



COMPARAÇÃO ENTRE OS MODELOS SFR E SMAP PARA PREVISÃO DE
VAZÃO DE CURTO PRAZO EM HIDRELÉTRICAS DO SISTEMA INTERLIGADO
NACIONAL

Bruno Moreno Rodrigo de Freitas

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Civil.

Orientador: Alexandre Gonçalves Evsukoff

Rio de Janeiro

Abril de 2018

COMPARAÇÃO ENTRE OS MODELOS SFR E SMAP PARA PREVISÃO DE
VAZÃO DE CURTO PRAZO EM HIDRELÉTRICAS DO SISTEMA INTERLIGADO
NACIONAL

Bruno Moreno Rodrigo de Freitas

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
CIÊNCIAS EM ENGENHARIA CIVIL.

Examinada por:

Prof. Alexandre Gonçalves Evsukoff, Ph.D.

Profª. Beatriz de Souza Leite Pires de Lima, D.Sc.

Prof. Marcio Cataldi, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
ABRIL DE 2018

Freitas, Bruno Moreno Rodrigo de

Comparação entre os Modelos SFR e SMAP para a Previsão de Vazão de Curto Prazo em Hidrelétricas do Sistema Interligado Nacional / Bruno Moreno Rodrigo de Freitas. – Rio de Janeiro: UFRJ/COPPE, 2018.

XIV, 90 p.: il.; 29,7 cm.

Orientador: Alexandre Gonçalves Evsukoff

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Civil, 2018.

Referências Bibliográficas: p. 86-88.

1. Aprendizado de Máquina. 2. Modelagem de Dados. 3. Modelagem Hidrológica. 4. Sistemas Fuzzy I. Evsukoff, Alexandre Gonçalves. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

“Embora ninguém possa voltar atrás e fazer um novo começo, qualquer um pode
começar agora e fazer um novo fim.”

Francisco Cândido Xavier

AGRADECIMENTOS

Agradeço, primeiramente, à vida que me foi dada para que eu possa traçar um caminho na tentativa de melhorar as condições de vida das gerações futuras.

À minha mãe que, com todo seu amor e carinho, apoia as minhas decisões e se torna cúmplice, sendo responsável significativamente das minhas conquistas.

À minha família em geral, que sempre me apoiou de alguma forma.

Aos meus amigos do Tijolino, Aldebaran, em Vila Isabel, Zona Norte do Rio de Janeiro, onde cresci, mencionando alguns: Felipe Carvalho (Fil), Marcos Valois (Kanga) Jocimar Ferreira (Negão), Fernando Dana (Pesado), Eric Miyahira (Samu), Diego Ramôa (Wando) entre outros.

Aos meus amigos irmãos: Vítor Mello (Frida), Guilherme Martins (Catador), Vítor Schwenck (Schw), Vítor Tavares (VT), Gilberto Jr. Mattos (Gil).

Aos amigos da UFF, os quais ainda mantenho fortes laços, mencionando alguns: Felipe Amaral, Fabio Coelho (Charlinho), Felipe Rosa, entre outros.

Aos professores da engenharia de recursos hídricos e meio ambiente da UFF, os quais me formaram e me deram bases sólidas para que eu pudesse começar a minha carreira, mencionando alguns: Professora Chou Sin Hwa, Professora Mônica da Hora, Professor Ricardo Félix, Professor Robson Saramago, Professor Antônio Henrique, entre outros.

Em especial, como professor da UFF, gostaria de agradecer ao professor Marcio Cataldi da disciplina de Modelagem de Sistemas Ambientais, a qual foi muito importante para o desenvolvimento do presente trabalho.

Aos colegas do PEC/COPPE, com quem eu tive o imenso prazer em trabalhar e discutir assuntos acadêmicos, mencionando alguns: José Guilherme, Daniel Lopes, Gustavo Ortiz, Gloria Nakamura, Manuel Vargas, entre outros.

Aos professores do PEC/COPPE, que me ensinaram bastante ao longo dessa trajetória, principalmente aos pertencentes da área de concentração de Sistemas Computacionais, mencionando alguns: Beatriz Lima e Solange Guimarães.

Ao meu orientador Alexandre Evsukoff que me ensinou e auxiliou em todo este trabalho, o qual tive o grande prazer em realizar.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

COMPARAÇÃO ENTRE OS MODELOS SFR E SMAP PARA PREVISÃO DE
VAZÃO DE CURTO PRAZO EM HIDRELÉTRICAS DO SISTEMA INTERLIGADO
NACIONAL

Bruno Moreno Rodrigo de Freitas

Abril/2018

Orientadores: Alexandre Gonçalves Evsukoff

Programa: Engenharia Civil

O Brasil é um dos países que apresenta grande vocação para a utilização da hidroeletricidade. A energia hidráulica apresenta a maior participação em relação à capacidade instalada em comparação com outras fontes. O ONS necessita obter informações do aporte de vazão que ocorrerá dez dias à frente para realizar operação do SIN de maneira que haja a otimização dos recursos energéticos do país, bem como o cálculo semanal do PLD, o preço do mercado spot de eletricidade. Para isso, modelos de previsão de vazão foram desenvolvidos com o intuito de auxiliar na tomada de decisão do ONS. Dentre os modelos utilizados na previsão de curto prazo pelo ONS estão o SFR e o SMAP. O modelo SFR é utilizado somente na bacia do rio Iguaçu, enquanto que o SMAP é utilizado nas bacias dos rios Grande, Paranapanema e Paranaíba. No presente trabalho, o desempenho dos dois modelos foram analisadas para a previsão de vazão dez dias à frente para algumas hidrelétricas conectadas ao SIN, pertencentes a diferentes bacias hidrográficas. O desempenho dos modelos dependeu da métrica de erro utilizada para a análise. Para determinadas hidrelétricas, o SFR apresentou melhor resultados e em outras o SMAP. Tal resultado nos leva a concluir que para cada bacia hidrográfica existirá um determinado modelo que terá um melhor desempenho.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

COMPARISON BETWEEN THE MODELS SFR AND SMAP FOR SHORT-TERM
FORECAST IN HYDROELECTRICS OF THE NATIONAL INTERCONNECTED
SYSTEM

Bruno Moreno Rodrigo de Freitas

April/2018

Advisor: Alexandre Gonçalves Evsukoff

Department: Civil Engineering

Brazil is one of the countries that has a great vocation for the use of hydroelectricity. Hydropower has the largest share of installed capacity compared to other sources. The ONS needs to obtain information on the flow contribution that will occur ten days ahead to carry out the operation of the SIN in a way that optimizes the country's energy resources, as well as the weekly calculation of the PLD, the spot market price of electricity. For this, flow forecasting models were developed with the purpose of assisting in the ONS's decision making. Among the models used in the short-term forecast by ONS are the SFR and SMAP. The SFR model is used only in the Iguaçú river basin, while the SMAP is used in the Grande, Paranapanema and Paranaíba river basins. In the present work, the performance of the two models were analyzed for the ten days ahead flow forecast for some hydroelectric connected to the SIN, belonging to different hydrographic basins. The performance of the models depended on the error metric used for the analysis. For some hydroelectric plants, the SFR presented better results and in others the SMAP. This result leads us to conclude that for each river basin there will be a certain model that will perform better.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiv
1. Introdução.....	1
1.1. Justificativa do tema	5
1.2. Objetivo	5
1.3. Estrutura da dissertação	5
2. Revisão bibliográfica.....	6
2.1. Modelagem de sistemas dinâmicos.....	6
2.2. Modelagem de sistemas hidrológicos chuva-vazão.....	13
2.3. Modelo <i>Soil Moisture Accounting Procedure</i> - SMAP.....	15
2.4.1 Calibração do modelo SMAP	20
2.4. Modelo Sistema Fuzzy Recorrente	23
2.5.1 Formulação matemática de Sistemas Fuzzy Recorrente	26
2.5.2 Identificação de SFR através de Algoritmos Genéticos	30
3. Materiais e métodos.....	33
3.1. Apresentação das regiões hidrográficas brasileiras	33
3.2. Apresentação das UHE estudadas.....	38
3.3. Pré-processamento dos dados	39
3.4. Análise exploratória dos dados	43
4. Resultados e discussões	54
4.1. Métricas de erro para o desempenho dos modelos	54

4.2.	Configuração experimental.....	55
4.2.1.	Treinamento - SFR.....	55
4.2.2.	Calibração - SMAP	57
4.2.3.	Teste dos modelos	58
4.3.	Avaliação do desempenho dos modelos	61
4.3.1.	Aplicação no conjunto de testes completo	61
4.3.2.	Simulação da operação real de previsão de vazão para 10 dias à frente.....	67
5.	Conclusões e trabalhos futuros	82
6.	Referências bibliográficas	86

LISTA DE FIGURAS

Figura 1.1: Capacidade instalada no SIN em dezembro de 2016.....	1
Figura 1.2: Participação de geração de energia elétrica por fonte.....	2
Figura 1.3: Representação do SIN e seus subsistemas	3
Figura 1.4: Modelos de previsão de vazão utilizados pelo ONS.....	4
Figura 2.1: Princípio da proporcionalidade	9
Figura 2.2: Princípio da superposição de efeitos.....	9
Figura 2.3: Classificação de modelos	11
Figura 2.4: Esquema do modelo SMAP na versão com quatro reservatórios.....	17
Figura 2.5: Funções de pertinência triangulares.....	27
Figura 2.6: Estrutura típica de um AG	31
Figura 3.1: Divisão hidrográfica do Brasil	37
Figura 3.2: Vazões mensais das UHE estudadas entre 1931-2015 Fonte: Elaboração própria a partir de dados do ONS (2017)	45
Figura 3.3: Bacias hidrográficas, dispersão das estações de chuva e localização das UHE	46
Figura 3.4: Precipitação transformada vs. vazão das UHE	48
Figura 3.5: Gráficos de auxílio à AED para a UHE Furnas	51
Figura 3.6: Gráficos de auxílio à AED para a UHE Machadinho	51
Figura 3.7: Gráficos de auxílio à AED para a UHE Mascarenhas	52
Figura 3.8: Gráficos de auxílio à AED para a UHE Três Marias.....	52
Figura 3.9: Distribuição do logaritmo natural das vazões das UHE analisadas.....	53
Figura 4.1: Matriz de entrada dos modelos	59
Figura 4.2: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Furnas	63

Figura 4.3: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Furnas	63
Figura 4.4: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Machadinho	64
Figura 4.5: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Machadinho	64
Figura 4.6: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Mascarenhas	65
Figura 4.7: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Mascarenhas	65
Figura 4.8: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Três Marias	66
Figura 4.9: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Três Marias	66
Figura 4.10: Resultados das Métricas de Erros para a UHE Furnas	68
Figura 4.11: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SFR para 1 dia à frente	69
Figura 4.12: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SMAP para 1 dia à frente	69
Figura 4.13: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SFR para 10 dias à frente.....	70
Figura 4.14: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SMAP para 10 dias à frente.....	70
Figura 4.15: Resultados das Métricas de Erros para a UHE Machadinho	72

Figura 4.16: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SFR para 1 dia à frente	72
Figura 4.17: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SMAP para 1 dia à frente	73
Figura 4.18: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SFR para 10 dias à frente.....	73
Figura 4.19: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SMAP para 10 dias à frente	74
Figura 4.20: Resultados das Métricas de Erros para a UHE Mascarenhas.....	75
Figura 4.21: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SFR para 1 dia à frente	76
Figura 4.22: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SMAP para 1 dia à frente	76
Figura 4.23: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SFR para 10 dias à frente.....	77
Figura 4.24: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SMAP para 10 dias à frente	77
Figura 4.25: Resultados das Métricas de Erros para a UHE Três Marias	79
Figura 4.26: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SFR para 1 dia à frente	80
Figura 4.27: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SMAP para 1 dia à frente	80
Figura 4.28: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SFR para 10 dias à frente.....	81

Figura 4.29: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SMAP para 10 dias à frente 81

LISTA DE TABELAS

Tabela 2.1: Faixas de variação dos parâmetros do modelo SMAP	21
Tabela 3.1: Características das UHE estudadas.....	39
Tabela 3.2: Número de estações para cada UHE.....	40
Tabela 3.3: Ano de operação verificado dentre as estações	41
Tabela 3.4: Período de análise e IVA dos conjuntos de dados de 2, 4 e 6 anos.....	42
Tabela 3.5: Período de análise e IVA dos conjuntos de dados de 8 e 10 anos.....	42
Tabela 3.6: Número de postos com IVA menor que 5%	42
Tabela 3.7: Estatísticas descritivas da variável de entrada precipitação transformada ..	49
Tabela 3.8: Estatísticas descritivas da variável de saída vazão	49
Tabela 4.1: Estrutura do SFR para cada bacia das UHE	57
Tabela 4.2: Resultado da FO para treinamento do SFR	57
Tabela 4.3: Resultado da calibração do modelo SMAP	58
Tabela 4.4: Inicialização do modelo SMAP para o conjunto de treinamento	58
Tabela 4.5: Resultado da FO para calibração do SMAP	58
Tabela 4.6: Resultados do SFR para as métricas de erro na aplicação do conjunto de testes completo	62
Tabela 4.7: Resultados do SMAP para as métricas de erro na aplicação do conjunto de testes completo	62
Tabela 4.8: Queda de desempenho dos modelos.....	62

1. INTRODUÇÃO

O Brasil é um dos países que apresenta grande vocação para a utilização da hidroeletricidade. O histórico do setor elétrico brasileiro, que iniciou nos primeiros anos do século 20, foi bastante influenciado pelo uso hidráulico para geração de energia elétrica. De fato, essa é a tecnologia que apresenta diversas vantagens comparativas em relação a outras, como a possibilidade de armazenamento de energia, baixo custo de operação, flexibilidade na entrada em operação das usinas, despachabilidade entre outras.

A matriz elétrica brasileira vem sofrendo grandes mudanças que iniciaram desde quando foi lançado o novo modelo do setor elétrico em 2004. A participação hidráulica na matriz elétrica vem sendo reduzida cada vez mais, muito pela dificuldade de implantar novas usinas hidrelétricas conectadas no Sistema Interligado Nacional – SIN, devido aos impactos socioambientais causados pela fonte. No entanto, a hidroeletricidade ainda apresenta a maior participação em relação à capacidade instalada em comparação com outras fontes, cerca de 60% segundo a Figura 1.1:

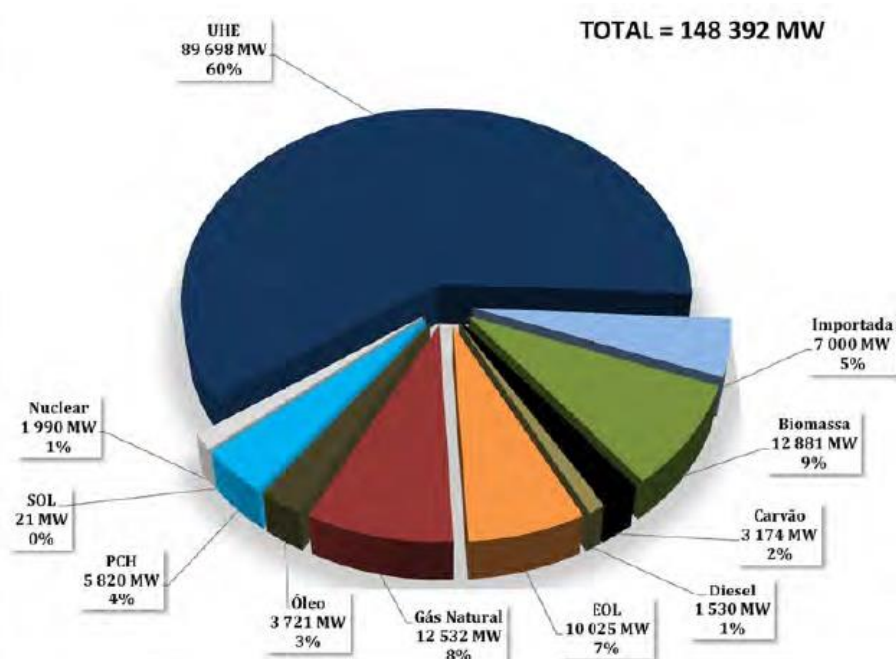


Figura 1.1: Capacidade instalada no SIN em dezembro de 2016

Fonte: (EPE, 2017)

Por outro lado, se for analisado o percentual anual de geração por cada fonte, a hidráulica apresenta ainda, um percentual bem superior que a participação na capacidade

instalada. Isso ocorre pelo fato de muitas usinas hidrelétricas apresentarem reservatórios de regularização que permite o armazenamento de energia para períodos de escassez hídrica ao longo do ano, fazendo com que o fator de capacidade da fonte fique mais próximo à unidade. Segundo o Operador Nacional do Sistema Elétrico - ONS (2017), a fonte hidráulica gerou 410.237 GWh de com participação de 76% do total de energia elétrica gerado em 2016 no Brasil, como demonstra a Figura 1.2:

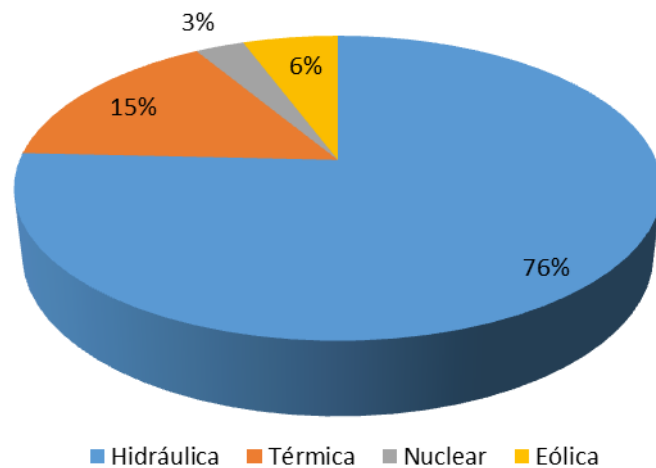


Figura 1.2: Participação de geração de energia elétrica por fonte

Fonte: Elaboração própria a partir de dados do ONS (2017)

Desde a reforma do setor elétrico brasileiro (MENDONÇA; DAHL, 1999), o ONS é autorizado a realizar atividades de coordenação e controle para a operação de geração e transmissão de energia elétrica no SIN. Este é composto por quatro subsistemas interconectados, S (Sul), Sudeste/ Centro-Oeste (SE/CO), Norte (N) e Nordeste (NE). Existem outros sistemas isolados que não fazem parte do SIN e que são situados em regiões remotas, principalmente na região norte do país. O subsistema mais importante do SIN é o SE/CO, o qual contabiliza aproximadamente 60% da demanda do país.



Figura 1.3: Representação do SIN e seus subsistemas

Fonte: ONS

A missão do ONS é garantir a continuidade, qualidade e economia do suprimento de energia elétrica para todos usuários do SIN. Para cumprir tal missão, o ONS deve planejar, programar e monitorar a operação das principais plantas de geração conectadas ao SIN. O planejamento operacional é realizado em três horizontes temporais:

- Médio-prazo com horizonte de 5 anos à frente e discretização mensal;
- Curto-prazo com horizonte de 12 meses à frente e discretização semanal/mensal; e
- Curtíssimo-prazo com horizonte de 1 semana e discretização de 30 minutos.

O planejamento de curto prazo produz um relatório técnico mensal chamado Programa Mensal de Operação – PMO com o objetivo de despacho para todas as plantas no SIN. Esse planejamento é revisado semanalmente em resposta às atualizações relacionadas a demanda esperada e previsão meteorológica.

A energia hidráulica ainda se mostra de grande importância no país. Toda a operação do SIN é feita com base na otimização do uso dos recursos hídricos nos reservatórios das usinas hidrelétricas dados, basicamente, a expectativa de crescimento da carga e o aporte das afluições nos reservatórios. Para o planejamento da operação do SIN é de grande interesse a previsão dos volumes afluentes, realizada constantemente e independentemente das condições hidrológicas, para melhor regular os níveis dos reservatórios de forma a maximizar a produção de energia. Portanto, existe um grande

empenho no aprimoramento dos modelos de previsão das afluições de vazões aos diferentes aproveitamentos hidrelétricos do setor elétrico, que servirá como entrada nos diferentes modelos do planejamento da operação, pois um bom sistema de previsão leva a uma maior eficácia operacional, seja sob o ponto de vista econômico, seja da segurança, dos benefícios sociais ou da preservação ambiental (CARVALHO, 2001).

Na programação diária da operação do SIN, concretizada centralizadamente pelo ONS, é necessário realizar a previsão de afluições nas usinas hidrelétricas a nível diário para os próximos 7 dias, de forma a otimizar os recursos energéticos do país. Tal previsão é dada como modelo de entrada para o cálculo do Preço de Liquidação das Diferenças – PLD, que é o preço do mercado SPOT de energia elétrica existente no Brasil e que serve como índice para as diversas transações financeiras, tanto no ambiente de contratação de energia regulado quanto no livre. Assim, quanto mais preciso forem os modelos de previsão de vazão utilizados na operação eletro-energética do SIN, maior aderência à realidade terá o PLD, reduzindo os riscos sistêmicos.

Hoje, diversos modelos são utilizados na previsão de afluições diárias no SIN. Como cada bacia hidrográfica apresenta suas particularidades, cada abordagem de modelagem chuva-vazão apresenta desempenhos diferenciados dependendo da bacia estudada. A fim de aumentar a acurácia na modelagem, o ONS utiliza diversos modelos para a operação do SIN, cada qual com seus desafios. A Figura 1.4 apresenta os modelos utilizados para cada bacia. Entre os modelos, é possível encontrar dois: o Sistema Fuzzy Recorrente – SFR e o Soil Moisture Accounting Procedure – SMAP, os quais serão alvo da presente dissertação.

Usina / Trecho / Bacia	Modelo
Incremental T.Marias-Sobradinho	NEUROSF/CPINS
Uruguai (B.Grande, C.Novos e Itá)	MPCV
Incremental à UHE Itaipu	SMAPMEL
Iguaçu	FUZZY
Incremental Itumbiara –S.Simão	MGB-IPH
Grande	SMAP
Paranaíba (até Itumbiara)	SMAP
Três Marias	NEURO3M
Parapanema	SMAP
Demais bacias	PREVIVAZ

Figura 1.4: Modelos de previsão de vazão utilizados pelo ONS

Fonte: (ONS, 2016)

1.1. JUSTIFICATIVA DO TEMA

Como mostra a Figura 1.4, os modelos alvo da presente dissertação, o SFR (referenciado como Fuzzy) e o SMAP, também já são utilizados na programação da operação eletroenergética do SIN. O SFR é somente utilizado na previsão diária de vazão da bacia do rio Iguaçu localizada no Sul do Brasil. A bacia do rio Iguaçu, apresenta comportamento hidrológico complexo, além de possuir um grande parque gerador instalado, fundamental, não só no suprimento energético do subsistema sul, mas também na definição dos intercâmbios de energia com o subsistema SE/CO. O SMAP, por sua vez, também é utilizado na operação do SIN nas bacias do rio Grande, Paranapanema e Paranaíba. É necessário um estudo sobre o desempenho desses modelos em outras bacias.

1.2. OBJETIVO

O objetivo principal da presente pesquisa é comparar o desempenho dos modelos SFR e SMAP para a utilização na simulação de previsão de vazão diária 10 dias à frente em hidrelétricas conectadas ao SIN localizadas em diferentes bacias hidrográficas brasileiras.

1.3. ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está dividida em cinco capítulos. No capítulo dois foi feita uma revisão bibliográfica sobre modelagem chuva-vazão, descrevendo também os modelos utilizados no presente trabalho. O capítulo três apresenta o estudo de caso, bem como a análise exploratória dos dados utilizados na modelagem. A análise dos resultados é apresentada no capítulo quatro e a conclusão e trabalhos futuros são discutidos no capítulo cinco.

2. REVISÃO BIBLIOGRÁFICA

2.1. MODELAGEM DE SISTEMAS DINÂMICOS

O estudo de sistemas dinâmicos envolve a modelagem matemática, a análise e a simulação tanto de sistemas físicos de interesse da engenharia, tais como sistemas mecânicos, pneumáticos, elétricos, térmicos e hidráulicos, como pode ser aplicada a outros tipos de sistemas, tais como sistemas biológicos, econômicos, ecológicos, de informação etc. Para o entendimento sobre modelagem de sistemas dinâmicos é necessário definir alguns conceitos primordiais.

Definição de Sistema

A definição do termo sistema varia de autor para autor e com a área de estudo. O artigo de (BACKLUND, 2000), traz uma discussão sobre a definição de sistema, avaliando a contribuição de diversos autores. Em LANGERFORS (1995) *apud* (BACKLUND, 2000), um sistema é um conjunto de entidades com relações entre elas (tradução livre). MILLER (1995) *apud* (BACKLUND, 2000), apresenta uma definição similar, declarando que um sistema é um conjunto de unidades que interagem com relacionamentos entre elas (tradução livre). As definições supracitadas, ainda são bastante abrangentes, porém já fazem menção à interação entre partes diferentes.

Para o presente trabalho, uma definição mais relacionada à teoria de sistemas e controle em engenharia é necessária. MELLODGE (2016) apresenta a definição que um sistema envolve um grupo de partes relacionadas funcionando juntas, sendo uma entidade que apresentam entradas e saídas. Ou seja, um sistema recebe a entrada e produz a saída baseada em seu estado inicial e na própria entrada. Tal abordagem é mais direta para o que será realizado na dissertação.

Definição de Modelagem

Muitos sistemas apresentam níveis de complexidade tais que os seres humanos não conseguem entender completamente em detalhe o seu processo. Abstrações, então, são necessárias para se tentar compreender e/ou controlar alguns aspectos de seus comportamentos. Modelos são, dessa maneira, a abstração de sistemas complexos que descrevem os componentes básicos e mais importantes, com o intuito de simplificá-los

para o entendimento de suas relações. Como no presente trabalho será dissertado sobre modelos matemáticos, abordaremos somente esses tipos de modelos.

Modelos matemáticos aplicados ao estudo de sistemas podem ser classificados de diversas formas. Se um modelo conseguir transparecer os processos existentes em um sistema através da teoria física, então esse modelo pode ser classificado como modelo teórico, ou de base física, e todo o processo é entendido apenas por suas equações matemáticas. Esses modelos também são conhecidos por *knowledge-driven models* (tradução livre, modelos orientados por conhecimento). Outra nomenclatura bastante utilizada na literatura é modelo caixa-branca (*white-box models*), por fazer alusão ao entendimento claro pela matemática da física do processo (XU, 2002).

Por outro lado, existem modelos matemáticos aplicados a sistemas que apresentam alto grau de acurácia, porém não se pode interpretar o processo fisicamente. Tais modelos possuem parâmetros que remetem pouca ou nenhuma significância física e que podem ser estimados apenas pelas medidas concomitantes das variáveis de entrada e saída, sendo chamados de modelos empíricos. Tais modelos também são conhecidos por *data-driven models* (tradução livre, modelos orientados por dados). Estes também são nomeados modelos caixa-preta (*black-box models*) pelo fato de não transparecer os processos do sistema (XU, 2002).

Em uma terceira vertente, há na literatura de modelagem de sistemas os chamados modelos conceituais. Estes são o intermediário entre os modelos teóricos e empíricos. Consideram algumas leis físicas, porém de uma forma muito simplificada, e possuem parâmetros que não apresentam qualquer significado proveniente da teoria física. Tais modelos podem ser chamados, seguindo o mesmo raciocínio da descrição dos tipos de modelos anteriores, de modelo caixa-cinza (*grey-box models*), por apresentar alguma intepretabilidade física, porém de forma incompleta (XU, 2002).

Definição do Termo Dinâmico

A definição do termo dinâmico se refere a um fenômeno que produz padrões de variação no tempo, como descrito por (LUENBERGER, 1979). A representação matemática para um sistema que varia no tempo é através de equações diferenciais ou a diferenças. A diferença entre uma ou outra é o fato de o comportamento do sistema ocorrer no tempo contínuo, para equações diferenciais (Equação 2.1), ou discreto, para equações a diferenças (Equação 2.2).

$$\frac{d^n}{dt^n} y(t) = a_1 \frac{d^{n-1}}{dt^{n-1}} y(t) + \dots + a_n y(t) + \frac{d^n}{dt^n} u(t) + (b_1 - a_1) \frac{d^{n-1}}{dt^{n-1}} u(t) + \dots + (b_n - a_n) u(t) \quad (2.1)$$

$$y(n+k) = a_1 y(n) + \dots + a_n y(n) + u(n+k) + (b_1 - a_1) u(n+k-1) + \dots + (b_n - a_n) u(n) \quad (2.2)$$

Onde u e y são, respectivamente, entrada e saída, e a_n e b_n são parâmetros do sistema.

Modelos de Sistemas Lineares e Não-Lineares

Outra classificação sobre um sistema é se ele apresenta características lineares ou não-lineares. Um sistema linear, de forma simplificada, apresenta o princípio da superposição de efeitos. Isto implica que, em qualquer sistema multivariável linear, pode-se analisar individualmente o efeito de cada variável de entrada na saída e depois sobrepor seus efeitos. Se essa regra não for obedecida, o sistema é dado como não-linear e a complexidade de análise aumenta. São duas regras:

- Proporcionalidade (Figura 2.1): a saída *constante* $\times y$ corresponde a entrada constante $\times u$;
- Superposição de efeitos (Figura 2.2): a saída $y_1 + y_2$ corresponde a entrada $u_1 + u_2$.

Classificação de Modelos de Sistemas a partir do Número de Entradas e Saídas

Um sistema pode apresentar únicas, ou múltiplas, entradas e saídas. Em engenharia de sistemas, os sistemas são classificados em relação ao número de entradas e saídas para melhor descrição do problema alvo através das seguintes abreviações:

- SISO – Single Input Single Output, para sistemas com únicas entradas e saídas;
- MIMO – Multiple Input Multiple Output, para sistemas com múltiplas entradas e saídas;
- MISO – Multiple Input Single Output, para sistemas de múltiplas entradas e única saída;
- SIMO – Single Input Multiple Output, para sistemas com única entrada e múltiplas saídas.

