

IDENTIFICAÇÃO, MONITORAMENTO E PREVISÃO DE TEMPESTADES ELÉTRICAS UTILIZANDO MÉTODOS NUMÉRICOS

Identification, Tracking and Nowcasting of Electrical Storms Using Numerical Methods

Mariana Kleina ¹

Luiz Carlos Matioli ¹

Eduardo Alvim Leite ²

¹ Universidade Federal do Paraná – UFPR, Programa de Pós-Graduação em Métodos Numéricos em Engenharia – PPGMNE. Centro Politécnico da Universidade Federal do Paraná, Curitiba, Paraná, Brasil.

² Sistema Meteorológico do Paraná – SIMEPAR. Centro Politécnico da Universidade Federal do Paraná, Curitiba, Paraná, Brasil.

Email: marianakleina11@gmail.com; matioli@ufpr.br; alvim@simepar.br

Resumo:

Nesta pesquisa é apresentado um sistema de identificação, monitoramento e previsão de tempestades elétricas baseado em clusterização de descargas atmosféricas e conexões espaciais/temporais de núcleos elétricos detectados. Durante o processo, diversas variáveis meteorológicas são calculadas e então projetadas uma hora à frente através da técnica de extrapolação de dados. A validação da metodologia proposta é feita por meio da comparação dos dados observados com a previsão realizada. A metodologia é validada e aplicada em uma região que abrange a linha de transmissão de energia elétrica mais importante do Brasil, bastante vulnerável às condições climáticas, afetada especialmente por descargas atmosféricas. Exemplos práticos são apresentados a fim de visualizar o produto do sistema proposto, onde duas situações distintas são analisadas.

Palavras-chave: Descargas Atmosféricas; Tempestades Elétricas; Previsão a Curto Prazo; Variáveis Meteorológicas.

Abstract:

This research presents a method of identification, tracking and nowcasting of electrical storms based on lightning clustering and spatial/temporal connections of electrical centers detected. During the process, different meteorological variables are calculated and then projected one hour ahead by data extrapolation technique. The validation of the proposed methodology is made by comparing the observed data with the prediction held. The methodology is validated and applied on a region covering the more important brazilian electric power transmission line, highly vulnerable to weather conditions, especially affected by lightning strokes. Practical examples are presented in order to view the product of the proposed system, where two different situations are

analyzed.

Keywords: Lightning Strokes; Electrical Storms; Short-Term Forecast; Meteorological Variables.

1. Introdução

Tempestades acompanhadas de ventos fortes, alta precipitação, granizo e/ou descargas atmosféricas podem provocar inúmeros danos a vida terrestre, sobretudo em localidades onde estas condições são favoráveis em virtude de fatores climáticos e territoriais como é o caso do Brasil. Em se tratando de descargas atmosféricas, o Brasil é líder mundial em ocorrências, devido ao seu clima predominantemente tropical e grande dimensão territorial. Particularmente, as descargas que atingem o solo são responsáveis por numerosos danos em diversos setores econômicos, além de provocar prejuízos e mortes de pessoas.

Se um sistema é capaz de prever, com certo período de antecedência, a ocorrência de um evento severo em determinada localidade, tomadas de decisão podem ser realizadas e danos podem ser evitados. Por exemplo, se no setor elétrico é sabido que uma tempestade elétrica se aproxima de linhas de transmissão, subestações e/ou transformadores, medidas preventivas podem ser realizadas a fim de evitar situações de má operação ou agravamento de interrupções sistêmicas.

Diversas pesquisas no âmbito de monitoramento e previsão de tempestades foram realizadas ao longo dos últimos anos, algumas utilizando dados de descargas atmosféricas e outras utilizando dados de radar. O algoritmo TITAN (*Thunderstorm Identification, Tracking, Analysis, and Nowcasting*) desenvolvido por Dixon e Wiener (1993) é um conhecido método de monitoramento e previsão a curto prazo com base em dados de radar. Tempestades são identificadas por regiões contíguas que excedem limites de refletividade e volume; seu trajeto é traçado através da resolução de um problema de otimização que visa a minimização da distância percorrida e a diferença no volume, fundamentada na hipótese de que em pequenos intervalos de tempo uma tempestade não caminha longas distâncias e não altera muito sua forma. A previsão de uma tempestade é feita para algumas variáveis baseada em um modelo de regressão linear com ponderação exponencialmente decrescente dos pesos.

Outro algoritmo que se baseia apenas em dados de radar (composição de imagens de três radares da Suíça) para detecção, monitoramento e previsão de sistemas convectivos é o TRT (*Thunderstorms Radar Tracking*) criado por Hering *et al.* (2004), cuja identificação é realizada com a utilização de três limiares para a refletividade, o monitoramento é feito por sucessivas imagens de radar levando em conta a velocidade do deslocamento da célula de tempestade bem como área de abrangência, e finalmente a previsão do movimento das tempestades de até uma hora é realizada por meio da extrapolação da velocidade de deslocamento ponderada.

Em Steinacker *et al.* (2000) são apresentados métodos automáticos de monitoramento de células convectivas e complexas na região da Áustria, que fazem uso de dados de radar e de descargas atmosféricas do tipo nuvem-solo, respectivamente. Falhas no monitoramento de células convectivas ocorrem devido a frequentes colapsos nas estações de radar, e por esse motivo dados de descargas também são demandados. O princípio da metodologia empregada é a utilização de

um filtro gaussiano, cuja largura é definida por um parâmetro, em que valores pequenos para esse parâmetro resultam na identificação de células convectivas e valores maiores detectam somente células complexas. Para conectar células de períodos consecutivos, é calculado o melhor vetor deslocamento para cada célula detectada. Dois campos discretos são comparados quando criada uma malha retangular com medições em intervalos de tempo: densidade de descargas e taxa de precipitação. Casos de estudo são retratados onde são evidenciadas as semelhanças e discrepâncias dos dois sistemas empregados.

Dados de descargas nuvem-solo e de radar também são utilizados no trabalho de Bonelli e Marcacci (2008) para criar dois sistemas capazes de prever e emitir alertas quando tempestades severas se aproximam do norte da Itália. A cada intervalo de tempo, são contabilizados o número de descargas e precipitação para um certo ponto de grade (são criadas duas malhas de tamanhos diferentes para os diferentes tipos de dados). Uma célula de tempestade é identificada por meio de um procedimento iniciando do máximo valor calculado em toda a grade (dados de radar ou descargas) e buscam-se pontos ao redor do ponto analisado que configurem a mesma célula de tempestade. Quando o próximo valor do ponto encontrado na estrutura não ultrapassa um limite fixado, então a célula de tempestade é definida e diversos atributos podem ser calculados. Outras células são detectadas da mesma forma, retirando-se da malha os pontos já pertencentes a outras células de tempestades. Exemplos são apresentados e diferenças entre os sistemas são justificadas devido às incertezas na determinação no núcleo da célula para o cálculo dos atributos das tempestades, especialmente nos casos com pequeno número de descargas e/ou grandes células de tempestades. Extrapolação para 30 minutos à frente da posição utilizando a velocidade calculada fornece a previsão de deslocamento da célula de tempestade.

Betz *et al.* (2008) utilizam dados de descargas atmosféricas dos tipos intra-nuvens e nuvens-solo para identificar e monitorar células de tempestades em uma região da Alemanha. A identificação é realizada por meio de densidade de descargas, em intervalos de tempo, em que uma célula é dita suspeita quando o número de descargas por área supera um valor mínimo, e do mesmo modo bordas são reconhecidas quando este número fica abaixo deste mínimo. Uma célula se relaciona com células do intervalo de tempo anterior se velocidade e localização esperadas estão dentro de limites selecionados. É mencionado que uma simples extrapolação baseada em três ou quatro intervalos de tempo consecutivos de cinco ou dez minutos produzem previsões eficientes, especialmente quando tempestades apresentam vida longa.

Descargas nuvem-solo são os dados utilizados no trabalho de Strauss *et al.* (2013) para identificar e rastrear células de tempestades eletricamente ativas em uma região do estado de São Paulo. Fez-se uso de clusterização espacial e temporal de descargas para detectar diferentes células de atividades elétricas e então é aplicado um método de estimação de densidade chamado KDE (*Kernel Density Estimation*) a cada cluster, resultando em um conjunto de campos de densidade de ocorrências de descargas.

Woolford e Braun (2007) desenvolveram uma metodologia para mapear a probabilidade da ocorrência de incêndios florestais em Ontário utilizando informações de descargas atmosféricas nuvem-solo. A metodologia consiste em clusterizar as descargas no espaço e no tempo a fim de identificar centros de tempestades elétricas e assim associá-los a incêndios florestais registrados. O método de clusterização foi o *Convergent Data Sharpening* que se baseia em um ajuste dos dados tornando-os mais agrupados para perto de modas locais (centros das tempestades), amenizando a falta e/ou incorreta localização dos dados de descargas. Também é apresentado o processo de rastreamento de tempestades elétricas através do trajeto que os centros dos clusters percorrem no espaço e no tempo.

Este último trabalho citado foi a base para o começo e desenvolvimento da presente pesquisa, pela similaridade de dados, analogia de propósito que é identificação e rastreamento de tempestades elétricas e facilidade da aplicação do método proposto.

1.1 Objetivo

A principal meta desta pesquisa é propor um sistema de detecção, monitoramento e previsão a curto prazo de tempestades elétricas baseado apenas em dados de descargas atmosféricas nuvem-solo. A metodologia empregada para identificação e monitoramento das tempestades é a clusterização espacial e temporal das descargas. Para previsão de diversos atributos das tempestades monitoradas são aplicadas técnicas de extrapolação de dados através de ajuste de funções exponenciais.

A utilização do sistema de modo operacional, especialmente em regiões que envolvem algum tipo de risco, como é o caso de linhas de transmissão de energia, permite acompanhar o desenvolvimento e vigiar o comportamento de tempestades elétricas, além da previsão de seus atributos, podendo auxiliar em tomada de decisões de operação do sistema.

2. Materiais e Métodos

Para o desenvolvimento da metodologia proposta, três etapas consecutivas são implementadas: a primeira é a coleta de dados confiáveis e precisos, a segunda é a detecção e monitoramento das tempestades elétricas e a terceira é a projeção uma hora à frente das tempestades monitoradas na segunda etapa. A seguir, a descrição de cada uma destas fases, bem como suas formas de aplicação.

