Revista Brasileira de Geografia Física

Homepage: www.ufpe.br/rbgfe

Aplicação da análise cluster na identificação de regiões homogêneas para incidência de descargas atmosféricas do tipo nuvem-solo no estado de São Paulo, Sudeste do Brasil

Ana Paula Paes dos Santos¹, Francisco José Lopes de Lima², Everaldo Barreiros de Souza³, Osmar Pinto Júnior⁴, Sérgio Rodrigo Quadro dos Santos⁵

¹ Doutoranda no Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos-SP. ana.santos@inpe.br (autor correspondente). ² Pesquisador do Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos-SP. francisco.lopes.lima@gmail.com. ³ Pesquisador Associado no Instituto Tecnológico Vale (ITV), Belém-PA. everaldo.barreiros.souza@itv.org. ⁴ Pesquisador titular do Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos-SP. osmar@dge.inpe.br. ⁵ Doutorando no Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos-SP. sergiosannts@yahoo.com.br.

Artigo recebido em 30/09/2016 e aceito em 20/10/2016

RESUMO

A climatologia de mesoescala das descargas atmosféricas pode apresentar expressivas variações em seu interior, devido à influência de determinados aspectos fisiográficos ou antrópicos que podem interferir nas trocas energéticas ou transporte de massa da circulação local, diferenciando a formação e atuação de subsistemas convectivos. Neste sentido, o objetivo deste trabalho é de identificar regiões homogêneas (RH's) de incidência de descargas atmosféricas no Estado de São Paulo, por meio da mineração de dados, com uso da técnica estatística de análise de *cluster*. Para isso, serão utilizados dados das redes de detecção de descargas atmosféricas RINDAT e BrasilDAT, compreendendo o período do verão entre os anos de 1999 a 2014. Os resultados revelaram a existência de seis RH's, no Estado, com máxima densidade sobre a RH4 (região metropolitana de São Paulo – RMSP) composta em sua maioria por cobertura de solo com áreas construídas e elevação em torno de 800 m, e mínimos sobre a RH6 (oeste do Estado), compreendendo uma região mais arborizada e plana. Apesar da RH4 ser expressivamente menor em dimensões geográficas que a RH6, as tendências revelaram magnitudes semelhantes entre ambas, o que mostra o potencial das tempestades elétricas sobre a RMSP, visto que, esta é expressivamente menor que a RH6.

Palavras-chave: regiões homogêneas, descargas atmosféricas, São Paulo.

Application of cluster analysis in the identification of homogeneous regions for the cloud-to-ground lightning incidence in the state of São Paulo, Southeastern Brazil

ABSTRACT

The climatology of mesoscale atmospheric discharges may show significant variations, due to the influence of certain physiographic and anthropic aspects that may interfere with the energy exchange or mass transport of local circulation, differentiating the training and performance of convective subsystems. Thus, the objective of this work is to identify homogeneous regions (HR's) of incidence lightning in the State of São Paulo, through data mining, using the statistical technique of cluster analysis. For this, data of lightning detection networks RINDAT and BrasilDAT will be used, comprising the summer period between 1999 to 2014. The results revealed the existence of six HR's in the State, with maximum density on HR4 (Metropolitan Region of São Paulo - MRSP) composed mostly by ground cover with built-up areas and elevation around 800 m, and minimum on RH6 (west of the State), comprising a region more wooded and flat. Although the RH4 be significantly smaller in the geographical dimensions than RH6, trends showed similar magnitudes between both, which shows the potential of electrical storms on the MRSP, since this is significantly lower than the HR6.

Keywords: Homogeneous Regions, Lightning, São Paulo.

Introdução

Na última década, estudos indicaram mudanças em parâmetros climáticos sobre diversas áreas continentais do globo, em grande parte, com tendências positivas (Del Río et al., 2010; IPCC, 2013; Santos et al., 2016a). Essas alterações são, em parte, justificadas pelo aumento de extremos

climáticos, tanto pela maior frequência de eventos extremos, originados por processos convectivos, quanto por períodos prolongados destes casos, ocasionando diversos impactos à sociedade (IPCC, 2007, 2013).

