

JOSÉ TORRES COURA NETO

**PROCESSAMENTO DE VALORES ATÍPICOS EM REDES ELÉTRICAS
INTELIGENTES BASEADO EM ALGORITMOS NEURO-FUZZY**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica - PPGEE, da Universidade Federal da Paraíba - UFPB, como requisito para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador:
Prof. Dr. Juan Moises Maurício Villanueva

JOÃO PESSOA

2018

JOSÉ TORRES COURA NETO

**PROCESSAMENTO DE VALORES ATÍPICOS EM REDES ELÉTRICAS
INTELIGENTES BASEADO EM ALGORITMOS NEURO-FUZZY**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica - PPGEE, da Universidade Federal da Paraíba - UFPB, como requisito para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador:
Prof. Dr. Juan Moises Maurício Villanueva

JOÃO PESSOA

2018

Catálogo na publicação
Seção de Catalogação e Classificação

N469p Neto, José Torres Coura.

Processamento de valores atípicos em redes elétricas inteligentes baseado em algoritmos neuro-fuzzy / José Torres Coura Neto. - João Pessoa, 2018.

71 f.

Orientação: Juan Moises Mauricio Villanueva.
Dissertação (Mestrado) - UFPB/CEAR.

1. Redes Elétricas Inteligentes. 2. Medidores Inteligentes. 3. Valores Atípicos. 4. RNA. 5. ANFIS. I. Villanueva, Juan Moises Mauricio. II. Título.

UFPB/BC

UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB
CENTRO DE ENERGIAS ALTERNATIVAS E RENOVÁVEIS – CEAR
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA – PPGE

A Comissão Examinadora, abaixo assinada, aprova a Dissertação

PROCESSAMENTO DE VALORES ATÍPICOS EM REDES ELÉTRICAS
INTELIGENTES BASEADO EM ALGORITMOS NEURO-FUZZY

Elaborada por

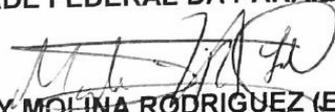
JOSÉ TORRES COURA NETO

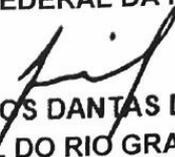
como requisito para obtenção do grau de
Mestre em Engenharia Elétrica.

COMISSÃO EXAMINADORA


PROF. DR. JUAN MOISES MAURÍCIO VILLANUEVA (Orientador)
UNIVERSIDADE FEDERAL DA PARAÍBA - UFPB


PROF. DR. HELON DAVID MACEDO DE BRAZ (Examinador Interno)
UNIVERSIDADE FEDERAL DA PARAÍBA - UFPB


PROF. DR. YURI PERCY MOLINA RODRÍGUEZ (Examinador Interno)
UNIVERSIDADE FEDERAL DA PARAÍBA - UFPB


PROF. DR. IVANOVITCH MEDEIROS DANTAS DA SILVA (Examinador Externo)
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE - UFRN

João Pessoa/PB
Novembro de 2018.

Dedico este trabalho à minha amada esposa Camila,
por estar ao meu lado nesse campo de batalhas que é a vida e por
compartilhar comigo sonhos por dias melhores.

AGRADECIMENTOS

Agradeço à Deus pelo dom da vida e por todas as oportunidades concedidas a mim.

À minha família por todo auxílio empenhado para que eu tivesse condições de estudar e lutar pelos meus objetivos. Especialmente, aos meus pais Clóvis Fernandes e Maria do Socorro Abrantes Fernandes e meus irmãos por todo amor a mim dado e pelos ensinamentos para a vida.

Ao meu amor e esposa Camila Rocha Vieira Torres, inspiração e esperança de dias melhores. E sua família por todo apoio prestado.

Ao meu professor orientador Juan Moises Mauricio Villanueva por acreditar no meu potencial e auxiliar decisivamente pela conclusão deste projeto de vida.

Aos meus professores e colegas estudantes da Universidade Federal da Paraíba por compartilhar seus conhecimentos e apoios, especialmente, Euler Macêdo, Nady Rocha, Pedro Meira e Oton Matheus.

Aos meus colegas servidores públicos do Instituto Federal da Paraíba – Campus Picuí e Campus Cajazeiras, especialmente, Fernando Fernandes, e aos meus alunos pelo compartilhamento de novas experiências e lutas pela educação pública e de qualidade.

E a todas as pessoas que de alguma forma, me motivaram a concluir este trabalho e, com isso, tornar o mundo um pouco melhor.

“Pois quem quer amar a vida e ver dias felizes refreie a língua do mal e evitem que seus lábios falem dolosamente; aparte-se do mal, pratique o que é bom, busque a paz e emprenhe-se por alcança-la”.

1Pe 3.10

SUMÁRIO

LISTA DE ILUSTRAÇÕES	XI
LISTA DE TABELAS	XII
RESUMO	XIII
ABSTRACT	XIV
1 INTRODUÇÃO	16
1.1 MOTIVAÇÃO E RELEVÂNCIA DO TRABALHO	16
1.2 OBJETIVOS	18
1.2.1 Geral	18
1.2.2 Específicos	19
1.3 ORGANIZAÇÃO	19
2 FUNDAMENTAÇÃO TEÓRICA	22
2.1 SETOR ENERGÉTICO BRASILEIRO	22
2.2 REDES ELÉTRICAS INTELIGENTES	24
2.3 MEDIDORES INTELIGENTES	25
2.4 <i>BIG DATA ANALYTICS</i> E VALORES ATÍPICOS	28
2.5 ALGORITMOS DE INTELIGENCIA ARTIFICIAL	29
2.5.1 Sistemas de Inferência Fuzzy	30
2.5.2 Redes Neurais Artificiais	32
2.5.3 Sistemas ANFIS	33
2.5.3.1 Características ANFIS	33
2.5.3.2 Arquitetura ANFIS	34
3 METODOLOGIA	40
3.1 TOPOLOGIA PROPOSTA DE SMART METERS	40
3.2 PERFIL DAS CURVAS DE DEMANDA	41
3.3 TRATAMENTO E PREPARAÇÃO DOS DADOS	43
3.4 ALGORITMO DE INTERPOLAÇÃO	45
3.5 ALGORITMOS RNA E ANFIS	45
3.5.1 Algoritmo RNA	46
3.5.2 Algoritmo ANFIS	48
4 RESULTADOS E DISCUSSÕES	55
4.1 CENÁRIO 1: UM DIA DE MEDIÇÕES COM OUTLIERS	55
4.2 CENÁRIO 2: DEZ DIAS DE MEDIÇÕES COM OUTLIERS	61

5 CONCLUSÕES.....	67
5.1 PUBLICAÇÕES DA PESQUISA.....	68
5.2 TRABALHOS FUTUROS	68
REFERÊNCIAS	69

LISTA DE ILUSTRAÇÕES

FIGURA 1. MATRIZ ENERGÉTICA DO BRASIL.	22
FIGURA 2. SISTEMA DE COMPENSAÇÃO DE ENERGIA ELÉTRICA.	23
FIGURA 3. SETOR ENERGÉTICO BRASILEIRO.	24
FIGURA 4. TOPOLOGIA DE MEDIÇÃO EM UMA REDE ELÉTRICA INTELIGENTE.	25
FIGURA 5. DIAGRAMA EM BLOCOS DO MÓDULO DE MEDIÇÃO (MM).	26
FIGURA 6. DIAGRAMA EM BLOCOS DO CONCETRADOR PRIMÁRIO (CP).	27
FIGURA 7. PROCESSO TÍPICO DE MINERAÇÃO DE DADOS.	28
FIGURA 8. FUNÇÕES DE PERTINÊNCIA FUZZY.	30
FIGURA 9. ETAPAS DO SISTEMA DE INFERÊNCIA FUZZY.	32
FIGURA 10. ARQUITETURA BÁSICA DAS REDES NEURAIS ARTIFICIAIS (RNA).	33
FIGURA 11. MODELO DO SISTEMA DE INFERÊNCIA TAKAGI-SUGENO.	35
FIGURA 12. ARQUITETURA DO SISTEMA DE INFERÊNCIA ANFIS TAKAGI-SUGENO.	37
FIGURA 13. DIAGRAMA EM BLOCOS DE UMA NOVA TOPOLOGIA DE MEDIDORES INTELIGENTES.	41
FIGURA 14. CURVA DE DEMANDA DE POTÊNCIA.	42
FIGURA 15. CURVA DE DEMANDA COM INSERÇÃO DE OUTLIERS DO TIPO ZERO.	43
FIGURA 16. TRATAMENTO E PREPARAÇÃO DOS DADOS POR JANELAS DESLIZANTES.	44
FIGURA 17. MODELO DO ALGORITMO DE INTERPOLAÇÃO.	45
FIGURA 18. MODELO DOS ALGORITMOS DE RNA E ANFIS.	46
FIGURA 19. MODELO DA REDE NEURAL ARTIFICIAL.	48
FIGURA 20. ARQUITETURA DO SISTEMA ANFIS.	50
FIGURA 21. CONJUNTOS DE PERTINÊNCIA DAS ENTRADAS E SAÍDA ANFIS.	51
FIGURA 22. APLICAÇÃO DAS REGRAS DO SISTEMA ANFIS.	52
FIGURA 23. GRÁFICO DE SUPERFÍCIE PARA AS REGRAS DO SISTEMA ANFIS.	53
FIGURA 24. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER E CORREÇÕES PARA VERIFICAÇÃO.	56
FIGURA 25. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER TIPO ZERO E CORREÇÕES NO CENÁRIO 1.	57
FIGURA 26. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER TIPO PICO E CORREÇÕES NO CENÁRIO 1.	58
FIGURA 27. GRÁFICO DO ERRO RELATIVO ANFIS E SEU DESVIO PADRÃO NO CENÁRIO 1.	59
FIGURA 28. GRÁFICO DO ERRO RELATIVO RNA E SEU DESVIO PADRÃO NO CENÁRIO 1.	60
FIGURA 29. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER TIPO ZERO E CORREÇÕES NO CENÁRIO 2.	62
FIGURA 30. GRÁFICO DO ERRO RELATIVO ANFIS E SEU DESVIO PADRÃO NO CENÁRIO 2.	64
FIGURA 31. GRÁFICO DO ERRO RELATIVO RNA E SEU DESVIO PADRÃO NO CENÁRIO 2.	65

LISTA DE TABELAS

TABELA 1. ERROS RELATIVOS NA ESTIMAÇÃO DA JANELA DE VALORES.....	44
TABELA 2. ANÁLISE DO NÚMERO DE NEURÔNIOS PARA DUAS CAMADAS RNA.....	47
TABELA 3. ANÁLISE DO NÚMERO DE NEURÔNIOS PARA TRÊS CAMADAS RNA.	47
TABELA 4. ANÁLISE DO NÚMERO DE NEURÔNIOS DO CONJUNTO DE PERTINÊNCIA ANFIS..	49
TABELA 5. ANÁLISE DOS TIPOS DE CONJUNTO DE PERTINÊNCIA ANFIS.....	49
TABELA 6. ERROS RELATIVOS DA CORREÇÃO DE OUTLIERS EM UM DIA.....	58
TABELA 7. MÉDIA DO ERRO QUADRÁTICO PARA O CENÁRIO 1.	60
TABELA 8. ERROS RELATIVOS DA CORREÇÃO DE OUTLIERS PARA 7 DIAS.	62
TABELA 9. MÉDIA DO ERRO QUADRÁTICO PARA O CENÁRIO 2.	65

RESUMO

PROCESSAMENTO DE VALORES ATÍPICOS EM REDES ELÉTRICAS INTELIGENTES BASEADO EM ALGORITMOS NEURO-FUZZY

As redes elétricas inteligentes estão cada vez mais presentes para aperfeiçoar a eficiência energética e diversos equipamentos são fundamentais nesse processo, entre eles os medidores inteligentes. Tais dispositivos acumulam um fluxo enorme de informações passíveis de serem analisadas para auxiliar na tomada de decisão dos controladores do sistema elétrico. O que torna presente o conceito de Big Data Analytics, capaz de processar diversos dados e corrigir valores atípicos, denominados de outliers, por meio de algoritmos utilizando inteligência artificial, como a lógica Fuzzy e as Redes Neurais Artificiais. Como forma de melhorar os resultados existentes, o presente trabalho sugere a utilização de um algoritmo híbrido, o Adaptive Neuro-Fuzzy Inference System – ANFIS. Tal algoritmo mostrou melhores desempenhos na correção de outliers quando comparados com técnicas baseadas em Redes Neurais Artificiais - RNA e Interpolação Linear. Finalmente, serão apresentados resultados das estimações usando dados reais de demanda de energia para uma subestação de distribuição de energia elétrica.

Palavras-chave: Redes Elétricas Inteligentes, Medidores Inteligentes, Valores Atípicos, RNA, ANFIS.

ABSTRACT

PROCESSING OF OUTLIERS IN SMART GRIDS BASED ON NEURO-FUZZY ALGORITHMS

Smart grids are increasingly present to improve the energy efficiency of the electrical network and several equipment are fundamental in this process, among them smart meters. Such devices accumulate an enormous flow of information that can be analyzed to aid in the decision making of the electric system controllers. This makes the concept of Big Data Analytics present, capable of processing various data and correcting atypical values, called outliers, through algorithms using artificial intelligence, such as Fuzzy logic and Artificial Neural Networks. As a way of improving existing results, the present work suggests the use of a hybrid algorithm, the Adaptive Neuro Fuzzy Inference System (ANFIS). This algorithm showed better performance in the correction of outliers when compared with techniques based on Artificial Neural Networks and Linear Interpolation. Finally, results of the estimations will be presented using real energy demand data for a substation of electricity distribution.

Keywords: Smart grids, Smart meters, Outliers, ANN, ANFIS.

1 INTRODUÇÃO

1 INTRODUÇÃO

1.1 MOTIVAÇÃO E RELEVÂNCIA DO TRABALHO

Com o aumento da população mundial, é crescente a demanda por energia elétrica (DEVI, 2017). Conseqüentemente, as usinas de geração de energia continuam utilizando fontes de energia convencionais, assim como, estão buscando suprir a necessidade de energia a partir de fontes de energias alternativas e renováveis (SULAIMAN et. al., 2016). No entanto, para atender às novas demandas de energia é necessária a implantação de uma infraestrutura tecnológica e inteligente, a fim de transformar a atual estrutura em uma rede elétrica inteligente, ou smart grid.

As redes elétricas inteligentes são compostas por uma infraestrutura de equipamentos eletrônicos dotados de uma tecnologia de comunicação digital bidirecional com o intuito de aperfeiçoar a eficiência energética da rede elétrica (SULAIMAN et. al., 2016). Para o bom funcionamento dessa rede, é necessária que haja comunicação em tempo real entre os pontos da rede (instalações residenciais e comerciais) e os centros de controle (Di SARNO et. al., 2014). Alguns desses dispositivos são denominados de medidores inteligentes, ou smart meters, os quais são capazes de medir, processar e transmitir dados referentes ao consumo de energia elétrica (AMARAL et. al., 2014). Tais valores de medição compõem séries temporais, as quais podem ser analisadas para a extração de padrões de consumo de energia elétrica para auxílio na tomada de decisões.

Assim, as redes elétricas inteligentes possuem capacidade de coletar, transmitir e processar dados do sistema elétrico, construindo um banco de dados com grande volume de informação, que pode ser denominado de *Big Data* (WANG et. al., 2015). A partir desse banco de dados, é possível a extração de informações úteis por meio de análises baseadas em técnicas estatísticas ou de inteligência artificial, que podem melhorar continuamente a eficiência e o funcionamento das redes elétricas inteligentes (HOU et. al., 2016).

Para estudar a captação de informações em meio a grandes volumes de dados, um novo termo foi forjado pela comunidade científica, o *Big Data Analytics* (WANG et. al., 2015). Inúmeras vantagens podem ser elencadas quanto à utilização

de *Big Data Analytics* em redes elétricas inteligentes. Especificamente, na análise de séries temporais em curvas de demanda.

Para o consumidor final, o benefício de se adquirir e analisar dados de consumo de energia é extremamente importante para o seu planejamento. Pois a utilização de cargas se dá especialmente em situações de tarifas que variam de acordo com o horário de consumo. Portanto, há uma possibilidade real de melhor utilização da energia de acordo com o plano tarifário (WANG et. al., 2017). Enquanto que, para as concessionárias, existe a oportunidade de se planejar a geração e a distribuição de energia de acordo com uma demanda estimada, minimizando erros e possíveis faltas de energia. Assim como, é possível a estimação de perdas financeiras, sejam por meio de problemas técnicos ou perdas comerciais (SULAIMAN et. al., 2016).

Na análise de séries temporais, é necessário que os dados apresentem a maior exatidão possível a fim de que não possam resultar em análises falhas e sem validação. Tudo isso, porque os dados em redes elétricas inteligentes dispostos em séries temporais podem ser afetados por inúmeros fatores, tais como erros de comunicação, falha na operação e nos dispositivos de medição. No campo do *Big Data Analytics* e do *Data Mining*, ou mineração de dados, são definidos por valores atípicos, ou *outliers*, observações discrepantes presentes nas séries temporais (YECLÉ et. al. 2018; YE et. al., 2016; DINGPING, 2013).

Dentro do conceito de *Big Data Analytics*, diversos métodos de *Data Mining* são apresentados em dois blocos distintos (HOU et. al., 2016): algoritmos básicos e algoritmos integrados de mineração de dados. Nesse contexto, métodos estatísticos já foram utilizados para a detecção e correção de outliers em séries temporais presentes em redes elétricas inteligentes, especificamente em curvas de demanda de energia. Dentre as principais técnicas se encontram o método dos mínimos quadrados e do processo recursivo (OKON et. al., 2010). Assim como, diversas aplicações são baseadas no filtro de Kalman (ZONGXIANG et. al., 2016; VALVERDE et. al., 2016; e ZANG et. al., 2011) e em curvas probabilísticas (YE et. al., 2016). No campo de *soft-computing*, ou métodos computacionais, vários algoritmos já foram implementados com o mesmo fim, assim como, as Redes Neurais Artificiais – RNA (KAUR, 2016) e a lógica Fuzzy (SATHIRACHEEWIN, 2011).

