

IDENTIFICAÇÃO DE REGIÕES HIDROLOGICAMENTE HOMOGÊNEAS ATRAVÉS DE ANÁLISE FATORIAL DE COMPONENTES PRINCIPAIS

Orlando M. Strobel *

RESUMO

É utilizada a análise fatorial de componentes principais para delimitar sub-regiões homogêneas quanto à precipitação, onde são analisados os aspectos conjunto de simultaneidade de ocorrência, altura e variabilidade pluvial. A análise de componentes principais permite uma diminuição da dimensionalidade do problema, possibilitando estabelecer-se correlações entre as variáveis e estes componentes de tal forma que se torna possível identificar regiões semelhantes. A metodologia foi aplicada sobre uma matriz de correlação interestações relativa aos dados pluviométricos diários de 40 postos da 7.^a zona hidrográfica de São Paulo (48.000km²). As séries de 39 anos de dados de cada posto foram particionadas em 4 séries trimestrais e uma semestral (úmida). Como resultado, obteve-se 5 campos pluviais homogêneos, os quais não coincidem necessariamente com os limites físicos das sub-bacias hidrográficas.

ABSTRACT

The factor analysis of principal components is done to delimitate the homogeneous subregions in relation to the precipitation where the entire aspects of simultaneity of occurrences, pluvial variability and height are analysed. The analysis of the principal components makes the decrease of the extension of the problem possible, and establishes the correlation between the variables and these components in such a way that it becomes possible to identify similar regions. The methodology was applied in a matrix of interstation correlation related to the daily pluvial data of 40 stations of the 7th hydrographic zone of São Paulo State (48.000km²). The series of 39 years of data from each station were divided in four three-monthly series and one wet half year. As result, five homogeneous pluvial fields were obtained which are not coincident with the physical limits of the hydrographic subbasin.

INTRODUÇÃO

A identificação de regiões hidrologicamente semelhantes é um problema com o qual o hidrólogo freqüentemente se depara, mormente quando está estudando uma região com dados escassos e necessita apoiar suas decisões nas similaridades existentes com outra região mais rica em dados.

A análise deveria ser feita apenas com base na variável hidrológica em questão (vazão, evaporação, precipitação, etc.), pois duas regiões distintas podem ser consideradas homogêneas quanto a uma variável, mas não o ser com respeito à outra. Entretanto, a escassez de dados não permite uma confrontação direta, motivo pelo qual, nos laboriosos e exaustivos estudos de regionalização, utilizam-se relações físicas existentes entre a variável em questão e outras mais facilmente obtíveis. Este tipo de análise pode, eventualmente, apresentar resultados bem afastados daquele considerado "correto", mas decisões, com base nestes estudos, são certamente mais adequadas do que aqueles, comumente observadas no meio técnico, que se baseiam em perigosas extrapolações apoiadas no conceito subjetivo do "jeitão" das regiões envolvidas.

Outras vezes o hidrólogo necessita definir sub-regiões homogêneas com finalidade de implementar algum tipo de modelo matemático hidrológico e, neste caso, normalmente dispõe-se de um grande número de dados que devem ser debulhados ordeiramente. Esta análise conjunta dos dados, feita sobre o intervalo de tempo adotado, deve considerar os elementos estatísticos e circunstanciais considerados relevantes para a modelagem em questão.

No caso de precipitação pluvial, o que muito se tem feito é subdividir a região em conformidade com os limites físicos da sub-bacia, considerar a distribuição espacial dentro desta como sendo homogênea e, eventualmente, agregar sub-bacias segundo algum critério pessoal, por exemplo, em função de isoietas anuais, da declividade média, orientação de vertentes, etc.

Entretanto, a precipitação é muito mais decorrente de causas atmosféricas, como a circulação geral de massas, o campo de pressões, etc, do que terrenas, sendo que nestas, as mais relevantes são a altitude e orografia. Assim, a cômoda subdivisão por bacias hidrográficas não é necessariamente real. Desta forma, considerar homogênea a distribuição espacial da precipitação dentro duma mesma sub-bacia pode levar a erros grosseiros. O mesmo se pode afirmar com respeito a técnicas de completar dados faltantes em um posto com base nos "p" postos mais próximos, postos estes pertencentes ou não à sub-bacia, mesmo em se adotando um critério de pesos por área de influência.

* Mestre em Engenharia Hidráulica pela Escola Politécnica da Universidade de São Paulo. Professor do Centro Federal de Educação Tecnológica do Paraná, Universidade Federal do Paraná e Pontifícia Universidade Católica do Paraná.

A análise de homogeneidade deve, pois, ser baseada em elementos parcialmente independentes daqueles físicos da bacia. Assim, procura-se apoiar tal análise sobre características intrínsecas da chuva, como altura de precipitação, variabilidade e simultaneidade de ocorrência, já que apenas duas destas características não são suficientes para definir o campo pluvial.

Uma maneira de se efetuar a identificação de regiões climatologicamente homogêneas pode ser feita através de uma técnica estatística multivariada, chamada análise fatorial, ou, mais propriamente, de uma pequena modificação desta, chamada análise de componentes principais. Esta análise tem como propósito explicar relações observadas entre várias variáveis em termos de relações mais simples, as quais proporcionam uma melhor visualização da estrutura básica destas. No presente trabalho, as variáveis são a altura de precipitação nos "p" postos da região em estudo.

Apesar de ser um instrumento estatístico conhecido, já há algum tempo, as primeiras aplicações em hidrologia foram feitas por Snyder (1962), Wallis (1965) e Wong (1968) a relativamente pouco tempo atrás, onde a preocupação básica era o estabelecimento de relações entre vazão e elementos fisiográficos da bacia. Haan (1977) analisou regressões sobre esta metodologia a fim de obter um modelo preditivo.

METODOLOGIA

Os dados diários de chuva coletados ao longo dos anos (n observações) em cada um dos postos (p variáveis) de uma região são quase sempre correlacionáveis a algum grau. Esta correlação indica que alguma informação contida em uma variável (posto) está também contida em outras p-1 variáveis.

O objetivo da análise de componentes principais é o de transformar as "p" variáveis originais correlacionadas, em "q" variáveis não-correlacionadas ($q < p$) ou seja, em "q" componentes principais que são funções lineares das variáveis originais. Assim, se as observações estão num sistema caracterizado por um espaço p-dimensional, ocorre uma transformação deste para outro, q-dimensional ($q < p$), transformação esta realizada de tal forma que a projeção das observações originais no novo sistema seja nula (projeção ortogonal). Desta forma, o eixo original (variável p) é transformado no novo eixo (componente principal), o qual é estatisticamente independente, pois cada componente principal, por sua vez, é ortogonal ao seguinte. Quando $p = q$, surge o caso, mais geral, da chamada análise fatorial. Matematicamente tem-se:

$$\left\{ \begin{array}{l} \underline{Z} = \underline{X} \cdot \underline{A} \end{array} \right. \quad (1)$$

onde:

$$\left\{ \begin{array}{l} \underline{X} = \text{matriz } n \times p \text{ de } n \text{ observações normalizadas (altura diária de chuva) de } p \text{ postos da} \\ \text{Região;} \\ \underline{A} = \text{matriz } q \times p \text{ de coeficientes que definem a transformação linear;} \\ \underline{Z} = \text{matriz } n \times q \text{ de elementos dos componentes principais.} \end{array} \right.$$

Trata-se, pois, de reduzir um problema multidimensional de "p" para "q" variáveis de tal forma que ainda se tenha uma boa representatividade do problema original e que permita chegar-se a conclusões sobre o sistema. Deve-se, então, construir Z de tal forma que cada componente z_j (vetor coluna $n \times 1$) explique a máxima variância residual de X não explicada pelos prévios "j - 1" componentes. Assegura-se desta forma que os primeiros "q" componentes explicam a maior parte da variabilidade do sistema.

