57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

Impacto da Vegetação Arbórea no Fornecimento de Energia Elétrica

Andréia S. Santos, Lucas Teles Faria, Leonardo H. Macedo
Universidade Estadual Paulista (UNESP), Faculdade de Engenharia e Ciências
Av. dos Barrageiros, 1881, Centro, Rosana, 19272-100, São Paulo, Brasil
andreia.faria@unesp.br, lucas.teles@unesp.br, leohfmp@ieee.org

RESUMO

A proximidade da vegetação arbórea junto à rede aérea de distribuição é um dos fatores associados à ocorrência de falhas. As interrupções provocadas por esse tipo de interferência impactam negativamente os indicadores de confiabilidade e de qualidade de energia elétrica. Nesse contexto, este estudo aplica técnicas de análise exploratória de dados geográficos para avaliar a associação entre vegetação arbórea e as faltas ocasionadas por desligamentos em transformadores da rede de distribuição. Como resultado, são produzidos mapas temáticos que indicam as regiões da cidade onde há maior concentração de faltas. Adicionalmente, o algoritmo de agrupamento k-means é aplicado para identificar padrões espaciais via formação de *clusters* de subáreas. Este estudo fornece subsídios para o planejamento de ações preventivas pelas distribuidoras de energia, com destaque para as áreas mais vulneráveis às faltas devido à proximidade da rede elétrica com a vegetação arbórea.

PALAVRAS CHAVE. Análise exploratória de dados geográficos. Faltas. Vegetação arbórea.

AS&DS – Análise e Ciência de Dados

ABSTRACT

The proximity of trees to overhead power lines is one of the factors associated with power grid failures. The interruptions caused by this interference affect the reliability and quality indicators of the energy supply. This study applies exploratory spatial data analysis techniques to assess the association between trees and faults caused by shutdowns in distribution transformers. As a result, thematic maps are produced that indicate the city's regions with a higher concentration of faults. In addition, the k-means clustering algorithm is applied to identify spatial patterns. This study supports the planning of preventive actions by power utilities, highlighting the areas most vulnerable to faults due to their proximity to trees.

KEYWORDS. Exploratory spatial data analysis. Faults. Tree vegetation.

AS&DS – Data Science and Analytics

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

1. Introdução

Os sistemas de distribuição de energia elétrica (SDEE) estão sujeitos a diversas perturbações que podem ocasionar interrupções no fornecimento de energia elétrica. Essas interrupções, resultantes de múltiplos fatores, geram impactos negativos para as distribuidoras de energia, incluindo prejuízos financeiros e a degradação dos índices de confiabilidade e de qualidade da energia elétrica [Santos et al., 2025; Melagoda et al., 2021].

Os principais fatores responsáveis por aproximadamente 30% dos casos de interrupções no fornecimento de energia elétrica incluem a proximidade da vegetação arbórea com a rede elétrica, falhas em equipamentos e condições climáticas adversas, tais como: descargas atmosféricas, chuvas torrenciais, tempestades e rajadas de vento [ONS, 2023]. Dentre esses fatores, a vegetação arbórea representa um desafio para as distribuidoras de energia, principalmente quando se encontra nas proximidades das linhas aéreas de distribuição, sendo um elemento que aumenta o risco de perturbações na rede elétrica [Santos et al., 2024].

Durante tempestades com a presença de rajadas de vento, os galhos das árvores podem entrar em contato com a rede elétrica, provocando perturbações no sistema elétrico, ou até mesmo ocasionar o rompimento dos cabos ao caírem sobre eles, resultando em interrupções no fornecimento de energia. A sobreposição de galhos sobre os cabos das linhas aéreas pode resultar em interferências, como curtos-circuitos, falhas operacionais e danos à infraestrutura da rede elétrica. Adicionalmente, a exposição de cabos rompidos representa sérios riscos à segurança da população e acarreta prejuízos econômicos para as distribuidoras de energia [Doostan et al., 2020; Melagoda et al., 2021].

Na literatura especializada, há diversos estudos que avaliaram a associação entre vegetação arbórea e faltas no SDEE. Chen et al. [2015] propuseram um método preditivo baseado em uma estrutura de sistema de informação geográfica (SIG) para o gerenciamento de interrupções em sistemas de transmissão e distribuição, considerando variáveis climáticas, intensidade dos ventos e a presença de vegetação arbórea. Santos et al. [2024] realizaram uma análise exploratória ponderada geograficamente para investigar a correlação local envolvendo faltas ocasionadas por desligamentos em transformadores e a presença de vegetação arbórea. A fim de quantificar as relações espaciais locais, foram aplicadas métricas ponderadas geograficamente, tais como: média, desvio padrão e o coeficiente de correlação de Pearson.

Wedagedara et al. [2023] abordaram o risco associado a variáveis locais que podem ocasionar interrupções devido à interferência da vegetação durante tempestades. Cinco modelos de risco de falhas provocadas por vegetação arbórea foram desenvolvidos e comparados utilizando o algoritmo *random forest*. Parent et al. [2019] avaliaram a eficácia de um programa aprimorado de poda de árvores na redução das interrupções de energia relacionadas à vegetação no SDEE. Doostan et al. [2020] propuseram uma abordagem para a previsão de interrupções relacionadas à vegetação, combinando modelos estatísticos baseados em séries temporais com técnicas de aprendizado de máquina não linear. Gallaher et al. [2021] investigaram o impacto das operações de poda de árvores na resiliência das redes elétricas, com foco na frequência, extensão e duração das interrupções provocadas por vegetação.

