

SIMULAÇÃO DE DADOS DE PRECIPITAÇÃO UTILIZANDO ANÁLISE EM COMPONENTES PRINCIPAIS.

Célia Campos **BRAGA**¹, Maria Luciene Dias **DE MELO**².

RESUMO

O presente trabalho fundamenta-se em analisar em que medida as séries mensais de precipitação em locais do Estado de Alagoas podem ser avaliadas com base, em simulação de dados disponíveis de estações pertencentes a grupos homogêneos. Os grupos homogêneos foram determinados através da técnica de agrupamento hierárquico proposta por Ward 1963. A consistência dos dados mensais das séries pluviométricas foi obtida a partir da Análise em Componentes Principais para os grupos homogêneos previamente determinados. Os resultados da simulação efetuada evidenciam que os erros percentuais médios entre as séries observada e estimada giram em torno de 20%, deduzindo-se que o método de simulação empregado é de grande utilidade para o preenchimento de falhas, de dados faltosos, em regiões com regime pluviométrico regular.

INTRODUÇÃO

Situado na região Nordeste do Brasil, o Estado de Alagoas abrange uma área de 27.731 km², encontra-se limitado pelos estados de Pernambuco, Sergipe e por um pequeno trecho da Bahia e compreende, aproximadamente, 1,82% da região Nordeste.

Possui um relevo constituído por planícies, vales, serras e montanhas com altitude entre 800m e 1200m, como por exemplo, a Serra das Palmeiras. O território alagoano é afetado por diversos sistemas de tempo, tais como a Zona de Convergência Intertropical (ZCIT), Sistemas Frontais, Zona de Convergência do Atlântico Sul (ZCAS), ondas de leste, brisas marítimas/terrestres e ventos vale/montanha.

A variabilidade espacial da pluviosidade no estado de Alagoas não é muito acentuada, podendo ser caracterizada em áreas distintas, as quais são: a faixa do litoral, que possui índices superiores a 1800mm em algumas áreas. Nesta região a chuva é caracterizada por ter uma distribuição mais regular no decorrer do ano. O Agreste, também possui índices elevados, superiores a 1000mm durante o outono. Os índices mais baixos inferiores a 600mm são encontrados no Sertão do São Francisco, à Noroeste do estado (Figura 1).

1- Professora do Departamento de Ciências Atmosféricas – DCA/UFPB

2- Aluna da Graduação em Meteorologia da UFPB– Bolsista PIBIC / CNPq / UFPB

A escassez das estações climatológicas, e a distribuição, espaço-temporal, da precipitação no Estado de Alagoas e ainda tendo em vista necessidade de utilização de séries longas de dados nos estudos climatológicos, o presente trabalho tem como objetivo, identificar sub-regiões homogêneas em relação ao ciclo e total anual da precipitação média mensal, oferecendo, assim, uma contribuição ao estudo climático do Estado. A identificação destas sub-regiões foi feita pelo método da análise em multivariada, através da técnica da análise de agrupamento hierárquico, proposta por Ward (EVERITT, 1974), GONG and RICHARMAN, 1995; e componentes principais (BOUROCHE et al. 1980), (BRAGA et al. 1998), CEBALLOS & BRAGA (1995).

Este trabalho é um seguimento de estudos feitos para alguns estados do Nordeste, por (BRAGA et al.; 1996, 1998). Neste contexto, utilizou-se o Agrupamento de fatores e as propriedades da Análise em Componentes Principais (ACP) para efetuar o preenchimento de falhas, mostrando que é possível simular séries temporais de um determinado local, com base em médias mensais de dados disponíveis de estações circunvizinhas.

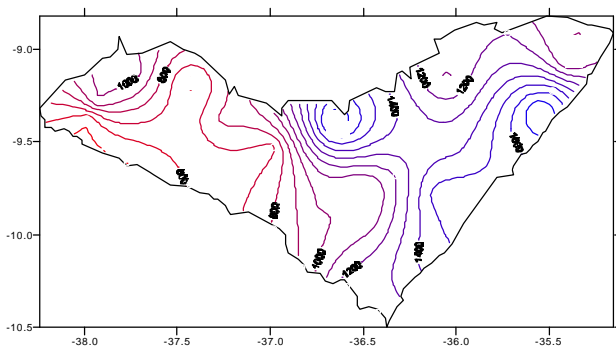


Fig.1 – Distribuição espacial dos totais médios anuais da precipitação do estado de Alagoas.