2.1 Região de Estudo e Dados

Os dados utilizados na pesquisa são provenientes da RINDAT, a Rede Integrada Nacional de Detecção de Descargas Atmosféricas (website: www.rindat.com.br) operada a partir de uma cooperação entre as empresas Furnas, CEMIG, INPE e SIMEPAR, identificando descargas nuvem-solo através de sensores especializados que enviam sinais para centrais de processamento que estimam a localização e características das descargas, e então são disponibilizadas para visualização em tempo real ou armazenadas para análises históricas. Os sensores da RINDAT estão instalados nas regiões Sul, Sudeste e Centro-Oeste do Brasil, cobrindo cerca de um terço do território nacional.

Descargas atmosféricas despertam grande interesse de estudo nas mais diversas áreas, e em especial no setor elétrico, onde diversos estudos já foram realizados, podendo-se citar os trabalhos de Diendorfer e Schulz (2003), Tun (2008) e Nucci (2009). No Brasil enormes prejuízos são contabilizados todos os anos devido a predominância aérea das linhas de energia. Kleina *et al.* (2014) aplicam técnicas de clusterização na região que engloba a linha de transmissão de energia LT 765 kV para revelar que tempestades elétricas apresentam comportamento específico de descargas em relação a variável pico de corrente do que apenas considerar o comportamento climatológico.

A LT 765 kV inicia-se em na subestação de Foz do Iguaçu (PR), passa pelas subestações de Ivaiporã (PR) e Itaberá (SP) e termina na subestação de Tijuco Preto (SP), conforme a Figura 1, levando a energia produzida na hidrelétrica de Itaipu para a proximidade do centro de consumo da região Sudeste do Brasil, sendo o sistema de transmissão de tensão mais elevada existente no Brasil (ITAIPU, 2010). Esta área da LT 765 kV será a região piloto para aplicação da metodologia proposta nesta pesquisa.

Para processamento e análise de dados, fez-se uso do *software* estatístico R (R Core Team, 2012) que possui códigos abertos, reproduzíveis e adaptáveis, de fácil manuseio para os usuários.

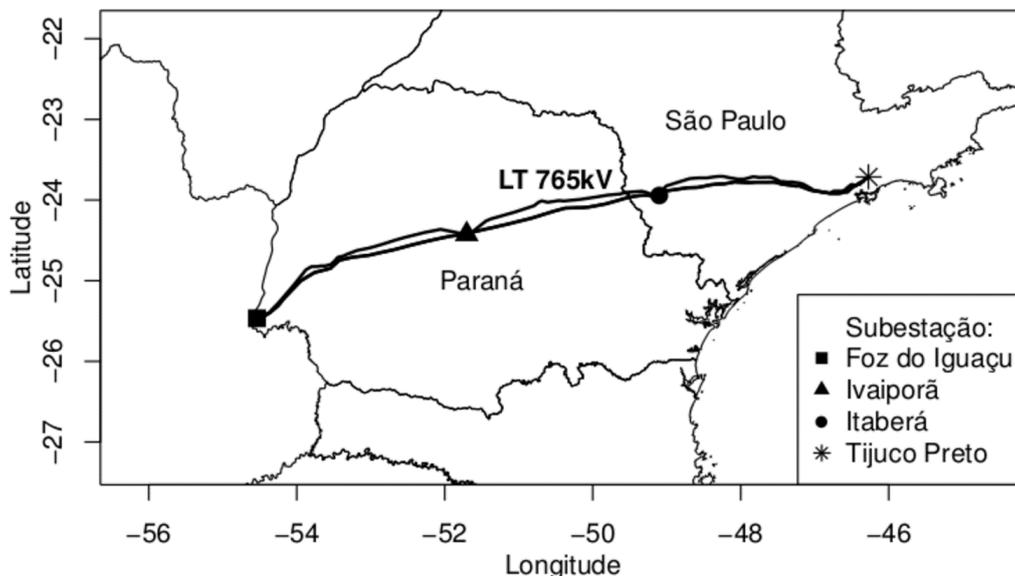


Figura 1: Região da linha de transmissão de energia LT 765 kV, que se inicia em Foz do Iguaçu (PR) e termina em Tijuco Preto (SP).

2.2 Identificação e Monitoramento de Tempestades Elétricas

O método *Data Sharpening* foi originalmente desenvolvido por Choi e Hall (1999) com a finalidade de reduzir o viés na estimativa de densidades. Uma estimativa para a densidade desconhecida $f(x)$ dos dados originais x_1, x_2, \dots, x_n 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). O parâmetro h desempenha o papel de suavizar a função a ser estimada, por isso também é chamado de parâmetro de suavização; valores pequenos de h geram uma aproximação bastante fiel aos dados, enquanto que valores grandes de h originam uma aproximação bastante suavizada para a função de probabilidade.

Em seu estudo, Choi e Hall (1999) obtém os dados ajustados $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$, onde

$$\hat{x}_j = \frac{\sum_{i=1}^n k_h(x_i - x_j) x_i}{\sum_{i=1}^n k_h(x_i - x_j)} \quad (2)$$

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

$$\hat{f}_s(x) = \frac{1}{n} \sum_{i=1}^n k_h(\hat{x}_i - x) \quad (3)$$

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

A Equação 1 tende a subestimar densidades nos picos e superestimar nos vales, e este ajuste faz com que cada dado original se desloque para mais perto de modos locais da função densidade.

Woolford e Braun (2006) propõem um teorema que diz que para h fixo e qualquer vetor inicial de observações x_0 , o algoritmo *Data Sharpening* de Choi e Hall (1999) converge para um único vetor \hat{x} . Esse teorema inspira o método *Convergent Data Sharpening*, que em síntese consiste em repetir o algoritmo de perturbação dos dados a fim de apontar esses modos locais como centro de clusters de dados com aderência a eles.

No *software* R, o método *Convergent Data Sharpening* está implementado dentro da biblioteca *CHsharp* através da função *sharp3dB* (R Core Team, 2012).

A detecção das tempestades elétricas é realizada por meio da clusterização da posição espacial de descargas atmosféricas através do método *Convergent Data Sharpening*. A aplicação emprega janela móvel de uma hora com passo de dez minutos, ou seja, 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 retirados no início e dez minutos acrescidos no final. Assim, o resultado da clusterização do período anterior pode ser utilizado na clusterização do período corrente, a fim de haver uma persistência nos clusters e não se iniciar um novo agrupamento totalmente independente do anterior. Essa estratégia permite uma maior capacidade de conexão entre clusters, promovendo o encadeamento temporal das tempestades e assim um acompanhamento mais realista das variáveis analisadas. O método *Convergent Data Sharpening* é capaz de promover essa persistência dos atributos das tempestades, uma vez que os dados aderem a núcleos de observações e estes núcleos tendem a permanecer correlatos se dados antigos apropriados permanecem e dados novos são introduzidos na amostra a ser clusterizada. Desta forma, os novos clusters gerados retratam o deslocamento temporal dos clusters passados. Por esse motivo este método foi escolhido para a aplicação proposta, além da robustez e da capacidade de lidar com grande quantidade de dados, especialmente no *software* R. Outra vantagem do método é o fato de não precisar informar previamente o número de clusters, fundamental nesta aplicação.

A clusterização aplicada conforme descrita origina retratos temporais, onde uma mesma descarga encontra-se em seis clusters distintos. Estes retratos temporais devem ser conectados (através de alguma medida de similaridade) uns aos outros para caracterizar o movimento das 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 50 km/10min ou equivalentemente 300 km/h. Em outras palavras, em 10 minutos é permitido um deslocamento dos centros dos clusters de janelas consecutivas de até 50 km. Esta velocidade é superior à encontrada na literatura para a região (Beneti, 2012) para não se perder as sequências das tempestades pois o processo de clusterização é bastante instável. Foram testadas variações de 10 a 100 km/10min para este parâmetro, porém valores de 10 a 40 km/10min conduzem a divisões inadequadas de uma tempestade e valores acima de 50 km/10min, apesar de não alterar significativamente as conexões criadas nesta aplicação, pode agregar tempestades distintas. Feito isso é concluído o procedimento de identificação e rastreamento das tempestades elétricas. Por convenção, 3 horas do passado das tempestades são monitoradas, visto que é um período de tempo suficiente para uma tempestade passar do estágio inicial de formação para o estágio de maturação. Assim, se uma tempestade durar mais que 3 horas, o passado mais antigo é descartado, dando lugar ao histórico mais recente da tempestade.

Os processos de fusão e cisão entre tempestades também são considerados. 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.

Contudo, para aplicação da metodologia proposta, inicialmente deve-se escolher o parâmetro h do método de clusterização. Este é o parâmetro do sistema proposto.

2.2 Previsão de Tempestades Elétricas

Com a identificação das tempestades elétricas, diversos atributos podem ser determinados ao longo do seu passado. Os atributos das tempestades calculados nesta pesquisa são:

- Posição central: corresponde a localização (longitude x latitude) do centro da tempestade elétrica, calculada através do valor médio das posições das descargas que compõem cada tempestade;
- Número de descargas: quantidade média de descargas por hora (descargas/hora) pertencente a cada tempestade;
- Pico de corrente: média do valor absoluto do pico de corrente das descargas (kA) que compõem cada tempestade;
- Distribuição espacial das descargas: representação no espaço de onde as descargas de cada tempestade se localizam (graus²).

Para retratar a disposição espacial das descargas dentro de uma tempestade é utilizada a estimativa de distribuição dos dados por uma função binormal, e devido as suas características é gerada uma elipse (chamada de elipse de incerteza) que indica uma região confiável, a um certo

nível, onde as descargas se encontram no espaço. A área desta elipse simboliza a região de abrangência espacial da tempestade. Esta elipse, e, portanto, a distribuição espacial de uma tempestade, é representada por meio de uma única matriz de covariância 2×2 , a qual é capaz de fornecer todos os atributos da elipse (eixos maior e menor e ângulo) que geram a área desta elipse. Esta matriz de covariância é calculada através da função *Mclust()* da biblioteca *mclust* do R, que é uma técnica de clusterização por modelos baseada na ideia de ajustar uma função densidade de probabilidade sobre os dados através de misturas de funções gaussianas finitas (Fraley, 2012), porém na aplicação proposta utilizou-se apenas uma função gaussiana a fim de representar adequadamente a disposição das descargas dentro de um único cluster (tempestade) já estruturado pelo *Convergent Data Sharpening*.

