APLICAÇÃO DA ANÁLISE FATORIAL EM COMPONENTES PRINCIPAIS A DADOS DE PRECIPITAÇÃO NO ESTADO DO AMAPÁ

Jonathan Castro Amanajás[1], Célia Campos Braga[2], Francisco José Lopes de Lima[1], Roni Valter de Souza Guedes[1]

RESUMO: Usando o método de Análise Fatorial (AF) em Componentes Principais (CPs), algumas características do padrão de precipitação pluvial (mm) são estabelecidas para o Estado do Amapá (Amazônia Oriental). Foram usados dados de precipitação pluvial mensal de 27 estações meteorológicas. As duas primeiras Componentes Principais (CPs), que descrevem 85,74% da variância total da precipitação observada, foram consideradas. Mostrou-se, graficamente e numericamente, que a primeira CP está relacionada com a atuação da Zona de Convergência Intertropical e a segunda está relacionada a tipos de tempo não-gradiente (não tem padrão bem definido ou associado a nenhum sistema de escala sinótica). Usando as cargas (escores) das CPs, três setores de precipitação homogênea foram determinados e comparados com a climatologia local. Finalmente, alguns detalhes do padrão de precipitação anual para estes setores são discutidos.

PALAVRAS-CHAVE: padrão de precipitação; rotação; grupos homogêneos

ABSTRACT: Using the of Factor Analysis (FA) in Principal Components (PCs) method, some characteristics of the precipitation pattern (mm) are established for the Amapá State (Amazônia Oriental). Data of monthly precipitation from 27 meteorological stations are used. The first two principal components (PCs), which describe 85.74% of the total precipitation field variance, are considered. It is shown, graphically and numerically, that the first PC is related to the performance of the Intertropical Convergence Zone and the second is related to non-gradient weather type (no clear pattern or associated with any system of synoptic scale). Using the loadings (scores) of the PCs, three homogeneous precipitation sectors are determined and compared with the local climatology. Finally, some details of the annual precipitation pattern for these sectors are discussed.

KEY WORDS: precipitation pattern; rotation; homogeneous groups

1. INTRODUÇÃO

O Estado do Amapá, localizado na faixa equatorial do globo, esta inserido no setor norte da Amazônia Oriental. Esta região caracteriza-se por altas temperaturas e elevados índices pluviométricos, tendo seu período chuvoso médio centrado de dezembro a maio. As chuvas nesta região são diretamente influenciadas pela migração sazonal da Zona de

[1] Pós-graduandos do Departamento de Ciências Atmosféricas, UFCG. Av. Aprígio Veloso, 882, Bairro Universitário. Campina Grande – PB, Brasil. CEP: 58.429-140. E-mail: jonathan.amanajas@hotmail.com; thescolopes@yahoo.com.br; roniguedes84@yahoo.com.br

[2] Professora do Departamento de Ciências Atmosféricas, UFCG. Av. Aprígio Veloso, 882, Bairro Universitário. Campina Grande – PB, Brasil. CEP: 58.429-140. E-mail: celia@dca.ufcg.edu.br

Convergência Intertropical (ZCIT) na direção meridional, caracterizada pela elevação do ar quente e úmido, formando um cinturão de nuvens e chuva convectiva (FIGUEROA & NOBRE, 1990; RAO & HADA, 1990).

Além da ZCIT outros mecanismos explicam o regime pluvial na região, os quais são resultantes da combinação das brisas marítimas e da fonte de vapor provinente da Floresta Amazônica (VIANELLO, 1991). Por outro lado, essa migração sazonal da ZCIT é fortemente relacionada as anomalias que ocorrem na circulação atmosférica e oceânica sobre os trópicos, a exemplo dos fenômenos El Niño e La Niña (KAYANO et. al., 1988; PEZZI & CAVALCANTI, 2001).

As altas temperaturas reinantes estão associadas à intensa radiação solar incidente na região, embora grande parte da energia seja convertida em calor latente de evaporação, uma vez que a região é, no globo, uma das principais fontes continentais de vapor d'água que será, em grande parte, advectado para as latitudes mais elevadas, sendo outra parte levada para a atmosfera superior e liberada na forma de calor sensível (VIANELLO, 1991).

O clima da região segundo a classificação de Köppen é do tipo megatérmico Af, caracterizando-a como clima tropical úmido, com chuvas em todas as estações e temperatura média no mês mais frio acima de 18°C, e período menos chuvoso de curta duração (NIMER, 1989; PEREIRA et al. 2002).

Considerando a necessidade de fazer uma avaliação mais consistente do comportamento espacial e temporal da precipitação, este estudo tem como objetivo utilizar técnicas estatísticas apropriadas que possibilite identificar a atuação dos sistemas atmosféricos que influenciam a precipitação da região. Para tal se utilizou as técnicas da Análise dos Componentes Principais e agrupamentos as quais fornecem informações bastante confiáveis sobre a variabilidade da variável meteorológica em questão.

