Revista Brasileira de Climatologia
ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

USO DOS ÍNDICES PCP E PCD NA DETERMINAÇÃO DE REGIÕES COM PRECIPITAÇÃO PLUVIAL HOMOGÊNEA

ARAÚJO, Jurema Maria Silva – jurema.araujo@hotmail.com
Engenharia Sanitária – Universidade Federal do Rio Grande do Norte

MAIA, Adelena Gonçalves – adelenam@gmail.com
Doutora – Universidade Federal do Rio Grande do Norte

NEVES, Josemir Araújo – josemir.neves@gmail.com
Doutor – Empresa de Pesquisa Agropecuária do Rio Grande do Norte.

RESUMO: Os resultados dos estudos de delimitação de áreas homogêneas de precipitação têm sido utilizados no planejamento das atividades econômicas, possibilitando o uso mais eficiente e racional dos recursos hídricos e também em regiões com dados de precipitação escassos. O objetivo desta pesquisa é determinar áreas homogêneas em termos do regime de precipitação no estado do Rio Grande do Norte. Foram utilizadas duas técnicas de agrupamento: a hierárquica de Ward e a não hierárquica de k-means. Verificou-se que a precipitação anual média e os índices PCP e PCD foi o conjunto de variáveis mais adequado para determinar os quatro grupos homogêneos. O grupo identificado na região leste do estado é caracterizado como a região de maior precipitação anual, melhor distribuição desta precipitação e concentração de mesma no mês de maio. Se dirigindo para o oeste do estado foi identificada uma região com período de maior precipitação adiantado para o início de maio, menor precipitação anual e uma maior concentração da precipitação. Na região central do estado o grupo identificado é caracterizado por apresentar uma elevada concentração da precipitação e um adiantamento das chuvas para o período do mês de abril, além de ser a região com menor precipitação anual. Na região oeste do estado as chuvas são adiantadas para o final de março, a região apresenta a pior distribuição da precipitação no estado e as precipitações anuais nesta região são mais elevadas do que na região central.

Palavras-chave: Análise de agrupamento; Áreas homogêneas; Índices de variabilidade.

THE USE OF PCP AND PCD INDICES TO DETERMINING PRECIPITATION HOMOGENEOUS AREAS

ABSTRACT: The results of studies of delimitation of homogeneous areas of precipitation have been used in the planning of economic activities, enabling more efficient and rational use of water resources and also in regions with scarce rainfall data. The objective of the research is to determine homogeneous areas in terms of the precipitation regime of Rio Grande do Norte State (Brazil), from the cluster analysis and techniques: hierarchical (Ward) and non-hierarchical (k-means). It was found that the average annual precipitation and PCP and PCD indices were the most appropriate set of variables for determining the four homogeneous groups. The hierarchical and non-hierarchical techniques showed similar results, with only 5 seasons differ. The four homogenous groups identified have the following characteristics. The group identified in the eastern region of the state is characterized as the region of highest annual rainfall, better distribution of this precipitation and concentration of the precipitation in the month of May. In the region more the western state was identified a region that has the rainy season advance to the beginning of May, the lower annual rainfall and a higher concentration of precipitation. The central region of the state is characterized by having a high concentration of precipitation and an advance of rainfall for the period of April, in addition it is the region with the lowest annual rainfall. In the western state region the rains are concentrated in late March, the region has the worst distribution of rainfall in the state and the annual precipitation in this region are higher than in the central region.

Keywords: Cluster analysis; Homogeneous areas; Variability indices.

1. INTRODUÇÃO

Áreas com precipitação pluvial homogênea são regiões com características similares em termos de algumas variáveis pluviais. Os resultados dos estudos de delimitação de áreas homogêneas de precipitação têm sido utilizados no planejamento das atividades econômicas, sobretudo da atividade agrícola, visando assim o uso mais eficiente e racional dos recursos hídricos.

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

Estudos nessa área também são realizados para regiões que possuem dados de precipitação escassos, para que o regime pluviométrico seja descrito de acordo com as similaridades existentes com outra região mais favorecida de dados.

Os agrupamentos através das variáveis de precipitações podem ser úteis ao criar classificações mais atuais do que as classificações climáticas existentes, visto que estas normalmente se utilizaram de dados antigos e levam em consideração outras variáveis, além da precipitação.

A determinação de áreas homogêneas apenas através da análise das precipitações médias anuais pode nos dar como resultado o mapa de isoietas, neste mapa não são consideradas as variações intranuais e interanuais das precipitações. Algumas variáveis são capazes de nos dar a informação da distribuição temporal da precipitação ao longo do ano, bem como da sua concentração neste período, como os índices "Grau de Concentração de Precipitação" (Precipitation Concentration Degree - PCD) e "Período de Concentração da Precipitação" (Precipitation Concentration Period - PCP), utilizados em Li et at. (2011) e Zhang e Qian (2003).

Para realizar o estudo de uma determinada área em que estejam envolvidas diversas variáveis meteorológicas é preciso utilizar determinada(s) técnica(s) estatística(s). Uma das mais utilizadas, em diversas áreas do conhecimento científico, é a análise multivariada, onde um dos objetivos pode ser a redução do número de variáveis utilizadas, ou ainda a possibilidade de se investigar o comportamento espacial e temporal das variáveis consideradas a partir da obtenção de grupos homogêneos.

Dentro da análise multivariada existem várias técnicas empregadas, como a análise de componentes principais, análise fatorial, correlação canônica, análise de agrupamento, dentre outras. A análise de agrupamento consiste em determinar o nível de similaridade ou dissimilaridade entre indivíduos, aplicando uma função de agrupamento a uma determinada variável, transformando um conjunto heterogêneo de unidades em grupos caracterizados pela homogeneidade interna e pela heterogeneidade externa.

Nos trabalhos desenvolvidos com o intuito de definir regiões homogêneas é indicada a utilização de vários métodos de análise de agrupamento, optando-se por aquele cujos resultados sejam mais coerentes com o fenômeno em estudo. Dessa forma, a aplicação de várias técnicas estatísticas se justifica pelo fato de que cada uma gera diferentes resultados, podendo-se optar por aquela que melhor reflita as características climáticas da área de estudo (NERY, 1996).

O método hierárquico de Ward, e o método não hierárquico k-means são métodos simples e largamente utilizados em estudos climatológicos com resultados satisfatórios. O método de Ward foi utilizado por Raziei et al., (2008), que determinou a regionalização da precipitação para o Oeste do Irã e da variabilidade regional da seca. Muñoz-Díaz e Rodrigo (2004), utilizaram o método de Ward com o objetivo de comparar os resultados do mesmo com os da Análise de Componentes Principais. Unal et al., (2003), testaram cinco diferentes técnicas de agrupamento para determinar zonas climáticas homogêneas através das variáveis temperaturas (média, máxima e mínima) e precipitação total, sendo o método de Ward indicado como o mais provável de produzir resultados aceitáveis. De acordo com Fechine e Galvíncio (2008), o método Ward é o mais utilizado em estudo de agrupamento. Diversos outros trabalhos também utilizaram a técnica de Ward para determinação de áreas homogêneas com dados de precipitação (LYRA et al., 2006; FECHINE e GALVÍNCIO, 2008; KELLER et al., 2005 e SILVA et al., 2009).

Seidel et al., (2008) compararam os resultados obtidos pelos métodos Ward e k-means, no agrupamento por similaridade de produtores de leite da região de Santa Maria, Rio Grande do Sul. Os resultados encontrados tiveram alta concordância, evidenciando a robustez dos agrupamentos formados pelos dois métodos. Anderson et al., (2009) utilizaram Ward e k-means para regionalizar a média sazonal de precipitação no sudoeste dos Estados Unidos, e ambas as técnicas apresentaram 4 grupos idênticos.

Herrera et al., (1996) compararam a classificação climática de Köppen (1931) e um método multivariado, utilizaram a análise de agrupamento usando médias, de dez variáveis meteorológicas que caracterizam o clima, de 16 estações do Estado de São Paulo para um período de 13 anos de observações. Os resultados obtidos foram concordantes com a classificação climática para o Estado de São Paulo. Becker (1992) fez a classificação climatológica para o Rio Grande do Sul através de regiões homogêneas de precipitação pluvial

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

e de temperatura decendiais e mensais. No mesmo trabalho, esses grupos foram também usados para simular séries temporais e preencher valores faltantes no conjunto de dados.

Apesar de termos a classificação climática para o estado do Rio Grande do Norte não temos uma classificação baseada apenas nos dados de precipitação, o que pode ser de grande utilidade principalmente para a implementação de políticas públicas de desenvolvimento econômico e social da região embasadas no fomento do desenvolvimento de setores fortemente dependente da disponibilidade hídrica.

Sendo assim, este trabalho tem como objetivo determinar as áreas homogêneas em termos de precipitação no estado do Rio Grande do Norte, esta análise terá como base os dados de precipitação total anual e da sua variabilidade temporal e período de concentração.

2. MATERIAL E MÉTODOS

Área de estudo

A área de estudo deste trabalho é o estado do Rio Grande do Norte, que possui uma área de 52.810,7 km², onde existem quatro diferentes climas, de acordo com a classificação climática de Thornthwaite. Sendo eles:

DdA'a': Clima semi-árido com pequeno ou nenhum excesso de água, megatérmico, com baixa variação estacional;

EdA'a': Clima árido com pequeno ou nenhuma excesso de água, megatérmico, com baixa variação estacional;

C1wA'a': Clima seco e sub úmido, com moderado excesso de água, megatérmico, com baixa variação estacional;

C1dA'a': Clima seco e sub úmido, com pequeno ou nenhum excesso de água, megatérmico, com baixa variação estacional.

No Rio Grande do Norte os maiores quantitativos de precipitação concentram-se na região litorânea, com registros de precipitações superiores a 1500 mm (figura 01).