Especificamente, levando-se em conta a aplicação de *Big Data Analytics* para o processamento, ou seja, identificação e correção, de valores atípicos baseada nas técnicas de Lógica Fuzzy e RNA aplicadas a curvas de previsão de demanda, um módulo complementar ao smart meter foi proposto em (ANDRADE et. al., 2018a). Deste modo, a aplicação RNA se mostrou mais eficiente em horários do dia próximos ao horário de pico de demanda (maior não linearidade), enquanto que, a Fuzzy em horários fora dos de pico (ANDRADE et. al., 2018b).

Tendo em vista as características da Lógica Fuzzy e das Redes Neurais Artificiais - RNA, destaca-se que as mesmas são complementares. Ou seja, as Redes Neurais Artificiais possuem a capacidade de aprender padrões a partir do treinamento supervisionado. Enquanto que, a Lógica Fuzzy propriamente dita possui a necessidade de contar com o conhecimento do especialista para a definição de regras.

Portanto, com o objetivo de corrigir valores atípicos em redes elétricas inteligentes utilizando algoritmos neuro-fuzzy, como modelo sinérgico para o aproveitamento das técnicas de inteligência artificial, o presente trabalho utiliza o Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS) para correção de valores atípicos identificados em medições de demanda presentes em redes elétricas inteligentes.

1.2 OBJETIVOS

1.2.1 Geral

- Identificar e corrigir valores atípicos em medições de demanda de energia em redes elétricas inteligentes por meio do desenvolvimento e aplicação de algoritmos baseados no Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS - *Adaptative Neuro Fuzzy Inference System*).

1.2.2 Específicos

- Utilizar as ferramentas de *Big Data Analytics* para analisar dados do setor elétrico em pontos de medição da demanda de energia de subestações com a finalidade de extrair padrões e fornecer melhorias na tomada de decisões;
- Contribuir com o desenvolvimento de novas topologias de medidores inteligentes de energia, especificamente para a correção de valores atípicos e no contexto de redes elétricas inteligentes;
- Desenvolver novos algoritmos para correção de valores atípicos baseados no conceito de *Big Data Analytics* e suas implementações por meio de técnicas de inteligência artificial, utilizando o Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS);
- Apresentar e discutir os resultados obtidos com o intuito de sugerir melhorias na análise e correção de dados do setor elétrico.

1.3 ORGANIZAÇÃO

Além deste primeiro capítulo de Introdução, este trabalho se subdivide em mais quatro capítulos, os quais serão descritos sucintamente a seguir:

No segundo capítulo, Fundamentação Teórica, uma revisão da literatura será realizada com o intuito de apresentar conceitos e características dos tópicos: Setor Energético Brasileiro, Redes Elétricas Inteligentes, Medidores Inteligentes, *Big Data Analytics* e Valores Atípicos e, por fim, os algoritmos de inteligência artificial, sistemas de Inferência Fuzzy, Sistemas ANFIS e Redes Neurais Artificiais.

No terceiro capítulo, Metodologia, serão apresentados a topologia proposta de medidores inteligentes, o perfil das curvas de demanda, as técnicas de *Big Data Analytics* para correção de valores atípicos em curvas de demanda, baseadas em algoritmos de Interpolação Linear, Redes Neurais Artificiais (RNA) e Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS).

No quarto capítulo, Resultados e Discussões, serão apresentadas análises estatísticas, com o objetivo de indicar os pontos positivos e negativos dos algoritmos utilizados na correção de valores atípicos em curvas de demanda, dispostas em dois cenários distintos.

Por fim, no quinto capítulo, Conclusão, será realizada uma análise do cumprimento dos objetivos propostos pelo trabalho e do avanço alcançado com a utilização de algoritmos híbridos, como o ANFIS, na correção de valores atípicos em curvas de demanda de energia. Assim como, as publicações da presente pesquisa e, por fim, os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

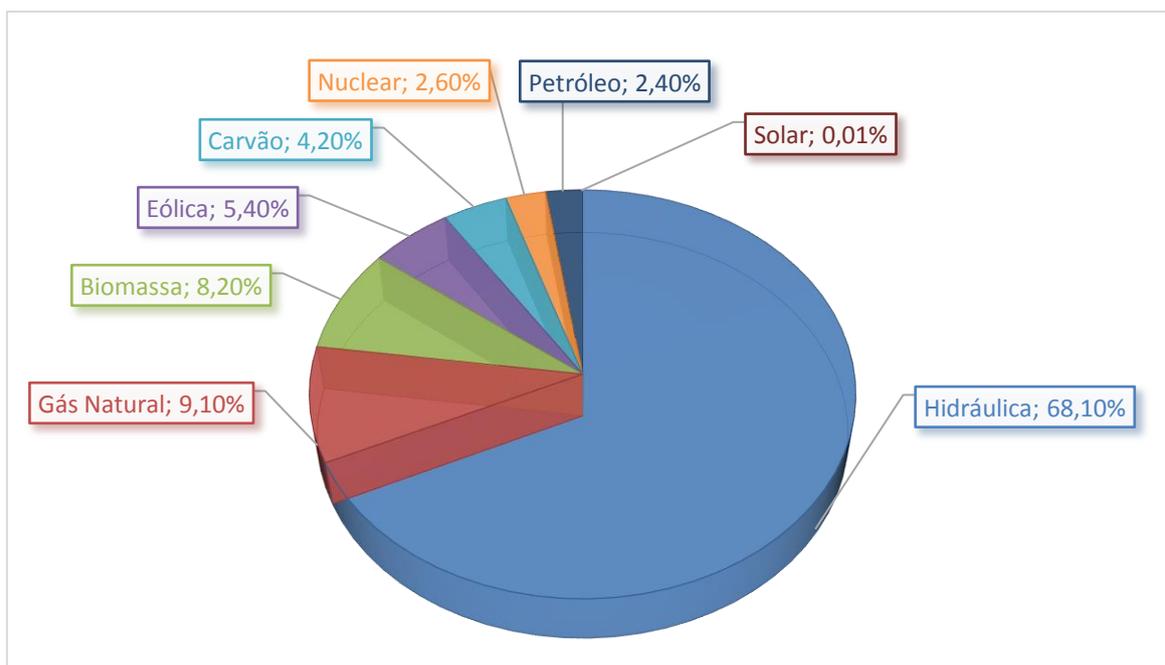
2 FUNDAMENTAÇÃO TEÓRICA

Com a finalidade de definir os conceitos fundamentais utilizados para o desenvolvimento deste trabalho, neste capítulo são apresentados e discutidos os seguintes tópicos: Setor Energético Brasileiro, Redes Elétricas Inteligentes, Medidores Inteligentes, Big Data Analytics e Valores Atípicos e, por fim, Algoritmos de Inteligência Artificial.

2.1 SETOR ENERGÉTICO BRASILEIRO

A matriz energética do Brasil é composta predominantemente pelo fornecimento de energia elétrica oriunda de usinas hidrelétricas, é o que indica o Balanço Energético Nacional publicado no ano de 2017. De acordo com o levantamento, conforme a Figura 1, a fonte hidráulica foi responsável por 68,1% da oferta interna de energia no ano de 2016. Enquanto que, a fonte eólica ofertou 5,4% e a solar apenas 0,01% da energia. Tais dados remontam ao longo percurso que o país ainda deve percorrer no caminho da micro e minigeração de energia por meio de fontes alternativas e renováveis (Portal EPE, 2017).

FIGURA 1. MATRIZ ENERGÉTICA DO BRASIL.

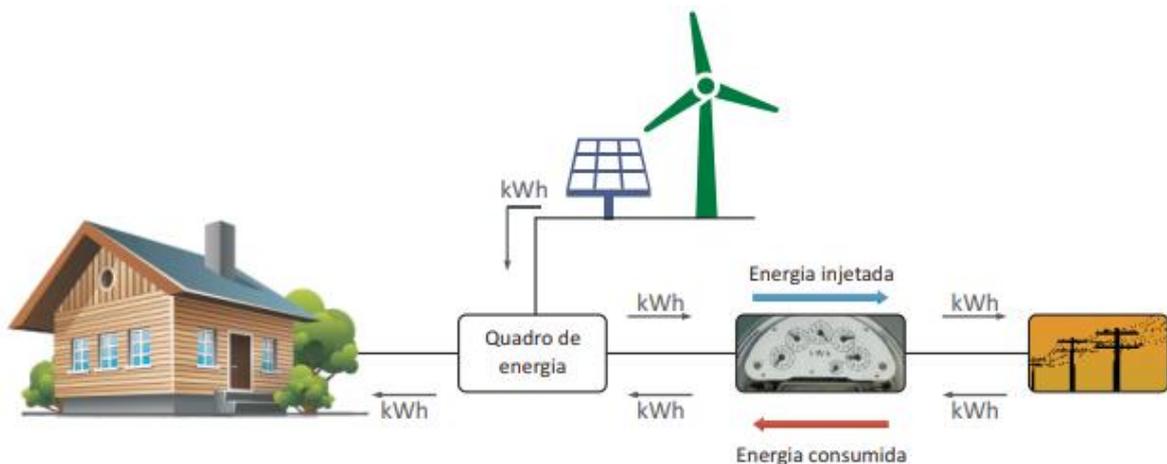


Fonte: Adaptada de Portal EPE, 2017.

Além das grandes usinas de geração de energia, com potência instalada superior a 1 MW, é possível a pequena produção de energia, seja por meio da microgeração, produzindo menos de 100 kW de potência, ou da minigeração, entre 100 kW e 1000 kW de potência. Tais usinas de geração de energia distribuída passaram a participar do sistema de compensação de energia elétrica (Resolução Normativa Nº 482, 2012), na qual o excedente gerado pela unidade consumidora passa a ser cedido à distribuidora local, o que é compensado em momento posterior com o consumo de energia elétrica pela unidade consumidora (ANEEL, 2014).

A Figura 2 ilustra o funcionamento do sistema de compensação de energia. Tal sistema indica que uma unidade residencial, por exemplo, pode gerar energia elétrica a partir de fontes renováveis, como a eólica e a solar, e por meio de um quadro de energia contabilizar a energia consumida e a energia injetada na rede elétrica. Dessa forma, caso haja excedente de energia, tais créditos serão compensados de acordo com a legislação vigente.

FIGURA 2. SISTEMA DE COMPENSAÇÃO DE ENERGIA ELÉTRICA.

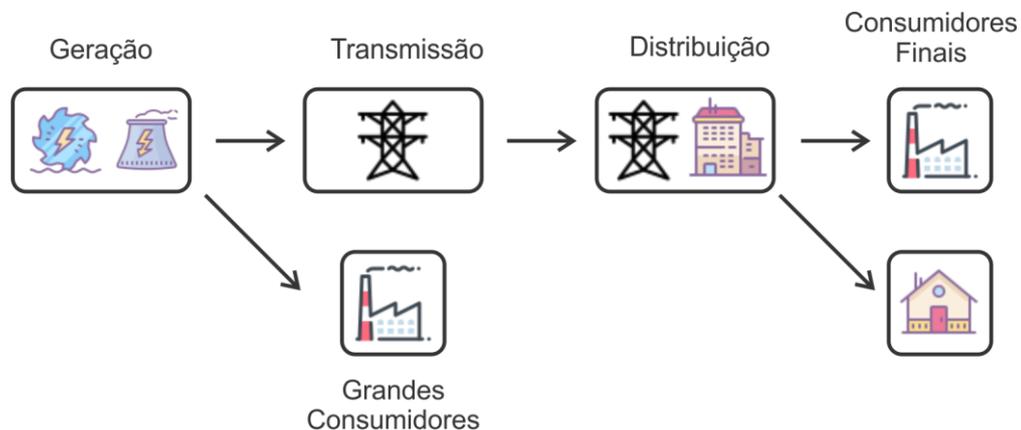


Fonte: ANEEL, 2014.

De acordo com o Atlas de Energia Elétrica do Brasil (ANEEL, 2008), o setor energético brasileiro é comandado pelo Sistema Interligado Nacional (SIN), responsável principal pela geração, transmissão e distribuição de energia elétrica no país. Conforme a Figura 3, as usinas de geração de energia estão interligadas com o sistema de transmissão, que podem fornecer energia diretamente para grandes consumidores ou para concessionárias de distribuição, que por sua vez, fornecem energia para os consumidores finais.

Com o avanço da tecnologia e o surgimento de novas formas de geração de energia, o emprego de uma rede elétrica inteligente se torna fundamental para a modernização do setor energético do país.

FIGURA 3. SETOR ENERGÉTICO BRASILEIRO.



Fonte: Adaptada de ANEEL, 2008.

2.2 REDES ELÉTRICAS INTELIGENTES

As redes elétricas inteligentes, ou smart grids, são resultados do avanço tecnológico em dispositivos eletrônicos, capazes de monitorar e controlar o fluxo de energia nos pontos de fornecimento e consumo de energia elétrica. Tal sistema é composto por diversos dispositivos, entre eles, os medidores inteligentes, ou smart meters (LEE et.al., 2018).

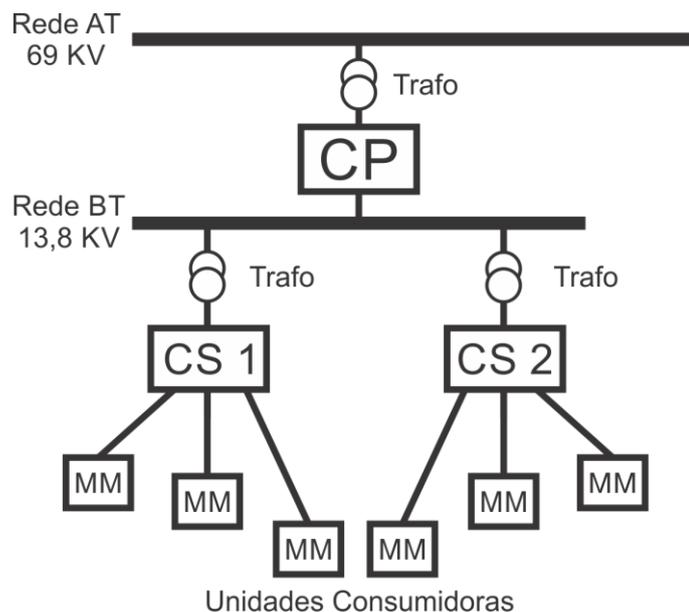
Na distribuição de energia elétrica para as mais diversas regiões do país, as subestações ocupam papel fundamental, sendo as mesmas administradas pelas concessionárias de distribuição de energia.

Como estudo de caso, conforme a Figura 4, uma subestação opera com tensão primária de 69 kV e secundária de 13,8 kV, com capacidade para 6,25 MVA de potência. Para o efetivo monitoramento, a rede elétrica inteligente possui três módulos principais (LANDIS GYR, 2011):

- Módulos de Medição (MM): presentes nas unidades consumidoras, responsáveis principalmente pelo monitoramento da energia consumida e gerada pela unidade;

- Concentradores Secundários (CS): localizados no topo dos postes, concatenando as informações e os controles dos Módulos de Medição (MM) próximos, e que se comunicam com um ponto central; e
- Concentrador Primário (CP): possui a função de comunicação com os Concentradores Secundários (CS) e com a central de controle na distribuidora de energia.

FIGURA 4. TOPOLOGIA DE MEDIÇÃO EM UMA REDE ELÉTRICA INTELIGENTE.



Fonte: Elaborada pelo Autor.

2.3 MEDIDORES INTELIGENTES

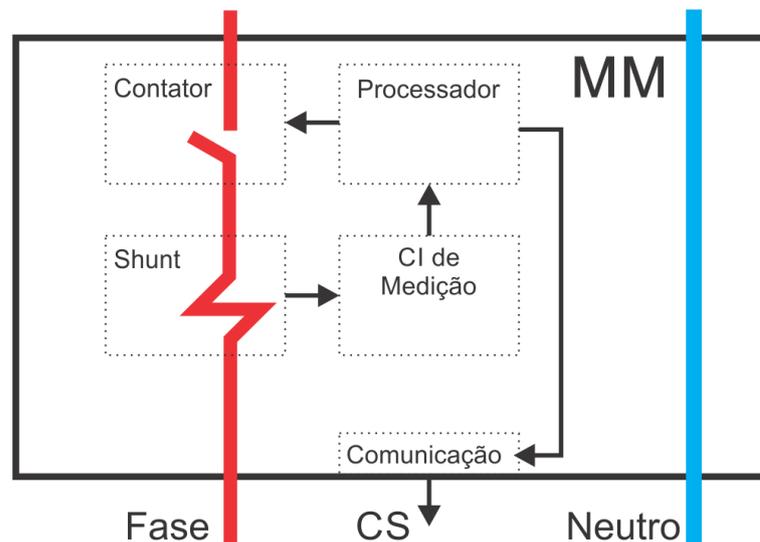
Na atualidade, uma nova classe de dispositivos está presente na rede elétrica, especificamente, em redes elétricas inteligentes, se trata da AMI – Infraestrutura Avançada de Medição, composta por medidores inteligentes, redes de comunicação e sistema de gerenciamento de dados (SULAIMAN et. al., 2016). Tais dispositivos podem ou não possuir processamento de informações por *soft-computing* ou inteligência artificial, sendo denominados de medidores inteligentes, ou smart meters. Tais dispositivos podem ser classificados por: Módulo de Medição (MM), Concentrador Secundário (CS) e Concentrador Primário (CP).

O Módulo de Medição (MM) tem como característica a realização da medição de energia e o corte ou religamento da unidade consumidora. De acordo com o

diagrama de blocos do MM, conforme a Figura 5, é possível destacar os seguintes componentes (LANDIS GYR, 2011):

- Sistema de medição do tipo *shunt* acoplado a um circuito integrado (CI) de medição;
- Processador - registro e processamento dos dados de energia;
- Registradores - acumulo de dados de medição;
- Contator - corte ou religação de pontos de energia; e
- Comunicação – ligação por meio de fibra óptica com o CS.

FIGURA 5. DIAGRAMA EM BLOCOS DO MÓDULO DE MEDIÇÃO (MM).



