



# **Redes Neurais Compostas Por Blocos de Regressões Sigmóides Não-Lineares Aplicadas à Previsão De Séries Temporais**

Trabalho de Conclusão de Curso

Engenharia de Computação

**Aluno: Filipe Lopes Mendes**  
**Orientador: Prof. Dr. Mêuser Jorge Silva Valença**



UNIVERSIDADE  
DE PERNAMBUCO

Universidade de Pernambuco  
Escola Politécnica de Pernambuco  
Graduação em Engenharia de Computação

Filipe Lopes Mendes

REDES NEURAIS COMPOSTAS POR  
BLOCOS DE REGRESSÕES  
SIGMÓIDES NÃO-LINEARES  
APLICADAS À PREVISÃO DE SÉRIES  
TEMPORAIS

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, junho de 2018.

## MONOGRAFIA DE FINAL DE CURSO

### Avaliação Final (para o presidente da banca)\*

No dia 12 de julho de 2018, às 9:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente **FILIPE LOPES MENDES**, orientado pelo professor **Mêuser Jorge Silva Valença**, sob título **Redes Neurais Compostas por Blocos de Regressões Sigmóides Não Lineares Aplicadas à Previsão de Séries Temporais**, a banca composta pelos professores:

**Sérgio Mário Lins Galdino**

**Mêuser Jorge Silva Valença**

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada       Aprovada com Restrições\*       Reprovada

e foi-lhe atribuída nota: 10,0 ( *dez* )

\*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 7 dias para entrega da versão final da monografia a contar da data deste documento.

*Sérgio Mário Lins Galdino*  
\_\_\_\_\_  
**SÉRGIO MÁRIO LINS GALDINO**

*Mêuser Jorge Silva Valença*  
\_\_\_\_\_  
**MÊUSER JORGE SILVA VALENÇA**

*Este trabalho é dedicado à minha mãe, namorada e amigos que me apoiaram e contribuíram com meu crescimento pessoal e intelectual.*

# Agradecimentos

Primeiramente gostaria de agradecer à minha família, em especial a minha mãe por todo o apoio durante esses cinco anos de dedicação ao curso. Aos amigos que fiz, Danilo Rocha, Erik Vinicius, Fernando Ribeiro, Luca Pimentel e Marcos Victor, os quais compartilhei noites em claro, várias horas de estudo e até discursões por ser rigoroso demais no desenvolvimento das atividades acadêmicas.

Gostaria também de agradecer ao Professor Dr. Byron Leite, que me deu a oportunidade de começar minhas pesquisas na Universidade através da iniciação científica e posteriormente minha carreira profissional, além do aprimoramento dos meus conhecimentos através das disciplinas que cursei e fui monitor, uma das quais, me permitiu conhecer minha namorada. Esta inclusive, Maria Paula, que, também agradeço pelo apoio e paciência pelas várias vezes que não pude acompanhá-la em eventos devido às atividades acadêmicas.

Por fim, gostaria de agradecer ao meu orientador, o professor Dr. Mêuser Valença, por me aceitar como orientando e compartilhar seus conhecimentos e experiências acadêmicas e profissionais.

# Resumo

A predição de séries temporais já é aplicada em diversas áreas, como: meteorologia, mercado financeiro, telecomunicações, hidrologia e outros. A maioria das técnicas existentes são baseadas em modelos estatísticos lineares que têm seus resultados prejudicados por conta da não-linearidade e complexidade da maioria das series temporais. Diante disso, algumas técnicas de Inteligência Artificial, como as Redes Neurais Artificiais, passaram a ser utilizadas como alternativa aos métodos estatísticos devido à capacidade de resolver problemas não lineares, gerando melhores resultados. Mas para que uma rede neural consiga gerar bons resultados é necessário que sua topologia seja corretamente determinada, o que não é uma tarefa trivial. Com isso, este trabalho propôs o uso das Redes Neurais Compostas por Blocos de Regressões Sigmoides Não-Lineares para prever o comportamento de séries temporais aproveitando-se, dentre outros fatores, da capacidade que estas redes têm em determinar uma arquitetura ótima de forma automática durante o seu treinamento. Os resultados obtidos na previsão de dados de três séries temporais foram muito interessantes, gerando, em quase todos os cenários de testes, um menor EMPA com o uso da técnica proposta em relação às técnicas estatísticas de Regressão Linear e o método de Holt-Winters que foram utilizadas para a validação do desempenho.

**Palavras-chave:** Séries Temporais, Previsão, Estatística, Inteligência Artificial, Redes Neurais Artificiais.

# Abstract

The prediction of time series is already applied in several areas, such as: meteorology, financial market, telecommunications, hydrology and others. Most of the existing techniques are based on linear statistical models that have their results ruined due to the nonlinearity and complexity of most time series. Therefore, some techniques of Artificial Intelligence, such as Artificial Neural Networks, have been used as an alternative to statistical methods due to the ability to solve non-linear problems, generating better results. But for a neural network to be able to generate good results it is necessary that its topology is correctly determined, which is not a trivial task. Then, this work proposed the use of the Non-Linear Sigmoidal Regression Blocks Networks to predict the behavior of time series taking advantage, among other factors, of the ability of these networks to determine an optimal architecture automatically during your training. The results obtained in the prediction of data from three time series were very interesting, generating, in almost all the test scenarios, a smaller MAPE with the use of the proposed technique in relation to the statistical techniques of Linear Regression and the Holt-Winters method that were used for the validation of performance.

**Key Words:** Time Series, Prediction, Statistics, Artificial Intelligence, Artificial Neural Networks.

# Sumário

<b>Índice de Figuras .....</b>	<b>x</b>
<b>Índice de Tabelas .....</b>	<b>xii</b>
<b>Tabela de Símbolos e Siglas .....</b>	<b>xiii</b>
<b>Capítulo 1 Introdução .....</b>	<b>1</b>
1.1 Motivação .....	1
1.2 Objetivos .....	3
1.2.1 Objetivo Geral .....	3
1.2.2 Objetivos Específicos .....	3
1.3 Estrutura do Documento .....	4
<b>Capítulo 2 Fundamentação Teórica .....</b>	<b>5</b>
2.1 Séries Temporais .....	5
2.1.1 Decomposição Clássica .....	7
2.1.2 Autocorrelação .....	8
2.2 Redes Neurais Artificiais .....	9
2.2.1 O Neurônio .....	10
2.2.2 <i>Perceptron</i> .....	12
2.2.3 ADALINE .....	13
2.2.4 Redes MLP .....	13
2.2.5 Redes NSRBN .....	15
<b>Capítulo 3 Materiais e Métodos .....</b>	<b>21</b>
3.1 Base de Dados .....	21

3.1.1	Dados de Vazão na Hidrelétrica de Sobradinho.....	21
3.1.2	Precipitação na Cidade do Recife .....	22
3.1.3	Variação Cambial do Real Perante Dólar.....	23
3.2	Pré-processamento dos Dados .....	24
3.2.1	Normalização dos Dados .....	24
3.2.2	Defasagem dos Dados.....	25
3.2.3	Correlação dos Dados.....	27
3.2.4	Divisão dos Dados .....	28
3.3	Parametrização das Redes NSRBN.....	29
3.4	Técnicas Comparativas.....	30
3.5	Testes Estatísticos .....	30
<b>Capítulo 4 Experimentos e Análise de Resultados .....</b>		<b>32</b>
4.1	Experimentos: Base de Vazão na Hidrelétrica de Sobradinho .....	32
4.1.1	Análise da Série .....	33
4.1.2	Resultados da Previsão: Redes NSRBN.....	35
4.1.3	Comparativo com Técnicas Estatísticas.....	37
4.2	Experimentos: Base da Precipitação na Cidade do Recife .....	38
4.2.1	Análise da Série .....	38
4.2.2	Resultados da Previsão: Redes NSRBN.....	41
4.2.3	Comparativo com Técnicas Estatísticas.....	41
4.3	Experimentos: Base da Variação Cambial Real contra Dólar .....	44
4.3.1	Análise da Série .....	44

4.3.2 Resultados da Previsão: Redes NSRBN.....	44
4.3.3 Comparativo com Técnicas Estatísticas.....	46
<b>Capítulo 5 Considerações Finais .....</b>	<b>47</b>
5.1 Conclusões.....	47
5.2 Trabalhos futuros .....	48
<b>Referências Bibliográficas.....</b>	<b>49</b>
<b>Anexo A <i>Backpropagation</i>.....</b>	<b>52</b>
<b>Apêndice A Validação Cruzada .....</b>	<b>56</b>

# Índice de Figuras

Figura 1 – Série mensal da produção de veículos automotores no Brasil de janeiro 1997 a dezembro de 2014.....	6
Figura 2 – Padrões cíclicos, sazonais e irregulares em séries temporais .....	8
Figura 3 – Neurônio biológico.....	10
Figura 4 – Neurônio de McCulloch e Pitts com limiar explícito .....	11
Figura 5 – Problema XOR. Famoso exemplo da inabilidade do <i>Perceptron</i> em resolver problemas não linearmente separáveis.....	14
Figura 6 – Estrutura de uma MLP tradicional com três camadas .....	15
Figura 7 – Arquitetura das redes NSRBN .....	17
Figura 8 – Arquitetura de um bloco polinomial .....	17
Figura 9 – Variação do EMQ com a complexidade do modelo.....	20
Figura 10 – Usina Hidrelétrica de Sobradinho.....	22
Figura 11 – Precipitação em mm/m <sup>3</sup> no mês de abril no Estado de Pernambuco.....	23
Figura 12 – Gráfico de autocorrelação da Tabela 4 com os limiares de referência...28	
Figura 13 – Série temporal de vazões diárias da hidrelétrica de sobradinho .....	34
Figura 14 – Autocorrelação da série temporal da Figura 13.....	35
Figura 15 – Previsão de vazão para um dia à frente utilizando as redes NSRBN.....	36
Figura 16 - Previsão de vazão para três dias à frente utilizando as redes NSRBN .36	
Figura 17 – Série temporal das precipitações médias mensais na cidade de Recife 39	
Figura 18 – Autocorrelação da série temporal da Figura 17.....	40

Figura 19 – Previsão da precipitação média mensal para um mês à frente utilizando as redes NSRBN .....	43
Figura 20 – Série temporal da variação do dólar perante o real.....	45
Figura 21 – Previsão do valor mensal do dólar em reais para um mês à frente utilizando as redes NSRBN .....	45
Figura 22 – Valores de uma rede com (a) <i>underfitting</i> , (b) <i>overfitting</i> e (c) treinamento interrompido com validação cruzada .....	56
Figura 23 – Representação gráfica do método de validação cruzada.....	57

# Índice de Tabelas

Tabela 1 – Exemplo de dados originais e normalizados da precipitação média mensal dos seis primeiros meses de 2013 .....	25
Tabela 2 – Defasagem de amostras nas vazões diárias da hidrelétrica de sobradinho .....	26
Tabela 3 – Resultado da defasagem aplicada na Tabela 2 .....	27
Tabela 4 – Coeficientes de correlação dos dados da Tabela 3.....	27
Tabela 5 – Resultados da previsão de vazão utilizando a rede NSRBN.....	37
Tabela 6 – Resultados da previsão de vazão utilizando as técnicas estatísticas.....	38
Tabela 7 – Resultados da previsão da precipitação utilizando a rede NSRBN .....	41
Tabela 8 - Resultados da previsão da precipitação utilizando a regressão linear .....	42
Tabela 9 – Resultados do EMPA na previsão do valor do dólar em reais utilizando a rede NSRBN.....	46
Tabela 10 – Resultados do EMPA durante as previsões do valor do dólar em reais utilizando as técnicas estatísticas .....	46

# Tabela de Símbolos e Siglas

ADALINE – *Adaptive Linear Element*

APAC – Agência Pernambucana de Águas e Clima

BDMEP – Banco de Dados Meteorológicos para Ensino e Pesquisa

CHESF – Companhia Hidrelétrica do São Francisco

EMQ – Erro médio quadrático

EPMA – Erro percentual médio absoluto

GMDH – *Group Method of Data Handling*

INMET – Instituto Nacional de Meteorologia

MLP – *Multilayer Perceptron*

NSRBN – *Non-Linear Sigmoidal Regression Blocks Networks*

RNA – Rede Neural Artificial

ST – Séries Temporais

# Capítulo 1

## Introdução

Este capítulo é dedicado à apresentação da monografia. Inicialmente é descrito a motivação pela qual o tema proposto foi escolhido. Em segundo, são retratados os objetivos traçados para com a finalização do projeto. E por fim, o detalhamento da estruturação dos demais capítulos desta monografia.

### 1.1 Motivação

A ideia de prever o futuro sempre foi um tema que atraiu a humanidade. A ideia de antever possíveis eventos antes que os primeiros sinais ocorram pode permitir que o homem se prepare melhor para este fenômeno, permitindo um melhor proveito dos benefícios ou se preparando para adversidades, dependendo da natureza do evento. A grande essência da previsão de eventos futuros é permitir o melhor planejamento para a tomada de decisões de acordo com os possíveis efeitos gerados pelo acontecimento, seja este favorável ou não.

Imaginando, por exemplo, de um gestor de uma linha de produção de uma fábrica de brinquedos no período que se aproxima do natal. Ele precisa definir a quantidade que será produzida de cada um dos itens que fazem parte do seu portfólio. Porém, antes de tomar essa decisão, é preciso analisar uma série de fatores, a tendência do mercado, o momento da economia, os custos de fabricação, e outros. Como estes fatores mudam dinamicamente qual a melhor saída para que o gestor tome sua decisão? Uma saída para esse tipo de problema que é constantemente empregada é tomar como base os números de anos anteriores, já que existe uma possível tendência que os números fiquem próximos aos passados. Para constatar que essa é a melhor saída, o gestor precisa estudar o comportamento da variável de interesse, a quantidade de itens fabricados, ao longo do tempo. A análise de séries

temporais contempla essa situação, “observar e extrair conhecimento sobre a movimentação recente de medidas e perceber fatores que interferem sobre eles” [1].

As Séries Temporais foram definidas como “uma sequência de dados quantitativos relativos a momentos específicos e estudados segundo sua distribuição no tempo” [1]. Diante desta definição, pode-se perceber a grande importância da análise e estudo das séries temporais e o quanto benéfico pode vir a ser utilizá-lo em aplicações cuja problemática assim o permita. A pesquisa e desenvolvimento de técnicas capazes de analisar e prever séries temporais já são realizadas há algumas décadas [2] trazendo grandes benefícios e possibilitando a criação de ferramentas cujo uso foram efetivamente aplicadas nos mais diferentes campos, tais como: medicina [3], meteorologia [4], controle industrial [5], mercado financeiro [6], telecomunicações [7] e eletrotécnica [8].