A influência da vegetação arbórea sobre a rede elétrica está diretamente relacionada às características geográficas das regiões onde se insere a infraestrutura da rede elétrica. Nesse contexto, este estudo adota uma abordagem envolvendo análise exploratória de dados geográficos (AEDG), a qual permite revelar a estrutura espacial das variáveis envolvidas e compreender os padrões subjacentes aos dados.

Como resultado da AEDG, são elaborados mapas temáticos que apresentam a distribuição espacial das variáveis em estudo. Adicionalmente, aplica-se o algoritmo k-means para identificar

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

agrupamentos ou *clusters* espaciais das subáreas mais vulneráveis às interrupções. Essa abordagem pode contribuir para a identificação e compreensão da distribuição espacial das falhas, auxiliando na detecção de padrões, bem como na delimitação de áreas críticas da rede elétrica que demandam maior atenção da distribuidora de energia para o planejamento de ações preventivas e corretivas.

2. Análise exploratória de dados geográficos

A AEDG consiste de um conjunto de técnicas voltadas à descrição, visualização e avaliação de informações georreferenciadas, permitindo revelar estruturas espaciais, padrões de distribuições e possíveis anomalias nos dados [Le Gallo e Ertur, 2003]. Essa abordagem incorpora diversas métricas da estatística descritiva espacial que possibilitam a identificação de *clusters*, *outliers* e a avaliação da autocorrelação espacial, além de fornecer subsídios para a formulação de hipóteses sobre a disposição espacial dos fenômenos observados [Monteiro et al., 2004; Haining, 2003]. Entre os principais instrumentos utilizados estão mapas temáticos, estatísticas locais de associação espacial, como o *índice de Moran* e estimativas de densidade, como o *mapa de kernel* ou mapa de calor. Essas ferramentas são essenciais para orientar as etapas posteriores de modelagem espacial e apoiar o planejamento e a tomada de decisão em diversas áreas do conhecimento.

2.1. Mapas temáticos

Os mapas temáticos ou mapas de calor são uma representação visual que utiliza variações de cor para indicar a densidade ou intensidade de um determinado fenômeno em uma área geográfica. Essa técnica é comumente aplicada para destacar áreas com maior concentração do evento em estudo, como as interrupções no fornecimento de energia elétrica.

Os mapas de calor podem ser classificados em duas categorias: mapa de calor espacial e em grade [Gu, 2022]. Mapas de calor são frequentemente construídos com base em métodos de interpolação espacial, como a estimativa de densidade por *kernel* ou *kernel density estimation* (KDE), que calcula a densidade de pontos em uma malha espacial contínua. Esse método permite suavizar a distribuição dos dados. Dessa forma, áreas de alta concentração ou *hotspots* e agrupamentos espaciais podem ser identificados.

A densidade é calculada com base no número de pontos dentro de uma determinada área e na distância entre eles. Quando há grande concentração de pontos próximos entre si, o valor de densidade estimada é maior, resultando em cores mais intensas no mapa. Por outro lado, áreas com poucos ou nenhum ponto são representadas com cores mais suaves ou ausentes, o que facilita a interpretação visual dos padrões espaciais e a identificação de regiões críticas [Santos et al., 2023].

2.2. Mapa de desvio padrão

O mapa de desvio padrão é um tipo de mapa temático utilizado para destacar valores extremos em um conjunto de dados espaciais. Nesse tipo de representação, a variável de interesse é transformada em unidades de desvio padrão, ou seja, os valores são padronizados por meio da padronização z, resultando em uma distribuição com média nula e desvio padrão unitário [Luc Anselin, 2020].

Essa abordagem permite identificar visualmente valores que se desviam significativamente da média, facilitando a detecção de *outliers*. A classificação dos dados no mapa depende do intervalo total de valores, determinado pelo número de unidades de desvio padrão que separam os valores extremos. Em geral, são utilizadas categorias que representam faixas específicas de desvios padrão. Observa-se que frequentemente podem existir faixas sem nenhuma observação, especialmente quando os dados estão concentrados em torno da média [Luc Anselin, 2020; Santos et al., 2023].

2.3. Clusterização de subáreas

Neste estudo, aplica-se o algoritmo k-means para agrupamento de subáreas ou setores censitários (SCs) de acordo com o número anual de faltas ocasionadas por interferência de vegetação arbórea nas linhas de distribuição. O algoritmo k-means permite identificar áreas com

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

comportamentos similares ao longo do tempo. Ele consiste em uma técnica de agrupamento não supervisionado que particiona um conjunto de dados em k grupos (*clusters*) distintos, com base nas características dos dados [Fränti e Sieranoja, 2018]. Assim, as observações dentro de um mesmo grupo são mais semelhantes entre si em comparação aos outros grupos.

A métrica de similaridade utilizada neste estudo é a distância Euclidiana, calculada entre cada ponto de dado (faltas) e o centroide do *cluster* correspondente (k). O algoritmo k-means busca iterativamente uma partição ótima dos dados em k grupos, minimizando a soma dos quadrados das distâncias no interior de cada *cluster*, ou seja, a variabilidade dentro de cada grupo é reduzida ao mover pontos de um *cluster* para outro até que se atinja a convergência [Hartigan e Wong, 1979].

