

UNIVERSIDADE FEDERAL DE ALFENAS

BETHÂNIA OLIVEIRA DE BRITO

**MODELAGEM INTELIGENTE PARA PREVISÃO DE SÉRIES DE VAZÕES
AFLUENTES**

Alfenas/MG

2016

BETHÂNIA OLIVEIRA DE BRITO

**MODELAGEM INTELIGENTE PARA PREVISÃO DE SÉRIES DE VAZÕES
AFLUENTES**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional.
Orientador: Prof. Dr. Ricardo Menezes Salgado.
Coorientador: Prof. Dr. Luiz Alberto Beijo.

Alfenas/MG

2016

Dados Internacionais de Catalogação-na-Publicação (CIP)
Biblioteca Central da Universidade Federal de Alfenas

Brito, Bethânia Oliveira de.

Modelagem inteligente para previsão de séries de vazões afluentes /
Bethânia Oliveira de Brito. -- Alfenas/MG, 2016.

101 f.

Orientador: Ricardo Menezes Salgado.

Dissertação (mestrado em Estatística Aplicada e Biometria) -
Universidade Federal de Alfenas, 2016.

Bibliografia.

1. Redes Neurais (Computação). 2. Algoritmos - Genética. 3. Previsão
Hidrológica. I. Salgado, Ricardo Menezes. II. Título.

CDD-519



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Alfenas / UNIFAL-MG
Programa de Pós-graduação em Estatística Aplicada e Biometria

Rua Gabriel Monteiro da Silva, 700. Alfenas - MG CEP 37130-000
Fone: (35) 3299-1392 (Secretaria) (35) 3299-1121(Coordenação)
<https://www.unifal-mg.edu.br/ppgeab/>



BETHÂNIA OLIVEIRA DE BRITO

“MODELAGEM INTELIGENTE PARA PREVISÃO DE SÉRIES DE VAZÕES
AFLUENTES”

A Banca Examinadora, abaixo assinada, aprova a
Dissertação apresentada como parte dos requisitos para
a obtenção do título de Mestre em Estatística Aplicada
e Biometria pela Universidade Federal de Alfenas.
Linha de Pesquisa: Modelagem Estatística e Estatística
Computacional.

Aprovado em: 26 de fevereiro de 2016.

Prof. Dr. Ricardo Menezes Salgado

Instituição: UNIFAL-MG

Assinatura:

Prof. Dr. Luiz Alberto Beijo

Instituição: UNIFAL-MG

Assinatura:

Prof. Dr. Takaaki Ohishi

Instituição: UNICAMP

Assinatura:

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me dar força e me proporcionar a realização deste sonho de forma tão abençoada.

A minha mãe Maria Consolação, por sempre me apoiar nesta caminhada e me mostrar a importância do estudo e do conhecimento em minha vida. A minha irmã Suelen que sempre soube me apoiar nos momentos mais difíceis com sua palavra amiga e sincera, obrigada por ser essa pessoa doce em minha vida. Amo vocês incondicionalmente.

Ao meu orientador Ricardo Menezes Salgado e ao meu coorientador Luiz Alberto Beijo por exercerem tão bem seus papéis. Muito obrigada pela dedicação, paciência e ensinamentos. Serei eternamente grata.

Ao meu amor Jackson, pela paciência, incentivo e carinho que sempre me proporcionou mesmo com a distância. Obrigada por tudo!

A minha avó Maria Aparecida pelas intermináveis orações. A minha madrinha de batismo Bete e meu primo/irmão Vinícius por serem sempre um apoio, uma segunda família. Vocês são um exemplo pra mim.

As minhas amigas de infância Valéria, Ana Carolina e Thácia, obrigada pela amizade.

As minhas amigas que elegi como verdadeiras irmãs Flávia e Aline, vocês são um presente de Deus em minha vida! Obrigada por todo incentivo de vocês!

Aos amigos que a graduação me deu e que permanecem até hoje: Juliana Ladeira, Bárbara, Juliana Fischer, Alice, Douglas, Isabela, Gaby. Obrigada por todos os momentos de alegria.

Aos colegas de mestrado que foram parte fundamental dessa caminhada: Isabela, Cristiane, Helen, Estela, Diailison, Bruna, Roger, Renata, Pórtia, Pablo, Larissa, Laís e Gustavo. Os momentos difíceis não seriam os mesmos se não pudesse contar com o apoio de vocês!

Aos professores Fabrício Avelar, Denismar Nogueira e Eric Ferreira pelos grandes ensinamentos durante o curso de mestrado, obrigada pela grande contribuição para meu crescimento intelectual e pessoal. Suas aulas foram essenciais para realização deste trabalho.

A todos aqueles que de alguma forma contribuíram para essa realização.

RESUMO

A geração de energia elétrica é assunto estratégico para o desenvolvimento econômico de qualquer nação e geralmente está ligada aos recursos naturais disponíveis. A exploração de tais fontes devem ser aproveitadas de maneira a maximizar os benefícios proporcionados e minimizar os impactos negativos ao meio ambiente e sociedade. A estratégia de previsão de séries de vazões consiste em estimar o fluxo de água com determinada antecedência visando minimizar as incertezas e os riscos auxiliando na redução dos fatores que prejudicam o planejamento das hidroelétricas e maximizando os resultados obtidos, pois a geração de energia elétrica a partir do sistema hidráulico depende principalmente das séries de vazões afluentes (TUCCI, 2002). Para realizar previsão de séries de vazões, encontram-se na literatura diversos modelos, dentre estes, as redes neurais artificiais, programação genética, modelos autorregressivos, entre outros. A fim de melhorar o desempenho das previsões de vazões propõe-se neste trabalho a construção de *ensembles*, que consiste em combinar componentes individuais. Neste trabalho, utilizou-se uma base dados do Operador Nacional de Sistemas (ONS) de duas usinas localizadas no Rio Grande: Água Vermelha e Itutinga. Os modelos que mais se destacaram como componentes individuais foram a rede neural artificial (RNA) com algoritmos de treinamento *Backpropagation* (BPM) e *Gradient Method* (GRAD) e a Programação Genética (PG). O *ensemble* BPM foi o que apresentou maior eficiência e capacidade de generalização. O MAPE da previsão dos modelos do período seco é menor que no período úmido. Não houve um modelo que se destacou em todos os casos quanto aos erros de previsão, sendo que os resultados dependem das características da usina e do período em estudo. Fazer previsões por períodos levaram a menores erros que quando considerado todo o ano. Após a combinação das componentes individuais, na maioria dos casos houve melhoria do desempenho, sendo que o melhor caso foi capaz de prover uma diminuição de até 14% do erro médio absoluto percentual (MAPE) em relação a melhor componente individual.

Palavra-chave: Redes Neurais Artificiais. Programação Genética. *Ensemble*.

ABSTRACT

The generation of power is of strategic importance for the economic development of any nation. The ability to generate power is fundamentally linked to the availability of natural resources. The exploitation of such resources should be guided by principles to maximize the benefit provided and minimize the negative impact on the environment and society. The generation of electricity from hydraulic system depends mainly on the water inflow series (TUCCI, 2002). The forecast strategy series streamflow estimates the water flow with the goal of minimizing uncertainties and risks while reducing factors that hinder the planning of hydroelectric energy production. There are several models in the literature for performing streamflow series forecasting. They include: artificial neural networks, genetic programming, and autoregressive models, among others. In this paper, we propose the construction of ensembles - the combination of individual components - in order to improve the performance of forecasts of streamflow rates. We used one database from the Operador Nacional do Sistema Elétrico (ONS) in two plants located in Rio Grande: Água Vermelha and Itutinga. The models that stood out were the artificial neural network (ANN) with the training algorithms Backpropagation (BPM), Gradient Method (GRAD), and genetic programming (GP). The ensemble BPM showed greater efficiency and generalizability. The forecast MAPE of models for dry periods is less than for the wet season. Model results depended upon the characteristics of the plant and the period under study. Making predictions by periods led to minor mistakes when taken throughout the year. After combining the individual components, there was up to a 14% reduction of the average percentage absolute error (MAPE).

Keyword: Artificial Neural Networks. Genetic Programming. Ensemble.

LISTA DE FIGURAS

Figura 1 - Representação gráfica das fontes de energia e sua contribuição para a matriz energética brasileira.	13
Figura 2 - Esquema ilustrativo da obtenção dos valores das séries históricas de vazões afluentes mensais e anuais.	18
Figura 3 - Representação gráfica da vazão e precipitação através do hietograma e Hidrograma e contribuição dos escoamentos superficial e de base para a vazão.	20
Figura 4 - Representação gráfica da vazão através dos hidrogramas comparativos de uma bacia radial e alongada	21
Figura 5 - Representação gráfica da vazão através dos hidrogramas comparativos de uma bacia rural e urbana.	22
Figura 6 - Série de vazão afluente média mensal para as usinas hidroelétricas Furnas e Igarapava no período de 1994 a 2003.	23
Figura 7 - Distribuição de frequência das séries de vazões para as usinas Furnas e Igarapava no período de 1931 a 2013.	24
Figura 8 - Média de longo termo (MLT) das séries de vazões das usinas de Furnas e Igarapava (1931 a 2013).	24
Figura 9 - Média de longo termo (MLT) das séries de vazões padronizadas para as usinas de Furnas e Igarapava (1931 a 2013)	25
Figura10 - Representação gráfica da variação do grau de incerteza da vazão quanto ao horizonte de previsão.	26
Figura11 - Correlograma para séries de vazões das usinas Furnas e Igarapava no período de 1931 a 2013.	27
Figura12 - Usinas hidroelétricas de reservatório e a fio d'água que compõem a bacia do Rio Grande e seu afluente Rio Pardo.	28
Figura13 - Medidas de posição e variabilidade mensal em m ³ s das séries de vazões para as usinas Furnas e Igarapava no período de 1931 a 2013.	30
Figura14 - Representação ilustrativa de um neurônio artificial com X _n padrões de entradas, uma função de ativação g(.) e uma saída y.	35

Figura15- Representação gráfica da função de ativação tangente hiperbólica mostrando o contra-domínio restrito no intervalo $[-1;1]$	36
Figura16- Representação ilustrativa de uma rede neural artificial do tipo <i>Multilayer Perceptron</i> com duas camadas intermediárias.	37
Figura17 - Ciclo “Criar-Testar-Modificar” utilizado na programação genética para selecionar os indivíduos que mais se adaptam ao problema.	44
Figura18 - Visão geral do algoritmo utilizado pela programação genética para chegar a uma solução satisfatória.	44
Figura19 - Exemplo da representação do programa $x^2 + 2$ da programação genética através da árvore sintática.	45
Figura20 - Exemplo ilustrativo do ganho de desempenho utilizando ensemble em problemas de previsão de séries temporais.	48
Figura21 - Esquema ilustrativo da construção de um <i>ensemble</i> em duas etapas: geração das componentes de previsão e a combinação das previsões obtidas através destas componentes.	48
Figura22 - Esquema da geração de conjuntos de treinamentos distintos para geração das componentes de previsão via <i>bagging</i>	49
Figura23 - Padrões de treino, validação e teste utilizados pelos modelos para prever o mês de janeiro de 2004.	60
Figura24 - Comparação dos dados brutos e normalizados da usina de Furnas do período de 1994 a 2003.	62
Figura25 - Esquema de obtenção dos resultados das previsões através das componentes individuais de previsão.	62
Figura26 - Esquema ilustrativo do <i>ensemble</i> utilizando oito e duas componentes individuais de previsão.	64
Figura27 - Esquema <i>ensemble</i> com entradas e combinador sendo os modelos redes neurais artificiais (RNA) e a programação genética (PG)	66
Figura28 - Série de vazão afluente média mensal para a usina hidroelétrica Água Vermelha no período de 1994 a 2003.	69
Figura29 - Distribuição de frequência da vazão no período de 1931 a 2013 para a usina Água Vermelha.	70

Figura30 - Série de vazão afluyente média mensal para a usina hidroelétrica Itutinga no período de 1994 a 2003.	71
Figura31 - Distribuição de frequência da vazão no período de 1931 a 2013 para a usina Itutinga.	72
Figura32 - Média de longo termo (MLT) para as usinas Água Vermelha e Itutinga utilizando dados de vazões do período de 1931 a 2013.	74
Figura33 - Média de longo termo (MLT) das séries de vazões padronizadas (m^3/s) do período de 1931 a 2013 para as usinas Água Vermelha e Itutinga.	75
Figura34 - MAPE (%) das previsões da vazão (m^3/s) para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.	76
Figura35 - MAPE (%) das previsões da vazão (m^3/s) realizadas através dos <i>ensembles</i> : MA, MP, MQ, MG e MIB para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.	77
Figura36 - MAPE (%) das previsões da vazão (m^3/s) para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.	79
Figura37 - MAPE (%) das previsões da vazão (m^3/s) realizadas através dos <i>ensembles</i> : MA, MP, MQ, MG e MIB para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.	80
Figura38 - MAPE (%) das previsões da vazão (m^3/s) para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.	81
Figura39 - MAPE (%) das previsões da vazão (m^3/s) realizadas através dos <i>ensembles</i> : MA, MP, MQ, MG e MIB para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.	82
Figura40 - Comparação do desempenho das componentes individuais com os valores reais da séire (2004 a 2013).	83
Figura41 - Regime de precipitação mensal no período de 1981 a 2010 na região da bacia do Rio Grande.	85
Figura42 - Medidas de posição e variabilidade da vazão (m^3/s) das fases de treino (1931 a 1993), validação (1994 a 2003) e teste (2004 a 2013) para a usina Água Vermelha.	97
Figura43 - Medidas de posição e variabilidade da vazão (m^3/s) das fases de treino (1931 a 1993), validação (1994 a 2003) e teste (2004 a 2013) para a usina Itutinga.	98

LISTA DE TABELAS

Tabela 1 - Fontes de energia no Brasil, número de usinas de cada fonte de energia, geração energética e porcentagem de contribuição de cada fonte para a matriz energética brasileira.	13
Tabela 2 - Medidas de posição para as séries de vazões mensais das usinas Furnas e Igarapava em m^3/s no período de 1931 a 2013	29
Tabela 3 - Exemplo de padrão utilizado pelos modelos para prever o mês de janeiro com uma dependência temporal de quatro.	59
Tabela 4 - Dependências temporais para cada mês do ano calculadas através da função de autocorrelação para as usinas Água Vermelha e Itutinga.	59
Tabela 5 - Parâmetros utilizados pelos modelos RNA e PG para obtenção das previsões das vazões no período de validação e teste.	61
Tabela 6 - <i>Ensembles</i> baseado em médias para combinar as componentes individuais de previsão	65
Tabela 7 - Medidas descritivas das séries de vazões (m^3/s) das fases de treino, validação e teste para os períodos seco e úmido das usinas Água Vermelha e Itutinga.	73
Tabela 8 - MAPE (%) das componentes de previsão individuais por ano para as usinas Água Vermelha e Itutinga.	84
Tabela 9 - MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 das usinas Água Vermelha e Itutinga no período seco.	99
Tabela10- MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 do período seco realizadas através dos ensembles: MA, MP, MQ, MG e MIB.	99
Tabela11- MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 das usinas Água Vermelha e Itutinga no período úmido.	100
Tabela12- MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 do período úmido realizadas através dos ensembles: MA, MP, MQ, MG e MIB.	100

Tabela13- MAPE (%) para as previsões da vazão (m ³ /s) correspondentes aos anos 2004 a 2013 das usinas Água Vermelha e Itutinga para todo o ano.	101
Tabela14- MAPE (%) para as previsões da vazão (m ³ /s) correspondentes aos anos 2004 a 2013 para todo o ano realizadas através dos <i>ensembles</i> : MA, MP, MQ, MG e MIB.	101

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS GERAIS.	15
1.2	OBJETIVOS ESPECÍFICOS.	15
1.3	ORGANIZAÇÃO DA DISSERTAÇÃO	16
2	SÉRIES DE VAZÕES AFLUENTES	17
2.1	VAZÕES E AFLUENTES.	17
2.2	PREVISÃO DE SÉRIES DE VAZÕES	25
3	REFERENCIAL TEÓRICO	31
3.1	REVISÃO BIBLIOGRÁFICA	31
3.2	METODOLOGIAS DE PREVISÃO.	34
3.2.1	Redes neurais artificiais (RNA)	35
3.2.2	Programação Genética (PG).	43
3.2.3	<i>Ensembles</i>	47
4	MATERIAL E MÉTODOS	57
4.1	DADOS DE VAZÃO.	57
4.2	COMPONENTES DE PREVISÃO.	57
4.3	<i>ENSEMBLES</i>	63
4.4	RECURSOS COMPUTACIONAIS	68
5	RESULTADOS	69
5.1	ANÁLISE DESCRITIVA DOS DADOS DE VAZÃO.	69
5.2	ESTUDOS DE CASOS	75
5.2.1	Resultados para o período seco.	75
5.2.2	Resultados para o período úmido.	78
5.2.3	Resultados gerais.	81
6	CONCLUSÃO E PERSPECTIVAS FUTURAS	86
	REFERÊNCIAS	87
	APÊNDICE	96

1 INTRODUÇÃO

A geração de energia elétrica é assunto estratégico para o desenvolvimento econômico de qualquer nação e geralmente está ligada aos recursos naturais disponíveis. A exploração de tais fontes devem ser aproveitadas de maneira a maximizar os benefícios proporcionados e minimizar os impactos negativos ao meio ambiente e à sociedade.

Dentre as diversas fontes de geração de energia: hídrica, biomassa, eólica, fóssil, nuclear e solar, a hídrica é a mais explorada no Brasil devido às condições climáticas e geográficas do país. A geografia do território brasileiro é marcada, em grande parte, por um relevo suave o que favoreceu a formação dos imensos recursos hídricos disponíveis. Desta forma o Brasil possui uma das mais extensas redes fluviais do mundo e por esta razão é um dos países com maior potencial hidráulico (AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL), 2016).

O clima brasileiro é diverso devido a extensão territorial, o relevo e a dinâmica das massas de ar. As massas de ar tem importância ímpar visto que influencia diretamente tanto na temperatura quanto na pluviosidade. O conhecimento de tais fatores são essenciais já que os recursos fluviais exibem regime de alimentação pluvial e devido ao clima predominantemente tropical, usualmente, as cheias ocorrem durante o verão com exceção a alguns rios da região nordeste cujas cheias ocorrem entre o outono e o inverno. Porém, é relevante destacar que o regime de alimentação pluvial pode mudar drasticamente em meio a fenômenos climáticos tais como o *El Niño*¹ e *La Niña*².

Na Tabela 1 e Figura 1 estão as contribuições de cada fonte de energia utilizada pela matriz de energia elétrica brasileira, em que fica claro a predominância das fontes hídricas que representa 61% de toda energia gerada. Vale destacar que essa fatia irá aumentar com a conclusão das obras de novas usinas.

¹*El Niño* é um fenômeno atmosférico-oceânico caracterizado por um aquecimento anormal das águas superficiais no oceano Pacífico Tropical, e que pode afetar o clima regional e global, mudando os padrões de vento a nível mundial, e afetando assim, os regimes de chuva em regiões tropicais e de latitudes médias (Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) (2016)).

²*La Niña* representa um fenômeno oceânico-atmosférico com características opostas ao EL Niño, e que caracteriza-se por um esfriamento anormal nas águas superficiais do Oceano Pacífico Tropical. Alguns dos impactos de *La Niña* tendem a ser opostos aos de *El Niño*, mas nem sempre uma região afetada pelo *El Niño* apresenta impactos significativos no tempo e clima devido à *La Niña* (CPTEC (2016)).

Tabela 1 – Fontes de energia no Brasil, número de usinas de cada fonte de energia, geração energética e porcentagem de contribuição de cada fonte para a matriz energética brasileira.

Fonte	Número de Usinas	KW	%
Biomassa	518	13268315	8,8573
Eólica	336	8150582	5,4407
Fóssil	2350	26150645	17,4562
Hídrica	1208	92054920	61,449
Nuclear	2	1990000	1,3283
Solar	33	21336	0,0142
Importação	-	8170000	5,4436

Fonte: ANEEL (2016).

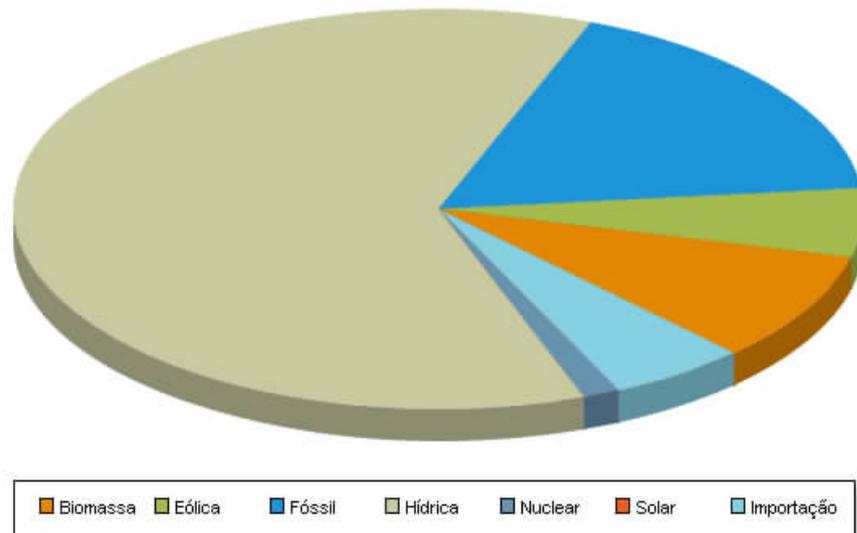


Figura 1 – Representação gráfica das fontes de energia e sua contribuição para a matriz energética brasileira.

Fonte: ANEEL (2016).

Como pode-se ver o Brasil dispõe de diversas fontes de energia, o que leva a uma posição mundial bastante competitiva, porém quanto maior a disponibilidade de recursos maior o compromisso com o planejamento energético. Logo, é imprescindível que o planejamento energético assegure a disponibilidade de energia necessária ao país com os menores custos e que seja o mais sustentável possível.

O principal fator para o planejamento energético dos países é a disponibilidade e a dinâmica dos recursos naturais existentes. Como no Brasil o recurso predominante na matriz energética é hídrico este se torna objeto de diversos estudos buscando a otimização em sua utilização como meio de gerar energia elétrica. Tal cuidado com os recursos hídricos são devidos a sazonalidade dos regimes pluviais principalmente quando sob efeito de condições

climáticas adversas como o *El niño* e *La niña*. O efeito desses eventos climáticos no Brasil podem aumentar as chuvas de maneira significativa ou diminuí-la drasticamente causando até mesmo um período de seca.

Durante os períodos de seca o potencial hídrico fica comprometido e a demanda energética deve ser atendida com as outras fontes (térmicas) disponíveis, que geralmente são mais caras e poluentes. Portanto, o planejamento de uma operação energética consiste em estabelecer uma política entre essas matrizes principais de modo a substituir ao máximo o sistema termoeletrico pelo hidroeletrico (SOARES FILHO, 1987).

Desta forma, a geração de energia envolve a tomada de decisões que é um processo complexo por envolver as incertezas climáticas e deve ser realizada de forma eficiente e com o maior grau de confiabilidade possível visando atender a toda demanda.

Os recursos hídricos são o insumo para as usinas hidroelétricas e são quantificados através da vazão dos rios, portanto este é o ponto central no processo de previsão para o planejamento das políticas energéticas.

Como o insumo principal das usinas hidroelétricas são os recursos hídricos e estes são quantificados através da vazão dos rios, portanto, este será o ponto central no processo de previsão para o planejamento das políticas energéticas.

A estratégia de previsão de séries de vazões consiste em estimar o fluxo de água com determinada antecedência visando minimizar as incertezas e os riscos auxiliando na redução dos fatores que prejudicam o planejamento das hidroelétricas (TUCCI, 2002). Portanto, realizar previsão de vazões é uma alternativa que ajuda maximizar os resultados obtidos, pois a geração de energia elétrica a partir do sistema hidráulico depende principalmente das séries de vazões afluentes.

Com a previsão de vazões, as hidrelétricas podem planejar suas decisões com a possibilidade de se anteciparem sobre algumas desconformidades que poderão vir a acontecer, tornando-as mais competitivas e aproveitando de forma mais adequada os recursos hídricos disponíveis. Além disso, as previsões servem como subsídio a outros setores usuários da água, como a irrigação, a pesca, o turismo, o lazer e a navegação (BRAVO et al., 2008). Vale lembrar que o setor elétrico brasileiro detém a maior parte do controle da quantidade de água disponível nos grandes reservatórios, que até pouco tempo destinavam-se exclusivamente à geração de energia, hoje são alvos de interesses diversos e conflitantes dos demais setores que utilizam a água. Nesse contexto, cresce a importância na área de planejamento e controle de sistemas de

recursos hídricos (LIMA, 2010).

Segundo Silva (2007) é importante ressaltar que realizar previsão pode ser um processo complicado, em que a complexidade dos fenômenos hídricos e a aleatoriedade das vazões são algumas das limitações que dificultam o planejamento da operação do sistema energético brasileiro, aliado a outros fatores como: topografia, tipo de solo, dinâmica de escoamento, etc. Portanto, para se conhecer a disponibilidade energética é possível realizar um estudo sobre previsão de vazões nos reservatório. Uma das formas para se alcançar tal objetivo é utilizar alguma abordagem computacional para beneficiar este processo, alguns modelos encontrados na literatura podem ser citados: Redes Neurais Artificiais estudada por Haykin (2007), Programação Genética desenvolvida por Koza (1992), Modelos Autorregressivos de Box e Jenkins (1976), entre outras, que são utilizadas com frequência na construção de modelos para solucionar problemas relacionados à previsão de séries temporais em diferentes áreas.

Entretanto, existe uma grande dificuldade em encontrar um modelo que seja capaz de sempre obter bons resultados de previsão, pois uma série pode mudar seu padrão de comportamento a partir de determinado instante tornando o modelo ineficaz. Neste sentido, os *ensembles* se tornam uma alternativa promissora, através da qual é possível combinar diferentes modelos de previsão com a finalidade de obter previsões mais precisas e acuradas.

1.1 OBJETIVOS GERAIS

O estudo desenvolvido neste trabalho tem como objetivo propor e comparar metodologias de previsão de séries de vazões médias mensais baseadas na combinação de previsores utilizando dados das usinas hidroelétricas Água Vermelha e Itutinga, pertencentes a bacia do Rio Grande.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos consistem em propor metodologias de previsão baseadas em redes neurais artificiais e Programação genética e avaliar seus erros de previsão para os períodos seco e úmido para as usinas de Água Vermelha e Itutinga. Comparar diferentes combinadores baseados em: redes neurais artificiais; programação genética, médias aritmética, média ponde-

rada, média geométrica, média quadrática e de modelo de regressão linear bayesiano analisando os erros de previsão para os períodos seco e úmido de séries de vazões para as usinas de Água Vermelha e Itutinga.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Este trabalho encontra-se dividido nos seguintes capítulos:

No Capítulo 2 são abordadas algumas definições importantes relacionadas ao significado de vazão, quais os fatores que a influenciam, como são geradas, qual a sua importância e porque há um elevado grau de dificuldade para realizar previsões para esse tipo de série.

O Capítulo 3 expõe a teoria dos modelos que serão utilizados neste trabalho para resolver o problema de previsão de séries de vazões e alguns estudos que fizeram esta tarefa de forma semelhante.

O Capítulo 4 descreve os fundamentos e metodologias utilizadas para a realização das previsões. São abordados os modelos como componentes individuais e em seguida as propostas de *ensembles*.

No Capítulo 5 encontram-se os resultados obtidos através dos modelos propostos.

No último capítulo são feitas as considerações finais sobre o trabalho e quais as conclusões, bem como a perspectiva de trabalhos futuros

2 SÉRIES DE VAZÕES E AFLUENTES

Nesta seção será abordado o problema de previsão de séries de vazões de uma forma geral. Na Subseção 2.1 será descrito o que são vazões, como são geradas, quais são os fatores que a influencia e porque são importantes. Na subseção seguinte será discutido sobre a previsão deste tipo de dado, o que é horizonte e a sua influência no processo de previsão e quais as dificuldades de realizar esta tarefa.

2.1 VAZÕES E AFLUENTES

A vazão representa o volume de água que atravessa uma seção transversal ao escoamento na unidade de tempo. Esse volume de água escoado na unidade de tempo é a principal grandeza a caracterizar o escoamento e suas unidades são normalmente expressas em m^3/s .

Segundo o Operador Nacional do Sistema Elétrico (ONS) (2016), o setor elétrico tem adotado o termo vazão natural para identificar a vazão que ocorreria em uma seção do rio, se não houvesse as ações antrópicas na sua bacia contribuinte, tais como regularizações de vazões realizadas por reservatórios, desvios de água, evaporações em reservatórios e usos consuntivos (irrigação, criação animal e abastecimentos urbano, rural e industrial). A vazão natural é obtida por meio de um processo de reconstituição, que considera a vazão observada no local e as informações relativas às ações antrópicas na bacia.

As séries de vazões são uma série temporal, ou seja, um conjunto de observações realizadas em períodos sucessivos no tempo, em que as observações vizinhas estão correlacionadas (MORETTIN; TOLOI, 2006). Estas observações podem ser coletadas em diferentes espaços de tempo, em geral, a discretização é diária, semanal ou mensal. As séries históricas são obtidas através de um processo baseado em medições e análises como mostra a Figura 2.

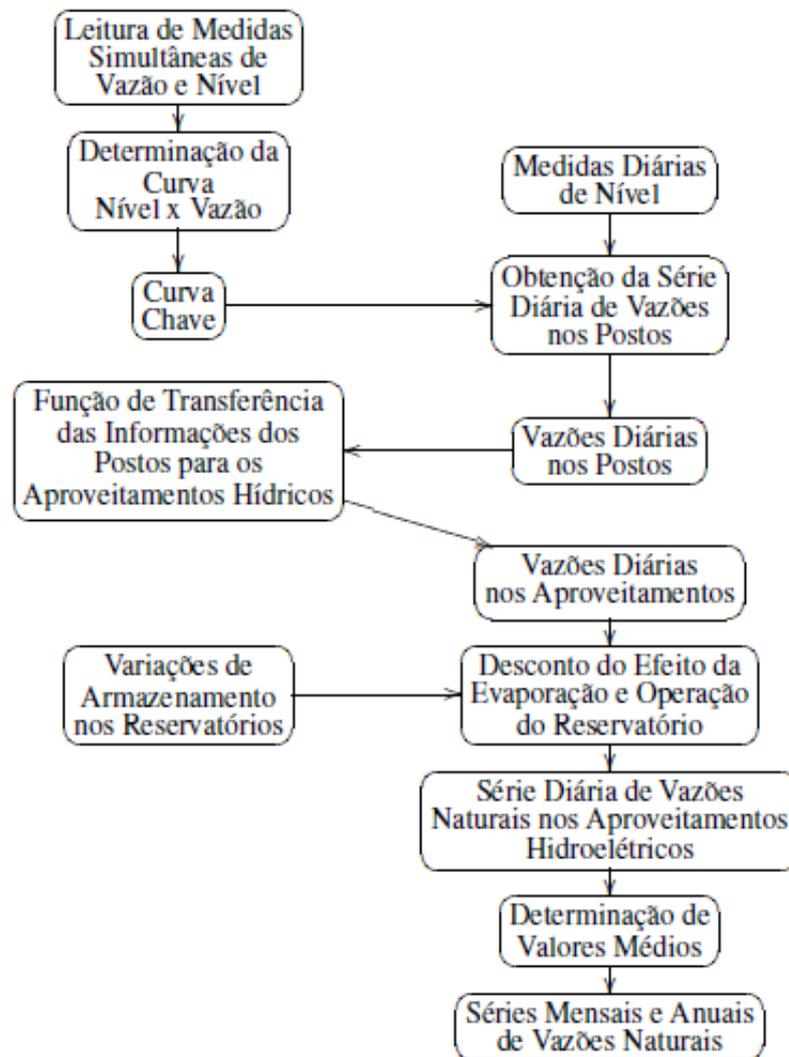


Figura 2 – Esquema ilustrativo da obtenção dos valores das séries históricas de vazões afluentes mensais e anuais.