A maioria das aplicações desenvolvidas que permitem a predição de valores futuros a partir da análise de séries temporais utilizam modelos estatísticos lineares, como os modelos de Box e Jenkins, onde valores existentes são utilizados para obter informações sobre a evolução da série, permitindo a predição através de uma função de aproximação [9]. Porém, o desempenho apresentado pelos métodos estatísticos clássicos são impactados por conta da não-linearidade e complexidade da maioria das séries temporais. Por conta disso, alguns métodos de inteligência artificial, em especial as Redes Neurais Artificiais, passaram a ser utilizados como abordagem alternativa na solução de problemas de predição de séries temporais, por conta de sua capacidade de resolução de problemas não lineares [10].

O uso das redes neurais artificiais no campo de pesquisa da predição de séries temporais têm mostrado bons resultados pois constituem uma técnica universal de regressão não-linear, que permitem resolver problemas que envolvem relacionamento complexo entre variáveis de entrada e de saída sem necessitar o conhecimento prévio de relações matemáticas entre estas [10]. Entretanto, para que uma rede neural artificial consiga realizar a aproximação de determinada função de forma correta e, ao mesmo tempo, possuir capacidade de generalização, deve-se estabelecer sua arquitetura coerente visando o problema que a mesma será aplicada [11]. Essa tarefa não é algo trivial e não existem procedimentos pré-definidos para fazê-lo. Uma forma

de contornar esse problema é através do uso de Redes Neurais com princípios cibernéticos, também chamadas redes construtivas. Este modelo utiliza abordagens híbridas, de redes neurais e conceitos estatísticos, com o objetivo de estimar uma arquitetura ótima de forma automática durante o processo de treinamento da rede [10].

Diante dos fatos expostos, este trabalho propõe o uso de uma rede neural artificial construtiva, mais precisamente as *Non-Linear Sigmoidal Regression Blocks Networks (NSRBN)*, para resolver problemas de predição em séries temporais.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

Este trabalho tem como objetivo principal aplicar o uso de uma rede neural construtiva para resolver problemas de predição em séries temporais e mostrar as vantagens que tornam o uso desse tipo de rede tão interessante. Para atingir tal objetivo será realizando a implementação do algoritmo das redes neurais *NSRBN* e em seguida, avaliar seu desempenho ao aplica-la na predição de valores em diferentes bases de dados de séries temporais.

### 1.2.2 Objetivos Específicos

Como objetivos específicos para o desenvolvimento deste trabalho, têm-se:

- Estudar e compreender a problemática da previsão de dados em séries temporais;
- Compreender as definições das redes neurais com princípios cibernéticos e como elas podem ser utilizadas para determinar uma arquitetura ótima para resolver problemas de previsão de series temporais;
- Implementar o algoritmo da rede neural construtiva *NSRBN*;
- Avaliar o desempenho dos resultados obtidos ao utilizar a rede neural escolhida na resolução de problemas reais que envolvem a predição de valores de series temporais;

## 1.3 Estrutura do Documento

Este documento está dividido em cinco capítulos: Introdução, Fundamentação Teórica, Materiais e Métodos, Experimentos e Análise de Resultados e Considerações Finais.

Este primeiro capítulo, contém as informações introdutórias sobre o trabalho onde foi descrito a motivação e objetivos para com a conclusão do trabalho além da estrutura do documento.

No segundo capítulo, está exposto a fundamentação teórica que contém todos os conceitos que foram necessários para a resolução do problema ao qual este trabalho se propõe.

No capítulo seguinte, o terceiro, os materiais utilizados e métodos necessários para a realização do trabalho foram relatados.

O quarto capítulo concentra todos os experimentos realizados e resultados obtidos com a conclusão dos objetivos deste trabalho.

Por fim, o quinto e último capítulo, contém as considerações finais sobre o desenvolvimento desta monografia e dos resultados que foram alcançados com a sua conclusão além de levantar possíveis pontos abordados em trabalhos futuros.

# Capítulo 2

## Fundamentação Teórica

Este capítulo tem como objetivo fundamentar os conceitos teóricos dos principais temas de estudo que foram alvo de estudo para o desenvolvimento desta monografia. Inicialmente é visto os princípios das Séries Temporais, objeto alvo do problema apresentado neste trabalho. E as Redes Neurais Artificiais, ferramenta proposta para a resolução do problema citado.

### 2.1 Séries Temporais

As Séries Temporais podem ser definidas como um conjunto de observações, sobre determinado fenômeno, feitas em sequência ao longo do tempo [1]. Podemos citar diversos exemplos de séries temporais que estão presentes de forma constante no mundo atual e que cujo estudo são de grande importância para a área em que a mesma se relaciona, podemos citar exemplos: taxa de desemprego mensal e preços diários de ações na Economia, análise de eletrocardiograma e eletroencefalograma na Medicina, estudo do número de novos casos anuais de gripe na Epidemiologia, temperaturas máximas e mínimas diárias na Meteorologia, dentre outros [2].

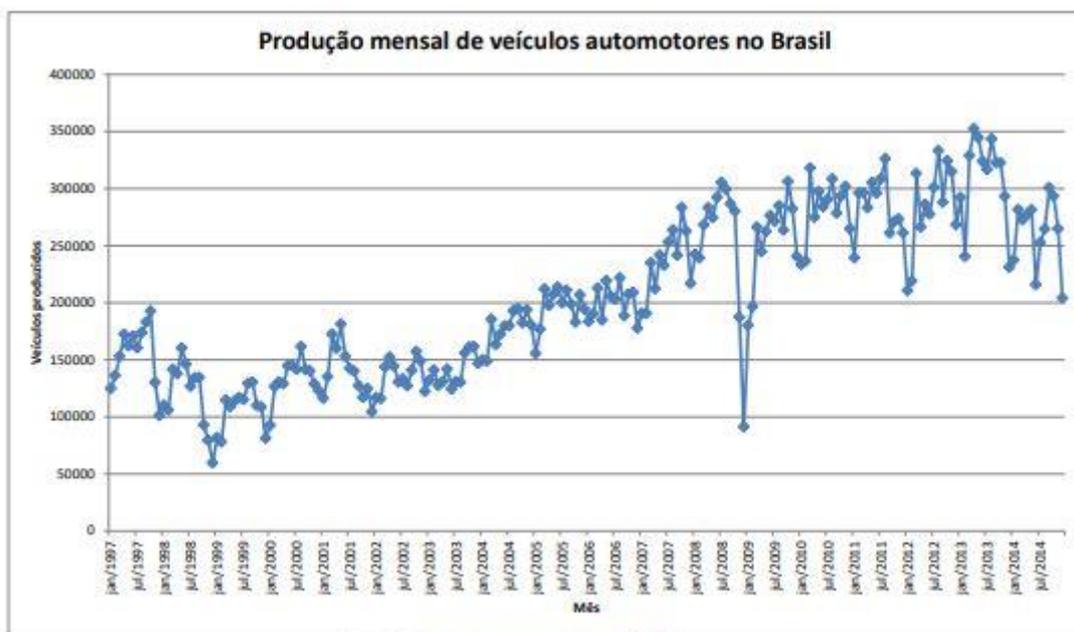
A principal característica presente em dados que compõem as séries temporais e que direciona o seu estudo é que existe um sistema causal, relativamente constante, que se relaciona com o tempo de modo que exerce algum tipo de influência sobre as amostras do passado e que pode continuar a exercer no futuro. Esse fenômeno costuma dar origem a padrões não aleatórios que podem ser detectados nas séries temporais através de técnicas descritivas, normalmente estatísticas.

A partir disso, pode-se entender como o principal objetivo do estudo e análise das séries temporais como sendo “identificar de padrões não aleatórios existentes em determinada variável de interesse de modo que a observação do seu comportamento

no passado possa permitir fazer previsões sobre futuro, direcionando a tomada de decisões” [1].

Durante a análise de uma ou mais séries temporais a representação das suas amostras através de gráficos de forma sequencial ao longo do tempo pode revelar padrões de comportamento importantes e ser fundamental para a detecção de padrões futuros. Podemos tomar como exemplo a figura a seguir:

Figura 1 – Série mensal da produção de veículos automotores no Brasil de janeiro 1997 a dezembro de 2014



Fonte: Adaptado de [12].

Na Figura 1 podemos observar alguns padrões presentes na série apenas com o auxílio da descrição gráfica. Existe uma tendência crescente no número de veículos produzidos com pequenas variações ao longo dos meses do ano, além de uma queda fora do padrão no fim de 2008 e início de 2009, por conta da crise econômica mundial vivida naquele período.

Como podemos ver através do exemplo anterior, analisar uma série temporal não é uma tarefa trivial, existem diversos fatores que precisam ser percebidos e levados em conta durante o processo. A ideia fundamental para resolver esse problema é utilizar algum modelo de análise que permita decompor a série em

padrões, possibilitando compreender o mecanismo gerador da série e posteriormente realizar previsões de possíveis valores futuros [1].

### 2.1.1 Decomposição Clássica

De acordo com o modelo clássico, várias propriedades comportamentais existentes em uma série temporal  $X_t$  podem ser descritas através da seguinte decomposição:

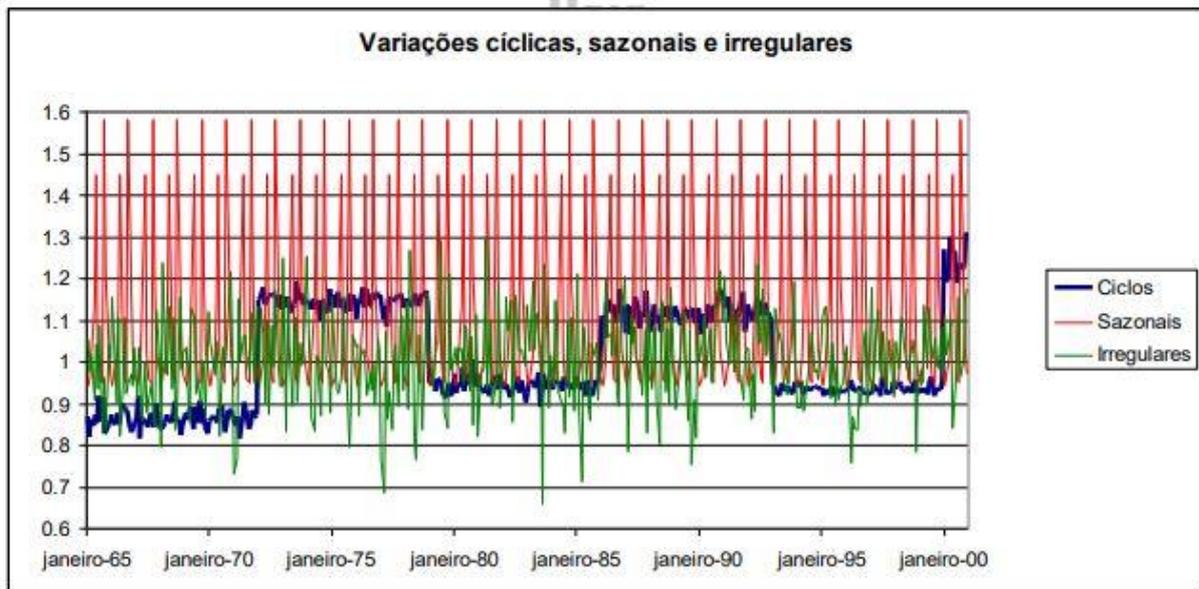
$$X_t = T_t + C_t + R_t \quad (2.1)$$

Onde  $T_t$  é uma componente de tendência,  $C_t$  uma componente que representa variações cíclicas ou sazonais e  $R_t$  é uma componente aleatória ou ruído [13]. Cada uma das componentes presentes na série pode ser entendida como se resume a seguir:

- i. **Tendência:** Comportamento de longo prazo da série, que pode ser causada por algum fenômeno, natural ou não, que afete a variável de interesse no longo prazo. Alguns exemplos desse tipo de comportamento que podem ser citados são: crescimento demográfico e mudanças de hábitos de consumo da poluição.
- ii. **Ciclos ou variações cíclicas:** São variações nos valores da variável de interesse com uma maior duração, que se repetem com certa periodicidade. Exemplos desse tipo de comportamento são: alterações na economia durante algum período de crescimento ou recessão e desastres naturais que se repetem com certa frequência em certos anos, normalmente no inverno.
- iii. **Sazonalidade ou variações sazonais:** São variações nos valores da variável de interesse com duração inferior a um ano e que se repetem todos os anos. Esse comportamento da série só irá ficar visível caso o período de amostragem dos dados seja compatível com a repetição identificada. Podemos citar alguns exemplos de séries com sazonalidade, como: datas festivas ou feriados que influenciam a economia de determinado local, como o natal e mudanças climáticas devido as estações do ano que influenciam na medição das temperaturas médias, por exemplo.
- iv. **Variações irregulares:** São variações que não tem explicação direta, normalmente são derivadas de fatos que não estão sendo esperados. Exemplos

desses acontecimentos são: desastres naturais sem periodicidade como terremotos e guerras.

Figura 2 – Padrões cíclicos, sazonais e irregulares em séries temporais



Fonte: Adaptado de [14].

É importante destacar que nem sempre as amostras de uma série temporal irão apresentar todos os padrões descritos. Uma vez que a série temporal seja decomposta, será possível identificar quais padrões comportamentais estão atuando nas amostras de interesse, além de permitir obter equações que as relacione de modo que seja possível realizar previsões sobre dados futuros [13].

### 2.1.2 Autocorrelação

A autocorrelação é uma importante técnica que permite identificar propriedades de uma série temporal. Esse método consiste em determinar um conjunto de valores, chamados de coeficientes de correlação amostral, que podem ser utilizados para conhecer a relação entre as observações atuais e passadas de uma variável de interesse. O objetivo do seu uso é medir o nível de correlação de uma observação de determinada variável em diferentes horizontes de tempo, ou seja, medir as correlações entre observações defasadas em períodos de tempo [1].

Assim, dadas  $n$  observações  $x_1, \dots, x_{n-1}$  de uma série temporal podemos calcular a correlação entre observações defasadas  $k$  períodos de tempo através da seguinte equação:

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

Onde  $\bar{x}$  é a média e  $n$  a quantidade de todas as observações da série temporal.

## 2.2 Redes Neurais Artificiais

O cérebro humano possui aproximadamente dez bilhões de neurônios que se conectam uns com os outros formando uma espécie de rede capaz de processar uma grande quantidade de informações e executar em torno de 50 trilhões de ligações sinápticas [10]. Por ser um sistema de processamento altamente paralelo, o cérebro humano possui a capacidade de processar uma grande quantidade de informações de uma maneira extremamente rápida, mesmo sendo composto por neurônios lentos, individualmente [11].

