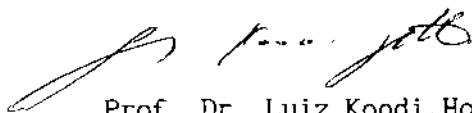


APLICAÇÃO DAS TÉCNICAS DE COMPONENTES PRINCIPAIS
E AGRUPAMENTOS EM PLUVIOMETRIA:
ANÁLISES DO NORDESTE PARAENSE E ESTADO DE SÃO PAULO

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Sérgio Antonio Zullo e aprovada pela Comissão Julgadora.

Campinas, 03 de julho de 1992.



Prof. Dr. Luiz Koodi Hotta

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em Estatística.

Dedico esse trabalho a meus pais,
Jurandir e Maria Aparecida.

ÍNDICE

Agradecimentos	i
Resumo	ii
Abstract	iii
Lista de figuras	v
Lista de quadros	vii
CAPÍTULO I - INTRODUÇÃO	1
CAPÍTULO II - REVISÃO BIBLIOGRÁFICA	5
CAPÍTULO III - NORDESTE PARAENSE	9
3.1 - Dados e descrição climática	11
3.1.1 - Situação geográfica e climática	11
3.1.2 - Dados utilizados	13
3.2 - Análise descritiva dos dados	13
3.2.1 - Características temporais	14
3.2.2 - Características espaciais	16
3.3 - Análise de componentes principais	18
3.4 - Regionalização pluviométrica	26
3.5 - Considerações gerais	30
3.5.1 - Análise de componentes principais	31
3.5.2 - Regionalização pluviométrica	33
3.6 - Conclusões	35
CAPÍTULO IV - ESTADO DE SÃO PAULO	37
4.1 - Dados e descrição climática	39
4.1.1 - Situação geográfica e climática	39
4.1.2 - Dados utilizados	41

4.2 - Análise descritiva dos dados	42
4.2.1 - Características temporais	42
4.2.2 - Características espaciais	46
4.2.3 - Consistência dos dados	49
4.3 - Análise de componentes principais	51
4.4 - Regionalização pluviométrica	56
4.5 - Conclusões	62
CAPÍTULO V - CONCLUSÕES	65
APÊNDICE A - ANÁLISE DE COMPONENTES PRINCIPAIS	68
APÊNDICE B - ANÁLISE DE AGRUPAMENTOS	74
APÊNDICE C - DESCRIÇÃO DAS ESTAÇÕES PLUVIOMÉTRICAS - NORDESTE PARAENSE	80
APÊNDICE D - DESCRIÇÃO DAS ESTAÇÕES PLUVIOMÉTRICAS - ESTADO DE SÃO PAULO	85
BIBLIOGRAFIA	92

AGRADECIMENTOS

Gostaria de agradecer às pessoas e entidades que de alguma forma contribuíram para a realização desse trabalho. Em especial sou grato a:

- Prof. Dr. Luiz Koodi Hotta, pela maneira incansável e valiosa com a qual orientou esse trabalho;

- Profs. Drs. Hilton S. Pinto, Regina Célia C. P. Moran e Belmer G. Negrillo, pelos importantes comentários e sugestões que fizeram;

- Depto. de Estatística do IMECC/UNICAMP e CEPAGRI/UNICAMP, através dos quais foi possível a execução dessa tese;

- CNPq, FAEP/UNICAMP e FINEP, pelo fornecimento dos recursos financeiros necessários;

- DAEE do Estado de São Paulo, EMBRAPA/CPATU, EMBRAPA/PIRELLI, INEMET e DNAEE, entidades que cederam os dados disponíveis para a execução das análises efetuadas;

- Jurandir Zullo Júnior, pelo apoio em todos os momentos;

- Profa. Dra. Tatiana D. de A. Sá, pela preciosa contribuição;

- Convest/UNICAMP;

- aos colegas Claudir, Edilene, Marcos, Maria do Rosário e Rosana.

RESUMO

As técnicas de componentes principais e agrupamentos são aplicadas a dados de pluviometria, visando o conhecimento dos padrões de chuva predominantes nas áreas investigadas e a determinação de regiões homogêneas. São analisados dois casos: nordeste paraense e Estado de São Paulo, Brasil.

Na análise do nordeste paraense são utilizados os dados de precipitação pluviométrica mensal de 22 estações, correspondentes ao período de 1983 a 1988. A análise de componentes principais consegue destacar, através dos 3 primeiros componentes, as características das chuvas no período de transição chuvoso-seco, no período seco e no período chuvoso, respectivamente. O método de agrupamento de Ward consegue, com 5 grupos, agrupar de maneira satisfatória as estações levando-se em consideração essas características.

Para o Estado de São Paulo são analisados os totais pluviométricos de cinco dias para o mês de setembro, de 1961 a 1988, sendo utilizadas as observações de 391 estações meteorológicas. Através dos mapas dos valores dos componentes principais consegue-se observar a influência de características espaciais na atuação das chuvas. A aplicação dos métodos de agrupamento de Ward e da ligação média mostra diferentes resultados quando comparados entre si. No entanto, levando-se em consideração as características de cada método, tem-se que eles procuram destacar diferentes informações do conjunto de dados.

As análises mostram, de uma maneira geral, a importância de se conhecerem as características específicas de cada método.

ABSTRACT

Principal component analysis is used to derive spatial and temporal patterns of rainfall data, and the cluster analysis is latter applied to find homogeneous regions. Two cases are analised: northeast of Para State and São Paulo State, both in Brazil.

The analysis of the northeast of Para State uses the monthly total profiles of the precipitation at 22 stations, covering the period from 1983 to 1988. The first three principal components emphasize the period of wet-dry transition, the amount of rainfall during the dry period and the amount of rainfall during the wet period, in this order. Ward's cluster method with five groups produces a regionalization closely related with Köppen classification and in agreement with the principal component analysis.

The São Paulo State analysis is based on total pentad rainfall during September. The data are collected at 391 meteorological stations covering the period from 1961 to 1988. The first principal component score is associated with the total rainfall. The gradient of the isolines of the second and third principal components scores mapping is, respectively, in the northeast-southwest and the northwest-southeast direction. The second component is related with frontal systems, but the third principal component is not immediatly interpretable. The application of the Ward's and the average linkage clustering methods yields different regionalization. In both cases, however, the characteristics of the clusters are interpretable showing different aspects of the data. The results show

that the objectives of the clustering plays an important role in the choice of the clustering method.

LISTA DE FIGURAS

Figura	Pág.
1 - Localização das regiões estudadas	3
2 - Mapa situando o nordeste paraense, com as estações uti- lizadas na análise	11
3 - Perfis de chuva das estações analisadas	14
4 - Características gerais das variáveis estudadas	15
5 - Total pluviométrico anual médio no nordeste paraense durante o período de 1983 a 1988	16
6 - Total pluviométrico médio do trimestre fevereiro-abril durante o período de 1983 a 1988	17
7 - Total pluviométrico médio do trimestre agosto-outubro durante o período de 1983 a 1988	17
8 - Diagrama scree	20
9 - Coeficientes dos seis primeiros componentes principais .	22
10 - Dispersão dos postos de acordo com os dois primeiros componentes principais	23
11 - Dendrograma gerado através do método de Ward	27
12 - Perfil de chuva por região homogênea	28
13 - Regiões pluviométricas homogêneas do nordeste paraense segundo o método de Ward	29
14 - Relevo do Estado de São Paulo	40
15 - Mapa situando o Estado de São Paulo, com as estações utilizadas na análise	41
16 - Média das chuvas nos quinqüídios de agosto a outubro	

durante o período analisado	43
17 - Médias das variáveis estudadas	44
18 - Diagrama box-plot após transformação logarítmica para os qüinqüídios de setembro (apenas com ocorrência de chuva)	45
19 - Média das chuvas por qüinqüídio, em setembro, durante os períodos de 1961/74 e 1975/88	46
20 - Precipitação pluviométrica média em setembro durante o período de 1961 a 1988	47
21 - Precipitação pluviométrica média em setembro durante o período de 1961 a 1988, por qüinqüídio	48
22 - Diagrama scree	52
23 - Coeficientes dos três primeiros componentes principais .	54
24 - Valores dos três primeiros componentes principais	55
25 - R^2 por número de grupos para o método de Ward	57
26 - Agrupamento proposto pelo método de Ward	58
27 - Chuvas médias por qüinqüídio para as regiões destacadas pelo método de Ward	59
28 - Agrupamento proposto pelo método da ligação média	60
29 - Comparação entre perfis de chuva para algumas estações destacadas pelo método da ligação média	61
30 - Exemplo da atuação da transformação de comp. principais para um caso bidimensional	68
31 - Perfis de chuva de 1983 a 1988 das estações pluviomé- tricas do nordeste paraense	81
32 - Perfis de chuva dos agrupamentos propostos pelo método de Ward	88

LISTA DE QUADROS

Quadro	Pág.
1 - Estrutura da matriz de correlação analisada	19
2 - Proporção da variabilidade total explicada pelos comp. principais	20
3 - Proporção da variabilidade explicada pelos autovetores para diferentes análises de comp. principais	32
4 - Correlação entre os autovetores obtidos a partir das matrizes de covariância e de correlação sem 83, com os obtidos a partir da matriz de correlação	32
5 - Frequência relativa de chuvas iguais a 0 mm, inferiores a 5 mm (inclusive 0 mm) e superiores a 20 mm, por quinqüídio	44
6 - Proporção da variabilidade total explicada pelos comp. principais	52

CAPÍTULO I

INTRODUÇÃO

Os elementos meteorológicos ocorrentes no globo terrestre têm as suas intensidades e distribuições reguladas por diversos fatores. Esses fatores, tais como latitude, circulação geral, barreiras orográficas, correntes marítimas, etc., condicionam a regularidade diferencial de parâmetros meteorológicos, para diferentes locais (Ometto, 1981, p. 389). Tendo em vista essa situação e o fato de que os fenômenos naturais apresentam grande influência na qualidade de vida humana, como elementos fundamentais a diversas atividades, torna-se essencial o conhecimento da distribuição temporal e espacial das variáveis climáticas.

Um dos principais fenômenos naturais, sobretudo nas regiões tropicais, é a precipitação pluviométrica. Dela dependem vários ramos de atividade humana, estando diretamente relacionada, entre outros, ao gerenciamento de recursos hídricos para a produção de energia elétrica e para o abastecimento de água. Em agricultura, a precipitação pluviométrica é um dos principais fatores determinantes do sucesso ou fracasso de atividades como o preparo de terra, plantio de culturas, irrigação e colheita.

Dentre as técnicas de Estatística empregadas para se estudar a precipitação pluviométrica umas das mais utilizadas modernamente têm sido as de análise multivariada, especialmente as análises de componentes principais e de agrupamentos.

O objetivo desse trabalho é avaliar uma metodologia a ser

utilizada no estudo da distribuição espacial e temporal da precipitação pluviométrica, visando a determinação de regiões homogêneas, através da análise de dois casos reais. Essa metodologia baseia-se na aplicação de técnicas descritivas aos dados para uma exposição inicial das características espaciais e temporais das chuvas nas áreas estudadas. O conhecimento dos padrões de chuva predominantes nas áreas sob investigação será buscado através da utilização da análise de componentes principais, que terá também a função de reduzir o número de variáveis a serem estudadas. A análise de agrupamentos será utilizada posteriormente para definir as regiões pluviometricamente homogêneas, levando-se em consideração as características mais gerais das chuvas destacadas pelos componentes principais.

A definição de regiões pluviometricamente homogêneas, além de propiciar um melhor conhecimento sobre o comportamento das chuvas na área sob consideração, pode ser de grande utilidade na tomada de decisão e adoção de técnicas conjuntas em atividades que sejam dependentes da pluviosidade. Nesse sentido, tem sido de grande valia a aplicação de técnicas de Estatística.

Foram analisados os casos do Estado de São Paulo e da região nordeste do Estado do Pará, Brasil (figura 1). Localizado na zona equatorial, o nordeste paraense apresenta características geográficas importantes que podem ser responsáveis pela diversificação nos regimes de chuva encontrados. Por outro lado, essa área possui uma carência de dados, sendo que foi possível analisar apenas as séries completas de 1983 a 1988, de 22 estações pluviométricas. As variáveis analisadas nesse caso foram os totais mensais de chuva.

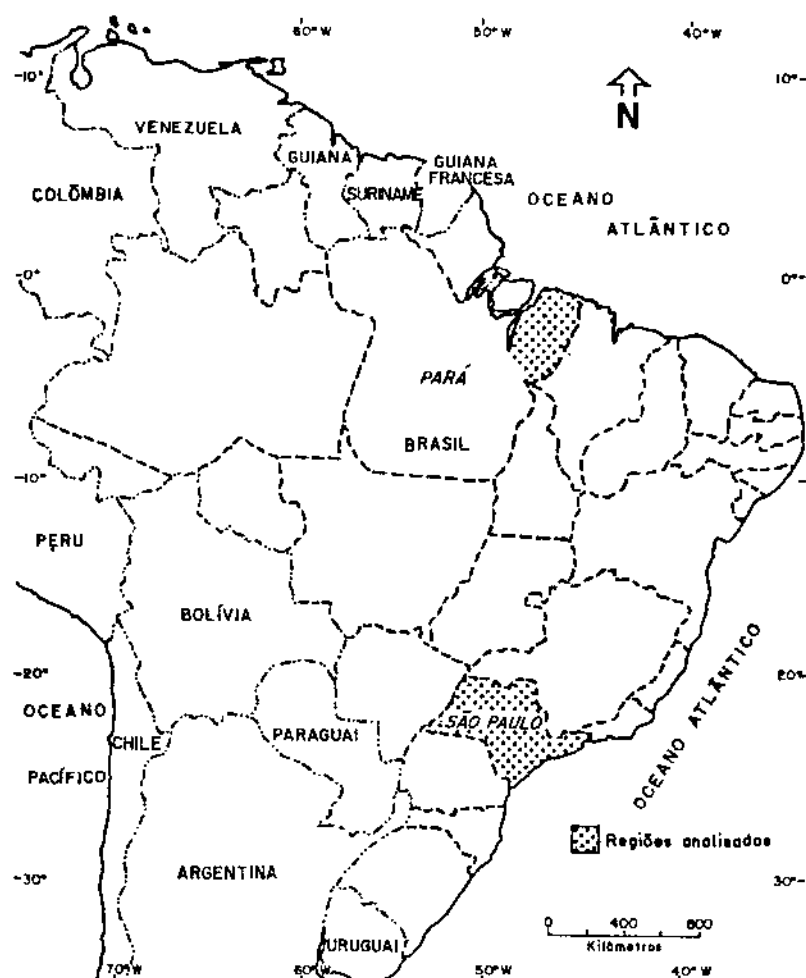


Figura 1 - Localização das regiões estudadas

O Estado de São Paulo, por sua vez, apresenta uma considerável disponibilidade de dados pluviométricos, sendo possível extrair 391 estações contendo séries pluviométricas completas durante o período de 1961 a 1988 para serem analisadas. Foram utilizados os dados referentes apenas ao mês de setembro, época de transição entre os períodos seco e chuvoso no estado, sendo as variáveis estudadas os totais pluviométricos de cada cinco dias (quinqüídios), visando um enfoque para as aplicações agrícolas.

As análises estatísticas foram executadas utilizando o programa SAS - versão 6.06, disponível no VAXcluster da UNICAMP. Os

mapas de isolinhas foram construídos através do programa SURFER - versão 4.08.

Uma revisão bibliográfica comentando sobre os principais textos que utilizam as técnicas estatísticas, particularmente a de componentes principais e a de agrupamentos, para a obtenção de regiões homogêneas será apresentada no capítulo 2. No capítulo 3 será vista a análise para regionalização envolvendo os dados de precipitação do nordeste paraense. A análise para o Estado de São Paulo será apresentada no capítulo 4, onde poderão ser vistas as regionalizações propostas por dois diferentes métodos de agrupamento. Nesse capítulo será vista a importância do conhecimento das características de cada método, que podem levar a agrupamentos muito distintos porém coerentes com as interpretações a serem feitas. As conclusões e sugestões serão vistas no capítulo 5.

Uma breve descrição dos métodos (análises de componentes principais e de agrupamentos) será feita nos apêndices A e B, respectivamente. No apêndice C será apresentada uma descrição das estações pluviométricas utilizadas na análise do nordeste paraense, assim como os perfis de chuva dessas estações. No apêndice D serão listados os códigos das estações meteorológicas localizadas no Estado de São Paulo cujos dados foram analisados no capítulo 4.

Esse trabalho não pretende fazer uma revisão exaustiva a respeito da aplicação dos métodos de análise multivariada no estudo de regionalização pluviométrica nem propor uma regionalização final, mas ilustrar e mostrar as amplas possibilidades de utilização de certas técnicas estatísticas, especialmente análises de componentes principais e de agrupamentos, na análise desse tipo de problema.

CAPÍTULO II

REVISÃO BIBLIOGRÁFICA

Uma das mais importantes ferramentas no estudo do comportamento de fenômenos meteorológicos e variáveis climáticas, a Estatística tem apresentado um grande número de técnicas para a resolução de problemas nessa área de aplicação.

Vários trabalhos podem ser encontrados utilizando técnicas estatísticas visando o conhecimento de características climáticas temporais e espaciais. Por exemplo, aplicações de métodos probabilísticos podem ser encontradas nos textos de Arruda & Pinto (1980), Stern (1982) e Guzman & Torrez (1985), entre outros; Boshell & Neild (1975) utilizam a análise de regressão para encontrar localidades que apresentem características agroclimáticas semelhantes às dos grandes produtores mundiais de chá; Chidley & Keys (1970) e Wei & McGuinness (1973) são alguns dos autores que utilizam as técnicas de geoestatística para a obtenção de estimativas de variáveis climáticas como a precipitação pluviométrica.

Por outro lado, os recursos de séries temporais podem ser vistos em textos como os de Buishand (1985), Asculai & Thieberger (1987) e Karl & Williams (1987). Já Castonguay & Dubé (1985) e Zwiers (1987) valem-se de testes de diferenças entre médias e análise de variância em estudos envolvendo variáveis fenológicas¹ e

¹ fenologia: estudo das relações dos processos biológicos periódicos, como a brotação, a floração e a frutificação, com o clima (Ferreira, 1989, p. 293).

meteorológicas.

Contudo, dado ao caráter das análises climatológicas, que envolvem um número de variáveis bastante grande, vêm sendo amplamente utilizadas atualmente as técnicas de análise multivariada. Dentre os métodos de análise multivariada existentes, destaque especial merecem as análises de componentes principais e agrupamentos, graças ao grande número de trabalhos que têm aplicado tais técnicas. Mais de oitenta aplicações de funções empíricas ortogonais, análise de componentes principais e análise de fatores comuns foram encontradas por Richman (1986) na literatura meteorológica desde 1980. Uma das razões apresentadas por ele para a proliferação do emprego desse tipo de técnica é o variado número de modos em que essas técnicas podem ser utilizadas.

Muitos também são os objetivos com que essas técnicas são aplicadas. Dentre os vários objetivos encontrados podemos citar: agrupar variáveis sinóticas (Christenson & Bryson, 1966); ajudar na previsão de parâmetros atmosféricos (Gilman, 1957); identificar modos coerentes de precipitação pluviométrica (Molteni, Bonelli & Bacci, 1983; Obregon & Nobre, 1990); obter grupos climáticos homogêneos (Gadgil & Joshi, 1983; Kalkstein, Tan & Skindlov, 1987). Além disso, na maioria dos trabalhos a redução dos dados também é colocada como meta a ser alcançada.

Um dos pontos básicos na aplicação das técnicas de componentes principais e análise fatorial, a escolha da matriz de dispersão, tem proporcionado algumas discussões. Alguns autores (Gadgil & Iyengar, 1980; Lyons, 1982) optaram pelo uso da matriz de covariância, que mantém a magnitude real das variâncias. Outros

autores, como Dyer (1975), preferiram utilizar a matriz de correlação, que pondera igualmente todas as variáveis. Resio & Hayden (1975) e Molteni, Bonelli & Bacci (1983) são alguns dos autores que adotaram a matriz de produtos cruzados em suas análises, incorporando tanto as informações da média como da variabilidade de cada variável. O caso de singularidade da matriz de dispersão também pode ser encontrado em Dyer (1975) e Gadgil & Iyengar (1980).

Também na literatura climatológica são propostos alguns critérios ou testes para a escolha do número de autovetores a serem retidos. Craddock & Flood (1969) e Cohen (1983) propuseram testes específicos para dados meteorológicos, que são variantes do teste scree de Cattell (1966). Preisendorfer & Barnett (1977) e Overland & Preisendorfer (1982) sugerem testes baseados nas técnicas de Monte Carlo para essa decisão. Uma comparação entre o comportamento de alguns desses critérios pode ser vista em Hakstian, Rogers & Cattell (1982) e Zwick & Velicer (1982).

Outra questão que tem provocado controvérsias é a aplicação de transformações lineares (rotação) aos autovetores. Richman (1986) recomenda sua utilização, justificando que alguns problemas podem ser reduzidos e até eliminados com a sua aplicação. Willmott (1978), entretanto, argumenta que não há sentido em fazer a rotação de componentes principais uma vez que os componentes não explicam um máximo de variância residual. Dentre os textos que utilizam a rotação de componentes principais podem ser citados os de Dyer (1975) e Molteni, Bonelli & Bacci (1983), enquanto que Paterson, Goodchild & Boyd (1978) e Jongman (1990) são alguns dos que não fazem uso desse procedimento.

Em estudos onde o objetivo consiste na obtenção de grupos semelhantes de observações e/ou determinação de regiões homogêneas, a análise de agrupamentos tem se apresentado como importante procedimento. Em grande parte dos trabalhos que têm esse objetivo pode-se vê-la atuando sobre os componentes principais obtidos a priori (Gadgil & Iyengar, 1980; Guiot, 1986; Jongman, 1990). Todavia, com a disponibilidade cada vez mais crescente dos recursos computacionais, está se tornando possível a aplicação da análise de agrupamentos diretamente às variáveis originais, sem a necessidade da utilização da análise de componentes principais para a redução inicial dos dados. Fernau & Samson (1990a e 1990b) podem ser citados como exemplos de autores que usam essa forma de análise.

No Brasil é recente a utilização das técnicas multivariadas para estudos em climatologia. Salvi (1984) estuda a tipologia climática do Estado de São Paulo através da aplicação da análise fatorial. Garayalde *et al.* (1986) e Justi da Silva, Chida & Manhaes (1988) utilizam a análise de componentes principais para fazer uma análise climática da região sul e Estado do Rio de Janeiro, respectivamente. Aspiazú *et al.* (1990) realiza um teste metodológico para a classificação climática do Estado de Minas Gerais baseado também na análise de componentes principais. Por sua vez Obregon & Nobre (1990) estudam os padrões de precipitação sobre a bacia do rio Amazonas a partir dos componentes principais rotacionados.

CAPÍTULO III

NORDESTE PARAENSE

A região nordeste do Estado do Pará constitui-se em uma das mais antigas áreas de exploração agrícola da região amazônica em solos de terra firme. Embora ocupe cerca de 110.000 km² de área, menos de 9% da superfície total do estado, abriga quase 57% da população estadual, que foi estimada em 4.200.000 habitantes em 1985 (Bastos, 1990, p. 10). Trata-se de uma área de grande importância na produção de diversas culturas agrícolas, visto que concentra cerca de 95% da produção brasileira de pimenta-do-reino, 50% da produção nacional de malva, além da quase totalidade da produção regional de mamão havaí, dendê, algodão e milho, entre as principais (Nascimento & Homma, 1984, p. 57).

Como em qualquer região equatorial, onde as oscilações de temperatura ao longo do ano são de reduzida magnitude, na região amazônica a precipitação pluviométrica é o principal elemento do clima a determinar diferenciações espaciais no potencial de produtividade agrícola e é o elemento meteorológico que mais contribui para as suas flutuações temporais. Torna-se, portanto, de fundamental importância para a agricultura no nordeste paraense o conhecimento dos padrões de chuva atuantes e da variação do seu comportamento no âmbito espacial.

Poucos trabalhos podem ser encontrados apresentando um enfoque de regionalização pluviométrica que englobe a área em consideração. Isso se deve ao fato de que são raros os locais que possuem séries de dados de tamanho suficiente para serem analisados,

assim como é baixa a concentração de estações a dispor de dados em períodos coincidentes (Bastos, 1990, p. 37). Há também que se considerar a questão da representatividade das estações pluviométricas na região amazônica (Molion & Dallarosa, 1990). Todos esses fatores vêm limitando a aplicação de técnicas estatísticas para essa finalidade.

Os métodos de análise descritiva, principalmente os indicadores de tendência central, tornaram-se um dos poucos recursos estatísticos utilizados. Dessa forma, eles têm sido empregados como parte integrante, e às vezes único, dos critérios de classificação climática adotados (Bastos, 1972; Diniz, 1986; Bastos, 1990), ou para a construção de mapas de isoietas² (Brasil.Ministério da Agricultura, 1969; Brasil.SUDAM, 1984).

Recentemente, com a momentânea disponibilidade de dados, torna-se mais atrativa e necessária a aplicação de técnicas estatísticas mais sofisticadas. Assim, já é possível encontrar alguns estudos utilizando métodos de análise multivariada para os dados dessa região, seja para o entendimento dos padrões espaciais de precipitação (Obregon & Nobre, 1990) ou para a comparação entre alguns métodos de classificação pluviométrica (Zullo *et al.*, 1989).

Nesse capítulo será estudado o comportamento da precipitação pluviométrica no nordeste paraense, com a finalidade de entender os padrões de chuva atuantes e determinar áreas homogêneas. Para isso

² isoietas: linha que, em um mapa, liga os pontos onde são iguais as alturas de precipitação pluviométrica recolhidas durante determinado intervalo de tempo (Ferreira, 1989, p.371).

serão utilizadas técnicas de análise multivariada, como a análise de componentes principais e a análise de agrupamentos, que serão avaliadas quanto ao seu desempenho.

