

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
**DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**JÔNATAS TRABUCO BELOTTI**

**PREVISÃO DE VAZÕES AFLUENTES UTILIZANDO REDES  
NEURAS ARTIFICIAIS E ENSEMBLES**

**DISSERTAÇÃO**

**PONTA GROSSA**  
**2019**

**JÔNATAS TRABUCO BELOTTI**

**PREVISÃO DE VAZÕES AFLUENTES UTILIZANDO REDES  
NEURAS ARTIFICIAIS E ENSEMBLES**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação ao Programa de Pós-Graduação em Ciência da Computação da Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa.

Orientador: Hugo Valadares Siqueira

Coorientador: Fábio Luiz Usberti

**PONTA GROSSA**

**2019**

Ficha catalográfica elaborada pelo Departamento de Biblioteca  
da Universidade Tecnológica Federal do Paraná, Campus Ponta Grossa  
n.33/19

B452 Belotti, Jônatas Trabuco

Previsão de vazões afluentes utilizando redes neurais artificiais e  
ensembles. / Jônatas Trabuco Belotti, 2018.

137 f. : il. ; 30 cm.

Orientador: Prof. Dr. Hugo Valadares Siqueira

Coorientador: Fábio Luiz Usberti

Dissertação (Mestrado em Ciência da Computação) - Programa de Pós-  
Graduação em Ciência da Computação. Universidade Tecnológica Federal do  
Paraná, Ponta Grossa, 2019.

1. Análise de séries temporais. 2. Previsão hidrológica. 3. Redes neurais  
(Computação). I. Siqueira, Hugo Valadares. II. Usberti, Fábio Luiz. III.  
Universidade Tecnológica Federal do Paraná. IV. Título.

CDD 004

Eelson Heraldo Ribeiro Junior. CRB-9/1413. 11/04/2019.



**Ministério da Educação**  
**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
**CÂMPUS PONTA GROSSA**  
Diretoria de Pesquisa e Pós-Graduação  
Programa de Pós-Graduação em Ciência da



## FOLHA DE APROVAÇÃO

Título de Dissertação Nº 10/2019

### PREVISÃO DE VAZÕES AFLUENTES UTILIZANDO REDES NEURAIS ARTIFICIAIS E ENSEMBLES

Por

Jônatas Trabuco Belotti

Esta dissertação foi apresentada às **08 horas e 30 minutos** do dia **15 de Fevereiro de 2019**, na sala de **videoconferência da DIRPPG**, como requisito parcial para a obtenção do título de MESTRE EM CIÊNCIA DA COMPUTAÇÃO, do Programa de Pós-Graduação em Ciência da Computação. O candidato foi arguido pela Banca Examinadora, composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho APROVADO.

---

**Prof. Dr. Paulo Salgado Gomes de Mattos Neto (UFPE)**

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Sheila Morais de Almeida (UTFPR)**

---

**Prof. Dr. Sergio Luiz Stevan Junior (UTFPR)**

---

**Prof. Dr. Hugo Valadares Siqueira (UTFPR)**  
*Orientador e presidente da banca*



Visto da Coordenadora:

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Sheila Morais de Almeida**  
Coordenadora do PPGCC  
UTFPR – Câmpus Ponta Grossa

A FOLHA DE APROVAÇÃO ASSINADA ENCONTRA-SE ARQUIVADA NA SECRETARIA DO CURSO.

## AGRADECIMENTOS

Primeiramente agradeço a Deus pela vida.

A minha família por todo o apoio e incentivo nas horas de dúvida: minha mãe Maria Lúcia, meu pai Jonatas Belotti e meu irmão Lucas. A minha noiva Letícia que sempre esteve do meu lado, me ajudando nas dificuldades e incentivando a seguir em frente.

Agradeço a todos os professores que durante esses anos repassaram seu conhecimento, aconselharam e principalmente tiveram paciência com minhas dificuldades. Em especial ao meu orientador Prof. Dr. Hugo Valadares Siqueira pela confiança que sempre depositou no meu trabalho, por toda a dedicação e horas de esforço empregadas para a realização dessa dissertação, muito mais que um orientador, um amigo que levo para a vida. Também um agradecimento especial ao meu coorientador Prof. Dr. Fábio Luiz Usberti.

Muito obrigado a todos os amigos que dividiram comigo momentos de felicidade. Aos companheiros do LICON, Biuk, Carlos, Fábio, Henrique, Lilian, Marco, Márcio, Solak, Theodor e Wallace serei eternamente grato pelo aprendizado diário, pelo trabalho em conjunto, pela amizade, pelas risadas, e pelo café.

Aos excelentíssimos moradores do Ap 13, Leandro, Luan e João muito obrigado por proporcionarem uma casa longe de casa, pela amizade e companheirismo sempre presentes.

A todos os membros da IPI Central de Ponta Grossa, não vou citar nomes aqui pois certamente cometeria alguma injustiça, eu simplesmente não tenho palavras para descrever tudo o que vocês fizeram por mim, levarei vocês em meu coração por toda a vida.

Agradeço a Universidade Tecnológica Federal do Paraná (UTFPR) pela bolsa concedida para a realização deste trabalho.

## RESUMO

BELOTTI, Jônatas Trabuco. **Previsão de vazões afluentes utilizando redes neurais artificiais e ensembles**. 2019. 137 f. Dissertação de mestrado em Ciência da Computação, Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2019.

A matriz energética brasileira é predominantemente composta por usinas hidrelétricas. Por este motivo, é importante garantir máxima eficiência no funcionamento dessas plantas, já que a consequência direta é um impacto significativo no custo da produção e na precificação de energia. Determinar a vazão afluente a uma usina hidrelétrica é uma etapa de fundamental importância na eficiência desta operação. Durante os anos, vários modelos lineares, como o Autorregressivo, e não lineares, como Redes Neurais Artificiais, têm sido utilizados na previsão de vazões afluentes. Com o objetivo de aperfeiçoar as técnicas existentes, este trabalho realizou a previsão de vazões mensais através da utilização de 2 modelos lineares: Autoregressivo e Autoregressivo de Médias Móveis; 10 arquiteturas de Redes Neurais Artificiais: MLP, RBF, ELM, ELM (CR), Elman, Jordan, ESN Jaeger, ESN Jaeger (CR), ESN Ozturk e ESN Ozturk (CR); e 6 *Ensembles*: combinadores de Média, Mediana, MLP, RBF, ELM e ELM (CR). O termo CR é relativo à presença do coeficiente de regularização. Os testes foram realizados nas séries históricas das usinas de Água Vermelha, Belo Monte, Ilha Solteira, Paulo Afonso e Tucuruí com horizontes de previsão de 1, 3, 6 e 12 passos a frente. Além disso, as entradas utilizadas pelos modelos neurais foram selecionadas através do método *Wrapper*. Ainda, foram propostas e testadas 3 estratégias de previsão fazendo uso de dados dos eventos climáticos El Niño e La Niña, tendo duas delas resultado em melhoras significativas nas previsões. Ao final verificou-se que o desempenho dos modelos neurais foi melhor que dos modelos lineares em todas as simulações, provando a superioridade dos preditores não lineares. Destaca-se a Rede Neural ELM como o melhor preditor.

**Palavras-chaves:** Análise de séries temporais. Previsão hidrológica. Redes neurais (Computação).

## ABSTRACT

BELOTTI, Jônatas Trabuco. **Prediction of affluent flows using artificial neural networks and ensembles**. 2019. 137 p. Master's degree thesis in Computer Science, Federal Technology University - Paraná. Ponta Grossa, 2019.

The Brazilian energy matrix is predominantly composed of hydroelectric plants. In this way, it is important to ensure maximum efficiency in the operation of these plants since the direct consequence is a significant impact on the cost of energy production and pricing. Determining the streamflow to a hydroelectric plant is a fundamental step in the efficiency of the operation. Over the years, several linear models, such as Autorregressive, and nonlinear, as Artificial Neural Networks have been used to predict streamflows. In order to improve the existing forecasting techniques, this work accomplished the forecast of monthly streamflows through the use of 2 linear models: Autoregressive and Autoregressive of Moving Averages; 10 Architectures of Artificial Neural Networks: MLP, RBF, ELM, ELM (CR), Elman, Jordan, ESN Jaeger, ESN Jaeger (CR), ESN Ozturk and ESN Ozturk (CR); and 6 Ensembles: Medium, Median, MLP, RBF, ELM and ELM (CR) combiners. The term CR is related to the presence of the regularization coefficient. The tests were carried out of the historical séries of the plants of Água Vermelha, Belo Monte, Ilha Solteira, Paulo Afonso and Tucuruí with forecasts horizons of 1, 3, 6 and 12 steps ahead. In addition, the inputs used by the neural models were selected using the Wrapper method. Also, we proposed and tested 3 forecasting strategies using data from the El Niño and La Niña climatic events, two of which resulted in significant improvements in the performances. We verified that the performance of the neural models were better than the linear models in all the simulations, proving the superiority of the nonlinear predictors. We highlight the ELM as the best predictor.

**Key-words:** Time-series analysis. Hidrological forecasting. Neural networks (Computer science).

## LISTA DE FIGURAS

Figura 1	– Modelo de neurônio artificial .....	28
Figura 2	– Identificação das partes de uma rede neural.....	33
Figura 3	– Arquitetura <i>feedforward</i> .....	34
Figura 4	– Arquitetura Recorrente.....	35
Figura 5	– Modelo de rede <i>Perceptron</i> de Múltiplas Camadas (MLP) .....	39
Figura 6	– Rede de função de base radial (RBF).....	41
Figura 7	– Rede neural de Elman .....	43
Figura 8	– Rede neural de Jordan.....	44
Figura 9	– Máquina de aprendizado extremo (ELM) .....	47
Figura 10	– Rede de estado de <i>echo</i> (ESN) .....	50
Figura 11	– Modelo genérico de um <i>Ensemble</i> .....	53
Figura 12	– Modelo <i>Ensemble</i> Estático.....	54
Figura 13	– Modelo <i>Ensemble</i> Dinâmico .....	54
Figura 14	– Modelo de classificação de padrões com Redes neurais.....	55
Figura 15	– Modelo previsão série temporal com Rede neural.....	56
Figura 16	– Localização das usinas hidrelétricas no mapa brasileiro.....	59
Figura 17	– Etapas de pré e pós-processamento.....	62
Figura 18	– Áreas de influência do El Niño e La Niña no Brasil.....	119
Figura 19	– Entradas do modelo de previsão com El Niño e La Niña.....	120



## LISTA DE GRÁFICOS

Gráfico 1	– Série de vazões médias mensais da usina de Furnas.....	19
Gráfico 2	– Média e variância mensal para a série da usina de Furnas. ....	20
Gráfico 3	– Série dessazonalizada da usina de Furnas.....	21
Gráfico 4	– Valores de autocorrelação para a série de Furnas dessazonalizada. ....	22
Gráfico 5	– Função de ativação degrau.....	30
Gráfico 6	– Função de ativação degrau bipolar.....	30
Gráfico 7	– Função de ativação rampa simétrica .....	31
Gráfico 8	– Função de ativação logística. ....	32
Gráfico 9	– Função de ativação Tangente Hiperbólica.....	32
Gráfico 10	– Função de ativação gaussiana. ....	42
Gráfico 11	– Vazões médias das usinas hidrelétricas. ....	60
Gráfico 12	– Série de vazões médias mensais da usina de Água Vermelha. ....	65
Gráfico 13	– Série dessazonalizada da usina de Água Vermelha.....	65
Gráfico 14	– Melhor previsão Água Vermelha de 2006 a 2015 – 1 passo a frente. ....	66
Gráfico 15	– Melhor previsão Água Vermelha de 2006 a 2015 – 3 passos a frente. ....	68
Gráfico 16	– Melhor previsão Água Vermelha de 2006 a 2015 – 6 passos a frente. ....	69
Gráfico 17	– Melhor previsão Água Vermelha de 2006 a 2015 – 12 passos a frente.....	70
Gráfico 18	– Evolução MSE para a usina de Água Vermelha. ....	71
Gráfico 19	– Dispersão dos resultados para a usina de Água Vermelha. ....	72
Gráfico 20	– Padrão das previsões para Água Vermelha com $P = 12$ . ....	73
Gráfico 21	– Série de vazões médias mensais da usina de Belo Monte.....	74
Gráfico 22	– Série dessazonalizada da usina de Belo Monte. ....	74
Gráfico 23	– Melhor previsão Belo Monte de 2006 a 2015 – 1 passo a frente. ....	76
Gráfico 24	– Melhor previsão Belo Monte de 2006 a 2015 – 3 passos a frente. ....	77
Gráfico 25	– Melhor previsão Belo Monte de 2006 a 2015 – 6 passos a frente. ....	78
Gráfico 26	– Melhor previsão Belo Monte de 2006 a 2015 – 12 passos a frente.....	80
Gráfico 27	– Evolução MSE para a usina de Belo Monte. ....	81
Gráfico 28	– Padrão das previsões para Belo Monte com $P = 12$ . ....	81
Gráfico 29	– Dispersão dos resultados para a usina de Belo Monte.....	82
Gráfico 30	– Série de vazões médias mensais da usina de Ilha Solteira. ....	83
Gráfico 31	– Série dessazonalizada da usina de Ilha Solteira. ....	83
Gráfico 32	– Melhor previsão Ilha Solteira de 2006 a 2015 – 1 passo a frente. ....	85
Gráfico 33	– Melhor previsão Ilha Solteira de 2006 a 2015 – 3 passos a frente.....	86
Gráfico 34	– Melhor previsão Ilha Solteira de 2006 a 2015 – 6 passos a frente.....	87
Gráfico 35	– Melhor previsão Ilha Solteira de 2006 a 2015 – 12 passos a frente. ....	88
Gráfico 36	– Dispersão dos resultados para a usina de Ilha Solteira. ....	89
Gráfico 37	– Padrão das previsões para Ilha Solteira com $P = 12$ .....	89
Gráfico 38	– Evolução MSE para a usina de Ilha Solteira. ....	90
Gráfico 39	– Série de vazões médias mensais do Complexo Hidrelétrico de Paulo Afonso.....	91
Gráfico 40	– Série dessazonalizada do Complexo Hidrelétrico de Paulo Afonso. ....	91
Gráfico 41	– Melhor previsão Paulo Afonso de 2006 a 2015 – 1 passo a frente.....	93
Gráfico 42	– Melhor previsão Paulo Afonso de 2006 a 2015 – 3 passos a frente. ....	94
Gráfico 43	– Melhor previsão Paulo Afonso de 2006 a 2015 – 6 passos a frente. ....	95
Gráfico 44	– Melhor previsão Paulo Afonso de 2006 a 2015 – 12 passos a frente. ....	96
Gráfico 45	– Evolução MSE para a usina de Paulo Afonso. ....	97

Gráfico 46	– Padrão das previsões para Paulo Afonso com $P = 12$ .	97
Gráfico 47	– Dispersão dos resultados para a usina de Paulo Afonso.	98
Gráfico 48	– Série de vazões médias mensais da usina de Tucuruí	99
Gráfico 49	– Série dessazonalizada da usina de Tucuruí	99
Gráfico 50	– Melhor previsão Tucuruí de 2006 a 2015 – 1 passo a frente.	100
Gráfico 51	– Melhor previsão Tucuruí de 2006 a 2015 – 3 passos a frente.	101
Gráfico 52	– Melhor previsão Tucuruí de 2006 a 2015 – 6 passos a frente.	103
Gráfico 53	– Melhor previsão Tucuruí de 2006 a 2015 – 12 passos a frente.	104
Gráfico 54	– Evolução MSE para a usina de Tucuruí.	105
Gráfico 55	– Padrão das previsões para Tucuruí com $P = 12$ .	105
Gráfico 56	– Dispersão dos resultados para a usina de Tucuruí.	106
Gráfico 57	– Porcentagens de melhor resultado por modelo de previsão.	109
Gráfico 58	– Percentual de uso por quantidade de neurônios.	112
Gráfico 59	– Percentual de neurônios menor que 50.	113
Gráfico 60	– Percentual de neurônios menor que 60.	113
Gráfico 61	– Percentual de neurônios menor que 70.	113
Gráfico 62	– Percentual de neurônios menor que 80.	114
Gráfico 63	– Percentual de neurônios menor que 90.	114
Gráfico 64	– Percentual do número de atrasos selecionado.	116
Gráfico 65	– Uso de cada atraso individual.	117
Gráfico 66	– Melhor previsão Tucuruí com intensidade do El Niño de 2006 a 2015 – 1 passo a frente.	122
Gráfico 67	– Melhor previsão Tucuruí com valor do El Niño de 2006 a 2015 – 1 passo a frente.	124
Gráfico 68	– Identificação dos erros nos picos da série Tucuruí de 2006 a 2015 – 1 passo a frente.	125
Gráfico 69	– Previsão com a ponderação do El Niño Tucuruí de 2006 a 2015 – 1 passo a frente.	127

## LISTA DE TABELAS

Tabela 1	– Resultados previsões Água Vermelha de 2006 a 2015 para P = 1. ....	66
Tabela 2	– Resultados previsões Água Vermelha de 2006 a 2015 para P = 3. ....	67
Tabela 3	– Resultados previsões Água Vermelha de 2006 a 2015 para P = 6. ....	68
Tabela 4	– Resultados previsões Água Vermelha de 2006 a 2015 para P = 12. ....	70
Tabela 5	– Evolução MSE Água Vermelha. ....	71
Tabela 6	– Resultados previsões Belo Monte de 2006 a 2015 para P = 1. ....	75
Tabela 7	– Resultados previsões Belo Monte de 2006 a 2015 para P = 3. ....	76
Tabela 8	– Resultados previsões Belo Monte de 2006 a 2015 para P = 6. ....	78
Tabela 9	– Resultados previsões Belo Monte de 2006 a 2015 para P = 12. ....	79
Tabela 10	– Evolução MSE Belo Monte. ....	80
Tabela 11	– Resultados previsões Ilha Solteira de 2006 a 2015 para P = 1. ....	84
Tabela 12	– Resultados previsões Ilha Solteira de 2006 a 2015 para P = 3. ....	85
Tabela 13	– Resultados previsões Ilha Solteira de 2006 a 2015 para P = 6. ....	86
Tabela 14	– Resultados previsões Ilha Solteira de 2006 a 2015 para P = 12. ....	88
Tabela 15	– Evolução MSE Ilha Solteira. ....	90
Tabela 16	– Resultados previsões Paulo Afonso de 2006 a 2015 para P = 1. ....	92
Tabela 17	– Resultados previsões Paulo Afonso de 2006 a 2015 para P = 3. ....	93
Tabela 18	– Resultados previsões Paulo Afonso de 2006 a 2015 para P = 6. ....	94
Tabela 19	– Resultados previsões Paulo Afonso de 2006 a 2015 para P = 12. ....	95
Tabela 20	– Evolução MSE Paulo Afonso. ....	96
Tabela 21	– Resultados previsões Tucuruí de 2006 a 2015 para P = 1. ....	100
Tabela 22	– Resultados previsões Tucuruí de 2006 a 2015 para P = 3. ....	101
Tabela 23	– Resultados previsões Tucuruí de 2006 a 2015 para P = 6. ....	102
Tabela 24	– Resultados previsões Tucuruí de 2006 a 2015 para P = 12. ....	103
Tabela 25	– Evolução MSE Tucuruí. ....	104
Tabela 26	– Melhor modelo por horizonte. ....	107
Tabela 27	– Resultado teste de Friedman. ....	110
Tabela 28	– Quantidade de neurônios selecionada por modelo neural e usina. ....	111
Tabela 29	– Atrasos selecionados por rede neural e usina. ....	115
Tabela 30	– Resultados Tucuruí com intensidade do El Niño de 2006 a 2015, P = 1. ...	121
Tabela 31	– Resultados Tucuruí com valor do El Niño de 2006 a 2015, P = 1. ....	123
Tabela 32	– Resultados Tucuruí com ponderação do El Niño de 2006 a 2015, P = 1. ...	126

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>12</b>
1.1 OBJETIVOS .....	13
1.1.1 Objetivo Geral .....	13
1.1.2 Objetivos Específicos.....	13
1.2 JUSTIFICATIVA .....	14
<b>2 ANÁLISE DE SÉRIES TEMPORAIS</b> .....	<b>15</b>
2.1 SÉRIES TEMPORAIS: CONCEITUAÇÃO .....	15
2.2 PROCESSOS ESTOCÁSTICOS .....	16
2.3 FERRAMENTAS DE ANÁLISE DE SÉRIES TEMPORAIS .....	16
2.3.1 Média e Variância .....	17
2.3.2 Autocovariância e Autocorrelação .....	17
2.4 SÉRIES DE VAZÕES MÉDIAS MENSAIS .....	18
2.4.1 Dessazonalização .....	20
2.5 ANÁLISE DA FUNÇÃO DE AUTOCORRELAÇÃO DESSAZONALIZADA .....	21
<b>3 MODELOS LINEARES DE PREVISÃO</b> .....	<b>23</b>
3.1 MODELO AUTO-REGRESSIVO (AR) .....	24
3.2 MODELO AUTO-REGRESSIVO DE MÉDIAS MÓVEIS (ARMA) .....	26
<b>4 REDES NEURAIS ARTIFICIAIS</b> .....	<b>27</b>
4.1 O NEURÔNIO ARTIFICIAL .....	28
4.1.1 Tipos de Função de Ativação .....	29
4.2 ARQUITETURA DE REDES NEURAIS .....	33
4.2.1 Arquitetura <i>Feedforward</i> .....	34
4.2.2 Arquitetura Recorrente .....	35
4.3 TREINAMENTO DA REDE .....	36
4.3.1 Treinamento Supervisionado.....	37
4.3.2 Validação cruzada .....	38
4.4 <i>PERCEPTRON</i> DE MÚLTIPLAS CAMADAS (MLP) .....	38
4.4.1 Treinamento MLP .....	39
4.5 REDES NEURAIS COM FUNÇÃO DE BASE RADIAL .....	40
4.5.1 Treinamento RBF.....	41
4.6 REDE NEURAL DE ELMAN.....	43
4.7 REDE NEURAL DE JORDAN .....	44
4.8 MÁQUINAS DESORGANIZADAS .....	45
4.8.1 Máquinas de Aprendizado Extremo (ELM) .....	46
4.8.1.1 Treinamento (ELM) .....	48
4.8.2 Redes Neurais Com Estado de Eco (ESN) .....	49
4.8.2.1 Propriedade de estado de eco .....	51
4.8.2.2 Construção do reservatório de dinâmicas .....	52
4.9 METODOLOGIA <i>ENSEMBLE</i> .....	52
4.10 REDES NEURAIS PARA PREVISÃO DE SÉRIES TEMPORAIS .....	55
4.10.1 Método <i>Wrapper</i> Para Seleção de Entradas.....	57
<b>5 ESTUDO DE CASOS (PREVISÃO DE SÉRIES DE VAZÕES)</b> .....	<b>59</b>
5.1 MODELOS DE PREVISÃO ABORDADOS .....	62
5.2 USINA HIDRELÉTRICA DE ÁGUA VERMELHA.....	64
5.3 USINA HIDRELÉTRICA DE BELO MONTE.....	74
5.4 USINA HIDRELÉTRICA DE ILHA SOLTEIRA .....	83

5.5	COMPLEXO HIDRELÉTRICO DE PAULO AFONSO .....	91
5.6	USINA HIDRELÉTRICA DE TUCURUÍ .....	98
5.7	ANÁLISE DOS RESULTADOS .....	107
5.8	TESTE DE FRIEDMAN .....	109
5.9	QUANTIDADE DE NEURÔNIOS SELECIONADA .....	110
5.10	ATRASOS SELECIONADOS .....	114
5.11	TENTATIVAS DE MELHORIA DOS RESULTADOS .....	117
5.11.1	Utilização das Intensidades do El Niño e La Niña .....	121
5.11.2	Utilização dos Valores Reais do El Niño e La Niña .....	123
5.11.3	Aproximação dos Picos .....	124
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>128</b>
6.1	TRABALHOS FUTUROS .....	129
	<b>REFERÊNCIAS .....</b>	<b>135</b>
	<b>APÊNDICE A - ARTIGOS PUBLICADOS EM ANAIS DE CONGRESSOS .....</b>	<b>137</b>

## 1 INTRODUÇÃO

No ano de 2017, 65,2% de toda energia elétrica produzida no Brasil foi resultado da geração hidráulica (EPE, 2018). Se comparado com o ano anterior, esse número teve uma redução percentual de 2,9%, o que não alterou o cenário da matriz energética brasileira, a qual continua sendo predominantemente composta por usinas hidrelétricas (EPE, 2018).

Em SOARES FILHO (1987) o planejamento da operação de um sistema de energia elétrica é definido com o objetivo de utilizar as fontes geradoras de energia da forma mais eficiente possível, diminuindo ao máximo o custo da geração. As fontes geradoras com menor custo de operação são as renováveis, pois seus combustíveis (água, luz do sol e vento) são oferecidos de forma gratuita pela natureza. Portanto, para diminuir o custo da produção, tais fontes devem ser utilizadas com máxima eficiência.

Como as usinas hidrelétricas possuem a maior representatividade dentro da matriz energética brasileira (EPE, 2018), garantir o seu funcionamento de forma eficiente apresentaria um impacto significativo no custo da produção de energia elétrica. Para tal, são necessárias informações precisas a respeito das suas vazões afluentes (BALLINI, 2000).

A previsão de vazões é um assunto comumente abordado no meio científico, sendo que várias metodologias lineares já foram propostas para o problema, como os trabalhos de Rasmussen et al. (1996), Siqueira (2009), TEIXEIRA JÚNIOR et al. (2016) e Malfatti, Cardoso e Hamburger (2018). Abordagens não lineares também já foram estudadas, como os trabalhos de Francelin, Ballini e Andrade (1996), Mason, Price e Tem'Me (1996), Zealand, Burn e Simonovic (1999), Siqueira (2013) e Dias, Cataldi e Ferreira (2017). Dentre estas, destaca-se com os melhores resultados a utilização de Redes Neurais Artificiais (RNA).

O modelo linear Autorregressivo (AR), da família Box & Jenkins é o utilizado na previsão de séries de vazões hidrológicas afluentes no planejamento da operação energética no Brasil pelo Sistema *Newave* (CEPEL, 2013). Por outro lado, metodologias que fazem uso de RNAs têm sido abordadas como alternativas viáveis para esta tarefa, tendo estas apresentado bons resultados como em Tang e Fishwick (1993), Srinivasan, Liew e Chang (1994), Dias, Cataldi e Ferreira (2017) e Oliveira et al. (2018).

Redes Neurais Artificiais (RNAs) foram elaboradas com inspiração no sistema nervoso central dos organismos superiores. São capazes de reconhecer regularidades e padrões de dados através do treinamento (aprendizado) e realizar generalizações baseadas no conhecimento adquirido (HAYKIN, 2001). Devido a isso, as RNAs têm sido muito utilizadas na previsão de séries temporais, inclusive na previsão de séries de vazões médias mensais à usinas hidrelétricas (SIQUEIRA, 2009).

Várias arquiteturas de RNAs já foram utilizadas para a previsão de vazões, como o *Perceptron* de múltiplas camadas (MLP) (FRANCELIN; BALLINI; ANDRADE, 1996), a Rede Neural com função base radial (RBF) (MASON; PRICE; TEM'ME, 1996), Rede neuro-*fuzzy*

(BALLINI, 2000), Rede neural com estados de eco (ESN) (SACCHI et al., 2007) e as Máquinas de aprendizado extremo (ELM) (SIQUEIRA, 2013).

Sharkey (1999) mostrou que para a previsão de séries temporais, os melhores resultados são alcançados através da combinação de diferentes modelos de previsão ao invés da seleção daquele que apresenta melhor desempenho individual. Existem diversas formas para combinar preditores, dentre elas, a metodologia *Ensemble*. Assim, para o problema proposto, tal combinação merece ser investigada.

Na literatura encontram-se aplicações de *Ensembles* ao problema da previsão de vazões. Silva, Tucci e Collischonn (2006) utilizaram *Ensembles* como parte de um modelo de previsão de vazões hidroclimático; Kasiviswanathan e Sudheer (2013) compararam os resultados obtidos por *Ensembles* de médias na previsão da vazão do rio Watershed (Índia); por sua vez, Fan et al. (2016) aplicaram *Ensembles* determinísticos para prever a vazão do reservatório de Três Marias no Brasil e Thober et al. (2018) realizaram previsões de vazões em vários rios da Europa utilizando tal método. Entretanto, os pesquisadores da área comumente exploram apenas *Ensembles* de média ou a utilização de Redes Neurais clássicas (MLP e RBF) como combinador, fazendo-se importante a exploração de outras arquiteturas neurais.

Dessa maneira, este trabalho realizou a previsão de séries de vazões médias mensais de importantes usinas hidrelétricas brasileiras através da utilização de Redes Neurais Artificiais, Modelos Lineares e da Metodologia *Ensemble*.

## 1.1 OBJETIVOS

As Seções 1.1.1 e 1.1.2 descrevem respectivamente, qual o objetivo desse trabalho e quais etapas foram realizadas para alcançá-lo.

### 1.1.1 Objetivo Geral

O objetivo principal desse trabalho foi o aperfeiçoamento das técnicas de previsão mensal de séries temporais, mais especificamente de séries de vazões médias mensais de importantes usinas hidrelétricas brasileiras, utilizando Redes Neurais Artificiais (RNAs).

### 1.1.2 Objetivos Específicos

Para que o objetivo geral fosse alcançado, os seguintes objetivos específicos foram cumpridos:

- Aplicação de métodos estatísticos para realizar o pré-processamento das séries das vazões médias mensais;
- Utilização dos modelos lineares Autoregressivo (AR) e Autoregressivo de Médias Móveis (ARMA) na previsão de vazões médias mensais;
- Aplicação do método *Wrapper* na seleção das variáveis utilizadas como entrada para as previsões realizadas;
- Realização da previsão com as redes neurais *feedforward*: MLP, RBF e ELM;
- Aplicação das seguintes redes neurais recorrentes: Elman, Jordan e ESN;
- Previsão através da utilização de metodologias *Ensemble* com combinadores de redes neurais clássicas;
- Utilização da Máquina de Aprendizado Extremo como combinador do *Ensemble*;
- Comparação dos resultados obtidos por todos os modelos testados;
- Comparação do número de neurônios selecionado para cada rede neural por usina;
- Comparação das entradas selecionadas para cada rede neural por usina.

## 1.2 JUSTIFICATIVA

Como mencionado, a matriz energética brasileira é composta predominantemente por usinas hidrelétricas, de modo que a operação eficiente destas tem impacto direto no custo da produção e na precificação da energia elétrica. Para isso, é necessário conceber informações precisas a respeito das vazões afluentes futuras.

A previsão de séries de vazões por meio da utilização de Redes Neurais Artificiais (RNAs) vem se mostrando uma alternativa viável aos modelos lineares tradicionais. Portanto, o aperfeiçoamento das metodologias para previsão tem impacto direto na diminuição do custo da geração de energia elétrica brasileira.

Além disso, como já foi dito, a literatura carece de estudos que explorem a previsão de vazões afluentes através da utilização de *Ensembles* com combinadores neurais. Este trabalho verificou a utilização de Máquinas de Aprendizado Extremo para este fim.



## 2 ANÁLISE DE SÉRIES TEMPORAIS

Diversos fenômenos da natureza podem ser modelados por uma série temporal com o objetivo de descrever o seu funcionamento ou realizar previsões para os comportamentos futuros (MORETTIN; TOLOI, 2006).

A principal tarefa ao lidar com séries temporais é definir um modelo capaz de representar o processo envolvido. Para tal, é necessário cumprir algumas etapas como: observação dos dados, pré-processamento, seleção dos atrasos mais significativos e determinação dos coeficientes presentes no modelo (SIQUEIRA, 2013).

Este capítulo apresenta conceitos essenciais acerca de séries temporais desde a sua definição em termos estatísticos, passando pelo comportamento de processos estocásticos até uma discussão sobre séries de vazões e análise da função de autocorrelação.

### 2.1 SÉRIES TEMPORAIS: CONCEITUAÇÃO

Uma série temporal é uma coleção de observações realizadas em sequência ao longo do tempo. Diferente dos modelos de regressão nos quais a ordem das observações não é relevante, em estudos sobre séries temporais a ordem delas tem extrema importância e não pode ser alterada. As amostras vizinhas são dependentes e exercem influência uma na outra. Dessa forma, problemas dessa natureza consistem basicamente em definir e modelar essa dependência entre as observações (EHLERS, 2005).

Box, Jenkins e Reinsel (2008) apresentam uma definição formal de séries temporais como sendo um conjunto de observações  $\mathbf{Z}$  de uma variável aleatória  $x_t \in \mathbf{Z}$ ,  $t \in \mathbf{T} \subset \mathbb{R}$ , geradas sequencialmente no tempo, sendo  $\mathbb{R}$  o conjunto dos números reais e  $\mathbf{T}$  um subconjunto arbitrário.

Dentre as principais razões para o estudo e a modelagem de séries temporais está a previsão de valores futuros, possibilitando a geração de sequências sintéticas probabilisticamente aceitáveis (PEREIRA; PAIS; SALES, 1986).

Como exemplos de séries temporais pode-se citar: a produção anual de aço no Brasil, as médias de temperaturas diárias de uma cidade, o valor diário de uma determinada ação na Bolsa de Valores e a quantidade mensal de energia elétrica produzida por uma usina hidrelétrica (SPIEGEL, 1993).

## 2.2 PROCESSOS ESTOCÁSTICOS

Morettin e Toloi (2006) definem um processo estocástico como uma família  $\mathbf{X} = \{x_t, t \in \mathbf{T}\}$ , tal que, para cada  $t \in \mathbf{T}$ ,  $x_t$  é uma variável aleatória (v. a.). Assumindo que  $\mathbf{T} \subset \mathbb{Z}_+$ , sendo este o conjunto dos números inteiros positivos, as observações que caracterizam uma série temporal podem ser descritas por uma v. a.  $\{x_t\}$ ,  $t \in \mathbf{T}$ , com função de probabilidade conjunta  $p(x_1, x_2, \dots, x_N)$  (BOX; JENKINS; REINSEL, 2008).

Processos estocásticos possuem um comportamento estatístico associado ao seu desenvolvimento, sendo possível classificá-los em função disto (EHLERS, 2005). Um processo estocástico  $\mathbf{X} = \{x_t, t \in \mathbf{T}\}$  é dito estritamente estacionário (ou fortemente estacionário) se todas as distribuições de probabilidade  $F(\cdot)$  dos elementos de  $\mathbf{X}$  permanecem as mesmas sob translações  $\tau$  no tempo, ou seja

$$F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) = F(x_1, \dots, x_n; t_1, \dots, t_n)$$

para quaisquer  $t_1, t_2, \dots, t_n, \tau \in \mathbf{T}$  (MORETTIN; TOLOI, 2006).

Em outras palavras, uma série temporal é dita estritamente estacionária, quando a origem dos tempos é deslocada  $\tau$  unidades e a distribuição de probabilidade continua a mesma, visto que ela depende apenas dos intervalos entre  $t_k$  e  $t_{k+1}$  (EHLERS, 2005).

Um processo estocástico  $\mathbf{X} = \{x_t, t \in \mathbf{T}\}$  é dito estacionário de segunda ordem (ou fracamente estacionário) se e somente se:

- A média é constante:  $E[x_t] = \mu_t = \mu$ , para todo  $t \in \mathbf{T}$ ;
- A variância é constante:  $Var[x_t] = E[(x_t - \mu)^2] = \sigma_t^2 = \sigma^2$ , para todo  $t \in \mathbf{T}$ ;
- A covariância  $Cov[x_t, x_{t+k}] = E[(x_t - \mu)(x_{t+k} - \mu)] = \gamma_k$  é uma função exclusiva de  $k$ .

Para tratar toda a dependência estatística entre as amostras é necessário determinar todas as distribuições conjuntas entre as amostras  $x_t$ . Entretanto, na prática isso é muito difícil uma vez que o que se tem é um número limitado de amostras, o que leva à utilização de séries estacionárias sob a égide da segunda ordem (MAGALHÃES, 2004).

## 2.3 FERRAMENTAS DE ANÁLISE DE SÉRIES TEMPORAIS

As distribuições de probabilidade de um processo estocástico possibilitam a derivação de algumas ferramentas estatísticas, as quais auxiliam na identificação e modelagem do processo. Nessa seção serão abordadas as seguintes ferramentas: média, variância, autocovariância e autocorrelação.

### 2.3.1 Média e Variância

Como a distribuição de probabilidade de um processo estocástico  $p(x_t)$  é a mesma para todo  $t \in \mathbf{T}$ , denota-se a distribuição de probabilidade de um processo estocástico como  $p(x)$ . Segundo Box, Jenkins e Reinsel (2008) a métrica que define o nível em torno do qual a série oscila é a média, sendo esta definida pela Equação 2.1.

$$\mu = E[x_t] = \int_{-\infty}^{+\infty} xp(x)dx \quad (2.1)$$

tal que  $E[\cdot]$  é o operador de esperança matemática.

Box, Jenkins e Reinsel (2008) também definem que a amplitude da oscilação da série é dada pela variância, definida pela Equação 2.2.

$$\sigma^2 = E[x_t - \mu]^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \quad (2.2)$$

Analisando as Equações 2.1 e 2.2 verifica-se que as integrais são definidas para uma distribuição de probabilidade constante em séries contínuas. No caso de séries discretas a média e a variância amostrais (ou estimadas) são obtidas pelas equações 2.3 e 2.4 respectivamente

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^N x_t \quad (2.3)$$

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{t=1}^N (x_t - \hat{\mu})^2 \quad (2.4)$$

sendo  $N$  o número total de amostras disponíveis.

### 2.3.2 Autocovariância e Autocorrelação

A autocovariância e a autocorrelação medem a dependência até segunda ordem entre duas observações sequenciais. Como estas são separadas por um intervalo constante  $k$ , elas são denotadas por  $x_t$  e  $x_{t+k}$ . A covariância entre  $x_t$  e  $x_{t+k}$  é chamada de autocovariância de passo  $k$  e definida pela Equação 2.5

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)] \quad (2.5)$$

na qual  $\gamma_k$  é a autocovariância de passo  $k$  e  $\mu$  é dado pela Equação 2.1.

De forma similar a autocorrelação de passo  $k$  é dada pela Equação 2.6.

$$\rho_k = \frac{E[(x_t - \mu)(x_{t+k} - \mu)]}{\sqrt{E[(x_t - \mu)^2]E[(x_{t+k} - \mu)^2]}} = \frac{\gamma_k}{\sigma^2} \quad (2.6)$$

Note que  $\sigma^2 = \gamma_0$ , o que permite que a Equação 2.6 possa ser reescrita, gerando a Equação 2.7.

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (2.7)$$

Como definido na Seção 2.2, processos fracamente estacionários têm variância igual para  $t$  e  $t + k$ , o que implica que  $\rho_0 = 1$  (BOX; JENKINS; REINSEL, 2008).

Para séries temporais estacionárias discretas, a autocorrelação e a autocovariância são calculados para cada  $k$ , podendo ser estimados pelo Método dos Momentos (BOX; JENKINS; REINSEL, 2008), no qual o termo  $\gamma_k$  é calculado como uma variável  $c_k$  de acordo com a Equação 2.8

$$c_k = \frac{1}{N - k} \sum_{t=1}^{N-k} (x_t - \hat{\mu})(x_{t+k} - \hat{\mu}), k = 0, 1, \dots, K \quad (2.8)$$

tal que  $N$  é o número total de amostras disponíveis,  $\hat{\mu}$  é a média estimada dada pela Equação 2.3 e  $K$  é o total de passos considerados.

De forma similar, fazendo uso das Expressões 2.7 e 2.8 o coeficiente de autocorrelação amostral  $r_k$  é dado por 2.9.

$$r_k = \frac{c_k}{c_0} \quad (2.9)$$

O valor de  $r_k$  deve ser definido para todo  $k \in \{0, 1, \dots, K\}$ , onde  $K$  depende do número de observações utilizadas para determinar  $c_k$ . Box, Jenkins e Reinsel (2008) afirmam que quando  $N \cong 100$  deve-se utilizar  $K$  menor que  $\frac{N}{4}$ .

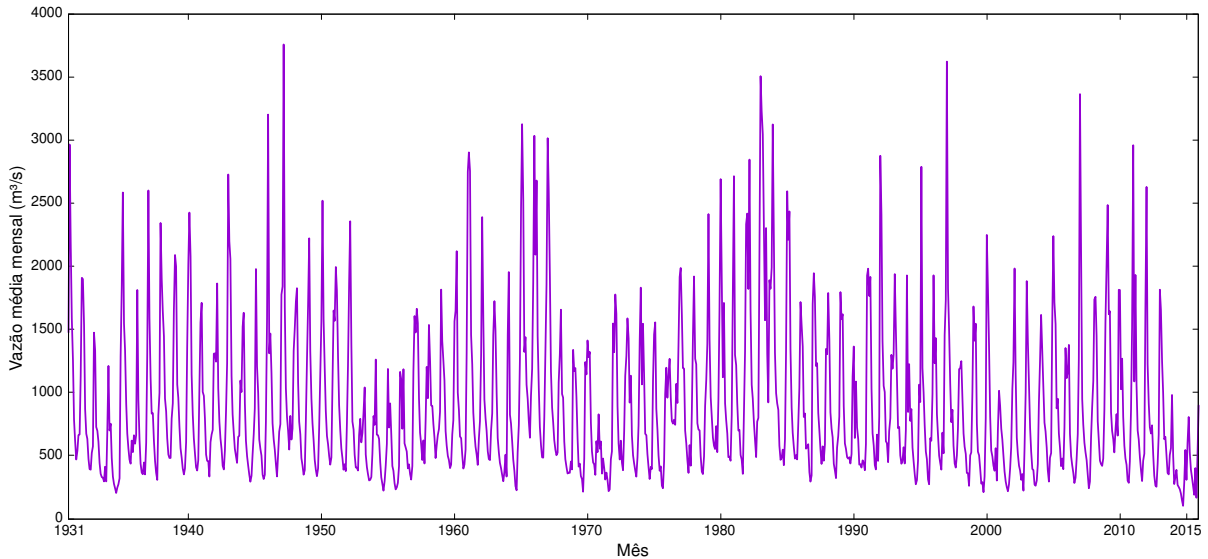
## 2.4 SÉRIES DE VAZÕES MÉDIAS MENSAIS

Séries de vazões são um tipo de série temporal em que cada observação diz respeito a vazão média mensal, semanal, diária ou horária. Utilizaremos neste trabalho as vazões médias mensais, com valores observados no tempo e denotados por  $\{x_1, x_2, \dots, x_N\}$ , sendo  $x_t$  a observação no instante de tempo  $t$ , tal que  $t \in \{1, 2, \dots, N\}$ .

Um exemplo desta é a série histórica da usina hidrelétrica de Furnas, disponibilizada pelo Operador Nacional do Sistema Elétrico (ONS), sendo uma das mais utilizadas em estu-

dos brasileiros. O Gráfico 1 apresenta todas as 1020 observações contidas na série, para a qual calcula-se  $\hat{\mu} = 912,1225 \text{ m}^3/\text{s}$  e variância  $\hat{\sigma} = 613,5036 \text{ m}^3/\text{s}^1$ .

**Gráfico 1 – Série de vazões médias mensais da usina de Furnas.**



**Fonte: Autoria própria.**

Segundo Morettin e Tolo (1987) fenômenos sazonais são aqueles que ocorrem de forma regular em períodos de tempo determinados, por exemplo de mês em mês ou de ano em ano. No caso de séries de vazões mensais a sazonalidade é anual visto que elas respeitam os ciclos de chuvas no decorrer do ano (BALLINI, 2000). Desse modo, tem-se que a vazão  $x_t$  é composta pelo valor estacionário mais a componente sazonal, como pode ser visto na Equação 2.10

$$x_t = s_t + \tilde{x}_t + a_t \quad (2.10)$$

sendo  $s_t$  a componente sazonal da observação  $t$ ,  $\tilde{x}_t$  o valor estacionário e  $a_t$  a componente aleatória de média zero, variância constante, chamada de ruído branco. Na prática  $a_t$  é o erro inerente à previsão.

Como as séries de vazões possuem sazonalidade anual, o entendimento das relações e padrões entre os dados de cada mês individual ajudam na compreensão geral da série. Este entendimento pode ser realizado através da análise da média e variância mensal da série, as quais podem ser obtidas pelas Equações 2.11 e 2.12, respectivamente

$$\hat{\mu}_m = \frac{1}{N} \sum_{i=1}^N x_{i,m} \quad (2.11)$$

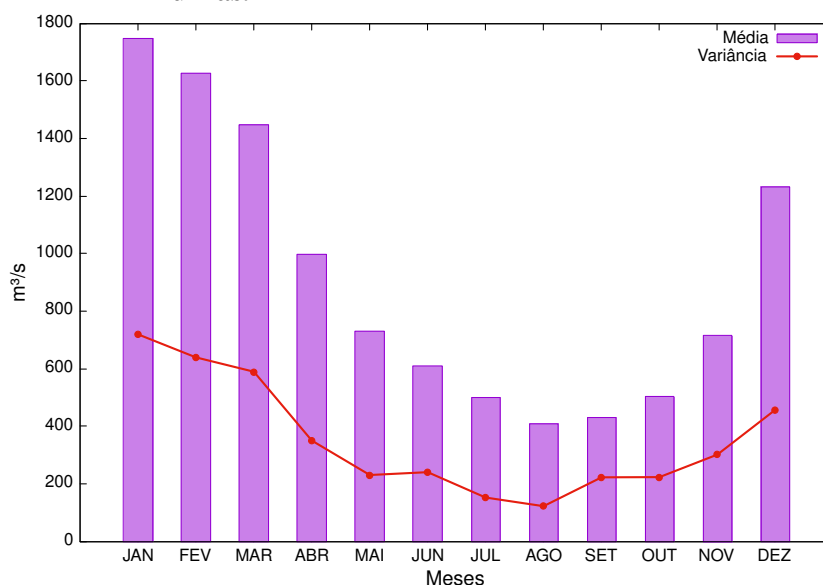
$$\hat{\sigma}_m = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{i,m} - \hat{\mu}_m)^2} \quad (2.12)$$

<sup>1</sup> Note que na Equação 2.4  $\sigma$  está elevado ao quadrado.

de modo que  $\hat{\mu}_m$  é a média do mês  $m$  ( $m = 1$  equivale a janeiro,  $m = 2$  equivale a fevereiro, etc.),  $N$  a quantidade de anos disponíveis nos dados,  $x_{i,m}$  a vazão no mês  $m$  do ano  $i$  e  $\hat{\sigma}_m$  a variância do mês  $m$ .

Para exemplificar, o Gráfico 2 apresenta as médias e as variâncias para a série da usina de Furnas.

**Gráfico 2 – Média e variância mensal para a série da usina de Furnas.**



**Fonte: Autoria própria.**

Como apresentado no Gráfico 2, o regime de chuvas do Brasil varia de acordo com as estações do ano, sendo que janeiro é o mês com a maior densidade pluviométrica e agosto com a menor (SIQUEIRA, 2013), o que comprova que a sazonalidade da série é resultante deste comportamento.

#### 2.4.1 Dessazonalização

A aplicação de modelos lineares exige que a componente sazonal da série seja retirada, tornando a série estacionária. Para tal é usado um processo estatístico conhecido como dessazonalização ou padronização (BALLINI, 2000). Com a remoção da componente sazonal a série fica com média zero e desvio padrão aproximadamente unitário, podendo ser classificada como estacionária no sentido amplo (SIQUEIRA et al., 2012). Nota-se que a adoção do tratamento é benéfica também para sua utilização com Redes Neurais Artificiais, uma vez que a série assume um comportamento padronizado, mais fácil de ser generalizado por uma RNA (SIQUEIRA, 2009).

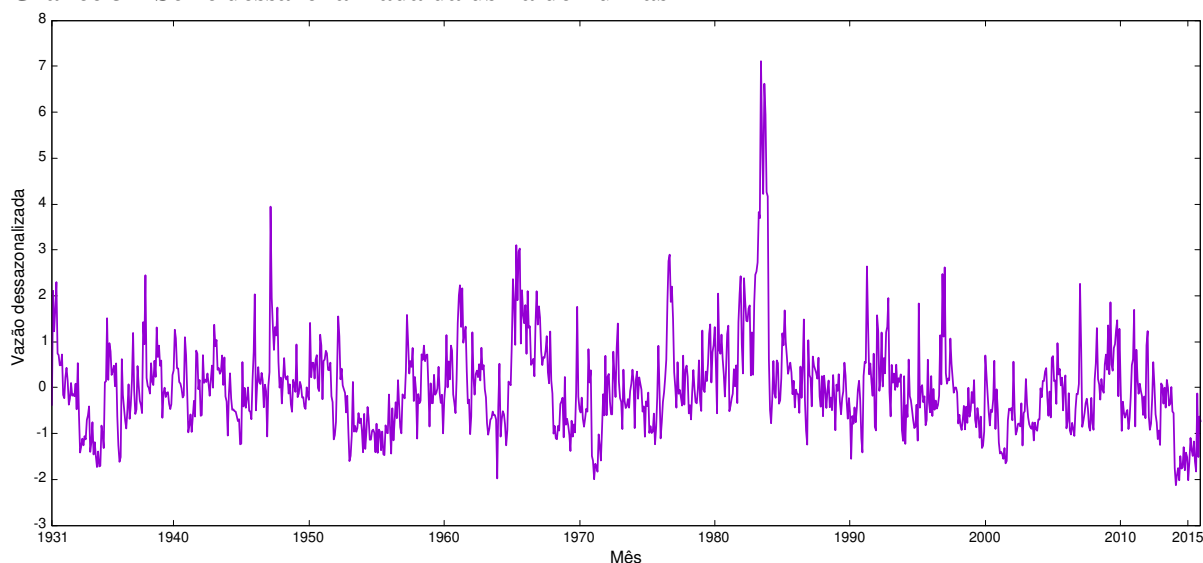
A vazão  $x_{i,m}$  dessazonalizada é denotada por  $x_{i,m}^{DE}$  e definida pela Equação 2.13

$$x_{i,m}^{DE} = \frac{x_{i,m} - \hat{\mu}_m}{\hat{\sigma}_m} \quad (2.13)$$

tal que  $\hat{\mu}_m$  é dado pela Equação 2.11 e  $\hat{\sigma}_m$  pela Equação 2.12.

O Gráfico 3 apresenta a série dessazonalizada da usina de Furnas através do uso da Equação 2.13.

**Gráfico 3 – Série dessazonalizada da usina de Furnas**



**Fonte: Autoria própria.**

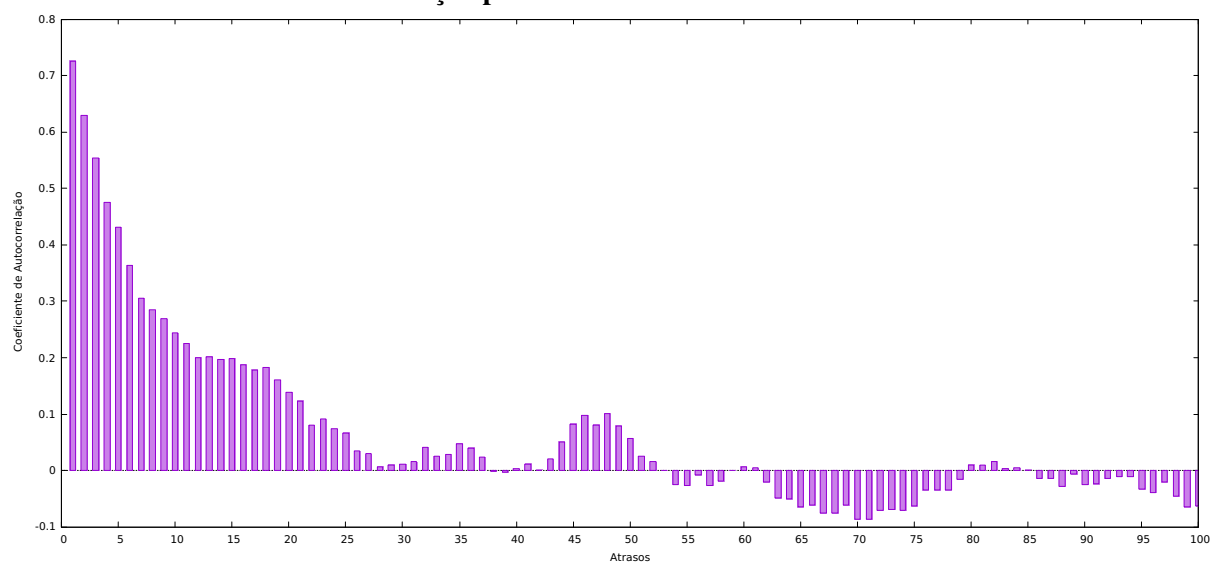
Note que a dessazonalização faz com que os novos valores da série estejam em um intervalo com amplitude menor do que a original. Entretanto, mesmo com a alteração dos valores, pode-se notar semelhanças no padrão de comportamento dos dados. Na série original (Gráfico 1), entre os anos 1960 e 1970 verifica-se um pico, seguido por uma grande queda e depois um pequeno aumento. Esse mesmo comportamento pode ser observado também na série dessazonalizada (Gráfico 3).

## 2.5 ANÁLISE DA FUNÇÃO DE AUTOCORRELAÇÃO DESSAZONALIZADA

A análise da função de autocorrelação permite avaliar o comportamento temporal e estatístico da série, auxiliando no seu entendimento. Aplicando as Equações 2.8 e 2.9 à série dessazonalizada de Furnas e calculando os 100 primeiros valores de  $r_k$ , é possível traçar o Gráfico 4.

Nele nota-se que o valor da autocorrelação diminui a média que  $k$  aumenta, como é esperado para dados com dependência temporal. Também verifica-se que essa variação de  $r_t$  para  $r_{t+k}$  apresenta um comportamento senoidal amortecido, típico de modelos autorregressivos (MORETTIN; TOLOI, 2006).

**Gráfico 4 – Valores de autocorrelação para a série de Furnas dessazonalizada.**



**Fonte: Autoria própria.**



### 3 MODELOS LINEARES DE PREVISÃO

Os modelos lineares de previsão foram os primeiros a serem utilizados na previsão de séries de vazões. Estes passaram a ganhar grande destaque a partir da década de 1960 e rapidamente se tornaram populares devido aos bons resultados obtidos e pela facilidade em sua implementação. Desde então muitos avanços tem sido feitos nessa área, sendo que hoje contamos com diversos modelos lineares, por exemplo: Auto-regressivo (AR), Médias Móveis (MA), Auto-regressivo de Médias Móveis (ARMA) e Auto-regressivo Integrado de Médias Móveis (ARIMA).

A definição dos modelos lineares de previsão por Box e Jenkins faz uso de conceitos acerca de filtros lineares (BOX; JENKINS; REINSEL, 2008). Deste modo, o elemento  $x_t$  de uma série temporal é resultado da aplicação de um filtro linear  $\Psi$  sobre um ruído branco gaussiano  $a_t$ . Formalmente esse pode ser representado pela Equação 3.1

$$x_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \dots + \Psi_n a_{t-n} = \mu + a_t + \sum_{k=1}^n \Psi_k a_{t-k} \quad (3.1)$$

sendo que  $\mu$  é a média da série,  $\Psi$  é o filtro linear e  $n$  é o número de elementos da série.

Podemos então definir um operador linear para realizar a soma ponderada dos termos  $a_t$  pelos respectivos termos  $\Psi$  resultando em  $x_t$ , adicionando ainda o termo  $k = 0$  para o elemento fora do somatório. Esse operador linear é denotado por  $\Psi$  e dado pela Equação 3.2.

$$\Psi = \mu + \sum_{k=0}^{\infty} \Psi_k a_{t-k} \quad (3.2)$$

Desse modo os modelos lineares podem ser considerados como casos particulares de um filtro linear de função de transferência  $\Psi$ , sendo  $x_t$  a saída. Especificamente, o ruído branco gaussiano  $a_t$  é transformado em uma série temporal através da aplicação de um filtro linear, sendo que a série temporal resultante possui correlação entre observações sequenciais. Considera-se  $a_t$  com distribuição normal, média zero e variância constante.

Outra forma de representar um modelo linear é através da ponderação dos sinais anteriores ( $x_{t-1}, x_{t-2}, \dots, x_1$ ) para a previsão do próximo elemento ( $x_t$ ). Para isso, é adicionado um ruído  $a_t$  e a média da série  $\mu$ . Essa segunda forma de representação é a mais comumente utilizada, sendo também a abordada por esse trabalho. A Equação 3.3 define formalmente essa representação

$$x_t = \mu + a_t + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots + \pi_t x_1 \quad (3.3)$$

tal que  $a_t$  é o ruído inerente ao  $t$ -ésimo termo e  $\pi_n$  é o peso atribuído ao  $(t - n)$ -ésimo termo da série.

### 3.1 MODELO AUTO-REGRESSIVO (AR)

Dado um valor qualquer  $x_t$  dentro de uma série temporal, o atraso  $p$  é definido como sendo  $x_{t-p}$ . Um processo Auto-Regressivo de ordem  $p$  (AR( $p$ )) é definido como a combinação linear dos  $p$  atrasos referentes à observação  $x_t$  com a adição de um ruído branco gaussiano  $a_t$  (BOX; JENKINS, 1976). Vale ressaltar que o Modelo Auto-Regressivo é formalizado para trabalhar apenas com séries temporais estacionárias (Seção 2.1).

A Equação 3.4 apresenta a definição formal deste último, em que nota-se sua semelhança com a Equação 3.3

$$\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + \dots + \phi_p \tilde{x}_{t-p} + a_t \quad (3.4)$$

sendo que  $\tilde{x}_t = x_t - \mu$ ,  $\phi_p$  é o coeficiente de ponderação referente ao atraso  $p$  e  $a_t$  é o ruído branco gaussiano.

Segundo Box, Jenkins e Reinsel (2008) a estacionaridade da série é mantida desde que os coeficientes de ponderação escolhidos tenham valores absolutos menores que 1, ou seja,  $|\phi_i| < 1, \forall i \in \{1, 2, \dots, p\}$ . Além disso, essa condição garante que uma mudança finita incremental na entrada do modelo resulte em uma outra mudança finita incremental na saída do mesmo.

No que se refere ao termo  $a_t$ , ele é considerado como o erro inerente ao processo de regressão, posteriormente, quando o modelo for utilizado para prever valores futuros ele é o erro da previsão. Assim, o cálculo dos coeficientes  $\phi_p$  ótimos deve ser feito com o objetivo de minimizar o erro  $a_t$  (HAYKIN, 2013). Esse cálculo pode ser feito de maneira analítica, o que é muito vantajoso em termos computacionais.

Para determinar os coeficientes  $\phi_p$  ótimos é necessário resolver uma relação de recorrência que nasce a partir da sua função de autocorrelação. Entretanto, é necessário antes fazer algumas manipulações na Equação 3.4, que descreve o modelo AR. Inicialmente multiplicamos ambos os lados da equação por  $x_{t-j}$ , sendo  $j \in \{1, 2, \dots, p\}$ . Em seguida, aplicamos o operador de esperança matemática também em ambos os lados, resultando na Equação 3.5

$$E(x_{t-j}x_t) = E(\phi_1 x_{t-j}x_{t-1}) + E(\phi_2 x_{t-j}x_{t-2}) + \dots + E(\phi_p x_{t-j}x_{t-p}) + E(a_t x_{t-j}) \quad (3.5)$$

Como  $E(a_t x_{t-j}) = 0, \forall j > 0$  e a média do processo também é zero, como mencionado no Capítulo 2. Podemos calcular os valores de autocorrelação entre os termos da série de acordo com a Equação 2.5 resultando na Equação 3.6.

