

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E
SISTEMAS**

CAROLINA DEINA

**MODELO PARA PREVISÃO DE DEMANDA DE ELETRICIDADE COM
REDES NEURAS ARTIFICIAIS INTEGRADAS A MÉTODOS
MULTICRITÉRIO**

DISSERTAÇÃO

PATO BRANCO

2020

CAROLINA DEINA

**MODELO PARA PREVISÃO DE DEMANDA DE ELETRICIDADE COM
REDES NEURAS ARTIFICIAIS INTEGRADAS A MÉTODOS
MULTICRITÉRIO**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Engenharia de Produção e Sistemas, do Programa de Pós-Graduação em Engenharia de Produção e Sistemas, da Universidade Tecnológica Federal do Paraná – *Campus Pato Branco* - PR.

Orientador: Prof. Dr. Flavio Trojan

Co-orientador: Prof. Dr. Hugo Siqueira

PATO BRANCO

2020

D324m Deina, Carolina.

Modelo para previsão de demanda de eletricidade com redes neurais artificiais integradas a métodos multicritério / Carolina Deina. – 2020.
119 f. : il.

Orientador: Prof. Dr. Flavio Trojan

Coorientador: Prof. Dr. Hugo Siqueira

Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná.
Programa de Pós-Graduação em Engenharia de Produção e Sistemas.
Pato Branco, PR, 2020.

Inclui bibliografia.

1. Energia - Consumo. 2. Variáveis. 3. Análise de séries temporais. 4. Modelos lineares (Estatística). I. Trojan, Flavio, orient. II. Siqueira, Hugo, Coorient. III. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia de Produção e Sistemas. IV. Título.

CDD 22. ed. 670.42

Ficha Catalográfica elaborada por
Suélem Belmudes Cardoso CRB9/1630
Biblioteca da UTFPR Campus Pato Branco



TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº 62

A Dissertação de Mestrado intitulada **“Modelo para previsão de demanda de eletricidade com redes neurais artificiais integradas a métodos multicritério”**, defendida em sessão pública pela candidata **Carolina Deina**, no dia 05 de março de 2020, foi julgada para a obtenção do título de Mestre em Engenharia de Produção e Sistemas, área de concentração Gestão dos Sistemas Produtivos, e aprovada em sua forma final, pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas.

BANCA EXAMINADORA:

Prof. Dr. Flavio Trojan - Presidente - UTFPR

Prof. Dr. Gilson Adamczuk Oliveira - UTFPR

Prof. Dr. José Donizetti de Lima – UTFPR

Prof. Dr. João Fausto Lorenzato de Oliveira - UPE

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Pato Branco, 28 de abril de 2020.

Prof. Dr. Fernando José Avancini Schenatto
Coordenador do Programa de Pós-Graduação em
Engenharia de Produção e Sistemas

AGRADECIMENTOS

Agradeço a todos aqueles que contribuíram direta e indiretamente para a realização deste trabalho, em especial:

Ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da UTFPR câmpus Pato Branco, que tem buscado sempre a melhoria do curso de pós-graduação.

À secretária do mestrado, Adriani Michelon, pela eficiência com o seu trabalho e pelo respeito e gentileza com que nos atende.

Aos órgãos/institutos EPE, SEEG, INMET, IPEA, IBGE e CCEE pela disponibilização das informações de forma diligente e acessível.

Aos membros da banca, por aceitar colaborar com a avaliação do trabalho.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela bolsa concedida durante o período de mestrado.

Aos meus amigos e colegas de mestrado, pelo compartilhamento de conhecimento, histórias e momentos de distrações.

Ao meu namorado Alexandre pela paciência, compreensão e pelas palavras de incentivo para que eu me mante-se firme durante a trajetória do mestrado.

As minhas irmãs Jéssica, Gabriela e Miquelina que me inspiram como pessoas e que sempre estiveram do meu lado para tudo sempre que fosse preciso.

Aos meus Pais Fátima e Pedro a quem devo minha vida, por estarem sempre ao meu lado, me incentivando desde criança a estudar.

Em especial, ao meu orientador Prof. Dr. Flávio Trojan e co-orientador Prof. Dr. Hugo Valadares Siqueira por toda confiança e dedicação depositadas nesse trabalho, contribuindo para a minha formação e amadurecimento enquanto pesquisadora.

AULA DE VOO

O conhecimento caminha
lento feito lagarta.
Primeiro não sabe que sabe
e voraz contenta-se com cotidiano orvalho
deixado nas folhas vividas das manhãs.

Depois pensa que sabe e se
fecha em si mesmo:
faz muralhas, cava trincheiras,
ergue barricadas.
Defendendo o que pensa saber
levanta certeza na forma de muro,
orgulha-se de seu casulo.

Até que maduro explode em voos
rindo do tempo que imagina saber
ou guardava preso o que sabia.
Voa alto sua ousadia
reconhecendo o suor dos séculos
no orvalho de cada dia.

Mas o voo mais belo
descobre um dia não ser eterno.
É tempo de acasalar:
voltar à terra com seus ovos
à espera de novas e prosaicas lagartas.

O conhecimento é assim: ri de si mesmo
E de suas certezas.
É meta de forma
metamorfose
movimento
fluir do tempo
que tanto cria como arrasa

a nos mostrar que para o voo
é preciso tanto o casulo
como a asa.

(Mauro Iasi)

RESUMO

DEINA, Carolina. **Modelo para previsão de demanda de eletricidade com redes neurais artificiais integradas a métodos multicritério**. 2020. 119 f. Dissertação (Mestrado em Engenharia de Produção e Sistemas) - Universidade Tecnológica Federal do Paraná. Pato Branco, 2020.

Para que não haja produção excessiva de eletricidade ou abaixo da demanda, o planejamento e a adoção de políticas eficientes são essenciais. Para isso, torna-se necessário aprimorar as técnicas de previsão de demanda de eletricidade em um país, pois uma produção que não atende à demanda causa racionamentos. Por outro lado, a produção excessiva gera desperdícios e custos elevados. Este trabalho teve por objetivo aprimorar as técnicas do processo de previsão de demanda de eletricidade, por meio de um modelo que integra redes neurais e o método multicritério ELECTRE I. Na literatura, constatou-se que ao incluir variáveis independentes no modelo de previsão, os resultados melhoram sensivelmente, porém ainda há uma carência de modelos eficientes que auxiliem neste processo. Assim, o modelo proposto apresenta cinco etapas distintas: 1) seleção das variáveis de entrada pelo método ELECTRE I, cujo o objetivo foi selecionar apenas as variáveis causais, excluindo as variáveis que podem estar apenas correlacionadas, porém sem influência direta no comportamento da demanda; 2) pré-processamento de dados com identificação do comportamento da demanda, considerando que poderá apresentar sazonalidade, tendência, ciclicidade e termos aleatórios e não estacionaridade; 3) previsão de demanda realizada com base nos modelos lineares de SEHW, AR e ARIMA e três modelos de Redes Neurais Artificiais, MLP, RBF e ELM; 4) pós-processamento de dados com a transformação dos dados nas escalas originais; 5) comparação dos modelos utilizados com base nas medidas de Erro: MSE, MAE e MAPE. Para testar o modelo proposto na metodologia, selecionou-se a série histórica de consumo de energia elétrica do Paraná. As previsões foram realizadas considerando 1, 3 e 6 passos à frente. Além disso, as entradas utilizadas pelos modelos neurais foram selecionadas por meio do método *Wrapper*. Como resultado, o ELECTRE I selecionou como variáveis explicativas a temperatura e evaporação média. Quando essas foram incluídas nos modelos preditivos constatou-se a melhora dos resultados de todas as RNAs. Além disso, para esse problema as RNAs tiveram desempenho superior comparado aos modelos lineares. Destaca-se as Redes ELM e RBF como os melhores preditores.

Palavras-chave: Consumo de Energia. Variáveis Causais. Variável dependente. Séries temporais. Modelos Lineares. Modelos Computacionais.

ABSTRACT

DEINA, Carolina. **Model for forecasting electricity demand with artificial neural networks integrated with multi-criteria methods**. 2020. 119 f. Dissertation (Master in Engineering of Production and Systems) - Federal Technological University of Paraná. Pato Branco, 2020.

In order to avoid excessive electricity production or below demand, efficient planning and adoption of policies are essential. For this, it is necessary to improve the techniques for forecasting electricity demand in a country, as a production that does not meet the demand for cause. On the other hand, excessive production generates waste and high costs. This work aimed to improve the techniques of the electricity demand forecasting process, using a model that integrates neural networks and the ELECTRE I multicriteria method. In the literature, see if you want to include independent variables in the forecasting model, the results improve significantly, but there is still a lack of efficient models that help in this process. Thus, the model presented has five distinct stages: 1) selection of input variables by the ELECTRE I method, whose objective was selected only as causal variables, excluding as variables that can only be correlated, but without direct influence demand behavior; 2) pre-processing of data with identification of demand behavior, considering the possibility of seasonal display, trend, cyclicity and random terms and non-stationarity; 3) demand forecast executed based on the linear models of SEHW, AR and ARIMA and three models of Artificial Neural Networks, MLP, RBF and ELM; 4) data post-processing with data transformation in the original dimensions; 5) Comparison of models used based on MSE, MAE and MAPE error measures. To test the proposed methodology model, select a historical series of electric energy consumption in Paraná. How they were performed considering 1, 3 and 6 steps ahead. In addition, the inputs used by the neural models were selected using the *Wrapper* method. As a result, or ELECTRE, I select as explanatory variables of temperature and average evaporation, when these were included in the constant predictive models, as improvements in the results of all as RNAs. In addition, for this problem, how RNAs performed better than linear models. It stands out as ELM and RBF Networks as the best editors.

Keywords: Energy consumption. Causal Variables. Dependent variable. Time series. Linear Models. Computational Models.

LISTA DE ILUSTRAÇÕES

Figura 1 - Artigos relacionados ao tema da pesquisa	19
Figura 2 - Estrutura geral da dissertação	24
Figura 3 – Conceitos básicos do apoio multicritério à decisão	26
Figura 4 - Tipos de problemáticas	29
Figura 5 - Situações básicas de preferências do decisor	30
Figura 6 - Situações consolidadas de preferências do decisor	30
Figura 7 - Características dos métodos da Família ELECTRE.....	32
Figura 8 - Exemplo de grafos de relações de sobreclassificação.....	34
Figura 9 - Escala de Saaty	35
Figura 10 - Matriz comparativa de critérios	36
Figura 11 - Matriz normalizada.....	36
Figura 12 - Índices de consistência aleatória	37
Figura 13 - Funcionamento genérico do neurônio biológico e artificial e seus processos sinápticos quando conectados a outros neurônios	44
Figura 14 - Função de Ativação de McCulloch e Pitts.....	46
Figura 15 - Função de Ativação Logística	46
Figura 16 - Função de Ativação Tangente Hiperbólica	47
Figura 17 - Função de Ativação Gaussiana	48
Figura 18 - Função de Ativação Linear	48
Figura 19 - Exemplo genérico da arquitetura de uma rede neural com múltiplas camadas e realimentação	50
Figura 20 - Modelo de rede Perceptron de Múltiplas Camadas (MLP).....	52
Figura 21 - Modelo de rede neural com função de base radial (RBF).....	54
Figura 22 - Fronteira limite com hiperplano e hipersférico	55
Figura 23 - Processo de ajuste utilizando o método k-means.....	56
Figura 24 - Modelo de rede Máquinas de Aprendizado Extremo (ELM).....	58
Figura 25 – Seleção de entradas de uma série temporal.....	60
Figura 26 - Estrutura do modelo.....	64
Figura 27 - Variáveis independentes elencadas na literatura.....	66
Figura 28 - Critérios elencados na literatura.....	66
Figura 29 - Condições da seleção do método ELECTRE I	67
Figura 30 - Tratamento dos dados	69
Figura 31 - Dados de consumo de energia do Estado do Paraná referente ao período de janeiro de 2004 até junho de 2019	75
Figura 32 - Dados de consumo de energia do Estado do Paraná referente ao período de janeiro de 2004 até junho de 2019, diferenciados e dessazonalizados	76
Figura 33 - Gráfico da função de autocorrelação parcial (FACP) para a série temporal diferenciada e dessazonalizada de consumo de energia elétrica no Paraná.....	77

Figura 34 - Gráfico da comparação dos dados reais com os dados previstos para 1 passo à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando os modelos SEHW, AR, ARIMA, MLP, RBF e ELM	79
Figura 35 - Gráfico da comparação dos dados reais com os dados previstos para 3 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando os modelos SEHW, AR, ARIMA, MLP, RBF e ELM	81
Figura 36 - Gráfico da comparação dos dados reais com os dados previstos para 6 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando os modelos SEHW, AR, ARIMA, MLP, RBF e ELM	82
Figura 37 - Gráfico Box-Plot com a dispersão do MSE(d) para as 30 simulações dos modelos lineares e RNAs	83
Figura 38 - Exemplo da Escala utilizada no critério g2.....	87
Figura 39 - Exemplo da Escala utilizada no critério g3.....	88
Figura 40 - Exemplo da Escala utilizada no critério g4.....	89
Figura 41 - Exemplo da Escala utilizada no critério g5.....	89
Figura 42 - Avaliação dos critérios realizada pelos decisores utilizando a escala de Saaty (1980).....	90
Figura 43 – Matriz de avaliação do AHP	91
Figura 44 – Pesos atribuídos aos critérios	91
Figura 45 – Ranking da pontuação dos critérios	91
Figura 46 – Ranking da pontuação dos critérios	92
Figura 47 – Matriz de concordância gerada pela comparação par a par de todas as alternativas	94
Figura 48 – Matriz de discordância gerada pela comparação par a par de todas as alternativas	95
Figura 49 – Matriz de seleção gerada pela condição definida pelos limiares de concordância e discordância em comparação com a matriz de concordância e discordância	96
Figura 50 – Grafo que representa o conjunto de Kernel das variáveis selecionadas para compor a previsão de demanda de energia elétrica	97
Figura 51 – Série temporal de evaporação referente ao período de janeiro de 2004 até junho de 2019.....	98
Figura 52 – Série temporal de temperatura referente ao período de janeiro de 2004 até junho de 2019.....	99
Figura 53 – Dados da evaporação e temperatura média do Estado do Paraná referente ao período de janeiro de 2004 até junho de 2019, dessazonalizados	100
Figura 54 – Gráfico da comparação dos dados reais com os dados previstos para 1 passo à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando RNAs MLP, RBF e ELM	102
Figura 55 – Gráfico da comparação dos dados reais com os dados previstos para 3 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando RNAs MLP, RBF e ELM	103
Figura 56 – Gráfico da comparação dos dados reais com os dados previstos para 6 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando RNAs MLP, RBF e ELM	104

Figura 57 – Gráfico Box-Plot com a dispersão do MSE(d) para as 30 simulações das RNAs 105

LISTA DE TABELAS

Tabela 1 – Testes não-paramétricos para análise de tendência e sazonalidade	75
Tabela 2 – Resultados das previsões para 1 passo à frente, comparado pelas medidas de erro MSE, MAE e MAPE	78
Tabela 3 – Ranking de desempenho das previsões para 1 passo à frente	79
Tabela 4 – Resultados das previsões para 3 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE	80
Tabela 5 – Ranking de desempenho das previsões para 3 passos à frente	80
Tabela 6 – Resultados das previsões para 6 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE	82
Tabela 7 – Ranking de desempenho das previsões para 6 passos à frente	82
Tabela 8 – Testes não-paramétricos para análise de tendência e sazonalidade	99
Tabela 9 – Resultados das previsões para 1 passo à frente, comparado pelas medidas de erro MSE, MAE e MAPE	101
Tabela 10 – Ranking de desempenho das previsões para 1 passo à frente	101
Tabela 11 – Resultados das previsões para 3 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE	102
Tabela 12 – Ranking de desempenho das previsões para 3 passos à frente	103
Tabela 13 – Resultados das previsões para 6 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE	104
Tabela 14 – Ranking de desempenho das previsões para 6 passos à frente	104
Tabela 15 – Ranking de desempenho geral das previsões para 1 passo à frente ..	106
Tabela 16 – Ranking de desempenho geral das previsões para 3 passos à frente	106
Tabela 17 – Ranking de desempenho geral das previsões para 6 passos à frente	106

SUMÁRIO

1 INTRODUÇÃO	15
1.1 OBJETIVOS DO TRABALHO	16
1.2 TRABALHOS RELACIONADOS	17
1.3 JUSTIFICATIVA	21
1.4 DELIMITAÇÃO DA PESQUISA	23
1.5 ORGANIZAÇÃO DO TRABALHO	24
2 MÉTODOS MULTICRITÉRIOS DE APOIO À DECISÃO	26
2.1 TIPO DE PROBLEMÁTICA	28
2.2 MODELAGEM E ESTRUTURA DAS PREFERÊNCIAS	29
2.3 MÉTODOS MULTICRITÉRIO DA FAMÍLIA ELECTRE	31
2.3.1 Electre I	33
2.3.1.1 Método AHP	35
3 MÉTODOS DE PREVISÃO UTILIZADOS NESSA PESQUISA	38
3.1 MODELOS LINEARES DE PREVISÃO	38
3.1.1 Suavização Exponencial de Holt-Winters (SEHW)	40
3.1.2 Autorregressivo (AR)	41
3.1.3 Autorregressivo Integrado de Médias Móveis (ARIMA)	41
3.2 REDES NEURAIS ARTIFICIAIS	42
3.2.1 Neurônios Biológicos Versus Artificiais	43
3.2.2 Funções de Ativação	45
3.2.3 Arquiteturas das Redes Neurais Artificiais e Processos de Treinamento	49
3.2.3.1 Perceptron de múltiplas camadas (MLP)	51
3.2.3.1.1 <i>Treinamento da rede MLP</i>	52
3.2.3.2 Redes neurais com função de base radial (RBF)	53
3.2.3.2.1 <i>Treinamento da rede RBF</i>	55
3.2.3.3 Máquinas de aprendizado extremo (ELM)	57
3.2.3.3.1 <i>Treinamento da rede ELM</i>	58
3.2.4 Método Wrapper Para Seleção de Entradas	60
4 METODOLOGIA	63
4.1 ENQUADRAMENTO METODOLÓGICO	63
4.2 ESTRUTURAÇÃO DO MODELO	64
4.2.1 Modelo para Seleção de Variáveis	65
4.2.2 Pré-processamento de Dados	68
4.2.3 Modelos Preditivos	70
4.2.4 Pós-Processamento de Dados	72
4.2.5 Desempenho dos Modelos Preditivos	72
5 RESULTADOS	74
5.1 PREVISÃO DA VARIÁVEL DEPENDENTE	74

5.1.1 Análise estatística e definição de parâmetros.....	74
5.1.2 Previsão da variável independente	78
5.2 SELEÇÃO DE VARIÁVEIS INDEPENDENTES	84
5.2.1 Aplicação do AHP	90
5.2.2 Aplicação do ELECTRE I	92
5.3 PREVISÃO COM VARIÁVEIS INDEPENDENTES	98
5.4 COMPARAÇÃO GERAL DAS PREVISÕES	105
6 CONCLUSÕES	108
TRABALHOS FUTURO	110
REFERÊNCIAS.....	111

1 INTRODUÇÃO

O consumo mundial de energia elétrica vem crescendo em uma escala global, representado hoje aproximadamente 21.771,69 bilhões kWh, sendo que os países desenvolvidos consomem 43,26% desse valor global (*Central Intelligence Agency - CIA, 2018*). Grande parte desse crescimento ocorreu nos setores residencial, comercial e de serviços públicos, sendo o setor industrial ainda o maior consumidor, mesmo com a sua participação reduzida ao longo dos anos em decorrência do aumento de eficiência energética (Agência Internacional de Energia - IEA, 2018).

Os demais setores que consomem eletricidade, como para usos agropecuários e de transportes são relativamente pequenos. No entanto, outro setor que apresentou forte crescimento foi o de transportes rodoviários. Essa expansão foi iniciada quando os veículos elétricos começaram a ser comercializados nos países desenvolvidos (IEA, 2018). Além desses países desenvolvidos, outros quatro que não compõe esse grupo representam juntos 38,39% do consumo global, a saber: China, Índia, Rússia e Brasil (CIA, 2018).

Nesses países a demanda industrial é predominante e responde por metade desse valor (CIA, 2018). Os países enfrentam desafios contínuos no setor energético. Assim, para que estes se mantenham em funcionamento é necessário aperfeiçoar decisões já tomadas e incorporar questões de futuro, como por exemplo a expansão da matriz energética, garantia da segurança energética, atendimento aos objetivos do desenvolvimento sustentável e garantia da justiça tarifária (Ministério de Minas e Energia - MME, 2019).

Como a eletricidade não poder ser estocada, os desequilíbrios entre a oferta e a demanda precisam ser evitados, pois implicam em elevados custos de produção que, conseqüentemente, serão repassados aos usuários (CHUJAI; KERDPRASOP; KERDPRASOP, 2013). Tanto o excesso quanto a redução da produção de eletricidade são nocivos à economia de um país, pois a produção excessiva gera desperdícios, enquanto a redução pode provocar racionamentos indesejados (TUTUN; CHOU; CANYLMAZ, 2015).

Nesse sentido, para que decisões sejam viáveis economicamente é essencial entender o comportamento da demanda por eletricidade, o que possibilita identificar previsões e tendências futuras a fim de alocar corretamente os investimentos no setor elétrico. A demanda por eletricidade pode ser prevista por diversos modelos

matemáticos e/ou estatísticos, porém desenvolver um modelo que apresente resultados coerentes não é a única atividade crítica desse processo (WU *et al.*, 2018).

Existem ainda, outros fatores que influenciam o consumo de eletricidade, que podem ser considerados como entradas para o modelo em determinadas situações (WU *et al.*, 2018). Ao mesmo tempo em que utilizar fatores independentes pode reduzir os erros, o modelo de previsão pode se tornar mais complexo, pois considerar variáveis que realmente influenciam a demanda requer um estudo mais detalhado, que envolve reconhecer critérios de influência (PAO, 2006), uma vez que essas influências são diferentes para cada ambiente (PAO, 2006; WU *et al.*, 2018).

Neste contexto, o estudo de modelos de previsão que integrem a análise multicritério cria a oportunidade de incorporar critérios até então não considerados pelos modelos existentes. Quanto mais eficiente um modelo de previsão, melhor será a clareza sobre os possíveis investimentos em fontes alternativas para o fornecimento de eletricidade e mais eficaz será o planejamento para o uso dos recursos disponíveis relativos à geração da eletricidade. Isso envolve tanto as dimensões econômica, social e ambiental concomitantemente, quanto à busca da eficiência energética para os processos produtivos.

Assim, no presente trabalho foram investigadas as ferramentas viáveis no desenvolvimento de um modelo para aperfeiçoar o processo de previsão de demanda por eletricidade. Inicialmente, selecionando os fatores que influenciam a demanda por um método multicritério, e posteriormente gerando previsões para demanda de energia elétrica com Redes Neurais Artificiais (RNAs) em uma aplicação específica, e por fim comparando os resultados com os modelos lineares de previsão.

1.1 OBJETIVOS DO TRABALHO

O estudo desenvolvido nessa dissertação apresenta como ponto de partida a análise de decisões relacionadas à previsão de demanda por eletricidade. Isso, devido à relevância da precisão no processo de previsão de demanda e para o planejamento energético. Visto que a ocorrência de diferenças entre a oferta e a demanda de qualquer produto ou serviço público implica em riscos para o desenvolvimento de qualquer país, a principal pergunta a ser respondida por essa pesquisa é: “**Quais**

variáveis independentes devem ser consideradas em um processo de previsão de demanda de eletricidade, e quais métodos de previsão podem apresentar maior acuracidade na previsão ou projeção de demanda futura?”

Para responder essa pergunta, neste trabalho é proposto um modelo de previsão de demanda que seleciona as variáveis independentes que influenciam diretamente na demanda, e a utilização de RNAs que são posteriormente comparadas com os tradicionais modelos de previsão. Assim, o objetivo principal desse estudo é **desenvolver um modelo integrado de previsão de demanda por eletricidade para aperfeiçoar os métodos já existentes.**

Os objetivos específicos desse trabalho são:

- ✓ Identificar na literatura as variáveis utilizadas para previsão de demanda e a motivação que os autores encontraram para a escolha dessas variáveis, bem como os métodos tradicionais de previsão utilizados globalmente;
- ✓ Selecionar as variáveis consideradas pelos decisores como influenciadoras na previsão de demanda por eletricidade e testar os modelos preditivos com a inclusão dessas variáveis identificando o desempenho;
- ✓ Comparar os resultados de previsão obtidos entre as RNAs e modelos lineares, e também analisar o desempenho das RNAs quando adicionadas das variáveis independentes.

1.2 TRABALHOS RELACIONADOS

De acordo com Vu, Muttaqi e Agalgaonkar (2015), as influências que atuam no consumo de eletricidade tendem a ser não lineares e não estacionárias, e é importante que a relação entre as variáveis de entrada com a variável de saída seja descoberta para a atividade de previsão. A relação entre as variáveis dependentes e independentes já foi investigada em várias pesquisas. Algumas dessas pesquisas estão resumidas em ordem cronológica na Figura 1, e que representam os estudos mais relevantes para o tema.

Sadownik e Barbosa (1999) realizaram previsões da demanda de eletricidade em três regiões no Brasil; São Paulo, Rio de Janeiro e Minas Gerais. As previsões foram realizadas pelo modelo dinâmico não linear e comparadas com previsões econométricas utilizadas pelas concessionárias brasileiras, onde foi considerado o Produto Interno Bruto (PIB) como variável independente. Conforme os autores, o modelo foi capaz de ajustar automaticamente pequenas mudanças estruturais nos dados, sem a necessidade de inclusão de dados do PIB.

Darbellay e Slama (2000) apresentaram uma comparação entre as previsões realizadas com RNAs e o modelo linear ARIMA para o consumo de eletricidade na Tchecoslováquia. Concluíram que para a modelagem univariada, as capacidades de previsão nos dois modelos eram bem similares. Para modelagem multivariada, ao adicionar a variável temperatura como uma entrada externa permitiu integrar mais informações nas RNAs, apresentando uma leve melhora em relação aos modelos lineares.

Pao (2006) utilizou Regressão Log-linear Múltipla (LNREG), Regressão da Superfície de Resposta (RSREG) e Autorregressivo de Média Móvel com Entradas Exógenas (ARMAX) comparados com as RNAs para realizar previsões do consumo de eletricidade em Taiwan. O autor considerou variáveis econômicas influenciando a demanda, sendo constatado uma maior influência das variáveis população e renda nacional. As RNAs apresentaram resultados mais precisos, segundo o autor, esse resultado é decorrente da capacidade que elas têm para captar efeitos de integração não-linear por meio de um processo de aprendizado.

Adam, Elahee e Dauhoo (2011) utilizaram RNAs para prever a demanda de eletricidade das ilhas Maurício, na África Ocidental. As variáveis de entrada foram selecionadas por um processo de difusão de Gompertz não homogêneo (NHGDP) com base em uma abordagem de Algoritmo Genético (GA). Essas variáveis são: PIB e os fatores climáticos, como temperatura, horas de sol e umidade. Segundo os autores, os resultados mostraram que o modelo forneceu previsões confiáveis para a demanda mensal de eletricidade.

Hernández *et al.* (2013) utilizaram apenas a rede MLP e diferentes variáveis meteorológicas para prever a demanda de energia em Sória na Espanha. Os autores realizaram uma análise de autocorrelação das variáveis utilizando o coeficiente de *Pearson*, e identificaram que a temperatura, a umidade relativa e a radiação solar tinham correlação com a demanda de eletricidade. Previsões adicionando

progressivamente variáveis de entrada foram realizadas e constataram que o modelo que utilizou a variável de radiação solar apresentou maior precisão.

Figura 1 - Artigos relacionados ao tema da pesquisa

ANO	AUTORES	Metodologia	Variáveis Dependentes	Variáveis Independentes
1999	Sadownik e Barbosa	Modelos dinâmicos não lineares e econométricos	- Consumo	- PIB
2000	Darbellay e Slama	Redes Neurais Artificiais e ARIMA	- Consumo	- Temperatura
2006	Pao	Modelos estatísticos LNREG, RSREG, ARMAX e RNA	- Consumo	- Renda nacional, - População, PIB - Índice preços consumidor
2011	Adam, Elahee e Dauhoo	Rede Neurais Artificiais	- Consumo	- PIB, - Temperatura, - Horas de sol e umidade
2013	Hernández et al.	Rede Neurais Artificiais	- Consumo	Precipitação, temperatura do ar, velocidade média do vento, direção do vento, umidade relativa, pressão e radiação solar
2015	Cui e Peng	ARIMAX, AR, ARMA e Rede Neurais Artificiais	- Consumo	- Temperatura,
2015	Vu, Muttaqi e Agalgaonkar	Modelos de regressão múltipla	- Consumo	População, PIB, preço, velocidade do vento, umidade, evaporação, dias chuvosos e horas de insolação.
2016	Torrini et al.	Abordagem de lógica difusa Fuzzy	- Consumo	- PIB - População
2016	Suganthi e Samuel	Modelos econométricos	- Consumo	PIB, preço da energia e população.
2016	Wang, Zhang e Nie	Algoritmo Artificial Bee Colony, Redes Neurais Artificiais e regressão quadrática	- Consumo	PIB, investimentos, população, urbanização, consumo das famílias e emissão de carbono.
2018	Mohammed	Modelo de regressão linear logarítmico e Redes Neurais Artificiais	- Consumo	População, produto nacional bruto, índice de preços ao consumidor, efeito de guerra e temperatura.
2018	Wu et al.	Modelos de predição cinza com convolução integral, acumulação de ordens e cinza tradicional	- Consumo	PIB, renda disponível per capita urbana, população total, valor da produção industrial e investimento em ativos fixos

Fonte: Autoria própria (2019).

Cui e Peng (2015) analisaram a relação entre carga elétrica e temperatura diária em uma cidade da China. Um modelo Autorregressivo Integrado de Média Móvel com Entradas Exógenas (ARIMAX) aprimorado foi proposto para realizar as previsões de demanda de eletricidade da cidade. Os autores verificaram que a quantidade de informações do modelo ARIMAX aprimorado foi menor que nos métodos clássicos, e o erro relativo também foi menor que os modelos AR, Autorregressivo de Média Móvel (ARMA) e RNA simples.

Vu, Muttaqi e Agalgaonkar (2015) utilizaram o modelo estatístico de regressão múltipla para prever a demanda de energia do estado de New South Wales na Austrália, tendo como entradas variáveis climáticas e socioeconômicas selecionadas via de análises estatísticas. As demandas mensais previstas foram obtidas com o modelo proposto e similares às demandas reais de eletricidade, com as variações de erros de previsão dentro dos limites aceitáveis.

Torrini *et al.* (2016) utilizaram uma abordagem de lógica difusa para fornecer as previsões anuais de demanda de eletricidade a longo prazo do Brasil, utilizando as variáveis: população e valor adicional do PIB. Duas outras abordagens, além do modelo *Fuzzy*, foram usadas para comparações: um método de dois parâmetros de Holt e as projeções oficiais do setor fornecidas por uma empresa de pesquisa de energia. Nessa comparação, os resultados para os modelos *Fuzzy* tiveram um resultado de previsão mais acurado.

Suganthi e Samuel (2016) desenvolveram modelos econométricos para prever o consumo de energia utilizando dados do PIB, preço da energia e taxa de crescimento da população. A relação das variáveis independentes com o consumo de energia foi verificada via modelos estatísticos. Os autores não utilizaram modelos para comparação, mas consideraram as previsões satisfatórias, e sugeriram que os administradores do país precisariam formular políticas para a utilização máxima das fontes de energia renováveis.

Wang, Zhang e Nie (2016) sugeriram um algoritmo *Artificial Bee Colony* (ABC) melhorado, combinado com Regressão Linear Progressiva (MLR) para explorar o mecanismo de influência de vários fatores (econômicos, sociais, industriais e ambientais) no consumo de eletricidade chinês e prever a demanda futura. Os resultados indicaram que esse modelo proposto supera os modelos tradicionais como o RNA simples e o modelo de regressão quadrática.

Mohammed (2018) buscou prever o consumo de energia do Iraque utilizando um modelo de regressão linear logarítmico comparado a um modelo de RNA. Para entradas, utilizou dados históricos de população, PIB, índice de preços, temperatura, e efeitos da guerra de 2003. O autor comparou seus resultados com outros estudos e constatou que os pesos dos fatores influentes têm impactos diferente na previsão de outros países. Neste trabalho, o modelo de regressão logarítmico linear foi apropriado porque foi eficiente em estimar a demanda de carga.

Wu *et al.* (2018) realizaram previsões de demanda elétrica para a província de Shandong na China utilizando um novo modelo de previsão multivariada cinza GMC (1, N) com acumulação de ordens fracionais que considerou variáveis socioeconômicas como independentes no modelo. Como resultado, o modelo forneceu desempenhos de previsão notáveis em comparação com o modelo tradicional de previsão cinza.

1.3 JUSTIFICATIVA

Como visto na seção anterior, diferentes variáveis meteorológicas, socioeconômicas e ambientais foram incluídas na maioria dos problemas relacionados a previsão de demanda por eletricidade, porém, o que se vê na literatura, é que a seleção dessas variáveis independentes tem sido motivadas principalmente pela escolha dos próprios pesquisadores e portanto, não garante que esses fatores sejam o melhor conjunto para a resolução dos problemas de previsão.

Mesmo que em alguns estudos os autores considerem seus resultados satisfatórios, estes poderiam ser melhorados com a retirada ou aquisição de informações ao modelo. Essa constatação, é semelhante a que foi encontrada por Vu, Muttaqi e Agalgaonkar (2015). Segundo os autores, o uso de poucas variáveis pode tornar o modelo fraco, enquanto considerar o maior número possível pode tornar o modelo computacionalmente intensivo, pois é atribuído ao problema dados que são desnecessários.

No trabalho de Sadownik e Barbosa (1999), por exemplo, houve pioras ao incluir variáveis, Hernández *et al.* (2013) adicionaram progressivamente variáveis e verificaram que nem todas elas tinham influências no modelo. No trabalho de

Darbellay e Slama (2000) os resultados comparando RNAs e modelos lineares foram similares, sendo que ao incluir a variável temperatura houve uma leve melhora, mas talvez, esse modelo poderia ter sido melhorado significativamente se outras variáveis listadas na literatura também fossem incluídas ao modelo.