3.1 - DADOS E DESCRIÇÃO CLIMÁTICA

3.1.1 - Situação geográfica e climática

O nordeste paraense (figura 2) possui algumas características geográficas muito interessantes que podem influenciar tanto no comportamento quanto nas diferenciações apresentadas pelas chuvas nessa área. Essa região está localizada na zona equatorial, entre os paralelos 0° e 6° de latitude sul. Situada na planície Amazônica, apresenta um relevo em que predominam as áreas com altitude inferior a 200 metros. É banhada ao norte pelo oceano Atlântico, além de ser cortada por vários rios pertencentes à bacia Amazônica, com

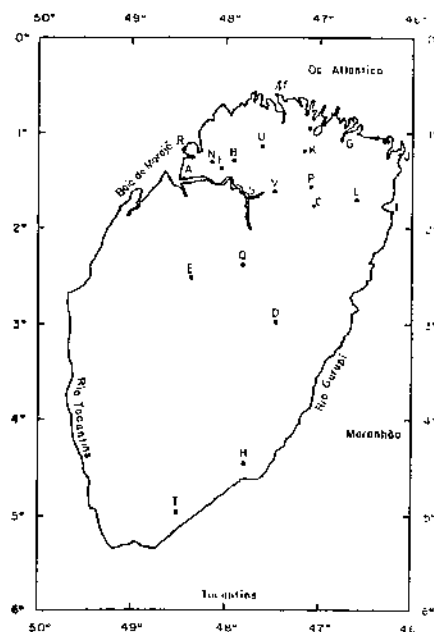


Figura 2 - Mapa situando o nordeste paraense, com as estações utilizadas na análise

destaque para o rio Tocantins, que delimita a porção ocidental dessa área, e à foz do Amazonas a noroeste (Adas, 1985).

A precipitação pluviométrica anual varia entre totais de 2000 mm e 3000 mm. O período mais chuvoso ocorre durante os meses de janeiro a abril, enquanto que o período menos chuvoso abrange desde o trimestre julho-setembro, na parte meridional, até o trimestre setembro-novembro, na parte setentrional (Brasil.Ministério da Agricultura, 1969; Brasil.SUDAM, 1984).

As condições de tempo nessa região são afetadas por um amplo espectro de fenômenos meteorológicos que variam desde a circulação de grande escala até sistemas convectivos de mesoescala, sendo que linhas de instabilidade³, se associadas à linha marítima, podem também explicar a série alternada de máximos e mínimos de totais pluviométricos anuais observados. A interação entre esses fenômenos desempenha papel importante na determinação da precipitação diária, bem como na variabilidade interanual sobre a região (Molion & Kousky, 1986).

O nordeste paraense está situado numa área onde o impacto do fenômeno El Niño - Oscilação Sul é severo, em alguns episódios. Esse fenômeno consiste numa modificação na circulação atmosférica global e oceânica nas bacias dos oceanos Pacífico e Índico, acarretando

³ linhas de instabilidade: formações cumuliformes que se desenvolvem e se deslocam ao longo de pequenas depressões barométricas, geralmente associadas às ondulações frontais, podendo resultar em chuvaradas intensas e localizadas, ocorrendo principalmente no verão (Vianello & Alves, 1991, p. 330).

variações meteorológicas em todo o mundo, entre as quais se incluem a redução de chuvas no nordeste da Amazônia. Os dois últimos episódios desse fenômeno foram registrados em 1982/83 e 1987, com intensidades e conseqüências diferentes. Em 1983 ocorreu um dos episódios mais intensos do século e que atingiu a região, com a redução de mais de 30% das chuvas durante o período de janeiro a maio em algumas estações. Em 1987 o fenômeno foi sentido no Brasil principalmente por uma diminuição das chuvas na região semi-árida do Nordeste em abril e maio (Climanálise, 1991).

3.1.2 - Dados utilizados

Nessa análise foram utilizados dados de precipitação pluviométrica de 22 estações, disponíveis pela EMBRAPA/CPATU (figura 2). Foram analisadas as séries correspondentes ao período de janeiro de 1983 a dezembro de 1988. Os valores diários observados foram sumarizados em totais mensais, resultando na obtenção de 72 variáveis, expressas em milímetros. A posição geográfica e o perfil de chuvas desses postos são exibidos no apêndice C.

Embora sejam disponíveis poucos anos de dados, esse período consegue ser abrangente a respeito do comportamento das chuvas na região, pois possui anos em que ocorrem períodos mais secos que os normais, casos de 1983 e 1987, assim como épocas mais chuvosas que o normal, como 1986.

3.2 - ANÁLISE DESCRITIVA DOS DADOS

A obtenção de algumas estatísticas descritivas é feita,

inicialmente, com o objetivo de auxiliar no entendimento do comportamento temporal e espacial das chuvas no nordeste paraense.

3.2.1 - Características temporais

Os perfis de chuva das estações analisadas são mostrados na figura 3. Para conhecer algumas das principais características temporais das chuvas nessa região, foram calculadas algumas medidas como a média, o desvio-padrão e o coeficiente de variação das variáveis estudadas, mostradas na figura 4.

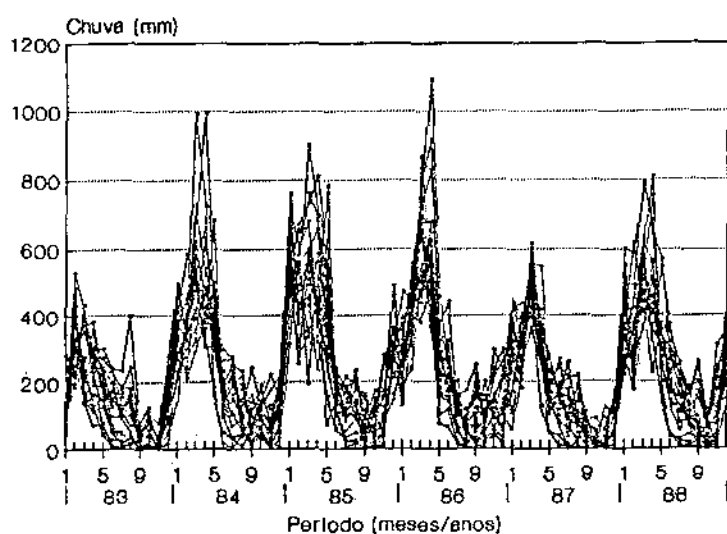
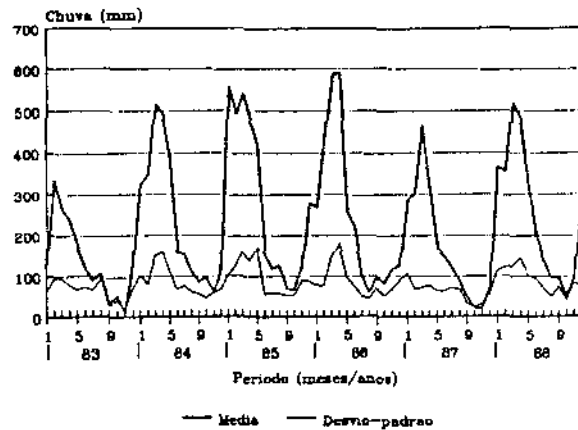


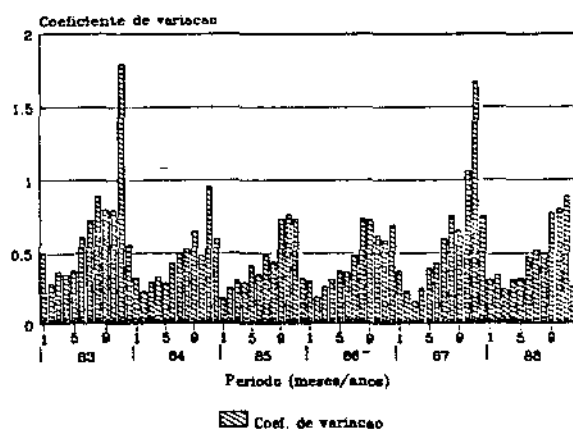
Figura 3 - Perfis de chuva das estações analisadas

Pode-se observar pela figura 4.a que, durante essa época, a ocorrência dos períodos mais chuvoso e menos chuvoso apresentou-se de forma semelhante ao descrito por Brasil.SUDAM (1984). Também se observa que as médias dos meses de 1983 são inferiores quando comparadas às dos mesmos períodos dos demais anos, o que pode ser

explicado pela ocorrência do fenômeno El Niño. Há uma redução nos totais médios de chuva dos meses de 1987, provavelmente devido também ao El Niño, porém de forma não tão acentuada como em 1983.



(a) Média e desvio-padrão



(b) Coeficiente de variação

Figura 4 - Características gerais das variáveis estudadas

O comportamento dos desvios-padrões (ver figura 4.a) ocorre de forma análoga ao das médias. Por outro lado, o que se observa é que o coeficiente de variação também apresenta um padrão anual bem definido, embora com um comportamento oposto ao do desvio-padrão (figura 4.b). Seus maiores valores são observados nos meses mais secos, principalmente nos dos anos de atuação do El Niño, enquanto que os valores mais baixos são registrados nos meses mais úmidos.

3.2.2 - Características espaciais

As figuras 5, 6 e 7 mostram, respectivamente, o comportamento espacial das chuvas médias anuais, chuvas médias no trimestre mais úmido (fevereiro-abril) e chuvas médias no trimestre mais seco (agosto-outubro) no nordeste paraense, durante o período analisado.

As curvas de isolinha foram traçadas utilizando o método kriging de geoestatística. A área localizada entre 46° e 50° de longitude oeste e entre 0° e 6° de latitude sul, na qual se situa o nordeste paraense, foi construída uma malha de pontos com a distância entre eles igual a 30' em cada direção. A estimativa em cada ponto da malha foi obtida a partir das observações das três estações mais próximas, localizadas dentro de um raio de 90'. As linhas de contorno foram interpoladas utilizando um spline cúbico, com fator de tensão de

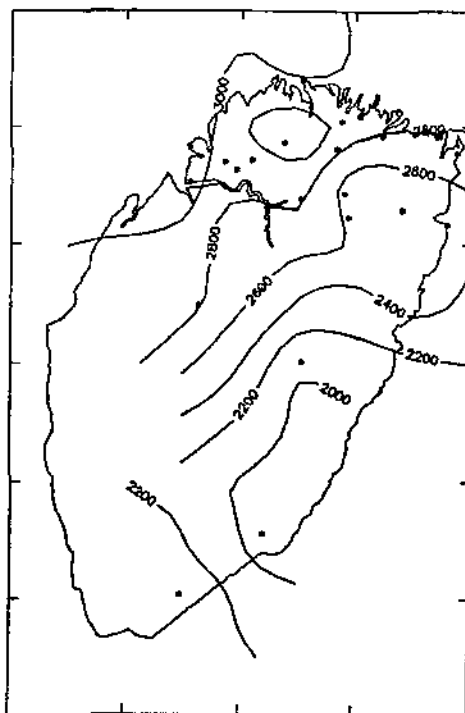


Figura 5 - Total pluviométrico anual médio no nordeste paraense durante o período de 1983 a 1988

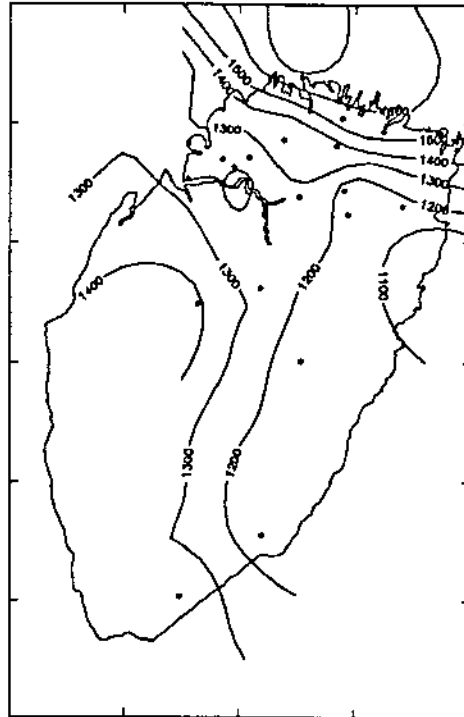


Figura 6 - Total pluviométrico médio do trimestre fevereiro-abril durante o período de 1983 a 1988

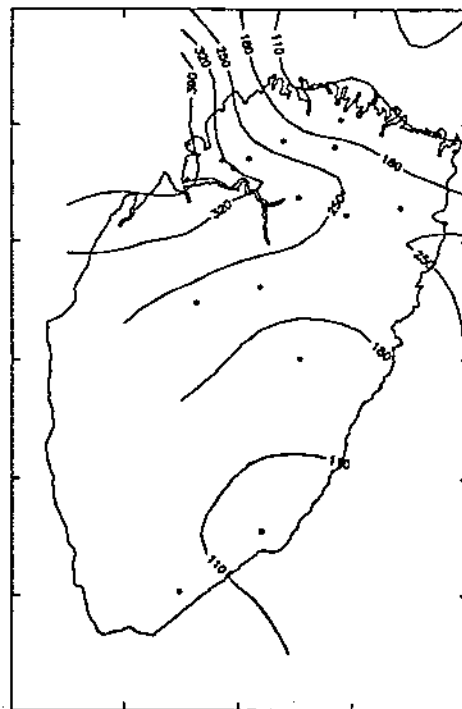


Figura 7 - Total pluviométrico médio do trimestre agosto-outubro durante o período de 1983 a 1988

grau 1. A utilização desses parâmetros deve-se principalmente ao número de estações utilizadas na análise.

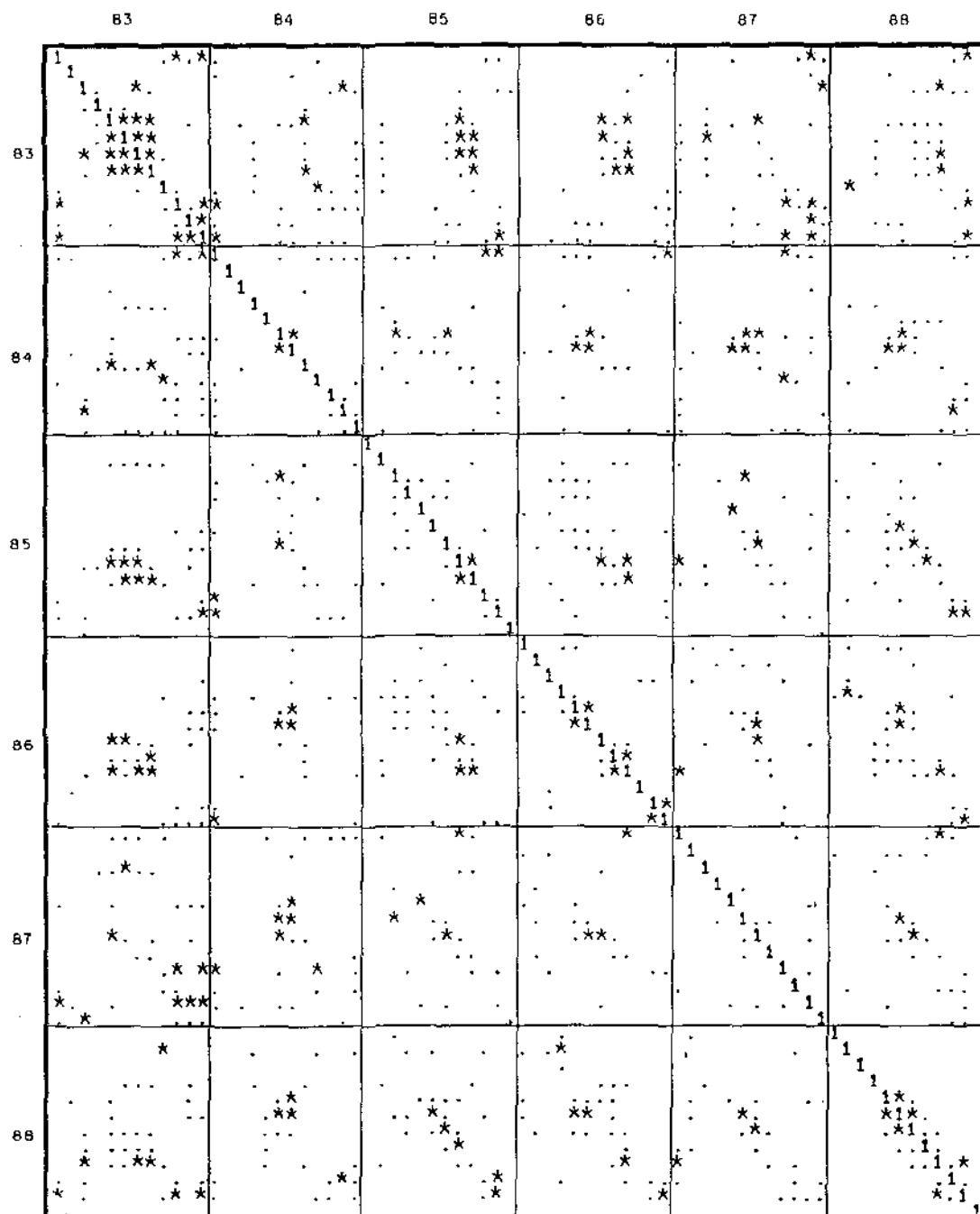
Nos três casos nota-se que as chuvas apresentam um claro comportamento de tendência na direção leste-oeste na parte meridional e na direção norte-sul na parte setentrional. Dessa forma foi encontrado que as regiões litorânea e noroeste foram as que apresentaram a maior quantidade de chuvas anuais e também no período mais chuvoso, com a região sudeste apresentando os menores índices nas duas situações. Já no trimestre mais seco o comportamento da precipitação nessa área modifica-se apenas na região litorânea, que passa a apresentar um perfil mais pronunciado de estiagem.

3.3 - ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais foi aplicada à matriz de correlação das variáveis observadas com o objetivo de conhecer características e padrões de chuva atuantes no nordeste paraense. Um resumo sobre as principais características da análise de componentes principais pode ser encontrado no apêndice A.

Antes de ser realizada essa análise foi construído o quadro 1, a partir da matriz de correlação, com a finalidade de conhecer melhor a sua estrutura. A cada elemento da matriz de correlação foi associado um símbolo, dependendo da grandeza do seu valor absoluto: (*) se dentro do intervalo entre 0.7 e 1.0, exceto os da diagonal principal; (.) se entre 0.5 e 0.7; em branco, caso contrário. Pode-se observar que as correlações mais altas foram encontradas com maior frequência nos meses secos que nos meses chuvosos. Também pode-se

verificar uma alta correlação entre os meses mais próximos de um mesmo ano, assim como entre os mesmos meses de anos diferentes, sempre no período menos chuvoso. Com isso pode-se dizer que as chuvas nessa região apresentam um comportamento de sazonalidade estável,



Quadro 1 - Estrutura da matriz de correlação analisada

considerando-se as baixas correlações observadas nos meses chuvosos como consequência do El Niño.

Com a aplicação da análise de componentes principais, a variabilidade total dos dados é completamente explicada pelos primeiros 21 componentes. Isso se deve ao fato de que, embora a matriz de correlação tenha dimensões 72x72, ela tem posto menor ou igual a 22, face ao número de unidades observacionais. Ver discussão sobre a singularidade da matriz de correlação no apêndice A.

Comp.	Autovalor	Prop.	Prop. Ac
1	18.568	0.258	0.258
2	17.472	0.243	0.501
3	7.255	0.101	0.601
4	4.492	0.062	0.664
5	3.515	0.049	0.713
6	2.926	0.041	0.753
7	2.348	0.033	0.786
8	2.318	0.032	0.818
9	1.898	0.026	0.844
10	1.657	0.023	0.867
11	1.482	0.021	0.888
12	1.433	0.020	0.908
13	1.117	0.016	0.923
14	1.003	0.014	0.937
15	0.919	0.013	0.950
16	0.818	0.011	0.961
17	0.691	0.010	0.971
18	0.633	0.009	0.980
19	0.579	0.008	0.988
20	0.498	0.007	0.995
21	0.378	0.005	1.000
22	0.000	0.000	1.000

Quadro 2 - Proporção da variabilidade total explicada pelos comp.principais

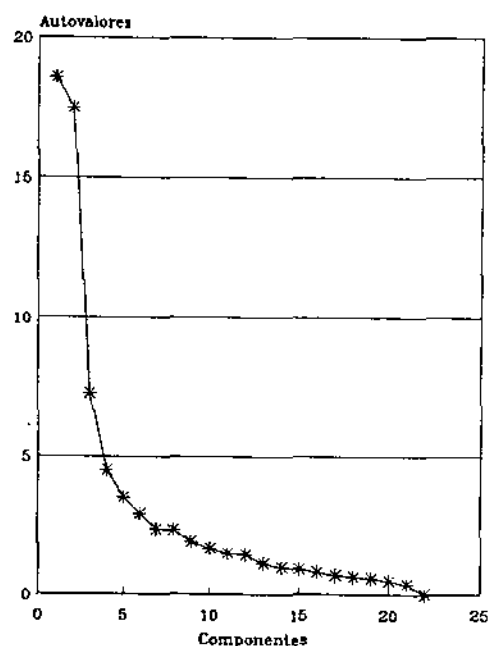


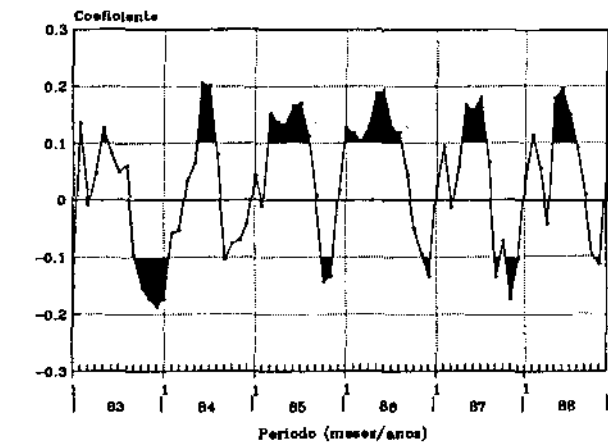
Figura 8 - Diagrama scree

Os autovalores encontrados, assim como a proporção da variabilidade explicada por cada um deles, são mostrados no quadro 2. Através desse quadro pode-se ver que o segundo componente explica

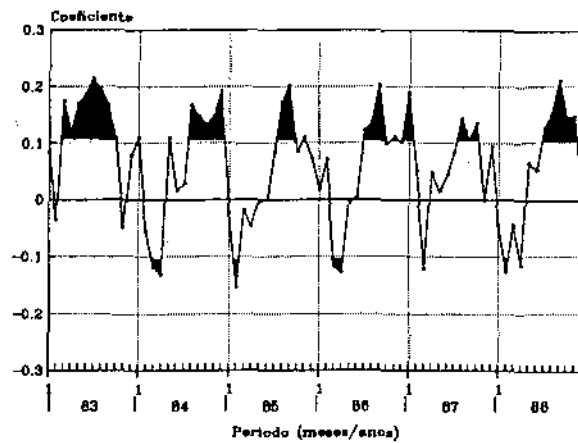
quase que a mesma proporção que o primeiro, e juntos explicam 50% da variabilidade total dos dados. Considerando-se que os seis primeiros componentes principais explicam 75% da variabilidade total, optou-se pela retenção desses componentes para um estudo mais detalhado. O diagrama scree (figura 8), traçado a partir dos autovalores obtidos, também serve como suporte para essa escolha.

Os coeficientes dos seis primeiros componentes principais são exibidos na figura 9. As áreas hachuradas mostram os períodos que são destacados por cada componente. A determinação desses períodos foi feita adotando uma aproximação do critério dos sinais, utilizado por Jolliffe (1986, p. 61-63). Dentro dessa aproximação, para cada componente principal, são destacados os períodos que tiverem coeficientes com valor absoluto maior que a metade do máximo coeficiente (também em valor absoluto) do componente. A dispersão das estações de acordo com os dois primeiros componentes principais é mostrada na figura 10.

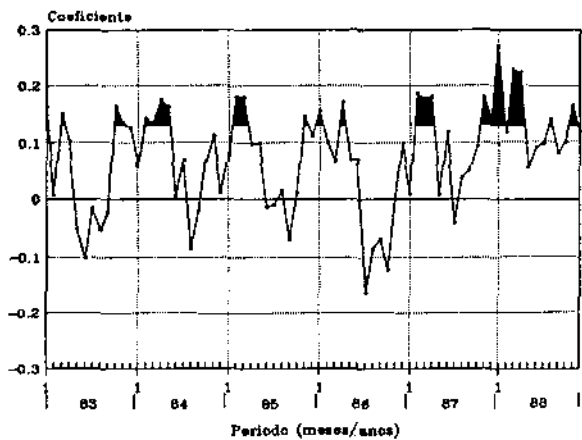
A partir da figura 9.a pode-se observar que os coeficientes do primeiro componente principal apresentam um padrão anual bem definido. Esse componente procura destacar com coeficientes positivos o período correspondente aos meses de maio a agosto de todos os anos, à exceção do ano de 1983, que apresenta pouca influência. Há também um destaque para os meses iniciais do ano de 1986. Os coeficientes negativos são observados durante os meses de setembro a dezembro, sobretudo nos anos de 1983 e 1987. Como, à exceção dos anos de 1983 e 1986, a maioria dos coeficientes que se destacam são positivos, então o primeiro componente principal procura explicar principalmente o comportamento das chuvas na fase de transição chuvoso-seco (maio a



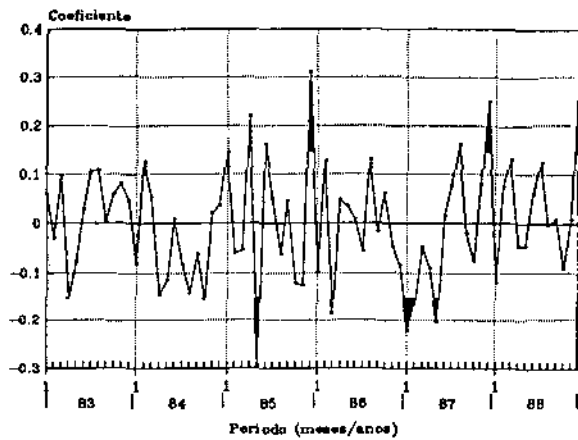
(a) Primeiro componente principal



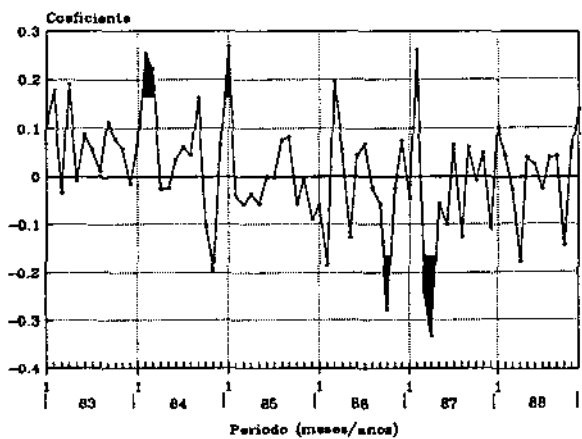
(b) Segundo componente principal



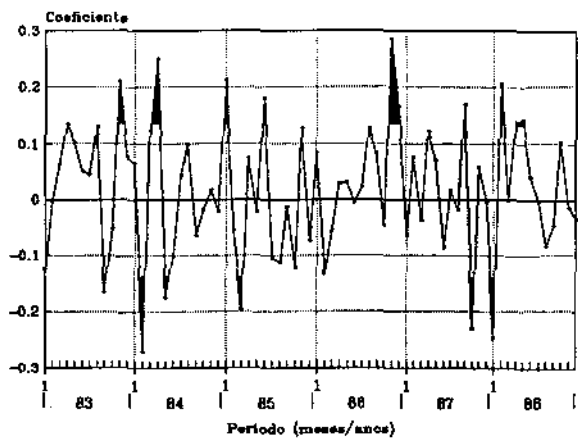
(c) Terceiro componente principal



(d) Quarto componente principal



(e) Quinto componente principal



(f) Sexto componente principal

Figura 9 - Coeficientes dos seis primeiros componentes principais

agosto). A distinção feita aos anos de 1983 e 1987 durante o período de setembro a dezembro deve-se ao fato de que nesses anos foram observados os menores índices pluviométricos das séries de dados.