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}, \forall p \leq j > 0 \quad (3.6)$$

Foi descrito no Capítulo 2 que existe uma relação entre covariância e correlação. Portanto ao dividirmos a Equação 3.6 por  $\gamma_0$  chegamos na Equação 3.7, que é a relação de recorrência que define os coeficientes  $\phi_p$  ótimos.

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, \forall j > 0 \quad (3.7)$$

Se expandirmos esta relação para  $j = 1, 2, \dots, p$  obteremos o conjunto de equações lineares 3.8 que definem  $\phi_1, \phi_2, \dots, \phi_p$  em função de  $\rho_1, \rho_2, \dots, \rho_p$ .

$$\begin{aligned} \rho_1 &= \phi_1 \rho_0 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \rho_0 + \dots + \phi_p \rho_{p-2} \\ \rho_3 &= \phi_1 \rho_2 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-3} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \rho_0 \end{aligned} \quad (3.8)$$

Esse conjunto recebe o nome de equações de *Yule-Walker* e sua função é determinar os coeficientes ótimos para um modelo AR(p) (BOX; JENKINS; REINSEL, 2008). Lembrando que  $\rho_0 = 1$  podemos reescrever esse sistema pelas seguintes matrizes:

$$\mathbf{P}_p = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix}$$

$$\boldsymbol{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}$$

$$\boldsymbol{\Phi}_p = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

Para determinar os coeficientes do modelo AR(p) basta resolver a multiplicação de matrizes dada pela Equação 3.9.

$$\boldsymbol{\Phi}_p = \mathbf{P}_p^{-1} \boldsymbol{\rho}_p \quad (3.9)$$

Por fim, o valor previsto pelo modelo AR(p) é dado pela Equação 3.10

$$\tilde{x}_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} \quad (3.10)$$

tal que,  $\tilde{x}_t$  é o valor previsto para o  $t$ -ésimo elemento da série,  $\phi_i$  é o coeficiente referente ao atraso  $x_{t-i}$  e  $x_i$  é o  $i$ -ésimo elemento da série que será utilizado como atraso para prever o valor de  $x_t$ .

### 3.2 MODELO AUTO-REGRESSIVO DE MÉDIAS MÓVEIS (ARMA)

Outro modelo linear comumente utilizado na previsão de séries temporais é o Médias Móveis (MA, do inglês *Moving Average*). Nele, diferentemente do Auto-Regressivo visto na Seção 3.1, que combina de forma ponderada os valores passados da série, são combinados sinais de ruídos brancos (BUENO, 2011). Um modelo MA é dito de ordem  $q$  se para prever o sinal  $x_t$  são utilizados  $q$  amostras de sinais de ruído branco, como na Equação 3.11, a qual apresenta um modelo  $MA(q)$ .

$$x_t = -\theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + a_t \quad (3.11)$$

sendo  $\theta_t, \forall t \in 1, 2, \dots, q$  os parâmetros do modelo.

Por sua vez um Modelo Auto-regressivo de Médias Móveis (ARMA, do inglês *Autoregressive Moving and Average*) é a junção do modelo Auto-Regressivo com o modelo de Médias Móveis. Desse modo, para realizar previsões o modelo ARMA de ordem  $p, q$  faz uso dos  $p$  sinais anteriores (modelo AR) e de  $q$  sinais de ruído branco (MA). Matematicamente um Modelo  $ARMA(p, q)$  é descrito pela Equação 3.12

$$x_t = \phi_1 \tilde{x}_{t-1} + \phi_2 \tilde{x}_{t-2} + \dots + \phi_p \tilde{x}_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + a_t \quad (3.12)$$

com  $\phi_t$  e  $\theta_t$  sendo os parâmetros do modelo.

Diferentemente do modelo AR, o cálculo dos coeficiente do ARMA é feito através da resolução de equações-não lineares. Entretanto, se a escolha desses coeficientes for adequada é possível alcançar um preditor linear ótimo (MORETTIN; TOLOI, 2006).

## 4 REDES NEURAIIS ARTIFICIAIS

Redes Neurais Artificiais (RNA) são classificadas como sistemas distribuídos compostos por unidades simples de processamento de dados. Essas unidades são chamadas de neurônios artificiais e são capazes de computar funções matemáticas que na maioria dos casos são não lineares. Os neurônios artificiais são conectados por conexões normalmente unidirecionais podendo ser dispostos em uma ou mais camadas (BRAGA; CARVALHO; LUDERMIR, 2007).

Devido a sua capacidade de aprendizado, mediante a aplicação de um método de treinamento, e generalização do conhecimento adquirido por meio da solução de instâncias do problema para as quais não se conhece resposta, as redes neurais são amplamente utilizadas em várias áreas da ciência e engenharia (SILVA; SPATTI; FLAUZINO, 2010).

Dahl et al. (2013) utilizaram uma estratégia baseada em RNAs para criar um filtro de *Malwares*. Os autores treinaram os modelos com 2,6 milhões de amostras. Como resultado, obtiveram 2 filtros, o primeiro utilizando apenas uma rede neural que obteve uma média de acerto de 99,51% e o segundo, utilizando uma combinação de várias redes neurais, com média de acerto de 99,58%.

Ainda no campo da classificação de padrões, Molchanov et al. (2016) realizaram a classificação de gestos das mãos através de uma Rede Neural Convolutacional 3D. Os proponentes construíram um sistema online de classificação de gestos que alcançou uma taxa de acerto de 83,4% nos testes, ficando muito perto do desempenho humano, que foi de 88,3%.

Outra área em que as RNAs tem sido muito aplicadas é na previsão de séries temporais. Krizhevsky, Sutskever e Hinton (2012) utilizaram-nas para a previsão da radiação solar sobre a Itália. Nos testes realizados, a rede MLP obteve uma taxa de acerto de 98% nos dias ensolarados e 94% nos dias nublados.

O mercado de ações é conhecido pela sua alta volatilidade, o que torna a tarefa de prever o valor de uma ação algo extremamente complexo. Entretanto, abordagens com o uso de RNAs vem obtendo bons resultados nessa tarefa. Kristjanpoller e Minutolo (2015) realizaram a previsão do preço do ouro. Os testes foram feitos com o preço em Dolar e Euro e o melhor resultado obtido alcançou taxa de acerto de 75%.

Neste capítulo serão apresentados os conceitos a respeito de RNAs, desde a concepção do Neurônio Artificial na Seção 4.1, passando pelos tipos de arquiteturas na Seção 4.2, o treinamento na Seção 4.3 e descrevendo as redes utilizadas nesse trabalho entre as seções 4.4 e 4.9. Por fim, a Seção 4.10 descreve como Redes Neurais Artificiais são utilizadas para prever valores futuros em séries temporais.

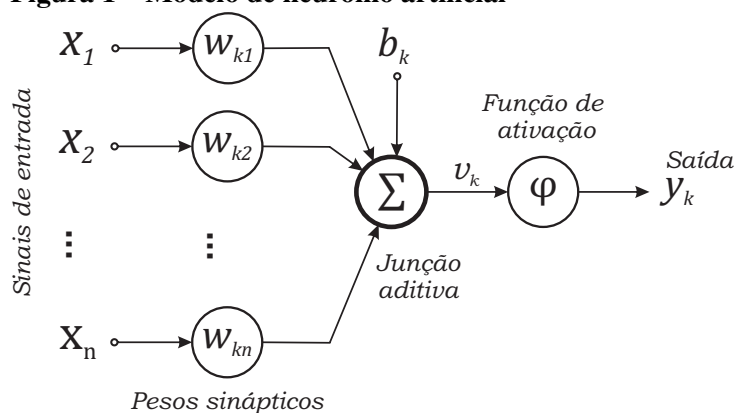
#### 4.1 O NEURÔNIO ARTIFICIAL

Em Silva, Spatti e Flauzino (2010) neurônios artificiais são definidos como modelos simplificados dos neurônios biológicos, inspirados a partir da análise da geração e propagação de impulsos elétricos pela membrana celular dos neurônios proposto por Hodgkin e Huxley (1952).

Neurônios artificiais são em geral não-lineares, sua saída é contínua e realizam funções simples, como coletar os valores de todas as suas entradas e misturá-los de acordo com sua função operacional e produzir uma resposta levando em consideração sua função de ativação (SILVA; SPATTI; FLAUZINO, 2010). O primeiro neurônio artificial foi concebido por McCulloch e Pitts (1943), o qual tratava-se de uma simplificação do que se conhecia na época a respeito do funcionamento dos neurônios biológicos (BRAGA; CARVALHO; LUDERMIR, 2007).

A Figura 1 mostra um dentre os diversos modelos de neurônio artificial existentes, denominado neurônio  $k$ , que é largamente usado como base para projetos de RNAs. Nela é possível ver as entradas do neurônio, os pesos de cada entrada, a função que combina as entradas, a função de ativação e a saída.

**Figura 1 – Modelo de neurônio artificial**



**Fonte: Adaptado de Haykin (2001).**

O conjunto de entradas do neurônio  $k$ , também conhecido como elos de conexões, é definido por  $\mathbf{x}(n_k) = \{x_1, x_2, x_3, \dots, x_n\}$ , onde  $n$  é o número total de entradas do neurônio  $n_k$ .

Cada entrada  $x_i \in \mathbf{x}(n_k)$  é ponderada por um peso sináptico  $w_{ki}$  associado, o qual é multiplicado pelo valor recebido na entrada. Na Figura 1 é possível ver que o valor da entrada  $x_1$  é aplicado ao peso sináptico  $w_{k1}$  e o sinal correspondente é passado à junção aditiva.

Por sua vez, a Junção Aditiva (ou Combinador Linear) é responsável por somar todos os valores recebidos das entradas ponderadas pelos seus respectivos pesos sinápticos.

O neurônio  $k$  possui ainda um valor de bias, também chamado de entrada de polarização ou ainda limiar de ativação, definido como  $b_k$ . Esse valor é acrescido na soma obtida pela Junção Aditiva, podendo assumir tanto valores positivos quanto negativos. O mais comum é utilizar  $b_k = 1$ . Este também pode ser interpretado como mais um peso do neurônio.

Após a adição do bias, o valor  $v_k$  é submetido a função de ativação  $\varphi(\cdot)$ . Esta, por sua vez, restringe a amplitude do resultado a um valor finito e insere um grau de não-linearidade à saída do neurônio. Tal amplitude é comumente definida como um intervalo unitário  $[0, 1]$  ou ainda  $[-1, 1]$ .

Pode-se então definir matematicamente o neurônio artificial  $k$  pela Equação 4.1

$$y_k = \varphi \left( \left[ \sum_{i=1}^n x_i w_{ki} \right] + b_k \right) \quad (4.1)$$

em que  $x_1, x_2, \dots, x_n$  são os sinais de entrada,  $w_{k1}, w_{k2}, \dots, w_{kn}$  são os pesos sinápticos,  $b_k$  é o bias,  $\varphi(\cdot)$  é a função de ativação e  $y_k$  é o valor de saída do neurônio  $k$ .

#### 4.1.1 Tipos de Função de Ativação

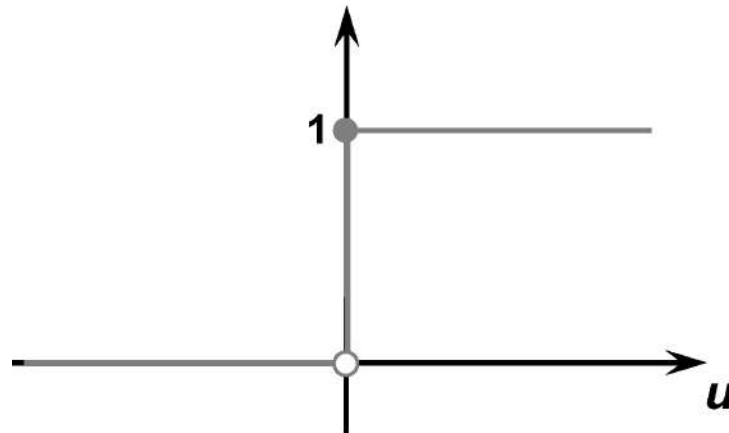
Silva, Spatti e Flauzino (2010) definem a função de ativação como sendo responsável por limitar a saída do neurônio dentro de um intervalo de valores razoáveis a serem assumidos pela sua própria imagem funcional. Imagine um problema de classificação em que uma rede neural é usada para classificar amostras em duas classes: a função de ativação transforma a saída numérica do Combinador Linear em números que representam essas duas classes (por exemplo 0 ou 1).

Essas funções são divididas em dois grupos: parcialmente diferenciáveis e totalmente diferenciáveis. O primeiro grupo compreende as funções em que pelo menos um ponto não possui derivada de primeira ordem definida. As três principais são: degrau, degrau bipolar e rampa simétrica. A Equação 4.2 descreve a função degrau. Nela a saída numérica da rede é convertida para uma saída binária. Valores maiores ou iguais a 0 assumem o valor de 1 e, caso contrário, 0.

$$\varphi(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases} \quad (4.2)$$

O Gráfico 5 ilustra a representação gráfica da função degrau. Nele é claro ver seu comportamento binário.

Outra função de ativação parcialmente diferenciável é a degrau bipolar, também chamada de função sinal. Os valores provenientes dela são valores unitários positivos ou negativos, ou ainda valores nulos. Deste modo, se o valor for menor que 0 a função retornará o valor  $-1$ , ou 1, em caso contrário. Se o valor for exatamente 0 a saída é reproduzida. A Equação 4.3 descreve

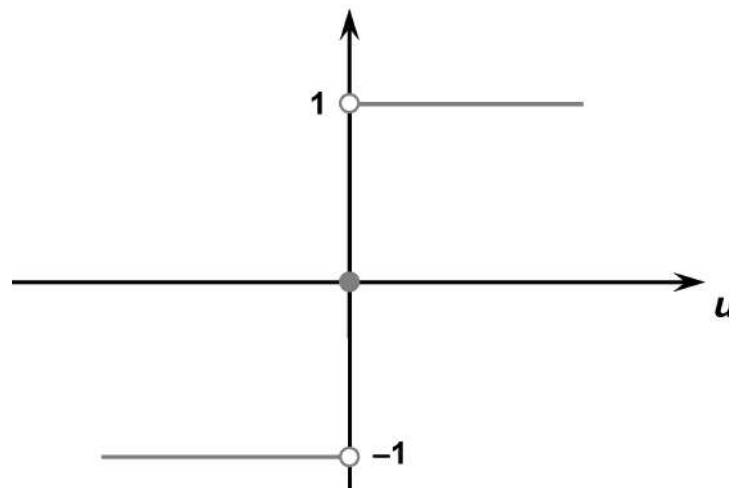
**Gráfico 5 – Função de ativação degrau.**

Fonte: (SILVA; SPATTI; FLAUZINO, 2010).

o funcionamento dessa função.

$$\varphi(u) = \begin{cases} 1, & \text{se } u > 0 \\ 0, & \text{se } u = 0 \\ -1, & \text{se } u < 0 \end{cases} \quad (4.3)$$

Quando aplicada em problemas de classificação, a função degrau bipolar pode ser alterada para retornar apenas valores unitários positivos ou negativos. O Gráfico 6 ilustra este comportamento.

**Gráfico 6 – Função de ativação degrau bipolar.**

Fonte: (SILVA; SPATTI; FLAUZINO, 2010).

A última função de ativação parcialmente diferenciável é a rampa simétrica. Para esta é necessário definir um intervalo  $[-a, a]$ . Se o limiar de ativação estiver dentro desse intervalo não é feita qualquer modificação em seu valor. Caso contrário ele deve ser alterado para o valor

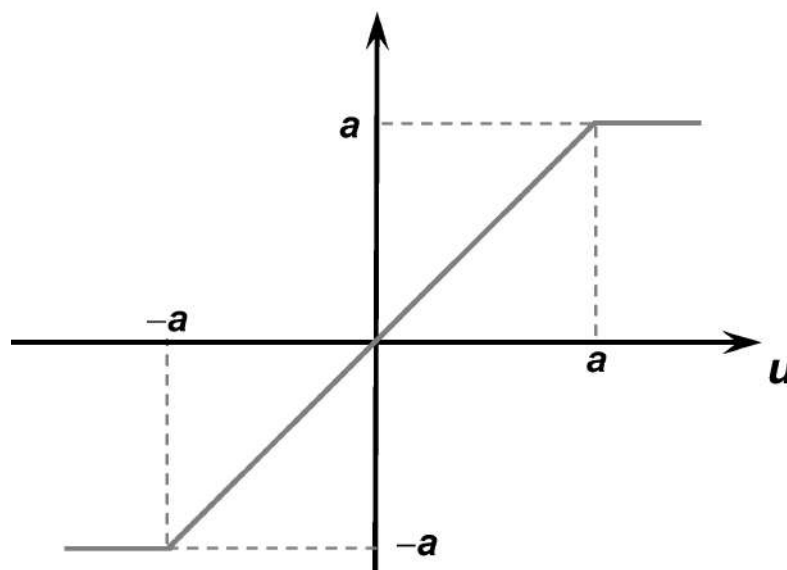


mais próximo dentro do intervalo  $[-a, a]$ . A Equação 4.4 apresenta sua definição matemática.

$$\varphi(u) = \begin{cases} a, & \text{se } u > a \\ u, & \text{se } -a \leq u \leq a \\ -a, & \text{se } u < -a \end{cases} \quad (4.4)$$

O Gráfico 7 mostra o comportamento da função.

**Gráfico 7 – Função de ativação rampa simétrica**



**Fonte: (SILVA; SPATTI; FLAUZINO, 2010).**

Em contrapartida das funções parcialmente diferenciáveis, o segundo grupo de funções de ativação, chamadas de funções totalmente diferenciáveis, compreende aquelas em que as derivadas de primeira ordem existem e são conhecidas para todos os pontos do seu domínio. As duas funções desse tipo que mais são utilizadas no contexto neural são a logística e tangente hiperbólica.

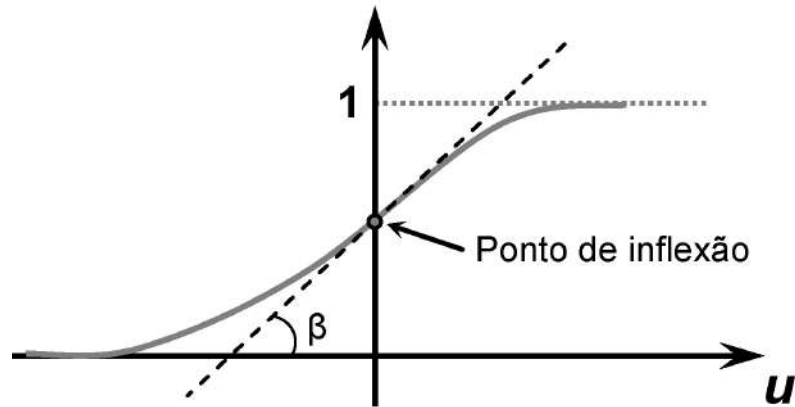
A função logística produz sempre valores reais compreendidos entre o intervalo de 0 e 1, sendo definida pela Equação 4.5

$$\varphi(u) = \frac{1}{1 + e^{-\beta u}} \quad (4.5)$$

em que  $\beta$  é um valor real previamente definido, associado ao nível de inclinação na função em relação ao seu ponto de inflexão. O papel realizado por  $\beta$  fica mais evidente ao analisar o Gráfico 8.

Pelo Gráfico 8 verifica-se o formato geométrico da função logística em que, quanto maior for o valor assumido por  $\beta$  (tendendo ao infinito), mais o comportamento da função se assemelhará com o da função degrau. Entretanto vale ressaltar que diferentemente da última, que não é diferenciável em todos os pontos, a função logística possui derivada de primeira ordem em todos os pontos do seu domínio.

**Gráfico 8 – Função de ativação logística.**



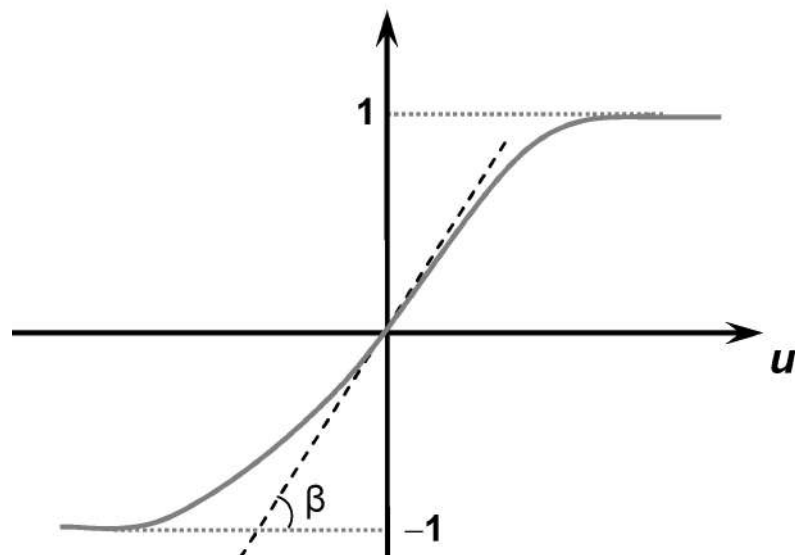
Fonte: (SILVA; SPATTI; FLAUZINO, 2010).

No mesmo grupo das funções totalmente diferenciáveis, a segunda função é a Tangente Hiperbólica, tendo uma curva similar a da função logística suas saídas estão compreendidas no intervalo  $[-1, 1]$  sendo definida pela Equação 4.6.

$$\varphi(u) = \frac{1 - e^{-\beta u}}{1 + e^{-\beta u}} \quad (4.6)$$

Aqui novamente o parâmetro  $\beta$  está associado à inclinação da curva em relação a seu ponto de inflexão, tal comportamento é mais evidente ao observar o Gráfico 9 que apresenta a curva da função Tangente Hiperbólica.

**Gráfico 9 – Função de ativação Tangente Hiperbólica.**



Fonte: (SILVA; SPATTI; FLAUZINO, 2010).

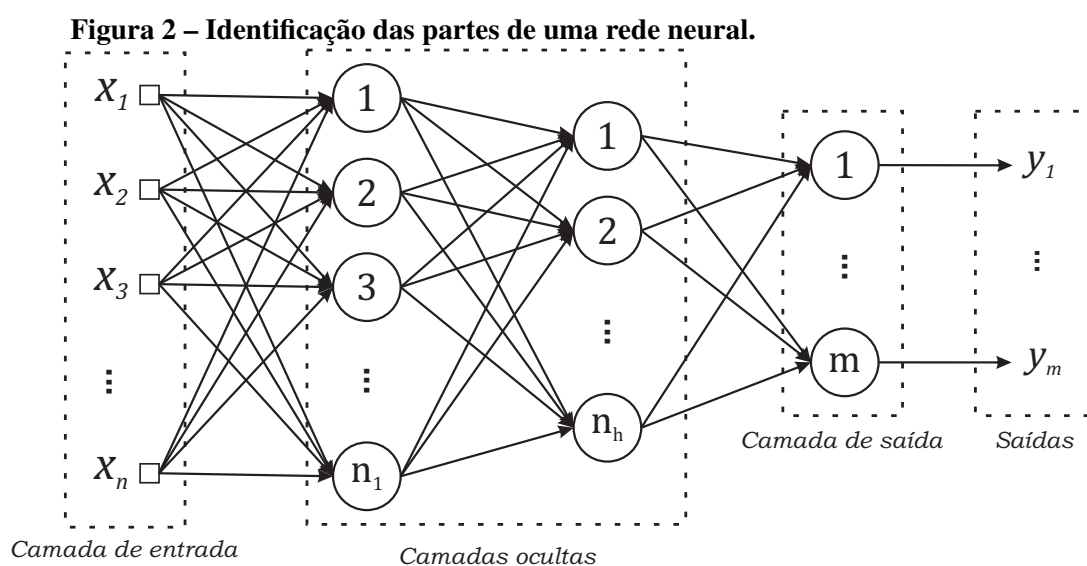
Assim como ocorre para a função logística, analisando o Gráfico 9 nota-se que quanto maior for o valor de  $\beta$  maior será a inclinação da função, de modo que para valores muito elevados de  $\beta$  o comportamento da função tangente hiperbólica se assemelhará ao da função degrau bipolar.

## 4.2 ARQUITETURA DE REDES NEURAIS

A forma como os neurônios artificiais estão arranjados, ou dispostos, uns em relação aos outros é chamada de arquitetura de rede neural. Esses arranjos são essencialmente estruturados através do direcionamento das conexões sinápticas dos neurônios (SILVA; SPATTI; FLAUZINO, 2010).

A topologia de uma rede, por sua vez, compreende as diferentes formas de composição estrutural que a mesma pode assumir dentro de uma arquitetura. Como por exemplo, pode-se ter duas topologias dentro de uma mesma arquitetura, sendo a primeira topologia composta por 20 neurônios enquanto a segunda faz uso de apenas 10. Este termo pode ainda fazer referência a função de ativação de cada neurônio da rede ou da utilização ou não do bias (SILVA; SPATTI; FLAUZINO, 2010).

Uma Rede Neural Artificial pode ser dividida em 3 partes: camada de entrada, camadas intermediárias e camada de saída (HAYKIN, 2008). Uma proposta genérica de RNA é mostrada na Figura 2, em que tais camadas são destacadas.



**Fonte:** Adaptado de Silva, Spatti e Flauzino (2010).

A descrição de cada uma das camadas segue o disposto a seguir:

- **Camada de entrada** - É a camada responsável por receber os sinais de entrada da rede,  $\mathbf{x} = \{x_1, x_2, x_3, \dots, x_n\}$ . Estes usualmente são normalizados para que seus valores compreendam o intervalo de abrangência da função de ativação, já que é sabido que as operações matemáticas realizadas pela rede apresentam uma maior precisão quando as entradas são normalizadas. É importante ressaltar que a camada de entrada geralmente não possui neurônios.
- **Camadas escondidas, intermediárias, ocultas ou invisíveis** - São as que possuem neurônios, mas que não são responsáveis por apresentar a saída da rede. Uma rede neural pode

possuir diversas ou nenhuma camada oculta. A maior parte do processamento interno realizado pela RNA é feito pelas camadas ocultas, sendo estas as responsáveis pela extração das características associadas ao processo ou sistema a ser inferido e a inserção de não-linearidade aos dados em processamento.

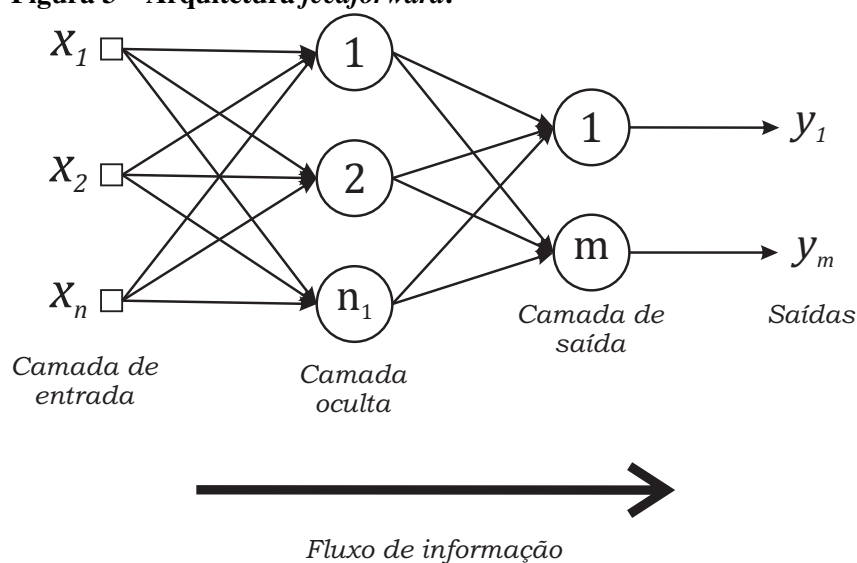
- **Camada de saída** - Responsável pela produção e apresentação dos resultados finais, os quais são gerados mediante os resultados das camadas anteriores. A camada de saída é a última camada de uma RNA, também sendo composta por neurônios. Analisando a Figura 2 verifica-se que a quantidade de neurônios presentes nesta camada corresponde a quantidade de saídas da rede.

Note que apenas as camadas de entrada e saída são obrigatórias em uma rede neural, sendo possível a construção de uma RNA com apenas 1 camada neural que é ao mesmo tempo de entrada e de saída. Como mencionado, uma rede neural pode não ter camadas ocultas ou ter qualquer quantidade de camadas ocultas. Ainda, a quantidade de neurônios e as funções de ativação de cada camada podem ser diferentes entre si.

#### 4.2.1 Arquitetura *Feedforward*

A arquitetura *feedforward* (não-recorrente) compreende o sentido em que o fluxo de informações segue pela rede, tendo origem na camada de entrada e seguindo em direção a camada de saída, seguindo em apenas um sentido. A Figura 3 apresenta uma rede neural *feedforward* genérica.

**Figura 3 – Arquitetura *feedforward*.**



**Fonte: Adaptado de Silva, Spatti e Flauzino (2010).**

É possível identificar o conjunto de entradas da rede composto por  $\mathbf{x} = \{x_1, x_2, \dots,$

$x_n$ }, uma camada oculta ou intermediária com  $n_1$  neurônios e uma camada de saída com  $m$  neurônios.

A rede da Figura 3 contém apenas uma camada oculta, mas a arquitetura *feedforward* independe da quantidade de camadas ou de neurônios. Note ainda que as ligações entre os neurônios são realizadas por setas direcionadas, de forma que o sentido delas representa o fluxo de informação da rede e que o mesmo ocorre em apenas um sentido.

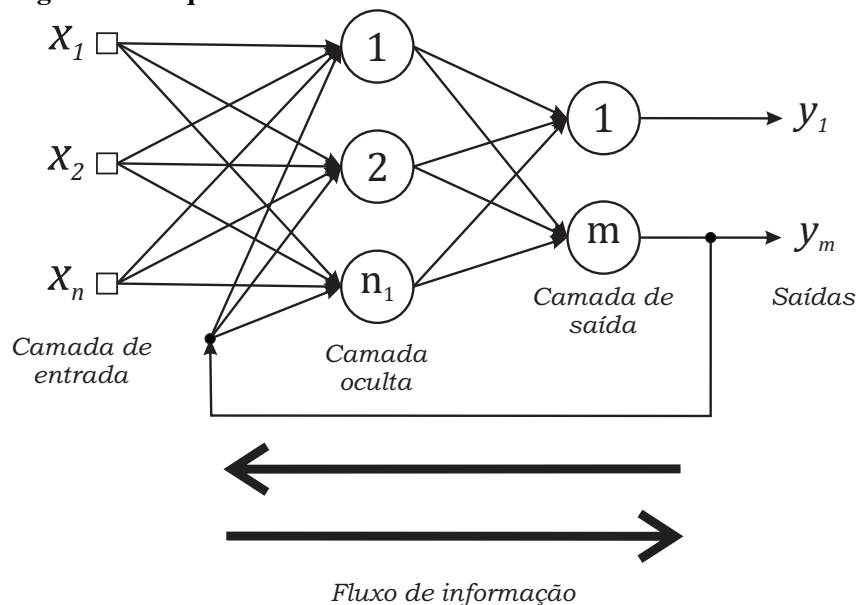
#### 4.2.2 Arquitetura Recorrente

Ao contrário das redes *feedforward*, a arquitetura recorrente (também chamada de realimentada) permite que a informação se propague entre os neurônios em mais de um sentido. Uma rede neural é considerada recorrente se ao menos a saída de um neurônio é utilizada como entrada para outro neurônio de uma camada anterior ou da mesma camada (BRAGA; CARVALHO; LUDERMIR, 2007).

Silva, Spatti e Flauzino (2010) afirmam que a característica recorrente de uma rede a qualifica para ser empregada em processamento dinâmico de informação, ou seja, ela tende a obter bons resultados quando aplicada a modelagem de um sistema variante em relação ao tempo, como em otimização e identificação de sistemas, controle de processos e previsão de séries temporais.

A Figura 4 apresenta um modelo genérico de rede recorrente, em que a saída de um dos neurônios da camada final é usada como entrada pelos neurônios da camada oculta.

**Figura 4 – Arquitetura Recorrente.**



**Fonte: Autoria própria.**

Vale ressaltar que, apesar de na rede da Figura 4 haver apenas uma realimentação, em

uma rede recorrente podem haver quantas forem necessárias. Tampouco existe a obrigatoriedade das realimentações partirem da camada de saída ou terem destino na camada de entrada. Outro detalhe é que apesar da realimentação ser utilizada como entrada por alguns neurônios, essa não faz parte da camada de entrada da rede neural.

### 4.3 TREINAMENTO DA REDE

A principal razão que levou as Redes Neurais Artificiais a serem tão amplamente utilizadas na solução dos mais diversos problemas é a sua capacidade de aprender a partir da apresentação de amostras que representam o comportamento de um sistema. Após ter "aprendido" os padrões e as regras, a mesma é capaz de generalizar esse conhecimento, fornecendo respostas para amostras que não foram utilizadas durante o aprendizado (SIQUEIRA, 2013).

Silva, Spatti e Flauzino (2010) definem o treinamento de uma rede neural como a aplicação de passos ordenados necessários para ajustar os pesos sinápticos e limiares de ativação de seus neurônios, com o objetivo de alcançar a generalização de soluções a serem produzidas pela rede. Esse conjunto de passos é chamado de algoritmo de aprendizagem.

As amostras utilizadas no treinamento são chamadas de conjunto de treinamento, o qual é composto com cerca de 60% a 90% do total de amostras disponíveis, sendo utilizado essencialmente no ajuste dos pesos sinápticos.

Por sua vez, o conjunto de teste é composto pelas amostras restantes (de 10% a 40% das amostras disponíveis) com o objetivo de verificar se a generalização das soluções obtidas está com um nível de erro aceitável para a aplicação em questão (SILVA; SPATTI; FLAUZINO, 2010). Vale ressaltar que as amostras do conjunto de teste não podem ter sido utilizadas no treinamento, desse modo a RNA será testada com um conjunto de amostras ao qual ela nunca teve acesso. Este processo é realizado após o ajuste dos pesos.

Uma etapa primordial que deve ser realizada antes do treinamento da rede é a normalização dos dados, como foi dito na Seção 4.1.1 a função de ativação tem o objetivo de limitar a resposta de uma RNA a valores razoáveis que representem de maneira adequada as possíveis soluções do problema. A normalização consiste em modificar os dados de entrada de modo que os mesmos sejam compreendidos no mesmo intervalo que a função de ativação da rede neural, esse processo torna a rede mais estável e melhora os resultados obtidos.

Com o objetivo de normalizar o valor  $x$  do intervalo  $[min, max]$  para o intervalo  $[min^*, max^*]$  é apresentada a Equação 4.7

$$x_{norma} = \frac{(x - min)(max^* - min^*)}{max - min} + min^* \quad (4.7)$$

tal que,  $x_{norma}$  é o valor normalizada para o novo intervalo  $[min^*, max^*]$ ,  $x$  é o valor original no intervalo  $[min, max]$ ,  $min$  é o menor valor do intervalo original,  $max$  é o maior valor do

intervalo original,  $min^*$  é o menor valor do novo intervalo e  $max^*$  é o maior valor do novo intervalo.

Em termos de custo computacional, o processamento utilizado por uma rede neural é quase que totalmente aplicado durante o ajuste dos pesos sinápticos, visto que os algoritmos de treinamento executam em sua maioria iterações sucessivas até que o erro esteja em um patamar aceitável. Uma vez que os pesos sinápticos estão todos definidos, dada uma entrada qualquer, calcular a saída é simplesmente realizar algumas operações de multiplicação e soma, o que tem um custo computacional baixo.

#### 4.3.1 Treinamento Supervisionado

O treinamento supervisionado induz a ideia de um professor que tem a missão de corrigir a resposta da rede na direção da saída correta. Para tanto, faz-se necessário conhecer qual a resposta correta para cada amostra de treinamento. Esta é chamada de saída esperada (desejada) (HAYKIN, 2008).

Durante o treinamento supervisionado, a cada sinal de entrada inserido na rede é calculada a diferença da saída da rede em relação a saída esperada, sendo esta utilizada no ajuste dos pesos sinápticos. O treinamento da rede é considerado completo quando a defasagem entre estas saídas atinge um nível aceitável. Tal valor varia de acordo com a aplicação e deve ser previamente definido pelo projetista antes de iniciar-se o treinamento (HAYKIN, 2008).

Note que, na prática, o treinamento tem o objetivo de minimizar a diferença entre as saídas da rede e as saídas desejadas. Logo, existe a possibilidade da rede ser capaz de responder com nível de certeza adequado apenas às amostras do subconjunto de treinamento, ou seja, aquelas amostras que ela já teve acesso (SILVA; SPATTI; FLAUZINO, 2010). Inclusive, se uma rede tiver um número elevado de neurônios em comparação ao número de amostras de entrada, o erro de saída no treinamento tende a ser zero se ela for exaustivamente ajustada. Embora esta afirmação pareça interessante, ao testar-se a rede com amostras inéditas, a resposta final tende a ser muito ruim. Diz-se, então, que a mesma está sobre-treinada e que perdeu o seu poder de generalização (HAYKIN, 2008).

Outro fator que pode levar uma rede neural a perder o seu poder de generalização é a utilização de amostras que não representam de maneira significativa o sistema a ser mapeado. Um dos motivos mais comuns para esse problema é a utilização de poucas amostras no treinamento (SILVA; SPATTI; FLAUZINO, 2010).

Para evitar tal comportamento indesejado, existem técnicas que podem ser aplicadas para contrabalançar esse efeito, por exemplo, a validação cruzada.

### 4.3.2 Validação cruzada

Ao ajustar os pesos sinápticos de uma rede neural durante o processo de treinamento, a saída da rede é regulada em função das amostras do conjunto de treinamento. Quanto mais se treinar uma rede (alto número de iterações), mais suas saídas serão próximas ao sinal desejado deste conjunto. Na prática isso leva a rede a responder de forma insatisfatória para dados desconhecidos, ou seja, para todas as amostras que não fizerem parte do conjunto de treinamento a resposta da rede tende a ser insatisfatória (HAYKIN, 2008).

A técnica de validação cruzada é uma ferramenta que ajuda na maximização da capacidade de generalização da rede, por meio de um método sistemático que indica o melhor ponto, ou número de iterações, em que a rede deve parar o ajuste dos pesos. Consiste em, ao final de cada época de treinamento, submeter a rede a um conjunto de dados desconhecido. Para isso, o conjunto de teste é dividido em dois subconjuntos, o primeiro, chamado de conjunto de validação que é utilizado na validação cruzada durante o treinamento, o segundo, o conjunto de teste será utilizado para medir o desempenho da rede após o processo de treinamento. O conjunto de pesos sinápticos escolhido é aquele que apresentar o melhor desempenho para o conjunto de validação e não para o conjunto de treinamento, ou seja, aquele em que a rede apresentar o menor erro mediante amostras que não foram utilizadas para ajustar seus pesos (HAYKIN, 2008).

Assim, apenas após ter selecionado a configuração de pesos que apresenta o melhor desempenho para o conjunto de validação, a rede é exposta ao conjunto de teste. Vale ressaltar que para que esse processo funcione de maneira eficiente, cada amostra deve estar localizada em apenas um dos conjuntos.

## 4.4 PERCEPTRON DE MÚLTIPLAS CAMADAS (MLP)

A rede neural *Perceptron* de Múltiplas Camadas (MLP, do inglês *Multilayer Perceptron*) é constituída por um conjunto de neurônios artificiais dispostos em várias camadas, de modo que o sinal de entrada se propaga adiante através da rede, camada por camada. A camada de entrada, como o próprio nome já diz, recebe os sinais de entrada, a camada de saída entrega o resultado obtido e entre elas podem existir quantas camadas ocultas forem necessárias (HAYKIN, 2001).

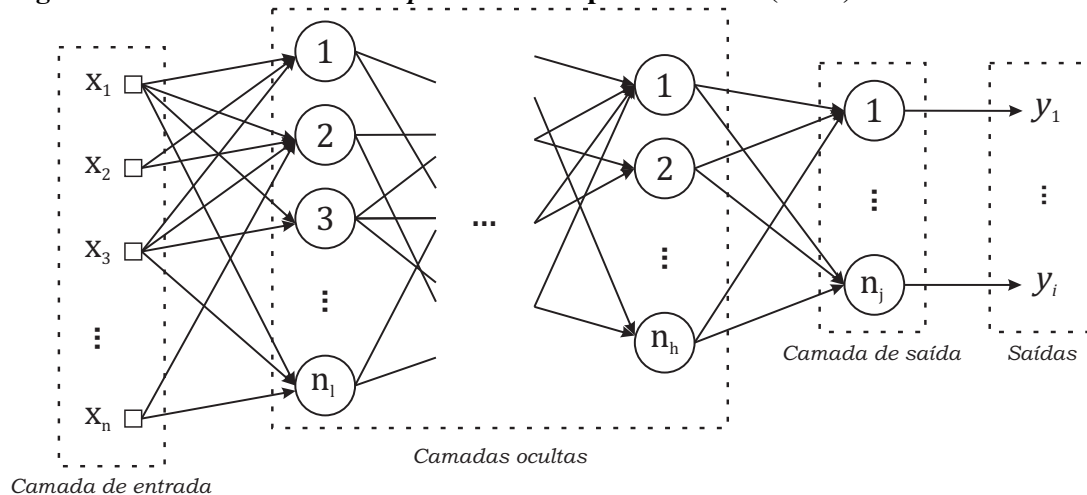
No trabalho de Cybenko (1989), o autor prova que redes neurais com duas camadas ocultas podem aproximar qualquer função linear ou não-linear. Por sua vez, redes de camada única também podem aproximar qualquer função, desde que ela seja contínua, limitada em um espaço compacto e diferenciável em todos os pontos.

As MLPs são consideradas uma das arquiteturas mais versáteis quanto a aplicabilidade, sendo utilizadas em diversas áreas do conhecimento como em aproximação universal de funções,



reconhecimento de padrões, identificação e controle de processos, previsão de séries temporais e otimização de sistemas (SILVA; SPATTI; FLAUZINO, 2010). A Figura 5 mostra um modelo genérico de MLP.

**Figura 5 – Modelo de rede *Perceptron* de Múltiplas Camadas (MLP)**



**Fonte: Adaptado de Silva, Spatti e Flauzino (2010).**

Nele é possível identificar a camada de entrada, as ocultas e por fim a de saída. Note que cada neurônio  $n_j$  da camada de saída entrega um resultado  $y_j$  do processo a ser mapeado. Assim, a quantidade de neurônios desta camada é dada pelo número de saídas necessárias para compor uma instância do problema (SILVA; SPATTI; FLAUZINO, 2010). Vale lembrar que, como evidência a Figura 5, a camada de entrada de uma MLP não é uma camada neural, ou seja, não possui neurônios.

Ainda analisando o modelo genérico de MLP, nota-se que os sinais fornecidos pela camada de entrada são recebidos pela primeira camada oculta para que sejam processados. O resultado da saída desta é utilizado como entrada para a segunda camada oculta, e assim sucessivamente até que a última camada oculta entregue seu resultado como entrada para a de saída, que por sua vez entrega o resultado final da rede. Assim, não importa a quantidade de camadas em uma MLP, a propagação dos sinais de entrada ocorre sempre em um único sentido (arquitetura *feedforward*), da entrada em direção à saída (SILVA; SPATTI; FLAUZINO, 2010).

#### 4.4.1 Treinamento MLP

Como estabelecido anteriormente, o processo de treinamento consiste no ajuste dos pesos sinápticos com o objetivo de encontrar o conjunto que atinja o melhor mapeamento a respeito do evento desejado. O método mais conhecido para MLPs é o do gradiente ótimo, em que o vetor gradiente é calculado por meio do algoritmo de retropropagação ou *backpropagation* criado por Werbos (1974) em sua tese de doutorado, mas popularizado por Rumelhart, Hinton e Williams (1986). Este processo é feito de forma supervisionada.

O algoritmo consiste em duas fases. Inicialmente, os dados de entrada são propagados pela rede a fim de obter as suas saídas. Em seguida, estes valores são comparados com os esperados e assim é obtida uma medida de erro. Na segunda etapa é realizado o caminho oposto, da camada de saída para a camada de entrada. Neste caso, todos os pesos sinápticos são ajustados de acordo com a regra de correção de erro assumida para que a saída dada pela rede na iteração seguinte seja mais próxima à esperada (SILVA; SPATTI; FLAUZINO, 2010).

O sinal de erro de um neurônio  $j$  na iteração  $t$  é dado pela Equação 4.8

$$e_j(t) = d_j(t) - y_j(t) \quad (4.8)$$

onde  $e_j(t)$  é o erro,  $d_j(t)$  é o resultado esperado (saída desejada) e  $y_j(t)$  é a saída dada pela rede.

A cada iteração deseja-se minimizar o erro de saída. A métrica mais usual utilizada é o Erro Quadrático Médio (MSE, do inglês *Mean Square Error*), dado pela Equação 4.9

$$MSE(t) = \frac{1}{N} \sum_{j=1}^N (d_j(t) - y_j(t))^2 \quad (4.9)$$

na qual  $N$  é o número amostras no treinamento.

Por fim as regras para atualização dos pesos sinápticos e dos bias de cada neurônio são dadas pelas Equações 4.10 e 4.11, respectivamente

$$w_{ij}^m(t+1) = w_{ij}^m(t) - \alpha \frac{\partial MSE(t)}{\partial w_{ij}^m(t)} \quad (4.10)$$

$$b_j^m(t+1) = b_j^m(t) - \alpha \frac{\partial MSE(t)}{\partial b_j^m(t)} \quad (4.11)$$

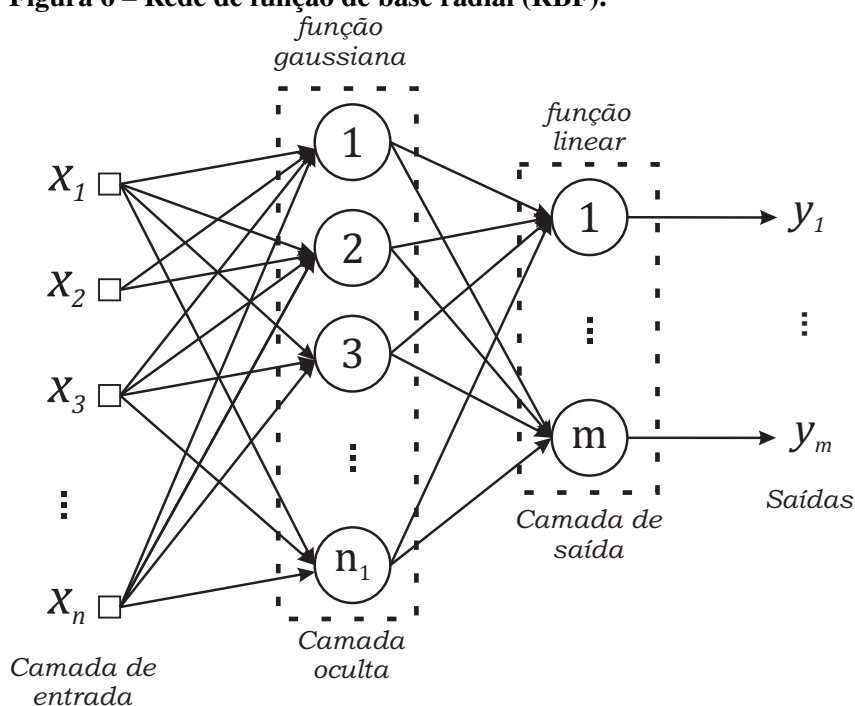
em que  $w_{ij}^m(t)$  é o peso sináptico da entrada  $i$  do neurônio  $j$  da camada  $m$  na iteração  $t$ ,  $\partial E(t)$  é a derivada parcial do MSE e  $b_j^m(t)$  é o bias do neurônio  $j$  da camada  $m$  na iteração  $t$ .

#### 4.5 REDES NEURAIS COM FUNÇÃO DE BASE RADIAL

As redes neurais com função de base radial (RBF, do inglês *Radial Basis Function*) são um método computacional *feedforward* de camadas múltiplas. Entretanto, diferentemente das MLPs, as redes RBFs possuem apenas 2 camadas, sendo uma oculta e outra de saída. Na primeira, todas as funções de ativação são do tipo gaussiana (SILVA; SPATTI; FLAUZINO, 2010). A Figura 6 apresenta a configuração genérica de uma RBF.

Na Figura 6, é possível verificar as duas camadas neurais, a função de ativação da camada oculta do tipo gaussiana e a função de ativação da camada de saída, a qual é linear. Outro detalhe importante é a ausência do bias nos neurônios da camada oculta, de forma que apenas os neurônios da saída possuem este componente.

**Figura 6 – Rede de função de base radial (RBF).**



**Fonte: Adaptado de Silva, Spatti e Flauzino (2010).**

O treinamento das RBFs é realizado em duas etapas. Primeiramente são calculados os pesos da camada intermediária e em seguida se ajustam os pesos da camada de saída. Isto é um claro contraste em comparação com a MLP.

#### 4.5.1 Treinamento RBF

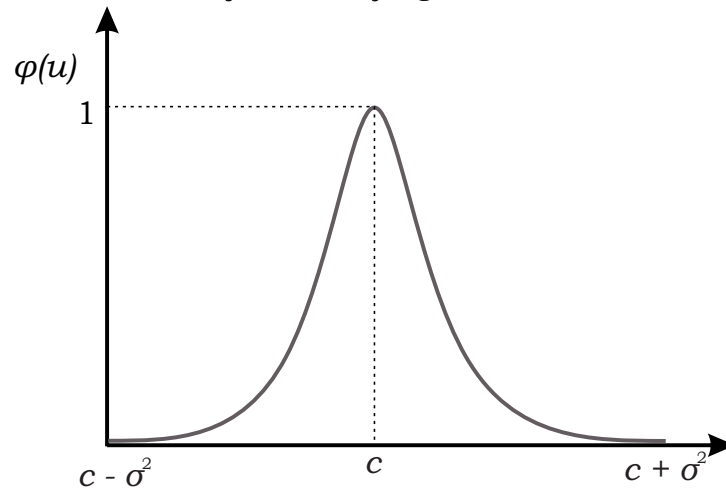
Como dito, a função de ativação dos neurônios da camada intermediária de uma RBF é de base radial, sendo a gaussiana uma das mais empregadas. Analogamente, nota-se que quanto maior for a variância maior será a extensão da base da função. A Equação 4.12 define uma função gaussiana

$$\varphi(u) = e^{-\frac{(u-c)^2}{2\sigma^2}} \quad (4.12)$$

na qual  $c$  é o centro da gaussiana e  $\sigma^2$  a sua variância em função do centro, de modo que  $c - \sigma$  e  $c + \sigma$  formam as extremidades da função. O Gráfico 10 ilustra a curva de uma função gaussiana.

Na prática, cada neurônio da camada intermediária acaba por ser uma função gaussiana, em que seu centro é definido pelos seus pesos sinápticos. Assim, o processo de treinamento ajusta o centro e o valor da variância da base de cada função.

Inicialmente são atribuídos os valores das entradas aos pesos dos neurônios da camada intermediária, de modo que os pesos do neurônio  $n_1$  são os valores da entrada  $x_1$ , os pesos do neurônio  $n_2$  são os pesos da entrada  $x_2$  e assim por diante.

**Gráfico 10 – Função de ativação gaussiana.**

Fonte: Adaptado de Silva, Spatti e Flauzino (2010).

Em seguida é criado um grupo de entradas  $\Omega^n$  para cada neurônio  $n$  da camada intermediária. Todas as entradas do conjunto de treinamento são adicionadas aos conjuntos  $\Omega^n$  pela seguinte regra: uma entrada  $x_i$  é adicionada ao grupo  $\Omega^n$  do neurônio  $n$  se a base da função do neurônio  $n$  for a mais próxima da entrada  $x_i$ . Para isso, comumente é utilizada a distância euclidiana.

Uma vez que todas as entradas foram inseridas em  $\Omega^n$  os centros das gaussianas dos neurônios são atualizados para corresponder ao centro de todas as entradas contidas no seu grupo  $\Omega^n$ . Após essa atualização dos centros, todos os grupos  $\Omega^n$  são esvaziados e todas as entradas são adicionadas novamente no grupo  $\Omega^n$  mais próximo. Esses passos são repetidos até que não haja mais mudança nos grupos  $\Omega^n$ .

Após calcular o centro de todas as gaussianas é necessário calcular suas variâncias, as quais são calculadas pela Equação 4.13

$$\sigma_j^2 = \frac{1}{|\Omega^j|} \sum_{x^k \in \Omega^j} \sum_{i=1}^n (x_i^k - w_{ji}^1)^2 \quad (4.13)$$

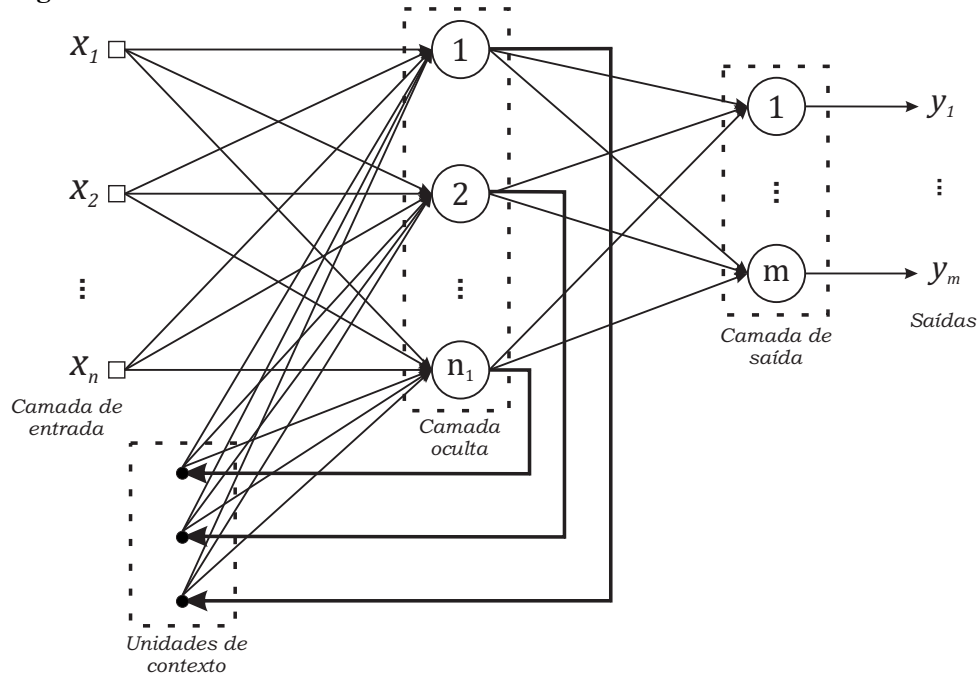
na qual  $\sigma_j^2$  é a variância da base da função gaussiana do neurônio  $j$ ,  $|\Omega^j|$  é o tamanho do conjunto  $\Omega$  do neurônio  $j$ ,  $x^k$  é a amostra  $k$  do conjunto  $\Omega$ ,  $n$  é o número de entradas de cada amostra,  $x_i^k$  é a entrada  $i$  da amostra  $k$  e  $w_{ji}^1$  é o peso do neurônio  $j$  da camada intermediária para a entrada  $i$ .

É importante ressaltar que o treinamento da camada de saída da RBF só pode ser realizado após o término do treinamento da camada intermediária. Uma vez que os pesos sinápticos e as variâncias da camada intermediária foram definidos, os pesos da camada de saída são ajustados da mesma forma que ocorre na MLP, podendo também ser utilizado o algoritmo de retropropagação do erro descrito pelas equações 4.10 e 4.11.

#### 4.6 REDE NEURAL DE ELMAN

A Rede Neural de Elman é uma arquitetura de rede neural recorrente criada por Elman (1990) a partir de uma rede MLP. Nela a camada de entrada é dividida em 2 partes, a primeira compreende as entradas da rede propriamente ditas e a segunda, por sua vez, chamada de unidade de contexto, consiste nas saídas da camada oculta. A Figura 7 ilustra uma rede recorrente de Elman em que é possível observar as realimentações saindo dos neurônios da camada de entrada e sendo reutilizados como entrada da rede.

**Figura 7 – Rede neural de Elman**



**Fonte: Adaptado de Silva, Spatti e Flauzino (2010).**

Perceba que o sinal de saída de um neurônio não é produzido enquanto as entradas não passam pelo combinador linear do mesmo. Logo, o sinal de saída de um neurônio no instante  $t$  não pode ser usado como entrada para as camadas anteriores ou para a própria camada no mesmo instante  $t$ . Portanto, as unidades de contexto armazenam as saídas do instante  $t - 1$  para serem usadas como entradas no instante  $t$ . Restando entretanto o caso da primeira execução da rede, onde  $t = 0$ , como para esse instante não existem valores anteriores de saídas armazenados nas unidades de contexto as mesmas tem valor igual a zero.

Desse modo, a saída do neurônio  $k$  da camada oculta de uma rede recorrente de Elman no instante  $t$  é dada pela Equação 4.14. Para os demais neurônios da rede as saídas são calculadas normalmente conforme a Equação 4.1

$$y_k(t) = \varphi \left( \left[ \sum_{i=1}^n x_i(t) w_{ki} \right] + \left[ \sum_{i=1}^N y_k(t-1) w_{ki}^* \right] + b_k \right) \quad (4.14)$$

tal que  $y_k(t)$  é a saída do neurônio  $k$  da camada oculta no instante  $t$ ,  $x_i(t)$  é a entrada  $i$  da rede no instante  $t$ ,  $w_{ki}$  é o peso da entrada  $i$  no neurônio  $k$  da camada oculta,  $N$  é a quantidade de neurônios da camada oculta e  $w_{ki}^*$  é o peso da unidade de contexto  $i$  do neurônio  $k$  da camada oculta.

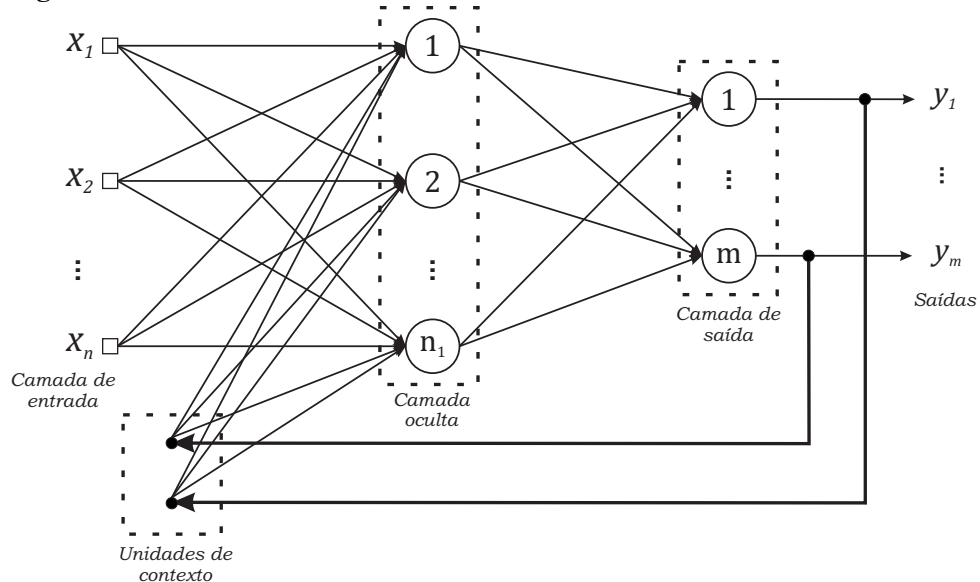
Como as unidades de contexto de uma rede de Elman são tratadas como entradas, os mesmos também têm pesos sinápticos associados a rede e podem ser ajustados da mesma forma que uma MLP, pelo algoritmo *backpropagation*.

