

**UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS**

**Sofia Moreira de Andrade Lopes**

**Desenvolvimento de método baseado em redes neurais  
profundas para estimação de demanda de curto prazo**

**São Carlos**

**2020**



**Sofia Moreira de Andrade Lopes**

**Desenvolvimento de método baseado em redes neurais profundas para estimação de demanda de curto prazo**

Dissertação de mestrado apresentada ao Departamento de Engenharia Elétrica e de Computação da Escola de Engenharia de São Carlos, da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Ciências, Programa de Engenharia Elétrica. Área de concentração: Sistemas Dinâmicos.

Orientador: Prof. Rogério Andrade Flauzino

**Trata-se da versão corrigida da dissertação. A versão original se encontra disponível na EESC/USP que aloja o Programa de Pós-Graduação de Engenharia Elétrica.**

**São Carlos  
2020**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS  
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da  
EESC/USP com os dados inseridos pelo(a) autor(a).

L864d           Lopes, Sofia Moreira de Andrade  
                  Desenvolvimento de método baseado em redes neurais  
                  profundas para estimação de demanda de curto prazo /  
                  Sofia Moreira de Andrade Lopes; orientador Rogério  
                  Andrade Flauzino. São Carlos, 2020.

                  Dissertação (Mestrado) - Programa de  
                  Pós-Graduação em Engenharia Elétrica e Área de  
                  Concentração em Sistemas Dinâmicos -- Escola de  
                  Engenharia de São Carlos da Universidade de São Paulo,  
                  2020.

                  1. Previsão de demanda de potência ativa. 2.  
                  Análise de séries temporais. 3. Redes neurais  
                  artificiais. 4. Redes Autoencoder. 5. Sistemas  
                  dinâmicos. I. Título.

## FOLHA DE JULGAMENTO

Candidata: Engenheira **SOFIA MOREIRA DE ANDRADE LOPES**.

Título da dissertação: "Desenvolvimento de método baseado em redes neurais profundas para estimação de demanda de curto prazo".

Data da defesa: 22/07/2020

### Comissão Julgadora:

### Resultado

Prof. Associado **Rogério Andrade Flauzino**  
**(Orientador)**  
(Escola de Engenharia de São Carlos – EESC/USP)

Aprovada

Prof. Dr. **Hélder de Paula**  
(Universidade Federal de Uberlândia/UFU)

Aprovada

Prof. Dr. **Ricardo de Andrade Lira Rabêlo**  
(Universidade Federal do Piauí/UFPI)

Aprovada

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica:  
Prof. Titular **Ivan Nunes da Silva**

Presidente da Comissão de Pós-Graduação:  
Prof. Titular **Murilo Araujo Romero**



## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, pelas graças diárias que coloca em minha vida.

Ao meu orientador, Prof. Dr. Rogério Andrade Flauzino, pela confiança depositada em mim e pelo auxílio no desenvolvimento deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pela bolsa de mestrado, sem a qual esta pesquisa não seria possível.

À minha família e amigos, pelo amor e apoio, e por me incentivarem a sempre buscar meus objetivos. Em especial, ao Lucas Sales, pelo companheirismo ao longo desses anos.





*"Todos os modelos são errados, mas alguns são úteis."  
George E. P. Box*



## RESUMO

LOPES, S. M. A. **Desenvolvimento de método baseado em redes neurais profundas para estimação de demanda de curto prazo.** 2020. 96p. Dissertação de mestrado - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2020.

A previsão de demanda de potência ativa é um quesito importante para o planejamento da operação do setor elétrico. A informação sobre a demanda futura auxilia na manutenção da qualidade de energia, influencia economicamente o setor, sendo vital para realizar o balanço entre geração e demanda, e afeta diretamente as relações de compra e venda de energia. Dentre os métodos encontrados na literatura, as técnicas de *soft computing* (redes neurais artificiais (RNA), sistemas fuzzy (SF) e máquinas de vetores de suporte (MVS)) se destacam em certos problemas por apresentarem resultados com baixos erros e maior aderência a diferentes tipos de cenários, principalmente quando comparadas com técnicas de estimação linear. Neste trabalho, o método proposto consiste na criação de uma estrutura de RNA profunda (RNAP) para realizar a previsão da demanda para um horizonte de 30 minutos. A utilização da técnica de RNA, ao invés de uma técnica linear, foi justificada através da realização de um teste de hipótese com o problema proposto. Para avaliar o desempenho do método proposto, sua previsão será comparada com a previsão realizada por uma técnica tradicional de RNA. A avaliação dos dois modelos foi realizada utilizando uma série temporal de demanda, obtida a partir de um alimentador conectado a uma subestação; foram utilizadas medições para o período de um ano. Considerando a acurácia e a aderência do modelo de estimação, os resultados obtidos indicam a superioridade da RNAP em relação à técnica tradicional. Desta forma, o método proposto apresentou, na maioria dos casos analisados, modelos mais simples, com erros menores e maior grau de aderência.

**Palavras-chave:** Análise de séries temporais, Previsão de demanda de potência ativa, Redes Autoencoder, Redes neurais artificiais, Sistemas dinâmicos.



## ABSTRACT

LOPES, S. M. A. **Proposal of a deep neural network structure for short-term load forecasting**. 2020. 96p. Dissertação de mestrado - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2020.

Load forecasting is a very important tool for planning the operation of electrical systems. The knowledge about future load, helps to maintain energy quality and economically affects the energy sector, by influencing the balance between generation and load, what directly affects the energy purchase-sale relations. Among the methods found in literature, soft computing techniques (such as artificial neural networks (ANN), fuzzy systems (FS) and support vector machines (SVM)) stand out for presenting results with lower error and greater adherence to different types of problems. In this work, the proposed method consists in the development of a deep ANN structure for short term load forecasting, considering a prediction for 30 minutes in the future. The methodology's performance will be evaluated regarding a benchmarking method, that consists in a traditional ANN. Both methods were analyzed for the forecasting of a load time series, that was obtained from a feeder connected to a power substation. The results indicate the superiority of the proposed method in relation to the traditional method. In most cases, the proposed method returned simpler models, with smaller errors. In addition, the proposed method also did not present local minimum convergence problems, that typically occur for traditional ANNs.

**Keywords:** Artificial neural networks, Autoencoder networks, Dynamic systems, Load forecasting, Time series analysis, .



## LISTA DE FIGURAS

Figura 1 – Modelo do neurônio artificial . . . . .	36
Figura 2 – Funções de ativação para RNAs - A, B e C pertencem à classe de funções parcialmente diferenciáveis; D, E, F e G pertencem à classe de funções totalmente diferenciáveis . . . . .	37
Figura 3 – Estrutura geral de uma rede PMC . . . . .	40
Figura 4 – Estrutura geral de uma Time-delay neural network . . . . .	42
Figura 5 – Rede autoencoder genérica . . . . .	44
Figura 6 – Demanda do alimentador para o período de 01/08/2017 a 31/07/2018 .	49
Figura 7 – Perfil da demanda do alimentador para uma semana sem feriados - Semana de 15/10/2017 à 21/10/2017 . . . . .	51
Figura 8 – Perfil da demanda do alimentador para uma semana com um feriado - Semana de 08/10/2017 à 14/10/2017 . . . . .	51
Figura 9 – Séries temporais para cada estação climática. . . . .	52
Figura 10 – Comparação da demanda do alimentador para dias de diferentes estações	53
Figura 11 – Função de autocorrelação avaliada para as séries das estações . . . . .	54
Figura 12 – Função de autocorrelação parcial avaliada para as séries das estações .	54
Figura 13 – Decomposição da série temporal de primavera . . . . .	57
Figura 14 – Decomposição da série temporal de verão . . . . .	58
Figura 15 – Decomposição da série temporal de outono . . . . .	58
Figura 16 – Decomposição da série temporal de inverno . . . . .	59
Figura 17 – Fluxograma do método utilizado para o tratamento e a decomposição da série temporal de demanda de potência ativa. . . . .	61
Figura 18 – Função de autocorrelação para uma série de ruído branco . . . . .	65
Figura 19 – Esquemático da estrutura de rede neural profunda proposta . . . . .	68
Figura 20 – Resultado da previsão da rede TDNN para série de primavera . . . . .	72
Figura 21 – Histograma do erro de previsão para a topologia selecionada - Primavera	72
Figura 22 – Função de autocorrelação para a série de resíduos de previsão - Primavera	73
Figura 23 – Resultado da previsão da rede neural profunda para série de primavera	74
Figura 24 – Histograma do erro de previsão para a topologia selecionada - Primavera	74
Figura 25 – Função de autocorrelação para a série de resíduos de previsão - Primavera	75
Figura 26 – Resultado da previsão da rede TDNN para série de verão . . . . .	76
Figura 27 – Histograma do erro de previsão para a topologia selecionada - Verão . .	76
Figura 28 – Função de autocorrelação para a série de resíduos de previsão - Verão .	77
Figura 29 – Resultado da previsão da rede neural profunda para série de verão . . .	78
Figura 30 – Histograma do erro de previsão para a topologia selecionada - Verão . .	78
Figura 31 – Função de autocorrelação para a série de resíduos de previsão - Verão .	79

Figura 32 – Resultado da previsão da rede TDNN para série de outono . . . . .	80
Figura 33 – Histograma do erro de previsão para a topologia selecionada - Outono .	80
Figura 34 – Função de autocorrelação para a série de resíduos de previsão - Outono	81
Figura 35 – Resultado da previsão da rede neural profunda para série de outono . .	82
Figura 36 – Histograma do erro de previsão para a topologia selecionada - Outono .	82
Figura 37 – Função de autocorrelação para a série de resíduos de previsão - Outono	83
Figura 38 – Resultado da previsão da rede TDNN para série de inverno . . . . .	84
Figura 39 – Histograma do erro de previsão para a topologia selecionada - Inverno .	84
Figura 40 – Função de autocorrelação para a série de resíduos de previsão - Inverno	85
Figura 41 – Resultado da previsão da rede neural profunda para série de inverno . .	86
Figura 42 – Histograma do erro de previsão para a topologia selecionada - Inverno .	86
Figura 43 – Função de autocorrelação para a série de resíduos de previsão - Inverno	87



## LISTA DE TABELAS

Tabela 1 – Dados estatísticos das séries temporais . . . . .	53
Tabela 2 – Resultados de previsão da rede TDNN - Primavera . . . . .	71
Tabela 3 – Resultados de previsão da rede neural profunda - Primavera . . . . .	73
Tabela 4 – Resultados de previsão da rede TDNN - Verão . . . . .	75
Tabela 5 – Resultados de previsão da rede neural profunda - Verão . . . . .	77
Tabela 6 – Resultados de previsão da rede TDNN - Outono . . . . .	79
Tabela 7 – Resultados de previsão da rede neural profunda - Outono . . . . .	81
Tabela 8 – Resultados de previsão da rede TDNN - Inverno . . . . .	83
Tabela 9 – Resultados de previsão da rede neural profunda - Inverno . . . . .	85
Tabela 10 – Resumo dos resultados de previsão . . . . .	88



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
<b>1.1</b>	<b>Introdução</b>	<b>19</b>
<b>1.2</b>	<b>Motivação e relevância da pesquisa</b>	<b>21</b>
<b>1.3</b>	<b>Objetivos da dissertação</b>	<b>22</b>
<b>1.4</b>	<b>Organização da Dissertação</b>	<b>23</b>
<b>1.5</b>	<b>Trabalhos publicados</b>	<b>24</b>
<b>2</b>	<b>PRINCIPAIS ASPECTOS RELACIONADOS À ESTIMAÇÃO DE DEMANDA DE ENERGIA ELÉTRICA E APLICAÇÕES NO SETOR ELÉTRICO.</b>	<b>25</b>
<b>2.1</b>	<b>Introdução</b>	<b>25</b>
<b>2.2</b>	<b>Conceitos básicos sobre séries temporais</b>	<b>25</b>
<b>2.3</b>	<b>Previsão de séries temporais</b>	<b>27</b>
2.3.1	Modelos lineares de previsão de séries temporais	28
2.3.1.1	Modelo Autoregressivo	29
2.3.1.2	Modelo Média Móvel	29
2.3.1.3	Modelo Autoregressivo de Média Móvel	29
2.3.1.4	Modelo Autoregressivo Integrado de Média Móvel (ARIMA)	30
2.3.2	Modelos não lineares de previsão de séries temporais	31
2.3.2.1	Redes Neurais Artificiais	31
2.3.2.2	Sistemas de inferência fuzzy	31
2.3.2.3	Máquinas de vetores de suporte	32
<b>2.4</b>	<b>Estado da arte de previsão de séries temporais</b>	<b>32</b>
<b>2.5</b>	<b>Conclusões parciais</b>	<b>33</b>
<b>3</b>	<b>REDES NEURAIAS ARTIFICIAIS PARA ESTIMAÇÃO DE DEMANDA</b>	<b>35</b>
<b>3.1</b>	<b>Introdução</b>	<b>35</b>
<b>3.2</b>	<b>Aspectos básicos das RNA</b>	<b>35</b>
<b>3.3</b>	<b>Paradigmas de aprendizado e construção do modelo de RNA</b>	<b>38</b>
<b>3.4</b>	<b>Redes neurais artificiais superficiais</b>	<b>40</b>
3.4.1	Perceptron Multicamadas	40
3.4.1.1	<i>Time-delay neural network</i>	42
<b>3.5</b>	<b>Redes neurais profundas</b>	<b>42</b>
3.5.1	Redes Autoencoder	43
<b>3.6</b>	<b>Redes Neurais Artificiais para previsão de séries temporais</b>	<b>46</b>
<b>3.7</b>	<b>Conclusões parciais</b>	<b>48</b>

<b>4</b>	<b>DESCRIÇÃO DOS DADOS - SÉRIE DE DEMANDA DE POTÊNCIA ATIVA</b>	<b>49</b>
<b>4.1</b>	<b>Análise da série temporal de demanda</b>	<b>49</b>
<b>4.2</b>	<b>Processamento da série temporal de demanda</b>	<b>55</b>
4.2.1	Normalização e remoção de dados corrompidos	55
4.2.1.1	Decomposição da série temporal	56
4.2.2	Construção dos conjuntos de treinamento e de teste	60
<b>4.3</b>	<b>Considerações parciais</b>	<b>60</b>
<b>5</b>	<b>CONSTRUÇÃO DOS MODELOS DE PREVISÃO</b>	<b>63</b>
<b>5.1</b>	<b>Indicadores de desempenho</b>	<b>63</b>
5.1.1	Métricas de erro	63
5.1.2	Diagnóstico dos resíduos	64
5.1.2.1	Histograma dos resíduos	65
5.1.3	Função de autocorrelação dos resíduos	65
<b>5.2</b>	<b>Construção do conjunto de treinamento e de teste</b>	<b>66</b>
<b>5.3</b>	<b>Implementação da rede neural TDNN</b>	<b>67</b>
<b>5.4</b>	<b>Implementação da rede neural profunda</b>	<b>68</b>
<b>5.5</b>	<b>Considerações parciais</b>	<b>70</b>
<b>6</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>71</b>
<b>6.1</b>	<b>Série temporal de demanda de energia - Primavera</b>	<b>71</b>
6.1.1	Previsão com a Time Delay Neural Network	71
6.1.2	Previsão com a rede neural profunda	73
<b>6.2</b>	<b>Série temporal de demanda de energia - Verão</b>	<b>75</b>
6.2.1	Previsão com a Time Delay Neural Network	75
6.2.2	Previsão com a rede neural profunda	77
<b>6.3</b>	<b>Série temporal de demanda de energia - Outono</b>	<b>79</b>
6.3.1	Previsão com a Time Delay Neural Network	79
6.3.2	Previsão com a rede neural profunda	81
<b>6.4</b>	<b>Série temporal de demanda de energia - Inverno</b>	<b>83</b>
6.4.1	Previsão com a Time Delay Neural Network	83
6.4.2	Previsão com a rede neural profunda	85
<b>6.5</b>	<b>Análise dos resultados</b>	<b>87</b>
<b>7</b>	<b>CONCLUSÕES</b>	<b>91</b>
	<b>REFERÊNCIAS</b>	<b>93</b>

# 1 INTRODUÇÃO À DISSERTAÇÃO DE MESTRADO

## 1.1 Introdução

O sistema elétrico de potência é constituído de cinco partes principais - a geração, a distribuição, a transmissão, o comércio de energia e a regulação - cada uma possuindo características e níveis de qualidade particulares que devem ser atendidos. Em relação à distribuição, um aspecto importante é a análise da qualidade do serviço, que prioriza a continuidade do fornecimento de energia elétrica ofertada pelas concessionárias (ANEEL, 2018). Neste sentido, para atender tais níveis de qualidade, um quesito essencial é a previsão de demanda de potência ativa. Este tipo de previsão, que será referida como previsão de demanda, é essencial para a operação e o planejamento do sistema, auxiliando no cumprimento dos níveis de qualidade de energia ao permite ao operador tomar decisões e planejar a operação de forma consciente sobre o estado atual e futuro do sistema (ANEEL, 2015; ANEEL, 2018).

Neste contexto, pode-se observar a operação do SEP como um problema de otimização que visa reduzir uma função do custo de operação, sobre a qual a previsão de demanda exerce influência. Por isto, ao se analisar os resultados obtidos, busca-se o método com menor nível de erro, pois este retornará os resultados com menor risco associado, o que é desejável na operação do sistema (MORAES, 2014).

Pode-se observar a importância e a necessidade de uma previsão de demanda realizada com bons níveis de precisão. Todavia, por se tratar de um problema com características altamente não lineares, é um tema de pesquisa que ainda apresenta dificuldades e muitos estudos são desenvolvidos buscando resultados com cada vez com maior acurácia (MOHAMMAD; LEE; KIM, 2018). Na literatura, encontram-se diversas técnicas aplicadas para previsão de demanda, podendo ser classificadas em três classes distintas: técnicas tradicionais de previsão, técnicas tradicionais modificadas e técnicas de *soft computing* (SINGH et al., 2012). Neste trabalho, as técnicas pertencentes à terceira classe serão exploradas para a abordar o problema de previsão proposto.

*Soft computing* é um ramo da área de inteligência artificial que engloba os métodos de algoritmos genéticos, de sistemas fuzzy, de redes neurais artificiais (RNA), de árvores de decisão, indução de regras, entre outros (SINGH et al., 2012). Além disso, com o avanço da capacidade computacional, novos métodos como os de *deep learning* têm chamado a atenção de pesquisadores da área, dentre estes, as redes neurais profundas se destacam para certas aplicações, apresentando resultados com maior acurácia em relação às RNAs tradicionais (BENGIO et al., 2007). Neste trabalho, optou-se pelo uso das técnicas de RNAs tradicionais e profundas, pela sua capacidade de lidar com extensos conjuntos

de dados e de aprender sobre as relações de dependência temporal existentes entre os elementos destes conjuntos.

As RNAs são modelos computacionais inspirados na estrutura do cérebro humano, o qual é essencialmente constituído por uma rede com inúmeras unidades de processamento, neurônios, altamente conectadas. As principais características destas redes são a capacidade de aprendizado e de generalização. Desta forma, as RNAs buscam aprender automaticamente comportamentos a partir de padrões ou exemplos que lhe são apresentados. Trata-se de uma técnica robusta, sendo que um dos grandes atrativos das RNAs é a capacidade de lidar com mapeamentos não-lineares. A RNA aprender a partir das informações contidas no conjunto de dados, ajustando seus parâmetros internos para representar as relações não lineares entre as entradas e as saídas deste conjunto (SILVA; SPATTI; FLAUZINO, 2015; HAYKIN, 1998).

Conforme houve evolução da capacidade computacional, novas abordagens de RNA se tornaram viáveis, tais como as redes neurais artificiais profundas (RNAP) (ou *Deep Neural Networks* (DNN), em inglês). A principal vantagem das redes neurais profundas é a sua capacidade de lidar com problemas extremamente complexos e com não linearidades e mesmo assim obter resultados precisos. Isto é possível graças aos diferentes níveis de não-linearidade da própria rede, que a tornam capaz de modelar funções altamente não-lineares e variantes (BENGIO et al., 2007). Estas redes também possuem uma capacidade maior que a das redes superficiais para abstrair relações e características presentes no conjunto de dados (HOSSEN et al., 2017). Embora possuam mais eficiência de previsão, o treinamento das RNAP pode ser custoso, exigindo um elevado esforço computacional. Por este motivo, diferentes metodologias foram desenvolvidas para simplificar o processo de treinamento destas redes, tornando-o viável para aplicações práticas (BENGIO et al., 2007).

Neste estudo, as técnicas de RNA são utilizadas para abordar o problema de previsão da demanda para um horizonte de curto prazo. Este problema será considerado pela ótica da previsão de sistemas variantes no tempo, pois os dados de demanda disponíveis constituem uma série temporal. Pode-se definir uma série temporal como uma realização de um processo estocástico, sendo que é justamente este processo gerador da série que a RNA tem objetivo de modelar (BOX et al., 2016).

Para a previsão da série temporal de demanda serão aplicadas duas RNAs. A primeira consiste em uma rede perceptron multicamadas (PMC) na configuração de *Time-delay neural network* (TDNN), utilizada com fins comparativos. A segunda rede proposta se trata de uma estrutura de rede neural profunda, formada a partir da concatenação de redes da arquitetura autoencoder com uma rede PMC. O funcionamento da rede autoencoder se baseia na repetição dos parâmetros de entrada da rede em sua saída. Para realizar esta cópia, a rede pode diminuir ou aumentar a dimensão dos dados de entrada. Durante este

---

processo de mudança de dimensão, ela extrai dos dados suas características principais, ajustando seus parâmetros de acordo com as informações aprendidas. Por este motivo, optou-se por esta arquitetura, pois, espera-se que este processo afete positivamente o nível de aprendizado da estrutura de rede final e facilite o processo de treinamento da RNAP.

A série temporal de demanda utilizada para estudo de caso foi obtida em uma subestação interligada ao sistema de distribuição brasileiro. O objetivo principal do trabalho é averiguar o desempenho das redes neurais na tarefa de previsão de curto prazo para esta série, de forma a avaliar a aplicabilidade prática da técnica proposta como ferramenta de apoio ao planejamento para empresas operadoras do sistema elétrico de potência (SEP).

## 1.2 Motivação e relevância da pesquisa

O problema de previsão de demanda é presente na operação diária de diversos ramos do setor de energia, afetando desde ramos operacionais e de planejamento até ramos econômicos. Do ponto de vista operacional e de planejamento de operação, a previsão da demanda é vital pois as empresas deste setor têm como objetivo principal garantir o suprimento de energia de forma segura e estável para o consumidor final. A previsão da demanda energética auxilia a economia financeira durante a operação e reduz os custos de manutenção, de forma que, existe uma correlação positiva entre a economia de recursos financeiros e a acurácia da previsão ([ALMESHAI EI; SOLTAN, 2011](#)).

Atualmente, um aspecto que dificulta o planejamento operacional e o balanço entre o nível de geração e o de demanda do sistema é a crescente inserção de fontes de geração distribuída, cujo perfil de geração é intermitente, alterando o dinamismo do sistema de potência e tornando-o mais complexo. Do lado das cargas, também existem aspectos que dificultam a previsão da demanda, como a conexão de cargas que alteram de forma brusca e anormal o perfil de consumo e que, portanto, mudam o comportamento esperado do sistema. Por este motivo, dados de previsão são relevantes, não apenas do lado da previsão de demanda, mas também na previsão de geração.

Esta realidade tem impulsionado estudos que trabalham com a previsão em pequenas áreas, ou até mesmo focadas no consumo de equipamentos específicos. Este novo tipo de abordagem, denominado de previsão de demanda hierárquica, pode ser aplicada em vários níveis de consumo e pode considerar diferentes horizontes de previsão e ocorreu principalmente devido à presença dos medidores inteligentes bidirecionais, que têm papel de destaque dentro do contexto de *smart grids* ([HONG; FAN, 2016](#)).

Além do aspecto de planejamento operacional, a previsão de demanda tem um importante viés econômico, pois tendo em vista que se busca o equilíbrio entre demanda e geração no sistema, o planejamento da geração com base na previsão de demanda é vital para economia. Ainda, a previsão da demanda de potência é um parâmetro importante

para a regulação do mercado de energia, afetando diretamente o preço da energia ofertada (HE, 2017).

Ainda em relação ao mercado de energia, os dados de previsão de demanda auxiliam empresas de distribuição no processo de contratação de demanda; empresas de geração no planejamento e ajuste nível de geração de energia; além de ser um fator essencial para o planejamento da expansão estrutural do sistema elétrico (SINGH; SINGH, 2013). No Brasil, em particular, as diretrizes para o desenvolvimento de estudos de previsão de demanda que visam auxiliar na expansão física do SEP são determinadas pelo módulo 2 do documento de Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRODIST) (ANEEL, 2015).

Dados de previsão de curto prazo são principalmente utilizados para realizar o balanço entre demanda e geração de energia (KUSTER; REZGUI; MOURSHED, 2017). Além disso, este tipo de previsão auxilia no processo de tomada de decisão do operador do SEP, sendo útil para o planejamento do fornecimento, da reserva e do despacho de energia (FALLAH et al., 2019). Enquanto que, dados de previsão com horizonte de médio prazo, ou seja, para um período entre uma semana e um ano no futuro, são muito utilizados pelo mercado de energia, principalmente para definição de preços de combustíveis e para a negociação de contratos com empresas de energia (GONZÁLEZ-ROMERA; JARAMILLO-MORÁN; CARMONA-FERNÁNDEZ, 2006). Por sua vez, previsões de longo prazo auxiliam na expansão do sistema, são úteis para estudos de análise de retorno de investimentos e para análises de lucros (KUSTER; REZGUI; MOURSHED, 2017).

Tendo em vista a importância do tema, nota-se a relevância de se construir modelos de previsão que realizem esta tarefa com baixo erro associado. Este é o objetivo principal deste trabalho, definido formalmente na próxima seção.

### 1.3 Objetivos da dissertação

Esta pesquisa visa uma técnica de previsão para o horizonte de curto prazo, que ofereça resultados com bons níveis de acurácia para o problema definido na seção anterior, ou seja, a previsão de demanda de potência. Assim, o objetivo principal deste trabalho pode ser definido como: **Desenvolvimento de uma estrutura utilizando técnicas de redes neurais artificiais profundas para realizar a previsão da demanda de potência ativa considerando um horizonte de previsão de curto prazo.**

A validação da técnica será feita pelo estudo de caso com a previsão da série temporal de demanda ativa obtida a partir de um alimentador conectado a uma subestação do estado de São Paulo, Brasil.

Por fim, ressalta-se que o diferencial deste estudo consiste na metodologia utilizada para o desenvolvimento do modelo de previsão. Com isto, espera-se que a técnica proposta



funcione como uma ferramenta de auxílio na definição de topologias e parâmetros iniciais das RNAs, contornando problemas encontrados na abordagem tradicional, tais como a convergência para mínimos locais devido à inicialização aleatória dos pesos sinápticos. Espera-se ainda que a ferramenta contribua com estudos futuros na área de sistemas inteligentes aplicados à previsão de dados provenientes de diversos tipos de sistemas dinâmicos, não se restringindo apenas à área de demanda de energia.

Desta forma, as contribuições principais desta pesquisa podem ser detalhadas como:

- Utilização da abordagem de construção de rede neural profunda por meio do uso de redes Autoencoder para o problema de previsão de demanda de potência, tendo em vista que este problema não foi analisado através desta abordagem na literatura;
- Apresentação de um método de simples implementação, que dispensa o uso de algoritmos de otimização para seleção de topologias e que ainda sim apresenta níveis de precisão melhores do que os dos métodos tradicionais;
- Apresentação do processo de análise estatística feita na série temporal de demanda de potência para sua correta utilização no modelo de previsão.

#### 1.4 Organização da Dissertação

Este trabalho está dividido em sete capítulos, considerando o presente capítulo de introdução. O [Capítulo 2](#) apresenta uma revisão bibliográfica sobre o tema de estimação de demanda, abordando o conceito de demanda como uma série temporal. Além de apresentar os principais métodos utilizados para realizar a previsão deste tipo de série.

Uma revisão teórica também é apresentada no [Capítulo 3](#), onde os principais conceitos sobre redes neurais artificiais e profundas são apresentados. Neste capítulo também se encontra uma revisão de literatura sobre RNAs aplicadas para o problema de previsão de demanda. Em seguida, o [Capítulo 4](#) introduz ao leitor a série temporal de demanda escolhida como estudo de caso para a validação do método proposto. Além disso, são apresentadas as análises e ajustes aplicados aos dados da série para sua utilização no trabalho.

O [Capítulo 5](#) detalha a implementação do método de RNA proposto. Os resultados dos métodos implementados neste capítulo, são apresentados no [Capítulo 6](#) e uma análise sobre eles é desenvolvida. Por fim, o [Capítulo 7](#) contém as conclusões inferidas com base no trabalho desenvolvido e propostas consideradas para trabalhos futuros.

Este documento apresenta as realizações parciais da pesquisa de mestrado, desta forma, os capítulos de metodologia, de resultados e de conclusões (capítulos [5](#), [6](#) e [7](#), respectivamente) apresentam apenas considerações sobre os resultados parciais e ainda serão alterados para a inclusão dos resultados finais da pesquisa.

## 1.5 Trabalhos publicados

Até o desenvolvimento deste documento, dois trabalhos frutos do tema desta dissertação haviam sido publicados ou aceitos para a publicação. A relação destes trabalhos é apresentada a seguir:

- LOPES, S. M. A.; FLAUZINO, R. A.: A Comparative Study of Load Forecasting Using Deep Neural Networks. In: 13th Latin-American Congress on Electricity Generation and Transmission (CLAGTEE 2019), 2019, Santiago. Proceedings of the 13th Latin-American Congress on Electricity Generation and Transmission - CLAGTEE 2019;
- LOPES, S. M. A.; FLAUZINO, R. A.: Análise da Aderência do Modelo de Redes Neurais Artificiais para Previsão de Demanda de Energia Utilizando a Função de Autocorrelação. In: Conferência Brasileira de Dinâmica, Controle e Aplicações (DINCON 2019), 2019, São Carlos.

