

Análise da Intensidade do Pico de Corrente de Descargas Elétricas Associadas à Tempestades Identificadas por Técnicas de Clusterização

Mariana Kleina, **Luiz C. Matioli,**

Programa de Pós Graduação em Métodos Numéricos em Engenharia, UFPR
81531-980, Curitiba, PR

E-mail: marianakleina11@gmail.com, matioli@ufpr.br

Eduardo A. Leite

Instituto Tecnológico SIMEPAR
81531-980, Curitiba, PR
E-mail: alvim@simepar.br

Resumo: *Este trabalho tem por objetivo analisar a variabilidade da intensidade do pico de corrente de descargas atmosféricas à nível de cada evento de tempestade, formadas por técnicas de clusterização, buscando identificar se o comportamento desta variável apresenta dissonância em relação ao comportamento climatológico, o qual se considera a distribuição geral de probabilidade, estimado a partir de medições. Para isso, foram tomadas como base as descargas elétricas incidentes no ano de 2011 na região da linha de transmissão de energia LT 765 kV conhecida como Linhão de Itaipu. Estas descargas foram então agrupadas por algumas técnicas de clusterização e gerados índices estatísticos próprios que comprovaram o pressuposto.*

Palavras-chave: *Descargas atmosféricas, Variabilidade do pico de corrente, Técnicas de clusterização*

1 Introdução

O Brasil é líder mundial na ocorrência de descargas atmosféricas que são produzidas devido ao acúmulo de cargas elétricas na atmosfera gerados por sistemas meteorológicos de diversos tipos e configurações, denominados genericamente de tempestades. Existem diversos tipos de descargas, sendo a de maior interesse as que atingem o solo, denominadas de “nuvem-solo”, responsáveis tanto por afetar a segurança humana, quanto por grandes prejuízos em diversos setores econômicos, com destaque para o setor elétrico. Neste setor elas são responsáveis por danos nos equipamentos e desligamentos não programados nos sistemas de transmissão e distribuição de energia, podendo gerar apagões de grande abrangência e impactos econômicos.

Uma característica fundamental da descarga atmosférica, responsável pelo seu potencial de impacto e dano no sistema elétrico e outros setores é a intensidade de seu pico de corrente. Existe grande variabilidade espacial e temporal associada a esta variável, mas seu conhecimento é limitado, sendo que a prática corrente na engenharia elétrica é considerar a distribuição geral de probabilidade desta variável, dita “climatológica”, estimada a partir de medições em todo o planeta. A investigação do comportamento da intensidade do pico de corrente e sua variabilidade no espaço e no tempo pode subsidiar medidas de planejamento, projeto e proteção operacional mais adequadas e confiáveis do sistema elétrico.

1.1 As Descargas Atmosféricas e o seu Pico de Corrente

Uma descarga atmosférica pode ser simples (composta apenas pela “descarga de retorno”) ou múltipla (composta pelas “descargas de retorno subsequentes”). As descargas de retorno e subsequentes são chamadas de *strokes* e o conjunto desses strokes recebe o nome de *flash*. O que diferencia a polaridade de uma descarga é o local da nuvem onde se origina. Os negativos saem da parte baixa da nuvem, onde se concentra a energia negativa, já os positivos saem do topo da nuvem, onde ficam as partículas carregadas positivamente. Descargas negativas são mais frequentes do que positivas, sendo que as positivas apresentam caráter mais destrutivo.

As distribuições estatísticas de pico de corrente são inferidas a partir de dados experimentais, registrados por altas torres instrumentadas. Historicamente, a distribuição lognormal tem sido usada para descrever a distribuição do pico de corrente das descargas atmosféricas.

No trabalho de Berger et al. (1975) foi proposto aproximar o pico de corrente da distribuição encontrada por uma lognormal com mediana de $\mu_I = 31.1$ e desvio-padrão logarítmico de $\sigma_I = 0.484$ respectivamente. O CIGRÉ (1991) optou por aproximar o valor da mediana e o desvio padrão logarítmico da distribuição de frequência lognormal dos valores de pico de corrente de raios subsequentes, e encontrou $\mu_I = 12.3$ e $\sigma_I = 0.53$.

