
Processamento de *outliers* nos dados do sistema elétrico

Pedro Henrique Meira de Andrade
Juan Mauricio Moises Villanueva
Helon David de Macêdo Braz

**PROCESSAMENTO DE *OUTLIERS* NOS
DADOS DO SISTEMA ELÉTRICO**



UNIVERSIDADE FEDERAL DA PARAÍBA

Reitor

VALDINEY VELOSO GOUVEIA

Vice-reitora

LIANA FILGUEIRA ALBUQUERQUE

Pró-reitor PRPG

GUILHERME ATAÍDE DIAS



EDITORA UFPB

Diretor

REINALDO FARIAS PAIVA DE LUCENA

Chefe de produção

JOSÉ AUGUSTO DOS SANTOS FILHO

Conselho editorial

Adailson Pereira de Souza (Ciências Agrárias)
Eliana Vasconcelos da Silva Esrael (Linguística, Letras e Artes)
Fabiana Sena da Silva (Interdisciplinar)
Gisele Rocha Côrtes (Ciências Sociais Aplicadas)
Ilda Antonieta Salata Toscano (Ciências Exatas e da Terra)
Luana Rodrigues de Almeida (Ciências da Saúde)
Maria de Lourdes Barreto Gomes (Engenharias)
Maria Patrícia Lopes Goldfarb (Ciências Humanas)
Maria Regina Vasconcelos Barbosa (Ciências Biológicas)

Conselho científico

Maria Aurora Cuevas-Cerveró (Universidad Complutense Madrid/ES)
José Miguel de Abreu (UC/PT)
Joan Manuel Rodriguez Diaz (Universidade Técnica de Manabí/EC)
José Manuel Peixoto Caldas (USP/SP)
Letícia Palazzi Perez (Unesp/Marília/SP)
Anete Roese (PUC Minas/MG)
Rosângela Rodrigues Borges (UNIFAL/MG)
Silvana Aparecida Borsetti Gregorio Vidotti (Unesp/Marília/SP)
Leilah Santiago Bufrem (UFPR/PR)
Marta Maria Leone Lima (UNEB/BA)
Lia Machado Fiuza Fialho (UECE/CE)
Valdonilson Barbosa dos Santos (UFCEG/PB)

Editora filiada à:



Pedro Henrique Meira de Andrade
Juan Mauricio Moises Villanueva
Helon David de Macêdo Braz

**PROCESSAMENTO DE *OUTLIERS* NOS
DADOS DO SISTEMA ELÉTRICO**

EDITORA UFPB
JOÃO PESSOA
2020

Direitos autorais 2020 – Editora UFPB
Efetuado o Depósito Legal na Biblioteca Nacional, conforme a
Lei nº 10.994, de 14 de dezembro de 2004.

Todos os direitos reservados à Editora UFPB.

É proibida a reprodução total ou parcial, de qualquer forma ou por qualquer meio. A violação dos direitos autorais (Lei nº 9.610/1998) é crime estabelecido no artigo 184 do Código Penal.
O conteúdo desta publicação é de inteira responsabilidade do autor.

Projeto Gráfico
Editora UFPB

Catálogo na publicação
Seção de Catalogação e Classificação

A553p Andrade, Pedro Henrique Meira de.
Processamento de outliers nos dados do sistema elétrico /
Pedro Henrique Meira de Andrade, Juan Mauricio Moises Villanueva,
Helon David de Macêdo Braz. – João Pessoa: Editora UFPB, 2020.

111 p. : il.
Recurso digital
Formato: PDF
Requisito do sistema: Adobe Acrobat Reader
ISBN 978-65-5492-008-7

1. Engenharia elétrica. 2. Sistema elétrico de potência.
3. Outliers – Detecção e correção. I. Villanueva, Juan Mauricio Moises.
II. Braz, Helon David de Macêdo. III. Título.

UFPB/BC

CDU 621.3

Livro aprovado para publicação através do Edital Nº 01/2020/
Editora Universitária/UFPB - Programa de Publicação de E-books.

EDITORA UFPB
Cidade Universitária, Campus I,
Prédio da editora Universitária, s/n
João Pessoa – PB. CEP 58.051-970

<http://www.editora.ufpb.br>
E-mail: editora@ufpb.br
Fone: (83) 3216-7147

AGRADECIMENTOS

Gostaria de agradecer a Deus por seu infinito Amor por mim e por todos presentes dados durante esta jornada.

À minha mãe, meu pai e meu irmão, o meu mais profundo agradecimento e amor. Só foi possível graças a vocês. Essa conquista é nossa. A vocês, todo o meu amor.

À minha namorada Aléxia que foi uma fiel e insistente incentivadora e que acreditou que tudo daria certo quando eu mesmo não acreditei.

Ao professor Juan, por sua paciência e tranquilidade e por seus conselhos, dicas e sugestões que foram engrandecedores.

Ao professor Helon que agregou bastante ao trabalho devido sua vasta experiência na área.

Aos meus amigos de universidade: André, Larissa e Jéssica que compartilharam essa caminhada ao meu lado.

Aos Clorofilhos de Deus, meus eternos amigos que sempre me levam para próximo de Deus.

Aos meus amigos do prédio e da Igreja que são fonte ilimitada de apoio, força e compreensão.

Por fim, agradeço a todos que direta ou indiretamente fizeram parte deste livro.

*“A felicidade pode ser encontrada
mesmo nos momentos mais sombrios,
se você lembrar de acender a Luz.”*

(Alvo Dumbledore)



SUMÁRIO

PREFÁCIO	8
1 SISTEMAS ELÉTRICOS DE POTÊNCIA	12
2 PROCESSAMENTO DE <i>OUTLIERS</i> NOS DADOS DO SISTEMA ELÉTRICO....	18
3 PROCESSAMENTO DE <i>OUTLIERS</i> EM DADOS DO SISTEMA ELÉTRICO.....	48
4 ESTUDO DE CASO: SUBESTAÇÃO REAL.....	76
POSFÁCIO	96
PUBLICAÇÕES DOS AUTORES NA ÁREA.....	98
REFERÊNCIAS	101
SOBRE OS AUTORES	110

PREFÁCIO

UM BREVE RECADO

Quando criança um dos meus sonhos era ser cientista, antes de ser astronauta e depois de ser policial. A vida passou e optei por Engenharia Elétrica, porém descobri-me cientista quando fui obrigado a utilizar o método científico para testar hipóteses e teorias sobre um problema do cotidiano. Realizei, portanto, um sonho de criança enquanto fazia a pesquisa do mestrado. Mais tarde, me tornei Perito Criminal de Engenharia Elétrica, tornando mais um sonho real ao entrar para área da Segurança Pública que lida com a ciência e a técnica. E no momento, estou aguardando algum convite da NASA para realizar o último dos meus sonhos infantis.

Esse livro, portanto, é mais um fruto de um sonho e de uma pesquisa bem-sucedida que teve ainda 2 artigos em periódicos, um deles sendo o principal da área de sistemas elétricos de potência, além de 6 artigos em congressos nacionais e internacionais.

A pesquisa durante o mestrado me levou a entrar no mundo da vanguarda da engenharia e da tecnologia, entrei no mundo onde as descobertas acabaram de acontecer.

Com o desenvolvimento cada vez maior de dispositivos capazes de conexão à internet, o conceito *de Internet of Things* (IoT) ganha força na comunidade científica. E, portanto, surgem incontáveis aplicações práticas para problemas cotidianos. Outra

área que ganhou bastante relevância foi a Inteligência Artificial (IA) que também passa a fazer parte do dia-a-dia das pessoas de forma muitas imperceptível.

Os dispositivos conectados (IoT) e dotados de inteligência artificial (IA) necessitam e geram um grande volume de dados para treinamento dos modelos e, depois, como resultados das aplicações. Então, entramos em um novo conceito que permeará a tecnologia daqui para frente: *Big Data*. Estamos imersos em uma quantidade colossal de dados gerados que não temos como tratá-los em sua integralidade. A geração dos dados está nas redes sociais, nos veículos, nas compras pela internet, nas transações bancárias, e, claro, nos sistemas de potência. Entretanto, a mera geração de dados brutos não representa significativo avanço. Os dados brutos, se não utilizados, podem ser até inconvenientes, diante de problemas de armazenamento ou da própria coleta. A grande questão é o tratamento dos dados para que deles possam ser retiradas as informações. A informação, sim, é valiosa e traz o verdadeiro conhecimento acerca do problema. Os dados são o petróleo do século XXI e a informação é o combustível.

Esse livro, portanto, está alocado em três grandes áreas da comunidade científica e voltadas para uma aplicação essencial para a sociedade: energia elétrica. Tentaremos ressaltar as tendências que guiam a implementação das Redes Elétricas Inteligentes, mas o foco principal será no tratamento dos dados das subestações elétricas. Serão expostos os algoritmos

desenvolvidos na pesquisa para o processamento de valores atípicos a partir de dados de subestações reais.

Caro(a) leitor(a), na minha humilde opinião, o sistema elétrico de potência é uma das maiores realizações da humanidade. Uma hidrelétrica a milhares de quilômetros gera energia para escrita desde livro no computador e para sua leitura em outro equipamento digital que cedo ou tarde precisará ser ligado a rede para recarga. A agregação de inteligência artificial a esse sistema possui repercussões que só podemos imaginar. E, sem querer ser muito pretensioso, quero fazê-lo(a) vislumbrar esse enorme *potencial*.

DESTAQUES

- Contextualização do Sistema Elétrico de Potência e Infraestrutura Avançada de Medição.
- Algoritmo de previsão do valor futuro que será medido para possível correção de *outlier*.
- Técnica de Correção Fuzzy: construção do algoritmo a partir dos incrementos diferenciais, definição dos conjuntos *fuzzy* e construção das regras do sistema de inferência a partir dos conhecimentos dos perfis de carga.
- Técnica de Correção RNA: aplicação de redes neurais artificiais para estimação de valor futuro baseada nas amostras anteriores com janela reduzida.
- Diagnóstico do sistema de medição: determinação do estado de operação dos medidores da subestação,

baseados no cálculo do fator de ocorrência que é representado pelo número de amostras errôneas no tempo.

- Injeção aleatória de *outliers* para teste: algoritmo para gerar um número aleatório em índices aleatórios de *outliers* nas amostras originais para poder corrigi-las e avaliar as técnicas desenvolvidas.

ORGANIZAÇÃO DO LIVRO

Neste trabalho, no Capítulo 2, serão discutidas as principais características dos medidores inteligentes, métodos para detecção e correção de *outliers* abordados na literatura, conceitos básicos sobre Redes Neurais, Lógica *Fuzzy* e ARIMA; por fim, são apresentadas as curvas de demanda da subestação e a proposta do fator de ocorrência para diagnosticar o sistema de medição quanto à quantidade de *outliers*. No Capítulo 3 serão explicados os algoritmos desenvolvidos a partir de técnicas de inteligência artificial. No Capítulo 4 serão apresentados os principais resultados e as discussões sobre as técnicas desenvolvidas. E, por fim, no Capítulo 5 serão apresentadas as considerações finais.

1. SISTEMAS ELÉTRICOS DE POTÊNCIA

O sistema elétrico está evoluindo para redes elétricas inteligentes, conhecidas na literatura como *Smart Grid*. Isto é, uma rede com uma infraestrutura baseada em tecnologias digitais e com o intuito de melhorar sua eficiência operacional da transmissão e distribuição da energia elétrica (SARNO et al., 2014).

As *Smart Grids* são caracterizadas por integrarem tecnologias de informação e comunicação para interconectar o sistema elétrico de um país ou região (GHOSH; GHOSE; MOHANTA, 2013). De acordo com Katic (2016), o conceito de inteligência é agregado às diferentes áreas do sistema elétrico, pois inclui:

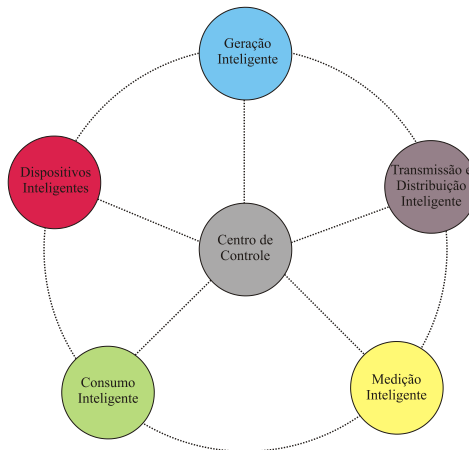
- (i) Geração inteligente: monitoramento e otimização do controle de energia gerada.
- (ii) Medição inteligente: automação, gerenciamento e tratamento dos dados coletados.
- (iii) Transmissão e distribuição inteligente: flexibilidade para manobras em possíveis quedas.
- (iv) Consumo inteligente: monitoramento e gerenciamento dos gastos a partir do conhecimento da carga;

- (v) Dispositivos inteligentes: carros elétricos, carregadores e dispositivos de armazenamento de energia.

As diferentes áreas são ilustradas na Figura 1 e no centro está a Central de Controle e inclui, dentre outros campos:

- Sistema Supervisório;
- Sistema de Gerenciamento de Energia;
- Sistema de Gerenciamento de Faltas;
- Análise de Operações e;
- Estimacões e Planejamento;

FIGURA 1. ÁREAS DA REDE ELÉTRICA INTELIGENTE.



Fonte: Elaborada pelo Autor.

As redes inteligentes contam com sensores, atuadores, processamento em tempo real, novas interfaces e novos protocolos de comunicação que permitem ao sistema introduzir características de auto-recuperação com capacidade de detectar, analisar e corrigir falhas na rede de forma automática. Essas características fortalecem a infraestrutura do sistema elétrico de potência dando-lhe mais flexibilidade, segurança, estabilidade e confiabilidade (FANG et al., 2012).

A concepção inicial da Rede Elétrica Inteligente começou a partir da ideia de Infraestrutura Avançada de Medição (IAM) em que o principal objetivo era conhecer integralmente a rede e melhorar o lado da demanda, isto é, os consumidores domésticos e comerciais, que possuem relevante papel no setor elétrico. Contudo, as novas necessidades e as novas demandas fizeram com que a ideia da IAM fosse expandida para todo o escopo da rede elétrica. Entretanto, para se ter uma infraestrutura avançada de medições é necessário ter Medidores Inteligentes, também conhecidos na literatura como *Smart Meters* (XENIAS et al., 2015).

Os medidores inteligentes são medidores eletrônicos com capacidade de comunicação bidirecional e com funcionalidades especiais nas medições. Nas redes elétricas inteligentes, as medições não devem ser feitas de forma isolada, mas conectadas aos sistemas de coleta de dados com plataformas de análises e processamento e com funções de monitoramento e apoio na tomada de decisões (SAMAD; ANNASWAMY, 2017).

Os medidores das subestações permitem a captura contínua dos dados em intervalos predeterminados e com as medições são construídos os bancos de dados de interesse da concessionária de energia. Quando não se tinha tais banco de dados, ficava a cargo de um operador da distribuidora as medições em campo e eram feitas de forma manual, contudo a amostragem era bem menor e estava sujeita às imprecisões tanto do sensor como imperícia ou negligência do funcionário (MULLER; FRANCO, 2014). Os medidores inteligentes fornecem uma maior quantidade de medições permitindo aumentar o nível de conhecimento sob o perfil de carga dos consumidores e promovendo um maior conhecimento da operação e planejamento das subestações (BAUMEISTER, 2010).

As distribuidoras de energia têm bastante interesse sobre os dados coletados nas subestações como as curvas de demanda e potência de sua região de distribuição. Com os dados adquiridos é feito um histórico e através dele são desenvolvidas análises internas sobre a subestação e sobre sua área de cobertura. Com esses estudos é possível desenvolver ou aperfeiçoar os métodos de operação, ter maior controle do sistema elétrico além de facilitar a manutenção corretiva e preventiva. Estas medidas por parte das concessionárias são base para um planejamento estratégico e uso mais racional na alocação de recursos (MEDEIROS, 2016).

Diante desta conjuntura, a exatidão e confiabilidade nas medições da curva de carga são fundamentais para as atividades de operação e planejamento de sistemas elétricos de potência. Contudo, os dados de demanda e potência podem ser afetados por fatores como erros de comunicação, manobras de chaves, queda de energia ou problemas de instabilidade na transdução da medida pelo sensor. Quando esses eventos ocorrem podem ser gerados valores discrepantes do comportamento padrão da curva. Esses valores atípicos e indesejáveis são conhecidos na literatura como *outliers* (MELO; CASTRO, 2014).

As medições são realizadas são realizadas sequencialmente, com um intervalo fixo e são armazenadas em um banco de dados criando uma série de amostras ao longo do tempo. A qualidade dos dados coletados deve ser um aspecto buscado para o histórico das medições, pois para bons estudos de previsão de séries temporais é importante uma integridade mínima dos dados ou os resultados podem ser equivocados ou inconclusivos (YANG et al., 2011).

Em busca de estratégias que permitam realizar o processamento de *outliers*, tem-se a motivação para pesquisa de novas metodologias para a estimação do valor faltante. Assim também, intensifica-se a análise dos efeitos dos *outliers* sob as curvas de demanda elétrica, procurando-se a maior exatidão possível nas informações disponibilizadas, atendendo assim as características de medidores inteligentes. Para a correção dos

outliers foram utilizadas técnicas de inteligência artificial que dispensam um complexo desenvolvimento matemático e estatístico.

Portanto, o principal objetivo deste livro consiste em desenvolver uma metodologia utilizando em técnicas de inteligência artificial que detecte e corrija os *outliers* das curvas de demanda das subestação

2. PROCESSAMENTO DE *OUTLIERS* NOS DADOS DO SISTEMA ELÉTRICO

Este capítulo contemplará as principais características dos medidores inteligentes presentes no setor elétrico de potência e as tendências para os que estão sendo desenvolvidos, presentes em uma rede de dados cada vez mais volumosa. Também serão apresentados métodos de detecção e correção de *outliers* em séries temporais presentes na literatura.

2.1 MEDIDORES INTELIGENTES

A medição do consumo de energia elétrica no Brasil ainda é feita com medidores eletromecânicos indutivos que operam contando o número de rotações que um disco de metal perfaz a uma velocidade proporcional a energia demandada. Nos anos de 1990, os medidores eletrônicos foram introduzidos com a capacidade de transmitir as medições remotamente, mas de maneira unidirecional. Então para melhorar a comunicação entre a companhia de energia e os consumidores, facilitando o tempo de cobrança, foi desenvolvida a primeira geração de medidores inteligentes em 2005; ainda pouco utilizados no Brasil, mas com um crescimento significativo no exterior. Os medidores inteligentes são a nova geração de dispositivos utilizados pelas

concessionárias de energia para fazer medições tanto no lado da demanda como do próprio sistema (GANDHI; BANSAL; 2013).