2. MATERIAL E MÉTODOS

Os dados de precipitação utilizados neste estudo são provenientes de 27 estações meteorológicas da Agência Nacional das Águas (ANA), a partir dos quais foram obtidas médias mensais e anuais de precipitação para o Estado do Amapá. A distribuição das estações é indicada por números na Fig. 1.

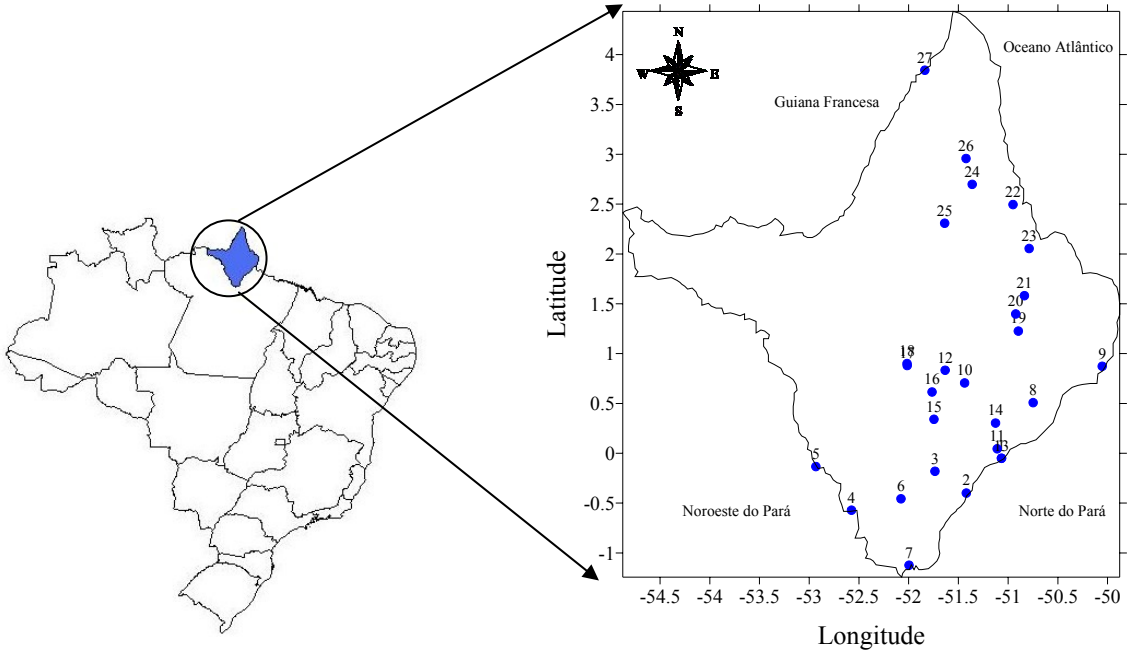

Figura 1. Mapa do Brasil destacando a distribuição espacial das estações meteorológicas da ANA no Estado do Amapá.

Neste estudo se utilizou fundamentalmente as técnicas de Análise Multivariada. Adotou-se a Análise Fatorial (AF) em Componentes Principais (CP) e Análise de Agrupamento nos modos espacial e temporal para definir os principais sistemas contribuintes de variabilidade interanual da precipitação pluvial e suas regiões homogêneas no Estado do Amapá, respectivamente.

2.1. Análise em Componentes Principais

A técnica de análise multivariada denominada Análise em Componentes Principais (ACP) foi introduzida pela primeira vez em estudos meteorológicos por Lorens, em 1956, que as denominou de Funções Ortogonais Empíricas (FOE) para destacar sua natureza não analítica (BRAGA, 2000). Este método tem por objetivo transformar p variáveis originais correlacionadas em componentes não correlacionadas ou ortogonais, sendo estes componentes funções lineares, em que p variáveis originais são medidos em n indivíduos (BOUROCHE & SAPORTA, 1982).

Empregou-se ACP porque essa técnica não só permite uma redução da dimensão da matriz de dados, ao mesmo que a máxima variância pode ser explicada pela classificação dos autovetores associados aos maiores autovalores da matriz de correlação, ou seja, os dados originais podem ser analisados a partir de um número pequeno de componentes independentes entre si (PREISENDORFER, 1988).

A coleta básica de informações na análise fatorial é a matriz de dados (PANDZIC, 1988). Em *n* observações existem *m* variáveis, então a matriz de dados normalizada (com média zero e variância um) de precipitação pluvial pode ser apresentada como *m* x *n*, e designada por ***Z***, a partir da qual se obtém a matriz de correlação ***R*** dada por:

$$R = \frac{1}{n-1}(Z)(Z)^t$$

onde $(Z)^t$ é a matriz transposta de *Z*.