Vu, Muttaqi e Agalgaonkar (2015) e Suganthi e Samuel (2016) utilizaram apenas modelos lineares, mas não fizeram comparações com outros modelos, o que não garante de fato que suas previsões são as mais apropriadas para resolver o problema em questão. Alguns trabalhos que buscaram conhecer a relação entre as variáveis, o fizeram por meio de testes estatísticos, como por exemplo, análise de correlação e multicolinearidade, sendo trabalhos dos autores Hernández *et al.* (2013), Vu, Muttaqi e Agalgaonkar (2015), Suganthi e Samuel (2016), Wu *et al.* (2018) e Mohammed (2018).

Apesar desses estudos selecionarem um grupo de fatores com uma justificativa, não necessariamente incluíram ao problema variáveis que são de fato “causais”, pois esses testes estatísticos não são capazes de explicar se a relação que existe entre elas implica na “causa” de uma alteração na demanda por eletricidade.

Como já exposto, a eletricidade é um recurso que não pode ser armazenado, nesse sentido é importante que a previsão de demanda seja realizada com uma alta precisão para que não ocorra desperdícios ou escassez desse recurso (CHUJAI; KERDPRASOP; KERDPRASOP, 2013). Na literatura há constatação de que a inclusão de fatores independentes no problema pode melhorar os resultados de previsão, mas selecionar um conjunto de variáveis que de fato causam a demanda ainda é uma lacuna que carece de estudos. Nesse sentido, propor uma ferramenta que auxilie de forma efetiva a seleção de variáveis independentes para a previsão de demanda tem relevância para a área, além de que não há evidências de métodos de apoio a decisão multicritério utilizados para esse propósito.

Além dos fatores para entrada no modelo, a escolha do método de previsão que será utilizado também é uma etapa importante nesse processo. A utilização de RNAs vem se mostrando uma alternativa viável para essa tarefa, além de que ainda há oportunidades para explorar a utilização de redes como por exemplo, ELM.

1.4 DELIMITAÇÃO DA PESQUISA

Para que a revisão sistêmica da literatura (RSL) pudesse ser realizada inicialmente limitou-se as buscas utilizando palavras-chave relacionadas aos dois eixos da pesquisa, a saber: previsão de demanda e setor de energia elétrica. As buscas foram realizadas nas bases de dados *Scopus* e *Web of Science*, as quais estão disponíveis no portal de periódicos da CAPES e que foram selecionadas de acordo com o alinhamento ao tema da pesquisa, disponibilidade para realizar buscas por títulos, resumos, palavras-chave e expressões booleanas.

Ambas bases de dados têm seu uso justificado por criarem medidas amplamente utilizadas na avaliação da qualidade de periódicos. A base *Scopus* faz uso da métrica *Scientific Journal Rankings (SJR)*, enquanto a base *Web of Science* utiliza a métrica *Journal Citation Reports (JCR)*. Outra limitação em relação a pesquisa bibliográfica se deu em relação ao horizonte temporal utilizado para buscas, o qual restringiu-se ao período de 1990 até 2018, objetivando reduzir a busca para dados e informações atualizadas referente ao tema.

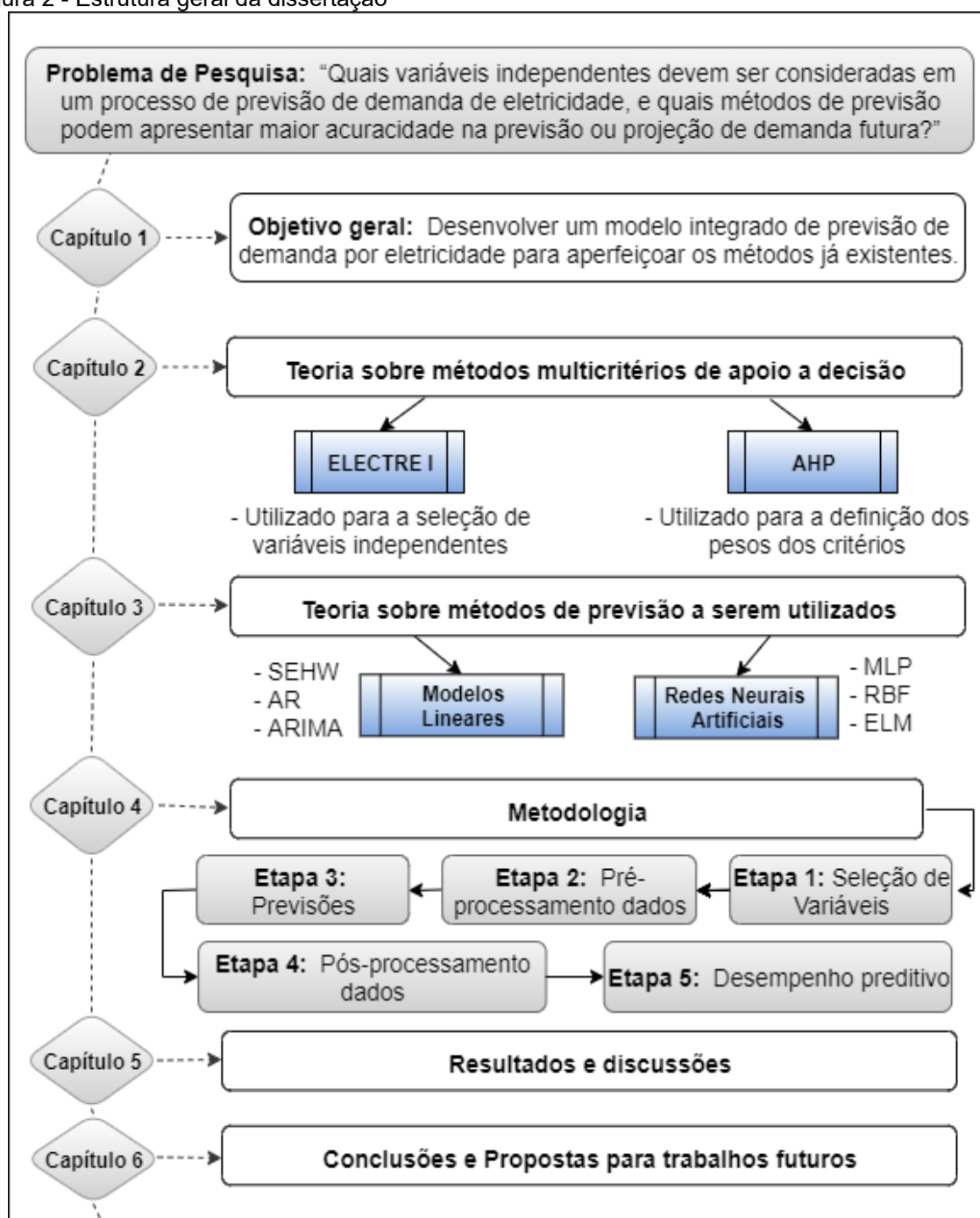
Visto na literatura a oportunidade de incluir variáveis exógenas no processo de previsão, para que isso fosse realizado de forma estruturada, os métodos multicritérios selecionados para essa tarefa restringiram-se aos *ELimination Et Choice Traidusaint la REalité I (ELECTRE I)* e *Analytic Hierarchy Process (AHP)*. Com relação aos métodos de previsão, dentre os diversos modelos lineares e redes neurais existentes, essa pesquisa se limitou as técnicas: SEHW, AR, ARIMA, MLP, RBF e ELM.

Para comparação das técnicas de previsão, as medidas de acurácia também foram limitadas para: Erro Quadrático Médio (MSE – *Mean Squared Error*), Erro Absoluto Médio (MAE – *Mean Absolute Error*) e Erro Percentual Absoluto Médio (MAPE – *Mean Absolute Percentual Error*). E por fim, para que o modelo proposto pudesse ser validado, o campo de aplicação limitou-se ao consumo de energia do estado Paraná.

1.5 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em 6 capítulos, na Figura 2 é apresentado a estrutura geral dessa dissertação.

Figura 2 - Estrutura geral da dissertação



Fonte: Autoria própria (2019).

No Capítulo 1 é apresentado a introdução ao tema, artigos relacionados, objetivo geral e específicos, justificativa e delimitação da pesquisa. No capítulo 2 são apresentados conceitos sobre métodos multicritérios de apoio a decisão, bem como o método ELECTRE I utilizado para a seleção de variáveis independentes e o AHP que serviu como propósito para definir o grau de importância dos critérios.

O capítulo 3 discorre sobre os métodos de previsão que foram utilizados para a previsão de demanda de eletricidade deste trabalho, divididos em: modelo linear e RNAs. Em relação aos modelos lineares SEHW, AR e ARIMA estão apresentados os conceitos e componentes que estão envolvidos em sua construção. Para os modelos de RNAs foram apresentados conceitos iniciais, analogia dos neurônios biológicos e artificiais, principais funções de ativação dos neurônios e arquiteturas e processos de treinamentos das redes MLP, RBF e ELM.

Também apresentou o método *Wrapper* para seleção progressiva de entradas as RNAs. No capítulo 4 foi apresentado os materiais e métodos utilizados para o desenvolvimento desse estudo. Já o capítulo 5 apresentou uma aplicação prática da metodologia proposta, bem como os resultados e discussões dessa aplicação. E por último, o capítulo 6 traz as conclusões da pesquisa e propostas para direcionamentos de futuras pesquisas.

2 MÉTODOS MULTICRITÉRIOS DE APOIO À DECISÃO

Diariamente o ser humano é exposto a situações sobre as quais precisará se decidir. Embora essa seja uma atividade comum e rotineira, ao contrário do que parece, pode tornar-se complexa e conflitante a medida em que diferentes possibilidades de escolhas (aqui chamadas de alternativas) e formas de avaliação que refletem as preferências do decisor (aqui chamadas de critérios) estão inseridas no problema. Dificilmente, problemas que envolvem certo grau de complexidade encontrarão uma alternativa que se destaque para todos os critérios (TROJAN, 2012).

Dessa forma, os conflitos e as dificuldades que o decisor pode encontrar nesse processo são evitadas/reduzidas com o auxílio de métodos multicritérios de apoio a decisão (GOMES; ARAYA; CARIGNANO, 2004; TROJAN, 2012). Na Figura 3, são introduzidos alguns conceitos básicos envolvidos em processos de decisão, para que estes sejam familiarizados no decorrer desse trabalho. São eles: atores, decisor, analista, alternativas, critérios e pesos.

Figura 3 – Conceitos básicos do apoio multicritério à decisão

Componentes	Descrição
Atores	Indivíduos que tem interesse na decisão a ser tomada, pois são atingidos diretamente ou indiretamente pelas consequências das decisões (CAMPOS, 2011).
Decisor	Indivíduo ou grupo de indivíduos que possui no papel mais importante no processo de tomada de decisão, pois são eles que proporcionam o juízo de valor final, através da avaliação de alternativas do problema de acordo com sua relação de preferência (GOMES, 2007).
Analista	Indivíduo ou grupo de indivíduos que tem como função auxiliar no processo decisório. São eles os responsáveis por sistematizarem o processo e modelar as preferências do decisor da maneira mais objetiva possível (GOMES; ARAYA; CARIGNANO, 2004).
Alternativas (A)	É o conjunto de opções possíveis em que o decisor irá fazer sua escolha. Deve-se considerar todas as alternativas, mesmo que não seja viável sua implementação, sempre que houver algum interesse em relacioná-la ao processo decisório (ROY, 1991).
Crítérios (K)	São parâmetros de avaliação para um conjunto de alternativas, podendo abranger natureza subjetiva ou objetiva. Sendo fundamental que estes reflitam as preferências reais do decisor (TSOUKIAS, 2008).
Pesos	Para cada critério atribui-se um peso que expressa o grau de importância de cada critério (BELTON; STEWART, 2002).

Fonte: Autoria própria (2019).

Importante mencionar, que para esse trabalho as variáveis que podem influenciar a demanda de energia serão aqui chamadas de alternativas, como exemplos, tem-se: temperatura, PIB, preço, entre outras. O que se sabe, é que nem todas as alternativas listadas na literatura influenciam a demanda de energia do estado do Paraná (recorte espacial), e por isso, devem ser avaliadas com base em critérios, como por exemplo: quantidade de dados disponíveis, correlação com a demanda de energia e assim por diante.

Tendo os conceitos básicos aqui introduzidos, outra informação importante a se expor diz respeito a classificação dos métodos multicritérios utilizados nessa pesquisa. Já que diferentes classificações estão disponíveis na literatura, para esse trabalho adotou-se a de Roy (1996), o qual os agrupa em três abordagens principais, a saber: critério único de síntese, julgamento interativo e sobreclassificação ou subordinação.

Na abordagem do critério único de síntese é agregado diferentes pontos de vista em uma única função de síntese, ranqueando soluções e permitindo que problemáticas de escolha e ordenação sejam tratadas por esses métodos. Destaca-se nessa família o método *Multiple Attribute Utility Theory* (MAUT) e o AHP (SILVA, 2016).

Já na abordagem do julgamento interativo faz-se uso de cálculos matemáticos, interativos e sucessivos para fornecer soluções de compromisso. Nessa família, busca-se uma alternativa que seja claramente superior a todos os objetivos estabelecidos, através de tentativas e erros e estruturas de programação matemática multiobjectivo. Métodos como *STEp Method* (STEM) e o *Interval Criterion Weights* (ICW) são exemplos de aplicação (ALMEIDA, 2011).

Na abordagem de sobreclassificação ou subordinação são representadas as preferências do decisor. Nessa família, as alternativas são comparadas par a par e verificado qual das alternativas é superior em cada etapa para os critérios estabelecidos, tendo como melhor alternativa aquela que apresentar superioridade na maioria dos critérios (ALMEIDA, 2012).

Conforme Haralambopoulos e Polatidis (2003) métodos dessa família são os mais indicados para solucionar problemas relacionados à energia e planejamento ambiental. Dentro dessa abordagem, destacam-se os métodos *ELimination Et Choice Traidusaint la REalité* (ELECTRE) e o *Preference Ranking Organization METHod for Enrichment Evaluations* (PROMETHEE) (CAMPOS, 2011).

Como já apresentado no Tópico 1.2 desse trabalho, a seleção de variáveis que influenciam a demanda por energia (alternativas), é motivada na maior parte dos casos pelo próprio decisor, sem que tenham critérios claros para isso. Portanto, é nesse contexto que os métodos de análise multicritério podem contribuir, fazendo a seleção de um grupo de variáveis influentes para previsão de demanda por eletricidade.

Uma abordagem de sobreclassificação, com base na família ELECTRE, mais especificamente o ELECTRE I é utilizada para a tarefa de seleção, assuntos que serão abordados com mais detalhes nos Tópicos 2.3 e 2.4. Além do entendimento do método, é importante compreender a escolha do mesmo para a solução desse problema, portanto, também será apresentado nas próximas seções alguns fatores que influenciam essa escolha, como: o tipo de problemática (Tópico 2.1) e a modelagem e estrutura das preferências do decisor (Tópico 2.2).

2.1 TIPO DE PROBLEMÁTICA

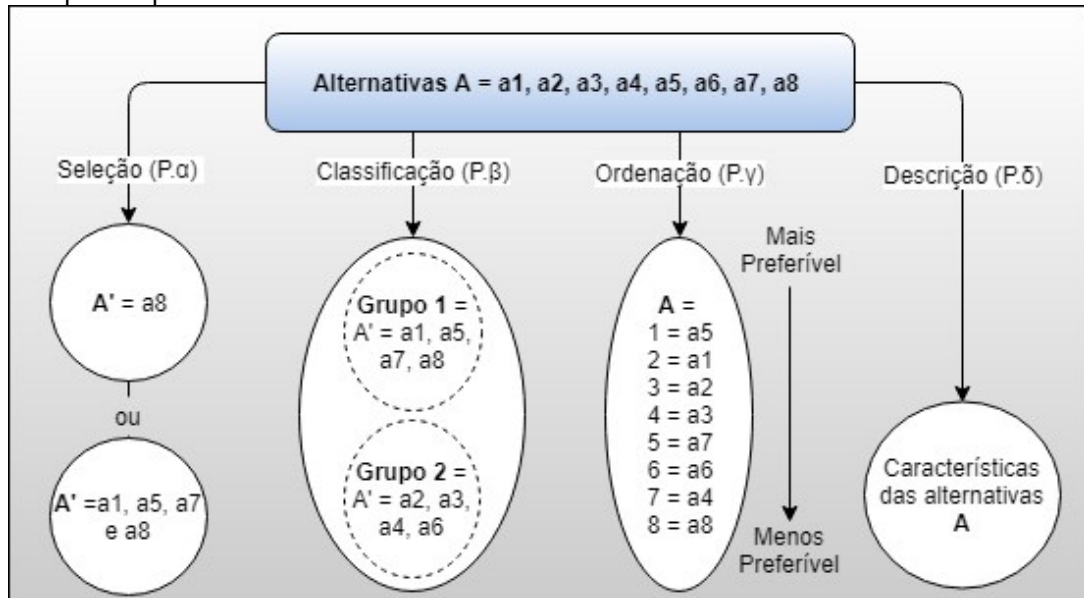
Durante a estruturação do processo decisório, o qual envolve a escolha do método multicritério de apoio a decisão, é primordial entender o tipo de problemática que se pretende resolver para que um método multicritério possa ser direcionado (CAMPOS, 2011). Na Figura 4 é ilustrado de forma genérica esses quatro tipos de problemáticas.

Roy (1996) e Gomes e Gomes (2012) descrevem essas problemáticas como:

- a) **Problemática de seleção (P. α):** tem como objetivo selecionar a melhor alternativa ou subconjunto de alternativas A' dentro do conjunto A . Essa problemática se caracteriza como escolha;
- b) **Problemática de classificação (P. β):** busca-se alocar o conjunto de alternativas A em determinadas classes, as quais são definidas por normas previamente estabelecidas. Procedimento de triagem/classificação;
- c) **Problemática de ordenação (P. γ):** tem como objetivo estabelecer uma ordem para cada ação contida dentro de um grupo de alternativas em classe de equivalências; e

- d) Problemática de descrição (P.δ):** descreve detalhadamente as alternativas e suas possíveis consequências, para que o decisor possa compreender melhor as características de cada ação.

Figura 4 - Tipos de problemáticas



Fonte: Adaptado de Campos (2011).

É importante mencionar que esses tipos de problemáticas não são independentes entre si, pois é possível fazer uso de mais de uma em problemas multicritério (GOMES; ARAYA; CARIGNANO, 2004).

2.2 MODELAGEM E ESTRUTURA DAS PREFERÊNCIAS

A modelagem de preferência tem como objetivo estudar a relação entre duas alternativas mediante a preferência ou indiferença do decisor. As comparações entre as alternativas são expressas via relações binárias. Algumas propriedades dessa relação são descritas por Gomes, Araya e Carignano (2004), se considerando que a relação binária (\mathfrak{R}) no conjunto (A), em que a notação é $a\mathfrak{R}b$ para $(a, b) \in \mathfrak{R}$, tem-se:

- a) Relação binária reflexiva** se $(a\mathfrak{R}a)$;
- b) Relação binária irreflexiva** se $(\text{não } a\mathfrak{R}a)$;
- c) Relação binária simétrica** se $(a\mathfrak{R}b)$ ou $(b\mathfrak{R}a)$;

d) **Relação binária assimétrica** se $(a \succ b)$ não $(b \succ a)$; e

e) **Relação binária transitiva** se $(a \succ b)$ e $(b \succ c)$ implicam em $(a \succ c)$, para todo a, b, c pertencente a A .

Quando o decisor consegue expressar sua preferência ao comparar duas alternativas (a, b) quatro situações básicas de preferência podem ser manifestadas, como exposto na Figura 5. Essas quatro situações básicas de preferências deram origem para outras cinco situações, conhecidas por situações consolidadas de preferências (Figura 6).

Figura 5 - Situações básicas de preferências do decisor

SITUAÇÃO	DESCRIÇÃO	EXPRESSÃO	DESCRIÇÃO
Indiferença (I)	Decisor indiferente entre as duas alternativas, razões que justifiquem a equivalência entre os elementos.	aIb	Simétrica ($a \succ b$ ou $b \succ a$) e reflexiva ($a \succ a$)
Preferência Estrita (P)	Há razões claras e objetivas que justifiquem a preferência estrita e sem dúvidas de uma alternativa em relação a outra.	aPb	Assimétrica ($a \succ b$ não $b \succ a$) e irreflexiva (não $a \succ a$)
Preferência Fraca (Q)	O decisor não consegue se decidir se prefere uma alternativa em relação a outra ou se essas são indiferentes.	aQb	Assimétrica ($a \succ b$ não $b \succ a$) irreflexiva (não $a \succ a$)
Incomparabilidade (R ou NC)	Ausência de razões claras e positivas que justifiquem as situações anteriores.	aRb	Simétrica ($a \succ b$ ou $b \succ a$) e irreflexiva (não $a \succ a$)

Fonte: Adaptado de Gomes, Araya e Carignano (2004).

Figura 6 - Situações consolidadas de preferências do decisor

Situação	Descrição	Expressão	Condição, Relação Binária
Não-preferência (\sim)	Situações em que as alternativas são indiferentes ou incomparáveis para o decisor. Decisor incapaz de diferenciá-las.	$a \sim b$	$a \sim b =$ se e somente se aIb ou aRb
Preferência (no sentido amplo) (\succ)	Decisor incapaz de definir se há preferência estrita ou fraca entre duas alternativas.	$a \succ b$	$a \succ b =$ se e somente se aPb ou aQb
Presunção de preferência (J)	Quando o decisor tem uma preferência fraca por uma alternativa e que, no limite, ela pode chegar a indiferença	aJb	$aJb =$ se e somente se aQb ou aIb
K-preferência (K)	O decisor tem preferência estrita por uma das duas alternativas ou identifica uma incomparabilidade entre elas.	aKb	$aKb =$ se e somente se aPb ou aRb
Superação (S)	Inexistência de divisão significativa entre situações de preferência estrita, fraca e indiferença. Decisor incapaz de distingui-las.	aSb	$aSb =$ se e somente se aPb ou aQb ou aIb

Fonte: Adaptado de Gomes, Araya e Carignano (2004).

Após essa modelagem, um sistema de estruturas de preferências sobre o conjunto de alternativas pode ser estabelecido, gerando relações de ordenamentos do tipo: ordem completa; pré-ordem completa; quase ordem e ordem de intervalo; pré-ordem parcial; e pseudo-ordem (MOUSSEAU; SLOWINSKI; ZIELNIEWICZ, 2000). Uma estrutura de ordem completa ocorre quando a comparação entre um par de alternativas corresponde a noção intuitiva de classificação dessas alternativas sem possibilidade de empate por similaridade (admite preferência entre a e b, por exemplo) (VINCKE, 1992).

A pré-ordem completa corresponde a uma noção intuitiva de classificação das alternativas na qual é possível acontecer um empate por similaridade (preferência e indiferença, sem incomparabilidades) (ALMEIDA, 2011). A quase ordem e a ordem de intervalo consideram a possibilidade de que a relação de simetria não é perfeitamente transitiva em casos extremos (definidos pelo limite da indiferença q). Diferente da ordem de intervalo, a quase-ordem possui um limite de indiferença constante (VINCKE, 1992).

Já a pré-ordem parcial é considerada uma generalização da pré-ordem completa, permitindo que um conjunto de alternativas possa ser ordenado do “melhor” ao “pior” sem que haja ligação (apesar de manter intransitividade, permite incomparabilidade na classificação) (CAMPOS, 2011). E, por último, a pseudo-ordem, na qual são admitidos três tipos de relações binárias: preferência estrita (P), preferência fraca (Q) e indiferença (I). Essas relações podem ser delimitadas pelos limiares de preferência (p) em que é claro a preferência em uma relação binária, e um limiar de indiferença (q) na qual é clara a indiferença (VINCKE, 1992).

2.3 MÉTODOS MULTICRITÉRIO DA FAMÍLIA ELECTRE

Os métodos ELECTRE tem como abordagem a sobreclassificação, eles são fundamentados no conceito de concordância para medir as vantagens de uma alternativa em relação as outras (exemplo: alternativa “a” preferível à alternativa “b”), e discordância para medir a desvantagem da alternativa em relação as demais (exemplo: ausência de critérios com intensidade de preferência da alternativa “b” em

relação à “a”), assim, esses métodos permitem identificar quais ações do problema são dominadas (ZOPOUDINIS; DOUMPOS, 2002; ALMEIDA, 2012).

Essa família, bem como suas características são resumidas na Figura 7. A versão ELECTRE I, como definida na problemática de seleção, tem como objetivo selecionar a melhor ou melhores alternativas consideradas satisfatórias (ROY, 1968). Já o método ELECTRE II tem como entrada os dados obtidos dos resultados do ELECTRE I. Tendo como problemática a ordenação, busca-se ranquear as alternativas da melhor até à pior (ROY; BARTIER, 1973).

Figura 7 - Características dos métodos da Família ELECTRE

Família ELECTRE	Autor (ano)	Problemática	Tipo de Critério
ELECTRE I	Roy (1968)	Seleção (P. α)	Simple
ELECTRE II	Roy e Bartier (1973)	Ordenação (P. γ)	Simple
ELECTRE III	Roy (1978)	Ordenação (P. γ)	Pseudo
ELECTRE IV	Roy e Hugonnard (1982)	Ordenação (P. γ)	Pseudo
ELECTRE IS	Roy e Skalka (1985)	Seleção (P. α)	Pseudo
ELECTRE TRI	Yu Wei (1992)	Classificação (P. β)	Pseudo

Fonte: Adaptado de Gomes, Araya e Carignano (2004).

A versão ELECTRE III é utilizada quando há apenas um decisor. Seu objetivo é ordenar alternativas, porém sua utilização destina-se em casos que contém pseudo-critérios, em que é possível estabelecer condições de preferência e indiferença caso o decisor seja incapaz ou não quiser comparar duas alternativas (ROY, 1978). O ELECTRE IV também é utilizado para problemas de ordenação com pseudo-critérios. Porém, não se atribui pesos de importância para os critérios, o que pode ser bastante útil quando há dificuldade de se chegar em um consenso para ponderar os atributos (ROY; HUGONNARD, 1982).

No ELECTRE IS realiza-se escolha de alternativas, com estrutura de pseudo-critérios, há uma concordância plena de que uma alternativa “a”, mesmo que tenha um desempenho um pouco inferior em relação a uma alternativa “b” é no mínimo tão boa quanto à “b” (ROY; SKALKKA, 1985). E por fim, o ELECTRE TRI agrupa as alternativas em classes já pré-definidas, através da comparação de cada alternativa e os limites de cada classe (YU, 1992).

2.3.1 Electre I

Como visto anteriormente, o ELECTRE I busca solucionar problemáticas de Seleção ($P.\alpha$). Proposto por Roy (1968), esse método tem como objetivo reduzir o conjunto de alternativas A para o menor subconjunto A' possível, de forma que, qualquer ação que não estiver dentro desse subconjunto é sobreclassificada por pelo menos uma alternativa de A' . Portanto, pelos critérios que foram estabelecidos nessa análise, A' contém as alternativas mais satisfatórias como respostas ao problema que se pretende resolver (VINCKE, 1992).

Essas relações de sobreclassificação, são obtidas pela comparação par a par das alternativas. Por exemplo, se considerado que um conjunto de alternativas A é composto por (a, b, \dots, n) , para cada par ordenado (exemplo: (a, b)) de ações é associado um índice de concordância $C(a, b)$ e um de discordância $D(a, b)$ (CAMPOS, 2011). $C(a, b)$ assume valores que variam entre 0 e 1, podendo ser entendido como uma medida que indica o quanto a alternativa “ a ” supera a alternativa “ b ”, ou seja o quanto “ a ” sobreclassifica “ b ”. Esse índice é calculado conforme a Equação (1) (FIGUEIRA; MOUSSEAU; ROY, 2005).

$$C(a, b) = \frac{1}{P} \sum_{(j: g_j(a) \geq g_j(b))} P_j, \text{ no qual } P = \sum_{j=1}^n P_j \quad (1)$$

em que, $g_j(a)$ representa o desempenho da alternativa a no critério j ; $g_j(b)$ é o desempenho da alternativa b no critério j ; e P_j é o peso do critério j , o qual é atribuído de acordo com o grau de importância do critério, sendo que a soma dos pesos dos critérios deve ser igual a 1. Em contrapartida, o índice de discordância $D(a, b)$ mede a desvantagem da alternativa “ a ” em relação a “ b ”, ou seja, mede a força dos critérios em favor de “ b ”, calculando a maior distância entre as alternativas, no qual “ b ” é preferível à “ a ”. Seu resultado, também assume valores que variam entre 0 e 1, e seu cálculo é realizado pela Equação (2) (CAMPOS, 2011).

$$D(a, b) = \begin{cases} 0 & \text{se } g_j(a) \geq g_j(b) \forall j \\ \frac{1}{\delta} \max_j [g_j(b) - g_j(a)] & \end{cases} \quad (2)$$

em que:

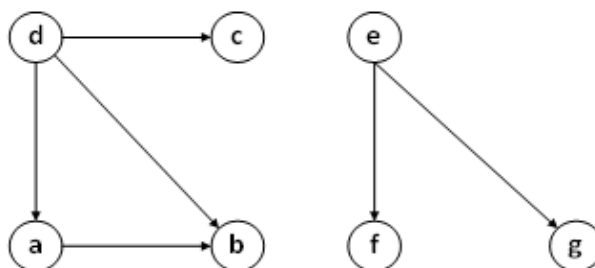
$$\delta = \max_{c,d,j} [g_j(c) - g_j(d)] \quad (3)$$

δ representa a máxima diferença para cada critério, $g_j(c)$ é a maior avaliação no critério j e $g_j(d)$ é a menor avaliação no critério j . Para que as relações de sobreclassificação sejam encontradas, o decisor deve estabelecer um valor para um limiar de concordância (c), relativamente grande, e um limiar de discordância (d), relativamente pequeno (CAMPOS, 2011). Conforme Vincke (1992), uma alternativa “ a ” sobreclassifica uma “ b ”, se:

$$aSb \text{ se e somente se } = \begin{cases} C(a,b) \geq c \\ D(a,b) \leq d \end{cases} \quad (4)$$

Sabendo-se que os índices de concordância e discordância variam entre 0 e 1, seus limiares também deve variar dentro desses parâmetros. Após essa construção do modelo, para obter o subconjunto de alternativas A' como o de melhor compromisso, conhecido também por *Kernel*, investiga-se suas relações de sobreclassificação da seguinte forma: é selecionado um subconjunto de alternativas A' que não foram sobreclassificadas por nenhuma outra alternativa, dentro do conjunto A (GOMES; ARAYA; CARIGNANO, 2004). Na forma de grafos, esse subconjunto de *kernel*, pode ser representado genericamente, como na Figura 8.

Figura 8 - Exemplo de grafos de relações de sobreclassificação



Fonte: Adaptado em Gomes, Araya e Carignano (2004).

Cada círculo corresponde a uma alternativa (a, b, c, d, e, f, g), e cada seta indica a relação de sobreclassificação (aSb), (dSa), (dSb), (dSc), (eSf), (eSg), sendo que para esse caso, o kernel é formado pelas alternativas “ d ” e “ e ”, que não são sobreclassificadas por nenhuma outra. É importante mencionar que a soma dos pesos

dos critérios deve ser igual a 1, esses pesos poderão ser estabelecidos pelo método AHP, na comparação par a par, realizadas pelos decisores, para melhor entendimento, ele está exposto na próxima seção.

2.3.1.1 Método AHP

Uma das etapas mais relevantes para o processo de tomada de decisão multicritério é a atribuição dos pesos aos critérios, pois as alternativas dispostas para solução dos problemas são avaliadas considerando os critérios e sua importância (ALMEIDA *et al.*, 2016). Dentre os métodos apresentados na literatura para tomada de decisão, o método AHP é muito utilizado também na atribuição de pesos para os critérios (REIS; LADEIRA; FERNANDES, 2013; KHANAM; SIDDIQUI; TALIB, 2015).

Desenvolvido por Saaty (1980), esse método tem como finalidade orientar para decisões mais objetivas. Ele se inicia pela decomposição de um problema em uma hierarquia de critérios, permitindo uma visão global do problema. Para a determinação dos pesos dos critérios, Saaty (1980) propôs a utilização de uma escala de importância entre os critérios, que são comparados para a par, como apresentada na Figura 9.

Figura 9 - Escala de Saaty

Escala	Avaliação	Inverso	Comentário
Igual Importância	1	1	Os dois critérios contribuem igualmente para os objetivos.
Importância Moderada	3	1/3	A experiência e o julgamento favorecem um critério levemente sobre o outro.
Mais Importante	5	1/5	A experiência e o julgamento favorecem um critério fortemente em relação a outro.
Muito Importante	7	1/7	Um critério fortemente favorecido em relação a outro e pode ser demonstrado na prática.
Importância Extrema	9	1/9	Um critério é favorecido em relação a outro com mais alto grau de certeza.
Valores Intermediários	2, 4, 6 e 8		Quando se procura condições de compromisso entre duas definições. É necessário acordo.

Fonte: Saaty (1980).

Em que 1 representa a “indiferença” ou igual importância de um critério em relação a outro, 9 significa a extrema importância de um critério em relação a outro e

entre os valores dessa os níveis intermediários de importância. O julgamento realizado pelos decisores é feito par a par com base nessa escala, e responde a duas perguntas: Qual critério é mais importante, e qual a intensidade dessa importância de um em relação ao outro?. Com base na escala da Figura 9, é construída uma matriz de comparação (Figura 10).

Figura 10 - Matriz comparativa de critérios

		Critério 1	Critério 2
		Critério 1	Critério 2
A =	Critério 1	1	Avaliação numérica
	Critério 2	1 / avaliação numérica	1
	Soma	(1+1/avaliação numérica)	(Avaliação numérica + 1)

Fonte: Adaptado de Saaty (1980).

Os valores da diagonal serão sempre 1, pois representam a importância do critério em relação a ele mesmo, a avaliação numérica corresponde ao julgamento realizado pelos decisores, com base na escala proposta por Saaty, e 1 dividido pela avaliação numérica representa o inverso das avaliações. Com isso podem ser definidos os pesos dos critérios, conforme as equações 1 e 2. Posteriormente, obtêm-se uma mesma matriz normalizada (Figura 11), dividindo cada elemento da Matriz da Figura 10 pela soma da coluna a que pertence.

