

Agrupamento espaço-temporal de descargas elétricas atmosféricas para seleção de eventos de atividade convectiva severa.

João Victor Cal Garcia¹, Dr. Stephan Stephany², Dr. Augusto Brandão d'Oliveira³

¹Programa de Doutorado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

²Laboratório Associado de Computação e Matemática Aplicada – LAC
Instituto Nacional de Pesquisas Espaciais – INPE

³Centro Nacional de Monitoramento e Alertas de Desastres Naturais - CEMADEN

sawamano@gmail.com, stephan@lac.inpe.br, augusto.oliveira@cemaden.gov.br

***Abstract.** Selecting events of severe convective activity is usually performed manually from sets of meteorological data or meteorological images. This paper proposes an automatic technique based on minimum spanning tree graphs grouping occurrences of atmospheric electrical discharges. It is assumed that these clusters are associated with such events and their evolution in space-time can be correlated with rainfall.*

***Resumo.** A seleção de eventos de atividade convectiva intensa é usualmente manual a partir de dados ou imagens meteorológicas. O presente trabalho propõe uma técnica automática baseada em grafos de árvores mínimas que agrupa ocorrências de descargas elétricas atmosféricas. Assume-se que estes agrupamentos estejam associados a este tipo de eventos e que sua evolução espaço-temporal possa ser correlacionada com a precipitação atmosférica.*

Palavras-chave: *Descargas elétricas atmosféricas, Radares meteorológicos, Grafos*

1. Introdução

Estudos relacionados com o comportamento e evolução espaço-temporal de eventos de atividade convectiva severa normalmente são realizados com base em um conjunto de dados proveniente de diferentes instrumentos que descrevem características diversas dos sistemas observados, como taxa de precipitação ou atividade elétrica, e nos dão importantes indícios sobre o funcionamento e desenvolvimento de cada evento analisado.

Cada evento pode aparecer na forma de célula isolada ou em agrupamentos convectivos. Esses agrupamentos podem ser agrupamentos desorganizados, linhas de instabilidade, ou ainda como super-célula quando uma das células de um agrupamento atinge grandes proporções [Rakov and Uman 2003].

Os eventos de maior interesse, no entanto, são aqueles que podem causar maiores danos e apresentam risco a segurança humana. Esses eventos são caracterizados por intensa atividade elétrica atmosférica, alta taxa de precipitação e ventos fortes. O estudo desse tipo de evento é fundamental em diversas áreas, como defesa civil e agricultura [Zhu et al. 2002].

O repositório de dados para esse tipo de estudo muitas vezes inclui leituras de longos períodos de tempo em regiões geograficamente muito extensas. Há interesse nas observações de mais de um tipo de sensor sobre o mesmo evento, o que muitas vezes implica em se lidar com diferentes formatos de armazenamento de dados com resoluções espaço-temporal distintas [Baker 2007]. Esse trabalho muitas vezes é feito de maneira manual, o que acaba limitando o estudo a uma quantidade reduzida de eventos estudados. Especificamente para trabalhos relacionando a ocorrência de descargas elétricas atmosféricas (DEA) com a massa precipitada em eventos de atividade convectiva severa, é necessário que ambas descargas e precipitação ocorram em cada evento. Formas de se detectar esse tipo de tempestade, em especial as super-células, são normalmente associadas ao *nowcasting*, baseando-se nas leituras compostas de radares Doppler [Johnson et al. 1998] e em variações de temperatura observadas por satélite [Mosher 2002].

DEAs são mais comuns em eventos de maior intensidade, e a existência de correlação entre ocorrência de precipitação e DEAs [Morin et al. 2006] permite estabelecer a hipótese de que seja possível detectar atividade convectiva severa observando apenas os dados de DEAs, como por exemplo o uso de alterações da taxa de ocorrência de DEAs em um evento [Gatlin and Goodman 2010]. Uma forma de se estimar a distribuição espaço-temporal da massa precipitada em função das DEAs é através do uso do modelo de Tapia [Tapia et al. 1998].