No sistema elétrico do futuro, os medidores terão papel significativo no monitoramento da demanda em tempo real. Os sistemas de sensores inteligentes fornecerão dados de forma automática e poderão ser capazes de tomar decisões. Segundo Morello et al. (2017) os medidores inteligentes deverão ter as seguintes características:

- Programação e controles remotos;
- Interoperação entre os diversos medidores da rede.
- Análise da qualidade de energia;
- Processamento de dados e tomada de decisões.

Essas melhorias permitirão que o próprio sistema reconfigure de forma automática a rede para otimizar a eficiência energética e reduzir a duração de interrupções, aumentando assim a confiabilidade e estabilidade do sistema (CACIOTTA et al., 2012).

Segundo Xenias et al. (2015), o avanço na infraestrutura de medição nas redes elétricas tem como grande consequência positiva um maior gerenciamento do lado da demanda. Isto é, os consumidores domésticos e de pequeno porte que possuem um relevante impacto tornam-se atores ativos no setor energético. O consumidor passa a medir e analisar os dados, podendo moldar seu comportamento através deles.

Com o desenvolvimento de redes de sensores, comunicação sem fio, sistemas em nuvem; a concepção tecnológica da Internet das Coisas tem sido agregada aos medidores inteligentes (SHOPOV, 2017). O termo Internet das Coisas (*Internet of Things - IoT*), refere-se a uma rede em que a capacidade de comunicação não será restrita apenas aos dispositivos móveis, mas expandida para todas as coisas que coexistem como veículos, prédios, câmeras, eletrodomésticos e tudo que possa ter uma tecnologia embarcada com sensores e conexão capazes de gerar, coletar e transmitir dados (LEE; BAE; KIM; 2017).

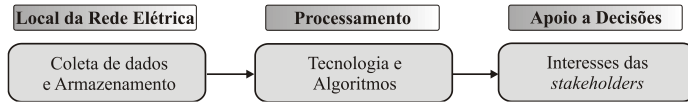
Os medidores inteligentes são bons exemplos de aplicações da Internet das Coisas, pois integram a capacidade de comunicação com as necessidades do setor elétrico para monitoração e controle automático de todas as áreas, desde a geração da energia até o consumo (SHOPOV, 2017). Ou seja, a inteligência nas medições faz parte de um ambiente muito complexo que a própria rede elétrica (inteligente), pois envolve fatores sociais, políticos e econômicos.

Os componentes-chaves ilustrados na Figura 2 fornecem o contexto que envolve a tecnologia de medições inteligentes. Na subestação (Local da Rede Elétrica) é feita a aquisição e o armazenamento dos dados brutos que são processados através de algoritmos que extraem informações importantes acerca da rede

elétrica. As informações são analisadas e servirão de apoio a decisões para as *stakeholders* que são os agentes ativos da sociedade em relação ao Sistema Elétrico. Os agentes da sociedade mais interessados no avanço da infraestrutura da medição são: as concessionárias de energia, o consumidor doméstico, o consumidor industrial e o governo. Então, inicialmente os dados são coletados pelos medidores inteligentes tanto junto à demanda internamente ao sistema elétrico. No lado dos consumidores é interessante saber seu consumo ao longo das diferentes horas do dia e ao longo das estações, pois mesmo tendo perfis mais previsíveis é possível moldar o comportamento para evitar gastos desnecessários (ALAHAKOON; YU, 2016).

Para as concessionárias, o conhecimento de sua região de distribuição é fator primordial para apoiar e embasar decisões, visando uma melhoria nos serviços e um aumento no lucro a partir de uma correta alocação de recursos. E o governo é capaz de desenvolver diretrizes que possam servir de base para uma política pública de gestão de energia elétrica a partir de um amplo conhecimento do sistema.

FIGURA 2. COMPONENTES-CHAVE DA INTELIGÊNCIA NA MEDIÇÃO DE ENERGIA.



Fonte: Adaptado de Alahakoon e Yu (2016).

Diante da importância relatada dos medidores inteligentes para as Redes Elétricas Inteligentes é necessário que os dados coletados sejam íntegros e corretos. As medições erradas (*outliers*) podem ocorrer devido à:

- Erros de comunicação;
- Manobras de chaves;
- Queda de energia;
- Problemas de instabilidade da medida pelo sensor.

E estas medições erradas podem resultar em grandes perdas econômicas tanto para as concessionárias como para os consumidores.

Para aumentar a confiabilidade das medições, Varodayan e Gao (2010) propõem usar um outro medidor para gerar uma redundância. Contudo não é um método muito eficiente, pois aumenta o custo ao se dobrar o número de sensores. Além disso, se um dos medidores apresentar um valor diferente será possível determinar que houve um erro, mas haverá ambiguidade. Esta ambiguidade poderia ser resolvida com o acréscimo de mais um medidor, contudo aumentaria ainda mais o custo.

Mesbah (2017) desenvolveu uma nova arquitetura com modificações nas disposições dos medidores ao longo da linha que proporciona ainda assim redundância das medições, entretanto apresenta desafios para sua implementação prática. Portanto, uma das principais características para o preenchimento das medições incorretas é a praticabilidade.

2.2 ALGORITMOS DE PROCESSAMENTO DE *OUTLIERS*

Diversas metodologias para o processamento de *outliers* têm sido desenvolvidas para diversos domínios de aplicações como engenharia, economia, meteorologia, dentre outros. E para cada domínio de aplicação tende a ser desenvolvido um método único de acordo com o tipo de série temporal e o tipo de *outlier* (YE et al., 2015).

Uma série temporal é caracterizada por uma coleção de observações feitas sequencialmente ao longo do tempo (BOX; JENKINS; REINSEL, 1994). São dados observados a cada hora (por exemplo: rendimento de um processo químico ou valores de potência de uma subestação), diariamente (por exemplo: preços das ações ou relatórios meteorológicos), mensalmente (por exemplo: taxa de desemprego ou inflação mensal) ou até anualmente (por exemplo: PIB de um país ou taxa de IDH).

Segundo Moretin e Toloi (2004), as séries temporais são definidas como o conjunto de observações $\{Y(t), t \in T\}$ com Y

sendo o valor de interesse e T o conjunto de índices relacionados ao tempo. São divididas em três tipos principais:

1. **Discreta:** o conjunto de índices é dito finito, isto é, $T = \{t_1, t_2, \dots, t_n\}$. Exemplo são as importações mensais de uma empresa.
2. **Contínua:** o conjunto de índices é assumido como infinito, $T = \{t: t_1 < t < t_2\}$. Exemplo é o registro da maré em determinada localidade durante 1 ano.
3. **Multivariada:** Cada observação na série temporal é um vetor de números. As observações são $Y_1(t), \dots, Y_k(t)$; $t \in T$. Exemplo é o vetor que possui o valor da receita mensal $Y_1(t)$ e o valor dos gastos $Y_2(t)$.

As séries temporais, geralmente, caracterizam sistemas complexos, e muitas vezes estão sujeitas a ocorrências inesperadas. Esses eventos incontroláveis geram os *outliers*, isto é, valores discrepantes das demais medições da série. E em sistemas de potência, esses valores devem ser detectados e corrigidos para que o histórico dos dados possua qualidade e integridade mínima para que estudos e análise sejam realizados com segurança (NASCIMENTO et al., 2014).

Ye et al. (2016) propuseram uma nova abordagem para detectar e corrigir *outliers* nos dados de potência de um parque eólico. O método proposto é baseado na função de probabilidade da curva e nas características dos *outliers*. Além disso, o método se

baseia na medição da velocidade do vento e da correlação entre os parques eólicos adjacentes para corrigir os *outliers* identificados.

Também focado em parques eólicos, Yesilbudak (2016) propõe a abordagem de clusterização particional para detecção de *outliers*. Este método consiste em separar por grupos toda a faixa das medições aplicando uma técnica de clusterização denominada *k*-média, calcular o centroide de cada grupo e verificar a distância das medições ao centroide respectivo de sua faixa.

Diante da abrangência das Redes Elétricas Inteligentes, especificamente sobre o conceito de Casa Inteligente (*Smart Home*), Jakkula e Cook (2010) propuseram um método estatístico para detectar *outliers* em um ambiente doméstico inteligente com o objetivo de conhecer o comportamento do sistema, conservar energia e torná-lo mais eficiente.

Neagu, Grigoras e Scarlatache (2017) propuseram um método para detecção de *outliers* em curvas de demanda de energia. A abordagem é realizada em dois estágios: no primeiro, é utilizada uma técnica de mineração de dados para extração das curvas de demanda indicadores calculados pelos medidores inteligentes. No segundo estágio, os *outliers* são detectados a partir dos indicadores utilizando processamento estatístico.

A mineração de dados é definida por Nia (2012) como “um processo que extrai a informação e o conhecimento implícito, desconhecidos antecipadamente, mas potencialmente úteis, a

partir de dados volumosos, incompletos, distorcidos, ruidosos e aleatórios gerados na aplicação prática."

Os principais indicadores extraídos das curvas foram valor máximo e mínimo da carga, fator de carga, fator de perdas e fator de demanda. E a partir do processamento estatístico sobre eles é que foram detectados os *outliers*. A distribuição gaussiana foi gerada a partir um grande volume de medições e é representada pela média das amostras (μ) e pelo desvio padrão (σ):

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

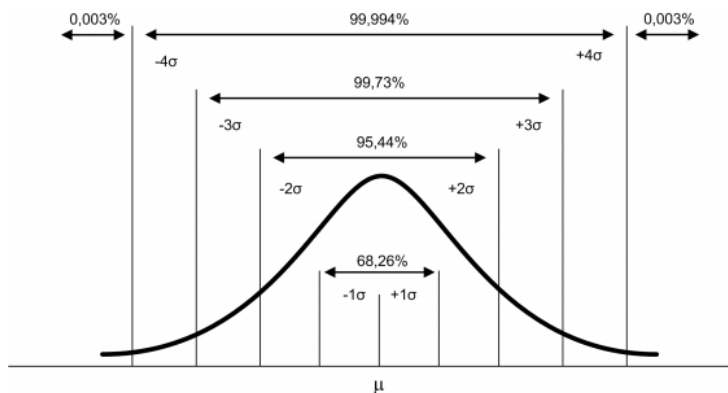
Na distribuição gaussiana, a amostra medida tem 68,26% de probabilidade de estar contida no intervalo $[\mu - \sigma; \mu + \sigma]$ e 95,44% é a probabilidade de estar contida no intervalo $[\mu - 2\sigma; \mu + 2\sigma]$ como ilustrado na Figura 3. Isto é, conforme aumenta o intervalo de confiança, também aumenta a probabilidade do valor medido estar nesse intervalo. Foram considerados como *outliers* (θ) os valores que estavam fora deste intervalo:

$$|\theta| > \mu + 2\sigma \quad (2)$$

No trabalho de Neagu, Grigoras e Scarlatache (2017), os *outliers* foram eliminados do conjunto de dados. Entretanto, foi verificado que com poucos parâmetros é possível identificar se uma medição é um valor discrepante ou não.

Nascimento et al. (2014) avaliaram e testaram métodos baseados em técnicas convencionais para a detecção de *outliers*: *Extreme Studentized Deviate* (ESD), *Generalized Extreme Studentized Deviate* (GESD), Z-Score, Z-Score Modificado, *Test Box Plot*, Thompson, *Adjusted Box Plot* e *Exponential Smoothing* (ExpSM). Todos os métodos testados apresentaram dificuldades nos ajustes dos parâmetros, pois existe uma dependência dos resultados de acordo com a variação da degradação da série. Isto é, à medida que mais *outliers* eram injetados o método se tornava menos eficaz. Dos supracitados, o que apresentou o melhor desempenho foi o Z-Score Modificado.

FIGURA 3. DISTRIBUIÇÃO GAUSSIANA.



Fonte: Retirada do PortalAction:
 <<http://www.portalaction.com.br/probabilidades/62-distribuicao-normal>>.
 Acesso em 25/05/2020.

O Método Z-Score é baseado na distribuição normal (gaussiana) que se $X \sim N(\mu, \sigma^2)$, então $Z = \frac{X - \mu}{\sigma} \sim N(0,1)$ e é definido por:

$$Z_{score} = \frac{x_i - \bar{x}}{s} \quad (3)$$

em que x_i é a medição atual, \bar{x} é a média das medições analisadas, e s é o desvio padrão das n medições dado por:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

Uma regra geral considera *outliers* as amostras com $|Z_{score}|$ maior que 3, contudo este valor depende do comportamento do sistema. Contudo, \bar{x} e s podem ser muito afetados pelos *outliers*. E a alternativa é uma modificação tornando o método mais robusto. No Método Z-Score Modificado, o \bar{x} é substituído pela média amostral \tilde{x} e s pelo desvio absoluto da mediana (MAD – *Median Absolute Deviation*), dado por:

$$MAD = median\{|x_i - \tilde{x}|\} \quad (5)$$

Assim, o Método Z-Score Modificado é expresso por:

$$M_i = \frac{0,6745}{MAD} (|x_i - \tilde{x}|) \quad (6)$$

em que: $E(MAD) = 0,6745\sigma$ para um grande conjunto de dados.

Para correção de *outliers*, foram testados, por Nascimento et al. (2014), métodos de médias aritméticas e ponderadas envolvendo as amostras das semanas e dos dias anteriores e o que

apresentou melhor resultado considerava como valor de preenchimento a média dos três dias anteriores no mesmo horário da amostra faltante.

Errico e Murru (1999) usaram uma abordagem Bayesiana aliada à lógica *fuzzy* para a identificação e tratamento dos outliers. Considerando x como o valor medido, μ como a média das amostras; assim tem-se que a distância é dada por $d(x, \mu) = |x - \mu|$ e a incerteza é expressa por $\bar{\sigma} = 100\sigma/\mu$. Então a distância *fuzzy* é obtida após a fuzzificação da distância $d = d(x, \mu)$, de acordo com:

Se $d(x, \mu) \geq 4\sigma$, a distância é muito longe.

Se $3\sigma \leq d(x, \mu) \leq 5\sigma$, a distância é longe.

Se $2\sigma \leq d(x, \mu) \leq 4\sigma$, a distância é média.

Se $d(x, \mu) \leq 3\sigma$, a distância é pequena.

E incerteza *fuzzy* é obtida por:

Se $\bar{\sigma} \geq 1\%$, a incerteza *fuzzy* é ampla;

Se $0,5\% \leq \bar{\sigma} \leq 1,5\%$, a incerteza é moderada.

Se $\bar{\sigma} \leq 1\%$ então a incerteza é pequena.

E quanto ao valor da saída que representa a amplitude do *outlier*:

Se $\rho \geq 0,5$ então a amplitude é alta.

Se $0,25 \leq \rho \leq 0,75$ então a amplitude é intermediária.

Se $\rho \leq 0,5$ então a amplitude é baixa.

O resultado $\rho(x)$ é utilizado para determinar o valor do peso (w) usado na estimação de acordo com as seguintes regras:

(w_1) se $d(x, \mu) \geq 5\sigma$, então $w(x) = 0$ (*outlier*).

(w_2) se $d(x, \mu) \leq 2\sigma$, então $w(x) = 1$ (medição normal)

(w_3) caso contrário $w(x) = 1 - \rho(x)$ (*outlier*)

Um pré-requisito necessário para uma previsão acurada de um sistema elétrico fotovoltaico é a alta qualidade dos dados. Com esse objetivo, Yu et al. (2016) utilizaram o Método dos Mínimos Quadrados para a correção de *outliers* nas curvas de potência fotovoltaica.

Para a detecção é obtida a média \bar{x} , o erro residual dado por $v_i = x_i - \bar{x}$. Neste procedimento foi calculado o desvio padrão quadrático de acordo com o Método de Bessel:

$$\sigma = \sqrt{\frac{\sum v_i^2}{n-1}} \quad (7)$$

Então:

Se $|x_i - \bar{x}| > 3\sigma$ a medição é considerada um *outlier* e;

Se $|x_i - \bar{x}| \leq 3\sigma$ é um dado normal.

Sarno (2014) propõe um algoritmo que utiliza um filtro de Kalman com um ajuste *fuzzy* para detecção e correção de *outliers*. Para cada medição feita denotada por z_k em que é associada uma variância fixa ($\bar{\sigma}^2$) que representa a falta de exatidão do instrumento de medição. Similarmente, σ_k^2 é a

incerteza associada ao valor estimado x_k , atualizada a cada passo de k para $k \in [0, L]$ para L igual ao número de medições total:

$$\sigma_k^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_{k-1}^2} \right)^{-1}, \quad 0 \leq k \leq L \quad (8)$$

então a estimação do estado da medição k é dada por:

$$x_k = \frac{\sigma^2 x_k + \sigma_{k-1}^2 z_k}{\sigma^2 + \sigma_{k-1}^2}, \quad 0 \leq k \leq L \quad (9)$$

em que,

$$\bar{x}_k = x_{k-1} \quad (10)$$

Então z_k é considerado um *outlier* quando $|z_k - \bar{x}_k| > 2\bar{\sigma}$ e um peso w é definido da seguinte forma:

$$\begin{cases} w(z_k) = 1, & \text{se } |z_k - \bar{x}_k| > 5\bar{\sigma} \\ 0 \leq w(z_k) \leq 1, & \text{se } 2\bar{\sigma} \leq |z_k - \bar{x}_k| < 5\bar{\sigma} \\ w(z_k) = 0, & \text{caso contrário.} \end{cases} \quad (11)$$

Para a condição $2\bar{\sigma} \leq |z_k - \bar{x}_k| < 5\bar{\sigma}$ o valor do peso w é definido de acordo com regras *fuzzy* (ERRICO; MURRU, 1999). Considerando que em uma distribuição gaussiana a probabilidade de uma amostra estar contida no intervalo $[\mu - 3\sigma, \mu + 3\sigma]$ é de 99%, se a condição $|z_k - \bar{x}_k| < 2\bar{\sigma}$, for ativada, então z_k não é considerado um *outlier*, visto que não está longe da primeira estimação. Mas se a condição $|z_k - \bar{x}_k| \geq 5\bar{\sigma}$ for ativada, então a medição z_k será considerada um *outlier*. E se a distância entre a

medição e a estimação possuir valores intermediários o ajuste *fuzzy* é adotado. A Equação (2.9) é modificada para:

$$x_k = \frac{\sigma^{-2} x_k + \sigma_{k-1}^2 w(z_k) z_k}{\sigma^{-2} + \sigma_{k-1}^2 w(z_k)}, \quad 0 \leq k \leq L \quad (12)$$

Desta forma a medição z_k é ponderada pelo peso w . E é feito um ajuste de acordo com a distância do ponto medido. Isto é, se o ponto é um *outlier*, o valor do peso será 1 e x_k é dado pelo cálculo do Filtro de Kalman convencional. Se o valor medido estiver em uma faixa intermediária, o peso é um valor entre 0 e 1 para corrigi-lo. E se o valor não for um *outlier*, o valor do peso é zero e o x_k se torna igual ao valor medido z_k .

Para a correção dos *outliers* proposta por Sarno (2014), o período de tempo estudado é dividido em grupos e é aplicado sobre cada grupo separadamente o Método dos Mínimos Quadrados para o preenchimento dos dados. Se o método é aplicado diretamente sobre a curva completa, tem-se o aumento significativo do erro, pois o algoritmo se tornava menos eficaz.

