

## Uso de Análise Multivariada na Regionalização de Cheias do Estado de São Paulo

Francisco de Assis Salviano de Sousa, Rodrigo Cezar Limeira

Centro de Ciências e Tecnologia UFCG, Campina Grande, PB.

fassis@dca.ufcg.edu.br, rodrigocezarlimeira@yahoo.com.br

Recebido: 24/11/04 – revisado: 15/12/05 – aceito: 12/09/06

---

### RESUMO

*Em hidrologia os métodos ou práticas de regionalização de cheias estão tradicionalmente vinculados às características físicas de bacias hidrográficas. É comum se delimitar sub-regiões hidrologicamente homogêneas com base nessas características. Essa metodologia exige, para se obter bons resultados, que a distribuição espacial dos postos fluviométricos seja relativamente uniforme. Essa exigência, especialmente no Brasil, torna o método pouco prático e passível de erros, visto que a densidade de estações fluviométricas no Brasil é baixa. O objetivo deste trabalho é o de propor uma metodologia alternativa, com base na análise multivariada, para delimitar sub-regiões hidrologicamente homogêneas, e em seguida regionalizar as cheias nessas sub-regiões. Uma das vantagens da metodologia aqui proposta é a de que ela utiliza as variáveis com perspectivas espacial e temporal e não apenas espacial, como nos métodos tradicionais de regionalização de cheias. Essa metodologia, da forma como é proposta aqui, independe da distribuição espacial das estações fluviométricas. Para aplicá-la foi escolhida uma parte da área geográfica do Estado de São Paulo. Como resultado obteve-se as delimitações de quatro sub-regiões hidrologicamente homogêneas a partir dos totais mensais precipitados. Essas delimitações explicam 70,4 % de toda a variância do sistema. De modo geral, os resultados em análise multivariada são tanto melhores quanto maior for a variância intergrupos.*

**Palavras-chave:** Análise multivariada; regionalização de cheias.

---

### INTRODUÇÃO

A cheia é um fenômeno que ocorre aleatoriamente no tempo e apresenta grande variação de um local para outro. As condições hidrológicas que produzem a cheia podem ser naturais ou artificiais. As naturais são aquelas cuja ocorrência é propiciada pela bacia hidrográfica em seu estado natural. Algumas dessas condições são relevos, tipos de solo, precipitação, cobertura vegetal, capacidade de drenagem etc.

As condições artificiais da bacia são àquelas provocadas pela ação do homem, tais como desmatamento, urbanização e obras hidráulicas em geral. Geralmente, as conseqüências da cheia são: doenças, mortes, prejuízos ao meio-ambiente e, dependendo de sua magnitude, grandes prejuízos econômicos.

A sociedade, de um modo geral, demanda proteção contra os riscos de cheia. Os governos municipais, estaduais e federal deveriam responder a essas demandas com planejamento e implementação de medidas estruturais e não-estruturais de pro-

teção contra as cheias a fim de controlar e mitigar seus efeitos. A implementação dessas proteções requer o conhecimento da magnitude e da frequência de recorrência associada à cheia. Tal conhecimento é também necessário a projetos, construções e operações de estruturas hidráulicas que servem a propósitos de abastecimento de água, geração de energia elétrica, navegação etc.; e tem o intuito de protegê-las de falhas e prejuízos.

Na análise de frequência de cheia, os valores das vazões máximas diárias anuais representam a variável aleatória ou evento cheia. Esses valores são ajustadas a uma função distribuição de probabilidades e, em seguida, os quantis de cheia para um dado período de retorno podem ser estimados. Mas, nesse tipo de estudo, quase sempre surgem problemas com a escassez de informações, visto que os registros históricos de vazões são usualmente curtos para garantir estimativas confiáveis de probabilidade de excedência.

