

Análise de zonas homogêneas em séries temporais de precipitação no Estado da Bahia

Camila da Silva Dourado ^(1*); Stanley Robson de Medeiros Oliveira ^(1,2); Ana Maria Heuminski de Avila ⁽³⁾

⁽¹⁾ Universidade de Campinas (UNICAMP), Faculdade de Engenharia Agrícola, Av. Cândido Rondon, 501, 13083-875 Campinas (SP), Brasil.

⁽²⁾ Embrapa Informática Agropecuária, Av. André Tosello, 209, 13083-886 Campinas (SP), Brasil.

⁽³⁾ UNICAMP, Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura, Cidade Universitária Zeferino Vaz, s/n, 13083-970 Campinas (SP), Brasil.

^(*) Autora correspondente: camila.dourado@feagri.unicamp.br

Recebido: 23/abr./2013; Aceito: 5/jun./2013

Resumo

O objetivo deste trabalho foi identificar zonas pluviometricamente homogêneas no Estado da Bahia e analisar as condições climáticas de cada zona entre 1981 e 2010. Foi aplicada a técnica de mineração de dados, *Clusterização* (agrupamento de dados), por meio do uso do algoritmo *k-means*, para transformação das séries históricas de precipitação em cinco zonas pluviometricamente homogêneas, em resposta à orografia, maritimidade e sistemas meteorológicos atuantes na região. Foram utilizados dados de médias mensais de precipitação de 92 estações meteorológicas. Os resultados apontam que as zonas mais secas estão situadas na parte central, de norte a sul do estado, principalmente ao norte com os menores volumes anuais, em torno de 480 mm. A zona localizada ao norte do estado é contrastante com a faixa litorânea, em que são observados os maiores volumes anuais de precipitação (1.380 mm aproximadamente). A alta variabilidade pluviométrica ocorre em quase todas as zonas, principalmente em duas do semiárido com coeficientes de variação (CV) iguais a 42 e 28%. Diferencia-se dessa característica a zona pertencente à faixa litorânea, que apresenta regularidade de chuvas durante todo o ano e CV de 15%. As estações chuvosas e secas estão bem definidas. Os valores de precipitação da estação chuvosa representam em torno de 81% dos totais anuais, com destaque para as zonas situadas no centro-oeste e oeste do estado, com 95 e 96% dos totais anuais.

Palavras-chave: mineração de dados, clusterização, variabilidade pluviométrica.

Analysis of rainfall homogeneous areas in time series of precipitation in the State of Bahia, Brazil

Abstract

The aim of this study was to identify rainfall homogeneous areas in the State of Bahia, Brazil and analyze the climatic conditions of each area for the period between 1981 and 2010. It was applied a data mining technique, clustering (grouping of data), by using the *k-means* algorithm for transforming time series of precipitation in five rainfall homogeneous areas, in response to topography, maritime dimension, and weather systems operating in the region of study. Data of average monthly rainfall of 92 meteorological stations were used. The results indicate that the driest areas are situated in the central part of the state, from north to south, mainly in the north with the lowest annual volumes, around 480 mm. The area located in the north of the state contrasts with that one located on the coast, where the largest volumes of annual rainfall of the study were observed (approximately 1.380 mm). The high rainfall variability occurs in almost all areas, especially in two of those of semiarid ones with Coefficients of Variation (CV) reaching 42 and 28%. This characteristic differs from the area belonging to the coastal area, which presents regular rainfall during all the year and a CV of 15%. The rainy and dry seasons are well defined. Precipitation values of the rainy season accounts for about 81% of the annual total, with emphasis on the zones located in the central-west and west of the state with 95 and 96% of the annual total.

Key words: data mining, clustering, rainfall variability.

1. INTRODUÇÃO

As condições climáticas, suas variações e flutuações, têm sido motivos do aumento de pesquisas científicas que visam à quantificação e o monitoramento mais preciso de variáveis ambientais. As variações da distribuição

espaço-temporal das chuvas, associadas aos baixos totais anuais pluviométricos, impactam diretamente a agricultura de sequeiro. O conhecimento dessa variável agrometeorológica pode orientar a tomada de decisão sobre as medidas necessárias para minimizar danos (ROMANI et al., 2010).

Em particular, o Estado da Bahia, caracterizado por uma diversidade de domínios climáticos, apresenta um quadro de alta variabilidade pluviométrica, decorrente da conjugação de diferentes sistemas meteorológicos que atuam nessa região, tais como, Zona de Convergência Intertropical (ZCIT), Vórtices Ciclônicos (VCAN), Sistemas Frontais (FPA), Zona de Convergência do Atlântico Sul (ZCAS), brisas marítimas/terrestres e ventos vale/montanha, e um relevo constituído por planícies, vales, serras e montanhas (MOLION e BERNARDO, 2002; CAVALCANTI et al., 2009).

O estado apresenta cinco tipos climáticos distribuídos nos sentidos Norte-Sul e Leste-Oeste, conforme a classificação de THORNTHWAITE e MATHER (1955), que dão origem às zonas árida, semiárida, subúmida a seca, subúmida a úmida e úmida. Os totais médios anuais da precipitação pluvial variam de 300 a 2.000 mm, podendo ainda ultrapassar valores de 2.000 mm em algumas regiões de clima úmido (KOUSKY, 1979).