Pela figura 10 pode-se verificar que as estações que apresentam os maiores valores positivos para o primeiro componente encontram-se na região litorânea (O, G, J e M), enquanto que os que apresentam os valores negativos mais extremos encontram-se na parte meridional (T, D, H, Q, E e F). Observando-se os perfis de chuva dessas estações, no apêndice C, consegue-se notar que as estações localizadas na região litorânea apresentam durante o período de maio a agosto, que foi destacado com coeficientes positivos, totais maiores

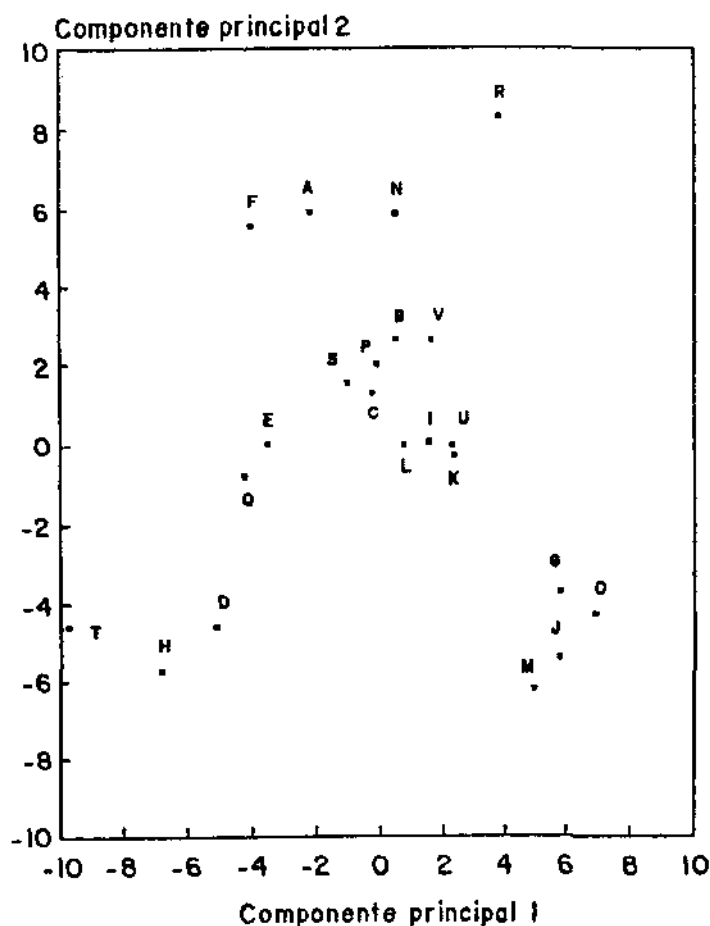


Figura 10 - Dispersão dos postos de acordo com os dois primeiros componentes principais

que os das demais estações. Durante os meses de setembro a dezembro, porém, ocorre uma inversão nessa situação, sendo que as estações localizadas no litoral são as que apresentam ausência de chuvas por períodos mais prolongados, assim como uma defasagem no início do período chuvoso.

O fato das estações situadas na região meridional apresentarem valores negativos pode ser explicado também pelo período entre maio e agosto, pois nessa época em que ocorre a diminuição das chuvas, ela é observada primeiramente nessa área. No período setembro-dezembro, tanto as quantidades quanto a fase de ascensão das chuvas se dão de forma semelhante em todas as estações não-litorâneas. Portanto, os maiores valores desse componente principal são encontrados nas estações em que o período seco inicia mais tarde.

O surgimento de um padrão temporal bem definido também pode ser observado no segundo componente principal (figura 9.b). Nesse caso, os coeficientes positivos foram designados ao período entre os meses de julho a dezembro de todos os anos, assim como aos meses de março a junho de 1983. Os coeficientes negativos foram dados aos meses de fevereiro a abril, à exceção do mesmo período do ano de 1983. Esse componente, então, procura distinguir principalmente o comportamento do período menos chuvoso. O comportamento diferenciado do ano de 1983 pode ser explicado pela ocorrência do El Niño.

As estações contendo os maiores valores positivos encontram-se na região de Belém (R, N, A e F) enquanto que as com valores negativos encontram-se nas áreas meridional (H, D e T) e litorânea (M, J, O e G) do nordeste paraense (figura 10). Pelos perfis de chuvas dessas estações (apêndice C) pode-se encontrar que os postos

localizados na região de Belém apresentam quantidades superiores de chuva em relação aos demais durante os períodos destacados com coeficientes positivos. Por outro lado, o mesmo não ocorre com as estações das regiões meridional e litorânea, que apresentam os menores índices de chuva nessa época do ano. Portanto, os menores valores desse componente principal são observados nas localidades que têm um perfil de seca mais pronunciado.

O terceiro componente (figura 9.c) procura associar coeficientes positivos ao período de novembro a abril, para todos os anos. Todos os demais meses de 1988 também são designados com coeficientes positivos. Por outro lado, apenas os meses de julho a outubro de 1986 são destacados com coeficientes negativos. Uma possível interpretação para esse componente é a de que ele ressalta com coeficientes positivos o período mais chuvoso (janeiro-abril), desde os meses em que se verifica um aumento da precipitação.

As estações destacadas com valores positivos são aquelas que apresentam as maiores quantidades de chuva no período chuvoso, como é o caso das estações localizadas no litoral (M, O, J e G), e de estações que apresentam altos índices pluviométricos em alguns desses meses, como E, T, A, F e Q. Por outro lado, a maioria das estações destacadas com valores negativos encontra-se na região central (I, P, L, K, U, C e S).

A partir do quarto componente principal (figuras 9.d, 9.e, 9.f), os coeficientes não passam a apresentar um comportamento bem definido, como nos casos anteriores, procurando explicar o perfil das chuvas em apenas alguns meses isolados.

De qualquer forma, a aplicação da análise de componentes

principais atingiu seus objetivos visto que as principais características das chuvas no nordeste paraense foram destacadas e diferenciadas quanto à sua atuação espacial e temporal. Essas características são: chuvas no período de transição chuvoso-seco, chuvas no período seco e chuvas no período chuvoso.

3.4 - REGIONALIZAÇÃO PLUVIOMÉTRICA

Uma vez conhecidas as principais características da precipitação pluviométrica no nordeste paraense, torna-se necessário para o processo de regionalização distinguir as estações que apresentam perfis de chuva semelhantes face a essas características. Para essa finalidade foi utilizada a análise de agrupamentos, aplicada sobre as variáveis originais.

Para a determinação das regiões homogêneas foram aplicados três métodos hierárquicos: ligação média, centróide e de Ward. O processo de agrupamento foi realizado de forma divisiva, considerando-se inicialmente todas as estações como pertencentes a um único grupo. Em seguida, esse agrupamento foi sendo particionado até se atingir um estágio em que os perfis de chuva das estações separadas por novos grupos não apresentassem diferenças sensíveis a ponto de justificar essa separação.

Dentre os métodos aplicados o que apresentou melhor performance, agrupando as estações com perfis de chuva mais semelhantes entre si, foi o método de Ward, cujo dendrograma é mostrado na figura 11 (ver discussão no item 3.5.2). A partir desse método foi possível destacar cinco regiões distintas. O perfil médio

de chuvas dessas regiões é mostrado na figura 12, e suas características mais importantes são as seguintes:

- **região oeste:** formada pelas estações E e Q, apresenta totais de 600 mm nos meses chuvosos. O período seco começa no mês de junho e se prolonga até o mês de dezembro, praticamente, porém não há uma ausência completa de chuvas;
- **região de Belém:** apresenta características de chuva parecidas com a da região oeste tanto na época em que ocorrem os períodos secos e chuvosos, como também nos totais pluviométricos dos meses chuvosos. Porém, apresenta os maiores índices pluviométricos de todo o nordeste paraense nos meses secos, com valores em torno de 100 a 200 mm, à exceção dos anos de ocorrência do El Niño. As estações que pertencem a essa área são A, F, B, N e R;

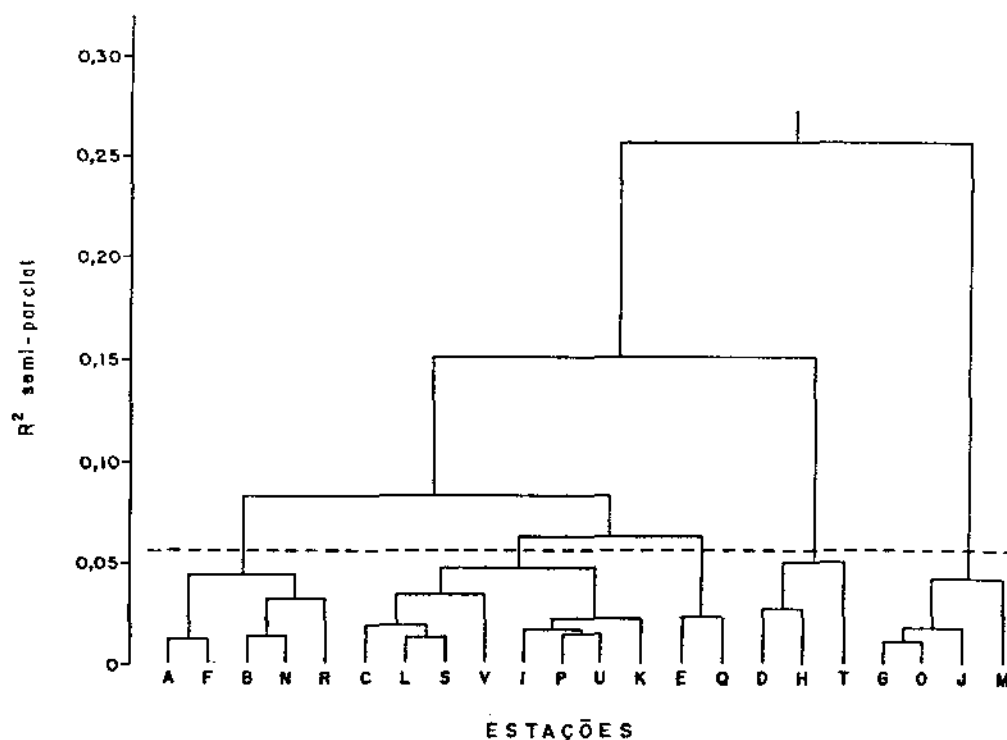
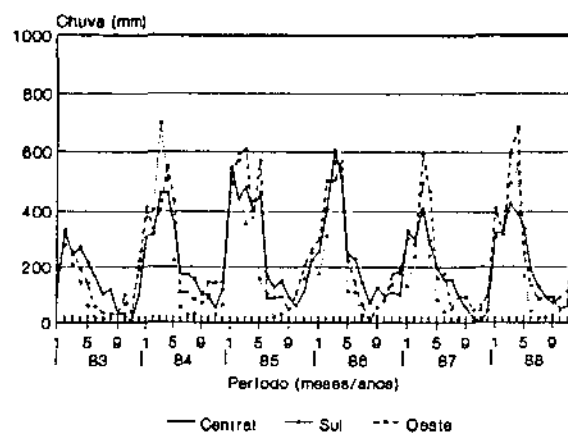
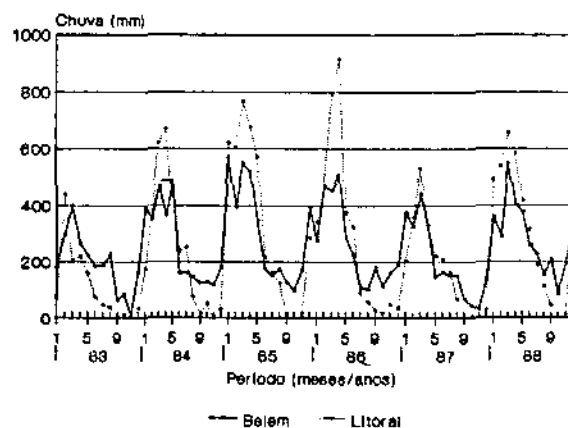


Figura 11 - Dendrograma gerado através do método de Ward

- **região central:** formada pelas estações C, I, K, L, P, S, U e V, apresenta baixas quantidades de chuva na época chuvosa e ausência de precipitação em apenas alguns dos meses do período seco;
- **região sul:** essa região, formada pelas estações T, D e H, possui um perfil de chuvas idêntico ao da região oeste, exceto no período seco, quando se verifica a ausência total de chuvas em vários meses;
- **região litorânea:** constituída pelas estações G, J, O e M. Caracteriza-se por apresentar altas quantidades de precipitação no período chuvoso, podendo atingir até 1000 mm em alguns casos, e ausência de chuvas no período seco. Também apresenta um atraso no começo e término da época chuvosa em relação aos demais locais.



(a) Regiões : sul, central e oeste



(b) Regiões : Belem e litoral

Figura 12 - Perfil de chuva por região homogênea

A distribuição das estações pluviométricas de acordo com a regionalização proposta acima é mostrada na figura 13. Um fato a ser observado é o de que os agrupamentos destacados apresentam uma continuidade espacial, embora essa restrição não tenha sido imposta no algoritmo.

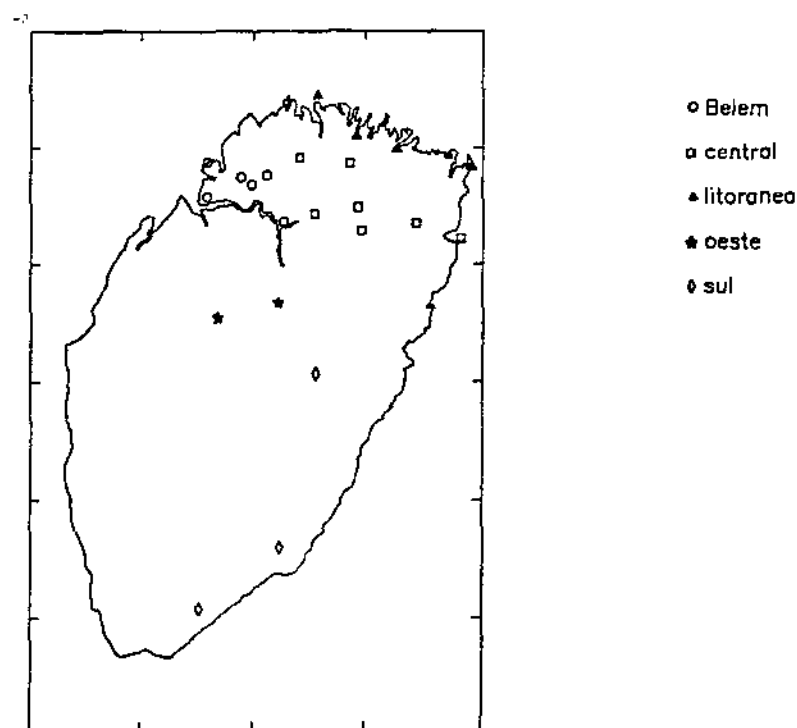


Figura 13 - Regiões pluviométricas homogêneas do nordeste paraense segundo o método de Ward

Tendo sido comentadas as principais características a respeito da regionalização pluviométrica proposta, uma comparação com os resultados obtidos para essa área através do critério de classificação climática de Köppen faz-se importante.

De acordo com o critério de Köppen, que leva em consideração apenas a média da chuva total durante os meses mais chuvoso e mais seco, essa área apresenta três subregiões distintas (Diniz, 1986). Uma das regiões destacadas é a região de Belém (Afi), caracterizada por

apresentar uma alta precipitação anual, com valores mensais superiores a 60 mm mesmo nos meses mais secos. Outra região determinada foi a próxima do oceano Atlântico (Ami), que também apresenta elevadas quantidades anuais de chuva, porém com pelo menos um mês com precipitação abaixo de 60 mm. A última área definida foi a da porção sul (Awi), com uma precipitação anual de moderada a alta, porém com uma estação seca bem marcante.

Comparando as regiões obtidas pelos dois métodos pode-se notar que, embora as características gerais das chuvas sejam semelhantes nos dois casos, a regionalização proposta pela análise de agrupamentos conseguiu um detalhamento maior do comportamento das chuvas. Isso pode ser explicado pelo fato de que o critério climático de Köppen leva em consideração apenas valores de precipitação nos períodos de comportamento mais extremo, sem considerar a época em que esses começam e terminam. O processo de regionalização feito através das técnicas multivariadas, entretanto, trabalha com toda a informação disponível pelos dados, apresentando a capacidade de selecionar aquelas variáveis com maior poder de discriminação.

3.5 - CONSIDERAÇÕES GERAIS

No decorrer da análise surgiram várias alternativas quanto à forma de empregar as técnicas utilizadas, devido à flexibilidade apresentada por essas. As principais razões pela utilização das técnicas da forma como elas foram aplicadas já foram comentadas. Existem outras razões, porém, que podem reforçar a escolha feita e que serão discutidas nesse item.

3.5.1 - Análise de componentes principais

Na análise de componentes principais efetuada optou-se pela utilização da matriz de correlação em relação a outras alternativas possíveis. Como as variáveis utilizadas representam o mesmo fenómeno, a precipitação pluviométrica, e encontram-se nas mesmas unidades, a aplicação da análise à matriz de covariância não deve apresentar consideráveis restrições. Porém, uma circunstância em que não se torna recomendável a utilização da matriz de covariância ocorre quando as variáveis apresentam variabilidade em diferentes escalas, o que parece ocorrer nesse caso, onde se observam desvios-padrões acima de 100 mm em alguns meses chuvosos, e outros bem próximos a zero, em meses secos.

Uma comparação entre os resultados obtidos a partir das duas análises, com a matriz de correlação e com a de covariância, traz à vista outros fatos. Inicialmente, a proporção da variabilidade total explicada pelos seis primeiros componentes principais não apresenta grandes diferenças nos dois casos, exceto para o primeiro e segundo componentes (ver quadro 3). Também se observa que, embora a forma dos componentes principais seja parecida, algumas distorções fazem com que as interpretações sejam diferentes (ver quadro 4). O primeiro componente principal da análise da matriz de covariância, por exemplo, procura realçar o período chuvoso dos anos em que não foi observado o El Niño, destacando fortemente as estações da região litorânea, que apresentam totais de chuva muito superiores aos dos demais postos nessas épocas. Também em outros momentos da análise conseguiu-se detectar a vulnerabilidade desse procedimento frente à diferença de

escala apresentada pelas variáveis, o que reafirma a preferência pela utilização da matriz de correlação.

Comp.	Comp. Principais via		
	Correl.	Cov.	Correl. sem 83
1	0.258	0.285	0.261
2	0.243	0.204	0.222
3	0.101	0.102	0.103
4	0.062	0.081	0.069
5	0.049	0.053	0.052
6	0.041	0.049	0.043

Quadro 3 - Proporção da variabilidade explicada pelos autovetores para diferentes análises de comp. principais.

Comp.	Correlação entre os autovetores	
	Cov.	Correl. sem 83
1	0.6376	0.9943
2	0.6507	0.9959
3	0.6205	0.9558
4	0.8613	0.9878
5	0.8196	0.9864
6	0.7620	0.9534

Quadro 4 - Correlação entre os autovetores obtidos a partir das matrizes de covariância e de correlação sem 83, com os obtidos a partir da matriz de correlação.

Outra análise de interesse consistiu em verificar se a presença de dados de um ano atípico como o de 1983 não estaria distorcendo os resultados. A reedição da análise inicial sem considerar as variáveis correspondentes a esse período pode dar algumas indicações a este respeito. Quanto às proporções explicadas pelos componentes principais, elas são muito parecidas nos dois casos (ver quadro 3). A explicação dada por cada componente apresenta uma grande concordância com os resultados do item 3.3, quando se considera apenas os coeficientes do período de 1984 a 1988 (ver quadro 4), de onde se deduz que no item 3.3 os primeiros componentes principais realmente explicam as características mais gerais das chuvas, mesmo com a presença de dados de um ano atípico como 1983.

3.5.2 - Regionalização pluviométrica

Uma vez que o procedimento de agrupamento das estações homogêneas constitui-se no passo mais importante na determinação das áreas pluviometricamente semelhantes, nesse processo, e dadas as características das técnicas hierárquicas de agrupamento (ver apêndice B), tornou-se necessária a aplicação de vários métodos e a opção pelo que apresentasse os melhores resultados. Para essa avaliação foram levados em consideração os seguintes critérios: comparação entre os perfis de chuva das estações; número de grupos indicados; comparação entre os agrupamentos obtidos a partir das variáveis originais com os obtidos a partir dos componentes principais.

Algumas razões podem justificar a escolha desses critérios. Uma delas é a de que os perfis de chuva das estações podem ser facilmente comparados uma vez que o número de unidades observacionais utilizadas não é grande. Dessa forma procura-se diretamente atender aos objetivos da análise de conglomerados, de agrupar elementos com características semelhantes, separando daqueles com características diferentes. O número de grupos gerados até que os postos dentro de um mesmo grupo não apresentem diferenças consideráveis no seu comportamento também pode auxiliar a favor de uma das técnicas. A comparação feita entre os agrupamentos obtidos a partir dos componentes principais e a partir das variáveis originais teve como interesse verificar se as principais características das chuvas estão sendo representadas pelos agrupamentos.

Essa avaliação foi feita ao método do centróide, ao método de Ward, um dos mais utilizados em classificação climática (Kalkstein,

Tan & Skindlov, 1987), e ao método da ligação média, um dos métodos de melhor performance em estudos de simulação (SAS, 1985). Os resultados obtidos foram os seguintes:

- Ward: as maiores diferenças entre os perfis de chuva são explicadas com a obtenção de cinco grupos. As características destacadas pelos dois primeiros componentes são também consideradas por esse método, estando esses grupos em bom acordo com a dispersão das estações exibida na figura 10. O mesmo não acontece com o terceiro componente. O dendrograma e os grupos resultantes são muito semelhantes quando comparados com a análise utilizando os componentes principais;

- ligação média: com cinco grupos também apresenta diferenciação satisfatória. As diferenças de comportamento explicadas pelas características do primeiro e terceiro componentes são realçadas, porém para o segundo componente isso não ocorre. Apenas três dos cinco grupos permanecem inalterados quando utilizando os componentes principais. Quando se consideram sete grupos os resultados dos métodos de Ward e da ligação média são iguais, porém tal nível de discriminação já foge aos objetivos do trabalho;

- centróide: necessita de mais de cinco grupos para destacar as características distintas dos postos. Com cinco grupos consegue destacar bem as características do primeiro e segundo componentes, mas não as do terceiro. Os resultados obtidos aqui são diferentes dos encontrados com os componentes principais.

Como foram observadas a ocorrência de grupos individuais e a de encadeamento no método do centróide, a adoção dos resultados dessa técnica foi automaticamente descartada. Por outro lado, a deficiência

do método de Ward em construir grupos de tamanhos semelhantes (ver apêndice B) não foi observada aqui, e como esse método apresentou mais vantagens que o da ligação média, então essa foi a técnica considerada de melhor performance. A aplicação de outros métodos de agrupamento, inclusive o de k-médias, não trouxe melhores contribuições. Em todos os métodos utilizados ficou clara a distinção entre o comportamento das estações litorâneas (G, J, O e M) e das localizadas no sul (T, D e H) em relação aos demais postos.

3.6 - CONCLUSÕES

A análise de componentes principais aplicada sobre a matriz de correlação das variáveis utilizadas conseguiu destacar as principais características das chuvas atuantes sobre o nordeste paraense. Cerca de 75% da variabilidade total dos dados ficou concentrada nos seis primeiros componentes principais.

O primeiro componente, que explica aproximadamente 26% da variabilidade total, procura destacar o comportamento das chuvas no período de transição chuvoso-seco. O segundo componente principal procura explicar principalmente o comportamento das chuvas no período menos chuvoso, retendo mais de 24% da variabilidade total dos dados. Já o terceiro componente principal procura destacar o comportamento das chuvas no período mais chuvoso, explicando cerca de 10% da variabilidade total dos dados.

A presença de dados de um ano de comportamento atípico como 1983 não impediu que a análise de componentes principais destacasse as características mais gerais das chuvas, podendo servir para trazer

informações adicionais para o entendimento dessas nessa região.

Posteriormente, a análise de agrupamentos foi aplicada sobre as variáveis originais para a obtenção das áreas pluviometricamente homogêneas. Dentre os vários métodos de agrupamento utilizados o que apresentou melhor performance foi o método de Ward, discriminando cinco agrupamentos. Embora nenhuma informação a respeito tivesse sido utilizada, as estações agrupadas nessas regiões apresentaram uma continuidade espacial. Uma comparação com o critério de classificação climática de Köppen favoreceu a regionalização obtida através dos métodos de análise multivariada, que forneceram resultados mais informativos.

