



## **Uso da entropia e da análise de agrupamento na avaliação da disponibilidade potencial de recursos hídricos do Nordeste do Brasil** (doi:10.4136/ambi-agua.146)

**Roni Valter de Souza Guedes<sup>1</sup>; Solange S. de Sousa<sup>2</sup>; Francisco de Assis S. de Sousa<sup>1</sup>**

<sup>1</sup>Universidade Federal de Campina Grande - UFCG

E-mail: roniguedes84@yahoo.com.br; fassis@dca.ufcg.edu.br

<sup>2</sup>Universidade Estadual da Paraíba - UEPB

E-mail: solange.salviano2@hotmail.com

### **RESUMO**

O objetivo deste estudo é definir, com base na teoria da informação da entropia de Shannon e na análise de agrupamento, métricas para representar a aleatoriedade da intensidade e da ocorrência da chuva mensal e usar essas métricas para avaliar a disponibilidade potencial de recursos hídricos (DPRH) na região Nordeste do Brasil. Aqui a entropia é expressa no contexto de probabilidade de ocorrência da “intensidade de variáveis hidrológicas”, isto é, no contexto da estrutura da IE (Intensidade de Entropia). O critério para realização da amostragem dos dados de chuva observados foi o período de registro. Pluviômetros com observações inferiores a 25 anos ficaram de fora deste estudo. Foram utilizados dados de totais mensais precipitados de 874 postos pluviométricos da região Nordeste do Brasil. Destaca-se neste estudo a perfeita adequação da metodologia aplicada à área de estudo. Os resultados apresentaram uma configuração muito boa com os tipos climáticos de Köppen, e com os sistemas meteorológicos atuantes em cada área.

**Palavras-chave:** Teoria da informação de Shannon; precipitação; k-média.

## **Use of entropy and clustering analysis for the evaluation of water resources potential availability in the Northeastern Brazil**

### **ABSTRACT**

The aim of this study was to define, based on Shannon entropy theory and on cluster analysis and metrics to represent the randomness of the intensity and occurrence of monthly rainfall and using these metrics to evaluate the potential availability of water resources (PAWR) in northeastern Brazil. Here the entropy is expressed in the context of probability of occurrence of the "intensity of hydrological variables," i.e, within the structure of Entropy Intensity. The criterion to perform the sampling of rainfall data was the period of data collection. Rain gauges with less than 25 years of observations were left out of this study. Data of total monthly precipitation from 874 rain gauges in the Northeast of Brazil were analyzed. This study showed the perfect adaptation of the methodology applied for the studied area. The results showed a perfect configuration with the Köppen climatic types and the weather systems operating in each area.

**Keywords:** Information theory of Shannon; precipitation; k-means.

## **1. INTRODUÇÃO**

A chuva constitui a contribuição primária para o ciclo hidrológico, e pode, portanto, ser utilizada para representar a disponibilidade potencial de recursos hídricos (DPRH) de uma área. A irregularidade (ou incerteza) da intensidade e ocorrência de chuva no tempo é uma

das primeiras restrições para o desenvolvimento dos recursos hídricos e práticas de uso da água. Ao se desenvolver uma estratégia de âmbito nacional ou regional para os recursos hídricos, incluindo-se também nessa estratégia demandas atual e futura de água, atributos para quantificar irregularidades em todo o país ou região podem ser fatores dominantes para a tomada de decisão de prioridades para o desenvolvimento de grandes áreas ou demarcar os limites a fim de estabelecer a necessidade e a viabilidade de desenvolvimento. Uma medida de incerteza ou irregularidade de uma variável é a sua entropia. A entropia pode ser calculada se a função distribuição de probabilidade ou função densidade de probabilidade (FDP) da variável aleatória for dada de forma discreta ou contínua, usando-se a teoria de informação da entropia, definida por Shannon (1948). De acordo com essa teoria, a incerteza de uma variável ou a incerteza de sua distribuição de probabilidade é o valor esperado negativo do logaritmo da FDP da variável.

Caracteristicamente o acréscimo da aleatoriedade ou irregularidade de uma variável causa decréscimo no coeficiente de assimetria da distribuição de probabilidade e, dessa forma conduzindo a amostra para o aumento da entropia. A entropia assume valor máximo quando a distribuição de probabilidade é uniforme sem qualquer desvio, enquanto reduz-se a zero quando um particular valor da variável acontece com probabilidade um. A aplicação direta da teoria da entropia a dados observados de chuva pode quantificar suas irregularidades. Por exemplo, Al-Zahrani e Husain (1998) utilizaram a teoria da informação de Shannon para otimizar e detectar irregularidades em uma rede hidrometeorológica, localizada no sudeste da Arábia Saudita. Os resultados mostraram que das setenta estações que compõe a rede apenas 45 seriam necessárias, dessas, algumas deveriam ser remanejadas e as demais descartadas. Outro trabalho de aplicação direta da teoria de Shannon foi efetuado por Kawachi et al. (2001), que utilizaram essa teoria para estimar e construir mapas de entropia (isoentropia) em todo o território japonês. Obtiveram como resultado principal a disponibilidade potencial de recursos hídricos em todo o país. A variância, uma técnica convencional e muito utilizada, também pode quantificar irregularidades em torno do valor médio de uma distribuição de frequência. Porém, a entropia é mais poderosa e geral do que a variância. Como se sabe (Kagan et al., 1973), quando a entropia de uma variável aleatória é maximizada sob a restrição de que a variância e a média são conhecidas para uma série de valores da variável, a função densidade de probabilidade para essa série reduz-se a distribuição normal. Isso implicitamente significa que o uso da variância é limitado para o caso de a distribuição ser a normal. Em outras palavras, a metodologia baseada na entropia é aplicável a qualquer distribuição, sendo ela conhecida ou desconhecida a priori. Silva et al. (2003), avaliou a variabilidade espacial e temporal da precipitação pluvial para o estado da Paraíba, com base na teoria da entropia.