De modo geral, a atividade convectiva e a ocorrência de eventos extremos sobre determinadas regiões, estão associadas a vários fatores, influenciados desde escalas locais à grande escala. Em grande escala, pode-se citar o modo de variabilidade do Oceano Pacífico equatorial, por meio do fenômeno El Niño Oscilação-Sul (ENOS), que em sua fase quente (anomalias positivas da temperatura da superfície do mar – TSM) pode favorecer a formação de nuvens convectivas no Sudeste do Brasil. Já em escala local têm-se a influência da composição de superfície, como áreas montanhosas e/ou urbanizadas que podem favorecer a formação e desenvolvimento da convecção.

No clima regional, a configuração do terreno, o tipo de solo e sua cobertura vegetal, são considerados como aspectos da localidade, sujeito a pequenas mudanças no tempo, determinando o clima que predomina em determinado lugar, da ordem de centenas de quilômetros quadrados, e pode ser denominado por ‘clima local’ (Oliver e Fairbridge, 1987).

Santos et al. (2016a) analisaram a densidade de descargas atmosféricas sobre o Estado de São Paulo durante o verão, e observaram que há uma maior incidência de descargas sobre a região metropolitana de São Paulo e áreas ao redor, e densidades menores sobre o oeste do Estado. Entretanto, os autores não determinaram os limites geográficos das áreas homogêneas em relação a incidência de descargas atmosféricas, e nem avaliaram a dinâmica dos eventos em cada região.

Neste contexto, uma vez já identificado o período de maior incidência de descargas sobre o Estado de São Paulo (verão) por Santos et al., 2016b, e analisada a dinâmica temporal deste fenômeno (Santos et al., 2016a), torna-se necessário um estudo mais aprofundado, em escala local, para se determinar as regiões homogêneas de incidência de descargas, e examinar a dinâmica interregiões de ocorrência deste fenômeno.

Desta forma, o presente estudo visa identificar regiões homogêneas de densidade de descargas atmosféricas no Estado de São Paulo durante o verão, por meio da mineração de dados, através da aplicação da Análise de *Cluster*, além de avaliar tendências, e a ocorrência de evento extremo em cada região.

Uma das principais justificativas para que este tipo de trabalho fosse desenvolvido, é que estudos dessa natureza para este fenômeno ainda são muito incipientes. Todavia, de grande relevância para o entendimento e monitoramento do clima regional, podendo servir de embasamento científico, que possa ser usado por gestores na tomada de decisões ambientais que minimizem os

impactos causados pela incidência de descargas atmosféricas.

Material e métodos

A base de dados de descargas atmosféricas do tipo nuvem-solo (NS) utilizada neste trabalho abrange o trimestre de DJF (verão) de 1999 a 2014, provenientes da Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT) e da Rede Brasileira de Detecção de Descargas Atmosféricas (BrasilDAT) para o Estado de São Paulo, no Sudeste do Brasil, dos quais, os dados de 1999 a 2010 são da rede RINDAT e os dados de 2011 a 2014 são da rede BrasilDAT.

A série temporal foi analisada visando identificar regiões homogêneas de incidência de descargas NS, por meio da mineração de dados, a qual se baseia na utilização de diversas ferramentas computacionais em busca de extrair padrões implícitos em grandes quantidades de dados. Estas ferramentas empregam técnicas como algoritmos genéricos, análise de *cluster*, redes neurais, entre outras.

Neste trabalho, será utilizada a técnica estatística de Análise Multivariada da Análise de *Cluster*.

A Análise de *Cluster* (AC) vem sendo amplamente aplicada a pesquisas na área de geociências desde a década de 1960. Esta técnica tem como objetivo principal, classificar elementos de uma população que são conhecidos por seus aspectos, em grupos que sejam homogêneos intragrupos e heterogêneos intergrupos. A técnica caracteriza-se por ser uma ferramenta útil no processo de filtragem de dados, objetivando determinar grupos homogêneos e padrões que estejam ocultos nos dados (Halkidi, et al., 2001).