A distribuição pluvial decrescente dos totais médios anuais, das extremidades Leste (Planície Litorânea) e Oeste (Chapadões Ocidentais) para o interior semiárido, confirma a ação periférica dos sistemas dinâmicos. A Chapada Diamantina representa com maior expressividade o efeito orográfico, proporcionando um clima local no semiárido com chuvas anuais acima de 1.000 mm (TANAJURA et al., 2010). Tais características possibilitam a existência de regimes pluviais diferentes sobre o estado.

A pluralidade de regimes pluviais vinculada ao aumento da variabilidade interanual da precipitação (SILVA et al., 2012) mostra a necessidade de sistemas que consigam analisar, com confiabilidade, informações sobre a climatologia do estado.

Desta forma, a aplicação de técnicas de mineração, etapa principal do processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases – KDD*) surge como uma alternativa promissora para transformar grandes volumes de dados climáticos em conhecimento novo e útil (FAYYAD et al., 1996; ; HAN et al., 2011).

A escolha de uma técnica de mineração exige análise detalhada do problema em questão e a decisão sobre o tipo de padrão que será encontrado nos dados. Em particular, a *Clusterização* é uma técnica capaz de transformar séries históricas de dados pluviais em zonas homogêneas pluviométricas. Cada posto meteorológico corresponde a uma série temporal de dados de uma cidade, enquanto uma região homogênea é formada pelo agrupamento das séries temporais mais similares.

Muitos trabalhos nas áreas de agrometeorologia e climatologia enfatizam a aplicação de *clusterização* para a definição de regiões climáticas homogêneas em várias partes do Brasil. Abordado nos trabalhos de ANDRÉ et al. (2008) para o Rio de Janeiro, FECHINE e GALVINCIO (2008) para Pernambuco e BOSCHI, OLIVEIRA e ASSAD (2011) para o Rio Grande do Sul.

Neste contexto, o objetivo deste trabalho foi identificar zonas pluviometricamente homogêneas no estado da Bahia, e analisar as condições climáticas de cada zona para o período de 1981 a 2010.

2. MATERIAL E MÉTODOS

A área de estudo corresponde ao estado da Bahia, com limítrofe ao Norte o rio São Francisco, no município de Curaçá, divisa com Pernambuco, entre as coordenadas 8°32'00"S e 39°22'49"W. Ao sul, o limite extremo é a Barra do Riacho Doce, no município de Mucuri, divisa com o Espírito Santo, 18°20'07"S e 39°39'48"W. A Leste, o ponto extremo é a Barra do Rio Real, no município de Jandaíra, divisa com o Oceano Atlântico, 11°27'07"S e 37°20'37"W. O ponto extremo do Oeste é o divisor de águas, no município de Formosa do Rio Preto, divisa com o Tocantins, 11°17'21"S e 46°36'54"W.

A metodologia aplicada para execução do processo de KDD, teve como suporte as etapas do modelo CRISP-DM (*Cross Industry Standard Process for Data Mining*) descrita por CHAPMAN et al. (2000). Esta metodologia é amplamente adotada em projetos de mineração de dados para definição dos procedimentos (BOSCHI et al., 2011; BHAKTIKUL et al., 2012). Neste trabalho, foram aplicadas quatro etapas do modelo CRISP-DM: entendimento dos dados, preparação dos dados, modelagem e avaliação.

Na primeira etapa (entendimento dos dados), foram utilizadas séries históricas de precipitação pluviométrica diária, adquiridas no sistema de informações hidrológicas Hidroweb da Agência Nacional de Água (<http://hidroweb.ana.gov.br/>) (ANA, 2011). Optou-se por trabalhar com um período amostral de 30 anos, segundo padronização da Organização Meteorológica Mundial, para caracterizar o clima de uma região. Sendo assim, foram selecionadas 92 estações meteorológicas para os anos de 1981 a 2010, cobrindo praticamente todo o estado.

Foram utilizados dados consistidos, uma vez que dados brutos provenientes de estações meteorológicas apresentam problemas, a saber, erros de leitura, transcrição, digitação e ausência de dados em alguns períodos. A consistência dos dados adotada pela ANA é baseada no modelo matemático proposto pela Agência Nacional de Energia Elétrica - ANEEL (BRASIL, 1984).

Para implantação dos dados espaciais, como por exemplo confecção dos mapas de localização das estações meteorológicas (Figura 1), distribuição pluviométrica anual e identificação das zonas homogêneas, foi adotado o sistema de coordenadas Latitude e Longitude, *Datum* SAD69, e sistema de projeção *Universal Transversa de Mercator*. As informações geoespaciais foram fornecidas pela ANA (2011) e o *shape*, referente aos limites municipais do Estado da Bahia, pelo IBGE (2012).

Para confecção do mapa pluviométrico da Bahia, foi aplicado o método geoestático IDW (*Inverse Distance Weighted* – Ponderação do Inverso da Distância), que, de acordo com MARCUZZO, ANDRADE e MELO, (2011), é uma das técnicas de interpolação mais usadas para pontos espalhados espacialmente. Para isto, foram utilizadas as médias anuais de precipitação pluviométrica referente ao período de 1981 a 2010.