A solução para contornar o problema de escassez de dados, e extrair mais informações, é usar os registros de uma sub-região com comportamento de cheia similar. Essa solução é denominada Análise

de Frequência de Cheia Regional (AFCR) ou regionalização de vazões. A regionalização é uma ferramenta que possibilita uma alternativa para a determinação de vazões de projeto. Cunanne (1998) apresenta as principais características de 12 métodos de AFCR. Na aplicação de alguns desses métodos exige-se, para obtenção de bons resultados na quantificação de cheia, que a sub-região de estudo seja hidrologicamente homogênea. É comum encontrar na literatura especializada estudos de cheia que fazem uso de alguns métodos de AFCR, que requerem homogeneidade hidrológica da sub-região de estudo, que não consideram devidamente essa homogeneidade.

Para alguns autores, uma sub-região que apresenta clima ou pluviometria ou vazões de cheia similar pode ser denominada, respectivamente, por: climatologicamente homogênea, pluviometricamente homogênea e hidrologicamente homogênea. Aqui, a definição de sub-região hidrologicamente homogênea está relacionada às curvas de frequências de cheias. Se essas curvas, para uma dada sub-região, apresentarem mesma tendência, diz-se que há um forte indício de que essa sub-região seja hidrologicamente homogênea. Quando uma sub-região, supostamente homogênea, é delimitada pelos parâmetros fisiográficos das sub-bacias, critério muito utilizado, essa delimitação pode ser ajustada pela distribuição geográfica dos resíduos da equação de regressão, mas ainda assim pode apresentar problemas de *contigüidade e erros nas delimitações dos contornos*. Em adição Fill (1994) questiona a validade da AFCR, que requeira sub-região homogênea, quando aplicada em regiões com poucos postos fluviométricos.

No presente trabalho é apresentado, sem pretender esgotar o assunto, uma revisão dos métodos multivariados utilizados na engenharia hidrológica, para estudar o comportamento das cheias em termos das características físicas e delimitar sub-regiões hidrologicamente homogêneas. Nessa revisão, se busca subsídios para apresentar um método que aqui delimite as sub-regiões, afim de que possam ser utilizadas por alguns métodos de AFCR. A fim de melhorar os critérios que são úteis no processo de delimitação de sub-regiões hidrologicamente homogêneas ou estudar a evolução das mudanças na morfometria das bacias hidrográficas, pesquisadores como Abrahams (1972), White (1975), Wong (1979), Mimikou e Kaemaki (1985), Mosley (1981), Acreman (1985), Wiltshire (1986c), Acreman e Sinclair (1986) e Nathan e McMahon (1990), Ribeiro-Corréa et. al. (1995), Chiang et. al. (2002a), Chiang et. al. (2002b), Rao e Srinivas (2006), utilizaram as

características físicas de bacias hidrográficas e técnicas de análise multivariada. Nos estudos supracitados foram usadas as seguintes técnicas multivariada: Análise Canônica, Análise de Componentes Principais e Análise de Agrupamento. Apesar do inegável avanço, novos questionamentos exigem respostas dos estudiosos de recursos hídricos. São eles: 1) qual técnica multivariada pode ser eleita como a mais apropriada para delimitar sub-regiões hidrologicamente similares? 2) a partir da variabilidade da precipitação é possível representar a variabilidade da cheia? O objetivo do presente estudo busca as respostas desses questionamentos por meio de uma ferramenta matemática multivariada, denominada Análise de Fatores Comuns (AFC). A área do Estado de São Paulo, com exceção da área metropolitana da cidade de São Paulo e do Vale do Paraíba, foi utilizada para aplicação dessa metodologia alternativa.

## METODOLOGIA

A quantificação das variáveis hidrológicas resulta de uma ou mais observações ou de medidas em campo. Em função das imperfeições das medidas ou das observações; seja por falta de calibração dos instrumentos, seja por imperícia do operador ou observador; os valores das variáveis observadas, em geral, divergem dos verdadeiros valores por diferenças que recebem o nome de erros de observações. Esses erros podem ser classificados em erros de transcrição, erros grosseiros, erros sistemáticos e erros fortuitos.

