

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
CURSO DE MESTRADO EM METEOROLOGIA

SÉRIES TEMPORAIS DE IRRADIAÇÃO SOLAR: RECUPERAÇÃO DE
DADOS MEDIANTE COMPONENTES PRINCIPAIS

GILBERTO BARBOSA DINIZ

Campina Grande - PB.
Agosto de 1994

SÉRIES TEMPORAIS DE IRRADIAÇÃO SOLAR: RECUPERAÇÃO DE
DADOS MEDIANTE COMPONENTES PRINCIPAIS

Dissertação apresentada ao curso de mestrado
em meteorologia da Universidade Federal da
Paraíba, em cumprimento às exigências para
obtenção do grau de Mestre.

Área de concentração: Climatologia estatística
Orientador: Prof. Dr. Juan Carlos Ceballos

Campina Grande - PB.
1994



D585s Diniz, Gilberto Barbosa
 Series temporais de irradiacao solar : recuperacao de dados mediante componentes principais / Gilberto Barbosa Diniz. - Campina Grande, 1994.
 118 f.

 Dissertacao (Mestrado em Meteorologia) - Universidade Federal da Paraiba, Centro de Ciencias e Tecnologia.

 1. Radiacao Solar 2. Irradiacao Solar 3. Meteorologia 4. Dissertacao I. Ceballos, Juan Carlos, Dr. II. Universidade Federal da Paraiba - Campina Grande (PB)

CDU 551.521.1(043)

SÉRIES TEMPORAIS DE IRRADIAÇÃO SOLAR: RECUPERAÇÃO DE DADOS
MEDIANTE COMPONENTES PRINCIPAIS

GILBERTO BARBOSA DINIZ

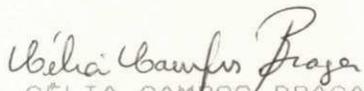
DISSERTAÇÃO

APROVADA

EM 09/08/94



JUAN CARLOS CEBALLOS
Orientador



CÉLIA CAMPOS BRAGA
Componente da Banca



CLÓVIS ANGELI SANSIGOLO
Componente da Banca

CAMPINA GRANDE
AGOSTO - 1994

SUMÁRIO

	pagina
DEDICATÓRIA	i
EM MEMÓRIA	ii
AGRADECIMENTOS	iii
RESUMO	iv
ABSTRACT	v
LISTA DE FIGURAS	vi
LISTA DE TABELAS E QUADROS	viii
1 - INTRODUÇÃO	01
2 - OBJETIVOS	03
2.1-Objetivo geral	03
2.2-Objetivo específico	03
3 - FUNDAMENTAÇÃO TEÓRICA	04
3.1-Análise de agrupamentos (Cluster analysis)	06
3.1.1-Função de agrupamento	08
3.1.2-Medidas derivadas da distância euclidiana	08
3.1.3-Métodos hierárquicos de agrupamento	09
3.1.4-Um exemplo de análise	12
3.2-Análise de componentes principais (ACP)	15
3.2.1-Fundamentação estatística da (ACP)	16
3.2.2-Alguns exemplos de aplicação da (ACP) em met.	18
3.2.3-Simulação de séries temporais	23
3.2.4-Séries sintéticas	24
3.3-Análise fatorial	26
4 - MATERIAIS E MÉTODOS	28
4.1-Descrição dos dados e área de estudo	28
4.2-Período de estudo	31
4.3-Agrupamento de estações	38

DEDICATÓRIA

À minha mãe, MARIA BARBOSA DINIZ, pelo amor infinito que se traduz nos pequenos gestos de renúncia e abnegação.

Dedico também à minha esposa, ELIANE DA SILVA ALCOFORADO DINIZ e filhas TASSIANA ALCOFORADO DINIZ e JULIANA ALCOFORADO DINIZ, como incentivo a prosseguir seus estudos. Aos meus colegas Professores da UFPel, que me incentivaram a afastar-me das atividades docentes e iniciar um curso de Pós-graduação.

Em memória.

Ao meu pai, FRANCISCO MARTINIANO DINIZ, pelo muito que representou para mim como exemplo de honestidade, caráter e integridade, que tanto valorizou o estudo e o saber. Sei que onde ele estiver estará se realizando, através de mim, por esse trabalho.

AGRADECIMENTOS

Agradeço a todos aqueles que de uma forma ou de outra contribuíram com seu incentivo e ajuda à conclusão deste.

Agradeço em especial ao Professor JUAN CARLOS CEBALLOS, pela imprescindível orientação e apoio material.

Ao Professor JUAREZ FERNANDES DE OLIVEIRA, pelo auxílio no uso e implantação do software SPSS/PC.

À professora CELIA CAMPOS BRAGA, pelas várias dúvidas tiradas com relação à análise multivariada.

Aos funcionários do DCA MIRIAM CARMEM COSTA, ROBERTO AMORIM GUEDES e EYRES DIANA VENTURA SILVA pelo auxílio na obtenção e tratamento dos dados de irradiação imprescindível ao trabalho.

À analista de sistemas ELIANE DA SILVA ALCOFORADO DINIZ, pela incansável dedicação e auxílio nos trabalhos relativos à informática.

Aos meus colegas de curso e professores do DCA que me auxiliaram na solução de pequenas dúvidas.

Aos profissionais da secretaria do Curso de mestrado em meteorologia.

Muito obrigado.

RESUMO

Foram simuladas séries temporais de irradiação global de duas estações da rede solarimétrica da UFPb (Barra de Santa Rosa e Pombal) usando técnicas de análise multivariada. Os dados usados foram das 16 estações da Rede Solarimétrica, mais 3 do Instituto Nacional de Meteorologia INEMET-MA localizadas nas regiões de fronteiras da Paraíba, nos Estados de Pernambuco, Ceará e Rio Grande do Norte. Nesta pesquisa, foram usados dados coletados num período de seis anos (1975 a 1980) com dois meses: um do período chuvoso da região de localização das estações (março), e outro do período seco (setembro). Previamente foram determinados grupos de estações com características homogêneas. O método de agrupamento adotado, dentre todos os testados, foi o de Ward o qual apresentou características satisfatórias de separação em grupos e parece ser o mais recomendável para o trabalho. Os grupos de estações homogêneas foram obtidos agrupando vários tipos de variáveis como por exemplo: dados brutos e reduzidos de irradiação, cargas fatoriais obtidas dos autovetores que explicaram 90% da variância e os três primeiros. Devido a limitação dos dados do período estudado, definiu-se cinco grupos apropriados às 19 estações, com exceção de alguns casos, que foi ainda menor.

Para as simulações das series temporais utilizou-se os dados do grupo pertencentes a Barra de Santa Rosa (BRS) e Pombal (POM). Aplicou-se, portanto, o método desenvolvido por CEBALLOS e BRAGA (1991) o qual se baseia na matriz de autovetores associados à correlação entre estações de um grupo homogêneo. Para os cálculos pertinentes, foram utilizados apenas os autovetores mais significativos da matriz (os que explicaram 90% da variância e os dois primeiros). As simulações apresentaram bons resultados do desvio da série simulada com relação à original entorno de 0,45, sendo portanto, da ordem dos próprios erros dos dados de irradiação das estações em estudo.

ABSTRACT

Time series of global irradiation of two stations (Barra de Santa Rosa and Pombal) of the solarimetric network of Paraíba were simulated using the technique of multivariate analysis. The data from 16 stations of solarimetric network of Paraíba and 3 stations of the National Institute of Meteorology (INEMET- MA) located in the neighbourhood of Paraíba in the states of Pernambuco, Ceará and Rio Grande do Norte were used in this study. The data collection at the above stations during the period of six years (1975 to 1980) for the months of March (in the rainy season) and September (in the dry season) were utilized in this study. Groups of stations with homogeneous characteristics were previously determined. The method of grouping adapted, among the methods tested, was that of Ward which presented satisfactory characteristics of separation into groups and it looks to be the recommendable method for this work. The homogeneous groups of stations were obtained by grouping various types of variables, for example: crude data and standardized data of irradiation. "factor loading" obtained from the eigenvalues which explained 90% of variance and the first the three eigenvalues. The number of appropriate groups for the 19 stations was five, but due to the limitations of data during the period of study, less number of groups were used in some cases.

A simulation of the time series was applied using the data of the group to which Barra de Santa Rosa (BSR) and Pombal (POM) pertain. Applying the method developed by CEBALLOS and BRAGA (1991) which is based on the matrix of the autovectors associated to the correlation between the stations of a homogeneous group. For the pertinent calculations, only the principal autovectors of the matrix (which explained 90% of the variance and the first two autovectors) were utilized. The simulations present good results of deviation of the order of 0,45 of the simulated series with relation to the original series. The magnitude of the derivation is of the order of the error encountered in the irradiation data of the stations studied.

LISTA DE FIGURAS

	Páginas
Figura 3.1 - Dendrograma obtido pelo método de Ward para o mês de março 1975	13
Figura 3.2 - Dendrograma obtido pelo método de Centróide para o mês de março 1975	14
Figura 4.1 - Mapa do estado da Paraíba	30
Figura 4.2 - Carta de irradiação média e desvio padrão do mês de março (1975 a 1980)	37
Figura 4.3 - Carta de irradiação média e desvio padrão do mês de setembro (1975 a 1980)	37
Figura 5.0.a - Estações agrupadas pelo método de Ward levando em consideração cinco grupos março 1975	46
Figura 5.0.b - Estações agrupadas pelo método de Ward levando em consideração cinco grupos setembro 1975	46
Figura 5.1 - Séries real e simulada de BSR para março de 1975	49
Figura 5.2 - Séries real e simulada de BSR para março de 1976	49
Figura 5.3 - Séries real e simulada de BSR para março de 1977	51
Figura 5.4 - Séries real e simulada de BSR para março de 1978	51
Figura 5.5 - Séries real e simulada de BSR para março de 1979	54
Figura 5.6 - Séries real e simulada de BSR para março de 1980	54
Figura 5.7 - Séries real e simulada de BSR para setembro de 1975	59
Figura 5.8 - Séries real e simulada de BSR para setembro de 1976	59
Figura 5.9 - Séries real e simulada de BSR para setembro de 1977	61
Figura 5.10 - Séries real e simulada de BSR para setembro de 1978	61
Figura 5.11 - Séries real e simulada de BSR para setembro de 1979	65
Figura 5.12 - Séries real e simulada de BSR para setembro de 1980	65
Figura 5.13 - Séries real e simulada de POM para março de 1975	68
Figura 5.14 - Séries real e simulada de POM para março de 1976	68
Figura 5.15 - Séries real e simulada de POM para março de 1977	70
Figura 5.16 - Séries real e simulada de POM para março de 1978	70
Figura 5.17 - Séries real e simulada de POM para março de 1979	73

	Página
Figura 5.18 - Séries real e simulada de POM para março de 1980	73
Figura 5.19 - Séries real e simulada de POM para setembro de 1975	76
Figura 5.20 - Séries real e simulada de POM para setembro de 1976	76
Figura 5.21 - Séries real e simulada de POM para setembro de 1977	78
Figura 5.22 - Séries real e simulada de POM para setembro de 1978	78
Figura 5.23 - Séries real e simulada de POM para setembro de 1979	81
Figura 5.24 - Séries real e simulada de POM para setembro de 1980	81
Figura 5.25 - Série real, simulada e dos desvios de BSR para setembro de 1979	88

LISTA DE TABELAS E QUADROS

	Página
Tabela 3.1 - Tabela ilustrativa dos trabalhos em meteorologia com ACP	05
Tabela 4.1 - Irradiação média mensal <Q> para o mês de março de 1975 a 1980	32
Tabela 4.2 - Irradiação média mensal <Q> para o mês de setembro de 1975 a 1980	33
Tabela 4.3 - Desvio padrão S para o mês de março de 1975 a 1980	34
Tabela 4.4 - Desvio padrão S para o mês de setembro de 1975 a 1980	35
Tabela 6.1 - Correlação entre as estações do grupo para o mês de setembro nos períodos de 1975 a 1980 e 1987 a 1989	87

TABELAS EM APÊNDICE

Tabela 5.1 - Composição dos grupos que incluem BSR, e simulações março 1975	97
Tabela 5.2 - Composição dos grupos que incluem BSR, e simulações março 1976	98
Tabela 5.3 - Composição dos grupos que incluem BSR, e simulações março 1977	99
Tabela 5.4 - Composição dos grupos que incluem BSR, e simulações março 1978	100
Tabela 5.5 - Composição dos grupos que incluem BSR, e simulações março 1979	101
Tabela 5.6 - Composição dos grupos que incluem BSR, e simulações março 1980	102
Tabela 5.7 - Composição dos grupos que incluem BSR, e simulações setembro 1975	103
Tabela 5.8 - Composição dos grupos que incluem BSR, e simulações setembro 1976	104
Tabela 5.9 - Composição dos grupos que incluem BSR, e simulações setembro 1977	105
Tabela 5.10 - Composição dos grupos que incluem BSR, e simulações setembro 1978	106
Tabela 5.11 - Composição dos grupos que incluem BSR, e simulações setembro 1979	107
Tabela 5.12 - Composição dos grupos que incluem BSR, e simulações setembro 1980	108
Tabela 5.1.1 - Composição dos grupos que incluem POM, e simulações março 1975	109

	Página
Tabela 5.2.1 - Composição dos grupos que incluem POM, e simulações março 1976	110
Tabela 5.3.1 - Composição dos grupos que incluem POM, e simulações março 1977	111
Tabela 5.4.1 - Composição dos grupos que incluem POM, e simulações março 1978	112
Tabela 5.5.1 - Composição dos grupos que incluem POM, e simulações março 1979	113
Tabela 5.6.1 - Composição dos grupos que incluem POM, e simulações março 1980	113
Tabela 5.7.1 - Composição dos grupos que incluem POM, e simulações setembro 1975	114
Tabela 5.8.1 - Composição dos grupos que incluem POM, e simulações setembro 1976	115
Tabela 5.9.1 - Composição dos grupos que incluem POM, e simulações setembro 1977	116
Tabela 5.10.1 - Composição dos grupos que incluem POM, e simulações setembro 1978	117
Tabela 5.11.1 - Composição dos grupos que incluem POM, e simulações setembro 1979	118
Tabela 5.12.1 - Composição dos grupos que incluem POM, e simulações setembro 1980	118

QUADROS

Quadro 1	Estações utilizadas no trabalho	29
Quadro 2	Métodos de agrupamentos e variáveis utilizadas no trabalho	47

1 - INTRODUÇÃO.

Em estudos climatológicos, é necessário conhecer a série histórica dos dados da variável objeto da pesquisa. Entretanto, as estações meteorológicas responsáveis pela coleta e armazenamento das informações nem sempre dispõem de séries completas de dados. Vários são os fatores que contribuem, como por exemplo a pane de instrumentos, falta do observador, desativação temporária da estação etc.

Se não bastassem os exemplos, a dificuldade de manter as estações meteorológicas no Brasil tem contribuído para a desativação de algumas delas, interrompendo assim, as séries temporais de algumas localidades deixando-as sem dados meteorológicos caracterizando uma perda tanto para o trabalho operacional como para o de pesquisa. Visando amenizar estes problemas propõe-se através deste trabalho examinar a possibilidade de recompor séries de irradiação global da rede solarimétrica da UFPB. Outras variáveis meteorológicas podem ser processadas de forma semelhante ao método a utilizar.

Empregou-se a metodologia proposta por CEBALLOS e BRAGA (1991), assim como as de (MARDIA, 1988) que incluem a Técnica de Análise de Agrupamento (TAA) e a Análise de Componentes Principais (ACP). Ambas constituem ferramentas importantes de grande utilidade, tanto para identificar regiões homogêneas do ponto de vista espacial e temporal, como para reduzir o número de variáveis que descrevem adequadamente séries temporais de uma região.

Uma das vantagens do método é que, mesmo onde não existam estações meteorológicas, podem ser geradas séries temporais de qualquer variável utilizando dados de estações vizinhas. Exige-se para isso que as estações de referência devam apresentar correlação razoável em suas séries temporais de forma a facilitar a homogenização das mesmas em grupos.

Entretanto, o método comentado foi utilizado para séries de dados em intervalos curtos (um mês). O presente trabalho se propõe estender a análise a um intervalo maior (seis anos), necessitando para isso verificar a estabilidade da qualidade da simulação ano-a-ano.

A verificação positiva permitiria reconstruir séries de estações da Rede da UFPb que, por algum motivo foram desativadas.

2 - OBJETIVOS

2.1 - Objetivo geral

Contribuir para o melhoramento do acervo de dados da Rede Solarimétrica da UFPB, e tentar descobrir métodos para um possível preenchimento de falhas existentes em suas séries de irradiação solar global diária com base nas técnicas objetivas de Análise de Agrupamento e Análise de Componentes Principais.

2.2 - Objetivos específicos

Pretende-se alcançar os seguintes objetivos específicos:

- Aplicar a Análise de Agrupamento aos dados de irradiação global para todas as estações da rede, como também aos fatores obtidos da aplicação de ACP aos dados reduzidos visando identificar regiões da Paraíba com comportamento homogêneo.
- Examinar a série temporal de duas estações da rede, verificando a possibilidade de preencher faltas de dados com base naqueles disponíveis de estações circunvizinhas.

3 - FUNDAMENTAÇÃO TEÓRICA

Quando se deseja informações de um conjunto de dados de uma certa variável ou às vezes de mais de uma no conjunto total dos dados de uma região, é usual recorrer-se à Análise Multivariada. Esta é uma técnica estatística usada para reduzir ao máximo o número de variáveis envolvidas em um problema com uma pequena perda de informações. Os principais objetivos desta técnica são:

- Reduzir a dimensão de uma matriz de dados.
- Investigar o comportamento espacial e temporal das variáveis consideradas.
- Obter grupos homogêneos das variáveis.

A Análise Fatorial é uma das técnicas usadas em Análise Multivariada que também contém, além de outras, a Análise de Componentes Principais (ACP) e as técnicas de Análise de Agrupamentos. Estas técnicas tem sido usadas há bastante tempo por vários pesquisadores em diferentes áreas do conhecimento como por exemplo biologia, ciências sociais, metalurgia, medicina e outras.

Em meteorologia a ACP foi usada primeiramente por LORENZ (1956) e depois por vários outros pesquisadores, conforme ilustrado pelos exemplos da tabela 3.1 a seguir.

Os métodos partem de uma matriz de dados $X(n \times p)$ cujas colunas correspondem a "p" unidades (indivíduos) fornecendo "n" características (valores numéricos) cada. Essas unidades podem ser estações meteorológicas, no caso de estudos climatológicos. As características seriam dados meteorológicos como pressão, temperatura, vento, irradiação etc, ordenados ou não segundo sequências cronológicas. No caso de uma única variável, estes dados podem constituir uma sequência cronológica, onde cada linha representa o valor da variável para um dia em um dia "n" no conjunto de "p" locais.

Tabela 3.1

Tabela ilustrativa dos trabalhos em meteorologia, autores e técnicas utilizadas em vários países.

variável	autor/ano	local	tec.
Precipitação	ISANTA (1984)	Brasil	ACP
Temp. e Vento	JUSTI et al. (1990)	Brasil	ACP
Irradiação	CEBALLOS e BRAGA (1991)	Brasil	TAA-ACP
Temp. Pot. Equiv.	JUSTI et al. (1992)	Brasil	ACP
Precipitação	PANDZIC (1988)	Iugoslávia	ACP
Alt. Geopotencial	TERRON (1990)	Cuba	ACP
Vento	BARROS e ESTEVAN (1983)	Argentina	ACP
I.S.S.P (*)	EDER, DAVIS e MANAHAM (1987)	E.U.A	ACP
P.T.P (**)	KUTZBACH (1967)	E.U.A	ACP
Precipitação	STIDD (1967)	E.U.A	ACP
Precip. e Temp.	BECKER (1992)	Brasil	TAA-ACP

(*) índice de seca severa de Palmer

(**) pressão ao nível do mar, temperatura da superfície e precipitação

Cada coluna forma uma série temporal da variável em estudo para o local (coluna) dado.

A matriz \mathbf{X} de dados é organizada da seguinte forma:

$$\mathbf{X}(n \times p) = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = (X_1, X_2, \dots, X_p)$$

Observe-se que a matriz $\mathbf{X}(n \times p)$ pode ser interpretada como um ordenamento de p vetores colunas (cada vetor equivalente à série temporal de cada uma das p estações), ou de n vetores linha (cada vetor indicando valores de uma determinada rede de estações numa época dada). O primeiro caso descreve principalmente o comportamento temporal de uma rede, enquanto que o segundo ponto de vista ilustra a situação espacial da variável em cada época.

3.1 - Análise de agrupamentos (cluster analysis)

Dado um conjunto de unidades, conhecidas somente por uma listagem de seus caracteres, pretende-se encontrar a melhor maneira de descrever seus padrões de similaridades mútuas. Um dos métodos utilizados com esta finalidade é a análise de agrupamentos (clusters analysis). Esta técnica não presuppõe a existência de grupos, os diversos passos desta ampla metodologia objetivam transformar um conjunto heterogêneo de unidades em grupos que se caracterizam pela homogeneidade interna e pela heterogeneidade externa. Conforme EVERITT (1979), classificação pode ser entendido como um processo para localizar entidades em classes inicialmente indefinidas, de modo que indivíduos da mesma classe sejam

similares, entre si, em algum sentido definido pelas variáveis consideradas. Essas classes de indivíduos similares serão os agrupamentos (CURI, 1982).

Existem várias tipos de técnicas de análise de agrupamento (TAA), a definida por EVERITT (1974) é a seguinte:

- 1 - Técnicas Hierárquicas
- 2 - Técnicas de Partição - Otimização
- 3 - Técnicas de Densidade
- 4 - Técnicas de Agrupamento com interseção
- 5 - Outras (técnicas que não se encaixam claramente dentro de qualquer das quatro técnicas prévias).

As técnicas de análise de agrupamentos podem ser as hierárquicas e as não-hierárquicas. Nas técnicas não hierárquicas efetua-se uma partição em um número de classes fixado a priori, e nas hierárquicas são produzidas seqüências de partições em classes cada vez mais vastas. As técnicas hierárquicas de agrupamento podem ser divididas em dois tipos:

- Aglomerativas, onde através de fusões sucessivas dos n elementos vão sendo obtidos $n-1$, $n-2$, etc, grupos, até reunir todos os elementos num único grupo; ou seja, no principio cada elemento forma um grupo individual e no final todos se fundem formando um único grupo.

- Divisivas, partem de um único grupo, e por divisão sucessiva vão sendo obtidos 2, 3, etc grupos.

O que caracteriza estes processos é que a reunião de dois grupos numa certa etapa, produz um dos agrupamentos da etapa superior, caracterizando o processo hierárquico.

3.1.1 - Função de agrupamento.

Na análise de agrupamentos, todos os processos de hierárquização são similares, iniciando-se pela determinação de uma função de agrupamento. Esta função é usada como critério para medir a distância entre dois objetos ou para quantificar o quanto eles são parecidos. Alguns autores chamam-na de coeficiente de parença, que pode se dividir em duas categorias: medida de similaridade e de dissimilaridade. Na primeira, quanto maior o valor observado mais parecidos são os objetos. Já para a segunda, quanto maior o valor observado menos parecidos (mais dissimilares) serão os objetos. Um exemplo de medida de similaridade é o coeficiente de correlação, e de dissimilaridade é a distância euclidiana. A maioria dos algoritmos utilizam medidas de similaridade e dissimilaridade entre os elementos de um conjunto, cujos os valores de similaridade (dissimilaridade) calculados com base na matriz de dados originais X são representados por uma matriz de distâncias ou de similaridade denominada matriz de parença.

3.1.2 - Medidas derivadas da distância euclidiana.

Considere-se um vetor \vec{X} de coordenadas (x_1, x_2, \dots, x_p) como descritor dos objetos que serão investigados os assemelhamentos. A medida mais conhecida para indicar a proximidade entre dois objetos A e B é a distância euclidiana definida por:

$$d(A,B) = \left[\sum_{i=1}^p (x_i(A) - x_i(B))^2 \right]^{1/2} \quad (3.2)$$

Quando se usam grandezas não diretamente comparáveis (com unidades e naturezas diferentes), a mudança de uma das unidades de medida pode alterar completamente o significado e o valor do coeficiente de parença. Essa é

uma das razões da padronização ou redução das variáveis. A redução é feita da seguinte forma: \vec{X} é transformado num outro vetor \vec{Z} , tal que

$$\vec{Z} = \frac{x_i(.) - \bar{x}_i}{s_i} \quad (3.3)$$

onde \bar{x}_i e s_i indicam respectivamente a média e desvio padrão da i -ésima coordenada. Feita a transformação a distância euclidiana passa a ser:

$$d(A,B) = \left[\sum_{i=1}^p (z_i(A) - z_i(B))^2 \right]^{1/2} \quad (3.4)$$

ou, substituindo (3.3) em (3.4), temos:

$$d(A,B) = \left[\sum_{i=1}^p \left(\frac{x_i(A) - x_i(B)}{s_i} \right)^2 \right]^{1/2} \quad (3.5)$$

que é a média dos desvios quadraticos dos padronizados.

3.1.3 - Métodos hierárquicos de agrupamento.

Nos métodos hierárquicos, existem várias técnicas ou critérios de agrupamentos. Dentre eles podemos destacar o método da Centróide, o da Ligação Simples, Ligação Completa e Ward. Estes métodos serão descritos aqui por terem sido utilizados na parte experimental deste trabalho.