Será apresentado no presente trabalho um método para buscar rapidamente de maneira automática eventos de atividade convectiva severa que permita a fácil configuração dos parâmetros para a seleção. O objetivo desse método é selecionar em conjuntos grandes de dados em busca de eventos de interesse, permitindo que um número grande de eventos possa ser localizado para estudos de maneira mais rápida.

2. Metodologia e dados

2.1. Dados utilizados

Os dados utilizados para os testes e validação são relativos ao sudeste Brasileiro no período de maio a dezembro de 2009. A região de estudo do é delimitada pelo alcance das observações dos dois radares do IPMet utilizados na estimação de precipitação. Os radares estão situados aos arredores das cidades de Bauru e Presidente Prudente, estado de São Paulo. O radar de Bauru está localizado em latitude de $22^{\circ}21'30''S$ e longitude $49^{\circ}1'42''W$, com altitude de 640 metros; o radar de Presidente Prudente está em latitude $22^{\circ}10'30''S$ e longitude $51^{\circ}22'30''W$, com altitude de 493 metros. Ambos radares operam banda S, que engloba frequências de 2,0 a 4,0 GHz, encobrendo UHF e SHF até a 3,0 GHz. Os dados de radar foram obtidos no formato CAPPI (Constant Altitude Plan Position Indicator) do IPMet tem resolução espacial de 1 km e altitude 3000m, cobrindo uma área correspondente a um círculo de raio 150 km com centro em cada radar, limite imposto pela confiabilidade das medidas para fins de estimação de precipitação, com uma fração dessa área sendo amostrada com sobreposição pelos dois radares.

Radares meteorológicos medem a energia retroespalhada pelos hidrometeoros, em dBZ. A taxa de precipitação local é estimada pela relação Z-R apresentada na Equação 1 abaixo, onde Z_r é o fator de refletividade, R é a taxa de precipitação ou pluviosidade

($\text{mm} \cdot \text{h}^{-1}$) e, finalmente, A e b são constantes empíricas, sendo no caso adotadas os valores $A = 32$, $b = 1,65$ referentes aos radares utilizados [Calheiros and Tepedino 2006].

$$Z_r = A \cdot R^b \quad (1)$$

Como R é expresso em $\text{mm} \cdot \text{h}^{-1}$ e assumindo-se que a taxa de precipitação seja constante em cada intervalo de 7,5 minutos ($1/8$ de hora) correspondente à resolução temporal do radar, pode-se aproximar R como sendo $Z/8$.

A rede de detecção de DEA do RINDAT é composta por uma rede de sensores que detectam a radiação eletromagnética emitida pelas DEAs na faixa de 10-300 kHz. O sistema utiliza as tecnologias denominadas Sistema de Localização e Rastreamento de Raios (*Lightning Positioning and Tracking System* - LPATS) e Localização da Direção Magnética (*Magnetic Direction Finder* - MDF). Os sensores cobrem a região sudeste do Brasil, possuindo alcance de cerca de 650 km e eficiência de detecção acima de 50%. Informações de tempo são providas pelo Sistema de Posicionamento Global (GPS). Um software específico filtra os dados adquiridos pelos sensores e faz o processamento referente à localização das DEAs, que tem uma precisão média de 500 metros, e gera uma lista de DEAs no formato UALF (Universal Ascii Lightning Format), em que cada registro/linha corresponde a um *stroke*, distinguindo o tipo de DEA (nuvem-nuvem ou nuvem-solo), bem como sua polaridade.

2.2. O modelo de Tapia

O modelo apresentado por Tapia [Tapia et al. 1998], estima a distribuição espaço-temporal da precipitação através da ocorrência de DEA, e é bastante útil para se estimar precipitação em regiões com cobertura insuficiente de radar, porém com cobertura de uma rede de detecção de DEA. O modelo é representado pela Equação 2 abaixo:

$$R(t, x) = C \sum_{t=1}^{N_t} Z f(t, T_i) g(x, X_i) \quad (2)$$

$R(t, x)$ taxa de precipitação no instante t e na posição x (mmh^{-1}),

N_t número de DEAs até o instante $t + \frac{\Delta t}{2}$,

T_i instante de ocorrência da i -ésima DEA,

X_i local de ocorrência da i -ésima DEA,

Z O RLR (razão precipitação/DEA) para o evento convectivo (kg por DEA),

C fator de conversão de unidades,

f distribuição temporal da precipitação dada uma DEA, descrita pela Equação 3,

g distribuição espacial da precipitação dada uma DEA, descrita pela Equação 4.