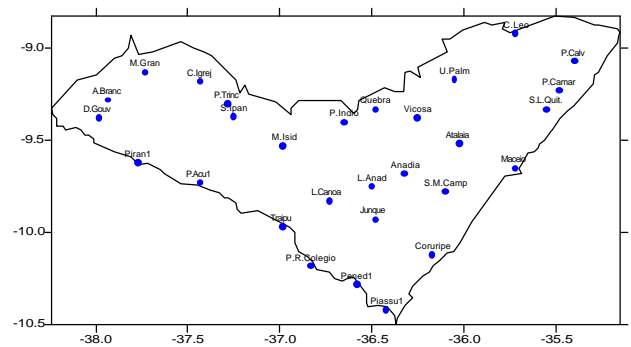


Fig. – Distribuição espacial das estações Meteorológicas do Estado de Alagoas.

MATERIAIS E MÉTODOS

Os dados utilizados, consistem de séries mensais de precipitação, de 29 estações meteorológicas distribuídas espacialmente no estado de Alagoas, que possuem no mínimo trinta (30) anos de observação. As médias mensais foram dispostas sob forma de matriz $X_{n \times p}$, onde os X_{ij} representam a i -ésima observação da j -ésima estação. Assim o número n de linhas corresponde aos 12 meses e o número p de colunas corresponde as 29 estações. Organizou-se um arquivo de totais médios mensais da precipitação, que serviram de entrada para software estatístico, para efetuar o cálculo das Componentes Principais (CP's), para em seguida, subdividir o Estado em sub-regiões homogêneas utilizando o método de agrupamento proposto por WARD (1963).

O objetivo da ACP consiste em transformar P variáveis originais (p estações) correlacionadas em p componentes ortogonais, sendo estas componentes funções lineares das

variáveis originais (HAAN, 1979). Uma característica da ACP é a redução do número de variáveis correlacionadas. A correlação existente entre as p variáveis indica que alguma informação contida numa variável, estará também contida em alguma das outras $p - 1$ variáveis. Desta forma, a ACP permite não só a redução da dimensão da matriz de dados, mas também investiga o comportamento espacial e temporal das variáveis envolvidas no problema, bem como detecta grupos de variáveis que apresentam comportamento homogêneo.

A técnica da ACP consiste em se projetar uma nuvem de indivíduos, situada num espaço afim de dimensão p , em um sub-espaço afim de dimensão q , determinado de tal forma que sua projeção sofra a menor deformação possível. O plano sobre o qual são efetuadas essas projeções é denominado de plano principal.

A partir de uma matriz de dados X , normalmente distribuída com $E(x_j) = m(j)$ e variância $V(x_j) = s$, $j=1 \dots p$, podemos obter a matriz de variância e covariância S através de:

$$S = \left(\frac{1}{n} \right) X^t \cdot X, \quad \text{onde } X^t \text{ é a transposta de } X. \quad (2.1)$$

Normalizando a matriz das variáveis X_1, X_2, \dots, X_p obtêm-se a matriz de correlação R , e neste caso:

$$S = R = \left(\frac{1}{n} \right) X' \cdot X^t \quad (2.2)$$

Sendo a matriz de covariância S uma matriz simétrica e positiva de dimensão $(p \times p)$, ela é diagonalizável por uma matriz A , de mudança de base denominada de matriz dos autovetores. A matriz diagonal D , cujos elementos diagonais são os autovalores de R , é obtida por:

$$D = A^{-1} \cdot R \cdot A \quad (2.3)$$

onde A^{-1} é a inversa da matriz A .

As componentes principais U_1, U_2, \dots, U_p , são obtidas por combinações lineares entre a transposta dos autovetores (A^t) e a matriz de observações (X), ou seja:

$$U = A^t \cdot X \quad \text{onde, } X = A \cdot U \quad (2.4)$$

Os valores de X_i do n -ésimo local podem ser estimados pela seguinte expressão:

$$X_i = a_{j1} \cdot U_1 + a_{j2} \cdot U_2 + \dots + a_{jk} \cdot U_k + \dots + a_{jp} \cdot U_p \quad (2.5)$$

Onde as variáveis de U_p são o conjunto dos autovetores de X ordenados por ordem decrescente dos autovalores mais significativos de a_k em A .