Fonte: FORTUNATO et al. (1990).

Pode-se determinar a vazão de um curso d'água através de uma representação gráfica denominada hidrograma. As análises de hidrogramas são importantes principalmente nos períodos de cheias, pois é neste período que obras hidráulicas relacionadas com as enchentes e, em particular, no dimensionamento de canais, reservatórios, vertedores e bueiros se tornam mais evidentes.

De acordo com Tucci, Porto e Barros (1995), um grande número de fatores influenciam no comportamento da vazão e conseqüentemente no seu hidrograma e por isso devem ser considerados. Os mais importantes são: o relevo, a cobertura da bacia, a duração e intensidade da precipitação, o tipo e natureza do solo e o nível da umidade nele presente. Será comentado brevemente sobre cada um destes fatores:

a) Precipitação

A precipitação é um fator que tem influência direta no comportamento da vazão. A partir do início da chuva, tempo t_0 (FIGURA 3) há um intervalo de tempo para que a vazão comece a se elevar e isso está relacionado com o deslocamento da água na superfície do terreno (TUCCI; PORTO; BARROS, 1995). Preenchidas as depressões acumuladoras e excedida a capacidade de infiltração do solo, inicia-se o escoamento superficial³. A partir do ponto A, na Figura 3, tem-se a elevação constante da vazão até atingir o valor máximo. No ponto I da Figura 3 tem-se um ponto de inflexão que caracteriza o fim da contribuição do escoamento superficial direto e, conseqüentemente, o início da predominância da contribuição do escoamento subterrâneo⁴. Além da representação do hidrograma na Figura 3 tem-se também o hietograma, que é uma representação gráfica da intensidade da chuva ao longo de sua duração.

³O escoamento superficial é o segmento do ciclo hidrológico caracterizado pelo deslocamento da água na superfície da terra e nos cursos d'água naturais (TUCCI; PORTO; BARROS, 1995).

⁴Escoamento subterrâneo é oriundo do acumulo de água em aquíferos, é responsável pela alimentação do curso de água durante períodos de estiagem (TUCCI; PORTO; BARROS, 1995).

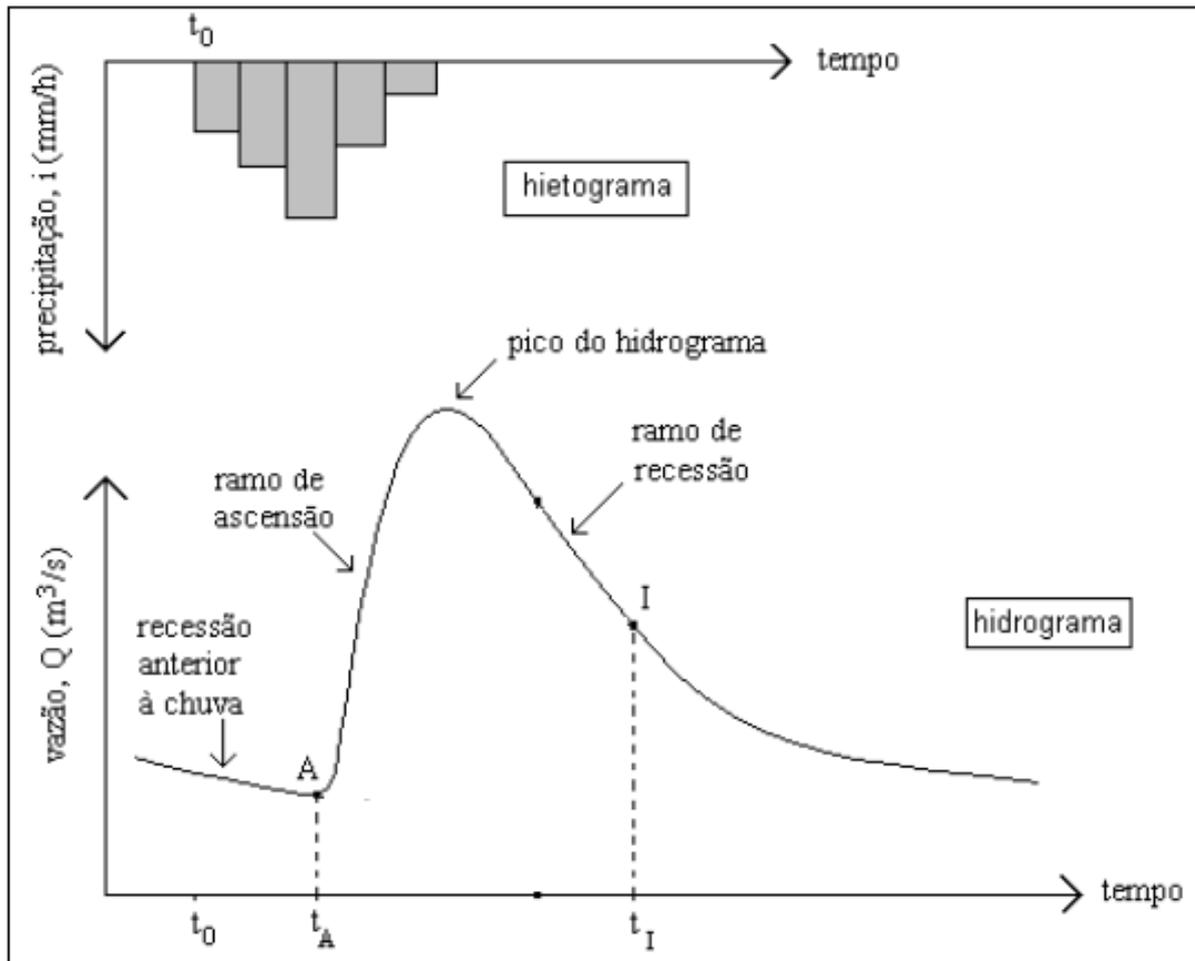


Figura 3 – Representação gráfica da vazão e precipitação através do hietograma e hidrograma e contribuição dos escoamentos superficial e de base para a vazão.

Fonte: TUCCI, PORTO e BARROS (1995).

A distribuição espacial da precipitação não é uniforme sobre toda a bacia. Em bacias hidrográficas pequenas ($A < 500\text{km}^2$), as precipitações de alta intensidade, pequena duração e atingindo pequena área são capazes de provocar grandes enchentes. Por outro lado, para bacias hidrográficas maiores, as precipitações mais importantes são as frontais, que atingem grandes áreas com intensidade média (TUCCI; PORTO; BARROS, 1995).

(b) Relevo

Uma característica de relevo que influencia no comportamento do hidrograma diz respeito ao formato da bacia. A forma da bacia hidrográfica é importante por influenciar no tempo de transição entre precipitação e escoamento. Isso é verificado por meio dos coeficientes (coeficiente de compactidade e coeficiente de forma) que permitem quantificar a influência da forma da bacia no modo de resposta à ocorrência de uma precipitação. Uma bacia de

formato radial concentra o escoamento, antecipando e aumentando o pico de vazão, já em uma bacia alongada ocorre o amortecimento da vazão, como ilustra a Figura 4 (TUCCI; PORTO; BARROS, 1995). Ao observar os hidrogramas na Figura 4, é possível observar que na bacia radial há a formação de pico, diferentemente da bacia alongada que possui uma curva mais suave.

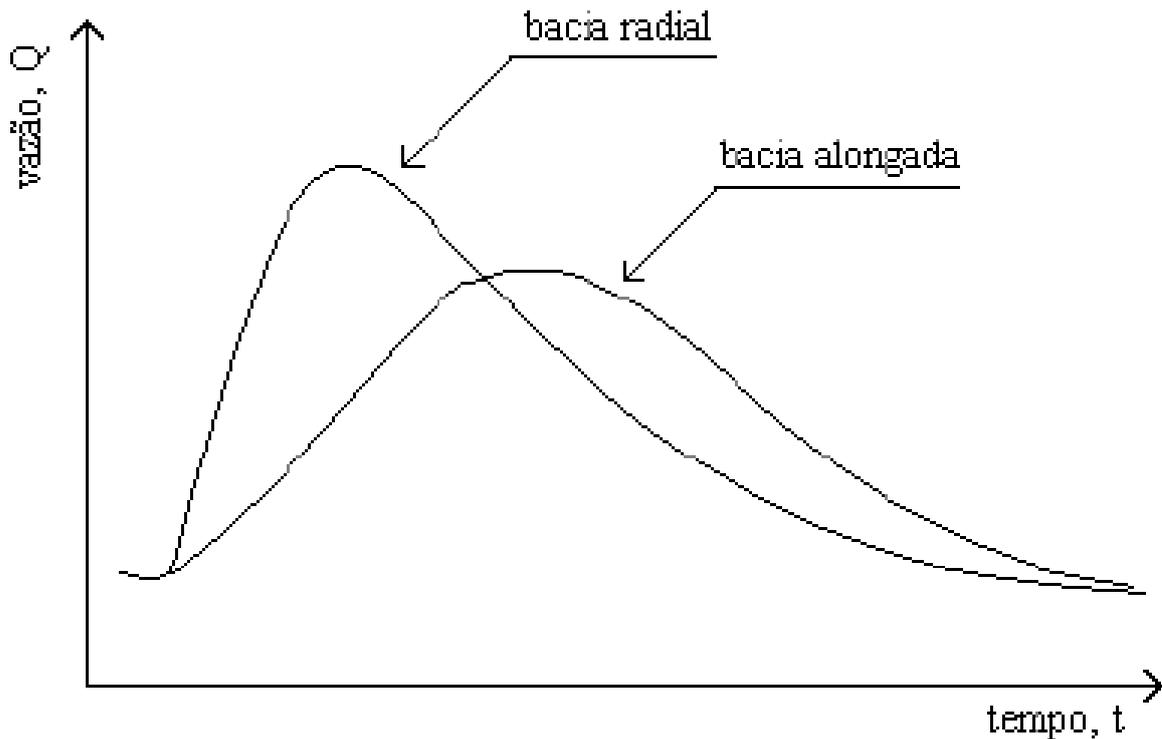


Figura 4 – Representação gráfica da vazão através dos hidrogramas comparativos de uma bacia radial e alongada.

Fonte: TUCCI, PORTO e BARROS (1995).

(c) Cobertura Vegetal

A cobertura vegetal influencia a vazão pois tende a retardar o escoamento superficial, facilita a infiltração e aumenta as perdas por evapotranspiração. Em bacias urbanas, onde a cobertura é constantemente alterada (a simples remoção da cobertura vegetal já torna a bacia mais impermeável), tem-se, assim, um aumento do volume do escoamento superficial e da vazão de pico (TUCCI; PORTO; BARROS, 1995). Essa diferença na vazão pode ser vista na Figura 5, onde pode-se ver que no hidrograma da bacia urbana a vazão é mais elevada que na bacia rural.

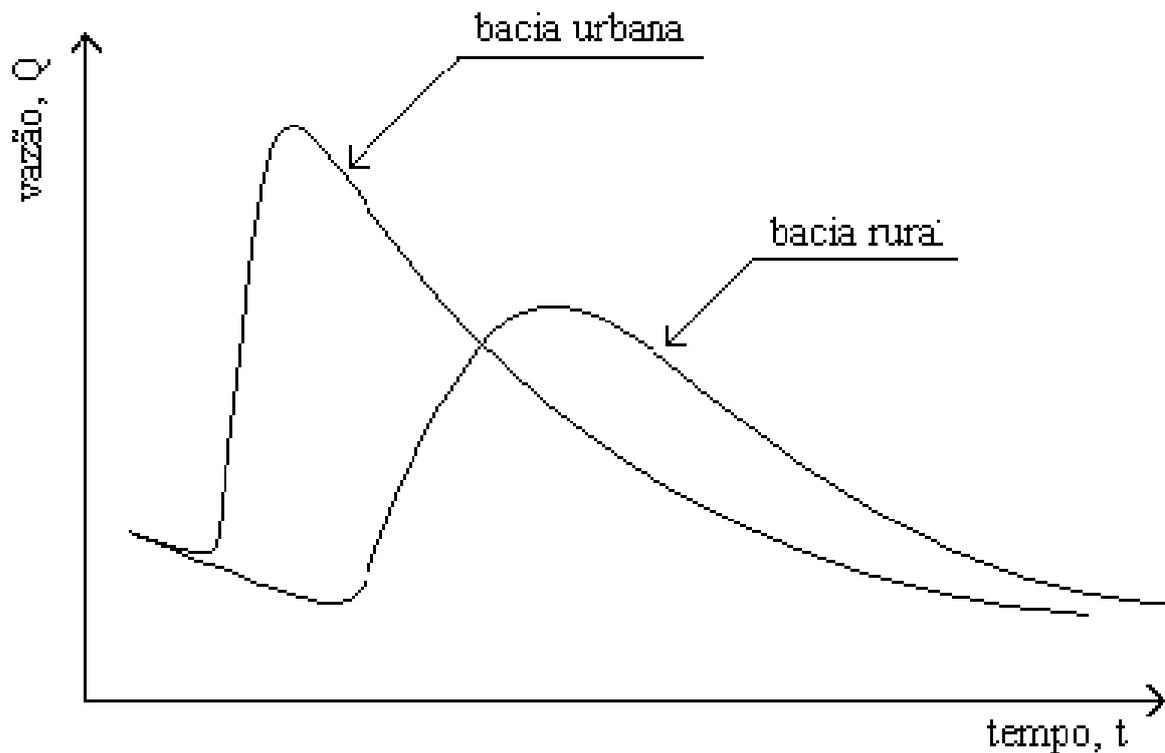


Figura 5 – Representação gráfica da vazão através dos hidrogramas comparativos de uma bacia rural e urbana.

Fonte: TUCCI, PORTO e BARROS (1995).

(d) Solo

O tipo, a natureza e o nível de umidade do solo têm influência na vazão, pois quando a umidade da camada superior do solo for pequena e o nível do lençol freático for baixo, boa parte da precipitação poderá ser retida, tornando o escoamento superficial e, portanto, a vazão reduzida (TUCCI; PORTO; BARROS, 1995).

Dentre todos estes fatores, deve-se atentar a contribuição destes para o comportamento ao longo do tempo. Como pode ser visto na Figura 6, que mostra a série para duas usinas do Rio Grande, a vazão apresenta um comportamento periódico, ou seja, uma oscilação entre um máximo e mínimo apresentando uma sazonalidade com um período de aproximadamente doze meses. Este comportamento está relacionado principalmente com o regime de precipitação. Em geral as chuvas de maior abundância ocorrem no verão, meses de novembro a abril, que é o período em que há um aumento na média da vazão, diferentemente dos meses de maio a outubro em que cessa o período úmido e há, conseqüentemente a diminuição no volume de água.

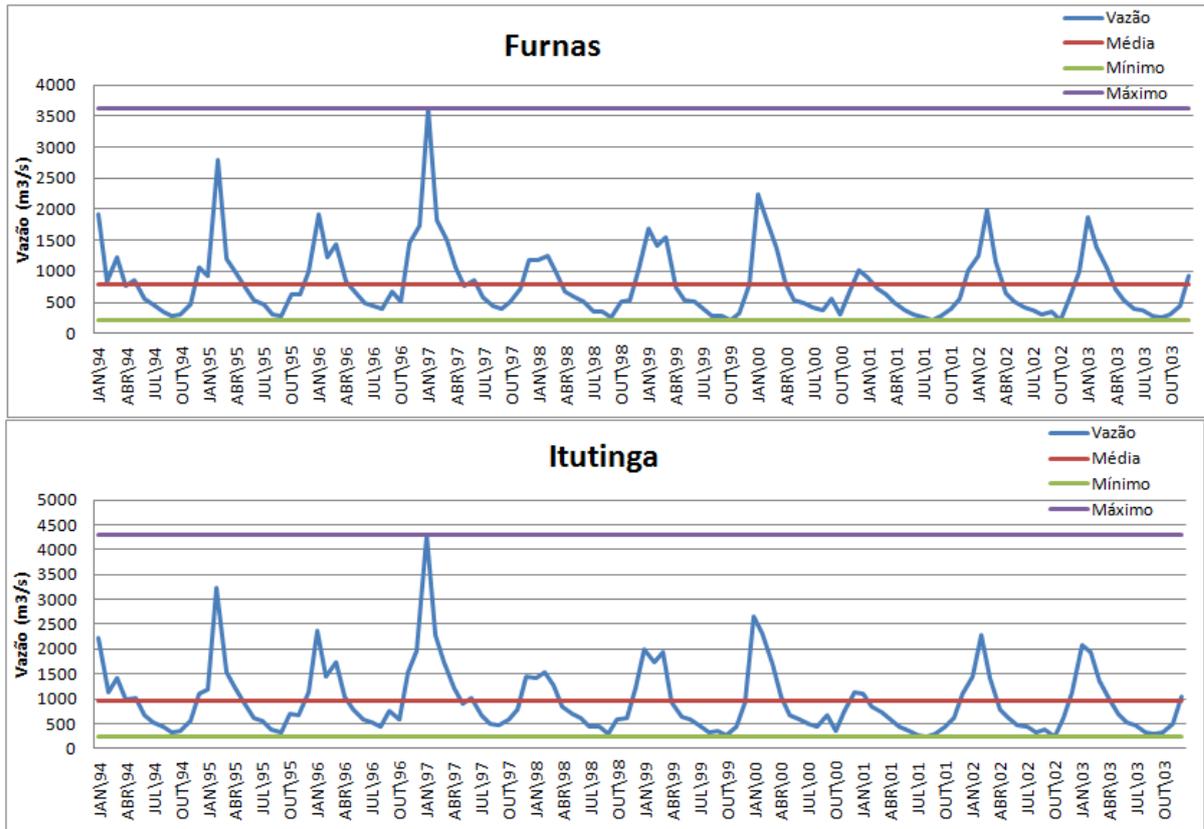


Figura 6 – Série de vazão afluente média mensal para as usinas hidroelétricas Furnas e Igarapava no período de 1994 a 2003.

Fonte: Própria do Autor.

A distribuição da vazão pode ser vista nos histogramas da Figura 7, na qual nota-se uma maior frequência a esquerda, sendo os valores elevados da vazão mais raros, o que caracteriza uma assimetria nos dados com uma cauda mais “pesada” a direita.

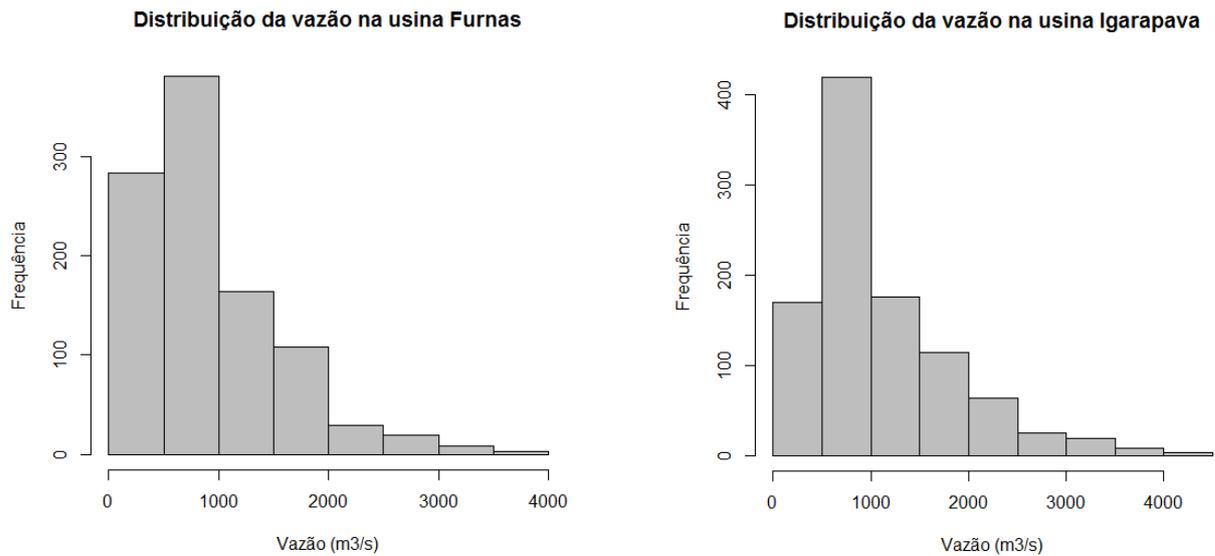


Figura 7 – Distribuição de frequência das séries de vazões para as usinas Furnas e Igarapava no período de 1931 a 2013.

Fonte: Própria do Autor.

Uma forma alternativa de analisar o comportamento da vazão é através da média de longo termo (MLT), que é a média aritmética das vazões naturais médias, correspondentes a um mesmo período, verificadas durante a série histórica de observações e é normalmente determinada para cada mês do ano, podendo, também, ser calculada para outros intervalos de tempo. A MLT da usina de Furnas e Igarapava podem ser vistas nas Figuras 8 e 9, na primeira a MLT é construída com os valores reais da série e na segunda com os dados padronizados, através da Equação 2.1:

$$\frac{x - \bar{x}}{\sigma}, \quad (2.1)$$

onde x é o valor real da série, \bar{x} é a média e σ é o desvio padrão.

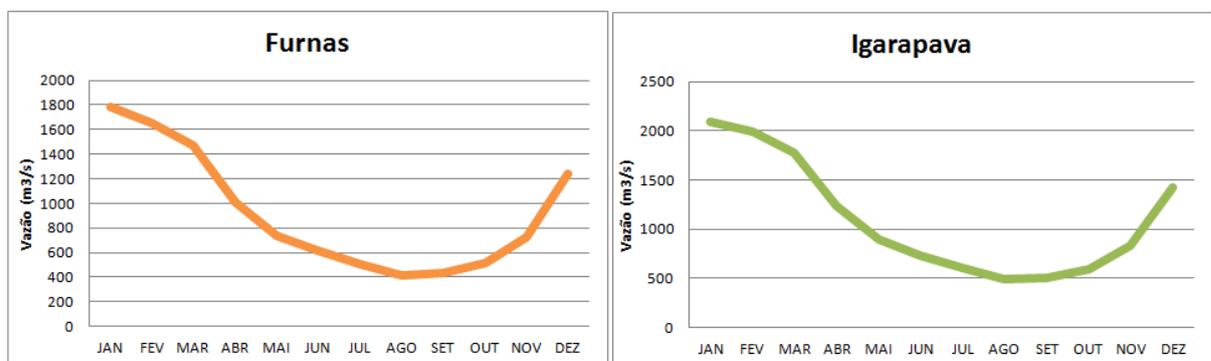


Figura 8 – Média de longo termo (MLT) das séries de vazões das usinas de Furnas e Igarapava (1931 a 2013).

Fonte: Própria do Autor.

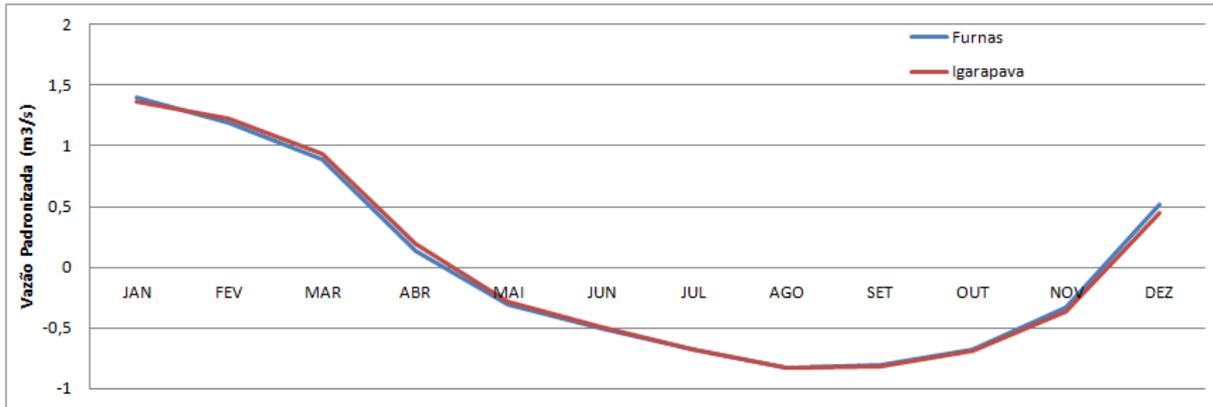


Figura 9 – Média de longo termo (MLT) das séries de vazões padronizadas para as usinas de Furnas e Igarapava (1931 a 2013).

Fonte: Própria do Autor.

A vazão, além de ser importante para o setor elétrico influencia também em outros aspectos como a pesca, a navegação, turismo e agricultura (BRAVO et al., 2008). Um período de seca ou cheia ao extremo pode afetar diretamente estes setores resultando em grandes prejuízos. Portanto, prever a vazão se torna uma tarefa de grande relevância para o planejamento econômico destas atividades.

2.2 PREVISÃO DE SÉRIES DE VAZÕES

A previsão de séries de vazões constitui o processo de estimar com base no histórico de vazões (vazão coletada no passado) um valor futuro da série. O objetivo de se realizar esta tarefa é principalmente o planejamento das operações nas usinas hidroelétricas, que consiste em substituir ao máximo o sistema termoelétrico pelo hidroelétrico, visto que possui um custo elevado e é altamente poluente.

Os modelos utilizados com o objetivo de prever valores desconhecidos são baseados em um histórico de dados passados. Um dos fatores importantes ao prever vazão é o horizonte de previsão, que é o momento futuro em que se deseja estimar o valor da série. Observando a Figura 10, apresentada por Soares Filho (1987), pode-se ver que o grau de incerteza da previsão cresce com a dimensão do horizonte. Para previsões de curto prazo, a incerteza é muito pequena (t_1), sendo crescente para previsões de médio e longo prazo, saturando após determinado tempo (t_2).

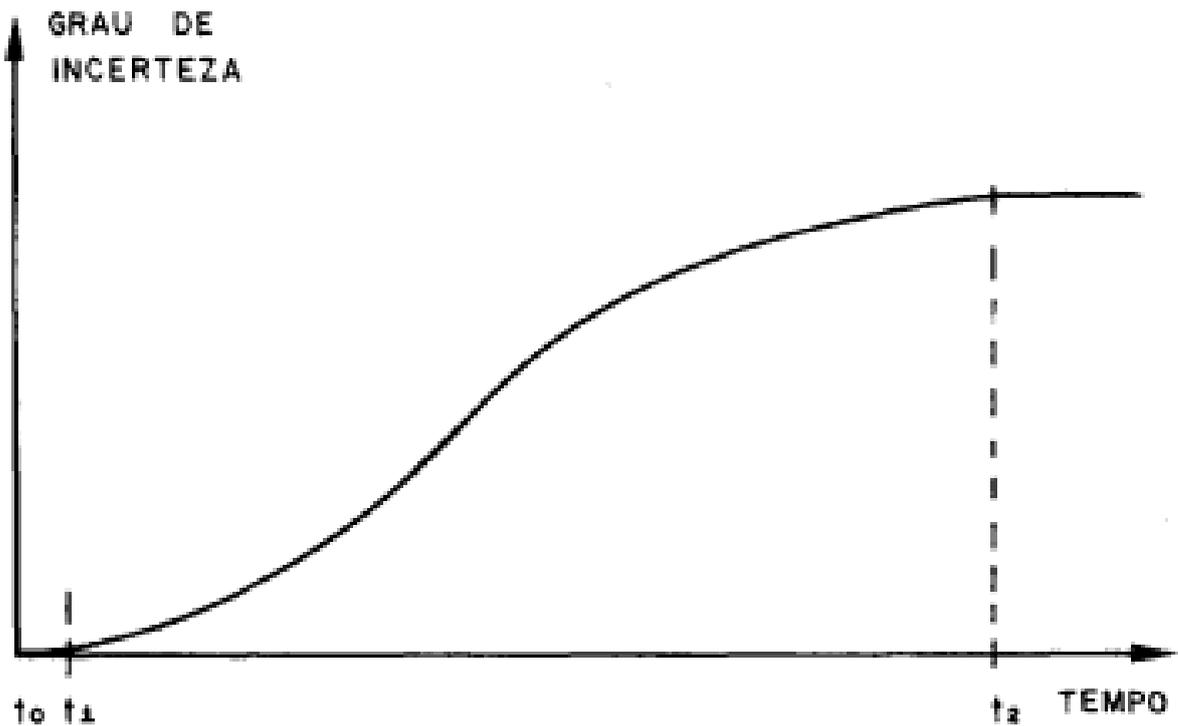


Figura 10 – Representação gráfica da variação do grau de incerteza da vazão quanto ao horizonte de previsão.

Fonte: SOARES FILHO (1987).

Outro fator que deve ser verificado ao se estudar uma série temporal é a correlação entre as observações que pode ser calculada através da Equação 2.2. O valor desta medida mostra a força da relação entre as observações. As estimativas do coeficiente de autocorrelação foram calculados e seus 29 primeiros valores podem ser vistos através do correlograma, na Figura 11. É possível observar que há uma oscilação de mesma frequência, esse tipo de comportamento é característico de uma uma série que apresenta sazonalidade, como foi visto também nos gráficos da Figura 6 da Subseção 2.2.

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (2.2)$$

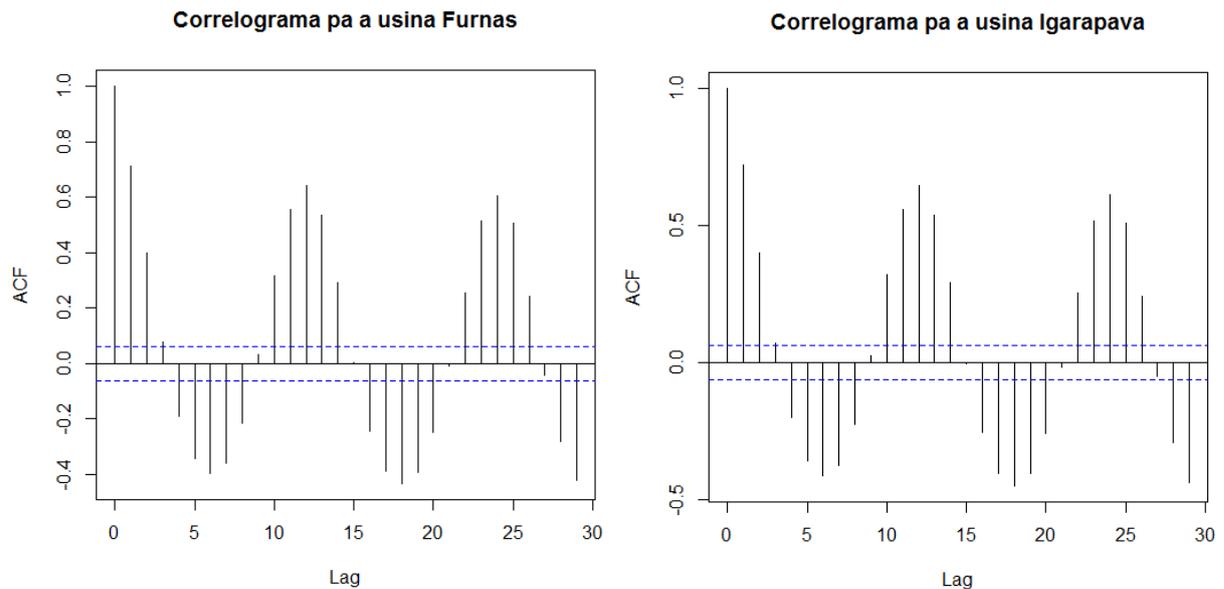


Figura 11 – Correlograma dos vinte e nove primeiros valores da correlação da série de vazões das usinas Furnas e Igarapava no período de 1931 a 2013.