CAPÍTULO IV

ESTADO DE SÃO PAULO

Considerado um dos estados mais desenvolvidos do país, São Paulo possui uma grande variedade de culturas agrícolas assentadas em seu território. Tais culturas, dentre as quais podem-se citar a cana-de-açúcar, milho, algodão, café e soja, servem não apenas para o abastecimento de sua população, estimada em mais de 30 milhões de habitantes, distribuída em cerca de 248.000 km² de área, como também representam importante parcela da produção agrícola nacional (São Paulo.SEADE, 1988).

Em áreas como o Estado de São Paulo, onde o desenvolvimento agrícola demanda um cronograma adequado a cada cultura, o conhecimento da distribuição da precipitação pluviométrica constitui-se em ferramenta fundamental ao seu ordenamento territorial e à maximização da produtividade (Sá *et al.*, 1989).

Vários trabalhos podem ser encontrados buscando conhecer o comportamento espacial das chuvas em São Paulo. Mapas de isoietas anuais, mensais e estacionais foram estudados nos textos de Setzer (1946; 1966), Schroder (1956), Blanco & Godoy (1967), São Paulo.DAEE (1972), São Paulo.CESP/CPFL (1979) e Nascimento & Pereira (1988). Todos esses trabalhos utilizaram dados pertencentes a um considerável número de estações pluviométricas e contendo não menos que 10 anos de informações. Entretanto, além de utilizarem basicamente medidas de tendência central, utilizaram intervalos de tempo maiores, mensais ou anuais, que não são suficientes e nem os mais adequados para tomadas

de decisão na área agrícola.

Por outro lado, os trabalhos que a isso se dispuseram estiveram limitados à utilização de localidades isoladas ou a um pequeno número de localidades. Tais trabalhos valeram-se basicamente da aplicação de métodos probabilísticos (Vivaldi, 1973; Arruda & Pinto, 1980; Camargo *et al.*, 1985 e 1988; Alfonsi *et al.*, 1989a e 1989b; Sansigolo, 1989) ou de séries temporais (Mello & Alfonsi, 1989).

As técnicas de análise multivariada têm sido adotadas alternativamente para a resolução desse tipo de problema. Sua capacidade de trabalhar com um considerável número de variáveis possibilita um estudo mais abrangente envolvendo tempos de resposta condizentes às necessidades agrícolas. Por sua vez, a evolução dos equipamentos e recursos computacionais também tem contribuído para a utilização de um maior número de pontos observacionais em análises estatísticas, permitindo uma avaliação mais adequada da distribuição espacial de variáveis climáticas, tais como a precipitação pluviométrica. Para o Estado de São Paulo pode-se encontrar alguns trabalhos aplicando esse tipo de técnica, envolvendo estudos sobre a tipologia climática do estado (Salvi, 1984), ou fazendo uma avaliação das chuvas em alguns períodos (Tarifa, 1974; Sá *et al.*, 1989; Zullo *et al.*, 1991).

O objetivo desse capítulo é estudar o comportamento espacial das chuvas no Estado de São Paulo, sobretudo através da aplicação das técnicas de componentes principais e análise de agrupamentos. Será avaliado o comportamento das chuvas durante o mês de setembro, época de transição entre a estação seca e chuvosa no estado (Climanálise,

1989). Também será proposta uma delimitação de regiões homogêneas em São Paulo em relação ao comportamento pluviométrico no referido mês.

É prevista a realização de análises semelhantes para os demais meses do ano para o Estado de São Paulo. Dentre alguns motivos para se fazer uma análise mês a mês pode-se citar o interesse em analisar separadamente as chuvas de acordo com os seus mecanismos geradores e a viabilização da utilização dos resultados obtidos considerando-se os procedimentos agrícolas a serem executados.

4.1 - DADOS E DESCRIÇÃO CLIMÁTICA

4.1.1 - Situação geográfica e climática

O Estado de São Paulo apresenta características geográficas que, dadas a sua diversificação, tornam-se importantes fatores responsáveis pela variação do comportamento da precipitação pluviométrica. Localizada na zona tropical, cortada pelo trópico de Capricórnio, essa região é banhada a sudeste pelo oceano Atlântico. Possui um relevo bastante diversificado, abrangendo desde áreas ao nível do mar até áreas superiores a 2700 m de altitude, predominando nessa região as áreas dos planaltos Meridional e Atlântico (ver figura 14). Destaque especial deve ser dado à serra do Mar, serra da Mantiqueira e vale do Paraíba (Adas, 1985).

As chuvas totais anuais variam entre níveis acima de 3000 mm, nas regiões serranas, até cerca de 1100 mm nas áreas baixas do planalto. Da mesma forma, a variação sazonal é caracterizada por verões com totais pluviométricos que vão desde os 400 mm, no mês de janeiro, até invernos com zero milímetro, no mês mais seco (Setzer,

1966, p. 16; Pinto, Zullo Jr. & Zullo, 1989). O trimestre mais chuvoso se verifica no período de dezembro a fevereiro em todo o estado, à exceção do litoral, no qual ocorre de janeiro a março. Já o trimestre mais seco acontece de junho a agosto no litoral e grande parte do interior, e de maio a julho em algumas poucas áreas do estado (Brasil.Ministério da Agricultura, 1969).

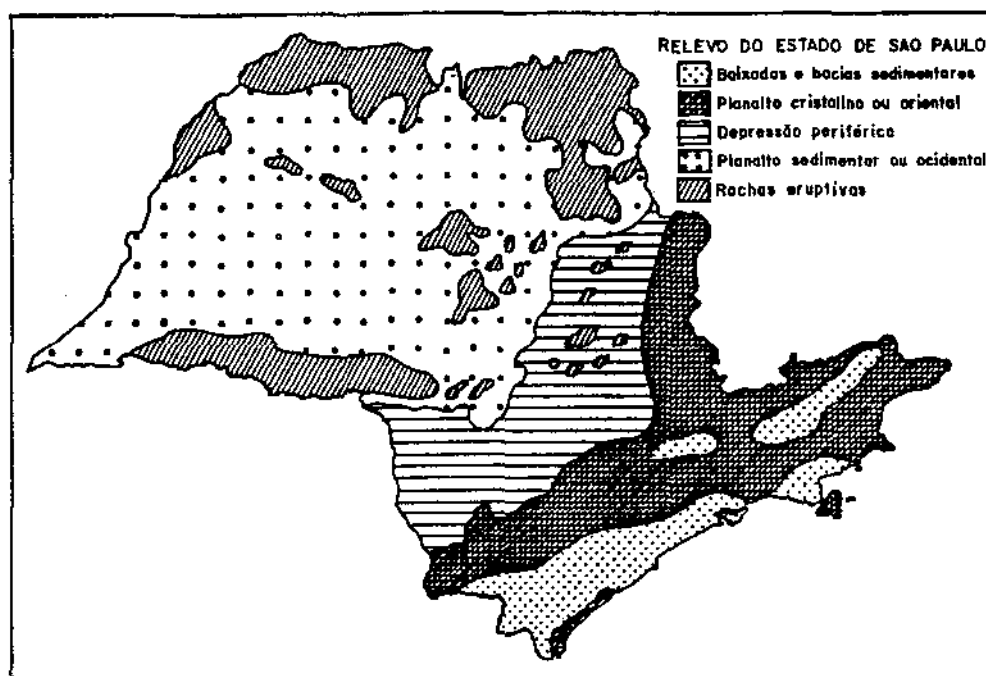


Figura 14 - Relevo do Estado de São Paulo

Os sistemas frontais predominam no inverno, com poucas chuvas e geadas freqüentes, enquanto que no verão as temperaturas elevadas e chuvas abundantes encontram-se associadas principalmente ao aquecimento superficial e às linhas de instabilidade. O relevo é um importante fator na distribuição das temperaturas e das precipitações, intensificando as frentes e linhas de instabilidade. Também atuam sobre a região sistemas atmosféricos de grande escala como o Anticiclone do Atlântico Sul, Correntes de Jato, etc (Vianello, 1991,

p. 433-434).

4.1.2 - Dados utilizados

Os dados utilizados nessa análise são referentes à precipitação pluviométrica de 391 estações, localizadas no Estado de São Paulo (ver figura 15), pertencentes à rede de postos do Departamento de Águas e Energia Elétrica do Estado de São Paulo, e disponíveis no banco de dados do CEPAGRI/UNICAMP (1988). Os códigos dessas estações encontram-se listados no apêndice D.

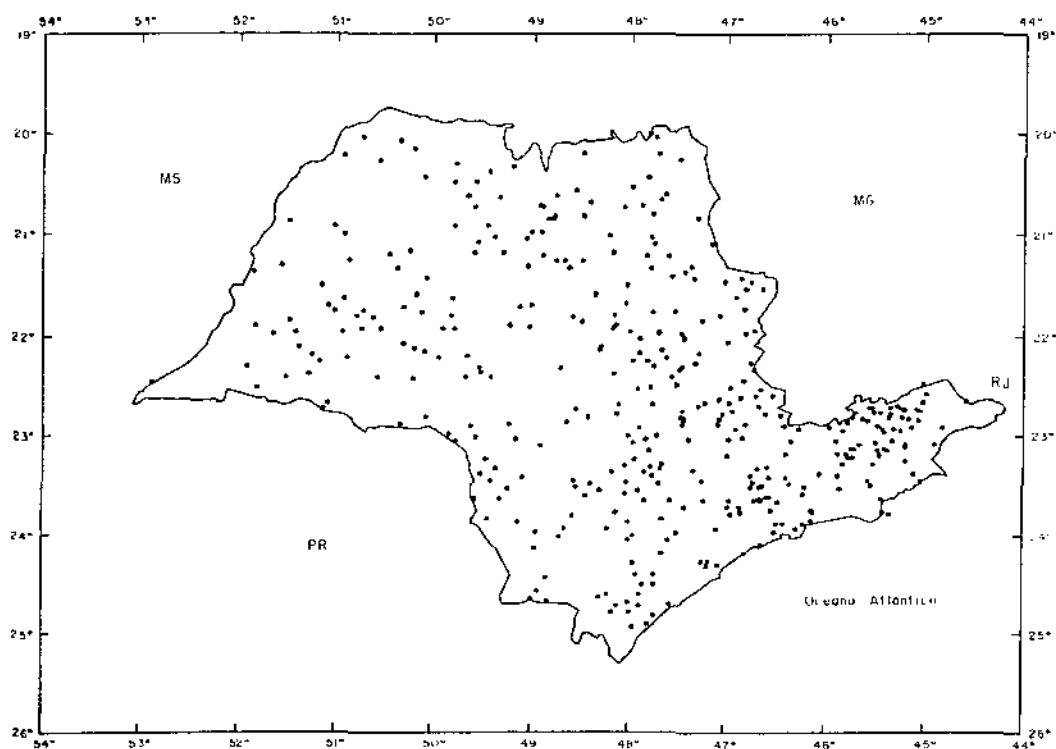


Figura 15 - Mapa situando o Estado de São Paulo, com as estações utilizadas na análise

Os valores diários observados durante o mês de setembro foram totalizados para períodos de cinco dias (quinqüídios). Foram considerados apenas os dados registrados de 1961 a 1988, num total de 28 anos. Como existem 6 quinqüídios por mês, tem-se então um total de

168 variáveis, todas expressas em milímetros.

O estudo do comportamento das chuvas em períodos de cinco dias procura fazer com que os seus resultados possam ser interpretados levando-se em consideração mais diretamente as necessidades hídricas e fisiológicas das plantas.

4.2 - ANÁLISE DESCRITIVA DOS DADOS

Da mesma forma que no capítulo anterior, estatísticas descritivas serão obtidas com o objetivo de procurar conhecer algumas características do comportamento temporal e espacial das chuvas em São Paulo, durante o período analisado. Espera-se, também, que elas possam contribuir de alguma forma para uma melhor compreensão das análises seguintes.

4.2.1 - Características temporais

Embora os valores considerados possam não refletir a média do estado, já que os postos não estão espalhados uniformemente, eles dão uma boa noção da quantidade de chuvas no Estado de São Paulo.

Assim, inicialmente procurou-se estudar o comportamento das chuvas no mês de setembro em relação ao comportamento nos meses precedente (agosto) e posterior (outubro). Para isso foi efetuado o cálculo das médias das chuvas nos quinqüídios correspondentes, utilizando os dados das 391 estações pluviométricas durante o período de 1961 a 1988. O gráfico exibido na figura 16, obtido a partir desses valores, mostra durante esse período uma tendência linear crescente das chuvas. As chuvas médias observadas nos quinqüídios de agosto

foram inferiores a 10 mm, enquanto que as observadas nos qüinqüídios de outubro foram superiores a 18 mm.

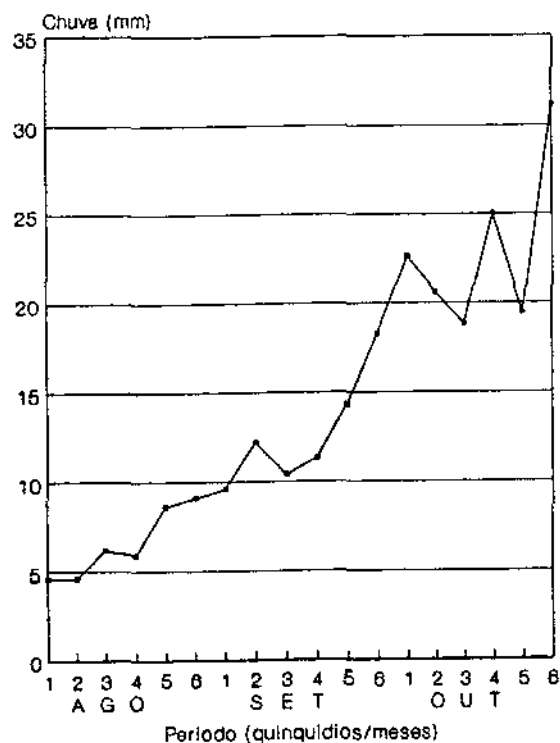


Figura 16 - Média das chuvas nos qüinqüídios de agosto a outubro durante o período analisado

A figura 17 mostra as médias das 168 variáveis estudadas, que correspondem aos totais de chuva dos qüinqüídios do mês de setembro. A partir do quadro 5 pode-se observar que o 5º e 6º qüinqüídios apresentam uma freqüência maior de casos com precipitação acima de 20 mm por qüinqüídio, em comparação com os demais. Por outro lado os quatro primeiros qüinqüídios apresentam uma freqüência em torno de 50% de casos com precipitação abaixo de 5 mm por período. Nos quatro primeiros qüinqüídios de setembro foram encontradas mais de 40% das observações com ausência de precipitação. No 5º e 6º qüinqüídios, porém, essas freqüências ficaram abaixo de 25%.

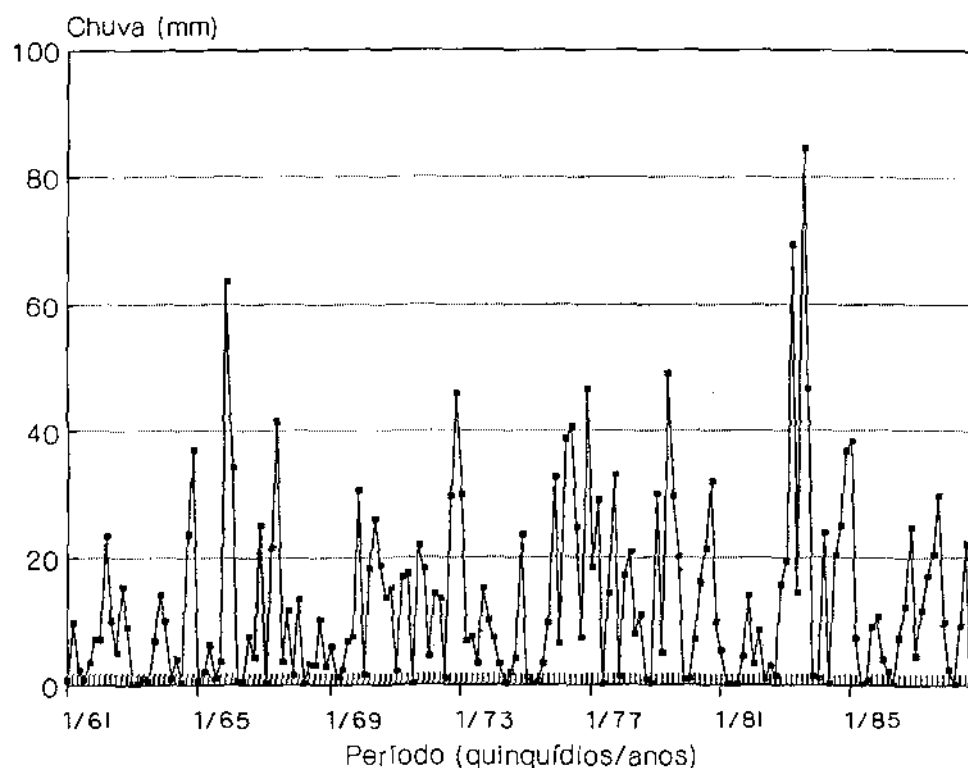


Figura 17 - Médias das variáveis estudadas

Frequência (%)	Quinquênio					
	1	2	3	4	5	6
iguais a 0 mm	43	41	46	43	25	22
abaixo de 5 mm	50	43	50	54	29	25
acima de 20 mm	14	25	14	25	32	36

Quadro 5 - Frequência relativa de chuvas iguais a 0 mm, inferiores a 5mm (inclusive 0mm) e superiores a 20 mm, por quinquênio

A figura 18 mostra os diagramas box-plot para os quinquênios de setembro, construídos a partir da transformação logaritmo natural das observações não nulas. À exceção apenas do 1º quinquênio, não se observa uma diferença acentuada entre o comportamento dos demais períodos. Assim, auxiliados pelos dados do quadro 5, pode-se

considerar que as maiores diferenças entre o comportamento das chuvas em cada quinqüídio estão nas freqüências em que essas ocorrem, do que propriamente na quantidade que apresentam.

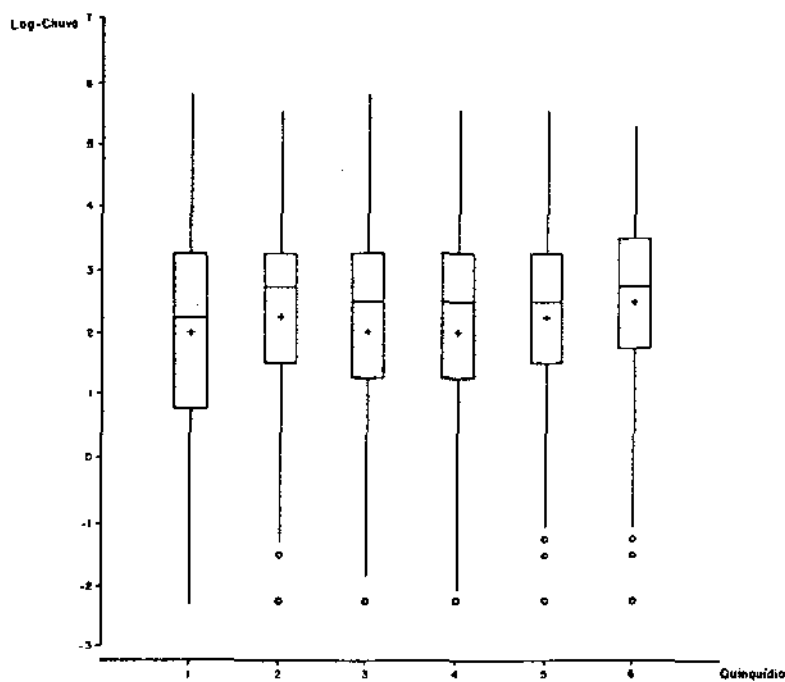


Figura 18 - Diagrama box-plot após transformação logarítmica para os quinqüídios de setembro (apenas com ocorrência de chuva)

Como há um considerável número de anos sendo analisado, procurou-se também verificar se houve mudanças no comportamento das chuvas no decorrer do tempo. Os valores observados, sempre no mês de setembro, foram separados em dois períodos (1961/74 e 1975/88), e calculadas as médias das chuvas por quinqüídio em cada período. A figura 19 mostra que os valores médios encontrados durante o período de 1975/88 foram superiores aos do período 1961/74, à exceção apenas do 6º quinqüídio. As maiores diferenças ocorrem no 2º, 4º e 6º quinqüídios. Durante o período de 1961 a 1974 o total médio de chuvas no mês de setembro foi de 64 mm, enquanto que de 1975 a 1988 esse valor foi de 88 mm.

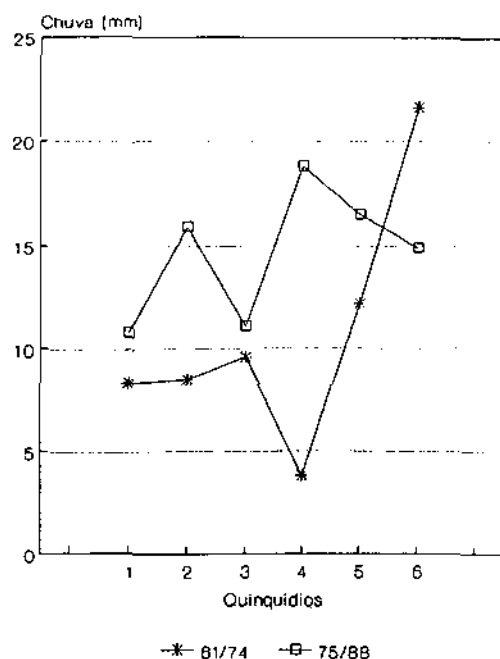


Figura 19 - Média das chuvas por quinquídio, em setembro, durante os períodos de 1961/74 e 1975/88

4.2.2 - Características espaciais

Para estudar a distribuição espacial das chuvas em São Paulo durante o mês de setembro foram mapeados seus valores médios mensais e por quinquídio, mostrados nas figuras 20 e 21 respectivamente.

A construção das curvas de isolinha deu-se através do método kriging de geoestatística. A malha de pontos foi construída dentro da área delimitada pelas coordenadas 19° e 26° de latitude sul e 44° e 54° de longitude oeste, onde está situado o Estado de São Paulo. A distância entre os pontos da malha foi de 30' em cada direção. Para a estimativa em cada ponto da malha foram considerados os cinco pontos mais próximos, localizados dentro de um raio de 30', sendo que as linhas de contorno foram interpoladas posteriormente mediante um

spline cúbico, com fator de tensão de grau 1.

A distribuição espacial das chuvas médias mensais em setembro (ver figura 20) apresenta um gradiente na direção latitudinal, com as menores médias sendo registradas na região norte do estado, e os maiores valores sendo observados na parte sul, sobretudo na faixa litorânea. Apenas a região do vale do Paraíba possui um comportamento destoante desse padrão, apresentando uma menor quantidade de chuvas em comparação com as regiões vizinhas.

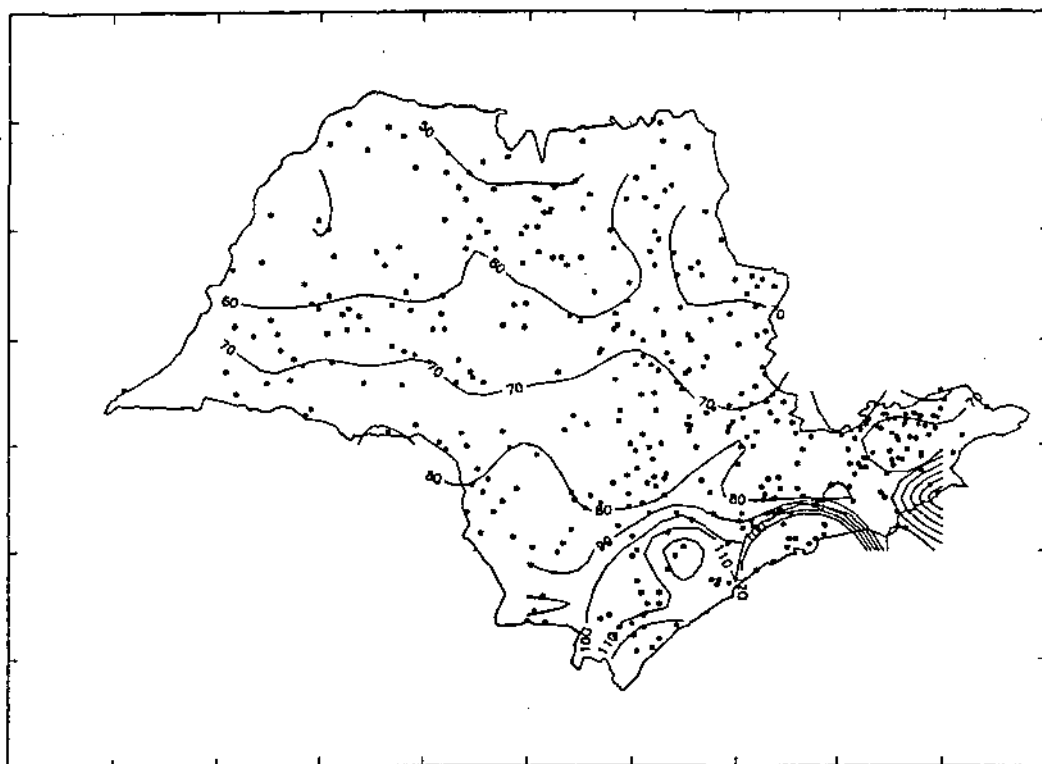
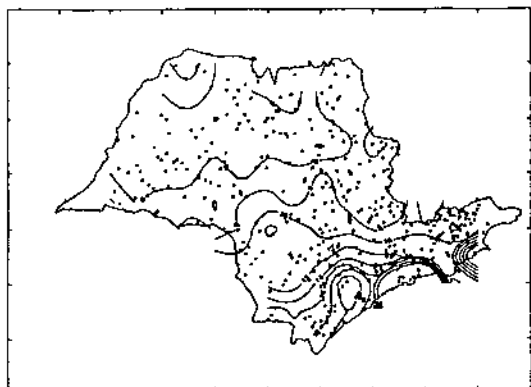
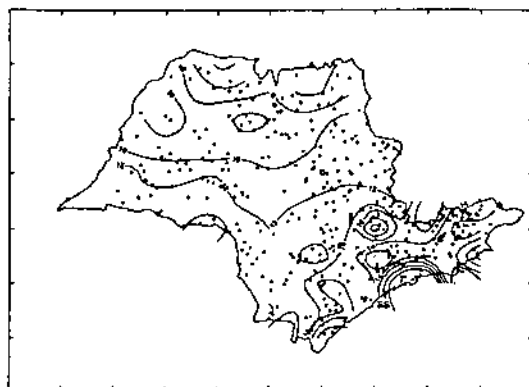


Figura 20 - Precipitação pluviométrica média em setembro durante o período de 1961 a 1988

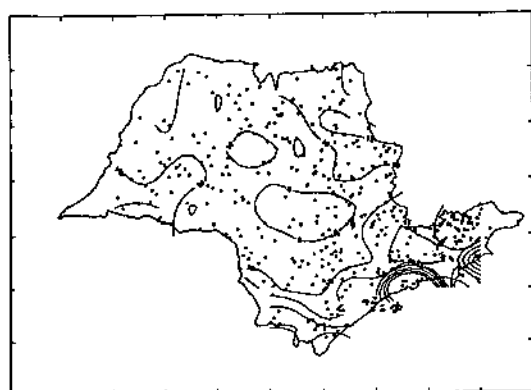
Quanto à distribuição espacial das chuvas por quinquídio (ver figuras 21.a a 21.f), elas apresentam de uma maneira geral padrões semelhantes ao da distribuição espacial mensal, predominando um gradiente de variação na direção latitudinal.