Figura 11 - Matriz normalizada

		Critério 1	Critério 2
		Critério 1	Critério 2
AN =	Critério 1	1 / SOMA	Avaliação numérica / SOMA
	Critério 2	(1 / avaliação numérica) / SOMA	1 / SOMA

Fonte: Adaptado de Saaty (1980).

Tendo a Figura 11, é calculado a média aritmética de cada linha dessa matriz normalizada, e com os resultados dessas médias tem-se os pesos dos critérios e

forma-se o vetor de prioridade w . Para identificar se os pesos estão corretos e se os julgamentos realizados pelos decisores possui razão de consistência (RC), usa-se a Equação 5:

$$RC = IC/IR \quad (5)$$

na qual, IC representa o índice de consistência e o IR é índice randômico para matrizes quadradas de ordem n . Os valores do IR são tabelados de acordo com a ordem da matriz, como na Figura 12. Já o índice de consistência é calculado como na Equação 6, que tem $\lambda_{m\acute{a}x}$ representando o maior autovalor da matriz de julgamentos, este que por sua vez, é obtido a partir da Equação 7.

Figura 12 - Índices de consistência aleatória

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.59

Fonte: Saaty (1991).

$$IC = (\lambda_{m\acute{a}x} - n)/(n - 1) \quad (6)$$

$$\lambda_{m\acute{a}x} = \frac{1}{n} \sum_{i=1}^n [Aw]_i / w_i \quad (7)$$

Como exemplificado na Equação 7, cada componente do vetor Aw é dividido pela sua respectiva componente em w e é realizado uma média aritmética desses quocientes. É importante ressaltar que $Aw = \lambda_{m\acute{a}x} w$, sendo que $w = (w_1, w_2, \dots, w_n)$ é o autovetor principal da matriz A e $\lambda_{m\acute{a}x}$ é o autovalor máximo correspondente. Com a razão de consistência calculada, segundo Saaty (2000) quanto maior for esse resultado, maior será a inconsistência, sendo que a condição de consistência dos julgamentos aceitável é $RC \leq 0,10$.

3 MÉTODOS DE PREVISÃO UTILIZADOS NESSA PESQUISA

O presente capítulo discute modelos lineares e RNAs para previsão de demanda, temática chave para o desenvolvimento desse trabalho. A apresentação desses métodos está dividida em duas seções secundárias: a primeira apresenta os conceitos e componentes envolvidos na construção dos modelos lineares de previsão SEHW, AR e ARIMA.

A segunda apresenta conceitos iniciais de RNAs, tendo seções terciárias sobre a analogia entre neurônios biológicos e artificiais, as funções de ativação comumente utilizadas na literatura, a arquitetura das RNAs e seu processo de treinamento, rede MLP e seu processo de treinamento, rede RBF e seu processo de treinamento e por fim, apresenta-se a rede ELM e seu processo de treinamento.

3.1 MODELOS LINEARES DE PREVISÃO

Para prever o futuro a cerca de uma série temporal estudada é necessário captar e representar o comportamento dos dados ao longo do tempo, esse comportamento pode ser desvendado através do uso de um modelo linear de previsão. Segundo Box, Jenkins e Reinsel (2008), o elemento previsto em de uma série temporal é resultante da aplicação de um filtro linear sobre um ruído branco gaussiano.

Na literatura, há diferentes técnicas lineares já consagradas para previsão. Para esse estudo, serão utilizadas apenas as técnicas univariadas, as quais tem seus valores previstos explicados apenas pelos valores passados de uma mesma série temporal (TAYLOR, 2008). Mais especificamente, as técnicas a serem utilizadas nesse estudo são das famílias de SE e ARIMA de Box-Jenkins.

As técnicas de previsão da família SE possuem baixo custo de aplicação e grande rapidez na convergência de resultados e ajustes, motivos pelos quais são amplamente utilizadas na literatura (PELLEGRINI, 2000; DAVIS *et al.*, 2003). Basicamente, consistem na aplicação de uma média ponderada nos dados da série temporal, sendo que as observações mais antigas levam pesos menores em relação

as observações mais recentes da série (MAKRIDAKIS *et al.*, 1998; MORETTIN; TOLOI, 2006).

As técnicas que compõe essa família são: Suavização Exponencial Simples (SES), sendo utilizada para casos em que apenas há ciclicidade na série; Suavização Exponencial linear de Holt (SEH), a qual além da ciclicidade é capaz de suavizar por meio de uma constante a tendência; e Suavização Exponencial de Holt-Winters (SEHW), sendo capaz de suavizar ciclicidade, tendência e sazonalidade que pode compor uma série temporal (MAKRIDAKIS *et al.*, 1998; MORETTIN; TOLOI, 2006).

As técnicas de previsão da família Box & Jenkins (1976) também possuem baixo custo de aplicação e rapidez de convergência, sendo fundamentada em quatro passos: primeiro, identificar um modelo apropriado para a série analisada; posteriormente estimar os parâmetros do modelo; verificar a adequação do modelo; e por último prever valores futuros da série.

As técnicas que compõe essa família são: modelo AR; modelo de Médias Móveis (MA); ARMA; e ARIMA, sendo que, as três primeiras necessitam que a série temporal esteja estacionária para que possam serem aplicadas, enquanto a última não se faz necessária essa condição (BOX; JENKINS, 1976). Como já mencionado, as técnicas de previsão a serem utilizadas nesse estudo são SEHW, AR e ARIMA.

O uso desses modelos lineares pode ser visto como vantajosos em decorrência da sua simplicidade/facilidade de aplicação, baixo custo computacional, rápida convergência de resultados e boa precisão a depender da série temporal analisada (MAKRIDAKIS; WHEELRIGHT; HYNDMAN, 1998; DAVIS *et al.*, 2003; MORETTIN; TOLOI 2006).

Como desvantagens, destaca-se a dificuldade de determinar os parâmetros exigidos pelos modelos como constantes de suavização da família SE e os parâmetros (p,d,q) da família ARIMA (MORETTIN; TOLOI 2006). Para um melhor entendimento, os modelos a serem utilizados serão apresentados, com mais detalhes, nas próximas seções.

3.1.1 Suavização Exponencial de Holt-Winters (SEHW)

O modelo SEHW pode ser utilizado para casos em que a série temporal é composta por ciclicidade, tendência e sazonalidade. Para se ajustar aos dados a serem previstos, esse modelo é composto por constantes que suavizam o ciclo (α), tendência (β) e sazonalidade (γ) que podem compor uma série. (MAKRIDAKIS *et al.*, 1998). Essas constantes funcionam como parâmetros de aproximação e podem assumir valores que variam entre 0 e 1 (MENEHINI *et al.*, 2018).

Além desses parâmetros de aproximação, o modelo também é dividido entre: multiplicativo, que é quando a amplitude de variação da série aumenta ou reduz ao longo do tempo; e aditivo, em que se tem a amplitude da variação sazonal e constante ao longo da série (MORETTIN; TOLOI, 2006).

As equações que suavizam a ciclicidade (Equação 8), tendência (Equação 9) e sazonalidade (Equação 10), compõe a Equação 11 que descreve a previsão pelo modelo Multiplicativo. Já as Equações 12, 13 e 14 também suavizam esses parâmetros, mas servem para compor a Equação 15 que descreve a previsão pelo modelo aditivo (MORETTIN; TOLOI, 2006):

$$\text{Ciclo: } \hat{Z}_t = \alpha \left(\frac{Z_t}{f_{t-s}} \right) + (1 - \alpha)(\hat{Z}_t + \hat{H}_{t-1}), \quad 0 < \alpha < 1, \quad t = s + 1, \dots, N \quad (8)$$

$$\text{Tendência: } \hat{H}_t = \beta (\hat{Z}_t - \hat{Z}_{t-1}) + (1 - \beta)\hat{H}_{t-1}, \quad 0 < \beta < 1, \quad t = s + 1, \dots, N \quad (9)$$

$$\text{Sazonalidade: } \hat{G}_t = \gamma \left(\frac{Z_t}{\hat{Z}_t} \right) + (1 - \gamma)\hat{G}_{t-s}, \quad 0 < \gamma < 1, \quad e \quad t = s + 1, \dots, N \quad (10)$$

$$\text{Previsão Mult: } \hat{Z}_{t+1}(h-1) = (\hat{Z}_{t+1} + (h-1)\hat{H}_{t+1}) \hat{G}_{t+1+h-2s}, \quad h = s + 2, \dots, 2s + 1 \quad (11)$$

$$\text{Ciclo: } \hat{Z}_t = \alpha (Z_t - \hat{G}_{t-s}) + (1 - \alpha)(\hat{Z}_{t-1} + \hat{H}_{t-1}), \quad 0 < \alpha < 1, \quad t = s + 1, \dots, N \quad (12)$$

$$\text{Tendência: } \hat{H}_t = \beta (\hat{Z}_t - \hat{Z}_{t-1}) + (1 - \beta)\hat{H}_{t-1}, \quad 0 < \beta < 1, \quad t = s + 1, \dots, N \quad (13)$$

$$\text{Sazonalidade: } \hat{G}_t = \gamma (Z_t - \hat{Z}_t) + (1 - \gamma)\hat{G}_{t-s}, \quad 0 < \gamma < 1, \quad e \quad t = s + 1, \dots, N \quad (14)$$

$$\text{Previsão Adit: } \hat{Z}_{t+1}(h-1) = \hat{Z}_{t+1} + (h-1)\hat{H}_{t+1} + \hat{G}_{t+1+h-2s}, \quad h = s + 2, \dots, 2s + 1 \quad (15)$$

em que \hat{Z}_t , \hat{H}_t e \hat{G}_t representam as estimativas de nível, tendência e sazonalidade, α , β e γ são as respectivas constantes de suavização, s corresponde ao comprimento sazonal, e $\hat{Z}_{t+1}(h-1)$ é o valor previsto para h passos à frente.

3.1.2 Autorregressivo (AR)

Para uma série temporal que apresente sua variância e média estáveis ao longo do tempo o modelo preditivo AR é uma alternativa viável (HYNDMAN; ATHANASOPOULOS, 2018). Sua representação matemática pode ser dada pela Equação (16), em que ε_t corresponde a um ruído branco (MORETTIN; TOLOI, 2006).

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (16)$$

Esse modelo pressupõe que a previsão da série é dada pela combinação linear $y_{t-1} + \dots + y_{t-p}$ dos termos passados da série, multiplicados pelos φ_p coeficientes livres do modelo e adicionados de um ε_t erro intrínseco ao processo. Nesse modelo, os valores dos coeficientes livres podem ser encontrados de forma determinística aplicando as equações de Yule-Walker, garantindo o mínimo MSE (BOX *et al.*, 1994).

3.1.3 Autorregressivo Integrado de Médias Móveis (ARIMA)

O método ARIMA foi desenvolvido na década de 70 pelos autores Box e Jenkins. Este que é um modelo linear de previsão, resulta da combinação dos componentes de autoregressão (AR), diferenciação (I) e média móvel (MA) (OSHODI *et al.* 2017). Diferente de outros modelos consolidados como o AR e o MA que requerem a condição de estacionariedade, este por sua vez é uma alternativa a ser aplicado em séries que possuem tendência (BOX *et al.*, 1994).

As componentes combinadas para realizar as previsões, segundo Box *et al.*, (2016), tem as seguintes funções: a componente AR assume o papel de modelar a relação dos dados atuais com os dados passados, assumindo uma ordem (p); a componente I corresponde ao número de diferenciações de ordem (d) necessárias

para remover a tendência e tornar a série estacionária; e a componente MA assume que os valores previstos dependem dos erros de previsão de períodos anteriores de ordem (q). O modelo ARIMA (p, d, q) é representado pela Equação (17) (BOX *et al.*, 1994):

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (17)$$

φ_0 representa uma constante no modelo estimado, $\varphi_1, \dots, \varphi_p$ são ponderações para os valores passados de y_t , ε_t corresponde aos ruídos brancos da série e $\theta_1, \dots, \theta_q$ possibilitam escrever a série em função de choques passados. Importante mencionar que uma das formas de determinar as ordens (p) e (q) do modelo ARIMA é através das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) (BOX *et al.*, 2016).

3.2 REDES NEURAIS ARTIFICIAIS

Inspirado no sistema nervoso dos seres vivos, as RNAs são modelos computacionais com capacidade para adquirir e manter conhecimento através de experiências (SILVA; SPATTI; FLAUZINO, 2016). Definida também como um conjunto de unidades de processamento de informações, compostos por neurônios artificiais, que são interligados por conexões sinápticas associadas a pesos, essas RNAs têm capacidade de solucionar diferentes tipos de problemas (HAYKIN, 2008).

Características como capacidade de aprendizado, adaptação por experiência, habilidade de generalização, organização de dados, tolerância a falhas, armazenamento distribuído e facilidade de prototipagem, conferem as RNAs potencial para aplicabilidade em diversas áreas do conhecimento, entre as quais se enquadra a previsão de séries temporais (SILVA; SPATTI; FLAUZINO, 2016). Se considerado que problemas de previsão são tarefas de mapeamento não linear estático, RNAs como a MLP podem ser atraentes para esse tipo de resolução (SIQUEIRA, 2013).

Essa possibilidade de aplicação em previsão de séries temporais é atribuída a capacidade de generalização, aproximação universal e metodologias de treinamentos eficientes que essa rede possui (HAYKIN, 2008). Uma rede que pode

ser uma alternativa ao uso da MLP para essa mesma tarefa, é a rede RBF, que pode ter vantagens em relação ao seu uso por possuir um padrão de estrutura mais simples e processo de treinamento mais rápido (SILVA; SPATTI; FLAUZINO, 2016).

Semelhante a rede MLP, mas com processo de ajustes de conexões mais simples e menor custo computacional, também pode ser utilizado para essa tarefa a rede ELM. Em problemas de previsão com dependência de temporal entre os dados, essas redes recorrentes com memória inerente podem ser úteis (SIQUEIRA, 2013).

Importante destacar, que para esse trabalho, a comparação das RNAs em relação aos modelos lineares será realizada em decorrência das vantagens da não linearidade que esses modelos têm, além da adaptabilidade, tolerância a falhas e capacidade de generalizar resultados (HAYKIN, 2008).

Embora esses modelos apresentem também algumas desvantagens em relação aos modelos lineares, como por exemplo: treinamentos demorados, resultados desconcertantes quando se há pouco volume de dados e perda da capacidade de generalização de soluções em decorrência de super treinamento (*overfitting*) (HAYKIN, 2008; SILVA; SPATTI; FLAUZINO, 2016).

Se constatada a superioridade na acurácia dos resultados de previsão, as RNAs podem ser a melhor alternativa como ferramenta para auxiliar nas tomadas de decisões a respeito da série temporal analisada. Para melhor entendimento, o restante do capítulo trará alguns aspectos biológicos e artificiais dos neurônios, classificações das RNAs e processo de treinamento, e de forma mais detalhada também será apresentado as redes MLP, RBF e ELM.

3.2.1 Neurônios Biológicos Versus Artificiais

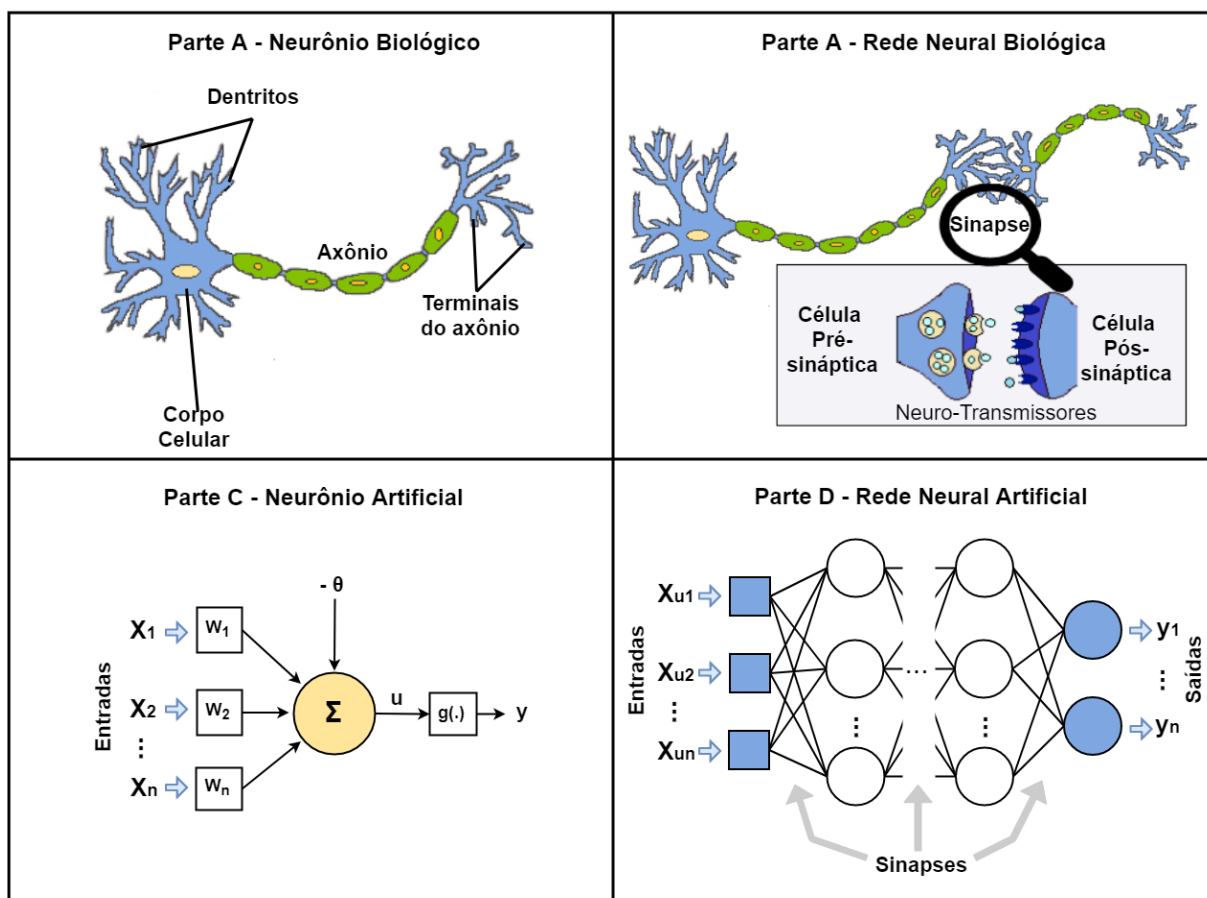
Os neurônios, também conhecidos como células nervosas, são unidades básicas para processamento de informações e estímulos do corpo humano, que tem como principal função a condução de impulsos nervosos (HAYKIN, 2008). Essas células são compostas de três partes principais, sendo: dendritos, corpo celular e axônio (Figura 13 – parte A). Os dendritos têm como função captar os estímulos (impulsos nervosos) vindos de outros neurônios e conduzi-los até o corpo celular.

Ali são processadas as informações e produzido um potencial de ativação que indica se o neurônio poderá disparar um impulso elétrico ao longo do seu axônio, que

ao receber esses impulsos, tem como função transmiti-los para outros neurônios através de suas terminações (BRAGA; CARVALHO; LUDERMIR, 2007). O canal de conexão entre a terminação axônica de um neurônio e o dendrito de outro é chamado de sinapse.

Deve-se destacar a inexistência de contato físico entre os neurônios nessa fase de transmissão sináptica, pois os elementos neurotransmissores liberados pelos axônios é que são responsáveis por ponderar a transmissão de informações para outros neurônios (SILVA; SPATTI; FLAUZINO, 2016). Embora as atividades realizadas pelos neurônios pareçam simples, quando estes estão trabalhando em conjunto, formam uma enorme rede de comunicação, chamada rede neural (Figura 13 – parte B).

Figura 13 - Funcionamento genérico do neurônio biológico e artificial e seus processos sinápticos quando conectados a outros neurônios



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Essa rede é responsável por todos os processamentos executados e gerenciados pelo cérebro humano (SILVA; SPATTI; FLAUZINO, 2016). De modo análogo ao neurônio biológico, McCulloch e Pitts (1943) propuseram um modelo artificial (Figura 13 – parte C), com um circuito binário simples que combina várias entradas e apenas um sinal de saída. Podendo ser descrito como na Equação (18) e (19).

$$u = \sum_{i=1}^n w_i \cdot x_i - \theta \quad (18)$$

$$y = g(u) \quad (19)$$

Matematicamente, os impulsos elétricos captados pelos dendritos no neurônio biológico, são representados por n sinais de entrada advindos do meio externo (x_1, x_2, \dots, x_n) (CAVALCANTE, 2018). Já as ponderações exercidas pelas junções sinápticas dos neurônios biológicos são representadas pelo conjunto de pesos sinápticos (w_1, w_2, \dots, w_n) que tem como função ponderar as variáveis de entrada, quantificando sua relevância em relação a funcionalidade do neurônio (ANOCHI, 2015).

A saída do corpo celular é representada pelo potencial de ativação (u), que é o valor produzido pelo combinador linear (\sum) subtraído pelo limiar de ativação representado por (θ), este que tem como função especificar o patamar apropriado para que a soma das entradas ponderadas possam gerar um valor de disparo em direção a saída y (axônio) do neurônio (CAVALCANTE, 2018).

Posteriormente, as informações de (u) passarão pela função de ativação (g(.)) que irá limitar a saída do neurônio dentro de um intervalo de valores assumidos (HAYKIN, 2008). Os resultados das saídas (y), são transformações das entradas recebidas, e quando se propagam para outros neurônios interligados, formam o que se conhece por RNAs (Figura 13 – parte D) (SIQUEIRA, 2013).

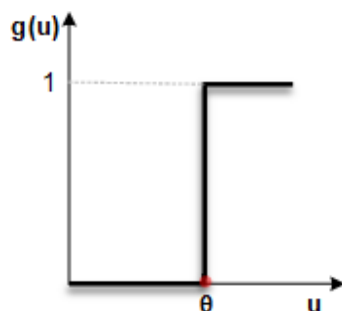
3.2.2 Funções de Ativação

No modelo de McCulloch e Pitts (1943) a função de ativação é dada pela função de limiar, em que um neurônio só é ativo (saída igual a 1) se a soma ponderada das entradas superar um certo limiar de ativação, caso contrário ele permanecerá

inativo (saída igual a 0) (Equação 20). A função $g(u)$ é representada graficamente na Figura 14.

$$g(u) = \begin{cases} 1 & \text{se } u \geq \theta \\ 0 & \text{caso contrário} \end{cases} \quad (20)$$

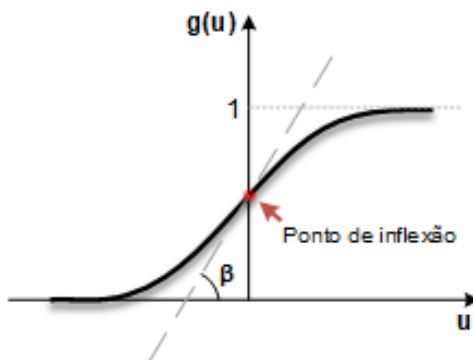
Figura 14 - Função de Ativação de McCulloch e Pitts



Fonte: Adaptado de McCulloch e Pitts (1943).

A partir desse modelo apresentado, vários outros foram derivados, que permitem a produção de uma saída qualquer, com diferentes funções de ativação. Entre as existentes, as mais utilizadas são do tipo logística e tangente hiperbólica, ambas pertencentes a família de funções chamadas sigmoideal (ANOCHI, 2015). A função logística é representada matematicamente pela Equação (21), sendo que o resultado de saída alcançado por essa aplicação sempre assumirá valores reais entre 0 e 1, graficamente pode ser visualizado na Figura (15).

Figura 15 - Função de Ativação Logística



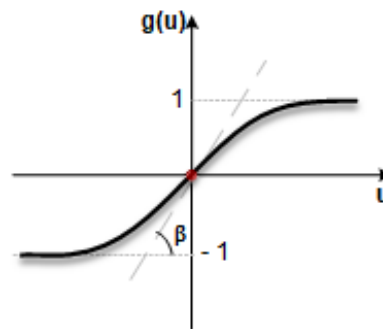
Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

$$g(u) = \frac{1}{1+e^{-\beta \cdot u}} \quad (21)$$

O parâmetro β é uma constante real associada ao nível de inclinação da função logística sigmoide referente ao seu ponto de inclinação e u é o valor da função de ativação do neurônio. Já a função tangente hiperbólica é do tipo bipolar, a qual permite valores positivos e negativos, que varia entre -1 e 1, representada pela Equação (22), podendo ser visualizada graficamente na Figura (16).

$$g(u) = \frac{1 - e^{-\beta \cdot u}}{1 + e^{-\beta \cdot u}} \quad (22)$$

Figura 16 - Função de Ativação Tangente Hiperbólica



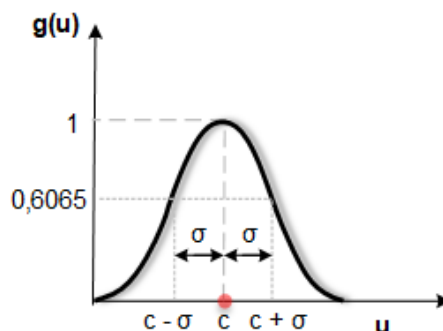
Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

O parâmetro β também está associado com a inclinação em seu ponto de inflexão e u é o valor da função de ativação do neurônio. Além dessas, existem as funções gaussiana e linear, que também são totalmente diferenciáveis, cujas derivadas de primeira ordem existem para todos os pontos de seu domínio de definição. Ambas comumente empregadas em RNAs como a RBF (LAZZARIN, 2019).

Quando utilizada a função gaussiana, o neurônio terá saída iguais para os valores do potencial de ativação u colocados na mesma distância do centro (média), sendo que a curva é simétrica a esse centro (SILVA; SPATTI; FLAUZINO, 2016). Essa função é expressa pela Equação (23), tendo sua representação gráfica ilustrada na Figura 17.

$$g(u) = -e^{-\frac{(u-c)^2}{2\sigma^2}} \quad (23)$$

Figura 17 - Função de Ativação Gaussiana



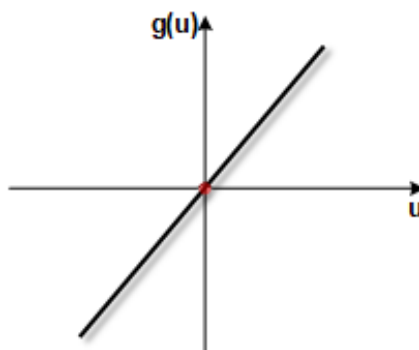
Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

O parâmetro c representa o parâmetro que define o centro da função gaussiana e σ é o desvio padrão associado a função, denotando o quão dispersa é a curva em relação ao seu centro. Pela Figura 17, observa-se que o valor do desvio padrão σ está diretamente associado com o ponto de inflexão da função gaussiana, com σ^2 indicando a sua variância (SILVA; SPATTI; FLAUZINO, 2016).

Por último, tem-se a função linear ou função identidade, sendo que esta apresenta resultados de saídas do neurônio idênticos aos valores do seu potencial de ativação (u). Sua representação matemática é definida pela Equação 24 e seu gráfico é ilustrado na Figura 18.

$$g(u) = u \quad (24)$$

Figura 18 - Função de Ativação Linear



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Uma das aplicações dessa função está na utilização de RNAs para a atividade de mapeamento do comportamento das variáveis de entrada e saída de um processo particular (aproximador universal de funções) (SILVA; SPATTI; FLAUZINO, 2016).

3.2.3 Arquiteturas das Redes Neurais Artificiais e Processos de Treinamento

A forma como os neurônios estão dispostos e arranjados uns em relação aos outros é conhecido por arquitetura de uma RNA (SILVA; SPATTI; FLAUZINO, 2016). Uma RNA pode ser dividida em três partes, sendo: camada de entrada, a qual é responsável pelo recebimento das informações de entrada; camadas intermediárias, compostas de neurônios, responsáveis por realizar quase todo o processamento interno referente ao funcionamento da rede; e camada de saída, também constituídas de neurônios, que são responsáveis pela produção e apresentação dos resultados finais (HAYKIN, 2008).

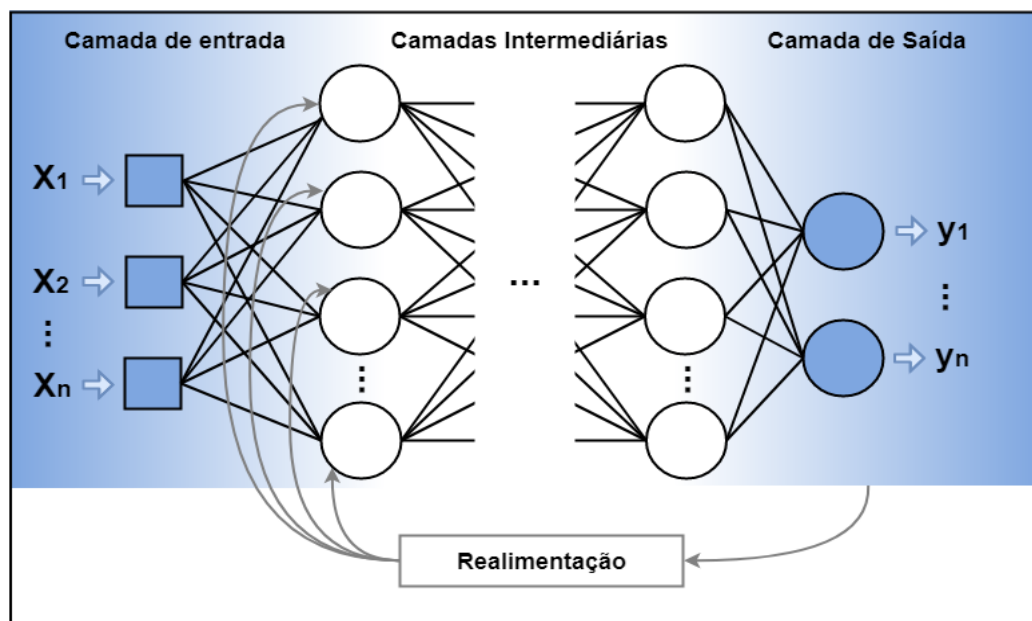
Quando uma rede possui apenas a camada de entrada e saída, ela é chamada de rede de camada simples, mas quando ela é constituída também de uma ou mais camadas de neurônios escondidas (camadas intermediárias), ela é conhecida por rede de camadas múltiplas (SILVA; SPATTI; FLAUZINO, 2016). Além disso, a arquitetura de uma RNA também é classificada quanto ao seu padrão de interconexão, podendo ser: *feedforward* ou recorrente.

O fluxo de informações de uma rede *feedforward* segue sempre adiante (unidirecional), ou seja, da camada de entrada para a camada de saída até que se estabeleça a resposta da rede (SIQUEIRA, 2013). Já as redes recorrentes possuem conexões de realimentação entre os neurônios, as informações de uma camada são transmitidas para as suas anteriores ou entre os neurônios dessa mesma camada, sendo processadas de forma dinâmica, criando uma espécie de memória interna (BRAGA; CARVALHO; LUDERMIR, 2007). Exemplo das possíveis arquiteturas descritas aqui nessa seção, podem ser visualizadas de forma genérica na Figura 19.

Além de diferentes arquiteturas de redes, também é possível que estas assumam diferentes composições estruturais. Dessa forma, uma mesma arquitetura pode ter neurônios com diferentes funções de ativação, ou ainda camadas com diferentes números de neurônios (SILVA; SPATTI; FLAUZINO, 2016). A definição da arquitetura de uma rede, restringirá o tipo de problema que pode ser resolvido por ela.

Por exemplo, a rede *Perceptron* proposta por Rosenblatt em 1958, que só é capaz de resolver problemas linearmente separáveis (CAVALCANTE, 2018).

Figura 19 - Exemplo genérico da arquitetura de uma rede neural com múltiplas camadas e realimentação



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Quando determinada a arquitetura de uma rede para um problema específico, o próximo passo é realizar o treinamento dela, que consiste na aplicação de um conjunto de passos ordenados com o objetivo de ajustar os pesos e os limiares dos neurônios. Esse processo de ajustes é que confere aprendizado a rede, que é uma das suas características mais relevantes (SILVA; SPATTI; FLAUZINO, 2016). Quando a rede é treinada e aprende o relacionamento entre as entradas e saídas, ela se torna hábil para generalizar soluções (HAYKIN, 2008).

A partir disso, ela é capaz de gerar resultados próximos aos valores desejados, mesmo que novos sinais de entradas sejam inseridos na rede (SIQUEIRA, 2013). Esse processo de aprendizado pode ocorrer através de treinamentos supervisionados, não-supervisionados ou por reforço. O treinamento supervisionado ocorre como se tivesse um “professor” que ensina para a rede a resposta correta para cada amostra de dados que é apresentada em sua entrada, ou seja, sua funcionalidade consiste em ter disponível dados de saídas desejadas (sinal de referência) para cada amostra de dados de sinais de entrada (HAYKIN, 2008).

Através desse sinal de referência é que se alteram/ajustam os pesos sinápticos até encontrar um resultado que esteja próximo do desejado, quando encontrado, a rede é considerada treinada e autônoma o suficiente para continuar a aprender com o ambiente por si própria (HAYKIN, 2008). Já o treinamento não-supervisionado tem disponível apenas os padrões de entrada, a rede consegue se organizar (ajustes auto organizados) em relação aos padrões existentes do conjunto de amostras, até obter padrões específicos de resposta (SILVA; SPATTI; FLAUZINO, 2016).