Para elaboração do cálculo da interpolação do valor de um ponto com base no método do IDW, foi utilizada a seguinte equação matemática MARCUZZO, ANDRADE e MELO (2011):

$$Z(X) = \frac{\sum_{i=1}^n \omega_i Z(X_i)}{\sum_{i=1}^n \omega_i} \quad (1)$$

Em que, $Z(X)$ é o valor do ponto que se deseja interpolar; n é a quantidade de pontos próximos utilizados na interpolação do ponto X ; $Z(X_i)$ é o valor do ponto X_i ; e ω_i é o peso do valor de X_i sobre o ponto X . Para se determinar ω_i , foi utilizada a seguinte equação:

$$\omega_i = \frac{1}{h(X, X_i)^p} \quad (2)$$

Sendo, $h(X, X_i)$ a distância entre o ponto X e o ponto X_i ; e p o parâmetro de potência, igual a dois.

A fase de preparação dos dados teve início com a construção de uma planilha composta por 92 linhas (registros), condizentes às estações meteorológicas, e 10.957 colunas (atributos),

relacionadas aos valores diários de precipitação. A disposição dos dias nas colunas e das estações nas linhas foi organizada com o intuito de subsidiar a geração de agrupamentos das estações meteorológicas conforme a similaridade de padrões pluviiais.

Com dados em escala diária, foi construído outro banco de dados com valores de precipitação acumulados em escala mensal. Essa transformação foi necessária devido à particularidade de alta variabilidade da variável estudada, visto que a técnica de agrupamento busca padrões de similaridade entre os objetos para formação dos *clusters*.

Uma etapa essencial na aplicação das técnicas de clusterização, quando se utiliza atributos com diferentes ordens de grandeza, é a normalização dos dados. No entanto, para o presente trabalho esta etapa não foi necessária, uma vez que o único atributo considerado foi a precipitação pluviométrica mensal, em que todos os acumulados mensais têm a mesma distribuição de dados.

Na etapa da modelagem, a técnica escolhida para seleção das zonas homogêneas foi a análise de *clusters* (agrupamento de dados), com o método de particionamento. Esta técnica consiste em agrupar objetos (itens de dados) que contenham características similares, ou seja, particionar uma dada população de objetos ou itens em conjuntos similares (KUMAR et al., 2009). A similaridade entre objetos (séries temporais) é medida pela distância Euclideana entre dois objetos (HAN et al., 2011). Quanto menor for a distância entre dois objetos, maior será a similaridade entre eles. Para análise de *clusters*, foi adotado o método de particionamento utilizando o algoritmo *k-means*. O programa computacional utilizado nesta fase foi o WEKA versão 3.6.3. (WITTEN et al., 2011).

No que se refere à metodologia de particionamento, de acordo com REZENDE et al. (2005), os objetos são alocados nos *clusters* conforme a medida de similaridade adotada, onde se busca a maximização da distância entre *clusters* e a minimização da distância entre os objetos do mesmo grupo. O agrupamento ocorre com a especificação de uma medida de distância de acordo com o algoritmo, conforme a representatividade dos dados, ou quantitativa ou qualitativa.

Foram testados valores para k (n.º de grupos) = 2, 3, 4 e 5 para definição do número de zonas homogêneas mais condizente com a realidade dos dados. A escolha desse intervalo deu-se com base em trabalhos já realizados sobre a distribuição da precipitação pluvial da região e comparação com o mapa da pluviometria anual (MOLION e BERNARDO, 2002; CAVALCANTI et al., 2009).

A última etapa aplicada do modelo CRISP-DM foi a avaliação. Nesta etapa foram discutidos os resultados obtidos, com a definição do número de *clusters* e identificação dos mesmos. Para uma melhor visualização dos resultados, os *clusters* foram espacializados, considerando-se a latitude e longitude de suas respectivas estações. Nesta etapa do processo, a participação de especialistas do domínio foi fundamental para a escolha do melhor número de zonas homogêneas.

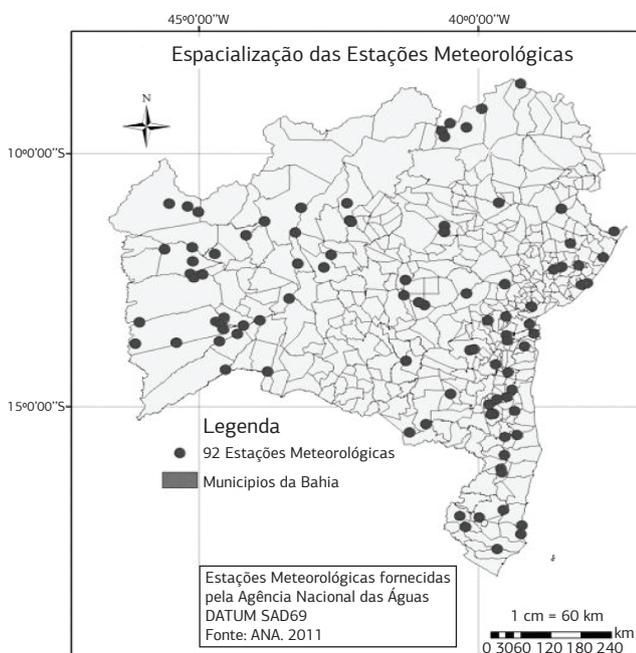


Figura 1. Espacialização das estações meteorológicas do Estado da Bahia.