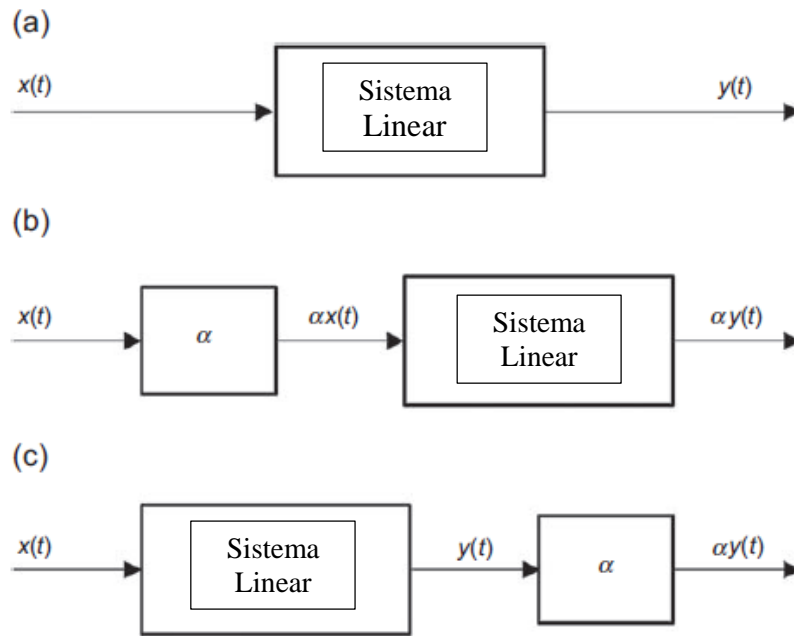


Figura 2.1: Princípio da proporcionalidade

Fonte: Adaptado de (MELLODGE, 2016)

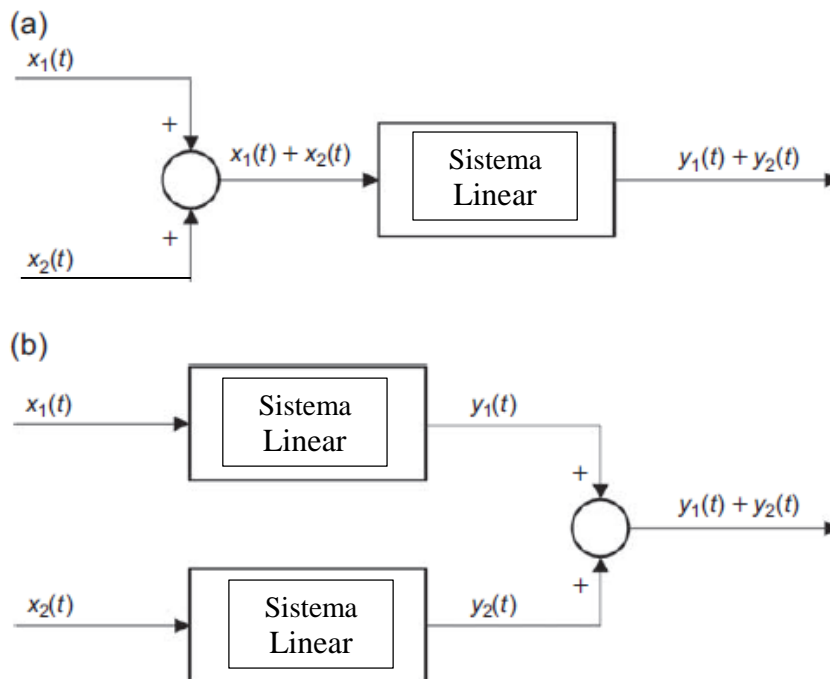


Figura 2.2: Princípio da superposição de efeitos

Fonte: Adaptado de (MELLODGE, 2016)

Modelos de Sistemas Variantes e Invariantes no Tempo

Sistemas também podem ser variantes ou invariantes no tempo. Se um sistema for invariante no tempo, uma saída, dada uma entrada em determinado tempo t , será igual a uma saída, dada uma entrada em um tempo $t + i$. Ou seja, um sistema será invariante no tempo se os seus parâmetros não mudarem com o tempo, caso contrário, o sistema será considerado variável no tempo.

Modelos de Sistemas com e sem Memória Longa

Um processo de memória longa é caracterizado pela dependência significativa entre observações distantes (em tempo e espaço). Neste caso, existe persistência nas autocorrelações observadas, que decaem lentamente para zero (SOLDÁ, 2008). Para tal, é necessário um modelo que atenda às necessidades deste tipo de processo.

Modelos de Sistemas Distribuídos e Concentrados

Em sistemas nos quais variáveis espaciais são verificadas, a diferença entre modelos distribuídos e concentrados pode ser interessante se compreender. A identificação de uma escala de sofisticação começando por modelos concentrados que tratam o sistema completo como se fosse inteiramente homogêneo, chegando a modelos distribuídos nos quais o sistema são divididos em área unitárias elementares como uma rede.

Modelos de Sistemas Estocásticos e Determinísticos

Em muitos problemas do mundo real, há incertezas na modelagem de sistemas dinâmicos, de maneira que, traçar suas trajetórias se torna uma tarefa complexa. Nem todos os parâmetros e entradas necessários para a modelagem de sistemas dinâmicos são sempre conhecidos. Além disso, o sistema pode sofrer algum tipo de ruído, que não se pode compreender em plenitude. Quando isso ocorre, diz-se que o sistema é estocástico e sua modelagem será realizada a partir de determinado nível de incerteza através de modelos utilizando estatística. No contrário, quando as entradas e parâmetros são conhecidos, diz-se que o sistema é determinístico, sendo estes, geralmente, menos custosos na modelagem.

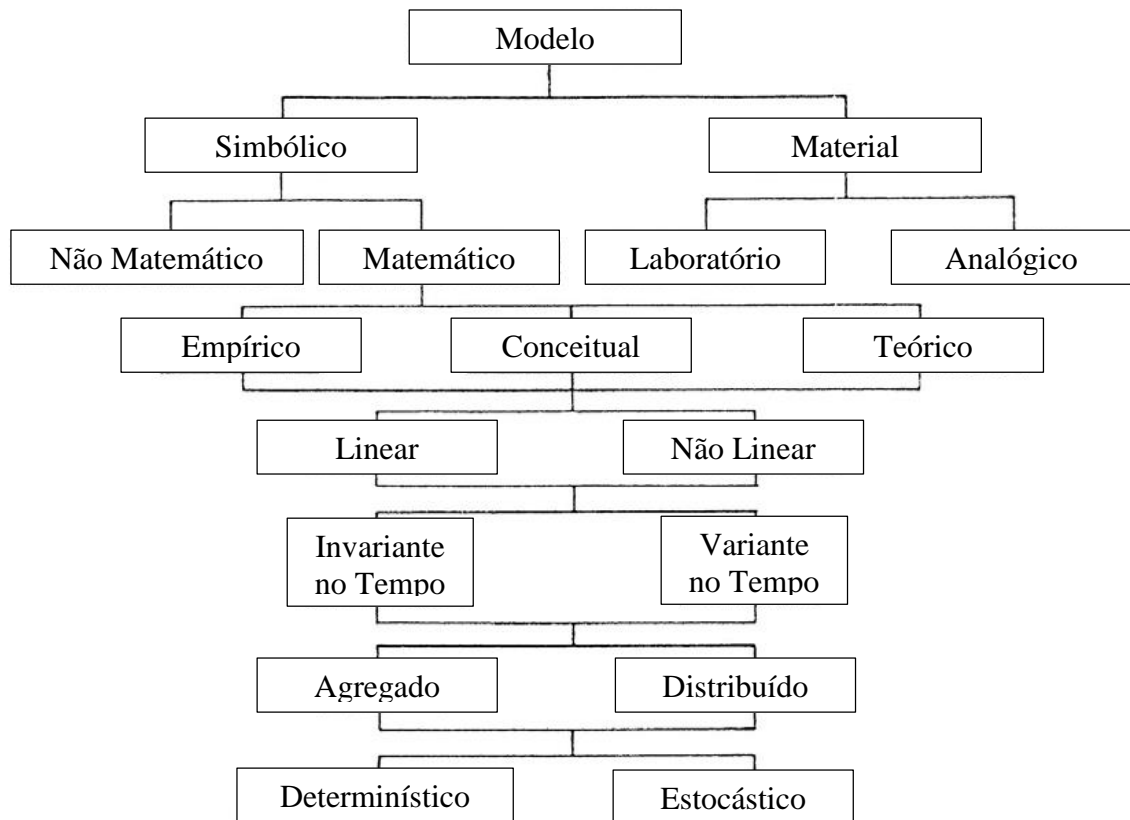


Figura 2.3: Classificação de modelos

Fonte: Adaptado de (XU, 2002)

Modelagem de Sistemas por Espaço de Estado

Outra abordagem interessante para a modelagem de sistemas dinâmicos é pela representação por espaços de estado. De certa forma, a representação de sistemas dinâmicos por espaço de estados é similar à representação por equações diferenciais. No entanto, a representação por espaço de estados retrata uma demonstração mais organizada e simplificada para modelagem de sistemas dinâmicos, pois a partir das equações diferenciais ou a diferenças, é realizado um rearranjo sob a forma de matrizes e variáveis de estado do sistema.

A forma generalista de um modelo linear invariante no tempo por espaço de estado para sistemas no tempo contínuo é dada pela Equação 2.3 e para sistemas no tempo discreto pela Equação 2.4.

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad (2.3)$$

$$\begin{cases} x[n+1] = Ax[n] + Bu[n] \\ y[n] = Cx[n] + Du[n] \end{cases} \quad (2.4)$$

onde (i) x é o vetor de estado contendo N estados de N -ésima ordem do sistema; (ii) u é o vetor de entrada contendo p entradas no sistema; (iii) y é o vetor de saída contendo q saídas do sistema; (iv) A é a matriz de estados; uma matriz $N \times N$ que relaciona os estados com suas derivadas (contínuo) ou o valor seguinte (discreto); (v) B é a matriz de entrada; uma matriz $N \times p$ que relaciona as entradas com a derivada dos estados (contínuo) ou o valor seguinte (discreto), sendo p o número de entradas; (vi) C é a matriz de saída; uma matriz $q \times N$ que relaciona os estados com as saídas, sendo q o número de saídas; (vii) D é a matriz de transição direta; uma matriz $q \times p$ que relaciona as entradas com as saídas.

É interessante enfatizar que para sistemas variantes no tempo as matrizes da modelagem por espaço de estados, que representam os parâmetros do modelo, serão representados por $A(t)$, $B(t)$, $C(t)$, $D(t)$, para o caso contínuo, e $A[n]$, $B[n]$, $C[n]$, $D[n]$, para o caso discreto. Outra questão interessante de se perceber é que, para sistemas SISO, a matriz B é reduzida para uma matriz $N \times 1$, a matriz C é reduzida para uma matriz $1 \times N$, e a matriz D é reduzida a um escalar d .

Primeiramente para entender a abordagem de modelagem de sistemas dinâmicos por espaço de estados é necessário definir o termo estado. MELLODGE (2016) traz algumas reflexões sobre o termo e tenta definir como, em tradução livre: “estado se refere a uma noção abstrata de onde o sistema se encontra em um espaço geométrico”. Ainda MELLODGE (2016) há uma discussão do quão difícil é ter uma precisa definição do termo estado na área de modelagem de sistemas. A definição dada anteriormente é bastante vaga, porém é exatamente isso que faz com que a modelagem de sistemas dinâmicos por espaço de estado seja tão poderosa, pois pode-se escolher o estado da maneira que mais for apropriada para o analista.

Para ter uma definição um pouco mais elucidativa, CANNON (2003) afirma que para se entender o conceito de estado pode-se fazer uma analogia com o termo configuração. O estado de um sistema, então, pode ser definido como um conjunto independente de coordenadas de posicionamento mais suas derivadas, o que implica em

configuração mais velocidade. A configuração informa onde o sistema está, porém um estado informa ambos, onde o sistema está e quão rápido e em que direção o sistema está se deslocando.

O conceito de variável de estado é mais claro e direto. Variáveis de estado são o menor subconjunto possível que pode representar o estado de um sistema inteiro em qualquer instante de tempo dado. O menor número de variáveis de estado necessários para representar um dado sistema é igual à ordem da equação diferencial que define o sistema.

Todas essas definições provenientes da teoria de sistemas podem ser utilizadas para a modelagem de qualquer tipo de sistema e são suficientes para o entendimento do presente trabalho. Um dos sistemas que são correntemente modelados a partir de tais definições são os sistemas hidrológicos chuva-vazão, alvo da dissertação. Estes serão discutidos na seção a seguir.

2.2. MODELAGEM DE SISTEMAS HIDROLÓGICOS CHUVA-VAZÃO

A modelagem de sistemas hidrológicos chuva-vazão tem se mostrado importante desde o início das pesquisas da área. O controle de cheias é uma área deveras importante onde a modelagem de sistemas chuva-vazão é frequentemente utilizada para construção de cenários e seus riscos associados com relação à ocupação da área de interesse. No planejamento de transporte hidroviário, também é necessário a cenarização de afluências para analisar se os cursos d'água terão profundidade suficiente para que as embarcações possam transitar. Estudos sobre disponibilidade hídrica também são conduzidos utilizando modelagem de sistemas chuva-vazão, seja para o abastecimento público de água, irrigação ou até para geração de energia elétrica através de usinas hidrelétricas.

De acordo com LINSLEY (1982) *apud* (MISHRA; SINGH, 2003), é possível que a história da modelagem chuva-vazão tenha se iniciado em 1674 com a publicação do relatório do engenheiro francês P. Perreault sobre mensurações quantitativas em hidrologia, que comparou a precipitação anual medida (P_a) e a vazão anual estimada (Q_a) do rio Sena em Paris. Com suas pesquisas, Perreault conseguiu estabelecer a relação P_a - Q_a apresentada na Equação 2.5.

$$Q_a = 1/6 P_a \quad (2.5)$$

Apesar de ter havido diversas tentativas em relacionar chuva e vazão, o primeiro modelo chuva-vazão amplamente utilizado foi o Método Racional (BEVEN, 2012), o qual ainda é empregado nos dias de hoje em muitas aplicações da engenharia, tendo sido desenvolvido pelo engenheiro irlandês Thomas Mulvany em 1851. TODINI (2007) enfatiza que o Método Racional é uma exposição clara do conceito de tempo de concentração e sua relação com a vazão máxima. O modelo estima a vazão de pico, porém não informa volume de cheia, sendo significativo fisicamente somente em pequenas e impenetráveis bacias, nas quais a vazão é efetivamente um processo puramente cinemático. A Equação 2.6 define o Método Racional.

$$Q_p = CA\bar{R} \quad (2.6)$$

Onde: (i) Q_p é a vazão de pico; (ii) C é um coeficiente empírico de escoamento superficial (*runoff*) que depende das características do solo da área de captação; (iii) A é a área de captação; e (iv) \bar{R} é intensidade média de chuva da bacia.

De acordo com DOOGE (1973), durante a última parte do século 19 e no início do século 20, a maioria dos engenheiros utilizava, ou fórmulas empíricas derivadas de casos particulares e aplicadas a outros casos sob a premissa de que as condições eram similares o suficiente, ou usavam o método racional. Assim, modelos chuva-vazão que podiam ser aplicados a áreas de captação maiores foram notados e muitas modificações e melhorias no Método Racional foram introduzidas a fim de lidar com as características das áreas de estudo, bem como com as distribuições não uniformes da chuva, tanto no tempo como no espaço (XU, 2002).

Em 1932, outro importante e amplamente usado modelo chuva-vazão foi desenvolvido pelo engenheiro americano Leroy K. Sherman com sua Teoria da Hidrógrafa Unitária. O conceito da hidrógrafa unitária, de acordo com TODINI (2007), era baseado no princípio da superposição de efeitos, já discutida na seção 2.1, e permitia não só o cálculo do pico de descarga, porém previa a hidrógrafa completa a partir de um evento de chuva amostrado em intervalos constantes. XU (2002) também declara que, apesar de não compreendido ainda naquele momento, o princípio da superposição

implicava em algumas premissas que a bacia estudada se comportava como um sistema causativo, linear, tempo invariante em respeito à transformação chuva-vazão.

TODINI (1988) comentou que um grande passo veio durante os anos 50, quando hidrólogos começaram a estudar a fundo a abordagem da engenharia de sistemas, utilizada para a análise de sistemas dinâmicos complexos. Este foi o período em que modelos lineares conceituais surgiram. TODINI (1988) ainda menciona que os modelos desenvolvidos por Box and Jenkins proporcionaram aos hidrólogos um método alternativo de previsão de vazão. SPOLIA e CHANDER (1974) foram um dos primeiros cientistas a simularem um modelo de previsão de vazão utilizando a abordagem de Box and Jenkins, o modelo ARMA, e compararam à sua eficiência com modelos conceituais.

Algumas contribuições surgiram envolvendo modelos físicos, como o modelo distribuído espacialmente AFFDEF (MORETTI; MONTANARI, 2007). Esse tipo de modelo é interessante sob o ponto de vista da identificação de mecanismos físicos envolvidos no processo de escoamento de um rio, porém apresenta algumas desvantagens, pois necessita de um grande número de parâmetros para descrever o solo, a vegetação e características necessárias da bacia para a modelagem do fenômeno chuva-vazão.

Outros modelos baseados em dados também ganharam muita popularidade durante os anos 90 devido à sua capacidade de simular sistemas complexos e não lineares, com custos computacionais baixos. Segundo DE VOS (2009), modelos como Redes Neurais Artificiais - RNA, *Support Vector Machines* – SVM, bem como modelos Fuzzy também apresentaram bons desempenhos e começaram a ser utilizados por hidrólogos para modelagem de sistemas chuva-vazão.

2.3. MODELO SOIL MOISTURE ACCOUNTING

PROCEDURE - SMAP

O Soil Moisture Accounting Procedure – SMAP é um modelo determinístico, conceitual e semi-concentrado de simulação hidrológica do tipo transformação chuva-vazão (LOPES, 1999). Foi desenvolvido em 1981 por Lopes J.E.G., Braga B.P.F. e Conejo J.G.L., e apresentado no International Symposium on Rainfall-Runoff Modeling realizado em Mississippi, U.S.A. e publicado pela Water Resources Publications (1982). Foi originalmente desenvolvido para intervalo de tempo diário e posteriormente foram

apresentadas as versões horária e mensal, com algumas adaptações em sua estrutura. Neste trabalho foi utilizada a versão diária (LOPES; BRAGA; CONEJO, 1982).

O SMAP vem sendo utilizado em muitos casos, sendo inclusive utilizado na previsão de vazão de algumas bacias hidrográficas no Brasil pelo ONS, desde 2008, para o planejamento da operação do SIN. Segundo ONS (2008) o SMAP apresenta duas vantagens comparativas para a modelagem de sistemas hidrológicos chuva-vazão:

- Facilidade de entendimento da metodologia e do funcionamento do modelo e de seus parâmetros, o que permite a realização de alguns ajustes/aprimoramentos, quando necessário.
- Facilidade na obtenção dos dados de entrada necessários e na aplicação para a grande maioria das bacias do SIN.

A matemática por trás da versão original do SMAP é constituída de três reservatórios lineares hipotéticos representando: o reservatório do solo (R_{solo}); o reservatório da superfície (R_{sup}), correspondente ao escoamento superficial da bacia; e o reservatório subterrâneo (R_{sub}), correspondente ao escoamento subterrâneo da bacia (escoamento de base) (LOPES, 1999).

Para bacias com significativas planícies de inundação, onde, em eventos de fortes chuvas, são verificados importantes extravasamentos pelas margens e amortecimentos nos picos das cheias, torna-se conveniente a inclusão de um quarto reservatório linear para representar o armazenamento e o escoamento de água nessas planícies. Outra vantagem da utilização deste esquema de quatro reservatórios é que, caso se queira inibir o funcionamento do quarto reservatório, basta fixar um valor relativamente elevado para o parâmetro H . Dessa forma, pode-se utilizar o esquema do modelo SMAP apresentado na Figura 2.4 para qualquer sub-bacia (ONS, 2008).

As variáveis de estado de cada um dos quatro reservatórios são atualizadas a cada instante de tempo, de acordo com as Equações 2.7 à 2.11.

$$R_{solo}(t + 1) = R_{solo}(t) + P(t) - Es(t) - Er(t) - Rec(t) \quad (2.7)$$

$$R_{sub}(t + 1) = R_{sub}(t) + Rec(t) - Eb(t) \quad (2.8)$$

$$R_{sup}(t + 1) = R_{sup}(t) + Es(t) - Marg(t) - Ed(t) \quad (2.9)$$

$$R_{sup2}(t + 1) = R_{sup2}(t - 1) + Marg(t) - Ed2(t) \quad (2.10)$$

$$\text{Se } R_{solo}(t) \leq Str \rightarrow R_{solo}(t) = R_{solo}(t) \quad (2.11)$$

$$\text{Senão } R_{sup}(t) = R_{sup}(t) + R_{solo}(t) - Str \ \& \ R_{solo}(t) = Str$$

Onde: (i) $R_{solo}(t)$ é o reservatório do solo no instante de tempo t ; (ii) $R_{sup}(t)$ é o reservatório da superfície da bacia no instante de tempo t ; (iii) $R_{sub}(t)$ é o reservatório subterrâneo (zona saturada) no instante de tempo t ; (iv) $P(t)$ é precipitação média na bacia, a ser considerada pelo modelo no instante de tempo t (mm); (v) $Es(t)$ é o escoamento superficial no instante de tempo n (mm); (vi) $Ed(t)$ é o escoamento direto no instante de tempo t ; (vii) $Er(t)$ é a evapotranspiração real no instante de tempo t ; (viii) $Rec(t)$ é a recarga subterrânea no instante de tempo t ; (ix) $Eb(t)$ é o escoamento básico no instante de tempo t ; (x) Str capacidade de saturação do solo (mm); (xi) t instante de tempo. Vale ressaltar que na descrição do modelo SMAP t é representado como sendo no tempo discreto.

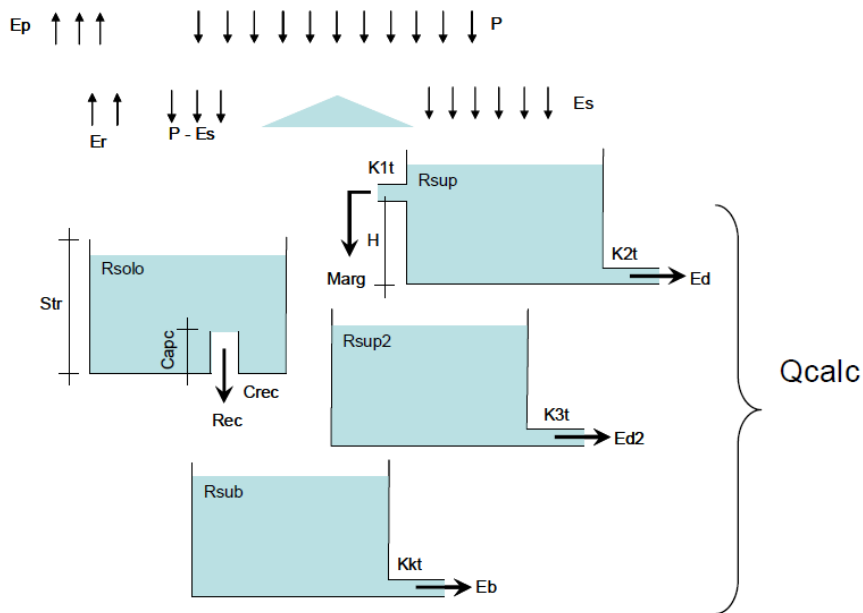


Figura 2.4: Esquema do modelo SMAP na versão com quatro reservatórios

Fonte: (ONS; HICON, 2012)

O modelo SMAP com quatro reservatórios é composto de sete funções de transferência apresentadas nas Equações 2.12 à 2.19.

$$\text{Se } P(t) > Ai \rightarrow S = Str - R_{solo}(t - 1) \text{ \& } Es(t) = \frac{(P(t)-Ai)^2}{P(t)-Ai+S} \quad (2.12)$$

$$\text{Senão } Es(t) = 0$$

$$\text{Se } (P(t) - Es(t)) > Ep(t) \rightarrow Er(t) = Ep(t) \quad (2.13)$$

$$\text{Senão } Er(t) = (P(t) - Es(t)) + (Ep(t) - (P(t) - Es(t))) \cdot Tu(t)$$

$$\text{Se } R_{solo}(t - 1) \leq \frac{Capc}{100} \cdot Str \rightarrow Rec(t) = 0 \quad (2.14)$$

$$\text{Senão } Rec(t) = \frac{Crec}{100} \cdot Tu(t) \cdot (R_{solo}(t - 1) - \frac{Capc}{100} \cdot Str)$$

$$\text{Se } R_{sup}(t - 1) > H \rightarrow Marg(t) = (R_{sup}(t - 1) - H) \cdot (1 - 0.5^{\frac{1}{K1t}}) \quad (2.15)$$

$$\text{Senão } Marg(t) = 0$$

$$Ed(n) = R_{sup}(t - 1) \cdot (1 - 0.5^{\frac{1}{K2t}}) \quad (2.16)$$

$$Ed2(n) = R_{sup2}(t - 1) \cdot (1 - 0.5^{\frac{1}{K3t}}) \quad (2.17)$$

$$Eb(n) = R_{sub}(t - 1) \cdot (1 - 0.5^{\frac{1}{Kkt}}) \quad (2.18)$$

$$Tu(n) = \frac{R_{solo}(t - 1)}{Str} \quad (2.19)$$

Onde: (i) Ai é a abstração inicial (mm); (ii) $Ep(t)$ é a evapotranspiração potencial (mm) no instante de tempo t (iii) $Tu(t)$ é o teor de umidade do solo (adimensional) no instante de tempo n ; (iv) $Capc$ é a capacidade de campo (%); (v) $Crec$ o parâmetro de recarga subterrânea (%); (vi) H é a altura representativa para início de escoamento em planícies (mm); (vii) $K1t$ é a constante de recessão do escoamento para planícies (dia); (viii) $K2t$ é a constante de recessão do escoamento superficial (dia); (ix) $K3t$ é a constante de recessão do escoamento da superfície/planícies (dia); e (x) Kkt é a constante de recessão do escoamento básico (dia).

O cálculo da vazão é dado pela Equação 2.20:

$$Q_{calc}(t) = \frac{Ed(t) + Ed2(t) + Eb(t)}{86,4} \cdot Ad \quad (2.20)$$

Onde: (i) $Q_{calc}(t)$ é a vazão total calculada pelo modelo no instante de tempo n (m^3/s); (ii) Ad é a área de drenagem da bacia considerada (km^2)

As constantes de recessão - $K1t, K2t, K3t, Kkt$ - são associadas à duração do intervalo, medido em dias, no qual a vazão do correspondente reservatório cai à metade de seu valor (não considerando nova recarga nesse período). O eventual transbordo do reservatório do solo é transformado em escoamento superficial.

Os dados de entrada do modelo são os totais diários de chuva - $P(t)$, os totais diários de evapotranspiração potencial - $Ep(t)$, e as vazões médias diárias observadas em cada bacia considerada no estudo - $Qobs(t)$. O total diário de chuva - $P(t)$, a ser considerado pelo modelo no instante de tempo t , em cada bacia, é calculado a partir das Equações 2.21 e 2.22:

$$Pb(t) = \sum_{i=1,2,3\dots}^z P_i(t) \cdot ke_i \quad (2.21)$$

$$P(t) = \sum_{i=-3}^1 Pb(t-i) \cdot kt(i) \quad (2.22)$$

Onde: (i) $\sum_{i=1,2,3\dots}^z ke_i = 1$; (ii) $\sum_{i=-3}^1 kt(i) = 1$; (iii) $Pb(t)$ é a precipitação média observada na bacia, no instante de tempo t (mm); (iv) $P_i(t)$ é precipitação observada em cada posto pluviométrico da bacia, no instante de tempo t (mm); (v) ke_i são coeficientes de representação espacial de cada posto pluviométrico, ou seja, o peso de cada posto no cálculo da precipitação média $Pb(t)$ observada na bacia; (vi) $kt(i)$ são coeficientes de representação temporal, ou seja, os pesos utilizados para o cálculo da precipitação média $P(t)$ na bacia, a ser considerada pelo modelo.

Os coeficientes de representação temporal foram utilizados para o cálculo da precipitação média na bacia a ser considerada pelo modelo devido aos seguintes aspectos:

- Em geral, as medidas de precipitação nos postos pluviométricos são realizadas às 7:00h, ou seja, a maior parte da precipitação ocorrida no dia t só é medida e computada no dia $t + 1$. Dessa forma, principalmente em bacias com tempo de concentração relativamente pequeno, a vazão média do dia t pode ser influenciada pela precipitação medida no dia $t + 1$.
- Em bacias com tempos de concentração maiores, pode haver uma defasagem temporal entre os picos de precipitação e de vazão. Neste caso, modelos concentrados ou mesmo semiconcentrados, como o SMAP, têm mais dificuldades

de reproduzir esta defasagem, caso a precipitação média observada na bacia no instante de tempo t seja considerada integralmente como a precipitação a ser utilizada pelo modelo no instante de tempo t .

Os totais diários de evapotranspiração potencial - $Ep(t)$ são obtidos a partir dos valores mensais de evapotranspiração potencial estimados pelo modelo SISEVAPO (MÜLLER, 2001) *apud* (ONS, 2008).

2.4.1 Calibração do Modelo SMAP

Os parâmetros a serem ajustados no processo de calibração do modelo SMAP estão dispostos na Tabela 2.1. Segundo (ONS; HICON, 2012) vale ressaltar que os seis primeiros parâmetros (Ai , Str , $Capc$, $Crec$, $K2t$ e Kkt) são os mesmos a calibrar quando se utiliza a versão original do modelo SMAP com três reservatórios. Os três parâmetros seguintes (H , $K1t$, $K3t$) são decorrentes da incorporação do quarto reservatório, que representa o eventual extravasamento pelas margens e escoamento pelas planícies de inundação.

Os coeficientes ke , kt e kep devem ser calibrados com vistas a uma melhor representatividade dos dados de precipitação média da sub-bacia ($P(t)$) e de evapotranspiração potencial média da sub-bacia ($Ep(t)$), a serem utilizados pelo modelo.

Os valores mínimos e máximos dos coeficientes de representação espacial das estações pluviométricas (ke) dependem da quantidade de estações existentes na sub-bacia e em sua proximidade, bem como da localização e situação relativa de cada estação.

O valor mínimo do parâmetro H (altura representativa para início de escoamento em planícies) deve ser estimado com base nos resultados preliminares da calibração e na experiência do hidrólogo na sub-bacia. No presente trabalho inibimos a utilização do quarto reservatório adotando um valor alto para H (por exemplo: 200,0mm) (ONS; HICON, 2012).