#### 4.7 REDE NEURAL DE JORDAN

A partir de uma rede MLP, Jordan (1986) criou a primeira rede neural recorrente. Esta foi inicialmente utilizada para reconhecimento de séries temporais, mas atualmente ela é aplicada a todos os tipos de problemas comumente resolvidos com RNAs.

Diferentemente da rede de Elman apresentada na seção anterior (4.6), as unidades de contexto da rede recorrente de Jordan são alimentadas pelas saídas dos neurônios da camada de saída e não da camada oculta da rede. A Figura 8 ilustra esse modelo.

**Figura 8 – Rede neural de Jordan.**



**Fonte: Adaptado de Silva, Spatti e Flauzino (2010).**

Da mesma forma que na rede de Elman as unidades de contexto guardam as saídas dos neurônios da camada oculta no instante  $t - 1$ , aqui as unidades de contexto armazenam as saídas da camada de saída no instante  $t - 1$  para serem utilizadas como entradas para a camada oculta no instante  $t$ . Assim, a saída dos neurônios da camada oculta é dada pela Equação 4.15

$$y_k(t) = \varphi \left( \left[ \sum_{i=1}^n x_i(t) w_{ki} \right] + \left[ \sum_{i=1}^M y_k(t-1) w_{ki}^* \right] + b_k \right) \quad (4.15)$$

sendo  $y_k(t)$  a saída do neurônio  $k$  da camada oculta no instante  $t$ ,  $x_i(t)$  a entrada  $i$  da rede neural no instante  $t$ ,  $w_{ki}$  é o peso sináptico do neurônio  $k$  associado a entrada  $i$  no instante  $t$ ,  $M$  é o número de neurônios da camada de saída da rede,  $y_k(t-1)$  é a saída do neurônio  $k$  da camada de saída no instante  $t-1$  e  $w_{ki}^*$  é o peso sináptico do neurônio  $k$  associado a unidade de contexto  $i$ .

#### 4.8 MÁQUINAS DESORGANIZADAS

O conceito de máquina desorganizada foi inserido por Alan Turing (1948). Neste trabalho o autor sugeria que o córtex humano infantil era uma "máquina desorganizada". Ele também as definiu como mecanismos aleatórios em sua construção, mas capazes de realizar tarefas particulares mediante treinamento.

Boccatto (2013) descreve que Alan Turing concebeu o conceito de duas redes desorganizadas (ou máquinas desorganizadas), chamadas de rede tipo A e rede tipo B. Também foi proposta uma modificação para a rede tipo B que foi chamada de rede tipo BI, como segue:

- **Rede tipo A** - É a rede desorganizada mais simples proposta por Turing, sendo composta por exatamente duas entradas  $x_1(t)$  e  $x_2(t)$ , um número suficientemente grande de neurônios  $N$  e apenas uma única saída  $y(t)$ . Cada neurônio pode assumir apenas o estado 1 ou 0 em cada instante de tempo  $t$ , de forma que a saída (que é o seu estado) é computada pela função booleana *NAND*. Nessa rede não existem camadas, de modo que cada neurônio pode se ligar a quantos outros forem necessário sem nenhum tipo de padrão ou regra. Logo, o bom funcionamento dessa rede depende da escolha da topologia adequada e do número de neurônios correto.
- **Rede tipo B** - Essa rede segue os mesmos conceitos da rede tipo A, salvo que no meio da ligação entre dois neurônios existe uma rede tipo A contendo apenas 3 elementos, sendo este um tipo de neurônio de interferência. Esse último possibilita que as ligações úteis sejam reforçadas enquanto as inúteis sejam eliminadas. Possuindo apenas 3 comportamentos possíveis, o neurônio de interferência pode: (i) inverter o sinal enviado de um neurônio para o outro; (ii) interromper a transmissão de qualquer informações, mediante o envio de um valor fixo para o neurônio de destino ou (iii) realizar os comportamentos (i) e (ii) de forma alternada.
- **Rede tipo BI** - A última rede desorganizada introduz a possibilidade da intervenção de um agente externo (um professor, similar ao conceito de treinamento supervisionado visto na Seção 4.3). Nessa rede são adicionadas 2 entradas em cada neurônio de interferência, de forma que os valores dessas podem ser alterados por um agente externo que pode ajustar a rede de forma a corrigir o seu funcionamento.

Apesar das redes desorganizadas propostas por Turing apresentarem arquitetura e funcionamento simples elas são capazes de produzir comportamentos bastante complexos, sendo que a rede tipo A é capaz de reproduzir qualquer função lógica. Isso enfatiza as contribuições do autor para as áreas de redes neurais e aprendizado de máquina (BOCCATO et al., 2014).

No trabalho de Boccato (2013), o autor generalizou o conceito de máquinas desorganizadas para as redes em que pelo menos alguns de seus pesos sinápticos não são ajustados, ficando estes estabelecidos de forma aleatória, ou em desordem. A aleatoriedade desses pesos traz implícita a ideia de desorganização, já que parte da rede permanece sem qualquer tipo de ajuste. A seguir descreve-se dois tipos de máquinas desorganizadas, as Máquinas de Aprendizado Extremo na Seção 4.8.1 e as redes neurais com estado de eco na Seção 4.8.2.

#### 4.8.1 Máquinas de Aprendizado Extremo (ELM)

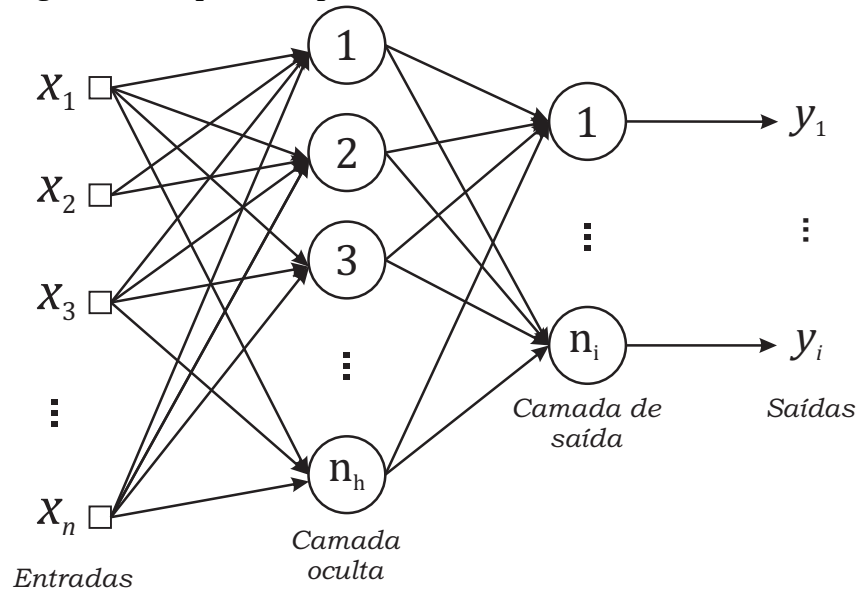
As Máquinas de aprendizado extremo (ELM, do inglês *Extreme Learning Machine*) são um tipo de rede neural *feedforward* com apenas uma camada oculta propostas por Huang, Zhu e Siew (2004). Os pesos sinápticos da camada intermediária são escolhidos de forma aleatória e permanecem sem ajuste. Durante o processo de treinamento apenas os pesos da camada de saída são atualizados. Esta aleatoriedade dos pesos da camada oculta que dá à rede um caráter desorganizado.

Durante a fase de treinamento os valores ótimos para os pesos sinápticos da camada de saída podem ser determinados mediante um método analítico, que consiste na resolução de um problema de regressão linear (HUANG; ZHU; SIEW, 2006). Embora essa não seja a única solução possível, sua utilização é benéfica pois evita o uso de algoritmos iterativos no treinamento da rede. Isto faz com que o custo computacional do treinamento seja baixo quando comparado ao de outras arquiteturas, como a MLP (SIQUEIRA et al., 2012).

No trabalho de Bartlett (1998), o autor apresenta um resultado de extrema importância para a área. Nele o autor conclui que controlar a norma dos pesos sinápticos é mais importante para a capacidade de generalização da rede do que controlar o tamanho da rede. A partir dessa constatação pode-se inferir que o número efetivo de neurônios da camada intermediária será definido pela configuração dos pesos sinápticos da camada de saída. Por sua vez, nas ELMs a garantia da boa capacidade de generalização é dada pelos pesos da camada de saída, de modo que os pesos da intermediária podem então ser definidos de forma aleatória. A representação de uma ELM é apresentada na Figura 9.

Na Figura 9 é possível ver o vetor de entradas denotado por  $\mathbf{x} = [x_1, x_2, x_3, \dots, x_n]$ , os  $n_1$  neurônios da camada oculta,  $m$  neurônios da camada de saída e por fim as saídas da rede  $y_m$ . Apesar de não constar, cada neurônio da camada intermediária possui um valor de bias associado a ele.



**Figura 9 – Máquina de aprendizado extremo (ELM)**

**Fonte: Adaptado de Siqueira (2013).**

Como é usual, as ELMs aceitam quantas saídas e quantas entradas forem necessárias. Sua arquitetura é semelhante a da rede MLP, de modo que no artigo original Huang, Zhu e Siew (2004) afirmam que a ELM é, de fato, um novo algoritmo de treinamento para esta rede. Entretanto, apenas 2 anos mais tarde Huang, Zhu e Siew (2006) justificaram que em função da obrigatoriedade da ELM possuir exatamente 2 camadas neurais ela vem a ser uma nova arquitetura de rede neural e não um algoritmo de treinamento. Nos dias atuais as duas são vistas como propostas distintas, apesar dessa conclusão não ser uma unanimidade entre os pesquisadores da área. Pela Figura 9 é possível notar a semelhança da ELM com o modelo de MLP mostrado na Figura 5.

A função de ativação de uma ELM também pode ser arbitrária desde que a mesma seja infinitamente diferenciável em qualquer intervalo. Na literatura pode-se encontrar propostas utilizando desde funções trigonométricas até a função sinal.

A execução da rede ELM não se diferencia em relação a de outras redes neurais de múltiplas camadas. Primeiramente é necessário computar a saída da camada intermediária para esta ser utilizada como entrada pela camada de saída, similar ao funcionamento das MLPs. O sinal de saída da camada oculta é dado pela Equação 4.16

$$y_k^{hid} = \varphi_k^{hid} \left( \left[ \sum_{i=1}^n x_i w_{ki}^{hid} \right] + b_k^{hid} \right) \quad (4.16)$$

na qual  $y_k^{hid}$  é o sinal de saída do neurônio  $k$  da camada oculta<sup>1</sup> na iteração  $t$ ,  $\varphi_k^{hid}(\cdot)$  é a função de ativação do neurônio  $k$  da camada oculta,  $n$  é o total de entradas da ELM,  $x_i$  é o sinal de entrada,  $w_{ki}^{hid}$  é o peso sináptico e  $b_k^{hid}$  é o valor do bias correspondente.

<sup>1</sup>  $y^{hid}$  – hid vem do inglês *hidden*.

De forma similar, o sinal de saída da camada de saída da ELM é dado pela Equação 4.17

$$y_k^{out} = \varphi_k^{out} \left( \sum_{i=1}^{n_1} y_i^{hid} w_{ki}^{out} \right) \quad (4.17)$$

em que  $y_k^{out}$  é o sinal de saída do neurônio  $k$  da camada de saída<sup>2</sup>,  $\varphi_k^{out}$  é a função de ativação do neurônio  $k$  da camada de saída,  $n_1$  é o total de neurônios da camada oculta,  $y_i^{hid}$  é o sinal de saída do neurônio  $i$  e  $w_{ki}^{out}$  é o peso sináptico do neurônio  $k$ .

#### 4.8.1.1 Treinamento (ELM)

Como já foi dito, o treinamento da ELM é realizado apenas nos pesos sinápticos da camada de saída. Assim, tanto os pesos quanto os bias para a camada intermediária são definidos de forma aleatória com valores entre 0 e 1.

No trabalho de Huang, Zhu e Siew (2006), os autores provaram que uma ELM possui capacidade de aproximação universal, uma vez que o erro de aproximação produzido pela rede pode sempre ser diminuído mediante a adição de um neurônio à camada intermediária, via determinação rigorosa dos pesos da camada de saída. Dessa maneira, os autores sugerem que treinar uma ELM seja equivalente a resolver o problema de otimização apresentado pela Equação 4.18

$$\mathbf{w}_k^* = \arg_{\mathbf{w}_k \in \mathbb{R}^{n_1+1}} \min \|\mathbf{w}_k\|^2 + C_k \times J(\mathbf{w}_k) \quad (4.18)$$

no qual  $k$  é o índice da saída,  $n_1$  é o número de neurônios da camada intermediária,  $\|\cdot\|^2$  é a norma euclidiana,  $C_k$  é um coeficiente de ponderação ou regularização a ser definido por exemplo por métodos de busca unidimensional e  $J(\mathbf{w}_k)$  é dado pela Equação 4.19

$$J(\mathbf{w}_k) = \frac{1}{2} \sum_{l=1}^n \left( \sum_{j=1}^{n_1} w_{kj} f(\mathbf{v}_j, b_j, \mathbf{x}_j) + w_{k0} - S_{kl} \right) \quad (4.19)$$

em que  $n$  é o número de amostras disponíveis para o treinamento.

Considere  $\mathbf{W}^{out}$  como a matriz com os pesos sinápticos de todos os neurônios da camada de saída, construída da seguinte forma:

$$\mathbf{W}^{out} = \begin{bmatrix} w_{1,1}^{out} & w_{1,2}^{out} & \dots & w_{1,i}^{out} \\ w_{2,1}^{out} & w_{2,2}^{out} & \dots & w_{2,i}^{out} \\ \vdots & \ddots & \ddots & \vdots \\ w_{k,1}^{out} & w_{k,2}^{out} & \dots & w_{k,i}^{out} \end{bmatrix}$$

na qual  $w_{ki}^{out}$  representa o peso sináptico da entrada  $i$  do neurônio  $k$  da camada de saída.

<sup>2</sup>  $y^{out}$  – out vem do inglês *output*.

Assim sendo, treinar a ELM consiste em determinar qual a matriz  $\mathbf{W}^{\text{out}}$  com o menor erro para o vetor  $\mathbf{d}$  o qual contém todas as saídas esperadas. Dessa forma, o processo de treinamento da ELM é resumido em resolver o operador pseudo-inversa de *Moore-Penrose* conforme a Equação 4.20

$$\mathbf{W}^{\text{out}} = (\mathbf{X}_{\text{hid}}^{\text{T}} \mathbf{X}_{\text{hid}})^{-1} \mathbf{X}_{\text{hid}}^{\text{T}} \mathbf{d} \quad (4.20)$$

onde  $\mathbf{X}_{\text{hid}}$  é a matriz com as saídas da camada oculta construída da seguinte forma:

$$\mathbf{X}_{\text{hid}} = \begin{bmatrix} y_{1,1}^{\text{hid}} & y_{1,2}^{\text{hid}} & \cdots & y_{1,t}^{\text{hid}} \\ y_{2,1}^{\text{hid}} & y_{2,2}^{\text{hid}} & \cdots & y_{2,t}^{\text{hid}} \\ \vdots & \ddots & \ddots & \vdots \\ y_{k,1}^{\text{hid}} & y_{k,2}^{\text{hid}} & \cdots & y_{k,t}^{\text{hid}} \end{bmatrix}$$

sendo  $y_{k,t}^{\text{hid}}$  a saída do neurônio  $k$  da camada oculta na iteração  $t$ .

Essa solução analítica garante que o treinamento da ELM seja muito mais eficiente computacionalmente. Entretanto, desempenho da rede pode ser melhorado com a utilização de um coeficiente de regularização  $C$  proposto por Huang et al. (2012). Adicionando-o obtêm-se a Equação 4.21

$$\mathbf{W}^{\text{out}} = \left( \frac{1}{C} + \mathbf{X}_{\text{hid}}^{\text{T}} \mathbf{X}_{\text{hid}} \right)^{-1} \mathbf{X}_{\text{hid}}^{\text{T}} \mathbf{d} \quad (4.21)$$

Huang et al. (2012) sugerem que  $C = 2^\lambda$ , em que  $\lambda \in \{-25, -24, \dots, 25, 26\}$ . Para utilizá-lo basta testar todos valores possíveis para  $C$  e escolher aquele que apresentar o menor erro quando comparado com o conjunto de validação. Junto ao conjunto de validação pois esse estudo utilizou a técnica de validação cruzada para evitar o sobre-treinamento, caso o projetista opte por não utilizar tal técnica os valores do Coeficiente de Regularização podem ser testados sobre o próprio conjunto de treinamento.

#### 4.8.2 Redes Neurais Com Estado de Eco (ESN)

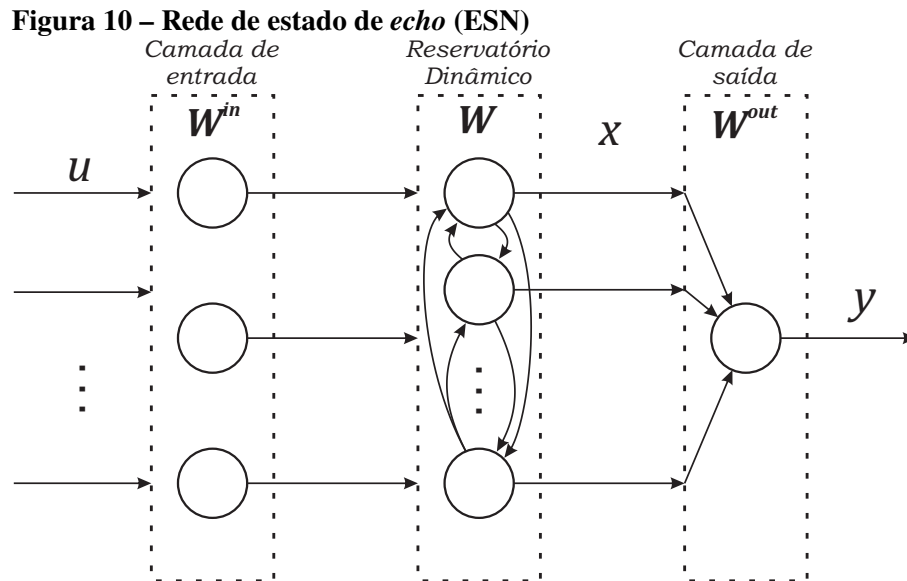
Redes Neurais Recorrentes (RNN, do inglês *Recurrent Neural Network*) possuem laços de realimentação entre os neurônios, que geram uma memória intrínseca capaz de armazenar o histórico dos dados de entrada e repassá-lo em certa medida para as respostas subsequentes. Entretanto, em uma rede neural recorrente clássica, essa inserção de *feedback* pode trazer grandes dificuldades no estabelecimento dos pesos sinápticos já que podem acarretar em convergência lenta, possível convergência para ótimos locais de qualidade ruim, instabilidade e complexidade computacional elevada (HAYKIN, 2001).

Como meio de evitar tais dificuldades Jaeger (2001) propôs as redes neurais com estados de eco (ESN, do inglês *Echo state network*) como sendo um novo tipo de RNN. Diferente das

redes *feedforward*, em que uma mesma entrada sempre resultará na mesma saída, as redes RNN possibilitam saídas diferentes para a mesma entrada, uma vez que esta dependerá do estado em que a rede se encontra. Assim, o nome *estado de eco* foi inspirado no fato de que um estado da rede ESN é baseado no eco dos estados anteriores (HAYKIN, 2008). Ademais, as ESNs também são consideradas aproximadores universais (SCHAEFER; ZIMMERMANN, 2007).

As ESNs possuem 3 (três) camadas: de entrada, a segunda, que é chamada de reservatório de dinâmicas, constituída por neurônios totalmente interconectados, o que gera uma característica não-linear, e a terceira camada, a de saída, responsável por combinar as saídas do reservatório de dinâmicas. Esta última, por sua vez, corresponde a parte linear da rede. Diferente de outras propostas de RNNs, que podem possuir realimentação em qualquer camada, as ESNs propostas por Jaeger (2001) neste trabalho apenas possuem realimentação no reservatório de dinâmicas.

Na Figura 10 é possível ver que os neurônios do reservatório de dinâmicas são conectados entre si, enquanto os neurônios da camada de saída combinam o resultado do reservatório de dinâmicas e fornecem a saída da rede.



**Fonte:** Adaptado de Siqueira (2013)

Nota-se pela Figura 10 que as entradas da rede fazem parte do vetor de entradas  $\mathbf{u}_t = [u_1, u_2, u_3, \dots, u_{t-K+1}]^T$ . Por sua vez os coeficientes da camada de entrada  $\mathbf{W}^{in} \in \mathbb{R}^{N \times K}$  ponderam linearmente as entradas e as passam ao reservatório de unidades não-lineares totalmente interconectadas  $\mathbf{W} \in \mathbb{R}^{N \times N}$ . O reservatório então gera as ativações  $\mathbf{x}_t = [x_t^1, x_t^2, x_t^3, \dots, x_t^N]^T$  que são os estados da rede. Estes são atualizados a cada nova entrada, sendo que os novos estados para a entrada  $t + 1$  são dados pela Equação 4.22 (OZTURK; XU; PRÍNCIPE, 2007).

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{W}^{in}\mathbf{u}_{t+1} + \mathbf{W}\mathbf{x}_t) \quad (4.22)$$

na qual  $\mathbf{x}_{t+1}$  são os estados na entrada  $t + 1$ ,  $\mathbf{f}(\cdot)$  representa as ativações dos neurônios do

reservatório  $f(\cdot) = (f_1(\cdot), f_2(\cdot), f_3(\cdot), \dots, f_N(\cdot))$ ,  $K$  representa o número de entradas e  $N$  é o número de neurônios no reservatório. Convencionou-se que quando  $t = 0$  os estados da rede são nulos (SIQUEIRA, 2013).

Por sua vez, o vetor de saída da rede  $\mathbf{y}_{t+1}$  é dado pela Equação 4.23

$$\mathbf{y}_{t+1} = \mathbf{W}^{\text{out}} \mathbf{x}_{t+1} \quad (4.23)$$

sendo  $\mathbf{W}^{\text{out}} \in \mathbb{R}^{L \times N}$  a matriz com os pesos sinápticos da camada de saída e  $L$  o número de saídas da rede.

Assim como nas ELMs, os pesos sinápticos do reservatório de dinâmicas das ESN não são ajustados durante o treinamento. Também se faz uso do operador pseudo-inversa de *Moore-Penrose* para realizar o ajuste da camada de saída a partir dos estados de eco  $\mathbf{X} \in \mathbb{R}^{K \times N}$ , conforme a Equação 4.24

$$\mathbf{W}^{\text{out}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{d} \quad (4.24)$$

onde  $\mathbf{d} \in \mathbb{R}^{L \times K}$  é o sinal desejado para a saída da rede.

Vale ressaltar que assim como ocorre nas ELMs, os pesos da ESN também podem ser melhorados com a adição do Coeficiente de Regularização. Desse modo os pesos da camada de saída de uma ESN também podem ser obtidos através da Equação 4.21.

#### 4.8.2.1 Propriedade de estado de eco

Jaeger (2001) comprovou a propriedade de estado de eco (PEC) constando que os estados  $\mathbf{x}_t$  são assintoticamente independentes da condição inicial. Isso significa que os mesmos valores de entrada não resultam nos mesmos sinais de saída, mas uma mesma sequência temporal de entradas converge sempre a sinais de saída similares, de modo que a rede possui estados de eco.

Em seu trabalho Jaeger (2001) ainda definiu 2 condições para existência dessa propriedade:

1. Os sinais de entrada devem ser extraídos de um espaço compacto  $B$ ;
2. Os estados da rede devem sempre estar contidos dentro de um conjunto compacto de estados admissíveis  $A \subset \mathbb{R}^N$ .

Se essas condições forem satisfeitas, o estado do reservatório irá depender assintoticamente somente do histórico de entrada e a rede será considerada uma rede com estados de eco (JAEGER, 2003).

#### 4.8.2.2 Construção do reservatório de dinâmicas

Como visto na seção anterior (4.8.2.1), o reservatório de dinâmicas de uma ESN deve necessariamente obedecer a propriedade de estado de eco. Na literatura existem 2 propostas principais para sua construção, de modo que este obedeça a PEC e, conseqüentemente, a rede funcione de forma correta.

A primeira proposta é a de Jaeger (2001). De uma maneira simples o reservatório é criado a partir de uma matriz de pesos com certo grau de esparsidade. Posteriormente foi provado que um padrão esparsa de conexões favorece o desacoplamento dos grupos de neurônios, induzindo o desenvolvimento de dinâmicas individuais e pouco relacionadas.

Jaeger descreveu várias formas de criar a matriz de pesos, sendo que a principal delas define 3 valores possíveis. Para decidir qual será seu valor é feito um sorteio seguindo as probabilidades descritas na Equação 4.25.

$$W_{ki}^{in} = \begin{cases} 0,4 & \text{com probabilidade de } 0,025 \\ -0,4 & \text{com probabilidade de } 0,025 \\ 0 & \text{com probabilidade de } 0,95 \end{cases} \quad (4.25)$$

A segunda proposta foi feita por Ozturk, Xu e Príncipe (2007). Nesta, o objetivo dos autores foi elaborar um reservatório rico do ponto de vista da entropia média dos estados de eco. O resultado foi uma estratégia em que os autovalores respeitam uma distribuição uniforme no círculo unitário, criando uma matriz canônica como mostra a Equação 4.26.

$$W_{ki}^{in} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -r^N \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (4.26)$$

na qual  $r$  é o raio espectral unitário, e  $N$  é a quantidade de neurônios presente no reservatório de dinâmicas. Vale dizer aqui que  $r$  é um parâmetro livre no intervalo  $[0, 1]$ , portanto seu valor varia de problema para problema e deve ser determinado antes que a rede possa ser treinada.

## 4.9 METODOLOGIA *ENSEMBLE*

Um *Ensemble*, ou Máquina de Comitê, combina o resultado de várias redes neurais treinadas individualmente com o objetivo de melhorar a resposta final do sistema. A ideia por trás dessa metodologia é que redes neurais diferentes entre si (seja pela arquitetura, número de

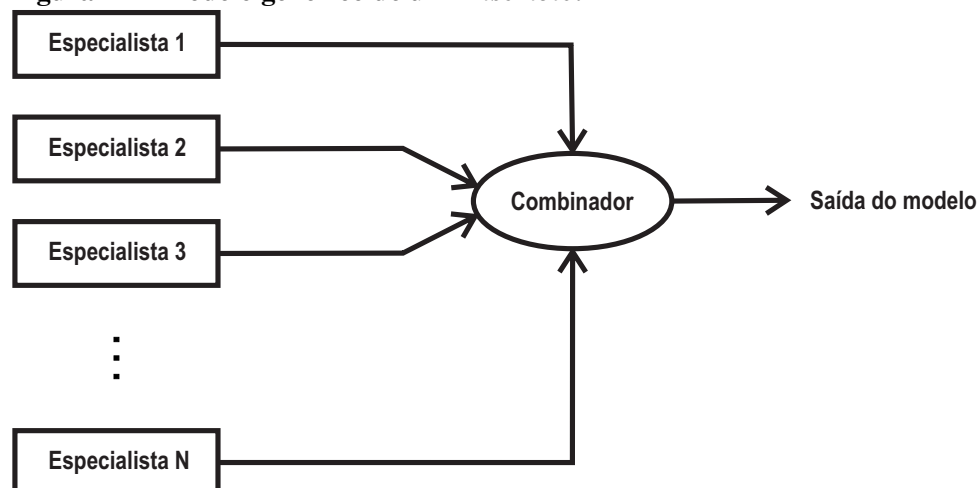
neurônios ou método de treinamento) produzem resultados diferentes quando aplicadas as mesmas entradas, de modo que cada uma pode apresentar melhores respostas para uma determinada faixa dos dados, enquanto outra se sai melhor em outra faixa. De posse das saídas de todas essas redes, também chamadas de especialistas, algum método de combinação (média, votação ou ainda uma outra rede neural) é aplicado para produzir a saída final do *Ensemble*.

A ideia de *Ensemble*, remete ao trabalho de Nilsson (1965). A estrutura proposta pelo autor compreende uma primeira camada de redes Perceptron<sup>3</sup> simples, que tem suas saídas combinadas por um último Perceptron de votação.

Hansen e Salamon (1990) mostraram que a combinação de diferentes Redes Neurais Artificiais treinadas separadamente tende a apresentar uma melhora significativa na generalização do sistema, superando o resultado da melhor RNA treinada individualmente. Em virtude dos bons resultados apresentados, os *Ensembles* foram utilizados na tentativa de resolução de diversos problemas em várias áreas, como reconhecimento de face (LU et al., 2006), reconhecimento de caracteres (BERTOLAMI; BUNKE, 2005), aproximação de funções (LIMA; COELHO; ZUBEN, 2002) e previsão de séries temporais (WICHARD; OGORZALEK, 2004).

Na Figura 11 é apresentado um modelo genérico de *Ensemble*. É possível verificar que as saídas de todos os especialistas, que foram devidamente treinados para o problema, são combinadas a fim de melhorar a solução geral.

**Figura 11 – Modelo genérico de um *Ensemble*.**



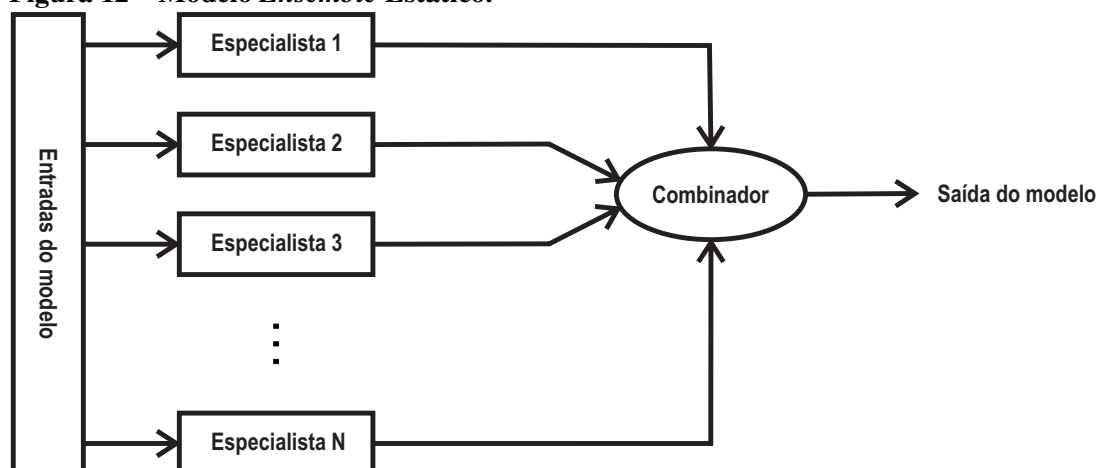
**Fonte: Autoria própria.**

Segundo Haykin (2001) *Ensembles* são aproximadores universais e podem ser divididos em duas categorias:

- **Estruturas estáticas** – Nessa categoria, as respostas de vários especialistas, ou previsores, são combinadas por um mecanismo que não tem acesso às entradas do modelo geral. Desse modo, o combinador tem acesso apenas às saídas dos especialistas (HAYKIN, 2001). A Figura 12 mostra a estrutura de um *Ensemble* estático.

<sup>3</sup> Diferentemente da MLP, a rede Perceptron possui apenas uma camada neural, idealizada por Rosenblatt (1958)

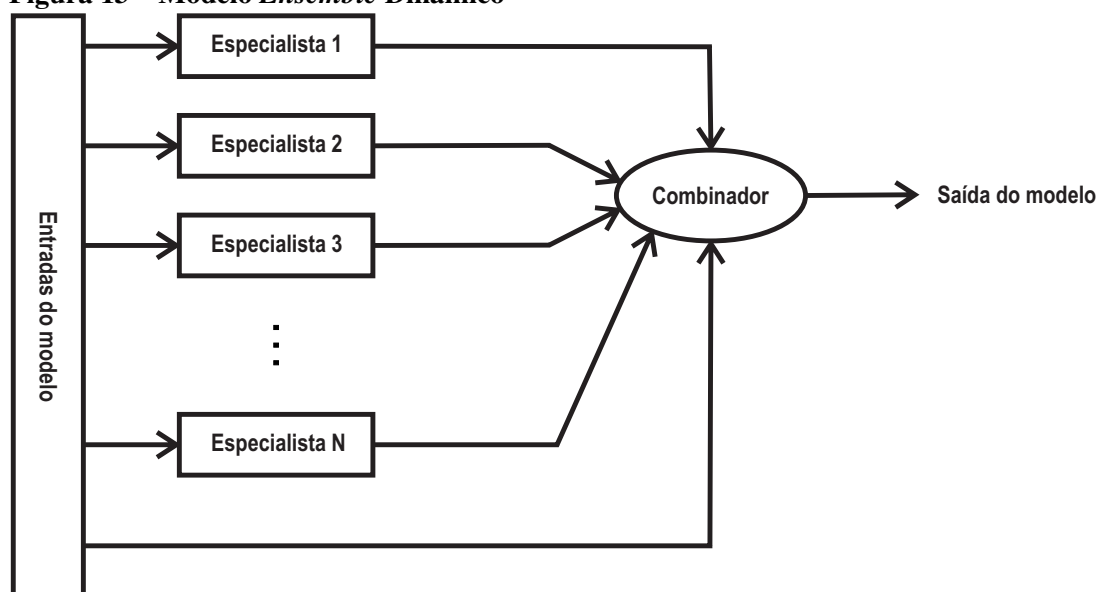
**Figura 12 – Modelo *Ensemble* Estático.**



Fonte: Autoria própria.

- **Estruturas dinâmicas** – Nesta segunda categoria as entradas do modelo são aplicadas aos especialistas que compõe o *Ensemble* como no caso das Estruturas estáticas, mas também ao combinador. Este deve combinar as respostas dadas pelos especialistas com as respectivas entradas do problema (HAYKIN, 2001). Na Figura 13 é apresentado um *Ensemble* com estrutura dinâmica.

**Figura 13 – Modelo *Ensemble* Dinâmico**



Fonte: Autoria própria.

Vale destacar que a combinação de vários especialistas por meio de um *Ensemble* não exclui a necessidade de que os preditores apresentem bons resultados individuais. O objetivo de uma metodologia *Ensemble* é melhorar os bons resultados já existentes, de modo que a condição essencial para que ele seja mais preciso é que seus modelos sejam precisos e diversos (HANSEN; SALAMON, 1990).

---

é a forma mais simples de configuração de uma rede neural artificial.



Tendo isso em vista, para que o resultado de saída do *Ensemble* seja melhor que os individuais é necessário que haja diversidade de erro nos seus componentes. Cada um deve apresentar um bom resultado individualmente e quando comparados os erros devem ser significativamente distintos. É de extrema importância que os componentes apresentem erros em partes diferentes do problema (PERRONE, 1993).

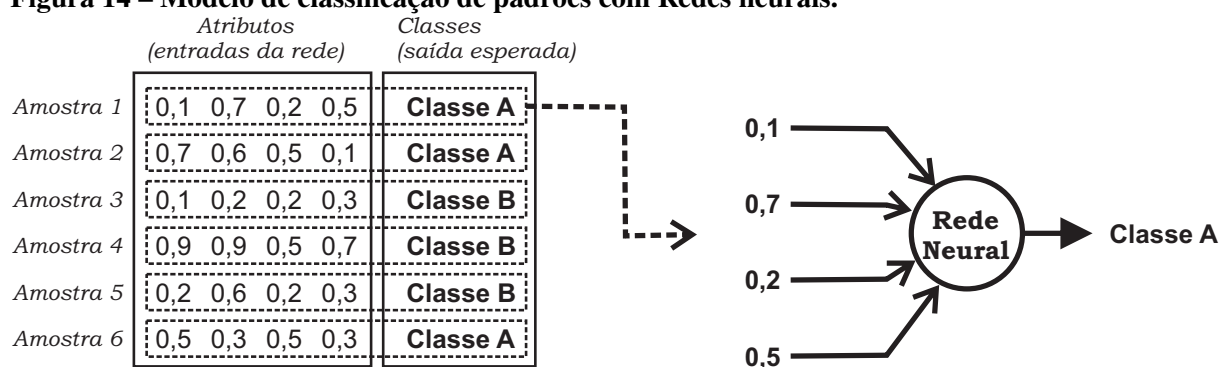
Essa necessidade de diversidade é de certa forma intuitiva, visto que caso todos apresentem a mesma saída não haverá espaço para que o combinador evidencie uma componente específica em determinadas entradas, uma vez que, combinadores com os mesmos erros implicam que os mesmos também tem os mesmos acertos.

Em termos de diversidade de erro, não se faz referência à taxa de acurácia dos especialistas do *Ensemble*. Diversidade aqui significa que diferentes componentes erram em diferentes partes do problema. Assim, para cada parte pode-se elencar o especialista com menor erro, a fim de favorecer sua resposta em relação às demais componentes.

#### 4.10 REDES NEURAIAS PARA PREVISÃO DE SÉRIES TEMPORAIS

Como discutido nas seções anteriores, as Redes Neurais Artificiais são metodologias bastante robustas e adequadas para resolver diferentes tipos de problemas reais. Um dos exemplos de sua generalidade pode ser a aplicação na classificação de padrões. Suponha-se que se queira classificar um grupo de pessoas de acordo com certos atributos físicos. Neste caso, tem-se que cada amostra de treinamento é composta pelas características observadas no indivíduo representado por essa amostra e pela sua respectiva classe. A Figura 14 apresenta um modelo do funcionamento de uma rede neural como classificadora de padrões.

**Figura 14 – Modelo de classificação de padrões com Redes neurais.**



**Fonte: Autoria própria.**

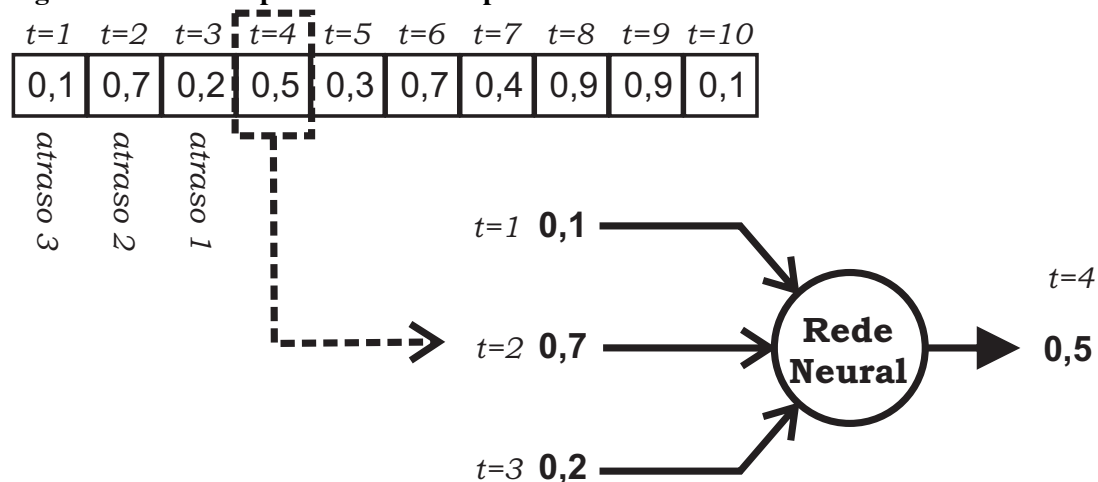
Na Figura 14 fica evidente que os atributos contidos nas amostras são as entradas da rede, enquanto que a saída é a classe correspondente.

Por sua vez, em uma série temporal não existem atributos que se relacionam com o elemento da série. O que há são os próprios dados da série que se deseja prever. Neste contexto,

ambas as aplicações são distintas, de modo que para o segundo caso certas adaptações devem ser feitas na forma de inserção dos dados de entrada.

A definição de série temporal implica na relação de um dado da série com os seus elementos anteriores, de modo que o valor de uma amostra  $x_t$  é dependente do valor das amostras passadas  $x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-l}$  em que  $l$  é chamado de atraso. Em virtude dessa relação, são os atrasos que são utilizados como entrada. A Figura 15 ilustra esta ideia.

**Figura 15 – Modelo previsão série temporal com Rede neural.**



**Fonte: Autoria própria.**

Como pode ser visto na Figura 15 para prever o elemento  $t = 4$  foram utilizados como entradas no exemplo os elementos  $t = 1, t = 2$  e  $t = 3$ . Nesse caso, diz-se que a previsão foi realizada com 3 atrasos.

Observa-se que a quantidade de atrasos e quais deles serão utilizados é um parâmetro livre, o qual deve ser ajustado de acordo com cada série. Vale lembrar que, os melhores atrasos podem ser distintos para redes distintas com diferentes quantidades de neurônios. Para a mesma série temporal, a previsão com 20 neurônios pode alcançar um melhor resultado com 10 atrasos enquanto a mesma série, com a mesma rede, mas agora com 30 neurônios pode alcançar o melhor resultado com 3 atrasos.

No exemplo da Figura 15 verifica-se que a previsão foi realizada para o dado imediatamente posterior aos atrasos selecionados. Entretanto, em situações reais, muitas vezes é necessário prever não só os valores de amanhã, mas também dos próximos dias ou semanas. A esse alvo de previsão (amanhã, daqui 3 dias, semana que vem, etc) chama-se previsão multi-passos. Se é desejável prever o próximo dado da série, a previsão é dita 1 passo a frente. Mas, se o alvo for prever o 4º elemento adiante, a previsão é dita 4 passos a frente e assim sucessivamente.

Considere, como exemplo, que uma rede neural está sendo utilizada para realizar previsões para a série de preços de uma ação na bolsa de valores e que tem-se a disposição os preços diários da mesma até a data de hoje. Entretanto, neste exemplo, interessa realizar uma previsão para a semana que vem, ou seja, 7 passos adiante. Considerando que a previsão será realizada com 3 atrasos, como prever 7 passos a frente?

Uma alternativa é a utilização da técnica de previsão recursiva para  $P$  passos. Inicialmente são utilizados os dados existentes para ajustar a rede para prever 1 passo a frente. Em seguida, o resultado desta é inserido na série como se fosse um valor real e não uma previsão. Com esse novo valor agora é possível prever 2 passos adiante. Esse procedimento é repetido até que a rede seja capaz de prever o horizonte necessário (SIQUEIRA, 2013).

A seguir, vamos discutir brevemente uma forma relevante para selecionar os atrasos mais adequados de previsão.

#### 4.10.1 Método *Wrapper* Para Seleção de Entradas

Em trabalhos de previsão, um tópico de grande relevância é determinar quais os melhores atrasos devem ser utilizados para realização da tarefa, de modo a alcançar desempenho máximo (SIQUEIRA, 2013). Diferente do senso comum, inserir uma grande quantidade de atrasos nem sempre leva a menores erros (GUYON; ELISSEEFF, 2003).

Contudo, testar todas as possíveis combinações é uma tarefa que demanda muito esforço computacional, o que pode tornar o treinamento da rede uma tarefa extremamente demorada ou até inviável. O método *Wrapper* com seleção progressiva é uma forma de contornar esse problema e que já apresentou bons resultados na literatura (SIQUEIRA, 2013). O Pseudocódigo 1 apresenta a sequência de passos a ser seguida na execução do método.

---

#### Pseudocódigo 1: MÉTODO WRAPPER.

---

```

1 início
2   inicie a lista  $c$  com todas as possíveis entradas
3   inicie  $S$  e  $s$  como vazio
4   repita
5      $e_{nova} \leftarrow c[0]$ 
6     para cada entrada  $e$  em  $c$  faça
7       se o erro com as entradas  $s \cup e$  for menor que com  $s \cup e_{nova}$  então
8          $e_{nova} \leftarrow e$ 
9       fim
10    fim
11    remova  $e_{nova}$  de  $c$ 
12    adicione  $e_{nova}$  em  $s$ 
13    se o erro com as entradas  $s$  for menor que com  $S$  então
14       $S \leftarrow s$ 
15    fim
16  até  $|c| == 0$ ;
17  retorne  $S$ 
18 fim

```

---

Fonte: Autoria própria.

Descrevendo o Pseudocódigo 1 de forma mais genérica, em primeiro lugar é necessá-

rio que o projetista defina o número máximo de entradas  $N$ . O procedimento é iniciado com um subconjunto de entradas  $s$  (atrasos) vazio. Uma a uma, cada variável é inserida na rede separadamente, como um esquema de entrada única. Aquela que levar ao melhor desempenho é selecionada e passa a compôr o subconjunto  $s$  e o erro correspondente é armazenado. O passo a seguir é escolher o segundo atraso. A rede, então, será agora municiada com duas entradas, a inicialmente escolhida e uma a uma serão inseridas as restantes. Novamente, a segunda entrada que combinada à primeira leve ao menor erro é selecionada e o erro de saída guardado. Esta metodologia é repetida até que todas as entradas sejam fixadas ao subconjunto. Ao final, aquele subconjunto  $s$  que durante o processo tiver levado ao menor erro geral será o escolhido (GUYON; ELISSEEFF, 2003).

Ainda observando o Pseudocódigo 1 fica evidente que o método *Wrapper* não considera todas as possibilidades para cada tamanho de entrada, conforme a quantidade de entradas aumenta menos possibilidades são testadas. Para entradas de tamanho 1 são consideradas  $N$  possibilidades, para entradas de tamanho 2 são consideradas  $N - 1$ , para tamanho 3 são  $N - 2$  possibilidades, assim sucessivamente até considerar apenas 1 possibilidade para entradas de tamanho  $N$ . Isso dá ao método um custo computacional de ordem  $O(N^2)$ , enquanto que verificar todas as possibilidades teria custo de ordem  $O(N!)$ .

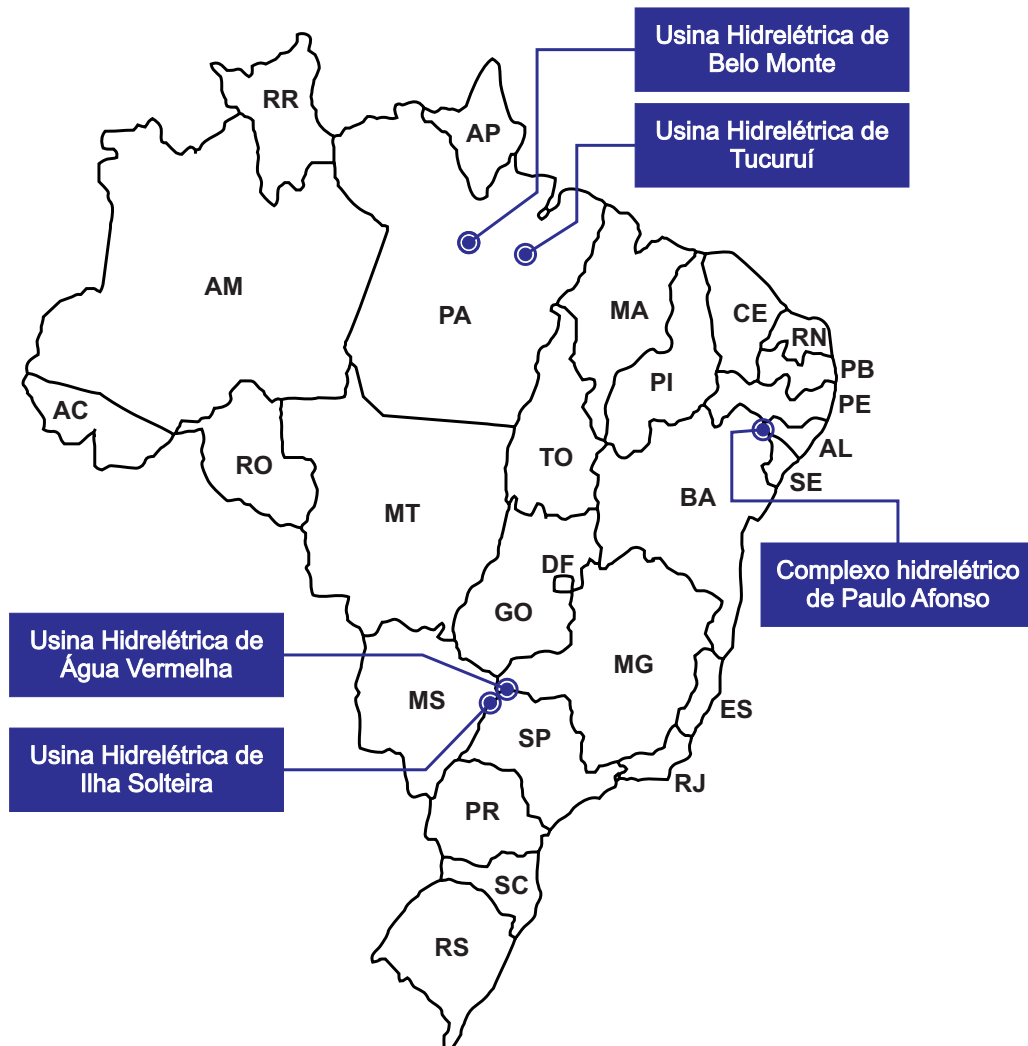
Vale lembrar que como o *Wrapper* com seleção progressiva não verifica todas as possibilidades de configuração para o conjunto de entradas da rede, portanto não garante que a configuração escolhida seja a ótima (GUYON; ELISSEEFF, 2003; SIQUEIRA, 2013).

## 5 ESTUDO DE CASOS (PREVISÃO DE SÉRIES DE VAZÕES)

O presente capítulo discute a aplicação das metodologias de previsão descritas nas séries de vazões afluentes à usinas hidrelétricas. Com o objetivo de estimular a pesquisa a respeito na área e permitir que outros agentes tenham acesso aos dados, o Operador Nacional do Sistema Elétrico (ONS) disponibiliza em seu site os dados históricos das séries de vazões médias mensais e diárias referentes a todas as usinas hidrelétricas brasileiras, juntamente com um relatório sobre a atualização das mesmas (ONS, 2018).

As previsões realizadas nesse estudo correspondem às séries de vazões mensais das usinas de Água Vermelha, Belo Monte, Ilha Solteira, Paulo Afonso e Tucuruí. Foram escolhidas essas 5 usinas pois cada uma delas possui níveis de vazões afluentes diferentes, o que possibilita uma análise mais robusta dos desempenhos dos modelos propostos. Além disso, elas estão localizadas em diferentes regiões do Brasil, como mostra a Figura 16.

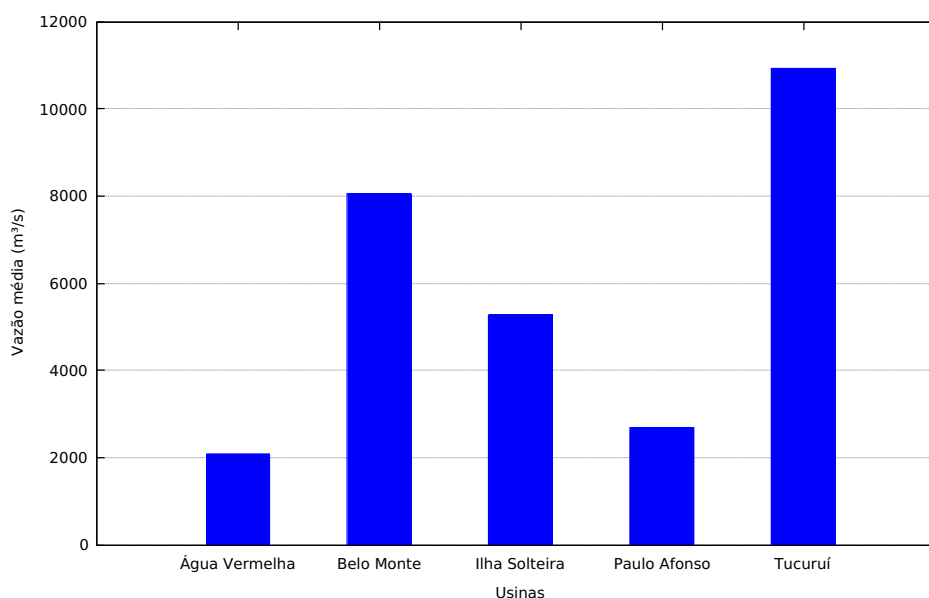
**Figura 16 – Localização das usinas hidrelétricas no mapa brasileiro.**



Fonte: Autoria própria.

Todas as 5 séries possuem dados de Janeiro de 1931 a Dezembro de 2015, totalizando 1020 amostras divididas em 85 anos. No Gráfico 11 são apresentadas as médias das vazões mensais de cada usina. É possível conferir que os níveis de cada planta tem valores variados.

**Gráfico 11 – Vazões médias das usinas hidrelétricas.**



**Fonte: Autoria própria.**

Ainda observando o Gráfico 11, nota-se que a usina com a menor média é a de Água Vermelha, seguida de Paulo Afonso e Ilha Solteira. Por fim as com os maiores níveis são Belo Monte e Tucuruí, respectivamente. A série proveniente de Água Vermelha é uma das que apresentam comportamento mais difícil de ser mapeado, o que tornou-a uma espécie de prova de qualidade para qualquer modelo de previsão desenvolvido.

Para aplicação dos modelos, os dados foram separados em três conjuntos, como segue:

- **Treinamento** – Dados entre os anos de 1931 e 1995, totalizando 65 anos. Esse conjunto é utilizado para o ajuste dos modelos, de modo que os pesos sinápticos das redes neurais e os coeficientes dos modelos lineares são definidos em função dele;
- **Validação** – Amostras compreendidas entre os anos de 1996 e 2005, totalizando 10 anos. Como mencionado na Seção 4.3.2 ao treinar uma rede neural existe a possibilidade do modelo ficar sobre-treinado. Para evitar esse problema foi aplicada a técnica de validação cruzada aos modelos utilizando este conjunto;
- **Teste** – Dados entre os anos de 2006 e 2015, totalizando 10 anos. Após o processo de treinamento ter sido realizado, cada modelo desenvolvido foi aplicado ao conjunto de teste, para o qual o modelo não teve acesso anteriormente, a fim de calcular as métricas de erro escolhidas para determinar a qualidade de cada resposta.

Como medida de desempenho foi utilizado o Erro Quadrático Médio (MSE, do inglês *Mean Square Error*) conforme a Equação 5.1

$$MSE = \frac{1}{N} \sum_{t=1}^N (d_t - y_t)^2 \quad (5.1)$$

tal que,  $N$  é o total de amostras,  $d_t$  é o valor real observado, nesse caso a vazão real medida na usina hidrelétrica, e  $y_t$  é o valor previsto pela rede neural ou pelo modelo linear.

O MSE é uma medida muito utilizada na literatura para comparar a eficiência de estimadores diferentes. Uma das suas características é uma penalização mais forte para erros maiores, enquanto que erros menores não tem influência tão grande no valor final. Desse modo o MSE penaliza mais um modelo que apresente poucos erros grandes do que um modelo com muitos erros pequenos (MORETTIN; TOLOI, 2006).

Além do MSE, também é apresentado o Erro Absoluto Médio (MAE, do inglês *Mean Absolute Error*) conforme a Equação 5.2. Esta métrica indica o afastamento dos valores observados em relação aos valores previstos penalizando igualmente erros pequenos e grandes (MORETTIN; TOLOI, 2006).

$$MAE = \frac{1}{N} \sum_{t=1}^N |d_t - y_t| \quad (5.2)$$

A última medida de erro calculada é o Erro Percentual Absoluto Médio (MAPE, do inglês *Mean Absolute Percentual Error*) que, como o próprio nome já diz, apresenta uma média percentual<sup>1</sup> dos erros cometidos pelo modelo para cada amostra. A Equação 5.3 descreve matematicamente o MAPE.

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{d_t - y_t}{d_t} \right| \quad (5.3)$$

No que diz respeito às entradas utilizadas, seguindo a metodologia usualmente empregada pelo o Setor Elétrico Brasileiro (SEB), foram utilizados até 6 atrasos para todas as previsões realizadas (SIQUEIRA, 2013). Além disso, foram realizadas previsões recursivas para os horizontes de previsão  $P = 1$  (próximo mês),  $P = 3$  (próxima estação),  $P = 6$  (próximo semestre) e  $P = 12$  (próximo ano). O método *Wrapper* com seleção progressiva foi o escolhido para determinar quais entradas devem ser realmente utilizadas por cada modelo.

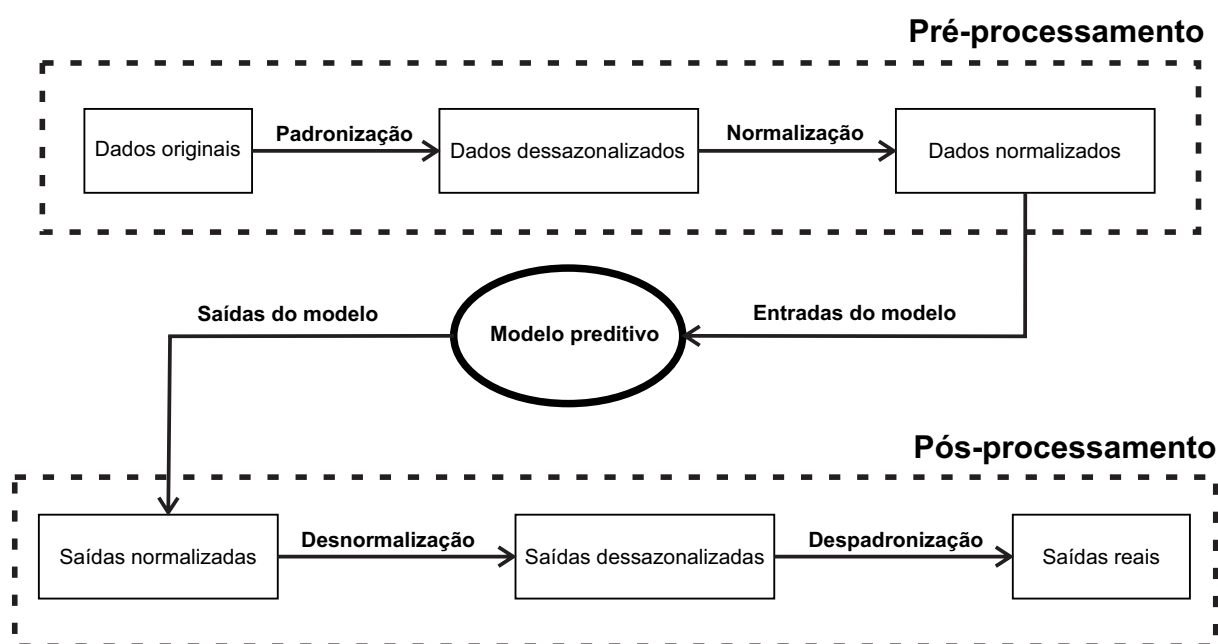
Como mencionado na Seção 2.4, séries de vazões médias mensais apresentam um comportamento sazonal, de modo que, para que um modelo preditivo possa alcançar um melhor desempenho é necessário remover a sazonalidade. Assim, foi aplicado o método da padronização descrito pela Equação 2.13 para as 5 séries selecionadas.