## 2 PRINCIPAIS ASPECTOS RELACIONADOS À ESTIMAÇÃO DE DEMANDA DE ENERGIA ELÉTRICA E APLICAÇÕES NO SETOR ELÉTRICO.

### 2.1 Introdução

A demanda de energia pode ser definida como a média do fluxo de potência ativa presente em um componente do sistema de distribuição, durante um intervalo de tempo específico (ANEEL, 2015). Considerando, o tratamento da demanda como uma série temporal, técnicas estatísticas provenientes da teoria de séries temporais podem ser aplicadas para realizar a modelagem do comportamento dinâmico da série temporal.

Formalmente, uma série temporal é definida como uma sequência de observações ordenadas no tempo e, normalmente, separadas por intervalos iguais (ADHIKARI; AGRAWAL, 2013). Pode-se afirmar que uma série temporal é uma realização de um processo estocástico. Em outras palavras, uma série com  $N$  elementos consiste em uma amostra de tamanho  $N$  retirada de uma população infinita, sendo esta população composta por todas as medições que podem ser geradas pelo processo estocástico que gerou a série (BOX et al., 2016).

Neste capítulo, inicialmente serão abordados os principais conceitos necessários para a compreensão do trabalho da teoria de séries temporais. Em seguida, uma análise explicativa sobre os modelos de previsão de séries temporais será realizada, abordando tanto modelos lineares, quanto os não lineares. Além disso, é apresentada uma revisão literária de estudos que realizaram a previsão da série temporal de demanda, incluindo estudos com diferentes técnicas e horizontes de previsão. Por fim, são apresentadas as conclusões parciais sobre o capítulo.

### 2.2 Conceitos básicos sobre séries temporais

As séries temporais estão presentes em diversas aplicações e em estudos de diferente áreas, tais como na economia, como em séries que acompanham as variações de preços de ações (KARA; BOYACIOGLU; BAYKAN, 2011); na área da saúde, em séries sobre evolução de taxas de natalidade e mortalidade ou sobre índices de ocorrência de doenças para estudos epidemiológicos (BUI et al., 2018); e na engenharia, em séries de consumo, geração ou demanda de energia, sendo que a previsão dos níveis de geração se tornou essencial com o aumento da inserção de fontes de geração distribuída nas redes de distribuição (KARDAKOS et al., 2013). Além disso, ainda na área de engenharia elétrica, as séries temporais podem ser utilizadas para obtenção de informações sobre o sistema de distribuição, como por exemplo, informações sobre o perfil de consumo dos clientes conectados à rede (ESPINOZA et al., 2005).

Uma série temporal  $\mathbf{z}$  contendo  $N$  elementos pode ser representada matematicamente como se segue:

$$\mathbf{z} = z_1, z_2, \dots, z_N \quad (2.1)$$

Uma característica intrínseca das séries temporais é a dependência existente entre observações adjacentes e cabe ao ramo de análise de séries temporais o desenvolvimento de modelos dinâmicos e estocásticos que realizem o estudo de tal dependência. É com base nessa característica e na existência de tais modelos, que se pode realizar a previsão dos valores futuros da série (BOX et al., 2016).

Usualmente, considera-se que uma série temporal é composta por quatro componentes: sazonalidade, tendência, ciclo e componentes residuais. A sazonalidade é um padrão periódico que aparece na série recorrentemente, normalmente ocasionado por fatores externos, tais como condições climáticas, épocas festivas, entre outros. Um exemplo é o aumento do uso de equipamentos de conforto térmico durante o verão para regiões quentes, ou aumento do nível de vendas durante as festas de final de ano (ADHIKARI; AGRAWAL, 2013).

A tendência, por sua vez, indica o comportamento da série para um longo período, isto é, indica se ao nível da série é crescente, decrescente ou permanece estável e qual a velocidade destas mudanças. Uma série pode apresentar tendência constante, linear, quadrática, dentre outras (HYNDMAN; ATHANASOPOULOS, 2018).

Os ciclos de uma série temporal também são causados por agentes exógenos que afetam a série, de forma semelhante às variações sazonais. Todavia, a duração de um ciclo costuma ser maior do que a duração de uma variação sazonal, além disso, os ciclos não apresentam duração uniforme, o que ocorre com a componente de sazonalidade. Quando removem-se as componentes de tendência, sazonalidade e ciclo da série, o sinal resultante é denominado de componente residual (HYNDMAN; ATHANASOPOULOS, 2018). Nesta componente, normalmente, se encontrará a parte estocástica da série, ou seja, a parte puramente aleatória que não pode ser prevista. Todavia, também é possível encontrar características determinísticas remanescentes na componente residual.

Existem diferentes aspectos que caracterizam uma série temporal, tais como esta ser contínua ou discreta, univariada ou multivariada. Para o primeiro caso, uma série é contínua se a grandeza medida é obtida de forma contínua no tempo, enquanto que, uma série é discreta se os valores para esta grandeza são obtidos em instantes discretos no tempo, normalmente espaçados por intervalos iguais. Para o segundo caso, a diferença está em quantas variáveis estão sendo medidas e armazenadas a partir de um determinado sistema (ADHIKARI; AGRAWAL, 2013).

## 2.3 Previsão de séries temporais

Um dos principais objetivos da análise de séries temporais é a previsão dos valores futuros desta série. Com este propósito, são construídos modelos que representem as características internas da série e o seu processo gerador, de forma a prever os valores futuros das variáveis com base no passado. Ou seja, quando realiza-se a análise de uma série temporal, busca-se encontrar as mudanças que ocorreram na série e em seu comportamento de forma a prever quais mudanças serão apresentadas no futuro ([MAHALAKSHMI; SRIDEVI; RAJARAM, 2016](#)).

Com base em suas características, uma série pode ser modelada por modelos determinísticos ou estocásticos. Dependendo do tipo de fenômeno cujo comportamento se quer estudar, é possível criar um modelo baseado em leis físicas, de forma que seja possível calcular o valor exato de uma variável temporal deste fenômeno para qualquer instante futuro de tempo. Caso este cálculo seja exato, este modelo é dito totalmente determinístico ([BOX et al., 2016](#)). Todavia, para fenômenos com variáveis temporais é comum que haja a interferência de fatores desconhecidos, desta forma, a criação de modelos determinísticos torna-se falha.

Para tratar de fenômenos com tais características e realizar a previsão de seus valores futuros, são utilizados os modelos de probabilidade, ou estocásticos. Estes modelos realizam a previsão de valores futuros calculando a probabilidade do valor futuro estar entre dois limites específicos. Tendo em vista que séries temporais são realizações de processos estocásticos, sua previsão deve ser realizada a partir de modelos estocásticos.

A previsão da série temporal de demanda é classificada na literatura de acordo com o tamanho do horizonte de previsão considerado, sendo este definido pelo número de instantes futuros para os quais é feita a previsão. O tipo de aplicação dos resultados da previsão dependerá fortemente do horizonte previsão selecionado, além disso, este fator também influencia na acurácia da previsão, pois, quanto maior o horizonte de previsão, maior o grau de incerteza associado aos valores previstos.

Segundo ([ALMESHAI EI; SOLTAN, 2011](#)), esta classificação ocorre para os seguintes intervalos:

- Previsão a curto prazo: Previsão realizada para o período de até uma semana no futuro. Com estes dados de previsão, são planejados os despachos de energia e os níveis de geração;
- Previsão a médio prazo: Previsão para o período de uma semana à um ano no futuro. Com estes dados, são organizadas as compras de combustíveis;
- Previsão à longo prazo: Horizontes de previsão acima de um ano no futuro. Dados de previsão à longo prazo são úteis para auxiliar no desenvolvimento do sistema, isto

é, planejamento de novas unidades geradoras, de expansão de linhas de distribuição e de transmissão, entre outros empreendimentos.

Para realizar a previsão são utilizados diferentes modelos, de forma que a escolha do melhor modelo depende do horizonte de previsão que se necessita e do tipo de série temporal que se deseja analisar. A construção ou escolha do modelo utilizado deve respeitar alguns aspectos (HYNDMAN; ATHANASOPOULOS, 2018):

- O modelo deve abranger todas as relações de dependência existentes na série;
- Os resíduos da previsão não devem apresentar correlação serial entre si;
- Deve-se buscar um modelo que explique o comportamento da série com o menor número de parâmetros possíveis.

Os dois primeiros aspectos são relacionados, pois, se o modelo de previsão for capaz de abstrair todas as relações existentes na série, não existirão relações remanescentes no conjunto de ruídos para que estes apresentem algum nível de autocorrelação. Ressaltando que os resíduos neste contexto são os erros do método de previsão, ou seja, a diferença entre o que foi previsto e o que era esperado.

Em relação ao último aspecto, este se relaciona com o conceito de parcimônia, que prega que a escolha do modelo deve equilibrar a acurácia dos resultados, com o número de parâmetros livres do modelo, cujo valor deve ser o menor possível (BOX et al., 2016). Ao seguir este conceito, são evitados fenômenos como modelos especialistas (fenômeno de *over fitting*), que não têm capacidade de generalização para situações desconhecidas, e modelos com pouco conhecimento, que são incapazes de aprender as relações existentes no próprio sistema avaliado.

Os modelos estocásticos de séries temporais podem ser divididos em lineares e não lineares, sendo que a separação entre os modelos é feita pelo tipo de dependência existente entre o instante atual da série e os passados, ou seja, considerando se essa dependência é linear ou não (ADHIKARI; AGRAWAL, 2013). Dentre os modelos lineares, pode-se citar o clássico modelo ARIMA, muito utilizado devido à sua simplicidade e bons resultados. Uma explicação detalhada sobre os principais modelos lineares e não lineares é apresentada nas próximas seções.

### 2.3.1 Modelos lineares de previsão de séries temporais

O modelo linear de estimação mais utilizado é o Modelo Autoregressivo Integrado de Média Móvel (ARIMA), sendo que este se desdobra em inúmeros sub modelos, tais como: Autoregressivo, Média Móvel e o Autoregressivo de Média Móvel. Para a construção do modelo ARIMA, normalmente se segue o método Box-Jenkins. Este modelo pode ser

aplicado a diversos tipos de séries temporais, desde que estas sejam lineares e possuam distribuição próxima da normal (ADHIKARI; AGRAWAL, 2013). Uma série é considerada linear se a relação entre seus instantes puder ser expressa como uma função linear. Uma breve explicação sobre a composição de tais modelos é apresentada a seguir.

### 2.3.1.1 Modelo Autoregressivo

Como indicado em seu nome, este modelo realiza a previsão de uma variável com base em uma regressão desta mesma variável, ou seja, este modelo assume que o instante atual da série depende linearmente de seus instantes passados (HYNDMAN; ATHANASOPOULOS, 2018). Desta forma, a previsão é feita com base na Eq. (2.2).

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + e_t \quad (2.2)$$

Onde  $y_t$  é o valor da série para o instante  $t$ ,  $e_t$  é o erro associado a este instante, sendo representado por um termo de ruído branco, e a constante  $c$  define a média da série. Na equação (2.2), os parâmetros deste modelo são dados pelos termos  $\phi_i$  e a ordem do modelo é definida por  $p$ .

### 2.3.1.2 Modelo Média Móvel

Este modelo se assemelha ao anterior, todavia, apesar de considerar o erro associado para o instante de previsão e o nível  $c$  da série, neste modelo o instante atual é associado com os erros dos instantes passados e não com os próprios instantes.

$$y_t = c + \sum_{j=1}^q \theta_j e_{t-j} + e_t \quad (2.3)$$

Esta equação define um modelo de ordem  $q$ , com parâmetros dados por  $\theta_j$ . Assim como para o modelo AR, os termos de erro representam um processo de ruído branco.

### 2.3.1.3 Modelo Autoregressivo de Média Móvel

O modelo ARMA é formado pela união dos modelos AR e MA, sendo matematicamente descrito por:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + e_t \quad (2.4)$$

Este tipo de modelo somente pode ser aplicado para séries temporais estacionárias, sendo estas, séries cujas propriedades não são afetadas pelo instante de tempo no qual se avalia a série. Isto é, a variável avaliada apresentará as mesmas propriedades, independentemente do instante de tempo observado. Séries estacionárias oscilam em torno de

um nível constante, não apresentado tendência e nem sazonalidade. Outro indicativo da não-estacionaridade de uma série temporal é sua função de autocorrelação, que tende a apresentar o coeficiente de *lag1* com valor elevado e positivo (HYNDMAN; ATHANASOPOULOS, 2018). Séries temporais não estacionárias podem se tornar estacionárias através da diferenciação.

#### 2.3.1.4 Modelo Autoregressivo Integrado de Média Móvel (ARIMA)

O modelo ARIMA, também conhecido como modelo Box-Jenkins, é formado a partir do modelo ARMA de forma a permitir a aplicação para séries temporais não-estacionárias.

$$ARIMA(p, q, d) = AR(p) + I(d) + MA(q) \quad (2.5)$$

Sendo:

- $p$  - Ordem do termo auto regressivo;
- $q$  - Ordem do termo de média móvel;
- $d$  - Número de vezes que a série foi diferenciada para se tornar estacionária;
- $AR(p)$  - Indica que o valor ( $y_t$ ) a ser previsto, depende de  $p$  valores passados desta variável. Foi definido na Eq. (2.2);
- $MA(q)$  - Indica que o valor ( $y_t$ ) a ser previsto, depende de  $q$  valores passados de erro. Foi definido na Eq. (2.3).

Assim, o modelo total ARIMA pode ser escrito como:

$$y_t = c + \sum_{i=1}^p \phi_i \cdot y_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + e_t \quad (2.6)$$

Onde  $\phi_i$  e  $\theta_j$  são, respectivamente, os coeficientes auto regressivo e de média móvel. A escolha do modelo é feita com base na função de autocorrelação (FAC) e na função de autocorrelação parcial (FACP) (JHA; SINHA, 2014). O uso das funções FAC e FACP pode ser feito quando se trata de modelos  $ARIMA(p, 0, d)$  ou  $ARIMA(0, q, d)$ , para encontrar os valores de  $q$  e  $p$ , respectivamente. Todavia, esta abordagem não funciona quando ambos os valores são não nulos (HYNDMAN; ATHANASOPOULOS, 2018).

Identifica-se um modelo do tipo  $ARIMA(p, 0, d)$ , quando os coeficientes da FAC decaem exponencialmente ou de forma senoidal e quando a FACP apresenta valores significativos somente até o coeficiente  $p$ . O modelo  $ARIMA(0, q, d)$ , por sua vez, é identificado quando os os coeficientes da FACP decaem exponencialmente ou de forma senoidal e quando a ACP apresenta valores significativos somente até o coeficiente  $q$ .



### 2.3.2 Modelos não lineares de previsão de séries temporais

Os modelos não lineares para séries temporais contornam a necessidade, exigida pelos modelos lineares, de linearidade da série. Além disso, estes modelos não requerem uma determinada distribuição da série. Dentre esta classe de modelos estão as técnicas de soft computing, como as redes neurais artificiais (RNA), os sistemas de inferência fuzzy (SIF) e as máquinas de vetores de suporte (MVS). A seguir estes três métodos serão apresentados e suas aplicações para a previsão voltada ao setor de energia elétrica serão avaliadas.

#### 2.3.2.1 Redes Neurais Artificiais

As redes neurais artificiais são modelos com capacidade de aprendizado, que se baseiam no conjunto de dados disponíveis e não no processo gerador dos dados. Por este motivo, tais redes podem modelar tanto problemas lineares quanto problemas com não linearidades (ZHANG; QI, 2005). A criação da estrutura do modelo é realizada de forma empírica pelo especialista, mas o ajuste de seus parâmetros livres é feita de forma sistemática através de um algoritmo de treinamento. Este algoritmo normalmente visa minimizar uma função custo, usualmente definida como a função de erro calculada entre a resposta obtida pela rede e a resposta esperada.

Pelo processo de aprendizado ser guiado pelo conjunto de dados da rede, a constituição deste é primordial para o bom desempenho da rede. O conjunto de treinamento deve representar todo o universo do sistema a ser modelado. Além disso, é de suma importância realizar um tratamento prévio dos dados antes destes serem utilizados como entrada da RNA. No contexto de séries temporais, este tratamento inclui não apenas a normalização dos elementos da série, como também a decomposição de tal série, retirando as parcelas de sazonalidade e tendência.

Este ponto é de suma importância, pois, devido a capacidade das RNA de aprender relações não-lineares, diversos estudos avaliaram a hipótese de que as RNA poderiam realizar diretamente a previsão de uma série temporal, isto é, sem que fosse feito nenhum tratamento prévio com a série. Todavia, de acordo com (ZHANG; QI, 2005), esta hipótese não se sustenta e os resultados de previsão obtidos para séries temporais sem tais parcelas eram melhores do que os obtidos diretamente.

#### 2.3.2.2 Sistemas de inferência fuzzy

Assim como as RNAs, os SIF também operam mapeando as relações existentes entre as variáveis de entrada e as de saída. A principal diferença entre as técnicas é que os SIF não necessitam que seus dados (tanto de entrada quanto de saída) sejam numéricos. Sendo que as variáveis podem ser definidas por expressões como "alto", "baixo", que indicam o nível da variável tratada. Além disso, as relações entre as entradas e saídas de um sistema

fuzzy são determinadas por expressões linguísticas condicionais de "Se-Então", as quais muitas vezes são obtidas de forma empírica por especialistas (RANAWEERA; HUBELE; KARADY, 1996).

A vantagem dos sistemas fuzzy está em sua capacidade de lidar com informações vagas e imprecisas e conseguir representar tais informações de forma matemática. Sendo que um dos motivos dos sistemas fuzzy terem tais habilidades, é justamente o uso de regras linguísticas em sua constituição (ALI et al., 2016; MUKHOPADHYAY et al., 2017).

No contexto de previsão de demanda, os sistemas fuzzy são capazes de agregar na previsão as incertezas associadas com a demanda de energia, tais como as variações de agentes exógenos que afetam a demanda (condições climáticas, ocorrência de eventos, entre outros). Muitas vezes os fatores externos afetam a demanda, mas a relação entre eles não pode ser definida de forma exata numericamente. Para isso, os sistemas fuzzy utilizam funções de pertencimento para identificar o relacionamento entre as variáveis (ALI et al., 2016).

### 2.3.2.3 Máquinas de vetores de suporte

Máquinas de vetores de suporte são técnicas baseadas na teoria de aprendizado estatístico e na teoria de aprendizado de Vapnik. É um tipo de método muito utilizado na área de aprendizado de máquinas e pode ser aplicado para diversos problemas como classificação e previsões. Para obter o aprendizado sobre a série temporal que irá prever, a MVS é treinada por um Regressor de vetores de suporte, que se trata de uma metodologia que utiliza dados passados da série para obter conhecimento (NIU; WANG; WU, 2010).

Assim como as RNAs, o aprendizado das MVS também consiste em um processo de minimização de uma função risco. Todavia, para as MVS o risco considerado está associado com a estrutura do modelo e não com os erros do processo de treinamento. Com esta forma de aprendizado, a MVS busca um modelo de estrutura ótima. Ressalta-se que para este tipo de método, os resultados obtidos são únicos e representam o resultado ótimo global, o que não ocorre para os outros métodos citados (HONG, 2009).

## 2.4 Estado da arte de previsão de séries temporais

Esta seção reúne estudos de previsão de demanda de potência ativa que aplicaram os métodos apresentados nas seções anteriores. No contexto dos modelos lineares, os modelos ARIMA, por exemplo, foram utilizados em (SINGH; SINGH, 2013) para realizar a previsão para um período de um dia a uma semana no futuro, considerando dados históricos de demanda.

Considerando os modelos não-lineares, dentre os métodos mais utilizados na literatura, encontram-se as diversas arquiteturas de redes neurais. Tendo em vista que este

trabalho visa a aplicação deste tipo de técnicas, o [Capítulo 3](#) apresentará não apenas os conceitos básicos sobre as RNAs, como também serão apresentados estudos relevantes que utilizam técnicas de redes neurais superficiais e profundas para a previsão de demanda.

Considerando as outras técnicas de *soft computing*, em ([ALI et al., 2016](#)) um sistema fuzzy realiza a previsão de demanda para longo prazo para o período de um ano a frente, utilizando como entradas dados de temperatura e umidade. Os bons resultados do trabalho e o alto nível de acurácia do modelo indicam validade da hipótese de dependência da demanda com as variáveis climáticas. Estas mesmas variáveis climáticas também foram utilizadas em ([MUKHOPADHYAY et al., 2017](#)), onde as funções de pertinência modelaram as variáveis considerando o horário do dia - baseando-se na hipótese de que, dependendo da hora do dia, a influência da variável sobre a demanda de energia varia.

Em ([NIU; WANG; WU, 2010](#)), a previsão de demanda para curto prazo é realizada através de uma MVS. O estudo propõe um pré-processamento dos dados utilizando o algoritmo de otimização de colônia de formigas, cujo papel principal foi a seleção dos atributos que influenciavam os dados de demanda. Este trabalho mostrou que o modelo proposto apresenta melhor desempenho que uma MVS comum e que uma RNA treinada pelo algoritmo back-propagation.

Um modelo de previsão híbrido constituído por um modelo ARIMA e uma MVS foi proposto em ([NIE et al., 2012](#)). A proposta utilizou o primeiro modelo para previsão da parcela linear da série de demanda, enquanto o segundo realizou a previsão da parcela não linear, que correspondia ao erro referente à primeira previsão. Uma metodologia similar foi desenvolvida em ([VELASCO et al., 2018](#)), neste, o modelo ARIMA foi utilizado junto com uma RNA. Em ambos os estudos, os modelos híbridos foram capazes de unir os pontos fortes dos métodos individuais, apresentando maior acurácia de previsão.

## 2.5 Conclusões parciais

Este capítulo apresentou conceitos básicos sobre séries temporais e seus modelos de previsão. Explicou-se o processo de seleção e construção de um modelo, nos quais os parâmetros livres deste são estimados com base nos dados da série, de forma que o modelo passe a representar o processo estocástico que gerou tais dados. As convenções utilizadas para a construção desses modelos também foram analisadas, como a utilização do princípio da parcimônia, por exemplo.

Em seguida, as principais técnicas utilizadas na previsão de séries temporais foram apresentadas, abrangendo desde os métodos lineares até os métodos não-lineares, que envolvem técnicas de *soft computing*. Por fim, foi realizada uma breve revisão literária sobre os modelos de previsão, dando destaques para estudos relevantes nesta área de pesquisa.



## 3 REDES NEURAIS ARTIFICIAIS PARA ESTIMAÇÃO DE DEMANDA

### 3.1 Introdução

As redes neurais artificiais são métodos da classe de *soft computing*, cujo aprendizado e funcionamento são baseados nos dados históricos obtidos a partir do sistema a ser modelado (JHA; SINHA, 2014). Estas redes possuem inspiração biológica na capacidade cognitiva do cérebro humano de resolver problemas com base em experiências passadas. Para isso, as RNAs buscam mapear as relações e os padrões presentes no conjunto de dados que recebem, de forma a compreender o processo gerador destes dados. Assim, como outros modelos, o ajuste dos parâmetros da RNA, de acordo com o processo gerador dos dados, permite que seu modelo seja utilizado para prever o comportamento futuro do sistema modelado.

Atualmente, as RNAs são utilizadas em aplicações de diversas áreas do conhecimento, tais como engenharia, medicina, economia, ciências agrárias, entre outras. Nesses campos, as redes são aplicadas de diferentes formas, sendo utilizadas para classificação de padrões, mapeamento de relações, aproximação de funções e para previsão, sendo que esta última aplicação é a considerada neste trabalho (SILVA; SPATTI; FLAUZINO, 2015).

Neste capítulo, os principais conceitos sobre RNAs serão discutidos, abordando seus aspectos construtivos e seu funcionamento. Posteriormente, uma revisão dos principais tipos de RNAs utilizados na literatura é apresentada, introduzindo as redes neurais artificiais profundas. Em seguida, uma análise do estado da arte é feita, apresentando as principais pesquisas que aplicam as técnicas de RNAs e RNAP para a previsão de demanda de energia. A última seção do capítulo contém as conclusões e considerações sobre o tema abordado.

### 3.2 Aspectos básicos das RNA

As principais características das RNAs, que atraíram tanto a atenção dos pesquisadores ao longo dos anos são: sua capacidade de aprendizado e de adaptação por meio da experiência e sua capacidade de generalização. Em relação ao primeiro aspecto, a RNA extrai seu conhecimento a partir de padrões que lhe são apresentados, por esse motivo, considera-se que a rede é capaz de aprender. Este aprendizado ocorre por meio do ajuste dos parâmetros livres da RNA, que se adaptam aos exemplos que a rede recebe.

Depois da etapa de aprendizado, na qual a rede mapeia as relações existentes nos dados, o modelo gerado é capaz de responder tal qual o sistema modelado para diversas situações, mesmo que estas não tenham sido apresentadas para a rede durante o seu aprendizado. Esta capacidade é chamada de generalização, e expressa o quanto a rede é

capaz de extrapolar o conhecimento que já possui.

Os parâmetros livres principais da rede são seus pesos sinápticos, cujo ajuste é realizado durante a etapa de aprendizado da rede, que ocorre no processo chamado de treinamento. A fim de compreender o processo de treinamento e o funcionamento da rede, primeiramente deve-se entender a constituição de sua unidade básica, o neurônio artificial, cujo aspecto característico é mostrado na [Figura 1](#).

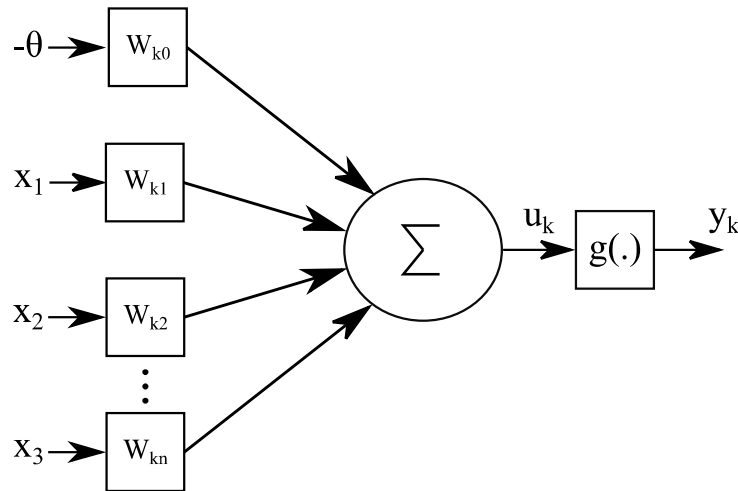


Figura 1: Modelo do neurônio artificial

O modelo do neurônio artificial é um modelo clássico que constitui até RNAs complexas. Em seu funcionamento, o neurônio recebe um vetor  $\mathbf{X}$  de parâmetros de entrada, além de receber a constante  $\theta$ , denominada de limiar de ativação. O vetor de entradas pode ser proveniente do meio que está sendo modelado ou pode ser a saída de outros neurônios da rede. Já o parâmetro  $\theta$  determina o patamar necessário para que o vetor de entrada produza uma determinada saída, o que é caracterizado por seu sinal negativo.

Em seguida, as entradas do neurônio (considerando também  $\theta$  como entrada, sendo a variável  $x_0$ ) são ponderadas por meio do vetor  $\mathbf{W}$ , que consiste no vetor de pesos sinápticos. O papel do  $\mathbf{W}$  é ponderar as entradas alterando o nível de importância de cada uma e os elementos deste vetor são os parâmetros livres da rede que são ajustados durante o aprendizado. Após serem ponderadas, as entradas são somadas, originando o sinal  $u_k$ , denominado de potencial de ativação e matematicamente descrito por:

$$u_k = \sum_{i=0}^n W_{ki} \cdot X_i \quad (3.1)$$

O potencial de ativação  $u_k$  servirá de argumento para a função de ativação  $g(\cdot)$  da rede. A saída desta função também será a saída  $y$  do neurônio, que pode tanto ser a saída

da RNA para o meio externo, quanto pode funcionar como entrada outro neurônio. Assim, a expressão de  $y$  em função de  $g(\cdot)$ , é expressa matematicamente por:

$$y_k = g(u_k) \quad (3.2)$$

Diversas funções matemáticas podem ser utilizadas como a função de ativação  $g(\cdot)$ , sendo que estas podem ser divididas em duas classes: funções parcialmente diferenciáveis e funções totalmente diferenciáveis, como mostrado na [Figura 2](#).

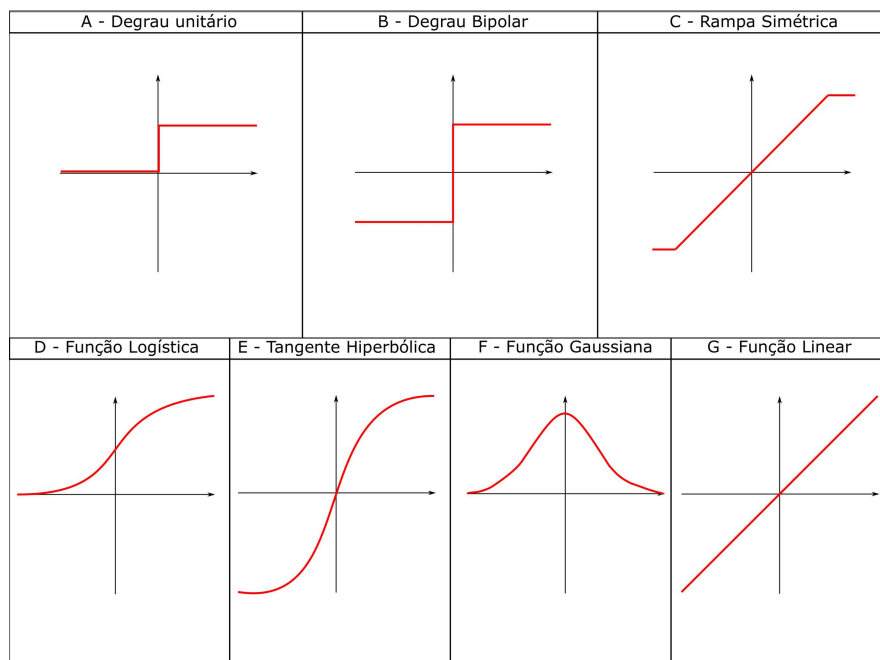


Figura 2: Funções de ativação para RNAs - A, B e C pertencem à classe de funções parcialmente diferenciáveis; D, E, F e G pertencem à classe de funções totalmente diferenciáveis

Na [Figura 2](#), as funções parcialmente diferenciáveis mostradas são: degrau unitário, degrau bipolar e rampa simétrica. Já na classe de totalmente diferenciáveis, tem-se as funções logística, tangente hiperbólica, gaussiana e linear. A seleção da classe de função utilizada depende do tipo de arquitetura de rede neural escolhido para a aplicação, enquanto que, a seleção de qual função dentro da classe depende do projetista e do problema a ser tratado.