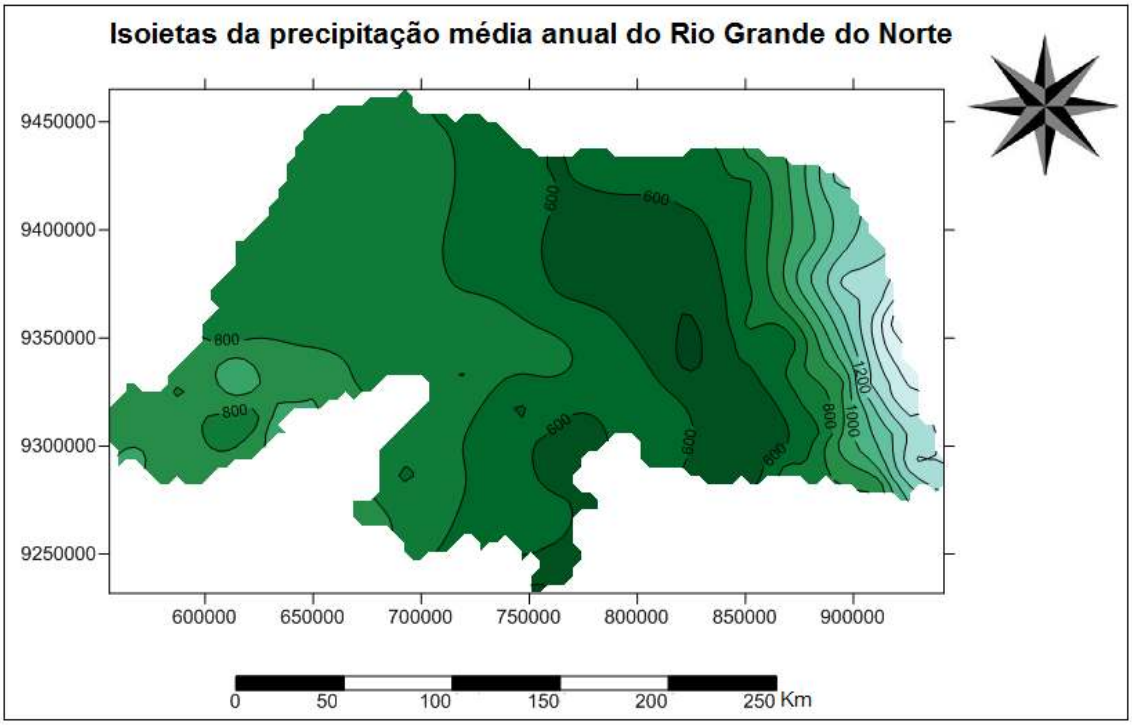


Figura 01 – Mapa das isoietas para o RN. Fonte: elaborado pelos autores.

Segundo o Anuário Estatístico do IDEMA (2010) o relevo do estado (Figura 02) é composto por: Planície Costeira, Planícies Fluviais, Tabuleiros Costeiros, Depressão Sub

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

Litorânea, Planalto da Borborema, Depressão Sertaneja, Chapada do Apodi, Chapada da Serra Verde.

A seguir é apresentado um mapa com a altimetria do estado do Rio Grande do Norte:

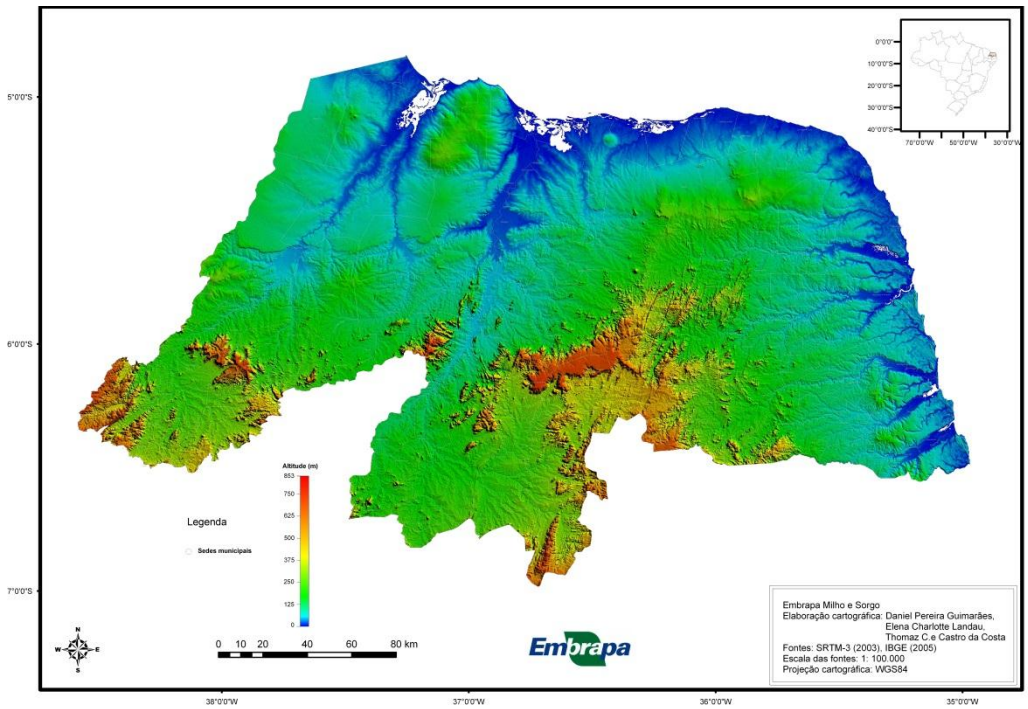


Figura 02 – Altimetria do Rio Grande do Norte. Fonte: Agritempo, 2011.

De acordo com Goovaerts (2000) a altitude tem grande efeito sobre a precipitação, influenciando em sua intensidade. Cardim (2001) afirma que diversos autores correlacionam a variabilidade espacial de chuvas com a altitude da região.

Caracterização da precipitação

Os dados de precipitação diária utilizados neste trabalho foram cedidos pela Empresa de Pesquisa Agropecuária do RN (EMPARN) e pela (ANA, 2011). Só foram utilizadas as estações que apresentam no mínimo 30 anos de dados, tendo sido totalizadas 144 estações pluviométricas distribuídas na área de estudo, com dados de 1963 a 2010.

Para a caracterização da distribuição temporal da precipitação na área de estudo serão utilizados dois índices: o Grau de Concentração de Precipitação (Precipitation Concentration Degree - PCD) e o Período de Concentração de Precipitação (Precipitation Concentration Period - PCP). O PCP refere-se ao mês em que o total de precipitação dentro de um ano se concentra e o PCD reflete o grau em que a precipitação total anual está concentrada num período de 12 meses. (XIE et al., 2005).

Li et at. (2011), estudou a variabilidade espacial e temporal de precipitação em Xinjiang, na China. O índice PCP determinou que a precipitação se concentra no verão, e a estação das chuvas chega mais cedo no oeste do que no leste ocidental de Xinjiang. O PCD indicou que a precipitação no norte da região foi mais dispersa dentro de um ano do que no sul. Zhang e Qian (2003) também aplicaram estes índices para analisarem as características da distribuição espacial e temporal e variação de precipitação em uma região da China.

Para o cálculo do PCD e do PCP parte-se do pressuposto que as precipitações mensais são grandezas vetoriais, orientadas em um círculo trigonométrico, onde a direção destes vetores é determinada em função do mês de ocorrência. A Tabela 01 apresenta a correspondência entre os valores de PCP e os meses do ano.

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

Tabela 01: Meses correspondentes a cada valor de PCP

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
PCP	0º	30º	60º	90º	120º	150º	180º	210º	240º	270º	300º	330º

Fonte: Li, et at.(2011)

Para se obter os valores de PCP e PCD calcula-se o vetor resultante das precipitações mensais, sendo PCP o ângulo deste vetor e PCD a relação entre o módulo do vetor resultante e a precipitação total anual. O cálculo destes índices é feito de acordo com as equações apresentadas a seguir (LI et at., 2011):

$$R_i = \sum r_{ij}$$

$$R_{xi} = \sum r_{ij}.sen\theta_j$$

$$R_{yi} = \sum r_{ij}.cos\theta_j$$

$$PCP_{ij} = \arctan(\frac{R_{xi}}{R_{yi}})$$

$$PCD_{ij} = \frac{\sqrt{{R_{xi}}^2 + {R_{Yi}}^2}}{R_i}$$

Onde "i" é o ano, e "j" representa o mês (j = 1, 2, ..., 12) em um ano; "r_{ij}" representa a precipitação total mensal no mês "j" no ano "i", e "θ_i" é o azimute do mês "j". PCP_{ij} representa o período (meses), no qual o total precipitação do ano "i" está concentrado e "PCD_{ij}" representa o grau em que a precipitação total do ano "i" está concentrada em 12 meses (LI, et at., 2011).

Análise de Agrupamento

A análise de agrupamento é uma técnica estatística de análise multivariada que tem como objetivo a análise de grupos ou variáveis através da similaridade de características, sejam elas geográficas, propriedades físicas ou estatísticas, ou seja, cada grupo contém a menor variância das variáveis (menor dissimilaridade) (HASSAN; PING, 2012). Os indivíduos de uma população em grupos devem ser homogêneos dentro de cada grupo formado e heterogêneos entre grupos.

O fato da Análise de Agrupamento não restringir o banco de dados de entrada para determinadas distribuições estatísticas (Gaussian) e levando em consideração as coincidências com resultados de técnicas como Análise de Componentes Principais, permitem afirmar que, pelo menos numa primeira abordagem, a Análise de Agrupamento é uma ferramenta adequada para descrever a variabilidade da precipitação (MUÑOZ-DÍAZ; RODRIGO, 2004).

Ward

Na Análise de Agrupamento hierárquica a disposição dos indivíduos em grupos distintos é feita a partir de uma função de agrupamento chamada distância ou similaridade e de um critério matemático de agrupamento. Na análise de agrupamento, todos os processos de hierarquização são similares, iniciando-se pela determinação da função de agrupamento. A função é empregada como critério para determinar a distância entre dois pontos x_i e x_j ou para definir o quanto eles são semelhantes (WILKS, 2006).

Os métodos hierárquicos são diferentes em relação à maneira com que a distância entre as entradas é calculada e como os dois pontos mais próximos são definidos. Os diferentes algoritmos utilizam critérios diferentes para formação de grupos, portanto, o número e as características dos agrupamentos podem ser distintos. Porém, todos os métodos hierárquicos seguem as seguintes fases: a medida de distância entre as estações climáticas é calculada; as duas melhores entradas são fundidas para formar um novo grupo; a distância

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

entre todas as entradas é recalculada e o segundo e o quarto passo são repetidos até que todas as entradas estejam em um grupo (UNAL, et al, 2003).

Um exemplo de medida de similaridade é o coeficiente de correlação e de dissimilaridade é a distância euclidiana. A maioria dos algoritmos utiliza medidas de similaridade e dissimilaridade entre os elementos de um conjunto, cujos valores de similaridade(dissimilaridade) calculados com base na matriz de dados originais X(nxp) são representados por uma matriz de distâncias ou de similaridade denominada matriz de parecença.

Neste trabalho foi utilizada a distância euclidiana, que é expressa pela seguinte equação:

$$d_{ij} = \left[\sum_{k=1}^{p} (x_{ik} - x_{jk})^2\right]^{1/2}$$

Onde "x_{ik}" e "x_{jk}" são, respectivamente, os valores da k-ésima variável das observações p-dimensionais para os indivíduos "i" e "j". Esta medida de distância tem a propriedade atraente, pois o coeficiente d_{ij} pode ser interpretado como distâncias físicas entre dois pontos p-dimensionais $x_{'i} = (x_{i1},... X_{ip})$ e $x_{'j} = (x_{j1},... X_{jp})$ no espaço euclidiano *(*EVERITT et al., 2011).

