



## Identificação e Análise das Características de Tempestades Elétricas Utilizando Métodos Numéricos de Agrupamento

Identification and Analysis of Electrical Storms Features Using Clustering Numerical Methods

Mariana Kleina<sup>1</sup>; Luiz Carlos Matioli<sup>1</sup>; Eduardo Alvim Leite<sup>2</sup> & Alana Renata Ribeiro<sup>1</sup>

<sup>1</sup> Universidade Federal do Paraná, Programa de Pós Graduação em Métodos Numéricos em Engenharia.

Av. Coronel Francisco Heráclito dos Santos, Centro Politécnico, Bloco LAME/CESEC, Caixa Postal 19011, 81531-980, Curitiba, PR.

<sup>2</sup> Sistema Meteorológico do Paraná (SIMEPAR).

Av. Coronel Francisco Heráclito dos Santos, Centro Politécnico, Caixa Postal 19100, 81531-980, Curitiba, PR.

E-mails: marianakleina11@gmail.com; matioli@ufpr.br; alanar89@gmail.com; alvim@simepar.br

Recebido em: 23/10/2015      Aprovado em: 18/12/2015

DOI: [http://dx.doi.org/10.11137/2015\\_2\\_91\\_103](http://dx.doi.org/10.11137/2015_2_91_103)

### Resumo

Descargas atmosféricas são fenômenos naturais responsáveis por um grande número de perturbações e desligamentos nos sistemas elétricos. A investigação de seu comportamento pode subsidiar estratégias de operação e projetos de instalação e proteção de linhas de transmissão de energia. O escopo desta pesquisa abrange o estudo do desenvolvimento conjunto das descargas atmosféricas, no espaço e no tempo, como produto de eventos meteorológicos estruturados denominados “tempestades elétricas”, e a análise do seu comportamento nas proximidades de um sistema elétrico selecionado. Para identificar e rastrear estas tempestades é utilizado o método de clusterização *Convergent Data Sharpening*, e para análise do comportamento das trajetórias formadas por estas tempestades em relação a características meteorológicas é empregado o método *Self-Organizing Map* (SOM). Ambos os métodos estão implementados no software R. A pesquisa revelou que tempestades elétricas que ocasionaram falhas no sistema elétrico estudado tendem a possuir maior número de descargas atmosféricas que as demais tempestades e também não possuem pico de corrente médio superior que tempestades comuns, o que poderia ser esperado pela relevância desta variável em perturbações ao sistema elétrico.

**Palavras-chave:** Tempestades Elétricas; *Convergent Data Sharpening*; *Self-Organizing Map*; Linha de Transmissão de Energia

### Abstract

Lightning strokes are natural phenomena responsible for a large number of disturbances and interruptions in electrical systems. The investigation of their behavior can support operation strategies and installation and protection projects of power transmission lines. This research scope includes the study of joint development of lightning strokes, in space and time, as result of structured weather events called “electrical storms”, and the analysis of their behavior near a selected electrical system. To identify and track these storms a clustering method of *Convergent Data Sharpening* is used, and to analyze the trajectories behavior formed by these storms regarding meteorological features the *Self-Organizing Map* (SOM) is applied. Both methods are implemented in software package R. The study revealed that electrical storms that caused failures in the electrical system studied tended to have more lightning strokes than other storms and also had no peak current average greater than regular storms, what could be expected by the relevance of this variable in the electrical system disturbances.

**Keywords:** Electrical Storms; *Convergent Data Sharpening*; *Self-Organizing Map*; Power Transmission Line

## 1 Introdução

O Brasil é um país com vasta extensão territorial e devido à predominância do clima tropical o torna o país com maior incidência de descargas no mundo. Áreas de São Paulo e Paraná apresentam densidade espacial de 2,5 descargas/km<sup>2</sup>/mês, similar ao valor encontrado na Flórida, EUA, conhecida como a capital dos relâmpagos dos Estados Unidos (Beneti, 2012). Descargas, especialmente as do tipo nuvem-solo, que atingem diretamente a superfície terrestre, podem causar diversos prejuízos materiais e muitas vezes mortes de pessoas.

O setor elétrico brasileiro é bastante vulnerável às condições atmosféricas de tempo, dado que 99% das linhas de energia são aéreas. Dentre as principais causas de instabilidade neste setor estão as descargas atmosféricas. Estima-se que cerca de 70% dos desligamentos na transmissão e 40% dos problemas de distribuição de energia são ocasionados por descargas (INPE, 2015). Sendo assim, surge o interesse em estudar e desenvolver novas metodologias que minimizem os danos e impactos causados por descargas, em especial no setor elétrico.

Muitas vezes é difícil interpretar e analisar informações de um grande volume de dados, fazendo-se necessário algum método que auxilie no entendimento destes dados. A análise de cluster tem por principal objetivo a compreensão da estrutura natural em um conjunto de dados. Descargas atmosféricas se encaixam perfeitamente nesta concepção: de maneira isolada elas podem ocultar informações valiosas, visto que quando agrupadas (por alguma medida de similaridade) no espaço e no tempo dão origem ao que nesta pesquisa serão denominadas de tempestades elétricas, estas sim mais qualificadas a expor características importantes por trás de sistemas meteorológicos atuantes.

No trabalho de Woolford & Braun (2006) é estudado o relacionamento entre incêndios florestais e descargas atmosféricas em Ontário no Canadá através de clusterização de dados de descargas no espaço e no tempo. É utilizada a técnica de clusterização chamada *Convergent Data Sharpening* para identificação e monitoramento dos centros das tempestades elétricas.

Em se tratando do SOM, diversas pesquisas na área meteorológica foram desenvolvidas nos

últimos anos. A pesquisa desenvolvida por Sá *et al.* (2012) tem por objetivo avaliar se o SOM é uma boa ferramenta para reconhecimentos de padrões climáticos severos na região leste da Amazônia no Brasil. Esta região, localizada em baixas latitudes, possui condições climáticas extremas especialmente com intensa atividade elétrica. Foram 8 anos de dados utilizados, com diversas variáveis atmosféricas como entrada para o método. O SOM foi comparado com duas variantes do método: *Temporal Kohonen Map* (TKM) e *Recurrent Self-Organizing Map* (RSOM). O resultado obtido foi a confirmação de que as redes utilizadas apresentam bons resultados como classificadoras de eventos extremos, sendo que a rede RSOM apresentou melhor desempenho do que as demais redes.

Já no trabalho de Jensen & Nowotarski (2012) é realizada a aplicação do SOM para discriminação de tempestades em supercélulas ou tornados, com base na forma de perfis verticais de variáveis relevantes, tais como variáveis do vento (velocidade, direção, vorticidade, etc.) e variáveis termodinâmicas (umidade relativa, estabilidade, etc.). A região de estudo foi os Estados Unidos entre abril de 1999 a junho de 2001 e janeiro de 2003 a março de 2005. Variaram-se os níveis de altitude e o número de neurônios no treinamento. Além de bons resultados na classificação do tipo de tempestade, o SOM também foi capaz de discernir padrões de tempo, localização e regimes sazonais.

Sang *et al.* (2008) utilizam o SOM para representar um conjunto de estados atmosféricos de dados meteorológicos representados na forma de 11 variáveis coletadas diariamente em 77 quadrículas de grade por um período de 31 anos no sul da África. Mostrou-se que o SOM é eficaz no agrupamento de dados meteorológicos diários, porém incapaz de prever a evolução temporal a curto prazo destes estados atmosféricos. Contudo, se assumido que os padrões do SOM descrevem bem a meteorologia regional e que os mesmos estados de tempo continuam a operar futuramente, é possível detectar alterações no clima a partir de uma análise de dados, sob a forma de uma menor incidência uniforme dos diferentes estados do que é observado atualmente.