1 - Método da Ligação Simples ou vizinho mais próximo (Single linkage method)

Este método usa como medida de parença entre dois grupos tanto a similaridade quanto a distância. Os grupos inicialmente constituído de grupos individuais (onde cada objeto é um grupo) se fundem de acordo com a distância

entre seus membros mais próximos. Os grupos com menor distância se fundem primeiro. Por este método, então, a distância entre grupos é definida como a distância entre seus membros mais próximos. Assim, dados os conjuntos de objetos X e Y , a distância entre eles será definida por:

$$d(X,Y) = \min \{ d(i,j): i \in X \text{ e } j \in Y \}. \quad (3.6)$$

No caso de similaridade(definida por s),

$$s(X,Y) = \max \{ s(i,j): i \in X \text{ e } j \in Y \}$$

2 - Método da Ligação Completa ou vizinho mais distante (Complete Linkage Method)

Este método é exatamente o oposto do da Ligação Simples, apesar de usar tanto a medida de distância quanto a de similaridade para definir a parecença. Aqui os grupos são formados fundindo os membros mais distantes entre os grupos. No caso da parecença ser definida pela distância, esta será calculada por:

$$d(X,Y) = \max \{ d(i,j): i \in X \text{ e } j \in Y \} \quad (3.8)$$

3 -Método da Centróide

Este processo é o mais direto, pois substitui cada fusão de objetos por um único ponto representado pelas coordenadas de seu centro. A distância entre os grupos é definida pela distância entre os centros. Em cada etapa procura-se fundir grupos que tenham a menor distância entre si.

4 - Método de Ward

Dado um conjunto de N indivíduos definidos por uma variável x , definimos a inércia desse conjunto como:

$$I = \sum_{i=1}^n X_i^2 - \frac{1}{n}(\sum X_i)^2 = n(\bar{X}^2 - \bar{X}^2) = \sum (X_i - \bar{X})^2 \quad (3.9)$$

onde: \bar{X} = média da variável X , e \bar{X}^2 = média de seu valor quadrático.

Se os indivíduos estão distribuídos em grupos, a inércia do conjunto é avaliada como a soma das inércias de cada grupo com relação à sua própria média. Ao se considerar uma situação inicial de N grupos individuais, a inércia total é nula.

Segundo EVERITT (1974), WARD em 1963 propôs que em qualquer estágio de uma análise a perda de informações que resulta do agrupamento de indivíduos em grupos pode ser medido pelo incremento da inércia do conjunto. Em cada passo da análise, a união de todos os possíveis pares dos grupos será considerada, e escolhida aquela que resulte na menor perda de informação total. Os dois grupos cuja fusão resulte no mínimo incremento da inércia são combinados. A equação (3.9) é aplicada diretamente aos dados para se obter os valores do nível de agregação.

A evolução da inércia em etapas sucessivas de agrupamentos, pode ser utilizada para definir o número de grupos em que um conjunto de indivíduos deveria dividir-se. Em geral, observa-se que a inércia em alguma sequência da agregação aumenta rapidamente indicando a partir daí a existência de um número de grupos a ser obtido. Tem-se portanto, que esta transição pode ser adotada para auxiliar na

determinação do número de grupos. Entretanto, nem sempre essa transição está bem definida; nesse caso algum critério subjetivo deve ser adotado.

3.1.4 - Um exemplo de análise.

As etapas descritas anteriormente, embora instrutivas acerca do processo de agrupar não facilitam a interpretação dos resultados. Necessita-se de instrumentos mais apropriados, e um deles é o dendrograma, que é uma representação gráfica muito usada na AA para apresentar o resultado de um agrupamento. As figuras 3.1 e 3.2 mostram dendrogramas obtidos em diferentes métodos de agrupamentos considerando como variável os fatores que explicaram 90% da variância total dos dados da rede, e como medida de proximidade a distância euclidiana. A escala vertical à esquerda indica o nível de similaridade, pois a cada partição corresponde um valor numérico que representa o nível no qual ocorre os agrupamentos; quanto mais elevado o índice, mais heterogêneas as partes agrupadas. Esse índice é também chamado nível de agregação. No eixo horizontal são marcados os elementos em uma ordem conveniente de formação dos grupos.

Conhecendo-se o dendrograma, pode-se determinar o número de grupos homogêneos de duas maneiras. Uma delas é através de cortes transversais no dendrograma feito de forma subjetiva e a critério do pesquisador, outra é pelo critério matemático da inércia onde o número de grupos é obtido pela curva do gráfico da inércia.

Figura-3.1

Dendrograma obtido pelo método de Ward para o mês de março de 1975

NORE

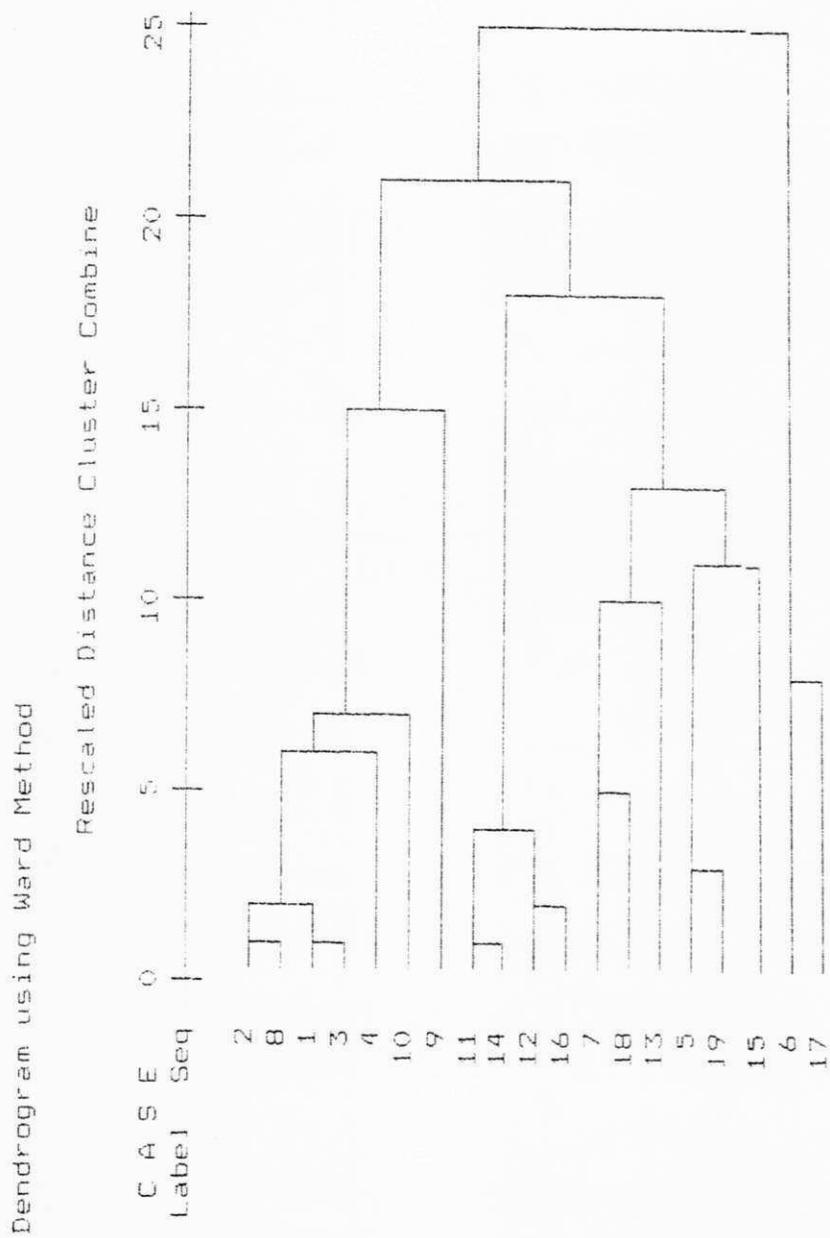
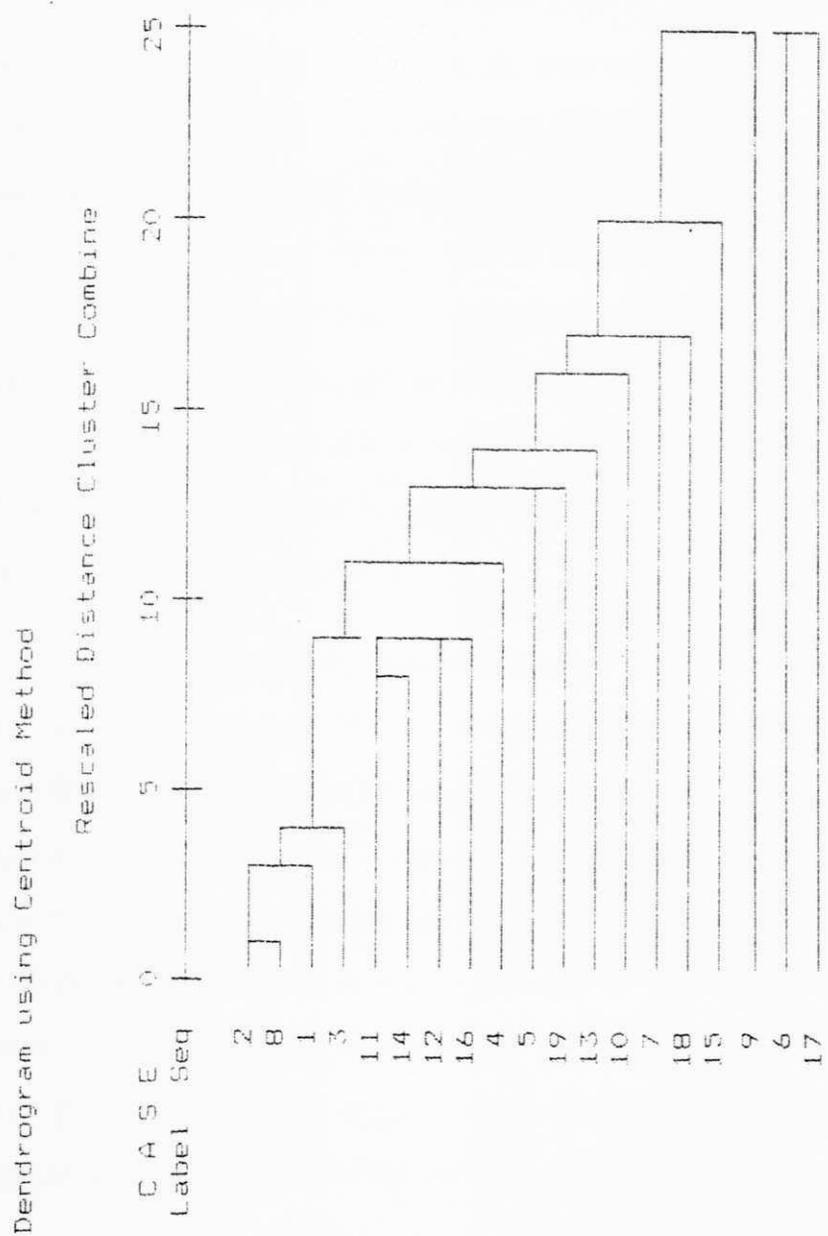


Figura-3.2

Dendrograma obtido pelo método da centroide para o mês de março de 1975

MORE



3.2 - Análise de componentes principais (A C P).

A idéia central da Análise de Componentes Principais - (ACP, no que segue) é reduzir a dimensão de um conjunto de dados composto de um grande número de variáveis interrelacionadas conservando as variações presentes no conjunto de dados. Isto é possível transformando-o em um novo grupo de variáveis, os componentes principais, os quais são não correlacionados. Introduzida por LORENZ (1956) em estudos meteorológicos e por ele denominada análise de funções ortogonais empíricas, parte de uma matriz de dados com dimensão $n \times p$ constituída de n "variáveis" medidas sobre p "indivíduos". Em análise climatológica, pode-se efetuar estudos no modo-T, sendo neste caso o tempo considerado como variável (meses do ano por exemplo) e os indivíduos como os locais de observação (estações meteorológicas) onde são obtidos os valores das variáveis meteorológicas em estudo. Pode-se ainda efetuar estudos no modo-S; neste caso, as "variáveis" seriam os locais de observação e os "indivíduos" poderiam ser os instantes de tempo (meses ou dias sucessivos) nos quais são observadas as grandezas meteorológicas analisadas (SILVA et al, 1992).

O enfoque da técnica de análise de componente principais, adotado por BOUROCHE e SAPORTA (1980) estabelece que a informação associada a M indivíduos (constituída por N informações) é representado por uma nuvem de N pontos num espaço vetorial Γ com dimensão M . A ACP consiste na projeção dessa nuvem de pontos situada em um subespaço de dimensão P de forma que sua projeção sofra a menor deformação possível. O plano sobre o qual são efetuadas essas projeções é denominado plano principal e a direção chamada direção principal. CEBALLOS e BRAGA (1991) representam a nuvem, considerando como direção principal em Γ aquela definida por uma reta de quadrados mínimos. A

solução do problema conduz naturalmente a M direções principais. Esta técnica transforma as variáveis que inicialmente são correlacionadas, em outras não correlacionadas entre si.

3.2.1 - Fundamentação estatística da A C P.

Seja Y_t , $t = 1, 2, \dots, N$, uma série temporal de uma variável meteorológica, reduzida para um determinado local representada pelo vetor coluna $\vec{Y} = (Y_1, Y_2, \dots, Y_N)_t$. Considerando uma série conjunta de M locais (estações meteorológicas) de uma região, a evolução temporal do vetor linha virá descrito pelo vetor $\vec{Y} = (Y_1, Y_2, \dots, Y_M)_t$. A matriz de dados para as M estações para um certo período de tempo (dias do mês por exemplo) será:

$$Y = \begin{pmatrix} Y_{11} & Y_{12} & \dots & Y_{1m} & \dots & Y_{1M} \\ Y_{21} & Y_{22} & \dots & Y_{2m} & \dots & Y_{2M} \\ \vdots & \vdots & & \vdots & & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{nm} & \dots & Y_{nM} \end{pmatrix}$$

A matriz de covariância de Y pode ser obtida por:

$$V = \frac{1}{n} Y^t Y \quad (3.10)$$

onde Y^t é a transposta da matriz Y . Se os dados da matriz Y estiverem centrados e normalizados, a matriz de covariância será igual à matriz de correlação R .

$$V = R = \frac{1}{n} Y^t Y \quad (3.11)$$

Sendo R uma matriz simétrica $M \times M$, pode ser diagonalizada por uma matriz ortogonal A , denominada matriz de direções principais. A matriz diagonal D cujos elementos diagonais são autovalores de R , é obtida pela equação.

$$\mathbf{D} = \mathbf{A}^{-1} \mathbf{R} \mathbf{A} \quad (3.12)$$

onde \mathbf{A}^{-1} é a matriz inversa de \mathbf{A} .

Como \mathbf{A} é uma matriz ortogonal, a matriz inversa é igual à transposta $\mathbf{A}^{-1} = \mathbf{A}^t$; assim,

$$\mathbf{D} = \mathbf{A}^t \mathbf{R} \mathbf{A} \quad (3.13)$$

A matriz \mathbf{A} descreve um operador mudança de base a um novo sistema de referência composto pelos autovetores normalizados de \mathbf{R} . Os componentes principais $(z_1, z_2, \dots, z_m, \dots, z_M)_t$ das variáveis $y_1, y_2, \dots, y_m, \dots, y_M$, compõem imagem dos vetores \vec{Y} logo,

$$\vec{Z} = \mathbf{A}^t \vec{Y} \quad (3.14)$$

$$\vec{Y} = \mathbf{A} \vec{Z} \quad (3.15)$$

Os valores y_m do m -ésimo local (estação meteorológica) serão reproduzidos exatamente pela expressão:

$$y_m = a_{m1}z_1 + a_{m2}z_2 + \dots + a_{mk}z_k + \dots + a_{mM}z_M \quad (3.16)$$

As variáveis $z_1, z_2, \dots, z_m, \dots, z_M$ são os componentes principais do conjunto de vetores \vec{Y} . Três características são importantes:

1) O conjunto (z_1, z_2, \dots, z_M) evolui no tempo; isto é: para cada instante t existe um conjunto diferente.

2) A variância do componente z_k (calculado com base em médias temporais) vale:

$$\langle z_k \rangle = 0, \text{Var}(z_k) = \langle z_k^2 \rangle = \lambda_k$$

onde λ_k é o k-ésimo autovalor

3) Ordenandos os autovalores λ_k em ordem decrescente percebe-se que as variáveis y_m podem ser aproximadas considerando na expressão (3.16) os primeiros $K < M$ termos, de forma que os K primeiros componentes z_k acumulem uma porcentagem desejada da variância total. É neste sentido que a ACP permite reduzir o número inicial de m variáveis (Y_m), a um número $K < M$ de componentes significativos.

Em particular, a TAA permite escolher grupos considerados homogêneos, incluindo cada um número $k_j < M$ suficiente para descrever o comportamento de uma variável na região homogênea. Assim, a terceira característica mencionada acima permite, através da ACP aplicada a cada grupo em particular, aumentar ainda mais, estas vantagens da TAA e ACP combinadas.

3.2.2 - Alguns exemplos de aplicação da ACP em meteorologia.

Após o trabalho de LORENZ (1956) usando ACP em meteorologia em que ele chamou de FOE (funções ortogonais empíricas) usando-as com o objetivo de previsores climáticos tendo como variável a pressão ao nível do mar vários trabalhos usando esta técnica foram feitos no Brasil com os mais diversos objetivos. Não só LORENZ trabalhou com funções ortogonais empíricas, mas também ALDAZ (1984) aplicou também a técnica à precipitação do Nordeste do Brasil visando calcular as direções principais da variação cronológica dos totais medios mensais em trinta anos. Considerando uma matriz de dados composta por

dozes variáveis (meses) e cento e uma unidades (estações meteorológicas) as quatro primeiras direções principais explicaram 97% da variância total da série.

Visando mais especificamente o uso de ACP na redução de variáveis meteorológicas no estudo climático de certas regiões e a recomposição de séries de dados de estações meteorológicas podemos destacar os seguintes trabalhos:

"Aplicacion de los componentes principais al analisis de los campos meteorológicos".

MARIA E. PÉREZ TERRÓN (1990)

Neste trabalho a autora usa a técnica de componentes principais para reduzir o número de variáveis meteorológicas no estudo de previsão do tempo para uma determinada área (entre 20 e 40° de latitude norte e entre 70 e 110° de longitude oeste com intervalos de 10°). Os dados reduzidos foram a altura geopotencial de 700 hPa e a precipitação. A técnica usada para escolha do número de autovalores foi a proposta por DYER (1975) e MARDIA (1979) que estabeleceram aplicar o critério de Kaiser, o qual leva a reter só aqueles autovalores que são maiores que a unidade.

De acordo com a análise realizada, chegou-se a conclusão de que esta técnica pode ser aplicada a problemas de classificação e otimização de dados meteorológicos para diferentes fins, levando a resultados objetivos que não estão em discrepância com as características físicas dos campos estudados.

"On the evaluation of wind power from short wind records"

VICENTE R. BARROS and EDUARDO A. ESTEVAN (1993)

Para julgar a conveniencia de se instalar em um local um sistema de conversão de vento em energia é necessária uma série longa do potencial do vento de um local. Como nem sempre é possível ter essa série, os autores propuseram um método para estimar o potencial médio do vento de um determinado local e ano baseado em somente três meses de dados. Este método esta baseado no proposto por BARROS e SERÓ (1981) de estimar a velocidade média anual do vento e suas variações baseado em um tempo de medição de somente dois ou três meses. Este método fundamenta-se no presuposto de que a velocidade do vento tem alguns graus de correlação espacial, e que não depende de uma só estação mas também de informação de vários locais com afinidade climatológica entre as mesmas. Por isso a técnica foi chamada de técnica de redução climatológica múltipla (TRCM). Esta técnica está baseada na análise de componentes principais como segue:

Sendo V uma matriz de dados com m colunas. É sempre possível descrever V como:

$$V = AW \quad (1)$$

onde A = matriz de autovetores

A expressão (1) é equivalênte a dizer que o item i da série j pode ser escrito.

$$V_{ij} = \sum_{k=1}^k a_{ik} w_{kj}, \quad i=1 \dots R, \quad j=1 \dots M \quad (2)$$

onde R = número de i -termos na série

k = número de autovetores

Se as séries não são todas independentes, uma representação razoavelmente acurada pode ser conseguida por:

$$V_{ij} = \sum_{t=1}^N a_{it} w_{jt}, \quad i=1 \dots N, \quad j=1 \dots M \quad (3)$$

onde $N < R$. A diferença entre o valor verdadeiro e o representado por (3) é um certo ε_{ij} .

A técnica explanada aqui, considera o caso no qual para uma matriz particular $j=q$ na região, existe somente um curto periodo de dados.

$V_{iq} = L \dots P$, onde $L \geq 1$ e $P \leq R$. Em tal caso, V_{iq} pode também ser expressa por (3). Os valores de W_{kq} são encontrados minimizando

$$F_q = \sum_{i=L}^P \varepsilon_{i,q}^2$$

Isto conduz a um sistema linear de N -equações com N -incógnitas W_{kq} , $K = 1 \dots N$, o qual pode ser facilmente resolvido. Portanto, V_{iq} é virtualmente extendida desde $i = L \dots P$ para $i = 1 \dots R$. Esta série extendida expressa os dados completos.

"Simulação de séries temporais de irradiação solar: Uma aplicação ao Estado da Paraíba"

JUAN C. CEBALLOS e CÉLIA C. BRAGA (1991)

Visando dotar alguns locais de uma região de séries temporais de irradiação. Os autores, propõem recompor dados faltosos de estações existentes baseados em informações de estações vizinhas, as quais poderiam apresentar elevado coeficiente de correlação entre elas e a estação em análise.

São usadas técnicas de análise multivariada com a determinação de grupos de estações homogêneas, bem como a aplicação de análise de componentes principais. Foi feito um estudo usando dados da Rede Solarimétrica da UFPb e de estações vizinhas pertencentes à rede do INMET-MA.

O teste de simulação foi feito para os meses de fevereiro e agosto do ano de 1976, para as estações de PAT e PIC com uma matriz de autovetores obtida dos autovalores que explicaram aproximadamente 90% da variância (os 6 primeiros explicaram 89% para o mês de fevereiro, e 87% para agosto) e não só os que tinham valores maiores que a unidade. A simulação das séries foi feita seguindo uma idéia desenvolvida por BARROS e ESTEVAN (1982) que define um método de obter o conjunto de estimadores dos componentes principais pelo método dos mínimos quadrados da diferença entre a parte conhecida da série e sua estimativa.

"Determinação de Regiões Homogêneas de Precipitação e Temperatura no Estado do Rio Grande do Sul".

CARMEM TEREZINHA BECKER (1992)

Apesar do objetivo principal ser a regionalização de subregiões com comportamento homogêneo de precipitação e temperatura, neste trabalho também é

feito um teste de simulação de série climatológica. A região estudada foi todo o estado do Rio Grande do Sul para um período de 30 anos (1931 a 1960) com 41 estações. A simulação foi feita para a estação de Vacaria, pertencente à região serrana do estado, o método adotado foi o proposto por CEBALLOS e BRAGA (1991) e a variável simulada foi precipitação.

3.2.3 - Simulação de séries temporais.

Considere-se uma série temporal y_t de uma variável meteorológica, em que seja desconhecido o dado de um certo dia da m -ésima estação. Como $\mathbf{Z} = \mathbf{A}^t \bar{\mathbf{Y}}$ todos os componentes principais z_k são desconhecidos para este dia, e a incógnita y_m não poderá ser avaliada mediante (3.16).

CEBALLOS e BRAGA (1991) utilizaram um critério de simulação no qual estimadores dos componentes principais podem ser obtidos com base no método dos mínimos quadrados, utilizando a informação disponível das $M-1$ estações restantes. Este critério consiste em definir um vetor $\bar{\mathbf{z}}$ no qual a norma da diferença entre a parte conhecida \mathbf{Y}' do vetor $\bar{\mathbf{Y}}$ e sua estimativa \mathbf{Y}'' seja mínima.

$$D^2 = (\mathbf{Y}'' - \mathbf{Y}') (\mathbf{Y}'' - \mathbf{Y}') = \sum_{m=1}^M (\sum_{k=1}^K a_{mk} z'_k - y^m)^2 = \min \quad (3.17)$$

$$\frac{\partial D^2}{\partial z'_j} = 0 \quad j=1,2,\dots,K$$

onde z'_k são estimadores dos CPs. Nesta equação, supõe-se que o desenvolvimento de \mathbf{Y}'' contém apenas $K < M$ componentes principais que acumulam uma fração suficiente da variância. Ou seja, o vetor $\bar{\mathbf{z}}$ é de dimensão $K < M$.

O critério de mínimos quadrados aplicado à expressão (3.17) fornece um sistema linear algébrico de ordem K para as incógnitas z_k donde resultam as relações

$$\sum_k V_{jk} Z_k = U_j \quad (3.18)$$

$$\mathbf{V} \vec{z} = \mathbf{U} \quad (3.19)$$

A dimensão da matriz \mathbf{V} é $K \times K$, bem menor do que da matriz \mathbf{A} . Para calcular \mathbf{V} pode-se definir uma matriz auxiliar \mathbf{B} com dimensão $(M-1) \times K$ resultante da matriz \mathbf{A} incompleta sem a m -ésima linha e considerando só as primeiras K colunas. A matriz auxiliar \mathbf{U} é construída pelo produto incompleto do vetor \vec{Y} com as colunas da matriz \mathbf{A} :

$$\mathbf{V} = \mathbf{B}^t \mathbf{B} \quad (3.20)$$

onde \mathbf{B}^t é a transposta de \mathbf{B}

Portanto, uma generalização do procedimento para uma série temporal pode ser escrita:

$$\vec{z} = \mathbf{V}^{-1} \mathbf{U} \quad (3.21)$$

Assim, a aplicação de $\mathbf{A}^* \mathbf{Z}^*$ reproduz aproximadamente as componentes conhecidas da série temporal \vec{Y} , e permite também estimar a série desconhecida.