3. RESULTADOS E DISCUSSÃO

A representação geoespacial da distribuição pluviométrica anual no Estado da Bahia, por meio do método geostatístico IDW, permitiu uma melhor visualização da disparidade do volume pluvial existente no território baiano (Figura 2).

A aplicação da técnica de agrupamento permitiu a identificação das diferentes regiões pluviométricas do estado da Bahia, de acordo com a similaridade do volume pluvial, calculada por meio da distância Euclidiana entre cada par de séries temporais (Figura 3). Para esta etapa do trabalho, foi adotado como padrão do estudo a escala mensal do banco de dados.

Optou-se pela divisão de cinco zonas pluviometricamente homogêneas, uma vez que este número está em consonância com pareceres de especialistas da área e registros na literatura referentes à distribuição, a orografia, a maritimidade e aos sistemas pluviométricos que atingem o estado (CAVALCANTI et al., 2009; SILVA et al., 2012).

Os resultados mostram que o algoritmo *k-means* gerou os *clusters* de forma coerente com a distribuição pluviométrica espaço temporal do estado da Bahia. Resultados semelhantes foram encontrados ao comparar com o mapa pluviométrico do estado referente ao período de 1981 a 2010, obtido por meio da técnica de interpolação IDW (Figura 2).

Métodos de agrupamentos foram utilizados em vários trabalhos, em diferentes partes do Brasil, tais como: Alagoas (SOUZA et al., 1992), Pernambuco (FECHINE et al., 2008), Rio de Janeiro (ANDRÉ et al., 2008) e Rio Grande do Sul (BOSCHI et al., 2011).

De forma semelhante aos trabalhos de ANDRÉ et al. (2008), que utilizaram o *k-means* em dados meteorológicos do estado do Rio de Janeiro para identificar regiões pluviometricamente homogêneas, sob forte influência dos fatores maritimidade e orografia, a delimitação das zonas homogêneas no estado da Bahia também apresentou fortes influências desses fatores climáticos.

Observa-se que a disposição dos *clusters* no território baiano evidencia a distribuição decrescente nos volumes de chuva, das extremidades Leste (Planície Litorânea) e Oeste (Chapadões Ocidentais) para o interior do estado da Bahia. Desta forma, confirma a ação periférica dos sistemas dinâmicos ZCIT, FPA, ZCAS e VCAN, que possibilitam a existência de regimes pluviais diferentes e superpostos sobre o território baiano, verificados da mesma forma por CAVALCANTI et al. (2009).

O *cluster 0*, representado por seis estações meteorológicas, situa-se ao Norte do estado. Os municípios deste grupo pertencem à região do semiárido nordestino, submetidos aos fortes rigores do clima árido, acometidos por longos períodos de estiagem, elevadas temperaturas, altas taxas de evapotranspiração e alta disponibilidade energética de luz solar (TEIXEIRA et al., 2010). Sobre este *cluster*, atuam a Zona

de Convergência Inter Tropical e os Vórtices Ciclônicos (MOLION e BERNARDO, 2002).

O *cluster 1*, representado por 21 estações meteorológicas, abrange a região mais setentrional, expandindo-se por toda parte central do Norte ao Sul do estado. Ocupa a maior área do território baiano, onde estão localizados os municípios do semiárido pertencentes ao Polígono das Secas. São responsáveis pela distribuição pluviométrica deste grupo os VCAN, ao Norte, e a FPA nas demais áreas.

O *cluster 2*, composto por 21 estações meteorológicas, delimita a região de transição do semiárido para o cerrado do Oeste baiano. Os municípios pertencentes a este grupo não são contíguos geograficamente como nos demais *clusters*, isto ocorre devido à expressividade do

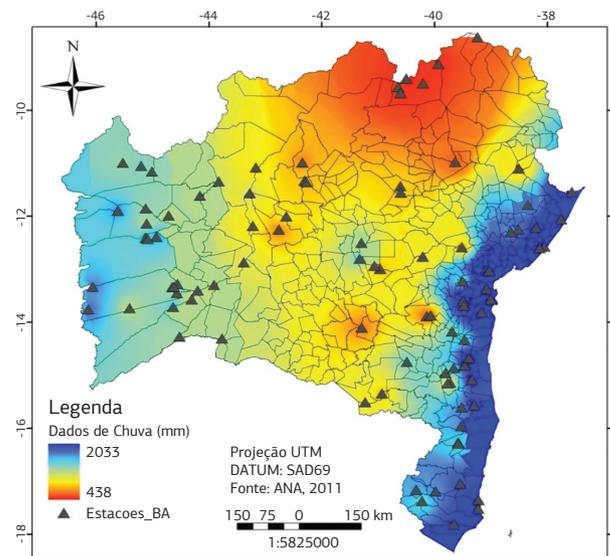


Figura 2. Mapa pluviométrico do Estado da Bahia - 1981 a 2010.