Não é possível a visualização dos componentes principais e sua posição relativa às variáveis originais, o que é comum em todas as transformações de dimensão no hiperespaço, uma vez que são obtidos através de abstratos conceitos estatísticos e matemáticos. Apresentaremos algumas características desta metodologia, sendo que informações adicionais podem ser encontradas em Harman (1976) e Haan (1977). O algoritmo utilizado se apóia basicamente no cálculo matricial e nas propriedades da covariância estatística. As etapas seriam:

Obtenção da matriz inicial

Apesar do problema poder ser resolvido com os dados "in natura", formando uma matriz $n \times p$, recomenda-se utilizar como matriz básica a matriz R ($p \times p$) das correlações entre as variáveis normalizadas. Obviamente, a matriz R é simétrica ($r_{j,k} = r_{k,j}$), positiva semidefinida (a diagonal contém a unidade) e, portanto, é dita matriz Gramiana. Observe-se que, trabalhando sobre uma matriz nestas condições, assegura-se a inclusão das três características pluviais julgadas relevantes.

Obtenção dos autovetores

Trata-se de obter a solução da equação:

$$\left\{ \begin{array}{l} |R - \lambda_j I| = 0 \\ j = 1, 2, \dots, p \end{array} \right. \quad (2)$$

onde:

$$\left\{ \begin{array}{l} I = \text{matriz identidade.} \\ \lambda_j = \text{raízes características, mais comumente chamadas de autovalores.} \\ \text{Em essência, é um operador Lagrangeano introduzido num problema de maximização da variância.} \end{array} \right.$$

Para $j = 1$ (primeira variável) teríamos:

$$\left| \begin{array}{cccccc} (r_{1,1} - \lambda_1) & r_{1,2} & r_{1,3} & \dots & r_{1,p} \\ r_{2,1} & (r_{2,2} - \lambda_1) & r_{2,3} & \dots & r_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p,1} & r_{p,2} & r_{p,3} & \dots & (r_{p,p} - \lambda_1) \end{array} \right| = 0$$

Resolvendo-se, surgem p soluções para λ_1 . Como a maximização da variância está embutida na equação (2), adota-se o maior dos "p" valores de λ_1 . Repete-se o exposto e determina-se todos os autovalores que formam o autovetor λ_j .

Obtenção da matriz de transformação

Na dedução das equações desta metodologia, para assegurar que o componente principal seja ortogonal e que sucessivamente vá explicando o máximo da variância residual, chega-se à expressão:

$$\left\{ \begin{array}{l} (R - \lambda_j I) a_j = 0 \\ j = 1, 2, \dots, p \end{array} \right. \quad (3)$$

Utilizando o resultado da etapa anterior, forma-se um sistema de equações, somando à restrição de ortogonalidade de matrizes $(a^j \cdot a_j = 1)$ que é equivalente a:

$$\left\{ \begin{array}{l} a_j = \sum_{k=1}^p a_{k,j}^2 = 1 \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{array} \right. \quad (4)$$

o que assegura independência de componentes principais. (Se a análise fatorial, j iria até p).

Obtenção do valor numérico dos componentes principais

Conforme equação (1), teríamos:

$$\left\{ \begin{array}{l} z_{ij} = \sum_{k=1}^p \left[\frac{x_{i,k} - \bar{x}_k}{S_k} \right] a_{k,j} \\ i = 1, 2, \dots, x \\ j = 1, 2, \dots, q \\ k = 1, 2, \dots, p \end{array} \right. \quad (5)$$

onde:

$$\left\{ \begin{array}{l} x_{i,k} \text{ é uma observação qualquer,} \\ \bar{x}_k \text{ é a média destas para o posto em questão,} \\ S_k \text{ é o desvio padrão.} \end{array} \right.$$

Obtenção da matriz de correlação entre as variáveis originais e os componentes principais

É obtida por:

$$\left\{ \begin{array}{l} \text{COR}(x_k, z_j) = \frac{\lambda_j^{1/2} a_{k,j}}{S_k} \end{array} \right. \quad \begin{array}{l} j = 1, 2, \dots, q \\ k = 1, 2, \dots, p \end{array} \quad (6)$$

Assim, estabelece-se uma matriz dita matriz de fatores de peso (não rotacionados), sobre a qual pode-se tentar estabelecer alguma significância física, pois um alto fator deve indicar uma alta correlação com a variável em questão, ou indicar uma forte similaridade linear entre a componente e esta variável, sugerindo que ambas são medidas de algum atributo hidrológico comum.

Rotação Ortogonal

A matriz obtida na etapa anterior é, na maioria das vezes, de difícil análise, pois uma variável pode apresentar uma alta correlação com mais de um componente principal. Para uma maior facilidade de análise, seria desejável que uma variável qualquer tivesse uma alta correlação com um componente principal e baixa correlação com os componentes restantes, ou seja, apresentar uma estrutura fatorial mais simples.

Faz-se então uma transformação ortogonal desta matriz de fatores de peso através de uma rotação, o que faz mudar as correlações entre as variáveis e componentes, mantendo contudo, a independência entre componentes no hiperespaço.

Existem infinitas maneiras de se efetuar este procedimento. Entretanto, no sentido de obter-se mais facilmente esta estrutura fatorial mais simples, estabeleceram-se algumas regras empíricas sobre as quais têm sido propostos vários esquemas de rotação, sendo que o mais utilizado é o chamado esquema de rotação varimax. Este esquema é mais apropriado em estudos de homogeneidade (agrupamento de variáveis), pois procura fazer com que haja uma maximização da variância do quadrado dos fatores em cada coluna, ou seja, de cada componente principal. Matematicamente, o esquema é:

$$\left\{ \begin{array}{l} \text{Max } p \sum_{k=1}^p \sum_{j=1}^q \left(\frac{a_{k,j}}{h_j} \right)^4 - \sum_{k=1}^p \left(\sum_{j=1}^q \frac{a_{k,j}^2}{h_j^2} \right)^2 \end{array} \right. \quad (7)$$

onde os $a_{k,j}/h_j$ são os fatores normalizados e h_j o desvio padrão. O termo h_j^2 (variância) também é chamado de peso comum global da variável.

O esquema de rotação acima descrito pode ser exposto graficamente num problema genérico que envolva 2 componentes principais e 7 variáveis, conforme exposto na figura 1. Por condições implícitas

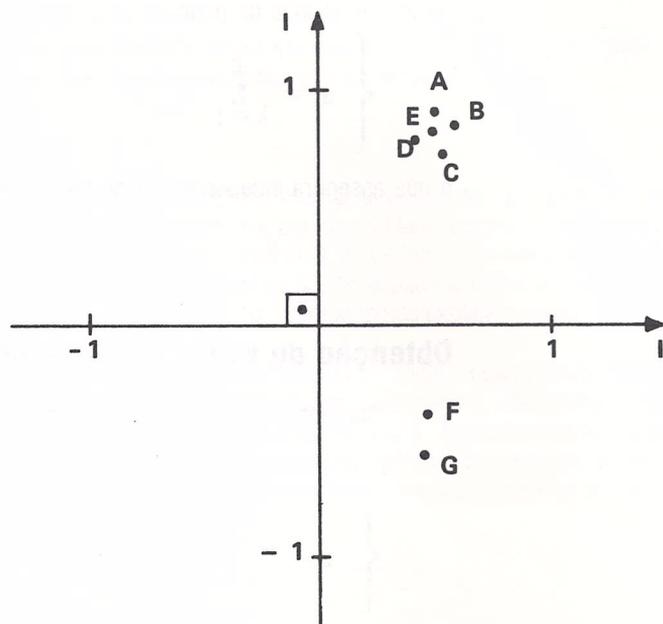


FIGURA 1. Eixos não rotacionados.

na metodologia, os fatores assumem valores entre -1 e +1. Os pontos surgem no plano componente principal I x componente principal II, através de projeções abstratas do hiperespaço de dimensão 7. Nota-se que as variáveis se agregam e que esta agregação pode ser analiticamente observável através da existência de fatores da mesma ordem de grandeza. Assim, as variáveis A, B, C, D e E estão agrega-

das distintamente das variáveis F e G. Contudo fica difícil relacionar aquelas primeiras 5 variáveis com um dos componentes, já que estão numa condição bissetriz, ou seja, com fatores próximos a 0,5 em ambos os eixos, impossibilitando uma análise mais completa. A realização de uma rotação ortogonal deste sistema de eixos faz com que ambos os grupos apresentem um maior valor segundo um componente principal e menor segundo o outro, como mostrado na figura 2.