Nos métodos hierárquicos existem diversos critérios de agrupamentos, dentre eles o método de Ward (WARD, 1963). Ao realizar um agrupamento entre indivíduos, seja ele em qualquer estágio, informações são perdidas, porém podem ser medidas, segundo Ward (1963), pela soma total do quadrado dos desvios (SQD) de cada ponto, da média do conglomerado ao qual ele pertence. Os grupos formados são aqueles em que a fusão resulte no mínimo incremento da inércia do conjunto.

Pelo método de Ward grupos são criados em cada etapa do procedimento, de forma que a solução resultante tenha o menor SQD (equação 2) dentro de grupos. Nessas etapas, são consideradas as uniões de todos os possíveis pares de grupos e, os dois que resultam em menor aumento de SQD são agrupados até que todos os grupos formem um único, reunindo todos os indivíduos (LYRA et al., 2006).

A soma dos quadrados dos desvios de *p* indivíduos organizados em *k* grupos é definida como:

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

Onde *n* é o número total de elementos do agrupamento e x_i é o n-ésimo elemento do agrupamento.

O método de Ward pode ser resumido nas seguintes fases, para cada série: 1) cálculo das médias das variáveis; 2) obtenção da distância euclidiana entre essas médias; 3) soma das distâncias para todos os indivíduos e 4) minimização da variância dentro dos grupos (SILVA et al., 2001).

O método de Ward tem tendência em criar grupos de tamanhos semelhantes e pequenos, fato este que pode distorcer a realidade de determinado local, sendo uma desvantagem apontada pelos pesquisadores da área de Climatologia. Apesar desta desvantagem, esse método tem se mostrado bastante eficiente, pois os resultados obtidos, após sua aplicação, confirmam o número de clusters pré-estabelecidos, com base em outras informações climatológicas para áreas de estudos (SILVA, 2001).

K-means

A outra metodologia de agrupamento utilizada neste trabalho foi o método k-means, que tem como objetivo separar um conjunto de dados em grupos consistentes, cada um é agrupado em um determinado grupo, que inicialmente são formados aleatoriamente, e o

processo de agrupamento continua até que seja obtida a maior semelhança dentro de cada grupo (MUÑOZ-DIAZ; RODRIGO, 2004).

O método k-means é um método não hierárquico por repartição. Consiste de uma metodologia onde, dado um número de grupos previamente determinado, é feito o cálculo de pontos que representam os "centros" destes grupos e que são espalhados homogeneamente no conjunto de respostas obtidas através do questionário e movidos, heuristicamente, até alcançar um equilíbrio estático (HAIR et al., 2005).

Após esse processo há uma divisão de todos os casos obtidos pelos "k" grupos preestabelecidos e a melhor partição dos "n" casos será aquela que otimize o critério escolhido. Em outras palavras, o método inicia-se usando os valores dos primeiros "k" casos como estimativas temporárias das médias dos grupos.

Os centróides iniciais são desenvolvidos através da designação de cada caso ao grupo de centro mais próximo. Dessa forma, a média altera-se, alterando assim o centróide. Um processo iterativo é usado para achar os centróides finais de cada grupo. A cada passo, os casos são agrupados ao grupo de centro mais próximo, e novamente as médias são calculadas. Este procedimento permanece até que não haja mais alterações nas médias ou que um número pré-determinado de iterações aconteça, encerrando-se o processo (HAIR et al.,2005).

Contrariamente ao procedimento de agrupamento hierárquico, o processo não-hierárquico, o algoritmo k-means, exige que o número de clusters seja conhecido e que os objetos sejam distribuídos entre eles. Este algoritmo é amplamente utilizado nos casos em que uma informação a priori sobre a natureza das medições está disponível. Um exemplo é a classificação dos tipos de aerossóis (OMAR et al., 2005).

Gordon e Norris (2010), que estudaram as propriedades de nuvens oceânicas, afirmam que definir o número de grupos para o método k-means é o aspecto mais subjetivo deste agrupamento. Os autores definiram o número de grupos analisando as propriedades das variáveis utilizadas e as características locais da região em estudo. Dessa forma, a análise conjunta de dois tipos de agrupamento, um hierárquico e outro não, é importante na definição do número de grupos.

Ramos (2001) utilizou o método k-means e o método de Ward para estudar as variações das precipitações pluviais no Mediterrâneo, as quais são responsáveis pelo clima da região. Os métodos foram aplicados às precipitações pluviais diárias do período de 1889 a 1999 do Alt Penedès, região Nordeste da Espanha. Foram analisados, também, valores de precipitações pluviais anuais das estações da primavera e outono. A utilização de mais de um método de agrupamento, permitiu extrair um número maior de informações. O uso da técnica hierárquica permite visualizar a combinação das observações para formar os grupos, o que não é possível na técnica não hierárquica.

André et al., (2008) utilizaram o método de Ward para determinar quantas regiões com homogeneidade pluvial existiam no Rio de Janeiro e o método de k-means para realizar o agrupamento propriamente dito.

Para o agrupamento de eventos meteorológicos, a metodologia hierárquica de Ward forneceu resultados parecidos com o aglomerado produzido por meio de k-means, portanto, pode ser usada como uma alternativa (SHAFER; DOSWELL, 2010)

Índices de validação do agrupamento

Como afirma Brock et al., (2008), as métricas utilizadas para validação do número de agrupamentos são divididas em dois grandes grupos: Interna e Estabilidade. Na validação interna os índices existentes são: índice de Dunn, índice de Conectividade e a largura de Silhouette (Everitt et al., 2001).

Na validação interna foram selecionados medidas que reflitam a compactação, conectividade e separação dos agrupamentos gerados. A conectividade refere-se à extensão na qual uma instância está alocada a um mesmo agrupamento dos seus vizinhos mais próximos. A compactação avalia a homogeneidade do agrupamento, usualmente calculando a variância intra-agrupamento, enquanto a separação quantifica o grau de separação dos agrupamentos, usualmente pela medida da distância dos centróides. Uma vez que a compactação e a separação apresentam tendências opostas, ou seja, compactação aumenta

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

com o número de agrupamentos e a separação diminui, uma opção é unir as duas métricas. Duas medidas que representam uma combinação não linear de compactação e separação são o índice de Dunn e a largura de Silhouette (EVERITT et al., 2001).

O índice Dunn é a razão entre a menor distância entre instâncias que não estejam no mesmo agrupamento e a maior distância intra-agrupamento. O valor do índice Dunn varia de 0 a 1 e quanto mais próximo de 1 melhor é o resultado (BROCK et al., 2008).

A largura de Silhouette também é recomendada por Everitt et al. (2001). Segundo Rao e Srinivas (2006), a largura de Silhouette para um vetor de característica é uma medida de quão semelhante aquele vetor de característica é para os vetores de características do seu próprio agrupamento comparado com os vetores de características dos outros agrupamentos.

Para cada instância i é definido um índice S(i) ϵ [-1, 1] que mede a diferença entre b(i) e a(i), onde a(i) é a dissimilaridade média da instância i em relação ao seu agrupamento e b(i) é a dissimilaridade média da instância i em relação a todas instâncias no agrupamento mais próximo. Quando S(i) está próximo de 1[a] instância i está mais próxima do seu agrupamento do que do agrupamento vizinho, e, portanto representa uma boa alocação. Quando S(i) está próximo de -1, a instância está mal alocada. Os autores da proposta indicam ainda que valores acima de 0,5 representam um bom resultado e que valores abaixo de 0,2 podem indicar ausência de estruturação clara dos dados (CARDOSO JUNIOR et al., 2010).

A conectividade refere-se a extensão na qual uma instância está alocada a um mesmo agrupamento dos seus vizinhos mais próximos. É definida pela equação 3:

$$Conn\ (C) = \sum_{i=1}^{N} \sum_{j=1}^{L} x_i nn_i(j)$$

onde N representa o número total de observações e L o número de dimensões, $nn_i(j)$ é o j-ésimo vizinho mais próximo da instância i na dimensão j, e x_i, $nn_i(j) = 0$ se i e j estiverem no mesmo agrupamento e 1/j caso contrário.

A conectividade tem valores entre $0 \leq Conn(C) \leq \infty$, e é uma métrica que deve ser minimizada, ou seja, quanto menor o valor melhor será a estrutura proposta pelo algoritmo (EVERITT et al., 2001). Everitt et al. (2001), também alertam não ser prudente confiar em apenas uma das métricas para selecionar o número ideal de agrupamentos.

A validação de estabilidade tem como métricas: Average Proportion of Nonoverlap - APN, Average Distance - AD, Average Distance between Means - ADM, e Figure of Merit – FOM. A APN mede a proporção média de instâncias não alocadas ao mesmo agrupamento quando a matriz completa é utilizada e quando uma coluna é retirada. Seja o agrupamento que contém a instância i usando a matriz completa e quando uma coluna i foi retirada.

Considerando o número total de agrupamentos igual a C, define-se:

$$APN\ (C) = \frac{1}{MN} \sum_{i=1}^{N} \sum_{l=1}^{N} \left(1 - \frac{n\left(C^{i,l} \cap C^{i,0}\right)}{C^{i,0}}\right)$$

Com APN(C) ϵ [0,1], sendo que valores próximos a zero correspondem a um agrupamento consistente (BROCK et al., 2008).

AD mede a distância média entre instâncias de um mesmo agrupamento baseada na matriz completa e com a retirada de uma coluna, é definida por:

$$AD\ (C) = \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\frac{1}{n(C^{i,0})n(C^{i,l})}\right) \left[\sum_{i\ \epsilon\ C^{i,0}, j\ \epsilon\ C^{i,l}} dist(i,j)\right]$$

com , AD (C) ϵ [0, ∞], sendo que valores próximos a zero os melhores resultados.
ADM mede a distância média entre centros dos agrupamentos, e é definida por:

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

$$ADM\ (C) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(dist(\ x_{C^{i,l}}, x_{C^{i,o}})\right)$$

com ADM (C) ϵ [0, ∞], e da mesma forma que AD, quanto menor melhor.
E finalmente FOM mede a variância média intra-agrupamento das instancias na coluna suprimida, onde o agrupamento está baseado nas colunas restantes. É estimado o erro médio usando previsões baseadas nas médias dos agrupamentos. Para uma coluna l deixada de fora, temos:

$$FOM\ (l, C) = \sqrt{\frac{1}{N}\sum_{k=1}^{K}\sum_{i\epsilon Ck(l)}\left(dist(x_{i,l}, x_{Ck(l)})\right)}$$

Onde $x_{i,l}$ é o valor da i-ésima instancia da l-ésima coluna no agrupamento $C_{k(l)}$ e $x_{Ck(l)}$ é a média do agrupamento. Finalmente FOM é multiplicado por um fator RAIZ(N/N-K), para minimizar a tendência de FOM decrescer à medida que K cresce. O escore final é calculado pela média de todas as colunas removidas, e os valores obtidos estão entre zero e ∞, com os menores valores representando a melhor configuração (BROCK et al., 2008).

