



PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

JOÃO LUCAS DE CASTRO LISBOA



DISSERTAÇÃO

AVALIAÇÃO DOS IMPACTOS DE MÚLTIPLOS AGENTES NA GESTÃO
DA POLUIÇÃO SALINA EM SUBESTAÇÕES DE ALTA TENSÃO



SÃO CRISTÓVÃO
2025

JOÃO LUCAS DE CASTRO LISBOA

AVALIAÇÃO DOS IMPACTOS DE MÚLTIPLOS AGENTES NA GESTÃO DA
POLUIÇÃO SALINA EM SUBESTAÇÕES DE ALTA TENSÃO

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica (PROEE) da Universidade Federal de Sergipe como parte dos requisitos obrigatórios para a obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Orientador:

Professor Tarso Vilela Ferreira, D. Sc.

São Cristóvão
2025



UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA-PROEE

TERMO DE APROVAÇÃO

“Avaliação dos impactos de múltiplos agentes na gestão da poluição salina em subestações de alta tensão”

Discente:

João Lucas de Castro Lisboa

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Sergipe, como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica.

Aprovada pela banca examinadora composta por:

Documento assinado digitalmente
 **DOUGLAS BRESSAN RIFFEL**
Data: 29/08/2025 12:23:09-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Douglas Bressan Riffel (PROEE/UFS)
Presidente

Documento assinado digitalmente
 **JOSE PISSOLATO FILHO**
Data: 29/08/2025 16:16:20-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. José Pissolato Filho (UNICAMP)
Examinador Externo

Documento assinado digitalmente
 **Bruno Albuquerque Dias**
Data: 29/08/2025 15:49:19-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Bruno Albuquerque Dias (IFPE)
Examinador Externo

Documento assinado digitalmente
 **JOAO LUCAS DE CASTRO LISBOA**
Data: 29/08/2025 21:35:32-0300
Verifique em <https://validar.iti.gov.br>

João Lucas de Castro Lisboa
Discente

Cidade Universitária “Prof. José Aloísio de Campos”, 28 de agosto de 2025.

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
UNIVERSIDADE FEDERAL DE SERGIPE**

L769a Lisboa, João Lucas de Castro.
Avaliação dos impactos de múltiplos agentes na gestão da
poluição salina em subestações de alta tensão / João Lucas de
Castro Lisboa; orientador Tarso Vilela Ferreira. – São Cristóvão,
SE, 2025.
43 f. : il.

Dissertação (mestrado em Engenharia Elétrica) – Universidade
Federal de Sergipe, 2025.

1. Engenharia elétrica. 2. Isoladores e isolamentos elétricos. 3.
Subestações elétricas. 4. Salinidade. I. Ferreira, Tarso Vilela, orient.
II. Título.

CDU 621.3

RESUMO

A crescente dependência da sociedade moderna por energia elétrica torna essencial a manutenção da operação dos serviços de energia elétrica com qualidade e confiabilidade. Nesse contexto, este trabalho tem como objetivo analisar múltiplas variáveis para identificar sinais que indiquem a degradação do desempenho de isolamentos elétricos em subestações de alta tensão, com foco na influência de variáveis meteorológicas, como temperatura e precipitação, sobre os sinais de corrente de fuga. A pesquisa foi realizada com dados coletados em ambiente real de aplicação, em isolamentos da subestação (500 kV) da Usina Termelétrica Porto de Sergipe I, pertencente à Eneva S.A. A subestação está localizada a 1,5 km do mar, no município da Barra dos Coqueiros, Sergipe, Brasil. Nesta pesquisa foi realizada uma análise para verificar-se como atributos da corrente de fuga, associados às condições meteorológicas, podem indicar as condições de isolamento e como o fenômeno pode ser compreendido a partir de uma abordagem multivariada. Para isso, foram analisados períodos de estiagem ao longo de um ano, utilizando-se ferramentas de análise de tendências em curvas temporais, matriz de covariância, análise de componentes principais e agrupamento de dados. Os resultados evidenciaram a existência de padrões distintos, possibilitando identificar a contribuição de cada variável analisada e classificar os dados em agrupamentos principais. A metodologia adotada se mostrou eficaz na caracterização do fenômeno e contribui para estratégias de monitoramento preditivo em ambientes de alta tensão.

Palavras-chave: Isoladores. Corrente de Fuga. Monitoramento Preditivo. Análise de Múltiplas Variáveis. Agrupamentos.

ABSTRACT

The increasing dependence of modern society on electric power makes it essential to maintain the operation of electric power services with quality and reliability. In this context, this work aims to analyze multiple variables to identify signals that indicate the degradation of the performance of electrical insulation in high-voltage substations, with a focus on the influence of meteorological variables, such as temperature and precipitation, on leakage current signals. The research was conducted with data collected in a real application environment, on the insulation of the (500 kV) substation of the Usina Termelétrica Porto de Sergipe I, owned by Eneva S.A. The substation is located 1.5 km from the sea, in the municipality of Barra dos Coqueiros, Sergipe, Brazil. In this research, an analysis was carried out to verify how attributes of the leakage current, associated with meteorological conditions, can indicate the insulation conditions and how the phenomenon can be understood from a multivariate approach. For this purpose, dry periods over a year were analyzed, using trend analysis tools in time series, covariance matrix, principal component analysis, and data clustering. The results evidenced the existence of distinct patterns, making it possible to identify the contribution of each variable analyzed and classify the data into main clusters. The adopted methodology proved to be effective in characterizing the phenomenon and contributes to predictive monitoring strategies in high-voltage environments.

Keywords: Insulators. Leakage Current. Predictive Monitoring. Multivariate Analysis. Clustering.

LISTA DE ILUSTRAÇÕES

Figura 1 Localização da UTEPSI relativamente ao oceano atlântico.	6
Figura 2 Exemplo de alinhamento entre duas curvas utilizando o algoritmo DTW (<i>Dynamic Time Warping</i>).	9
Figura 3 Ajuste iterativo da curva média entre duas sequências a partir do método DBA.	10
Figura 4 Importância da padronização dos dados para o PCA. Principais direções P1 e P2 para	11
Figura 5 Fluxograma do Algoritmo <i>K-Means Clustering</i>	13
Figura 6 Fluxograma do procedimento metodológico.	16
Figura 7 Disposição dos sensores de corrente de fuga na UTEPSI, Barra dos Coqueiros.	17
Figura 8 Identificação dos períodos de estiagem compreendidos entre junho de 2024 e junho de 2025	19
Figura 9 Processamento de sinais de corrente.	21
Figura 10 Suavização de uma curva com média móvel.	22
Figura 11 Média global da corrente de fuga por horas do dia.	23
Figura 12 Ajuste das curvas características globais obtidas por DBA.	28
Figura 13 Distribuição das componentes principais (PC1 e PC2).	33
Figura 14 Gráfico de correlação entre as componentes principais do PCA.	34
Figura 15 Gráficos de dispersão dos <i>clusters</i> individuais.	37
Figura 16 Ajuste de curvas entre irms e dh para os clusters 0 e 1.	38
Figura 17 ajuste linear entre temperatura e corrente eficaz para os cluster 0 e 1.	38
Figura 18 ajuste das curvas entre temperatura e distorção harmônica para os cluster 0 e 1.	39

LISTA DE TABELAS

Tabela 1 Períodos de estiagem considerados para análise.....	20
Tabela 2 Quantidade percentual de dados válidos dos sensores por período de análise.....	23
Tabela 3 Classificação do coeficiente de correlação entre duas variáveis.....	26
Tabela 4 Parâmetros para comparação das curvas de corrente de fuga individuais com a curva característica global.....	29
Tabela 5 Matriz dos coeficientes de correlação de Pearson, Spearman e Kendall das variáveis analisadas.	31
Tabela 6 Contribuição das variáveis na representação do PCA.	32
Tabela 7 características individuais de cada agrupamento.	35
Tabela 8 Quantidade de ocorrências em cada <i>cluster</i> , por período e sensor.....	36

SUMÁRIO

1	Introdução	5
1.1	Objetivos	6
1.2	Estrutura do Texto.....	7
2	Fundamentação Teórica	8
2.1	Isoladores	8
2.2	<i>Dynamic Time Warping Barycenter Averaging (DBA)</i>	9
2.3	<i>Principal Component Analysis (PCA)</i>	11
2.4	<i>K-means clustering</i>	12
3	Revisão Bibliográfica	14
4	Metodologia	16
4.1	Aquisição e composição da base de dados	16
4.1.1	Aquisição de dados meteorológicos	17
4.1.2	Aquisição de dados de corrente	18
4.2	Processamento e tratamento dos dados	18
4.2.1	Processamento dos sinais meteorológicos	18
4.2.2	Processamento dos sinais de corrente de fuga	20
4.3	Análise exploratória dos dados.....	24
4.3.1	Análise temporal	25
4.3.2	Análise por correlação	25
4.4	Identificação de padrões.....	26
4.4.1	<i>Principal Component Analysis (PCA)</i>	27
4.4.2	Análise por agrupamentos (<i>Clustering</i>).....	27
5	Resultados e Discussões	28
5.1	Análise exploratória	28
5.2	Análise por agrupamentos	34
6	Conclusão.....	40
	Referências Bibliográficas.....	41

1 INTRODUÇÃO

A demanda e a dependência da energia elétrica na sociedade atual implicam na necessidade de fornecimento contínuo e na confiabilidade do sistema elétrico. São muitos os equipamentos envolvidos e com diferentes funções. Entre eles estão os isoladores, que são responsáveis pelo suporte mecânico e pelo isolamento elétrico dos outros equipamentos. A falha dos isoladores pode causar danos ao sistema, resultando em desabastecimento de energia e prejuízos financeiros (DOLNÍK, *et. al.*, 2022)

Quando estão em operação os isoladores estão continuamente expostos à poluição do ambiente, como resíduos industriais, pesticidas agrícolas e maresia (AHMED, *et al.*, 2022). Essa poluição causa danos e acúmulo de resíduos nos isoladores, resultando na ocorrência de *flashover* e perda total do equipamento. O nível de contaminação pode ser determinado através da análise de *equivalent salt deposit density* (ESDD) e *non-soluble deposit density* (NSDD) (MONTTOYA, RAMIREZ, & MONTTOYA, 2004).

A contaminação do isolador associada à presença de umidade degrada a performance dielétrica do isolamento (SALEM, *et al.*, 2022). Essa combinação resulta na dissolução dos sais contaminantes, formando uma camada condutora sobre a superfície do isolador. Por sua vez, aumentando a corrente de fuga e favorecendo a ocorrência de faíscas que podem ficar mais intensas até a ocorrência do *flashover* (SALEM, *et al.*, 2021).

Para evitar falhas inesperadas, faz-se necessária a aplicação de medidas preventivas, como inspeções e monitoramento do equipamento. Dessa forma, identificam-se os equipamentos que precisarão de ações corretivas. Diversas técnicas foram desenvolvidas para analisar as condições dos isoladores, destacando-se a inspeção visual, a termografia, a análise de campo elétrico e a medição da corrente de fuga.

Neste trabalho é realizada uma análise dos dados de corrente de fuga registrados em campo (em ambiente real de aplicação) em isolamentos da subestação (SE) de 500 kV da Usina Termelétrica Porto de Sergipe I (UTEPS I), pertencente à Eneva S.A. A UTEPS I está localizada a 1,5 km do mar, no município da Barra dos Coqueiros, Sergipe, Brasil, conforme pode-se observar na Figura 1.

Figura 1 Localização da UTEPSI relativamente ao oceano atlântico.



Fonte: autoria própria.

Este trabalho foi desenvolvido com apoio do Laboratório de Alta Tensão (LAT) do Departamento de Engenharia Elétrica (DEL), e do Laboratório de Automação, Controle e Simulação (LACS) do Núcleo de Pesquisa em Petróleo e Gás (NUPEG), ambos da Universidade Federal de Sergipe (UFS). As atividades são parte do projeto GImpSI – Gestão de Impactos da Salinidade em Isolamentos, realizado no âmbito do programa P&D da Agência Nacional de Energia Elétrica (ANEEL), código PD-11278-0001-2021, financiado pela ENEVA S.A. e gerido pelo INESC P&D Brasil.

O objetivo do projeto é conceber um sistema de monitoramento e gestão dos efeitos do acúmulo de poluição nos isolamentos. Como insumos de informação para tal sistema empregam-se dados de corrente de fuga, dados climatológicos, informações das grandezas elétricas dos barramentos da SE e imagens registradas por câmeras sensíveis a infravermelho e ultravioleta.

1.1 OBJETIVOS

O objetivo geral deste trabalho é buscar correlações entre atributos extraídos de sinais de corrente de fuga medida em isolamentos, indicadores de sua condição operacional, e fatores meteorológicos na região de instalação dos mesmos, em uma subestação de 500 kV.

Como objetivos específicos dessa pesquisa, pode-se citar:

- Construir uma base de dados composta por grandezas que impactam no desempenho dos isolamentos (temperatura ambiente e precipitação) de uma subestação de 500 kV;

- Propor uma metodologia de análise a partir de redução de dimensionalidade dos dados;
- Propor uma metodologia de análise por agrupamentos, buscando identificar características semelhantes no conjunto de dados.

1.2 ESTRUTURA DO TEXTO

A presente dissertação está organizada em 6 capítulos. O Capítulo 2 corresponde à fundamentação teórica, onde são apresentados os conceitos, técnicas e ferramentas utilizadas no desenvolvimento da pesquisa.