A função objetivo do algoritmo apresentada em (1) busca minimizar a variabilidade interna dos *clusters* ou a dissimilaridade entre os pontos do mesmo grupo [Anselin, 2024].

$$\text{minimizar} \quad \frac{1}{2}\sum_{h=1}^{k}\sum_{i\in h}\sum_{j\in h}\|x_i - x_j\|^2. \tag{1}$$

Para cada *cluster* C_h somam-se todas as distâncias ao quadrado entre os pares de observações i e j que pertencem ao mesmo grupo. O objetivo é minimizar essa soma, garantindo que os elementos dentro de cada *cluster* sejam o mais semelhantes possível.

A expressão em (1) pode ser reformulada usando a distância entre os pontos e o centroide do seu respectivo *cluster* [Anselin, 2024]:

$$\text{minimizar} \quad \sum_{h=1}^{k} n_h \sum_{i\in h}(x_i - \bar{x}_h)^2, \tag{2}$$

onde n_h é o número de observações no *cluster* h; $\bar{x}_h$ é o vetor centroide (média) do *cluster* h.

Neste contexto, as variáveis analisadas correspondem ao número de faltas por SCs ocasionadas devido à vegetação arbórea. A aplicação do algoritmo k-means permite agrupar setores com perfis semelhantes de ocorrência de faltas ao longo dos anos analisados, viabilizando a identificação de padrões espaciais e temporais de falhas na rede elétrica. Esses padrões podem dar suporte ao planejamento de estratégias de manutenção preventiva e corretiva.

3. Testes e Resultados

Nesta seção, são apresentados os resultados para uma cidade cuja área urbana analisada está subdividida em 301 SCs. Os SCs correspondem a unidades geográficas definidas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para fins de levantamento demográfico. Os arquivos vetoriais *shapefiles* disponibilizados pelo IBGE são compatíveis com todos os SIGs, o que possibilita a realização de análises espaciais detalhadas, integradas ao contexto urbano e adequadas ao mapeamento das variáveis de interesse, como vegetação arbórea e as interrupções de energia, consideradas neste estudo.

As análises e simulações foram realizadas utilizando o QGIS (versão 3.30), GeoDa (versão 1.20.0.36) e a linguagem de programação Python. Todas as simulações foram executadas em um computador com processador Intel Core i7 (1,8–2,3 GHz) e 16 GB de memória RAM.

3.1. Descrição da base de dados

Neste estudo, são analisadas as faltas associadas aos desligamentos de transformadores provocados por interferência da vegetação arbórea. Foram consideradas 287, 327 e 241 ocorrências registradas em 2010, 2011 e 2012, respectivamente. Esses dados históricos pertencem a uma rede de distribuição localizada em um município do interior paulista, Brasil. A Figura 1 apresenta alguns fatores relevantes associados a interrupções na cidade analisada. Observa-se que a vegetação

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

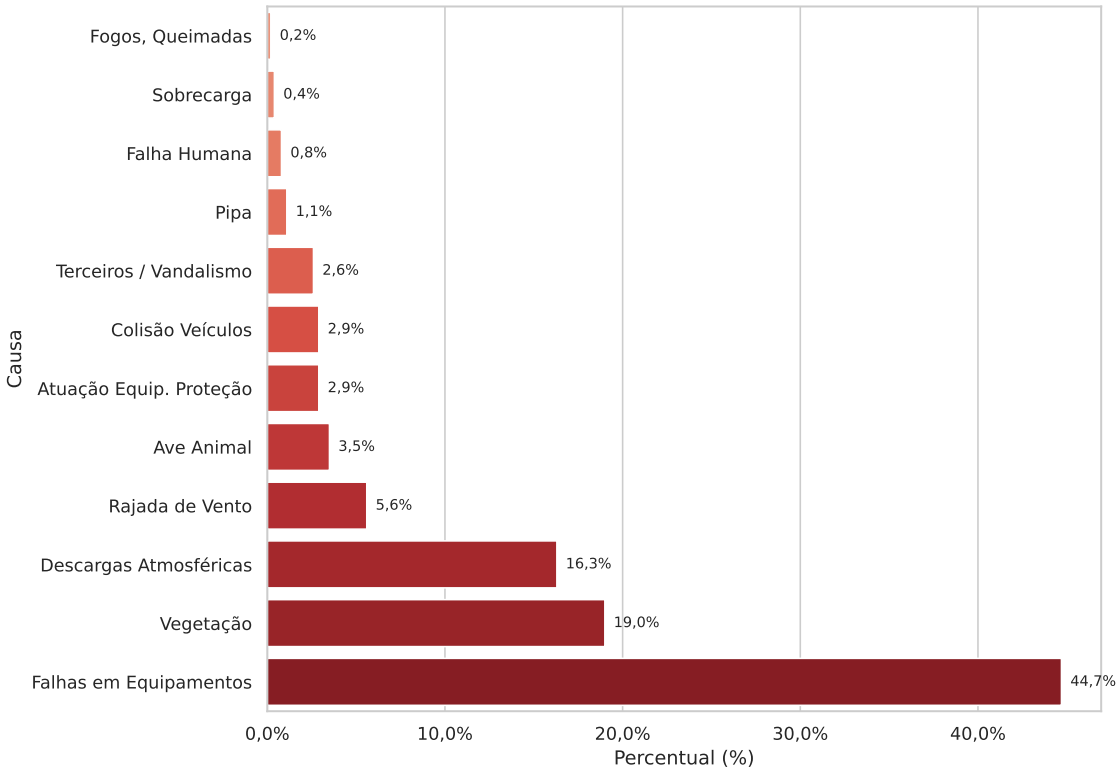


Figura 1: Principais fatores causadores de interrupções no fornecimento de energia elétrica.

