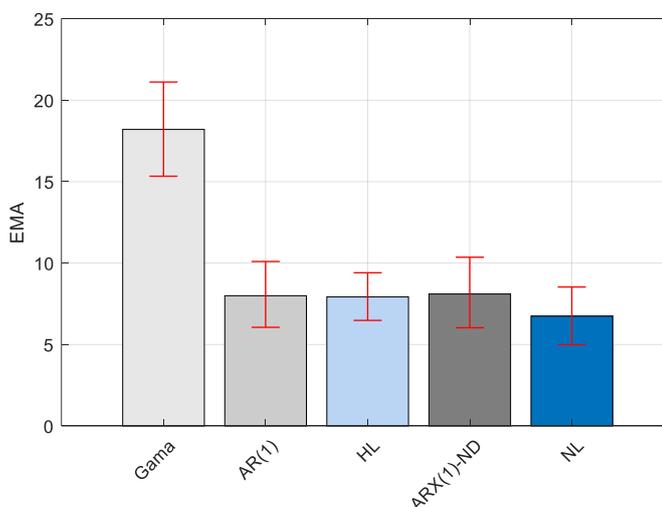


Figura C. 12 Erro Médio Absoluto (EMA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos em que $CRPSS \geq CRPSSm$ do modelo Gama.

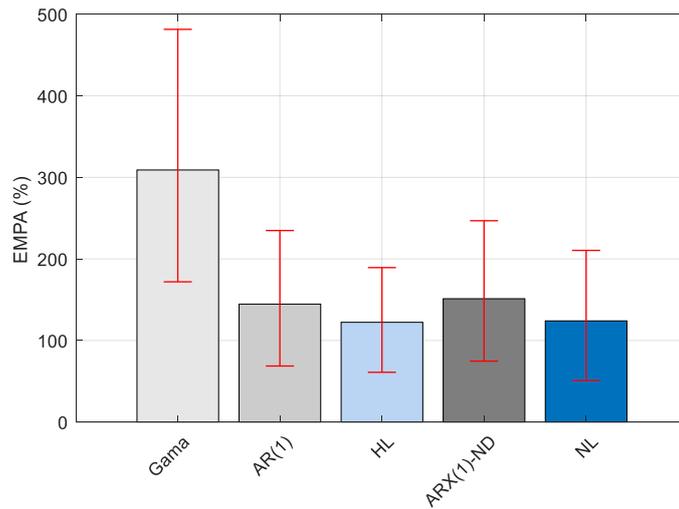


Figura C. 13 Erro Médio Percentual Absoluto (EMPA) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

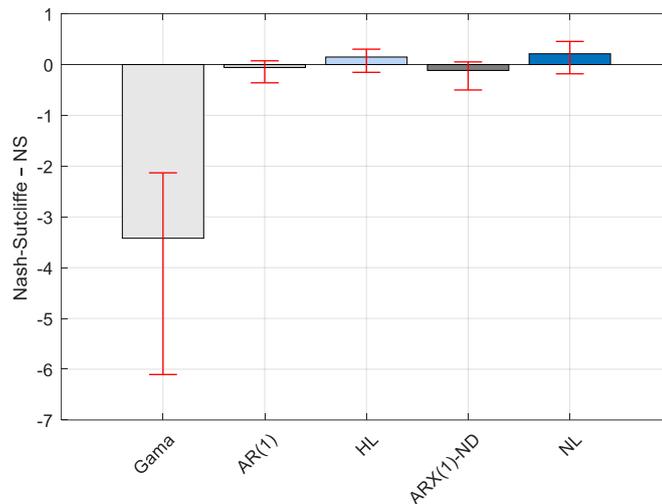


Figura C. 14 Coeficiente de Nash-Sutcliffe (NS) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