No Capítulo 3 é apresentada a revisão bibliográfica realizada sobre os métodos de monitoramento de equipamentos elétricos de alta tensão, a partir de dados de corrente de fuga e as correlações com variáveis meteorológicas.

No Capítulo 4 é descrita a metodologia utilizada para aquisição, processamento, tratamento e análise dos dados de corrente de fuga no contexto desta pesquisa. Posteriormente, os resultados das análises são discutidos no Capítulo 5.

Por fim, no Capítulo 6 são apresentadas as conclusões conforme as discussões apresentadas no capítulo anterior, bem como sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção são discutidos os conceitos fundamentais relacionados ao monitoramento de equipamentos elétricos, métodos de análise de séries temporais e técnicas de agrupamento que serviram como base para o desenvolvimento desta pesquisa.

2.1 ISOLADORES

Isoladores são equipamentos que têm a função de manter a distância de isolamento entre duas estruturas com potenciais diferentes, impedindo uma circulação indesejável de corrente, além de servirem como sustentação para cabos energizados (MORAES, 2016).

Dessa maneira, é fundamental que suas condições de operação sejam mantidas adequadas para preservar a estabilidade do sistema. Entretanto, as falhas em isoladores são as principais causas de desligamentos do sistema de potência (MENDONÇA, 2016).

São diversos os fatores que influenciam na integridade dos isoladores, sendo alguns deles: descargas elétricas superficiais e parciais, exposição à radiação ultravioleta, umidade, poluição salina, entre outros. A influência desses agentes varia a depender do tipo de isolador utilizado, pois materiais distintos apresentam diferentes níveis de resistência e durabilidade frente a estas condições (BARBOSA, 2025).

O acúmulo de poluição, em especial, potencializa a ação desses fatores. Uma das principais consequências é a oxidação de elementos metálicos existentes nas terminações do isolador. Outra consequência está associada à formação de uma camada superficial condutora, que pode provocar o aumento da corrente de fuga (BARBOSA, 2025).

O monitoramento e a lavagem periódica são essenciais para preservar a integridade destes equipamentos e prevenir falhas. Existem diversas abordagens para avaliar as condições dos isolamentos, dentre elas destacam-se a inspeção visual, a análise por ESDD (*Equivalent Salt Deposit Density*) e NSDD (*Non-Soluble Deposit Density*), a termografia, detecção de descargas corona e a inspeção por corrente de fuga.

O aumento da corrente de fuga pode estar associado a uma redução das características isolantes do componente, o que pode ocasionar descargas superficiais e, nas situações mais severas, descargas disruptivas (*flashover*) (BARBOSA, 2025).

A presença de uma componente capacitiva é uma das características do sinal de corrente de fuga, evidenciada pelo o adiantamento em relação ao sinal de tensão. Além disso, também apresenta características não-lineares, tendo em vista a não linearidade inerente aos materiais dielétricos diante da aplicação da tensão (MENDONÇA, 2016; BARBOSA, 2025).

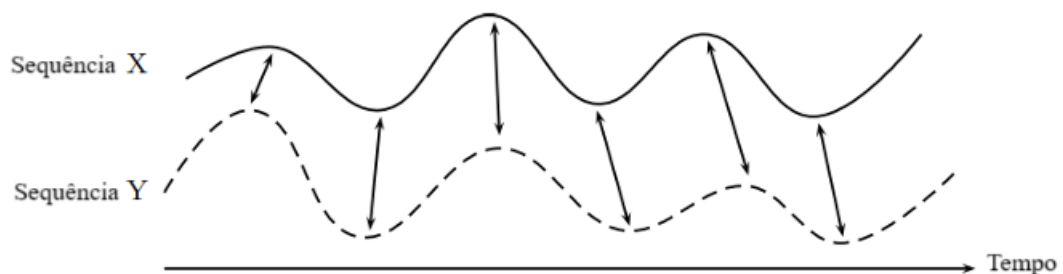
Sendo assim, a amplitude do sinal de corrente de fuga, as harmônicas presentes e suas amplitudes e as relações entre elas fornecem informações para o monitoramento de isoladores (MENDONÇA, 2016).

A partir dessa análise, parâmetros como corrente eficaz, distribuição espectral obtida por Transformada Rápida de Fourier e conteúdo harmônico tornam-se ferramentas fundamentais para a avaliação quantitativa da condição do isolamento.

2.2 *DYNAMIC TIME WARPING BARYCENTER AVERAGING* (DBA)

Trata-se de um método de alinhamento global proposto por PETITJEAN *et. al.* (2010) que consiste em refinar iterativamente uma sequência de dados inicial (potencialmente arbitrária), a fim de minimizar a distância média entre ela e as sequências de entrada. Nesse algoritmo, a medida de distância entre duas curvas que deseja ser minimizada é obtida a partir do DTW (*Dynamic Time Warping*), que é um algoritmo de alinhamento ótimo capaz de identificar similaridades de forma flexível, alinhando as coordenadas de ambas as sequências, como ilustrado na Figura 2.

Figura 2 Exemplo de alinhamento entre duas curvas utilizando o algoritmo DTW (*Dynamic Time Warping*).



Fonte: adaptada de Müller (2007).

O método inicia uma sequência inicial com o baricentro entre as todas as curvas disponíveis, de acordo com a eq. (1):

$$\text{baricentro}\{X_1, \dots, X_n\} = \frac{X_1 + \dots + X_n}{n}, \quad (1)$$

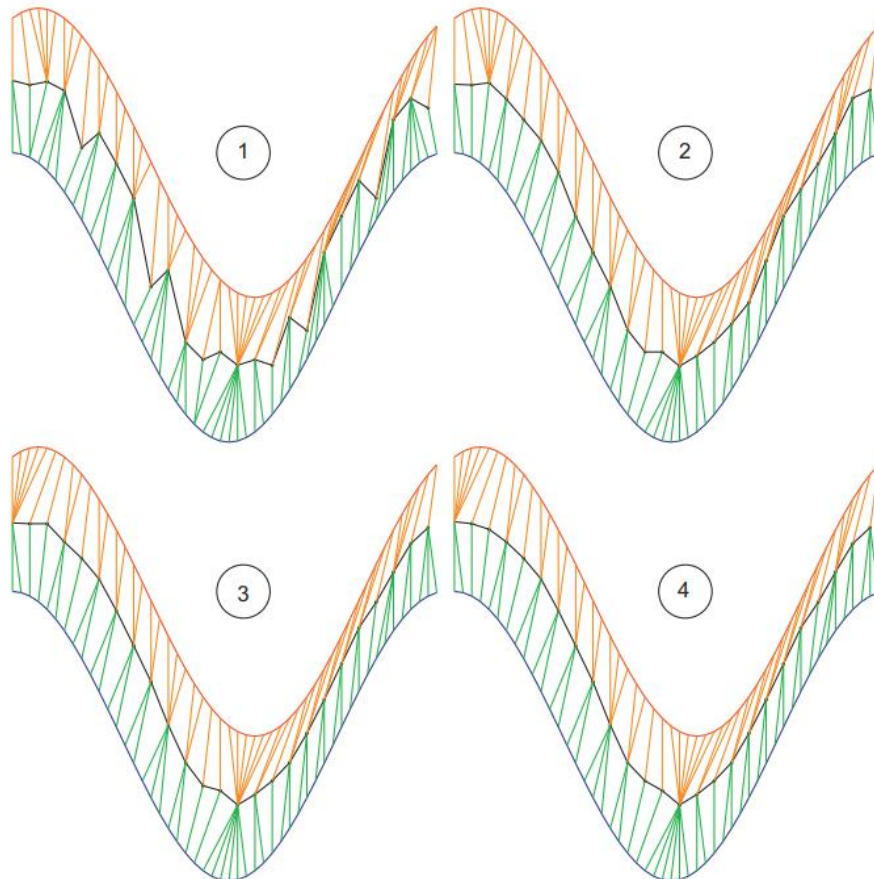
em que X_n representa a “enésima” amostra da série e n representa o total de elementos na série.

Em seguida, para cada iteração são realizadas duas etapas:

1. Aplica-se DTW entre cada curva individual e a curva média temporária para identificar associações entre coordenadas das curvas;
2. Atualiza cada coordenada da curva média temporária com o baricentro calculado a partir das coordenadas associadas a ela no passo anterior.

O algoritmo continua iterando até que a curva global pare de mudar significativamente, de acordo com um limiar de tolerância. O processo é ilustrado na Figura 3.

Figura 3 Ajuste iterativo da curva média entre duas sequências a partir do método DBA.



Fonte: PETITJEAN *et. al.* (2010).

No âmbito desta pesquisa, utiliza-se o DBA na análise temporal do fenômeno, a partir da comparação entre curvas de corrente de sensores diferentes, buscando observar similaridades entre elas.

2.3 PRINCIPAL COMPONENT ANALYSIS (PCA)

Trata-se de uma ferramenta de análise de dados, simples e não paramétrica, com potencial de redução de dimensionalidade e extração de características de conjuntos de dados complexos.

A ideia central do PCA é reduzir as dimensões de um conjunto de dados composto por várias variáveis correlatas, enquanto mantém o máximo possível da variação presente no conjunto. Isso é alcançado por meio da transformação em um novo conjunto de variáveis, os componentes principais (PCs), que não possuem correlação entre si e são ordenadas de forma que a primeira componente retenha a maior variação presente em todas as variáveis originais. (JOLLIFFE, 2022, p. 1)

No contexto desta pesquisa, justifica-se a aplicação desta ferramenta por sua capacidade de simplificar a complexidade dos dados sem perda significativa de informação e pela possibilidade de identificar padrões que não são evidentes nos dados brutos.

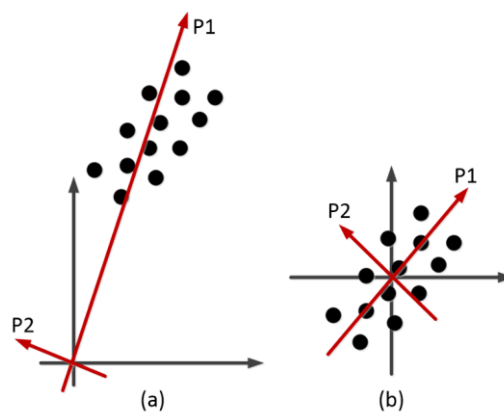
Suponha que X seja um conjunto de dados de dimensão $n \times p$, em que n é a quantidade de amostras de x , e p é a quantidade de variáveis associadas à x .

Inicialmente, recomenda-se padronizar os dados para média zero e variância unitária. Obtém-se o novo conjunto de dados Z conforme a eq. (2):

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad (2)$$

em que $\bar{x}_j$ e σ_j são a média e o desvio padrão de cada coluna j da matriz X , respectivamente. Sem a padronização, o resultado do PCA ficará enviesado pela variável de maior escala, conforme pode ser visualizado na Figura 4.

Figura 4 Importância da padronização dos dados para o PCA. Principais direções P1 e P2 para (a) dados não centrados em zero e (b) dados centrados em zero.



Fonte: (GHOJOGH, B. et. al., 2019).

Busca-se uma matriz V , capaz de transformar Z em um novo conjunto de dados Y , conforme a eq. (3):

$$Y = ZV, \quad (3)$$

com quantidade menor de variáveis, ortogonais entre si, mas mantendo suas características de maior variância.

Para isso, calcula-se a matriz de covariância C do conjunto de dados padronizados, conforme a eq. (4):

$$C = \frac{1}{n} Z^T Z. \quad (4)$$

Em seguida é realizada a decomposição espectral da matriz de covariância, onde serão obtidos os autovalores λ_k de C , a partir da eq. (5):

$$C - \lambda I = 0, \quad (5)$$

em que I é a matriz identidade. Os autovalores são ordenados de forma decrescente, de modo a representar a maior variância possível com os primeiros componentes.

Além disso, obtém-se os autovetores v_k associados aos autovalores selecionados λ_k que, por sua vez, indicam a direção da maior variância do novo conjunto. A projeção dos dados nas novas componentes principais é dada pela eq. (3), em que V é a matriz composta por esses autovetores (GHOJOGH, B. *et. al.*, 2019).

A contribuição de cada componente principal é expressa em porcentagem pela proporção da variância total que a componente k é responsável. Esse parâmetro é utilizado para determinar a quantidade de componentes que deve ser retida. Em muitos casos, adota-se modelos que expliquem pelo menos 80% da variação total (HONGYU, K. *et. al.*, 2015).

Apesar da aparente simplicidade da técnica, ela é amplamente utilizada em diversas áreas como agricultura, biologia, química, meteorologia, economia, engenharia, entre outras (JOLLIFFE, 2022). Nesta pesquisa, ela é utilizada como etapa preliminar a fim de reduzir a dimensionalidade e posteriormente combinar com técnicas de agrupamentos.

2.4 *K-MEANS CLUSTERING*

O problema fundamental do agrupamento (*clustering*) é agrupar dados que são similares entre si. Essa ferramenta tem sido utilizada e aprimorada em diversas áreas como mineração de dados, análise estatística, reconhecimento de padrões e problemas de otimização (BRADLEY, P *et. al.*, 1998).