A AC parte de uma matriz de dados, onde as linhas correspondem a “*n*” unidades, fornecendo “*p*” características (valores numéricos). Neste trabalho, as ‘unidades’ foram definidas por pontos de grades compostas por 5x5 km, e as ‘características’ são densidades de descargas atmosféricas em cada ponto de grade. Cada ponto de grade tem um ID para identificação, e por síntetização textual, a partir daqui, os ID’s serão denominados por “estações”.

Na AC, todos os processos de hierarquização são similares, iniciando pela determinação de uma função de agrupamento. Essa função é utilizada como critério para medir a distância entre dois objetos ou para estabelecer o quanto eles são parecidos, denominada como coeficiente de parença. Este coeficiente pode ser dividido em duas classes: medida de similaridade e dissimilaridade. Na primeira classe, quanto maior

os valores observados, mais parecidos são os objetos. Já na segunda classe, quanto maior os valores observados, menos parecidos serão os objetos.

Para a determinação dos grupos homogêneos, pode ser utilizada tanto a medida de dissimilaridade, composta pelo cálculo da distância euclidiana, quanto à medida de similaridade, composta pelo coeficiente de correlação. Para o agrupamento, existem dois tipos de métodos, os hierárquicos e os não hierárquicos. No primeiro, são possíveis várias técnicas ou algoritmos de classificação de grupos.

Neste trabalho, aplicou-se o método hierárquico aglomerativo de Ward (1963), tendo como medida de dissimilaridade a distância euclidiana (Everitt e Dunn, 1991), definida por:

$$d_e = |x_i - x_j| = \left[\sum_{k=1}^P (x_{i,k} - x_{j,k})^2 \right]^{0,5} \quad (\text{Eq.1})$$

Onde d_e é a distância euclidiana; e $x_{i,k}$ e $x_{j,k}$ são as variáveis k dos pontos i e j , respectivamente.

O método de Ward forma grupos, minimizando a dissimilaridade, ou minimizando o total das somas dos quadrados dos desvios. Em cada fase do procedimento, são formados grupos, de forma que, o resultado tenha a menor soma de quadrados dos desvios dentro dos grupos. Nessas fases, são consideradas as junções de todos os possíveis pares de grupos e, os dois que resultam em menor aumento da soma de quadrados dos desvios são agrupados, até que todos os grupos formem um único, reunindo todos os indivíduos (Everitt e Dunn, 1991). A soma dos quadrados dos desvios é determinada da seguinte forma:

$$SQD = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (\text{Eq.2})$$

Após a definição dos grupos homogêneos de incidência de relâmpagos NS, foi avaliada a série temporal de cada grupo homogêneo, visando avaliar a dinâmica temporal intergrupos e a tendência de cada região.

Para análise da tendência, foi utilizado o teste estatístico não paramétrico de Mann-Kendall, ao qual foi utilizado devido o mesmo não necessitar que a série de dados apresente uma distribuição normal (gaussiana) a qual, é uma característica dos dados de relâmpagos, e devido também ao seu resultado ser menos influenciado por valores

outliers, visto que seu cálculo é baseado no sinal das diferenças, e não diretamente nos valores da variável.

Criado originalmente por Mann (1945) e reformulado por Kendall (1948), o teste consiste em comparar cada valor da série de dados com os valores seguintes, calculando o número de vezes em que os termos restantes são maiores que o valor analisado. O teste não paramétrico é usado para analisar a existência de uma tendência monotônica crescente ou decrescente e é o método mais apropriado para analisar alterações climáticas em séries de dados climatológicos (Goossens e Berger, 1986). O teste estatístico de Mann-Kendall é determinado da seguinte forma:

$$Z = \begin{cases} \frac{S - 1}{\sqrt{\text{VAR}(S)}} & \text{se } S > 0 \\ 0 & \text{se } S = 0 \\ \frac{S + 1}{\sqrt{\text{VAR}(S)}} & \text{se } S < 0 \end{cases} \quad (\text{Eq.3})$$