Sendo ***R*** é uma matriz simétrica positiva de dimensão (k x k), ela é diagonalizável por uma matriz ***A***, de mudança de base, denominada de autovetores. A matriz diagonal ***D*** cujos elementos diagonais são os autovalores (λ) de ***R***, é expressa por:

$$D = A^{-1}RA$$

Pela ortogonalidade dos autovetores, a inversa de ***A*** ($\boldsymbol{A^{-1}}$) é igual a sua transposta ($\boldsymbol{A^t}$). Logo, as CPs Z_1, Z_2, ..., Z_n são obtidas por combinações lineares entre a transposta dos autovetores ($\boldsymbol{A^t}$) e a matriz de observações (***Y***), ou seja:

$$Z = A^tY$$

$$Y = AZ$$

Cada linha de ***Z*** corresponde a uma CP que forma as séries temporais associadas aos autovalores. Os valores de Y do n-ésimo local podem ser calculados por:

$$Y_i = a_{j1}Z_1 + a_{j2}Z_2 + ... + a_{jk}Z_k + a_{np}Z_p$$

A solução desta equação é única. Esta solução considera a variação total presente no conjunto das variáveis iniciais, onde a CP1 explica o máximo possível da variância dos dados iniciais, enquanto a CP2 explica o máximo possível da variância ainda não explicada e assim por diante, até a última CPm que contribui com a menor parcela de explicação da variância total dos dados iniciais.

No caso deste estudo cada CP tem uma porção da variância total dos dados mensais da precipitação pluvial, e são ordenadas por ordem decrescente dos autovalores mais significativos de a_1 em ***A***, dada por:

$$Z_i = \sum_{j=1}^{n} a_{j,i}Y_i$$

A variância total do sistema (***V***) é definida como a soma das variâncias das variáveis observadas, assim ***V*** é dada por:

$$V = traço\ S = \sum_{i=1}^{p} S_{ii} = \sum_{i=1}^{p} \lambda_i$$

em que S é a variância das variáveis observadas e λ_i são os autovalores. O traço da matriz pode ser entendido também como o somatório da diagonal principal da matriz de correlação.

A variância explicada por cada componente é:

$$\alpha_i = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \cdot 100\ (\%)$$

O número escolhido de CPs foi baseado no critério de truncamento de Kaiser, que considera como mais significativos os autovalores cujos valores sejam superiores a unidade (GARAYALDE et al., 1986).

A distribuição espacial dos fatores comuns temporais facilita as interpretações físicas correspondentes às CPs. Se a interpretação física for o principal objetivo da ACP, então se faz a rotação da solução inicial para uma nova coordenada de vetores, através de uma transformação linear (FULGOSI, 1979; RICHMAN, 1986; WIKS, 2006).

Neste trabalho, utilizou-se uma rotação ortogonal mantendo a correlação entre as componentes igual a zero, portanto independentes. Um dos métodos de rotação mais utilizados que maximiza a variância em cada componente é o VARIMAX, dado por:

$$Var = \sum_{j=1}^{r} \left(m \sum_{i=1}^{m} (a_{ij})^2 - \left(\sum_{i=1}^{m} a_{ij}^2 \right)^2 \right) \Big/ m^2$$

onde *r* é o número de CPs que representam as informações físicas e *m* o número de variáveis.

O principal objetivo da utilização de CPs rotacionadas é se obter a maior concentração dos dados originais da série temporal nas CPs pela máxima variância, e não à característica normalizada da série temporal como ocorre nas CPs não rotacionadas (CORRAR et. al., 2007).

2.2. Análise de Agrupamentos (Clusters)

Esta técnica estatística de analise multivariada tem como objetivo principal classificar os indivíduos de uma população que são conhecidos por suas características, em grupos que sejam homogêneos intra-grupos e heterogêneos intergrupos.

Aqui, a técnica de Análise de Agrupamento escolhida foi a hierárquica, em que são produzidas sequências de partições em classes cada vez mais vastas, e a união de dois grupos numa determinada etapa, produz um dos agrupamentos da etapa superior o que caracteriza o processo hierárquico.

A classificação dos indivíduos em grupos distintos é feita a partir de uma função de agrupamento denominada distância ou similaridade e de um critério matemático de agrupamento (BOUROCHE & SAPORTA, 1982).

Na análise de agrupamento, todos os processos de hierarquização são similares, iniciando-se pela determinação da função de agrupamento. De acordo com Wilks (2006) esta função é usada como critério para medir a distância entre dois pontos x_i e x_j ou para estabelecer o quanto eles são parecidos. Aqui, para medir a dissimilaridade entre dois indivíduos x_i e x_j utilizou-se a distância euclidiana, expressada por:

$$d(x_i, x_j) = \left| x_i - x_j \right| = \left[\sum_{k=1}^{n} P_i (x_{i,k} - x_{j,k})^2 \right]^{1/2}$$

onde, $x_{j,k}$ representa os valores mensais dos locais x_i e x_j e P_i é o peso associado a cada indivíduo.