Em vista do exposto acima, há necessidade de se detectar os erros nas amostras de dados aqui utilizadas e, em seguida, efetuar a análise de consistência. Rao e Hsieh (1991) propuseram uma metodologia para estimar os valores das variáveis em locais sem medições ou observações, denominada Funções Ortogonais Empíricas (FOE). A idéia básica desse método é a de representar a série de dados hidrológicos padronizada em termos de suas componentes espaciais e temporais. A vantagem desse método quando comparado com os métodos tradicionais de preenchimento de falhas é que os valores são estimados em função de pesos espacial (menor distância entre o posto com valores faltosos e os postos de sua vizinhança) e temporal (coeficientes resultantes de um sistema de combinação linear com base na amostra de dados). Neste estudo a FOE será utilizada para preencher os valores faltosos

do período base escolhido e, quando possível, verificar a consistência dos valores observados.

**Análise de fatores comuns**

A origem da análise de fatores é atribuída a Charles Spearman. Ele desenvolveu em 1904 uma teoria psicológica envolvendo um número específico de fatores no trabalho “General intelligence, objectively determined and measured”, publicado pelo American Journal of Psychology *apud* Harman (1960). A teoria foi progredindo e Thurstone (1965) apresentou-a expandida para fatores múltiplos. A metodologia de Spearman — Thurstone domina, até hoje a aplicação da análise de fatores. Essa técnica, apesar de ter sido proposta no início do século 20, somente teve larga aplicação a partir da década de 60, início da revolução quantitativa das ciências sociais e naturais.

As técnicas matemáticas inerentes à análise de fatores, certamente, não ficaram limitadas às aplicações psicológicas. Devido ao desenvolvimento e divulgação, ela foi sendo incorporada por outros ramos da ciência, tais como: ciência política, comércio, biologia, economia, astronomia e meteorologia. Especificamente, a análise de fatores foi introduzida em hidrologia por Snyder (1962), Wong (1963) e Wallis (1965) *apud* Matalas e Reihner (1967), e despertou interesse considerável entre os hidrologistas.

A análise de fatores é uma técnica matemático-estatística multivariada. Seu objetivo é o de simplificar os relacionamentos, diversos e complexos, que existem entre séries temporais observadas. Essa técnica descobre dimensões ou fatores comuns que ligam essas séries e em consequência, produz introspecção na estrutura implícita dos dados históricos.

Existe, por exemplo, em hidrologia, uma relação implícita da vazão com a intensidade da precipitação, com a cobertura vegetal da bacia hidrográfica, com o relevo, com o clima etc. A análise de fatores procura descobrir o grau de correlação entre essas variáveis a fim de apresentar um conjunto comum que melhor explique a vazão.

De forma semelhante existe uma relação implícita da precipitação, que ocorre na bacia hidrográfica, com a evaporação, evapotranspiração, com a direção e velocidade do vento, com a umidade do ar, com a temperatura, com a radiação solar, com a pressão atmosférica etc. Adicionalmente, há ainda inter-relações entre essas mesmas variáveis. Todas essas intrincadas relações são motivadas pela dinâmica da atmosfera e dão origem ao mecanismo gerador dos valores da variável precipitação. De

modo mais simplificado, e fazendo analogia com a análise de correlação e regressão linear, pode-se dizer que o fator, da análise de fatores, é uma reta de regressão linear que explica os valores comuns dispersos em torno dela. É por isso que alguns autores denominam a análise de fatores de *Análise de Fatores Comuns* (AFC). A essência da AFC é mostrada na Figura 1.