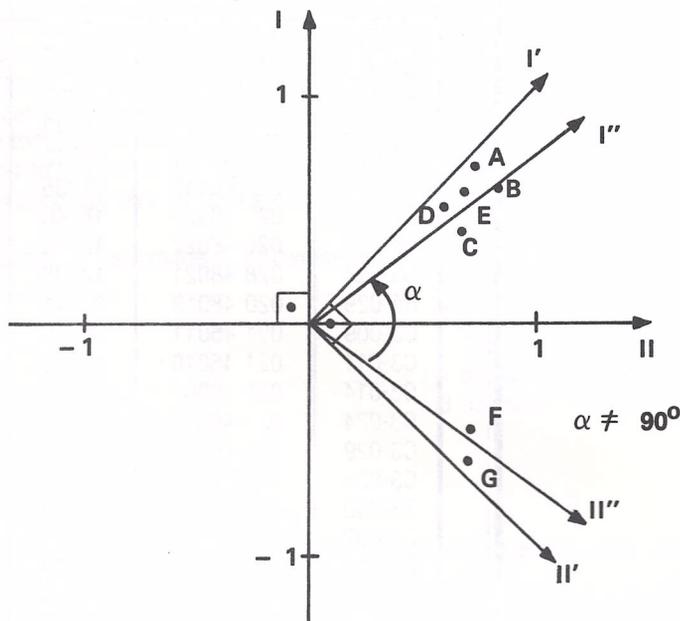


FIGURA 2. Rotação de eixos (' = ortogonal rotacionado, '' = oblíqua).

Rotação Oblíqua

Em algumas aplicações desta metodologia [ver Matalas e Reiher (1967)] observou-se que a ortogonalidade entre componentes não era necessariamente válida ou exigível, propondo-se, então, esquemas de rotação oblíqua que permitem uma melhor agregação de variáveis. O esquema genérico desta rotação oblíqua é o de minimizar os produtos cruzados dos fatores de peso em relação ao eixo (componente principal) mais próximo. Observando a figura 2, é fácil concluir que os fatores seriam maximizados com relação a um componente e minimizados em relação ao outro, para cada variável em questão. (I'' e II'' são os eixos oblíquos).

Análise da Matriz Final

A matriz final é uma matriz de fatores de peso $p \times q$. Para cada p variável observa-se qual é o máximo fator, já que, dentro do contexto desta análise, a variável teria maior afinidade com o componente principal no qual tivesse um alto fator. Assim tem-se, para cada componente principal, um grupo de variáveis agregadas, ou seja, que presumivelmente possuem elementos em comum.

CASO ESTUDADO

Aplicou-se a metodologia descrita a dados pluviométricos de 40 postos da 7ª Zona Hidrográfica do Estado de São Paulo constantes da tabela 1, postos estes espalhados por uma área de aproximadamente 48.000m², conforme figura 3. Nesta figura foram somente analisados os postos ressaltados por '•', já que os outros são postos com instalação bem mais recente.

TABELA 1.
Postos da 7ª Zona Hidrográfica de São Paulo

PREFIXO DO POSTO		DATA DE INSTALAÇÃO	LATITUDE	LONGITUDE	ALTITUDE	NOME DO POSTO
DAEE-SP	DNAEE					
B4-001	020 47017	05/35	20° 31'	47° 24'	1004	Franca
B4-002	020 47008	09/31	20° 12'	47° 43'	865	Buritzal
B4-003	020 47021	09/31	20° 39'	47° 41'	620	Usina Dourados
B4-005	020 47028	01/31	20° 50'	47° 18'	755	Usina Esmeril

TABELA 1. (Continuação)
Postos da 7ª Zona Hidrográfica de São Paulo.

PREFIXO DO POSTO		DATA DE	LATITUDE	LONGITUDE	ALTI- TUDE	NOME DO POSTO
DAEE-SP	DNAEE	INSTALAÇÃO				
B4-012	020 47027	09/40	20° 48'	47° 46'	730	Fazenda Conquista
B4-015	020 47025	05/37	20° 44'	47° 53'	735	Orlândia
B4-018	020 47018	08/37	20° 32'	47° 59'	600	Faz. Sta. Cecília
B4-026	020 47002	05/42	20° 02'	47° 45'	578	Igarapava
B4-029	020 47001	05/42	20° 00'	47° 49'	670	Faz. S. Geraldo
B5-002	020 48015	01/36	20° 34'	48° 34'	518	Barretos
B5-004	020 48023	18/40	20° 44'	48° 04'	540	Morro Agudo
B5-024	020 48027	12/40	20° 49'	48° 29'	567	Faz. Marambaia
B5-028	028 48021	12/38	20° 43'	48° 33'	610	Colina
B5-029	020 48019	05/41	20° 41'	48° 25'	530	Jaborandi
C3-009	021 46011	08/36	21° 50'	46° 54'	690	V. Gde. do Sul
C3-011	021 46010	11/36	21° 43'	46° 49'	1000	S. S. da Gama
C3-014	021 46007	10/37	21° 32'	46° 39'	884	Caconde
C3-024	021 46004	04/40	21° 28'	46° 49'	825	Itaquara
C3-029	021 46015	04/40	21° 32'	46° 48'	740	Fazenda Guaxupé
C3-034	021 46012	11/42	21° 56'	46° 43'	823	A. da Prata
C3-040	021 46003	04/40	21° 26'	46° 51'	860	Fazenda Açude
C4-007	021 47014	11/36	21° 21'	44° 47'	805	Cravinhos
C4-021	021 47037	02/39	21° 57'	47° 42'	826	Faz. M. Alegre
C4-029	021 47033	10/36	21° 50'	47° 15'	644	S. C. das Palmeiras
C4-033	021 47096	08/39	22° 01'	47° 26'	696	Pirassununga
C4-034	021 47006	07/39	21° 13'	47° 50'	518	R. Preto
C4-040	021 47023	06/40	21° 28'	47° 01'	665	Mococa
C4-043	021 47123	06/42	21° 26'	47° 20'	605	Usina Amália
C4-052	021 47094	09/42	21° 22'	47° 26'	590	F. S. Lourenço
C5-009	021 48029	08/36	21° 31'	48° 09'	604	Mutuca
C5-020	021 48028	01/36	21° 30'	48° 02'	506	Guataparã
C5-040	021 48001	04/41	21° 01'	48° 13'	503	Pitangueiras
D3-003	022 46007	09/31	22° 17'	46° 45'	710	Usina Pinhal
D3-012	022 46019	11/36	22° 36'	46° 42'	920	Serra Negra
D3-015	022 46013	11/36	22° 27'	46° 49'	630	Itapira
D3-022	022 46002	08/39	22° 03'	46° 49'	660	Aguai
D4-027	022 47017	10/36	22° 18'	47° 27'	713	Faz. Belmonte
D4-029	022 47011	10/36	22° 15'	47° 27'	635	Faz. Santana
D4-030	022 47005	10/36	22° 10'	47° 17'	575	Cresciumal
D4-032	022 47003	10/36	22° 07'	47° 25'	605	Faz. Bonanza

OBS.: Altitude em metros.