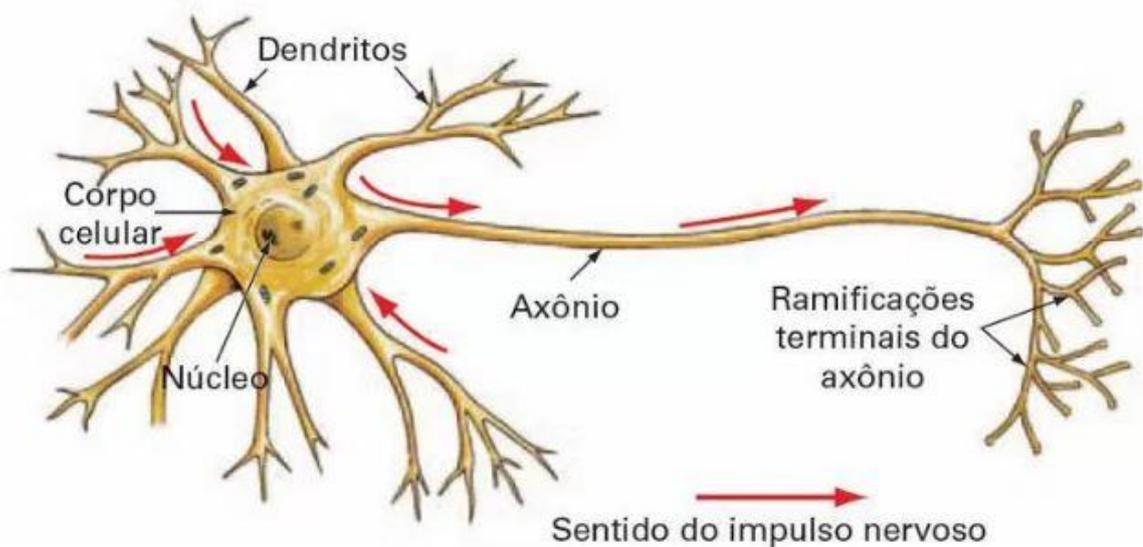
Diante disso, as RNAs surgiram como modelos que se baseiam no cérebro humano e nas suas redes neurais biológicas e sua capacidade de processar informações [10]. De maneira geral, as RNAs podem ser definidas como “um sistema composto por unidades de processamento interconectados, os neurônios, que ao serem agrupados em camadas formam uma espécie de rede, sendo responsáveis pela capacidade de memória e não linearidade da estrutura” [15].

A primeira ideia de um neurônio artificial surgiu em 1943 a partir de um estudo proposto pelo neurofisiologista McCulloch e do matemático Pitts [11] que sugeriu a construção de uma máquina inspirada no cérebro humano. Devido a este estudo, os pesquisadores ficaram conhecidos como os responsáveis pelo desenvolvimento da fundamentação da neurocomputação [10]. A partir disto, vários estudiosos continuaram a pesquisar o tema, desenvolvendo seus conceitos e contribuindo com a evolução da área. Os principais estudos, que deram origem a modelos que são muito conhecidos atualmente, serão descritos nos tópicos seguintes.

### 2.2.1 O Neurônio

O neurônio é a unidade básica do cérebro humano, tem como principal atividade a transmissão de informações, uma vez que possui propriedades de excitabilidade e condução de transmissões nervosas. O neurônio é formado por três partes fundamentais: um corpo celular, de onde saem algumas ramificações chamadas de dendritos, por uma outra ramificação mais extensa originária do mesmo corpo celular, chamado de axônio e pelos terminais nervosos, localizados nas extremidades dos axônios, que são responsáveis pela transmissão de informações para outros neurônios. A transmissão de informação entre neurônios é definida como sinapse e ocorre sempre do axônio de um neurônio para os dendritos do neurônio seguinte [16].

Figura 3 – Neurônio biológico



Fonte: Adaptado de [16].

Um fator importante para o entendimento do funcionamento do neurônio biológico está relacionado no modo em que as sinapses ocorrem nos mesmos. A intensidade do estímulo realizado entre os neurônios determina se o impulso nervoso irá ou não ser disparado, esta intensidade é chamada de limiar excitatório [16]. Se o estímulo for muito fraco e sua intensidade inferior ao limiar excitatório não acontecerá o impulso nervoso, se o impulso for superior ao limiar, a ação do neurônio sempre será a mesma. Esse comportamento ficou conhecido como a Lei do Tudo ou Nada [11].

A partir da estrutura e funcionamento do neurônio biológico os pesquisadores desenvolveram uma primeira forma de neurônio artificial, chamado de neurônio matemático [11]. Como dito anteriormente, este primeiro modelo foi realizada por McCulloch e Pitts em 1943.

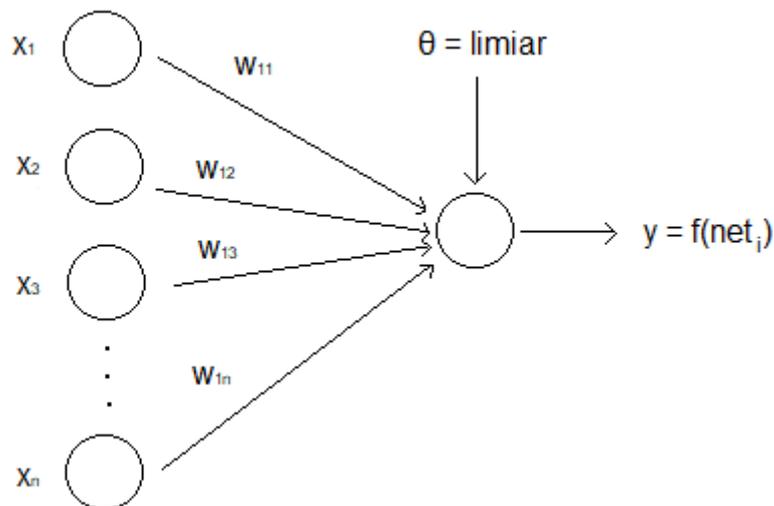
O modelo desenvolvido para o neurônio matemático, procura representar o neurônio biológico através de uma regra de propagação e de uma função de ativação. A transmissão do sinal através desse modelo de neurônio acontece com um limiar de forma análoga à Lei do Tudo ou Nada [11].

Seja  $x_1, x_2, x_3 \dots, x_n$ , como as variáveis de entrada do neurônio de saída  $i$ . A entrada líquida  $net_i$  é dada pela seguinte regra de propagação:

$$net_i = \sum_{j=1}^n w_{ij}x_j - \theta \quad (2.3)$$

Onde  $w_{ij}$  são os pesos sinápticos e  $\theta$  o limiar.

Figura 4 – Neurônio de McCulloch e Pitts com limiar explícito



Fonte: Adaptado de [11]

Baseado no neurônio biológico, a função de ativação é do tipo degrau, logo:

$$f(net_i) = \begin{cases} 1, \forall net_i \geq 0 \\ 0, \forall net_i < 0 \end{cases} \quad (2.4)$$

### 2.2.2 Perceptron

A primeira aplicação prática das RNAs foi através do *Perceptron*, desenvolvido por Frank Rosenblatt em 1958 [11]. Esta rede neural foi capaz de realizar reconhecimento de padrões através de uma regra de aprendizagem que usou como fundamentação a proposta do psicólogo Donald Hebb, “Quando um neurônio recebe um estímulo de outro, e ambos estão altamente ativos, o peso entre estes deve ser fortalecido, caso contrário enfraquecido” [11]. A partir disso, esta rede neural adquiriu uma importante característica existente no cérebro humano, a capacidade de aprender, o que gerou um grande interesse nas redes neurais artificiais pelos estudiosos da época.

O *Perceptron* é uma rede neural simples e segue o modelo de McCulloch e Pitts Figura 4, é composto por uma camada de entrada e uma camada de saída, onde cada entrada está ligada à saída através de uma conexão ponderada, chamada de peso sináptico, assim como o neurônio biológico [11]. O ajuste dos pesos sinápticos, inicialmente com valores aleatórios, é realizado durante o processo de treinamento através dos seguintes passos:

1. Verifica-se se o treinamento deve continuar a partir de critérios de seleção pré-definidos;
2. Um exemplo que represente o problema em questão é apresentado à rede;
3. A partir do exemplo, calcula-se a saída da rede utilizando as equações (2.3) e (2.4);
4. Caso o resultado esteja correto, volta para 1;
5. Caso contrário, atualiza-se os pesos através da seguinte expressão:

$$\Delta w_{ij} = \mu x_i (d_j - y_j) \quad (2.5)$$

Onde,  $\Delta w_{ij}$  é o ajuste a ser realizado nos pesos,  $\mu$  a taxa de aprendizagem,  $x_i$  o exemplo apresentado à rede,  $d_j$  o valor desejado e  $y_j$  a saída calculada da rede.

### 2.2.3 ADALINE

A rede ADALINE, proposta por Widrow e Hoff em 1960, é uma extensão do *Perceptron* que permite o uso de uma função contínua para representar a função de saída do neurônio da camada de saída [10].

Dentre as funções mais utilizadas estão:

$$\text{Linear} \quad f(\text{net}) = \text{net} \quad (2.6)$$

$$\text{Sigmoide logística} \quad f(\text{net}) = \frac{1}{1 + e^{-\text{net}}} \quad (2.7)$$

$$\text{Tangente hiperbólica} \quad f(\text{net}) = \frac{e^{\text{net}} - e^{-\text{net}}}{e^{\text{net}} + e^{-\text{net}}} \quad (2.8)$$

O treinamento da rede ADALINE é realizado através da regra delta que, por sua vez, é baseado no método dos mínimos quadrados. Esta regra determina o conjunto de pesos ótimos pelo uso do método do gradiente descendente, conhecida técnica de otimização que busca minimizar o erro médio quadrático durante o processo de treinamento [10]. A regra de aprendizagem da rede ADALINE é dada pela seguinte expressão:

$$\Delta w_{ij} = \alpha(d_i - y_i) x_i f'(net_i) \quad (2.9)$$

Assim, o novo valor para peso é dado por:

$$w_{ij}(\text{novo}) = w_{ij}(\text{antigo}) + \alpha(d_i - y_i) x_i f'(net_i) \quad (2.10)$$

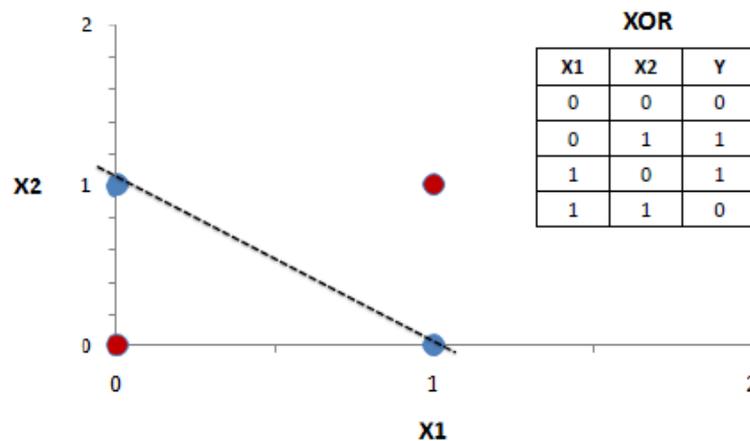
Sendo  $w_{ij}$  é o peso sináptico,  $\alpha$  a taxa de aprendizagem,  $d_i$  o valor desejado,  $y_i$  a saída calculada pela rede,  $x_j$  a variável de entrada,  $f'(net_i)$  a derivada da função de saída com relação a entrada líquida e  $i$  o índice do neurônio conectado ao peso.

### 2.2.4 Redes MLP

Com o desenvolvimento da rede *Perceptron* e de sua extensão, a rede ADALINE, as redes neurais artificiais atraíram o interesse dos pesquisadores por serem capazes de resolver problemas de classificação utilizando a sua capacidade de adquirir conhecimento através de uma regra de aprendizagem [10]. Com a

publicação de Minsky e Papert em 1969, revelou-se através de um simples exemplo, como o problema XOR, que as redes baseadas no *Perceptron* conseguiram resolver apenas problemas linearmente separáveis. Essa constatação acarretou fortes críticas ao modelo, levando à perda de interesse e corte de verba para o financiamento de pesquisas [11].

Figura 5 – Problema XOR. Famoso exemplo da incapacidade do *Perceptron* em resolver problemas não linearmente separáveis



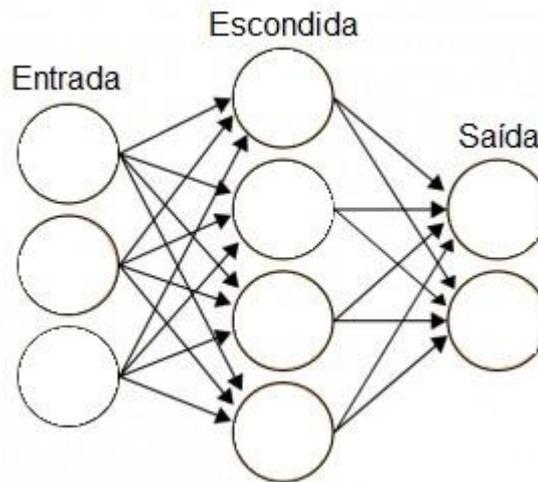
Fonte: Adaptado de [17]

O interesse pelas redes neurais ressurgiu com a publicação sobre as propriedades associativas das redes neurais pelo físico John Hopfield em 1982 e com o desenvolvimento do algoritmo de treinamento *Backpropagation*, iniciado por Paul Werbos em 1970 e popularizado com Rumelhart e McClelland em 1986 [11]. Com a retomada das pesquisas na área, foi visto que ao adicionar uma nova camada entre a entrada e a saída de uma rede *Perceptron* juntamente com o uso de um algoritmo de aprendizado apropriado, como o *Backpropagation*, a rede neural passaria a resolver problemas não-lineares, acabando com a grande restrição que estava a anos limitando o desenvolvimento das pesquisas na área de redes neurais artificiais [10].

Assim, as redes neurais *Multilayer Perceptron* se destacaram e ficaram amplamente conhecidas, sendo definidas como uma rede do tipo *feedforward*, onde o processamento da informação se dá no sentido progressivo através dos neurônios de camadas vizinhas [10]. A principal e mais importante das propriedades das redes MLP diz que com apenas uma camada escondida e uma função de ativação do tipo sigmoide ela será capaz de aproximar qualquer função contínua existente [11].

A estrutura de uma rede MLP requer que exista no mínimo três camadas, a de entrada, a intermediária ou escondida e a de saída, cada uma contendo pelo menos um neurônio [15].

Figura 6 – Estrutura de uma MLP tradicional com três camadas



Fonte: Adaptado [15]

Seguindo um mesmo padrão de funcionamento, a arquitetura de uma rede MLP pode ser generalizada para múltiplas camadas com pelo menos um neurônio cada.