Tabela 2.1: Faixas de variação dos parâmetros do modelo SMAP

Parâmetros	Descrição	Unidade	Limite Superior	Limite Inferior
Ai	abstração inicial	mm	1,00	5,00
Str	capacidade de saturação do solo	mm	50,00	2000,00
Capc	capacidade de campo	%	30,00	50,00
Crec	parâmetro de recarga subterrânea	%	0,00	100,00
K2t	constante de recessão do escoamento superficial	dia	0,20	10,00
Kkt	constante de recessão do escoamento básico	dia	30,00	180,00
H	altura representativa para início de escoamento em planícies	mm	Variável	Variável
K1t	constante de recessão do escoamento planícies	dia	0,20	20,00
K3t	constante de recessão do escoamento superficial/planícies	dia	0,20	10,00
ke ₁ ;ke ₂ ;...;ke _n	coeficientes de representação espacial de cada posto pluviométrico	-	Variável	Variável
kt ₍₋₃₎ ;kt ₍₋₂₎ ;...;kt ₍₊₁₎	coeficientes de representação temporal	-	0,00	1,00
kep	coeficiente de ajuste da evapotranspiração potencial média da sub-bacia	-	0,80	1,20

Fonte: (ONS; HICON, 2012)

Para iniciar a calibração, de acordo com ONS e HICON (2012), é necessário fornecer valores iniciais para as variáveis de estado do modelo, ou seja, valores de $R_{solo}(0)$, $R_{sub}(0)$, $R_{sup}(0)$ e $R_{sup2}(0)$. Estes valores são calculados pelas seguintes expressões:

$$R_{solo}(0) = \frac{Tuin}{100} \cdot Str \quad (2.23)$$

$$R_{sup}(0) = \frac{Supin}{(1 - 0.5^{K2t}) \cdot Ad} \cdot 86,4 \quad (2.24)$$

$$R_{sup2}(0) = \frac{Sup2in}{(1 - 0.5^{K3t}) * Ad} \cdot 86,4 \quad (2.25)$$

$$R_{sub}(0) = \frac{Ebin}{(1 - 0.5^{Kkt}) * Ad} \cdot 86,4 \quad (2.26)$$

Onde: (i) *Tuin* é o teor de umidade inicial do solo (%); (ii) *Ebin* é a vazão básica inicial (m³/s); (iii) *Supin* é a vazão superficial inicial, proveniente do reservatório da superfície (m³/s); (iv) *Sup2in* é a vazão superficial inicial, proveniente do reservatório da superfície/planície (m³/s).

A correta estimativa dos valores de *Tuin*, *Ebin*, *Supin* e *Sup2in* é importante para o bom desempenho do modelo. Para isso, deve-se sempre iniciar o período de calibração no meio para o final do período seco do ano, com valores baixos de umidade do solo e de vazão superficial. Dessa forma, a vazão básica inicial (*Ebin*) deverá ser próxima da vazão observada inicial (*Qobs*(0)). Para a estimativa do valor de *Supin*, deve-se verificar a quantidade de chuva ocorrida nos dias anteriores. O valor de *Sup2in*, em geral, é nulo (ONS; HICON, 2012).

A calibração dos parâmetros foi realizada com uso da rotina *Solver* da planilha eletrônica MS Excel, de forma semi-automática e iterativa, ajustando-se, a cada passo, os parâmetros do modelo, buscando-se a minimização do valor da função objetivo, sem perder o sentido físico representativo de cada parâmetro ou variável do modelo.

Segundo ONS e HICON (2012) a Função Objetivo utilizada na calibração do modelo (otimização dos coeficientes) baseou-se na minimização da Função Objetivo (*FO*), a qual foi construída a partir de duas métricas de erros, incluindo Coeficiente de Eficiência (*CE*) e o Coeficiente do Erro Relativo (*CER*), sendo definida da seguinte forma:

$$FO = \sqrt{(1 - CE)^2 + (1 - CER)^2} \quad (2.27^1)$$

Onde:

¹ A *FO* também pode ser escrita como $FO = \sqrt{(1 - NSE)^2 + (MAPE)^2}$

$$CE = \frac{\sum(Qobs(t) - \overline{Qobs})^2 - \sum(Qobs(t) - Qcalc(t))^2}{\sum(Qobs(t) - \overline{Qobs})^2} \quad (2.28)$$

$$CER = 1 - \frac{1}{n} \cdot \sum_{t=1}^n \frac{abs(Qobs(t) - Qcalc(t))}{Qobs(t)} \quad (2.29)$$

Onde: (i) *FO* é a Função Objetivo multicritério; (ii) *CE* é o coeficiente de eficiência Nash-Sutcliffe; (iii) *CER* é o coeficiente de erro relativo; (iv) *Qobs(t)* é a vazão observada no instante *t* (m³/s); (v) *Qcalc(t)* é a vazão calculada no instante *t* (m³/s); (vi) \overline{Qobs} é a média da vazão observada no período considerado (m³/s).

O coeficiente de eficiência (*CE*) apresenta valores sempre inferiores à unidade, podendo variar na faixa de valores compreendidos entre $-\infty$ e 1. Valores elevados do coeficiente de eficiência (próximos à unidade) indicam elevada associação entre os dados observados e calculados, representando um bom ajuste do modelo aos dados de campo. A obtenção de um coeficiente de eficiência igual à unidade representaria o ajuste perfeito entre as vazões observadas e calculadas. Devido à sua formulação, o coeficiente de eficiência é mais sensível aos desvios nas vazões mais elevadas (ONS; HICON, 2012).

O coeficiente de erro relativo (*CER*) também apresenta valores sempre inferiores à unidade, podendo variar na faixa de valores compreendidos entre $-\infty$ e 1. Da mesma forma, a obtenção de um coeficiente de erro relativo igual à unidade representaria o ajuste perfeito entre as vazões observadas e calculadas. Devido à sua formulação matemática, o coeficiente de erro relativo é mais sensível aos desvios nas vazões mais baixas (ONS; HICON, 2012).

A função objetivo proposta para calibração representa a distância entre 2 (dois) vetores, podendo variar entre 0 e $+\infty$, sendo a obtenção de um valor igual a 0 (zero), distância mínima, representaria o ajuste perfeito entre as vazões observadas e calculadas. Devido à sua formulação, esta expressão deverá ser sensível aos desvios em todo o período do ano, tanto nas vazões mais baixas, quanto nas vazões mais elevadas (ONS; HICON, 2012).

2.4. MODELO SISTEMA FUZZY RECORRENTE

Em (ZADEH, 1965) pela primeira vez foi introduzido o conceito de conjuntos fuzzy. O autor argumentou que no mundo físico real os critérios de definição da função

de pertinência não são precisamente definidos. O autor citou que, claramente, cavalos, cachorros e pássaros podiam pertencer ao grupo de animais, mas rochas, fluidos, ou vegetais não podiam pertencer ao mesmo grupo. Porém havia alguns seres mais complexos de definir a função de pertinência, como as estrelas do mar ou as bactérias, devido às suas características. Dessa forma, alguns conjuntos como “dos homens altos” ou “das mulheres bonitas” não constituem conjuntos no sentido matemático usual da época, mas que de fato essa incerteza faz parte importante do pensamento humano, principalmente na área de reconhecimento de padrões, comunicação de informação e abstração.

ZADEH (1965) definiu uma nova possibilidade de classificação de objetos em relação a conjuntos. Em vez de classificar os objetos binariamente, sendo 0 não pertencente e 1 pertencente, o autor criou os conjuntos fuzzy que se caracterizam por ser uma classe de objetos com graus de pertinência contínuos. Tal conjunto é caracterizado por uma função de pertinência que atribui a cada objeto um grau de pertinência variando de 0 a 1. Dessa forma, os conjuntos clássicos podem ser entendidos como um caso dos conjuntos fuzzy.

O trabalho de (ZADEH, 1965) foi muito importante para o início da exploração da lógica fuzzy aplicada a problemas do mundo real. Segundo DOS SANTOS (2013), a teoria de conjuntos fuzzy foi muito importante para o desenvolvimento da teoria do raciocínio aproximado, também elaborada por Zadeh em 1976. DOS SANTOS (2013) ainda afirma que tais conceitos definiram uma nova estrutura que permitiu modelar níveis de incerteza na representação e no processamento de informações através da ideia de raciocínio aproximado. Com isso, foi possível desenvolver modelos que a descrição matemática é substituída por relações qualitativas expressas por termos linguísticos.

A teoria fuzzy foi aplicada pela indústria e academia através da interpretação de seus respectivos sistemas como Sistemas Fuzzy. Praticamente, aos sistemas de especialistas tradicionais, foram incorporados os conceitos fuzzy que puderam oferecer dois importantes atributos como: (i) o sistema é representado por conceitos linguísticos, imprecisos e vagos, por meio de conjuntos fuzzy; (ii) sua natureza como aproximador universal (KOSKO, 1994), permitindo que o sistema seja utilizado para modelar teoricamente, qualquer problema aproximável por uma função real contínua, linear ou não-linear (COUTINHO, 2015).

Ao final da década de 70 e início dos anos 80, a teoria de conjuntos fuzzy começou a ser utilizada em aplicações reais. Um bom exemplo, foi o trabalho de (KING;

MAMDANI, 1977) que aplicou os conceitos da teoria de conjuntos fuzzy em um processo industrial. A ideia central se desenvolve com a afirmativa de que processos industriais complexos eram difíceis de controlar automaticamente, devido ao comportamento não linear, variável no tempo, e da baixa qualidade das medições na época. Dessa forma, algumas variáveis como pressão, temperaturas, entre outras, podiam ser mensuradas, enquanto que outras variáveis como qualidade e quantidade dos produtos eram controladas por decisões humanas, imprecisas, heurísticas e de difícil compreensão, dos operadores das fábricas, baseadas em suas experiências. Com isso, políticas de controle de qualidade foram aplicadas linguisticamente como um conjunto de regras heurísticas e imprecisas de decisão.

Outra aplicação interessante sob o ponto de vista de modelagem foi o trabalho de (TONG, 1980) que aplicou um modelo fuzzy a dados experimentais referentes a monitoramento ambiental. O primeiro conjunto de dados foi o de quantidade de dióxido de carbono emitido de um forno a gás a partir da taxa de variação do uso de combustível. Já o segundo foi na área de hidrologia, a qualidade da água de determinado ponto em um rio em relação à demanda bioquímica de oxigênio (DBO) e o oxigênio dissolvido (OD) a partir de dados coletados de quatro variáveis: DBO e OD a montante, vazão e temperatura. Em ambos, foi realizada a tentativa de se chegar a um modelo fuzzy que representasse os sistemas em questão.

GORRINI e BERSINI (1994) introduziram um controlador fuzzy que podia ser aplicado a sistemas dinâmicos, o Sistema Fuzzy Recorrente – SFR. Os autores abordam o tema declarando que aproximadores universais, tais como as Redes Neurais Artificiais – RNA e Sistemas Fuzzy, poderiam ser utilizados para modelar processos não-lineares com melhor desempenho. A identificação do SFR de (GORRINI; BERSINI, 1994), no entanto, era realizada através de um método gradiente para a otimização dos parâmetros do modelo.

Uma alternativa de SFR foi apresentada por (EVSUKOFF; EBECKEN, 2004) que utilizava Algoritmos Genéticos – AG como método de otimização e identificação dos parâmetros do modelo. AG são simples, robustos e muitas aplicações podem ser encontradas na internet. Além disso, AG podem superar problemas com mínimos locais, problema frequentemente encontrado com modelos de otimização baseados em gradiente. O SFR proposto por (EVSUKOFF; EBECKEN, 2004) será o utilizado na presente dissertação e sua formulação matemática será descrita a seguir.

2.5.1 Formulação Matemática de Sistemas Fuzzy Recorrente

Considere um sistema dinâmico não-linear com a representação por espaço de estado como na Equação 2.30:

$$\begin{cases} \mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) \\ \mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t)) \end{cases} \quad (2.30)$$

Onde: (i) $\mathbf{x}(t)$ são as variáveis de estado; (ii) $\mathbf{u}(t)$ as variáveis de entrada; e (iii) \mathbf{f} e \mathbf{g} são as funções de transição e de saída, respectivamente. Para a atual aplicação, o SFR é utilizado para modelar a função de transição, enquanto que a função de saída é considerada uma função linear de estados, tal como: $\hat{\mathbf{y}}(t) = \mathbf{K}\mathbf{x}(t)$

Há na literatura, dois modelos SFR, o tipo TSK, introduzido por (GORRINI; BERSINI, 1994), bem como o tipo Mamdani, explanado em (EVSUKOFF; EBECKEN, 2004). Nesse trabalho será utilizado somente o tipo TSK, pela menor complexidade na otimização dos parâmetros do modelo. As regras do modelo tipo TSK são escritas de acordo com a Equação 2.31:

$$\text{Se } \mathbf{x}(t) \text{ é } A_i \text{ e } u(t) \text{ é } B_j \text{ então } x_0(t+1) = \theta_r \quad (2.31)$$

Onde $\mathbf{x}(t) = (x_0(t), \dots, x_{n-1}(t))$ é o vetor de n variáveis de estado. Os conjuntos fuzzy A_i e B_j são definidos no domínio das variáveis de estado e de entrada, respectivamente, e θ_r o parâmetro de saída para cada uma das regras $r = 1, \dots, M$.

Todas as variáveis são consideradas nas premissas das regras, porém somente a variável de estado $x_0(t)$ é calculada na saída. Se forem necessários mais estados para ajustar a dinâmica do sistema, novas variáveis de estado são adicionadas como cópias de valores defasados da variável de estado $x_0(t)$, tal como na Equação 2.32:

$$x_i(t) = x_0(t-i), i = 1, \dots, n-1 \quad (2.32)$$

O conjunto dos conjuntos $\mathbf{B} = \{B_j, j = 1, \dots, q\}$ define a partição fuzzy no domínio da variável de entrada. Como todas as variáveis de estado são cópias defasadas

da variável de estado $x_0(t)$, a mesma partição fuzzy $\mathbf{A} = \{A_i, i = 1, \dots, p\}$ pode ser usada para calcular a fuzzificação de todas as variáveis de estado.

Uma partição fuzzy é considerada forte, se a soma de todos os graus de pertinência de determinado número *crispy*² for igual à unidade. Tal adoção é interessante para a modelagem de Sistemas Fuzzy, pois auxilia na interpretação dos resultados. Dessa forma, as funções de pertinência são consideradas triangulares e normalizadas, tal que as seguintes condições da Equação 2.33 são satisfeitas.

$$\begin{aligned} \sum_i \mu_{A_i}(x) &= 1 \quad \forall x \\ \forall i, \exists x: \mu_{A_i}(x) &= 1 \end{aligned} \quad (2.33)$$

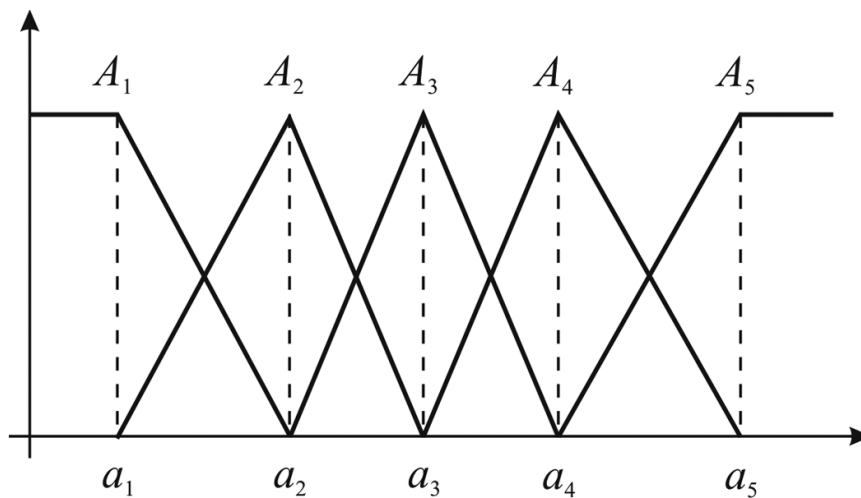


Figura 2.5: Funções de pertinência triangulares

Fonte: (GAMA et al., 2008)

Existem diversos outros tipos de função de pertinência que podem ser encontradas facilmente na literatura. No entanto, o uso de funções triangulares reduz a complexidade, bem como o custo computacional, e não necessariamente apresentam piores resultados que outras funções de pertinência mais complexas. Funções de pertinência triangulares e normalizadas são definidas por um vetor $\mathbf{a} = \{a_1, \dots, a_n\}$ que definem os vértices das funções de pertinência e são conhecidos por protótipos. Por exemplo, considere a partição fuzzy demonstrada na Figura 2.5, para $p = 5$ conjuntos fuzzy. A fuzzificação das

² Números *crispy* são aqueles que, ao contrário dos números fuzzy, se adequam a teoria clássica, ou *crispy*, de conjuntos. Em outras palavras, são números únicos comuns.

variáveis estado (Equação 2.34) e entrada (Equação 2.35) produzem os vetores de fuzzificação:

$$\boldsymbol{\mu}_A(x(t)) = (\mu_{A_1}(x(t)), \dots, \mu_{A_p}(x(t))) \quad (2.34)$$

e

$$\boldsymbol{\mu}_B(u(t)) = (\mu_{B_1}(u(t)), \dots, \mu_{B_q}(u(t))) \quad (2.35)$$

A saída do modelo SFR é a variável de estado $x(t + 1)$, onde é calculada como padrão em modelos TSK de Sistemas Fuzzy, como a Equação 2.36:

$$x(t + 1) = \mathbf{w}(t)\boldsymbol{\theta} \quad (2.36)$$

Onde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$ é o vetor de parâmetros de saída e $\mathbf{w}(t)$ é o vetor o qual seus componentes são os valores de ativação das premissas das regras.

Os componentes do vetor $\mathbf{w}(t)$ são calculadas pela junção dos conjuntos fuzzy das premissas. Diversas formas podem ser realizadas para calcular o vetor $\mathbf{w}(t)$, como em (EVSUKOFF; EBECKEN, 2004), porém, como em (GAMA *et al.*, 2008), o operador produto será usado na forma do produto tensorial de Kronecker³, dado pela Equação 2.37:

$$\mathbf{w}(t) = \boldsymbol{\mu}_A(x(t)) \otimes \boldsymbol{\mu}_B(u(t)) \quad (2.37)$$

Dessa forma, todas as combinações dos conjuntos fuzzy são consideradas nas premissas das regras, tal que o modelo possa cobrir todo o domínio. O número de regras M é definido pelo número de conjuntos fuzzy das partições fuzzy das variáveis de entrada e estados, considerando p conjuntos fuzzy na partição fuzzy da variável de estado e q conjuntos fuzzy na partição fuzzy das variáveis de entrada. O número total de regras fuzzy em SFR SISO de primeira ordem⁴ do tipo TSK é $M = p \cdot q$.

Modelos SFR de maiores ordens podem alcançar dinâmicas mais complexas. Nesse caso, n variáveis de estado são consideradas no vetor $\mathbf{x}(t) = (x_0(t), \dots, x_{n-1}(t))$.

³ O resultado do produto tensorial de Kronecker de dois vetores $\mathbf{a} = (a_1, \dots, a_n)$ $\mathbf{b} = (b_1, \dots, b_m)$ é o vetor $\mathbf{a} \otimes \mathbf{b} = (a_1\mathbf{b}, \dots, a_n\mathbf{b}) = (a_1b_1, a_2b_2, \dots, a_nb_m)$

⁴ Que apresenta somente uma variável de estado

Em cada simulação do tempo, a fim de reduzir o número de parâmetros a serem otimizados, variáveis de estado de maiores ordens são atualizadas como cópias defasadas da variável de estado $x_0(t)$, de acordo com a Equação 2.38:

$$x_k(t) = x_0(t - k), \quad k = 1 \dots n - 1 \quad (2.38)$$

Dessa forma, todas as variáveis de estado são consideradas nas premissas, porém somente o valor da variável de estado $x_0(t)$ é calculado na saída do SFR, tal que as regras são escritas na forma da Equação 2.39:

$$\text{Se } \mathbf{x}(t) \text{ é } C_i \text{ e } u(t) \text{ é } B_j \text{ então } x_0(t + 1) = \theta_r \quad (2.39)$$

O conjunto fuzzy C_i na regra anterior é um conjunto fuzzy multidimensional, calculado como o produto Cartesiano dos conjuntos fuzzy do domínio das variáveis de estado. Todas as combinações devem ser consideradas de tal forma que o domínio inteiro das variáveis de estado seja coberto. O conjunto do rótulo simbólico $\mathbf{C} = \{C_i, i = 1 \dots p^n\}$ é então o produto Cartesiano dos elementos do conjunto do rótulo \mathbf{A} da variável de estado, tal como na Equação 2.40:

$$\mathbf{C} = \mathbf{A} \times \dots \times \mathbf{A} = \mathbf{A}^n \quad (2.40)$$

Como as variáveis de estados são cópias defasadas da variável de estado $x_0(t)$, a mesma partição fuzzy pode ser utilizada para calcular todos os vetores de pertinência das variáveis de estado. O vetor de pertinência $\boldsymbol{\mu}_{\mathbf{C}}(\mathbf{x}(t))$ (Equação 2.41) também é calculado como o produto tensorial de Kronecker das componentes correspondentes em cada domínio da variável de estado.

$$\boldsymbol{\mu}_{\mathbf{C}}(\mathbf{x}(t)) = \boldsymbol{\mu}_{\mathbf{A}}(x_{n-1}(t)) \otimes \dots \otimes \boldsymbol{\mu}_{\mathbf{A}}(x_0(t)) \quad (2.41)$$

Onde $\boldsymbol{\mu}_{\mathbf{A}}(x_k(t))$ é o vetor de fuzzificação da variável de estado $x_k(t)$ calculado como Equação 2.34.

O mesmo raciocínio pode ser adotado para estender o modelo SFR SISO do tipo TSK para modelos SFR MIMO do tipo TSK, tais que as regras são escritas de acordo com a Equação 2.42:

$$\text{Se } \mathbf{x}(t) \text{ é } C_i \text{ e } \mathbf{u}(t) \text{ é } B_j \text{ então } x_0(t+1) = \theta_r \quad (2.42)$$

Onde $\mathbf{u}(t) = (u_1(t), \dots, u_m(t))$ é o vetor de variáveis de entrada.

O vetor de pertinência da entrada $\mu_B(\mathbf{u}(t))$ pode ser calculado como a combinação de conjuntos fuzzy no domínio das variáveis de entrada, como no caso das variáveis de estado, porém resultarão geralmente em um número grande de regras. Para simplificar as notações, o número de conjuntos fuzzy na partição fuzzy da variável de entrada, seja uni ou multidimensional, é considerado por q . O número total de regras em um SFR MIMO do tipo TSK é então dado por $M = p^n \cdot q$, onde n é a ordem do modelo e p o número de conjuntos fuzzy na partição fuzzy da variável de estado. A saída do modelo fuzzy $x_0(t+1)$ é então calculada através da Equação 2.43:

$$x(t+1) = \mathbf{w}(t)\boldsymbol{\theta} \quad (2.43)$$

Onde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$ é o vetor de parâmetros de saída e $\mathbf{w}(t)$ é o vetor o qual seus componentes são os valores de ativação das premissas das regras calculado como na Equação 2.44:

$$\mathbf{w}(t) = \mu_A(\mathbf{x}(t)) \otimes \mu_B(\mathbf{u}(t)) \quad (2.44)$$

2.5.2 Identificação de SFR através de Algoritmos Genéticos

De acordo com EVSUKOFF e EBECKEN (2004), em geral, a identificação de processos complexos segue uma metodologia composta de três passos:

- i. Identificação da estrutura;
- ii. Estimação dos parâmetros a partir de um conjunto de dados representativo do comportamento do sistema chamado de conjunto de treinamento;
- iii. Validação do modelo, considerando objetivos de modelagem: previsão, simulação, diagnóstico, etc.

A identificação da estrutura para um modelo fuzzy consiste em determinar o tipo de modelo e o número e localização dos conjuntos fuzzy no domínio de cada variável. Os parâmetros do modelo são geralmente associados com os parâmetros das conclusões das regras. A validação do modelo deve contemplar a precisão do modelo, mas também certificar que o modelo é legível por especialistas do domínio.

A abordagem mais comum verificada na literatura para estimativa dos parâmetros é adotar algoritmos baseados em gradiente. Porém, o uso de AG para a identificação de SFR foi proposta atingindo bons resultados (GAMA *et al.*, 2008).

Algoritmos Genéticos

Algoritmos Genéticos – AG tem sido explorados amplamente em diversas aplicações relacionadas a Sistemas Fuzzy. O procedimento de resolução de problemas através de AG é inspirado na evolução natural. As soluções candidatas para um dado problema são codificadas em um dado indivíduo. Uma população inicial de indivíduos gerada randomicamente é desenvolvida em direção a uma solução. A cada iteração, chamada por geração na área, os indivíduos são avaliados pela aptidão, ou fitness, em relação à solução e um mecanismo de seleção garante que somente os melhores indivíduos são mantidos para a próxima geração através da aplicação de operadores de recombinação como a mutação e cruzamentos (*crossover*). Operadores de recombinação guiados de forma randômica permitem os indivíduos a explorar o espaço de solução por completo, evitando, assim, mínimos locais. Apesar de apresentar diversas variantes, uma estrutura AG pode ser encontrada de acordo com o seguinte algoritmo:

```
Procedimento AG  
  inicia  
    t = 0;  
    inicializa P(t);  
    avalia estruturas em P(t)  
    enquanto {não termina}  
      inicia  
        t = t + 1;  
        seleciona P(t) de P(t - 1);  
        recombina estruturas em P(t);  
        avalia estruturas em P(t);  
      fim  
    fim
```

Figura 2.6: Estrutura típica de um AG

Fonte: Adaptado de (GAMA *et al.*, 2008)

Quando um AG é aplicado a um novo problema, duas questões devem ser definidas pelo usuário: (i) os parâmetros do modelo a serem estimados; e (ii) a função de objetivo.

O AG, no presente trabalho, é utilizado para identificar o vetor de parâmetros θ e a função objetivo utilizada para tal é a métrica de erro Mean Squared Error – MSE definida por pela Equação 2.45. Diversas estruturas são testadas por exaustão e aquela que apresenta o menor valor de MSE (Equação 2.45) é a escolhida.

$$MSE = \frac{1}{N} \sum_{t=1 \dots N} (y(t) - v(t))^2 \quad (2.45)$$

3. MATERIAIS E MÉTODOS

3.1. APRESENTAÇÃO DAS REGIÕES HIDROGRÁFICAS BRASILEIRAS

O Brasil possui uma das mais extensas e diversificadas redes fluviais do mundo, dividida em 12 regiões hidrográficas⁵: Bacia Amazônica, Bacia Tocantins Araguaia, Bacia do Paraguai, Bacia Atlântico Nordeste Ocidental, Bacia Atlântico Nordeste Oriental, Bacia do Paraná, Bacia do Parnaíba, Bacia do São Francisco, Bacia do Atlântico Leste, Bacia do Atlântico Sudeste, Bacia do Atlântico Sul e Bacia do Uruguai (BRASIL, 2009) e podem ser acompanhadas na Figura 3.1.

Bacia Amazônica

Considerada a rede hidrográfica mais extensa do mundo, a Bacia Amazônica ocupa uma área total de 7.008.370 km². Esta área vai desde as nascentes, nos Andes Peruanos, até sua foz (local onde o rio deságua) no Oceano Atlântico - 64,88% (ou 3.843.402 km²) desse total ficam em território brasileiro e o restante está dividido entre a Colômbia (16,14%), Bolívia (15,61%), Equador (2,31%), Guiana (1,35%), Peru (0,60%) e Venezuela (0,11%). No Brasil, a Bacia Amazônica é compartilhada por sete Estados - Acre, Amazonas, Amapá, Rondônia, Roraima, Pará e Mato Grosso. Seus principais rios são Javari, Purus, Madeira, Tapajós e Xingú (pela margem direita) e Iça, Japurá, Negro, Trombetas, Paru e o Jarí (pela margem esquerda) (BRASIL, 2009).

Bacia Tocantins-Araguaia

Com uma área total de 967.059 km², segundo BRASIL (2009), a Bacia Tocantins-Araguaia ocupa 11% do território nacional. Grande parte está na Região Centro-Oeste, nos Estados de Goiás, Tocantins, Pará, Maranhão, Mato Grosso e Distrito Federal. Como o próprio nome diz, os dois principais rios dessa bacia são o Tocantins e o Araguaia. O Tocantins nasce no planalto de Goiás, a cerca de 1.000 metros de altitude. Com 1.960 km de extensão até sua junção com o rio Araguaia, ele tem como principais afluentes (rios menores que deságuam no rio principal) os rios Bagagem, Tocantinzinho, Paranã, dos

⁵ Uma rede hidrográfica é o conjunto formado pelo rio principal e todos os seus afluentes e subafluentes.

Sonos, Manoel Alves Grande e Farinha (margem direita) e rio Santa Tereza (margem esquerda). Em seus 2.600 km, o Araguaia abriga a maior ilha fluvial do mundo – a Ilha do Bananal – com 350 km de comprimento e 80 km de largura.

Bacia do Paraguai

De acordo com BRASIL (2009), o rio Paraguai nasce na Chapada dos Parecis, no Mato Grosso. Ao longo do seu percurso rumo ao sul, recebe vários afluentes importantes como o Cuiabá, o São Lourenço, o Taquari, o Miranda e o Negro. Sua bacia hidrográfica abrange uma área de 1.095.000 km², sendo 33% no Brasil – Mato Grosso e Mato Grosso do Sul - e o restante na Argentina, Bolívia e Paraguai. A região se divide em duas áreas principais hidrográficas: Planalto (215.963 km²), com terras acima de 200 metros de altitude, e Pantanal (147.629 km²), que são terras abaixo de 200 metros de altitude, com baixa capacidade de drenagem e sujeitas a grandes inundações. Considerado uma das maiores extensões úmidas contínuas do planeta, o Pantanal funciona como um grande reservatório que retém a maior parte da água oriunda do Planalto e regulariza a vazão do rio Paraguai. A baixa capacidade de drenagem dos rios e lagoas que se formam no Pantanal, juntamente com a influência do clima da região, faz com que cerca de 60% da água proveniente do Planalto seja perdida por evaporação.

Bacia Atlântico Nordeste Ocidental

Localizada no Estado do Maranhão e em uma pequena porção oriental do Pará, fazem parte da região hidrográfica do Atlântico Nordeste Ocidental os rios Gurupi, Turiaçu, Pericumã, Mearim, Itapecuru, Munim e a região do litoral do Maranhão. Com uma área de 254.100 km², a bacia atinge 233 municípios, sendo 9% no Pará e 91% no Maranhão (BRASIL, 2009).

Bacia Atlântico Nordeste Oriental

Segundo BRASIL (2009), a Bacia do Atlântico Nordeste Oriental não tem grandes rios e, por isso, apresenta baixa disponibilidade de água em relação à demanda local, principalmente em períodos de estiagem. Seus principais rios são o Capibaribe, Paraíba, Jaguaribe e Acaraú. Os 287.348 km² (3% do território brasileiro) dessa bacia atingem cinco Estados do Nordeste e suas capitais (Ceará, Rio Grande do Norte, Paraíba, Pernambuco e Alagoas), dezenas de núcleos urbanos e um grande parque industrial. Além disso, a região reúne diversas bacias costeiras de pouca extensão. No litoral do Ceará, Rio

Grande do Norte, Paraíba e Pernambuco podem ser encontrados estuários (parte de um rio que se encontra em contato com o mar), manguezais e lagoas costeiras. O litoral de Alagoas inclui o delta do rio São Francisco, compartilhado com Sergipe, e o Complexo Estuarino-Lagunar Mundaú / Manguaba.