Sheridan & Lee (2011) relatam uma revisão literária minuciosa de muitas outras aplicações climatológicas do SOM ao longo dos últimos anos.

## 1.1 Objetivo

A principal meta da pesquisa é verificar se tempestades elétricas possuem comportamento e características que indiquem se são potencialmente perigosas ao sistema elétrico selecionado. Para isso duas etapas são realizadas: a primeira é a identificação e rastreamento de tempestades que será realizada através da clusterização de descargas atmosféricas através do método *Convergent Data Sharpening*, já a segunda etapa consiste em aplicar o método SOM para reconhecimento de padrões a fim de originar modelos de tempestades que representem bem o grande volume de tempestades identificadas provenientes de 14 anos de dados, e verificar se estes podem estar associados a alguma perturbação no setor elétrico. É importante ressaltar que esta análise de tempestades elétricas será feita somente com informações oriundas dos próprios dados, sem fazer qualquer referência a eventos meteorológicos atuantes e estudos climatológicos na região, o que demandaria maior volume de dados, bem como dados de outra natureza.

## 2 Materiais e Métodos

### 2.1 Região de Estudo e Dados

Conhecer o comportamento e características de sistemas meteorológicos que atuam em regiões que abrangem linhas de transmissão de energia pode significar importante avanço para o setor, pois medidas operacionais podem ser tomadas a fim de minimizar riscos de desligamentos e danos aos equipamentos. Por esta razão a região selecionada para o estudo foi a área que abrange a linha de transmissão que leva a energia produzida na hidrelétrica de Itaipu para a proximidade do centro de consumo da região Sudeste do Brasil, conhecida como LT 765 kV (Figura 1). Este sistema é composto de três linhas de transmissão entre as subestações de Foz do Iguaçu (PR) e Tijuco Preto (SP), cada uma com extensão de aproximadamente 900 km. É o sistema de transmissão de tensão mais elevada existente no Brasil, desde que começou sua operação em 1986 (ITAIPU, 2015).

De 1999 a 2012 foram registradas 29.344.598 descargas nuvem-solo na região que engloba a linha LT 765 kV, provenientes da Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT),

mantidos pelo SIMEPAR em cooperação com Furnas, CEMIG e INPE. Este registro compreende as seguintes informações das descargas atmosféricas: posição geográfica (longitude/latitude), data, polaridade e intensidade do pico de corrente (em kA). Os dados armazenados contemplam apenas descargas com pico de corrente igual ou superior a 10 kA em valor absoluto, visto que intensidades menores que esta indicam descargas intra-nuvens e é sugerido que sejam filtradas para eliminar ruídos nas informações (Beneti, 2012).

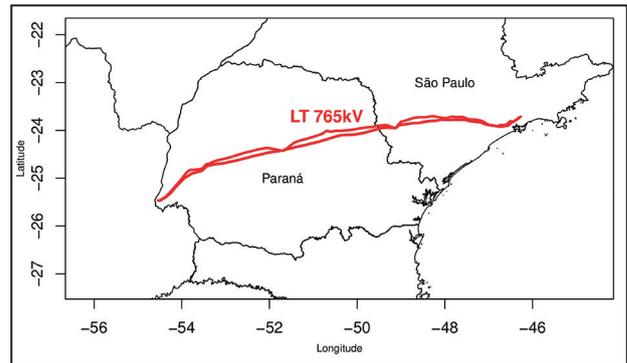


Figura 1 Área de estudo que abrange a linha de energia LT 765 kV (em vermelho).

### 2.2 Método de Clusterização: *Convergent Data Sharpening*

Choi & Hall (1999) introduziram originalmente o método *Data Sharpening* para redução de viés na estimação de densidades. Considere os dados originais  $x_1, x_2, \dots, x_n$  com densidade desconhecida  $f(x)$ . Uma estimativa para a densidade em um ponto  $x$  é dada por

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n k_h(x_i - x) \quad (1)$$

onde  $k_h$  é uma função densidade de probabilidade simétrica, também chamada de função *kernel*, com parâmetro de escala  $h$  (largura da banda).

Choi & Hall (1999) mostram que por regressão constante local de  $x$  em  $x$ , são obtidos os dados ajustados  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$  onde

$$\hat{x}_j = \frac{\sum_{i=1}^n k_h(x_i - x_j) x_i}{\sum_{i=1}^n k_h(x_i - x_j)}$$

e assim a densidade estimada é feita sobre os dados ajustados, dada por

$$\hat{f}_s(x) = \frac{1}{n} \sum_{i=1}^n k_h(\hat{x}_i - x)$$

onde  $\hat{f}_s(x)$  tem menor viés do que  $\hat{f}(x)$ .

Sendo assim, a ideia principal do método é que cada dado original mova-se para mais perto de modos locais, visto que (1) tende a subestimar densidades nos picos e superestimar nos vales. Woolford & Braun (2006) propuseram iterar o algoritmo de perturbação dos dados a fim de apontar esses modos locais como centro de clusters de dados com aderência a eles. No trabalho citado anteriormente, os autores apresentam um teorema juntamente com sua prova, o qual diz que para  $h$  fixo e qualquer vetor inicial de observações  $x_0$ , o algoritmo *Data Sharpening* de Choi & Hall (1999) converge para um único vetor  $\hat{x}$ . A esse método deu-se o nome de *Convergent Data Sharpening*.

Existem diversas funções *kernel* que podem ser utilizadas na estimativa da densidade, tais como as funções gaussiana, *biweight*, triangular, entre outras (Wand & Jones, 1994). Woolford & Braun implementaram o método *Convergent Data Sharpening* no *software* R utilizando a função *kernel* gaussiana, a qual será utilizada nesta pesquisa. A função do R utilizada para clusterização foi a *sharp2dB* do pacote *CHsharp* (R, 2012). O *software* R foi utilizado na pesquisa devido a sua potencialidade na análise de dados e incorporação de rotinas atualizadas e desenvolvidas por usuários e pesquisadores do mundo todo.

Assim, o algoritmo *Convergent Data Sharpening* será utilizado nesta pesquisa como o método de agrupamento de descargas atmosféricas (espacial e temporalmente) a fim de associá-las a tempestades elétricas, podendo estas revelar informações mais importantes sobre eventos meteorológicos do que as próprias descargas de maneira isolada.

### 2.2.1 Escolha do Método de Clusterização

O método *Convergent Data Sharpening* foi escolhido para agrupar as descargas atmosféricas primeiramente pela capacidade do método em lidar com grande volume de dados, visto que não é um problema trivial de se tratar devido à quantidade de informações disponíveis e pelo processo que estas

informações são expostas. Outros fatores decisivos na escolha deste método foi a não imposição prévia do número de clusters e a não formação de clusters com números de pontos e formatos similares, posto que tempestades elétricas não obedecem estes fatores. Adicionalmente, através de experimentos o método também se mostrou eficiente no encadeamento temporal dos clusters para rastreamento das tempestades elétricas.

### 2.2.2 Escolha do Parâmetro de Clusterização

De acordo com Woolford & Braun (2006), se o parâmetro largura de banda no método *Convergent Data Sharpening* é muito pequeno então ruídos não são removidos, e se for muito grande, os pontos começam a aderir ao modo global.

O parâmetro largura de banda espacial utilizado nesta pesquisa foi 0,1° (aproximadamente 10 km), escolhido através de testes experimentais e inspeção visual. Analisando os núcleos das tempestades identificadas com esse valor de parâmetro, nota-se que estes apresentam em média 40 km de largura, o que é compatível com os núcleos das tempestades da região (Beneti, 2012).