### 2.2.5 Redes NSRBN

As Redes Neurais Compostas por Blocos de Regressões Sigmoideas Não-lineares, do inglês *Non-linear Sigmoidal Regression Blocks Networks*, são uma classe de redes *feedforward* de alta ordem propostas por Mêuser Valença em 1999 [10]. Essas redes possuem uma forte capacidade de mapeamento e conseguem aproximar qualquer função contínua definida em um conjunto  $R^n$  [10].

Duas importantes características presentes nas redes NSRBN merecem destaque. A primeira é que as mesmas possuem uma única camada escondida mantendo uma estrutura compacta e simples a segunda é que pelo uso de um algoritmo baseado no método de aprendizado construtivo é possível determinar uma arquitetura ótima de forma automática durante o processo de treinamento da rede [10].

A principal motivação para o desenvolvimento da arquitetura das redes NSRBN foi de desenvolver um método disciplinado que, fazendo uso da abordagem

construtiva, estabelecesse um mecanismo para incrementar o crescimento da rede e selecionasse o modelo de complexidade ótima. Ao mesmo tempo, possuindo a grande capacidade de mapeamento de uma rede neural de alta ordem com apenas uma camada escondida além de evitar o incremento explosivo do número de termos requeridos como acontece em outras redes construtivas como as do tipo GMDH [10].

Existem diferentes modelos de redes construtivas, alguns dos critérios utilizados para classifica-los, são [10]:

- i. Limite para crescimento da rede;
- ii. Conectividade dos neurônios;
- iii. Estrutura adicionada à rede durante o crescimento;
- iv. Ajuste das conexões;

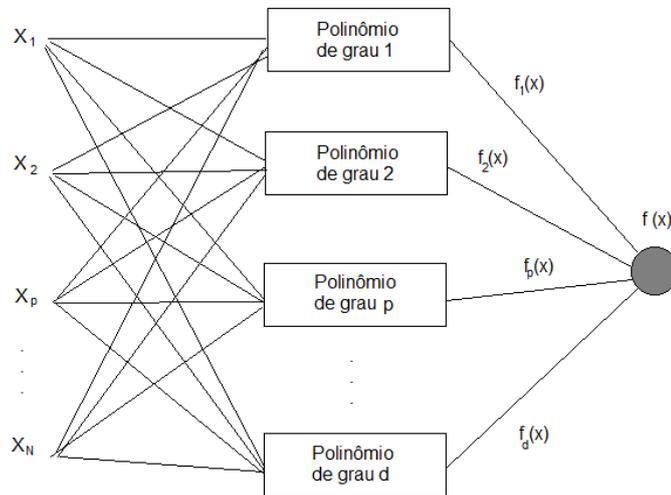
A definição da arquitetura da rede utilizando o modelo NSRBN é de construí-la utilizando blocos de polinômios homogêneos com funções de ativação específicas para cada uma das camadas. Para os neurônios da camada escondida pode ser utilizado uma função sigmoide tradicional como a tangente hiperbólica. Porém, na camada de saída, é necessário utilizar uma função de ativação mais específica, já que nesta camada os neurônios irão realizar uma regressão logística não-linear em relação com os neurônios que ficam na camada anterior [10]. Utilizando este modelo de disposição para as funções de ativação, as redes NSRBN conseguem aproximar qualquer função contínua, além de conseguirem lidar com funções *booleanas*, que não são suportadas pelas redes tradicionais presentes no grupo GMDH [10].

Uma rede NSRBN é uma rede combinatória composta pela soma de  $p$  blocos, onde  $p = 1, 2, \dots, d$  como pode ser visto na Figura 7. Em cada um dos blocos, os neurônios da camada escondida realizam uma regressão linear das entradas, assim como acontece nas redes MLP e os neurônios da camada de saída realizam uma regressão do tipo sigmoide não-linear [10]. Essa estrutura possui uma grande capacidade de mapeamento não-linear.

A estrutura de cada um dos blocos é formada por uma única camada escondida tendo  $N_h$  neurônios. Com isso, para o bloco de grau  $p = 1$  existirá apenas um neurônios na camada escondida, o bloco de grau  $p = 2$  terá 2 neurônios na camada

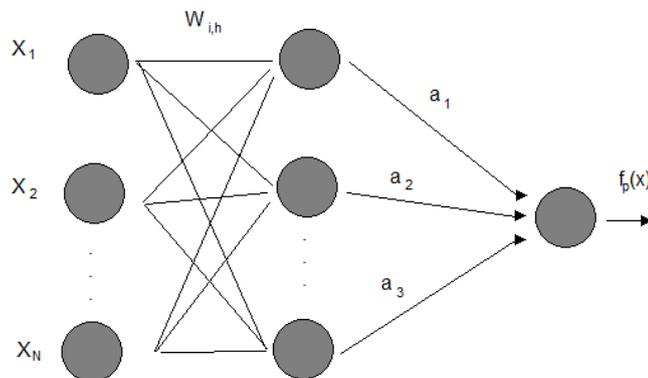
escondida, e assim por diante. Cada bloco é uma rede *feedforward* completamente conectada, como pode ser visto na Figura 8.

Figura 7 – Arquitetura das redes NSRBN



Fonte: Adaptado de [10]

Figura 8 – Arquitetura de um bloco polinomial



Fonte: Adaptado de [10]

Tomando o bloco acima como base, podemos ver que se trata de um bloco de grau  $p$ , com  $i = 1, 2, \dots, N$  variáveis de entrada e um único neurônio de saída. Sendo  $x = [1, x_1, x_2, \dots, x_n]^T$  o vetor de variáveis de entrada da rede, onde  $x_i$  representa um elemento qualquer desse conjunto. Os pesos sinápticos que ponderam as entradas são definidos como  $w = [w_{0h}, w_{1h}, w_{2h}, \dots, w_{Nh}]^T$  com  $h = 1, 2, \dots, p$ .

Podemos obter a equação da função de ativação para os neurônios da camada escondida como:

$$f_{at}(net(h)) = \sigma_{net(h)} + \theta_h \quad (2.11)$$

onde  $\sigma_{net(h)}$  representa uma função de ativação apropriada, como a tangente hiperbólica, e  $\theta_h$  a compensação do bloco anterior. A entrada líquida  $net(h)$  é definida conforme em (2.3) com uma leve adaptação, tornando-se:

$$net(h) = \sum_{i=0}^N w_{ih}x_i, h = 1, 2, \dots, p \quad (2.12)$$

Com isso, a saída dos neurônios da camada de saída, os quais realizam a regressão logística não-linear, do bloco podem ser representados pela seguinte equação:

$$f_p(x) = a_1[f_{at}(net(1))]^1 + a_2[f_{at}(net(w))]^2 + \dots + a_p[f_{at}(net(p))]^p \quad (2.13)$$

sendo  $a = [a_1, a_2, \dots, a_p]^T$  o vetor dos pesos que ponderam as ligações entre os neurônios da camada escondida e os da saída.

Conforme exposto na Figura 7 a arquitetura de uma rede NSRBN é formada pela soma dos blocos existentes. Assim, utilizando esta rede é possível aproximar uma função em  $R^N$  através da seguinte equação:

$$f(x) = \sigma_{net(0)} \sum_{p=1}^d f_p(x) \quad (2.14)$$

onde  $\sigma_{net(0)}$  é uma função de ativação adequada para o neurônio de saída.

Uma outra característica presente na rede NSRBN é de que apenas o último bloco adicionado precisa passar pelo processo de treinamento, e os anteriormente adicionados e já treinados permanecem com seus pesos fixos [10]. Esse comportamento permite o treinamento de forma rápida evitando a explosão combinatória da quantidade de neurônios de cada bloco durante o treinamento [10]. Como a construção da rede NSRBN é feita de modo incremental onde a arquitetura

ótima é definida automaticamente durante o treinamento, é possível representar a equação da aproximação da função desejada da seguinte maneira:

$$f(x) = \sigma_{net(0)} \sum_{p=1}^{d-1} f_p(x) + f_d(x) \quad (2.15)$$

No qual os blocos  $f_1, f_2, \dots, f_{d-1}$  permanecem seus pesos congelados e apenas o bloco  $f_d$  necessita ser treinado. Com isso, a função de saída pode ser escrita da seguinte forma:

$$f(x) = \sigma_{net(0)} \{ c_0 + c_1 [\sigma_{net(1)} + \theta_1]^1 + c_2 [\sigma_{net(2)} + \theta_2]^2 + \dots + c_d [\sigma_{net(d)} + \theta_d]^d \} \quad (2.16)$$

de forma que o vetor  $c = [c_1, c_2, \dots, c_d]$  representam os pesos dos neurônios da camada escondida para a camada de saída do último bloco adicionado e a constante  $c_0$  representa a soma das saídas de todos os blocos cujos pesos foram congelados.

A equação (2.16) é idêntica à da técnica estatística de regressão logística não-linear, quando a função de ativação é a sigmoide logística tradicional (2.7).

O algoritmo construtivo para as redes NSRBN tem como base um método de aprendizagem construtiva, também chamado de auto-organização na teoria das redes GMDH [10]. O princípio de funcionamento deste algoritmo pode ser definido da seguinte forma: durante o crescimento da complexidade do modelo, certos critérios de seleção são utilizados como referencial para, a partir da análise de seu valor, encontrar um valor mínimo, que irá fornecer um modelo de complexidade ótima [10].

A Figura 9 demonstra a variação da complexidade do modelo, os número de blocos das redes NSRBN, durante a execução do algoritmo construtivo em comparação a variação de um critério de seleção, como o EMQ.

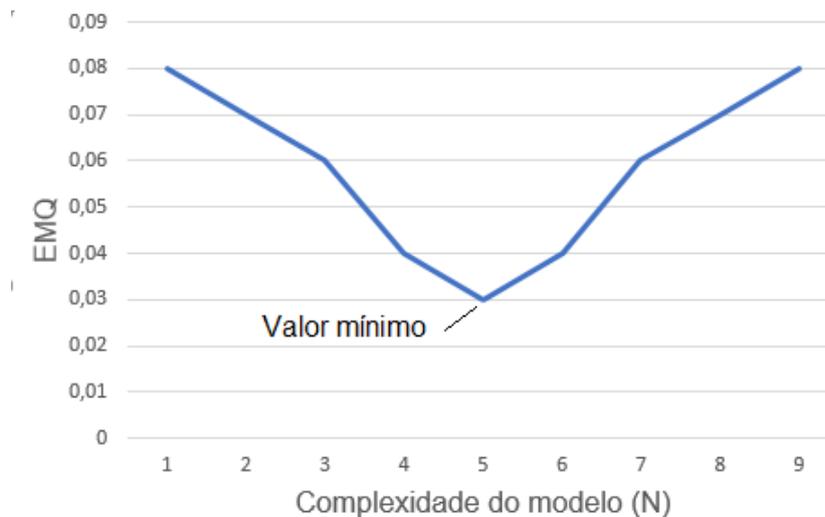
Seguinte esse modelo, o algoritmo construtivo para a definição da arquitetura das redes NSRBN é feito através dos seguintes passos:

- i. Escolha de dois critérios de seleção, um para determinar os pesos e outro para definir a complexidade. Definir outros parâmetros da rede como, número de

ciclos mínimo e máximo, limiares para critérios de seleção e para incremento da complexidade;

- ii. Adicionar um bloco à rede;
- iii. Utilizando os exemplos do problema, apresentar os exemplos à rede:
  - a. Calcular a saída da rede;
  - b. Otimizar pesos utilizando o primeiro critério de seleção;
- iv. Ao fim de cada ciclo, calcular segundo critério de seleção;
- v. Verificar critérios de para a otimização dos pesos (primeiro critério de seleção). Se o critério for atendido, ir para passo vi, caso contrário voltar iii;
- vi. Calcular valor para critério de parada, em função do segundo critério de seleção;
- vii. Caso o critério seja atendido, finaliza o treinamento e o modelo de complexidade ótima foi encontrado. Caso contrário, voltar para passo ii;

Figura 9 – Variação do EMQ com a complexidade do modelo



Fonte: Adaptado de [10]

# Capítulo 3

## Materiais e Métodos

Neste capítulo será descrito a metodologia utilizada durante o desenvolvimento do projeto abordado nesta monografia. No tópico 3.1 serão descritas as bases de dados utilizadas e as características de cada uma. Em seguida, no 3.2 estará exposto todas as operações de pré-processamento que foram feitas nos dados. No seguinte, 3.3, será detalhado todos os parâmetros existentes na rede NSRBN e os respectivos valores que foram utilizados durante os testes. No tópico 3.4 estão descritas as técnicas utilizadas para a comparação dos resultados obtidos. Por fim, o último tópico, 3.5, apresenta os testes estatísticos que foram utilizados para a comprovação dos resultados encontrados.

### 3.1 Base de Dados

Para realizar os experimentos de desempenho das técnicas de previsão abordadas neste projeto de monografia, foram escolhidas três bases de dados que representam informações de series temporais reais. As três bases de dados utilizadas são: números de vazão da hidrelétrica de sobradinho, dados da precipitação em Recife e a variação cambial do real em relação ao dólar. As seções a seguir irão descrever de forma detalhada cada uma dessas bases além de destacar pontos que justificaram sua escolha.

#### 3.1.1 Dados de Vazão na Hidrelétrica de Sobradinho

A Usina Hidrelétrica de Sobradinho está localizada nas cidades de Sobradinho e Casa Nova, na Bahia e fica a uma distância de aproximadamente 470 quilômetros do complexo hidro energético de Paulo Afonso também no estado da Bahia. A usina tem uma potência instalada de 1050 megawatts e conta com seis máquinas geradoras [18]. A hidrelétrica está posicionada no rio São Francisco a 748 quilômetros de sua

foz e possui como funções a geração de energia elétrica e de regularizar os recursos hídricos da região em momentos de estiagem.

Figura 10 – Usina Hidrelétrica de Sobradinho



Fonte: Adaptado de [18]

O reservatório de Sobradinho possui cerca de 320 quilômetros de extensão, uma superfície de espelho d'água de 4214 quilômetros quadrados e uma capacidade de armazenamento de 34,1 bilhões de metros cúbicos, assim sendo o maior lago artificial do Brasil. Com isso, a usina garante uma vazão de 2060 metros cúbicos por segundo o que permite a operação de todas as usinas da CHESF localizadas ao longo do rio São Francisco em momentos de estiagem.

Diante de tamanha importância na área de geração elétrica no país, foram utilizados para os testes de previsão dados da vazão diária da barragem captados entre os anos de 1931 e 2015.