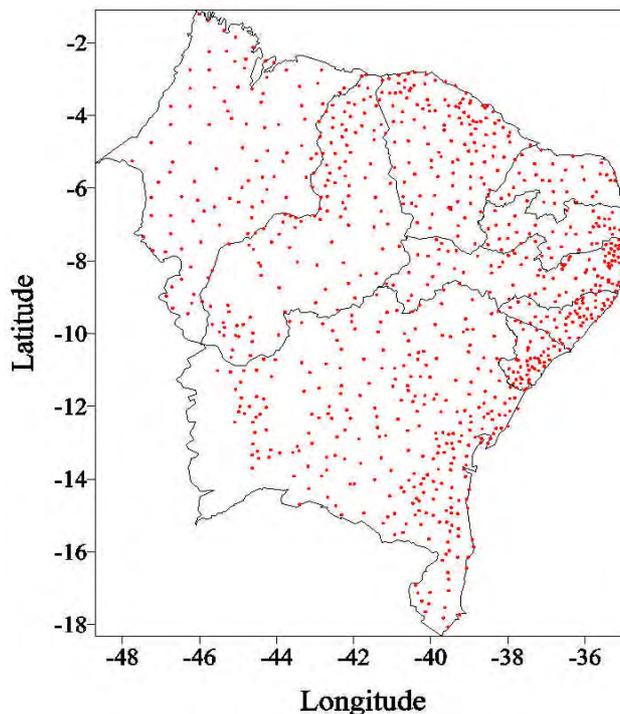
O objetivo deste estudo é definir, com base na teoria da informação da entropia de Shannon e na análise de agrupamento, métricas para representar a aleatoriedade da intensidade e da ocorrência da chuva mensal e usar essas métricas para avaliar a disponibilidade potencial de recursos hídricos (DPRH) na região Nordeste do Brasil.

## **2. MATERIAL E MÉTODOS.**

Foram utilizados dados de totais mensais precipitados de 874 postos pluviométricos da região Nordeste do Brasil. Os softwares usados foram Excel 2003 nos ajuste dos dados e construção dos gráficos, o Surfer 2007 na plotagem dos mapas e o SPSS 13.0 (Statistical Package for the Social Sciences) para efetuar a divisão de agrupamento pelo método K - Means. A Figura 1 exibe a distribuição geográfica dos postos pluviométricos na região estudada.

Neste estudo, primeiro a intensidade de chuva mensal (i.e. o total de chuva durante um mês) é considerada como uma variável aleatória. E a relação entre a intensidade e sua

probabilidade de ocorrência (i.e. frequência), dentro de um determinado período de tempo é computada por formulação baseada na entropia. A entropia é então obtida e será denominada “intensidade de entropia (IE)”. Segundo, a taxa da chuva mensal em relação à soma (totais mensais precipitados durante um ano (i.e.chuva anual)) é considerada como outra variável aleatória. Essas intensidades relativas de chuvas durante um ano refletem o caráter probabilístico de ocorrência da chuva no ano. Visto que nessas taxas estão implicitamente consideradas as taxas de distribuição das chuvas durante todos os meses do ano, a entropia assim calculada é chamada de “distribuição de entropia (DE)”.



**Figura 1.** Distribuição espacial dos postos pluviométricos utilizados.

Para se determinar a DPRH de uma área, usando IE e DE, é necessário distribuir espacialmente DE e IE de forma significativa. Para isso foi empregado o método do agrupamento não hierárquico *k-means*. Essa metodologia foi aplicada para delimitar atributos regionais da DPRH na região estudada. Singh (1997) faz extensa revisão de aplicações da teoria da informação da entropia para vários problemas relacionados à hidrologia e aos recursos hídricos. Em todas as aplicações, contudo, a entropia é expressa no contexto de probabilidade de ocorrência da “intensidade de variáveis hidrológicas”, isto é, no contexto da estrutura da IE. Focalizando a variabilidade temporal da chuva diária, Maruyama e Kawachi (1998) e Kawachi et al. (2001) empregaram a informação da entropia para estimar o grau de incerteza da ocorrência de chuva no tempo ou estimar a partição temporal da chuva de um ano para outro, baseando-se na função densidade de probabilidade dessa chuva, aleatoriamente dividida em tempos fragmentados. Em geral, é mais fácil se obterem dados de chuva mensais do que diários. De qualquer forma, a DE em base mensal, como apresentada aqui, é também uma alternativa viável para base diária.

## 2.1. Caracterização da região estudada

A maior parte de seu território é formada por extenso planalto, antigo e aplainado pela erosão. Em função das diferentes características físicas que apresenta, a região Nordeste encontra-se dividida em quatro sub-regiões: Zona da Mata, Agreste, Sertão e o Polígono das Secas. (IBGE, 2002).

**Zona da Mata** - É a mais povoada, a que concentra o maior número de indústrias e a mais urbanizada. Estende-se do estado do Rio Grande do Norte ao sul do estado da Bahia, numa faixa litorânea de até 200 km de largura. Possui clima tropical úmido, com chuvas mais frequentes na época do outono e inverno, exceto no sul do estado da Bahia, onde se distribuem uniformemente por todo o ano. O solo dessa área é fértil e a vegetação natural é a Mata Atlântica, já praticamente extinta e substituída por lavouras de cana-de-açúcar, desde o início da colonização do país.

**Agreste** - É a área de transição entre a Zona da Mata, região úmida e cheia de brejos, e o sertão semiárido. Nessa sub-região, os terrenos mais férteis são ocupados por minifúndios, onde predominam as culturas de subsistência e a pecuária leiteira.

**Sertão** - Extensa área de clima semiárido, nos estados do Rio Grande do Norte e Ceará o sertão chega até o litoral. Os solos desta sub-região são rasos e pedregosos, as chuvas escassas e mal distribuídas e as atividades agrícolas sofrem grande limitação. A vegetação típica do sertão é a caatinga. Nas partes mais úmidas existem bosques de palmeiras, especialmente a carnaubeira. O rio São Francisco é o maior rio da região e única fonte perene de água para as populações que habitam as suas margens. Nele existem várias represas e usinas hidrelétricas, como a de Sobradinho, em Juazeiro, estado da Bahia, e a de Paulo Afonso, na divisa dos estados da Bahia e Pernambuco. O sertão apresenta muitos rios temporários e baixo índice pluviométrico, que também tem como característica a irregularidade, pois ocorrem longos períodos de estiagem.