### 2.3 Método de Reconhecimento de Padrões: *Self-Organizing Map*

*Self-Organizing Map* é uma rede neural de treinamento não supervisionado que faz uso de uma estrutura topológica entre os dados e os grupos. Essa rede é formada por uma estrutura de neurônios, onde cada dado tem seus atributos conectados a todos os neurônios da rede. Um peso inicialmente aleatório é designado entre a ligação neurônio-atributo. Os pesos dos neurônios servem como modelo de padrão de entrada para os grupos. Durante o treinamento, o neurônio cujos pesos são mais próximos do dado apresentado à rede é denominado “neurônio vencedor” e é escolhido para representá-lo. O neurônio vencedor e sua vizinhança (determinada por um raio) com relação ao mapa tem seus pesos atualizados com o intuito de se especializar na identificação de atributos, embora a vizinhança não possua necessariamente proximidade dos pesos em relação ao padrão de entrada (Kohonen, 1990, 2001).

## Algoritmo

0. Iniciar os pesos dos  $n$  neurônios da rede com valores aleatórios baixos:  $w_{ij}$

1. Apresentar cada entrada  $\mathbf{x}$  para a rede, e executar os passos 2 e 3:

2. Determinar o neurônio vencedor  $i$  que possui a menor distância do peso sináptico  $w_j$  com o vetor .

$$d_i = \sum_{j=1}^n (x_j - w_{ij})^2$$

3. Ajustar os pesos do neurônio vencedor e de todos os neurônios que pertencem a uma vizinhança centrada nele,  $V_i$ .

$$w_{ij}^{atual} = w_{ij}^{anterior} + \alpha(x_j - w_{ij}^{anterior})$$

onde  $i \in V_i$

4. Ajustar a taxa de aprendizado  $\alpha$  e o raio de vizinhança. Se não existirem mais mudanças substanciais no mapa, pare; caso contrário, retorne ao passo 1.

No *software* R, foi utilizada a função *som* do pacote *Kohonen*. Esta implementação utiliza os seguintes parâmetros para realizar o treinamento da rede: pesos iniciais aleatórios, apresentação aleatória das amostras na entrada da rede, distância euclidiana quadrática para calcular desigualdades entre neurônios, função de vizinhança do tipo circular centrada no neurônio vencedor e definida pelo raio de vizinhança, decréscimos lineares para o raio de vizinhança e taxa de aprendizagem, número de épocas de treinamento igual a 100, e formato hexagonal para o arranjo bidimensional de neurônios (Kohonen, 2001). O tamanho do mapa (número de neurônios) é outro parâmetro fundamental para a definição do mapa e será discutido posteriormente.

## 3 Projeto de Aplicação

### 3.1 Identificação das Tempestades

A identificação das tempestades elétricas foi feita através da clusterização das descargas atmosféricas pela posição espacial através do

método *Convergent Data Sharpening* sendo considerada janela móvel de uma hora com passo de dez minutos. Em outras palavras, um conjunto de descargas é selecionado e clusterizado no período de uma hora, e após a clusterização a nova janela temporal considerada é a anterior com dez minutos eliminados no início e dez minutos acrescidos no final.

Este processo de clusterização origina retratos temporais, visto que uma mesma descarga está presente em seis clusters distintos. Estes retratos temporais devem ser conectados (através de alguma medida de similaridade) uns aos outros para dar origem às tempestades. Na pesquisa, clusters de janelas consecutivas são unidos se a velocidade de deslocamento de um núcleo para outro não ultrapassa um limite máximo, estipulado em 40 km/10min ou equivalentemente 240 km/h. Em outras palavras, em 10 minutos é permitido um deslocamento dos centros dos clusters de janelas consecutivas de até 40 km. Esta velocidade é superior à encontrada na literatura para a região (Beneti, 2012) devido à instabilidade do processo, sendo assim admissível se considerar esta folga. Feito isso é concluído o procedimento de identificação e rastreamento das tempestades elétricas.

#### 3.1.1 Fusões e Cisões de Tempestades

Os processos de fusão e cisão entre tempestades também é considerado. Se duas ou mais tempestades estão ativas no tempo  $t_1$  e no tempo  $t_2$  elas se fundem, o histórico da nova tempestade provém daquela cujo centro é mais próximo do centro da nova tempestade.

Se uma tempestade está ativa em  $t_1$  porém em  $t_2$  ela se desmembra em duas ou mais tempestades, o histórico da tempestade velha é mantido àquela nova tempestade cujo centro está mais próximo do centro da tempestade velha. As tempestades restantes iniciarão um novo histórico.

#### 3.1.2 Resultados e Simplificações

Como resultado da metodologia descrita acima, foram identificadas 636.012 tempestades elétricas com os 14 anos de dados na região da LT 765 kV.

Assim como feito em Dixon & Wiener (1993), algumas tempestades foram descartadas do estudo. São elas: tempestades que percorreram um caminho pequeno ou ficaram paradas, geralmente compostas por poucas descargas; tempestades cujo número máximo de descargas contabilizadas não foi superior a 20; e principalmente tempestades triviais, que não duraram mais do que uma hora. Assim, o número de tempestades reduziu-se a 130.120. Com a finalidade de se analisar algumas peculiaridades, três cenários de estudo foram criados: *Incondicional*, *Linhão* e *Falhas*.

- *Incondicional*: são todas as tempestades que atuaram na região de estudo, totalizando 130.120 tempestades;

- *Linhão*: são as tempestades que se desenvolveram nas proximidades da linha de energia e a elipse de 95% de confiança das tempestades contém um trecho da linha. A quantidade de tempestades classificadas neste cenário foi 16.870;

- *Falhas*: são as tempestades que contém descargas possivelmente associadas a falhas no sistema elétrico ao longo do período estudado. Estas descargas ocorreram próximas temporalmente das falhas e também nas imediações do desligamento da LT 765 kV. Com isso, 113 tempestades foram classificadas neste cenário.

Analisar e explorar uma grande quantidade de dados pode não ser uma tarefa fácil sendo muitas vezes impossível obter conclusões relevantes. É o que acontece quando se quer examinar as características de milhares de tempestades elétricas. Visualmente elas representam um emaranhado de trajetórias, de todos os possíveis comprimentos, direções e sentidos, de uma forma desordenada. A vista disso buscou-se uma estratégia que pudesse reduzir significativamente a complexidade da análise dos dados, auxiliando visualmente a detecção de padrões de tempestades: o método SOM.

### 3.2 Reconhecimento de Padrões de Tempestades

Uma tempestade, até então, é representada por uma sequência de pontos (longitude e latitude) que são os centros dos clusters no momento de sua formação, representados pelas setas na Figura 2. O número de pontos é o mesmo que a quantidade de clusters que foram unidos para a caracterização

da tempestade, ou seja, se uma tempestade durou 80 minutos provinda da conexão de 8 clusters consecutivos no tempo (10 minutos cada passo de tempo), então o número de pontos (setas na Figura 2) será 8. Assim, cada tempestade possui um número próprio de pontos que a representa. Como as entradas para o método SOM devem ter a mesma dimensão, buscou-se outra forma de representação das tempestades.

Foi realizado um ajuste polinomial de terceiro grau, no tempo, da trajetória de cada tempestade para  $x$  (longitude) e  $y$  (latitude) de forma independente:

$$x = a_0 + a_1t + a_2t^2 + a_3t^3$$

$$y = b_0 + b_1t + b_2t^2 + b_3t^3$$

onde

$$t = t_i, i = 1, \dots, n, t \in [0, 1] \text{ e } \Delta t = 10 \text{ minutos}$$

A Figura 2 mostra um exemplo de uma trajetória e sua representação pelo polinômio cúbico ajustado. Os pontos inicial e final da trajetória real e do modelo ajustado são coincidentes.