Uma observação importante com respeito ao pico de corrente é que, em toda a pesquisa, tanto na etapa de monitoramento quanto na etapa de previsão, optou-se por trabalhar com o logaritmo desta variável, pelo motivo de que é sabido que o pico de corrente das descargas segue a distribuição lognormal (Berger et al., 1975) e, portanto, o logaritmo desta variável é aproximadamente normal. Este fato favorece alguns cálculos e torna o problema mais simples. Entretanto, nas figuras que seguem, a própria média do pico de corrente será apresentada, pela familiarização com esta variável.

Todos os atributos são calculados a cada passo de tempo na etapa de clusterização, e o acompanhamento deles permite monitorá-los ao longo do tempo. A previsão, para um determinado período de tempo à frente, indicará o comportamento do evento meteorológico estudado que é a tempestade elétrica.

Um método amplamente utilizado na literatura para realizar previsão de variáveis meteorológicas (Dixon e Wiener (1993), Hering *et al.* (2004), Bonelli e Marcacci (2008)) é a extrapolação de dados, que consiste basicamente em aproximar um valor desconhecido fora de um intervalo de pontos conhecidos. Uma função é ajustada a um conjunto de observações (uma série histórica por exemplo) e esse padrão é estendido para o futuro. É claro que o resultado da extrapolação de dados é mais confiável quando o horizonte da previsão é menor e quando a série histórica de dados apresenta comportamento mais regular (uniforme). Para uma boa previsão, informações essenciais sobre o futuro da série de dados estão contidas na série histórica e pressupõem-se que tendências passadas se estenderão para o futuro. Se o futuro apresentar novos efeitos não observados no passado, eventualmente a extrapolação será inadequada.

Diversas funções são utilizadas para ajustar um conjunto de dados, tais como funções lineares, exponenciais e logísticas, porém a escolha da função ajuste depende essencialmente dos dados, que podem ser contínuos, suaves, periódicos, entre outros. Uma análise exploratória nos dados históricos permite uma melhor compreensão de tendências e assim um modelo mais apropriado pode ser escolhido.

A extrapolação de dados no tempo será a técnica utilizada para previsão à curto prazo de alguns atributos das tempestades elétricas. Conforme mencionado, o conhecimento prévio dos dados utilizados é a chave para uma extrapolação adequada, por isso uma análise visual foi realizada das variáveis que serão extrapoladas: posição central, número médio de descargas, pico de corrente médio das descargas e distribuição espacial das descargas das tempestades elétricas.

Para as variáveis relacionadas a posição central da tempestade, isto é, a latitude e a longitude, foram ajustados os seguintes modelos extrapoladores, respectivamente:

$$f(t) = e^{a_1 t + b_1} \quad (4)$$

$$f(t) = e^{a_2 t + b_2} \quad (5)$$

onde t representa a variável tempo, a_1 e b_1 são coeficientes obtidos de um ajuste linear que utiliza a transformação semi-log dos dados, ou seja, $\log(lat_1, \dots, lat_n) \sim (t_1, \dots, t_n)$ em que lat_1, \dots, lat_n são dados passados e já conhecidos. De forma análoga a_2 e b_2 são obtidos de $\log(lon_1, \dots, lon_n) \sim (t_1, \dots, t_n)$. Latitude e longitude são extrapoladas independentes umas das outras. Nesta pesquisa, quando se menciona o termo logaritmo (ou log na sua forma abreviada), fica subentendido que é referente ao logaritmo natural.

Para as demais variáveis, a função extrapoladora utilizada foi:

$$f(t) = e^{a \log(t)+b} \quad (6)$$

onde novamente t é o tempo, a e b são coeficientes provenientes de um ajuste linear que utiliza a transformação log-log dos dados, ou seja, $\log(y_1, \dots, y_n) \sim \log(t_1, \dots, t_n)$, y sendo as demais variáveis previstas (número de descargas, pico de corrente e elementos da matriz de covariância da elipse que representa a distribuição espacial da tempestade) de maneira independente. A escolha das funções extrapoladoras utilizadas na pesquisa foi realizada por meio de testes empíricos com análise visual dos dados, sendo escolhidas as funções expressas em (4), (5) e (6). No *software* R, o ajuste linear é feito através da função $lm()$ do pacote *stats* (R Core Team, 2012).

Vale ressaltar que os coeficientes do ajuste linear não são previamente calculados uma única vez e então aplicados para projetar as variáveis, mas sim calculados em tempo de processamento para cada tempestade (modo dinâmico), utilizando seu histórico recente e construindo um modelo no tempo, para então projetá-las uma hora à frente.

O número de observações transcorridas utilizadas na extrapolação dos dados é uma questão a se analisar: poucas observações podem gerar uma extrapolação pobre, enquanto que muitas observações podem ocasionar extrapolação inadequada devido a irrelevância da utilização de dados antigos. Testes numéricos comprovaram a boa previsibilidade das variáveis analisadas utilizando quatro observações passadas, ou seja, nos tempos -30, -20, -10 e 0, totalizando 30 minutos de histórico da tempestade.

3. Resultados e Discussões

3.1 Encontrando o Parâmetro do Sistema

Conforme apresentado, o sistema de detecção, monitoramento e previsão de tempestades elétricas é dependente do parâmetro de suavização h do método de clusterização. A escolha deste parâmetro deu-se por meio da análise da eficiência de monitoramento e previsão do sistema, baseado em otimização multicritério.

Para representar as tempestades no espaço, gerou-se uma malha regular de 10x10 km sobre a região da LT 765 kV e optou-se por representar cada tempestade por uma distribuição de probabilidade do tipo normal bivariada. Através desta representação bivariada e com as previsões dos atributos das tempestades, é possível estimar a probabilidade de ocorrência de descargas para cada ponto de grade (quadrícula) através da integração numérica desta função de probabilidade.

Com esta estimativa, calculam-se o número esperado de descargas por quadrícula, assim como demais atributos. Portanto, a comparação entre valores previstos e observados é realizada sobre cada ponto de grade.

O problema de otimização multicritério com pesos p_i previamente definidos é:

$$\min F(I_n, I_{PC}, I_p, I_d) = \sqrt[2]{p_1(I_n)^2 + p_2(I_{PC})^2 + p_3(I_p)^2 + p_4(I_d)^2}$$

em que os significados dos índices (que dependem de h) são descritos na Tabela 1.

Tabela 1: Significado dos índices utilizados na função objetivo F do problema de otimização.

Índice	Fórmula	Significado
I_n	$1 - \frac{EQM_{n_{ref}}}{EQM_n}$	Índice baseado no número esperado de descargas por quadrícula, onde EQM_n é o erro quadrático médio entre os valores previstos e observados do número de descargas por quadrícula; $EQM_{n_{ref}} = 0,081$ é o erro que o sistema comete acertando exatamente a previsão do número de descargas das tempestades.
I_{PC}	$1 - \frac{EQM_{PC_{ref}}}{EQM_{PC}}$	Índice baseado na média do logaritmo do valor absoluto do pico de corrente das descargas incidentes por quadrícula, em que EQM_{PC} é o erro quadrático médio entre os valores previstos e observados desta variável por quadrícula; $EQM_{PC_{ref}} = 0,272$ é o erro que o sistema comete acertando exatamente a previsão do pico de corrente das descargas das tempestades.
I_p	$1 - \frac{AUC_p}{AUC_{p_{ref}}}$	Índice baseado na probabilidade de ocorrência de descargas nas quadrículas, onde AUC_p é área sob a curva ROC (Jolliffe e Stephenson, 2003) que compara o resultado da previsão (probabilidade de ocorrer descargas na quadrícula) com as verdadeiras ocorrências (ocorreu (1) ou não (0) descargas na quadrícula); $AUC_{p_{ref}} = 0,992$ é a maior área sob a curva ROC assumindo que a previsão foi exata.
I_d	$\frac{duração_{ref}}{duração}$	Índice baseado na duração das tempestades, onde $duração$ é o tempo médio de existência da tempestade; $duração_{ref} = 18$ representa a menor duração de uma única tempestade que abrange todas as descargas incidentes (18 momentos de 10 minutos = 3 horas).

Os índices de referências de EQM não são nulos e $AUC_{p_{ref}}$ não é igual a 1 pelo fato de que, mesmo que a previsão para as tempestades seja exatamente a real ocorrida, se têm o erro da representação elíptica das tempestades e a comparação entre valores previstos e observados é feita com base na grade gerada, isto é, por quadrícula.

Almejam-se I_n , I_{PC} , I_p e I_d pequenos, por isso a minimização da função objetivo. Os três primeiros índices estão relacionados com a parte de previsão do sistema, já o último índice é referente ao monitoramento das tempestades. Os pesos p_1, p_2 e p_3 foram ajustados com o valor $2/9$ (totalizando peso de $2/3$) e p_4 com o valor $1/3$, significando uma importância maior na etapa de previsão do que na etapa de monitoramento do sistema.

Como o processo de otimização exige bastante esforço computacional (são realizadas clusterizações, conexões entre clusters, previsões, acúmulo do histórico das tempestades e

comparação das previsões com os valores observados, levando em média 15 minutos para completar o processo no *software* R com dados de 1 dia, com aproximadamente 8.000 descargas, utilizando $h = 0,2$), um período amostral de vinte dias foi escolhido para a otimização – cinco dias seguidos para cada estação do ano a fim de representar a variabilidade climática ao longo de um ano. O resultado da otimização é ilustrado na Figura 2, onde são retratados os índices da Tabela 1, juntamente com a função objetivo do problema, com h variando de 0,1 em 0,1. O resultado estimado da otimização ocorreu em $h = 0,2$.