### 3.1.2 Precipitação na Cidade do Recife

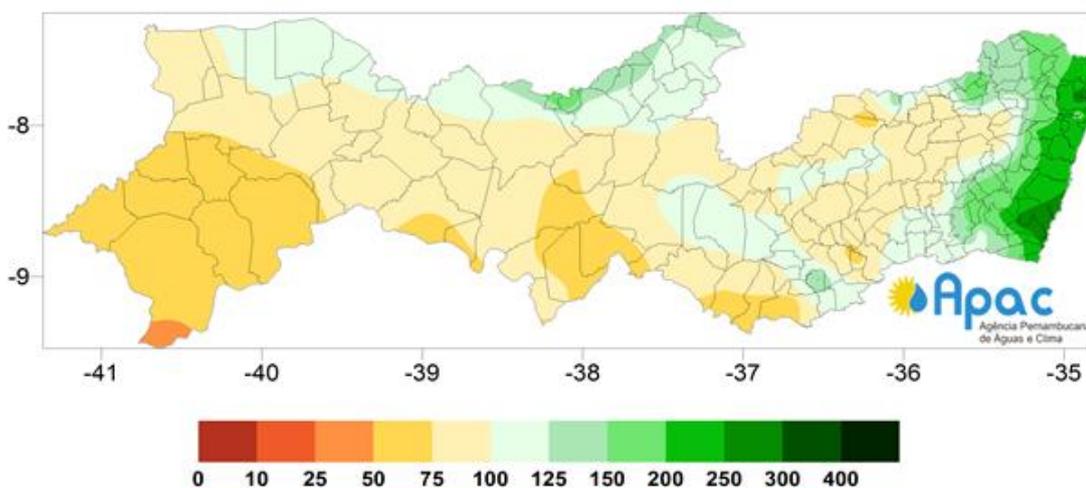
A Agencia Pernambucana de Águas e Clima é um órgão governamental que realiza uma série de estudos e monitoramentos com intuito de antecipar possíveis eventos meteorológicos no estado de Pernambuco e tomar ações que permitam as autoridades e a população se preparar da melhor forma possível [19].

As ocorrências de problemas climáticos podem causar diversos danos que são classificados pela doutrina nacional de defesa civil como humanos, materiais e ambientais [20]. Os danos humanos são aqueles onde os impactos de desastres

geram desalojados, desaparecidos, mortos ou doentes. Os materiais, por sua vez, são aqueles onde danos em edificações, instalações ou outros bens são danificados e necessitam de um valor considerável para a recuperação. E por fim, os ambientais são aqueles que geram algum problema para o meio ambiente, como poluição, desmatamento, degradação do solo ou queimadas.

Diante desses fatos fica claro que os estudos realizados são de grande importância e possibilitam que o Estado realize esforços que amenizem possíveis problemas. Um desses estudos é o monitoramento da precipitação e seu estudo permite que a APAC tome as medidas necessárias de acordo os eventos registrados.

Figura 11 – Precipitação em mm/m<sup>3</sup> no mês de abril no Estado de Pernambuco



Fonte: Adaptado de [19]

Diante da grande importância do tema visto, esse problema foi escolhido como um dos tópicos associados a previsão das series temporais que serão vistos nesta monografia. A base de dados faz parte do BDMEP fornecida pela INMET e possui a série das precipitações médias mensais registradas na cidade do Recife do ano de 1961 a 2017.

### 3.1.3 Variação Cambial do Real Perante Dólar

O dólar americano pode ser considerado como a principal e mais importante moeda do mundo, uma vez que produzida pela maior economia do mundo, os Estados Unidos [21]. A importância do dólar é tamanho que, além de ser utilizada para realizar pagamentos no seu país de origem, é tomada como referência para a maioria das

transações financeiras internacionais [22]. Por conta disso, uma vez que a maioria dos países do mundo possuem uma moeda própria torna necessário realizar uma operação de câmbio para que transações de comercio exterior sejam possíveis.

Esse tipo de transação, juntamente com outros fatores, influencia diretamente no peso internacional que uma moeda possui fazendo com que seu valor varie em relação a outras. De maneira geral, o valor de uma moeda varia de acordo com a oferta e demanda e questões políticas nacionais e internacionais.

Diante desse cenário exposto, a série que relata a variação do real perante o dólar será abordada como um dos problemas a serem analisados através da predição de series temporais. A base de dados possui o valor mensal do real em relação ao dólar comercial desde o início de sua circulação, em julho de 1994 até maio de 2018.

## 3.2 Pré-processamento dos Dados

Após definir quais bases de dados serão utilizadas, um passo importante antes de iniciar os testes é realizar o pré-processamento dessas informações. Dependendo da natureza da base de dados, nessa etapa podem ocorrer algumas operações que consistem na remoção, complementação ou transformação dos dados. Essa etapa é útil pois auxilia que as técnicas escolhidas obtenham melhores resultados durante os experimentos de predição sobre as bases de dados.

A seguir, serão descritas as operações de pré-processamento utilizadas em cada uma das três bases de dados selecionadas além da motivação pelo uso de cada uma.

### 3.2.1 Normalização dos Dados

A normalização dos dados consiste em um processo de transformação de como o dado está representado. Esse processo normalmente se resume em utilizar alguma formulação para limitar o valor do dado em uma escala desejada.

Dentre as diversas fórmulas matemáticas existentes para a normalização dos dados, a escolhida para o uso foi a transformação linear, definida pela equação a seguir:

$$y_i = (b - a) \frac{x_i - x_{min}}{x_{max} - x_{min}} + a \quad (3.1)$$

Onde  $y_i$  é o valor convertido para o dado de entrada  $x_i$ ,  $x_{min}$  e  $x_{max}$  são, respectivamente, os menores e maiores valores existentes para a variável em questão, enquanto  $a$  e  $b$  são os limites mínimos e máximos baseados na função de ativação que será utilizada na técnica de previsão.

O uso dos limites baseados na função de ativação são uteis pois permite evitar valores onde a derivada da função de ativação seja zero, esse cenário pode ocorrer durante os cálculos existentes nas redes NSRBN [11].

Tabela 1 – Exemplo de dados originais e normalizados da precipitação média mensal dos seis primeiros meses de 2013

Mês/Ano	Valor Original ( $\frac{mm}{m^2}$ )	Valor Normalizado
01/2013	95,7	-0,676
02/2013	47,7	-0,788
03/2013	98,8	-0,669
04/2013	229	-0,364
05/2013	316,9	-0,159
06/2013	491,4	0,248

Fonte: Autoria própria

Nos experimentos onde a rede neural NSRBN será utilizada, a função de ativação escolhida para a camada intermediária foi a tangente hiperbólica. Assim, com o uso dessa função resultou em valores de saída definidos no intervalo entre -1 e 1. Durante a normalização, foram utilizados os limites  $a = -0,9$  e  $b = 0,9$ .

### 3.2.2 Defasagem dos Dados

É uma operação que consiste em um processo auto regressivo sobre uma variável de interesse onde se realiza um deslocamento no tempo inserindo *lags*, de

modo que, seja possível gerar uma lista onde um determinado valor seja consequência de  $k$  valores passados, onde  $k$  representa a quantidade de *lags* ou deslocamentos realizados. A Tabela 2 é um exemplo da aplicação desse processo em uma das bases de dados utilizadas, a defasagem realizada é de  $k = 3$  :

Tabela 2 – Defasagem de amostras nas vazões diárias da hidrelétrica de sobradinho

$k = 3$	$k = 2$	$k = 1$	$k = 0$
			-0,524
		-0,524	-0,540
	-0,524	-0,540	-0,560
-0,524	-0,540	-0,560	-0,588
-0,540	-0,560	-0,588	-0,611
-0,560	-0,588	-0,611	-0,633
-0,588	-0,611	-0,633	-0,652
-0,611	-0,633	-0,652	
-0,633	-0,652		
-0,652			

Fonte: Autoria própria

Na Tabela 2, a técnica de defasagem foi aplicada em sete amostras de vazões diárias da hidrelétrica de sobradinho. Em  $k = 0$  as amostras estão ordenadas da mais antiga para a mais nova. Em  $k = 1$ , foi realizado o primeiro atraso, dessa forma todos os dados são atrasados em uma unidade, simulando um *lag* na sequência temporal. Em  $k = 2$  mais um atraso é inserido deslocando a lista em mais uma unidade. E então o último atraso, com  $k = 3$ , é realizado.

A Tabela 3 representa o resultado após a defasagem dos dados, removendo os elementos com colunas em branco, se obtém as amostras que podem ser usadas no processo de previsão servindo como entradas e saídas desejadas para as técnicas escolhidas. O número de entradas e de saídas é determinado pela quantidade de dados futuros que se deseja prever, caso a configuração escolhida seja de utilizar os três últimos valores para prever o próximo, as três primeiras colunas iriam representar as entradas e a última coluna a saída desejada.

Tabela 3 – Resultado da defasagem aplicada na Tabela 2

-0,524	-0,540	-0,560	-0,588
-0,540	-0,560	-0,588	-0,611
-0,560	-0,588	-0,611	-0,633
-0,588	-0,611	-0,633	-0,652

Fonte: Autoria própria

As configurações de defasagem para cada uma das três bases de dados utilizadas serão descritas nos experimentos realizados que estão detalhados no capítulo seguinte.

### 3.2.3 Correlação dos Dados

A correlação de dados é um processo utilizado para determinar o quanto uma variável influencia no valor futuro de uma outra [13]. Conforme descrito no tópico 2.1.2, se tratando de series temporais, a correlação pode ser realizada, além de com outras variáveis, com própria variável em análise, porém com amostras defasadas no tempo.

O processo da autocorrelação é normalmente utilizado após a defasagem dos dados para determinar quais amostras deslocadas no tempo possuem um coeficiente de correção interessante, possuindo um maior peso na previsão e justificando seu uso. Um exemplo do cálculo da autocorrelação pode ser visto na e foi realizado tendo como entrada os dados da Tabela 3.

Tabela 4 – Coeficientes de correlação dos dados da Tabela 3

$k$	Coeficiente de Autocorrelação
1	0,6018
2	0,1889
3	-0,1713

Fonte: Autoria própria

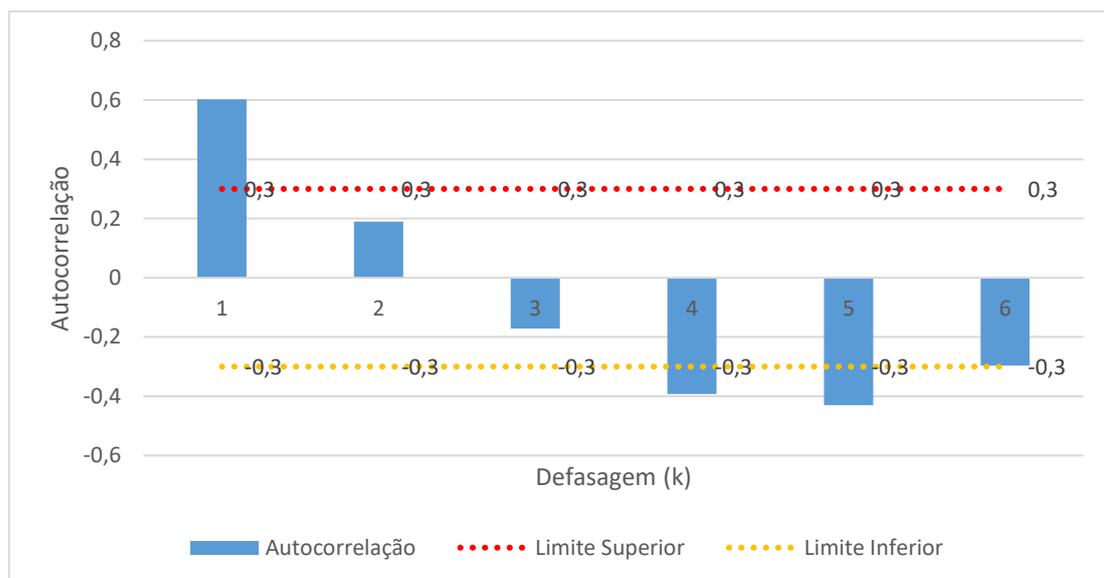
A correlação entre variáveis pode variar entre -1 e 1, onde:

- i. -1: correlação inversamente proporcional máxima;
- ii. 0: nenhuma correlação;

iii. 1: correlação proporcional máxima;

Os limiares que determinam quais variáveis serão utilizadas podem variar de acordo com a natureza da base de dados e a quantidade de amostras existentes. Para os experimentos realizados nesta monografia foram removidas as variáveis com o índice de correlação entre -0,3 e +0,3.

Figura 12 – Gráfico de autocorrelação da Tabela 4 com os limiares de referência



Fonte: Autoria própria

### 3.2.4 Divisão dos Dados

Quando se realiza um treinamento de algum algoritmo de aprendizado de máquina, geralmente se deseja obter um modelo que possua uma maior capacidade de generalização possível. Isso significa que, quando apresentado um exemplo real nunca visto previamente durante o treinamento, a resposta apresentada pela arquitetura deve ser correta.

Para se atingir esse comportamento, durante o treinamento da rede NSRBN através do Algoritmo do *Backpropagation* foi utilizado uma técnica conhecida como Validação Cruzada como critério de seleção responsável pelo momento de parada de treinamento. Os procedimentos de validação cruzada são amplamente utilizadas devido a seu fácil entendimento e implementação e bons resultados, evitando

problemas de super ajuste (*overfitting*) e sub ajuste (*underfitting*) do modelo em questão [10].

Esta abordagem costuma dividir a base de dados em três grupos com funções específicas para cada uma, são elas:

- i. Treinamento: com 50% dos dados;
- ii. Validação: com 25% dos dados;
- iii. Teste: com 25% dos dados;

Essa divisão dos dados foi inicialmente adotada para as três bases de dados utilizadas durante os experimentos quando o modelo em teste foi a rede NSRBN, porém em alguns casos específicos, devido à pouca quantidade de dados, outras porcentagens para a divisão foram utilizadas. Para mais detalhes de como a técnica de validação cruzada funciona consultar o Apêndice A.

### 3.3 Parametrização das Redes NSRBN

Uma vez terminado a preparação dos dados a próxima etapa no andamento do projeto é iniciar os experimentos configurando o modelo e apresentando os exemplos de cada uma das bases de dados.