Fonte: Elaborada pelo Autor.

O Concentrador Secundário (CS) possui comunicação com todos os MM e se comunica com o CP de forma remota por meio de uma CPU. O CS possui os seguintes componentes (LANDIS GYR, 2011):

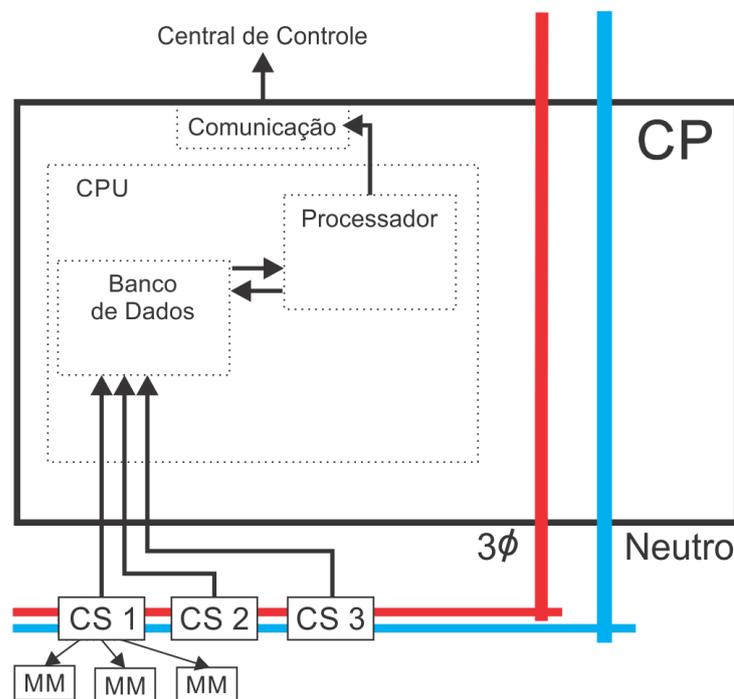
- Barramento - distribuição de energia;
- Disjuntor - proteção e uso em manutenção;
- Conectores para vários Módulos de Medição (MM); e
- CPU - controle e comunicação com os Módulos de Medição (MM) e Concentrador Primário (CP).

Por fim, o Concentrador Primário (CP) é responsável pela comunicação entre as CS e a central de controle remota. Um esquema do CP é ilustrado na Figura 6. Diversos dispositivos podem ser acoplados ao CP, dentre eles, um bloco funcional

para identificação e correção de valores atípicos para melhoria dos dados adquiridos. O diagrama de blocos do CP é composto pelos seguintes componentes (LANDIS GYR, 2011):

- Contator - proteção e segurança;
- Comunicação – ligação com os CS e central de controle (sem fio); e
- CPU - módulo responsável por *polling* de dados de leitura, alarmes e comandos para os Concentradores Secundários (CS).

FIGURA 6. DIAGRAMA EM BLOCOS DO CONCENTRADOR PRIMÁRIO (CP).



Fonte: Elaborada pelo Autor.

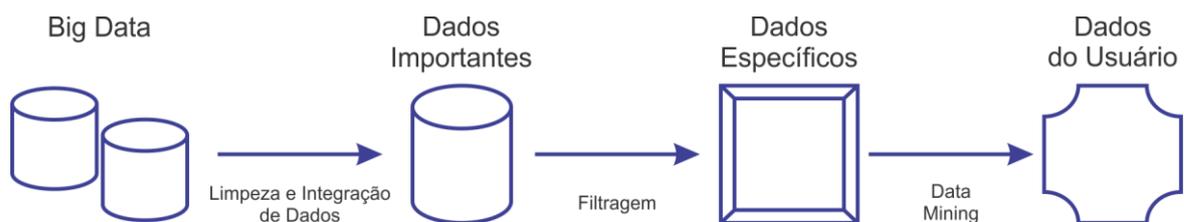
Portanto, os medidores inteligentes atuais são capazes de gerenciar as redes elétricas por meio de captação de dados de consumo, corte ou religação de pontos de energia, dentre diversas outras funções. Entretanto, os dados obtidos podem possuir informações corrompidas ou destoantes da situação real, o que induz o sistema a erros. A proposta do presente trabalho condiz justamente na análise dos dados de demanda de energia e a consequente identificação e correção de valores atípicos.

2.4 BIG DATA ANALYTICS E VALORES ATÍPICOS

A partir do acelerado desenvolvimento das redes elétricas inteligentes, com a aplicação dos conceitos de *Internet of Things* (IoT), tecnologia em nuvem e tecnologia móvel, uma enorme quantidade de dados do setor elétrico tem sido armazenada com potencial para extração de informações relevantes. Para essa possibilidade dá-se o nome de *Big Data Analytics* ou *Data Mining*, ou seja, mineração de dados (WANG, 2015).

O objetivo principal da utilização da mineração de dados em redes elétricas inteligentes é encontrar informações relevantes nos dados de energia. Um esquema, ilustrado na Figura 7, representa uma estrutura típica do processo de mineração de dados (HOU et. al. 2016). Inicialmente, uma densa quantidade de dados é captada pelos dispositivos de medição, *Big Data*. Após passar por um processo de limpeza e integração de dados, tais variáveis são armazenadas em um banco de dados. Em seguida, há uma nova filtragem de dados para a escolha de conjuntos de dados específicos para a mineração de dados, nesse caso, tais dados são os valores atípicos. Por fim, o usuário consegue extrair as informações relevantes para a sua necessidade e possível tomada de decisão.

FIGURA 7. PROCESSO TÍPICO DE MINERAÇÃO DE DADOS.



Fonte: Adaptada de HOU et. al. 2016.

Para a aplicação efetiva do *Big Data Analytics*, diversos algoritmos são postos como ferramentas de extração de informações relevantes em grandes bancos de dados. Entre eles, estão (HOU et. al. 2016):

- Algoritmos básicos de mineração de dados: classificação e análises de dados, Análise de *Cluster*, e Análise de Associação;
- Algoritmos integrados de mineração de dados: Redes Neurais Artificiais - RNA, *Rough Set*, e Lógica Fuzzy.

Tais ferramentas são fundamentais para a tomada de decisões nos investimentos do setor elétrico e monitoramento de condições de equipamentos de potência. Especificamente, as técnicas de mineração de dados se aplicam à construção de curvas de previsão de preços e tarifas na rede elétrica e previsão de demanda de cargas em períodos variados (HUANG, 2004; ZHAN, 2014). Ainda no campo da mineração de dados, uma grande preocupação é a existência de valores atípicos, ou *outliers*, cujas características são inconsistentes com a grande maioria dos dados da série temporal (DINGPING, 2013).

O presente trabalho se propõe a identificar e corrigir valores atípicos, ou seja, dados que estejam fora de um intervalo de confiança estimado pelos algoritmos de identificação, sejam eles do tipo zero, pico ou apenas fora de um intervalo de confiança, que podem ser causados por falhas de comunicação e defeitos nos dispositivos de medição. Nesse caso, ao se detectar o valor atípico, um outro algoritmo é utilizado para corrigir o dado com um valor estimado oriundo de técnicas de inteligência artificial.

Alguns algoritmos de inteligência artificial foram testados em outros trabalhos de pesquisa, como a lógica Fuzzy e as Redes Neurais Artificiais - RNA (ANDRADE et. al., 2018a; ANDRADE et. al., 2018b). Porém, a necessidade de criação de regras por um especialista dificulta a utilização da lógica Fuzzy. Enquanto que, os resultados inconstantes da RNA em horários de pico e fora dele, indica o uso de uma técnica híbrida para a melhoria da correção de valores atípicos.

Com o propósito de melhorar os resultados apresentados na literatura, o presente trabalho aplica o conceito de *Big Data Analytics* e desenvolve um algoritmo baseado no Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS - *Adaptative Neuro Fuzzy Inference System*) capaz de identificar e corrigir valores atípicos em curvas de demanda de energia presentes em redes elétricas inteligentes.

2.5 ALGORITMOS DE INTELIGENCIA ARTIFICIAL

Neste tópico, serão abordados conceitos iniciais dos algoritmos de inteligência artificial. Inicialmente, serão apresentados os Sistemas de Inferência Fuzzy, as Redes Neurais Artificiais e, por fim, os Sistemas ANFIS.

2.5.1 Sistemas de Inferência Fuzzy

Como base para a teoria dos conjuntos fuzzy, encontra-se o conceito de grau de pertinência (ZADEH, 1965), na qual a um objeto é associado um grau de pertinência com valor numérico entre zero e um (ANDRADE, 2010).

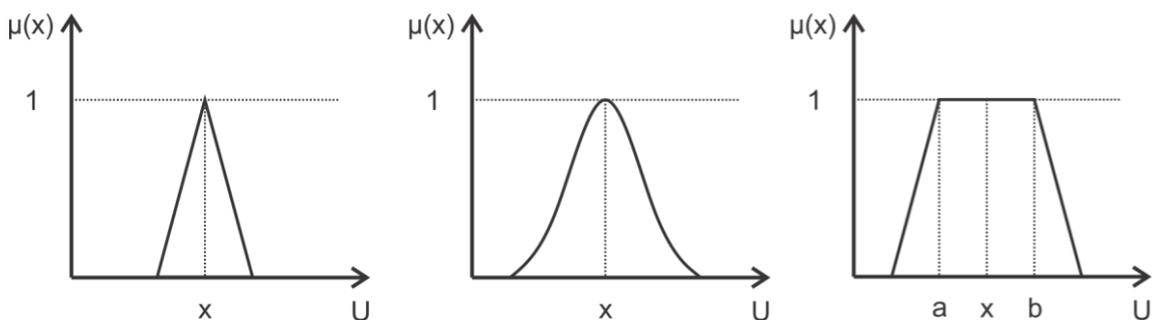
O grau de pertinência na teoria clássica de conjuntos é dado da seguinte forma: dado um conjunto universo U e um elemento $x \in U$, o grau de pertinência $\mu_A(x)$ do elemento x , sendo o conjunto $A \subseteq U$, é:

$$\mu_A(x) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases} \quad (1)$$

Enquanto que, a função de pertinência para os conjuntos fuzzy não é representada de forma binária, ou seja, o grau de pertinência pode assumir valores entre zero e um (ZADEH, 1973).

Como exemplo, estão ilustrados na Figura 8, três tipos de funções de pertinência, são elas: função triangular, função gaussiana e função trapezoidal, respectivamente. Nas funções triangular e gaussiana há uma indicação de que, quanto mais próximo do valor de x , maior o seu grau de pertinência em relação ao conjunto A . Enquanto que, para a função trapezoidal, se o elemento estiver entre os valores de a e b , há um grau de pertinência igual a 1.

FIGURA 8. FUNÇÕES DE PERTINÊNCIA FUZZY.



Fonte: Adaptada de ANDRADE, 2010.

Dessa forma, os conjuntos fuzzy são representados por pares ordenados, na qual o elemento é associado ao seu grau de pertinência em relação ao conjunto A , conforme expressão 2.

$$A = \{ (x, \mu_A(x)) \mid x \in U \} \quad (2)$$

Para a definição de regras, o método mais utilizado é o composto por antecedentes e consequentes. Nas quais, os antecedentes representam as condições do sistema, enquanto que, os consequentes, indicam os resultados para tais condições, conforme a expressão 3. De modo que, no sistema de inferência fuzzy pode haver a agregação de inúmeras regras, que são ponderadas com um valor relativo à importância de cada regra para o resultado final (ZADEH, 1973).

$$\textit{Se} < \textit{antecedentes} > \textit{ ENTÃO} < \textit{consequentes} > \quad (3)$$

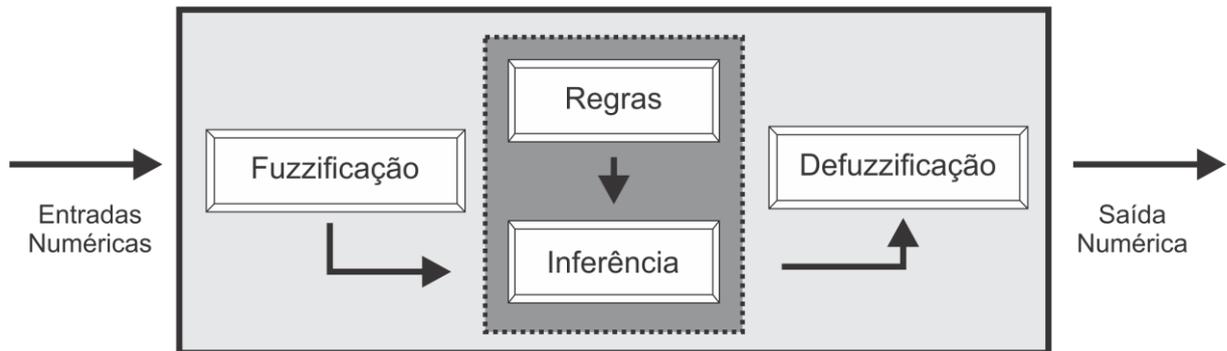
Dessa forma, o sistema de inferência fuzzy pode ser representado por três etapas, são elas: fuzzificação, inferência e defuzzificação, conforme apresentado na Figura 9 (ANDRADE, 2010).

Na etapa de fuzzificação, os dados numéricos de entrada do sistema passam por um processo que determina as variáveis linguísticas a que pertencem os dados numéricos e seus respectivos graus de pertinência. Dessa forma, variáveis numéricas são transformadas em variáveis linguísticas, como por exemplo: alto, médio e baixo. E, em seguida, associadas a funções de pertinência de acordo com seu grau de pertencimento a estas funções.

Na etapa de inferência, as variáveis linguísticas são combinadas afim de gerar valores linguísticos de saída de acordo com as regras de inferência fuzzy fornecidas por especialistas ou extraídas de dados numéricos. No fornecimento de regras com dependência do conhecimento de um especialista, há uma certa dificuldade. Porém, com a proposição de sistemas híbridos, existe a possibilidade da interação entre os sistemas de inferência fuzzy e as redes neurais artificiais.

Na etapa de defuzzificação, finalmente, há uma interpretação do conjunto de saída, formado pelo sistema de inferência, para a determinação de saídas numéricas do sistema de inferência fuzzy.

FIGURA 9. ETAPAS DO SISTEMA DE INFERÊNCIA FUZZY.



Fonte: Adaptada de ANDRADE, 2010.

2.5.2 Redes Neurais Artificiais

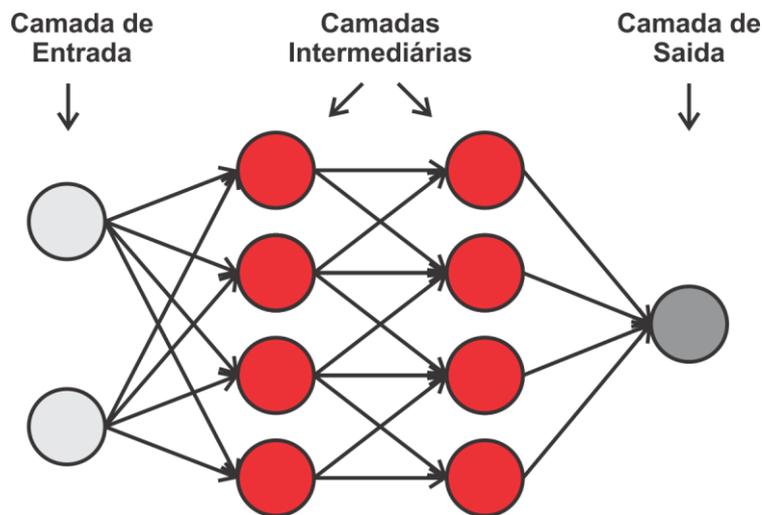
O modelo matemático das Redes Neurais Artificiais (RNA) foi proposto inicialmente no ano de 1943 por McCulloch e Pitts (McCULLOCH et. al., 1943). Nesse trabalho, o sistema computacional proposto foi inspirado na estrutura neural de organismos inteligentes. Essa rede interconectada de neurônios é bastante utilizada em problemas de classificação de dados (GHANBARI et. al., 2016).

A arquitetura básica neural é organizada em camadas, nas quais os elementos podem estar conectados às camadas posteriores com atribuição de pesos, que são ajustados a partir de experiências anteriores. Conforme a Figura 10, a Camada de Entrada possui os padrões apresentados à rede, as Camadas Intermediárias, ou ocultas, possuem as características de processamento por meio de conexões ponderadas e, por fim, a camada de saída representa o resultado final da rede (GHANBARI et. al., 2016; McCULLOCH et. al., 1943).

O processamento de informações nas Redes Neurais Artificiais (RNA) é baseado no seguinte algoritmo: inicialmente, os sinais de entrada são introduzidos no sistema por meio da Camada de Entrada; a partir de então, nas Camadas Intermediárias, cada sinal é multiplicado por um número, representando um peso, que indica o quanto aquele dado irá influenciar no resultado final; em seguida, uma soma ponderada dos sinais produz um nível de atividade; e, por fim, na Camada de Saída, uma resposta final é produzida na saída (GHANBARI et. al., 2016; McCULLOCH et. al., 1943).

O algoritmo *Backpropagation*, por exemplo, baseado em Redes Neurais Artificiais (RNA) pode ser utilizado para classificar comportamentos normais e anômalos em dados de consumo de energia elétrica em redes elétricas inteligentes (GHANBARI et. al., 2016).

FIGURA 10. ARQUITETURA BÁSICA DAS REDES NEURAIS ARTIFICIAIS (RNA).



Fonte: Elaborada pelo Autor.

2.5.3 Sistemas ANFIS

O Sistema ANFIS (*Adaptive Neuro-fuzzy Inference System*) foi proposto por JANG no ano de 1993 e representa um sistema de inferência fuzzy formado por redes adaptativas com capacidade de aprendizado supervisionada (JANG, 1993).

2.5.3.1 Características ANFIS

O sistema híbrido ANFIS associa a possibilidade de extração de conhecimento na forma de regras da Lógica Fuzzy e a capacidade de aprendizado das Redes Neurais Artificiais (SANCHEZ, 2009; SULAIMAN et. al., 2016).