Tanto a distância euclidiana quanto o coeficiente de correlação podem ser usados na determinação dos grupos. Para este fim existem dois métodos de agrupamento, os hierárquicos e os não hierárquicos. Nos métodos hierárquicos vários critérios de agrupamentos são possíveis, dentre eles optou-se pelo método de Ward (1963).

O método de Ward procura por partições que minimizem a perda associada a cada agrupamento (EVERITT, 1974). Essa perda é quantificada pela diferença entre a soma dos erros quadráticos de cada padrão e a média da partição em que está contido. 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.

Nos métodos hierárquicos, o dendrograma que é a representação gráfica de uma árvore de classificação, representa a ordem em que os dados foram agrupados. A obtenção dos grupos pode ser feita pelo dendrograma a partir do critério da inércia (BOUROCHE & SAPORTA, 1982) e conhecimento prévio da região em estudo.

3. RESULTADOS E DISCUSSÃO

A seguir são apresentados os resultados obtidos da aplicação do método da ACP e agrupamentos aos dados de precipitação pluvial no Estado do Amapá. A Fig. 2 mostra a variabilidade espacial dos totais médios anuais da precipitação no Estado.

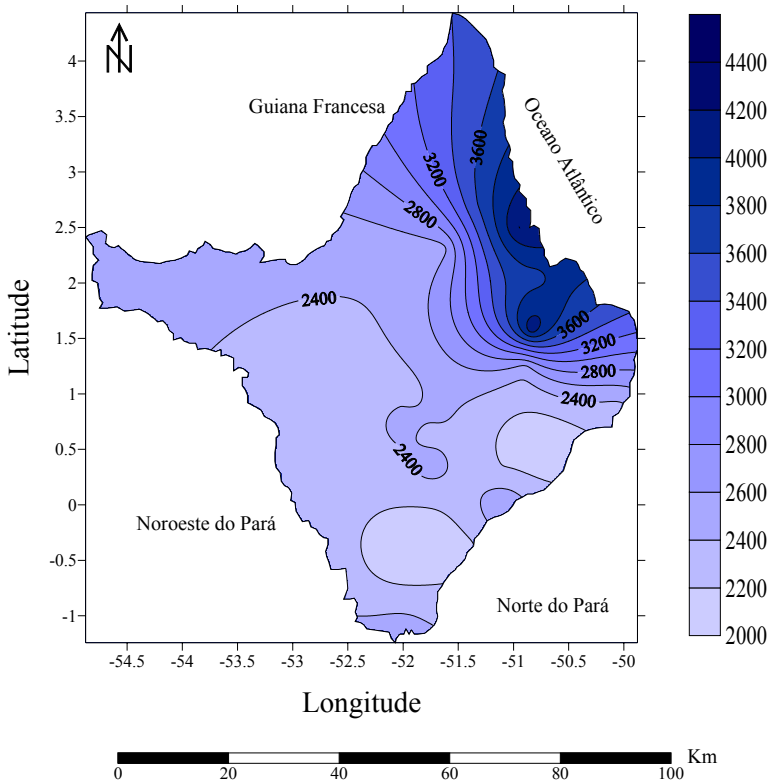

Figura 2. Espacialização dos totais médios anuais da precipitação (mm) no estado do Amapá.

A associação entre as variações espaciais da precipitação no Estado identificam a existência de, pelo menos, dois regimes pluviométricos. Observa-se que os maiores totais pluviométricos ocorreram no setor norte-nordeste do estado, declinando do setor costeiro para o interior. Os menores valores foram encontrados na região sudoeste. Essa distribuição caracteriza-se pela atuação de fenômenos meteorológicos que atuam na Amazônia Oriental ao longo do ano, os quais se destacam a ZCIT e Linhas de Instabilidade (COHEN, 1995).

A análise fatorial em componentes principais rotacionada, aplicada aos dados médios mensais da precipitação no Estado produziram padrões espaciais associados a cada fator comum (pesos). A Tab. 1 mostra o peso de cada variável, representada pelos meses do ano, em relação ao conjunto de dados médios mensais da precipitação pluvial. Foram considerados os dois primeiros fatores comuns rotacionados segundo o critério de truncamento de Kaiser.

Tabela 1. Sequência dos autovalores na ordem decrescente e a contribuição (%) à variância total dos dados médios mensais rotacionados da precipitação pluvial no Estado do Amapá.

Fator	Autovalor	Variância Explicada (%)	Variância Acumulada (%)
1	6,41	53,44	53,44
2	3,88	32,30	85,74
□	□	□	□
12	0	0	100

Os dois primeiros fatores comuns rotacionados, que explicam 85,74% da variância total da série, são mostrados na Fig. 3. O primeiro fator comum temporal rotacionado da precipitação, explica 53,44% da variância total, com correlações positivas, superiores a 0,8, com ciclo anual bem definido predominando nos meses de dezembro a junho, e com valor mínimo em setembro. O segundo fator comum apresenta máximas correlações positivas de agosto a novembro (superiores a 0,8) e mínimas entre fevereiro e março. Este fator, explica 32,3% da variância total e tem ciclo anual bem definido.