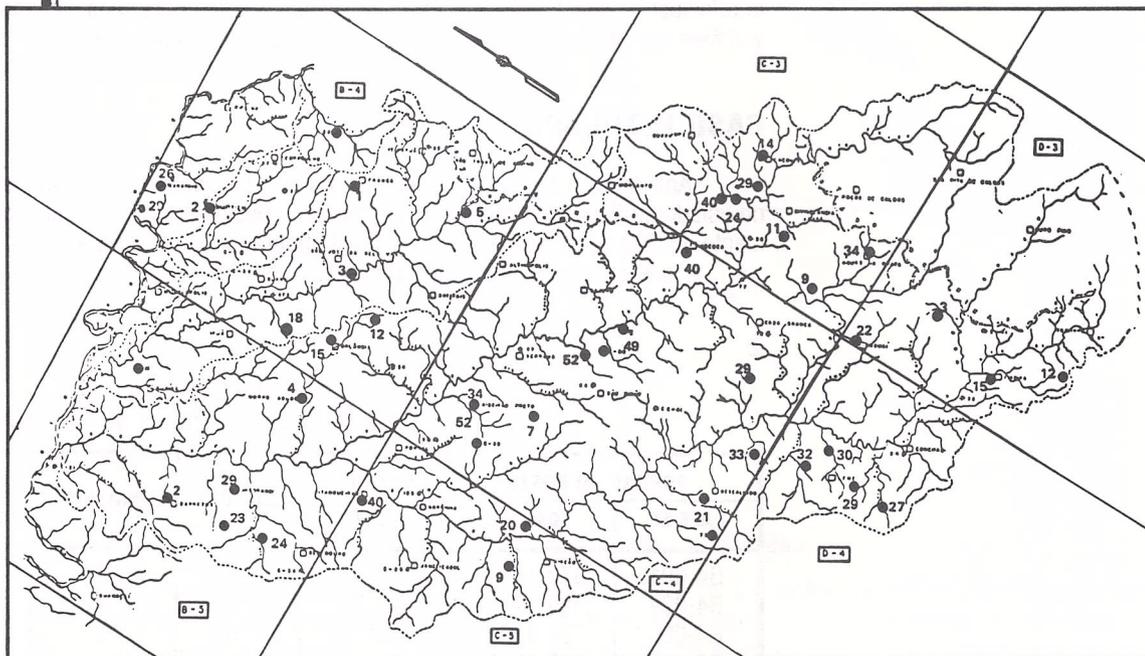


FIGURA 3. Postos analisados da 7ª Zona Hidrográfica de São Paulo.

A série histórica analisada foi de 39 anos (1943 — 1981). Para verificar-se a variabilidade da homogeneidade ao longo dos meses do ano, partiu-se a série em outras, trimestrais (jan-mar; abr-jun; jul-set e out-dez), e uma do semestre úmido (out-mar); assim, por exemplo, a série do 1.º trimestre (jan-mar) conteria todos os 1.ºs trimestres de todos os anos e, como se trata de uma análise multivariada, de todas as 40 estações em questão.

Inicialmente, para cada série, foram eliminados todos os dias em que ocorreram dados faltantes em pelo menos um dos 40 postos, já que o elemento simultaneidade de ocorrência é importante e não poderia ser suprido por técnicas usuais. Mesmo que pudesse ser suprido pelas técnicas usuais, não seria um procedimento correto já que se está, justamente, querendo estabelecer sobre quais variáveis (postos) deve-se tomar uma decisão.

Também, em todas as séries, foram eliminados os dias em que não choveu na bacia, ou que o montante precipitado não tenha sido maior que arbitrários 0,3mm. Isto porque o estudo é de chuvas e a permanência de dias secos na série afetaria a estrutura de correlação inter-estações (a correlação aumentaria).

Depurados os dados desta maneira, obteve-se as seguintes séries:

1.º trimestre	1.937 dias
2.º trimestre	1.147 dias
3.º trimestre	743 dias
4.º trimestre	1.645 dias
Semestre úmido	3.557 dias

RESULTADOS

Para cada uma destas séries estabeleceu-se a matriz de correlação inter-estações, ou seja, obteve-se uma matriz R(40x40).

Conforme sugestão de Morrison (1967), adotou-se uma análise sobre 5 componentes principais; conseqüentemente haverá uma subdivisão da região em estudo em 5 sub-regiões homogêneas. Obteve-se, então, os autovalores e a primeira matriz de fatores de peso (não rotacionados). Na tabela 2 é exposta esta matriz para o 1.º trimestre, com os respectivos autovalores. Conforme exposto anteriormente, o primeiro componente possui o maior autovalor ($\lambda_1 = 12,34$) e, conseqüentemente, explica o máximo da variância ($12,34 \times 100/40 \cong 31\%$). Também a variância de um posto explicada por um componente principal é o quadrado do fator de peso respectivo. Assim, 14,6% ($= 0,382^2 \times 100$) da variância do 6.º posto é explicada pelo 2.º componente. Desta forma, a % da variância deste posto explicada por todos os 5 componentes é:

$$\left\{ h_6^2 = [(-0,603)^2 + 0,382^2 + 0,094^2 + 0,092^2 + 0,069^2] \times 100 \cong 53\% \right.$$

TABELA 2.
Matriz dos Fatores de Peso (não-rotacionada) — 1.º Trimestre

VARIÁVEL		COMPONENTES PRINCIPAIS				
N.º	POSTO	I	II	III	IV	V
1	B4-29	- 0.566	0.379	0.147	0.227	0.361
2	B4-26	- 0.529	0.353	0.159	0.269	0.356
3	B4-02	- 0.573	0.344	0.098	0.218	0.315
4	B4-01	- 0.576	0.266	0.056	0.154	0.156
5	B4-18	- 0.596	0.423	0.012	0.017	0.054
6	B4-03	- 0.503	0.382	0.094	0.092	0.069
7	B4-15	- 0.373	- 0.088	- 0.503	0.148	0.139
8	B4-12	- 0.383	0.317	0.017	- 0.009	0.104
9	B4-05	- 0.624	0.226	0.133	0.173	- 0.118
10	B5-02	- 0.513	0.328	- 0.138	- 0.280	0.034
11	B5-29	- 0.500	0.195	- 0.215	- 0.221	- 0.030
12	B5-28	- 0.592	0.331	- 0.131	- 0.319	- 0.120
13	B5-04	- 0.455	0.093	- 0.395	- 0.054	0.133
14	B5-24	- 0.504	0.234	- 0.134	- 0.327	- 0.121
15	C3-40	- 0.685	0.018	0.197	0.241	- 0.289
16	C3-24	- 0.508	- 0.244	- 0.390	0.299	- 0.074
17	C3-14	- 0.616	- 0.042	0.152	0.332	- 0.311
18	C3-11	- 0.518	- 0.113	0.019	0.292	- 0.381
19	C3-09	- 0.434	- 0.172	- 0.001	0.172	0.077
20	C3-34	- 0.442	- 0.233	0.032	0.204	- 0.023
21	C3-29	- 0.528	- 0.147	- 0.263	0.291	- 0.188

TABELA 2. (Continuação)
Matriz dos Fatores de Peso (não-rotacionada) — 1º Trimestre