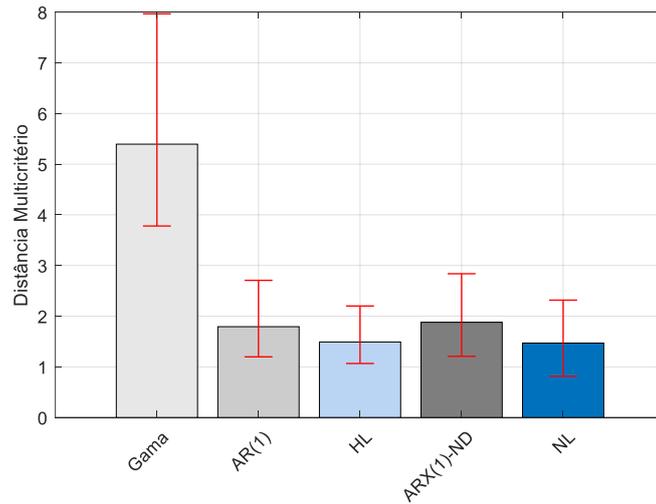


Figura C. 15 Distância Multicritério (DM) para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

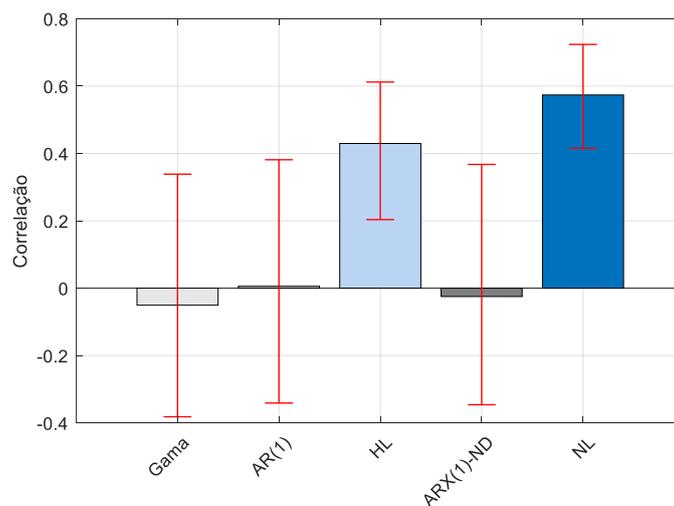


Figura C. 16 Correlação para os modelos sem informação climática, Gama, AR(1) e HL, e modelos com o indicador ND, ARX(1) e NL. As linhas vermelhas representam os intervalos de confiança de 95% calculados via Bootstrap. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

