

APLICAÇÃO DOS COMPONENTES PRINCIPAIS NA SIMULAÇÃO E CONSISTÊNCIA DE SÉRIES TEMPORAIS

*Célia Campos Braga¹, Maria José Herculano Macedo²,
Bernardo Barbosa da Silva¹, Ramon Campos Braga³*

¹ Meteorologistas, Professores Associados da Unidade Acadêmica de Ciências Atmosféricas, UFCG/Campina Grande – PB, email: celia@dca.ufcg.edu.br, bernardo@dca.ufcg.edu.br.

² Licenciada em Matemática, M.Sc., Pós-Graduada, Unidade Acadêmica de Ciências Atmosféricas, UFCG/Campina Grande – PB, email: mariejhm@hotmail.com.

³ Divisão de Satélite e Sistemas Ambientais – INPE – SP, email: ramon.braga@cptec.inpe.br

Artigo recebido em 02/07/2011 e aceito em 22/08/2011

RESUMO

Empregou-se a Análise das Componentes Principais - ACP para estimar séries temporais de precipitação pluviométricas no Estado do Rio Grande do Norte. Foram utilizados dados mensais de 67 estações meteorológicas no período de 1963 a 2007. Com o emprego do método de classificação de grupos K-means identificou quatro (4) regiões com características próprias e distintas da precipitação: Agreste (G1) e Litoral (G2), com chuvas associadas aos sistemas de leste; Alto Apodi e Serra Central (G3), que além da atuação da ZCIT existe um efeito orográfico acentuado e Central e Norte (G4), com chuvas influenciadas principalmente pela ZCIT. Selecionaram-se duas regiões com regimes pluviométricos diferentes (litoral e semiárido) para estimar séries temporais de precipitação utilizando o método da ACP. O erro percentual médio na região litorânea foi de 12% e de 19,8% a região semiárida, revelando eficácia na estimativa de dados faltosos, principalmente em regiões com regime pluviométrico mais regular.

Palavras-Chave: Regimes pluviométricos, Nordeste do Brasil, análise multivariada, preenchimento de falha.s

APPLICATION OF PRINCIPAL COMPONENTS IN SIMULATION AND CONSISTENCY OF TIME SERIES

ABSTRACT

The Principal Component Analyse - PCA was applied in the estimation of time series of rainfall in Rio Grande do Norte State, Brazil . Monthly data from 67 meteorological stations in the period 1963 to 2007 were used in the study. According to the technique of K-means classification four distinct groups according to rainfall characteristics were identified: Agreste (G1) and Coastal East (G2) groups, associated with eastern synoptic systems; High Apodi and Sierra Central (G3) group, which apart from the ITCZ influence has a strong orographic effect; and Central and North (G4) group, with rainfall influenced mainly by the ITCZ. Two regions with different rainfall regimes (Coastal East - G2 and High Apodi and Sierra Central - G3) were selected to estimate time series of precipitation using the PCM. The average percentage error in the Coastal East group and High Apodi and Sierra Central group were respectively equal to 12% and 19.8%, revealing that the PCA has high efficiency for missing data assessment, particularly in regions with more regular rainfall.

Keywords: Pluviometric regimes, Northeast of Brazil, multivariate analysis, missing data

INTRODUÇÃO

O clima de uma determinada região é uma resposta aos fenômenos oceânicos-atmosféricos, à orografia, à distribuição dos continentes e oceanos e à latitude, podendo ser modificado no decorrer do tempo através das atividades humanas.

A caracterização climática do Nordeste Brasileiro (NEB) é complexa, apresentando grande variabilidade espacial e temporal da precipitação, elevada taxa de evaporação e temperatura média elevada variando entre 24° e 28°C. Apresenta desde o clima semi-árido no interior da região, até o clima tropical em sua costa leste. O máximo de precipitação no norte do Nordeste, principalmente o estado do Ceará, oeste do Rio Grande do Norte, e interior da Paraíba e Pernambuco, ocorre no período de fevereiro a maio, deve-se ao deslocamento da ZCIT para latitudes mais ao sul, a qual influencia diretamente na qualidade do período chuvoso da região (Uvo, 1989). Enquanto que ao sul-sudeste e oeste da região, a máxima precipitação é influenciada pelas frentes frias oriundas das latitudes médias do hemisfério sul que alcançam latitudes mais baixas nos meses de novembro a fevereiro (Kousky, 1979).