VARIÁVEL		COMPONENTES PRINCIPAIS				
Nº	POSTO	I	II	III	IV	V
22	C4-34	- 0.607	0.146	- 0.045	- 0.146	- 0.158
23	C4-07	- 0.578	0.040	- 0.158	- 0.105	- 0.188
24	C4-40	- 0.670	0.000	0.222	0.206	- 0.215
25	C4-29	- 0.629	- 0.218	0.149	- 0.066	- 0.034
26	C4-21	- 0.528	- 0.313	- 0.096	- 0.108	0.169
27	C4-52	- 0.520	- 0.176	- 0.526	0.110	0.053
28	C4-33	- 0.632	- 0.247	0.210	- 0.185	0.075
29	C4-43	- 0.496	- 0.231	- 0.526	0.141	0.061
30	C5-40	- 0.606	0.255	- 0.023	- 0.312	- 0.190
31	C5-20	- 0.591	0.047	0.051	- 0.292	- 0.194
32	C5-09	- 0.613	0.006	0.022	- 0.283	- 0.169
33	D3-22	- 0.604	- 0.282	0.276	- 0.039	- 0.001
34	D3-03	- 0.591	- 0.310	0.286	0.006	0.045
35	D3-15	- 0.549	- 0.318	0.228	- 0.009	0.041
36	D3-12	- 0.470	- 0.289	0.090	- 0.032	0.236
37	D4-32	- 0.573	- 0.321	0.142	- 0.193	0.132
38	D4-30	- 0.498	- 0.355	- 0.001	- 0.196	0.173
39	D4-29	- 0.555	- 0.327	0.245	- 0.180	0.148
40	D4-27	- 0.423	- 0.304	- 0.136	- 0.198	0.230
Autovalor		12,340	2,607	1,904	1,662	1,285
% variância total		30,85	6,52	4,76	4,16	3,21

OBS.: A ordenação dos postos seguiu a ordem crescente do código do DNAEE.

valor este chamado de peso comum global da variável. Observe-se que, neste caso, mais de 50% são explicados somente pelos 2 primeiros componentes. Na mesma tabela 2, nota-se que o autovalor é decorrente da soma do quadrado dos fatores de peso, ou seja, para o 1º componente principal:

$$\lambda_1 = (-0,782)^2 + (-0,765)^2 + \dots + (-0,049)^2 = 12,34$$

Em se analisando o maior fator de peso para cada posto, verifica-se que alguns destes não estão bem definidos. Efetuou-se, então, uma rotação ortogonal varimax (tabela 3 — 1º trimestre), mas ainda persistem algumas dúvidas já que, por exemplo, o 9º posto tanto pode ser relacionado com o 1º como com o 5º componente principal. Assim realizou-se uma rotação oblíqua, cujos fatores de peso para o 1º trimestre estão representados na tabela 4, numa matriz padrão sobre a qual a conclusão fica simplificada.

TABELA 3.
Matriz dos Fatores de Peso — 1º Trimestre, após rotação ortogonal

VARIÁVEL		COMPONENTES PRINCIPAIS				
Nº	POSTO	I	II	III	IV	V
1	B4-29	- 0.782	- 0.141	- 0.078	- 0.131	- 0.115
2	B4-26	- 0.765	- 0.128	- 0.072	- 0.074	- 0.130
3	B4-02	- 0.727	- 0.137	- 0.123	- 0.150	- 0.136
4	B4-01	- 0.565	- 0.144	- 0.145	- 0.221	- 0.210
5	B4-18	- 0.549	- 0.050	- 0.012	- 0.096	- 0.194
6	B4-03	- 0.580	- 0.091	- 0.063	- 0.346	- 0.250
7	B4-15	- 0.150	- 0.051	- 0.638	- 0.093	- 0.014
8	B4-12	- 0.502	- 0.136	- 0.116	- 0.380	- 0.151
9	B4-05	- 0.434	- 0.134	- 0.078	- 0.289	- 0.453
10	B5-02	- 0.311	- 0.099	- 0.140	- 0.586	0.012
11	B5-29	- 0.200	- 0.110	- 0.244	- 0.519	- 0.049
12	B5-28	- 0.251	- 0.105	- 0.131	- 0.699	- 0.110
13	B5-04	- 0.235	- 0.081	- 0.473	- 0.325	0.036
14	B5-24	- 0.152	- 0.123	- 0.127	- 0.822	- 0.078
15	C3-40	- 0.287	- 0.248	- 0.100	- 0.226	- 0.666
16	C3-24	- 0.081	- 0.150	- 0.647	- 0.037	- 0.339
17	C3-14	- 0.233	- 0.199	- 0.151	- 0.123	- 0.692

TABELA 3. (Continuação)
Matriz dos Fatores de Peso — 1º Trimestre, após rotação ortogonal

VARIÁVEL		COMPONENTES PRINCIPAIS				
Nº	POSTO	I	II	III	IV	V
18	C3-11	- 0.069	- 0.136	- 0.228	- 0.123	- 0.649
19	C3-09	- 0.203	- 0.299	- 0.259	- 0.009	- 0.235
20	C3-34	- 0.140	- 0.308	- 0.244	0.023	- 0.337
21	C3-29	- 0.105	- 0.114	- 0.503	- 0.106	- 0.448
22	C4-34	- 0.211	- 0.197	- 0.156	- 0.508	- 0.270
23	C4-07	- 0.113	- 0.165	- 0.278	- 0.453	- 0.279
24	C4-40	- 0.303	- 0.295	- 0.079	- 0.205	- 0.600
25	C4-29	- 0.149	- 0.526	- 0.140	- 0.227	- 0.317
26	C4-21	- 0.098	- 0.517	- 0.352	- 0.149	- 0.055
27	C4-52	- 0.100	- 0.159	- 0.714	- 0.183	- 0.119
28	C4-33	- 0.164	- 0.640	- 0.081	- 0.248	- 0.119
29	C4-43	- 0.074	- 0.175	- 0.730	- 0.125	- 0.123
30	C5-40	- 0.199	- 0.178	- 0.030	- 0.671	- 0.204
31	C5-20	- 0.094	- 0.323	- 0.055	- 0.547	- 0.247
32	C5-09	- 0.092	- 0.355	- 0.105	- 0.531	- 0.240
33	D3-22	- 0.155	- 0.604	- 0.054	- 0.134	- 0.336
34	D3-03	- 0.176	- 0.618	- 0.065	- 0.069	- 0.327
35	D3-15	- 0.133	- 0.584	- 0.097	- 0.069	- 0.289
36	D3-12	- 0.186	- 0.539	- 0.196	- 0.022	- 0.070
37	D4-32	- 0.110	- 0.652	- 0.143	- 0.189	- 0.120
38	D4-30	- 0.050	- 0.594	- 0.231	- 0.135	- 0.025
39	D4-29	- 0.132	- 0.684	- 0.060	- 0.143	- 0.135
40	D4-27	- 0.049	- 0.490	- 0.330	- 0.152	0.089

TABELA 4.
Matriz Padrão — 1º Trimestre, após rotação oblíqua

VARIÁVEL		COMPONENTES PRINCIPAIS				
Nº	POSTO	I	II	III	IV	V
1	B4-29	- 0.072	0.778	- 0.019	- 0.028	- 0.029
2	B4-26	- 0.063	0.769	- 0.018	0.031	- 0.052
3	B4-02	- 0.062	0.712	- 0.066	- 0.030	- 0.052
4	B4-01	- 0.059	0.517	- 0.084	- 0.136	- 0.138
5	B4-18	0.053	0.472	- 0.030	- 0.378	- 0.121
6	B4-03	0.008	0.511	0.008	- 0.277	- 0.183
7	B4-15	0.019	0.112	- 0.646	- 0.029	0.055
8	B4-12	- 0.050	0.433	- 0.051	- 0.321	- 0.071
9	B4-05	- 0.018	0.336	0.001	- 0.211	- 0.414
10	B5-02	- 0.025	0.219	- 0.084	- 0.579	0.102
11	B5-29	- 0.031	0.100	- 0.196	- 0.505	0.031
12	B5-28	- 0.008	0.119	- 0.060	- 0.698	- 0.027
13	B5-04	- 0.009	0.175	- 0.455	- 0.284	0.125
14	B5-24	- 0.045	0.032	- 0.066	- 0.627	- 0.004
15	C3-40	- 0.120	0.153	- 0.006	- 0.133	- 0.647
16	C3-24	- 0.046	- 0.002	- 0.629	0.058	- 0.296
17	C3-14	- 0.072	0.110	- 0.071	- 0.026	- 0.688
18	C3-11	- 0.013	- 0.057	- 0.168	- 0.046	- 0.657
19	C3-09	- 0.251	0.159	- 0.217	0.095	- 0.187
20	C3-34	- 0.251	0.080	- 0.196	0.110	- 0.030
21	C3-29	0.003	0.002	- 0.470	- 0.020	- 0.421
22	C4-34	- 0.099	0.081	- 0.081	- 0.475	- 0.205
23	C4-07	- 0.085	- 0.018	- 0.216	- 0.417	- 0.220
24	C4-40	- 0.182	0.182	0.016	- 0.111	- 0.571
25	C4-29	- 0.472	0.044	- 0.051	- 0.150	- 0.244
26	C4-21	- 0.493	0.033	- 0.298	- 0.076	- 0.043
27	C4-52	- 0.069	0.022	- 0.703	- 0.108	- 0.039
28	C4-33	- 0.614	0.070	0.016	- 0.174	- 0.105
29	C4-43	- 0.090	0.002	- 0.723	- 0.046	- 0.047
30	C5-40	- 0.083	0.055	0.024	- 0.666	- 0.131
31	C5-20	- 0.248	- 0.047	0.033	- 0.529	- 0.179
32	C5-09	- 0.280	- 0.049	- 0.018	- 0.506	- 0.166