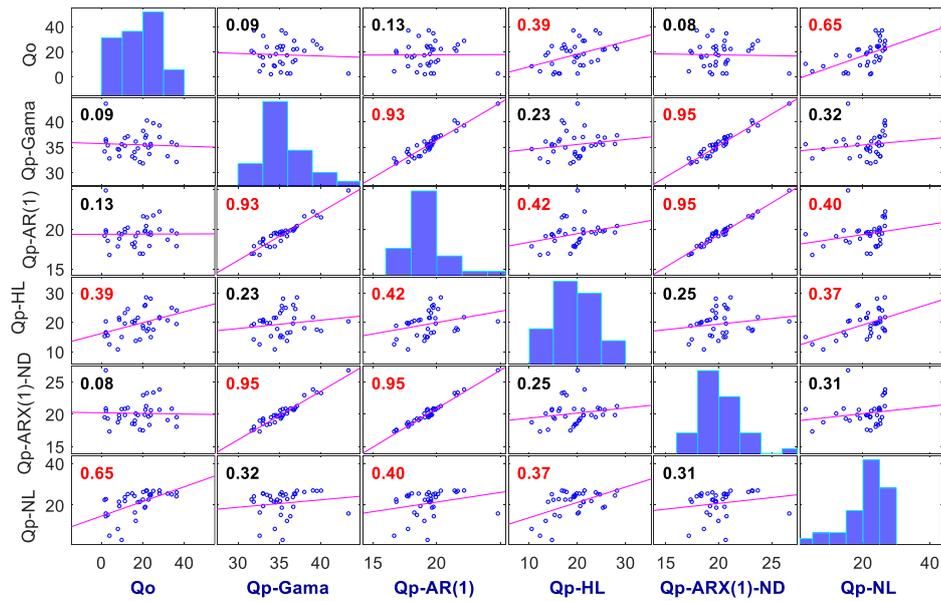


Figura C. 17 - Matriz de correlação tipo Spearman entre as vazões observadas e as médias previstas nos 60 anos do período de validação para pelos modelos Gama, HL e NL e modelos autoregressivos. Os valores em vermelho indicam uma correlação estatisticamente diferente de zero. Valores para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama. Qo: vazão observada; Qp: vazão prevista; Gama: modelo Gama; HL: HMM Log-Normal de 2 estados ocultos; NL: NHMM Log-Normal de 2 estados ocultos que inclui o indicador NINO3 de dezembro do ano anterior.

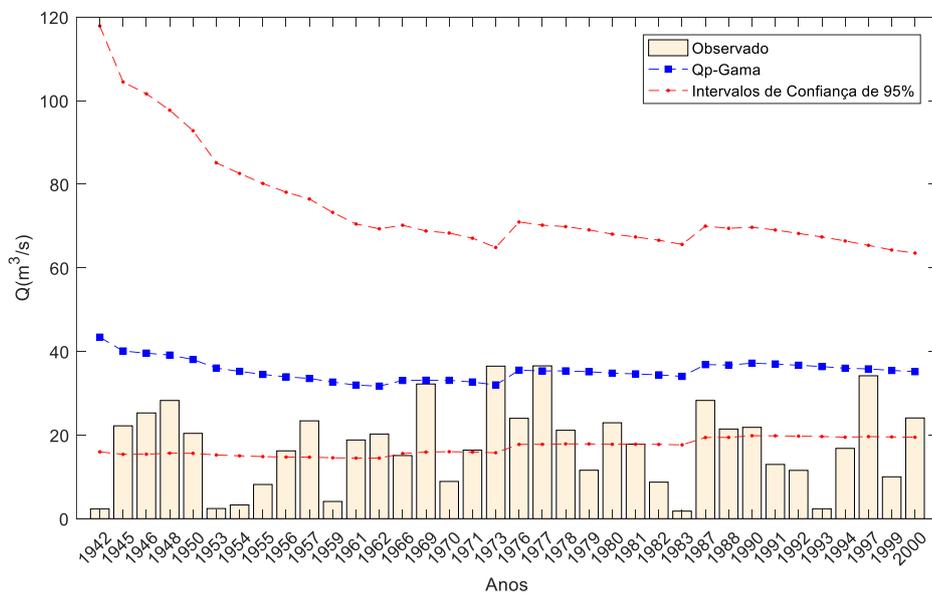


Figura C. 18 - Valores das observações e previsões produzidas pelo modelo Gama (Qp-Gama) para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

