

**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  
NEURAIS ARTIFICIAIS**

**EXAME DE QUALIFICAÇÃO**

**PONTA GROSSA  
2018**

**JÔNATAS TRABUCO BELOTTI**

**PREVISÃO DE VAZÕES AFLUENTES UTILIZANDO REDES  
NEURAIS ARTIFICIAIS**

Exame de qualificação apresentado ao Programa de Pós-Graduação em Ciência da Computação da Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa.

Orientador: Hugo Valadares Siqueira  
Co-orientador: Fábio Luiz Usberti

**PONTA GROSSA  
2018**

## **RESUMO**

A matriz energética brasileira é predominantemente composta por usinas hidrelétricas. De modo que garantir máxima eficiência no funcionamento dessas plantas elétricas apresentaria um impacto significativo no custo da produção de energia elétrica. Determinar a vazão afluente a uma usina hidrelétrica é uma tarefa de fundamental importância na eficiência da geração de energia elétrica. Durante os anos, vários modelos lineares como o Modelo Linear Autorregressivo e não lineares como Redes Neurais Artificiais foram utilizados no problema da previsão de vazões afluentes. Esse trabalho realizou a previsão de vazões mensais afluentes através da utilização de 2 arquiteturas de redes neurais: *Perceptron* de Múltiplas Camadas (MLP) e Redes de Função de Base Radial (RBF). 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. Como resultado parcial verificou-se que a MLP obteve melhores previsões em 85% dos testes realizados enquanto a RBF foi melhor em apenas 15% dos testes.

**Palavras-chaves:** Previsão de séries temporais. Previsão de séries de vazões. Redes Neurais Artificiais.

## ABSTRACT

The brazilian energy matrix is predominantly composed of hydroelectric power plants. So ensuring maximum efficiency in the operation of these power plants would have a significant impact on the cost of electricity production. Determining the affluent flow to a hydroelectric plant is a fundamental task in the efficiency of electric power generation. During the years several linear models such as the Linear Model Autoregressive and nonlinear as Artificial Neural Networks were used in the problem of the prediction of tributaries. This work has predicted affluent monthly flows through the use of 2 neural network architectures: Multilayer Perceptron (MLP) and Radial Basis Function Networks (RBF). The tests were carried out in the historical series of the plants of Água Vermelha, Belo Monte, Ilha Solteira, Paulo Afonso and Tucuruí with forecasts horizons of 1, 3, 6 and 12 steps forward. As a partial result it was found that the MLP obtained better predictions at 85% of the tests performed while the RBF was better at only 15% of the tests.

**Key-words:** Streamflow series forecasting. River flow forecast. Neural network.

## LISTA DE ILUSTRAÇÕES

Figura 1	– Modelo de neurônio artificial .....	20
Figura 2	– Identificação das partes de uma rede neural .....	21
Figura 3	– Arquitetura <i>feedforward</i> .....	22
Figura 4	– Modelo de rede <i>Perceptron</i> de Múltiplas Camadas .....	25
Figura 5	– Rede de função de base radial (RBF) .....	27
Figura 6	– Máquina de aprendizado extremo (ELM) .....	31
Figura 7	– Rede de estado de <i>echo</i> (ESN) .....	34
Figura 8	– Modelo de classificação de padrões com Redes neurais .....	36
Figura 9	– Modelo previsão série temporal com Rede neural .....	36
Figura 10	– Localização das usinas hidrelétricas no mapa brasileiro .....	39
Gráfico 1	– Série de vazões médias mensais da usina de Furnas .....	15
Gráfico 2	– Média e variância mensal para a série da usina de Furnas .....	16
Gráfico 3	– Série dessazonalizada da usina de Furnas .....	17
Gráfico 4	– Valores de autocorrelação para a série de Furnas dessazonalizada .....	18
Gráfico 5	– Série de vazões médias mensais da usina de Água Vermelha .....	41
Gráfico 6	– Série dessazonalizada da usina de Água Vermelha .....	41
Gráfico 7	– Resultado melhores previsões Água Vermelha de 2006 a 2015 – 1 passo a frente .....	43
Gráfico 8	– Resultado melhores previsões Água Vermelha de 2006 a 2015 – 3 passos a frente .....	43
Gráfico 9	– Resultado melhores previsões Água Vermelha de 2006 a 2015 – 6 passos a frente .....	44
Gráfico 10	– Resultado melhores previsões Água Vermelha de 2006 a 2015 – 12 passos a frente .....	44
Gráfico 11	– Série de vazões médias mensais da usina de Belo Monte .....	45
Gráfico 12	– Série dessazonalizada da usina de Belo Monte .....	45
Gráfico 13	– Resultado melhores previsões Belo Monte de 2006 a 2015 – 1 passo a frente .....	46
Gráfico 14	– Resultado melhores previsões Belo Monte de 2006 a 2015 – 3 passos a frente .....	47
Gráfico 15	– Resultado melhores previsões Belo Monte de 2006 a 2015 – 6 passos a frente .....	47
Gráfico 16	– Resultado melhores previsões Belo Monte de 2006 a 2015 – 12 passos a frente .....	48
Gráfico 17	– Série de vazões médias mensais da usina de Ilha Solteira .....	48
Gráfico 18	– Série dessazonalizada da usina de Ilha Solteira .....	49
Gráfico 19	– Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 1 passo a frente .....	50
Gráfico 20	– Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 3 passos a frente .....	50
Gráfico 21	– Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 6 passos a frente .....	51
Gráfico 22	– Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 12 passos a frente .....	51

Gráfico 23	– Série de vazões médias mensais do Complexo Hidrelétrico de Paulo Afonso	52
Gráfico 24	– Série dessazonalizada do Complexo Hidrelétrico de Paulo Afonso .....	52
Gráfico 25	– Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 1 passo a frente .....	53
Gráfico 26	– Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 3 passos a frente .....	54
Gráfico 27	– Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 6 passos a frente .....	54
Gráfico 28	– Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 12 passos a frente .....	55
Gráfico 29	– Série de vazões médias mensais da usina de Tucuruí .....	55
Gráfico 30	– Série dessazonalizada da usina de Tucuruí .....	56
Gráfico 31	– Resultado melhores previsões Tucuruí de 2006 a 2015 – 1 passo a frente ..	57
Gráfico 32	– Resultado melhores previsões Tucuruí de 2006 a 2015 – 3 passos a frente ..	57
Gráfico 33	– Resultado melhores previsões Tucuruí de 2006 a 2015 – 6 passos a frente ..	58
Gráfico 34	– Resultado melhores previsões Tucuruí de 2006 a 2015 – 12 passos a frente ..	58
Gráfico 35	– Vazões médias das usinas hidrelétricas .....	59

## **LISTA DE TABELAS**

Tabela 1	– Resultados previsões Água Vermelha de 2006 a 2015 .....	42
Tabela 2	– Resultados previsões Belo Monte de 2006 a 2015 .....	46
Tabela 3	– Resultados previsões Ilha Solteira de 2006 a 2015 .....	49
Tabela 4	– Resultados previsões Paulo Afonso de 2006 a 2015.....	53
Tabela 5	– Resultados previsões Tucuruí de 2006 a 2015 .....	56

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>8</b>
1.1 OBJETIVOS .....	9
1.1.1 Objetivo geral .....	9
1.1.2 Objetivos específicos .....	9
1.2 JUSTIFICATIVA .....	9
1.3 ORGANIZAÇÃO DO TRABALHO.....	10
<b>2 ANÁLISE DE SÉRIES TEMPORAIS .....</b>	<b>11</b>
2.1 SÉRIES TEMPORAIS: CONCEITUAÇÃO .....	11
2.2 PROCESSOS ESTOCÁSTICOS .....	12
2.3 FERRAMENTAS DE ANÁLISE DE SÉRIES TEMPORAIS .....	12
2.3.1 Média e variância.....	13
2.3.2 Autocovariância e autocorrelação .....	13
2.4 SÉRIES DE VAZÕES MÉDIAS MENSAIS.....	14
2.4.1 Padronização.....	16
2.5 ANÁLISE DA FUNÇÃO DE AUTOCORRELAÇÃO DESSAZONALIZADA .....	17
<b>3 REDES NEURAIS ARTIFICIAIS .....</b>	<b>19</b>
3.1 O NEURÔNIO ARTIFICIAL.....	19
3.2 ARQUITETURA DE REDES NEURAIS .....	20
3.2.1 Arquitetura <i>feedforward</i> .....	22
3.3 TREINAMENTO DA REDE .....	22
3.3.1 Treinamento supervisionado .....	23
3.3.2 Validação cruzada .....	24
3.4 <i>PERCEPTRON</i> DE MÚLTIPLAS CAMADAS (MLP) .....	24
3.4.1 Treinamento do MLP .....	25
3.5 REDES NEURAIS COM FUNÇÃO DE BASE RADIAL .....	26
3.5.1 Treinamento camada intermediária .....	27
3.5.2 Treinamento camada de saída.....	28
3.6 MÁQUINAS DESORGANIZADAS .....	28
3.6.1 Máquinas de aprendizado extremo (ELM) .....	30
3.6.1.1 Treinamento (ELM) .....	31
3.6.1.2 Execução da ELM .....	32
3.6.2 Redes neurais com estados de eco (ESN) .....	33
3.7 REDES NEURAIS PARA PREVISÃO DE SÉRIES TEMPORAIS.....	35
3.8 MÉTODO WRAPPER PARA SELEÇÃO DE ENTRADAS .....	37
<b>4 ESTUDO DE CASOS (PREVISÃO DE SÉRIES DE VAZÕES) .....</b>	<b>39</b>
4.1 USINA HIDRELÉTRICA DE ÁGUA VERMELHA.....	40
4.2 USINA HIDRELÉTRICA DE BELO MONTE.....	45
4.3 USINA HIDRELÉTRICA DE ILHA SOLTEIRA .....	48
4.4 COMPLEXO HIDRELÉTRICO DE PAULO AFONSO .....	52
4.5 USINA HIDRELÉTRICA DE TUCURUÍ .....	55
4.6 COMPARAÇÃO DOS RESULTADOS .....	58
<b>5 CONCLUSÃO .....</b>	<b>60</b>
5.1 CRONOGRAMA PROPOSTO .....	60
<b>REFERÊNCIAS.....</b>	<b>64</b>

## 1 INTRODUÇÃO

No ano de 2016<sup>1</sup>, 68,1% de toda energia elétrica produzida no Brasil foi resultado da geração hidráulica. Se comparado com o ano anterior, esse número teve um aumento de 5,9%, contribuindo ainda mais para a consolidação do cenário da matriz energética brasileira que continua sendo predominantemente composta por usinas hidrelétricas (EPE, 2017).

Em 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 de energia elétrica com menor custo 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, 2017), 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 (RASMUSSEN *et al.*, 1996); (BOX; JENKINS, 1976) e não-lineares (FRANCELIN; BALLINI; ANDRADE, 1996); (MASON; PRICE; TEM'ME, 1996); (ZEALAND; BURN; SIMONOVIC, 1999)), já foram propostas para o problema. Nesta última categoria encontram-se as Redes Neurais Artificiais.

Redes Neurais Artificiais (RNA) 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 baseada no conhecimento adquirido (HAYKIN, 2001). Devido a isso, as RNA tem sido muito utilizadas na previsão de séries temporais, inclusive na previsão de séries de vazões médias mensais a usinas hidrelétricas (SIQUEIRA, 2009).

O modelo linear Autorregressivo (AR), da família Box & Jenkins é o utilizado na previsão de séries de vazões hidrológicas afluentes utilizadas no planejamento da operação energética no Brasil pelo Sistema *Newave* (CEPEL, 2013). Entretanto, metodologias que fazem uso de RNA tem sido estudadas como alternativas viáveis para esta tarefa, tendo estas apresentado bons resultados como em Tang e Fishwick (1993) e Srinivasan, Liew e Chang (1994).

A partir dessa premissa, tem havido uma expansão na quantidade de estudos na área. Várias arquiteturas de RNAs já foram utilizadas, como o *Perceptron* de múltiplas camadas (MLP - do inglês *multilayer perceptron*) (FRANCELIN; BALLINI; ANDRADE, 1996), a Rede de função base radial (RBF) (MASON; PRICE; TEM'ME, 1996), Rede neuro-fuzzy (BALLINI,

---

<sup>1</sup> Foram utilizados os dados do ano de 2016 pois os dados do ano de 2017 ainda não foram divulgados pelo Ministério de Minas e Energia.

2000) e Rede neural com estados de eco (SACCHI *et al.*, 2007) e máquinas de aprendizado extremo (ELM) (SIQUEIRA, 2013).

## 1.1 OBJETIVOS

As Seções 1.1.1 e 1.1.2 descrevem respectivamente, qual o objetivo desse trabalho e quais etapas devem ser cumpridas para alcançar tal objetivo.

### 1.1.1 Objetivo geral

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

### 1.1.2 Objetivos específicos

Para que o objetivo central desse trabalho possa ser alcançado os seguintes objetivos específicos devem ser satisfeitos:

- Estudo dos modelos utilizados na previsão de séries de vazões;
- Estudo aprofundado de arquiteturas de Redes Neurais Artificiais (RNA) e de suas utilizações;
- Estudos de métodos de pré-processamento das séries;
- Realização de testes com as técnicas estudadas;
- Análise comparativa das redes investigadas.

## 1.2 JUSTIFICATIVA

Como mencionado, a matriz energética brasileira é composta predominantemente por usinas hidrelétricas, de modo que a operação eficiente dessas usinas tem impacto direto no custo da produção de 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 (RNA) vem se mostrando uma alternativa viável aos modelos lineares tradicionais. Portanto, o aperfeiçoamento da previsão de séries de vazões utilizando RNA tem impacto direto na diminuição do custo da geração de energia elétrica brasileira.

### 1.3 ORGANIZAÇÃO DO TRABALHO

No Capítulo 2 são apresentados os conceitos básicos sobre a análise de séries temporais, desde a definição de uma série temporal (Seção 2.1) e a definição de um processo estocástico (Seção 2.2) até a análise da função de autocorrelação de uma série de vazão (Seção 2.5).

Uma apresentação detalhada sobre Redes Neurais Artificiais (RNA) são mostradas no Capítulo 3. Nele também são descritos algumas arquiteturas: Perceptron de Múltiplas Camadas na Seção 3.4, Redes de Função de Base Radial na Seção 3.5, Máquinas de aprendizado extremo na Seção 3.6.1 e Redes Neurais em estado de *Echo* na Seção 3.6.2.

Por fim, o Capítulo 4 apresenta os resultados das previsões realizadas para a série de Água Vermelha na Seção 4.1, Belo Monte na Seção 4.2, Ilha Solteira na Seção 4.3, Paulo Afonso na Seção 4.4 e Tucuruí na Seção 4.5. Também no Capítulo 4 é realizada uma comparação dos resultados obtidos na Seção 4.6.

## 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 e conceitos básicos 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 das observações tem extrema importância e não pode ser alterada. As observações 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  $x_t$ ,  $t \in T \subset \mathbb{R}$ , de uma variável aleatória, geradas sequencialmente no tempo, sendo  $\mathbb{R}$  o conjunto dos números reais e  $T$  um conjunto 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  $\mathbf{X} = \{x_t, t \in T\}$ , tal que, para cada  $t \in T$ ,  $x_t$  é uma variável aleatória (v. a.). Assumindo que  $\mathbf{X} \in \mathbb{Z}_+$ , sendo este o conjunto dos números inteiros positivos, as observações que caracterizam uma série temporal podem ser descritas por uma v. a.  $\{x_t\}, t \in \mathbf{X}$ , com função de probabilidade conjunta  $p(z_1, z_2, \dots, x_N)$  (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  $\mathbf{X} = \{x_t, t \in T\}$  é dito estritamente estacionário (ou fortemente estacionário) se todas as distribuições de probabilidade dos elementos de  $\mathbf{X}$  permanecem as mesmas sob translações 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  $t_1, t_2, \dots, t_n, \tau \in T$  (MORETTIN; TOLOI, 2006).

Em outras palavras se 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  $\mathbf{X} = \{x_t, t \in T\}$  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  $t \in T$ ;
- A variância é constante:  $Var[x_t] = E[(x_t - \mu)^2] = \sigma_t^2 = \sigma^2$ , para todo  $t \in T$ ;
- 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  $x_t$ . Entretanto, na prática isso é muito difícil uma vez que o que se tem é um número limitado de amostras, o que leva a 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 alguma ferramentas estatísticas, as quais auxiliam na identificação e modelagem do processo. Nessa seção serão abordadas as seguintes ferramentas estatísticas: 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  $t \in T$ , 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.

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

onde  $E[\cdot]$  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.

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

Analizando as Equações 2.1 e 2.2 verifica-se que as integrais são definidas para uma distribuição de probabilidade constante e 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

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

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

onde  $N$  é 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 as observações são separadas por um intervalo constante  $k$ , elas são denotadas de  $x_t$  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 tem 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), na qual o termo  $\gamma_k$  é calculado como uma variável  $c_k$  de acordo com a Equação 2.8.

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

onde  $N$  é o número total de amostras disponíveis e  $\hat{\mu}$  é a média estimada dada pela Equação 2.3.

De forma similar, fazendo uso das Equações 2.7 e 2.8 o coeficiente de autocorrelação amostral  $r_k$  é dado pela Equação

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

O valor de  $r_k$  deve ser calculado 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  $\frac{N}{4}$ .

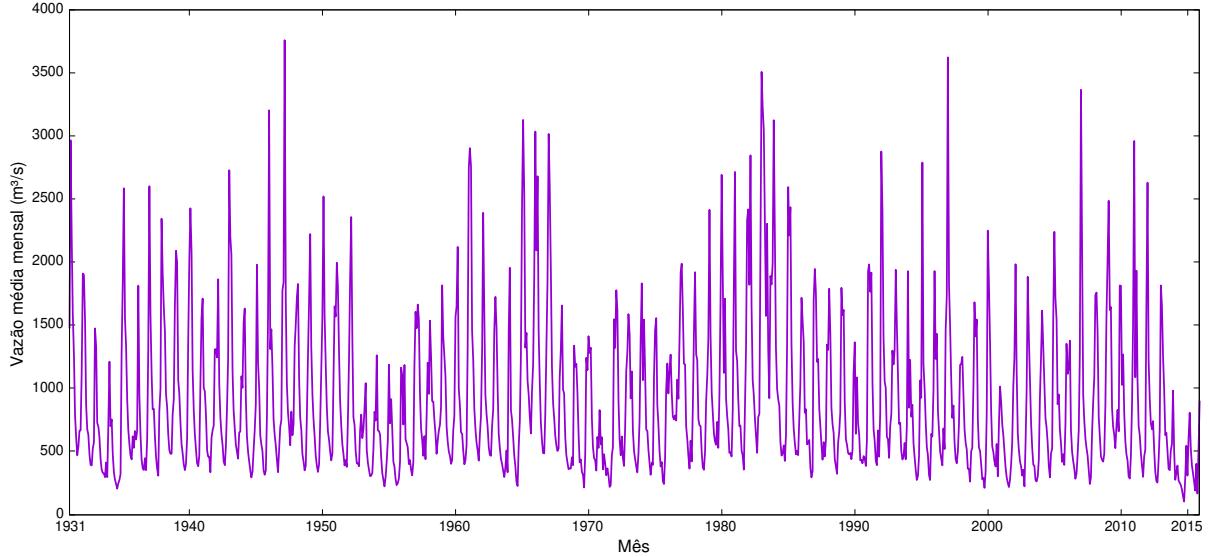
## 2.4 SÉRIES DE VAZÕES MÉDIAS MENSAIS

Séries de vazões são um tipo de série temporal, na qual cada observação diz respeito a vazão média mensal, semanal, diária ou horária. Utilizaremos 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 de série de vazão média mensal e a série histórica da usina hidrelétrica de Furnas, disponibilizada pelo Operador Nacional do Sistema Elétrico (ONS), sendo uma das mais utilizadas em estudos brasileiros de previsões de séries de vazões. O Gráfico 1 apresenta

todas as 1020 observações contidas na série, para a qual calculasse  $\hat{\mu} = 912,1225 \text{ m}^3/\text{s}$  e  $\hat{\sigma} = 613,5036 \text{ m}^3/\text{s}^1$ .

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



**Fonte:** Autoria própria.

Segundo Morettin e Toloi (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. 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.