TABELA 4. (Continuação)
Matriz Padrão — 1º Trimestre, após rotação oblíqua

VARIÁVEL		COMPONENTES PRINCIPAIS				
Nº	POSTO	I	II	III	IV	V
33	D3-22	- 0.567	0.063	0.042	- 0.048	- 0.269
34	D3-03	- 0.567	0.097	0.027	0.025	- 0.269
35	D3-15	- 0.556	0.057	- 0.014	0.018	- 0.224
36	D3-12	- 0.534	0.151	- 0.138	0.061	0.017
37	D4-32	- 0.640	0.032	- 0.059	- 0.117	- 0.023
38	D4-30	- 0.591	- 0.013	- 0.189	- 0.092	0.076
39	D4-29	- 0.683	0.062	0.028	- 0.069	- 0.042
40	D4-27	- 0.491	0.003	- 0.290	- 0.099	0.191

Selecionando-se o máximo módulo dos fatores de peso para cada posto, identifica-se, para cada componente principal, um conjunto de postos que, identificados num mapa, demonstram uma característica gregária espacialmente ordeira, conforme figura 4 — 1º trimestre. A separação entre regiões foi traçada artisticamente, podendo ser refinada em função do relevo do terreno.

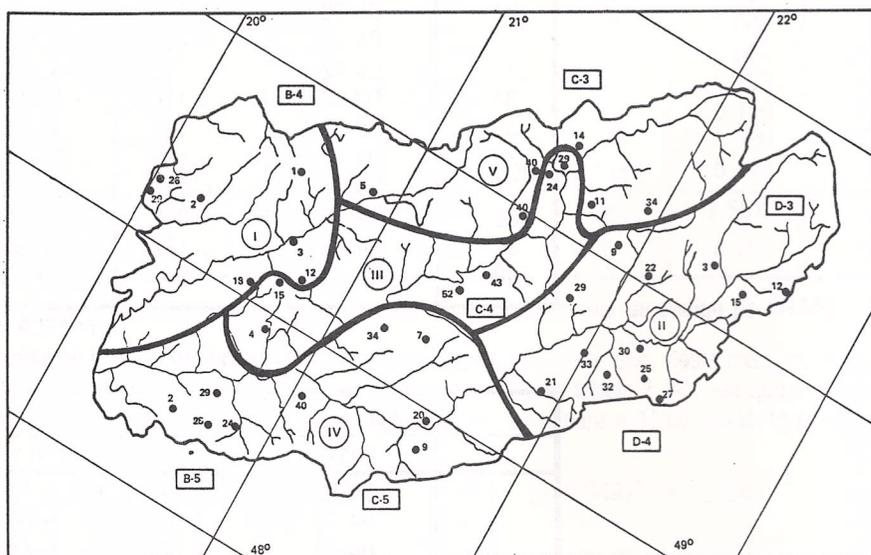


FIGURA 4. Regiões homogêneas — 1º Trimestre.

Ressalta-se, olhando a figura 4, o posicionamento dos postos e regiões homogêneas. Dificilmente um hidrólogo, mesmo que experiente na região e conhecendo os acidentes do terreno, faria tal subdivisão. Observe-se algumas situações relativas interessantes como é o caso dos postos (B-4) 12, 15 e 18, e dos (C-3) 24, 29 e 40. Note-se também a diferenciação existente entre estas regiões e os limites físicos das sub-bacias, resultado este já esperado.

Como o mecanismo gerador das chuvas apresenta uma certa variabilidade anual, é de se esperar que o quadro de homogeneidade também se modifique. Em se fazendo a mesma análise para os outros trimestres e, para o semestre úmido, chega-se à situação exposta nas figuras 5, 6, 7 e 8 que, apesar de apresentar mudanças não muito pronunciadas, confirma esta expectativa.

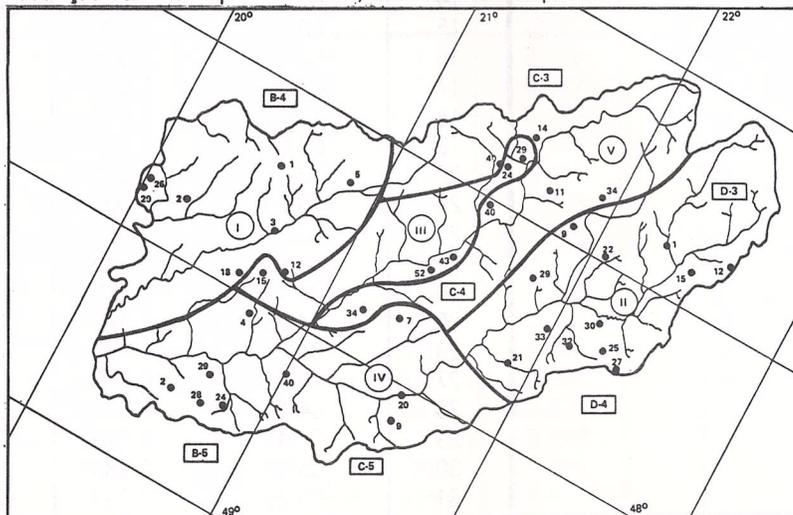


FIGURA 5. Regiões homogêneas — 2º Trimestre.