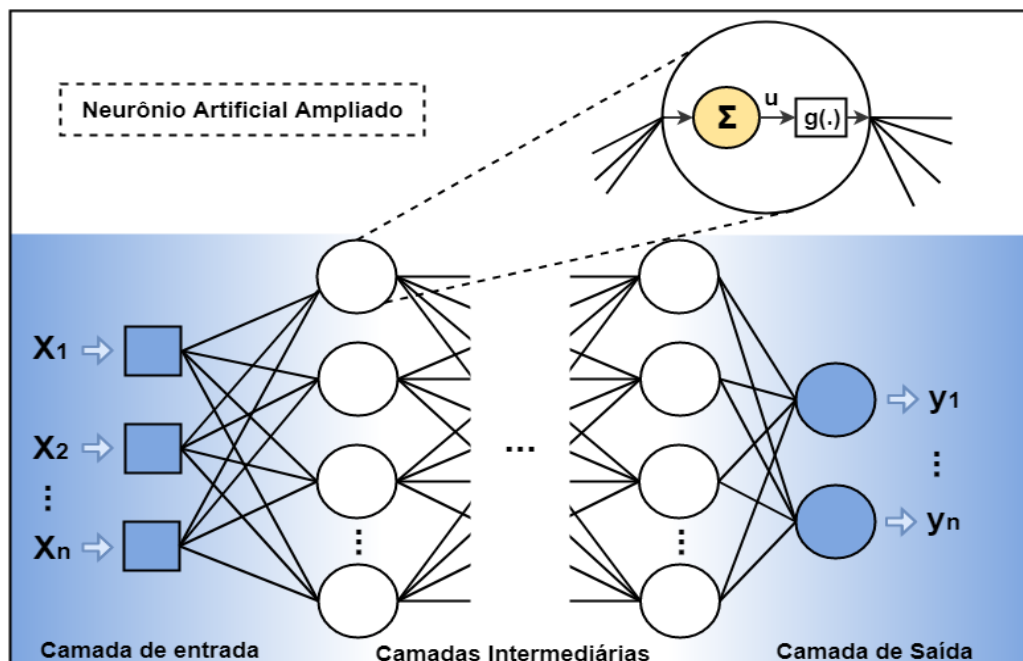
Por último, o treinamento por reforço pode ser considerado uma variação do supervisionado, tendo disponível os sinais de referência para cada amostra de dados de sinais de entrada, as respostas fornecidas pela rede são avaliadas por um indicador de desempenho (sinal de reforço), este que por sua vez direciona os ajustes dos pesos. Esse processo de treinamento é realizado por tentativa e erro (HAYKIN, 2008).

3.2.3.1 Perceptron de múltiplas camadas (MLP)

Constituídas de uma ou mais camadas de neurônios escondidas (camadas intermediárias) entre a camada de entrada e saída, a rede MLP é considerada uma das arquiteturas mais versáteis em relação ao seu potencial de aplicabilidade (HAYKIN, 2008). Essa característica conferida a rede é consequência da sua flexibilidade de aplicação em diferentes áreas do conhecimento como em otimização de sistemas, identificação e controles de processos, aproximação universal de funções, reconhecimento de padrões e previsão de séries temporais (SILVA; SPATTI; FLAUZINO, 2016). Na Figura 20 é ilustrado de forma genérica a estrutura de uma rede MLP.

Observa-se (Figura 20) que os estímulos ou sinais de entrada (X_1, X_2, \dots, X_n) são apresentados a rede na camada de entrada, os quais se propagam para as camadas intermediárias, que por sua vez extraem as informações referentes aos comportamentos desses estímulos e os codificam através de pesos sinápticos (W_1, W_2, \dots, W_n) e limiares (θ), e por último, os neurônios da camada de saída recebem os estímulos da última camada oculta, e produz um padrão de resposta que será a saída da rede (y) (CASTRO, 2006; SILVA; SPATTI; FLAUZINO, 2016).

Figura 20 - Modelo de rede Perceptron de Múltiplas Camadas (MLP)



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Comparado ao que foi visto na seção 3.2.3 deste estudo, percebe-se que nessa rede o fluxo de informações se propaga sempre em uma única direção, não há comunicação entre neurônios de uma mesma camada ou em direção a camadas anteriores, ou seja, não se caracteriza ao padrão de interconexão conhecido por realimentação, mas sim, ao *feedforward* (CASTRO, 2006).

É importante mencionar, que a configuração topológica da rede (número de camadas intermediárias e seus respectivos neurônios) dependerá do tipo de problema a ser tratado, não há uma regra que determine uma quantidade ideal (ANOCHI, 2015). Porém, alguns fatores como a disposição espacial dos sinais de entrada e os valores sinápticos que inicialmente são atribuídos no processo, auxiliam na definição de sua topologia (SILVA; SPATTI; FLAUZINO, 2016).

3.2.3.1.1 Treinamento da rede MLP

Como mencionado na seção 3.2.3, os ajustes dos pesos sinápticos e do limiar de cada neurônio é definido através do processo de treinamento, uma rede MLP aprenderá por treinamento supervisionado, a qual utiliza o algoritmo de aprendizado de retropropagação de erro, também conhecido como *backpropagation*

(AYYANATHAN; KANNAMMAL, 2014; SILVA; SPATTI; FLAUZINO, 2016). O processo supervisionado utilizando esse algoritmo é realizado em duas fases.

A primeira, conhecida por *forward*, é iniciada com a introdução e propagação dos dados a rede a fim de obter uma resposta de saída, com o valor dos pesos fixos. Na sequência, essa resposta é comparada com os valores desejados e uma medida de erro é obtida (HAYKIN, 2008). Uma das medidas que mais se utiliza na literatura para medição do desempenho de resposta da rede é MSE (SIQUEIRA, 2013), sua Equação 28 está disposta no Capítulo 4. Para cada iteração realizada na fase de treinamento da rede, deseja-se minimizar esse erro.

Tendo encontrado o erro pelo MSE na primeira iteração, a segunda fase conhecida como *backward* é então empregada, a qual executa os ajustes dos valores de pesos de acordo com a regra de correção de erro assumida, fazendo sucessivas iterações até que a saída dada pela rede, se aproxime do valor desejado (HAYKIN, 2008). Um algoritmo de primeira ordem, como o *backpropagation*, atualizará seus pesos sinápticos e os bias de cada neurônio utilizando as Equações (25) e (26) (SIQUEIRA, 2009).

$$w_{ij}^m(t+1) = w_{ij}^m(t) - \alpha \frac{\partial MSE(t)}{\partial w_{ij}^m(t)} \quad (25)$$

$$b_j^m(t+1) = b_j^m(t) - \alpha \frac{\partial MSE(t)}{\partial b_j^m(t)} \quad (26)$$

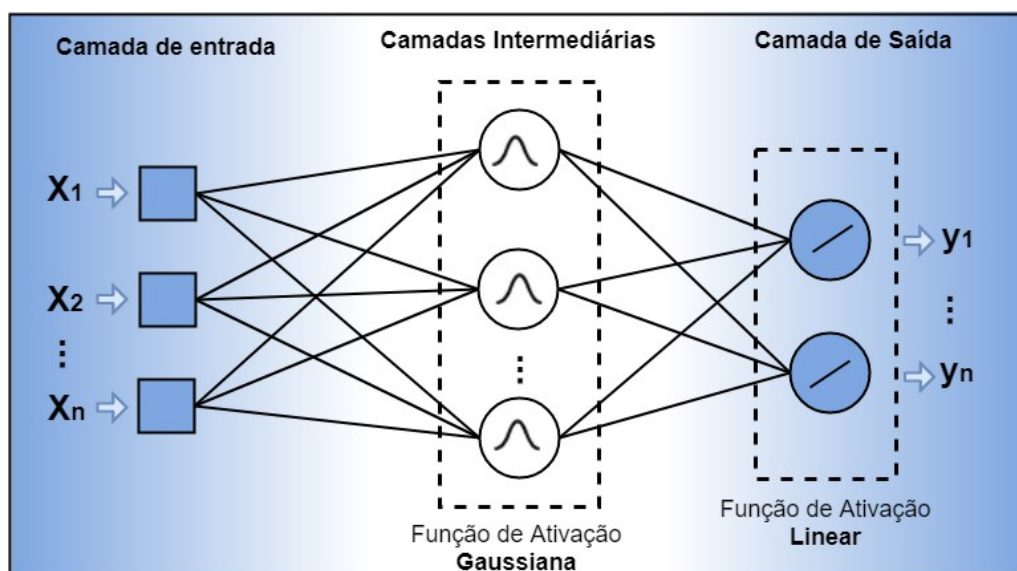
$w_{ij}^m(t)$ é igual ao peso sináptico da entrada i do neurônio j da camada m na iteração t , e $b_j^m(t)$ é o bias do neurônio j da camada m na iteração t . A derivada da função ∂MSE para cada camada é obtida pela regra da cadeia, a qual retropropaga o erro encontrado em direção a camada de entrada, e α representa o passo de otimização do algoritmo.

3.2.3.2 Redes neurais com função de base radial (RBF)

Devido a arquitetura de múltiplas camadas e padrão de interconexão *feedforward*, as redes RBF podem ser uma alternativa ao uso das redes MLPs, uma vez que possuem padrão de estrutura mais simplificado e processo de treinamento

mais ágil (HAYKIN, 2008). A Figura 21 expõe a estrutura típica de uma RBF, a qual apresenta uma única camada intermediária, diferindo-se das redes MLPs, que podem apresentar diversas camadas intermediárias (SILVA; SPATTI; FLAUZINO, 2016).

Figura 21 - Modelo de rede neural com função de base radial (RBF)



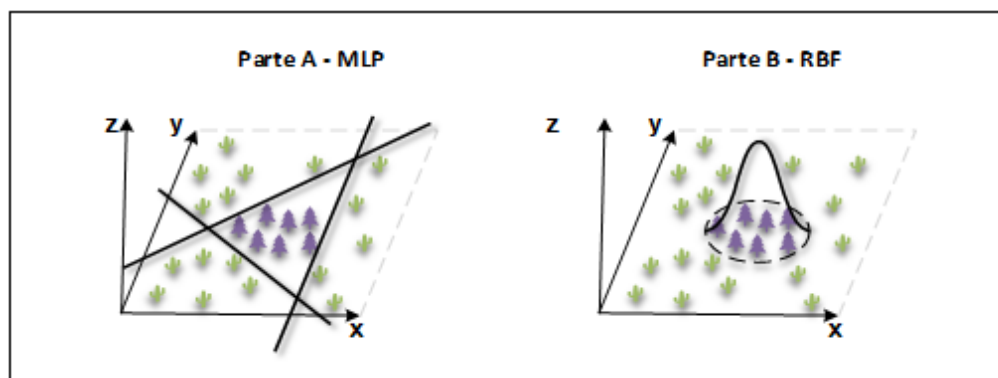
Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Nessa camada intermediária, a definição da quantidade de neurônios necessários para a resolução de um problema, poderá ser definida através de técnicas de validação cruzada (LAZZARIN, 2019). Todos os neurônios presentes nessa camada oculta, sempre serão ativados por funções de base radial, das quais comumente emprega-se a função gaussiana. Já os neurônios da camada de saída serão ativados por função linear (BELOTTI, 2019).

Uma comparação do funcionamento da rede RBF em relação a MLP, baseada no trabalho de Silva, Spatti e Flauzino (2016), pode ser visualizada na Figura 22, na qual expõe um exemplo de classificação de padrões constituído de duas entradas X_1 e X_2 .

Nota-se (Figura 22) que a rede MLP realizou essa tarefa delimitando as classes por hiperplanos, enquanto a rede RBF com funções de ativação gaussiana, realizou essa tarefa com um campo receptivo hipersférico. Nesse exemplo, a fronteira de separabilidade da rede MLP é formada por três retas, indicando a existência de três neurônios na camada intermediária, ao passo que a rede RBF agrupou os padrões de uma mesma classe em um único campo receptivo radial, indicando a existência de apenas um neurônio na camada intermediária.

Figura 22 - Fronteira limite com hiperplano e hipersférico



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

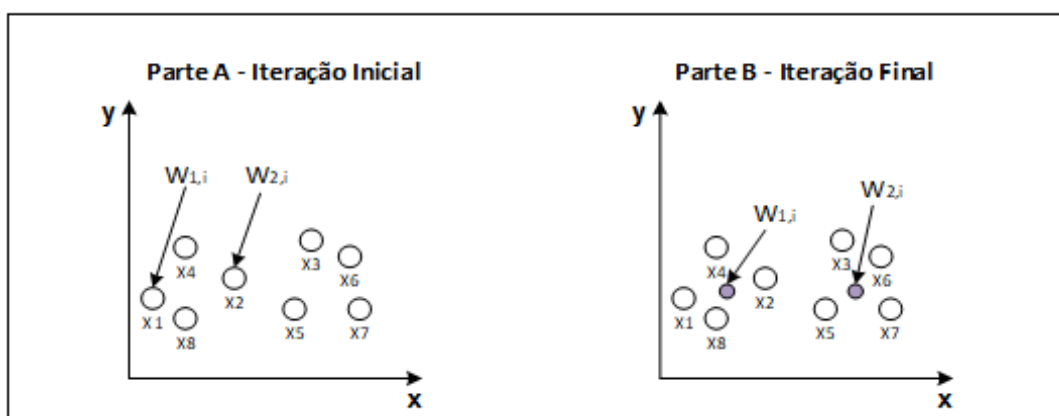
É importante mencionar que o campo receptivo hipersférico trabalha em todo domínio real, entretanto, a separação de amostras de uma mesma classe será delimitada através da variância dos dados, que indica o quão disperso está o potencial de ativação em relação ao centro da função gaussiana (SILVA; SPATTI; FLAUZINO, 2016).

3.2.3.2.1 Treinamento da rede RBF

O treinamento das redes RBFs é realizado em duas etapas. A primeira, calcula os pesos dos neurônios da camada intermediária através de um método de aprendizagem não supervisionado, que está diretamente relacionado com funções de base radial (HAYKIN, 2008). Se considerado que cada neurônio da camada intermediária é representado por uma função gaussiana (Equação 23), que tem seu centro definido pelos pesos sinápticos dos neurônios, o objetivo dessa etapa é posicionar os centros das gaussianas da forma mais adequada possível (SILVA; SPATTI; FLAUZINO, 2016).

Para facilitar o entendimento a respeito dessa primeira etapa de treinamento, considere um problema cujo objetivo seja classificar as amostras $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ com o uso de uma RBF constituída de dois neurônios na camada oculta. A disposição espacial dessas amostras é exemplificada na Figura 23. Inicialmente, os valores das entradas x_1 e x_2 são atribuídos aos pesos sinápticos dos neurônios n_1 e n_2 (Figura 23 – Parte A).

Figura 23 - Processo de ajuste utilizando o método k-means



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Se a camada oculta dessa rede fosse constituída de 4 neurônios, por exemplo, os pesos sinápticos de n_3 e n_4 também receberiam os valores das entradas x_3 e x_4 , e assim sucessivamente (BELOTTI, 2019). Dando sequência, as distâncias de cada amostra em relação aos pesos sinápticos dos neurônios (distância euclidiana) são calculadas e agrupadas de acordo com sua proximidade (SILVA; SPATTI; FLAUZINO, 2016).

Nesse exemplo, os conjuntos de amostras são formados por: conjunto 1 = (x_1, x_4, x_8) e conjunto 2 = $(x_2, x_3, x_5, x_6, x_7)$ (Figura 23 – Parte A). Tendo todas as entradas inseridas em seus respectivos conjuntos iniciais, os centros das gaussianas dos neurônios são atualizados para corresponder ao centro de todas as entradas contidas em cada conjunto. Após isso, os conjuntos são esvaziados e as entradas são novamente agrupadas em conjuntos mais próximos aos novos centros.

Esse processo é repetido, até que não haja mais mudanças de amostras para outros conjuntos (BELOTTI, 2019). No exemplo, os conjuntos finais das amostras são formados por: conjunto 1 = (x_1, x_2, x_4, x_8) e conjunto 2 = (x_3, x_5, x_6, x_7) (Figura 23 – Parte B). Após o cálculo desses centros das gaussianas, calcula-se suas variâncias, utilizando a Equação 27 (BELOTTI, 2019).

$$\sigma_j^2 = \frac{1}{|\Omega^j|} \sum_{x^k \in \Omega^j} \sum_{i=1}^n (x_i^k - w_{ji}^1)^2 \quad (27)$$

em que σ_j^2 representa a variância da base da função gaussiana do neurônio j , $|\Omega^j|$ é o tamanho do conjunto do neurônio j , x^k é a amostra k do conjunto Ω , x_i^k é a entrada i da amostra k , e w_{ji}^1 é o peso do neurônio j para a camada i . A saída de cada neurônio j da camada intermediária é expressa na Equação 28 (LAZZARIN, 2019).

$$g_j^{(1)}(u_j^1) = g_j^{(1)}(x) = e^{-\frac{\sum_{i=1}^n (x_i - w_{ji}^1)^2}{2\sigma_j^2}}, \quad j = 1, \dots, n_1 \quad (28)$$

Após o término do treinamento da camada intermediária, aplica-se a segunda etapa de treinamento, que é o ajuste dos pesos da camada de saída. Essa etapa ocorre como na rede MLP, podendo também ser utilizado o algoritmo de retropropagação do erro, o qual já foi descrito na seção 3.2.3.1.1 (SILVA; SPATTI; FLAUZINO, 2016).

Aqui, foi exemplificado um problema de classificação devido a sua simplicidade de entendimento. Não diferente ocorre o processo de treinamento da RBF para resolução de problemáticas de previsões, que utilizam uma certa quantidade de neurônios para mapear (e não classificar) os dados da série temporal (SILVA; SPATTI; FLAUZINO, 2016).

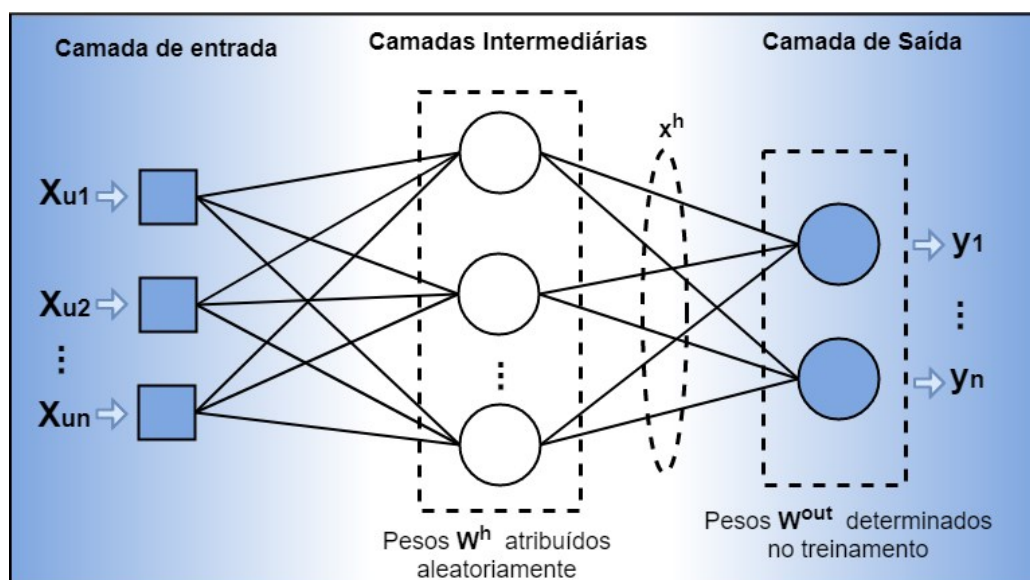
3.2.3.3 Máquinas de aprendizado extremo (ELM)

A rede ELM foi proposta por Huang, Zhu e Siew, (2004) com o objetivo de reduzir gargalos no processo de treinamento e aumentar a eficiência de aprendizagem das RNAs até então conhecidas. Para isso, os autores propuseram uma rede com arquitetura semelhante a MLP, com padrão de interconexão *feedforward*, mas com apenas uma camada intermediária, que se diferiu pelo seu processo de treinamento (LAZZARIN, 2019). De forma genérica, essa arquitetura é exemplificada na Figura 24.

Percebe-se (Figura 24) que os pesos dessa camada intermediária são determinados de maneira aleatória e permanecem sem ajuste, enquanto apenas os pesos da camada de saída são atualizados durante o processo de treinamento. Características estas que conferem baixo custo computacional no processo de

treinamento, e grande agilidade no processo de convergência dos resultados, se comparado as outras arquiteturas até então conhecidas (HUANG; ZHU; SIEW, 2004).

Figura 24 - Modelo de rede Máquinas de Aprendizado Extremo (ELM)



Fonte: Adaptado de Silva, Spatti e Flauzino (2016).

Na camada de saída, os pesos podem ser determinados através de um método analítico, que consiste na resolução de um problema de regressão linear (HUANG; ZHU; SIEW, 2006). Dessa forma, isenta-se da necessidade de calcular derivadas, retropropagação de sinais de erro ou utilização de algoritmos iterativos (SIQUEIRA *et al.*, 2012).

3.2.3.3.1 Treinamento da rede ELM

Como comentado anteriormente, o ajuste dos pesos durante o processo de treinamento é realizado apenas para a camada de saída, enquanto os pesos e bias da camada intermediária são atribuídos aleatoriamente com valores que variam entre 0 e 1, e permanecem sem alteração (BELOTTI, 2019). As saídas da camada intermediária são utilizadas como entradas para a camada de saída, e seu sinal x_t^h é dado pela Equação 29 (SILVA; SPATTI; FLAUZINO, 2016).

$$\mathbf{x}_t^h = f^h(\mathbf{W}^h \mathbf{X} \mathbf{u}_t + \mathbf{b}) \quad (29)$$

em que: $f^h(\cdot)$ é a função de ativação dos neurônios da camada intermediária; $\mathbf{X}\mathbf{u}_t = [Xu_1, Xu_2, \dots, Xu_{t-k+1}]^T$ é o vetor das entradas que tem seus componentes passados à camada intermediária; $\mathbf{W}^h \in \mathfrak{R}^{N_h \times K}$, são pesos gerados de forma aleatória; e b é o bias de cada neurônio. Após encontradas as saídas da camada intermediária, estas são multiplicadas pela matriz de pesos sinápticos da camada de saída \mathbf{W}^{out} , para que o valor de saída da rede seja produzido (Equação 30) (SILVA; SPATTI; FLAUZINO, 2016).

$$y_t = \mathbf{W}^{\text{out}} \mathbf{x}_t^h \quad (30)$$

Dessa forma, o treinamento para a rede ELM, consistirá em encontrar a matriz de pesos (\mathbf{W}^{out}) com o menor erro para o vetor d , que contém todas as saídas desejadas. O operador pseudo-inversa de *Moore-Penrose* (Equação 31) é uma opção interessante para resolver essa tarefa de treinamento (SIQUEIRA *et al.*, 2012).

$$\mathbf{W}^{\text{out}} = (\mathbf{x}_h^T \mathbf{x}_h)^{-1} \mathbf{x}_h^T \mathbf{d} \quad (31)$$

em que: $\mathbf{x}^h \in \mathfrak{R}^{T_s \times N_k}$ representa a matriz que contém a saída da camada intermediária, T_s é o número de amostras do treinamento e $(\mathbf{x}_h^T \mathbf{x}_h)^{-1} \mathbf{x}_h^T$ é a pseudo-inversa de \mathbf{x}_h . Embora essa solução analítica garanta um treinamento computacionalmente eficiente para a rede, o seu desempenho e capacidade de generalização podem ainda ser melhorado, utilizando o coeficiente de regularização C , proposto por Huang *et al.* (2012).

$$\mathbf{W}^{\text{out}} = \left(\frac{1}{C} + \mathbf{x}_h^T \mathbf{x}_h \right)^{-1} \mathbf{x}_h^T \mathbf{d} \quad (32)$$

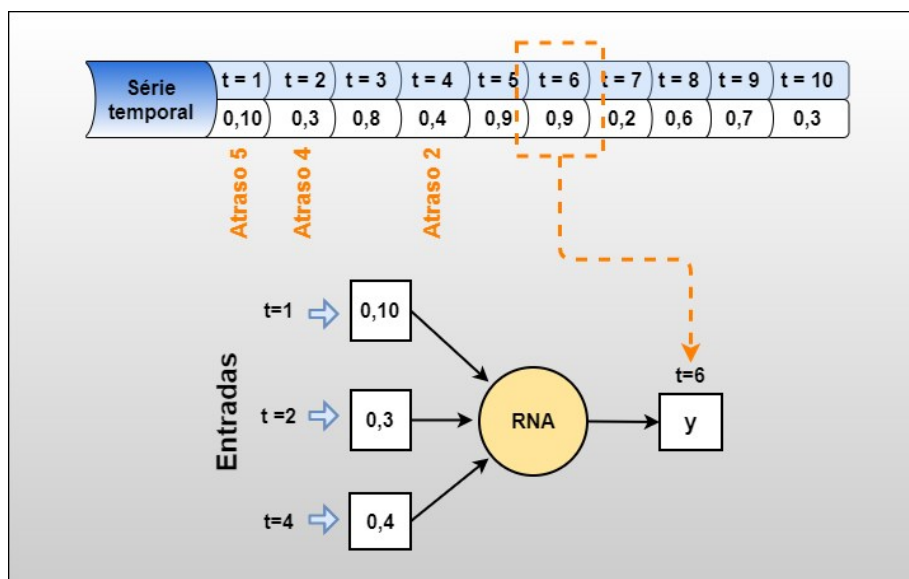
Para a determinação do valor de C , os autores indicam testar todos os valores possíveis para C e escolher aquele com o menor erro quando comparado ao conjunto de validação, inédito a rede. Existem 52 possíveis valores de vetor $\lambda = \{-26, -25, \dots, 24, 25\}$ de forma que $C = 2^\lambda$ (LAZZARIN, 2019).

3.2.4 Método Wrapper Para Seleção de Entradas

Para que as RNAs sejam utilizadas para a tarefa de previsão é necessário inicialmente selecionar as entradas que irão compor o modelo. Se tratando de análise de séries temporais, a saída das RNAs (previsão da demanda de energia, por exemplo) será dada com base na dependência que esse dado a ser previsto tem com os próprios dados passados do consumo de energia observados em relação ao tempo (SILVA; SPATTI; FLAUZINO, 2016).

Para auxiliar no entendimento do que são consideradas entradas (atrasos) de uma série temporal, tem-se a Figura 25. Sendo que, para prever o elemento $t = 6$ foram utilizados como entradas os elementos $t = 1$, $t = 2$, e $t = 4$. Para esse exemplo, pode-se dizer que o modelo considerou como entradas 3 atrasos (atrasos 2, 4 e 5) para realizar a previsão.

Figura 25 – Seleção de entradas de uma série temporal



Fonte: Autoria própria (2019).

Mas selecionar quais atrasos e em quais quantidades necessárias que quando processados podem apresentar o melhor desempenho de previsão não é uma atividade tão simples, visto que as entradas consideradas para uma mesma série temporal podem ser distintas quando variada a quantidade de neurônios de uma mesma rede, ou ainda quando utilizadas diferentes RNAs para prever uma mesma série temporal (BELOTTI, 2019).

Diferente do que pode parecer, inserir o maior número de atrasos possíveis nem sempre leva a melhores desempenhos de previsão, por outro lado, testar todas as possíveis combinações de entradas pode ser um processo inviável, já que demanda muito tempo computacional, tornando o processo de treinamento de uma RNA extremamente demorado (GUYON; ELISSEEFF, 2003).

Uma das possíveis alternativas na resolução desse problema, que na literatura já apresentou bons resultados é conhecido como método *Wrapper*, que faz seleção progressiva de possíveis entradas. Para que esse método seja empregado, inicialmente é definida a quantidade máxima de atrasos que será testado (SIQUEIRA, 2013).

Como exemplo, suponha que um projetista definiu que a quantidade máxima de atrasos a ser testado é igual a 4, o método *Wrapper* irá realizar as previsões considerando inicialmente o atraso 1 como entrada, depois o 2, 3 e 4 separadamente. O atraso que obtiver o melhor desempenho nessa primeira iteração é selecionado como o melhor subconjunto da primeira iteração e passa a compor de forma fixa o subconjunto para a segunda iteração (GUYON; ELISSEEFF, 2003).

Nesse mesmo exemplo, acredita-se que o atraso 2 teve o melhor desempenho nos resultados de previsões, em uma segunda iteração será testado então os seguintes conjuntos de atrasos: (2,1); (2,3); (2,4). Dessa forma, o melhor resultado dessa segunda iteração irá compor de forma fixa a terceira iteração. Supondo-se que o resultado tenha sido (2,4), por exemplo, em uma terceira iteração testa-se os seguintes conjuntos de atrasos: (2,4,1); (2,4,3) (BELOTTI, 2019).

Se o melhor resultado dessa terceira iteração for (2,4,1), por exemplo, na quarta e última iteração é testado o desempenho da RNA considerando então os atrasos (2,4,1,3). Percebe-se que esse processo de testar progressivamente as entradas, é realizado até que todas as possíveis entradas tenham sido fixadas ao subconjunto (BELOTTI, 2019).

Ao final, aquele subconjunto que durante o processo tiver levado ao menor erro geral será o escolhido (GUYON; ELISSEEFF, 2003). Ou seja, no exemplo dado, após testar as possíveis entradas de forma progressiva, seleciona-se o subconjunto (dentre essas possibilidades (2); (2,4); (2,4,1); (2,4,1,3)), que obteve o melhor desempenho geral de previsão.

Importante mencionar, que o *Wrapper* seleciona os atrasos mais relevantes para a previsão de forma progressiva, portanto não verifica todas as possibilidades de

configurações como entradas mais relevantes para as redes e por isso não garante que a configuração escolhida seja a ótima (GUYON; ELISSEEFF, 2003; SIQUEIRA, 2013). Para o exemplo considerando até 4 possíveis atrasos, por exemplo, utilizando o método *Wrapper*, fez-se apenas 10 iterações, se fosse testada todas as possíveis combinações de entradas ($4^4 = 256$), o tempo de processamento seria aumentado em até 25,6 vezes.

4 METODOLOGIA

Nesse capítulo, inicialmente é abordado o enquadramento metodológico da pesquisa (Seção 4.1), e posteriormente a estrutura do modelo utilizado para realizar a previsão de demanda por eletricidade (Seção 4.2).

4.1 ENQUADRAMENTO METODOLÓGICO

Para enquadramento metodológico da pesquisa utilizou-se as definições dadas por Gil (2008). Sendo assim, do ponto de vista dos objetivos, essa pesquisa enquadra-se como explicativa, pois fez uso da literatura técnica relacionando artigos já consolidados para a construção da base teórica, e posteriormente buscou proporcionar maiores informações sobre a inclusão de variáveis explicativas no processo de previsão de demanda de energia elétrica.

Sendo essa pesquisa uma junção dos conhecimentos já dispostos na literatura e sua ampliação, quanto a sua natureza, essa pesquisa classifica-se como aplicada, pois gerou conhecimentos para aplicação prática do modelo proposto. Em relação a abordagem do problema, essa pesquisa caracterizou-se como qualitativa e quantitativa.

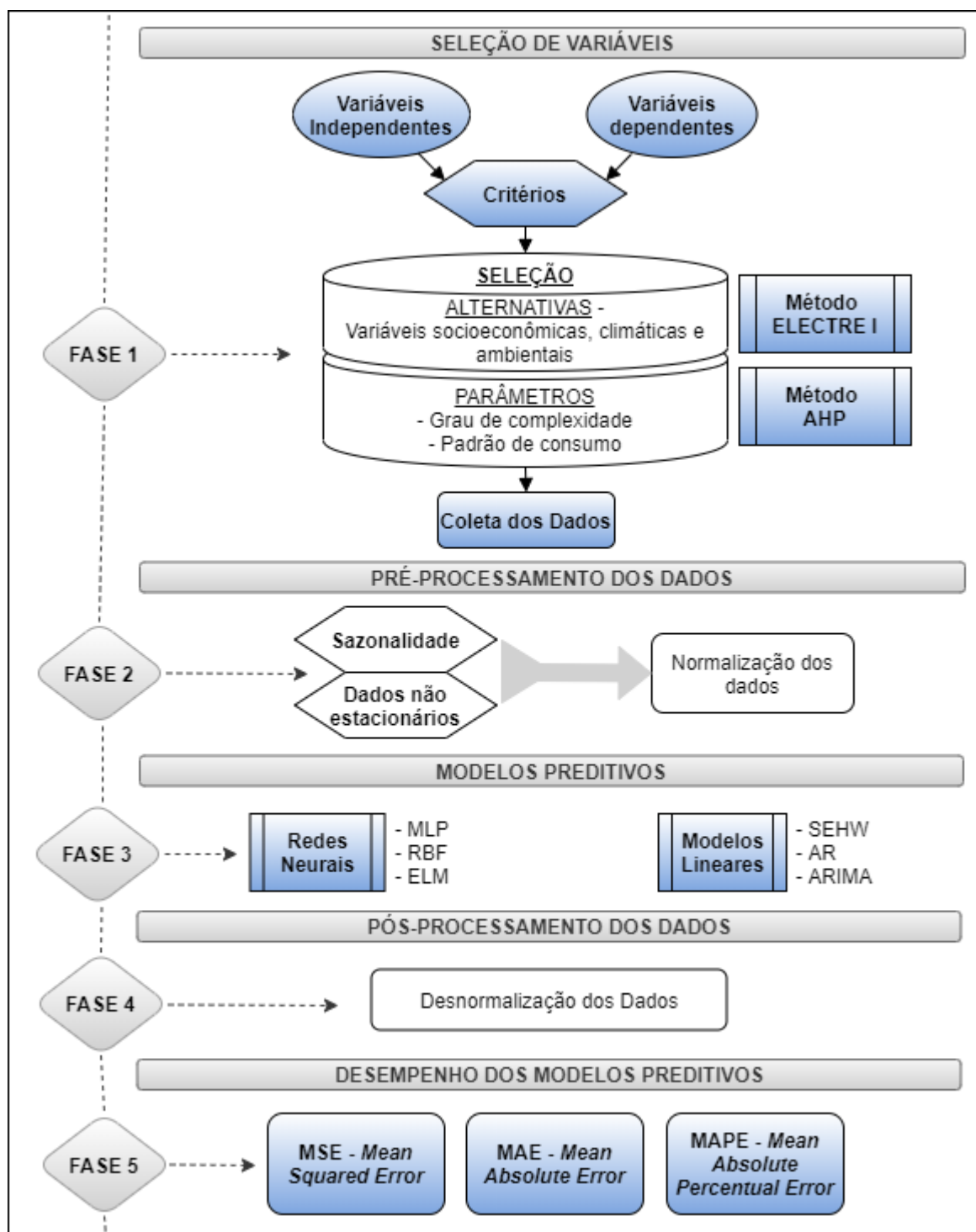
Qualitativa, pois o método multicritério utilizado para seleção de variáveis considerou dados, experiências, percepções e julgamentos de especialistas da área, ou seja, buscou entender o fenômeno específico em profundidade, aproximando-se da realidade do contexto que até então não eram expostos de forma numérica. E quantitativa, pois além de converter a opinião dos especialistas para dados numéricos, também faz uso de modelos preditivos exclusivamente matemáticos/estatísticos.

Do ponto de vista dos procedimentos técnicos, essa pesquisa enquadra-se como bibliográfica, pois foi fundamentada na exploração de estudos científicos sobre demanda de energia elétrica. Documental, pois as previsões foram elaboradas a partir de materiais científicos e dados históricos já publicados. E operacional, pois forneceu simulações e ferramentas quantitativas ao processo de decisão.