A configuração da rede NSRBN consiste na definição de uma série de parâmetros utilizados durante seu treinamento. Uma das vantagens dessa rede, por conta da sua característica construtiva, é que não é necessário definir a quantidade de camadas escondidas e de neurônios em cada uma, conforme foi explicado no tópico 2.2.5. Os outros parâmetros existentes que precisam ser configurados são os seguintes:

- i. Primeiro critério de seleção: utilizado para atualização dos pesos durante o treinamento. O algoritmo *backpropagation* foi escolhido;
- ii. Segundo critério de seleção: utilizado para a definição da complexidade do modelo como critério de parada. O método escolhido foi a validação cruzada tendo como função objetivo o EMQ;

- iii. Função de ativação da camada escondida: a função escolhida foi a tangente hiperbólica;
- iv. Número mínimo e máximo de ciclos ou épocas: define valores mínimos e máximos para ciclos de treinamento, funcionando em conjunto com o método de validação cruzada. O valor utilizado variou de acordo com cada base de dados, o valor mínimo ficou em um intervalo de quinze a trinta ciclos e o máximo, ficou fixado em 500 para todas as bases;
- v. Semente para geração pesos sinápticos: valor utilizado para a geração aleatória dos valores iniciais dos pesos. O valor utilizado variou de acordo com a base de dados em análise;
- vi. Limiar para adição de novo bloco: valor que determina o limiar onde a adição de um novo bloco não irá trazer benefícios na definição da complexidade do modelo;

### 3.4 Técnicas Comparativas

Com o objetivo de avaliar o desempenho da predição de séries temporais utilizando a rede NSRBN, foram escolhidas outras técnicas para realizar uma análise comparativa através de testes estatísticos. As técnicas escolhidas foram a Regressão Linear e o método de previsão por suavização exponencial *Holt-Winters*.

Os experimentos com essas técnicas foram realizados utilizando o Weka, uma ferramenta que contém uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados [23].

### 3.5 Testes Estatísticos

Para avaliar os resultados obtidos durante os experimentos com a rede NSRBN em comparação com as outras abordagens selecionadas, foram realizados testes avaliando o erro encontrado com o uso de cada técnica. A avaliação do erro obtido tem como objetivo comprovar matematicamente os resultados encontrados e assim sustentando a solução proposta nesta monografia.

Para cada cenário de análise, envolvendo cada uma das três bases de dados descritas anteriormente, foi calculado o erro porcentual médio absoluto (EPMA) sobre o conjunto de testes da base. Os resultados obtidos foram calculados utilizando a seguinte equação:

$$EMPA = \frac{\sum_{i=1}^n |d_i - y_i|}{n} \cdot 100\% \quad (3.1)$$

Onde  $d_i$  representa o valor desejado,  $y_i$  o valor encontrado e  $n$  o número de simulações. O valor encontrado com o cálculo do EMPA foi utilizado para representar o desempenho final obtido com o uso de cada uma das técnicas para a previsão dos dados.

A descrição dos cenários de testes para cada uma das bases de dados e os resultados obtidos estão detalhados no capítulo seguinte.

# Capítulo 4

## Experimentos e Análise de Resultados

Este capítulo apresenta em detalhes todos os experimentos realizados e a análise dos resultados obtidos. O processo utilizado consistiu em analisar a natureza e problemática dos dados, realizar o pré-processamento dessas informações, treinar a rede NSRBN com esses dados e, em seguida, realizar uma análise do desempenho do modelo em comparação com outras técnicas estatísticas que foram submetidas aos mesmo conjuntos de dados.

Este processo foi realizado para as três bases de dados descritas no tópico 3.1 de modo que foi possível verificar os resultados da aplicação das redes construtivas NSRBN para a previsão de séries temporais que representam problemas reais onde cada série possui suas próprias características.

### 4.1 Experimentos: Base de Vazão na Hidrelétrica de Sobradinho

Esta seção está dividida em três subseções de modo que os resultados encontrados fiquem mais organizados e fáceis de compreender. O primeiro tópico refere-se a análise da série temporal a vazão diária na hidrelétrica de sobradinho. Em seguida, serão apresentados os resultados da previsão da série pelo uso da rede NSRBN. E por fim, o último tópico apresenta uma comparação dos resultados encontrados com a rede NSRBN com as técnicas estáticas propostas em 3.4.

#### 4.1.1 Análise da Série

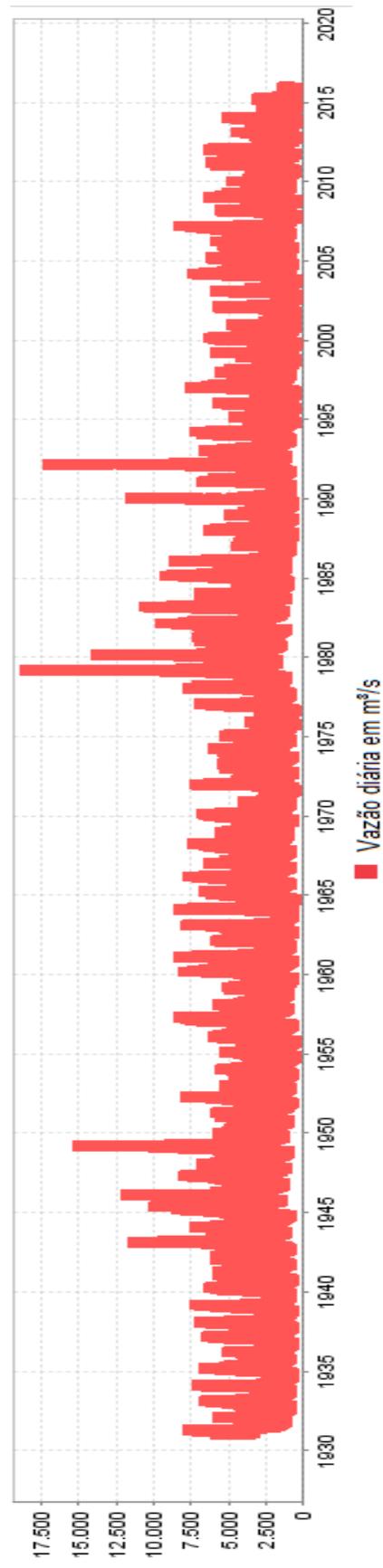
O primeiro passo antes de realizar qualquer tipo de experimento foi analisar a série temporal para compreender suas características e assim atingir melhores resultados. A Figura 13 representa a distribuição dos dados ao longo do tempo da base em questão.

Com a visualização gráfica é possível perceber alguns padrões existentes na série, a primeira é que a mesma possui sazonalidade, algo esperado em fenômenos que sofrem influência das estações do ano, como é o caso da chuva que afeta diretamente na vazão da hidrelétrica. Outro ponto é que a série possui uma certa tendência de vazão que situa no intervalo de 0 e  $7500 \text{ m}^3/\text{s}$ , com alguns picos fora desse valor mas que não chegam a representar ciclos.

O próximo passo utilizado durante a análise da série foi verificar a autocorrelação para auxiliar na estimação da quantidade de deslocamentos necessários para realizar a predição além de identificar quais descolamentos possuem um grau de correlação dentro dos valores determinados como aceitáveis para o uso no modelo.

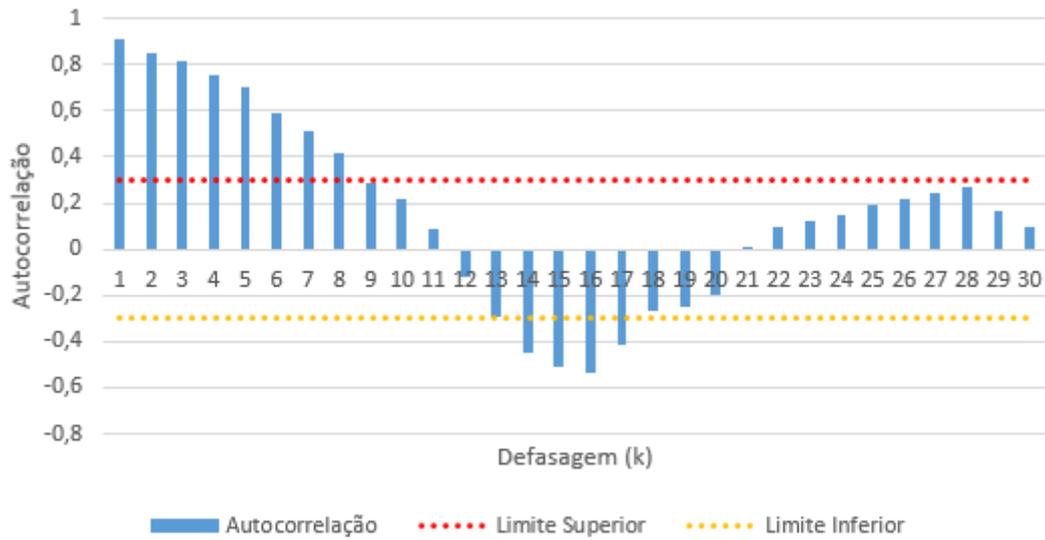
O resultado desse processo pode ser visto na Figura 14, onde podemos verificar que o grau de autocorrelação decresce constantemente de acordo com a quantidade de *lags* inseridos. Esse comportamento demonstra a tendência existente na série, nesse caso pouca informação pode ser extraída da análise da autocorrelação ao menos que a componente da tendência seja removida da série e os testes refeitos. Após este processo e os testes refeitos foi determinado uma quantidade de doze entradas para as técnicas e iniciados os experimentos.

Figura 13 – Série temporal de vazões diárias da hidrelétrica de sobradinho



Fonte: Autoria própria

Figura 14 – Autocorrelação da série temporal da Figura 13



Fonte: Autoria própria

#### 4.1.2 Resultados da Previsão: Redes NSRBN

Os experimentos com a rede NSRBN foram realizados com doze variáveis de entrada onde cada uma representa o valor da vazão registrado para os últimos doze dias. O número de saídas da rede, representando o valor da vazão previsto para os dias seguintes, foi inicialmente de um e gradativamente ampliado até que o valor do erro na previsão deixasse de ser interessante, com quatorze dias à frente.

O resultado para a previsão de um dia a frente sobre o conjunto de teste foi de um EMPA de 1,24 % e pode ser visto graficamente na Figura 15.

Em seguida, foram sendo adicionadas mais saídas a rede até o valor máximo dito anteriormente. Os resultados obtidos podem ser vistos de forma completa na Tabela 5 e graficamente, com três dias de previsão na Figura 16.

Figura 15 – Previsão de vazão para um dia à frente utilizando as redes NSRBN



Fonte 1 – Autoria própria

Figura 16 - Previsão de vazão para três dias à frente utilizando as redes NSRBN



Fonte: Autoria própria

Tabela 5 – Resultados da previsão de vazão utilizando a rede NSRBN

Previsão (dia à frente)	EMPA (%)
1	1,24
2	2,87
3	4,93
4	7,23
5	9,67
6	12,20
7	14,78
8	17,41
9	20,11
10	22,86
11	25,67
12	28,56
13	31,55
14	34,65

Fonte: Autoria própria

#### 4.1.3 Comparativo com Técnicas Estatísticas

Uma vez que os experimentos com a rede NSRBN foram finalizados, os testes foram novamente realizados utilizando as técnicas estatísticas descritas no tópico 3.5. Os dados de entrada foram os mesmos que os utilizados na rede neural e passaram pelos mesmos procedimentos de pré-processamento, porém de forma automática através do uso do Weka, isso foi feito para aumentar a confiabilidade da comparação dos resultados.

O resumo com os valores do EPMA encontrados com a previsão da vazão com o uso das técnicas estatísticas pode ser visto na Tabela 6, onde ficou constatado que o melhor resultado na previsão foi com o uso da rede NSRBN.

Tabela 6 – Resultados da previsão de vazão utilizando as técnicas estatísticas

Previsão (dia à frente)	EMPA (%)	
	Regressão Linear	<i>Holt-Winters</i>
1	9,99	17,41
2	15,88	18,63
3	21,34	19,86
4	27,40	21,07
5	34,73	22,28
6	44,01	23,49
7	56,10	24,73
8	72,18	26,05
9	93,81	27,52
10	115,31	29,24
11	132,25	31,27
12	151,72	33,65
13	181,46	44,61
14	201,14	46,52

Fonte: Autoria própria

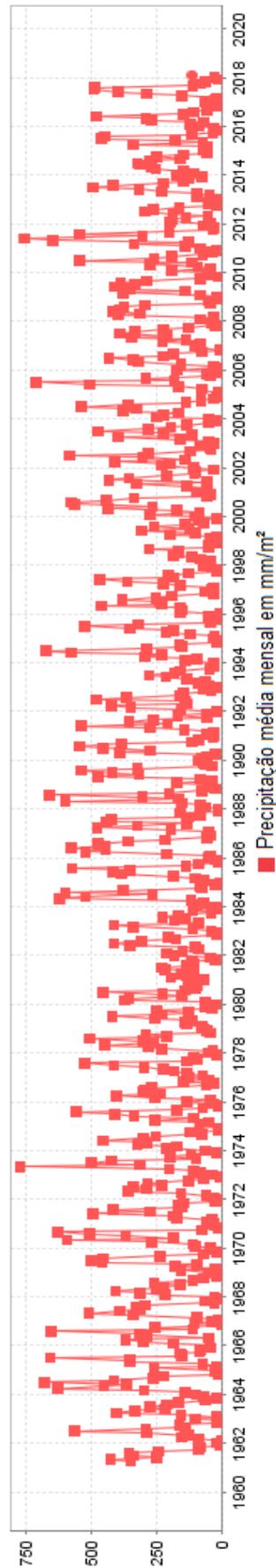
## 4.2 Experimentos: Base da Precipitação na Cidade do Recife

Esta seção expõe o processo e os resultados dos experimentos de previsão aplicados sobre a base de dados que contém a precipitação média mensal na cidade do Recife. As subseções a seguir estão dispostas assim como as do tópico 4.1.

### 4.2.1 Análise da Série

A análise da série temporal da precipitação média mensal medidos em Recife foi realizado da mesma forma que na base de dados anterior, iniciando com a análise gráfica, vista na Figura 17.

Figura 17 – Série temporal das precipitações médias mensais na cidade de Recife

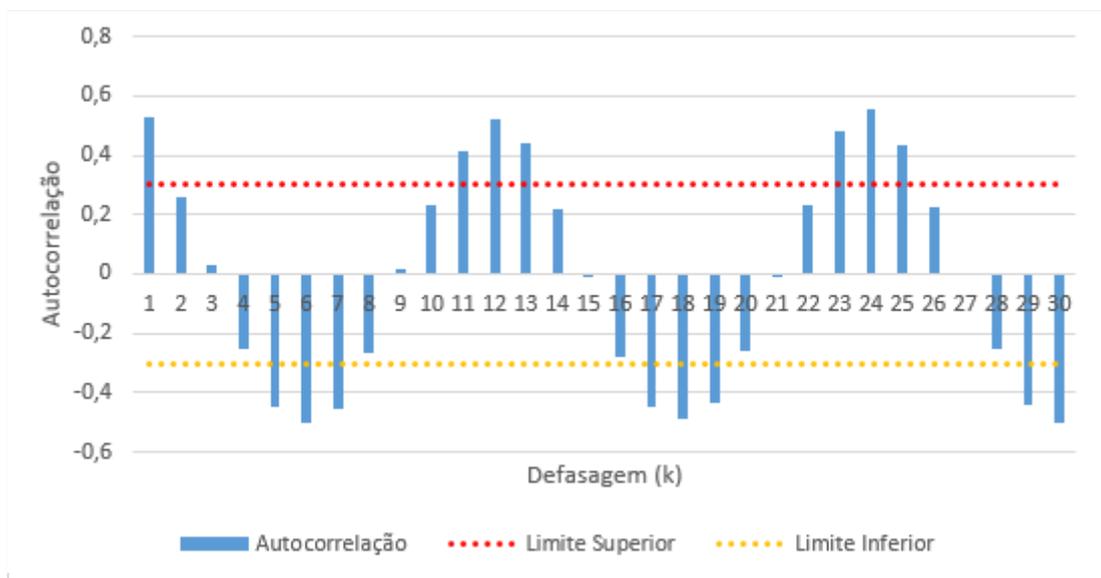


Fonte: Autoria própria

A série em questão também possui sazonalidade, algo esperado pois, assim como a base de dados anterior, se trata de um fenômeno que sofre influências das estações do ano. Na série é possível ver que existem dois grupos bem definidos onde a precipitação é média ou alta, no período de fevereiro a agosto, e o período onde a precipitação é baixa ou inexistente, no restante do ano.