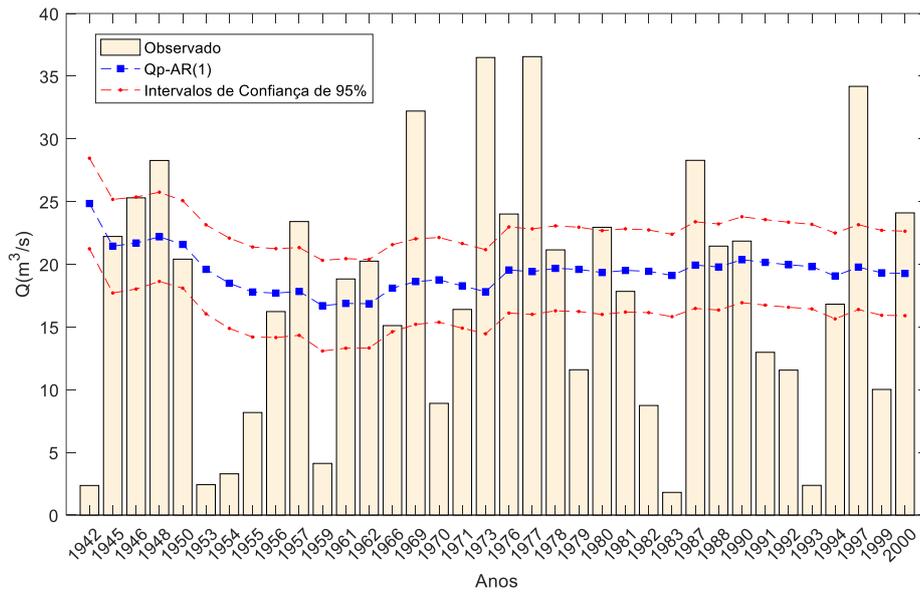


Figura C. 19 - Valores das observações e previsões produzidas pelo modelo AR(1) (Qp-AR(1)) para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

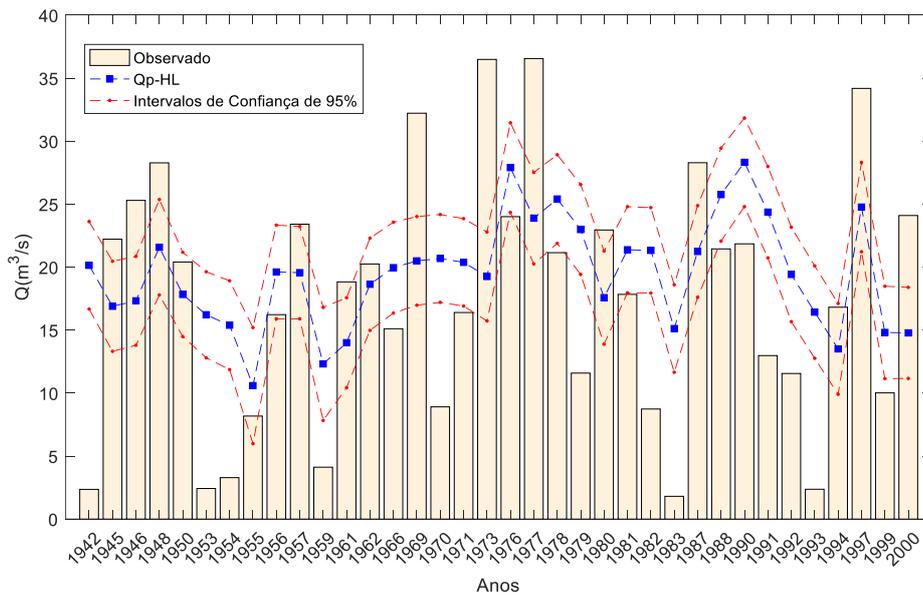


Figura C. 20 - Valores das observações e previsões produzidas pelo modelo HL (Qp-HL) para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

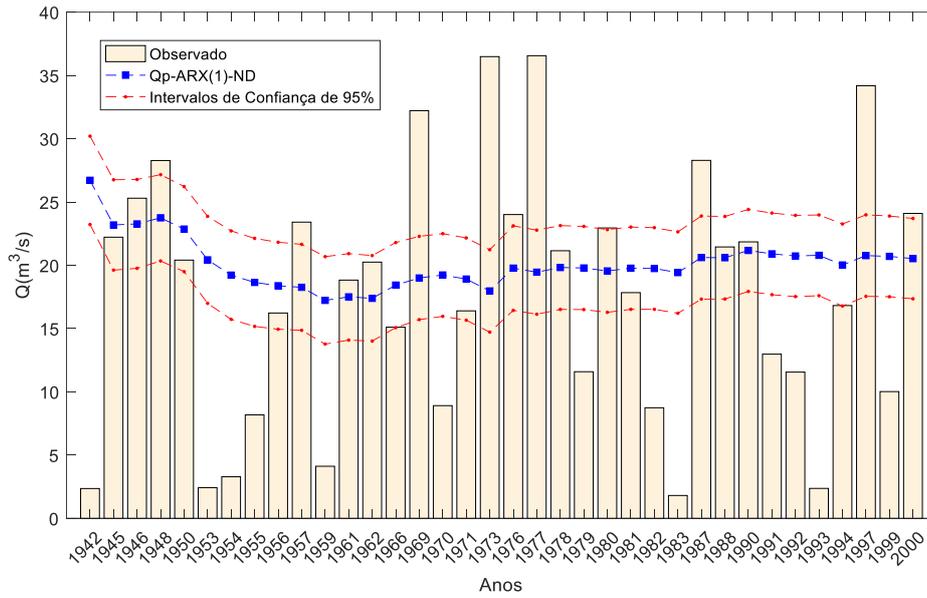


Figura C. 21 - Valores das observações e previsões produzidas pelo modelo ARX(1)-ND (Qp- ARX(1)-ND) para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