O monitoramento e controle do sistema elétrico requer uma avaliação confiável dos estados do sistema. Nos centros de controle mais modernos a avaliação é produzida com a estimação dos estados. Diante deste cenário, Okon e Wilkosz (2010) usaram o Método dos Mínimos Quadrados Ponderados (*Weighted Least Squares* – WLS) para estimação dos estados nas coordenadas polar

e retangular no sistema de potência. Então, com a estimação do estado é possível identificar se uma nova amostra é um *outlier*.

Foram analisadas as características do estimador WLS em relação ao número de estimações e a exatidão dos resultados. O resultado das pesquisas mostrou que o carregamento do sistema de potência e a redundância nos dados têm um impacto relevante na estimação dos estados.

Zhang et al. (2011) apresentam um filtro de Kalman com dois estágios para estimar as magnitudes e ângulos de fase das tensões. No primeiro estágio há a estimação dos estados estáticos e no segundo estágio dos estados dinâmicos.

O filtro de Kalman é usado em uma grande variedade de aplicações, sendo um método eficiente para estimar um processo de forma que minimize a média do erro quadrado quando o modelo do processo e das medições são acurados e obedecem determinadas propriedades estatísticas. O modelo do sistema pode ser representado por meio de um sistema em espaço de estados linear.

Um sistema linear pode ser modelado por um par de equações lineares:

$$x_k = Ax_{k-1} + w_{k-1} \quad (13)$$

$$z_k = Hx_k + v_k \quad (14)$$

Em que x_k é o vetor das variáveis de estados, A é a matriz de transição, w_{k-1} é a incerteza associada aos erros de previsão, z_k é

o vetor dos dados adquiridos, H é a matriz de medições e $v_k v_k$ é a incerteza associada aos erros de medição.

O método do Filtro de Kalman Descentralizado (*Uncentend Kalman Filter – UKF*) para estimação dinâmica de estados foi desenvolvido por Valverde e Terzija (2011) sendo baseado na Transformação Descentralizada combinada com a teoria do Filtro de Kalman. O desafio no processo de estimação é tratar o modelo matemático não-linear da rede como linear para que possa ser aplicado o algoritmo de estimação. Contudo com a transformação descentralizada essa limitação é superada e é possível melhorar a acurácia da estimação dos estados com uma implementação simples.

Os dados errôneos degradam a série temporal e consequentemente a exatidão do estimador e é necessário que sejam identificados. Então a vantagem de se ter um estimador de estados é poder avaliar a presença destas anomalias causadas por dados errôneos ou mudanças súbitas de estados.

Outras abordagens para detectar e corrigir os *outliers* se baseiam na aplicação das técnicas de Inteligência Artificial. Considerando as vantagens oferecidas pelas Redes Neurais Artificiais (RNAs) para a solução de problemas não-lineares, Melo e Castro (2014) propuseram a utilização de uma rede Perceptron Multicamadas (*Multi Layer Perceptron – MLP*) autoassociativa para o problema de identificação de *outliers* em dados de demanda

no setor elétrico. Uma rede MLP autoassociativa é um caso especial de uma Rede Neural na qual o vetor de entrada é igual ao vetor de saída que resulta em um mapeamento chamado armazenamento de vetor.

As leituras dos medidores do sistema elétrico geralmente são coletadas pelo sistema supervisório SCADA (*Supervisory Control and Data Acquisition*) e processadas por um estimador para filtrar o ruído e detectar erros. As medições contêm erros relacionados ao ruído, mas também dados errôneos, conhecidos na literatura como *bad data*, causado por faltas ou erros de comunicação (YANG et al., 2011).

Alves, Martins e Klautau (2017) investigaram a ocorrência de outliers existentes no *Big Data* gerado pelo sistema SCADA. O termo *Big Data* se refere a um grande e complexo volume de dados – estruturados ou não estruturados – que precisam de técnicas especiais para serem processados, analisados e assimilados (JAPEC et al., 2018).

O trabalho de Alves, Martins e Klautau (2017) se concentrou na mineração dos dados gerado pelo Sistema SCADA em busca de *outliers*, pois os dados do sistema contêm informações relevantes para a operação, manutenção e segurança; podendo assim melhorar a qualidade do serviço prestado pelas companhias do setor elétrico.

A integralidade correta dos dados do sistema de potência é tão importante que existe uma preocupação crescente com *cyber*-ataques que possam introduzir medições maliciosas no sistema. O ataque, por exemplo, pode ser direto sobre os medidores inteligentes de uma subestação ou sobre os computadores que armazenam os dados para injetar os dados maliciosos sendo injetadas as medições erradas (LIU; REITER; NING, 2009).

Para detectar a presença de ataques cibernéticos, Chakhchoukh et al. (2016) propõem uma técnica de inteligência artificial denominada estimação de densidade de taxa (*Density Ratio Estimation*) que fornece contramedidas nos ataques.

Ainda na correção de *outliers* em séries temporais, Li et al. (2017) desenvolveram um algoritmo baseado na predição com janela deslizante para tratar dados de temperatura de sensores com aplicações meteorológicas. Em que inicialmente a série é segmentada baseada na janela e então o modelo de previsão é feito de acordo com o histórico dos dados para estimar o dado futuro. E se a diferença do dado futuro com o dado medido for muito grande, será julgado se a medição é um *outlier* ou não de acordo com um valor limite, denominado valor de limiar (*threshold*).

A janela de previsão, determinada por $\eta_i^{(k)} = \{d_{i-2k}, d_{i-2k+1}, \dots, d_{i-1}\}$ em que k é a largura da janela, é

usada como parâmetro de entrada para o modelo de previsão de um passo-a-frente e o algoritmo de previsão é dado por:

$$d_{i+1} = M(\eta_i^{(k)}) \quad (15)$$

em que $M()$ é o modelo de previsão autoregressivo (AR). O Modelo AR é um método estatístico para prever o comportamento de séries temporais a partir do histórico das variáveis, assumindo que elas são lineares.

Uma abordagem diferente sobre *outliers* foi feita por Guo et al. (2012) em que o foco de detecção dos valores atípicos foi sobre o eixo das abscissas (eixo x), isto é, os dados de potência do sistema elétrico não estavam de acordo com a periodicidade usual. A curva de potência apresenta uma forma periódica (por exemplo, diária, semanal, mensal) e as medições se repetem regularmente ao longo de um intervalo fixo. Em Estatística, esse comportamento periódico é denominado sazonalidade e os valores faltantes de *outliers-x*. E quando ocorrem erros dos sensores, podem ocorrer irregularidades em que faltam dados no intervalo de tempo associado.

Farrokhifard, Hatami e Parniani (2015) desenvolveram uma aplicação *online* para detecção de *outliers* em dados de sistema de potência. A cada passo do algoritmo o valor estimado para a próxima amostra era calculado. O dado era recebido e verificado se era uma medição válida. Se sim, era integrado ao histórico. Caso contrário, era substituído pelo valor calculado na

iteração anterior. Assim, o conjunto dos dados era preservado em sua integralidade.

Rao, Akella e Guley (2011) desenvolveram um método para identificar *outliers* em um conjunto de sensores usando o sinal da rede de energia (*Power Line Communication* – PLC). A comunicação é feita através da infraestrutura da rede elétrica transportando sinais a partir de uma interface padrão. Foram analisadas as intensidades de sinal de 14524 medidores em uma área residencial durante 54 semanas e destes 677 apresentaram anomalias. Sobre cada sinal era calculada a distribuição de probabilidade $P(S_i)$ em que S_i é o nível de intensidade do sinal, T é o total de medições em uma semana e D é o número da intensidade de sinais distintos:

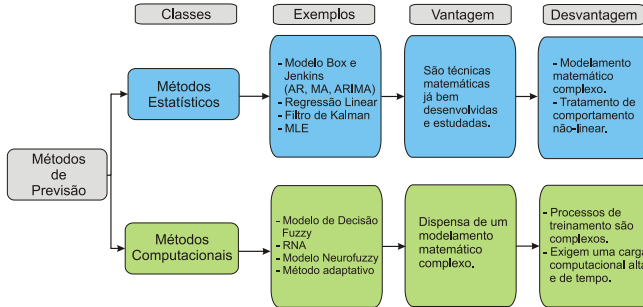
$$P(S_i) = \frac{N_i + 1}{T + D} \quad (16)$$

Os *outliers* provocam uma diminuição da intensidade do sinal e conseqüentemente uma diminuição da distribuição de probabilidade. Portanto, as medições que indicavam baixos valores de $P(S_i)$ eram as apresentavam medições errôneas.

A técnica utilizada para detecção e correção dos *outliers* depende muito do tipo de valor atípico, do comportamento e do nível de degradação da série temporal. Então para cada aplicação é necessário fazer um estudo sobre qual a melhor abordagem. E a previsão de séries temporais é contemplada na literatura através de diversos métodos que são divididos por Guirelli (2006) em

duas classes: os métodos estatísticos e os métodos de inteligência artificial ou computacionais; como são apresentados na Figura 4.

FIGURA 4. CLASSIFICAÇÃO DOS MÉTODOS DE PREVISÃO DE ACORDO COM GUIRELLI (2006).



Fonte: Elaborada pelo autor.

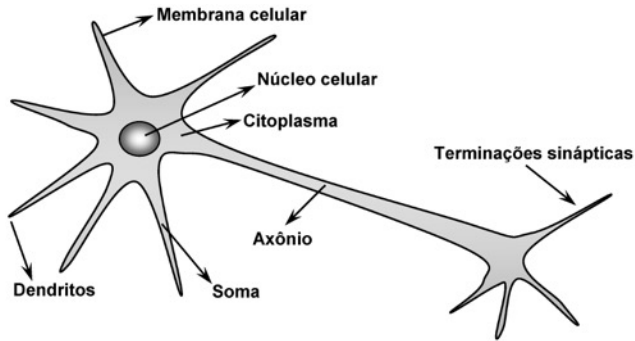
2.3 REDES NEURAIS ARTIFICIAIS

Uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso (HAYKIN, 2001). As redes neurais artificiais (RNA) se configuram como soluções para solução com comportamentos não lineares ou com uma grande quantidade de entradas.

No cérebro humano, o processamento de informações é regido por elementos processadores biológicos que operam em paralelo com o objetivo de pensar e memorizar denominados neurônios, conforme Figura 5. O cérebro humano é constituído por aproximadamente 100 bilhões de neurônios e cada neurônio

é interligado em média com outros 6000 neurônios em ligações denominadas sinapses (DA SILVA, 2010).

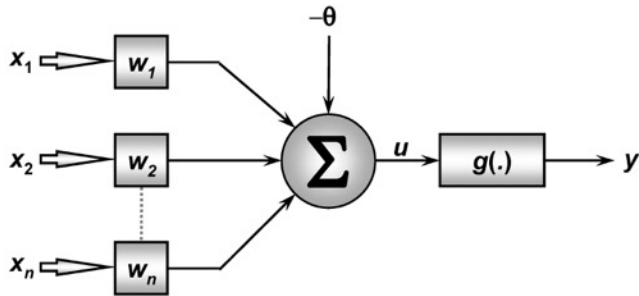
FIGURA 5. NEURÔNIO DO CÉREBRO HUMANO E SEUS CONSTITUINTES.



Fonte: Teodoro, Arthur & Gomes, Otávio. (2018).

Portanto, foi desenvolvido um modelo computacional baseado no comportamento dos neurônios e do cérebro humano para a resolução de problemas. O modelo matemático de um neurônio artificial é ilustrado pela Figura 6 com x_1, x_2, \dots, x_n sendo as entradas da rede neural, w_1, w_2, \dots, w_n sendo os pesos sinápticos, o bloco Σ o combinador linear, $-\theta$ é o limiar de ativação, u é o potencial de ativação, $g(.)$ a função de ativação e por fim, o sinal de saída dado por y .

FIGURA 6. NEURÔNIO ARTIFICIAL E SEUS CONSTITUINTES.



Fonte: Teodoro, Arthur & Gomes, Otávio. (2018).

As redes neurais se assemelham ao cérebro humano em dois aspectos:

1. O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.

2. Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

Para Haykin (2001), o uso de redes neurais oferece as seguintes propriedades úteis e capacidades:

- *Não-linearidade*: as redes neurais são opções para resolução de problemas não-lineares.
- *Mapeamento de Entrada-Saída*: refere-se à modificação dos pesos sinápticos de uma rede neural pela aplicação de um conjunto de *amostras de treinamento* rotuladas ou *exemplos da tarefa*, em que cada exemplo consiste de um

sinal de entrada único e de uma resposta desejada correspondente.

- *Adaptabilidade*: as RNA têm a capacidade de adaptar seus pesos sinápticos às modificações do ambiente, podendo ser retreinada para atender modificações operativas do sistema.
- *Informação contextual*: cada neurônio é potencialmente afetado pela atividade de todos os outros neurônios da rede e o conhecimento é representado pela própria estrutura e estado de ativação de uma rede neural.

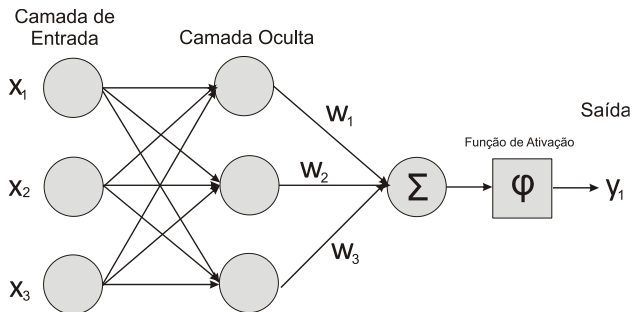
Segundo Zanini (2000), o termo “arquitetura” em redes neurais é utilizado para definir a forma com que os neurônios e suas conexões são arranjadas. As redes neurais, em geral, apresentam um nível de entrada, um nível de saída e camadas de unidades ocultas. As chamadas ocultas não são ligadas diretamente às entradas ou às saídas, mas são conectadas através dos pesos sinápticos. Na maioria das aplicações, basta a utilização de uma camada oculta (PEDREIRA, 2000).

As redes neurais podem ser classificadas em dois tipos: não-recorrentes e recorrentes. As redes não-recorrentes não possuem realimentação da saída para as es entradas e são consideradas “sem memória” (HAYKIN, 2000). As redes não-recorrentes são denominadas *feedforward* em que o sinal é sempre

propagado para as próximas camadas até a saída, como ilustrado na Figura 7.

As redes recorrentes possuem ao menos um laço de realimentação fazendo com que as redes respondam às entradas de forma dinâmica. A cada nova entrada, a saída é calculada e é realimentada para modificar a entrada, sendo consideradas redes “com memória”. A realimentação melhora de forma significativa a capacidade de aprendizagem e o desempenho da rede.

FIGURA 7. ARQUITETURA DE UMA REDE NEURAL COM APENAS UMA CAMADA OCULTA.



Fonte: Elaborada pelo autor.

2.4 LÓGICA FUZZY

A “lógica *Fuzzy*”, também denominada lógica difusa ou nebulosa, foi desenvolvida em 1965 por Lotfi A. Zadeh com a proposta da Teoria dos Conjuntos Difusos. A lógica convencional é binária ao considerar proposições como “Verdadeira” ou “Falsa”, “Sim” ou “Não”, “Ligado” ou “Desligado”. Enquanto que a lógica

Fuzzy se aproxima mais do pensamento humano ao tentar modelar, através do senso das palavras, informações que podem ser vagas ou incertas. Isto é, na lógica *Fuzzy*, o valor-verdade de uma proposição é um subconjunto *Fuzzy* ao contrário da lógica binária que só assume dois valores: 0 ou 1 (SHAW; SIMÕES, 2004).

Lee (1990) define um conjunto *Fuzzy* F em um universo como uma função de pertinência μ_F que assume valores no intervalo $[0; 1]$ denominado $\mu_F: U \in [0; 1]$. Então um conjunto *Fuzzy* em U pode ser representado por pares ordenados do elemento genérico e de seu valor de pertinência dado pela função:

$$F = \{(u, \mu_F(u)) \mid u \in U\} \quad (17)$$

Para um conjunto universo U contínuo, o conjunto *fuzzy* F pode ser escrito por:

$$F = \int_U \mu_F(u) / u \quad (18)$$

Portanto, a função de pertinência $\mu_F(u)$ indica o grau de compatibilidade entre u e o conceito expresso por U . Se $\mu_F(u) = 1$, u é totalmente compatível com U . E se $\mu_F(u) = 0$, u é totalmente incompatível com U . O valor u será parcialmente compatível com U se $0 < \mu_F(u) < 1$.

A lógica *fuzzy* está baseada em expressões linguísticas e não em números, ou seja, os valores de pertinência são dados através de palavras. Isto favorece o tratamento de imprecisões e simplifica a solução de problemas. Por exemplo, se através da lógica

booleana assumirmos que a corrente elétrica é considerada perigosa para o ser humano se possuir valor maior que 50mA e é considerada inofensiva se possuir magnitude menor que 50mA. Portanto, uma corrente de 49,9mA seria considerada inofensiva, representando uma incoerência. Então uma forma de modelar esse problema poderia ser dividir o intervalo da corrente em conjuntos como Imperceptível, Baixa, Mediana e Perigosa.

Devido a essas características de se aproximar do raciocínio humano, a lógica *fuzzy* vem sendo aplicada em diversas áreas como: controle industrial, sistemas especialistas; controle de processos; modelamento de sistemas parcialmente abertos; apoio na tomada de decisão; robótica;

2.5 MODELO AUTORREGRESSIVO ARIMA

Para o estudo de séries temporais é importante definir se a série é estacionária ou não, para, a partir daí, estabelecer a estrutura do modelo probabilístico que estimará a série. De acordo com Diggle (1992), uma série é definida como estacionária quando suas observações ocorrem aleatoriamente ao redor de uma média constante, ou seja, não há tendência. Isto é, $E(Z(t)) = E(Z(t + \Delta t)) = \mu$ em que μ é a média; e $Var(Z(t)) = Var(Z(t + \Delta t)) = cte$ (MORETIN; TOLOI, 2004).

Diggle (1992) define, ainda, a função de autocorrelação que, em cada período p (*lag p*) da série, é calculado o

coeficiente de correlação entre as observações t e $t + p$. Neste caso, se t e $t + p$ são independentes, a correlação entre t e $t + p$ é zero.

Uma variedade de métodos de estimação pode ser utilizada para dados de séries temporais e não existe um modelo que é aplicado de forma universal. Entretanto, o modelo ARIMA tem dominado muitas áreas de previsão. Este modelo tem uma funcionalidade atrativa que promove uma adequação de acordo com os dados. O modelo ARIMA contém três partes, denominadas autorregressiva (AR), Integradora (I) e média móvel (MA – *Moving Average*); e estabelece a relação entre o valor atual e os valores anteriores (HE et al., 2012).