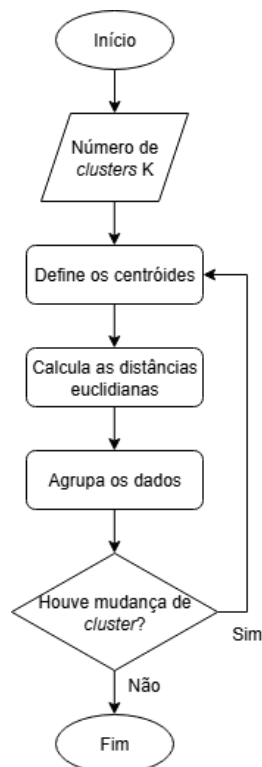
No âmbito desta pesquisa, essa estratégia é utilizada para verificar relações entre as grandezas em conjuntos de dados com características semelhantes, visto que se trata de um sistema composto por diversos sensores, instalados em equipamentos diferentes, e com diversas variáveis analisadas.

O algoritmo de *clustering* utilizado neste trabalho denomina-se *K-Means*. Este algoritmo acha a solução ótima local minimizando a soma da distância euclidiana entre cada ponto do conjunto de dados e o centroide do *cluster* mais próximo (BRADLEY, P *et. al.*, 1998).

Inicialmente define-se a quantidade de *clusters*. Em seguida, inicializam-se os centroides em posições aleatórias e então calcula-se a distância euclidiana de cada ponto do conjunto de dados à todos os centroides disponíveis. O agrupamento é definido pela proximidade dos dados aos centroides, em que cada amostra fica associado ao *cluster* de menor distância (NIENNATTRAKUL *et. al.*, 2007).

Em seguida, calcula-se o novo centroide do *cluster*, sendo este a média de todas as amostras classificadas naquele agrupamento. Repete-se o procedimento até que mais nenhuma amostra mude de *cluster* (NIENNATTRAKUL *et. al.*, 2007). O fluxograma do algoritmo é apresentado na Figura 5. A partir a finalização do algoritmo, a análise dos dados é realizada a partir de cada agrupamento.

Figura 5 Fluxograma do Algoritmo *K-Means Clustering*.



Fonte: autoria própria.

3 REVISÃO BIBLIOGRÁFICA

Diversos trabalhos demonstram que o acúmulo de poluição altera significativamente as características da corrente de fuga, tornando-a um parâmetro eficaz para monitoramento da integridade dos isoladores.

Pesquisas comparativas entre isoladores poluídos e isoladores limpos, revelam mudanças nas características da forma de onda da corrente, com análises no domínio do tempo (magnitude e defasagem) e no domínio da frequência (harmônicas), conforme evidenciado por JAYASUNDARA *et. al.*, 2008, PAKPAHAN *et. al.* 2008, BASHIR *et. al.* 2009 e ZENGHFA *et. al.* 2018.

JAYASUNDARA *et. al.* (2008) e PAKPAHAN *et. al.* (2008), observaram que isoladores poluídos apresentam maior magnitude da corrente e defasagem acentuada em relação à tensão aplicada, quando comparados a isoladores limpos, indicando redução das características capacitivas do material. JAYASUNDARA *et. al.* (2008) também propôs uma tipologia de 5 categorias de corrente de fuga, baseada na não-linearidade do sinal, que varia conforme o grau de poluição e névoa.

Além da magnitude da corrente, BASHIR *et. al.* (2009) e ZENGHFA *et. al.* (2018) destacaram que as harmônicas ímpares, de terceira e quinta ordens, aumentam com a degradação do isolamento, sugerindo sua utilização como indicadores para monitoramento.

O ambiente também tem influência na forma de onda da corrente de fuga, como apontam os trabalhos de PAKPAHAN *et. al.* (2008), RAMIREZ *et. al.* (2012), BANIK *et. al.* (2015), MOHAMMADNABI, *et. al.* e (2021), MANJANG, *et. al.* (2021).

PAKPAHAN *et. al.* (2008) e MOHAMMADNABI *et. al.* (2021) indicam influência da umidade na forma de onda da corrente de fuga, com efeito potencializado em isoladores poluídos. Nestes casos, a magnitude da corrente de fuga e das harmônicas aumentam quando submetidos à ambientes úmidos.

Por sua vez, MANJANG *et. al.* (2021) atentam para as propriedades hidrofóbicas dos isoladores poliméricos. Em suas avaliações, identificaram que a temperatura, a umidade e a precipitação não têm efeito significativo nas flutuações de corrente de fuga para esse tipo de isolador, associando a hidrofobicidade à uma característica de resiliência do equipamento. Entretanto, a radiação ultravioleta possui efeito degradante dessa propriedade neste caso.

Nos trabalhos de RAMIREZ *et. al.* (2012) e BANIK *et. al.* (2015) identificou-se que tipos de poluição diferentes também influenciam em características diferentes da corrente. E que, neste caso, isoladores com magnitudes de correntes mais baixas não apresentam, necessariamente, melhor performance.

Diante da influência de diversos fatores sobre a forma de onda da corrente de fuga, diversos estudos passaram a identificar padrões, desenvolver técnicas de análise e elaborar modelos de previsão do desempenho dos isoladores.

AMARH *et. al.* (2002) propõem uma abordagem estatística para analisar a corrente de fuga em isoladores poluídos, visando aprimorar o monitoramento de sua condição e prevenir o *flashover*. Essa abordagem foi justificada pelo caráter estocástico do *flashover*, devido à natureza não linear das descargas superficiais. Em vez de analisar os valores de pico instantâneos da corrente, o estudo concentra no envelope do sinal, que fornece informações mais consistentes sobre o desempenho do isolador. A técnica examina a frequência com que o envelope da corrente de fuga cruza um determinado limiar de amplitude, que aumenta significativamente com a aproximação do *flashover*. Entretanto, os parâmetros do modelo variam com o tempo, com os fatores ambientais, com o nível e a natureza da poluição e com a condição dos isolamentos.

No trabalho de FARAMARZI *et. al.* (2016), sugere-se o monitoramento pela razão entre as amplitudes da terceira e quinta harmônicas. A partir de testes em laboratório, verificou-se que, em isoladores limpos e sob condições normais, a amplitude da quinta harmônica da corrente de fuga é consistentemente maior que a da terceira harmônica. Com a presença de poluição e umidade, a amplitude da terceira harmônica aumenta significativamente, superando o índice da quinta harmônica e aumentando a probabilidade da ocorrência de *flashover*. Entretanto, a própria natureza não linear do fenômeno associado à poluição e umidade é uma limitação desta técnica. Além disso, o ambiente controlado em laboratório não reflete exatamente as condições reais em campo.

No estudo desenvolvido por CASTILLO-SIERRA *et. al.* (2021) foi proposto um modelo de previsão a partir da regressão quadrática, visando encontrar o momento ideal para a lavagem dos isolamentos. O modelo foi determinado a partir da observação dos padrões de corrente durante os dias, e considera que a corrente de fuga aumenta progressivamente com o acúmulo de poluição. Também são aplicadas técnicas de controle e otimização, para minimizar alarmes falsos e lavagens desnecessárias.

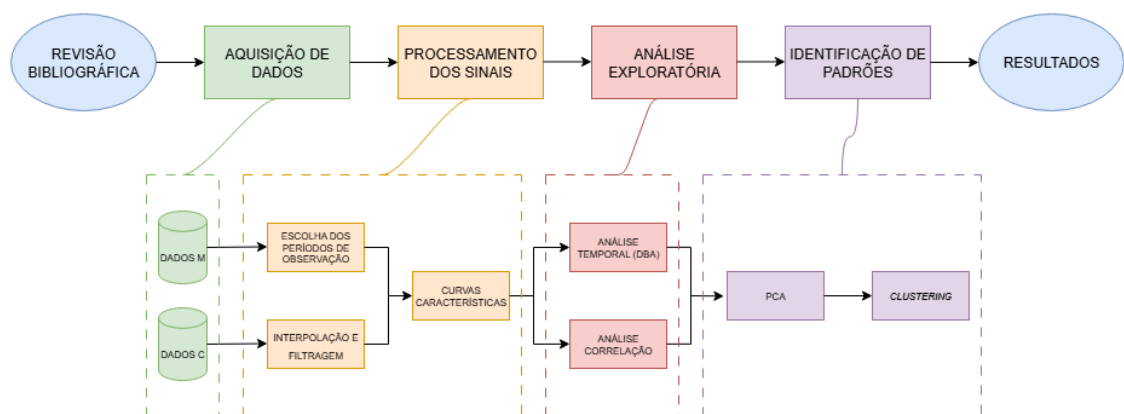
4 METODOLOGIA

Neste trabalho apresenta-se uma avaliação das condições de isolamento em equipamentos elétricos de alta tensão monitorados a partir de sinais de corrente de fuga, e correlações destas condições com as variáveis meteorológicas. Esta análise foi conduzida em quatro etapas principais:

- Aquisição, composição e estruturação de base de dados de corrente e parâmetros meteorológicos;
- Processamento e tratamento dos dados, incluindo a redução do volume, a identificação e o tratamento de dados faltantes e inválidos;
- Análise exploratória dos dados, com foco em aspectos temporais e na correlação entre as variáveis.
- Identificação de padrões a partir de PCA e análise por agrupamentos (*clustering*).

O fluxograma do procedimento metodológico está apresentado na Figura 6 e será detalhado nas subseções correspondentes a cada uma das etapas.

Figura 6 Fluxograma do procedimento metodológico.



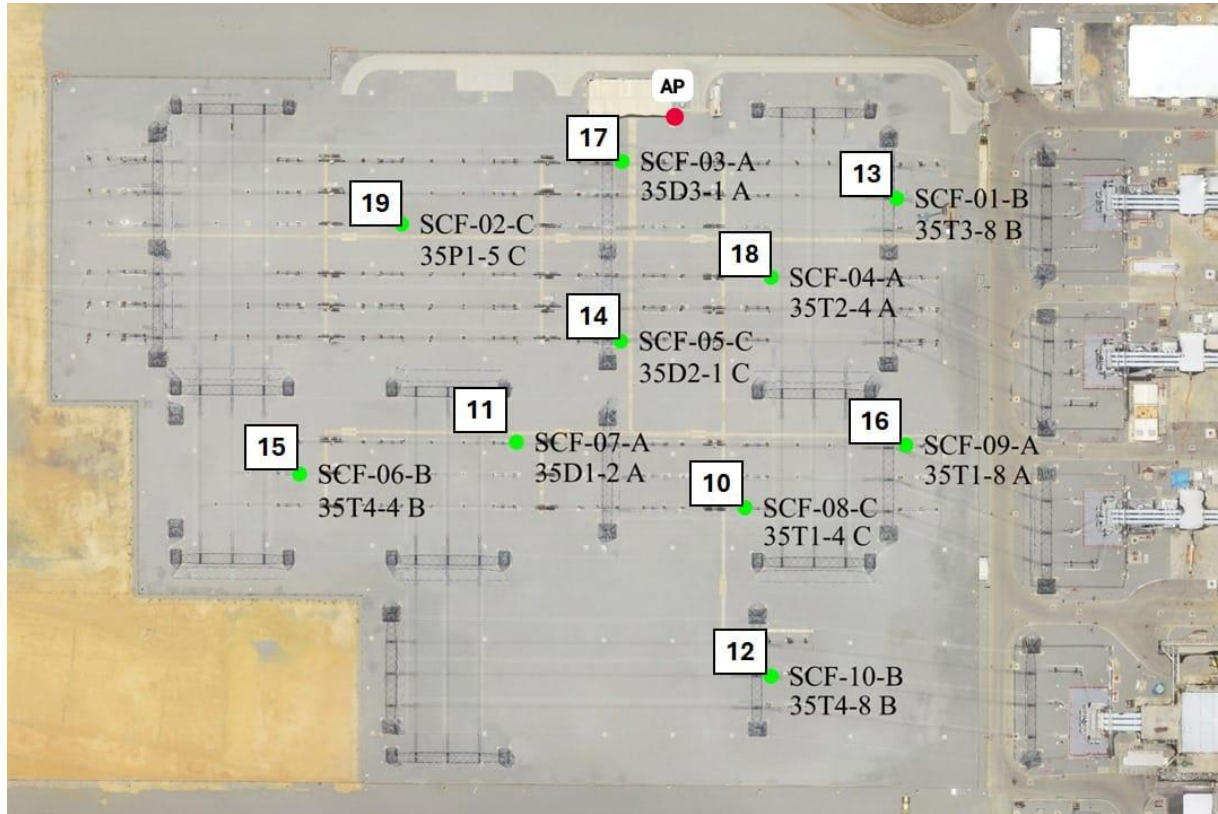
Fonte: autoria própria

4.1 AQUISIÇÃO E COMPOSIÇÃO DA BASE DE DADOS

Para a realização deste trabalho, classificou-se como dados meteorológicos, aqueles provenientes de estações meteorológicas, e como dados de corrente, aqueles adquiridos por

unidades microprocessadas com sensores de corrente de fuga (UMSCF). Na Figura 7 apresenta-se a disposição das UMSCF e da estação meteorológica do GImpSI (alocada na posição “AP”). O período analisado compreende de junho de 2024 a junho de 2025.

Figura 7 Disposição dos sensores de corrente de fuga na UTEPSI, Barra dos Coqueiros.



Fonte: autoria própria.

4.1.1 AQUISIÇÃO DE DADOS METEOROLÓGICOS

Os dados meteorológicos foram obtidos pela estação meteorológica do GImpSI (EMG) e pela estação meteorológica TRMM.12199. Esta última é parte do sistema Agritempo, um sistema de monitoramento agrometeorológico desenvolvido e mantido pela Embrapa Agricultura Digital em parceria com o Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (Cepagri) da Unicamp. As duas estações meteorológicas ficam na Barra dos Coqueiros, mesmo município da SE.