O método de classificação utilizado é um método hierárquico, que produz uma seqüência de partições em classes cada vez mais vastas e obtém a construção de uma seqüência de partição em $n, n-1, n-2, \dots$, classes embutidas uma nas outras no sentido de uma distância e um critério de agrupamento (BOUROCHE et. al., 1980; EVERITT, 1974). Foi utilizado critério de agrupamento

de Ward (1963) para medir a qualidade de uma partição. A distância utilizada para medir a similaridade ou dissimilaridade entre os indivíduos (estações), S_i e S_j , foi a euclidiana.

O critério de simulação de série temporal $\{X_t\}$ foi o adotado por CEBALLOS e 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 \vec{Z} , cuja norma da diferença entre a parte conhecida X' do vetor \vec{X} e sua estimativa X'' seja mínima; assim:

$$D^2 = (X'' - X') \cdot (X'' - X') = \sum_{m \neq k} \left[\sum a_{mk} z'_m - x'_m \right]^2, k = 1, 2, \dots, K; \quad \frac{\partial D^2}{\partial z'_j} = 0, j = 1, 2, \dots, K. \quad (2.6)$$

O sistema algébrico de ordem K para as incógnitas z_k^* pode ser escrito como:

$$\sum_{k=1}^K v_{jk} z_k^* = u_j \Rightarrow VZ^* = U \Rightarrow Z^* = V^{-1}U. \quad (2.7)$$

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 .

$$V = B^t B, \quad Z^* = V^{-1}U \quad \text{ou} \quad S = (B^t B)^{-1} B^t, \quad (2.9)$$

$$Z^* = SX', \quad x_i = LSX'. \quad (3.0)$$

X' é a parte conhecida do vetor \vec{X} .

$$L = (a_{i1}, \dots, a_{ik}) \text{ é a } i - \text{ésima linha da matriz } A. \quad (3.1)$$

RESULTADOS E DISCUSSÃO

A aplicação da análise em componentes principais (ACP) aos valores médios mensais da precipitação resultou na retenção dos 2 primeiros autovetores associados aos dois primeiros autovalores mais significativos.

As duas primeiras CP's explicam aproximadamente 98% da variância total da série, e observa-se que a primeira CP apresenta forte componente anual com máximo no outono e inverno entre maio e julho e mínimo na primavera e verão (Figura 3). A configuração espacial do primeiro fator comum temporal (correlações), mostra contraste entre a parte sul e norte, com a porção a oeste do Estado (em torno de $9,5^\circ$). Estas variações estão possivelmente associadas ao relevo local (região do Vale do São Francisco).

A segunda CP apresenta comportamento sazonal bastante regular durante o ano inteiro. O fator comum temporal desta 2ª CP, evidencia uma dessemelhança de regime pluviométrico entre a

parte oeste (acima citado) e as outras regiões do Estado, onde as precipitações são bem mais acentuadas.

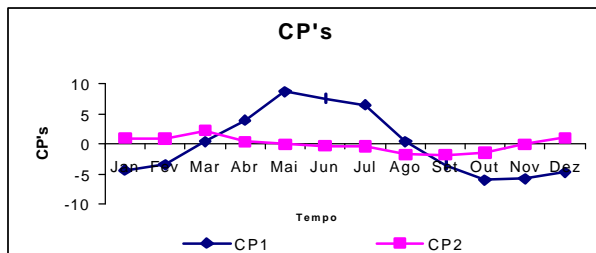


Fig. 3 - Comportamento temporal das duas primeiras Componentes Principais.

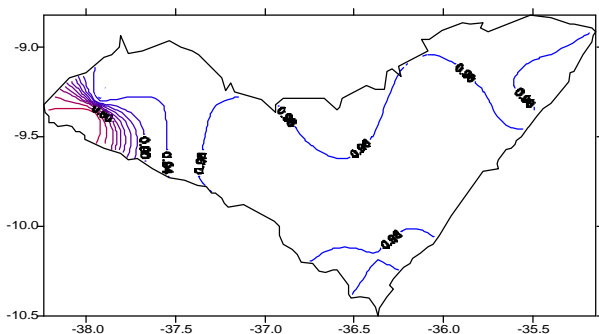


Fig.4 – Distribuição do 1º fator de ponderação, explicando 93,1% da variância total dos dados médios mensais para o Estado de Alagoas.