A contribuição da Lógica Fuzzy ao sistema híbrido pode ser dividida em quatro partes, são elas:

- Modelo Fuzzy: determina o formato das regras fuzzy, sendo o modelo de inferência (Mandami, Takagi-Sugeno, etc);

- Funções de Pertinência: determina os graus de pertinência associados a cada variável (Triangular, Trapezoidal, Gaussiana, etc);
- Particionamento do Espaço de Entradas e Saídas (E/S): define internamente regiões fuzzy de espaço, relacionados por meio de regras fuzzy (Fuzzy grid, BSP, Quad Tree, etc); e
- Método de Defuzzificação: resultado final numérico definido após avaliação do conjunto de regras fuzzy (Média Ponderada, Centro de Soma, Centro de Área, etc).

Assim como, as características das Redes Neurais Artificiais relevantes para o sistema híbrido ANFIS são as seguintes:

- Tipo de Aprendizado: determinante na atualização dos parâmetros da estrutura, podendo ser a cada dado apresentado ou após a apresentação de todos os dados do conjunto de treinamento (*Online* ou *Offline*, respectivamente);
- Identificação da estrutura: define o número adequado de regras fuzzy e o particionamento satisfatório das entradas e saídas. Tal identificação pode ser realizado com o auxílio de um especialista ou método auto-organizado antes do início da atualização dos parâmetros (Identificação Prévia), ou sem o conhecimento prévio por meio incremental ou decremental (Identificação Automática); e
- Identificação dos parâmetros: relacionado ao ajuste dos pesos fuzzy referentes às funções de pertinência dos antecedentes e consequentes das regras fuzzy (Antecedentes e Consequentes).

2.5.3.2 Arquitetura ANFIS

Um dos modelos mais comuns para aplicação do sistema ANFIS é o modelo de Takagi-Sugeno (JANG, 1995; SULAIMAN et. al., 2016). Tal modelo é baseado na atribuição de valores ponderados às regras ativadas no modelo de fuzzificação. Dessa forma, as seguintes etapas são executadas: fuzzificação das entradas, encontro das regras ativadas e determinação dos valores das funções de saída. Por fim, a partir de tais valores de saída, após ponderação dos mesmos, é gerada uma saída numérica (ANDRADE, 2010).

Assim, um sistema de inferência fuzzy composto por duas entradas x e y e uma saída z , de primeira ordem, conforme está ilustrado na Figura 11, a partir do modelo Takagi-Sugeno, é definido da seguinte forma:

Regra 1: Se x é A_1 e y é B_1 então $f_1 = p_1x + q_1y + r_1$

Regra 2: Se x é A_2 e y é B_2 então $f_2 = p_2x + q_2y + r_2$

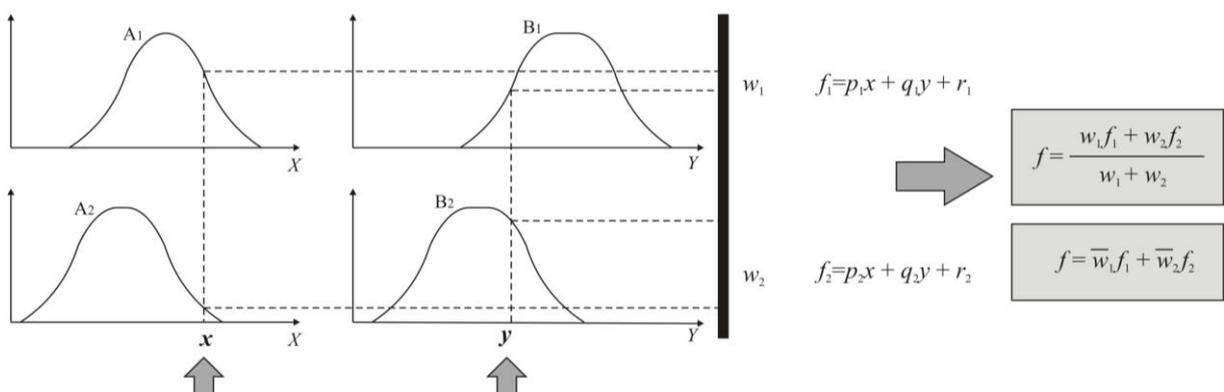
Sendo A_1 e A_2 funções de pertinência, f_1 e f_2 funções de ativação compostas pelos parâmetros p , q e r e dependentes das entradas x e y .

Dessa forma, os graus de pertinência w_i são determinados pela análise de o quanto as entradas x e y fazem parte das funções de pertinência A_1 e A_2 . Em seguida, as funções de ativação f_1 e f_2 de cada regra gerada são classificadas como saída, tendo como resultado final a média ponderada de todos esses resultados, conforme expressão 4 a seguir:

$$f = \frac{\sum_{i=1}^N w_i \cdot f_i}{\sum_{i=1}^N w_i} \quad (4)$$

na qual f é a saída final, N é a quantidade de regras ativadas e w_i é o grau de pertinência de cada regra.

FIGURA 11. MODELO DO SISTEMA DE INFERÊNCIA TAKAGI-SUGENO.



Fonte: JANG, 1993.

Para a definição da arquitetura do Sistema ANFIS, foi utilizado o modelo do tipo Takagi-Sugeno, ilustrado na Figura 12. Tal modelo possui na sua estrutura um

total de cinco camadas, além da disposição de entradas e saídas. As camadas e suas respectivas funções serão descritas a seguir:

- Camada 1: a partir das entradas x e y , a saída dessa camada são os graus de pertinência das entradas, baseada na premissa de cada regra (SANCHEZ, 2009). Todos os nós i da camada são nós adaptativos, tendo como nós de saída (JANG, 1995):

$$\begin{aligned} O_{l,i} &= \mu_{A_i}(x), & \text{para } i = 1,2 \\ O_{l,i} &= \mu_{B_{i-2}}(y), & \text{para } i = 3,4 \end{aligned} \quad (5)$$

na qual o nó de saída i da camada l é $O_{l,i}$, com x (ou y) como entrada do nó e A_i (ou B_{i-2}) é o conjunto fuzzy associado a este nó. Como exemplo, A_i pode ser caracterizado pela função gaussiana, como a seguir:

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^{b_i}} \quad (6)$$

tendo como parâmetros do conjunto a_i , b_i e c_i . Sendo a_i a metade da largura da função de pertinência, b_i é a inclinação no ponto 0,5 da função de pertinência e c_i determina o centro da função de pertinência (ANDRADE, 2010).

- Camada 2: nesse ponto, é calculado o grau de pertinência ao qual é submetido o conseqüente de cada regra (SANCHEZ, 2009). Cada nó desta camada é fixo e nomeado pelo operador \prod , indicando que há uma multiplicação entre os sinais de entrada, como a seguir:

$$O_{2,i} = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y), \quad i = 1, 2. \quad (7)$$

- Camada 3: uma normalização dos graus de ativação das regras é feita (SANCHEZ, 2009). Assim, cada nó nesta camada é nomeado por N , na qual o i -ésimo nó calcula a razão entre a ativação da i -ésima regra pela ativação de todas as regras:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \quad (8)$$

- Camada 4: a saída de cada neurônio é calculada pelo produto da saída normalizada da camada anterior e o grau de ativação do consequente (SANCHEZ, 2009). Todos os nós i da camada são nós adaptativos, com um nó de função:

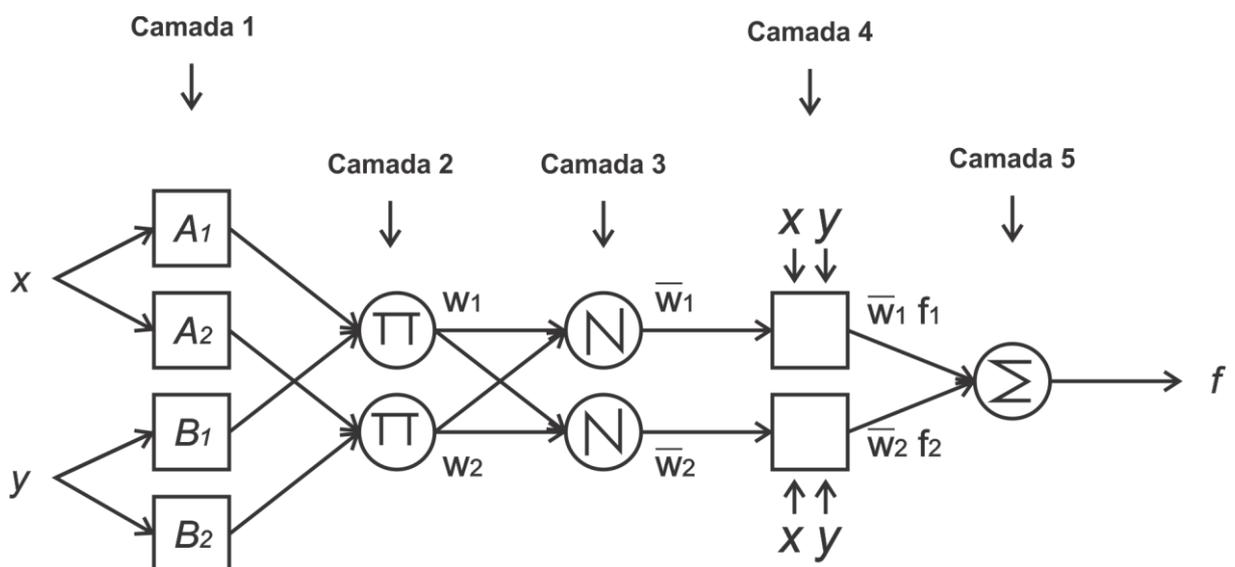
$$O_{4,i} = \bar{w}_i \cdot f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (9)$$

na qual \bar{w}_i é a saída da camada 3 e $\{p_i, q_i, r_i\}$ os parâmetros do conjunto. Tais parâmetros serão referenciados como parâmetros consequentes.

- Camada 5: tal camada fornece a saída precisa do sistema ANFIS (SANCHEZ, 2009). O nó único desta camada é um nó fixo definido como a somatória Σ , que calcula a saída total como a soma de todos os sinais de entrada relacionados com os parâmetros encontrados:

$$O_{5,i} = \sum_i \bar{w}_i \cdot f_i = \frac{\sum_i w_i \cdot f_i}{\sum_i w_i} \quad (10)$$

FIGURA 12. ARQUITETURA DO SISTEMA DE INFERÊNCIA ANFIS TAKAGI-SUGENO.



Fonte: Adaptada de JANG, 1993.

Na determinação do espaço de particionamento das Entradas e Saídas (E/S) do sistema ANFIS, é utilizado o método Fuzzy Grid Partition, na qual duas etapas são realizadas de forma repetitiva até atingidos os critérios de parada, tais como Tolerância de Erro e Número de Épocas (SANCHEZ, 2009):

- Etapa 1: os parâmetros dos consequentes são ajustados pelo método MQO (Mínimos Quadrados Ordinários), permanecendo fixos os antecedentes;
- Etapa 2: os parâmetros dos consequentes são ajustados pelo método do GD (Gradiente Decrescente), permanecendo fixos os consequentes.

Assim, está posta uma rede adaptativa com a função de um modelo fuzzy Takagi-Sugeno por meio de uma rede neural.

3 METODOLOGIA

3 METODOLOGIA

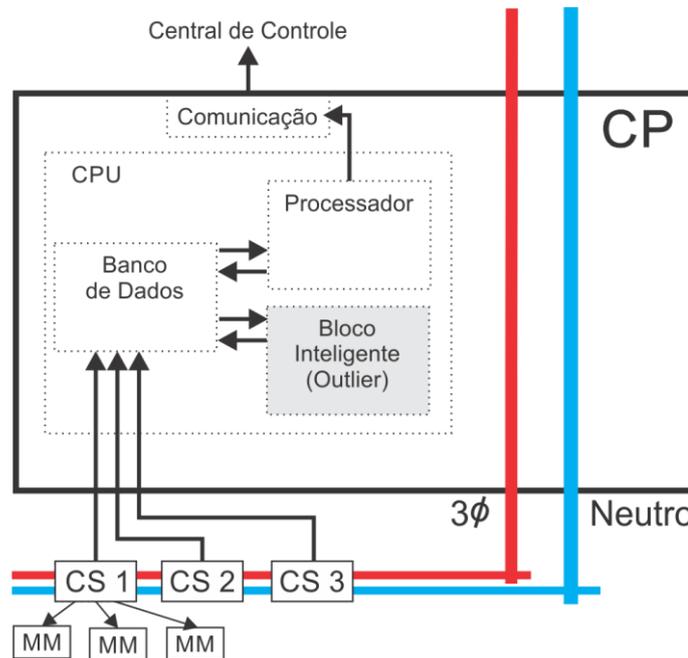
Nesta seção, será apresentada uma nova topologia de medidores de energia com inteligência computacional para identificação e correção de valores atípicos para medições de demanda de energia utilizando um algoritmo baseado no Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS). Complementarmente, com a finalidade de comparação, serão apresentados os procedimentos de identificação e correção de valores atípicos usando algoritmos baseados nas técnicas de interpolação linear e Redes Neurais Artificiais (RNA).

3.1 TOPOLOGIA PROPOSTA DE SMART METERS

Na seção 2.3, foi apresentada uma topologia padrão para o dispositivo de medição Concentrador Primário (CP), responsável pela comunicação entre os Concentradores Secundários (CS), presentes nos postes e que reúne informações e controla os Módulos de Medição (MM) instalados nas unidades consumidoras, e a central de controle remota, localizada na concessionária de energia.

A fim de que seja possível a correção e identificação de outliers, o presente trabalho sugere uma nova topologia para o Concentrador Primário (CP). Desta vez, sendo composto com um Bloco Inteligente, na qual seja possível a implementação de algoritmos inteligentes, especificamente o Sistema ANFIS ou Redes Neurais Artificiais. Um esquema da nova topologia do CP é ilustrado na Figura 13.

FIGURA 13. DIAGRAMA EM BLOCOS DE UMA NOVA TOPOLOGIA DE MEDIDORES INTELIGENTES.



Fonte: Elaborada pelo Autor.

3.2 PERFIL DAS CURVAS DE DEMANDA

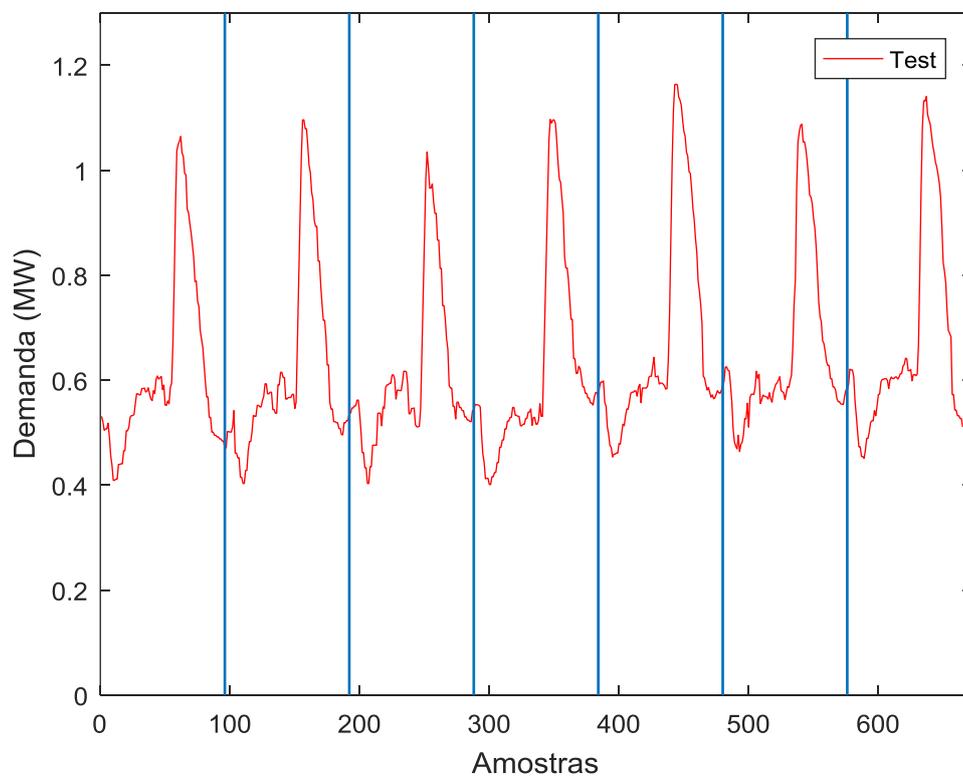
Neste tópico, será apresentado um perfil das curvas de demanda de energia geradas por meio da aquisição de dados em uma subestação. Tais dados são reais e foram fornecidos pela companhia de energia local referente à uma região do Estado da Paraíba. A partir de um sistema de aquisição e monitoramento de dados, com uma taxa de amostragem de 4 amostras por hora, o número total de amostras coletadas para o registro diário totaliza 96 medições.

A Figura 14 ilustra a representação de uma curva de demanda com a quantidade de 672 amostras, referente a 7 dias. Observa-se que há um padrão entre os diferentes dias da semana, com horários de pico se concentrando no início do período noturno, 18 horas. Os valores de potência dessa curva estão todos de acordo com a normalidade, ou seja, não possuem valores atípicos de pico ou zero.

A Figura 15 ilustra uma curva de demanda semelhante, no entanto, com a inserção de valores atípicos do tipo zero em 10% dos valores, ou seja, em 67 amostras.