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

onde  $s_t$  é a componente sazonal da observação  $t$ ,  $\tilde{x}_t$  é o valor estacionário da observação  $t$  e  $a_t$  é a componente aleatória de média zero, variância constante e 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 ajudam na sua compreensão. Este entendimento pode ser realizado através da análise da média e variância de cada mês da série, que podem ser obtidas pelas equações 2.11 e 2.12, respectivamente

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

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

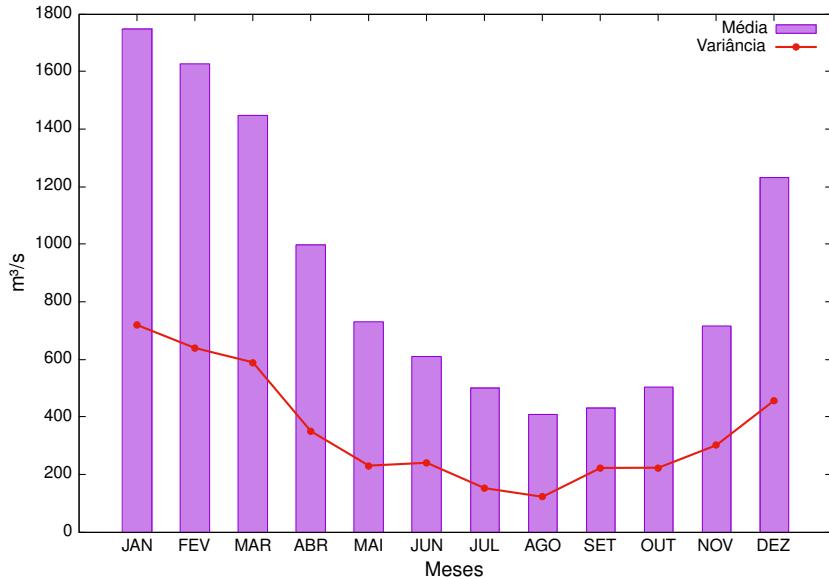
---

<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 usinas de Furnas.

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



**Fonte:** Autoria própria.

Como mostra o 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 Padronização

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 benéfico também para sua utilização com redes neurais (SIQUEIRA, 2009).

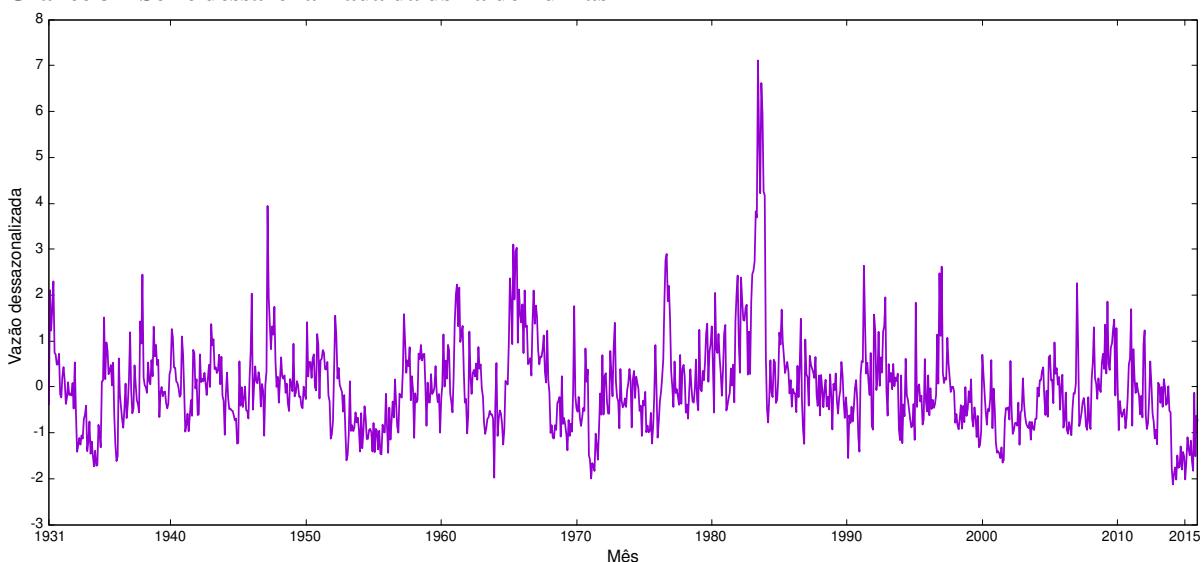
A vazão  $x_{i,m}$  dessazonalizada é denotada por  $x_{i,m}^{PA}$  e definida pela Equação 2.13.

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

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

O Gráfico 3 apresenta a série dessazonalizada da usina de Furnas.

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



**Fonte:** Autoria própria.

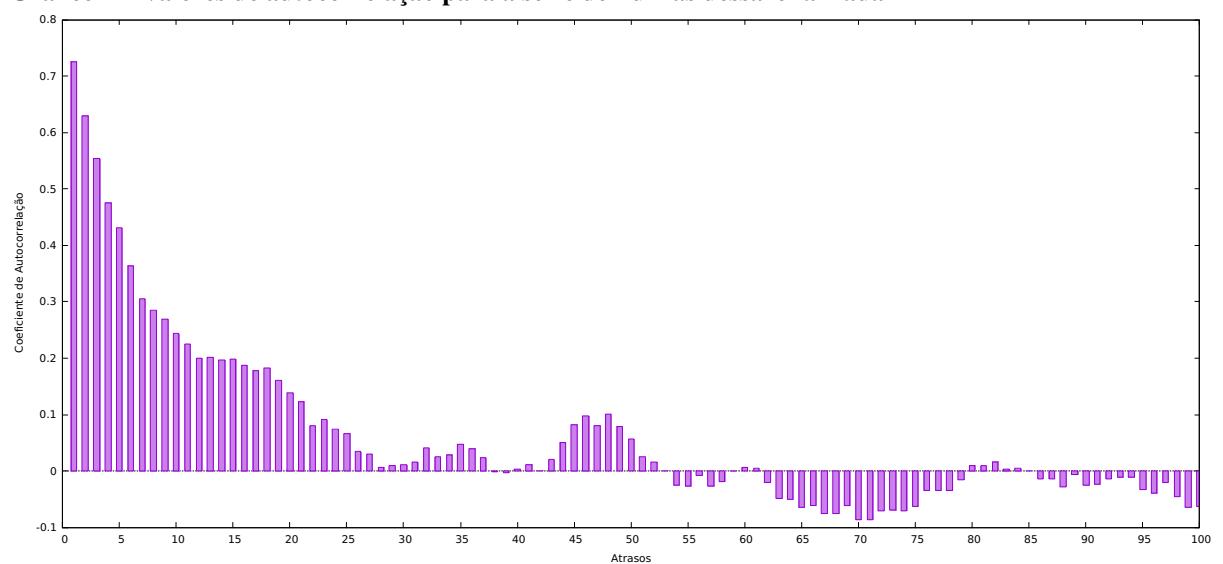
Note que a dessazonalização faz com que os novos valores da série estejam em um intervalo com amplitude menor do que a original. Entretanto, mesmo com a alteração dos valores, pode-se notar o semelhanças no padrão de comportamento nos dados. Na série original, logo depois da observação 400 nota-se um pico, seguido por uma grande queda e depois um pequeno aumento. Esse mesmo comportamento pode ser observado também na série padronizada.

## 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 e calculados os 100 primeiros valores de  $r_k$ , é possível traçar o Gráfico 4.

Nele é possível notar que o valor da autocorrelação diminui a média que  $k$  aumenta, como é esperado para dados com dependência temporal. Também nota-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 REDES NEURAIS 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 unidireccionais podendo ser dispostos em uma ou mais camadas (BRAGA; de 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).

#### 3.1 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).

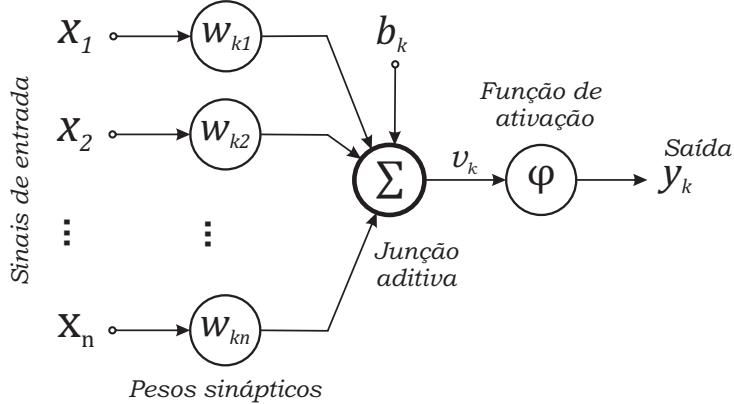
Neurônios artificiais são em geral não-lineares, sua saída é contínua e realizam funções simples, como coletar os valores de todas as suas entradas e misturá-los de acordo com sua função operacional e produzir uma resposta levando em consideração sua função de ativação (SILVA; SPATTI; FLAUZINO, 2010). 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; de Carvalho; LUDERMIR, 2007).

A Figura 1 mostra um dentre os diversos modelos de neurônio artificial, denominado neurônio  $k$ , que é largamente usado como base para projetos de Redes Neurais Artificiais. Nele 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.

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

**Figura 1 – Modelo de neurônio artificial**

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

Junção aditiva, podendo ser tanto valores positivos quanto negativos. O mais comum é utilizar  $b_k = 1$ , o qual também pode ter um peso associado.

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  $[-1, 1]$ .

Pode-se então definir matematicamente o neurônio artificial  $k$  pela Equação 3.1.

$$y_k = \varphi \left( \left[ \sum_{i=1}^n x_i w_{ki} \right] + b_k \right) \quad (3.1)$$

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  $k$ .

### 3.2 ARQUITETURA DE REDES NEURAIS

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

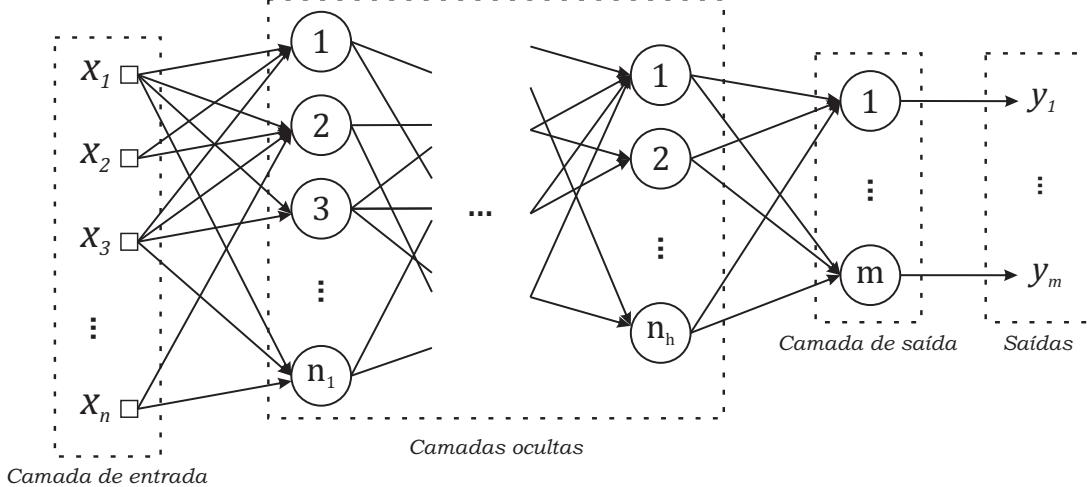
A topologia de uma rede, por sua vez, compreende as diferentes formas de composição estrutural que a mesma pode assumir dentro de uma arquitetura. Como exemplo pode-se ter duas topologias dentro de uma mesma arquitetura, sendo a primeira topologia composta por 20 neurônios enquanto a segunda faz uso de 10. A topologia de uma rede neural pode ainda fazer referência a função de ativação de cada neurônio da rede (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):

- **Camada de entrada** - É a camada responsável por receber os sinais de entrada da rede,

$\mathbf{x} = \{x_1, x_2, x_3, \dots, x_n\}$ . Tais sinais normalmente são normalizados para que seus valores compreendam o intervalo de abrangência da função de ativação. As operações matemáticas realizadas pela rede apresentam uma maior precisão quando os valores dos sinais de entrada são normalizados. É importante ressaltar que a camada de entrada não possui neurônios. Na Figura 2 é apresentada uma rede neural genérica onde é possível identificar a camada de entrada.

**Figura 2 – Identificação das partes de uma rede neural**



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

- **Camadas escondidas, intermediárias, ocultas ou invisíveis** - São as camadas que possuem neurônios e que não representam a saída da rede. Uma rede neural pode possuir diversas ou nenhuma camada oculta. A maior parte do processamento interno realizado pela rede neural é feito pelas camadas ocultas, sendo estas as responsáveis pela extração das características associadas ao processo ou sistema a ser inferido. Na Figura 2 é possível identificar quais são as camadas ocultas.
- **Camada de saída** - Responsável pela produção e apresentação dos resultados finais da rede, os quais são gerados mediante os resultados das camadas anteriores. A camada de saída é a última camada de uma rede neural, também sendo composta por neurônios. A Figura 2 realça a camada de saída de uma rede neural. Analisando a Figura 2 verifica-se que a quantidade de neurônios presentes na camada de saída corresponde à quantidade de saídas da mesma.

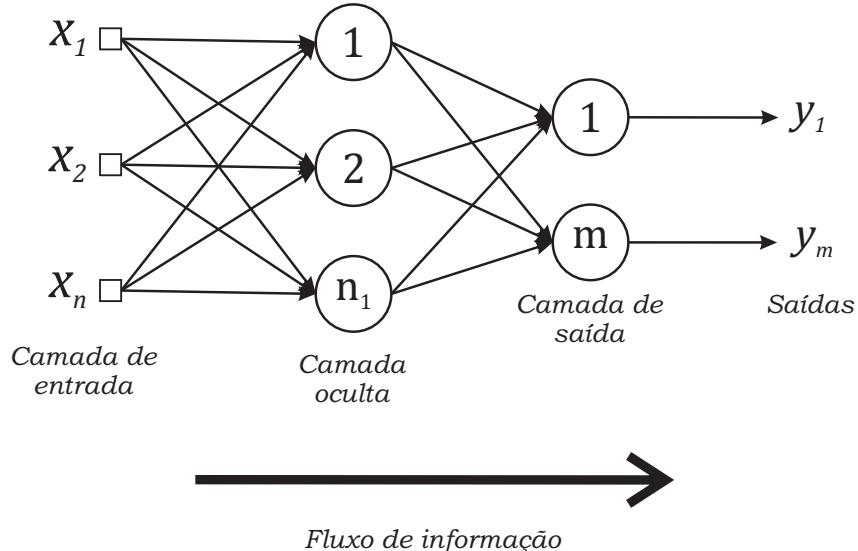
Note que apenas as camadas de entrada e de saída são obrigatórias em uma rede neural. Como mencionado, uma rede neural pode não ter camadas ocultas ou ter qualquer quantidade de camadas ocultas, mas apenas uma camada de entrada e uma camada de saída. Ainda, a quantidade de neurônios e as funções de ativação de cada camada oculta podem ser diferentes entre si.

### 3.2.1 Arquitetura *feedforward*

A arquitetura *feedforward* (não-recorrente) comprehende o sentido em que o fluxo de informações segue pela rede. Nela o fluxo de informação segue sempre em uma única direção, tendo origem na camada de entrada e seguindo em direção a camada de saída.

A Figura 3 apresenta uma rede neural da arquitetura *feedforward* mostrando o sentido do fluxo de informação da rede.

**Figura 3 – Arquitetura *feedforward***



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

Na Figura 3 é possível identificar o conjunto de entradas da rede composto por  $X = \{x_1, x_2, \dots, x_n\}$ , 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, entretanto a arquitetura *feedforward* independe da quantidade de camadas da rede ou da quantidade de neurônios em cada camada. Note ainda que na Figura 3 as ligações entre os neurônios são realizadas por setas direcionadas, de forma que o sentido da seta representa o sentido do fluxo de informação da rede.

## 3.3 TREINAMENTO DA REDE

A principal razão que levou as redes neurais 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 deste, a mesma é capaz de generalizar esse conhecimento, fornecendo respostas para amostras que não foram utilizadas durante o aprendizado.

Silva, Spatti e Flauzino (2010) definem o treinamento de uma rede neural como a apli-

cação 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 da rede neural são chamadas de conjunto de treinamento, o qual é composto com cerca de 60 a 90% das amostras totais e utilizado essencialmente no ajuste dos pesos sinaíticos.

Por sua vez o conjunto de teste é composto pelas amostras restantes com o objetivo de verificar se a generalização das soluções obtidas pela rede neural está com um nível de erro aceitável para a aplicação em questão (SILVA; SPATTI; FLAUZINO, 2010). Após a execução do algoritmo de aprendizagem (treinamento) a rede neural está pronta para ser utilizada na aplicação real.

O processamento utilizado por uma rede neural é quase que totalmente aplicado durante o treinamento da rede, visto que a execução dela é realizada com custo computacional linear.

### 3.3.1 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), dessa forma cada amostra do conjunto de treinamento é composta pelos sinais de entradas e as respectivas saídas esperadas.

Durante o treinamento supervisionado, a cada sinal de entrada inserido na rede é calculada qual 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 as saídas da rede e as saídas esperadas do conjunto de treinamento atinge um nível aceitável, previamente definido pelo projetor.

Note que, na prática, o treinamento da rede 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 para as amostras do subconjunto de treinamento. Dessa forma, diz-se que a rede perdeu o seu poder de generalização.

Para evitar tal comportamento indesejado, existem técnicas que podem ser aplicadas para contrabalançar esse efeito. Outro fator que pode levar a 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. Neste caso, a rede não apresentará resultados satisfatórios quando a mesma for aplicada na prática.

### 3.3.2 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, o conjunto de teste deverá apresentar alto grau de erro, perdendo sua capacidade de generalização.

Quando uma rede neural perde o seu poder de generalização dizemos que a mesma está sobre-treinada. Para evitar tal fenômeno, pode-se aplicar o processo de validação cruzada durante o seu treinamento.

A validação cruzada consiste em, ao final de cada época de treinamento, submetê-la a um conjunto de dados desconhecido para validar o seu poder de generalização. Assim, o conjunto de teste que foi mencionado na Seção 3.3 é dividido em dois subconjuntos, um o validação e o de teste propriamente dito.

Ao final do treinamento o conjunto de pesos sinápticos escolhido é aquele que apresentar o melhor desempenho para o conjunto de validação, ou seja, aquele em que a rede apresentar o menor erro. Dessa forma tem-se a certeza de escolher uma configuração de pesos para a rede que tenha excelente poder de generalização.

Ainda com a finalidade de comprovar o poder de generalização da rede, ao final do treinamento, ou seja, após ter selecionado a configuração de pesos que apresenta o melhor desempenho para o conjunto de validação, a mesma é exposta ao conjunto de teste. Dessa forma seu desempenho é verificado junto a um conjunto de dados inédito.

## 3.4 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 nome explica, recebe os sinais de entrada, a camada de saída entrega o resultado obtido pela rede e entre elas dessas podem existir quantas camadas ocultas forem necessárias (HAYKIN, 2001).

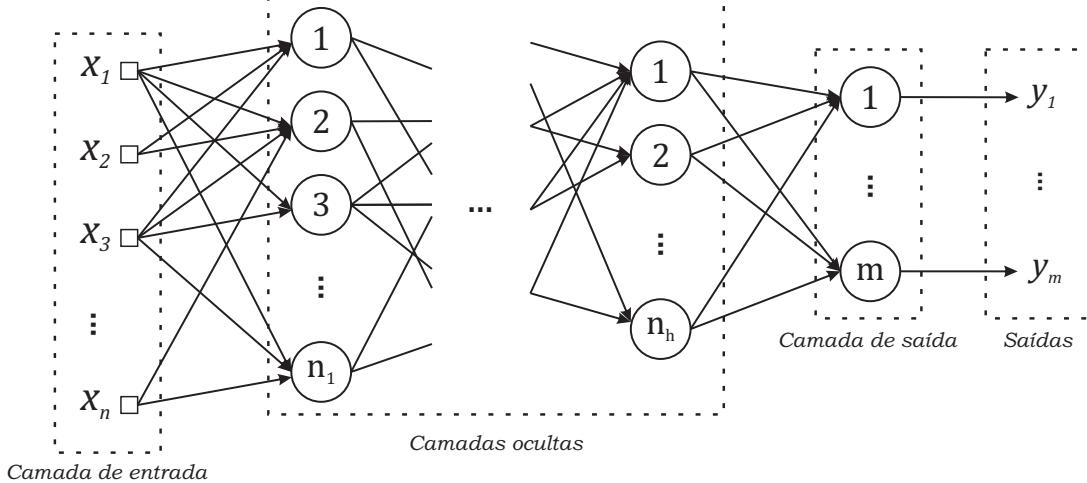
As MLPs são consideradas uma das arquiteturas mais versáteis quanto a aplicabilidade, sendo utilizadas em diversas áreas do conhecimento como em aproximação universal de funções, reconhecimento de padrões, identificação e controle de processos, previsão de séries temporais e otimização de sistemas (SILVA; SPATTI; FLAUZINO, 2010).

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.

A Figura 4 mostra um modelo genérico de MLP. Nele é possível identificar a camada de entradas da rede, as camadas ocultas e por fim a camada 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 é dada pelo número de saídas necessárias para compor uma instância do problema (SILVA; SPATTI; FLAUZINO, 2010).