3.2.4 - Séries sintéticas.

A série sintética de uma variável meteorológica qualquer, consiste no estabelecimento de uma série desta grandeza para um determinado local onde não

existe estação meteorológica. A fundamentação matemática é a mesma utilizada nas recomposições de séries com dados faltosos já descrita, mudando apenas a matriz auxiliar **B** que é constituída, no caso da série sintética, da matriz **A** incompleta sem a *m*-ésima linha que será uma estação (criada) obtida através da análise escalar de isolinhas de correlação entre estações contituíntes de grupos homogêneos. A série sintética pode ser também obtida para uma estação que tenha passado um grande período de tempo sem coletar dados, desejando-se recuperar seus dados em um certo período de tempo. Neste caso poder-se-ia usar uma interpolação temporal do coeficiente de correlação do grupo de estações homogêneas para se obter, previamente as outras componentes da matriz **A** referentes a essa estação.

3.3 - Análise fatorial.

A análise de fatores é uma técnica estatística usada para identificar um número relativamente pequeno de fatores que procura-se identificar como responsáveis da relação entre variáveis de uma população. Os fatores configuram um grupo menor que o de variáveis, são independentes e interpretáveis, mas não necessariamente observáveis.

Segundo POSSOLI (1984), nesta técnica o que se faz é substituir um conjunto inicial de p variáveis X_1, X_2, \dots, X_p , correlacionadas, por um conjunto menor de fatores comuns que podem ser não correlacionados (fatores ortogonais) ou correlacionados (fatores oblíquos). Deseja-se determinar um número mínimo de fatores necessários para explicar a maior parte da variância do conjunto original dos dados. Alguma variância que não é explicada pelos fatores comuns pode ser descrita por um termo residual ou "fator específico" da variável.

A análise fatorial como modelo matemático supõe que cada variável observada é representada como uma função linear de um menor número m de fatores comuns, mais uma componente de variação residual (fatores específicos). Assim,

$$X_i = \sum_{j=1}^m \phi_{ij} f_j + \varepsilon_i, \quad i = 1, 2, \dots, p \quad (3.22)$$

onde ϕ_{ij} é a carga fatorial do j -ésimo fator sobre x_i da i -ésima variável e refletem a importância do j -ésimo fator na composição da i -ésima variável.

Os f_j são chamados fatores comuns, e as quantidades ε_i são os fatores específicos, descrevendo a variação residual específica da i -ésima variável

A expressão (3.21) pode ser escrita na forma matricial.

$$\vec{X} = \Phi \vec{F} + \vec{E} \quad (3.23)$$

onde:

\vec{X} é o vetor das variáveis observadas

\vec{F} é o vetor dos fatores comuns

e \vec{E} o vetor dos fatores específicos

Φ é a matriz de cargas fatoriais.

Neste tipo de modelo, o número de fatores pode ser fixado de forma a concentrar uma certa porcentagem da variância total do sistema. O método para avaliar as cargas fatoriais φ_{ij} envolve um conjunto de condições-vinculos, como ortogonalidade entre os fatores e entre as variações residuais. Hipóteses adicionais acerca da distribuição possível de valores (f_j , e_j) podem levar a estimar Φ mediante o critério de máxima verossimilhança. Neste esquema, a recuperação de valores numéricos dos fatores é problemática.

Um outro critério de solução é a ACP. Ela inclui de forma natural a ortogonalidade entre fatores, supõe a existência de P fatores e reduz a zero o valor das variações específicas. Neste caso, usa-se a frase "análise fatorial em componentes principais" (manual do usuário do SPSS/PC "Statistical Package for the Social Sciences"), e os fatores podem ser avaliados diretamente mediante a equação (3.14). Este último é o método utilizado neste presente trabalho. A rotina FACTOR do SPSS (Statistical Package for the Social Sciences) pode ser usada porque realiza análise fatorial em componentes principais.

4 - MATERIAIS E MÉTODOS.

4.1 - Descrição dos dados e área de estudo.

Os dados de irradiação global da Paraíba utilizados neste trabalho foram da Rede Solarimétrica da UFPb. Com relação às estações de outros estados, utilizaram-se dados provenientes do Zoneamento de Potencial Solar do Nordeste (AZEVEDO *et al*,1981) obtidos de estações da Rede Climatológica do INMET-MA. O Quadro 1 a seguir apresenta o nome e a localização de cada estação, e um código composto por três letras abreviando seu nome. Como os dados de irradiação global são obtidos através de integração gráfica de actinogramas por meio de planimetria, espera-se um erro da ordem de 5% em cada estimativa de irradiação diária. Já que o processo de obtenção dos dados foi o mesmo, em todos os casos a precisão é semelhante para dados provenientes das duas Redes.

O estado da Paraíba está localizado na Região Nordeste e limitado ao norte pelo estado do Rio Grande do Norte, ao sul pelo estado de Pernambuco, a leste pelo Oceano Atlântico e ao oeste pelo estado do Ceará. A fig. (4.1) mostra a área estudada. O estado, como praticamente todos os do Nordeste, tem regiões climáticas bem distintas como por exemplo: Brejo, Sertão, Agreste, Zona da Mata e outras.

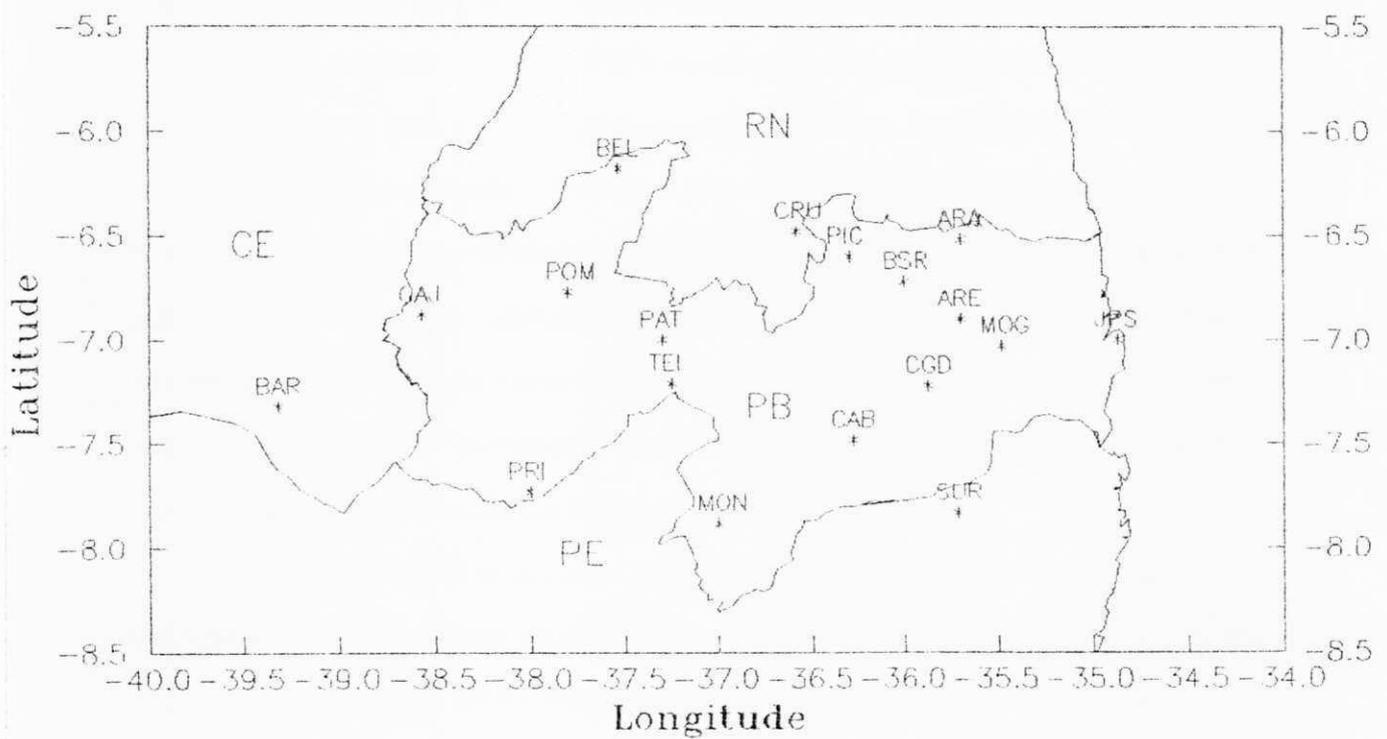
As estações que não pertencem à Rede da UFPb permitem obter informações de fronteira, e estão muito próximas à divisa do Estado. Portanto, a região estudada tem dimensões de 600 Km no sentido Leste - Oeste, e de pouco mais de 200 Km na direção Norte - Sul, cobrindo uma área da ordem de 120.000 Km².

Quadro 1 Estações que foram utilizadas no trabalho. A ordem e a numeração seguem critério alfabético.

CÓDIGO	ESTAÇÃO	LAT	LON	ALT
	Paraíba	(S)	(W)	(M)
01-ARA	ARARUNA	06° 31'	35° 42'	580
02-ARE	ARELA	05° 58'	35° 41'	619
03-BSR	BARRA DE SANTA ROSA	06° 43'	36° 03'	457
04-BEL	BELÉM DO BREJO DO CRUZ	06° 11'	37° 32'	200
05-CAB	CABACEIRAS	07° 29'	36° 17'	390
06-CAJ	CAJAZEIRAS	06° 53'	38° 34'	298
07-CGD	CAMPINA GRANDE	07° 13'	35° 53'	551
08-ESP	ESPERANÇA	07° 01'	35° 51'	634
09-JPS	JOÃO PESSOA	07° 06'	34° 52'	007
10-MOG	MOGEIRO	07° 18'	35° 29'	118
11-MON	MONTEIRO	07° 53'	37° 04'	599
12-PAT	PATOS	07° 00'	37° 18'	249
13-PIC	PICUÍ	06° 30'	37° 07'	440
14-POM	POMBAL	06° 46'	37° 48'	184
15-PRI	PRINCESA ISABEL	07° 44'	38° 00'	683
16-TEI	TEIXEIRA	07° 13'	37° 15'	768
	Ceará			
17-BAR	BARBALHA	07° 19'	39° 18'	408
	Rio Grande do Norte			
18-CRU	CRUZETA	06° 29'	36° 35'	226
	Pernambuco			
19-SUR	SURUBIM	07° 50'	35° 43'	380

Figura 4.1

O estado da Paraíba limitado pelos estados do Rio Grande do Norte, Pernambuco, Ceará e pelo oceano Atlântico



4.2 - Período de estudo

Os dados de irradiação global são coletados tanto nas estações climatológicas principais como em estações agrometeorológicas. Estas estações localizadas em vários pontos do estado funcionam há bastante tempo mas os actinógrafos foram instalados a partir da década de 1970 e a produção de dados tanto de irradiação global quanto difusa pela Rede Solarimétrica da UFPb começou no ano de 1975 e continua até os dias atuais. Por vários motivos, o número de estações tem variado com o tempo; algumas deixaram de funcionar; outras foram transferidas, estações novas surgiram, e com isso não se manteve um número constante de estações para um período extenso de estudo. Devido a esta variação no número de estações, assim como à conveniência de situar um período de estudo com um número razoável de estações, foi escolhido o intervalo que se estende de 1975 a 1980. Neste período o número de estações ainda varia bastante, começando com 19 estações e chegando (para alguns anos) a 7 estações. As tabelas 4.1 a 4.4 apresentam as médias e os desvios padrão de irradiação da rede para os meses propostos.

Das estações da rede foram escolhidas duas para terem suas séries temporais simuladas: Barra de Santa Rosa, na região do Cariri, e Pombal, no alto sertão da Paraíba. Dois meses foram escolhidos: um do período chuvoso da região, e outro do período seco. Segundo dados do Departamento de Recursos Naturais da Divisão de Hidrologia da SUDENE (Dados pluviométricos mensais, Volume II) para o período de 1911 a 1960 a época chuvosa dessas localidades ocorre nos meses de fevereiro, março e abril. O período seco ocorre nos meses de agosto, setembro e outubro. Escolheu-se os meses de março e setembro dos períodos chuvoso e seco respectivamente.

A unidade de irradiação usada neste trabalho é o langley ($1ly = 1 \text{ cal/cm}^2$)

TABELA 4.1 - Irradiação média mensal $\langle Q \rangle$ para o mês de março do período de 1975 a 1980. Unidades langley = cal/cm^2

Código	ANO					
	75	76	77	78	79	80
01-ara	495	446	449	485	518	
02-are	542	488	462	438		
03-bsr	421	454	485	476	474	440
04-bel	539	550	512	523	512	531
05-cab	506	502	531	503	496	
06-caj	441	543	478	535	548	474
07-egd	517		496	444	485	409
08-esp	537	456	533		476	
09-jps	491	461	466		534	529
10-mog	469	465	456		457	
11-mon	525	491	468	458		499
12-pat	548	535	538	537		
13-pic	473	476	458	524	487	
14-pom	505	521	537	543	549	539
15-pri	529	486	477	498	527	
16-tei	496	517	502	439		
17-bar	421	480	457			
18-cru	540	485	482			
19-sur	519	501	499			

TABELA 4.2 - Irradiação média mensal $\langle Q \rangle$ para o mês de setembro do período de 1975 a 1980. Unidades langley = cal/cm²

Código	ANO					
	75	76	77	78	79	80
01-ara	509	439	449	463		
02-are	488	546	409			
03-bsr	467	466	450	447	468	474
04-bel	535	546	507	515	469	478
05-cab	484	512	469		407	533
06-caj	490	578	450	548	553	564
07-egd	445	469	377	427	395	396
08-esp	490	500	432	370	354	
09-jps	497	495	477	490	432	482
10-mog	463	470	447	422		
11-mon	518	537	458	473		484
12-pat	524	537	567	525		
13-pic	535	516	496	481		
14-pom	532	575	515	534	519	
15-pri	548	545	509	536		
16-tei	545	528	497			
17-bar	538	510	491			
18-cru	563	528	527			
19-sur	490		474			

TABELA 4.3 - desvio padrão S para o mês de março do período de 1975 a 1980.
Unidades langley = cal/cm²

Código	ANO					
	75	76	77	78	79	80
01-ara	74	93	93	77	78	
02-are	87	93	80	104		
03-bsr	68	103	78	60	60	77
04-bel	70	71	58	51	72	43
05-cab	64	76	72	103	64	
06-caj	66	68	75	40	49	70
07-cgd	65		72	90	80	80
08-esp	74	106	72		68	
09-jps	109	76	97	71	93	
10-mog	62	103	72		77	
11-mon	61	83	75	85		66
12-pat	56	89	58	69		
13-pic	86	98	67	61	56	
14-pom	61	68	64	46	62	60
15-pri	77	87	98	90	80	
16-tei	84	91	75	108		
17-bar	80	66	128			
18-cru	87	86	59			
19-sur	53	71	81			

TABELA 4.4 - desvio padrão S para o mês de setembro do período de 1975 a 1980. Unidades langley = cal/cm²

Código	ANO					
	75	76	77	78	79	80
01-ara	74	72	71	79		
02-are	92	63	94			
03-bsr	65	72	70	66	86	87
04-bel	68	38	67	93	89	90
05-cab	65	78	91		61	98
06-caj	59	30	70	73	79	53
07-cgd	77	66	75	106	55	78
08-esp	84	63	67	78	67	
09-jps	61	19	81	66	60	75
10-mog	70	57	69	72		
11-mon	81	46	75	94		69
12-pat	67	56	56	95		
13-pic	72	59	63	58		
14-pom	65	34	67	85	83	
15-pri	58	40	74	77		
16-tei	68	56	66			
17-bar	51	63	62			
18-cru	89	48	42			
19-sur	103		84			

embora a recomendação atual seja o uso de unidades do SI, ou seja Joule/m^2 . Como será visto na seção 4.4, o método de trabalho utiliza unidades adimensionais, independentes do sistema de medida.

Isolinhas da distribuição espacial da irradiação média mensal $\langle Q \rangle$ (tabelas 4.1 e 4.2) e do desvio padrão mensal S (tabelas 4.3 e 4.4) são apresentadas nas figuras 4.3 e 4.4 para os meses de março e setembro do período estudado. No traçado das isolinhas, estas devem ser identificadas para facilitar a análise e interpretação dos resultados expostos, por isso, elas devem ser numeradas com valores múltiplos do intervalo entre as isolinhas, o intervalo escolhido para a carta de irradiação foi de 25 ly (langley) que corresponde ao erro médio de medição de Q . Com relação ao desvio padrão, este foi traçado com intervalos de 20 ly .

Figura 4.2 - Carta de irradiação média e desvio padrão do mês de março para o período de 1975 a 1980.

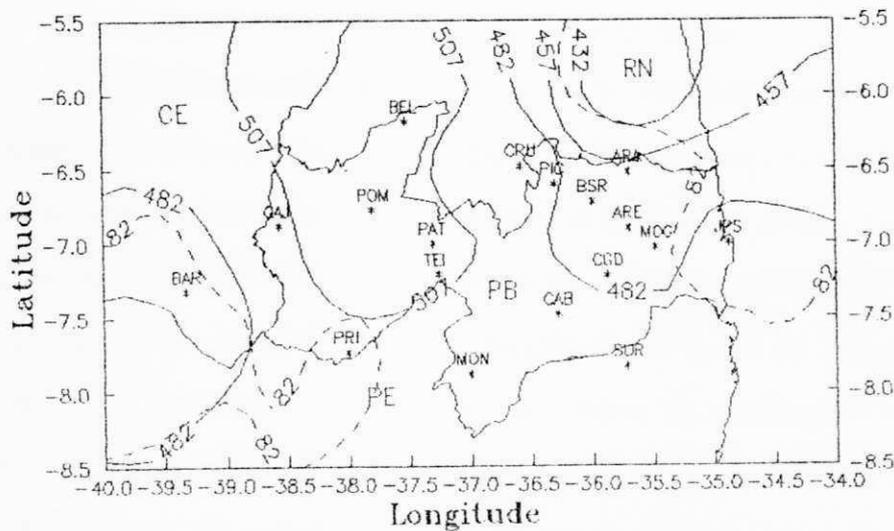
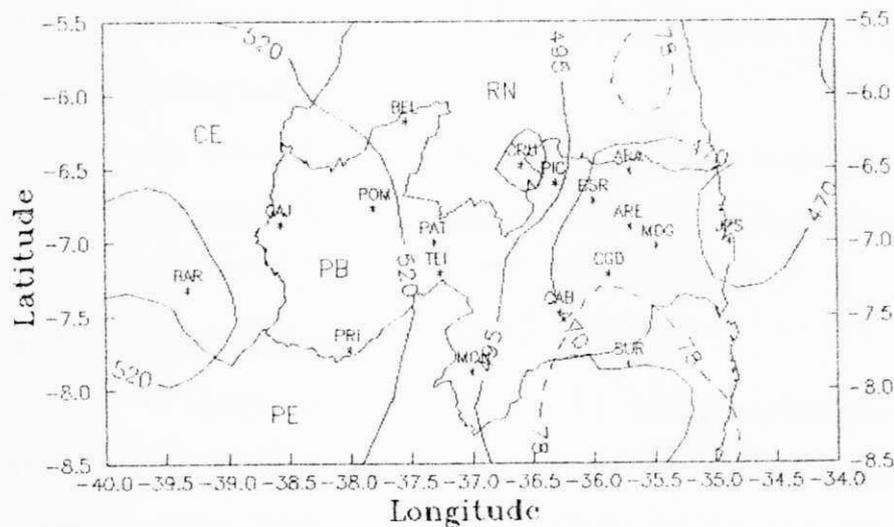


Figura 4.3 - Carta de irradiação média e desvio padrão do mês de setembro para o período de 1975 a 1980.



4.3 - Agrupamento de estações.

4.3.1 - Determinação de Grupos de estações Homogêneas.

Em um processo de simulação de uma série temporal de uma variável qualquer para uma determinada estação, a análise de componentes principais (ACP) pode ser aplicada para o conjunto das estações da rede em estudo. Isto resulta em um grande número de autovalores, visto que as estações que formam a rede não são todas homogêneas entre si. Por isso, escolhe-se um grupo de estações homogêneas entre si que contenha a estação que será objeto da simulação. Além dos vários critérios de agrupamento existentes, as variáveis utilizadas podem ser expressas de várias maneiras. No presente trabalho, examinou-se resultados para:

- 1 - dados brutos do conjunto de estações da rede
- 2 - dados reduzidos deste mesmo conjunto de estações
- 3 - as cargas fatoriais obtidas ao avaliar componentes principais (CP's) dos dados do conjunto de estações da rede ou, ainda, as referentes aqueles que explicam 90% da variância, ou apenas as três primeiras.

Os cálculos necessários foram realizados mediante o subprograma CLUSTERS do utilitário SPSS ("Statistical Package for Social Sciences").

4.3.2 - Critérios de agrupamento.

O agrupamento das estações foi feito primeiramente usando como variáveis todos os tipos descritos acima. As variáveis brutas foram dispostas em uma matriz de dados $X(n \times p)$ onde X_{ij} representa a i -ésima observação da j -ésima estação com $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$. Ou seja, cada coluna da matriz corresponde às observações de uma estação; no caso dos dados reduzidos, usou-se uma matriz $Y(n \times p)$ com as mesmas características da matriz X . Para os fatores (no caso das variáveis serem fatores obtidos na ACP), obteve-se uma matriz $F(n \times p)$. Em todos os

casos, n (linhas das matrizes) corresponde ao número de estações, p (colunas das matrizes), no caso da matriz F , corresponde ao número de fatores.

Como nem a quantidade de estações nem o número de dias do mês com observações permanecem constantes ao longo do período em estudo, ao eliminar linhas e/ou colunas de dados faltosos a dimensão da matriz de dados varia muito. No caso analisado, abrangem de um máximo de 19 estações com 25 dias de observação a um mínimo de 7 estações com 17 dias de dados observados. No caso de agrupamento com fatores, o número de colunas varia caso a caso, devido ao fato de que 90% da variância nem sempre é explicado com o mesmo número de fatores.

O agrupamento das estações foi feito através da técnica hierárquica de agrupamento usando o Método Aglomerativo. Esta técnica foi preferida, porque não define previamente o número de grupos permitindo pesquisar o tamanho mais adequado dos mesmos.

Foram testados alguns dos métodos aglomerativos descritos na seção 3.1.3, como nenhum desses métodos apresentou grupos que delineassem mais ou menos as características climáticas do estado, escolheu-se para o trabalho os de Ward e Centróide.

4.3.3 - Determinação do número de grupos

No método aglomerativo o processo se inicia com cada unidade fazendo parte de um grupo individual, e vão se fundindo umas às outras agrupando-se primeiramente aquelas mais "semelhantes" até formar um único grupo com todas as unidades. Esta sequência de operações é ilustrada graficamente num dendrograma. Nas figuras 3.1 e 3.2 foram apresentados os dendrogramas dos agrupamentos feitos pelos métodos de Ward e Centróide usando as mesmas variáveis (os 10 primeiros fatores que explicaram 90% da variância total). Pode-se observar que apesar de se agrupar as mesmas variáveis os dendrogramas apresentam

características bem distintas. Enquanto no método de Ward as estações são apresentadas inicialmente agrupadas em pequenos grupos mais ou menos separados, no da centróide, existe uma tendência a que um grupo inicial cresça incorporando estações uma-a-uma. Algumas estações permanecem separadas para se agruparem apenas nas fusões finais. Neste sentido, o método de Ward parece mais apropriado para definir grupos.

A determinação do número de grupos adequado é muito importante no processo de simulação de séries visto que, à medida que avança o nível de agregação, aumenta a quantidade de estações por grupo e há uma perda de homogeneidade entre elas já que se agrupam primeiro as mais similares. Considerando esse princípio, foi estabelecido a priori quantos grupos seriam determinados independente do critério de agrupamento a ser utilizado. A forma adotada para a obtenção do número de grupos foi o de cortes transversais no dendrograma, feito de forma subjetiva a critério do pesquisador levando em consideração a quantidade de estações dos grupos que serviria de base para as simulações, obteve-se portanto, agrupamentos com um número de grupos adequado à matriz de dados.

4.4 - Análise de Componentes Principais (ACP)

A ACP foi usada duas vezes no trabalho. Primeiramente foi aplicada aos dados reduzidos de todas as estações da rede, a fim de obter as cargas fatoriais que seriam usadas como "variáveis" num processo de agrupamento das estações procurando definir o grupo apropriado para a simulação. Após determinar tal conjunto, incluindo a estação a simular, a ACP foi aplicada a esse grupo para determinar seus autovetores que seriam utilizados na simulação das séries de dados

das estações escolhidas. Na obtenção das cargas fatoriais foi usado o subprograma FACTOR do utilitário SPSS ("Statistical Package for Social Sciences").

4.4.1 - Simulação de séries temporais.