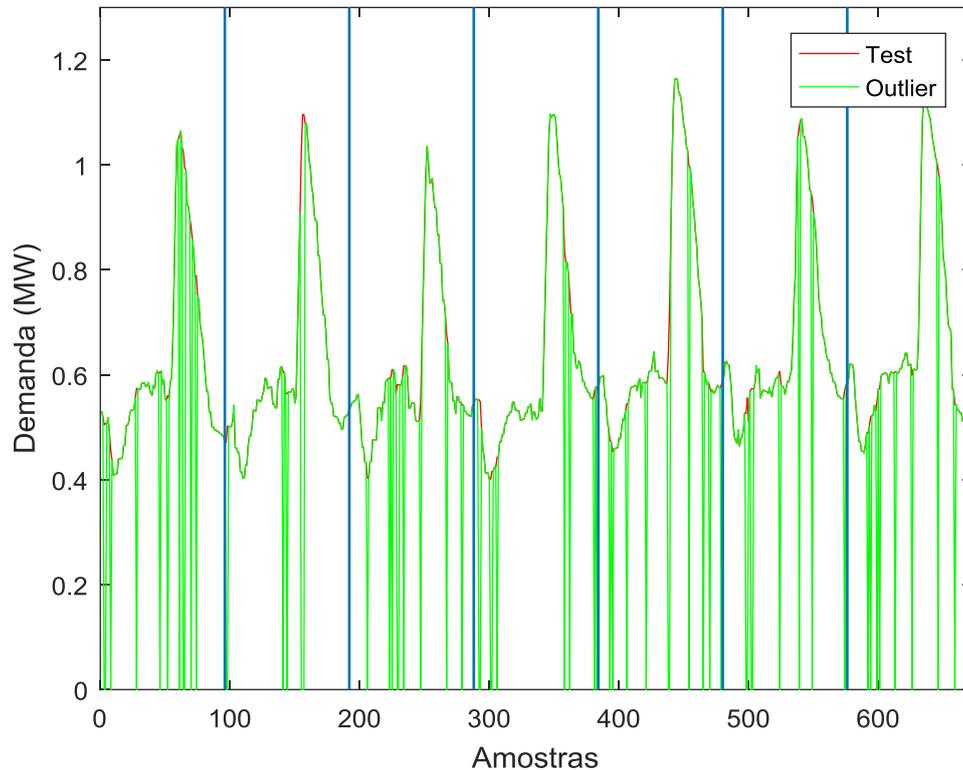
Assim, percebe-se que os valores atípicos, ou outliers, produzem curvas com padrões distorcidos, que podem influenciar negativamente no levantamento de informações por meio da mineração de dados. Por exemplo, ao realizar análises estatísticas, serão fornecidos resultados tendenciosos, pois entre os dados válidos, se encontram diversos valores nulos. Desta forma, o presente trabalho contribui para a identificação desses valores atípicos e sua consequente correção por meio de algoritmos de inteligência artificial, especificamente o ANFIS.

FIGURA 14. CURVA DE DEMANDA DE POTÊNCIA.



Fonte: Elaborada pelo Autor.

FIGURA 15. CURVA DE DEMANDA COM INSERÇÃO DE OUTLIERS DO TIPO ZERO.

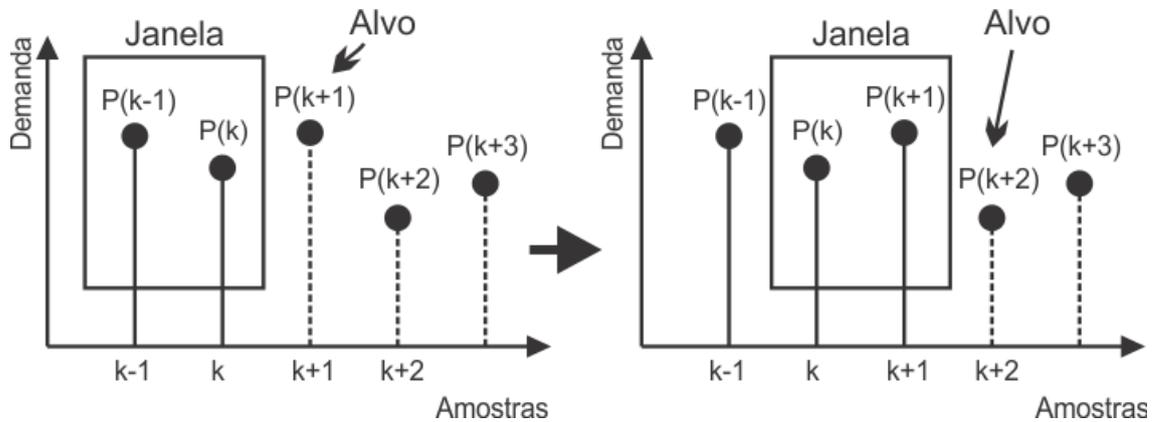


Fonte: Elaborada pelo Autor.

3.3 TRATAMENTO E PREPARAÇÃO DOS DADOS

A partir deste histórico de demanda, foi construído um banco de dados composto por duas matrizes que correspondem às entradas e saídas do módulo de correção de outliers. Para a construção das referidas matrizes, foi utilizada a metodologia de janelas deslizantes, como ilustrado na Figura 16, nas quais as amostras passada, $P_{(k-1)}$, e presente, $P_{(k)}$, de potência são utilizadas para a estimação de um valor futuro, $P_{(k+1)}$. A medida que o valor futuro é encontrado, a janela se desloca no sentido crescente e utiliza os valores seguintes, $P_{(k)}$ e $P_{(k+1)}$ para a estimação do valor futuro, $P_{(k+2)}$. Verifica-se que, para previsões de curtíssimo prazo, os valores mais próximos são mais significativos para a previsão do valor seguinte (ANDRADE, 2018a).

FIGURA 16. TRATAMENTO E PREPARAÇÃO DOS DADOS POR JANELAS DESLIZANTES.



Fonte: Elaborada pelo autor.

Para a determinação do tamanho da janela de valores utilizados como entrada para prever uma saída, foram simulados os valores de janela com 2 e 3 valores, sendo constatado por meio da análise de Erros Relativos Absolutos, conforme Tabela 1, que o valor ótimo referente a erro máximo percentual absoluto e desvio padrão foi com 2 amostras. Para tal simulação, foram utilizados os dados descritos no item 3.2, na qual houve a inserção de outliers em 10% das amostras totais. Os algoritmos utilizados para teste foram Interpolação Linear (Inter), Redes Neurais Artificiais (RNA) e Sistema de Inferência Neuro-Fuzzy Adaptativo (ANFIS).

TABELA 1. ERROS RELATIVOS NA ESTIMAÇÃO DA JANELA DE VALORES.

JANELA (AMOSTRAS)	ERROS RELATIVOS ABSOLUTOS (%)		
	<i>Inter</i>	<i>RNA</i>	<i>ANFIS</i>
2	4,5577	7,6979	3,7616
3	6.4272	31.778	26.384

Fonte: Dados da Pesquisa.

Assim, foram geradas matrizes com os valores de entrada (Input) e saída (Target). Devido ao fato de que são 96 amostras por dia, as matrizes possuem 4697 exemplos disponíveis para a construção de sistemas recursivos. Dessa forma, o banco de dados foi dividido da seguinte maneira, 80% para treinamento e 20% para teste.

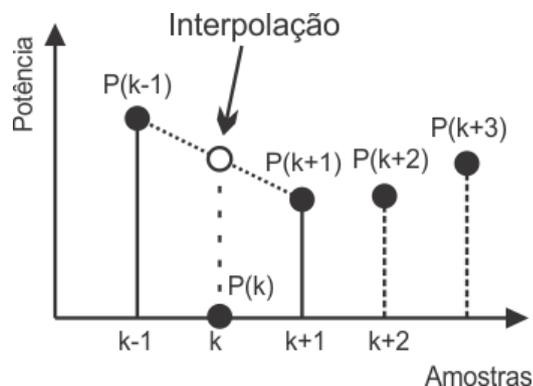
3.4 ALGORITMO DE INTERPOLAÇÃO

O primeiro algoritmo desenvolvido para a correção de outliers foi o de interpolação linear, modo computacional mais rápido utilizado usualmente na resolução de valores atípicos nas mais diversas aplicações.

Um modelo do algoritmo de Interpolação pode ser visualizado na Figura 17, no qual ao ser detectado o outlier, o valor a ser preenchido, $P_{(k)}$, é igual ao resultado da média aritmética entre a amostra passada, $P_{(k-1)}$, e a amostra futura, $P_{(k+1)}$.

O principal problema desse método é na ocasião de outliers seguidos, pois possui grande chance de aumento do erro relativo percentual. Para a solução parcial desse ponto, foi incrementado ao algoritmo a detecção de valor atípico nos valores passados $P_{(k-1)}$, $P_{(k-2)}$, e $P_{(k-3)}$, assim como nos valores futuros $P_{(k+1)}$, $P_{(k+2)}$, e $P_{(k+3)}$, para que, caso o valor adjacente a $P_{(k)}$ seja atípico, o próximo valor é verificado e assim sucessivamente, até que seja utilizado um valor válido. Dessa forma, há solução válida para o intervalo de dados de até 90 minutos, ou seja, seis amostras adjacentes.

FIGURA 17. MODELO DO ALGORITMO DE INTERPOLAÇÃO.

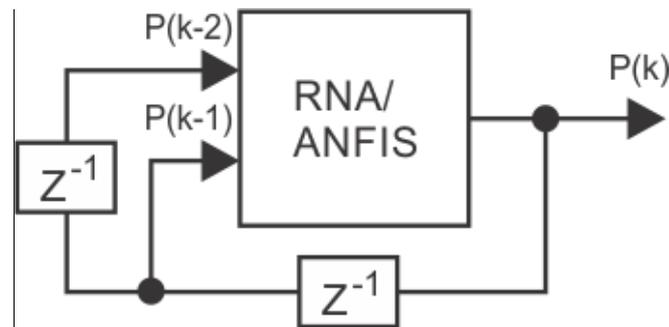


Fonte: Elaborada pelo autor.

3.5 ALGORITMOS RNA E ANFIS

Para os algoritmos baseados em RNA e ANFIS, foram utilizadas duas amostras passadas (entrada) e uma futura (saída) para treinamento e teste, conforme esquema de preparação de dados apresentado na Figura 18. Diferentemente da proposta de interpolação linear, que utiliza uma amostra passada e outra futura para a estimação do valor desejado.

FIGURA 18. MODELO DOS ALGORITMOS DE RNA E ANFIS.



Fonte: Elaborada pelo autor.

3.5.1 Algoritmo RNA

Na determinação dos parâmetros utilizados para RNA, foram feitos alguns testes para encontrar as quantidades de neurônios na rede. O método do máximo erro percentual absoluto foi utilizado, conforme as Tabelas 2 e 3.

Na Tabela 2, há uma análise do número de neurônios para duas camadas RNA. Os testes foram iniciados com a quantidade de 2 neurônios em cada camada e se estendeu até a quantidade de 17 e 18 neurônios nas camadas 1 e 2, respectivamente. Assim como, na Tabela 3, a mesma análise foi realizada variando-se a quantidade de neurônios em 3 camadas, iniciando os testes com a quantidade de 2 neurônios nas três camadas, até os números de 15, 16 e 17 neurônios nas camadas 1, 2 e 3, respectivamente.

Dentre todos os Máximos Erros Relativos Absolutos (%) encontrados, o melhor resultado encontrado se deu na utilização da quantidade de 15 neurônios em ambas as camadas, com o valor de 3,8327 %.

TABELA 2. ANÁLISE DO NÚMERO DE NEURÔNIOS PARA DUAS CAMADAS RNA.

Camada 1	Camada 2	Máximos Erros Relativos Absolutos (%)
		RNA
2	2	8,7038
2	3	8,0180
3	3	5,8179
3	4	5,3783
4	4	6,6070
4	5	6,1818
5	5	4,8391
5	6	7,6172
6	6	6,5627
6	7	7,3283
7	7	8,6393
7	8	6,3965
8	8	6,8186
8	9	8,0485
9	9	3,9991
9	10	4,3467
10	10	6,0743
10	11	4,6137
11	11	6,9338
11	12	6,6940
12	12	4,7725
12	13	5,7112
13	13	7,6803
13	14	4,5639
14	14	5,0140
14	15	4,4656
15	15	3,8327
15	16	5,2702
16	16	7,0126
16	17	6,2271
17	17	5,3846
17	18	9,7408

Fonte: Dados da Pesquisa.

TABELA 3. ANÁLISE DO NÚMERO DE NEURÔNIOS PARA TRÊS CAMADAS RNA.

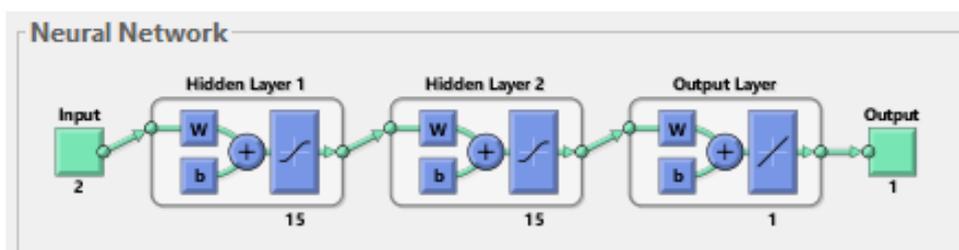
Camada 1	Camada 2	Camada 3	Máximos Erros Relativos Absolutos (%)
			RNA
2	2	2	7,5442
2	2	3	5,9023
2	3	4	8,4474
3	3	3	6,7976
3	3	4	7,7223
3	4	5	6,3720
4	4	4	5,3767
4	4	5	7,5324
4	5	6	5,7126
5	5	5	5,4092
5	5	6	4,0506
5	6	7	3,9603
6	6	6	4,9817
6	6	7	5,3503
6	7	8	5,5171
7	7	7	6,6171

7	7	8	6,0540
7	8	9	6,6168
8	8	8	9,3186
8	8	9	6,6228
8	9	10	4,9229
9	9	9	5,5759
9	9	10	7,1068
9	10	11	9,8683
10	10	10	4,7712
10	10	11	5,7192
10	11	12	5,1333
11	11	11	4,5130
11	11	12	6,7532
11	12	13	5,9258
12	12	12	3,8615
12	12	13	6,3884
12	13	14	5,0531
13	13	13	6,0719
13	13	14	5,6834
13	14	15	5,2846
14	14	14	6,3296
14	14	15	5,7860
14	15	16	10,4660
15	15	15	8,1646
15	15	16	4,1450
15	16	17	5,2787

Fonte: Dados da Pesquisa.

Dessa forma, para a RNA, conforme a Figura 19, foram utilizadas 2 camadas ocultas, com 15 neurônios nas Camadas 1 e 2 e, por fim, 1 neurônio de saída. Como parâmetros de parada no treinamento foram determinados o número máximo de épocas em 100 e erro mínimo de $10e-3$.

FIGURA 19. MODELO DA REDE NEURAL ARTIFICIAL.



Fonte: Elaborada pelo autor.

3.5.2 Algoritmo ANFIS

Na determinação dos parâmetros utilizados para o algoritmo ANFIS, o método do máximo erro percentual absoluto também foi utilizado. O número de neurônios dos conjuntos de pertinências foi determinado conforme Tabela 4, visto que a quantidade

de neurônios foi variada entre os números 2 a 6, e o melhor resultado encontrado foi de 2 neurônios.

Enquanto que, para a determinação do tipo de conjunto de pertinência foram realizados diversos testes, conforme a Tabela 5, com os principais tipos apresentados nesse trabalho, tais como triangular (trimf), trapezoidal (trapmf), e gaussiana (gbellmf, gaussmf e gauss2mf), sendo o resultado mais satisfatório o tipo gauss2mf.

TABELA 4. ANÁLISE DO NÚMERO DE NEURÔNIOS DO CONJUNTO DE PERTINÊNCIA ANFIS.

Neurônios	Máximos Erros Relativos Absolutos (%)
	ANFIS
2	3.6957
3	3.7444
4	3.7523
5	3.7096
6	4.3003

Fonte: Dados da Pesquisa.

TABELA 5. ANÁLISE DOS TIPOS DE CONJUNTO DE PERTINÊNCIA ANFIS.

Conjunto de Pertinência	Máximos Erros Relativos Absolutos (%)
	ANFIS
trimf	4.6854
trapmf	3.6805
gbellmf	3.6957
gaussmf	3.6873
gauss2mf	3.5117

Fonte: Dados da Pesquisa.

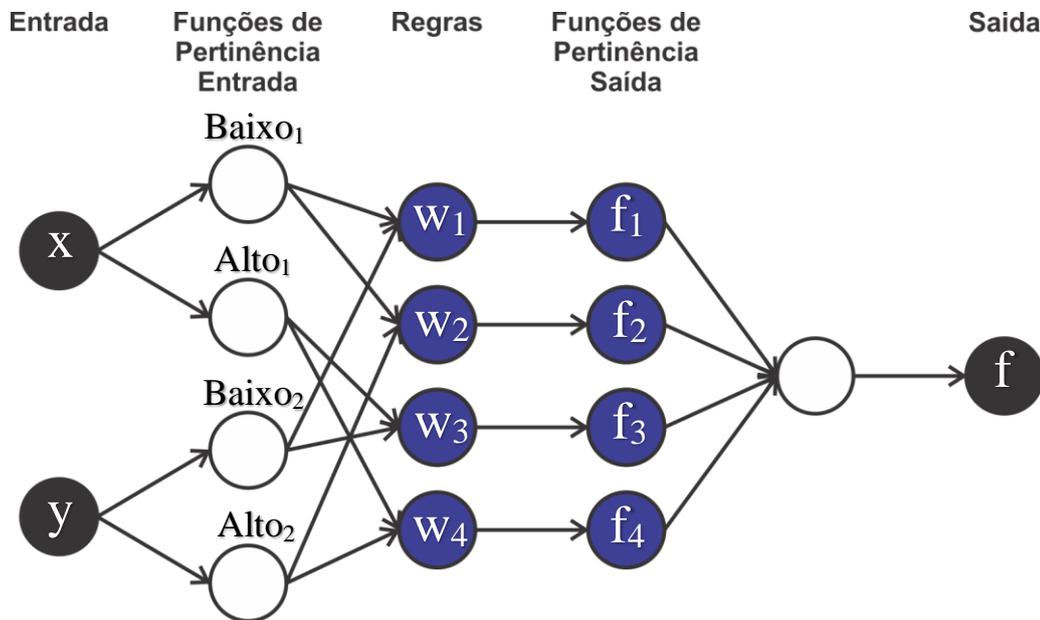
Dessa forma, para o sistema ANFIS foi escolhido o método *Grid Partition* com 2 conjuntos de pertinência para ambas as entradas do tipo 'gauss2mf'.