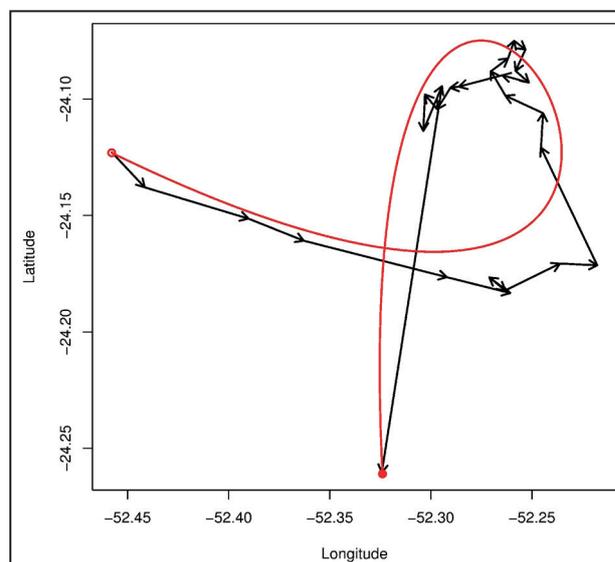


Figura 2 Exemplo de trajetória de tempestade (linha preta com setas) e sua representação polinomial (linha vermelha), onde os círculos vazado e preenchido são o início e o fim do modelo ajustado que coincidem com o início e o final da trajetória real.

Com um modelo ajustado para cada trajetória de tempestade, foram obtidos quatro pontos equidistantes temporalmente, sendo dois deles os pontos inicial e final. Obtidos estes quatro pontos,

pôde-se calcular o comprimento das três semi-retas que ligam os pontos e os três ângulos entre o eixo y e estas semi-retas no sentido horário, conforme ilustrado na Figura 3.

Portanto, a entrada para o método SOM, para representar cada uma das tempestades, é composta das seguintes informações: primeiro ponto (ponto inicial), distância entre o primeiro e o segundo ponto, ângulo entre a direção Norte e a semi-reta que liga o primeiro ao segundo ponto, distância entre o segundo e o terceiro ponto, ângulo entre a direção Norte e a semi-reta que liga o segundo ao terceiro ponto, distância entre o terceiro e quarto ponto (ponto final) e finalmente o ângulo entre direção Norte e a semi-reta que liga o terceiro ao quarto ponto. Desta forma, tempestades elétricas representadas pelo mesmo padrão devem ter origens próximas e suas trajetórias serem similares.

Desta forma, a representação matricial da entrada do método SOM é:

$$\begin{bmatrix} x_{0_1} & y_{0_1} & distância_{1_1} & \hat{a}_{1_1} & distância_{2_1} & \hat{a}_{2_1} & distância_{3_1} & \hat{a}_{3_1} \\ x_{0_2} & y_{0_2} & distância_{1_2} & \hat{a}_{1_2} & distância_{2_2} & \hat{a}_{2_2} & distância_{3_2} & \hat{a}_{3_2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{0_n} & y_{0_n} & distância_{1_n} & \hat{a}_{1_n} & distância_{2_n} & \hat{a}_{2_n} & distância_{3_n} & \hat{a}_{3_n} \end{bmatrix}$$

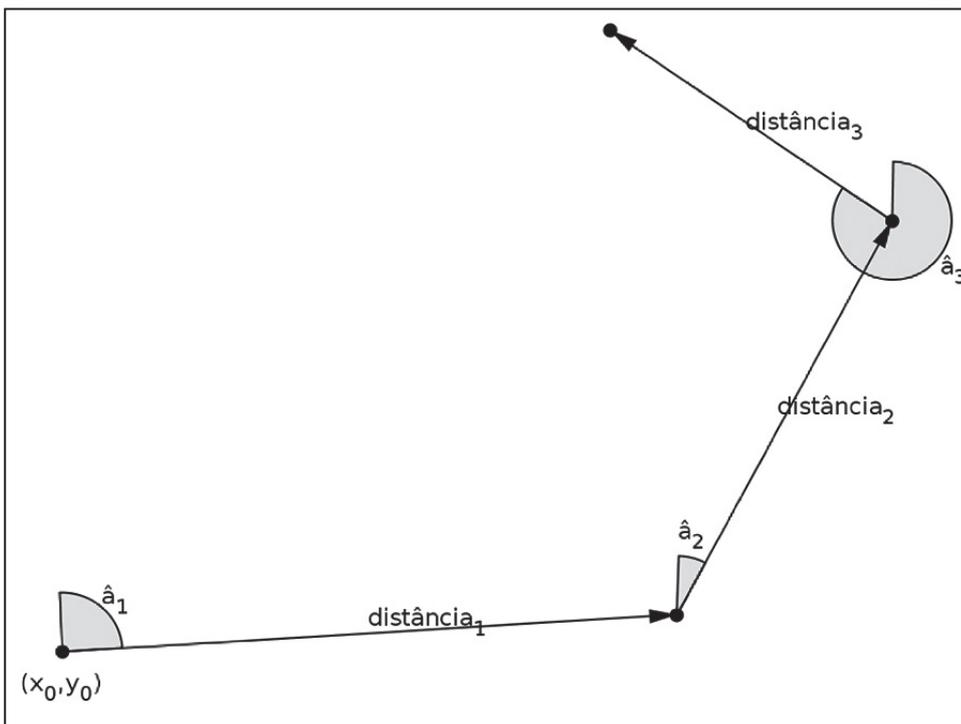


Figura 3 Ilustração das variáveis de entrada que representam as trajetórias das tempestades para o método SOM.

onde cada linha desta matriz representa uma tempestade,  $n=130.120$  para o cenário *Incondicional*,  $n=16.870$  para o cenário *Linhão* e  $n=113$  para o cenário *Falhas*.

Por serem de grandezas diferentes, os dados foram normalizados através da regra Min-Max:

$$\frac{z - z_{min}}{z_{max} - z_{min}}$$

onde  $z$  representa cada coluna da matriz e  $z_{min}$  e  $z_{max}$  são os valores mínimos e máximos de cada coluna, respectivamente. Assim os dados de entrada para o SOM sempre estarão entre 0 e 1.

### 3.2.1 Escolha do Número de Padrões

A escolha do número de padrões para cada cenário foi realizada através da maximização do *Bayesian Information Criterion* (BIC), que consiste na seleção de modelos, entre um conjunto finito, baseado na verossimilhança (Fraley & Raftery, 1998).

Por meio do método BIC, o número de padrões resultantes para o cenário *Incondicional* foi 220, para o cenário *Linhão* foi 78 e para o cenário *Falhas* foi 40. Como dito anteriormente, o tipo de grade utilizada no SOM foi a hexagonal, em sua forma mais quadrada possível, isto é, 220 padrões = 22x10, 78 padrões = 13x6 e 40 padrões = 8x5.

Após obtidos os padrões de tempestades, retomou-se a representação destes padrões pelos seus polinômios cúbicos aproximados, para melhor visualização e análise dos resultados.

#### 4 Resultados e Discussões

Identificados os padrões de tempestades elétricas para os três cenários, tem-se uma simplificação e redução nos dados a serem analisados, estes sim mais plausíveis a revelar informações importantes do comportamento das tempestades que atuam nas proximidades da LT 765 kV. A Figura 4 mostra 3 exemplos (dos 220 padrões definidos no cenário *Incondicional*) da representação de tempestades com trajetórias similares por seus respectivos padrões obtidos pelo SOM.