Fonte: Própria do autor.

O tipo de usina e sua localização também podem ser fatores importantes para a realização das previsões e podem influenciar neste processo. Nas usinas de reservatório ocorre a regularização do rio, ou seja, durante os períodos de cheia, além da do fluxo da água utilizada para a produção de energia, pode ocorrer o acúmulo de água no reservatório. Durante os períodos secos, a água acumulada, além da decorrente do fluxo natural, é utilizada para gerar energia. Por isso, usinas com reservatórios de regularização têm como característica importante o fato que a produção de energia é mais constante (fator de capacidade elevado).

Nas usinas a fio d'água o reservatório tem dimensão reduzida, insuficiente para permitir a regularização do rio, ou seja, a produção de energia será determinada pelo fluxo de água disponível no período. Deste modo, nas épocas de cheias a usina produz muita energia e nos períodos de seca a produção é bastante reduzida, visto que não há capacidade de acúmulo.

O volume de água é mais constante em usinas que possuem reservatório e consequentemente a variabilidade é menor, ao contrário das usinas a fio d'água em que não há a possibilidade de estocar água aumento a variabilidade.

Na Figura 12 encontram-se as usinas que compõem a bacia do Rio Grande, das quais os dados de vazão de duas delas serão utilizadas neste estudo: Água Vermelha e Itutinga.

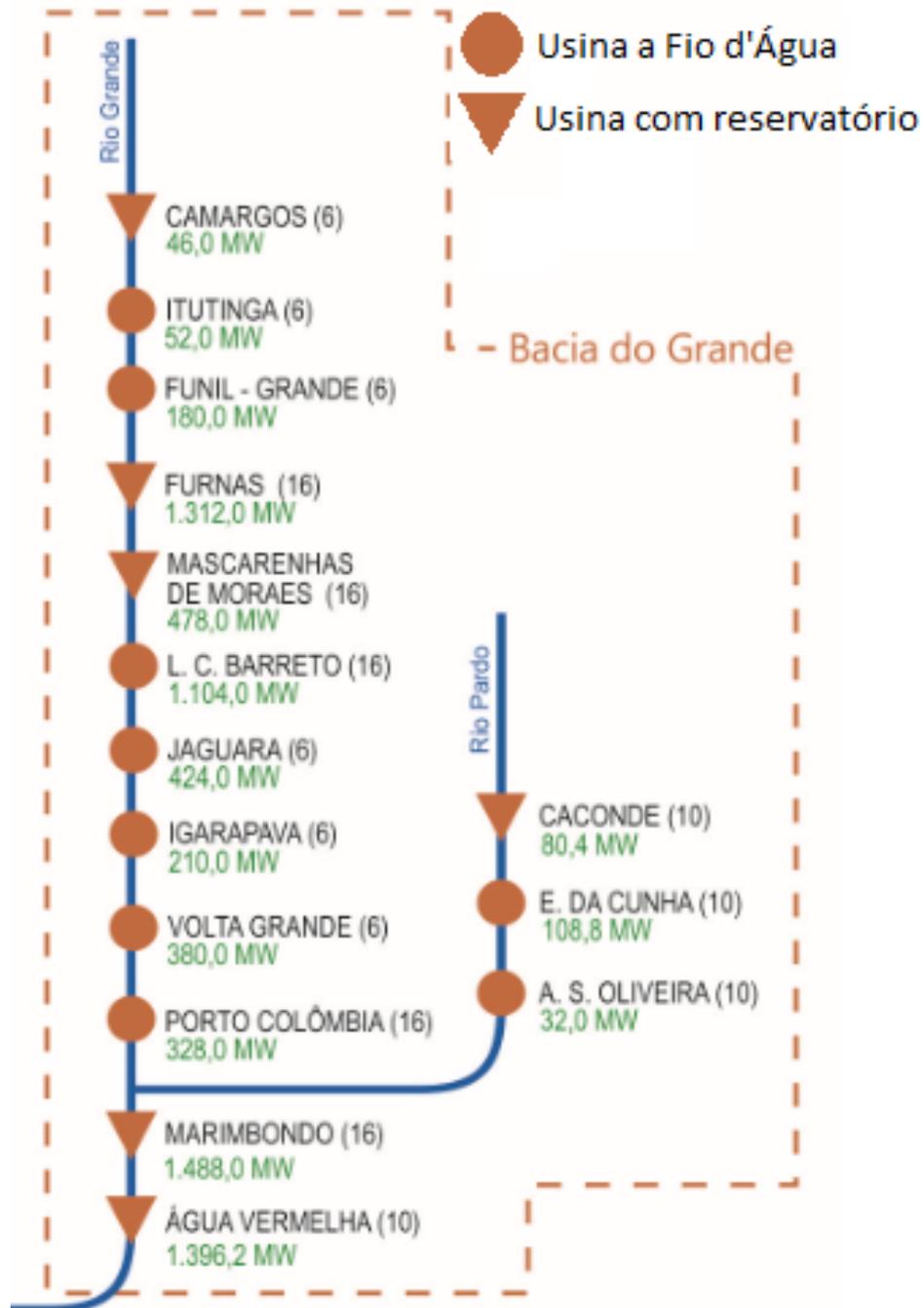


Figura 12 – Usinas hidroelétricas de reservatório e a fio d'água que compõem a bacia do Rio Grande e seu afluente Rio Pardo.

Fonte: ONS (2016).

Prever a vazão não é uma tarefa trivial. Como já discutido na seção anterior, existem diversos fatores naturais e antrópicos que variam ao longo dos anos, de modo que a vazão apresente alta variabilidade, com coeficiente de variação em torno de 40%, como pode ser visto na Tabela 2.

Tabela 2 – Medidas de posição para as séries de vazões mensais das usinas Furnas e Igarapava em m^3/s no período de 1931 a 2013.

Furnas						
Mês	Mínimo	Máximo	Amplitude	Média	Desv Pad	Coef. Var.
Janeiro	594,00	3621,00	3027,00	1781,64	695,45	39,03%
Fevereiro	357,00	3230,00	2873,00	1653,27	618,25	37,40%
Março	477,00	3757,00	3280,00	1470,22	578,58	39,35%
Abril	401,00	2327,00	1926,00	1010,20	343,57	34,01%
Maio	310,00	1572,00	1262,00	738,35	223,73	30,30%
Junho	276,00	2303,00	2027,00	615,84	237,35	38,54%
Julho	240,00	1308,00	1068,00	505,88	148,59	29,37%
Agosto	204,00	921,00	717,00	415,60	118,45	28,50%
Setembro	214,00	1889,00	1675,00	433,33	222,21	51,28%
Outubro	210,00	1822,00	1612,00	511,05	218,49	42,75%
Novembro	295,00	1989,00	1694,00	721,99	300,28	41,59%
Dezembro	334,00	3123,00	2789,00	1244,52	454,23	36,50%

Igarapava						
Mês	Mínimo	Máximo	Amplitude	Média	Desv Pad	Coef. Var.
Janeiro	722,00	4291,00	3569,00	2095,58	811,55	38,73%
Fevereiro	477,00	4106,00	3629,00	1993,68	746,95	37,47%
Março	628,00	4429,00	3801,00	1777,08	684,67	38,53%
Abril	514,00	2803,00	2289,00	1236,06	412,53	33,37%
Maio	408,00	1849,00	1441,00	891,41	267,15	29,97%
Junho	331,00	2605,00	2274,00	736,49	270,32	36,70%
Julho	277,00	1525,00	1248,00	603,16	175,08	29,03%
Agosto	239,00	1060,00	821,00	493,79	135,89	27,52%
Setembro	265,00	2028,00	1763,00	505,55	240,81	47,63%
Outubro	229,00	2125,00	1896,00	595,74	253,57	42,56%
Novembro	351,00	2327,00	1976,00	831,99	341,88	41,09%
Dezembro	413,00	3506,00	3093,00	1422,70	506,19	35,58%

Fonte: Própria do autor.

Pode-se observar também que o máximo da vazão, aconteceu no mês de Março (3757 m^3/s para Furnas e 4429 m^3/s para Igarapava) e o mínimo em Agosto e Outubro (204 m^3/s para Furnas e 229 m^3/s para Igarapava), o que reforça a relação entre a vazão e os períodos secos e úmidos. A Figura 13 evidencia também este fato, onde é possível ver a variação da média e do desvio padrão ao longo dos meses do ano.

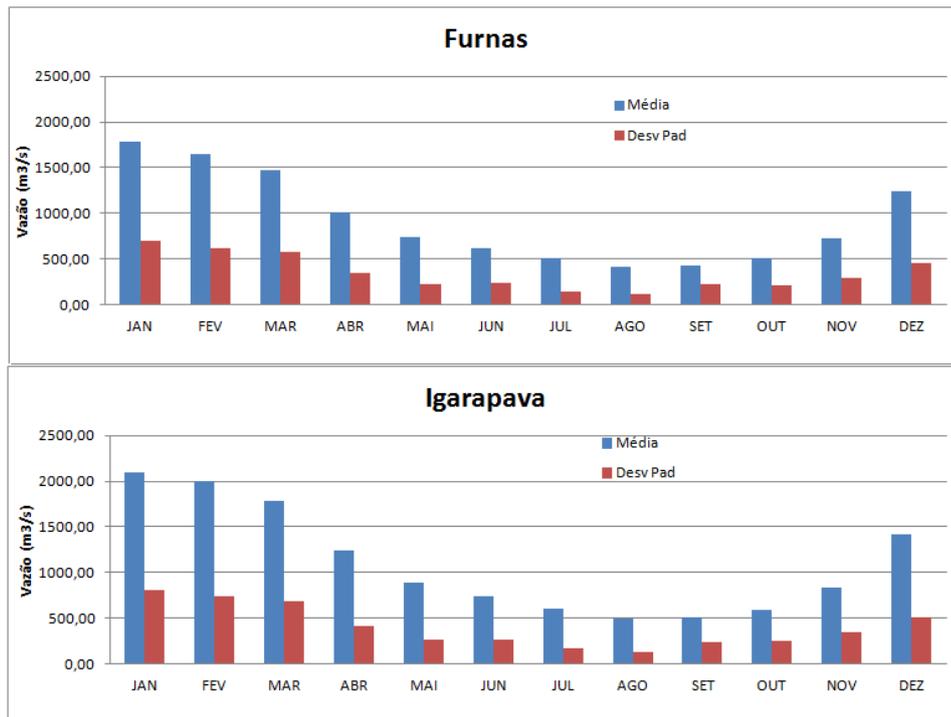


Figura 13 – Medidas de posição e variabilidade mensal em m^3/s das séries de vazões para as usinas Furnas e Igarapava no período de 1931 a 2013.
 Fonte: Própria do autor.

3 REFERENCIAL TEÓRICO

Esta seção tem como objetivo a descrição e análise da bibliografia e apontar os modelos mais abordados em previsão de vazões. Na Subseção 3.1 são mostrados os trabalhos relacionados com o problema de previsão de vazões e na Subseção 3.2 faz-se a apresentação das teorias envolvidas neste processo. Além disso, são apresentados os conceitos e formulações que contribuíram para a criação dos modelos de previsão de vazão propostos neste trabalho.

3.1 REVISÃO BIBLIOGRÁFICA

A previsão de séries de vazões é um típico problema de séries temporais que tem se tornado alvo de pesquisas à medida que aumenta a necessidade de planejamento. Mudanças climáticas vêm acontecendo e com isso há o aumento da preocupação com a quantidade de recursos e a qualidade destes que vem se tornando escassos.

A partir da década de 1990, os modelos baseados em RNA começaram a ser utilizados para as tarefas de previsão obtendo sucesso entre pesquisadores. Estes modelos fazem parte de uma área denominada inteligência artificial e são baseados na capacidade de aprendizado do cérebro humano. Algumas vantagens das RNA são: aprendem a partir de exemplos e não necessitam de conhecimento a priori sobre a modelagem do sistema a ser previsto, além disso, após ter aprendido as características de determinado conjunto de dados, as RNA tem uma alta capacidade de generalização.

Entre os trabalhos na área é possível destacar o artigo proposto por Gomes e Montenegro (2010) no qual foi apresentada uma metodologia para realizar previsão das vazões da bacia do Rio São Francisco seguindo um horizonte diário e semanal. Foram utilizadas redes neurais artificiais dos tipos *Multilayer Perceptron* (MLP) e *composite nonlinear sigmoidal regression blocks networks* (NSRBN). Neste trabalho, dados de estações pluviométricas, fluviométricas e de vazão natural do reservatório de Três Marias, do período de janeiro/1987 a dezembro/2004 foram empregadas. Os erros obtidos foram em média de 16,9% e 17,1% para as redes MLP e NSRBN respectivamente, considerando o horizonte de previsão diário, que é inferior em comparação com os erros obtidos por outros sistemas como Cemig e PrevivazH que mostram

um erro de 18,8% e 41,4%.

Outro estudo que propôs fazer previsão para a bacia do rio São Francisco foi apresentado por Cruz, Rodrigues e Versiani (2010). Neste trabalho foram usados dois modelos de previsão diária de vazão, um que utiliza as técnicas de RNA e diferenças primeiras da função de transferência (DPFT). Os dados utilizados no trabalho compõem uma série histórica diária cujo período e o horizonte de previsão não foram explicitados. A metodologia DPFT apresentou um desempenho ligeiramente superior ao modelo RNA.

Kisi (2009) propôs a comparação do modelo *wavelet* conjunto, que é a combinação entre os modelos *wavelet* e RNA. A série era diária e os dados eram da região Thrace, na Turquia. Os resultados do modelo sugerido aumentaram significativamente a precisão das previsões quando comparados ao modelo neural individual, o erro médio absoluto reduziu em 31% em relação a RNA.

Besaw, Rizzo e Bierman (2010) usaram RNA para fazer previsão de vazões diárias e por hora, os dados utilizados foram da bacia do rio Winooski, localizada no noroeste de Vermont no EUA. O método mostrou bom desempenho com erros em torno de 20%.

Bravo et al. (2008) fizeram uso de RNA para estimar a vazão de afluência ao reservatório de Furnas (Minas Gerais, Brasil) com base em previsões quantitativas de chuvas do modelo regional ETA. As previsões das vazões realizadas possuem um horizonte de doze dias com intervalo diário. Os resultados levam à conclusão de que considerar as previsões das chuvas melhoram as previsões da vazão, com erros próximos a 11% e 13%.

RNA do tipo MLP foram empregadas por Lima, Lima e Salgado (2011) para fazer previsão um passo a frente e n passos, os dados se referem à quatro usinas do Rio Grande sendo elas: Camargos, Itutinga, Funil Grande e Furnas. Os erros obtidos foram em torno de 19%.

Assim como as RNA, a PG é uma área da inteligência artificial que vem ganhando espaço como meio de fazer previsões para séries de vazões. Essa metodologia é baseada na seleção natural de Darwin (1859) e foi desenvolvida por Koza (1989).

Como destaque na área pode ser citado o trabalho de Wang et al. (2009) no qual a técnica supracitada é utilizada para fazer previsão de séries de vazões a longo prazo no rio Manwan. A série histórica utilizada é mensal e corresponde ao período de janeiro de 1953 a dezembro de 2003. O desempenho da previsão fazendo uso de PG foi comparado com as RNA, sendo que a PG apresentou um erro relativo médio 7,97% menor que a RNA.

A técnica de *Support Vector Machines* (SVM) que é uma técnica de aprendizado de máquina vem recebendo crescente atenção nos últimos anos. As SVMs vem sendo utilizadas em diversas tarefas, dentre estas a previsão de séries de vazões. Esta técnica foi empregada por Kisi e Cimen (2011) em conjunto com *wavelet* para fazer previsão de vazões. Foram utilizados dados da estação Gerdelli no Rio Canakdere e Estação Isakoy no Rio Goksudere na região oriental do Mar Negro da Turquia. A série utilizada é mensal e corresponde ao período de 1960 a 1999. Os resultados mostraram que a combinação dos dois métodos melhorou a precisão da previsão, considerando que a raiz do erro médio relativo foi de 15,7% para 13,9%.

Makkeasorn, Chang e Zhou (2008) fizeram uso da PG para fazer previsão de uma série de vazão diária, na bacia hidrográfica do semi-árido do sul do Texas, considerando um horizonte de curto prazo. Os resultados da PG foram comparados com o modelo de RNA, e o primeiro mostrou melhores resultados.

Os modelos de Box e Jenkins (1976), encontrados em diversos trabalhos, são processos estocásticos utilizados para descrever uma série temporal. A construção destes modelos depende de vários fatores, tais como o comportamento do fenômeno em estudo ou o conhecimento *a priori* que temos da sua natureza e o objetivo da análise.

O trabalho de Bayer e Castro (2012) propõe uma metodologia baseada no método de Box e Jenkins para fazer previsão das vazões no rio Potiribu que é um afluente da margem esquerda do rio Ijuí que, por sua vez, é afluente do rio Uruguai, para tal adotou-se uma série histórica de médias mensais obtida no Sistema de Informações Hidrológica (HidroWeb). Foi aplicado o modelo sazonal autorregressivo integrado de médias móveis (SARIMA) para fazer previsões de seis meses à frente e os erros de previsão foram na ordem de 16%.

Fadiga, Lopes e Martins (2008) utilizaram modelação hidrológica e modelos estocásticos para fazer previsão de séries de vazões naturais na usina de Itaipu. Após a obtenção das previsões de cada modelo, foi utilizado uma combinação linear dos resultados individuais. O erro relativo médio (ERM) variou entre 18% e 25%.

Andrade et al.(2012) realizaram um estudo de séries de vazões médias mensais, abordando o modelo autorregressivo periódico (PAR) e considerando um horizonte de previsão que variou de 1 a 12 meses. As previsões foram realizadas para três usinas hidroelétricas: Furnas, Emborcação e Sobradinho. Os resultados do trabalho mostram que meses que apresentam maior coeficiente de variação, como por exemplo janeiro, fevereiro e março, são mais difíceis de serem previstos e por isso apresentam erros de previsão mais elevados. O desempenho dos modelos

foram medidos através da razão entre a raiz quadrada do erro quadrático médio (RMSE) e a média de longo termo (MLT), para os meses mais húmidos os erros foram em torno de 40% e para os meses secos em torno de 20%.

Lucas et al. (2009) aplicaram um modelo hidrológico determinístico mensal de dois parâmetros e o modelo estocástico autorregressivo integrado de médias móveis (ARIMA) para simular a vazão mensal das bacias do Xingu, no estado do Pará. O modelo estocástico representou de maneira satisfatória a dinâmica da série.

O trabalho de Modarres e Eslamian (2006) mostrou o desempenho do modelo multiplicativo autorregressivo integrado e de médias móveis para realizar previsão de vazões de janeiro de 2000 a dezembro de 2001, no rio Zayandehrud. Foram testados dois modelos, sendo que o $ARIMA(1,1,0) \times (0,1,1)$ mostrou bom desempenho.

Uma técnica que pode ser bastante útil no ramo de previsão e que vem sendo utilizada são os *ensembles*, que consiste em combinar previsões através de uma equação ou modelo. Um exemplo pode ser visto em Moreira et al. (2011), onde foram utilizados modelos inteligentes para fazer previsão de carga e em seguida as previsões obtidas desses modelos foram combinadas através de três combinadores: média aritmética, média ponderada e um modelo não-linear baseado em RNA. Os resultados deste estudo mostraram que através de *ensembles* é possível obter resultados mais precisos do que quando são consideradas as componentes individuais, reduzindo os erros em torno de 0,15%, 0,99% e 0,2%.

Salgado et al.(2006) também fizeram uso de *ensemble* no mesmo contexto de previsão de energia. As componentes individuais foram as RNA e SVM. Após obtidas as componentes individuais, estas foram combinadas através da própria RNA. O erro de previsão obtido pelo combinador foi de 25% menor que a melhor componente individual.

3.2 METODOLOGIAS DE PREVISÃO

Nesta subseção será apresentado uma breve revisão de conceitos relevantes que foram utilizados neste trabalho.

3.2.1 Redes neurais artificiais (RNA)

Os trabalhos realizados na área de Inteligência Artificial têm como objetivo capacitar computadores para que estes executem tarefas baseadas no conhecimento, raciocínio e inteligência humana. Os sistemas inteligentes são capazes de manipular símbolos que representem entidades do mundo real e por essa razão eles são capazes de trabalhar com grande eficiência com o conhecimento (REZENDE, 2005).

As RNA é um destes modelos baseados no funcionamento do cérebro humano e é um sistema computacional constituído por unidades conhecidas como neurônios. Os neurônios são elementos processadores interligados, trabalhando em paralelo para desempenhar uma determinada tarefa. Os modelos RNA's constituem uma importante técnica estatística não-linear capaz de resolver uma gama de problemas de grande complexidade. As principais áreas de atuação são para classificação de padrões e previsão. Segundo Zang, Patuwo e Hu (1998) esta metodologia tem sido utilizada em diversas áreas do conhecimento, dentre estas, negócios, indústria e ciências.

O neurônio é a unidade sensorial de uma RNA e um modelo geral do neurônio artificial pode ser representado pela Figura 14, onde x_1, x_2, \dots, x_n são as entradas fornecidas à rede, estas são ponderadas através dos pesos ou parâmetros w_1, w_2, \dots, w_n que representam a "memória" da rede. Isto é, a experiência ganha como resultado das n-apresentações dos padrões. São os pesos que combinam a não-lineariedade para que a mesma fique distribuída pela rede. Após a ponderação das observações, como mostra a Equação 3.1, estas vão passar pela função de ativação $g(\cdot)$ resultando em uma saída y .

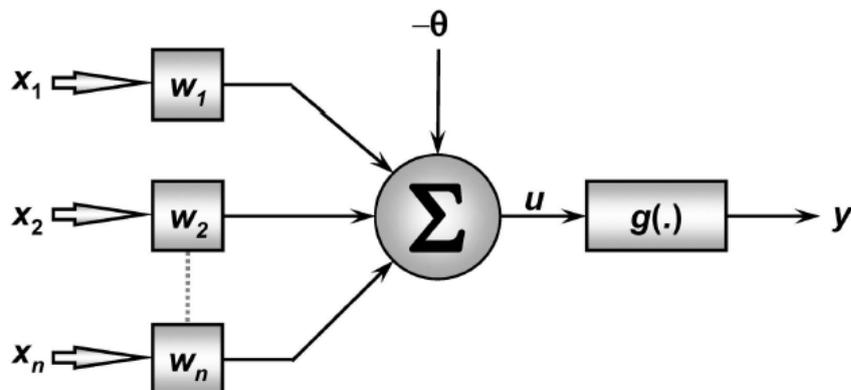


Figura 14 – Representação ilustrativa de um neurônio artificial com X_n padrões de entradas, uma função de ativação $g(\cdot)$ e uma saída y

Fonte: Própria do autor.

$$u = \sum_{i=1}^n w_i * x_i - \theta, \quad (3.1)$$

onde θ é o bias, que é um neurônio especial que permite uma melhor adaptação ao conhecimento fornecido à rede.

Função de Ativação

A função de ativação é geralmente utilizada com dois propósitos: restringir a amplitude do intervalo de saída do neurônio e introduzir a não linearidade no modelo. Sua escolha está relacionado com o processo de normalização dos dados, uma abordagem através da qual é possível padronizar os dados dentro de um limite pré-definido.

Por exemplo se os dados passaram pela normalização que restringiu os valores no intervalo $[-1; 1]$ faz sentido usar a função tangente hiperbólica, pois seu contra-domínio também assume valores neste intervalo. Esta função pode ser vista na Equação 3.2 e seu gráfico na Figura 15.

$$g(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} \quad (3.2)$$

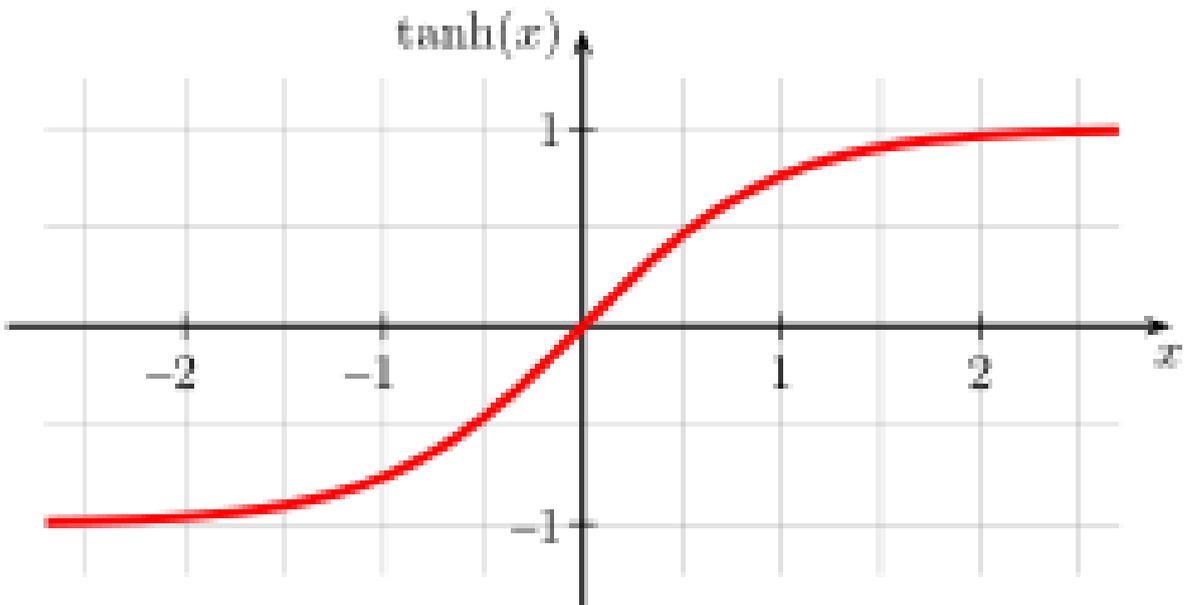


Figura 15 – Representação gráfica da função de ativação tangente hiperbólica mostrando o contra-domínio restrito no intervalo $[-1;1]$.

Fonte: Própria do autor.

Arquitetura da RNA

Existem basicamente dois tipos de RNA: as redes recorrentes e as não recorrentes. Neste trabalho serão abordadas apenas as não recorrentes que são aquelas que não possuem realimentação de suas saídas para as suas entradas e, por isso, são ditas “sem memória”. A estrutura dessas redes pode ser formada por uma camada única ou por multi-camadas. No caso de redes em camadas existe um conjunto de neurônios de entrada, uma camada de saída e uma ou mais camadas intermediária ou oculta.

Neste trabalho, serão utilizadas as RNA do tipo *Multilayer Perceptron* (MLP), que é uma das estruturas mais utilizadas atualmente em problemas de previsão. Pode também ser chamada de rede *Feedforward*, cujos dados fluem sempre para frente (no sentido das setas). Na Figura 16 tem-se uma representação de uma RNA MLP com duas camadas intermediárias.

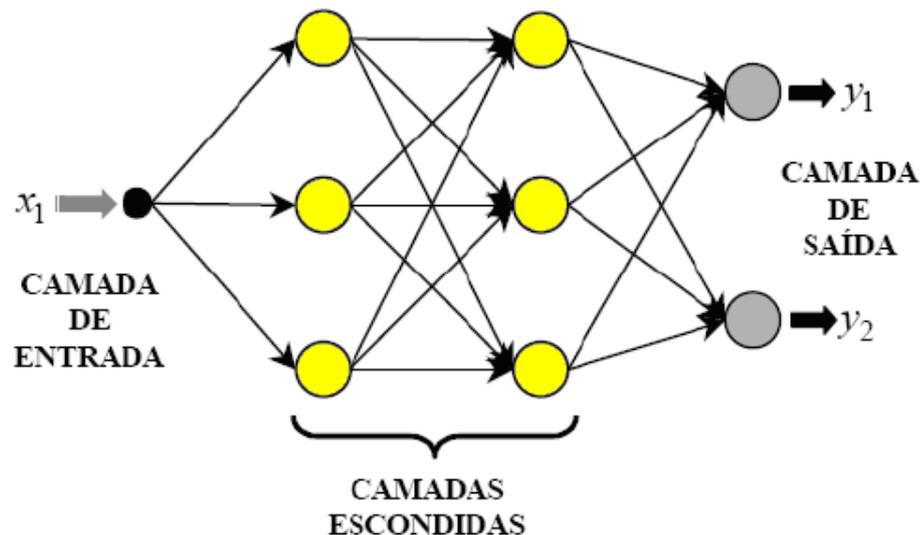


Figura 16 – Representação ilustrativa de uma rede neural artificial do tipo *Multilayer Perceptron* com duas camadas intermediárias.

Fonte: Própria do autor.

O número de neurônios nas camadas de entrada/saída depende da dimensão do vetor de entrada x e saída y que varia para cada conjunto de dados. Já para a camada intermediária pode-se determinar o número de neurônios empiricamente (tentativa e erro) bem como através do Teorema de Kolmogorov.

Teorema de Kolmogorov-Nielsen Dada uma função contínua arbitrária $f : (0,1)^n \rightarrow \mathbb{R}^m, f(x) = y$ existe sempre para f , uma implementação exata com uma rede neural de três camadas, sendo a camada de entrada um vetor de dimensão n , a camada oculta composta por $(2n + 1)$ neurônios, e a camada de saída com m neurônios representando as m componentes

do vetor y (KOVÁČZ, 1997).

Treinamento das RNA's

Diz-se que uma RNA aprende quando ela alcança, através de algoritmos de treinamento, um conjunto de pesos (ou parâmetros) que minimize o erro de classificação do conjunto de dados de treinamento (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989). O erro a ser minimizado é a função da diferença entre o valor desejado e a previsão que a rede está dando com o conjunto de parâmetros de cada ciclo, como mostra a Equação 3.3:

$$\varepsilon = \frac{1}{2} \sum_{k=1}^{n_3} (\text{desejado}_k - \text{calculado}_k)^2, \quad (3.3)$$

onde n_3 é o número de neurônios da camada de saída da rede.

O aprendizado relaciona-se a maneira pela qual a modificação dos parâmetros ocorre. Neste processo os parâmetros são ajustados através de um processo estimulado pelo ambiente no qual a rede está inserida. O algoritmo de aprendizagem refere-se a regras bem definidas para a solução de um problema de aprendizado e diferem entre si pela forma como se dá o ajuste dos pesos.