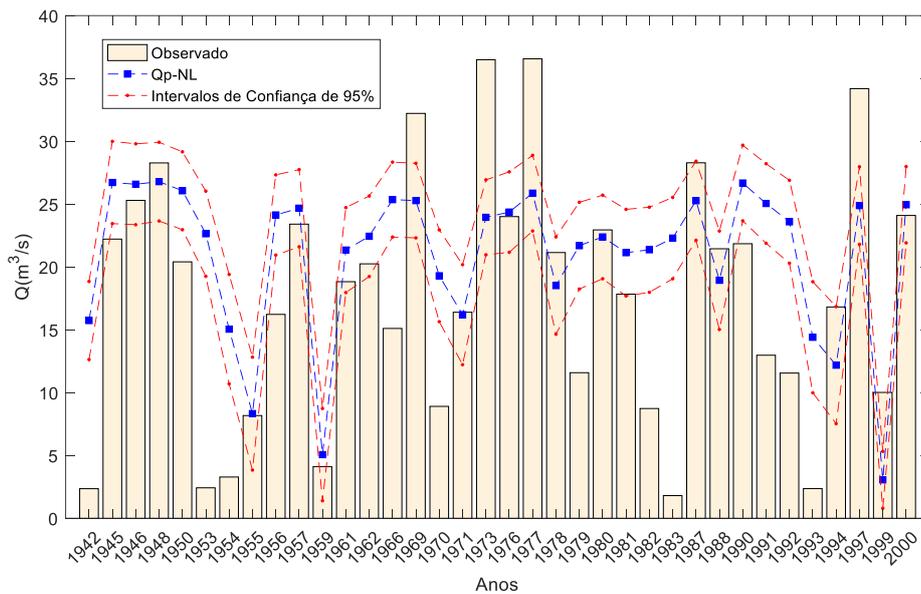


Figura C. 22 - Valores das observações e previsões produzidas pelo modelo NL (Qp- NL) para os anos em que $CRPSS \geq CRPSS_m$ do modelo Gama.

ANEXO A

ANEXO A – MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS PARA HMMs

A.1 - MAXIMIZAÇÃO DIRETA DA VEROSSIMILHANÇA

A expressão da verossimilhança é diferente no caso contínuo e discreto. Por exemplo, no caso discreto e estacionário equivale a:

$$\begin{aligned} L(\theta_1, \dots, \theta_m, \delta_1, \dots, \delta_m, \gamma_{11}, \dots, \gamma_{1m}, \dots, \gamma_{mm} | r_1, \dots, r_N) \\ = \prod_{t=1}^N \sum_{i=1}^m \sum_{j=1}^m \delta_i \gamma_{ij} p_i(r_t, \theta_i) \end{aligned} \quad (\text{A.1})$$

Em que θ é o vetor de parâmetros das distribuições dependentes. Zucchini e MacDonald (2009) mencionam que no caso contínuo, para algumas combinações de parâmetros contidos em θ a verossimilhança é ilimitada. Por exemplo, em alguns casos de HMMs com distribuições dependentes Normais, a verossimilhança pode alcançar valores arbitrariamente grandes. Portanto, a verossimilhança, contínua, deve ser calculada por intervalos que incluem os valores observados.

Esses autores mostram várias aplicações de HMMs com distribuições Normais, em que a verossimilhança é calculada de forma discreta e contínua. Os resultados obtidos em ambos os casos são similares, sendo que em alguns casos a forma discreta entrega melhores resultados, inclusive próximos aos obtidos com o algoritmo maximização da expectativa.