Agrupamentos Propostos

Neste trabalho foram realizadas algumas simulações para a definição das variáveis de entrada a serem utilizadas para o agrupamento. Primeiramente foram analisados os resultados dos seguintes agrupamentos:

- agrupamento 1: método de agrupamento de Ward e as variáveis utilizadas foram as médias das precipitações anuais e os índices PCP e PCD;
- agrupamenmto 2: método de agrupamento de Ward e as variáveis utilizadas foram as doze precipitações médias mensais.

Após esta análise preliminar foi determinado o número de grupos a ser adotado em função dos resultados dos índices de validação do agrupamento. O agrupamento então foi realizado para o número de grupos identificado e a escolha das variáveis que melhor agrupavam as estações foi definido a partir de uma análise da similaridade de comportamento da variabilidade da precipitação mensal das estações localizadas em cada grupo e seus índices PCP e PCD. Com a definição do número de grupos e das variáveis a serem adotadas foi realizado o agrupamento 3.

- Agrupamento 3: método k-means para 4 grupos e as variáveis utilizadas foram as médias das precipitações anuais e os índices PCP e PCD;

O software utilizado para o desenvolvimento da pesquisa foi o Programa R - The R Project for Statistical Computing (R DEVELOPMENT CORE TEAM, 2011). Para realizar a validação foi utilizado o pacote "clValid".

3. RESULTADOS E DISCUSSÃO

Variabilidade Temporal da Precipitação

Aos resultados apresentados a seguir referem-se aos valores de PCD e PCP encontrados para as 144 estações pluviométricas estudadas e a espacialização desta informação para o estado do Rio Grande do Norte. A interpolação dos dados resultando nos mapas das figuras 03 e 04, foi realizada através do software "Surfer", com o uso da ferramenta "kriging".

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

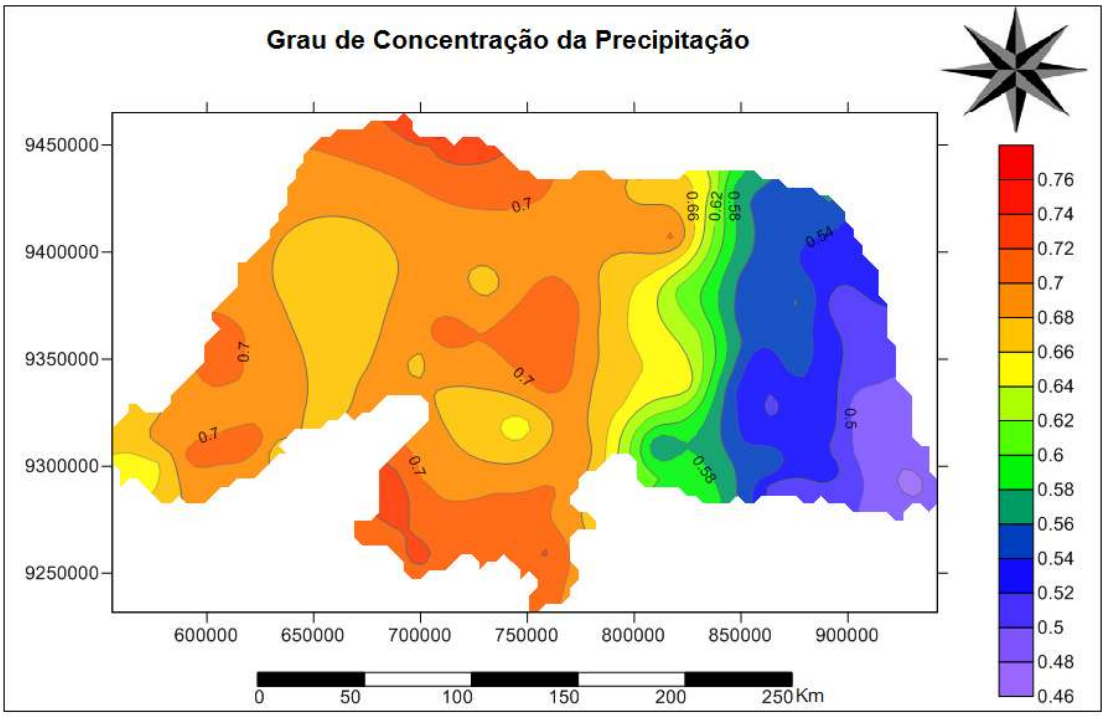


Figura 03 – Grau de Concentração de Precipitação para o Estado do RN.

Como podemos observar, na região leste da área de estudo as precipitações são melhores distribuídas ao longo do ano, ou seja, possuem menores valores de PCD. À medida que nos dirigimos para a região oeste, os índices aumentam, refletindo áreas com más distribuições de precipitação. Municípios localizados na região central e na região oeste potiguar apresentam valores de PCP acima de 0,72.

Existem regiões no alto oeste e na região do Seridó (localizado no centro-sul do estado) onde as precipitações são melhores distribuídas, como valores de PCD em torno de 0,65. Esses locais correspondem a estações localizadas em municípios com elevadas atitudes e sujeitas ao efeito orográfico. No alto oeste temos melhor distribuição das precipitações no município de Venha Ver (local onde se encontra a Serra do Coqueiro) e a Serra de Luís Gomes.

Outra região que se destaca por ter uma melhor distribuição de precipitação em relação às áreas vizinhas é a da Chapada do Apodi, possui PCD em torno de 0,67, onde encontramos regiões mais propícias às práticas agrícolas.

O litoral Sul do estado também merece destaque, pois é a região com melhor distribuição de chuvas. O que pressupõe, nessas localidades, a ação das Perturbações Ondulatórias dos Alísios, em maior intensidade.

A figura (figura 08) a seguir apresenta os resultados obtidos com a outra variável utilizada para a área de estudo, o índice PCP.

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

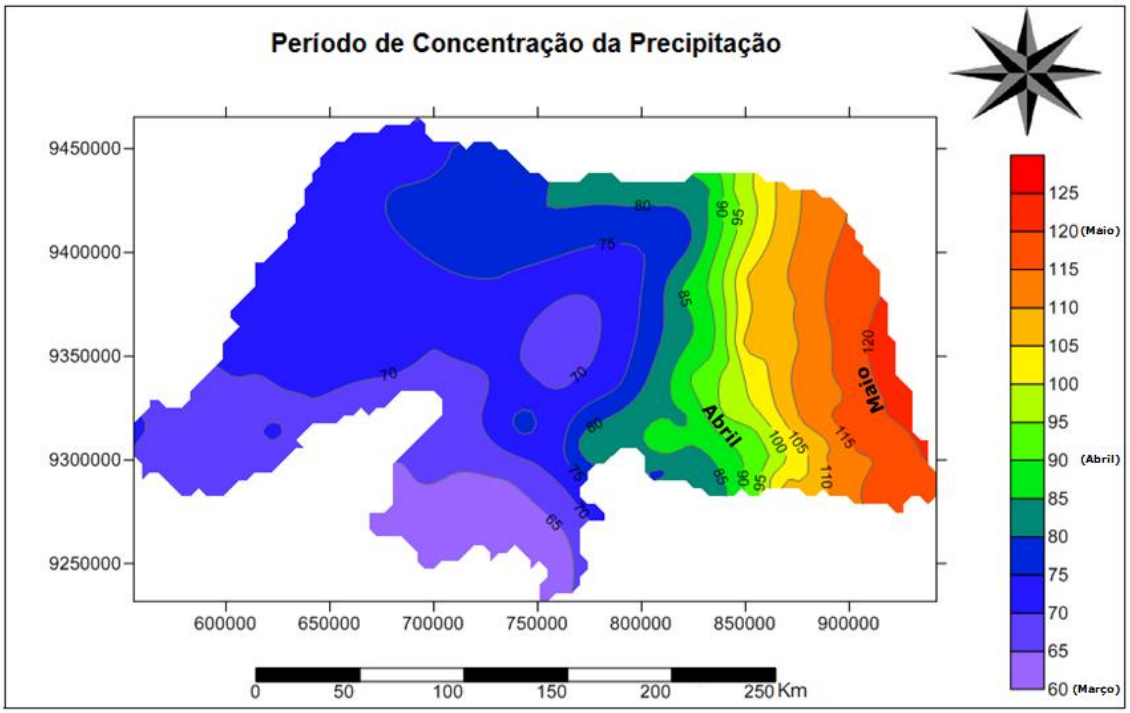


Figura 04 – Período de Concentração da Precipitação para o Estado do RN.

As áreas localizadas no leste do estado tem como período mais chuvoso o mês de maio (120º, Tabela 01), pois nessa região (leste e agreste) além da atuação da Zona de Convergência Intertropical (ZCIT) temos a ocorrência das Perturbações Ondulatórias dos Alísios (POA's), que atuam de maio a agosto.

Esse período vai sendo alterado para abril (90º) e março (60º) a medida que se desloca para o oeste do estado, pois neste caso o fator determinante de chuvas é o deslocamento da ZCIT. A atuação da ZCIT é no período de fevereiro a maio na região semiárida central, oeste e norte do estado.

Índices de validação do agrupamento

Para que fosse definido o número de grupos foram aplicados os testes de validação, para os Agrupamentos 1 e 2, e os resultados estão expressos nas figuras 05, 06, 07 e 08.