Os procedimentos de treinamento e aprendizado podem ser classificados em dois tipos: supervisionado e não supervisionado. Neste trabalho será abordado apenas o primeiro: os exemplos utilizados para o treinamento são do tipo "rotulados", ou seja, existe um vetor de entrada e um vetor alvo específico que se deseja na saída. No processo aplica-se o vetor de entrada e, então, a saída fornecida pela rede é comparada com o vetor de resposta desejado. Dessa comparação obtém-se um sinal de erro que é realimentado através da rede e os pesos são atualizados segundo um algoritmo determinado a fim de minimizar este erro. Existem diversos algoritmos de treinamento para o aprendizado supervisionado e alguns deles serão abordados neste trabalho.

Algoritmos de Treinamento

O processo de treinamento supervisionado de redes neurais artificiais multicamadas é equivalente a um problema de otimização não-linear irrestrito, onde uma função de erro global é minimizada a partir do ajuste de parâmetros da rede neural (pesos).

Os algoritmos utilizam somente o gradiente local da função ou então métodos que utilizam também as derivadas de segunda ordem. No primeiro caso, a função é aproximada

pelo primeiro (constante) e segundo (linear) termos da expansão de Taylor; no segundo caso, o terceiro termo (quadrático) também é considerado, (Equação 3.4).

$$g(x) = g(x^*) + \nabla g(x)^T|_{x=x^*}(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 g(x)^T|_{x=x^*}(x - x^*) + \dots \quad (3.4)$$

Em problemas de aproximação utilizando um número finito de dados amostrados e definido um modelo de aproximação $\hat{g}(\cdot, \theta)$, a distância entre a função a ser aproximada e sua aproximação $dist(g(\cdot), \hat{g}(\cdot, \theta))$ é uma função apenas do vetor de parâmetros $\theta \in \mathbb{R}^p$. Tomando a norma euclidiana como a medida de distância, produz-se a seguinte expressão (Equação 3.5):

$$J(\theta) = \frac{1}{N} \sum_{l=1}^n (g(x) - \hat{g}(x, \theta))^2. \quad (3.5)$$

O problema de aproximação passa a ser um problema de otimização cuja solução é o vetor $\theta^* \in \mathbb{R}^p$ que minimiza $J(\theta)$, ou seja,

$$\theta^* = \operatorname{argmin} J(\theta). \quad (3.6)$$

Para a maioria dos modelos de aproximação $\hat{g}(\cdot, \theta)$, o problema de otimização tem a desvantagem de ser não-linear e não-convexo. Estas características impedem a existência de uma solução analítica, mas permitem obter processos iterativos de solução, a partir de uma condição inicial θ_0 , na forma:

$$\theta_{i+1} = \theta_i + \alpha_i d_i \quad (3.7)$$

onde $\theta_i \in \mathbb{R}^p \rightarrow$ é o vetor de parâmetros, $\alpha_i \in \mathbb{R}^+ \rightarrow$ é um escalar que define o passo do ajuste e $d_i \in \mathbb{R}^p \rightarrow$ é a direção de ajuste.

A distinção entre cada um dos algoritmos procede no cálculo da direção de ajuste dos parâmetros e do passo a a cada iteração permite estabelecer distinção entre eles.

Os algoritmos podem ser classificados baseado na ordem da informação que eles devem calcular, ou seja, a ordem das derivas da Equação 3.5. Aqueles que fazem uso de derivada primeira são denominados métodos de primeira ordem e os de segunda ordem fazem uso de derivada segunda.

Métodos de Primeira Ordem.

O erro quadrático médio a ser minimizado pode ser apresentado considerando-se seus termos até segunda ordem:

$$J_{quad}(\theta) = J(\theta_i) + \nabla J(\theta_i)^T(\theta - \theta_i) + (\theta - \theta_i)^T \nabla^2 J(\theta_i)(\theta - \theta_i) \quad (3.8)$$

onde $\nabla J(\theta_i) \rightarrow$ é o vetor gradiente e $\nabla^2 J(\theta_i) \rightarrow$ é a matriz hessiana.

ambos calculados no ponto $\theta = \theta_i$ e $J_{quad}(\theta)$ representa a aproximação até segunda ordem de $J(\theta)$. Nos métodos de primeira ordem apenas os termos constante e linear em θ da expansão em série de Taylor são considerados.

Algoritmo de Retropropagação (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989)

Quando a rede está em um estado θ_i , o gradiente $\nabla J(\theta_i)$ é calculado e um passo de minimização na direção oposta ao gradiente $d = -\nabla J(\theta)$ é efetuado. A regra de aprendizado é dada pela Equação 3.9:

$$\theta_{i+1} = \theta_i + \alpha_i d_i + \beta_i \Delta \theta_{i-1}, i \geq 0. \quad (3.9)$$

No algoritmo padrão, a minimização é geralmente feita com um passo α fixo. A determinação do passo α é fundamental, pois para valores muito baixos, o tempo de treinamento pode tornar-se exageradamente alto, e para valores muito altos os parâmetros podem divergir (HAYKIN, 1994). A velocidade de convergência é geralmente melhorada adicionando-se um termo de momento (RUMELHART et. al., 1986).

Método Gradiente (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989)

O método gradiente é também um método de primeira ordem e o mais simples para a obtenção da direção d_i , que na i -ésima iteração é definida como a direção de módulo unitário de maior decrescimento da função J :

$$d = -\frac{\nabla J(\theta)}{\|\nabla J(\theta)\|}. \quad (3.10)$$

A lei de ajuste do método do gradiente é, então, dada por:

$$\theta_{i+1} = \theta_i - \alpha_i \frac{\nabla J(\theta)}{\|\nabla J(\theta)\|}. \quad (3.11)$$

Métodos de Segunda Ordem.

Os métodos de segunda ordem são considerados a maneira mais eficiente de se fazer o treinamento de redes neurais do tipo MLP (SHEPHERD, 1997). Estes métodos utilizam um procedimento de busca unidimensional para encontrar um α_i que seja solução ótima do problema.

Davidon Fletcher Powell (DFP) (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989)

Este método é classificado como método quase-Newton e consiste em fazer uma aproximação iterativa da inversa da matriz hessiana de forma que:

$$\lim_{i \rightarrow \infty} H_i = \nabla^2 J(\theta)^{-1}. \quad (3.12)$$

A construção da inversa se dá da seguinte forma:

$$H_{i+1} = H_i + \frac{p_i p_i^T}{p_i^T q_i} - \frac{H_i q_i q_i^T H_i}{q_i^T H_i q_i} \quad (3.13)$$

onde

$$p_i = \alpha_i d_i \quad (3.14)$$

e

$$q_i = \nabla^2 J(\theta) p_i. \quad (3.15)$$

Broyden Fletcher Goldfarb Shanno (BFGS) (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989)

A diferença básica deste método para o anterior (DFP) é a forma da aproximação da inversa da matriz hessiana. A expressão que permite determinar a aproximação da inversa da

hessiana é apresentada na Equação 3.16:

$$H_{i+1} = H_i + \frac{p_i p_i^T}{p_i^T q_i} \left[1 + \frac{q_i^T H_i q_i}{p_i^T q_i} \right] - \frac{H_i q_i p_i^T + p_i q_i^T H_i}{p_i^T q_i}. \quad (3.16)$$

Os vetores p_i e q_i são determinados pelas Equações 3.14 e 3.15.

Polak Ribière (PR) (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989)

Estes métodos também utilizam informações de segunda ordem para escolher a direção de busca d_i , o passo α_i e o termo momento β_i . E são baseados no método para tratar problemas quadráticos:

$$J(\theta) = \frac{1}{2} \theta^T Q \theta - b^T \theta. \quad (3.17)$$

A lei de ajuste do método das direções conjugadas é dada por:

$$\theta_{i+1} = \theta_i - \frac{d_i^T \nabla J(\theta_i)}{d_i^T Q d_i} \quad (3.18)$$

onde $d_0 = -\nabla J(\theta_0)$, $d_{i+1} = -\nabla J(\theta_{i+1}) + \beta_i d_i$, $\beta_i = \frac{\nabla J(\theta_{i+1})^T Q d_i}{d_i^T Q d_i}$ e a matriz Q é aproximada pela matriz hessiana calculada no ponto θ_i .

Uma dessas aproximações é dada pelo método PR:

$$\beta_i = \frac{g_{i+1}^T (g_{i+1} - g_i)}{g_i^T g_i}. \quad (3.19)$$

Fletcher & Reeves (FR) (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989)

Outra forma de realizar a aproximação da matriz Q pela hessiana é através do método FR:

$$\beta_i = \frac{\|g_{i+1}\|^2}{\|g_i\|^2}. \quad (3.20)$$

Gradiente Conjugado Escalonado (SCGM) (BAZARAA; SHERALI; SHETTY, 1993, LUENBERGER, 1989)

O algoritmo SCGM é uma modificação do método SCG e permite o cálculo exato da

informação de segunda ordem. Este método, assim como o PR e o FR apresenta uma variação do algoritmo gradiente conjugado e a matriz Q deve ser aproximada pela hessiana:

$$\beta_i = \frac{d_i^T \nabla J(\theta_i)}{d_i^T \nabla^2 J(\theta_i) d_i}. \quad (3.21)$$

Pearlmutter (1994) propôs a utilização de um operador diferencial capaz de calcular exatamente o produto entre a matriz hessiana e qualquer vetor desejado, sem a necessidade de calcular e armazenar a matriz hessiana.

3.2.2 Programação genética (PG)

A PG foi desenvolvida por Koza (1992) com base nos trabalhos de Holland (1975) em algoritmos genéticos, esta técnica é baseada na teoria de Darwin(1859) inspirada na competição entre indivíduos pela sobrevivência.

Neste método os indivíduos são programas formados pela combinação do conjunto de funções (operadores) e terminais (variáveis e constantes) adequados ao domínio do problema, aos quais são associados um valor de mérito (*fitness*), representando o quanto ele é capaz de resolver o problema. O objetivo é encontrar uma solução no espaço de todos os programas possíveis.

A PG pode ser descrita como um ciclo “criar-testar-modificar”, como mostra a Figura 17. Primeiramente os programas são criados baseados no conhecimento do domínio do problema e em seguida são testados para verificar sua funcionalidade. Se os resultados não forem satisfatórios, modificações são feitas para melhorá-los. Este ciclo é repetido até que uma solução satisfatória seja encontrada ou algum critério seja satisfeito.

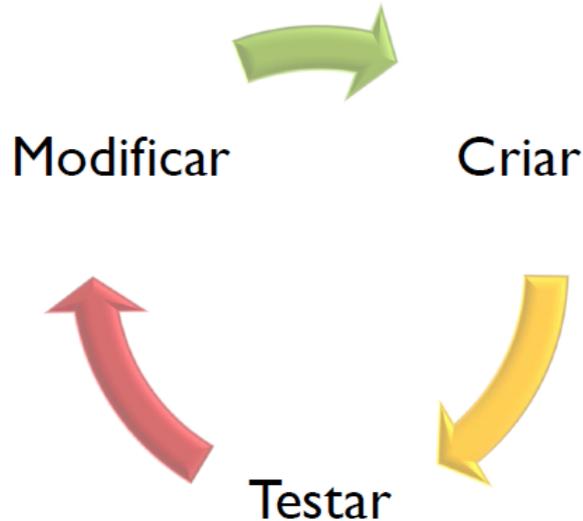


Figura 17 – Ciclo “Criar-Testar-Modificar” utilizado na programação genética para selecionar os indivíduos que mais se adaptam ao problema.
Fonte: Própria do autor.

A Figura 18 ilustra uma visão geral do algoritmo de PG. Pode-se observar como é possível chegar em uma solução aceitável para o problema.

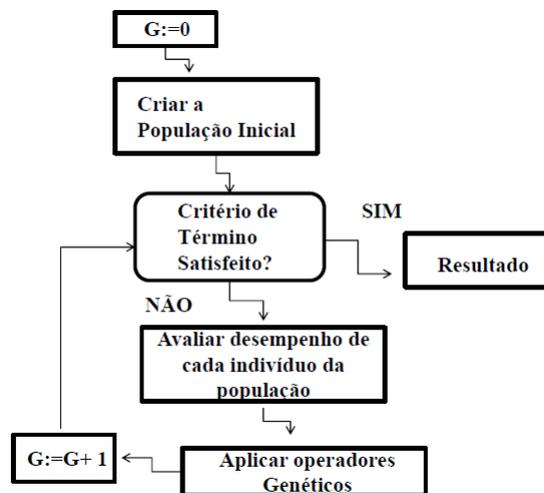


Figura 18 – Visão geral do algoritmo utilizado pela programação genética para chegar a uma solução satisfatória.
Fonte: Própria do autor.

Representação dos Programas

Os programas são criados através da combinação do conjunto de funções e terminais. Parte-se dos seguintes conjuntos:

a) **F (Conjunto das Funções)**: pode conter operadores aritméticos (+, -, *, /), funções matemáticas (sen, cos, log), operadores lógicos (e, ou), dentre outros;

b) **T (Conjunto de Terminais)**: é composto pelas variáveis e constantes.

Por exemplo, supondo $F = \{+, -, *, /\}$ e $T = \{x, 2\}$ o conjunto de funções e terminais respectivamente, uma possível representação matemática seria $x^2 + 2$ e pode ser representado pela árvore de sintaxe abstrata (FIGURA 19). O espaço de busca é determinado por todas as árvores que possam ser criadas pela combinação de elementos dos conjuntos F e T.

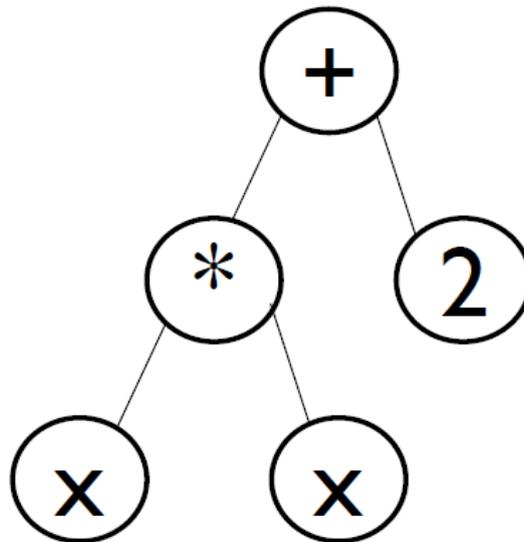


Figura 19 – Exemplo da representação do programa $x^2 + 2$ da programação genética através da árvore sintática.

Fonte: Própria do autor.

Função de Aptidão

Os programas que melhor resolvem os problemas recebem melhores valores de *fitness* e, conseqüentemente, terão maior chance de serem selecionados para reproduzir.

Usualmente, para proceder à avaliação de *fitness*, é fornecido um conjunto de casos de treinamento contendo valores de entrada e saída a serem aprendidos. A cada programa é fornecido os valores de entrada e confronta-se a sua resposta ao valor esperado de saída. Quanto mais próxima a resposta do programa estiver do valor de saída, melhor é o programa. Desta forma, a avaliação de *fitness* estabelece uma forma de diferenciar os melhores dos piores.

Os métodos mais comuns usados para avaliação de *fitness* são:

a) Aptidão Nata (row fitness): é a avaliação do erro cometido, isto é, a soma de todas as diferenças absolutas entre o resultado obtido pelo programa e o seu valor correto.

b) Aptidão Padronizada (standardized fitness): a avaliação da aptidão padronizada é feita através de uma função de adaptação do valor da aptidão nata de forma que quanto melhor

o programa, menor deve ser a aptidão padronizada.

c) Aptidão Ajustada (adjusted fitness): é obtida através da aptidão padronizada do indivíduo i na geração t , então a aptidão ajustada $a(i,t)$ é:

$$a(i,t) = \frac{1}{1 + S(i,t)} \quad (3.22)$$

$a(i,t)$ varia entre 0 e 1, e maiores valores representam os melhores indivíduos. E $S(i,t)$ é a aptidão padronizada.

d) Aptidão Normalizada (normalized fitness): Depende da aptidão ajustada e é obtida da seguinte forma:

$$n(i,t) = \frac{a(i,t)}{\sum_{k=1}^m a(k,t)}. \quad (3.23)$$

A soma de todas as aptidões normalizadas dentro de uma população é 1.

Métodos de Seleção

Os métodos de seleção tem o objetivo de escolher quais os programas deverão sofrer a ação dos operadores genéticos e compor uma nova geração. A seleção deve preferenciar os programas que apresentem os melhores valores de *fitness*. Existem diversos métodos de seleção, porém neste trabalho foi abordado apenas a seleção por torneio.

Neste método, t indivíduos são escolhidos aleatoriamente da população e o melhor deles é escolhido. Este processo é repetido até que se tenha uma nova população. O valor de t é conhecido como o tamanho do torneio.

Operadores Genéticos

Após a seleção dos programas, são aplicados os operadores genéticos para compor uma nova geração. Existem três principais operadores genéticos:

a) Reprodução: um programa é selecionado e copiado para a próxima geração sem sofrer nenhuma mudança de estrutura.

b) Cruzamento (Crossover): dois programas são selecionados e recombinaos para gerar outros dois programas. Um ponto aleatório de cruzamento é escolhido em cada programa-pai e as árvores abaixo destes pontos são trocadas.

c) Mutação: um programa é selecionado e um de seus nós é escolhido aleatoriamente. A árvore cuja raiz é o nó selecionado é então eliminada e substituída por uma nova árvore gerada aleatoriamente.

Critério de Término

O critério de término é responsável por interromper o laço de repetição do processo evolutivo que, idealmente, não teria fim. O critério mais comum é limitar o número máximo de gerações ou até que uma solução satisfatória seja encontrada. Porém, existem critérios baseados no próprio acompanhamento do processo evolutivo, isto é, enquanto houver melhoria na média da população, o processo evolutivo prossegue.

3.2.3 *Ensembles*

Ensemble é uma palavra que se origina do inglês e significa “agrupamentos” e consiste em um conjunto de componentes que devem ser combinadas na proposição de um única solução para o problema. Este paradigma originou-se do trabalho de Hansen e Salamon (1990), que mostraram que a habilidade de generalização pode ser significativamente melhorada por meio da composição de várias redes neurais artificiais, ou seja, treinamento independente de várias redes neurais artificiais e posterior composição das saídas individuais.

Dentre as diversas áreas de aplicação dos *ensembles* podem ser citadas: a classificação de padrões: reconhecimento de face (GUTTA; WECHSLER, 1996, HUANG et al., 2000), reconhecimento óptico de caracteres (DRUCKER, 1993), análise de imagens (CHERKAUER, 1996), diagnóstico médico (CUNNINGHAM et al., 2000). Assim como em problemas de regressão: aproximação de funções na forma de mapeamentos de entradas-saídas (HASHEM; SCHMEISER, 1995, LIMA; COELHO; VON ZUBEN, 2002), predição de séries temporais (INOUE; NARIHISA, 2000).

A Figura 20 ilustra o caso do problema de predição de séries temporais, supondo os componentes na forma de RNAs. Na Figura 20, há 3 componentes com desempenhos diferentes para o conjunto de teste (predição). Pode-se perceber que os dois primeiros conseguiram aproximar melhor os maiores picos da série, e que o último teve melhor desempenho nos valores baixos. Se combinarmos essas predições utilizando a média simples, obteremos o resultado apresentado ao lado direito da Figura 20. Conforme pode ser observado, o resultado obtido via *ensemble* apresenta um desempenho superior a qualquer um dos componentes tomados de forma isolada (VILLANUEVA, 2006).

Em geral, um *ensemble* é construído em dois passos: geração dos vários componentes e então a combinação da saída proposta pelos componentes, como esquematizado na Figura 21.

Quanto à geração dos componentes, as abordagens predominantes são *bagging* e *boosting*. *Bagging* (*Bootstrap aggregating*) foi proposto por Breiman (1996), baseado em amostragem *bootstrap* (EFRON; TIBSHIRANI, 1993, BREIMAN, 1996). Trata-se de uma técnica de geração de conjuntos distintos de treinamento a partir de um único conjunto, via reamostragem

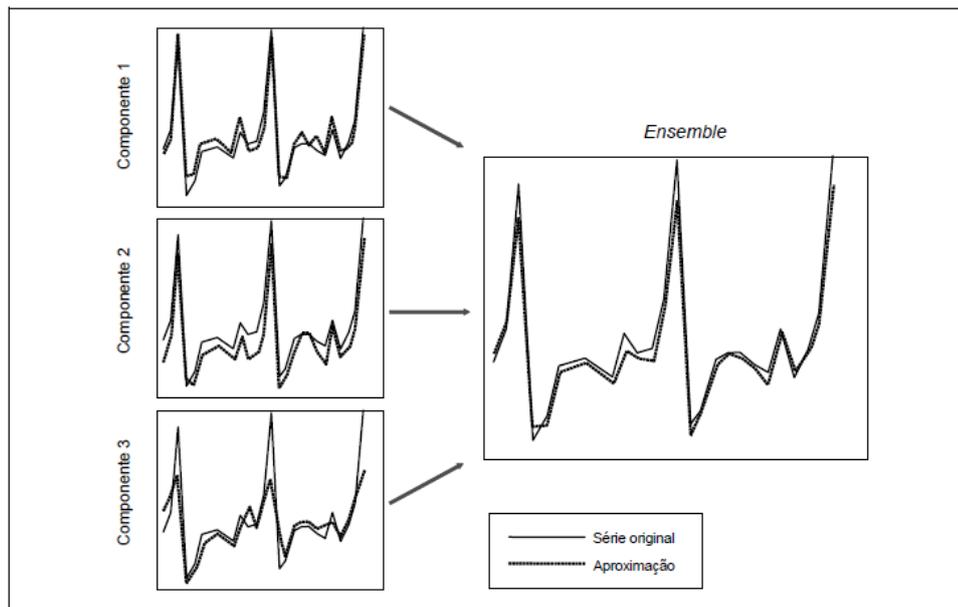


Figura 20 – Exemplo ilustrativo do ganho de desempenho utilizando *ensemble* em problemas de previsão de séries temporais.

Fonte: VILLANUEVA (2006).

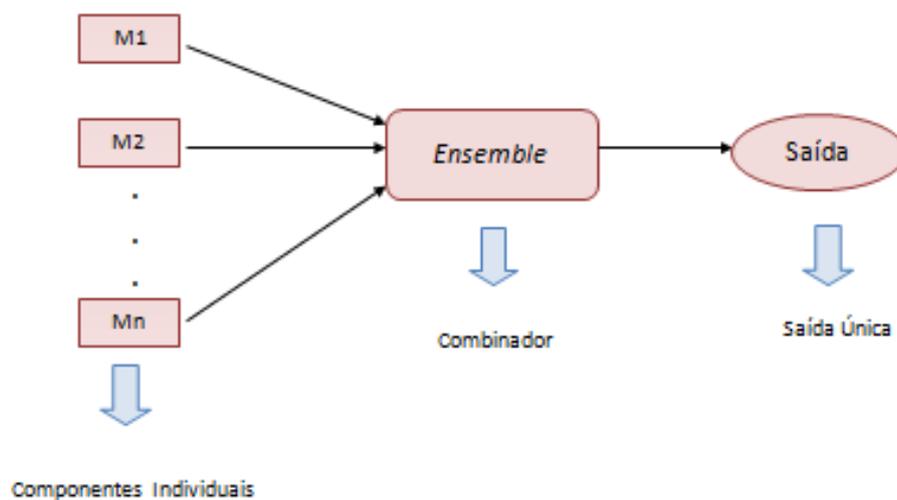


Figura 21 – Esquema ilustrativo da construção de um *ensemble* em duas etapas: geração das componentes de previsão e a combinação das previsões obtidas através destas componentes.

Fonte: Própria do autor.

com reposição. Estes conjuntos de treinamento serão usados para treinar cada rede neural do tipo MLP que compõe o *ensemble*. A Figura 22 mostra graficamente este processo. Considere que se parta de um único conjunto de treinamento, com N amostras, para treinar m componentes do tipo MLP. O objetivo é construir m conjuntos de treinamento com N amostras cada um. Supondo uma distribuição de probabilidade uniforme para seleção das amostras a partir do conjunto fonte, a existência de reposição de amostras selecionadas implica que algumas amostras podem ser selecionadas mais de uma vez. Com isso, apenas um subconjunto das amostras do conjunto fonte estará em cada um dos m conjuntos, o que implica na existência de amostras repetidas, já que o total de amostras é N (VILLANUEVA, 2006).

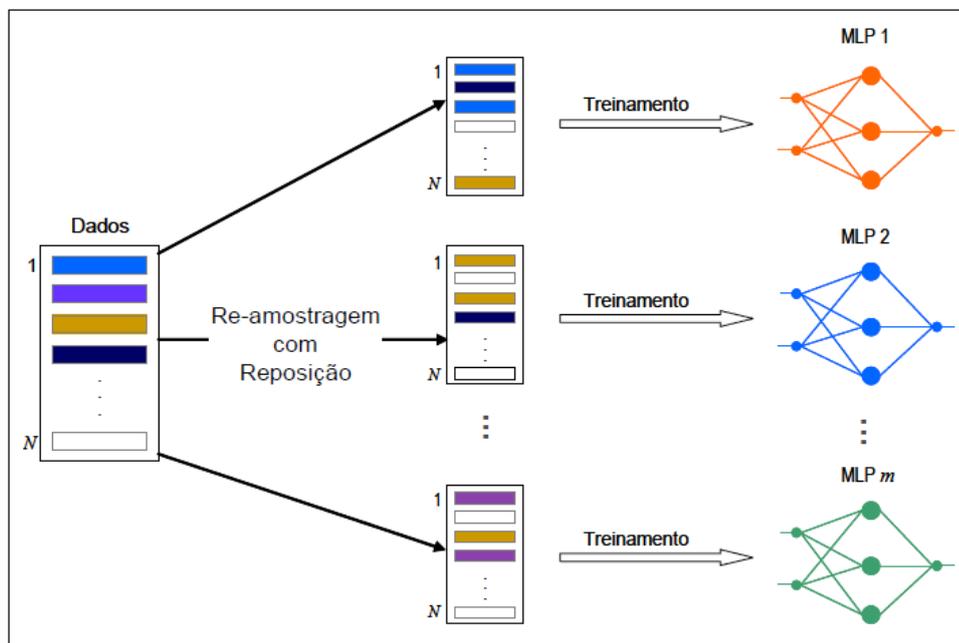


Figura 22 – Esquema da geração de conjuntos de treinamentos distintos para geração das componentes de previsão via *bagging*.
Fonte: VILLANUEVA (2006).

Boosting foi proposto por Shapire (1990) e aperfeiçoado por Freund e Schapire(1995). Nesta abordagem, os vários conjuntos de treinamento não são gerados a partir de uma amostragem uniforme com reposição, como no caso do *bagging*. A probabilidade de escolha de uma amostra depende da contribuição desta para o erro de treinamento dos componentes já treinados, isto é, caso uma amostra não tenha sido corretamente classificada pelos componentes já gerados, a probabilidade de escolha desta aumenta em relação às demais amostras, quando do treinamento de novos componentes. Conseqüentemente, esta amostra terá uma chance maior de ser escolhida para compor o conjunto de dados do próximo componente a ser gerado. Portanto, apenas o primeiro componente do *ensemble* é treinado a partir de uma amostragem uniforme

do conjunto de dados original. É necessário, assim, que os vários componentes do *ensemble* sejam treinados sequencialmente, visando redefinir a probabilidade de escolha das amostras na geração dos próximos conjuntos de treinamento.

Uma condição para que um *ensemble* tenha capacidade de apresentar um desempenho superior a qualquer um de seus componentes individuais é fazer com que os classificadores que irão compor o *ensemble* tenham um bom desempenho isoladamente e sejam diversos entre si (HANSHEM; SALAMON, 1990).

O objetivo então é encontrar redes neurais de bom desempenho e que generalizam de forma descorrelacionada. Há vários parâmetros de treinamento que podem ser manipulados tendo este objetivo em mente. Estes incluem: condição inicial, dados de treinamento, tipo de rede e algoritmo de treinamento (LIMA, 2004).

No entanto, a simples adoção de parâmetros e aspectos estruturais distintos para as redes neurais componentes promove uma tendência, mas não garante a obtenção de padrões distintos de comportamento para as diversas soluções geradas, ou seja, não garante a obtenção de redes neurais de bom desempenho e que generalizam de forma descorrelacionada. Logo, os métodos mais avançados de geração de componentes devem ser dedicados à maximização da divergência de comportamento, ao mesmo tempo em que se busca garantir bom desempenho (LIMA, 2004).

Os métodos para conseguir componentes que tenham um desempenho satisfatório podem ser divididos em duas classes: (1) geração pelo pré-processamento de parâmetros e aspectos estruturais de redes neurais artificiais; e (2) geração pelo pré-processamento dos dados de treinamento.

a) Pré-processamento de parâmetros e aspectos estruturais

Definição do conjunto de pesos iniciais: como os algoritmos de treinamento das redes neurais partem de um conjunto inicial de pesos, os quais são então submetidos a um processo iterativo de ajuste, a adoção de condições iniciais distintas representa pontos de partida distintos para o processo de ajuste de pesos, o que pode propiciar a convergência para um conjunto de pesos distinto, ou seja, a convergência do treinamento pode se dar para um ótimo local diferente e dependente da localização da condição inicial ou ponto de partida. Com isso, espera-se que as redes neurais resultantes generalizem de forma diferente, mesmo que sejam mantidos idênticos os dados de treinamento, a arquitetura da rede neural e o algoritmo de treinamento.

Definição da arquitetura da rede neural: como o comportamento de uma rede neural depende fortemente de sua arquitetura (basicamente vinculada à estratégia de conexão e ao número e tipo de neurônios), a adoção de arquiteturas distintas pode conduzir a redes neurais que generalizam de forma diferente, mesmo que sejam mantidos idênticos os dados de treinamento e o algoritmo de treinamento. Com arquiteturas distintas, os comportamentos de múltiplas redes neurais treinadas junto ao mesmo conjunto de dados podem ser bastante desconcorrelacionados, pois as estruturas internas de processamento serão distintas.

Definição do algoritmo de treinamento: como os algoritmos de treinamento das redes neurais empregam um processo iterativo de ajuste, mesmo que se parta de uma mesma condição inicial, o processo de ajuste iterativo de pesos pode diferir bastante, de um algoritmo de treinamento para outro. Sendo assim, há a possibilidade de convergência para um conjunto de pesos distinto, ou seja, a convergência do treinamento pode se dar para um ótimo local diferente. Logo, a adoção de procedimentos de otimização distintos pode conduzir a redes neurais que generalizam de forma diferente, mesmo que sejam mantidos idênticos os dados de treinamento, os pesos iniciais e a arquitetura da rede neural.

b) Pré-processamento dos dados de treinamento

Reamostragem dos dados: uma abordagem comum para a criação de um conjunto de componentes com capacidade de generalização distinta é a reamostragem dos dados, caracterizada pela obtenção de vários subconjuntos de treinamento distintos a partir de um único conjunto de treinamento. Dentre os métodos de reamostragem usados para este propósito está a técnica *bagging* (*bootstrap aggregating*) que envolve amostragem de dados com reposição (BREIMAN, 1996).