Em seguida, realizando os cálculos da autocorrelação da defasagem da base podemos identificar claramente a sazonalidade da série e as variáveis que contemplam o grau mínimo de correlação desejado conforme visto na Figura 18. Visualizando o período de um ano, é possível ver que apenas seis *lags* contemplam o valor mínimo de correlação, seja este positivo ou negativo. Com isso, ficou definida a quantidade de entradas utilizadas para a rede NSRBN.

Figura 18 – Autocorrelação da série temporal da Figura 17.



Fonte: Autoria própria

Uma técnica bastante utilizada pela literatura é de representar a sazonalidade dos dados como entradas com valores binários ou com senos e cossenos, podendo gerar melhores resultados [10].

#### 4.2.2 Resultados da Previsão: Redes NSRBN

Os experimentos com a rede NSRBN seguiram o mesmo padrão realizado anteriormente, com as entradas que representam defasagens de dados prévios sendo utilizadas para a previsão de possíveis valores futuros e com isso o EMPA sendo calculado e utilizado como métrica de desempenho.

Os resultados da previsão para um mês à frente utilizando o conjunto de dados de teste resultaram em um EMPA de 5,73% e podem ser visto graficamente na Figura 19. Em seguida, foram sendo adicionadas mais saídas a rede, realizando a previsão para a precipitação média dos meses seguintes. Esse processo foi realizado até atingir a quantidade de seis saídas. A Tabela 7 contém os EMPA calculados a partir dos resultados para essas previsões.

Tabela 7 – Resultados da previsão da precipitação utilizando a rede NSRBN

Previsão (meses à frente)	1	2	3	4	5	6
EMPA (%)	5,73	6,29	6,63	7,02	7,06	7,13

Fonte: Autoria própria

#### 4.2.3 Comparativo com Técnicas Estatísticas

Seguindo mesmo processo que o realizado na base de dados anterior, foram realizadas as previsões utilizando as técnicas estatísticas para a comparação dos resultados. O resumo com os valores do EPMA encontrados com a previsão da precipitação com o uso da técnica de regressão linear e a de *Holt-Winters* pode ser visto na Tabela 8.

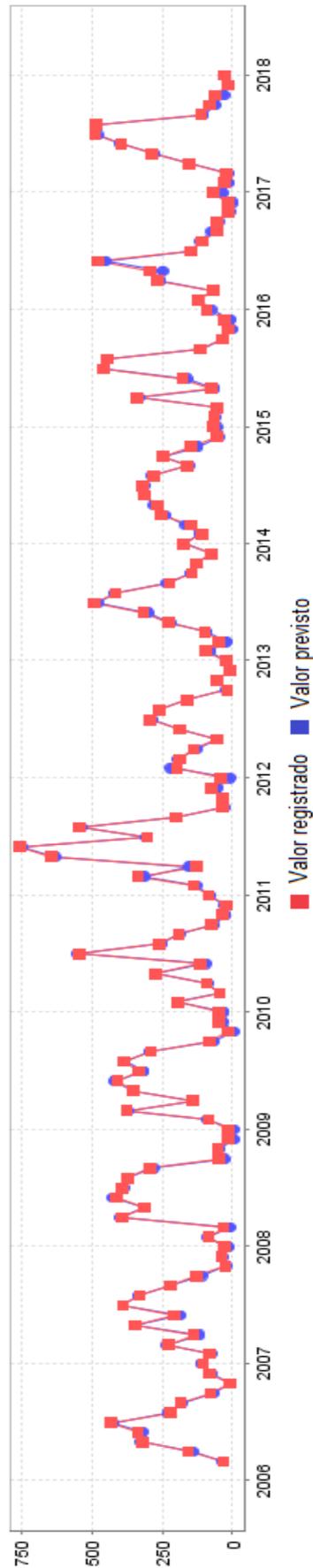
. Mais uma vez ficou constatado que os melhores resultados na previsão foram encontrados com o uso da rede NSRBN.

Tabela 8 - Resultados da previsão da precipitação utilizando a regressão linear

Previsão (mês à frente)	EMPA (%)	
	Regressão Linear	<i>Holt-Winters</i>
1	12,42	18,58
2	15,82	23,55
3	16,45	24,46
4	17,02	36,71
5	18,51	37,69
6	19,41	42,82

Fonte: Autoria própria

Figura 19 – Previsão da precipitação média mensal para um mês à frente utilizando as redes NSRBN



Fonte: Autoria própria

## 4.3 Experimentos: Base da Variação Cambial Real contra Dólar

Esta seção expõe o processo e os resultados dos experimentos de previsão da última base de dados utilizada nesta monografia, esta contém os dados da variação cambial do real perante o dólar, o período utilizado foi do ano de criação do real em 1994 a 2018. As subseções a seguir estão dispostas assim como as dos tópicos 4.1 e 4.2.

### 4.3.1 Análise da Série

Assim como feito nas outras duas bases de dados o processo de previsão se iniciou com o estudo da série temporal. A análise gráfica pode ser vista Figura 20.

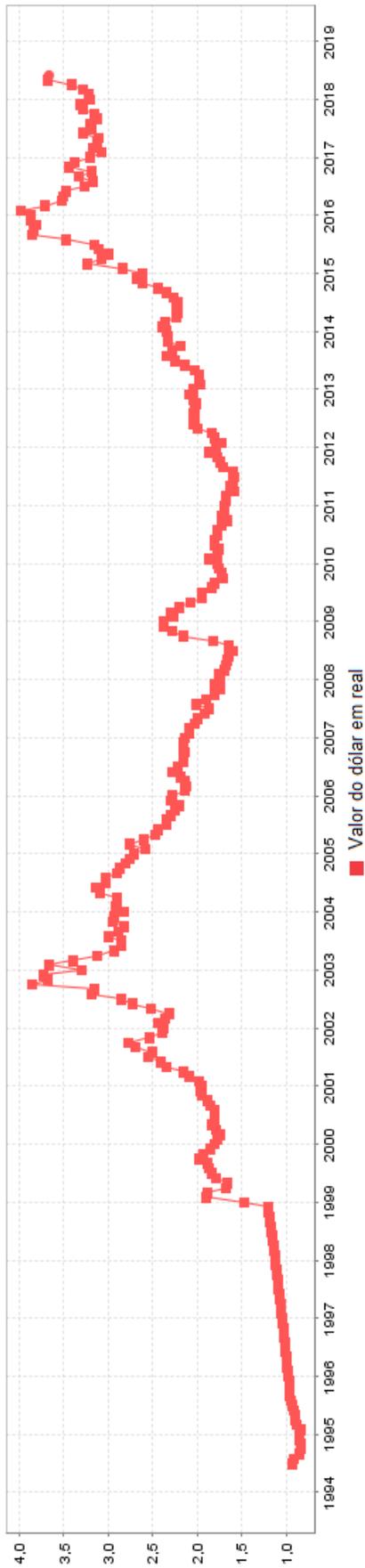
Diferentemente das outras bases vistas anteriormente, a base do valor do dólar em reais não possui sazonalidade em vez desse padrão é possível ver uma tendência inicial de crescimento no valor, porém após o ano de 1999, a série tem variabilidade acentuada com constantes mudanças de tendências de crescimento e decrescimento sendo afetada também por padrões cíclicos derivados de variações da econômica mundial como pode ser visto no ano de 2008, com a crise americana, e a partir de 2016 com a recessão econômica vivenciada no Brasil.

Diante do comportamento irregular da série foi escolhido como a quantidade de doze entradas, um valor normalmente utilizado para séries com dados de amostras mensais [1].

### 4.3.2 Resultados da Previsão: Redes NSRBN

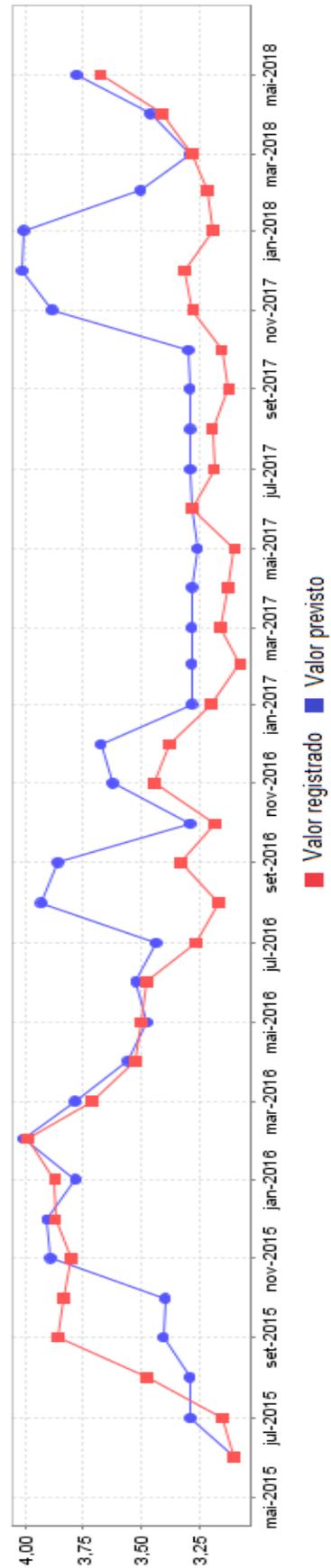
Os experimentos realizados seguiram os mesmos moldes que os vistos anteriormente. As doze entradas foram utilizadas para, inicialmente, prever o valor do dólar em reais, um mês a frente e posteriormente incrementado até a quantidade de seis meses.

Figura 20 – Série temporal da variação do dólar perante o real



Fonte: Autoria própria

Figura 21 – Previsão do valor mensal do dólar em reais para um mês à frente utilizando as redes NSRBN



Fonte: Autoria própria

Os resultados calculados com o uso da rede NSRBN podem ser vistos na Figura 21 com a previsão de um mês à frente. E na Tabela 9 com os resultados da previsão de seis meses à frente.

Tabela 9 – Resultados do EMPA na previsão do valor do dólar em reais utilizando a rede NSRBN

Previsão (meses à frente)	1	2	3	4	5	6
EMPA (%)	3,28	5,05	6,22	7,26	8,16	8,98

Fonte: Autoria própria

### 4.3.3 Comparativo com Técnicas Estatísticas

Como realizado nos experimentos com as bases anteriores, a próxima etapa foi realizar a comparação do desempenho obtido com a rede NSRBN com as técnicas estatísticas escolhidas. A métrica de comparação continuou sendo o EMPA. Os resultados do EMPA nas previsões da técnicas estatísticas propostas estão detalhadas na Tabela 10.

Diferentemente dos outros experimentos, o resultado obtido com o uso da rede NSRBN foi superior apenas a técnica de *Holt-Winters* e inferior a técnica da Regressão Linear, esse fato ocorreu possivelmente devido à pouca quantidade de dados disponíveis para o treinamento.

Tabela 10 – Resultados do EMPA durante as previsões do valor do dólar em reais utilizando as técnicas estatísticas

Previsão (mês à frente)	EMPA (%)	
	Regressão Linear	<i>Holt-Winters</i>
1	3,26	12,37
2	4,86	13,68
3	6,02	14,90
4	7,13	16,13
5	8,20	17,18
6	9,08	17,94

Fonte: Autoria própria

# Capítulo 5

## Considerações Finais

Este capítulo apresenta as considerações finais para com o desenvolvimento desta monografia e é dividido em duas seções: a 5.1 contendo as conclusões do trabalho e a 5.2 com trabalhos futuros.

### 5.1 Conclusões

Com a percepção do problema apresentado, este projeto de monografia propôs o uso das redes NSRBN para a resolução de problemas de previsão de séries temporais. Essa ideia foi fundamentada por conta dos princípios cibernéticos de auto-organização presentes nesta rede que permitem determinar de forma automática uma arquitetura ótima durante o treinamento, mostrando-se uma ferramenta poderosa devido a sua fácil utilização.

Com a conclusão desta monografia, observando os experimentos realizados, foi possível perceber que o uso das redes NSRBN foi possível realizar de forma satisfatória a previsão de séries temporais que representam problemas do mundo real como foi o caso da previsão de vazão da hidrelétrica de sobradinho, da precipitação média na cidade de Recife e a variação cambial do dólar em relação ao real. Para fundamentar os resultados, foram realizadas comparações com técnicas estatísticas clássicas onde ficou constatado um melhor desempenho da rede NSRBN em quase todos os cenários testados.

Os resultados da predição das vazões de Sobradinho, para um dia a frente, foram de 1,24% da rede NSRBN contra 9,99% e 17,41% da regressão linear e o método de *Holt-Winters*, respectivamente. De 5,73% na rede NSRBN contra 12,42% na regressão Linear e 18,58% no método de *Holt-Winters* para um mês a frente nas precipitações médias mensais em Recife. E para a previsão da relação cambial do

dólar-real para um mês a frente foi de 3,28% das redes NSRBN contra 3,26% da regressão linear e 12,37% de *Holt-Winters*, sendo este o único cenário onde os melhores resultados não foram obtidos com o uso da rede NSRBN.

## 5.2 Trabalhos futuros

Os possíveis trabalhos futuros consistem em buscar formas para melhorar os resultados obtidos na previsão das séries temporais utilizando a rede NSBN. Ideias iniciais para atingir este objetivo incluem melhorar o pré-processamento dos dados, otimizar as variáveis de entradas da rede e testar diferentes funções de ativação na camada escondida, que foi fixa em todos os experimentos realizados.