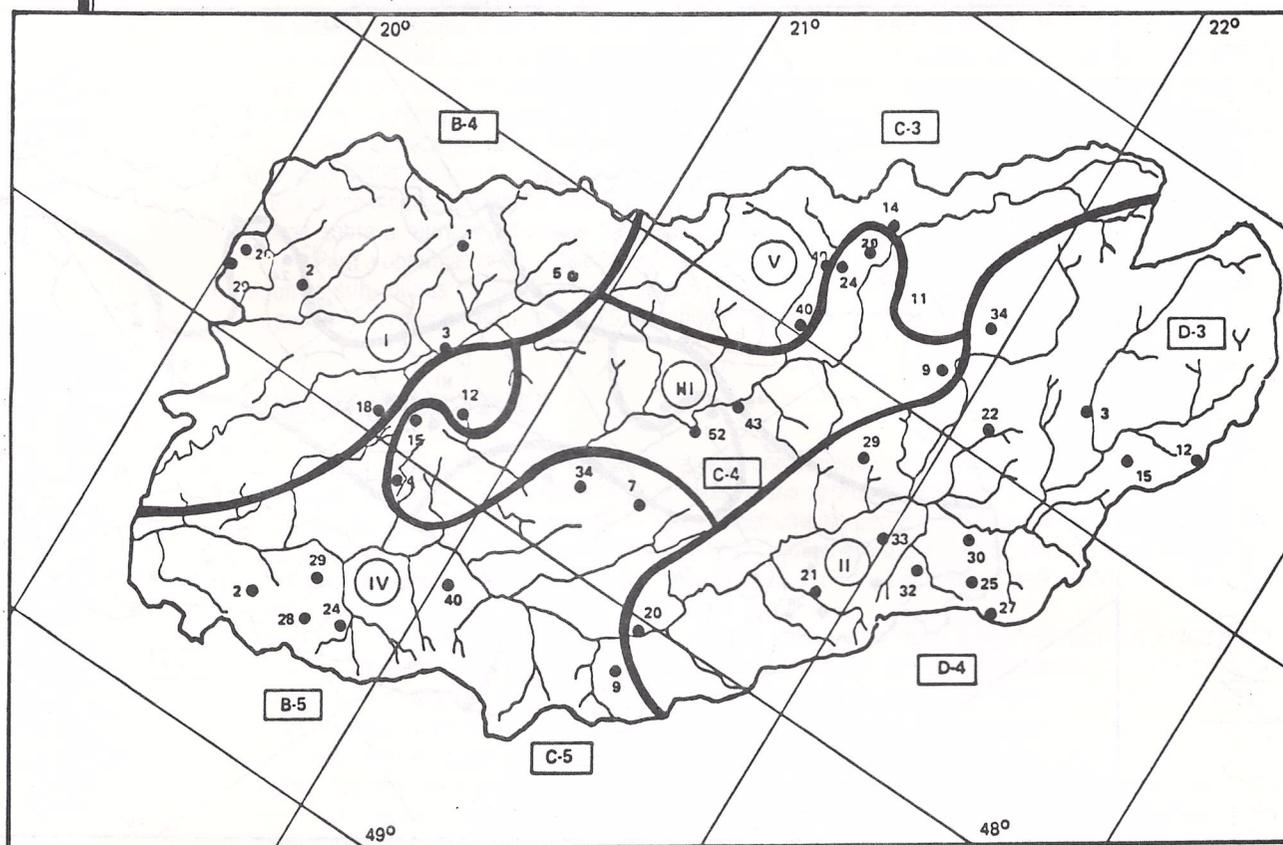


FIGURA 6. Regiões homogêneas — 3º Trimestre.

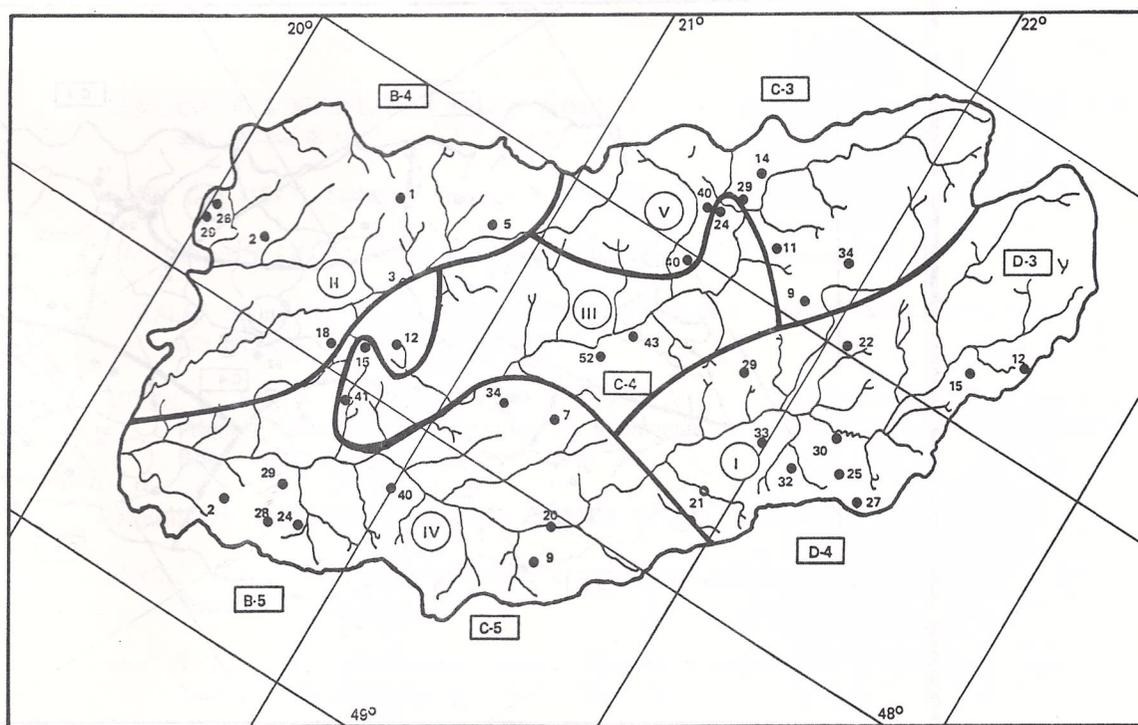


FIGURA 7. Regiões homogêneas — 4º Trimestre.

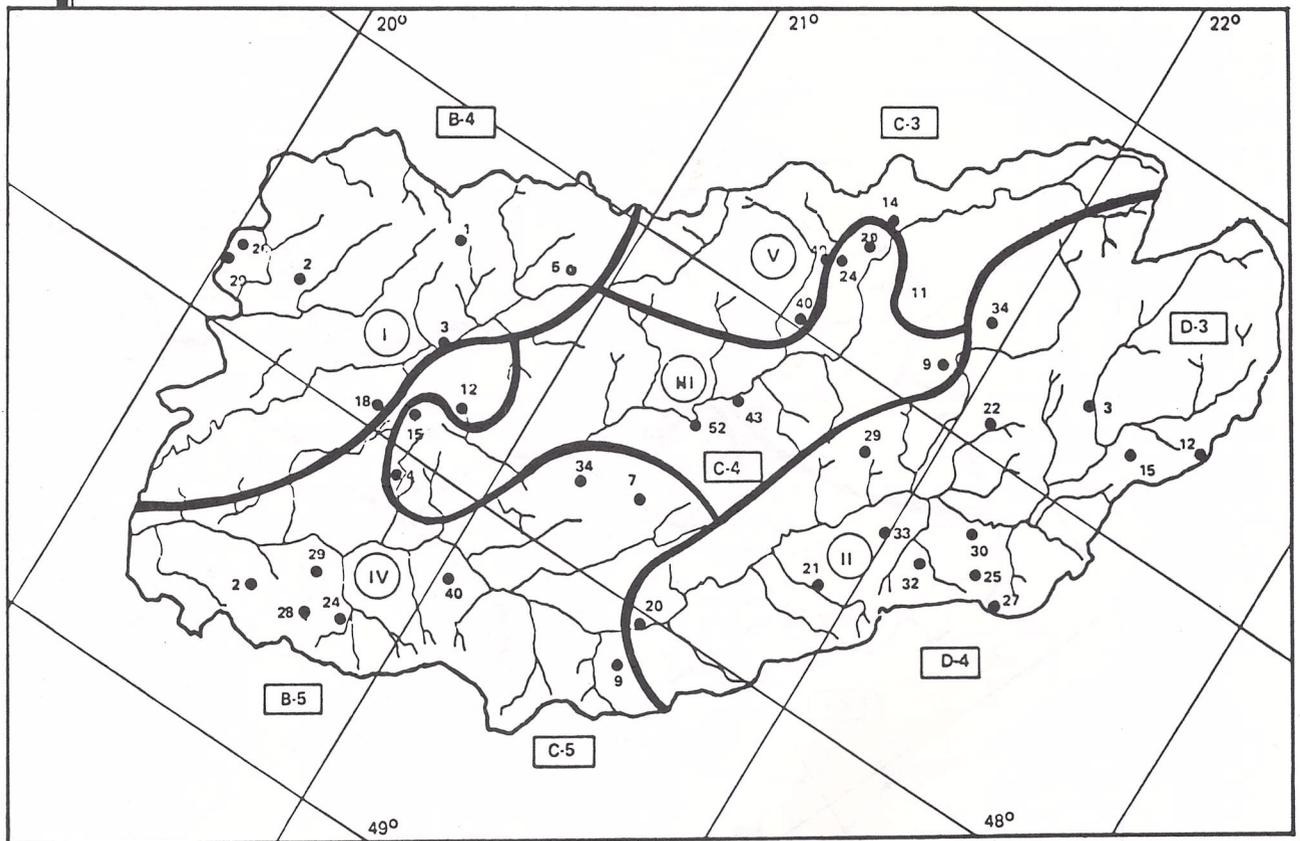


FIGURA 8. Regiões homogêneas — Semestre Úmido.