Boosting ou reamostragem adaptativa: Vários estudos empíricos sustentam a eficácia do algoritmo *boosting* (DRUCKER et al., 1994). Freund e Schapire (1996) propuseram uma das mais utilizadas versões de *boosting*, chamada *Adaboost* (*Adaptive boosting*), em que os conjuntos de treinamento são reamostrados de forma adaptativa, de tal modo que amostras que mais contribuem para o erro de treinamento dos componentes já treinados têm aumentada sua

probabilidade de comporem o conjunto de treinamento a ser empregado na síntese do próximo componente. Como é deixado transparecer, os componentes do *ensemble* devem ser obtidos de maneira sequencial. Breiman (1996) explora algumas das diferenças entre os algoritmos *Adaboost* e *bagging*, concluindo, com base em evidência empírica e analítica, que o algoritmo *Adaboost* obtém melhor êxito que o *bagging* na redução da variância do erro de treinamento.

Conjuntos de treinamento disjuntos: um método similar aos dois já mencionados acima envolve a obtenção de conjuntos de treinamento disjuntos ou mutuamente exclusivos, isto é, amostragem sem repetição (SHARKEY et al., 1996). Não há então nenhuma sobreposição de dados usados para treinar os componentes diferentes. Como observado por Tumer e Ghosh (1996), o problema é que o tamanho dos conjuntos de treinamento vai ficar reduzido, e este fator pode resultar em desempenho de generalização ruim.

Variáveis distintas e transformação de variáveis: em algumas aplicações específicas, é possível obter conjuntos de treinamento distintos pela simples consideração de um elenco de variáveis distintas para o vetor de entrada (SHARKEY et al., 1996). Desse modo, cada componente do *ensemble* toma um subconjunto de variáveis de entrada distinto dos demais, embora possa haver variáveis comuns entre eles. Além da possibilidade de existirem variáveis provenientes de fontes sensoriais distintas na composição destes conjuntos, os diferentes conjuntos de variáveis podem ser gerados pela aplicação de métodos de poda de variáveis (TUMER; GHOSH, 1996), injeção de ruído (RAVIV; INTRATOR, 1999), ou então as variáveis distintas podem representar também alguma combinação linear ou não-linear de variáveis tomadas isoladamente em outros conjuntos, dentre as quais se destaca a análise de componentes principais lineares e não-lineares (SHARKEY; SHARKEY, 1997).

Apesar da falta de uma teoria unificada de *ensemble*, existem razões teóricas para combinar múltiplos componentes. Não há nenhuma garantia que o desempenho de combinadores será sempre superior ao das componentes individuais. Porém, diversos estudos mostram que na maioria das vezes isso acontece.

Um trabalho de grande relevância na área é o de Salgado et al. (2006), no qual é proposta uma metodologia de previsão de carga através da combinação dos resultados dos modelos de RNA e SVM. Neste trabalho, é identificado um erro cometido pelo *ensemble* 25% menor do que o melhor resultado obtido pela melhor componente de previsão de forma individual.

Um estudo que propõe a utilização do *ensemble* para melhorar os resultados de previsão é o de Li, Wang e Han (2009), em que são utilizadas combinações de RNA através de um algoritmo evolucionário. Os experimentos observados demonstraram que a abordagem proposta oferece resultados promissores, além de conferir uma maior precisão e estabilidade.

Nesta subsecção sobre *ensembles* foram mostradas algumas estratégias simples de combinação já utilizadas por outros pesquisadores e que tiveram êxito ao empregar esta metodologia. A idéia deste trabalho é propor outras formas de combinação através de RNA's e modelos de regressão linear múltipla com abordagem bayesiana, além dos métodos mais simples como as médias.

Inferência Bayesiana

A inferência bayesiana é caracterizada por se basear em um mecanismo que faz uso de informação anterior, a qual é denominada informação *a priori*. Ao contrário da inferência clássica, na inferência bayesiana a quantidade de interesse θ não é constante, mas sim aleatório (BOX; TIÃO, 1992). Portanto, ao se utilizar uma informação *a priori* é necessário especificar qual a distribuição do parâmetro θ . Essa distribuição representa, probabilisticamente, a informação anterior ao experimento que se tem sobre o parâmetro (PAULINO; AMARAL; MURTEIRA, 2003).

Segundo Paulino, Amaral e Murteira (2003), a teoria da inferência bayesiana é baseada no Teorema de Bayes. Utilizando-se a informação *a priori* $P(\theta)$ e a função de verossimilhança $L(Y_n|\theta)$ é possível obter a distribuição *a posteriori* através do Teorema de Bayes:

$$P(\theta|Y_n) = \frac{L(Y_n|\theta)P(\theta)}{\int L(Y_n|\theta)P(\theta)d\theta} \quad (3.24)$$

sendo $Y_n = \{y_1, y_2, \dots, y_n\}$. Como o denominador não depende de θ tem-se:

$$P(\theta|Y_n) \propto L(Y_n|\theta)P(\theta) \quad (3.25)$$

ou seja,

$$Posteriori \propto Verossimilhança \times Priori.$$

A função de *verossimilhança* pode ser definida da seguinte forma: Seja $f(y|\theta)$ uma função densidade de probabilidade (fdp) conjunta da amostra independente e identicamente distribuída $Y = \{Y_1, \dots, Y_n\}$. Então, dado que $Y = y$ é observado, a função de θ definida por

$$L(\theta|y) = f(y|\theta) \quad (3.26)$$

é chama de função de verossimilhança (CASELLA; BERGER, 2010).

A distribuição de probabilidade *a priori* reflete o conhecimento que se tem sobre os parâmetros anteriormente à realização do experimento ou à observação dos dados. Os parâmetros desta distribuição são denominados hiperparâmetros para diferenciá-los do parâmetro de interesse θ .

Existem diferentes especificações da distribuição *a priori*, uma delas é a *priori* conjugada que ocorre quando as distribuições *a priori* e *a posteriori* pertencem a mesma classe de distribuições, mudando apenas os hiperparâmetros.

Outra especificação de informação *a priori* é a não informativa, nesse tipo de *priori* espera-se que a informação dos dados seja dominante. Uma idéia de não informação seria o uso da distribuição *a priori* uniforme. Neste caso, nenhum valor particular tem preferência, pois θ está variando em um subconjunto da reta. Segundo Casella e Berger (2010), a distribuição uniforme é definida pela massa uniformemente espalhada sobre um intervalo $[a,b]$, sua fdp é dada por:

$$f(\theta) = \frac{1}{b-a} I_{[a,b]}(\theta). \quad (3.27)$$

Após obter a distribuição *a posteriori* através da Equação 3.25, qualquer processo inferencial em relação a qualquer elemento de θ deve ser realizado através da distribuição marginal desse elemento, que é obtida integrando-se a distribuição *a posteriori* conjunta dos parâmetros em relação a todos os outros elementos que a constituem (ROSA,1998).

Em geral a integração da distribuição *a posteriori* conjunta não é analítica, e por isso faz-se necessário o uso de algoritmos iterativos para a realização dessa tarefa, tais como o algoritmo de Metropolis-Hastings e o Amostrador de Gibbs, os quais fazem parte de uma classe de simulações iterativas baseadas em cadeias de Markov (MCMC).

A ideia da cadeia de Markov é simular um passeio aleatório no espaço de θ que converge para uma distribuição estacionária, que é a distribuição de interesse do problema. Ou seja, é um processo estocástico tal que a distribuição de Y_t dados todos os valores anteriores $\{Y_0, Y_1, \dots, Y_{t-1}\}$ dependa apenas de Y_{t-1} (GAMERMAN, 1997).

Os métodos MCMC exigem que a cadeia seja: homogênea, ou seja, as probabilidades de transição de um estado para outro são invariantes; irredutível, ou seja, cada estado pode ser

atingido a partir de qualquer outro em um número finito de iterações; aperiódica, ou seja, não haja estados absorventes.

Os métodos MCMC que será utilizado neste trabalho é o amostrador de Gibbs. De acordo com Gamerman (1997), neste método, não existe mecanismo de aceitação e rejeição, isto é, a cadeia sempre se moverá para um novo valor. As transições de um valor para outro são feitas de acordo com as distribuições condicionais completas, que é a distribuição da *i-ésima* componente de θ condicionada em todas as outras componentes. E é obtida da seguinte maneira:

$$\pi(\theta_i|\theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta)d\theta_i}. \quad (3.28)$$

Segundo Gamerman (1997), o amostrador de Gibbs pode ser descrito da seguinte forma:

1. Inicie o contador de iterações da cadeia $j=1$ e escolha valores iniciais $\theta^0 = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$.
2. Obtenha um novo valor $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_p^{(j)})$ a partir de θ^{j-1} por meio de sucessivas gerações de valores

$$\begin{aligned} \theta_1^{(j)} &\propto \pi(\theta_1|\theta_2^{(j-1)}, \dots, \theta_p^{(j-1)}) \\ \theta_2^{(j)} &\propto \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)}) \\ &\vdots \\ \theta_p^{(j)} &\propto \pi(\theta_p|\theta_1^{(j)}, \dots, \theta_{p-1}^{(j)}). \end{aligned}$$

3. Mude o contador j para $j + 1$ e retorne ao passo 2 até atingir a convergência.

Uma questão importante de ordem prática é como os valores iniciais influenciam o comportamento da cadeia. A idéia é que conforme o número de iterações aumenta, a cadeia gradualmente esquece os valores iniciais e eventualmente converge para uma distribuição de equilíbrio. Assim, em aplicações práticas é comum que as iterações iniciais sejam descartadas, como se formassem uma amostra de aquecimento.

Como o método MCMC são iterativos faz-se necessário verificar se houve a convergência da cadeia para a distribuição alvo (distribuição *a posteriori*) (GAMERMAN; LOPES, 2006). Para verificar a convergência da cadeia existem os métodos formais e informais.

Os métodos informais foram propostos por Gelfand e Smith (1990), e se trata de uma técnica gráfica que consiste em observar a trajetória da cadeia ao longo das iterações. Se após um período inicial, o gráfico apresentar repetidamente o mesmo comportamento, então pode-se

concluir que houve convergência da cadeia.

Os métodos formais mais utilizados são: Raftery e Lewis (1992), Geweke (1992), Heidelberger e Welch (1983).

a) Raftery e Lewis (1992):

Este critério permite estimar quantas iterações são necessárias para que a cadeia apresente convergência à distribuição estacionária. O método fornece:

- burn-in, que é o número de iterações que devem ser descartadas (queima dos valores iniciais da cadeia);
- número de iterações que devem ser computadas (fator de dependência, que deve ser menor que cinco);
- distância mínima de uma iteração à outra para se obter a subamostra aproximadamente independente ("thin").

b) Geweke (1992):

Este método é um teste em que a cadeia é dividida em duas partes, a primeira corresponde aos 10% iniciais e a segunda os 50% finais da cadeia. O teste consiste verificar se os valores da primeira parte são iguais em média aos da segunda parte.

c) Heidelberger e Welch (1983):

Este método testa a hipótese de estacionariedade da distribuição da cadeia de Markov. Se a hipótese de estacionariedade for rejeitada descarta-se os 10% iniciais das iterações e repete-se o teste. Se a hipótese mais uma vez for rejeitada então descarta-se mais 10%. Repete-se este processo até que se obtenha uma proporção de 50% (10%, 20%, ..., 50%). Se novamente rejeitar-se a hipótese ou caso não seja possível aplicar o teste, isto indica falta de estacionariedade, fazendo-se necessário um maior número de iterações. Por outro lado, se o teste for satisfatório, o número de iterações descartadas inicialmente é indicado como tamanho da amostra de aquecimento (burn-in) (NOGUEIRA, 2004).

4 MATERIAL E MÉTODOS

Nesta seção são apresentadas as componentes de previsão, com o objetivo de mostrar quais as configurações dos modelos foram utilizadas para a realização das previsões. Também são mostrados quais os *ensembles* utilizados e qual a medida de desempenho utilizada para avaliar as previsões de cada um dos modelos. Na última seção são apresentados os recursos computacionais empregados para realizar estas tarefas.

4.1 DADOS DE VAZÃO

Este trabalho consiste em realizar previsões de vazões para duas usinas do Rio Grande: Água Vermelha e Itutinga. Os dados são do Operador Nacional do Sistema Elétrico (ONS) e correspondem a vazões médias mensais do período de 1931 a 2013.

O conjunto de dados foi dividido em três partes. O período de 1931 a 1993 para o treinamento que é utilizado para o aprendizado dos modelos. De 1994 a 2003 é a fase de validação e serve para verificar durante o treinamento como evolui o erro do modelo evitando o *overfitting*, ou seja, quando o modelo se especializa excessivamente em um conjunto de treino e perde a capacidade de generalizar. O período de 2004 a 2013 corresponde a fase de teste, através do qual é possível verificar o desempenho do modelo para amostras desconhecidas.

Devido os meses secos e úmidos serem bem definidos e a diferença no comportamento da vazão para estes períodos ao longo do ano, optou-se por avaliar o desempenho dos modelos para os períodos de chuva e estiagem separadamente, sendo o período seco equivalente aos meses de maio a outubro e o úmido, novembro a abril.

4.2 COMPONENTES DE PREVISÃO

As componentes de previsão empregadas nesse trabalho são a Programação Genética (PG) e as Redes Neurais Artificiais (RNA). Para que as RNAs sejam capazes de aprender,

são necessários algoritmos de treinamento. Neste trabalho foram utilizados sete algoritmos diferentes:

- a) Retropropagação (BPM);
- b) Método Gradiente (GRAD);
- c) Davidon Fletcher Powell (DFP);
- d) Broyden Fletcher Goldfarb Shanno (BFGS);
- e) Polak Ribière (PR);
- f) Fletcher & Reeves (FR);
- g) Gradiente Conjugado Escalonado (SCGM).

portanto, tem-se oito diferentes componentes individuais de previsão.

Um parâmetro importante para se realizar previsões é o número de atrasos, neste trabalho foram utilizados pelo modelo meses imediatamente anteriores ao mês a ser previsto, que correspondem aos meses que devem ser levados em consideração para modelar a dependência. Para determinação do número de atrasos utilizou-se a função de auto-correlação (FAC):

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}. \quad (4.1)$$

O número de atrasos compõem os padrões de entrada, neste caso os padrões utilizados pelo modelo foram os meses imediatamente anteriores ao mês desejado a ser previsto. Desta forma, o número de atrasos corresponde à quantidade de meses que serão levados em consideração para a composição do padrão. A Tabela 3 exemplifica um padrão utilizado pelo sistema com 4 atrasos, onde a previsão será realizada para o mês de janeiro e os meses de setembro até dezembro são utilizados como entradas para o modelo. Neste caso m representa apenas uma generalização do padrão utilizado pelo modelo, no qual $m - n$ são as componentes do padrão utilizadas como entrada e m é a saída desejada do modelo.

Tabela 3 – Exemplo de padrão utilizado pelos modelos para prever o mês de janeiro com uma dependência temporal de quatro.

		Entradas			Saídas Desejada
Mês	Set.	Out.	Nov.	Dez.	Jan
Série	m-4	m-3	m-2	m-1	m

Fonte: Própria do autor.

A dependência temporal foi calculada para cada mês do ano e os valores utilizados encontram-se na Tabela 4.

Tabela 4 – Dependências temporais para cada mês do ano calculadas através da função de autocorrelação para as usinas Água Vermelha e Itutinga.

Mês	Dependência Temporal	
	Água Vermelha	Itutinga
Janeiro	4	10
Fevereiro	6	9
Março	8	5
Abril	1	4
Maio	1	4
Junho	1	1
Julho	1	1
Agosto	1	1
Setembro	7	4
Outubro	11	1
Novembro	4	6
Dezembro	1	2

Fonte: Própria do autor.

Como exemplo, vamos supor que se queira realizar a previsão para o mês de janeiro de 2004, os dados utilizados para o treino do modelo são os meses precedentes a janeiro de 1932 até janeiro de 1993, para conseguir os dados de validação são utilizados os dez anos que precedem o ano anterior ao desejado, neste caso, são utilizados os dados de janeiro de 1994 até janeiro 2003 e os dados utilizados para testar a rede é o ano anterior ao ano que será previsto, ou seja, os dados dos meses de setembro de 2003 até dezembro de 2003. A Figura 23 ilustra de forma clara os dados utilizados para realizar o treino, a validação e o teste do modelo de previsão.

Entrada Treino				Saída
<i>Set/1931</i>	<i>Out/1931</i>	<i>Nov/1931</i>	<i>Dez/1931</i>	<i>Jan/1932</i>
<i>Set/1932</i>	<i>Out/1932</i>	<i>Nov/1932</i>	<i>Dez/1932</i>	<i>Jan/1933</i>
<i>Set/1933</i>	<i>Out/1933</i>	<i>Nov/1933</i>	<i>Dez/1933</i>	<i>Jan/1934</i>
⋮	⋮	⋮	⋮	⋮
<i>Set/1992</i>	<i>Out/1992</i>	<i>Nov/1992</i>	<i>Dez/1992</i>	<i>Jan/1993</i>

Entrada Validação				Saída
<i>Set/1993</i>	<i>Out/1993</i>	<i>Nov/1993</i>	<i>Dez/1993</i>	<i>Jan/1994</i>
<i>Set/1994</i>	<i>Out/1994</i>	<i>Nov/1994</i>	<i>Dez/1994</i>	<i>Jan/1995</i>
<i>Set/1995</i>	<i>Out/1995</i>	<i>Nov/1995</i>	<i>Dez/1995</i>	<i>Jan/1996</i>
⋮	⋮	⋮	⋮	⋮
<i>Set/2002</i>	<i>Out/2002</i>	<i>Nov/2002</i>	<i>Dez/2002</i>	<i>Jan/2003</i>

Entrada Teste				Saída
<i>Set/2003</i>	<i>Out/2003</i>	<i>Nov/2003</i>	<i>Dez/2003</i>	Previsão

Figura 23 – Padrões de treino, validação e teste utilizados pelos modelos para prever o mês de janeiro de 2004.

Fonte: Própria do autor.

Nos modelos de RNA, para um melhor desempenho dos algoritmos foi realizada uma busca dos parâmetros, ou seja, quais as configurações do modelo com o qual seria possível obter um menor erro de previsão da série. Os intervalos de busca para os parâmetros da RNA e os parâmetros utilizados pela PG se encontram na Tabela 5.

Tabela 5 – Parâmetros utilizados pelos modelos RNA e PG para obtenção das previsões das vazões no período de validação e teste.

RNA	
Número de neurônios na camada intermediária	[3; 15]
Taxa de Aprendizagem	[0; 1]
Termo <i>Momentum</i>	[0; 1]
Entradas	[1; 11]
PG	
Tamanho da População	100
Número de Gerações	50
Operadores Genéticos	{+,-,*,sen,cos,log,sqrt}
Entradas	[1; 11]

Fonte: Própria do autor.

A função de ativação utilizada neste trabalho foi a tangente hiperbólica que possui o contra-domínio no intervalo $[0; 1]$ como já foi discutido no Capítulo 3. As observações da base de dados devem pertencer também a este intervalo e por isso é necessário que os dados passem por um processo de normalização, neste caso o procedimento utilizado foi dividir todos os elementos pelo valor máximo da série, como mostra a Equação 4.2. Desta forma, os dados ficaram padronizados entre os valores 0 e 1. Este processo não acarretou perda da representatividade dos dados normalizados em relação aos dados brutos e para demonstrar visualmente que os dados não perdem suas características, a Figura 24 mostra os gráficos dos dados brutos e normalizados. Foram utilizados para gerar estes gráficos os dados referente a Usina de Furnas no período de janeiro de 1994 até dezembro de 2003.

$$z = \frac{x}{\text{máximo}} \quad (4.2)$$

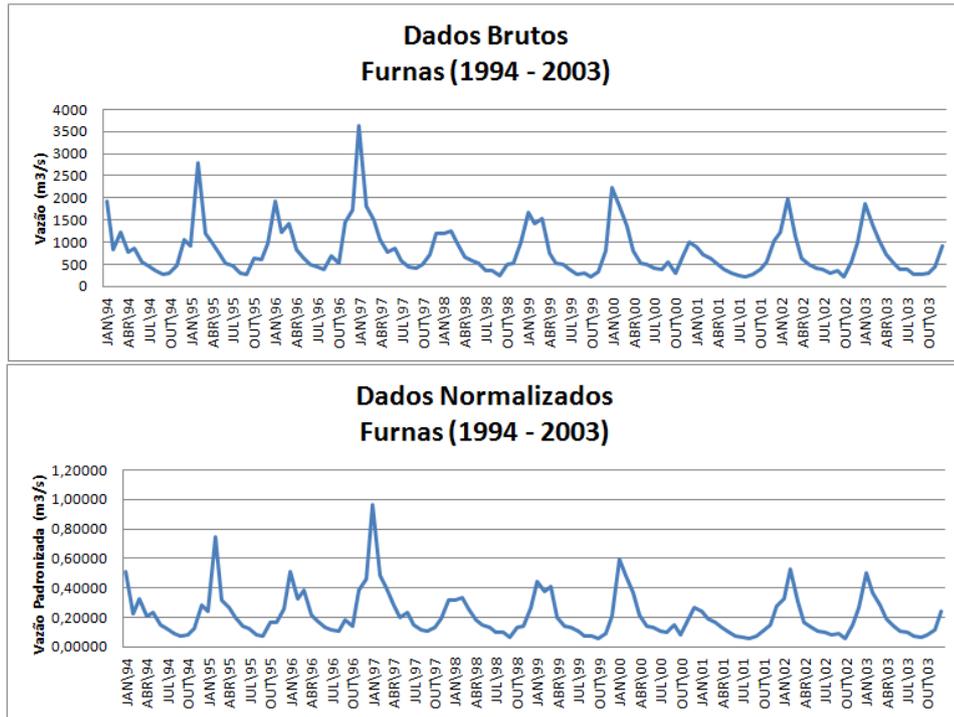


Figura 24 – Comparação dos dados bruto e normalizados da usina de Furnas do período de 1994 a 2003.

Fonte: Própria do autor.

A Figura 25 mostra os procedimentos para a obtenção dos resultados. Como é possível perceber, inicialmente foram realizadas 500 previsões para o conjunto de validação das usinas em estudo, através destas previsões realizou-se uma análise para encontrar os valores dos parâmetros que mais se adaptaram ao modelo, foi considerado o modelo com os parâmetros mais adaptados o que possuiu menor erro percentual. Após esta análise, foram realizados novos experimentos já com os parâmetros definidos para o período de teste.

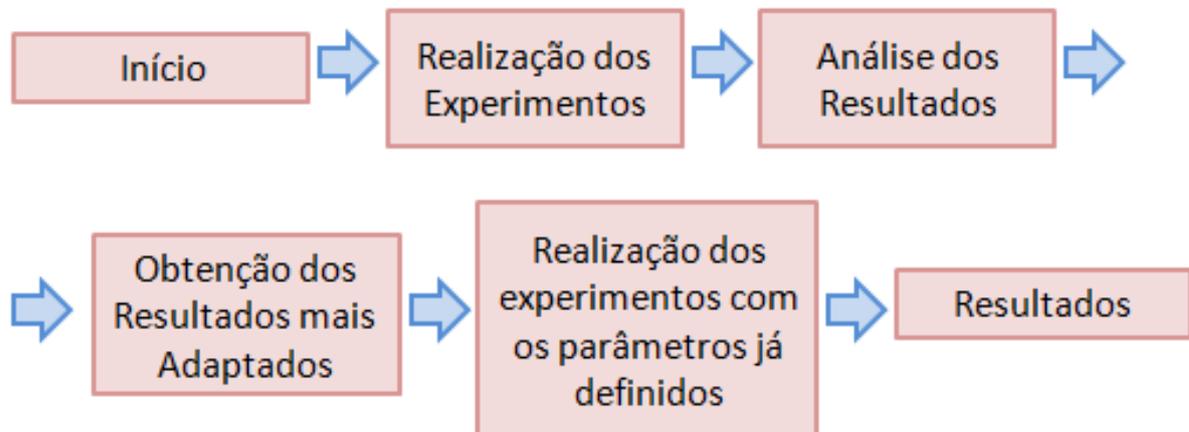


Figura 25 – Esquema de obtenção dos resultados das previsões através das componentes individuais de previsão.

Fonte: Própria do autor.

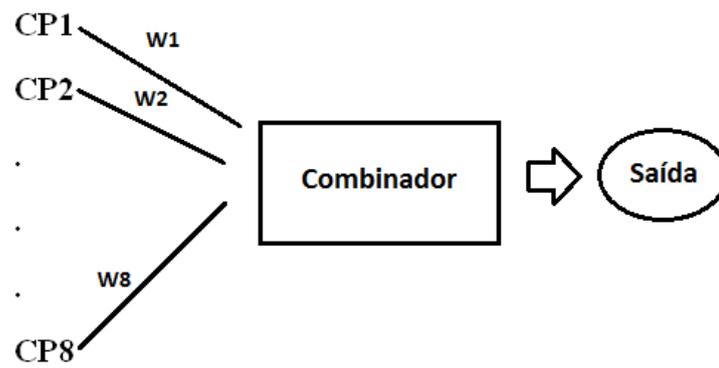
4.3 ENSEMBLES

Existem diversas formas de combinar componentes individuais. Suponha que existam r previsões $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_r$, a questão é como combinar estas previsões distintas em um único predictor \hat{y} com o objetivo de obter estimativas mais precisas. A forma geral de um combinador pode ser definido da seguinte forma:

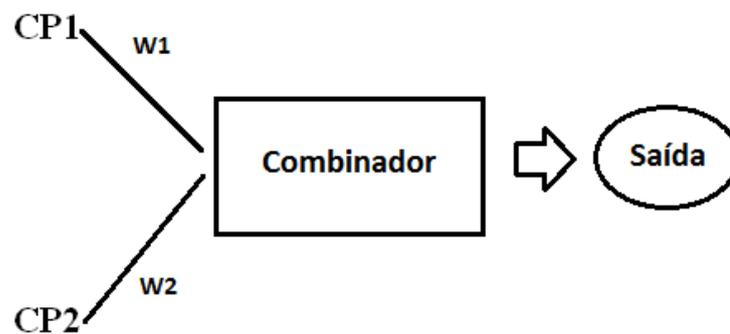
$$\hat{y} = \sum_{i=1}^r w_i \hat{x}_i, \quad (4.3)$$

onde w_i representa o peso atribuído a cada previsão \hat{x}_i .

Obtidas as previsões através das oito componentes individuais aplicando os modelos RNA e PG, estas foram combinadas através de treze diferentes *ensembles* que serão descritos a seguir. Primeiramente, combinou-se todas as oito componentes como mostra a Figura 26 (a) e posteriormente combinou-se as duas melhores componentes da fase de validação, ou seja, as duas componentes que apresentaram menor erro de previsão no período de validação como mostra o esquema da Figura 26 (b).



(a) Esquema do *ensemble* com oito componentes de previsão.



(b) Esquema do *ensemble* com duas componentes de previsão.

Figura 26 – Esquema ilustrativo do *ensemble* utilizando oito e duas componentes individuais de previsão.

Fonte: Própria do autor.

a) Médias

Os primeiros quatro combinadores utilizados neste trabalho foram: Média Aritmética (MA), Média Ponderada (MP), Média Geométrica (MG), Média Quadrática (MQ) e estão descritas na Tabela 6.

Tabela 6 – *Ensembles* baseado em médias para combinar as componentes individuais de previsão.

Médias Utilizadas	Descrição
Média Aritmética	$MA = \frac{\hat{x}_1 + \hat{x}_2 + \dots + \hat{x}_n}{n}$
Média Ponderada	$MP = \frac{\rho_1 \hat{x}_1 + \rho_2 \hat{x}_2 + \dots + \rho_n \hat{x}_n}{\rho_1 + \rho_2 + \dots + \rho_n}$
Média Geométrica	$MG = \sqrt[n]{\hat{x}_1 \cdot \hat{x}_2 \cdot \dots \cdot \hat{x}_n}$
Média Quadrática	$MQ = \sqrt{\frac{\hat{x}_1^2 + \hat{x}_2^2 + \dots + \hat{x}_n^2}{n}}$

Fonte: Própria do autor.

Observando a Tabela 6, $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ são as previsões das n componentes individuais e n é o total de componentes a serem combinadas. Na média ponderada, $\rho_1, \rho_2, \dots, \rho_n$ são os pesos de cada componente, cujos valores foram distribuídos empiricamente de acordo com o desempenho de cada modelo no período de validação de forma que aqueles que tiveram melhor desempenho receberam o peso de 0,9, os demais modelos receberam peso igual a 0,014. No caso, em que foram combinadas as duas melhores componentes do período de validação, os pesos escolhidos foram 0,9 e 0,1 para a primeira e segunda melhor componentes, respectivamente.

b) RNA e PG

A RNA e a PG além de terem sido empregadas como componentes individuais de previsão, foram utilizadas também como *ensemble*. Observando o esquema apresentado na Figura 27, E assume os seguintes modelos como *ensembles*: RNA com os sete algoritmos de treinamento já apresentados na Seção 4.2 e a PG.

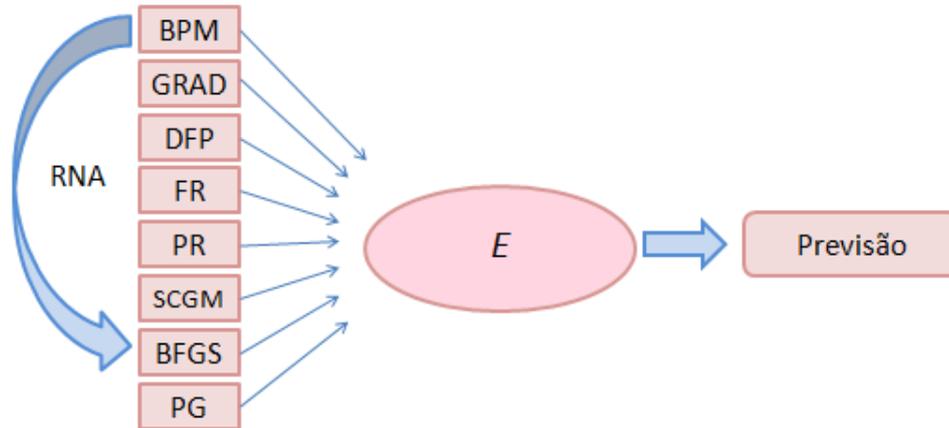


Figura 27 – Esquema *ensemble* com entradas e combinador sendo os modelos redes neurais artificiais (RNA) e a programação genética (PG).

Fonte: Fonte: Própria do autor.