Os máximos de precipitação no outono e inverno, atuante na faixa litorânea leste do Nordeste, desde o Estado do Rio Grande do Norte até a Bahia é

influenciado, pela predominância de movimentos ascendentes ao longo da costa devido à convergência dos ventos alísios de sudeste, que sopram do mar para o continente, brisas marítimas e terrestres, além da ação de frentes frias remanescentes.

O El Niño - Oscilação sul (ENOS) é citado na literatura como um dos principais fenômenos meteorológicos responsáveis por extremos climáticos no NEB (Silva, 2007). O oceano pacífico tropical sofre um aquecimento anômalo de águas geralmente frias a leste deste oceano, em consequência disso provoca mudanças na circulação de grande escala na atmosfera.

O Rio Grande do Norte situa-se no extremo nordeste brasileiro, apresenta uma área total de 52.796,8 km² (0,62% do território brasileiro). Limita-se ao norte e a leste com o oceano atlântico (410 km de extensão de praias), ao sul com o estado da Paraíba e a oeste com o estado do Ceará. Apresenta diversos tipos de climas, Clima Úmido no litoral leste diminuindo no sentido costa-interior, o Sub-úmido no extremo oeste do Estado e nas áreas serranas, o Semi-árido em quase todo interior e parte do litoral norte e o árido na parte central.

A maior parte do estado Rio Grande do Norte é caracterizada por grandes flutuações espaciais e temporais da precipitação, principalmente na região

mais semi-árida. A identificação do potencial hidrológico no Nordeste brasileiro tem sido prejudicada em virtude das interrupções das séries pluviométricas de alguns postos que compõem a rede. Então, para contornar as dificuldades inerentes as séries de dados, é plausível tentar aplicar técnicas, que possibilitasse obter padrões espaciais e temporais da precipitação, assim como fazer preenchimento de dados faltosos de séries pluviométricas. Neste sentido, utilizaram-se as técnicas da análise de agrupamentos e dos Componentes Principais que permitem trabalhar com amostras que envolvem um grande número de variáveis, reduzindo ao máximo o volume total do número de variáveis iniciais, com perda mínima das informações, que são indispensáveis na investigação das variáveis envolvidas no problema em questão. Estas técnicas têm grande vantagem sobre outros estudos realizados na região, porque são métodos capazes de fazer análises mais consistentes e detalhadas da variabilidade espacial e temporal de variáveis meteorológicas (Ceballos & Braga, 1995; Gong & Richman, 1995; Wilks, 2006; Bouroche & Saporta, 1983).

Nos últimos anos inúmeros trabalhos climatológicos têm utilizado as técnicas de análise multivariada em diferentes estudos. Eastman & Fulk (1993) aplicaram a

técnica para obtenção das condições ambientais da África. Gong & Richman (1995) utilizaram vários métodos de classificação com diferentes medidas de similaridade para a América do Norte e observou que dentre os métodos os que apresentaram melhores resultados foram os de Ward, K-Means e Método Aglomerativo dos Núcleos. Ceballos & Braga (1995) aplicaram a análise de componentes principais em regiões homogêneas de radiação global diária para estimar dados faltosos locais. Os resultados obtidos foram satisfatórios, pois os erros encontrados entre séries estimadas e observadas não ultrapassaram 5%. Braga et al. (2006) empregaram as técnicas da análise de agrupamento para determinar grupos homogêneos e fração vegetação média anual no leste da Bahia nas décadas de 1980 e 1990. Unal et. al. (2003) aplicaram cinco métodos de agrupamentos a dados mensais de temperaturas (média, mínima e máxima), sendo o método de Ward o método que apresentou melhor poder discriminante. Recentemente, Amanajás e Braga, 2010, utilizam a ACP para analisar a variabilidade espacial da chuva no estado Amapá. Os autores mostram resultados coerentes com o regime climático da região.

Neste contexto, o objetivo deste estudo foi utilizar as propriedades da Análise dos Componentes Principais para

preenchimentos de falhas, mostrando que é possível simular séries temporais de um determinado local, baseado em médias mensais de dados disponíveis de estações circunvizinhas.