A distribuição lognormal é inferida ao pico de corrente do flash, que é geralmente tomado como sendo a intensidade do pico de corrente da primeira descarga, pois geralmente é a mais forte por conseguir romper a rigidez dielétrica do ar. Neste estudo, o pico de corrente do flash é tomado como uma média dos picos das descargas que o compõe, ou seja, a média dos picos dos strokes. Isto faz com que o pico seja menor do que com a técnica descrita anteriormente, pois como já mencionado, as descargas de retorno subsequentes tendem a ter intensidade da corrente menores do que a primeira. Todavia tomando o pico de corrente desta forma, ele também segue a distribuição lognormal, pois a maioria dos raios são raios simples (ou seja, somente composto pela descarga de retorno) e os raios múltiplos são providos da mesma natureza ou fenômeno que a primeira descarga, logo apresentam mesmo âmbito intrínseco que o flash.

2 Materias e Localização

2.1 Região de Estudo

Descargas atmosféricas são responsáveis por falhas, desligamentos, variações de tensão e introdução de transientes no sistema elétrico, sendo responsáveis por prejuízos, materiais e humanos, para concessionárias de energia e consumidores (Silva Neto, 2004).

Conhecer as características das tempestades que afetam diretamente este setor é de grande valia, aumentando a capacidade de previsão e alertas. Por este motivo, a região piloto para esta pesquisa é uma área que abrange a linha de transmissão de energia mais importante do Brasil, a linha LT 765 kV, conhecida como Linhão de Itaipu, responsável pela transmissão de aproximadamente um sexto da energia consumida no Brasil. O sistema é composto de três linhas de transmissão entre as subestações de Foz do Iguaçu (PR) e Tijuco Preto (SP), cada uma com cerca de 2 mil torres de transmissão e extensão de aproximadamente 900 km.

2.2 Dados de Descargas Atmosféricas

Os dados de descargas utilizados nesta pesquisa foram provenientes da Rede Nacional de Detecção de Descargas Atmosféricas (RINDAT), mantido pelo Instituto Tecnológico SIMEPAR em cooperação com Furnas, CEMIG e INPE, com informações de dia e horário, localização (latitude/longitude) e intensidade do pico de corrente (em kA) de todas as descargas incidentes em 2011 sobre a região do Linhão de Itaipu. Essa pesquisa totalizou 1.448.466 descargas.

Para processamento e análise da metodologia proposta, fez-se uso do da linguagem e ambiente para computação estatística “R” (R Core Team, 2012).

3 Agrupamento de Descargas Atmosféricas

Para que uma tempestade elétrica seja caracterizada, fez-se uso da metodologia de agrupamentos das descargas por clusterização, uma técnica que favorece interpretar características primordiais de grandes conjuntos de dados. O conceito de clusterização é, basicamente, juntar dados similares no mesmo grupo, também chamado de *cluster*, assim dados de um mesmo cluster tem mais características em comum entre si do que com dados de outros clusters (Rasmussen, 2013).

Um algoritmo de clusterização deve levar em conta diversos fatores, tais como: a forma com que os dados estão sendo representados, como medir a similaridade entre dados e entre clusters, como estimar a qualidade do resultado obtido pelo método, entre outros. A maneira como esses fatores são abordados e a escolha dos parâmetros iniciais definem os diferentes algoritmos de clusterização.

Basicamente, foram clusterizadas as descargas atmosféricas incidentes na região piloto no intervalo de uma hora, para todo o ano de 2011. É importante ressaltar que são necessárias técnicas de clusterização onde não se conheça o número de clusters, pois é evidente que esse número varia de acordo com o número de descargas que incidiram no tempo considerado, o qual é altamente dependente dos períodos do dia, estações do ano, entre outros. A quantidade de descargas varia tanto que em 2011 há dias em que não se tem registro de nenhuma descarga e há dias em que o número de descargas registradas ultrapassa 50 mil. Os tipos de clusterização abordados nesta pesquisa são descritos a seguir.

3.1 Clusterização Hierárquica

A principal característica dos algoritmos hierárquicos é a variação do número de clusters a cada iteração. O funcionamento destes algoritmos utiliza o conhecimento de uma determinada alocação dos objetos em clusters e, a partir daí, considerando a distância entre os elementos dos clusters, forma os seus sucessores. Sua denominação vem do fato da evolução do algoritmo poder ser representada como uma árvore, com cada nível sendo construído a cada iteração (Johnson & Wichern, 2007).

Na versão aglomerativa, cada objeto representa um cluster. A cada iteração dois clusters são escolhidos e mesclados, formando um cluster maior. O que caracteriza os algoritmos de produzir agrupamentos é o critério usado para definir a distância entre grupos. Calculadas as distâncias entre todos os pares de pontos dos clusters, em que cada ponto pertence a clusters distintos, se a distância mínima entre todos os pares de pontos é escolhida, tem-se a mínima distância, se for escolhida a distância máxima entre todos os pares de dados, tem-se a máxima distância, e se for escolhida a distância entre os centroides dos clusters, tem-se a distância pelo centroide (centroide é valor médio dos objetos que pertencem a determinado cluster).