**Polígono das Secas** - Delimitada em 1951 para combater as secas do Nordeste, essa área abrange praticamente todos os estados do Nordeste, com exceção do Maranhão e o litoral leste da região. As secas de 1979 a 1984 e 1989 a 1990 atingiram 1510 municípios do Nordeste brasileiro. O combate tradicional às secas vem sendo feito com a construção de açudes e distribuição de verbas aos prefeitos dos municípios atingidos. Recentemente, no entanto, o governo federal começou a implementar projetos na região, que visam à solução definitiva do problema de convivência do homem nordestino com a seca. Dentre tais projetos destaca-se o Projeto Áridas, financiado pelo Banco Mundial.

A faixa de transição entre o sertão semiárido e a região Amazônica denomina-se Meio-Norte, apresentando clima bem mais úmido e vegetação exuberante à medida que avança para o oeste. A vegetação natural dessa área é a mata dos cocais, onde se encontra a palmeira babaçu. A população da região Nordeste totaliza 44.768.201 habitantes, o que representa 28,9% do total do país. Sua densidade demográfica é de 28,05 habitantes por km<sup>2</sup> e a maior parte da população se concentra na zona urbana (60,6%). As principais metrópoles regionais são as cidades de Salvador, capital do estado da Bahia, Recife, capital do estado de Pernambuco, e Fortaleza, capital do estado do Ceará, (IBGE, 2002).

## 2.2. Seleção de pluviômetros

Aqui se apresenta o critério para realização da amostragem dos dados observados, para se avaliar a distribuição espacial da DPRH em termos da chuva. Os dados disponíveis de totais mensais de chuva foram considerados. Pluviômetros com amostras de dados inferiores a 25 anos de observações ficaram de fora deste estudo. Os arquivos de dados de chuvas diárias foram manipulados para se obterem valores mensais de chuvas. Usando-se esses valores mensais, as estimativas da IE e da DE foram estimadas para cada local ou pluviômetro. Dados adicionais foram obtidos em Vose et al. (1992).

## 2.3. Intensidade de entropia (IE)

Shannon (1948) definiu a entropia  $H$  para uma série de  $p_1, p_2, \dots, p_n$  probabilidades como:

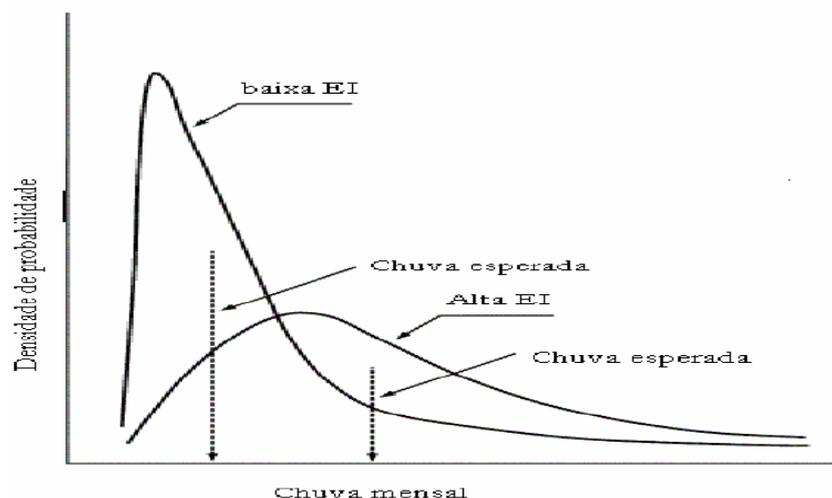
$$H = -\sum_{i=1}^n P_i \log P_i \quad [1]$$

Quando se considera a intensidade ou total mensal de chuva como uma variável aleatória e  $p_i$  como sua probabilidade de ocorrência em uma série temporal de chuva, a entropia de intensidade (IE) pode ser calculada usando a equação (1). As probabilidades  $p_i$  para cada pluviômetro serão expressas de forma discreta levando em conta todos os valores de chuvas mensais disponíveis e suas ocorrências de probabilidades. A IE foi avaliada como segue:

- Reuniram-se todos os anos completos de observações mensais de chuva disponíveis em uma localidade ou pluviômetro. Assim, 12 meses x m anos = N dados mensais de chuva, ou seja, com esse cálculo se obtém o tamanho da amostra N para cada localidade;
- Dividiram-se as amostras de dados de chuvas mensais em n classes de intervalos iguais;
- Contaram-se as frequências  $f_i$  para cada classe i e, em seguida, construíram-se tabelas de distribuição de frequências;
- Calcularam-se as frequências relativas ( $f_i/N$ ) para cada classe i e, em seguida obteve-se a função densidade de probabilidade de forma discreta para todas as faixas de chuvas mensais;
- Calcularam-se as entropias IE usando a equação (2) em termos dessas frequências relativas.

$$IE = -\sum_{i=1}^n (f_i / N) \log_2 (f_i / N) \quad [2]$$

em que n é o número de classes e  $f_i$  é a frequência para classe i. Aqui, a unidade da IE é “bit” com o logarítmico na base 2. IE, definida em uma faixa semi-infinita de  $0 < IE < \infty$  é uma medida para decifrar a intensidade aleatória da chuva mensal. Intensidade menos aleatória é medida por IE menor, indicando uma distribuição de frequência de ocorrências de chuvas mensal mais assimétrica. De forma contrária, intensidades mais aleatórias resultam em IE maior, estendendo-se sobre uma faixa maior de chuvas mensais. Porém, pode ser notado que a função de densidade probabilidade das intensidades das chuvas é sempre definida sobre uma abscissa positiva, incluindo sua origem zero, devido a não-negatividade da chuva. É fácil verificar que um acréscimo em IE resulta em acréscimo no valor esperado da chuva mensal, aplainando o gráfico da função (Figura 2). Isso sugere que IE é positivamente correlacionada à chuva mensal total esperada e pode ser, portanto, uma alternativa para o total de chuva agregada.