Duas importantes características foram tomadas no estudo: o número médio de descargas e o pico de corrente médio das tempestades que compõem cada padrão estruturado. O número médio de descargas foi calculado da seguinte forma: para cada tempestade foi contabilizado o número de descargas total durante sua vida e para o padrão foi atribuído a média do número de descargas de cada tempestade representada por este padrão. Em relação ao pico de corrente, a cada tempestade atribuiu-se um pico de corrente médio formado pela média dos valores absolutos dos picos das descargas que constituem a tempestade; com isso para o padrão faz-se a média destas médias dos picos das tempestades que são representadas por este padrão. As figuras a seguir ilustram as trajetórias dos padrões de tempestades elétricas gerados pelo SOM para os três cenários estudados para as duas características referidas. Previamente é possível notar que a tendência do deslocamento das tempestades é basicamente do oeste para o leste, o que é coerente com estudos climatológicos de tempestades realizados na região (Sakamoto, 2009; Beneti, 2012).

A Figura 5 apresenta o pico de corrente médio (em kA) dos padrões de tempestades para o cenário *Incondicional*, onde nota-se maior intensidade

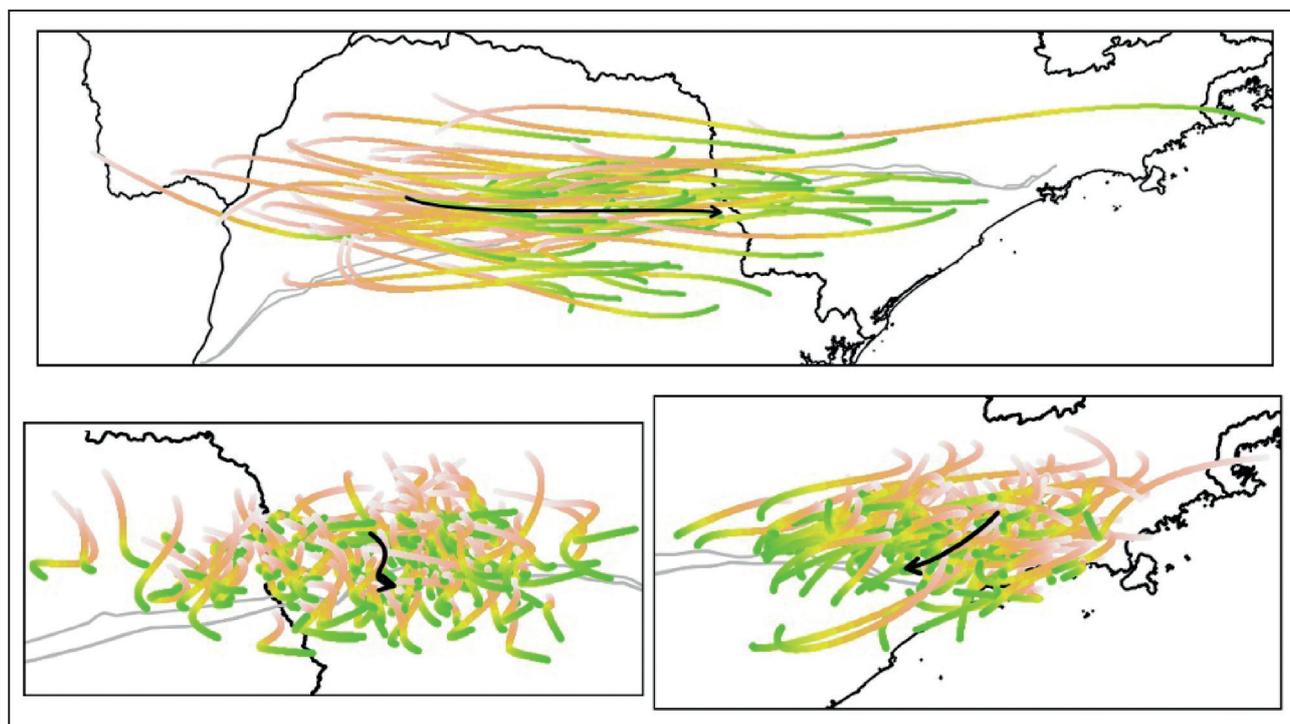


Figura 4 Ilustração de 3 dos 220 padrões (seta preta) do cenário *Incondicional* providos pelo SOM que representam tempestades reais (rosa claro: início; verde: fim da tempestade).

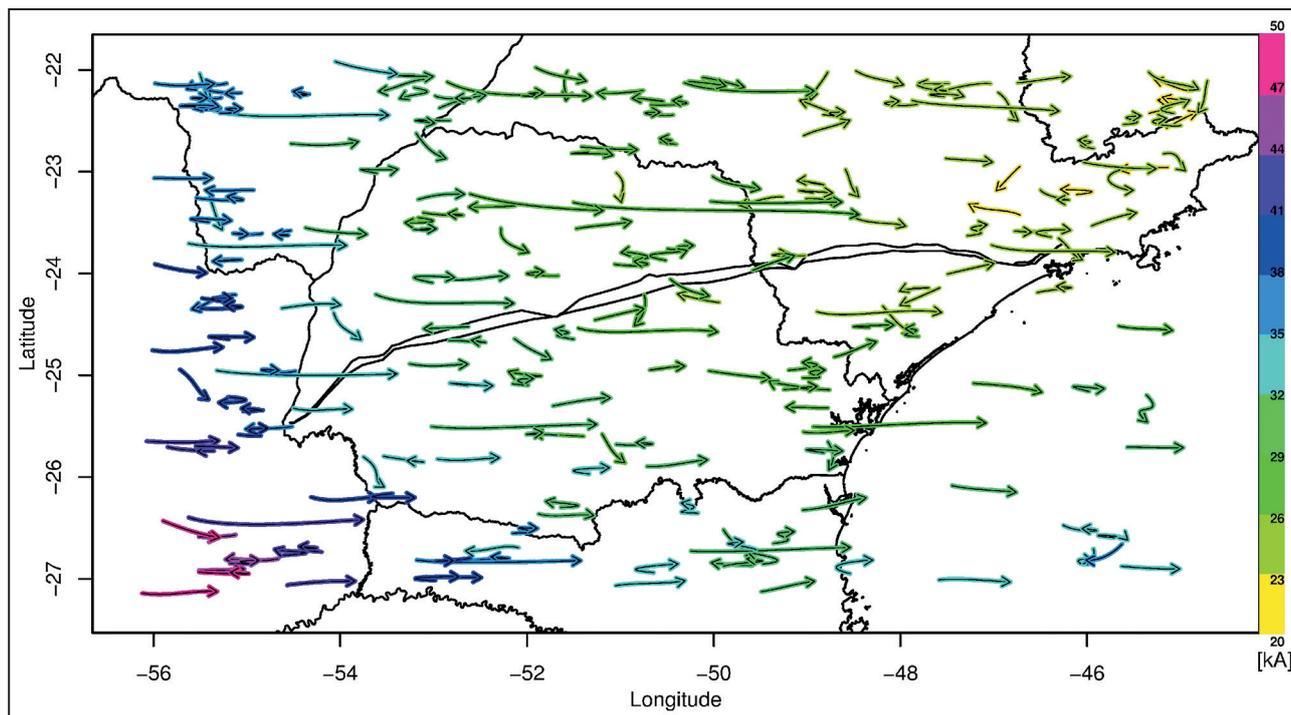


Figura 5 Pico de corrente médio dos padrões de tempestades para o cenário *Incondicional*.

desta variável no sudoeste da região, ocorrendo tempestades que registraram média de pico de corrente de 50 kA em valor absoluto, enfraquecendo a medida que se caminha no sentido nordeste.