Após determinado um grupo composto por n' estações (incluindo aquela que seria simulada) aplicou-se ACP aos dados destas estações para obter a matriz quadrada de fatores, $F(n' \times p)$. Os autovetores normalizados (a_1, a_2, \dots, a_p) foram obtidos dividindo cada coluna de fatores (f) pela raiz quadrada do autovalor (λ) correspondente ou seja $f_1/\sqrt{\lambda_1}, f_2/\sqrt{\lambda_2}, \dots, f_p/\sqrt{\lambda_p}$. O número K de autovetores significativos, $K < P$, foi escolhido com base em observar o acúmulo de autovetores até explicar 90% da variância total do grupo. Esta porcentagem foi respeitada aproximadamente. A razão para este nível de exigência é explicada na seção 4.4.2. Com o critério estabelecido, ficou definida uma matriz de trabalho $B(N \times K)$ proveniente da matriz inicial $A(n' \times p)$, tendo-se eliminado as linhas correspondentes às estações a serem simuladas (uma ou mais), de forma que $N < n'$. Escolheu-se simular uma estação por vez (portanto, $N = n' - 1$). Procedendo-se a manipulações com matrizes, tal como descrito na seção (3.2.3) encontrou-se uma matriz linha chamada "forma linear" L cujo número de colunas é igual a N . A sequência de operações para se obter a série simulada é a seguinte:

- determina-se a transposta de B ; B^t
- realiza-se a operação

$$B^t \cdot B = Q \quad (4.1)$$

- determina-se a inversa de Q ; Q^{-1}
- realiza-se a operação

$$Q^{-1} \cdot B^t = S \quad (4.2)$$

- multiplicando-se por S a matriz linha formada pelos dados da estação a ser simulada, encontra-se L (dependendo da ordem de K, que é função do número de autovetores significativos, determina-se quantos Ls serão obtidos).
- multiplicando-se L por Y*, (Y incompleta, só com as estações agrupadas) sem a estação que será simulada, encontra-se a série procurada.

Vale salientar que a série simulada tem um número de dados igual ao de colunas da matriz Y(Nxp) obtida com as estações agrupadas. Se estas tiverem 25 dias de dados, a série simulada terá também 25 dias de dados simulados.

Para se obter uma série simulada com 31 dias (relativo ao mês de março da pesquisa), determina-se a matriz L (conforme descrito anteriormente) usando uma matriz de trabalho B(N'xK) proveniente da matriz de dados agrupados e com estações cuja séries temporais tenham 31 dias $N' \leq N$, onde N' corresponde ao número de estações com séries de 31 dias, Fazendo-se o produto de L por Y*(N'xP) onde P corresponde a 31 dias de observação, simula-se o mês completo. Mas, se o número de estações do grupo com 31 dias de dados for menor, isso leva a uma diminuição do número de estações do grupo originalmente obtido pelos métodos de agrupamento já descritos.

4.4.2 - Critério para julgar o desempenho do método.

O erro esperado em determinações experimentais de irradiação diária é de 5% das tabelas 4.1 e 4.2, percebe-se que isso significa aproximadamente 25ly para a região estudada (500ly em média). Considerando-se agora a série de dados (X), a série simulada (Xs) e a dos desvios da simulação ($X - Xs = e$), a variância dos dados "brutos" e dos reduzidos devem cumprir com a relação:

$$\text{Var}\{X\} = \text{Var}\{X_s\} + \text{Var}\{e\}$$

$$\text{Var}\{Y\} = \text{Var}\{X_s/S\} + \text{Var}\{e/S\}$$

onde S = desvio padrão da série de dados (da ordem de 75ly). Nesta equação, observa-se que $\text{Var}\{Y\} = 1$ e $\text{Var}\{e/S\} = 1/9$.

O critério básico para uma simulação com significado físico, é não ultrapassar a precisão inerente da própria medida. Conclui-se portanto, que $\text{Var}\{e/S\} = 1/9$ é um limite adequado para o erro associado à simulação, configurando em torno de 10% da variância da série original. Consequentemente, as componentes principais utilizadas no método não deveriam acumular mais de 8/9 de $\text{Var}\{X/S\}$, ou seja, em torno de 90%. Este é, precisamente, o critério mencionado na seção 4.4.1 para definir a ordem K de aproximação.

Uma vez realizada a simulação, é preciso estabelecer um critério de comparação para verificar a qualidade dos resultados obtidos. O critério adotado foi o de avaliar a grandeza típica do desvio da série simulada com relação à original,

$$E(\text{erro}) = \bar{y}(\text{simulado}) - \bar{y} \quad (4.3)$$

Observe-se que os valores \bar{y} são da variável padronizada. A partir de série E , seu desvio padrão (SE) fornece uma estimativa da qualidade da série simulada. Obviamente valores de SE em torno de 1/3 são completamente satisfatórios desde que equivalentes à própria precisão das medidas.

5 - RESULTADOS.

Foram analisados os meses de março e setembro para o período proposto (1975 - 1980) com as estações de Barra de Santa Rosa e Pombal. Inicialmente, o mês de março de 1975 e em particular Barra de Santa Rosa foram submetidos a análise detalhada a fim de melhor delimitar o método de trabalho.

5.1 - Grupamento de estações: março 1975.

Neste ano, a rede contava com 19 estações. A ACP aplicada às irradiações padronizadas forneceu fatores que acumulavam 91,8% da variância até a ordem 10. Os três primeiros acumulavam 60,7%

Foram usados quatro tipos de variáveis e dois métodos de agrupamento: Ward e Centróide. As variantes testadas são ilustradas pelo quadro 2 a seguir. Como o método principal de agrupamento a ser usado no trabalho era o de Ward, este foi usado com todas as variáveis. O da Centróide, considerado aqui um método de suporte, foi usado só para efeito de comparação e com as variáveis obtidas da aplicação de ACP.

Com relação aos grupos formados, o número desejado era 5 para todas as variáveis e os dois métodos de agrupamento. Este número foi estabelecido após testes feitos com diferentes quantidades de grupos. Foi observado que, como seria feita uma análise para vários anos, quanto mais estações por grupo, mais difícil ficaria encontrar as mesmas nos mesmos grupos para os anos seguintes. Por outro lado, com poucas estações por grupo (consequentemente número maior de grupos) corria-se o risco do desaparecimento de algumas delas quando da diminuição da matriz original de dados. Se percebeu também, que com 5 grupos (pelo menos para 1975 que continha todas as estações da rede) delineava-se mais ou menos as regiões

climáticas do Estado. As figuras 5.0.a e 5.0.b a seguir, mostram as estações agrupadas pelo método de ward, para os meses de março e setembro do ano de 1975 usando como variável os fatores que explicaram 90% da variância total dos dados, levando-se em consideração 5 grupos. Pode ser visto nesta figura, a distribuição das estações dentro dos grupos formados, bem como a posição geográfica de cada um deles. Ao aplicarmos o método da centróide, resultou que se fosse mantido o mesmo número de grupos, teríamos uma excessiva quantidade de estações no grupo que continha Barra de Santa Rosa. Isso poderia prejudicar a simulação já que introduziríamos no grupo estações relativamente heterogêneas às existentes. Para este método, o número de grupos estabelecido com os dez primeiros fatores (agrupamento 5) foi 12, e com os três primeiros, (agrupamento 6) foi 7. Todos os grupos e estações resultantes de cada agrupamento, podem ser vistos na tabela 5.1, incluída no Apêndice. Em cada caso, apresenta-se apenas o grupo que contém BSR. Com relação à semelhança entre os grupos resultantes, pode ser verificado que algumas estações aparecem em vários deles como é o caso de Esperança, Areia e Araruna. Já outras só aparecem apenas em um tipo de grupo, como é o caso de João Pessoa que só aparece no agrupamento de tipo (1).

FIGURA 5.0.a - Estações agrupadas pelo método de ward, levando em consideração

5 grupos. Março de 1975

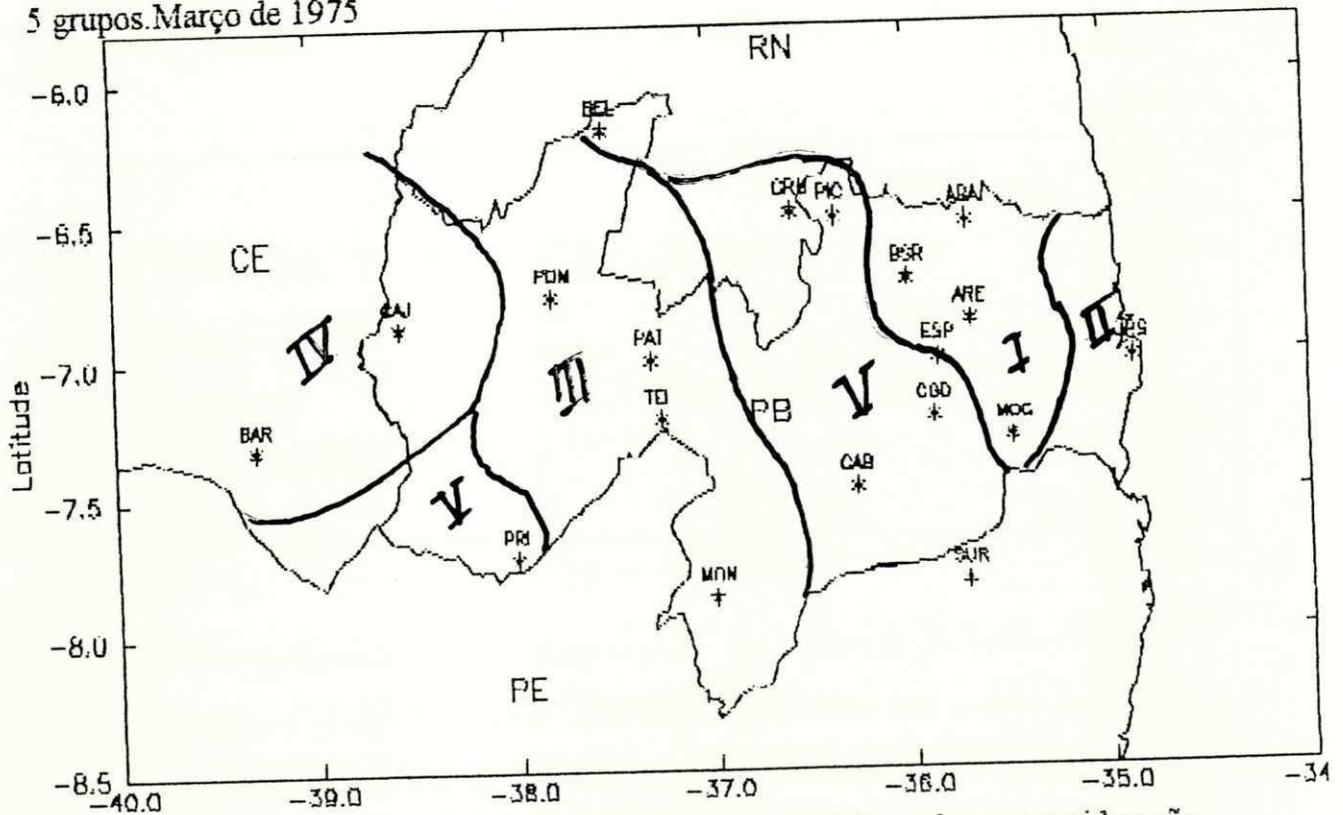
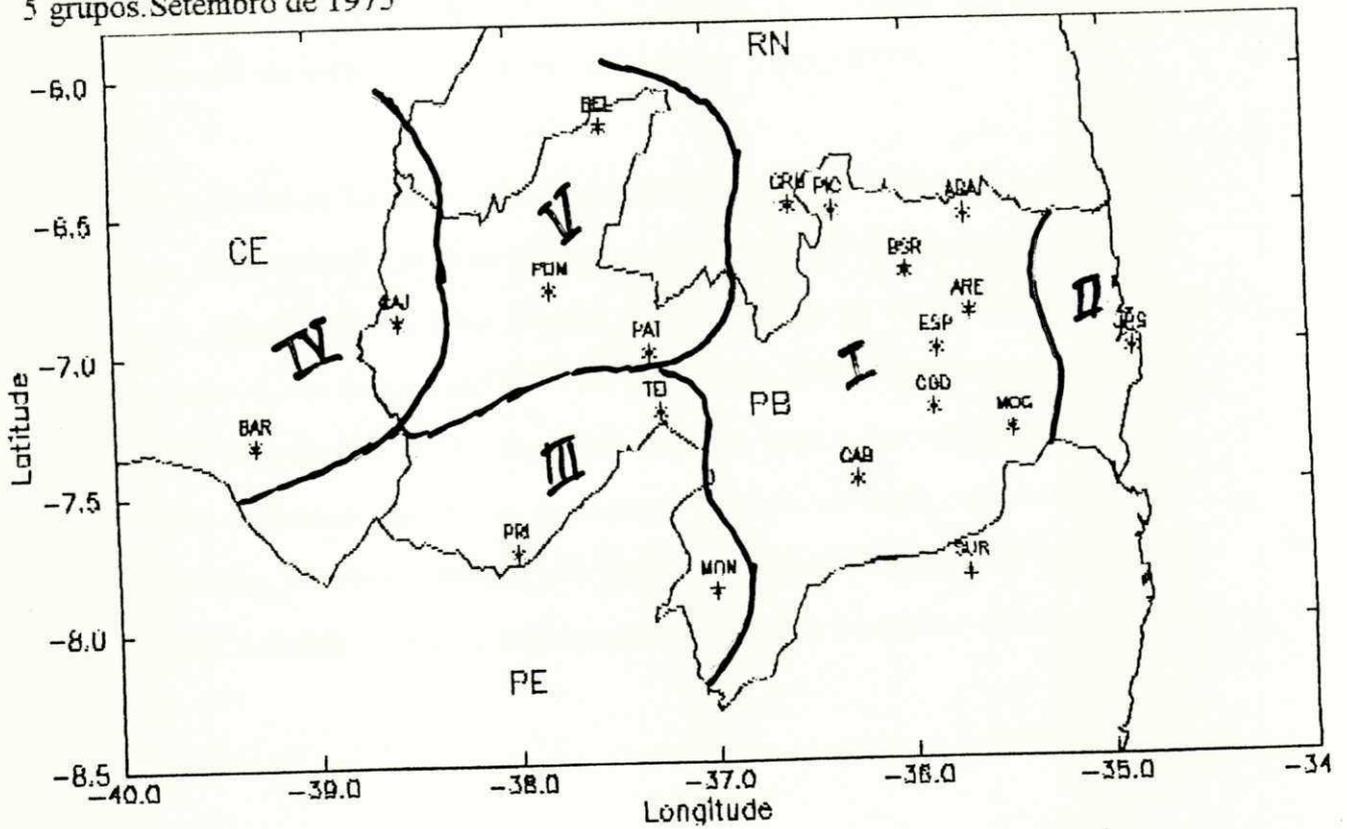


FIGURA 5.0.b - Estações agrupadas pelo método de ward, levando em consideração

5 grupos. Setembro de 1975



Quadro 2

Métodos de agrupamentos utilizados e tipo de variáveis.

Tipo de variáveis	Agrupamentos	
	WARD	CENTRÓIDE
dados brutos	(1)	
dados reduzidos	(2)	
10 primeiros fatores	(3)	(5)
3 primeiros fatores	(4)	(6)

Aliás, esse é o agrupamento que contém o grupo mais diferente de todos os outros. Nos agrupamentos (1) e (2) há uma diferença significativa nos grupos formados, pode ser verificado que apesar de ter sido usado o mesmo método de agrupamento a natureza das variáveis influenciou bastante nos resultados.

5.2 - Simulação da série de Barra de Santa Rosa: março 1975.

Como já foi dito, simulou-se primeiro a série de Barra de Santa Rosa. De todas as simulações feitas, com todos os agrupamentos, a melhor foi a do (3).

Na simulação feita usando a aproximação de ordem $K=3$ (os três primeiros autovetores explicando 90% da variância), obtivemos um desvio padrão do erro (SE) da ordem de 0,58. Na simulação feita com a aproximação de ordem $K=2$ (os dois primeiros autovetores explicaram 81% da variância), obtivemos (SE) = 0,59. Portanto, uma aproximação não tão exigente fornece qualidade semelhante da simulação. A figura 5.1 ilustra a série real e a simulada do melhor resultado.

Apresenta-se a seguir resultados obtidos para os anos posteriores a 1975

5.3 - Análise de março 1976 a 1980, Barra de Santa Rosa.

Para 1976 não foi repetido todo o processo feito para 1975, pois o objetivo era no primeiro ano da série a ser estudada fazer todos os métodos de agrupamento, simulações, e posteriormente (para os anos seguintes) repetir o critério de obtenção dos grupos mas simular só os que tivessem apresentado os melhores resultados na simulação feita para 1975.

FIGURA 5.1 - Série real e simulada para março de 1975. A linha contínua representa a série real e a tracejada a simulada

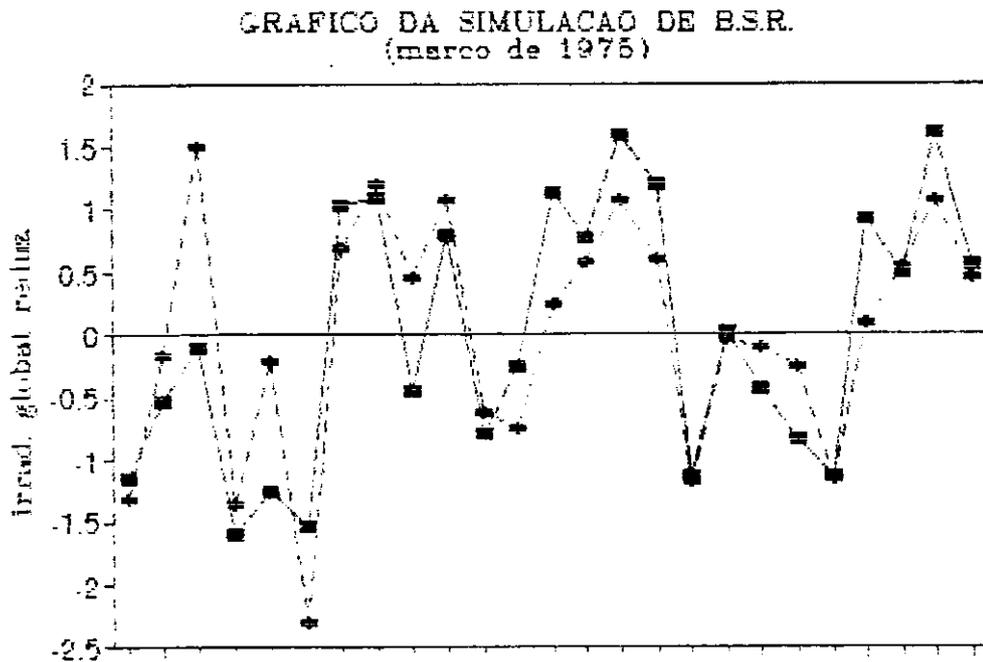
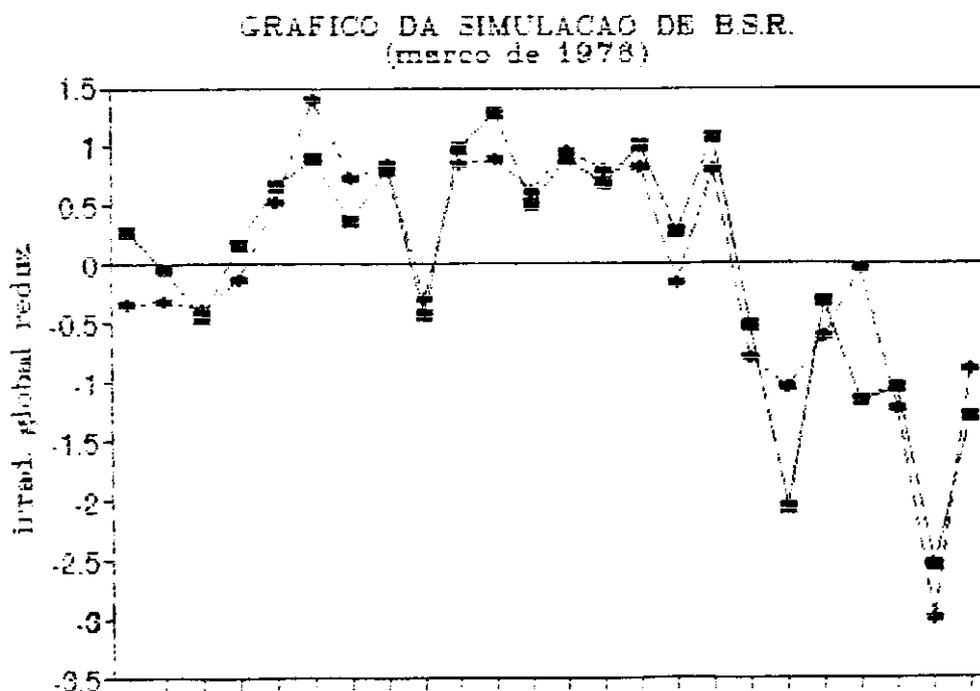


FIGURA 5.2 - Série real e simulada para março de 1976. A linha contínua representa a série real e a tracejada a simulada.



- Formação dos grupos.

A tabela 5.2., no apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6).

Neste ano, o número de estações da rede foi praticamente o mesmo que em 1975, faltando apenas uma. Portanto, o número de grupos foi mantido para cada método separadamente (Ward 5 grupos, Centróide 12 e 7 grupos). Com relação aos grupos formados, percebe-se que são todos semelhantes, ou seja, mantendo mais ou menos as mesmas estações em todos os grupos. Pode ser visto também que as estações de Araruna, Areia e Esperança aparecem em quase todos os grupos e que o método da centróide apresenta as mesmas estações dos grupos obtidos nos agrupamentos dos tipos (5) e (6).

- Simulação da série.

A figura 5.2 apresenta as series real e simulada do melhor resultado. Nas simulações feitas com as estações dos agrupamentos de tipo (3) e (4), os melhores resultados foram obtidos com as estações do (3). Com a aproximação de ordem $K=3$ (os três primeiros explicaram 90% da variância) obtivemos um desvio padrão do erro (SE) = 0,43, e com a aproximação de ordem $K=2$ (os dois primeiros explicaram 83% da variância) o desvio foi de 0,43.

- março de 1977.

No mês de março de 1977 o número de estações foi igual ao de 1975, e com isso o número de grupos foi mantido para quase todos os agrupamentos, variando somente os dos agrupamentos (5) e (6) que foram respectivamente, de 11 e 7 grupos.

FIGURA 5.3 - Série real e simulada para março de 1977. A linha contínua representa a série real e a tracejada a simulada

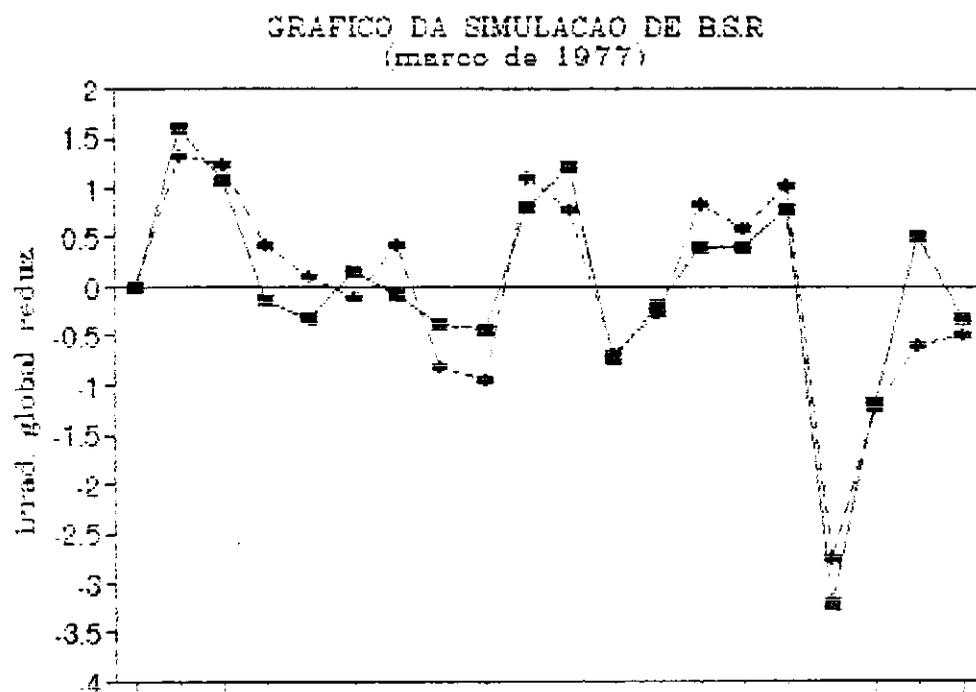
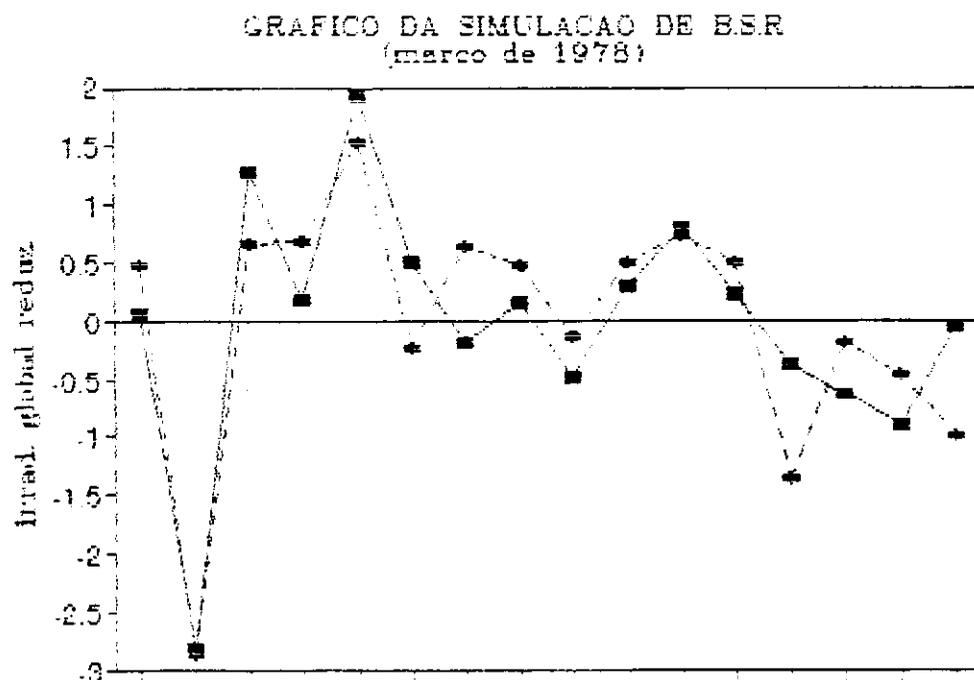


FIGURA 5.4 - Série real e simulada para março de 1978. A linha contínua representa a série real e a tracejada a simulada.