Tapia supõe que a precipitação é uniformemente distribuída temporalmente em um intervalo de tempo Δt centrado no instante de ocorrência da DEA, conforme a Equação 3. Seu trabalho também apresenta um valor $\Delta t = 5$ minutos baseado em experimentos.

$$f(t, T_i) = \begin{cases} 1 & \text{se } |t - T_i| < t + \frac{\Delta t}{2} \\ 0 & \text{caso contrário} \end{cases} \quad (3)$$

A variação da correlação espacial entre DEA e precipitação sugere o uso de uma distribuição uniforme para a massa precipitada. No modelo proposto por Tapia, a precipitação é distribuída em um círculo de Δx de diâmetro centrado no local de ocorrência da DEA, conforme a Equação 4. Tapia utiliza $\Delta x = 10$ km.

$$g(x, X_i) = \begin{cases} 1 & \text{se } |x - X_i| < \frac{\Delta x}{2} \\ 0 & \text{caso contrário} \end{cases} \quad (4)$$

Assim, o modelo de Tapia pode ser utilizado para estimar a taxa de precipitação a partir de dados de uma rede de detecção de DEA, desde que o RLR seja conhecido e também as distribuições espaciais e temporais da precipitação em relação a cada DEA [Tapia et al. 1998]. Existem trabalhos relativos ao cálculo do RLR, mas valores encontrados variam geografica e sazonalmente [Soula and Chauzy 2001, Sist et al. 2010].

2.3. Esquema de agrupamento de DEAs

A partir da hipótese de que o modelo de Tapia gera uma aproximação razoável da distribuição espaço-temporal da ocorrência de precipitação em um evento de atividade convectiva severa, cria-se então um grafo representando esse mesmo modelo. Seja $G = (V, E)$ o grafo da ocorrência de DEA, onde V é o conjunto da ocorrência de DEA e E é o conjunto de arestas que indicam relacionamento entre as DEAs. V é gerado considerando a ocorrência de cada DEA como um nó $V(i) = v_i = (t_i, x_i)$ do grafo, onde t_i é o instante de ocorrência da i -ésima DEA e x_i é o local de sua ocorrência.

A regra para a criação de uma aresta $E(v_a, v_b) = e_{a,b}$ entre os vértices v_a e v_b é dada pela Equação 5:

$$e_{a,b} = \begin{cases} \text{Existe} & \text{se } (|t_a - t_b| \leq \Delta t_G) \text{ e } (|x_a - x_b| \leq \Delta x_G) \\ \text{Não existe} & \text{caso contrário} \end{cases} \quad (5)$$

onde Δx_G é uma distância e Δt_G é um intervalo de tempo, ambos definidos antes da aplicação do algoritmo, mas não necessariamente relacionados com os parâmetros de nome similar de Tapia apesar de representarem o mesmo tipo de informação. As DEAs associadas com uma DEA i ocorreram em instantes de tempo posteriores à sua ocorrência. Em termos mais práticos, de uma DEA i ocorreu, se forem detectadas outras DEAs durante uma janela temporal de Δt_G anterior e posteriormente ao instante de ocorrência de i e a uma distância máxima de Δx_G do local onde ocorreu i , essas DEAs possuem uma aresta conectando-as a i .

Normalmente mais de um grafo será formado em cada evento, dadas as dimensões geográficas apresentadas pelos sensores. Cada um desses grafos é um núcleo de atividade elétrica, mas deve-se levar em conta ele ainda não foi relacionado com precipitação. Cada um desses grafos possui enorme redundância de arestas, já que as DEA normalmente possuirão mais do que uma outra DEA ocorrendo em sua janela espaço-temporal. Dependendo dos valores de Δt_G e Δx_G , é possível que em um evento de grande intensidade possua uma quantidade de arestas muito grande, o que pode dificultar o processamento e o armazenamento dessas estruturas. Pode-se re-estruturar o grafo obtido como uma árvore mínima, removendo todas as arestas redundantes [CS et al. 2010].