Tanto o modelo AR, quanto o MA, quanto o ARMA são utilizados para séries estacionárias. Entretanto, para processos não-estacionários homogêneos, isto é, os que flutuam ao redor de um nível por um período de tempo depois muda de nível e passa a flutuar sobre este, e assim sucessivamente. Uma das maneiras de analisá-lo é incorporando um processo de diferenças do modelo ARMA em que d é a ordem das diferenças necessárias para tirar a tendência da série (MORETIN; TOLOI, 1985).

De acordo com Carvalho, Teixeira e Matos (2016), a equação que descreve o modelo ARIMA (p,d,q) é:

$$\hat{y}^d(t) = c + \Phi_1 y^d(t-1) + \dots + \Phi_p y^d(t-p) + \theta_1 e(t-1) + \dots + \theta_q e(t-q) + e(t) \quad (19)$$

em que $\hat{y}^d(t)$ é o valor estimado, p é a ordem da parte autorregressiva, d é o grau de diferença da série temporal original, q é a ordem da média móvel, $e(t-q) = y(t-q) - \hat{y}(t-q)$, $e(t)$ é o ruído branco, c é uma constante e Φ são os coeficientes da parte autorregressiva e θ são os coeficientes da parte média móvel.

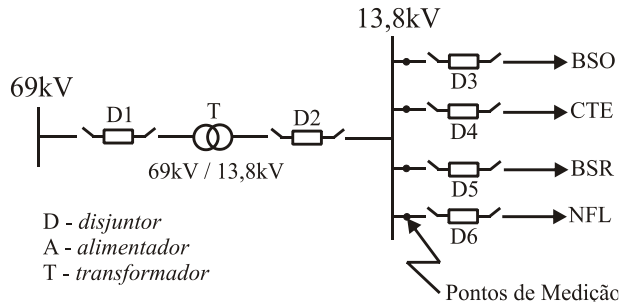
3. PROCESSAMENTO DE *OUTLIERS* EM DADOS DO SISTEMA ELÉTRICO

Este capítulo apresenta a técnica de detecção utilizada e as três técnicas de correção para os dados de interesse. Foram implementadas três técnicas para a correção dos valores atípicos. Duas técnicas foram baseadas em inteligência artificial e utilizam os dados passados dos perfis de carga ao invés de utilizar somente os valores atuais: a primeira é uma abordagem usando a Lógica *Fuzzy* e a segunda usando Redes Neurais Artificiais. A terceira técnica é baseada no modelo autorregressivo do tipo ARIMA que modela a variável de saída de acordo com os valores anteriores.

3.1 CURVAS DE DEMANDA

As curvas de demanda foram medidas nas barras dos alimentadores de uma subestação de distribuição real localizada no estado da Paraíba. Os pontos de medição estão posicionados nos troncos dos quatro alimentadores, como ilustrado no modelo simplificado da subestação abaixadora da Figura 8.

FIGURA 8. MODELO SIMPLIFICADO DA SUBESTAÇÃO ESTUDADA.

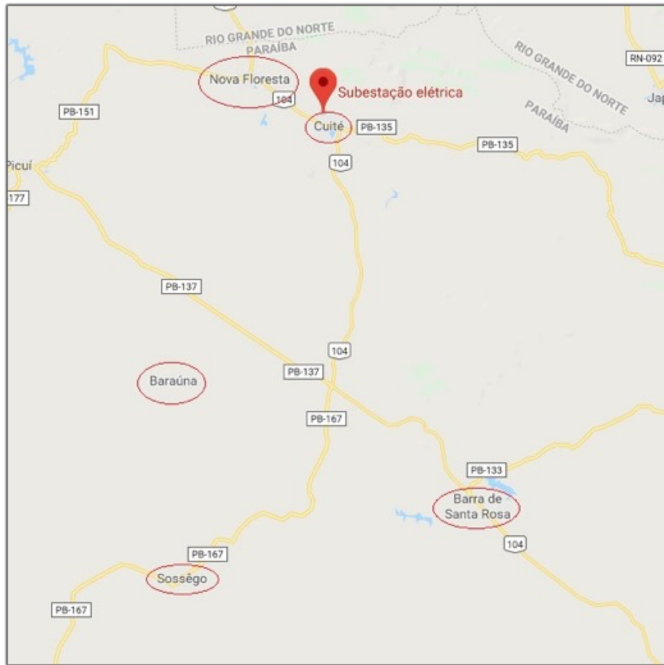


Fonte: Elaborada pelo Autor.

Os alimentadores são denominados neste trabalho de acordo com a região de distribuição: BSO (Baraúna e Sossego), CTE (Cuité), BSR (Barra de Santa Rosa) e NFL (Nova Floresta), ilustradas na Figura 9.

As subestações são monitoradas pelo Sistema de Aquisição de Dados (SCADA – *Supervisory Control and Data Acquisition*) e os dados coletados são tensão (V), corrente (I), e potência (P); a cada 15 (quinze) minutos; sendo no total 96 medições por dia para cada grandeza.

FIGURA 9. ÁREA COM AS CIDADES ALIMENTADAS PELA SUBESTAÇÃO DE CUITÉ.

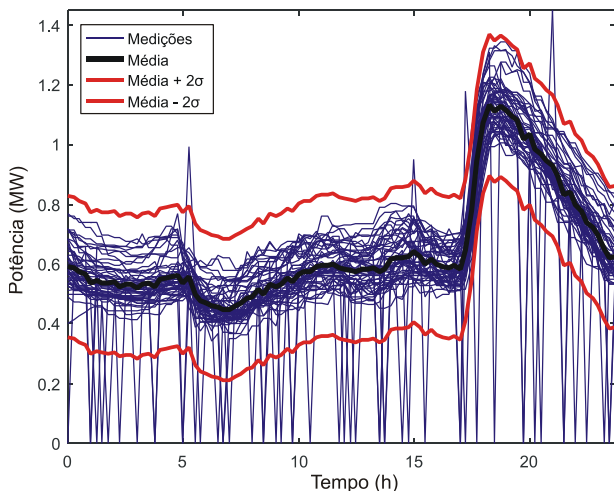


Fonte: Adaptada do Google Maps.

Os *outliers* identificados nesses conjuntos de dados e, que serão identificados na metodologia proposta, são de dois tipos principais: zeros ou picos (*spikes*). Quando há um problema no sensor ou quando há uma falta no fornecimento de energia para o sistema de medição pode haver ausência de dados indicando erroneamente que os valores são zeros, corrompendo o histórico das medições. Em outros casos podem haver os picos (*spikes*) que são observações com um valor muito superior aos outros pontos do comportamento padrão da curva, devido instabilidade nos sensores (NASCIMENTO et al., 2014).

Na Figura 10 ilustram-se a superposição das curvas de carga de todas as segundas feiras no ano de 2010, a curva média e as curvas da média mais e menos o dobro do desvio padrão. Neste alimentador, em geral, não existem muitos *outliers* em conjunto e isso favorece a aplicabilidade de um método de estimação de curtíssimo prazo. Para grandes intervalos com ausência de dados, outras técnicas devem ser utilizadas. Observa-se que esta curva pode ser utilizada para definir um intervalo de dispersão das medições para um dia específico de medição, para cada instante de tempo.

FIGURA 10. CURVAS DE TODAS AS SEGUNDAS-FEIRAS DO ANO 2010 NO ALIMENTADOR ESTUDADO.



Fonte: Elaborada pelo autor.

Outros tipos de *outliers* também podem estar presentes nas medições, como por exemplo, transferências de carga de uma subestação para outra que afetam um conjunto contínuo de

medições que variam com a duração da redistribuição do fluxo de potência.

Diante deste contexto, as curvas de carga frequentemente encontram-se deturpadas com *outliers*, os quais devem ser detectados e corrigidos para plena utilização nas atividades relacionadas como operação, manutenção e planejamento do sistema elétrico de potência.

É possível ainda agregar a essa nova topologia um sistema de alerta para a central da concessionária quando houver um grande conjunto de valores atípicos, indicando que o sistema está com algum problema permanente que pode ser um sensor defeituoso ou uma queda de energia.

3.2 METODOLOGIA

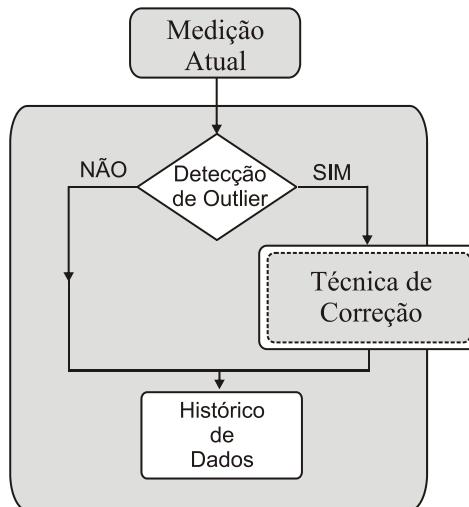
O objetivo deste trabalho consiste no desenvolvimento do algoritmo de detecção e correção de *outliers* e serão considerados como estudo de caso: os vales (ausência de dados) e os picos (*spikes*).

Os algoritmos de correção de *outliers* para curtíssimo prazo levam em consideração o fato de que a estimação do valor ausente deve seguir o padrão apresentado nas últimas leituras observadas, e não para a relação da demanda com fatores que a influenciariam como condições meteorológicas, sociais e econômicas. Estas considerações são válidas principalmente quando são desenvolvidos sistemas de previsão de demanda de

minutos ou horas à frente e com medições em tempo real que requerem alta exatidão (CHARYTONIUK, 2000).

O algoritmo de detecção e correção de *outliers* desenvolvido é ilustrado na Figura 11 em que a técnica utilizada para correção dos *outliers* pode ser substituída. Isto é, o algoritmo é o mesmo, contudo a metodologia de correção é modular, pois pode-se substituí-la de acordo com as características do perfil de carga.

FIGURA 11. ALGORITMO DE DETECÇÃO E CORREÇÃO DE *OUTLIERS*.



Fonte: Elaborado pelo autor.

Foram desenvolvidas uma técnica de detecção de *outliers* e três técnicas de correção de *outliers*. A detecção foi feita de acordo com os padrões típicos dos perfis de carga para cada dia da semana. E para a correção, a primeira abordagem é baseada em

Lógica *Fuzzy* e estabelece relações entre os incrementos diferenciais da curva de carga das três últimas amostras. Assim o intuito é definir quanto será o acréscimo atual ao valor que foi medido. A segunda abordagem utiliza Redes Neurais Artificiais com uma janela de valores anteriores reduzida para estimar o próximo valor de entrada. E por fim, é feita uma modelo autoregressivo do tipo ARIMA para estimação. Todas as técnicas desenvolvidas neste trabalho foram implementadas no software MATLAB®.

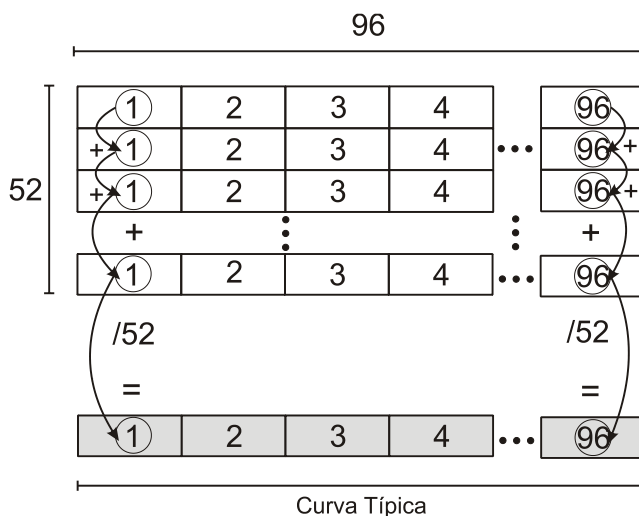
3.3 ALGORITMO DE DETECÇÃO DE *OUTLIERS*

As curvas utilizadas para construção dos padrões típicos foram retiradas do período de 1 ano da subestação e além de serem sazonais, não apresentam tendência. Inicialmente, foi feita a separação dos dados do período de um ano de acordo com o dia da semana, isto é, todos os domingos foram agregados em uma única matriz, assim consecutivamente para os outros seis dias. As matrizes tinham dimensão 96x52, pois foram 52 semanas com 96 amostras para cada.

Então sobre cada uma dessas matrizes foi calculado índice a índice, a sua respectiva média, como ilustrado na Figura 12. No índice 1, referente ao horário 00h, foram somados todos os valores das 52 semanas e dividido por 52. Prosseguindo, desta forma, para todos os índices do dia fechando em 96 referente às

23h45. O cálculo da curva média, denominada Curva Típica, foi feito para cada dia da semana. E sobre cada conjunto de índices também foi calculado o desvio padrão respectivo.

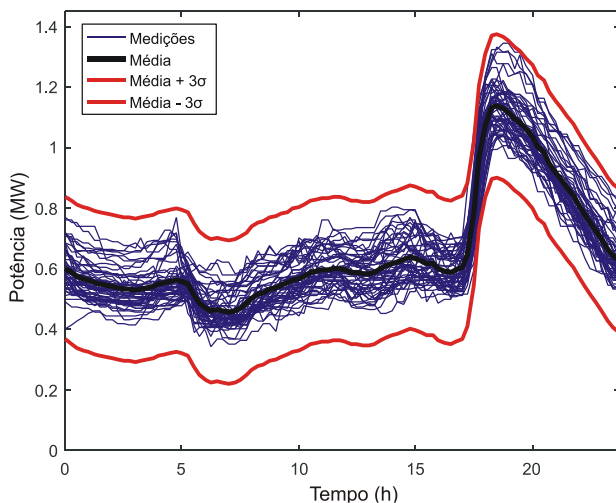
FIGURA 12. CONSTRUÇÃO DAS CURVAS TÍPICAS A PARTIR DA MÉDIA DE CADA ÍNDICE DO RESPECTVO DIA DA SEMANA.



Fonte: Elaborada pelo autor.

A Figura 13 ilustra todas as sextas-feiras do ano de 2010 em que a curva de cor preta é a média de todos os dias e é denominada Sexta-Feira Típica. As curvas em vermelho representam a média mais e menos o triplo do desvio padrão, pois para este intervalo uma distribuição gaussiana compreende 99,73% de todas as amostras, sendo então este o critério que será adotado para a detecção dos *outliers*.

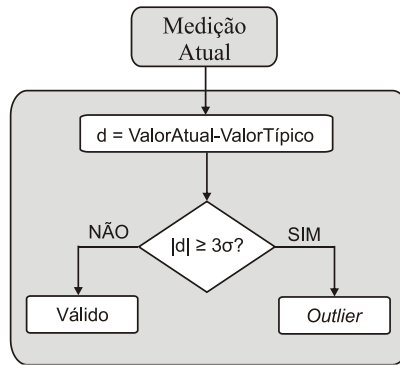
FIGURA 13. CURVAS DE TODAS AS SEXTAS-FEIRAS DO ANO DE 2010, CURVA DA SEXTA FEIRA TÍPICA E MAIS E MENOS 3 VEZES O DESVIO PADRÃO.



Fonte: Elaborada pelo autor.

Conforme a Figura 14, que ilustra o algoritmo de detecção de *outliers*; para cada novo valor de entrada, o algoritmo de detecção verifica o dia da semana e o índice respectivo dentro do intervalo [1; 96] referente ao intervalo das 00h00 às 23h45. Por exemplo, se o *outlier* ocorrer no índice 5 (1h da manhã) de um domingo, então será feita a subtração do valor atual medido pelo índice 5 (1h da manhã) da curva do domingo típico. Caso o módulo da diferença (d) seja maior que o triplo do desvio padrão máximo (σ), então o valor medido será considerado um *outlier*.

FIGURA 14. ALGORITMO DE DETECÇÃO DE *OUTLIERS* DESENVOLVIDO.



Fonte: Elaborada pelo autor.

Portanto, a detecção consiste em avaliar se o valor medido está dentro do intervalo do dia respectivo e este algoritmo permite detectar tanto *outliers* do tipo vale como do tipo pico. A Figura 15 ilustra a curva medida no período de 1 dia e a curva típica correspondente e neste perfil de carga foram injetados 4 *outliers* (2 picos e 2 vales).

Os quatro valores são detectados de acordo com os valores limites ilustrados na Figura 16, em que cada ponto é a diferença entre o valor medido e o valor típico da curva e, portanto, os que estão fora da área de cobertura são considerados *outliers*.

Uma desvantagem deste algoritmo é que em alguns casos algumas amostras válidas ficaram fora da faixa limite e foram consideradas medições errôneas mesmo sem sê-las, forçando uma correção que não deveria ser realizada.

FIGURA 15. CURVA TÍPICA E CURVA MEDIDA COM 4 *OUTLIERS* PARA DETECÇÃO.

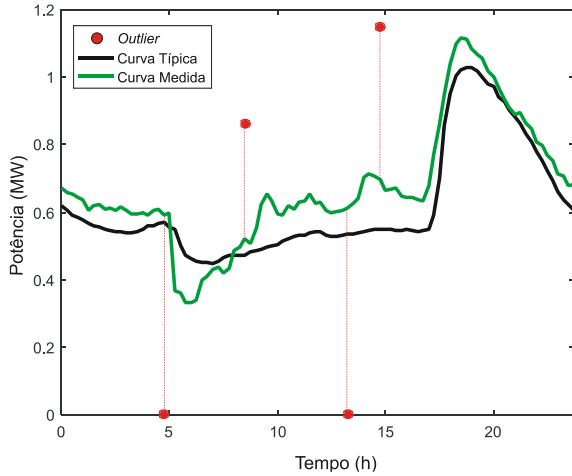
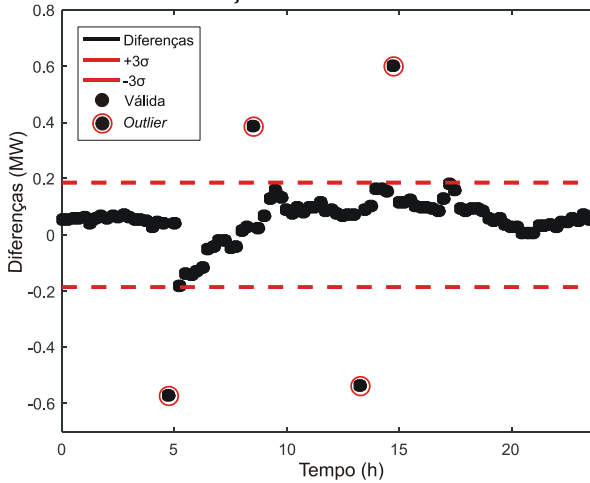


FIGURA 16. DIFERENÇAS ENTRE O VALOR MEDIDO E O VALOR TÍPICO PARA DETECÇÃO DOS *OUTLIERS*.



Fonte: Elaboradas pelo autor.

3.4 ALGORITMO DE CORREÇÃO DE *OUTLIERS* BASEADO EM LÓGICA *FUZZY*

A lógica *fuzzy* apresenta uma grande vantagem que é não exigir um complexo desenvolvimento matemático e estatístico. A partir dos conhecimentos do comportamento do sistema, o especialista é capaz de modelar o valor da saída em função das entradas.