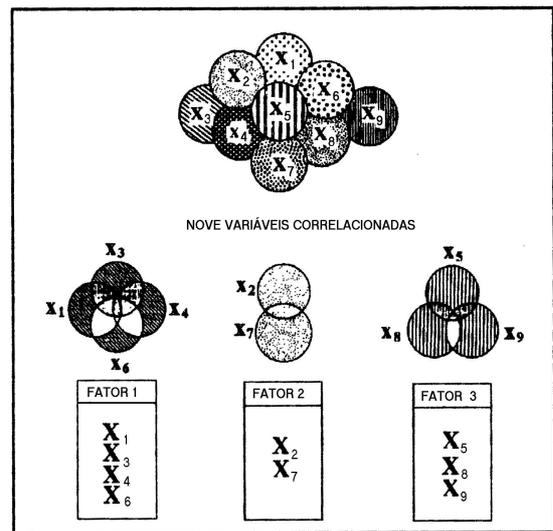


Figura 1 Nove variáveis reduzidas a três fatores

Nela existem nove variáveis  $X_1, X_2, \dots, X_9$  que “pesam” sobre três fatores comuns não-observados. As variáveis  $X_1, X_3, X_4,$  e  $X_6$  estão agrupadas no fator 1, isso significa que elas são bem correlacionadas e constituem o primeiro fator. Similarmente, as variáveis  $X_2, X_7$  definem o segundo fator e as variáveis  $X_5, X_8$  e  $X_9$  formam o terceiro fator. Portanto, cada sub-série de variáveis pode ser imaginada como o reflexo de uma compreensão implícita dos fatores. Assim, ao invés de lidar com as características associadas às nove variáveis, só é preciso, agora, considerar três fatores.

**A função da AFC**

O modelo de Análise de Fatores Comuns (AFC) assume que as variáveis consistem de duas partes: uma comum e outra única. A parte comum é aquela da variação da variável que é repartida com as partes comuns das outras variáveis, enquanto que a parte única é aquela da variação da variável que é específica à variável apenas. As partes comuns de todas as variáveis definem um espaço vetorial co-

mum. É assumido que as partes únicas das variáveis são incorrelacionadas, e que há, também, incorrelações entre as partes únicas e comuns. O objetivo da AFC é definir as dimensões desse espaço vetorial. Essas dimensões são chamadas de fatores comuns. Em outras palavras, a AFC estuda as inter-relações entre as variáveis originais, a fim de encontrar uma nova série de variáveis, menor que a série original, que expresse aquilo que é comum entre as variáveis originais. O estudo da AFC pode ter as seguintes funções: i) reduzir o número de variáveis originais, mantendo a informação tão original quanto possível, ou seja, não alterando muita a variância dos dados; ii) identificar grupos de indivíduos inter-relacionados; iii) reescrever a série de dados originais de uma forma alternativa – quando há problemas de colinearidade das variáveis independentes -, visto que os fatores são ortogonais e, portanto, independentes. Sem perda de generalidade, a função da AFC, aqui, é a de identificar grupos de variáveis intercorrelacionadas e agrupar indivíduos (postos pluviométricos) com comportamento semelhantes.

**O modelo AFC**

Pode-se descrever o uso da AFC – com relação à análise de dados – em dois diferentes contextos. Por exemplo, o analista dispõe, a priori, de informações teóricas sobre a estrutura comum implícita dos dados, e deseja confirmar ou negar uma hipótese. O uso da AFC, nesse caso, é chamado de confirmatório. De modo diferente, quando o analista não dispõe de nenhuma hipótese teórica, ele procura uma estrutura comum implícita nos dados. O uso da AFC, desta forma, é denominado exploratório, e será aqui estudado. A Figura 2 mostra um modelo de AFC exploratório.