Para iniciar o treinamento, foi utilizado o método de otimização híbrido, que utiliza a combinação dos métodos de *backpropagation* e mínimos quadrados. E, por fim, os parâmetros de parada do treinamento, número de épocas igual a 20 e erro desejado igual a zero.

A arquitetura do sistema ANFIS pode ser visualizada na Figura 20. Na camada 1 há dois neurônios como entradas (x e y); na camada 2 existem 4 neurônios correspondentes às funções de pertinência de entrada (*Baixo1*, *Alto1*, *Baixo2* e *Alto2*); na camada 3, há 4 neurônios, indicando a normalização dos graus de ativação das regras (w_1 , w_2 , w_3 e w_4); na camada 4, existem 4 neurônios que indicam as

funções de pertinência de saída (f_1, f_2, f_3 e f_4), nesse caso, constantes; e, por fim, na camada 5, há uma saída precisa do sistema ANFIS (f).

FIGURA 20. ARQUITETURA DO SISTEMA ANFIS.



Fonte: Elaborada pelo autor.

Após a determinação da arquitetura ANFIS e do treinamento da rede, foram encontradas 4 regras para o presente sistema, com peso 1, elencadas a seguir:

Regra 1: *Se* (x é *Baixo1*) e (y é *Baixo2*) então (f é f_1) (1)

Regra 2: *Se* (x é *Baixo1*) e (y é *Alto2*) então (f é f_2) (1)

Regra 3: *Se* (x é *Alto1*) e (y é *Baixo2*) então (f é f_3) (1)

Regra 4: *Se* (x é *Alto1*) e (y é *Alto2*) então (f é f_4) (1)

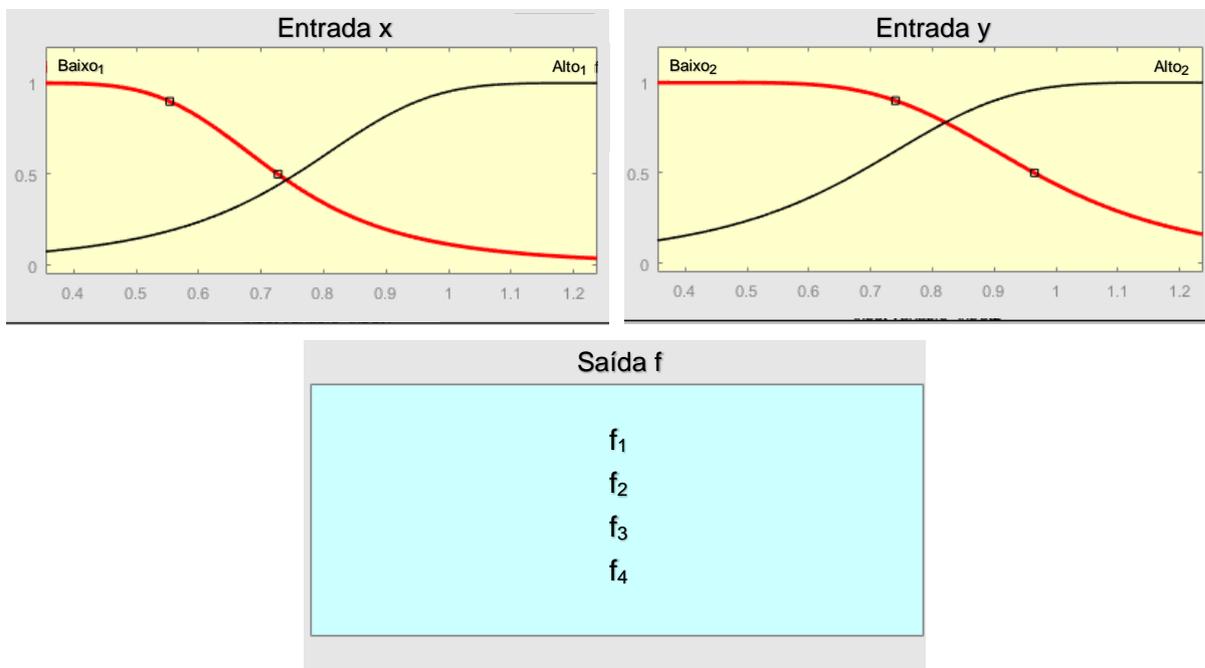
Nesse caso, tais regras indicam como a rede está estabelecida. Por exemplo, a Regra 1 indica que, se a x está contida no conjunto de pertinência *Baixo1* e a entrada y está contida no conjunto de pertinência *Baixo2* (antecedentes), então a saída f estará no conjunto de pertinência de saída f_1 (consequente). Dessa forma, tais regras podem ser esquematizadas da seguinte forma:

$$\text{Regra 1: } \begin{cases} \text{Antecedente} : [1 \ 1] \\ \text{Consequente} : 1 \\ \text{Peso} : 1 \\ \text{Conexão} : 1 \end{cases} \qquad \text{Regra 2: } \begin{cases} \text{Antecedente} : [1 \ 2] \\ \text{Consequente} : 2 \\ \text{Peso} : 1 \\ \text{Conexão} : 1 \end{cases}$$

$$\text{Regra 3: } \begin{cases} \text{Antecedente : [2 1]} \\ \text{Consequente: 3} \\ \text{Peso:} & : 1 \\ \text{Conexão} & : 1 \end{cases} \quad \text{Regra 4: } \begin{cases} \text{Antecedente : [2 2]} \\ \text{Consequente: 4} \\ \text{Peso:} & : 1 \\ \text{Conexão} & : 1 \end{cases}$$

Os conjuntos de pertinência das entradas e saída se encontram na Figura 21, conforme se observa, há duas funções de pertinência (*Baixo1* e *Alto1*) e (*Baixo2* e *Alto2*) do tipo “gauss2” em cada uma das entradas (*x* e *y*) e quatro funções de pertinência na saída do tipo constante (*f1*, *f2*, *f3* e *f4*).

FIGURA 21. CONJUNTOS DE PERTINÊNCIA DAS ENTRADAS E SAÍDA ANFIS.



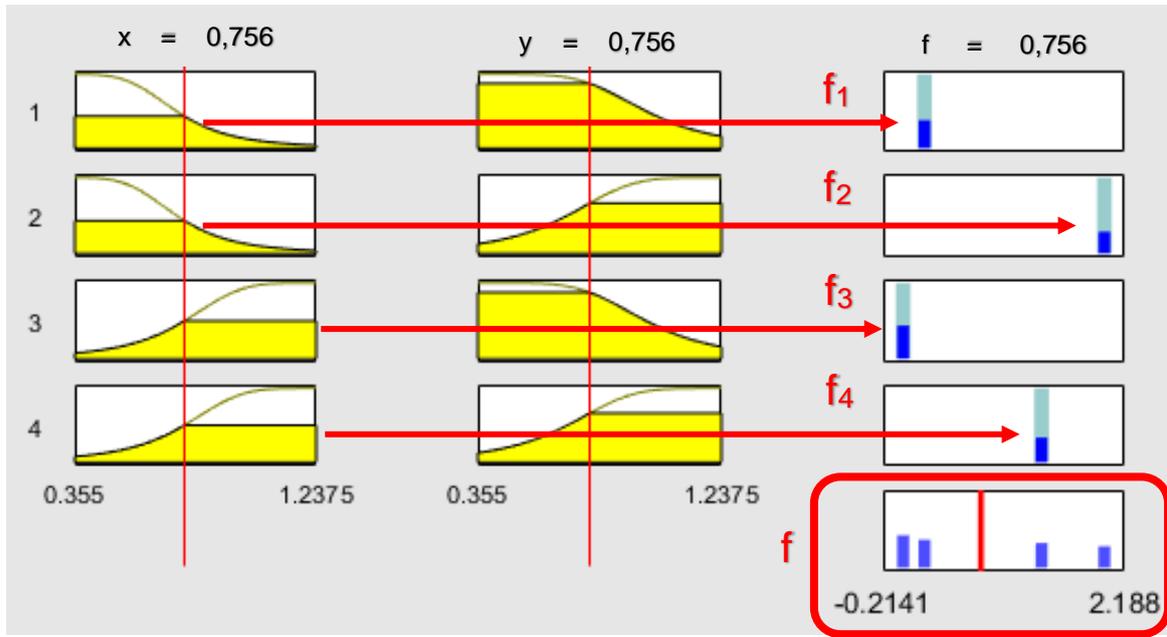
Fonte: Elaborada pelo autor.

Para a visualização das regras ANFIS e de suas aplicações nas entradas para determinação da saída, foi gerada a Figura 22. Percebe-se claramente as quatro regras aplicadas às duas entradas (*x* e *y*), assim como, as funções de pertinência de saída constantes (*f1*, *f2*, *f3* e *f4*) e, por fim, o resultado preciso da saída (*f*), como um valor ponderado das saídas anteriores. Para a determinação das funções de saída (*f_i*), o operador “e” foi utilizado na formação nas regras, dessa forma, utiliza-se o valor mínimo do grau de pertinência.

O exemplo com entradas e saídas da Figura 22 ilustra a aplicação das regras do sistema ANFIS. Observa-se que tanto a entrada *x*, como *y*, estão contidas no

intervalo de 0,355 a 1,2375. E que os graus de pertinência de menor valor em cada regra indicam os valores das saídas (f_1 , f_2 , f_3 e f_4), assim como, o valor final da saída numérica (f).

FIGURA 22. APLICAÇÃO DAS REGRAS DO SISTEMA ANFIS.



Fonte: Elaborada pelo autor.

De forma complementar, apresenta-se o gráfico de superfície na Figura 23 para as regras do sistema ANFIS, com duas entradas, indicadas por x e y e uma saída, f . Esse gráfico de superfície indica a variação da saída conforme a variação das duas entradas do sistema.

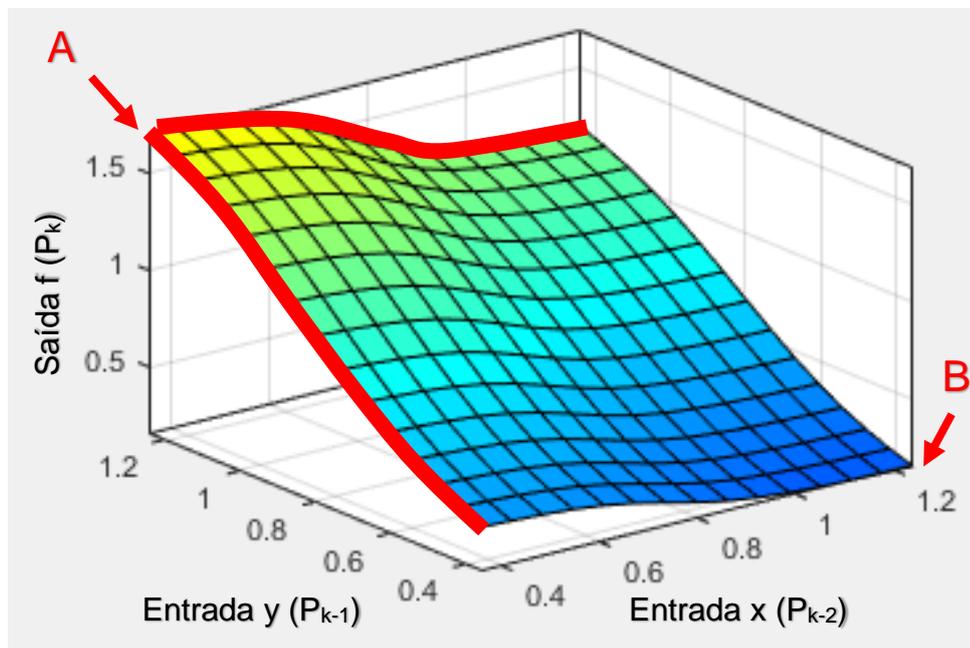
Levando-se em consideração o modelo adotado do sistema ANFIS citado na Figura 18, na qual duas amostras passadas são utilizadas como entradas para a previsão de uma amostra futura como saída, sendo a entrada x com dois atrasos de tempo e a entrada y com um atraso de tempo, algumas conclusões importantes desse gráfico de superfície serão citadas abaixo, tais como:

- A entrada y (amostra mais próxima do valor de saída) é mais relevante do que a entrada x na determinação do valor de saída f ;
- O valor máximo da saída f (1,7), representado pelo ponto A, é atingido quando está no valor mínimo a entrada x (0,355) e no valor máximo a entrada y (1,24);

- Enquanto que, o valor mínimo da saída f (0,151), indicado pelo ponto B, é atingido quando está no valor máximo a entrada x (1,24) e no valor mínimo a entrada y (0,355).

Tais informações extraídas do gráfico da Figura 23, ratifica a informação de que a entrada mais recente y é mais relevante na determinação da saída numérica f , o que não indica irrelevância da entrada mais pretérita x .

FIGURA 23. GRÁFICO DE SUPERFÍCIE PARA AS REGRAS DO SISTEMA ANFIS.



Fonte: Elaborada pelo autor.

4 RESULTADOS E DISCUSSÕES

4 RESULTADOS E DISCUSSÕES

Com a finalidade de verificar o desempenho do algoritmo de identificação e correção de outliers de curvas de demanda de energia baseado no modelo ANFIS, foram criados dois cenários para comparação entre os valores de potência separados para teste (Teste), valores com outlier (Outlier), valores estimados pelos algoritmos de INTERPOLAÇÃO, RNA e ANFIS. Nos dois cenários, foram inseridos em ordem aleatória 10% das amostras com outliers dos tipos zero e pico. Os cenários foram os seguintes:

- Cenário 1: utilização de 96 amostras, referentes a um dia de medições com inserção de outliers.
- Cenário 2: utilização de 938 amostras, referentes a dez dias de medições com inserção de outliers.

4.1 CENÁRIO 1: UM DIA DE MEDIÇÕES COM OUTLIERS

Inicialmente, como forma de atestar o funcionamento do algoritmo de identificação de outliers, um gráfico, conforme a Figura 24, foi gerado para representar seis curvas: Teste, Outliers, Interpolação, RNA, ANFIS e Desvio. A curva de teste foi feita com valores reais. Enquanto que, a curva Outlier utiliza a inserção de valores do tipo zero, pico ou atípicos, ou seja, fora do intervalo de confiança. As curvas que representam as correções dos outliers por meio de algoritmos estão expressas em Interpolação, RNA e ANFIS. Por fim, as curvas de Desvio, representam o intervalo de confiança referente a um acréscimo do desvio padrão, seja positivo ou negativo, ao valor real de teste.

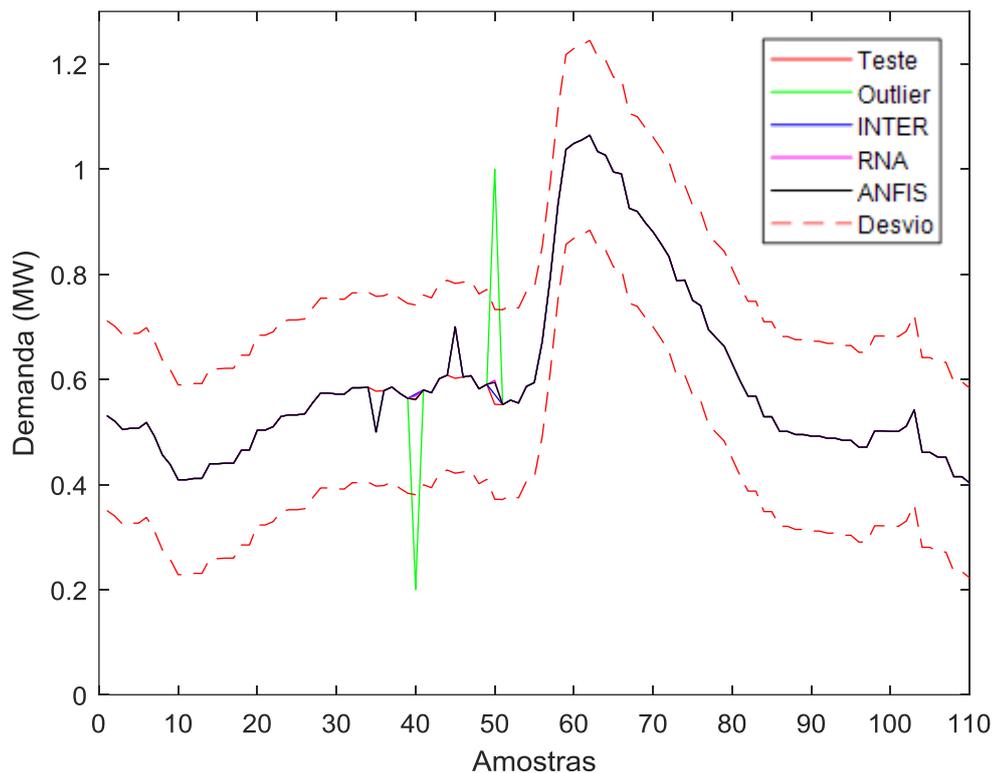
Assim, foram inseridos os valores seguintes de outliers:

- Valores destoantes do padrão, porém dentro do intervalo de confiança (desvio):
 - Amostra 35: valor 0.5 MW;
 - Amostra 45: valor 0.7 MW.
- Valor atípicos, fora do intervalo de confiança:
 - Amostra 40: valor 0.2 MW;
 - Amostra 50: valor 1.0 MW.

Nas duas primeiras situações, como as amostras (35 e 45) estão dentro do intervalo de confiança, não há correção dos seus valores pelos algoritmos propostos.

De modo contrário, para as duas situações seguintes, as amostras (40 e 50) estão fora do padrão e por isso são consideradas valores atípicos, sendo prontamente corrigidas por valores dos algoritmos de Interpolação, RNA e ANFIS.

FIGURA 24. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER E CORREÇÕES PARA VERIFICAÇÃO.



Fonte: Elaborada pelo autor.