**Figura 4 – Modelo de rede Perceptron de Múltiplas Camadas**



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

Ainda analisando o modelo genérico de MLP da Figura 4 nota-se que os sinais de entrada fornecidos pela camada de entrada são recebidos pela primeira camada oculta para que sejam processados. O resultado da saída desta camada 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 camada de saída, a qual devolve o resultado obtido pela rede. Assim, não importa a quantidade de camadas ocultas em uma MLP, a propagação dos sinais de entrada ocorre sempre em um único sentido, da camada de entrada em direção à camada de saída (SILVA; SPATTI; FLAUZINO, 2010).

### 3.4.1 Treinamento do MLP

Como estabelecido anteriormente, processo de treinamento consiste no ajuste dos pesos sinápticos com o objetivo de encontrar o conjunto deles que atinge o melhor mapeamento a respeito do evento desejado. O algoritmo mais conhecido de treinamento das MLPs é o gradiente ótimo, cujo vetor gradiente é calculado por meio do método de retropropagação ou *backpropagation* criado por Rumelhart, Hinton e Williams (1986). Este processo é do tipo supervisionado.

O algoritmo consiste em duas fases. Inicialmente, os dados de entrada são propagados

pela rede a fim de obter as 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 3.2.

$$e_j(t) = d_j(t) - y_j(t) \quad (3.2)$$

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

A cada iteração deseja-se minimizar o erro de saída. A métrica mais usual utilizada é o erro quadrático médio instantâneo (do inglês *mean square error - MSE*), dado pela Equação 3.3.

$$MSE(t) = \frac{1}{N} \sum_{j=1}^N (d_j(t) - y_j(t))^2 \quad (3.3)$$

na qual  $N$  é o número de neurônios da rede.

Por fim as regras para atualização dos pesos sinápticos e dos bias de cada neurônio são dadas pelas Equações 3.4 e 3.5, respectivamente

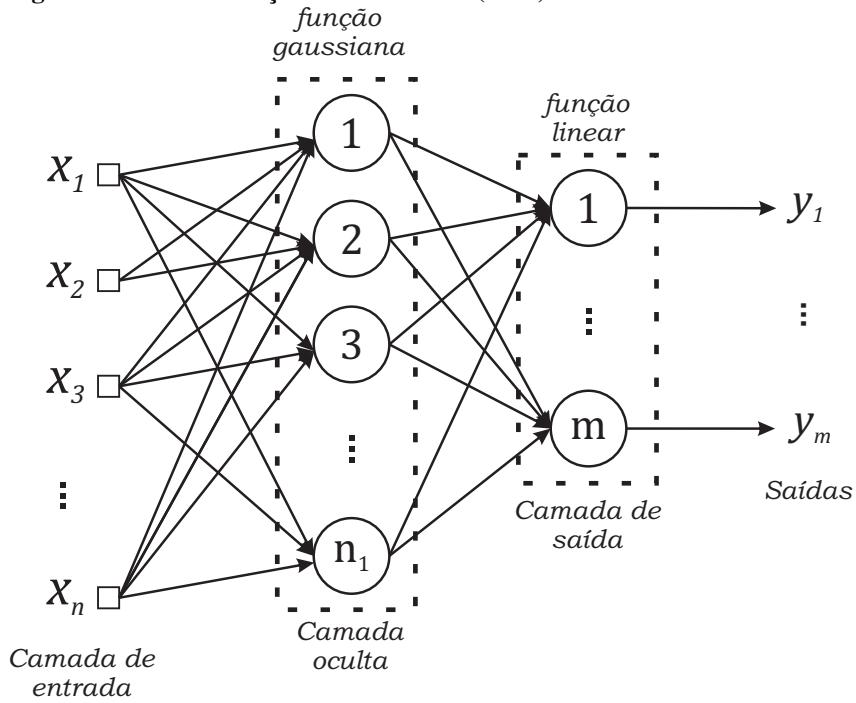
$$w_{ij}^m(t+1) = w_{ij}^m(t) - \alpha \frac{\partial MSE(t)}{\partial w_{ij}^m(t)} \quad (3.4)$$

$$b_j^m(t+1) = b_j^m(t) - \alpha \frac{\partial MSE(t)}{\partial b_j^m(t)} \quad (3.5)$$

onde  $w_{ij}^m(t)$  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  $t$ .

### 3.5 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 multiplas. 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 5 apresenta uma configuração genérica de uma RBF. Nela, 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 camada de saída possuem esta componente.

**Figura 5 – Rede de função de base radial (RBF)**

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

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

### 3.5.1 Treinamento camada intermediária

Como dito, a função de ativação dos neurônios da camada intermediária de uma RBF são de base radial, sendo a gaussiana uma das mais empregadas. A Equação 3.6 define uma função gaussiana

$$\varphi(u) = e^{-\frac{(u-c)^2}{2\sigma^2}} \quad (3.6)$$

na qual  $c$  define o centro da gaussiana e  $\sigma^2$  a sua variância em função do centro, de modo que  $c - \sigma^2$  e  $c + \sigma^2$  formam as extremidades da função. Analogamente, nota-se que quanto maior for a variância maior será a extensão da base da função.

Na prática, cada neurônio da camada intermediária acaba por ser uma função gaussiana, em que seu centro é definido pelos pesos de suas entradas. Assim, o processo de treinamento necessita ajustar o valor da variância da base de cada função, ou seja, de cada neurônio da camada intermediária.

Inicialmente são atribuídos os valores das entradas aos pesos dos neurônios da camada intermediária, de modo que os pesos do neurônios  $n_1$  são os valores da entrada  $x_1$ , os pesos do

neurônios  $n_2$  são os pesos da entrada  $x_2$  e assim por diante.

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  $x_i$ .

Uma vez que todas as entradas foram inseridas em  $\Omega^n$  os centros das gaussianos dos neurônios são atualizados para corresponder ao centro de todas as entradas contidas no seu grupo  $\Omega^n$ . Após essa atualização dos centros das funções de  $\Omega^n$  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 3.7.

$$\sigma_j^2 = \frac{1}{|\Omega^j|} \sum_{x^k \in \Omega^j} \sum_{i=1}^n (x_i^k - w_{ji}^1)^2 \quad (3.7)$$

na qual  $\sigma_j^2$  é a variância da base da função gaussiana da 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  $i$ .

### 3.5.2 Treinamento camada de saída

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 na MLP, pelas equações 3.4 e 3.5.

## 3.6 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 definiu essas máquinas como mecanismos aleatórios em sua construção, mas capazes de realizar tarefas particulares mediante treinamento.

Como pode ser visto em Boccato (2013), 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 (BOCCATO, 2013):

- **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 o estado 0 em cada instante de tempo  $t$ , de forma que a saída de cada neurônio (que é o seu estado) é computada pela função booleana *NAND* (que é a negação da função lógica *AND*). 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 neurônios, 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 3.3). Nessa rede são adicionadas 2 entradas em cada neurônio de interferência, de forma que os valores dessas entradas podem ser alterados por um agente externo, ou seja, um agente externo 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.

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 3.6.1 e as redes neurais com estado de eco na Seção 3.6.2.

### 3.6.1 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. Neste caso, apenas os pesos da camada de saída são atualizados. Esta aleatoriedade que dá a 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 aliado ao fato de que os neurônios ocultos não se ajustam, faz com que o custo computacional do treinamento seja extremamente baixo.

No trabalho de Bartlett (1998), o autor apresenta um resultado de extrema importância para a área de redes neurais. 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.

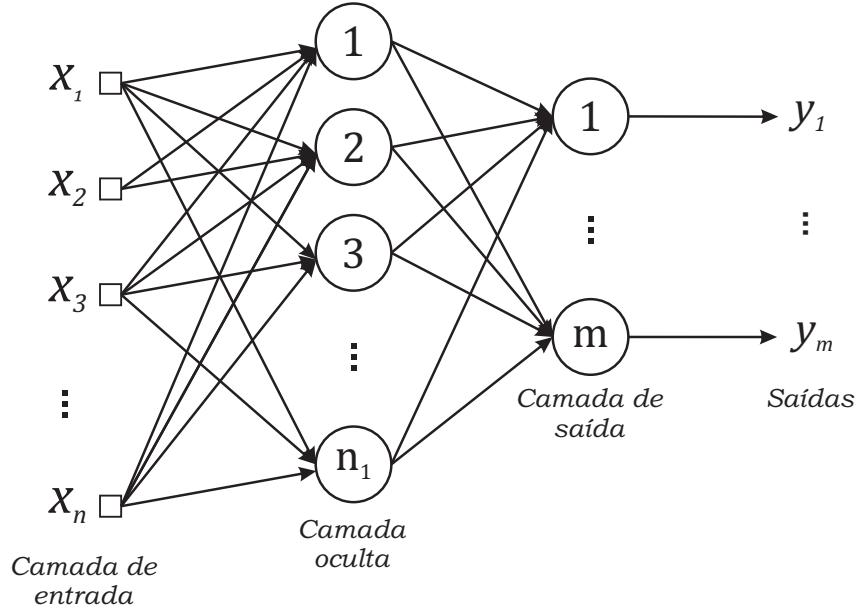
Por sua vez, nas ELMs a garantia da boa capacidade de generalização é dada pelos pesos da camada de saída, de modo que os pesos da camada intermediária podem então ser definidos de forma aleatória.

Como usual, a ELMs aceitam quantas saídas e 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 aprendizado para esta rede. Entretanto, nos dias atuais as duas são vistas como propostas distintas de redes neurais. Uma ELM é apresentada na Figura 6.

Na Figura 6 é possível ver o vetor de entradas denotado por  $\mathbf{x} = [x_1, x_2, x_3, \dots, x_n]$ , os neurônios da camada oculta  $1, 2, 3, \dots, n_1$ , os neurônios da camada de saída  $1, 2, \dots, m$  e por fim as saídas da rede  $y_1, y_2, y_3, \dots, y_m$ . Apesar de não constar na Figura 6, cada neurônio da camada intermediária da ELM possui um valor de bias associado a ele, sendo que a camada de saída não possui bias.

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 com a utilização de funções desde funções trigonométricas até a função sinal.

**Figura 6 – Máquina de aprendizado extremo (ELM)**



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

### 3.6.1.1 Treinamento (ELM)

O treinamento da ELM é realizado apenas nos pesos sinápticos da camada de saída. Assim tanto os pesos sinápticos quanto os valores dos bias para a camada intermediária são definidos de forma aleatória.

No trabalho de Huang, Zhu e Siew (2004), 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 a 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 3.8

$$\mathbf{w}_k^* = \arg_{\mathbf{w}_k \in \Re^{n_1+1}} \min \| \mathbf{w}_k \|^2 + C_k \times J(\mathbf{w}_k) \quad (3.8)$$

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 3.9.

$$J(\mathbf{w}_k) = \frac{1}{2} \sum_{l=1}^n \left( \sum_{j=1}^{n_1} w_{kj} f(\mathbf{v}_j, b_j, \mathbf{x}_j) + w_{k0} - S_{kl} \right) \quad (3.9)$$

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

Considere  $\mathbf{W}^{\text{out}}$  como a matriz com os pesos sinápticos de todos os neurônios da ca-

mada de saída, construída da seguinte forma:

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

na qual  $w_{ki}^{\text{out}}$  representa o peso sináptico da entrada  $i$  do neurônio  $k$  da camada de saída,  $\mathbf{d}$  e o vetor contendo todas as saídas esperadas.

Assim sendo, treinar a ELM consiste em determinar qual a matriz  $\mathbf{W}^{\text{out}}$  com o menor erro para o vetor  $\mathbf{d}$ . Dessa forma, o processo de treinamento da ELM consiste em resolver o operador pseudo-inversa de *Moore-Penrose* conforme a Equação 3.10

$$\mathbf{W}^{\text{out}} = (\mathbf{X}_{\text{hid}}^T \mathbf{X}_{\text{hid}})^{-1} \mathbf{X}_{\text{hid}}^T \mathbf{d} \quad (3.10)$$

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

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

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. O desempenho da rede pode ser melhorado com a utilização de um coeficiente de regularização  $C$  proposto por Huang *et al.* (2012). Adicionando esse coeficiente obtém-se a Equação 3.11

$$\mathbf{W}^{\text{out}} = \left( \frac{1}{C} + \mathbf{X}_{\text{hid}}^T \mathbf{X}_{\text{hid}} \right)^{-1} \mathbf{X}_{\text{hid}}^T \mathbf{d} \quad (3.11)$$

Huang *et al.* (2012) definiram o coeficiente de regularização como sendo  $C = 2^\lambda$  e  $\lambda$  é o vetor  $\lambda = \{-25, -24, \dots, 25, 26\}$ . 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.

### 3.6.1.2 Execução da ELM

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 para a camada de saída, similar ao funcionamento das MLPs.

O sinal de saída da camada oculta da ELM é dado pela Equação 3.12

$$y_k^{hid} = \varphi_k^{hid} \left( \left[ \sum_{i=1}^n x_i w_{ki}^{hid} \right] + b_k^{hid} \right) \quad (3.12)$$

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.

De forma similar, o sinal de saída da camada de saída da ELM é dado pela Equação 3.13.

$$y_k^{out} = \varphi_k^{out} \left( \sum_{i=1}^{n_1} y_i^{hid} w_{ki}^{out} \right) \quad (3.13)$$

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  $k$ .

### 3.6.2 Redes neurais com estados de eco (ESN)

Redes neurais recorrentes (RNN) 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* podem 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 aproximadores universais (SCHAEFER; ZIMMERMANN, 2007).

As ESNs possuem 3 (três) camadas: a primeira é camada de entrada, a segunda 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, é responsável por combinar as saídas do reservatório de dinâmicas. Esta última, por sua vez, corresponde a parte linear da

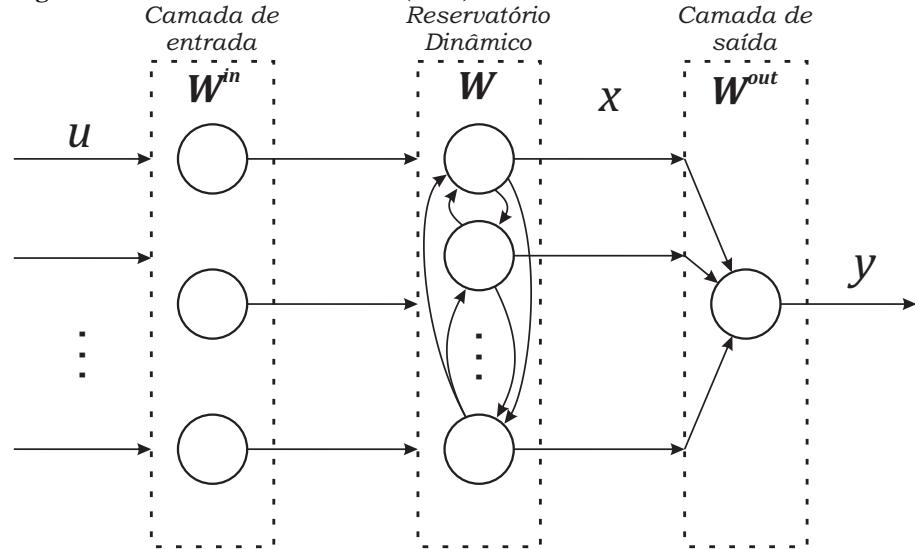
<sup>1</sup>  $y^{hid}$  – *hid* vem do inglês *hidden*.

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

rede. Diferente de outras propostas de RNNs, que podem possuir realimentação em qualquer camada, as ESNs propostas por Jaeger (2001) possuem realimentação apenas no reservatório de dinâmicas.

Na Figura 7 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.

**Figura 7 – Rede de estado de echo (ESN)**



Fonte: Adaptado de Siqueira (2013).

Nota-se pela Figura 7 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  $\mathbf{W}^{in} \in \Re^{N \times K}$  ponderam linearmente as entradas e as passam ao reservatório de unidade não-lineares totalmente interconectadas  $\mathbf{W} \in \Re^{N \times N}$ . 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. Os estados da rede são atualizados a cada nova entrada, sendo que os novos estados da rede para a entrada  $t + 1$  são dados pela Equação 3.14 (OZTURK; XU; PRÍNCIPE, 2007).

$$\mathbf{x}_{t+1} = f(\mathbf{W}^{in}\mathbf{u}_{t+1} + \mathbf{W}\mathbf{x}_t) \quad (3.14)$$

na qual  $\mathbf{x}_{t+1}$  são os estados da rede na entrada  $t + 1$ ,  $f(\cdot)$  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. Convenciona-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 3.15.

$$\mathbf{y}_{t+1} = \mathbf{W}^{out}\mathbf{x}_{t+1} \quad (3.15)$$

sendo  $\mathbf{W}^{out} \in \Re^{L \times N}$  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 dos pesos sinápticos da camada de saída a partir dos estados de eco  $\mathbf{X} \in \Re^{K \times N}$  conforme a Equação 3.16.

$$\mathbf{W}^{\text{out}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{d} \quad (3.16)$$

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

É importante ressaltar ainda que a matriz  $\mathbf{W}$  deve obedecer a propriedade de estado de eco. Jaeger (2001) comprovou a propriedade de estado de eco, o autor constatou 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, entretanto 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 validar a propriedade de estado de eco:

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  $A \subset \Re^N$ .

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

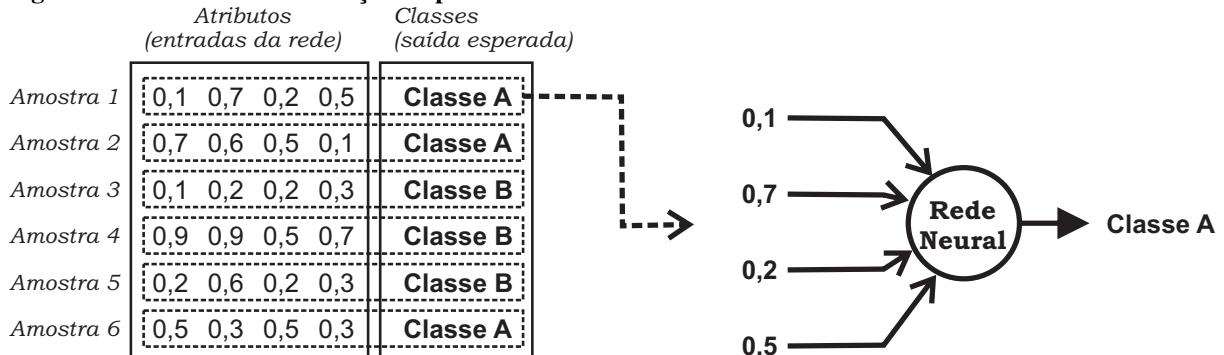
### 3.7 REDES NEURAIS PARA PREVISÃO DE SÉRIES TEMPORAIS

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 8 apresenta um modelo do funcionamento de uma rede neural como classificadora de padrões.

Na Figura 8 fica evidente que os atributos contidos na amostras são as entradas da rede, enquanto que a saída da rede é classe da amostra.

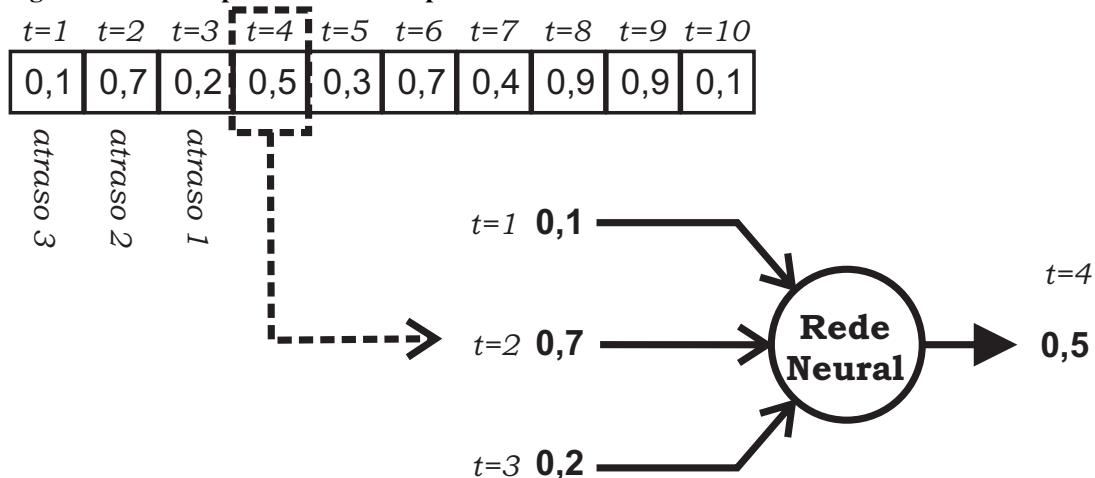
Por sua vez, em uma série temporal não existem atributos que se relacionam com o elemento da série. O que há são os próprios dados da série que se deseja prever. Neste contexto,

**Figura 8 – Modelo de classificação de padrões com Redes neurais**

Fonte: Autoria própria.

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

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  $x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-k}$  onde  $k$  é chamado de atraso. Em virtude dessa relação direta são os atrasos que são utilizados como entrada da rede neural para a previsão de séries temporais. A Figura 9 ilustra esta ideia.