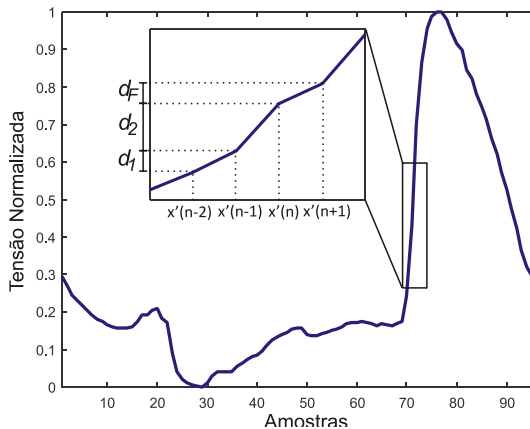
3.4.1 *Desenvolvimento do Modelo Fuzzy*

Inicialmente, é feita uma separação do histórico de demanda de energia nos sete dias da semana e o valor máximo e o valor mínimo são calculados sobre cada um deles. Com o valor máximo e o valor mínimo de cada dia da semana é feita a normalização dos valores para o intervalo entre 0 e 1:

$$x'(n) = \frac{x(n) - \min(x)}{\max(x) - \min(x)} \quad (20)$$

em que $x'(n)$ é o valor medido normalizado, $x(n)$ é o valor de entrada atual e $\min(x)$ e $\max(x)$ são funções que retornam os valores de entrada mínimo e máximo, respectivos ao dia da semana da amostra em estudo. A normalização é importante, pois assim o bloco *fuzzy* pode ser utilizado por qualquer outra subestação que possua o mesmo perfil de carga.

FIGURA 17. GRÁFICO COM A INDICAÇÃO DOS INCREMENTOS DE ENTRADA E SAÍDA DO BLOCO FUZZY.



Fonte: Elaborada pelo autor.

A técnica de correção *Fuzzy* é baseada nos incrementos diferenciais dos pontos medidos e a partir da magnitude e do sinal das entradas é dado o incremento diferencial de saída, como mostrado na Figura 17.

FIGURA 18. MODELO FUZZY DESENVOLVIDO COM OS INCREMENTOS DIFERENCIAIS DE ENTRADA E SAÍDA.



As entradas do modelo *fuzzy*, ilustrado na Figura 18, são os incrementos anteriores, denominados *deltas* d_1 e d_2 para calcular o valor posterior:

$$d_1 = x'(t-1) - x'(t-2) \quad (21)$$

$$d_2 = x'(t) - x'(t-1) \quad (22)$$

em que $x'(t-1)$ é o valor anterior normalizado, $x'(t-2)$ é o valor que precede o valor anterior e $x'(t)$ é o valor de entrada atual normalizado. A saída do bloco é o incremento (d_F) que deve ser somado ao valor atual resultando no valor estimado $x'(t+1)$:

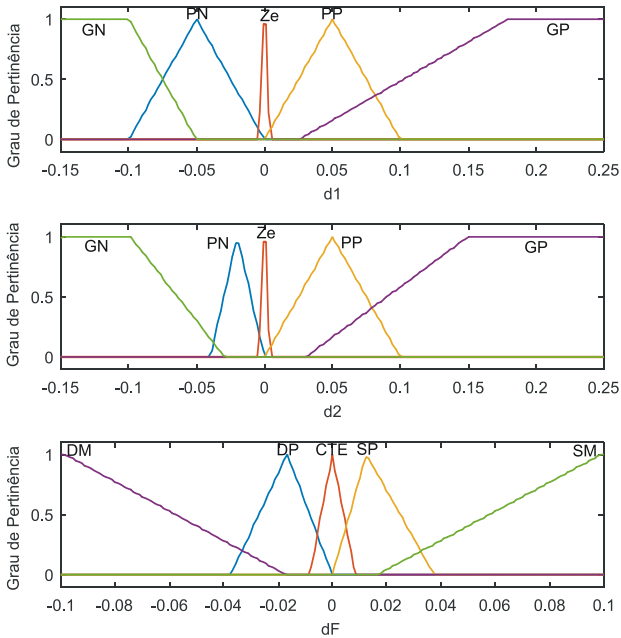
$$x'(t+1) = x'(t) + d_F \quad (23)$$

O dado adquirido passa pelo algoritmo de detecção de *outlier*. Se não for, o mesmo valor é integrado ao histórico, como ilustrado na Figura 11. Caso seja um valor atípico, o dado é substituído pelo valor estimado na iteração anterior. A cada passo, o algoritmo normaliza o dado usando os parâmetros do respectivo dia de semana, calcula os incrementos e passa como entrada para o bloco *Fuzzy* que retorna o incremento de saída “delta” (d_F) que será somado ao valor atual. Por fim, o sinal é reconstruído aplicando a função inversa da normalização. Esse resultado é o valor estimado para a próxima iteração e se a amostra subsequente for um *outlier* será substituída por ele.

O bloco *fuzzy* foi implementado utilizando a *toolbox Fuzzy* do MATLAB® em que após modelagem do sistema é possível definir os conjuntos das entradas e saída, além de definir as regras. Cada incremento de entrada (d_1 e d_2) foi dividido em cinco conjuntos *fuzzy*: Grande Negativo (GN), Pequeno Negativo (PN), Zero (Ze), Pequeno Positivo (PP) e Grande Positivo (GP). E o valor de saída

era estabelecido de acordo com a dinâmica do sistema: Desce Muito (DM), Desce Pouco (DP), Constante (CTE), Sobe Pouco (SP) e Sobe Muito (SM), de acordo com a Figura 19.

FIGURA 19. CONJUNTOS FUZZY CRIADOS PARA OS INCREMENTOS DIFERENCIAIS DE ENTRADA E SAÍDA.



Fonte: Elaborado pelo autor.

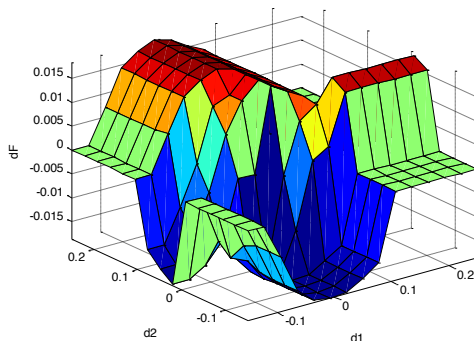
A superfície do sistema de inferência *Fuzzy* desenvolvido é apresentada na Figura 20 que ilustra os intervalos das variáveis de entrada e o intervalo do incremento de saída do bloco *fuzzy*. Através da superfície, gerada pelo MATLAB, é possível com-

prender como as entradas e as regras influenciam no valor da saída (d_F).

Foram construídas 25 regras (5 conjuntos de $d_1 \times 5$ conjuntos de d_2), expostas na Tabela 1, de acordo com o estudo do comportamento dos perfis das curvas de carga. Os principais focos de análise eram os pontos angulosos, isto é, quando havia inversões de sentido e os incrementos passavam de positivo para negativo ou vice-versa, principalmente nos horários de ponta. Se a regra for mal programada, o preenchimento pode ser feito com outro valor atípico, inviabilizando o algoritmo.

Algumas considerações foram incluídas na modelagem do bloco: se d_1 e d_2 forem do conjunto Grande Positivo (GP), indica que já houve a subida do pico do dia e o próximo valor subiria pouco ou seria constante. Da mesma forma na descida, para d_1 e d_2 no conjunto GN o valor deve permanecer constante, pois entra em uma zona de maior constância.

FIGURA 20. SUPERFÍCIE DE INFERÊNCIA FUZZY.



Fonte: Elaborada pelo autor.

Outra observação feita após os estudos sobre as curvas é que quando o d_2 é zero a saída deve permanecer constante porque tende a diminuir o erro. Isso acontece porque tanto a curva pode subir suavemente mesmo com d_1 negativo como pode descer suavemente mesmo com o d_1 positivo. Portanto, as regras devem ser bem construídas baseadas no conhecimento do problema e da experiência do especialista que está construindo o sistema de inferência *fuzzy*, pois se houver alguma inconsistência a correção pode ser feita com outros valores atípicos.

TABELA 1. REGRAS FUZZY.

d_1	d_2	d_F
GN	GN	DP
GN	PN	CTE
GN	Ze	DP
GN	PP	DP
GN	GP	CTE
PN	GN	DP
PN	PN	DP
PN	Ze	CTE
PN	PP	CTE
PN	GP	SP
Ze	GN	DM
Ze	PN	DP
Ze	Ze	CTE
Ze	PP	SP
Ze	GP	SM
PP	GN	DP
PP	PN	CTE
PP	Ze	CTE
PP	PP	SP
PP	GP	SP
GP	GN	CTE
GP	PN	CTE
GP	Ze	SP
GP	PP	SP
GP	GP	SP

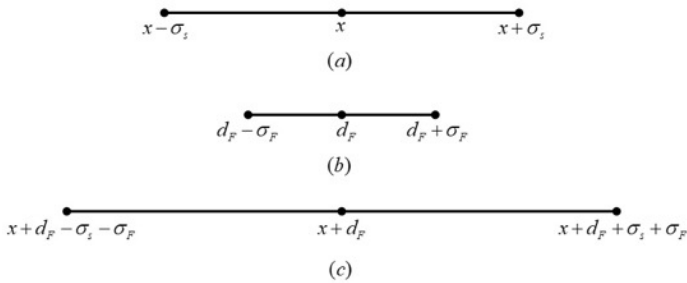
3.4.2 Avaliação de Incertezas

Com a finalidade de avaliar quantitativamente o resultado da abordagem de correção *Fuzzy*, pode-se associar uma incerteza à estimação. O resultado da correção *Fuzzy* foi expresso na forma intervalar, como ilustrado na Figura 20. Essa maneira indica mais confiabilidade no grau de incerteza calculada, pois o diâmetro da solução do intervalo é uma indicação do erro relacionado ao sistema *fuzzy* e ao sensor. Em Matemática Intervalar para dois intervalos $X, Y \in \mathbb{R}$ com $X = [x_1; x_2]$ e $Y = [y_1; y_2]$, a operação de adição é dada por:

$$X + Y = [(x_1 + y_1); (x_2 + y_2)] \quad (24)$$

Desta forma, de um ponto de vista metrológico, o resultado de estimação (x) torna-se mais completo. A Figura 20(a) considera a medição dentro de um intervalo com a incerteza associada ao sensor que foi adotada por $\sigma_5 = 0,03$ que é a maior classe de exatidão medições em potência, de acordo com a NBR 6855/2009 da ABNT.

FIGURA 21. INTERVALO DA MEDIÇÃO, INTERVALO DO INCREMENTO E SOMA DOS INTERVALOS.



Fonte: Elaborada pelo autor.

A Figura 20(b) ilustra o incremento de saída do bloco fuzzy (d_F) dentro do intervalo representado pela incerteza do algoritmo *fuzzy*, que de acordo com Zhai e Mendel (2011) é dada por:

$$\sigma_F = \sqrt{\frac{1}{N} \sum_i^N [\bar{x} - x_i]^2 \mu(x_i)} \quad (25)$$

em que:

$$\bar{x} = \frac{\sum_i^N x_i \mu(x_i)}{\sum_i^N \mu(x_i)} \quad (26)$$

para N igual ao número de elementos, \bar{x} a média (centro geométrico) e $\mu(x_i)$ como o valor do grau de pertinência do conjunto *fuzzy*.

Portanto, a incerteza *fuzzy* é dada por:

$$\sigma_{TF} = \sigma_F + \sigma_S \quad (27)$$

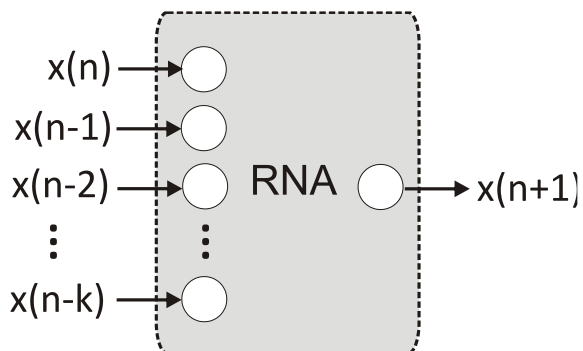
e é representada no intervalo da Figura 20(c) que ilustra a soma dos intervalos.

3.5 ALGORITMO DE CORREÇÃO DE *OUTLIERS* BASEADO EM REDES NEURAI ARTIFICIAIS

Assim como a lógica *Fuzzy*, trabalhar com Redes Neurais Artificiais dispensa um desenvolvimento matemático e estatístico complexo. E, diferentemente da Lógica *Fuzzy*, não é necessário conhecer de forma profunda a dinâmica do sistema.

O algoritmo recebe como entradas as amostras anteriores e a saída é o valor estimado para o próximo instante. A Figura 22 ilustra a entrada expressa pelas medições denotadas por $x(n)$, $x(n - 1)$, ..., $x(n - k)$ em que k é o tamanho da janela e a saída dada por $x(n + 1)$, construindo-se assim um modelo autorregressivo.

FIGURA 22. BLOCO DA REDE NEURAL COM AS AMOSTRAS ANTERIORES COMO ENTRADAS.

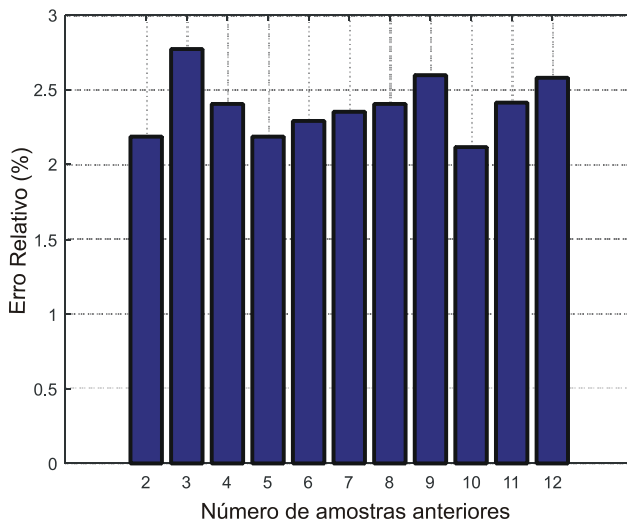


Fonte: Elaborada pelo autor.

Foram feitos testes em uma semana de dados com injeção aleatória de *outliers* usando janelas de dados anteriores de 2 até 12 amostras. Foram calculados os erros relativos de cada um desses testes. Apesar dos erros serem próximos, a janela de melhor desempenho foi a que apresentava 10 amostras, como ilustrado na Figura 23 que retrata os erros relativos de acordo com o número de amostras anteriores.

O algoritmo foi feito em três etapas: tratamento dos dados, treinamento da rede e aplicação nas curvas de demanda.

FIGURA 23. ERRO RELATIVO DA TÉCNICA RNA DE ACORDO COM O NÚMERO DE AMOSTRAS ANTERIORES.



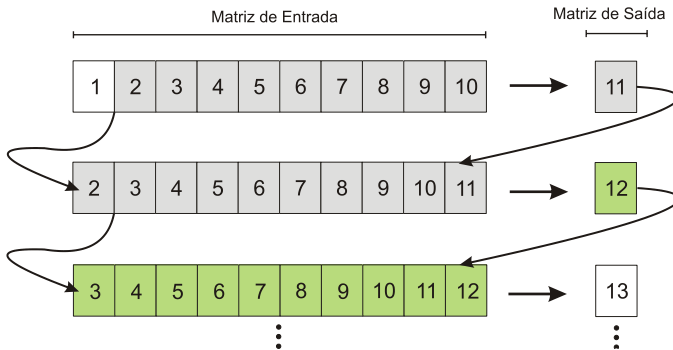
Fonte: Elaborada pelo Autor.

3.5.1 Tratamento dos dados

Os dados para treinamento foram retirados aleatoriamente do histórico disponível e inicialmente foi feita a separação de todos os dias da semana do período de um ano resultando em uma matriz para cada dia da semana.

Para criação das matrizes de entrada e saída (objetivo) da rede neural foram selecionados 20 dias do histórico de dados. Na matriz de entrada se apresentam os valores de 10 amostras anteriores e na matriz de saída o valor subsequente, como ilustrado na Figura 24.

FIGURA 24. TRATAMENTO DOS DADOS COM A CRIAÇÃO DAS MATRIZES DE ENTRADA E DE SAÍDA.



Fonte: Elaborado pelo autor.

3.5.2 Treinamento e Validação

As entradas foram organizadas em 80% para treinamento, 10% para validação e 10% para teste. Foi utilizado o treinamento Levenberg-Marquardt com 10 neurônios na primeira camada e 1 na camada escondida e função de ativação do tipo tangente hiperbólica.

Foi escolhido o método de Levenberg-Marquardt por ser bastante efetivo e rápido para redes neurais pequenas. É um método que se concentra na resolução de problemas de mínimos quadrados não-lineares e é usado para encontrar o mínimo local do erro da função, não sendo necessariamente o mínimo global.

3.5.3 Avaliação de Incertezas

Para calcular a incerteza da abordagem RNA é necessário considerar o comportamento não-linear do sistema. Cada entrada é transformada em um modelo gaussiano utilizando a incerteza do sensor (σ_s) como desvio padrão. Então, de acordo com Gal e Ghahramani (2016), a rede neural é ativada e o modelo da incerteza pode ser aproximado por:

$$\sigma_R = \sqrt{\frac{1}{B} \sum_{b=1}^B [\hat{y}^*(b) - \bar{y}^*]^2} \quad (28)$$

onde B é o número de repetições, \hat{y}^* é a saída da rede neural e:

$$\bar{y}^* = \frac{1}{B} \sum \hat{y}^*(b) \quad (29)$$

3.6 ALGORITMO DE CORREÇÃO DE *OUTLIERS* BASEADO ARIMA

Há uma grande variedade de métodos de estimação que podem ser utilizados para a previsão de séries temporais. E como o perfil das curvas de carga é bastante similar, uma opção interessante é utilizar a técnica ARIMA que tem sido amplamente utilizada para previsão e será usada neste trabalho como comparativo em relação às outras duas técnicas.

3.6.1 Modelagem

Uma série temporal é dita sazonal quando os eventos que ocorrem durante o intervalo de tempo de estudo se repetem em períodos idênticos de tempo, por exemplo pode ser uma repetição anual, mensal, diária (MORENTIN; TOLOI, 2004). As curvas estudadas possuem uma sazonalidade diária, referentes às 96 medições realizadas durante o dia.

A implementação do modelo ARIMA é feita no MATLAB® e, além dos três parâmetros p, d e q , são considerados os parâmetros da sazonalidade característica do perfil de carga das subestações resultando na forma $ARIMA(p, d, q)(P, D, Q)_s$, em que p e q são os coeficientes não-sazonais do ARIMA, d é o número de diferenças não-sazonais, P é o número de coeficientes autorregressivos multiplicativos, D é o número de diferenças sazonais, Q é o número de coeficientes da média móvel multiplicativos para um período sazonal s .