Bacia do Paraná

A região onde está localizada a bacia do Paraná é de grande importância para o País e tem o maior desenvolvimento econômico do país e atinge 32% da população brasileira. Ocupa 10% do território nacional (879.860 km²) e se divide entre os Estados de São Paulo, Paraná, Mato Grosso do Sul, Minas Gerais, Goiás, Santa Catarina e Distrito Federal. A bacia recebe esse nome por ter o rio Paraná como seu principal formador. Com uma extensão de 2750 km até sua foz, o Paraná tem como principais afluentes o Paranaíba e o Grande. Essa região hidrográfica se subdivide em seis grandes rios: Grande, Iguaçu, Paranaíba, Paranapanema, Paraná e Tietê, apresentando uma vazão média correspondente a 6,5% do total do país. A bacia do Paraná também é a que possui a maior capacidade de produção (59,3% do total nacional) e demanda (75% do consumo nacional) de energia do país. Existem 176 usinas hidrelétricas na região, com destaque para Itaipu, Furnas, Porto Primavera e Marimbondo (BRASIL, 2009).

Bacia do Parnaíba

Com 344.112 km² de área (3,9% do território nacional), de acordo com BRASIL (2009), a Bacia do Parnaíba ocupa 99% do Piauí, 19% do Maranhão e 10% do Ceará. No Piauí, a água subterrânea representa a principal fonte de abastecimento da população. Em áreas semiáridas, nas quais muitos rios são intermitentes (ou seja, descontínuos, que terminam e recomeçam por intervalos), é a única alternativa para os habitantes. Parte da Bacia do Parnaíba é marcada por um elevado índice de pobreza, e a proporção da população que se encontra em zonas rurais (40%) é alta em relação à média nacional (18,2%). Nessa região, a utilização média de água por hectare é superior à média do Brasil. Um dos motivos para isso é a intensa perda de água para a atmosfera, causada pela evaporação a partir do solo e pela transpiração das plantas.

Bacia do São Francisco

Conhecido como o rio da integração nacional, o São Francisco tem sido cenário de fatos históricos do país. Sua região hidrográfica abrange sete Estados: Minas Gerais,

Distrito Federal, Goiás, Bahia, Pernambuco, Alagoas e Sergipe. Com cerca de 2.700 km de extensão, o São Francisco nasce na Serra da Canastra (MG) e corre para o norte, seguindo até Pernambuco, onde muda o percurso para o Sudeste e desagua no Oceano Atlântico entre Alagoas e Sergipe. Ao todo são 168 afluentes, dos quais 99 constantes e 69 intermitentes. As hidrelétricas da bacia do São Francisco são responsáveis por grande parte do abastecimento de energia da Região Nordeste. São 33 usinas em operação – nove no próprio rio São Francisco. Além disso, as barragens também são usadas para abastecimento, lazer e irrigação (BRASIL, 2009).

Bacia do Atlântico Leste

Com uma área que corresponde a 8% do país (374.677 km²), a região hidrográfica do Atlântico Leste inclui parte dos Estados de Sergipe, Bahia, Minas Gerais e Espírito Santo. Atinge 526 cidades, alguns grandes núcleos urbanos e um parque industrial. Entre seus principais rios estão o Paraguaçu, Contas, Salinas, Pardo, Jequitinhonha e Mucuri. Além disso, nas bacias costeiras, entre Sergipe e Espírito Santo, também existe uma grande diversidade de rios, córregos e riachos (BRASIL, 2009).

Bacia do Atlântico Sudeste

Região mais populosa do país, o Sudeste também possui o maior pólo econômico e industrial do Brasil, de acordo com BRASIL (2009). Por isso, a região hidrográfica do atlântico sudeste – distribuída pelos Estados do Espírito Santo, Minas Gerais, Rio de Janeiro, São Paulo e o litoral do Paraná - possui grande importância no cenário econômico nacional. Com uma área de 229.972 km², equivalente a 2,7% do território brasileiro, seus principais rios são o Paraíba do Sul e Doce, com respectivamente 1.150 e 853 km. Além desses, vários outros rios de menor porte formam as seguintes bacias: São Mateus, Santa Maria, Reis Magos, Benevente, Itabapoana, Itapemirim, Jacu, Ribeira e litorais do Rio de Janeiro e de São Paulo. Por ser a mais populosa e industrializada, a região tem uma grande demanda de água (10% do total nacional), sendo 41% para a área urbana e 15% para a área industrial.

Bacia do Atlântico Sul

A região hidrográfica Atlântico Sul tem início na divisa dos Estados de São Paulo e Paraná e se estende até o Arroio Chuí, no extremo sul do país. Com uma área total de 185.856 km² (2% do país) a região abrange partes dos Estados do Paraná, Santa Catarina

e Rio Grande do Sul. Na bacia do Atlântico Sul, predominam rios de pequeno porte que escoam diretamente para o mar. As exceções mais importantes são os rios Itajaí e Capivari, em Santa Catarina, que apresentam maior volume de água. Na região do Rio Grande do Sul são encontrados rios de grande porte, como o Taquari-Antas, Jacuí, Vacacaí e Camaquã (BRASIL, 2009).

Bacia do Uruguai

Com 2.200 km de extensão, segundo com BRASIL (2009), o rio Uruguai nasce na junção dos rios Pelotas e Peixe, e segue em direção ao oeste dividindo os Estados do Rio Grande do Sul e Santa Catarina. Em seu caminho, ele também se une com o rio Peperi-Guaçu, servindo de fronteira entre Brasil e Argentina. Seguindo na direção sudoeste, o Uruguai se une com o rio Quaraí (que limita o Brasil e o Uruguai) e daí toma a direção sul, passando a dividir Argentina e Uruguai até a sua foz. A região hidrográfica do Uruguai tem grande importância para o país, pois atende a agroindústria e tem grande potencial hidrelétrico. Junto com as regiões hidrográficas do Paraná e Paraguai, ela forma a grande bacia do Prata. A bacia do Uruguai se divide entre os Estados do Rio Grande do Sul e Santa Catarina. Sua área total 385.000 km², sendo que 45% em território nacional.



Figura 3.1: Divisão hidrográfica do Brasil

Fonte: (CNRH, 2003)

3.2. APRESENTAÇÃO DAS UHE ESTUDADAS

Nesta seção será realizada uma breve descrição das UHE utilizadas no presente trabalho. Foram escolhidas quatro UHE localizadas em diferentes regiões hidrográficas para realizar a modelagem e comparar seus resultados e que o resumo das informações pode ser observado na Tabela 3.1.

UHE Furnas

A construção da UHE de Furnas entrou em operação por completo em 1965. A barragem está localizada no curso médio do rio Grande, no trecho denominado "Corredeiras das Furnas", entre os municípios de São José da Barra e São João Batista do Glória, em Minas Gerais. No início da década de 70, foi iniciada sua ampliação para a instalação das sétima e oitava unidades, totalizando 1.216 MW, o que colocou a obra entre uma das maiores da América Latina. A Usina de Furnas, além de se constituir em um marco de instalação de grandes hidrelétricas no Brasil, possibilitou a regularização do rio Grande e a construção de mais oito usinas, aproveitando, integralmente, um potencial de mais de 6.000 MW instalados (FURNAS, 2017).

UHE Machadinho

A UHE Machadinho está localizada na divisa dos estados de Santa Catarina e do Rio Grande do Sul, entre os municípios de Piratuba (SC) e Maximiliano de Almeida (RS). O empreendimento começou a operar em fevereiro de 2002, utilizando o potencial hidrelétrico do Rio Pelotas, principal afluente do Rio Uruguai, que forma uma das maiores bacias hidrográficas do Sul do Brasil. Com três unidades geradoras de 380MW, a UHE Machadinho tem potência instalada de 1.140MW, o que corresponde a cerca de 37% da demanda de energia do estado de Santa Catarina ou 23% do total consumido pelo estado do Rio Grande do Sul (CONSÓRCIO MACHADINHO, 2017).

UHE Mascarenhas

A UHE Mascarenhas situa-se no Baixo Curso do rio Doce, no distrito de Mascarenhas, município de Baixo Guandu-ES, iniciando sua operação em 1974. A UHE Mascarenhas opera a fio d'água gerando 189 MW de potência instalada (ANEEL, 2017) com 4 unidades geradoras. Possui um reservatório com 3,9 km² de área e volume aproximado de 21.800.000 m³.

UHE Três Marias

A Usina Hidrelétrica de Três Marias foi inaugurada em 1962 com potência instalada de 396 MW. A barragem, que tem 2.700 metros de comprimento e forma um reservatório de 21 bilhões de metros cúbicos de água, a 2.221 km acima da foz do rio é considerada de grande importância para o Brasil. Localizada na parte central de Minas Gerais, compreende os municípios de São Gonçalo do Abaeté, Felixlândia, Morada Nova de Minas, Biquinhas, Paineiras, Pompéu, Martinho Campos e Abaeté além de Três Marias tendo partes de seus municípios alagados e formando o Circuito Turístico do Lago de Três Marias (CEMIG, 2017).

Tabela 3.1: Características das UHE estudadas

UHE	Capacidade (MW)	UF	Long	Lat	Divisão Hidrográfica	Rio	Extensão da Bacia (km ²)
Furnas	1.216,00	MG	-46,32	-20,67	Paraná	Grande	51.881,93
Machadinho	1.140,00	SC	-51,79	-27,53	Uruguai	Pelotas	32.023,57
Mascarenhas	189,00	ES	-41,01	-19,53	Atl Leste	Doce	73.645,76
Três Marias	396,00	MG	-45,26	-18,21	S. Francisco	São Francisco	50.766,93

Fonte: Elaboração própria

3.3. PRÉ-PROCESSAMENTO DOS DADOS

Os dados utilizados no presente trabalho foram coletados de dois websites: (i) a chuva diária do website da Agência Nacional de Águas (ANA, 2017), o Hidroweb, e (ii) a média diária de vazão natural do website do ONS (ONS, 2017). A vazão natural é fornecida de forma estruturada em formato xls do software MS Excel. Todavia, a aquisição de dados de chuva média diária é mais desafiadora, pois está em um banco de dados que pelo website só é possível baixa-los um de cada vez. Por se tratar de muitos postos, a aquisição de dados foi realizada de maneira automática através de um código na linguagem R com técnicas de *Web Scrapping*.

A localização das estações pluviométricas foi retirada do Hidroweb e foi contabilizada a quantidade de estações inseridas na área de contribuição de cada UHE analisada, com auxílio da ferramenta “Interseção” do software Quantum GIS. Em um primeiro momento, já é possível perceber que a quantidade de estações inseridas na área de contribuição de cada UHE é bem variável. Isso ocorre, pois, a exploração da informação de dados pluviométricos ocorrida no Brasil foi realizada de forma assimétrica

e regiões que possuíam maior interesse outrora começaram a ser exploradas primeiramente. Assim, as regiões Sudeste e Sul do Brasil apresentam mais estações pluviométricas instaladas que as demais regiões.

Tabela 3.2: Número de estações para cada UHE

UHE	Quantidade de Estações
Furnas	334
Machadinho	235
Mascarenhas	358
Três Marias	282

Fonte: Elaboração própria

Dentre as estações totais contabilizadas anteriormente, foram extraídas somente as que possuíam dados consistentes e disponíveis para uso gratuito.

Para a escolha do melhor conjunto de dados para a realização da modelagem, existem *tradeoffs* como um conjunto de dados que apresente: (i) dados mais recentes possíveis; (ii) a menor quantidade de falhas possível (menor índice de valores ausentes - IVA); (iii) a princípio, a maior quantidade de dados possível para dar maior margem de opções para o analista.

Para o item (i) a explanação se dá pelo fato de que, apesar de muitos sistemas ambientais serem representados por uma modelagem de sistemas invariantes no tempo, sabe-se que isso é uma simplificação. Então, quanto mais recentes forem os dados, mais próximos do sistema real serão os parâmetros do modelo. No item (ii), para a modelagem, é necessário obter as informações que são representadas pelos dados. Se não houver dados suficientes, dependendo do caso, é possível aplicar técnicas de mineração de dados para inserir valores aos dados ausentes. No entanto, quando isso ocorre, os dados inseridos são sintéticos e não correspondem ao processo do sistema real e, por isso, deve ser evitado ao máximo. Já para o item (iii), quanto mais dados forem coletados dos postos, maior será a margem de atuação para realizar a seleção de dados que realmente interesse.

O período de operação para cada estação pluviométrica pode ser diferente entre si para uma mesma bacia hidrográfica. Determinadas estações são muito antigas e pode ocorrer de algumas terem começado a operação no início do século 1900, por exemplo, enquanto que outras podem ter começado a operar recentemente. Além disso, dados de chuva apresentam falhas que ocorreram por algum motivo de parada na operação das

estações, por exemplo, ou até perda do equipamento etc. Para o presente trabalho, o ideal é que o conjunto de dados utilizado esteja na interseção dentre as estações mais antigas e mais recentes, dado uma janela temporal ideal estipulada para a modelagem.

A definição do horizonte da janela temporal ideal para a modelagem dependerá de alguns fatores e enfrenta outros *tradeoffs*. O ideal é que a janela temporal de análise seja de pelo menos um ano, para que o modelo identifique a sazonalidade de um ano hidrológico. Em hidrologia, há também a ciclicidade da chuva que ocorre devido a alguns fenômenos climáticos responsáveis por alterações nos padrões de precipitação no Brasil, e que podem apresentar intervalos de ocorrência variados. Assim, uma janela temporal maior, além de identificar a sazonalidade, também identificaria a ciclicidade do processo. Todavia, pode aumentar o custo computacional significativamente.

Analisando todas as estações inseridas na área de contribuição de cada UHE, foram extraídos os períodos de operação, representado pelo ano de início e fim. Ou seja, o ano de início é momento que pelo menos uma estação pluviométrica começou a operar, sendo o ano de fim o ano em que pelo menos uma estação finalizou a operação. A *Tabela 3.3* informa os anos de início e fim para cada UHE.

Tabela 3.3: Ano de operação verificado dentre as estações

UHE	Período de Operação
Furnas	1924-2012
Machadinho	1942-2006
Mascarenhas	1940-2005
Três Marias	1940-2006

Fonte: Elaboração própria

Algumas janelas temporais foram testadas a fim de observar o índice de valores ausentes (IVA) do conjunto de dados, ou seja, a razão da quantidade de valores ausentes pelo total de dados. Pelos IVA dos conjuntos de dados de diferentes janelas temporais, é perceptível, segundo a *Tabela 3.4* e a *Tabela 3.5*, que quanto maior for a janela temporal, maior será o IVA. No entanto, terá mais dados para treinamento e teste do modelo. Dessa forma, a janela temporal de dez anos foi escolhida, pois: (i) apresenta uma quantidade razoável de dados e margem para modelagem hidrológica; (ii) possui extensão suficiente para a sazonalidade de dez anos hidrológicos diferentes; e (iii) apresenta um IVA razoável.

Tabela 3.4: Período de análise e IVA dos conjuntos de dados de 2, 4 e 6 anos

UHE	2 anos		4 anos		6 anos	
	Período	IVA	Período	IVA	Período	IVA
Furnas	25/04/1997 - 24/04/1999	44,50%	29/12/1995 - 26/12/1999	45,02%	30/12/1993 - 27/12/1999	46,42%
Machadinho	30/12/2003 - 01/01/2006	18,10%	30/12/2001 - 01/01/2006	20,81%	31/12/1999 - 01/01/2006	22,68%
Mascarenhas	19/03/1997 - 24/03/1999	28,51%	30/03/1995 - 01/04/1999	28,62%	06/07/1993 - 03/08/1999	29,13%
Três Marias	01/01/2001 - 04/01/2003	30,15%	01/01/2001 - 03/01/2005	30,37%	02/11/1999 - 30/10/2005	30,56%

Fonte: Elaboração própria

Tabela 3.5: Período de análise e IVA dos conjuntos de dados de 8 e 10 anos

UHE	8 anos		10 anos	
	Período	IVA	Período	IVA
Furnas	30/12/1991 - 26/12/1999	47,15%	30/12/1989 - 26/12/1999	47,65%
Machadinho	30/12/1997 - 31/12/2005	24,12%	31/12/1995 - 26/12/2005	24,72%
Mascarenhas	29/10/1991 - 28/10/1999	29,51%	29/10/1991 - 29/10/2001	29,80%
Três Marias	02/11/1997 - 30/10/2005	30,93%	20/10/1995 - 20/10/2005	31,26%

Fonte: Elaboração própria

Um código foi desenvolvido em R para a verificação do período. O funcionamento do código se dá pelo percurso de uma janela de 3654 dias consecutivos que abrange todas as estações e assim contabiliza a quantidade de valores ausentes no interior dessa janela. A janela de 3654 dias com menor quantidade de valores ausentes será a utilizada para a modelagem.

Após a seleção da janela, foi analisada a quantidade de valores ausentes de cada variável, representadas pelas estações pluviométricas. Foi estipulado um valor de no máximo 5% do IVA para seleção das variáveis. Aquelas que apresentaram índice superior foram excluídas do conjunto de dados. Ao final dessa etapa, cada UHE apresentou determinado número de postos como mostra a

Tabela 3.6:

Tabela 3.6: Número de postos com IVA menor que 5%

UHE	Quantidade de Estações
Furnas	66
Machadinho	24
Mascarenhas	59
Três Marias	53

Fonte: Elaboração própria

A inserção dos valores ausentes foi realizada através de um algoritmo bastante difundido em *Data Mining*, o *k – Nearest Neighbour – kNN*, ou seja, o *k*-ésimo Vizinho Mais Próximo. Basicamente, o algoritmo calcula a distância de Gower⁶ entre todas as variáveis. Cada variável, então, irá apresentar *k* vizinhos mais próximos. Quando encontrado um valor ausente no conjunto de dados, este será preenchido a partir de uma métrica calculada a partir dos *k* vizinhos mais próximos. Tal métrica fica a critério do analista e pode ser a média, mediana, moda, etc. A linguagem R apresenta o pacote VIM para manipulação e inserção de valores ausentes em conjuntos de dados. Para mais detalhe verificar em (TEMPL *et al.*, 2016). No presente trabalho, foram escolhidos *k* sendo igual a 3 e a métrica utilizada foi a mediana.

O presente trabalho visa comparar o modelo Sistema Fuzzy Recorrente – SFR com um modelo bastante utilizado em modelagem hidrológica, o Soil Moisture Accounting Procedure – SMAP. O SFR pode ser elaborado a partir de modelos SISO ou MISO, de maneira que, quanto maior o número de entradas, maior o custo computacional da modelagem. Dessa forma, para o presente trabalho optou-se por realizar a modelagem SISO para os modelos SFR. Para tal, os dados de precipitação de todos os postos de cada bacia hidrográfica foram transformados em valores que pudessem ser comparados com os valores de vazão natural. A precipitação de cada posto é medida a partir da soma diária do total de milímetros (mm). Para transformar os dados de precipitação em um valor da mesma unidade que a vazão aplicou-se a Equação 3.1:

$$Precip_{transform} = \frac{\overline{Precip}(t) \cdot Ad}{86,4} \quad (3.1)$$

Onde: (i) $Prec_{Transf}(t)$ é a precipitação transformada no tempo *t*; (ii) $\overline{Precip}(t)$ é a média de precipitação no tempo *t* de todos os postos da bacia selecionados para a realização da modelagem; (iii) *Ad* é a área de drenagem da bacia hidrográfica analisada.

3.4. ANÁLISE EXPLORATÓRIA DOS DADOS

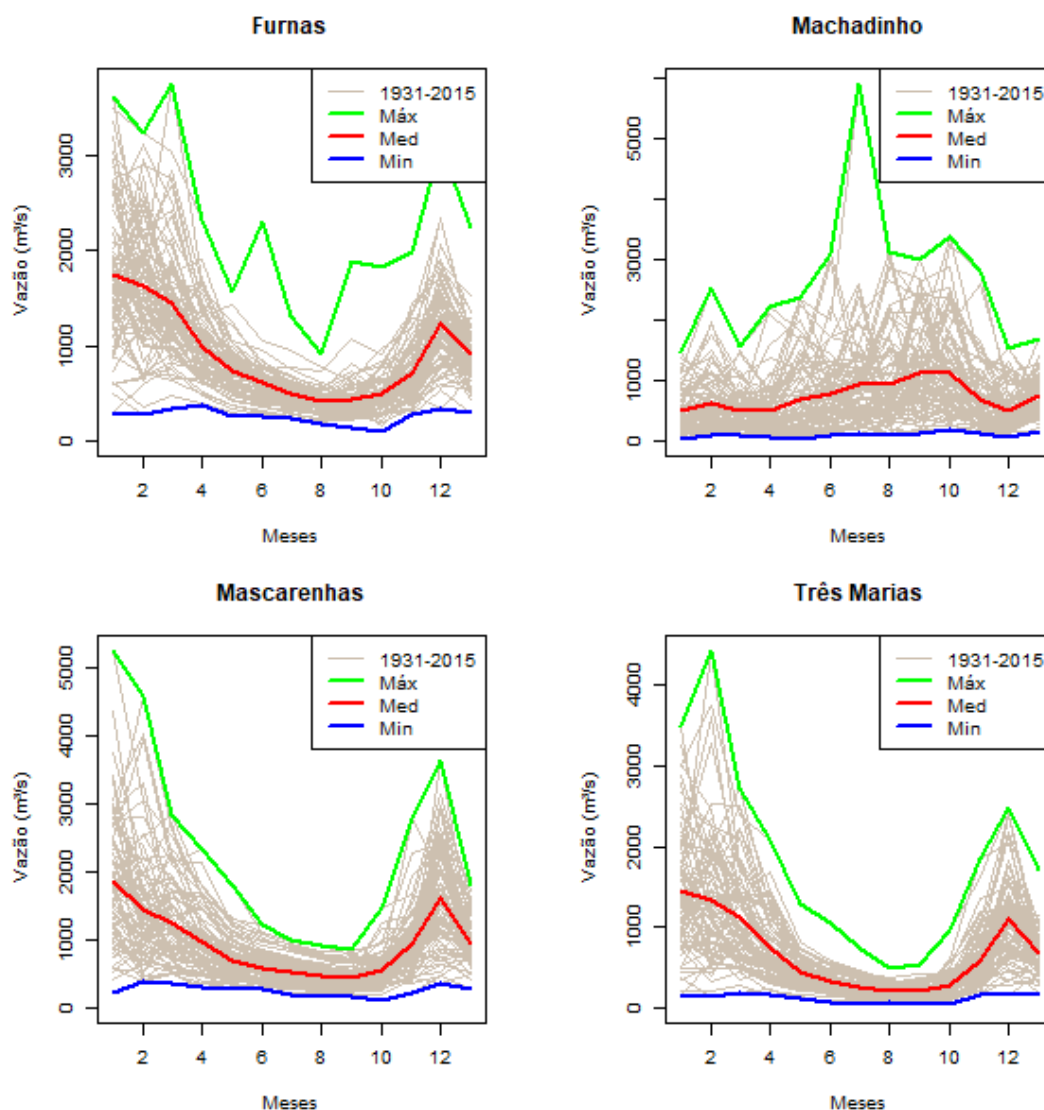
A Análise Exploratória de Dados (AED) tem o intuito de averiguar os dados previamente à aplicação de qualquer técnica estatística. Assim, o analista pode ter um

⁶ Mais detalhes sobre a distância de Gower em (GOWER, 1971)

entendimento mínimo de seu conjunto de dados e das inter-relações efetivas entre as variáveis analisadas. Tal etapa é primordial, de tal forma que uma análise descritiva detalhada permite ao analista familiarizar-se com os dados, organizá-los e sintetizá-los de forma a obter as informações necessárias do conjunto de dados para responder as questões que estão sendo estudadas.

A localização das UHE é fator importante para se ter ideia de como ocorre os regimes de chuva e subsequentemente a sazonalidade das vazões em cada UHE. Na maioria das regiões hidrográficas do Brasil, o maior regime de chuvas se inicia durante o mês de dezembro e se encerra em março, ou seja, durante o verão. O período de estiagem, dessa forma, ocorre durante os demais meses, sendo que o mínimo de chuvas acontece no inverno em junho/julho. Isso ocorre para, de certa forma, todas as grandes regiões hidrográficas, excetuando as que se localizam mais ao sul do Brasil, como: Uruguai, Atlântico Sul e a parte Sul da divisão hidrográfica do Paraná. Estas apresentam um regime de chuvas não bem delineado e sua ocorrência pode variar significativamente de um ano para outro. Com isso, não há um movimento sazonal a se extrair das vazões nessas localidades.

Para comprovar tal processo, as séries de vazão natural mensal de cada UHE estudada foram extraídas de (ONS, 2017) e plotadas de acordo com a Figura 3.2. Em um primeiro momento, é possível perceber que as séries de vazão natural mensal das UHE seguem a tendência das regiões hidrográficas brasileiras, com maiores vazões ocorrendo durante o verão e menores durante o inverno, excetuando, somente, a UHE Machadinho. Esta, como explanado na seção 3.2, não apresenta qualquer tendência sazonal e o pico e o mínimo de vazão mensal pode ocorrer em qualquer período do ano. Isso se dá, justamente, pelo fato de a UHE Machadinho estar localizada na bacia do rio Uruguai, na região sul do Brasil.



*Figura 3.2: Vazões mensais das UHE estudadas entre 1931-2015
 Fonte: Elaboração própria a partir de dados do ONS (2017)*

Na seção 3.3, foi explanado o pré-processamento dos dados e, por conseguinte, as estações pluviométricas adotadas em cada área de contribuição da bacia hidrográfica das UHE. O número de estações variou bastante para cada área estudada. Além disso, a distribuição das estações também foi diferenciada. Tal fato, de certa forma, pode impactar significativamente na modelagem do sistema chuva-vazão para cada UHE. Isso pode ocorrer, pois, o modelo SFR será do tipo SISO, de maneira que, a chuva ocorrida será transformada em uma média das estações, como explanado na Equação 3.1. Assim, se as estações não estiverem bem distribuídas ao longo da área de contribuição para cada UHE, a precipitação transformada não está inserindo a informação das demais áreas onde não há estações.

Na Figura 3.3, é possível observar que a dispersão das estações de chuva em Três Marias a parte da bacia mais próxima do exutório, ao norte. As demais UHE apresentaram boa dispersão das estações ao longo das bacias.

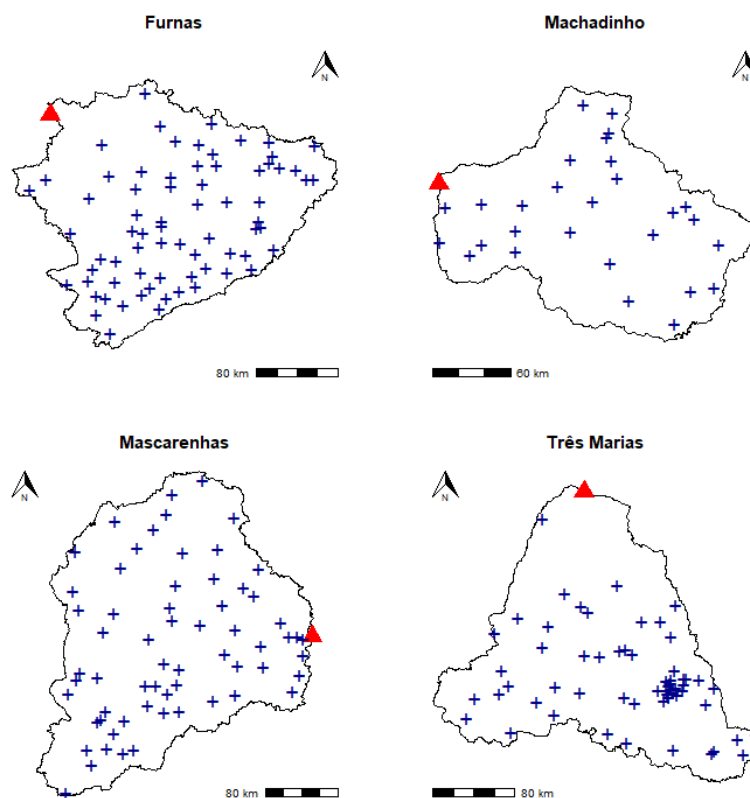


Figura 3.3: Bacias hidrográficas, dispersão das estações de chuva e localização das UHE⁷

Fonte: Elaboração própria

A primeira etapa de uma AED deve ser utilizar ferramentas gráficas para entender o comportamento das variáveis do conjunto de dados. No caso do presente trabalho, são utilizadas duas variáveis para o modelo SFR, a precipitação transformada como entrada, e outra de saída sendo a vazão natural. Como visto no início desta seção, no Brasil, a maioria das bacias hidrográficas apresenta um maior regime de precipitação durante o verão e menor no inverno, excetuando as bacias do sul. Através do Figura 3.4, pode-se observar o comportamento da precipitação transformada e da vazão natural da área de contribuição da bacia hidrográfica de cada UHE estudada. Como mencionado na seção

⁷ Triângulo vermelho representa as UHE e a cruz azul as estações de chuva

3.3, foram utilizados, 10 anos (3650 registros) de dados para cada UHE, de maneira que se pudesse inserir, pelo menos 4 períodos chuvosos na modelagem.

Para as UHE Furnas, Mascarenhas e Três Marias é possível observar que há, realmente, pelo menos, 4 períodos de maior excitação da variável de entrada chuva no sistema chuva-vazão. Ou seja, em determinados instantes de tempo, há maior probabilidade de chuva que em outros. Isso repercute em uma resposta da variável de saída vazão de maior volume nesses períodos. No entanto, para a UHE Machadinho, a chuva se comporta como uma variável aleatória com probabilidade de ocorrência uniforme ao longo do ano.

O formato da hidrógrafa⁸ também é algo importante a ser analisado. Sabe-se que a configuração da hidrógrafa de cada bacia será função de diversos fatores como o relevo, a cobertura da bacia, modificações artificiais no rio, distribuição, duração e intensidade da chuva, solo etc. Analisando a Figura 3.4, as hidrógrafas de Furnas e Mascarenhas, apresentam o tempo de ascensão prévia ao início das chuvas inferior ao tempo de recessão após as chuvas com uma forte ascensão ao pico de descarga e uma suavização da recessão. Ao contrário, Três Marias e Machadinho apresentam um tempo de recessão e ascensão similar, porém as duas primeiras apresentam um tempo de ascensão e recessão relativamente maior que as últimas. O fato de os tempos de recessão e ascensão serem similares mostra a tendência de o sistema chuva-vazão se comportar de uma maneira mais linear.

⁸ Denomina-se hidrógrafa à representação gráfica da variação da vazão em relação ao tempo.