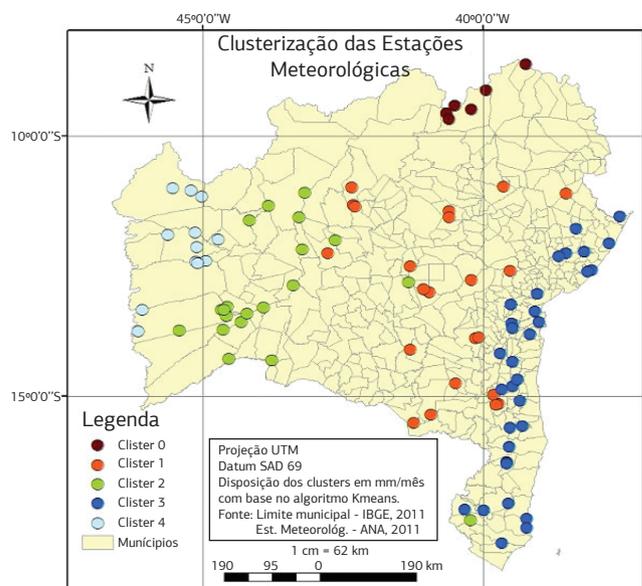


Figura 3. Agrupamento das zonas pluviometricamente homogêneas da Bahia.

efeito orográfico da Chapada Diamantina, formando um micro clima no interior da região semiárida. Outro fator é a presença dos mecanismos de ação da FPA filiados ao regime climático da região Sudeste do país, que influenciam o volume pluviométrico do Sudeste e Sul do estado (CAVALCANTI et al., 2009).

O *cluster 3*, composto por 32 estações meteorológicas, é representado pelos municípios da costa litorânea do estado, apresentando grande volume de chuva e relativa regularidade pluvial durante todo o ano. Sobre este *cluster* atuam os sistemas meteorológicos de FPA e as brisas marítimas. Representa a zona com maior índice pluviométrico, tal característica atribui a esta região a condição de baixo risco de seca, segundo os estudos de Risco de Seca do estado da Bahia realizado pela SEI (2000).

Por fim, o *cluster 4*, representado por 12 estações meteorológicas, corresponde à parte ocidental do Oeste da Bahia, com períodos chuvosos e secos bem definidos ao longo do ano. É a região do cerrado baiano cuja concentração das atividades agropecuárias proporciona aos municípios desse grupo o maior PIB agrícola do Estado da Bahia. Possui o segundo maior volume pluviométrico e, sobre este grupo, atuam os Sistemas Frontais e a Zona de Convergência do Atlântico Sul.

Houve alta heterogeneidade na distribuição espacial das chuvas no estado, com precipitação média variando de 477 mm (*cluster 0*) a 1.377mm (*cluster 3*) entre zonas pluviométricas (Tabela 1, Figura 4). Destaca-se uma diferença de 900 mm de chuva na média anual entre os *clusters 0* e 3. A disposição espacial dos valores da variação pluviométrica corrobora com os estudos da SEI (2000) sobre a distribuição pluviométrica, apresentando maiores volumes de

chuva nas extremidades em direção ao interior do estado. Os *clusters 0* e 1 apresentam os menores volumes pluviométricos e os maiores coeficientes de variação 42% e 28%, respectivamente (Tabela 1), o que evidencia as características climáticas das regiões semiáridas, vistas também por SILVA, PEREIRA e ALMEIDA (2012), ao estudar a variabilidade pluviométrica no semiárido do Nordeste brasileiro.

Já o *cluster 3* apresenta o menor coeficiente de variação (15%), representando a região de maior altura e regularidade pluviométrica durante toda a série histórica. Observação semelhante encontra-se em TANAJURA, GENZ e ARAÚJO (2010). Nos *clusters 2* e 4, o coeficiente de variação corresponde a 24 e 21%, respectivamente, indicando uma baixa variação no volume pluviométrico. O *cluster 2* corresponde à região mais próxima do semiárido, o que representa comportamento de variabilidade mais elevado em relação ao *cluster 4*, que faz parte de uma região de cerrado e o segundo maior em níveis pluviométricos anuais.

A figura 5 mostra as distribuições mensais da precipitação média de cada *cluster* determinada pelo algoritmo *k-means*. Desta forma, é possível comparar as regiões avaliando onde há maior ou menor variabilidade da média entre os *clusters*.

Observa-se que, durante a estação chuvosa, para os *clusters 0*, 1, 2 e 4 há uma considerável dispersão entre os valores mensais. Entretanto, no período seco, apenas o *cluster 1* apresenta variações consideráveis em relação aos demais. Verifica-se que a estação chuvosa do *cluster 3* é composta por outros meses em relação aos demais *clusters*.

A estação chuvosa do *cluster 0* compreende aos meses de dezembro a abril, com volume máximo no mês de março (Figura 5). Isso ocorre devido à influência da ZCIT, que atinge a região nesse período, observado também por KOUSKY (1979) e CAVALCANTI (2009). A precipitação na estação chuvosa deste *cluster* é 90% do total anual, caracterizando uma região de estações bem definidas ao longo do ano (Tabela 2). Compartilha desta mesma característica os *clusters 2* e 4, que apresentam no período chuvoso um total da precipitação anual de 95 e 96%, respectivamente (Tabela 2). Apresentam semelhança entre os meses que compreendem o período chuvoso, outubro a abril, com volumes máximos nos meses de novembro a dezembro (Figura 5). O *cluster 4* representa o grupo com o maior volume pluvial no período chuvoso em relação aos demais, alcançando um total superior a 1.000 mm.