Uma série medidas foi adotada a fim de reduzir o tempo de processamento durante a criação das arestas dos grafos e também gerar uma árvore mínima diretamente. Cada vértice a do grafo deve buscar por novos vértices b em uma janela de tempo posterior ao instante de ocorrência de a apenas, ou seja, $(t_a - t_b) \leq \Delta t_G$, ao invés de $|t_a - t_b| \leq \Delta t_G$ conforme anteriormente estipulado. Assumindo que os DEA estão organizados de maneira cronológica nos dados, é possível reduzir a quantidade de operações necessárias pela metade com esse simples procedimento. Ele também evita arestas duplamente redundantes, ou seja, uma aresta $e_{a,b}$ e uma aresta $e_{b,a}$, exceto se ambas ocorrerem no mesmo exato instante. Outra medida adotada é a de que uma aresta $e_{a,b}$ só pode ser criada se não houver nenhum outro caminho no grafo existente que ligue-as. De maneira simplificada, se duas arestas já fazem parte de um mesmo grafo devido a outras arestas, não é necessário estabelecer uma aresta diretamente entre elas. Essa medida evita a criação de problemas de redundância e permite que duas árvores possam ser fundidas facilmente se for criada uma aresta entre elas. O grafo gerado é uma árvore mínima cuja raiz é a DEA mais antiga do agrupamento.

A Figura 1 exemplifica esse comportamento como se as DEAs fossem ocorrendo em tempo real, com cada imagem sendo separada da outra por aproximadamente Δt_G e com Δx_G sendo a distância de quadrados adjacentes: 1A mostra a ocorrência da primeira DEA em $v_0 = (x = (3, 3), t = A)$; 1B indica a ocorrência de 3 novas DEAs $v_1 = (x = (1, 3), t = B)$, $v_2 = (x = (4, 4), t = B)$, $v_3 = (x = (6, 1), t = B)$, e da criação da aresta $e_{0,2}$; 1C e 1D continuam mostrando a formação de árvores.

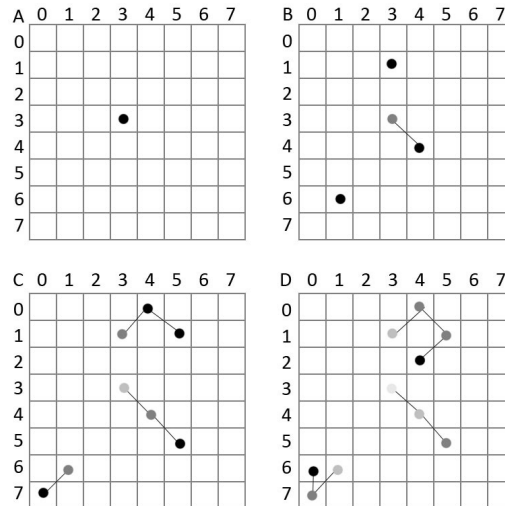


Figura 1. Esquema ilustrativo do algoritmo de grafos proposto para agrupamento de DEAs.

Por fim, a cada uma das DEAs de cada árvore é então associada a precipitação correspondente, estimada a partir dos dados de radar definidos espacial e temporalmente conforme a janela definida pelas Equações 3 e 4. Deve-se estabelecer agora características para filtrar DEA isoladas e também eventos de menor intensidade, que não caracterizam atividade convectiva severa. Se a quantidade de DEAs for muito baixa em um grafo, ou seja, se houve poucos vértices, então o grafo em questão possivelmente não representa o tipo de evento buscado. Também se a precipitação relacionada com o grafo, que pode ser estimada observando os dados de radar na janela espaço-temporal das DEA que compõem o

grafo, for baixa, o evento observado também não é intenso o suficiente para ser selecionado.