arbórea é a segunda causa responsável por interrupções no fornecimento de energia elétrica, sendo associada a 19% das ocorrências registradas.

3.2. Análise exploratória de dados geográficos

3.2.1. Mapas temáticos

Na Figura 2 são apresentados mapas temáticos correspondentes aos anos de 2010 até 2012. Eles mostram a distribuição espacial das interrupções de energia devido à vegetação arbórea na cidade em estudo. O principal objetivo dessa análise consiste em identificar os *pontos quentes*, isto é, as áreas com alta concentração de ocorrências. Dessa forma, quanto maior o valor obtido na estimativa de densidade por *kernel*, maior será a aglomeração de interrupções registradas em um determinado SC.

No mapa correspondente ao ano de 2010 (Figura 2a), observam-se *pontos quentes* concentrados principalmente na região central da cidade. Existem algumas áreas com elevada densidade de faltas nas regiões centro-oeste e nordeste. Em 2011 (Figura 2b), o padrão espacial permanece semelhante; no entanto, com maior intensidade na região central e uma leve expansão das *áreas quentes* em direção ao centro-noroeste. Finalmente, em 2012 (Figura 2c), nota-se uma distribuição espacial mais difusa, com menor concentração de faltas, embora ainda persistam focos relevantes nas regiões central e centro-oeste, indicando uma suavização no padrão de aglomeração das interrupções. Como a análise considera faltas associadas a desligamentos em transformadores da rede de distribuição, ressalta-se que a região central possui alta densidade populacional e; portanto, concentra um maior número de transformadores. Esse fato justifica a maior ocorrência de interrupções em SCs localizados nessa região.

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

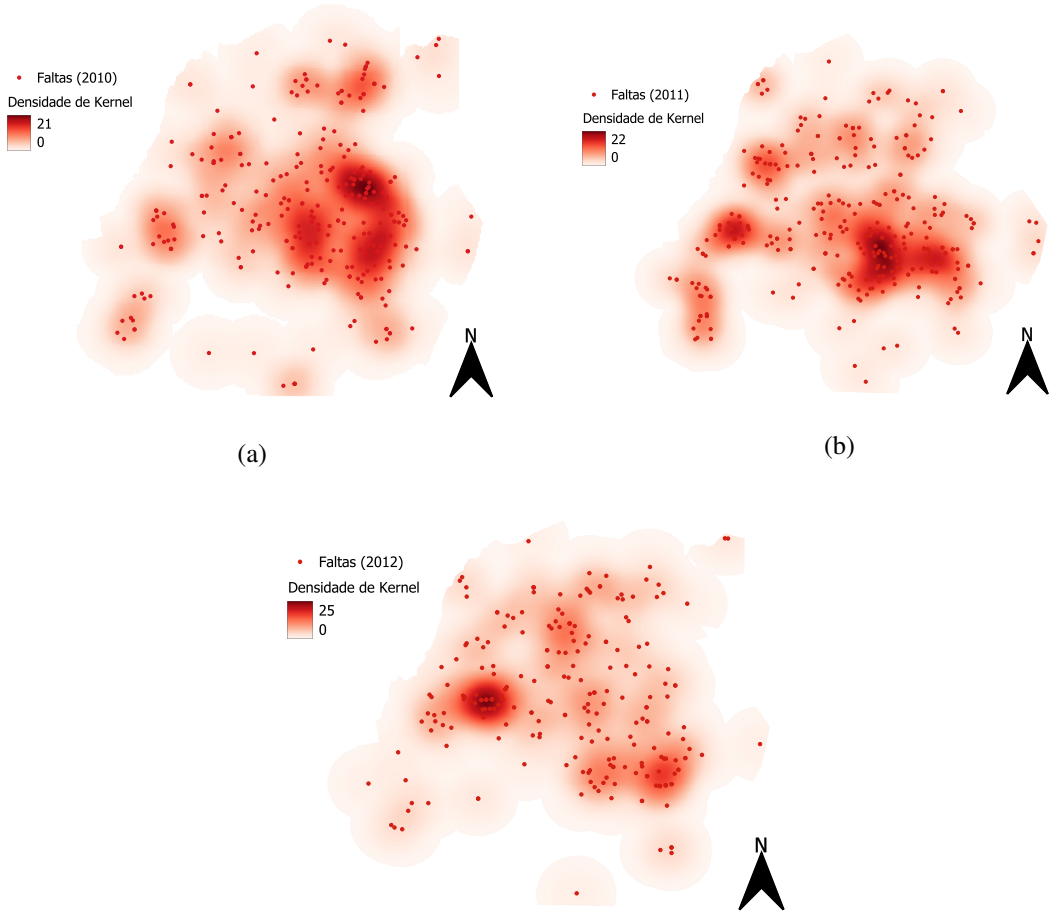


Figura 2: Mapa de calor: (a) 2010, (b) 2011 e (c) 2012.