Observa-se que os volumes pluviais do período chuvoso do *cluster 1* correspondem a 81% da precipitação total. Os sistemas meteorológicos FPA, ZCAS e VCANs são responsáveis pela produção de chuvas nessa região e contribuem para ocorrência de chuvas entre os meses de novembro a abril, com máximo em dezembro.

A precipitação dos períodos chuvosos e secos do *cluster 3* é bem elevada, com destaque para volumes acima de 459 mm na estação seca, revelando regularidade na distribuição pluviométrica durante todo o ano. Apresenta 66% do montante pluviométrico no período chuvoso em relação ao total das chuvas anuais (Tabela 2). Os meses que integram a estação

Tabela 1. Valores acumulados anuais de precipitações, médias, desvio-padrão, coeficiente de variação, máximo e mínimo para o período de 1981 a 2010

Cluster	Média	DP	CV (%)	Mediana	Máximo	Mínimo
0	477	201	42	465	1148	160
1	672	189	28	664	1147	323
2	873	210	24	886	1337	512
3	1377	209	15	1378	1861	900
4	1059	219	21	1034	1531	695

DP: desvio-padrão; CV: coeficiente de variação.

Tabela 2. Contribuição da precipitação nos períodos chuvosos e secos para precipitação anual

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Prec. Anual (mm)	477	672	873	1377	1059
Prec. Chuvoso (mm)	432	544	828	918	1014
Prec. Seco (mm)	45	127	45	459	45
% Per. Chuvoso	90	81	95	66	96
% Per. Seco	10	19	5	34	4

chuvosa desta região são março a julho, novembro e dezembro, com volumes máximos nos meses de dezembro e abril.

Considerando a contribuição da precipitação para as estações chuvosa e seca em cada *cluster* (Tabela 2), verifica-se que o Estado da Bahia apresenta estações bem

definidas e que o período chuvoso é responsável por grande parte da precipitação anual. Com destaque para os *clusters* 0, 2 e 4 com precipitação média do período chuvoso acima de 800 mm (*clusters* 2 e 4); 432 mm (*cluster* 0) para estação chuvosa e com 45 mm para estação seca (*clusters* 0, 2 e 4).