Uma RNA normalmente é constituída de diversos neurônios, que são conectados entre si. A forma de conexão entre os neurônios, o número de neurônios e sua distribuição em camadas consecutivas, define o que se chama de topologia da RNA. Já a forma que ocorre o fluxo de informações entre os neurônios, determina a arquitetura das redes.

Basicamente, pode-se dividir as redes, de acordo com a sua arquitetura, em duas classes: *redes feedforward* e *redes recorrentes*. Sendo que o primeiro tipo caracteriza redes

cuja informação segue da camada de entrada até a camada de saída de forma direta e sem realimentações entre os neurônios. Ou seja, não há conexões entre neurônios de uma camada posterior com os de uma camada anterior. Além disso, também não há conexões entre neurônios de uma mesma camada. Já para as redes recorrentes, tais conexões existem, podendo ocorrer em diversos níveis, como conexões entre a camada de saída e a de entrada, ou entre a saída e as camadas intermediárias.

Além do tipo de arquitetura, as RNAs também podem ser classificadas em relação ao dinamismo presente em sua constituição, de forma que as redes podem ser estáticas ou dinâmicas. As RNA dinâmicas apresentam bons resultados quando aplicadas na previsão de séries temporais, pois levam em consideração a dependência dos dados em relação ao tempo. Desta forma, existem duas abordagens com RNA com dinamismo que podem ser utilizadas para modelagem de séries temporais (JHA; SINHA, 2014):

- Modelagem por meio de RNA recorrentes. Onde a dependência temporal é inserida na rede por meio de conexões de realimentação entre a camada de saída e as camadas anteriores;
- Modelagem por meio de uma RNA estática dotada de propriedades dinâmicas. Onde o dinamismo é inserido na rede através de memórias de curto ou longo prazo.

No segundo caso, pode-se citar a arquitetura Time-delay Neural Network (TDNN). Esta rede consiste em uma rede perceptron multicamadas (que se trata de uma RNA estática), na qual foi inserido uma memória de curto prazo através do uso de um atraso de tempo em sua camada de entrada (JHA; SINHA, 2014).

Por fim, as redes neurais também podem ser divididas entre redes superficiais e redes profundas. Mais detalhes sobre este tipo de redes serão apresentados nas próximas seções.

### **3.3 Paradigmas de aprendizado e construção do modelo de RNA**

Como ressaltado anteriormente, a RNA aprende com base em exemplos extraídos do sistema a ser modelado. As características deste conjunto de exemplos definem o tipo de aprendizado que será implementado, de forma que existem dois tipos principais de aprendizado: o aprendizado supervisionado e o não supervisionado. A diferença principal entre estes tipos, também chamados de paradigmas de aprendizado, é a de que para o primeiro, o conjunto de exemplos possui pares de entradas e saídas desejadas, enquanto que o segundo é composto apenas pelos valores das entradas.

No aprendizado supervisionado, a rede recebe um conjunto de entradas, gera a sua saída e pode compará-la com a saída que era desejada. Assim, a rede é capaz de saber o seu erro e de ajustar seus parâmetros de forma proporcional a este erro. Na versão não



---

supervisionada, a rede não sabe qual deveria ser sua saída e, conseqüentemente, não sabe o quão longe sua saída está do valor desejado. Para este tipo de abordagem, o ajuste dos parâmetros internos da rede é feito com base em outras métricas que não o erro da saída, métricas estas que priorizam as características internas do conjunto de entradas, como a da distância entre as amostras de entrada, por exemplo.

Estes paradigmas de aprendizado fazem parte da etapa de aprendizado da rede, também conhecida como etapa de treinamento. É nesta fase que a rede extrai todo seu conhecimento ajustando seu parâmetros internos. Em suma, o processo de ajuste de um modelo de RNA tem duas etapas principais: o treinamento, e o teste. Enquanto o primeiro foi explicado acima, o segundo visa determinar o nível de generalização da rede para situações novas, ou seja, para as quais ela não foi treinada.

Assim, após os ajustes dos parâmetros da rede durante o treinamento, esta é colocada em operação na fase de teste e suas saídas são avaliadas para entradas desconhecidas. Para realizar estes dois processos, o conjunto de dados do sistema é dividido em dois, o conjunto de treinamento e o de teste, sendo que as amostras presentes em um não podem estar no outro. Na literatura, esta divisão é feita normalmente separando de 60% a 90% para o conjunto de treinamento e o restante para o de teste.

Os parâmetros da RNA, i.e., seus pesos sinápticos e limiares são diretamente relacionados com a performance da rede, pois é nestes parâmetros que o conhecimento da rede é armazenado. Durante a etapa de treinamento, estes parâmetros são normalmente inicializados de forma aleatória, de forma que a cada inicialização, a rede irá realizar a busca de sua solução em uma região diferente da superfície da função erro. O problema associado à esta abordagem de inicialização é a convergência da rede para mínimos locais. Uma forma de contornar este problema é a definição dos parâmetros iniciais considerando características do sistema sendo modelado.

O desempenho do aprendizado da rede também é afetado pela escolha de sua topologia. Topologias com neurônios insuficientes tendem a apresentar um elevado erro durante a fase de treinamento, pois não são capazes de aprender as relações do problema. Este fenômeno é conhecido como *underfitting* e pode ser resolvido com o aumento do número de neurônios da rede. Todavia, o excesso de neurônios também é prejudicial, causando o fenômeno de *overfitting*, que é caracterizado pela redução significativa do erro de treinamento, mas do aumento expressivo do erro na fase de teste, o que significa que a capacidade de generalização da rede foi reduzida e que esta se tornou especialista no processo, decorando os padrões e os resultados.

Para evitar a ocorrência de tais fenômenos, utiliza-se a técnica de validação cruzada, na qual diversas topologias de rede são avaliadas e a rede com melhor desempenho global é selecionada. Neste processo, leva-se em conta a acurácia das redes e a parcimônia do modelo, assim, busca-se escolher o melhor modelo que também seja simples.

### 3.4 Redes neurais artificiais superficiais

As RNA superficiais consistem nas arquiteturas tradicionais de redes neurais, muito utilizadas para diversas aplicações nas últimas décadas. A rede neural perceptron foi a primeira RNA superficial criada, sendo um classificador linear composto de um neurônio artificial. O interesse pela rede perceptron impulsionou a pesquisa na área, permitindo a criação de outras arquiteturas superficiais, como as redes Adaline e posteriormente a rede perceptron multicamadas.

Neste trabalho, a rede perceptron multicamadas foi selecionada como método de benchmarking, para o problema de previsão de demanda. Tendo em vista o alto grau de sucesso desta rede em diversas aplicações de previsão, espera-se que a comparação com o nível de desempenho dela seja um bom indicador sobre o nível de sucesso ou de fracasso da metodologia proposta. A seguir, uma breve explicação sobre os conceitos básicos e o funcionamento desta rede é apresentada.

#### 3.4.1 Perceptron Multicamadas

A rede perceptron multicamadas (PMC) é uma das arquiteturas mais famosas de RNA, consistindo em uma rede feed-forward sem realimentações entre suas camadas ou seus neurônios. Esta arquitetura é caracterizada pela presença de uma camada de entrada, algumas camadas escondidas e uma camada de saída, como pode ser observado na [Figura 3](#). A camada de entrada não é uma camada neural e apenas recebe os dados de entrada provenientes do sistema. Tais dados são ponderados pelos pesos sinápticos existentes entre a camada de entrada e a primeira camada escondida e depois servem como entrada para os neurônios desta primeira camada escondida.

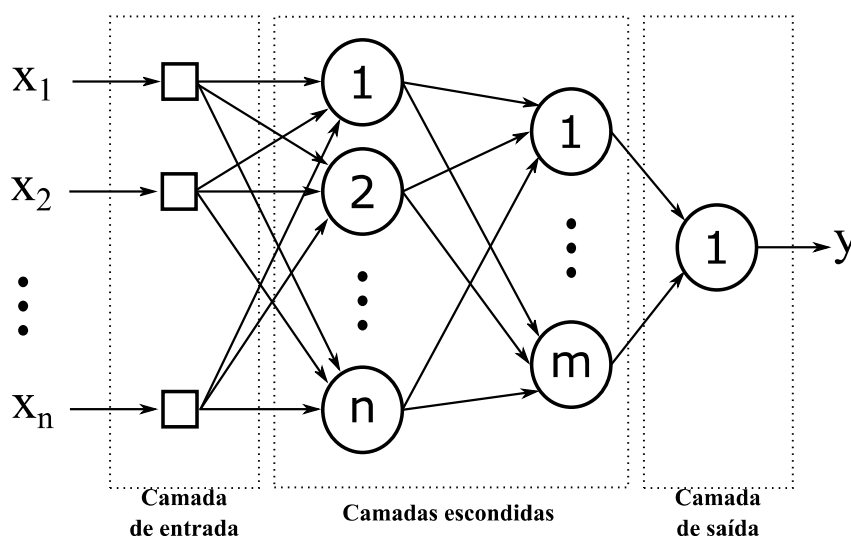


Figura 3: Estrutura geral de uma rede PMC

O número de camadas escondidas que constituem uma rede PMC, assim como o

---

número de neurônios que compõe tais camadas, depende do projetista e normalmente é selecionado de forma empírica. Esta escolha de topologia, usualmente é realizada pelo o processo de validação cruzada. Neste, várias topologias distintas são treinadas e testadas, variando-se também os dados definidos para os conjuntos de treinamento e de teste, a escolha é realizada com base na topologia com maior sucesso na etapa de teste, ou seja, a topologia com maior capacidade de generalização para dados novos.

O treinamento da rede PMC é realizado de forma supervisionada, sendo que o algoritmo de aprendizado tradicional deste tipo de rede é o algoritmo *Back-propagation* (BP). O funcionamento deste algoritmo consiste em duas etapas:

- Fase *forward*: Nesta etapa, a rede, cujos parâmetros livres foram iniciados aleatoriamente, recebe em sua camada de entrada uma amostra do conjunto de treinamento. Esta amostra passa pela rede seguindo o fluxo da camada de entrada até a camada de saída. Na camada de saída, a rede apresenta a saída calculada para a amostra recebida e o erro existente entre a saída obtida e a saída desejada para aquela amostra é calculado;
- Fase *backward*: Nesta fase, o erro calculado retorna à rede para ajuste dos pesos sinápticos. Com base no gradiente do erro, o algoritmo ajusta inicialmente os pesos da última camada escondida. Em seguida, os pesos das camadas escondidas anteriores são ajustados, sendo que, os pesos de uma camada sempre serão ajustados em relação aos pesos da camada imediatamente posterior.

O processo descrito repete-se para todas as amostras do conjunto de treinamento e este conjunto é reapresentado à rede diversas vezes até o fim do treinamento. Cada vez que o conjunto de treinamento completo é apresentado para a RNA, conta-se uma época de treinamento. O processo de treinamento de uma PMC utilizando o algoritmo BP normalmente utiliza muitas épocas, pois este processo somente é encerrado quando o erro verificado na saída da RNA é inferior à tolerância do problema, ou quando o número máximo de épocas é atingido.

O algoritmo BP é o algoritmo base para o treinamento de redes PMC, todavia, este método apresenta problemas. Os principal problema do algoritmo BP é a sua convergência lenta e a possibilidade de ocorrência de instabilidade da rede em torno de um mínimo da função erro.

Para contornar estes problemas, muitos outros algoritmos foram propostos, dentre eles, pode-se citar o algoritmo de Levenberg-Maquardt (LM), utilizado para a implementação das redes neste trabalho. Este algoritmo atua combinando o método de gradiente descendente (utilizado pelo BP) e o método de Gauss-Newton. A mudança do algoritmo

entre os métodos ocorre com base na localização na superfície da função erro e garante uma convergência mais rápida (HAGAN; MENHAJ, 1994).

#### 3.4.1.1 Time-delay neural network

A *Time-delay neural network* (TDNN) é uma rede que pode ser construída a partir de uma rede perceptron multicamadas (PMC) quando se estrutura seu conjunto de dados para que as entradas da rede sejam os instantes passados da variável que se deseja prever (SILVA; SPATTI; FLAUZINO, 2015). A configuração TDNN permite à rede PMC assumir um comportamento dinâmico, devido ao atraso temporal inserido em sua camada de entrada, relacionando as amostras de acordo com a dependência temporal presente entre elas.

A estrutura desta rede é apresentada na Figura 4, onde o número de neurônios na primeira camada escondida é  $n$  e o tamanho do vetor de entrada é  $np$ , ou seja,  $np$  instantes passados serão utilizados para prever o instante presente. O treinamento da rede é realizado de forma idêntica ao treinamento da rede PMC, sendo que a única diferença entre as duas é o conjunto de dados, no qual as saídas desejadas da rede passam a ser suas próprias entradas futuras.

Escolheu-se esta arquitetura como método de benchmarking pela sua boa performance quando aplicada para a previsão de sistemas que apresentam dependência temporal entre suas variáveis.

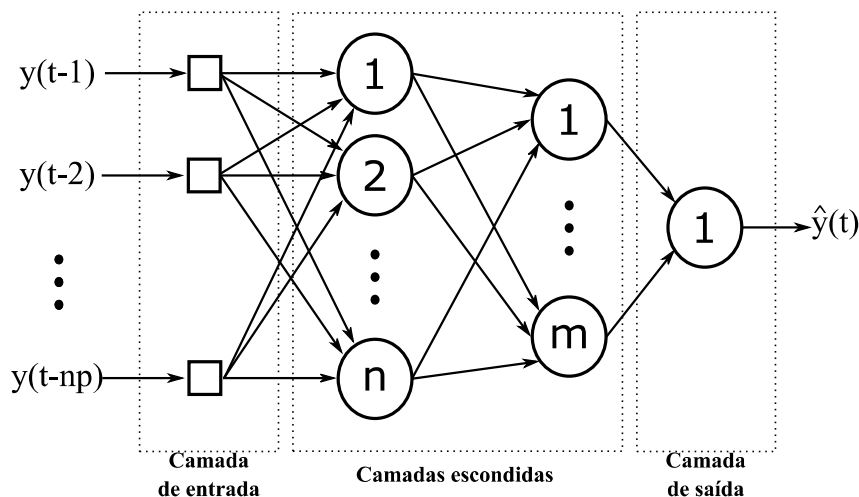


Figura 4: Estrutura geral de uma Time-delay neural network

### 3.5 Redes neurais profundas

O conceito de aprendizado profundo, ou *deep learning* é explicado por (GOOD-FELLOW; BENGIO; COURVILLE, 2016) como uma abordagem de aprendizado para que modelos computacionais possam aprender com base em experiências e entender o

---

mundo através de conceitos hierárquicos. O termo de profundo advém desta hierarquia, que caracteriza que cada conceito pode ser definido a partir da combinação de outros conceitos mais simples, o que cria uma cadeia de aprendizado profunda. Desta forma, o computador é capaz de aprender conceitos complexos, inferindo-os a partir dos mais simples.

Neste campo de inteligência artificial, que visa dotar o computador de capacidades que imitam as do ser humano, um problema encontrado foi o de como traduzir o conhecimento informal e intuitivo do ser humano em uma linguagem inteligível para a máquina. Ou seja, como programar o computador para este aprender a realizar tarefas triviais, porém subjetivas, presentes no dia a dia dos seres humanos. Para solucionar este problema, o conceito de machine learning prega que a máquina deve obter seu próprio conhecimento extraíndo-o a partir de dados que lhe são apresentados.

Desta forma, o conhecimento é extraído do conjunto de dados que é apresentado à máquina. Este conjunto de dados pode ser formado por diversos atributos que estão presentes no sistema sobre o qual se quer aprender. O papel da máquina é encontrar as relações entre estes atributos para poder inferir sobre o comportamento deste sistema no futuro.

Neste trabalho, optou-se pela utilização de uma abordagem de rede neural profunda, devido a esta maior capacidade de generalização e de abstração de conhecimento (LECUN; BENGIO; HINTON, 2015). A arquitetura escolhida para compor a RNAP foi a da rede Autoencoder, esta será treinada com o conjunto de dados para extrair conhecimento sobre as relações presentes neste conjunto. Uma rede neural profunda formada pelas camadas enconder já treinadas será construída e depois o treinamento (*fine tuning*) desta nova rede será realizado, partindo dos parâmetros já ajustados pelas autoencoders. Espera-se que esta abordagem contorne problemas comuns de RNAs, como a inicialização aleatória de parâmetros, que leva à convergência para mínimos locais e a seleção empírica de topologias, que pode causar problemas de generalização.

### 3.5.1 Redes Autoencoder

As redes da arquitetura Autoencoder têm seu funcionamento baseado basicamente em copiar sua entrada em sua saída. Pode-se considerar que este tipo de rede é formado por dois blocos complementares, um bloco de codificação e um bloco de decodificação, sendo estes unidos pela camada escondida da rede, como apresentado na [Figura 5](#).

Uma rede autoencoder não é projetada para aprender com exatidão como copiar sua entrada em sua saída, ao invés disto, deseja-se que a rede consiga extrair do conjunto de dados apenas as relações e informações que sejam relevantes para obtenção de um bom grau de reconstrução. Este processo permite selecionar aspectos do conjunto de dados que realmente são importantes para sua caracterização, ignorando informações redundantes ou

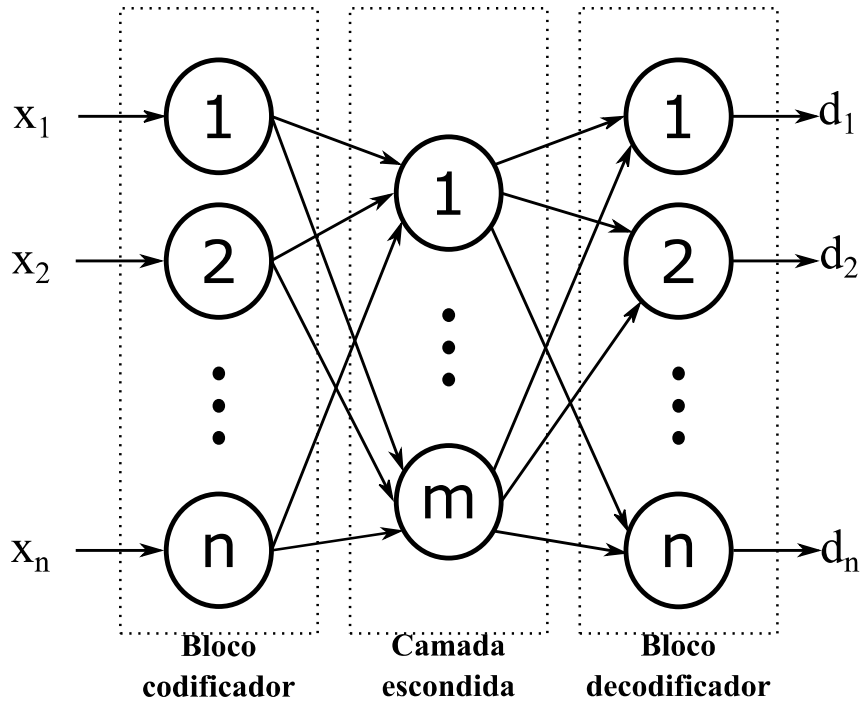


Figura 5: Rede autoencoder genérica

pouco necessárias (GOODFELLOW; BENGIO; COURVILLE, 2016).

Se for permitido à rede, a partir dos parâmetros que lhe são fornecidos, ter um desempenho muito bom na tarefa de copiar as amostras da entrada para a saída, é capaz da rede não aprender nada significativo sobre o conjunto de dados utilizado. Desta forma, nota-se que o interesse durante o treinamento da rede está no nível de conhecimento contido na camada escondida e não na saída da rede.

Tendo em vista este objetivo, uma abordagem é a de definir a dimensão da camada escondida de forma que esta seja menor que a dimensão do vetor de entradas. Uma rede autoencoder com estas características é chamada de *undercomplete* (GOODFELLOW; BENGIO; COURVILLE, 2016).

Considerando este tipo de rede, o aprendizado se torna um processo de otimização cujo objetivo é minimizar a função de custo apresentada na equação (3.3). O objetivo desta função é penalizar a saída da rede ( $g^{(2)}(g^{(1)}(x))$ ) por ser diferente da entrada ( $x$ ). Neste tipo de treinamento, indica-se o uso de funções não lineares para  $g^{(1)}$  e  $g^{(2)}$ , por permitirem que a rede aprenda mais sobre os dados, incluindo relações não lineares.

$$L(x, g^{(2)}(g^{(1)}(x))) \quad (3.3)$$

Para as *undercomplete* autoencoders, o aprendizado pode ser prejudicado caso os blocos de codificação e de decodificação recebam parâmetros que permitam a reconstrução exata do conjunto de dados. Este problema de aprendizado pobre também pode ocorrer em

outras situações, como nos casos da dimensão da camada escondida ser igual ou superior à da camada de entrada. Nestes casos, a rede é capaz de representar a entrada na saída com exatidão, até mesmo utilizando funções de ativação lineares. No último caso, onde a dimensão da camada escondida é superior, a rede autoencoder é denominada de overcomplete (GOODFELLOW; BENGIO; COURVILLE, 2016). Para contornar os problemas utilizam-se as redes autoencoder regularizadas. Estas constituem uma abordagem que obriga a rede a aprender sobre os dados por meio da alteração da função custo e não através de limitações de capacidade dos blocos da rede ou de reduções da dimensão da camada escondida.

Uma das alterações que podem ser feitas é a de adicionar à função custo uma penalidade baseada na esparsialidade da reconstrução obtida pela rede, esta abordagem gera as redes autoencoders esparsas. Neste contexto, a função custo pode ser escrita da seguinte forma:

$$L(x, g^{(2)}(g^{(1)}(x))) + \Omega(h) \quad (3.4)$$

Onde  $\Omega(h)$  é a penalidade por esparsialidade da representação na camada escondida  $h$ . O bloco codificador pode ser formado por inúmeras camadas, no exemplo da [Figura 5](#) optou-se por uma única camada por motivos de didática, o número de neurônios na camada de entrada da rede é igual ao número de atributos do vetor de entrada ( $x$ ). Após esta camada inicial, o vetor de entradas ( $x$ ) com dimensão  $n$  é transformado no vetor ( $z$ ), com dimensão  $m$ , que representa a saída da camada escondida. Este vetor  $z$  possui dimensão reduzida em relação à  $x$ . Desta forma, infere-se que  $m < n$ , este processo garante uma redução dimensional do vetor de atributos. O vetor  $z$  pode ser matematicamente expresso pela equação (3.5).

$$z = g^{(1)}(W^{(1)}x + b^{(1)}) \quad (3.5)$$

Onde  $g$  é a função de transferência,  $W$  é a matriz de pesos sinápticos e  $b$  é o vetor de potenciais de ativação dos neurônios.

Posteriormente, este vetor de dimensão reduzida é utilizado como entrada para o bloco decodificador da rede. Este bloco tem como objetivo receber como entrada o vetor de dimensão reduzida e recuperar o vetor de entradas original em sua saída. Neste trabalho o vetor de saída será referido como vetor reconstruído,  $d$ , que é descrito matematicamente pela equação (3.6).

$$d = g^{(2)}(W^{(2)}x + b^{(2)}) \quad (3.6)$$

O treinamento da rede ocorre com base na métrica de erro obtida com base na diferença entre o vetor reconstruído e o vetor original. Todavia, ressalta-se que o treinamento deste tipo de rede é do tipo não-supervisionado, ou seja, não utiliza amostras de entradas com suas respectivas saídas desejadas. Todavia, com base no funcionamento da rede, sabe-se que a saída desejada para uma determinada amostra de entrada é a própria amostra.

### 3.6 Redes Neurais Artificiais para previsão de séries temporais

Nesta seção, serão abordados alguns, dentre os inúmeros, trabalhos relevantes utilizando as técnicas de RNAs. Para análise de estudos de revisão sobre a aplicação de tais técnicas, recomenda-se as referências (RAZA; KHOSRAVI, 2015; HIPPERT; PEDREIRA; SOUZA, 2001). Além disso, recomenda-se o estudo desenvolvido por (FALLAH et al., 2019), que faz uma rica pesquisa bibliográfica sobre as metodologias utilizadas em estudos de previsão de demanda em curto prazo utilizando técnicas de *soft computing*. De acordo com o autor, na literatura os trabalhos podem ser divididos em quatro metodologias distintas, são estas: a de dia similar, a de seleção de variáveis, a previsão hierárquica e a seleção da estação climática. A seleção da metodologia de previsão utilizada depende de fatores como disponibilidade de dados históricos, horizonte de previsão, nível de precisão dos dados disponíveis, nível de precisão necessário para a previsão, entre outros. Além disso, a metodologia diz respeito a forma de implementação do método de previsão e não apenas à técnica utilizada (FALLAH et al., 2019). Neste trabalho utiliza-se a metodologia de seleção de variáveis, pois se considera a dependência temporal entre os elementos da série de potência. E esta metodologia é implementada através da técnica de RNAs.

No contexto de RNAs superficiais, em (GONZÁLEZ-ROMERA; JARAMILLO-MORÁN; CARMONA-FERNÁNDEZ, 2006) duas RNAs são utilizadas para realizar a previsão de uma série temporal de demanda de energia. O trabalho realiza uma divisão na série original, separando-a em uma série formada por sua componente de tendência e em uma série residual. Cada série foi prevista por uma rede e o resultado da previsão consistiu na soma dos resultados individuais das redes. Durante o trabalho, duas arquiteturas de RNAs foram avaliadas, a arquitetura PMC e a arquitetura de função de base radial. Os resultados do estudo mostram que a separação da série retorna melhores resultados do que a previsão realizada para a série original.

Em (BIANCHI et al., 2017) um estudo comparativo é feito, avaliando cinco diferentes arquiteturas de redes neurais recorrentes para a previsão da demanda. O estudo apresenta uma rica quantidade de resultados, tendo aplicado as redes a diferentes conjuntos de dados. Os resultados mostram que redes cujo treinamento é baseado no método do gradiente descendente, apresentam aprendizado mais lento e complexo, mas que após o treinamento, são capazes de obter um alto nível de conhecimento e de generalização sobre o



---

problema tratado. Além disso, também é indicado que redes com arquiteturas e construções mais complexas, apresentaram resultados semelhantes aos das mais simples, o que indica a falta de necessidade de recorrer a arquiteturas complexas em certos casos. Outro estudo comparativo também é encontrado em (DUDEK, 2016), onde cinco arquiteturas de RNAs são avaliadas para a previsão e depois comparadas com métodos clássicos, como o modelo ARIMA e o método de amortecimento exponencial.

Uma arquitetura de rede neural profunda é aplicada em (Chen et al., 2019). Neste estudo, o modelo de previsão é capaz de prever a demanda de 24 horas no futuro e se baseia em uma rede profunda residual. O método aplicado utilizou como entrada dados não tratados de demanda, temperatura e outras variáveis dependentes, apresentando uma boa capacidade de generalização para o problema de previsão, sendo testado com três diferentes conjuntos de dados. Além de realizar a previsão tradicional, em (Chen et al., 2019) também é realizada a previsão probabilística, adotando o uso do algoritmo de Monte Carlo no modelo.

O modelo de rede neural profunda também é utilizado em (KUO; HUANG, 2018), onde uma estrutura primeiramente realiza a seleção de atributos, através de camadas convolucionais, e posteriormente realiza a previsão de demanda utilizando uma estrutura de rede totalmente conectada. A performance do modelo proposto foi avaliada em comparação com cinco diferentes técnicas de *soft computing*, sendo elas: MVS, árvores de decisão, método de random forest, PMC e redes neurais do tipo long-short term memory. Os resultados do estudo indicam que para a previsão da demanda para um horizonte de três dias no futuro, o modelo de rede profunda proposto apresentou maior acurácia que os métodos tradicionais de *soft computing*.

Por fim, destaca-se dois estudos que apresentam um método similar ao desenvolvido nesta pesquisa. Em (TASNIM et al., 2017) e em (JIAO et al., 2018), redes autoencoder são utilizadas para realizar a previsão de potência eólica gerada. Em (TASNIM et al., 2017) as camadas da rede autoencoder são conectadas em forma de cascata para extrair informações do conjunto de entrada. Em seguida, estas informações são utilizadas como entrada para dois modelos de previsão, um modelo utilizando regressão linear e um modelo utilizando uma máquina de vetor de suporte. O estudo ainda compara a previsão realizada com as informações obtidas pela rede autoencoder com a previsão a partir de informações obtidas por meio de análise estatística, sendo que o método utilizando a rede autoencoder apresenta melhores resultados.

Em (JIAO et al., 2018), a previsão da potência eólica é realizada a partir de redes neurais profundas formadas por camadas encoder pré-treinadas. A etapa de ajuste da rede é feita utilizando o algoritmo de treinamento *back-propagation* e o diferencial do estudo é o uso do algoritmo de otimização *particle swarm optimization* para selecionar o número de neurônios das camadas encoder e o valor da taxa de aprendizado do algoritmo

de treinamento.

Considerando os trabalhos presentes na literatura, os diferenciais desta pesquisa se encontram na utilização da abordagem de construção de rede neural profunda por meio do uso de redes autoencoder para o problema de previsão de demanda de potência, pois este problema não foi analisado através desta abordagem na literatura; na implementação simples do método, que dispensa o uso de algoritmos de otimização para seleção de topologias e que ainda sim apresenta níveis de precisão melhores do que os dos métodos tradicionais; e na análise estatística feita na série temporal de demanda de potência para sua utilização no modelo de previsão. Sendo que esta análise também prepara os dados para sua utilização com outras técnicas de *soft computing*.