MATERIAL E MÉTODOS

Dados

Baseado em totais médios mensais de um longo período de observações (1963-2007) de precipitação, para 67 estações localizadas no Estado do Rio Grande do Norte (Figura 1) aplicou-se o método das

componentes principais para simular séries temporais. Esses dados foram cedidos pela empresa de Pesquisa Agropecuária do Rio Grande do Norte (Emparn). Para auxiliar na interpretação dos resultados utilizou-se o mapa das microrregiões do estado do Rio Grande do Norte que denota cada região de acordo com suas características peculiares às quais são fundamentais na análise dos resultados (Figura 2).

Figura - Localização e distribuição espacial das estações selecionadas no estado do Rio Grande do Norte.

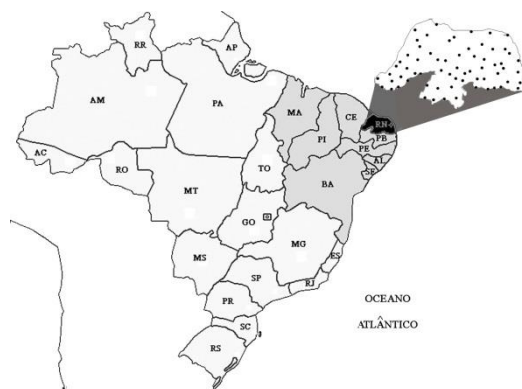


Figura 2 - Divisão das microrregiões, considerando os aspectos físicos, econômicos e demográficos do Estado do Rio Grande do Norte. Fonte: Aspectos Físicos, 2010



Análise de componentes principais (ACP)

A Análise em componentes principais tem por objetivo transformar p variáveis originais correlacionadas X em outras p variáveis não correlacionadas ou ortogonais em U , que explica uma parcela substancial das variáveis originais as quais são denominadas componentes principais. Onde a primeira componente principal explica a maior parcela da variabilidade total dos dados, a segunda explica a segunda maior parcela e assim sucessivamente. Neste sentido a ACP permite reduzir a dimensão de uma matriz de dados, investigar o comportamento espaço-temporal das variáveis envolvidas no problema, assim como detectar grupos de variáveis de comportamento similar (Richman, 1986; Ceballos & Braga, 1995).

A partir de uma matriz de dados normalizados da precipitação pluvial X (n indivíduos $\times p$ caracteres), se pode obter a matriz de correlação R através de:

$$R = \frac{1}{n} X.X^t \quad (1.1)$$

X^t é a transposta de X .

Sendo R uma matriz simétrica e positiva de dimensão ($p \times p$), ela é diagonalizável por uma matriz A de mudança de base, chamada de matriz de autovetores. Obtêm-se a matriz diagonal

D , cujos elementos diagonais são os autovalores de R , dada por:

$$D = A^{-1}.R.A \quad (1.2)$$

onde A^{-1} é a inversa da matriz A .

Assim, as CP's $U_1, U_2, \dots, U_p$, são obtidas por combinações lineares entre a transposta dos autovetores de A^t e a matriz de observações X , ou seja:

$$U = A^t.X \quad (1.3)$$

Cada componente U_k descreve o comportamento mensal do conjunto de estações. Os coeficientes a_{jk} representam os pesos associados a cada componente U_k e são ordenados por ordem decrescente dos autovalores mais significativos de A .

O número escolhido de Componentes Principais foi baseado no critério de truncamento de Kaiser, que considera como mais significativos os autovalores cujos valores sejam superiores à unidade (Garayalde et al., 1986). A distribuição espacial das CP's representa a contribuição destes em relação à variância total do conjunto de dados da precipitação em cada localidade. Os fatores comuns são obtidos pela relação entre a matriz dos autovetores (A) e a raiz quadrada do autovalor (λ), ou seja, o primeiro fator comum é obtido pela relação entre o primeiro autovetor (a_{11}) e o primeiro autovalor (λ_1) e assim sucessivamente.

Análise de Agrupamento

A partição dos grupos foi obtido através do método de K-means baseado na distância euclidiana e no critério de inércia (Bouroche & Saporta, 1983; Everitt, 1993; Gong & Richman, 1995). Logo a distância euclidiana é dada por:

$$d(s_i, s_j) = \left[\sum_{q=1}^p P_i (s_{iq} - s_{jq})^2 \right]^{\frac{1}{2}} \quad (1.6)$$

Em que S_i e S_j são pontos do espaço vetorial e I_p ; S_{iq} representa os valores mensais de cada estação nos locais S_i e S_j ; e P_i é o peso associado a cada individuo (estação).