**Figura 2.** Acréscimo da chuva esperada com o aumento de IE.

**Fonte:** Kawachi et al. (2001).

## 2.4. Distribuição de entropia (DE)

Usando a equação (1), Kawachi et al. (2001) definiram a distribuição de entropia (DE) para chuva diária. De forma similar, a DE para chuva mensal pode ser definida. Seja  $r_i$  a chuva mensal agregada durante o  $i$ -ésimo mês de um ano. Assim, a chuva agregada durante o ano  $R$  pode ser expressa pela soma de  $r_i$ , com  $i=1$  até 12:

$$R = \sum_{i=1}^{12} r_i \quad [3]$$

em que os valores de  $r_i$  podem ser zero para alguns dias e diferentes de zero para outros.

Então cada  $r_i$  desse vai representar uma série de precipitação e assim podemos encontrar a frequência relativa da precipitação ( $p_i$ ) que será obtida dividindo-se  $r_i$  pelo tamanho total da amostra ( $R$ ), ou seja:

$$p_i = \frac{r_i}{R} \quad [4]$$

Agora, empregando-se a teoria da informação de Shannon, podemos encontrar ED que será a distribuição representativa da variabilidade temporal da chuva mensal durante um ano. ED pode ser escrito como:

$$DE = -\sum_{i=1}^{12} p_i \log_2 p_i = -\sum_{i=1}^{12} (r_i / R) \log_2 (r_i / R) \quad [5]$$

em que a unidade de DE também é “bit” com o logarítmico na base 2. A IE, previamente definida, é exclusivamente estimada no respectivo pluviômetro ou localidade, enquanto DE definida para um ano é estimada para cada um dos anos considerado para os respectivos pluviômetros ou locais. Portanto, os valores de DE obtidos para certo pluviômetro são simplesmente representados pela média desses valores. DE estima a aleatoriedade interanual (ou igualdade temporal) de eventos de chuvas mensais no tempo. Por definição, a equação (4) explica que, quando o total de chuva anual é uniformemente distribuído para cada um dos 12 meses, com probabilidade igual a  $1/12$ , DE assume seu valor máximo ( $H = \log_2 12$ ). O valor mínimo DE ( $H=0$ ) ocorre quando a distribuição é feita para apenas um dos 12 meses com uma probabilidade igual a um. Isso indica que DE assume um valor em uma faixa finita de  $0 - \log_2 12$ . Quanto maior o valor de DE tanto maior será a DPRH no tempo, com menor variabilidade mensal ou sazonal da chuva.

## 2.5. Padronização

Baseado nas equações (2) e (5), duas séries de valores IE e DE podem ser estimadas para todas as localidades. Duas outras diferentes séries: IE\* e DE\* podem também ser obtidas com base nas duas primeiras, respectivamente. A padronização dessas séries ajusta suas diferentes propriedades estocásticas, de modo que o valor da média é igual a zero e o do desvio-padrão igual a um. A equação (6) resume essa operação estatística de normalização da variável:

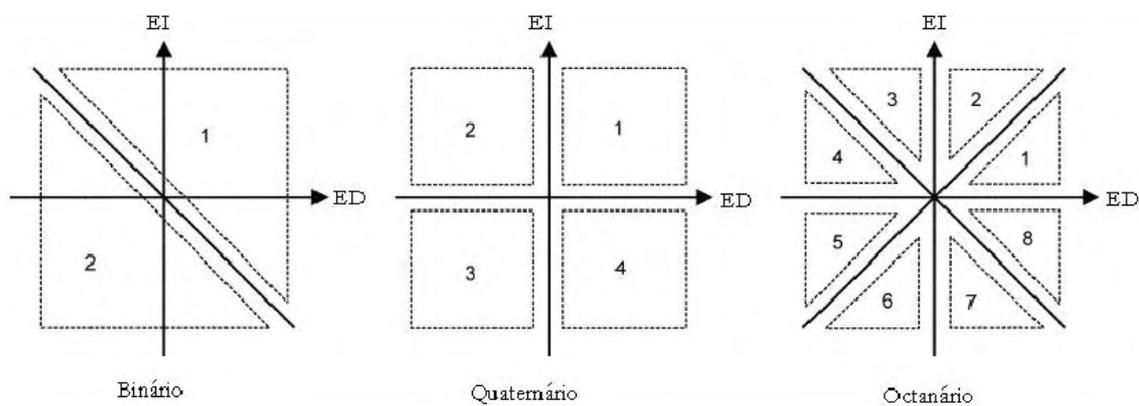
$$X' = \frac{X - m_x}{\sigma_x} \quad [6]$$

em que  $X'$  é o valor padronizado de IE ou DE,  $X$  é o valor estimado em “bit” (IE ou DE), e  $m_x$  e  $\sigma_x$  são a média e o desvio-padrão para a  $X^*$  (IE\* ou DE\*), respectivamente. Por essa

padronização, um diagrama de dispersão pode ser esboçado para ilustrar o local-dependência da relação entre IE e DE. Desse modo, a DPRH pode ser avaliada em sentido relativo para a área de estudo. Doravante, as anotações IE e DE serão usadas para entropias padronizadas.

## 2.6. Métodos de agrupamento

Diferentes pares de valores de IE\* e DE\* para diferentes locais da região foram plotados em diagrama de dispersão. As coordenadas de um ponto neste diagrama caracterizam a DPRH no local correspondente ao pluviômetro. Nota-se que os eixos do diagrama, ao usar valores padronizados de IE e DE, se tornam linhas dos valores médios de IE e DE, respectivamente. O agrupamento dos pontos pode conduzir a uma classificação significativa e delimitar zonas de recursos hídricos. Quanto maior o número de partições, tanto mais refinado será o agrupamento e a análise. Algumas partições de agrupamentos são mostradas na Figura 3, para três casos, de dois, quatro e oito agrupamentos.



**Figura 3.** Tipo de partições obtidas por meio de métodos de agrupamento.

O segundo agrupamento categoriza os pontos contra os valores médios de IE e DE ou contra  $IE=0$  e  $DE=0$  em quatro grupos. Portanto, o terceiro quadrante, no qual os valores de IE e DE são menores do que suas respectivas médias será, em média, categorizada como uma área pobre em recursos hídricos, apresentando baixa oferta e alta variabilidade da chuva. O primeiro quadrante é caracterizado como uma área abundante e perene de chuva. O segundo quadrante é uma zona onde a chuva é relativamente abundante, mas concentrada no tempo. O quarto quadrante é uma área com chuvas perenes, mas de curta duração. Explicações semelhantes podem ser dadas para outros tipos de partições.