Foram priorizados os dados da estação do GimpSI, com complementação dos dados da estação TRMM.12199. As variáveis meteorológicas de entrada para a análise são definidas como: temperatura máxima (T_{max}), temperatura mínima (T_{min}), temperatura média (T_{med}) e precipitação total diária (P), em que cada leitura representa uma amostra por dia.

4.1.2 AQUISIÇÃO DE DADOS DE CORRENTE

A base de dados de corrente foi construída por meio de uma rede de UMSCF instaladas nos condutores de aterramento de dez colunas isolantes de chaves seccionadoras da UTEPSI. A aquisição dos sinais de corrente foi realizada a cada cinco minutos, com taxa de amostragem de 8 mil amostras por segundo, durante 10 segundos.

Conforme CARVALHO *et al.* (2023) e OLIVEIRA (2023), os sensores desenvolvidos foram baseados em um transformador de corrente com núcleo de liga nanocristalina, e possuem capacidade de medir correntes de fuga na faixa de centenas de microampères até dezenas de miliampères, considerando uma faixa de frequência entre 60 e 540 Hz.

Ao todo foram instalados 10 sensores em locais diferentes da SE, e cada sensor foi identificado com um número entre 10 e 19. A disposição dos sensores na SE foi definida conforme ilustrada na Figura 7.

Após serem obtidos pelos sensores, os sinais de corrente são processados por um sistema de condicionamento e, em seguida, por um sistema de comunicação. O sistema de condicionamento realiza o ajuste da resposta em frequência, calibração adaptativa e filtro *anti-aliasing*. O sistema de comunicação armazena os dados de corrente em uma memória local e os envia, via rede sem fio, para um servidor remoto, onde ficam disponíveis para processamento e análise.

4.2 PROCESSAMENTO E TRATAMENTO DOS DADOS

Nesta etapa os dados de entrada foram processados a fim de extraírem-se características dos sinais e reduzir o volume de dados na análise. O processamento dos dados meteorológicos e dos dados de corrente será detalhado a seguir nesta seção.

4.2.1 PROCESSAMENTO DOS SINAIS METEOROLÓGICOS

A análise do fenômeno foi delimitada em períodos de estiagem. Portanto, os objetivos do processamento dos sinais meteorológicos são: padronizar os dados e identificar estes períodos de estiagem.

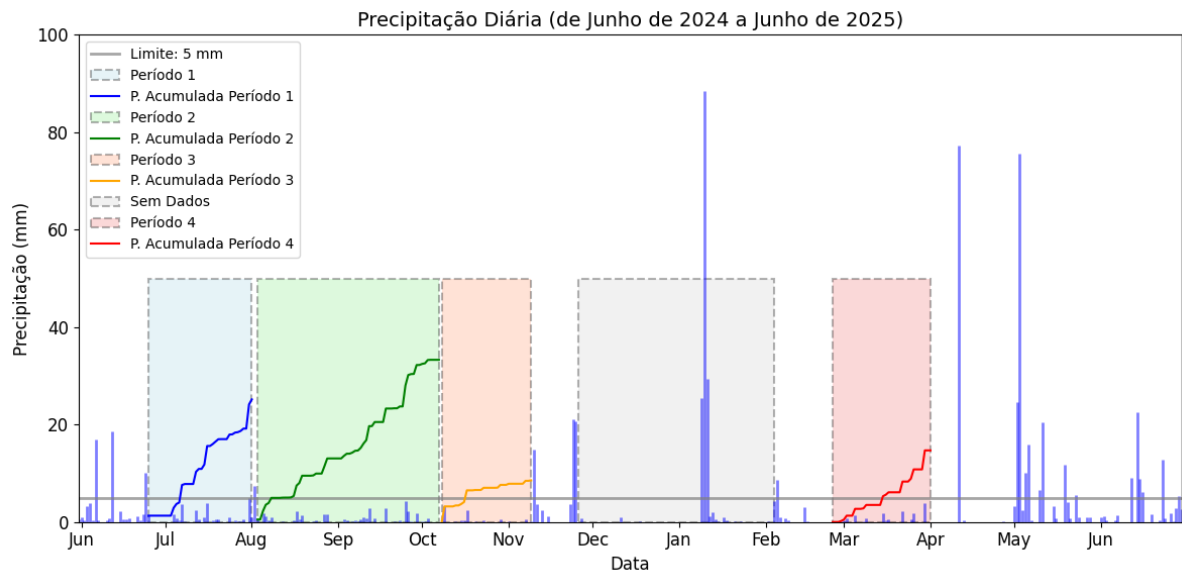
De acordo com JIANG *et. al.* (2017), chuvas com precipitação diárias acima de 5 mm/dia possuem a capacidade de provocar limpeza em isoladores, sendo capazes então de

reduzir o acúmulo de poluição depositada. Determinou-se então esse limiar de precipitação para categorizar os dias em que houve chuva forte (≥ 5 mm).

Tendo em vista que o fenômeno de deposição de poluição nos isolamentos tem uma dinâmica lenta, o período de estiagem deve ser longo o suficiente para que seja possível analisá-lo de forma mais confiável. Definiu-se, portanto, o período de estiagem como o intervalo de tempo entre duas chuvas fortes, desde que a duração deste intervalo seja superior a 7 dias.

Os eventos de precipitação e a delimitação dos períodos de estiagem escolhidos para análise podem ser observados na Figura 8.

Figura 8 Identificação dos períodos de estiagem compreendidos entre junho de 2024 e junho de 2025



Fonte: autoria própria.

Os dados das duas estações meteorológicas (EMG e TRMM.12199) foram uniformizados e padronizados (periodicidade da aquisição, unidades, etc.), de maneira que se tornaram estruturalmente idênticos para efeitos de análise.

A partir da Figura 8 observa-se que os períodos 2 e 3, destacados pelas regiões em verde e laranja, respectivamente, pertencem ao mesmo período de estiagem. No entanto, a separação em dois períodos diferentes foi necessária devido à eventual indisponibilidade de dados de sensores de corrente. Este mesmo fator impossibilitou a análise dos períodos de estiagem compreendidos na região delimitada pela cor cinza, e também motivou o recorte temporal do período de estiagem em que a região de cor vermelha está inserida. Os períodos de análise considerados para a análise são apresentados na Tabela 1.

Tabela 1 Períodos de estiagem considerados para análise.

Período	Data de Início	Data de Fim	Duração
1	24 de junho de 2024	01 de agosto de 2024	39 dias
2	3 de agosto de 2024	7 de outubro de 2024	66 dias
3	8 de outubro de 2024	9 de novembro de 2024	33 dias
4	25 de fevereiro de 2025	1 de abril de 2025	36 dias

Fonte: autoria própria.

Em seguida, foram calculados os parâmetros: amplitude térmica do dia (ΔT) e precipitação total acumulada (P_{AC}). A amplitude térmica do dia k foi calculada a partir da diferença entre a temperatura máxima (T_{max}) e a temperatura mínima (T_{min}) daquele dia, como na eq. (6):

$$\Delta T_k = T_{max,k} - T_{min,k}, \quad (6)$$

em que ΔT_k representa a amplitude térmica do dia [°C], $T_{max,k}$ representa a temperatura máxima do dia [°C] e $T_{min,k}$ representa a temperatura mínima do dia [°C]. A precipitação total acumulada do dia k ($P_{AC,k}$), representa a precipitação total acumulada até aquele dia [mm], desde o início do período de estiagem, como na eq. (7):

$$P_{AC,k} = \sum_{n=1}^k P_k, \quad (7)$$

Ao final do processamento dos dados meteorológicos foram obtidos os períodos de estiagem e as informações diárias de temperatura máxima, temperatura mínima, temperatura média, amplitude térmica e precipitação total e precipitação total acumulada. Em seguida será detalhado o processamento dos sinais de corrente.

4.2.2 PROCESSAMENTO DOS SINAIS DE CORRENTE DE FUGA

O principal desafio desta etapa é a consolidação dos sinais de corrente de fuga em um único valor diário, visando manter a compatibilidade com os dados meteorológicos. O sensor faz uma aquisição a cada 5 minutos, com amostragem de 8 mil amostras por segundo durante 10 segundos. O sinal adquirido deve ter 80 mil amostras por aquisição, com um total de 1440 leituras por dia.

Para cada sinal com 80 mil amostras, calculou-se o valor eficaz (I_{rms}) do sinal de corrente a partir da eq. (8):

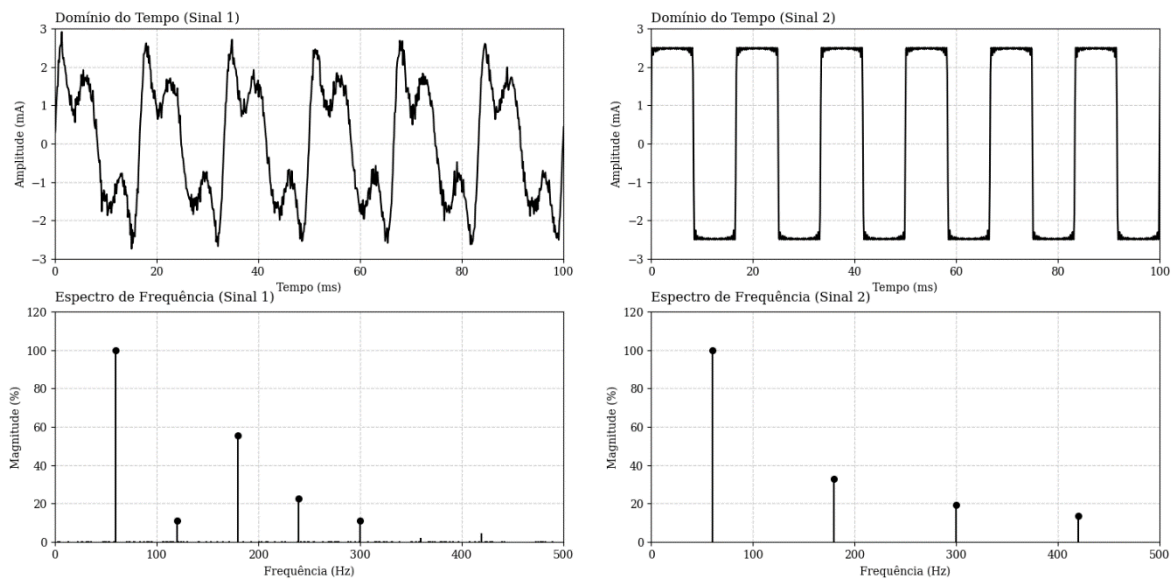
$$I_{rms} = \sqrt{\frac{1}{N} \sum_{k=1}^N i_k^2}, \quad (8)$$

em que k é o índice da amostra, N é a quantidade total de amostras e i_k é o valor de corrente na k -ésima posição.

Além do valor eficaz, também foram obtidas as amplitudes das componentes harmônicas do sinal de corrente (I_h), incluindo a componente fundamental (60 Hz) e as harmônicas ímpares até a 7ª harmônica, por meio da aplicação da FFT (*Fast Fourier Transform*).

Idealmente, há 1440 leituras diárias, entretanto algumas foram descartadas por apresentarem inconsistências como: ausência de leitura, quantidade de amostras inferior a 80 mil, ou saturação do sinal de corrente. Para ilustração, são apresentados dois sinais de corrente na Figura 9.

Figura 9 Processamento de sinais de corrente.



Fonte: autoria própria.

Na Figura 9, o Sinal 1 é representado como um caso válido para análise pois não apresenta saturação, enquanto o Sinal 2 é considerado saturado e seus parâmetros obtidos com o processamento não condizem com a informação real da corrente, sendo então descartado das amostras.

Extraídas as características das formas de onda dos sinais de corrente, têm-se agora variáveis de corrente eficaz e composição harmônica para cada aquisição de 5 minutos, totalizando 1440 medidas por dia. Nesta etapa do processamento foram aplicadas técnicas de

filtragem e suavização das curvas, incluindo: remoção de *outliers*, interpolação dos dados faltantes e média móvel simples, conforme a eq. (9):

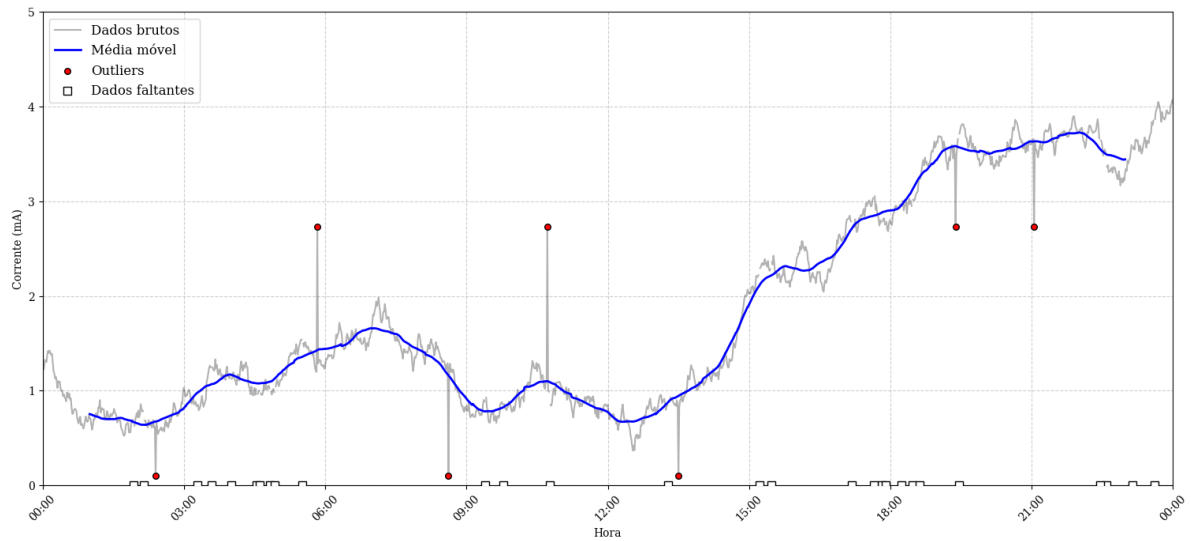
$$\bar{i}_k = \frac{1}{2N + 1} \sum_{k=-N}^N i_k. \quad (9)$$

Nesta etapa, calcula-se a média dos valores de corrente em torno do índice k , considerando o intervalo de $N - k$ até $N + k$. Esse resultado foi então associado ao elemento i_k do novo vetor de correntes ajustadas. Para identificar *outliers*, também foi calculado o desvio-padrão local (σ_k) nestes mesmos intervalos. Com os resultados da média e desvio-padrão para cada índice k , foi definido como *outlier* qualquer valor de corrente que fosse superior à soma da média local com o desvio-padrão local, ou inferior à média local subtraída do desvio-padrão local, ou seja, qualquer valor que ultrapasse a condição definida na eq. (10):

$$|\bar{i}_k - i_k| > \sigma_k. \quad (10)$$

A interpolação dos dados em torno do elemento k quando ele não existe também foi feita utilizando a média móvel, desde que a quantidade de dados faltantes na janela observada seja inferior a 20% do tamanho total da janela. Em caso de ainda haver dados faltantes, foi aplicada a interpolação linear, baseada na técnica *spline*. Na Figura 10 é apresentada uma curva para ilustração do procedimento da média móvel, com identificação de outliers e dados faltantes.