4.2 ESTRUTURAÇÃO DO MODELO

O modelo proposto para realizar a previsão de demanda de energia é composto por cinco fases e está exposto na Figura 26.

Figura 26 - Estrutura do modelo



Fonte: Autoria própria (2019).

A primeira fase corresponde a seleção de variáveis independentes que influenciam a demanda através do método multicritério de apoio a decisão ELECTRE I. A segunda corresponde a coleta e pré-processamento dos dados de entrada para o modelo, caso seja verificado que os dados possuem padrões não estacionários como por exemplo, sazonalidade, ciclicidade e tendência.

A terceira realiza as previsões com os modelos lineares e RNAs. Na quarta, um pós-processamento dos dados é realizado, para que estes sejam votem ao padrão com sazonalidade e tendência e então possam ser aplicados as medidas de erros (quinta fase) para verificar e comparar os resultados de previsão. Detalhadamente, essas fases são explicadas nas próximas seções.

4.2.1 Modelo para Seleção de Variáveis

Para apoiar a construção de um modelo de previsão de demanda por eletricidade (variável dependente) que considere variáveis independentes como alternativas de entradas, deve-se definir as alternativas, critérios, peso dos critérios, decisores do contexto e facilitador que será responsável pela aplicação do método. Dessa forma, primeiramente, foram elencadas na literatura as principais variáveis independentes (Figura 27) que foram consideradas influenciadoras por diferentes trabalhos já expostos na Seção 1.2 desse estudo.

Definidas as alternativas, considerando que elas foram classificadas na literatura como variáveis socioeconômicas, climáticas e ambientais, o próximo passo foi selecionar os critérios (Figura 28) correspondentes as características dos problemas de previsão de demanda que podem ter precisões aumentadas com o uso de variáveis explicativas.

A validação dos critérios encontrados na literatura e/ou indicação de um novo critério que tenha relevância para o problema, foi realizada pelos decisores do contexto. Aos quais foram expostos questionamentos como: Na sua opinião, os critérios levantados pela literatura são relevantes? Existem outros que possam ser relevantes para o problema? Se sim, quais?.

Figura 27 - Variáveis independentes elencadas na literatura

Tipos	Variáveis Independentes	Alternativas
Variáveis Socioeconômicas	Renda nacional (renda disponível Per capita)	a1
	População	a2
	Produto Interno Bruto (PIB)	a3
	Preço da eletricidade	a4
	Investimento em ativos fixos	a5
	Investimento estrangeiro direto	a6
	Valor da produção industrial	a7
	Nível de urbanização	a8
Variáveis Climáticas	Temperatura	a9
	Radiação solar	a10
	Umidade relativa	a11
	Precipitação	a12
	Direção do vento	a13
	Velocidade do vento	a14
	Pressão	a15
	Evaporação	a16
	Dias chuvosos	a17
Variável Ambiental	Emissão de carbono	a18

Fonte: Autoria própria (2019).

Figura 28 - Critérios elencados na literatura

Parâmetros	Critérios	Características	
Grau de complexidade	Verificação da disponibilidade dos dados	g1	Quantidade de informações, dados numéricos, dados não numéricos e análise numérica
	Correlação com a variável dependente	g2	Correlação da demanda de eletricidade com as variáveis independentes
Padrão de consumo	Verificação do padrão de consumo por setor	g3	Setores industrial, comercial, residencial, rural e outros
	Verificação do padrão de consumo por região	g4	Características geográficas, econômicas e sociais da região
	Verificação do padrão de consumo por calendário	g5	Férias, dias da semana, feriados e outros

Fonte: Autoria própria (2019).

Após essa etapa, foi definido pesos que refletiram a importância de cada critério. Nessa pesquisa, os pesos dos critérios foram estabelecidos pelo método AHP,

na comparação par a par, realizadas pelos decisores que participaram desse estudo: 2 especialistas sobre o comportamento de demanda de energia.

O uso do AHP é comumente utilizado na literatura, justifica-se seu uso nesse trabalho pela simplicidade de aplicação, capacidade de avaliar critérios de natureza qualitativa e realizar comparação pareada (escala Saaty) dos critérios, pela qual é atribuído valores de 1 a 9, que determina a importância relativa do critério “i” em relação ao “j” e vice versa (ROSA; STEINER; COLMENERO, 2015).

Os decisores do contexto são responsáveis pela decisão e é fundamental que tenham conhecimento técnico sobre o problema abordado. Os analistas que é representado pelos pesquisadores, são os responsáveis por conduzir a aplicação do modelo multicritério, realizar as análises, auxiliar os decisores durante o processo.

Tendo esses conceitos iniciais definidos, o próximo passo foi escolher o método multicritério para seleção de variáveis. Como visto na revisão de literatura, essa escolha depende de diferentes fatores como o contexto a ser analisando, tipo de problemática e estruturas de preferências do decisor (VINCKE, 1992). Na Figura 29 é apresentado as relações que estabeleceram a escolha do método ELECTRE I para a realização dessa pesquisa.

Figura 29 - Condições da seleção do método ELECTRE I

Passos para a escolha	Características
Decisor	Especialistas de demanda de eletricidade
Relações de Preferência	Reflexiva, irreflexiva, simétrica e assimétrica
Estruturas de Preferência	Preferência forte ou estrita, Indiferença
Resultados	Pré-ordem completa
Critérios	Características do problema e variáveis independentes
Alternativas	Variáveis Independentes que influenciam a demanda por eletricidade
Problemática	Seleção (P.α)
Método Multicritério	ELECTRE I

Fonte: Autoria própria (2019).

Como o propósito dessa etapa foi selecionar um único grupo de alternativas, elimina-se os métodos que são utilizados para as atividades de classificação, ordenação e descrição. Nesse ponto, o modelo propõe contribuir para a seleção de um grupo de variáveis ótimas que expliquem a demanda por energia.

4.2.2 Pré-processamento de Dados

Os modelos preditivos apresentam distintas exigências no pré-processamento dos dados antes que se inicie o processo de previsão. Por exemplo: os métodos da família SE, como comentado anteriormente, são compostos por constantes que são capazes de suavizar distintas características que podem compor uma série temporal, e por isso não exigem uma condição de estacionariedade da série, antes que a mesma entre no modelo preditivo.

Se a série de dados for composta por ciclos, um modelo SES contém uma constante (α) que suavizará esse comportamento, caso seja identificado a presença de tendência, um modelo SEH com constante de suavização é aplicado (β), e se além de tendência e ciclo a série apresentar também sazonalidade, é aplicado o modelo de SEHW que suavizará esse comportamento pela constante (γ) (MORETTIN; TOLOI, 2006).

Já o método ARIMA, possui a fase de diferenciação (I) dos dados, de respectiva ordem (d), e por isso, caso seja identificado que a série possui tendência, diferenciações para torná-la estacionária são realizadas pelo próprio modelo. Mas se por exemplo, for identificado que além de tendência a série é composta também por sazonalidade, um pré-processamento que retire essa característica pode ser realizado antes que os dados entrem no modelo preditivo Box *et al.*, (2016).

Por último, além da condição de estacionariedade dos dados para o modelo AR, optou-se por deixar os dados de entradas estacionários também para as RNAs MLP, RBF e ELM, portanto, características que distorcem essa condição serão retiradas da série temporal por um pré-processamento antes que se inicie sua previsão (BELOTTI, 2019).

Dessa forma, o pré-processamento dos dados consistiu inicialmente em uma análise estatística e gráfica da série temporal com o objetivo de identificar seu comportamento, que de acordo com Bouzada (2012) pode conter tendência, ciclicidade, sazonalidade e termo aleatório (flutuação). Conforme Morettin e Tolo (2006) a presença de tendência pode ser verificada por um dos testes mais comumente utilizados na literatura, a saber: teste estatístico de Cox-Stuart, o qual baseia-se nas seguintes hipóteses:

$H_0 = p\text{-valor} > 5\%$ Não há tendência na série;

$H_1 = p\text{-valor} < 5\%$ Há tendência na série.

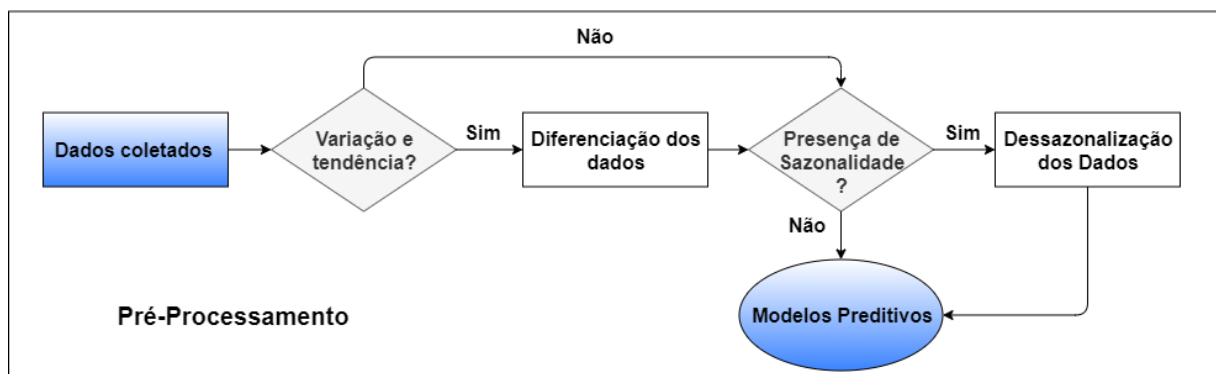
Enquanto a presença de sazonalidade pode ser verificada também por um teste não-paramétrico usualmente empregado na literatura, a saber: teste de Friedman (MORETTIN E TOLOI, 2006), o qual também considera duas hipóteses:

$H_0 = p\text{-valor} > 5\%$ Não há sazonalidade na série;

$H_1 = p\text{-valor} < 5\%$ Há sazonalidade na série.

Após identificado o comportamento da série, um pré-processamento dos dados foi realizado de acordo com as características apresentadas, como exposto na Figura 30.

Figura 30 - Tratamento dos dados



Fonte: Autoria própria (2019).

Caso seja identificado que a série é munida por tendência e ciclicidade faz-se uma ou mais diferenciações (Equação 33) até torná-la estacionária.

$$y'_t = y_{t+1} - y_t \quad (33)$$

em que, y'_t é a diferença entre o valor atual da série menos o valor anterior. Caso também seja identificado uma componente sazonal, um processo estatístico conhecido como dessazonalização é aplicado nos dados (BALLINI, 2000), este é expresso na Equação 34.

$$z_i = \frac{x_i - \mu_m}{\sigma_m} \quad (34)$$

em que, x_i é o dado que está sendo dessazonalizado. Supondo que esse dado seja o consumo de energia no mês de janeiro, por exemplo, μ_m é a média de janeiro de todos os anos disponíveis na série analisada e σ_m é o desvio padrão de todos os janeiros. Tendo os dados estacionários para os respectivos modelos que exigem essa condição, segue-se para a etapa de previsão de demanda, exposta na seguinte seção.

4.2.3 Modelos Preditivos

Para aplicação dos métodos de previsão, é necessário que os dados sejam divididos em três conjuntos: treinamento, validação e teste. A amostra de dados utilizadas para treinamento, deve ser composta por aproximadamente 60% até 90% do total de amostras disponíveis, conseqüentemente, o conjunto para validação e teste é composto pelas amostras restantes (de 10% a 40% das amostras disponíveis) (SILVA; SPATTI; FLAUZINO, 2016).

Com o processo de treinamento e validação, é possível identificar se a generalização das soluções está com um erro aceitável para antecipar o comportamento da série em situações reais. Conforme Belotti (2019), as amostras do conjunto de teste, não podem ter sido utilizadas no treinamento, desse modo, o modelo será testado com um conjunto de amostras aos quais ele nunca teve acessibilidade.

Sabendo-se que a série temporal endógena correspondente a demanda de energia do estado do Paraná é composta por 186 dados que corresponde ao período mensal de janeiro de 2004 até junho de 2019, para essa pesquisa separou-se os dados em três conjuntos, como segue:

- a) Treinamento:** Dados entre o período de janeiro de 2004 até dezembro de 2015, totalizando 144 dados;
- b) Validação:** Dados entre o período de janeiro de 2016 até dezembro de 2017, totalizando 24 dados;

c) Teste: Dados entre o período de janeiro de 2018 até junho de 2019, totalizando 18 dados;

Para os casos em que foram incluídas variáveis exógenas no processo de previsão, a quantidade de dados utilizadas para treinamento e validação foram respectivamente as mesmas. Tendo separado os dados, outro passo importante no processo de previsão diz respeito a estimação dos parâmetros exigidos por cada modelo preditivo.

Em relação ao modelo SEHW, parâmetros como as constantes de suavização α , β e γ precisaram ser determinados. Segundo Montgomery *et al.* (1990), não há um método exato para a definição desses valores. Portanto, para essa pesquisa, realizou-se testes paramétricos até encontrar valores que apresentassem os melhores ajustes para as previsões.

Já os modelos AR e ARIMA tiveram seus coeficientes (p) e (q) estimados a partir do gráfico de autocorrelação parcial (PACF) entre a série de dados e seus atrasos. A determinação desses coeficientes de AR e MA puderam ser determinados pela análise gráfica. Sendo a determinação desses termos necessárias para corrigir qualquer autocorrelação que pode permanecer na série já estacionária (BOX *et al.*, 2016).

Já a determinação dos atrasos mais relevantes que foram utilizados como entradas para as RNAs MLP, RBF e ELM foram determinados a partir da utilização do método *Wrapper*. Para que o método fosse iniciado utilizou-se até 6 atrasos, sendo esse valor usualmente empregada pelo setor Elétrico Brasileiro (SEB) (SIQUEIRA, 2013; BELOTTI, 2019), além de que o horizonte de previsão foi determinado para ser realizado considerando até 6 passos à frente.

Em relação a arquitetura das redes, todas foram construídas com apenas uma camada intermediária e com quantidade de neurônios iniciando em 5, com incrementos de 5 em 5, até que se chegasse a 200 neurônios. Os treinamentos iterativos foram realizados considerando como critério de parada uma melhora no MSE menor que 10^{-6} ou um número máximo de 2000 épocas de treinamento.

Tendo as possíveis configurações definidas para todos os modelos, estas foram simuladas 30 vezes, escolhendo-se a melhor execução (menor MSE) junto ao conjunto de teste. Como os dados eram mensais, as previsões foram realizadas considerando os seguintes horizontes de passos à frente: P = 1 (próximo mês), P = 3 (próximos 3 meses), P = 6 (próximos 6 meses). Desconsiderou-se realizar previsões

para $P = 12$ (próximo ano), pois o conjunto de dados separados para teste correspondeu ao período de janeiro de 2018 até junho de 2019.

Por fim, para que as previsões pelos modelos lineares SEHW, AR e ARIMA pudessem ser realizadas, utilizou-se o auxílio do *software* Minitab®, enquanto as configurações das RNAs MLP, RBF e ELM foram desenvolvidas utilizando a linguagem Python®, possibilitando também realizar as previsões.

4.2.4 Pós-Processamento de Dados

Tendo as previsões realizadas, os dados precisaram retornar ao padrão original para que os desempenhos resultantes das previsões pudessem ser comparados. Dessa forma, aplicou-se a Equação 35 para voltar ao padrão sazonal e a Equação 36 para voltar ao comportamento ausente de diferenciação.

$$x_i = (z_i \cdot \sigma_m) + \mu_m \quad (35)$$

$$y_t = y_{t-1} + y'_t \quad (36)$$

em que x_i representa os dados sazonalizados, z_i são os dados dessazonalizados, μ_m é a média e σ_m é o desvio padrão de cada ano. y_t representa os dados sem diferenciação, y_{t-1} é o dado anterior sem diferenciação que é somado ao dado posterior previsto y'_t o qual está diferenciado. Tendo os dados previstos já colocados no padrão original, faz-se comparação dos desempenhos obtidos pelos modelos de previsão, como exposto na próxima seção.

4.2.5 Desempenho dos Modelos Preditivos

Com os dados já pós-processados, os desempenhos das previsões realizadas foram comparados utilizando as seguintes medidas de erros: MSE, MAE e o MAPE, respectivamente expressos pelas Equações 37, 38 e 39.

$$MSE = \frac{1}{N_S} \sum_{t=1}^{N_S} (d_t - y_t)^2 \quad (37)$$

$$MAE = \frac{1}{N_S} \sum_{t=1}^{N_S} |d_t - y_i| \quad (38)$$

$$MAPE = \frac{1}{N_S} \sum_{t=1}^{N_S} \left| \frac{d_t - y_i}{d_t} \right| \quad (39)$$

Nessas equações, N_S representa o número de amostras, d_t é o resultado desejado e y_i é a saída dada pelos modelos. Os desempenhos resultantes de cada modelo estão expostos nas seções seguintes.

A medida MSE é comumente empregada em RNAs como critério de parada em processos de treinamento, e por esse motivo, considerou-se utilizá-la também para a comparação do conjunto de teste, visto que essa medida eleva os erros ao quadrado, automaticamente penalizará os erros maiores mais fortemente (HAYKIN, 2008; MOHAMMED, 2018).

Para comparação, optou-se por fazer uso das medidas MAPE e MAE, pois além de serem bastante empregadas na literatura, são de fácil entendimento e não fazem distinção entre erros grandes e pequenos, visto que o MAE representa a média entre o valor previsto e o real, e o MAPE apresenta esse mesmo resultado em termos percentuais, ambos podem trazer resultados de ranking diferente do MSE (VU; MUTTAQI; AGALGAONKAR, 2015; TUTUN; CHOU; CANYLMAZ, 2015).

5 RESULTADOS

Nesse capítulo está exposto as previsões pelos modelos lineares e RNAs para a variável dependente (Seção 5.1) referente ao consumo de energia elétrica no Paraná. Posteriormente, apresentou-se a estrutura para a aplicação do modelo multicritério de classificação das variáveis explicativas (independentes) para as previsões (Seção 5.2), bem como as previsões realizadas com a inclusão dessas variáveis exógenas (independentes) no modelo preditivo (Seção 5.3).

5.1 PREVISÃO DA VARIÁVEL DEPENDENTE

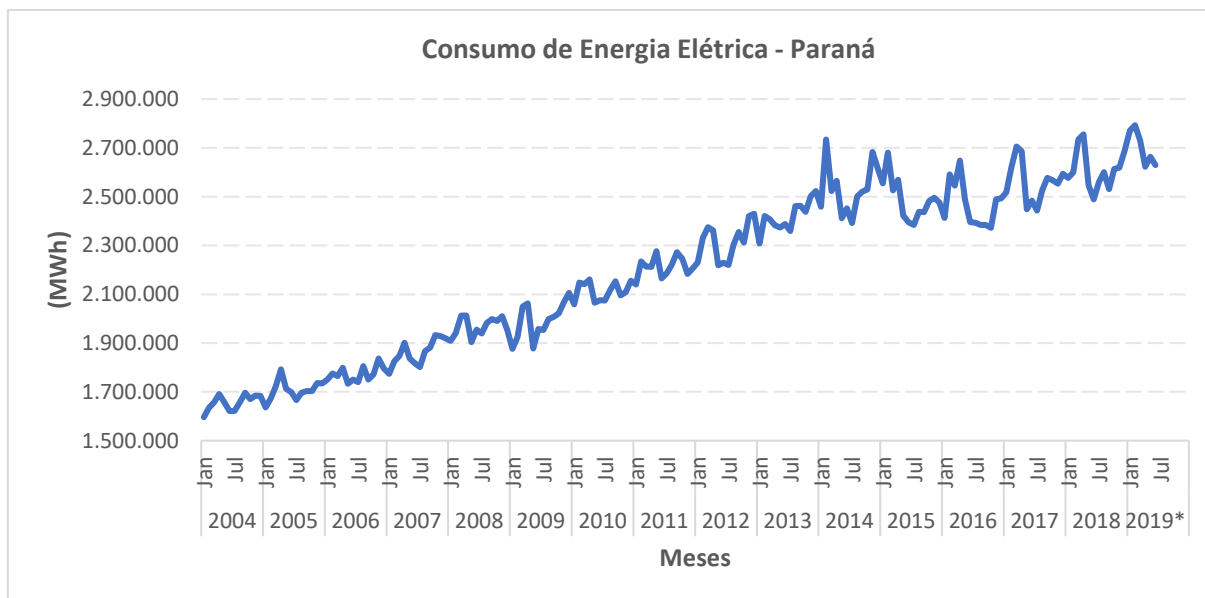
A presente seção discute uma aplicação prática dos modelos preditivos lineares e RNAs. Para concretização desse estudo, realizou-se uma coleta de dados brutos sobre consumo de energia elétrica no estado do Paraná, os quais estavam disponíveis na base de dados do site da Empresa de Pesquisa Energética do governo federal (EPE, 2019).

Estes dados referem-se ao consumo mensal de energia elétrica na rede (MWh) do Paraná, entre o período de janeiro de 2004 até junho de 2019, totalizando 186 amostras mensais de consumo de energia elétrica. Após essa coleta, seguiu-se para os procedimentos de análise estatística e definição dos parâmetros dos modelos (seção 5.1.1) e previsão da série temporal e comparação dos desempenhos preditivos (seção 5.1.2).

5.1.1 Análise estatística e definição de parâmetros

Tendo os dados coletados, inicialmente foi identificada a regularidade dos mesmos e então fez-se a plotagem destes em um gráfico, com o objetivo de identificar as componentes da série de consumo de eletricidade do estado do Paraná (Figura 31). Pela análise gráfica, verificou-se que a série de dados é composta por tendência de crescimento, ciclos que variam ao longo do tempo e padrões de sazonalidade que se repetem.

Figura 31 - Dados de consumo de energia do Estado do Paraná referente ao período de janeiro de 2004 até junho de 2019



Fonte: Empresa de Pesquisa Energética (2019).

Diante dessa análise, objetivando confirmar ou não a presença de tendência e sazonalidade, aplicou-se os testes estatísticos não-paramétricos de Cox-Stuart e Friedman (Tabela 1).

Tabela 1 – Testes não-paramétricos para análise de tendência e sazonalidade

testes	Série Original	Série diferenciada e dessazonalizada
	<i>P-Valor</i> < 5%	<i>P-Valor</i> > 5%
<i>Cox-Stuart</i>	0	0,251322103
<i>Friedman</i>	3,49E-15	0,994761543

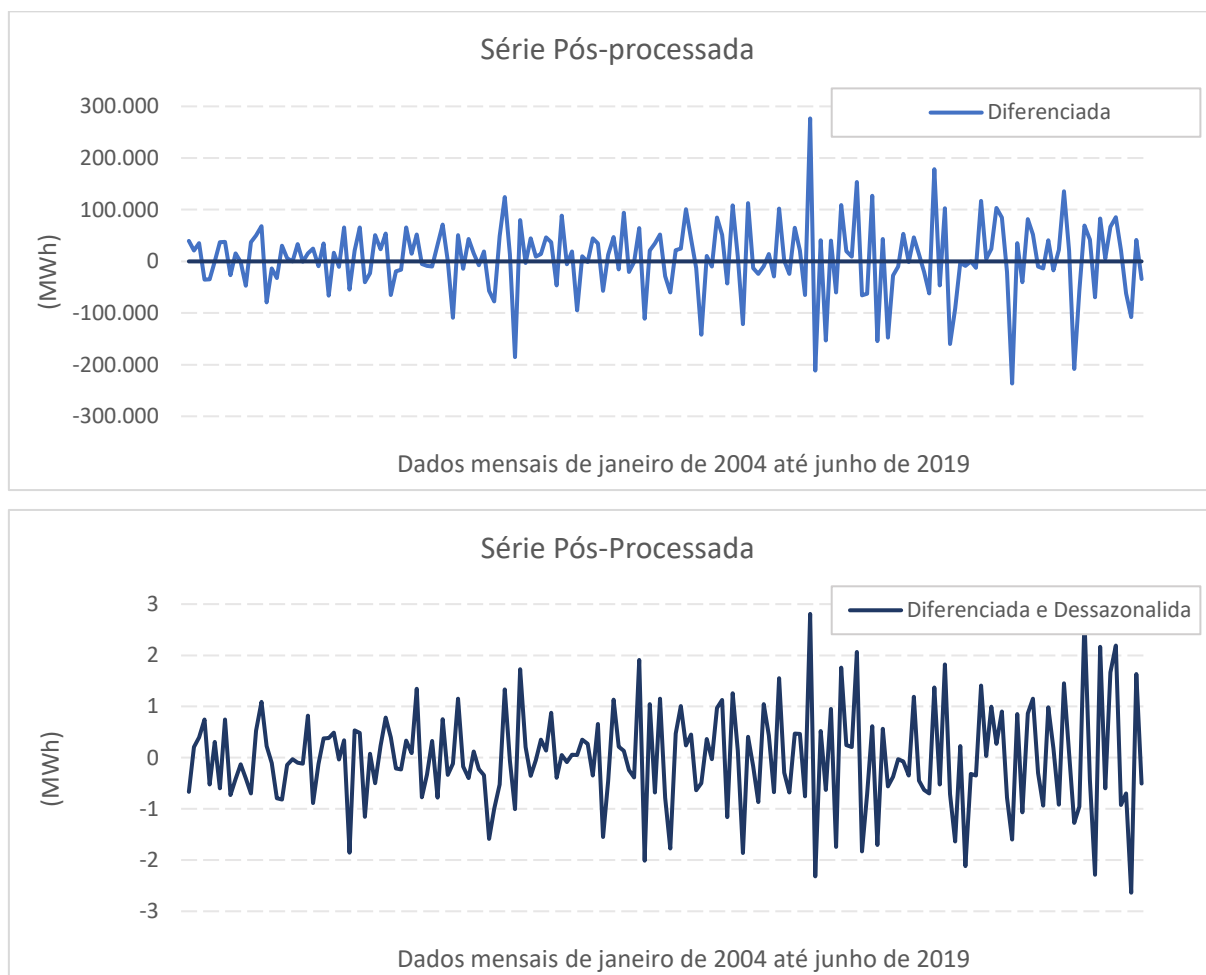
Fonte: Autoria própria (2019).

Tendo os resultados expostos na Tabela 1, foi possível verificar que o *P-valor* tanto para o teste de tendência quanto para o teste de sazonalidade para a série analisada foi menor que 5%, descartando-se a hipótese nula para ambos os casos, portanto foi identificado a presença de tendência e sazonalidade na série, confirmando a análise gráfica da Figura 31.

Verificado isso, aplicou-se as Equações 33 e 34, apresentadas na Seção 4.2.2 deste estudo, com o objetivo de retirar a componente sazonal e de tendência e tornar a série estacionária. Os resultados dessa aplicação também estão apresentados na Tabela 1, o qual verifica-se que para ambos os casos o *P-valor* foi maior que 5%, indicando que apenas uma diferenciação e a dessazonalização foram suficientes para

tornar a série estacionária. Graficamente, esses resultados podem ser visualizados na Figura 32.

Figura 32 - Dados de consumo de energia do Estado do Paraná referente ao período de janeiro de 2004 até junho de 2019, diferenciados e dessazonalizados



Fonte: Autoria própria (2019).

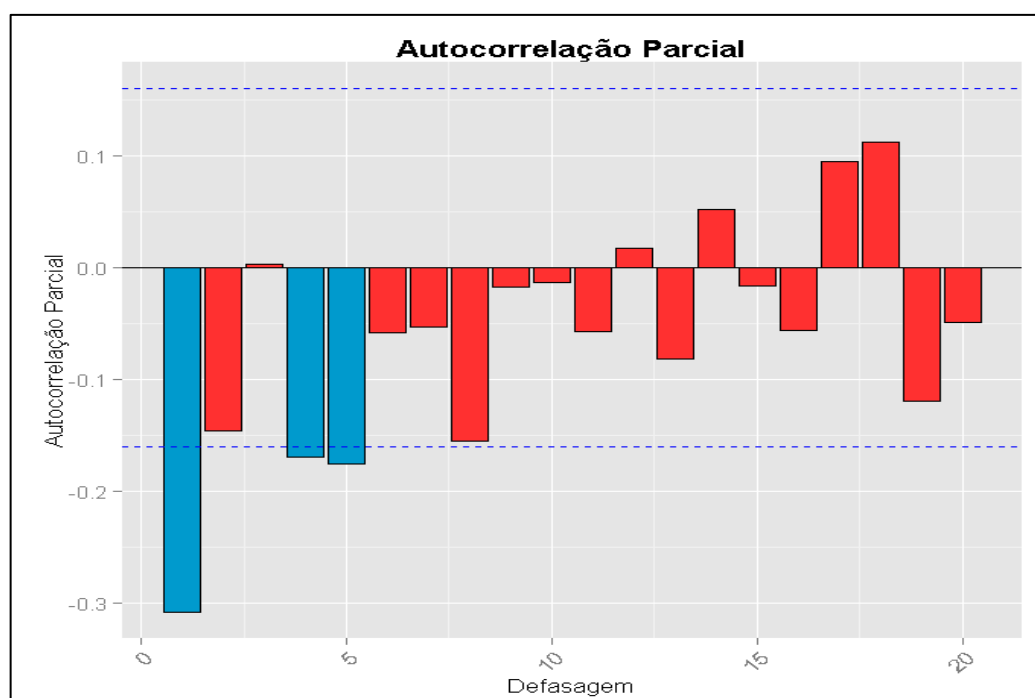
Observa-se que os valores da série estacionária variam entre 3 e -3, oscilando em torno de zero indicando a condição de estacionariedade nos dados. Tendo a série analisada, o próximo passo foi definir os parâmetros para previsão de cada modelo. O modelo de SEHW, como mencionado anteriormente, é utilizado para prever séries que apresentam componentes com ciclicidade, tendência e sazonalidade (MENEHINI *et al.*, 2018). Sendo essas, identificadas na série temporal de energia elétrica do Paraná (Figura 30, Tabela 1).

Como esse modelo apresenta constantes que suavizam esses parâmetros cíclicos (α), sazonais (γ) e com tendência (β), utilizou-se a série apresentada na Figura 20 para realização da previsão. Fez-se testes paramétricos até encontrar os valores

que suavizassem essas constantes e resultassem no melhor desempenho preditivo para o modelo, testando tanto para multiplicativo quanto aditivo. Os valores encontrados foram: SEHW aditivo $\alpha = 0,63$, $\beta = 0,12$, $\gamma = 0,45$.

Após a atribuição ao SEHW, o próximo passo foi definir os parâmetros para os modelos AR (p) e ARIMA (p, d, q). Para isso, construiu-se o gráfico para análise da função de autocorrelação parcial (FACP) (Figura 33) para a série já estacionária (Figura 32). Com o auxílio desse gráfico foi possível que os parâmetros p e q do AR e ARIMA fossem estimados, já o parâmetro d recebeu o valor 1, pois correspondeu ao número de diferenciações necessárias para tornar a série estacionária.

Figura 33 - Gráfico da função de autocorrelação parcial (FACP) para a série temporal diferenciada e dessazonalida de consumo de energia elétrica no Paraná



Fonte: Autoria própria (2019).

Pela análise da Figura 33, é possível perceber que a série temporal possui coeficiente de autocorrelação parcial significativo até o quinto atraso, com um grau de confiança de 95%. Com os testes realizados considerando as possíveis configurações de 1 até 5 atrasos para p e q , o modelo AR (2) e ARIMA (5,1,5) alcançaram os menores erros de previsão. Para ajustar os modelos o estimador de máxima verossimilhança foi utilizado.

Quanto as RNAs, como exposto na seção 4.2.3, foi variada empiricamente a quantidade de neurônios na camada intermediária começando com 5 neurônios, com

incremento de 5, até 200 neurônios. Para cada variação de neurônio aplicou-se o método *wrapper* que fez uma varredura progressivamente considerando até 6 atrasos. Dessa forma, as configurações das RNAs que obtiveram os melhores resultados de previsão foram: MLP de 5 neurônios na camada intermediária e com o terceiro atraso como entrada para previsão; RBF com 10 neurônios e atraso 6; ELM com 25 neurônios e atrasos 1 e 6.

Com os parâmetros dos modelos definidos, fez-se a previsão de demanda por energia elétrica no Paraná para 1, 3 e 6 passos à frente, referentes aos 18 dados (meses dos anos de 2018 e 2019) de teste. Na próxima seção está exposto os resultados obtidos das previsões, comparado aos dados separados para teste pelas medidas: Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE) e Erro Percentual Absoluto Médio (MAPE). Neles, os valores encontrados nas medidas de erro são referentes a melhor das 30 simulações realizadas para cada modelo.

5.1.2 Previsão da variável independente

Com os parâmetros definidos na seção anterior, as previsões para 1 passo à frente ($P = 1$) para a série de consumo de energia no Paraná foi realizada. Na Tabela 2 é apresentado os resultados das medidas MSE com valores reais, MSE (d) com valores diferenciados e dessazonalizados, MAE real, MAE (d) diferenciado e dessazonalizado e o percentual do MAPE apenas na escala real, pois como trata-se de porcentagem, sua escala não é sensível quando se faz a diferenciação e dessazonalização. A Tabela 3 traz o ranking desses resultados.

Tabela 2 – Resultados das previsões para 1 passo à frente, comparado pelas medidas de erro MSE, MAE e MAPE

1 passo à frente						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
SEHW	$\alpha = 0,63, \beta = 0,12, \gamma = 0,45$	5679573386,83	0,2566	62124,83	0,4286	2,3265
AR	$p = 2$	5247636249,96	0,2371	60064,66	0,4144	2,2765
ARIMA	$p = 5, d = 1, q = 5$	5860003521,64	0,2647	65994,23	0,4553	2,5035
MLP	neurônios = 5, atraso = 3	6557391893,34	0,3152	70231,17	0,4796	2,6589
RBF	neurônios = 10, atraso = 6	4634913071,81	0,2094	55409,00	0,3823	2,0994
ELM	neurônios = 25, atrasos = 1 e 6	4599769468,03	0,1945	48604,73	0,3463	1,8538

Fonte: Autoria própria (2019).