**Figura 9 – Modelo previsão série temporal com Rede neural**

Fonte: Autoria própria.

Como pode ser visto na Figura 9 para prever o elemento  $t = 4$  foram utilizados 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 para um problema de previsão 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 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 9 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 deseja-se prever o próximo dado da série, a previsão é dita 1 passo a frente, se deseja-se prever o 4 elemento a frente a previsão é dita 4 passos a frente e assim sucessivamente.

Imagine 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 ação até a data de hoje. Entretanto, neste exemplo, interessa realizar uma previsão do preço da ação na 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  $n$  passos a frente. Inicialmente são utilizados os dados existentes para ajustar a rede para prever 1 passo a frente. Em seguida, o resultado da previsão 1 passos adiante é inserido na série como se fosse um valor real da série e não uma previsão. Com esse novo valor agora é possível prever 2 passos a frente. Esse procedimento é repetido até que a rede seja capaz de prever o horizonte necessário (SIQUEIRA, 2013).

### 3.8 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. Diferente do que se possa imaginar, inserir uma grande quantidade de atrasos nem sempre leva a menores erros. Neste sentido, o método *wrapper* testa de forma sistemática diversas combinações para a melhor configuração do conjunto de entradas (SIQUEIRA, 2013).

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).

Inicialmente é necessário que o projetista defina o número máximo de entradas. O procedimento é iniciado com um subconjunto de entradas (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 subconjunto 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 que durante o processo tiver levado ao menor erro geral será o

escolhido.

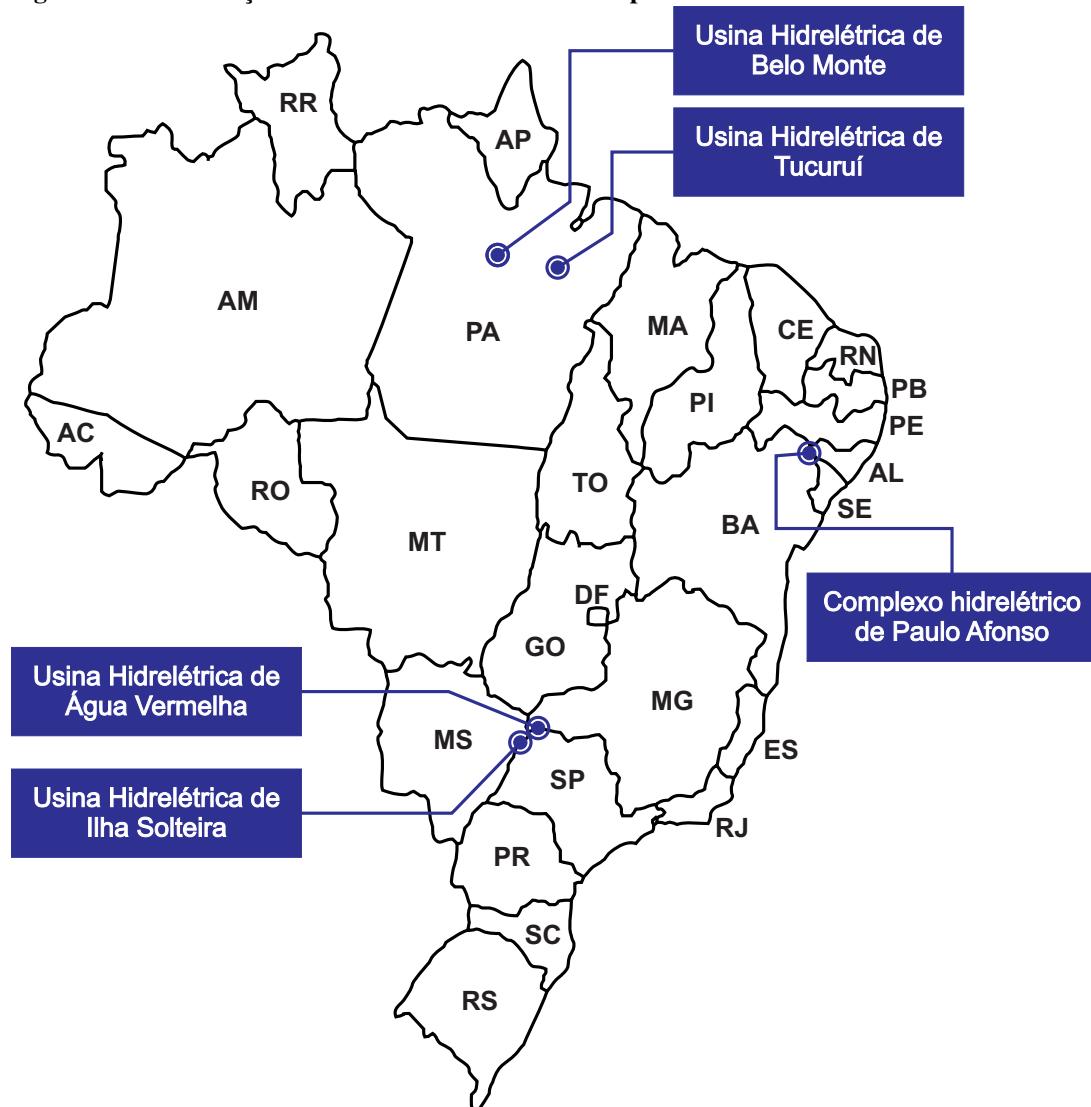
Vale lembrar que como o *wrapper* com seleção progressiva não verifica todas as possibilidade de configuração para o conjunto de entradas da rede ele não garante que a configuração escolhida seja a configuração ótima.

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

Com o objetivo de estimular a pesquisa a respeito da previsão de séries de vazões, o Operador Nacional do Sistema Elétrico (ONS) disponibiliza em seu site os dados históricos referentes a série de vazões médias mensais e diárias, juntamente com um relatório sobre a atualização de séries históricas de vazões (ONS, 2018).

As previsões realizadas nesse estudo correspondem as 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 possuem níveis de vazões afluentes diferentes, o que possibilita uma análise mais robusta dos desempenhos dos modelos. A Figura 10 mostra a localização de cada usina hidrelétrica no mapa brasileiro.

**Figura 10 – Localização das usinas hidrelétricas no mapa brasileiro**



**Fonte:** Autoria própria.

A série proveniente da usina da Água Vermelha é uma das séries com comportamento mais difícil de ser mapeado, o que tornou essa série uma espécie de prova de qualidade para os

modelos de previsão desenvolvidos.

Para as 5 usinas foram realizadas previsões utilizando uma rede MLP e uma RBF, visando assim determinar qual das duas arquiteturas é a mais indicada para cada planta.

Todas as séries possuem dados de Janeiro de 1931 a Dezembro de 2015, totalizando 85 anos. Os dados foram separados entre treinamento, validação e teste da seguinte forma:

- **Treinamento** – Dados entre os anos de 1931 e 1995, totalizando 75 anos.
- **Validação** – Dados entre os anos de 1996 e 2005, totalizando 10 anos.
- **Teste** – Dados entre os anos de 2006 e 2015, totalizando 10 anos.

No que diz respeito as MLPs treinadas, foi utilizada apenas 1 camada intermediária, na qual todos os neurônios contam com função de ativação do tipo tangente hiperbólica.

Para as duas redes foi utilizado o processo de validação cruzada no treinamento e a condição de parada do treinamento foi uma precisão de  $10^{-6}$  com um máximo de épocas de 2000. Além disso as duas redes foram testadas quanto a quantidade de neurônios, sendo executadas para cada número de neurônios contido no vetor  $V_N = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 40, 50, 60, 70, 80, 90, 100]$ .

Seguindo a metodologia usualmente empregada pelo Setor Elétrico Brasileiro (SEB), serão utilizados até 6 atrasos para a previsão. O método *wrapper* com seleção progressiva foi o escolhido para determinar a configuração de entrada de cada rede.

Como medida de desempenho foi utilizado o erro quadrático médio (MSE, do inglês *mean square error*) conforme a Equação 4.1

$$MSE = \frac{1}{N} \sum_{t=1}^N (d_t - y_t)^2 \quad (4.1)$$

Além do MSE, também é apresentado o Erro absoluto médio (MAE, do inglês *mean absolute error*) conforme a Equação 4.2.

$$MAE = \frac{1}{N} \sum_{t=1}^N |d_t - y_t| \quad (4.2)$$

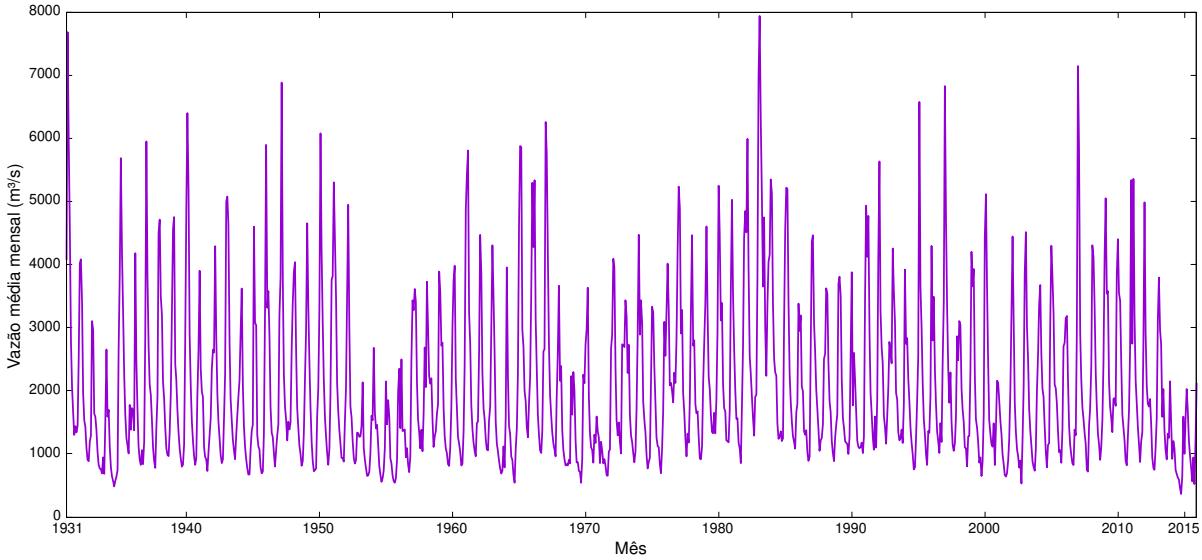
Foram realizadas previsões com 1, 3, 6 e 12 passos a frente.

#### 4.1 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  $\hat{\mu} = 2077,3284 \text{ m}^3/\text{s}$  e desvio padrão  $\hat{\sigma} = 1295,7097 \text{ m}^3/\text{s}$ , calculados de acordo com as equações 2.3 e 2.4, respectivamente. Todas as 1020 observações contidas na séries são apresentadas no Gráfico 5.

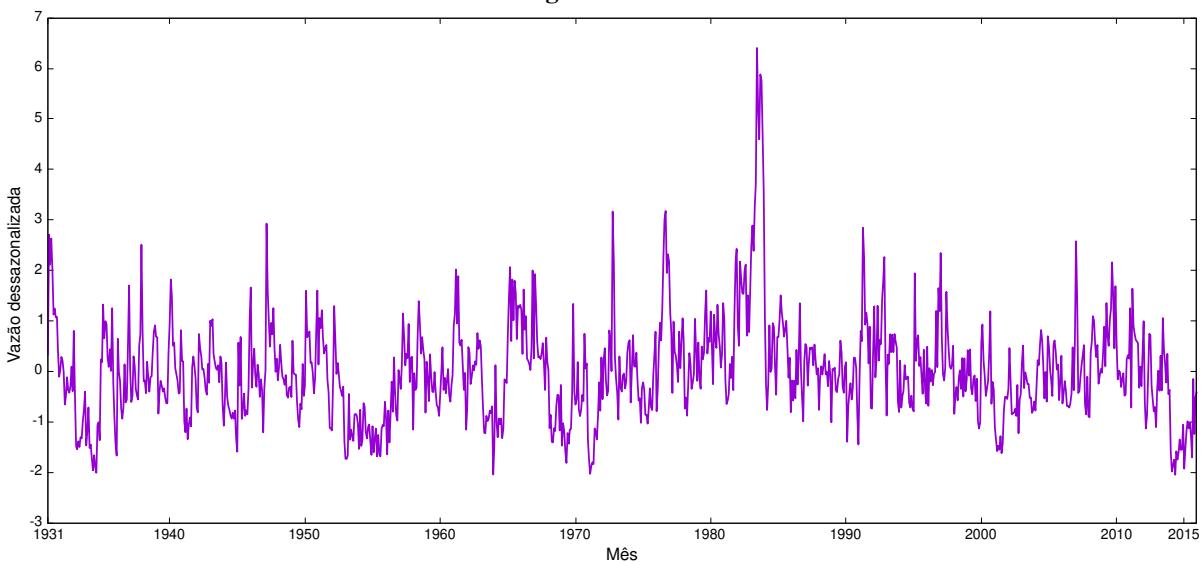
**Gráfico 5 – Série de vazões médias mensais da usina de Água Vermelha**



**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 6 mostra a série dessazonalizada da usina de Água Vermelha, de acordo com a Equação 2.13.

**Gráfico 6 – Série dessazonalizada da usina de Água Vermelha**



**Fonte:** Autoria própria.

Para a realização do treinamento tanto da MLP quanto da RBF a componente sazonal foi retirada da série. Ao final da execução de cada rede, o valor previsto encontra-se sem a

componente sazonal, sendo necessária a reinserção do mesmo nas saídas previstas, para que os mesmos compreendam os valores reais. Além disso, os valores de entrada foram normalizados no intervalo  $[-1, 1]$ .

Dentre todas as variações de MLPs testadas a que apresentou melhor desempenho (menor MSE junto ao conjunto de teste) foi a rede com 30 neurônios na camada intermediária e utilizando os atrasos 1 e 3 como entradas. Seu treinamento foi realizado em 26 épocas.

Por sua vez, a melhor RBF dentre todas as variações testadas foi aquela com 28 neurônios na camada intermediária e utilizando os atrasos 1 e 2 como entradas. Seu treinamento ótimo foi realizado em 200 épocas.

A Tabela 1 apresenta os resultados das previsões realizadas pela melhor MLP e pela melhor RBF para a série de Água Vermelha entre os anos de 2006 e 2015 e para 1, 3, 6 e 12 passos a frente.

Na Tabela 1 é possível verificar 4 medidas de erro para cada previsão:

- **MSE (Dess)** – erro quadrático médio (Equação 4.1) calculado para a saída da rede dessazonalizada;
- **MSE (Real)** – erro quadrático médio (Equação 4.1) calculado para a saída da rede com a adição da componente sazonal, ou seja, calculado para o valor real;
- **MAE (Dess)** – erro absoluto médio (Equação 4.2) calculado para a saída da rede dessazonalizada;
- **MAE (Real)** – erro absoluto médio (Equação 4.2) calculado para a saída da rede com a adição da componente sazonal, ou seja, calculado para o valor real;

**Tabela 1 – Resultados previsões Água Vermelha de 2006 a 2015**

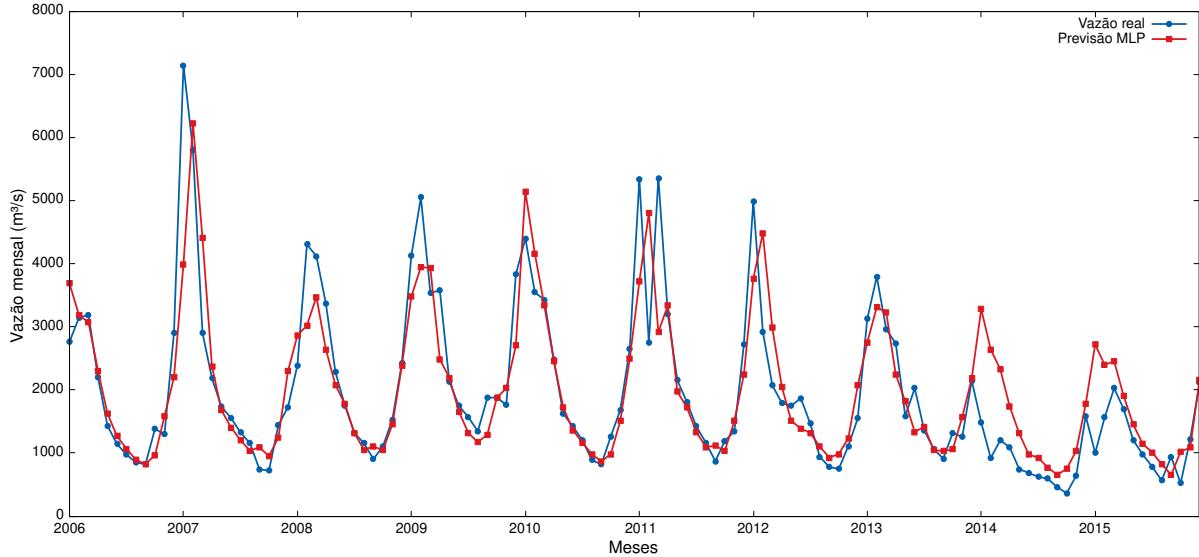
		MSE (Dess)	MSE (Real)	MAE (Dess)	MAE (Real)
<b>1 passo</b>	MLP	0, 460201	470.873, 28	0, 532545	432, 25
	RBF	0, 489075	509.674, 53	0, 555630	456, 37
<b>3 passos</b>	MLP	0, 696479	636.717, 63	0, 670420	522, 18
	RBF	0, 858814	725.035, 54	0, 738309	565, 79
<b>6 passos</b>	MLP	0, 839391	729.692, 63	0, 726681	561, 66
	RBF	0, 893429	751.041, 03	0, 760026	577, 27
<b>12 passos</b>	MLP	0, 844266	646.189, 45	0, 744603	559, 42
	RBF	0, 865863	656.205, 37	0, 750796	562, 65

**Fonte:** Autoria própria.

Analizando a Tabela 1 nota-se que quanto maior o horizonte de previsão, maiores são os valores dos erros obtidos pelas redes, tanto para a MLP quanto para a RBF. Esse fato pode ser explicado pela propagação do erro devido a previsão recursiva e é um resultado esperado em problemas de previsão (SIQUEIRA, 2013).

Pela Tabela 1 verifica-se que a rede MLP teve o melhor desempenho para as previsões em todos os horizontes sugeridos. O Gráfico 7 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Água Vermelha de 2006 a 2015 com horizonte de 1 passo a frente.

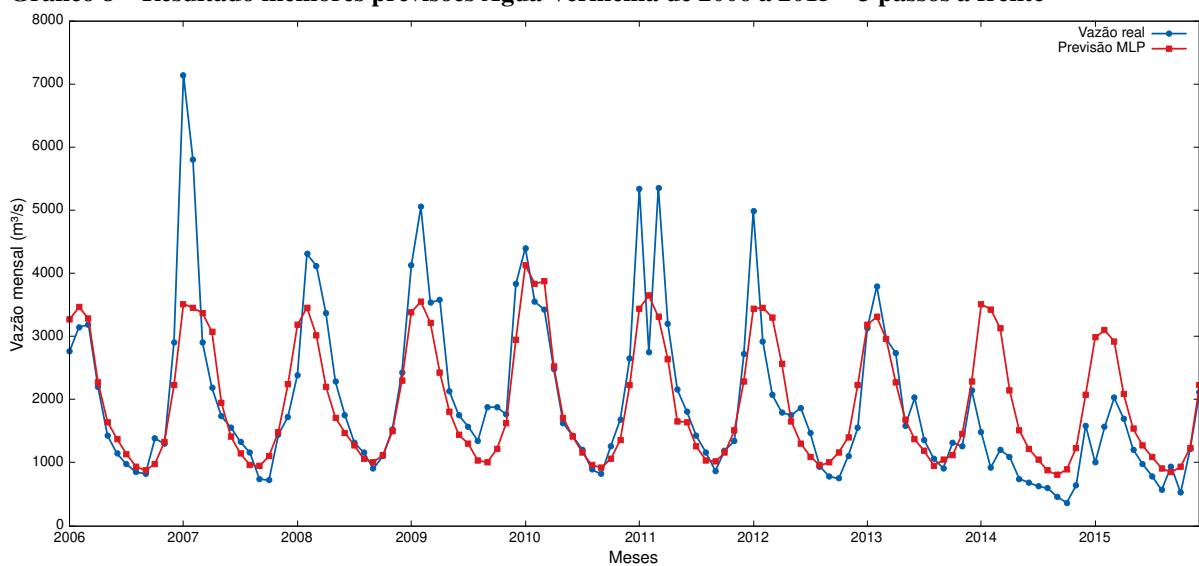
**Gráfico 7 – Resultado melhores previsões Água Vermelha de 2006 a 2015 – 1 passo a frente**



**Fonte:** Autoria própria.

Também pelos dados da Tabela 1 nota-se que as previsões realizadas pela MLP foram melhores que as da RBF para o horizonte de 3 passos a frente. O Gráfico 8 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Água Vermelha de 2006 a 2015 com horizonte de 3 passos a frente.