Para utilizar a RNA e a PG como *ensembles* procede-se de forma semelhante quando estes modelos foram empregados como componentes individuais, seguindo algumas etapas. Primeiramente dividiu-se o conjunto de dados em três partes, a primeira correspondente ao período de treino, a segunda ao período de validação e a terceira ao período de teste. Em seguida, realizou-se a previsão do conjunto de validação através das componentes de previsão. Estas previsões foram empregadas como padrão de treino para o combinador. Assim, o modelo é capaz de ponderar as componentes de forma que, aqueles modelos que mostraram um desempenho melhor recebam um peso maior. Desta forma, obteve-se as previsões para o período de teste.

c) Modelo de Inferência Bayesiana (MIB)

O modelo de regressão linear múltipla com uma abordagem bayesiana foi empregado como um combinador dos oito modelos inteligentes. Os coeficientes $\beta_1, \beta_2, \dots, \beta_8$, estimados pelo método bayesiano, representam os pesos de cada modelo M_1, M_2, \dots, M_8 , respectivamente.

$$Y_p = \beta_1 M_1 + \beta_2 M_2 + \beta_3 M_3 + \beta_4 M_4 + \beta_5 M_5 + \beta_6 M_6 + \beta_7 M_7 + \beta_8 M_8 + \varepsilon \quad (4.4)$$

ou

$$Y_p = M\beta + \varepsilon \quad (4.5)$$

em que Y_p é um vetor de valores previstos 10 anos x 1 mês; M é uma matriz de previsões obtidas

através das componentes de previsão 10 anos x 8 modelos e ε é o vetor de erros aleatórios 10 x 1 que segue uma distribuição normal multivariada com média zero e variância $\sigma^2 I_n$:

$$\varepsilon \sim N_n(0, \sigma^2 I_n).$$

O modelo indica então que $Y|\beta, \sigma^2 \sim N_n(M\beta, \sigma^2 I_n)$, então a função de verossimilhança é dada por:

$$L(Y|\beta, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} [kS^2 + (\beta' - \hat{\beta}') M' M (\beta - \hat{\beta})] \right\} \quad (4.6)$$

em que $kS^2 = (Y - M\hat{\beta})'(Y - M\hat{\beta})$ e $\hat{\beta} = (M'M)^{-1}MY$ é a estimativa de mínimos quadrados (ou de máxima verossimilhança) de β .

Os parâmetros β e σ^2 são desconhecidos e por isso é necessário estabelecer *priori* para ambos. Assumindo que a *priori* para β é uma normal:

$$\beta \sim N(b_0, B_0^{-1})$$

ou

$$p(\beta) = \frac{|B_0|}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (\beta - b_0)' B_0 (\beta - b_0) \right\}.$$

Na inferência bayesiana trabalha-se com a precisão $\tau = \frac{1}{\sigma^2}$, que é o inverso da variância (GAMERMAN, 1997). E para esse parâmetro a *priori* utilizada foi a distribuição gama:

$$\tau \sim Gama(a, b)$$

ou

$$p(\tau) = \frac{b^a}{\Gamma(a)} \tau^{a-1} \exp\{-b\tau\}$$

onde $a = \frac{c_0}{2}$ e $b = \frac{d_0}{2}$. Os hiperparâmetros a serem definidos são: b_0 é a informação da média de β anterior aos dados, que neste trabalho é um vetor, B_0 é a informação da precisão de β anterior aos dados. Da mesma forma, este valor pode ser um escalar ou uma matriz quadrada com dimensão igual ao número de parâmetros de β . c_0 é o hiperparâmetro relacionado à forma da distribuição Gama pela sua precisão. d_0 é o hiperparâmetro relacionado à taxa de variação da distribuição Gama pela sua precisão.

Portanto, a distribuição a *priori* conjunta de β e τ será uma Normal-Gama:

$$p(\beta, \tau) = p(\beta) \times p(\tau) \propto b^a \tau^{a-1} \exp \left\{ -b\tau - \frac{1}{2}(\beta - b_0)' B_0(\beta - b_0) \right\}.$$

A *priori* utilizada foi uma Normal com uma variância alta, sendo portanto uma *priori* vaga. A fim de obter a informação a *priori* para os parâmetros, ajustou-se inicialmente um modelo de regressão linear frequentista utilizando a precisão, que é o inverso do erro $MAPE_{M_i}$ obtido no conjunto de validação (1994 a 2003), como pode ser visto na Equação 4.7.

$$W_i M_i = \frac{1}{MAPE_{M_i}} \quad (4.7)$$

Calculando a distribuição a *posteriori* tem-se:

$$p(\beta, \tau | Y) = p(\beta, \tau) \times L(Y | \beta, \tau) \propto b^a \tau^{(a+\frac{n}{2})-1} \exp \left\{ -b\tau - \frac{\tau}{2} [kS^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})] + \frac{1}{2}(\beta - b_0)' B_0(\beta - b_0) \right\}$$

Para verificar a convergência das cadeias foram utilizados os critérios de Raftery e Lewis (1992), Geweke (1992) e Heidelberg e Welch (1983) já descritos na Seção 3, na Subseção 3.2.3.

O desempenho dos modelos testados foi comparado através do erro de previsão de cada um, calculados através da Equação 4.8, onde x_i é a vazão observada, \hat{x}_i é a vazão prevista e n é o número de previsões.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\%. \quad (4.8)$$

4.4 RECURSOS COMPUTACIONAIS

Os cálculos deste trabalho foram realizados computacionalmente através de dois *softwares*. Para obter as previsões através dos modelos de inteligência artificial usou-se o MATLAB. O R (R CORE TEAM, 2014) foi empregado para a construção do *ensemble* MIB e para a obtenção da *posteriori* foi utilizado o pacote MCMCpack também do *software* R.

5 RESULTADOS

Esta seção evidencia os resultados obtidos pelos modelos individuais de previsão e pelos *ensembles* propostos neste trabalho. Estes resultados serão apresentados separadamente para os períodos secos e úmidos devido a grande diferenças na média e variabilidade da vazão ao longo dos meses já discutidos na Seção 2. Na Subseção 5.1 é feita uma análise dos dados por meio das medidas de posição e variabilidade das séries. E em seguida, na Subseção 5.2 são apresentados os estudos de casos abordados neste trabalho.

5.1 ANÁLISE DESCRITIVA DOS DADOS DE VAZÃO

Esta subseção se refere a análise descritiva dos dados de vazão das usinas Água Vermelha e Itutinga.

Na Figura 28 tem-se a representação gráfica dos dados de vazão para a usina hidroelétrica Água Vermelha no período de 1994 a 2003. Como pode ser visto, possuem um comportamento periódico, ou seja, uma oscilação entre um máximo de $6827 \text{ m}^3/\text{s}$ e um mínimo de $530 \text{ m}^3/\text{s}$. Isto caracteriza a sazonalidade da série, com um período de aproximadamente 12 meses.

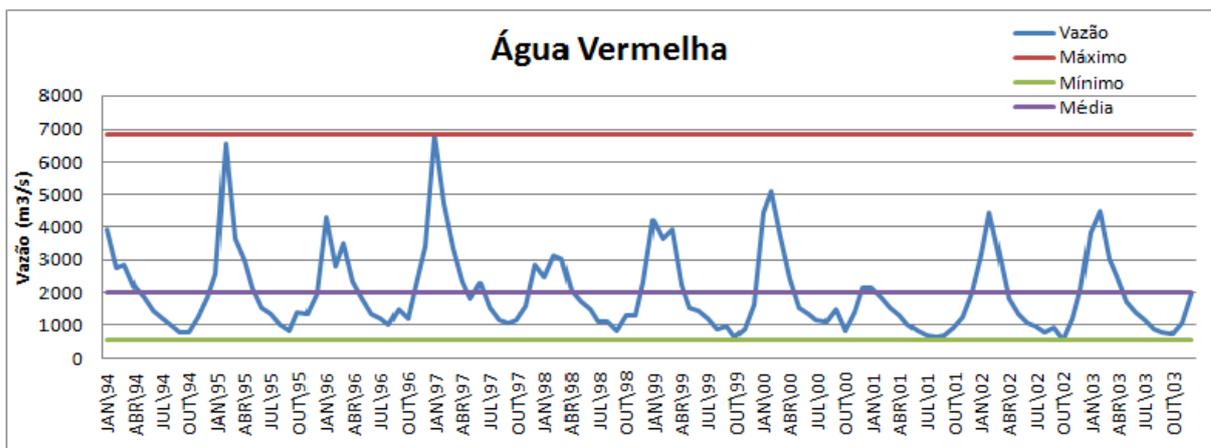


Figura 28 – Série de vazão afluyente média mensal para a usina hidroelétrica Água Vermelha no período de 1994 a 2003.

Fonte: Própria do autor.

A Figura 29 apresenta a distribuição de frequência da vazão no período de 1931 a 2013 para a usina Água Vermelha. Pode-se ver que a maior frequência está a esquerda, nos valores de

vazão menores, com uma calda mais pesada a direita caracterizando uma assimetria dos dados.

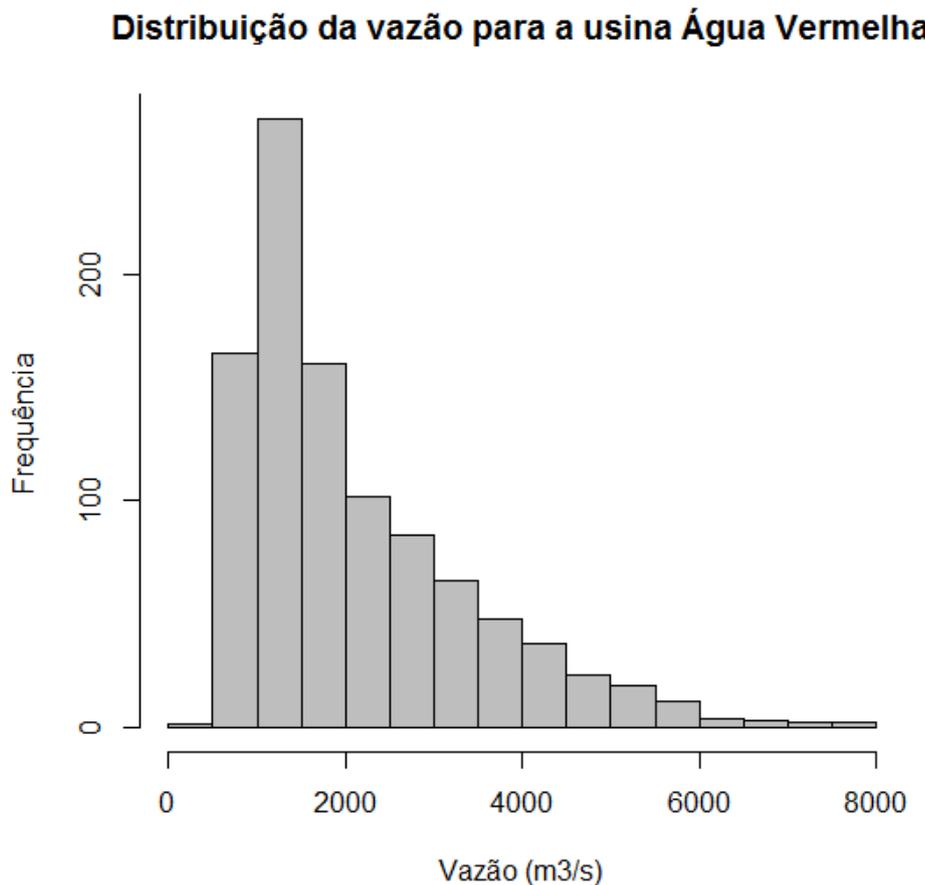


Figura 29 – Distribuição de frequência da vazão no período de 1931 a 2013 para a usina Água Vermelha.

Fonte: Própria do autor.

O comportamento da vazão na usina Itutinga é semelhante ao de Água Vermelha, como pode ser visto na Figura 30, que apresenta os dados para o período de 1994 a 2003. Os dados oscilam em torno da média de $111,83 \text{ m}^3/\text{s}$, com um máximo de $449 \text{ m}^3/\text{s}$ e mínimo de $34 \text{ m}^3/\text{s}$. Comparando as Figuras 28 e 30 é possível perceber a diferença nos valores da vazão para as duas usinas que são diferentes quanto ao tipo, sendo a Água Vermelha de reservatório e portanto há o acúmulo de água nos reservatórios e Itutinga que é uma usina à fio d'água, utiliza somente a água disponível no período.

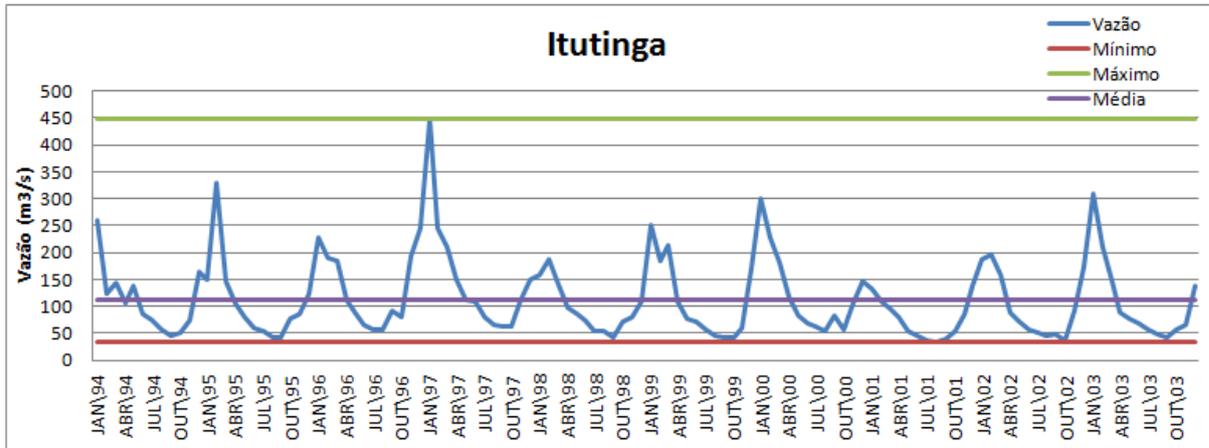


Figura 30 – Série de vazão afluente média mensal para a usina hidrelétrica Itutinga no período de 1994 a 2003. Fonte: Própria do autor.

Esta diferença fica clara também quando comparamos o histograma das Figuras 29 e 31, onde é possível perceber uma maior frequência para vazões mais baixas na usina de Itutinga. Isto acontece pelo fato de usinas sem reservatório terem suas vazões dependentes da quantidade da água que disponível. No histograma da Figura 29, percebe-se que há frequência maior para valores mais altos de vazão, já que Água Vermelha é uma usina de reservatório e permite o acúmulo de água.

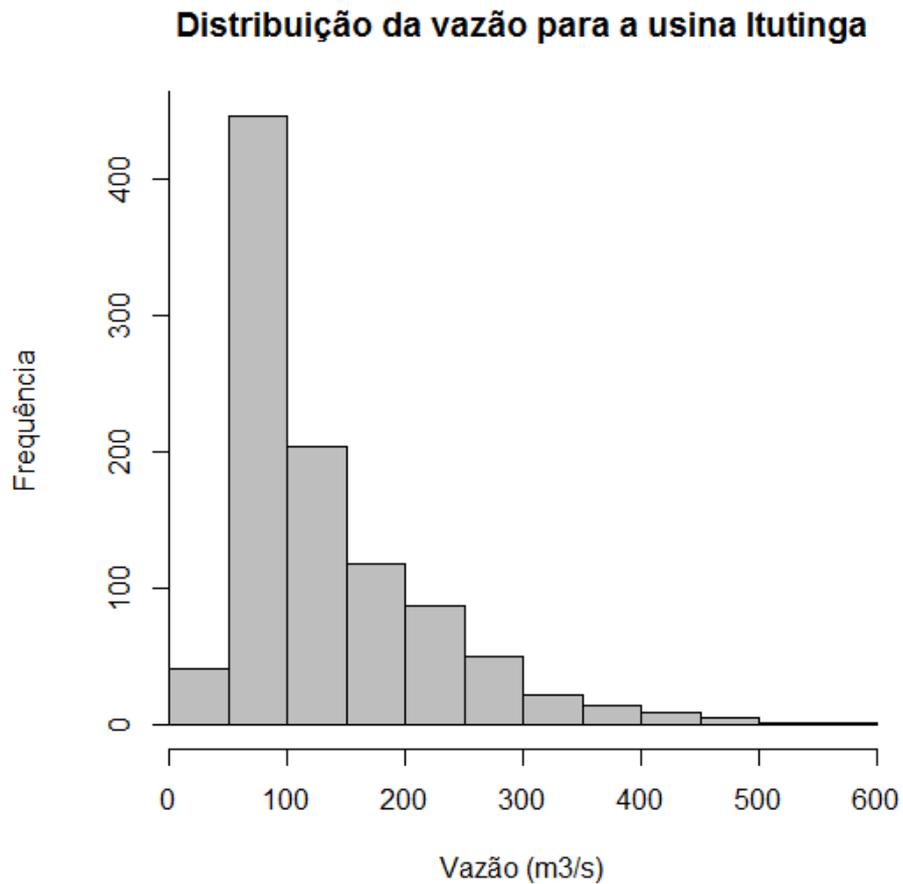


Figura 31 – Distribuição de frequência da vazão no período de 1931 a 2013 para a usina Itutinga.
 Fonte: Própria do autor.

Para o melhor entendimento da variável estudada são apresentadas na Tabela 7, algumas medidas descritivas: o mínimo, o máximo, a amplitude, a média, a mediana e o desvio padrão das séries de vazões das usinas em estudo para os períodos seco e úmido. Além disso, são mostradas estas mesmas medidas para as fases de treinamento, validação e teste já descritas na Seção 4, na Subseção 4.1.

Tabela 7 – Medidas descritivas das séries de vazões (m^3/s) das fases de treino, validação e teste para os períodos seco e úmido das usinas Água Vermelha e Itutinga.

Água Vermelha							
Período Seco							
Fase	Mín.	Máx.	Amp.	Méd.	Mediana	DP.	CV (%)
Treino	484,00	4748,00	4264,00	1297,81	1191,00	537,73	45,15%
Validação	530,00	2291,00	1761,00	1176,00	1137,00	302,10	26,57%
Teste	718,31	2282,25	1563,94	1350,18	1323,48	423,85	32,03%
Período Úmido							
Fase	Mín.	Máx.	Amp.	Méd.	Mediana	DP.	CV (%)
Treino	684,00	7938,00	7254,00	2922,71	2715,00	1364,48	50,26%
Validação	873,00	6827,00	5954,00	2782,83	2478,50	996,02	40,19%
Teste	1110,08	7143,59	6033,50	2999,42	2906,72	1253,52	43,12%
Ano Inteiro							
Fase	Mín.	Máx.	Amp.	Méd.	Mediana	DP.	CV (%)
Treino	484,00	7938,00	7454,00	2110,26	1641,50	1317,19	62,42%
Validação	530,00	6827,00	6297,00	1979,42	1535,50	1235,15	62,40%
Teste	718,31	7143,59	6425,28	2174,80	1776,98	1246,53	57,32%
Itutinga							
Período Seco							
Fase	Mín.	Máx.	Amp.	Méd.	Mediana	DP.	CV (%)
Treino	34,00	332,00	298,00	82,70	76,00	33,93	44,65%
Validação	34,00	138,00	104,00	63,22	58,00	15,34	26,45%
Teste	42,48	125,03	82,55	75,65	71,89	20,35	28,31%
Período Úmido							
Fase	Mín.	Máx.	Amp.	Méd.	Mediana	DP.	CV (%)
Treino	48,00	576,00	528,00	190,34	168,50	89,23	52,95%
Validação	59,00	449,00	390,00	160,43	147,00	54,65	37,18%
Teste	82,83	524,77	441,94	188,83	176,62	85,20	48,24%
Ano Inteiro							
Fase	Mín.	Máx.	Amp.	Méd.	Mediana	DP.	CV (%)
Treino	34,00	576,00	542,00	136,46	106,00	86,36	63,29%
Validação	34,00	449,00	415,00	111,83	85,00	72,14	64,51%
Teste	42,48	524,77	482,29	132,24	101,36	83,87	63,42%

Fonte: Própria do autor.

Percebe-se na Tabela 7 que nos períodos úmidos a média e o desvio padrão são maiores que no período seco tanto para Água Vermelha como para Itutinga, o que foi constatado também por Balline, Andrade e Soares (2003), Lima, Lima e Salgado (2011) e Salgado et al. (2012), que mostram em seus estudos que além da média, há uma diferença no desvio padrão ao longo dos meses, sendo maior nos meses chuvosos. Este fato, pode ser observado também no MLT, na Figura 32. Isto pode ocasionar uma perda no processo de previsão, o que justifica a apresentação

dos resultados deste trabalho separadamente para cada período.

Deve-se ressaltar que existe uma diferença na média da vazão das usinas Água Vermelha e Itutinga em todas as fases (treinamento, validação e teste), assim como a amplitude, sendo mais elevados na primeira que é uma usina de reservatório e por isso apresenta uma quantidade de água disponível maior que Itutinga, já que é possível o armazenamento de água, ao contrário de Itutinga que tem sua vazão dependente do volume de água disponível naquele período.

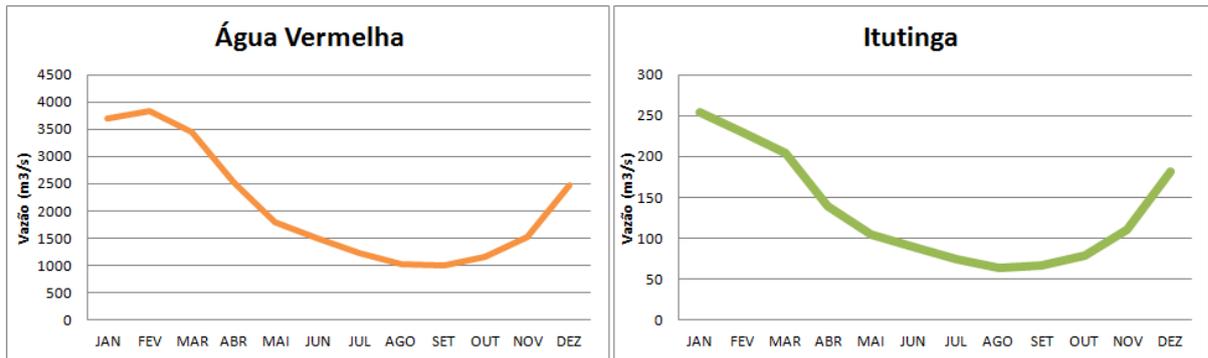


Figura 32 – Média de longo termo (MLT) para as usinas Água Vermelha e Itutinga utilizando dados de vazões do período de 1931 a 2013.

Fonte: Própria do autor.

A Figura 33 mostra a MLT das séries no qual é possível ver como a vazão se comporta ao longo do ano e como esta diminui nos meses de estiagem. Para a construção do MLT da Figura 5.6, os dados foram padronizados através da Equação 5.1. Como pode ser observado o comportamento da curva se mantém mesmo após a padronização, sendo possível a sobreposição destas.

$$z = \frac{x - \bar{x}}{\sigma}, \quad (5.1)$$

em que z é o valor padronizado; x é o valor real da série; \bar{x} é a média dos dados e σ é o desvio padrão dos dados.

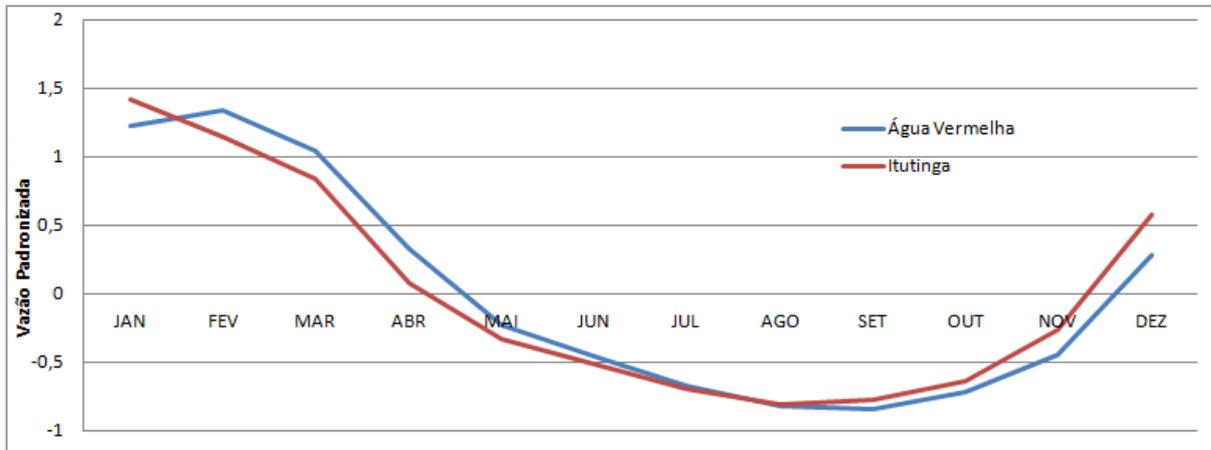


Figura 33 – Média de longo termo (MLT) das séries de vazões padronizadas do período de 1931 a 2013 para as usinas Água Vermelha e Itutinga.

Fonte: Própria do autor.

5.2 ESTUDOS DE CASOS

Os resultados serão apresentados através de tabelas e gráficos para as componentes individuais e os *ensembles* em três subseções: uma para o período seco, uma para o período úmido e outra para os resultados gerais.

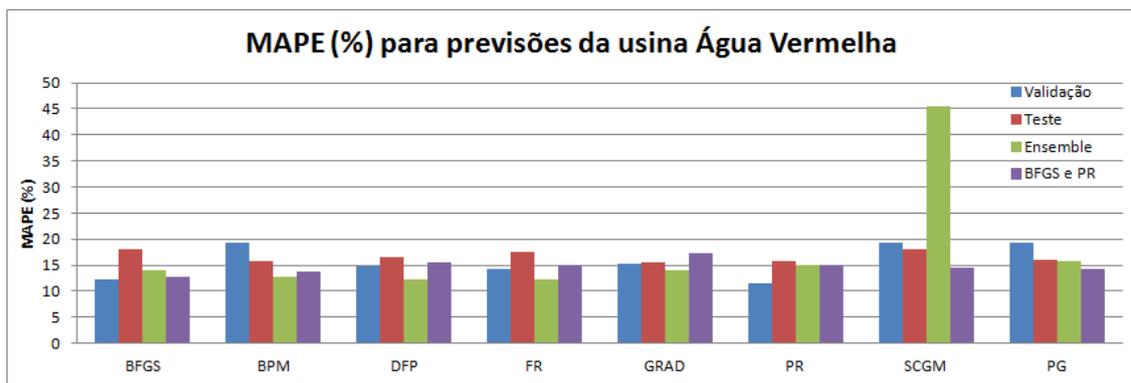
5.2.1 Resultados para o período seco

Para obter as previsões através das componentes individuais utilizou-se os modelos descritos na Seção 4.2 do Capítulo 4 (Seção 4.2) e logo após obter estes resultados, eles foram combinados através da metodologia de *ensembles*.

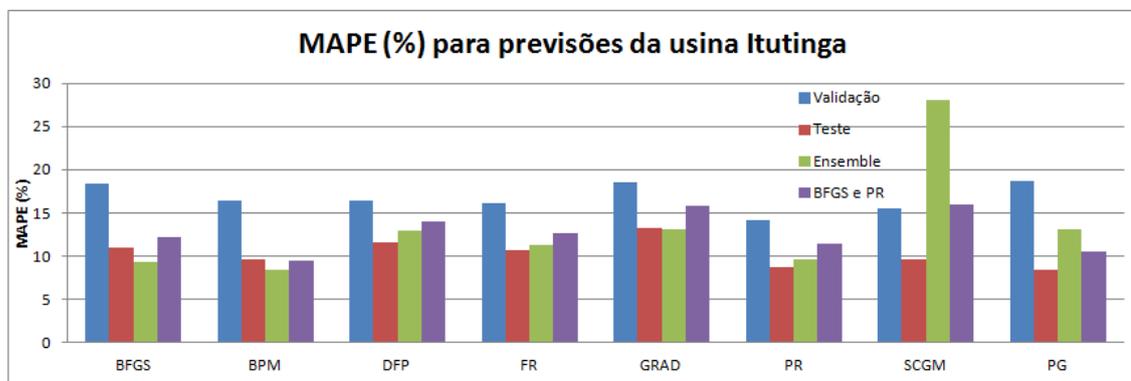
Na Figura 34 encontram-se os resultados para as duas usinas em estudo considerando o período seco. Para Água Vermelha, a componente através da qual foi possível obter melhores resultados foi a RNA com algoritmo de treinamento GRAD, com MAPE de 15,40%. Para esta mesma usina, o *ensemble* que superou o desempenho da melhor componente individual foi o RNA com algoritmo DFP, com MAPE de 12,18%. Já para a usina Itutinga a melhor componente individual foi a PG, com MAPE de 8,47%. Para esta usina, o combinador que apresentou melhores resultados foi a RNA com algoritmo de treinamento BPM, com MAPE de 8,35%.

Quando compara-se os dois tipos de *ensembles* (com oito componentes e com duas)

(FIGURA 34) pode-se perceber que houve melhora no desempenho das previsões em três casos para a usina Água Vermelha: RNA com algoritmos de treinamento BFGS e SCGM e na PG e em dois casos para Itutinga: RNA com algoritmo SCGM e PG. Para os demais modelos não houve melhora e isto pode ser explicado pelo fato de que os períodos de validação e teste serem diferentes e não necessariamente o melhor modelo no período de validação é o melhor no período de teste.



(a)



(b)

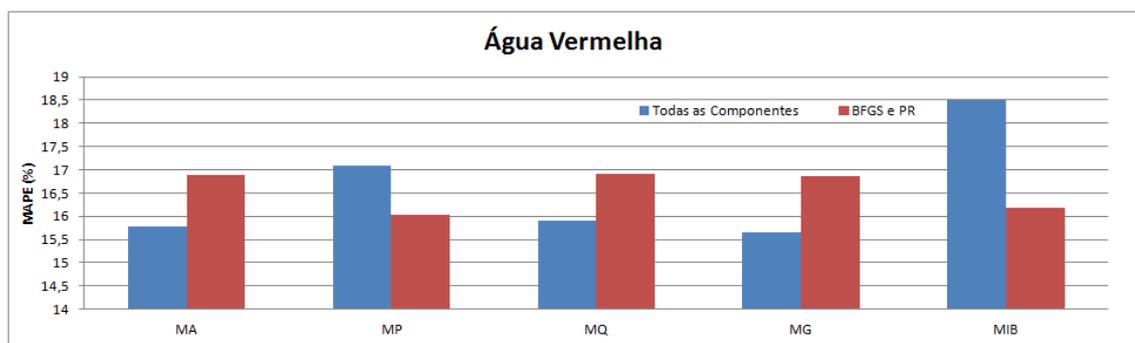
Figura 34 – MAPE (%) das previsões da vazão (m^3/s) para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.

Fonte: Própria do autor.