<sup>1</sup> Vale observar aqui que os valores obtidos pela Equação 5.3 são compreendidos entre o intervalo  $[0, 1]$ , portanto, para se obter o valor percentual é necessário multiplicar esse valor por 100.

Antes dos dados serem inseridos nos modelos ainda existe uma outra etapa a ser cumprida, a normalização. Como discutido na Seção 4.3 valores de entradas com magnitudes muito maiores que os dos limites da saída de uma rede neural podem gerar resultados incoerentes, visto que eles geram uma tendência da resposta da rede sempre ser o maior valor possível dentro da sua respectiva função de ativação. No caso das redes neurais, observa-se que, como descrito na Seção 4.3 os dados de entrada de uma rede neural devem ser normalizados a fim de ficarem na mesma magnitude da função de ativação antes de serem inseridos na rede. Desse modo, após a série ser padronizada todos os valores foram normalizados de acordo com a Equação 4.7 antes de serem apresentados aos modelos preditivos.

Por fim, após as previsões serem realizadas, é necessário desnormalizar o valor previsto que pertencerá ao intervalo da função de ativação da rede neural e reinserir a componente sazonal para que enfim possa-se obter o valor real previsto. Essas etapas antes e depois da previsão recebem o nome de Pré-processamento e Pós-processamento respectivamente. A Figura 17 mostra detalhadamente todo o processo.

**Figura 17 – Etapas de pré e pós-processamento.**



Fonte: Autoria própria.

## 5.1 MODELOS DE PREVISÃO ABORDADOS

Nesse trabalho foram considerados no total 18 modelos de previsão, dentre lineares e não lineares. Estes que foram aplicados ao problema da previsão de vazões médias mensais afluentes a usinas hidrelétricas, sendo 2 métodos lineares da família Box & Jenkins, 10 Redes Neurais Artificiais e 6 *Ensembles*.



As ordens dos modelos lineares foram definidas através de tentativa e erro. Foram testadas todas as possibilidades de ordens, tendo sido escolhida aquela com menor MSE junto ao conjunto de validação<sup>2</sup>. Ainda com relação aos modelos AR e ARMA, foi mencionado na Seção 3.1 que os mesmos somente podem ser aplicados em séries temporais estacionárias. Como todas as séries testadas foram dessazonalizadas através da Padronização (Seção 2.4.1), as mesmas agora são (fracamente) estacionárias e portanto podem ser utilizadas de forma direta.

No que diz respeito às redes neurais e aos *Ensembles*, foi aplicada a validação cruzada a fim de evitar um sobre-treinamento, conforme discutido na Seção 4.3.2. Para todas as redes neurais foi considerada a função Tangente Hiperbólica descrita na Seção 4.1.1, com  $\beta = 1$ . Os treinamentos iterativos foram realizados considerando como critério de parada uma melhora no MSE menor que  $10^{-6}$  ou um número máximo de 2000 épocas. Além disso, todas as redes foram testadas quanto ao número de neurônios, sendo treinadas de 5 até 200, com incremento de 5 neurônios. Todas as redes que comportam o Coeficiente de Regularização foram testadas com todas as possibilidades e escolhido novamente aquele com menor MSE junto ao conjunto de validação. Foram consideradas como entradas candidatas para os *Ensembles* todos os 12 demais modelos de previsão desenvolvidos, os modelos lineares e as Redes Neurais Artificiais, sendo aplicado novamente o método *wrapper* para selecionar quais especialistas resultam nas melhores previsões.

Todos os modelos de previsão propostos por este trabalho foram desenvolvidos na linguagem de programação Java. A configuração de cada um é:

- **AR( $q$ )** – Modelo Auto-regressivo de ordem  $q$ ;
- **ARMA( $p, q$ )** – Modelo Auto-regressivo de Médias Móveis de ordem  $p, q$ ;
- **MLP** – Rede neural *Perceptron* de Múltiplas Camadas com apenas uma camada intermediária e treinada pelo algoritmo gradiente descendente e derivadas calculadas via *back-propagation*. Além disso, utiliza-se taxa de aprendizagem de 0,1 e sem a utilização de momentum;
- **RBF** – Rede neural de função de base radial com a camada de saída treinada pelo algoritmo gradiente descendente e *backpropagation*, com taxa de aprendizagem de 0,1 e as gaussianas da camada intermediária ajustadas pelo algoritmo *K-Means*;
- **ELM** – Máquina de aprendizado extremo sem a utilização do Coeficiente de Regularização;
- **ELM (CR)** – Máquina de aprendizado extremo com a utilização do Coeficiente de Regularização;
- **Elman** – Rede recorrente de Elman treinada pelo algoritmo gradiente descendente e *back-propagation* e com taxa de aprendizagem de 0,1;

<sup>2</sup> Foi escolhido o conjunto de validação para ajuste de parâmetros livres devido a aplicação da validação cruzada.

- **Jordan** – Rede recorrente de Jordan também treinada com o algoritmo gradiente descendente e *backpropagation* e com taxa de aprendizagem de 0,1;
- **ESN Jaeger** – Rede neural com estado de eco com reservatório de dinâmicas de Jaeger (2001) e sem Coeficiente de Regularização;
- **ESN Jaeger (CR)** – Rede neural com estado de eco e reservatório de dinâmicas de Jaeger (2001), mas com Coeficiente de Regularização;
- **ESN Ozturk** – Rede neural com estado de eco, reservatório de dinâmicas proposto por Ozturk, Xu e Príncipe (2007) e sem Coeficiente de Regularização;
- **ESN Ozturk (CR)** – Rede neural com estado de eco, reservatório de dinâmicas de Ozturk, Xu e Príncipe (2007) e com Coeficiente de Regularização;
- **Ensemble Média** – *Ensemble* com um combinador que realiza uma média aritmética das entradas;
- **Ensemble Mediana** – *Ensemble* com combinador que devolve como saída o valor da mediana das entradas;
- **Ensemble MLP** – *Ensemble* cujo combinador é um *Perceptron* de Múltiplas Camadas com apenas uma camada oculta. O combinador MLP é treinado pelo algoritmo gradiente descendente e *backpropagation*, com taxa de aprendizagem de 0,1 e sem a utilização de momentum;
- **Ensemble RBF** – *Ensemble* com combinador de RBF, treinada da mesma forma que o modelo RBF anterior;
- **Ensemble ELM** – *Ensemble* com combinador ELM sem Coeficiente de Regularização;
- **Ensemble ELM (CR)** – *Ensemble* com combinador ELM e Coeficiente de Regularização.

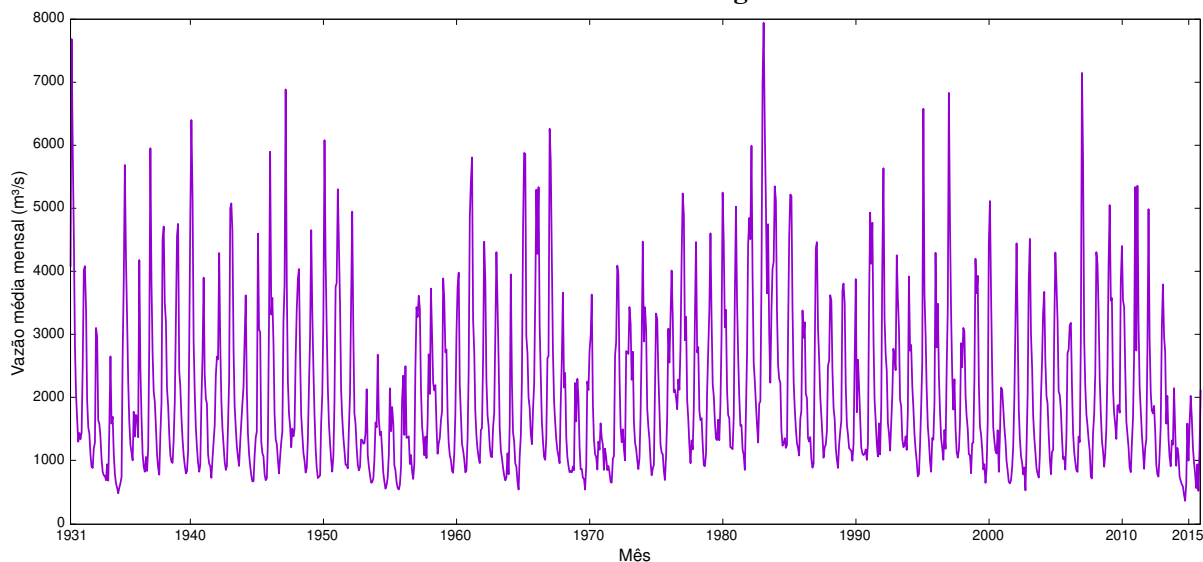
Com o objetivo de obter dados amostrais, para cada configuração de entradas e cada quantidade de neurônios todos os modelos propostos foram executados 30 vezes, tendo sido escolhido a melhor execução, ou seja, aquele com menor MSE junto ao conjunto de teste.

## 5.2 USINA HIDRELÉTRICA DE ÁGUA VERMELHA

Abastecida pelo Rio Grande, a usina hidrelétrica de Água Vermelha fica localizada entre os municípios de Ouroeste e Iturama, na divisa dos estados de São Paulo e Minas Gerais. Sua série histórica apresenta as vazões médias mensais de janeiro de 1931 a dezembro de 2015, compreendendo 1020 observações distribuídas entre 85 anos. Apresenta média histórica de  $\hat{\mu} =$

2077,3284  $m^3/s$  e desvio padrão  $\hat{\sigma} = 1295,7097 m^3/s$ , calculados de acordo com as equações 2.3 e 2.4, respectivamente. Todas as 1020 observações contidas na série são apresentadas no Gráfico 12.

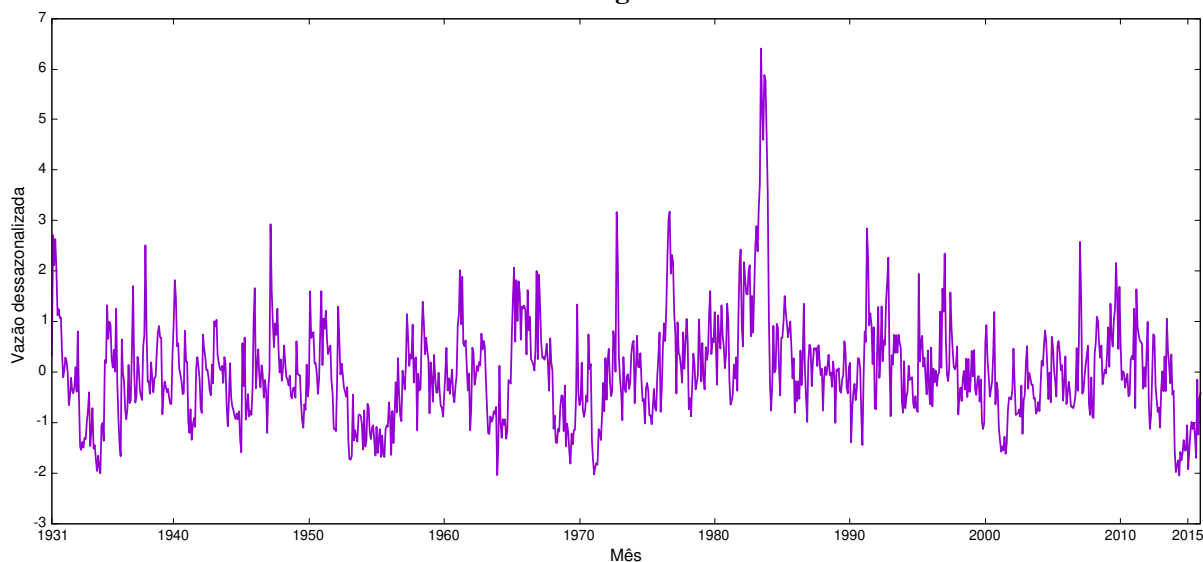
**Gráfico 12 – Série de vazões médias mensais da usina de Água Vermelha.**



**Fonte: Autoria própria.**

Como mencionado na Seção 2.4.1 é necessário que a componente sazonal da série seja retirada antes da realização da previsão. O Gráfico 13 mostra a série dessazonalizada da usina de Água Vermelha, de acordo com a Equação 2.13.

**Gráfico 13 – Série dessazonalizada da usina de Água Vermelha.**



**Fonte: Autoria própria.**

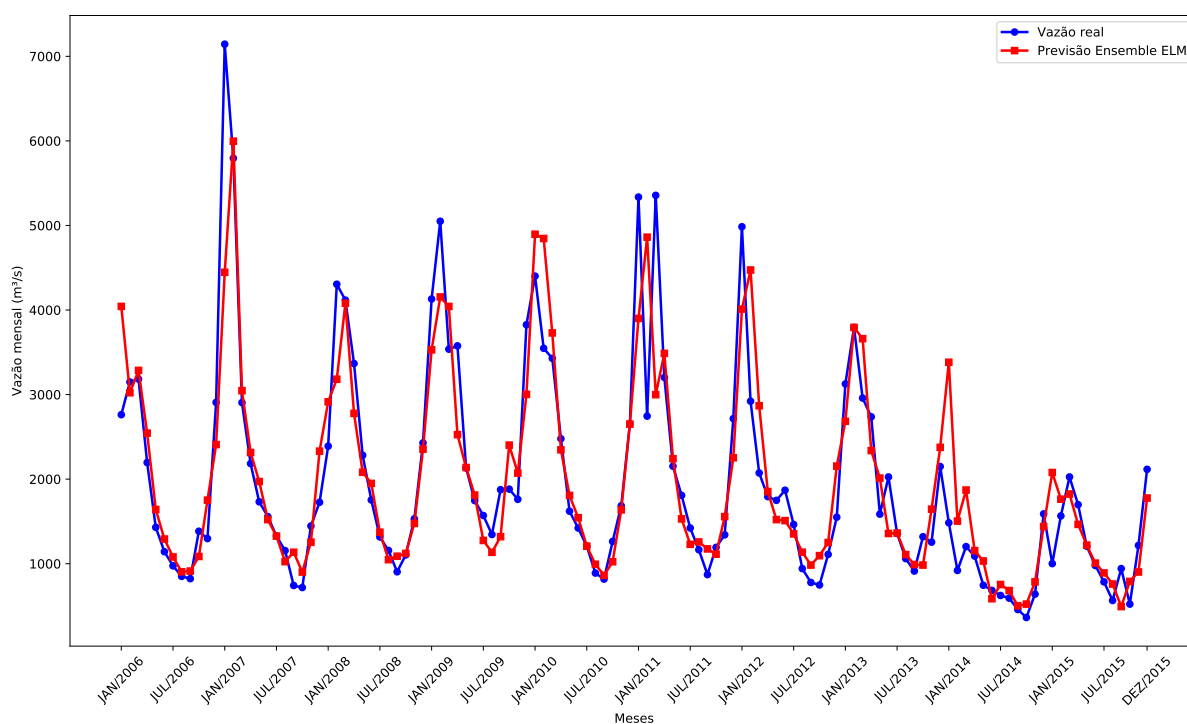
Na Tabela 1 são apresentados os resultados das simulações realizadas pelos modelos de previsão propostos para a usina de Água Vermelha com horizonte de previsão  $P = 1$ . Na tabela são mostrados o número de neurônios escolhido para cada rede neural, o MSE real, o MSE dos valores dessazonalizados, o MAE real, o MAE dessazonalizado e o MAPE real.

**Tabela 1 – Resultados previsões Água Vermelha de 2006 a 2015 para P = 1.**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	468.513,81	0,4292	417,01	0,4939	21,03%
ARMA(2,2)	0	459.868,32	0,4160	418,56	0,4920	21,35%
MLP	10	435.047,55	0,4320	417,44	0,5193	24,16%
RBF	95	433.015,37	0,4904	437,38	0,5641	25,46%
ELM	30	413.488,42	0,3899	390,31	0,4695	20,11%
ELM (CR)	80	393.462,78	0,3906	386,60	0,4816	20,52%
Elman	5	391.192,04	0,4018	390,95	0,4870	20,30%
Jordan	5	417.081,01	0,4180	401,15	0,4991	22,87%
ESN Jaeger	80	383.174,66	0,3763	374,12	0,4632	19,55%
ESN Jaeger (CR)	65	379.043,86	0,3782	372,12	0,4629	19,60%
ESN Ozturk	35	401.843,39	0,3773	381,18	0,4588	19,70%
ESN Ozturk (CR)	35	397.982,11	0,3773	381,66	0,4615	19,93%
Ensemble Média	0	374.741,57	0,3686	370,85	0,4577	19,56%
Ensemble Mediana	0	379.043,86	0,3782	372,12	0,4629	19,60%
Ensemble MLP	5	381.381,74	0,3959	386,26	0,4897	21,83%
Ensemble RBF	15	378.694,99	0,3913	391,66	0,4908	22,81%
<b>Ensemble ELM</b>	<b>40</b>	<b>360.574,55</b>	<b>0,3757</b>	<b>374,27</b>	<b>0,4744</b>	<b>19,70%</b>
Ensemble ELM (CR)	60	362.264,36	0,3873	379,11	0,4833	20,25%

Fonte: Autoria própria.

Os resultados da Tabela 1 mostram que a melhor previsão realizada foi obtida pelo *Ensemble* com combinador de ELM pois foi o modelo que apresentou o menor MSE para o conjunto de teste. Sua previsão, que foi a melhor para este cenário é apresentada no Gráfico 14.

**Gráfico 14 – Melhor previsão Água Vermelha de 2006 a 2015 – 1 passo a frente.**

Fonte: Autoria própria.

Ainda observando os dados da Tabela 1, verifica-se que o pior modelo foi o Autore-

gressivo, tendo obtido um MSE 30% maior que o melhor modelo. Nota-se também pelos dados da Tabela 1 que o modelo com menor MSE no domínio real não é necessariamente o mesmo com menor MSE dessazonalizado, visto que neste caso, o *Ensemble* com combinador de Média aritmética simples foi o melhor com um MSE dessazonalizado de 0,3686 contra 0,3757 do *Ensemble* ELM.

Outra observação a ser feita sobre os dados da Tabela 1 é com relação ao MAPE. Quando comparamos essa medida do melhor modelo com pior não vemos uma diferença de pouco mais de 1 ponto percentual, enquanto para o MSE essa diferença chega 30%. Como já foi dito no início desse capítulo, o MSE penaliza mais fortemente erros maiores, de forma que ele não é uma média simples dos erros cometidos pelo modelo, enquanto que o MAPE é uma média percentual simples que não faz distinção entre erros pequenos e erros grandes.

Seguindo, a Tabela 2 apresenta os resultados obtidos por cada modelo proposto para a usina de Água Vermelha com  $P = 3$ .

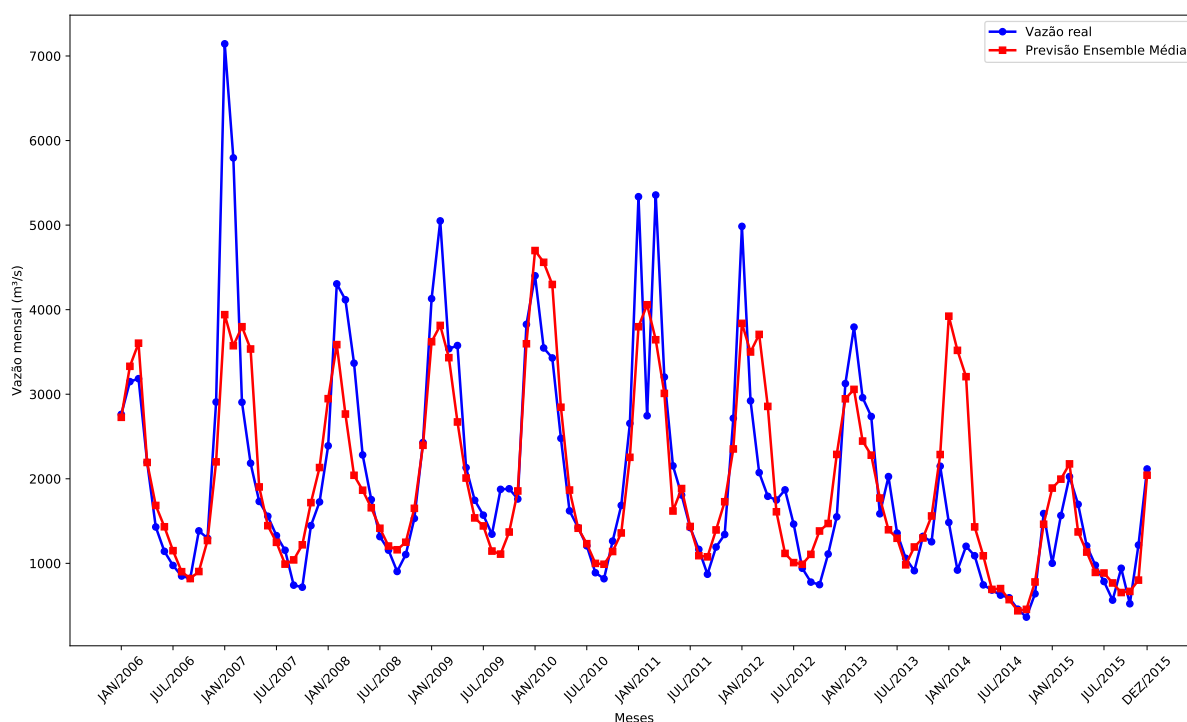
**Tabela 2 – Resultados previsões Água Vermelha de 2006 a 2015 para  $P = 3$ .**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	635.623,14	0,6905	525,75	0,6517	28,63%
ARMA(2,2)	0	652.368,14	0,6568	522,76	0,6375	28,20%
MLP	10	592.399,82	0,6705	510,71	0,6588	31,79%
RBF	95	616.761,67	0,7877	563,78	0,7500	35,97%
ELM	30	548.179,42	0,5525	472,96	0,5822	26,13%
ELM (CR)	80	548.521,40	0,5861	484,45	0,6108	28,05%
Elman	5	590.235,98	0,6518	504,45	0,6478	30,66%
Jordan	5	642.432,88	0,7444	536,07	0,6969	34,57%
ESN Jaeger	80	576.896,70	0,5940	489,68	0,6120	27,98%
ESN Jaeger (CR)	65	565.927,30	0,5894	489,06	0,6115	27,60%
ESN Ozturk	35	745.343,33	0,9746	593,25	0,8008	33,82%
ESN Ozturk (CR)	35	833.991,16	1,18	640,58	0,8806	36,81%
<b>Ensemble Média</b>	<b>0</b>	<b>543.746,08</b>	<b>0,5589</b>	<b>465,66</b>	<b>0,5712</b>	<b>25,10%</b>
<i>Ensemble</i> Mediana	0	574.877,03	0,5909	482,57	0,5954	26,24%
<i>Ensemble</i> MLP	5	595.528,64	0,6240	498,01	0,6209	27,97%
<i>Ensemble</i> RBF	15	582.641,85	0,6122	488,34	0,6149	27,99%
<i>Ensemble</i> ELM	40	598.571,40	0,6316	495,63	0,6167	26,49%
<i>Ensemble</i> ELM (CR)	60	596.396,64	0,6391	495,19	0,6191	26,09%

**Fonte: Autoria própria.**

Pelos dados da Tabela 2 observa-se que mais uma vez o melhor modelo de previsão foi um *Ensemble*, mas dessa vez o combinador de Média Aritmética simples, obtendo um MSE de 543.746,08 para conjunto de teste. Por sua vez, o pior preditor foi a Rede Neural com Estado de Eco e proposta de reservatório de dinâmicas de Ozturk, Xu e Príncipe (2007) com a utilização do Coeficiente de Regularização, que apresentou MSE de 833.991,16. Novamente, o resultado do MSE dessazonalizado foi diferente, o melhor modelo foi a rede ELM enquanto que o pior foi a rede ENS Ozturk (CR). O Gráfico 15 apresenta a melhor previsão realizada para usina de Água Vermelha com  $P = 3$ .

Gráfico 15 – Melhor previsão Água Vermelha de 2006 a 2015 – 3 passos a frente.



Fonte: Autoria própria.

A seguir, a Tabela 3 mostra os resultados das previsões realizadas para  $P = 6$ .

Tabela 3 – Resultados previsões Água Vermelha de 2006 a 2015 para  $P = 6$ .

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	920.536,06	0,9560	613,24	0,7587	34,87%
ARMA(2,2)	0	857.987,67	0,8951	588,20	0,7379	33,31%
MLP	10	705.060,11	0,8360	563,53	0,7312	36,06%
RBF	95	1.003.756,28	1,30	679,76	0,9019	43,29%
<b>ELM</b>	<b>30</b>	<b>602.358,84</b>	<b>0,6664</b>	<b>503,40</b>	<b>0,6417</b>	<b>29,75%</b>
ELM (CR)	80	643.385,78	0,7177	544,42	0,6906	33,53%
Elman	5	719.246,60	0,8161	563,72	0,7222	36,47%
Jordan	5	726.929,58	0,8792	569,66	0,7444	37,69%
ESN Jaeger	80	650.640,11	0,7129	522,22	0,6651	31,79%
ESN Jaeger (CR)	65	632.523,34	0,6990	519,92	0,6665	32,14%
ESN Ozturk	35	671.232,69	0,8152	554,56	0,7274	33,88%
ESN Ozturk (CR)	35	695.143,95	0,8861	571,95	0,7598	35,95%
Ensemble Média	0	718.287,32	0,7420	524,54	0,6518	28,54%
Ensemble Mediana	0	806.882,94	0,8172	547,73	0,6770	29,94%
Ensemble MLP	5	796.168,68	0,8097	548,92	0,6815	30,61%
Ensemble RBF	15	774.132,08	0,7849	537,02	0,6636	30,59%
Ensemble ELM	40	834.303,17	0,8658	563,19	0,7081	30,90%
Ensemble ELM (CR)	60	837.861,29	1,00	613,46	0,7915	37,49%

Fonte: Autoria própria.

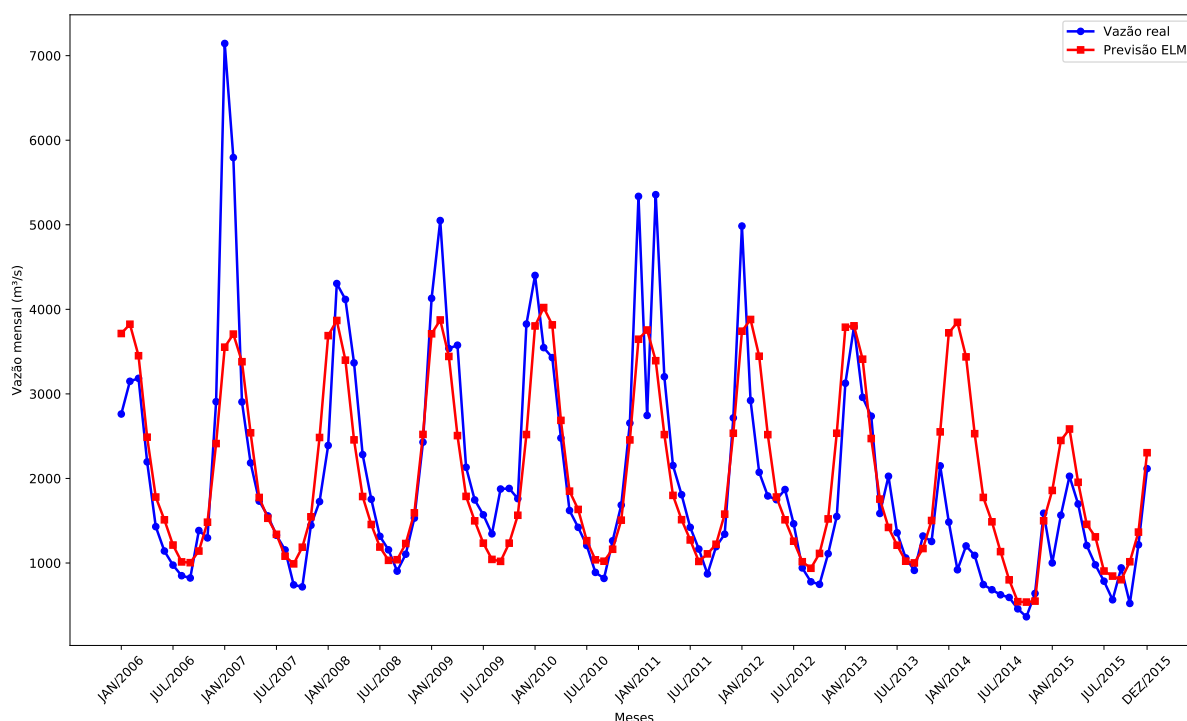
Analisando os dados da Tabela 3 verifica-se que o melhor modelo de previsão para o horizonte de  $P = 6$  foi a rede neural ELM sem Coeficiente de Regularização, com MSE real de 602.358,84. Diferentemente dos horizontes  $P = 1$  e  $P = 3$ , dessa vez não foi um *Ensem-*

ble que obteve o melhor desempenho, indicando que apesar de combinar o resultado de vários preditores os *Ensembles* nem sempre alcançam uma melhora no resultado geral, comprovando a necessidade de não medir esforços em desenvolver os melhores preditores individuais possíveis.

Em contrapartida, o pior resultado foi o da rede RBF com MSE real de 1.003.756, 28, tendo um resultado 66% pior que o do melhor preditor, a rede neural ELM. Aqui novamente há similaridade entre o MSE real e o dessazonalizado, uma vez que sob o viés do MSE dessazonalizado o preditor com melhor desempenho foi novamente a rede neural ELM, o mesmo ocorreu com as demais métricas analisadas.

O Gráfico 16 mostra a melhor previsão realizada para a série da usina de Água Vermelha com  $P = 6$ .

**Gráfico 16 – Melhor previsão Água Vermelha de 2006 a 2015 – 6 passos a frente.**



**Fonte: Autoria própria.**

Por fim, a Tabela 4 apresenta os resultados obtidos por todos os modelos para as previsões realizadas com 12 passos a frente para a usina de Água Vermelha.

Pelos dados da Tabela 4 nota-se que o modelo que realizou a melhor previsão foi a rede neural ESN de Jaeger sem Coeficiente de Regularização, tendo obtido MSE real junto ao conjunto de teste de 679.957, 56. Por outro lado o modelo que realizou a pior previsão para esse horizonte foi a rede neural RBF, com MSE real de 2.139.100, 52, um resultado 214% pior que o da ESN Jaeger.

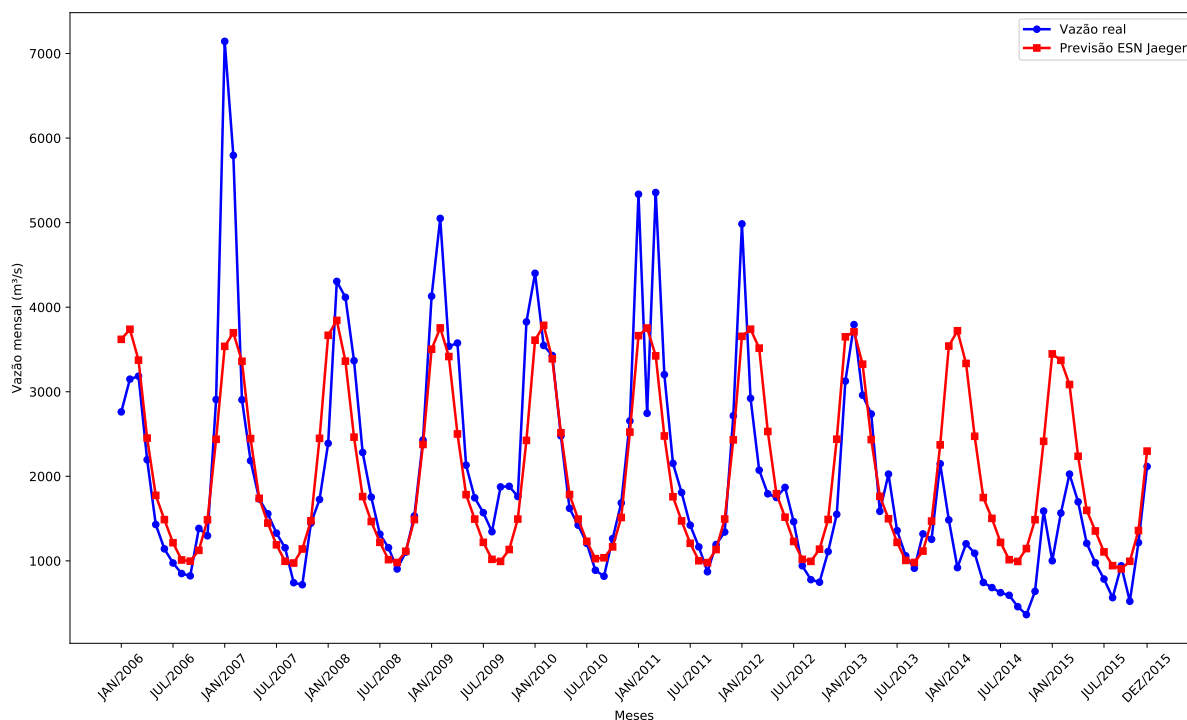
Novamente como ocorrido com  $P = 6$ , observa-se que ao considerar as previsões com horizonte de 12 passos o modelo predito com menor MSE dessazonalizado é o mesmo com menor MSE real, a rede neural ESN Jaeger.

**Tabela 4 – Resultados previsões Água Vermelha de 2006 a 2015 para P = 12.**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	1.032.390,38	1,31	675,30	0,8881	45,94%
ARMA(2,2)	0	844.191,84	1,08	616,40	0,8098	41,23%
MLP	10	717.730,32	0,8534	564,50	0,7359	36,99%
RBF	95	2.139.100,52	3,15	925,26	1,29	66,62%
ELM	30	695.473,23	0,8123	548,99	0,7102	36,13%
ELM (CR)	80	735.096,12	0,8840	577,66	0,7528	38,97%
Elman	5	733.779,79	0,8746	572,89	0,7460	38,25%
Jordan	5	711.201,57	0,8455	560,76	0,7306	36,64%
<b>ESN Jaeger</b>	<b>80</b>	<b>679.957,56</b>	<b>0,8021</b>	<b>545,17</b>	<b>0,7091</b>	<b>35,37%</b>
ESN Jaeger (CR)	65	690.611,82	0,8196	553,96	0,7192	36,84%
ESN Ozturk	35	726.758,25	0,8643	569,35	0,7408	37,54%
ESN Ozturk (CR)	35	749.393,17	0,8869	580,48	0,7526	38,56%
Ensemble Média	0	736.530,62	0,9079	561,82	0,7432	35,44%
Ensemble Mediana	0	761.767,76	0,9891	568,98	0,7684	36,75%
Ensemble MLP	5	779.261,85	0,9861	572,75	0,7648	36,83%
Ensemble RBF	15	738.982,18	0,9638	552,92	0,7471	36,26%
Ensemble ELM	40	800.998,37	1,05	592,73	0,8051	37,91%
Ensemble ELM (CR)	60	1.252.464,38	1,70	713,98	0,9585	46,67%

Fonte: Autoria própria.

O Gráfico 17 mostra a melhor previsão realizada pelos modelos propostos para a série da usina de Água Vermelha com horizonte de previsão de 12 passos a frente.

**Gráfico 17 – Melhor previsão Água Vermelha de 2006 a 2015 – 12 passos a frente.**

Fonte: Autoria própria.

Comparando os resultados obtidos para todos os horizontes de previsão das tabelas 1, 2, 3 e 4 é possível observar que o aumento no horizonte de previsão resulta no aumento do erro



das previsões realizadas pelos modelos. Com o intuito de detalhar esse fenômeno, a Tabela 5 apresenta o valor do melhor e do pior MSE real para cada horizonte de previsão. Além disso, ela também traz o percentual de aumento do MSE em relação ao horizonte anterior e em relação a  $P = 1$ .

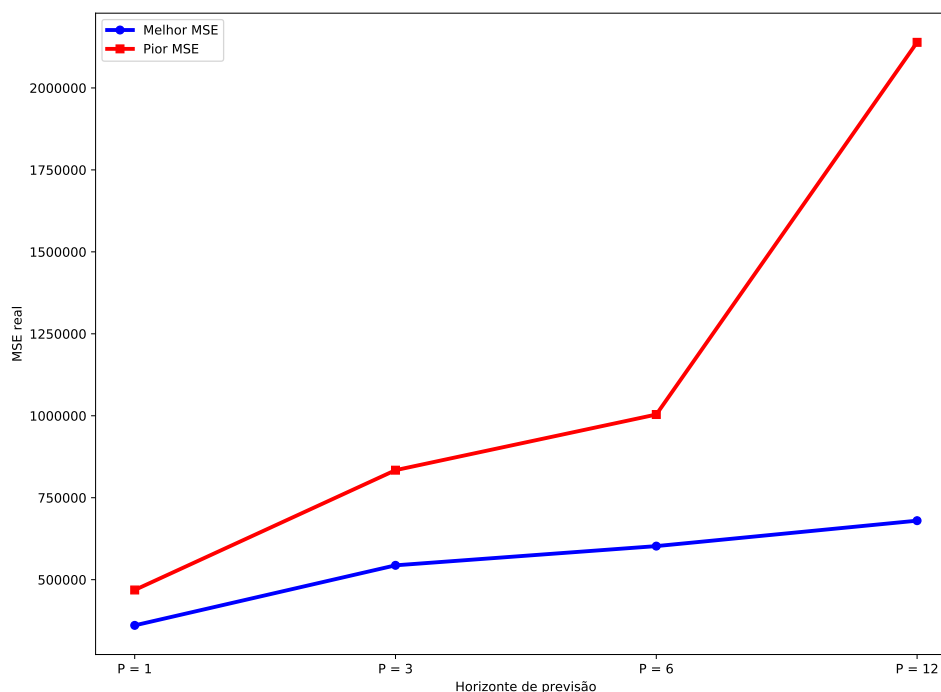
**Tabela 5 – Evolução MSE Água Vermelha.**

Horizonte	Melhor MSE	Anterior	P = 1	Pior MSE	Anterior	P = 1
$P = 1$	360.574,55	–	–	468.513,81	–	–
$P = 3$	543.746,08	50,80%	50,80%	833.991,16	78,01%	78,01%
$P = 6$	602.358,84	10,78%	67,06%	1.003.756,28	20,36%	114,24%
$P = 12$	679.957,56	12,88%	88,58%	2.139.100,52	113,11%	356,57%

**Fonte: Autoria própria.**

Os dados apresentados na Tabela 5 deixam claro que a medida em que o horizonte de previsão cresce o desempenho dos modelos de previsão diminui, apresentando erros cada vez maiores. Esse comportamento já era esperado, uma vez que estudos com séries temporais realizados por Ballini (2000) mostraram que a dependência entre as amostras diminui a medida que o horizonte se eleva. O Gráfico 18 mostra a evolução do melhor e do pior MSE para as previsões realizadas para a usina de Água Vermelha com todos os horizontes de previsão.

**Gráfico 18 – Evolução MSE para a usina de Água Vermelha.**



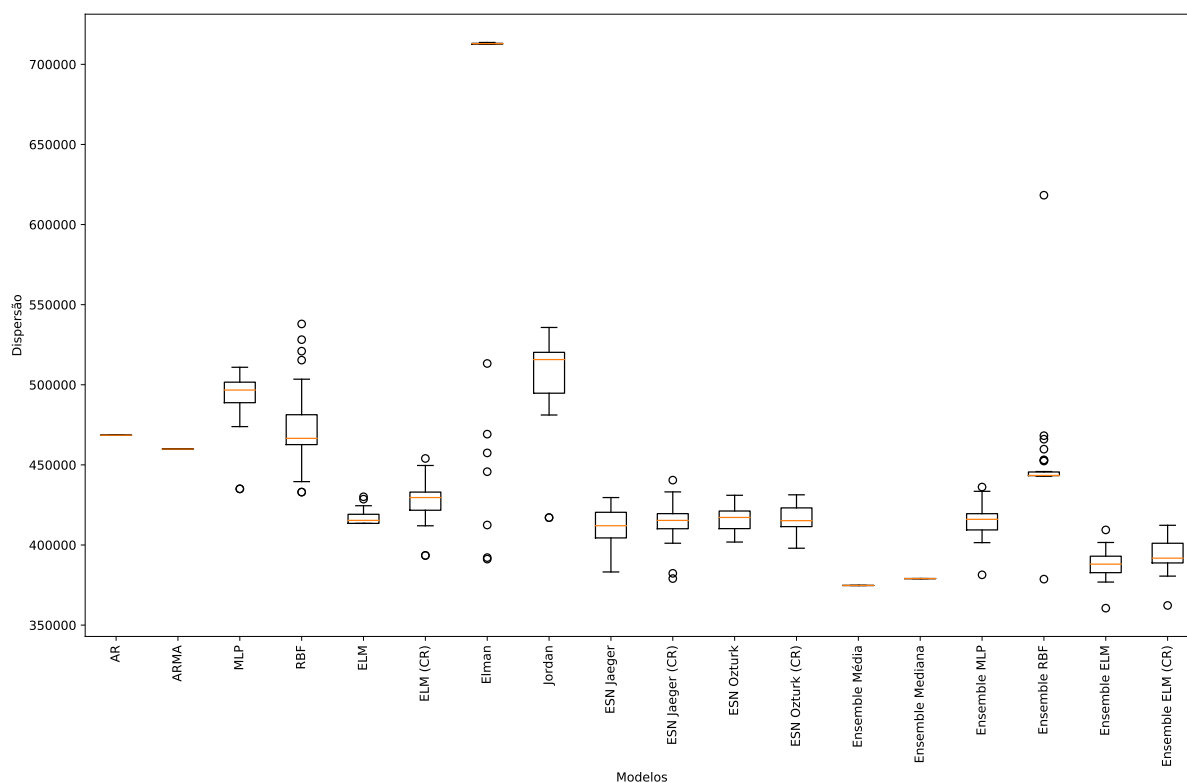
**Fonte: Autoria própria.**

Pelos dados da Tabela 5 e pelo Gráfico 18 verificamos que o maior aumento do MSE foi do horizonte  $P = 1$  para o  $P = 3$ , tendo um incremento percentual de 50%, enquanto para  $P = 6$  e  $P = 12$  o aumento foi menor que 15%. Já com relação ao pior MSE, o maior incremento percentual foi de  $P = 6$  para  $P = 12$ , de 113%.

Ainda analisando os dados da Tabela 5 vale ressaltar que a melhor previsão realizada para  $P = 12$  obteve um erro 88% maior que para  $P = 1$ . Outro fator que colabora para isso é a utilização da previsão recursiva. Como foi discutido na Seção 4.10 essa abordagem faz a utilização de previsões anteriores como entrada para previsões futuras. Como os modelos acabam utilizando uma previsão que já contém uma componente de erro como entrada esse erro é propagado.

Como foi mencionado anteriormente, cada modelo proposto por esse trabalho foi treinado 30 vezes com o objetivo de obter dados estatísticos quanto a reprodutibilidade dos resultados obtidos. Com esse objetivo, construiu-se o Gráfico 19 do tipo boxplot, o qual mostra a dispersão apresentada pelas execuções de todos os modelos para a usina de Água Vermelha.

**Gráfico 19 – Dispersão dos resultados para a usina de Água Vermelha.**



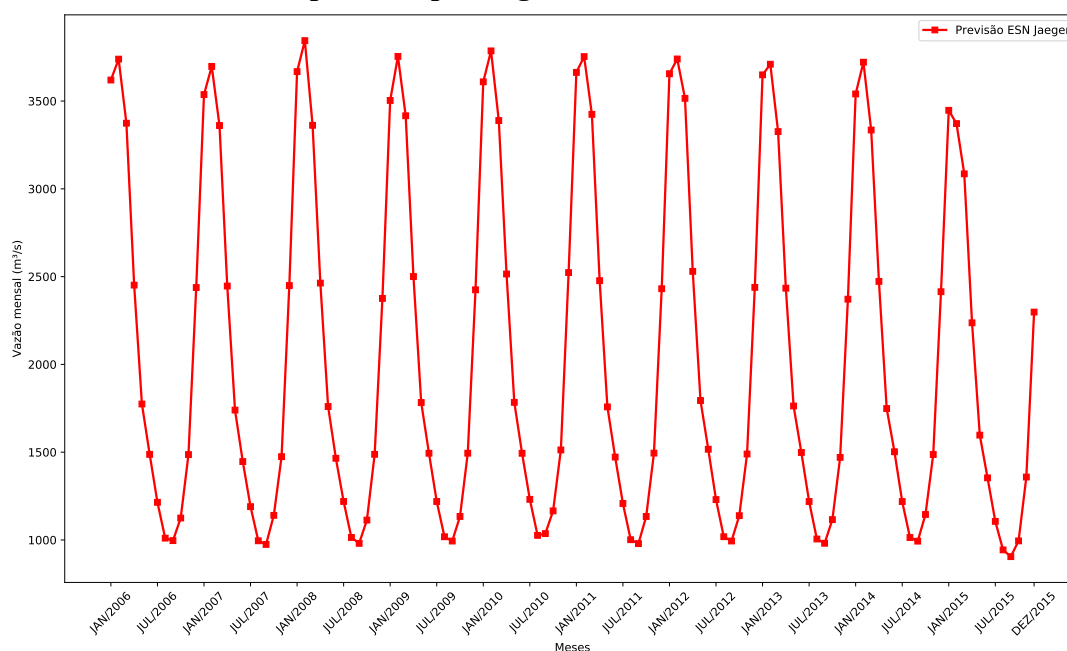
**Fonte: Autoria própria.**

Analisando o Gráfico 19 nota-se que os modelos AR, ARMA, *Ensemble de Média* e *Ensemble Mediana* não apresentaram dispersão, tendo mostrado o mesmo resultado em todas as 30 execuções. Esse resultado já era esperado, uma vez que o AR possui solução determinística (equações de Yule-Walker) e o ARMA tem seus coeficientes determinados através de estimadores de máxima verossimilhança. Por outro lado, as redes neurais apresentam a necessidade de treinamento, sendo que em todos os casos ocorre inicialização aleatória dos pesos. Dessa forma, mesmo com a utilização dos mesmos parâmetros de treinamento e do mesmo número de épocas o resultado final de cada treinamento é distinto.

Com relação a dispersão dos resultados das redes neurais verifica-se que a maior foi a da rede RBF, enquanto a menor foi a da rede de Elman.

Outro comportamento que ocorre nas previsões com horizontes maiores que  $P = 1$  é a tendência dos modelos realizarem previsões que expressem o comportamento cíclico principal da série. Observando as curvas das previsões nos gráficos 15, 16 e 17 verifica-se que conforme o horizonte de previsão aumenta a curva dos valores previstos apresenta um comportamento cada vez mais cíclico e padronizado. Para deixar esse comportamento claro o Gráfico 20 mostra apenas a curva das previsões realizadas pela ESN Jaeger para a usina de Água Vermelha com  $P = 12$ .

**Gráfico 20 – Padrão das previsões para Água Vermelha com  $P = 12$ .**



**Fonte: Autoria própria.**

O Gráfico 20 mostra o comportamento senoidal da curva de previsões e revela que o mesmo padrão se repete por toda extensão da série prevista. Tal comportamento apresenta um ciclo anual, com picos e vales sempre nas mesmas épocas do ano. Na Seção 2.4 foi dito que séries de vazões mensais apresentam sazonalidade anual devido ao ciclo das chuvas. Apesar deste ter seu comportamento geral conhecido, existem inúmeros fatores externos com comportamento aleatório que influenciam diretamente na densidade pluviométrica de uma região. Dessa forma, sabe-se com certeza que no mês de janeiro chove mais que em dezembro, mas tendo em mãos os dados das chuvas de janeiro de um ano não podemos dizer com certeza quanto irá chover em dezembro. Esses fatores externos que dão aleatoriedade à série e fazem com que o problema seja de difícil solução.

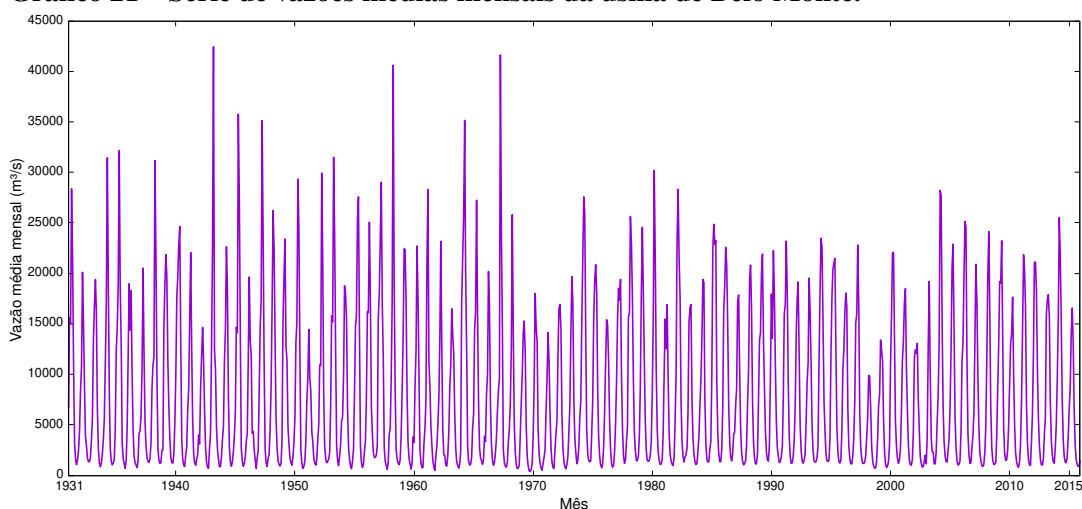
A previsão recursiva insere os novos valores previstos, mas a componente aleatória não é considerada nas previsões com  $P > 1$ , o que gera esse comportamento estático apresentado no Gráfico 20. Esta tendência já foi verificada em outros trabalhos de Siqueira (2013) e Ballini (2000) e chama-se média de longo termo.

### 5.3 USINA HIDRELÉTRICA DE BELO MONTE

A usina hidrelétrica de Belo Monte é abastecida pelo Rio Xingu e fica localizada próximo ao município de Altamira, no sudoeste do estado do Pará. Sua construção foi iniciada em Junho de 2011 e tem término previsto para o ano de 2019. Trata-se, portanto, de uma usina que ainda não se encontra em operação. Apesar disso, o Ministério de Minas e Energia sempre realizou medições da vazão afluyente em vários pontos do Rio Xingu, de modo que a usina de Belo Monte também possui uma série histórica.

Tal série apresenta média  $\hat{\mu} = 8.045,84 \text{ m}^3/\text{s}$  e desvio padrão  $\hat{\sigma} = 7.769,75 \text{ m}^3/\text{s}$ , sendo apresentada no Gráfico 21.

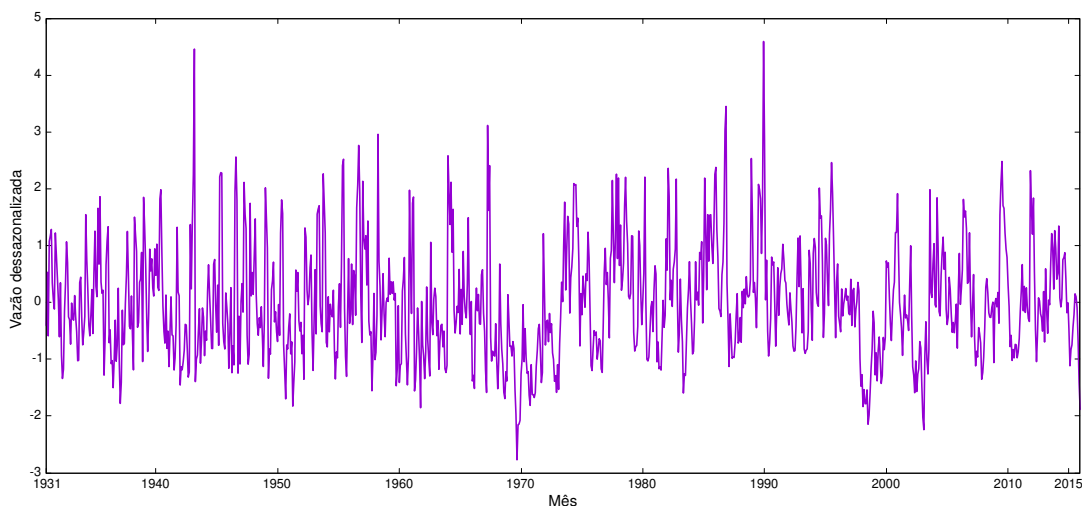
**Gráfico 21 – Série de vazões médias mensais da usina de Belo Monte.**



**Fonte: Autoria própria.**

O Gráfico 22 mostra a série dessazonalizada de acordo com a Equação 2.13 para o mesmo posto.

**Gráfico 22 – Série dessazonalizada da usina de Belo Monte.**



**Fonte: Autoria própria.**

A Tabela 6 mostra os resultados das previsões realizadas por todos os modelos para o conjunto de teste da usina de Belo Monte com horizonte de previsão de  $P = 1$ . Nela é possível verificar a quantidade de neurônios selecionada para cada modelo e o valor das métricas de erro calculadas para as previsões.

**Tabela 6 – Resultados previsões Belo Monte de 2006 a 2015 para  $P = 1$ .**

<b>Modelo</b>	<b>Neur</b>	<b>MSE</b>	<b>MSE (d)</b>	<b>MAE</b>	<b>MAE (d)</b>	<b>MAPE</b>
AR(3)	0	5.136.070,02	0,3681	1.333,52	0,4711	16,14%
ARMA(4,3)	0	5.560.980,02	0,3978	1.423,45	0,4911	16,47%
MLP	10	4.611.515,04	0,3956	1.258,04	0,4869	16,77%
RBF	25	4.189.519,18	0,3976	1.236,87	0,4898	16,95%
ELM	35	3.932.745,86	0,4455	1.267,65	0,5302	17,71%
ELM (CR)	5	4.263.133,64	0,3620	1.247,08	0,4633	15,10%
Elman	5	4.348.164,47	0,4310	1.246,96	0,4929	15,05%
Jordan	5	4.841.287,43	0,3795	1.317,76	0,4833	17,00%
ESN Jaeger	5	3.990.917,22	0,3619	1.202,66	0,4594	15,14%
ESN Jaeger (CR)	5	4.063.963,41	0,3703	1.197,11	0,4610	15,08%
ESN Ozturk	5	4.135.861,20	0,3986	1.292,16	0,4992	16,69%
ESN Ozturk (CR)	5	4.054.930,75	0,3769	1.205,12	0,4630	15,25%
<i>Ensemble Média</i>	0	3.664.436,73	0,3738	1.192,59	0,4793	15,77%
<i>Ensemble Mediana</i>	0	3.654.331,87	0,3379	1.167,74	0,4453	14,32%
<i>Ensemble MLP</i>	5	3.683.236,20	0,3495	1.190,90	0,4552	15,16%
<i>Ensemble RBF</i>	15	3.716.658,52	0,4082	1.219,07	0,5241	17,88%
<b><i>Ensemble ELM</i></b>	<b>5</b>	<b>3.496.325,52</b>	<b>0,3583</b>	<b>1.154,20</b>	<b>0,4642</b>	<b>15,06%</b>
<i>Ensemble ELM (CR)</i>	5	3.537.238,59	0,3522	1.151,14	0,4590	15,08%

**Fonte: Autoria própria.**

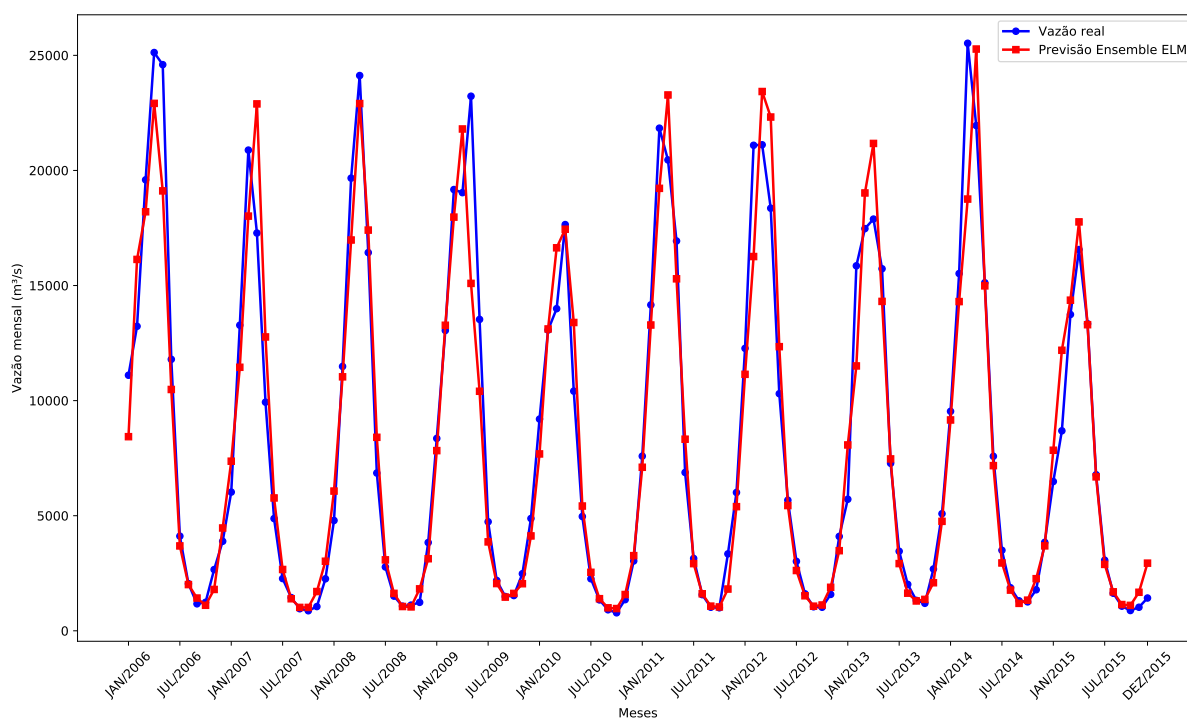
Analisando os dados da Tabela 6 verifica-se que o melhor modelo de previsão foi o *Ensemble* com combinador de ELM sem Coeficiente de Regularização, o qual obteve um MSE real de 3.496.325,52. Em contrapartida, o pior desempenho foi do modelo ARMA com MSE real de 5.560.980,02.

Quando analisamos os dados da Tabela 6 dando ênfase ao MSE dessazonalizado em vez do real o resultado é diferente. Nesse caso, o modelo que apresentou as melhores previsões foi o *Ensemble* com combinador de Mediana simples tendo obtido MSE dessazonalizado de 0,3379. Neste caso, o pior modelo também foi diferente, sendo ele a ELM com MSE dessazonalizado de 0,4455.

Para corroborar com o fato de que existe uma variação no desempenho dos modelos dependendo de qual métrica de erro for considerada, podemos dar atenção agora ao MAE, se olharmos os dados da Tabela 6 pelo viés dessa métrica verifica-se que o melhor modelo de previsão dessa vez foi o *Ensemble* com combinador de ELM e fazendo uso do Coeficiente de Regularização.

O Gráfico 23 apresenta a melhor previsão realizada para a usina de Belo Monte com horizonte de previsão de 1 passo a frente, que foi obtida através do *Ensemble* com combinador de ELM.

Gráfico 23 – Melhor previsão Belo Monte de 2006 a 2015 – 1 passo a frente.



Fonte: Autoria própria.

A seguir, a Tabela 7 apresenta os resultados das previsões realizadas para a usina de Belo Monte com horizonte de previsão de  $P = 3$ . Aqui novamente é possível observar o número de neurônios selecionados para cada modelo preditivo e os valores de todas as métricas de erro calculadas.