Figura 10 Suavização de uma curva com média móvel.

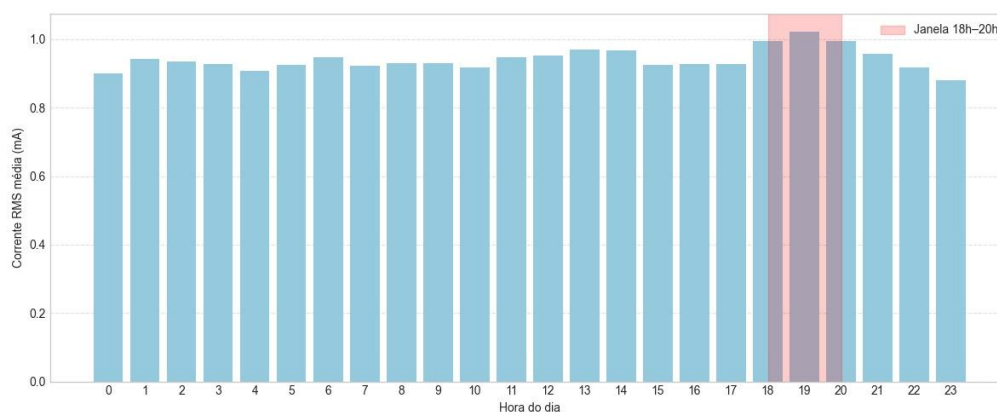


Fonte: autoria própria.

Para finalizar o processamento das curvas de corrente, foi definida uma janela de observação para representar o dia de medição. A janela em questão compreende os horários de

18h a 20h, pois é quando acontece a maior atividade na corrente de fuga devido aos picos de umidade, como pode ser visualizado na Figura 11.

Figura 11 Média global da corrente de fuga por horas do dia.



Fonte: autoria própria.

Por fim, para cada dia foram calculados os valores médios da corrente eficaz e componentes harmônicos compreendidos no intervalo de observação definido como representativo.

Os dados das 10 UMSCF foram coletados desde sua instalação na SE, em junho de 2024. No entanto, as unidades 16 e 18 apresentaram problemas de funcionamento e foram retirados para manutenção e outras análises, o que justifica suas ausências nas discussões posteriores.

Com o objetivo de ter uma quantidade de amostras que fossem minimamente representativas para cada período analisado, ficou estabelecido que seriam considerados apenas os sensores que apresentassem pelo menos 75% da quantidade total de amostras válidas esperadas para cada período. Na Tabela 2 é apresentada a representação amostral de cada sensor para cada período analisado.

Tabela 2 Quantidade percentual de dados válidos dos sensores por período de análise.

Período	S10	S11	S12	S13	S14	S15	S17	S19
1	94,7%	94,7%	94,7%	60,5%	78,9%	94,7%	21,1%	94,7%
2	98,5%	83,3%	78,8%	97,0%	83,3%	80,3%	43,9%	98,5%
3	12,1%	3,0%	97,0%	100,0%	3,0%	87,9%	60,6%	3,0%
4	0,0%	0,0%	98,9%	94,4%	0,0%	58,3%	100,0%	0,0%

Fonte: autoria própria.

A partir da análise da Tabela 2, observa-se que os Períodos 3 e 4 têm pouca representação quanto ao número de sensores disponíveis, embora ainda sejam considerados neste trabalho. Observa-se também que entre o Período 2 e o Período 3, a quantidade de dados dos sensores 10, 11, 14 e 19 reduzem de maneira significativa e, por esse motivo, o período de estiagem compreendido entre agosto e novembro foi separado em dois. Nota-se também que apenas o Sensor 12 atende ao critério de representação em todos os períodos analisados e, por este motivo, optou-se por não avaliar os sensores individualmente neste trabalho. Assim, com base no critério de representação adotado, os sensores considerados para cada período analisado foram:

- Período 1 - Sensores 10, 11, 12, 14, 15 e 19;
- Período 2 - Sensores 10, 11, 12, 13, 14, 15 e 19;
- Período 3 - Sensores 12, 13 e 15;
- Período 4 - Sensores 12, 13 e 17.

Na próxima etapa busca-se investigar a relação entre as variáveis obtidas, o seu comportamento ao longo do tempo e as interações entre si.

4.3 ANÁLISE EXPLORATÓRIA DOS DADOS

A exploração dos dados foi realizada em duas etapas: a análise temporal, buscando compreender a evolução das variáveis ao longo dos dias; e a análise por correlação, para verificar como as variáveis se relacionam entre si.

Cada sensor está instalado em uma coluna de isolamento diferente e em locais diferentes na SE. Portanto, é esperado que as magnitudes de corrente também sejam diferentes para cada sensor, apesar da tensão aplicada em todas as colunas ser a mesma. Dessa forma, a comparação direta entre valores de amplitude não é trivial.

A fim de comparar as curvas buscando observar tendências de crescimento e comportamento dinâmico, faz-se necessário normalizar os dados. A técnica de padronização selecionada para realizar este trabalho foi a padronização por z-score, onde a variável X será transformada na nova variável Z a partir da eq. (11):

$$Z = \frac{X - \mu}{o}, \quad (11)$$

sendo μ e o os parâmetros da variável X : a média e o desvio-padrão, respectivamente. Esta normalização transforma o sinal original em um sinal com média nula e desvio padrão unitário.

4.3.1 ANÁLISE TEMPORAL

Idealmente, cada período de estiagem previamente definido deveria conter 10 curvas de corrente representando os dados de cada sensor. Contudo, para cada variável de interesse (corrente eficaz, componentes harmônicas, distorção harmônica etc.), podem haver até 720 combinações entre pares de curvas por período de estiagem, o que torna volumosa a análise exploratória. Diante disso, busca-se obter uma curva média que consiga representar o comportamento global dos sensores no período de estiagem analisado.

Conforme DATTA *et. al* (2020), obter curvas globais a partir de médias de distâncias ponto-a-ponto, embora computacionalmente simples, são sensíveis a ruídos e desalinhamentos temporais, além de não conseguirem lidar com variações locais no tempo e na amplitude.

Em virtude disso, utilizou-se nesse trabalho a técnica DBA (*DTW Barycenter Averaging*), conforme exposto na Subseção 2.2, para obter as curvas de corrente globais. A curva média global da corrente (I_{ref}) será utilizada como sinal de referência para o cálculo da curva residual (ΔI). Esta curva representa a diferença entre a curva original normalizada e a curva modelo, como na eq. (12):

$$\Delta I = I - I_{ref}. \quad (12)$$

Além disso, uma segunda forma de comparação entre os dados de corrente e a curva modelo é realizada por meio da inclinação das curvas quando ajustadas em um modelo linear. Dessa forma é possível identificar quais curvas tiveram maior tendência de crescimento em relação à média global.

4.3.2 ANÁLISE POR CORRELAÇÃO

Nesta etapa do procedimento objetiva-se analisar a relação de interdependência entre as grandezas disponíveis, são elas:

1. Grandezas elétricas: corrente eficaz, componente harmônica fundamental e componentes harmônicas de terceira, quinta e sétima ordem;
2. Grandezas meteorológicas: temperatura média, amplitude térmica, precipitação e precipitação acumulada.

Neste trabalho foram avaliados três tipos de coeficientes de correlação: o coeficiente de correlação de Pearson r_p definido na eq. (13), o coeficiente de correlação de Spearman p_s definido na eq. (14), e o coeficiente de correlação de Kendall τ_k definido na eq. (15):

$$r_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \cdot \left[\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \right]}, \quad (13)$$

$$p_s = 1 - \frac{6 \sum_{i=1}^n (z - k)^2}{n(n^2 - 1)}, \quad (14)$$

$$\tau_k = \frac{2(n_c - n_d)}{n(n - 1)}. \quad (15)$$

A interpretação destes resultados é considerada:

- Positiva - quando as duas variáveis aumentam juntas;
- Negativa - indica que à medida que uma variável aumenta, a outra diminui; ou
- Nula - quando não existe interdependência estatística entre as variáveis.

Adicionalmente, nesta análise busca-se medir a força ou o grau de associação linear ou não linear entre duas variáveis. Para isso, será utilizado o coeficiente de correlação entre duas variáveis (C_{CC}), que pode assumir resultados entre -1, indicando correlação negativa perfeita, e 1 se for positiva e perfeita (MORETTIN, 2023; MARTINS, G. A.; DOMINGUES, 2017) apud (HENRIQUE, 2025). A classificação será feita de acordo com a Tabela 3.

Tabela 3 Classificação do coeficiente de correlação entre duas variáveis.

Graus	Classificação
$0,8 \leq C_{CC} \leq 1,0$	Muito Forte
$0,6 \leq C_{CC} < 0,8$	Forte
$0,4 \leq C_{CC} < 0,6$	Moderada
$0,2 \leq C_{CC} < 0,4$	Fraca
$0,0 \leq C_{CC} < 0,2$	Inexistente

Fonte: adaptado de (MORETTIN, 2023; MARTINS, G. A.; DOMINGUES, 2017) apud (HENRIQUE, 2025).

4.4 IDENTIFICAÇÃO DE PADRÕES

A partir da análise exploratória foi possível identificar alguns padrões iniciais e relações entre as grandezas, mas a complexidade do fenômeno aumenta com a quantidade de variáveis envolvidas na análise. Com o objetivo de sintetizar as informações disponíveis e agrupar comportamentos semelhantes, foram empregadas as técnicas *Principal Component Analysis* (PCA) e a análise por agrupamentos (*Clustering*).

4.4.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

A princípio idealiza-se reduzir a quantidade de dimensões para, no máximo, 3 componentes com uma representação acima de 75%. Para uma avaliação inicial da viabilidade dessa redução, todas as variáveis disponíveis foram selecionadas para compor o conjunto de entradas do PCA. Em seguida, considerando os resultados da análise exploratória e as observações nas tentativas de representação por PCA com combinações diferentes das grandezas disponíveis, ficou estabelecido que o conjunto de variáveis seria composto por:

- Corrente eficaz (I_{rms}) pós processamento da média móvel e interpolação, visto que é a principal grandeza de interesse na análise;
- Valor absoluto da curva residual da corrente eficaz (ΔI), para verificar a influência da variação da curva de cada sensor em torno da média global;
- Indicador de distorção harmônica (DH), calculado a partir da eq. (16):

$$DH = \sqrt{\left(\frac{I_{h3}}{I_{h1}}\right)^2 + \left(\frac{I_{h5}}{I_{h1}}\right)^2 + \left(\frac{I_{h7}}{I_{h1}}\right)^2}, \quad (16)$$

utilizando apenas as componentes harmônicas de 3ª, 5ª e 7ª ordens;

- Temperatura Média (T_{med});
- Precipitação Acumulada (P_{AC}).

Antes de aplicar o PCA, todas as variáveis devem estar normalizadas com média nula e desvio padrão unitário. Em seguida, aplica-se o PCA e retém os componentes de maior importância. Com esta composição de variáveis de entrada foi possível representar aproximadamente 66,3% da variância total utilizando apenas duas componentes principais.

4.4.2 ANÁLISE POR AGRUPAMENTOS (CLUSTERING)

Para mapear o resultado do PCA em agrupamentos com características semelhantes, foi utilizado o algoritmo *K-means* sobre os dados projetados no plano PC1 x PC2. A escolha de classificação em 3 *clusters* foi definida a partir da inspeção visual dos resultados. Para a inicialização dos centroides foi adotada uma configuração inicial com vetores unitários e equidistantes, com defasagem de 120° entre eles.

Com a finalização desta etapa, conclui-se o procedimento metodológico deste trabalho. A seguir são apresentados os resultados e as interpretações.