### **3.7 Conclusões parciais**

Neste capítulo, conceitos básicos sobre RNAs foram apresentados, incluindo sua constituição, paradigmas de aprendizado e aspectos considerados para a construção deste tipo de modelo. Em seguida, tratou-se da arquitetura PMC, escolhida como método de comparação, onde explicou-se sobre os principais métodos de treinamento desta rede e sobre a construção de sua estrutura dinâmica, que foi implementada neste trabalho.

A arquitetura da rede autoencoder, utilizada para construir a rede neural profunda proposta neste trabalho, também foi apresentada e seu funcionamento foi detalhado. Posteriormente, uma revisão de literatura apresentou trabalhos relevantes do tema. Esta revisão abordou estudos que possuem relação com o tema de pesquisa e o método proposto neste trabalho. Além disso, as contribuições deste trabalho para à área de estudo foram ressaltadas.

## 4 DESCRIÇÃO DOS DADOS - SÉRIE DE DEMANDA DE POTÊNCIA ATIVA

Neste trabalho, os métodos de previsão de séries temporais propostos serão analisados para o estudo de caso da previsão da demanda de potência ativa para um alimentador de uma subestação do estado de São Paulo, Brasil. Este capítulo apresentará as características desta série, analisando seu comportamento diante de diferentes fatores, tais como influência de feriados e das estações climáticas. Além de expor o processamento realizado nos dados para que estes pudessem ser utilizados nos modelos de previsão.

### 4.1 Análise da série temporal de demanda

O perfil da série temporal é apresentado na [Figura 6](#), sendo esta composta por 17520 medições, que constituem 365 dias. A obtenção dos dados foi realizada entre o período de 01 de Agosto de 2017 e 31 de Julho de 2018. Para cada dia, foram efetuadas 48 medições, com intervalos de tempos iguais de 30 minutos, considerando o período entre às 00h00m39s e às 23h30m39s.

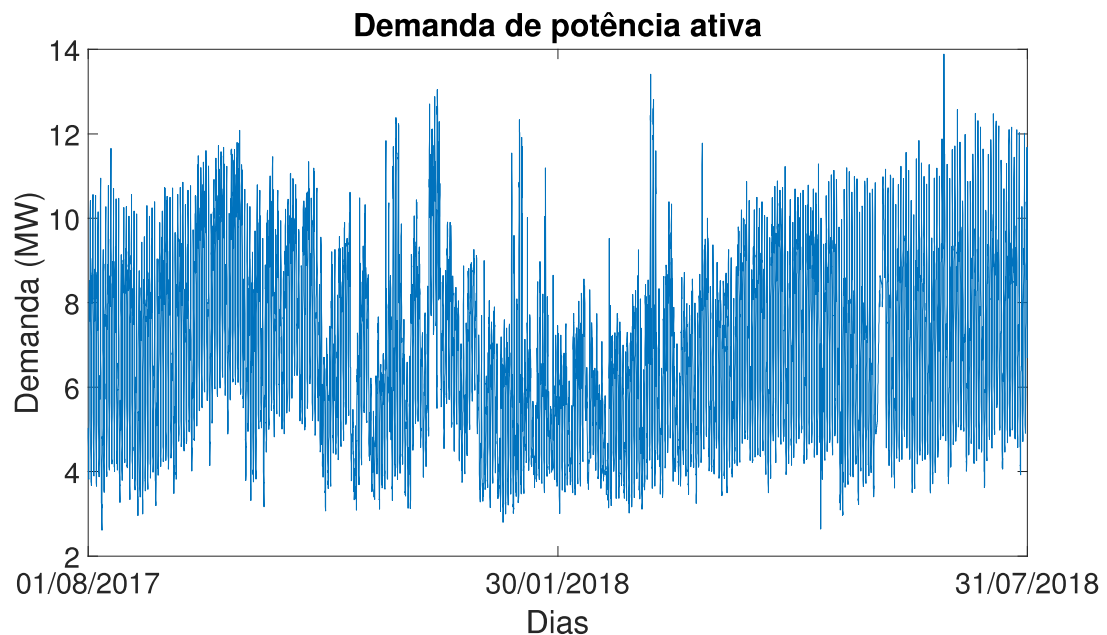


Figura 6: Demanda do alimentador para o período de 01/08/2017 a 31/07/2018

Esta série apresenta fatores de sazonalidade que influenciam seu comportamento e que variam de acordo com o horizonte de observação. Pode-se considerar uma sazonalidade mensal, ocasionada principalmente devido às mudanças climáticas entre os meses e ocorrência de feriados. Também existe uma sazonalidade semanal, na qual os dias da semana apresentam comportamentos característicos, principalmente devido a atividades

comerciais e industriais. Além disso, existe também a sazonalidade diária, na qual cada hora do dia apresenta um comportamento característico.

Para a análise das características do conjunto de dados, primeiramente realizou-se a caracterização dos dias presentes na série, retornando os seguintes padrões: 261 dias de semana; 104 dias de final de semana; 11 feriados. Com base nesta caracterização, foi possível observar os padrões e os perfis presentes na série.

Uma série de demanda com perfil residencial apresenta uma curva de carga diária que se inicia com um baixo índice de demanda no período entre as 00h00 e 06h00, sobe para um patamar médio durante o período entre 06h00 e 17h00 e tem um pico de consumo por volta das 18h00 até às 21h00. Uma curva de carga com perfil comercial e industrial, por sua vez, também apresenta uma demanda baixa entre as 00h00 e as 06h00, todavia, a partir deste horário, a demanda atinge um patamar elevado que se mantém até as 18h00.

Analisando o perfil de consumo diário apresentado na [Figura 7](#), infere-se que o perfil de demanda tem características industriais, comerciais e residenciais. Esta conclusão deve-se aos picos de consumo de energia no início e no meio do dia, que são ocasionados pelo comércio e pela indústria, e ao pico característico no início da noite, proveniente das unidades consumidoras residenciais. A elevação de demanda durante o dia pode ser devida ao acionamento de máquinas elétricas e ao uso extensivo de equipamentos de conforto térmico no meio do dia, por indústrias e comércios, respectivamente.

Tendo em vista este perfil, considerou-se a hipótese de que a ocorrência de feriados poderia afetar o perfil da demanda que seria normal para aquele dia da semana, principalmente devido à mudança dos hábitos das pessoas e do funcionamento do comércio e da indústria. Assim, espera-se que o nível de demanda caia devido ao não funcionamento da indústria e do comércio para um feriado. Para avaliar esta hipótese, foram observados os perfis dos feriados ocorridos no período coberto pela série. Na [Figura 8](#) é apresentado o perfil de uma semana no qual ocorre um feriado na Quinta-feira. Ao comparar a demanda desta quinta-feira com a demanda da Quinta-feira da [Figura 7](#), nota-se que há uma redução significativa da demanda solicitada no alimentador para o dia com feriado. Ainda, pode-se observar que o perfil de demanda para a Quinta-feira com feriado se assemelha com o perfil apresentado para os dias de final de semana, nos quais o único pico de demanda expressivo ocorre no final do dia, se assemelhando a uma curva de carga tradicionalmente residencial.

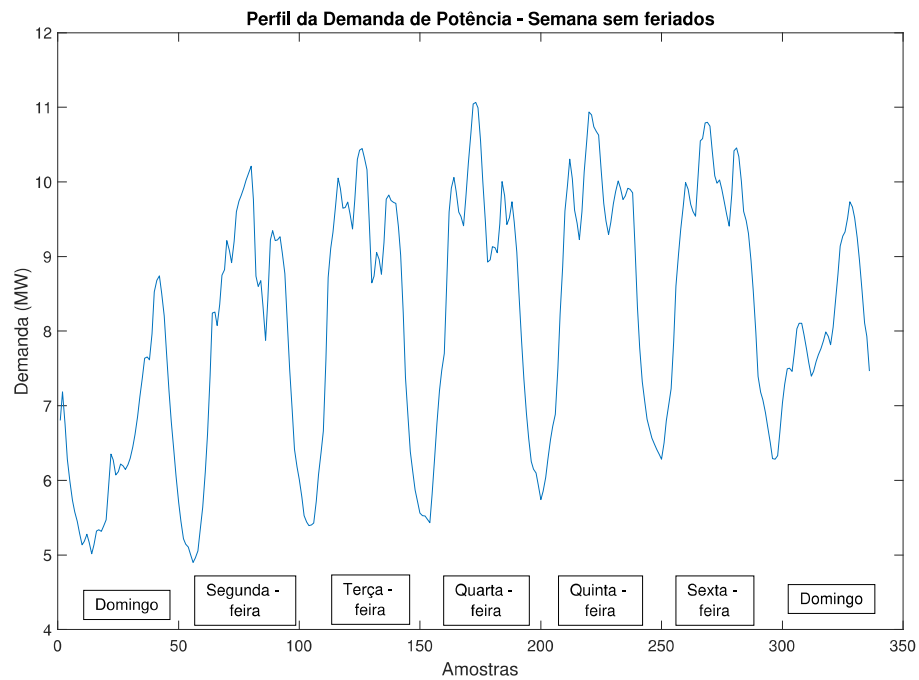


Figura 7: Perfil da demanda do alimentador para uma semana sem feriados - Semana de 15/10/2017 à 21/10/2017

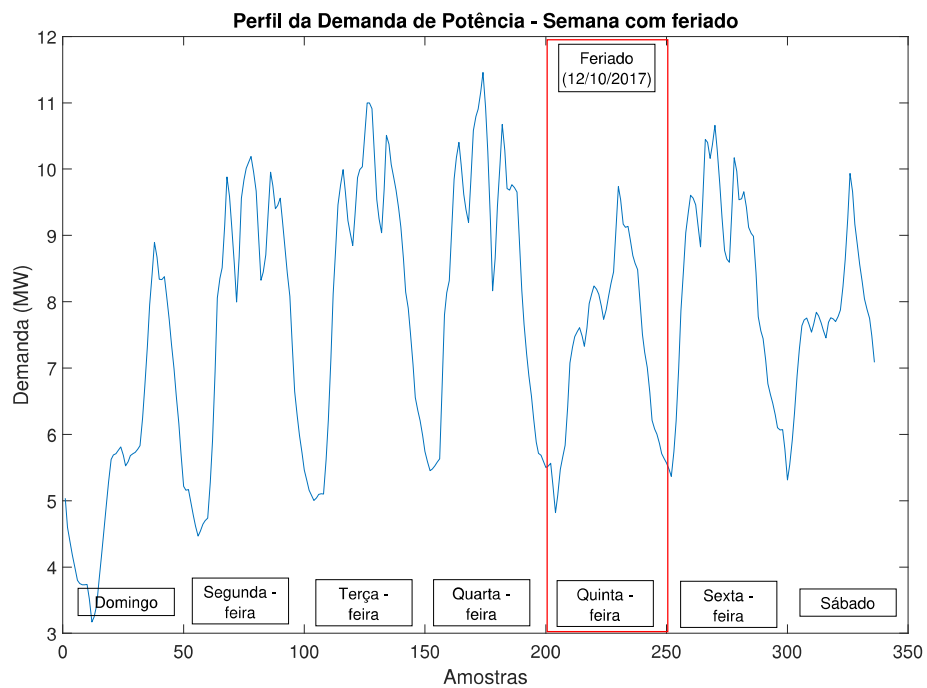


Figura 8: Perfil da demanda do alimentador para uma semana com um feriado - Semana de 08/10/2017 à 14/10/2017

Além disso, também observou-se a influência das estações climáticas sobre os

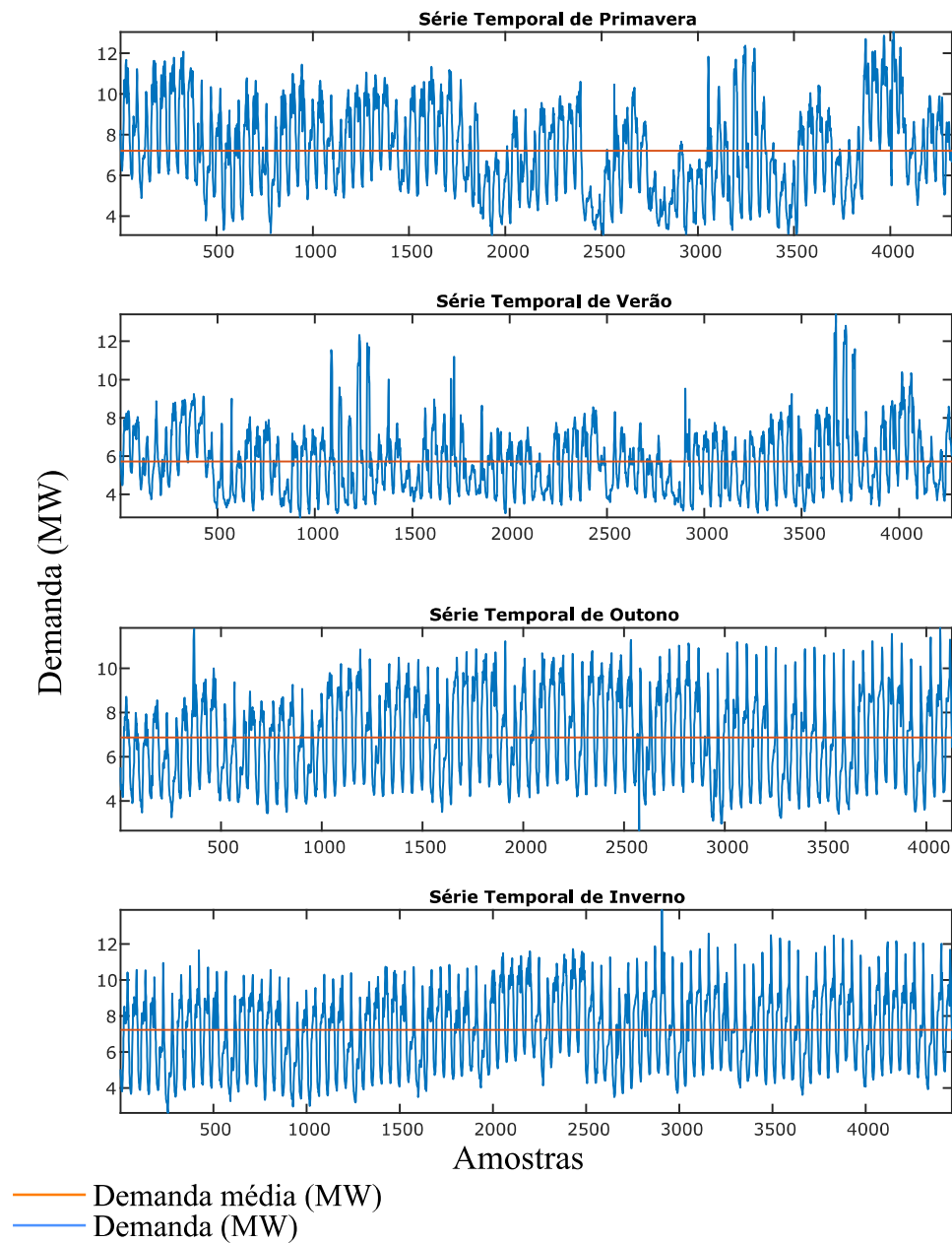


Figura 9: Séries temporais para cada estação climática.

níveis de demanda, tendo em vista a existência de uma componente sazonal dependente das estações. Assim, os dias da série também foram classificados em relação às estações climáticas, considerado o início de cada estação como 22/09/2017, 21/12/2017, 20/03/2018 e 21/06/2018, para a primavera, verão, outono e inverno, respectivamente. Os gráficos da [Figura 9](#) apresentam o perfil de cada uma das quatro séries, sendo que a linha laranja representa a média da demanda em MW. Além disso, os dados estatísticos de cada uma das séries estão reunidos na [Tabela 1](#).

Para verificar a hipótese de influência das estações, foram analisados os perfis de demanda para as diferentes estações climáticas. A [Figura 10](#) apresenta o comportamento

Tabela 1: Dados estatísticos das séries temporais

Série Temporal	Nº de Amostras	Média (MW)	Desvio Padrão (MW)
Série temporal de primavera	4320	7.21	2.14
Série temporal de verão	4272	5.73	1.67
Série temporal de outono	4464	6.88	1.96
Série temporal de inverno	4464	7.22	2.09

da demanda para quatro Segundas-feiras típicas, cada uma analisada para uma das quatro estações. Ressalta-se que um dia típico de uma determinada classe é considerado como um dia formado pela média de todos os dias daquela classe.

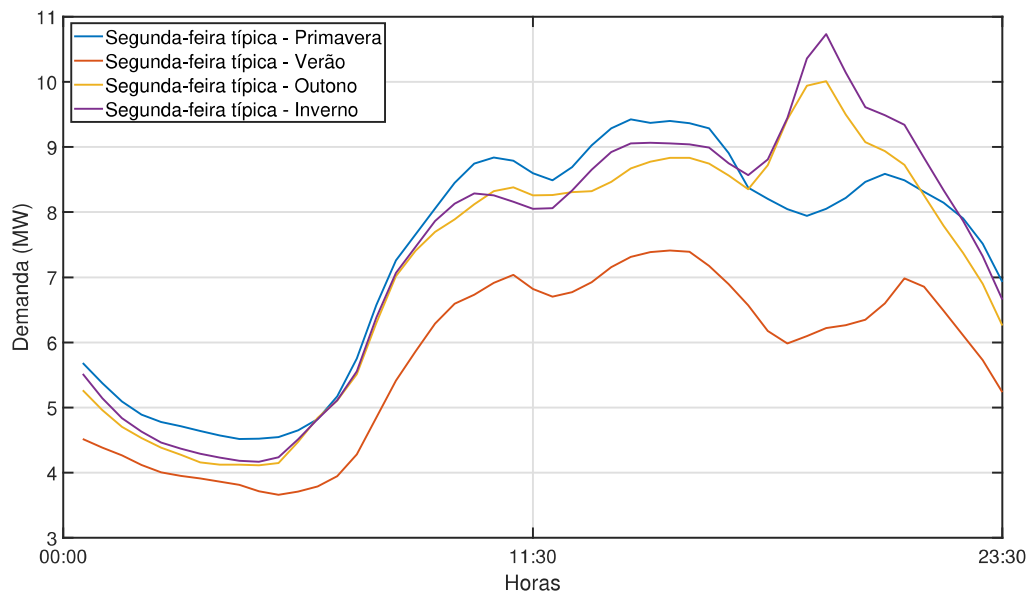


Figura 10: Comparação da demanda do alimentador para dias de diferentes estações

Infer-se da [Figura 10](#) que a demanda de energia é mais elevada para as estações com temperatura mais amena, inverno e outono. Isto pode decorrer do fato de que neste período o uso do chuveiro elétrico é elevado, o que agrega mais um elemento de consumo no perfil residencial, afetando principalmente o pico de energia no início da noite. Além disso, a análise do gráfico condiz com os dados apresentados na [Tabela 1](#), que indicam que os dias de verão possuem uma média de consumo menor, enquanto que a série de inverno e de primavera possuem consumos mais elevados.

Por fim, para avaliar o nível de relação temporal entre os instantes presentes e os

passados das séries das estações, realizou-se a análise das funções de autocorrelação e de autocorrelação parcial. As figuras 11 e 12 apresentam, respectivamente, os autocorrelogramas obtidos para estas funções.

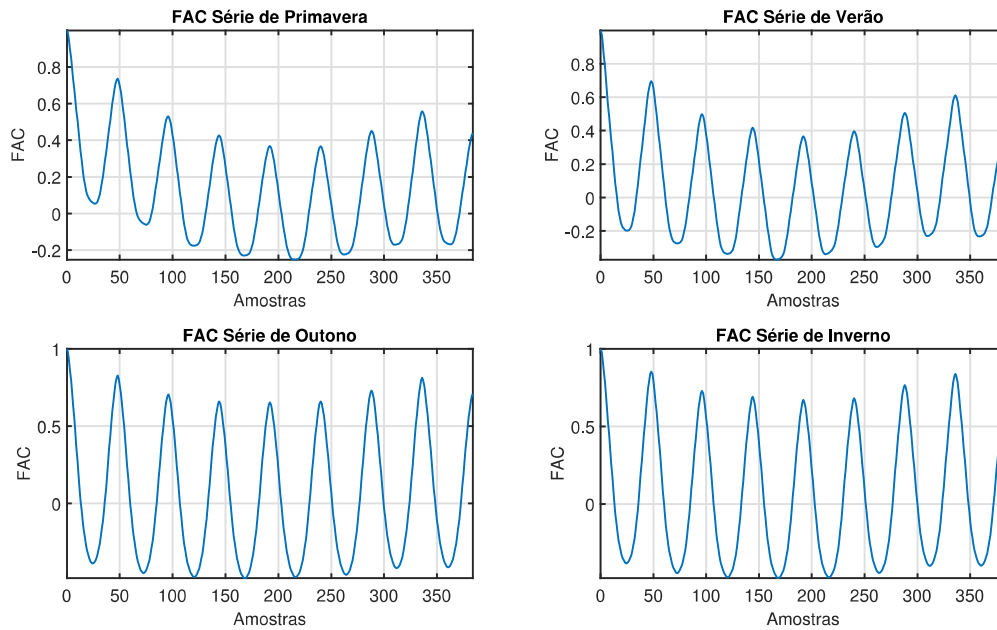


Figura 11: Função de autocorrelação avaliada para as séries das estações

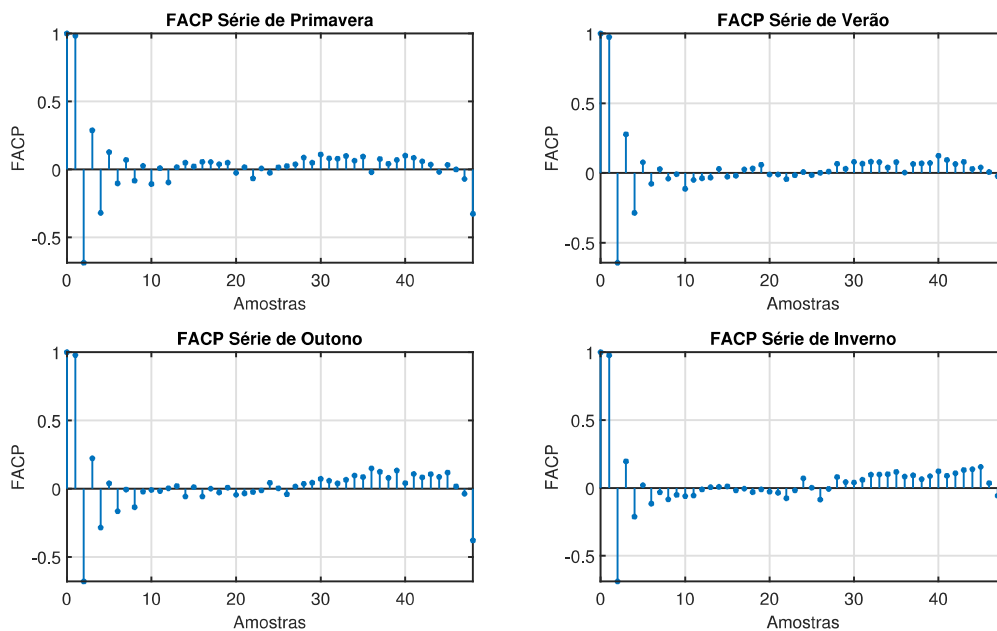


Figura 12: Função de autocorrelação parcial avaliada para as séries das estações



Todos os autocorrelogramas apresentam a ocorrência de picos para certos *lags*, cuja presença é atribuída à existência de sazonalidade nos dados. Desta forma, os coeficiente de autocorrelação serão maiores para os que são múltiplos da frequência sazonal.

O fator de sazonalidade é mais expressivo de forma semanal e diária, sendo que a sazonalidade diária possui frequência igual a 48, e pode ser observada no *lag 48* e em seus múltiplos. A influência da sazonalidade semanal é vista no *lag 336* e em seus múltiplos, neste *lag*, além da influência da sazonalidade diária, a semanal também é observada. Isto indica que a relação entre a demanda para às 20h00 de duas Segundas-feiras de semanas consecutivas é maior do que a relação para às 20h00 de uma segunda e às 20h00 de uma Quarta-feira da mesma semana, como é visto nos gráficos.

O coeficiente de autocorrelação calculado entre os instantes  $y_t$  e  $y_{t-k}$  é composto pelo valor direto de autocorrelação entre estes instantes e pelos valores indiretos referentes à dependência linear com os instantes  $y_1$  até  $y_{t-k+}$ . A função de autocorrelação parcial apresentada na [Figura 12](#) denota a autocorrelação para o *lag k* quando estas dependências lineares são removidas. A análise dos gráficos da FACP indica que o instante presente da série apresenta uma relação significativa com os instantes imediatamente anteriores, o que já era esperado, tendo em vista que, às vezes, a melhor previsão de uma variável para um instante é o valor da variável para o instante passado. Esta abordagem geralmente ocorre para séries totalmente estocásticas, como a série de passeio aleatório (HYNDMAN; ATHANASOPOULOS, 2018).

## 4.2 Processamento da série temporal de demanda

Esta seção apresentará os processos de tratamento de dados realizados na série de demanda para que esta pudesse posteriormente ser modelada pelo método de previsão proposto.

### 4.2.1 Normalização e remoção de dados corrompidos

Inicialmente, dados corrompidos ou com características anômalas foram retirados da série ou substituídos. Esta situação ocorreu para três casos, sendo estes:

- Para o dia 17/02/2018, foram realizadas medidas duplicadas para os instantes de 23h00m39s e 23h30m39s. Como solução, optou-se por considerar as médias das medidas duplicadas de cada instante como a medida daquele determinado instante;
- Para o dia 15/10/2017, não foram aferidos os instantes de 00h00m39s e 00h30m39s. Para contornar este problema, foram atribuídos a estes instantes os valores médios obtidos a partir destes instantes no dia anterior e no dia posterior ao dia da falha. Esta abordagem foi considerada pois, para estes instantes, todos os perfis de dias apresentam níveis de demanda similares;

- A semana do dia 03/06/2018 até 09/06/2018 apresentou comportamento anômalo, caracterizando uma possível falha do equipamento de medição. Por este motivo, optou-se por retirar tais dados da série, pois sua presença poderia afetar negativamente o processo de aprendizado do método de previsão, que poderia extrair e considerar as características do comportamento anormal como parte das características do sistema.

Os problemas relatados acima podem ser categorizados como erros grosseiros e provavelmente ocorreram por mau funcionamento momentâneo do equipamento de medição. Após a remoção e correção dos dados anormais, realizou-se a normalização da série, que consiste em um processo importante quando se trabalha com algoritmos de aprendizado. O primeiro motivo é que, quando se possui algoritmos com diversos atributos de entrada, a normalização dos dados equaliza o peso destes atributos. Ou seja, se um atributo tiver uma ordem de grandeza de  $10^3$  e o outro de  $10^{-6}$  e eles não forem normalizados, o aprendizado será tendencioso, dando mais importância para o atributo com maior ordem de grandeza, o que não é desejável.

O segundo motivo para se realizar a normalização dos dados, é o de adequar os dados para a faixa de variação dinâmica das funções utilizadas no algoritmo de aprendizado. Este é o motivo que tornou necessária a normalização neste estudo, visto que apenas um atributo de entrada é utilizado, a demanda. Esta necessidade decorre do fato de que, durante seu aprendizado, os métodos de previsão, como os de RNA, utilizam funções matemáticas que possuem saturação a partir de certos valores. Exemplos destas funções são a função tangente hiperbólica e a logística, cujas entradas saturam, entre -1 e 1. Por este motivo, deve-se normalizar os dados entre estes limites de saturação. Considerando que serão utilizadas técnicas de RNA, este tipo de normalização foi feita, seguindo a equação (4.1), baseada no teorema de Tales (SILVA; SPATTI; FLAUZINO, 2015).

$$z = 2 \cdot \left( \frac{x - x_{min}}{x_{max} - x_{min}} \right) - 1 \quad (4.1)$$

Antes da normalização, a série de demanda original foi dividida nas quatro séries das estações climáticas apresentadas na Figura 9. Após a separação, as quatro séries foram normalizadas. Assim, cada uma das séries terá seu próprio modelo de previsão e o aprendizado do modelo ocorrerá levando em consideração a diferença de perfil existente entre elas. Após a normalização das séries, estas passaram por um processo de decomposição para que posteriormente pudesse se realizar a previsão. A seção a seguir apresenta os detalhes da metodologia utilizada para esta decomposição.

#### 4.2.1.1 Decomposição da série temporal

A decomposição da série temporal é necessária para melhorar o desempenho do modelo de previsão, mesmo este sendo um modelo de RNAs. Neste trabalho, optou-se

por decompor a série com base na metodologia desenvolvida por (AGUIRRE et al., 2008). Nesta metodologia, a série temporal é dividida em duas parcelas, uma representando o comportamento típico e esperado da série e outra representando as variações devidas à fatores como sazonalidade ou agentes externos. A primeira parcela resulta em uma série composta por dias médios e a segunda, em uma série residual.

Assim, primeiramente obteve-se a série média construída a partir dos dias típicos da série original. Com base nos diferentes perfis de demanda identificados, foram criados modelos de dias típicos para cada dia da semana e para cada estação, resultando em 28 modelos, como por exemplo, *Segunda-feira típica de primavera e Sábado típico de verão*. Para determinar cada um dos dias típicos foi utilizada a média da demanda de todos os dias deste perfil existentes em uma determinada estação. Ou seja, para construção do modelo da *Segunda-feira típica de primavera* foram calculadas as médias das demandas para cada instante de todas as segundas-feiras da primavera. Estas médias foram ordenadas ao longo de um dia para formar o dia típico da segunda-feria de primavera.

Em seguida, os modelos de dias típicos foram ordenados de acordo com os dias da semana, formando quatro séries de dias típicos (séries médias), uma para cada estação. Após obter as séries médias para cada estação, estas foram retiradas das respectivas séries originais, resultando em quatro séries residuais. Nas figuras 13, 14, 15 e 16 pode-se observar os gráficos para as séries original, média e residual para cada estação.