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

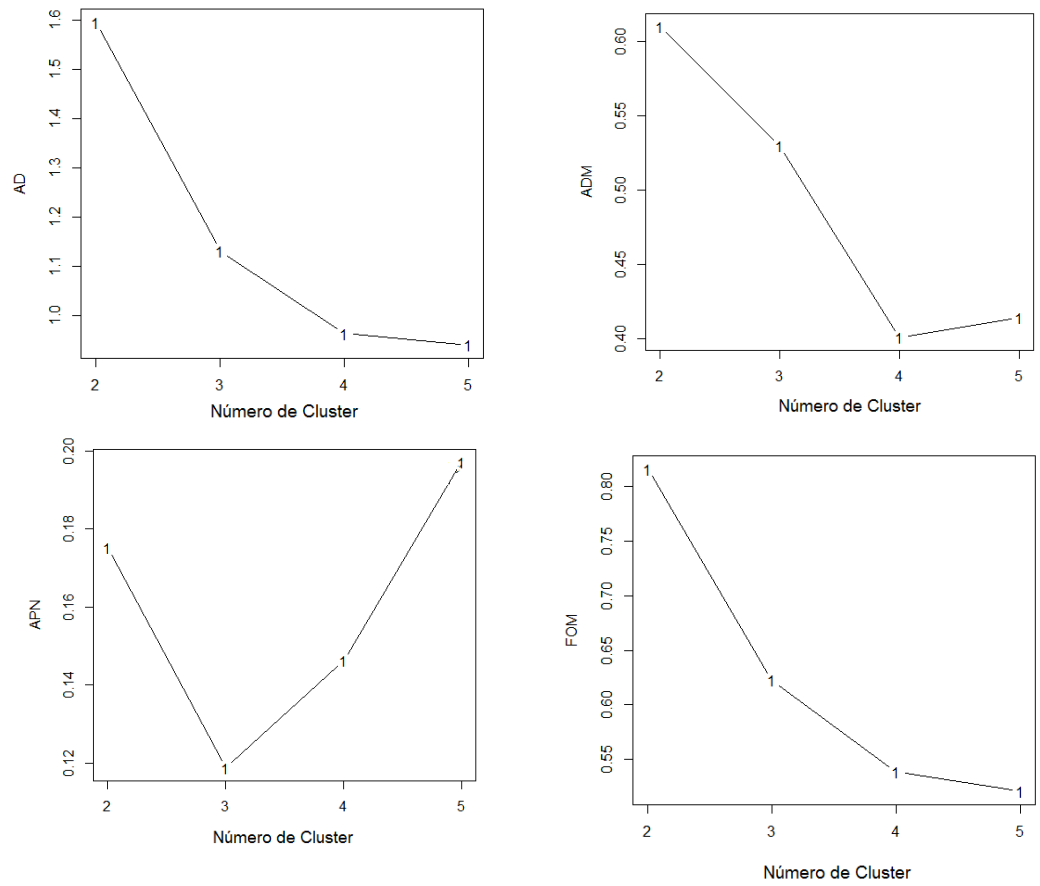


Figura 05 – Validação de estabilidade (Agrupamento 1).

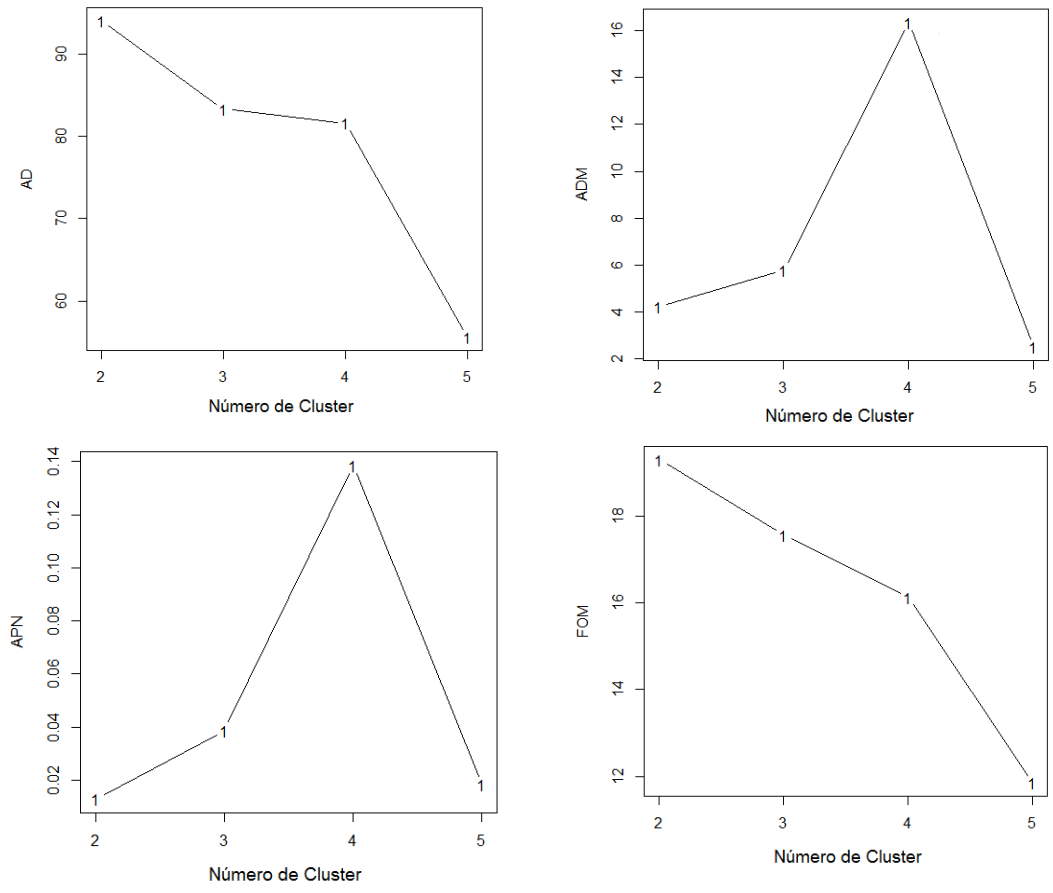


Figura 06 – Validação de estabilidade (Agrupamento 2).

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

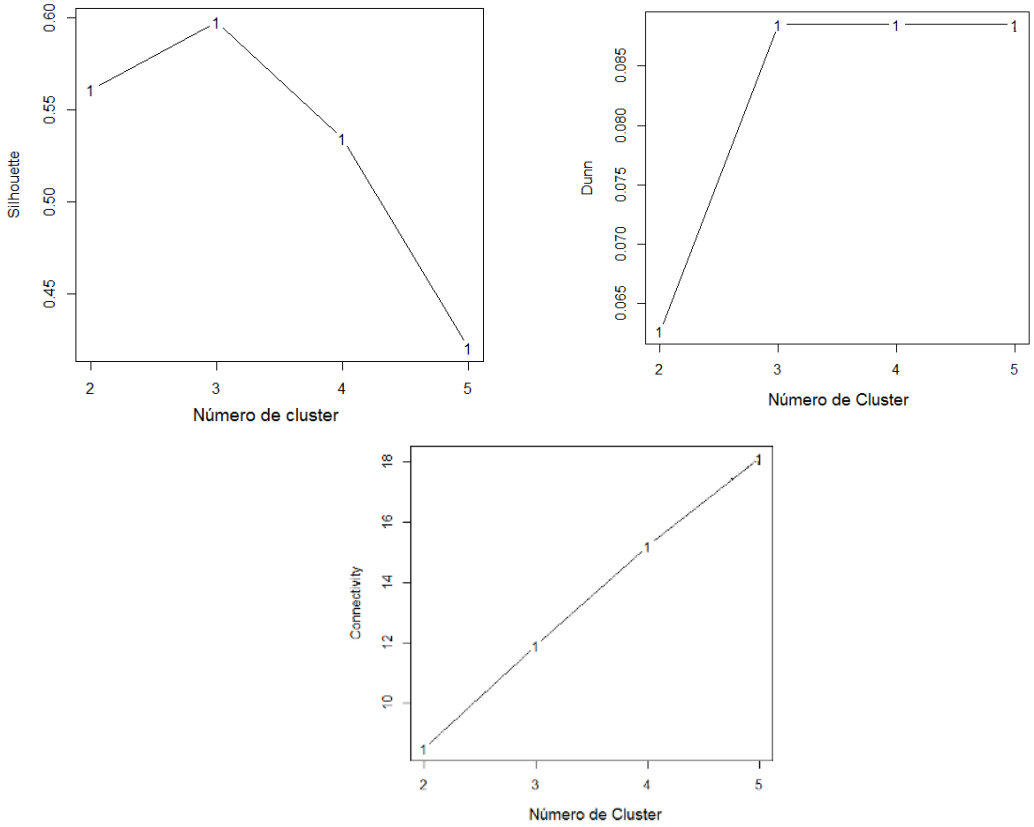


Figura 07 – Validação interna (Agrupamento 1).

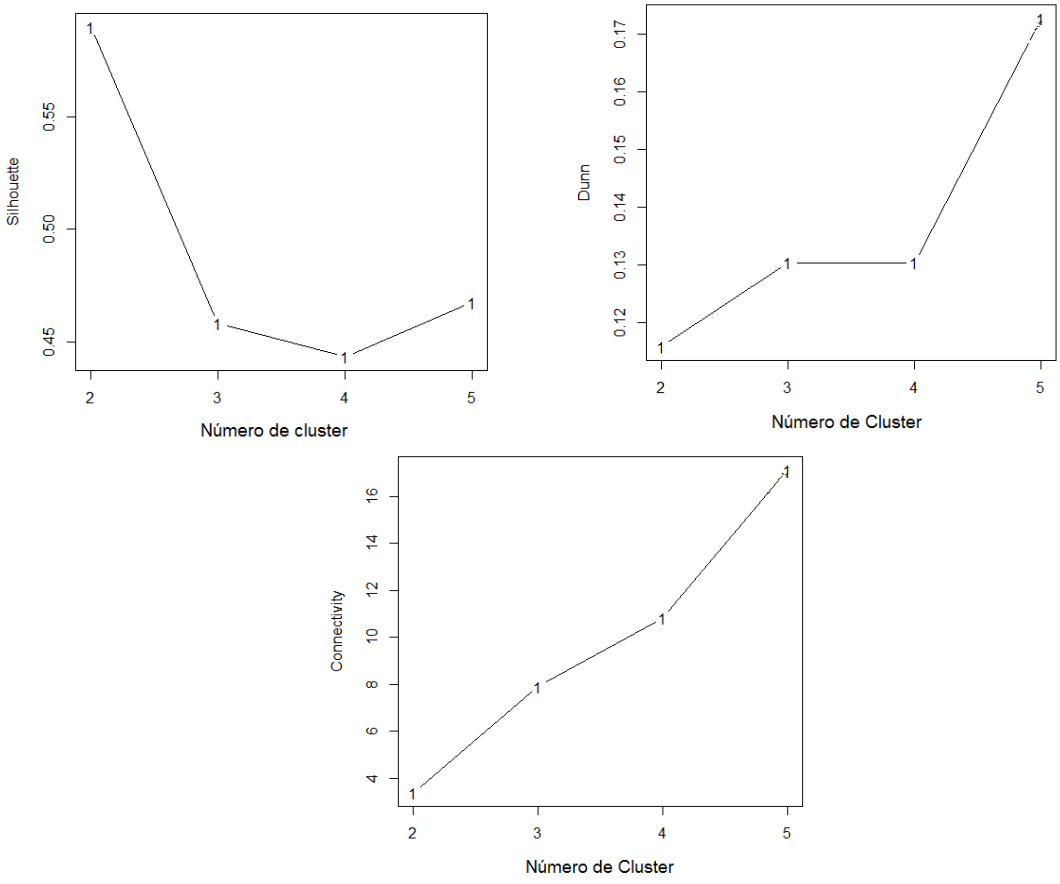


Figura 08 – Validação interna (Agrupamento 2).