- Formação dos grupos.

A tabela 5.3., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Os grupos obtidos para este ano, apesar de conterem mais estações do que 1976, tiveram quatro (dos seis existentes) formados pelas mesmas estações o que prova que apesar dos vários tipos de variáveis usadas para agrupamento, e dos métodos aplicados, estas se mantiveram. É verificado também que a estação de Cajazeiras, que até então não tinha aparecido em nenhum dos grupos formados, aparece em todos independente do método de agrupamento e das variáveis utilizadas.

- Simulação da série.

A figura 5.3 mostra a Série real e a simulada do melhor resultado. As simulações até aqui tinham sido feitas para os agrupamentos (3) e (4), neste caso, apresenta-se a simulação só com as estações do agrupamento (3). Este procedimento deve-se ao fato de que a simulação feita para o (4) daria o mesmo resultado pois as estações que compõem ambos os grupos são as mesmas. Então, tivemos para a aproximação de ordem $K=3$ (os três primeiros autovetores explicaram 90% da variância) um desvio padrão do erro (SE) de 0.42, e com $K=2$ (os dois primeiros autovetores explicaram 85% da variância), $(SE)= 0.48$.

- março de 1978.

Neste ano houve uma redução significativa do número de estações que compõem a rede de dados, mas mesmo assim o número de grupos, usando o método de Ward, foi mantido variando somente quando se usou o da Centróide, para o qual se estabeleceram 4 grupos.

- Formação dos grupos.

A tabela 5.4., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Os grupos apresentam um número pequeno de estações chegando a ter apenas duas (Araruna e Barra de Santa Rosa) no grupo obtido do agrupamento (1). As estações mantidas em cada grupo também variam, não apresentando-se nenhum grupo com as mesmas estações nos agrupamentos feitos com o método de Ward. Já com os da Centróide os grupos foram semelhantes.

- Simulação da série.

A figura 5.4 mostra a série real e simulada do melhor resultado. Devido à diminuição de estações por grupo, os dois primeiros autovetores já explicaram 91,2% da variância não sendo possível usarmos simulação com matriz de aproximação de ordem $K > 2$, já que o 3º explicou 97,7% da variância a simulação foi feita usando uma aproximação de ordem $K = 2$ (os dois primeiros explicaram 91% da variância) no agrupamento (3) e 88% no (4). Os desvios padrões dos erros das duas simulações foram, respectivamente, 0,53 e 0,57.

- março de 1979.

A quantidade de estações da Rede de dados continuou diminuindo e para este ano foram 12. Neste caso, para manter uma quantidade razoável de estações por grupo formado foi preciso diminuir o número de grupos pois com 5 a quantidade de estações que se agrupava com Barra de Santa Rosa era pequena, (em torno de duas). Então foram adotados 4 grupos para os agrupamentos (1) (2) (3) (4) e 4 e 7 grupos nos agrupamentos (5) e (6).

FIGURA 5.5 - Série real e simulada para março de 1979. A linha contínua representa a série real e a tracejada a simulada

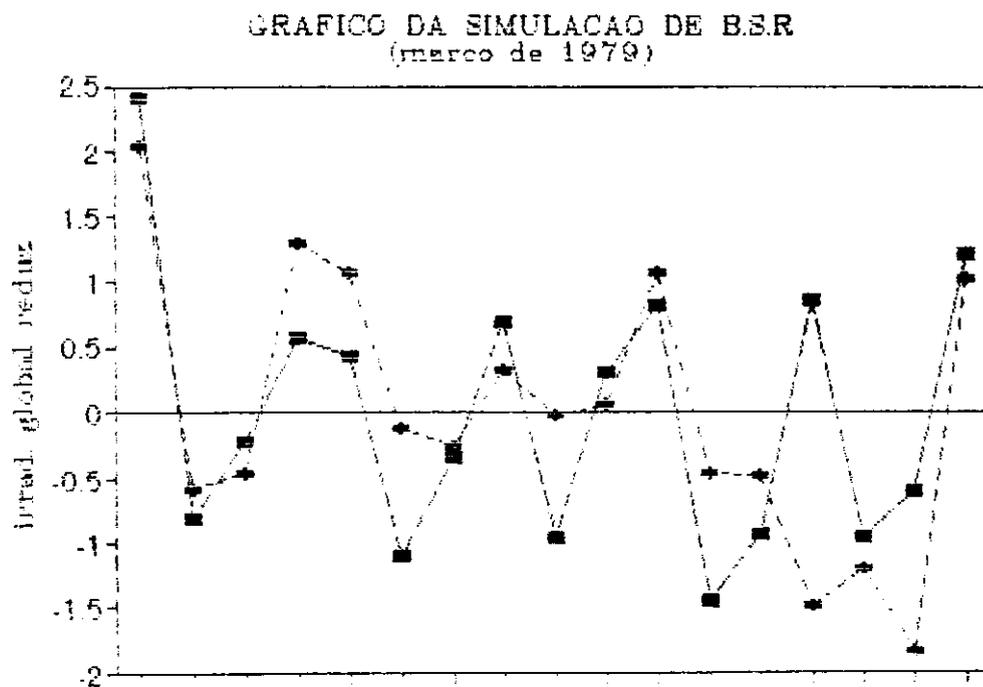
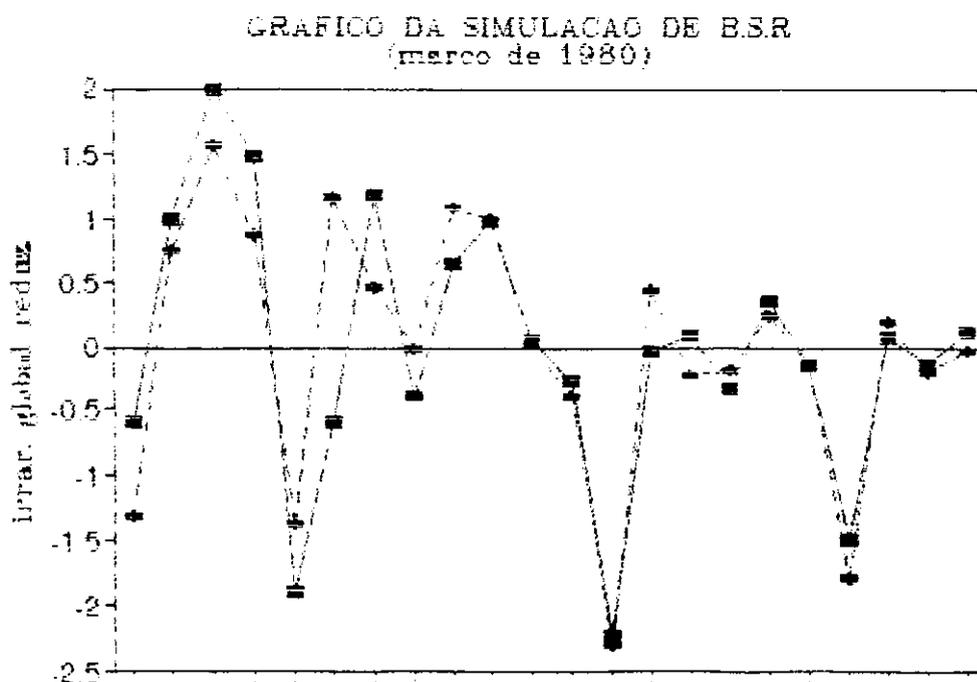


FIGURA 5.6 - Série real e simulada para março de 1980. A linha contínua representa a série real e a tracejada a simulada.



- Formação dos grupos.

A tabela 5.5., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Com a diminuição do número de grupos obteve-se uma quantidade razoável de estações, até um pouco acima do esperado, chegando a ter 8 (das 12 existentes) por grupo como foi o caso do grupo obtido no agrupamento (3).

A figura 5.5 mostra as séries real e simulada do melhor resultado. Com o aumento da quantidade de estações nos grupos formados, houve uma perda de homogeneidade interna dos mesmos fazendo com que as simulações não fossem boas. Foi tentada uma terceira simulação com o grupo do Agrupamento (2) mas não apresentou um resultado satisfatório. Assim, os resultados das simulações foram as seguintes: no agrupamento (2), a simulação feita com as aproximações de ordem $K=3$ (os três primeiros explicando 89% da variância) e de ordem $K=2$ (os dois primeiros explicando 83% da variância), apresentaram desvios padrões dos erros de 0,85 e 0,73 respectivamente. Na simulação feita com as estações do agrupamento (3) com as aproximações de ordem $K=3$ e $K=2$ (os três e os dois primeiros explicaram 92% e 84% da variância), foram obtidos os piores resultados nos desvios padrões dos erros, que foram 1,0 e 0,83. Nas simulações feitas com as estações do agrupamento (4), encontrou-se um desvio padrão do erro (SE) de 0,84 para a aproximação de ordem $K=3$ (os três primeiros explicaram 87% da variância). E de 0,75 para simulação com aproximação de ordem $K=2$ que explicaram 77% da variância.

- março de 1980.

Neste ano tivemos o menor número de estações da Rede em toda série estudada, limitado a 7 estações (quase um terço do total que foi de 19 em 1975), apenas 7 estações. Assim, foi preciso diminuir o número de grupos, que antes era de 5, para 2 nos Agrupamentos (1) (2) (3) (4) e para 4 e 3 nos Agrupamentos (5) e (6).

- Formação dos grupos.

A tabela 5.6., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Como já era esperado, os grupos formados tiveram um pequeno número de estações em cada um e apresentaram, com exceção do Agrupamento (1), uma boa similaridade entre eles.

- Simulação da série.

A figura 5.6 apresenta as séries real e simulada do resultado. Apesar de os grupos terem quase todas as mesmas estações e apresentarem boa similaridade entre eles, era de se esperar uma boa simulação, mas isso não aconteceu e o desvio padrão do erro (SE) na simulação com aproximação de ordem $K=2$ explicando 91% da variância, foi de 0,58.

5.4 - Grupamento de estações: setembro 1975.

Após definir o método de trabalho usado para o mês de março de 1975, este foi estendido para o de setembro.

Neste ano, a rede contava também com o mesmo número de estações que março de 1975. A análise de componentes principais aplicada às irradiações padronizadas forneceu uma matriz de fatores com dimensão menor do

que para março de 1975, já que 90,2% da variância foi explicada por 6 fatores. Os três primeiros acumulavam 77,6%. Foram usadas as mesmas variáveis, os mesmos métodos de agrupamento e o mesmo número de grupos.

Com relação aos grupos formados, conforme pode ser vistos na tabela 5.7 incluída no apêndice, a característica observada foi praticamente a mesma que em março tendo a destacar a semelhança dos agrupamentos dos tipos (2) e (3) os quais tiveram as mesmas estações. Levando em consideração a constituição dos grupos, neste mês a quantidade de estações repetidas em vários deles é maior, podendo ser citadas Areia, Cabaceiras, Campina Grande, Esperança, Cruzeta e Mogeiro (que aparecem em todos os grupos).

5.5 - Simulação da série de Barra de Santa Rosa: setembro 1975.

Como já foi dito anteriormente, a simulação feita com o mês de março de 1975 foi com todos os agrupamentos; a partir daí simulou-se somente os que tiveram melhor resultado. Isto nos leva a apresentar, daqui para frente, só as simulações feitas com os agrupamentos dos tipos (3) e (4) considerando sempre os autovetores que explicaram em torno de 90% da variância por um lado, e os dois primeiros por outro.

A figura 5.7 mostra as séries real e simulada do melhor resultado. Na simulação feita com o agrupamento (3), o qual apresentou o melhor resultado, obtivemos um desvio padrão do erro (SE) igual a 0,44 com a aproximação de ordem $K=3$ (os três primeiros autovetores explicaram 93% da variância). Na simulação feita com a aproximação de ordem $K=2$ (os dois primeiros autovetores explicando 89% da variância) obtivemos (SE) igual a 0,48. Portanto, nas duas simulações feitas para esse mês os resultados foram melhores do que os de março.

5.6 - Análise de setembro 1976 a 1980, Barra de Santa rosa.

Nesta sequência de anos o procedimento foi o mesmo adotado anteriormente; discutiremos apenas a formação dos grupos e a simulação das séries.

- Formação dos grupos

A tabela 5.8., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Em 1976 o número de estações da rede se manteve praticamente o mesmo faltando apenas uma estação, e assim pôde-se conservar 5 grupos para o método de Ward, variando um pouco os agrupamentos pelo método da Centróide. Com relação aos grupos formados, percebe-se a semelhança das estações que compõem os grupos do agrupamento (2) e (3), o mesmo acontecendo com os (4) e (6), enquanto nos dois primeiros mudam as variáveis, nos dois últimos mudam os métodos de agrupamento. Duas estações, com exceção da que está sendo simulada, aparecem em todos os grupos que são: Esperança e Mogeiro.

- Simulação da Série

A figura 5.8 mostra as séries real e simulada do melhor resultado. Nas simulações feitas com os dois agrupamentos pré-estabelecidos(3) e (4), o melhor resultado foi com as estações do (3). Com a aproximação de ordem $K=4$ (os quatro primeiros explicando 91% da variância) obtivemos um desvio padrão do erro de 0,61, e com a aproximação de ordem $K=2$ (os dois primeiros explicando 87% da variância) o desvio foi de 0,63. Portanto, um erro maior do que para março do mesmo ano.

FIGURA 5.7 - Série real e simulada para setembro de 1975. A linha contínua representa a série real e a tracejada a simulada.

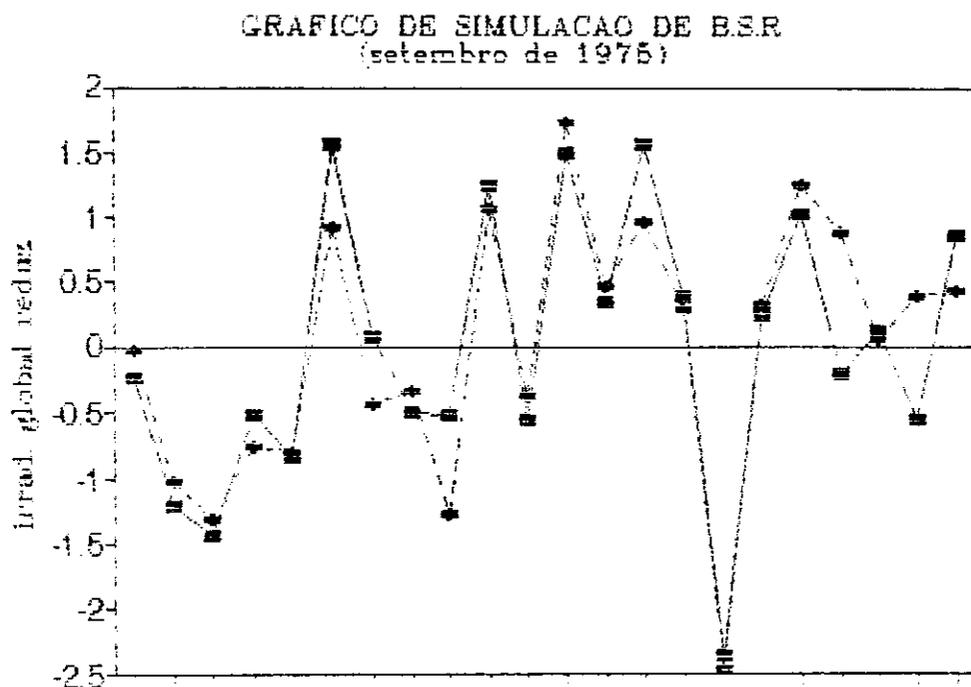
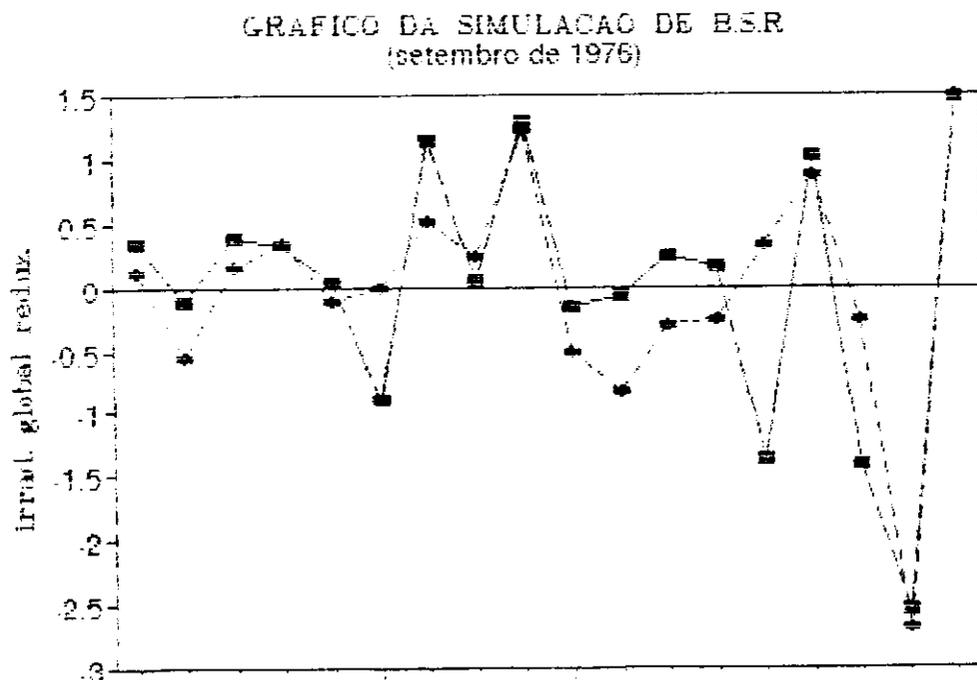


FIGURA 5.8 - Série real e simulada para setembro de 1976. A linha contínua representa a série real e a tracejada a simulada.



- setembro de 1977

No mês de setembro de 1977 o número de estações foi o mesmo de 1975; portanto, foi possível manter o número de grupos variando só os dos agrupamentos com método da Centróide, que foi de 7 para o agrupamento (5) e de 6 para o agrupamento (6)

- Formação dos grupos

A tabela 5.9., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Para os grupos desse mês, percebe-se também a semelhança entre eles, como é o caso dos grupos dos agrupamentos (3) e (5) que são iguais, e os (4) e (6). Em ambos, as variáveis são as mesmas mas mudam os métodos de agrupamento. Com relação às estações comuns a todos os grupos, se verifica que Araruna e Cajazeiras aparecem em todos eles.

- Simulação da série

A figura 5.9 mostra as séries real e simulada do melhor resultado. Nas simulações feitas com as estações dos agrupamentos (3) e (4) tivemos para a aproximação de ordem $K=2$, no agrupamento (3), dois autovetores explicando 95% da variância, não sendo necessário fazer simulação com a aproximação de ordem $K>2$, obtivemos aqui a melhor simulação com $(SE) = 0,18$. Já na simulação feita com as estações do agrupamento (4), tivemos para a aproximação de ordem $K=3$ (os três primeiros autovetores explicando 89% da variância) um desvio padrão do erro (SE) de 0,39 e com $K=2$ (os dois primeiros autovetores explicaram 83% da variância), (SE) igual a 0,42. Portanto, resultados melhores que os de março.

FIGURA 5.9 - Série real e simulada para setembro de 1977. A linha contínua representa a série real e a tracejada a simulada.

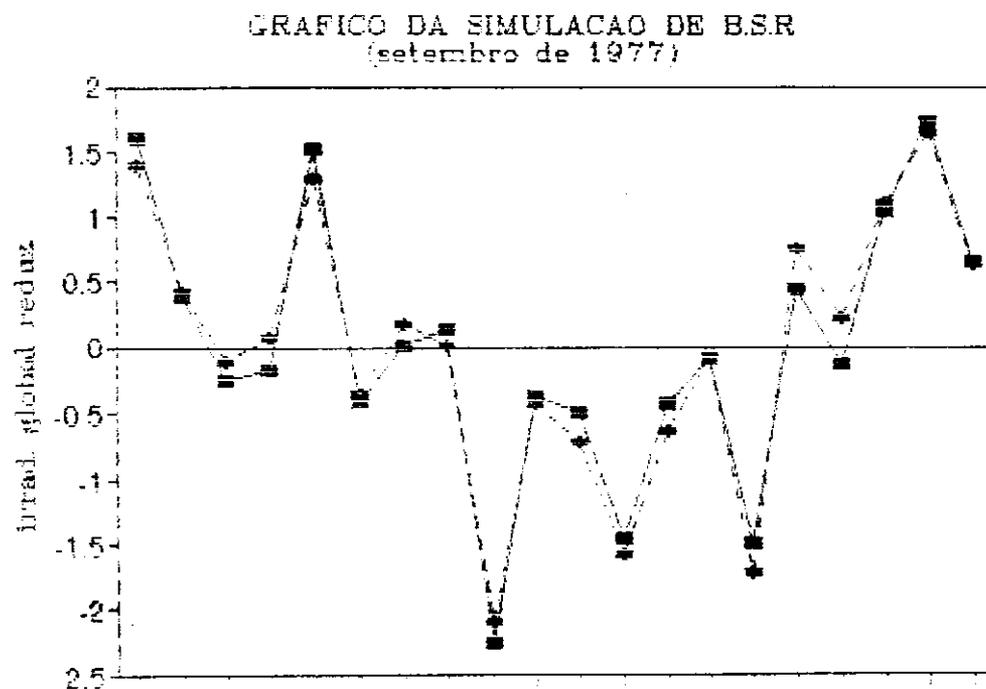
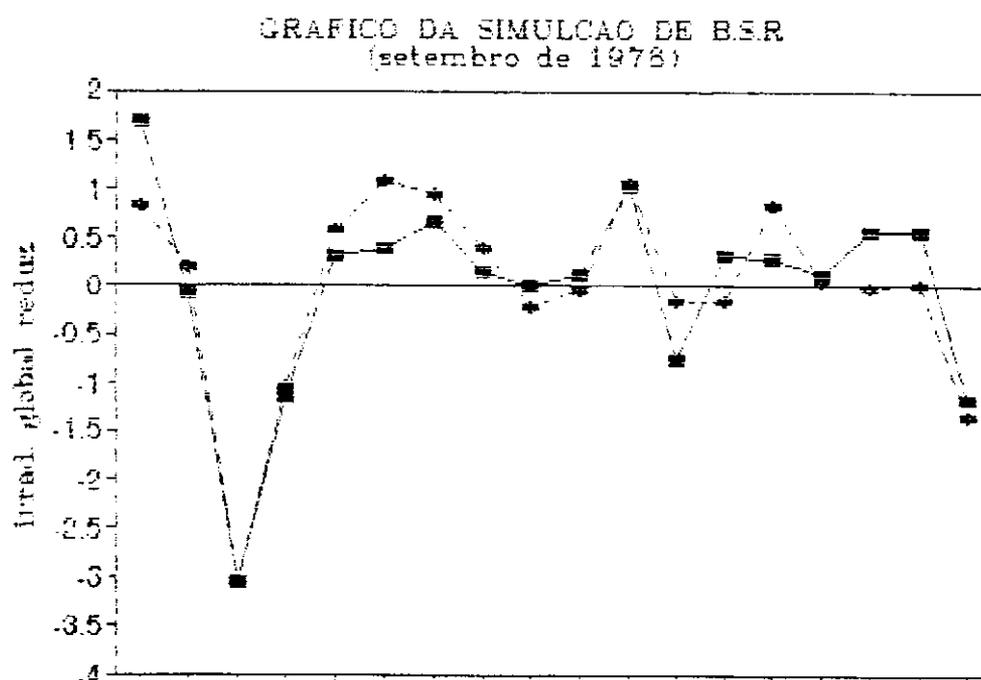


FIGURA 5.10 - Série real e simulada para setembro de 1978. A linha contínua representa a série real e a tracejada a simulada.



- setembro de 1978

Neste ano, já que houve redução do número de estações para o mês de setembro, quando tivemos apenas 14. Esta redução levou a variar o número de grupos com o método de Ward, ficando com 4 grupos nos agrupamentos (3) e (4), enquanto que os agrupamentos com método da Centróide, que antes estavam sendo de 6 a 7 grupos, passaram a ter 5.