**Gráfico 8 – Resultado melhores previsões Água Vermelha de 2006 a 2015 – 3 passos a frente**

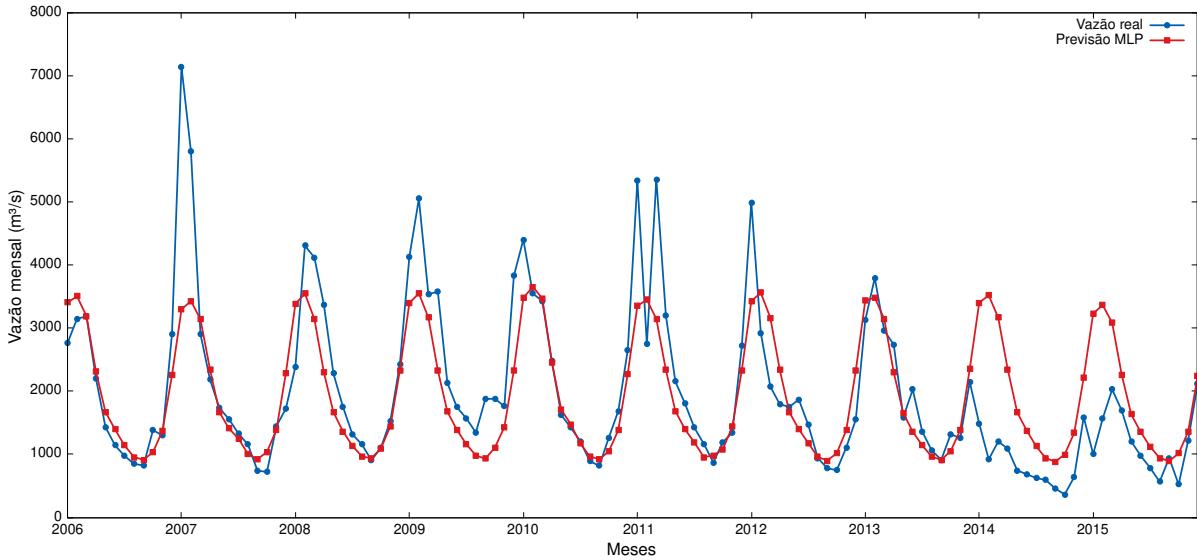


**Fonte:** Autoria própria.

Para o horizonte de previsão de 6 passos a frente a MLP também realizou previsões melhores que a RBF, como pode ser visto nos dados da Tabela 1. O Gráfico 9 apresenta uma

comparação dos valores previstos pela MLP e os valores reais da série de Água Vermelha de 2006 a 2015 com horizonte de 6 passos a frente.

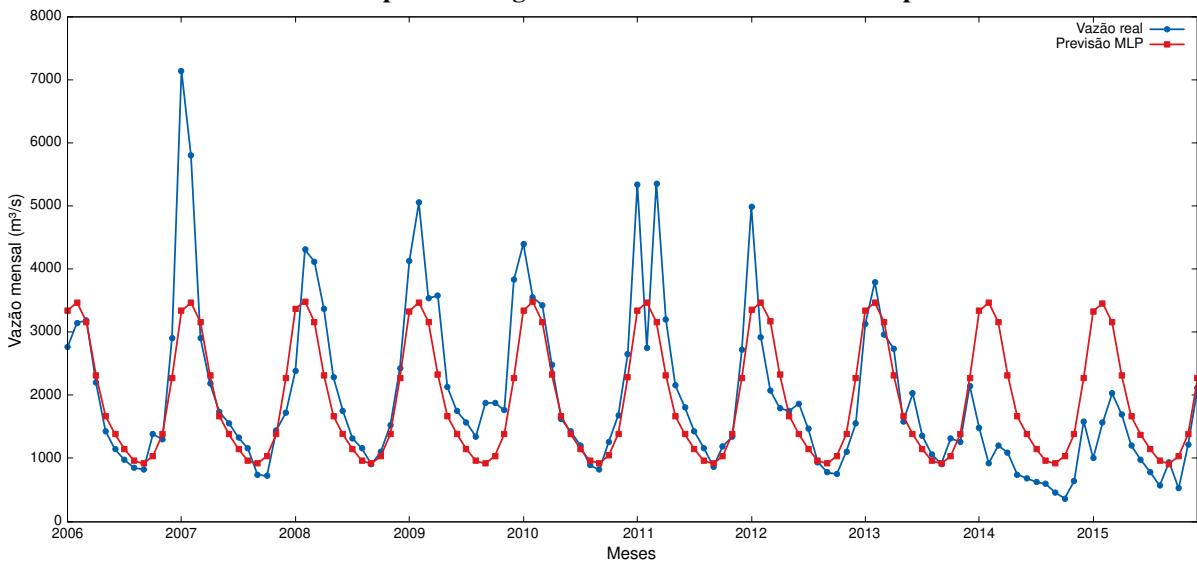
**Gráfico 9 – Resultado melhores previsões Água Vermelha de 2006 a 2015 – 6 passos a frente**



**Fonte:** Autoria própria.

O resultado da previsão com a MLP para a série de Água Vermelha com horizonte de 12 passos a frente pode ser observado no Gráfico 10.

**Gráfico 10 – Resultado melhores previsões Água Vermelha de 2006 a 2015 – 12 passos a frente**



**Fonte:** Autoria própria.

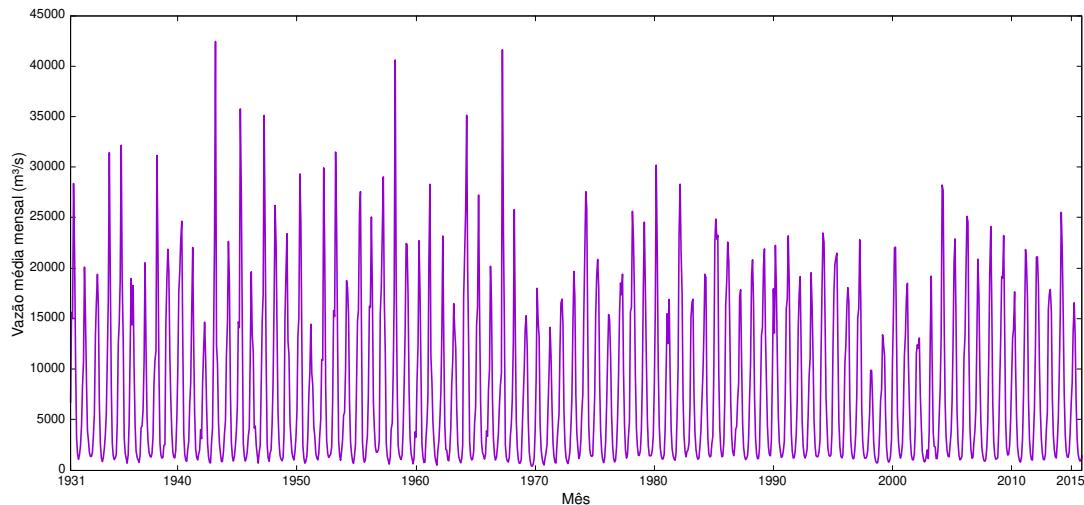
Por fim nota-se analisando os resultados apresentados na Tabela 1 que a rede MLP foi melhor que a RBF para todas as previsões realizadas na série histórica da usina hidrelétrica de Água Vermelha, para todos os horizontes selecionados.

## 4.2 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 hidrelétrica que ainda não se encontra em operação. Apesar disso, o Ministério de Minas e Energia sempre realizou medições da vazão afluente em vários pontos do Rio Xingu, de modo que como as demais 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  $\hat{\sigma} = 7.769,75 \text{ m}^3/\text{s}$ . A série completa é apresentada no Gráfico 11.

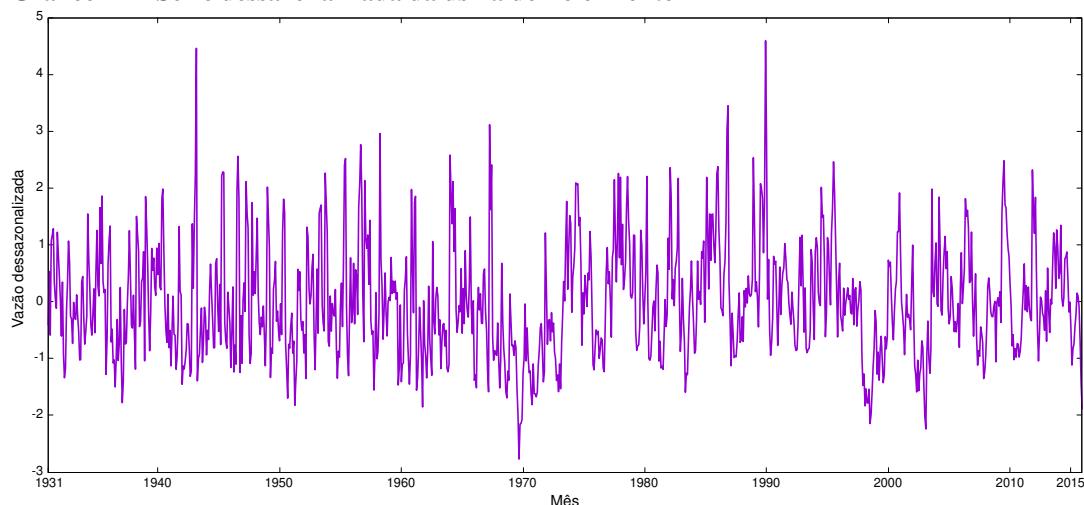
**Gráfico 11 – Série de vazões médias mensais da usina de Belo Monte**



**Fonte:** Autoria própria.

O Gráfico 12 mostra a série dessazonalizada da usina de Belo Monte, de acordo com a Equação 2.13.

**Gráfico 12 – Série dessazonalizada da usina de Belo Monte**



**Fonte:** Autoria própria.

Dentre todas as variações de MLPs testadas a que apresentou melhor desempenho (menor MSE junto ao conjunto de teste) foi a rede com 4 neurônios na camada intermediária e utilizando os atrasos 1 e 6 como entradas. Seu treinamento foi realizado em 124 épocas.

Por sua vez, a melhor RBF dentre todas as variações testadas foi aquela com 28 neurônios na camada intermediária e utilizando os atrasos 1, 3, 5 e 6 como entradas. Seu treinamento foi realizado em 88 épocas.

A Tabela 2 apresenta os resultados das previsões realizadas pela melhor MLP e pela melhor RBF para a série de Belo Monte entre os anos de 2006 e 2015 com horizontes de previsão de 1, 3, 6 e 12 passos a frente.

**Tabela 2 – Resultados previsões Belo Monte de 2006 a 2015**

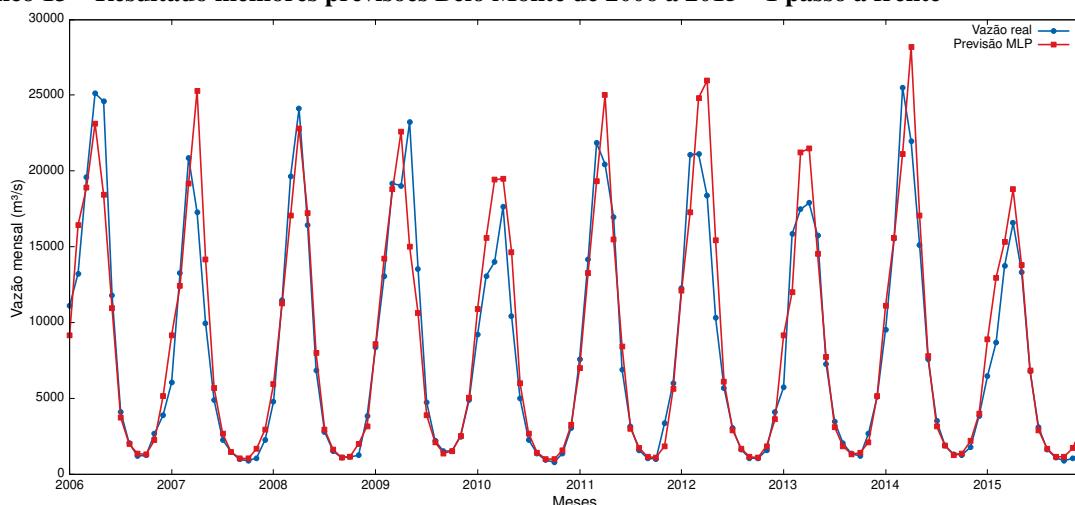
		MSE (Dess)	MSE (Real)	MAE (Dess)	MAE (Real)
<b>1 passo</b>	<b>MLP</b>	<b>0, 381561</b>	<b>5.192.729, 70</b>	<b>0, 478673</b>	<b>1.357, 95</b>
	<b>RBF</b>	0, 501284	5.622.556, 61	0, 566989	1.488, 45
<b>3 passos</b>	<b>MLP</b>	0, 741540	8.319.717, 61	0, 707134	1.825, 82
	<b>RBF</b>	<b>0, 704263</b>	<b>6.987.203, 74</b>	<b>0, 667368</b>	<b>1.663, 36</b>
<b>6 passos</b>	<b>MLP</b>	0, 914698	10.174.613, 94	0, 795821	2.007, 06
	<b>RBF</b>	<b>0, 790317</b>	<b>7.908.640, 21</b>	<b>0, 717699</b>	<b>1.735, 16</b>
<b>12 passos</b>	<b>MLP</b>	1, 069925	10.762.669, 05	0, 883740	2.174, 12
	<b>RBF</b>	<b>1, 027412</b>	<b>8.893.615, 40</b>	<b>0, 794355</b>	<b>1.899, 42</b>

**Fonte:** Autoria própria.

Mais uma vez, nota-se pela Tabela 2 que com o aumento do horizonte de previsão há um aumento nas medidas de erro devido a propagação do erro, que é uma consequência da previsão recursiva.

Pela Tabela 2 verifica-se que a rede MLP obteve melhores resultados que os apresentados pela rede RBF para as previsões de 1 passo a frente. O Gráfico 13 apresenta as previsões da MLP para a série de Belo Monte de 2006 a 2015 com horizonte de 1 passo a frente.

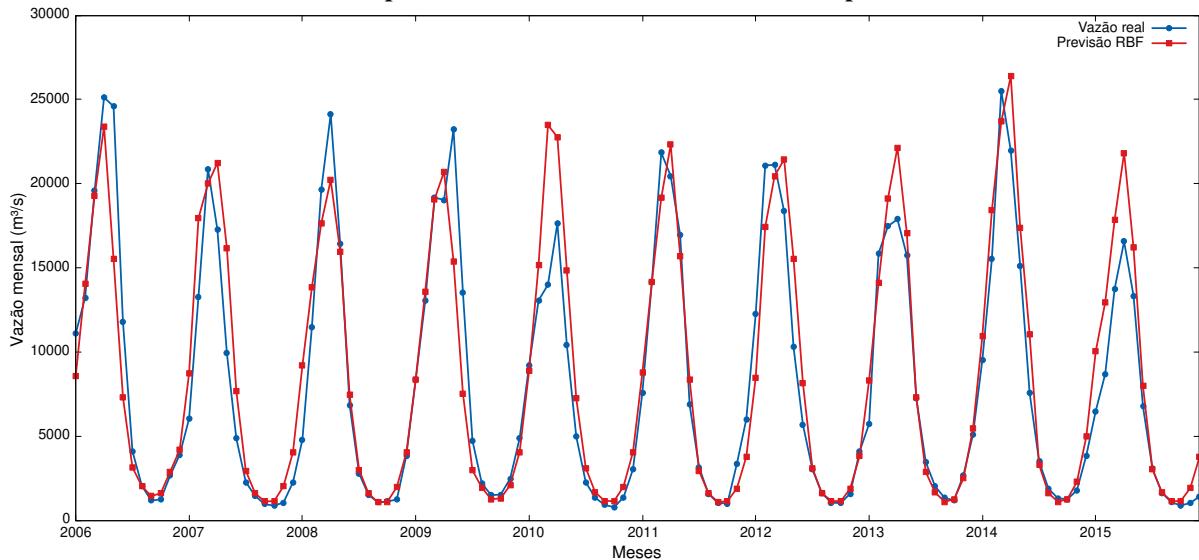
**Gráfico 13 – Resultado melhores previsões Belo Monte de 2006 a 2015 – 1 passo a frente**



**Fonte:** Autoria própria.

De acordo com a Tabela 2 nota-se que, diferentemente da previsão 1 passo a frente, para a previsão 3 passos a frente foi a RBF que obteve os melhores resultados. O Gráfico 14 apresenta as previsões realizadas pela RBF para a série de Belo Monte com 3 passos a frente.

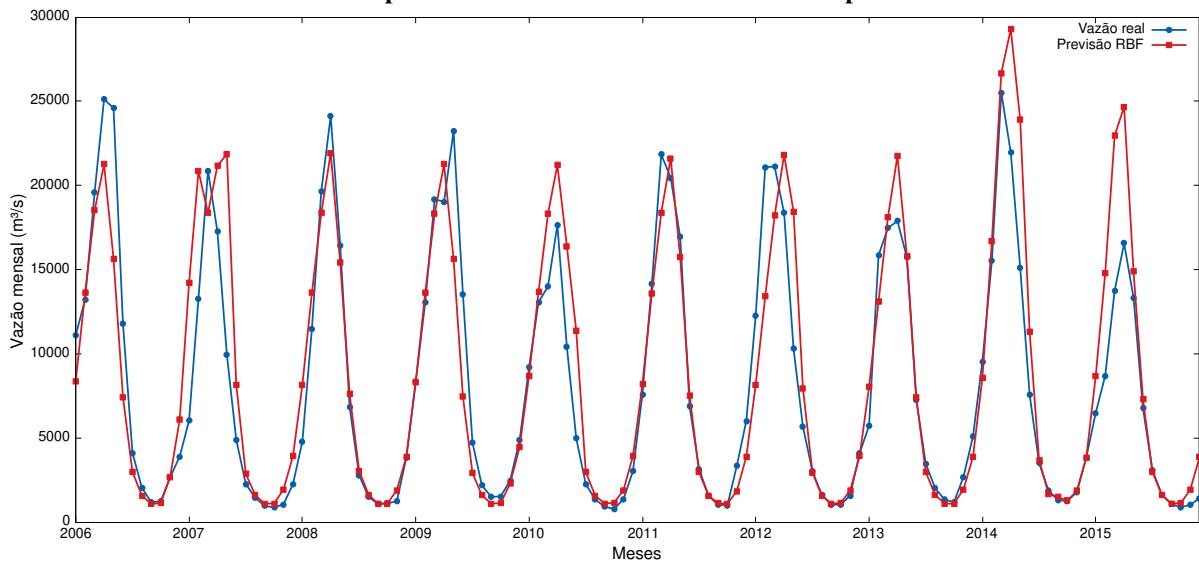
**Gráfico 14 – Resultado melhores previsões Belo Monte de 2006 a 2015 – 3 passos a frente**



**Fonte:** Autoria própria.

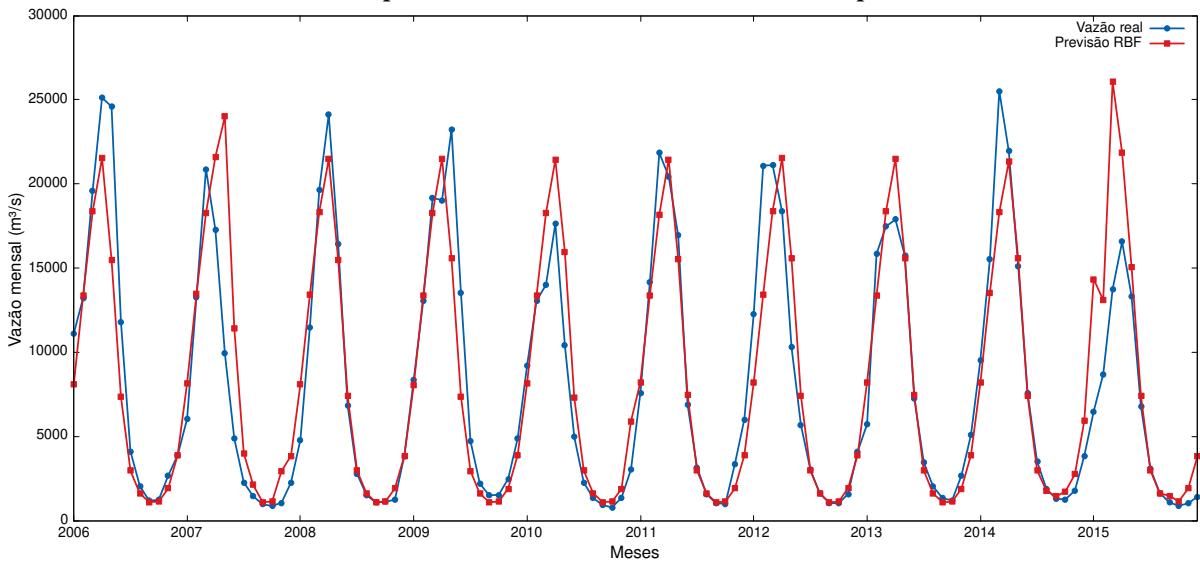
De igual modo, pelos dados da Tabela 2 verifica-se que a rede RBF mais uma vez obteve melhores resultados do que as apresentadas pela MLP, dessa vez para previsões realizadas com horizonte de previsão de 6 passos a frente. O Gráfico 15 apresenta as previsões realizadas pela RBF para a série de Belo Monte com 6 passos a frente.