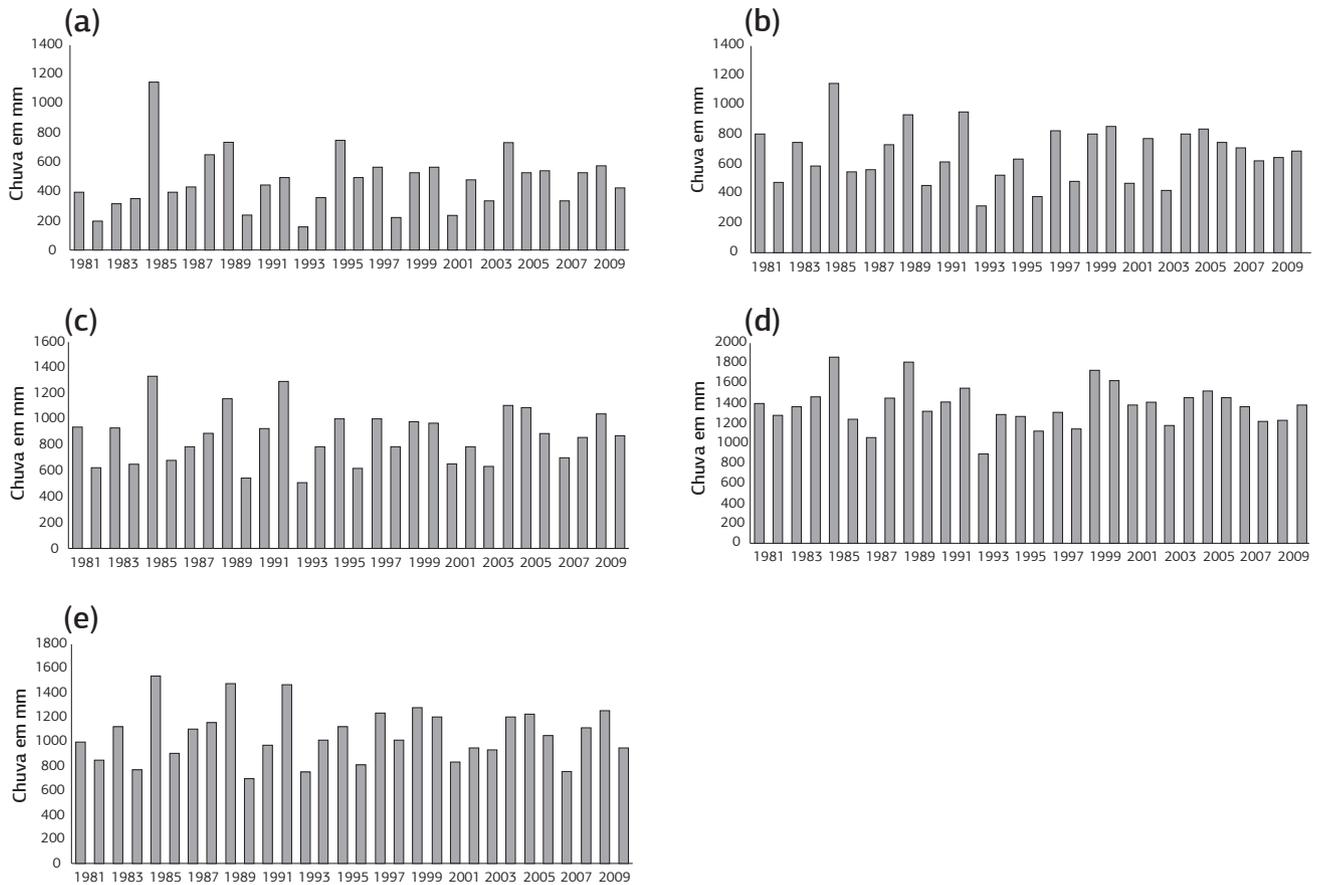


Figura 4. (a-e) Distribuição da precipitação pluviométrica nos cinco *clusters* (0, 1, 2, 3 e 4), para o período de 1981 a 2010 do Estado da Bahia.

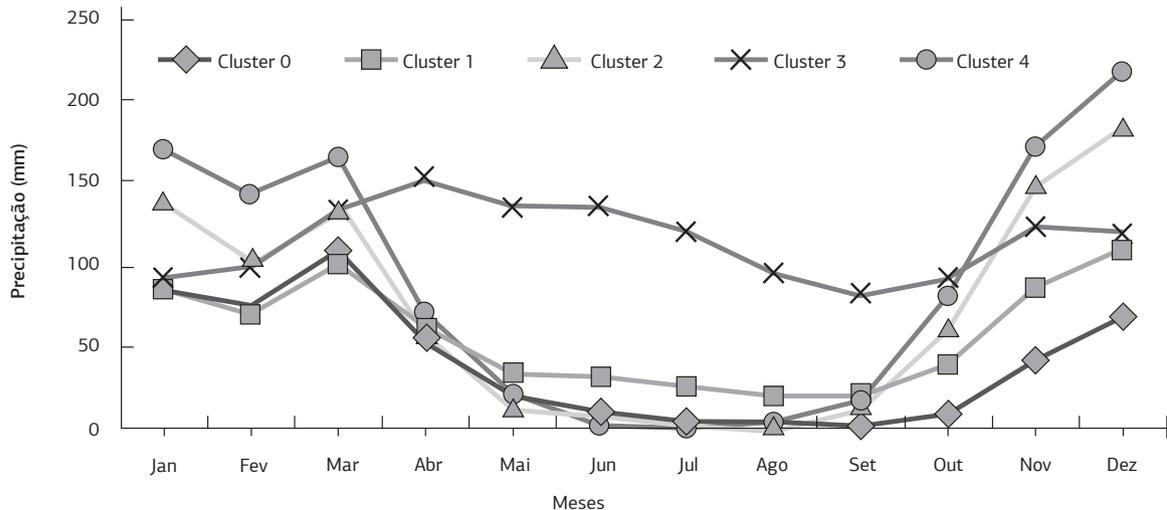


Figura 5. Precipitação média mensal em todos os *clusters* (0, 1, 2, 3 e 4) do Estado da Bahia, referente ao período de 1981 a 2010.

4. CONCLUSÃO

Foram identificadas cinco zonas pluviometricamente homogêneas no Estado da Bahia, com os dados de precipitação correspondente ao período de 1981 a 2010. A técnica de mineração de dados, análise de *cluster* (agrupamento de dados), delimitou com coerência as regiões pluviométricas do estado, mostrando-se eficiente para esta aplicação. O *cluster*, situado no centro-oeste do estado, difere dos demais quanto à disposição geográfica dos seus municípios. Integra a este *cluster* municípios do centro do estado, região semiárida, o que pode estar associado à interferência da orografia da região da Chapada Diamantina. Os valores de acumulados anuais da precipitação apresentaram alta variabilidade na distribuição pluviométrica em todos os *clusters*. A maior variação foi registrada nos *clusters* que correspondem ao semiárido baiano, que apresenta como característica climática alta variabilidade pluviométrica. A menor variação foi registrada na faixa litorânea, que apresenta como característica a regularidade e altos volumes pluviométricos durante todo ano. Em todo estado, as estações secas e chuvosas são bem definidas, sendo os volumes pluviométricos do período chuvoso responsáveis por grande parte dos totais anuais de precipitação.