Algumas métricas pertencentes à estatística descritiva são apresentadas na Tabela 1 referentes ao número de faltas associadas à vegetação arbórea por SC para os anos de 2010 a 2012. Observa-se que, em média, cada SC registrou aproximadamente uma falta por ano. No entanto, em 2011, destaca-se um caso atípico em que um único SC concentrou nove interrupções associadas à interferência da vegetação. Esse resultado pode indicar a presença de características locais específicas que favorecem a ocorrência dessas falhas, como a existência de vegetação densa ou muito próxima à rede elétrica.

A Tabela 2 apresenta o resumo estatístico com a duração das interrupções por SC para os anos de 2010 a 2012. Observa-se que o tempo médio de interrupção variou de forma moderada ao longo do período analisado, sendo de 513,72; 549,18 e 582,97 minutos em 2010, 2011 e 2012, respectivamente. Em 2010 ocorreu o maior tempo de interrupção registrado de 78h52min. Os valores de desvio padrão indicam elevada dispersão em todos os anos, evidenciando a ocorrência de interrupções com duração variável entre os SCs. Esses resultados evidenciam a necessidade de estratégias direcionadas para manutenção corretiva e preventiva em regiões críticas, principalmente naquelas sujeitas às faltas mais prolongadas.

A Figura 3a apresenta a distribuição espacial do tempo médio de interrupção em minutos

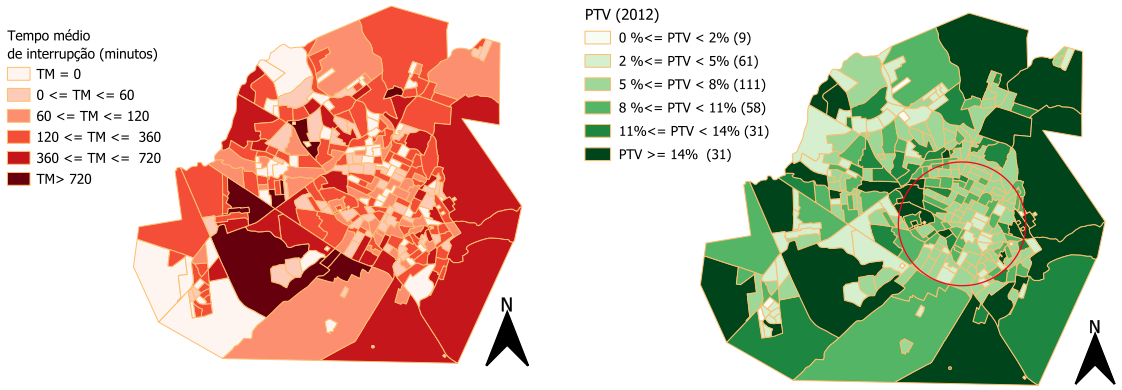
DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

Tabela 1: Resumo estatístico de faltas por setor censitário

Parâmetros	**2010**	**2011**	**2012**
Máx.	7	9	6
Mín.	0	0	0
Média	1	1,1	0,8
Desvio padrão	1,23	1,48	1,16
Total de faltas	**287**	**327**	**241**

Tabela 2: Resumo estatístico com o tempo de interrupção por setor censitário

Parâmetros	**2010**	**2011**	**2012**
Máx.	4.732,0	2.736,0	2.902,0
Mín.	7	5	8
Média	513,72	549,18	582,97
Desvio padrão	736,17	565,84	604,38



(a) (b)

Figura 3: Distribuição espacial: (a) tempo médio de interrupções considerando os anos de 2010 até 2012, (b) vegetação arbórea.

ao longo da área urbana da cidade analisada, onde se consideram dados históricos de três anos. O tempo médio é obtido de acordo com (3). Observa-se que as regiões periféricas concentram o maior número de SCs com elevados tempos médios de interrupção, com destaque para as áreas localizadas nas regiões oeste e sudoeste da cidade, onde alguns SCs registram interrupções no fornecimento de energia superiores a 720 min.

$$TM_i = \frac{\sum_{j=1}^{n} T_{i,j}}{n}, \tag{3}$$

onde TM_i representa o tempo médio de interrupção no SC_i ao longo dos anos analisados, $T_{i,j}$ consiste no tempo total de interrupção registrado no SC_i no ano j e n é o número total de anos analisados.

A distribuição espacial para a variável percentual de vegetação arbórea (PVA) por SC é exibida na Figura 3b, correspondente a 2012. Os mapas para os outros anos não foram apresentados, pois a cobertura de vegetação arbórea tende a apresentar variações muito lentas ao longo do

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

tempo. Os SCs com os maiores valores de PVA estão concentrados nas áreas periféricas da cidade, predominantemente nas regiões leste, nordeste e sudeste, com alguns SCs localizados no centro e no extremo oeste. Conforme esperado, a região central, destacada pelo círculo vermelho, apresenta baixos índices de PVA, devido à elevada densidade populacional.