A homogeneidade das regiões foi ajustada baseada no teste de significância de Bartlett (Bartlett, 1937; Snedecor & Cochran, 1989) dado por:

$$T = \frac{(N-K) \ln(S_p^2) - \sum_{i=1}^K (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(K-1)} \left(\sum_{i=1}^K \left(\frac{1}{n_i - 1} \right) - \frac{1}{N-K} \right)} \quad (1.7)$$

Em que S_i^2 é a variância de cada grupo, N é o total da amostra, n_i o tamanho da amostra do grupo, K é o número de Grupos e S_p^2 é a variância total dada pela equação:

$$S_p^2 = \frac{1}{N-K} \sum_i (n_i - 1) S_i^2 \quad (1.8)$$

Se $T < \chi_{\alpha; K-1}^2$ para o nível de significância α de 5%, tem-se evidências para aceitar a hipótese nula, ou seja, que as variâncias são iguais.

Simulação de séries temporais

Seja uma série temporal $\{X_t\}$ de uma variável meteorológica que se desconheça o dado de um determinado mês da m -ésima localidade (estação). Como $Z = A^t X$, todos as CP's Z_k são desconhecidas para este mês, e a incógnita x_m não pode ser avaliada mediante:

$$x_m = a_{m1} z_1 + \dots + a_{mk} z_k + \dots + a_{mM} z_M. \text{ Então}$$

a simulação da série temporal X_t pode ser feito pelo critério adotado por Ceballos & Braga (1995), no qual o conjunto de estimadores das CP's pode ser obtido com base no método dos mínimos quadrados, a partir de informações disponíveis das $M-1$ estações restantes. As condições introduzidas consistem em definir o vetor Z , cuja norma da diferença entre a parte conhecida X' do vetor X e sua estimativa X'' seja mínima, assim:

$$D^2 = (X'' - X').(X'' - X') = \sum_{m \neq k} \left[\sum a_{mk} z'_m - x'_m \right]^2, \quad (1.9)$$

$$k = 1, 2, \dots, k$$

$$\frac{\partial D^2}{\partial z'_j} = 0, \quad j = 1, 2, \dots, k \quad (1.10)$$

Z_k' são os estimadores das Cp's, X'' contém $k < M$ que acumulam uma fração da variância total dos dados da precipitação. O sistema algébrico de ordem K para as incógnitas Z_k^* pode ser escrito como:

$$\sum_{k=1}^k v_{jk} z_k^* = u_j \Rightarrow VZ^* = U \Rightarrow Z^* = V^{-1}U. \quad (1.11)$$

A matriz $V(K \times K)$ é menor do que a matriz A . O cálculo de V é obtido com o auxílio da matriz B de dimensão $(M - 1) \times K$, resultante de A incompleta sem a m -ésima linha, logo a matriz V é dada por:

$$V = B^t B \quad (1.12)$$

Generalizando o procedimento a série temporal pode ser escrita da seguinte maneira:

$$Z^* = V^{-1}U \text{ ou} \quad (1.13)$$

$$S = (B^t B)^{-1} B^t,$$

$$Z^* = SX', x_i^* = LSX' \quad (1.14)$$

Em que X' é a parte conhecida do vetor $\vec{X}$ e $L = (a_{i1}, \dots, a_{ik})$ é a i -ésima linha da matriz A .

RESULTADOS E DISCUSSÃO

Foram retidas as três primeiras CP's que explicaram aproximadamente 94,0% da variância total dos dados mensais de precipitação conforme critério sugerido por Kaiser (Garayalde et al., 1986). A primeira CP que explica 52,02% da variância total da série apresenta altas correlações positiva superiores a 0,8 nos meses de maio a outubro e mínimas em

janeiro e fevereiro (Figura 3). A configuração espacial associada a esta CP tem contribuições superiores à 1 na faixa litorânea leste, decrescendo em direção ao interior com valores negativos nas regiões sul e central do Estado (Figura 4a). Este fator delimita a costa leste do Estado com período mais chuvoso de maio a setembro e está relacionada à atuação dos sistemas de brisa e ventos alísios. Estudos mostram que mais de 50% da precipitação nesta região ocorrem no período noturno devido ao contraste térmico entre continente/oceano e circulações de brisas mais intensas (Kousky, 1980; Rao et al., 1993).