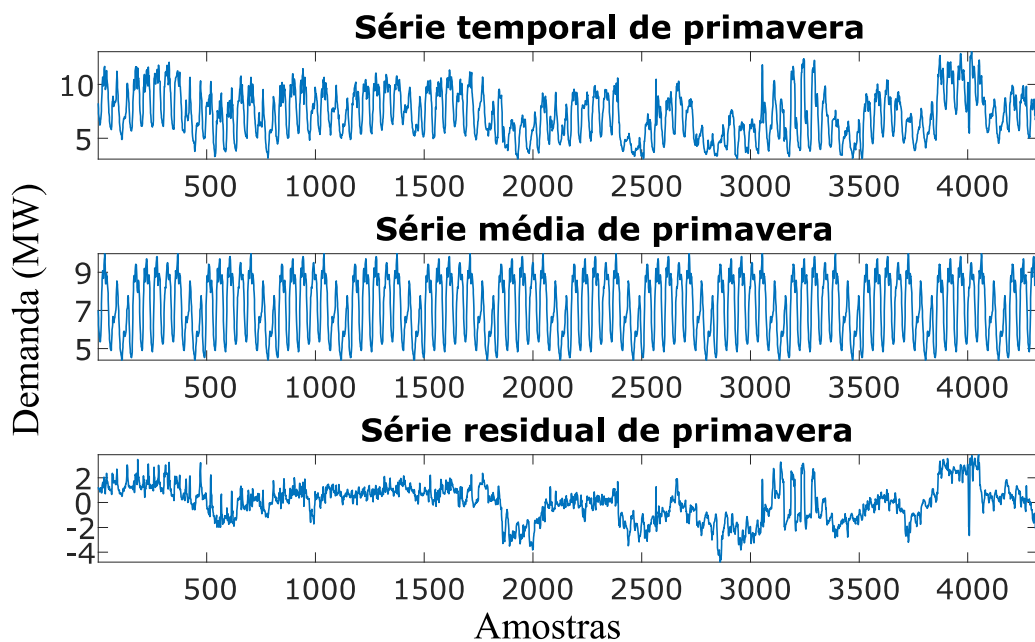


Figura 13: Decomposição da série temporal de primavera

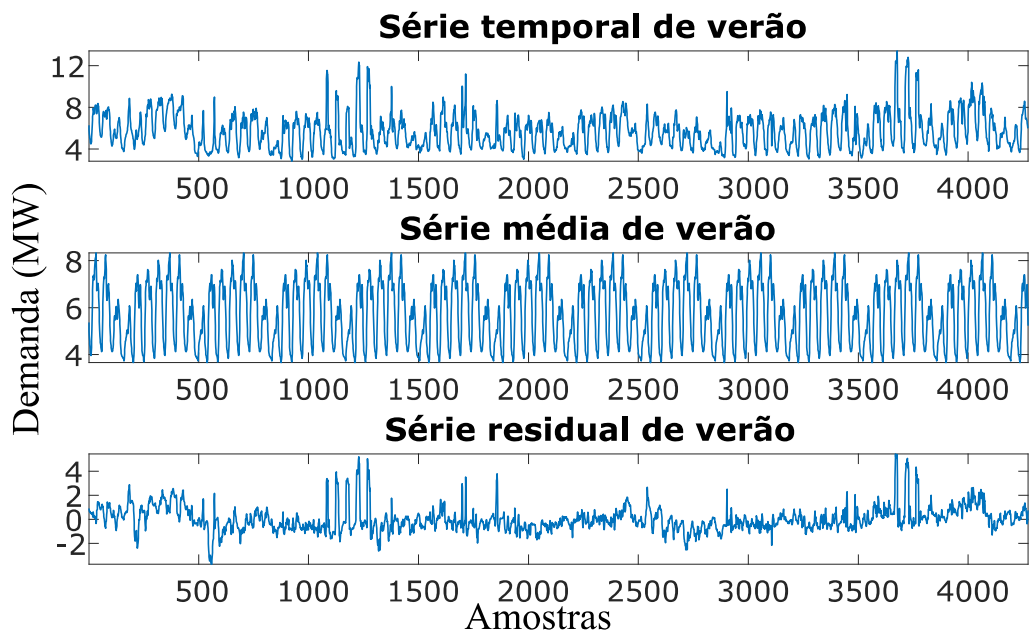


Figura 14: Decomposição da série temporal de verão

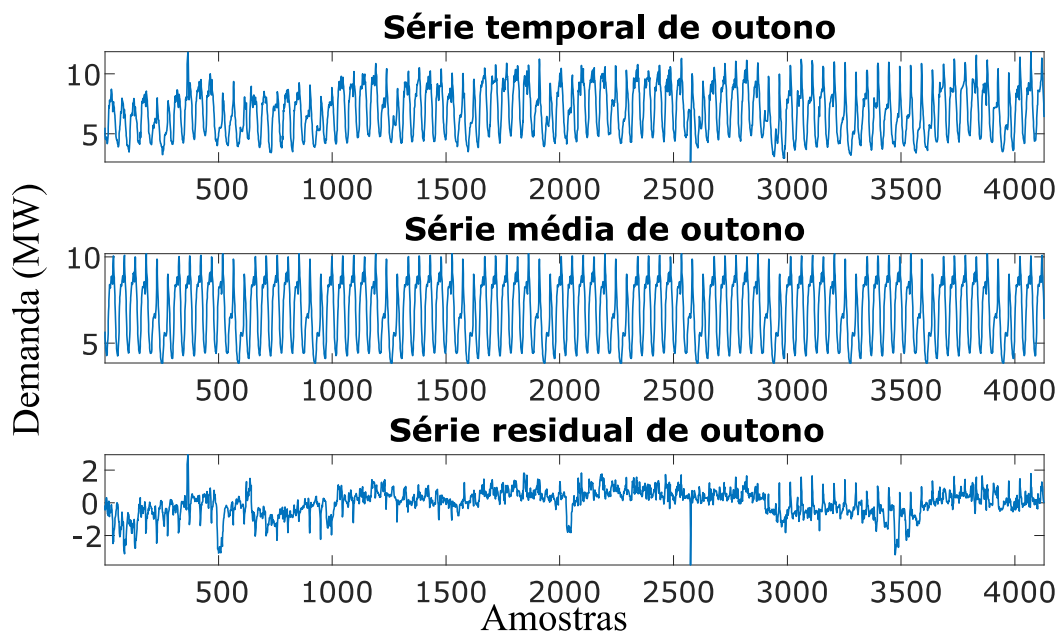


Figura 15: Decomposição da série temporal de outono

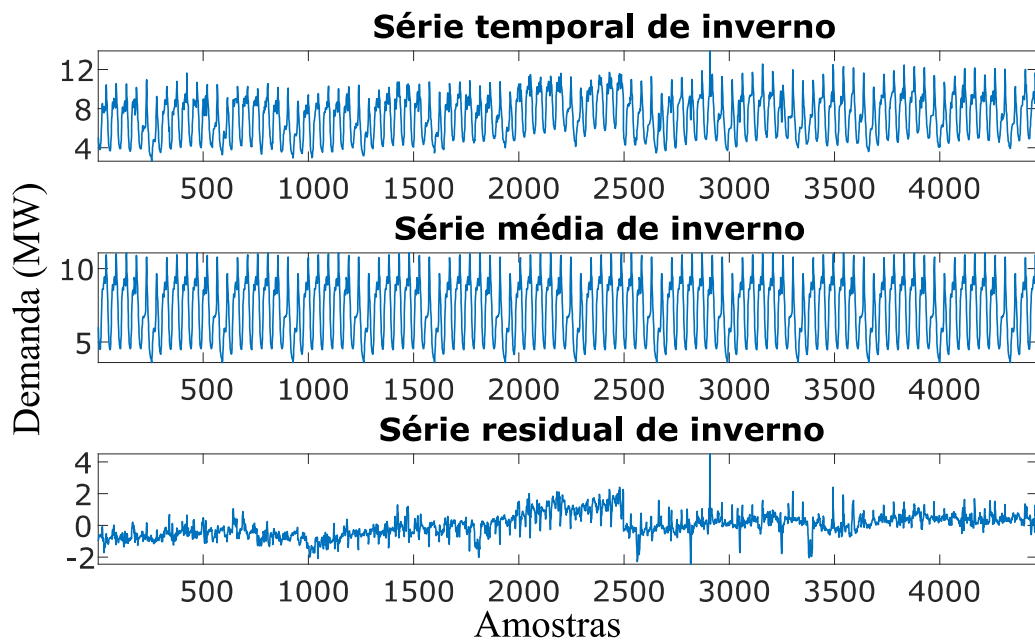


Figura 16: Decomposição da série temporal de inverno

De acordo com (AGUIRRE et al., 2008), a etapa de definição da série média é denominada de *pattern mapping* e, para alguns casos, este é o melhor modelo de previsão para a série original que se irá encontrar. Esta situação ocorre quando a série residual possui comportamento puramente estocástico, o que indica que não há como criar um modelo para prever seus valores futuros. Neste tipo de situação, um simples modelo linear é capaz de prever a componente média da série, e esforços adicionais para construção de modelos complexos, como um modelo de RNA, são dispensáveis.

Por este motivo, deve-se testar a série residual para descobrir se existem aspectos determinísticos em seu comportamento e se, conseqüentemente, deve-se construir um modelo não linear capaz de prever tais aspectos. O teste de determinismo da série residual foi feito seguindo o procedimento descrito em (AGUIRRE et al., 2008), consistindo em um teste de hipóteses, no qual a hipótese  $H_0$ , que se desejava rejeitar, é a de que a série seria não-determinística.

Para o teste de hipótese seguiu-se o seguinte procedimento:

1. Primeiramente, criou-se uma série puramente estocástica para a qual a hipótese seria verdadeira, para isto, os elementos da série residual original foram permutados de forma aleatória;
2. Em seguida, realizou-se a previsão da série estocástica e da série residual utilizando uma RNA do tipo perceptron multicamadas;

3. Observou-se o erro de previsão para ambas as séries, caso o erro de previsão para a série estocástica fosse muito maior que o erro para a série residual, a hipótese poderia ser rejeitada e a existência de determinismo na série residual poderia ser considerada.

Com base no teste, foi possível determinar a existência de determinismo nas quatro séries temporais residuais. Com isto, pôde-se prosseguir para a construção do modelo de previsão. O processo de decomposição das séries temporais de demanda é resumido no fluxograma apresentado na [Figura 17](#).

#### 4.2.2 Construção dos conjuntos de treinamento e de teste

Após o teste de determinismo, o primeiro passo para a criação do modelo de previsão, considerando que serão utilizadas técnicas de redes neurais artificiais, foi a criação dos conjuntos de treinamento e de teste.

Primeiramente, foi feita a divisão dos dados originais de cada série entre os conjuntos. Optou-se por utilizar 80% das amostras para o treinamento e o restante para o teste, este valor foi escolhido com base na literatura ([SILVA; SPATTI; FLAUZINO, 2015](#)). Como o treinamento das RNA selecionadas é realizado de forma supervisionada, o conjunto de treinamento deve ser composto por dados que sejam pares de entradas e saídas desejadas, como explicado no [Capítulo 3](#).

Além disso, apenas os valores passados de demanda de potência ativa serão utilizados como entrada, sendo que a quantidade  $np$  de amostras passadas utilizadas foi determinada de forma empírica, com base no teste de vários valores de  $np$  e na escolha da rede com melhor desempenho. Assim, considerando que o horizonte de previsão é de 30 minutos no futuro, ou seja, previsão de uma amostra a frente, as amostras de treinamento seguem o formato apresentado em [\(4.2\)](#).

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-np)) \quad (4.2)$$

### 4.3 Considerações parciais

Neste capítulo realizou-se uma análise da série temporal de demanda utilizada como estudo de caso do trabalho. Inicialmente, avaliou-se o perfil da série, averiguando suas características de consumo próprias de uma região com setores de indústria e comércio ativos. Além disso, verificou-se a influência de fatores externos, tais como a ocorrência de feriados e a variação das estações climáticas. Em relação às estações climáticas, optou-se por dividir a série original em quatro séries, uma para cada estação, a fim de retirar a influência das estações dos aspectos que deveriam ser previstos pelo modelo. Em seguida, apresentou-se o método de decomposição aplicado nas séries temporais para que estas

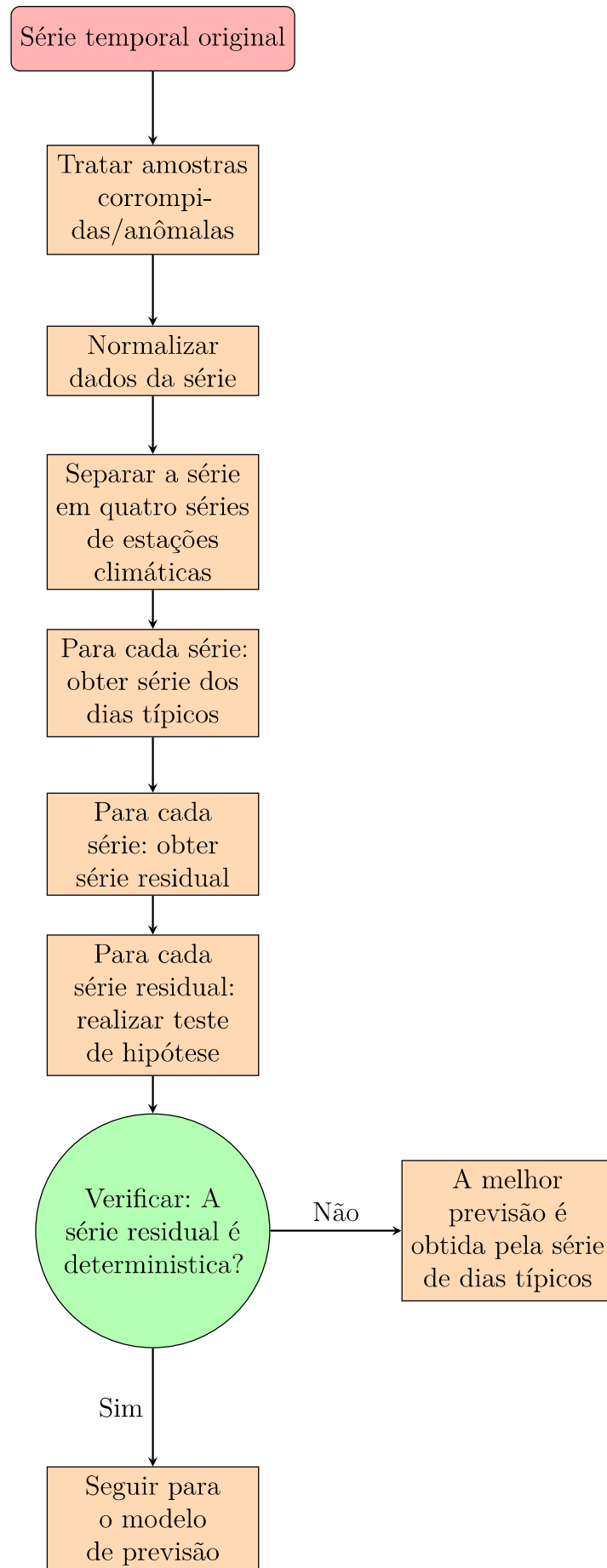


Figura 17: Fluxograma do método utilizado para o tratamento e a decomposição da série temporal de demanda de potência ativa.

pudessem ser previstas pelo modelo de RNAs. Por fim, realizou-se uma breve descrição sobre a criação dos conjuntos de treinamento e de teste.



## 5 CONSTRUÇÃO DOS MODELOS DE PREVISÃO

Este capítulo é dividido em quatro seções e visa apresentar de forma detalhada a metodologia utilizada no trabalho. A primeira seção introduz os conceitos de indicadores de desempenho e discorre sobre técnicas e métricas utilizadas para avaliar métodos de previsão. Estes indicadores serão posteriormente utilizados para avaliar os modelos propostos neste trabalho.

Em seguida, a construção do conjunto de treinamento e de teste, abordada na seção anterior, é detalhada. Nas seções seguintes, é descrita a implementação dos modelos de redes neurais artificiais, abordando tanto a arquitetura superficial, quanto a profunda.

### 5.1 Indicadores de desempenho

Uma etapa importante do processo de previsão de séries temporais é a avaliação da acurácia da previsão. Para este propósito, usualmente utilizam-se os indicadores de desempenho, de forma que seja possível comparar e avaliar diferentes métodos e modelos por meio da análise de um indicador em comum. Esta seção apresentará as características dos indicadores de desempenho utilizados nesta pesquisa, sendo eles: as métricas de erro, a análise do histograma do erro e a avaliação da função de autocorrelação.

#### 5.1.1 Métricas de erro

Usualmente, para os métodos de previsão de séries temporais, os indicadores utilizados são as métricas de erro. Sendo as mais comuns indicadas a seguir:

- Erro médio (*mean error (ME)*, em inglês)

$$ME = \frac{1}{N} \sum_{k=1}^N x(k) - \hat{x}(k) \quad (5.1)$$

- Erro médio absoluto (*mean absolute error (MAE)*, em inglês)

$$MAE = \frac{1}{N} \sum_{k=1}^N |x(k) - \hat{x}(k)| \quad (5.2)$$

- Erro percentual absoluto médio (*mean absolute percentage error (MAPE)*, em inglês)

$$MAPE = \frac{1}{N} \sum_{k=1}^N \left| \frac{x(k) - \hat{x}(k)}{x(k)} \cdot 100 \right| \quad (5.3)$$

- Erro médio percentual (*mean percentage error (MPE)*, em inglês)

$$MPE = \frac{1}{N} \sum_{k=1}^N \frac{x(k) - \hat{x}(k)}{x(k)} \cdot 100 \quad (5.4)$$

- Erro quadrático médio (*mean square error (MSE)*, em inglês)

$$MSE = \frac{1}{N} \sum_{k=1}^N (x(k) - \hat{x}(k))^2 \quad (5.5)$$

- Raíz do erro quadrático médio (*root mean square error (RMSE)*, em inglês)

$$RMSE = \sqrt{MSE} \quad (5.6)$$

Onde, nas equações de (5.1)-(5.6),  $x(k)$  indica o elemento  $k$  da série temporal e  $\hat{x}(k)$  é o valor previsto para este elemento. Os indicadores apresentados nestas equações são muito utilizados para comparar métodos de previsão. Todavia, o projetista também está interessado em outros aspectos, tais como saber a distribuição dos níveis de erro entre as amostras e verificar se seu método conseguiu prever o melhor possível todas as características determinísticas da série. Para fornecer essas informações, os métodos apresentados a seguir podem ser utilizados.

### 5.1.2 Diagnóstico dos resíduos

Para iniciar esta análise, primeiramente é interessante definir dois conceitos, o de resíduos e o de ruído branco. Dentro do contexto de modelos de previsão, o resíduo é definido como a diferença entre o valor esperado para uma determinada variável e o valor previsto para esta, ou seja, o resíduo é o erro do modelo de previsão e pode ser matematicamente descrito como:

$$e_t(k) = x(k) - \hat{x}(k) \quad (5.7)$$

Considerando o desejo de encontrar um modelo de previsão que se ajuste da melhor forma possível ao processo gerador da série, espera-se que os resíduos encontrados após a previsão sejam livres de aspectos determinísticos que não foram detectados pelo modelo. Mais precisamente, deseja-se que os resíduos sigam o comportamento de um ruído branco. Sendo que um ruído branco é definido como uma série cujos elementos não apresentam correlação entre si e que possui média zero e variância finita (HYNDMAN; ATHANASOPOULOS, 2018).

Além do comportamento semelhante ao de um ruído branco, outros aspectos que podem ser analisados, mas que não são estritamente necessários, tais como a distribuição dos resíduos ser similar à normal e sua variância ser constante. Desta forma, um modelo de previsão cujos resíduos não seguem tal padrão deve ser melhorado, pois a existência de autocorrelação indica a existência de relações restantes entre os dados e a presença de uma média não nula sugere que o modelo de previsão está enviesado. Para verificar a média dos resíduos e a existência de relações, são utilizados o histograma dos resíduos e a função de autocorrelação, respectivamente.

### 5.1.2.1 Histograma dos resíduos

O histograma dos resíduos é um gráfico que indica a distribuição da frequência com a qual cada valor de erro apareceu durante a previsão, considerando todas as amostras. Para um método com bons níveis de acurácia, espera-se um histograma com distribuição próxima à distribuição normal, o que indica que, para a maioria das amostras, a diferença entre o valor previsto e o esperado é próxima de zero. Neste caso, a média dos resíduos é próxima de zero, como para um ruído branco, e indica que o modelo de previsão utilizado não está enviesado.

### 5.1.3 Função de autocorrelação dos resíduos

Para verificar a aderência do modelo de previsão à série temporal, utilizou-se a análise da função de autocorrelação dos resíduos da previsão, obtidos na etapa de teste do método. Como dito anteriormente, deseja-se que a FAC dos resíduos seja similar à FAC de uma série de ruído branco, cujo correlograma possui o aspecto apresentado na [Figura 18](#).

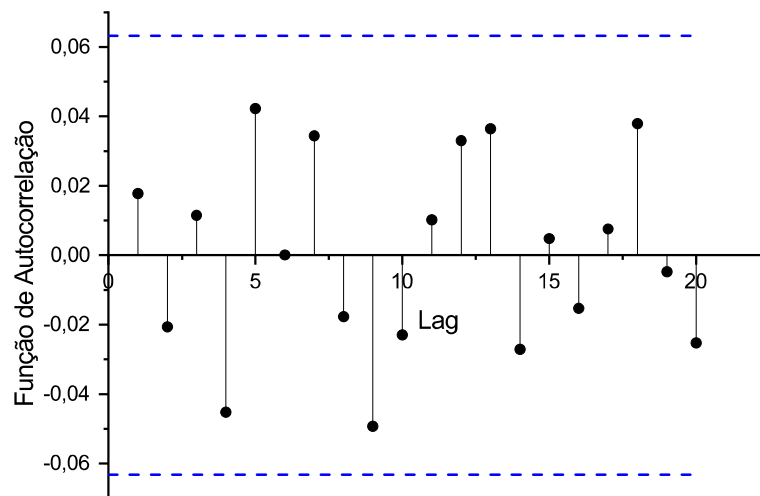


Figura 18: Função de autocorrelação para uma série de ruído branco

Neste gráfico, nota-se que as FAC para todos os atrasos não nulos, estão dentro do intervalo de confiança, ou seja, são próximos de zero. Para uma série ser considerada como ruído branco, esta deve possuir pelo menos 95% de seus coeficientes dentro do intervalo de confiança, cujos limites podem ser calculados por  $\alpha = \pm \frac{2}{\sqrt{N}}$  (HYNDMAN; ATHANASOPOULOS, 2018). Caso a série temporal dos erros do teste seja considerada um ruído branco, pode-se afirmar que o modelo realizou a previsão de toda a parcela da série que era possível de ser prevista. Sendo, o restante é puramente estocástico, ou seja,

não pode ser previsto. Para a análise da aderência dos modelos de previsão, o coeficiente  $\beta$  foi proposto, sendo que este representa a porcentagem de coeficientes da FAC que excedeu os limites definidos por  $\pm\alpha$ .

## 5.2 Construção do conjunto de treinamento e de teste

A previsão da série de demanda caracterizada no [Capítulo 4](#) será realizada por meio de RNAs do tipo perceptron multicamadas. Como abordado anteriormente, este tipo de arquitetura é treinada de forma supervisionada, necessitando de um conjunto de treinamento composto por amostras constituídas de pares de entrada e saída desejada. O conjunto de teste por sua vez, deve possuir o mesmo número de atributos do conjunto de treinamento, mas não exige o componente de saídas desejadas. Entretanto, para fins de análise do desempenho da rede, o conjunto de teste utilizado também apresenta suas respectivas saídas desejadas.

O conjunto de dados original utilizado na construção dos conjuntos de treinamento e de testes é a própria série temporal de demanda de energia. Esta série é composta por 17184 elementos e seu formato geral é apresentado a seguir:

$$\text{Série de demanda} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_{17182} \\ x_{17183} \\ x_{17184} \end{bmatrix} = \begin{bmatrix} 5.038 \\ 4.803 \\ 4.575 \\ 4.365 \\ \vdots \\ 4.274 \\ 4.472 \\ 4.685 \end{bmatrix} \quad (5.8)$$

Como dito anteriormente, a separação dos dados foi realizada considerando 80% para o conjunto de treinamento e o restante para o conjunto de teste. Desta forma, o nível de aprendizado da rede será testado a partir das amostras inéditas na etapa de teste. O número de atributos que as RNAs receberam como entrada foi variado de acordo com o parâmetro  $np$ . Este parâmetro definiu quantas amostras passadas seriam consideradas na previsão do instante atual. Desta forma, para exemplificar o formato do conjunto de treinamento, considerando arbitrariamente  $np = 2$ , tem-se a seguinte matriz de entradas:

$$\mathbf{x} = \begin{bmatrix} x_{t-np} & x_{t-1} & \dots & x_{N-np} \\ x_{t-1} & x_t & \dots & x_{N-1} \end{bmatrix} = \begin{bmatrix} 5.038 & 4.803 & \dots & 4.274 \\ 4.803 & 4.575 & \dots & 4.472 \end{bmatrix} \quad (5.9)$$

Onde cada coluna representa uma amostra de entrada e cada linha representa um atributo. Como o treinamento é supervisionado, cada amostra da matriz de entradas

possui sua respectiva saída desejada. As saídas desejadas formam o vetor  $\mathbf{d}$ , apresentado a seguir:

$$\mathbf{d} = [x_t \quad x_{t+1} \quad \dots \quad x_N] = [4.575 \quad 4.365 \quad \dots \quad 4.685] \quad (5.10)$$

Sendo  $\mathbf{d}$  um vetor com o mesmo número de colunas que a matriz  $\mathbf{x}$  e  $N$  é o número total de elementos da série selecionados para a etapa de treinamento. Com este procedimento, o número de padrões de treinamento variou de acordo com a escolha de  $np$ , seguindo a seguinte regra:

$$n_{treinamento} = N - np \quad (5.11)$$

Para a criação dos conjunto de treinamento e teste para as RNAs de cada estação, seguiu-se o mesmo procedimento.

### 5.3 Implementação da rede neural TDNN

A rede neural TDNN consiste em uma rede perceptron multicamadas utilizada na configuração de TDNN através da inserção de um atraso temporal na camada de entrada da rede, como explicado no [Capítulo 3](#). Para a implementação desta rede utilizou-se o software matemático Matlab, através da função *feedforwardnet*, que cria uma RNA da arquitetura PMC. Para configurar esta rede como uma TDNN, foi feita uma alteração no conjunto de treinamento para inserir o atraso temporal.

O processo de treinamento da rede ocorreu dentro de uma função criada para a pesquisa denominada *tdnntreinamento*. Esta função executou a transformação do conjunto de treinamento e de teste para inserir o atraso temporal, posteriormente criou a rede PMC com o uso da função do matlab *feedforwardnet*, treinou esta rede com o conjunto construído por meio da função *train* do Matlab e, por fim, testou a rede e avaliou sua performance através da função nativa *perform*. O conjunto de dados utilizado para o treinamento e teste da rede é o composto pela série temporal residual, cuja obtenção é detalhada no capítulo anterior.

Para evitar problemas como *underfitting*, *overfitting* e convergência para mínimos locais, várias topologias diferentes foram testadas. Para cada valor de  $np$  avaliado, foram criadas cinco topologias, sendo que cada uma destas foi treinada e testada cinco vezes. Os resultados apresentados na [Capítulo 6](#) representam o melhor treinamento da melhor topologia para cada  $np$  avaliado.

Para a seleção da melhor topologia, avaliou-se a métrica de erro MAPE, considerando os valores de previsão da série residual somados aos valores da série média. Desta

forma, o valor do erro foi avaliado para o resultado final do modelo de previsão, e não apenas para a previsão da parcela residual da série original.

#### 5.4 Implementação da rede neural profunda

Como explicado anteriormente, a rede neural profunda proposta neste trabalho será composta por camadas encoder, provenientes de redes autoencoder treinadas, e por uma camada final de uma rede PMC. A estrutura do método proposto é apresentada na Figura 19.

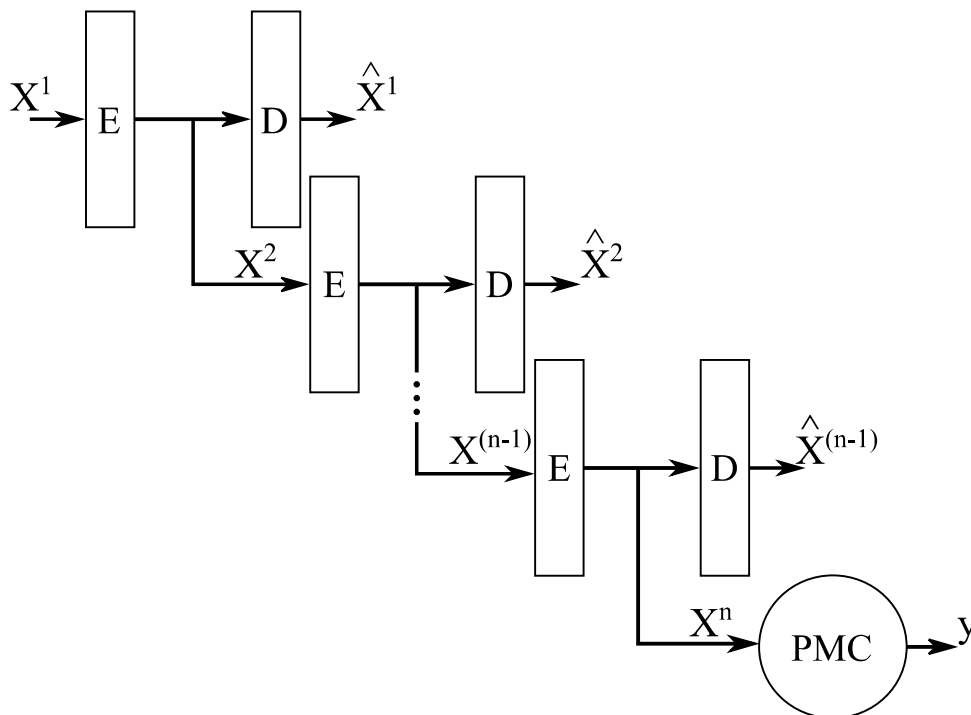


Figura 19: Esquemático da estrutura de rede neural profunda proposta

Assim como a rede TDNN, a rede proposta também foi implementada por meio do software Matlab, através da função *trainAutoencoder(X)*. As etapas da implementação da estrutura proposta podem ser definidas como:

1. O conjunto de dados, já tratado e com  $np$  amostras, é apresentado para uma rede autoencoder. Esta rede possui  $np$  neurônios nas camadas de entrada e saída e  $m$  neurônios em sua camada escondida. Utilizando a função *trainAutoencoder(X)*, esta rede é treinada, considerando que o tamanho  $m$  será menor do que o valor de  $np$ ;
2. As saídas da camada encoder da rede são utilizadas como entradas para uma outra rede autoencoder. Nesta etapa, a rede recebe um conjunto de entradas com dimensão  $m$  e o reduz para uma dimensão  $m_1$ . Este processo é repetido por um determinado

número de vezes, de acordo com a quantidade de camadas escondidas se deseja para a rede final;

3. As camadas encoder criadas são conectadas à uma rede da arquitetura PMC, com uma camada escondida com um neurônio que utiliza uma função de ativação linear. Esta rede recebe como entrada a saída da última camada encoder e sua saída será a resposta do modelo de previsão. A união das camadas é realizada por meio da função *stack* do Matlab;
4. Por fim, a rede completa é treinada de forma supervisionada. Todavia, os parâmetros da rede não são iniciados de forma aleatória, sendo utilizados os pesos já calculados para os treinamentos individuais.