Na Figura 4, são apresentados os mapas de desvio padrão das faltas. Eles são úteis para identificar valores discrepantes de desligamentos provocados por vegetação na região em estudo. Na Figura 4a, correspondente ao ano de 2010, observa-se que a maioria das regiões (145 SCs) apresentam desvio padrão entre 0 e 0,95, indicando uma baixa concentração de desligamentos. No entanto, destacam-se 16 SCS pertencentes às regiões sudoeste, sudeste e nordeste, com valores correspondentes a 3,5 desvios padrão da média. Assim, há uma forte concentração de faltas nessas localidades.

No ano de 2011 (Figura 4b), observa-se uma mudança sutil na distribuição dos desligamentos. As áreas com maior desvio padrão (acima de 4) permanecem localizadas em algumas localidades da região sudoeste e alguns novos agrupamentos na região centro-oeste. O aumento no número de SCs com valores entre 2,57 e 4 (31 SCs) indica que os desligamentos são mais frequentes e espacialmente mais concentrados em comparação ao ano anterior. Ressalta-se que, em 2011, houve maior número de faltas registrado em comparação a 2010 e 2012.

O mapa correspondente ao ano de 2012 (Figura 4c) apresenta uma pequena variação em comparação aos anos anteriores. A maior parte dos SCs permanecem nas categorias intermediárias de desvio padrão, principalmente entre 0 e 0,8. No entanto, algumas áreas das regiões sudoeste e centro-sul permanecem como elevado valor de desvio padrão, refletindo padrões consistentes de anomalias espaciais ao longo dos anos avaliados.

3.2.2. Aplicação do k-means para agrupamento de setores censitários

O algoritmo k-means foi aplicado para agrupar os SCs com base no número anual de faltas ocasionadas por vegetação arbórea. Essa abordagem permite a visualização, por meio de mapas temáticos, de áreas com comportamentos semelhantes ao longo do tempo, ou seja, com número de ocorrências de faltas similares. Adotam-se quatro *clusters* ($k = 4$), que proporcionou uma boa separação entre os dados, conforme os critérios de desempenho do algoritmo. O menor valor obtido para a razão entre a soma dos quadrados entre os *clusters* (BSS) e a soma total dos quadrados (TSS) corresponde a 0,9263, indicando que aproximadamente 92,63% da variância total dos dados é explicada pela separação em *clusters*. Quanto maior o valor dessa razão, melhor é a capacidade de discriminação dos agrupamentos formados, refletindo maior homogeneidade no interior de cada *cluster* e maior heterogeneidade entre os *clusters*.

No algoritmo k-means, a métrica de similaridade adotada é a distância Euclidiana. Os dados foram previamente padronizados por meio do método *z-score*. Essa transformação garante que todas as variáveis contribuam de forma equilibrada para o processo de agrupamento, evitando distorções provocadas por diferenças de escala. No entanto, os resultados finais são fornecidos em unidades originais, o que permite comparar os valores dos centroides dos *clusters* com as médias reais das faltas observadas em cada ano.

Na Tabela 1 são apresentadas as médias de faltas registradas sendo de 1,0; 1,1; e 0,8 em 2010, 2011 e 2012, respectivamente. Ao comparar esses valores com as médias dos centros dos *clusters* (Tabela 3 e Figura 6), observa-se que os *clusters* 3 e 4 apresentam valores significativamente superiores à média geral. Em 2010, por exemplo, o *cluster* 3 registrou média de 2,28 faltas, o que corresponde ao dobro da média geral. Por outro lado, o *cluster* 4 atingiu 4,69 faltas, sendo cerca de quatro vezes maior do que a média geral.

No ano de 2011, três *clusters* apresentaram médias superiores à média geral: o *cluster* 2 apresenta um valor ligeiramente acima da média sendo de 1,43, enquanto os *clusters* 3 e 4 registra-

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br



Figura 4: Desvio padrão: (a) 2010, (b) 2011 e (c) 2012.

ram médias de 3,35 e 6,2, respectivamente. Em 2012, há manutenção do padrão, onde os *clusters* 3 e 4 novamente apresentam médias superiores à média geral com valores registrados de 2,33 e 4,7 para os *clusters* 3 e 4, respectivamente.

Os mapas temáticos com a distribuição espacial dos agrupamentos que compartilham padrões semelhantes de ocorrência de faltas são apresentados na Figura 5. Observa-se que, o *cluster* 1 concentra o maior número de SCs sendo 145, 148 e 169 em 2010, 2011 e 2012, respectivamente. Esse agrupamento representa as regiões com ausência de faltas ocasionadas por vegetação arbórea.

Ressalta-se que, neste estudo, foram consideradas apenas as faltas ocasionadas por desligamentos em transformadores da rede de distribuição decorrentes da interferência de vegetação arbórea, ou seja, outras causas de desligamento não foram incluídas. Esse recorte justifica o elevado número de SCs sem registros de faltas ao longo dos anos analisados.

Adicionalmente, nota-se uma tendência espacial importante: os setores classificados nos *clusters* 3 e 4, que apresentam as maiores médias de faltas, estão localizados majoritariamente em regiões periféricas da área urbana. Esse padrão persiste ao longo dos três anos avaliados. Isso é um indício de que as áreas periféricas estão mais suscetíveis à ocorrência de faltas ocasionadas por vegetação arbórea, possivelmente devido à proximidade da zona rural com maior presença de áreas verdes ou arborização sem poda sazonal.