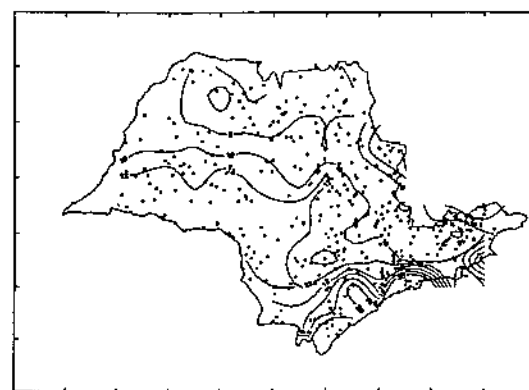
(a) Primeiro quinquídio



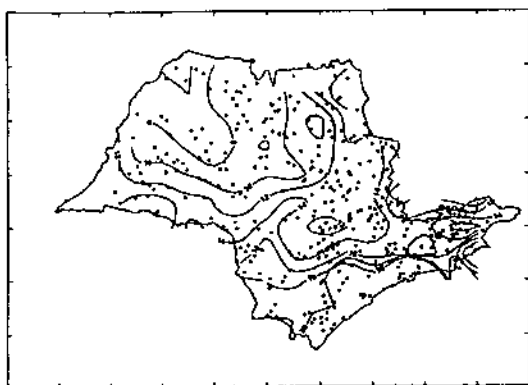
(b) Segundo quinquídio



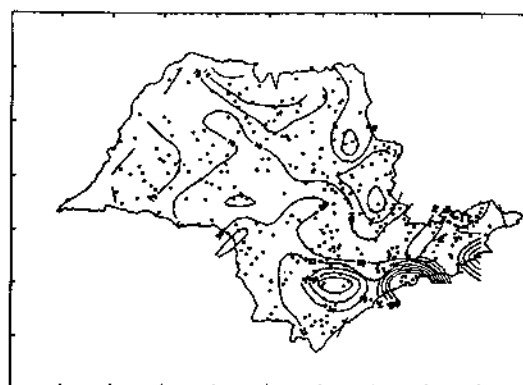
(c) Terceiro quinquídio



(d) Quarto quinquídio



(e) Quinto quinquídio



(f) Sexto quinquídio

Figura 21 - Precipitação pluviométrica média em setembro durante o período de 1961 a 1988, por quinquídio

Esses mapas, principalmente o das chuvas médias mensais, mostram o comportamento das chuvas altamente correlacionado com a variação do relevo no estado (ver figura 14), confirmando-se como um importante fator na sua distribuição espacial.

4.2.3 - Consistência dos dados

Antes da execução das análises de componentes principais e de agrupamentos, os dados foram submetidos a um teste para a detecção de possíveis outliers, visto que principalmente os métodos de agrupamento são sensíveis à presença desses (Johnson & Wichern, 1982, p. 554).

Foi utilizado um teste empírico e univariado, sem rigor estatístico, para classificar como outlier apenas as observações que fossem claramente aberrantes. Uma análise mais acurada exigiria a utilização de técnicas multivariadas para detecção de outliers.

Esse teste consistiu na identificação das estações que apresentaram pelo menos um quinqüídio no mês de setembro com precipitação superior a 150 mm. Para cada uma dessas estações, 27 num total de 391, foram realizadas análises descritivas dos totais de chuva de cada quinqüídio, e anotados os valores da diferença entre o 1º e 3º quartis.

Em seguida, as observações foram ordenadas de forma decrescente de acordo com a quantidade de chuvas apresentada. Caso a diferença entre a primeira e a segunda maiores observações fosse maior que duas vezes a diferença entre quartis então a primeira observação seria descartada. Tal procedimento se repetiu para as observações seguintes, apenas para as superiores a 150 mm, até que a diferença

entre elas fosse menor que duas vezes a diferença entre quartis. As observações descartadas foram então substituídas pelo valor da maior observação restante.

Como exemplo vamos analisar o caso da estação E3-042R, localizada em Santos. A diferença entre quartis apresentada por essa estação foi de 79.3 mm, sendo as cinco maiores precipitações registradas nos quinqüídios de setembro, durante os anos estudados, iguais a 553.7, 265.1, 243.0, 224.2 e 224.0 mm. A diferença entre a primeira e a segunda maiores observações foi de 288.6 mm, portanto maior que o dobro da diferença entre quartis, sendo a primeira observação descartada. A diferença entre a segunda e terceira observações foi de 22.1 mm, sendo então mantida a segunda observação. No lugar da observação que foi descartada (553.7 mm) foi utilizado o valor da segunda observação (265.1 mm), que passou a ser o maior valor dentre os considerados.

Essa metodologia procurou destacar tanto as observações que estão mais sujeitas a erros no processo de coleta e depuração dos dados como aquelas que, mesmo corretas, possam estar associadas à ocorrência de fenômenos pouco freqüentes nessa época do ano. Esse é o caso, por exemplo, dos 553.7 mm de chuva registrados na estação E3-042R, no período de 21 a 25 de setembro de 1972, enquanto que a segunda maior quantidade de chuvas registrada foi de 265.1 mm. Segundo o boletim pluviométrico de São Paulo.DAEE (1974, p. 491), nessa estação foram registrados 379.5 mm de chuva apenas no dia 22.

Dessas 27 estações analisadas mais atentamente, 15 tiveram suas observações alteradas conforme o procedimento anterior. Dentre essas 15 estações, duas tiveram três observações modificadas e outras

duas tiveram duas observações modificadas, enquanto que as demais estações sofreram apenas uma alteração nos dados. Essas 21 alterações processadas foram realizadas em 18 quinqüídios diferentes, sendo que em apenas três desses quinqüídios foi feita a alteração de duas observações no mesmo quinqüídio. Isso contribui para a hipótese de que as observações alteradas possam ser realmente devido à ocorrência de fenômenos observados isoladamente ou à manipulação incorreta dos dados.

Assim, procurou-se evitar que a presença de dados extremos influísse na performance dos métodos de agrupamento, visto que há interesse num processo de regionalização que agrupe as estações de acordo com suas características mais gerais. Não é esperado que tais modificações influenciem significativamente os resultados das análises posteriores de forma a alterar o sentido das conclusões a serem obtidas. Análises posteriores confirmaram que as mudanças realizadas tiveram influências desprezíveis.

4.3 - ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais foi aplicada à matriz de correlação das variáveis utilizadas com o objetivo de destacar os padrões de chuva atuantes em São Paulo durante o mês de setembro, responsáveis pelo comportamento diferenciado da precipitação pluviométrica nas diversas estações meteorológicas.

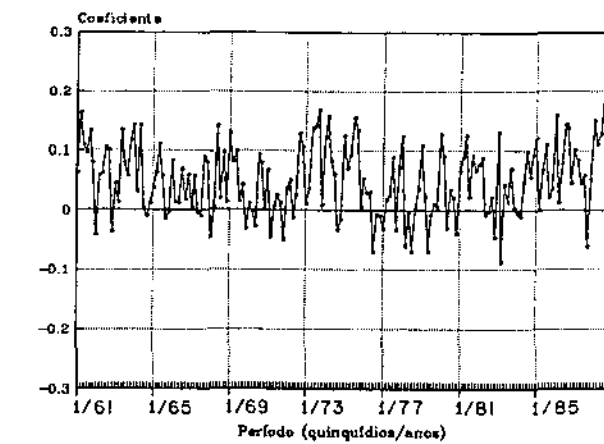
A matriz de correlação entre as variáveis foi obtida a partir da matriz de dados observacionais (391x168), com as devidas alterações comentadas no item 4.2.3. A opção por utilizar essa matriz,

autovetores a cada uma das variáveis utilizadas para os três primeiros componentes principais. O primeiro componente principal associa coeficientes positivos à grande maioria das variáveis, buscando destacar o comportamento geral das chuvas durante o mês de setembro. Essa interpretação também se confirma quando os valores do primeiro componente principal para cada estação são mapeados (ver figura 24.a), com a obtenção de padrões semelhantes ao das chuvas médias nesse mês, exibidos na figura 20.

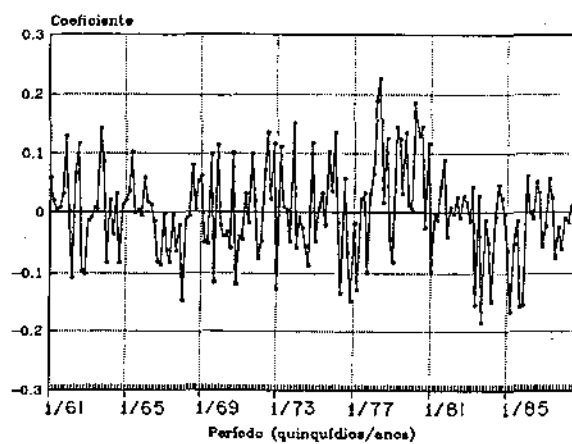
Para o segundo e terceiro componentes principais há uma certa dificuldade em se fazer uma interpretação baseado nos coeficientes obtidos através dos autovetores (figuras 23.b e 23.c), não sendo possível encontrar algum período que se destacasse dos demais. Entretanto, quando os valores dos componentes principais são mapeados (figuras 24.b e 24.c), consegue-se observar que esses valores apresentam um comportamento bem definido.

No caso do segundo componente principal, esses valores apresentam uma tendência na direção sudoeste-nordeste do estado, com os maiores valores positivos sendo observados na parte sudoeste e os valores negativos na parte nordeste. Uma possível interpretação para esse componente pode estar associada à atuação de frentes polares no Estado de São Paulo.

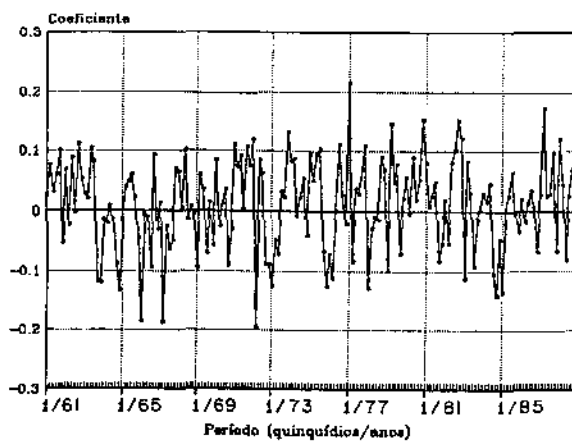
Para o terceiro componente principal observa-se uma tendência na direção sudeste-noroeste, com os valores positivos sendo observados nas porções norte e oeste do estado, e os valores negativos sendo registrados sobretudo na parte sul. No entanto não foi encontrada uma interpretação climática para explicar esse comportamento.



(a) Primeiro componente principal

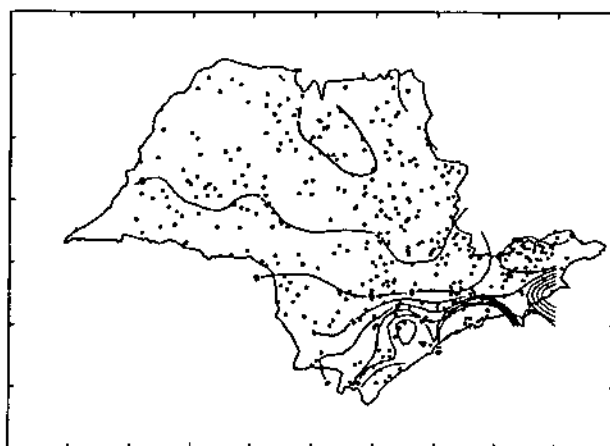


(b) Segundo componente principal

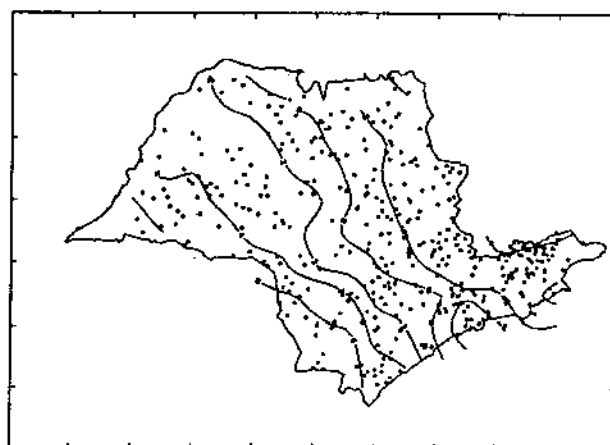


(c) Terceiro componente principal

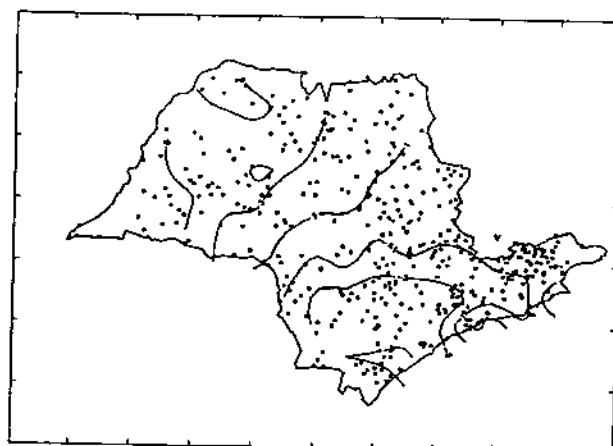
Figura 23 - Coeficientes dos três primeiros componentes principais



(a) Primeiro componente principal



(b) Segundo componente principal



(c) Terceiro componente principal

Figura 24 - Valores dos três primeiros componentes principais

Para os demais componentes não se conseguiu uma interpretação clara, seja através dos coeficientes de cada variável ou através do mapeamento dos valores de cada estação.

A análise de componentes principais executada utilizando as variáveis originais (sem as modificações executadas no item 4.2.3) apresentou resultados semelhantes tanto na proporção explicada pelos primeiros autovalores como na explicação dada pelos autovetores a esses componentes.

4.4 - REGIONALIZAÇÃO PLUVIOMÉTRICA

A análise de agrupamentos foi aplicada às variáveis estudadas, contendo as alterações processadas no item 4.2.3, com a finalidade de obter regiões que apresentem características pluviométricas semelhantes.

Como comentado no apêndice B, a escolha do método de agrupamento a ser utilizado depende dos objetivos do estudo. Isso será confirmado nesse capítulo com a aplicação dos métodos hierárquicos de agrupamento de Ward e da ligação média. O método de Ward, que apresenta a característica de construir grupos com tamanhos semelhantes, foi aplicado com a finalidade de obter regiões que destaquem as tendências mais gerais das chuvas, sem estar preocupado com o comportamento de microrregiões. Já o método da ligação média, que tem a característica de minimizar a soma de quadrados dentro dos grupos e maximizar a soma de quadrados entre eles, foi utilizado com essa finalidade.

A figura 25 mostra o gráfico de R^2 pelo número de grupos

para o método de Ward. Com base nesse gráfico e no mapeamento das regiões para diferentes níveis de discriminação, optou-se pelo agrupamento contendo cinco regiões distintas, que conseguiram destacar de maneira mais resumida as principais diferenças nos perfis de chuva das estações. O agrupamento proposto pelo método de Ward e as chuvas médias dessas regiões por quinquídio no mês de setembro podem ser vistos nas figuras 26 e 27, respectivamente (ver também o apêndice D).

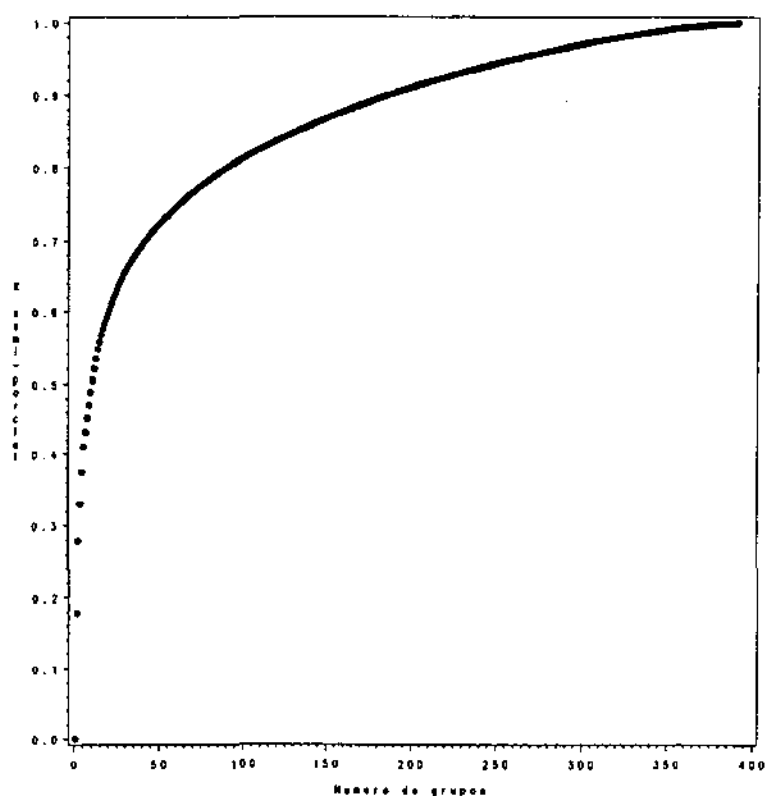


Figura 25 - R^2 por número de grupos para o método de Ward

As principais características dessas regiões são as seguintes:

- região litorânea (1): contendo 31 estações que abrangem quase toda a faixa litorânea, apresenta altos índices de chuvas no mês e um pequeno

número de quinqüídios com ausência completa de precipitação;

- **região litorânea (2)**: formada por 7 estações localizadas na faixa litorânea, porém sem se apresentarem dispostas continuamente, também possui poucos períodos com ausência de chuvas. No entanto nos demais períodos apresenta, em média, cerca de o dobro da quantidade de chuvas que as demais estações litorâneas (ver figura 27);

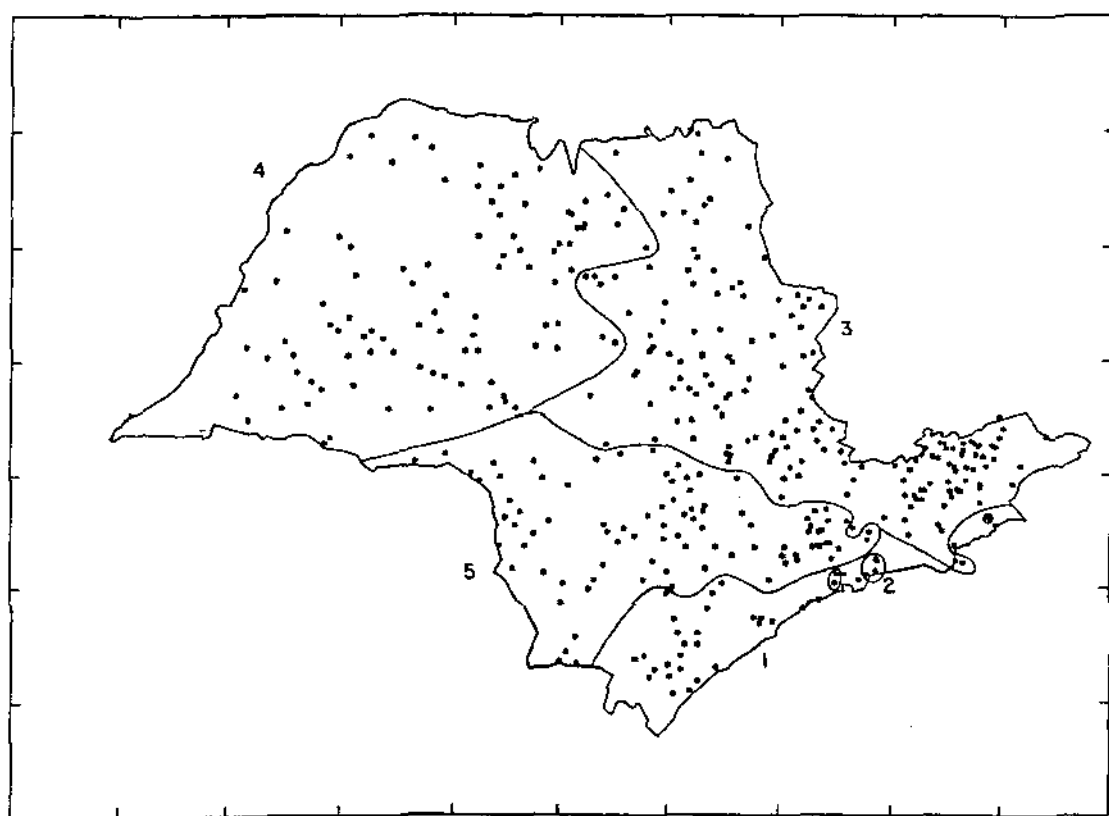


Figura 26 - Agrupamento proposto pelo método de Ward

- **regiões leste (3), noroeste (4) e sudoeste (5)**: apresentam um número maior de períodos com ausência de precipitação em comparação com as regiões litorâneas. No entanto não há uma grande diferença entre essas três regiões, entre si, quanto a essa característica. Uma maior

diferenciação surge quando comparados apenas os quinquídios em que ocorrem chuvas, onde a região leste (169 estações) apresenta médias maiores que a região noroeste (94 estações) e menores que a região sudoeste (90 estações).

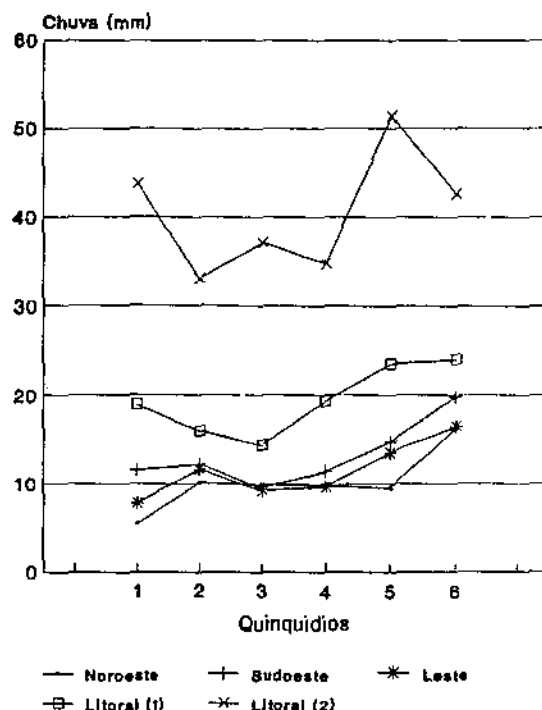


Figura 27 - Chuvas médias por quinquídio para as regiões destacadas pelo método de Ward

O agrupamento proposto pelo método da ligação média apresenta características muito diferentes do proposto pelo método de Ward. Por esse método é necessário um nível de discriminação de 20 grupos até que duas grandes regiões sejam destacadas. Uma dessas regiões destaca apenas o extremo sul do estado e a outra abrange quase todas as estações do interior (ver figura 28). Os demais agrupamentos são constituídos por estações isoladas (16 estações) ou por pequenos grupos de estações litorâneas. Portanto, temos que toda a faixa litorânea do estado é discriminada, porém de uma forma não tão simples como no método de Ward.