A segunda CP que explica 37,0% da variância total dos dados da precipitação apresenta correlações superiores a 0,8 de dezembro a abril e negativa de julho a setembro (Figura 3). O padrão espacial desta CP tem contribuições positivas superiores a 0,4 no oeste, sudoeste e noroeste. No restante da região, a contribuição da variância é pequena, ou seja, indicando período de pouca chuva como pode ser visto na Figura 4b. Esta CP pode estar associada com a precipitação influenciada pela atuação da zona de convergência (ZCIT) nos meses de dezembro (Hastenrath, 1985; Uvo, 1989; Mota, 1997). Finalmente, a terceira CP que explica 5,04% da variância total tem correlação temporal pouco significativa de

0,4 apenas em novembro. O padrão espacial desta CP apresenta contribuições superiores a 1 em dois pequenos núcleos no sudeste e extremo sudoeste. No restante da região as contribuições da variância são mínimas, indicando portanto o período de estiagem da região. Esses dois núcleos podem está relacionados com a orografia local provocando chuvas isoladas em novembro (Figura 4c). Estes três primeiros fatores permitiram detectar a existencia de regiões mais específicas e de comportamento distintos com relação a precipitação média mensal. Os resultados obtidos colaboram com os encontrados preliminarmente por Braga (1992), que na ocasião evidenciou discrepâncias significativas na pluviosidade no litoral leste e norte, assim como no oeste do Estado.

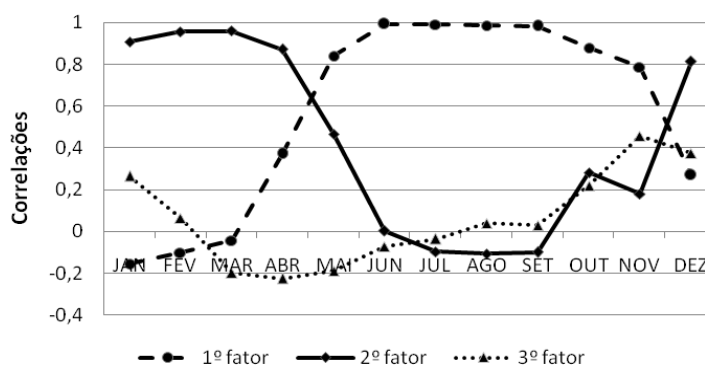
Evidentemente a ACP constitui apenas um instrumento matemático para o estudo de variáveis meteorológicas, no

entanto o conhecimento da climatologia e sistemas atuantes na região é indispensável para associar os resultados da ACP com o comportamento climatológico da região.

A aplicação do método de K-means (Gong & Richman, 1995) aos três primeiros fatores comuns temporais da precipitação dividiu o Estado do Rio Grande do Norte em quatro regiões homogêneas, Figura 5. Suas principais características estão sintetizadas na tabela 1 e descrição mais detalhadas a seguir.

O Grupo 1, situado no Agreste e Curimataú, possui 20 estações com precipitação média anual da ordem 579 mm apresentando clima semiárido e trimestre mais chuvoso de fevereiro a abril. Esta região é a mais castigada pela seca no estado, devido a maior influência dos ventos alísios secos do Nordeste, que incidem no Litoral Norte se interiorizando pelo território potiguar (Aspectos Físicos, 2010).

Figura 3 - Correlações entre os três primeiros fatores comuns temporais que explicam 94,04% da variância total dos dados.



Grupo 2 localizado no litoral leste, possui climas úmido e sub-úmido com precipitação média anual em torno de 1181mm (Tabela 1), período chuvoso de março a julho. Esta região é influenciada pela zona de convergência intertropical

(ZCIT) e sistemas de leste (brisas e ondas de leste) (Uvo, 1989; Kousky, 1980).

Grupo 3 está situado no sul e sudoeste do estado, constituído por 23 estações com total médio anual de 787mm caracterizando as zonas do Alto Apodi, Caicó e Serras Centrais (Figura 3).

Tabela 1 - Principais características das regiões homogêneas da precipitação pluvial no Rio Grande do Norte.

Regiões	Estações (%)	Localização	Média anual pluviométrica (mm)	Precipitação mínima(mm)	Precipitação máxima(mm)
G1	29,9	Centro	579,4	449,0	718,8
G2	13,4	Litoral leste	1181,6	825,7	1640,9
G3	34,3	Sudoeste	787,4	614,2	997,0
G4	22,4	Litoral Norte	710,7	609,1	797,5

O Grupo 4 se estende deste do litoral norte até a zona Mossoroense, Serras Central (Figura 3 e 5) este grupo possui clima semi-árido com total médio anual de 710,7mm e período chuvoso concentrado nos meses de fevereiro e março, cujas precipitações estão diretamente influenciadas pela ZCIT e sistemas de brisas.