### 2.6.1. Método de agrupamento não hierárquico *K-means*

O método *K-means* possui um algoritmo de aprendizagem que organiza  $n$  objetos em  $k$  partições em que cada uma representa um grupo. O funcionamento dele é descrito por dividir os objetos em  $k$  grupos e, a partir da similaridade do valor da média dos atributos numéricos, agrupa os demais objetos em torno desses grupos previamente indicados.

É um método de partição que fornece indicações mais precisas sobre o número de conglomerados a ser formado. Esse método é um dos mais utilizados quando se têm muitos objetos para agrupar, com pequenas variações. O critério mais utilizado para detectar homogeneidade dentro do grupo e heterogeneidade entre os grupos é o da soma dos quadrados residual baseado na Análise de Variância (Bussab et al., 1990). Assim, quanto menor for esse valor, mais homogêneos serão os elementos dentro de cada grupo e melhor será a partição.

Abaixo são apresentados os passos envolvidos no algoritmo *k*-médias, em que *k* é o número de *clusters* (grupos) desejado e informado a priori:

- (1) Determinar as posições iniciais dos *k* centroides dos grupos;
- (2) Alocar cada elemento ao grupo do centroide mais próximo;
- (3) Recalcular os centros dos grupos a partir dos elementos alocados;
- (4) Repetir os passos de 2 a 4 segundo algum critério de convergência.

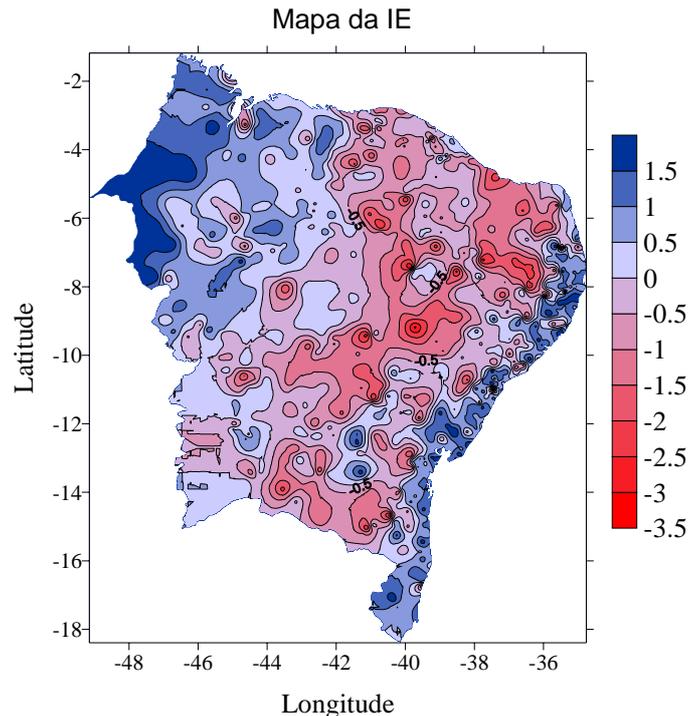
É possível mostrar que o método *k-means* produz um conjunto de grupos que minimiza o erro quadrático com relação aos centros de gravidade de cada grupo. Esse método só produz bons resultados quando os grupos são “nuvens” compactas de dados, bem separadas umas das outras. A vantagem do método é sua eficiência em tratar grandes conjuntos de dados. Suas desvantagens estão no fato do usuário ter que fornecer o número de grupos e no fato de ser sensível a ruídos, já que objetos com valores altos podem causar grande alteração no centro de gravidade dos grupos e assim, distorcer a distribuição dos dados neles. Mais detalhes sobre os métodos de agrupamento e sobre o método *k-means* pode ser encontrado em Hartigan (1975) e Hartigan e Wong (1979).

### 3. RESULTADOS E DISCUSSÃO

Com base nas equações (2) e (5) calcularam-se a Intensidade de Entropia (IE) e Distribuição de Entropia (DE), respectivamente. A equação (6) foi utilizada para padronizar essas informações. A Figura 3 exibe as possíveis partições utilizadas neste estudo.

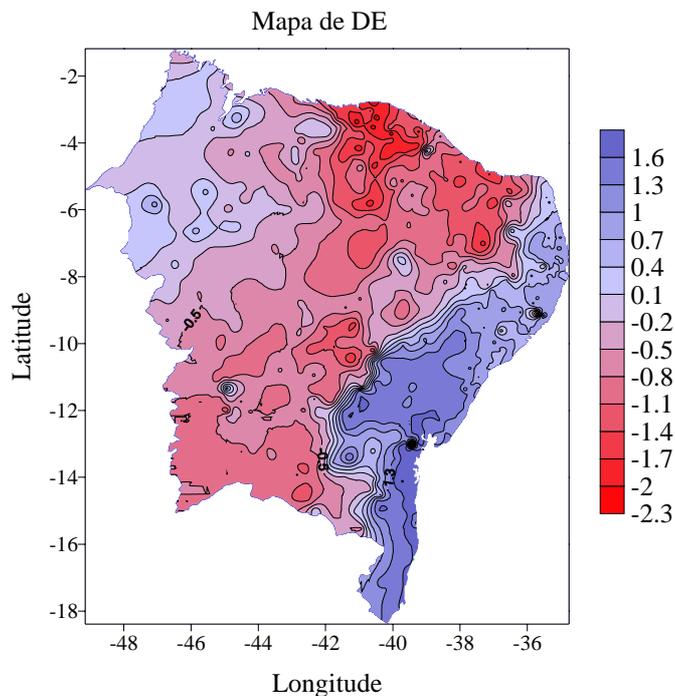
Analisando o mapa da IE (Figura 4), nota-se que os maiores valores estão localizados no litoral leste e no oeste do Estado do Maranhão. Já os menores valores de IE estão localizados na região semiárida, compreendidos desde o sul e centro do Estado da Bahia até o litoral norte do Rio Grande do Norte, Ceará e Piauí.