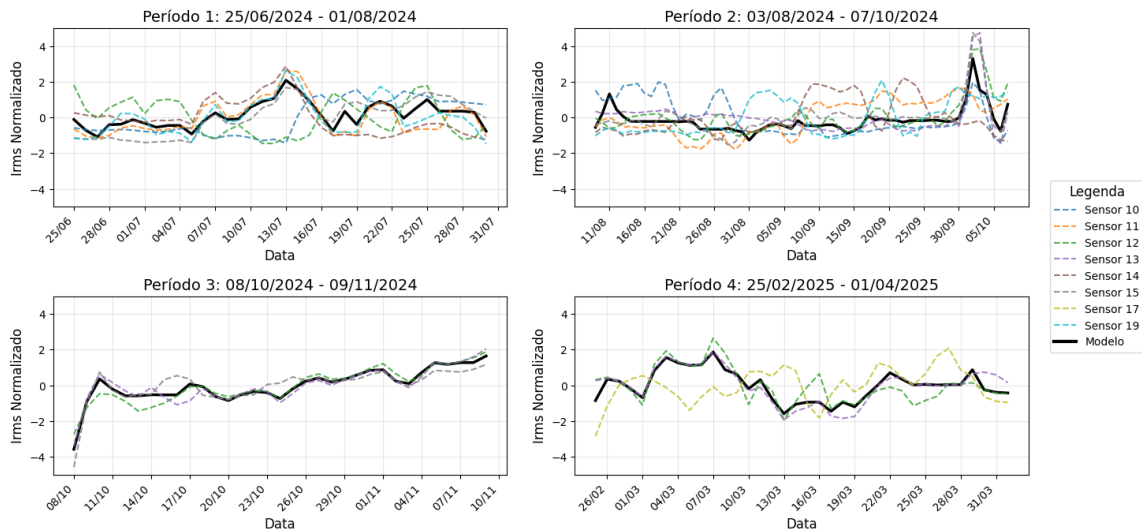
5 RESULTADOS E DISCUSSÕES

Nesta seção são apresentados os resultados obtidos a partir da observação dos dados de corrente de fuga e dados meteorológicos, conforme o procedimento metodológico discutido na Seção 4. Inicialmente, apresenta-se o resultado global referente às análises exploratórias iniciais e por PCA. Na sequência, são apresentadas as hipóteses formuladas a partir do resultado dos agrupamentos e a discussão dos principais achados.

5.1 ANÁLISE EXPLORATÓRIA

A partir da metodologia de exploração proposta na Subseção 4.3.1 foram obtidas as curvas globais de cada período de estiagem apresentadas na Figura 12.

Figura 12 Ajuste das curvas características globais obtidas por DBA.



Fonte: autoria própria.

Além disso, os resultados dos parâmetros de erro quadrado médio (*Root Mean Square Error* – RMSE) e inclinação das curvas utilizados para análise de comparação entre as curvas são apresentados na Tabela 4.

Tabela 4 Parâmetros para comparação das curvas de corrente de fuga individuais com a curva característica global.

Período:		1	2	3	4
Inclinação (β)	Referência:	0,028	0,011	0,077	-0,019
	UMSCF 10:	0,072	-0,032	-	-
	UMSCF 11:	0,017	0,042	-	-
	UMSCF 12:	-0,020	0,029	0,093	-0,040
	UMSCF 13:	-	0,000	0,076	-0,018
	UMSCF 14:	-0,038	0,019	-	-
	UMSCF 15:	0,075	0,024	0,063	-
	UMSCF 17:	-	-	-	0,031
	UMSCF 19:	0,020	0,032	-	-
RMSE (ε^2)	UMSCF 10:	1,158	0,977	-	-
	UMSCF 11:	0,578	1,001	-	-
	UMSCF 12:	1,447	0,660	0,340	0,578
	UMSCF 13:	-	0,683	0,265	0,399
	UMSCF 14:	0,922	1,292	-	-
	UMSCF 15:	0,631	0,619	0,429	-
	UMSCF 17:	-	-	-	1,289
	UMSCF 19:	0,522	1,061	-	-

Fonte: autoria própria.

A partir da análise das curvas características globais de cada período, nota-se uma tendência de crescimento da corrente nos três primeiros períodos analisados, com destaque para o Período 3 que apresenta maior inclinação. Vale destacar que o Período 3 continua imediatamente após o Período 2, sem ocorrência de chuva, visto que a separação dos períodos de análise foi motivada pela ausência de dados de alguns sensores. Além disso, observa-se que a curva característica do Período 4 apresenta uma oscilação completa com um pico entre os dias 20 de fevereiro e 10 de março, e um vale entre 10 de março e 22 de março, impactando no resultado da inclinação do modelo.

No Período 1, os maiores valores de RMSE estão associados às UMSCF 10, 12 e 14 e, com exceção das unidades 12 e 14, todos apresentam inclinação positiva. A curva da UMSCF 14 acompanha a curva característica, mas inicia acima dela e, após o pico citado, a curva de corrente passa a ficar abaixo até o final do período, refletindo no resultado de sua inclinação. Nota-se também que as curvas das UMSCF 10 e 12 não acompanham a curva global, principalmente entre os dias 10 e 16 de julho, em que a corrente de todos os outros sensores apresenta um pico. Buscando investigar este evento, verificou-se que a precipitação nos dias 12 e 16 foi superior a 2,5 mm, que pode estar associada ao aumento da corrente neste intervalo.

No Período 2, apenas a UMSCF 10 não acompanha a tendência de inclinação da curva global e, desta vez, os maiores valores de RMSE estão associados às UMSCF 10, 11, 14 e 19. As curvas de corrente destes sensores também apresentam comportamentos muito distintos entre si, que é refletido no ajuste da curva característica global. Neste período nota-se mais uma vez a ocorrência de um pico, identificado nos dias seguintes ao dia 30 de setembro. De acordo com os dados meteorológicos, não houve precipitação significativa entre 30 de setembro e 5 de outubro, entretanto, no dia 25 de setembro a precipitação total foi superior a 4 mm, e superior a 2 mm no dia seguinte. Neste caso também se sugere uma relação entre esses eventos, tendo em vista a proximidade dos dias em que houve maior precipitação, e o ajuste da curva modelo não ter sido tão representativo.

No Período 3 o ajuste da curva característica foi satisfatório, com inclinação indicando crescimento. Considerando que o Período 3 é a parte final do maior período de estiagem do ano observado, sendo ele uma continuação direta do Período 2 sem ocorrência de chuva, o resultado indicando aumento da corrente era esperado.

No Período 4, a curva da UMSCF 17 é a que mais apresenta distorção em relação à curva característica, embora não pareça ter causado muito impacto no resultado do ajuste, visto que as outras duas curvas têm perfis semelhantes à global obtida. Para este conjunto com poucas curvas, um ajuste por média ponto a ponto poderia ter mascarado essa diferença. Investigando o perfil oscilatório citado anteriormente, não foi detectado nenhum evento de precipitação considerável na janela de tempo correspondente à oscilação. Embora tenham-se identificado duas ocorrências de precipitação superiores a 2 mm no período observado: uma no dia 22 de março e outra no dia 29 de março, esta última com precipitação próxima a 4 mm, mas desta vez sem grandes variações na curva global.

A análise de dias específicos com base nas curvas de corrente e nos dados de precipitação sugere uma possível relação entre esses fatores. A seguir são apresentadas os

resultados e discussões acerca da exploração dos dados por correlação conforme descrito na Subseção 4.3.2. Na Tabela 5 são apresentados os graus de correlação obtidos.

Tabela 5 Matriz dos coeficientes de correlação de Pearson, Spearman e Kendall das variáveis analisadas.

	I_{rms}	$I_{h,1}$	$I_{h,3}$	$I_{h,5}$	$I_{h,7}$	T_{med}	ΔT	P_{AC}
I_{rms}	1	0,987	0,675	0,639	0,477	0,139	0,014	-0,004
		0,985	0,660	0,700	0,420	0,144	0,020	0,004
		0,910	0,481	0,505	0,301	0,101	0,013	0,004
$I_{h,1}$	1		0,584	0,555	0,376	0,185	0,032	-0,028
			0,596	0,651	0,355	0,163	0,022	-0,010
			0,418	0,455	0,242	0,115	0,014	-0,008
$I_{h,3}$	1			0,929	0,900	-0,244	-0,154	0,172
				0,897	0,874	-0,260	-0,178	0,177
				0,741	0,699	-0,174	-0,119	0,119
$I_{h,5}$	1				0,880	-0,210	-0,115	0,147
					0,784	-0,303	-0,129	0,087
					0,599	-0,205	-0,085	0,057
$I_{h,7}$	1					-0,249	-0,155	0,181
						-0,393	-0,237	0,232
						-0,272	-0,160	0,160
T_{med}	1						0,014	0,004
							-0,086	0,014
							-0,090	0,006
ΔT	1							-0,230
								-0,265
								-0,172
P_{AC}								1

Fonte: autoria própria.

Identificou-se que as características elétricas extraídas da forma de onda da corrente de fuga apresentam correlação positiva entre si, com diminuição do grau de correlação à medida que aumenta a ordem do harmônico. Assim, o par com maior correlação é aquele formado pela corrente eficaz e a componente harmônica fundamental, como era esperado. Por outro lado, o par com o grau de correlação mais fraca é composto pela componente harmônica fundamental e a de sétima ordem.

As componentes harmônicas ímpares, quando comparadas entre si, seguem o padrão semelhante quanto à diminuição da correlação com a ordem, embora com menor intensidade. O grau de correlação entre elas não foi classificado abaixo de forte em nenhum dos parâmetros utilizados.

Quanto às variáveis meteorológicas, todas as combinações entre elas apresentaram graus de correlação classificados como inexistentes, com exceção do par formado entre amplitude térmica e precipitação acumulada, que apresentou correlação negativa fraca.

Na comparação entre grandezas elétricas e meteorológicas, observou-se que a temperatura média apresenta correlação negativa fraca com todas as componentes harmônicas ímpares analisadas.

Na sequência será discutido o resultado obtido conforme o procedimento metodológico da Subseção 4.4, que trata da identificação de padrões a partir da análise das componentes principais (PCA) e análise por agrupamentos (*clustering*).

Com o objetivo de reduzir a dimensionalidade e buscar padrões que não foram identificados anteriormente, aplicou-se a técnica de PCA ao conjunto total de dados, conforme descrito na Subseção 4.4.1. Os resultados das contribuições de cada variável analisada estão apresentados na Tabela 6.

Tabela 6 Contribuição das variáveis na representação do PCA.

Variável	PC1	PC2
I_{rms}	-0,311	-0,194
ΔI	0,058	-0,163
DH	0,827	0,427
T_{med}	-0,465	0,868
P_{AC}	-0,010	-0,028

Fonte: autoria própria.

A partir deste resultado, observa-se que a componente principal 1 (PC1) aumenta significativamente com o acréscimo do índice de distorção harmônica (DH), enquanto as contribuições da corrente eficaz (I_{rms}) e da temperatura média diária (T_{med}) atuam no sentido de diminuir esta componente. Quanto à componente principal 2 (PC2), a maior influência positiva é exercida pela temperatura, seguida pela distorção harmônica, enquanto a corrente eficaz e a variação em torno da tendência global da corrente (ΔI) apontam na direção oposta.

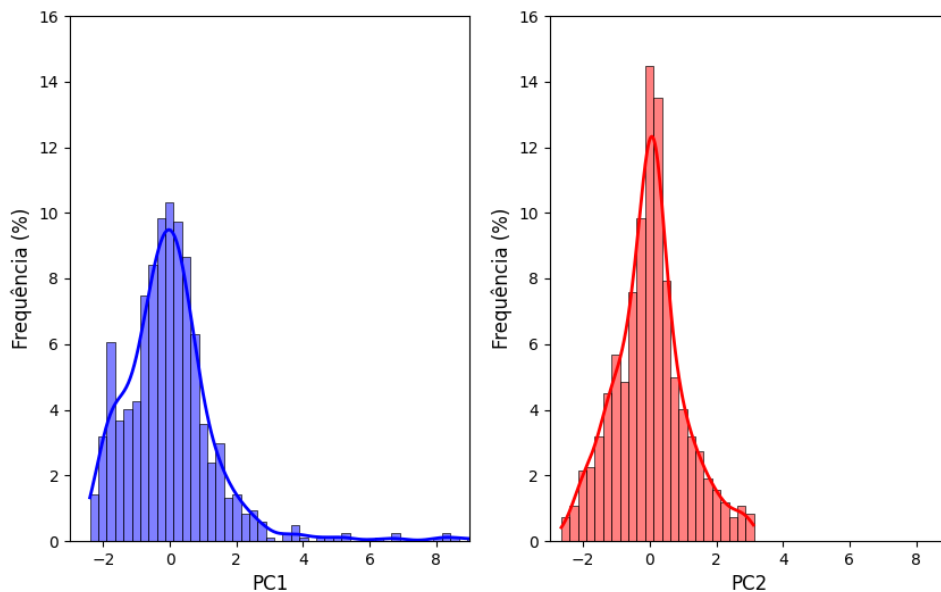
Nesse contexto, a composição dos autovetores indica a possibilidade de um agrupamento de dados sob influência de maiores temperaturas apontando no sentido do segundo quadrante. Do mesmo modo, espera-se que os dados com maiores índices de distorção tendam a se distribuir pelo primeiro quadrante.

Como os autovetores associados a I_{rms} e ΔI apresentam as menores magnitudes e contribuições semelhantes nas componentes principais, adota-se uma combinação linear entre estes dois autovetores para a interpretação dos resultados. Dessa forma, espera-se que o aumento da corrente reflita num arranjo de dados apontando para o terceiro quadrante.