**Gráfico 15 – Resultado melhores previsões Belo Monte de 2006 a 2015 – 6 passos a frente**



**Fonte:** Autoria própria.

Pelos resultados da Tabela 2 verifica-se que para as previsões de 12 passos a frente a rede RBF mais uma vez obteve melhores resultados. O Gráfico 16 apresenta as previsões realizadas pela RBF para a série de Belo Monte entre os anos de 2006 e 2015 com horizonte de previsão de 12 passos a frente.

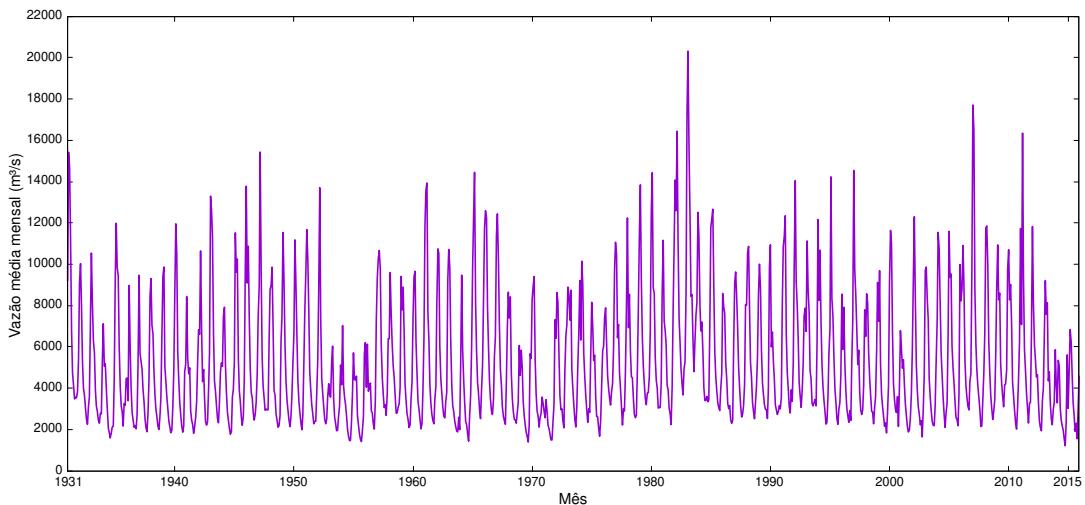
**Gráfico 16 – Resultado melhores previsões Belo Monte de 2006 a 2015 – 12 passos a frente**

**Fonte:** Autoria própria.

Pelos resultados da Tabela 2 verifica-se que a rede MLP foi melhor que a RBF na série de Belo Monte apenas para as previsões com 1 passo a frente enquanto que a RBF obteve os melhores resultados em todos os demais horizontes de previsão (3, 6 e 12 passos a frente).

#### 4.3 USINA HIDRELÉTRICA DE ILHA SOLTEIRA

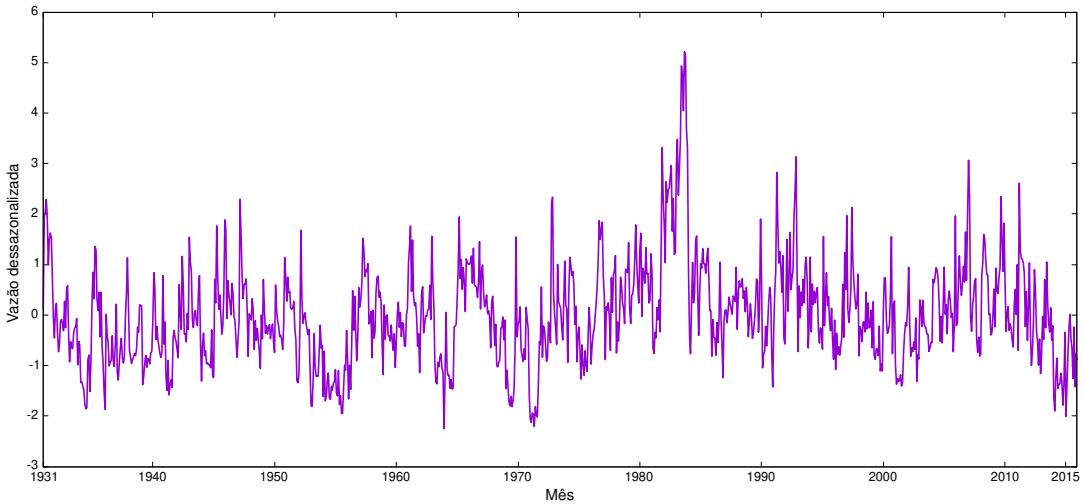
A usina hidrelétrica de Ilha Solteira é a maior do Estado de São Paulo e a sexta maior usina do Brasil. Sendo abastecida pelo Rio Paraná, fica localizada entre os municípios de Ilha Solteira e Selvíria, na divisa dos estados de São Paulo e Mato Grosso do Sul. Sua série histórica apresenta média  $\hat{\mu} = 5.281,94 \text{ m}^3/\text{s}$  e desvio padrão  $\hat{\sigma} = 3.100,62 \text{ m}^3/\text{s}$ , sendo apresentada no Gráfico 17.

**Gráfico 17 – 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 18 mostra a série dessazonalizada da usina de Ilha Solteira, de acordo com a Equação 2.13.

**Gráfico 18 – Série dessazonalizada da usina de Ilha Solteira**



**Fonte:** Autoria própria.

Dentre todas as variações de MLPs testadas a que apresentou melhor desempenho (menor MSE junto ao conjunto de teste) foi a rede com 24 neurônios na camada intermediária e utilizando os atrasos 1 e 2 como entradas. Seu treinamento foi realizado em 14 épocas.

Por sua vez, a melhor RBF dentre todas as variações testadas foi aquela com 16 neurônios na camada intermediária e utilizando o atraso 1 como entrada. Seu treinamento foi realizado em 151 épocas.

A Tabela 3 apresenta os resultados das previsões realizadas pela melhor MLP e pela melhor RBF para a série de Ilha Solteira entre os anos de 2006 e 2015.

**Tabela 3 – Resultados previsões Ilha Solteira de 2006 a 2015**

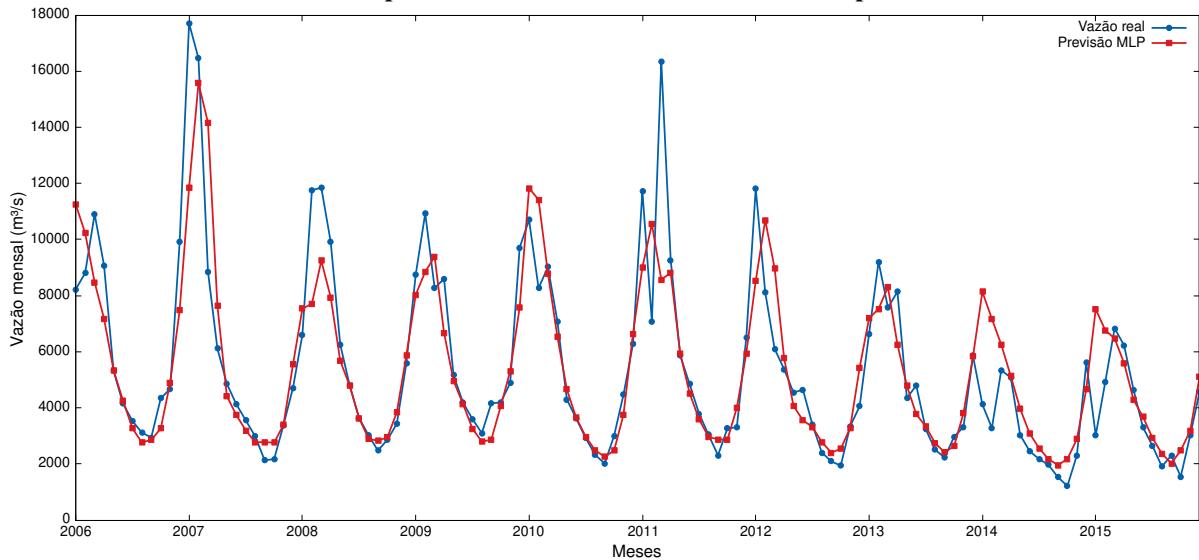
		MSE (Dess)	MSE (Real)	MAE (Dess)	MAE (Real)
<b>1 passo</b>	<b>MLP</b>	<b>0, 515536</b>	<b>2.679.918, 60</b>	<b>0, 546141</b>	<b>981, 15</b>
	RBF	0, 564354	3.015.473, 13	0, 572742	1.020, 05
<b>3 passos</b>	<b>MLP</b>	<b>0, 805773</b>	<b>3.563.557, 64</b>	<b>0, 711329</b>	<b>1.195, 39</b>
	RBF	0, 910331	3.894.962, 90	0, 783177	1.309, 73
<b>6 passos</b>	<b>MLP</b>	<b>0, 972036</b>	<b>4.197.011, 56</b>	<b>0, 764147</b>	<b>1.271, 08</b>
	RBF	1, 000049	4.392.726, 47	0, 815634	1.371, 13
<b>12 passos</b>	<b>MLP</b>	<b>0, 888516</b>	<b>3.595.310, 60</b>	<b>0, 739434</b>	<b>1.211, 60</b>
	RBF	0, 976444	4.312.704, 00	0, 802783	1.355, 59

**Fonte:** Autoria própria.

Mais uma vez, nota-se pela Tabela 3 que com o aumento do horizonte de previsão há um aumento nas medidas de erro devido a propagação do erro, que é uma consequência da previsão recursiva.

Pela Tabela 3 verifica-se que a rede MLP teve o melhor desempenho para as previsões de 1 passo a frente. O Gráfico 19 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Ilha Solteira entre os anos de 2006 e 2015 com horizonte de previsão de 1 passo a frente.

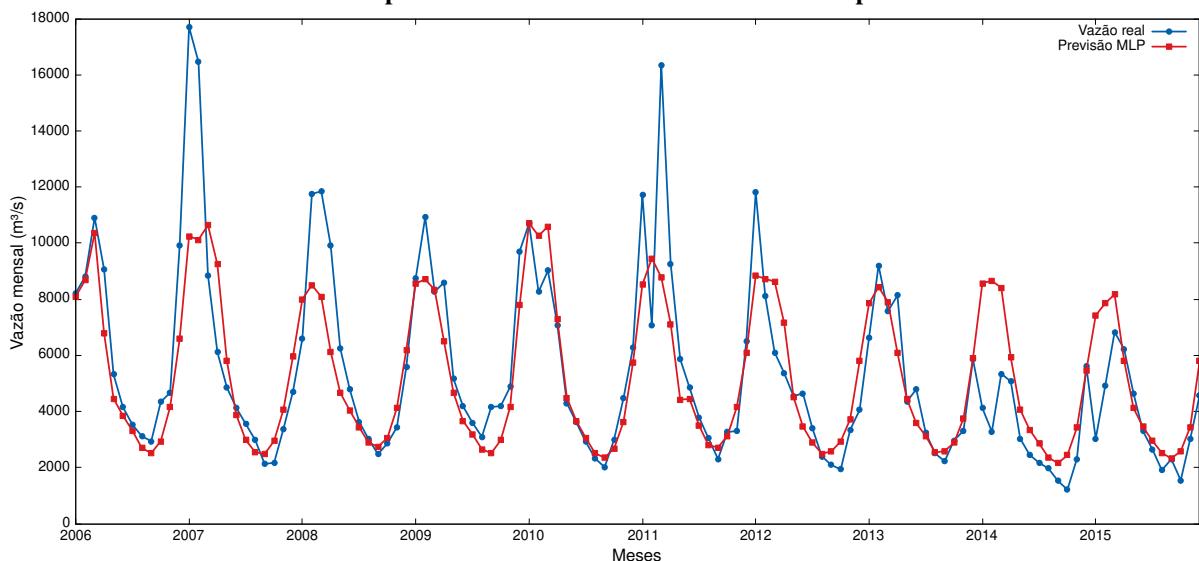
**Gráfico 19 – Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 1 passo a frente**



**Fonte:** Autoria própria.

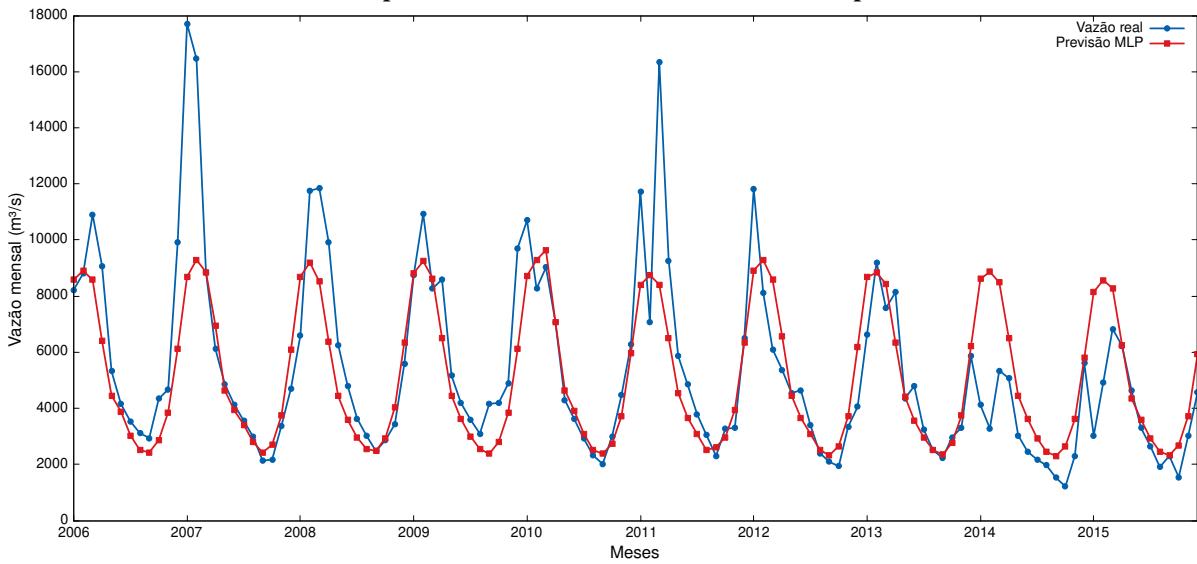
De acordo com a Tabela 3 nota-se que a MLP também obteve os melhores resultados para a previsão 3 passos a frente. O Gráfico 20 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Ilha Solteira entre os anos de 2006 e 2015 com horizonte de previsão de 3 passos a frente.

**Gráfico 20 – Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 3 passos a frente**



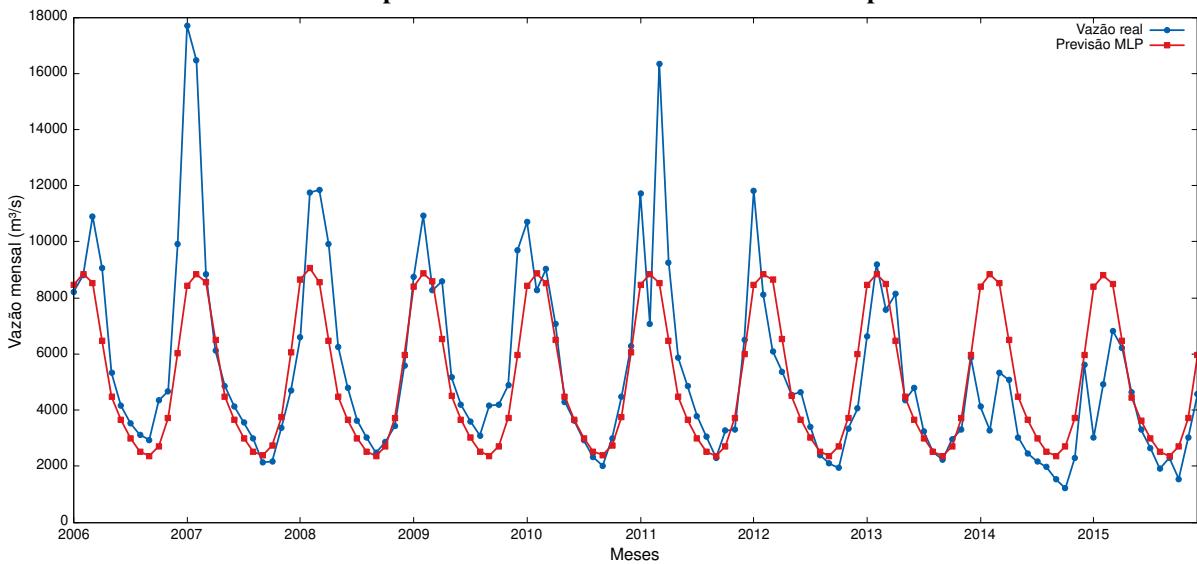
**Fonte:** Autoria própria.

A MLP também obteve os melhores resultados para as previsões 6 passos a frente. Seu resultado pode ser observado no Gráfico 21.

**Gráfico 21 – Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 6 passos a frente**

**Fonte:** Autoria própria.

Os resultados da Tabela 3 indicam que a MLP também obteve os melhores resultados para as previsões 12 passos a frente. Seu resultado para a série de Ilha Solteira com horizonte de 12 passos a frente pode ser observado no Gráfico 22.

**Gráfico 22 – Resultado melhores previsões Ilha Solteira de 2006 a 2015 – 12 passos a frente**

**Fonte:** Autoria própria.

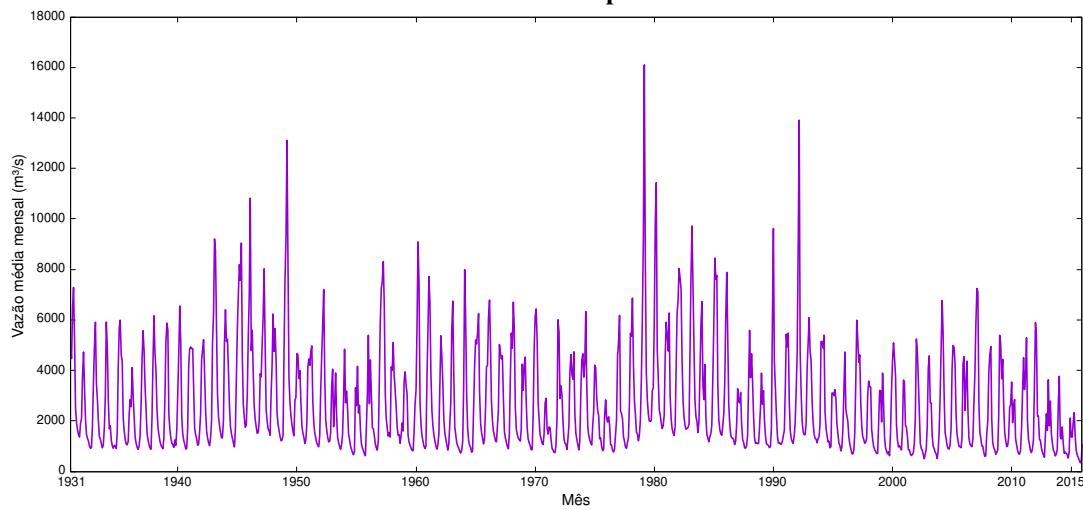
Por fim nota-se analisando os resultados apresentados na Tabela 3 que as previsões realizadas pela MLP foram melhores que as realizadas pela RBF para a série de Ilha Solteira em todos os horizontes de previsão, 1, 3, 6 e 12 passos.

#### 4.4 COMPLEXO HIDRELÉTRICO DE PAULO AFONSO

O Complexo Hidrelétrico de Paulo Afonso é um conjunto de usinas, localizado na cidade de Paulo Afonso no estado da 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 apresenta 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 23.