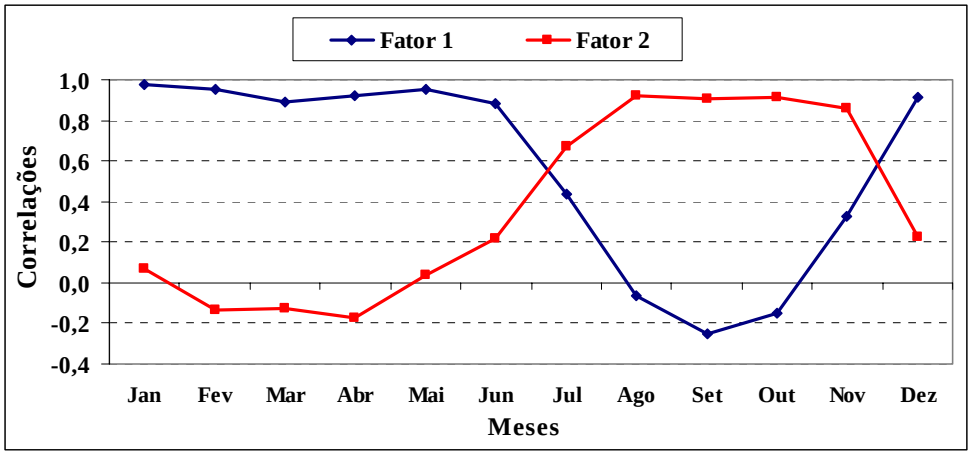

Figura 3. Cargas fatoriais rotacionadas (correlações) para os dois fatores comuns de precipitação (mm) que explicam 85,74% do total (53,44%+32,3%) no estado do Amapá.

O padrão espacial associado ao primeiro fator (Fig. 4a) mostra contribuições (escores) baixas e negativas inferiores a 0,4 em todo sudeste, sul e sudoeste do Estado. Os valores positivos aumentam gradativamente na direção da Mesorregião do Norte do Amapá, atingindo seus máximos superiores a 2,0 nas proximidades do setor costeiro atlântico. A distribuição espacial deste fator, com as maiores correlações de dezembro a junho, está em fase com as maiores precipitações nesses meses na região leste. No restante da região os totais pluviométricos são menores (correlações negativas de agosto a outubro). Possivelmente as chuvas associadas a este padrão estejam relacionadas com a atuação e deslocamento norte-sul da ZCIT, principal sistema gerador de chuva na região.

O segundo fator (Fig. 4b) apresenta padrões espaciais com contribuições positivas significativas (escores maiores que 0,8) sobre o sudoeste, e valores negativos altos (maiores que -1,0) no leste do Estado. Este fator pode estar associado a tipos de tempo não-gradiente, ou seja, não tem padrão bem definido ou associado a nenhum sistema de escala sinótica. A distribuição espacial deste fator, com as maiores correlações de agosto a novembro, está em fase com as menores precipitações nesses meses na região leste.