Com esta proposta de construção de rede neural, espera-se que a rede seja capaz de abstrair conhecimento dos dados durante os processos de treinamento individuais de cada camada, atingindo um grau maior de conhecimento e capacidade de generalização sobre o problema tratado. Os passos de 1-3 correspondem a etapa de pré-treinamento da rede, enquanto que o passo 4 caracteriza a etapa de ajuste fino, onde a rede já construída tem o ajuste de seus parâmetros.

Para a implementação desta rede, foram consideradas duas camadas escondidas compostas pelas camadas encoder. A definição do número de camadas escondidas se baseou no equilíbrio entre a complexidade do treinamento e a acurácia da previsão. Desta forma, como não foi encontrada melhora significativa na acurácia da previsão para o aumento do número de camadas, limitou-se ao uso de duas camadas.

Assim como foi feito para a rede TDNN, vários valores de  $np$  foram testados. Estes valores variaram de 2 até 12 instantes passados da série. Para a construção de cada rede, a definição dos parâmetros  $m$  e  $m1$  foi feita de acordo com o valor de  $np$  avaliado. Tais parâmetros indicam a dimensão das camadas escondidas da rede autoencoder e, conseqüentemente, o número de neurônios das camadas escondidas da rede final. O valor de  $m$ , que é relativo à primeira camada escondida, deve ter valor máximo igual a  $np - 1$ , pois deve ser menor que  $np$ , enquanto que  $m1$  tem valor máximo de  $m - 1$ . Os valores mínimos destes parâmetros foram definidos com base no custo do treinamento da rede, pois, durante a implementação, observou-se que valores muito pequenos para estes parâmetros tornavam o treinamento inviável, de modo que a rede era incapaz de reduzir a dimensão dos dados. Desta forma, determinou-se que os valores mínimos para  $m$  e  $m1$  seriam iguais à 50% de  $np$  e de  $m$ , respectivamente.

Outro aspecto relevante é o fato de que não foi possível implementar a metodologia para  $np = 2$ , pois este valor fornecia poucos dados para a rede autoencoder e esta não era capaz de abstrair informações importantes para reduzir a dimensionalidade dos

dados de forma satisfatória. Desta forma, para este valor de  $np$ , o treinamento da rede autoencoder não atendia os critérios de parada pré-estabelecidos. O critério de parada para o treinamento da rede autoencoder foi o erro quadrático médio da rede durante o treinamento. Este critério foi definido como 0.01 para a rede da primeira camada e como 0.05 para a rede da segunda camada. Estes valores foram escolhidos por retornarem redes que apresentavam bom grau de abstração e que posteriormente resultavam em bons níveis de acurácia para a previsão. Além disso, seguindo estudos da literatura, definiu-se o número máximo de épocas de treinamento como 100 e 50 épocas para a rede da primeira e da segunda camada escondida, respectivamente.

Após definir os limites dos parâmetros utilizados, várias topologias foram avaliadas por meio da alteração sistemática dos parâmetros. Desta forma, considerando  $np = 6$ , por exemplo,  $m$  foi avaliado para o intervalo entre 3 e 5 e  $m1$  foi avaliado entre 3 e 4. Variando os parâmetros de forma sistemática infere-se que para  $np = 6$ , 6 topologias foram avaliadas no total. Ressalta-se ainda que cada uma das topologias foi treinada 5 vezes, assim como feito para a rede TDNN.

## 5.5 Considerações parciais

Este capítulo apresentou detalhes sobre a implementação do método proposto e do método de comparação. Descrevendo as funções utilizadas para treinar as redes e os processos de construção adotados. Além disso, apresentou-se uma detalhada análise sobre os indicadores de desempenho utilizados para avaliar as metodologias propostas. A construção e o perfil dos conjuntos de treinamento e de teste também foram abordados.



## 6 RESULTADOS E DISCUSSÕES

Esta seção do trabalho será dividida em cinco partes, sendo as quatro primeiras referentes à previsão das séries de demanda para cada uma das quatro estações e a última consistindo de um compilado dos resultados individuais, reunindo também as considerações e conclusões sobre tais resultados.

Ressalta-se que, mesmo a previsão tendo sido realizada com a parcela residual das séries temporais das estações, os resultados serão apresentados na forma da demanda desnormalizada. Para isso, após a previsão da parcela residual pela rede, esta foi desnormalizada e posteriormente somada à parcela média prevista pelo método dos dias típicos. Optou-se por esta abordagem para que a análise do erro considerasse o erro associado com o resultado final, que é o utilizado em situações práticas.

### 6.1 Série temporal de demanda de energia - Primavera

Esta seção apresenta os resultados para as previsões da série de primavera realizadas pela rede TDNN e pela rede neural profunda proposta.

#### 6.1.1 Previsão com a Time Delay Neural Network

Os processos de treinamento e teste para previsão desta série seguiram os procedimentos indicados na [Capítulo 5](#). Assim, a [Tabela 2](#) apresenta os resultados obtidos para os melhores treinamentos das melhores topologias para cada um dos valores de  $np$  avaliados.

Tabela 2: Resultados de previsão da rede TDNN - Primavera

$np$	Topologia	MAPE (%)	$(\beta)$ (%)
2	[7 10]	1.83	6.94
4	[4 5]	1.73	10.07
6	[2 9]	1.71	7.06
8	[2 5]	1.68	4.63
10	[2 1]	1.75	11.69
12	[3 1]	1.74	7.64

Nota-se que dentre os resultados, a topologia com melhor desempenho global foi a com 2 neurônios na primeira camada escondida e 5 neurônios na segunda camada, sendo esta referente ao modelo com  $np = 8$ . Além de atingir menor erro percentual, esta topologia também foi a única cujos resíduos de previsão apresentaram comportamento próximo ao comportamento da série de ruído branco ( $\beta < 5\%$ ). As [Figuras 20](#), [21](#) e [22](#) apresentam, respectivamente, os resultados de previsão, o histograma dos resíduos da previsão e a

análise da FAC para os resíduos da previsão. Sendo todos referentes à topologia selecionada com melhor desempenho global.

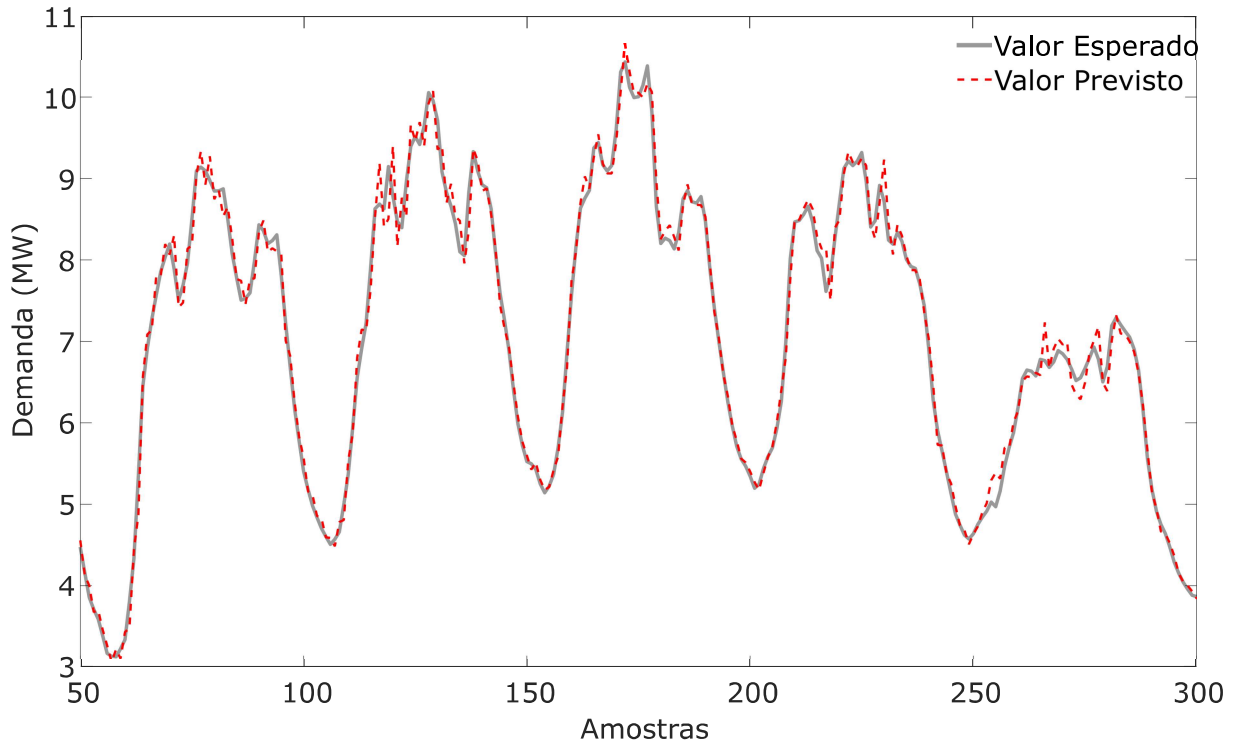


Figura 20: Resultado da previsão da rede TDNN para série de primavera

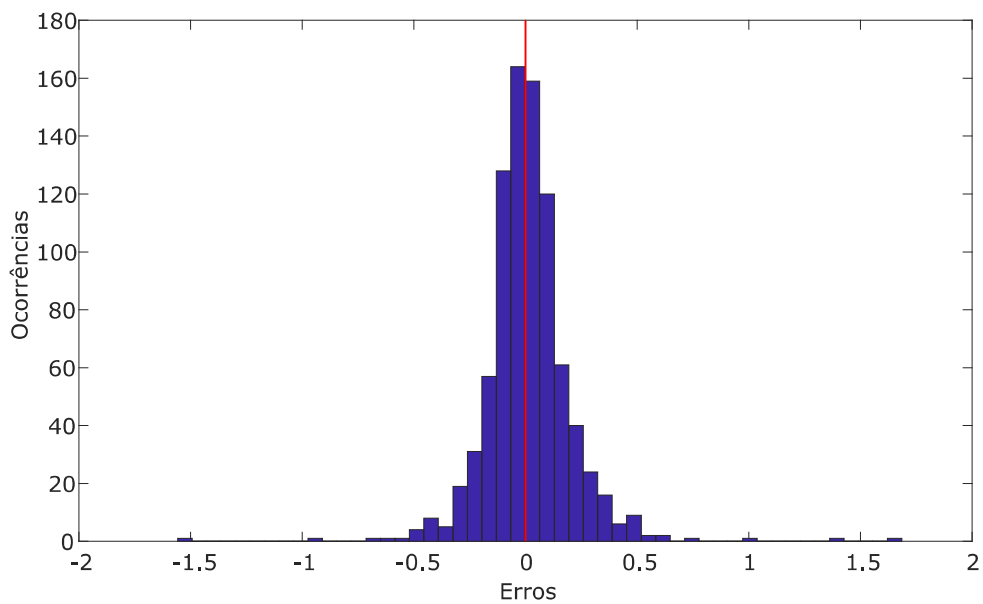


Figura 21: Histograma do erro de previsão para a topologia selecionada - Primavera

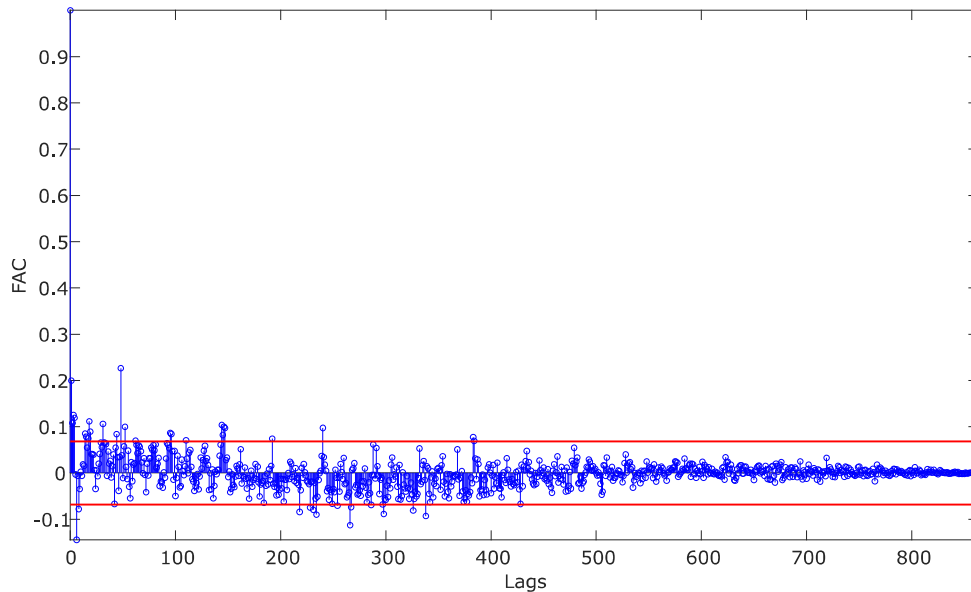


Figura 22: Função de autocorrelação para a série de resíduos de previsão - Primavera

### 6.1.2 Previsão com a rede neural profunda

Os resultados apresentados nesta seção foram obtidos a partir da implementação da rede neural profunda apresentada no [Capítulo 5](#). A [Tabela 3](#) apresenta os valores do erro e do parâmetro  $\beta$  obtidos para os treinamentos com menor erro das topologias selecionadas para cada um dos valores de  $np$  avaliados. Nota-se que para o caso da rede neural profunda, não foi possível avaliar o valor de  $np = 2$ , como já indicado na [seção 5.4](#).

Tabela 3: Resultados de previsão da rede neural profunda - Primavera

$np$	Topologia	MAPE (%)	$(\beta)$ (%)
2	-	-	-
4	[3 2]	15.02	57.06
6	[5 4]	1.67	4.40
8	[4 4]	1.67	4.40
10	[5 4]	1.66	3.24
12	[6 7]	1.69	4.51

Inferre-se da [Tabela 3](#) que a topologia com melhor desempenho global foi a obtida para  $np = 10$ . Esta topologia possui 5 neurônios em sua primeira camada escondida e 4 neurônios em sua segunda camada, sendo escolhida como a melhor por retornar previsões com o menor valor de MAPE ( $MAPE = 1.66\%$ ) e o menor valor de  $\beta$  ( $\beta = 3.24\%$ ).

Além disso, nota-se que para  $np = 4$ , a rede neural profunda apresentou valores elevados para os parâmetros avaliados. Este comportamento indica que a rede autoencoder não foi capaz de abstrair conhecimento durante o processo de pré-treinamento devido ao número reduzido de amostras de entrada. Este comportamento caracteriza uma limitação do modelo proposto, que para determinadas séries temporais pode exigir um conjunto

maior de dados de entrada, e não possui relação com o problema de convergência para mínimos locais, típico para séries temporais superficiais.

Ressalta-se ainda que, para todos os outros valores de  $np$ , a rede apresentou alto grau de aderência, com valores baixos de erro com  $\beta < 5\%$ , o que indica que os resíduos de previsão apresentaram comportamento próximo ao comportamento da série de ruído branco. As figuras 23, 24 e 25 a seguir apresentam, respectivamente, os resultados de previsão, o histograma dos resíduos da previsão e a análise da FAC para os resíduos da previsão para a melhor topologia da rede profunda.

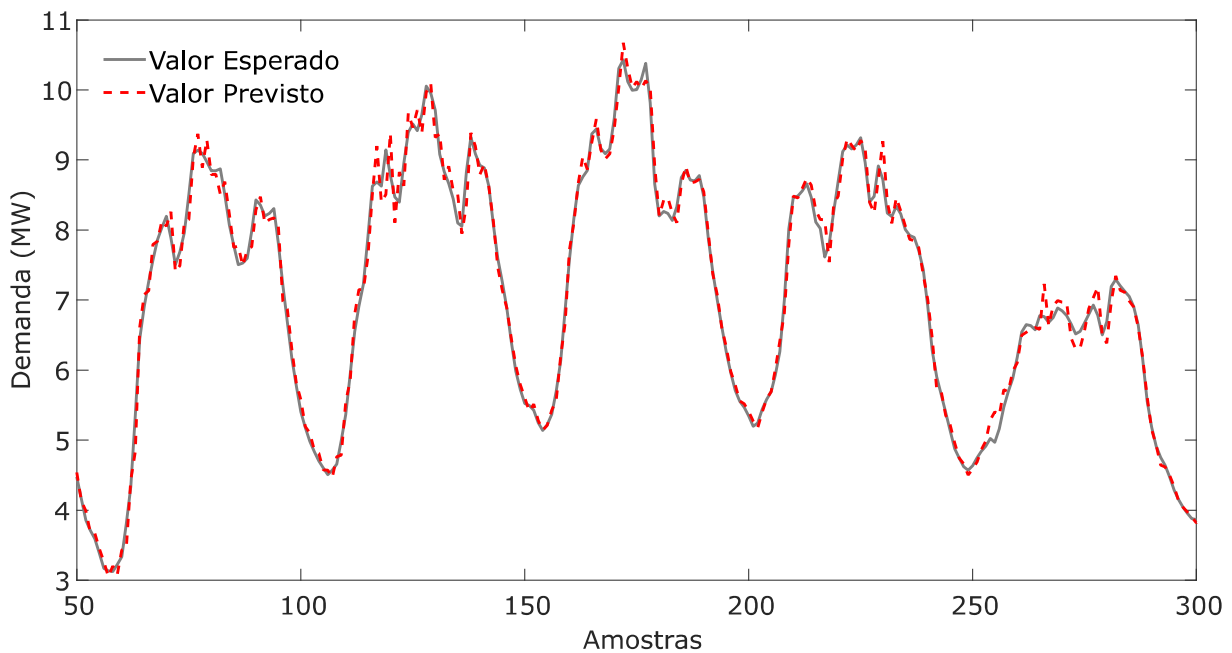


Figura 23: Resultado da previsão da rede neural profunda para série de primavera

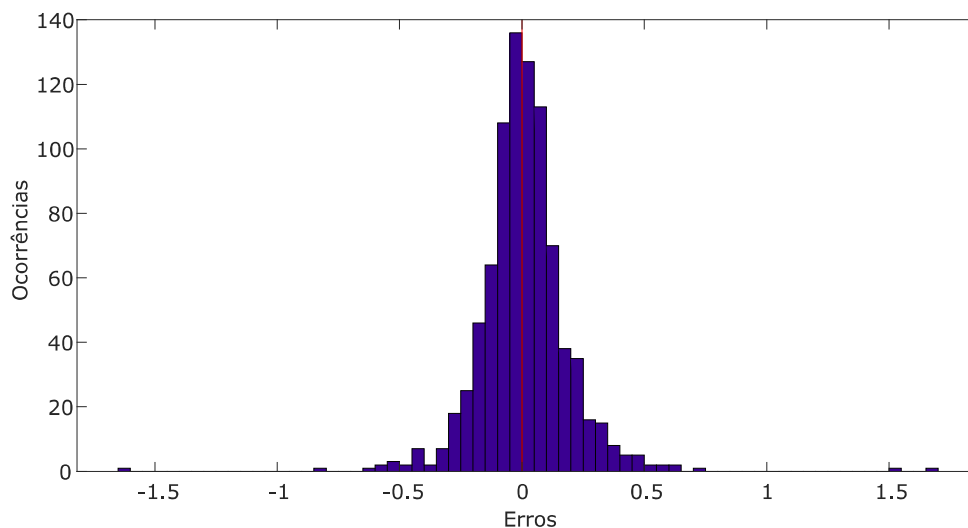


Figura 24: Histograma do erro de previsão para a topologia selecionada - Primavera

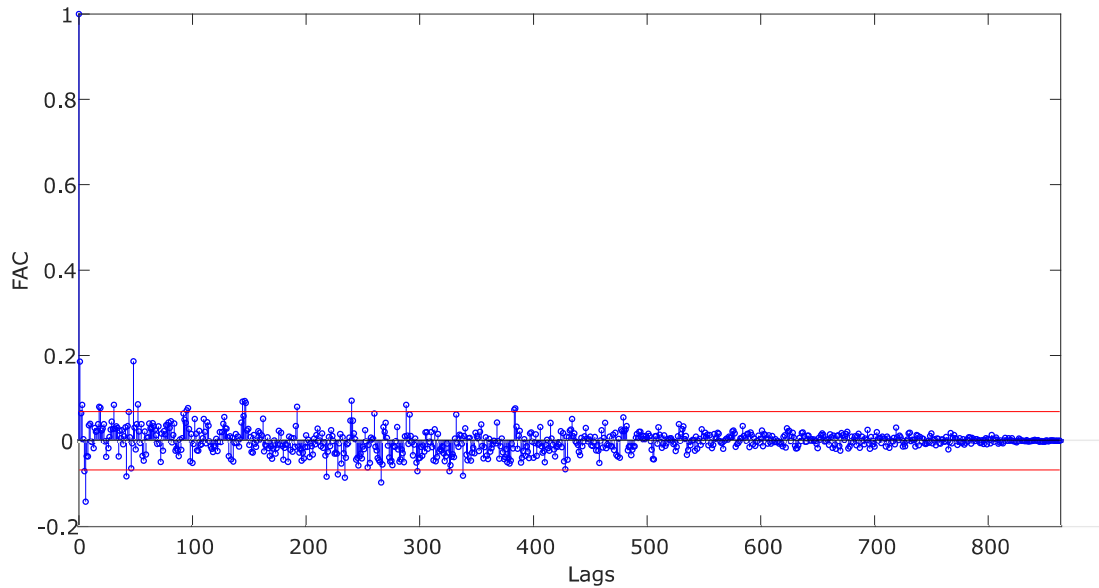


Figura 25: Função de autocorrelação para a série de resíduos de previsão - Primavera

## 6.2 Série temporal de demanda de energia - Verão

Esta seção apresenta os resultados de previsão da série temporal de verão, obtidos com base nos modelos de previsão utilizando a rede TDNN e a rede neural profunda implementada.

### 6.2.1 Previsão com a Time Delay Neural Network

A [Tabela 4](#) apresenta os resultados obtidos para os valores de  $np$  avaliados considerando a série temporal de verão.

Tabela 4: Resultados de previsão da rede TDNN - Verão

$np$	Topologia	MAPE (%)	$(\beta)$ (%)
2	[10 1]	2.56	4.44
4	[1 6]	2.33	3.51
6	[2 5]	2.33	3.51
8	[7 2]	2.33	3.16
10	[3 8]	2.35	4.56
12	[5 7]	2.38	4.09

Infere-se da [Tabela 4](#) que a topologia com melhor desempenho global foi a com 7 neurônios na primeira camada escondida e 2 neurônios na segunda camada, sendo esta também referente ao modelo com  $np = 8$ , assim como para a série de primavera. Além de atingir menor erro percentual, esta topologia também foi a que apresentou menor valor para o parâmetro  $\beta$ . Todavia, ao contrário do que aconteceu para a série de primavera, no caso da série de verão todos os modelos foram capazes de captar as principais características

determinísticas da série, apresentando  $\beta < 5\%$ . Os resultados de previsão relativos à topologia com melhor desempenho global são apresentados nas figuras 26, 27 e 28.

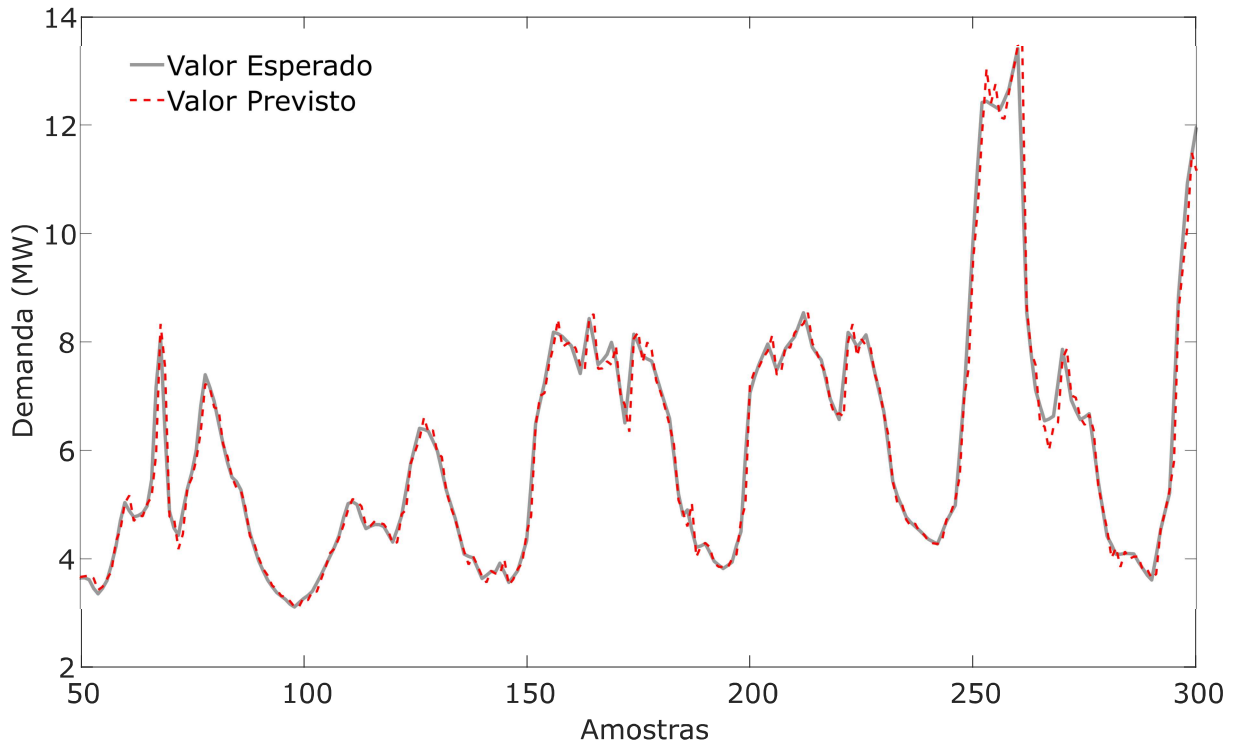


Figura 26: Resultado da previsão da rede TDNN para série de verão

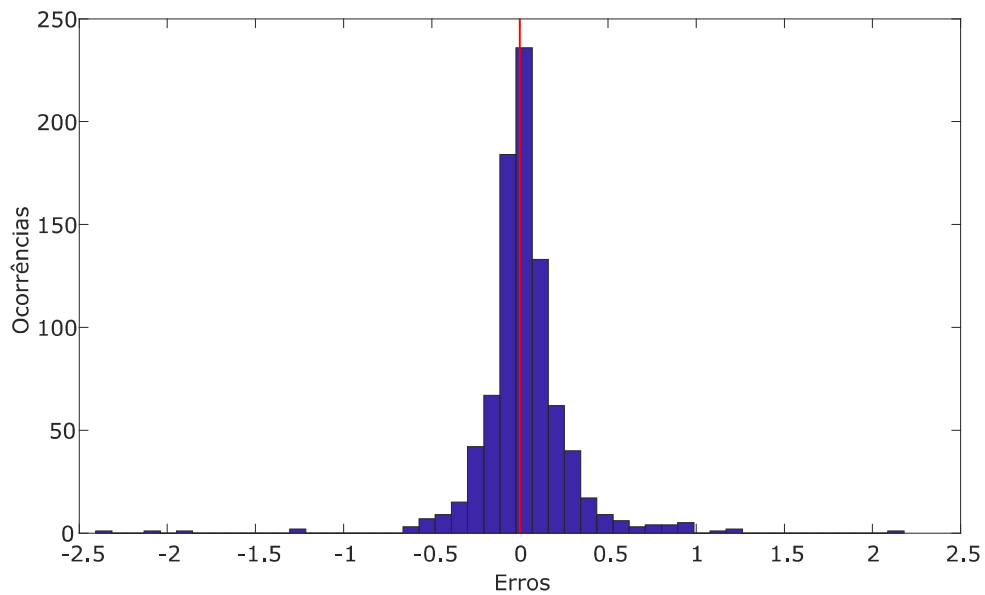


Figura 27: Histograma do erro de previsão para a topologia selecionada - Verão

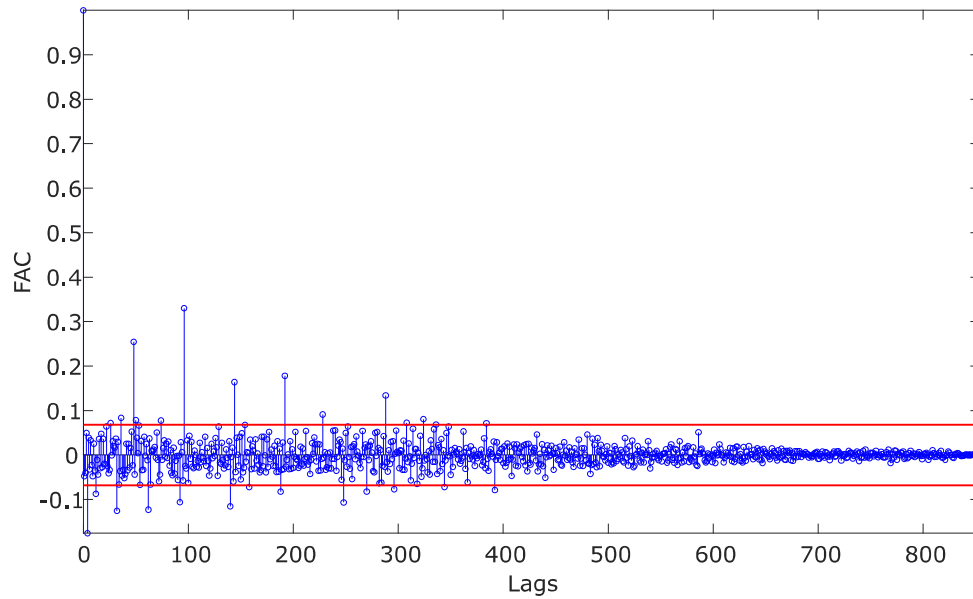


Figura 28: Função de autocorrelação para a série de resíduos de previsão - Verão

### 6.2.2 Previsão com a rede neural profunda

A [Tabela 5](#) apresenta os resultados obtidos com a rede neural profunda para a série temporal de verão.