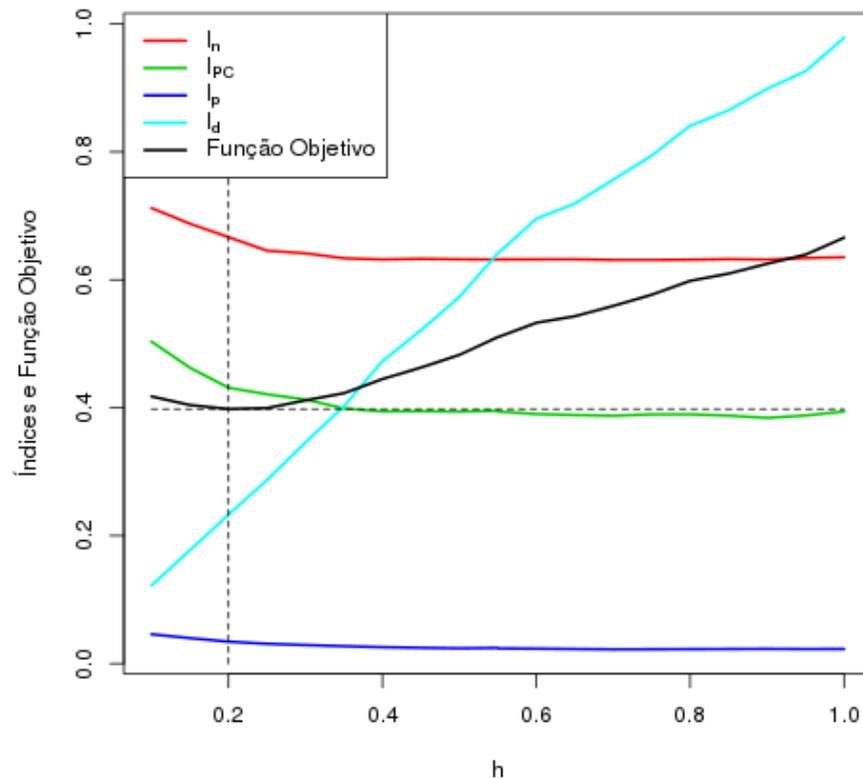


Figura 2: Valores dos índices e função objetivo do problema, cuja melhor solução foi $h = 0,2$.

3.2 Sistema Pré-Operacional

Encontrado um bom parâmetro para o sistema ($h=0,2$), é possível aplicá-lo para detectar, monitorar e prever tempestades elétricas de modo pré-operacional, podendo uma tempestade elétrica ser acompanhada, inspecionando diversas características ao longo do seu desenvolvimento.

O sistema proposto produz características tanto do passado da tempestade quanto do seu futuro (uma hora à frente), sendo possível acompanhar visualmente algumas destas informações. Como já mencionado, uma tempestade tem histórico máximo de 3 horas no sistema apresentado. Nas figuras a seguir, o passo de 10 minutos na consideração da janela móvel da clusterização foi modificado para 1 minuto com o intuito de discretizar melhor o tempo e originar traçados mais contínuos. Porém esta modificação não afeta em nada o que já foi realizado, uma vez que o passo

de 1 minuto não foi utilizado na etapa de ajuste do parâmetro devido ao aumento do esforço computacional.

Para ilustrar o resultado do sistema proposto, dois exemplos reais serão apresentados. O primeiro caso de estudo é referente a um período em que foi registrada uma falha no sistema elétrico piloto e o segundo caso de estudo é um dia usual com descargas, sem a informação de falha ou não no sistema.

Caso de estudo 1: foi selecionado o dia 29 de Outubro de 2008, em torno do horário 13:03, em que foi registrado um desligamento de energia da LT 765 KV, no trecho entre Foz do Iguaçu e Ivaiporã, com provável causa previamente definida como sendo descarga atmosférica. Nas Figuras 3 à 8, o período de 09:33 à 12:33 foi monitorado e a previsão para a hora seguinte foi realizada (isto é, previsão para às 13:33, sendo que o momento da falha fica exatamente na metade do período de previsão). A Figura 3 mostra todas as tempestades elétricas que foram acompanhadas no período mencionado. O trajeto na cor rosa representa o caminho percorrido por cada tempestade; a seta azul é o resultado da previsão da localização da tempestade uma hora adiante e a elipse azul representa onde 50% das descargas de uma hora à frente foram previstas a incidir. Exatamente às 12:33 caíram 7.037 descargas na região e 22 tempestades elétricas foram identificadas e estavam ativas no momento. Algumas tempestades apresentam um longo passado e outras encontram-se no estágio inicial de vida (círculo azul com um ponto no centro). Nota-se que uma tempestade (assinalada com asterisco verde) se aproxima do trecho mencionado onde ocorreu a falha, por isso será analisada mais detalhadamente.

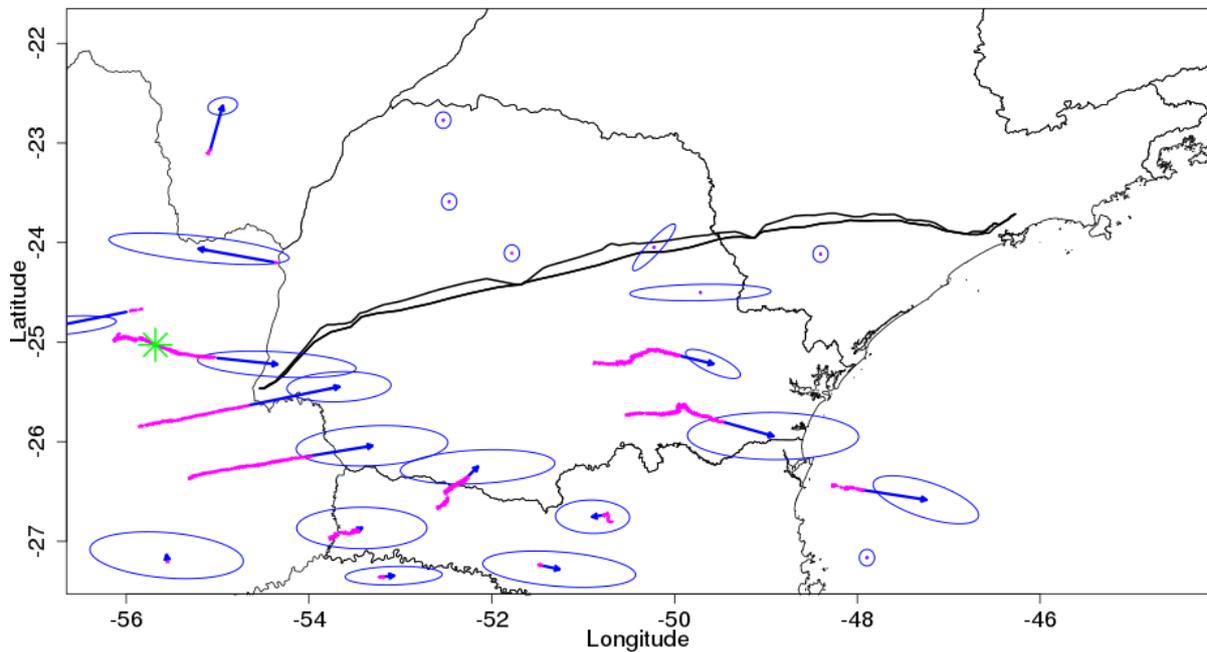


Figura 3: Tempestades elétricas identificadas no dia 29/10/2008 das 09:33 às 12:33 (trajeto rosa) e suas respectivas previsões uma hora à frente (setas e elipses azuis).

A Figura 4 ilustra a ampliação da imagem da tempestade assinalada na figura anterior com algumas informações adicionais. As elipses em gradiente de cores (mais claras são mais antigas) mostram com 50% de certeza onde as descargas que integram a tempestade incidiram minuto a minuto. É possível notar que esta tempestade em sua fase inicial de monitoramento apresentou um trajeto um pouco instável, porém em um certo momento começou a apresentar regularidade,

com trajetória do oeste para o leste. Ocorreram em média 2.833 descargas/hora às 13:33 e a elipse vermelha indica a distribuição espacial de tais descargas com 50% de confiança. Foram previstas 3.438 descargas/hora para o momento e a elipse azul ilustra a área onde estas descargas foram previstas incidir com 50% de confiança. Atenta-se que o percurso da tempestade é no sentido da linha de energia, mostrado tanto na previsão como no trajeto real.

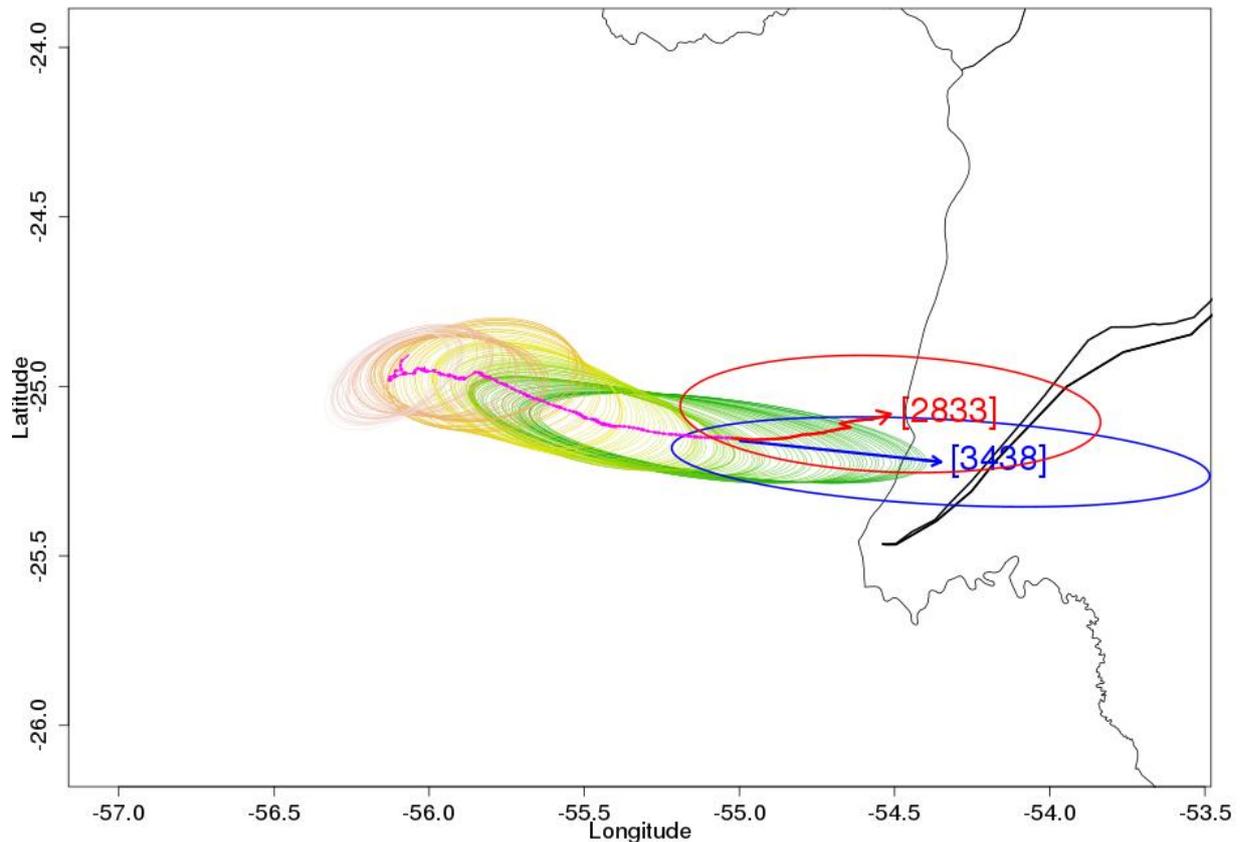


Figura 4: Trajetória de uma tempestade elétrica identificada na região (rosa), o trajeto real (vermelho) e sua respectiva previsão uma hora à frente (azul).