Tabela 3 – Ranking de desempenho das previsões para 1 passo à frente

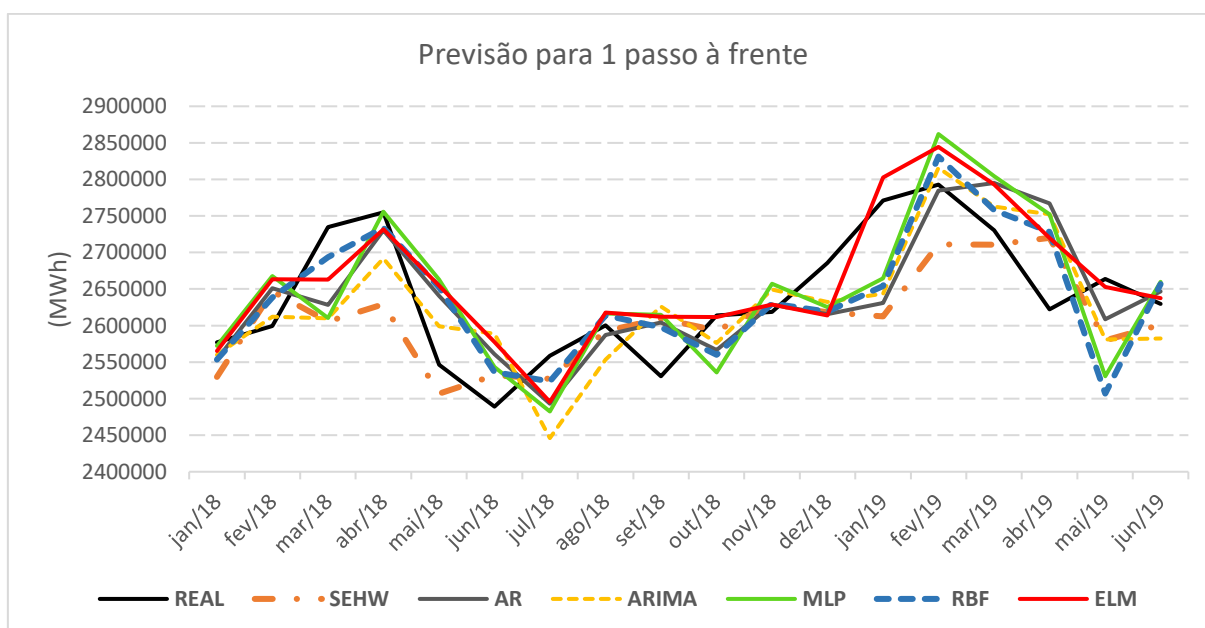
Ranking de desempenho						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
SEHW	$\alpha = 0,63, \beta = 0,12, \gamma = 0,45$	4	4	4	4	4
AR	$p = 2$	3	3	3	3	3
ARIMA	$p = 5, d = 1, q = 5$	5	5	5	5	5
MLP	neurônios = 5, atraso = 3	6	6	6	6	6
RBF	neurônios = 10, atraso = 6	2	2	2	2	2
ELM	neurônios = 25, atrasos = 1 e 6	1	1	1	1	1

Fonte: Autoria própria (2019).

Os resultados da Tabela 2 e 3 apresenta que a melhor previsão realizada foi obtida pela RNA ELM, seguida da RNA RBF, para todas as medidas de erro analisadas. Enquanto o pior resultado de previsão foi obtido pela rede MLP.

Considerando o MSE no domínio real, o erro do segundo colocado em relação ao primeiro em termos percentual foi de 0,76%, e realizado a mesma comparação do terceiro, quarto, quinto e sexto modelo colocado em relação ao desempenho do MSE da rede ELM, esses erros foram respectivamente de 14,08%, 23,48%, 27,40% e 42,56%. A Figura 34 apresenta as previsões realizadas pelos modelos e os dados reais separados para teste.

Figura 34 - Gráfico da comparação dos dados reais com os dados previstos para 1 passo à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando os modelos SEHW, AR, ARIMA, MLP, RBF e ELM



Fonte: Autoria própria (2019).

Embora visualmente os resultados pareçam próximos pela Figura 34, a escala da série analisada tem consumo mensal na casa de milhões de (MWh) para todo o estado do Paraná, a diferença entre o valor real e previsto que está aproximadamente na casa de mil (MWh), quando elevado ao quadrado e dividido pelo número de previsões (cálculo do MSE) pode chegar a casa de bilhões de (MWh) por mês para todo o estado. Isso porque, como comentado por Belotti (2019), o MSE penaliza mais fortemente erros maiores se comparado as outras medidas de desempenho.

Por se tratar de valores grandes a serem previstos pode se dizer então que o modelo ELM obteve um resultado bastante superior de previsão em relação aos outros modelos. Dando sequência, fez-se também previsões para a mesma série considerando o horizonte de 3 passos à frente, tendo seus resultados comparados pelas medidas MSE, MSE(d), MAE, MAE(d) e MAPE (Tabela 4) e fazendo o Ranking desses desempenhos (Tabela 5).

Tabela 4 – Resultados das previsões para 3 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE

3 passos à frente						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
SEHW	$\alpha = 0,63, \beta = 0,12, \gamma = 0,45$	7485077440,82	0,3292	72273,26	0,4987	2,7328
AR	$p = 2$	7013687861,15	0,3247	70897,07	0,4836	2,6770
ARIMA	$p = 5, d = 1, q = 5$	8924931364,90	0,4853	80795,54	0,5770	3,0485
MLP	neurônios = 5, atraso = 3	6573580787,03	0,3162	70331,85	0,4806	2,6625
RBF	neurônios = 10, atraso = 6	6528243517,99	0,2871	62256,53	0,4296	2,3556
ELM	neurônios = 25, atrasos = 1 e 6	6366837641,56	0,2740	64652,19	0,4296	2,4487

Fonte: Autoria própria (2019).

Tabela 5 – Ranking de desempenho das previsões para 3 passos à frente

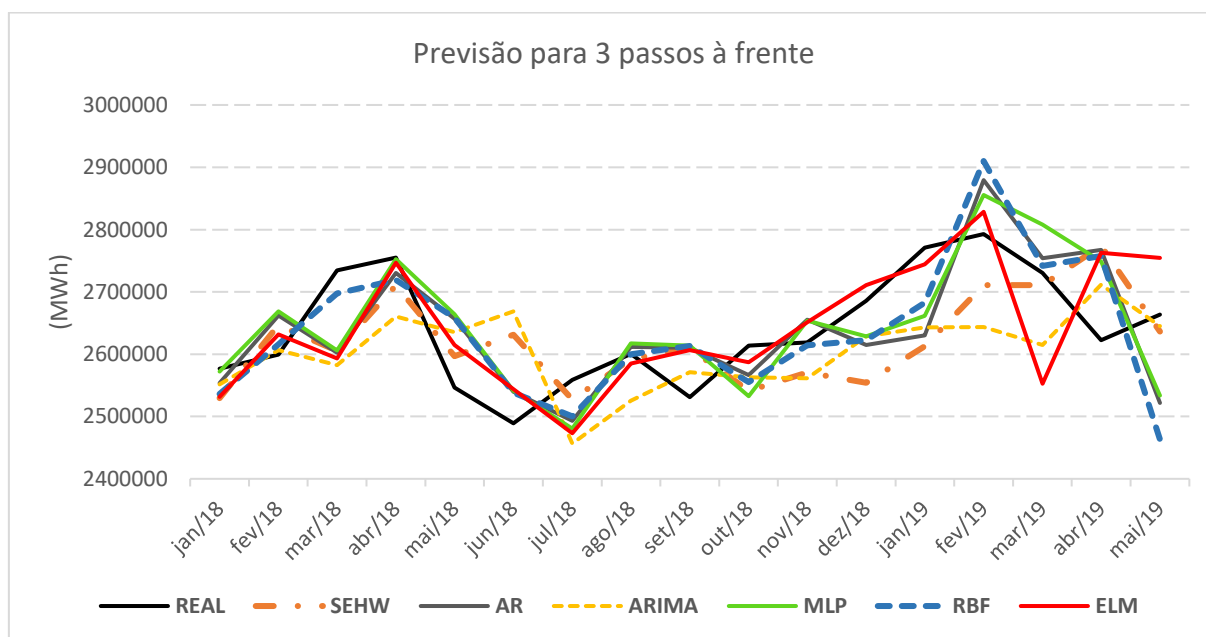
Ranking de desempenho						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
SEHW	$\alpha = 0,63, \beta = 0,12, \gamma = 0,45$	5	5	5	5	5
AR	$p = 2$	4	4	4	4	4
ARIMA	$p = 5, d = 1, q = 5$	6	6	6	6	6
MLP	neurônios = 5, atraso = 3	3	3	3	3	3
RBF	neurônios = 10, atraso = 6	2	2	1	1	1
ELM	neurônios = 25, atrasos = 1 e 6	1	1	2	2	2

Fonte: Autoria própria (2019).

Pela análise das Tabelas 4 e 5 a rede ELM novamente teve desempenho de previsão superior as demais, se comparado o MSE. Porém esse Ranking é invertido com a RBF quando comparado as outras medidas de desempenho MAE e MAPE. Lembrando que o MAE corresponde ao erro absoluto médio e o MAPE é uma média percentual simples. Ambos não fazem distinção de erros pequenos e erros grandes, como o caso do MSE que os eleva os erros ao quadrado.

Se comparado o MSE no domínio real, o erro da rede RBF em relação a ELM para termos em percentual foi de 4,62%, não muito distante da rede MLP em relação a ELM que foi de 5,91%. Já os modelos AR, SEHW e ARIMA obtiveram um desempenho muito inferior se comparados a rede ELM, sendo decrescidos respectivamente em 18,50%, 31,98% e 73,15%. A Figura 35 apresenta as previsões realizadas pelos modelos e os dados reais separados para teste.

Figura 35 - Gráfico da comparação dos dados reais com os dados previstos para 3 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando os modelos SEHW, AR, ARIMA, MLP, RBF e ELM



Fonte: Autoria própria (2019).

Por fim, fez-se também previsões para a mesma série considerando o horizonte de 6 passos à frente, tendo seus resultados comparados pelas medidas MSE, MSE(d), MAE, MAE(d) e MAPE (Tabela 6) e fazendo o Ranking desses desempenhos (Tabela 7). A Figura 36 traz as previsões realizadas pelos modelos e os dados reais separados para teste.

Tabela 6 – Resultados das previsões para 6 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE

6 passos à frente						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
SEHW	$\alpha = 0,63, \beta = 0,12, \gamma = 0,45$	8283400542,03	0,3916	76488,21	0,5268	2,8931
AR	$p = 2$	6840287879,73	0,3175	71062,50	0,4815	2,6849
ARIMA	$p = 5, d = 1, q = 5$	8174710719,13	0,3865	73181,97	0,5041	2,7387
MLP	neurônios = 5, atraso = 3	6686402960,82	0,3168	69819,32	0,4766	2,6426
RBF	neurônios = 10, atraso = 6	6595147324,83	0,3118	68586,05	0,4724	2,5950
ELM	neurônios = 25, atrasos = 1 e 6	6406990346,37	0,3020	63890,27	0,4487	2,4310

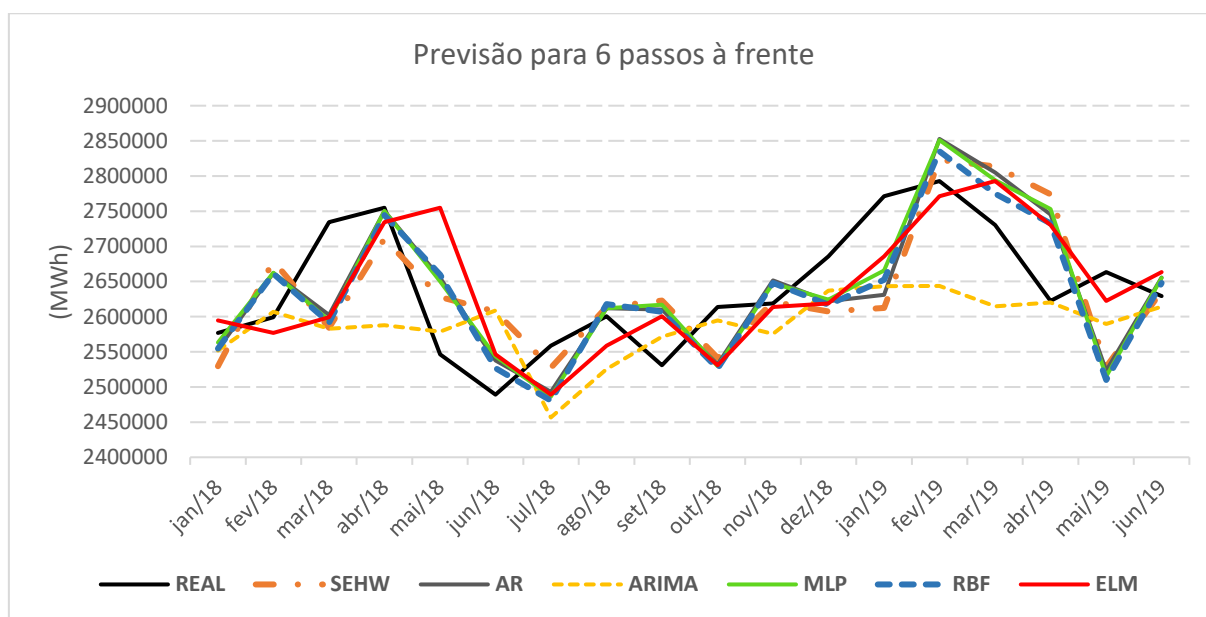
Fonte: Autoria própria (2019).

Tabela 7 – Ranking de desempenho das previsões para 6 passos à frente

Ranking de desempenho						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
SEHW	$\alpha = 0,63, \beta = 0,12, \gamma = 0,45$	6	6	6	6	6
AR	$p = 2$	4	4	4	4	4
ARIMA	$p = 5, d = 1, q = 5$	5	5	5	5	5
MLP	neurônios = 5, atraso = 3	3	3	3	3	3
RBF	neurônios = 10, atraso = 6	2	2	2	2	2
ELM	neurônios = 25, atrasos = 1 e 6	1	1	1	1	1

Fonte: Autoria própria (2019).

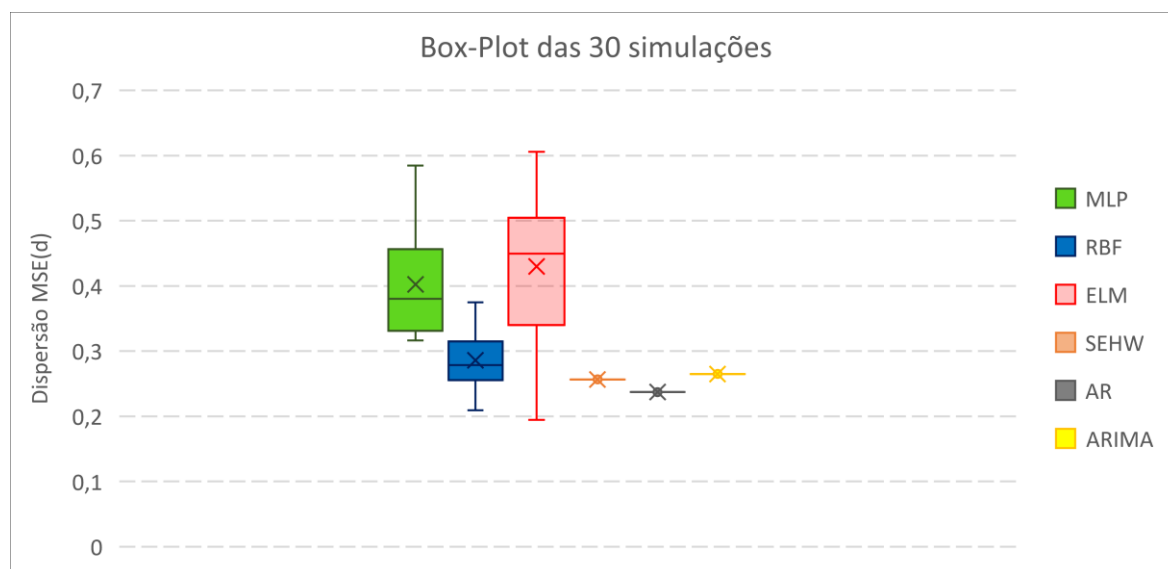
Figura 36 - Gráfico da comparação dos dados reais com os dados previstos para 6 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando os modelos SEHW, AR, ARIMA, MLP, RBF e ELM



Fonte: Autoria própria (2019).

Percebe-se que as redes ELM e RBF mantiveram a posição no ranking com os melhores desempenhos de previsões para todos os passos à frente, já o modelo de SEHW para 6 passos à frente obteve o pior desempenho de previsão. Quando se compara o MSE do domínio real da rede ELM em relação aos modelos RBF, MLP, AR, ARIMA e SEHW, em termos percentuais o desempenho é reduzido respectivamente em 5,38%, 7,99%, 12,39%, 50,55% e 53,66%. Como já mencionado anteriormente, os modelos de previsões propostos foram executados 30 vezes, isso para identificar a sua dispersão nos resultados. A Figura 37 apresenta o gráfico de Box-plot dessas simulações.

Figura 37 - Gráfico Box-Plot com a dispersão do MSE(d) para as 30 simulações dos modelos lineares e RNAs



Fonte: Autoria própria (2019).

Analisando a Figura 37, percebe-se que os modelos lineares SEHW, AR e ARIMA não apresentaram dispersão nas 30 simulações, resultado esse já esperado, pois estes tiveram seus coeficientes estimados por máxima verossimilhança, tendo permanecido os mesmos para todas as execuções. O número de neurônios e atrasos estimados para as redes também permaneceram os mesmos nas 30 simulações, porém, para que as RNAs adquiram a capacidade de generalização e mapeamento dos dados é necessário que se faça o treinamento delas.

Lembrando que nessa etapa do processo, as redes têm seus pesos sinápticos inicializados de forma aleatória e ajustados no treinamento, e por isso, apresenta distintos resultados em todas as suas execuções. Observa-se no gráfico

que o modelo preditivo que apresentou maior dispersão (amplitude interquartílica) em seus resultados foi a rede ELM, seguido da MLP e RBF. Outra característica encontrada, diz respeito a assimetria da distribuição, isso pode ser visualizado pela posição da linha mediana, em que para a rede ELM encontra-se próxima do quartil superior, e as redes RBF e MLP próximas ao quartil inferior.

A mediana próxima ao quartil superior, indica que existem amostras com dados discrepantes com valores menores, isso explica o fato do valor da média (que é representado por \bar{x} no gráfico) ser menor do que o da mediana. Por outro lado, a mediana próxima do quartil inferior indica uma discrepância de dados com valores maiores, e conseqüentemente a média aparece no gráfico acima da mediana.

Importante observar, que a rede RBF apresentou pouca assimetria, uma vez que a mediana e média aparece bastante próximas no gráfico de dispersão. Outra observação, diz respeito aos outliers, que são “pontos desgarrados” que podem aparecer fora do desenho do box-plot, e não se fizeram presentes em nenhuma das dispersões apresentadas.

5.2 SELEÇÃO DE VARIÁVEIS INDEPENDENTES

Para testar o modelo proposto, selecionou-se como variável endógena a série temporal referente ao consumo de energia elétrica do estado do Paraná. Como variáveis exógenas (alternativas do problema), pesquisou-se as séries temporais referentes ao Quadro 9 do Capítulo 4. Essas séries foram encontradas nas bases de dados a saber: Sistema de Estimativa de Emissão de Gases de Efeito Estufa – SEEG; Instituto Nacional de Meteorologia – INMET; Instituto de Pesquisa Econômica Aplicada – IPEA; Instituto Brasileiro de Geografia e Estatística – IBGE; e Câmara de comercialização de energia elétrica – CCEE.

Em relação aos critérios utilizados nessa pesquisa (Quadro 10 do Capítulo 4), estes foram validados pelos decisores e analistas envolvidos no processo. Os critérios utilizados nesse estudo apresentam naturezas qualitativa e quantitativa. Em relação aos critérios de natureza qualitativa, a mensuração deles foi desenvolvida com base na percepção e experiência dos decisores envolvidos no processo.

A família de critérios utilizada pelo método ELECTRE I é do tipo simples, fazendo uso de índices de concordância e discordância. O método busca selecionar alternativas baseando-se na sobreclassificação, incorporando preferências e indiferenças estabelecidas pelos decisores frente ao problema e alternativas. Através das comparações de alternativas em relação aos critérios, o decisor estabelece pesos e avaliações referentes as alternativas e critérios do problema (GOMES; ARAYA; CARIGNANO, 2004). Tendo essas definições, detalhes referentes aos critérios utilizados nesse estudo são apresentados a seguir:

Critério g1 – Verificação da disponibilidade de dados: como as previsões são realizadas com base em modelos matemáticos e computacionais, é necessário que se tenham para entradas nos modelos preditivos dados disponíveis e de qualidade em relação as alternativas do problema. Dessa forma, esse critério tem como finalidade verificar a apresentação dos dados frente a resolução do problema, se eles podem contribuir de forma satisfatória para aplicação nas entradas dos modelos preditivos.

Esse critério está relacionado aos seguintes critérios encontrados na literatura: quantidade de informações, quantidade de dados disponíveis, dados numéricos, dados não numéricos, análise numérica. Como esse critério é do tipo quantitativo e está relacionado a disponibilidade dos dados para treinamento dos modelos, a pontuação atribuída as alternativas são representadas pela Equação 31:

$$gd = \frac{qnt\ vi}{qnt\ vd} \times 100 \quad (31)$$

em que gd representa o grau de disponibilidade dos dados, $qnt\ vi$ corresponde a quantidade de dados disponíveis da variável independente que está sendo analisada, e $qnt\ vd$ refere-se a quantidade de dados disponíveis para treinamento e validação (dados de janeiro de 2004 até dezembro de 2017) da série dependente (consumo de energia elétrica no Paraná).

Critério g2 – Correlação com a variável dependente: esse critério busca realizar a correlação da variável “consumo de energia elétrica” com as variáveis exógenas (alternativas). Essa correlação implica na dependência ou associação estatística entre duas variáveis, com ou sem causalidade. O critério está mencionado na literatura como: correlação entre variáveis dependentes e independentes; análise

de causas; análise de cenários; seleção de variáveis explicativas. Esse critério tem natureza quantitativa e para a sua pontuação utilizou-se o coeficiente de correlação linear de Pearson (r).

Através desse coeficiente foi medido o grau de correlação entre o consumo de energia elétrica do Paraná em relação as demais alternativas, podendo assumir valores que variam entre 1 e -1. Quanto mais próximos os valores estiverem de 1 (correlação positiva) e -1 (correlação negativa) maior é a correlação entre as variáveis, valores próximos de zero indicam a não dependência linear entre as variáveis (JUDGE *et al.*, 1988).

Porém, Judge *et al.* (1988) atenta a casos em que as variáveis com uma alta correlação podem apresentar problemas relacionados a colinearidade, ou seja, altos coeficientes de correlação estimados com erros padrão muito elevados e baixos níveis de significância, mesmo que sejam conjuntamente significativos e com o grau de ajuste da regressão elevado. Quando a correlação entre as variáveis é quase perfeita, pode-se dizer que elas medem praticamente a mesma coisa.

Dessa forma, uma das medidas comumente utilizadas para identificar alta colinearidade entre duas variáveis é o próprio coeficiente de correlação (r), este que por sua vez considera que (r) maior que 0,80 ou 0,90 indica problema sério de colinearidade (JUDGE *et al.*, 1988). Por esse motivo, a nota máxima (100) foi atribuída para coeficientes com valores 0,8 e -0,8. As Equações 32 e 33 expressam a pontuação atribuída as alternativas, sendo que a Equação 32 é utilizada para (r) com valores positivos e a Equação 33 é utilizada para (r) com valores negativos.

$$d = \frac{r - 0}{0,8} \times 100 \quad (32)$$

$$d = \frac{r - 0}{-0,8} \times 100 \quad (33)$$

Sendo que essas Equações medem os percentuais das distâncias (d) dos valores dos coeficientes de correlação em relação ao valor zero, ou seja, quanto mais próximos de zero forem os resultados dos coeficientes de correlação, menor será a pontuação atribuída a essas alternativas, sendo que, os valores -0,8 e 0,8 apresentam

pontuação igual a 100. Na Figura 38 é exposto um exemplo a respeito da pontuação atribuída as alternativas.

Figura 38 - Exemplo da Escala utilizada no critério g2

Valor	Descrição
100	Relação linear positiva $r = 0,8$
	Relação linear negativa $r = -0,8$
Pontuação proporcional à distância	Relação linear positiva $0,8 > r > 0$
	Relação linear e negativa $-0,8 < r < 0$
0	Não correlação $r = 0$
	Relação linear positiva $r > 0,8$
	Relação linear negativa $r < -0,8$

Fonte: Autoria própria (2019).

Critério g3 – Verificação do padrão de consumo por setor: algumas variáveis podem ter uma influência sobre a demanda a depender de determinados contextos, variando de acordo com o setor. Na literatura, os critérios relacionados ao g3 são: setor industrial; comercial; residencial; rural, transporte e outros. Na série temporal de consumo de energia elétrica do Paraná é apresentada a porcentagem correspondente ao consumo de cada setor, a saber: 41,68% representa o consumo industrial; 23,72% o residencial; 18,36% o comercial; e 16,23% outros setores (EPE, 2019).

Como a própria EPE traz essa informação, a escala elaborada para esse critério baseou-se nessa porcentagem (Figura 39), ou seja, se os decisores considerarem que a variável independente que está sendo analisada tem influência no consumo de energia apenas dos setores industrial e comercial, por exemplo, a nota atribuída a essa variável corresponde a soma percentual de consumo desses setores: $(41,68\% + 18,36\% = 60,04\%)$.

Figura 39 - Exemplo da Escala utilizada no critério g3

Valor	Descrição
100%	Influencia todos os setores
83,77%	Influencia os setores residencial, industrial, comercial
81,64%	Influencia os setores residencial, industrial, outros
58,32%	Influencia os setores residencial, comercial, outros
76,28%	Influencia os setores industrial, comercial, outros
65,41%	Influencia os setores residencial e industrial
42,08%	Influencia os setores residencial e comercial
39,96%	Influencia os setores residencial e outros
60,04%	Influencia os setores industrial e comercial
57,92%	Influencia os setores industrial e outros
34,59%	Influencia os setores comercial e outros
23,72%	Influencia apenas o setor residencial
41,68%	Influencia apenas o setor industrial
18,36%	Influencia apenas o setor comercial
16,23%	Influencia outros setores
0%	Não Influencia nenhum setor

Fonte: Autoria própria (2019).

Esse critério é de natureza qualitativa, pois o valor atribuído para cada alternativa do contexto analisado, para que a tabela de decisão do ELECTRE I pudesse ser montada foi dado com base na percepção e experiência dos decisores envolvidos no processo. Os critérios g4 e g5 também são de natureza qualitativa, por não terem parâmetros para medição estes foram avaliados com base na escala ordinal de *Likert* de 5 pontos.

Critério g4 – Verificação do padrão de consumo por região: na literatura, os critérios relacionados ao g4 são: características geográficas, econômicas e sociais da região. Esse critério determina a influência que as variáveis explicativas podem apresentar em decorrência das características da região avaliada, uma mesma variável pode ter influência altíssima ou baixíssima em regiões diferentes devido as características dos locais serem distintas, na Figura 40 é apresentada a escala de *Likert* com a descrição para esse critério.

Figura 40 - Exemplo da Escala utilizada no critério g4

Valor	Descrição
5	Altíssima
4	Alta
3	Média
2	Baixa
1	Baixíssima

Fonte: Autoria própria (2019).

Critério g5 – Verificação do padrão de consumo por calendário: as variáveis que influenciam a demanda podem apresentar frequências diferentes a depender do calendário. Na literatura, os critérios relacionados ao g5 são: férias, dias da semana, finais de semana, feriados, outros. Nesse critério é determinado o quão recorrente ou não uma variável explicativa pode influenciar na demanda em decorrência do padrão de calendário. Na Figura 41 é apresentada a escala de *Likert* com a descrição para esse critério.

Figura 41 - Exemplo da Escala utilizada no critério g5

Valor	Descrição
5	Altíssima
4	Alta
3	Média
2	Baixa
1	Baixíssima

Fonte: Autoria própria (2019).

Com os critérios estabelecidos para a seleção das variáveis, o próximo passo foi aplicar o método AHP com o objetivo de identificar o grau de importância de cada critério que foi estabelecido a priori, essa aplicação é apresentada na próxima seção desse estudo.

5.2.1 Aplicação do AHP

Para identificar a relevância relativa de cada critério foi aplicado o método AHP já exposto na seção 2.3.1.1 desse estudo. Com base na escala de Saaty (1980), os dois decisores realizaram a comparação pareada referente aos 5 critérios. Com a análise par a par, os decisores responderam qual dos critérios era o mais importante e qual a intensidade de importância que um apresentava em relação ao outro (Figura 42).

Figura 42 - Avaliação dos critérios realizada pelos decisores utilizando a escala de Saaty (1980)

Comparação critérios		Valor	Descrição
g1	g2	1/4	Intermediário entre importância moderada e mais importante
g1	g3	3	Importância moderada
g1	g4	5	Mais importante
g1	g5	3	Importância moderada
g2	g3	5	Mais importante
g2	g4	7	Muito importante
g2	g5	5	Mais importante
g3	g4	2	Intermediário entre igual importância e importância moderada
g3	g5	1	Igual importância
g4	g5	1/2	Intermediário entre igual importância e importância moderada

Fonte: Autoria própria (2019).

A partir dessa comparação, foi possível obter a matriz de julgamento (Figura 43), cujo peso relativo ao julgamento dos decisores foi considerado igualitário, possuindo a mesma importância dentro do processo decisório. Tendo a construção dessa matriz, seus dados foram normalizados através da divisão entre o valor do julgamento obtivo para cada critério dividido pela soma total da sua respectiva coluna (Figura 44).

Tendo a Figura 44, realizou-se o somatório correspondente a avaliação de cada critério em relação aos correspondentes e com isso foi expresso os valores referentes aos vetores de preferência de cada critério, estes expostos por sua vez na Figura 45, deixando evidente que os critérios apontados pelos decisores como os mais importantes foram o grau de correlação com a variável dependente e a verificação da disponibilidade de dados.

Figura 43 – Matriz de avaliação do AHP

Critérios	g1	g2	g3	g4	g5
g1	1	1/4	3	5	3
g2	4	1	5	7	5
g3	1/3	1/5	1	2	1
g4	1/5	1/7	1/2	1	1/2
g5	1/3	1/5	1/1	2	1
Soma	4,9	0,8	9,5	16,0	9,5

Fonte: Autoria própria (2019).

Figura 44 – Pesos atribuídos aos critérios

Critérios	g1	g2	g3	g4	g5
g1	0,17	0,14	0,29	0,29	0,29
g2	0,68	0,56	0,48	0,41	0,48
g3	0,06	0,11	0,10	0,12	0,10
g4	0,03	0,08	0,05	0,06	0,05
g5	0,06	0,11	0,10	0,12	0,10

Fonte: Autoria própria (2019).

Figura 45 – Ranking da pontuação dos critérios

Ranking	Critério	Pontuação	Descrição
1º	g2	0,52	Correlação com a variável dependente
2º	g1	0,24	Verificação da disponibilidade de dados
3º	g3	0,10	Verificação de padrão de consumo por setor
4º	g5	0,10	Verificação de padrão de consumo por calendário
5º	g4	0,05	Verificação de padrão de consumo por região

Fonte: Autoria própria (2019).

Com os pesos dos critérios atribuídos, por fim, calculou-se a razão de consistência dos resultados para verificar a coerência nos julgamentos dos decisores. Como já mencionado, a condição de consistência dos julgamentos aceitável é $RC \leq 0,10$, limite esse definido por Saaty (2005), que afirma que quanto maior for a RC, maior será a inconsistência. O resultado do cálculo para esse problema foi de 0,0308, portanto seguiu-se para a etapa da seleção de variáveis pelo ELECTRE I.

5.2.2 Aplicação do ELECTRE I

Para que o ELECTRE I pudesse ser aplicado, uma matriz de avaliação foi construída com base nas escalas quantitativas já definidas anteriormente e nas escalas qualitativas, sendo que essas foram atribuídas de acordo com as premissas estabelecidas pelos decisores.

Lembrando que os dados de g2 precisaram passar pelas equações 32 e 33 para que a retirar condições de colinearidade e para que também os valores que representam uma correlação negativa tivessem uma atribuição positiva na avaliação. Fez-se também a normalização dos dados através da equação de benefício (Equação 34) para que as alternativas em relação a todos os critérios tivessem avaliações equivalentes.

$$Benefício = \frac{(V_{ij} - Min V_j)}{(Max V_j - Min V_j)} \quad (32)$$

em que V_{ij} representa o valor da alternativa que está sendo avaliada, $(Min V_j)$ é o mínimo valor das alternativas para o respectivo critério, $Max V_j$ é o máximo valor das alternativas para o critério analisado. Como resultado, tem-se a matriz de decisão normalizada (Figura 46).

Figura 46 – Ranking da pontuação dos critérios

Alternativa /critério	a1	a2	a3	a4	a5	a6	a7	a8	a9
g1	0,07	0,07	0,29	1,00	0,00	0,00	0,00	0,00	1,00
g2	0,00	0,00	0,00	0,48	0,00	0,00	0,00	0,00	0,56
g3	1,00	0,40	1,00	0,58	0,00	0,60	0,58	0,58	1,00
g4	0,75	0,50	1,00	0,75	0,25	0,50	1,00	1,00	1,00
g5	0,00	0,33	0,00	0,67	0,00	0,33	0,00	0,67	1,00
Alternativa /critério	a10	a11	a12	a13	a14	a15	a16	a17	a18
g1	1,00	1,00	1,00	0,78	1,00	1,00	0,92	0,72	0,08
g2	0,35	0,20	0,37	0,00	0,13	0,25	1,00	0,24	0,00
g3	0,58	0,42	0,42	0,00	0,00	0,00	0,00	0,42	0,60
g4	0,25	0,00	0,25	0,00	0,00	0,00	0,00	0,25	0,50
g5	0,33	0,00	0,33	0,00	0,00	0,00	0,00	0,33	0,33

Fonte: Autoria própria (2019).

E para que fosse estabelecido os resultados da aplicação do ELECTRE I, inicialmente calculou-se a matriz de concordância (Figura 47). Cada índice que compõe essa matriz, foi calculado como mostra o cálculo da alternativa 1 em relação a 2, considerando todos os critérios (Equação 33).

$$C_{1,2} = g_1 (se A_1 \geq A_2) + g_2 (se A_1 \geq A_2) + g_3 (se A_1 \geq A_2) + g_4 (se A_1 \geq A_2) + g_5 (se A_1 \geq A_2) / \Sigma g \quad (32)$$

Portanto, o índice de concordância onde são somados os pesos dos critérios nos quais a alternativa 'a1' supera a 'a2', por exemplo foi de $C_{1,2} = 0,24 (a1 = a2) + 0,52 (a1 = a2) + 0,1 (a1 > a2) + 0,05 (a1 > a2) + 0 (a1 < a2) / \Sigma g = 0,90$, e assim seguiu-se o cálculo na comparação par a par de todas as alternativas até que se estabelecesse a matriz.