A verossimilhança discreta pode ser maximizada diretamente através de métodos de maximização numérica. Porém, é necessário realizar um escalonamento do vetor α_t para evitar problemas de sub fluxo. Isso se deve ao fato de que α_t é um produto de probabilidades e tende progressivamente a zero com o aumento de t . Similarmente, no caso contínuo existe super fluxo (Leroux e Puterman, 1992).

No caso discreto e estacionário, α_t pode ser escalonado da seguinte forma:

$$\phi_t = \frac{\alpha_t}{\omega_t} \quad (\text{A.2})$$

Em que $\omega_t = \sum_i \alpha_t(i) = \alpha_t \mathbf{1}'$ e se tem que:

$$\omega_0 = \alpha_0 \mathbf{1}' = \delta \mathbf{1}' = 1$$

$$\phi_0 = \delta$$

$$\omega_t \phi_t = \omega_{t-1} \phi_{t-1} \mathbf{B}_t$$

$$L_N = \alpha_N \mathbf{1}' = \omega_N (\phi_N \mathbf{1}') = \omega_N$$

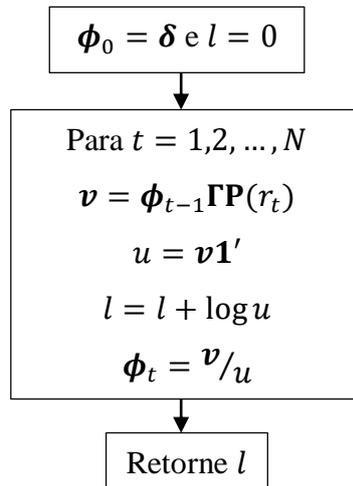
Em $\mathbf{B}_t = \mathbf{\Gamma P}(r_t)$ e $L_N = \omega_N = \prod_{t=1}^N (\omega_t / \omega_{t-1})$ e da equação 4.28 se tem que:

$$\omega_t = \omega_{t-1} (\phi_{t-1} \mathbf{B}_t \mathbf{1}')$$

Dessa forma:

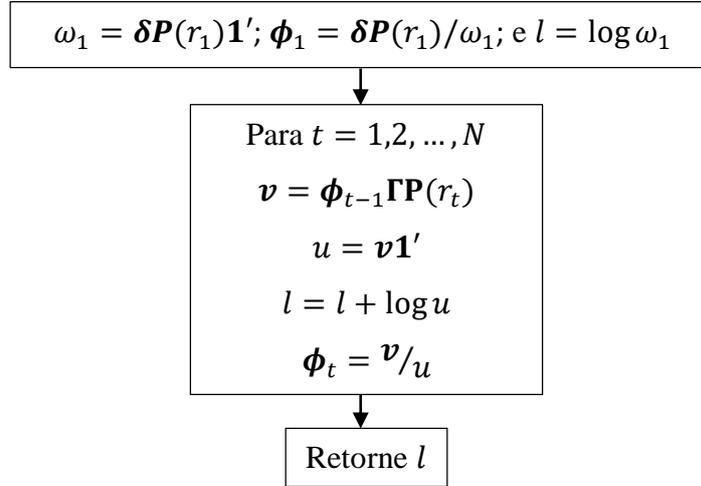
$$\log L_N = \sum_{t=1}^N \log(\omega_t / \omega_{t-1}) = \sum_{t=1}^N \log(\phi_{t-1} \mathbf{B}_t \mathbf{1}')$$

Em forma de algoritmo o $\log L_N$ pode ser calculado como:



Em que $\mathbf{\Gamma}$ e $\mathbf{P}(r_t)$ são matrizes de $m \times m$, v e ϕ_t são vetores de comprimento m , u é um escalar e l é um escalar no qual se acumula o $\log L_N$. O valor desejado será o valor final de

$\log L_N$. Esse procedimento evita sub fluxo na maioria dos casos. No caso não estacionário se deve realizar a seguinte modificação no algoritmo:



Vale a pena ressaltar que esse último algoritmo também aplica no caso estacionário.