- Formação dos grupos

A tabela 5.10., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Apesar da redução do número de estações da rede, as composições dos grupos apresentaram um número razoável de estações, em torno de 5. Tivemos três agrupamentos com grupos formados pelas mesmas estações, que foram (2), (3) e (5), e também os dos agrupamentos (4) e (6). Com relação as estações que aparecem em todos os grupos, tivemos apenas 2, que foram Araruna e João Pessoa.

- Simulação da série

A figura 5.10 mostra as séries real e simulada do melhor resultado. No agrupamento (3) os dois primeiros autovetores já explicavam 94% da variância, e não se justificam a simulação com aproximação $K > 2$. Todavia, na simulação com a matriz de aproximação de ordem $K=2$ tivemos um desvio padrão do erro (SE) de 0,41, igual ao erro obtido com a simulação feita com as estações do agrupamento (4) com uma aproximação de ordem $K=3$ (os três primeiros autovetores explicaram 94% da variância). Para as estações do agrupamento (4) mas com uma aproximação $K=2$ (os dois primeiros autovetores explicando 88% da variância), tivemos (SE) = 0,43. Estes resultados também foram melhores do que os de março.)

- setembro de 1979

Neste ano houve uma queda brusca no número de estações da rede, muito maior do que no mês de março que teve 12 estações. Em setembro o número baixou para 8, menos da metade do total da rede, o que levou à redução da quantidade de grupos, que passou para 3 com todas as variáveis e os dois métodos de agrupamento.

- Formação dos grupos

A tabela 5.11., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Com a diminuição do total de estações da rede, houve também uma redução no número de componentes por grupo, como foi o caso do agrupamento (1) para o qual tivemos apenas duas estações no grupo. Isto também afetou a composição dos mesmos que não teve nenhuma estação comum a todos.

- Simulação da série

A figura 5.11 mostra as séries real e simulada do melhor resultado. Como as estações simuladas foram as dos agrupamentos (3) e (4) e estes apresentaram grupos semelhantes, foi feita a simulação só para o grupo de estações do agrupamento (3). Com a redução do número de estações dois autovetores foi suficiente para explicar 93% da variância, assim tivemos com uma matriz de aproximação de ordem $K=2$, um desvio padrão do erro de 0,48 que deverá ser o mesmo para a simulação com as estações do (4) já que são as mesmas. Esta simulação também foi melhor que a março.

- setembro de 1980

Neste ano o número de estações da rede foi o mesmo de 1979, portanto foi mantido o número de grupo do ano anterior.

- Formação dos grupos

A tabela 5.12., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Apesar de conter poucas estações, estas se distribuíram uniformemente nos grupos mantendo, não só a mesma quantidade de estações, como também se mantiveram praticamente inalterados mantendo as mesmas estações em todos os grupos de quase todos os agrupamentos, com exceção só do (1).

- Simulação da série

A figura 5.12 mostra as séries real e simulada do melhor resultado. Como as estações foram as mesmas nos dois agrupamentos a serem simulados, simulou-se as do (3), e devido a poucas estações existentes, dois autovetores explicaram mais de 90% da variância não havendo necessidade de se usar a matriz de aproximação $K=3$. Na simulação feita com $K=2$ (os dois primeiros autovetores explicaram 93% da variância) tivemos um desvio padrão do erro de 0,16.

FIGURA 5.11 - Série real e simulada para setembro de 1979. A linha contínua representa a série real e a tracejada a simulada.

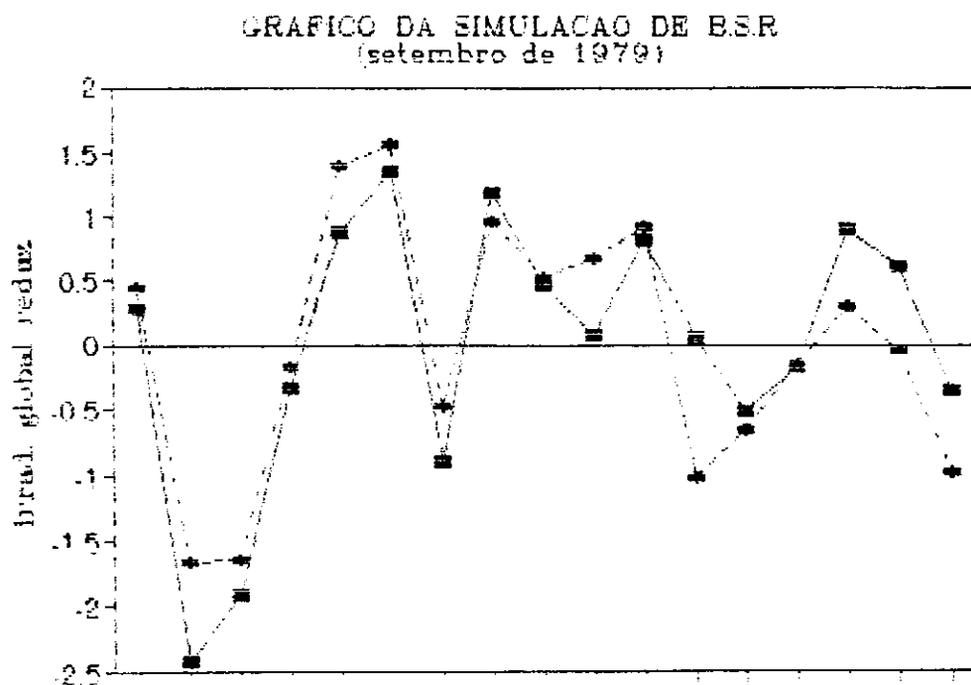
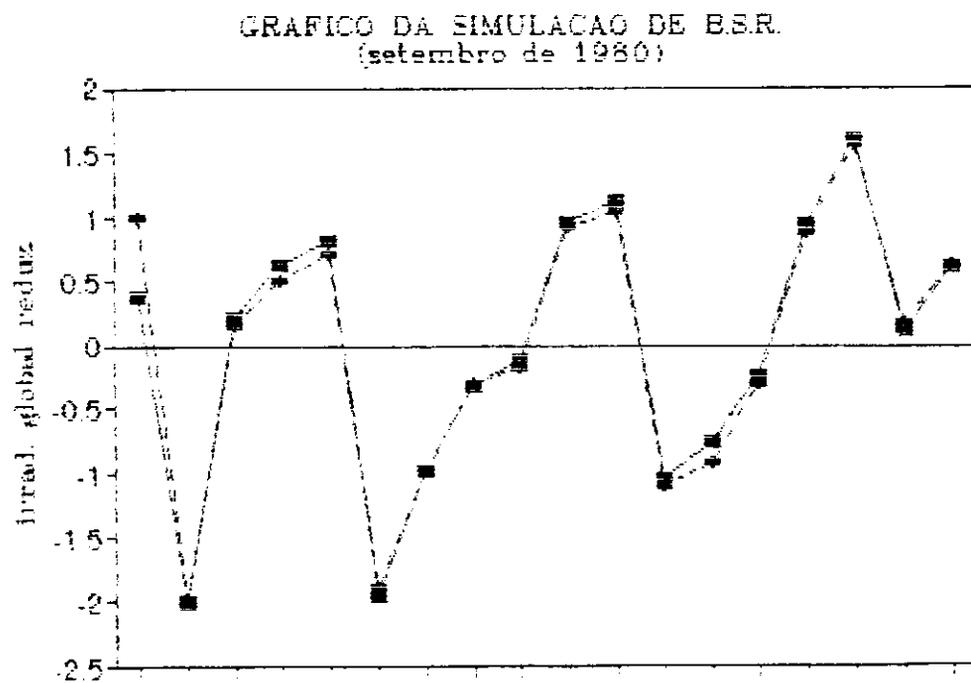


FIGURA 5.12 - Série real e simulada para setembro de 1980. A linha contínua representa a série real e a tracejada a simulada.



5.7 - Grupamento de estações: março 1975 - Pombal.

Como o procedimento adotado para simulação dos dados de Barra de Santa Rosa foi estendido para Pombal, se usou os mesmos dendrogramas para separar os grupos homogêneos de estações que continham Pombal. Uma visão geral nestes dendrogramas se percebe, em quase todos, que formam-se grupos mais ou menos separados de estações em torno tanto de BSR quanto de POM. Dificilmente estas duas estações estão em grupos iguais, pois localizam-se em regiões climaticamente distintas.

- Formação dos grupos

A tabela 5.1.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). A principal característica dos grupos formados neste ano é a diferença, em quase todos eles, com relação as estações que os constituem. Tendo só dois apresentado as mesmas estações, que foram os grupos dos agrupamentos (2) e (3), que apesar de terem sido agrupados pelo mesmo método, as variáveis adotadas foram diferentes.

- Simulação da série

A figura 5.1.1 mostra as séries real e simulada do melhor resultado. Como as estações dos agrupamentos (3) e (4) eram distintas, simulou-se Pombal com os dois grupos de estações. No agrupamento (3) usando uma aproximação de ordem $K=3$ (os três primeiros autovetores explicando 93% da variância) obtivemos um desvio padrão do erro (SE) de 0,84 e a aproximação de ordem $K=2$ (os dois primeiros autovetores explicando 85% da variância) obteve-se um (SE) de 0,67, na simulação do agrupamento (4) obteve-se um resultado melhor com a matriz de

aproximação de ordem $K=2$ (os dois primeiros autovetores explicaram 77% da variância), o desvio padrão do erro foi de 0,64

Apresentam-se a seguir resultados obtidos para os anos posteriores a 1975.

5.8 - Análise de março 1976 a 1980, Pombal.

A tabela 5.2.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Em 1976 apesar da matriz de dados originais contar com 18 estações, não foi possível manter o mesmo número de grupos que vinham até então sendo adotado para os agrupamentos feitos tanto para março quanto para setembro. Pois como dito anteriormente, Pombal se agrupa com outras estações que são menos numerosas do que as que se agrupam com Barra de Santa Rosa e por isso, foi estabelecido um número menor de grupos (três).

- Simulação da série

A figura 5.2.1 mostra as séries real e simulada do melhor resultado. Na simulação com as estações do agrupamento (3) usando uma matriz de aproximação de ordem $K=3$ (os três primeiros autovetores explicando 93% da variância) se obteve um resultado que não foi bom, um desvio padrão do erro de 0,93. Na simulação com as estações do agrupamento (4) usando uma matriz de aproximação também de ordem $K=3$ o resultado foi um desvio padrão do erro de 0,52.

- março de 1977

A tabela 5.3.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Nos grupos formados com os dados deste ano se percebe que muitos deles contêm as mesmas estações, como é o caso dos grupos nos agrupamentos (2), (4) e (6) e que estas estações pertencem, quase todas, à mesma região climática, Sertão, todavia, no agrupamento (1)

FIGURA 5.15 - Série real e simulada para março de 1977. A linha contínua representa a série real e a tracejada a simulada.

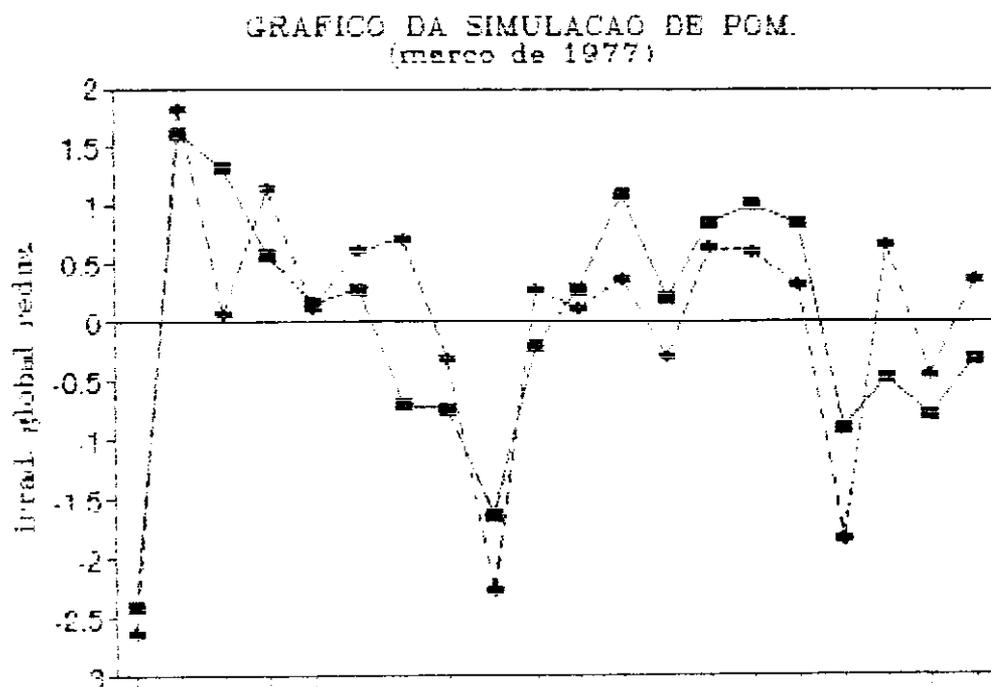
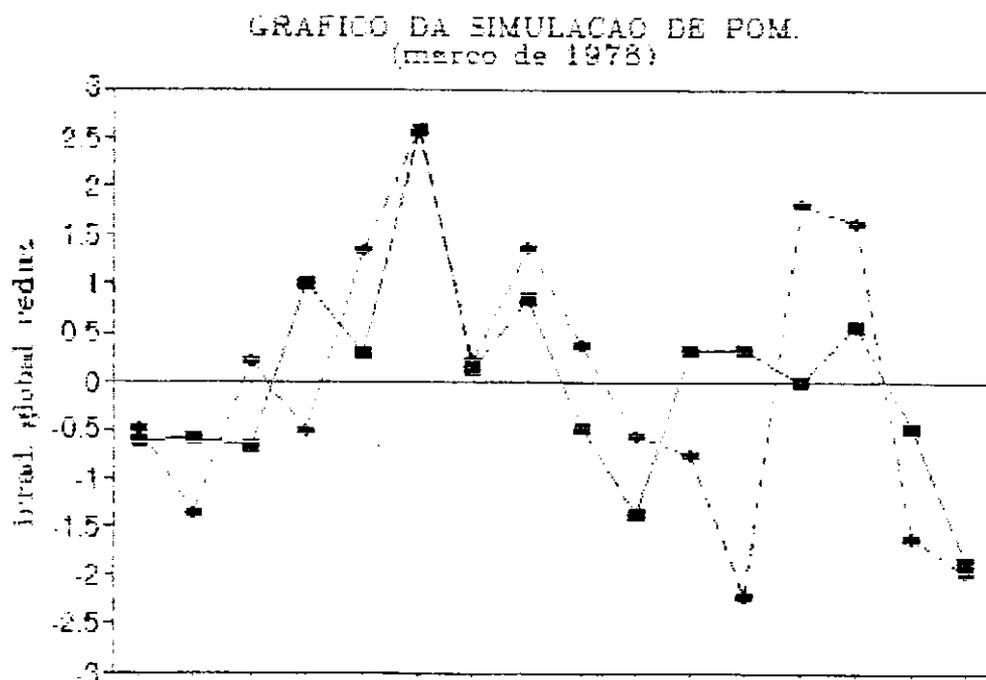


FIGURA 5.16 - Série real e simulada para março de 1978. A linha contínua representa a série real e a tracejada a simulada.



aparecem as estações de Esperança, Campina Grande e Cabaceiras pertencentes a regiões climáticas distintas e diferentes da de Pombal.

- Simulação da série

A figura 5.3.1 mostra as séries real e simulada do melhor resultado. Das quatro simulações feitas para os dois agrupamentos, a melhor foi a obtida com o (4) usando a matriz de aproximação $K=3$ (os três primeiros autovetores explicando 87% da variância) que teve um desvio padrão do erro (SE) de 0,67.

- março de 1978

A tabela 5.4.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Com a redução do número de estações foi necessário diminuir ainda mais o número de grupos que passou de três para dois. Apesar dos grupos formados conterem uma quantidade razoável de estações, os agrupados pelo método da centróide ficaram constituídos por apenas uma estação. Nos outros grupos, embora com mais estações, são todos distintos, não tendo estações comuns.

- Simulação da série

A figura 5.4.1 mostra as séries real e simulada do melhor resultado. De todas as séries simuladas em todos os anos o pior resultado foi o obtido para esse, nenhum valor do desvio padrão do erro nas quatro simulações, com as estações dos dois agrupamentos, ficou abaixo da unidade.

-março de 1979

A tabela 5.5.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Neste ano a quantidade de estação por grupo foi

muito pequena, tendo apenas o grupo obtido do agrupamento (1) com quatro estações.

- Simulação da série

A figura 5.5.1 mostra as séries real e simulada do melhor resultado. Com duas estações por grupo não deu para simular série com o método proposto no trabalho, mas como o critério a ser seguido era o de simular as estações dos agrupamentos (3) e (4), no (3) tivemos estações suficientes para fazer a simulação. Com uma matriz de aproximação de ordem $K=2$ (os dois primeiros autovetores explicando 99% da variância) tivemos um desvio padrão do erro igual a 0,90.

- março de 1980

A tabela 5.6.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Mesmo com um número reduzido de grupos, houve um aumento na quantidade de estações dos mesmos e quase todos foram compostos pelas mesmas estações, como foi o caso dos obtidos com os agrupamentos (2), (3), (4) e (6).

- Simulação da série

A figura 5.6.1 mostra as séries real e simulada do melhor resultado. Com as mesmas estações fazendo parte dos agrupamentos (3) e (4), simulou-se apenas o (3) cujo os dois primeiros autovetores explicaram 89% da variância, obtendo-se um desvio padrão do erro (SE) de 0,79.

FIGURA 5.17 - Série real e simulada para março de 1979. A linha contínua representa a série real e a tracejada a simulada.

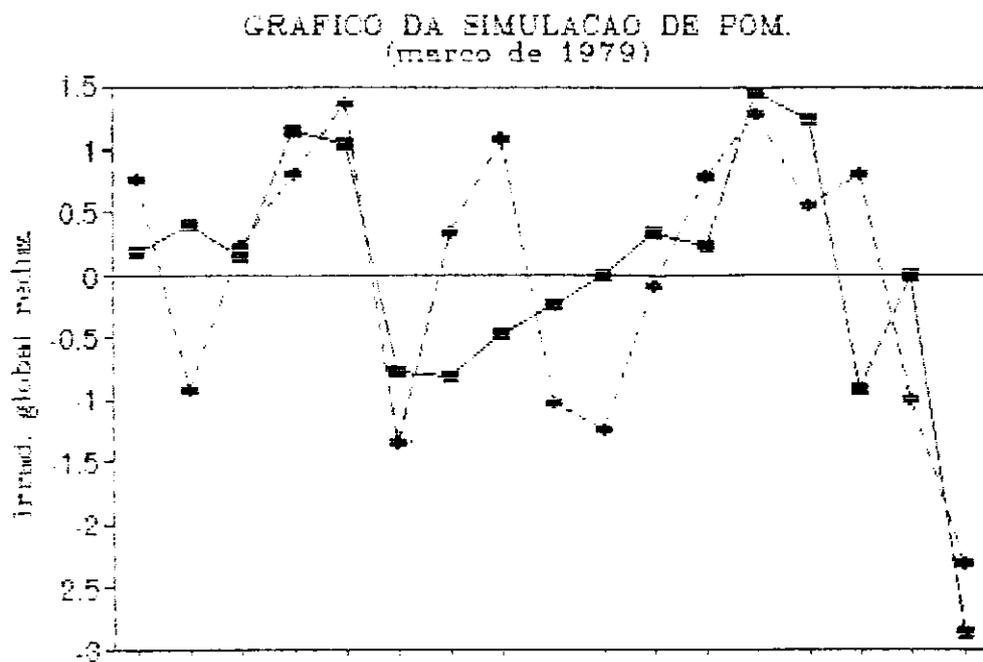
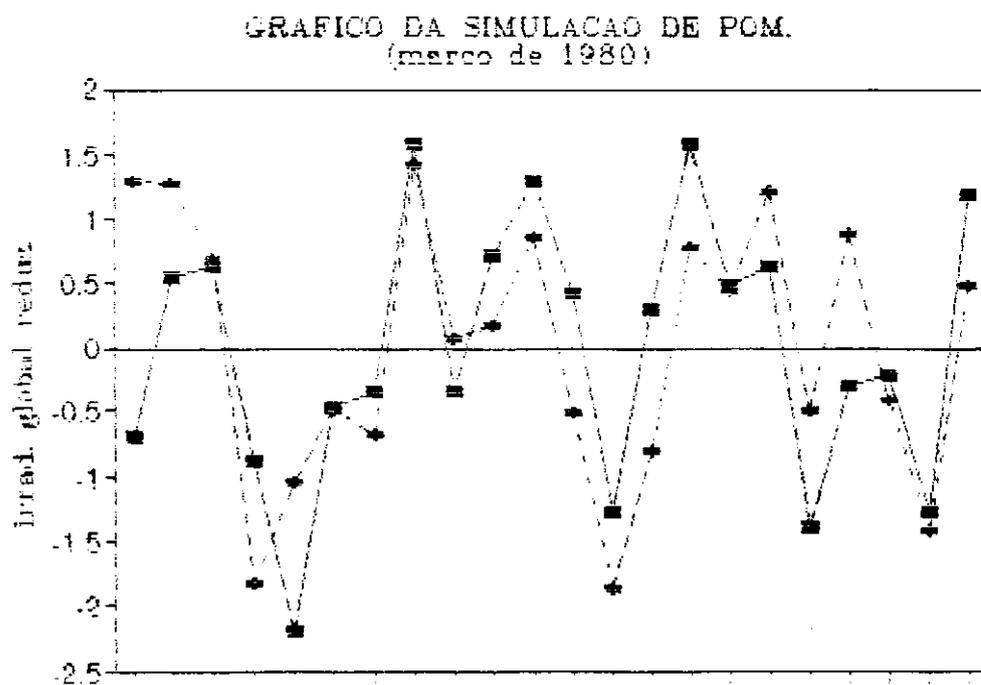


FIGURA 5.18 - Série real e simulada para março de 1980. A linha contínua representa a série real e a tracejada a simulada.



5.9 - Grupamento de estações: setembro 1975, Pombal.

A tabela 5.7.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Neste mês, diferentemente do que aconteceu em março, pôde ser adotado o número de grupos estabelecido previamente. As estações agrupadas seguiram uma característica regional ou seja, quase todas as estações que se agruparam com Pombal pertenciam à mesma região climática (sertão) do estado.

- Simulação da série

A figura 5.7.1 mostra as séries real e simulada do melhor resultado. Não foi feita simulação usando a matriz de aproximação de ordem $K=3$, pois nos dois agrupamentos (3) e (4) os dois primeiros autovetores explicaram 96,4 e 96,7% da variância não havendo portanto, necessidade da aproximação de ordem três. As duas simulações com $K=2$ apresentaram bons resultados e o melhor deles foi o do agrupamento (4) que teve um desvio padrão do erro de 0,41.

5.10 - Análise de setembro 1976 a 1980, Pombal.

A tabela 5.8.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Em 1976 o número de grupos manteve-se o mesmo, e pôde-se perceber uma quantidade muito grande de estações em cada um deles (mais de dez) inclusive estações que se agrupavam tanto com Pombal quanto com Barra de Santa Rosa. O único grupo que diferenciou totalmente de todos foi o do agrupamento (1) que teve somente três.

- Simulação da série

A figura 5.8.1 mostra as séries real e simulada do melhor resultado. Mesmo com muitas estações por grupo, inclusive BSR, as simulações

FIGURA 5.19 - Série real e simulada para setembro de 1975. A linha contínua representa a série real e a tracejada a simulada.