Como visto na Metodologia, valores baixos de IE indicam menor aleatoriedade da intensidade das chuvas. Esse fato conduz a uma distribuição simétrica de frequência de chuvas mensais. As maiores IE dos totais mensais precipitados podem se estender por meses, variando de um local para outro. Acredita-se que as áreas que apresentam alta IE podem ser influenciadas pelos sistemas meteorológicos atuantes nessas regiões, como Ondas de Leste, Brisas Marítimas e ZCIT (Zona de Convergência Intertropical) e os VCANs (Vórtice de Circulação dos Altos Níveis). Nota-se ainda que na direção do interior da região semiárida, os valores de IE diminuem, destaca-se, portanto, alta variabilidade na intensidade das chuvas nessa área.



**Figura 4.** Mapa da intensidade de entropia (IE).

Quando a Distribuição de Entropia (DE), Figura 5, dos totais mensais precipitados possui valores altos indica que o tempo da DPRH é maior. Ou seja, as áreas com DE maior apresentam menor variabilidade da chuva mensal ou sazonal. Aqui se pode ver claramente que os maiores valores da entropia de distribuição estão no litoral leste com destaque para o sul da Bahia, justificando a maior DPRH que era esperada. Além dessa área, tem-se melhores condições para aproveitamento da água no lado oeste do estado do Maranhão.



**Figura 5.** Mapa da distribuição de entropia.

Essa melhor distribuição nas regiões citadas se explica pelos maiores valores de precipitação e pela melhor frequência de sistemas meteorológicos atuantes. Como esses

sistemas perdem força dentro do continente, a região central apresenta valores inferiores às demais. Assim, podemos dizer que há uma região de alta variabilidade de chuva na parte central e litoral norte (CE e RN), enquanto que no litoral leste e oeste do MA a variabilidade é baixa, com chuvas mais concentradas e frequentes.

A Figura 6 exhibe a dispersão padronizada binária dos valores de IE e DE. Segundo o método, a DPRH apresenta-se da seguinte forma: as áreas e pontos na cor azul são áreas ou locais de chuvas abundantes, com boa distribuição e relativamente concentradas no tempo. Já as áreas e locais na cor laranja são áreas e ou locais com alta variabilidade das chuvas, pois  $IE$  e  $DE < 0$ . As áreas ou pontos na cor laranja representam, ainda, chuvas de curta duração ou concentradas no tempo. Desse modo, pode-se entender que a cor vermelha (áreas e locais) é representativa de áreas ou locais pobres de Disponibilidade Potencial de Recursos Hídricos - DPRH.

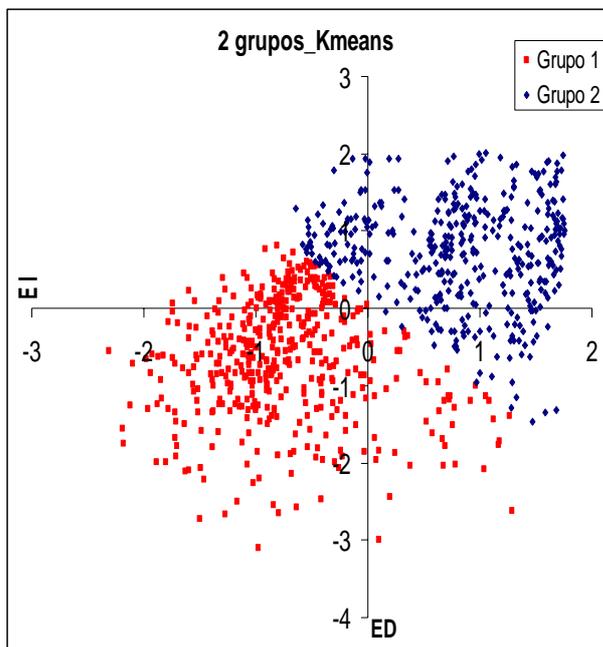


Figura 6. Dispersão padronizada binária.

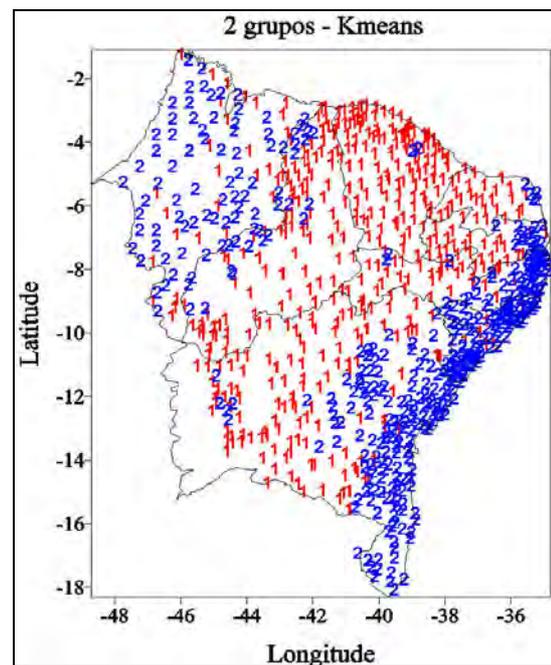


Figura 7. DPRH no Nordeste do Brasil.

A Figura 7 mostra a distribuição geográfica da Disponibilidade Potencial de Recursos Hídricos – DPRH. Nota-se que desde o sul do Estado da Bahia até o norte do Estado do Rio Grande do Norte, destacando-se a Chapada Diamantina e Raso da Catarina, na Bahia, e pontos isolados no interior dos Estados de Pernambuco e da Paraíba, a DPRH é muito boa (áreas ou locais de chuvas abundantes, com boa distribuição e relativamente concentradas no tempo). O Estado do Maranhão apresenta pontos de DPRH classificada como muito boa distribuída por todo o estado até as margens do Rio Parnaíba, mas também possui alguns pontos em vermelho onde a classificação não é boa. As chuvas do Estado do Maranhão são fortemente influenciadas pelo transporte de umidade da Floresta Amazônica, pela descida da ZCIT de norte. Na Chapada do Araripe, no Estado de Pernambuco, destacam-se pontos isolados de boa DPRH. No Estado do Ceará destacam-se pontos isolados de boa DPRH na Serra de Baturité.

A Figura 8 exhibe a dispersão padronizada de IE e DE, com base na partição quartenária. A Figura 9 aloca esses pontos geograficamente para constituir áreas ou pontos isolados e classificá-las segundo os critérios dispostos na Tabela 1. Nessa nova partição quartenária, nota-se o refinamento da distribuição de pontos e classificação de áreas da região em estudo. Nota-se que nela ocorreram subdivisões e melhoria da classificação quando comparada com a classificação anterior da DPRH, com base na partição binária.