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

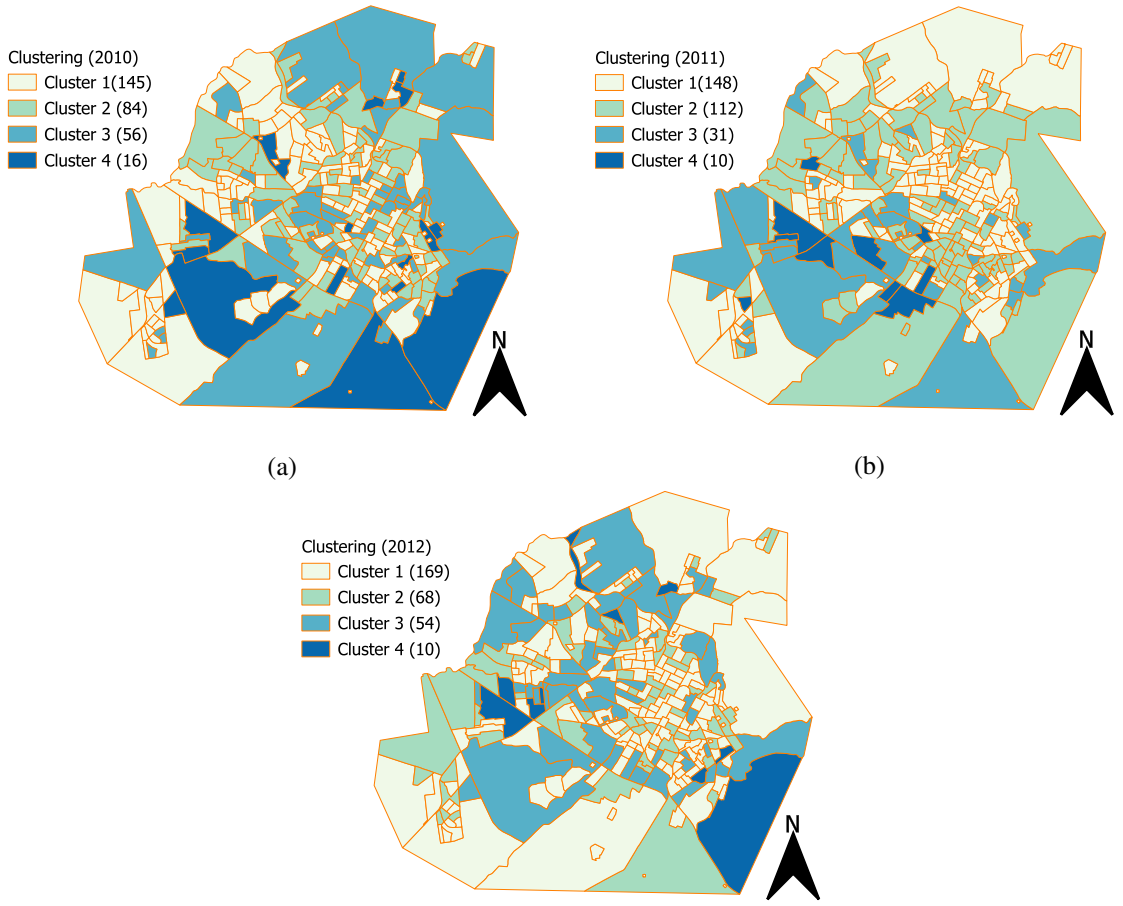


Figura 5: *Clusters* faltas via k-means: (a) 2010, (b) 2011 e (c) 2012.

Tabela 3: Média de faltas anuais para cada *cluster*

Clusters	**2010**	**2011**	**2012**
Cluster 1	0	0	0
Cluster 2	1	1,43	1
Cluster 3	2,28	3,35	2,33
Cluster 4	4,69	6,2	4,7

4. Conclusões

Neste estudo, realizou-se uma análise exploratória de dados geográficos com o objetivo de identificar tendências espaciais associadas às faltas ocasionadas pela interferência da vegetação arbórea no sistema de distribuição de energia elétrica. Para isso, foram elaborados mapas temáticos, que evidenciaram as regiões com maior concentração de interrupções, e mapas de desvio padrão, que indicaram o grau de variação no número de faltas entre os setores analisados.

O algoritmo k-means foi aplicado para identificar padrões de agrupamento ou *clusters*, permitindo a formação de conjuntos de setores censitários com número semelhante de ocorrências de faltas. Essa abordagem facilitou a interpretação de padrões espaciais recorrentes. Os resultados indicaram uma alta concentração de faltas nas regiões periféricas da cidade, indicando a presença