O pico de corrente médio dos padrões de tempestades para o cenário *Linhão* (Figura 6) tem comportamento similar ao do cenário *Incondicional*. Na cidade de Foz do Iguaçu, onde se inicia a LT 765 kV, é onde ocorre os casos de padrões de tempestades com maior pico de corrente médio.

Já no cenário *Falhas*, os padrões de tempestades com maior pico de corrente médio se apresentam dispersos pela região (Figura 7). Nota-se que não se constatou maior intensidade do pico de corrente médio em comparação com os outros cenários, como poderia ser esperado (Westinghouse, 1964). Assim, os padrões mostram que, para os três cenários, a intensidade do pico de corrente das tempestades é similar, não havendo expressivas diferenças.

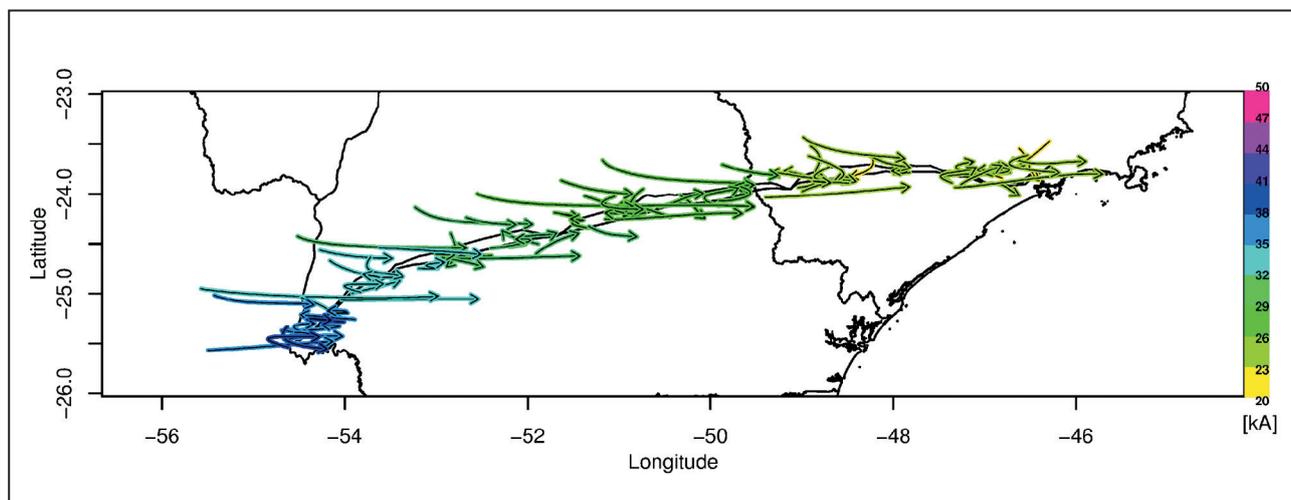


Figura 6 Pico de corrente médio dos padrões de tempestades para o cenário *Linhão*.

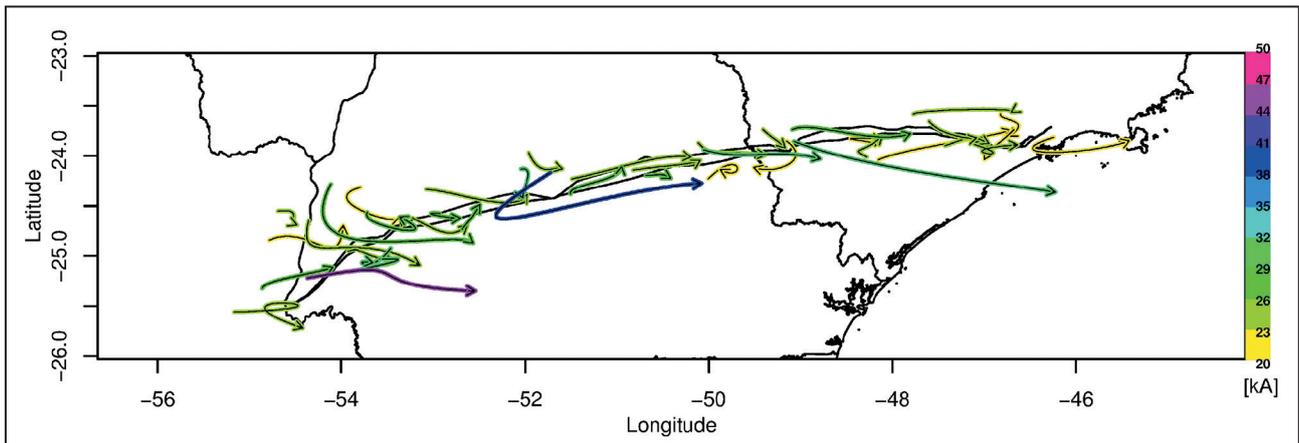


Figura 7 Pico de corrente médio dos padrões de tempestades para o cenário *Falhas*.

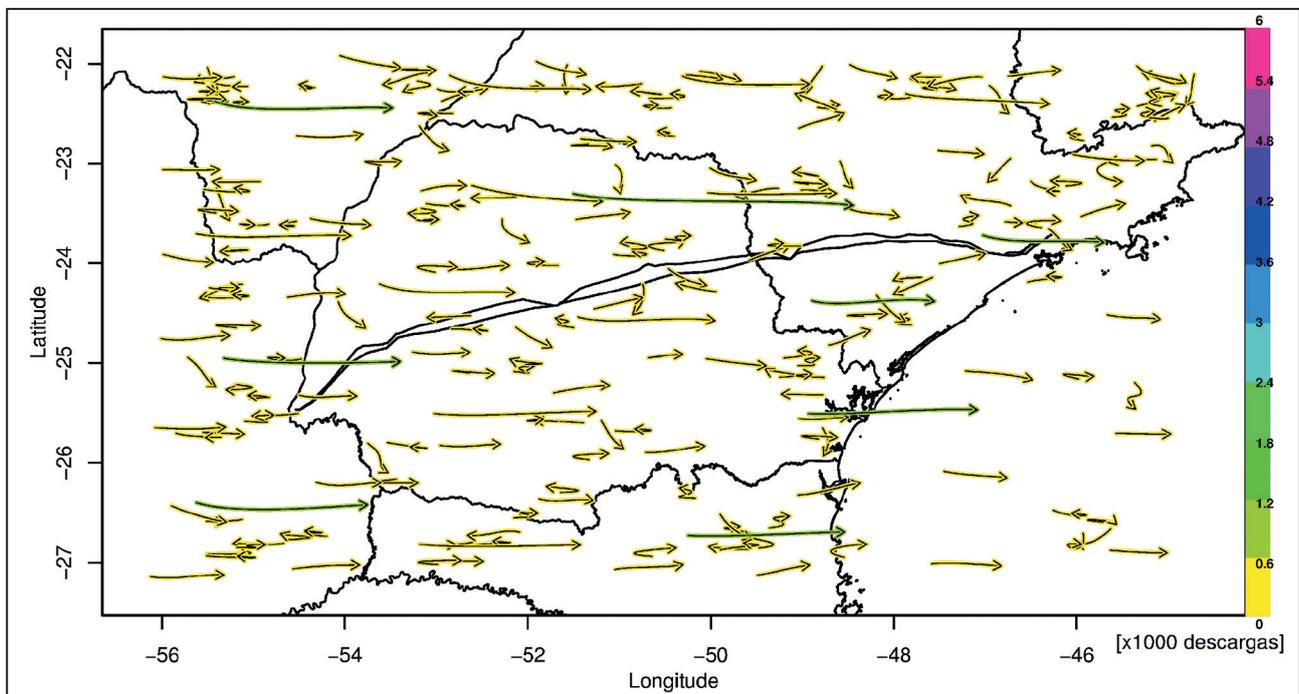


Figura 8 Número de descargas médio dos padrões de tempestades para o cenário *Incondicional*.