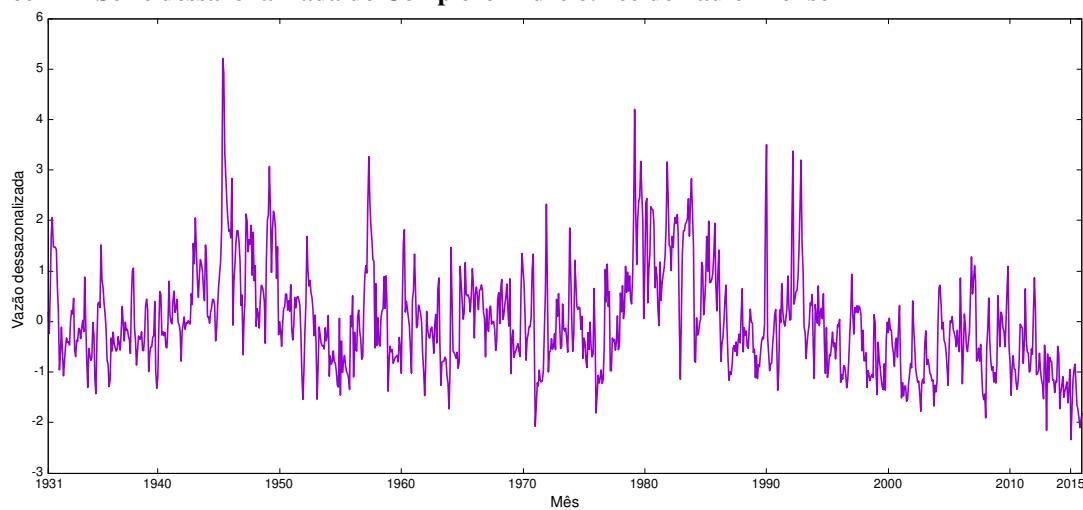
**Gráfico 23 – Série de vazões médias mensais do Complexo Hidrelétrico de Paulo Afonso**



**Fonte:** Autoria própria.

Mais uma vez se faz necessária a retirada da componente sazonal da série para um melhor entendimento de seu comportamento. O Gráfico 24 mostra a série dessazonalizada da usina de Água Vermelha, de acordo com a Equação 2.13.

**Gráfico 24 – Série dessazonalizada do Complexo Hidrelétrico de Paulo Afonso**



**Fonte:** Autoria própria.

Dentre todas as variações de MLPs testadas a que apresentou melhor desempenho foi

a rede com 26 neurônios na camada intermediária e utilizando o atraso 1 como entrada. Seu treinamento foi realizado em 9 épocas.

Por sua vez, a melhor RBF dentre todas as variações testadas foi aquela com 14 neurônios na camada intermediária e utilizando os atrasos 1 e 4 como entradas. Seu treinamento foi realizado em 196 épocas.

A Tabela 4 apresenta os resultados das previsões realizadas pela melhor MLP e pela melhor RBF para a série de Paulo Afonso entre os anos de 2006 e 2015.

**Tabela 4 – Resultados previsões Paulo Afonso de 2006 a 2015**

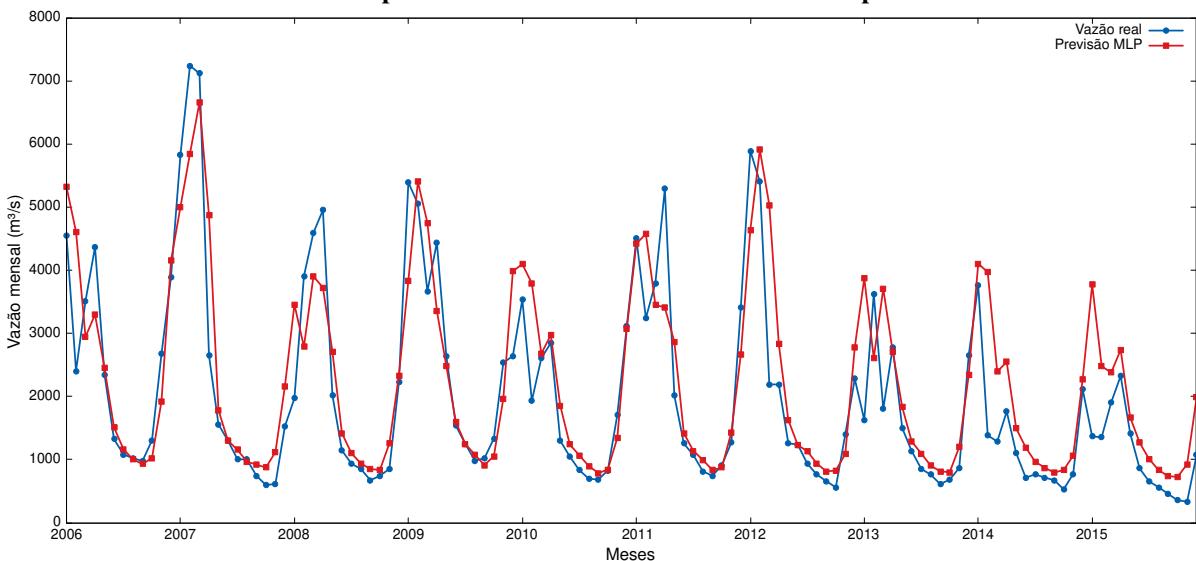
		MSE (Dess)	MSE (Real)	MAE (Dess)	MAE (Real)
<b>1 passo</b>	MLP	<b>0, 310617</b>	<b>653.831, 78</b>	<b>0, 447934</b>	<b>520, 79</b>
	RBF	0, 423874	831.148, 83	0, 552960	610, 91
<b>3 passos</b>	MLP	<b>0, 602407</b>	<b>1.047.995, 88</b>	<b>0, 653315</b>	<b>703, 63</b>
	RBF	0, 859322	1.409.855, 65	0, 801390	830, 82
<b>6 passos</b>	MLP	<b>0, 707444</b>	<b>1.113.357, 62</b>	<b>0, 728197</b>	<b>758, 24</b>
	RBF	1, 226485	1.828.424, 48	0, 983628	1.001, 12
<b>12 passos</b>	MLP	<b>0, 743317</b>	<b>1.214.032, 15</b>	<b>0, 738363</b>	<b>784, 95</b>
	RBF	1, 359673	2.157.155, 83	1, 026072	1.078, 79

**Fonte:** Autoria própria.

Mais uma vez, nota-se pela Tabela 4 que com o aumento do horizonte de previsão há um aumento nas medidas de erro devido a propagação do erro pela previsão recursiva.

Pela Tabela 4 verifica-se que a rede MLP obteve melhores resultados que os apresentados pela rede RBF para as previsões realizadas com horizontes de previsão de 1 passo a frente. O Gráfico 25 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Paulo Afonso de 2006 a 2015 com horizonte de 1 passo a frente.

**Gráfico 25 – Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 1 passo a frente**

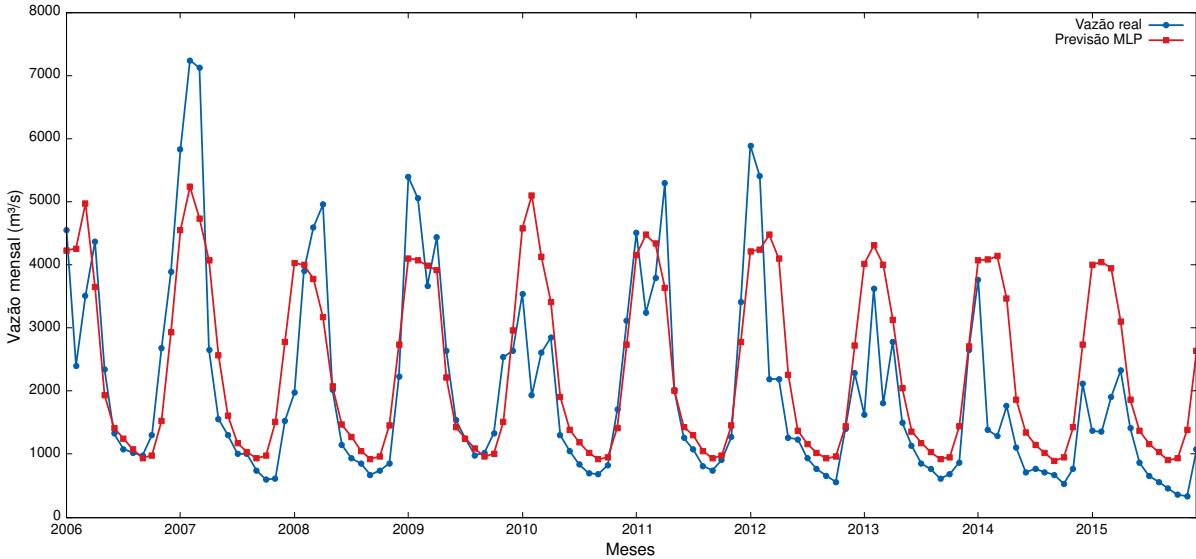


**Fonte:** Autoria própria.

A MLP também apresentou melhores resultados que a RBF para as previsões realizadas

com horizonte de previsão de 3 passos a frente como mostra a Tabela 4. O Gráfico 26 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Paulo Afonso de 2006 a 2015 com horizonte de 3 passos a frente.

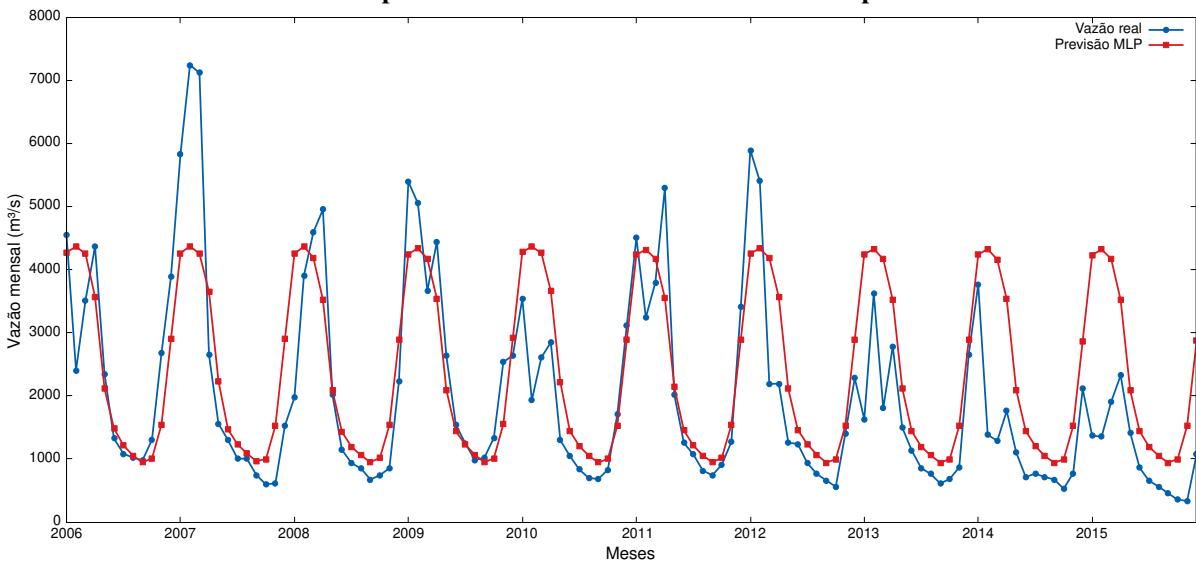
**Gráfico 26 – Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 3 passos a frente**



**Fonte:** Autoria própria.

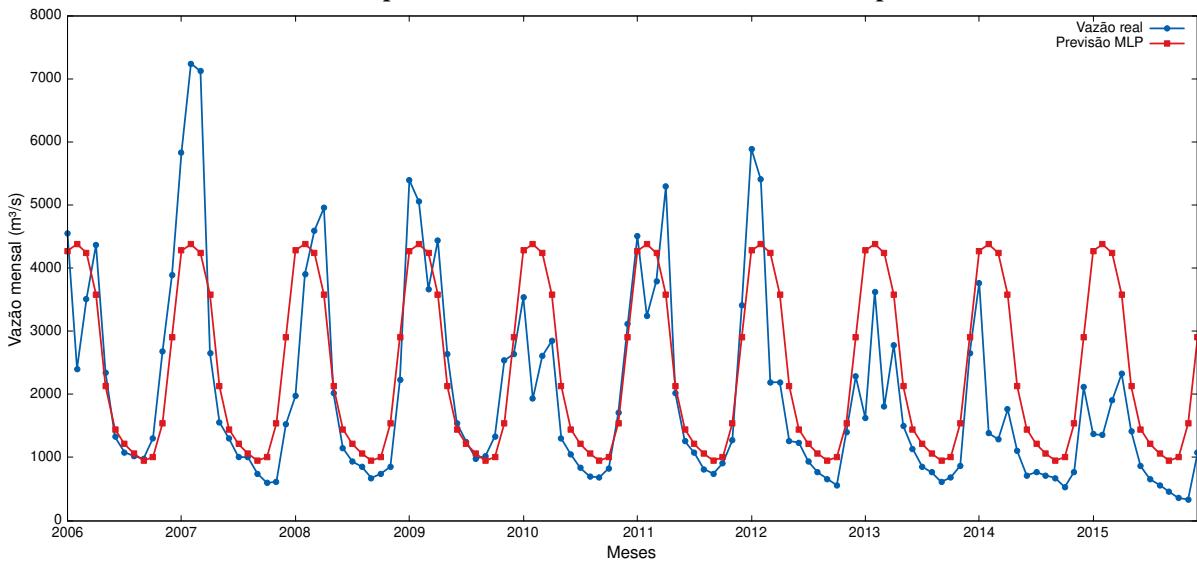
De acordo com a Tabela 4 a MLP também apresentou melhores resultados que os obtidos pela rede RBF para as previsões realizadas com horizonte de previsão de 6 passos a frente. O Gráfico 27 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Paulo Afonso de 2006 a 2015 com horizonte de 6 passos a frente.

**Gráfico 27 – Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 6 passos a frente**



**Fonte:** Autoria própria.

A MLP também obteve melhores resultados que os apresentados pela RBF para as previsões realizadas com horizonte de previsão de 12 passos a frente como mostram os dados da Tabela 4. O Gráfico 28 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Paulo Afonso de 2006 a 2015 com horizonte de 12 passos a frente.

**Gráfico 28 – Resultado melhores previsões Paulo Afonso de 2006 a 2015 – 12 passos a frente**

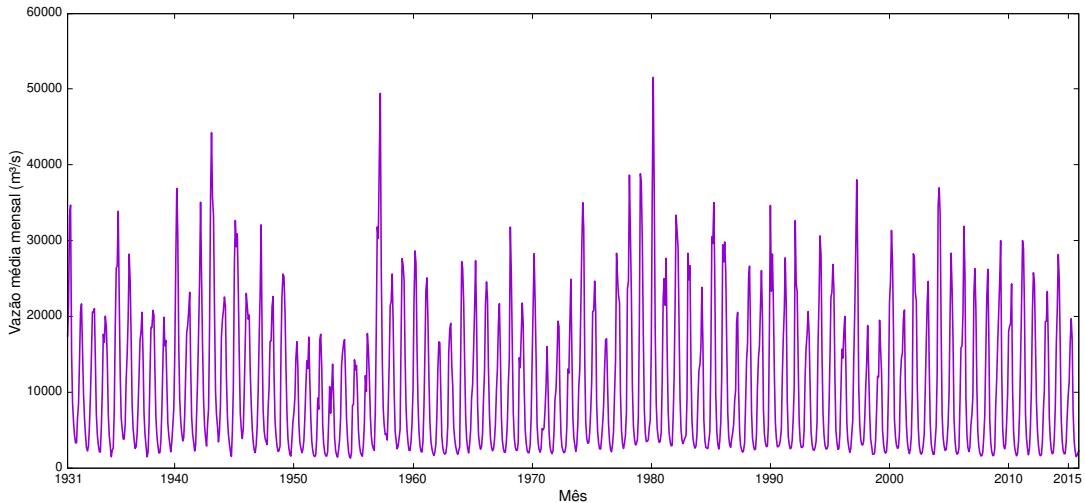
**Fonte:** Autoria própria.

Por fim analisando os dados da Tabela 4 verifica-se que a rede MLP foi melhor que a RBF em todas as previsões realizadas para a série de Paulo Afonso.

#### 4.5 USINA HIDRELÉTRICA DE TUCURUÍ

Sendo abastecida pelo Rio Tocantins, a Usina Hidrelétrica de Tucuruí é a maior usina hidroelétrica 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  $\hat{\sigma} = 9.182,29 \text{ m}^3/\text{s}$ , sendo apresentada no Gráfico 29.

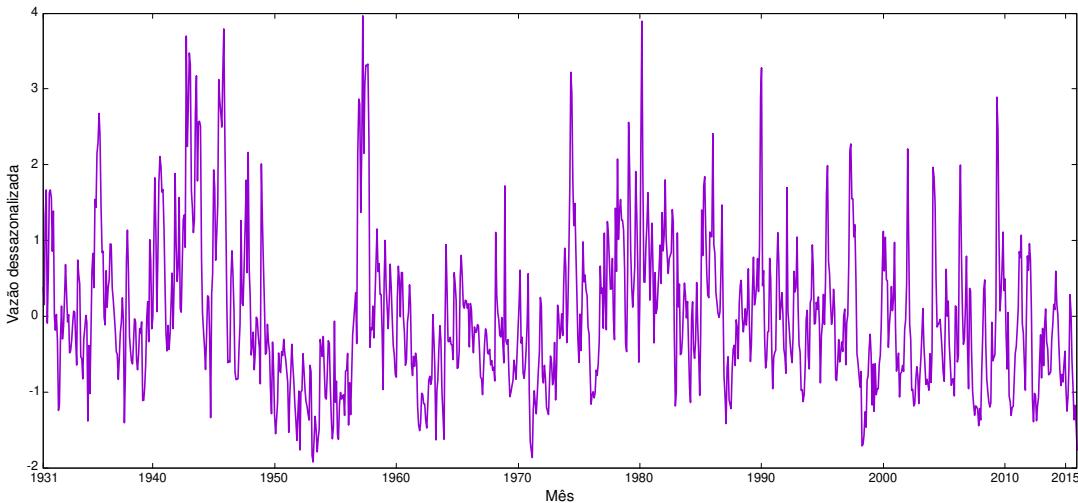
**Gráfico 29 – Série de vazões médias mensais da usina de Tucuruí**

**Fonte:** Autoria própria.

Mais uma vez se faz necessária a retirada da componente sazonal da série para obter

um melhor entendimento do seu comportamento. O Gráfico 6 mostra a série dessazonalizada da usina de Tucuruí, de acordo com a Equação 2.13.

**Gráfico 30 – Série dessazonalizada da usina de Tucuruí**



**Fonte:** Autoria própria.

Dentre todas as variações de MLPs testadas a que apresentou melhor desempenho (menor MSE junto ao conjunto de teste) foi a rede com 2 neurônios na camada intermediária e utilizando os atrasos 1, 2, 5, 6 e 4 como entradas. Seu treinamento foi realizado em 189 épocas.

Por sua vez, a melhor RBF dentre todas as variações testadas foi aquela com 10 neurônios na camada intermediária e utilizando o atraso 1 como entrada. Seu treinamento foi realizado em 60 épocas.

A Tabela 5 apresenta os resultados das previsões realizadas pela melhor MLP e pela melhor RBF para a série de Tucuruí entre os anos de 2006 e 2015 com os horizontes de previsão de 1, 3, 6 e 12 passos a frente.

**Tabela 5 – Resultados previsões Tucuruí de 2006 a 2015**

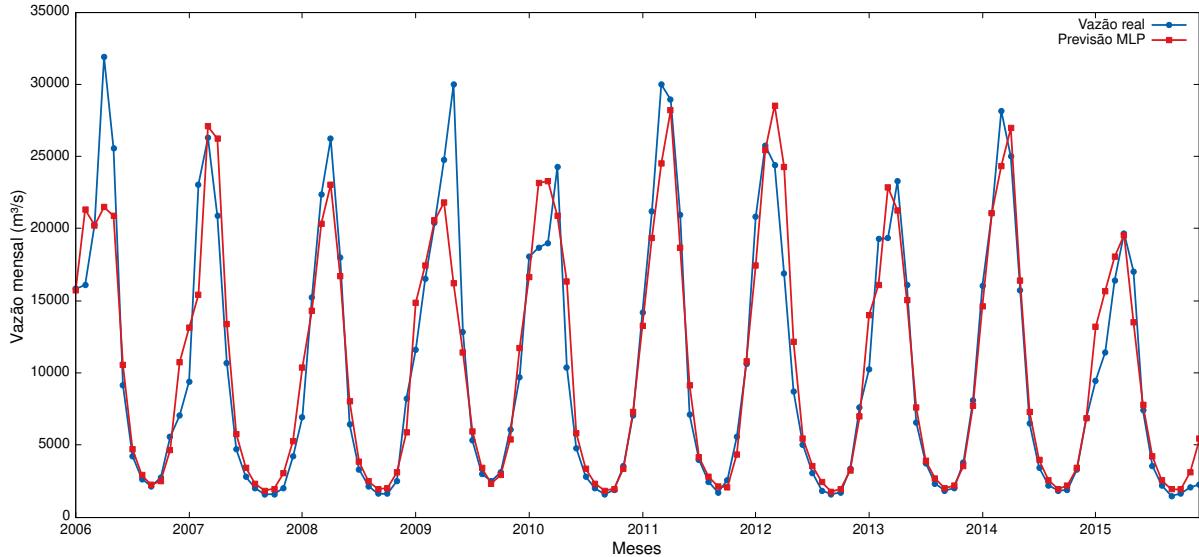
		MSE (Dess)	MSE (Real)	MAE (Dess)	MAE (Real)
<b>1 passo</b>	<b>MLP</b>	<b>0, 307137</b>	<b>6.436.393, 41</b>	<b>0, 442522</b>	<b>1.573, 66</b>
	RBF	0, 377722	7.714.626, 24	0, 495276	1.735, 42
<b>3 passos</b>	<b>MLP</b>	<b>0, 728643</b>	<b>10.847.421, 73</b>	<b>0, 693971</b>	<b>2.149, 68</b>
	RBF	0, 842290	11.396.106, 09	0, 760398	2.281, 25
<b>6 passos</b>	<b>MLP</b>	<b>0, 835925</b>	<b>11.551.915, 99</b>	<b>0, 743534</b>	<b>2.295, 12</b>
	RBF	0, 967217	12.319.596, 30	0, 819796	2.385, 27
<b>12 passos</b>	<b>MLP</b>	<b>0, 890338</b>	<b>11.967.742, 49</b>	<b>0, 782142</b>	<b>2.335, 13</b>
	RBF	0, 987807	12.325.127, 76	0, 842732	2.481, 38

**Fonte:** Autoria própria.

Mais uma vez, nota-se pela Tabela 5 que com o aumento do horizonte de previsão há um aumento nas medidas de erro devido a propagação do erro, que é uma consequência da previsão recursiva.