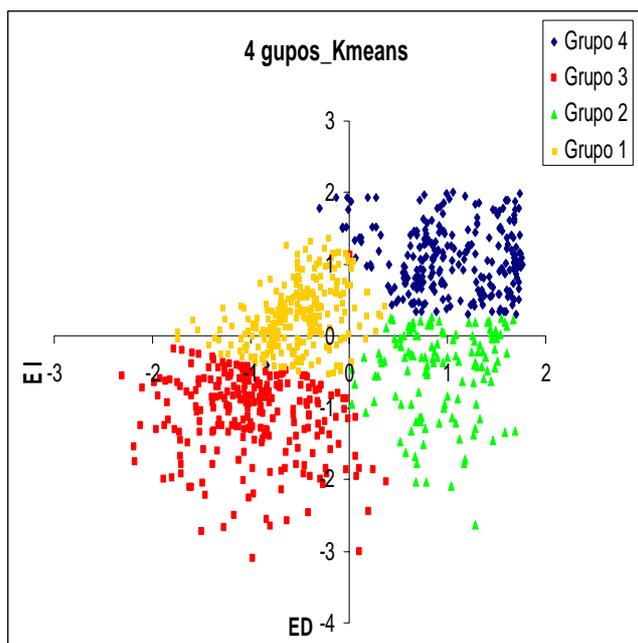


Figura 8. Dispersão padronizada quartenária.

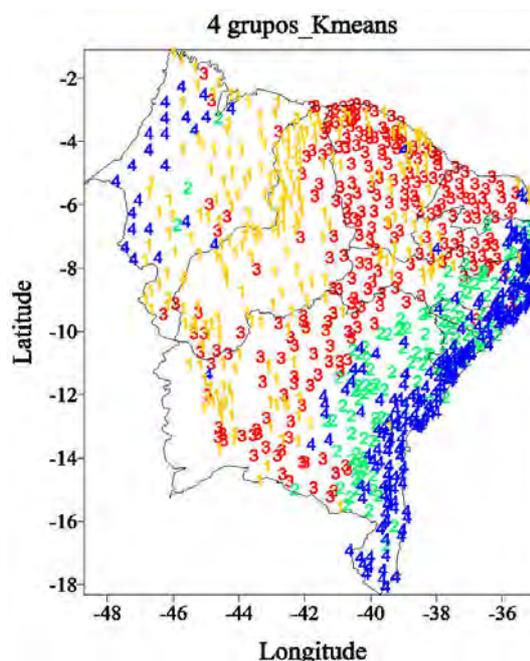


Figura 9. DPRH no Nordeste do Brasil.

Novamente, a cor azul predomina, desde o sul do Estado da Bahia até o leste do Estado do Rio Grande do Norte, mas com estreitamento dessa faixa, do litoral para o interior do continente. Segundo a classificação da Tabela 1, em toda essa zona ou área litorânea, as chuvas são abundantes e perenes, portanto, a DPRH na área é muito boa.

A fronteira da região semiárida, vizinha da faixa litorânea (verde claro); caracteriza-se, agora, como zona de chuvas perenes, mas de curta duração. Basicamente, essa zona se situa na Chapada Diamantina, Raso da Catarina, no Estado da Bahia e Agreste do Estado de Pernambuco. Pontos isolados podem ser vistos no Agreste do Estado da Paraíba, no interior de Pernambuco, mais precisamente, na Chapada do Araripe e interior do Estado do Maranhão.

Tabela 1. Classificação das chuvas de acordo com a dispersão padronizada de IE e DE.

Quadrante	Cor	Sinal de IE	Sinal de DE	Classificação das chuvas
Primeiro	Azul	+	+	Zona abundante e perene de chuvas
Segundo	Laranja	+	-	Zona com chuva relativamente abundante, mas concentrada no tempo.
Terceiro	Vermelho	-	-	Zona pobre em recursos hídricos com baixa oferta e alta variabilidade das chuvas
Quarto	Verde Claro	-	+	Zona com chuvas perenes, mas de curta duração.

Em quase toda região semiárida predomina a classificação na cor vermelha (zona pobre em recursos hídricos com baixa oferta e alta variabilidade das chuvas). Alguns pontos isolados podem ser notados no âmbito do Estado do Maranhão. Partes do oeste do Estado da Bahia, leste do Estado do Maranhão e o Estado do Piauí, principalmente ao longo das margens do Rio Parnaíba, podem ser classificadas como zona com chuvas relativamente abundantes, mas concentradas no tempo (cor laranja).

Nas Figuras 10 e 11, pode ser vista a divisão da DPRH em oito grupos (octanário), em que se nota um refinamento ainda maior, que indica algumas características interessantes da

região. Pode-se afirmar que o litoral sul da Bahia possui a melhor relação de intensidade e distribuição temporal da chuva, enquanto os litorais de Sergipe, de Alagoas, de Pernambuco e da Paraíba e ainda o lado oeste do Maranhão, possuem intensidade semelhantes, mas com menor grau de distribuição temporal.

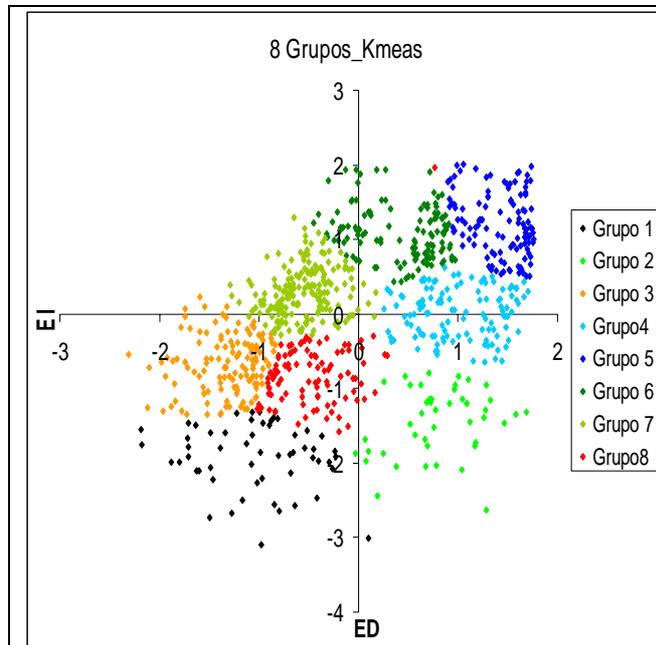


Figura 10. Dispersão padronizada octanária.