Tabela 5: Resultados de previsão da rede neural profunda - Verão

np	Topologia	MAPE (%)	$(\beta)$ (%)
2	-	-	-
4	[3 2]	2.40	4.56
6	[3 2]	2.35	3.51
8	[4 3]	2.35	3.63
10	[5 4]	2.37	4.09
12	[6 2]	2.38	4.56

Para esta série, a topologia com melhor desempenho global foi a obtida para  $np = 6$ . Esta topologia possui 3 neurônios na primeira camada escondida e 2 neurônios na segunda camada. Apesar de não possuir o menor valor de MAPE dentre as topologias avaliadas, esta topologia foi escolhida como melhor por ter apresentado o menor valor para o parâmetro  $\beta$  e por ter obtido uma métrica de erro aproximadamente igual a da topologia com menor MAPE (obtida para  $np = 8$ ). Desta forma, ambas poderiam ser utilizadas para a previsão da série de verão, por terem obtido parâmetros similares e que indicam boa aderência ao problema e a escolha foi feita priorizando a parcimônia do modelo, ao se escolher a topologia com menos parâmetros.

Outro aspecto interessante observado para a previsão desta série é que, mesmo para valores pequenos de  $np$ , todos os modelos da rede neural profunda apresentaram

desempenhos semelhantes para a previsão desta série. Além disso, ao contrário do que ocorreu para a previsão da série de primavera, todas os modelos foram capazes de captar as principais características determinísticas da série, apresentando  $\beta < 5\%$ . As figuras 29, 30 e 31 apresentam detalhes da previsão realizada pela melhor topologia da rede neural profunda.

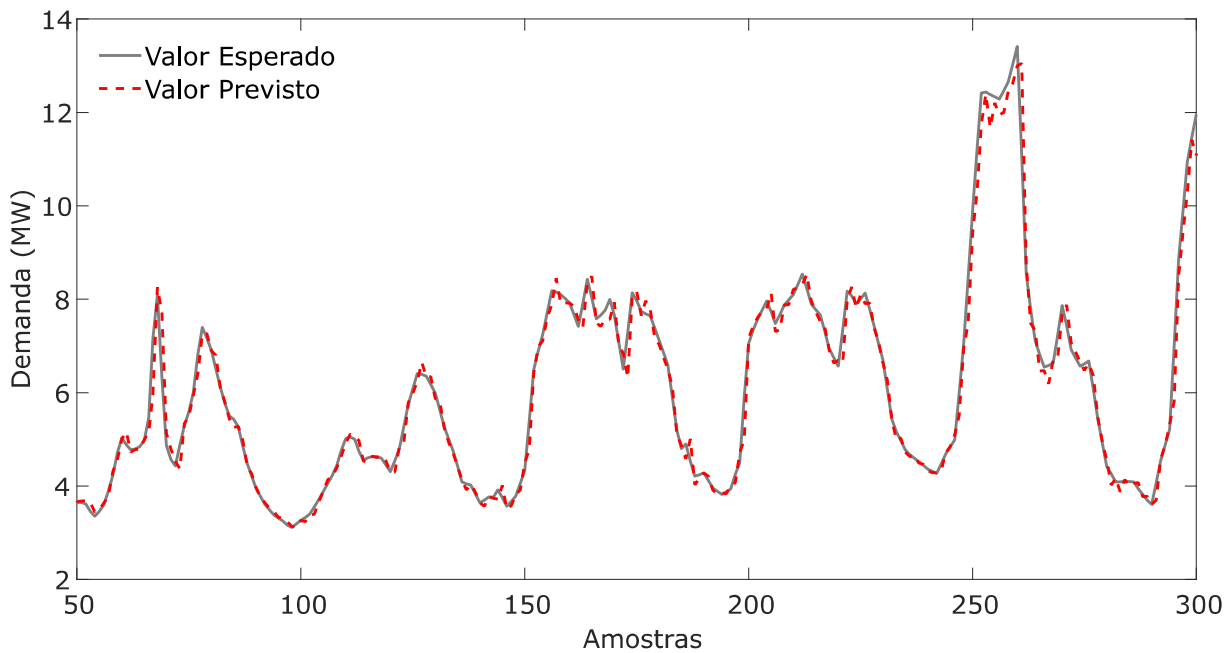


Figura 29: Resultado da previsão da rede neural profunda para série de verão

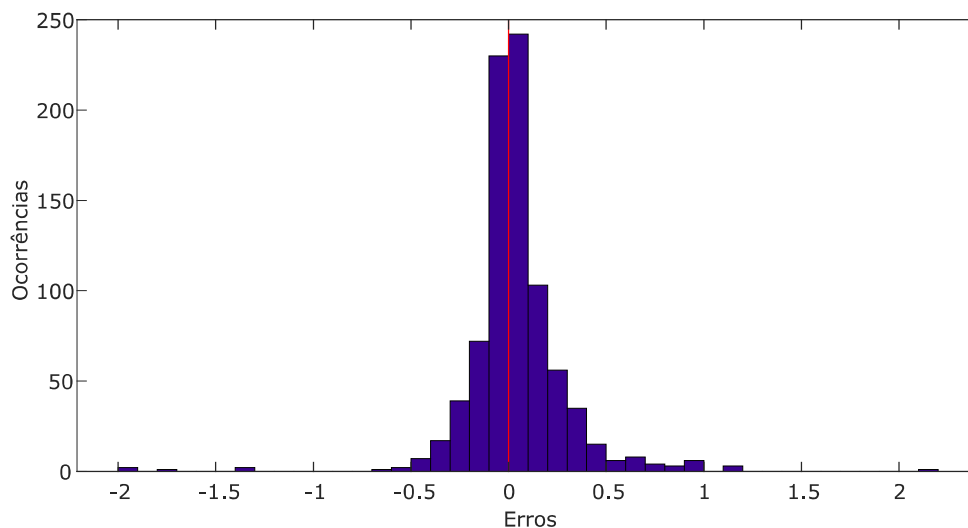


Figura 30: Histograma do erro de previsão para a topologia selecionada - Verão



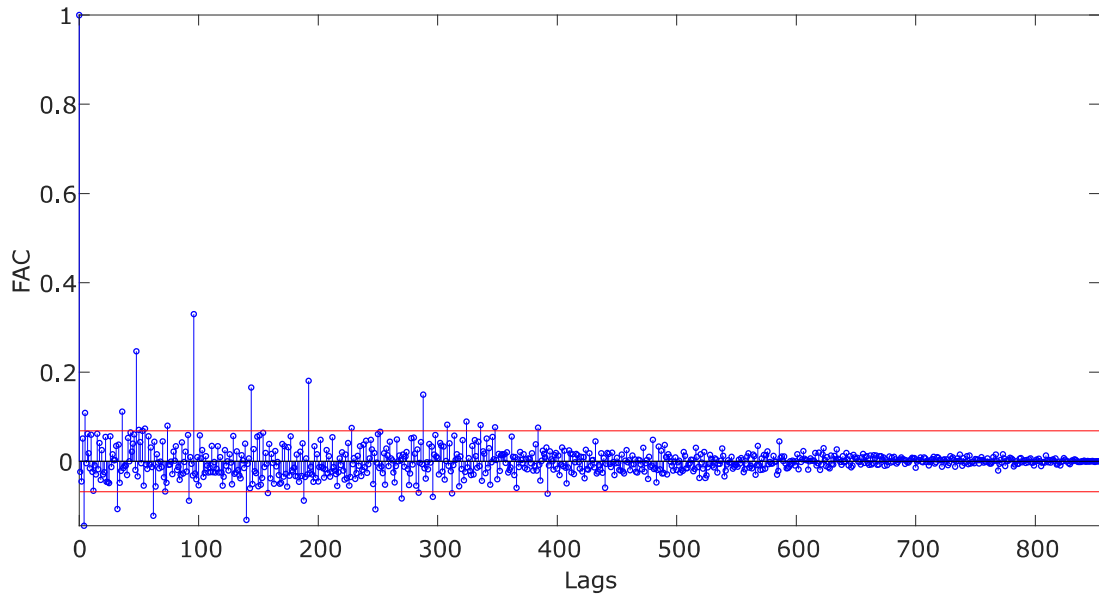


Figura 31: Função de autocorrelação para a série de resíduos de previsão - Verão

### 6.3 Série temporal de demanda de energia - Outono

Esta seção apresenta os resultados de previsão da série de outono, obtidos com base nos modelos construídos com a rede TDNN e com a rede neural profunda.

#### 6.3.1 Previsão com a Time Delay Neural Network

Os resultados obtidos considerando a série temporal de outono são apresentados na [Tabela 6](#).

Tabela 6: Resultados de previsão da rede TDNN - Outono

$np$	Topologia	MAPE (%)	$(\beta)$ (%)
2	[8 10]	2.02	4.96
4	[7 8]	1.92	4.00
6	[9 2]	1.91	4.96
8	[7 3]	1.94	4.60
10	[6 8]	1.91	3.27
12	[7 3]	1.91	4.12

Com base na tabela, define-se que a topologia com melhor desempenho global foi a com 6 neurônios na primeira camada escondida e 8 neurônios na segunda camada, sendo esta referente ao modelo com  $np = 10$ . Esta topologia apresentou um MAPE de 1.91% e atingiu o menor valor para o parâmetro  $\beta$  ( $\beta = 3.27\%$ ). Assim como para a série de verão, no caso da série de outono, todos os modelos foram capazes de captar as características determinísticas da série, apresentando  $\beta < 5\%$ . Os resultados de previsão, o histograma dos resíduos e a FAC dos resíduos, relativos à topologia com melhor desempenho global, são apresentados nas figuras [32](#), [33](#) e [34](#), respectivamente.

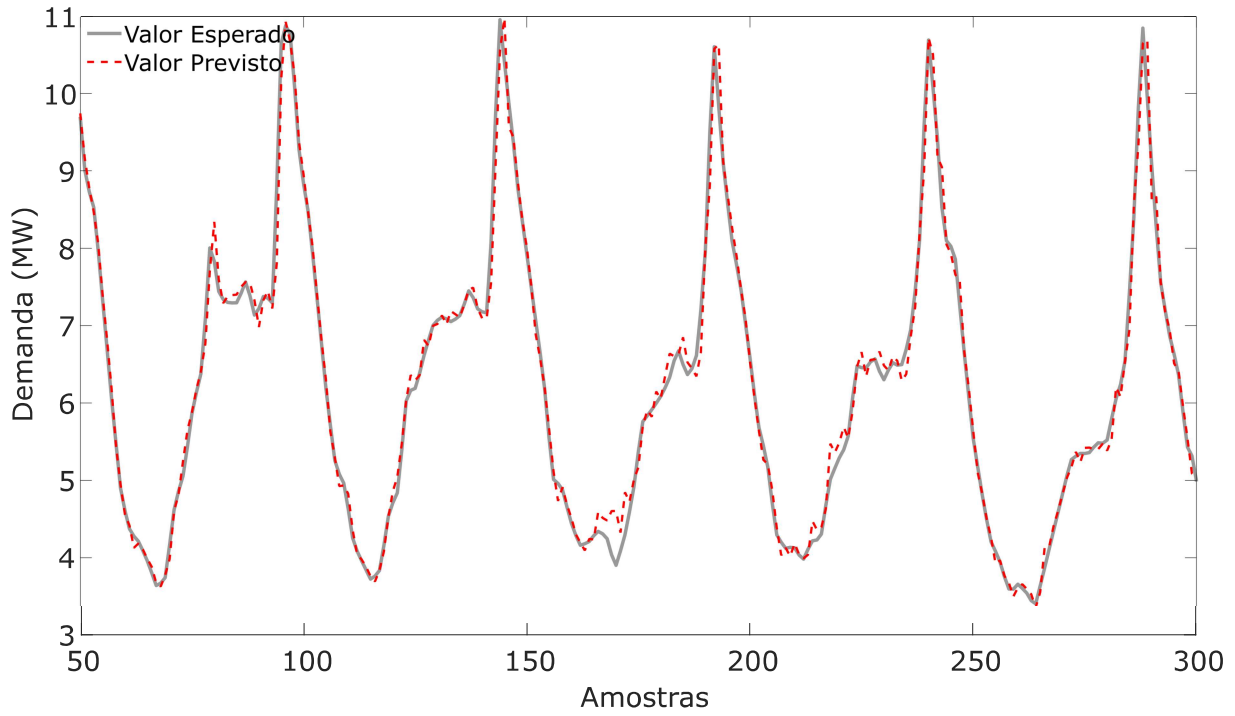


Figura 32: Resultado da previsão da rede TDNN para série de outono

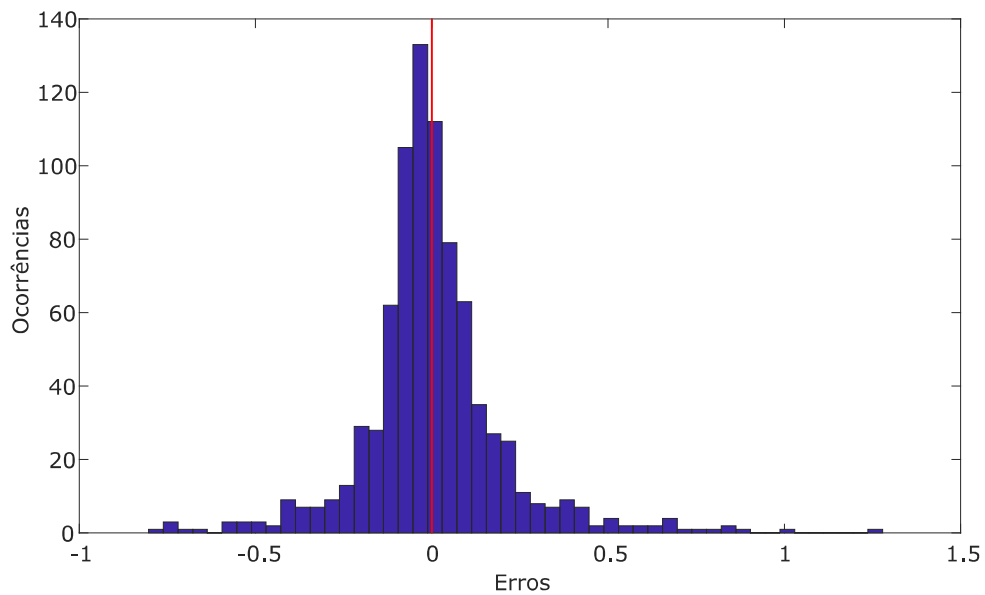


Figura 33: Histograma do erro de previsão para a topologia selecionada - Outono

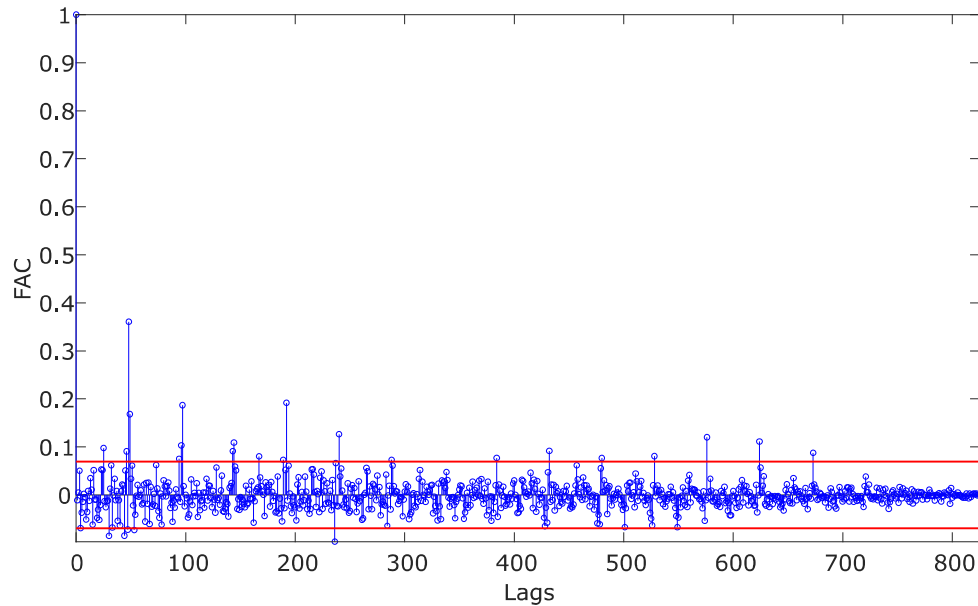


Figura 34: Função de autocorrelação para a série de resíduos de previsão - Outono

### 6.3.2 Previsão com a rede neural profunda

Os resultados obtidos pela rede neural profunda para a série temporal de outono são apresentados na [Tabela 7](#).

Tabela 7: Resultados de previsão da rede neural profunda - Outono

np	Topologia	MAPE (%)	$(\beta)$ (%)
2	-	-	-
4	[3 2]	8.72	60.65
6	[5 3]	1.92	4.48
8	[5 4]	1.92	4.48
10	[6 3]	1.91	3.87
12	[10 9]	1.91	4.12

A partir da tabela, infere-se que a topologia com melhor desempenho foi a com 6 neurônios na primeira camada escondida e 3 neurônios na segunda camada, avaliada para  $np = 10$ . Para esta topologia o erro obtido foi de  $MAPE = 1.91\%$  e o parâmetro  $\beta$  foi de  $3.87\%$ . Nota-se que, assim como para a série de primavera, o modelo de rede profunda apresentou um resultado fraco para  $np = 4$ , com um valor de  $\beta$  elevado, chegando a  $60.65\%$ . Todavia, para os outros valores de  $np$  avaliados todos os modelos apresentaram  $\beta < 5\%$ , sendo capazes de captar as características determinísticas da série. As figuras [35](#), [36](#) e [37](#) apresentam, respectivamente, os resultados de previsão, o histograma dos resíduos e a FAC dos resíduos, relativos à melhor topologia.

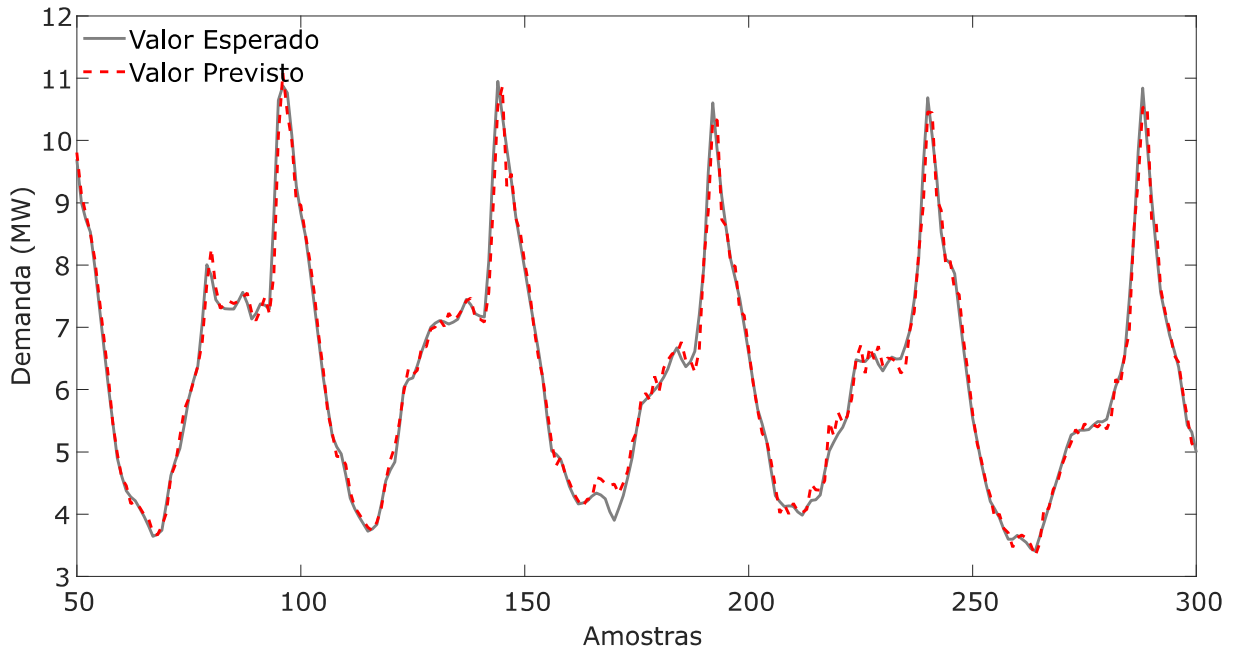


Figura 35: Resultado da previsão da rede neural profunda para série de outono

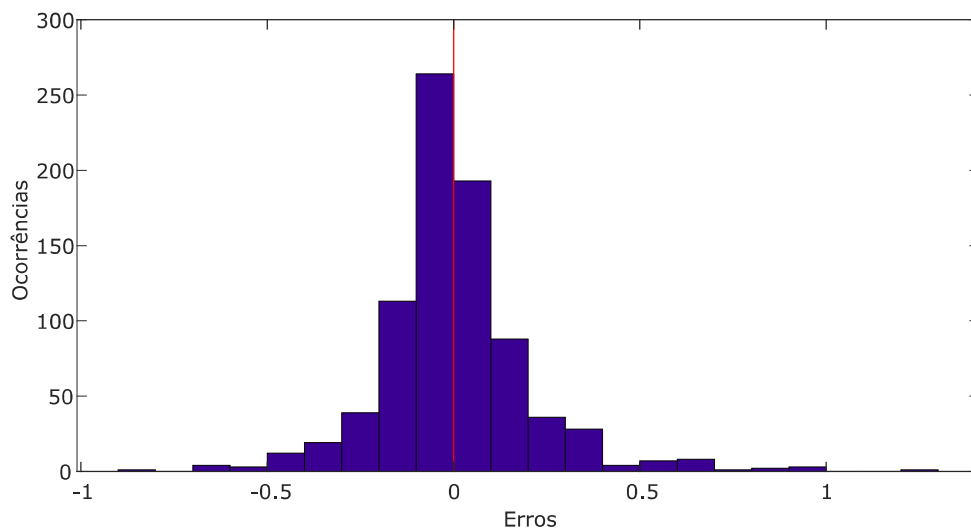


Figura 36: Histograma do erro de previsão para a topologia selecionada - Outono

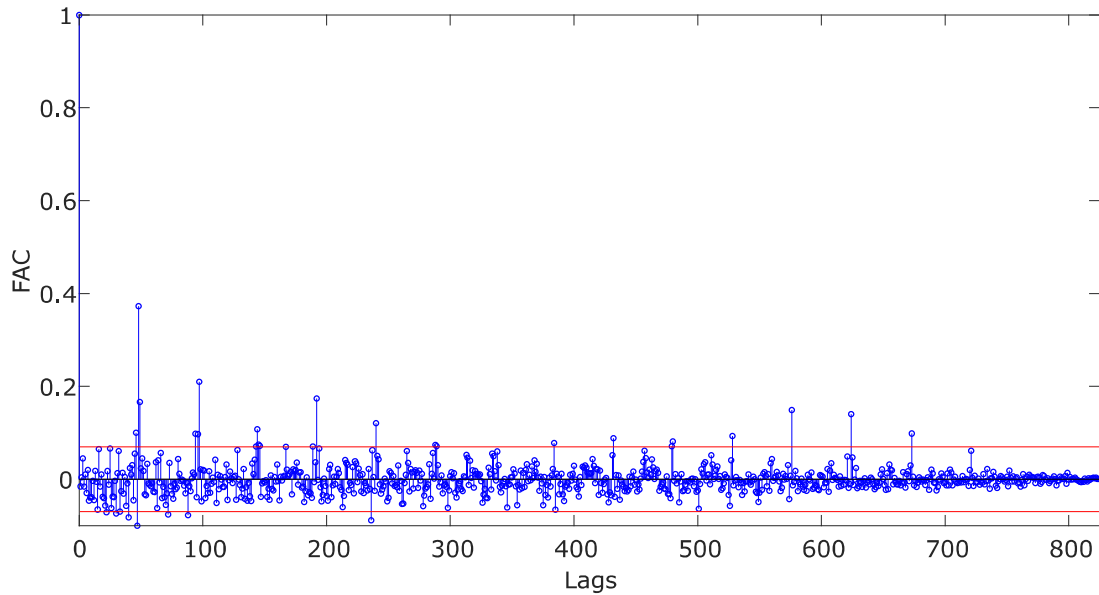


Figura 37: Função de autocorrelação para a série de resíduos de previsão - Outono

#### 6.4 Série temporal de demanda de energia - Inverno

Esta seção apresenta os resultados obtidos para a previsão da série de inverno, obtidos pela rede TDNN e pela rede neural profunda.

##### 6.4.1 Previsão com a Time Delay Neural Network

Os resultados obtidos para a série temporal de inverno estão reunidos na [Tabela 8](#).

Tabela 8: Resultados de previsão da rede TDNN - Inverno

np	Topologia	MAPE (%)	$(\beta)$ (%)
2	[2 5]	1.57	6.83
4	[8 4]	1.46	5.82
6	[3 6]	1.46	5.71
8	[9 5]	1.45	5.04
10	[9 3]	1.46	4.93
12	[1 3]	1.47	5.60

Para o caso desta série, a seleção da topologia com melhor desempenho global não é trivial, pois ao contrário das outras séries, a topologia com menor valor de MAPE, não é a topologia com menor valor de  $\beta$ . Tendo em vista que se deseja um modelo que apresente um bom grau de aderência ao problema tratado, sendo capaz de aprender a maior quantidade possível de características determinísticas da série, e considerando que é insignificativa a diferença entre o MAPE do modelo com menor MAPE e do modelo com menor  $\beta$ , deve-se escolher a topologia relativa a  $np = 10$ . Esta topologia possui 9 neurônios na primeira camada escondida e 3 neurônios na segunda camada, sendo a única que foi

capaz de retornar uma série de resíduos de previsão com comportamento similar ao de uma série de ruído branco. Os resultados de previsão para esta topologia são apresentados nas figuras 38, 39 e 40, respectivamente.