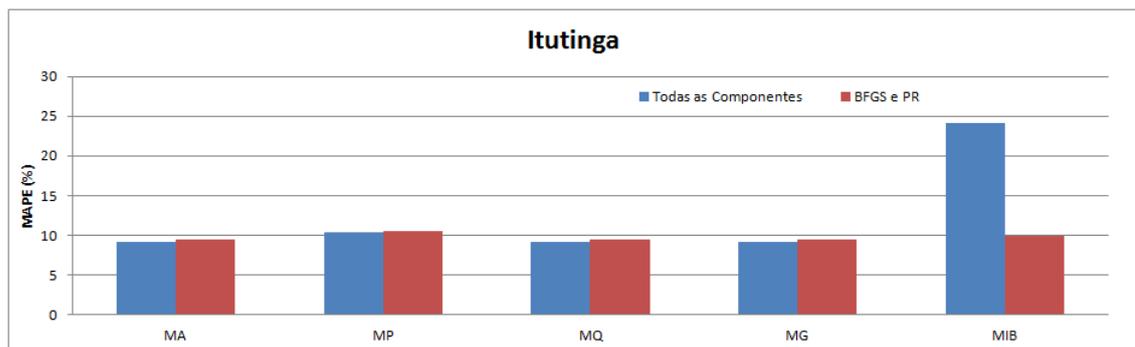
Como pode ser visto através da Figura 34, o desempenho dos modelos nas fases de validação e teste são diferentes, de forma que o melhor modelo na fase de validação seja distinto do melhor modelo na fase de teste. Isto acontece pelo fato de que os períodos correspondentes a estas fases serem diferentes, apresentando suas próprias características e particularidades. Se for observado o coeficiente de variação para estas fases no período seco na Seção 5.1 (TABELA 7), pode-se perceber que este é menor na fase de validação, havendo diferenças também para as

demais medidas descritivas. Estas características influenciam no desempenho dos modelos e na busca de parâmetros, fazendo com que não seja possível eleger um único modelo que preveja o comportamento das séries de vazões de maneira geral.

Na Figura 35 estão os resultados dos *ensembles* MA, MP, MQ, MG e o MIB para o período seco. Como é possível observar, ao fazer uso destes combinadores foi possível obter resultados superiores aos das componentes individuais, ou seja MAPE menores que nas componentes. Porém, o desempenho dos combinadores inteligentes ainda são superiores, pois estes modelos são capazes de atribuir pesos as componentes de previsão de forma mais uniforme que os demais combinadores. Além disso, sabe-se que as médias são influenciadas por valores extremos, logo se algum modelo errou muito ao prever isso vai influenciar também no desempenho destes combinadores.



(a)



(b)

Figura 35 – MAPE (%) das previsões da vazão (m^3/s) realizadas através dos *ensembles*: MA, MP, MQ, MG e MIB para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.

Fonte: Própria do autor.

Comparando os dois tipos de *ensembles* é possível observar que para as médias a diferença nos erros não foi significativa, porém para o MIB ao selecionar as melhores componentes houve uma diferença de aproximadamente 2% e 14% no MAPE das usinas Água Vermelha

e Itutinga respectivamente. Este resultado é explicado pelo fato de que as componentes de desempenho inferior estavam influenciando na eficiência do combinador.

Como é possível observar, o MAPE para a usina Água Vermelha é maior que para usina de Itutinga e está relacionado com o volume da vazão que abastece as usinas. Na Subseção 5.1, onde foi feita uma análise descritiva dos dados fica claro que a variação dos valores mínimo, máximo e conseqüentemente da amplitude da vazão de uma usina para outra é muito grande. O volume de água, por sua vez, é influenciado pela localização destas usinas, o tipo e os fatores ambientais já discutidos na Seção 2.

5.2.2 Resultados para o período úmido

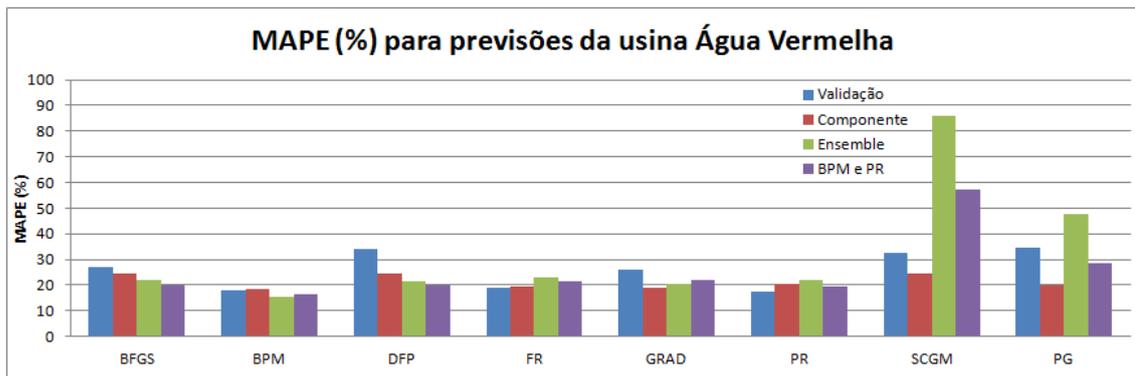
Esta subseção aponta os resultados do período úmido para as usinas em estudo de forma semelhante a Subseção 5.2.1. A Figura 36 mostra os resultados dos modelos. Para Água Vermelha o modelo que mais se destacou foi a RNA com algoritmo de treinamento BPM, tanto como componente individual quanto *ensemble*, com MAPE de 18,46% e 15,24% respectivamente. Para Itutinga a melhor componente foi a RNA com algoritmo de treinamento GRAD, com MAPE de 23,31%. Como *ensemble* o melhor modelo foi a RNA com algoritmo BPM, com MAPE de 20,94%.

Comparando-se os dois tipos de *ensembles* (com oito componentes e com duas) na Figura 36 percebe-se que houve melhora no desempenho das previsões em seis casos na usina Água Vermelha: RNA com algoritmos de treinamento BFGS, DFP, FR, PR e SCGM e na PG e em dois casos para Itutinga: RNA com algoritmo de treinamento SCGM e PG. Para os demais modelos não houve melhora e isto pode ser explicado pelo fato de que os períodos de validação e teste serem diferentes e não necessariamente o melhor modelo no período de validação é o melhor no período de teste.

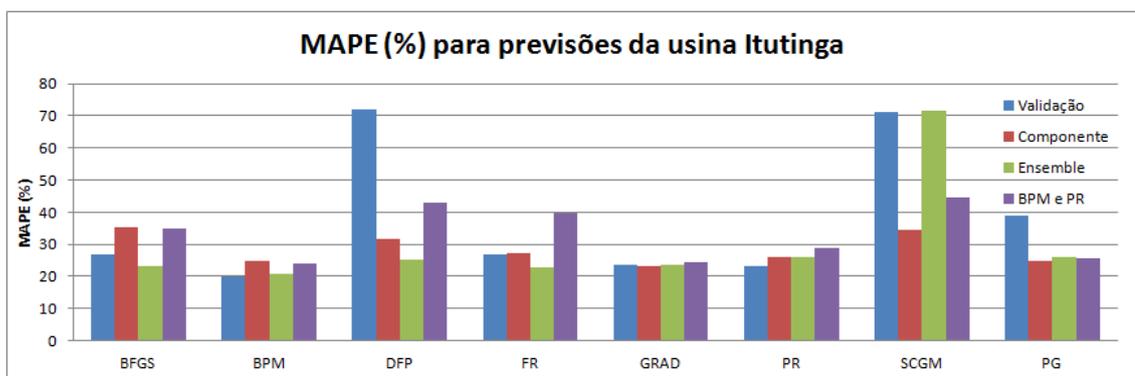
Apesar de diminuir o erro de previsão em alguns casos após utilizar a estratégia de combinar as duas melhores componentes da fase de validação, os resultados para estes *ensembles* não superou o desempenho do melhor combinador utilizando as oito componentes de previsão, pois ao se utilizar duas componentes pode-se estar priorizando um modelo com bom desempenho na fase de validação porém o mesmo não ocorre na fase de teste.

Na usina Água Vermelha o comportamento dos modelos nas fases de validação e teste

foram mais semelhantes no período úmido que no seco, já que o coeficiente de variação destes também foram mais próximos neste período. Já para Itutinga os melhores modelos na fase de validação se diferenciaram da fase de teste. Assim como no período seco, não foi possível eleger um modelo de melhor desempenho para ambas as usinas devidos as particularidades de cada uma e dos fatores que influenciam nas suas vazões, pois os períodos correspondentes as fases de validação e teste são diferentes apresentando suas próprias características particulares.



(a)



(b)

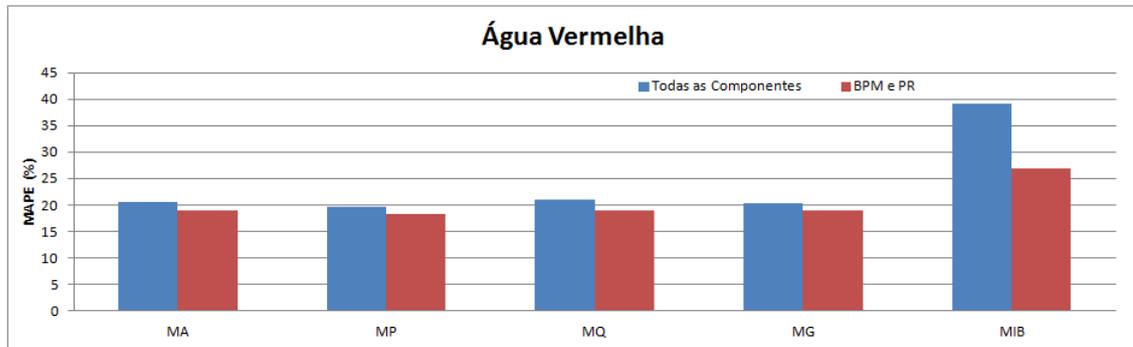
Figura 36 – MAPE (%) das previsões da vazão (m^3/s) para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.

Fonte: Própria do autor.

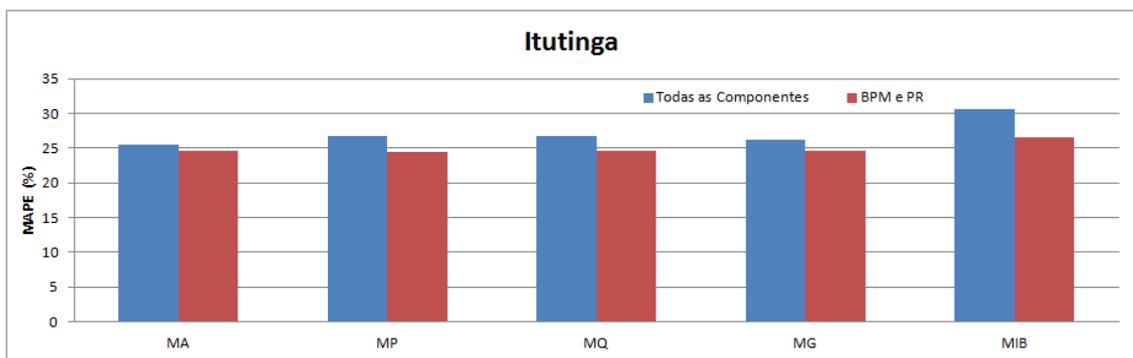
Diferentemente do período seco, no período úmido os erros de previsão são maiores para a usina de Itutinga, pois o coeficiente de variação para esta usina é superior nas fases de treino e teste, já que Itutinga é uma usina a fio d'água e depende da quantidade de água disponível no período e esta quantidade aumenta significativamente nos meses de cheia aumentando a variabilidade.

Na Figura 37 estão os resultados dos *ensembles*: MA, MP, MG, MQ e MIB com as oito componentes de previsão e as duas melhores componentes de previsão na fase de validação.

Como pode ser observado, a melhora no MAPE após utilizar a estratégia de combinar as duas melhores componentes da fase de validação foi ainda maior que no período seco (FIGURA 35), principalmente para o *ensemble* MIB, no qual a diferença nos erros foram de aproximadamente 4% para Itutinga e 12% para Água Vermelha.



(a)



(b)

Figura 37 – MAPE (%) das previsões da vazão (m^3/s) realizadas através dos *ensembles*: MA, MP, MQ, MG e MIB para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.

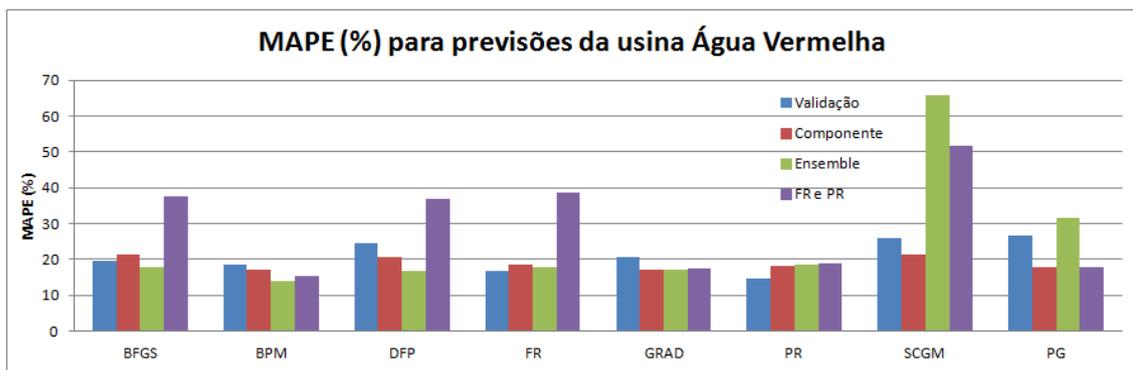
Fonte: Própria do autor.

Comparando os erros de previsão dos períodos secos e úmidos das Subseções 5.2.1 e 5.2.2, é possível perceber como os erros do período úmido são maiores e como a variabilidade que também é maior neste período influencia nas previsões, fazendo com que esse processo seja dificultado, justificando a análise de resultados separadamente para cada período.

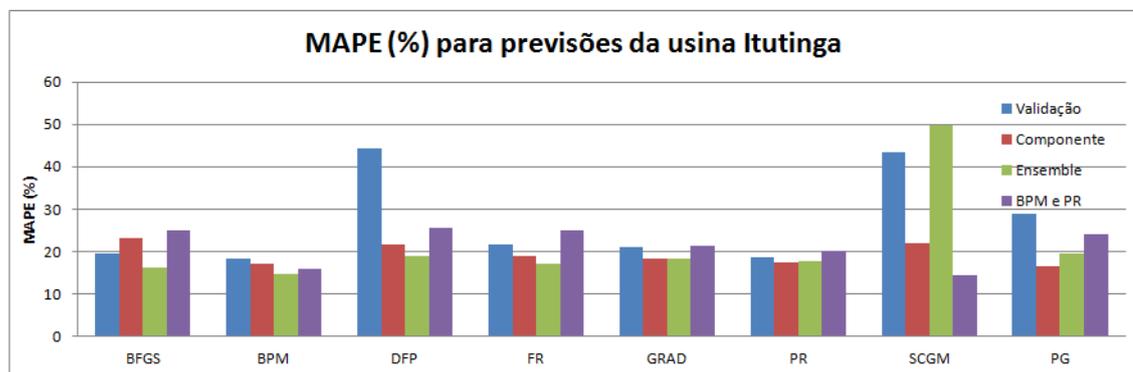
5.2.3 Resultados gerais

Ao determinar o erro de previsão de todos os meses do ano sem distinguir os meses secos dos úmidos, obteve-se os resultados encontrados na Figura 38. Para Água Vermelha, a componente de previsão com menor MAPE foi o modelo RNA com algoritmo de treinamento GRAD (17,12%) e o melhor *ensemble* foi o modelo RNA com algoritmo de treinamento BPM com MAPE de 13,97%. Já para Itutinga a melhor componente e *ensemble* foram PG e RNA com algoritmo BPM respectivamente com MAPE de 16,61% e 14,65%.

Ao comparar os dois tipos de *ensembles* pode-se observar que combinar as duas melhores componentes na fase de validação trouxe melhora apenas para o modelo RNA com algoritmo SCGM no caso da usina Água Vermelha e para Itutinga no modelo RNA também com algoritmo SCGM e na PG.



(a)



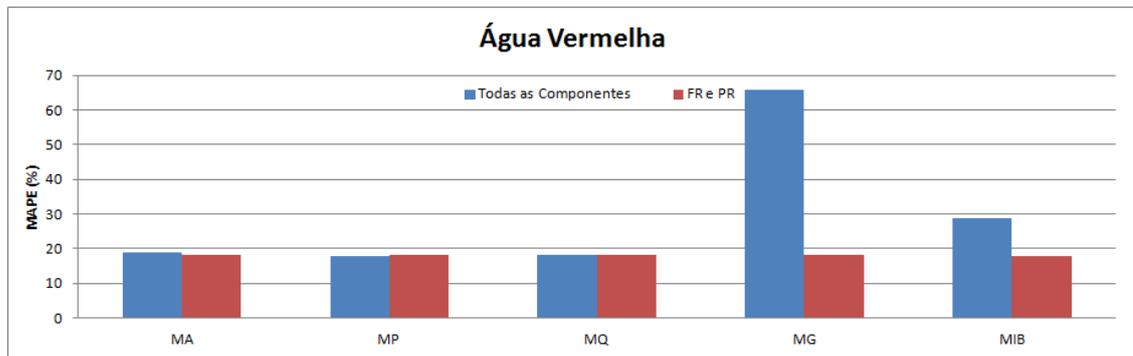
(b)

Figura 38 – MAPE (%) das previsões da vazão (m^3/s) para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.

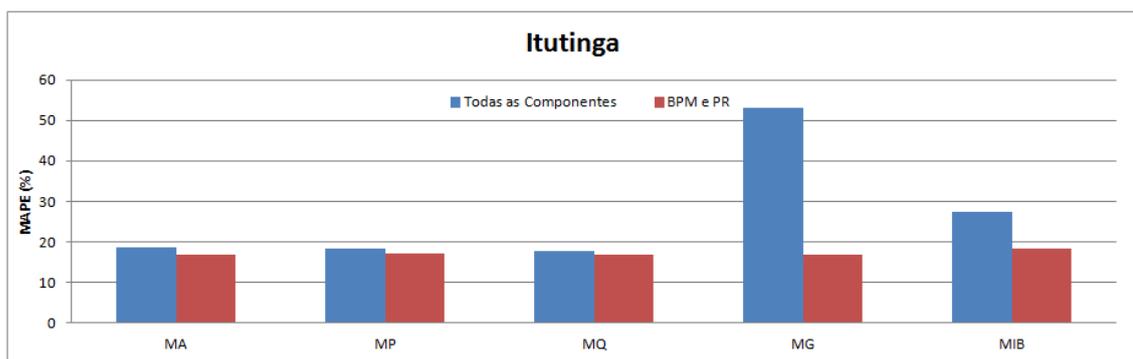
Fonte: Própria do autor.

As previsões realizadas através dos demais *ensembles* estão na Figura 39. Como é

possível perceber através da comparação dos dois *ensembles* houve melhora após utilizar a estratégia de combinar as duas melhores componentes na fase de validação sendo mais evidente na MG (47,6% para Água Vermelha e 36,39% para Itutinga) e no MIB (10,95% para Água Vermelha e 9,07% para Itutinga).



(a)



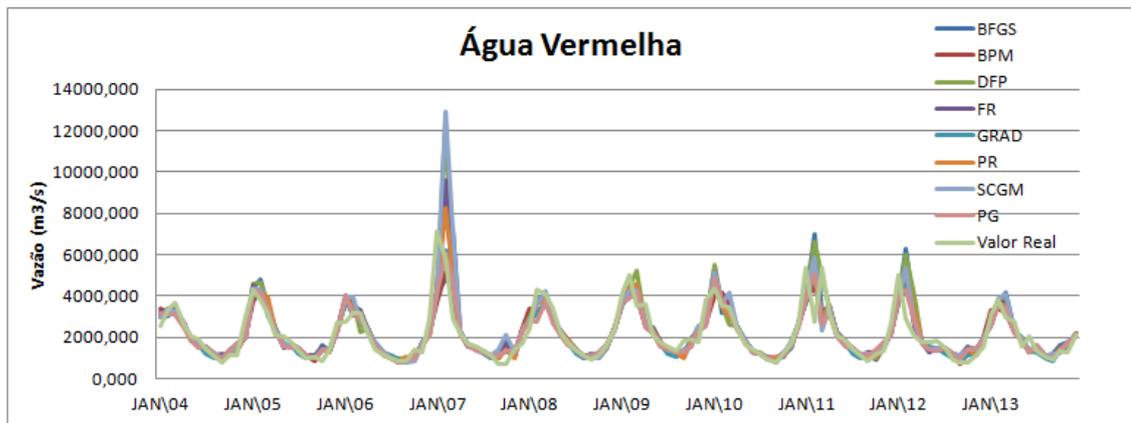
(b)

Figura 39 – MAPE (%) das previsões da vazão (m^3/s) realizadas através dos *ensembles*: MA, MP, MQ, MG e MIB para as usinas Água Vermelha e Itutinga no período de 2004 a 2013.

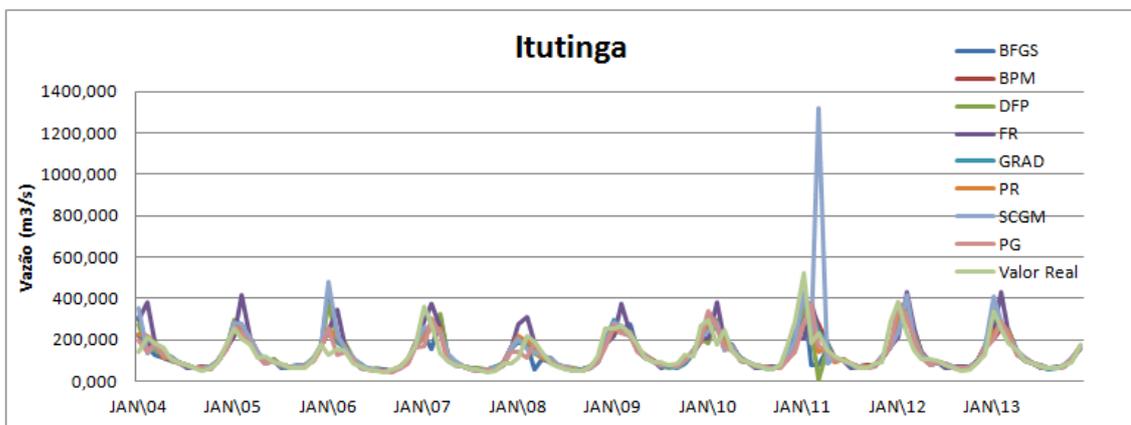
Fonte: Própria do autor.

Como é possível observar através dos resultados das Subseções 5.2.1, 5.2.2 e 5.2.3, a hipótese de que a medida de desempenho dos modelos é influenciada pela variação sazonal da série se confirma. Os erros de previsão foram maiores que no período seco, já que foram influenciados pelos resultados do período úmido, cuja variabilidade e a média são maiores.

Os algoritmos se adaptaram bem as variações da série real como pode ser visto na Figura 40. Um fator relevante e que deve ser destacado são os picos que acontecem nas séries, estes podem influenciar diretamente no aprendizado dos modelos, fazendo com que os erros de previsão nestes períodos sejam maiores, influenciando na medida de desempenho das previsões.



(a)



(b)

Figura 40 – Comparação do desempenho das componentes individuais com os valores reais da séire (2004 a 2013).

Fonte: Própria do autor.

Na Figura 40 pode ser observado que no ano de 2007 ocorre uma variação no desempenho dos algoritmos, isto pode ser observado também na Tabela 8, em que os erros se apresentam maiores para estes ano em todos os modelos. Este fato pode estar relacionado com o volume de precipitação para este período.

Tabela 8 – MAPE (%) das componentes de previsão individuais por ano para as usinas Água Vermelha e Itutinga.

Água Vermelha								
Ano	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
2004	17,02	13,85	15,45	14,97	13,63	14,47	15,51	19,55
2005	16,99	16,59	16,56	14,78	12,32	15,23	14,13	14,04
2006	17,76	13,19	20,38	16,21	16,12	16,20	18,37	13,42
2007	37,44	30,30	32,73	35,82	29,71	32,42	55,27	24,99
2008	8,88	13,06	10,70	8,01	12,51	8,62	7,87	13,58
2009	18,13	16,52	20,88	18,72	19,47	18,34	17,75	18,98
2010	9,57	8,84	10,27	7,00	8,06	8,37	7,21	6,49
2011	27,93	19,07	24,54	22,54	22,23	19,41	24,51	23,54
2012	37,28	25,95	37,55	31,15	23,61	32,36	33,09	29,15
2013	21,08	13,95	16,59	15,77	13,58	16,16	19,32	16,80

Itutinga								
Ano	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
2004	23,08	13,86	19,84	25,53	13,47	12,55	22,90	13,60
2005	15,42	11,27	13,04	20,09	10,85	12,22	13,14	8,61
2006	29,51	15,21	30,22	26,63	18,58	16,43	35,36	17,35
2007	33,74	25,84	33,13	27,46	26,79	25,44	28,41	19,99
2008	23,59	17,48	20,65	24,27	17,35	19,34	19,80	15,24
2009	14,46	12,45	12,01	15,84	15,59	11,84	10,66	11,81
2010	20,56	16,90	21,81	26,34	14,77	18,15	19,68	19,50
2011	24,42	22,42	24,47	25,56	20,74	22,56	53,70	23,76
2012	24,88	23,45	22,76	28,96	19,97	18,63	22,84	14,18
2013	9,64	11,21	8,11	17,74	11,13	11,35	9,39	11,73

Fonte: Própria do autor.

Na Figura 41 pode ser visto o regime de precipitação da região 108, que é a região da Bacia do Rio Grande. A curva em vermelho representa a média de precipitação mensal para o ano de 2007 que em comparação com a média de chuva do período de 1981 a 2010 está bem acima principalmente no mês de janeiro. Este fator pode ter contribuído para a elevação dos erros neste período, justificando o desempenho dos modelos neste ano como foi visto na Tabela 8.

Boxplot de precipitação mensal (1981-2010): Região 108

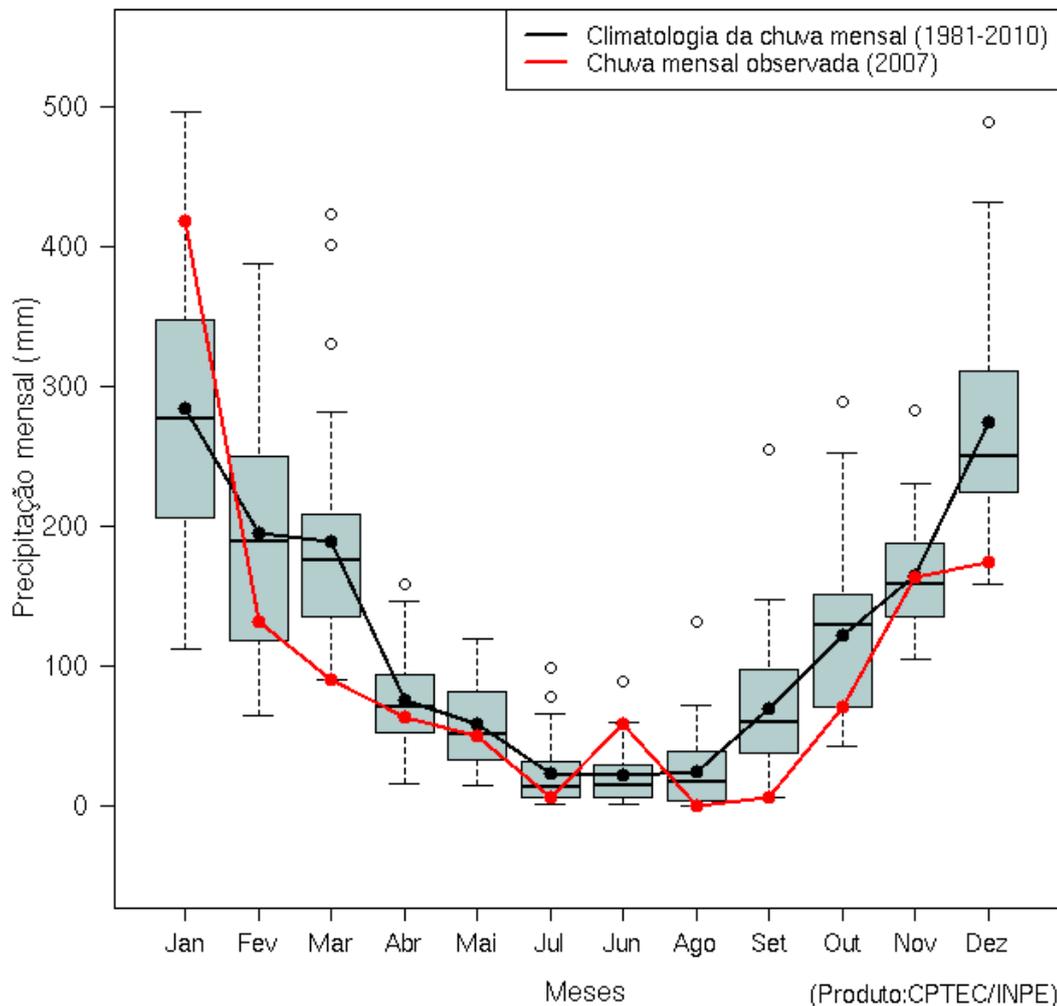


Figura 41 – Regime de precipitação mensal no período de 1981 a 2010 na região da bacia do Rio Grande.

Fonte: CPTEC/INPE (2016).

O desempenho dos modelos como componentes e *ensembles* varia de uma usina para outra, não sendo possível eleger um dos modelos que seja o mais eficiente para todos os casos. Ao compará-los conclui-se que os resultados foram satisfatórios e com erros aceitáveis comparados a trabalhos semelhantes da literatura, como o de Lima, Lima e Salgado (2011) e o de Balline, Andrade e Soares (2003). Para este tipo de dado, deve-se tomar o cuidado de analisar os períodos secos e úmidos separadamente, pois fica evidente como a variabilidade e a sazonalidade influenciam nas previsões ao longo do ano.

6 CONCLUSÃO E PERSPECTIVAS FUTURAS

Este trabalho apresentou o estudo de previsões de séries de vazões para as usinas Água Vermelha e Itutinga através dos modelos: RNA e PG, além disso, a metodologia de *ensembles* foi introduzida com o objetivo de obter previsões mais precisas e acuradas. O desempenho das previsões obtidas pelas componentes individuais e *ensembles* foram comparadas através do MAPE.

É importante enfatizar que não houve um modelo que se destacou em todos os casos quanto aos erros de previsão, os resultados dependem das características da usina e também do período em estudo.

De modo geral, o MAPE da previsão dos modelos no período seco é menor que no período úmido, isso se deve ao fato de esse período apresentar menor amplitude de variação.

Através das análises realizadas foi possível observar que as redes neurais artificiais do tipo MLP e a programação genética são modelos promissores para realizar previsões. Os algoritmos de treinamento mostraram bom desempenho, sendo que os erros de previsão são aceitáveis para o tipo de dado trabalhado, muitas vezes sendo inferiores a outras metodologias utilizadas na literatura.

Fazer previsões por períodos levaram a cometer menores erros que quando considerado todo o ano. As melhores componentes e combinadores foram diferentes para mesma localidade quando analisou-se os períodos secos e úmidos separadamente.

A estratégia de combinar previsões obtidas através de diferentes modelos resultou na diminuição do MAPE, mostrando ser uma técnica promissora para previsão de séries de vazões, possibilitando erros de previsão 14% mais baixos.

Como propostas futuras é possível que sejam aplicados outros modelos computacionais como componentes individuais e como *ensemble*. Além disso, pode-se realizar as previsões utilizando um horizonte diário e semanal através dos modelos estudados neste trabalho.