Foram testados os três tipos de distâncias, sendo que as distâncias mínima e máxima geraram grupos que não caracterizaram bem espacialmente as tempestades elétricas. Por isso, nesta pesquisa o método do centroide foi adotado, logo a distância entre grupos é definida pela distância entre os centros. Em cada etapa procura-se fundir grupos que tenham a menor distância entre si (Hair et al., 2009). A distância máxima para fundir dois agrupamentos será chamada de *DistMax* nesta pesquisa.

3.2 Clusterização por Densidade

Os métodos baseados em densidade permitem descobrir grupos de formatos arbitrários. Consideram grupos como sendo regiões densas de objetos no espaço de dados que são separados por regiões de baixa densidade, que geralmente representam ruídos (Han & Kamber, 2006). Um exemplo de algoritmo baseado em densidade é o algoritmo DBSCAN (Ester et. al., 1996).

Para o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) um grupo é definido como um conjunto máximo de pontos conectados por densidade (Oliveira, 2007). Este método requer que dois parâmetros iniciais sejam informados: o *Raio* (distância entre um objeto

e seus vizinhos) e o objeto central, chamado *MinPts*. Para gerar os clusters, o DBSCAN testa o raio de cada ponto da base de dados. Se o raio de um objeto p contém mais de um ponto central (*MinPts*), então um novo cluster é criado para o objeto p . Os objetos no raio de p são então adicionados ao novo cluster. Durante o processo, um objeto central que já pertence a um cluster pode ser encontrado em outro cluster. Nesse caso, os dois clusters são agrupados em um só e o processo se encerra quando nenhum novo ponto for adicionado a qualquer cluster.

3.3 Clusterização Sharp

Dados $x_1, x_2, \dots, x_n \sim f(x)$, a técnica usada para estimar a densidade desconhecida $f(x)$ é a estimativa do núcleo da densidade: $\hat{f}(x) = \frac{1}{n} \sum_i k_{bw}(x_i - x)$ onde k_{bw} é a função densidade de probabilidade simétrica, com parâmetro bw (a largura da banda).

Esta estimativa é tendenciosa: ela subestima os picos e vales da densidade. Choi e Hall (1999) introduziram uma técnica chamada *Data Sharpening* que reduz esses vies movendo observações através de modos locais. Este método ajusta os dados distantes da moda ou a alguma distância das caldas e não altera os demais.

Esta técnica foi utilizada como algoritmo de clusterização por Woolford & Braun (2006) para agrupar relâmpagos e também será utilizada nesta pesquisa.

4 Análise dos Resultados

Foram então clusterizadas as descargas incidentes na região de estudo no ano de 2011 de uma em uma hora, através da sua posição geográfica, utilizando os métodos de clusterização descritos acima, variando seus parâmetros. Os parâmetros são: *DistMax* para o método hierárquico, *Raio* e *MinPts* para o algoritmo DBSCAN e bw para o método Sharp, os quais foram descritos na Seção 3.

A Figura 1 mostra uma ilustração das diferentes formas de clusterização dos raios que caíram na região estudada, das 20 às 21 horas do dia 28 de janeiro de 2011, totalizando 1483 ocorrências. Os clusters são separados por cores. Na Figura 1a, os raios foram clusterizados pelo método hierárquico com parâmetro $DistMax = 100$; na Figura 1b é utilizado o algoritmo DBSCAN com $Raio = 100$ e $MinPts = 20$, onde os círculos cinzas não preenchidos representam ruídos; e na Figura 1c, o método Sharp com parâmetro $bw = 0.5$ é utilizado.

Para que se possa comparar o comportamento das tempestades formadas por clusterização com seu comportamento climatológico em termos da intensidade do pico de corrente, foram geradas tempestades aleatórias através de simulação Monte Carlo.