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

Na tabela a seguir (tabela 02) são apresentados os dois números de grupos mais indicados para cada teste de validação realizado. Para os índices de validação interna, os melhores resultados, no caso do índice de Dunn e Silhouette, são aqueles que em eu seus valores são maximizados, e no caso de conectividade, os menores valores representam os melhores agrupamentos. Todos os índices de validação de estabilidade apresentam melhores resultados quando seus valores são minimizados.

Tabela 02: Tipo de validação e os números de grupos mais indicados

Testes de Validação	Tipo	Agrupamento 1	Agrupamento 2
Validação Interna	Silhouette	2, 3	2, 5
	Dunn	3, 4	4, 5
	Conectividade	2, 3	2, 3
Validação de Estabilidade	AD	4, 5	4, 5
	ADM	4, 5	2, 5
	APN	3, 4	2, 5
	FOM	4, 5	4, 5

De acordo com os resultados apresentados na tabela 02, para o agrupamento 1 foi considerado que quatro seria o número de grupos mais adequado, pois apareceu de forma mais recorrente nas validações interna e de estabilidade. Já para o agrupamento 2 o melhor seria o uso de cinco grupos. O número de grupos escolhidos neste trabalho foi quatro, em função dessa divisão refletir melhor a realidade da área de estudo em relação à sua classificação climática (Thornthwaite), que divide o estado em 4 diferentes tipos de clima.

Os agrupamentos com a separação dos quatro grupos são apresentados nas figuras 09 e 10, onde a figura 09 apresenta os resultados para o uso das variáveis: PCD, PCP e precipitações médias anuais (Agrupamento 1) e a figura 10 apresenta o agrupamento realizado com 12 variáveis que se referem as precipitações médias mensais (Agrupamento 2).

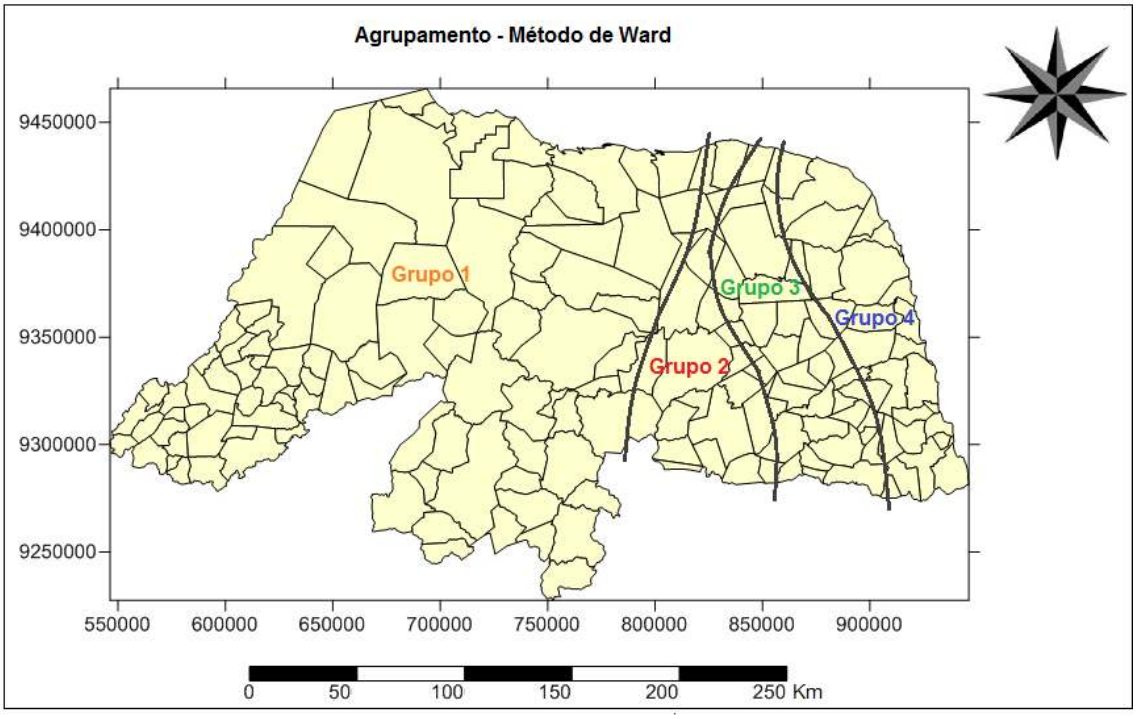


Figura 09 – Agrupamento das estações (Agrupamento 1)

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

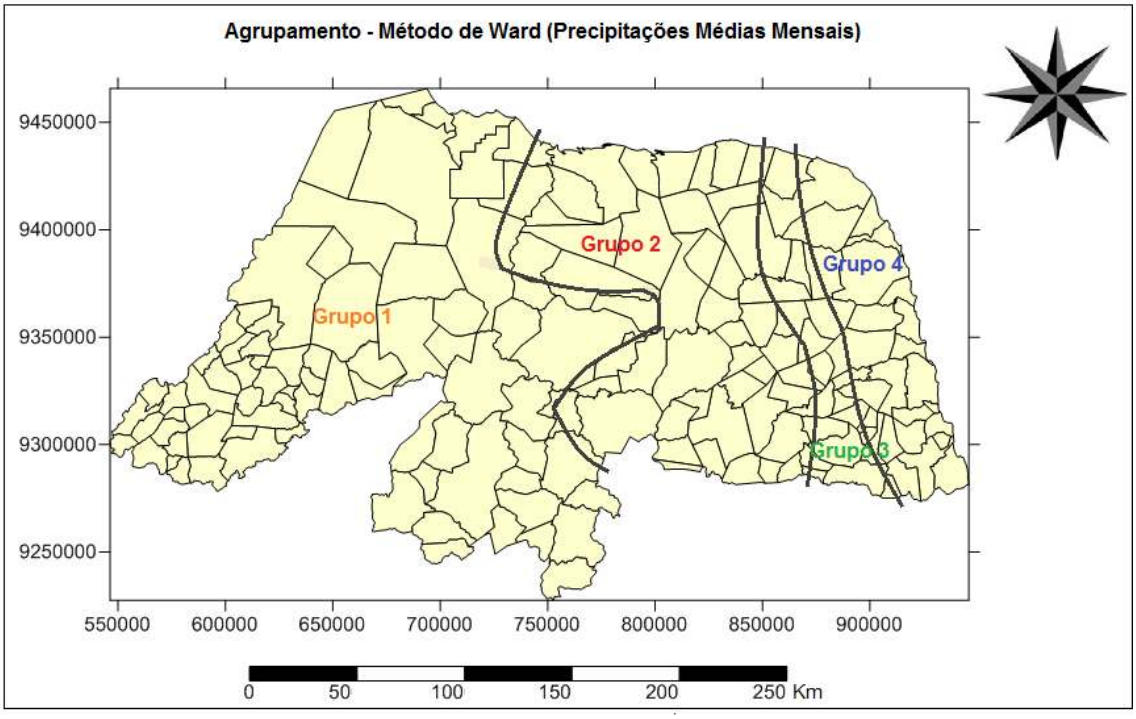


Figura 10 – Agrupamento das estações (Agrupamento 2).

Estações localizadas nos grupos 1 e 3 no agrupamento 1, migraram para o grupo 2, quando o agrupamento foi feito com as precipitações médias mensais (agrupamento 2). Os municípios do grupo 3 que migraram foram: Bento Fernandes, Santa Maria, Riachuelo, Serrinha, Passa e Fica, Senador Eloi de Souza, Serra de São Bento, Boa Saúde, Serra Caiada, São Paulo do Potengi, João Câmara e São Pedro. Já os do grupo 1 foram: Angicos, Jandaíra, Pedro Avelino, São Vicente, Lagoa Nova, Guamaré, Macau, Acari, Currais Novos, Equador, Cerro Corá, Assú, Tenente Laurentino Cruz, Bodó, Fernando Pedroza, Galinhos, Caiçara do Norte, Alto do Rodrigues, Porto do Mangue e Itajá. Taipu pertencia ao grupo 4 e na segunda aglomeração migou para o grupo 3.

A figura 11 apresenta a análise da migração de algumas cidades do grupo 1 para o grupo 2, na cor preta são representados os municípios que migraram, em laranja estão os municípios do grupo 1 que permaneceram neste mesmo grupo e em vermelho os municípios que eram e permaneceram no grupo 2. A figura permite observar o padrão de precipitação dos municípios que migraram, e se este se assemelha mais ao padrão do grupo 1 ou do grupo 2.

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

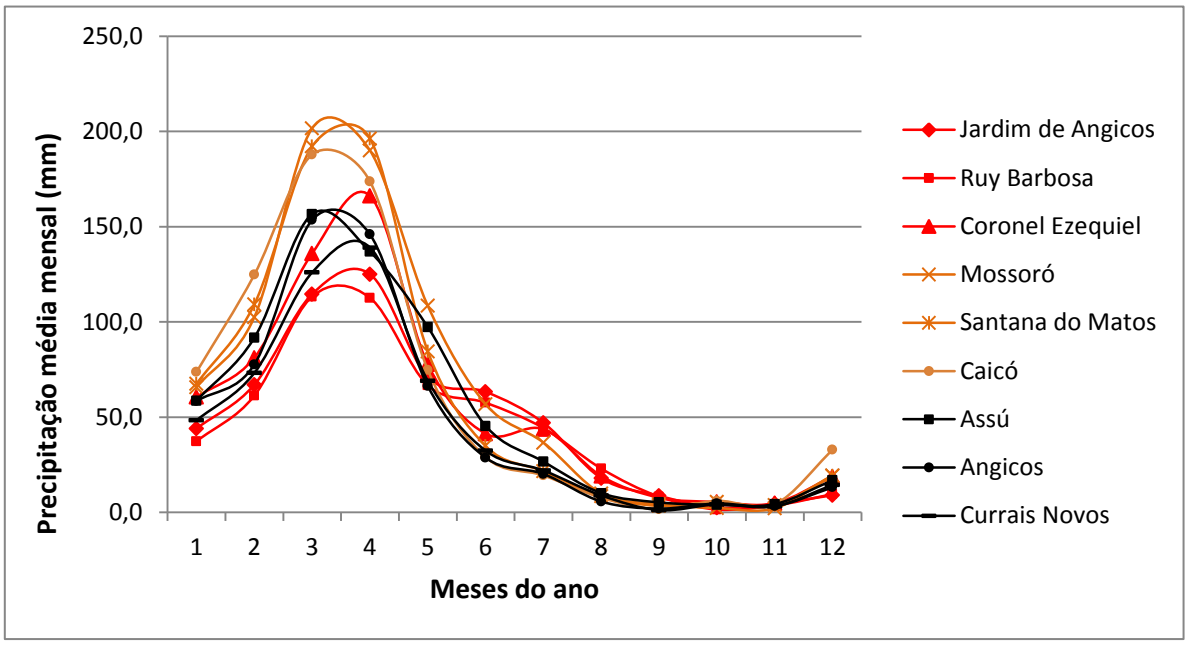


Figura 11 – Precipitações médias mensais das cidades que migraram, suas vizinhas e do grupo 3.