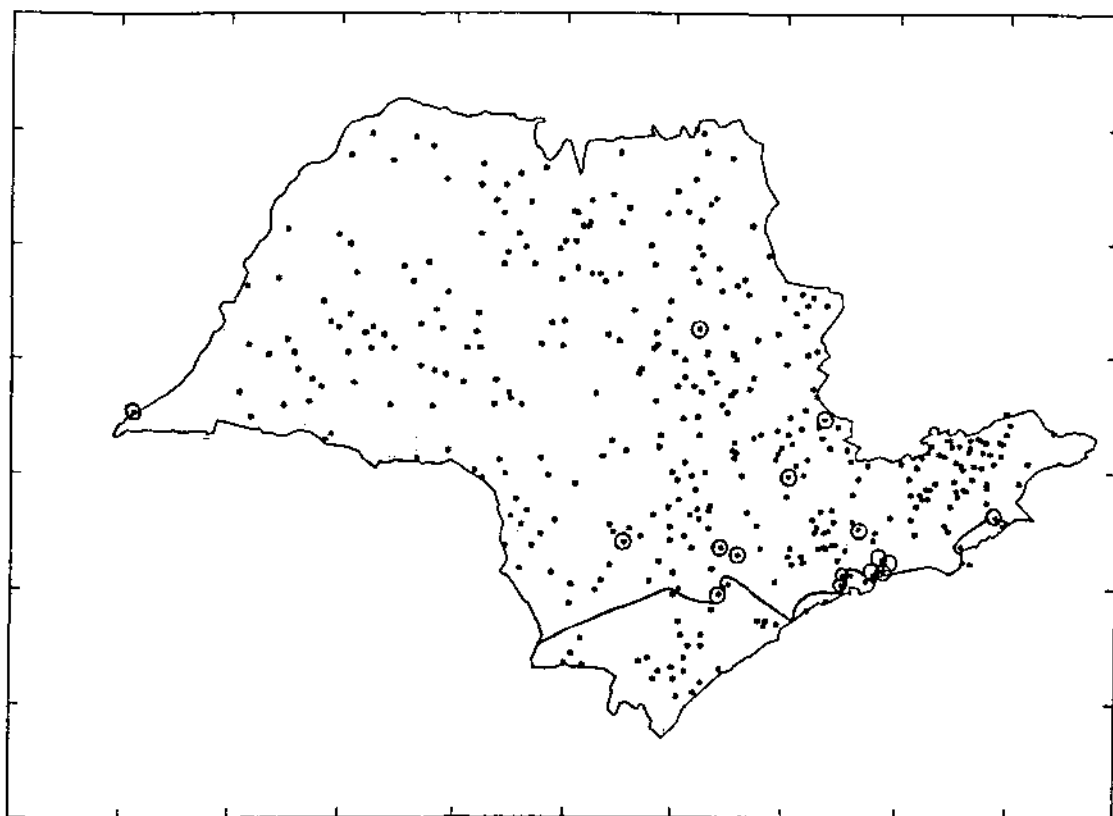
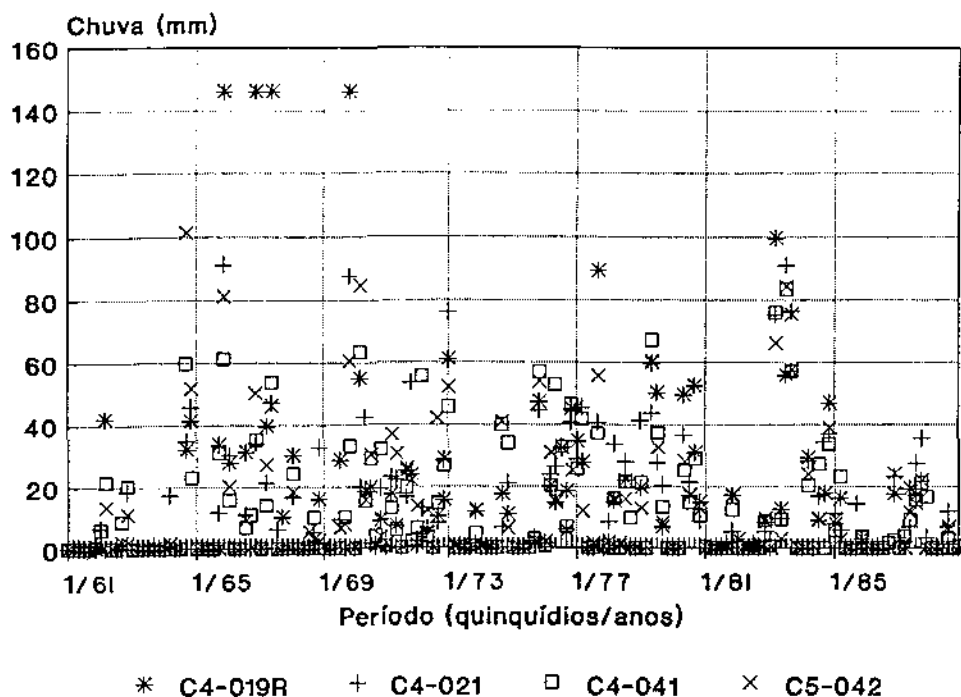
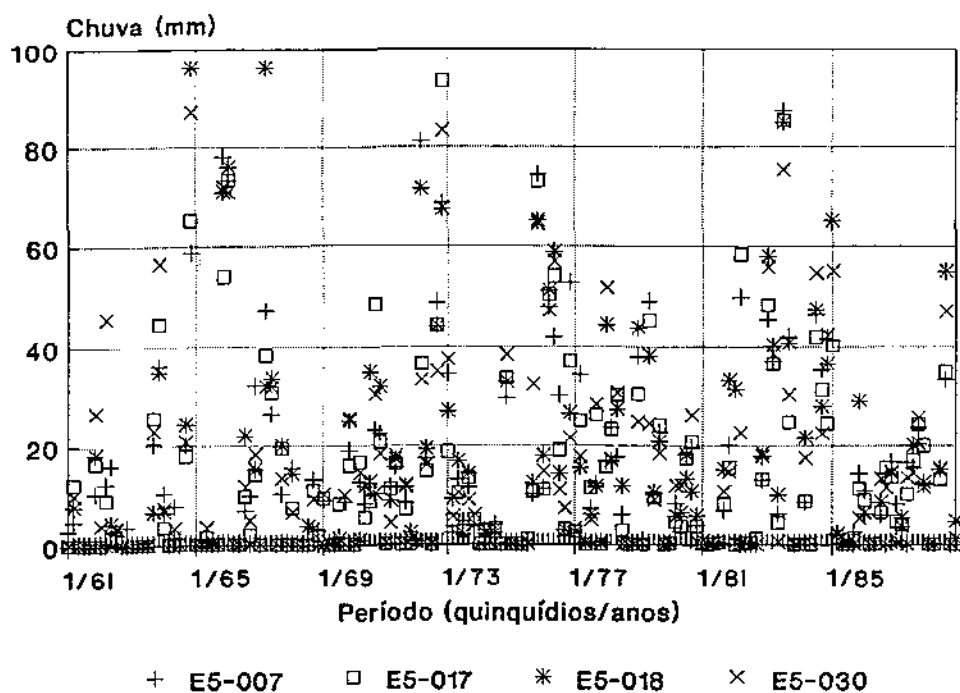


Figura 28 - Agrupamento proposto pelo método da ligação média

Quanto às estações localizadas no interior e que se constituem em grupos unitários, tal comportamento distinto pode ser devido a fatores puramente locais como a proximidade de um rio ou à diferença de altitude. Tomando como exemplo a estação C4-019R, localizada em São Carlos, e comparando o perfil de chuvas dessa estação com o perfil médio de chuvas das três estações mais próximas (C4-021, C4-041 e C5-042), mostrados na figura 29.a, pode-se notar claramente o padrão semelhante apresentado entre esses três perfis e a diferença entre esses três e o primeiro, o que justifica o fato dessas



(a) C4-019R, C4-021, C4-041 e C5-042



(b) E5-007, E5-017, E5-018 e E5-030

Figura 29 - Comparação entre perfis de chuva para algumas estações destacadas pelo método da ligação média

estações estarem agrupadas separadamente. Observe, no entanto, que as grandes diferenças ocorrem em poucos períodos. Na maioria dos períodos a diferença entre a estação de São Carlos e as estações vizinhas não é tão grande. Um comportamento diferenciado também é notado quando se comparam os perfis de chuva da estação E5-018, localizada em Angatuba, com os das estações E5-007, E5-017 e E5-030, que a cercam (ver figura 29.b). Novamente, à exceção de poucos períodos, as diferenças não são grandes se comparadas com as variabilidades entre as estações.

Dessas 16 estações que se constituíram em grupos unitários, 10 tiveram suas observações alteradas pelo procedimento de consistência dos dados, ou seja, mesmo com as modificações elas ainda se destacam.

Assim, tem-se que os dois métodos de agrupamento analisados procuram destacar diferentes características das estações. Enquanto o método de Ward agrupou as estações levando em consideração o perfil geral de chuvas, o método da ligação média, ao contrário, procurou distinguir aquelas estações que, em alguns períodos, apresentaram um comportamento destoante das demais, mesmo sendo esses períodos pouco freqüentes.

4.5 - CONCLUSÕES

O procedimento de consistência dos dados, aplicado antes das análises de componentes principais e de agrupamentos, destacou 27 estações dentre as 391 estudadas que apresentaram pelo menos um quinqüídio com precipitação pluviométrica superior a 150 mm no mês de setembro.

A partir de uma análise descritiva realizada aos dados de cada uma dessas 27 estações foram feitas alterações às observações superiores a 150 mm e que apresentaram uma diferença maior que o dobro da diferença entre quartis da observação seguinte. Foram feitas um total de 21 alterações, executadas em 15 estações distintas.

A análise de componentes principais foi aplicada à matriz de correlação das variáveis estudadas, contendo as alterações comentadas acima, e explica 75% da variabilidade total dos dados com a retenção de 35 componentes. Os três primeiros componentes explicam 23.9% , 6.7% e 5.8% da variabilidade total, respectivamente.

O mapa dos valores do primeiro componente principal mostrou uma distribuição espacial semelhante ao das chuvas médias no Estado de São Paulo durante o mês de setembro. O segundo componente apresenta uma tendência na direção sudoeste-nordeste, podendo estar relacionada à atuação de frentes polares no estado. O terceiro componente principal apresenta uma tendência na direção sudeste-noroeste, porém não foi encontrada uma interpretação climatológica para esse comportamento.

A análise de agrupamentos foi aplicada às variáveis estudadas através dos métodos hierárquicos de Ward e da ligação média, os quais apresentaram resultados diferentes.

O agrupamento proposto pelo método de Ward, com um nível de discriminação de cinco grupos, destacou regiões que mostraram-se espacialmente contínuas e apresentam padrões de chuva distintos.

Já o agrupamento proposto pelo método da ligação média necessita de um nível de discriminação de 20 grupos para destacar apenas duas grandes regiões em todo o estado. Desses 20 grupos, 16 são

formados por estações isoladas e que apresentaram um comportamento destoante daqueles apresentados pelas estações vizinhas.

Embora os agrupamentos obtidos pelos dois métodos apresentem resultados muito diferentes, esses resultados podem ser validados de acordo com a interpretação a ser feita, levando-se em consideração as características de cada método. Ambos são, portanto, bastante úteis na análise e interpretação dos dados.

CAPÍTULO V

CONCLUSÕES

O objetivo desse trabalho não foi fazer uma revisão exhaustiva das aplicações de análise multivariada em pluviometria, e nem propor uma regionalização pluviométrica final das áreas estudadas, mas mostrar algumas aplicações das técnicas estatísticas, como a de componentes principais e a de agrupamentos, no estudo desse tipo de problema.

Cada conjunto de dados e objetivo demanda um tratamento específico de forma a extrair os melhores resultados. Daí a necessidade de se estudarem claramente as características de cada técnica e se definirem os objetivos.

Nos dois casos estudados a análise de componentes principais conseguiu detectar através dos primeiros componentes os principais padrões de chuva atuantes nas regiões analisadas. No caso do nordeste paraense observou-se padrões temporais e espaciais. Esses padrões foram detectados através dos gráficos dos coeficientes dos componentes principais e do diagrama de dispersão dos valores desses componentes.

Já para o Estado de São Paulo, esse procedimento não propiciou uma interpretação clara dos componentes principais devido às características das variáveis estudadas. Isso pode ser explicado pelo fato de terem sido observados apenas valores no mês de setembro. Embora nesse mês se verifique um aumento nas quantidades de chuva no estado, a variação do padrão temporal não se mostrou tão importante como a variação do padrão espacial, sendo que a construção dos mapas

dos valores das estações para cada componente principal possibilitou uma interpretação mais clara desses componentes.

Nas duas situações a análise com a matriz de correlação mostrou-se a mais indicada pois evita possíveis distorções provocadas pelo comportamento das estações litorâneas.

Na análise do nordeste paraense, que possui poucas observações disponíveis, sendo possível comparar diretamente o perfil de chuvas de cada estação, o método de agrupamento hierárquico de Ward apresentou a melhor performance dentre os utilizados.

Para a análise do Estado de São Paulo, a aplicação de dois métodos diferentes (Ward e ligação média) levou a resultados diferentes. Porém, ao levar em consideração as características de cada método tem-se que tais resultados procuram destacar diferentes características climatológicas. Em suma, métodos diferentes podem trazer diferentes informações. Assim, devem ser utilizados os métodos cujas características possam atender aos objetivos da análise.

Em grande parte das análises os agrupamentos obtidos mostraram-se espacialmente contínuos, o que representa um ponto favorável à aplicação dessa técnica.

De uma maneira geral os resultados obtidos podem ser considerados muito bons e coerentes com as possíveis interpretações climatológicas. Devido ao caráter exploratório das técnicas utilizadas torna-se muito importante a interação entre as análises estatísticas com as informações fornecidas por especialistas em climatologia.

Não restam dúvidas que, mesmo no âmbito de aplicação de técnicas de componentes principais e análise de agrupamentos, existem ainda várias possibilidades de se aprofundar a análise. Por exemplo,

poderiam ser realizadas transformações nas variáveis, testados diferentes tipos de agregações ou fazer uma análise mais completa de outliers. No entanto, como enfatizado anteriormente, não faz parte do objetivo realizar um estudo exaustivo; além do que, as análises apresentadas já mostram, de forma clara, a utilidade da aplicação dessas técnicas estatísticas em pluviometria.

APÊNDICE A

ANÁLISE DE COMPONENTES PRINCIPAIS

A idéia central da análise de componentes principais é reduzir a dimensionalidade de um conjunto de dados em que há um grande número de variáveis interrelacionadas, retendo, tanto quanto possível, a variabilidade apresentada por esse conjunto (Jolliffe, 1986, p. 1). Basicamente, a análise de componentes principais consiste na aplicação de uma transformação ortogonal às variáveis originais, denominada transformação de componentes principais, que visa rearranjar a variabilidade total dos dados de forma a concentrá-la em apenas algumas das novas variáveis. Essa transformação é feita também com o objetivo de facilitar a interpretação do comportamento dos dados.

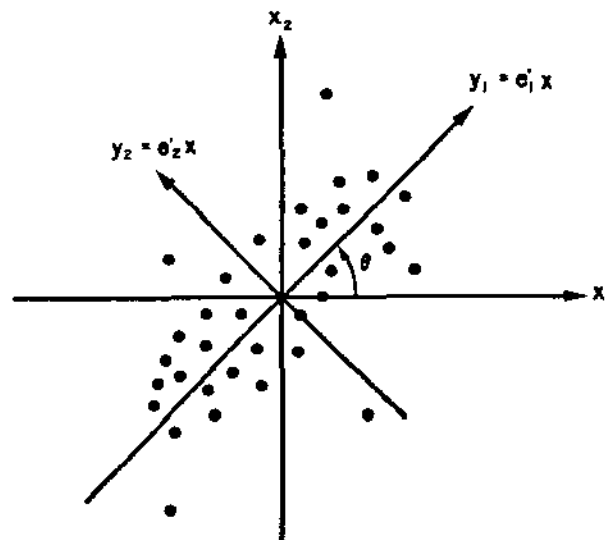


Figura 30 - Exemplo da atuação da transformação de componentes principais para um caso bidimensional

Um exemplo de como atua a análise de componentes principais, para o caso em que se tem apenas duas variáveis, pode ser visto na

figura 30. Na situação encontrada tem-se que a variabilidade total, que inicialmente estava distribuída quase que igualmente entre as duas variáveis (x_1 e x_2), com a aplicação da transformação de componentes principais ficou concentrada em uma das novas variáveis obtidas (y_1), enquanto que a dispersão apresentada pelos dados em relação à outra variável (y_2) foi baixa.

Uma definição de transformação de componentes principais (Mardia, Kent & Bibby, 1979, p. 214), dentro de um contexto populacional, é a seguinte:

"Seja x um vetor aleatório de dimensão p , com esperança μ e matriz de covariância Σ , então a transformação de componentes principais é uma transformação linear dada por $y = \Gamma' (x - \mu)$, onde Γ é uma matriz ortogonal satisfazendo a $\Gamma' \Sigma \Gamma = \Lambda$, e Λ é diagonal e com elementos $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Ao i -ésimo elemento de y (y_i) usa-se a denominação de i -ésimo componente principal de x ."

Pelo teorema de decomposição espectral, uma vez que Σ é uma matriz simétrica, então Γ é a matriz cujas colunas são os autovetores normalizados de Σ , e Λ é a matriz diagonal formada pelos seus autovalores. Tem-se, também, que o i -ésimo componente principal é uma combinação linear normalizada de x , não correlacionada com os $i-1$ primeiros componentes principais, que apresenta máxima variância, sendo o primeiro componente aquela combinação linear que apresenta a maior variância entre todas as combinações lineares possíveis. Várias propriedades emergem da definição acima, dentre as quais pode-se destacar as seguintes:

$$i \quad - E(y_i) = 0, \text{ para } i = 1, 2, \dots, p;$$

$$ii \quad - V(y_i) = \lambda_i, \text{ para } i = 1, 2, \dots, p;$$

$$iii - \text{Cov}(y_i, y_j) = 0, \text{ para } i, j = 1, 2, \dots, p, i \neq j;$$

$$iv - V(y_1) \geq V(y_2) \geq \dots \geq V(y_p) \geq 0;$$

$$v - V(y_1) + V(y_2) + \dots + V(y_p) = \text{traço } \Sigma.$$

A partir dessas propriedades torna-se mais fácil visualizar a forma com que a transformação de componentes principais atua no processo de redução da dimensionalidade dos dados. A variabilidade explicada por cada um dos componentes principais é igual ao valor do seu respectivo autovalor. Como os autovalores encontram-se ordenados decrescentemente, a maior parte da variabilidade dos dados pode se concentrar nos primeiros componentes. Sendo baixa a proporção da variabilidade explicada pelos últimos componentes, então esses poderão ser ignorados, sem a perda considerável de informação.

No caso em que não são conhecidas todas as observações do vetor aleatório $\mathbf{x}$, assim como a sua média μ e sua matriz de covariância Σ , mas são conhecidas as observações $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ pertencentes a uma amostra dessa população, e conseqüentemente as estimativas $\mathbf{m}$ e $\mathbf{S}$ do vetor de médias e matriz de covariância, respectivamente, então a transformação de componentes principais pode ser definida de maneira análoga ao do caso populacional como $\mathbf{Y} = (\mathbf{X} - \mathbf{1}\mathbf{m}') \mathbf{G}$, onde $\mathbf{G}'\mathbf{S}\mathbf{G} = \mathbf{L}$ é diagonal. Aqui $\mathbf{Y}$ é uma matriz $n \times p$ constituída pelos componentes principais, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ é a matriz de dados amostrais, $\mathbf{1}' = (1, \dots, 1)'$ é o vetor coluna $n \times 1$, $\mathbf{G}$ é a matriz ortogonal cujas colunas são os autovetores normalizados de $\mathbf{S}$, e $\mathbf{L}$ é a matriz diagonal formada pelos autovalores de $\mathbf{S}$. As propriedades comentadas anteriormente continuam válidas desde que feitas as devidas modificações dos termos populacionais para os termos amostrais.

Em muitas situações, principalmente quando as variáveis originais encontram-se em unidades diferentes, ou nas mesmas unidades mas apresentando escalas de variabilidade muito diferentes, torna-se mais adequado utilizar variáveis padronizadas. Isso é análogo a utilizar a matriz de correlação em lugar da matriz de covariância. Uma vez que as matrizes de correlação e covariância não necessariamente satisfazem a uma relação linear entre si, então seus autovetores também não obedecerão a uma relação desse tipo. Portanto, os resultados e interpretações de análises realizadas pelos dois casos podem levar a conclusões diferentes.

Uma situação particular pode ser encontrada quando o número de observações for menor que o número de variáveis estudadas ($n < p$). Nesses casos tem-se que as matrizes de covariância e correlação serão singulares. Alguns pesquisadores desaconselham a utilização de componentes principais nesses casos (Flury & Riedwyl, 1988, p. 197-198). Entretanto, outros pesquisadores são favoráveis à aplicação de componentes principais nessas situações (Kshirsagar, 1972, p. 462-463). Essas situações geralmente ocorrem quando se trata de eventos raros, como no caso de estudo de doenças raras (Dempster, 1963). Na literatura meteorológica Dyer (1975) e Gadgil & Iyengar (1980) podem ser citados como exemplos de casos em que é observada a singularidade da matriz de dispersão.

Uma medida da importância das variáveis originais na explicação de um componente principal pode ser dado pelo coeficiente de correlação entre eles. Nesse caso pode-se mostrar que essa correlação é dada por $\rho_{1j} = \text{corr}(x_1, y_j) = \gamma_{1j} * \sqrt{\lambda_j} / \sqrt{\sigma_{11}}$ quando utilizada a matriz de covariância (onde γ_{1j} é o i -ésimo elemento do

autovetor j , e σ_{1j} é a variância de x_1), e $\rho_{1j} = r_{1j} \cdot \sqrt{\lambda_j}$ se aplicada à matriz de correlação. Nesse último caso, uma vez que para um mesmo componente principal o valor de λ_j não se altera, então essa correlação será proporcional ao valor do coeficiente da variável original para esse componente.

Em situações em que as matrizes de covariância ou de correlação apresentam alguns padrões conhecidos, os componentes principais podem ser expressos em formas simples. Se a matriz de covariância for diagonal, então seus autovetores serão os vetores elementares, e conseqüentemente o conjunto de componentes principais será exatamente o conjunto de variáveis originais. Com isso, nenhuma simplificação será obtida com a aplicação dessa análise. Por outro lado, se as variáveis originais forem igualmente correlacionadas então o primeiro componente dará a mesma importância a todas as variáveis originais, se a análise for aplicada à matriz de correlação. Os demais componentes explicarão, cada um, a mesma proporção da variabilidade dos dados.

Como um dos objetivos da análise de componentes principais é a redução no número de variáveis a serem consideradas, porém sem a perda significativa da informação contida na variabilidade dos dados, então a seleção dos componentes principais que serão retidos representa um dos passos mais importantes dentro de toda a análise. Como não há nenhum modelo estatístico sendo considerado, essa não será uma decisão estatística. Porém existem muitos critérios que são propostos por vários autores e que podem contribuir para essa tomada de decisão.

Um critério proposto consiste em reter os componentes cuja

variação acumulada explique uma alta proporção da variabilidade total dos dados. Outro critério, o critério de Kaiser, consiste em considerar apenas os componentes cujos autovalores sejam maiores que a média dos autovalores. Um critério proposto por Cattell (1966), e denominado de teste scree, consiste em graficar os autovalores em função da sua ordem. Aquele valor de ordem tal que as inclinações das retas que unem os pontos graficados sejam acentuadas à esquerda, mas não o sejam à direita, é que indicará o número de componentes a ser retido.

Quando se considera uma amostra aleatória de dados normais, uma série de testes de hipóteses pode ser realizado para decidir quantos componentes devem ser retidos. Dentre alguns testes, podem ser citados os que consideram as hipóteses nulas de que os $(p-k)$ últimos autovalores sejam iguais e a de que $(\lambda_1 + \dots + \lambda_k) / (\lambda_1 + \dots + \lambda_p) = \psi$, onde ψ é uma proporção definida. Algumas críticas podem ser encontradas a respeito desses critérios, tais como a de que o critério de Kaiser tende a incluir poucos componentes quando se trabalha com menos de 20 variáveis, e de que o teste scree tende a incluir muitos componentes.

Uma discussão muito mais ampla sobre a análise de componentes principais pode ser feita envolvendo outros temas de pouco interesse frente ao desenvolvimento das análises executadas nessa tese. Uma revisão mais aprofundada para esse assunto pode ser encontrada nos capítulos específicos apresentados por Mardia, Kent & Bibby (1979, cap. 8), Johnson & Wichern (1982, cap. 8), e Jolliffe (1986), e que serviram de referência para esse apêndice.

APÊNDICE B

ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos consiste em uma técnica estatística cujo objetivo principal é encontrar e separar objetos ou indivíduos em grupos. Idealmente, procura-se que os componentes de um mesmo grupo apresentem características muito similares entre si, enquanto que os componentes de grupos diferentes apresentem características distintas que justifiquem essa separação.

Na prática, tem-se disponível uma amostra de tamanho n de objetos ou indivíduos, medidos sobre p variáveis, na qual a análise de agrupamentos procurará aplicar um esquema de classificação que agrupe esses elementos em g grupos distintos, onde esse número de grupos e as suas características também deverão ser determinadas (Everitt, 1974).

Os principais passos dentro de uma análise de agrupamentos, baseado em Bussab, Miazaki & Andrade (1990, p. 1), são os seguintes:

- i - definição de objetivos, critérios, variáveis e elementos;
- ii - obtenção e tratamento dos dados;
- iii - escolha de critérios de parença entre os elementos;
- iv - adoção de um algoritmo de agrupamento;
- v - apresentação, avaliação e interpretação dos resultados.

Dentre essas fases, uma das mais importantes consiste em quantificar a diferença apresentada entre dois elementos. Para essa finalidade é utilizada uma medida, chamada de coeficiente de parença, e que pode ser dividida em duas categorias de acordo com a

sua forma de atuação: similaridade e dissimilaridade. No primeiro caso, quanto maior o valor encontrado mais semelhantes serão considerados os elementos, enquanto que no segundo caso o comportamento é o oposto. Um grande número de coeficientes de parença pode ser encontrado, inclusive levando em consideração o tipo de variável que esteja sendo analisada e até mesmo a área de aplicação em que se esteja trabalhando. Dentre os mais usuais pode-se citar a distância euclidiana, a distância de Mahalanobis e o coeficiente de correlação.

Uma vez escolhida uma medida de parença, a etapa seguinte será a de adotar um algoritmo de agrupamento entre os elementos. Aqui também encontraremos uma grande variedade de métodos disponíveis. Esses métodos de agrupamento podem ser classificados, basicamente, em três tipos:

- hierárquico: os elementos são classificados em grupos em diferentes etapas, de tal forma que, em uma certa etapa, o agrupamento obtido será originado da união de dois agrupamentos da etapa anterior (aglomerativo), ou serão obtidos dois agrupamentos a partir da divisão de um dos agrupamentos da etapa antecedente (divisivo). Como exemplo têm-se os métodos de Ward, do centróide e da ligação média;

- partição: os agrupamentos obtidos produzem uma partição do conjunto de objetos de modo a satisfazer as condições de coesão interna entre os elementos de um mesmo grupo e isolamento entre diferentes grupos. Nessa técnica o número de grupos resultantes é considerado conhecido a princípio, e há a necessidade de alocar os elementos a agrupamentos iniciais, os quais irão se alterando no decorrer da análise. Um dos métodos mais utilizados é o método das

k-médias;

- cobertura: são técnicas nas quais os agrupamentos obtidos podem se sobrepor um ao outro.