Tabela 7 – Resultados previsões Belo Monte de 2006 a 2015 para  $P = 3$ .

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	9.481.744,90	0,7818	1.936,64	0,7168	26,40%
ARMA(4,3)	0	10.338.028,24	0,7803	1.975,14	0,6895	23,93%
MLP	10	6.610.976,00	0,7674	1.694,82	0,7196	26,26%
RBF	25	7.374.107,90	0,7756	1.731,87	0,7170	24,91%
<b>ELM</b>	<b>35</b>	<b>5.211.473,53</b>	<b>0,7751</b>	<b>1.525,03</b>	<b>0,7000</b>	<b>23,65%</b>
ELM (CR)	5	5.977.477,68	0,6716	1.544,47	0,6391	21,83%
Elman	5	12.461.369,17	1,62	2.126,66	0,9735	27,14%
Jordan	5	8.029.604,93	0,7721	1.853,97	0,7347	27,23%
ESN Jaeger	5	5.744.660,75	0,6405	1.509,01	0,6169	20,68%
ESN Jaeger (CR)	5	6.268.470,74	0,6559	1.577,00	0,6319	21,66%
ESN Ozturk	5	6.121.782,43	0,6100	1.580,77	0,6106	20,92%
ESN Ozturk (CR)	5	5.576.790,17	0,6437	1.471,20	0,6092	20,07%
Ensemble Média	0	6.583.650,15	0,6255	1.559,46	0,5976	19,97%
Ensemble Mediana	0	7.892.943,44	0,6883	1.690,42	0,6285	20,40%
Ensemble MLP	5	9.631.438,35	0,7448	1.784,76	0,6433	20,72%
Ensemble RBF	15	6.756.906,30	0,7051	1.590,44	0,6347	19,81%
Ensemble ELM	5	7.360.295,02	0,6748	1.628,12	0,6182	20,02%
Ensemble ELM (CR)	5	7.038.931,44	0,6529	1.602,92	0,6077	19,83%

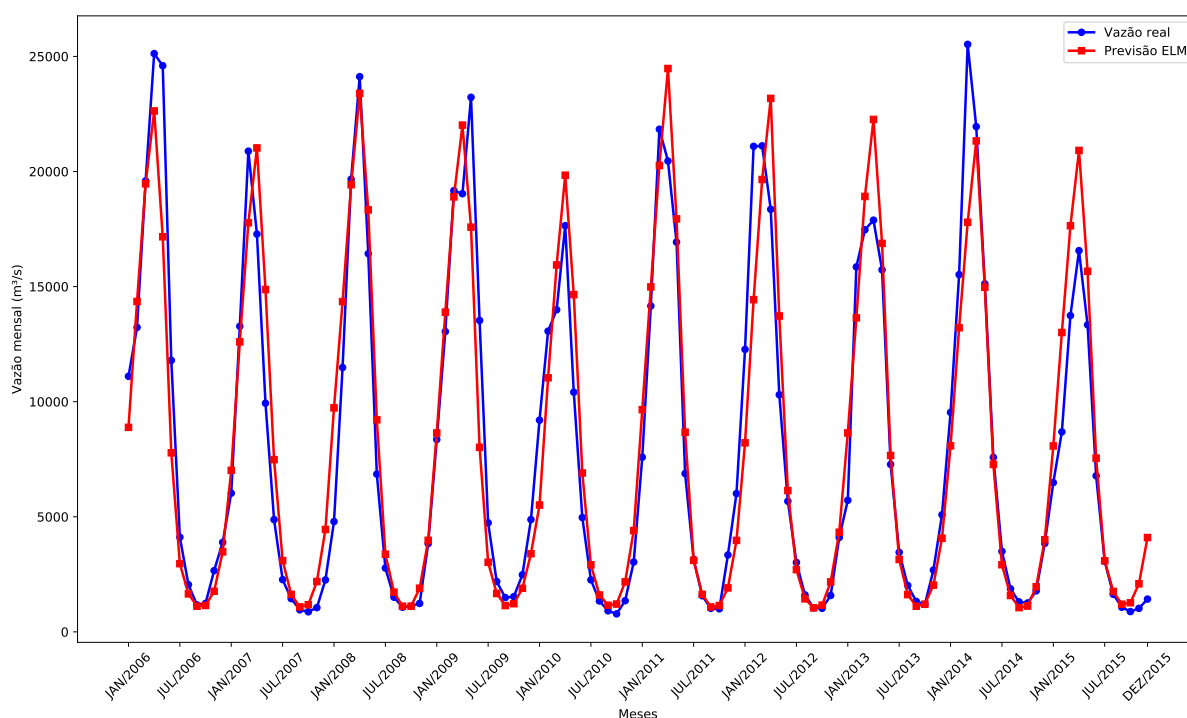
Fonte: Autoria própria.

Pelos dados da Tabela 7 verifica-se que dentre todos os modelos testados, aquele que apresentou o melhor desempenho foi a rede neural ELM sem Coeficiente de Regularização, com MSE real de 5.211.473, 53. Por sua vez, o pior desempenho foi da rede Elman, com MSE real de 12.461.369, 17, um resultado insatisfatório, sendo 139% maior do que o do melhor modelo testado. No ambiente dessazonalizado entretanto, o melhor modelo foi a ESN de Ozturk, Xu e Príncipe (2007) sem a utilização de Coeficiente de Regularização.

Mais uma vez pode-se verificar que ao mudar a métrica de erro principal o resultado geral pode ser diferente. Para o horizonte de previsão de 3 passos a frente, ao olhar-se para o MAPE como a métrica de erro dominante encontramos um *Ensemble* como sendo o melhor modelo preditivo, mais especificamente o *Ensemble* com combinador de RBF, tendo obtido um MAPE de 19, 81%.

O Gráfico 24 mostra a melhor previsão realizada para o conjunto de teste da usina de Belo Monte com horizonte de previsão de  $P = 3$ , tal previsão foi realizada pela rede neural ELM.

**Gráfico 24 – Melhor previsão Belo Monte de 2006 a 2015 – 3 passos a frente.**



**Fonte: Autoria própria.**

Avançando sobre os horizontes de previsão, na Tabela 8 são apresentados os resultados para a planta de Belo Monte com horizonte de previsão de  $P = 6$ .

Os dados da Tabela 8 mostram que assim como ocorreu com  $P = 3$ , na previsão de 6 passos a frente o modelo com melhor desempenho foi novamente a rede ELM, com MSE real de 5.390.642, 11. Por outro lado, as piores previsões foram aquelas realizadas pelo *Ensemble* com combinador de MLP, tendo obtido MSE real de 14.838.424, 95.

A análise do MSE dessazonalizado mostra como melhor preditor a rede ESN Ozturk

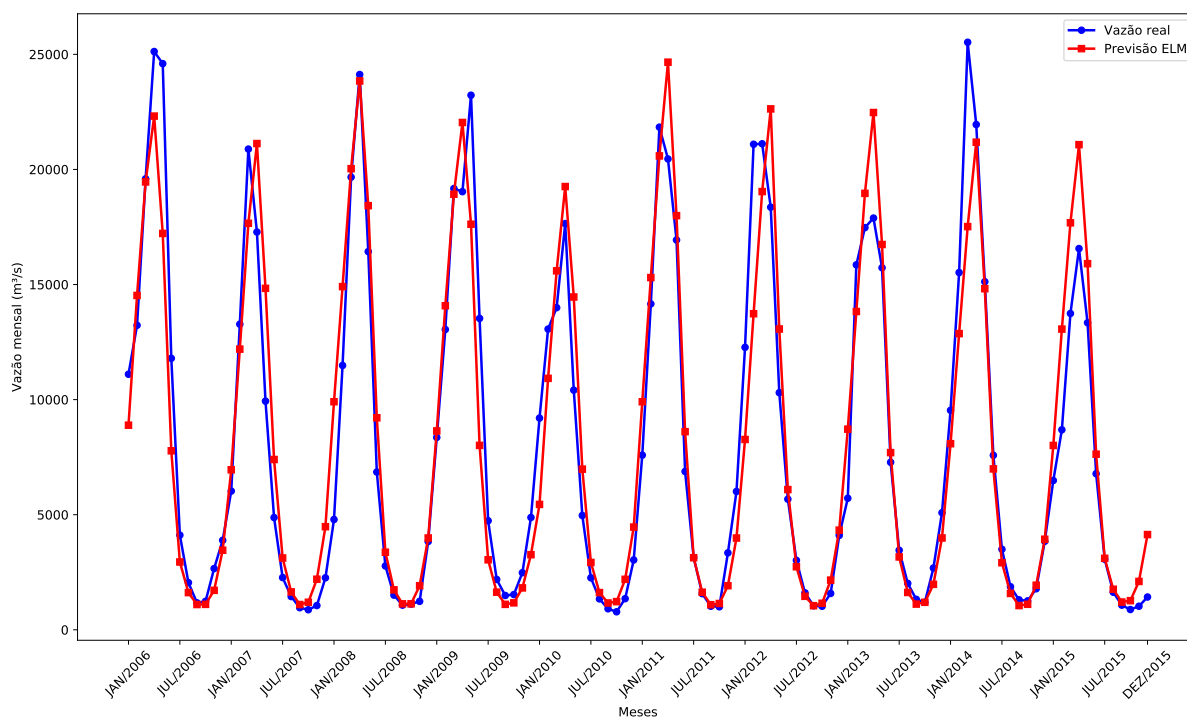
**Tabela 8 – Resultados previsões Belo Monte de 2006 a 2015 para P = 6.**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	13.220.246, 31	1,07	2.261, 87	0, 8595	32, 10%
ARMA(4,3)	0	14.598.946, 90	1,07	2.341, 76	0, 8285	28, 73%
MLP	10	8.503.685, 65	0, 9158	1.990, 49	0, 8098	29, 96%
RBF	25	9.560.433, 48	0, 9534	1.938, 90	0, 7693	27, 65%
<b>ELM</b>	<b>35</b>	<b>5.390.642, 11</b>	<b>0, 8129</b>	<b>1.554, 93</b>	<b>0, 7191</b>	<b>24, 19%</b>
ELM (CR)	5	6.266.406, 91	0, 7252	1.589, 44	0, 6722	23, 03%
Elman	5	6.837.675, 11	1,04	1.605, 38	0, 7740	23, 32%
Jordan	5	10.176.393, 17	0, 9748	2.109, 99	0, 8430	31, 39%
ESN Jaeger	5	5.778.132, 12	0, 7207	1.515, 88	0, 6560	21, 52%
ESN Jaeger (CR)	5	5.681.748, 34	0, 7446	1.506, 14	0, 6615	21, 17%
ESN Ozturk	5	5.906.402, 73	0, 6860	1.531, 26	0, 6502	21, 66%
ESN Ozturk (CR)	5	5.764.898, 14	0, 7325	1.505, 27	0, 6533	21, 03%
Ensemble Média	0	10.217.687, 60	0, 8305	1.963, 32	0, 7119	23, 26%
Ensemble Mediana	0	12.782.750, 96	0, 9963	2.213, 76	0, 7806	24, 98%
Ensemble MLP	5	14.838.424, 95	1,16	2.406, 98	0, 8475	26, 57%
Ensemble RBF	15	11.651.888, 44	1,03	2.069, 25	0, 7966	24, 99%
Ensemble ELM	5	11.834.559, 77	0, 9613	2.119, 24	0, 7614	24, 18%
Ensemble ELM (CR)	5	11.409.184, 74	0, 9323	2.078, 66	0, 7490	24, 08%

Fonte: Autoria própria.

sem a utilização de Coeficiente de Regularização, assim como ocorre com o horizonte de 3 passos a frente.

O Gráfico 25 apresenta a melhor previsão para este cenário, que foi realizada pela rede neural ELM.

**Gráfico 25 – Melhor previsão Belo Monte de 2006 a 2015 – 6 passos a frente.**

Fonte: Autoria própria.



Por fim, a Tabela 9 apresenta os resultados de todas as previsões realizadas para o último horizonte testado para a usina de Belo Monte, o horizonte de 12 passos a frente.

**Tabela 9 – Resultados previsões Belo Monte de 2006 a 2015 para P = 12.**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	13.856.487,04	1,31	2.507,36	0,9821	37,10%
ARMA(4,3)	0	10.811.736,13	1,31	2.072,80	0,8915	32,40%
MLP	10	8.834.763,12	0,8732	1.982,22	0,7911	28,99%
RBF	25	10.802.574,68	1,14	2.086,96	0,8721	31,34%
ELM	35	6.258.443,28	0,7798	1.582,07	0,6811	22,80%
ELM (CR)	5	6.209.801,53	0,7287	1.585,01	0,6741	23,13%
Elman	5	7.177.995,01	0,9005	1.738,34	0,7474	25,43%
Jordan	5	10.698.937,70	1,03	2.183,05	0,8693	32,50%
ESN Jaeger	5	5.938.616,01	0,7261	1.531,70	0,6563	21,72%
ESN Jaeger (CR)	5	5.942.372,26	0,7407	1.521,83	0,6550	21,16%
ESN Ozturk	5	6.391.806,49	0,7184	1.606,68	0,6727	23,41%
<b>ESN Ozturk (CR)</b>	<b>5</b>	<b>5.911.871,83</b>	<b>0,7356</b>	<b>1.517,36</b>	<b>0,6527</b>	<b>21,07%</b>
<i>Ensemble Média</i>	0	9.363.944,04	1,12	1.914,92	0,8249	27,57%
<i>Ensemble Mediana</i>	0	11.326.906,60	1,39	2.165,58	0,9363	30,86%
<i>Ensemble MLP</i>	5	12.489.587,18	1,61	2.319,99	1,01	33,19%
<i>Ensemble RBF</i>	15	12.834.385,61	1,23	2.243,33	0,8622	27,33%
<i>Ensemble ELM</i>	5	10.523.981,04	1,30	2.085,63	0,9029	29,35%
<i>Ensemble ELM (CR)</i>	5	10.465.521,50	1,27	2.012,56	0,8758	28,88%

**Fonte: Autoria própria.**

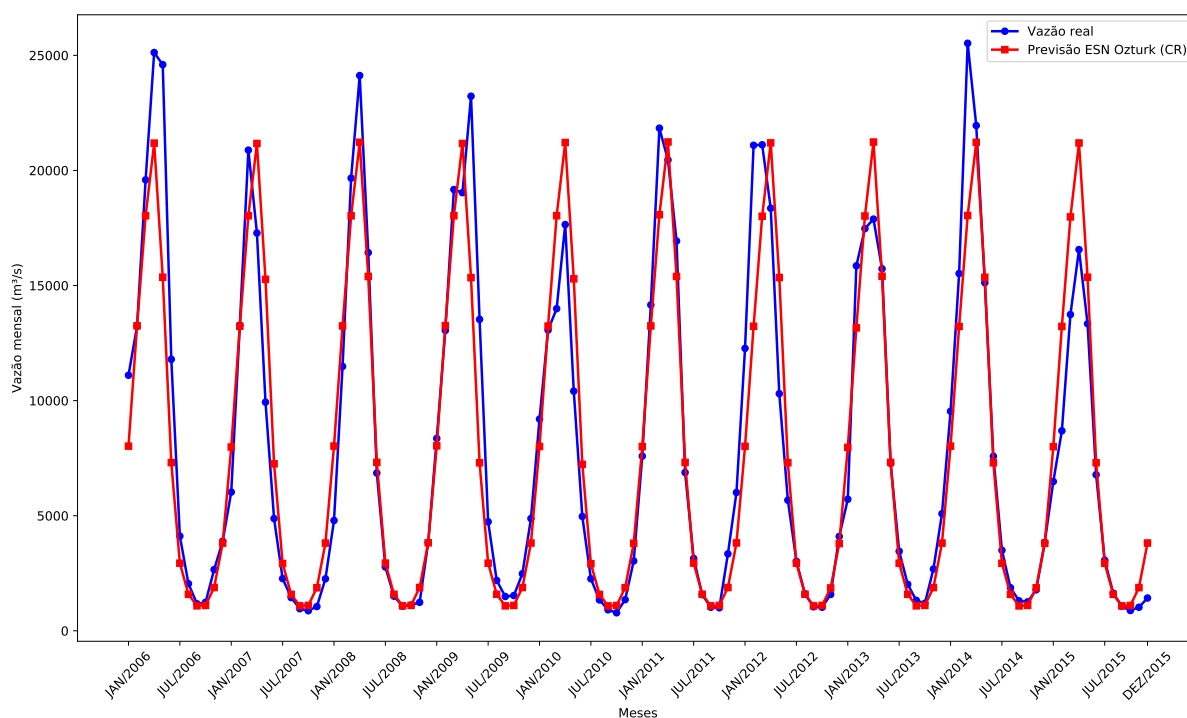
Ao analisar os dados da Tabela 9 verifica-se que o melhor modelo preditivo para  $P = 12$  é a rede ESN Ozturk com Coeficiente de Regularização, a qual apresenta MSE real de 5.911.871,83. Na Seção 4.8.2 foi dito que as Redes Neurais com Estado de Eco possuem memória intrínseca, sendo que as respostas dadas pela rede neural dependem não somente das entradas fornecidas ao modelo, mas também das respostas anteriores, o que pode favorecer essa arquitetura ao ser aplicada em séries temporais.

Observando o MSE dessazonalizado, o melhor preditor foi o mesmo, ESN de Ozturk, Xu e Príncipe (2007), mas dessa vez sem o Coeficiente de Regularização. Por sua vez, no ambiente dessazonalizado o pior modelo é o *Ensemble* com combinador de MLP, tendo obtido um MSE real 135% pior.

Vale ressaltar, que considerando o MAE como a principal métrica de erro a ESN Ozturk (CR) também alcançou o melhor resultado. Nesse mesmo cenário o pior modelo foi mais uma vez o Auto-regressivo.

O Gráfico 26 mostra a melhor previsão realizada para a usina de Belo Monte com horizonte de previsão de 12 passos a frente.

Como dito na Seção 4.10 e exemplificado pela Tabela 5 e Gráfico 18, a previsão recursiva propaga o erro obtido pelas previsões para os valores futuros, de modo que quanto maior o horizonte de previsão maior tende a ser o erro inerente a previsão. Com o objetivo de analisar esse comportamento, a Tabela 10 contém o melhor e pior MSE real por horizonte para a usina

**Gráfico 26 – Melhor previsão Belo Monte de 2006 a 2015 – 12 passos a frente.**

**Fonte: Autoria própria.**

de Belo Monte, além do aumento percentual em relação ao horizonte anterior e em relação a  $P = 1$ .

**Tabela 10 – Evolução MSE Belo Monte.**

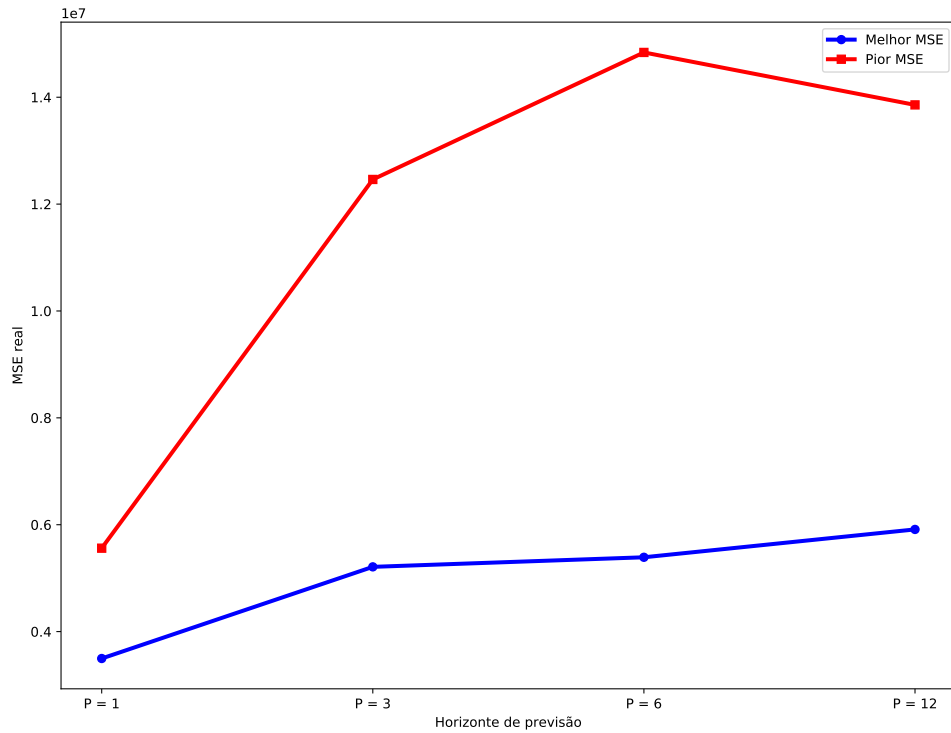
Horizonte	Melhor MSE	Anterior	$P = 1$	Pior MSE	Anterior	$P = 1$
$P = 1$	3.496.325, 52	–	–	5.560.980, 02	–	–
$P = 3$	5.211.473, 53	49, 06%	49, 06%	12.461.369, 17	124, 09%	124, 09%
$P = 6$	5.390.642, 11	3, 44%	54, 18%	14.838.424, 95	19, 08%	166, 83%
$P = 12$	5.911.871, 83	9, 67%	69, 09%	13.856.487, 04	–6, 62%	149, 17%

**Fonte: Autoria própria.**

Pelos dados da Tabela 10 verifica-se que em se tratando do aumento em relação ao horizonte anterior o maior incremento tanto do melhor quanto do pior caso foi de  $P = 1$  para  $P = 3$ , com 49, 06% e 124, 09% respectivamente. Os aumentos de  $P = 3$  para  $P = 6$  e de  $P = 6$  para  $P = 12$  foram mais modestos, sendo 3, 44% e 9, 67% respectivamente para o melhor MSE e 19, 08% e –6, 62% para o pior MSE.

Quando se dá atenção ao aumento do MSE em relação a  $P = 1$  verifica-se que é sempre crescente, enquanto que para o pior MSE ele começa crescente mas apresenta uma queda de 6, 62% de  $P = 6$  para  $P = 12$ . Com o objetivo de esclarecer esse comportamento o Gráfico 27 apresenta a curva de evolução do MSE real em relação a cada um dos horizontes de previsão testados.

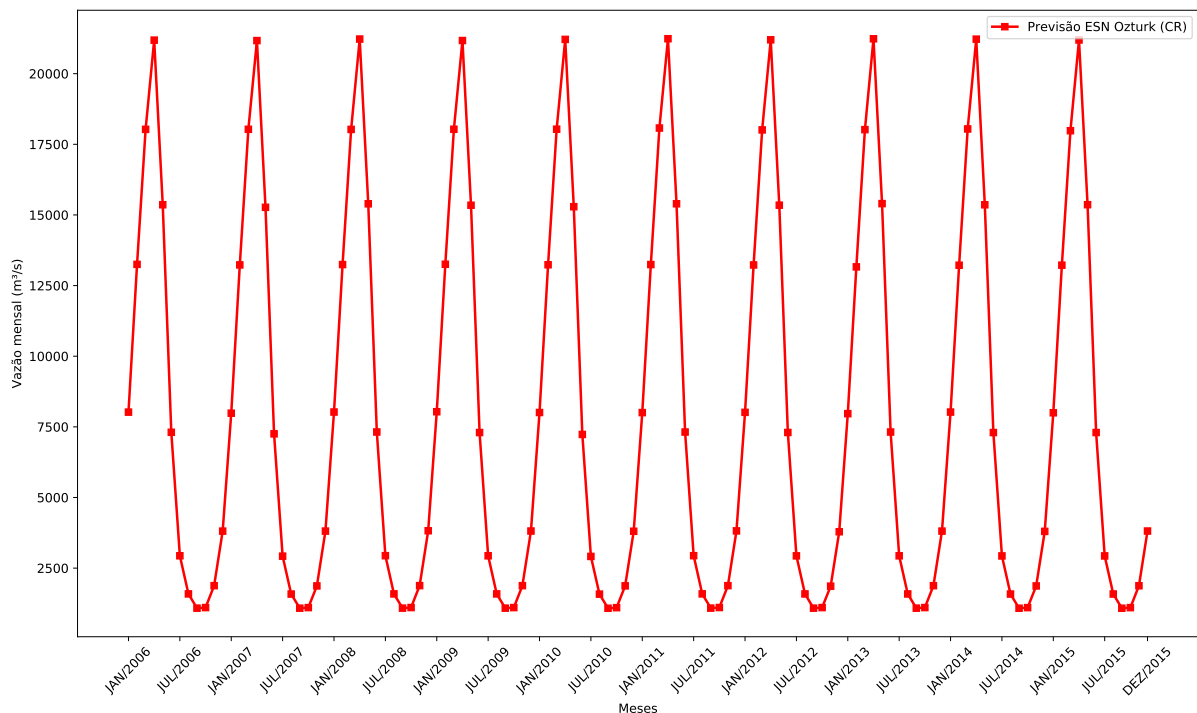
Analisando as curvas do Gráfico 27 fica evidente que o melhor MSE começa com um crescimento expressivo, seguido por 2 crescimentos mais brandos. Por sua vez, o pior MSE

**Gráfico 27 – Evolução MSE para a usina de Belo Monte.**

**Fonte: Autoria própria.**

começa com um crescimento bem mais abrupto, e ao final uma queda.

No Gráfico 28 é apresentada apenas a curva da previsão realizada pela ESN Ozturk (CR) para a usina de Belo Monte com horizonte de previsão de 12 passos a frente.

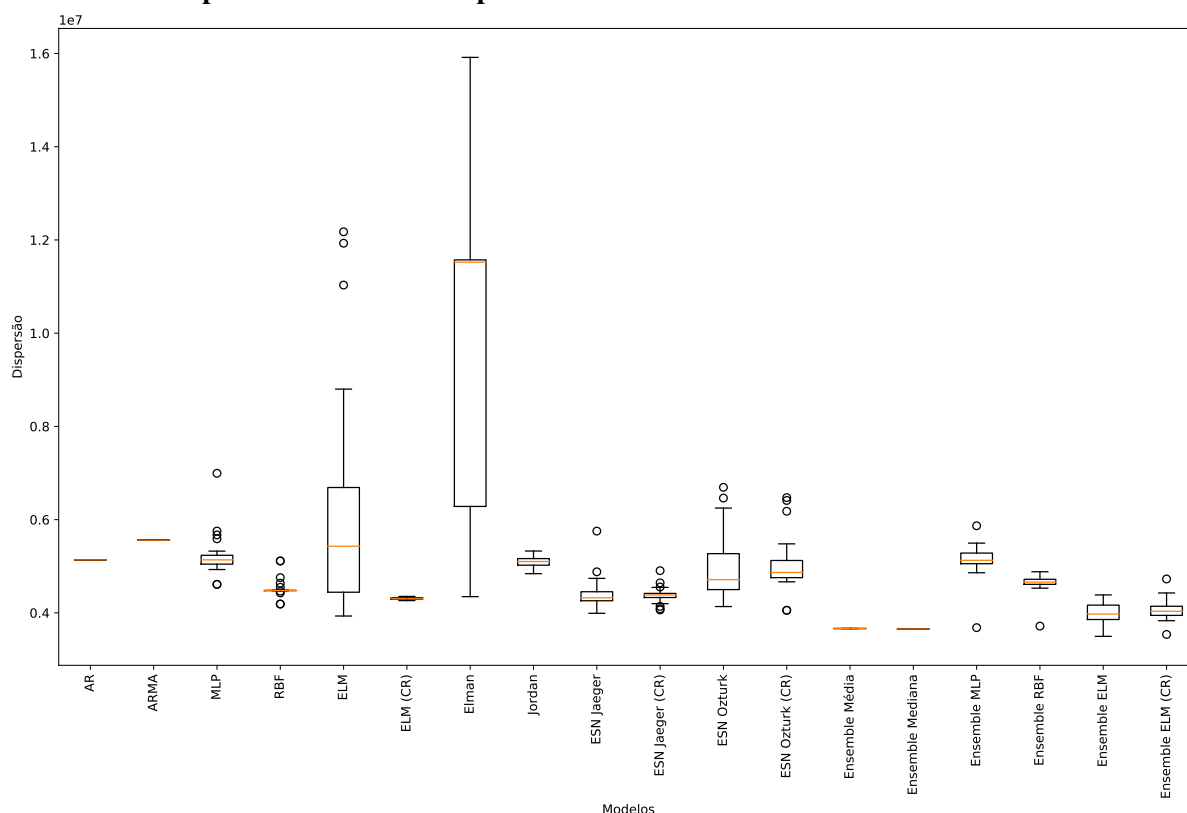
**Gráfico 28 – Padrão das previsões para Belo Monte com P = 12.**

**Fonte: Autoria própria.**

A curva do Gráfico 28 permite observar mais uma vez um mesmo comportamento referente a previsão recursiva já descrito na Seção 5.2. Conforme o horizonte de previsão avança são utilizados apenas valores já previstos como entradas para o modelo preditivo o que faz com que o comportamento aleatório da série tenda a não ser considerado quando  $P > 1$ , gerando previsões estáticas e padronizadas como as do Gráfico 28, uma tendência já conhecida e chamada de média de longo termo.

O Gráfico 29 apresenta a dispersão dos resultados obtidos pelas 30 execuções de cada um dos preditores.

**Gráfico 29 – Dispersão dos resultados para a usina de Belo Monte.**



**Fonte: Autoria própria.**

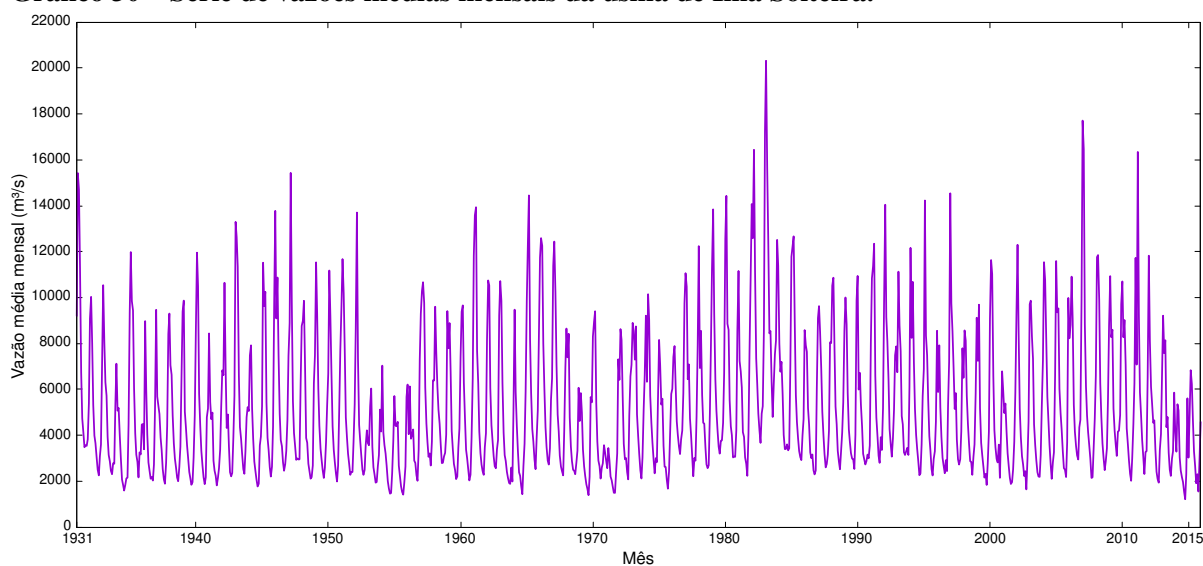
Como já era esperado, os modelos lineares (AR e ARMA) que têm seus coeficientes obtidos através de modelos determinísticos e os modelos neurais que não precisam de treinamento (*Ensemble Média* e *Ensemble Mediana*) não apresentaram qualquer tipo de dispersão.

No Gráfico 29 é evidente que o modelo que apresenta a maior dispersão é a rede de Elman, seguido pela ELM sem Coeficiente de Regularização e pela ESN Ozturk. Por outro lado, a menor dispersão foi apresentada pela RBF, seguida pela rede ELM (CR) e pela rede ESN Jaeger (CR).

## 5.4 USINA HIDRELÉTRICA DE ILHA SOLTEIRA

A usina hidrelétrica de Ilha Solteira é a maior do Estado de São Paulo e a sexta maior usina hidrelétrica do Brasil. É abastecida pelo Rio Paraná e fica localizada entre os municípios de Ilha Solteira e Selvíria, na divisa dos estados de São Paulo e Mato Grosso do Sul. Sua série histórica apresenta média  $\hat{\mu} = 5.281,94 \text{ m}^3/\text{s}$  e desvio padrão  $\hat{\sigma} = 3.100,62 \text{ m}^3/\text{s}$ , sendo apresentada no Gráfico 30.

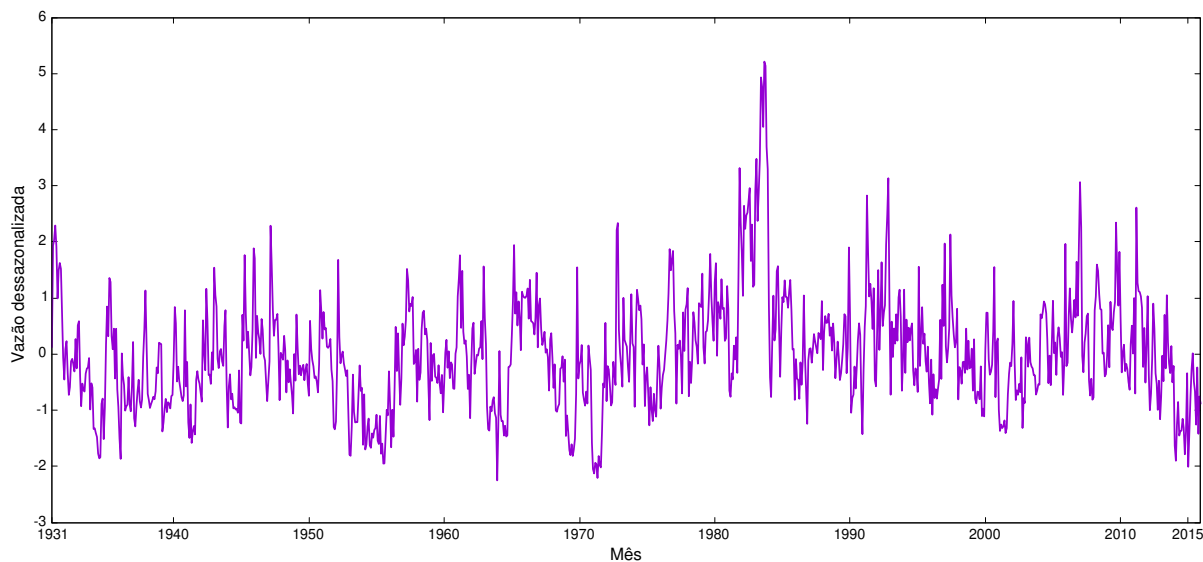
**Gráfico 30 – Série de vazões médias mensais da usina de Ilha Solteira.**



**Fonte: Autoria própria.**

Novamente se faz necessária a retirada da componente sazonal da série para um melhor entendimento do comportamento da mesma. O Gráfico 31 mostra a série dessazonalizada de acordo com a Equação 2.13.

**Gráfico 31 – Série dessazonalizada da usina de Ilha Solteira.**



**Fonte: Autoria própria.**

A Tabela 11 contém os resultados das previsões realizadas por todos os modelos propostos para a usina de Ilha Solteira com horizonte de previsão de 1 passo a frente. Também são apresentados na tabela o número de neurônios selecionado para cada modelo preditivo juntamente com o valor de todas as métricas de erro calculadas.

**Tabela 11 – Resultados previsões Ilha Solteira de 2006 a 2015 para  $P = 1$ .**

<b>Modelo</b>	<b>Neur</b>	<b>MSE</b>	<b>MSE (d)</b>	<b>MAE</b>	<b>MAE (d)</b>	<b>MAPE</b>
AR(6)	0	2.966.223,64	0,5451	1.038,59	0,5587	18,31%
ARMA(2,2)	0	2.820.456,34	0,5311	1.018,31	0,5519	17,98%
MLP	20	2.673.739,64	0,5366	1.005,22	0,5611	18,72%
RBF	75	2.646.716,44	0,5745	1.023,36	0,5887	18,96%
ELM	30	2.545.002,11	0,4973	962,84	0,5286	17,19%
ELM (CR)	95	2.498.945,94	0,5240	1.008,87	0,5714	18,25%
Elman	5	2.689.735,15	0,5713	1.011,99	0,5878	18,37%
Jordan	10	2.722.458,65	0,5524	996,08	0,5676	18,88%
ESN Jaeger	5	2.511.512,57	0,4914	994,33	0,5399	17,52%
ESN Jaeger (CR)	10	2.568.883,38	0,4999	986,43	0,5426	17,53%
ESN Ozturk	20	2.337.398,44	0,4663	944,79	0,5213	17,00%
ESN Ozturk (CR)	60	2.313.534,09	0,4768	944,94	0,5324	17,20%
<i>Ensemble Média</i>	0	2.277.545,05	0,4593	925,48	0,5137	16,63%
<i>Ensemble Mediana</i>	0	2.313.534,09	0,4768	944,94	0,5324	17,20%
<i>Ensemble MLP</i>	15	2.312.233,01	0,4945	933,62	0,5404	17,46%
<i>Ensemble RBF</i>	55	2.006.994,79	0,4903	904,30	0,5411	17,56%
<i>Ensemble ELM</i>	40	2.084.515,90	0,4546	891,63	0,5151	16,46%
<b><i>Ensemble ELM (CR)</i></b>	<b>50</b>	<b>1.992.611,39</b>	<b>0,4491</b>	<b>865,73</b>	<b>0,5093</b>	<b>16,21%</b>

**Fonte: Autoria própria.**

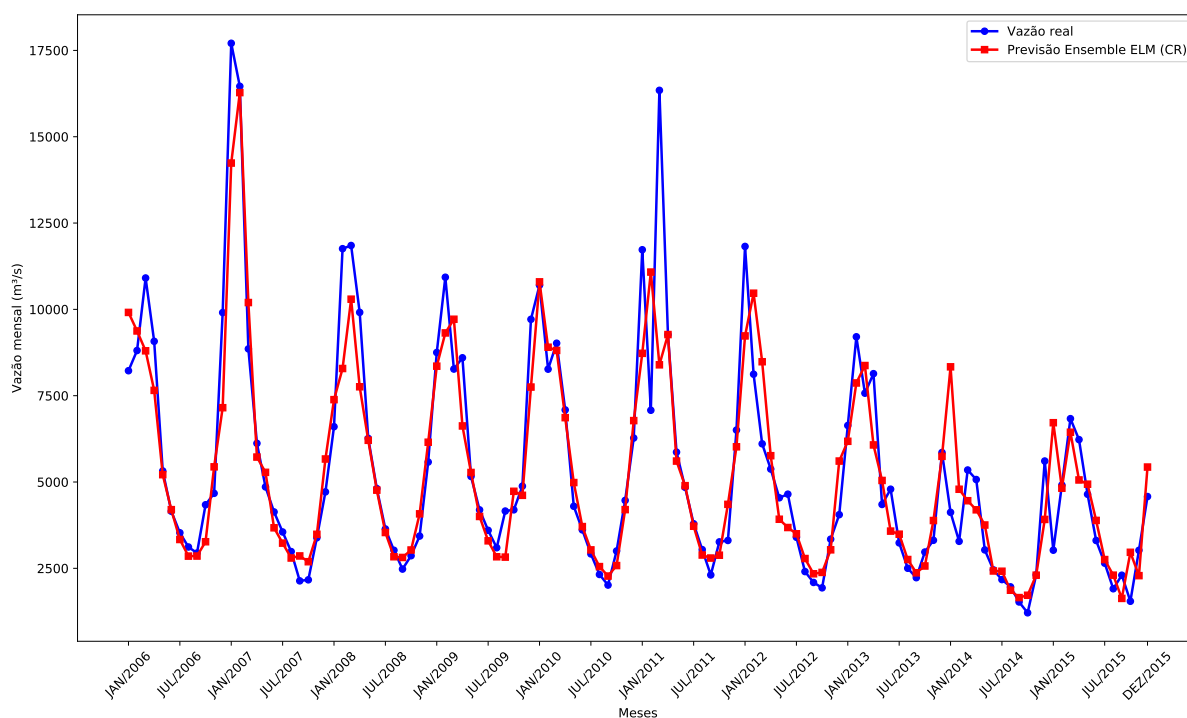
Pelos dados da Tabela 11 nota-se que novamente assim como ocorreu com as usinas de Água Vermelha (Seção 5.2) e Belo Monte (Seção 5.3) com  $P = 1$ , o melhor modelo de previsão foi um *Ensemble*, sendo neste caso com combinador de ELM e utilizando Coeficiente de Regularização, que obteve um MSE real de 1.992.611,39. Por sua vez, o pior desempenho foi mais uma vez do modelo linear Auto-regressivo, suas previsões alcançaram MSE real de 2.966.223,64.

Analisando o ambiente dessazonalizado, tem-se que o modelo preditivo que apresentou as melhores previsões foi também o *Ensemble ELM (CR)*, com MSE dessazonalizado de 0,4491. Por sua vez, as previsões com pior desempenho nesse domínio foram as da rede RBF, com MSE dessazonalizado de 0,5745.

Ainda olhando para as outras métricas de erro apresentadas na Tabela 11 é possível notar que o *Ensemble ELM (CR)* teve as menores taxas de erro em todas as métricas calculadas. Vale ressaltar aqui que nos 4 horizontes de previsão testados tanto para Água Vermelha (Seção 5.2) quanto para Belo Monte (Seção 5.3) ainda não havia ocorrido de um modelo ser o melhor em todas as métricas de erro.

O Gráfico 32 apresenta a melhor previsão realizada para a usina de Ilha Solteira com horizonte de  $P = 1$ .

Gráfico 32 – Melhor previsão Ilha Solteira de 2006 a 2015 – 1 passo a frente.



Fonte: Autoria própria.

Na Tabela 12 são apresentados os resultados das simulações dos modelos propostos para a usina de Ilha Solteira com horizonte de 3 passos a frente.

Tabela 12 – Resultados previsões Ilha Solteira de 2006 a 2015 para P = 3.

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(6)	0	3.607.452,30	0,8363	1.260,98	0,7240	23,54%
ARMA(2,2)	0	3.575.263,58	0,8337	1.248,26	0,7173	22,94%
MLP	20	3.385.476,99	0,7607	1.179,97	0,6921	22,62%
RBF	75	3.910.685,33	0,9068	1.323,71	0,7815	26,10%
ELM	30	3.316.801,94	0,7265	1.163,04	0,6709	21,72%
ELM (CR)	95	3.445.130,56	0,7750	1.200,88	0,6987	22,71%
Elman	5	3.719.407,93	0,8549	1.197,06	0,7227	22,51%
Jordan	10	3.557.423,14	0,8219	1.244,23	0,7373	25,44%
ESN Jaeger	5	3.259.442,70	0,7219	1.183,22	0,6734	21,61%
ESN Jaeger (CR)	10	3.294.093,73	0,7489	1.141,20	0,6682	20,95%
ESN Ozturk	20	3.428.320,79	0,7850	1.182,33	0,7004	22,03%
ESN Ozturk (CR)	60	3.646.528,58	0,8801	1.201,16	0,7182	22,38%
Ensemble Média	0	3.316.183,09	0,7599	1.155,80	0,6714	21,23%
Ensemble Mediana	0	3.230.939,25	0,7472	1.149,71	0,6697	21,01%
<b>Ensemble MLP</b>	<b>15</b>	<b>3.196.000,87</b>	<b>0,7135</b>	<b>1.142,92</b>	<b>0,6617</b>	<b>21,29%</b>
Ensemble RBF	55	3.468.793,60	0,8388	1.213,91	0,7137	22,96%
Ensemble ELM	40	3.377.696,82	0,7866	1.179,89	0,6866	21,28%
Ensemble ELM (CR)	50	3.265.153,09	0,7606	1.172,55	0,6850	21,16%

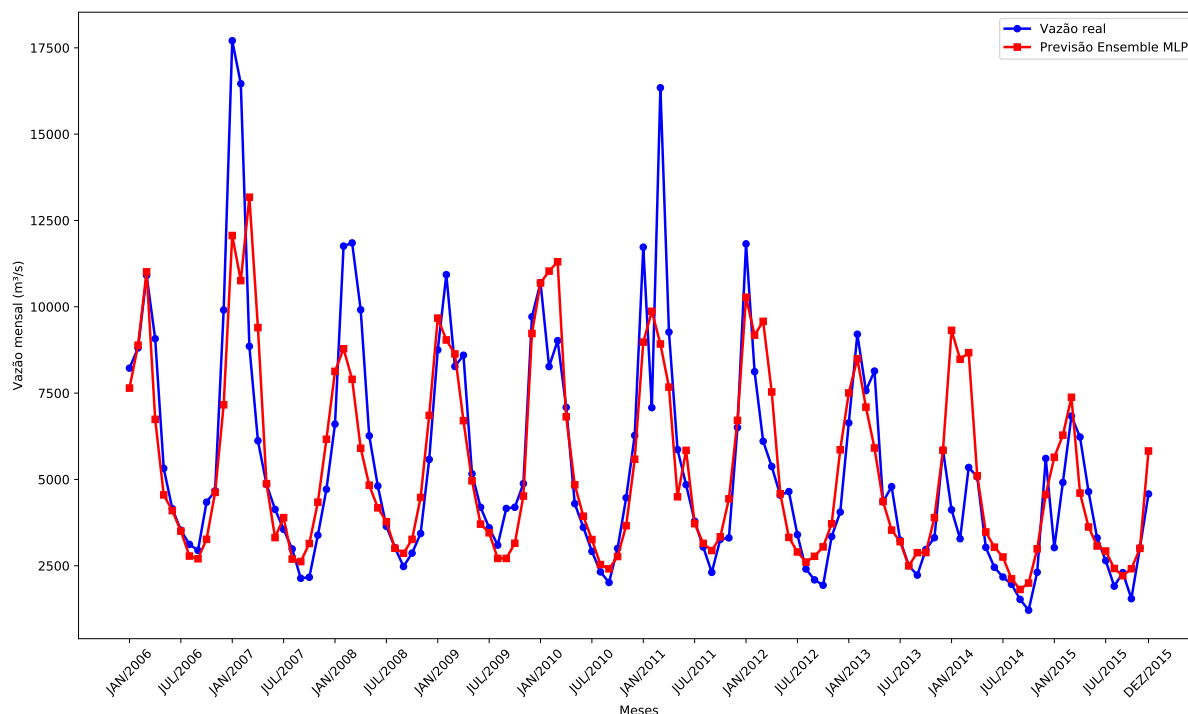
Fonte: Autoria própria.

Analisando os desempenhos, vê-se que tanto para o MSE real quanto para o dessazonalizado, o modelo com melhor desempenho foi o *Ensemble* com MLP. De igual modo, o pior

modelo foi a RBF tanto considerando o MSE real quanto o dessazonalizado.

O Gráfico 33 apresenta a melhor previsão para a usina de Ilha Solteira com  $P = 3$ .

**Gráfico 33 – Melhor previsão Ilha Solteira de 2006 a 2015 – 3 passos a frente.**



Fonte: Autoria própria.

A Tabela 13 contém os resultados das previsões para Ilha Solteira com  $P = 6$ .

**Tabela 13 – Resultados previsões Ilha Solteira de 2006 a 2015 para  $P = 6$ .**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(6)	0	4.511.751,90	1,11	1.494,40	0,8887	28,97%
ARMA(2,2)	0	4.473.867,42	1,12	1.483,68	0,8864	28,30%
MLP	20	3.920.672,52	0,9025	1.266,84	0,7530	24,47%
RBF	75	4.307.512,83	1,01	1.353,31	0,8173	26,96%
<b>ELM</b>	<b>30</b>	<b>3.761.082,94</b>	<b>0,8480</b>	<b>1.278,29</b>	<b>0,7497</b>	<b>24,57%</b>
ELM (CR)	95	3.823.661,64	0,8934	1.275,38	0,7536	24,94%
Elman	5	4.154.039,41	0,9698	1.281,16	0,7740	24,69%
Jordan	10	4.169.869,49	0,9662	1.312,63	0,7865	26,50%
ESN Jaeger	5	3.838.690,48	0,8814	1.316,70	0,7691	24,77%
ESN Jaeger (CR)	10	3.884.498,96	0,8727	1.230,27	0,7237	22,80%
ESN Ozturk	20	3.938.797,73	0,9629	1.295,02	0,7804	24,50%
ESN Ozturk (CR)	60	3.864.185,96	0,9479	1.269,58	0,7758	24,36%
Ensemble Média	0	4.188.451,09	1,05	1.389,50	0,8303	25,91%
Ensemble Mediana	0	4.131.637,37	1,06	1.389,14	0,8334	26,03%
Ensemble MLP	15	3.982.036,00	0,9874	1.348,78	0,8023	25,39%
Ensemble RBF	55	3.934.481,31	1,04	1.334,20	0,8054	24,93%
Ensemble ELM	40	4.419.019,18	1,17	1.421,19	0,8636	26,33%
Ensemble ELM (CR)	50	4.586.375,57	1,31	1.454,34	0,8890	27,22%

Fonte: Autoria própria.



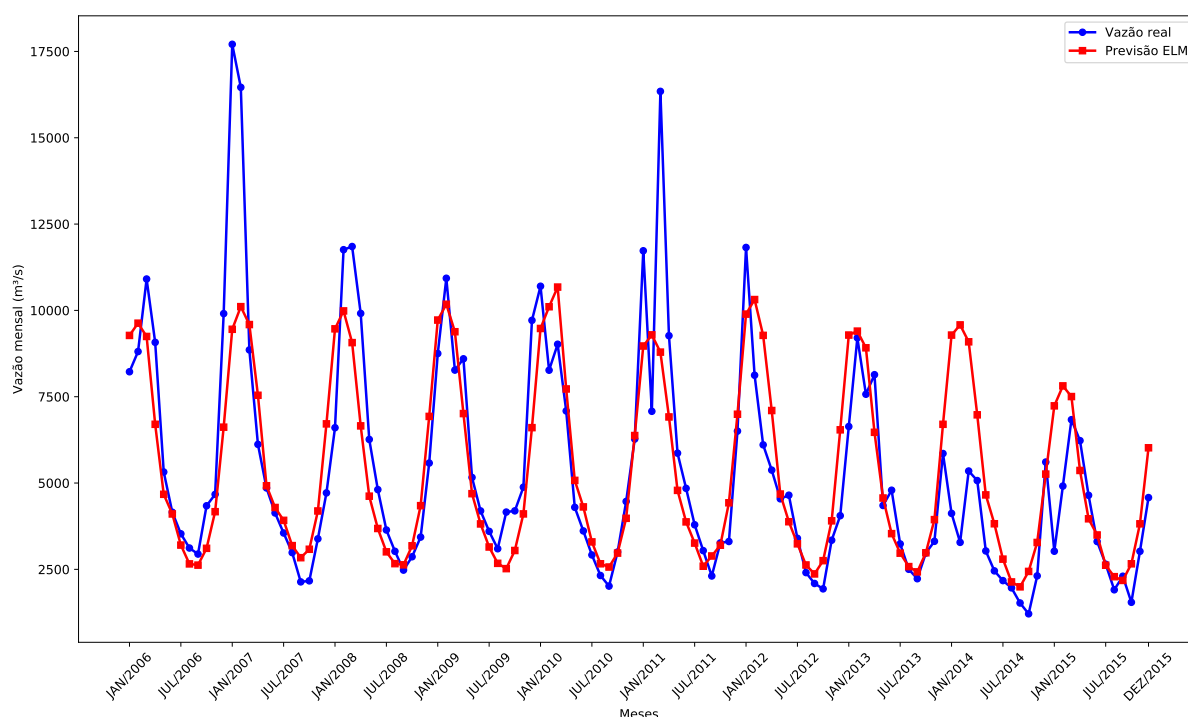
A Tabela 13 permite observar que o modelo que apresentou as melhores previsões para esse horizonte foi a rede ELM, com MSE real de 3.761.082, 94. O pior desempenho entretanto foi o do *Ensemble* com combinador de ELM e Coeficiente de Regularização que teve um erro 21% maior que o do melhor modelo.

No ambiente dessazonalizado o melhor preditor continua sendo a ELM, com MSE dessazonalizado de 0, 8480. Por outro lado, o pior desempenho foi do *Ensemble* com combinador de ELM (CR), tendo apresentado MSE dessazonalizado de 1, 31.

Analisando o MAE como sendo a métrica de erro predominante, destaca-se a rede neural MLP como o modelo preditivo com melhor desempenho, tendo MAE de 1.266, 84 contra 1.494, 40 do modelo Auto-regressivo, que por sua vez foi o preditor que realizou as piores previsões.

O Gráfico 34 apresenta a melhor previsão realizada para a usina de Ilha Solteira com  $P = 6$ .

**Gráfico 34 – Melhor previsão Ilha Solteira de 2006 a 2015 – 6 passos a frente.**



**Fonte: Autoria própria.**

Por fim, a Tabela 14 traz os resultados obtidos para a usina de Ilha Solteira, com horizonte de 12 passos à frente.

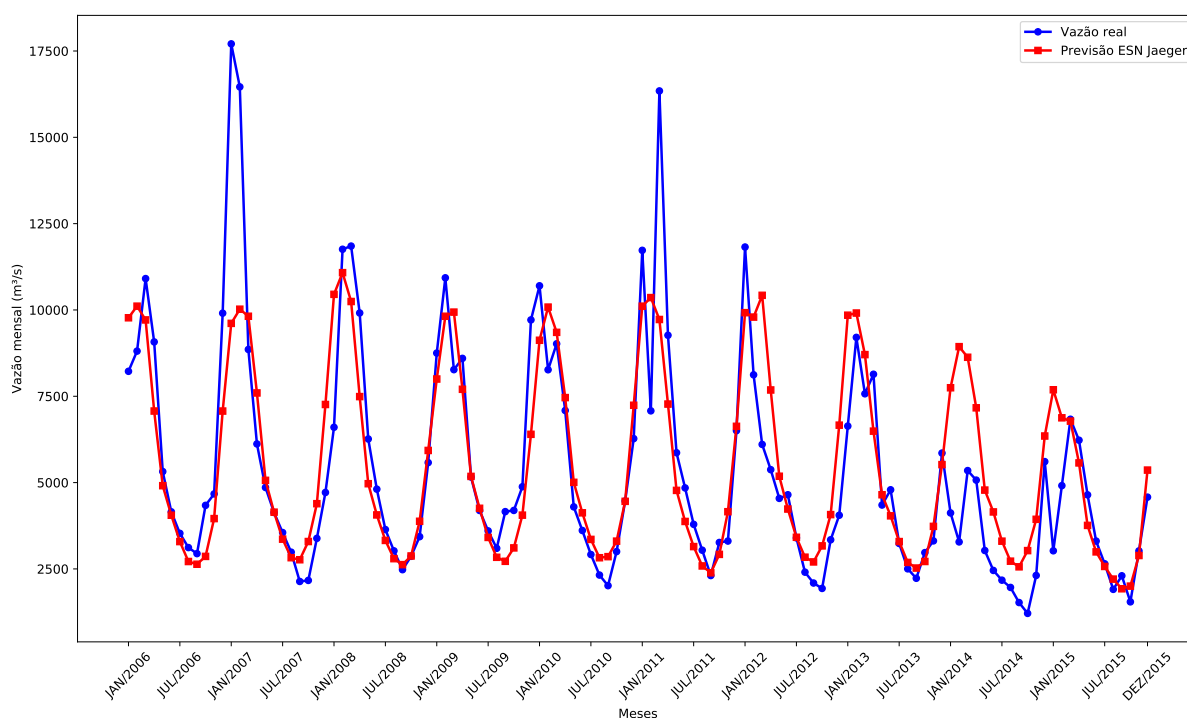
Os dados da Tabela 14 mostram que tanto para o MSE real quanto para o dessazonalizado o modelo com melhor desempenho foi a rede ESN com a proposta de reservatório de dinâmicas de Jaeger (2001) e sem a utilização do Coeficiente de Regularização. Já o pior foi o *Ensemble* ELM tendo apresentado MSE real de 6.998.311, 41, com erro 101% maior que o melhor modelo.

O Gráfico 35 mostra a melhor previsão para Ilha Solteira com  $P = 12$ .

**Tabela 14 – Resultados previsões Ilha Solteira de 2006 a 2015 para P = 12.**

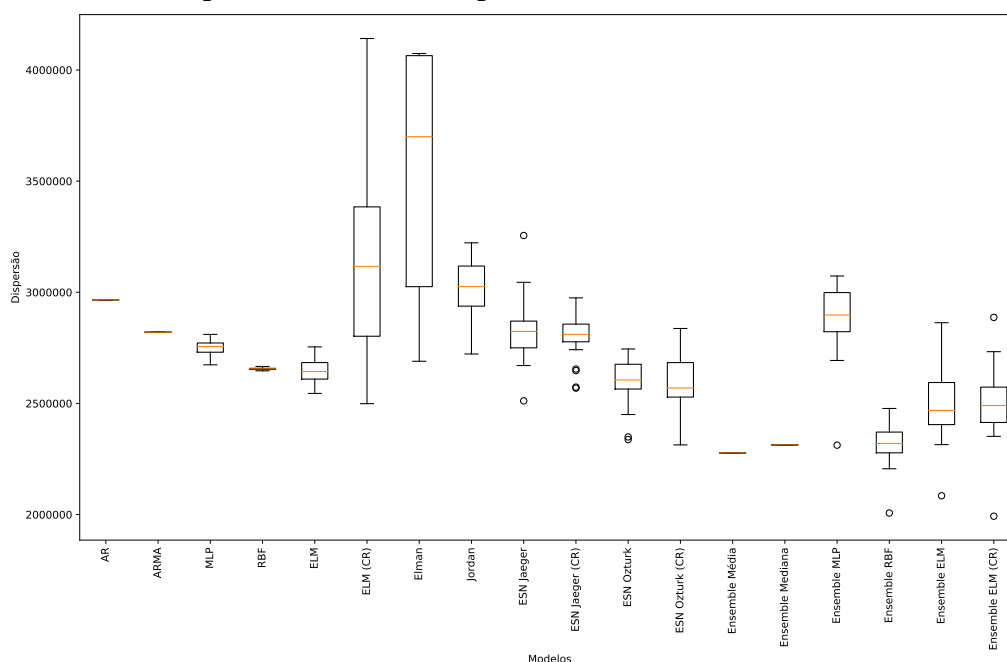
Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(6)	0	4.773.379,18	1,19	1.429,78	0,8596	30,24%
ARMA(2,2)	0	4.494.073,05	1,13	1.405,17	0,8506	29,05%
MLP	20	4.049.775,43	0,9433	1.270,28	0,7666	24,82%
RBF	75	4.283.245,08	1,11	1.319,23	0,8073	26,66%
ELM	30	3.873.269,31	0,8967	1.270,47	0,7580	25,70%
ELM (CR)	95	3.815.701,76	0,8860	1.272,35	0,7556	26,03%
Elman	5	4.116.367,32	0,9615	1.276,40	0,7717	24,67%
Jordan	10	4.070.205,28	0,9373	1.306,93	0,7801	26,42%
<b>ESN Jaeger</b>	<b>5</b>	<b>3.477.087,73</b>	<b>0,8497</b>	<b>1.219,86</b>	<b>0,7312</b>	<b>24,73%</b>
ESN Jaeger (CR)	10	3.969.786,68	0,9293	1.251,88	0,7555	24,20%
ESN Ozturk	20	3.712.781,93	0,8854	1.212,07	0,7375	23,14%
ESN Ozturk (CR)	60	4.670.725,87	0,9975	1.334,24	0,7819	25,67%
Ensemble Média	0	4.920.660,69	1,17	1.377,99	0,8255	26,78%
Ensemble Mediana	0	4.970.254,98	1,16	1.369,89	0,8211	26,80%
Ensemble MLP	15	4.739.091,94	1,09	1.333,70	0,7933	26,03%
Ensemble RBF	55	6.385.242,27	1,49	1.572,62	0,9334	30,64%
Ensemble ELM	40	6.998.311,41	2,46	1.652,36	1,02	33,01%
Ensemble ELM (CR)	50	6.612.087,53	1,45	1.489,25	0,8871	28,71%

Fonte: Autoria própria.