# Referências Bibliográficas

- [1] MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2<sup>a</sup>. ed. [S.I.]: Blucher, 2006.
- [2] GOOIJER, J. G. D.; HYNDMAN, R. J. 25 Years of Time Series Forecasting. **International Journal of Forecasting** **22**, p. 443-473, 2006.
- [3] WALDMAN, A. J.; PAIVA, B. C. Effectiveness of influenza vaccination and its impact on health inequalities. **International Journal of Epidemiology**, v. 36, p. 1319-26, December 2007.
- [4] SHUKLA, A. K.; GARDE, Y. A. Forecast of Weather Parameters Using Time Series Data. **Mausam**, v. 65, p. 209-520, 2014.
- [5] KATO, R.; NAGAO, T. Stock Market Prediction Based On Interrelated Time Series Data. **In Proceedings of the IEEE Symposium on Computers & Informatics**, Penang, Malaysia, 18-20 March 2012.
- [6] BERNAL, A.; FOK, S. **Financial Market Time Series Prediction with Recurrent Neural Networks**. Stanford University. Stanford , CA, USA, p. 5. 2012.
- [7] SONI, S. K.; SINGH, D. P. Reducing the Data Transmission in WSNs Using Time Series Prediction Model. **In Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing**, Hong Kong, China, 13-15 August 2012.
- [8] CEPEL. Modelo de Previsão de Vazões Semanais Aplicado ao Sistema Hidroelétrico Brasileiro - Modelo Previvaz. **Manual de Referência**.
- [9] MORGAN, B. J. T. The Cubic Logistic Model for Quantal Assay Data. **Applied Statistics**, p. 105-113.

- [10] VALENÇA, M. J. S. **Aplicando Redes Neurais: Um Guia Completo**. 1ª. ed. Olinda, PE: Livro Rápido, 2005.
- [11] VALENÇA, M. J. S. **Fundamentos das Redes Neurais**. 2ª. ed. Olinda: Do Autor, 2016.
- [12] ANFAVEA. Associação Nacional dos Fabricantes de Veículos Automotores. Disponível em: <<http://www.anfavea.com.br/tabelas.html>>. Acesso em: 07 abr. 2018.
- [13] EHLERS, R. S. **Análise de Séries Temporais**. 4ª. ed. São Paulo: [s.n.], 2007.
- [14] REIS, M. M. **Departamento de Informática e Estatística da UFSC**. Disponível em: <<http://www.inf.ufsc.br/~marcelo.menezes.reis>>. Acesso em: 10 abr. 2018.
- [15] REED, R.; MARKS, R. J. **Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks**. [S.l.]: A Bradford Book, 1999.
- [16] NETTER, F. H. **Atlas de Anatomia Humana**. 6º. ed. [S.l.]: [s.n.], 2015.
- [17] SAYAD, S. saedsayad. **An Introduction to Data Science**. Disponível em: <[http://www.saedsayad.com/artificial\\_neural\\_network\\_bkp.htm](http://www.saedsayad.com/artificial_neural_network_bkp.htm)>. Acesso em: 20 abr. 2018.
- [18] CHESF - Sistema de Geração. **Chesf**. Disponível em: <<https://www.chesf.gov.br/SistemaChesf/Pages/SistemaGeracao/Sobradinho.aspx>>. Acesso em: 02 maio 2018.
- [19] APAC - Agência Pernambucana de Águas e Clima. **Meteorologia**. Disponível em: <<http://www.apac.pe.gov.br/meteorologia/>>. Acesso em: 07 maio 2018.
- [20] ALVES, A.; RODRIGUES, L. **Manual Técnico de Defesa Civil**. Secretaria Executiva de Defesa Civil. Recife, p. 44. 2012.

- [21] INFOMONEY. **Importância do dólar e do mercado**. Acesso em: 08 maio 2018.
- [22] PORTAL IBRE. **IBRE - Instituto Brasileiro de Economia**. Disponível em: <<http://portalibre.fgv.br/>>. Acesso em: 08 maio 2018.
- [23] UNIVERSITY OF WAIKATO. Weka 3: Data Mining Software in Java. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 10 maio 2018.
- [24] MEST, I. **Generalização e Padronização**. São Paulo: [s.n.], p. 18.

# Anexo A

## ***Backpropagation***

O algoritmo de aprendizado Backpropagation foi proposto por Paul Werbos em 1970 e popularizado através de uma publicação de David Rumelhart e James McClelland em 1986 [10]. Este algoritmo foi um dos grandes fatores para o ressurgimento do interesse dos pesquisadores nas redes neurais artificiais, uma vez que forneceu uma alternativa para o treinamento de redes baseadas no *Perceptron* com mais de uma camada, e com isto, removendo a limitação destas redes em resolver problemas não linearmente separáveis [11].

Este anexo tem como objetivo descrever detalhadamente este algoritmo tão importante para a área de redes neurais artificiais e que foi utilizado durante a resolução do problema apresentado nesta monografia e foi adaptada de [10].

A legenda das expressões presentes nas equações existentes na explicação do *backpropagation* estão definidas a seguir:

$i$  : índice do neurônio da camada de entrada;

$j$  : índice do neurônio da camada de saída;

$x_{ni}$  : valor de saída do neurônio  $i$  correspondente ao exemplo de entrada  $n$  que será entrada do neurônio  $j$ ;

$d_{nj}$  : valor desejado de saída do exemplo  $n$  para o neurônio  $j$  na camada de saída;

$net$ : entrada líquida do neurônio;

$f(net)$ : função de ativação para a entrada  $net$ ;

$y_{nj}$  : valor calculado de saída do exemplo  $n$  para o neurônio  $j$  na camada de saída;

$\mu$  : taxa de aprendizagem;

$\delta$  : sensibilidade;

O algoritmo *backpropagation* tem como objetivo minimizar uma função objetivo qualquer, previamente definida, durante o processo de aprendizagem da rede neural. Para realizar esse procedimento é utilizado o método do gradiente descendente, calculado para todos os  $n$  exemplos e todos os neurônios de saída  $j$  [10]. Com isso, o erro que será minimizado é definido como:

$$E_n = \sum_n \sum_j e_n \quad (1)$$

Considerando a função objetivo como sendo o erro médio quadrático, a expressão se torna:

$$\frac{1}{n} \sum_n \sum_j (d_{nj} - y_{nj})^2 \quad (2)$$

Utilizando o método do gradiente descendente, o ajuste dos pesos deve ser equivalente ao sentido contrário do gradiente da função objetivo relacionada com o erro, ou seja:

$$\nabla w_{ij} = -\mu \frac{\partial E_n}{\partial w_{ij}} \quad (3)$$

Porém, nas redes MLP, o erro é uma função indireta nos pesos das camadas escondidas, logo é necessário utilizar a regra da cadeia no cálculo da derivada, resultando em:

$$\frac{\partial E_n}{\partial w_{ij}} = \frac{\partial E_n}{\partial net_{nj}} \frac{\partial net_{nj}}{\partial w_{ij}} \quad (4)$$

Onde  $net_{nj}$  é determinada utilizando a equação (2.3). Substituindo o valor da derivada parcial de  $net$  e atribuindo um símbolo para derivada do erro em relação aos pesos, tem-se:

$$\nabla w_{ij} = \mu \delta_{nj} x_{ni} \quad (5)$$

Onde a sensibilidade é definida como:

$$\delta_{nj} = -\frac{\partial E_n}{\partial net_{nj}} = -\frac{\partial E_n}{\partial y_{nj}} \frac{\partial y_{nj}}{\partial net_{nj}} \quad (6)$$

$$\delta_{nj} = f'_j(net_{nj})(d_{nj} - y_{nj}) \quad (7)$$

Como o erro é calculado apenas na camada de saída da MLP, o ajuste ocorre apenas quando  $j$  é um neurônio da camada escondida, cujo valor de  $\delta$  não é conhecido. Logo, deve-se utilizar um valor proporcional do erro calculado na camada posterior. Considerando  $k$  como o índice da camada de saída:

$$\frac{\partial E_n}{\partial y_{nj}} = \sum_k \frac{\partial E_n}{\partial net_{nk}} \frac{\partial net_{nk}}{\partial y_{nj}} \quad (8)$$

Uma vez que a derivada parcial da entrada  $net$  com relação a cada entrada é o peso sináptico do neurônio e incide  $j$  e considerando a equação (6) para  $k$ , resulta em:

$$\frac{\partial E_n}{\partial y_{nj}} = -\sum_k \delta_{nk} w_{jk} \quad (9)$$

Substituindo a equação (9) em (6) tem-se:

$$\delta_{nj} = f'_j(net_{nj}) \sum_k \delta_{nk} w_{jk} \quad (10)$$

Com isso, pode-se resumir a expressão de atualização dos pesos sinápticos dos neurônios como sendo:

$$\nabla w_{ij}^n = \mu \delta_{nj} y_{ij} + \nabla w_{ij}^{n-1} \quad (11)$$

Onde o valor da sensibilidade muda de acordo com a camada onde os pesos estão sendo atualizados. A equação (7) é utilizada caso seja a camada de saída e a (10) caso seja outra camada.

O funcionamento do algoritmo *backpropagation* possui duas etapas, a *forward* onde acontece a propagação do sinal do cálculo das entradas para a saída e a

*backward* que utiliza o erro de saída para realiza os cálculos da camada saída para a de entrada. De modo geral o algoritmo executa as seguintes etapas:

- i. Apresentação das entradas da rede;
- ii. Propagação do sinal calculando a saída da rede;
- iii. Cálculo do erro;
- iv. Caso o erro seja diferente de zero ou de um limiar pré-definido, realiza-se a repropropagação do sinal atualizando os pesos pelo uso das equações (7), (10) e (11);
- v. Verificar critério de parada do treinamento, caso não tenha sido atingido, volta para passo 1;
- vi. Salvar pesos ótimos atuais e finalizar treino da rede;

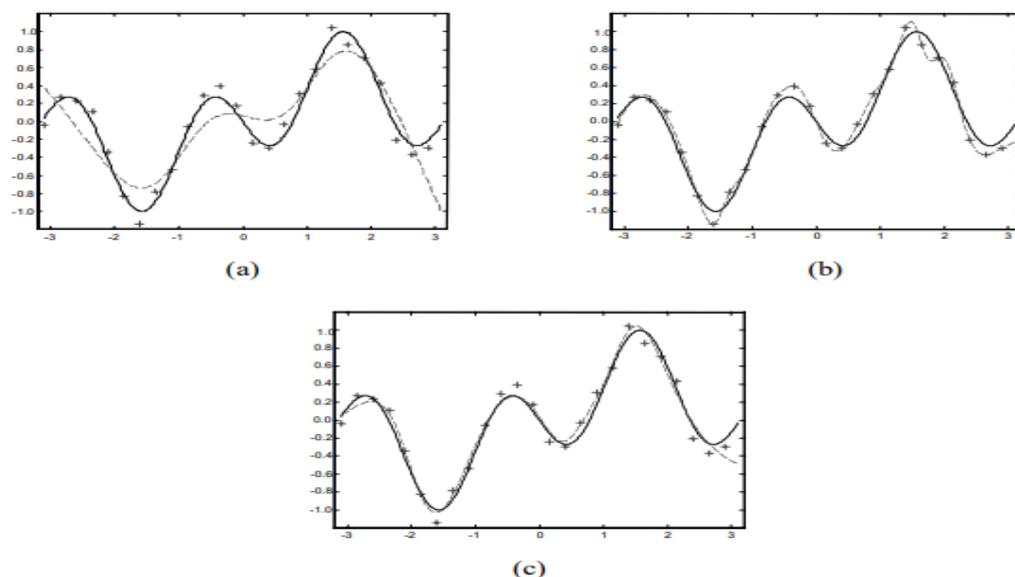
# Apêndice A

## Validação Cruzada

Um importante fator a ser considerado durante o processo de treinamento de uma rede neural é o critério de parada. Algumas abordagens são comumente utilizadas são o número máximo de ciclos ou um valor mínimo para o erro do conjunto de treinamento. Porém, esses métodos não levam em conta o processo iterativo existente nos algoritmos de aprendizados utilizados pelas redes neurais, como o *backpropagation*, tornando esses métodos pouco interessantes [15].

Se o treinamento de uma rede neural for interrompido em um momento indevido, podem acontecer dois tipos de problemas o *overfitting* e o *underfitting*. O primeiro problema, *overfitting*, acontece quando uma rede é treinada em excesso tornando-a super ajustada aos dados de treinamento fazendo com que o modelo simplesmente decore os exemplos. Enquanto no *underfitting* ocorre o problema oposto, quando o treinamento é interrompido cedo demais, a rede ficará sub ajustada apresentando baixo poder de generalização [10].

Figura 22 – Valores de uma rede com (a) *underfitting*, (b) *overfitting* e (c) treinamento interrompido com validação cruzada



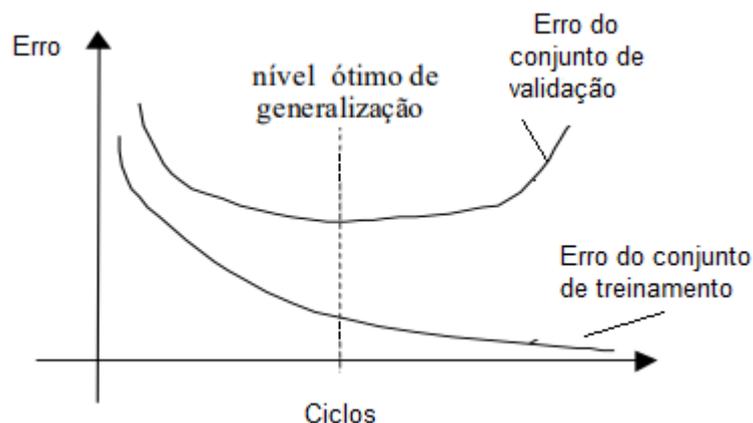
Fonte: Adaptado de [24]

Diante disto, a validação cruzada é apresentada como um método que permite determinar um momento considerado ótimo para que o treinamento seja interrompido e, com isso, fazendo com que a rede apresente uma boa capacidade de generalização para quando for apresentada a novos dados. O funcionamento do método de validação cruzada inicia separando uma parte da base de dados (normalmente entre 10% e 25%) para calcular uma função de erro que é utilizada como referencial para interromper o treinamento quando este valor atingir um mínimo local ou quando ficar constante [24].

O processo de validação cruzada pode ser realizado da seguinte forma:

- i. Dividir os dados em três conjuntos, de treinamento, validação e teste;
- ii. Treinar a rede utilizando apenas o conjunto de treinamento e calcular a função erro a cada  $k$  ciclos;
- iii. Apresentar o conjunto de validação à rede, porém sem atualizar os pesos sinápticos;
- iv. Interromper o treinamento quando o valor da função erro for maior do que era  $k$  ciclos atrás ou então quando ficar estacionada dentro de intervalo de limiar anteriormente definido;

Figura 23 – Representação gráfica do método de validação cruzada



Fonte: Adaptado de [24]