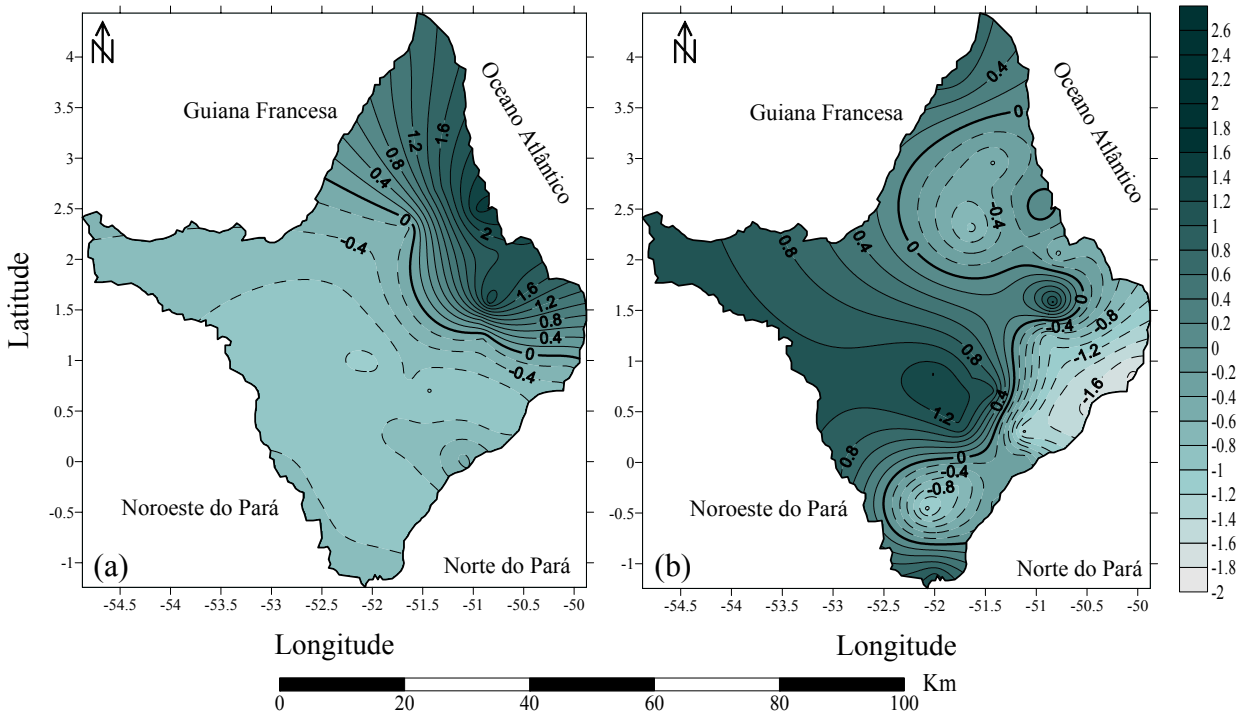

Figura 4. Padrões espaciais para a primeira (a) e segunda (b) componentes principais associadas aos dois primeiros fatores comuns dos totais médios mensais da precipitação pluvial (mm) no Estado do Amapá.

Para a determinação dos grupos de precipitação homogênea adotou-se a técnica aglomerativa hierárquica, baseada no critério de inércia intra-grupos. Através da curva de inércia (Fig. 5) têm-se uma análise subjetiva quanto ao número de grupos de precipitação homogênea existentes na área de estudo, em que se sugere número de grupos igual a três.

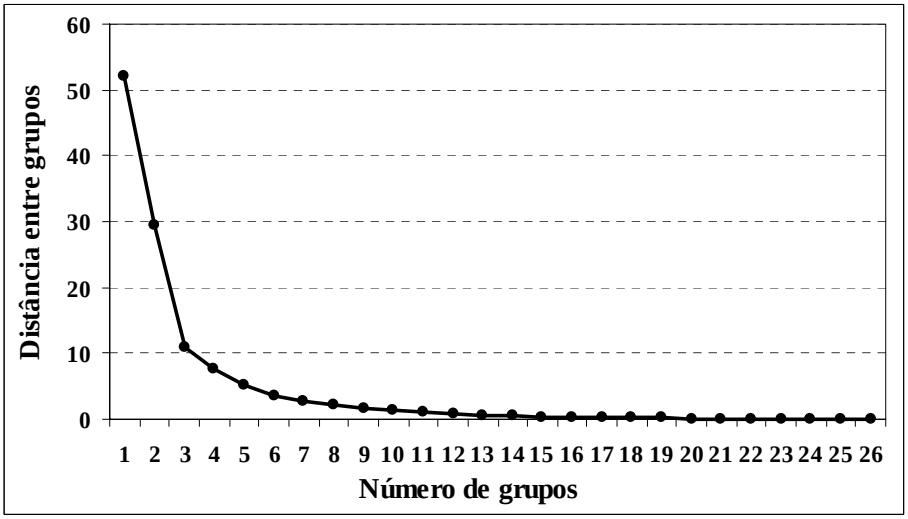

Figura 5. Curva de Inércia intra-grupos.

A sequência sucessiva de agrupamentos encontra-se sob a forma de dendrograma, em que na abscissa estão as estações e na ordenada os níveis de agregação (distâncias) entre os grupos (Fig. 6). O número objetivo de grupos homogêneos foi obtido através de cortes transversais no dendrograma, efetuado de acordo com o critério matemático de inércia e conhecimento prévio da climatologia da região. Usando este critério chegou-se a um dendrograma resultante do agrupamento das componentes da precipitação das 27 estações analisadas, em que o corte foi feito na distância 10,8 o que corresponde a três grupos de precipitação homogênea.