Para a modelagem foram definidos alguns parâmetros na construção da topologia do modelo:

```
MdlS = arima('Constant',0,'D',1,'Seasonality', 96,'MALags',1,'SMALags',1);
```

Comando: arima		
Parâmetro	Valor	Explicação
<i>Constant</i>	0	Constante de <i>offset</i>
<i>D</i>	1	Grau da diferenciação
<i>Seasonality</i>	96	Inteiro não-negativo que indica o grau do operador de atraso de diferenciação sazonal
<i>MALags</i>	1	Vetor de inteiros positivos relacionados aos atrasos do polinômio da média móvel
<i>SMALags</i>	1	Vetor de inteiros positivos relacionados aos atrasos da do polinômio da média móvel sazonal

A escolha adequada dos parâmetros do modelo ARIMA produzirá resíduos com distribuição de média igual a 0 e nenhuma correlação de tempo. Em aplicações práticas, a ordem do modelo é selecionada de acordo com Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). A ideia é escolher o modelo ARIMA que minimize tais critérios (CARVALHO et al., 2016).

Os parâmetros para estimação são determinados através do Método de Máxima Verossimilhança a partir do modelo *MdlS* e dos dados históricos do alimentador:

```
EstMdlS = estimate(MdlS, M_BSR);
```

Comando: estimate		
Parâmetro	Valor	Explicação
<i>MdlS</i>	-	Modelo de estimação criado
<i>M_BSR</i>	1	Dados históricos do alimentador

Por fim, o modelo com os parâmetros de estimação é gerado e será utilizado com a função de *forecast*:

```
[yS, yMSES] = forecast (EstMdlS, window, 'Y0', iaS);
```

Comando: <i>forecast</i>		
Parâmetro	Valor	Explicação
<i>EstMdlS</i>	-	Modelo de estimação criado no comando anterior
<i>window</i>	1	Janela de estimação futura
<i>'Y0'</i>	-	Indica que serão fornecidas as amostras iniciais
<i>iaS</i>	96	96 medições anteriores

Os argumentos de *saída* são o valor estimado (*yS*) e o erro médio quadrático (*yMSES*).

3.6.2 Avaliação das Incertezas

A função utilizada anteriormente retorna o erro médio quadrático para cada ponto estimado. Então como só foi feita a previsão de um valor, o retorno de “*forecast*” será o valor estimado e o erro médio quadrático, em que a raiz quadrada do *yMSES* é o desvio padrão da previsão.

3.7 FATOR DE OCORRÊNCIA E DIAGNÓSTICO DO SISTEMA DE MEDIÇÃO

Para cada alimentador é gerado um relatório pelo módulo proposto para o Sistema Supervisório indicando: o número de *outliers* e seus horários, além dos valores corrigidos; no período definido pelo usuário.

Além da detecção e correção dos *outliers*, o algoritmo gera, para cada alimentador, um relatório com o número de falhas e seus horários, além de um diagnóstico do sistema de medição do alimentador obtido via consulta direta à Tabela 2 em que f_o é o fator de ocorrências dado por:

$$f_o = \frac{n_o}{N} \quad (30)$$

com n_o sendo o número de ocorrência de *outliers* e N sendo o número total de amostras no intervalo de tempo estudado.

A estratégia de diagnóstico é considerada estática e individualizada, pois observa-se apenas a realidade dos alimentadores da subestação tratada. Isto é, pode ser que o sistema de medição de um alimentador receba o diagnóstico de “Aceitável”, mas comparando com outras subestações da região seu resultado pode ser considerado negativo, caso as outras subestações possuam uma quantidade de *outliers* significativamente menor. Então, o sistema de medição pode receber uma avaliação local como “Aceitável”, mas globalmente ser considerado negativo.

TABELA 2. FATOR DE OCORRÊNCIA DE *OUTLIERS* E DIAGNÓSTICO DO SISTEMA DE MEDIÇÃO DO ALIMENTADOR.

Fator de ocorrência (f_o)	Diagnóstico
$f_o = 0$	Ótimo
$f_o \leq 5\%$	Aceitável
$5 < f_o \leq 10\%$	Tolerável
$10 < f_o \leq 30\%$	Inaceitável
$f_o > 30\%$	Crítico

4. ESTUDO DE CASO: SUBESTAÇÃO REAL

Neste Capítulo são apresentados resultados e as discussões para metodologia de detecção e correção de *outliers* considerando medições de uma subestação real de 69kV/13,8kV localizada na cidade de Cuité no estado da Paraíba, Brasil. O alimentador estudado neste capítulo será o que atende a cidade de Barra de Santa Rosa e é representado neste trabalho como o alimentador BSR.

Os resultados são apresentados em três cenários de testes: no período de 1 dia, 1 ano e 6 anos. A avaliação em um dia permite conhecer melhor o comportamento do perfil da carga e dos algoritmos de detecção e de correção. O algoritmo será testado para as três abordagens desenvolvidas no período de um dia para o conhecimento da curva de carga e o comportamento das técnicas. Depois serão testadas as técnicas no período de 1 ano. E, por fim, as abordagens de detecção e correção são expostas a um período de 6 anos com a finalidade de avaliar o desempenho estatístico e a robustez dos métodos propostos.

A temperatura no local da subestação estudada não apresenta grandes variações ao longo do ano, logo o perfil de carga não se altera significativamente durante as estações. Desta forma, a análise torna-se mais simples do que se houvesse uma grande divergência entre as temperaturas de verão e inverno.

4.1 INJEÇÃO ALEATÓRIA DE *OUTLIERS*

Para avaliar a exatidão das técnicas de correção dos *outliers*, foram injetados sobre os dados históricos de boa qualidade valores iguais a zero e valores de magnitude superior ao padrão da curva, simulando os *outliers* do tipo vale e do tipo pico, respectivamente. Então os valores originais são comparados aos valores dados pelas técnicas de correção através do cálculo do erro relativo (E_{rel}) dado por:

$$E_{rel} = 100 \left| \frac{V - C}{V} \right| \quad (31)$$

em que V é o valor verdadeiro e C é o valor calculado pelas técnicas de correção.

A injeção de *outliers* é feita de forma aleatória utilizando a função ***randi*** do MATLAB® que adota os seguintes parâmetros de entrada:

<i>Comando: randi</i>	
Parâmetro	Explicação
Valor inicial	Limite inferior da faixa de valores que podem ser gerados.
Valor final	Limite superior da faixa de valores que podem ser gerados.
Número de linhas	Número de linhas do vetor de saída
Número de colunas	Número de colunas do vetor de saída

No primeiro momento é definido aleatoriamente o número de outliers que são gerados representado pela variável `n_teste_BSR` e que possui os seguintes parâmetros de entrada:

```
n_teste_BSR = randi([nI nF], 1, 1);
```

Retorno: n_teste_BSR (1x1)	
Parâmetro	Valor
Valor inicial (nI)	0
Valor final (nF)	Número total de amostras
Número de linhas	1
Número de colunas	1

A saída da função ***randi*** denominada de `n_teste_BSR` é apenas um número inteiro compreendido no intervalo de 0 até o número total máximo de amostras. No caso do intervalo de 1 ano, por exemplo, o número máximo é 35040. Após definido o número de *outliers*, é calculada o vetor (`rS`) que conterà todos os índices das medições errôneas.

```
rS = randi([n_inicial nLoop], 1, n_teste_BSR);
```

Retorno: rS (1xn_teste_BSR)	
Parâmetro	Valor
Valor inicial (n_inicial)	1
Valor final (nLoop)	Número total de amostras
Número de linhas	1
Número de colunas (n_teste_BSR)	Número total de <i>outliers</i> calculado anteriormente

Portanto, o vetor de saída (`rS`) será um vetor com números aleatórios compreendidos no intervalo entre o valor de índice 1 até o último valor do número total de amostras, sendo um total de `n_teste_BSR` medições errôneas. Finalmente, é criada uma

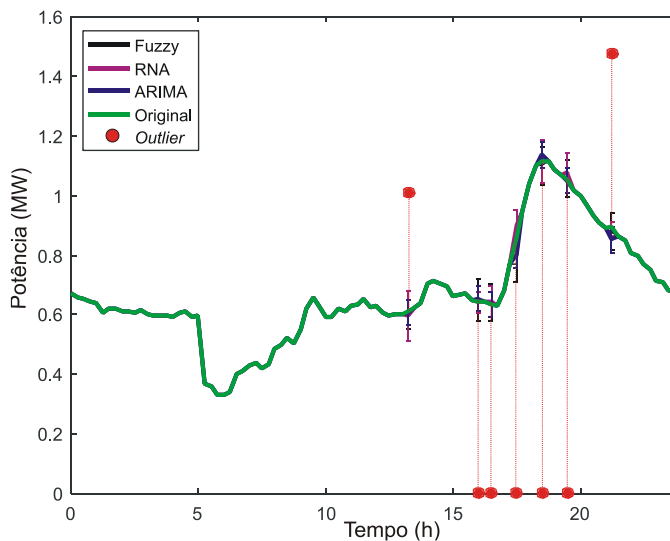
máscara (cópia) dos dados históricos originais e é feito um laço de repetição *for* para zerar todos os valores dessa máscara com índices iguais aos elementos do vetor rS. A máscara é criada para que os dados históricos sejam mantidos íntegros para comparação posterior através do erro relativo dado pela Eq. (12).

4.2 PRIMEIRO CENÁRIO: 1 DIA

No primeiro cenário de teste, é analisada a injeção de 7 *outliers* aleatórios não-consecutivos na curva do alimentador BSR, ilustrada na Figura 25. Para cada valor novo de entrada foi feita a subtração com o valor respectivo de mesmo índice do dia típico que neste exemplo foi um domingo. Portanto, a cada nova medição era verificada se o módulo desta diferença era maior ou menor que o triplo do desvio padrão médio.

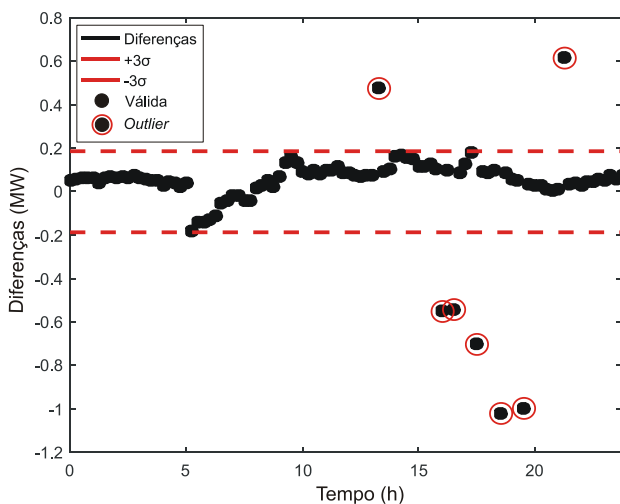
Os valores que se encontraram fora do intervalo determinado são marcados como *outliers*, como ilustrado na Figura 26, que apresenta as diferenças, a faixa de medições válidas e os 7 pontos externos que são detectados como medições errôneas.

FIGURA 25. PRIMEIRO CENÁRIO: INJEÇÃO DE 7 OUTLIERS ALEATÓRIOS NO PERÍODO DE UM DIA.



Fonte: Elaborada pelo autor.

FIGURA 26. GRÁFICO DAS DIFERENÇAS ENTRE O VALOR MEDIDO E O VALOR TÍPICO COM O TRIPLO DO DESVIO PADRÃO COMO VALOR LIMITE.



Fonte: Elaborada pelo Autor.

A Tabela 3 indica os erros relativos médios de cada uma das técnicas desenvolvidas e para esse cenário a abordagem RNA foi levemente superior. Os *outliers* ocorridos às 17h30 e às 18h30 fazem parte da região do perfil de carga em que há uma maior variação da derivada e que são zonas críticas para as técnicas de correção. A abordagem que apresenta o pior desempenho nas regiões de maior variação é técnica *Fuzzy* que responde inferiormente às grandes mudanças, como pode ser observada pela Figura 25 e pelo erro relativo igual a 10,43% às 17h30.

TABELA 3. ERROS RELATIVOS DAS TÉCNICAS DE CORREÇÃO PARA 7 *OUTLIERS* NO PERÍODO DE 1 DIA.

Horário	Índice	Tipo	Erro Relativo Fuzzy $\pm \sigma_F$ (%)	Erro Relativo RNA $\pm \sigma_R$ (%)	Erro Relativo ARIMA $\pm \sigma_A$ (%)
13h15	54	Pico	0,4676 \pm 3,2200	2,2417 \pm 4,8000	0,9801 \pm 2,1600
16h00	65	Vale	0,4647 \pm 3,4900	0,5977 \pm 1,900	1,3321 \pm 2,1600
16h30	67	Vale	0,9584 \pm 3,2100	0,9642 \pm 2,9200	0,3613 \pm 2,1600
17h30	71	Vale	10,4306 \pm 3,0700	4,8391 \pm 3,1200	6,7530 \pm 2,1600
18h30	75	Vale	1,6289 \pm 3,0700	0,2757 \pm 3,0900	1,7095 \pm 2,1600
19h30	79	Vale	0,0250 \pm 3,2100	1,8107 \pm 3,2700	0,7192 \pm 2,1600
21h15	86	Pico	1,8084 \pm 3,1900	4,2331 \pm 2,7300	4,9235 \pm 2,1600
Média Erro Relativo (%)			2,2548 \pm 3,2100	2,1375 \pm 3,1200	2,3970 \pm 2,1600

4.3 SEGUNDO CENÁRIO: 1 ANO

No segundo cenário de teste, foi feita a injeção aleatória de *outliers* no intervalo de um ano da curva do Alimentador BSR da subestação estudada. A Tabela 4 é similar ao relatório gerado pela metodologia proposta após o ciclo de um ano e contém: a Subestação e o alimentador, o código do alimentador referente a região de distribuição, número de falhas por ano, o fator de ocorrência e o diagnóstico do sistema de medição de acordo com a Tabela 2 exposta no Capítulo 4. O fator de ocorrência foi igual a 11,90%, isto é, foram 4171 *outliers* em um total de 35040 amostras do período de um ano (365×96) e é considerado “Inaceitável” de acordo com o sistema de diagnóstico proposto. Além desta avaliação, é gerado um relatório para cada alimentador com a hora e dia em que ocorreu cada *outlier*.

TABELA 4. RESUMO DO RELATÓRIO GERADO PELO ALGORITMO PROPOSTO SOBRE O SISTEMA DE MEDIÇÃO DO ALIMENTADOR BSR.

Subestação/ Alimentador	Região	Número de falhas	Fator de ocorrência (%)	Diagnóstico
CTE-21L3	BSR	4171	11,90	Inaceitável

Vale ressaltar que para cada *outlier* é acionada uma *flag* para o Sistema SCADA que fica habilitada enquanto não houver dados válidos. Se o erro persistir, o Sistema Supervisório permanecerá sinalizando indicando que os dados estão sendo gerados pelo algoritmo de correção.

As curvas corrigidas do segundo cenário de teste são apresentadas nas Figuras 27-29. Sobre os valores históricos foram injetados *outliers* e estes conjuntos de dados foram submetidos às técnicas de correção. Portanto, os dados simulados estão em vermelho marcados pela presença de *outliers* e os dados corrigidos são indicados em cor diferente.

FIGURA 27. SEGUNDO CENÁRIO DE TESTE: PERÍODO DE UM ANO COM A TÉCNICA DE CORREÇÃO FUZZY.

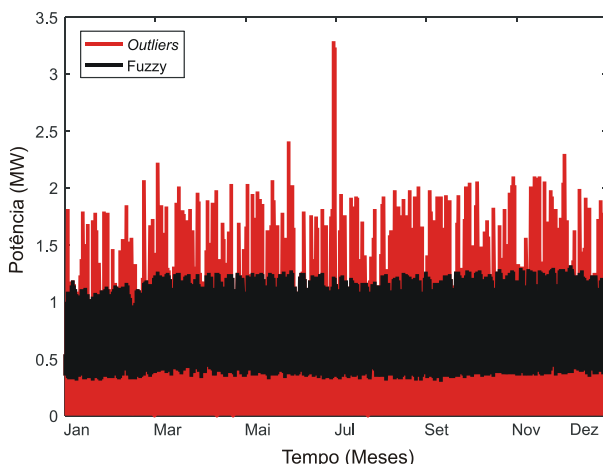
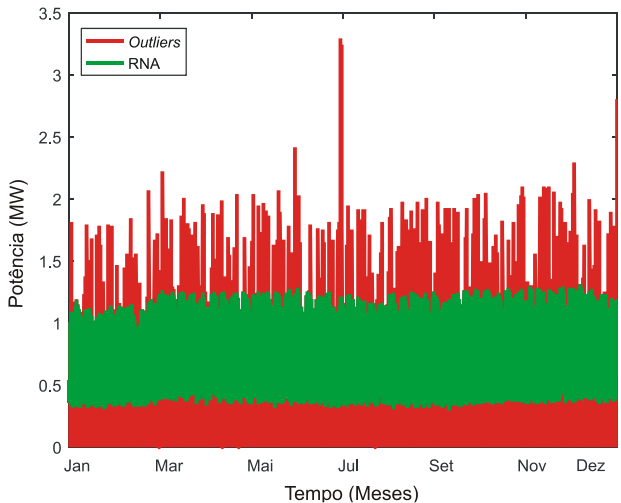
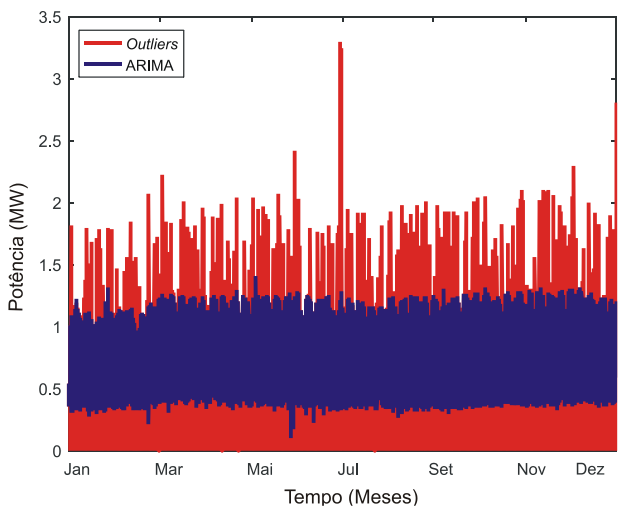


FIGURA 28. SEGUNDO CENÁRIO DE TESTE: PERÍODO DE UM ANO COM A TÉCNICA DE CORREÇÃO RNA.



Fonte: Elaboradas pelo autor.

FIGURA 29. SEGUNDO CENÁRIO DE TESTE: PERÍODO DE UM ANO COM A TÉCNICA DE CORREÇÃO ARIMA.



Fonte: Elaborada pelo autor.