A Figura 5 exibe o comportamento do número médio de descargas por hora desta tempestade onde é possível notar o crescente aumento desta variável, chegando a mais de 2.000 descargas/hora às 12:33 (último minuto monitorado e indicado no tempo 0). Para as 13:33 foram previstas 3438 descargas/hora (curva azul), porém o valor observado foi menor, 2.833 descargas por hora (curva vermelha), conforme já relatado anteriormente. Nota-se que a curva rosa inicia aproximadamente no tempo -150, significando que esta tempestade não tem um passado de 180 minutos, e sim 150 minutos (ou seja, ela iniciou por volta de 10:00), porém é indiferente para a previsão que utiliza 30 minutos do passado da tempestade.

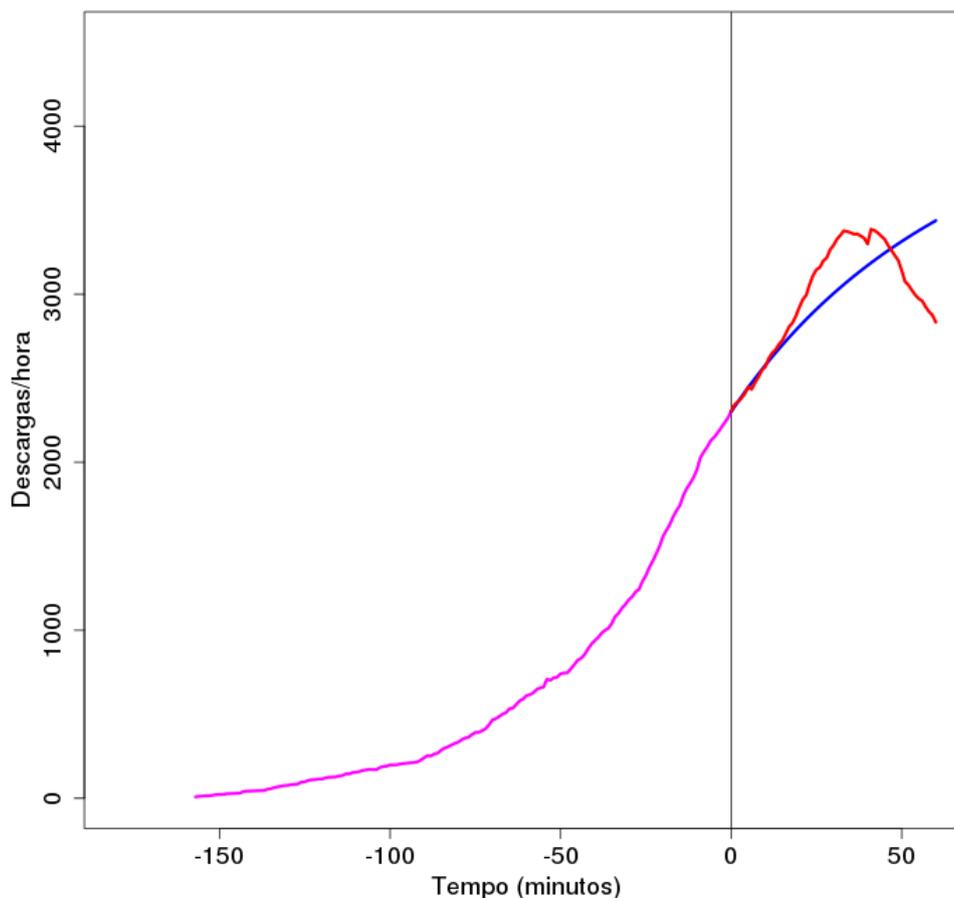


Figura 5: Comportamento do número médio de descargas por hora de uma tempestade elétrica (curva rosa), sua previsão (curva azul) e observação (curva vermelha) uma hora à frente.

A Figura 6 apresenta o comportamento da área da elipse de 95% de confiança da tempestade analisada, representando a abrangência espacial das descargas. Esta variável, assim como o número de descargas, apresenta uma evolução (curva rosa), atingindo quase 1 grau² (ou equivalentemente 111.000 km²) às 12:33. Já para às 13:33, tanto a previsão (curva azul) quanto a observação (curva vermelha) foram de aproximadamente 1,5 graus², significando que as descargas futuras são mais dispersas sobre a região.

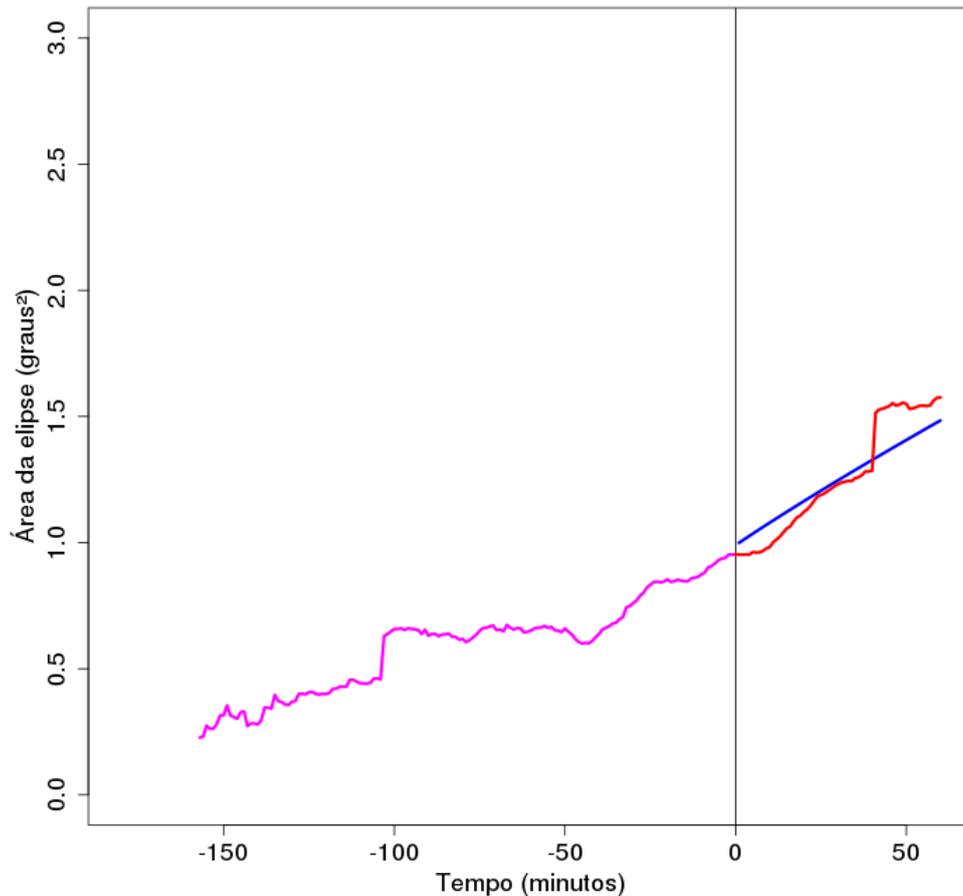


Figura 6: Comportamento da área da elipse de incerteza de 95% de uma tempestade elétrica (curva rosa), sua previsão (curva azul) e observação (curva vermelha) uma hora à frente.

A Figura 7 exibe o comportamento da média do valor absoluto do pico de corrente das descargas que integraram a tempestade ao longo do período analisado e sua respectiva extrapolação. Durante o período monitorado da vida desta tempestade (curva rosa), houve um máximo de pouco menos de 35 kA logo no início do rastreamento, decrescendo após esse momento, até atingir valor próximo a 23 kA no último minuto de monitoramento. A previsão (curva azul) e a observação (curva vermelha) uma hora à frente resultaram em valores próximos a 19 kA. Estes resultados mostram que o valor médio do pico de corrente das descargas diminuiu mais de 10 kA com a evolução da tempestade.

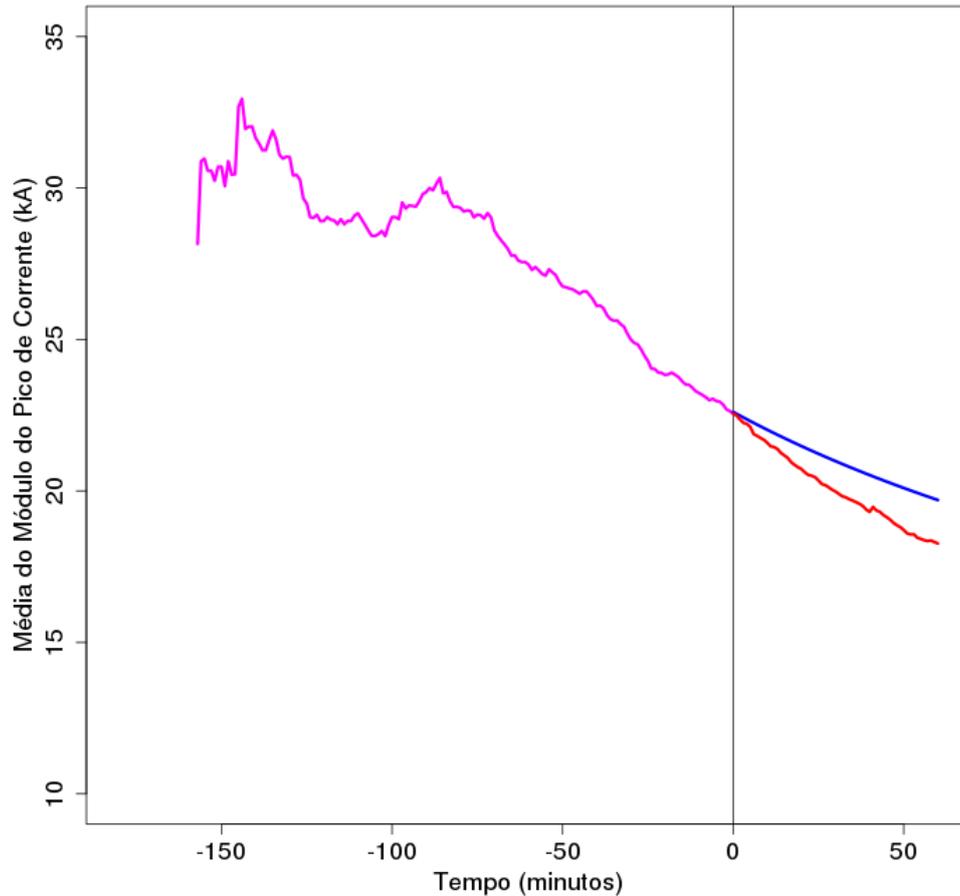


Figura 7: Comportamento da média do valor absoluto do pico de corrente das descargas que compõem a tempestade elétrica (curva rosa), sua previsão (curva azul) e observação (curva vermelha) uma hora à frente.