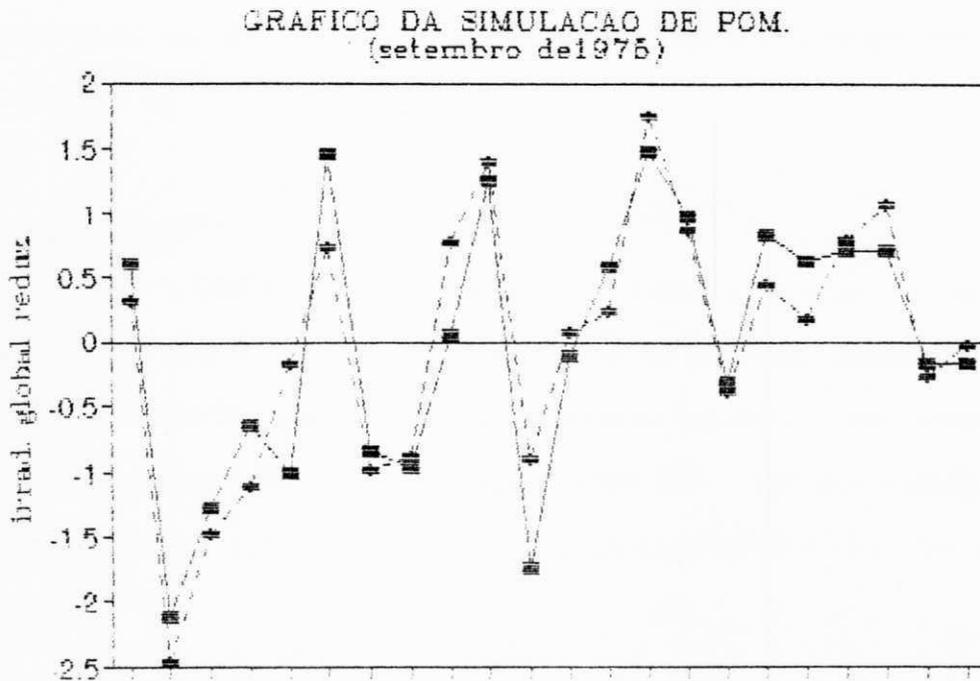
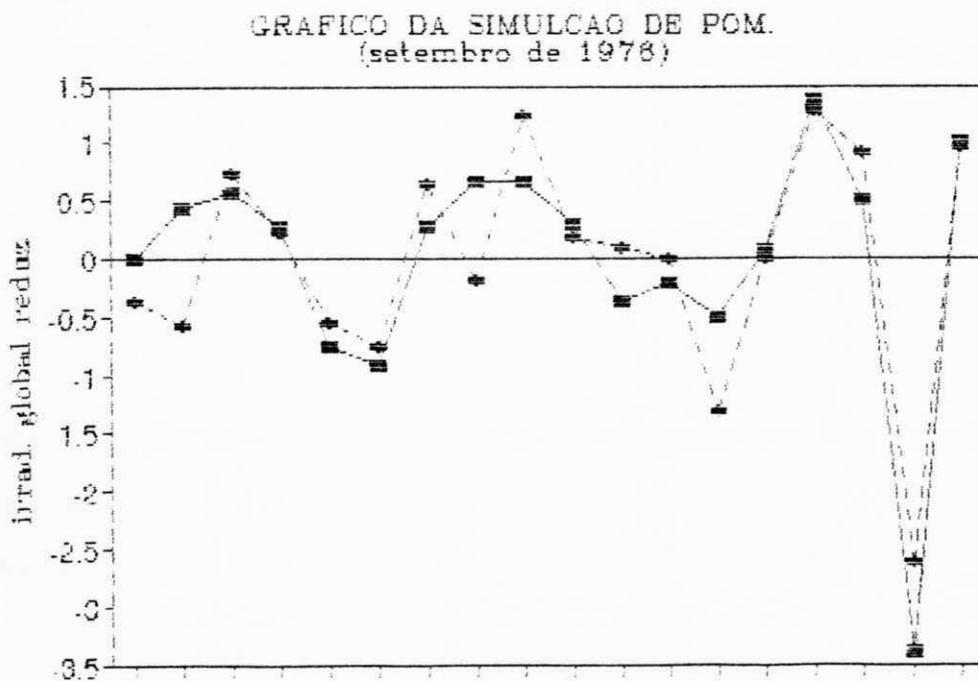


FIGURA 5.20 - Série real e simulada para setembro de 1976. A linha contínua representa a série real e a tracejada a simulada.



apresentaram resultados satisfatórios. O melhor deles foi na obtida com as estações do agrupamento (4), usando uma matriz de aproximação de ordem $K=4$ (os quatro primeiros autovetores explicando 92,3% da variância) obteve-se um desvio padrão do erro de 0.48.

- setembro de 1977.

A tabela 5.9.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Apesar de manter o mesmo número de grupos de 1976, houve uma diminuição das estações em cada um deles. E houve também, uns grupos compostos pelas mesmas estações como foi o caso dos formados com os agrupamentos (2), (3), (4) e (6) cuja composição continha Cabaceiras, Pombal e Surubim.

- Simulação da série.

A figura 5.9.1 mostra as séries real e simulada do melhor resultado. Como as estações dos agrupamentos (3) e (4) são as mesmas, a simulação foi feita só com as do (3). Usando uma aproximação de ordem $K=2$ (os dois primeiros autovetores explicando 95% da variância) resultou num desvio padrão do erro de 0,55.

- setembro de 1978.

A tabela 5.10.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Os grupos encontrados para este ano são quase todos iguais, com excessão dos obtidos nos agrupamentos (4) e (6), nos quais aparece a estação de Picuí. Nos demais, as estações são as mesmas.

FIGURA 5.21 - Série real e simulada para setembro de 1977. A linha contínua representa a série real e a tracejada a simulada.

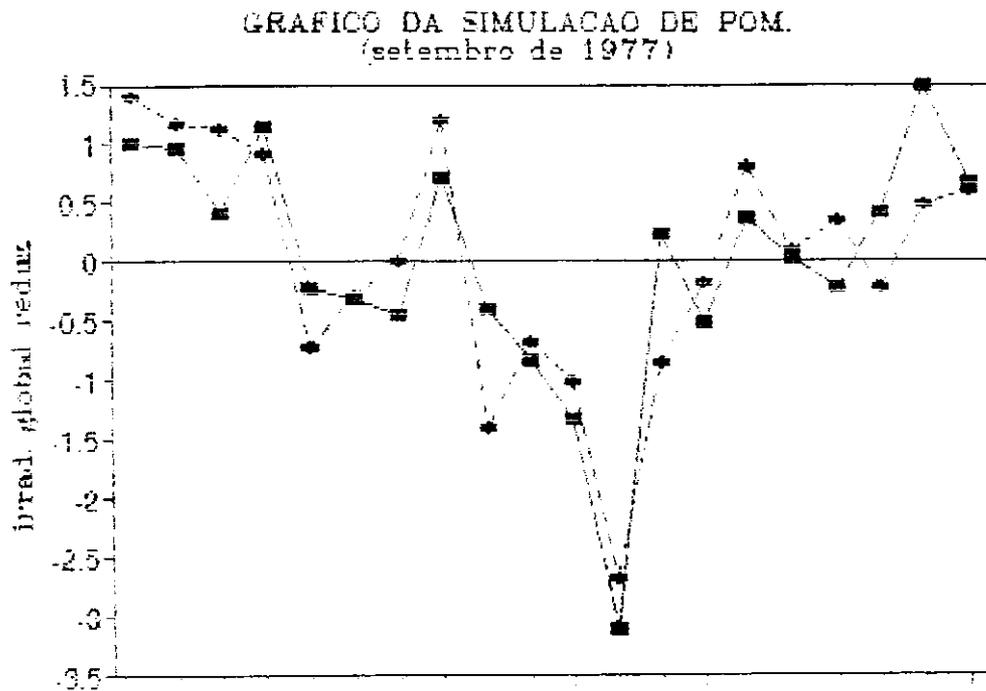
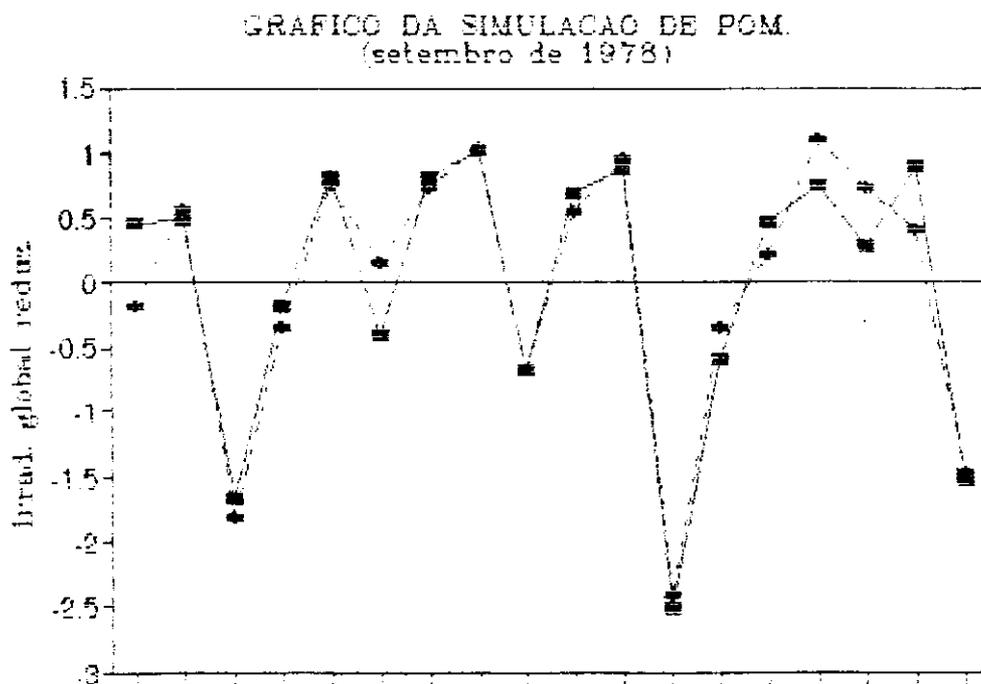


FIGURA 5.22 - Série real e simulada para setembro de 1978. A linha contínua representa a série real e a tracejada a simulada.



- Simulação da série.

A figura 5.10.1 mostra as séries real e simulada do melhor resultado. Apesar de conter muitos grupos com as mesmas estações, os que seriam simulados eram formados por estações diferentes. Portanto, as simulações feitas com (3) e (4) apresentaram excelentes resultados, a melhor delas foi a do agrupamento (4), com uma aproximação de ordem $K=3$ (os três primeiros autovetores explicando 93,8% da variância) obteve-se um desvio padrão do erro de 0,28.

- setembro de 1979.

A tabela 5.11.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Com a diminuição da quantidade de estações da matriz original de dados foi necessário diminuir o número de grupos para três e, em alguns casos, a dois. Alguns grupos ficaram compostos por somente duas estações, como foi o caso dos grupos nos agrupamentos (1), (5) e (6) que ficaram formados por Cajazeiras e Pombal. Os outros grupos dos agrupamentos (3) e (4) tiveram as mesmas estações.

- Simulação da série.

A figura 5.11.1 mostra as séries real e simulada do melhor resultado. Na simulação com as estações do agrupamento (3) os dois primeiros autovetores explicaram 89% da variância e o desvio padrão do erro foi de 0.52.

- setembro - de 1980.

A tabela 5.12.1., no Apêndice, apresenta os resultados obtidos com os agrupamentos dos tipos (1) a (6). Para este ano os grupos, com exceção do primeiro, foram compostos pelas mesmas estações, inclusive com BSR fazendo parte de todos.

- Simulação da série.

A figura 5.12.1 mostra as séries real e simulada do melhor resultado. Usando uma matriz de aproximação $K=2$ (os dois primeiros autovetores explicando 93% da variância) tivemos um desvio padrão do erro de 0.43 que foi diferente do desvio obtido na simulação dos dados de BSR com o mesmo grupo.

FIGURA 5.23 - Série real e simulada para setembro de 1979. A linha contínua representa a série real e a tracejada a simulada.

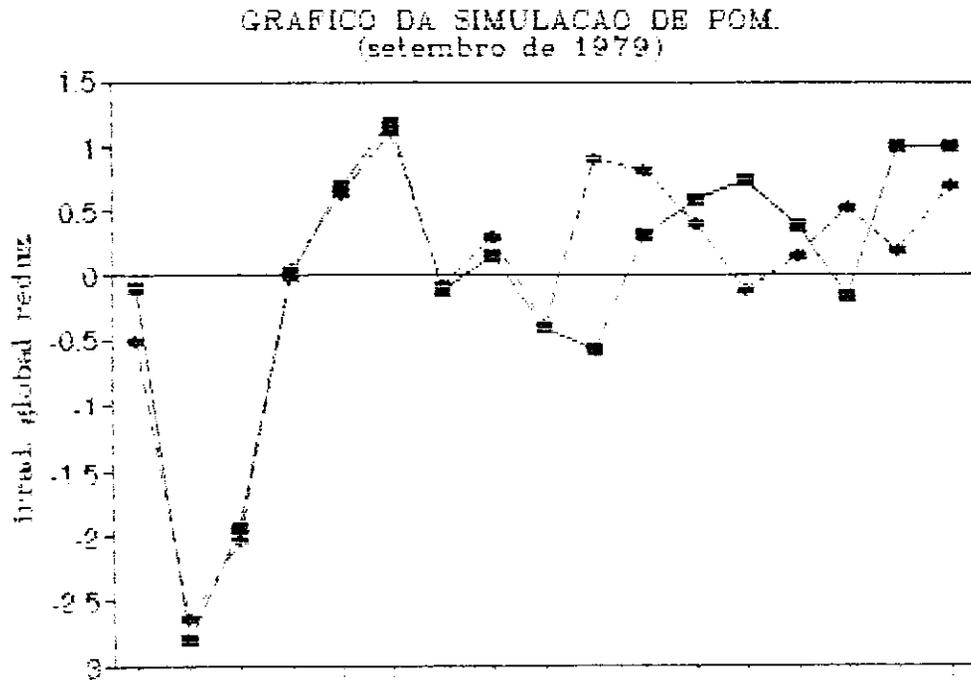
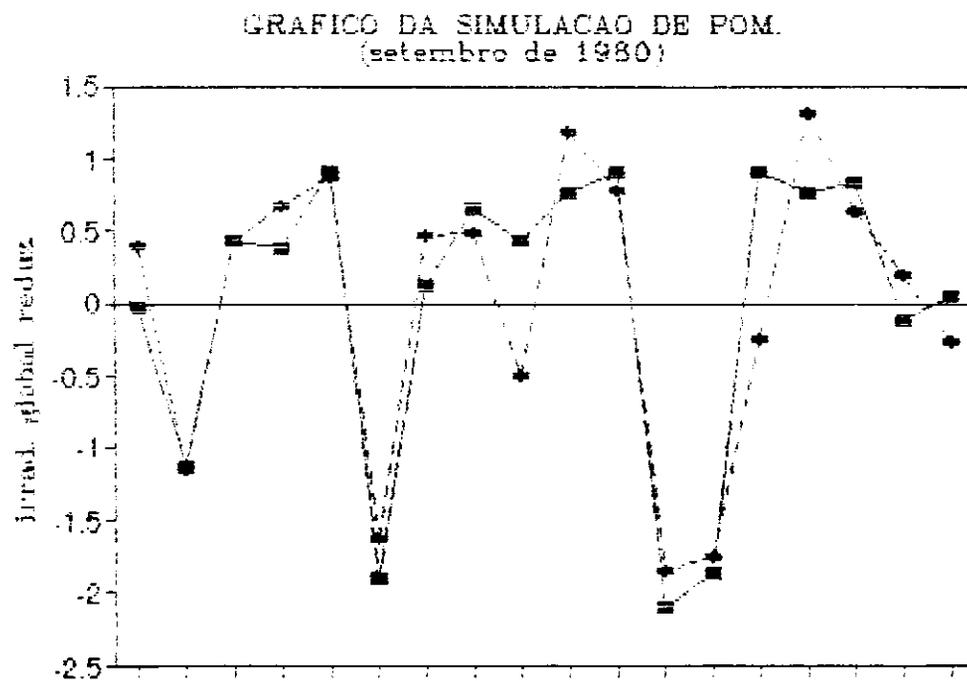


FIGURA 5.24 - Série real e simulada para setembro de 1980. A linha contínua representa a série real e a tracejada a simulada.



5.11 - Grupo permanente para simulação dos dados de Barra de Santa Rosa.

Nas seções anteriores, foi obtido para cada ano do período estudado (1975 a 1980) um conjunto de estações homogêneas agrupadas pelo método de Ward, usando como variáveis os fatores que explicaram 90% da variância. Apenas os três primeiros fatores do grupo serviria de base para todas as simulações. É conveniente ter um grupo permanente de estações que possa ser usado para simular os dados da estação em qualquer ano deste período. O critério para definir essas estações consistiu em determinar, dentre todos os grupos dos agrupamentos (3) e (4), um número de estações (aproximadamente três ou quatro), que em todos os anos fizessem parte do mesmo grupo que continha a estação a ser simulada. A escolha desse número, baseou-se no fato de que simulações feitas tanto com 6 quanto com 3 ou 4 estações não alteravam a qualidade das mesmas. Além disso, como a matriz original de dados ano a ano diminuía progressivamente a quantidade de estações, ficava muito difícil encontrar um grupo que se mantivesse constante durante todo o período.

Devido à diminuição da matriz original de dados, citada acima, e a conseqüente falta de determinadas estações em alguns anos, escolheu-se aquelas que apareciam mais vezes em todo o período. Os resultados foram:

- Mês de março.
- Araruna
- Areia
- Barra de Santa Rosa
- Belém.

- Mês de setembro

- Araruna
- Barra de Santa Rosa
- Campina Grande
- Picuí

5.12 - Grupo permanente para simulação dos dados de Pombal.

- Mês de março

seguinte grupo permanente para simulação dos dados de Pombal:

- Patos
- Pombal
- Princesa Isabel
- Teixeira

- Mês de setembro

Neste mês houve uma grande variabilidade nas estações dos grupos que continha Pombal ao longo do período estudado, dificultando a determinação de um grupo que servisse de base para a simulação proposta. Mesmo assim foi possível encontrar estações que, embora não tivesse presente nos grupos em todos os anos, apareciam em três ou quatro anos da série. Assim, as estações escolhidas para o grupo permanente de simulação dos dados de Pombal são:

- Belém do Brejo do Cruz
- Cabaceiras
- Patos

- Pombal.

- Comparação de resultados.

Fazendo uma simulação de Barra de Santa Rosa com os grupos permanentes, para os meses de março de 1975 e 1976 e comparando os valores dos desvios padrão do erro (SE) obtidos com os da simulações com os grupos originais do agrupamento (3). Chegou-se ao seguintes resultados:

- para março de 1975 com grupo permanente, (SE) = 0,62
- para março de 1975 com grupo original, (SE) = 0,59
- para março de 1976 com grupo permanente, (SE) = 0,48
- para março de 1976 com grupo original, (SE) = 0,43

5.13 - Teste de estabilidade da matriz de correlação entre estações de um mesmo grupo.

Nas seções precedentes, a simulação de séries temporais de estações foi comparada com os valores verdadeiros. Entretanto, o método aplicado tenciona avaliar séries de uma estação no caso em que efetivamente não existia informação sobre ela. Portanto, ao se definir um grupo homogêneo de estações de uma rede, é importante saber até que ponto pode ser usado esse grupo (ou parte dele) para processar simulações em intervalos sem dados.

Como foi visto, os parâmetros básicos para construir uma simulação são os coeficientes de correlação de Pearson, organizados numa matriz de correlação do grupo de estações. Mesmo considerando-se um mês específico, pode se esperar que essa matriz apresente flutuações em anos sucessivos. Para um ano particular, no qual não existam dados, parece razoável utilizar como referência *a matriz média de correlação*. Nesse caso, seria razoável que a série temporal de coeficientes (mensais, flutuando ano-a-ano) seja relativamente estável.

No caso da rede solarimétrica da UFPB, existem três períodos distintos, que denominaremos A, B, C; A: [1975-1980] e C: [1987-1989] têm séries de dados processados e publicados, enquanto que B: [1981-1986] virtualmente carece de dados (embora não de actinogramas). A questão que se coloca é a seguinte: se for adotada uma matriz média de coeficiente de correlação, construída com base nas médias aritméticas dos valores para os anos com dados, ela seria estatisticamente (ou climatologicamente) aceitável?

Uma outra forma de colocar a questão é a seguinte. Considerando o coeficiente de correlação entre duas estações dadas, para o intervalo A obtém-se uma média \bar{X}_1 (para N_1 anos com dados), enquanto que para o C é obtida a média \bar{X}_2 (para N_2 anos com dados). Deseja-se testar a hipótese H_0 contra sua alternativa H_1 :

H_0 = "não há, essencialmente, diferença importante entre \bar{X}_1 e \bar{X}_2 "

H_1 = "a diferença entre \bar{X}_1 e \bar{X}_2 é significativa"

O teste apropriado para esta decisão é o da "t" de Student, com $v=N_1+N_2-2$ grau de liberdade. O estatístico t é avaliado segundo as equações

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{1/n_1 + 1/n_2}} \quad (5.1)$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (5.2)$$

onde s_p é uma média ponderada dos desvios padrões amostrais para \bar{X}_1 e \bar{X}_2 , usando como peso o número de grau de liberdade de cada um.

- Escolha de grupo

nos itens 5.11 e 5.12 deste capítulo, foram estabelecidos grupos permanentes de estações que poderiam ser usadas para simulações de Barra de Santa Rosa e de Pombal. Todavia, para os intervalos A e C o número de estações com dados disponíveis varia bastante. Somente a primeira tinha dados no período C completo. por outro lado, diversas estações não tinham dados durante um período razoavelmente extenso. Decidiu-se limitar a análise a um subgrupo do agrupamento de tipo (três), composto pelas estações de Barra de Santa Rosa, Cabaceiras, Campina Grande e Esperança. A tabela 6.1 apresenta informação sobre as correlações estimadas em cada período.

TABELA 6.1

Correlação entre Barra de Santa Rosa e as outras três estações no mês de setembro.

	Ano						
estações	1975	1976	1977	1978	1979	1980	Média
BSR/CAB	0.52	0.79	0.29	0.54	0.61		0.55
BSR/CGD	0.43		0.68	0.47	0.59	0.85	0.60
BSR/ESP	0.65	0.79	0.82		0.57		0.71

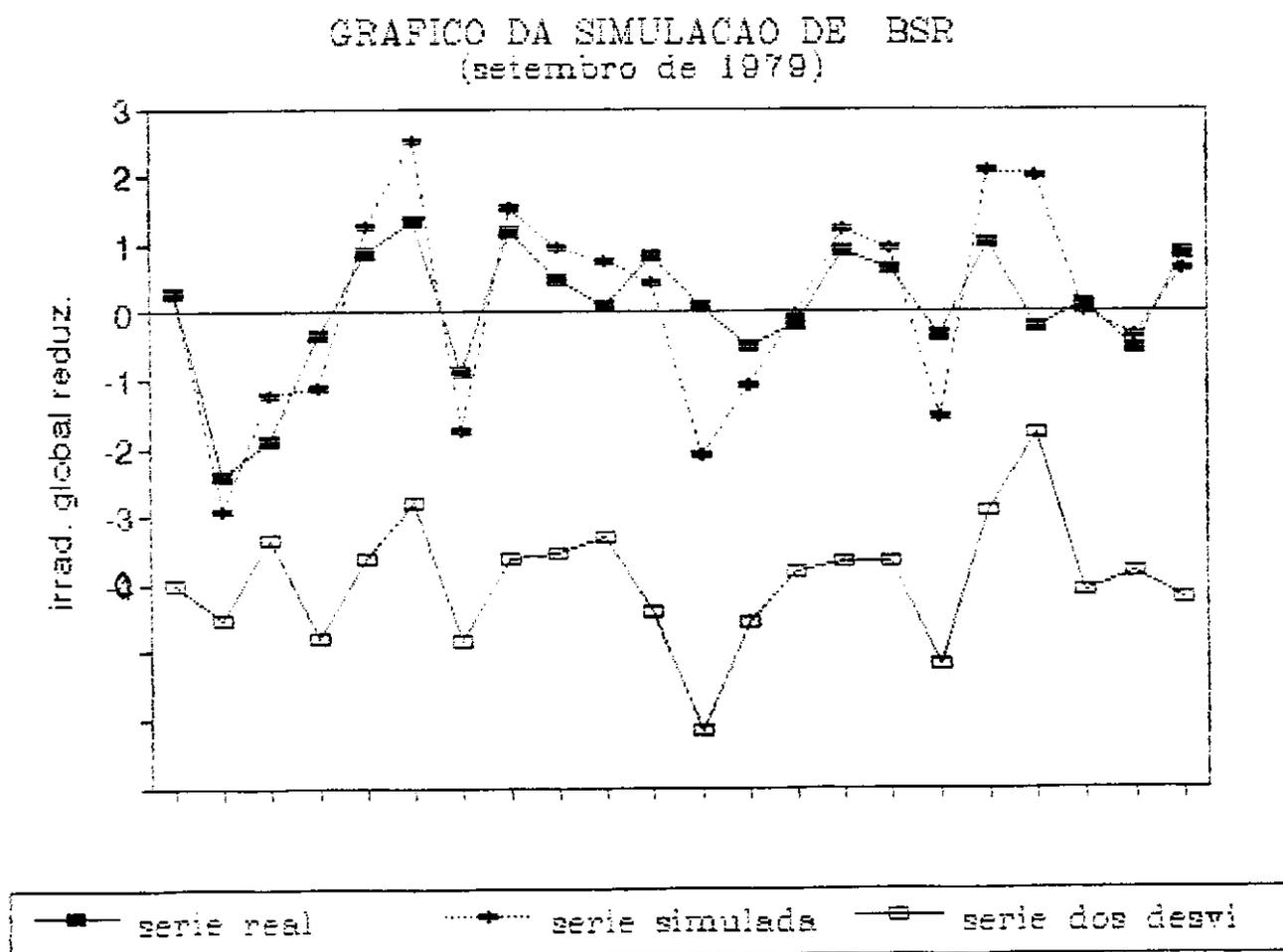
	Ano			
	1987	1988	1989	Média
BSR/CAB	0.23	0.57	0.34	0.38
BSR/CGD	0.55	0.69	0.53	0.59
BSR/ESP	0.73	0.60	0.51	0.61

Usando as equações (5.1) e (5.2), foi avaliado o estatístico t. Com base no número de grau de liberdade, foram determinados valores críticos para o nível de significância $\alpha = 0.05$. Os valores críticos (t_{crit}) encontrados foram 1.94(CAB), 1.94 (CGD) e 2.05 (ESP). Os valores calculados foram, respectivamente, [1.38;0.094;0.086]. Conclui-se que as diferenças entre \bar{X}_1 e \bar{X}_2 não eram significativas para as correlações com CAB, CGD e ESP, a nível de 5% de significância. Portanto, não houve tendência ao longo de 15 anos. Este resultado sugere que valores de referência para o intervalo B poderiam ser as médias dos coeficientes dos intervalos A + C.