Apenas analisando os gráficos das curvas simuladas e corrigidas não é possível se ter um bom entendimento da eficácia da metodologia proposta. Portanto, são apresentados os gráficos dos erros relativos percentuais das três técnicas de correção propostas nas Figuras 30-32.

Neste cenário, a técnica de correção que apresentou o menor erro relativo médio foi novamente a abordagem RNA com um erro relativo médio de 3,50%. E a técnica ARIMA, utilizada como comparativo, foi a que apresentou o pior desempenho com um erro relativo médio de 4,15%. A abordagem *Fuzzy* obteve resultado intermediário dentre as três técnicas com um erro relativo médio de 3,99%.

TABELA 5. ERROS RELATIVOS DAS TÉCNICAS DE CORREÇÃO DE *OUTLIERS* PARA O PERÍODO DE UM ANO.

Alimentador	Erro Relativo Médio Fuzzy $\pm \sigma_F$ (%)	Erro Relativo Médio RNA $\pm \sigma_R$ (%)	Erro Relativo Médio ARIMA $\pm \sigma_A$ (%)
21L3-BSR	3,9906 \pm 3,6900	3,5012 \pm 3,9700	4,1487 \pm 3,1000

Os erros relativos máximos são expostos na Tabela 6 em que o maior erro ocorreu na técnica RNA com 53,67%.

Na abordagem *fuzzy*, o erro aumenta quando há vários *outliers* em sequência. Já na modelagem ARIMA tende a repetir o comportamento do dia anterior e caso esse dia seja atípico (feriado ou evento que modifique significativamente a demanda da cidade), o erro será maior para essas atipicidades, isto é,

quando as 96 medições anteriores possuem características diferentes da amostra faltante.

Em alguns casos, a técnica de detecção pode avaliar medições corretas como *outliers*, pois a diferença entre o valor medido e o valor típico nestas situações é maior que o valor limite estabelecido, podendo aumentar o número de amostras consideradas como *outliers*.

FIGURA 30. GRÁFICO DOS ERROS RELATIVOS (%) E MÉDIA DOS ERROS RELATIVOS (%) DA TÉCNICA DE CORREÇÃO FUZZY.

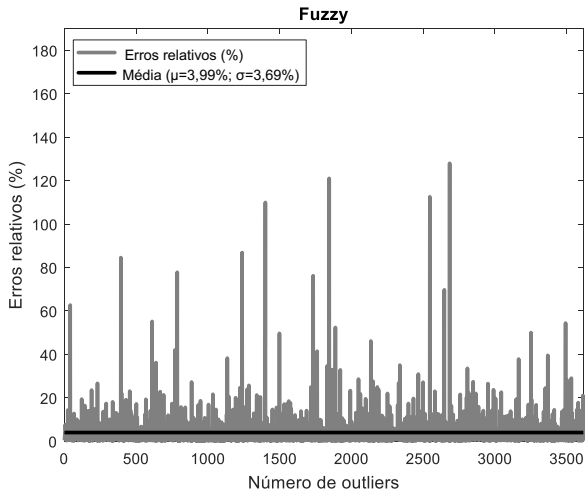


FIGURA 31. GRÁFICO DOS ERROS RELATIVOS (%) E MÉDIA DOS ERROS RELATIVOS (%) DA TÉCNICA DE CORREÇÃO RNA.

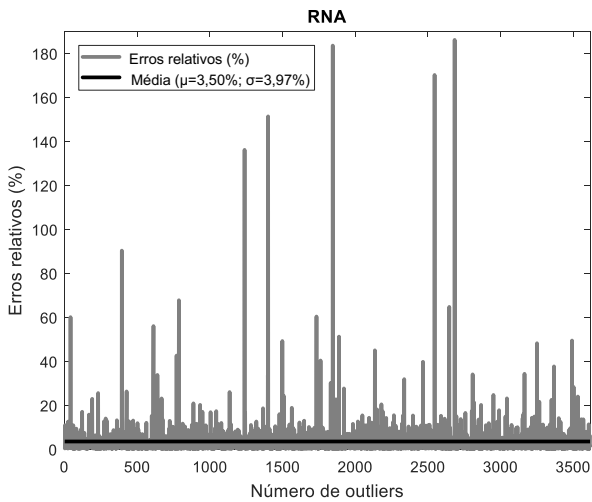
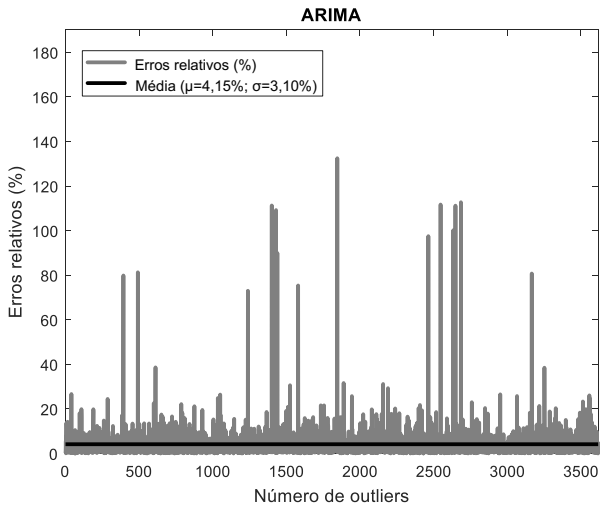


FIGURA 32. GRÁFICO DOS ERROS RELATIVOS (%) E MÉDIA DOS ERROS RELATIVOS (%) DA TÉCNICA DE CORREÇÃO ARIMA NO PERÍODO DE UM ANO.



Fonte: Elaboradas pelo autor.

TABELA 6. ERROS RELATIVOS MÁXIMOS DAS TÉCNICAS DE CORREÇÃO DE OUTLIERS PARA O PERÍODO DE UM ANO.

Alimentador	Erro Relativo Máximo Fuzzy $\pm \sigma_F$ (%)	Erro Relativo Máximo RNA $\pm \sigma_R$ (%)	Erro Relativo Máximo ARIMA $\pm \sigma_A$ (%)
21L3-BSR	127,97 \pm 3.1700	186,14 \pm 3.8500	132,45 \pm 2.3000

4.4 Terceiro Cenário: 6 anos

O terceiro cenário de testes foi realizado para um período de 6 anos (2008-2013) em que as técnicas são analisadas quanto à robustez para responderem a uma possível tendência associada ao longo período de tempo.

A Tabela 7 é o relatório gerado pelo programa que indica o número de *outliers* presentes no período de 6 anos. Neste teste,

os dados estavam deturpados com uma quantidade considerável de dados errôneos apresentando 11,56% das amostras como valores errôneos. E em um caso real *online*, deveria ser considerada a correção ou troca do equipamento de medição, pois o diagnóstico associado corresponde a uma situação “Inaceitável”.

As curvas corrigidas para o período de 6 anos são apresentadas nas Figuras 33-35 e pode-se perceber que não há uma tendência muito evidente, o que favorece as técnicas de correção desenvolvidas.

TABELA 7. RESUMO DO RELATÓRIO GERADO PELO ALGORITMO PROPOSTO SOBRE O SISTEMA DE MEDIÇÃO DO ALIMENTADOR BSR PARA O PERÍODO DE SEIS ANOS.

Subestação/ Alimentador	Região	Número de falhas	Fator de ocorrência (%)	Diagnóstico
CTE-21L3	BSR	24323	11,56	Inaceitável

As técnicas de correção utilizadas mantiveram o erro relativo médio próximo aos valores encontrados no período de 1 dia e de 1 ano. As regras da abordagem *fuzzy* foram construídas para o perfil de carga desta subestação então desde que ele não se altere, a técnica permanecerá respondendo positivamente. A abordagem RNA por utilizar uma janela de amostras anteriores reduzida favorece a previsão a curtíssimo prazo, assim como a abordagem ARIMA que tem como entrada do modelo autorregressivo 96 medições anteriores.

FIGURA 33. TERCEIRO CENÁRIO DE TESTE: PERÍODO DE SEIS ANOS COM A TÉCNICA DE CORREÇÃO FUZZY.

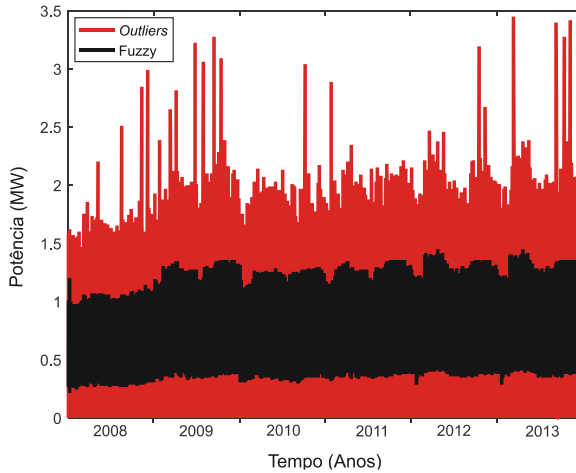
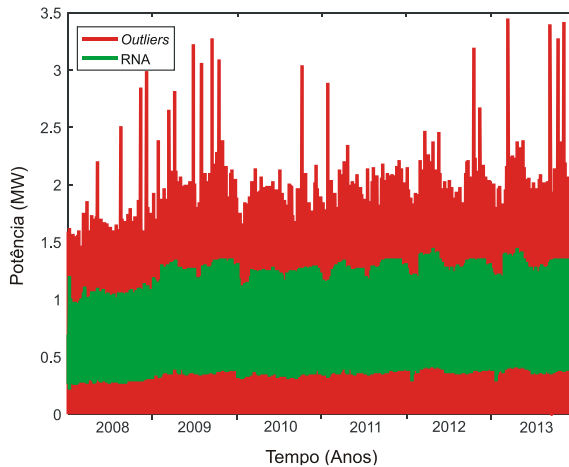
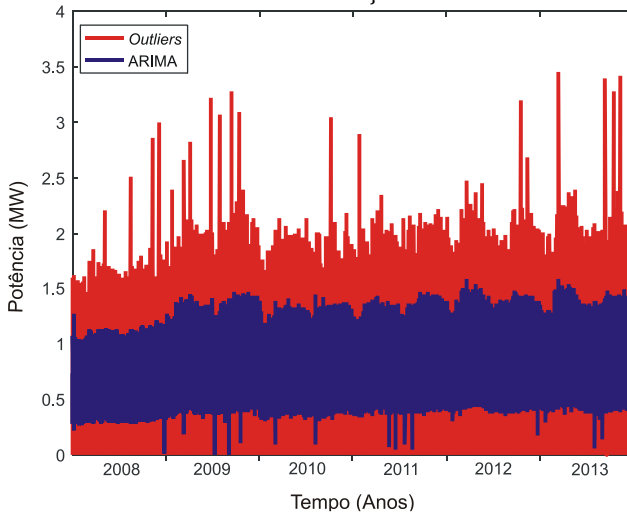


FIGURA 34. TERCEIRO CENÁRIO DE TESTE: PERÍODO DE SEIS ANOS COM A TÉCNICA DE CORREÇÃO RNA.



Fonte: Elaboradas pelo autor.

FIGURA 35. TERCEIRO CENÁRIO DE TESTE: PERÍODO DE SEIS ANOS COM A TÉCNICA DE CORREÇÃO ARIMA.



Fonte: Elaborada pelo autor.

A Tabela 8 indica os erros relativos médios das três técnicas implementadas no período de 6 anos associados às médias das incertezas calculadas. A abordagem RNA apresentou desempenho levemente superior a técnica *Fuzzy* e ARIMA. O erro relativo médio da RNA foi 4,46% enquanto que o pior desempenho foi a abordagem ARIMA novamente com 4,71%, bem próximo da abordagem *Fuzzy* que apresentou 4,68%.

Os erros relativos das três técnicas de correção são apresentados nas Figuras 36-38 para um melhor entendimento sobre a metodologia aplicada. O eixo das abscissas indica o número total de amostras errôneas que para o terceiro cenário de testes foi 24323. Portanto, esse é o número de erros relativos calculados e são encontrados em um grande intervalo de valores,

contudo no geral a média é consideravelmente baixa (menor que 5%).

TABELA 8. ERROS RELATIVOS DAS TÉCNICAS DE CORREÇÃO DE *OUTLIERS* PARA O PERÍODO DE SEIS ANOS.

Alimentador	Erro Relativo Médio <i>Fuzzy</i> $\pm \sigma_F$ (%)	Erro Relativo Médio RNA $\pm \sigma_R$ (%)	Erro Relativo Médio ARIMA $\pm \sigma_A$ (%)
21L3-BSR	4.6845 \pm 3.6900	4.4587 \pm 3.9300	4.7151 \pm 3.0300

FIGURA 36. GRÁFICO DOS ERROS RELATIVOS (%) E MÉDIA DOS ERROS RELATIVOS (%) DA TÉCNICA DE CORREÇÃO *FUZZY* NO PERÍODO DE SEIS ANOS.

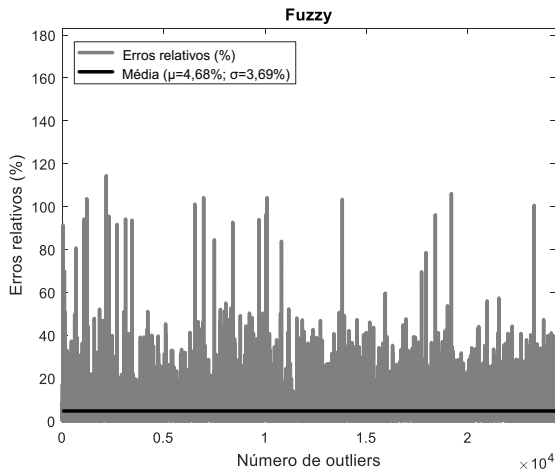


FIGURA 37. GRÁFICO DOS ERROS RELATIVOS (%) E MÉDIA DOS ERROS RELATIVOS (%) DA TÉCNICA DE CORREÇÃO RNA NO PERÍODO DE SEIS ANOS.

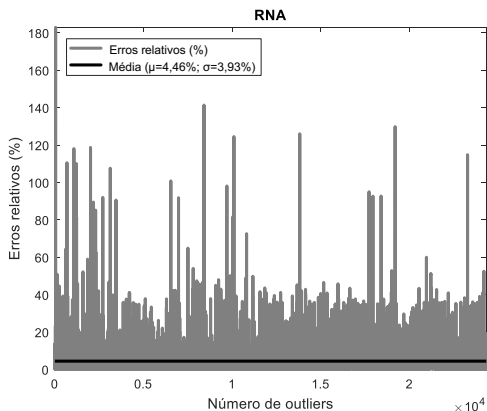
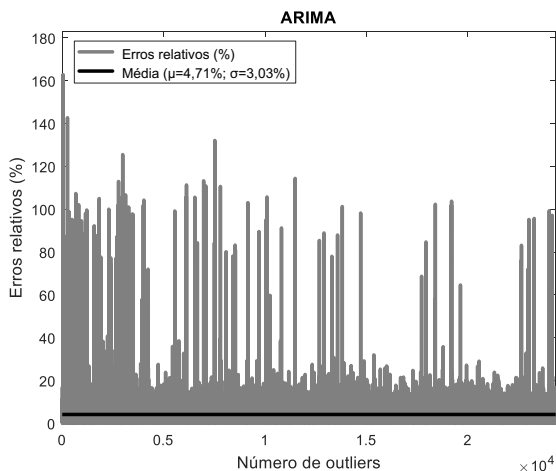


FIGURA 38. GRÁFICO DOS ERROS RELATIVOS (%) E MÉDIA DOS ERROS RELATIVOS (%) DA TÉCNICA DE CORREÇÃO ARIMA NO PERÍODO DE SEIS ANOS.



Fonte: Elaboradas pelo autor.

Os erros relativos máximos, expostos na Tabela 9, ocorreram em pontos diferentes do cenário anterior e para este teste o que apresentou menor valor foi o erro relativo da técnica

Fuzzy com 114,42% enquanto que o maior das três técnicas foi o da abordagem RNA apresentando erro relativo de 182,95%.

Os altos erros são causados em dias atípicos e em pontos que as técnicas deveriam indicar uma derivada positiva, mas indicam derivadas negativas aumentando consideravelmente o erro.

TABELA 9. ERROS RELATIVOS MÁXIMOS DAS TÉCNICAS DE CORREÇÃO DE *OUTLIERS* PARA O PERÍODO DE SEIS ANOS.

Alimentador	Erro Relativo Máximo <i>Fuzzy</i> ± σ_F (%)	Erro Relativo Máximo RNA ± σ_R (%)	Erro Relativo Máximo ARIMA ± σ_A (%)
21L3-BSR	114,4224 ± 3.4200	182,9515 ± 3.3300	162,8603 ± 3.1200

As técnicas RNA e *Fuzzy* são enquadradas de acordo com a classificação de Guirelli (2006) como Métodos Computacionais e têm por característica uma menor dependência de modelagem matemática e, após a fase de treinamento para RNA e de construção das regras para a lógica *fuzzy*, a utilização é eficiente por não exigir muito esforço computacional seja do medidor inteligente ou do *software* de processamento dos dados adquiridos.

Nos três algoritmos desenvolvidos, após 1h30 de ausência de dados, o programa passa a salvar os valores do mesmo horário e mesmo dia da semana anterior. Entretanto a *flag* de *outlier* se manteve ainda habilitada indicando um sinal de alerta

informando que houve um erro e que os dados, apesar de próximos, não são mais os dados verdadeiros. A concessionária pode usar esta informação de onde foi o problema como um conhecimento para o Centro de Controle de Operações e ativar equipes técnicas para encaminhamento ao local e corrigi-lo.

O *ranking* do diagnóstico foi definido analisando exclusivamente esta subestação e nos dois períodos testados o Diagnóstico foi “Inaceitável” por apresentar valores superiores a 10% e inferiores a 20%, entretanto os valores de corte foram arbitrados de acordo com a presença e o efeito dos *outliers* na série temporal. Um trabalho mais completo seria a realização de uma análise de todas as subestações da região, avaliando a presença de *outliers* e criando *clusters* com os piores e melhores casos. Desta forma, a concessionária de energia poderia direcionar os recursos para as subestações mais comprometidas, efetuando a correção dos sistemas de medição dos piores casos encontrados.

POSFÁCIO

É fundamental para o bom serviço das distribuidoras de energia ter o conhecimento preciso de seu sistema elétrico. Portanto, a falta desses dados pode prejudicar planejamentos futuros. Então o desenvolvimento do módulo inteligente para a detecção e correção de *outliers* do Sistema Supervisório de uma subestação implica em vantagens associadas à qualidade das curvas de demanda e aumento da robustez do sistema SCADA.