Como admite-se que o pico de corrente das descargas segue uma distribuição lognormal, então o logaritmo do valor absoluto do pico de corrente (LAPC) é aproximadamente normal. Assim, pela normalidade desta variável, almeja-se diferenciar tempestades clusterizadas de tempestades aleatórias pela média e desvio padrão do LAPC dentro de cada grupo formado no sentido de que para a primeira técnica espera-se que haja grande variabilidade em relação à média e pouca variabilidade para o desvio padrão desta variável dentro de cada tempestade. Já para a segunda técnica este comportamento pode não ser observado, não conseguindo separar descargas com altos e baixos valores para o seu pico de corrente. E para confirmar o pressuposto, os seguintes índices foram calculados:

$$\mu_\sigma = \frac{\mu(\sigma_{clusters})}{\mu(\sigma_{aleatórios})} \quad \text{e} \quad \sigma_\mu = \frac{\sigma(\mu_{clusters})}{\sigma(\mu_{aleatórios})}$$

onde μ_σ mede a proporção de quanto a média dos desvios padrões do LAPC de tempestades geradas por clusterização varia em relação à média dos desvios das tempestades geradas aleatoriamente. Como o LAPC segue a distribuição normal, espera-se que os desvios dos clusters sejam pequenos em média. Já σ_μ mede a proporção de quanto o desvio padrão das médias do LAPC dos clusters varia em relação ao desvio padrão das médias das tempestades aleatórias. Resumidamente, espera-se que os métodos de clusterização consigam separar descargas que possuem

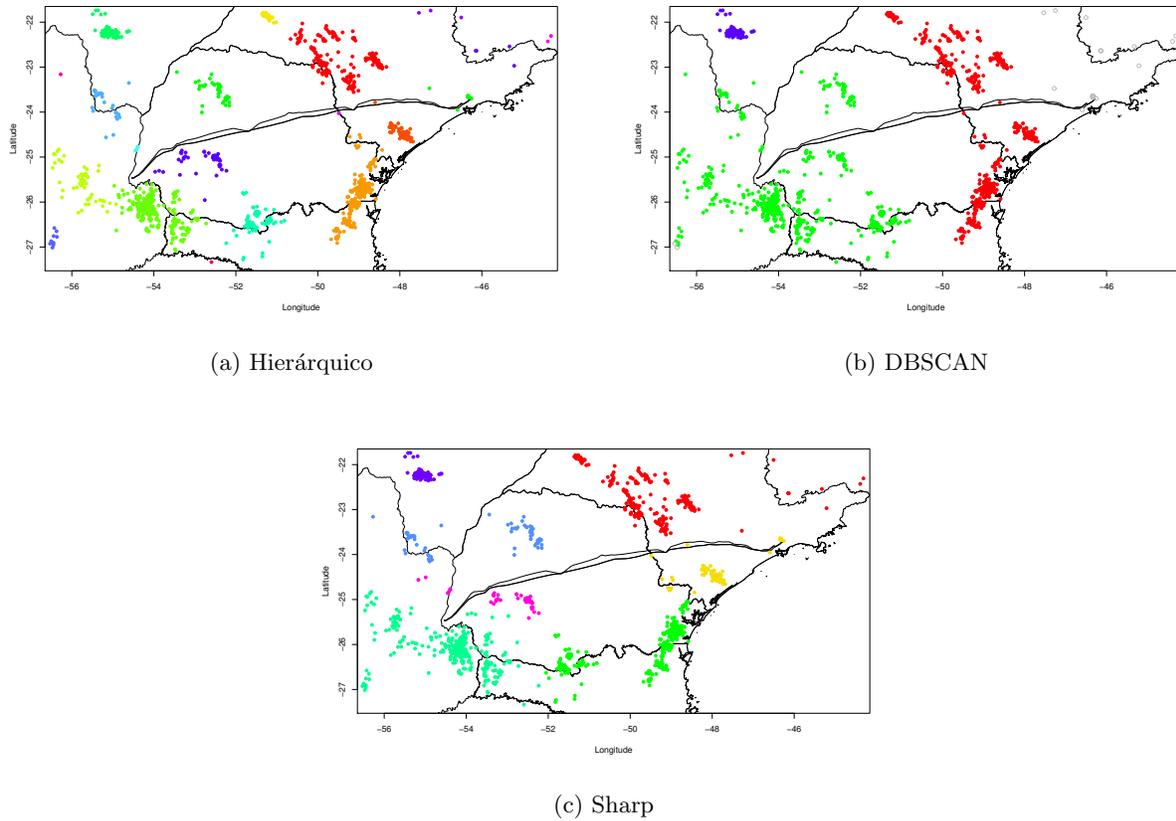


Figura 1: Clusterização de 1483 raios que caíram na região do Linhão de Itaipu no dia 28/01/2011 das 20 às 21 horas, pelo método hierárquico (1a), algoritmo DBSCAN (1b) e Sharp (1c).

picos de corrente altos das descargas que possuem baixos picos de corrente, e pela normalidade associada à essa variável, estes índices irão evidenciar este fato.