Pela Tabela 5 verifica-se que a rede MLP teve o melhor desempenho para as previsões de 1 passo a frente. O Gráfico 31 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Tucuruí de 2006 a 2015 com horizonte de previsão de 1 passo a frente.

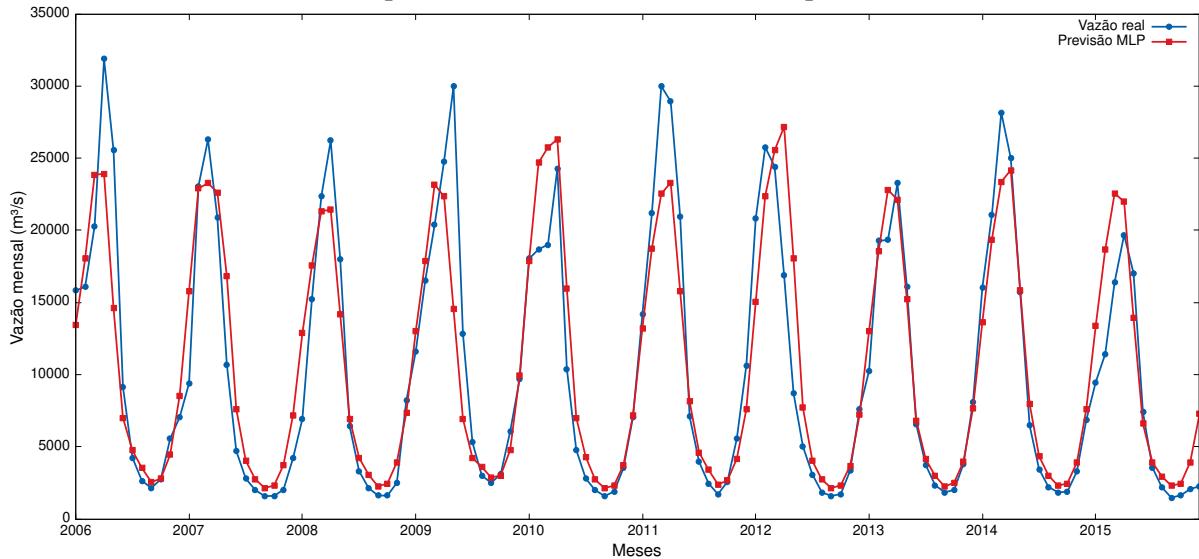
**Gráfico 31 – Resultado melhores previsões Tucuruí de 2006 a 2015 – 1 passo a frente**



**Fonte:** Autoria própria.

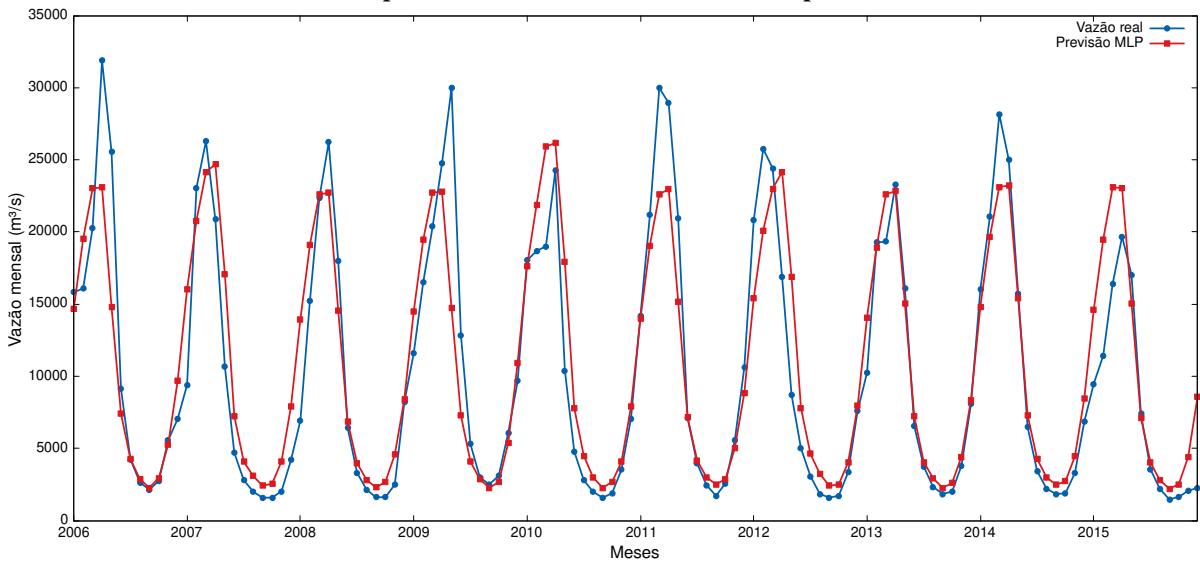
De acordo com a Tabela 5 a MLP também foi melhor que a RBF para as previsões de 3 passos a frente. O Gráfico 32 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Tucuruí de 2006 a 2015 com horizonte de 3 passos a frente.

**Gráfico 32 – Resultado melhores previsões Tucuruí de 2006 a 2015 – 3 passos a frente**



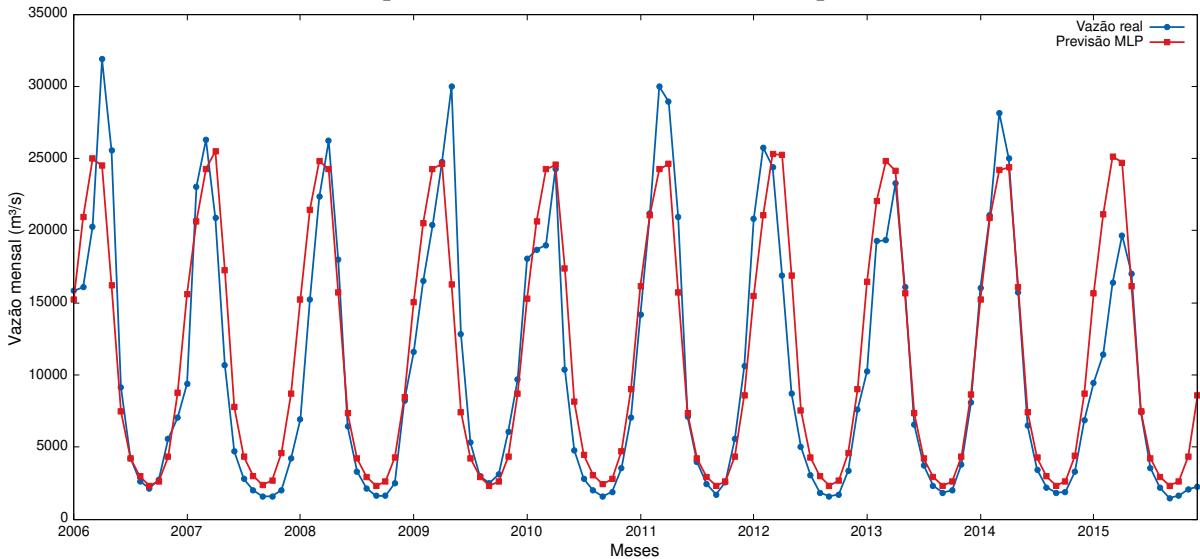
**Fonte:** Autoria própria.

A MLP também foi melhor que a RBF nas previsões 6 passos a frente como mostra a Tabela 5. O Gráfico 33 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Tucuruí de 2006 a 2015 com horizonte de 6 passos a frente.

**Gráfico 33 – Resultado melhores previsões Tucuruí de 2006 a 2015 – 6 passos a frente**

**Fonte:** Autoria própria.

Pela Tabela 5 pode-se observar que a MLP também foi melhor que a RBF para as previsões 12 passos a frente. O Gráfico 34 apresenta uma comparação dos valores previstos pela MLP e os valores reais da série de Tucuruí de 2006 a 2015 com horizonte de 12 passos a frente.

**Gráfico 34 – Resultado melhores previsões Tucuruí de 2006 a 2015 – 12 passos a frente**

**Fonte:** Autoria própria.

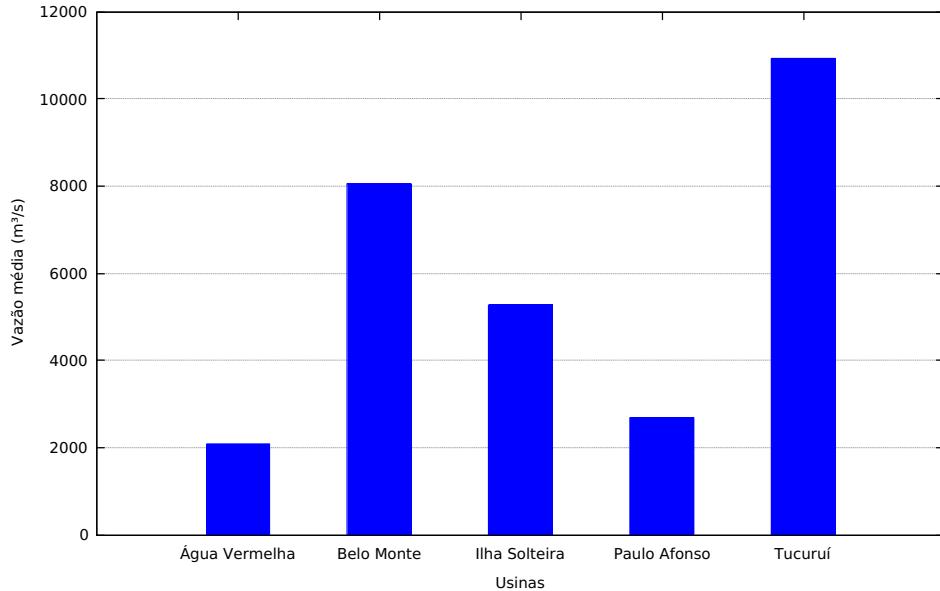
Por fim analisando os dados da Tabela 5 verifica-se que a rede MLP foi melhor que a RBF em todas as previsões realizadas para a série de Tucuruí.

#### 4.6 COMPARAÇÃO 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 35, a qual mostra com a vazão média de cada usina.

**Gráfico 35 – Vazões médias das usinas hidrelétricas**



**Fonte:** Autoria própria.

Note pelo Gráfico 35 que as únicas usinas com valor de vazão média próximos são Água Vermelha (com  $2.077 m^3/s$ ) e Paulo Afonso (com  $2.698 m^3/s$ ). Isso indica que as redes neurais aplicadas foram testadas para cenários diferentes, validando assim seu poder de generalização em diversas situações reais.

As previsões para as 5 séries e os 4 horizontes propostos apresentam no Quadro 1 qual a rede que alcançou o melhor desempenho, para cada caso.

**Quadro 1 – Melhor rede para previsão de cada série**

Série	1 passo	3 passos	6 passos	12 passos
<b>Água Vermelha</b>	MLP	MLP	MLP	MLP
<b>Belo Monte</b>	MLP	RBF	RBF	RBF
<b>Ilha Solteira</b>	MLP	MLP	MLP	MLP
<b>Paulo Afonso</b>	MLP	MLP	MLP	MLP
<b>Tucuruí</b>	MLP	MLP	MLP	MLP

**Fonte:** Autoria própria.

Analizando o Quadro 1 pode-se notar que a rede RBF foi melhor apenas para as previsões da série de Belo Monte com horizonte de previsão de 3, 6 e 12 passos adiante, totalizando 15% das previsões realizadas. A MLP foi melhor em todos os demais cenários, ou 85% das previsões.

Este é um forte indicativo da predominância da MLP como uma arquitetura mais adequada para resolver o problema da previsão de vazões afluentes.

## 5 CONCLUSÃO

Este documento foi elaborado como parte dos requisitos necessários para o exame de qualificação do autor. Entretanto, o desenvolvimento da dissertação de mestrado ainda envolverá algumas etapas, dentre as quais pode-se citar:

- Aplicação e análise de desempenho das máquinas desorganizadas no problema;
- Aplicação de metodologia linear, como modelos autorregressivos (AR), para fins de comparação;
- Análise do uso de outras metodologias de seleção de variáveis, como filtros;
- Utilização de modelos de *ensamble*, sobretudo utilizando outras redes neurais para fazer a combinação dos resultados.
- Aplicação das técnicas propostas em séries de vazões diárias e semanais.

### 5.1 CRONOGRAMA PROPOSTO

A seguir, o Quadro 2 apresenta o cronograma proposto para realização das etapas finais de elaboração dessa dissertação até a data de defesa.

**Quadro 2 – Cronograma das etapas finais**

<b>2018</b>					
<b>Fevereiro</b>	<b>Março</b>	<b>Abril</b>	<b>Maio</b>	<b>Junho</b>	<b>Julho</b>
Exame de qualificação.	Continuação dos estudos sobre máquinas desorganizadas.	Estudo de metodologias lineares.	Estudo de novas metodologias para seleção de variáveis.	Estudo de modelos de <i>ensamble</i> .	Análise de todos os resultados obtidos.
Início dos estudos sobre máquinas desorganizadas.	Aplicação de metodologias lineares.	Aplicação das metodologias para seleção de variáveis estudadas.	Aplicação dos modelos de <i>ensamble</i> estudados.	Escrita da Dissertação.	Defesa final da dissertação.
	Escrita da Dissertação.	Escrita da Dissertação.	Escrita da Dissertação.	Escrita da Dissertação.	

**Fonte:** Autoria própria.

## REFERÊNCIAS

- BALLINI, Rosangela. **Analise e previsões de vasões utilizando modelos de series temporais, redes neurais e redes neurais nebulosas.** 169 f. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2000.
- BARTLETT, Peter 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.
- BOCCATO, Levy. **cNovas propostas e aplicações de redes neurais com estados de eco.** 226 f. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2013.
- BOX, George Edward Pelham; JENKINS, Gwilym Meirion. **Time series analysis: forecasting and control, revised ed.** [S.l.]: Holden-Day, 1976.
- BOX, George Edward Pelham; JENKINS, Gwilym Meirion; REINSEL, Gregory C. **Time Series Analysis: Forecasting and Control.** 4. ed. New Jersey: John Wiley & Sons, 2008. ISBN 978-0-470-27284-8.
- BRAGA, Antônio de Pádua; de Carvalho, André Ponce de Leon Ferreira; LUDELMIR, Teresa Bernarda. **Redes neurais artificiais: teoria e aplicações.** 2. ed. Rio de Janeiro: LTC Editora, 2007. ISBN 978-85-216-1564-4.
- 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.
- CYBENKO, George. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals, and Systems (MCSS)**, Springer, v. 2, n. 4, p. 303–314, 1989.
- EHLERS, Ricardo Sandes. **Análise de Séries Temporais.** 3. ed. Curitiba: Departamento de Estatística, UFFPR, 2005. Disponível em: <<http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>>. Acesso em: 14 jun. 2017.
- EPE, Empresa de Pesquisa Energética. **Balanço Energético Nacional 2017: Ano base 2016.** Rio de Janeiro: Ministério de Minas e Energia, 2017. Disponível em: <[https://ben.epe.gov.br/downloads/Relatorio\\_Final\\_BEN\\_2017.pdf](https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2017.pdf)>. Acesso em: 27 jan. 2018.
- FILHO, Secundino Soares. Planejamento da operação de sistemas hidrotérmicos. **SBA-Controle e Automação**, v. 1, n. 2, p. 122–131, 1987.
- FRANCELIN, R; BALLINI, R; ANDRADE, MG. 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.
- HAYKIN, Simon. **Redes Neurais: principios e prática.** 2. ed. Porto Alegre: Bookman, 2001. ISBN 978-85-7307-718-6.

- \_\_\_\_\_. **Neural Networks and Learning Machines**. 3. ed. New York: Pearson Prentice Hall, 2008. ISBN 978-0131471399.
- HODGKIN, Alan L; HUXLEY, Andrew 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, Guang-Bin *et al.* 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, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. 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.
- \_\_\_\_\_. Extreme learning machine: theory and applications. **Neurocomputing**, Elsevier, v. 70, n. 1, p. 489–501, 2006.
- JAEGER, Herbert. 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.
- \_\_\_\_\_. Adaptive nonlinear system identification with echo state networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2003. p. 609–616.
- MAGALHÃES, Marina Hirota. **Redes neurais, metodologias de agrupamento e combinação de previsores aplicados a previsão de vazões naturais**. 107 f. Dissertação (Mestrado em Engenharia Eletrica) — Universidade Estadual de Campinas, Campinas, 2004.
- MASON, JC; PRICE, RK; 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, Warren S; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.
- MORETTIN, Pedro A; TOLOI, Clélia. **Análise de séries temporais**. 2. ed. São Paulo: Blucher, 2006.
- MORETTIN, Pedro Alberto; TOLOI, CM de C. Previsão de séries temporais. **São Paulo: Atual**, 1987.
- 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, Mustafa C; XU, Dongming; PRÍNCIPE, José C. Analysis and design of echo state networks for function approximation. **Neural computation**, MIT Press, v. 19, n. 1, p. 111–138, 2007.
- PEREIRA, Basilio de Bragança; PAIS, Maria Bernadette Zanusso; 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.

- RASMUSSEN, Peter F *et al.* Estimation and validation of contemporaneous parma models for streamflow simulation. **Water Resources Research**, Wiley Online Library, v. 32, n. 10, p. 3151–3160, 1996.
- RUMELHART, David E; HINTON, Geoffrey E; WILLIAMS, Ronald J. Learning representations by back-propagating errors. **Cognitive modeling**, v. 5, n. 3, p. 1, 1986.
- SACCHI, Rodrigo *et al.* Water inflow forecasting using the echo state network: A brazilian case study. In: IEEE. **Neural Networks, 2007. IJCNN 2007. International Joint Conference on.** [S.I.], 2007. p. 2403–2408.
- SCHAEFER, Anton Maximilian; ZIMMERMANN, Hans-Georg. Recurrent neural networks are universal approximators. **International journal of neural systems**, World Scientific, v. 17, n. 04, p. 253–263, 2007.
- SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. **Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas - Curso Pratico**. 1. ed. São Paulo: ARTLIBER, 2010. ISBN 978-85-88098-53-4.
- SIQUEIRA, Hugo Valadares. **Previsão de séries de vazões com redes neurais artificiais e modelos lineares ajustados por algoritmos bio-inspirados**. 156 f. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2009.
- SIQUEIRA, Hugo Valadares. **Máquinas Desorganizadas para Previsão de Séries de Vazões**. 218 f. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2013.
- SIQUEIRA, Hugo Valadares *et al.* Echo state networks in seasonal streamflow series prediction. **Learning and Nonlinear Models**, v. 10, p. 181–191, 2012.
- SPIEGEL, Murray R. **Estatística**. 3. ed. São Paulo: Makron Books, 1993.
- SRINIVASAN, Dipti; LIEW, AC; CHANG, CS. A neural network short-term load forecaster. **Electric Power Systems Research**, Elsevier, v. 28, n. 3, p. 227–234, 1994.
- TANG, Zaiyong; FISHWICK, Paul A. Feedforward neural nets as models for time series forecasting. **ORSA journal on computing**, Informs, v. 5, n. 4, p. 374–385, 1993.
- TURING, Alan Mathison. Intelligent machinery. **National Physical Laboratory**, 1948. Disponível em: <[http://www.alanturing.net/turing\\_archive/archive/I/I32/L32-001.html](http://www.alanturing.net/turing_archive/archive/I/I32/L32-001.html)>. Acesso em: 16 jun. 2017.
- ZEALAND, Cameron M; BURN, Donald H; SIMONOVIC, Slobodan P. Short term streamflow forecasting using artificial neural networks. **Journal of hydrology**, Elsevier, v. 214, n. 1, p. 32–48, 1999.