onde,

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_k) \quad (\text{Eq.4})$$

Em que x_j e x_k são valores (anuais/sazonais/mensais) nos anos j e k (para $j > k$), respectivamente; n é o tamanho da série de dados; e a função *sign* é a variância de S , denotada por $[\text{VAR}(S)]$, assume o valor de 1 quando $x_j - x_k > 0$; 0 quando $x_j - x_k = 0$; e -1 quando $x_j - x_k < 0$, e é definida por:

$$\text{VAR}(S) = \frac{n(n-1)(2n+5)}{18} \quad (\text{Eq.5})$$

Nos casos em que há repetições de dados, a variância assume a expressão:

$$\text{VAR}(S) = \frac{1}{18} [n(n-1)(2n+5) - \sum_{p=1}^g t_p(t_p-1)(2t_p+5)] \quad (\text{Eq.6})$$

onde g é o número de grupos com dados repetidos e t_p é o número de dados no p -ésimo grupo. Para análise da tendência utiliza-se o valor de Z , no qual, o valor positivo ($Z > 0$) refere-se a uma tendência crescente e o valor negativo ($Z < 0$) refere-se a uma tendência decrescente.

Estimou-se também a inclinação e magnitude da tendência dos dados observados das descargas atmosféricas, utilizando o método não paramétrico de Sen (Sen, 1968; Sneyers, 1975). O método utiliza um modelo linear para estimar a inclinação da tendência e da variância dos resíduos, sendo constante no tempo. O teste foi utilizado por ser insensível aos valores de *outliers*, provendo uma medida mais realística das tendências em uma série temporal de dados, o que o torna mais rigoroso que a usual regressão linear. O método é determinado por:

$$f(t) = Qt - B \quad (\text{Eq.7})$$

onde Q é o valor referente a inclinação da reta e B é uma constante.

Para obter a estimativa de inclinação Q , primeiro calculam-se as inclinações de todos os pares de dados, definidos por:

$$Q_i = \frac{x_j - x_k}{j - k} \quad (\text{Eq.8})$$

para $j > k$.

Caso a série temporal seja composta por n valores x_j , obtêm-se $N=n(n-1)/2$ estimativas de inclinação Q_i . Deste modo, a inclinação estimada pelo método Sen é a mediana dos valores N de Q_i , os quais são classificados em ordem crescente e a estimativa Sen é dada por:

$$Q = Q_{[(N+1)/2]}$$

se N for ímpar, e

$$Q = \frac{1}{2}(Q_{[N/2]} + Q_{[(N+2)/2]})$$

se N for par.

(Eq.9)

Para obter uma estimativa de B na Eq. 7, são considerados n valores da diferença $x_i - Q_i$.

Outros trabalhos que também utilizaram os métodos de Mann-Kendall e da estimativa de inclinação Sen foram Fan e Wang (2011), Silva e Derecznski (2014), Santos et al., (2016a) e Santos et al., (2016b).

Além disso, foi identificado o evento extremo de cada grupo homogêneo, utilizando a técnica dos percentis (quantis), proposta por

Pinkayan (1966). Amplamente utilizada por Santos et al., (2013); Santos et al., (2016a); e Santos et al., (2016b), esta técnica baseia-se na distribuição da frequência acumulada, sendo que a aproximação da função densidade de probabilidade que descreve o fenômeno é tanto melhor quanto maior é o número de observações disponíveis. Os intervalos de cada percentil representam as probabilidades ou frequências esperadas para cada um dos eventos que podem ocorrer na sequência da série temporal de uma variável x . Esta metodologia consiste estritamente na distribuição em ordenamento crescente de uma série amostral contínua. Neste trabalho a série está constituída pelos totais de relâmpagos NS do trimestre de DJF, sendo atribuída uma probabilidade p univocamente a cada valor amostral. Assim, a série temporal pode ser distribuída sob a forma $\{x_1, x_2, x_3, \dots, x_n\}$, na qual x_1 representa o menor valor e x_n o maior valor da série. O índice n , portanto, representa o tamanho da amostra. Por fim, são calculados os quantis sendo que, neste caso em particular, percentis, visto que a divisão considera em cem partes. A seguir, são obtidas as ordens quantílicas e a classificação em grupos propriamente dita (Santos et al., 2016). Este trabalho utilizou a ordem quantílica $Q_{0,85}$ para estabelecer a classe “extrema” em relação aos valores observados (x_i) de relâmpagos NS.