Finalmente, na Figura 8 são exibidas as mesmas informações apresentadas na Figura 3, isto é, o trajeto das 22 tempestades identificadas na região em um período de 3 horas e a previsão da localização uma hora à frente, porém com a informação do número esperado de descargas para às 13:33 em uma grade regular de 10x10 km. Analisando a mesma tempestade investigada anteriormente, nota-se que ela provoca um número bastante elevado de descargas (20 descargas para uma quadrícula de 100 km²) para uma área contígua ao seu desenvolvimento na próxima hora. Outras duas tempestades apresentam condições similares à analisada em relação ao número esperado de descargas por quadrícula. As demais tempestades apresentaram previsões de aproximadamente 1 a 15 descargas a cada 100 km².

Por meio das Figuras 3 a 8, pode-se concluir que havia uma tempestade eletricamente ativa se desenvolvendo nas proximidades do trecho Foz do Iguaçu – Ivaiporã da LT 765 kV, e que às 13:03 (momento exato do desligamento) era esperada uma forte atividade elétrica sobre a linha. Se operado em tempo real, este sistema de identificação, monitoramento e previsão de tempestades elétricas pode representar uma poderosa ferramenta de apoio em tomadas de decisão para que falhas como esta possam ser evitadas, quando possível.

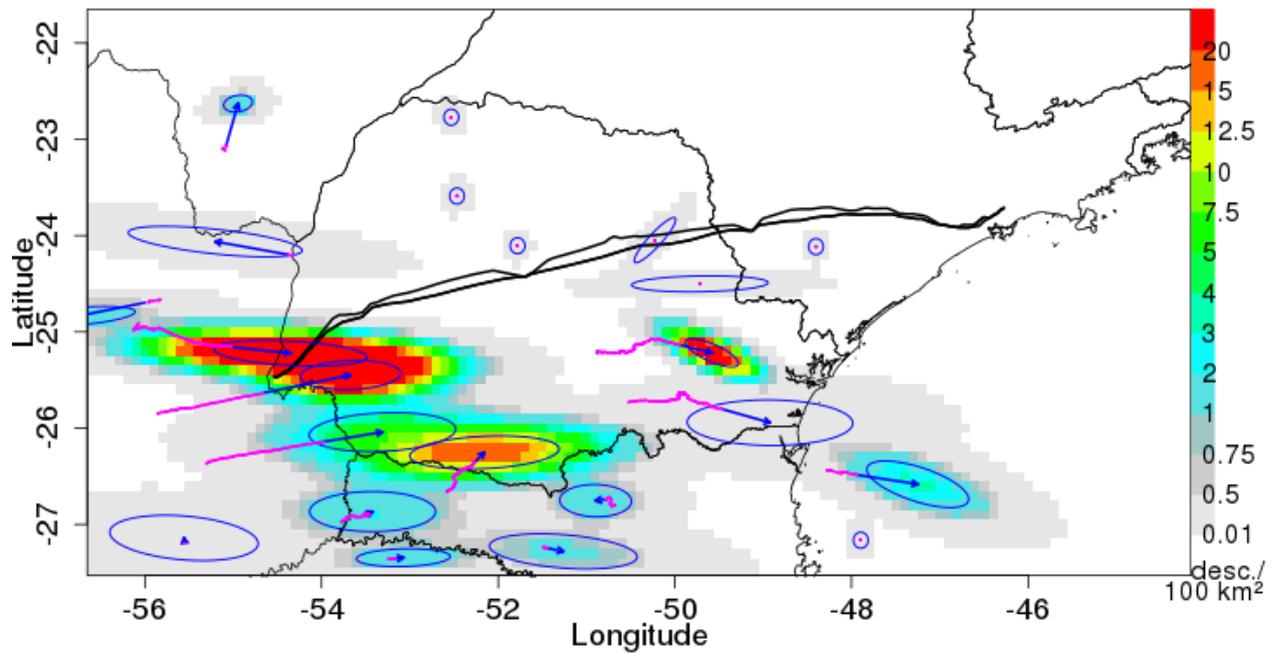


Figura 8: Número esperado de descargas para a próxima hora em 10x10 km para a região piloto.

Caso de estudo 2: o segundo caso estudado é referente ao dia 10 de Julho de 2015, onde tempestades elétricas foram identificadas e rastreadas das 11:00 até 14:00 horas e projetadas para as 15:00. Não há informação se houve ou não desligamento de energia da linha neste período. A Figura 9 ilustra 11.484 descargas organizadas em 29 tempestades elétricas ativas às 14:00, com suas previsões para às 15:00.

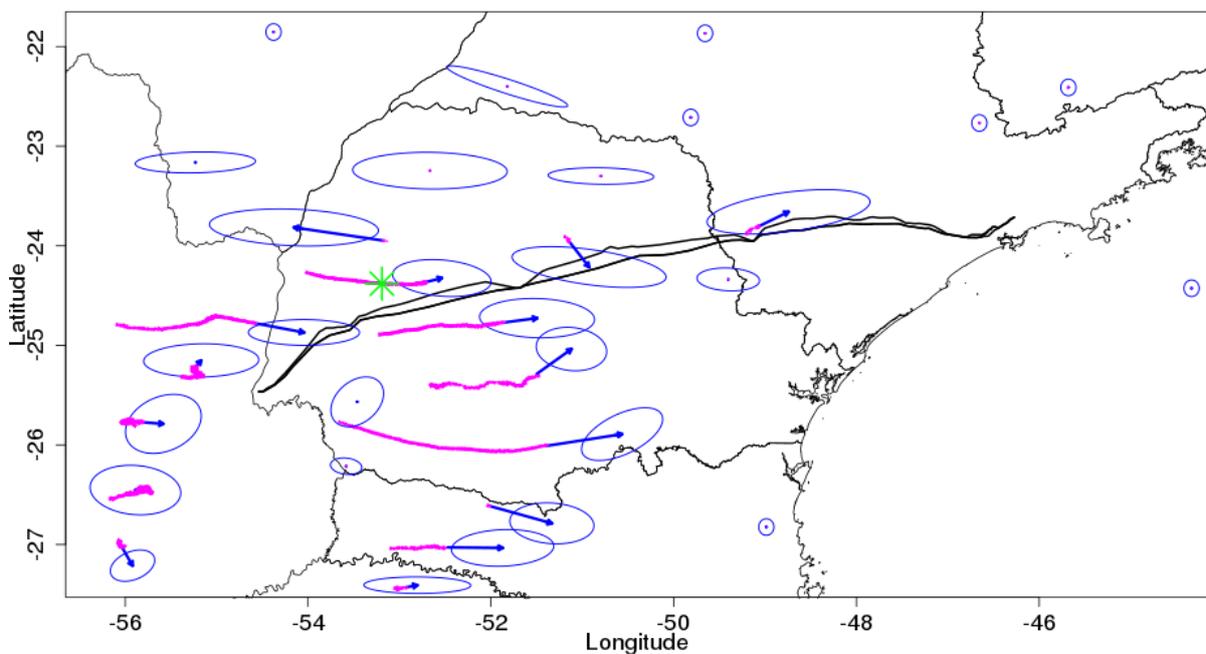


Figura 9: Tempestades elétricas identificadas no dia 10/07/2015 das 11:00 às 14:00 (trajeto rosa) e suas respectivas previsões uma hora à frente (setas e elipses azuis).

A Figura 10 ilustra a ampliação da tempestade marcada na Figura 9, uma tempestade que têm previsão de percurso nos arredores da LT 765 kV. A trajetória da tempestade é bastante regular (trajeto rosa), e as elipses em cores degradê revelam que o espalhamento das descargas foi maior na fase inicial desta tempestade. Trajetórias prevista e observada foram próximas, e representação espacial das descargas (elipses) se mostraram bastante similares, entretanto a previsão do número médio de descargas por hora (2.442 descargas/hora) superestimou o real ocorrido (1.032 descargas/hora).

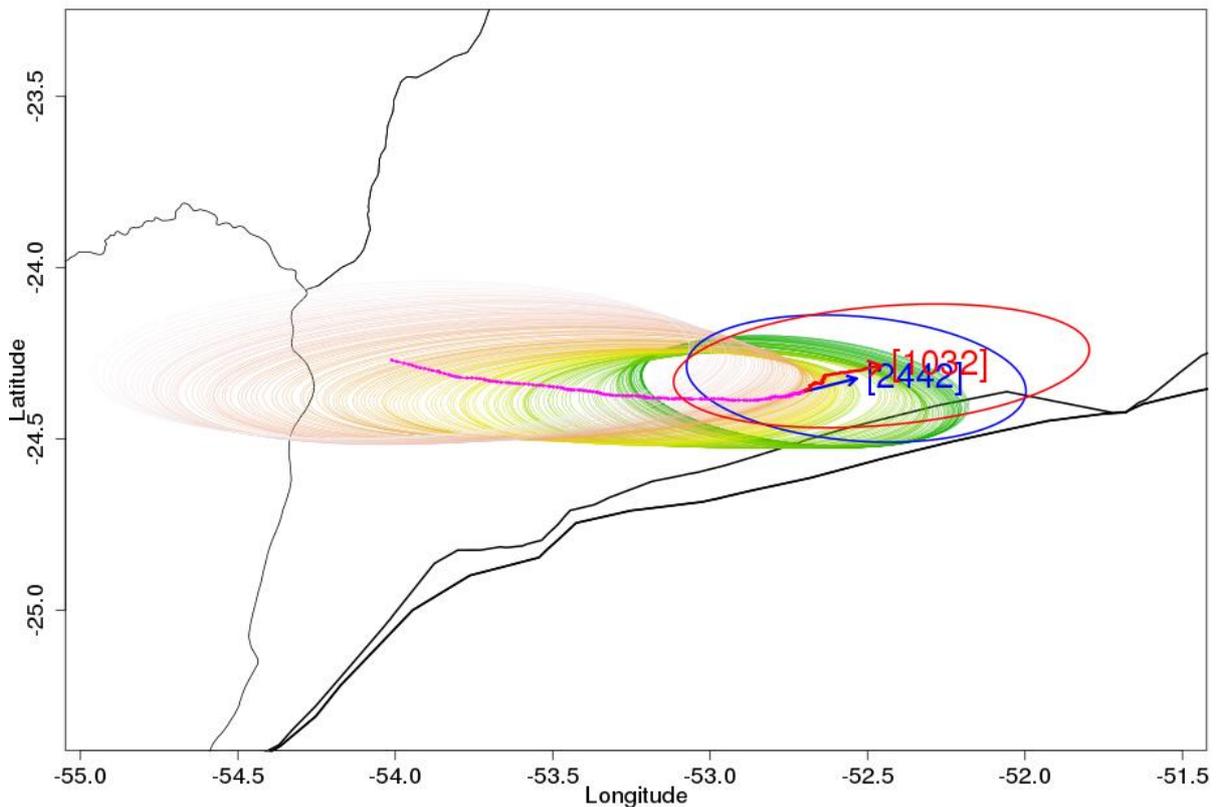


Figura 10: Trajetória de uma tempestade elétrica identificada na região (rosa), o trajeto real (vermelho) e sua respectiva previsão uma hora à frente (azul).