A Figura 8 mostra as trajetórias das tempestades com a representação do número médio de descargas para o cenário *Incondicional*. Nos 14 anos estudados, o número máximo contabilizado em um padrão de tempestade foi de 1.800 descargas em média (setas de cor verde na Figura 8).

No cenário *Linhão* a situação não é muito diferente (Figura 9) do cenário *Incondicional*. O número médio de descargas para as tempestades desta classe é semelhante ao caso anterior.

Entretanto no cenário *Falhas* é possível notar que houve registro de padrões de tempestades com

número médio de descargas superior aos demais cenários (Figura 10). Visualmente é notado que tempestades que podem ter ocasionado falhas neste sistema elétrico tendem a possuir mais descargas do que tempestades que não foram apontadas como possíveis causadoras de falhas.

A Figura 11 apresenta um diagrama de caixas (Braga, 2010), uma forma de representação que identifica onde estão localizados 50% dos dados (a base da caixa representa o quartil inferior e o topo da caixa o quartil superior dos valores observados), a mediana, a média e valores extremos (segmentos

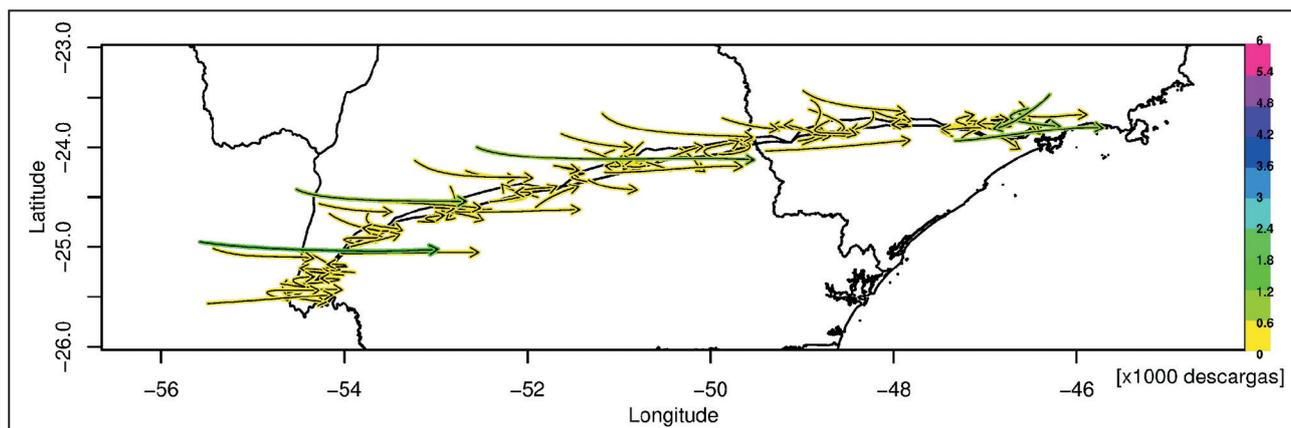


Figura 9 Número de descargas médio dos padrões de tempestades para o cenário *Linhão*.

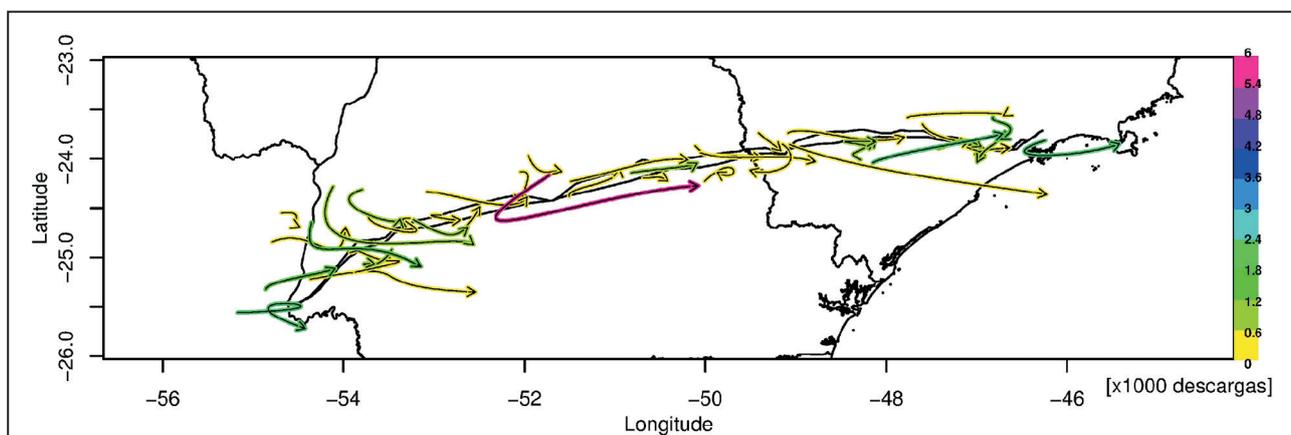


Figura 10 Número de descargas médio dos padrões de tempestades para o cenário *Falhas*.

de retas conectados ao topo e base da caixa que representam 1,5 vezes a diferença entre o quartil inferior e o superior). Nesta figura são reveladas características da distribuição do conjunto das tempestades e respectivos padrões, para as variáveis pico de corrente médio e número de descargas, respectivamente. Conforme era esperado, a dispersão dos dados é maior no conjunto das tempestades, mas os padrões capturaram as medidas centrais (média e mediana), indicando a boa representação das tempestades pelos seus padrões.

Utilizando a técnica *Bootstrap* (Davison & Hinkley, 1997), foram analisadas as distribuições dos dados em questão, para investigar se há diferenças significativas entre os picos de corrente e também entre os números de descargas nos cenários estudados. Como resultados, não foi constatado que o pico de corrente no caso *Falhas* é maior que nos demais cenários e também foi confirmado que o número de descargas em *Falhas* é maior do que nos outros cenários, ao nível de confiança de 99%.

## 5 Conclusões

Nesta pesquisa fez-se uso de diversas ferramentas matemáticas para tratar e analisar uma grande quantidade de dados de descargas atmosféricas incidentes nas proximidades de uma linha de energia elétrica selecionada. Tal região foi escolhida pelo fato que, muitas vezes, descargas são responsáveis por desligamentos não-programados em linhas de energia e danos em equipamentos, acarretando em grandes prejuízos financeiros para consumidores e companhias que administram as linhas. Conhecer o comportamento dos sistemas meteorológicos que agem nesta região pode resultar em tomadas de decisões que minimizem estes danos citados.