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

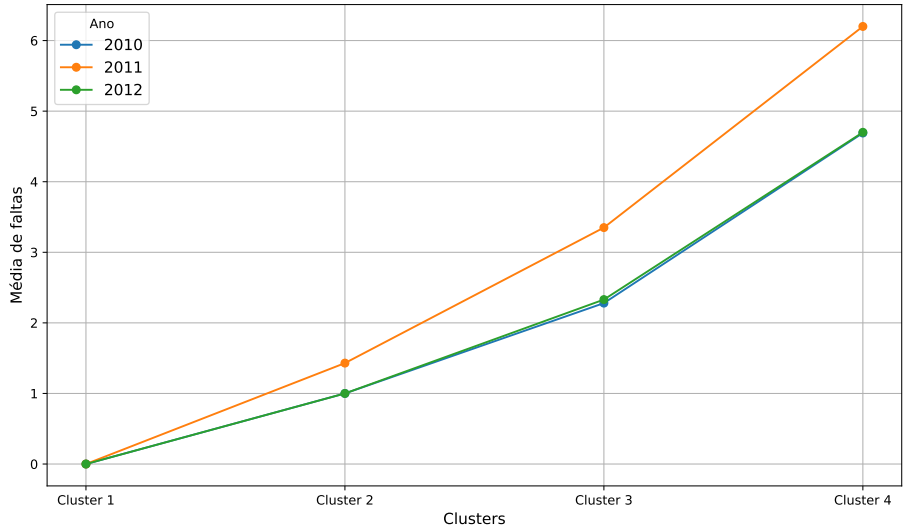


Figura 6: Média de faltas anuais para cada *cluster*.

de características locais que possivelmente contribuem para essas ocorrências. Destaca-se que essas regiões estão na proximidade da zona rural e têm os maiores percentuais de cobertura por vegetação arbórea.

Em estudos futuros, os autores pretendem incorporar outras variáveis explicativas e avaliar correlações locais dessas variáveis com as interrupções de energia elétrica. Os resultados deste estudo dão suporte à tomada de decisão para o planejamento de ações para manutenção corretiva e preventiva, auxiliando na priorização de áreas críticas e promovendo maior confiabilidade e qualidade no fornecimento de energia elétrica.

Agradecimentos

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil, processo nº 2024/20578-1, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processos 409062/2023-5 e 311430/2025-2.

Referências

Anselin, L. (2024). *An Introduction to Spatial Data Science with GeoDa: Clustering Spatial Data*, volume 2. CRC Press.

Chen, P.-C., Dokic, T., Stokes, N., Goldberg, D. W., e Kezunovic, M. (2015). Predicting weather-associated impacts in outage management utilizing the GIS framework. In *2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, p. 417–422.

Doostan, M., Sohrabi, R., e Chowdhury, B. (2020). A data-driven approach for predicting vegetation-related outages in power distribution systems. *International Transactions on Electrical Energy Systems*, 30:12154. ISSN 20507038.

Fränti, P. e Sieranoja, S. (2018). K-means properties on six clustering benchmark datasets. *Applied intelligence*, 48:4743–4759.

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br

Gallaher, A., Graziano, M., e Fiaschetti, M. (2021). Legacy and shockwaves: A spatial analysis of strengthening resilience of the power grid in connecticut. *Energy Policy*, 159:112582–112582.

Gu, Z. (2022). Complex heatmap visualization. *iMeta*, 1. ISSN 2770596X.

Haining, R. P. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press.

Hartigan, J. A. e Wong, M. A. (1979). A k-means clustering algorithm. *Journal of the royal statistical society. Series c (applied statistics)*, 28(1):100–108.

Le Gallo, J. e Ertur, C. (2003). Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. *Papers in regional science*, 82:175–201.

Luc Anselin (2020). Standard deviation map. URL https://geodacenter.github.io. Acesso em: 15 02 2025.

Melagoda, A. U., Karunarathna, T. D., Nisaharan, G., Amarasinghe, P. A., e Abeygunawardane, S. K. (2021). Application of machine learning algorithms for predicting vegetation related outages in power distribution systems. In *2021 3rd International Conference on Electrical Engineering, EECon 2021*, p. 25–30.

Monteiro, A. M. V., Câmara, G., Carvalho, M., e Druck, S. (2004). Análise espacial de dados geográficos. *Brasília: Embrapa*.

ONS (2023). Qualidade do suprimento. URL https://www.ons.org.br/paginas/resultados-da-operacao/qualidade-do-suprimento.

Parent, J. R., Meyer, T. H., Volin, J. C., Fahey, R. T., e Witharana, C. (2019). An analysis of enhanced tree trimming effectiveness on reducing power outages. *Journal of environmental management*, 241:397–406.

Santos, A. S., Faria, L. T., Lopes, M. L. M., e Minussi, C. R. (2023). Power distribution systems' vulnerability by regions caused by electrical discharges. *Energies*, 16(23):7790.

Santos, A. S., da Silva, R. J., Montenegro, P. A., Faria, L. T., Lopes, M. L. M., e Minussi, C. R. (2025). Integrating autoencoders to improve fault classification with PV system insertion. *Electric Power Systems Research*, 242:111426.

Santos, A. S., Faria, L. T., Boschi, L. S., Lopes, M. L. M., e Minussi, C. R. (2024). Estimation of vulnerable areas to faults caused by tree vegetation in power distribution systems. *IEEE Access*, 12:48472–48484. ISSN 21693536.

Wedagedara, H., Witharana, C., Fahey, R., Cerrai, D., Joshi, D., e Parent, J. (2023). Modeling the impact of local environmental variables on tree-related power outages along distribution powerlines. *Electric Power Systems Research*, 221:109486.

DOI: 10.59254/sbpo-2025-212134
https://proceedings.science/p/212134?lang=pt-br