Simulação de séries temporais

A Figura 5 ilustra as regiões homogêneas do estado do Rio Grande do Norte obtidas para dados mensais de 67 estações meteorológicas selecionadas e agrupadas pelo método de K-means (Gong

e Richman,1995; Braga et al., 2006; Wilks, 2006). O conjunto de estimadores das CP's são obtidos com base no método de mínimos quadrados conforme descrito na sessão 2.3. Foram selecionados dois grupos com características pluviométricas distintas, ou seja, na região seca no interior (semi-árida) e outra mais úmida na parte do litoral leste do Estado. A estes grupos homogêneos aplicou-se a ACP objetivando simular série de dados faltosos de um ano específico. Na Tabela 1 apresentam-se as variâncias das estações pertencentes ao grupo 2 e grupo 3). Simularam-se séries temporais para estação de Touros e Caraúbas, respectivamente.

Figura 4 a, b, c - Padrões espaciais das três primeiras CP's da precipitação do Rio Grande do Norte que explicam da 94.04% da variância total dos dados

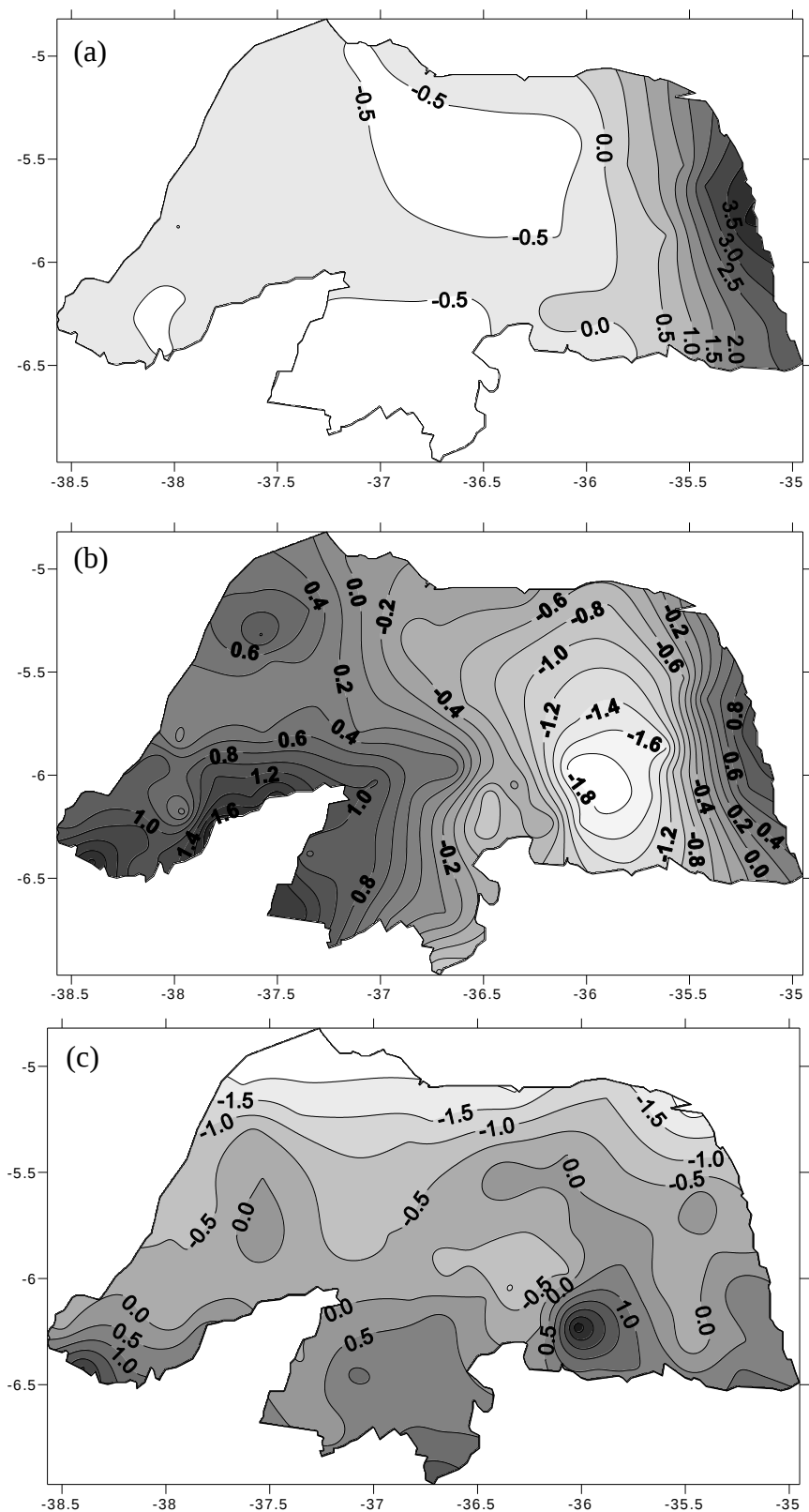
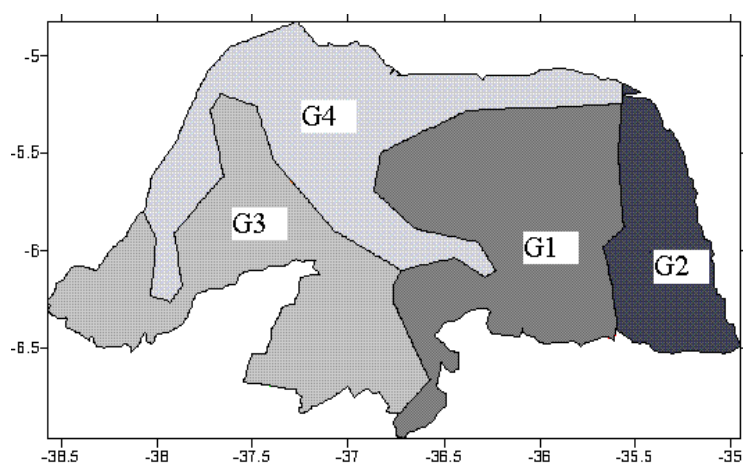