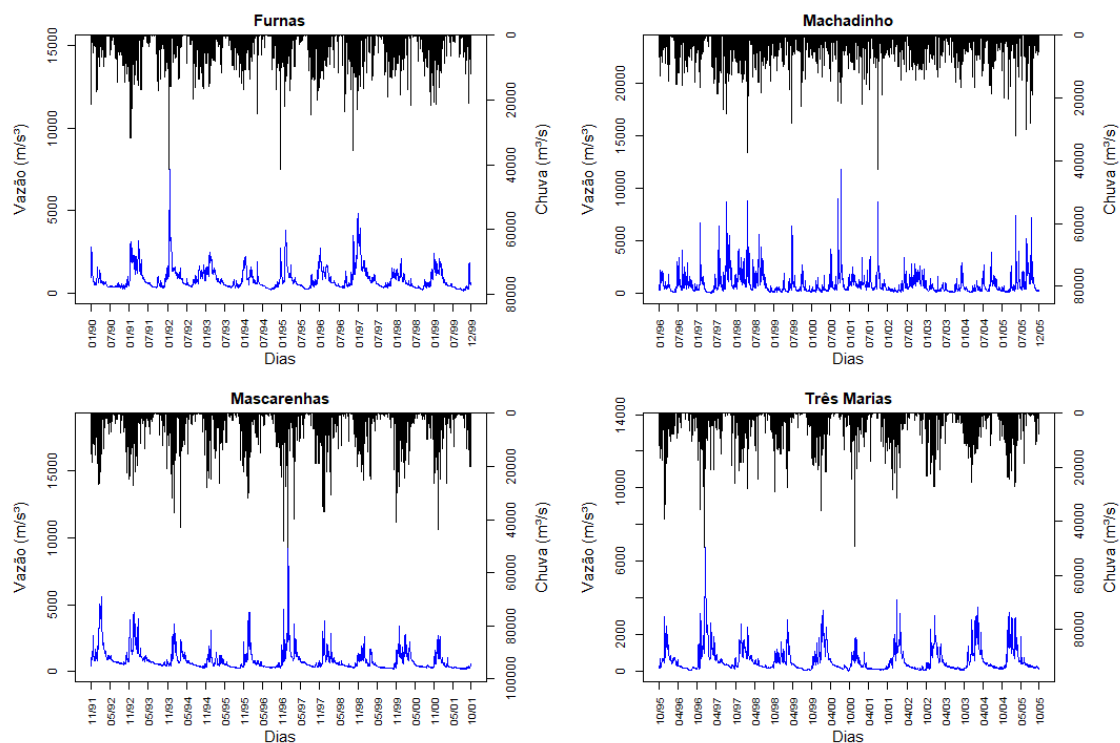


Figura 3.4: Precipitação transformada vs. vazão das UHE

Fonte: Elaboração própria

Um segundo passo na AED é descrever as variáveis estatisticamente. Para tal, foram extraídas diversas estatísticas descritivas das variáveis de entrada e saída (Tabela 3.7 e Tabela 3.8). Em geral, sistemas chuva-vazão apresentam entradas e saídas bem assimétricas. Analisando a variável de entrada, esta apresenta um alto valor de assimetria para todas as UHE devido a valores iguais a zero, ou seja ausência de chuva. Assim, a distribuição dos dados de entrada para cada UHE apresentam uma assimetria positiva, o que nos remete a uma forma com uma longa cauda para direita. A curtose de cada UHE apresentou também altos valores, bem maiores que a de uma curva gaussiana a qual apresenta o valor de curtose igual a 3. Isso significa todas apresentaram a estatística curtose afunilada e bem concentrada. Os valores de máximo foram bem superiores aos valores de medida central, média e mediana, o que o fato de haver eventos extremos durante o período de análise. A mesma análise pode ser feita para o primeiro e o segundo quartil, os quais apresentam um grande distanciamento entre eles. O desvio padrão também apresentou resultados maiores que as medidas de valor central média e mediana o que caracteriza grandes incertezas na ocorrência dos eventos de precipitação transformada diária.

A precipitação transformada é uma maneira de estabelecer a quantidade de água que serviu como entrada em toda área de contribuição na mesma unidade que a variável de saída vazão. Sistemas chuva-vazão funcionam de maneira que a área de contribuição atenua a variável de entrada chuva. O grau de atenuação dependerá das características ambientais da área de contribuição da bacia. A variável vazão natural apresenta, dessa forma, estatísticas descritivas mais suaves que a precipitação transformada. Para verificar tal afirmação, a distância entre as medidas de valor central e o desvio padrão são menores que na análise da variável de entrada. A distância entre os máximos e mínimos e a distância entre o primeiro e o terceiro quartil também confirmam a afirmação anterior. Para todos os casos, a assimetria foi positiva, fazendo com a curva de distribuição apresente uma forma na qual há uma longa cauda a direita para cada UHE. A curtose também apresentou valores superiores que a de uma curva normal.

Tabela 3.7: Estatísticas descritivas da variável de entrada precipitação transformada

UHE	Média	Desv. Pad.	Min	Max	Q1	Mediana	Q3	Assimetria	Curtose
Furnas	2344.56	3949.91	0.00	41413.47	5.46	438.99	3187.80	3.05	14.62
Machadinho	1777.18	3457.30	0.00	42732.06	0.00	208.49	1962.86	3.52	18.98
Mascarenhas	2745.84	5317.26	0.00	50512.99	23.12	385.74	2755.07	3.19	13.33
Três Marias	2260.72	4507.50	0.00	49638.33	1.11	199.00	2440.12	3.46	17.30

Fonte: Elaboração própria

Tabela 3.8: Estatísticas descritivas da variável de saída vazão

UHE	Média	Desv. Pad.	Min	Max	Q1	Mediana	Q3	Assimetria	Curtose
Furnas	886.56	680.26	174.00	7497.00	449.00	651.50	1120.25	2.79	13.58
Machadinho	888.90	957.22	78.00	11802.00	324.00	577.50	1104.00	3.48	19.32
Mascarenhas	854.10	762.83	193.00	9263.00	388.00	609.00	1019.75	3.24	17.11
Três Marias	619.07	644.26	48.00	6762.00	212.00	376.00	776.75	2.82	14.10

Fonte: Elaboração própria

De certa forma, pode-se observar que o processo de transformação da chuva em vazão dos sistemas hidrológicos apresenta uma suavização do comportamento frenético da chuva. Isso dependerá da bacia hidrográfica analisada, porém cada uma apresenta características particulares. Os histogramas das variáveis de entrada e saída para cada UHE analisada refletem tal questão (Figura 3.5 a Figura 3.8). A distribuição da chuva pode apresentar o formato de uma função exponencial ou chi quadrado de até 2 graus de liberdade, por exemplo, onde a maioria dos registros se encontra nos valores mais baixos.

A distribuição da vazão também pode apresentar um formato de uma distribuição chi quadrado. Porém, para uma afirmação mais acurada, um estudo mais aprofundado deve ser realizado.

A dispersão da variável de entrada chuva com a de saída vazão também apresenta um comportamento interessante. De fato, os gráficos foram elaborados com a precipitação transformada no tempo t . Como a maioria das estações de chuva utilizadas na análise não se encontravam próximas a UHE em questão, ou seja, se localizavam mais a montante, a resposta da entrada não seria no tempo t , mas sim em um tempo $t + i$, dependendo da bacia analisada. No entanto, algumas bacias das UHE analisadas apresentavam uma menor dispersão dos dados da entrada no tempo t em relação à saída no tempo t , como é o caso de Machadinho.

O tempo de resposta também será diferente para cada bacia e UHE, por diversos fatores já supracitados. Isso dependerá também da localização das estações escolhidas para análise. No entanto, tal influência é reduzida se as estações estiverem bem dispersas ao longo da bacia hidrográfica, de maneira que represente bem a distribuição espacial da chuva. O gráfico de correlação cruzada traduz bem o grau de influência que cada defasagem da chuva causa na vazão. Por exemplo, para UHE Machadinho Figura 3.6 é possível extrair que a chuva que cai ao longo da bacia terá impacto dois dias depois e que outras defasagens tem tanto impacto assim. Ou seja, de certa forma, podemos inferir que é uma bacia com concentração rápida, pois obtivemos uma boa dispersão da localização das estações (Figura 3.3). As demais bacias apresentam uma boa dispersão das estações e apresentaram um grau de impacto maior da chuva na defasagem de três dias.

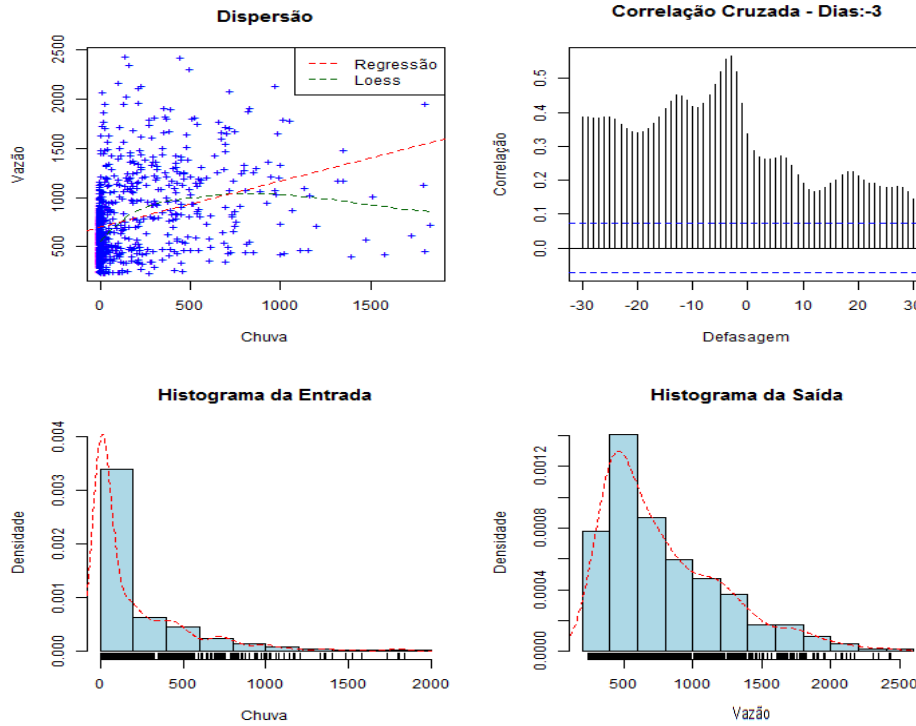


Figura 3.5: Gráficos de auxílio à AED para a UHE Furnas

Fonte: Elaboração própria

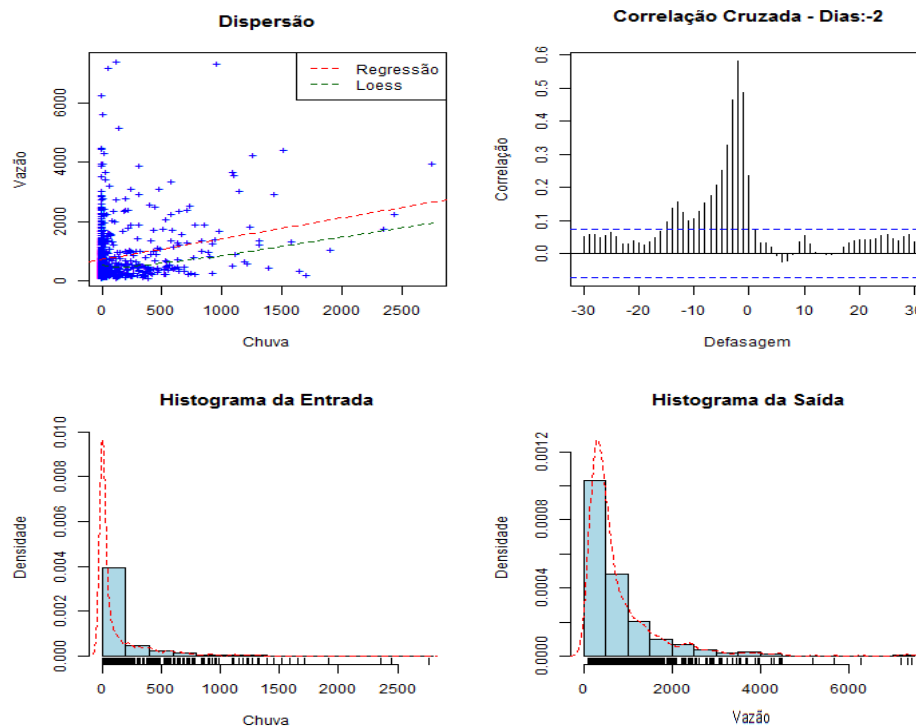


Figura 3.6: Gráficos de auxílio à AED para a UHE Machadinho

Fonte: Elaboração própria

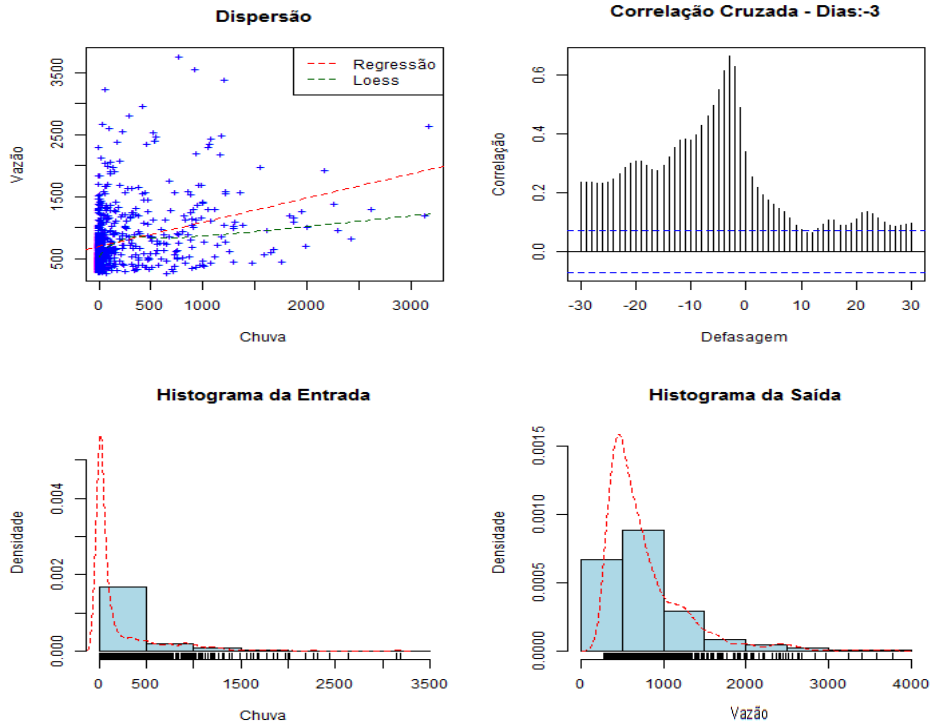


Figura 3.7: Gráficos de auxílio à AED para a UHE Mascarenhas

Fonte: Elaboração própria

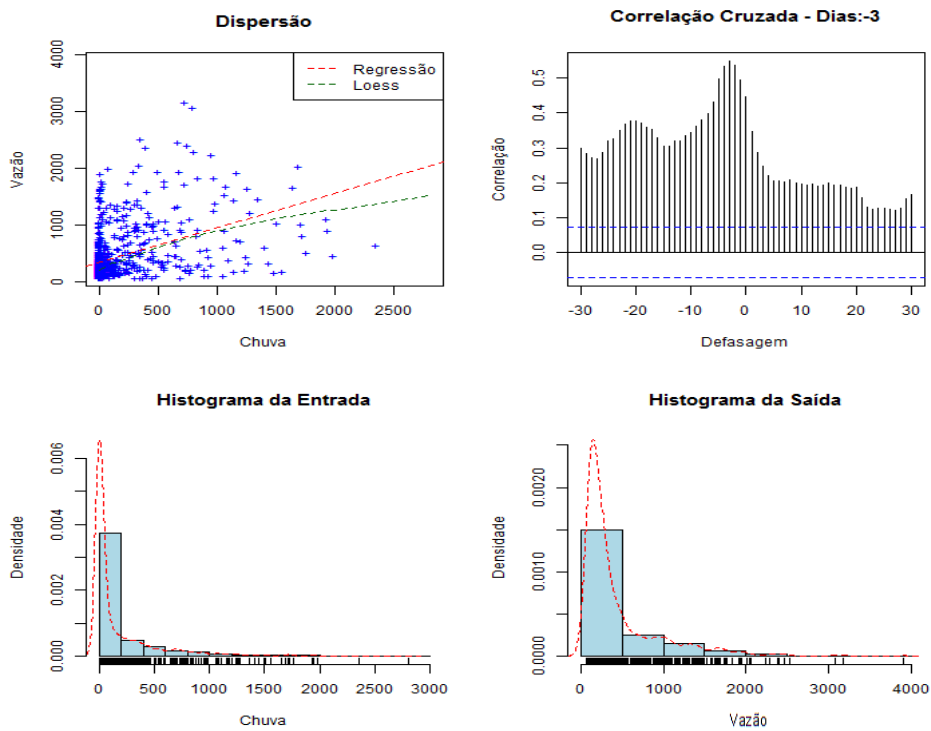


Figura 3.8: Gráficos de auxílio à AED para a UHE Três Marias

Fonte: Elaboração própria

Do conjunto de dados de dez anos, escolheu-se sete anos consecutivos. Os primeiros quatro anos serão o conjunto de treinamento e os dois anos seguintes o conjunto de teste. O objetivo do presente trabalho não é desenvolver um modelo de previsão para eventos extremos. Ao contrário, deseja-se treinar o modelo para prever padrões de maiores probabilidade de ocorrência. Dessa forma, a distribuição dos dados de treinamento e teste devem ser similares. Como exposto anteriormente, a distribuição dos dados diários de vazão são bem assimétricos. Por essa razão, para ter um maior detalhamento, é interessante comparar a distribuição de dados assimétricos aplicando o logaritmo natural aos dados para verificar se há grande discrepância entre os conjuntos de dados de treinamento e teste. Se as distribuições do conjunto de treinamento e testes forem parecidas, isso significa que os dois conjuntos pertencem ao mesmo processo. A Figura 3.9 apresenta tais resultados, o que ratifica que os dois conjuntos podem ser usados para a configuração experimental:

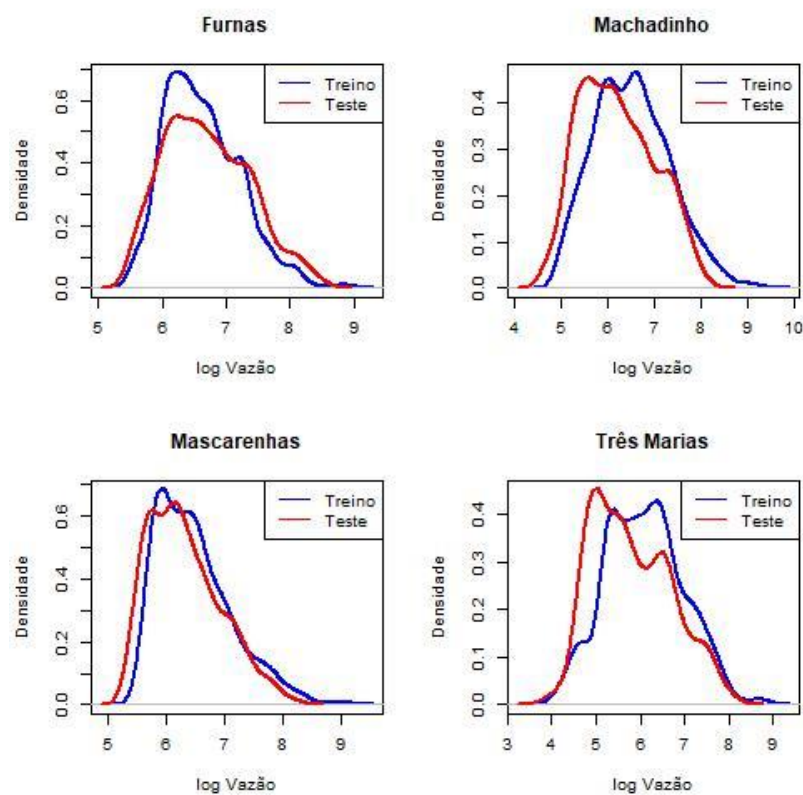


Figura 3.9: Distribuição do logaritmo natural das vazões das UHE analisadas

Fonte: Elaboração própria

4. RESULTADOS E DISCUSSÕES

4.1. MÉTRICAS DE ERRO PARA O DESEMPENHO DOS MODELOS

A avaliação do desempenho dos modelos utilizados no presente trabalho foi realizada de acordo com as seguintes métricas de erro:

- RMSE – Root Mean Squared Error, definido pela Equação 4.1:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1...N} (y(t) - v(t))^2} \quad (4.1)$$

- MAPE – Mean Average Percentage Error, definido pela Equação 4.2:

$$MAPE = \frac{100}{N} \sum_{t=1...N} \left| \frac{y(t) - v(t)}{v(t)} \right| \quad (4.2)$$

- Coeficiente de eficiência Nash-Sutcliffe ou também conhecido como coeficiente de determinação, definido pela Equação 4.3:

$$R^2 = 1 - \frac{\sum_{t=1...N} (y(t) - v(t))^2}{\sum_{t=1...N} (v(t) - \bar{v})^2} \quad (4.3)$$

- Coeficiente de eficiência Nash-Sutcliffe como valores logarítmicos, definido pela Equação 4.4:

$$R_{log}^2 = 1 - \frac{\sum_{t=1...N} (\log(y(t)) - \log(v(t)))^2}{\sum_{t=1...N} (\log(v(t)) - \overline{\log(v)})^2} \quad (4.4)$$

Os termos \bar{v} e $\overline{\log(v)}$ se referem à média da vazão observada e a média do logaritmo natural da vazão observada, respectivamente.

A métrica RMSE representa o erro na mesma escala que da variável de saída vazão (m^3/s) e é mais sensível a erros de valores altos devido ao soma dos quadrados. É útil quando valor do erro é de acordo com uma dada tolerância, que é dependente da aplicação. Uma RMSE menor que 5% a 10 % do valor máximo é considerado como bom ou aceitável desempenho. A métrica MAPE apresenta uma mensuração relativa e é útil para uma avaliação qualitativa da acurácia do modelo. Apesar de também ser dependente da complexidade da aplicação, a métrica MAPE não deve ser maior que 20% ou 30% para ser um modelo aceitável. Os coeficientes de eficiência Nash-Sutcliffe (R^2 e R_{log}^2) representam a qualidade do modelo levando em consideração a variância da variável de saída vazão e deve o mais próximo possível da unidade. O coeficiente Nash-Sutcliffe tradicional, sem valores logarítmicos, é mais sensível aos erros com altos valores da variável de saída vazão, ao passo que, o coeficiente Nash-Sutcliffe com valores logarítmicos é mais sensível a erros de baixos valores. Em geral, o coeficiente Nash-Sutcliffe com resultados iguais a 0,95, 0,85 e 0,75 representam um modelo excelente, bom e aceitável, respectivamente (EVSUKOFF; LIMA; EBECKEN, 2011).

4.2. CONFIGURAÇÃO EXPERIMENTAL

Para cada área de contribuição das UHE analisadas, um período foi escolhido para construir os conjuntos de dados. Para os dois modelos, o conjunto de dados de treinamento apresenta quatro anos de dados de precipitação e vazões diária e o de teste apresenta os dois anos subsequentes do mesmo tipo de dados. Os mesmos conjuntos de treinamento e de testes foram adotados para os dois modelos.

4.2.1. TREINAMENTO - SFR

Para o caso do modelo SFR, o conjunto de treinamento foi dividido em dois: 50% para a escolha da estrutura e identificação de parâmetros e 50% para validação. A escolha da estrutura do modelo se refere a escolha do número de estados ($nsta$), a escolha do número de conjuntos fuzzy para a variável de entrada (npu) e a escolha do número de conjuntos fuzzy para a variável de estados (npx). A escolha da estrutura do modelo nos permite também saber o total de parâmetros a ser identificado, ou seja, o tamanho do vetor θ , descrito na seção 2.5.1. Para a escolha da estrutura e identificação de parâmetros

foram verificadas por exaustão diversas estruturas nos dois primeiros anos do conjunto de treinamento. A modelagem realizada para o SFR se obteve através de um código em Matlab utilizado para acionar um executável desenvolvido na linguagem C com base na explanação descrita em 2.4. Foram analisadas estruturas como:

- $nsta = [1,2,3]$;
- $npu = [2,3,4,5]$; e
- $npv = [2,3,4,5]$.

Cada estrutura foi aplicada, então, ao conjunto de validação e verificou-se aquela que apresentou o menor erro de acordo com a métrica RMSE. Após a escolha da estrutura, o modelo SFR foi retreinado com os quatro anos de dados do conjunto de treinamento completo.

Também é necessário inserir no modelo os protótipos das variáveis de entrada e de estado. No presente trabalho, o vetor de protótipos para as variáveis de entrada terá como o primeiro componente o percentil mínimo ($Percentil_{min}(\mathbf{u})$) e como o último componente o percentil máximo $Percentil_{max}(\mathbf{u})$ de cada uma das variáveis. A quantidade dos componentes intermediários dependerá do npu adotado, de maneira que o espaçamento entre os componentes do vetor se dará através de:

$$Esp = \frac{(Percentil_{max}(\mathbf{u}) - Percentil_{min}(\mathbf{u}))}{(npu - 1)} \quad (4.5)$$

Onde Esp é o espaçamento entre os componentes do vetor de protótipos. Assim, por exemplo, se $npu = 4$, o $Percentil_{min}(\mathbf{u}) = 0\%$ e o $Percentil_{max}(\mathbf{u}) = 100\%$, $Esp = 33.33\%$ e o vetor de protótipos da entrada $px = [Percentil_{min=0}(\mathbf{u}) \text{ } Percentil_{33}(\mathbf{u}) \text{ } Percentil_{66}(\mathbf{u}) \text{ } Percentil_{max=100}(\mathbf{u})]$. O mesmo raciocínio serve para a variável de saída.

Todos os dados foram normalizados a partir do valor máximo de precipitação transformada do conjunto de treinamento, tanto para a variável de entrada quanto para a de estados, para que a informação de proporcionalidade da saída e entrada não fosse perdida.

As estruturas adotadas para cada UHE foram diferentes, de acordo com o resultado do desempenho. Os resultados das estruturas adotadas estão dispostos na Tabela

4.1 e o resultado da performance do treinamento de acordo com a Equação 2.27 se encontra na Tabela 4.2.

Tabela 4.1: Estrutura do SFR para cada bacia das UHE

UHE	nsta	npX	npU
Furnas	1	2	2
Machadinho	1	2	5
Mascarenhas	1	2	2
Três Marias	1	4	2

Fonte: Elaboração própria

Tabela 4.2: Resultado da FO para treinamento do SFR

	UHE			
	Furnas	Machadinho	Mascarenhas	Três Marias
FO	0,35	0,53	0,32	0,41

Fonte: Elaboração própria

4.2.2. CALIBRAÇÃO - SMAP

Já para o modelo SMAP, a calibração foi realizada com os mesmos quatro anos de dados do conjunto de treinamento utilizado para a modelagem do SFR com auxílio da ferramenta *Solver* da Planilha Eletrônica MS Excel. Este otimiza os parâmetros de determinada função dadas as restrições estabelecidas, observando uma função objetivo específica. O resultado da calibração pode ser acompanhado na Tabela 4.3. A inicialização da calibração é estabelecida pelo hidrólogo através do método tentativa-erro, observando, da mesma forma, a função objetivo específica e a forma da hidrógrafa calculada. Os parâmetros de inicialização apresentam significância física e por isso necessitam da experiência do hidrólogo para serem estabelecidos. O parâmetros de inicialização podem ser verificados na Tabela 4.4. Ao final, os resultados da calibração podem ser acompanhados na Tabela 4.5.

Tabela 4.3: Resultado da calibração do modelo SMAP

Parâmetros	UHE			
	Furnas	Machadinho	Mascarenhas	Três Marias
Ai	2,00	2,00	2,00	2,00
Str	65,00	65,00	150,01	239,99
Capc	20,00	22,00	24,98	50,00
Crec	20,00	3,99	6,09	4,96
K2t	8,00	2,80	5,98	8,00
Kkt	120,00	46,80	180,00	90,00
kep	1,12	1,20	1,03	1,20
kt(-3)	0,17	0,03	0,00	0,00
kt(-2)	0,14	0,00	0,00	0,05
kt(-1)	0,18	0,40	0,74	0,36
kt(0)	0,11	0,39	0,17	0,31
kt(+1)	0,40	0,18	0,09	0,28

Fonte: Elaboração própria

Tabela 4.4: Inicialização do modelo SMAP para o conjunto de treinamento

Parâmetros	UHE			
	Furnas	Machadinho	Mascarenhas	Três Marias
Tuin	100,00	100,00	50,00	0,00
Ebin	200,00	70,00	330,00	355,00
Supin	4,59	23,80	4,60	6,02

Fonte: Elaboração própria

Tabela 4.5: Resultado da FO para calibração do SMAP

FO	UHE			
	Furnas	Machadinho	Mascarenhas	Três Marias
	0,19	0,29	0,17	0,32

Fonte: Elaboração própria

4.2.3. TESTE DOS MODELOS

O teste dos modelos visa simular a operação real da previsão de vazão 10 dias a frente desenvolvida pelo ONS. No entanto, a metodologia de previsão depende de modelo para modelo.

Para a realizar a previsão de vazão utilizando o modelo SMAP, é necessário obter dados de precipitação e vazão diárias de 31 dias prévios ao primeiro dia de previsão. Esse conjunto de dados serve para inicializar o SMAP utilizando os parâmetros *Tuin*, *Ebin* e *Supin* observando algum tipo de função objetivo.