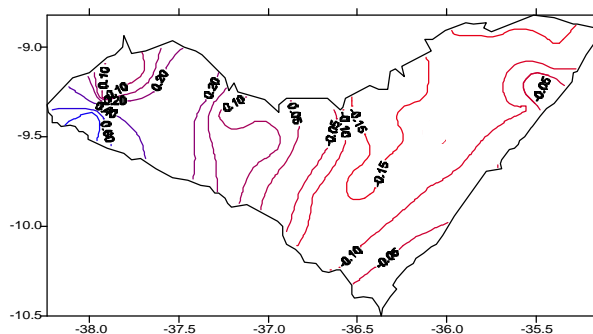


Fig.5 – Distribuição do 2º fator de ponderação explicando 4,6% da variância total dos dados Médios mensais para o Estado de Alagoas.

A delimitação das regiões homogêneas foi efetuada a partir do método de agrupamento de Ward, Ligação Simples e Completa (EVERITT, 1974). Embora se tenha utilizado vários métodos, optou-se pelo de classificação hierárquico ascendente baseado no critério de inércia intra-classe, proposto por Ward (1963), por ter apresentado melhores resultados.

Os grupos homogêneos de estações foram determinados por cortes transversais no dendrograma, baseado no critério de inércia de WARD (1963) e conhecimento da climatologia da região.

Levando em consideração o conhecimento climático e os procedimentos de análises, acima citados, optou-se em que dividiu o Estado de Alagoas em quatro (04) sub-regiões pluviometricamente homogêneas (figura 6).

Região1 (Zona da Mata): Anadia, Atalaia, Colônia Leopoldina, Junqueiro, Limoeiro de Anadia, Palmeira dos Índios, Quebragulho, União dos Palmares, Viçosa. Esta região possui precipitação entre 1000 e 1400mm, sofrendo influência das ondas de leste, ZCIT e também das brisas marítimas.

Região2 (Litoral e Agreste): Coruripe, Lagoa da Canoa, Maceió, Passo do Camaragibe, Penedo, Piassabussu, Porto Calvo, Porto Real do Colégio, São Luís do Quitunde, São Miguel dos Campos,

Traipú. Esta região é caracterizada pelo o maior índice de precipitação do Estado, variando de 1000 a 1800 mm, do agreste ao litoral.

Região3 (Sertão):Água Branca, Major Isidoro, Mata Grande, Pão de Açúcar, Poço das Trincheiras, Santana do Ipanema. Esta, possui uma variação de precipitação de 700 a 1000 mm. Nesta região encontra-se uma localidade com características do grupo4, isto deve-se a orografia local

Região4 (Sertão do São Francisco): Capia da Igrejinha, Delmiro Gouveia, Piranhas. É caracterizada pelo menor índice pluviométrico, variando entre 500 e 600 mm, isto está relacionado com a sua orografia, pois está situado no Vale do São Francisco.

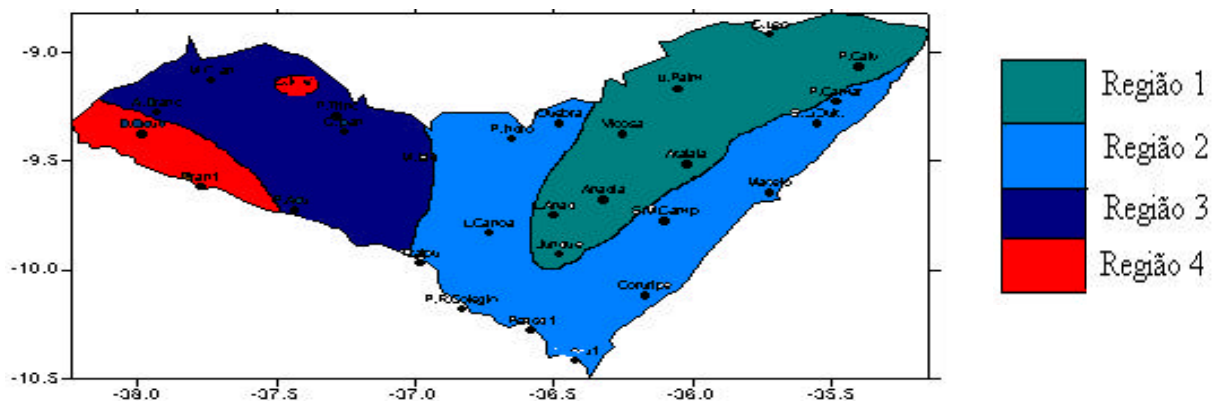


Fig.6 – Zoneamento da distribuição espacial da precipitação no estado de Alagoas, utilizando o método de agrupamento de Ward.