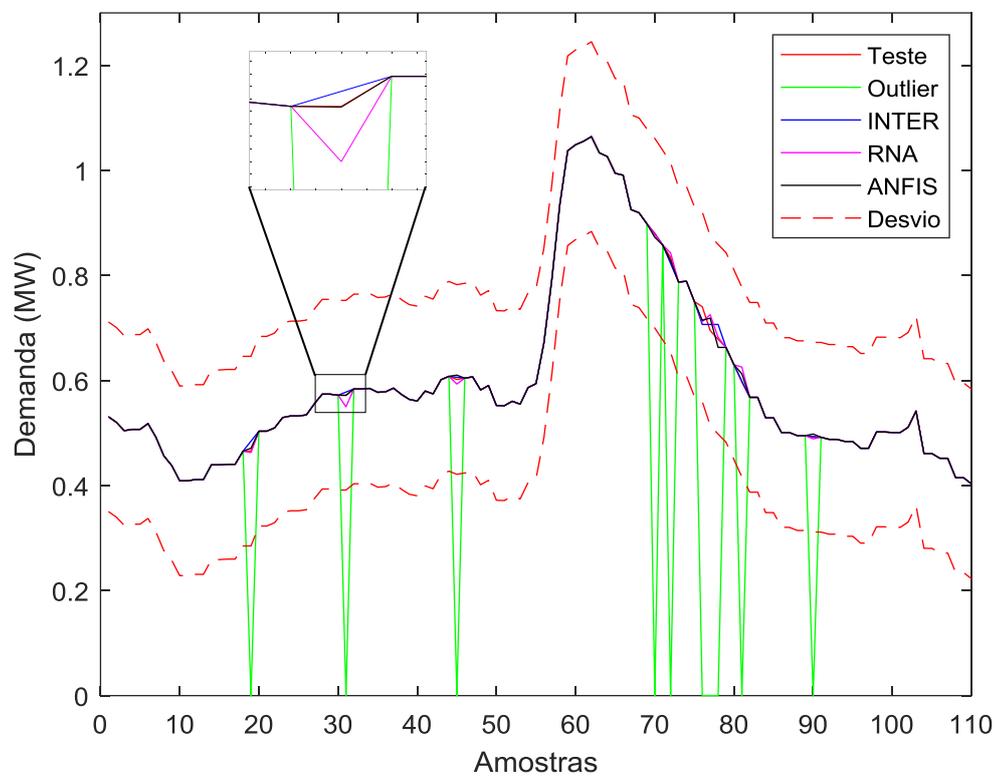
A partir da Figura 25, é possível observar a curva de amostras de potência ao longo de um dia com a inserção aleatória de outliers do tipo zero, o mesmo foi feito para a inserção de outliers do tipo pico, conforme Figura 26. Na Tabela 6, contém as avaliações dos erros relativos na estimação dos outliers para cada técnica, isto é, Interpolação, RNA e ANFIS. Deve-se observar nessa tabela que a primeira coluna corresponde aos valores dos índices das amostras que foram selecionados aleatoriamente como outliers (variando da amostra 1 até a 96).

Assim, foram obtidos como resultados para o erro relativo percentual máximo os seguintes valores:

- Interpolação: 4,5577 %;
- RNA: 4,4729 %; e
- ANFIS: 3,6957 %;

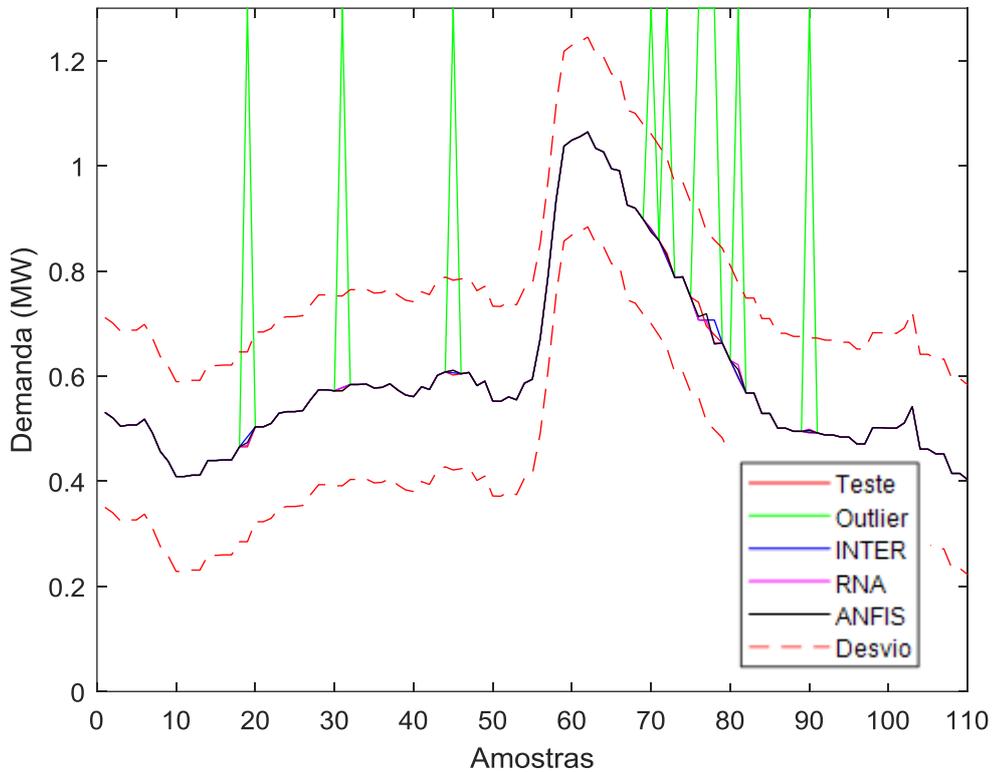
Pode-se observar o melhor desempenho do modelo ANFIS quando comparado com as outras técnicas. Embora a técnica de interpolação seja bastante rápida em questões de tempo de processamento e implementação, a mesma teve o menor desempenho para a correção de outliers nessas condições.

FIGURA 25. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER TIPO ZERO E CORREÇÕES NO CENÁRIO 1.



Fonte: Elaborada pelo autor.

FIGURA 26. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER TIPO PICO E CORREÇÕES NO CENÁRIO 1.



Fonte: Elaborada pelo autor.

TABELA 6. ERROS RELATIVOS DA CORREÇÃO DE OUTLIERS EM UM DIA.

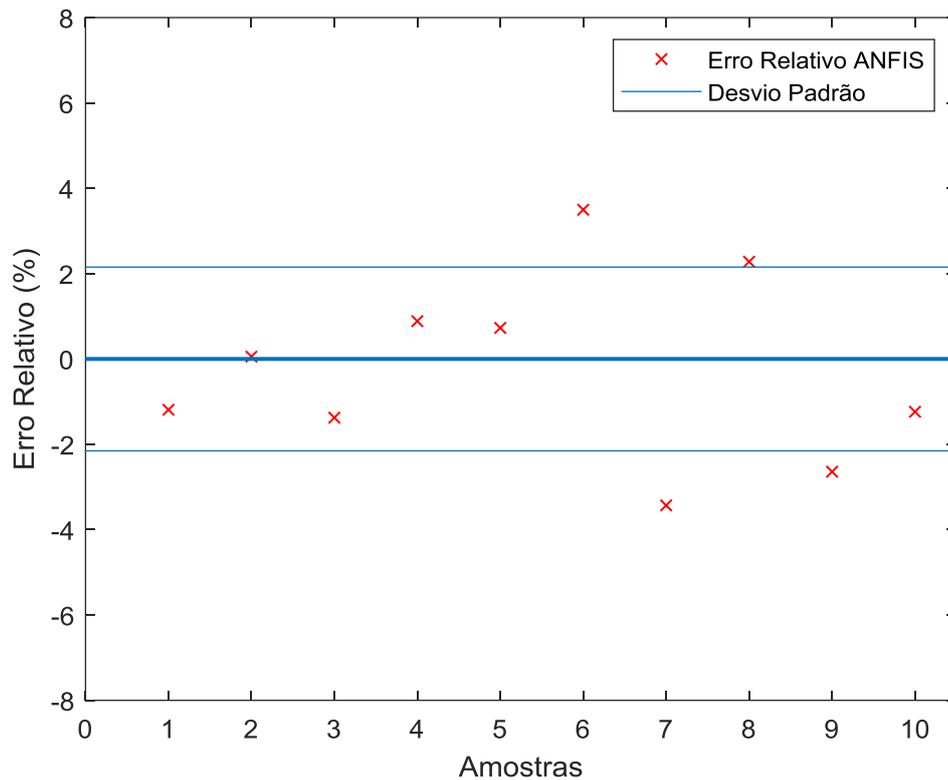
Índice	Erros Relativos Máximos Absolutos		
	Inter	RNA	ANFIS
19	4,0816	0,5870	1,1912
31	1,0490	3,8559	0,0548
45	0,7060	1,4495	1,3761
70	0,2555	0,2766	0,8858
72	1,3489	1,0794	0,7274
76	4,5577	3,7965	3,4964
77	1,6906	4,3603	3,4316
78	4,1636	0,5323	2,2788
81	0,3350	4,7959	2,6421
90	0,3049	0,6175	1,2383

Fonte: Dados da Pesquisa.

Com a finalidade de se definir um intervalo de confiança associado ao erro relativo de correção de outliers, foi determinado o desvio padrão de um conjunto de 10 outliers corrigidos (correspondentes a 1 dia).

Na Figura 27, estão plotados os valores de erros relativos da correção de outliers usado ANFIS e seu desvio padrão. O desvio padrão encontrado foi de 1,1705 % e erro máximo relativo absoluto de 3,4964 %.

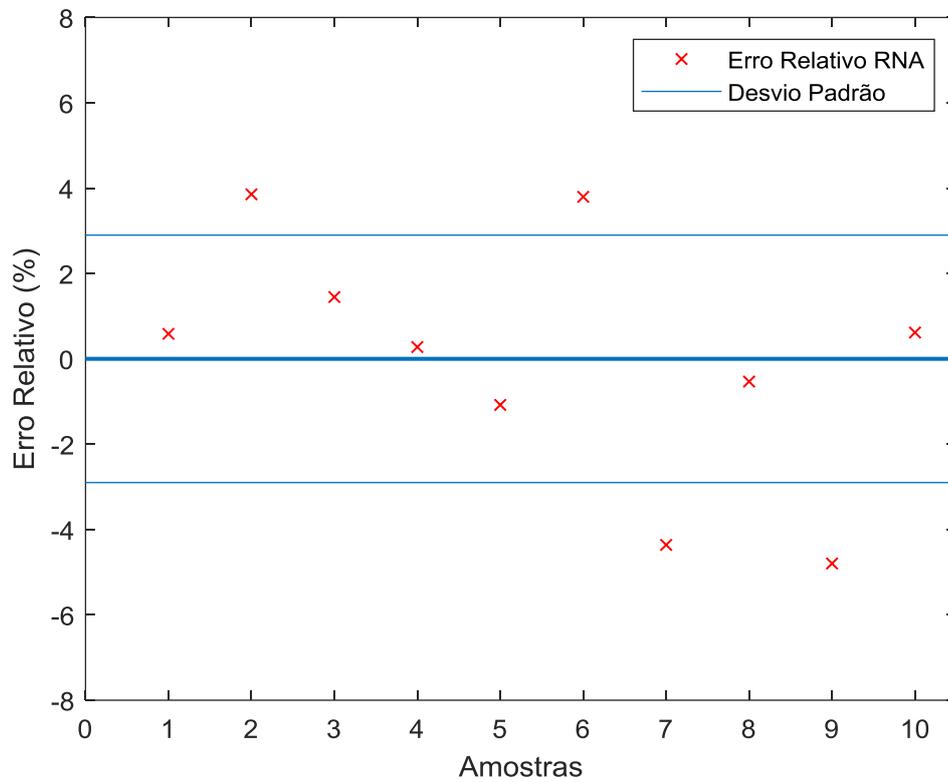
FIGURA 27. GRÁFICO DO ERRO RELATIVO ANFIS E SEU DESVIO PADRÃO NO CENÁRIO 1.



Fonte: Elaborada pelo autor.

Enquanto que, numa comparação com RNA, conforme a Figura 28, apesar de haver um intervalo de confiança com a uma quantidade de amostras maior, o desvio padrão é de 1,8276 % e erro máximo relativo absoluto de 4,4729 %. Deste modo, pode-se observar um aumento do tamanho do intervalo de confiança RNA de aproximadamente 56,1384 % quando comparado com o intervalo de confiança ANFIS.

FIGURA 28. GRÁFICO DO ERRO RELATIVO RNA E SEU DESVIO PADRÃO NO CENÁRIO 1.



Fonte: Elaborada pelo autor.

Um outro parâmetro estatístico foi utilizado para a estimação dos resultados dos algoritmos utilizado, a Média dos Erros Quadráticos. A partir da Tabela 7, observa-se que o melhor resultado foi o algoritmo ANFIS, em seguida, Interpolação e, por fim, o RNA.

TABELA 7. MÉDIA DO ERRO QUADRÁTICO PARA O CENÁRIO 1.

<i>Média do Erro Quadrático</i>	
<i>RNA</i>	<i>ANFIS</i>
0,00032072	0,00019612

Fonte: Dados da Pesquisa.

4.2 CENÁRIO 2: DEZ DIAS DE MEDIÇÕES COM OUTLIERS

De forma análoga ao Cenário 1, nessa etapa será realizada uma análise das identificações e correções de outliers com um conjunto de amostragem maior, com dados relativos a 7 dias.

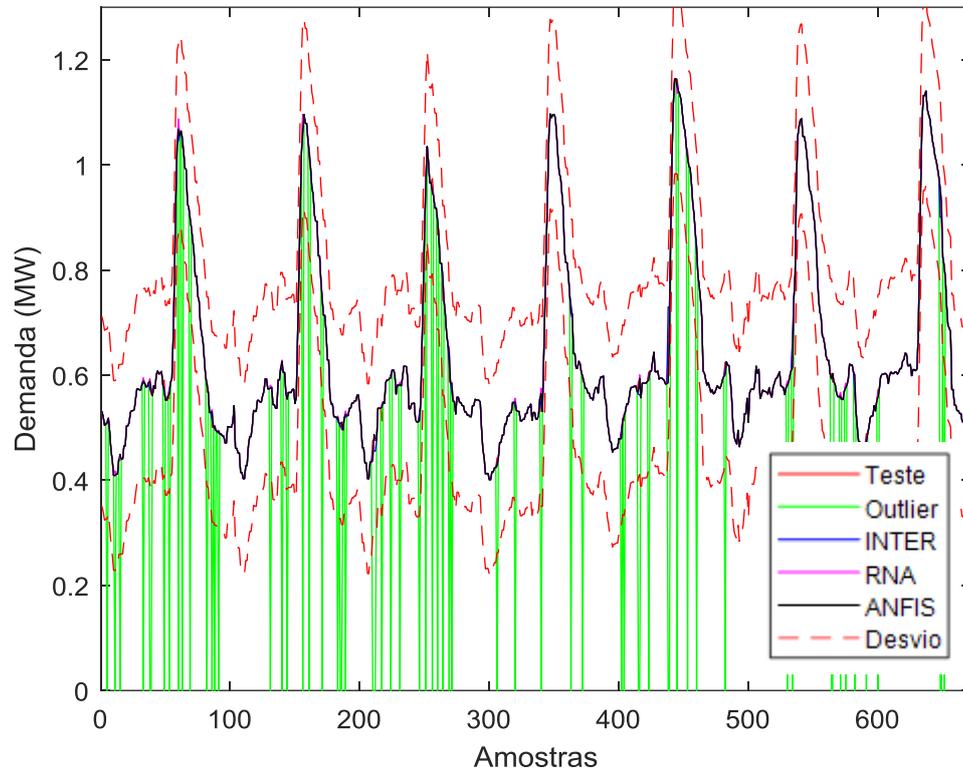
Na Figura 29, há uma curva de amostras de potência no decorrer de 7 dias com a inserção aleatória de outliers do tipo zero. A Tabela 8 possui as avaliações dos erros relativos absolutos na estimação dos outliers para cada técnica, isto é, Interpolação, RNA e ANFIS. Deve-se observar nessa tabela que a primeira coluna corresponde aos valores dos índices das amostras que foram selecionados aleatoriamente como outliers (variando da amostra 1 até a 672).

Assim, foram obtidos como resultados para o erro relativo máximo absoluto os seguintes valores:

- Interpolação: 14,2913 %;
- RNA: 11,7294 %; e
- ANFIS: 8,9947 %.

Desse modo, o algoritmo ANFIS se mostrou mais eficiente em relação aos de Interpolação e RNA, visto que apresentou menor erro relativo absoluto, 8,9947 % em um universo de 93 valores. Destaca-se que houve um aumento generalizado do erro relativo máximo absoluto com o aumento do número de amostras.

FIGURA 29. GRÁFICO DAS CURVAS DE DEMANDA COM VALORES DE TESTE, OUTLIER TIPO ZERO E CORREÇÕES NO CENÁRIO 2.



Fonte: Elaborada pelo autor.

TABELA 8. ERROS RELATIVOS DA CORREÇÃO DE OUTLIERS PARA 7 DIAS.