**Gráfico 35 – Melhor previsão Ilha Solteira de 2006 a 2015 – 12 passos a frente.**

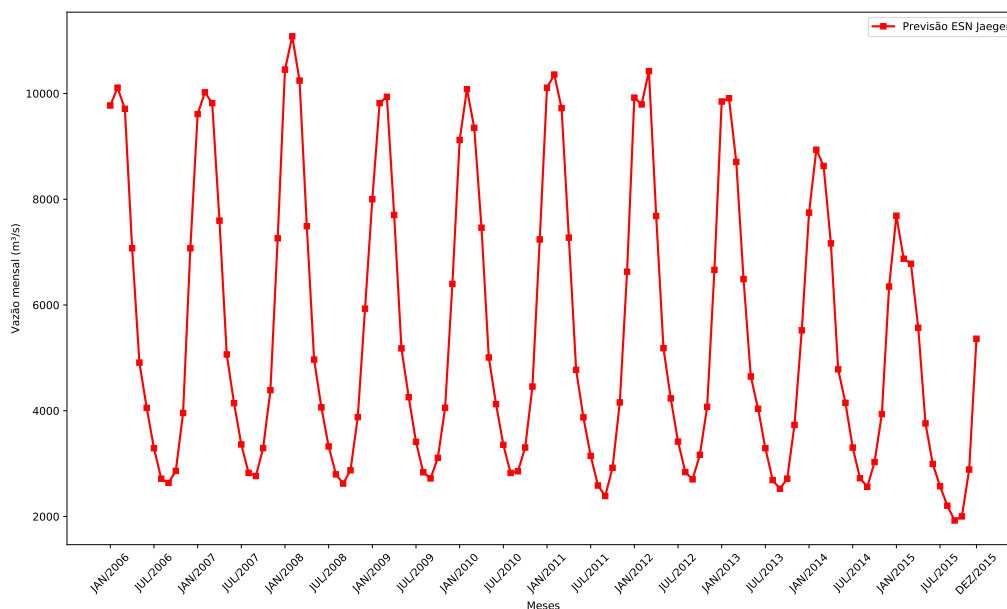
Fonte: Autoria própria.

No que diz respeito a dispersão dos resultados obtidos pelos modelos preditivos testados, tem-se o Gráfico 36 do tipo boxplot para as 30 execuções de cada preditor. Nele é possível notar que a menor dispersão foi a da rede neural MLP enquanto que a maior dispersão foi apresentada pela ELM com o uso do Coeficiente de Regularização.

**Gráfico 36 – Dispersão dos resultados para a usina de Ilha Solteira.**

**Fonte: Autoria própria.**

O Gráfico 37 apresenta apenas as previsões realizadas para Ilha Solteira com  $P = 12$ .

**Gráfico 37 – Padrão das previsões para Ilha Solteira com  $P = 12$ .**

**Fonte: Autoria própria.**

A curva presente no Gráfico 37 ressalta mais uma vez o comportamento cíclico e estático das previsões recursivas com horizontes mais elevados. No gráfico é possível verificar um padrão de comportamento da média de longo termo, apesar de um pico ter se destacado entre janeiro e julho de 2008 e a partir de 2014 haver uma diminuição no tamanho dos picos com 2015 apresentando um valor ainda menor que 2014. Apesar dos detalhes mencionados, no Gráfico 37 ainda é possível ver o mesmo comportamento dos gráficos 20 e 28.

Os testes realizados permitem verificar mais uma vez a propagação do erro de previsão conforme o horizonte de previsão avança. Para detalhar esse comportamento a Tabela 15 mostra o melhor e o pior MSE real para cada horizonte de previsão, juntamente com seu aumento percentual em relação ao horizonte anterior e em relação a  $P = 1$ .

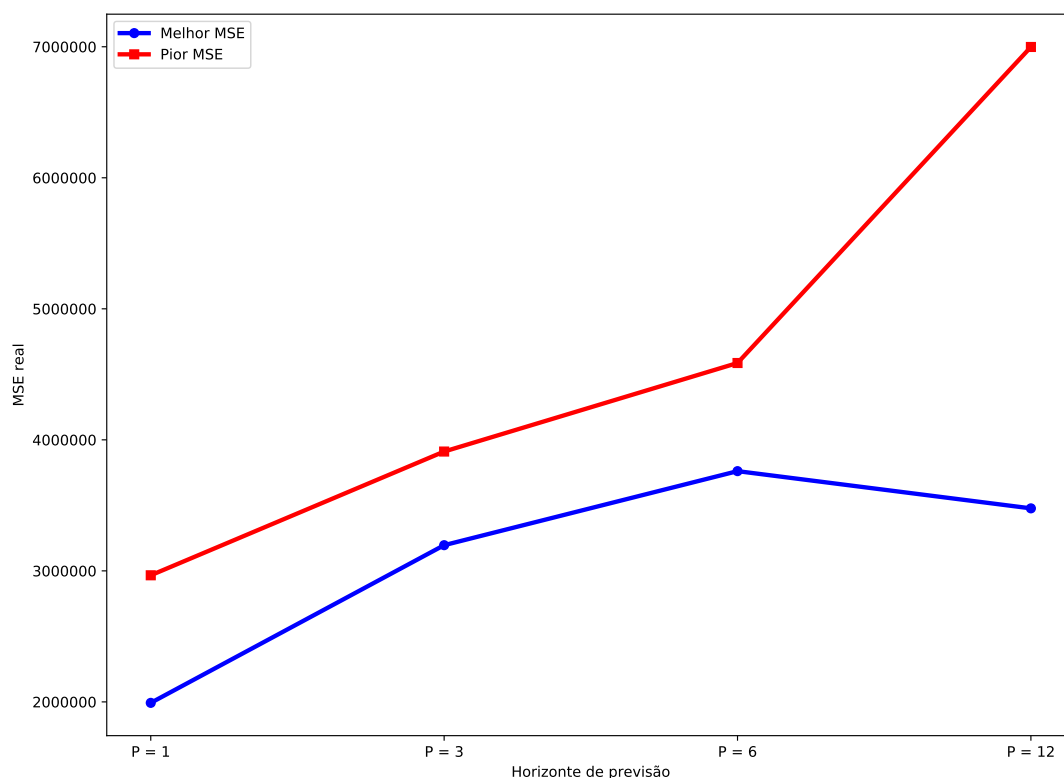
**Tabela 15 – Evolução MSE Ilha Solteira.**

Horizonte	Melhor MSE	Anterior	P = 1	Pior MSE	Anterior	P = 1
$P = 1$	1.992.611,39	–	–	2.966.223,64	–	–
$P = 3$	3.196.000,87	60,39%	60,39%	3.910.685,33	31,84%	31,84%
$P = 6$	3.761.082,94	17,68%	88,75%	4.586.375,57	17,28%	54,62%
$P = 12$	3.477.087,73	-7,55%	74,50%	6.998.311,41	52,59%	135,93%

**Fonte: Autoria própria.**

Pelos dados da Tabela 15 é possível verificar que o maior aumento do melhor MSE foi do horizonte  $P = 1$  para  $P = 3$ , com um elevação percentual de 60,39%. Além disso, também é possível ver que para o pior MSE o maior aumento foi de  $P = 6$  para  $P = 12$ , um incremento de 52,59%. Com o objetivo de complementar a interpretação desses dados o Gráfico 38 mostra as curvas da evolução desses erros.

**Gráfico 38 – Evolução MSE para a usina de Ilha Solteira.**



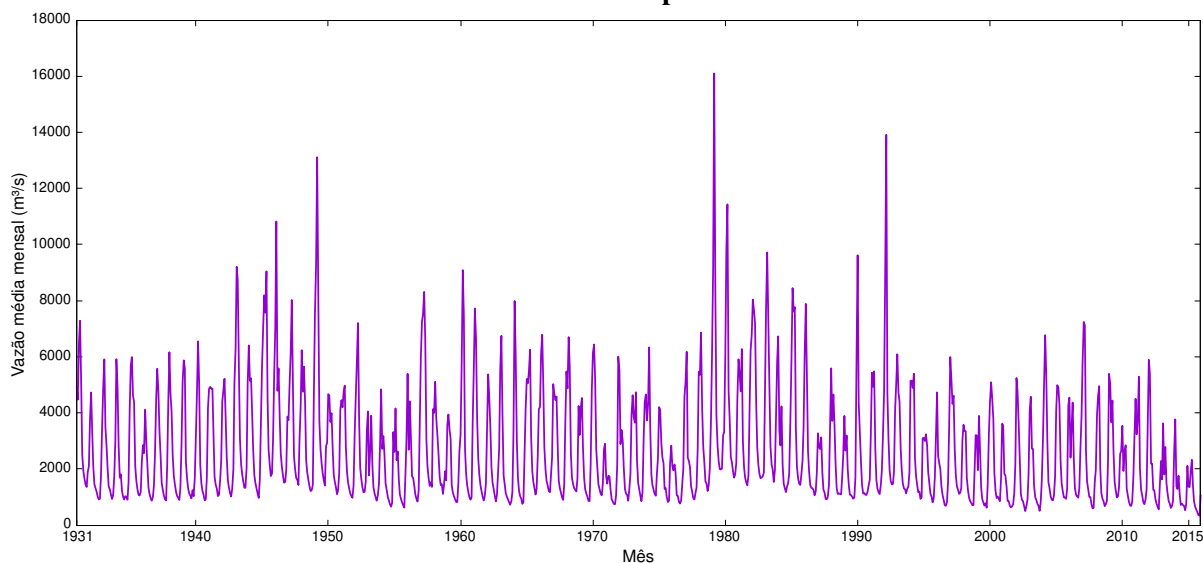
**Fonte: Autoria própria.**

É fácil perceber que o pior MSE segue crescente por todos os horizontes de previsão. Já o melhor MSE segue crescente até  $P = 6$  e, ao avançar de 6 para 12 passos à frente, o melhor MSE tem uma diminuição de 7,55%, o que faz com que o maior erro não coincida com o maior horizonte de previsão testado.

## 5.5 COMPLEXO HIDRELÉTRICO DE PAULO AFONSO

O Complexo Hidrelétrico de Paulo Afonso é um conjunto de usinas localizado na cidade de Paulo Afonso, na Bahia, formado pelas usinas de Paulo Afonso I, II, III, IV e Apolônio Sales. Trata-se de um complexo pois as 5 usinas fazem uso das águas da Cachoeira de Paulo Afonso, um desnível natural de 80 metros do Rio São Francisco. Sua série histórica tem média  $\hat{\mu} = 2.698,80 \text{ m}^3/\text{s}$  e desvio padrão de  $\hat{\sigma} = 2.026,48 \text{ m}^3/\text{s}$ , sendo apresentada no Gráfico 39.

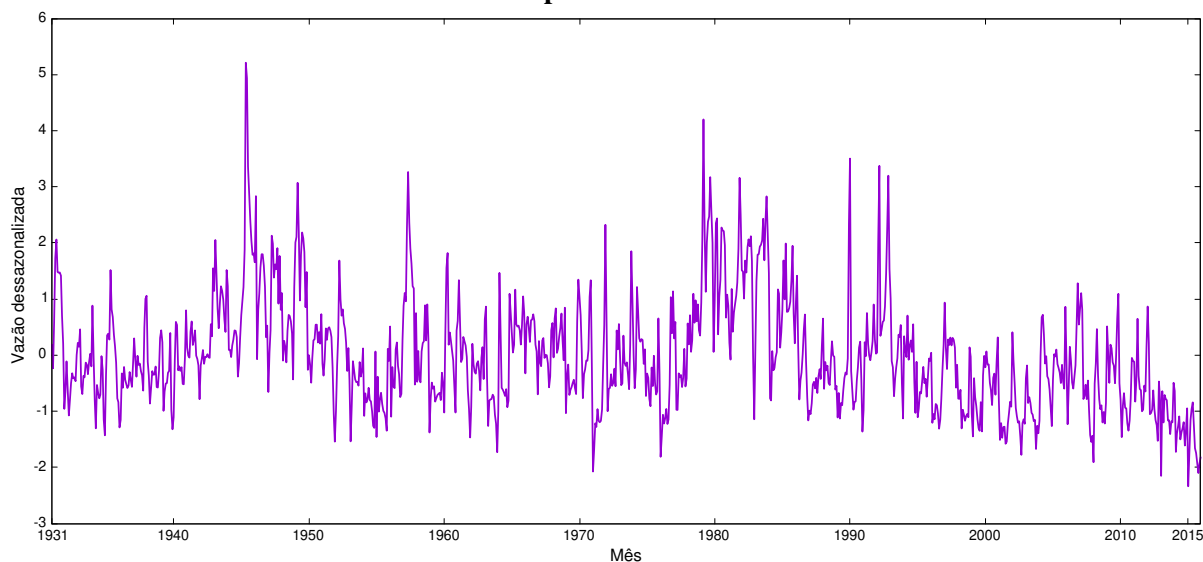
**Gráfico 39 – Série de vazões médias mensais do Complexo Hidrelétrico de Paulo Afonso.**



**Fonte: Autoria própria.**

Do mesmo modo que ocorreu com as usinas anteriores, aqui também se faz necessário remover a componente sazonal da série antes que a mesma possa ser utilizada. O Gráfico 40 mostra a série dessazonalizada da usina de Paulo Afonso.

**Gráfico 40 – Série dessazonalizada do Complexo Hidrelétrico de Paulo Afonso.**



**Fonte: Autoria própria.**

Na Tabela 16 são apresentados os resultados obtidos pelas melhores previsões realizadas por todos os modelos propostos para a usina de Paulo Afonso com horizonte de previsão de 1 passo a frente. A tabela mostra a quantidade de neurônios selecionada para cada modelo de previsão, além disso, também são apresentados os valores calculados de todas as métricas de erro para cada preditor.

**Tabela 16 – Resultados previsões Paulo Afonso de 2006 a 2015 para P = 1.**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(6)	0	726.145,70	0,2931	533,05	0,4194	26,22%
ARMA(2,1)	0	694.185,05	0,2874	513,10	0,4148	25,52%
MLP	40	635.673,56	0,3437	537,02	0,4806	32,29%
RBF	10	719.377,51	0,4103	561,22	0,5255	35,11%
ELM	5	602.176,48	0,3312	520,77	0,4725	30,30%
ELM (CR)	35	607.078,51	0,3237	522,82	0,4635	30,68%
Elman	5	617.664,77	0,2869	498,78	0,4255	26,78%
Jordan	5	673.001,75	0,3398	545,15	0,4784	32,11%
ESN Jaeger	75	566.929,67	0,2786	484,71	0,4149	26,22%
ESN Jaeger (CR)	45	627.573,85	0,3415	529,75	0,4716	32,42%
ESN Ozturk	80	576.338,26	0,2957	496,77	0,4369	28,30%
ESN Ozturk (CR)	80	581.609,27	0,3095	513,41	0,4532	29,58%
<i>Ensemble Média</i>	0	551.531,61	0,2908	487,33	0,4315	28,42%
<i>Ensemble Mediana</i>	0	552.252,81	0,2521	460,89	0,3786	23,17%
<i>Ensemble MLP</i>	35	548.535,13	0,2931	475,77	0,4202	27,17%
<i>Ensemble RBF</i>	10	554.508,82	0,2986	495,05	0,4391	29,85%
<b><i>Ensemble ELM</i></b>	<b>45</b>	<b>535.210,83</b>	<b>0,2714</b>	<b>472,88</b>	<b>0,4062</b>	<b>26,12%</b>
<i>Ensemble ELM (CR)</i>	55	535.865,26	0,3016	488,17	0,4439	29,57%

**Fonte: Autoria própria.**

Os resultados apresentados pela Tabela 16 indicam que o melhor modelo de previsão foi o *Ensemble* com combinador de ELM sem Coeficiente de Regularização, tendo obtido MSE real de 535.210,83. Já o pior preditor foi o modelo linear Auto-Regressivo, com MSE real de 726.145,70, um resultado 35% pior que o do *Ensemble ELM*.

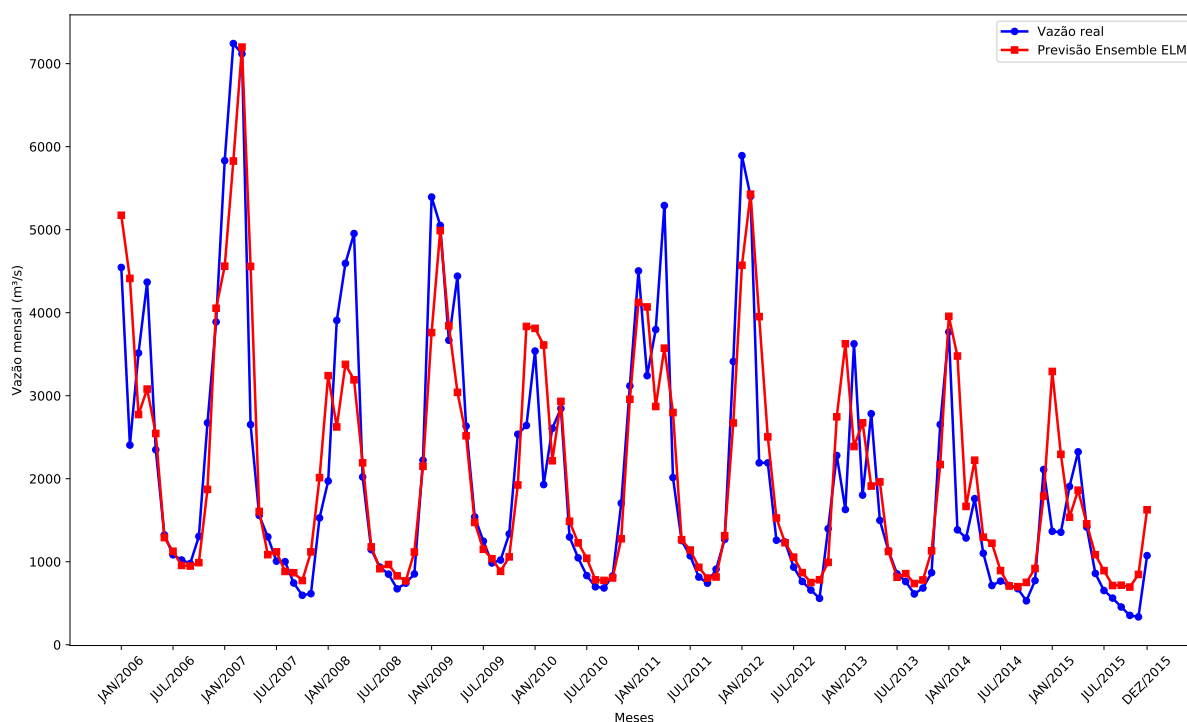
No ambiente dessazonalizado entretanto, o modelo que apresentou o melhor desempenho nas previsões realizadas foi o *Ensemble* com combinador de Mediana, com MSE dessazonalizado de 0,2521. Por outro lado o preditor que realizou as piores previsões sob o viés do MSE dessazonalizado foi a rede neural RBF, tendo um resultado 62,75% pior que o do *Ensemble Mediana*.

Analisando mais cuidadosamente os dados contidos na Tabela 16 verifica-se que dentre todas as métricas de erro calculadas, apenas para o MSE real o melhor preditor é o *Ensemble ELM*, para todas as outras métricas consideradas o modelo com o melhor resultado foi o *Ensemble Mediana*.

O Gráfico 41 apresenta a melhor previsão realizada para a usina de Paulo Afonso com horizonte de previsão de 1 passo a frente.

A Tabela 17 contém os resultados obtidos por todos os modelos de previsão para a usina

Gráfico 41 – Melhor previsão Paulo Afonso de 2006 a 2015 – 1 passo a frente.



Fonte: Autoria própria.

de Paulo Afonso com horizonte de previsão de  $P = 3$ .

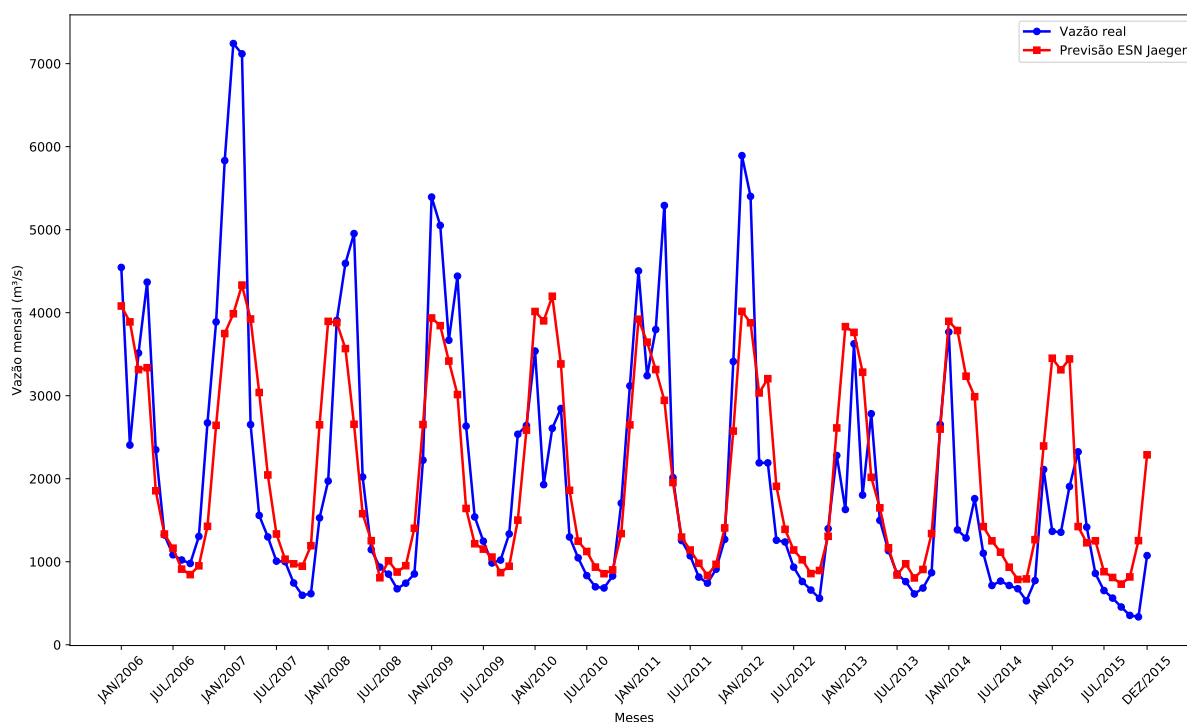
Tabela 17 – Resultados previsões Paulo Afonso de 2006 a 2015 para  $P = 3$ .

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(6)	0	1.129.419,58	0,5697	723,00	0,6383	41,33%
ARMA(2,1)	0	1.131.738,27	0,5453	710,37	0,6103	38,59%
MLP	40	1.058.546,46	0,6355	725,72	0,6728	46,77%
RBF	10	2.000.897,18	1,18	943,65	0,9218	65,31%
ELM	5	938.475,79	0,5768	694,37	0,6492	42,15%
ELM (CR)	35	1.089.089,42	0,6389	727,19	0,6774	47,27%
Elman	5	936.344,63	0,4566	651,82	0,5476	33,66%
Jordan	5	1.166.257,86	0,6914	769,75	0,7173	50,51%
<b>ESN Jaeger</b>	<b>75</b>	<b>929.239,63</b>	<b>0,5230</b>	<b>664,61</b>	<b>0,5993</b>	<b>38,90%</b>
ESN Jaeger (CR)	45	1.490.572,96	0,9615	838,66	0,8182	56,87%
ESN Ozturk	80	1.256.147,76	0,6778	712,61	0,6420	39,34%
ESN Ozturk (CR)	80	1.047.548,67	0,5800	662,61	0,6050	37,70%
Ensemble Média	0	1.230.760,56	0,5601	722,50	0,6121	40,33%
Ensemble Mediana	0	1.169.891,73	0,5154	711,51	0,5768	35,41%
Ensemble MLP	35	1.185.108,36	0,5330	696,01	0,5776	36,55%
Ensemble RBF	10	987.644,32	0,4806	665,60	0,5640	36,73%
Ensemble ELM	45	1.088.286,19	0,4940	692,60	0,5771	36,65%
Ensemble ELM (CR)	55	1.193.241,92	0,5780	711,54	0,6112	39,50%

Fonte: Autoria própria.

O melhor modelo foi a rede ESN Jaeger e o pior foi a RBF. O Gráfico 42 apresenta a melhor previsão para a usina de Paulo Afonso com  $P = 3$ .

Gráfico 42 – Melhor previsão Paulo Afonso de 2006 a 2015 – 3 passos a frente.



Fonte: Autoria própria.

Na Tabela 18 são apresentados os resultados dos modelos para a usina de Paulo Afonso com  $P = 6$ .

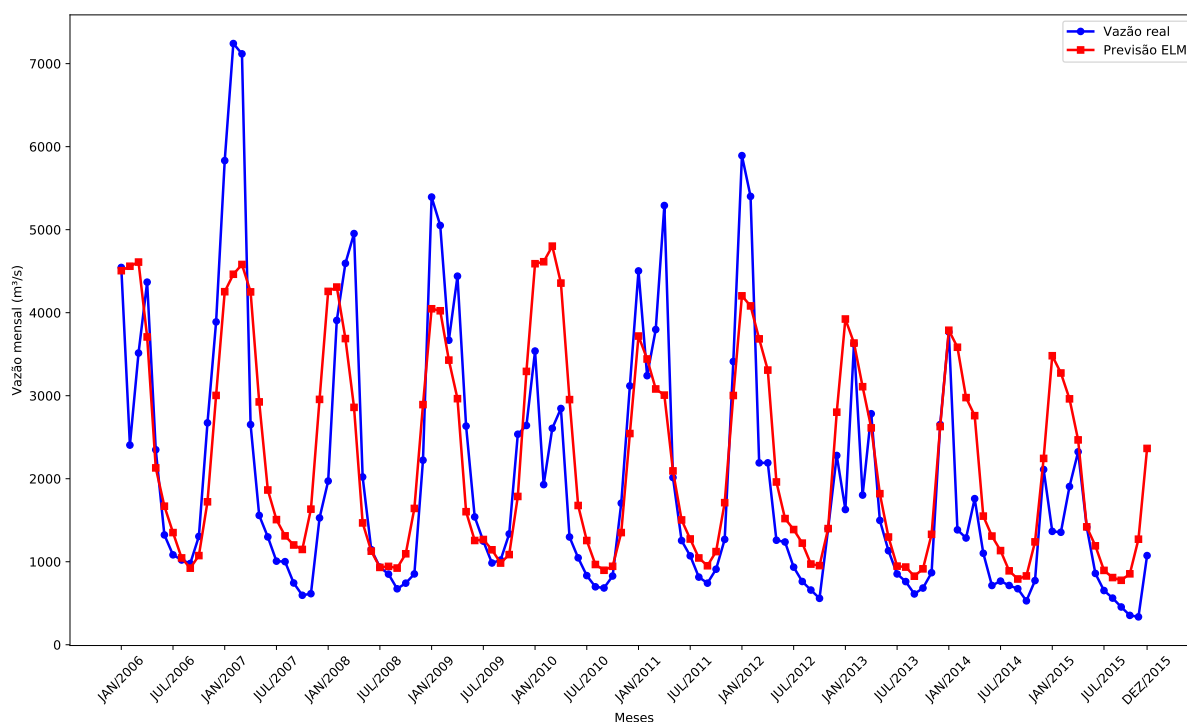
Tabela 18 – Resultados previsões Paulo Afonso de 2006 a 2015 para  $P = 6$ .

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(6)	0	1.088.288,61	0,7460	751,31	0,7244	47,56%
ARMA(2,1)	0	1.024.438,22	0,6806	730,15	0,6893	43,93%
MLP	40	1.113.602,56	0,6855	743,63	0,6962	48,67%
RBF	10	3.779.964,35	2,60	1.450,49	1,44	100,82%
<b>ELM</b>	<b>5</b>	<b>980.510,07</b>	<b>0,6357</b>	<b>709,47</b>	<b>0,6747</b>	<b>44,22%</b>
ELM (CR)	35	1.129.832,90	0,7220	769,17	0,7334	50,87%
Elman	5	981.800,01	0,5880	697,68	0,6469	44,07%
Jordan	5	1.209.656,52	0,7708	795,27	0,7561	53,36%
ESN Jaeger	75	1.106.672,85	0,6847	738,66	0,6951	48,70%
ESN Jaeger (CR)	45	1.278.004,66	0,8767	807,99	0,7848	54,77%
ESN Ozturk	80	1.040.841,92	0,6180	699,59	0,6457	44,68%
ESN Ozturk (CR)	80	1.083.715,18	0,7617	750,10	0,7336	49,67%
Ensemble Média	0	1.025.453,52	0,6182	726,01	0,6581	41,56%
Ensemble Mediana	0	1.078.563,81	0,5853	721,56	0,6340	37,07%
Ensemble MLP	35	1.063.431,44	0,5914	704,16	0,6203	37,30%
Ensemble RBF	10	1.058.247,06	0,6367	737,50	0,6734	42,36%
Ensemble ELM	45	989.046,79	0,5391	696,25	0,6081	37,26%
Ensemble ELM (CR)	55	997.954,04	0,5994	724,09	0,6520	40,47%

Fonte: Autoria própria.

Verifica-se que o melhor modelo de previsão para  $P = 6$  foi a rede ELM e o pior modelo foi a RBF. O Gráfico 43 apresenta a previsão realizada pela rede ELM.



**Gráfico 43 – Melhor previsão Paulo Afonso de 2006 a 2015 – 6 passos a frente.**

Fonte: Autoria própria.

A Tabela 19 apresenta os resultados para  $P = 12$ .

**Tabela 19 – Resultados previsões Paulo Afonso de 2006 a 2015 para  $P = 12$ .**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(6)	0	1.916.895,36	1,11	975,46	0,8989	63,04%
ARMA(2,1)	0	1.634.060,46	0,9137	867,97	0,7825	54,16%
<b>MLP</b>	<b>40</b>	<b>1.118.376,67</b>	<b>0,6888</b>	<b>744,90</b>	<b>0,6975</b>	<b>48,80%</b>
RBF	10	13.040.123,34	8,11	2.892,90	2,74	182,21%
ELM	5	1.290.254,85	0,7766	808,48	0,7497	52,64%
ELM (CR)	35	1.723.308,91	1,11	959,04	0,9214	66,08%
Elman	5	1.214.455,57	0,7626	786,67	0,7428	52,65%
Jordan	5	1.258.279,09	0,7960	802,56	0,7603	54,16%
ESN Jaeger	75	1.364.591,04	0,8514	848,93	0,7964	56,55%
ESN Jaeger (CR)	45	1.852.101,70	1,17	998,93	0,9541	67,77%
ESN Ozturk	80	1.250.520,94	0,7806	796,67	0,7508	53,28%
ESN Ozturk (CR)	80	1.348.961,53	0,8507	831,10	0,7773	54,59%
Ensemble Média	0	1.420.331,33	0,7761	782,28	0,6808	43,92%
Ensemble Mediana	0	1.334.761,67	0,7067	733,49	0,6234	38,54%
Ensemble MLP	35	1.311.740,31	0,7058	729,42	0,6247	38,82%
Ensemble RBF	10	1.393.776,65	0,7927	772,99	0,6866	44,54%
Ensemble ELM	45	1.267.630,78	0,6619	716,46	0,6123	38,03%
Ensemble ELM (CR)	55	1.346.719,34	0,7240	738,21	0,6461	41,19%

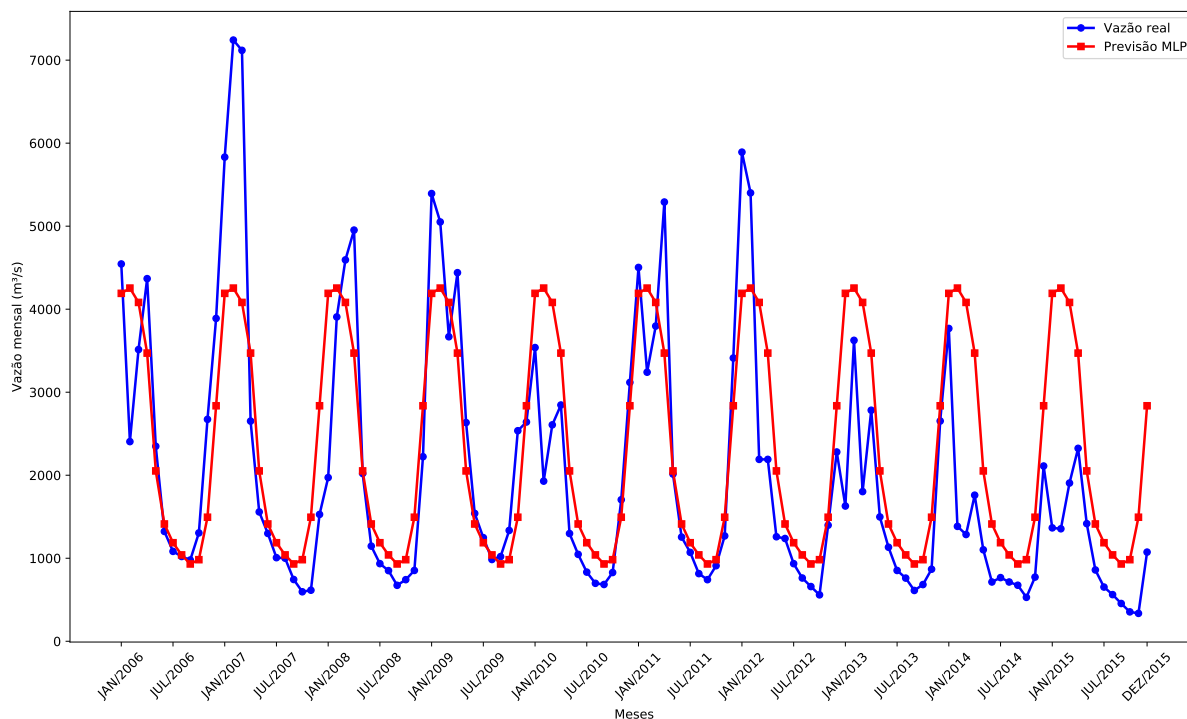
Fonte: Autoria própria.

Pelos dados da Tabela 19 é possível verificar que para o horizonte de 12 passos à frente o modelo com a melhor previsão foi a rede neural MLP. Entretanto, o maior destaque aqui é para

o pior modelo, já que as previsões da rede RBF alcançaram um MSE real de 13.040.123,34, um resultado 1.065% pior que o do melhor modelo.

O Gráfico 44 apresenta as previsões realizadas pela rede neural MLP para o conjunto de teste da usina de Paulo Afonso com horizonte de 12 passos a frente.

**Gráfico 44 – Melhor previsão Paulo Afonso de 2006 a 2015 – 12 passos a frente.**



**Fonte: Autoria própria.**

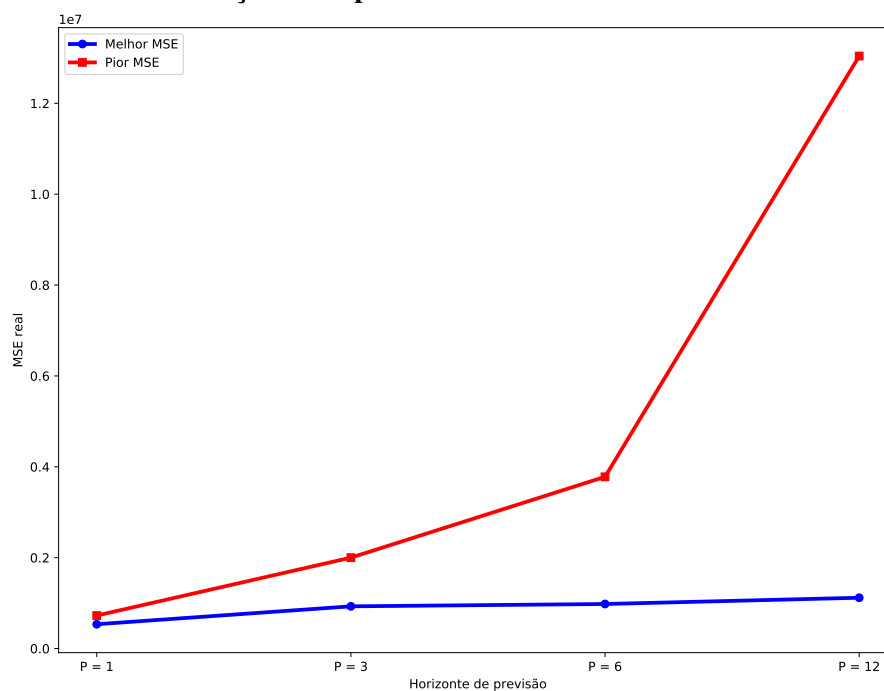
Ao analisar os dados dos resultados presentes nas tabelas 16, 17, 18 e 19 nota-se que, conforme o horizonte de previsão aumenta, também cresce o valor do MSE. A Tabela 20 apresenta o valor do melhor e do pior MSE por horizonte de previsão e também o valor percentual do aumento do MSE em relação ao horizonte anterior e em relação a  $P = 1$ .

**Tabela 20 – Evolução MSE Paulo Afonso.**

Horizonte	Melhor MSE	Anterior	P = 1	Pior MSE	Anterior	P = 1
$P = 1$	535.210,83	—	—	726.145,70	—	—
$P = 3$	929.239,63	73,62%	73,62%	2.000.897,18	175,55%	175,55%
$P = 6$	980.510,07	5,52%	83,20%	3.779.964,35	88,91%	420,55%
$P = 12$	1.118.376,67	14,06%	108,96%	13.040.123,34	244,98%	1.695,80%

**Fonte: Autoria própria.**

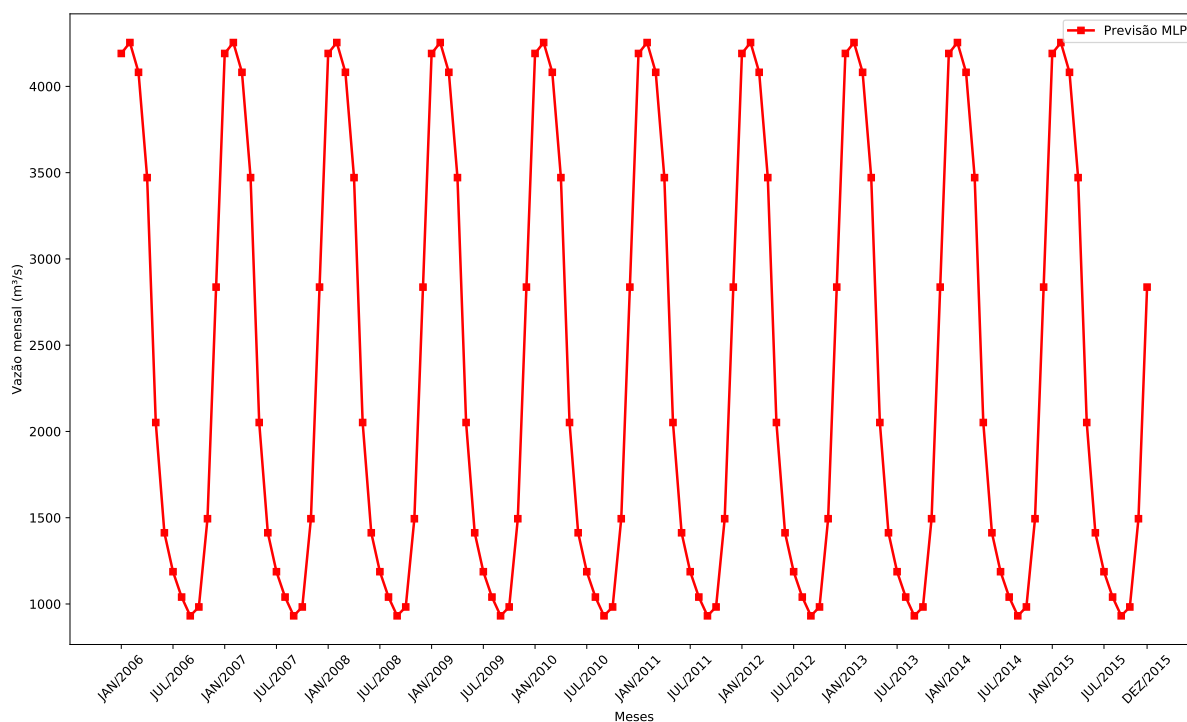
Pelos dados da Tabela 20 é possível observar que o melhor MSE aumenta a cada a horizonte de previsão, isso ocorre devido a propagação do erro pela previsão recursiva. Percebe-se também que em se tratando do melhor MSE o maior aumento foi de  $P = 1$  para  $P = 3$  com 73,62% de elevação. Já com relação ao pior MSE, o maior crescimento foi de  $P = 6$  para  $P = 12$  com 244,98% de aumento. O Gráfico 45 apresenta a evolução do melhor e do pior MSE entre os horizontes de previsão para a usina de Paulo Afonso.

**Gráfico 45 – Evolução MSE para a usina de Paulo Afonso.**

**Fonte: Autoria própria.**

As curvas do Gráfico 45 evidenciam quão expressivo foi o aumento do pior MSE de  $P = 6$  para  $P = 12$ . As duas curvas seguem crescendo conforme o horizonte avança, mas o ritmo de crescimento do melhor MSE é o menor.

O Gráfico 46 apresenta apenas as previsões realizadas para Paulo Afonso com  $P = 12$ .

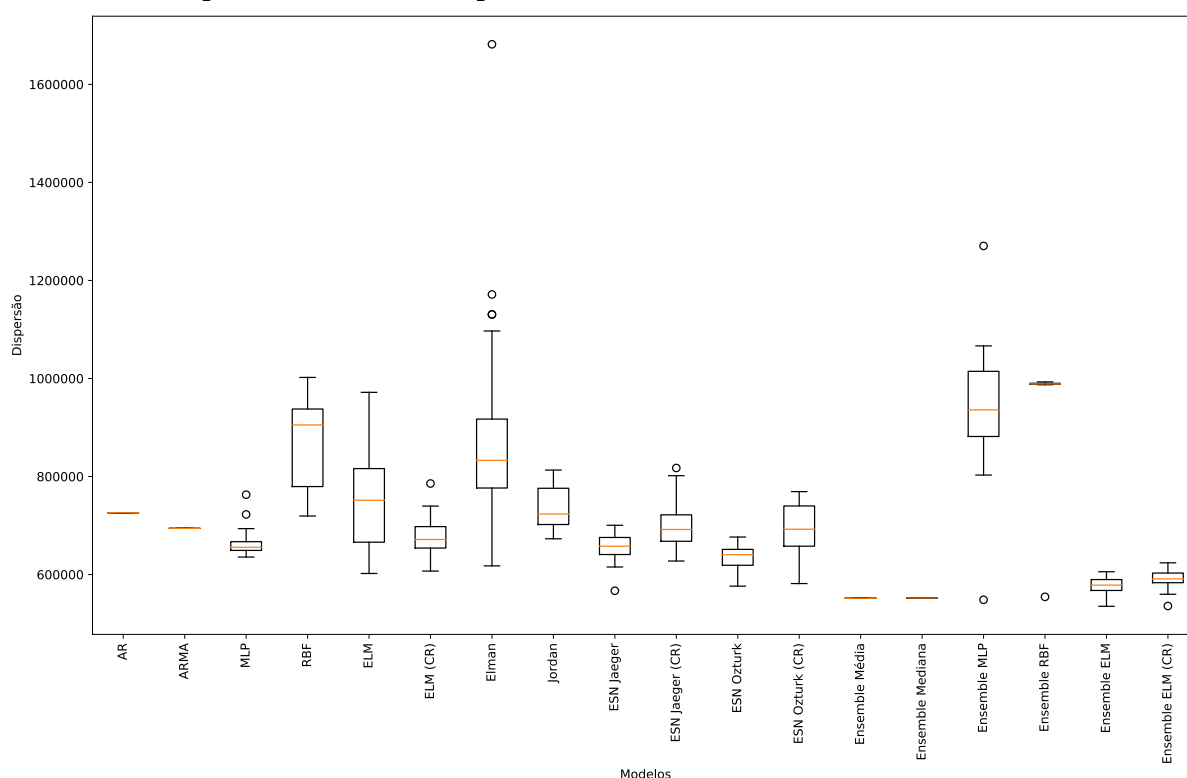
**Gráfico 46 – Padrão das previsões para Paulo Afonso com  $P = 12$ .**

**Fonte: Autoria própria.**

No Gráfico 46 destaca-se um padrão cíclico nas previsões realizadas, com picos e vales sempre nas mesmas épocas. Novamente é observado um padrão da média de longo termo. Uma vez que o horizonte de previsão aumenta, a tendência dos modelos é utilizar apenas valores previstos como entradas e não mais os valores reais, o que retira a influência dos eventos aleatórios externos nos novos valores previstos, deixando apenas o padrão base da série, que nesse caso é o nível de chuva da região.

Com relação a variação dos resultados obtidos pelos modelos de previsão, o Gráfico 47 apresenta a dispersão dos resultados das 30 execuções de cada modelo para a usina de Paulo Afonso.

**Gráfico 47 – Dispersão dos resultados para a usina de Paulo Afonso.**



**Fonte: Autoria própria.**

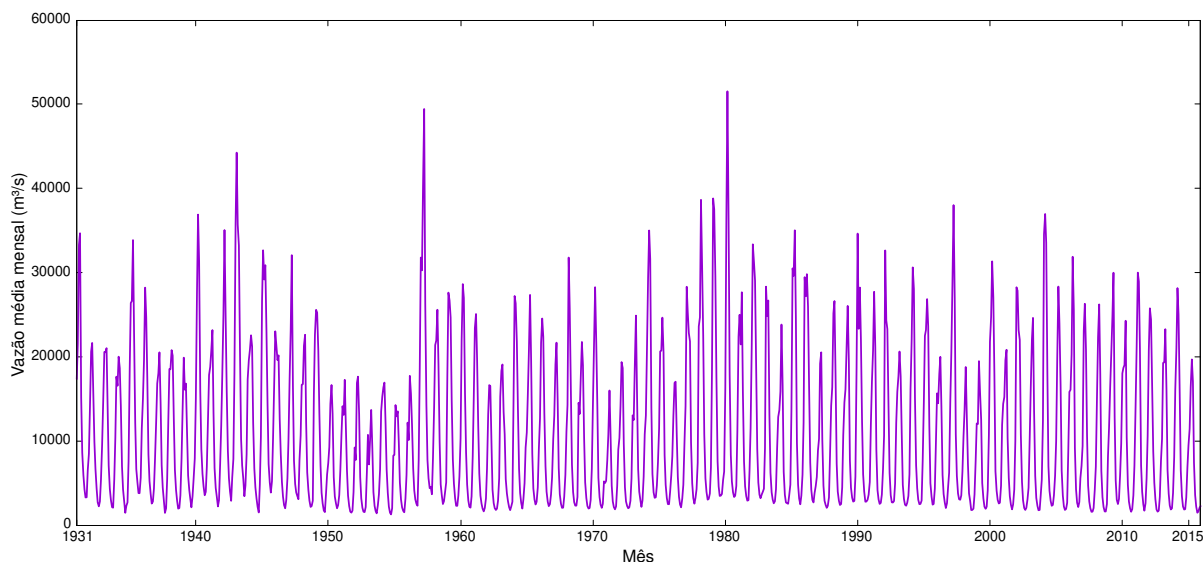
Vê-se que a menor variação de resultado foi da rede neural MLP, enquanto a maior dispersão foi da rede Elman. Como já era esperado os modelos lineares AR e ARMA e os *Ensembles* Média e Mediana não apresentaram variação alguma em suas execuções, uma vez que os coeficientes dos modelos lineares são obtidos de forma determinística.

## 5.6 USINA HIDRELÉTRICA DE TUCURUÍ

Abastecida pelo Rio Tocantins, a Usina Hidrelétrica de Tucuruí é a maior usina hidrelétrica 100% brasileira (visto que a usina de Itaipu é dividida entre o Brasil e o Paraguai). Fica localizada na cidade de Tucuruí no estado do Pará.

Sua série histórica apresenta média  $\hat{\mu} = 10.935,23 \text{ m}^3/\text{s}$  e desvio padrão de  $\hat{\sigma} = 9.182,29 \text{ m}^3/\text{s}$ , a qual é apresentada no Gráfico 48.

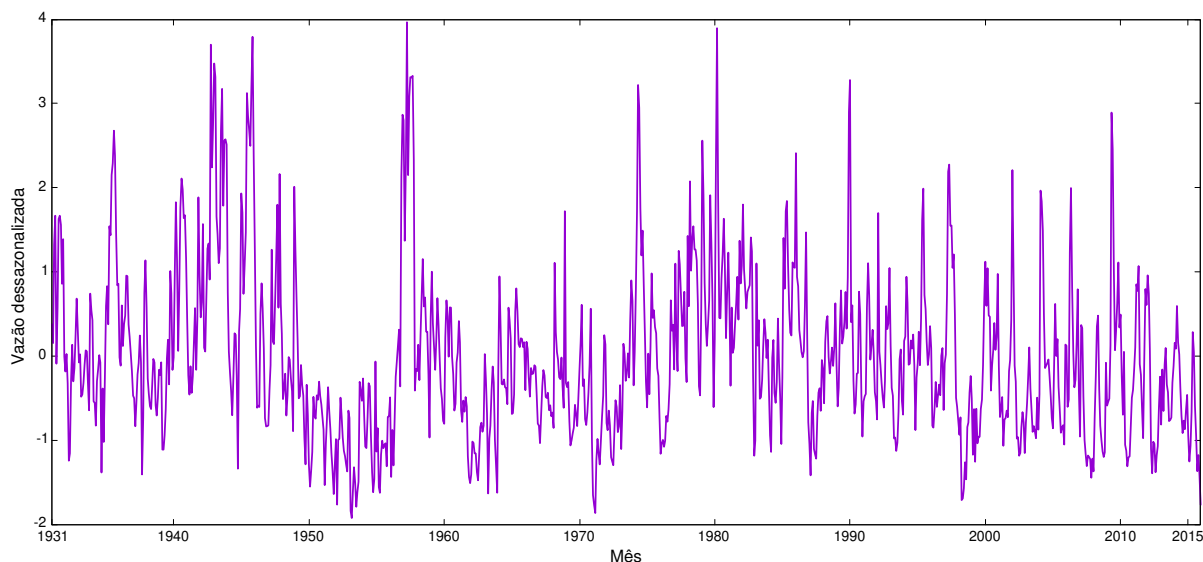
**Gráfico 48 – Série de vazões médias mensais da usina de Tucuruí**



**Fonte: Autoria própria.**

O Gráfico 13 mostra a série dessazonalizada da usina de Tucuruí, de acordo com a Equação 2.13.

**Gráfico 49 – Série dessazonalizada da usina de Tucuruí**



**Fonte: Autoria própria.**

A Tabela 21 contém os resultados das simulações realizadas para a usina de Tucuruí com horizonte de previsão de 1 passo à frente. Nela são apresentadas todas as métricas de erro calculadas (MSE, MAE e MAPE), além do número de neurônios selecionado para cada modelo neural.

Pelos dados da Tabela 21 verifica-se que a melhor previsão para Tucuruí foi realizada pelo *Ensemble RBF*, com MSE real de 5.776.403,72. Já o pior desempenho foi do modelo

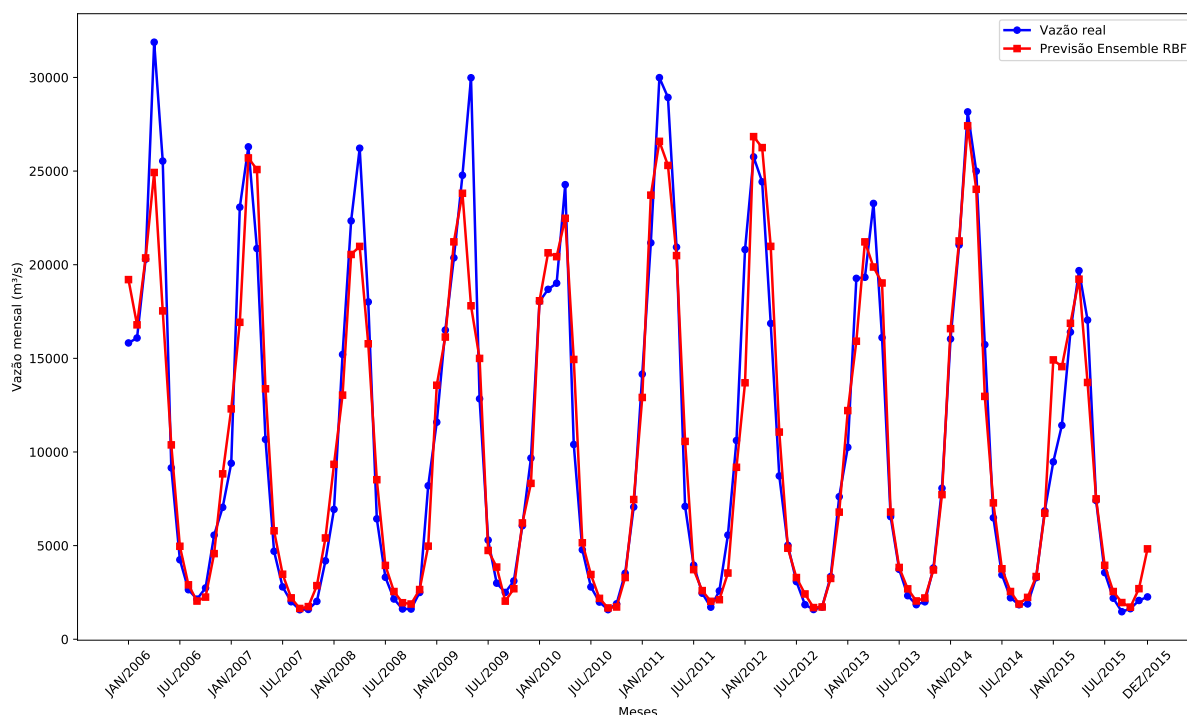
**Tabela 21 – Resultados previsões Tucuruí de 2006 a 2015 para  $P = 1$ .**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	7.446.258,52	0,3310	1.675,10	0,4385	16,25%
ARMA(4,1)	0	7.991.096,15	0,3383	1.725,76	0,4355	16,01%
MLP	5	7.100.370,96	0,3451	1.663,66	0,4706	18,88%
RBF	45	6.721.796,96	0,3592	1.608,63	0,4385	16,05%
ELM	25	6.824.285,76	0,3030	1.587,68	0,4164	15,48%
ELM (CR)	25	6.829.099,84	0,3066	1.592,28	0,4197	15,67%
Elman	5	7.351.576,91	0,4107	1.730,36	0,5255	21,55%
Jordan	5	6.980.914,40	0,3251	1.635,45	0,4533	18,02%
ESN Jaeger	15	6.936.607,01	0,3024	1.607,92	0,4202	15,53%
ESN Jaeger (CR)	15	6.975.358,52	0,3035	1.599,95	0,4156	15,50%
ESN Ozturk	85	6.757.516,84	0,3084	1.584,31	0,4270	16,31%
ESN Ozturk (CR)	5	7.119.853,31	0,3310	1.669,84	0,4587	17,59%
Ensemble Média	0	6.254.104,50	0,3089	1.563,79	0,4408	17,07%
Ensemble Mediana	0	6.757.516,84	0,3084	1.584,31	0,4270	16,31%
Ensemble MLP	5	6.317.164,23	0,3121	1.565,06	0,4439	17,22%
<b>Ensemble RBF</b>	<b>60</b>	<b>5.776.403,72</b>	<b>0,3120</b>	<b>1.472,85</b>	<b>0,4182</b>	<b>15,51%</b>
Ensemble ELM	10	5.923.084,05	0,2792	1.491,39	0,4072	15,09%
Ensemble ELM (CR)	10	5.859.606,76	0,2810	1.485,05	0,4106	15,22%

Fonte: Autoria própria.

ARMA, com MSE real de 7.991.096,15.

Ao analisar a Tabela 21 pelo viés do MSE dessazonalizado obtém-se um resultado diferente. Nesse contexto, o melhor foi o *Ensemble* com combinador de ELM e o pior foi a rede de Elman. O Gráfico 50 apresenta a melhor previsão para Tucuruí com  $P = 1$ .

**Gráfico 50 – Melhor previsão Tucuruí de 2006 a 2015 – 1 passo a frente.**

Fonte: Autoria própria.

Por sua vez, a Tabela 22 apresenta os resultados de Tucuruí com  $P = 3$ .