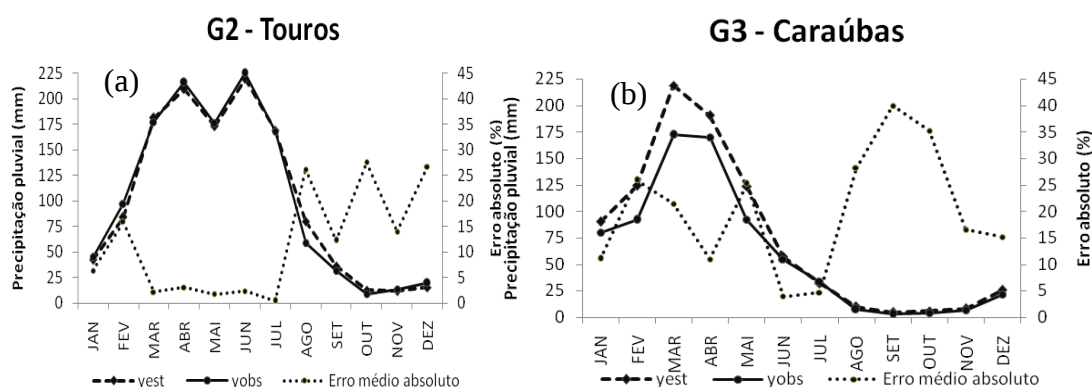
Figura 5- Regiões homogêneas de precipitação através do método de K-Means



Na Figura 6a ilustra avariação temporal da série real e estimada para estação de Touros. Verifica-se que o maior erro percentual associado às séries estimadas foi da ordem de 28% para o mês de outubro e os menores se observa nos meses de março a julho. Já na região semi-árida representada pela localidade Caraúbas, o maior erro percentual de 40% verificado no mês de novembro e menor nos meses de junho e julho. O erro médio percentual para estas localidades foram de 12 e 19,8% respectivamente.

Todas as aproximações visualizadas na Figura 6a e 6b denotam a eficácia da simulação das séries temporais da precipitação média mensal utilizando o método ACP de regressão dos mínimos quadrados proposto por Ceballos & Braga (1995). Isto significa que a série temporal de dados do m-ésimo local pode ser estimada pela ACP, principalmente para as regiões onde as chuvas são mais regulares, a exemplo de Touros (grupo 2).