Índice	Erros Relativos Absolutos (%)		
	Inter	RNA	ANFIS
5	1,0848	1,7018	0,6603
11	0,3056	0,9902	0,5553
25	0,1878	2,9970	0,5488
33	0,0856	3,3497	0,9048
38	0,0436	1,3857	2,6216
39	1,6401	2,0921	0,9077
49	3,9373	4,2249	3,1069
53	3,3333	2,6795	1,7349
62	2,2076	3,7024	1,8214
63	0,7257	0,1870	2,1014
69	0,1669	0,6956	0,1099
82	2,5528	3,3569	2,7514
86	2,1956	2,4499	6,0270
87	2,1956	1,9450	0,5463
91	0,4065	1,1444	0,5375
131	0,2600	4,3010	0,4489
140	0,8130	1,8024	1,8467
144	0,2660	2,7452	2,3329
156	3,2847	3,9814	2,4807
161	1,1893	1,4636	1,4528

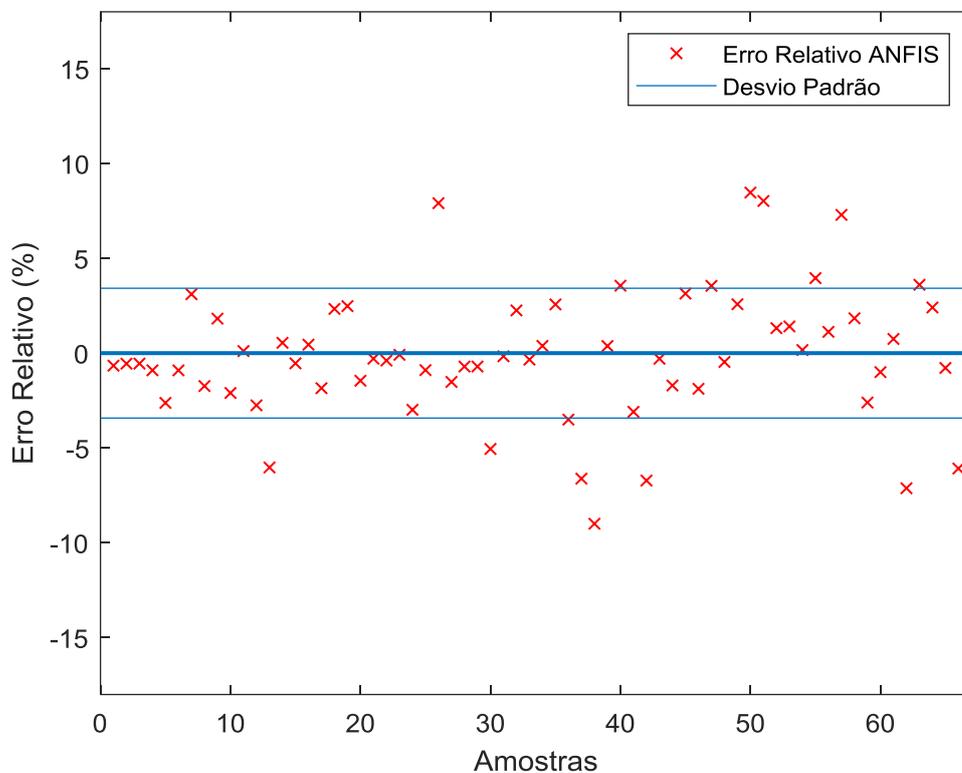
171	0,4032	2,5844	0,2914
183	1,0597	2,5296	0,3903
185	1,1811	2,1840	0,0855
186	1,2097	0,5267	2,9862
199	2,4021	3,5302	0,8961
210	4,3067	9,6942	7,9092
211	4,3067	0,6356	1,5173
212	4,3067	0,6309	0,7047
212	4,3067	0,6309	0,7047
217	5,8480	1,1733	5,0513
224	1,3468	3,0767	0,1649
234	1,7018	6,1588	2,2532
246	3,2227	2,1991	0,3390
251	2,9726	7,8708	0,3831
256	2,1561	1,3875	2,5703
260	2,9412	4,7296	3,5040
264	4,7844	5,8162	6,6162
269	5,4237	11,1411	8,9947
271	2,9915	3,3936	0,3775
306	2,1445	5,8802	3,5586
320	1,4071	1,0300	3,0942
340	2,2599	2,7796	6,7174
363	2,6536	0,1188	0,2956
372	1,5319	0,9440	1,7051
400	2,0855	4,3810	3,1408
404	2,0955	0,6549	1,8792
415	4,0460	6,7416	3,5489
416	4,8564	3,9017	0,4643
423	1,2386	6,0316	2,5794
438	14,2913	11,7294	8,4682
439	0,3783	10,9143	8,0238
445	0,9458	3,2349	1,3177
453	1,0467	1,3501	1,4096
460	1,6884	0,0720	0,1644
482	2,0817	7,1192	3,9573
530	1,1903	5,4876	1,1251
534	2,2668	10,0638	7,2922
564	1,3267	1,8131	1,8380
565	1,3629	0,0411	2,6036
571	0,4058	3,1477	0,9996
575	0,9888	5,2099	0,7476
582	1,6071	5,9930	7,1235
591	2,3006	5,2943	3,6055
600	1,7904	6,3499	2,4094
628	0,1638	2,2872	0,7803
650	1,7576	1,8875	6,0830
651	3,9628	5,2549	0,6340

Fonte: Dados da Pesquisa.

De modo semelhante ao Cenário 1, ao se definir um intervalo de confiança associado ao erro relativo de correção de outliers, foi determinado o desvio padrão para os valores de outliers corrigidos.

Na Figura 30, estão ilustrados os valores de erros relativos da correção de outliers usado ANFIS e seu desvio padrão, há uma indicação de que grande parte das amostras se encontram no intervalo de confiança, com desvio padrão de 2,3653 % e erro máximo relativo absoluto de 8,9947 %.

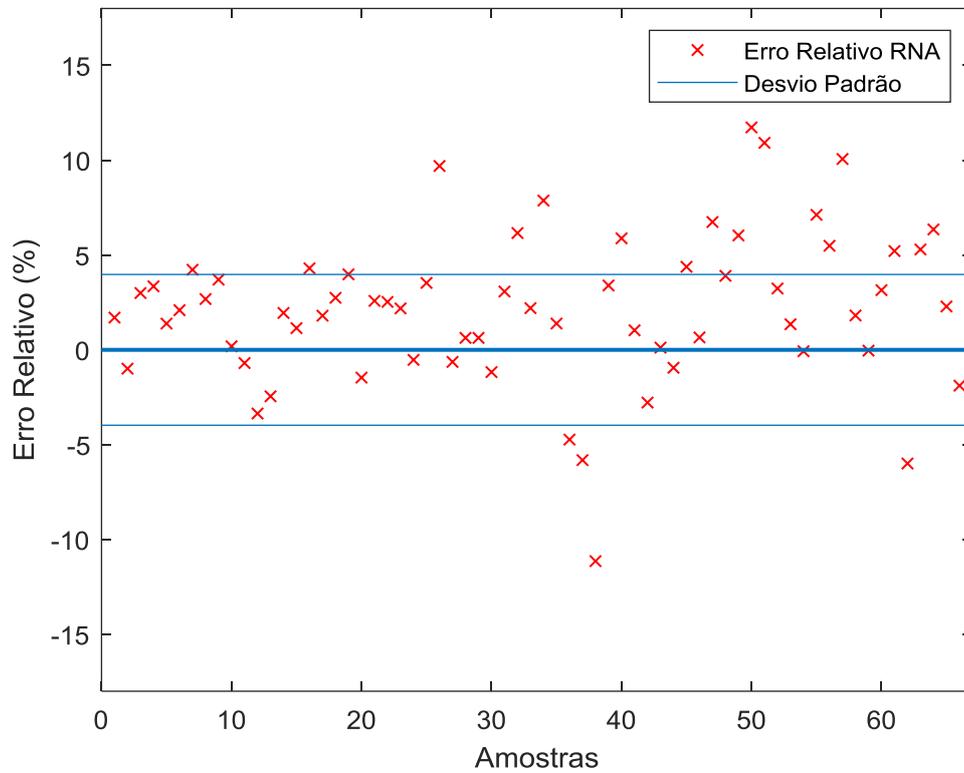
FIGURA 30. GRÁFICO DO ERRO RELATIVO ANFIS E SEU DESVIO PADRÃO NO CENÁRIO 2.



Fonte: Elaborada pelo autor.

Enquanto que, numa comparação com o algoritmo RNA, conforme a Figura 31, o desvio padrão é de 2,8350 % e erro máximo relativo absoluto de 11,7294 %. Deste modo, pode-se observar um aumento do intervalo de confiança de aproximadamente 19,8579 % quando comparado com o intervalo de confiança ANFIS.

FIGURA 31. GRÁFICO DO ERRO RELATIVO RNA E SEU DESVIO PADRÃO NO CENÁRIO 2.



Fonte: Elaborada pelo autor.

De maneira análoga, uma análise da Média dos Erros Quadráticos dos algoritmos foi realizada, conforme a Tabela 9. Desse modo, o melhor resultado foi o do algoritmo ANFIS e, em seguida, o RNA.

TABELA 9. MÉDIA DO ERRO QUADRÁTICO PARA O CENÁRIO 2.

<i>Média do Erro Quadrático</i>	
<i>RNA</i>	<i>ANFIS</i>
0,00088392	0,00046493

Fonte: Dados da Pesquisa.

5 CONCLUSÕES

5 CONCLUSÕES

Com o avanço da tecnologia, a rede elétrica tradicional está cedendo espaço para a utilização de um novo conceito, as redes elétricas inteligentes, ou smart grids, utilizando amplamente novos dispositivos chamados de medidores inteligentes, as smart meters, para o monitoramento e gerenciamento de cargas. Tais equipamentos trabalham com uma grande quantidade de informações, Big Data, e podem incorporar na sua arquitetura módulos inteligentes para a correção de dados espúrios chamados de outliers. O presente trabalho propõe a inserção de uma nova topologia de smart meters com um módulo para identificação e correção de outliers baseados em inteligência artificial.

Neste trabalho, também foi desenvolvido uma nova estratégia baseada em Big Data Analytics para o processamento de outliers aplicada a medições de demanda de energia em redes elétricas inteligentes. O algoritmo de identificação e correção de outlier é baseado na técnica híbrida de inteligência artificial ANFIS, Sistema de Inferência Neuro-Fuzzy Adaptativo, que combina as capacidades e vantagens individuais da Lógica Fuzzy, como a utilização da lógica de conjuntos fuzzy, e das Redes Neurais Artificiais – RNA, principalmente na determinação de regras sem a subjetividade de um especialista, mas sim a partir de um treinamento da rede.

A partir dos resultados apresentados, foi constatado que o método ANFIS proposto apresenta um menor erro relativo máximo percentual nos valores estimados, quando comparado com os algoritmos desenvolvidos usando RNA e Interpolação Linear. No Cenário 1, relativo a um dia de medições, o erro relativo máximo ANFIS foi de 3,4964 %, enquanto que o RNA de 4,7959 % e o Interpolação de 4,5577%. Assim como, no Cenário 2, apesar de haver um aumento generalizado nos valores, o erro relativo ANFIS foi de 8,9947 %, o RNA de 11,7294 % e o Interpolação de 14,2913 %. Ratificando a conclusão do trabalho de que o algoritmo ANFIS foi superior aos demais testados.

Foi observado também, que o desvio padrão dos erros relativos do modelo ANFIS é menor do que a de RNA, sendo este resultado promissor para a construção de intervalos de confiança das estimações feitas para a correção de outliers. Assim, no Cenário 1, em um dia, o desvio padrão ANFIS foi de 1,1705 %, enquanto que o RNA foi de 1,8276 %, indicando um aumento de 56,1384 % no intervalo de confiança.

E no Cenário 2, em dez dias, o desvio padrão do algoritmo ANFIS foi de 2,3653 %, em contrapartida, o do RNA foi de 2,8350 %, resultando em um aumento de 19,8579 % no intervalo de confiança.

Finalmente, pode-se concluir que a técnica ANFIS possui a característica positiva de se comportar bem em estimativas de correção de valores atípicos, até mesmo em situações de mudanças acentuadas ocorridas em curvas de demanda, ou seja, em horários de pico.

5.1 PUBLICAÇÕES DA PESQUISA

NETO, J. T. C., ANDRADE, P. H. M., VILLANUEVA, J. M. M., SANTOS, F. “*Big Data Analytics of Smart Grids using Artificial Intelligence for the Outliers Correction at Demand Measurements*”, in 3rd International Symposium on Instrumentation Systems, Circuits and Transducers (INSCIT 2018), Agosto de 2018.

5.2 TRABALHOS FUTUROS

Por meio desse trabalho, foi possível o desenvolvimento de algoritmos baseados em inteligência artificial capazes de corrigir valores atípicos detectados em curvas de demanda de energia elétrica. Dessa forma, a partir desses algoritmos, é possível que se utilize um sistema embarcado para a identificação e correção de modo *online* de tais valores atípicos por meio de hardwares de baixo custo, tais como Rapsberry ou Arduino.

REFERÊNCIAS

AMARAL, H. L. M. do, A. SOUZA, N. de, “Development of a low cost smart meter to collecting data and in-place tests”, 11th IEEE/IAS International Conference on Industry Applications, pp. 1-7. 2014.

ANDRADE, L. C. M., SILVA, I. N., Very Short-Term Load Forecasting Using a Hybrid Neuro-Fuzzy Approach. 11st Brazilian Symposium on Neural Networks. SBRN 2010. Brasil. IEEE Computer Society. 2010.

ANDRADE, P. H. M. de, VILLANUEVA, J. M. M., MACEDO, H. D. “Algoritmos de Correção de Outliers para Curvas de Potência Utilizando Inteligência Artificial”, SBSE Simpósio Brasileiro de Sistemas Elétricos, 2018.

ANDRADE, P. H. M. de, VILLANUEVA, J. M. M., MACEDO, H. D. “Complementary Module to Smart Meters based on Outliers Correction Using Artificial Intelligence”, I2MTC IEEE International Instrumentation and Measurement Technology Conference, 2018.

ANEEL. Cadernos Temáticos: Micro e Minigeração Distribuída. Brasília/DF. 2014.

ANEEL. RESOLUÇÃO NORMATIVA Nº 482, DE 17 DE ABRIL DE 2012. 2012.

DEVI, B. R, SUSMITHA, E. Energy Distribution and Pricing based on Load Demand Taxonomy in a Smart Grid Tree Network. International Conference on Inventive Computing and Informatics (ICICI). 2017.

Di SARNO, C., GAROFALO, A., CERULLO, G., MURRU, N., Di GIANDOMENICO, F., CHIARADONNA, S. “Power Grid Outlier Treatment through Kalman Filter”, 2014 IEEE International Symposium on Software Reliability Engineering Workshops, pp. 407-412, 2014.

DINGPING, L., KAITAO, Z., QIQI, Y., “Application of Data Stream Outlier Mining Techniques in Steam generator Safety Early Warning System of Nuclear Power Plant”,

Fifth International Conference on Measuring Technology and Mechatronics Automation, pp. 287-290, 2013.

GHANBARI, M., KINSNER, W., FERENS, K. Anomaly Detection in a Smart Grid Using Wavelet Transform, Variance Fractal Dimension and an Artificial Neural Network. IEEE Electrical Power and Energy Conference (EPEC). 2016.

HUANG, S.-J.; LIN, J.-M. "Enhancement of anomalous data mining in power system predicting-aided state estimation", IEEE Transactions on Power Systems, vol. 19, pp. 610-619. 2004.

JANG, J. -S. R. "ANFIS: adaptive-network-based fuzzy inference system", IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, pp. 665-685, 1993.

KANG, C., WNAG, Y., XUE, Y., MU, G., LIAO, R. Big Data Analytics in China's Electric Power Industry. IEEE POWER & ENERGY MAGAZINE. 2018.

KAUR, P. Outlier Detection using Kmeans and Fuzzy Min Max Neural Network in Network Data. 8th International Conference on Computational Intelligence and Communication Networks. 2016.

LANDIS GYR Equipamentos de Medição Ltda. SGP+M E13 Sistema de Gerenciamento de Perdas + Medição Manual do Usuário. 2011. Available: https://fccid.io/ANATEL/01744-11-03316/Manual-SGP_M-E13-CP/14A3AC9C-77AF-4E2D-85ED-95B45753ED02/PDF

LEE, C., SHIN, S. Fault Tolerance for Software-Defined Networking in Smart Grid. IEEE International Conference on Big Data and Smart Computing. 2018.

McCULLOCH, W. S., PITTS, W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics. vol. 5. pp. 115-133. 1943.

OKON, T., WILKOSZ, K. "Comparison of weighted least squares power system state estimation in polar and rectangular coordinate systems" in Environment and Electrical Engineering (EEEIC). pp. 140-143. 2010.

PORTAL EPE – Empresa de Pesquisa Energética. “Balanço Energético Nacional 2017 - Ano base 2016: Relatório Síntese” (2017). Available: https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2017.pdf

SANCHEZ, E. C. M. Controle por Aprendizado Acelerado e Neuro-Fuzzy de Sistemas Servo - Hidráulicos de Alta Frequência. Pontifícia Universidade Católica do Rio de Janeiro - PUC-RJ. 2009.

SATHIRACHEEWIN, S., SURAPATANA, V. Daily Typical Load Clustering of Residential Customers. Electrical Power Systems Power Engineering and Power System. 2011.

SULAIMAN, S. M., JEYANTHY, P. A., DEVARAJ, D. “Big Data Analytics of Smart Meter Data using Adaptive Neuro Fuzzy Inference System (ANFIS),” 2016 International Conference on Emerging Technological Trends (ICETT), pp. 1-5, 2016.

VALVERDE, G., TERZIJA, V. "Unscented kalman filter for power system dynamic state estimation", Generation, Transmission and Distribution, IET, vol. 5, no. 1, pp. 2937. 2011.

WANG, J., LI, C., ZHANG, S. “Big Data Impact Analysis of Smart Grid based on AHP method”, International Conference on Manufacturing Science and Engineering (ICMSE 2015), 2015.

YE, X., LU, Z., QIAO, Y., MIN, Y., O'MALLEY M. “Identification and Correction of Outliers in Wind Farm Time Series Power Data”, IEEE Transactions on Power Systems, vol. 31, pp. 4197-4205. 2016.

YE, X., ZONGXIANG, L., QIAO, Y., MIN, Y., O'MALLEY, M. "Identification and Correction of Outliers in Wind Farm Time Series Power Data". IEEE Transactions on Power Systems. 31. pp. 1-9. 2016.

ZADEH, L. A. Fuzzy Sets, Information and Control. Vol. 8. pp. 338-353. 1965.

ZADEH, L. A. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Transactions on Systems, Man and Cybernetics. Vol. SMC-3. pp. 28-44. 1973.

ZHAN, J., HUANG, J., NIU, L., PENG, X., DENG, D., CHENG, S. "Study of the key technologies of electric power big data and its application prospects in smart grid", IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1-4. 2014.

ZHANG, J., WELCH, G., BISHOP, G., HUANG, Z. "A two stage kalman filtering approach for robust and real time power systems state tracking". IEEE Transactions on Sustainable Energy, vol. 5, no. 2, pp. 629-636. 2011.