Um teste de simulação realizado com Barra de Santa Rosa para o mês de setembro de 1979 usando uma matriz de aproximação de ordem $K=2$ (os dois primeiros autovetores) encontrou-se um valor de $SE=0.89$. Apesar de ter sido

aproximadamente o dobro da média dos valores até então encontrados, e visto que a matriz de correlação apresentou valores discrepantes de um ano para outro, SE é ainda inferior a um desvio padrão da série de Barra de Sanra Rosa, da ordem de 70ly, pode ser considerado como um bom resultado. Com isso, conclui-se que a simulação é satisfatória. A figura 5.25, a seguir, ilustra o resultado da simulação.

Figura 5.25 apresenta a série real (linhas contínuas) e simulada (linhas pontilhadas) e a série dos desvios (linhas tracejadas) para o mês de setembro de 1979.



6 - CONCLUSÃO E RECOMENDAÇÕES

Foi testado um método de simulação de séries temporais de uma estação solarimétrica baseado no uso dos autovetores da matriz de correlação de estações vizinhas, abrangendo uma região considerada homogênea. O método utiliza aqueles autovetores que acumulam 90% da variância total do conjunto de dados reduzidos (padronizados). Este critério é adotado considerando que os erros instrumentais de medições isoladas acumulam 10% da variância típica das estações da rede.

A rede solarimétrica considerada foi a da UFPb, cobrindo o Estado da Paraíba. As regiões homogêneas foram definidas por aplicação de um método de agrupamento. O de Ward mostrou-se mais coerente. Para estabelecer um critério de corte, foi imposta a condição de que o número de estações no grupo regional de interesse específico fosse da ordem de cinco. Este critério evita tanto um excesso de estações com dados faltosos, como o trabalho com grupos excessivamente reduzidos. O resultado foi satisfatório, delimitando-se a Paraíba em cinco regiões coerentes com os regimes climáticos já conhecidos.

Com relação a se fazer uma associação entre os grupos obtidos ano a ano e as mudanças climáticas obtidas em um período, pode-se ressaltar que as mudanças também devem -se verificar na composição dos grupos já que as variáveis usadas representam uma situação meteorológica. E essas estações, se estiverem todas presentes no período total estudado, refletem esta mudança.

O teste sistemático no período de 1975-80 para dois meses (março e setembro) apresentou a dificuldade da grande variação no número de estações, começando por 19 em 1975 e chegando a 7 em 1980. Alguns anos reduziram o número de estações de interesse de um grupo, de forma que já os dois primeiros autovetores acumulavam mais de 90% da variância total dos dados.

Aplicado sistematicamente a duas estações da rede (Barra de Santa Rosa e Pombal) o método mostrou-se satisfatório, podendo reproduzir séries mensais de dados com erro quadrático médio da ordem dos próprios erros instrumentais (5% da média mensal, ou 25 ly diários), e metade do desvio-padrão da série. Em muitos casos, o uso de apenas os dois primeiros autovetores já permite esse grau de aproximação. A qualidade dos resultados fica mais evidente quando se considera que não se trabalhou com dados climatológicos, mas com valores diários. Usualmente, médias climatológicas produzem uma suavização das variações espaciais e temporais, facilitando os processos de interpolação. O uso de dados diários implica em variações mais abruptas entre estações e / ou entre dias diferentes.

Os resultados descritos precedentemente, referem-se a autovetores avaliados com base no conhecimento completo da matriz de correlação entre estações da região (o que implica possuir o conjunto completo dos dados, incluídos aqueles da estação simulada). Em princípio, pode-se pensar que a carência de dados de uma estação durante um mês daria impediria a aplicação do método. Entretanto, testes estatísticos mostram que os coeficientes de correlação para um mesmo mês não tiveram tendência significativa entre o período 1975-80 e 1987-89. O teste foi aplicado para as estações dos grupos que servem de referência para BSR e POM. Para considerar condições reais de carência de dados de uma estação, foi testado o uso da média de coeficientes de correlação para o período de 1975-80 aplicado a um ano dado. O resultado obtido na simulação produziu um erro quadrático médio de 0,89 permitindo concluir que apesar de ser aproximadamente o dobro do valor médio do erro encontrado nas simulações em condições controladas, pode ser considerado como um bom resultado.

Os resultados obtidos sugerem que a matriz de correlação de uma região considerada homogênea pode ser razoavelmente aproximada, no mês dado,

com base no conhecimento de dados precedentes e subsequentes. Estações transitórias podem ser instaladas em locais cuja simulação posterior é desejada, de forma que o regime de correlação com as estações vizinhas possa ser avaliado convenientemente. Uma vez obtida esta informação, a estação pode ser retirada e simulações futuras realizadas. Este método permite a recuperação de dados faltosos em estações solarimétricas. Um outro critério pode considerar, o fato de que o coeficiente de correlação deve ter uma distribuição espacial relativamente suave permite realizar interpolações lineares simples na região, permitindo reter os parâmetros necessários para o método proposto.

Finalmente, o tipo de estudo realizado pode utilizar-se para pesquisar o número mínimo de estações desejáveis para uma rede solarimétrica. Com efeito, quando a série temporal de uma estação poder ser simulada, de forma confiável, a partir de dados de estações vizinhas ela pode ser retirada da rede, reduzindo o total de estações e os custos decorrentes de sua manutenção.

Recomenda-se para trabalhos futuros o desenvolvimento de processo semelhante, usando regiões mais amplas, por exemplo a região Nordeste. Os dados do Zoneamento do Potencial de Radiação Solar do Nordeste (AZEVEDO et al., 1982) seriam de grande utilidade. Na região Sul do Brasil, os dados disponíveis em instituições oficiais de meteorologia podem permitir estudos similares.

7 -BIBLIOGRAFIA.

- ANDERSON, T.W. **An Introduction to Multivariate Statistical Analysis.** Canadá: John Wiley and Sons, 1958. 371p.
- BARROS, V.R; ESTEVAN, E.A. On the evaluation of wind power from short wind records. *Journal of Climate and Appl. Meteorology*; v.22, p1116-1123, jun1983.
- BASSAB, W.O; MLAZAKI, E.S; ANDRADE, D.F. **Introdução à análise de agrupamentos.** São Paulo, 1990. 104p.
- BECKER, C.T. **Determinação de regiões homogêneas de precipitação e temperatura no Estado do Rio Grande do Sul.** Campina Grande: Universidade Federal da Paraíba, 1992. 80p (Dissertação, mestrado em meteorologia).
- BRAGA, C.C; SILVA, B.B.da. Determinação de Regiões Pluviometricamente Homogênea no Estado da Paraíba. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 6, 1990, Salvador. *Anais...* Salvador: Sociedade Brasileira de Meteorologia, 1990. p200-205.
- BRAGA, C.C. Classificação de Regiões Pluviometricamente Homogênea através da Análise Multivariada. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 7, 1992, São Paulo. *Anais...* São Paulo: Sociedade Brasileira de Meteorologia, 1992. p2341-2343.
- BOUROCHE, J.M; SAPORTA, G. **Análise de dados.** trad. Marcos Penchel. Rio de Janeiro: Zahar Editores, 1982. 117p.
- BRIER, G.W; MELTESEN, G.T. Eigenvector Analysis for Prediction of Time Series. *Journal of Climate and Appl. Meteorology*; v.22, p1307-1312, 1976.

- CEBALLOS, J.C; BRAGA, C.C. **Simulação de Séries Temporais de Irradiação Solar: Uma aplicação ao Estado da Paraíba.** Serimet, DCA/UFPb, n.3, 92p, 1991.
- CEBALLOS, J.C; BRAGA, C.C. Simulation of time Series in a Solarimetric network. *Journal of Climatology*, 1994. No prelo.
- COSTA NETO, P.L.O. **Estatística.** São Paulo: Edgar Blucher, 1977. p34-134.
- CURI, P.R. Análise de agrupamento: Métodos sequenciais, aglomerativos e hierárquicos. *Ciência e Cultura*; v.35, n.10, p1416-1427, out. 1983.
- CURI, P.R. Análise de agrupamento complementada com ordenação pelos componentes principais e análise de variância multivariada. Um exemplo biológico. *Ciência e Cultura*; v.37, n.6, p879-887, jun. 1985.
- DIEHL, P; SILVA, M.G.A.J. Componentes Principais da Precipitação em Minas Gerais In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 7, 1992, São Paulo. *Anais...* São Paulo: Sociedade Brasileira de Meteorologia, 1992. p280-284.
- EDER, B. K; DAVIS, J.M; MONAHAN, J.F. Espatial Analysis of the Palmer Druogh Severity Index Over the South-Eastern United States. *Journal of Climatology* v. 7, p31-56, 1987.
- ESSENWANGER, O. **Applied Statistics in Atmospheric Science.** New York: Elsevier Scientific Publishing Company, 1976. p252-289.
- EVERITT, B. **Cluster Analysis.** London: Heinemann Education Books, 1974. 135p.
- FIGUEIREDO, M. L. **Geografia da Paraíba.** 1ed, Campina Grande: Gráfica Santa Fé, 1987. p1-35.
- GAMA, M.P. **Base de Análise de Agrupamentos.** Brasília: Universidade de Brasília, 1980. 229p (Dissertação, mestrado em estatística).

- HARDY, D.M; WALTON, J.J. Principal Components Analysis of Vector Wind Measurements. **Journal of Applied Meteorology**. v.17, p1153-1162, 1978.
- ISANTA, L.A. Classificação de Regiones Climaticas por médio de los vectores propios de la precipitacion. **Revista de Meteorologia**.1984.
- JOLLIFFE, I.T. **Principal Component Analysis**; Springer Series in Statistics. New York: Springer-Verlag, 1986. 217p.
- KUTZBACH, J.E. Empirical eigenvectors of sea level pressure, surface temperature and precipitation complexes over North America. **Journal of Applied Meteorology**. v.6, p791-828, 1967.
- LORENZ, E.N. **Empirical orthogonal functions and statistical weather prediction**. Massachusetts: Sci. Rrp. n1.Statistical Forecasting Project Department of Meteorology. 1956. 49p.
- MORETTIN, P.A; TOLOI, C.M. **Métodos Quantitativos, Séries temporais**. São Paulo: Atual. 1986. 133p.
- MORRISON. **Multivariate Statistical Methods**. Second edition, McGraw-Hill, 1967. 413p.
- PANDZIC, K. Principal Component Analysis of Precipitation in the Adriatic-Pannonian Area of Yugoslavia. **Journal of Climatology**. v.8, p358-370, 1988.
- PANDZIC, K; KISEGI, M. Principal Component Analysis of a Local Precipitation Field within the Global Circulation. **Beitr.Phys. Atmosph.** v.63, p101-116, 1990.
- PISSANI,J.F. Análise Estatística Multidimensional em Biologia. **Ciência e Cultura**; v.21, n.3, p619-631, set.1960.
- POSSOLI, S. Técnicas de Análise Multivariada para avaliação das condições de saúde dos municípios do Rio Grande do Sul,Brasil. **Rev. Saúde Publ.** v18, p288-299, 1984.

- SILVA, M.G.A.J.da; GARAYALDI, E.J.G; SANTOS, I.A. O uso de Autovetores empíricos no diagnóstico da interrelação entre configurações da pressão. In: CONGRESSO INTERAMERICANO DE METEOROLOGIA/CONGRESSO BRASILEIRO DE METEOROLOGIA, 4, 1986, Brasília. Anais...Brasília: Sociedade Brasileira de Meteorologia,1986. p149-154.
- SILVA, M.G.A.J.da; SANTOS, I.A.de; SILVA, A.S.da. Componentes Principais da estrutura troposférica da Região Equatorial Brasileira. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 7, 1992, São Paulo. Anais...São Paulo: Sociedade Brasileira de Meteorologia,1992. p205-209.
- SPIEGUEL, M.R. *Estatística*. Coleção Schãum. Trad. Pedro Consentino. São Paulo: McGraw-Hill do Brasil,Coleção Schãum. 1974. 580p.
- STID, C.K. The use of eigenvector for climatic estimates. *Journal of Applied Meteorology*. v.6, p255-264, 1967.
- TERRON, M.E.P. Aplicacion de los Componentes Principales al análisis de los Campos Meteorológicos. *Revista Cubana de Meteorologia*. v.3, p71-75, 1990.

APÊNDICE

TABELA 5.1

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1975.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	10 fatores WARD	3 fatores WARD	10 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	12 grupos	12 grupos
ARA BSR JPS MOG PIC	ARA ARE BSR BEL CAB ESP MOG PIC SUR	ARA ARE BSR BEL ESP MOG	ARE BSR BEL ESP PAT POM TEI	ARA ARE BSR ESP MON PAT POM TEI	ARE BSR CAB ESP PAT PIC POM TEI SUR
3atv--[87%] SE=0,70	4atv--[88%] SE=0,68	3atv--[90%] SE=0,58	4atv--[90%] SE=0,61	4atv--[90%] SE=0,70	5atv--[90%] SE=0,75
2atv--[77%] SE=0,70	2atv--[72%] SE=0,63	2atv--[81%] SE=0,59	2atv--[77%] SE=0,62	2atv--[70%] SE=0,70	2atv--[69%] SE=0,69

TABELA 5.2

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1976.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	8 fatores WARD	3 fatores WARD	8 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	12 grupos	7 grupos
ARA	ARA	ARA	ARA	ARA	ARA
	ARE	ARE	ARE	ARE	ARE
BSR	BSR	BSR	BSR	BSR	BSR
ESP	ESP	ESP	ESP	ESP	ESP
PIC	PIC	PIC	PIC		
			TEI		
		SUR			
		3atv--[90%] SE=0,43	3atv--[91%] SE=0,44		
		2atv--[83%] SE=0,43	2atv--[85%] SE=0,45		

TABELA 5.3

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1977.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	11 grupos	7 grupos
ARA	ARA	ARA	ARA	ARA	ARA
ARE	ARE	ARE	ARE	ARE	ARE
BSR	BSR	BSR	BSR	BSR	BSR
CAJ	CAJ	CAJ	CAJ	CAJ	CAJ
	CGD	CGD	CGD		CGD
	ESP	ESP	ESP	ESP	ESP
JPS					
MOG	MOG	MOG	MOG		MOG
PIC	PIC	PIC	PIC		PIC
	SUR	SUR	SUR		SUR
		3atv--[90%] SE=0,42			
		2atv--[85%] SE=0,48			

TABELA 5.4

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1978.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	5 fatores WARD	3 fatores WARD	5 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	4 grupos	4 grupos
ARA BSR	ARA ARE BSR	ARA ARE BSR CGD	ARA ARE BSR CAB CGD	ARA BSR BEL PIC	ARA BSR BEL PIC
		2atv--[91%] SE=0,53	2atv--[88%] SE=0,57		

TABELA 5.5

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1979.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	5 fatores WARD	3 fatores WARD	5 fatores CENTRÓIDE	3 fatores CENTRÓIDE
4 grupos	4 grupos	4 grupos	3 grupos	4 grupos	7 grupos
ARA BSR BEL CAB PIC	ARA BSR BEL CAB CGD ESP MOG PIC	ARA BSR BEL CAB PIC	ARA BSR BEL CAB PIC PRI	ARA BSR BEL CAB CGD ESP MOG PIC	BSR CAB PIC
		3atv--[92%] SE=1,0	3atv--[87%] SE=0,84		
		2atv--[84%] SE=0,48	2atv--[75%] SE=0,75		

TABELA 5.6

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1980.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	4 fatores WARD	3 fatores WARD	4 fatores CENTRÓIDE	3 fatores CENTRÓIDE
2 grupos	2 grupos	2 grupos	2 grupos	4 grupos	3 grupos
BSR	BSR	BSR	BSR	BSR	BSR
	BEL	BEL	BEL		
CGD	CGD	CGD	CGD	CGD	CGD
	JPS	JPS	JPS	JPS	JPS
		2atv--[91%] SE=0,58			

TABELA 5.7

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1975.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	12 grupos	7 grupos
ARE BSR CGD ESP MOG SUR	ARA ARE BSR CAB CGD ESP MOG PIC CRU SUR	ARA ARE BSR CAB CGD ESP MOG PIC CRU SUR	BSR CAB MOG PIC CRU	ARE BSR CGD ESP MOG CRU SUR	BSR CAB CGD MOG PIC CRU
		4atv--[93%] SE=0,44	3atv--[94%] SE=0,50		
		2atv--[89%] SE=0,48	2atv--[89%] SE=0,48		

TABELA 5.8

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1976.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	7 grupos	6 grupos
ARA BSR CGD ESP MOG	ARA BSR BEL CAB CGD ESP MOG MON PIC POM CRU	ARA BSR BEL CAB CGD ESP MOG MON PIC POM CRU	BSR ESP MOG	ARA BSR BEL CAB CGD ESP MOG PIC POM CRU	BSR ESP MOG
		4atv--[91%] SE=0,61	3atv--[99%] SE=0,74		
		2atv--[87%] SE=0,63			

TABELA 5.9

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1977.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	7 grupos	6 grupos
ARA	ARA	ARA	ARA	ARA	ARA
	ARE	ARE	ARE	ARE	ARE
BSR	BSR	BSR	BSR	BSR	BSR
CAJ	CAJ	CAJ	CAJ	CAJ	CAJ
	CGD		CGD		CGD
ESP			ESP		ESP
			JPS		JPS
			MOG		MOG
	PIC	PIC	PIC	PIC	PIC
			3atv--[89%] SE=0,39		
		2atv--[95%] SE=0,18	2atv--[95%] SE=0,42		

TABELA 5.10

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1978.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	5 fatores WARD	3 fatores WARD	5 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	4 grupos	4 grupos	5 grupos	5 grupos
ARA BSR JPS MON PIC	ARA BSR JPS PIC	ARA BSR JPS PIC	ARA BSR CGD JPS MOG MON	ARA BSR JPS PIC	ARA BSR CGD JPS MOG MON
			3atv--[94%] SE=0,41		
		2atv--[94%] SE=0,41	2atv--[88%] SE=0,43		

TABELA 5.11

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1979.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	4 fatores WARD	3 fatores WARD	4 fatores CENTRÓIDE	3 fatores CENTRÓIDE
3 grupos	3 grupos	3 grupos	3 grupos	3 grupos	3 grupos
BSR BEL	BSR CGD JPS	BSR BEL CAB ESP	BSR BEL CAB ESP	BSR CAB CGD ESP JPS	BSR CAB CGD ESP JPS
		2atv-[93%] SE=0,48			

TABELA 5.12

Composição dos grupos que incluem BSR, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1980.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	4 fatores WARD	3 fatores WARD	4 fatores CENTRÓIDE	3 fatores CENTRÓIDE
3 grupos	3 grupos	3 grupos	3 grupos	4 grupos	3 grupos
BSR BEL JPS MON	BSR BEL CAB CGD POM	BSR BEL CAB CGD POM	BSR BEL CAB CGD POM	BSR BEL CAB CGD POM	BSR BEL CAB CGD POM
		2atv-[93%] SE=0,16			

TABELA 5.1.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1975.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	10 fatores WARD	3 fatores WARD	10 fatores CENTRÓID E	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	12 grupos	7 grupos
ARE BEL CAB CGD ESP MON PAT POM TEI SUR	MON PAT POM TEI	MON PAT POM TEI	ARE BSR BEL ESP PAT POM TEI	ARA ARE BSR ESP MON PAT POM TEI	ARE BSR BEL CAB ESP PAT PIC POM TEI SUR
		3atv--[93%] SE=0,48	4atv--[90%] SE=0,68		
		2atv--[85%] SE=0,67	2atv--[77%] SE=0,64		

TABELA 5.2.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1976.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	8 fatores WARD	3 fatores WARD	8 fatores CENTRÓIDE	3 fatores CENTRÓIDE
3 grupos	3 grupos	3 grupos	3 grupos	2 grupos	3 grupos
BEL CAJ PAT POM TEI	BEL CAB PAT POM TEI CRU	 CAJ POM PRI BAR	BEL CAJ MON PAT POM CRU	ARA ARE BSR BEL CAB ESP JPS MOG MON PAT PIC POM PRI TEI CRU SUR	BEL CAJ MON PAT POM
		3atv--[93%] SE=0,93	3atv--[90%] SE=0,52		
		2atv--[81%] SE=0,84	2atv--[82%] SE=0,57		

TABELA 5.3.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1977.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
3 grupos	3 grupos	4 grupos	3 grupos	5 grupos	2 grupos
BEL CAB CGD ESP PAT POM SUR	 MON PAT POM PRI TEI BAR CRU	 MON PAT POM PRI TEI CRU	 MON PAT POM PRI TEI BAR CRU	 MON PAT POM PRI	 MON PAT POM PRI TEI BAR CRU
		3atv--[92%] SE=0,81	3atv--[87%] SE=0,67		
		2atv--[85%] SE=0,72	2atv--[80%] SE=0,68		

TABELA 5.4.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1978.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	5 fatores WARD	3 fatores WARD	5 fatores CENTRÓIDE	3 fatores CENTRÓIDE
2 grupos	2 grupos	2 grupos	3 grupos	2 grupos	2 grupos
BEL CAJ PAT PIC POM	 CAB CAJ CGD MON PAT POM PRI TEI	 CAB CAJ MON PAT POM PRI TEI	ARA ARE BSR CAB CGD POM	 POM	 POM
		3atv--[88%] SE=1,27	3atv--[91%] SE=1,11		
		2atv--[75%] SE=1,24	2atv--[81%] SE=1,18		

TABELA 5.5.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1979.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	5 fatores WARD	3 fatores WARD	5 fatores CENTRÓIDE	3 fatores CENTRÓIDE
2 grupos	2 grupos	2 grupos	3 grupos	3 grupos	2 grupos
CAJ JPS POM PRI	CAJ POM	CAJ POM PRI	CAJ POM	CAJ POM	CAJ POM
		2atv--[89%] SE=0,90			

TABELA 5.6.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para março de 1980.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	4 fatores WARD	3 fatores WARD	4 fatores CENTRÓIDE	3 fatores CENTRÓIDE
2 grupos	2 grupos	2 grupos	3 grupos	4 grupos	3 grupos
BEL CAJ JPS MON POM	CAJ MON POM	CAJ MON POM	CAJ MON POM	CAJ POM	CAJ MON POM
		2atv--[89%] SE=0,79			

TABELA 5.7.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1975.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	5 grupos	4 grupos
BEL MON PAT POM PRI TEI BAR	BEL PAT POM TEI	BEL PAT POM	 CAJ PAT POM	BEL MON PAT POM PRI TEI	 CAJ PAT POM
		2atv--[95%] SE=0,47	2atv--[97%] SE=0,41		

TABELA 5.8.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1976.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	7 grupos	6 grupos
CAJ	ARA BSR BEL CAB	ARA BSR BEL CAB	ARA BEL CAB	ARA BSR BEL CAB	ARA BEL CAB
POM	CGD ESP MOG MON	CGD ESP MOG MON	CGD MON	CGD ESP MOG	CGD MON
PRI	PIC POM	PIC POM	PAT PIC POM	PIC POM	PAT PIC POM
	CRU	CRU	TEI CRU	CRU	TEI CRU
		4atv--[91%] SE=0,63	4atv--[92%] SE=0,48		
		2atv--[87%] SE=0,63	2atv--[83%] SE=0,50		

TABELA 5.9.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1977.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	6 fatores WARD	3 fatores WARD	6 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	5 grupos	7 grupos	6 grupos
BEL	CAB	CAB	CAB	CAB CGD JPS MOG	CAB
PAT PIC POM TEI CRU	POM	POM	POM	POM	POM
	SUR	SUR	SUR	SUR	SUR
		2atv--[95%] SE=0,55			

TABELA 5.10.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1978.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	5 fatores WARD	3 fatores WARD	5 fatores CENTRÓIDE	3 fatores CENTRÓIDE
5 grupos	5 grupos	5 grupos	3 grupos	5 grupos	4 grupos
BEL PAT POM PRI CRU	BEL PAT POM PRI CRU	BEL PAT POM PRI CRU	BEL PAT PIC POM PRI CRU	BEL PAT POM PRI CRU	BEL PAT PIC POM PRI CRU
			3atv--[94%] SE=0,28		
		2atv--[91%] SE=0,32	2atv--[88%] SE=0,34		

TABELA 5.11.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1979.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	4 fatores WARD	3 fatores WARD	4 fatores CENTRÓIDE	3 fatores CENTRÓIDE
3 grupos	2 grupos	2 grupos	2 grupos	3 grupos	3 grupos
CAJ	BSR CAJ CGD JPS	CAJ CGD JPS	CAJ CGD JPS	CAJ	CAJ
POM	POM	POM	POM	POM	POM
		2atv--[88%] SE=0,52			

TABELA 5.12.1

Composição dos grupos que incluem POM, para diferentes métodos de agrupamentos e variáveis utilizadas. Simulações para setembro de 1980.

(1)	(2)	(3)	(4)	(5)	(6)
var. brutas WARD	var. reduz. WARD	4 fatores WARD	3 fatores WARD	4 fatores CENTRÓIDE	3 fatores CENTRÓIDE
3 grupos	3 grupos	3 grupos	3 grupos	4 grupos	3 grupos
CAB CAJ	BSR BEL CAB	BSR BEL CAB	BSR BEL CAB	BSR BEL CAB	BSR BEL CAB
POM	CGD POM	CGD POM	CGD POM	CGD POM	CGD POM
		2atv--[93%] SE=0,43			