A distribuição espacial dos fatores de ponderação seguida da análise de agrupamento hierárquico de Ward e análise das Cp's evidencia a existência de quatro (04) regiões pluviometricamente homogêneas no Estado de Alagoas.

A partir das regiões homogêneas determinadas através do método de classificação hierárquico (Figura 6), selecionou-se algumas estações pertencentes a regiões homogêneas distintas para proceder a simulação. A estes grupos homogêneos foi feita a simulação utilizando a ACP e método dos mínimos quadrados descritos acima, para a estimação de séries com dados faltosos.

As regiões homogêneas aplicou-se a ACP objetivando simular série de dados para algumas estações de cada grupo. A figura 7 representando a classe 1, ilustra a variação temporal das séries estimada e observada para Atalaia. Verifica-se que o erro percentual máxima associado as duas séries é da ordem de 12% em dezembro, enquanto que, o erro médio padrão foi de 3,5% (com 28 anos de dados de observações). A figura 8 mostra a variação temporal da séries estimada e observada para estação de Água Branca no sertão do Estado. O erro máximo foi da ordem 15%,

enquanto que o médio é de 6%. O que nos leva a concluir que o método empregado para dados médios mensais é de grande utilidade para o preenchimento de falhas em regiões com regime pluviométrico .

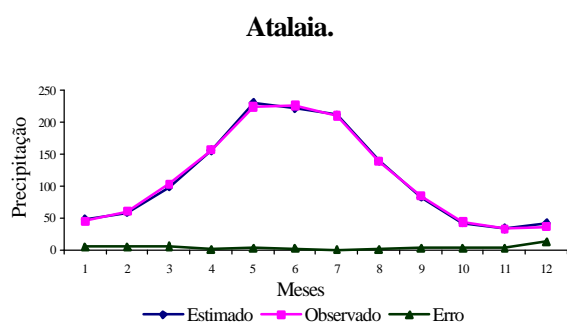


Fig. 7 - Séries de precipitação observada e estimada para a estação de Atalaia.

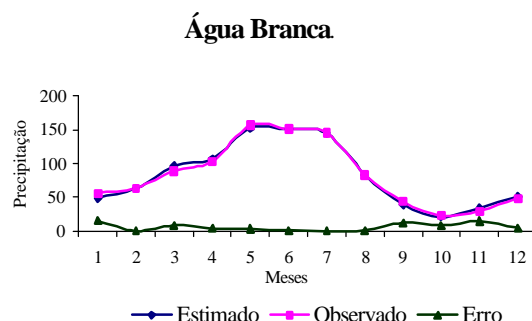


Fig. 8 - Séries de precipitação observada e estimada para a estação de Água Branca.

BIBLIOGRAFIA

- BOUROCHE, J.M. & SAPORTE, G. **L'Analyse des Données**. Presses Universitaire France, 127p., 1980.
- BRAGA, C.C., at al. Simulação e Consistência de Séries Temporais de Precipitação para o Estado de Pernambuco. IX Congresso Brasileiro de Meteorologia. Campos do Jordão. **Anais**, 312 – 315, 1996.
- BRAGA, C.C., MELO, M.L.D., MELO, E.C.S.. Análise Objetiva da Distribuição da Pluviometria no Estado da Bahia. X Congresso Brasileiro de Meteorologia –VIII Congresso da FLISMET. **CDROM**, 1998.
- CEBALLOS, J.C. e BRAGA, C.C. Missing data assasment in a solarimetric network. **International Journal of Climatology**, 15: 325 – 340, 1995.
- EVERITT, B. **Cluster Analysis**. Heinemann Educational Books, London, 135 p, 1974.
- GONG, X. and RICHARMAN, M.B. On the Application of Cluster Analysis to Growing Season Precipitation Data in North America Eat of the Rochies. **American Meteorological Society**. 8: 897-930, 1995.
- HAAN, C.T. Statical Methods in Hidrology. Iowa, **The Iowa State University Press**, 313p., 1979.
- NIMER, Edmon. **Climatologia do Brasil**. 2ª ed. Rio de Janeiro: IBGE, Departamento de Recursos Naturais e Estudos Ambientais, 422 p, 1989.
- WARD, J.H. Hierarchical grouping to otimize an objective function. **Journal American Association**, 58: 236 – 244 p, 1963.