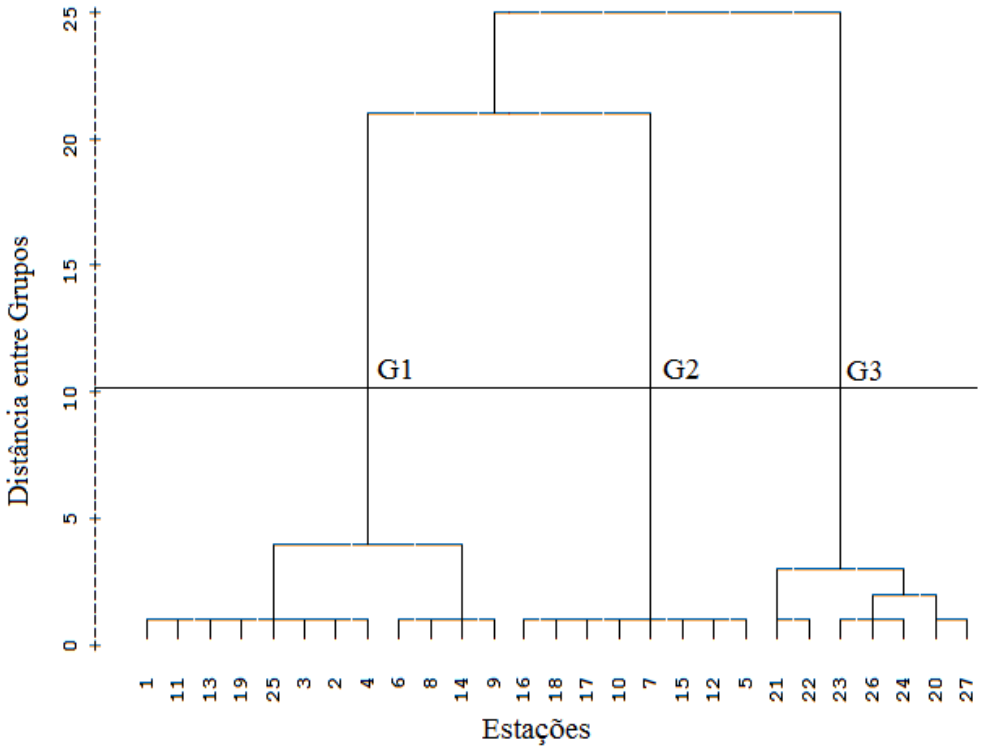

Figura 6. Dendrograma resultante da análise objetiva das 27 estações pluviométricas a partir dos fatores comuns temporais mais significativos usando o método aglomerativo hierárquico de Ward.

Considerando-se tais procedimentos, a região em estudo pode ser classificada em três (03) sub-regiões pluviometricamente homogêneas em relação a variabilidade espacial da precipitação. A delimitação das regiões pluviometricamente homogêneas estão ilustradas na Fig. 7. Suas principais características pluviométricas estão sintetizadas na Tab. 2 e descrição mais detalhadas a seguir.

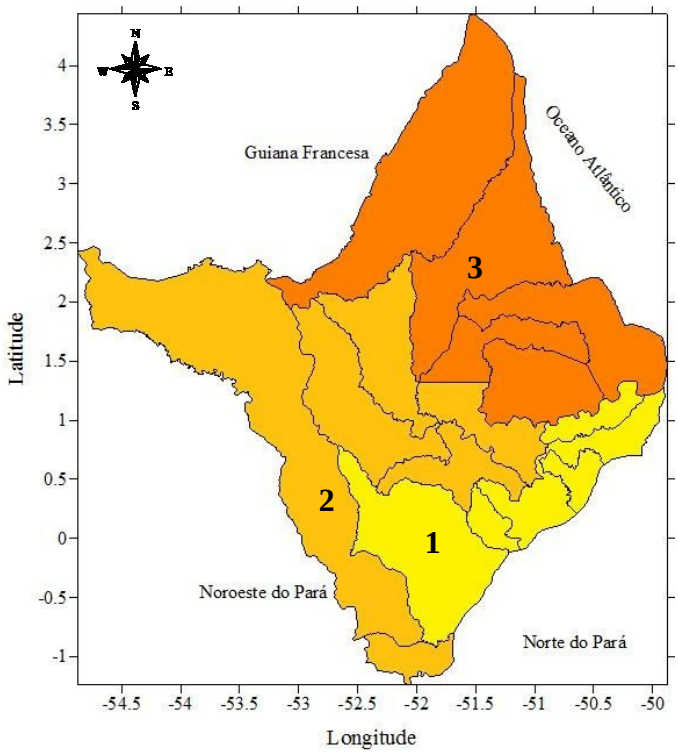

Figura 8. Regiões homogêneas em relação à variabilidade espacial da precipitação, obtidas a partir de uma análise de agrupamentos dos seus dois fatores comuns temporais para o Estado do Amapá.

Tabela 2. Síntese das principais características dos três (03) grupos de precipitações homogêneas no Estado do Amapá.