Podemos observar que o regime pluvial das estações que migraram (Assú, Angicos e Currais Novos) se parecem mais com os elementos do grupo 1 (Mossoró, Santana do Matos, Caicó), por possuírem apenas um pico de chuva. No entanto em termos dos valores de precipitação média mensal, as estações que migraram se assemelham mais ao grupo 2 (Jardim de Angicos, Ruy Barbosa, Coronel Ezequiel).

Também podemos verificar que o primeiro agrupamento reflete melhor a realidade local através das observações nos valores de PCD, PCP (Tabela 03). Os municípios do grupo 2 (em vermelho) possuem menores valores de PCD e maiores de PCP. Essa característica se deve ao fato de que nessa região começa a haver influência das POA's como mecanismo formador de chuva. Dessa forma as estações passam a ter outro padrão, com dois picos de chuva, o que faz com que as precipitações sejam mais bem distribuídas (menores valores de PCD) e que as concentrações de chuvas se atrasem para o mês de abril (maiores valores de PCP).

Tabela 03: Valores de PCD e PCP para as estações em análise

Cidade	PCD	PCP	Mês de concentração
Jardim de Angicos	0,590	90,9	Abril
Ruy Barbosa	0,595	92,1	Abril
Coronel Ezequiel	0,607	78,7	Março
Mossoró	0.691	75.676	Março
Santana do Matos	0.717	68.034	Março
Caicó	0.698	62.459	Março
Assú	0.671	74.459	Março
Angicos	0.717	68.797	Março
Currais Novos	0.694	84.333	Março

A mesma análise foi feita para as estações do grupo 3 que migraram para o grupo 2 (em preto), os municípios que eram e permaneceram no grupo 2 (em vermelho) e os municípios que eram e permaneceram no grupo 3 (em verde) (Figura 12).

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

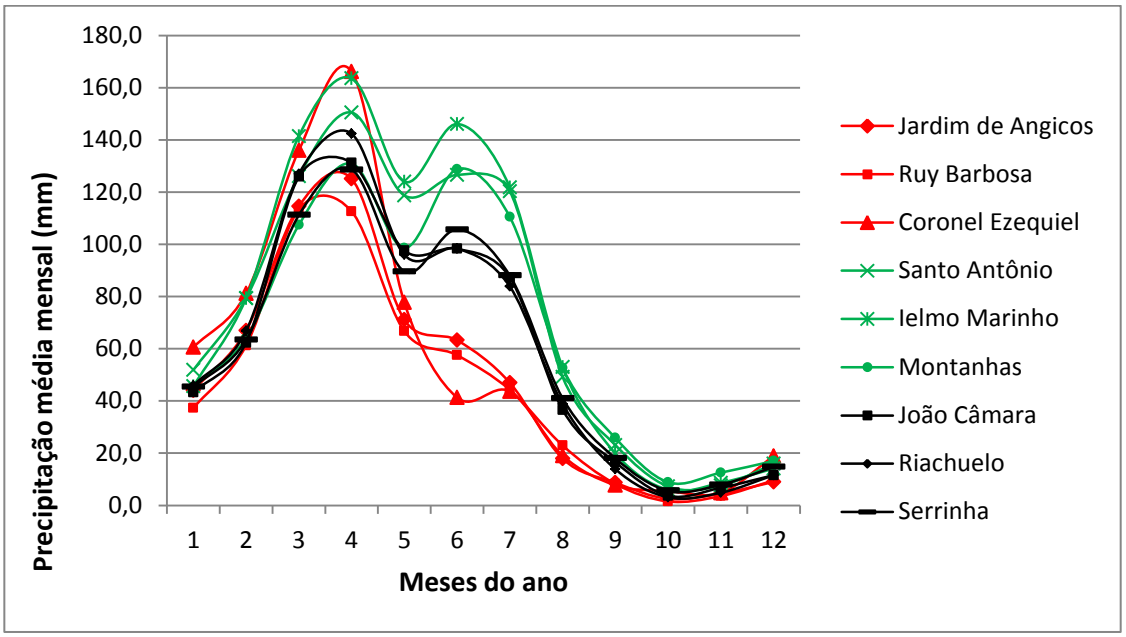


Figura 12 - Precipitações médias mensais das cidades que migraram, suas vizinhas e do grupo 2.

Neste segundo caso, podemos observar que as estações que migraram possuem características mais próximas das estações do agrupamento em que pertenciam no primeiro caso (agrupamento 1), pois a curvas em preto se assemelham mais com as curvas em verde, com a característica de dois picos de precipitação mais acentuados.

Analisando os valores médios de PCD, PCP neste segundo caso (tabela 04), também podemos observar uma maior semelhança entre as estações que migraram (João Câmara, Riachuelo e Serrinha) e o grupo ao qual elas pertenciam no primeiro agrupamento (Santo Antônio, Ielmo Marinho e Montanhas), com chuvas melhor distribuídas (menores valore de PCD) e com concentrações de chuvas se aproximando do mês de maio.

Tabela 04: Valores de PCD e PCP para as estações em análise

Cidade	PCD	PCP	Mês de concentração
Jardim de Angicos	0.590	90.929	Abril
Ruy Barbosa	0.595	92.097	Abril
Coronel Ezequiel	0.607	78.664	Março
Santo Antônio	0.533	111.000	Abril
Ielmo Marinho	0.543	111.307	Abril
Montanhas	0.494	116.655	Abril
João Câmara	0.550	106.557	Abril
Riachuelo	0.556	108.511	Abril
Serrinha	0.521	107.007	Abril

Em função das análises realizadas foram encontrados os melhores resultados para o primeiro agrupamento, onde foram utilizadas as variáveis de PCP, PCD e precipitação média anual. Sendo assim, para o agrupamento 3, foi aplicado o método de agrupamento k-means e as variáveis PCP, PCD e precipitação média anual, para a formação de 4 grupos (Figura 13).

ISSN: 1980-055x (Impressa) 2237-8642 (Eletrônica)

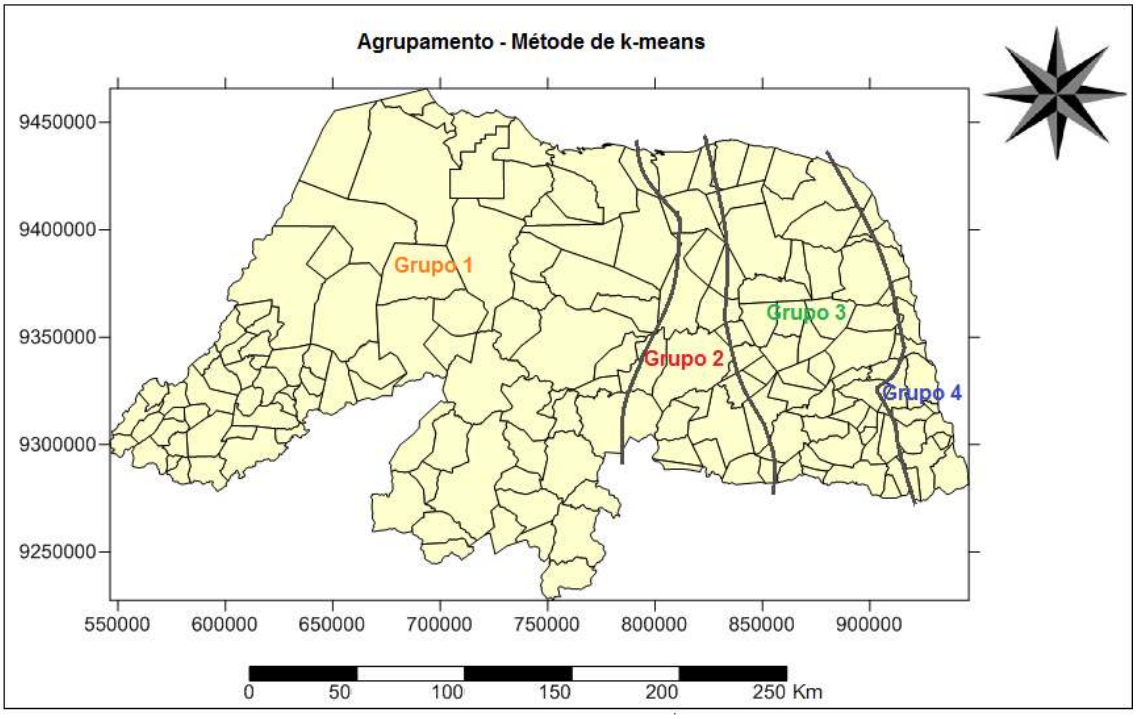


Figura 13 – Agrupamento através do método K-means para as variáveis PCD, PCP e precipitações médias anuais (Agrupamento 3).

Das 144 estações classificadas, 5 (3,47%) delas foram alocadas em grupos distintos, do agrupamento de Ward (Agrupamento 1, figura 09). Isso aconteceu com estações de Lagoa Nova, Galinhos e Caiçara do Norte, que em Ward pertenciam ao grupo 1 e em K-means migraram para o grupo 2. Já o município de Pedra Grande passou do grupo 2 para o grupo 3. E por fim, Taipú que fazia parte do grupo 4 migrou para o grupo 3.

A migração de apenas 5 estações evidencia a eficiência da robustez dos agrupamentos formados pelos métodos de Ward e K-means. A utilização dos dois métodos conjuntamente foi de grande utilidade, pois foi possível utilizar o número de agrupamentos encontrado pelo método Ward, para a definição do número de grupos que devem ser formados pelo método K-means.

A tabela 05 apresenta os valores médios das variáveis que caracterizam o regime pluvial de cada região homogênea, encontrada pelo método k-means (agrupamento 3).

Tabela 05: Média dos valores de PCD, PCP e precipitação média anual para as quatro regiões homogêneas

GRUPOS	PCD MÉDIO	PCP MÉDIO (°)	PRECIPITAÇÃO MÉDIA ANUAL (mm)
Grupo 1	0.69	70.4	751
Grupo 2	0.58	90.2	589
Grupo 3	0.53	109.9	803
Grupo 4	0.51	118.5	1339

Como podemos observar a região do grupo 1, onde se enquadraram 78 estações, é a que apresenta pior distribuição de chuvas do estado e como mecanismo principal formador de chuva temos apenas a ZCIT.