O conjunto de testes da variável de entrada foi transformado em uma matriz de 44 colunas por 686 onde a primeira linha são os primeiros 44 dados de entrada. A segunda linha também apresenta 44 dados, porém essa começará a partir do segundo dia de dados até o quadragésimo quinto dia e assim por diante. Ao final, a matriz de entrada será da seguinte forma:

$$\begin{array}{cccccc}
 u(t + 1) & u(t + 2) & \cdots & u(t + 43) & u(t + 44) & \\
 u(t + 2) & u(t + 3) & \cdots & u(t + 44) & u(t + 45) & \\
 \vdots & \vdots & \ddots & \vdots & \vdots & \\
 u(t + 685) & u(t + 686) & \cdots & u(t + 727) & u(t + 728) & \\
 u(t + 686) & u(t + 687) & \cdots & u(t + 728) & u(t + 729) &
 \end{array}$$

Figura 4.1: Matriz de entrada dos modelos

Fonte: Elaboração própria

As três primeiras colunas da matriz servem para calcular a chuva média de todos os pontos e seus respectivos pesos levando também em consideração as chuvas de três dias anteriores. As 31 colunas seguintes são utilizadas para a inicialização do modelo com a adoção dos parâmetros *Tuin*, *Ebin* e *Supin*. Normalmente, a escolha de tais parâmetros é realizada por um hidrólogo através do método tentativa-erro, observando uma função objetivo. No entanto, como cada linha da matriz será a simulação da operação real de previsão e são ao todo 686 linhas, resolveu-se automatizar o processo de escolha dos parâmetros de inicialização através de um algoritmo de otimização. Para tal, escolheu-se um algoritmo de inteligência artificial chamado de Evolução Diferencial, por ser um algoritmo de otimização heurística, e que por isso, de certa forma, se aproxima de um método tentativa erro, como usado originalmente. O algoritmo pode ser utilizado através do pacote DEoptim desenvolvido na linguagem de programação R. A função objetivo utilizada foi a mesma da equação 2.27. O pacote foi utilizado com os parâmetros *default*, sendo somente adotados o número da população (NP) igual a 30 e o número máximo de iterações igual a 250. Mais detalhes sobre o algoritmo podem ser encontrados em (ARDIA *et al.*, 2011, 2016; MULLEN *et al.*, 2011). Por fim, as colunas 35 a 44 são utilizadas como entrada para realizar a previsão de vazão.

A inicialização do modelo SFR é mais simples e dependerá do número de estados (*nsta*) da estrutura do modelo. Se *nsta* for igual a 1, por exemplo, a inicialização do modelo será realizada com o primeiro valor da variável de saída antes do primeiro dia de previsão. Já se *nsta* for igual a 2, a inicialização necessitará dois primeiros valores da variável de saída, antes do primeiro dia de previsão e assim por diante. Como todas as estruturas dos modelos realizados para cada UHE apresentaram *nsta* = 1, somente o primeiro valor da variável de saída vazão antes do primeiro dia de previsão será utilizado. Para que a comparação entre os modelos seja justa, é necessário que os modelos sejam testados com as mesmas condições e sejam usados os mesmos dados. Por isso, somente as colunas 35 a 44 serão usadas como entrada para o modelo SFR.

O presente trabalho visa reproduzir a operação de previsão de vazão 10 dias a frente. A operação de previsão de vazão utiliza dados de previsão de chuva como entrada, os quais podem apresentar erros significativos, devido aos erros de predição dos modelos meteorológicos. Com isso, para dar maior credibilidade aos modelos, foram realizadas 10 simulações da variável de entrada precipitação com base no conjunto de teste acrescentando um erro. Dessa forma, para cada conjunto de variáveis de entrada das áreas de contribuição de cada UHE, foi simulada uma nova matriz de entrada. A simulação dos dados de entrada seguiu uma distribuição normal de média igual à média de precipitação dos trinta dias anteriores para determinado posto de chuva e desvio padrão igual o desvio padrão de precipitação dos trinta dias anteriores para determinado posto multiplicado por um erro de 20%. Ou seja, a simulação da precipitação segue a Equação 4.6:

$$\hat{P}(t) \sim N\left(\frac{1}{30} \sum_{t=-30}^0 P(t), \sqrt{\frac{\sum_{t=-30}^0 (P(t) - \bar{P})^2}{29}} \cdot 1,2\right) \quad (4.6)$$

Sendo que $\hat{P}(t) = \max(0, \hat{P}(t))$, pois não existe valor negativo de chuva.

4.3. AVALIAÇÃO DO DESEMPENHO DOS MODELOS

4.3.1. APLICAÇÃO NO CONJUNTO DE TESTES COMPLETO

Para começar a análise do desempenho dos modelos SMAP e SFR é interessante aplica-los aos dados do conjunto de testes completo. A Tabela 4.6 e a Tabela 4.7, mostram os resultados das métricas de erro para a chuva simulada e observada como entrada para os dois modelos. Da Figura 4.2 a Figura 4.9, estão dispostos os gráficos dos resultados da modelagem de vazão das UHEs testadas no presente trabalho, tanto para chuva observada (em laranja) utilizada como entrada, como para chuva simulada (em vermelho).

Para as UHEs Furnas e Três Marias, a saída do modelo SMAP apresentou melhor desempenho que o SFR quando utilizada a chuva observada como entrada. No entanto, quando a chuva simulada foi utilizada na entrada, houve queda significativa de rendimento para os dois modelos, porém o modelo SMAP apresentou maior queda de rendimento que o SFR. Por exemplo, analisando a métrica NSE, para UHE Furnas utilizando o modelo SFR, o resultado passou de 0,82 com a chuva observada para 0,70 com a chuva simulada. No entanto, com o modelo SMAP, o resultado passou de 0,90 com a chuva observada, ou seja, um resultado melhor que SFR, para 0,43 com a chuva simulada. Tendo outra métrica de erro como exemplo, a UHE Furnas com o modelo SFR, a métrica log NSE passou de 0,73 para 0,68, para chuva observada e simulada, respectivamente, uma queda de 0,05. Para o modelo SMAP, a mesma métrica passou de 0,91 para 0,01, uma queda de 0,90. Tal resultado dá indícios que os erros em uma possível previsão meteorológica da chuva, representada aqui pela chuva simulada, podem afetar diferentemente os modelos e que o SMAP é mais sensível à possíveis erros de predição da chuva que o SFR, pois a queda de desempenho do SMAP é superior que a do SFR. Ainda, os erros da chuva simulada impactaram mais o modelo SMAP na métrica log NSE que na NSE. Com isso, podemos compreender que a chuva simulada impactou mais as saídas de baixa vazão que as de alta. Essa mesma reflexão serve também para a UHE Três Marias. Para as UHEs Machadinho e Mascarenhas, os resultados da chuva simulada para o modelo SMAP ainda foram melhores que para o SFR, porém, a queda de desempenho dos modelo SMAP para todas as métricas foi superior que o SFR. A Tabela 4.8 mostra a queda de desempenho dos modelos para cada métrica utilizada.

Tabela 4.6: Resultados do SFR para as métricas de erro na aplicação do conjunto de testes completo

UHE	RMSE	MAPE	NSE	log NSE
Chuva Observada				
Furnas	334,51	31,69	0,82	0,73
Machadinho	372,68	60,76	0,61	0,53
Mascarenhas	223,40	30,46	0,82	0,72
Três Marias	273,56	58,68	0,74	0,65
Chuva Simulada				
Furnas	436,21	37,80	0,70	0,68
Machadinho	412,17	72,33	0,52	0,43
Mascarenhas	305,24	36,23	0,67	0,66
Três Marias	342,73	78,17	0,59	0,50

Fonte: Elaboração própria

Tabela 4.7: Resultados do SMAP para as métricas de erro na aplicação do conjunto de testes completo

UHE	RMSE	MAPE	NSE	log NSE
Chuva Observada				
Furnas	255,82	16,10	0,90	0,91
Machadinho	233,31	30,90	0,85	0,81
Mascarenhas	189,80	13,99	0,87	0,92
Três Marias	215,73	46,69	0,84	0,75
Chuva Simulada				
Furnas	602,77	48,78	0,43	0,01
Machadinho	449,24	68,70	0,43	0,49
Mascarenhas	281,52	33,80	0,72	0,66
Três Marias	520,99	118,34	0,06	0,19

Fonte: Elaboração própria

Tabela 4.8: Queda de desempenho dos modelos

UHE	RMSE	MAPE	NSE	log NSE
SMAP				
Furnas	346,95	32,68	0,47	0,90
Machadinho	215,92	37,80	0,42	0,32
Mascarenhas	91,72	19,81	0,15	0,25
Três Marias	305,26	71,64	0,78	0,56
SFR				
Furnas	101,70	6,11	0,12	0,05
Machadinho	39,48	11,57	0,09	0,10
Mascarenhas	81,84	5,76	0,15	0,07
Três Marias	69,16	19,49	0,15	0,15

Fonte: Elaboração própria

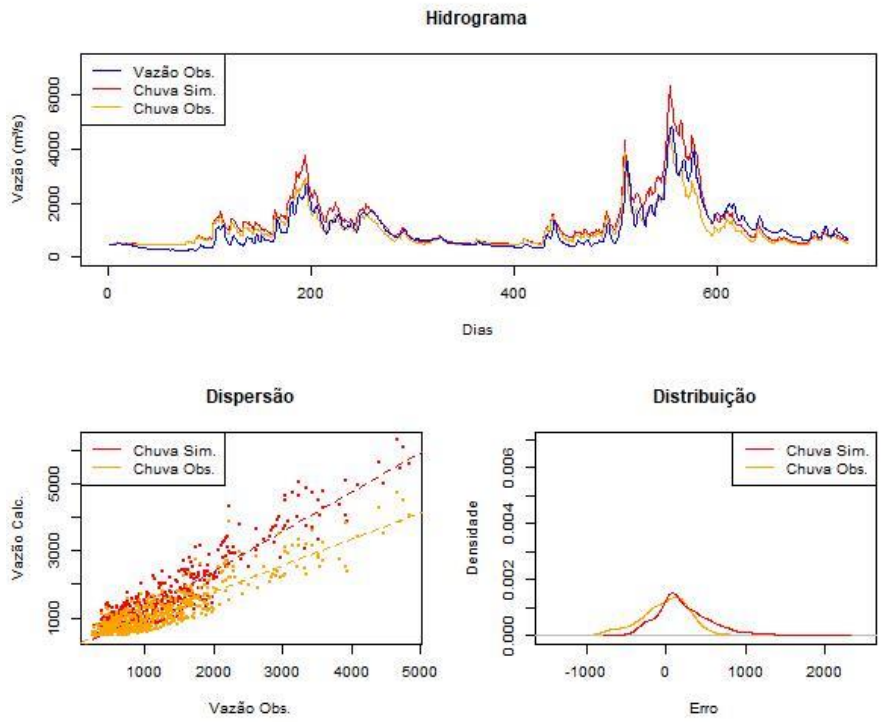


Figura 4.2: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Furnas

Fonte: Elaboração própria

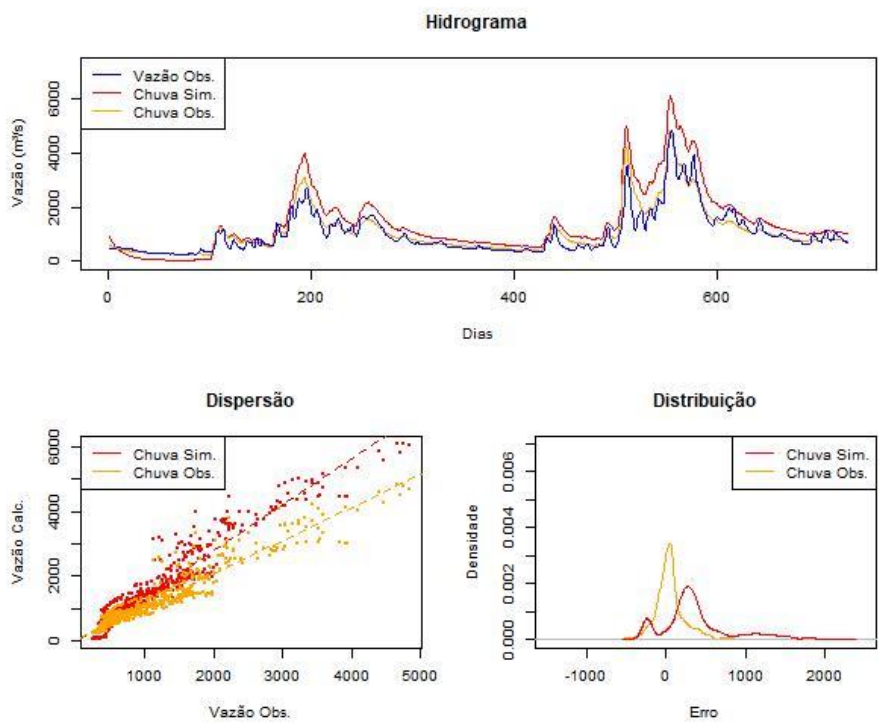


Figura 4.3: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Furnas

Fonte: Elaboração própria

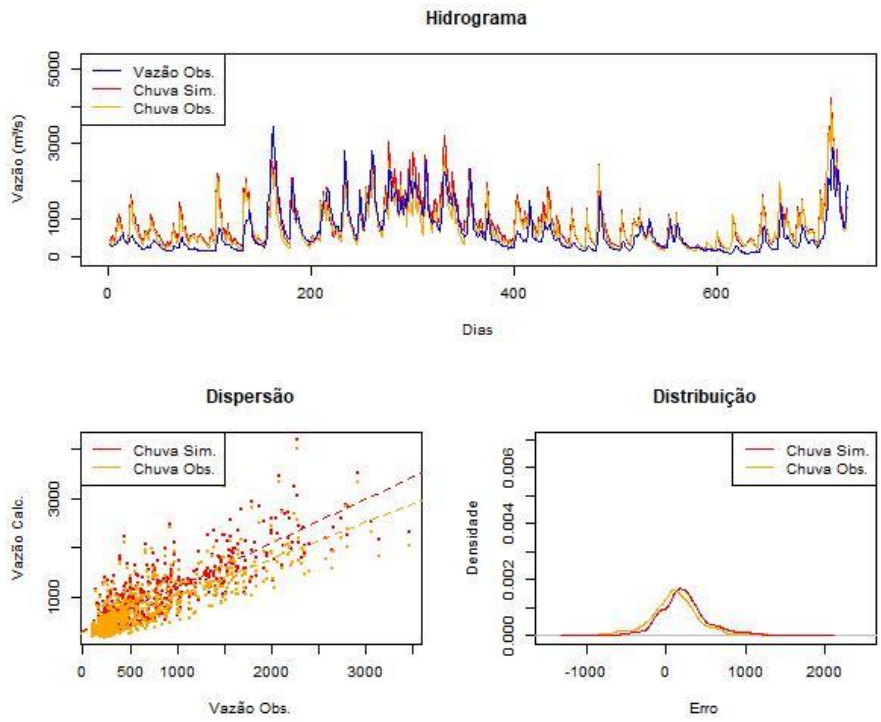


Figura 4.4: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Machadinho

Fonte: Elaboração própria

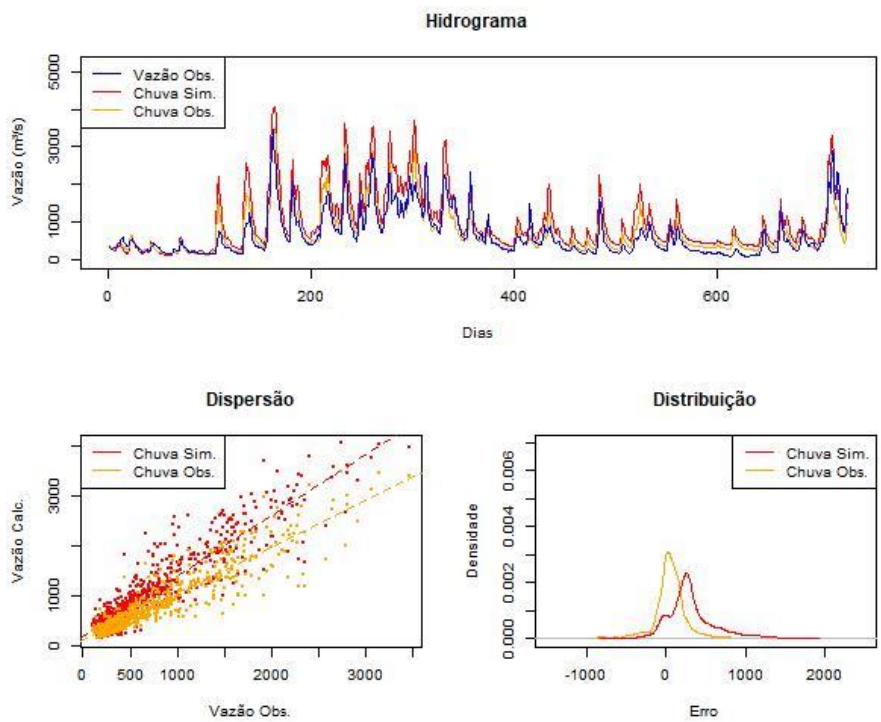


Figura 4.5: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Machadinho

Fonte: Elaboração própria

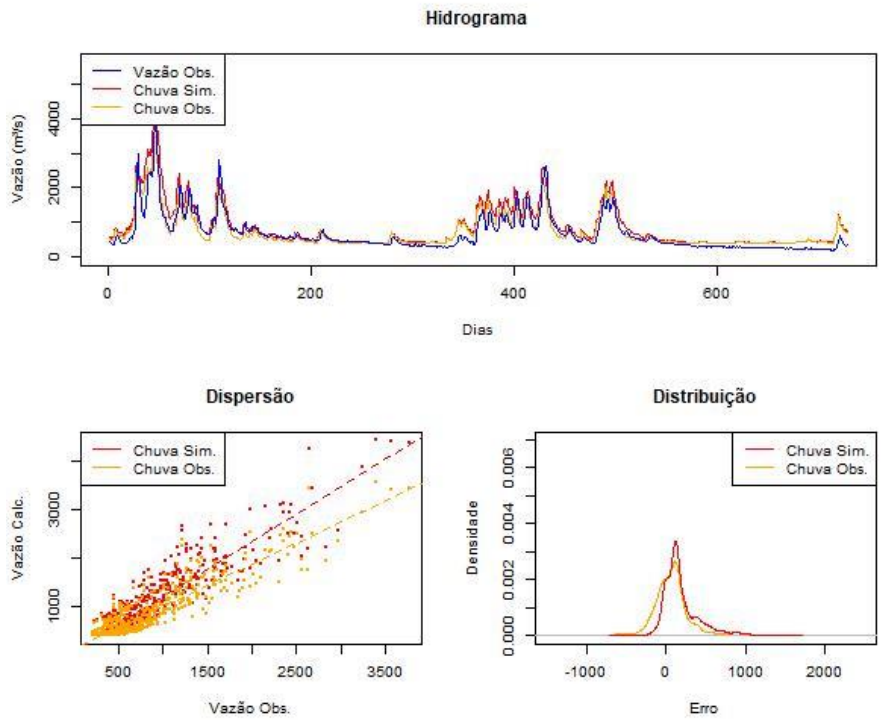


Figura 4.6: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Mascarenhas

Fonte: Elaboração própria

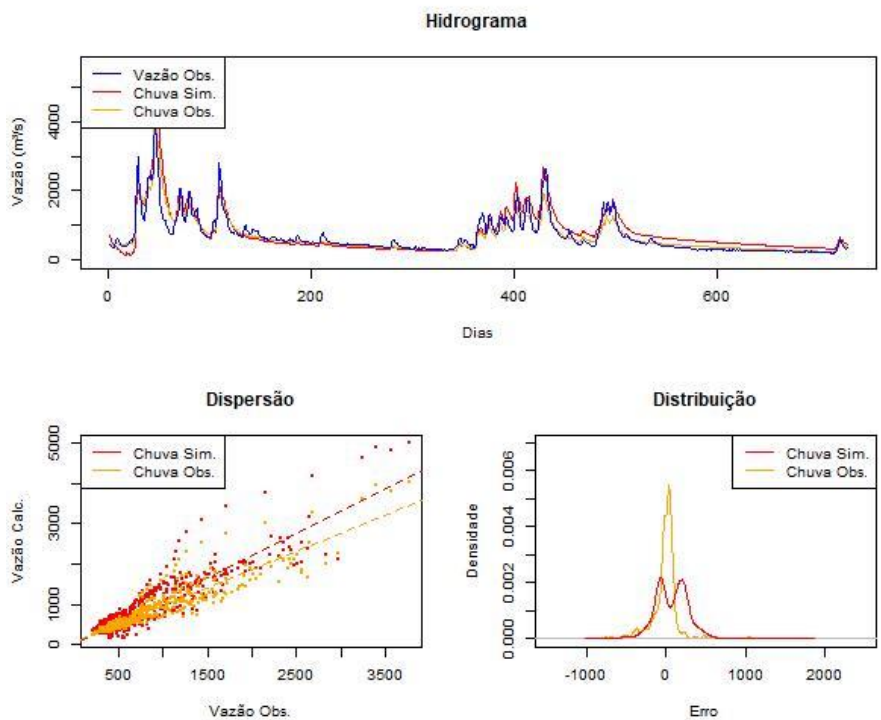


Figura 4.7: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Mascarenhas

Fonte: Elaboração própria

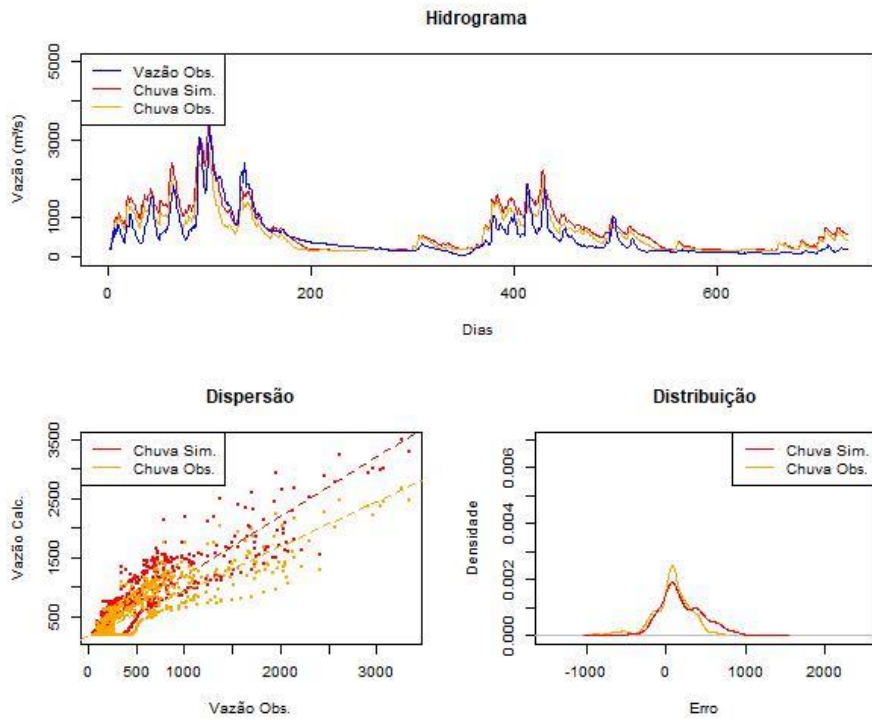


Figura 4.8: Hidrograma, dispersão e distribuição dos erros do modelo SFR para UHE Três Marias

Fonte: Elaboração própria

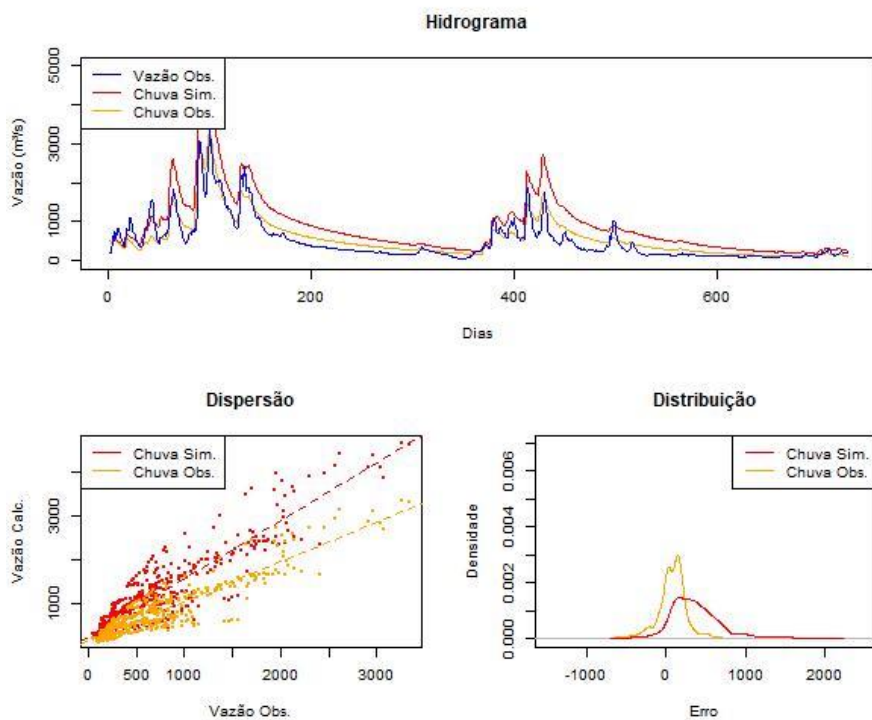


Figura 4.9: Hidrograma, dispersão e distribuição dos erros do modelo SMAP para UHE Três Marias

Fonte: Elaboração própria

4.3.2. SIMULAÇÃO DA OPERAÇÃO REAL DE PREVISÃO DE VAZÃO PARA 10 DIAS À FRENTE

Para esta etapa do teste, utilizou-se o procedimento descrito na seção 4.2.3, onde o vetor de entrada do conjunto de teste foi transformado em uma matriz, tanto para chuva observado como para simulada. Em tal matriz cada linha é a entrada de um horizonte de previsão de dez dias à frente. Esse teste é interessante para verificar o desempenho de modelos para diversos horizontes de previsão, pois determinado modelo pode ser exemplar para prever um dia à frente, por exemplo, porém seu desempenho cai significativamente para horizontes maiores. Por outro lado, outros modelos podem ser piores para prever horizontes menores e é compensado pelo desempenho em horizontes mais distantes. Para cada situação, determinados modelos são mais aconselháveis que outros.

É racional ponderar a hipótese que quanto maior o horizonte de previsão, menor será o desempenho para cada modelo. Dessa forma, para o exemplo do presente trabalho, o desempenho de um modelo para previsão de 10 dias à frente deverá ser inferior ao desempenho do mesmo modelo para um dia à frente. Tal fato se dá, pois todo modelo, por ser uma simplificação do processo de um sistema real, apresenta um erro intrínseco na saída. Assim, por exemplo, num horizonte de dois dias, este apresentará um resultado no qual terá o acúmulo do erro de previsão de um dia à frente somado ao erro de previsão do segundo dia e assim por diante.

Outro ponto razoável a se cogitar é que possíveis erros na variável de entrada em relação aos dados observados também vão refletir em uma queda de desempenho nos modelos, porém cada modelo vai responder diferentemente. O intuito de utilizar uma chuva simulada baseada na chuva observada somada a um erro é exatamente verificar a resposta de cada modelo para possíveis desvios na variável de entrada. Tal teste visa simular a operação real de previsão de vazão do ONS, pois o mesmo utiliza como entrada a previsão de chuva que é a saída de um modelo de previsão meteorológica e que, por ser também uma simplificação de um sistema real, apresenta erros na saída.

O desempenho dos modelos SMAP e SFR quando aplicados aos dados da UHE Furnas, de certa forma, atendem ao supracitado. De acordo com a Figura 4.10, o modelo SFR começa com um desempenho melhor nos horizontes mais curtos e vai aumentando a cada horizonte. Ainda, há uma divergência no desempenho entre a chuva observada e simulada utilizada como entrada, mais observada nos horizontes mais longínquos, onde

a saída utilizando a chuva simulada apresentou resultados piores. O modelo SMAP também obedeceu às expectativas expostas anteriormente, excetuando somente no fato que houve uma melhora no desempenho do modelo em relação ao último horizonte de previsão. Tal comportamento pode estar relacionado ao processo de simulação da chuva como variável de entrada, o qual talvez deva ser melhor representado.

Na comparação entre os modelos, o resultado não foi conclusivo. O SFR apresentou melhor desempenho nas métricas RMSE e NSE, ao passo que o SMAP foi melhor no MAPE e no log NSE. As métricas RMSE e NSE são mais sensíveis aos erros com altos valores da variável de saída vazão e o MAPE e log NSE são mais sensíveis a erros de baixos valores, como exposto na seção 4.1. Dessa forma, pode-se inferir que o modelo SFR teve melhor de acerto nos valores de baixa vazão e o SMAP nos valores altos. Tal resultado pode ser verificado nas Figura 4.11 a Figura 4.14. Também pode ser observado nas mesmas figuras, que a dispersão e a distribuição dos erros das saídas com chuvas simuladas e observadas são bem próximas entre si no horizonte de previsão de um dia a frente para os dois modelos (Figura 4.11 e Figura 4.12). No entanto, quando verificado no horizonte de previsão de dez dias à frente, há uma dispersão dos dados entre as chuvas simuladas e observada (Figura 4.13 e Figura 4.14).