A escolha da técnica de agrupamento a ser aplicada é geralmente dependente dos objetivos e características da análise considerada. Em estudos meteorológicos, onde os dados possuem propriedades particulares, e que geralmente não são fortemente nucleadas, as técnicas hierárquicas têm sido preferencialmente adotadas, sendo que essas são também consideradas as mais eficientes quando trabalhando com um grande número de indivíduos (Belbin, 1984). Uma avaliação da performance de três dos principais métodos de agrupamento hierárquico (Ward, centróide e ligação média), para utilização na classificação climatológica sinótica, é feita por Kalkstein, Tan & Skindlov (1987). As principais características apresentadas por esses métodos são as seguintes:

- Ward: une os agrupamentos de forma a minimizar, a cada etapa, a soma de quadrados dentro dos grupos. A medida de similaridade entre dois grupos é a soma de quadrados entre eles, somado sobre todas as variáveis;

- centróide: une os grupos com a menor distância entre centróides. A medida de similaridade é a distância euclidiana ao quadrado entre os centróides de dois grupos;

- ligação média: calcula a distância euclidiana quadrada de todos os possíveis pares de observações de dois grupos, sendo um elemento de cada grupo. A medida de similaridade consiste na distância euclidiana quadrada média.

Kalkstein, Tan & Skindlov (1987) concluem, a partir de uma

avaliação teórica dos métodos, que o método de Ward possui a tendência de agrupar um elemento ao grupo que tiver menos componentes, e conseqüentemente, o de apresentar grupos com números semelhantes de elementos. O método do centróide, por sua vez, apresenta a tendência de formar um grupo muito grande e muitos grupos pequenos, sendo susceptível ao problema de encadeamento. Quanto ao método da ligação média, apresenta a característica de minimizar a variância dentro dos grupos e maximizar a variância entre eles.

A comparação entre os três métodos foi feita também com dados reais de alguns índices sinóticos temporais, para classificação de dias meteorologicamente homogêneos, para duas cidades norte-americanas, e os resultados obtidos nos dois casos apresentaram concordância com a avaliação teórica. Isso não significa, porém, que os outros métodos não devam ser aplicados, visto que pode-se encontrar na prática situações em que a formação de grupos de tamanhos similares ou muito diferentes representa, muito mais que uma deficiência do método, uma característica da amostra analisada.

Uma das dificuldades que surgem no decorrer da análise, sobretudo quando aplicadas as técnicas hierárquicas, é a de determinar o número ideal de grupos existentes. Não há métodos satisfatórios para determinar o número de grupos para qualquer tipo de análise de agrupamentos (Everitt, 1979). Quando a análise apresenta um caráter exploratório e não se tem indícios a respeito desse número, a experiência do pesquisador na área de aplicação da análise será um fator fundamental nessa decisão. Contudo, é importante também a procura por grupos em que os elementos dentro deles sejam o mais semelhantes possível, e que os elementos de grupos diferentes

apresentem baixa similaridade.

Face ao grande número de opções de técnicas de agrupamento que podem ser executadas, e conseqüentemente de diferentes conjuntos de agrupamentos que podem ser obtidos, torna-se recomendável a aplicação de vários métodos ao conjunto de dados. A constituição do agrupamento mais racional poderá ser evidenciada por grande parte dos métodos, sendo importante sua validação a partir da opinião de especialistas. Esse procedimento também é importante para se evitar a adoção de soluções espúrias que podem ser obtidas a partir da aplicação automática de um particular método. Tal problema pode ocorrer com demasiada facilidade se os dados apresentarem algum tipo de estrutura não circular, ou quando se tenta impor uma estrutura de hierarquia aos dados, através de métodos hierárquicos, quando esses não a possuem.

Uma atenção especial deve ser dada ao processo de seleção das variáveis que serão utilizadas na análise. Deverão ser utilizadas as variáveis que realmente tenham um poder de discriminação entre os elementos considerados e que sejam de relevância frente aos objetivos que deverão ser atingidos pela análise. É importante, também, precaver-se quanto à presença de outliers, visto que os métodos de agrupamento são sensíveis a eles (Johnson & Wichern, 1982, p. 554). A aplicação inicial de análises como a de componentes principais tem grande poder para minimizar esse problema (Jolliffe, 1986, p. 163; Gnanadesikan, 1977; Rao, 1964).

Uma vasta abordagem sobre a análise de agrupamentos, tratando tanto dos temas comentados anteriormente como de outros assuntos de menor importância dentro do contexto dessa tese, pode ser

encontrada em Cormack (1971), Everitt (1974) e Bussab, Miazaki & Andrade (1990), textos esses que contribuíram para a exposição das idéias básicas apresentadas nesse apêndice.

APÊNDICE C

DESCRIÇÃO DAS ESTAÇÕES PLUVIOMÉTRICAS - NORDESTE PARAENSE

Cód.	Localidade	Município	Latit.	Longit.	Fonte
A	Belém	Belém	01°28'S	48°27'W	1
B	Castanhal	Castanhal	01°17'S	47°55'W	1
C	Capitão Poço	Capitão Poço	01°46'S	47°04'W	1
D	Paragominas	Paragominas	02°59'S	47°28'W	1
E	Tomé-açu	Tomé-açu	02°31'S	48°22'W	1
F	Pirelli	Benevides	01°22'S	48°03'W	2
G	Tracuateua	Bragança	01°04'S	46°46'W	3
H	Km zero PA-70	Paragominas	04°27'S	47°48'W	4
I	Alto Bonito	Vizeu	01°50'S	46°13'W	4
J	Vizeu	Vizeu	01°13'S	46°08'W	4
K	Capanema	Capanema	01°11'S	47°10'W	4
L	Tararua	Vizeu	01°42'S	46°36'W	4
M	Salinópolis	Salinópolis	00°36'S	47°28'W	4
N	Sta. Isabel do Pará	Sta. Isabel do Pará	01°18'S	48°09'W	4
O	Primavera	Primavera	00°57'S	47°07'W	4
P	Bom Jardim	Ourém	01°34'S	47°06'W	4
Q	Badajós	S. Domingos do Capim	02°23'S	47°49'W	4
R	Mosqueiro	Belém	01°10'S	48°27'W	4
S	S. Domingos do Capim	S. Domingos do Capim	01°41'S	47°46'W	4
T	Rondon do Pará	S. Domingos do Capim	04°58'S	48°31'W	4
U	Igarapé-açu	Igarapé-açu	01°08'S	47°37'W	4
V	Colônia Sto. Antônio	Irituia	01°37'S	47°29'W	4

Fontes dos dados:

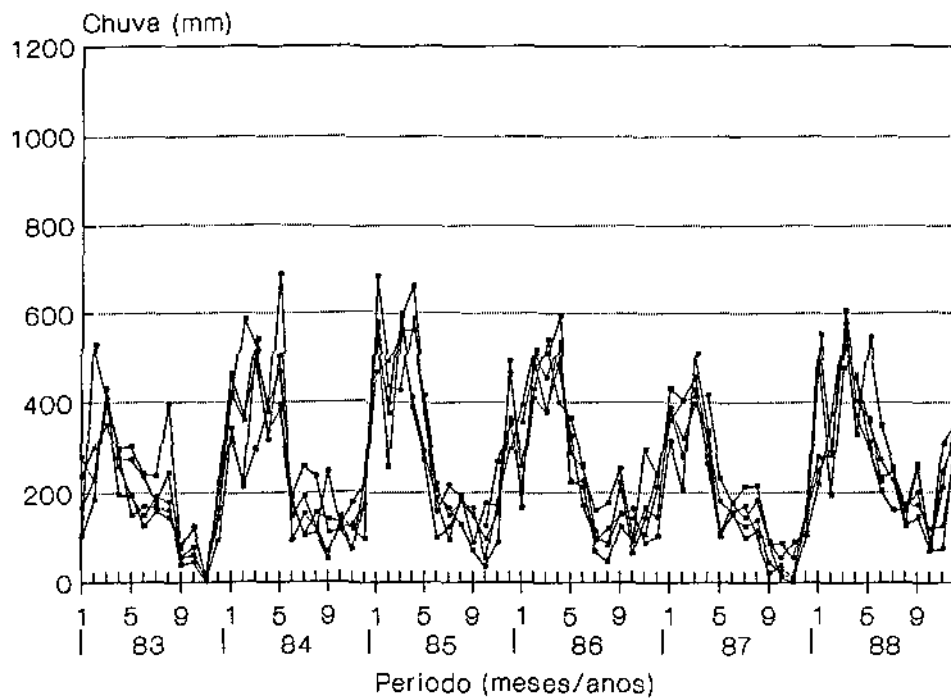
1 - EMBRAPA/CPATU

2 - EMBRAPA/PIRELLI

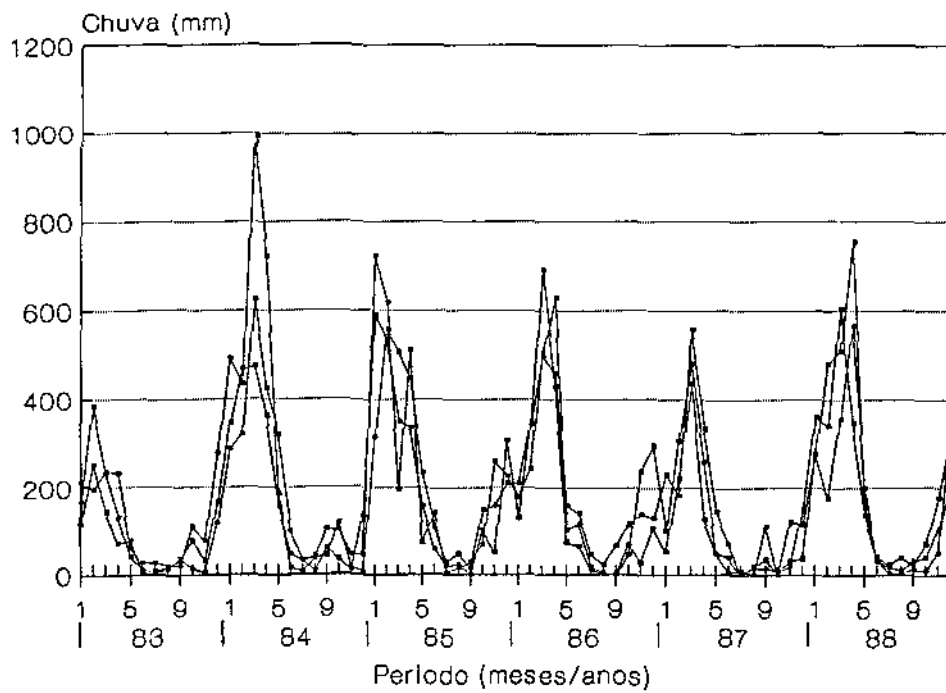
3 - INEMET

4 - DNAEE

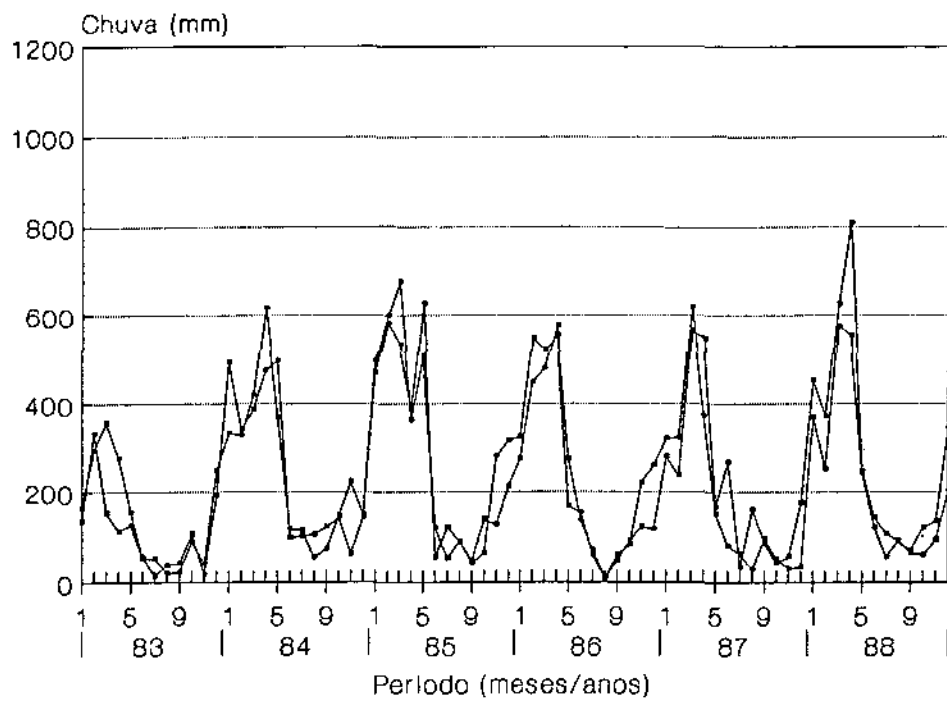
Figura 31 - Perfis de chuva de 1983 a 1988 das estações pluviométricas do nordeste paraense



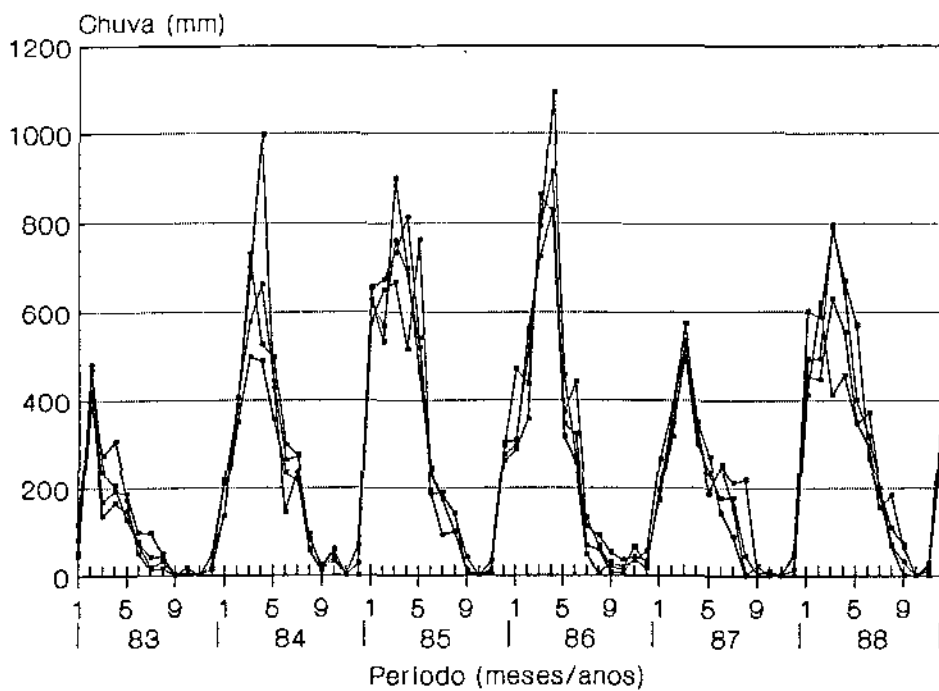
(a) Região de Belém
Postos: A, B, F, N e R



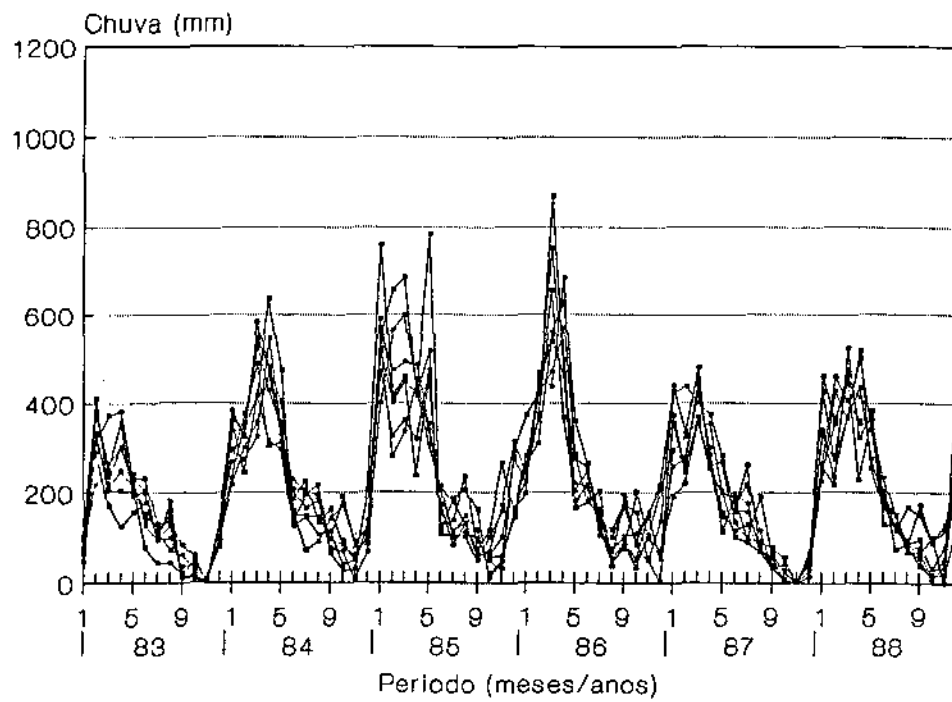
(b) Região sul
Postos: D, H e T



(c) Reglao oeste
Postos: E e Q



(d) Reglao Ilitoranea
Postos: G, J, M e O



(e) Região central
Postos: C, I, K, L, P, S, U e V

APÊNDICE D

DESCRIÇÃO DAS ESTAÇÕES PLUVIOMÉTRICAS - ESTADO DE SÃO PAULO

REDE PLUVIOMÉTRICA DO DAEE

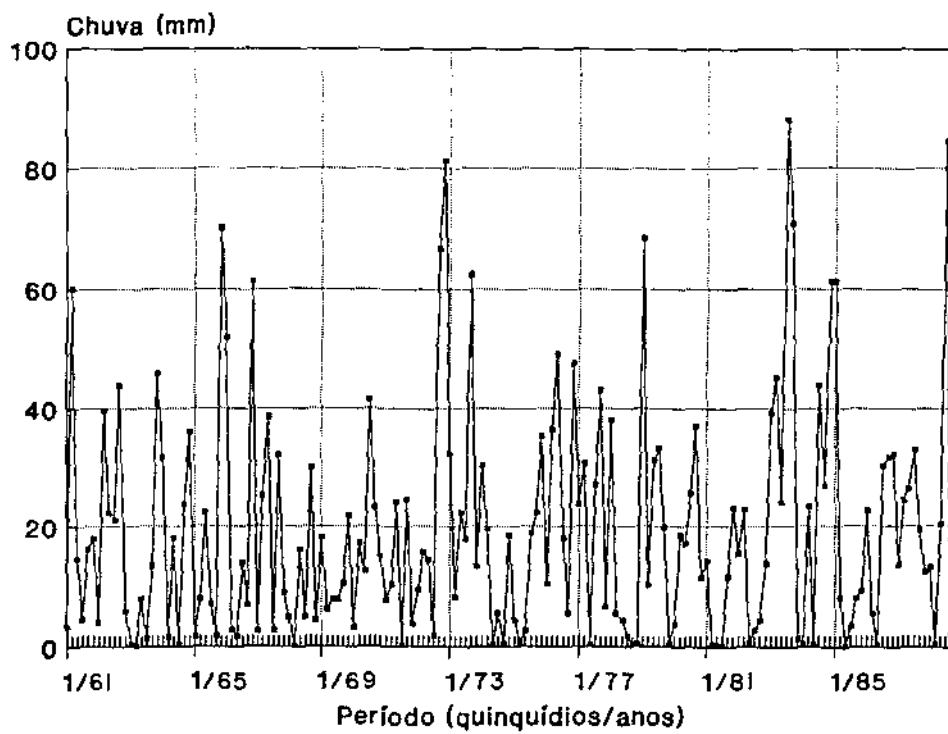
B4-002	B4-003	B4-005	B4-012	B4-015
B4-018	B4-021	B4-026	B4-029	B4-035R
B4-037	B5-002	B5-003	B5-004	B5-005
B5-012	B5-015	B5-016	B5-020	B5-024
B5-029	B5-034	B5-035	B6-001	B6-003
B6-006	B6-008	B6-009	B6-010	B6-022
B6-023	B6-030R	B6-033	B6-034R	B7-008
B7-011	B7-012	B7-013	B7-014	B7-024
B8-002	B8-012	C3-011	C3-014	C3-029
C3-030	C3-031	C3-034	C3-035R	C3-040
C4-001	C4-019R	C4-021	C4-029	C4-032
C4-033	C4-034	C4-040	C4-041	C4-043
C4-052	C4-054	C4-071	C4-072R	C4-075
C4-083R	C4-085	C4-087	C4-088	C5-016
C5-017R	C5-018	C5-020	C5-027	C5-029
C5-035	C5-040	C5-041	C5-042	C5-048
C5-050	C5-056	C5-070	C5-073	C5-074
C5-082	C5-096	C6-008	C6-023R	C6-031
C6-036	C6-040R	C6-041	C6-050R	C6-053
C6-056	C6-066	C6-071	C6-078	C7-001R
C7-006	C7-008	C7-009	C7-010	C7-011
C7-023	C7-034	C7-036	C7-043	C7-045
C7-046	C7-062	C7-064	C7-075	C7-079
C8-004	C8-008	C8-009	C8-011	C8-014
C8-018	C8-019R	C8-026R	C8-030	D1-003
D1-005R	D1-006	D1-008	D2-001	D2-004
D2-005	D2-007	D2-009	D2-013R	D2-015

D2-020	D2-021R	D2-026	D2-028	D2-029
D2-031	D2-034R	D2-035	D2-037	D2-041
D2-060	D2-063	D2-065R	D2-067	D2-068
D2-069	D2-070	D2-071	D2-072	D2-075
D2-076	D2-077R	D2-079	D3-002	D3-003
D3-009	D3-012	D3-014	D3-015	D3-018
D3-022	D3-027	D3-030R	D3-031	D3-035R
D3-036	D3-041	D3-042	D3-046	D3-052
D3-054	D4-004	D4-012	D4-014	D4-015
D4-016	D4-022	D4-027	D4-030R	D4-033
D4-034	D4-035	D4-036R	D4-040	D4-043
D4-044	D4-046	D4-047	D4-052R	D4-054
D4-059	D4-060	D4-061	D4-079	D4-082
D4-087	D4-088	D4-092	D5-006	D5-019
D5-023R	D5-029	D5-037	D5-039	D5-044
D5-047	D5-053	D6-005	D6-006R	D6-011
D6-018	D6-019	D6-020	D6-025	D6-035
D6-084	D7-003R	D7-012	D7-031	D7-032
D7-033	D7-036	D7-043	D7-046	D8-003
D8-004R	D8-006	D8-008	D8-013	D8-028
D8-040	D8-041R	D8-047	D9-001	E1-001
E2-001R	E2-002	E2-007	E2-009	E2-012
E2-028	E2-029R	E2-032	E2-034	E2-036
E2-039	E2-040	E2-045	E2-046R	E2-048
E2-049	E2-052R	E2-054	E2-055R	E2-057
E2-091	E2-092	E2-095	E2-097R	E2-098
E2-099	E2-102	E2-127R	E3-002	E3-005

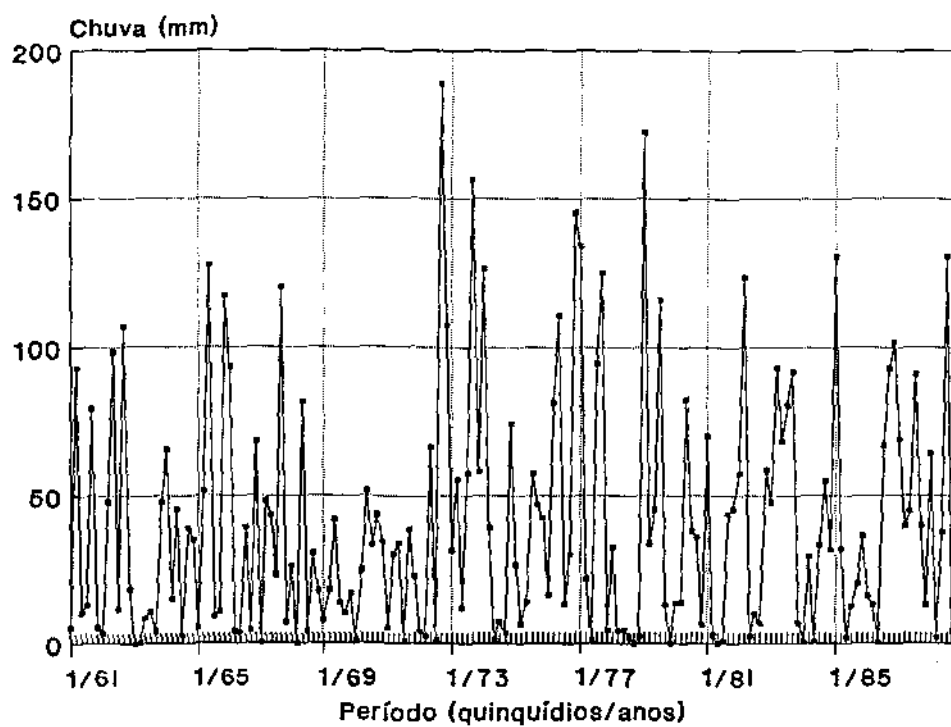
E3-006	E3-007	E3-010	E3-015	E3-016
E3-017R	E3-022	E3-027	E3-030	E3-032R
E3-033	E3-034R	E3-035	E3-040R	E3-041
E3-042R	E3-045R	E3-047R	E3-052	E3-053
E3-066	E3-067	E3-068R	E3-076R	E3-082
E3-085R	E3-094	E3-096R	E3-097	E3-099
E3-101	E3-106	E3-142R	E3-147	E3-148R
E3-153R	E4-001	E4-010R	E4-013R	E4-018
E4-019	E4-020	E4-023R	E4-025R	E4-026
E4-028R	E4-029	E4-031	E4-032	E4-036
E4-037	E4-043	E4-046R	E4-047R	E4-049
E4-050	E4-053	E4-055R	E4-059R	E5-001
E5-007	E5-014	E5-015	E5-017	E5-018
E5-019	E5-023	E5-027R	E5-030	E5-034
E5-045	E5-046	E5-047	E5-051R	E5-069
E6-002	E6-003	E6-007	E6-010	E6-011
E6-012	E6-013	E6-015	E6-016	E6-017
E6-020	E6-022	E6-030	F3-002	F3-005R
F4-001R	F4-004	F4-005	F4-006	F4-007
F4-011	F4-014	F4-015	F4-016	F4-017
F4-018	F4-019	F4-021	F4-022	F4-024
F4-025R	F4-026R	F4-028	F4-029R	F4-030
F4-031	F5-002	F5-005	F5-008	F5-011
F5-012	F5-013	F5-014	F5-016	F5-017
F5-028				

As informações como localização (latitude, longitude e altitude) e município ao qual pertence, entre outras, relativas a cada estação, podem ser encontradas nos boletins pluviométricos do DAEE, como o de São Paulo. DAEE (1974) por exemplo.