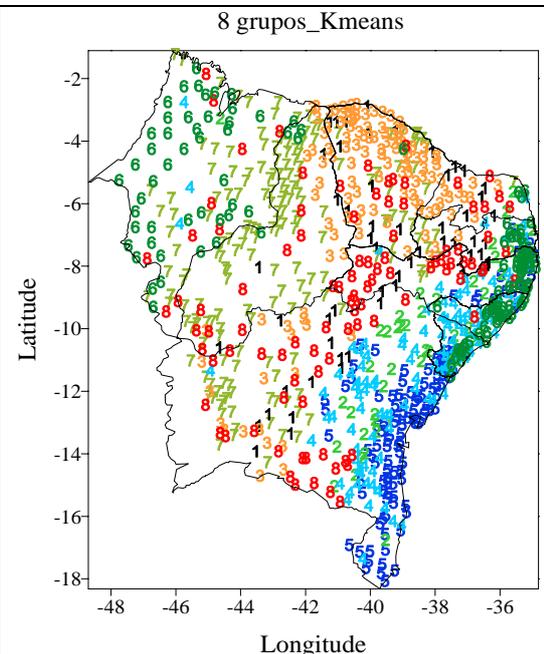


Figura 11. DPRH no Nordeste do Brasil.

A faixa em azul claro, que separa parte da faixa litorânea da parte central do Nordeste. Possui menor intensidade que as duas anteriores, mas se mantém com boa distribuição temporal, sendo assim, considerada com melhor DPRH que a faixa verde cana (segundo quadrante) com pontos distribuídos entre o leste do Maranhão, Estado do Piauí e oeste da Bahia, pois essa última aparece como razoável sem distribuição temporal.

Os pontos mais críticos, que apresentam péssima DPRH, estão pintados em preto, espalhados numa faixa central entre o sul da Bahia o litoral do Ceará e Rio Grande do Norte, misturando-se com os pontos vermelhos e laranja, formando o grupo de baixa intensidade, menor variabilidade e maior aleatoriedade da chuva.

## 4. CONCLUSÕES

Em todos os gráficos e mapas ficou evidente que o litoral leste do Nordeste e o oeste do Maranhão formam as regiões mais propícias, em termos de DPRH, ao cultivo de muitas culturas, pois possuem os melhores valores de intensidade e a melhor distribuição de chuvas. Com destaque para o litoral da Bahia, que tem a melhor DPRH da região Nordeste. Os piores resultados foram encontrados na parte central do Nordeste, sendo essa faixa a mais carente de recursos hídricos, necessitando de um maior investimento e implantação de projetos permanentes.

É interessante observar que os estados do Ceará e do Rio Grande do Norte não apresentaram bons resultados nem mesmo na faixa litorânea, como ocorre em outros estados. Esses dois estados apresentam apenas alguns pontos isolados de boa DPRH.

Destaca-se neste estudo a ótima adequação da metodologia aplicada à área de estudo. Os resultados apresentaram uma configuração muito boa com os tipos climáticos de Köppen, tropical úmido, úmido, semiárido e com os sistemas meteorológicos atuantes em cada área, conforme mencionado anteriormente, ZCIT, Ondas de Leste, VCANs, Brisas Marítimas, e outros sistemas locais de menor expressão.

## 5. AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

## 6. REFERÊNCIAS

- AL-ZAHRANI, M.; HUSAIN T. An algorithm for designing a precipitation network in the south-western region of Saudi Arabia. **Journal of Hydrology**, v. 205, n. 3/4, p. 205-216, 1998.
- BUSSAB, W. O.; MIAZAK, E. S.; ANDRADE, D. F. Introdução à análise de grupamentos. In: SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 9., 1990, São Paulo. **Anais...** São Paulo: IME – USP, 1990.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Anuário estatístico do Brasil. 2002. Disponível em: <<http://www.brasilrepublica.com/nordeste.htm>>. Acesso: 30 abr. 2010.
- HARTIGAN, J. A. **Clustering algorithms**. New York: Wiley, 1975. p. 84–112.
- HARTIGAN, J. A.; WONG, M. A. A K-means clustering algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, v. 28, p. 100–108, 1979.
- KAGAN, A. M.; LINNIK, YU. V.; RAO, C. R. **Characterization problems in mathematical statistics**. New York: Wiley, 1973. p. 408–410.
- KAWACHI, T.; MARUYAMA, T.; SINGH, V. P. Rainfall entropy for delineation of water resources zones in Japan. **Journal of Hydrology**, n. 246, p. 36–44, 2001.
- MARUYAMA, T.; KAWACHI, T. Evaluation of rainfall characteristics using entropy. **Journal of Rainwater Catchment systems**, v. 4, n. 1, p. 7–10, 1998.
- SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, p. 379–423; 623–656, 1948.
- SILVA, V. P. R.; CAVALCANTI, E. P.; NASCIMENTO, M. G.; CAMPOS, J. H. B. C. Análises da precipitação pluvial no Estado da Paraíba com base na teoria da entropia. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, PB, v. 7, n. 2, p. 269-274, 2003. Disponível em <<http://www.agriambi.com.br>>. Acesso: 30 abr. 2010.
- SINGH, V. P. The use of entropy in hydrology and water resources. **Hydrological Processes**, n. 11, p. 587–626, 1997.
- VOSE, R. S.; SCHMOYER, R. L.; STEURER, P. M.; PETERSON, T. C.; HEIM, R.; KARL, T. R.; EISCHEID, J. K. **The global historical climatology network: long-term monthly temperature, precipitation, sea level pressure, and station pressure**. Oak Ridge: Oak Ridge National Laboratory, 1992. Disponível em: <<http://cdiac.esd.ornl.gov/ghcn/ghcn.html>>. Acesso: 30 abr. 2010.