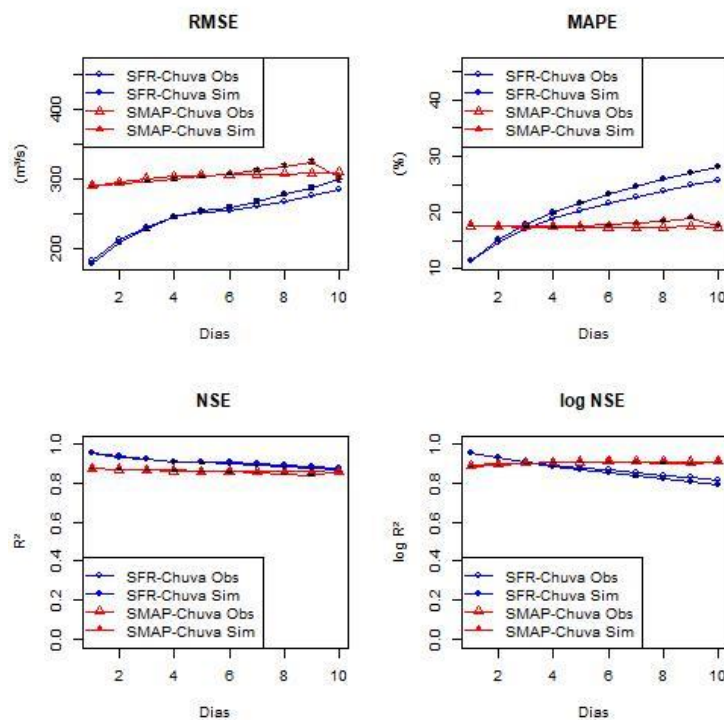


Figura 4.10: Resultados das Métricas de Erros para a UHE Furnas

Fonte: Elaboração própria

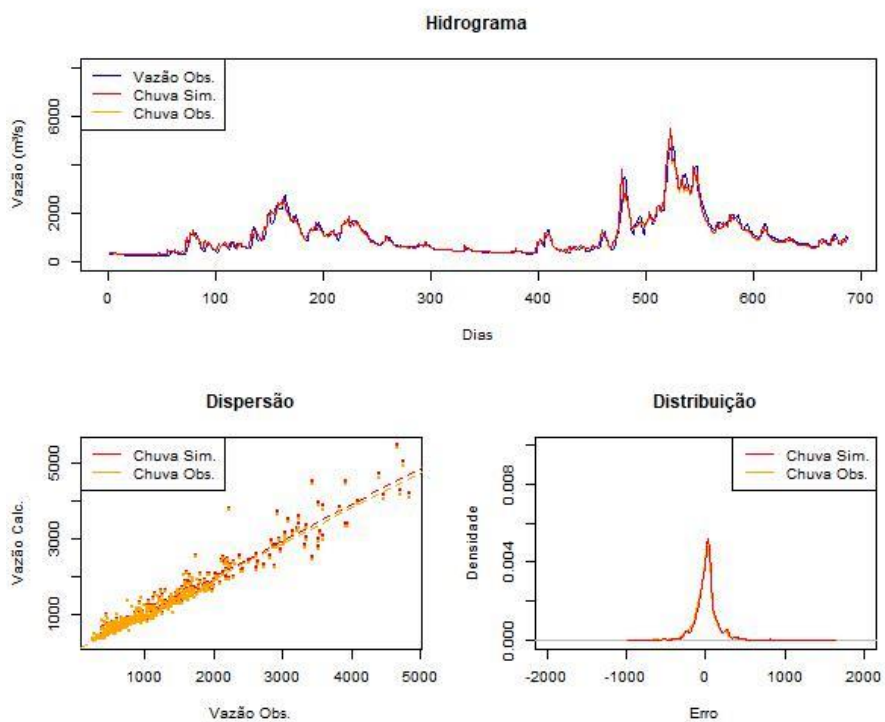


Figura 4.11: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SFR para 1 dia à frente

Fonte: Elaboração própria

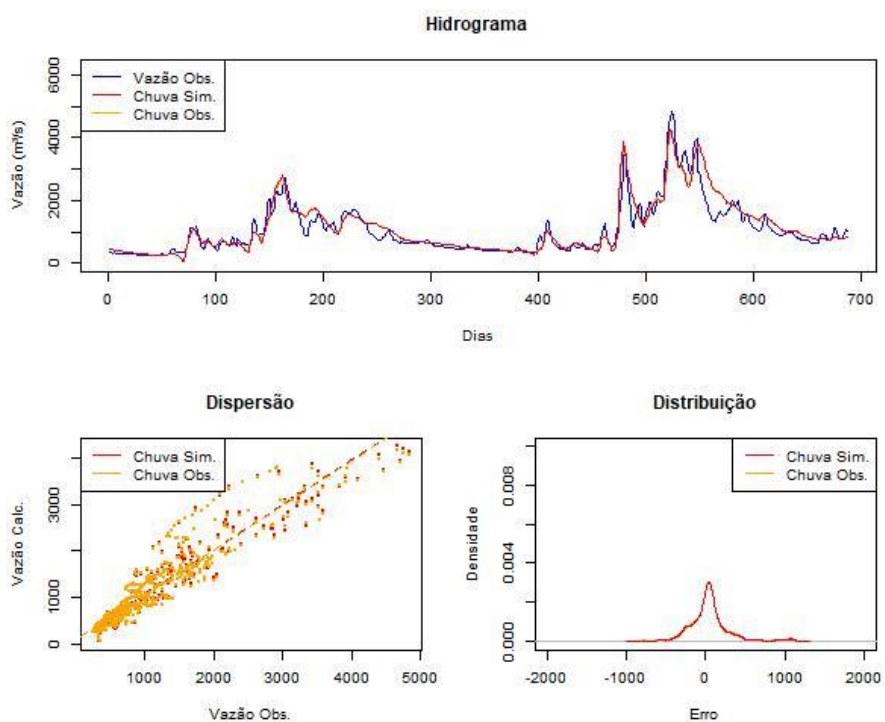


Figura 4.12: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SMAP para 1 dia à frente

Fonte: Elaboração própria

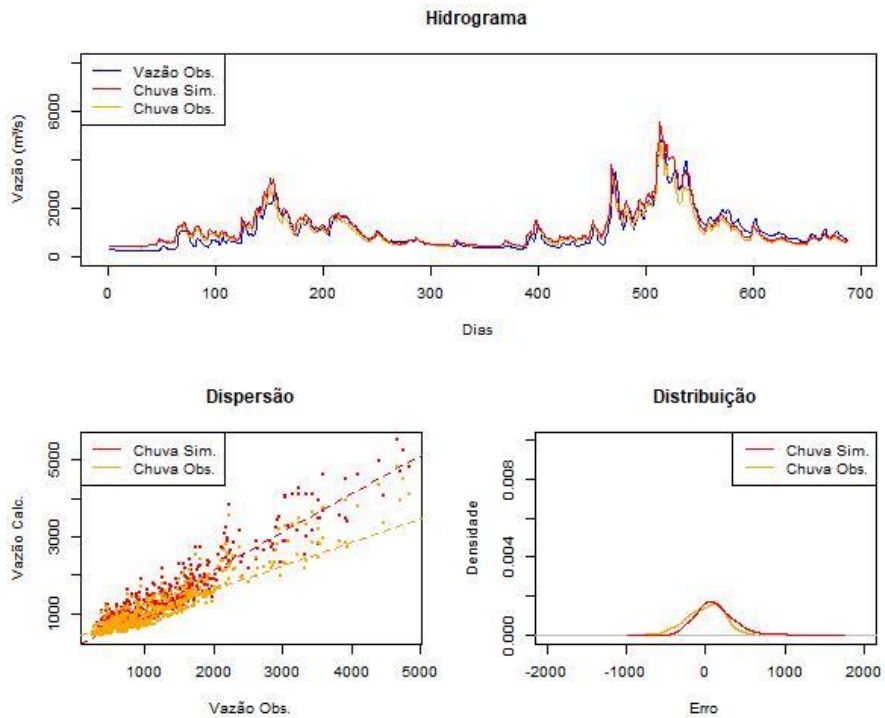


Figura 4.13: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SFR para 10 dias à frente

Fonte: Elaboração própria

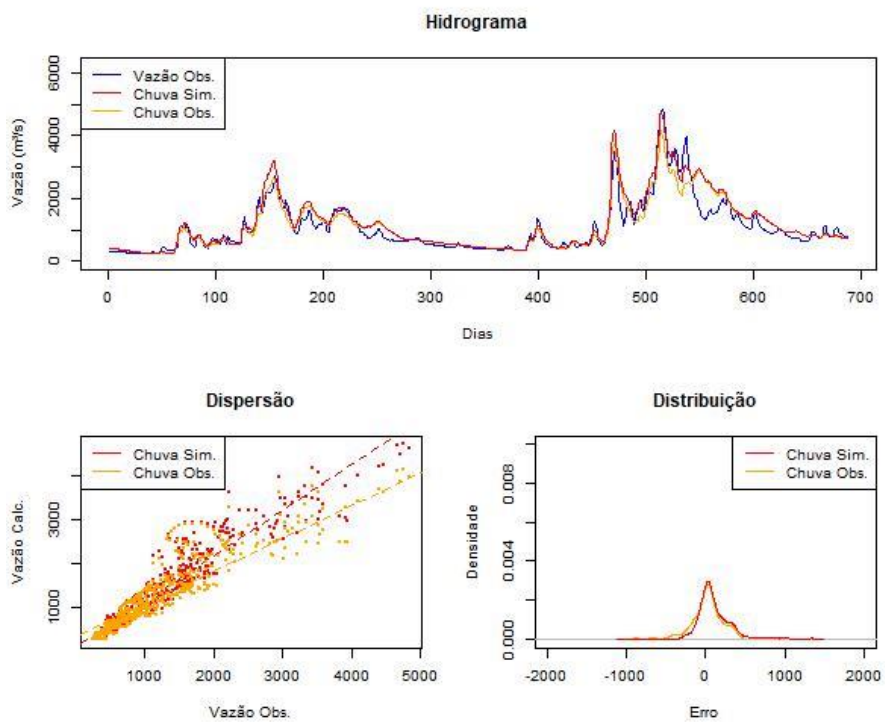


Figura 4.14: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Furnas com o SMAP para 10 dias à frente

Fonte: Elaboração própria

Analisando os resultados aplicados à UHE Machadinho, estes também seguem às expectativas levantadas no início da seção, excetuando o fato que há uma melhora de desempenho do modelo SMAP no décimo dia de previsão em relação ao nono utilizando a chuva simulada como entrada, como na análise da UHE Furnas (Figura 4.15).

O modelo SFR apresentou dificuldades para a modelagem com os dados da UHE Machadinho. O modelo SFR, não apresentou bons resultados, nem para chuva observada nem pra a simulada. Seu desempenho ultrapassa 60% na métrica MAPE e permanece em torno de 0,6 nas métricas NSE e log NSE para o horizonte de previsão de 10 dias à frente. Tais resultados, como explicitado na seção 4.1, não correspondem a um bom desempenho como modelo de previsão o que causa estranheza, pois o SFR é um modelo utilizado hoje oficialmente pelo ONS para a previsão de vazão na bacia do rio Iguaçu. Esta é vizinha à bacia do rio Uruguai onde se localiza a UHE Machadinho e apresenta características semelhantes às da bacia do rio Iguaçu em relação ao sistema chuva-vazão. Um fato importante, que não foi levado em consideração na modelagem do SFR neste trabalho é em relação a análise do relevo. Determinada bacias apresentam questões do relevo que podem interferir significativamente no sistema-chuva vazão. No caso da UHE de Machadinho, talvez se a bacia fosse separada em outras de menor área e um modelo SFR fosse desenvolvido para cada área, talvez os resultados poderiam melhorar bastante.

No entanto, o modelo SMAP obteve resultados superiores comparados ao modelo SFR. Tal afirmação é de certa forma surpreendente, pois o SMAP é um modelo linear, ao contrário do SFR. Para todas as métricas, este apresentou certa estabilidade nos horizontes de previsão analisando a saída com a chuva observada como entrada. Para a chuva simulada como entrada, houve queda de desempenho do modelo ao longo dos horizontes de previsão. A diferença entre o desempenho dos modelos pode ser verificada também nos gráficos de dispersão e distribuição dos erros da Figura 4.16 à Figura 4.19. A dispersão e distribuição dos erros para os dois modelos é similar quando analisado o horizonte de previsão de um dia a frente. Porém, no horizonte de dez dias, a dispersão dos erros de modelo SFR é significativamente maior que a dispersão dos erros do SMAP.

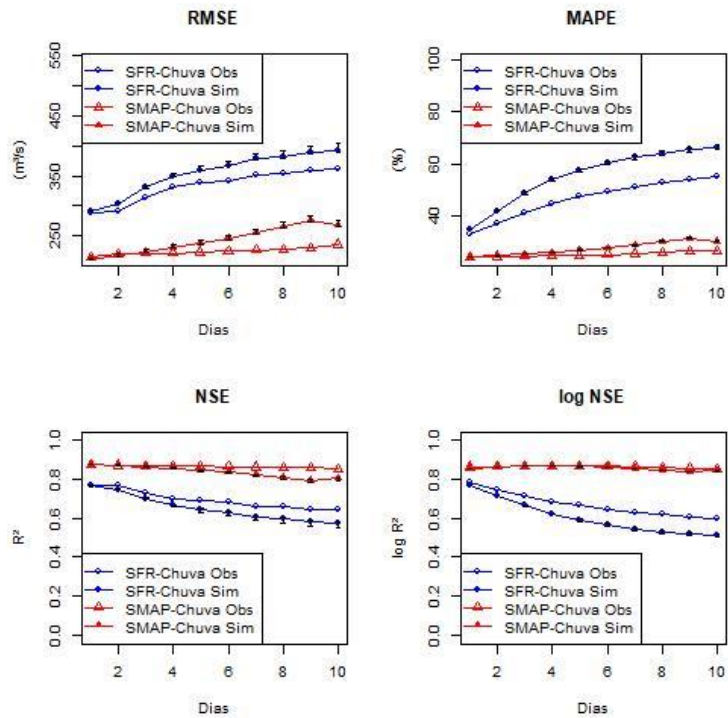


Figura 4.15: Resultados das Métricas de Erros para a UHE Machadinho

Fonte: Elaboração própria

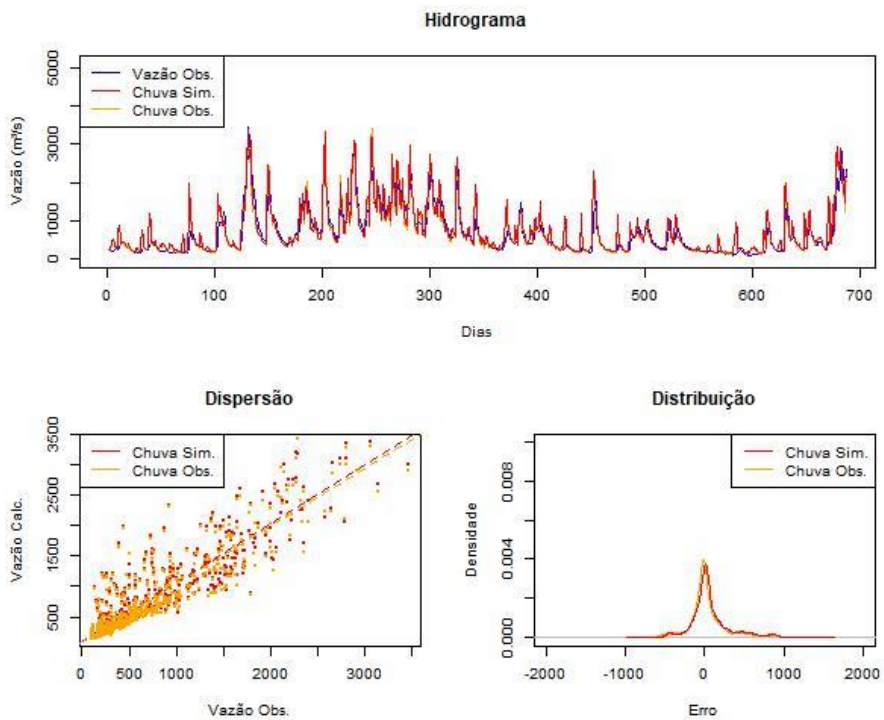


Figura 4.16: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SFR para 1 dia à frente

Fonte: Elaboração própria

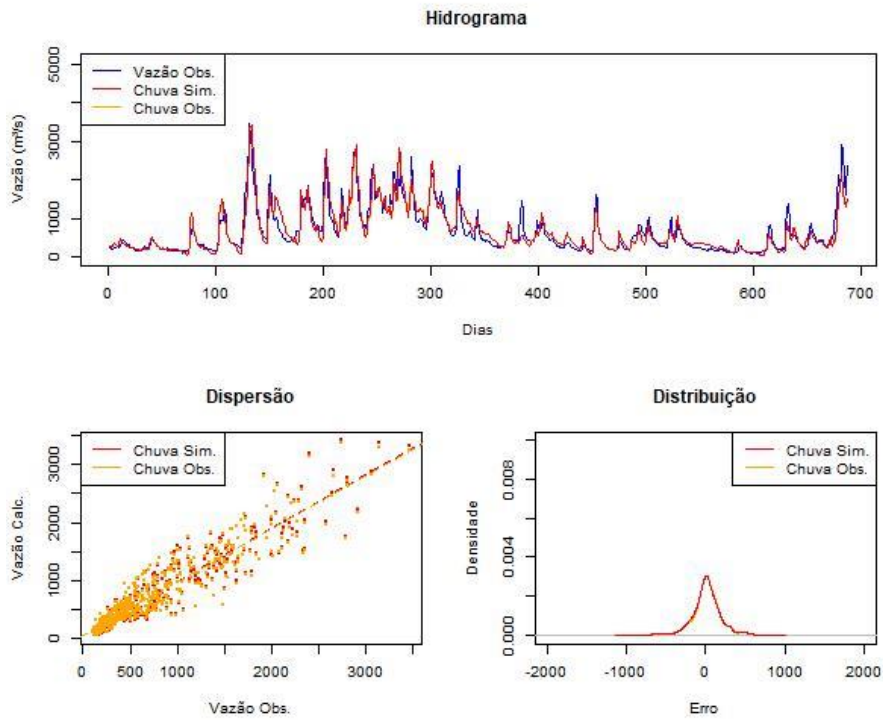


Figura 4.17: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SMAP para 1 dia à frente

Fonte: Elaboração própria

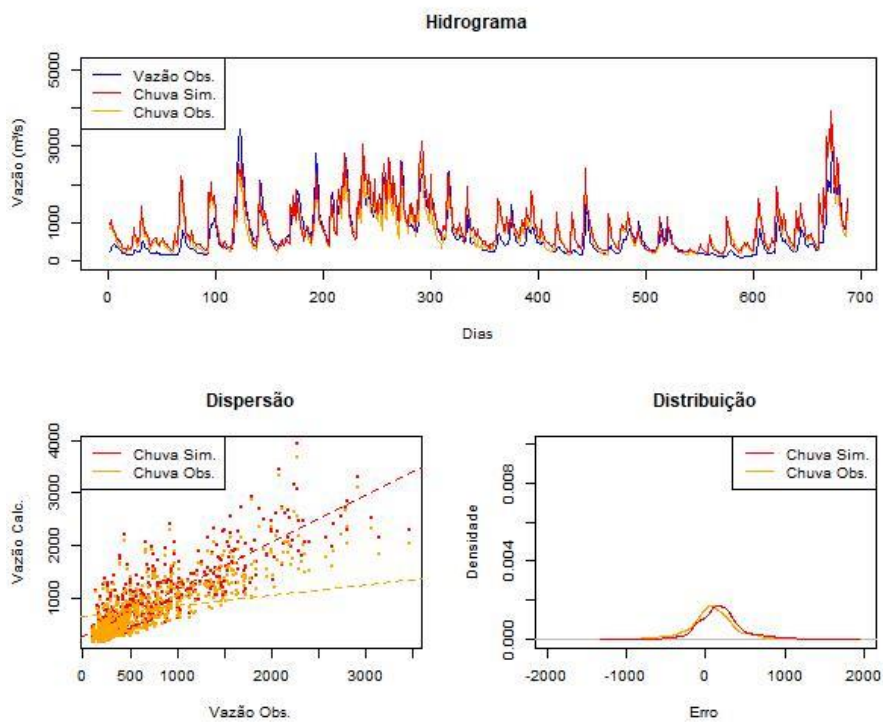


Figura 4.18: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SFR para 10 dias à frente

Fonte: Elaboração própria

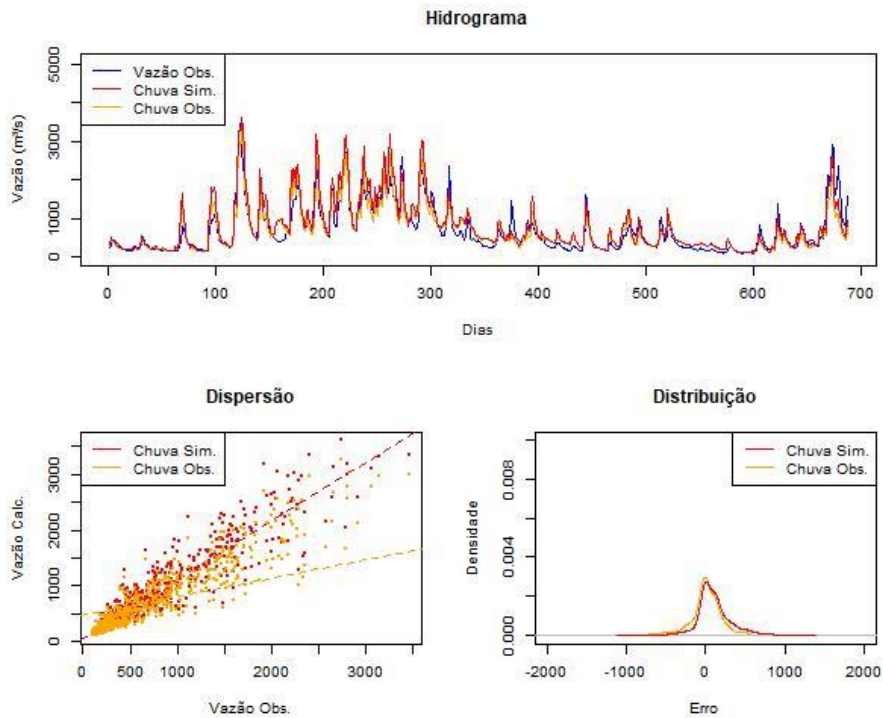


Figura 4.19: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Machadinho com o SMAP para 10 dias à frente

Fonte: Elaboração própria

Observando a aplicação aos dados da UHE Mascarenhas, os resultados do modelo SFR atenderam às expectativas, onde o desempenho do modelo piora ao longo do horizonte de previsão. Utilizando a chuva observada como entrada, o desempenho do modelo se reduz de forma mais acentuada nos três primeiros horizontes de previsão, ao passo que a partir do quarto dia, há uma estabilidade do desempenho, apesar de ainda ser crescente com o aumento do horizonte de previsão. Utilizando a chuva simulada como entrada, a redução do desempenho do modelo também é verificada com o aumento do horizonte de previsão, porém a partir do terceiro dia a curva de desempenho da chuva simulada começa a descolar da curva de desempenho da chuva observada. Em relação ao RMSE, a diferença entre as duas curvas é de aproximadamente $40 \text{ m}^3/\text{s}$ no décimo dia e, em relação ao MAPE, a diferença é de aproximadamente 2% no mesmo dia de previsão.

Já, verificando o modelo SMAP, os resultados não atenderam tanto à expectativa mencionadas no início da presente seção. O modelo apresentou certa estabilidade no desempenho com o crescimento do horizonte de previsão. Para o desempenho utilizando a chuva observada, observando o gráfico RMSE da Figura 4.20, a curva de certa forma aparenta crescer sutilmente até o sexto dia e começa a decrescer da mesma forma até o

décimo numa forma parabólica. No entanto, com a chuva simulada utilizada como entrada, o modelo tem um desempenho quase constante e pouco melhor que com a chuva observada. Tal resultado causa estranheza, pois o que se pode concluir é que o erro inserido na chuva melhorou o desempenho do modelo, o que não deveria se reproduzir.

Observando os resultado de previsão para um dia à frente nas Figura 4.21 e Figura 4.22, verifica-se que os gráficos de distribuição do erro de ambos os modelos é bem afunilado no valor zero, mostrando que os dois modelos foram bem acurados, identificando bem a dinâmica do sistema. No entanto, quando o horizonte de dez dias à frente é analisado, a curva de distribuição do erro para o modelo SFR fica mais achatada enquanto que o SMAP se mostra mais resiliente e permanece com a curva de distribuição do erro bem afunilada e centrada no valor zero.

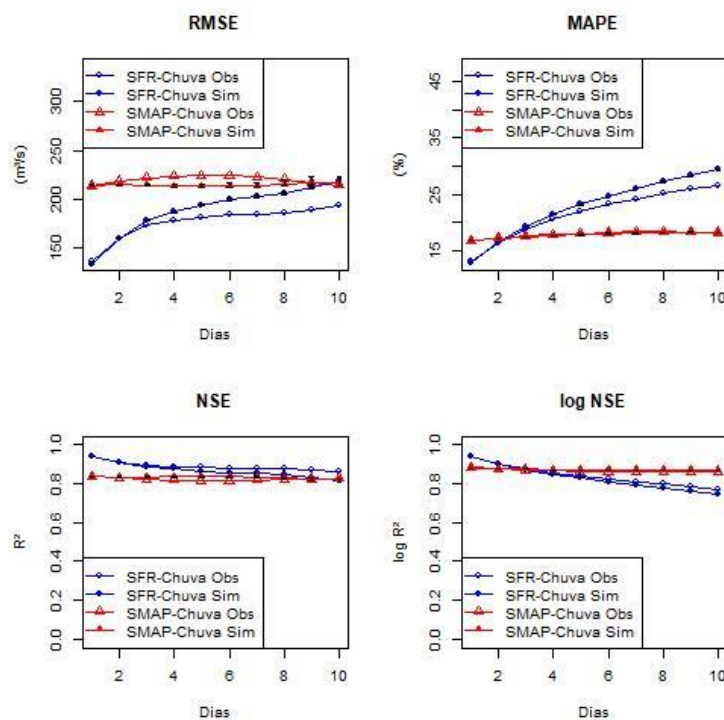


Figura 4.20: Resultados das Métricas de Erros para a UHE Mascarenhas

Fonte: Elaboração própria

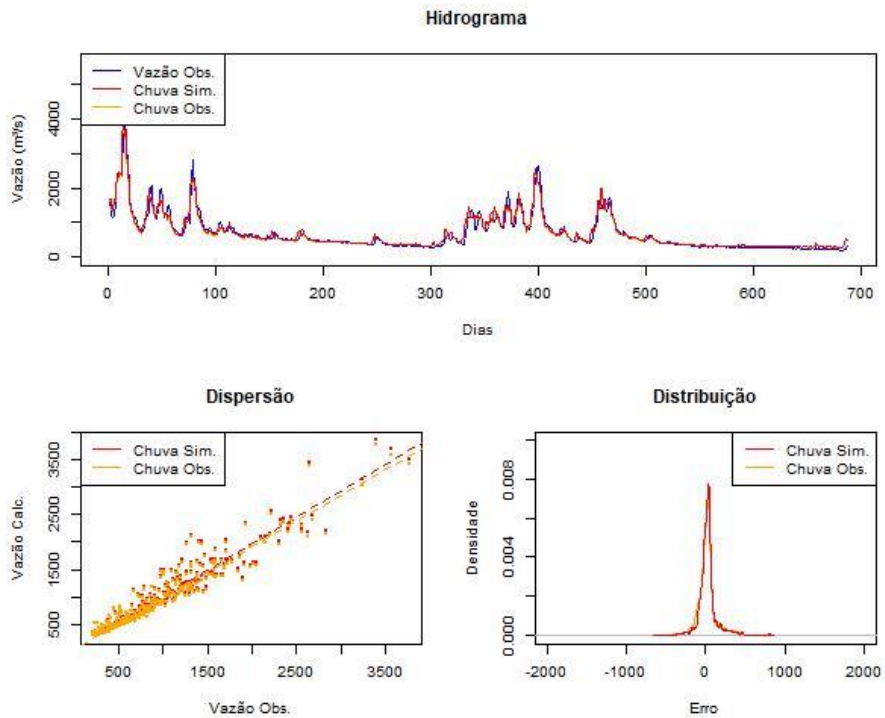


Figura 4.21: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SFR para 1 dia à frente

Fonte: Elaboração própria

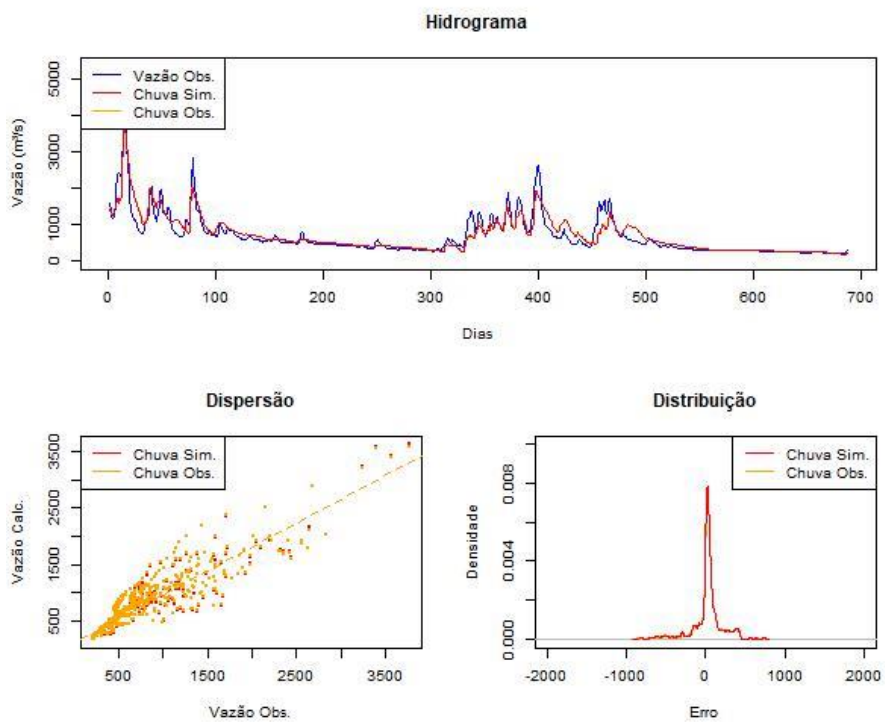


Figura 4.22: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SMAP para 1 dia à frente

Fonte: Elaboração própria

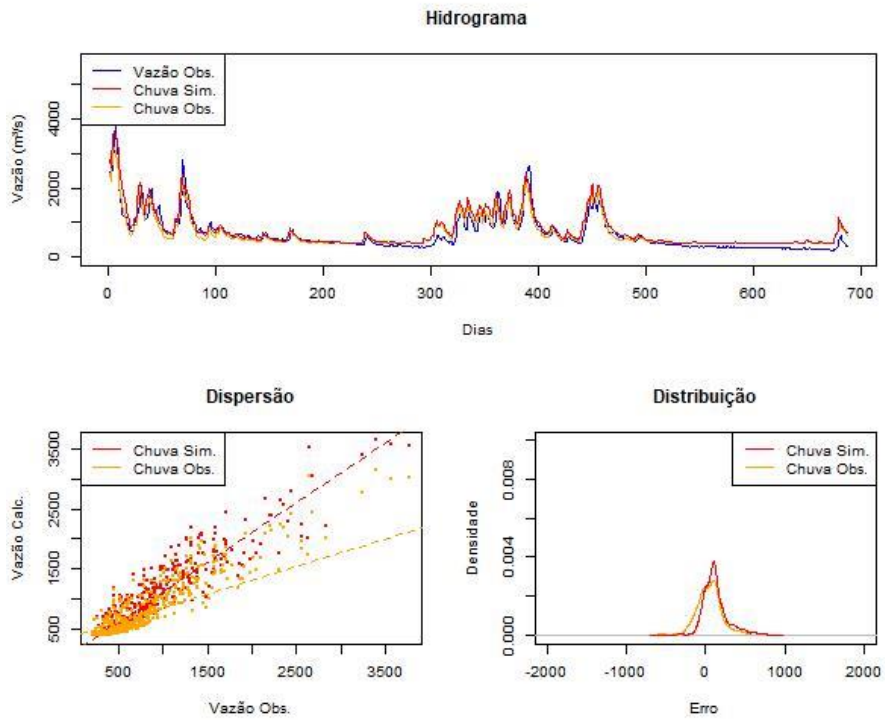


Figura 4.23: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SFR para 10 dias à frente

Fonte: Elaboração própria

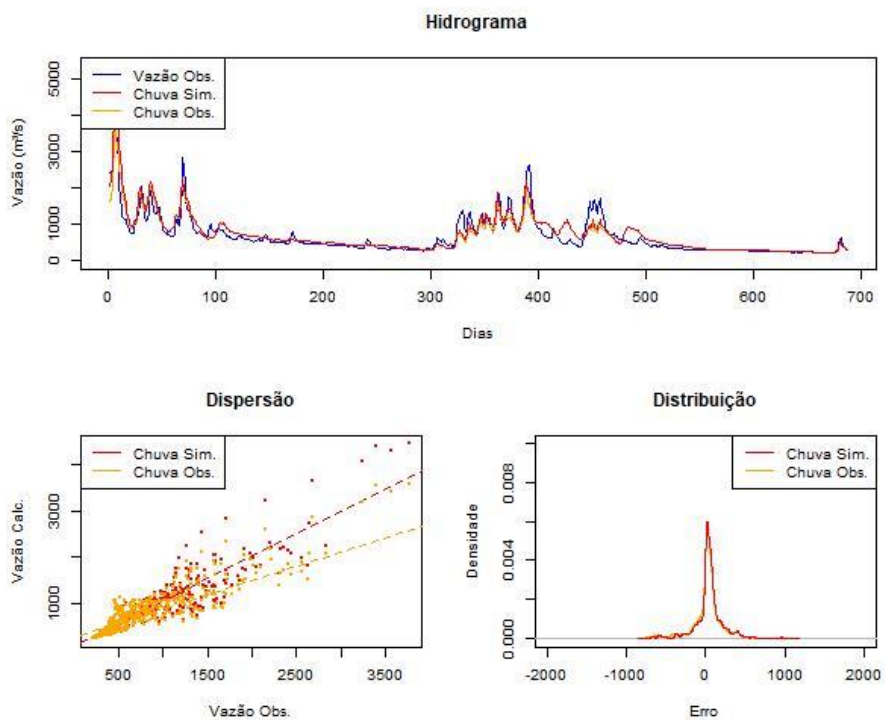


Figura 4.24: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Mascarenhas com o SMAP para 10 dias à frente

Fonte: Elaboração própria

Observando a aplicação da modelagem aos dados da UHE Três Marias, o modelo SFR atendeu às hipóteses discutidas no início desta seção, com a piora do desempenho do modelo com o crescimento do horizonte de previsão. Para a chuva observada como entrada e analisando a métrica RMSE, o modelo SFR inicia com um desempenho pouco superior a $100 \text{ m}^3/\text{s}$ no horizonte de previsão de um dia à frente, chegando à aproximadamente $220 \text{ m}^3/\text{s}$ no horizonte de dez dias. No entanto, verificando a métrica MAPE, o desempenho começa por volta de 20%, chegando a pouco mais que 40%. Para as métricas NSE e log NSE, o modelo mostra bom desempenho permanecendo acima de 0,8. Com a chuva simulada, para a RMSE, a curva de desempenho começa a descolar da curva de desempenho da chuva observada a partir do sétimo dia de previsão, sendo que no décimo dia a diferença entre as duas curvas não se mostra tão significativa. Na MAPE, a chuva simulada provoca um erro de quase 10% de diferença em relação a chuva observada, chegando a quase 50% de erro. Essa diferença ocorre pelas mesmas causas das aplicações anteriores, onde a MAPE apresenta maior peso para os erros nas vazões mais altas que a RMSE. A chuva simulada apresentou ainda bom desempenho na NSE, ficando acima de 0,8. Já na métrica log NSE, a chuva simulada ficou um pouco abaixo de 0,8, porém ainda se mostra um resultado aceitável.