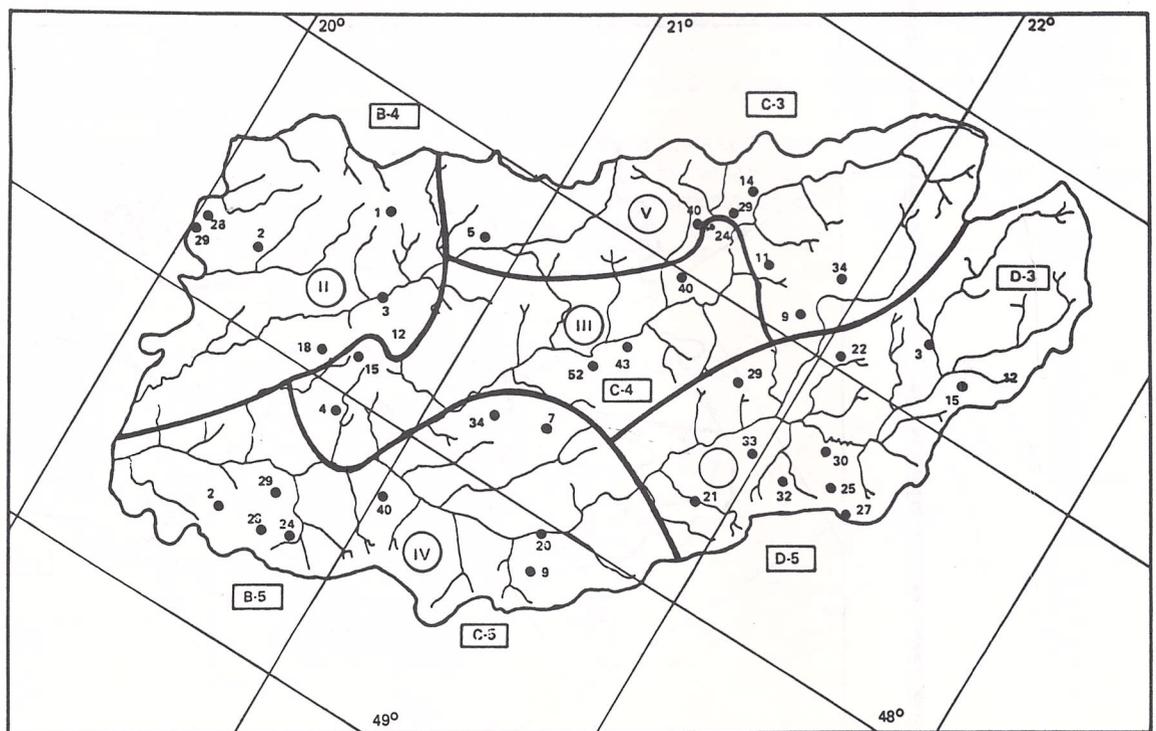


FIGURA 9. Variabilidade temporal da homogeneidade.

É interessante observar a troca de posição ocorrida com os componentes principais I e II no 4.º trimestre (e no semestre úmido, que engloba o 1.º e 4.º semestre). Admitindo-se que estes componentes principais estão ligados a alguma grandeza física, a troca pode sugerir uma mudança no mecanismo gerador das chuvas como por exemplo, uma maior preponderância nestes períodos (semestral e 4.º trimestre) de chuvas convectivas (ou frentes quentes) em detrimento das frontais frias.

Fica, entretanto, difícil relacionar diretamente um componente principal com alguma grandeza física, pois a metodologia foi desenvolvida com base a elementos relativamente abstratos. Estudos de natureza climatológica devem complementar esta análise e, eventualmente, permitir suposições mais realísticas sobre a representatividade física destes componentes principais.

Para a obtenção destes resultados utilizou-se, num computador IBM-370(4341), programas com algumas sub-rotinas do IMSL — International Mathematical and Statistical Libraries. Após debulhados os dados, o tempo de CPU foi de 75 segundos para cada trimestre.

COMENTÁRIOS FINAIS

Como foi exposto, a análise multivariada de componentes principais pode ser uma ferramenta útil na solução de problemas hidrológicos. Contudo, como apontado em Matalas e Reiher (1967), apesar deste tipo de análise dar a impressão de ser um instrumento poderoso devido sua aparente complexidade matemática, existem ainda dúvidas na sua utilização generalizada, principalmente dado que os componentes principais não são observáveis, seu significado físico é incerto e considerando que não se conhecem as características amostrais dos fatores de peso.

Mas, ressalvadas estas incertezas, é viável empregar a metodologia exposta nos estudos que envolvam variáveis temporais físicas, mormente se tais estudos são complementados por estudos climatológicos. Além de análises de homogeneidade [ver Beaudoin e Rousselle (1982)], esta metodologia pode ser empregada na racionalização de redes hidrológicas, já que alguns postos que apresentarem um baixo peso comum global (h_j^2), não acrescentam muita informação ao sistema. Conforme Morin et alii (1979), também pode ser empregada para se decidir sobre a interpolação ótima inter-estações.

Um campo promissor advém do emprego desta metodologia na análise de homogeneidade de campos de chuvas intensas, bastando para isto estabelecer-se um valor mínimo relativamente alto (ao invés do 0,3mm adotado) e proceder a mesma análise aqui descrita.

Para um modelista, a metodologia exposta é muito útil porque, além de identificar sub-regiões onde o campo pluvial é homogêneo, possibilitando estabelecer análises de regressão, suprir dados faltantes e ajustar modelos tipo chuva-vazão, proporciona uma melhor visão global do campo pluvial, o que permite a identificação de estações temporalmente imutáveis e aquelas cuja homogeneidade é variável no tempo. Tal pode ser observado na figura 9, que é a junção das figuras 4 a 8, onde os pontos plotados com ■ são os que mudaram de região pluvial em pelo menos um dos trimestres analisados.

BIBLIOGRAFIA

- BEAUDOIN, P. & ROUSSELLE, J. (1982), **A study of Space Variations of Precipitation by Factor Analysis**. Journal of Hydrology, 59; pages 123-138.
- HANN, C.T. (1977), **Statistical Methods in Hydrology**, Iowa State University Press, Ames, IOWA.
- HARMAN, H.H. (1976), **Modern Factor Analysis**, University of Chicago Press, Chicago, Ill.
- MATALAS, N.C. & REIHER, B.J. (1967), **Some Comments on the Use of Factor Analysis**, Water Resources Research 3(1): pages 213-224.
- MORIN, G.; FORTIN, J.P.; SOCHANSKA, W.; LARDEAU, J.P.; CHARBONNEAU, R. (1979), **Use of Principal Component Analysis to Identify Homogeneous Precipitation Stations of Optimal Interpolation**, Water Resources Research, 15(6). December.
- MORRINSON, D.F. (1967), **Multivariate Statistical Methods**, McGraw Hill, New York.
- SNYDER, W.M. (1962), **Some Possibilities for Multivariate Analysis in Hydrologic Studies**, J. Geophys. Research 67(13), pages 721-730.
- WALLIS, J.R. (1965), **Multivariate Statistical Methods in Hydrology — A Comparison using data of known functional Relationship**, Water Resources Research 1(4), pages 447-461.
- WONG, S.T. (1968), **A Multivariate Statistical Model for Predicting Mean Annual Flood in New England**, in, Spatial Analysis — A Reader in Statistical Geography. Prentice Hall.