A variável de precipitação apresentou pouca correlação com as demais grandezas e contribuiu de forma pouco significativa com a representação via PCA. A hipótese sugerida é que, por se tratarem de períodos de estiagem, a influência dessas variáveis com o fenômeno é pouco presente. Por este motivo a precipitação não será considerada na etapa da identificação de padrões.

Na Figura 13 apresenta-se o histograma da distribuição das componentes principais.

Figura 13 Distribuição das componentes principais (PC1 e PC2).



Fonte: autoria própria.

Nota-se que o histograma de PC1 foi limitado à esquerda, tendo em vista que os dados reais de entrada são limitados a valores positivos. Além disso, é possível observar a formação de uma cauda a partir do ponto 3, indicando a ocorrência de eventos infrequentes que podem ser caracterizados como *outliers*.

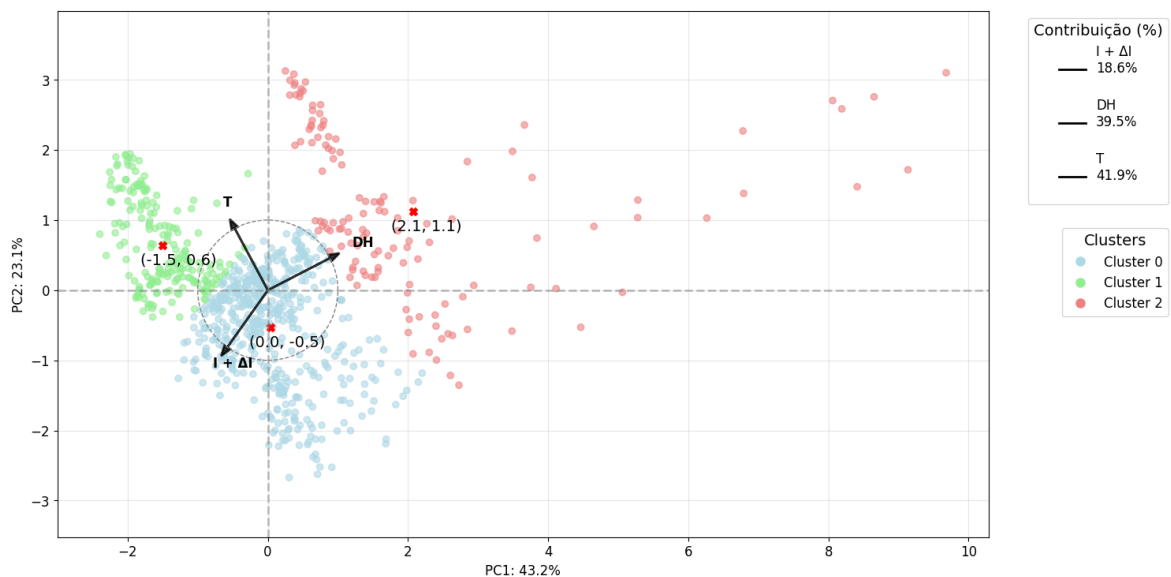
De forma semelhante, observa-se também uma limitação nos dois extremos do histograma de PC2, que pode ser atribuído à baixa variabilidade da temperatura média diária no ambiente analisado. Neste caso, aumentar a amostragem para valores horários poderia apresentar mais detalhes nesse resultado.

Em seguida, são apresentados os resultados e as discussões referentes à análise por agrupamentos conforme descrito na Subseção 4.4.2.

5.2 ANÁLISE POR AGRUPAMENTOS

Na Figura 14 apresenta-se o gráfico de correlação entre as duas componentes principais, contendo a orientação dos autovetores das variáveis analisadas, bem como o resultado do agrupamento por *clusters*.

Figura 14 Gráfico de correlação entre as componentes principais do PCA.



Fonte; autoria própria.

A partir da análise do gráfico, nota-se que o cluster 0 tem maior concentração de pontos em torno da origem, indicando equilíbrio entre as ponderações dos dados pelos autovetores das variáveis analisadas. No cluster 1, por sua vez, nota-se uma correlação negativa entre as componentes principais, interpretada pela influência da temperatura no sinal de corrente de fuga. Já o cluster 2 apresenta a maior dispersão entre os dados, sendo associado à cauda citada no histograma do PC1 da Figura 13. Nota-se também que, apesar do cluster 2 parecer indicar

uma relação positiva entre as componentes principais, a relação entre as variáveis de corrente e distorção parece ser negativa, com os vetores apontando em sentidos opostos.

Para verificar as características individuais de cada agrupamento a partir dos dados medidos, apresenta-se as informações de cada variável de interesse por cluster na Tabela 7, são elas: corrente eficaz média, valor de pico da terceira harmônica, distorção harmônica média, temperatura média, precipitação média e quantidade de ocorrências de precipitação acima de 2,5 mm. Este resultado foi obtido a partir dos dados em escalas reais, sem normalização.

Tabela 7 características individuais de cada agrupamento.

<i>Cluster</i>	I_{rms}	$I_{h,3}$	DH	T_{med}	P	$P \geq 2,5 \text{ mm}$
0	0,61 mA	0,28 mA	0,59	25,5 °C	0,48 mm	32
1	1,09 mA	0,18 mA	0,14	27,3 °C	0,31 mm	4
2	0,14 mA	0,13 mA	2,38	26,0 °C	0,56 mm	9

Fonte: autoria própria.

A análise dos resultados apresentados na Tabela 7 revela que as características meteorológicas do cluster 0 indicam uma associação ao clima de menor temperatura e maior ocorrência de precipitação moderada, que pode ser associada à maior umidade. Além disso, apesar de não ter o maior indicador de distorção harmônica, apresenta o maior valor médio de $I_{h,3}$.

No cluster 1 nota-se o maior valor de corrente e temperatura, mas com o menor indicador de distorção harmônica. Por sua vez, as características do cluster 2 apontam para correntes mais baixas e com maiores níveis de distorção harmônica. Estes resultados estão alinhados com o exposto anteriormente sobre a correlação negativa entre a corrente eficaz e a distorção harmônica, sugerindo que este indicador pode mascarar as variações de componentes harmônicas individuais com o aumento da corrente de fuga.

Nas características do cluster 2 observa-se a maior distorção harmônica, a maior precipitação média e pouca ocorrência de precipitação moderada, quando comparada aos demais agrupamentos. Entretanto, também se nota o menor valor de corrente, além de ter a distribuição mais dispersa.

A seguir será apresentada a quantidade de ocorrências de cada cluster por período analisado e sensor. Os dados estão dispostos na Tabela 8.

Tabela 8 Quantidade de ocorrências em cada *cluster*, por período e sensor.

C	P	S-10	S-11	S-12	S-13	S-14	S-15	S-17	S-19	Σ_p
C0	P1	32	31	31	-	35	11	-	14	154
	P2	55	53	38	56	56	47	-	46	351
	P3	-	-	3	2	-	2	-	-	7
	P4	-	-	0	0	-	-	0	-	0
	Σ_C	87	84	72	58	91	60	0	60	512
C1	P1	4	5	0	-	0	0	-	0	9
	P2	4	7	0	3	4	4	-	3	25
	P3	-	-	30	31	-	31	-	-	92
	P4	-	-	36	36	-	-	0	-	72
	Σ_C	8	12	66	70	4	35	0	3	198
C2	P1	0	0	5	-	1	25	-	22	53
	P2	1	0	22	1	0	9	-	11	44
	P3	-	-	0	0	-	0	-	-	0
	P4	-	-	0	0	-	-	36	-	36
	Σ_C	1	0	27	1	1	34	36	33	133
Σ_{total}		96	96	165	129	96	129	36	96	843

Fonte: autoria própria.

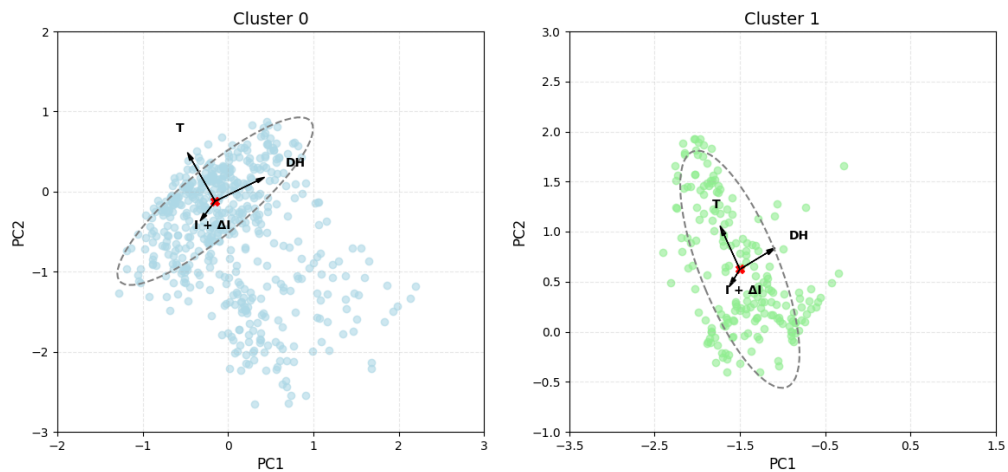
O cluster 0 concentra 60,7% das ocorrências totais dos dados, sendo predominantes nos períodos 1 e 2. Já o cluster 1 concentra 23,5% das ocorrências nos períodos 3 e 4. Essa observação indica elevação da corrente com o passar do tempo, possivelmente pelo efeito de deposição de poluição nos isolamentos.

O cluster 2 representa 15,8% das ocorrências, mas distribuídas entre os períodos 1, 2 e 4, porém fortemente associadas a sensores específicos. No Período 1, 88,7% dos dados pertencentes ao cluster 2 são provenientes dos sensores 15 e 19. No Período 2, 95,4% das ocorrências estão concentradas nos sensores 12, 15 e 19. E 100% dos dados classificados no cluster 2 correspondem ao sensor 17, no Período 4. Esses dados reforçam que o cluster 2 pode refletir particularidades das medições, representando comportamentos particulares com correntes baixas e distorções altas provenientes de sensores específicos.

Após esta análise dos resultados do PCA e da caracterização individual dos clusters, será realizada a verificação das relações entre a temperatura e a corrente eficaz, bem como entre a corrente eficaz e a distorção harmônica. O cluster 2 será desconsiderado nesta análise, sendo tratado como outlier.

Na Figura 15 são apresentados os gráficos de dispersão dos clusters 0 e 1, separadamente, com destaque para as regiões de maior densidade de dados delimitadas por elipses.

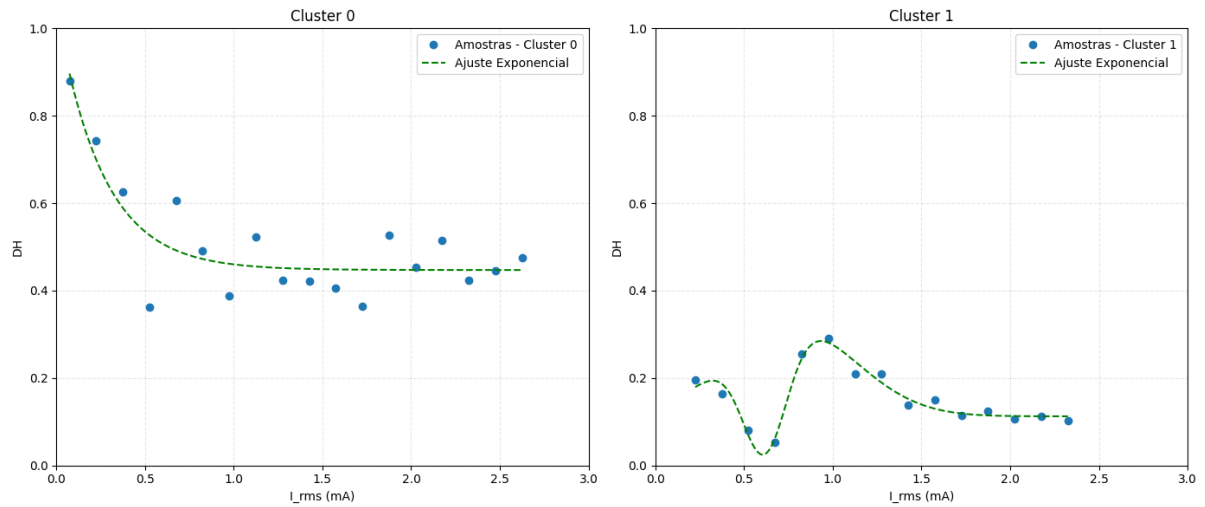
Figura 15 Gráficos de dispersão dos *clusters* individuais.



Fonte: autoria própria.

A região delimitada no *cluster* 0 indica correlação positiva entre PC1 e PC2, mas também revela uma relação negativa entre DH e I_{rms} . Para investigar essa relação, foi calculada a média de DH em intervalos de corrente ΔI . Os resultados do ajuste para os clusters 0 e 1 são apresentados na Figura 16.

Figura 16 Ajuste de curvas entre irms e dh para os clusters 0 e 1.