O método dos percentis permite selecionar de forma objetiva e rápida os eventos climáticos anômalos e de interesse para estudo em um determinado período. Ressalta-se que, no cálculo do trimestre DJF do primeiro ano em estudo, 1999, foram considerados apenas os meses de janeiro, fevereiro, uma vez que os dados de descargas registrado em dezembro de 1998 não constava na base de dados. Na análise dos resultados ficou evidente que o uso de dois meses apenas, não alterou o resultado da classe para este trimestre, já que para 1999 o número de descargas desse trimestre ficou bem próximo daquele observado nos anos seguintes.

Resultados

O dendrograma obtido pelo método de agrupamento hierárquico de Ward para determinar as RH's, é apresentado na Figura 1, tendo como medida de dissimilaridade a distância euclidiana e usando como variável de agrupamento, a densidade de descargas NS, detectadas durante o trimestre de DJF (verão) do período de 1999 a 2014, no Estado de São Paulo.

No eixo das abscissas é mostrada a ordem de formação dos grupos, composto pelas estações. No eixo das ordenadas do dendrograma, tem-se o nível de similaridade da densidade de descargas

atmosféricas por estações. Cada subdivisão corresponde a um valor numérico que indica o nível em que ocorrem os agrupamentos. Quanto mais elevado for o valor, mais heterogêneas são as estações agrupadas.

Foram testadas três principais divisões de grupos. A primeira partição resultou em quatro grandes grupos homogêneos, com o maior nível de similaridade em torno de 32. A segunda partição, composta por seis grupos homogêneos, alcançou o

maior nível de similaridade em torno de 12. E a terceira partição apresentou nove pequenos grupos homogêneos, com maior nível de similaridade de aproximadamente 08. Desta forma, optou-se por fazer o corte no dendrograma com o teste da segunda partição, com seis grupos homogêneos, devido o mesmo ser mais representativo, quando comparado ao mapa do padrão médio de densidade de descargas no verão, observado por Santos et al., (2016a).

Figura 1. Dendrograma de densidade de descargas NS no trimestre de DJF (verão), para o Estado de São Paulo, considerando o período de 1999-2014.

A distribuição espacial das regiões homogêneas é apresentada na Figura 2, onde cada cor corresponde a um grupo homogêneo. Dado que a proximidade física das localidades não garante a semelhança climática entre elas, ocorreram casos em que são observadas as mesmas regiões homogêneas em áreas distintas do Estado. Isto ocorre, pois, a descarga atmosférica é um fenômeno meteorológico de alta variabilidade espacial.

Observa-se que os grupos formados pela AC estão de acordo com o mapa climatológico do verão do Estado, apresentado em Santos et al., (2016a), onde há uma grande área sobre o oeste do Estado com valores menos intensos de incidência de descargas e aumenta à medida em que se aproxima da região metropolitana de São Paulo (RMSP).

Figura 2. Regiões homogêneas (RH's) de descargas NS no verão, para o Estado de São Paulo, considerando o período de 1999-2014, com resolução espacial de 5x5 km.

Considerando o contexto atual dos grupos homogêneos de densidade de descargas, discutidos na figura anterior, para se investigar adequadamente a dinâmica da densidade de descargas intergrupos, faz-se necessário a normalização das áreas para futuras comparações de incidência de descargas.

Desta forma, visando analisar a extensão geográfica de cada grupo homogêneo, a Figura 3a mostra a área correspondente a cada região homogênea em quilômetros quadrados.

Como observado na Figura 2, a RH6 apresenta o maior domínio geográfico, localizado no oeste de São Paulo. A segunda RH em relação a dimensão territorial é a RH2, que ocupa a área adjacente a RH6, e seguidas a esta, têm-se as RH3, RH5, RH1 e RH4, respectivamente.