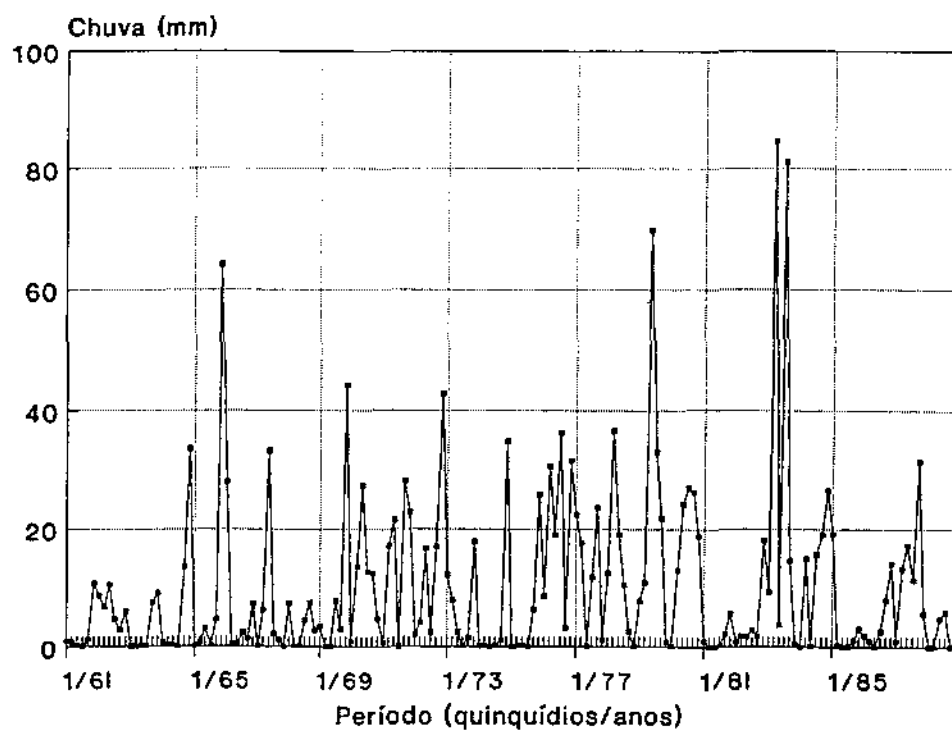
Figura 32 - Perfis de chuva dos agrupamentos propostos
pelo método de Ward



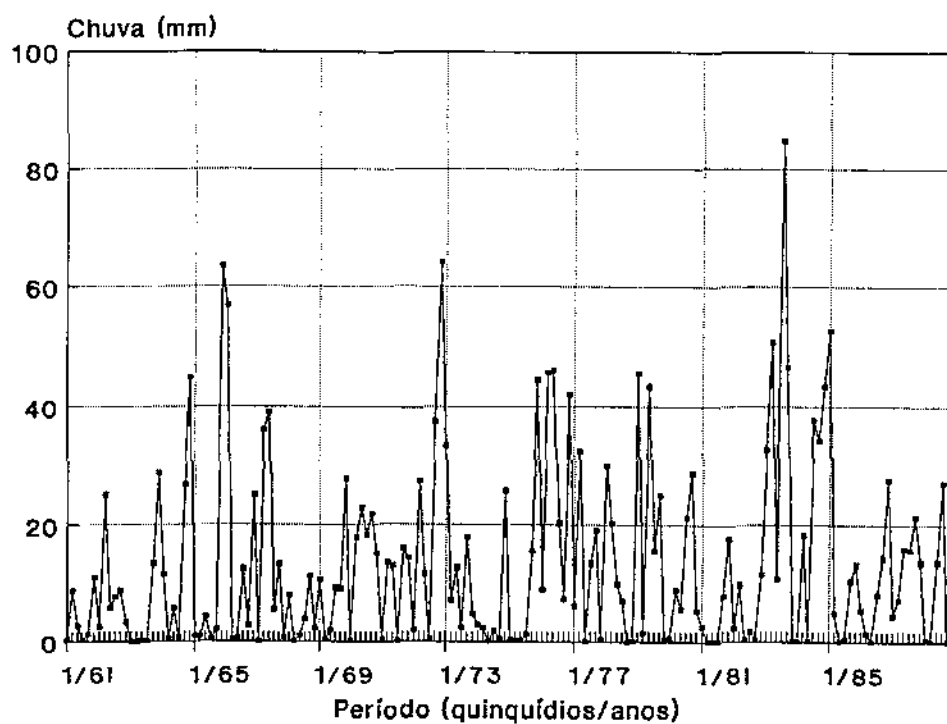
(a) Região litorânea (1)



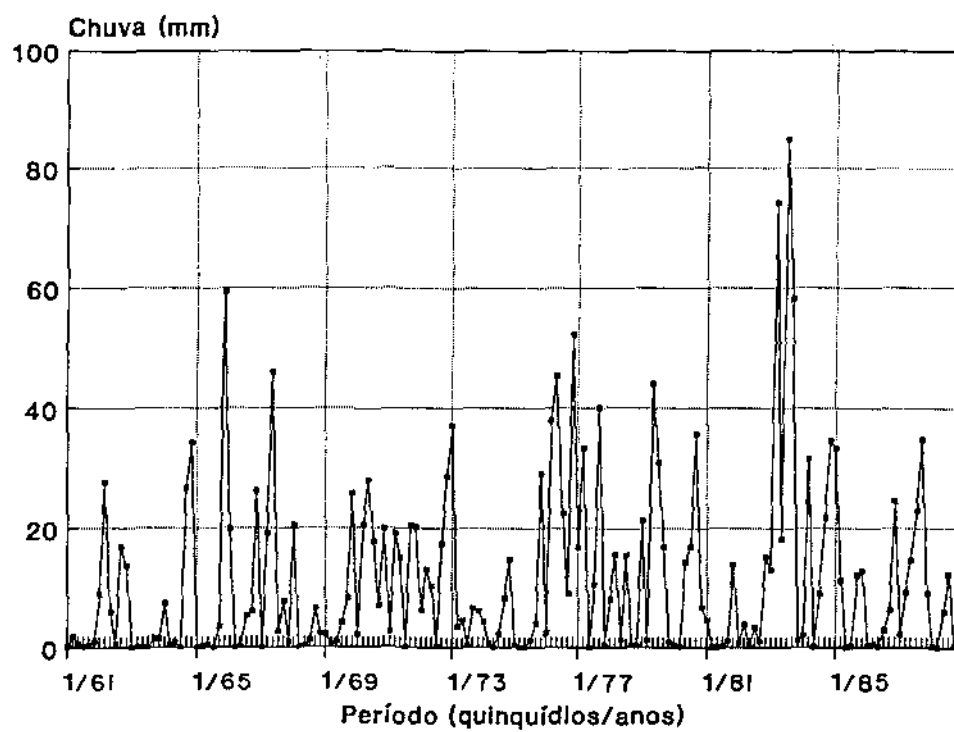
(b) Região litorânea (2)



(c) Região noroeste



(d) Região sudoeste



(e) Região leste

BIBLIOGRAFIA

- ADAS, M. **Panorama geográfico do Brasil: aspectos físicos, humanos e econômicos.** São Paulo, Moderna, 1985. 294 p.
- ALFONSI, R.R.; CAMARGO, M.B.P. de; CHIAVEGATTO, O.M.D.P. *et al.* Simulação de épocas de plantio baseadas no atendimento da demanda hídrica para culturas de soja, milho e arroz de sequeiro no Estado de São Paulo. In: CONGR. BRAS. de AGROMETEOROLOGIA, 6, 1989, Maceió. **Anais...** Maceió, INPE-SBA-UFAL, 1989a. p. 34-38.
- ALFONSI, R.R.; CAMARGO, M.B.P. de; CHIAVEGATTO, O.M.D.P. *et al.* Níveis de probabilidade de seca como subsídio à irrigação para o Estado de São Paulo. In: CONGR. BRAS. de AGROMETEOROLOGIA, 6, 1989, Maceió. **Anais...** Maceió, INPE-SBA-UFAL, 1989b. p. 39-42.
- ARRUDA, H.V. & PINTO, H.S. A simplified gamma probability model for analysis of the frequency distribution of rainfall in the region of Campinas, SP, Brazil. **Agric. Meteorol.**, v. 22, p. 101-108. 1980.
- ASCULAI, E. & THIEBERGER, R. On the randomness of anual average precipitation data. **J. Clim. Appl. Meteor.**, v. 26, p. 1601-1602, Nov. 1987.
- ASPIAZÚ, C.; RIBEIRO, G.A.; VIANELLO, R.L. *et al.* Análise dos componentes principais aplicada na classificação climática do Estado de Minas Gerais. Teste metodológico. **R. Árvore**, v. 14, n. 1, p. 1-15, jan./jun. 1990.
- BASTOS, T.X. O estado atual dos conhecimentos das condições de clima da amazônia brasileira. In: **Zoneamento agrícola da amazônia. 1a. aproximação.** Belém, IPEAN, 1972. p. 68-122. (IPEAN Boletim Técnico,

54).

- BASTOS, T.X. **Delineating agroclimatic zones for deforested areas in Pará State, Brazil.** Hawaii, University of Hawaii, 1990. 170 p. (Ph. D. dissertation).
- BELBIN, L. FUSE: A FORTRAN V program for agglomerative fusion for minicomputers. **Computers and Geosciences**, v. 10, p. 361-384. 1984.
- BLANCO, H.G. & GODOY, H. **Cartas das chuvas do Estado de São Paulo.** Campinas, IAC/SCA, 1967. 12 cartas.
- BOSHELL, F. & NEILD, R.E. A computer-statistical procedure to determine agroclimatic analogues for tea production in Colombia. **Agric. Meteorol.**, v. 15, p. 221-230. 1975.
- BRASIL.MINISTÉRIO DA AGRICULTURA. **Atlas Climatológico do Brasil.** Rio de Janeiro, Ministério da Agricultura, 1969. 100 p.
- BRASIL.SUDAM - Projeto de Hidrologia e Climatologia da Amazônia **Atlas Climatológico da Amazônia Brasileira.** Belém, SUDAM, 1984. 125 p. (Publicação, 339).
- BUISHAND, T.A. The effect of seasonal variation and serial correlation on the extreme value distribution of rainfall data. **J. Clim. Appl. Meteor.**, v. 24, p. 154-160. 1985.
- BUSSAB, W. de O.; MIAZAKI, E.S. & ANDRADE, D.F. de. **Introdução à Análise de Agrupamentos.** In: SINAPE, 9, 1990, São Paulo. São Paulo, IME/USP, 1990. 105 p.
- CAMARGO, M.B.P. de; ARRUDA, H.V.; PEDRO Jr.; M.J. *et al.* Melhores épocas de plantio de trigo no Estado de São Paulo baseadas na probabilidade de atendimento hídrico. **Bragantia**, Campinas (SP), v. 44, n. 1, p. 255-261. 1985.
- CAMARGO, M.B.P. de; ARRUDA, H.V.; PEDRO Jr.; M.J. *et al.*

- Probabilidade de atendimento da demanda hídrica da cultura do trigo pela precipitação pluvial no Estado de São Paulo. Campinas, IAC, 1988. 26 p. (IAC.Boletim Técnico, 120).
- CASTONGUAY, Y. & DUBÉ, P.A. Climatic analysis of a phenological zonation: a multivariate approach. *Agric. For. Meteorol.*, v. 35, p. 31-45. 1985.
- CATTELL, R.B. The scree test for the number of factors. *Multivar. Behav. Res.*, v. 1, p. 245-276. 1966.
- CEPAGRI/UNICAMP. Relatório das atividades desenvolvidas pelo projeto: Implantação de um banco de dados climatológicos como base a um sistema estadual de informações geográficas dirigido a zoneamento agrícola e monitoramento de fenômenos meteorológicos extremos. FAPESP(SP), 1988. 23 p.
- CHIDLEY, T.R.E. & KEYS, K.M. A rapid method of computing area rainfall. *J. of Hydrol.*, v. 12, n. 1, p. 15-24. 1970.
- CHRISTENSON, W.I. & BRYSON, R.A. An investigation of the potential of component analysis for weather classification. *Mon. Wea. Rev.*, v. 94, p. 697-709. 1966.
- CLIMANÁLISE - Boletim de Monitoramento e Análise Climática. *Climanálise*, São José dos Campos (SP), v. 4, n. 9, p. 1, set. 1989.
- CLIMANÁLISE - Boletim de Monitoramento e Análise Climática. *Climanálise*, São José dos Campos (SP), v. 6, n. 7, p. 3, jul. 1991.
- COHEN, S.J. Classification of 500mb height anomalies using obliquely rotated principal components. *J. Clim. Appl. Meteor.*, v. 22, p. 1975. 1983.
- CORMACK, R.M. A review of classification. *J. Roy. Statistical Soc., Series A*, v. 134, part 3, p. 321-367. 1971.

- CRADDOCK, J.M. & FLOOD, C.R. Eigenvectors for representing the 500mb geopotential surface over the Northern Hemisphere. *Quart. J. Roy. Meteor. Soc.*, v. 95, p. 576-593. 1969.
- DEMPSTER, A.P. Stepwise multivariate analysis of variance based on principal variables. *Biometrics*, v. 19, n. 3, p. 478-490, Sept. 1963.
- DINIZ, T.D. de A.S. Caracterização climática da Amazônia oriental. In: *Perspectivas sobre utilização e conservação do solo na Amazônia oriental*, 1986, Belém. *Anais...* Belém, EMBRAPA/CPATU-GTZ, 1986, p. 3-13.
- DYER, T.G.J. The assignment of rainfall stations into homogeneous groups: an application of principal component analysis. *Quart. J. Roy. Meteor. Soc.*, v. 101, p. 1005-1013. 1975.
- EVERITT, B. *Cluster Analysis*. London, Heinemann Educational Books, 1974. 122 p.
- EVERITT, B.S. Unresolved problems in cluster analysis. *Biometrics*, v. 35, n. 1, p. 169-182, Mar. 1979.
- FERNAU, M.E. & SAMSON, P.J. Use of cluster analysis to define periods of similar meteorology and precipitation chemistry in Eastern North America. Part I: transport patterns. *J. Appl. Meteor.*, v. 29, p. 735-750, Aug. 1990a.
- FERNAU, M.E. & SAMSON, P.J. Use of cluster analysis to define periods of similar meteorology and precipitation chemistry in Eastern North America. Part II: precipitation patterns and pollutant deposition. *J. Appl. Meteor.*, v. 29, p. 751-761, Aug. 1990b.
- FERREIRA, A.B. de H. *Dicionário Aurélio básico da língua portuguesa*.

- Rio de Janeiro, Nova Fronteira, 1989. 687 p.
- FLURY, B. & RIEDWYL, H. **Multivariate statistics: a practical approach.** London, Chapman and Hall, 1988. 296 p.
- GADGIL, S. & IYENGAR, R.N. Analysis of rainfall stations of the Indian peninsula. *Quart. J. Roy. Met. Soc.*, v. 106, p. 873-886. 1980.
- GADGIL, S. & JOSHI, N.V. Climatic clusters of the Indian region. *J. Climatol.*, v. 3, p. 47-63. 1983.
- GARAYALDE, E.J.G.; JUSTI da SILVA, M.G.A.; SANTOS, I.A. *et al.* Classificação meso-climática da região sul do Brasil pela análise de componentes principais. In: CONGR. BRAS. METEOROL., 4, 1986, Brasília. *Anais...* Brasília, SBMET, 1986. p. 119-124.
- GILMAN, D.L. apud: RICHMAN, M.B. Rotation of principal components. *J. Climatol.*, v. 6, p. 293-335. 1986.
- GNANADESIKAN, R. **Methods for statistical data analysis of multivariate observations.** USA, John Wiley & Sons, 1977. 311 p.
- GUIOT, J. Sur la détermination de régions climatiques quasi-homogènes. *Revue de Statistiques Appliquées*, v. 34, n. 2, p. 15-34. 1986.
- GUZMAN, A.G. & TORREZ, W.C. Daily rainfall probabilities: conditional upon prior occurrence and amount of rain. *J. Clim. Appl. Meteor.*, v. 24, n. 10, p. 1009-1014, Oct. 1985.
- HAKSTIAN, A.R.; ROGERS, W.T. & CATTELL, R.B. The behavior of number-of-factors rules with simulated data. *Multivar. Behav. Res.*, v. 17, p. 193. 1982.
- JOHNSON, R.A. & WICHERN, D.W. **Applied Multivariate Statistical Analysis.** New Jersey, Prentice-Hall, 1982. 594 p.

- JOLLIFFE, I.T. **Principal Component Analysis.** New York, Springer-Verlag, 1986. 271 p.
- JONGMAN, R.H.G. Ecological classification of the climate of the Rhine catchment. *Int. J. Biometeorol.*, v. 34, p. 194-203. 1990.
- JUSTI da SILVA, M.G.A.; CHIDA, I.O. & MANHAES, S.X. Análise climática conjugada temperatura-precipitação para o Estado do Rio de Janeiro. In: CONGR. BRAS. METEOROL., 5, 1988, Rio de Janeiro. *Anais...* Rio de Janeiro, SBMET, 1988. p. 23-28.
- KALKSTEIN, L.S.; TAN, G. & SKINDLOV, J.A. An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Clim. Appl. Meteor.*, v. 26, p. 717-730, June. 1987.
- KARL, T.R. & WILLIAMS Jr., C.N. An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Clim. Appl. Meteor.*, v. 26, p. 1744-1763, Dec. 1987.
- KSHIRSAGAR, A.M. **Multivariate Analysis.** New York, Marcell Dekker, 1972. 534 p.
- LYONS, S.W. Empirical orthogonal function analysis of Hawaiian rainfall. *J. Appl. Meteor.*, v. 21, p. 1713-1729, Nov. 1982.
- MARDIA, K.V.; KENT, J.T. & BIBBY, J.M. **Multivariate Analysis.** London, Academic Press, 1979. 521 p.
- MELLO, M.H. de A. & ALFONSI, R.R. O uso da reta dos mínimos quadrados para análise do comportamento pluviométrico. In: CONGR. BRAS. de AGROMETEOROLOGIA, 6, 1989, Maceió. *Anais...* Maceió, INPE-SBA-UFAL, 1989. p. 03-06.
- MOLION, L.C.B. & KOUSKY, V.E. Climatologia da dinâmica da troposfera sobre a Amazônia. In: Simpósio do Trópico Úmido, 1, 1986, Belém.

- Anais... Belém, 1986. p. 86-96.
- MOLION, L.C.B. & DALLAROSA, R.L.G. Pluviometria da Amazônia: são os dados confiáveis?. *Climanálise*, São José dos Campos (SP), v. 5, n. 3, p. 40-42, mar. 1990.
- MOLTENI, F.; BONELLI, P. & BACCI, P. Precipitation over Northern Italy: a description by means of principal component analysis. *J. Clim. Appl. Meteor.*, v. 22, p. 1738-1752. 1983.
- NASCIMENTO, C. & HOMMA, A. *Amazônia: Meio ambiente e tecnologia agrícola*. Belém, EMBRAPA/CPATU, 1984. 282 p.
- NASCIMENTO, C.M. & PEREIRA, M.A. de M.G. *Atlas climatológico do Estado de São Paulo (1977-1986)*. Campinas, Fundação Cargill, 1988. 93 p.
- OBREGON, G.O. & NOBRE, C.A. Principal Component Analysis of precipitation fields over the Amazon River Basin. *Climanálise*, São José dos Campos (SP), v. 5, n. 7, p. 35-46, jul. 1990.
- OMETTO, J.C. *Bioclimatologia vegetal*. São Paulo, Agronômica Ceres, 1981. 440 p.
- OVERLAND, J.E. & PREISENDORFER, R.W. A significance test for principal components applied to cyclone climatology. *Mon. Wea. Rev.*, v. 110, n. 1, p. 1-4, Jan. 1982.
- PATERSON, J.G.; GOODCHILD, N.A. & BOYD, W.J.R. Classifying environments for sampling purposes using a principal component analysis of climatic data. *Agric. Meteorol.*, v. 19, p. 349-362. 1978.
- PINTO, H.S.; ZULLO Jr., J. & ZULLO, S.A. Oscilações pluviométricas temporais no Estado de São Paulo. In: CONGR. BRAS. de AGROMETEOROLOGIA, 6, 1989, Maceió. Anais... Maceió, INPE-SBA-UFAL,

1989. p. 29-33.
- PREISENDORFER, R.W. & BARNETT, T.P. apud: RICHMAN, M.B. Rotation of principal components. *J. Climatol.*, v. 6, p. 293-335. 1986.
- RAO, C.R. The use and interpretation of principal component analysis in applied research. *Sankhyā, série A*, v. 26, p. 329-358, July. 1964.
- RESIO, D.T. & HAYDEN, B.P. Recent secular variations in mid-Atlantic winter extratropical storm climate. *J. Appl. Meteor.*, v. 14, p. 1223. 1975.
- RICHMAN, M.B. Rotation of principal components. *J. Climatol.*, v. 6, p. 293-335. 1986.
- SÁ, T.D. de A.; PINTO, H.S.; ZULLO, S.A. *et al.* Avaliação espacial e cartografia das chuvas de primavera no Estado de São Paulo através de técnicas de análise multivariada, como subsídio ao estabelecimento de cronogramas agrícolas regionais. In: CONGR. BRAS. de AGROMETEOROLOGIA, 6, 1989, Maceió. *Anais...* Maceió, INPE-SBA-UFAL, 1989. p. 24-28.
- SALVI, L.L. Tipologia climática do Estado de São Paulo segundo técnicas de quantificação. *R. Dep. Geogr.*, São Paulo (SP), p. 37-61. 1984.
- SANSIGOLO, C.A. Análise das precipitações diárias de Piracicaba, SP, visando planejamento agrícola. In: CONGR. BRAS. de AGROMETEOROLOGIA, 6, 1989, Maceió. *Anais...* Maceió, INPE-SBA-UFAL, 1989. p. 224-231.
- SÃO PAULO, CESP/CPFL. Irrigação. Dados climatológicos auxiliares para o projeto de sistema de irrigação. São Paulo, CESP/CPFL, 1979. 12 p.
- SÃO PAULO, DAEE. Atlas pluviométrico do Estado de São Paulo (período 1941-1970). São Paulo, Secretaria de Serviços de Obras Públicas -

- DAEE, 1972. 50 p.
- SÃO PAULO.DAEE. **Dados pluviométricos da rede básica do Estado de São Paulo - Ano: 1972.** São Paulo, Secretaria de Serviços de Obras Públicas - DAEE, 1974. 860 p. (Boletim Pluviométrico, 13).
- SÃO PAULO.SEADE. **Anuário Estatístico do Estado de São Paulo 1987.** São Paulo, SEADE, 1988. 538 p.
- SAS INSTITUTE INC. **SAS User's Guide: Statistics, version 5 Edition.** Cary (NC), SAS Inst., 1985. p. 45-69. SAS Clustering Procedures.
- SCHRODER, R. **Contribuição e curso anual das precipitações no Estado de São Paulo.** Bragantia, Campinas (SP), v. 15, n. 18, p. 193-249. 1956.
- SETZER, J. **Contribuição para o estudo da chuva do Estado de São Paulo.** São Paulo, Escola Profissional Salesiana, 1946. 239 p.
- SETZER, J. **Atlas climático e ecológico do Estado de São Paulo.** Comissão Interestadual das Bacias do Paraná-Uruguai - CESP, São Paulo, 1966. 60 p.
- STERN, R.D. **Computing a probability distribution for the start of the rains from a Markov chain model for precipitation.** J. Appl. Meteor., v.21, p. 420-423, Mar. 1982.
- TARIFA, J.R. **Fluxos polares e chuvas de primavera-verão no Estado de São Paulo, uma análise quantitativa do processo genético.** São Paulo, Depto. de Geografia - USP, 1974. 88 p. (Tese de Doutorado).
- VIANELLO, R.L. & ALVES, A.R. **Meteorologia básica e aplicações.** Viçosa (MG), UFV, 1991. 449 p.
- VIVALDI, L.J. **Utilização da distribuição gama em dados pluviométricos.** Piracicaba, ESALQ, 1973. 71 p. (Tese de Mestrado).
- WEI, T.C. & McGUINNESS, J.L. **Reciprocal distance squared method: a**

- computer technique for estimating areal precipitation. U.S. Department of Agriculture, Agriculture Research Service, 1973. 29 p.
- WILLMOTT, C.J. apud: KALKSTEIN, L.S.; TAN, G. & SKINDLOV, J.A. An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Clim. Appl. Meteor.*, v. 26, p. 717-730, June. 1987.
- ZULLO, S.A.; HOTTA, L.K.; SÁ, T.D. de A. *et al.* Precipitação no nordeste do Estado do Pará: uma descrição através de análise multivariada. In: Encontro de Análise Multivariada e Aplicações, 1, 1989, Campinas. Não publicado.
- ZULLO, S.A.; PINTO, H.S.; HOTTA, L.K. *et al.* Determinação de áreas pluviométricas homogêneas no Estado de São Paulo (BR) utilizando técnicas de análise multivariada. In: Curso Latino-Americano Sobre Variabilidad Climática Y Su Impacto Ambiental, 2, 1991, Santiago. *Resumos...* Santiago, 1991, p. 1.
- ZWICK, W.P. & VELICER, W.F. Factors influencing four rules for determining the number of components to retain. *Multivar. Behav. Res.*, v. 17, p. 253. 1982.
- ZWIERS, F.W. Statistical considerations for climate experiments. Part II : multivariate tests. *J. Clim. Appl. Meteor.*, v. 26, p. 477-487, Apr. 1987.