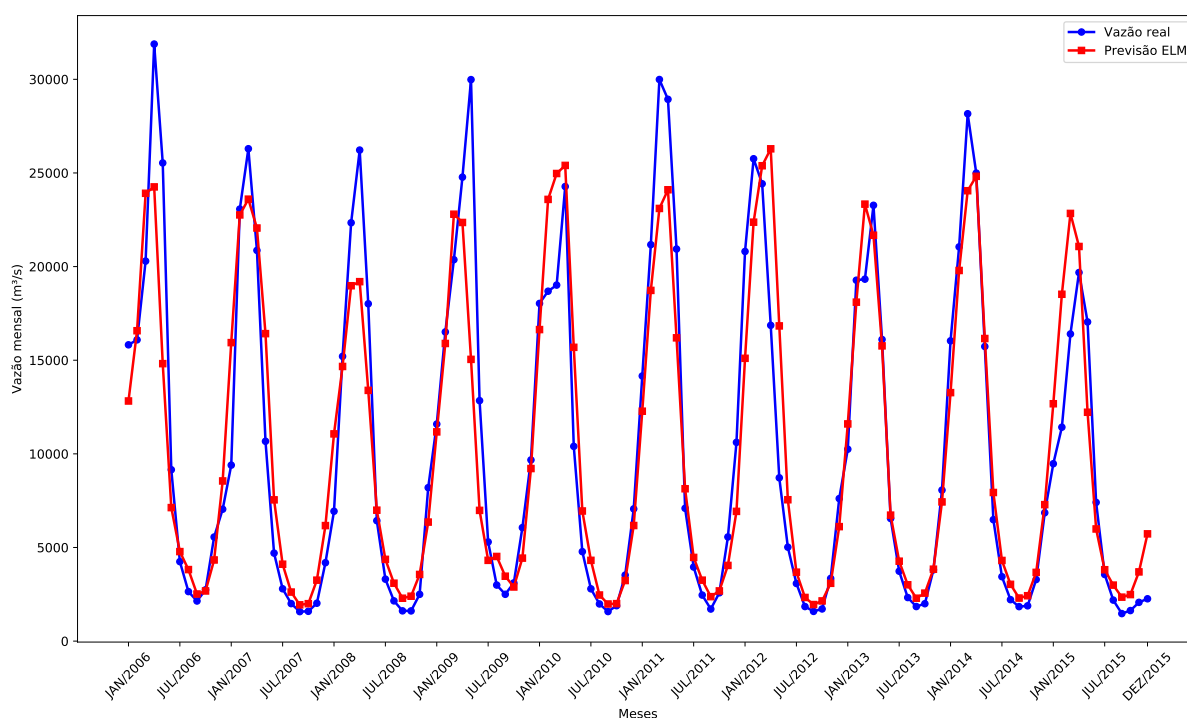
**Tabela 22 – Resultados previsões Tucuruí de 2006 a 2015 para  $P = 3$ .**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	13.406.528, 21	0, 9131	2.397, 81	0, 7710	29, 29%
ARMA(4,1)	0	13.974.802, 61	0, 9714	2.398, 09	0, 7866	30, 19%
MLP	5	11.346.681, 97	0, 7834	2.229, 27	0, 7259	29, 64%
RBF	45	11.966.687, 62	0, 7750	2.208, 67	0, 7040	28, 06%
<b>ELM</b>	<b>25</b>	<b>10.906.829, 60</b>	<b>0, 7399</b>	<b>2.161, 56</b>	<b>0, 6881</b>	<b>26, 20%</b>
ELM (CR)	25	11.134.575, 77	0, 7359	2.177, 69	0, 6883	26, 36%
Elman	5	13.167.672, 49	0, 8019	2.414, 67	0, 7442	31, 34%
Jordan	5	11.752.842, 37	0, 8432	2.298, 55	0, 7625	31, 77%
ESN Jaeger	15	11.639.311, 67	0, 6543	2.200, 17	0, 6357	23, 48%
ESN Jaeger (CR)	15	13.382.939, 20	0, 7647	2.309, 23	0, 6562	23, 32%
ESN Ozturk	85	17.883.748, 20	0, 8628	2.722, 08	0, 6926	23, 63%
ESN Ozturk (CR)	5	11.180.015, 34	0, 6629	2.159, 94	0, 6449	24, 77%
Ensemble Média	0	12.708.717, 47	0, 7938	2.305, 40	0, 7063	26, 33%
Ensemble Mediana	0	14.733.468, 98	0, 8444	2.514, 51	0, 7277	26, 49%
Ensemble MLP	5	12.726.659, 05	0, 8268	2.302, 56	0, 7146	26, 51%
Ensemble RBF	60	14.878.993, 97	0, 8811	2.433, 57	0, 7097	25, 56%
Ensemble ELM	10	15.279.206, 30	1, 09	2.539, 91	0, 7587	26, 49%
Ensemble ELM (CR)	10	14.830.513, 86	1, 01	2.506, 08	0, 7429	26, 16%

Fonte: Autoria própria.

Pela Tabela 22 nota-se que o melhor preditor foi a rede ELM. Já o modelo com o pior desempenho foi a ESN Ozturk. O Gráfico 51 apresenta a previsão para Tucuruí com  $P = 3$ .

**Gráfico 51 – Melhor previsão Tucuruí de 2006 a 2015 – 3 passos a frente.**



Fonte: Autoria própria.

Avançando mais um passo no horizonte de previsão, a Tabela 23 apresenta os resultados para Tcuruí com  $P = 6$ .

**Tabela 23 – Resultados previsões Tucuruí de 2006 a 2015 para  $P = 6$ .**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	15.727.475, 82	1, 20	2.676, 87	0, 8829	37, 52%
ARMA(4,1)	0	16.315.642, 78	1, 37	2.759, 14	0, 9513	40, 78%
<b>MLP</b>	<b>5</b>	<b>11.658.727, 90</b>	<b>0, 8131</b>	<b>2.290, 61</b>	<b>0, 7339</b>	<b>31, 43%</b>
RBF	45	12.696.569, 79	0, 9830	2.405, 42	0, 7957	33, 92%
ELM	25	11.989.453, 79	0, 8238	2.300, 67	0, 7285	30, 72%
ELM (CR)	25	12.143.842, 92	0, 8236	2.301, 84	0, 7237	30, 36%
Elman	5	14.163.188, 22	1, 10	2.642, 41	0, 8814	38, 54%
Jordan	5	12.264.013, 86	0, 8860	2.369, 35	0, 7752	33, 61%
ESN Jaeger	15	12.040.332, 15	0, 7548	2.271, 11	0, 6878	28, 16%
ESN Jaeger (CR)	15	12.071.588, 19	0, 7345	2.270, 12	0, 6721	27, 45%
ESN Ozturk	85	16.299.760, 49	0, 8256	2.529, 34	0, 7000	27, 08%
ESN Ozturk (CR)	5	11.770.003, 63	0, 7062	2.243, 59	0, 6693	27, 44%
<i>Ensemble Média</i>	0	16.469.919, 61	0, 9983	2.645, 48	0, 7881	31, 27%
<i>Ensemble Mediana</i>	0	20.303.247, 25	1, 12	2.970, 70	0, 8459	32, 68%
<i>Ensemble MLP</i>	5	15.554.587, 87	0, 9710	2.512, 95	0, 7641	30, 39%
<i>Ensemble RBF</i>	60	18.606.965, 66	1, 02	2.797, 43	0, 7906	29, 96%
<i>Ensemble ELM</i>	10	21.731.212, 23	1, 28	3.011, 98	0, 8791	33, 35%
<i>Ensemble ELM (CR)</i>	10	20.723.489, 57	1, 26	2.951, 88	0, 8740	33, 37%

**Fonte: Autoria própria.**

Os dados da Tabela 23 indicam que com  $P = 6$  o modelo de previsão com melhor desempenho foi a rede neural MLP, com MSE de 11.658.727, 90. Por sua vez, as piores previsões foram as realizadas pelo *Ensemble ELM*, tendo MSE de 21.731.212, 23, resultado 86% pior que o do melhor modelo.

Por outro lado, ao analisar os valores do MAE o resultado é outro. Neste caso o modelo com melhor desempenho foi a ESN Ozturk com Coeficiente de Regularização. Esse fato indica que as previsões realizadas pela rede podem até conter menos erros. Entretanto, esses erros são maiores que os erros da MLP, o que faz com que seu MSE fique maior mas seu MAE não.

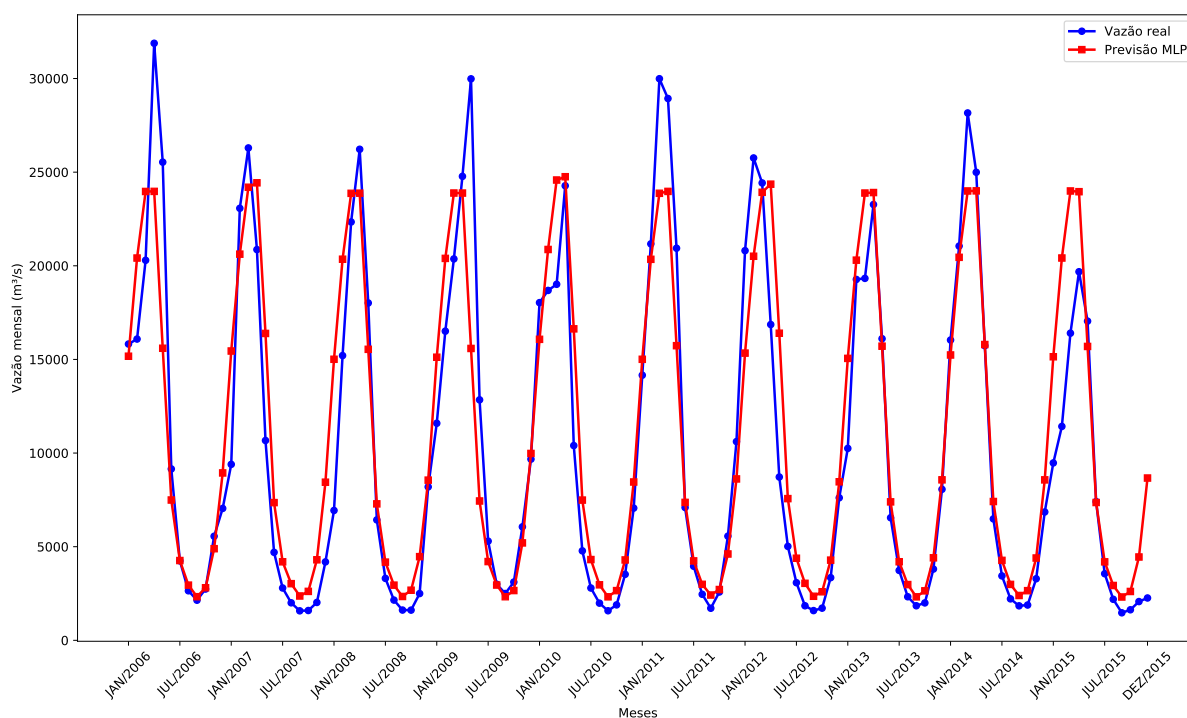
Já o Erro Percentual Absoluto Médio é uma medida mais simples, que compreende apenas uma média do percentual de erro de cada previsão realizada, não sendo possível extrair informações mais específicas sobre as previsões realizadas (por exemplo a magnitude dos erros) a partir da comparação dos erros, uma vez que um único erro muito grande pode acabar gerando o mesmo valor de MAPE que vários erros menores.

Outro fato que vale a pena mencionar é que ao considerar o ambiente dessazonalizado, tem-se que as melhores previsões foram realizadas pela ESN Jaeger tanto pelo MSE quanto pelo MAE.

O Gráfico 52 apresenta a previsão realizada pela MLP para o conjunto de teste da usina de Tucuruí e  $P = 6$ .

Por fim, a Tabela 24 apresenta os resultados das simulações realizadas para a usina de



**Gráfico 52 – Melhor previsão Tucuruí de 2006 a 2015 – 6 passos a frente.**

Fonte: Autoria própria.

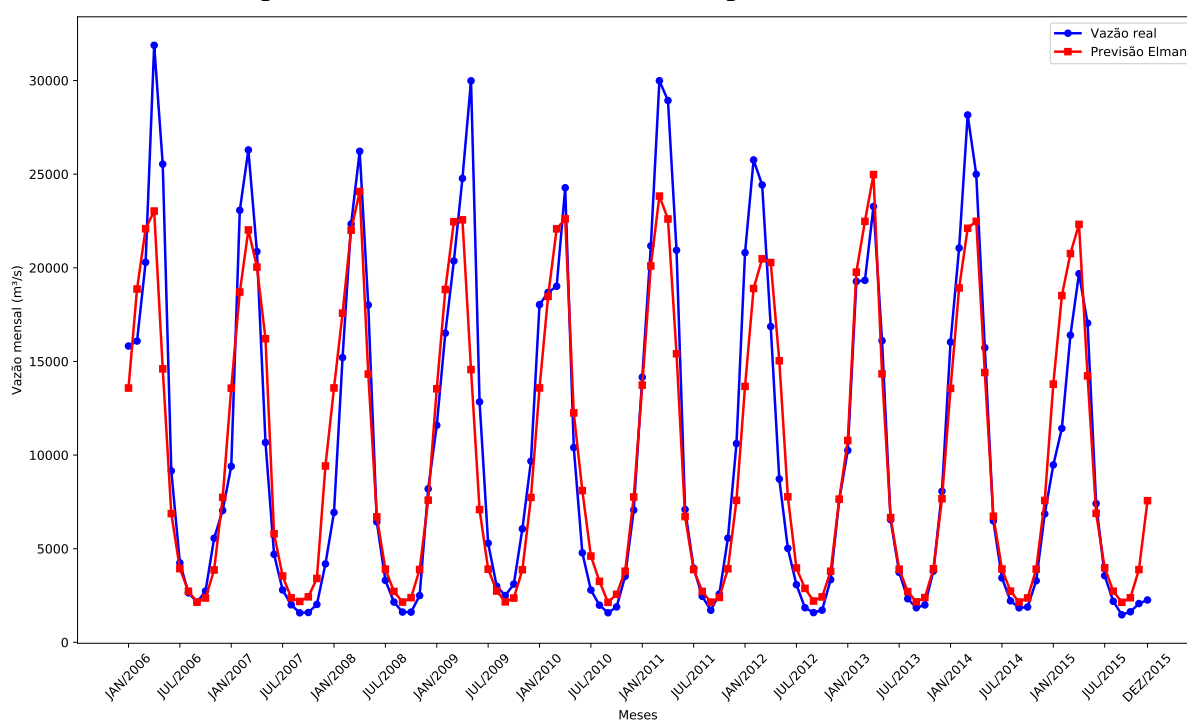
Tucuruí com  $P = 12$ .

**Tabela 24 – Resultados previsões Tucuruí de 2006 a 2015 para  $P = 12$ .**

Modelo	Neur	MSE	MSE (d)	MAE	MAE (d)	MAPE
AR(3)	0	22.161.343, 50	1, 46	3.240, 40	1, 02	45, 02%
ARMA(4,1)	0	29.987.913, 13	1, 91	3.883, 99	1, 20	53, 39%
MLP	5	11.743.205, 59	0, 8254	2.315, 67	0, 7465	32, 03%
RBF	45	15.087.729, 25	0, 9688	2.642, 20	0, 8202	35, 93%
ELM	25	12.304.667, 66	0, 8764	2.382, 91	0, 7739	33, 53%
ELM (CR)	25	12.284.951, 17	0, 8614	2.379, 63	0, 7661	33, 14%
<b>Elman</b>	<b>5</b>	<b>10.533.366, 45</b>	<b>0, 6894</b>	<b>2.117, 52</b>	<b>0, 6487</b>	<b>26, 03%</b>
Jordan	5	12.341.774, 82	0, 8899	2.394, 39	0, 7821	33, 90%
ESN Jaeger	15	11.766.645, 65	0, 8119	2.317, 05	0, 7397	31, 68%
ESN Jaeger (CR)	15	12.279.058, 45	0, 8578	2.377, 23	0, 7550	32, 18%
ESN Ozturk	85	11.238.408, 86	0, 7952	2.223, 79	0, 6976	29, 68%
ESN Ozturk (CR)	5	10.807.257, 91	0, 6965	2.171, 14	0, 6716	27, 53%
Ensemble Média	0	19.601.009, 99	1, 07	2.977, 82	0, 8155	33, 12%
Ensemble Mediana	0	20.367.666, 36	1, 18	3.091, 24	0, 8640	34, 57%
Ensemble MLP	5	21.292.183, 31	1, 07	3.027, 21	0, 8075	32, 60%
Ensemble RBF	60	22.318.965, 50	1, 21	2.991, 08	0, 8293	33, 70%
Ensemble ELM	10	31.221.147, 23	1, 52	3.447, 62	0, 8793	34, 32%
Ensemble ELM (CR)	10	34.497.724, 89	1, 58	3.577, 12	0, 8979	36, 75%

Fonte: Autoria própria.

É possível notar que, considerando tanto MSE quanto o MAE, o melhor desempenho foi a rede de Elman. O Gráfico 53 apresenta as previsões realizadas por esse preditor para a usina de Tucuruí com  $P = 12$ .

**Gráfico 53 – Melhor previsão Tucuruí de 2006 a 2015 – 12 passos a frente.**

**Fonte: Autoria própria.**

Observando cuidadosamente os dados das tabelas 21, 22, 23 e 24 nota-se um aumento no valor do MSE dos modelos de previsão, como já foi mencionado nas seções 5.2, 5.3, 5.4 e 5.5. Isso ocorre em função da propagação do erro presente na previsão recursiva. Para esclarecer tal comportamento para a usina de Tucuruí a Tabela 25 apresenta a evolução no valor do MSE do melhor e do pior modelo em cada horizonte de previsão, além dos percentuais de aumento em relação ao horizonte anterior e a  $P = 1$ .

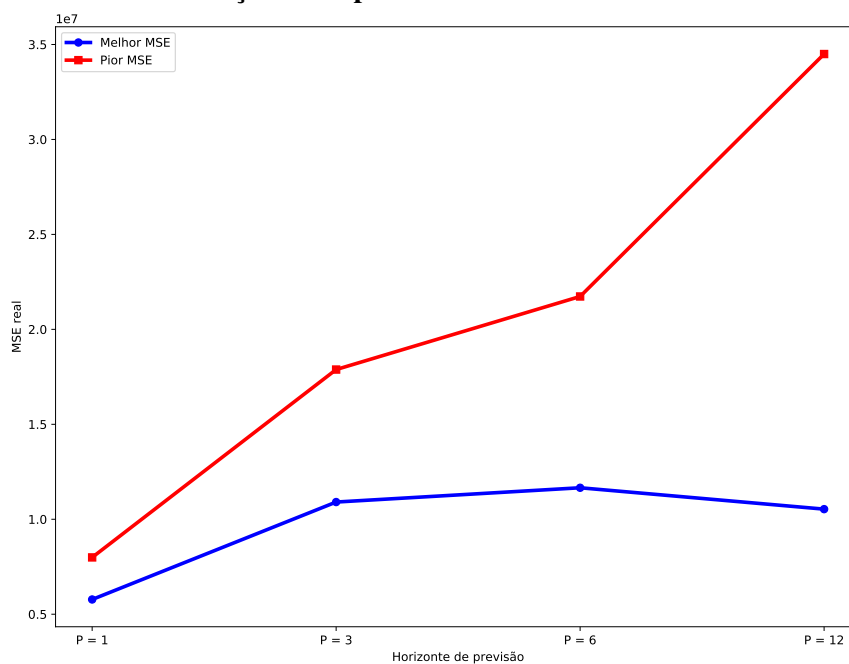
**Tabela 25 – Evolução MSE Tucuruí.**

Horizonte	Melhor MSE	Anterior	$P = 1$	Pior MSE	Anterior	$P = 1$
$P = 1$	5.776.403,72	—	—	7.991.096,15	—	—
$P = 3$	10.906.829,60	88,82%	88,82%	17.883.748,20	123,80%	123,80%
$P = 6$	11.658.727,90	6,89%	101,83%	21.731.212,23	21,51%	171,94%
$P = 12$	10.533.366,45	-9,65%	82,35%	34.497.724,89	58,75%	331,70%

**Fonte: Autoria própria.**

De acordo com a análise dos resultados contidos na Tabela 25 é possível observar que o valor do melhor MSE segue crescendo até  $P = 6$ , sendo o maior aumento de 88,82% entre  $P = 1$  e  $P = 3$ . Após o horizonte de 6 passos a frente o melhor MSE tem uma diminuição de 9,65% o que é contrário ao comportamento provocado pela propagação do erro de previsão causado pela técnica de previsão recursiva. A evolução do melhor e do pior MSE em relação ao horizonte de previsão em questão é apresentada no Gráfico 54.

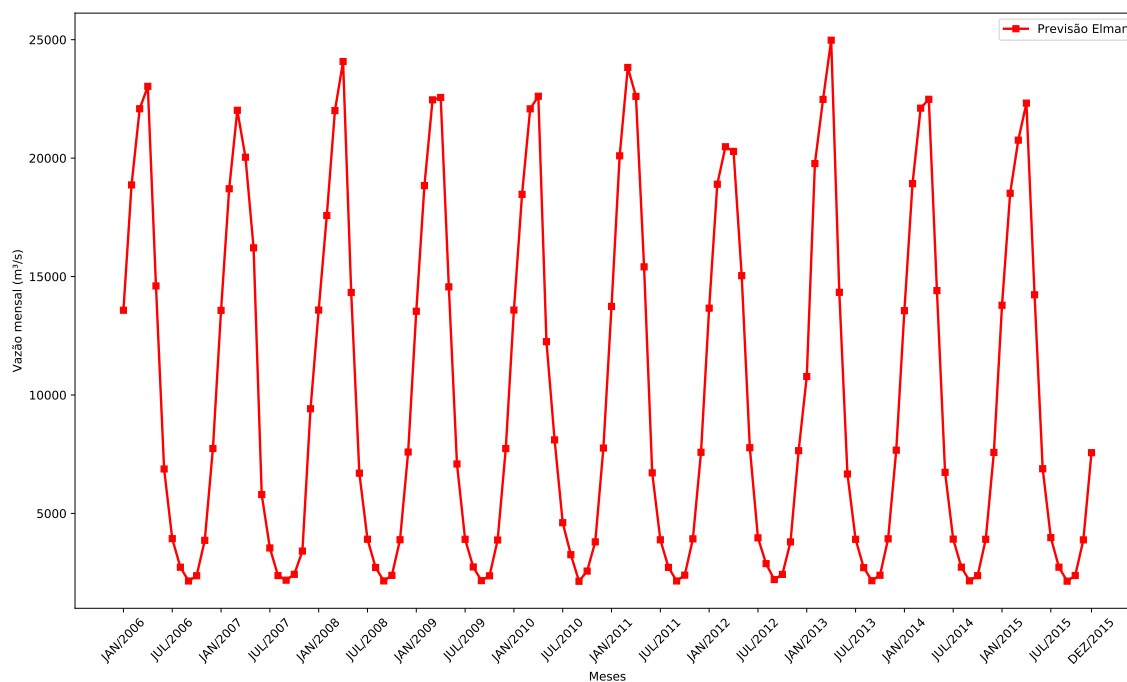
No que diz respeito a evolução do pior MSE para a usina de Tucuruí, ao analisar os dados da Tabela 25 e as curvas do Gráfico 54 nota-se que este segue aumentando por todos

**Gráfico 54 – Evolução MSE para a usina de Tucuruí.**

**Fonte: Autoria própria.**

os horizontes. Enfatizamos a elevação de 123, 80% quando o horizonte avança de  $P = 1$  para  $P = 3$ . Diferentemente do observado na evolução do melhor MSE, aqui existe um clássico comportamento decorrente da propagação de erro pela previsão recursiva.

O Gráfico 55 contém apenas as previsões realizadas pela rede de Elman para o conjunto de teste da usina de Tucuruí com horizonte de previsão de 12 passos à frente.

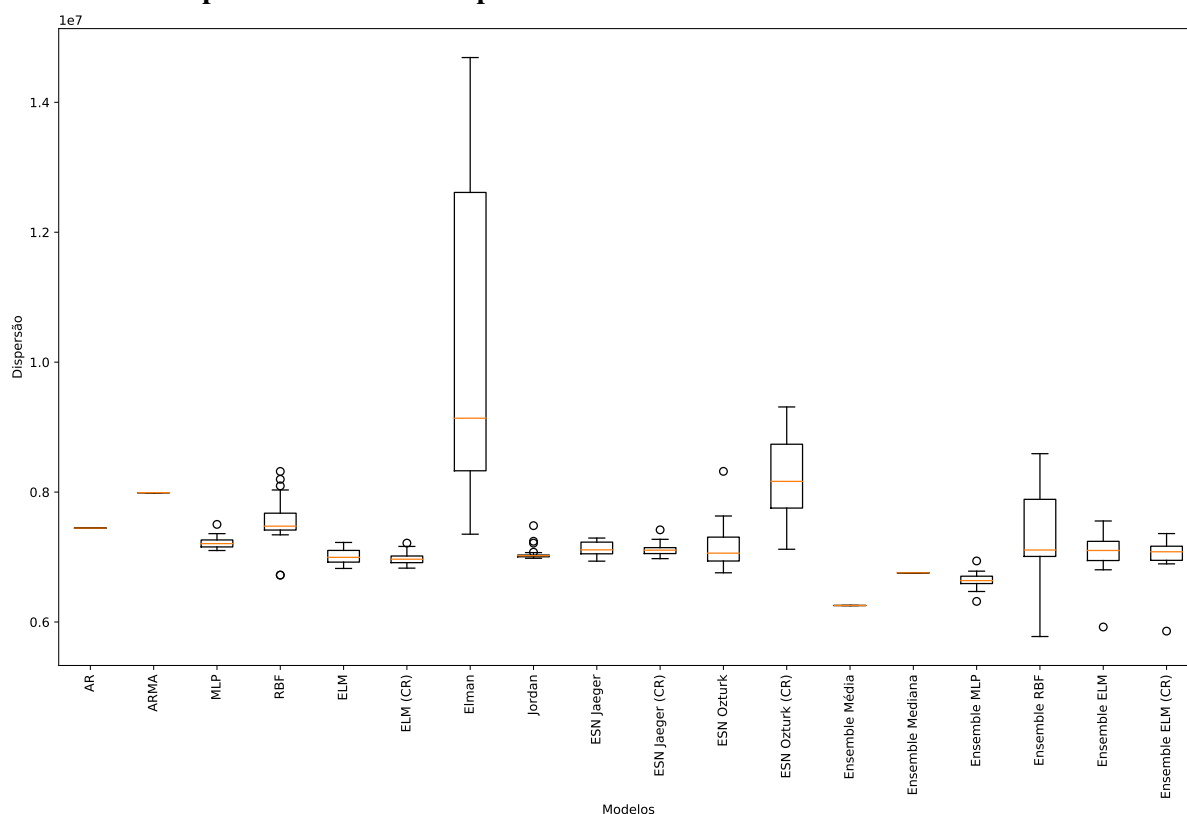
**Gráfico 55 – Padrão das previsões para Tucuruí com P = 12.**

**Fonte: Autoria própria.**

O padrão cíclico nas previsões realizadas está evidente no Gráfico 55, com picos e vales sempre nas mesmas épocas. Observando mais atentamente os gráficos 50, 51, 52 e 53 é possível notar que a cada horizonte de previsão esse comportamento vem se reforçando até que fica bem marcante com  $P = 12$ .

Como foi mencionado no início desse capítulo, cada modelo de previsão proposto foi treinado 30 vezes com o objetivo de verificar o quão distantes seriam os resultados obtidos. O Gráfico 56 apresenta portanto a dispersão dos resultados.

**Gráfico 56 – Dispersão dos resultados para a usina de Tucuruí.**



**Fonte: Autoria própria.**

Pelo gráfico anterior, nota-se que os modelos lineares (Auto-regressivo e Auto-regressivo de Médias Móveis) e os modelos neurais que não precisam de treinamento (*Ensemble Média* e *Ensemble Mediana*) não apresentaram qualquer dispersão nos seus resultados, como esperado.

Por outro lado, quando se analisa a dispersão dos modelos neurais, que tem seus pesos sinápticos ajustados através de treinamento, pode-se notar que a rede de Elman apresenta uma alta variação nos seus resultados, de modo que não há garantias muito específicas quanto a que resultado será alcançado para o problema em questão.

Ainda analisando o Gráfico 56, observa-se que o modelo que apresentou os resultados com menor variação foi a rede de Jordan, indicando que ao se treinar esse modelo com as mesmas configurações para o mesmo conjunto de dados se obtém sempre resultados próximos.

## 5.7 ANÁLISE DOS RESULTADOS

Como mencionado no início desse capítulo, a escolha das 5 séries foi realizada de forma que seus valores médios e comportamentos hidrológicos fossem variados. Pode-se exemplificar tal discrepância entre as vazões médias das 5 séries históricas pelo Gráfico 11, o qual mostra a vazão média de cada usina. Isso indica que as metodologias aplicadas foram testadas para cenários diferentes, validando assim seu poder de previsão em diversas situações reais.

Com o objetivo de verificar qual dos modelos de previsões propostos teve o melhor desempenho no maior número de testes, a Tabela 26 foi elaborada mostrando quantas vezes cada um foi o melhor para cada horizonte e em todas as usinas.

**Tabela 26 – Melhor modelo por horizonte.**

<b>Modelo</b>	<b>P = 1</b>	<b>P = 3</b>	<b>P = 6</b>	<b>P = 12</b>	<b>Total</b>	
AR	0	0	0	0	0	0
ARMA	0	0	0	0	0	
MLP	0	0	1	1	2	2
RBF	0	0	0	0	0	
ELM	0	<b>2</b>	<b>4</b>	0	<b>6</b>	6
ELM (CR)	0	0	0	0	0	
Elman	0	0	0	1	1	1
Jordan	0	0	0	0	0	
ESN Jaeger	0	1	0	<b>2</b>	3	4
ESN Jaeger (CR)	0	0	0	0	0	
ESN Ozturk	0	0	0	0	0	
ESN Ozturk (CR)	0	0	0	1	1	
<i>Ensemble Média</i>	0	1	0	0	1	7
<i>Ensemble Mediana</i>	0	0	0	0	0	
<i>Ensemble MLP</i>	0	1	0	0	1	
<i>Ensemble RBF</i>	1	0	0	0	1	
<i>Ensemble ELM</i>	<b>3</b>	0	0	0	3	
<i>Ensemble ELM (CR)</i>	1	0	0	0	1	

**Fonte: Autoria própria.**

A Tabela 26 permite observar que se tratando do horizonte de 1 passo a frente, para todas as usinas o melhor preditor foi sempre um *Ensemble*, dando destaque para o combinador baseado em ELM, que foi o melhor para 3 das 5 usinas testadas.

Esse resultado comprova o que já foi dito na Seção 4.9, que o uso da metodologia *Ensemble* como forma de combinar os resultados de preditores individuais pode levar a um incremento importante de desempenho.

Para  $P = 3$ , tem-se um resultado bem mais variado, em que diversas arquiteturas diferentes conseguiram se destacar ao menos uma vez. São elas: ELM, ESN Jaeger, *Ensemble* de Média e o *Ensemble* MLP. Enfatiza-se que a rede ELM foi melhor em 2 das 5 usinas testadas.

No que se refere ao horizonte  $P = 6$ , pelos dados da Tabela 26, tem-se novamente a rede neural ELM como melhor modelo de previsão, tendo sido o destaque para 4 dos 5 cenários,

sendo seguida pela rede MLP, que foi a melhor apenas para Tucuruí.

Por fim, analisando o último horizonte de previsão abordado, 12 passos a frente, verifica-se que 4 arquiteturas diferentes de rede neural foram o melhor modelo pelo menos uma vez: MLP, rede de Elman, ESN Jaeger e ENS Ozturk (CR). O destaque é a ESN de Jaeger (2001) sem a utilização do Coeficiente de Regularização.

Analisando os resultados gerais de cada arquitetura é possível notar que em se tratando dos horizontes de previsão mais curtos ( $P = 1$  e  $P = 3$ ) os *Ensembles* e as Máquinas de Aprendizado Extremo são as arquiteturas mais indicadas.

Já para horizontes mais longos ( $P = 6$  e  $P = 12$ ), observamos novamente que as propostas mais recomendada são as Máquinas de Aprendizado Extremo e as Redes Neurais com Estado de Eco, sendo as ELM a melhor para 6 passos a frente e as ESN para 12. Neste último caso, tal arquitetura foi melhor em 3 dos 5 cenários. Já foi mencionado anteriormente que esta rede possui memória associativa, o que a favorece quando se trata de problemas que possuem dependência temporal. Esses resultados confirmam mais uma vez o poder de generalização que as máquinas desorganizadas possuem, já que para esse problema sua performance superou as das redes clássicas (MLP, RBF, Elman e Jordan).

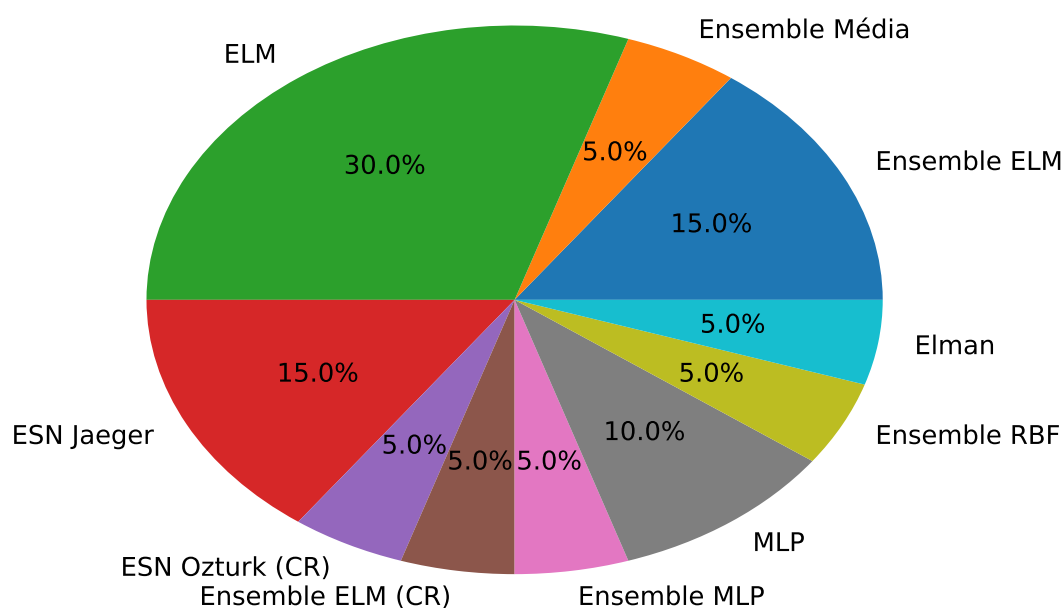
Vale ressaltar que os modelos lineares não superaram as redes neurais em nenhum dos 20 cenários. Isso indica que para o problema da previsão de vazões mensais as Redes Neurais Artificiais são mais adequadas, apesar destes ainda serem os muito utilizados atualmente.

Apesar da MLP ser uma arquitetura clássica e a rede neural mais utilizada, seus desempenhos não foram destaque, uma vez que dos 20 cenários realizados ela foi melhor em apenas 2 deles. O mesmo vale para a rede de Elman, que por sua vez é uma das redes recorrentes clássicas. Em geral, as arquiteturas mais populares foram superadas em todos os casos, o que prova que as novas são propostas de potencial e devem ser consideradas. Além disso, este é um indicativo que a área tem progredido com o passar dos anos, técnicas mais novas e melhores tem sido criadas, proporcionando avanços significativos em novos problemas.

Complementando os dados da Tabela 26 tem-se o Gráfico 57, que por sua vez apresenta em qual porcentagem dos testes realizados cada modelo de previsão obteve o melhor desempenho.

Pelo Gráfico 57 nota-se que o modelo de previsão que mais se destacou foi a rede neural ELM sem o Coeficiente de Regularização, tendo sido o melhor em 30% dos testes. O segundo modelo com o maior percentual foi a rede ESN com reservatório de dinâmicas proposto por Jaeger (2001) e, novamente, sem a utilização do Coeficiente de Regularização, com 15% das melhores previsões. Ainda na segunda colocação tem-se também o *Ensemble* com combinador de ELM, também com 15% dos melhores resultados. Por fim, em terceiro lugar está a rede neural MLP, tendo sido o melhor modelo de previsão em 10% dos testes realizados.

**Gráfico 57 – Porcentagens de melhor resultado por modelo de previsão.**



**Fonte: Autoria própria.**

## 5.8 TESTE DE FRIEDMAN

Ferreira (2010) afirma que considerar apenas as métricas de erro como suficientes para análises dos resultados obtidos por modelos de previsão pode levar o pesquisador a conclusões estatisticamente equivocadas. Além das métricas de erro o autor sugere a utilização de testes de hipótese como forma de garantir que os resultados obtidos pelos diversos modelos de previsão considerados são estatisticamente diferentes entre si.

Tais testes são divididos em dois grupos, paramétricos e não paramétricos. O primeiro faz uso dos parâmetros da distribuição de probabilidade dos dados para realizar seus cálculos estatísticos e assim validar ou não suas hipóteses. Tal grupo tende a ser mais rigoroso com os resultados e conseqüentemente também é mais confiável (MÜLLER; KRÜGER; KAVISKI, 1998). Sua maior desvantagem está na premissa de que a distribuição de probabilidade dos dados testados deve ser conhecida, de modo que se for utilizada a distribuição incorreta os resultados obtidos serão inválidos.

O segundo grupo de teste de hipótese, os não paramétricos, não fazem uso de qualquer informação a priori sobre a distribuição de probabilidade dos dados. Neste caso, os pressupostos são atribuídos apenas aos dados ordenados, o que gera cálculos estatísticos mais simples, mas conseqüentemente resultados menos rigorosos (MÜLLER; KRÜGER; KAVISKI, 1998). Apesar de serem menos exigentes, podem ser aplicados às mais diversas massas de dados.

Para esse trabalho foi selecionado o teste não paramétrico de Friedman (FRIEDMAN,

1937). Seu objetivo é determinar se os dados foram extraídos da mesma população, ou seja, este mensura a probabilidade de que as amostras tenham origem no mesmo processo gerador.

O teste de Friedman considera duas hipóteses para serem confirmadas ou rejeitadas:

- $H_0$  – os tratamentos são iguais;
- $H_1$  – pelo menos um tratamento em toda a população produz grandes efeitos.

Além das duas hipóteses apresentadas, o teste de Friedman assume os seguintes pressupostos sobre os dados:

1. Os blocos de dados são independentes entre si;
2. Não existe qualquer interação entre os blocos de dados;
3. As observações dentro de cada bloco podem ser ranqueadas.

Como resultado são extraídos dois valores: o primeiro é a Estatística de Friedman ( $X^2$ ) e o segundo é a Probabilidade Limite ( $p$ -valor). A comprovação ou rejeição da hipótese  $H_0$  depende do valor  $p$ -valor, o qual, quanto mais próximo de 0 maior é a probabilidade da hipótese  $H_0$  ser rejeitada.

O teste de Friedman foi aplicado com os dados do MSE real obtido em cada uma das 30 execuções de cada modelo preditivo proposto, sendo que para cada usina foi encontrado o  $p$ -valor que é apresentado na Tabela 27.

**Tabela 27 – Resultado teste de Friedman.**

<b>Usina</b>	<b>p-valor</b>
Água Vermelha	2,2128e-85
Belo Monte	9,2064e-76
Ilha Solteira	2,9835e-82
Paulo Afonso	1,5249e-80
Tucuruí	9,7062e-75

**Fonte: Autoria própria.**

Como pode ser observado os  $p$ -valores calculados foram muito próximos de 0, o que indica que a hipótese  $H_0$  foi rejeitada. Portanto, podemos assumir que alterar o preditor utilizado resulta em uma alteração significativa das previsões realizadas.

## 5.9 QUANTIDADE DE NEURÔNIOS SELECIONADA

Nesta seção será discutida a quantidade de neurônios selecionada para cada modelo neural de previsão. Como foi dito no início deste capítulo, todas as redes neurais foram testadas



em relação à melhor quantidade de neurônios com valores entre 5 e 200, com incremento de 5 em 5. Entretanto, como pode ser observado pelas tabelas 1, 6, 11, 16 e 21 a quantidade deles para os melhores desempenhos não passou de 100.

A Tabela 28 apresenta o número de unidades neurais selecionada para cada modelo. Os dados são separados por usina, e também são apresentadas as médias por modelo e por série.

**Tabela 28 – Quantidade de neurônios selecionada por modelo neural e usina.**

Modelo	Água V.	Belo M.	Ilha S.	Paulo A.	Tucuruí	Média
MLP	10	10	20	40	5	17
RBF	95	25	75	10	45	50
ELM	30	35	30	5	25	25
ELM (CR)	80	5	95	35	25	48
Elman	5	5	5	5	5	5
Jordan	5	5	10	5	5	6
ESN Jaeger	80	5	5	75	15	36
ESN Jaeger (CR)	65	5	10	45	15	28
ESN Ozturk	35	5	20	80	85	45
ESN Ozturk (CR)	35	5	60	80	5	37
<i>Ensemble</i> MLP	5	5	15	35	5	13
<i>Ensemble</i> RBF	15	15	55	10	60	31
<i>Ensemble</i> ELM	40	5	40	45	10	28
<i>Ensemble</i> ELM (CR)	60	5	50	55	10	36
Média	40	9	35	37	22	28

**Fonte: Autoria própria.**

Pelos dados da Tabela 28<sup>3</sup>, de modo geral, os números de neurônios selecionados foram relativamente baixos. Ao analisar-se as médias por usina pode-se observar que Tucuruí, que é a com maior vazão dentre as plantas selecionadas, ficou com uma média de apenas 22 neurônios. Em contrapartida, a usina de Água Vermelha, usina com menor vazão, foi a que apresentou a maior quantidade neurônios.

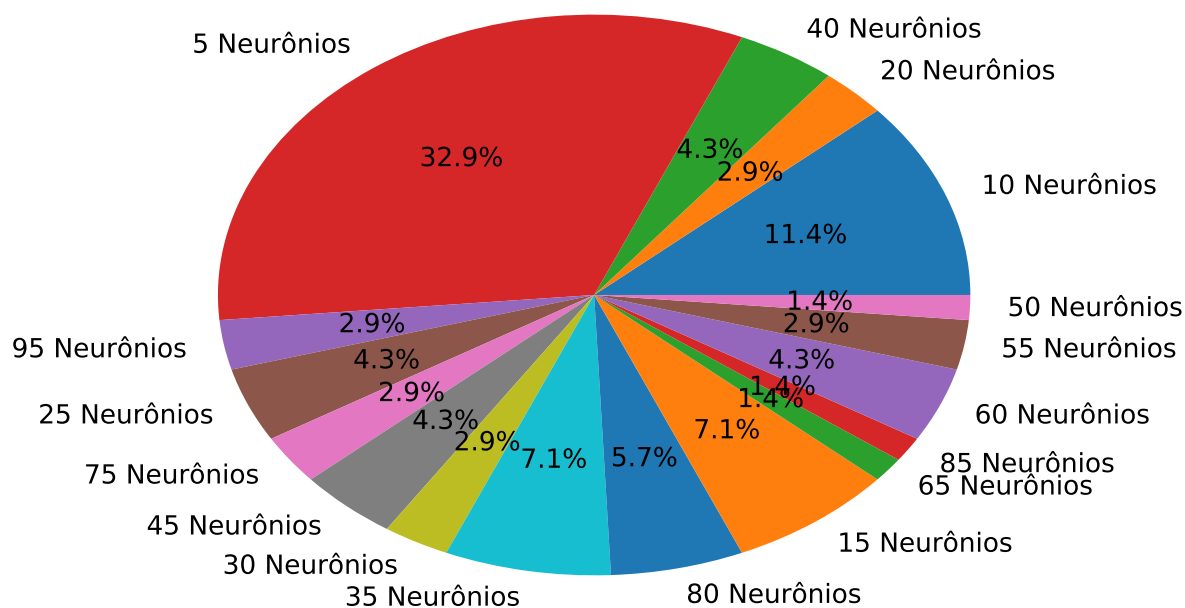
O último fato mencionado pode gerar a ideia de que quanto menor for a média da vazão afluyente maior deve ser o número de neurônios utilizado para generalizar seu comportamento. Entretanto, tal generalização não se faz de forma direta, uma vez que um exemplo já citado refuta essa hipótese: a usina de Tucuruí é a com maior média histórica, mas não é a usina com menor média de neurônios, sendo esta a de Belo Monte.

Ao comparar as médias de neurônios dos modelo de previsão destacam-se as redes recorrentes clássicas com os menores números, tendo a rede de Elman média de 5 e a rede de Jordan 6. As demais também apresentaram quantidades pequenas, mas seus valores não possibilitaram que qualquer conclusão fosse comprovada.

A partir dos dados da Tabela 28 foi criado o Gráfico 58.

Observando o Gráfico 58 fica evidente que o número de neurônios mais utilizado foi 5 (32, 9% dos casos), que é justamente o menor valor testado. Em segundo lugar vem 10 neurônios

<sup>3</sup> Vale ressaltar aqui que os valores referentes aos modelo AR, ARMA, *Ensemble* Média e *Ensemble* Mediana foram omitidos, uma vez que os mesmos não contém neurônios em suas estruturas.

**Gráfico 58 – Percentual de uso por quantidade de neurônios.**

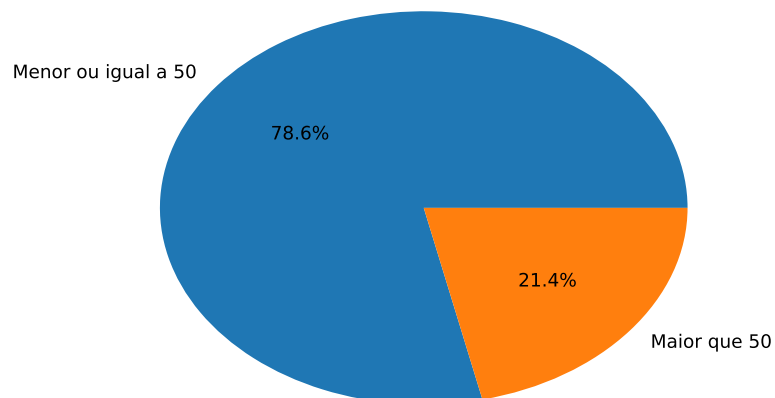
**Fonte: Autoria própria.**

com 11,4% dos casos e em terceiro lugar empatados com 7,1% vem os valores 15 e 35. Esses dados corroboram com a ideia de que para o problema da previsão de vazões mensais com as mesmas condições apresentadas neste trabalho não se faz necessário um grande número de neurônios para que as redes neurais consigam generalizar o seu comportamento, de modo que quanto maior o número de neurônios pior tende a ser o resultado.

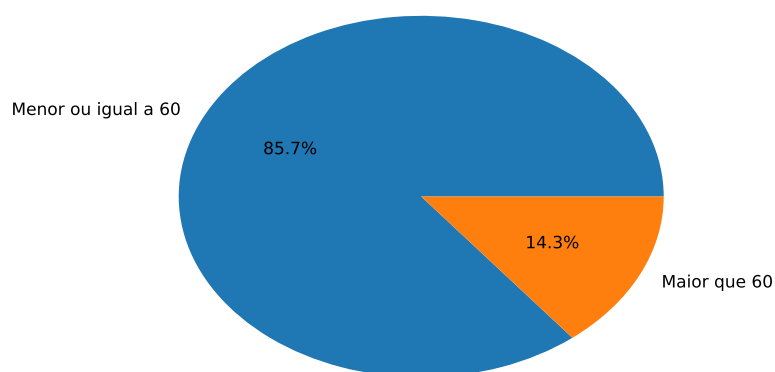
Voltando aos dados da Tabela 28, dessa vez com foco nas quantidades médias de neurônios, tanto por usina hidrelétrica quanto por modelo preditivo encontra-se mais um argumento a favor da ideia de que para este problema em específico não se faz necessário quantidades altas, uma vez que de todas as médias apresentadas na Tabela 28 nenhuma apresentou valor maior que 50. No que se refere aos valores médios por usina a maior média foi da usina de Água Vermelha com média de 40 neurônios por modelo preditivo.

Com o objetivo de estabelecer um limite superior para o número de neurônios, de modo que um projetista com pouco tempo para realizar o treinamento de seus modelos de previsão possa reduzir a quantidade de testes realizados, nos gráficos 59, 60, 61, 62 e 63 é apresentado o percentual de vezes em que foi selecionado um número de neurônios menor ou igual a 50, 60, 70, 80 e 90, respectivamente.

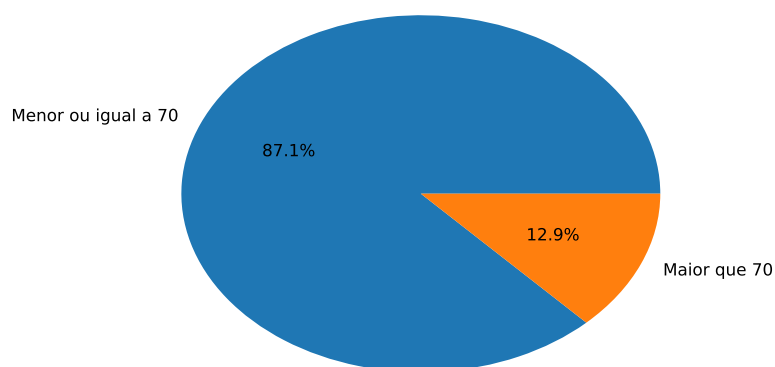
Analisando o Gráfico 61 verifica-se que testar a quantidade de neurônios até 70 já é o suficiente para determinar o melhor número em praticamente 90% dos casos (87,1%). Caso o projetista tenha a liberdade de considerar um limite superior recomenda-se 60 unidades neurais, pois, como é mostrado no Gráfico 60, em 85,7% dos casos esse número foi suficiente. Para

**Gráfico 59 – Percentual de neurônios menor que 50.**

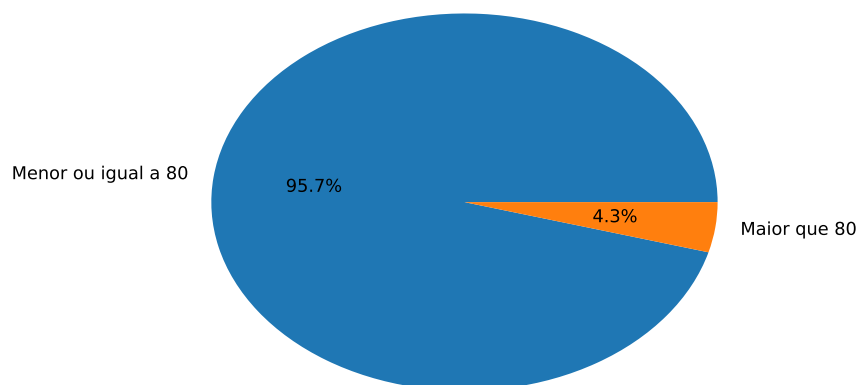
**Fonte: Autoria própria.**

**Gráfico 60 – Percentual de neurônios menor que 60.**

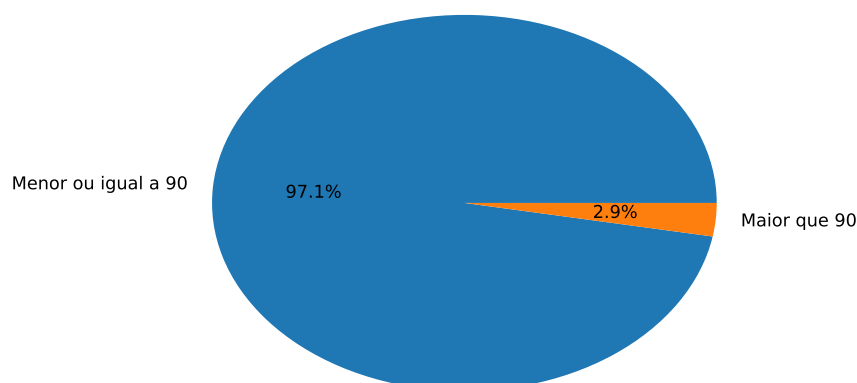
**Fonte: Autoria própria.**

**Gráfico 61 – Percentual de neurônios menor que 70.**

**Fonte: Autoria própria.**

**Gráfico 62 – Percentual de neurônios menor que 80.**

**Fonte: Autoria própria.**

**Gráfico 63 – Percentual de neurônios menor que 90.**

**Fonte: Autoria própria.**

praticamente 100% dos casos o limitante superior é de até 90 neurônios. Isso considerando as mesmas condições apresentadas nesse trabalho.

Um último argumento para comprovar a ideia de que não é necessário um alto número de neurônios para que as redes neurais sejam capazes de generalizar o comportamento das séries de vazões mensais é obtido ao analisar o Gráfico 59. Aqui é possível observar que em 78,6% dos casos o número de neurônios selecionados foi menor ou igual a 50.

## 5.10 ATRASOS SELECIONADOS

Na Seção 4.10.1 foi mencionado que em se tratando da previsão de séries temporais por meio da utilização de Redes Neurais Artificiais a escolha de quais atrasos serão utilizados como entradas pelos modelos é uma etapa tão importante quanto determinar a quantidade de neurônios

que cada arquitetura deve utilizar. Para isso, o método *Wrapper* foi utilizado para determinar quais atrasos geram as melhores previsões para cada um dos modelos. Nesta seção são realizadas análises a respeito da quantidade e de quais atrasos foram selecionados pelo método *Wrapper*.

Aqui vale destacar que como mencionado no Capítulo 5 o *Wrapper* não foi aplicado aos modelos lineares aqui propostos já que tanto o Modelo Autoregressivo quanto o Autoregressivo de Médias Móveis fazem uso apenas de atrasos sequências como entrada.

Por sua vez, as entradas de um modelo *Ensemble* aplicado a previsão de séries temporais não são os atrasos da série como ocorre com as Redes Neurais simples, mas sim o conjunto de especialistas que terão seus resultados combinados. Dessa forma, ao executar o *Wrapper* sobre as possíveis entradas de um *Ensemble* a configuração selecionada é a combinação de quais especialistas devem ser utilizados para realizar as previsões e não a combinação dos melhores atrasos.

Na Tabela 29 são apresentados os números de entradas escolhidas por cada rede neural treinada e por cada usina testada, além de quais atrasos específicos foram utilizados.

**Tabela 29 – Atrasos selecionados por rede neural e usina.**

<b>Modelo</b>	<b>Água Vermelha</b>	<b>Belo Monte</b>	<b>Ilha Solteira</b>
MLP	5 (1, 3, 4, 2, 6)	3 (1, 6, 4)	2 (1, 3)
RBF	5 (1, 4, 2, 5, 6)	2 (1, 4)	2 (1, 3)
ELM	3 (1, 3, 5)	2 (1, 6)	4 (1, 3, 2, 6)
ELM (CR)	6 (1, 3, 5, 2, 4, 6)	2 (1, 2)	6 (1, 3, 6, 4, 2, 5)
Elman	6 (1, 2, 4, 3, 6, 5)	4 (1, 2, 6, 5)	6 (1, 4, 2, 5, 6, 3)
Jordan	3 (1, 4, 3)	2 (1, 4)	5 (1, 2, 4, 6, 5)
ESN Jaeger	2 (1, 6)	3 (1, 2, 3)	4 (1, 3, 5, 2)
ESN Jaeger (CR)	2 (1, 6)	3 (1, 3, 2)	5 (1, 2, 3, 5, 6)
ESN Ozturk	3 (1, 2, 4)	4 (1, 3, 2, 4)	6 (1, 6, 3, 2, 5, 4)
ESN Ozturk (CR)	3 (1, 2, 4)	3 (1, 3, 6)	4 (1, 3, 2, 4)
<b>Modelo</b>	<b>Paulo Afonso</b>	<b>Tucuruí</b>	
MLP	2 (1, 2)	3 (1, 2, 3)	
RBF	2 (1, 2)	3 (1, 2, 4)	
ELM	5 (1, 6, 4, 5, 2)	3 (1, 2, 3)	
ELM (CR)	5 (1, 6, 4, 2, 5)	3 (1, 2, 3)	
Elman	4 (1, 2, 5, 3)	6 (1, 3, 6, 2, 5, 4)	
Jordan	4 (1, 4, 5, 3)	1 (1)	
ESN Jaeger	3 (1, 6, 3)	2 (1, 2)	
ESN Jaeger (CR)	3 (1, 6, 4)	2 (1, 2)	
ESN Ozturk	5 (6, 1, 5, 2, 3)	2 (1, 6)	
ESN Ozturk (CR)	5 (1, 3, 5, 2, 6)	2 (1, 5)	

**Fonte: Autoria própria.**

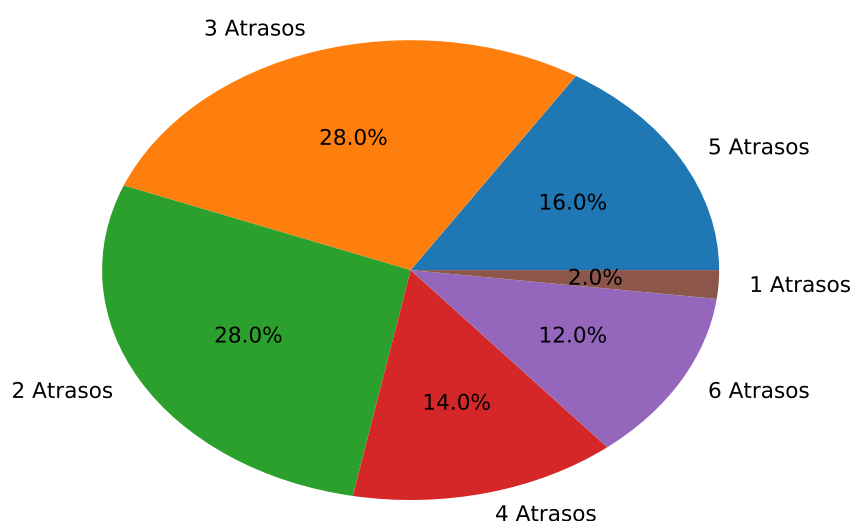
Em seu trabalho, STEDINGER JR (2001) diz que não faz sentido considerar atrasos não sequenciais na previsão de séries de vazões. Todavia, ao analisar-se os dados da Tabela 29 verifica-se que foram vários os casos em que os atrasos selecionados não são sequenciais, o que contradiz tal ideia. É justo dizer que o autor testou especificamente um modelo linear de previsão derivado do Modelo Auto-regressivo, o Modelo Periódico Auto-regressivo (PAR). Redes Neu-

rais Artificiais são estruturas não lineares com poder de mapeamento universal e comportamento complexo. Portanto, antecipar o resultado que determinadas entradas terão no desempenho da previsão não é uma tarefa trivial.

Pelos dados da Tabela 29 verifica-se que apesar de ser definido que os modelos preditivos poderiam utilizar até 6 atrasos, ao final da execução do método *Wrapper* foram poucas as vezes que todos os atrasos disponíveis foram utilizados. Nenhum dos modelos testados para as usinas de Belo Monte e Paulo Afonso utilizaram 6 atrasos. Se tratando de Tucuruí apenas a rede de Elman utilizou todos eles. Destaca-se ainda a rede de Jordan treinada para a usina de Tucuruí, a qual teve os melhores resultados utilizando apenas um dos atrasos.

Com o objetivo de facilitar as análises no que se refere a quantidade de entradas selecionadas, apresenta-se o Gráfico 64 com o percentual de uso por quantidade de atrasos. Aqui não foi considerado se os atrasos são sequenciais ou não.

**Gráfico 64 – Percentual do número de atrasos selecionado.**

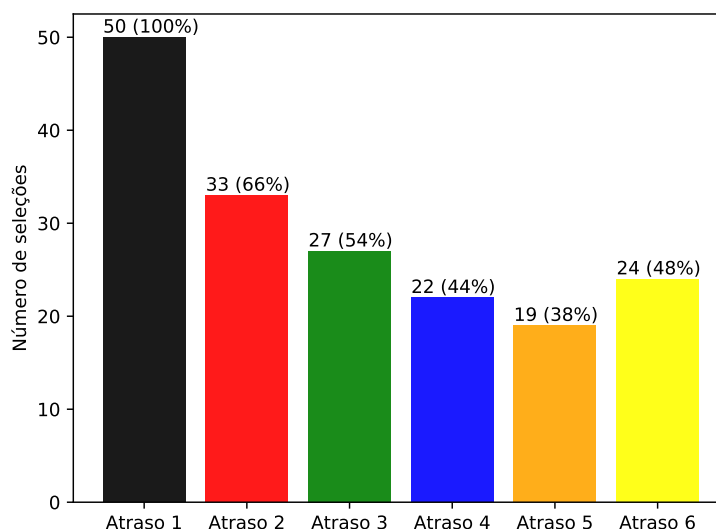


**Fonte: Autoria própria.**

Nota-se que mais da metade das redes (56%) utilizaram no máximo 3 entradas. Apenas em 28% dos testes foram selecionados 5 ou 6 atrasos. Um resultado ainda mais surpreendente é que em 2% dos casos as previsões foram realizadas fazendo uso de apenas 1 atraso.

Esses fatos sugerem que a quantidade de entradas não tem relação direta com a qualidade da previsão. Em vez disso, faz mais sentido utilizar aqueles que tenham uma maior correlação com o dado que se deseja prever. Partindo dessa afirmação é razoável pensar que o atraso mais significativo para a previsão é o atraso 1 (o imediatamente anterior ao valor previsto) e por sua vez, o atraso 6 o menos significativo ou com menor correlação. Com o intuito de validar essa hipótese o Gráfico 65 apresenta o percentual dos casos em que cada um dos atrasos individuais foi utilizado pelas previsões.