A Figura 3b apresenta a variabilidade temporal das descargas atmosféricas NS, com valores normalizados, para cada região homogênea do Estado de São Paulo, durante o verão. Verifica-

se que, de modo geral, a variabilidade da incidência de descargas é muito semelhante entre as RH's, com máximos e mínimos concomitantes, na maioria dos casos.

Nota-se que a maior concentração de descargas ocorre na RH4, composta por grandes centros urbanos como São Paulo, Guarulhos e Santo André (dentre outros), e elevação em torno de 1000 m. A menor concentração de descargas ocorre na RH6, composta por uma área mais plana (aproximadamente 200 m) e arborizada.

De toda a série, o verão dos anos de 2010 e 2013 se destacam pela grande incidência de descargas em todas as RH's, entretanto, somente o ano de 2013 foi classificado como evento extremo (EE), por apresentar valor acima do limiar do percentil 85. Nesta figura também é possível notar que a RH5 apresenta a variabilidade mais próximo da média de incidência de descargas de todas as RH's.

Figura 3. Regiões homogêneas de descargas atmosféricas NS no verão para o Estado de São Paulo, no período de 1999-2014, considerando: (a) área de cada RH; e (b) variabilidade temporal, com valores normalizados (descarga/área) do número de descargas por RH. A sigla ‘EE’ significa ‘evento extremo’.

Complementando a análise da dinâmica de incidência de descargas a Figura 4a mostra a distribuição das descargas sobre as RH's, por meio do box-plot.

Observa-se que a RH1 e a RH4 apresentam as maiores variabilidades e a RH6 a menor variabilidade. Já as RH's 2, 3 e 5 apresentam os valores de medianas próximas entre si. Além disso, nota-se que os *outliers* das RH's 1 e 5 se destacam expressivamente em comparação aos das demais RH's, e.g. o da RH6 que não chega ao mínimo da RH4, ressaltando o potencial das tempestades elétricas sobre a RH4.

A Figura 4b mostra os resultados dos testes de Mann-Kendall para as RH's. Ressalta-se que os valores apresentados nesta Figura não são normalizados, como os das demais figuras. A

tendência (Z) das RH's 1, 2, 3 e 6 apresentaram valores semelhantes entre si, entretanto, com magnitudes diferentes de, 719, 997, 1567 e 919 descargas/ano, respectivamente. Nota-se também a preponderância dos resultados do teste Z para a RH4 e RH5, com valores de 1,94 e 1,58, respectivamente, revelando a tendência crescente elevada da incidência de descargas, especialmente sobre a RH4, onde a mesma apresentou significância estatística de 90%.

Para esta RH, a magnitude da tendência foi semelhante à da RH6, a qual possui uma área (km²) muito maior que a RH4, conforme apresentando na Figura 3a, o que ressalta a o potencial das tempestades elétricas sobre a RMSP, visto que, esta é expressivamente menor que a RH6.

Figura 4. Regiões homogêneas de descargas atmosféricas NS no verão para o Estado de São Paulo, no período de 1999-2014, considerando: (a) box-plot (diagrama de caixa) de cada RH; e (b) valor não-normalizado dos resultados dos testes de Mann-Kendall (Z – tendência) e de Sen (Q – magnitude). No box-plot, a linha central representa o valor da mediana e o símbolo quadrado significa a média da série de dados. A linha inferior da caixa, delimita o primeiro quartil (Q1), que representa o valor de 25% dos dados e a linha superior da caixa, delimita o terceiro quartil (Q3), que representa o valor de 75% dos dados. As hastes horizontais representam os limites inferior e superior, e os círculos os valores atípicos (outliers).

Considerações finais

O presente trabalho abordou o estudo da mineração de dados para a identificação de regiões homogêneas de densidade de descargas atmosféricas no Estado de São Paulo, por meio da Análise de *Cluster*.