3. Resultados

Para os testes realizados, não foram considerados agrupamentos com menos de 20 DEAs nem agrupamentos com menos de 5 minutos de duração, já que esses agrupamentos normalmente descrevem eventos de baixa intensidade [Rakov and Uman 2003]. Também foram desconsiderados agrupamentos relacionados com massa precipitada média inferior a 10^6 kg por DEA. A complexidade do algoritmo foi $O(n^3)$; no entanto, como trata-se de um algoritmo de busca por árvore mínima, poderiam ser utilizados para essa parte do trabalho algoritmos de menor complexidade descritos em outros trabalhos [Pettie and Ramachandran 2002, Chazelle 2000].

Os parâmetros Δx_G e Δt_G foram inicialmente aproximados pelos valores $\Delta x_G = 10$ km e $\Delta t_G = 5$ minutos, que são o dobro dos valores Δx e Δt definidos no trabalho de Tapia [Tapia et al. 1998], e próximos aos valores utilizados por Gatlin [Gatlin and Goodman 2010]. Essa escolha foi feita pensando na situação da distância espaço-temporal limite de duas DEAs no modelo de Tapia para que possa haver interseção não-nula entre elas, e é justificada pelo fato do modelo de Tapia ser uma tentativa de reconstruir a distribuição da precipitação através de DEAs. A Tabela 1 apresenta os primeiros resultados. Com esses parâmetros iniciais, o maior agrupamento de DEAs válido possui aproximadamente 800 DEAs e durou pouco mais de 3h30min. Os eventos de interesse possuem em média 91 DEAs e duração de 9min30s.

Tabela 1. Resultados dos testes

Teste	1	2	3	4
Δt_G	5 min	10 min	5 min	10 min
Δx_G	10 km	10 km	20 km	10 km
Número máximo de DEAs por agrupamento	796	4164	7410	8793
Média de DEAs por agrupamento	91.18	121.04	144.67	176.84
Desvio Padrão	40.42	108.67	191.43	282.26
Agrupamento de maior duração	3.51 h	6.47 h	6.67 h	8.83 h
Duração média dos Agrupamentos	0.16 h	0.26 h	0.19 h	0.35 h
Desvio Padrão	0.20 h	0.35 h	0.28h	0.56 h

Explorando os efeitos de uma maior janela temporal, os testes foram feitos novamente para $\Delta t_G = 10$ minutos, mantendo $\Delta x_G = 10$ km. Os resultados, conforme apresentados no Teste 2 da Tabela 1, demonstram um aumento significativo de tanto a duração quanto a quantidade de DEAs do maior agrupamento válido. Esse aumento também é observado no tamanho e duração média dos agrupamentos de DEAs do teste.

O mesmo teste realizado agora com uma janela temporal $\Delta t_G = 5$ minutos, igual a inicial, e uma janela espacial $\Delta x_G = 20$ km, o dobro da inicial, é apresentado no Teste 3 da Tabela 1. A duração do maior agrupamento válido é bem similar a do teste com apenas a janela de tempo expandida, mas a quantidade de DEAs nele é consideravelmente maior. Também é possível notar que em média apesar de a quantidade de DEAs por agrupamento ser maior, a duração dos agrupamentos não é muito superior a do teste com janela de tempo maior.

No Teste 4 da Tabela 1, os testes foram realizados com janela espacial $\Delta x_G = 20$ km e temporal $\Delta t_G = 10$ minutos, ou seja, com janela espaço-temporal maior que no Teste 1. No maior agrupamento, a duração aproximou as 9 horas, com quase 9 mil DEAs.

Para as janelas espaço-temporais testadas, a maior parte dos DEAs ocorre em pequenos agrupamentos isolados, tipicamente relacionados com pequena massa precipitada (inferior a 10^6 kg). Muitas vezes esses pequenos agrupamentos foram encontrados próximos a agrupamentos de grande extensão, frequentemente conectados por regiões de baixa ocorrência de precipitação. Tais pequenos agrupamentos possivelmente estão relacionados com regiões de instabilidade elétrica menor, distantes do núcleo de atividade elétrica do evento observado [Finke 1999].

4. Conclusões

O presente trabalho apresentou um método para selecionar automaticamente eventos para estudos de atividade convectiva severa. O método proposto utiliza-se do agrupamento de DEAs utilizando algoritmo de busca por árvores mínimas, onde o parâmetro de criação de arestas é a existência de interseção não-nula entre janelas espaço-temporais centradas em DEAs. A janela espaço-temporal desses agrupamentos é então validada pela presença de precipitação, mas outros fenômenos meteorológicos relevantes podem ser utilizados com poucas alterações, apesar de não testados aqui.