Nas Figuras 11, 12 e 13 são ilustrados os resultados do monitoramento das 11:00 às 14:00 (curva rosa), previsão (curva azul) e observação (curva vermelha) para às 15:00 das variáveis número médio de descargas, área de abrangência da tempestade e valor médio do pico de corrente das descargas, respectivamente. A Figura 11 exibe a típica evolução do número de descargas de uma tempestade: apresenta comportamento crescente no início e após maturação tende a cair; a Figura 12 indica que as descargas são mais dispersas pela região no seu surgimento e tendem a se concentrar quando a tempestade é bem comportada e está no seu estágio maduro; e a Figura 13 apresenta o comportamento da média do pico de corrente, demonstrando que esta tempestade não sofreu grandes variações desta variável ao longo do seu rastreamento. Comparando as Figuras 7 e 13, é possível notar que a tempestade analisada no caso 1 foi bem mais intensa com relação ao pico de corrente do que esta tempestade em questão, podendo representar maior potencial de impacto e dano no sistema elétrico.

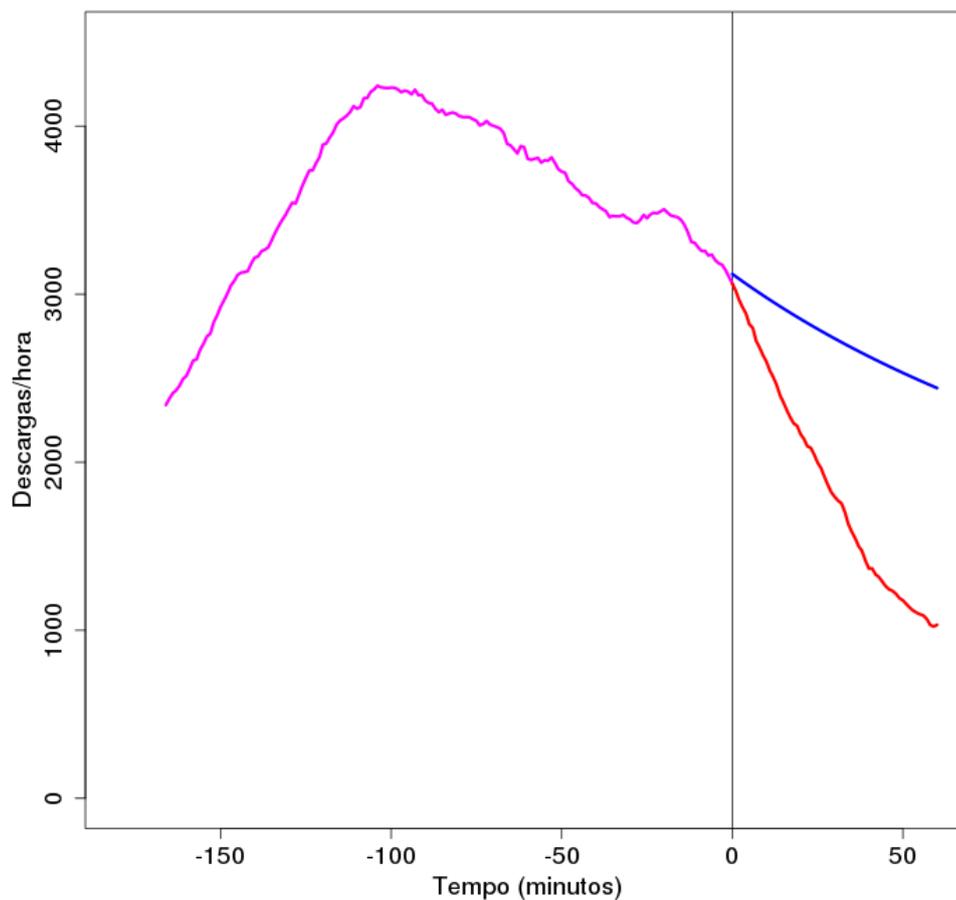


Figura 11: Comportamento do número médio de descargas por hora de uma tempestade elétrica (curva rosa), sua previsão (curva azul) e observação (curva vermelha) uma hora à frente.

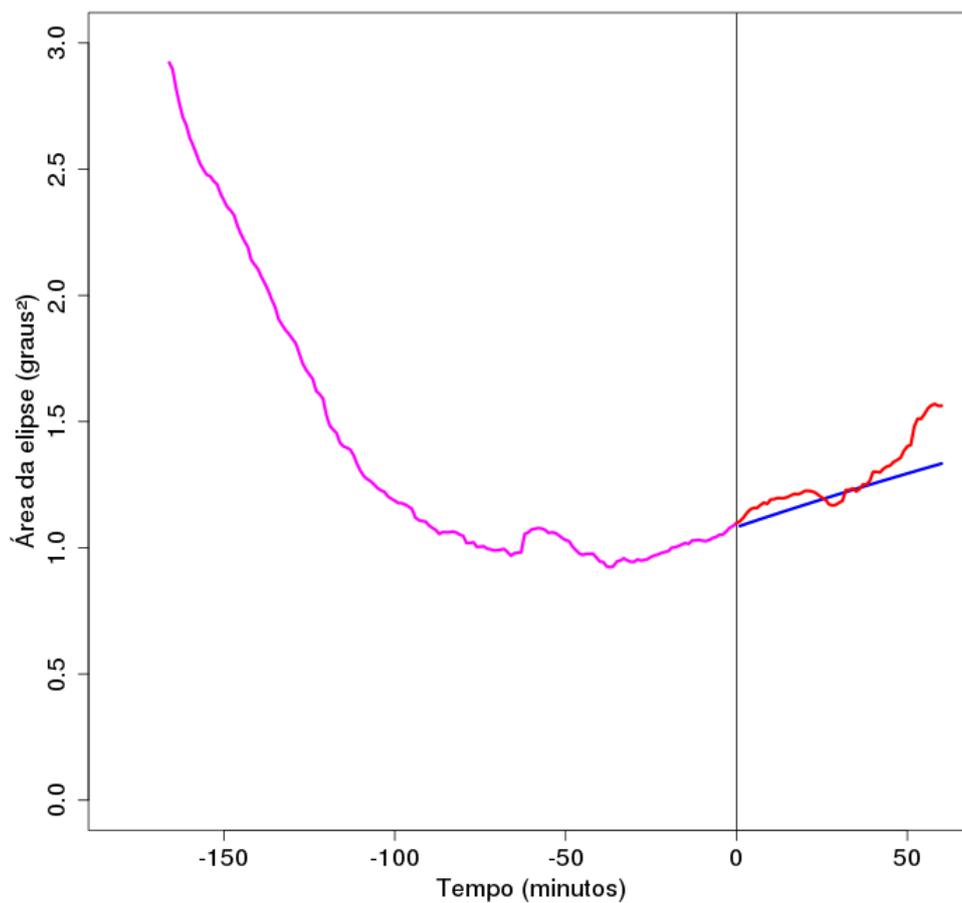


Figura 12: Comportamento da área da elipse de incerteza de 95% de uma tempestade elétrica (curva rosa), sua previsão (curva azul) e observação (curva vermelha) uma hora à frente.

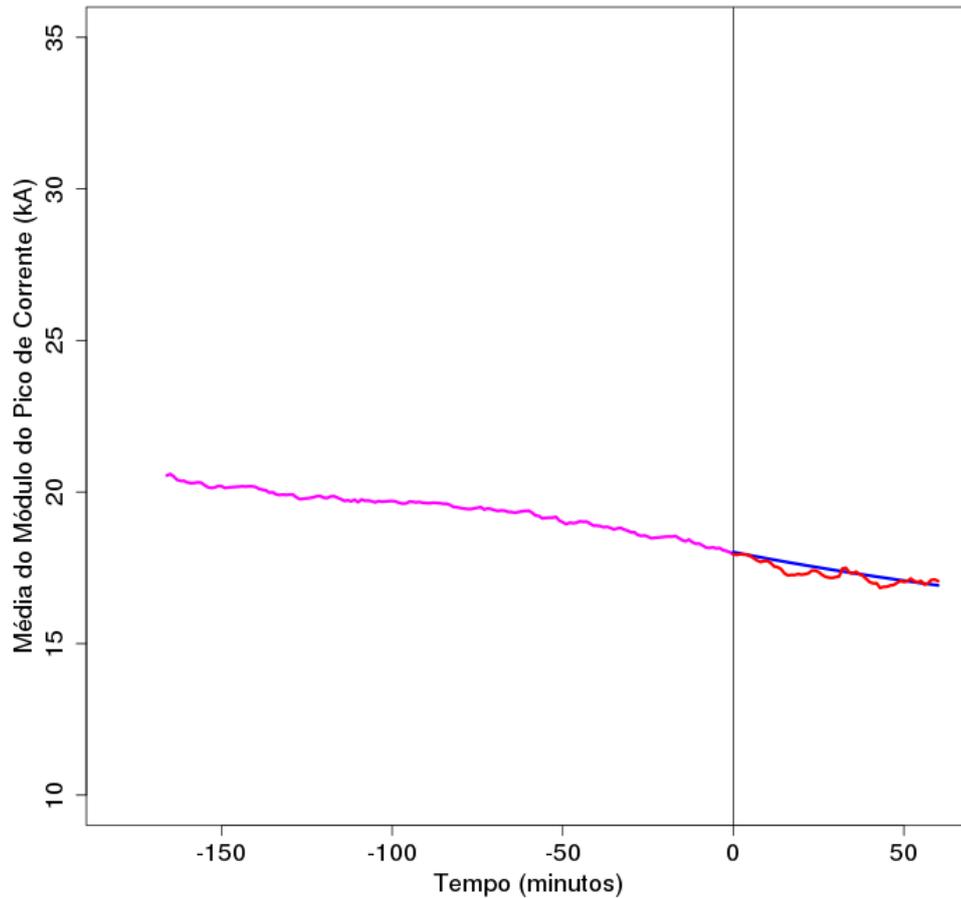


Figura 13: Comportamento da média do valor absoluto do pico de corrente das descargas que compõem a tempestade elétrica (curva rosa), sua previsão (curva azul) e observação (curva vermelha) uma hora à frente.