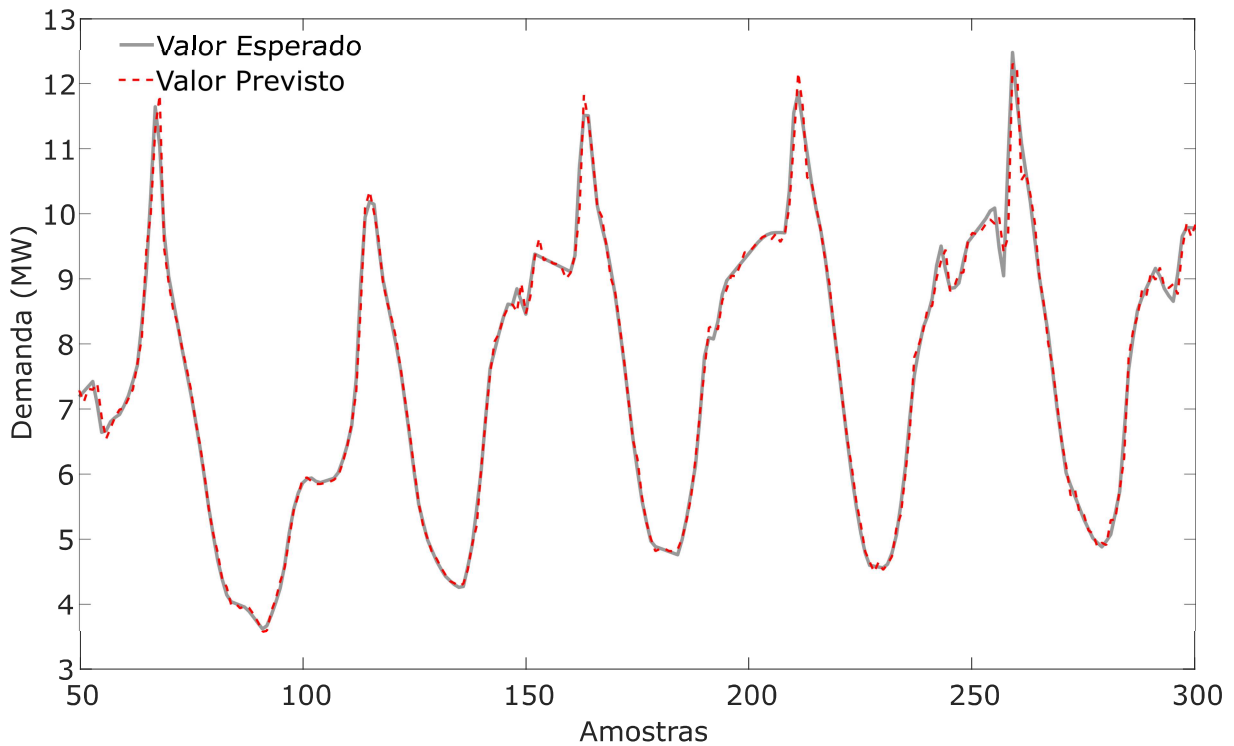


Figura 38: Resultado da previsão da rede TDNN para série de inverno

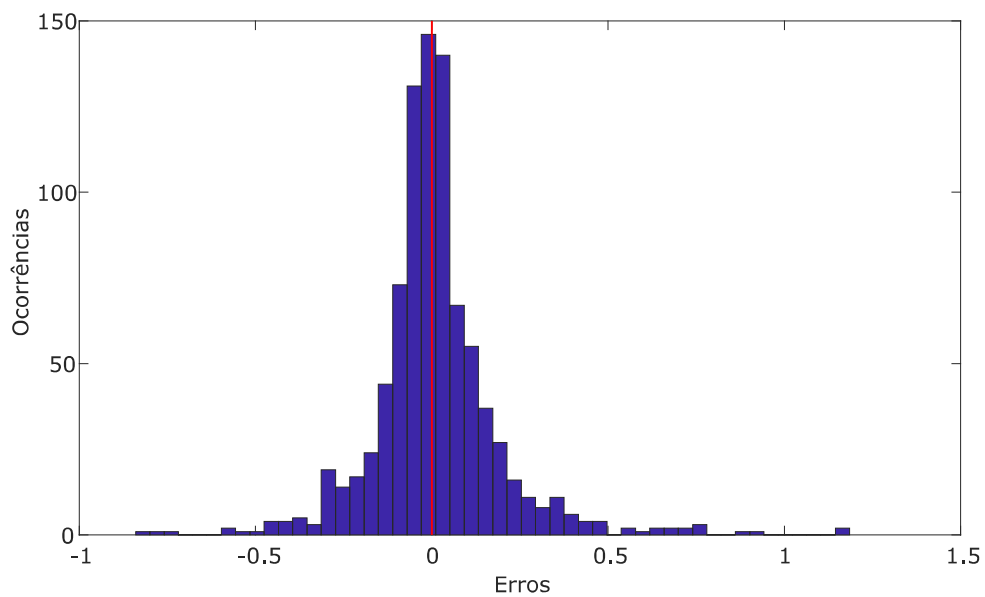


Figura 39: Histograma do erro de previsão para a topologia selecionada - Inverno

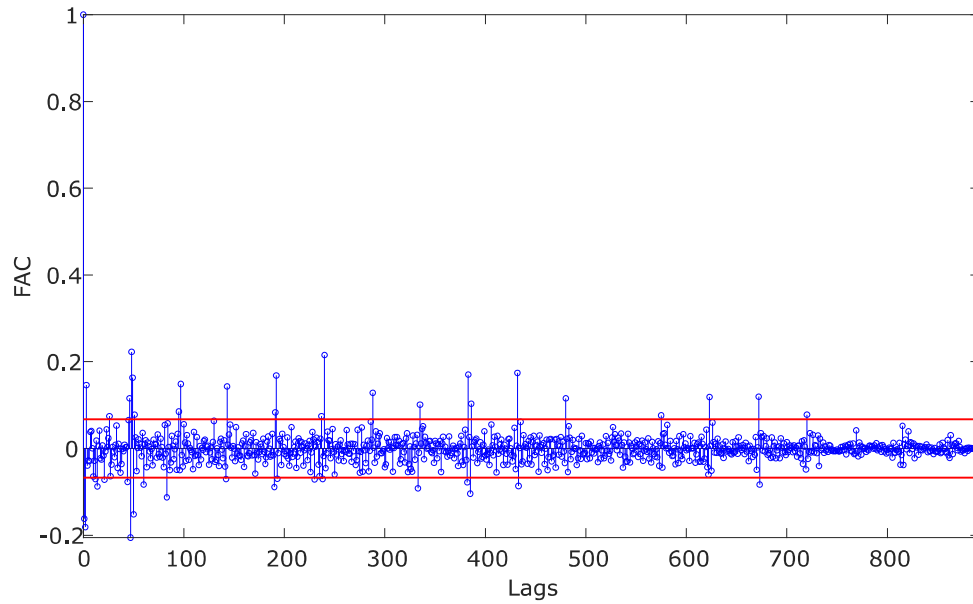


Figura 40: Função de autocorrelação para a série de resíduos de previsão - Inverno

#### 6.4.2 Previsão com a rede neural profunda

A [Tabela 9](#) apresenta os resultados de previsão obtidos pela rede neural profunda para a série temporal de inverno.

Tabela 9: Resultados de previsão da rede neural profunda - Inverno

np	Topologia	MAPE (%)	$(\beta)$ (%)
2	-	-	-
4	[3 2]	1.45	5.82
6	[4 3]	6.69	20.16
8	[7 4]	1.45	4.82
10	[5 6]	1.46	5.38
12	[9 8]	1.46	5.38

Como ocorrido para a série de verão, a topologia com menor valor de MAPE (topologia para  $np = 4$ ) não é a topologia com menor valor de  $\beta$  (topologia para  $np = 8$ ). Todavia, para este caso, a diferença entre os valores de  $\beta$  é significativa, principalmente pois uma das topologias apresenta  $\beta < 5\%$ , enquanto a outra apresenta um valor superior. Desta forma, considerando que a escolha do modelo de previsão deve levar em consideração o grau de aderência do modelo ao problema tratado e que os valores de MAPE encontrados pelas topologias são aproximadamente iguais, deve-se escolher a topologia relativa a  $np = 8$ . Esta possui 7 neurônios em sua primeira camada escondida e 4 neurônios na segunda. Os resultados de previsão para esta topologia são apresentados nas figuras [41](#), [42](#) e [43](#), respectivamente.

Ressalta-se ainda que, excetuando a topologia para  $np = 6$ , todas as topologias

retornaram valores de MAPE inferior à 2%. Neste caso, considera-se que a topologia com  $np = 6$  não foi capaz de abstrair conhecimento suficiente do conjunto de dados que lhe foi apresentado durante a etapa de pré-treinamento das camadas encoder.

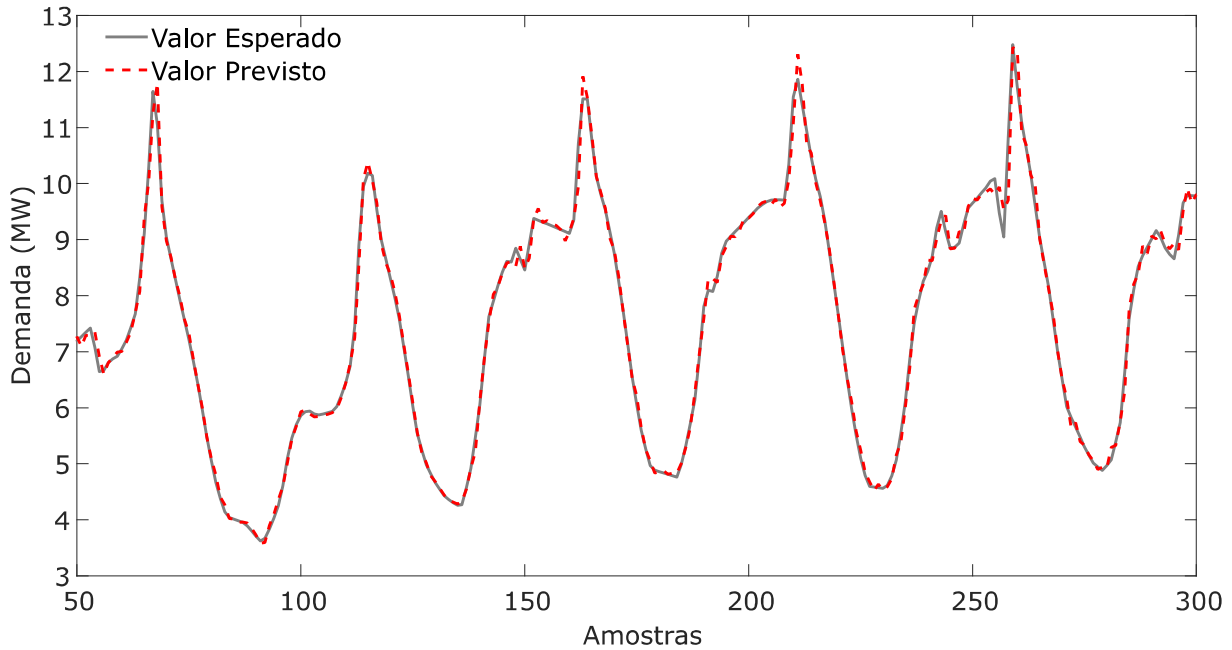


Figura 41: Resultado da previsão da rede neural profunda para série de inverno

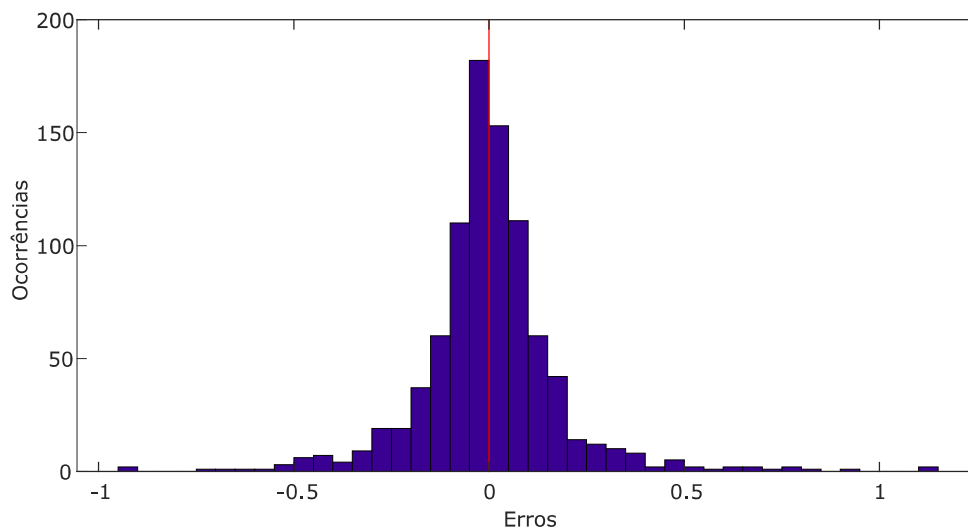


Figura 42: Histograma do erro de previsão para a topologia selecionada - Inverno



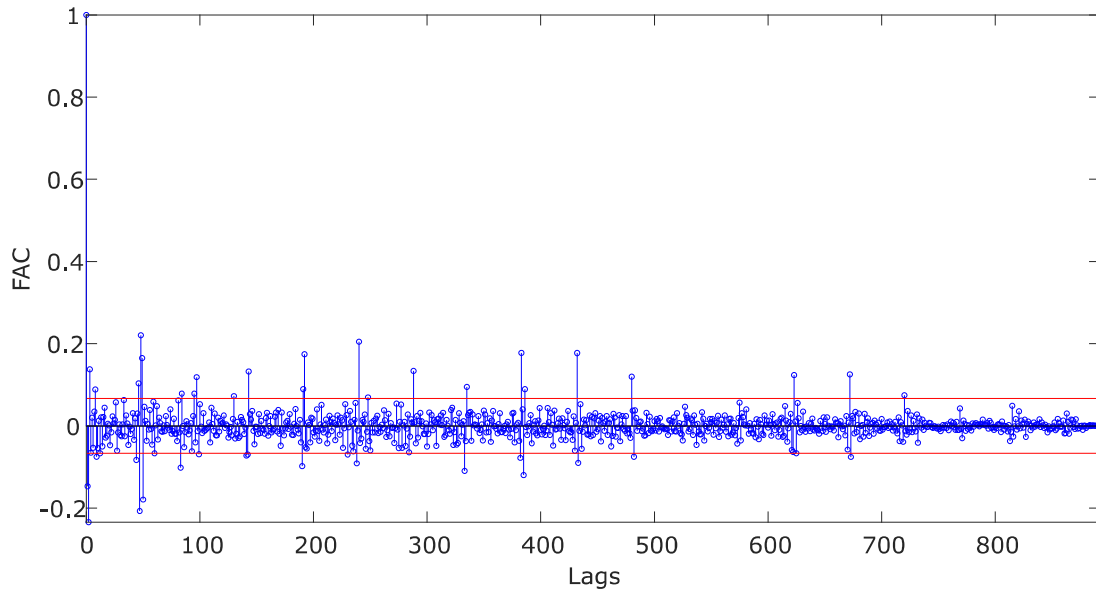


Figura 43: Função de autocorrelação para a série de resíduos de previsão - Inverno

## 6.5 Análise dos resultados

Esta seção apresenta a análise e as considerações sobre os resultados de previsão apresentados nas seções anteriores. Primeiramente, com base nas figuras 20, 26, 32, 38, 23, 29, 35, e 41, nota-se que para todas as séries e para ambos os modelos de previsão avaliados, a previsão final apresentou resultados com perfil muito similar ao esperado. Para alguns casos, em tais gráficos, é possível observar pequenas divergências nos pontos de pico e vale de demanda, como exemplo, pode-se observar a Figura 26. Estes picos, assim como os vales da demanda, foram previstos pelas redes na etapa de previsão da série residual de demanda, de forma que, a divergência entre os valores previstos e esperados é ocasionada devido a presença da parcela estocástica da série e indicam uma parcela da série que não pode ser prevista. Ressalta-se ainda que, visualmente, as previsões realizadas pela rede neural profunda proposta se assemelham mais ao valor esperado nestes pontos do gráfico, como pode ser percebido comparando-se as figuras 26 e 29.

A análise da FAC dos resíduos também indica a boa aderência dos modelos. Para todas as séries temporais foi possível encontrar modelos de previsão que capazes de aprender a maioria das relações determinísticas. Esses modelos retornaram resíduos de previsão com um comportamento similar ao do ruído branco, com mais de 95% dos coeficientes de autocorrelação permanecendo dentro dos limites de confiança.

Destaca-se o comportamento do modelo de previsão proposto para as séries de primavera, de outono e de inverno, que para valores reduzidos de  $np$ , apresentou um valor elevado do parâmetro  $\beta$ . Este comportamento é devido à característica intrínseca da construção do modelo da rede neural profunda, pois para a criação desta rede são utilizadas camadas encoder pré-treinadas obtidas a partir de redes neurais autoencoders. Dependendo

da série temporal analisada, para valores reduzidos de  $np$ , as redes autoencoder podem não ser capazes de extrair características relevantes do conjunto de dados, pois o conjunto de dados de entrada sendo apresentado não fornece tais informações. Por este motivo, as redes apresentam um treinamento fraco, que permanece na etapa de pós-treinamento e que refletem no resultado da previsão. Todavia, quando alimentadas com um conjunto maior de dados de entradas, tais redes apresentaram bons resultados, na maior parte dos casos retornando valores de  $\beta$  inferiores a 5%. Neste contexto, destaca-se o caso da previsão para a série de primavera, que para o modelo composto pela rede TDNN apresentou apenas uma topologia com  $\beta < 5\%$ , enquanto que, para a rede neural profunda, todas as topologias apresentaram  $\beta < 5\%$  (com exceção da topologia com  $np = 4$ , como explicado anteriormente).

Ainda em relação à análise dos gráficos apresentados, nota-se através das figuras 21, 27, 33, 39, 24, 30, 36 e 39, que todas as previsões apresentam erros cuja distribuição se assemelha à distribuição normal. Este comportamento indica a eficiência do modelo de previsão em incorporar as características determinísticas da rede, além de indicar que todos os modelos apresentaram erros próximos de zero. Nos gráficos, o valor de erro nulo é indicado pela linha vermelha, e as maiores porcentagens de ocorrência de erros estão em torno desta linha, indicando este comportamento.

Em relação aos resultados obtidos durante os processos de treinamento e teste das redes, a Tabela 10 apresenta um compilado dos resultados de previsão obtidos pelas melhores topologias para os modelos de previsão utilizando a TDNN e a RNAP proposta.

Tabela 10: Resumo dos resultados de previsão

	TDNN				RNAP			
	np	Topologia	MAPE (%)	$\beta$ (%)	np	Topologia	MAPE (%)	$\beta$ (%)
<b>Série temporal de primavera</b>	8	[2 5]	1.68	4.63	10	[5 4]	1.66	3.24
<b>Série temporal de verão</b>	8	[7 2]	2.33	3.16	6	[3 2]	2.35	3.51
<b>Série temporal de outono</b>	10	[6 8]	1.91	3.27	10	[6 3]	1.91	3.87
<b>Série temporal de inverno</b>	10	[9 3]	1.46	4.93	8	[7 4]	1.45	4.82

A partir da tabela, infere-se que a rede neural profunda proposta apresentou valores de MAPE inferiores ou iguais aos da rede TDNN para todas as séries temporais, com exceção da série de verão. Ressalta-se ainda, que para o caso no qual os valores de MAPE são iguais, série de outono, mesmo que a rede proposta tenha obtido um valor de  $\beta$  ligeiramente maior, ela também apresenta uma rede com estrutura mais simples. Neste caso, a rede proposta utiliza 3 neurônios em sua segunda camada escondida, ao invés de

8, como é o caso da rede TDNN. Considerando-se que deve-se optar pelo modelo mais simples disponível e que a diferença entre os parâmetros MAPE e  $\beta$  das redes é pequena, pode-se optar pela utilização da RNAP para previsão. Em relação à série de verão, o mesmo argumento pode ser utilizado, pois a rede RNAP apresenta valores de MAPE e de  $\beta$  semelhantes aos da rede TDNN, todavia a primeira apresenta uma estrutura mais simples, possuindo menos neurônios e menos parâmetros em sua topologia.

Em relação ao custo computacional de implementação das redes, ambas apresentam um custo baixo que permite aplicação prática dos métodos. Para a rede profunda proposta, o processo de construção da rede, feito por meio do pré-treinamento das redes autoencoder, e as etapas de treinamento e teste da rede custam em média 16 vezes mais que o processo de treinamento e de teste da rede TDNN. Todavia, considerando que o tempo médio de treinamento para uma rede TDNN genérica é de 1 segundo, o custo da metodologia proposta também é bastante reduzido. Além disso, ressalta-se que para a metodologia proposta há menos casos de convergência para mínimos locais ou para modelos de previsão que não se aderem ao problema tratado, como pode ser observado para o caso da rede de primavera, onde a rede TDNN apresentou dificuldade para convergir para um modelo com  $\beta$  inferior a 5%.

Com base nos resultados, pode-se concluir que ambos os modelos de previsão são válidos para a aplicação no problema de previsão de séries temporais de demanda. Observa-se ainda a superioridade do modelo proposto que, na maioria dos casos, apresentou resultados com erro menor e com maior aderência obtidos por modelos mais simples. Desta forma, o método proposto atende aos quesitos de um bom modelo de previsão ao atender os critérios de acurácia e parcimônia. Ainda, ressalta-se que ao contrário do que ocorreu para a rede TDNN, a RNAP proposta não apresentou problemas de convergência para determinadas séries, o que indica a validade do uso do pré-treinamento das camadas encoder para inicialização dos pesos sinápticos da rede como forma de contornar os problemas de convergência para mínimos locais.



## 7 CONCLUSÕES

A previsão de demanda é uma ferramenta importante para o setor elétrico, influenciando não apenas em aspectos de planejamento operacional do sistema, mas também tendo um grande viés econômico. Diversos métodos são encontrados na literatura para realizar a previsão da série temporal de demanda, sendo que as RNAs são um dos métodos mais populares. Todavia, tais redes podem apresentar problemas que são intrínsecos ao seu funcionamento, como a convergência para mínimos locais devido à inicialização aleatória dos pesos sinápticos da rede.

Tendo em vista esta motivação, neste trabalho foi proposta uma metodologia de rede neural artificial profunda para a previsão da série de demanda de potência ativa. Esta metodologia foi construída por meio do pré-treinamento de camadas encoder de redes neurais autoencoder. O objetivo do trabalho foi o de desenvolver a estrutura de RNA profunda e comparar o desempenho desta com os resultados obtidos pela RNA tradicional, do tipo TDNN. Sendo que esta última arquitetura foi selecionada como método comparativo por se tratar de um modelo clássico com bons resultados para a aplicação de previsão de séries temporais.

Com base nos resultados obtidos pelo trabalho, observa-se que ambas as redes neurais são válidas para a aplicação no problema de previsão proposto, obtendo resultados com erro percentual médio inferior à 3% para todas as séries avaliadas. Além disso, nota-se superioridade da rede desenvolvida no método proposto por apresentar resultados, em sua maioria, com menores valores de erro e com maior aderência. Ainda, as redes profundas implementadas também retornaram modelos mais simples, o que respeita o conceito de parcimônia para a seleção de modelos de previsão. Em relação ao objetivo do trabalho, este também foi atingido, pois o problema de convergência para mínimos locais não ocorreu para as redes profundas propostas, ao contrário do que ocorreu para as redes TDNN. Ainda, o custo computacional da metodologia proposta é viável para aplicações práticas, justificando seu uso em detrimento da abordagem tradicional.

Ressalta-se que este trabalho também contribuiu para o tema pesquisa, através do desenvolvimento de uma detalhada análise estatística da série temporal de demanda de potência, discutindo métodos para a preparação dos dados e para a construção de conjuntos de treinamento e de teste, que também podem ser utilizados por outros métodos da classe de *soft computing*. Por fim, a principal contribuição desta pesquisa é para o contexto da construção do modelo de RNAs para a previsão de demanda, através da apresentação de uma metodologia alternativa para construção da rede, na qual o treinamento das camadas é realizado de forma individual. Esta metodologia mostrou superioridade em relação à abordagem tradicional e pode ser utilizada em situações práticas na operação de sistemas

elétricos de potência.

## REFERÊNCIAS

ADHIKARI, R.; AGRAWAL, R. **An Introductory Study on Time series Modeling and Forecasting**. [S.l.: s.n.], 2013. ISBN 978-3-659-33508-2.

AGUIRRE, L. A. et al. Dynamical prediction and pattern mapping in short-term load forecasting. **International Journal of Electrical Power and Energy Systems**, v. 30, n. 1, p. 73–82, 2008. ISSN 01420615.

ALI, D. et al. Long-term load forecast modelling using a fuzzy logic approach. **Pacific Science Review A: Natural Science and Engineering**, Elsevier Ltd, v. 18, n. 2, p. 123–127, 2016. ISSN 24058823. Disponível em: <<http://dx.doi.org/10.1016/j.pusra.2016.09.011>>.

ALMESHAI EI, E.; SOLTAN, H. A methodology for electric power load forecasting. **Alexandria Engineering Journal**, Elsevier BV, v. 50, n. 2, p. 137–144, 6 2011. Disponível em: <<https://doi.org/10.1016/j.aej.2011.01.015>>.

ANEEL. **Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST: Módulo 2 – Planejamento da Expansão do Sistema de Distribuição**. [S.l.: s.n.], 2015. 36 p.

\_\_\_\_\_. **Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST: Módulo 8 – Qualidade da Energia Elétrica**. [S.l.: s.n.], 2018. 88 p.

BENGIO, Y. et al. Greedy layer-wise training of deep networks. **Advances in Neural Information Processing Systems**, n. 1, p. 153–160, 2007. ISSN 10495258.

BIANCHI, F. M. et al. An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting. may 2017. Disponível em: <<http://arxiv.org/abs/1705.04378><http://dx.doi.org/10.1007/978-3-319-70338-1>>.

BOX, B. G. E. P. et al. **Time Series Analysis: Forecasting and Control**. [S.l.: s.n.], 2016. v. 68. ISBN 978-1-118-67502-1.

BUI, C. et al. Time series forecasting for healthcare diagnosis and prognostics with the focus on cardiovascular diseases. In: . [S.l.: s.n.], 2018. p. 809–818.

Chen, K. et al. Short-term load forecasting with deep residual networks. **IEEE Transactions on Smart Grid**, v. 10, n. 4, p. 3943–3952, 2019.

DUDEK, G. Neural networks for pattern-based short-term load forecasting: A comparative study. **Neurocomputing**, Elsevier, v. 205, p. 64–74, 2016. ISSN 18728286. Disponível em: <<http://dx.doi.org/10.1016/j.neucom.2016.04.021>>.

ESPINOZA, M. et al. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. **IEEE Transactions on Power Systems**, v. 20, n. 3, p. 1622–1630, Aug 2005.

FALLAH, S. N. et al. Computational intelligence on short-term load forecasting: A methodological overview. **Energies**, v. 12, n. 3, 2019. ISSN 19961073.

GONZÁLEZ-ROMERA, E.; JARAMILLO-MORÁN, M. Á.; CARMONA-FERNÁNDEZ, D. Monthly Electric Energy Demand Forecasting Based on Trend Extraction. **IEEE TRANSACTIONS ON POWER SYSTEMS**, v. 21, n. 4, p. 1946–1953, 2006.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>.

HAGAN, M. T.; MENHAJ, M. B. Training feedforward networks with the marquardt algorithm. **IEEE Transactions on Neural Networks**, v. 5, n. 6, p. 989–993, Nov 1994.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. [S.l.]: Prentice Hall, 1998.

HE, W. Load forecasting via deep neural networks. **Procedia Computer Science**, v. 122, p. 308–314, 01 2017.

HIPPERT, H. S.; PEDREIRA, C. E.; SOUZA, R. C. Neural networks for short-term load forecasting: A review and evaluation. **IEEE Transactions on Power Systems**, 2001. ISSN 08858950.

HONG, T.; FAN, S. Probabilistic electric load forecasting: A tutorial review. **International Journal of Forecasting**, Elsevier B.V., v. 32, n. 3, p. 914–938, 2016. ISSN 01692070. Disponível em: <<http://dx.doi.org/10.1016/j.ijforecast.2015.11.011>>.

HONG, W. C. Electric load forecasting by support vector model. **Applied Mathematical Modelling**, Elsevier Inc., v. 33, n. 5, p. 2444–2454, 2009. ISSN 0307904X. Disponível em: <<http://dx.doi.org/10.1016/j.apm.2008.07.010>>.

HOSSEN, T. et al. Short-term load forecasting using deep neural networks (dnn). In: **2017 North American Power Symposium (NAPS)**. [S.l.: s.n.], 2017. p. 1–6.

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2nd. ed. Australia: OTexts, 2018.

JHA, G. K.; SINHA, K. Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in india. **Neural Computing and Applications**, v. 24, n. 3, p. 563–571, 3 2014. ISSN 1433-3058. Disponível em: <<https://doi.org/10.1007/s00521-012-1264-z>>.

JIAO, R. et al. A Model Combining Stacked Auto Encoder and Back Propagation Algorithm for Short-Term Wind Power Forecasting. **IEEE Access**, IEEE, v. 6, p. 17851–17858, 2018. ISSN 21693536.

KARA, Y.; BOYACIOGLU, M. A.; BAYKAN Ömer K. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. **Expert Systems with Applications**, v. 38, n. 5, p. 5311 – 5319, 2011. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417410011711>>.

KARDAKOS, E. G. et al. Application of time series and artificial neural network models in short-term forecasting of pv power generation. In: **2013 48th International Universities' Power Engineering Conference (UPEC)**. [S.l.: s.n.], 2013. p. 1–6.



KUO, P. H.; HUANG, C. J. A high precision artificial neural networks model for short-Term energy load forecasting. **Energies**, 2018. ISSN 19961073.

KUSTER, C.; REZGUI, Y.; MOURSHED, M. Electrical load forecasting models: A critical systematic review. **Sustainable Cities and Society**, Elsevier, v. 35, n. July, p. 257–270, 2017. ISSN 22106707. Disponível em: <<http://dx.doi.org/10.1016/j.scs.2017.08.009>>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Springer Nature, v. 521, n. 7553, p. 436–444, may 2015. Disponível em: <<https://doi.org/10.1038/nature14539>>.

MAHALAKSHMI, B. G.; SRIDEVI, S.; RAJARAM, S. A survey on forecasting of time series data. In: **2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)**. [S.l.: s.n.], 2016. p. 1–8.

MOHAMMAD, F.; LEE, K. B.; KIM, Y.-C. Short term load forecasting using deep neural networks. 11 2018.

MORAES, L. A. **Desenvolvimento de uma abordagem fuzzy para estimação de demanda de potência em um sistema de distribuição de energia elétrica**. 2014. 74 p. Tese (Dissertação) — Universidade de São Paulo, 2014.

MUKHOPADHYAY, P. et al. Electricity Load Forecasting Using Fuzzy Logic. **2017 7th International Conference on Power Systems (ICPS)**, IEEE, p. 812–819, 2017.

NIE, H. et al. Hybrid of ARIMA and SVMs for short-term load forecasting. **Energy Procedia**, v. 16, n. PART C, p. 1455–1460, 2012. ISSN 18766102. Disponível em: <<http://dx.doi.org/10.1016/j.egypro.2012.01.229>>.

NIU, D.; WANG, Y.; WU, D. D. Power load forecasting using support vector machine and ant colony optimization. **Expert Systems with Applications**, Elsevier Ltd, v. 37, n. 3, p. 2531–2539, 2010. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2009.08.019>>.

RANAWEERA, D. K.; HUBELE, N. F.; KARADY, G. G. Fuzzy logic for short term load forecasting. **International Journal of Electrical Power and Energy Systems**, v. 18, n. 4, p. 215–222, 1996. ISSN 01420615.

RAZA, M. Q.; KHOSRAVI, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 50, p. 1352–1372, 2015. ISSN 18790690. Disponível em: <<http://dx.doi.org/10.1016/j.rser.2015.04.065>>.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais Para Engenharia E Ciências Aplicadas. Curso Prático**. [S.l.]: Artliber, 2015.

SINGH, A. K. et al. Load forecasting techniques and methodologies: A review. In: **2012 2nd International Conference on Power, Control and Embedded Systems**. [S.l.: s.n.], 2012. p. 1–10.

SINGH, S.; SINGH, R. ARIMA Based Short Term Load Forecasting for Punjab Region. **International Journal of Science and Research**, v. 4, n. 6, p. 2319–7064, 2013. Disponível em: <[www.ijsr.net](http://www.ijsr.net)>.

TASNIM, S. et al. Autoencoder for wind power prediction. **Renewables: Wind, Water, and Solar**, Springer Singapore, v. 4, n. 1, 2017. ISSN 2198-994X. Disponível em: <https://doi.org/10.1186/s40807-017-0044-x>.

VELASCO, L. C. P. et al. Load forecasting using autoregressive integrated moving average and artificial neural network. **International Journal of Advanced Computer Science and Applications**, v. 9, n. 7, p. 23–29, 2018. ISSN 21565570.

ZHANG, G.; QI, M. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, v. 160, n. 2, p. 501 – 514, 2005. ISSN 0377-2217. Decision Support Systems in the Internet Age. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0377221703005484>.