A proposta de desenvolvimento do módulo complementar utilizou três técnicas de correção de *outliers* (*Fuzzy*, RNA e ARIMA), sendo a abordagem baseada em RNA a mais eficiente por conseguir acompanhar as variações da curva de demanda, principalmente, no horário de ponta. A técnica de correção *Fuzzy* perde desempenho nas inclinações e declínios súbitos, isto é, quando a derivada cresce ou decresce muito rapidamente. E esse cenário pode se agravar se houver muitos *outliers* em série. A abordagem ARIMA foi utilizada como comparativo por ser uma técnica clássica na literatura e apresentou resultados próximos às técnicas desenvolvidas. Nos estudos de casos apresentados, foi observado que as técnicas desenvolvidas apresentaram um erro relativo médio inferior a 5% para um período de análise de 6 anos. E a principal vantagem

encontrada de trabalhar com técnicas de inteligência artificial é a dispensa de uma análise estatística multivariada mais complexa.

Adicionalmente, os autores propõem um fator de ocorrência de *outliers*, que indica a frequência de ocorrência de outlier em um intervalo de tempo. A partir desta avaliação foi possível determinar o diagnóstico do sistema de medição, isto é, em função à frequência de ocorrências, concluindo-se que alguns dos pontos de medição operavam em estado crítico (fator de ocorrência $f_o > 30\%$). Portanto, esta informação fornecida pelo Sistema Supervisório pode ser usada como tomada de decisões para os especialistas para o aprimoramento dos sistemas de automação e controle.

No geral, as curvas de demanda da subestação estudada apresentam *outliers* de forma pontual desta maneira o algoritmo desenvolvido garante uma boa performance e confiabilidade dos dados. Para um grande volume de dados ausentes devem ser usadas outras técnicas de correção baseadas no histórico dos dados, nas condições climáticas, sociais e econômicas da região alimentada pela subestação tratada.

Em trabalhos futuros, com a finalidade de avaliar ainda mais a confiabilidade das metodologias propostas para a correção de *outlier*, serão desenvolvidas novas análises de desempenho contemplando mais parâmetros estatísticos; além de informações sobre a região alimentada pela subestação.

PUBLICAÇÕES DOS AUTORES NA ÁREA

Artigos completos publicados em periódicos

- ANDRADE, PEDRO H. M.; MAURICIO VILLANUEVA, JUAN MOISES ; MACEDO BRAZ, HELON DAVID. ***An Outliers Processing Module based on Artificial Intelligence for Substations Metering System.*** IEEE TRANSACTIONS ON POWER SYSTEMS, v. 1, p. 1-9, 2020.
- ANDRADE, P. H. M.; VILLANUEVA, JUAN M. M. ; BRAZ, HELON D. M. ***Outliers Treatment in Power Curves Using Hybrid Artificial Intelligence Technique.*** INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING, INFORMATION & CONTROL, v. 16, p. 513-525, 2020.

Artigo completos publicados em congressos

- Pedro H. M. de Andrade; VILLANUEVA, JUAN M. M. ***“Detecção e Correção de Outliers em Curvas de Potência Utilizando Técnica Híbrida de Inteligência Artificial”.*** In: 14º Simpósio Brasileiro de Automação Inteligente, 2019, Ouro Preto.
- DE ANDRADE, PEDRO H. M.; VILLANUEVA, JUAN M. M. ; MACEDO, HELON D. . ***“Complementary module to smart***

- meters based on outliers correction using artificial intelligence*". In: 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2018, Houston. 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2018.
- DE ANDRADE, PEDRO H. M.; VILLANUEVA, JUAN M. M. ; BRAZ, HELON D. M. "***Algoritmos de Correção de Outliers para Curvas de Potência utilizando inteligência artificial***". In: 2018 Simpósio Brasileiro de Sistemas Elétricos (SBSE), 2018, Niteroi. 2018 Simposio Brasileiro de Sistemas Elétricos (SBSE), 2018.
 - NETO, J. T. C. ; DE ANDRADE, P. H. M. ; VILANUEVA, J. M. ; SANTOS, F. A. O. "***Big Data Analytics of Smart Grids using Artificial Intelligence for the Outliers Correction at Demand Measurements***". In: 2018 3rd International Symposium on Instrumentation Systems, Circuits and Transducers (INSCIT), 2018, Bento Goncalves. 2018 3rd International Symposium on Instrumentation Systems, Circuits and Transducers (INSCIT), 2018.
 - Pedro H. M. de Andrade; VILLANUEVA, JUAN M. M.; BRAZ, H. D. M. "***Detecção e Correção Automática de Outliers para Sistemas Supervisórios em Subestações***". In: XXII Congresso Brasileiro de Automática, 2018, João Pessoa, 2018.

- Jéssica Madruga de Miranda Henriques; Pedro Henrique Meira de Andrade; Helon David de Macêdo Braz. “***Preditores aplicados na inicialização inteligente do método da soma de potências em série temporal***”. In: XXII Congresso Brasileiro de Automática, 2018, João Pessoa, 2018.

REFERÊNCIAS

ALAHAKOON, D.; YU, X. "**Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey**," in *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 425-436, Feb. 2016.

ALVES, W.; MARTINS, D.; BEZERRA, U.; KLAUTAU, A. "**A Hybrid Approach for Big Data Outlier Detection from Electric Power SCADA System**," in *IEEE Latin America Transactions*, vol. 15, no. 1, pp. 57-64, Jan. 2017.

BERA, S.; MISRA, S.; RODRIGUES, J. J. P. C. "**Cloud Computing Applications for Smart Grid: A Survey**," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1477-1494, Maio 2015.

BOX, G.E.P., JENKINS, G.M. e REINSEL, G.C. (1994) **Time Series Analysis**; Forecasting and Control. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey.

CACIOTTA, M.; GIARNETTI, S.; LECCESE, F.; TRINCA, D. "**Development of an USB data acquisition system for Power Quality and smart metering applications**," *2012 11th International Conference on Environment and Electrical Engineering*, Venice, 2012, pp. 835-839.

CARVALHO, L. M.; TEIXEIRA, J.; MATOS, M. "**Modeling wind power uncertainty in the long-term operational reserve adequacy assessment: A comparative analysis between the Naïve and the**

ARIMA forecasting models," 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing, 2016, pp. 1-6.

CHAKHCHOUKH, Y.; LIU, S.; SUGIYAMA, M.; ISHII, H. "**Statistical outlier detection for diagnosis of cyber attacks in power state estimation,**" 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, 2016, pp. 1-5.

CHARYTONIUK, W.; CHEN, M. "**Very Short-Term Load Forecasting Using Artificial Neural Networks,**" IEEE Trans. Power Systems, vol. 15, pp. 263-268, Feb. 2000.

da SILVA, I.N., Spatti, D.H. & Flauzino, R.A. (2010) "**Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas**", Artliber Editora Ltda., ISBN: 9788588098534.

D'ERRICO, G. E.; MURRU, N. "**Fuzzy treatment of candidate outliers in measurements,**" Advances in Fuzzy Systems, vol. 44, no. 6, pp. 1292-1296, 1999.)

DIGGLE PJ. *Time series: a biostatistical introduction.* Oxford:Oxford University Press; 1992.

EL-HALABI, M.; FARRAJ, A. K.; LY, H. D.; KUNDUR, D. "**A distortion theoretic perspective for redundant metering security in a smart grid,**" in *Proc. 25th IEEE Canadian Conf. Elect. Comput. Eng. (CCECE)*, Montreal, QC, Canada, 2012, pp. 1-5.

FANG, X.; MISRA, S.; XUE, G.; YANG, D. "**Smart grid-the new and improved power grid: A survey**", IEEE Commun. Surv. Tuts., vol. 14, no. 4, pp. 944-980, Dezembro, 2012.

FARROKHIFARD, M.; HATAMI, M.; PARNIANI, M. "**Novel approaches for online modal estimation of power systems using PMUs data contaminated with outliers.**" Electric Power Systems Research, 2015.

GAL, Y., GHAHRAMANI, Z., "**Dropout as a Bayesian approximation: representing model uncertainty in deep learning**". In *Proceedings of the 33rd International Conference on International Conference on Machine Learning – Vol. 48*, 2016.

GANDHI, K.; BANSAL, H. O. "**Smart Metering in electric power distribution system,**" *2013 International Conference on Control, Automation, Robotics and Embedded Systems (CARE)*, Jabalpur, 2013, pp. 1-6.

GHOSH, D.; GHOSE, T.; MOHANTA, D. K. "**Communication Feasibility Analysis for Smart Grid With Phasor Measurement Units,**" in *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1486-1496, Aug. 2013.

GUIRELLI, C. R. **Previsão da carga de curto prazo de áreas elétricas através de técnicas de inteligência artificial.** 2006. Biblioteca Digital de Teses e Dissertações da USP, 2006.

GUO, Z.; LI, W.; LAU, A.; INGA-ROJAS, T.; WANG, K. "**Detecting X-Outliers in Load Curve Data in Power Systems**," in IEEE Transactions on Power Systems, vol. 27, no. 2, pp. 875-884, May 2012.

HAYKIN, Simon. **Redes Neurais: Princípios e prática**. Porto Alegre, RS. Bookman, 2001.

HE, H.; CHEN, T. L., R.; XIAO, Y. e YANG, J. "**High frequency short-term demand forecasting model for distribution power grid based on ARIMA**," *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, Zhangjiajie, 2012, pp. 293-297.

JAKKULA, V.; COOK, D. "**Outlier Detection in Smart Environment Structured Power Datasets**," 2010 Sixth International Conference on Intelligent Environments, Kuala Lumpur, 2010, pp. 29-33.

JAPEC, L.; KREUTER, F.; BERG, M.; BIEMER, P.; P. DECKER, P.; LAMPE, C.; LANE, J.; O'NEIL, C.; USHER, A. "**Big Data in Survey Research: AAPOR Task Force Report**", *Public Opinion Quarterly*, vol. 79. Jan, 2015, pp. 839-880.

JIA, N.; WANG, J.S.; LI, N. "**Application of data mining in inteligente power consumption**", Intern. Conf. on Automatic Control and Artificial Intelligence (ACAI 2012), pp. 538-541, 2012.

KATIC, N. A.: "**Profitability of Smart Grid Solutions Applied in Power Grid**", THERMAL SCIENCE: Year 2016, Vol. 20, Suppl. 2, pp. S371-S382.

KHAN, N., YAQOUB, I., HASHEM, I. A. T., INAYAT, Z., MAHMOUD ALI, W. K., ALAM, M., GANI, A. (2014). **Big Data: Survey, Technologies, Opportunities, and Challenges**. The Scientific World Journal, 2014.

LEE, S.K.; BAE, M.; KIM, H. “**Future of IoT Networks: A Survey**”. *Appl. Sci.* 2017, 7, 1072.

Lee, C.C (1990). **Fuzzy Logic in Control Systems: Fuzzy Logic Controller, part I and II**. IEEE Trans. on Systems, Man and Cybernetics, vol. 20, pp 404-435

LI, Y. *et al.*, “**Service life estimation of smart electricity meters using operation data**,” *2017 Chinese Automation Congress (CAC)*, Jinan, China, 2017, pp. 4755-4759.

LIU, Y.; REITER, M. K.; NING, P. “**False data injection attacks against state estimation in electric power grids**,” in Proceedings of the 16th ACM conference on Computer and communications security, November 2009.].

MA, L.; GU, X.; WANG, B. (2017). Correction of Outliers in Temperature Time Series Based on Sliding Window Prediction in Meteorological Sensor Network.

MEDEIROS, R.A.O. “**Previsão de Demanda no Médio Prazo Utilizando Redes Neurais Artificiais em Sistemas de Distribuição de Energia**”, Dissertação de Mestrado, Dep. Eng. Elétrica, UFPB, 2016.

MELO, D. C. R.; CASTRO, A. R. G. **“Uma nova abordagem para Detecção e Correção de Outliers em Séries Temporais: Aplicação em Consumo de Energia Elétrica”**. Simpósio Brasileiro de Sistemas Elétricos SBSE, 2014)

MORELLO, R.; CAPUA, C. De; FULCO, G.; MUKHOPADHYAY, S. C. **“A Smart Power Meter to Monitor Energy Flow in Smart Grids: The Role of Advanced Sensing and IoT in the Electric Grid of the Future,”** in *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7828-7837, Dec.1, 1 2017.

MORETIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2ª Edição, 2004. Brasil.

MORETIN, P.A.; TOLOI, C.M.C. **Previsão de séries temporais**. 2ª. ed. São Paulo: Atual Editora; 1985.

NASCIMENTO, R. M.; OENING, A. P.; MARCILIO, D. C.; AOKI, A. R.; ROCHA JR., E. P.; SCHIOCHET, J. M.. **“Algoritmo de Detecção e Correção de Outliers para previsão de carga”**. Simpósio Brasileiro de Sistemas Elétricos SBSE, 2014)

NEAGU, B. C.; GRIGORAŞ, G.; SCARLATAACHE, F. **“Outliers discovery from Smart Meters data using a statistical based data mining approach,”** 2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, 2017, pp. 555-558.)

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), 2010, pp. 140-143.)

RAO, R.; AKELLA, S.; GULEY, G. "**Power line carrier (PLC) signal analysis of smart meters for outlier detection,**" 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, 2011, pp. 291-296.

SAMAD, T.; ANNASWAMY, M.A. "**Controls for Smart Grids: Architectures and Applications**". Proceedings of the IEEE. PP. 1-18, 2017.

SARNO, C. D.; GAROFALO, A.; CERULLO, G.; MURRU, N.; GIANDOMENICO, F. D.; CHIARADONNA, S.; "**Power Grid Outlier Treatment through Kalman Filter,**" 2014 IEEE International Symposium on Software Reliability Engineering Workshops, Naples, 2014, pp. 407-412.)

SHAW, I. S. e SIMÕES, M. G. **Controle e Modelagem Fuzzy** – E. Edgard Blúcher Ltda, 2004 - 2ª edição ISBN 85-212-0248-2.

SHOPOV, M. P. "**IoT gateway for smart metering in electrical power systems - software architecture,**" *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, 2017, pp. 974-978.

TEODORO, ARTHUR & GOMES, OTÁVIO. (2018). "**Desenvolvimento de Rede Neural Artificial para Distribuição de Chaves em Algoritmos**

de Criptografia Simétrica Utilizando Hardware Reconfigurável.
10.13140/RG.2.2.14506.26561.

VALVERDE, G.; TERZIJA, V. **“Unscented kalman filter for power system dynamic state estimation,”** Generation, Transmission and Distribution, IET, vol. 5, no. 1, pp. 29–37, 2011.)

VARODAYAN, D. P.; GAO, G. X. **“Redundant metering for integrity with information-theoretic confidentiality,”** in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Gaithersburg, MD, USA, Oct. 2010, pp. 345–349.

XENIAS D., AXON C. J., WHITMARSH L., CONNOR, P. M., BALTA-OZKAN, N., SPENCE, A., **“UK smart grid development: An expert assessment of the benefits, pitfalls and functions”**, In *Renewable Energy*, Volume 81, 2015, Pages 89-102.)

YANG, Q.; YANG, J.; YU, W.; ZHANG, N. **“On a hierarchical false data injection attack on power system state estimation,”** in *Global Telecommunications Conference (GLOBECOM 2011)*, 2011, pp. 1–5.)

YE, X.; ZONGXIANG, I.; QIAO, Y.; MIN, Y.; O'MALLEY, M. (2016). **“Identification and Correction of Outliers in Wind Farm Time Series Power Data”**. *IEEE Transactions on Power Systems*. 31, pp. 1-9)

YESILBUDAK, M. **“Partitional clustering-based outlier detection for power curve optimization of wind turbines,”** 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, 2016, pp. 1080-1084.

YU, L.; WANG, H.; CHE, J.; LU, J.; ZHENG, X. "**Outliers screening for photovoltaic electric power based on the least square method,**" 2016 Chinese Control and Decision Conference (CCDC), Yinchuan, 2016, pp. 2799-2804.)

ZANINI, A. **Redes neurais e regressão dinâmica: um modelo híbrido para previsão de curto prazo da demanda de gasolina automotiva no brasil.** 2000. PUC-Rio, 2000. Disponível em: <<http://www.maxwell.vrac.puc-rio.br/7457/7457.PDF>>. Acesso em: 13 jun. 2018.

ZHAI D., MENDEL, J. M., "**Uncertainty measures for general Type-2 fuzzy sets**", In Information Sciences, Volume 181, Issue 3, 2011, pp. 503-518.

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.

SOBRE OS AUTORES

Pedro Henrique Meira de Andrade Possui graduação em Engenharia Elétrica na UFPB (2016), Mestrado em Engenharia Elétrica na UFPB (2018) e Especialização em gestão de Pessoas - IESP (2018). Atualmente é Perito Oficial Criminal de Engenharia Elétrica no Instituto Técnico-Científico de Perícia (ITEP) no Rio Grande do Norte, possuindo cursos na área de Segurança Pública e Ciências Forenses. Atua como pesquisador independente junto ao Programa de Pós-Graduação da UFPB e está cursando o Doutorado como aluno-regular na UFRN no PPGEEC na área de Engenharia da Computação. Participou durante um ano em um PIBIC pesquisando na área de Conversores CA-CA no acionamento de máquinas pentafásicas. E durante dez meses trabalhou em um Projeto de Extensão (PROEXT) para Confecção de Kits Robóticos Educacionais para alunos de escolas públicas. Tem experiência na área de Tratamento de Dados e Inteligência Computacional Aplicada. Com áreas de interesse em: Redes Elétricas Inteligentes, Tratamento de dados, Perdas Não-Técnicas e Carros inteligentes.

Prof. Dr. Juan Moises Mauricio Villanueva, obteve o grau de Doutor em Engenharia Elétrica pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Mestre em Engenharia Elétrica pela Universidade Federal do Maranhão (UFMA) e Bacharel em Engenharia Eletrônica pela Universidad Nacional de Ingeniería UNI-Lima/Perú.

Membro ativo do Institute of Electrical and Electronics Engineers (IEEE) e da Sociedade Brasileira da Automática (SBA). Em 2018 recebeu a nomeação de Senior Member IEEE.

Professor do Departamento de Engenharia Elétrica do Centro de Energias Alternativas e Renováveis – CEAR da Universidade Federal da Paraíba (UFPB),

com áreas de interesse em: Análise de Dados (Big Data), Smart Meters e Fusão de Dados, Inteligência Artificial e suas Aplicações em Automação e Controle de Sistemas de Medição usando Transdutores Ultrassônicos.

Prof. Dr. Helon David de Macêdo Braz possui graduação em Engenharia Elétrica pela Universidade Federal da Paraíba (2002), mestrado em Engenharia Elétrica pela Universidade Federal de Campina Grande (2003) e doutorado em Engenharia Elétrica pela Universidade Federal de Campina Grande (2010). Atualmente é professor na Universidade Federal da Paraíba - UFPB, tendo sido engenheiro da Companhia Hidro Elétrica do São Francisco por 7 anos (2006-2012). Tem experiência na área de Engenharia Elétrica, com ênfase em Transmissão e Distribuição da Energia Elétrica, atuando principalmente nos seguintes temas: inteligência artificial, otimização do sistema elétrico de potência, configuração ótima de redes primárias, estimação de estado em redes de transmissão e distribuição.



Este livro foi diagramado pela
Editora UFPB em novembro de 2020