Figura 6 a,b - Séries pluviométricas estimadas e observadas e seus respectivos erros percentuais das estações representativas dos grupos G2(a), G3(b).



CONCLUSÕES

A aplicação da análise subjetiva da ACP e objetiva de agrupamento as séries temporais de precipitação permitiu identificar quatro (04) regiões homogêneas e associá-las aos diferentes sistemas meteorológicos que atuam direta ou indiretamente na precipitação do Estado. O critério de simulação de séries temporais X_t adotado por Ceballos & Braga 1995, onde o conjunto de estimadores das CP's podem ser obtidos com base no método dos mínimos quadrados, a partir de informações de M-1 estações, mostrou-se eficaz com um erro percentual médio para a região litorânea de 12% e 19,8% para a região semiárida.

REFERÊNCIAS BIBLIOGRÁFICAS

- ASPECTOS FÍSICOS. Disponível em: <<http://www.idema.rn.gov.br/perfilrn/Aspectos-fisicos.pdf>>. Acesso em: 20 de fevereiro de 2010.
- BARTLETT, M. S. **Properties of sufficiency and statistical tests.** Proceedings of the Royal Statistical Society Series A 160, p. 268-282, 1937.
- BOUROCHE, J. N. & SAPORTA, G. **L'Analyse des Données, Col. 'Que sais-je?'** Presses Universitaire de France, p. 127, 1983.
- BRAGA, C. C. Classificação de regiões pluviometricamente homogêneas Através da Analise em Multivariada. **VII Congresso Brasileiro de Meteorologia.** São Paulo, SP. Anais... v. 1, p. 338-341, 1992.
- BRAGA, C. C.; BRITO, J. B. B.; SANSIGOLO, C. A.; RAO, T. V. R. Tempo de resposta da vegetação às variabilidades sazonais da precipitação. **Revista Brasileira de Agrometeorologia**, v.14, n.3, p.252-260, 2006.
- CEBALLOS, J. C.; BRAGA, C. C. Missing Data Assessment in a Solarimetric Network. **International Journal of Climatology**, v.15, p. 325-340, 1995.
- EASTMAN, J.R., FULK, M. Long sequence time series evaluation using standardized principal components. **Photogrammetric Engineering & Remote Sensing**, v.59, n.9 p.1307-1312, 1993.
- 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, CONGRESSO BRASILEIRO DE METEOROLOGIA, 1986, Brasília. **Anais...**, v.1, p.119-124, 1986.
- GONG, X; RICHMAN, M. B. On the application of Cluster to Growing Season Precipitation Data in North America East of the Rockies. **American Meteorological Society**, abril, 1995.
- KOUSKY, V. E. Frontal influences on Northeast Brasil. **Monthly Weather Review**, v.107, n.9, p.1140-1153, 1979.
- KOUSKY, V. E. Diurnal rainfall variations in northeast Brazil. **Monthly Weather Review**, v.108, n.4, p.488-498, 1980.
- RAO, V. B., LIMA de, M. C. FRANQUITO, S. H. Seasonal and variations of rainfall over eastern Northeast Brasil. **Journal of Climate**, v.6, p.1723-1754, 1993.

RICHMAN, M. B. Review article on rotation of principal components. **Journal of Climatology**, v.6, p. 293-335, 1986.

SILVA, L. L. **Precipitações pluviais da pré-estação chuvosa e do período chuvoso e suas influências na produtividade agrícola da Paraíba.** 2007. Dissertação de mestrado em Meteorologia, UFCG (Universidade Federal de Campina Grande), Campina Grande, PB, 2007.

SNEDECOR G. W.; COCHRAN, W. G. **Statistical methods.** 8º edição, Iow State, University Press, 1989.

UNAL, Y.; KINDAP, T.; KARACA, M. Redefining the climate zones of Turkey

using cluster analysis. **International Journal of Climatology**, v.23, p. 1045-1055, 2003.

UVO, C. R. B. **Zona de Convergência Intertropical (ZCIT) e sua relação com a Precipitação na região norte e Nordeste do Brasil.** São José dos Campos, 1989. 82p. Dissertação (Mestrado em Meteorologia) - Instituto de Pesquisas Espaciais, 1989.

WILKS, D. S. **Statistical Methods in the Atmospheric Sciences.** 2ª Edition. California: Elsevier Science & Technology Books. Academic Press, 2006.