Tabela 1: Índices estatísticos obtidos com diferentes tipos de clusterização e parâmetros

Método de clusterização / parâmetro(s)	σ_{μ}	μ_{σ}
Hierárquica / 100	4.513	0.868
Hierárquica / 125	4.711	0.874
Hierárquica / 150	4.811	0.884
DBSCAN / 50 20	4.313	0.852
DBSCAN / 75 20	4.683	0.872
DBSCAN / 100 20	4.901	0.886
Sharp / 0.1	3.862	0.845
Sharp / 0.5	4.842	0.884

Na Tabela 1, são mostrados os resultados da metodologia proposta aplicada às três técnicas de clusterização descritas, para alguns parâmetros. Analisando, por exemplo, o resultado do algoritmo DBSCAN com parâmetros $Raio = 100$ e $MinPts = 20$, conclui-se que o desvio padrão das médias dos clusters é 4.9 vezes maior do que o desvio padrão das médias das tempestades aleatórias; e a média dos desvios padrões dos clusters é 11.4% menor do que a média dos desvios padrões das tempestades aleatórias. No geral, tempestades formadas por clusterização tem desvio padrão das médias de 3 a 5 vezes maior do que as aleatórias e média dos desvios padrões de 10 a 20 % menor do que as aleatórias. À vista disso, fica evidenciado que os métodos de clusterização são capazes de reproduzir as características das tempestades elétricas no que diz

respeito a mais importante variável da descarga: o pico de corrente.

5 Conclusões

Através do estudo realizado, pode-se comprovar o pressuposto de que tempestades elétricas se caracterizam como tipo específico de eventos meteorológicos, estão sob influência de condições atmosféricas próprias, assim apresentando comportamentos específicos de descargas em relação ao seu pico de corrente, diferindo do comportamento climatológico conhecido na literatura.

No contexto geral, independente da técnica de clusterização e dos parâmetros, é possível afirmar que o pico de corrente de descargas de tempestades agrupadas por clusterização apresentam características próprias, diferindo e muito das características do pico de corrente das descargas como um todo, que seguem a distribuição geral de probabilidade.

Com o desenvolvimento destes métodos de clusterização espera-se futuramente poder estudar e identificar as características das tempestades elétricas capazes de produzir descargas de alta intensidade de pico de corrente, aumentando a capacidade de previsão e alerta antecipado aos múltiplos setores impactados pela sua ocorrência, em especial, o setor elétrico.

Referências

- [1] K. Berger, R. B. Anderson & H. Kroeninger, Parameters of lightning flashes, *Electra*, 41, pp. 23–37, (1975).
- [2] CIGRÉ – INTERNATIONAL COUNCIL ON LARGE ELECTRIC SYSTEMS, Guide to Procedures for Estimating the Lightning Performance of Transmission Lines, em “CIGRÉ WG 33-01”, 1991.
- [3] E. Choi & P. Hall, Data sharpening as a prelude to density estimation, *Biometrika*, 86, pp. 941-947, (1999).
- [4] M. Ester, H-P. Kriegel, J. Sander, & X. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, em “International Conference on Knowledge Discovery and Data Mining (KDD-96)”, pp. 226-231, 1996.
- [5] J. F. Hair, W. C. Black, B. J. Babin & R. E. Anderson, “Multivariate Data Analysis”, Hardcover, 2009.
- [6] J. Han & M. Kamber, “Data Mining: Concepts and Techniques”, Elsevier, 2006.
- [7] R. A. Johnson & D. W. Wichern, “Applied Multivariate Statistical Analysis”, Hardcover, 2007.
- [8] T. B. S. Oliveira. “Clusterização de dados utilizando técnicas de redes complexas e computação bioinspirada”, Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação, USP, 2008.
- [9] R Core Team, “R: A language and environment for statistical computing”, R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org/>>, 2012.
- [10] E. Rasmussen, “Clustering Algorithms”, acesso em: 18 Junho 2013. Disponível em: <<http://orion.lcg.ufrj.br/Dr.Dobbs/books/book5/chap16.htm>>.
- [11] A. Silva Neto, “Tensões induzidas por descargas atmosféricas em redes de distribuição de baixa tensão”, Dissertação de Mestrado, Escola Politécnica, USP, 2004.
- [12] D. G. Woolford & W. J. Braun, “Convergent Data Sharpening for the Identification and Tracking of Spatial-Temporal Centers of Lightning Activity”, John Wiley & Sons, 2006.