A heurística para seleção dos dois parâmetros principais para o algoritmo de agrupamento (Δt_G , representando a janela temporal relativa a cada DEA, e Δx_G , que é relativa a janela espacial) ainda precisa ser definida e testada; possivelmente esses valores variarão conforme posição geográfica e por fatores sazonais, conforme ocorre com a relação entre DEAs e precipitação [Soula and Chauzy 2001].

O método se mostrou capaz de agrupar eventos nos quais ocorre simultaneamente precipitação e DEA. No entanto, ele retorna como resultado os núcleos de atividade elétrica; se houve vários núcleos separados ocorrendo simultaneamente sem atividade elétrica significativa entre eles, mas com precipitação, o método iria identificá-los como eventos distintos. Não foram realizados estudos para verificar se o sistema é capaz de identificar os diferentes tipos de tempestades. Esses problemas poderão ser abordados em trabalhos futuros.

Referências

- Baker, T. A. (2007). *An open-source program to animate and visualize the recorded temperature and relative humidity data from dataloggers including the building's three-dimensional geometry*. PhD thesis.
- Calheiros, R. V. and Tepedino, P. R. P. (2006). Daily Interval and Range Stratification in Rainfall Measurements with the Bauru Radar. In *Proceedings of ERAD (2006)–Fourth European Conference on Radar in Meteorology and Hydrology*, pages 18–21, Barcelona. Finnish Meteorological Institute.
- Chazelle, B. (2000). A minimum spanning tree algorithm with inverse-ackermann type complexity. *Journal of the ACM (JACM)*, 47(6):1028–1047.
- CS, V. et al. (2010). Minimum spanning tree algorithm. *International Journal of Computer Applications IJCA*, 1(8):39–45.

- Finke, U. (1999). Space-time correlations of lightning distributions. *Monthly weather review*, 127(8):1850–1861.
- Gatlin, P. and Goodman, S. (2010). A total lightning trending algorithm to identify severe thunderstorms. *Journal of atmospheric and oceanic technology*, 27(1):3–22.
- Johnson, J., MacKeen, P., Witt, A., Mitchell, E., Stumpf, G., Eilts, M., and Thomas, K. (1998). The storm cell identification and tracking algorithm: An enhanced wsr-88d algorithm. *Weather and Forecasting*, 13(2):263–276.
- Morin, E., Goodrich, D., Maddox, R., Gao, X., Gupta, H., and Sorooshian, S. (2006). Spatial patterns in thunderstorm rainfall events and their coupling with watershed hydrological response. *Advances in Water Resources*, 29(6):843–860.
- Mosher, F. (2002). Detection of deep convection around the globe. In *Preprints, 10th Conf. on Aviation, Range, and Aerospace Meteorology, Portland, OR, Amer. Meteor. Soc.*, pages 289–292.
- Pettie, S. and Ramachandran, V. (2002). An optimal minimum spanning tree algorithm. *Journal of the ACM (JACM)*, 49(1):16–34.
- Rakov, V. A. and Uman, M. A. (2003). *Lightning: physics and effects*. Cambridge Univ Pr.
- Sist, M., Zauli, F., Melfi, D., and Biron, D. (2010). A study about the correlation link between lightning data and meteorological data. In *2010 EUMETSAT Meteorological Satellite Conference*.
- Soula, S. and Chauzy, S. (2001). Some aspects of the correlation between lightning and rain activities in thunderstorms. *Atmospheric research*, 56(1-4):355–373.
- Tapia, A., Smith, J. A., and Dixon, M. (1998). Estimation of convective rainfall from lightning observations. *Journal of Applied Meteorology*, 37(11):1497–1509.
- Zhu, Y., Toth, Z., Wobus, R., Richardson, D., and Mylne, K. (2002). The economic value of ensemble-based weather forecasts. *Bulletin of the American Meteorological Society*, 83(1):73–83.