Para maximizar $\log L_N$ existem alguns pacotes em *softwares* como *R* ou *Matlab* que permitem minimizar funções sem restrições. Para empregar esses pacotes é necessário realizar transformações dos parâmetros δ e $\mathbf{\Gamma}$ e θ para que não possuam restrições, já que: $\sum_{i=1}^m \delta_i = 1$ e $\delta_i > 0$; $\sum_{j=1}^m \gamma_{i,j} = 1$ para e $\gamma_{ji} > 0$; e θ possui restrições que dependem do tipo de distribuição dependente, por exemplo, para distribuições Normais $\theta = [\mu_i, \sigma_i]$ com $\mu_i > 0$ e $\sigma_i > 0$.

Dessa forma, algumas das transformações são:

$$t\delta_i = \log \left(\frac{\delta_i}{1 - \sum_{j=2}^m \delta_j} \right) \quad i = 1, \dots, m \quad (\text{A.3})$$

Para $\mathbf{\Gamma}$ se define a matriz $t\mathbf{\Gamma}$ com $m(m-1)$ elementos $t\gamma_{ij}$, $i \neq j$ sem restrições, tal que $t\gamma_{ij} \in \mathbb{R}$, como:

$$t\mathbf{\Gamma} = \begin{pmatrix} - & \dots & t\gamma_{1m} \\ \vdots & - & \vdots \\ t\gamma_{m1} & \dots & - \end{pmatrix}$$

Uma estratégia para formar $t\mathbf{\Gamma}$ é a seguinte:

$$\mathbf{d} = \begin{pmatrix} \gamma_{11} \\ \vdots \\ \gamma_{mm} \end{pmatrix} \quad (\text{A.4})$$

$$t\gamma_{ji} = \log(\gamma_{ji}/d_j) \text{ para } j, i = 1, 2, \dots, m \quad (\text{A.5})$$

Em que a matriz \mathbf{d} contém os m elementos da diagonal principal de $\mathbf{\Gamma}$.

No caso de distribuições Normais, pode-se aplicar:

$$t\mu_i = \log \mu_i \quad (\text{A.6})$$

$$t\sigma_i = \log \sigma_i \quad (\text{A.7})$$

E para regressar aos parâmetros originais se realizam as seguintes transformações:

$$\delta_i = \frac{\exp(t\delta_i)}{1 - \sum_{j=2}^m \exp(t\delta_j)}$$

$$\delta_1 = 1 - \sum_{j=2}^m \delta_j$$

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

$$p\gamma_{ij} = \exp(t\gamma_{ij}), \text{ para } i \neq j$$

$$ds_i = \sum_{j=1}^m p\gamma_{i,j} \text{ para } i = 1, 2, \dots, m$$

$$\gamma_{ij} = p\gamma_{ij}/ds_i$$

$$\mu_i = \exp(t\mu_i)$$

$$\sigma_i = \exp(t\sigma_i)$$

A.2 - ALGORITMO MAXIMIZAÇÃO DA EXPECTATIVA

O algoritmo Maximização da Expectativa –ME– é comumente utilizado para maximizar a verossimilhança quando se tem dados faltantes ou não observados. No caso dos HMMs, os dados faltantes são os dados das Cadeias de Markov. O algoritmo se divide em dois passos: Expectativa E e Maximização M.

No passo E se calcula o $\log L_N$ condicionado às observações, empregando um valor inicial dos parâmetros. No passo M se maximiza o $\log L_N$ calculado no passo E e os valores encontrados dos parâmetros são utilizados em um novo passo E, o procedimento se repete até atingir um critério de convergência. Geralmente, a convergência é alcançada quando a diferença entre os parâmetros de uma a outra iteração é bastante pequena, por exemplo, 1×10^{-6} (Zucchini e MacDonald, 2009).

Para HMMs, o algoritmo ME se denomina algoritmo Baum–Welch e utiliza a relação que existe entre α_t , β_t e a verossimilhança. Dessa forma, o algoritmo utiliza o logaritmo da verossimilhança completa (LVC) que é a verossimilhança dos parâmetros de interesse δ , Γ e θ , baseada nos dados observados e os dados não observados.