REFERÊNCIAS

- ANA - Agência Nacional das Águas. Disponível em: <<http://hidroweb.ana.gov.br>>. Acesso em: ago. 2011.
- ANDRÉ, R.G.B.; MARQUES, V.S.; PINHEIRO, F.M.A.; FERRAUDO, A.S. Identificação de regiões pluviometricamente homogêneas no estado do Rio de Janeiro, utilizando valores mensais. *Revista Brasileira de Meteorologia*, v.23, p.501-509, 2008. DOI: 10.1590/S0102-77862008000400009
- BHAKTIKUL, K.; ANUJIT, R.; TOIM, J. Estimation of crop coefficient of corn ($K_{c,com}$) under climate change scenarios using data mining technique. *Environment Asia*, v.5, p.56-62, 2012.
- BOSCHI, R.S.; OLIVEIRA, S.R.M.; ASSAD, E.D. Técnicas de mineração de dados para análise pluvial decenal do Rio Grande do Sul. *Engenharia Agrícola*, v.31, p.1189-1201, 2011. DOI: 10.1590/S0100-69162011000600016
- BRASIL. Departamento Nacional de Águas e Energia Elétrica – DNAEE. Divisão de Controle de Recursos Hídricos. Sistemática para análise de consistência e homogeneização de dados pluviométricos, Brasília, 1984.
- CAVALCANTI, I.E.A.; FERREIRA, N.J.; DIAS, M.A.F.; JUSTI, M.G.A. Terra e clima no Brasil. São Paulo: Editora Oficina de Textos, p.464, 2009.
- CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. CRISP-DM 1.0: step-by-step data mining guide. Illinois: SPSS, p.78, 2000.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery: an overview. In: *Advances in knowledge discovery and data mining*, eds. U. M. Fayyad et al., AAAI/MIT Press. Menlo Park: American Association for Artificial Intelligence, Calif, pp. 1-34, 1996.
- FECHINE, J.A.L.; GALVINCIO, J.D. Agrupamento da precipitação mensal da bacia hidrográfica do rio Brígida-PE, através da multivariada. *Revista Brasileira de Geografia Física*, v.1, p.39-46, 2008.
- HAN, J.; KAMBER, M. Data mining: concepts and techniques. San Francisco: Morgan Kaufmann Publishers, p.770, 2011.
- IBGE- Instituto Brasileiro de Geografia e Estatística. Clima da Bahia. Disponível em: <www.ibge.gov.br>. Acesso em: jul. 2012.
- KOUSKY, V.E. Frontal influences on northeast Brazil. São Paulo: INPE, p.16, 1979. DOI: 10.1175/1520-0493(1979)107<1140:FIONB>2.0.CO;2
- KUMAR, D.N.; DHANYA, M.C.T. Data mining and its applications for modeling rainfall extremes. *Journal of Hydraulic Engineering*, v.15, p.25-51, 2009. DOI: 10.1080/09715010.2009.10514967
- MARCUZZO, F.F.N.; ANDRADE, L.R.; MELO, D.C.R. Métodos de interpolação matemática no mapeamento de chuvas no estado do Mato Grosso. *Revista Brasileira de Geografia Física*, v.4, p.793-804, 2011.
- MOLION, L.C.B.; BERNARDO, S.O. Uma revisão da dinâmica das chuvas no nordeste brasileiro. *Revista Brasileira de Meteorologia*, v.17, p.1-10, 2002.
- REZENDE, S.O.; PUGLIESI, J. B., MELANDA, E.A.; PAULA, M.F. Mineração de dados. In: *Sistemas inteligentes. Fundamentos e aplicações*. Barueri: Manole Ltda., p.527, 2005.
- ROMANI, L.A.S.; ÁVILA, A.M.H.; ZULLO, J.J.; TRAINA, C.J.; TRAINA, A.J.M. Mining relevant and extreme patterns on climate time series with CLIPSMiner. *Journal of Information and Data Management*, v.1, p.245-260, 2010.
- SILVA, V.P.R.; PEREIRA, E.R.R.; ALMEIDA, R.S.R. Estudo da variabilidade anual e intra-anual da precipitação na região nordeste do Brasil. *Revista Brasileira de Meteorologia*, v.27, p.163-172, 2012. DOI: 10.1590/S0102-77862012000200005
- SOUZA, J.L., AMORIM, R.E.C.; CARVALHO, S.M.R.; PEREIRA, J.O.; CURTI, P.R.C. Agrupamento de estações pluviométricas do estado de Alagoas, utilizando-se análise multivariada. *Revista Brasileira de Meteorologia*, v.7, p.603-612, 1992.
- SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA (SEI). Sistemas de dados estatísticos. 2000. Disponível em: <<http://www.sei.ba.gov.br/side/consulta>>. Acesso em: maio 2011.
- TANAJURA, C.A.S.; GENZ, F.; ARAÚJO, H.A. Mudanças climáticas e recursos hídricos na Bahia: validação da simulação do clima presente do HADRM3P e comparação com os cenários A2 e B2 para 2070 – 2100. *Revista Brasileira de Meteorologia*, v.25, p.345-358, 2010. DOI: 10.1590/S0102-77862010000300006
- TEIXEIRA, A.H.C.; MOURA, M.S.B; ANGELOTTI, F. Cultivo da mangueira. EMBRAPA SEMIÁRIDO – Sistema de Produção, v.2, 2010.
- THORNTHWAITE, C.W.; MATHER, J.R. The water balance. *Climatology*, v.8, p.1-40, 1955. DOI: 10.1097/00010694-195904000-00024
- WITTEN, I.H.; FRANK, E.; HALL, M.A. Data mining: practical machine learning tools and techniques. 3rd ed. San Francisco: Morgan Kaufmann, p.629, 2011. DOI: 10.1145/507338.507355