Os resultados mostraram que a técnica estatística de Análise de *Cluster* foi eficaz na identificação de padrões intrínsecos nos dados de descargas que ocorreram no Estado, apontando a existência de seis RH's, com máxima concentração de descargas sobre a RH4 (região metropolitana de São Paulo – RMSP) composta, em sua maioria, por áreas construídas e elevação em torno de 800 m. Os valores mínimos foram observados sobre a RH6 (oeste do Estado), compreendendo uma região mais arborizada e plana, com elevação em torno de 400 m. Apesar da RH4 ser expressivamente menor em dimensões geográficas que a RH6, as tendências revelaram magnitudes semelhantes entre ambas, o que revela o potencial das tempestades elétricas sobre a RMSP. As RH's 1 e 4 apresentaram as maiores variabilidades da incidência de descargas e a RH6 a menor. Já as RH's 2, 3 e 5 apresentam os valores de medianas próximas entre si.

Observou-se também que o verão do ano de 2013 apresentou o máximo de incidência de descargas atmosférica sobre todas as regiões do Estado, classificando este caso como o evento extremo mais representativo da série analisada.

Agradecimentos

Os autores agradecem ao Programa de Pós-Graduação em Ciência do Sistema Terrestre do Instituto Nacional de Pesquisas Espaciais (CCST/INPE) e ao apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Referências

- Del Río, S., Herrero, L., Fraileb, R., Penasa, A., 2011. Spatial distribution of recent rainfall trends in Spain (1961–2006). *Int. J. Climatol.* 31, 656-667.
- Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. *Journal of Intelligent Information Systems* 17, 107-145.
- Goossens, C., Berger, A., 1986. Annual and seasonal climatic variations over the northern hemisphere and Europe during the last century. *Annales Geophysicae* 4, 385-400.
- IPCC. Intergovernmental Panel in Climate Change, 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Disponível: <https://www.ipcc.ch/report/ar5/wg1/>. Acesso: 15 out. 2016.
- IPCC. Intergovernmental Panel in Climate Change, 2007. *Working Group II Contribution to the Intergovernmental Panel on Climate Change Fourth Assessment Report Climate Change: Climate Change Impacts, Adaptation and Vulnerability Summary for Policymakers*. Disponível: <https://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4-wg2-spm.pdf>. Acesso: 22 mai. 2012.
- Kendall, M.G., 1948. *Rank correlation methods*, 4 ed. Charles Griffin, London.
- Mann, H.B., 1945. Non-parametric tests against trend. *Econometrica* 13, 245-259.
- Oliver, J. F., Fairbridge, R. W., 1987 *Encyclopedia of climatology*. V. N. Reinhold, New York.
- Pinkayan, S., 1966. *Conditional Probabilities of occurrence of Wet and Dry Years Over a Large Continental Area*. Hydrology Papers, Fort Collins-Colorado-USA.
- Santos, A. P. P., Pinto Jr, O., Souza, E. B., Azambuja, R., Santos, S. R. Q., 2016a. Variabilidade espaço-temporal e identificação de Eventos Extremos de Descargas Atmosféricas no Estado de São Paulo durante o Verão. *Revista Brasileira de Geografia Física* 9, 346-352.
- Santos, A. P. P., Pinto Jr, O., De Souza, E. B., Azambuja, R., de Lima, F. J. L., Santos, S. R. Q., 2016b. *Eventos Climáticos Extremos de Descargas Atmosféricas sobre o Estado de São Paulo. Parte I: Aspectos Anuais e Sazonais*. *Revista Brasileira de Geografia Física* 9, 1346-1356.
- Sen, P. K., 1968. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association* 63, 1379-1389.
- Silva, W., Dereczynski, C. P., 2014. Caracterização Climatológica e Tendências Observadas em Extremos Climáticos no Estado do Rio de Janeiro. *Anuário do Instituto de Geociências* 37, 123-138.
- Sneyers, R., 1975. *Sur l'analyse statistique des séries d'observations*. Secrétariat de l'Organisation Météorologique Mondiale. Monografia. Geneve, OMM.
- Ward, J. H., 1963. Hierarchical grouping to optimize an objective function. *Journal American Association* 58, 236-244.