Grupo	Área (Km²)	Localização	Precipitação (mm)		
			Média	Máxima	Mínima
1	24.935	Macapá, Cutias, Itaubal, Santana e Mazagão	2277,8	359,8 (mar)	26,5 (out)
2	50.655	Porto Grande, Pedra Branca do Amapari, Serra do Navio, Ferreira Gomes, Laranjal do Jari e Vitória do Jari	2323,6	324,1 (abr)	66,3 (out)
3	67.227	Mesorregião do Norte do Amapá	2655,0	535,2 (abr)	35,5 (out)

Os grupos 1 e 2 correspondem juntos a aproximadamente 53% da área de estudo. Situam-se geograficamente na Mesorregião do Sul do Amapá. O grupo 3 esta localizado na Mesorregião do Norte do Amapá e compreende 47% da área do estado. Numa solução mais geral, as três regiões (grupos) homogêneas de precipitação pluvial poderiam ser fundidas em duas grandes regiões com características físicas semelhantes.

4. CONCLUSÕES

A Análise Fatorial em Componentes Principais das séries temporais de precipitação permitiu extrair informações importantes sobre o regime de chuvas da área em estudo. A representação espacial dos padrões pluviométricos possibilitou uma melhor compreensão dos sistemas atmosféricos responsáveis pela variabilidade sazonal desta variável.

O peso do primeiro fator comum mostrou que as maiores contribuições ocorreram na parte norte do Estado (Grupo 3) e estão associados a atuação da ZCIT. O peso associado ao segundo fator apresentou maiores contribuições na Mesorregião do Sul do Amapá (Grupos 1 e 2) e esta associado ao período menos chuvoso da região.

A classificação da precipitação pluvial usando o método aglomerativo de Ward segundo o critério da inércia possibilitou a obtenção de três sub-regiões (grupos) homogêneas no Estado do Amapá. Todos os grupos obtidos estão geograficamente compactados. Dessa forma, uma divisão da área em setores considerados homogêneos em relação à partida de precipitação média anual foi alcançada.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BOUROCHE, J- M; SAPORTA, G. **Análise de Dados**. Rio de Janeiro, Zahar, 1982. 117p.

BRAGA, C. C. **Inter-relação entre padrões de índice de vegetação e de pluviometria no Nordeste do Brasil.** Campina Grande, 2000. 129p. Tese (Doutorado em Recursos Naturais). Universidade Federal da Paraíba.

COHEN, J. C. P.; SILVA DIAS, M. A. F; NOBRE C. A. **Environmental conditions associated with Amazonian squall lines: A case study**. Monthy Weather Review, Boston, 123:3163-3174, 1995.

CORRAR, L. J; PAULO, E. e DIAS FILHO, J. M. **Análise Multivariada**. São Paulo, Atlas, 2007, 541p.

EVERITT, B.S. **Cluster analysis**. London, Heinemann Educational Books Ltd., 1974. 121p.

FIGUEROA, S.N.; NOBRE, C. A. **Precipitation distribution over Central and Western tropical South America**. Climanálise, São José dos Campos, 5:36-44, 1990.

FULGOSI, A. **Faktorska analisa.** Zagreb, Skolska knjiga, 1979. 367p.

GARAYALDE, E.J.G.; da SILVA, M.G.A.; TAVARES, A. de SÁ. **Classificação meso-climática da região sul do Brasil pela análise de Componentes Principais**. In: CONGRESSO INTERAMERICANO DE METEOROLOGIA, 1., Brasília, 1986. Anais. Brasília, Sociedade Brasileira de Meteorologia, 1986. p.119-124.

KAYANO, M. T.; RAO, V. B.; MOURA, A. D. **Tropical circulation and the associated rainfall anomalies during 1983-1984**. J. Climatol., 8:477-488, 1988.

NIMER, E. **Climatologia do Brasil.** 2. ed. Rio de Janeiro, IBGE, 1989. 422p.

PANDZIC, K. **Principal component analysis of precipitation in the Adriatic-Pannonian área of Yugoslavia.** J. Climatol., 8:357-370, 1988.

PEZZI, L. P.; CAVALCANTI, I. F. A. **The relative importance of ENSO and tropical Atlantic sea surface temperature anomalies for seasonal precipitation over South America: a numerical study.** Climate Dynamics, 17:205-212, 2001.

PEREIRA, A. R.; ANGELOCCI, L. R.; SENTELHAS, P. C. **Agrometeorologia: fundamentos e aplicações práticas**. Guaíba, Agropecuária, 2002.

PREISENDORFER, R. **Principal component analysis in meteorology and oceanography.** Amsterdan, Elsevier, 1988. 425p.

RAO, V.B.; HADA, K. **Characteristics of rainfall over Brazil: annual variations and connections with Southern oscillations.** Theoretical and Applied Climatology, 42:81-91, 1990.

RICHMAN, M. B. **Review article on rotation of principal components.** J. Climatol., 6:293-335, 1986.

VIANELLO, R. L. **Meteorologia Básica e Aplicações**. Viçosa, UFG, Impr. Univ., 1991.

WARD, J. H. **Hierarchical grouping of otimize an objetive function.** J. American Statistical Association, 58:236-244, 1963.

WILKS, D. S. **Statistical Methods in the Atmospheric Sciences.** 2. ed. London, Academic Press, 2006. 649p.