No segundo grupo temos 21 estações é uma região que apresenta o maior déficit hídrico do estado, demandando maiores cuidado em relação a gestão das águas. Através da análise das médias mensais realizadas nas figuras 11 e 12, podemos notar um efeito discreto dos dois sistemas de formação de chuva, já que em alguns municípios ainda podemos observar dois picos de chuva durante o ano, porém a ZCIT, que atua em todo o estado nos períodos de fevereiro a maio, é o sistema que tem uma maior influência na região.

O terceiro grupo é composto por 27 estações, neste caso temos uma melhor distribuição de chuva em relação aos grupos 1 e 2, e precipitações são concentradas entre os meses de abril e maio, porém neste caso a concentração de chuva se atrasa um pouco mais para o início de maio. Nesta área também podemos observar a atuação dos dois sistemas principais de formação de chuva no estado. Pois os dois primeiros grupos possuem estações com dois picos de chuva, característica que é explicada devido a atuação dos dois sistemas (ZCIT e POA's).

Por fim, o grupo 4 é formado por 18 estações, que é composto por municípios localizados na região leste do estado. É a porção do estado com precipitações mais bem distribuídas e as chuvas concentram-se em meados de maio. Os sistemas atuantes na região são as POA's e a ZCIT.

4. CONCLUSÕES

Podemos concluir que o estado do Rio Grande do Norte pode ser caracterizado por 4 (quatro) regiões homogêneas. Apesar dos dois agrupamentos obtidos pela técnica de Ward apresentarem resultados aceitáveis, o agrupamento com o uso das variáveis: precipitação média anual, PCD e PCP (agrupamento 1) foi considerado mais adequado que o agrupamento com o uso das doze precipitações médias mensais como variáveis (agrupamento 2); pois o agrupamento 1 foi capaz de reunir em um mesmo grupo as estações com padrões de precipitação mais semelhantes.

A técnica hierárquica, método de Ward, e a não hierárquica, método de k-means, apresentaram resultados semelhantes, com apenas 5 estações diferindo e os dois resultados podem ser utilizados como resultado final do agrupamento. Porém, é importante que ao realizar agrupamentos sejam utilizados mais de um método para que possa ser feita a comparação de qual reflete melhor as características da área em estudo.

Os quatro grupos homogêneos identificados separaram regiões com as seguintes características. O grupo identificado na região leste do estado é caracterizado como a região de maior precipitação anual, melhor distribuição desta precipitação e concentração de mesma no mês de maio, com a identificação de dois picos de precipitação em função da atuação dos sistemas meteorológicos POA's e a ZCIT. Se dirigindo para o oeste do estado foi identificada uma região com período de maior precipitação adiantado para o início de maio, menor precipitação anual e uma maior concentração da precipitação, em função do efeito do sistema ZCIT e enfraquecimento das POA´s. Na região central do estado o grupo identificado é caracterizado por apresentar uma elevada concentração da precipitação com o adiantamento das chuvas para o período do mês de abril, além de ser a região com menor precipitação anual, nela a atuação das POA's está bastante enfraquecida, tendo assim como sistema meteorológico atuante com maior intensidade a ZCIT. Na região oeste do estado as chuvas são adiantadas para o final de março, a região apresenta a pior distribuição da precipitação no estado em função da atuação mais forte apenas da ZCIT para a formação das precipitações, as precipitações anuais nesta região são mais elevadas do que na região central.

AGRADECIMENTOS

Os autores agradecem o apoio da CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - pela concessão da bolsa de Mestrado da primeira autora. Agradecemos também à Empresa de Pesquisa Agropecuária do Rio Grande do Norte (EMPARN) pelo fornecimento dos dados de precipitação.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AGRITEMPO, Sistema de Monitoramento Agrometeorológico. *Mapa de Altimetria do Rio Grande do Norte.* Disponível em: http://www.agritempo.gov.br/altimetria/RN.html A acesso em: 11 ago. 2011.

ANDERSON, B. T.; WANG, J.; GOPAL, S.; SALVUCCI, G. Influence of Daily Rainfall Characteristics on Regional Summertime Precipitation over the Southwestern United States. *American Meteorological Society,* 10: 1218-1230, 2009.

ANDRÉ, R. G. B.; MARQUES, V. da S.; PINHEIRO, F. M. A. P.; FERRAUDO, A. S. Identificação de regiões pluviometricamente homogêneas no estado do Rio de Janeiro, utilizando-se valores mensais. *Revista Brasileira de Meteorologia,* v.23, n.4, p. 501-509, 2008.

BECKER, C. T. et al. Regionalização da precipitação e temperatura no Estado do Rio Grande do Sul a partir da análise de agrupamento. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 7., 1992, São Paulo. *Anais...* São Paulo: SBMet, 1992. p. 225 - 229.

BROCK, G.; PIHUR, V.; DATTA, S; DATTA, S. (2008) clValid: An R Package for Cluster Validation Journal of Statistical Software 25(4) http://www.jstatsoft.org/v25/i04

CARDIM, M. Mapeamento do comportamento multivariado das principais variáveis climáticas de interesse agrícola do Estado de São Paulo. 2001. 124 f. ***Tese (Doutorado).*** Universidade Estadual Paulista, Faculdade de Ciências Agronômicas, Botucatu, 2001.

CARDOSO JUNIOR, M. M.; MADEIRA JUNIOR, A. G. Uso de técnicas de geração de agrupamentos para determinação no número de classes dos portos que movimentam granéis sólidos. In: XLII SBPO, 2010, Bento Gonçalves. *Anais...* Bento Gonçalves, 2010.

EVERITT B. S.; LANDAU S. E.; MORVEN, L. *Cluster Analysis.* Vol. 4, Arnold, London, 2001.

FECHINE, J. A. L.; GALVÍNCIO, J. D. Agrupamento da precipitação mensal da bacia hidrográfica do rio Brígida-PE, através da multivariada. *Revista Brasileira de Geografia Física.* Recife-PE. 1: 39-46, 2008.

GOOVAERTS, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology,* 228: 113–129, 2000.

GORDON, N. D.; NORRIS, J. R..Cluster analysis of midlatitude oceanic cloud regimes: mean properties and temperature sensitivity. *Atmos. Chem. Phys.,* 10: 6435–6459, 2010.

HAIR Jr., J.F. et al. *Análise Multivariada de Dados.* 5. ed. Porto Alegre: Bookman, 2005.

HASSAN, B. G. H.; PING, F. Formation of Homogenous Regions for Luanhe Basin – by Using L-Moments and Cluster Techniques. *International Journal of Environmental Science and Development*, 3: 205-210, 2012.

HERRERA, O. M. et al. Agrupamento de estações climatológicas localizadas no estado de São Paulo, Brasil, utilizando-se análise multivariada: comparação com classificação de Koppen. In: CONGRESSO ARGENTINO DE METEOROLOGIA, 7., 1996, Buenos Aires CONGRESSO LATINAMERICANO E IBÉRICO DE METEOROLOGIA, 7., 1996, Buenos Aires. *Anais...* Buenos Aires: Biblioteca Nacional, 1996. p. 233-234.

IDEMA, Instituto de Desenvolvimento Sustentável e Meio Ambiente. *Anuário Estatístico 2010.* Disponível em: http://www.idema.rn.gov.br/contentproducao/aplicacao/idema/socio_economicos/arquivos/Anuario-CDROM%202010/index.htm Acesso em: 11 ago. 2011.

LI, X.; JIANG, F.; LI, L. WANG, G. Spatial and temporal variability of precipitation concentration index, concentration degree and concentration period in Xinjiang, China. *Int. J. Climatol.* 31: 1679–1693, 2011.

LYRA, G. B.; GARCIA, B. I. L.; PIEDADE, S. M. S. SEDIYAMA, G. C.; SENTELHAS, P. C.. Regiões homogêneas e funções de distribuição de probabilidade da precipitação pluvial no Estado de Táchira, Venezuela. *Pesq. agropec. bras.,* Brasília, v.41, n.2, p.205-215, 2006.

MUÑOZ-DÍAZ, D.; F. S. RODRIGO, F. S.. Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and principal component analysis: comparison. *Annales Geophysicae,* 22: 1435–1448, 2004.

NERY, J. T.; VARGAS, W. M.; MARTINS, M. L. O. F. Caracterização da precipitação no estado do Paraná. *Revista Brasileira de Agrometeorologia,* v.4, n.2, p.81-89, 1996.

OMAR, A. H.; WON, J-G.; WINKER, D. M.; YOON, S-C.; DUBOVIK, O.; McCORMICK, M. P.. Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements. *Journal of geophysical research,* 110: 148-227, 2005.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

RAMOS, M. C. Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region*. Atmospheric Research*, 57: 123–138, 2001.

RAZIEI, T. at al. A precipitation-based regionalization for Western Iran and regional drought variability.Hydrolgy and Earth System Sciences, v.12, p.1309-1321, 2008. Disponívelem: http://www.hydrol-earth-syst-sci.net/12/1309/2008/hess-12-1309-2008.html Acesso em: 16 ago. 2012.

SHAFER, C. M.; DOSWELL III, C. A.: A multivariate index for ranking and classifying severe weather outbreaks. *Electronic J. Severe Storms Meteor.,* 5 (1), 1–39, 2010.

SILVA, E. S.; TRINCA, L. A.; NERY, J. T.. Áreas espacialmente homogêneas de níveis de precipitação nas regiões Sudeste e Sul do Brasil. *Revista Energia na Agricultura***,** v. 24, n.3, p. 34-59, 2009.

SILVA, F. M.; LERMEN, V. K.; NERY, J. T.. Variabilidade interanual da precipitação na bacia do rio Iguaçu. *Acta Scientiarum*, Maringá, v. 23, n. 6, p. 1439-1444, 2001.

UNAL, Y.; KINDAP, T.; KARACA, M. Redefining the climate zones of Turkey using cluster analysis. *Int. J. Climatol.* 23: 1045–1055, 2003.

WARD, J.H. Hierarchical grouping to otimize an objective function. *Journal American Association*, 58: 236 –244 p, 1963.

WILKS, D. S. Statistical Methods in the Atmospheric Sciences. *International Geophysics Series.* v. 91, Elsevier, 627 p. 2006.

XIE, Z. Q.; DU, Y.; JIANG, A. J.; DING, Y. G. Climatic trends of different intensity heavy precipitation events concentration in China. *Journal of Geographical Sciences*, 15: 459–465, 2005.

ZHANG, L. J.; QIAN, Y. F. A study on the feature of precipitation concentration and its relation to flood-producing in the Yangtze River Valley of China. *Chinese journal of geophysics. 47: 709-718, 2003**.***

Texto submetido à RBClima na data de 09/12/2014