O Gráfico 65 comprova a hipótese que o atraso 1 é o mais significativo, uma vez

**Gráfico 65 – Uso de cada atraso individual.**

**Fonte: Autoria própria.**

que o mesmo foi utilizado por todas (100%) as previsões realizadas nesse trabalho. Entretanto, a segunda parte da hipótese não se faz verdadeira, visto que o atraso menos utilizado foi o 5 em apenas 38% dos casos contra 48% do atraso 6. Dos atrasos mais significantes tem-se em segundo lugar o 2 com 66%, em terceiro o 3 com 54%, em quarto o 6 com 48%, em quinto lugar o atraso 4 com 44% e em último o 5 com 38%.

Apesar da hipótese de atrasos mais próximos serem mais significativos não ter se tornado uma regra válida, pelos dados apresentados na Tabela 29 e pelo Gráfico 65 é possível pelo menos observar que há uma tendência para esse comportamento.

## 5.11 TENTATIVAS DE MELHORIA DOS RESULTADOS

Após a realização de todos os testes propostos, alguns últimos esforços foram empregados na tentativa de melhorar as previsões já realizadas através da inserção de novas variáveis de entrada. Nessa seção é descrito o uso de dados referentes aos fenômenos meteorológicos El Niño e La Niña como entradas dos modelos.

O fenômeno meteorológico conhecido como El Niño é caracterizado pelo aquecimento das águas do Oceano Pacífico, mais especificamente na região Equatorial, de modo que tanto as águas superficiais como as sub-superficiais apresentam temperaturas maiores que o normal. Esse aumento na temperatura ocorre devido à mudança no vento nessa região. Durante sua ocorrência os ventos presentes no centro do Oceano Pacífico sopram com menos força ou mesmo têm seu sentido invertido (normalmente sopram da costa oeste da América do Sul em direção a costa leste da Austrália), o que resulta na diminuição da ressurgência de águas profundas. Isto, por sua vez gera um acúmulo de águas mais quentes na costa oeste da América do Sul (OLIVEIRA, 2001).

Como resultado do acúmulo de águas quentes no litoral oeste da América do Sul, há uma maior evaporação que gera grandes deslocamentos de massas de ar por todo o globo. Neste local ocorrem chuvas mais fortes e mais frequentes. Na Indonésia e Austrália o resultado é a diminuição das chuvas levando a um clima mais seco. Nos estados centrais dos Estados Unidos os invernos são menos rigorosos e na Europa ocorrem verões extremamente quentes (OLIVEIRA, 2001).

No que se refere ao Brasil, os efeitos do El Niño variam de acordo com a região e com a intensidade do evento, podendo causar:

- **Região Norte e Nordeste** – diminuição significativa nas chuvas, levando a um clima mais seco no sertão nordestino e na Amazônia;
- **Região Sudeste** – aumento nas temperaturas;
- **Região Sul** – aumento das chuvas e da temperatura.

O La Niña também é um fenômeno natural, entretanto seu comportamento é o oposto do apresentado pelo El Niño. Ao invés dos ventos que vão da costa oeste da América do Sul em direção a costa leste da Austrália soprarem com menos força eles ficam mais intensos, o que faz com que as águas quentes normalmente depositadas no Oceano Pacífico Central sejam empurradas para o oeste, ficando mais longe do litoral da América Latina (OLIVEIRA, 2001).

Como consequência, o que se deposita no oeste da costa da América do Sul é uma massa de águas frias, que apresentam uma menor evaporação e leva a uma menor formação de nuvens de chuva e, conseqüentemente, a um clima mais seco. Em termos gerais pode-se dizer que o El Niño é o aumento da temperatura das águas do Oceano Pacífico Equatorial enquanto que o La Niña é o resfriamento.

Os efeitos do La Niña também podem ser sentidos por todo o globo: na Colômbia, esse evento climático causa chuvas fortes e enchentes; em contrapartida, no Uruguai, causa uma diminuição de precipitação e conseqüentemente um clima mais seco; no Peru o resultado é uma tendência maior de seca. Na costa oeste dos Estados Unidos os anos de ocorrência do La Niña apresentam invernos mais rigorosos e uma tendência a dias mais frios durante todo o ano. Do outro lado do Oceano Pacífico, a costa leste da Ásia apresenta um aumento nas chuvas e temperaturas mais altas no leste do continente (OLIVEIRA, 2001).

Os efeitos do La Niña sobre o Brasil também variam de acordo com a região e com a intensidade do evento, podendo gerar:

- **Região Nordeste** – chegada de frentes frias e chuvas acima da média;
- **Região Norte** – tendência a chuvas torrenciais no norte e leste da Amazônia;
- **Região Sudeste** – durante o inverno as temperaturas ficam próximas a média;



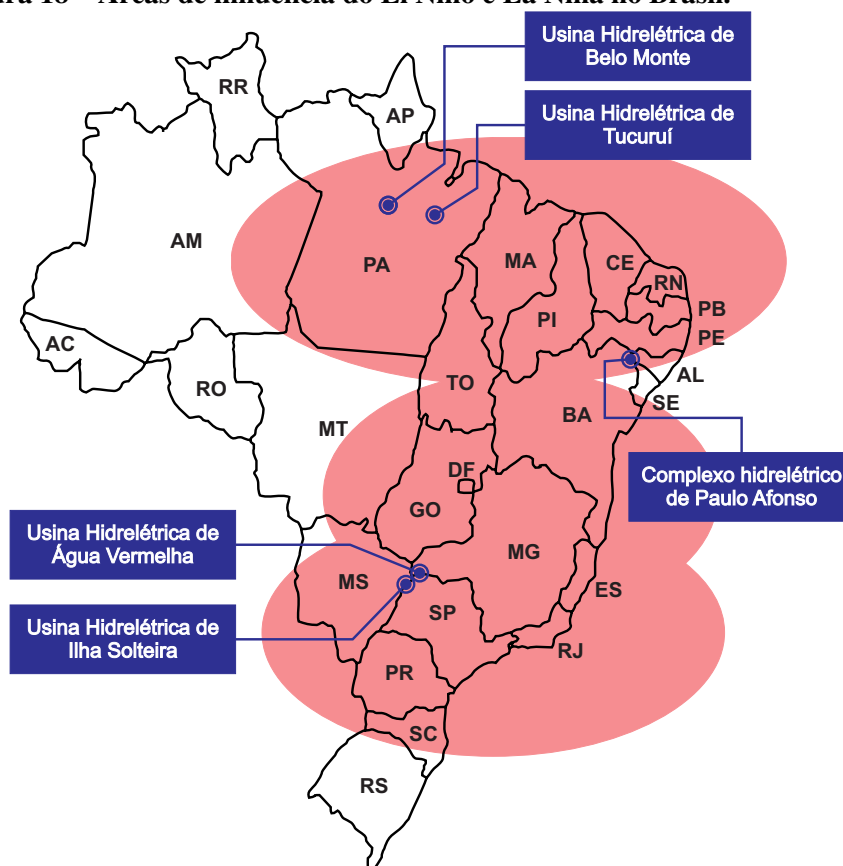
- **Região Sul** – passagens rápidas de frentes frias, chuvas acima da média nos estados do leste e estiagem no oeste da região.

Em geral, os episódios de El Niño e La Niña se alternam com uma frequência de 3 a 7 anos, havendo ainda intervalos em que nenhum dos dois ocorre e a temperatura das águas do Oceano Pacífico Equatorial fica então dentro da média. Por sua vez, a frequência entre as ocorrências do mesmo evento, ou seja, o tempo entre duas ocorrências do El Niño ou do La Niña varia de 1 a 10 anos, não sendo incomum casos nos quais o mesmo evento se repete várias vezes consecutivas.

Os episódios do El Niño são sempre mais intensos que os do La Niña, ou seja, a variação de temperatura provocada pelo fenômeno é sempre maior. Outro fato interessante é que nas últimas décadas a frequência com que ocorre o La Niña vem diminuindo, já sendo significativamente menor que o número de ocorrências do El Niño no mesmo período (CPTEC, 2018).

Como pode-se observar, o El Niño e o La Niña são eventos climáticos com influência direta sobre o padrão de chuvas brasileiro, consequentemente influenciando os níveis dos rios e suas vazões. Com o objetivo de facilitar o entendimento a respeito da atuação desses fenômenos sobre as vazões afluentes às usinas hidrelétricas tratadas nesse estudo, a Figura 18 apresenta o mapa com a localização das plantas hidráulicas consideradas nesse trabalho acrescido da marcação das áreas sob a influência do El Niño e do La Niña.

**Figura 18 – Áreas de influência do El Niño e La Niña no Brasil.**



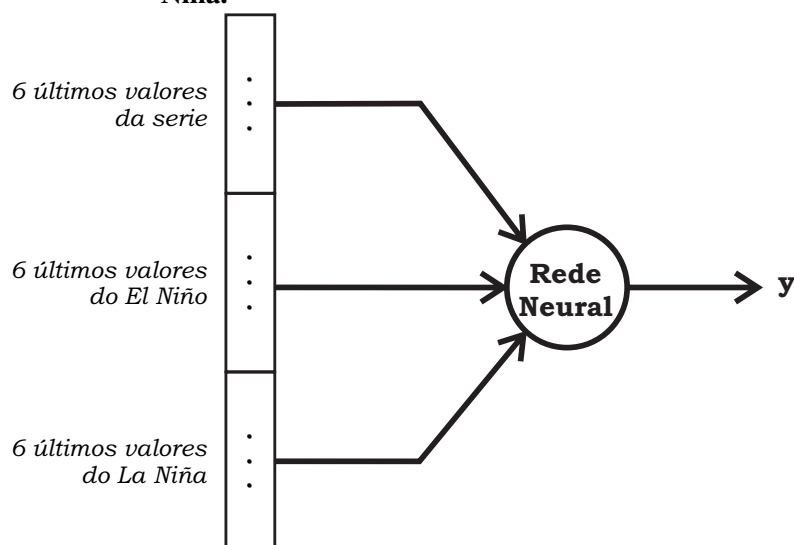
Fonte: Adaptado de (CPTEC, 2018).

Pela Figura 18 é evidente que todas as usinas aqui consideradas estão localizadas em áreas sob influência direta dos fenômenos meteorológicos em questão. Portanto, o uso dessas informações na previsão de suas vazões pode melhorar os resultados atingidos pelos preditores.

Com o objetivo de diminuir o tempo necessário para os testes foi considerada apenas a Usina Hidrelétrica de Tucuruí, pois é aquela com maior vazão média (Gráfico 11). Além disso foram ajustadas apenas as redes neurais ELM, ELM (CR), ESN Jaeger, ESN Jaeger (CR), ESN Ozturk e ESN Ozturk (CR) por possuírem treinamento mais rápido, e foram realizadas apenas previsões com horizonte de 1 passo a frente. Como nos testes principais, aqui o número de neurônios também foi variado de 5 a 200 com intervalos de 5 em 5 e foi aplicado o Método *Wrapper* para selecionar as melhores entradas.

Novamente é utilizado um número máximo de 6 atrasos da série, conforme definido pelo Setor Elétrico Brasileiro. Assim, também serão considerados os valores do El Niño e La Niña dos últimos 6 meses na entrada das redes neurais. A Figura 19 exemplifica como foram compostas as entradas com os atrasos da série e os valores do El Niño e La Niña.

**Figura 19 – Entradas do modelo de previsão com El Niño e La Niña.**



**Fonte: Autoria própria.**

Pela Figura 19 nota-se que agora cada entrada é composta pelos 6 atrasos já utilizados juntamente com os 6 últimos valores do efeito El Niño e os 6 últimos valores do efeito La Niña, totalizando 18 entradas. É importante mencionar que um maior número de entradas resulta em maior complexidade ao executar o método *Wrapper*. Como consequência isso leva a um maior tempo de treinamento das redes, uma vez que agora existem mais pesos sinápticos para serem definidos, justificando mais uma vez a utilização apenas dos preditores com treinamento analítico.

Do mesmo modo que ocorreu nos testes principais, aqui também foram realizadas as etapas de pré e pós-processamento dos dados (Figura 17), de maneira que tanto os valores das vazões mensais quanto os valores dos eventos El Niño e La Niña foram padronizados e norma-

lizados antes de serem inseridos nas redes neurais, e após as mesmas realizarem as previsões, os valores foram desnormalizados e despadronezados.

Foram realizadas 3 novas avaliações: a primeira com a utilização das intensidades dos eventos, a segunda com o valor da variação de temperatura do Oceano Pacífico Equatorial e a terceira realizou uma ponderação nas previsões próximas aos eventos. Os testes são descritos nas seções 5.11.1, 5.11.2 e 5.11.3, respectivamente.

### 5.11.1 Utilização das Intensidades do El Niño e La Niña

No Brasil, o Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) que faz parte do Instituto Nacional de Pesquisas Espaciais (INPE) é o setor governamental responsável pelo acompanhamento das ocorrências dos eventos El Niño e La Niña. Em seu site o CPTEC disponibiliza os dados históricos da intensidade de todas as ocorrências do El Niño <<http://enos.cptec.inpe.br/~renos/misc/elnino.html>><sup>4</sup> e do La Niña <<http://enos.cptec.inpe.br/~renos/misc/lanina.html>><sup>5</sup> desde 1892, ano em que os eventos começaram a ser medidos. Eles são classificados em três intensidades, forte, moderada e fraca, de acordo com a metodologia Trenberth (1997). Para que estes valores pudessem ser inseridos como entradas nas redes neurais para cada intensidade foi dado um valor numérico entre 1 e 3. Aos meses em que não há a incidência de nenhum dos dois eventos foi atribuído o valor 0.

Na Tabela 30 são apresentados os resultados obtidos com a adição dos valores da intensidade do El Niño e do La Niña nas previsões realizadas para a usina de Tucuruí com  $P = 1$ . Além do MSE e MAPE das novas previsões, a tabela também contém os resultados obtidos pelas redes neurais sem a utilização dessas novas variáveis, além da taxa de melhora no valor do MSE de cada modelo.

**Tabela 30 – Resultados Tucuruí com intensidade do El Niño de 2006 a 2015,  $P = 1$ .**

Modelo	Sem El Niño		Com El Niño		Melhora
	MSE	MAPE	MSE	MAPE	
ELM	6.824.285,76	15,48%	<b>6.557.780,67</b>	<b>15,16%</b>	<b>3,91%</b>
ELM (CR)	6.829.099,84	15,67%	6.597.086,27	15,48%	3,40%
ESN Jaeger	6.936.607,01	15,53%	6.696.143,06	15,12%	3,47%
ESN Jaeger (CR)	6.975.358,52	15,50%	6.805.443,89	16,97%	2,44%
ESN Ozturk	6.757.516,84	16,31%	6.834.150,62	15,86%	-1,1341%
ESN Ozturk (CR)	7.119.853,31	17,59%	6.652.053,42	16,64%	6,57%

**Fonte: Autoria própria.**

Voltando aos dados da Tabela 21 verifica-se que a melhor previsão realizada para a usina de Tucuruí com  $P = 1$  pelos modelos de previsão já descritos foi do *Ensemble* com combinador de RBF, tendo MSE real de 5.776.403,72. Pela Tabela 30 nota-se que, com a adição

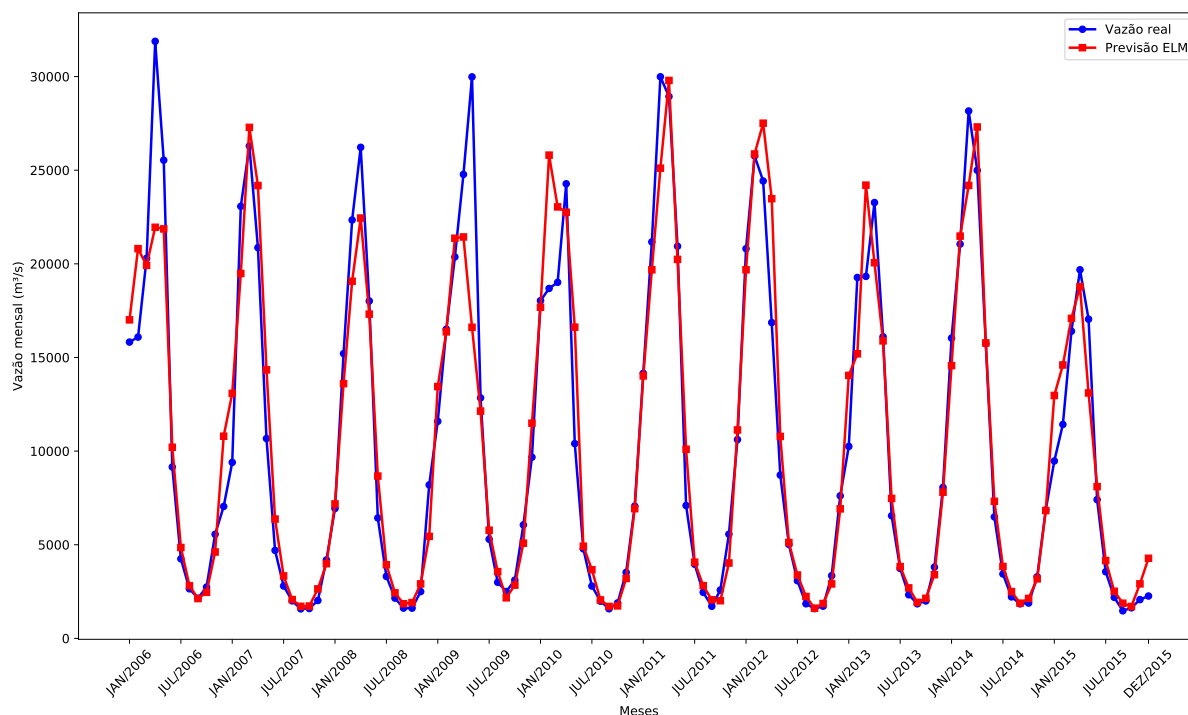
<sup>4</sup> Acessado em 9 de janeiro de 2019.

<sup>5</sup> Acessado em 9 de janeiro de 2019.

dos valores referentes a intensidade do El Niño e La Niña, a melhor previsão foi realizada pela ELM, tendo MSE real de 6.557.780,67, um valor mais elevado que o obtido anteriormente. Portanto, a melhor previsão ainda é do modelo sem as intensidades do El Niño.

O Gráfico 66 apresenta a melhor previsão realizada pelos modelos com a adição da intensidade do El Niño e La Niña para a usina de Tucuruí com  $P = 1$ .

**Gráfico 66 – Melhor previsão Tucuruí com intensidade do El Niño de 2006 a 2015 – 1 passo a frente.**



**Fonte: Autoria própria.**

Apesar do desempenho geral não ter melhorado, os testes mostraram que os resultados individuais das redes neurais tiveram pequenos aumentos de desempenho. Excluindo-se a rede ESN Ozturk que apresentou um resultado 1,1341% pior, com a adição da intensidade dos fenômenos, todos os outros modelos preditivos tiveram ganhos de performance, destacando-se a rede ESN Ozturk (CR) com o maior percentual de melhora no valor do MSE, de 6,57%.

A média de melhora no valor do MSE das redes testadas foi de 3,11%, que indica um possível incremento de desempenho. Infelizmente são necessários mais testes com esse cenário para que esta hipótese possa ser confirmada, de modo que tal tarefa ficará a cargo dos trabalhos futuros.

Todavia, embora a média de melhora dos resultados pareça baixa, a previsão de vazões afluentes é uma das principais etapas no planejamento da operação energética de um país. Dessa forma, aumentar a eficiência de uma previsão em 3,11% significa planejar com mais acurácia a operação a partir de fontes renováveis. A consequência é uma menor utilização das termoelétricas e das usinas nucleares, resultando em uma economia de milhares ou milhões de reais, além da possível diminuição do preço da energia elétrica para a população.

### 5.11.2 Utilização dos Valores Reais do El Niño e La Niña

Como mencionado na Seção 5.11 os fenômenos do El Niño e La Niña são caracterizados pelo aumento ou diminuição das temperaturas marinhas. Nessa segunda abordagem é exatamente o valor dessa variação de temperatura que foi utilizado como entrada para os modelos preditivos. A agência governamental americana *National Oceanic and Atmospheric Administration* (Administração Nacional Oceânica e Atmosférica) através do *National Weather Service* (Serviço Nacional de Meteorologia) realiza o acompanhamento das temperaturas do Oceano Pacífico Equatorial, disponibilizando em seu site <[https://origin.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ONI\\_v5.php](https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php)><sup>6</sup> os dados históricos de todas as oscilações de temperatura marinha que caracterizam as ocorrências dos eventos El Niño e La Niña desde o ano de 1950.

A série histórica de vazões da usina de Tucuruí tem início no ano de 1931, tendo 20 anos a mais de dados que as variações na temperatura marinha. Para que os resultados deste teste pudessem ser comparados com as demais previsões realizadas o conjunto de treinamento para os modelos com a adição do valor da variação da temperatura média oceânica foi considerado com dados de 1950 até 1995. Assim, os conjuntos de validação e teste ficam inalteradas possibilitando que os resultados possam ser comparados.

Na Tabela 31 são apresentados os resultados obtidos pelas previsões para a usina de Tucuruí com  $P = 1$  e com a adição do valor da variação da temperatura marinha como entrada. Também são reapresentados os resultados das previsões de cada modelo fazendo uso apenas dos atrasos da série de vazões e a taxa de melhora nos resultados individuais.

**Tabela 31 – Resultados Tucuruí com valor do El Niño de 2006 a 2015, P = 1.**

Modelo	Sem El Niño		Com El Niño		Melhora
	MSE	MAPE	MSE	MAPE	
ELM	6.824.285, 76	15, 48%	5.751.897, 20	14, 90%	15, 71%
<b>ELM (CR)</b>	6.829.099, 84	15, 67%	<b>5.631.678, 97</b>	<b>16, 10%</b>	<b>17, 53%</b>
ESN Jaeger	6.936.607, 01	15, 53%	6.104.716, 89	15, 30%	11, 99%
ESN Jaeger (CR)	6.975.358, 52	15, 50%	6.264.228, 96	16, 05%	10, 19%
ESN Ozturk	6.757.516, 84	16, 31%	6.615.189, 86	15, 53%	2, 11%
ESN Ozturk (CR)	7.119.853, 31	17, 59%	6.900.596, 27	16, 26%	3, 08%

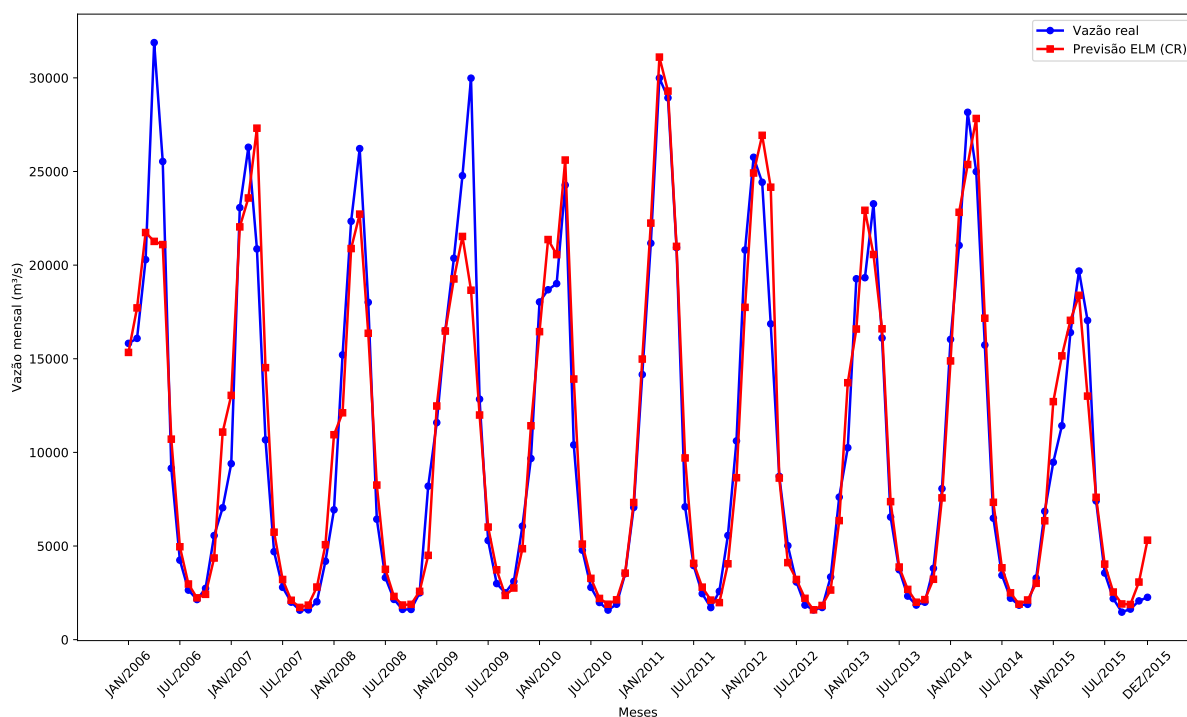
**Fonte: Autoria própria.**

Pelos dados da Tabela 31 observa-se que os resultados individuais de todas as redes neurais testadas apresentaram incremento de desempenho. O destaque fica para as redes ELM (CR) e ELM com 17, 53% e 15, 71% de melhora, respectivamente. Por sua vez, o menor percentual de melhora individual foi da rede ESN Ozturk, apresentando um desempenho comparado de 2, 11% superior com a utilização das variações de temperatura marinhas.

<sup>6</sup> Acessado em 9 de janeiro de 2019.

Além da melhora dos resultados individuais, o resultado geral alcançado por este teste superou o MSE de 5.776.403, 72 da melhor previsão realizada para a usina de Tucuruí. Na nova proposta, a rede ELM (CR) obteve um MSE de 5.631.678, 97. O Gráfico 67 apresenta a previsão realizada para a usina de Tucuruí com  $P = 1$  e a utilização dos atrasos e das variações de temperatura.

**Gráfico 67 – Melhor previsão Tucuruí com valor do El Niño de 2006 a 2015 – 1 passo a frente.**



**Fonte: Autoria própria.**

Comparando os testes descritos nessa seção com os da Seção 5.11.1 é evidente que a utilização dos valores da variação de temperatura das águas do Oceano Pacífico Equatorial se fez mais eficiente que a utilização da intensidade dos eventos do El Niño e La Niña, tendo não só apresentado os maiores aumentos no desempenho individual das redes neurais como também o melhor resultado geral para as previsões da usina de Tucuruí com 1 passo a frente.

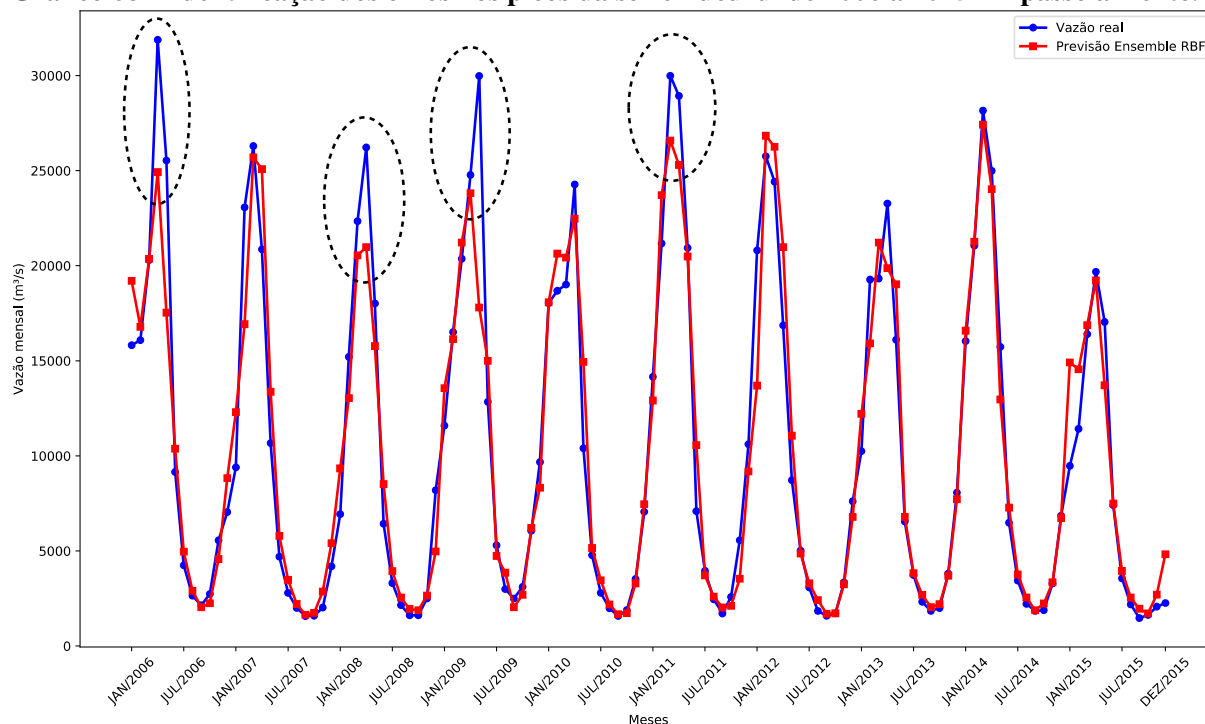
Apesar dos resultados apresentados aqui terem sido melhores que as previsões utilizando apenas os atrasos da série é necessária a realização de mais testes. Em média, essa técnica melhorou o resultado individual de cada modelo de previsão em 10, 10%. Como trabalho futuro a mesma pode ser aplicada aos demais modelos preditivos, podendo até resultar em previsões ainda melhores que as alcançadas aqui.

### 5.11.3 Aproximação dos Picos

A última tentativa de melhoria dos resultados realizada por esse trabalho foi a minimização do erro de previsão nos picos da série. Ao analisar o Gráfico 50, que mostra a melhor

previsão realizada para a usina de Tucuruí com  $P = 1$ , nota-se que em alguns picos da série o erro de previsão é maior que a média. O Gráfico 68 reinterpreta a melhor previsão e destaca os picos em que os valores previstos ficaram muito abaixo dos valores reais.

**Gráfico 68 – Identificação dos erros nos picos da série Tucuruí de 2006 a 2015 – 1 passo a frente.**



**Fonte: Autoria própria.**

Observando o Gráfico 68 é fácil ver 4 picos nos quais os valores previstos ficaram muito abaixo dos valores reais da série. Esse fato indica que nessas ocasiões houve um aumento incomum na vazão do rio, que, como já foi mencionado na Seção 5.11, pode ter sido causado pelos fenômenos meteorológicos El Niño e La Niña. Entretanto ao comparar os períodos desses picos na série com as datas das ocorrências do El Niño e La Niña verificou-se que os mesmos não ocorreram nos mesmos meses, mas de 4 a 8 meses antes da ocorrência dos fenômenos climáticos.

Com o objetivo de aumentar os valores previstos nesses picos que antecedem a ocorrência do El Niño foi adotada uma estratégia de ponderação. Após a realização da previsão pelo modelo, a mesma é multiplicada por um fator de acordo com a intensidade do evento que venha a ocorrer alguns meses depois. A Equação 5.4 descreve esse conceito

$$y_t^{final} = y_t(1 + (i_{t+MI}FP)) \quad (5.4)$$

de modo que  $y_t^{final}$  é o valor da previsão para o instante  $t$  depois da ponderação,  $y_t$  é a previsão original para o instante  $t$ ,  $i_{t+MI}$  é o valor da intensidade do El Niño (0: ausente, 1: fraca, 2: moderada e 3: forte) para o instante  $t + MI$ ,  $FP$  é o fator de ponderação (aqui definido como 0, 1) e  $MI$  é o número de meses antes da ocorrência do evento considerado na ponderação. Consideramos esse último valor como 5.

Note que na Equação 5.4 não é considerada a intensidade do La Niña. Isso se deve ao fato de que nos teste iniciais não foi possível estabelecer uma regra utilizando esse valor que fizesse com que as previsões alcançassem valores mais próximos dos picos desejados. Outro fato que vale a pena mencionar é que os valores *MI* e *FP* foram definidos através de testes empíricos, tendo sido consideradas várias possibilidades e escolhida aquela que levou o comportamento das previsões mais próximo do desejado.

Na Tabela 32 são apresentados os resultados obtidos pelo teste de ponderação da intensidade do El Niño para a usina de Tucuruí com  $P = 1$ .

**Tabela 32 – Resultados Tucuruí com ponderação do El Niño de 2006 a 2015, P = 1.**

Modelo	Sem El Niño		Com El Niño		Melhora
	MSE	MAPE	MSE	MAPE	
<b>ELM</b>	6.824.285,76	15,48%	<b>7.116.337,61</b>	<b>19,68%</b>	<b>-4,2796%</b>
ELM (CR)	6.829.099,84	15,67%	7.284.576,33	19,99%	-6,6696%
ESN Jaeger	6.936.607,01	15,53%	7.288.683,61	19,68%	-5,0756%
ESN Jaeger (CR)	6.975.358,52	15,50%	7.294.952,32	20,05%	-4,5818%
ESN Ozturk	6.757.516,84	16,31%	7.295.108,01	21,11%	-7,9555%
ESN Ozturk (CR)	7.119.853,31	17,59%	7.541.554,54	22,42%	-5,9229%

**Fonte: Autoria própria.**

Pelos dados da Tabela 32 nota-se que o resultado individual de todos os modelos de previsão piorou com a ponderação, sendo que a rede com a maior queda de desempenho foi a ESN Ozturk, com um MSE 7,9229% maior que o obtido sem a ponderação.

Das seis redes neurais testadas a que apresentou as melhores previsões foi a rede ELM, com MSE de 7.116.337,61. Entretanto esse valor é 23% maior que o obtido pelo melhor modelo preditivo sem ponderação, que foi de 5.776.403,72.

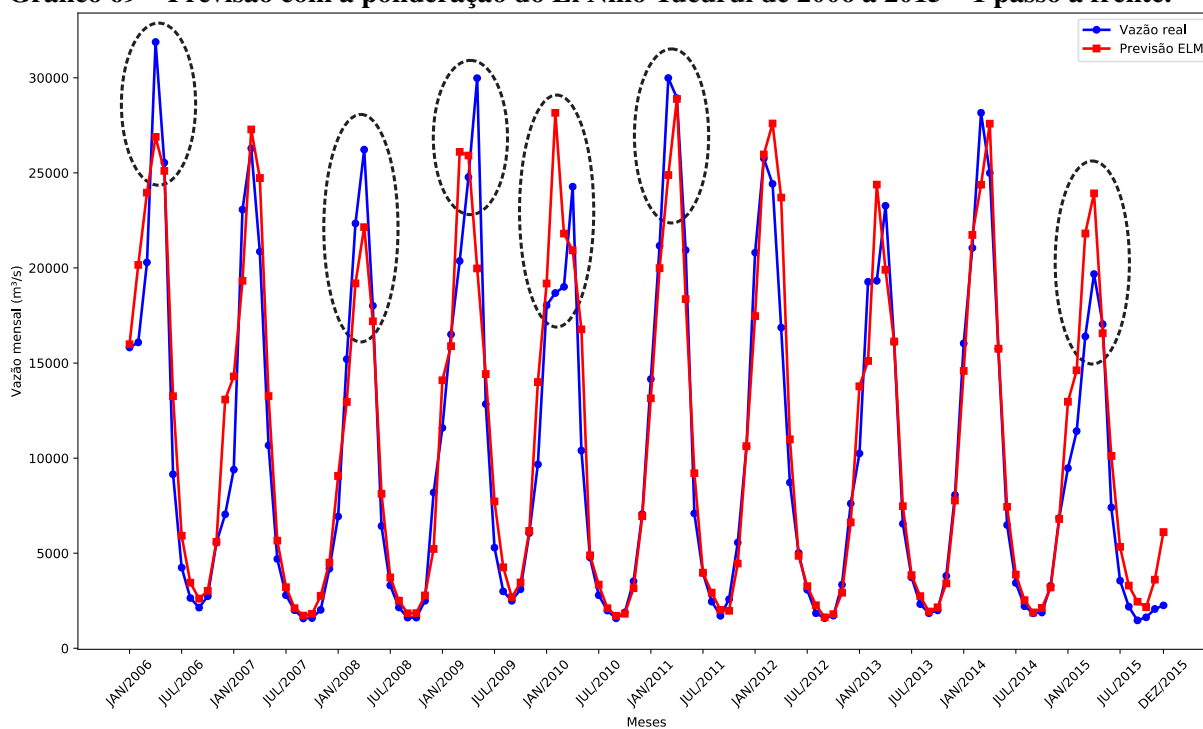
O Gráfico 69 apresenta o resultado da previsão realizada pela rede ELM com ponderação para a usina de Tucuruí com  $P = 1$ .

Comparando os gráficos 68 e 69 é possível observar que o uso da ponderação fez com que alguns dos picos apresentassem previsões com erros menores. No primeiro, verifica-se que nos três primeiros em destaque o valor da vazão real foi maior que o da prevista. Este comportamento se manteve nos mesmos picos do Gráfico 69 com a utilização da ponderação. Todavia, nota-se que no terceiro pico o valor da previsão teve um aumento com a ponderação. O quarto pico em destaque no Gráfico 68, agora no Gráfico 69 apresenta valores de vazão prevista na mesma faixa que as reais. Apesar de um dos picos ter apresentado uma melhora nos valores previstos, analisando o Gráfico 69 verifica-se que para aqueles entre 2010 e 2015 o aumento nos valores previstos foi mais alto que o necessário, ficando nos dois casos maior que as vazões reais.

Ao final dos testes verificou-se que a técnica de ponderação não teve bons resultados nas previsões gerais. Todos os modelos apresentaram desempenho inferior ao uso apenas dos atrasos da série.



**Gráfico 69 – Previsão com a ponderação do El Niño Tucuruí de 2006 a 2015 – 1 passo a frente.**



**Fonte: Autoria própria.**

Por fim, o objetivo de minimizar os erros nos picos foi parcialmente cumprido, mostrando que para tal fim a ponderação pode ser promissora. Novamente são necessários mais testes e ajustes para que um veredito definitivo dessa técnica possa ser dado.

Parcialmente, a inserção dos valores reais do El Niño e La Niña parece ser a proposta mais adequada.

## 6 CONCLUSÃO

A previsão de vazões afluentes é uma etapa muito importante no planejamento energético do país. Na literatura, frequentemente esse problema é tratado por modelos de previsão lineares da família Box & Jenkins. Apesar disso, Redes Neurais Artificiais (RNAs) vem sendo estudadas como alternativas viáveis para a tarefa. Como são modelos não lineares, elas trazem outro tipo de possibilidade de mapeamento do espaço.

Tendo isso em vista, este trabalho realizou a previsão de vazões médias mensais de importantes usinas hidrelétricas brasileiras através da utilização de RNAs - *Perceptron* de Múltiplas Camadas (MLP), redes neurais com Função de Base Radial (RBF), rede de Jordan, rede de Elman, máquinas de aprendizado extremo (ELM), redes neurais com estados de eco (ESN) - e da metodologia *Ensemble* utilizando as redes *feedforward* - MLP, RBF e ELM. Além disso os resultados dos modelos neurais foram comparados com os tradicionais modelos lineares AR e ARMA. Foram considerados quatro horizontes de previsão, a saber  $P = 1, 3, 6$  e  $12$  passos a frente. A seleção dos melhores atrasos para as previsões foi realizada pelo método *Wrapper* para os modelos neurais.

Os testes realizados mostraram que para as previsões com horizonte de 1 passo à frente os *Ensembles* são os modelos mais indicados, uma vez que eles apresentaram os melhores desempenhos em todas as simulações realizadas nesse cenário. Para os demais horizontes as máquinas desorganizadas (ELM e ESN) foram os melhores preditores, sendo o modelo com o maior número de melhores desempenhos a ELM. No que se refere aos modelos lineares este trabalho mostrou sua inferioridade em relação aos modelos neurais, uma vez que em nenhum dos 20 testes realizados o melhor modelo de previsão foi um modelo linear.

Também foram propostas 3 tentativas de melhorias das previsões realizadas através da inserção de dados dos fenômenos climáticos El Niño e La Niña, considerando a usina de Tucuruí e  $P = 1$ . A primeira proposta foi o uso da intensidade desses eventos, tendo resultado na melhoria significativa das previsões individuais dos modelos testados, mas não no resultado geral. A segunda tentativa foi o uso dos valores das variações de temperatura marinha, que se demonstraram a melhor estratégia. Por fim, a última proposta foi a ponderação das previsões dos modelos desenvolvidos com o objetivo de reduzir os erros nos picos da série. Após a realização dos testes verificou-se que para alguns picos os erros foram minimizados enquanto que para os demais os erros aumentaram, o que tornou o resultado geral das previsões pior.

Por fim, conclui-se que para o problema da previsão de séries de vazões médias mensais as Redes Neurais Artificiais tem desempenho superior em relação aos modelos lineares atualmente utilizados. Em especial destaca-se a rede ELM como sendo o melhor modelo preditivo testado, para os resultados gerais. Com relação a quantidade de neurônios, este estudo verificou que para o problema em questão não são necessários mais que 95 unidades. Também constatou-se que apesar dos modelos disporem de 6 atrasos para realizar as previsões, em 58% dos casos

a melhor configuração de entradas é alcançada utilizando até 3 atrasos.

Outra conclusão deste trabalho é o impacto positivo que a utilização das informações referentes aos fenômenos climáticos El Niño e La Niña causaram nas previsões, tendo apresentado melhoria significativa com a utilização da intensidade e do valor desses eventos.

## 6.1 TRABALHOS FUTUROS

Apesar da grande quantidade de testes realizados por este trabalho nota-se que o desempenho apresentado por cada modelo preditivo depende do comportamento da série que se está prevendo. Assim se faz necessário submeter todos os modelos preditivos aqui propostos às séries de outras usinas hidrelétricas, tanto usinas brasileiras como internacionais. Além da aplicação em outras séries mensais é possível ainda utilizar os preditores em séries de vazões diárias, horárias e minuto a minuto.

No que diz respeito ao método de seleção de entradas, este trabalho aplicou apenas o método *Wrapper*, cabendo ainda o teste dos modelos aqui propostos com outras metodologias de seleção de entradas encontradas na literatura, por exemplo filtros e *Stepwise*.

Com relação as Redes Neurais utilizadas por este trabalho cabe realizar variações de suas arquiteturas e algoritmos de treinamento. Na literatura encontram-se Máquinas Desorganizadas com camada de saída não linear que podem ser usadas no problema da previsão de vazões. É conhecido que o algoritmo de retropropagação do erro não é a única forma de treinamento para uma MLP. Dessa forma, outros algoritmos podem ser abordados ou mesmo metaheurísticas podem ser utilizadas.

Um dos diferenciais deste trabalho foi a utilização de uma rede neural não clássica como combinador de um *Ensemble*. Outras Redes Neurais Artificiais também podem ser aplicadas ao problema da previsão de vazões como as Redes Neurais com Estado de Eco. Ainda, com relação aos Ensembles, metodologias que buscam garantir a diversidade dos seus especialistas podem ser aplicados com o objetivo de melhorar os resultados já obtidos por este trabalho.

Os testes utilizando as informações do El Niño e do La Niña mostraram que essas abordagens são promissoras, portanto tais abordagens devem ser seguidas de modo a testar mais arquiteturas de Redes Neurais e *Ensembles*. A proposta que tentou minimizar os picos de erros nos períodos que antecedem o El Niño e o La Niña não se mostrou eficiente de forma geral, entretanto em alguns pontos específicos o erro foi minimizado. Isso possibilita que esta estratégia pode ser inserida como especialista em um modelo Ensemble com objetivo de aumentar a diversidade e assim melhorar seu resultado geral.

## REFERÊNCIAS

- BALLINI, R. **Análise e previsões de vazões utilizando modelos de séries temporais, redes neurais e redes neurais nebulosas**. 2000. 169 f. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2000.
- BARTLETT, P. L. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. **IEEE transactions on Information Theory**, IEEE, v. 44, n. 2, p. 525–536, 1998.
- BERTOLAMI, R.; BUNKE, H. Ensemble methods for handwritten text line recognition systems. In: IEEE. **Systems, Man and Cybernetics, 2005 IEEE International Conference on**. [S.l.], 2005. v. 3, p. 2334–2339.
- BOCCATO, L. **Novas propostas e aplicações de redes neurais com estados de eco**. 2013. 226 f. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2013.
- BOCCATO, L.; SOARES, E. S.; FERNANDES, M. M. L. P.; SORIANO, D. C.; ATTUX, R. Unorganized machines: from turing's ideas to modern connectionist approaches. In: **Natural Computing for Simulation and Knowledge Discovery**. [S.l.]: IGI Global, 2014. p. 221–236.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control, revised ed**. [S.l.]: Holden-Day, 1976.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 4. ed. New Jersey: John Wiley & Sons, 2008. ISBN 978-0-470-27284-8.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC Editora, 2007. ISBN 978-85-216-1564-4.
- BUENO, R. L. S. **Econometria de séries temporais**. 2. ed. [S.l.]: Cengage, 2011. ISBN 978-85-221-1157-2.
- CEPEL, Centro de Pesquisas Energéticas Estratégicas. **Manual de referência modelo NEWAVE**. [S.l.], 2013. Disponível em: <[http://simsee.org/simsee/biblioteca/Brasil/NW201203/ManualReferencia\\_Newave\\_comentado.pdf](http://simsee.org/simsee/biblioteca/Brasil/NW201203/ManualReferencia_Newave_comentado.pdf)>. Acesso em: 25 jun. 2017.
- CPTEC, Centro de Previsão de Tempo e Estudos Climáticos. **El Niño e La Niña - CPTEC/INPE**. INPE, Instituto Nacional de Pesquisas Espaciais, 2018. Disponível em: <<http://enos.cptec.inpe.br/>>. Acesso em: 05 jan. 2019.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals, and Systems (MCSS)**, Springer, v. 2, n. 4, p. 303–314, 1989.
- DAHL, G. E.; STOKES, J. W.; DENG, L.; YU, D. Large-scale malware classification using random projections and neural networks. In: IEEE. **Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on**. [S.l.], 2013. p. 3422–3426.
- DIAS, T. L.; CATALDI, M.; FERREIRA, V. H. Application of neural networks techniques and atmospheric modeling to elaborate inflow forecasts in the rio grande basin (mg), brazil. **Engenharia Sanitaria e Ambiental**, SciELO Brasil, v. 22, n. 1, p. 169–178, 2017.

- EHLERS, R. S. **Análise de séries temporais**. 3. ed. Curitiba: Departamento de Estatística, UFPR, 2005. Disponível em: <<http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>>. Acesso em: 14 jun. 2017.
- ELMAN, J. L. Finding structure in time. **Cognitive science**, Wiley Online Library, v. 14, n. 2, p. 179–211, 1990.
- EPE, Empresa de Pesquisa Energética. **Balço energético nacional 2018: ano base 2017**. Rio de Janeiro: Ministério de Minas e Energia, 2018. Disponível em: <<http://epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-303/topico-419/BEN2018.pdf>>. Acesso em: 20 out. 2018.
- FAN, F. M.; SCHWANENBERG, D.; ALVARADO, R.; REIS, A. A.; COLLISCHONN, W.; NAUMMAN, S. Performance of deterministic and probabilistic hydrological forecasts for the short-term optimization of a tropical hydropower reservoir. **Water Resources Management**, v. 30, n. 10, p. 3609–3625, Aug 2016. ISSN 1573-1650. Disponível em: <<https://doi.org/10.1007/s11269-016-1377-8>>.
- FERREIRA, A. M. Testes não-paramétricos. **Métodos Estatísticos e Delineamento**, 2010.
- FRANCELIN, R.; BALLINI, R.; ANDRADE, M. G. Back-propagation and box & jenkins approaches to streamflow forecasting, latin-iberian-american congress on operations research and system engineering-claio. **Simpósio Brasileiro de Pesquisa Operacional-SBPO, Rio de Janeiro-RJ**, p. 1307–1312, 1996.
- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. **Journal of the american statistical association**, Taylor & Francis, v. 32, n. 200, p. 675–701, 1937.
- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of machine learning research**, v. 3, n. Mar, p. 1157–1182, 2003.
- HANSEN, L. K.; SALAMON, P. Neural network ensembles. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 12, n. 10, p. 993–1001, 1990.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. ISBN 978-85-7307-718-6.
- HAYKIN, S. **Neural networks and learning machines**. 3. ed. New York: Pearson Prentice Hall, 2008. ISBN 978-0131471399.
- HAYKIN, S. O. **Adaptive filter theory**. [S.l.]: Pearson Higher, 2013. ISBN 978-0132671453.
- HODGKIN, A. L.; HUXLEY, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. **The Journal of physiology**, Wiley-Blackwell, v. 117, n. 4, p. 500, 1952.
- HUANG, G. B.; ZHU, Q. Y.; SIEW, C. K. Extreme learning machine: theory and applications. **Neurocomputing**, Elsevier, v. 70, n. 1, p. 489–501, 2006.
- HUANG, G. H.; ZHOU, H.; DING, X.; ZHANG, R. Extreme learning machine for regression and multiclass classification. **Transactions on Systems, MAN, and Cybernetics – PART B: Cybernetics**, IEEE, v. 42, n. 2, p. 513–529, 2012.

HUANG, G. H.; ZHU, Q. Y.; SIEW, C. K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE. **Proceedings of International Joint Conference on Neural Networks**. [S.l.], 2004. v. 2, p. 985–990.

JAEGER, H. The “echo state” approach to analysing and training recurrent neural networks—with an erratum note. **German National Research Center for Information Technology**, GMD Technical Report 148, n. 34, p. 13, 2001.

JAEGER, H. Adaptive nonlinear system identification with echo state networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2003. p. 609–616.

JORDAN, M. I. Attractor dynamics and parallelism in a connectionist sequential. **Cognitive Science**, 1986.

KASIVISWANATHAN, K. S.; SUDHEER, K. P. Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. **Stochastic environmental research and risk assessment**, Springer, v. 27, n. 1, p. 137–146, 2013.

KRISTJANPOLLER, W.; MINUTOLO, M. C. Gold price volatility: A forecasting approach using the artificial neural network–garch model. **Expert Systems with Applications**, v. 42, n. 20, p. 7245 – 7251, 2015. ISSN 0957-4174.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105.

LIMA, C. A. M.; COELHO, A. L. V.; ZUBEN, F. J. V. Ensembles of support vector machines for regression problems. In: IEEE. **Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on**. [S.l.], 2002. v. 3, p. 2381–2386.

LU, J.; PLATANIOTIS, K. N.; VENETSANOPOULOS, A. N.; LI, S. Z. Ensemble-based discriminant learning with boosting for face recognition. **IEEE transactions on neural networks**, Citeseer, v. 17, n. 1, p. 166–178, 2006.

MAGALHÃES, M. H. **Redes neurais, metodologias de agrupamento e combinação de previsores aplicados a previsão de vazões naturais**. 2004. 107 f. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2004.

MALFATTI, M. G. L.; CARDOSO, A. O.; HAMBURGER, D. S. Linear empirical model for streamflow forecast in itaipu hydroelectric dam–parana river basin. **Revista Brasileira de Meteorologia**, SciELO Brasil, v. 33, n. 2, p. 257–268, 2018.

MASON, J. C.; PRICE, R. K.; TEM'ME, A. A neural network model of rainfall-runoff using radial basis functions. **Journal of Hydraulic Research**, Taylor & Francis Group, v. 34, n. 4, p. 537–548, 1996.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.

MOLCHANOV, P.; YANG, X.; GUPTA, S.; KIM, K.; TYREE, S.; KAUTZ, J. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network. In: **The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016.

- MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. 2. ed. São Paulo: Blucher, 2006.
- MORETTIN, P. A.; TOLOI, C. M. C. Previsão de séries temporais. **São Paulo: Atual**, 1987.
- MÜLLER, I. I.; KRÜGER, C. M.; KAVISKI, E. Análise de estacionariedade de séries hidrológicas na bacia incremental de itaipu. **Revista Brasileira de Recursos Hídricos**, v. 3, n. 4, p. 51–71, 1998.
- NILSSON, N. J. **Learning machines: foundations of trainable pattern-classifying systems**. [S.l.]: McGraw-Hill, 1965.
- OLIVEIRA, B. A. S.; NAMETALA, M. S.; CASTORIL, B. R.; CATALANO, M. D.; NAMETALA, C. A. L. Avaliação de uma rede neural artificial como estimador temporal pluviométrico no sistema de abastecimento cantareira. **Revista de Informática Aplicada**, v. 14, n. 1, 2018.
- OLIVEIRA, G. S. **O El Niño e você o fenômeno climático**. [S.l.]: Transtec Editorial, 2001.
- ONS, Operador Nacional do Sistema Elétrico. **Dados hidrológicos / Vazões**. Ministério de Minas e Energia, 2018. Disponível em: <[http://ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados\\_hidrologicos\\_vazoes.aspx](http://ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_vazoes.aspx)>. Acesso em: 15 jan. 2018.
- OZTURK, M. C.; XU, D.; PRÍNCIPE, J. C. Analysis and design of echo state networks for function approximation. **Neural computation**, MIT Press, v. 19, n. 1, p. 111–138, 2007.
- PEREIRA, B. B.; PAIS, M. B. Z.; SALES, P. R. H. **Análise espectral de séries temporais - uma introdução para engenharia, economia e estatística**. Rio de Janeiro: Arte Final Leasing Editora Ltda, 1986.
- PERRONE, M. P. **Improving regression estimation: averaging methods for variance reduction with extensions to general convex measure optimization**. 1993. Tese (Doutorado) — Citeseer, 1993.
- RASMUSSEN, P. F.; SALAS, J. D.; FAGHERAZZI, L.; RASSAM, J. C.; BOBÉ, B. Estimation and validation of contemporaneous parma models for streamflow simulation. **Water Resources Research**, Wiley Online Library, v. 32, n. 10, p. 3151–3160, 1996.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Cognitive modeling**, v. 5, n. 3, p. 1, 1986.
- SACCHI, R.; OZTURK, M. C.; PRÍNCIPE, J. C.; CARNEIRO, A. A. F. M.; SILVA, I. N. Water inflow forecasting using the echo state network: a brazilian case study. In: **IEEE Neural Networks, 2007. IJCNN 2007. International Joint Conference on**. [S.l.], 2007. p. 2403–2408.
- SCHAEFER, A. M.; ZIMMERMANN, H. G. Recurrent neural networks are universal approximators. **International journal of neural systems**, World Scientific, v. 17, n. 04, p. 253–263, 2007.
- SHARKEY, A. J. C. **Combining artificial neural nets: ensemble and modular multi-net systems (perspectives in neural computing)**. Springer Verlag, 1999.

SILVA, B.; TUCCI, C.; COLLISCHONN, W. Previsão de vazão com modelos hidroclimáticos. v. 11, p. 15–29, 01 2006.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais para engenharia e ciências aplicadas - curso pratico**. 1. ed. São Paulo: ARTLIBER, 2010. ISBN 978-85-88098-53-4.

SIQUEIRA, H. V. **Previsão de series de vazões com redes neurais artificiais e modelos lineares ajustados por algoritmos bio-inspirados**. 2009. 156 f. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2009.

SIQUEIRA, H. V. **Máquinas desorganizadas para previsão de séries de vazões**. 2013. 218 f. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2013.

SIQUEIRA, H. V.; BOCCATO, L.; ATTUX, R. R. F.; LYRA FILHO, C. Echo state networks in seasonal streamflow series prediction. **Learning and Nonlinear Models**, v. 10, p. 181–191, 2012.

SOARES FILHO, S. Planejamento da operação de sistemas hidrotérmicos. **SBA-Controle e Automação**, v. 1, n. 2, p. 122–131, 1987.

SPIEGEL, M. R. **Estatística**. 3. ed. São Paulo: Makron Books, 1993.

SRINIVASAN, D.; LIEW, A. C.; CHANG, C. S. A neural network short-term load forecaster. **Electric Power Systems Research**, Elsevier, v. 28, n. 3, p. 227–234, 1994.

STEDINGER JR. **Report on the evaluation of cepel's par models: techical report**. [S.l.]: School of Civil and Environmental Engineering-Cornell University Ithaca . . . , 2001.

TANG, Z.; FISHWICK, P. A. Feedforward neural nets as models for time series forecasting. **ORSA journal on computing**, *Inform*s, v. 5, n. 4, p. 374–385, 1993.

TEIXEIRA JÚNIOR, L. A.; FARIA JÚNIOR, A. E.; PEREIRA, R. V. de B.; SOUZA, R. C.; FRANCO, E. M. C. Combinação linear wavelet sarima-rna com estágios múltiplos na previsão de séries temporais. **Blucher Marine Engineering Proceedings**, Blucher, v. 2, n. 1, p. 295–307, 2016.

THOBER, S.; KUMAR, R.; WANDERS, N.; MARX, A.; PAN, M.; RAKOVEC, O.; SAMANIEGO, L.; SHEFFIELD, J.; WOOD, E. F.; ZINK, M. Multi-model ensemble projections of european river floods and high flows at 1.5, 2, and 3 degrees global warming. **Environmental Research Letters**, IOP Publishing, v. 13, n. 1, p. 014003, jan 2018.

TRENBERTH, K. E. The definition of el nino. **Bulletin of the American Meteorological Society**, American Meteorological Society, v. 78, n. 12, p. 2771–2778, 1997.

TURING, A. M. Intelligent machinery. **National Physical Laboratory**, 1948. Disponível em: <[http://www.alanturing.net/turing\\_archive/archive/1/132/L32-001.html](http://www.alanturing.net/turing_archive/archive/1/132/L32-001.html)>. Acesso em: 16 jun. 2017.

WERBOS, P. J. **Beyond regression: new tools for prediction and analysis in the behavioral sciences**. 1974. Tese (Doutorado) — Harvard University, Cambridge, 1974.



WICHARD, J. D.; OGORZALEK, M. Time series prediction with ensemble models. In: IEEE. **Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on.** [S.l.], 2004. v. 2, p. 1625–1630.

ZEALAND, C. M.; BURN, D. H.; SIMONOVIC, S. P. Short term streamflow forecasting using artificial neural networks. **Journal of hydrology**, Elsevier, v. 214, n. 1, p. 32–48, 1999.

**APÊNDICE**

## APÊNDICE A - ARTIGOS PUBLICADOS EM ANAIS DE CONGRESSOS

BELOTTI, J. T.; LAZZARIN, L. N. A.; USBERTI, F. L.; SIQUEIRA, H. V. **Seasonal streamflow series forecasting using recurrent neural networks**. In: 2018 IEEE Latin American Conference on Computational Intelligence (LACCI), 2018, Guadalajara. 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI), 2018. p. 1.

LAZZARIN, L. N.; BELOTTI, J. T.; SILVA, L. V.; MARINHO, M. H. N.; ALVES, T. A.; TADANO, Y. S.; Siqueira, H. V. **Redes neurais feedforward aplicadas na avaliação do impacto da poluição atmosférica e variáveis climáticas na saúde humana**. In: 1st Iberic Conference on Theoretical and Experimental Mechanics and Materials / 11th National Congress on Experimental Mechanics, 2018, Porto - Portugal. Proceedings of the 1st Iberic Conference on Theoretical and Experimental Mechanics and Materials / 11th National Congress on Experimental Mechanics, 2018. v. 1. p. 275-284.

ARAÚJO, L. N.; BELOTTI, J. T.; ALVES, T. A.; TADANO, Y. S.; TROJAN, F.; SIQUEIRA, H. V. **Analysis of regularized echo state networks on the impact of air pollutants on human health**. In: International Conference On Computational Vision and Bio Inspired Computing 2018, 2018, Coimbatore. Proceedings of the ICCVBIC 2018, n. 89, 2018.