Por fim, a Figura 14 apresenta o número esperado de descargas por quadrícula 10x10 km uma hora adiante para as 29 tempestades identificadas. Percebe-se que três tempestades (uma delas foi analisada anteriormente) percorrem organizadamente no sentido sudoeste para nordeste da região piloto, cruzando a LT 765 KV, originando previsões de mais de 20 descargas por quadrícula para áreas adjacentes às tempestades. Não se sabe se houve desligamento na linha neste período analisado.

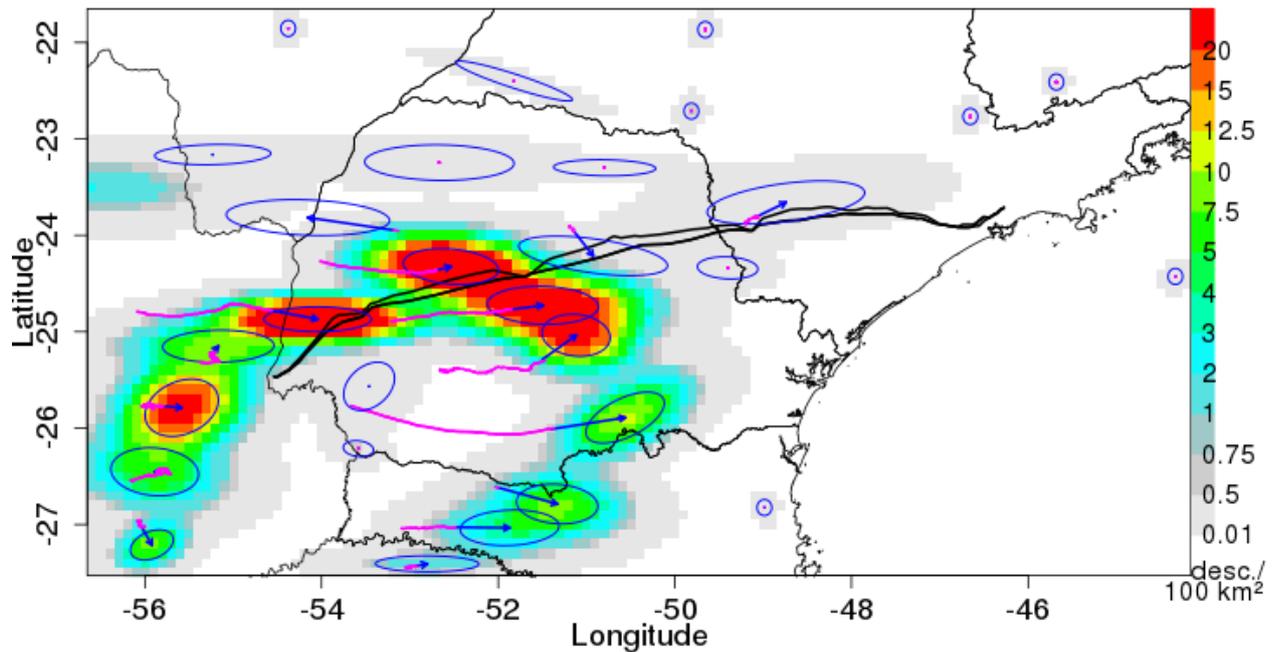


Figura 14: Número esperado de descargas para a próxima hora em 10x10 km para a região piloto.

4. Conclusões

Nesta pesquisa foi apresentado um novo sistema de detecção, monitoramento e previsão de tempestades elétricas cuja base de dados são algumas informações de descargas atmosféricas. Na etapa de identificação e monitoramento, fez-se uso de uma técnica de clusterização chamada *Convergent Data Sharpening*, já na etapa de previsão de variáveis das tempestades elétricas, utilizou-se extrapolação de dados. Para validar e calibrar o sistema proposto, um problema de otimização foi desenvolvido a fim de encontrar um valor para o parâmetro do sistema que aliasse bom monitoramento e previsibilidade. Após ajustado, o novo sistema pôde ser empregado na região piloto, visando o acompanhamento e vigilância de eventos com intensa atividade elétrica, particularmente nas proximidades da LT 765 kV, a linha de transmissão de tensão mais elevada do Brasil.

A aplicação do sistema proposto em situações reais permitiu a visualização dos resultados de períodos com intensa atividade elétrica nas proximidades da linha de transmissão presente na região piloto. Foi possível acompanhar visualmente a trajetória de todas as tempestades elétricas ativas por um período de 3 horas (tempo máximo de memória do sistema), bem como outros atributos, tais como o número médio de descargas e média do valor absoluto do pico de corrente de descargas por tempestade. Além do monitoramento das características do passado das tempestades elétricas, também foi possível acompanhar a previsão destas mesmas características uma hora à frente, possibilitando uma percepção da severidade e estágio de vida das tempestades.

De um modo geral, a técnica de clusterização se mostrou uma ferramenta com alto potencial para solucionar o problema de identificar tempestades elétricas, bem como a técnica de extrapolação

de dados para prever atributos das mesmas. Em relação ao sistema proposto, os resultados tanto no monitoramento quanto na previsão se mostraram adequados e coerentes com os eventos meteorológicos atuantes na região de estudo.

Como continuação da pesquisa, pretende-se ampliar a região de aplicação do sistema proposto, abrangendo mais áreas de interesse onde descargas atmosféricas causam prejuízos. Obviamente novos parâmetros deverão ser calculados e ajustados em conformidade com a região piloto.

Com a utilização de técnicas de computação paralela e adaptações na função de clusterização (levando em conta a frequência dos dados), será possível a aplicação do sistema proposto, a cada minuto, de modo operacional e assim acompanhar em tempo real o desenvolvimento e previsão das tempestades elétricas.

REFERÊNCIAS BIBLIOGRÁFICAS

Beneti, Cesar A. A. “Caracterização Hidrodinâmica e Elétrica de Sistemas Convectivos de Mesoescala.” *PhD diss.*, Universidade de São Paulo, 2012.

Berger, Karl; Anderson, R. B.; Kröninger, H. “Parameters of lightning flashes.” *Electra* 41 (1975): 23–37.

Betz, Hans Dieter; Schmidt, Kersten; Oettinger, Wolf P.; Montag, Brigitte. “Cell-tracking with lightning data from LINET.” *Advances in Geosciences* 17 (2008): 55–61.

Bonelli, Paolo; Marcacci, Pietro. “Thunderstorm nowcasting by means of lightning and radar data: algorithms and applications in northern Italy.” *Natural Hazards and Earth System Sciences* 8 (2008) 1187–1198.

Choi, Edwin; Hall, Peter. “Data sharpening as a prelude to density estimation.” *Biometrika* 86 (1999): 941–947.

Diendorfer, Gerhard; Schulz, Wolfgang. “Ground flash density and lightning exposure of power transmission lines.” In: *Power Tech Conference Proceedings*, IEEE, Bologna, 2003.

Dixon, Michael; Wiener, Gerry. “TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting – A Radar-based Methodology.” *Journal of Atmospheric and Oceanic Technology* 10 (1993): 785–797.

Fraley, Chris; Raftery, Adrian E.; Murphy, Thomas B.; Scrucca, Luca. “Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation.” *Technical Report* 597 (2012). Acesso: Jan 2015. Disponível em: <http://www.stat.washington.edu/research/reports/2012/tr597.pdf>.

Hering, Alessandro. M.; Morel, Christophe; Galli, Gianmario; Sénési, Stéphane; Ambrosetti, Paolo; Boscacci, Marco. “Nowcasting thunderstorms in the Alpine region using a radar based adaptive thresholding scheme.” *Proceedings of ERAD* (2004): 1–6.

ITAIPIU. “Integração ao Sistema Brasileiro.” 2010. Acesso em: Mai 2015. Disponível em: <https://www.itaipu.gov.br/print/117>.

Jolliffe, Ian T.; Stephenson, David B. *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*. England: Wiley, 2003.

Kleina, Mariana; Matioli, Luiz C.; Leite, Eduardo A. “Análise da Intensidade do Pico de Corrente de Descargas Elétricas Associadas à Tempestades Identificadas por Técnicas de

Clusterização.” *Proceeding Series of the Brazilian Society of Applied and Computational Mathematics* 2 (2014). Acesso: Jun 2015, doi: 10.5540/03.2014.002.01.0061.

Nucci, Carlo A. “A Survey on Cigré and IEEE Procedures for the Estimation of the Lightning Performance of Overhead Transmission and Distribution Lines.” In: *X International Symposium on Lightning Protection*, 151–165, Curitiba, 2009.

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/>.

Steinacker, Reinhold; Dorninger, Manfred; Wölfelmaier, Friedrich; Krannert, Thomas. “Automatic Tracking of Convective Cells and Cell Complexes from Lightning and Radar Data.” *Meteorology and Atmospheric Physics*, 72 (2000): 101–110.

Strauss, Cesar; Rosa, Marcelo B.; Stephany, Stephan. “Spatio-temporal clustering and density estimation of lightning data for the tracking of convective events.” *Atmospheric Research* 134 (2013): 87–99.

Tun, Zar N. “Protection of lightning caused interruptions on transmission line.” In: *International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* 2, 909–912, Krabi, 2008.

Woolford, Douglas G.; Braun, Willard J. “Convergent data sharpening for the identification and tracking of spatial temporal centers of lightning activity.” *Environmetrics* 18 (2007): 461–479.

Recebido em outubro de 2015.

Aceito em março de 2016.