Para o modelo SMAP, o fenômeno ocorrido na aplicação à UHE Mascarenhas também sucedeu. Na métrica RMSE, com a chuva observada, não houve piora crescente do desempenho do modelo. Ao contrário, a curva se assemelha sutilmente a uma forma parabólica. Onde os horizontes centrais de previsão são maiores que os extremos, variando de $250 \text{ m}^3/\text{s}$ à $280 \text{ m}^3/\text{s}$. A mesma conclusão pode ser feita para a métrica NSE, a qual permaneceu em torno de 0,8. Nas métricas MAPE e log NSE, o desempenho se mostra quase constante em torno de 30% e 0,85, respectivamente, mostrando um bom desempenho. A chuva simulada causou o mesmo fenômeno observado na UHE Mascarenhas, onde seu desempenho foi superior ao desempenho da chuva observada, provocando estranheza, principalmente para as métricas RMSE e NSE, com uma forma sutilmente parabólica. Estas tiveram o desempenho em torno de $250 \text{ m}^3/\text{s}$ e 0,8. Já para as métricas MAPE e log NSE não um descolamento tão grande entre as curvas de desempenho da chuva simulada e observada. A curva de distribuição do erro da previsão de vazão de um dia à frente foi mais afunilada para o SFR (Figura 4.26) ou seja, esta é mais leptocúrtica que a do SMAP (Figura 4.27). Já a curva de distribuição do erro do modelo SMAP mostra um maior desvio para a direita, o que leva a concluir que houve mais predição de vazão acima da observada. Para o horizonte de dez dias, as curvas de

distribuição do erro dos dois modelos tem certo viés para valores positivos, porém, a do SMAP (Figura 4.29) é mais leptocúrtica que do SFR (Figura 4.28) e não grande diferença entre a chuva simulada e observada como entrada.

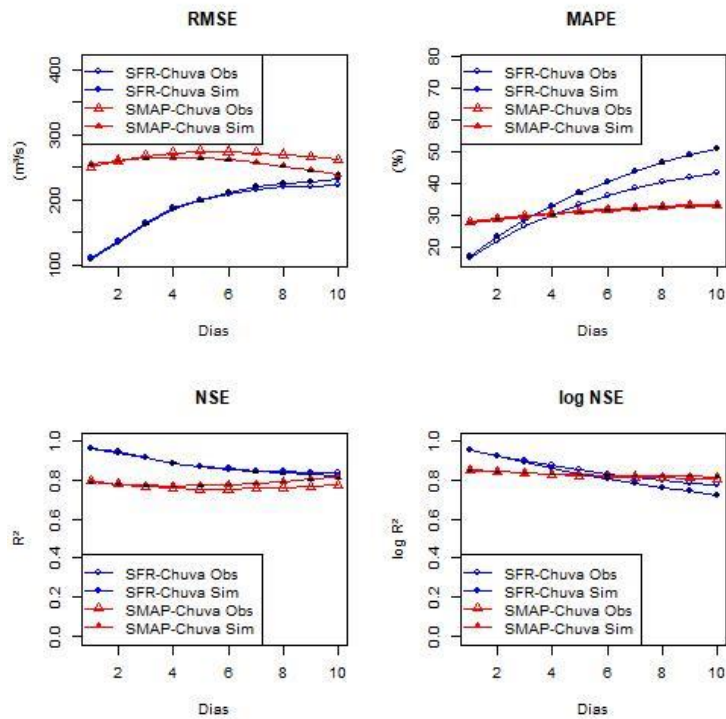


Figura 4.25: Resultados das Métricas de Erros para a UHE Três Marias

Fonte: Elaboração própria

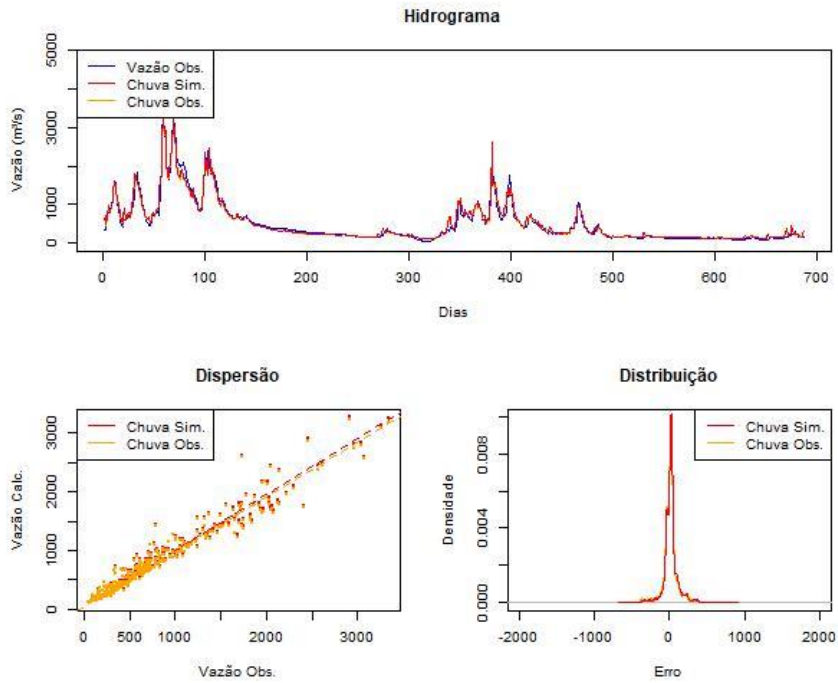


Figura 4.26: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SFR para 1 dia à frente

Fonte: Elaboração própria

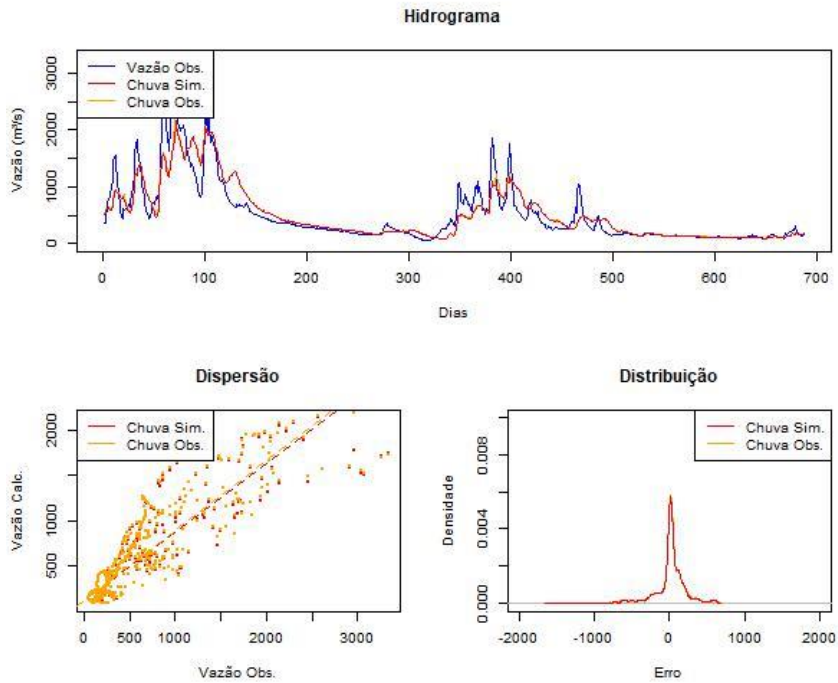


Figura 4.27: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SMAP para 1 dia à frente

Fonte: Elaboração própria

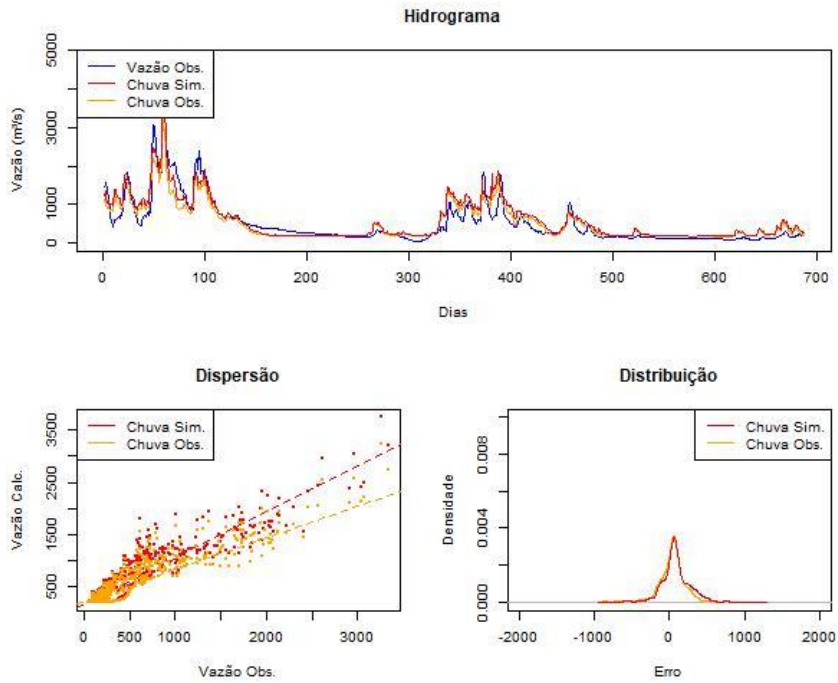


Figura 4.28: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SFR para 10 dias à frente

Fonte: Elaboração própria

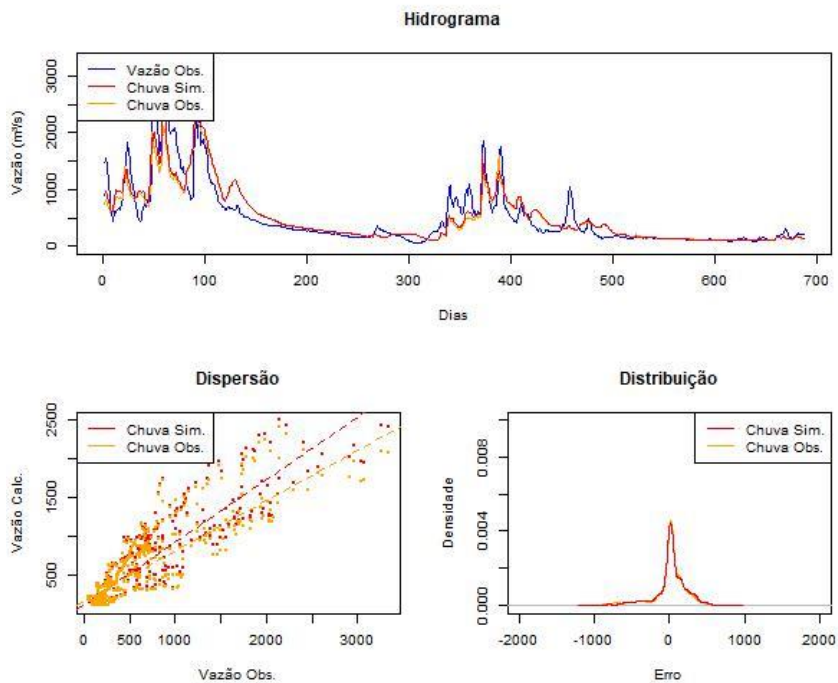


Figura 4.29: Hidrograma, Dispersão e Distribuição dos Erros para a UHE Três Marias com o SMAP para 10 dias à frente

Fonte: Elaboração própria

5. CONCLUSÕES E TRABALHOS FUTUROS

O Brasil é um país de vocação hidrelétrica que apresenta mais de 60% da capacidade instalada do seu parque baseado neste tipo de geração. A otimização da utilização dos recursos eletroenergéticos brasileiros é realizada pelo ONS através do despacho centralizados das usinas conectadas ao SIN. Por haver grande dependência dos recursos hídricos, o ONS necessita realizar a previsão de vazão afluente em cada uma das UHE do SIN. Esses resultados são, então, inseridos em um modelo de otimização eletroenergética o qual uma das saídas é o PLD, o preço semanal do mercado spot de energia elétrica brasileiro. Tal índice é base para diversas transações financeiras no mercado de energia elétrica no Brasil, tanto no ambiente regulado, como no livre.

O ONS utiliza diversos modelos de previsão de vazão em diferentes horizontes de previsão e discretização. Para a previsão de vazão 10 dias à frente, o ONS utiliza diversos modelos, os quais são baseados em diferentes abordagens. A utilização de cada modelo dependerá da localização da bacia hidrográfica. A razão disso se dá pelo fato que os modelos utilizados tem diferentes desempenhos para cada bacia hidrográfica. Assim, determinados modelos são mais indicados para a modelagem chuva-vazão em certas bacias que outros.

Os modelos SFR e SMAP, alvos da presente dissertação, são, hoje, utilizados pelo ONS para a previsão de vazão de 10 dias à frente e, conseqüentemente, auxiliam na tomada de decisão do despacho centralizado e otimização dos recursos eletroenergéticos do SIN. O SMAP é utilizado nas bacias dos rio Grande, Paranapanema e Paranaíba, na região Sudeste do Brasil, ao passo que o SFR é utilizado somente na bacia do rio Iguaçu, na região Sul do Brasil. Esses modelos, foram, então testados para a previsão de vazão dez dias à frente em UHE localizadas em outras bacias hidrográficas para verificar o seu desempenho e compara-los. As UHE testadas são Furnas (Paraná), Machadinho (Uruguai), Mascarenhas (Atlântico Leste) e Três Marias (São Francisco).

A operação real de previsão de vazão realizada pelo ONS utiliza como entrada a previsão de chuva realizada por um modelo meteorológico. Por ser um modelo, que por definição é uma simplificação do processo que ocorre em um sistema real, este apresenta também erros de predição. Para simular a operação real de previsão de vazão, erros foram introduzidos na variável de entrada do conjunto de testes para que se pudesse verificar também o efeito dos erros de previsão de chuva nos modelos de previsão de vazão.

O desempenho dos modelos dependeu das UHE testadas bem como das métricas de erro utilizadas, sendo no total quatro: RMSE, MAPE, NSE e log NSE. As métricas RMSE e NSE são mais sensíveis aos erros com altos valores da variável de saída vazão e as métricas MAPE e log NSE são mais sensíveis aos erros com baixos valores.

O modelo SFR mostrou bom desempenho quando aplicado às UHE Furnas, Mascarenhas e Três Marias para as métricas RMSE e NSE. Isso nos leva a concluir que o modelo acerta mais em valores mais baixos de vazão e erra mais em valores de vazão mais alta. Um fator que pode influenciar é o fato de o treinamento do modelo ser realizado com um AG na qual a função de ajuste é o MSE. Talvez, se a função de ajuste fosse alterada, o SFR poderia ter uma ponderação de erros melhor entre métricas mais sensíveis a valores altos e baixos da saída. O mesmo não ocorre para o modelo SMAP. A função objetivo na calibração do modelo é a distância entre duas métricas de erro, o NSE e o MAPE, os quais apresentam sensibilidades de tipos de erros diferentes. É provável que se a mesma função objetivo da calibração do SMAP fosse utilizada no treinamento do SFR, seus resultados poderiam ser melhores para as métricas MAPE e log NSE.

O desempenho do modelo SFR para a UHE Machadinho foi bem inferior quando aplicado às demais UHE. Tal resultado causa estranheza, pois esse modelo é usado hoje pelo ONS para a previsão de vazão na bacia do rio Iguaçu, vizinha à bacia do rio Uruguai e que apresenta características semelhantes em relação ao sistema chuva-vazão. Um fato importante, que não foi levado em consideração na modelagem do SFR neste trabalho é em relação a análise do relevo. Determinadas bacias apresentam questões do relevo que podem interferir significativamente no sistema-chuva vazão. No caso da UHE de Machadinho, talvez se a bacia fosse separada em outras de menor área e um modelo SFR fosse desenvolvido para cada área, talvez os resultados poderiam melhorar bastante. Isso deve ser melhor analisado em trabalhos futuros.

Outro ponto que poderia melhorar os resultados do SFR seria mudar o critério de seleção dos protótipos dos conjuntos fuzzy das variáveis de entrada e de estado. No presente trabalho, o método se deu por um espaçamento uniforme entre os percentis das variáveis que dependia da quantidade de conjuntos fuzzy. De fato, não se sabe ao certo se essa é a melhor maneira de escolher os protótipos. Talvez uma outra abordagem seria inseri-los na otimização do AG. Assim, além de realizar a identificação do vetor de parâmetros do sistema fuzzy, o AG também identificaria os percentis dos protótipos.

Outra abordagem que poderia melhorar o desempenho do SFR seria testa-lo como sistema MISO. Nesta dissertação, o SFR foi utilizado como SISO, transformando os

dados dos postos em um único vetor de entrada. Inserindo múltiplas entradas através, por exemplo, por agrupamento espacial, poderia melhorar o desempenho do modelo, já que o tempo de resposta da vazão é função da localização dos postos de chuva. Um inconveniente de se modelar um sistema MISO é que aumentaria o custo computacional.

Na operação real de previsão de vazão com o modelo SMAP, o ONS necessita de 31 dias anteriores ao primeiro dia de previsão para inicializar o modelo através dos parâmetros *Tuin*, *Ebin* e *Supin*. A escolha dos parâmetros é realizada por um hidrólogo experiente através do método tentativa-erro observando uma função objetivo, bem como a forma da hidrógrafa. Todavia, como metodologia de teste, no presente trabalho foram realizadas centenas de previsões de vazão diferentes, não sendo possível inicializar o SMAP manualmente. A inicialização foi então feita através da otimização do parâmetros de inicialização com auxílio de um algoritmo heurístico baseado em inteligência artificial chamado de Evolução Diferencial. Existem diversos outros algoritmos de otimização que poderiam ser utilizados, bem como diversas configurações que poderiam ser testadas para a inicialização do SMAP.

Na tentativa de se manter o modelo SMAP mais próximo da versão original, não foi inserida na modelagem na etapa de testes de previsão o método de reinicialização do modelo descrito em ONS e HICON (2012) chamado de “chuva perfeita”. Tal metodologia melhora o desempenho do modelo. Para trabalhos futuros, a inserção dessa etapa é altamente recomendada.

A hipótese levantada na qual o desempenho dos modelos de previsão de vazão pioram com o horizonte de previsão não necessariamente foi atendida no caso do modelo SMAP. Tal hipótese se dá, pois em muitos modelos o erro de previsão da saída nos horizontes mais distantes são somados aos erros da previsão de saída dos horizontes anteriores. Ou seja, no caso da previsão de vazão, por exemplo, no horizonte de três dias à frente, é necessário utilizar a previsão de dois dias à frente, que por sua vez utilizou a previsão de um dia e assim por diante. É o caso do modelo SFR e seus resultados mostraram isso. No entanto, para o caso do SMAP a hipótese não prevaleceu. Isso pode ser relativo ao fato que, diferente do SFR, o SMAP não utiliza as saídas previstas em horizontes mais próximos para prever horizontes mais distantes. Este utiliza valores de variáveis de estado anteriores para calcular variáveis de estados mais distantes. De alguma forma, isso deve impedir que o erro se propague ou o reduzir de maneira mais eficaz com o aumento dos horizontes de previsão.

Nos resultados do modelo SMAP para as UHE Mascarenhas e UHE Três Marias, a previsão de vazão utilizando a chuva simulada como entrada apresentaram melhores resultados que quando a chuva observada foi utilizada. Tal fato é de difícil explicação, já que se espera que o na entrada de qualquer modelo afetará seu desempenho, como o caso do SFR. De alguma forma, os ruídos introduzidos na chuva observada beneficiaram o resultado do SMAP. Talvez o fato de ser um modelo semi-distribuído possa causado tal resultado, porém deve ser averiguado com um detalhe maior.

As conclusões e questões levantadas ao longo desse trabalho poderão servir de estímulo para que futuros pesquisadores abordem o tema. Diversas outras abordagens poderiam ter sido diferentes no processo de modelagem, tanto do SFR quando do SMAP.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ANA, 2016. *Hidroweb*. Disponível em: <<http://hidroweb.ana.gov.br>>. Acesso em: 11 dez 2016.

ANEEL, 2017. *Banco de Informações de Geração - BIG*. Disponível em: <<http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>>. Acesso em: 15 jul. 2017.

ARDIA, David *et al.*, 2011. Differential Evolution with DEoptim. *The R Journal*, v. 3, n. 1, p. 27–34.

ARDIA, David *et al.*, 2016. *Package 'DEoptim': Global Optimization by Differential Evolution*.

BACKLUND, Alexander, 2000. The definition of system. *Kybernetes*, v. 29, n. 4, p. 444–451.

BEVEN, Keith, 2012. *Rainfall-runoff modelling*. Oxford: Wiley-Blackwell.

BRASIL, 2017. *Portal Brasil*. Disponível em: <<http://www.brasil.gov.br/meio-ambiente/2009/10/rios-e-bacias-do-brasil-formam-uma-das-maiores-redes-fluviais-do-mundo>>. Acesso em: 15 abr. 2017.

CANNON, Robert H. Jr., 2003. *Dynamics of Physical Systems*. Nova York: Dover Publications.

DE CARVALHO, Angela Regina Livino, 2001. *Previsão de Afluências Diárias para Horizontes de 1 a 7 dias através de uma abordagem estocástica - Modelo Previvazh*. Dissertação de Mestrado. Rio de Janeiro. UFRJ.

CEMIG, 2017 *Usinas Hidrelétricas da CEMIG*. Disponível em: <http://www.cemig.com.br/pt-br/a_cemig/Nossa_Historia/Paginas/Usinas_Hidreletricas.aspx>. Acesso em: 15 jul. 2017.

CNRH, 2003. *Resolução nº 32*. Disponível em: <<http://www.cnrh.gov.br>>. Acesso em: 17 jul. 2017.

CONSÓRCIO MACHADINHO. *Usina Hidrelétrica Machadinho*. Disponível em: <<http://www.machadinho.com.br/novo/usina.php?cod=83>>. Acesso em: 5 abr. 2017.

COUTINHO, Julio Ribeiro, 2015. *AutoMFIS: Um Sistema Fuzzy para Previsão de Séries Temporais Multivariadas*. Dissertação de Mestrado. Rio de Janeiro. PUC-RJ.

DE VOS, Nelis Jacob, 2009. *Computational Intelligence in Rainfall-Runoff Modeling*. Tese de Doutorado. Delft. TU Delft.

DOOGE, J. C. I., 1973. *Linear Theory of Hydrological Systems*. Washington

DOS SANTOS, Cristian Klen, 2013. *Identificação De Modelos Baseados Em Sistemas Fuzzy Recorrentes Com Duplo Feedback Usando Evolução Diferencial*. Tese de Doutorado. Rio de Janeiro. UFRJ.

EPE. *Consulta Pública nº 34 de 07/07/2017 - Relatório PDE 2026*. Rio de Janeiro 2017.

EVSUKOFF, Alexandre G; EBECKEN, Nelson F. F., 2004. Identification of Recurrent Fuzzy Systems with Genetic Algorithms. *Proceedings of IEEE International Conference on Fuzzy Systems*, p. 1703–1708.

EVSUKOFF, Alexandre G; DE LIMA, Beatriz S L P; EBECKEN, Nelson F F. Long-Term Runoff Modeling Using Rainfall Forecasts with Application to the Iguaçu River Basin. *Water Resources Management*, v. 25, n. 3, p. 963–985, 2011.

FURNAS. *Parque Gerador: Usina Hidrelétrica de FURNAS*. Disponível em: <http://www.furnas.com.br/hotsites/sistemaFurnas/usina_hidr_furnas.asp>. Acesso em: 15 maio 2017.

GAMA, Carlos A. *et al.* Parameter identification of recurrent fuzzy systems with fuzzy finite-state automata representation. *IEEE Transactions on Fuzzy Systems*, v. 16, n. 1, p. 213–224, 2008.

GORRINI, Vittorio; BERSINI, Hughes, 1994. Recurrent fuzzy systems. *Proceedings of the Third IEEE Conference on Fuzzy Systems*, p. 193–198.

GOWER, J. C., 1971. A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, v. 27, n. 4, p. 857–871.

KING, P.J.; MAMDANI, E.H., 1977. The application of fuzzy control systems to industrial processes. *Automatica*, v. 13, n. 3, p. 235–242.

KOSKO, Bart, 1994. Fuzzy systems as universal approximators. *IEEE Transactions on Computer*, v. 43, n. 11, p. 1329–1333.

LOPES, João Eduardo G, 1999. *Manual SMAP*.

LOPES, João Eduardo G.; BRAGA, Benedito B. F. Jr.; CONEJO, João Gilberto L., 1982. SMAP A Simplified Hydrologic Model. In: SINGH, P. VIJAY (Org.). *Applied Modeling in Catchment Hydrology*. Water Resources Publications. p. 167–176.

LUENBERGER, David G, 1979. *Introduction to Dynamic Systems: Theory, Models, and Applications*. Nova York: John Wiley & Sons.

MELLODGE, Patricia, 2016. *A practical approach to dynamical systems for engineers*. Cambridge: Woodhead Publishing.

MENDONÇA, Augusto F.; DAHL, Carol, 1999. The Brazilian electrical system reform. *Energy Policy*, v. 27, n. 2, p. 73–83.

MISHRA, Surendra Kumar; SINGH, Vijay P., 2003. *Soil Conservation Service Curve Number (SCS-CN) Methodology*. Springer Science+Business Media.

MORETTI, G; MONTANARI, A, 2007. AFFDEF: A spatially distributed grid based rainfall–runoff model for continuous time simulations of river discharge. *Environmental Modelling & Software*, v. 22, n. 6, p. 823–836.

MOURA, Rildo Gonçalves De *et al.*, 2010. Evaluation of the eta regional model using the analysis of CPTEC and NCEP. *Revista Brasileira de Meteorologia*, v. 25, n. 1, p. 46–53.

MULLEN, Katharine *et al.*, 2011 DEoptim : An R Package for Global Optimization by Differential Evolution. *Journal of Statistical Software*, v. 40, n. 6, p. 1–26.

ONS, 2008. *Metodologia Para a Semana À Frente Na Bacia Do Alto / Médio Rio Grande*.

ONS, 2016. *Relatório Executivo de Previsões de Vazões –2016*.

ONS. *Séries Históricas de Vazões*. Disponível em: <http://www.ons.org.br/operacao/vazoes_naturais.aspx>. Acesso em: 10 jun. 2017.

ONS; HICON, 2012. *Aplicação de Modelos de Previsão a Curto Prazo na Sub-bacia do Alto Rio Paranaíba - Relatório Final*. Rio de Janeiro.

SOLDÁ, Grazielly, 2008. *Modelos de memória longa, GARCH e GARCH com memória longa para séries financeiras*. 2008. Dissertação de Mestrado. São Paulo. USP.

SPOLIA, S. K.; CHANDER, S., 1974. Modelling of Surface Runoff Systems by an ARMA Model. *Journal of Hydrology*, v. 22, p. 317–332.

TEMPL, Matthias *et al.*, 2016 *Package ‘VIM’*.

TODINI, E., 2007 Hydrological catchment modelling: past, present and future. *Hydrological Earth System Science*, v. 11, n. 1, p. 468–482.

TODINI, E., 1988 Rainfall-runoff modeling - Past, present and future. *Journal of Hydrology*, v. 100, n. 1–3, p. 341–352.

TONG, R M., 1980. The evaluation of fuzzy models derived experimental data. *Fuzzy Sets and Systems*, v. 4, p. 1–12.

XU, Chong-yu , 2002. *Text Book of Hydrologic Models*. Uppsala. Uppsala University. v. 72.

ZADEH, L. A., 1965. Fuzzy sets. *Information and Control*, v. 8, n. 3, p. 338–353.