Já para os índices de discordância, fez-se primeiro o cálculo das escalas de cada critério, que corresponde ao valor da diferença do maior valor atribuído a alternativa menos o menor valor. Dessa forma, a escala de todos os critérios resultou em 1,0. Portanto, para que a matriz de discordância (Figura 48) pudesse ser construída, comparando a alternativa 'a1' em relação a 'a2', por exemplo, fez-se:

$$D_{1,2} = Máx[g_1 \frac{(a2 - a1)}{1}; g_2 \frac{(a2 - a1)}{1}; g_3 \frac{(a2 - a1)}{1}; g_4 \frac{(a2 - a1)}{1}; g_5 \frac{(a2 - a1)}{1}] \quad (32)$$

Por fim, para que as alternativas pudessem serem classificadas, criou-se uma matriz de seleção. Para o cálculo dessa matriz, os decisores definiram os limiares de concordância (0,7) e discordância (0,3). Como exposto anteriormente na Equação 4, os índices da matriz de concordância que obedeceram à condição de maior igual ao limiar de concordância e os índices da matriz de discordância que obedeceram à condição de menor igual ao limiar de discordância foi atribuído o valor 1, e para os casos contrários o valor atribuído foi 0, como mostra a Figura 49.

Figura 47 – Matriz de concordância gerada pela comparação par a par de todas as alternativas

MATRIZ DE CONCORDÂNCIA																		
C	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18
a1		0,90	0,71	0,15	1,00	0,90	0,71	0,85	0,10	0,15	0,25	0,15	0,76	0,25	0,25	0,25	0,15	0,66
a2	0,85		0,61	0,00	1,00	0,90	0,61	0,75	0,00	0,15	0,15	0,15	0,76	0,25	0,25	0,25	0,15	0,66
a3	1,00	0,90		0,15	1,00	0,90	0,76	0,90	0,15	0,15	0,25	0,15	0,76	0,25	0,25	0,25	0,15	0,90
a4	0,90	1,00	0,85		1,00	0,90	0,95	0,95	0,24	1,00	1,00	1,00	1,00	1,00	1,00	0,49	1,00	0,90
a5	0,61	0,51	0,61	0,00		0,75	0,61	0,75	0,00	0,05	0,15	0,05	0,76	0,25	0,25	0,25	0,05	0,51
a6	0,61	0,76	0,61	0,10	1,00		0,71	0,85	0,00	0,25	0,25	0,25	0,76	0,25	0,25	0,25	0,25	0,76
a7	0,90	0,90	0,90	0,39	1,00	0,80		0,90	0,29	0,39	0,49	0,39	1,00	0,49	0,49	0,49	0,39	0,80
a8	0,66	0,76	0,66	0,25	1,00	0,90	0,76		0,05	0,25	0,25	0,25	0,76	0,25	0,25	0,25	0,25	0,66
a9	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00		1,00	1,00	1,00	1,00	1,00	1,00	0,49	1,00	1,00
a10	0,85	0,95	0,85	0,34	1,00	0,85	0,95	0,85	0,24		1,00	0,49	1,00	1,00	1,00	0,49	1,00	0,85
a11	0,85	0,85	0,85	0,24	0,95	0,75	0,85	0,75	0,24	0,24		0,34	1,00	1,00	0,49	0,49	0,34	0,75
a12	0,85	0,95	0,85	0,24	1,00	0,85	0,85	0,75	0,24	0,90	1,00		1,00	1,00	1,00	0,49	1,00	0,85
a13	0,85	0,75	0,85	0,00	0,95	0,75	0,61	0,75	0,00	0,00	0,15	0,00		0,25	0,25	0,25	0,24	0,75
a14	0,85	0,75	0,85	0,24	0,95	0,75	0,85	0,75	0,24	0,24	0,39	0,24	1,00		0,49	0,49	0,24	0,75
a15	0,85	0,75	0,85	0,24	0,95	0,75	0,85	0,75	0,24	0,24	0,90	0,24	1,00	1,00		0,49	0,75	0,75
a16	0,85	0,75	0,85	0,51	0,95	0,75	0,61	0,75	0,51	0,51	0,66	0,51	1,00	0,76	0,76		0,75	0,75
a17	0,85	0,95	0,85	0,00	1,00	0,85	0,61	0,75	0,00	0,15	0,76	0,25	0,76	0,76	0,25	0,25		0,85
a18	0,85	1,00	0,61	0,10	1,00	1,00	0,71	0,85	0,00	0,25	0,25	0,25	0,76	0,25	0,25	0,25	0,25	

Fonte: Autoria própria (2019).

Figura 48 – Matriz de discordância gerada pela comparação par a par de todas as alternativas

MATRIZ DE DISCORDÂNCIA																		
C	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18
a1		0,33	0,25	0,93	0,00	0,33	0,93	0,67	1,00	0,93	0,93	0,93	0,71	0,93	0,93	1,00	0,65	0,33
a2	0,60		0,60	0,93	0,00	0,20	0,93	0,50	0,93	0,93	0,93	0,93	0,71	0,93	0,93	1,00	0,65	0,20
a3	0,00	0,33		0,71	0,00	0,33	0,71	0,67	1,00	0,71	0,71	0,71	0,49	0,71	0,71	1,00	0,43	0,33
a4	0,42	-0,18	0,42		-0,48	0,02	0,25	0,25	0,42	0,00	0,00	0,00	-0,22	0,00	0,00	0,52	-0,16	0,02
a5	1,00	0,40	1,00	1,00		0,60	1,00	0,75	1,00	1,00	1,00	1,00	0,78	1,00	1,00	1,00	0,72	0,60
a6	0,40	0,07	0,50	1,00	0,00		1,00	0,50	1,00	1,00	1,00	1,00	0,78	1,00	1,00	1,00	0,72	0,08
a7	0,42	0,33	0,42	0,67	0,00	0,33		0,67	1,00	0,37	0,20	0,37	0,00	0,13	0,25	1,00	0,33	0,33
a8	0,42	0,07	0,42	1,00	0,00	0,02	1,00		1,00	1,00	1,00	1,00	0,78	1,00	1,00	1,00	0,72	0,08
a9	0,00	-0,50	0,00	0,00	-0,56	-0,40	0,00	0,00		0,00	0,00	0,00	-0,22	0,00	0,00	0,44	-0,28	-0,40
a10	0,50	0,25	0,75	0,50	0,00	0,25	0,75	0,75	0,75		0,00	0,02	-0,22	0,00	0,00	0,65	0,00	0,25
a11	0,75	0,50	1,00	0,75	0,25	0,50	1,00	1,00	1,00	0,33		0,33	0,00	0,00	0,05	0,80	0,33	0,50
a12	0,58	0,25	0,75	0,50	0,00	0,25	0,75	0,75	0,75	0,16	0,00		-0,22	0,00	0,00	0,63	0,00	0,25
a13	1,00	0,50	1,00	0,75	0,25	0,60	1,00	1,00	1,00	0,58	0,42	0,42		0,22	0,25	1,00	0,42	0,60
a14	1,00	0,50	1,00	0,75	0,25	0,60	1,00	1,00	1,00	0,58	0,42	0,42	0,00		0,12	0,87	0,42	0,60
a15	1,00	0,50	1,00	0,75	0,25	0,60	1,00	1,00	1,00	0,58	0,42	0,42	0,00	0,00		0,75	0,42	0,60
a16	1,00	0,50	1,00	0,75	0,25	0,60	1,00	1,00	1,00	0,58	0,42	0,42	0,00	0,08	0,08		0,42	0,60
a17	0,58	0,25	0,75	0,50	0,00	0,25	0,75	0,75	0,75	0,28	0,28	0,28	0,06	0,28	0,28	0,76		0,25
a18	0,40	0,00	0,50	0,92	0,00	0,00	0,92	0,50	0,92	0,92	0,92	0,92	0,70	0,92	0,92	1,00	0,64	

Fonte: Autoria própria (2019).

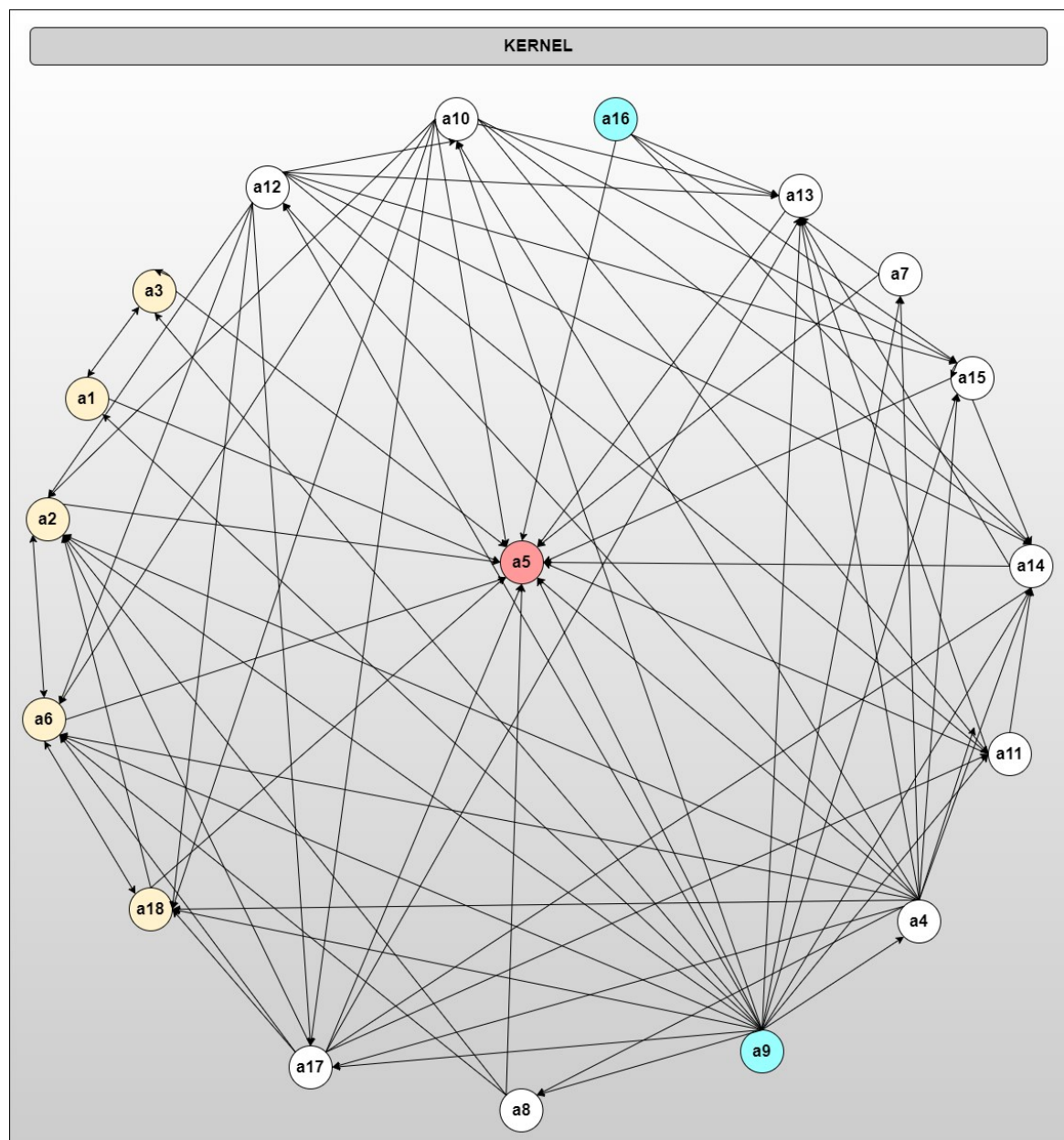
Figura 49 – Matriz de seleção gerada pela condição definida pelos limiares de concordância e discordância em comparação com a matriz de concordância e discordância

MATRIZ DE SELEÇÃO																		
C	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18
a1		0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
a2	0		0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
a3	1	0		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
a4	0	1	0		1	1	1	1	0	1	1	1	1	1	1	0	1	1
a5	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0
a6	0	1	0	0	1		0	0	0	0	0	0	0	0	0	0	0	1
a7	0	0	0	0	1	0		0	0	0	0	0	1	0	0	0	0	0
a8	0	1	0	0	1	1	0		0	0	0	0	0	0	0	0	0	0
a9	1	1	1	1	1	1	1	1		1	1	1	1	1	1	0	1	1
a10	0	1	0	0	1	1	0	0	0		1	0	1	1	1	0	1	1
a11	0	0	0	0	1	0	0	0	0	0		0	1	1	0	0	0	0
a12	0	1	0	0	1	1	0	0	0	1	1		1	1	1	0	1	1
a13	0	0	0	0	1	0	0	0	0	0	0	0		0	0	0	0	0
a14	0	0	0	0	1	0	0	0	0	0	0	0	1		0	0	0	0
a15	0	0	0	0	1	0	0	0	0	0	0	0	1	1		0	0	0
a16	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1		0	0
a17	0	1	0	0	1	1	0	0	0	0	1	0	1	1	0	0		1
a18	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	

Fonte: Autoria própria (2019).

Como resultado, as alternativas que foram selecionadas como variáveis independentes para compor a previsão, segundo os critérios estabelecidos foram a9 e a16 que corresponde respectivamente a temperatura e evaporação. Para facilitar a visualização desse conjunto de *kernel*, criou-se um grafo dos resultados (Figura 50).

Figura 50 – Grafo que representa o conjunto de Kernel das variáveis selecionadas para compor a previsão de demanda de energia elétrica



Fonte: Autoria própria (2019).

Pela análise visual do *kernel*, só podem fazer parte do subconjunto final as alternativas que não são sobreclassificadas por nenhuma alternativa, como as

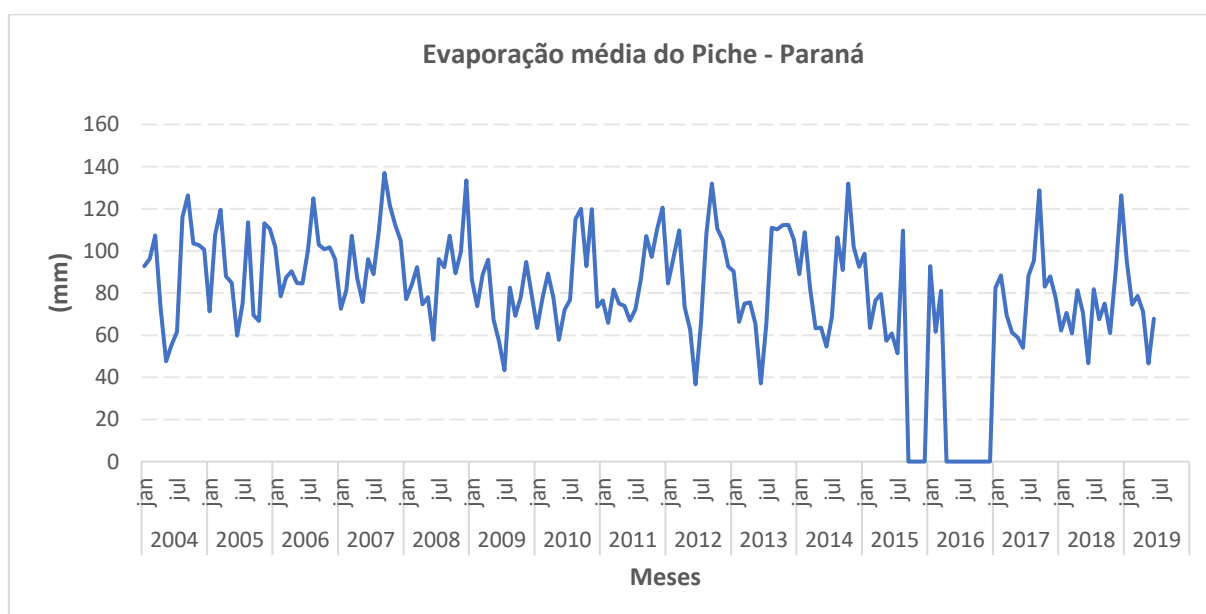
variáveis independentes a9 e a16 e as alternativas que não são sobreclassificadas pelas alternativas escolhidas anteriormente, porém todas as alternativas restantes não satisfazem essa condição.

Outra característica desse conjunto é composta pelas alternativas que sobreclassificam umas as outras resultando em um empate (a1 e a3; a2 e a6; a6 e a18), isso implica na indiferença entre as alternativas. E a alternativa a5 foi a única que não sobreclassificou nenhuma, notável também que não há casos de incomparabilidade onde uma alternativa não sobreclassifica nenhuma outra e nem é sobreclassificada.

5.3 PREVISÃO COM VARIÁVEIS INDEPENDENTES

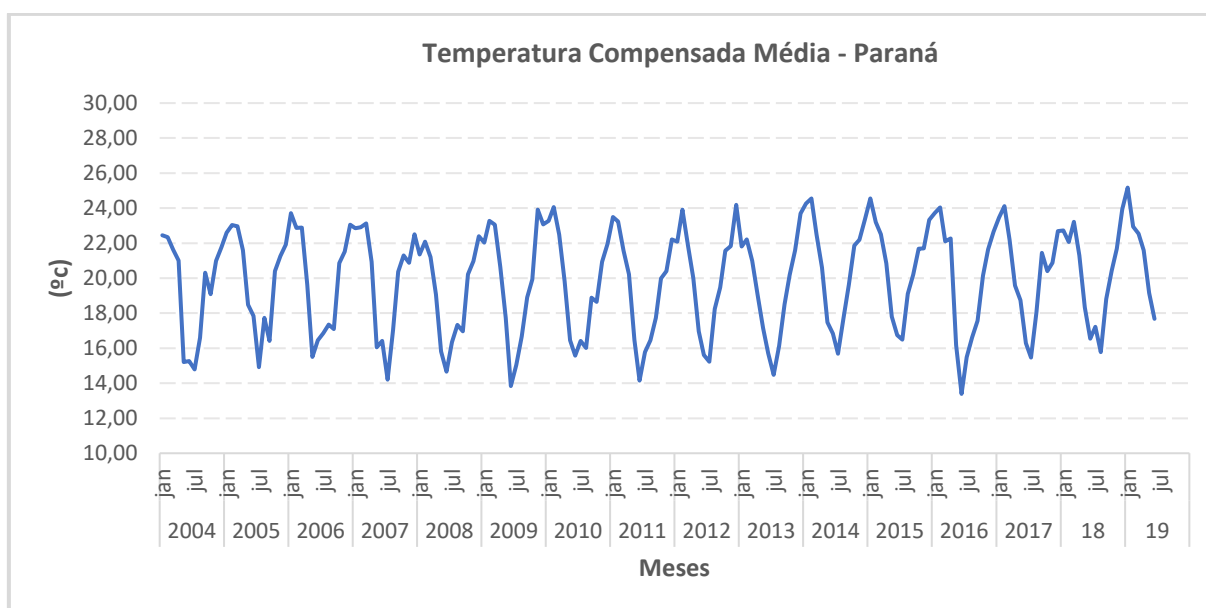
Com as variáveis independentes já selecionadas na seção anterior, coletou-se os dados dela para identificar sua regularidade. Para isso, plotou-se os gráficos dessas séries temporais de evaporação e temperatura em um gráfico (Figura 51 e 52). Por essa análise gráfica, é possível verificar a presença de sazonalidade em ambas as séries e ausência de tendência.

Figura 51 – Série temporal de evaporação referente ao período de janeiro de 2004 até junho de 2019



Fonte: Instituto Nacional de Meteorologia – INMET (2019).

Figura 52 – Série temporal de temperatura referente ao período de janeiro de 2004 até junho de 2019



Fonte: Instituto Nacional de Meteorologia – INMET (2019).

Diante dessa análise, objetivando confirmar ou não a ausência de tendência e presença de sazonalidade, aplicou-se os testes estatísticos não-paramétricos de Cox-Stuart e Friedman (Tabela 8).

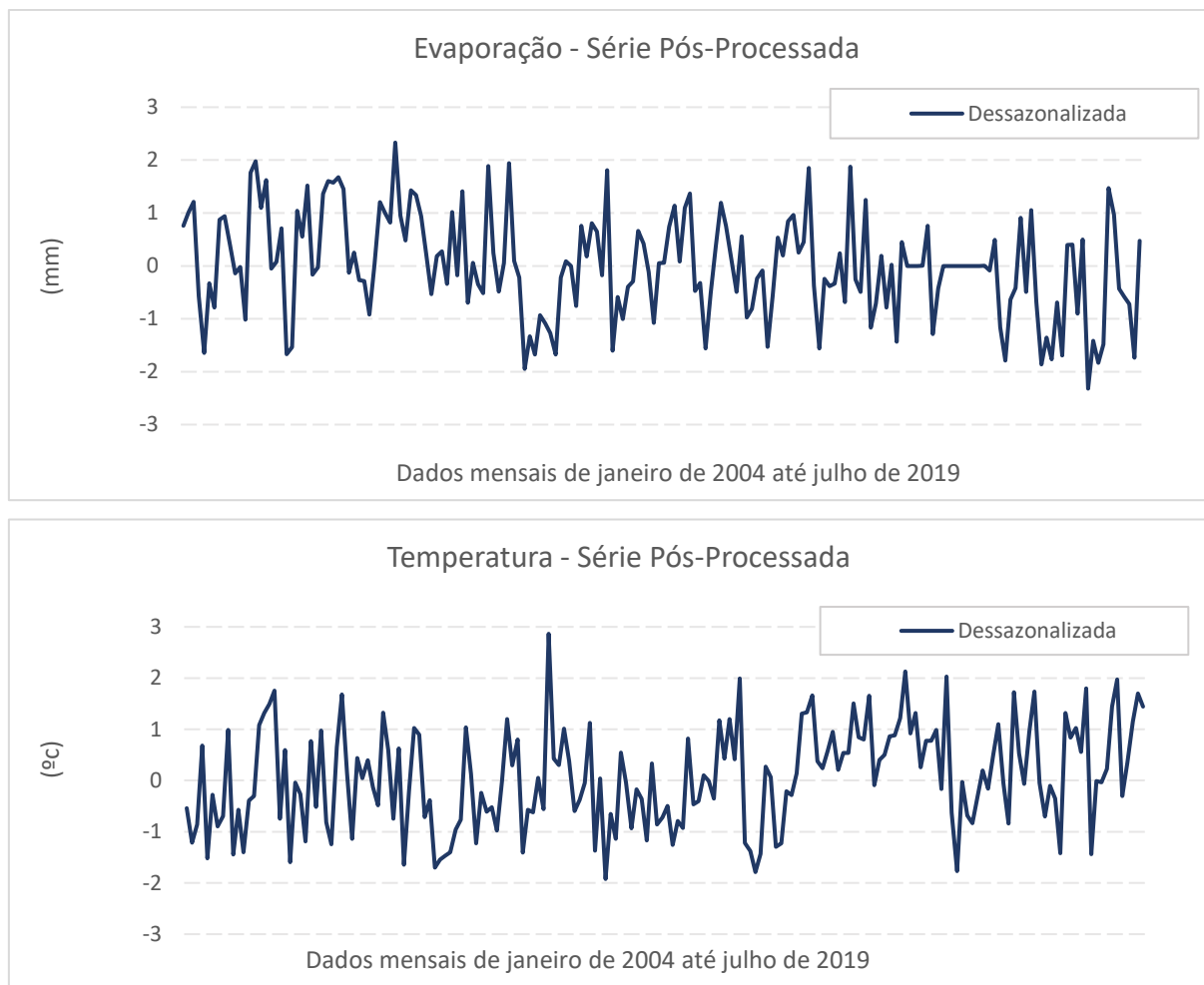
Tabela 8 – Testes não-paramétricos para análise de tendência e sazonalidade

Testes	Cox-Stuart <i>P-Valor</i> > 5%	Friedman <i>P-Valor</i> < 5%
Evaporação	0,146177029	7,58419E-08
Temperatura	1	1,37641E-26

Fonte: Autoria própria (2019).

Tendo os resultados expostos na Tabela 8, foi possível verificar que o *P-valor* para o teste de tendência de ambas as séries analisadas foi maior que 5%, indicando que não há tendência nas séries, já o *P-valor* para o teste de sazonalidade foi menor que 5%, portanto há sazonalidade em ambas as séries. Portanto, fez-se a retirada da sazonalidade para que a série se tornasse estacionária e os *P-valores* do teste de Friedman para evaporação e temperatura foram de respectivamente 0,997947317 e 0,999969254. O gráfico dessazonalizado de ambas as séries está exposto na Figura 53.

Figura 53 – Dados da evaporação e temperatura média do Estado do Paraná referente ao período de janeiro de 2004 até junho de 2019, dessazonalizados



Fonte: Autoria própria (2019).

Observa-se que os valores da série estacionária variam entre 3 e -3, oscilando em torno de zero indicando a condição de estacionariedade nos dados. Importante mencionar, que como não havia tendência na série de evaporação, os dados faltantes do período de setembro a dezembro de 2015 e abril à dezembro de 2016 foram substituído pelos respectivos valores médios correspondente aos meses analisados.

Tendo a série analisada, o próximo passo foi definir os parâmetros para previsão das RNAs com as variáveis independentes. Lembrando que a série temporal de consumo de energia elétrica também entrará na previsão, mas seu pré-processamento já foi realizado na seção 5.1.1. Outro ponto a ser lembrado é que como são 3 séries temporais para realizar a previsão a divisão do horizonte de treinamento (dados entre o período de janeiro de 2004 até dezembro de 2015), validação (dados entre o período de janeiro de 2016 até dezembro de 2017) e teste

(dados entre o período de janeiro de 2018 até junho de 2019) permaneceu a mesma, porém a quantidade de dados foi triplicada.

Como na seção 5.1.1, as RNAs também tiveram as quantidades de neurônios na camada intermediária variando de 5 em 5, até 200. Para cada variação de neurônios aplicou-se o método *wrapper* que fez uma varredura progressivamente considerando até 6 atrasos para cada série temporal. Dessa forma, as configurações das RNAs que obtiveram os melhores resultados de previsão foram:

MLP de 35 neurônios na camada intermediária e com o primeiro atraso de energia, sexto de temperatura e quarto de evaporação como entrada para previsão; RBF com 50 neurônios e atraso 1 energia, 6 temperatura e 4 evaporação; ELM com 45 neurônios e atrasos 1 e 6 de energia, 3 evaporação e 6 temperatura.

Com os parâmetros dos modelos definidos, fez-se a previsão de demanda por energia elétrica no Paraná para 1 passos à frente, referentes aos 18 dados (meses dos anos de 2018 e 2019) de teste. Na Tabela 9 está exposto os resultados obtidos das previsões, comparado aos dados separados para teste pelas medidas: Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE) e Erro Percentual Absoluto Médio (MAPE). Neles, os valores encontrados nas medidas de erro são referentes a melhor das 30 simulações realizadas para cada modelo. Já a tabela 10, traz o ranking do desempenho dos modelos.

Tabela 9 – Resultados das previsões para 1 passo à frente, comparado pelas medidas de erro MSE, MAE e MAPE

1 passo à frente						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
MLP	neurônios = 35, atrasos = 1, 6, 4	5854808140,90	0,2363	62140,20	0,4015	2,3501
RBF	neurônios = 50, atrasos = 2, 6, 6	3900550265,86	0,2142	50749,56	0,3643	1,9354
ELM	neurônios = 45, atraso = 1, 6, 3, 6	4154631401,10	0,2180	51866,49	0,3673	1,969425

Fonte: Autoria própria (2019).

Tabela 10 – Ranking de desempenho das previsões para 1 passo à frente

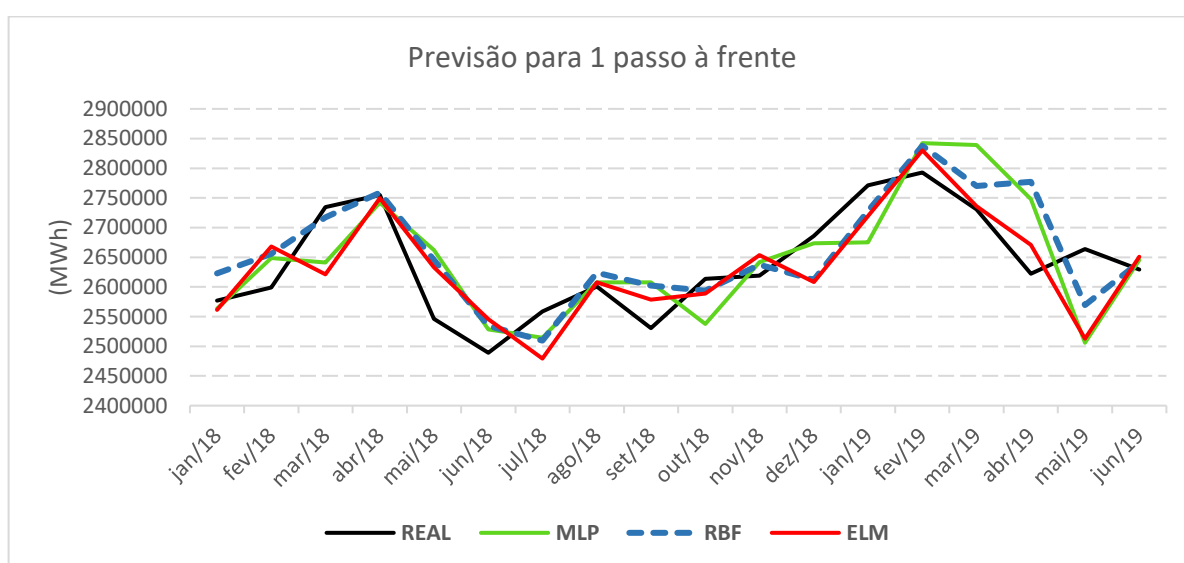
Ranking de desempenho						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
MLP	neurônios = 35, atrasos = 1, 6, 4	3,00	3,00	3,00	3,00	3,00
RBF	neurônios = 50, atrasos = 2, 6, 6	1,00	1,00	1,00	1,00	1,00
ELM	neurônios = 45, atraso = 1, 6, 3, 6	2,00	2,00	2,00	2,00	2,00

Fonte: Autoria própria (2019).

Os resultados da Tabela 2 e 3 apresenta que a melhor previsão realizada foi obtida pela RNA RBF, seguida da ELM para todas as medidas de erro analisadas, enquanto o pior resultado de previsão foi obtido pela rede MLP.

Considerando o MSE no domínio real, o erro do segundo colocado em relação ao primeiro em termos percentual foi de 6,51%, e realizado a mesma comparação do terceiro em relação ao primeiro foi de 50,10%. A Figura 54 apresenta as previsões realizadas pelos modelos e os dados reais separados para teste.

Figura 54 – Gráfico da comparação dos dados reais com os dados previstos para 1 passo à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando RNAs MLP, RBF e ELM



Fonte: Autoria própria (2019).

Dando sequência, fez-se também previsões para a mesma série considerando o horizonte de 3 passos à frente, tendo seus resultados comparados pelas medidas MSE, MSE(d), MAE, MAE(d) e MAPE (Tabela 11) e fazendo o Ranking desses desempenhos (Tabela 12).

Tabela 11 – Resultados das previsões para 3 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE

3 passos à frente						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
MLP	neurônios = 35, atrasos = 1, 6, 4	6248871037,89	0,2823	66536,95	0,4368	2,5146
RBF	neurônios = 50, atrasos = 2, 6, 6	6205419571,17	0,2761	67111,50	0,4544	2,5330
ELM	neurônios = 45, atraso = 1, 6, 3, 6	5536080184,82	0,2594	58071,25	0,4042	2,2150

Fonte: Autoria própria (2019).

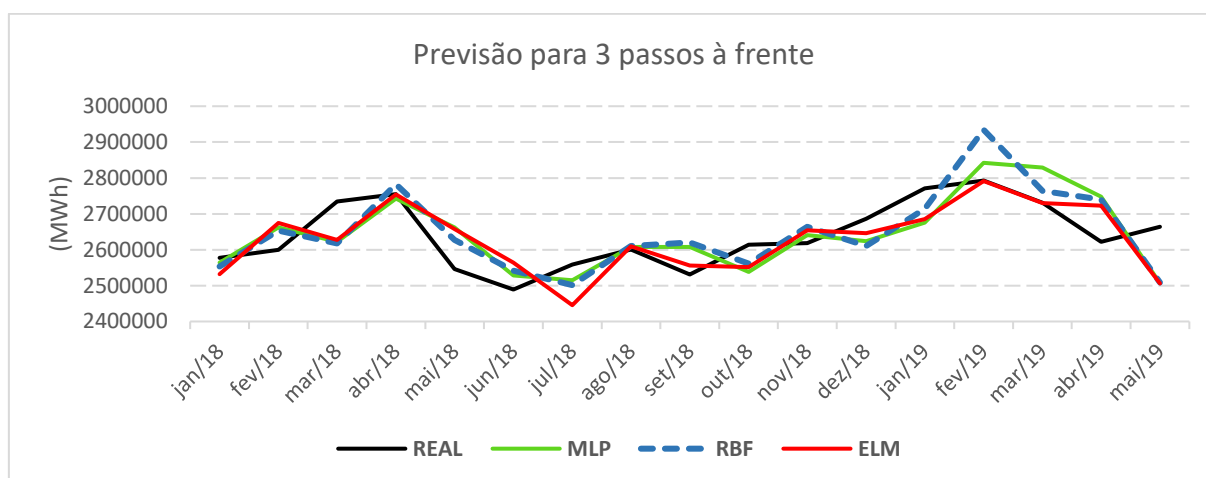
Tabela 12 – Ranking de desempenho das previsões para 3 passos à frente

Ranking de desempenho						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
MLP	neurônios = 35, atrasos = 1, 6, 4	3,00	3,00	2,00	2,00	2,00
RBF	neurônios = 50, atrasos = 2, 6, 6	2,00	2,00	3,00	3,00	3,00
ELM	neurônios = 45, atraso = 1, 6, 3, 6	1,00	1,00	1,00	1,00	1,00

Fonte: Autoria própria (2019).