Para facilitar a implementação do algoritmo ME é necessário realizar as seguintes definições para as Cadeias de Markov:

$$u_j(t) = 1 \text{ se e somente se } s_t = j \text{ para } t = 1, 2, \dots, N \quad (\text{A.8})$$

$$v_{jk}(t) = 1 \text{ se e somente se } s_{t-1} = j \text{ e } s_t = k \quad (\text{A.9})$$

Dessa forma, a LVC é:

$$\log(P(r_1, \dots, r_N | s_1, \dots, s_N)) = \log\left(\delta_{s_1} \prod_{t=2}^N \gamma_{s_{t-1}, s_t} \prod_{t=1}^N p_{s_t}(r_t)\right) \quad (\text{A.10})$$

$$= \log \delta_{s_1} + \sum_{t=2}^N \log \gamma_{s_{t-1}, s_t} + \sum_{t=1}^N \log p_{s_t}(r_t) \quad (\text{A.11})$$

$$= \sum_{j=1}^m u_j(1) \log \delta_j + \sum_{j=1}^m \sum_{k=1}^m \left(\sum_{t=2}^N v_{jk}(t) \right) \log \gamma_{jk} + \sum_{j=1}^m \sum_{t=1}^N u_j(t) \log p_j(r_t) \quad (\text{A.12})$$

$$= \textit{termo 1} + \textit{termo 2} + \textit{termo 3}$$

Logo, o **passo E** é o seguinte:

$$\hat{u}_j(t) = \Pr(S_t = j | r_1, \dots, r_N) = \frac{\alpha_t(j) \beta_t(j)}{L_N} \quad (\text{A.13})$$

$$\hat{v}_{jk}(t) = \Pr(S_{t-1} = j, S_t = k | r_1, \dots, r_N) = \frac{\alpha_{t-1}(j) \gamma_{jk} p_k(r_t) \beta_t(k)}{L_N} \quad (\text{A.14})$$

$$L_N = \log \left(\sum_{i=1}^m \alpha_N(i) \right) = c + \log \left(\sum_{i=1}^m \exp(\log(\alpha_N(i)) - c) \right) \quad (\text{A.15})$$

$$c = \max(\log \alpha_N(i)) \quad (\text{A.16})$$

Dessa maneira, o LVC é formado com $\hat{u}_j(t)$ e $\hat{v}_{jk}(t)$. Para a resolução é necessário que α_t e β_t sejam escalonados. Por outra parte, o valor de c é escolhido de tal forma que evita as chances de sub fluxo na exponenciação.

No **passo M** é maximizado o LVC, o que praticamente consiste em realizar três maximizações separadamente, já que o termo 1 depende de δ , o termo 2 de Γ e o termo 3 de θ . A solução para δ e Γ é a seguinte:

$$\delta_j = \frac{\hat{u}_j(1)}{\sum_{j=1}^m \hat{u}_j(1)} = \hat{u}_j(1) \quad (\text{A.17})$$

$$\gamma_{jk} = \frac{f_{jk}}{\sum_{k=1}^m f_{jk}} \quad (\text{A.18})$$

$$f_{jk} = \sum_{t=2}^N \hat{v}_{jk}(t) \quad (\text{A.19})$$

A maximização do termo 3 é mais complicada e depende do tipo das FDPs dependentes dos estados ocultos. Em alguns casos, por exemplo, quando as FDPs são distribuições Gama é necessário realizar uma maximização numérica.

No caso de FDPs Normais a maximização do termo 3 é a seguinte:

$$\hat{\mu}_j = \frac{\sum_{t=1}^N \hat{u}_j(t) r_t}{\sum_{t=1}^N \hat{u}_j(t)} \quad (\text{A.20})$$

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^N \hat{u}_j(t) (r_t - \hat{\mu}_j)^2}{\sum_{t=1}^N \hat{u}_j(t)} \quad (\text{A.21})$$