REFERÊNCIAS

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL). **Matriz de energia elétrica**. Disponível em: <<http://www.aneel.gov.br/>>. Acesso em: 19 set. 2016.
- ANDRADE, M. G. et al. Análise do erro de previsão de vazões mensais com diferentes horizontes de previsão. **Revista Controle & Automação**, v. 23, n. 3, 2012.
- BALLINE, R.; ANDRADE, M. G.; SOARES, S. Previsão Médias Mensais Usando Redes Neurais Nebulosas. **Revista Controle & Automação**, v. 14, n. 3, 2003.
- BAYER, D. M.; CASTRO, N. M. R. Modelagem e previsão de vazões médias mensais do rio Potiribu utilizando modelos de séries temporais. **Revista Brasileira de Recursos Hídricos**, v. 17, n. 2, p. 229-239, 2012.
- BAZARAA, M.; SHERALI, H. D.; SHETTY, C. M. **Nonlinear Programming: Theory and Algorithms**. 2.ed. John Wiley & Sons Inc., 1993, p. 265-282.
- BESAW, L. E.; RIZZO, D. M.; BIERMAN, P. R. Advances in ungauged streamflow prediction using artificial neural networks. **Journal of Hydrology**, v. 386, p. 27-37, 2010.
- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis, Forecasting and Control**. 4.ed. California, USA, San Francisco: Holden Day, 1976, p. 746
- BOX, G. E. P.; TIAO, G. C.. **Bayesian inference in statistical analysis**. 2.ed. New York: John Wiley, 1992, p. 608.
- BRAVO, J. M. et al. Previsões de curto prazo de vazão afluente ao reservatório de Furnas utilizando redes neurais artificiais. **Revista Brasileira de Recursos Hídricos**, v. 13, n. 2, p. 77-88, 2008.

BREIMAN, L. **Bagging predictors**. Machine Learning, vol. 24, pp. 123-140, 1996.

CASELLA, G.; BERGER, R. L. **Inferência Estatística**. 2.ed. São Paulo: Cengage Learning, 2010.

CENTRO DE PREVISÃO DE TEMPO E ESTUDOS CLIMÁTICOS (CPTEC). **Evolução mensal e sazonal das chuvas**. Disponível em: <<http://clima1.cptec.inpe.br/evolucao/pt>>. Acesso em: 09 fev. 2016.

CHERKAUER, K. J. **Human expert level performance on a scientific image analysis task by a system using combined artificial neural networks**. In P. Chan, S. Stolfo, D. Wolpert 320 (Eds), Proceeding AAAI-96 Workshop on Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms, Portland, OR, AAAI Pres, Menlo Park, CA, p. 15-21, 1996.

CRUZ, M. F. M.; RODRIGUES, L. D.; VERSIANI, B. R. Previsão de vazões com a metodologia DPFT e com redes neurais artificiais. **Revista Brasileira de Recursos Hídricos**, v. 15, n. 1, p. 121-132, 2010.

CUNNINGHAM, P.; CARNEY, J.; JACOB, S. Stability problems with artificial neural networks and the ensemble solution. **Artificial Intelligence in Medicine**, v. 20, n. 3, p. 217-225, 2000.

DARWIN, C. **On the Origin of Species by Means for Natural Selection**. London: Jhon Murray, 1859.

DRUCKER, H.; SHAPIRE, R.; SIMARD, P. **Improving performance in neural networks using a boosting algorithm**. In S. J. Hanson, J. D. Cowan, C. L. Giles (Eds.), Advances in Neural Information Processing System 5, Denver, CO, Morgan Kaufmann, San Mateo, CA, p. 42-49, 1993.

DRUCKER, H. et al. Boosting and other ensemble methods. **Neural Computation**, v. 6, n. 6, pp. 1289-1301, 1994.

EFRON, E.; TIBSHIRANI, R. **An introduction to the Bootstrap**. Chapman & Hall, New York, 1993.

FADIGA Jr, F. M.; LOPES, J. E. G.; MARTINS, J. R. S. Modelos de Previsão de Vazões para a Bacia Incremental à UHE Itaipu. **Revista Brasileira de Recursos Hídricos**, v. 13, n. 2, p. 65-75, 2008.

FORTUNATO, L. A. M. et al. **Introdução ao Planejamento de Expansão e Operação de Sistemas de Produção de Energia Elétrica**. Niterói,RJ: Editora Universitária, 1990.

FREUND, Y.; SHAPIRE, R. E. **A decision-theoretic generalization of on-line learning and an application to boosting**. Proceedings EuroCOLT-94, Barcelona, Spain, Springer-Verlag, Berlin, 1995, p. 23-27.

FREUND, Y.; SHAPIRE, R. E. **Experiments with a new boosting algorithm**. Proceedings of the Thirteenth International Conference on Machine Learning, p. 149-156. Morgan Kaufmann, 1996.

GAMERMAN, D. **Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. Texts in Statistical Sciences. Chapman and Hall, London, 1997.

GAMERMAN, D.; LOPES, H. **Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. Texts in Statistical Science Series. CRC Press, 2006.

GELFAND, A. E.; SMITH, A. F. M. Sampling-Based Approaches to Calculating Marginal Densities. **Journal of the American Statistical Association**, v. 85, p. 398-409, 1990.

GEWEKE, J. **Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments**. N: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. (Ed.). Bayesian Statistics 4. New York: Oxford University Press, p. 625-631, 1992.

GOMES, L. F. C.; MONTENEGRO, S. M. G. L. Modelo baseado na técnica de redes neurais para previsão de vazões na bacia do rio São Francisco. **Revista Brasileira de Recursos Hídricos**, v. 15, n. 1, p. 05-15, 2010.

GUTTA, S.; WECHSLER, H. **Face recognition using hybrid classifier systems**. Processing ICNN-96, Washington, DC, IEEE Computer Society Press, Los Alamitos, CA, p.1017-1022, 1996.

HANSEN, L. K.; SALAMON, P. Neural network ensembles. **IEEE Transactions Pattern Analysis and Machine Intelligence**, v. 12, n. 10, p. 993-1001, 1990.

HASHEM, S.; SCHMEISER, B. Improving model accuracy using optimal linear combinations of trained neural networks, **IEEE Transactions on Neural Networks**, v. 6, n. 3, p. 792-794, 1995.

HASHEM S. Optimal linear combinations of neural networks. **Neural Networks**, v. 10, n. 4, p. 599-614, 1997.

HAYKIN, S. **Neural Networks-A Comprehensive Foundation**, 1994.

HAYKIN, S. **Redes Neurais: Princípios e Práticas**. 2 ed. Editora Bookman, 2007.

HEIDELBERGER, P.; WELCH, P. Simulation Run Length Control in the Presence of an Initial Transient. **Operations Research**, v. 31, p. 1109-1144, 1983.

HOLLAND, J. H. **Adaptation in natural and artificial systems**. The University of Michigan Press, Ann Arbor, MI, 1975.

HUANG, F. J. et al. **Pose invariant face recognition**. Processing 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, IEEE Computer Society Press, Los Alamitos, CA, p. 245-250, 2000.

INOUE, H.; NARIHISA, H. Predicting Chaotic Time Series by Ensemble Self-Generating, **International Joint Conference on Neural Networks IJCNN**, v. 2, p. 231-236, 2000.

KISI, O. Neural Networks and Wavelet Conjunction Model for Intermittent Streamflow Forecasting. **Journal of Hydrologic Engineering**, v. 14, p. 773-782, 2009.

KISI, O.; CIMEN, M. A. Wavelet-support vector machine conjunction model for monthly streamflow forecasting. **Journal of Hydrology**, v. 399, p. 132-140, 2011.

KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. Livraria da física, 1997, p. 107

KOZA, J. R. Hierarchical genetic algorithms operating on populations of computer programs. **In Proceedings of the 11th International Joint Conference on Artificial Intelligence**. CA: Morgan Kaufmann. v. 1, p. 768-774, 1989.

KOZA, J. R. **Genetic Programming: On the Programming of Computers by Means of Natural Selection**. MIT Press, Cambridge, USA, 1992.

LI, Y.; WANG, D. F.; HAN, P. Selective Ensemble Using Discrete Differential Evolution Algorithm For Short-Term Load Forecasting. **Eighth International Conference On Machine Learning And Cybernetics**, Baoding, p. 12-15 July 2009.

LIMA, C. A. M.; COELHO, A. L. V.; VON ZUBEN, F. J. Ensembles of Support Vector Machines for Regression Problems. **Proceedings of International Joint Conference on Neural Networks IJCNN**, p. 2381-2386, 2002.

LIMA, C. A. M. **Uma abordagem unificada empregando máquinas de vetores suporte**. 2004. 378 f. Tese (Doutorado em Engenharia Elétrica). Universidade Estadual de Campinas, Campinas - SP, 2004.

LIMA, D. B. **Modelo inteligente de previsão de vazões afluentes para o auxílio à tomada de decisão no setor elétrico**. 2010. 108 p. Monografia(Bacharelado em Ciência da Computação). Universidade Federal de Alfenas, Alfenas - MG, 2010,

LIMA, D. B.; LIMA, M. D. C. E.; SALGADO, R. M. An Empirical Analysis of MLP Neural Networks Applied to Streamflow Forecasting. **IEEE International Conference on Intelligent Computing and Intelligent Systems**, v. 9, n. 3, p. 295-301, 2011.

LUCAS, E. W. M et al. Modelagem hidrológica determinística e estocástica aplicada à região hidrográfica do Xingu, Pará. **Revista Brasileira de Meteorologia**, v. 24, n. 3, p. 308-322, 2009.

LUENBERGER, D. G. **Linear and Nonlinear Programming**. 2. ed., 1989.

MAKKEASORN, A.; CHANG, N. B.; ZHOU, X. Short-term streamflow forecasting with global climate change implications: A comparative study between genetic programming and neural network models. **Journal of Hydrology**, v. 352, p. 336-354, 2008.

MODARRES, R.; ESLAMIAN, S. S. Streamflow time series modeling of Zayandehrud river. **Iranian Journal of Science & Technology**, v. 30, n. 4, 2006.

MOREIRA, M. O. et al. Modelo Baseado em Combinação de Previsores para Previsão de Séries Temporais de Carga Elétrica. **Revista GEINTEC**. São Cristóvão/SE, v. 1, n. 2, p. 20-31, 2011.

MORETTIN, P. A., TOLOI, C. M. C. **Análise de Séries Temporais**. 2. ed. São Paulo: Editora Edgard Blucher,2006, p. 511.

NOGUEIRA, D. A. **Proposta e avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov: casos uni e multivariados**. 2004. 135 f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária). Universidade Federal de Lavras, Lavra - MG, 2004.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). **Séries Históricas de Vazões**. Disponível em: <<http://www.ons.org.br/operacao/vazoesnaturais.aspx>> Acesso em: 29 jul. 2016.

PAULINO, C. D.; AMARAL, M. A. e MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkians. p.445, 2003.

PEARLMUTTER, B. A. Fast Exact Calculation by the Hessian. **Neural Computation**, v. 6, p. 147-160, 1994, URL:

R Core Team: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em <http://www.r-project.org>. Acesso em 30 de janeiro de 2014.

RAFTERY, A. E.; LEWIS, S. Comment: One long run with diagnostics: implementation strategies for markov chain monte carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 493-497, 1992.

RAVIV, Y.; INTRATOR, N. **Variance Reduction via Noise and Bias in constraints**. In *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems*, ed. Amanda Sharkey, Springer-Verlag London Ltd, 1999.

REZENDE, S. O. **Sistemas Inteligentes**. 1. ed. Rio de Janeiro: Editora Manole Ltda, 2005, p.513.

ROSA, G. J. M. **Análise Bayesiana de modelos lineares mistos robustos via Amostrador de Gibbs**. 1998. 57f. Tese (Doutorado em Estatística e Experimentação Agronômica) - Universidade de São Paulo, Piracicaba, SP.

RUMELHART, D., E.; MCCLELLAND, J. L., and the PDP Research Group. **Parallel Distributed Processing: Exploration in the Microstructure of Cognition.**, MIT Press, v. 1. , Cambridge, Massachusetts, 1986.

SALGADO, R. M. et al. A Hybrid Ensemble Model Applied to the Short-Term Load Forecasting Problem. **In: International Joint Conference on Neural Networks**, 2006. Vancouver. Canadá: Sheraton Vancouver Wall Centre Hotel, p.16-21, 2006.

SALGADO, R. M. et al. Modelos de Inteligência Computacional para Geração de Séries Sintéticas de Vazões Médias Mensais. **Learning and Nonlinear Models**, v. 10, n. 3, p. 166-174, 2012.

SCHAPIRE, R. E. The strength of weak learnability. **Machine Learning**, v. 5, p. 197-227, 1990.

SHARKEY, A. J. C.; SHARKEY, N. E.; CHANDROTH, G. O. Neural nets and diversity. **Neural Computing and Application**, v. 4, p. 218-227, 1996.

SHARKEY, A. J. C.; SHARKEY, N. E. Combining diverse neural nets. **The Knowledge Engineering Review**, v. 12, n. 3, p. 231-247, 1997.

SHEPHERD, A. J. **Second-Order Methods for Neural Networks Fast and reliable Methods for Multi-Layer Perceptrons**, Springer, 1997.

SILVA, H. A. O. **Power Map Explorer: Uma Ferramenta para visualização e previsões de vazões**. 2007. 112 p. Dissertação (Mestrado em Ciências Matemáticas e de Computação). Universidade de São Paulo (USP), São Carlos, 2007.

SOARES FILHO, S. Planejamento de operações de sistemas hidrotérmicos. **Controle de Automação**, v. 1, n. 2, p. 122-131, 1987.

TUCCI, C. E. M.; PORTO, R. L. L.; BARROS, M. T. **Drenagem Urbana**. Editora da Universidade/ ABRH/ UFRGS, 1995.

TUCCI, C. E. **Regionalização de Vazões**. Porto Alegre: UFRGS/ANELL/ANA/ABRH, 202, p.250.

TUMER, K.; GHOSH, J. Error correlation and error reduction in ensemble classifiers. **Connection Science. Special Issue on Combining Artificial Neural: Ensemble Approaches**, v. 8, n. 3, p. 385-404, 1996.

VILLANUEVA, J. P. W. **Comitê de Máquinas em Predição de Séries Temporais**. 2006. 178f. Tese Mestrado Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas, 2006.

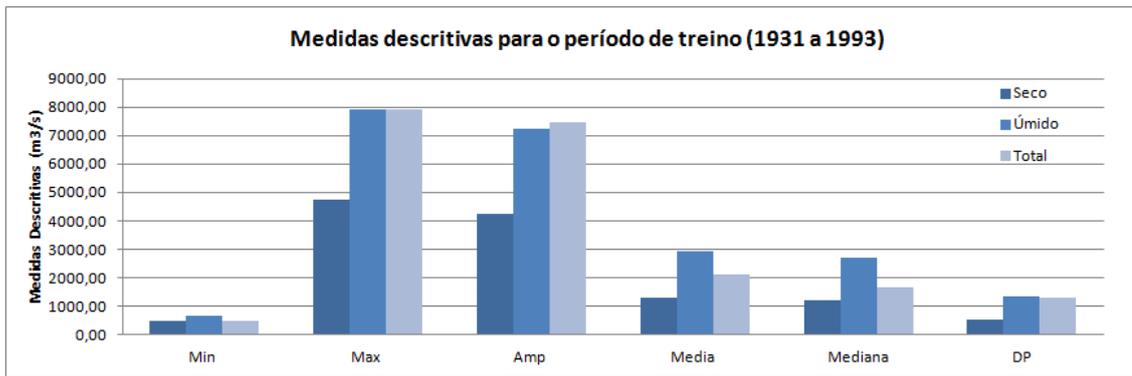
WANG, W. et al. Genetic programming for modelling long-term hydrological time series. **IEE-Fifth International Conference on Natural Computation**, 2009.

ZHANG, Q., PATUWO, B. e HU, M. Forecasting with Artificial Neural Networks: the State of the Art. **International Journal of Forecasting**, p. 35-62, 1998.

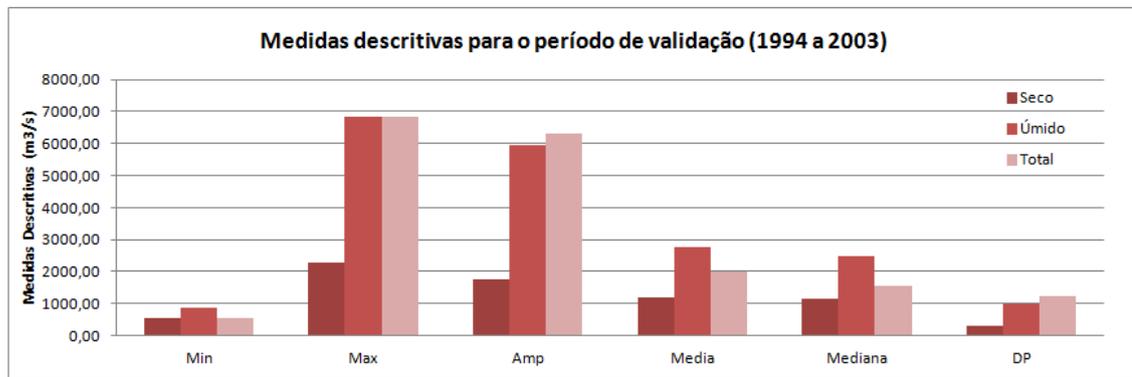
ZHOU, Z. H.; WU, J.; TANG, W. **Ensembling Neural Networks: Many Could Be Better Than All**. *Artificial Intelligence*, v.137, n.1-2, p.239-263, 2002.

APÊNDICE

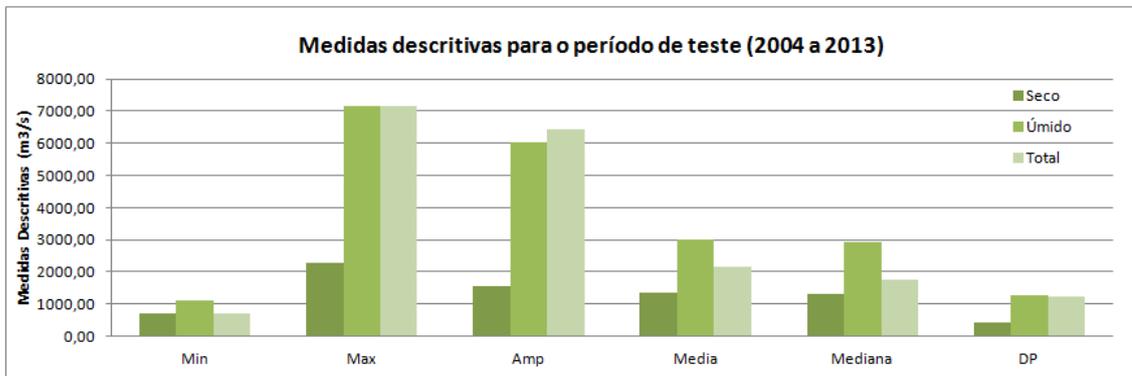
APÊNDICE A - Gráficos das medidas descritivas da vazão (m^3/s) para as usinas Água Vermelha e Itutinga



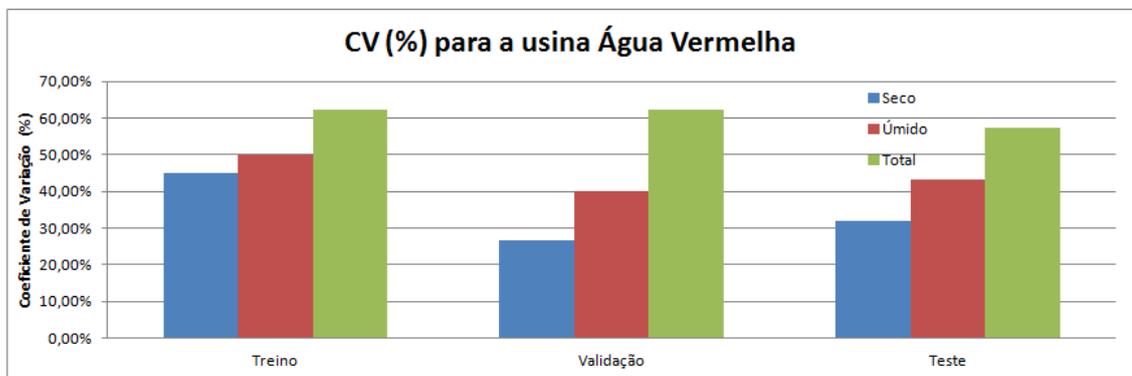
(a)



(b)

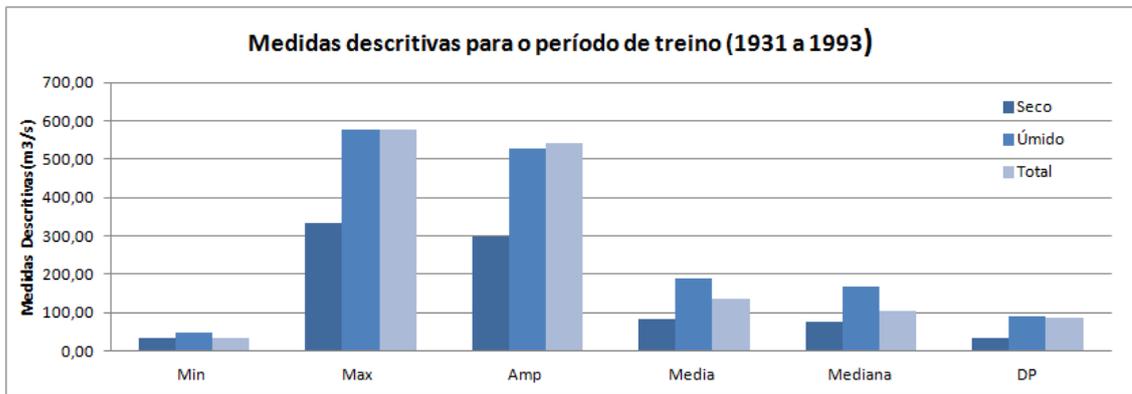


(c)

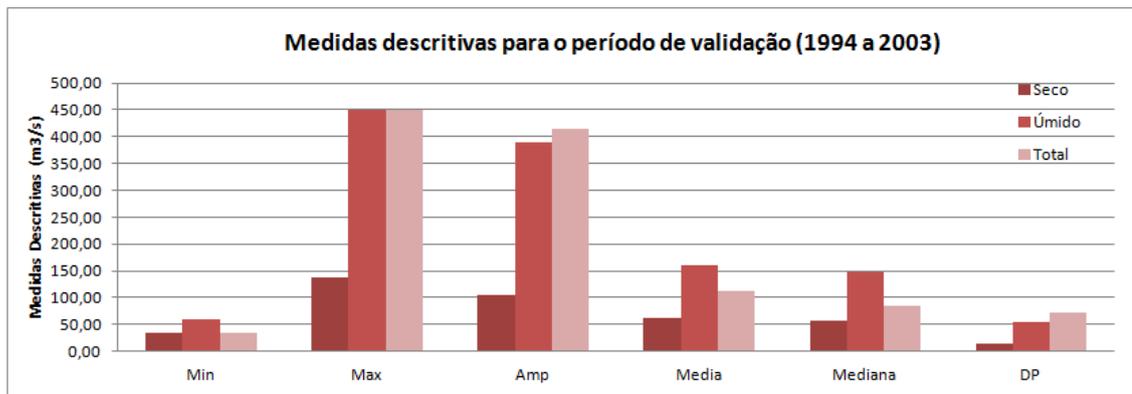


(d)

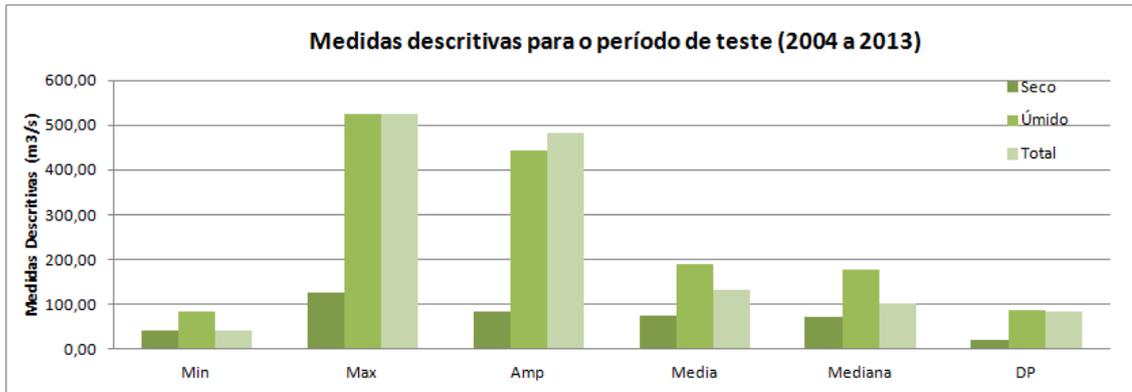
Figura 42 – Medidas de posição e variabilidade da vazão (m^3/s) das fases de treino (1931 a 1993), validação (1994 a 2003) e teste (2004 a 2013) para a usina Água Vermelha.



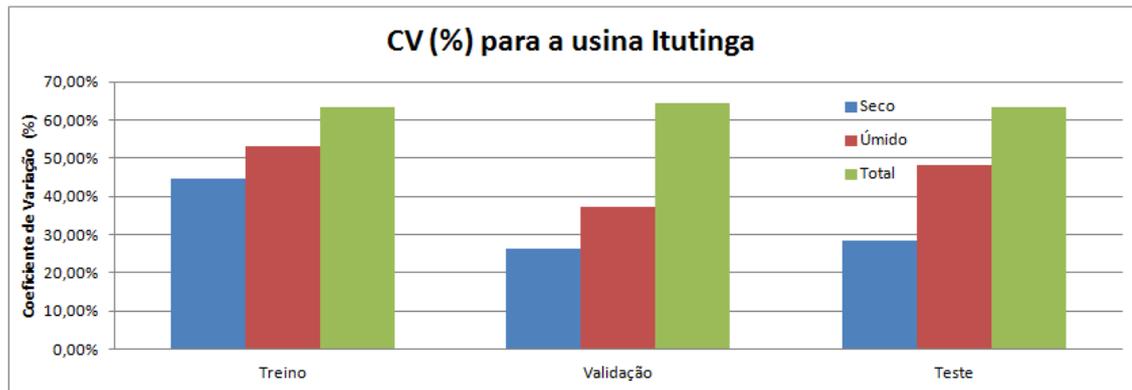
(a)



(b)



(c)



(d)

Figura 43 – Medidas de posição e variabilidade da vazão (m^3/s) das fases de treino (1931 a 1993), validação (1994 a 2003) e teste (2004 a 2013) para a usina Itutinga.

APÊNDICE B - Tabelas contendo os resultados das previsões das vazões (m^3/s) para as usinas Água Vermelha e Itutinga

Tabela 9 – MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 das usinas Água Vermelha e Itutinga no período seco.

Água Vermelha								
	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
Validação	12,14	19,27	14,75	14,38	15,17	11,59	19,36	19,19
Componente	18,11	15,81	16,57	17,46	15,40	15,82	18,11	16,04
<i>Ensemble</i>	13,91	12,70	12,18	12,32	14,12	14,96	45,44	15,71
BFGS e PR	12,62	13,69	15,63	15,09	17,20	14,97	14,63	14,32
Itutinga								
	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
Validação	18,33	16,35	16,46	16,07	18,48	14,08	15,49	18,74
Componente	10,95	9,63	11,64	10,63	13,17	8,75	9,61	8,47
<i>Ensemble</i>	9,31	8,35	12,95	11,24	13,05	9,57	28,01	13,03
BFGS e PR	12,20	9,42	14,03	12,56	15,76	11,42	15,98	10,59

Fonte: Própria do autor.

Tabela 10 – MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 do período seco realizadas através dos *ensembles*: MA, MP, MQ, MG e MIB.

Água Vermelha					
<i>Ensemble</i>	MA	MP	MQ	MG	MIB
Todas as componentes	15,77	17,08	15,90	15,64	18,51
BFGS e PR	16,89	16,02	16,92	16,85	16,17
Itutinga					
<i>Ensemble</i>	MA	MP	MQ	MG	MIB
Todas as componentes	9,11	10,37	9,18	9,16	24,07
BFGS e PR	9,44	10,58	9,47	9,41	9,95

Fonte: Própria do autor.

Tabela 11 – MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 das usinas Água Vermelha e Itutinga no período úmido.

Água Vermelha								
	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
Validação	26,79	18,03	34,26	18,89	25,86	17,54	32,54	34,34
Componente	24,30	18,46	24,56	19,54	18,85	20,50	24,51	19,81
<i>Ensemble</i>	22,11	15,24	21,21	23,19	20,27	21,91	85,92	47,57
BPM e PR	19,77	16,43	19,76	21,21	21,86	19,48	57,02	28,44
Itutinga								
	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
Validação	26,79	20,19	72,05	26,97	23,57	23,38	71,04	39,05
Componente	35,13	24,67	31,62	27,17	23,31	26,11	34,48	24,75
<i>Ensemble</i>	23,23	20,94	25,10	22,93	23,67	26,06	71,52	26,19
BPM e PR	34,94	23,82	42,89	39,66	24,55	28,89	44,50	25,62

Fonte: Própria do autor.

Tabela 12 – MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 do período úmido realizadas através dos *ensembles*: MA, MP, MQ, MG e MIB.

Água Vermelha					
<i>Ensemble</i>	MA	MP	MQ	MG	MIB
Todas as componentes	20,58	19,56	20,88	20,29	39,22
BPM e PR	18,93	18,38	19,01	18,85	26,84
Itutinga					
<i>Ensemble</i>	MA	MP	MQ	MG	MIB
Todas as componentes	25,52	26,68	26,71	26,14	30,63
BPM e PR	24,67	24,46	24,7	24,65	26,59

Fonte: Própria do autor.

Tabela 13 – MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 das usinas Água Vermelha e Itutinga para todo o ano.

Água Vermelha								
	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
Validação	19,46	18,65	24,50	16,63	20,51	14,57	25,95	26,77
Componente	21,21	17,13	20,57	18,50	17,12	18,16	21,31	17,92
Ensemble	18,01	13,97	16,70	17,75	17,19	18,43	65,68	31,64
FR e PR	37,59	15,47	36,74	38,74	17,34	18,91	51,56	17,90
Itutinga								
	BFGS	BPM	DFP	FR	GRAD	PR	SCGM	PG
Validação	19,46	18,27	44,25	21,52	21,02	18,73	43,26	28,90
Componente	23,04	17,15	21,63	18,90	18,24	17,43	22,04	16,61
Ensemble	16,27	14,65	19,03	17,08	18,36	17,81	49,76	19,61
BPM e PR	24,91	15,96	25,65	25,06	21,46	19,99	14,30	24,04

Fonte: Própria do autor.

Tabela 14 – MAPE (%) para as previsões da vazão (m^3/s) correspondentes aos anos 2004 a 2013 para todo o ano realizadas através dos *ensembles*: MA, MP, MQ, MG e MIB.

Água Vermelha					
Ensemble	MA	MP	MQ	MG	MIB
Todas as componentes	18,99	17,93	18,06	65,65	28,87
FR e PR	18,1	18,12	18,15	18,05	17,92
Itutinga					
Ensemble	MA	MP	MQ	MG	MIB
Todas as componentes	18,47	18,40	17,60	53,23	27,35
BPM e PR	16,86	16,98	16,89	16,84	18,28

Fonte: Própria do autor.