Inicialmente foi realizada uma organização espacial e temporal dos dados de descargas atmosféricas, originando as chamadas tempestades elétricas e então pôde-se associar tempestades similares (em trajetórias) em seus respectivos padrões de tempestades. Duas variáveis dos padrões

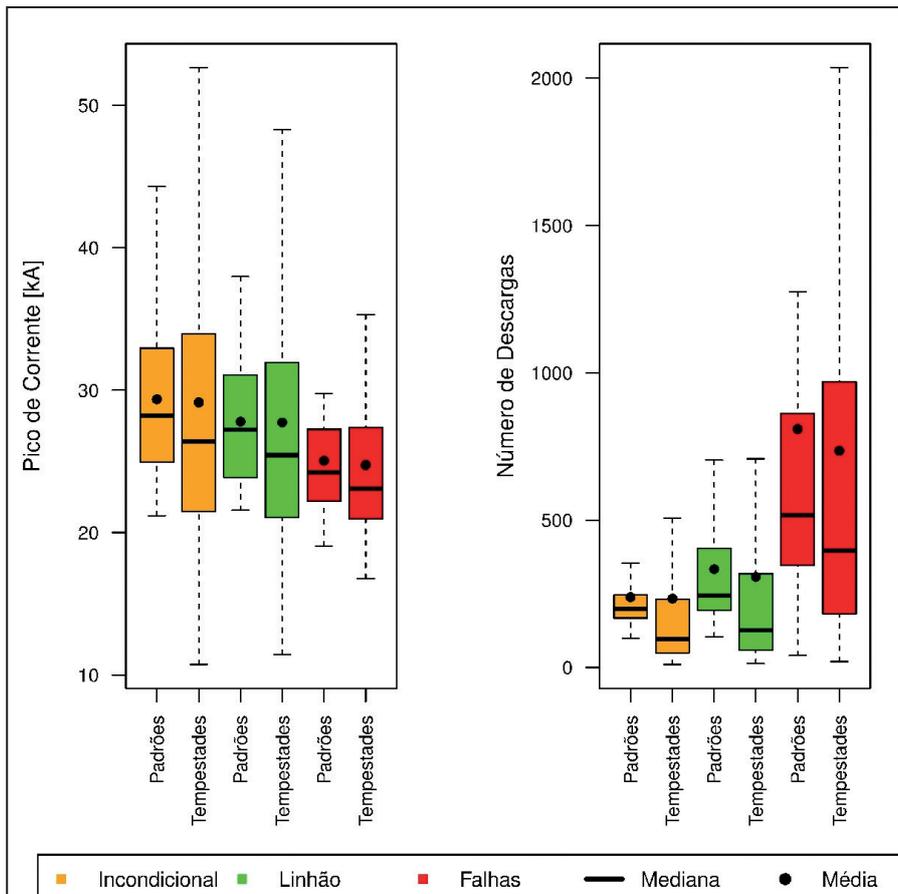


Figura 11 Comparação do pico de corrente médio e do número de descargas médio das tempestades e dos respectivos padrões para os três cenários estudados.

de tempestades foram investigadas a fim de averiguar se estas tem relação com falhas no sistema elétrico: pico de corrente e número de descargas.

Com relação a variável pico de corrente, notou-se que tempestades em situações de falhas tem magnitude similar a situações onde não se observaram falhas, sendo sudoeste a região com tempestades elétricas de maior intensidade desta variável. Em se tratando do número de descargas, a circunstância é outra: quando há falhas no sistema elétrico, as tempestades apresentaram maior número de descargas do que em situações sem falhas.

Esta pesquisa revelou a efetiva representação das tempestades elétricas por meio da clusterização de descargas atmosféricas com o método *Convergent Data Sharpening* e também a robustez e o hábil desempenho do SOM para identificação de padrões de tempestades. A organização de tempestades com trajetórias semelhantes em padrões não permitiu a constatação de que tempestades com maior potencial de risco a falhas do sistema elétrico que

abrange a LT 765 kV são aquelas que possuem maior pico de corrente, mas sim aquelas com maior número de descargas.

Todos os resultados obtidos nesta pesquisa procederam de aplicações de ferramentas matemáticas dos próprios dados de descargas, isentando qualquer consideração a eventos atuantes já estudados e conhecidos. Sendo assim, as conclusões do trabalho são intrínsecas das descargas atmosféricas, revelando a potencialidade e alertando para estudos mais aprofundados deste fenômeno natural, especialmente em regiões onde seu impacto pode causar grandes danos, como é o caso de linhas de energia.

## 6 Referências

- Beneti, C.A.A. 2012. *Caracterização Hidrodinâmica e Elétrica de Sistemas Convectivos de Mesoescala*. Programa de Pós Graduação em Meteorologia, Universidade de São Paulo, Tese de Doutorado, 131 p.
- Braga, L.P.V. 2010. *Compreendendo Probabilidade e Estatística*.

- Rio de Janeiro, E-papers, 230 p.
- Choi, E. & Hall, P. 1999. Data sharpening as a prelude to density estimation. *Biometrika*, 86(4): 941-947.
- Davison, A.C. & Hinkley, D.V. 1997. *Bootstrap Methods and their Application*. New York, Cambridge University Press, 582 p.
- Dixon, M. & Wiener, G. 1993. TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting – A Radar-based Methodology. *Journal of Atmospheric and Oceanic Technology*, 10(6): 785-797.
- Fraley, C. & Raftery, A. E. 1998. How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal*, 41(8): 578-588.
- INPE – Instituto Nacional de Pesquisas Espaciais. *Sistema Elétrico*. Disponível em: <[http://www.inpe.br/webelat/homepage/menu/infor/relampagos.e.efeitos/sistema\\_eletrico.php](http://www.inpe.br/webelat/homepage/menu/infor/relampagos.e.efeitos/sistema_eletrico.php)>. Acesso em: 05 fevereiro 2015.
- ITAIPU. *Integração ao Sistema Brasileiro*. Disponível em: <<http://www.itaipu.gov.br/energia/integracao-ao-sistema-brasileiro>>. Acesso em: 20 fevereiro 2015.
- Jensen, A.A. & Nowotarski, C. 2012. *Application of a self-organizing map statistical technique to a RUC supercell proximity sounding database*. In: CONFERENCE ON SEVERE LOCAL STORMS, 26, Nashville, Tennessee, American Meteorological Society.
- Kohonen, T. 1990. The Self-Organizing Map. *Proceedings of the IEEE*, 78(9): 1464-1480.
- Kohonen, T. 2001. *Self-organizing maps*. Berlin, Springer, 458 p.
- R Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <<http://www.R-project.org/>>.
- Sá, J.A.; Rocha, B.; Almeida, A. & Souza, J.R. 2012. Recurrent Self-Organizing Map for Severe Weather Patterns Recognition. In: ELHEFNAWI, M. & MYSARA, M. (eds.). *Recurrent Neural Networks and Soft Computing*. Editora Intech, p. 151-174.
- Sakamoto, M.S. 2009. *Sistemas Convectivos de Mesoescala observados na Região Subtropical da América do Sul durante o SALLJEX*. Programa de Pós Graduação em Meteorologia, Universidade de São Paulo, Tese de Doutorado, 243 p.
- Sang, B.H.; Gelfand, A.E.; Lennard, C.; Hegerl, G. & Hewitson, B. 2008. Interpreting Self-Organizing Maps Through Space-Time Data Models. *The Annals of Applied Statistics*, 2(4): 1194-1216.
- Sheridan, S.C. & Lee, C.C. 2011. The self-organizing map in synoptic climatological research. *Progress in Physical Geography*, 35(1): 109-119.
- Westinghouse, C.S.E. 1964. *Electrical transmission and distribution reference book*. Westinghouse, Pennsylvania, 824 p.
- Woolford, D.G. & Braun, W.J. 2006. Convergent data sharpening for identification and tracking of spatial and temporal centers of lightning activity. *Environmetrics*, 18: 461-479.
- Wand, M.P. & Jones, M.C. 1994. *Kernel Smoothing*. CRC Press, 224 p.