Se comparado o MSE das Tabelas 11 e 12 a rede ELM teve o desempenho superior em relação aos outros modelos, a rede RBF ocupou a segunda posição, mas o posicionamento entre a RBF e MLP é invertido quando comparado as outras medidas de erro. Lembrando, que como já comentado anteriormente a medida de erro MSE penaliza mais fortemente erros maiores e por isso pode ter havido a inversão nos rankings quando se analisa por outras medidas de erro. A Figura 55 apresenta as previsões realizadas pelos modelos e os dados reais separados para teste.

Figura 55 – Gráfico da comparação dos dados reais com os dados previstos para 3 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando RNAs MLP, RBF e ELM



Fonte: Autoria própria (2019).

Se comparado o MSE no domínio real, o erro da rede RBF em relação a ELM para termos em percentual foi de 12,09%, já a MLP em relação a ELM teve um erro acrescido de 12,88%. E por fim, fez-se também previsões para a mesma série considerando o horizonte de 6 passos à frente, tendo seus resultados comparados pelas medidas MSE, MSE(d), MAE, MAE(d) e MAPE (Tabela 13) e fazendo o Ranking desses desempenhos (Tabela 14). A Figura 56 traz as previsões realizadas pelos modelos e os dados reais separados para teste.

Tabela 13 – Resultados das previsões para 6 passos à frente, comparado pelas medidas de erro MSE, MAE e MAPE

6 passos à frente						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
MLP	neurônios = 35, atrasos = 1, 6, 4	6433456508,91	0,3195	68395,90	0,4847	2,5895
RBF	neurônios = 50, atrasos = 2, 6, 6	6358076170,74	0,3180	68783,06	0,4873	2,6132
ELM	neurônios = 45, atraso = 1, 6, 3, 6	6049034166,40	0,3165	70008,28	0,4886	2,6619

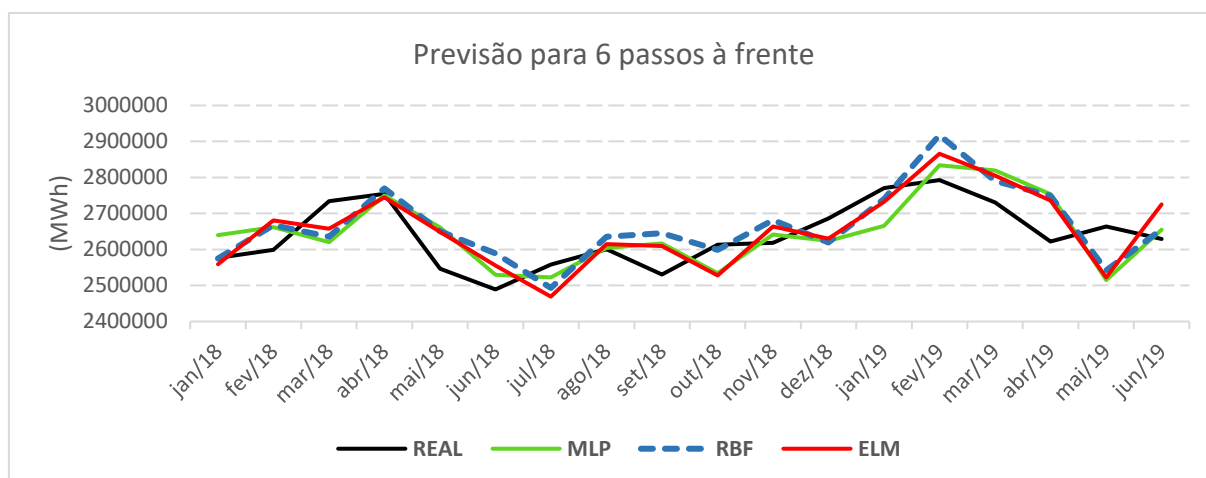
Fonte: Autoria própria (2019).

Tabela 14 – Ranking de desempenho das previsões para 6 passos à frente

Ranking de desempenho						
Modelo	Parâmetros do modelo	MSE	MSE (d)	MAE	MAE (d)	MAPE
MLP	neurônios = 35, atrasos = 1, 6, 4	3,00	3,00	1,00	1,00	1,00
RBF	neurônios = 50, atrasos = 2, 6, 6	2,00	2,00	2,00	2,00	2,00
ELM	neurônios = 45, atraso = 1, 6, 3, 6	1,00	1,00	3,00	3,00	3,00

Fonte: Autoria própria (2019).

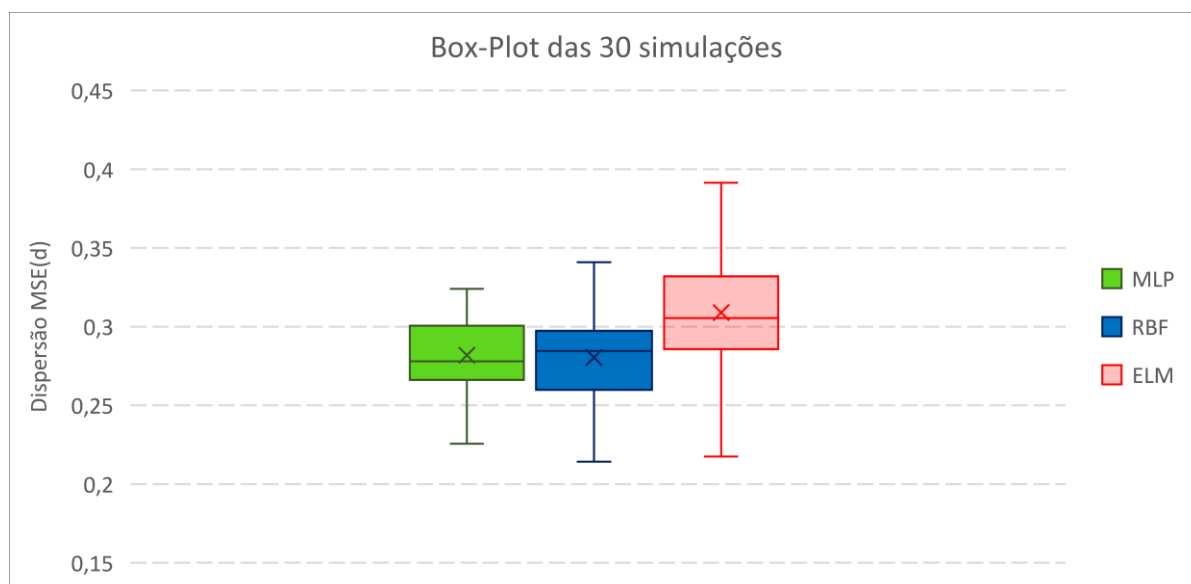
Figura 56 – Gráfico da comparação dos dados reais com os dados previstos para 6 passos à frente referentes aos 18 meses dos anos de 2018 e 2019, utilizando RNAs MLP, RBF e ELM



Fonte: Autoria própria (2019).

Novamente a rede ELM obteve um desempenho superior de previsão para todas as medidas analisadas. Percebe-se que as redes ELM e RBF mantiveram a posição no ranking com os melhores desempenhos de previsões para todos os passos à frente. Quando se compara o MSE do domínio real da rede ELM em relação aos modelos RBF e MLP, em termos percentuais o desempenho é reduzido respectivamente em 7,44%, 9,25%. Com o objetivo de verificar a dispersão das 30 simulações, fez-se também o apresenta o gráfico de Box-plot para todos os modelos (Figura 57).

Figura 57 – Gráfico Box-Plot com a dispersão do MSE(d) para as 30 simulações das RNAs



Fonte: Autoria própria (2019).

Observa-se no gráfico 57 que o modelo preditivo que apresentou maior dispersão em seus resultados foi a rede ELM, seguido da RBF e MLP. Em relação a assimetria da distribuição, percebe-se que a rede ELM tem uma distribuição praticamente uniforme, pois a linha da mediana além de estar próxima centralizada, está muito perto da média, porém com uma leve tendência para o quartil inferior, indicando que pode haver dados discrepantes com valores maiores.

A rede MLP também apresenta dados discrepantes maiores, pois a mediana está próxima do quartil inferior e tem valor menor do que a média das amostras, já a rede RBF apresenta discrepância de dados menores pois a mediana encontra-se próxima ao quartil superior e com valor maior que a média. Nenhuma das redes apresentou outlier em seus resultados.

5.4 COMPARAÇÃO GERAL DAS PREVISÕES

Com o objetivo de realizar uma análise geral, o desempenho de todos os modelos considerando as previsões para todos os passos à frente realizadas está exposto nas Tabelas 15, 16 e 17.

Tabela 15 – Ranking de desempenho geral das previsões para 1 passo à frente

Ranking dos resultados de MSE - 1 passo à frente		
Modelo	MSE	Posição
RBF - var. indep.	3900550266	1
ELM - var. indep.	4154631401	2
ELM	4599769468	3
RBF	4634913072	4
AR	5247636250	5
SEHW	5679573387	6
MLP - var. indep.	5854808141	7
ARIMA	5860003522	8
MLP	6557391893	9

Fonte: Autoria própria (2019).

Tabela 16 – Ranking de desempenho geral das previsões para 3 passos à frente

Ranking dos resultados de MSE - 3 passos à frente		
Modelo	MSE	Posição
ELM - var. indep.	5536080185	1
RBF - var. indep.	6205419571	2
MLP - var. indep.	6248871038	3
ELM	6366837642	4
RBF	6528243518	5
MLP	6573580787	6
AR	7013687861	7
SEHW	7485077441	8
ARIMA	8924931365	9

Fonte: Autoria própria (2019).

Tabela 17 – Ranking de desempenho geral das previsões para 6 passos à frente

Ranking dos resultados de MSE - 6 passos à frente		
Modelo	MSE	Posição
ELM - var. indep.	6049034166	1
RBF - var. indep.	6358076171	2
ELM	6406990346	3
MLP - var. indep.	6433456509	4
RBF	6595147325	5
MLP	6686402961	6
AR	6840287880	7
ARIMA	8174710719	8
SEHW	8283400542	9

Fonte: Autoria própria (2019).

Em uma análise inicial, considerando apenas as previsões de demanda de energia elétrica que não incluem as variáveis explicativas, é possível perceber que as previsões para um passo à frente de todos os modelos obtiveram os melhores

desempenhos de previsão, com exceção da rede MLP. Porém, mesmo que a rede MLP não tenha tido um desempenho tão bom em relação aos outros modelos, a medida em que é aumentado os passos à frente esse Ranking é invertido. O desempenho da rede MLP para 6 passos à frente, por exemplo, é superior ao desempenho dos modelos AR, ARIMA e SEHW para 3 e 6 passos à frente.

A rede ELM teve desempenho superior em todos os passos à frente se comparado aos outros modelos. No desempenho geral, suas previsões até mesmo para os próximos 6 meses superou o desempenho de todos os outros modelos para 3 e 6 passos à frente. Importante notar, que a medida em que os modelos MLP, ELM, RBF e SEHW aumentam seu horizonte de previsão o seu desempenho foi reduzido, diferente dos modelos AR e ARIMA que tiveram o desempenho da previsão para os próximos 6 meses superior ao de 3 passos à frente.

Em uma análise geral, como esperado, os modelos lineares se comparado as RNAs obtiveram resultados inferiores de previsão. Conforme Silva *et al.* (2018a) as RNAs por terem uma grande capacidade de generalização e mapeamento, ajusta-se bem seus dados previstos aos valores reais. Outro ponto a ser comentado, é que a rede ELM que obteve desempenho superior de previsão, possui custo computacional mais baixo e desempenho superior no tempo de processamento em relação as outras redes.

Isso porque sua camada intermediária permanece sem ajuste durante o treinamento, e a camada de saída tem seu treinamento baseado em um método determinístico. Importante mencionar que os modelos lineares, mesmo que para esse caso não tenham tido resultados tão bons em relação aos outros modelos, são mais simples e rápidos para serem implementados e executados, podendo em outros casos apresentar resultados equivalentes de previsão e custos mais vantajosos.

Depois dessa análise, percebe-se que, quando é incluído ao problema as variáveis explicativas, todas as RNAs tiveram seus desempenhos preditivos melhorados para todos os passos à frente, sendo que no Ranking geral as redes RBF e ELM com variáveis explicativas ocuparam respectivamente a primeira e segunda posição, deixando evidências de que a utilização do ELECTRE I para seleção de variáveis foi eficiente na melhoria do desempenho geral das previsões.

6 CONCLUSÕES

A previsão de demanda de energia elétrica é uma etapa importante para o planejamento energético de um país. É importante que se tenha um planejamento adequado para que não ocorram desperdícios ou escassez desse recurso. Na literatura, diferentes autores buscaram resolver esse problema com modelos de previsão lineares, também acrescentando outras variáveis independentes de previsão.

Porém as variáveis que são selecionadas como explicativas muitas vezes resultam de apenas uma análise de correlação ou até mesmo o próprio conhecimento dos autores, o que não garante que as variáveis que realmente explicam a demanda sejam selecionadas, já que diferentes critérios estão relacionados nesse processo e correlação não implica necessariamente na causa.

Para que as variáveis selecionadas para previsão tenham efetividade no processo preditivo é importante reconhecer os critérios de decisão que estão envolvidos nesse processo. A insuficiência de dados, a correlação com a energia, o padrão de consumo por setor, calendário e região foram critérios considerados como os mais importantes para resolução desse problema, lembrando que estes foram validados pelos decisores.

Esse problema de previsão, envolveu possível 18 alternativas listadas na literatura e 5 critérios com grau de importância diferentes no processo. Como é um problema relativamente grande, a abordagem multicritério apresenta vantagens para uma tomada de decisão assertiva, especialmente utilizando o método ELECTRE I para selecionar as variáveis que tem influências sobre a demanda de energia elétrica.

Lembrando, que até o momento não havia indícios na literatura de métodos multicritérios utilizados para esse propósito. Tendo isso em vista, esse trabalho realizou a previsão de demanda de energia elétrica para o Paraná, considerando inicialmente apenas a série temporal de consumo mensal de energia elétrica, utilizando os modelos preditivos lineares SEHW, AR e ARIMA e as RNAs MLP, RBF e ELM.

Posteriormente, aplicou-se o ELECTRE I para as alternativas e critérios expostos na literatura, possibilitando selecionar as alternativas temperatura e evaporação como possíveis influências na demanda de energia elétrica no Paraná.

Com a seleção dessas variáveis, fez-se a previsão para a demanda mensal de energia elétrica no Paraná, considerando também a temperatura e evaporação. Essas previsões foram realizadas pelas RNAs MLP, RBF e ELM, para o período entre janeiro de 2018 e junho de 2019. Como apresentou 18 dados disponíveis para teste, o horizonte considerado para previsão foi de 1 (próximo mês), 3 e 6 passos à frente.

Como resultados, quando comparado as previsões entre os modelos lineares e RNAs considerando apenas a série temporal de consumo de energia elétrica, as RNAs obtiveram desempenhos superiores em uma análise geral de todos os horizontes de previsão, com exceção da rede MLP para 1 passo à frente, que teve o pior desempenho nesse respectivo horizonte de previsão.

Porém, a medida em que se aumentou os horizontes para 3 e 6 passos à frente, essa rede, mesmo com seu desempenho sendo reduzido, foi superior aos modelos lineares para os horizontes 3 e 6 de previsão. O desempenho superior das RNAs em relação aos modelos lineares já era esperado, visto que as RNAs por terem essa não linearidade conseguem generalizar e mapear séries temporais que possuem uma maior variabilidade nos dados.

Talvez, se a série temporal em questão apresentasse uma certa estacionariedade os modelos lineares teriam resultados suficientes ou até melhores para a resolução do problema, visto que seu tempo operacional e custo são menores em relação as RNAs. Outro ponto a ser comentado, é que as redes RBF e ELM que obtiveram resultados superiores em relação a rede MLP, possuem custo computacional mais baixo em relação a MLP.

A primeira, em decorrência da sua arquitetura, e a segunda porque além de ter uma arquitetura mais simplificada tem também um processamento mais rápido já que a camada de saída tem seu treinamento baseado em um método determinístico. Depois de realizar as previsões considerando apenas a série temporal de consumo de energia elétrica no Paraná, acrescentou-se como entradas para os modelos preditivos dados das variáveis selecionadas pelo ELECTRE I.

Com o acréscimo desses dados, verificou-se, sem exceções, a melhora no desempenho de todas as RNAs considerando todos os passos à frente, sendo que as redes RBF e ELM ocuparam a primeira e segunda posição considerando o desempenho geral de todos os modelos testados para todos os horizontes de previsão. Porém a arquitetura das redes considerando essas variáveis foram compostas por mais neurônios e atrasos.

Importante lembrar que apesar de ter 18 alternativas no processo apenas duas foram selecionadas como explicativas da demanda, quando considerado os 5 critérios e seus graus de importância para a avaliação. Assim, nem sempre tem dados disponíveis para todas as alternativas, ou estas estão correlacionadas com a demanda de energia, ou apresentam ainda influências sobre o padrão de consumo por setor, calendário e região.

Portanto, apesar de ter apresentado resultados positivos para a aplicação no Paraná, essas mesmas variáveis se aplicadas a outros contextos poderiam resultar em resultados diferentes de previsão, pois além das características dos dados, cada local de aplicação tem outras características distintas que precisariam ser consideradas por especialistas que entendem sobre demanda de energia elétrica e que conhecem o local de aplicação.

TRABALHOS FUTURO

O modelo apresentado nesse estudo pode ser aprimorado e empregado em aplicações similares, como também expandir para outros estados do Brasil ou até mesmo outros contextos de aplicações. Em relação a seleção de variáveis, ampliar e aprimorar estudos sobre outros critérios a serem considerados no processo, bem como os pesos definidos para cada critério, também é uma proposta para trabalhos futuros.

Quanto ao método de seleção de entradas das RNAs nesse trabalho aplicou-se apenas o método *Wrapper*, porém, ainda pode ser testadas outras metodologias como por exemplo filtros e *Stepwise*. Poderá também ser realizado previsões considerando 12 passos à frente e entradas sendo testadas para até 12 atrasos. Além disso, podem ser utilizadas outras arquiteturas de redes e algoritmos de treinamento.

Na rede MLP por exemplo, utilizou-se para treinamento o algoritmo de retropropagação de erro, mas na literatura, reconhece que essa não é a única forma de treinar a rede, portanto, outros algoritmos podem ser abordados ou mesmo metaheurísticas podem ser utilizadas. E por fim, redes neurais não clássicas como combinador de um Ensemble também surgem como oportunidades a serem trabalhadas.

REFERÊNCIAS

ADAM, N. R. B.; ELAHEE, M. K.; DAUHOO, M. Z. Forecasting of peak electricity demand in Mauritius using the non-homogeneous Gompertz diffusion process. **Energy**. v. 36, p. 6763-6769, 2011.

ALMEIDA, A.T.; ALMEIDA, J.A.; COSTA, A.P.C.S.; ALMEIDA-FILHO, A.T. A new method for elicitation of criteria weights in additive models: Flexible and interactive tradeoff. **European Journal of Operational Research**, v. 250, p 179–191, 2016.

ALMEIDA, A. T. **O conhecimento e o uso de métodos multicritério de apoio a decisão**. Recife: Editora Universitária da UFPE, 2011.

ALMEIDA, J. A. **Modelo de multicritério para seleção de portfólio de projetos de sistemas de informação**. 2012. 115 f. Tese (Doutorado) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Pernambuco. Recife, 2012.

ANOCHI, J. A. **Previsão climática de precipitação por redes neurais autoconfiguradas**. 2015. Tese (Doutorado). Curso de Pós-Graduação em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais – INPE. São José dos Campos, 2015.

AYYANATHAN, N.; KANNAMMAL, A. Share price time series forecasting for effective supply chain information exchange. **International Journal of Logistics Systems and Management**. v. 18, p. 139-158, 2014.

BALLINI, R. **Análise e previsões de vazões utilizando modelos de series temporais, redes neurais e redes neurais nebulosas**. 2000. Tese (Doutorado). Programa de Pós-Graduação em Engenharia Elétrica, Universidade Estadual de Campinas, Campinas, 2000.

BELOTTI, J. T. **Previsão de vazões afluentes utilizando redes neurais artificiais e ensembles**. 2019. Dissertação (Mestrado). Programa de Pós-Graduação em Ciência da Computação, Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2019.

BELTON, V.; STEWART, T. J. **Multiple Criteria Decision Analysis: an integrated approach**. Norwell: Kluwer Academic Publishers, 2002.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1976.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Times series analysis: forecasting and control**. 3ª Ed. San Francisco: Holden-Day, 1994.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 4. ed. New Jersey: John Wiley & Sons, 2008.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Times Series Analysis – Forecasting and Control**. 5 ed. New Jersey: Wiley, 2016.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2 ed. Rio de Janeiro: LTC Editora, 2007.

CAMPOS, V. R. **Modelo de Apoio a Decisão Multicritério para Priorização de Projetos em Saneamento**. 2011. Tese (Doutorado). Programa de Pós-Graduação em Engenharia de Produção, Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, 2011.

CASTRO, L. N. **Fundamentals of natural computing: Basic concepts, algorithms and applications**. Boca Raton: Chapman& Hall, 2006.

CAVALCANTE, W. M. **Redes neurais artificiais como ferramenta de auxílio na operação e controle da estação de tratamento de água de palmeira dos índios**. 2018. Dissertação (Mestrado). Programa de Pós-Graduação em Engenharia Industrial, Universidade Federal da Bahia. Salvador, 2018.

CHUJAI, P, KERDPRASOP N, KERDPRASOP K. Time series analysis of household electric consumption with ARIMA and ARMA models. In: **Proc. IMECS Conf**. Hong Kong, 2013.

CIA, Central Intelligence Agency. **Eletricidade, mapa de consumo mundial**. USA: CIA World Factbook, 2018. Disponível em: <
<https://www.cia.gov/library/publications/the-world-factbook/>>. Acesso em: 10 ago. 2019.

CUI, H.; PENG, X. Short-Term City Electric Load Forecasting with Considering Temperature Effects: An Improved ARIMAX Model. **Mathematical Problems in Engineering**. v. 2015, p. 1-10, 2015.

DARBELLAY, G.A.; SLAMA M. Forecasting the short-term demand for electricity: do neural networks stand a better chance. **International Journal Forecasting**. v. 16, p. 71-83, 2000.

DAVIS, M. M.; AQUILANO, N. J.; CHASE, R. B. **Fundamentals of operations management**. Boston: McGraw-Hill Irwin, 2003.

EPE, Empresa de Pesquisa Energética. **Publicações/dados abertos**. Rio de Janeiro: Ministério de Minas e Energia, 2019. Disponível em: <<http://www.epe.gov.br/pt/publicacoes-dados-abertos>>. Acesso em: 11 jun. 2019.

FIGUEIRA, J.; MOUSSEAU, V.; ROY, B. **ELECTRE Methods**, em FIGUEIRA, J.; GRECO, S.; EHRGOTT, M. (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer Science + Business Media, New York, 133-162, 2005.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

GOMES, L. F. A. M.; ARAYA, M. C. G.; CARIGNANO, C. **Tomada de decisões em cenários complexos**: introdução aos métodos discretos do apoio multicritério à decisão. São Paulo: Pioneira Thomson Learning, 2004.

GOMES, L. F. A. M. **Teoria da Decisão**. São Paulo: Thomson Learning, 2007.

GOMES, L. F. A. M.; GOMES, C. F. S. **Tomada de Decisão Gerencial**: o Enfoque Multicritério. Rio de Janeiro: Atlas, 2012.

GUIRADO, L. **Comparação do desempenho de Modelos Lineares Generalizados (MLG) e Modelos Aditivos Generalizados (MAG) na predição de dados financeiros em credit score**. 2010. Dissertação (Mestrado) – Programa de Pós-Graduação em Estatística, Universidade Federal de São Carlos. São Carlos, 2010.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of machine learning research**, v. 3, n. Mar, p. 1157–1182, 2003.

HARALAMBOPOULOS, D. A.; POLATIDIS, H. Renewable energy projects: structuring a multicriteria group decision-making framework. **Renewable energy**, v.28 p. 961-973, 2003.

HAYKIN, S. **Neural Networks and Learning Machines**. 3 rd. New York: Prentice Hall, 2008.

HERNÁNDEZ, L.; BALADRÓN, C.; AGUIAR, J. M.; CALAVIA, L.; CARRO, B.; ESGUEVILLAS, A. S.; GARCÍA, P.; LLORET, J. Experimental Analysis of the Input Variables' Relevance to Forecast Next Day's Aggregated Electric Demand Using Neural Networks. **Energies**. v. 6, p. 2927-2948, 2013.

HUANG, G. B.; ZHU Q. Y.; SIEW, C. K. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. **Proceedings of International Joint Conference on Neural Networks**. p. 985-990, 2004.

HUANG, G. B.; ZHU, Q. Y.; SIEW, C. K. Extreme learning machine: theory and applications. **Neurocomputing, Elsevier**, v. 70, n. 1, p. 489–501, 2006.

HUANG, G. H.; ZHOU, H.; DING, X.; ZHANG, R. Extreme learning machine for regression and multiclass classification. **Transactions on Systems, MAN, and Cybernetics – PART B: Cybernetics, IEEE**. v. 42, n. 2, p. 513–529, 2012.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. Australia: Melbourne, 2018.

IEA, Agência Internacional de Energia. **Participação da OCDE e consumo de energia**. 2019. Disponível em: < <https://www.iea.org/>>. Acesso em: 11 jun. 2019.

IEA, Agência Internacional de Energia. **Estatísticas de eletricidade**. 2018. Disponível em: < <https://www.iea.org/statistics/electricity/>>. Acesso em: 11 jun. 2019.

JUDGE, G. G.; HILL, R. C.; GRIFFITHS, E. W.; LUTKEPOHL, H.; LEE, T. **Introduction to the Theory and Practice of Econometrics**. 2ª ed. USA: John Wiley & Sons, 1988.

KHANAM, S.; SIDDIQUI, J.; TALIB, F. Prioritising the TQM Enablers and IT Resources in the ICT Industry: An AHP Approach. World Academy of Science,

Engineering and Technology. **International Journal of Social, Behavioral, Educational, Economic and Management Engineering**, v. 9, n. 5, 2015.

LAZZARIN, L. N. A. **Método ensemble baseado em redes neurais artificiais para estimação de internações por doenças respiratórias**. 2019. Dissertação (Mestrado). Programa de Pós-Graduação em Ciência da Computação, Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2019.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. J. **Forecasting methods and applications**. 3 ed. New York: John Wiley & Sons, 1998.

MCCULLAGH, P.; NELDER, J. A. **Generalized linear models**. New York: Chapman and Hall, 1989.

MCCULLOCH, W.; W. PITTS. A Logical Calculus of Ideas Immanent in Nervous Activity. **Bulletin of Mathematical Biophysics**, p. 115–133, 1943.

MENEGHINI, M.; ANZANELLO, M.; KAHMANN, A.; TORTORELLA, G. L. Ajuste de previsão de demanda quantitativa com base em fatores qualitativos: estudo de caso em um restaurante fast food. **Sistemas & Gestão**. v. 13, p. 68-80, 2018.

MME, Ministério de Minas e Energia. Estatísticas do setor energético brasileiro. 2019. Disponível em: < <http://www.mme.gov.br/>>. Acesso em: 11 jun. 2019.

MOHAMMED, N. A. Modelling of unsuppressed electrical demand forecasting in Iraq for long term. **Energy**. v. 162, p. 354-363, 2018.

MONTGOMERY, D.; JOHNSON, L.; GARDINER, J. **Forecasting and Time Series Analysis**. New York: McGraw-Hill, 1990.

MORETTIN, P.A.; TOLOI, C.M.C., **Análise de Séries Temporais**. Ed. Blucher, 2006.

MOUSSEAU, V.; SLOWINSKI, R.; ZIELNIEWICZ, P. A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support. **Computers & Operations Research**, v. 27, p.757-777, 2000.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. **J. R. Statist. Soc. A**, v. 135, p. 370-384, 1972.

NEUHAUS, J.; MCCULLOCH, C. Generalized linear models. **Wiley Interdisciplinary Reviews: Computational Statistics, Wiley Online Library**. v. 3, n. 5, p. 407–413, 2011.

OSHODI, O. S.; EJOHWOMU, O. A.; FAMAKIN, I. O.; CORTEZ, P. Comparing univariate techniques for tender price index forecasting: Box-Jenkins and neural network model. **Construction Economics and Building**. v. 17, p. 109–123, 2017.

PAO, H. T. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. **Energy**. v. 31, p. 2129–2141, 2006.

PELLEGRINI, F. R. **Metodologia para a implementação de sistemas de previsão de demanda**. 2000. Dissertação (Mestrado). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, Porto Alegre - RS. 2000.

REIS, L. P.; LADEIRA, M. B.; FERNANDES, J. M. Contribuição do método analytic hierarchy process (ahp) para auxílio ao processo decisório de terceirizar ou internalizar atividades no contexto de uma empresa de base tecnológica. **Revista Produção Online**, v.13, n. 4, p. 1325-1354, 2013.

ROSA, C.R.M.; STEINER, M.T.A.; COLMENERO, J.C. Utilização de processo de análise hierárquica para definição estrutural e operacional de centros de distribuição: uma aplicação a uma empresa do ramo alimentício. **Gestão e Produção**, v.22, n. 4, p. 935-950, 2015.

ROY, B. Classement et choix en présence de points de vue multiples. **RAIRO Operations Research-Recherche Opérationnelle**, v. 2, n 1, p. 57-75, 1968.

ROY, B. ELECTRE III: Un algorithme de classements fondé sur une représentation floue des préférences en présence de criteres multiples. **Cahiers du CERO**, v. 20, n. 1, p. 3-24, 1978

ROY, B. The outranking approach and the foundations of ELECTRE methods. **Theory and Decision**, v. 31, p. 49-73, 1991.

ROY, B. **Multicriteria methodology for decision aiding**. Dordrecht: Kluwer Academic, 1996.

ROY, B.; BERTIER, P. La méthode ELECTRE II. **SEMA (Metra internacional), Note de travail**. p. 142, 1973.

ROY, B.; HUGONNARD, J. C. Ranking of suburban line extension projects on the Paris metro system by multicriteria method. **Transpn. Res.** v.16, n.4, p.301-312, 1982.

ROY, B.; SKALKA, J. M. ELECTRE IS: Aspéctis methodologiques et guide d'utilisation. Cahier du LAMSADE. **Université de Paris – Dauphine**. Paris, fev. 1985.

SAATY, T. L. **The Analytic Hierarchy Process**. New York: McGraw-Hill International, 1980.

SAATY, T. L. “**Método de Análise Hierárquica**”, Tradução de Wainer da Silveira e Silva. São Paulo: McGraw-Hill, Makron. 1991.

SAATY, T. L. **Decision making for leaders**. Pitts burg, USA: WS. Publications, 2000.

SADOWNIK, R.; BARBOSA, E. P. Short-term Forecasting of Industrial Electricity Consumption in Brazil. **Journal of Forecasting**. v.18, p. 215-224, 1999.

SILVA, A. C. G. C. **Modelo de Avaliação Multicritério para Gestão Organizacional, aplicado ao Polo Exportador de Frutas do Vale de São Francisco**. 2016. Tese (Doutorado). Programa de Pós-Graduação em Engenharia Industrial, Universidade Federal da Bahia, Salvador, 2016.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas: Fundamentos Teóricos e Aspectos Práticos**. 2 ed. São Paulo: Artliber, 2016.

SILVA, N.; TAY, I.; OKIDA, S.; STEVAN JR., S. L.; SIQUEIRA, H. Neural Networks for Predicting Prices of Sugarcane Derivatives. **Sugar Tech**. DOI: 10.1007/s12355-018-0648-5. 2018a.

SIQUEIRA, H. V. **Previsão de séries de vazões com redes neurais artificiais em modelos lineares ajustados por algoritmos bio-inspirados**. 2009. Dissertação (Mestrado). Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, CAMPINAS, 2009.

SIQUEIRA, H. V. **Máquinas Desorganizadas para Previsão de Séries de Vazões**. 2013. Tese (Doutorado). Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, CAMPINAS, 2013.

SIQUEIRA, H. V.; BOCCATO, L.; ATTUX, R. R. F.; LYRA FILHO, C. Echo state networks in seasonal streamflow series prediction. **Learning and Nonlinear Models**, v. 10, p. 181–191, 2012.

SUGANTHI, L.; SAMUEL, A. A. Modelling and forecasting energy consumption in INDIA: Influence of socioeconomic variables. **Energy Sources, Part B: Economics, Planning, and Policy**. v. 11:5, p. 404-411, 2016.

TAYLOR, J. A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center. **Management Science**. v. 54, n. 2, pp. 253–265, 2008.

TORRINI, F.C.; SOUZA, R. C.; OLIVEIRA, F. L. C.; PESSANHA, J. F. M. Long term electricity consumption forecast in Brazil: A fuzzy logic approach. **Socio-Economic Planning Sciences**, v. 54, p. 18-27, 2016.

TROJAN, F. **Modelos Multicritério para Apoiar Decisões na Gestão da Manutenção de Redes de Distribuição de Água para a redução de Custos e Perdas**. 2012. Tese (Doutorado) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Pernambuco. Recife, 2012.

TSOUKIAS, A. From decision theory to decision aiding methodology. **European Journal of Operational Research**, v. 187, p.138-161, 2008.

TUTUN, S.; CHOU, C. A.; CANYLMAZ, E. A new forecasting framework for volatile behavior in net electricity consumption: A case study in Turkey. **Energy**. v. 93, p. 2406-2422, 2015.

VINCKE, P. **Multicriteria decision aid**. New York: John Wiley e Sons, 1992.

VU, D. H.; MUTTAQI, K. M.; E AGALGAONKAR, A. P. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*. v. 140, p. 385–394, 2015.

WANG, J.; ZHANG, J.; NIE, J. An Improved Artificial Colony Algorithm Model for Forecasting Chinese Electricity Consumption and Analyzing Effect Mechanism. *Mathematical Problems in Engineering*. v. 2016, p. 1-14, 2016.

WU, L.; GAO, X.; XIAO, Y.; YANG, Y.; CHEN, X. Using a novel multi variable grey model to forecast the electricity consumption of Shandong Province in China. *Energy*. v. 157, p. 327-335, 2018.

YU, W. **Aide multicritère à la décision dans le cadre de la problématique du tri:** Concepts, méthodes et applications, PhD thesis, Université Paris-Dauphine, 1992.

ZOPOUNIDIS, C.; DOUMPOS, M. Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research*, v. 138, p. 229-246, 2002.