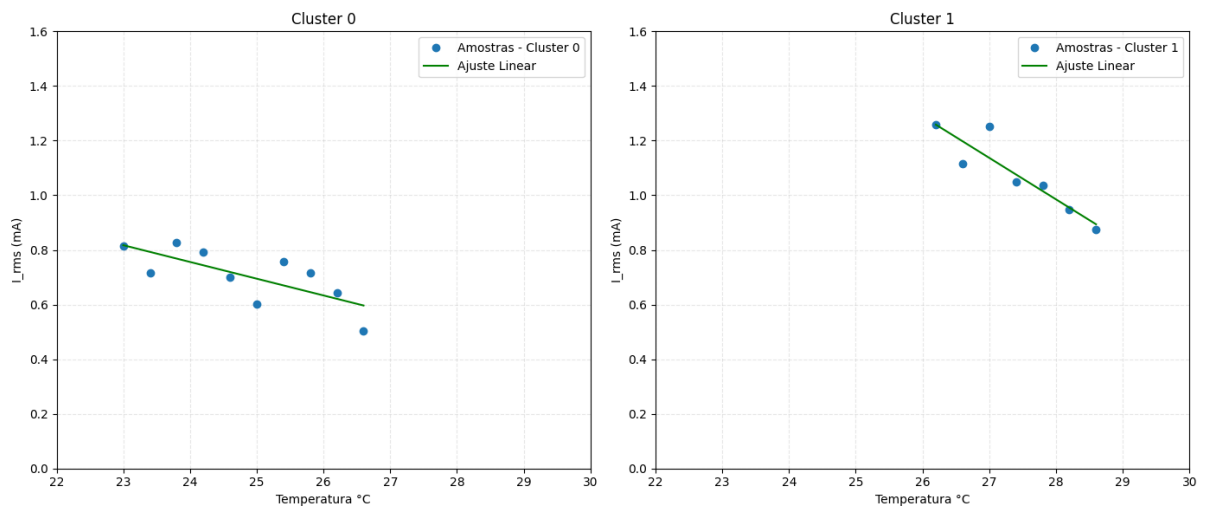


Fonte: autoria própria.

Os perfis obtidos confirmam a relação negativa entre a corrente e a distorção harmônica. No entanto, observa-se que para correntes mais baixas, o padrão difere entre os dois clusters.

A região delimitada no cluster 1 apresentada na Figura 15 indica correlação negativa entre as componentes principais, mas nenhum autovetor se opõe diretamente ao de temperatura. Por essa razão ambas as combinações serão analisadas individualmente. Na Figura 17 é apresentado o ajuste para a relação entre temperatura e corrente eficaz.

Figura 17 ajuste linear entre temperatura e corrente eficaz para os cluster 0 e 1.

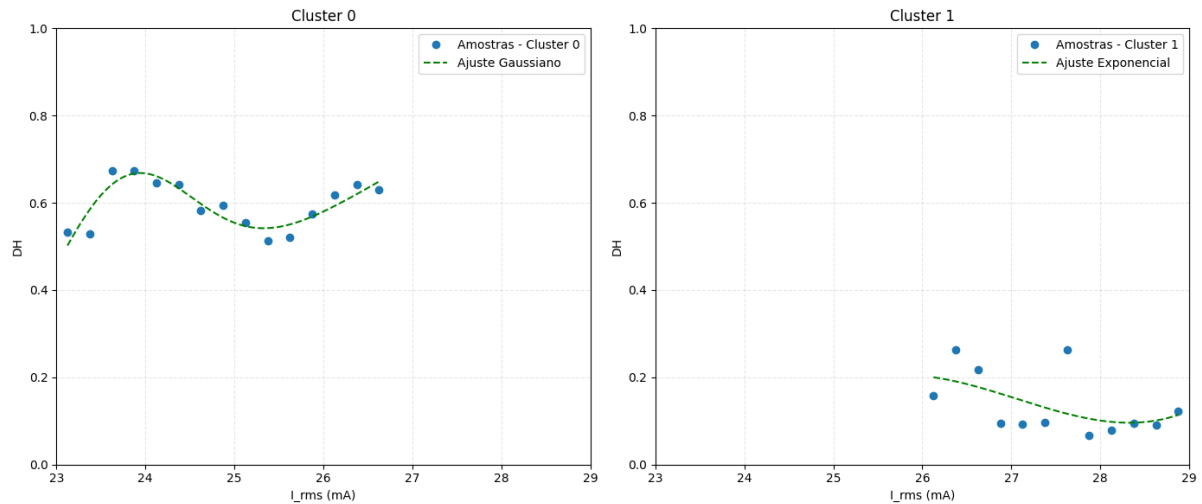


Fonte: autoria própria.

Embora os perfis identificados não sejam necessariamente lineares, observa-se uma tendência decrescente para ambos os clusters, sugerindo, por exemplo, que em dias mais frios,

a corrente de fuga tende a aumentar. Vale destacar que, no contexto do Nordeste, os dias mais frios estão associados à maiores índices de umidade. Em seguida é apresentado o resultado da análise da temperatura com a DH na Figura 18.

Figura 18 ajuste das curvas entre temperatura e distorção harmônica para os cluster 0 e 1.



Fonte: autoria própria.

Os perfis identificados nessa relação não são lineares. A curva referente ao cluster 0 não apresenta tendência clara de crescimento ou decrescimento. Já no cluster 1, embora aparente ser decrescente, os resultados ainda não apresentam uma relação conclusiva.

6 CONCLUSÃO

Neste trabalho foi apresentada uma análise da corrente de fuga de isolamentos de uma subestação de 500 kV, considerando a influência de variáveis meteorológicas e utilizando parâmetros extraídos da forma de onda, como corrente eficaz e componentes harmônicas.

A metodologia adotada, baseada na coleta e processamento de dados e na redução de dimensionalidade e análise por agrupamentos, possibilitou uma avaliação do comportamento elétrico associado à temperatura e eventos de precipitação. No entanto, modelar tais comportamentos ainda é um desafio. Informações de precipitação durante o dia permitiriam identificar se o evento se trata de uma chuva intensa em um período curto, ou uma chuva mais distribuída ao longo do dia com menor intensidade. Outras variáveis importantes neste cenário, como a umidade relativa do ar, também permitiriam avaliar melhor os comportamentos.

Os resultados evidenciaram o perfil não linear do fenômeno e se mostraram consistentes com a literatura, embora algumas limitações tenham sido identificadas, as quais sugerem necessidade de maior quantidade de dados disponíveis e informações mais detalhadas ao longo dos dias de observação.

Recomenda-se que estudos futuros ampliem a base de dados, incorporando mais variáveis como umidade, radiação ultravioleta e velocidade do vento.

REFERÊNCIAS BIBLIOGRÁFICAS

AHMED, Rizwan et al. Field-dependent pollution model under polluted environments for outdoor polymeric insulators. *Polymers*, v. 14, n. 3, p. 516, 2022.

AMARH, Felix; KARADY, George G.; SUNDARARAJAN, Raji. Linear stochastic analysis of polluted insulator leakage current. **IEEE Transactions on Power Delivery**, v. 17, n. 4, p. 1063-1069, 2002.

BANIK, Apu; DALAI, Sovan; CHATTERJEE, Biswendu. Studies the effect of Equivalent Salt Deposit Density on leakage current and flashover voltage of artificially contaminated disc insulators. In: 2015 1st Conference on Power, Dielectric and Energy Management at NERIST (ICPDEN). IEEE, 2015. p. 1-5.

BARBOSA, Vandilson Rodrigo do Nascimento. Estimador inteligente do momento ótimo de limpeza de isoladores de vidro a partir da medição da corrente de fuga. 2025. 157 f. Tese (Doutorado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Centro de Ciências e Tecnologia, Universidade Federal de Campina Grande, Paraíba, Brasil, 2025.

BASHIR, Nouruddeen; AHMAD, Hussein. Leakage current harmonic components: Indexes to insulator ageing. In: TENCON 2009-2009 IEEE Region 10 Conference. IEEE, 2009. p. 1-5.

BRADLEY, Paul S. et al. Refining initial points for k-means clustering. In: ICML. 1998. p. 91-99.

CARVALHO, Stéphane et al. Current transformer-based system for measuring leakage current. In: 2023 7th International Symposium on Instrumentation Systems, Circuits and Transducers (INSCIT). IEEE, 2023. p. 1-6.

CASTILLO-SIERRA, R. *et al.* **A novel method for prediction of washing cycles of electrical insulators in high pollution environments.** *International Journal of Electrical Power & Energy Systems*, v. 130, p. 107026, 2021.

DATTA, Shreyasi; KARMAKAR, Chandan K.; PALANISWAMI, Marimuthu. Averaging methods using dynamic time warping for time series classification. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2020. p. 2794-2798.

DOLNÍK, Bystrík et al. Sensing method using multiple quantities for diagnostic of insulators in different ambient conditions. **Sensors**, v. 22, n. 4, p. 1376, 2022.

FARAMARZI, Mousalreza et al. Automatic system for detection critical conditions in overhead lines distribution insulators based on leakage current analysis. In: **2016 IEEE/IAS 52nd Industrial and Commercial Power Systems Technical Conference (I&CPS)**. IEEE, 2016. p. 1-7.

GHOJOGH, Benyamin; KARRAY, Fakhri; CROWLEY, Mark. Principal component analysis using structural similarity index for images. In: International Conference on Image Analysis and Recognition. Cham: Springer International Publishing, 2019. p. 77-88.

HENRIQUE, Clayrton Monteiro et al. Análise multivariada de eventos e variações em sistemas de transmissão de energia elétrica com base em processos de digitalização dos sistemas de potência. 2025.

HONGYU, Kuang; SANDANIELO, Vera Lúcia Martins; DE OLIVEIRA JUNIOR, Gilmar Jorge. Análise de componentes principais: resumo teórico, aplicação e interpretação. E&S Engineering and science, v. 5, n. 1, p. 83-90, 2016.

JAYASUNDARA, H. I. S.; SUDARSHANI, W. P. S.; FERNANDO, M. A. R. M. Leakage current patterns on high voltage insulators: Analysis on frequency and time-frequency domains. In: 2008 IEEE Region 10 and the Third international Conference on Industrial and Information Systems. IEEE, 2008. p. 1-6.

JIANG, Zidan et al. Investigating the effect of rainfall parameters on the self-cleaning of polluted suspension insulators: insight from Southern China. Energies, v. 10, n. 5, p. 601, 2017.

JOLLIFFE, Ian. Principal component analysis. In: International encyclopedia of statistical science. Springer, Berlin, Heidelberg, 2011. p. 1094-1096.

MANJANG, Salama et al. Study of leakage current characteristics of high voltage insulators ageing after dry season in polluted environmental conditions. In: 2021 3rd International Conference on High Voltage Engineering and Power Systems (ICHVEPS). IEEE, 2021. p. 074-078.

MARTINS, Gilberto A.; DOMINGUES, Osmar. Estatística Geral e Aplicada. 6. ed. [S.l.]: Rio de Janeiro: Atlas, 2017. 261p.

MENDONÇA, Yorkismar de Andrade. Análise da corrente de fuga de isoladores poliméricos para classificação das condições de degradação com auxílio de lógica fuzzy. 2016. 84f. (Dissertação de Mestrado) Programa de Pós-Graduação em Engenharia Elétrica, Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande - Paraíba - Brasil, 2016.

MOHAMMADNABI, S.; RAHMANI, Kh. Influence of humidity and contamination on the leakage current of 230-kV composite insulator. Electric Power Systems Research, v. 194, p. 107083, 2021.

MONTOYA, Gerardo; RAMIREZ, Isaias; MONTOYA, J. I. Correlation among ESDD, NSDD and leakage current in distribution insulators. **IEE Proceedings-Generation, Transmission and Distribution**, v. 151, n. 3, p. 334-340, 2004.

MORAES, Luca Alves. Análise do comportamento espectral de isoladores em linhas de transmissão. 2016. 51f. (Trabalho de Conclusão de Curso - Monografia), Curso de Bacharelado em Engenharia Elétrica, Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande – Paraíba - Brasil, 2016.

MORETTIN, Pedro A. Estatística Básica. 10. ed. [S.l.]: Rio de Janeiro: Saraiva Uni, 2023. 513p.

NIENNATTRAKUL, Vit; RATANAMAHATANA, Chotirat Ann. On clustering multimedia time series data using k-means and dynamic time warping. In: 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07). IEEE, 2007. p. 733-738.

OLIVEIRA, Graziella Bedenik de. Sistema de medição de picos de corrente de fuga para isoladores de alta tensão. 2023. 60 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Sergipe, São Cristóvão, 2023.

PAKPAHAN, Parouli M. et al. Multivariate analysis for investigation of outdoor insulator conditions. In: 2008 International Conference on Condition Monitoring and Diagnosis. IEEE, 2008. p. 642-645.

PETITJEAN, François; KETTERLIN, Alain; GANÇARSKI, Pierre. A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition*, v. 44, n. 3, p. 678-693, 2011.

RAMIREZ, Isaias; HERNANDEZ, Ramiro; MONTOYA, Gerardo. Measurement of leakage current for monitoring the performance of outdoor insulators in polluted environments. *IEEE Electrical Insulation Magazine*, v. 28, n. 4, p. 29-34, 2012.

SALEM, Ali Ahmed et al. Pollution flashover characteristics of coated insulators under different profiles of coating damage. **Coatings**, v. 11, n. 10, p. 1194, 2021.

SALEM, Ali Ahmed et al. Pollution flashover voltage of transmission line insulators: Systematic review of experimental works. **IEEE Access**, v. 10, p. 10416-10444, 2022.

ZHENGFA, Liu et al. Study on leakage current characteristics and influence factors of 110kV polluted composite insulators. In: 2018 12th International Conference on the Properties and Applications of Dielectric Materials (ICPADM). IEEE, 2018. p. 896-900.