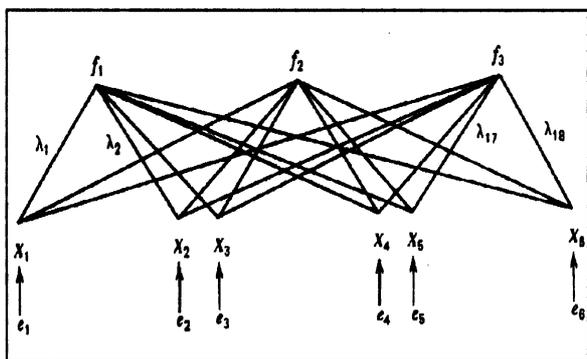


Figura 2 Modelo de AFC exploratório

Os termos  $f_1, f_2$  e  $f_3$  representam fatores incorrelacionados (ortogonais) não - observados (FNOs). Cada um dos  $X_i$  é uma variável observada que estima, até certo grau, os três FNOs. A relação entre os FNOs e as variáveis observadas é dada pelas cargas dos fatores ( $l_i$ ). Se todas as variáveis forem padronizadas e os FNOs forem incorrelacionados, as cargas dos fatores ( $l_i$ ) darão as correlações ordinárias simples, entre as variáveis observadas ( $X_i$ ) e os FNOs.

Visto que, geralmente, os FNOs não podem explicar cem por cento da variância de cada variável, explicam apenas as partes comuns; um erro, para representar a parte única da variável, simbolizando por  $e_i$ , é incluído para capturar a porção restante da variância da variável. Os erros são considerados independentes, não-correlacionados, com relação aos FNOs e a eles próprios. Algebricamente, as variáveis observadas  $X_i$  podem ser representadas como combinações lineares das cargas dos fatores ( $l_i$ ) e dos FNOs, como nas equações:

$$\begin{aligned}
 X_1 &= \lambda_1 f_1 + \lambda_7 f_2 + \lambda_{13} f_3 + e_1 \\
 X_2 &= \lambda_2 f_1 + \lambda_8 f_2 + \lambda_{14} f_3 + e_2 \\
 X_3 &= \lambda_3 f_1 + \lambda_9 f_2 + \lambda_{15} f_3 + e_3 \\
 X_4 &= \lambda_4 f_1 + \lambda_{10} f_2 + \lambda_{16} f_3 + e_4 \\
 X_5 &= \lambda_5 f_1 + \lambda_{11} f_2 + \lambda_{17} f_3 + e_5 \\
 X_6 &= \lambda_6 f_1 + \lambda_{12} f_2 + \lambda_{18} f_3 + e_6
 \end{aligned}
 \tag{1}$$

Notar que na AFC exploratória cada uma das  $X_i$  variáveis “pesa” sobre cada um dos FNOs. A decisão quanto a quais variáveis estimarão que FNOs, só poderá ser tomada após as estimativas das cargas dos fatores ( $\lambda_i$ ). O modelo básico de Análise de Fatores Comuns (AFC) é, usualmente, segundo RUMMEL (1970), expresso na forma vetorial por:

$$X = \Lambda f + e
 \tag{2}$$

em que  $X$  é o vetor coluna p-dimensional das variáveis observadas,  $x_1, x_2, \dots, x_p$ ;  $f$  é vetor coluna q-dimensional das variáveis não-observadas ou FNOs  $f_1, f_2, \dots, f_q$ ;  $e$  é o vetor coluna p-dimensional das

variáveis não-observadas ou erros,  $e_1, e_2, \dots, e_p$  e  $\Lambda$  é a matriz de cargas dos fatores ( $\lambda_i$ ), de dimensão  $p \times q$ . O modelo linear (2), expresso em termos de  $q$  fatores comuns e seus  $p$  fatores únicos, pode se escrito como segue:

$$X_i = \sum_{j=1}^q \lambda_{i,j} f_j + e_i \quad (3)$$

A equação (3) quando expandida é chamada de padrão dos fatores ou padrão. A matriz das cargas dos fatores de dimensão  $p \times q$ , com  $q < p$ , que produz as correlações entre as  $i$ -ésimas variáveis observadas e os  $j$ -ésimos fatores  $f_j$  é denominada de matriz estrutura dos fatores ou matriz estrutura. As definições supracitadas são necessárias para formar a base das soluções do problema proposto. Notar que na equação (3) as variáveis observadas, escritas em função dos FNOs são indicadoras refletivas, isto é, estão sujeitas às estimativas dos erros,  $e_i$ .

**A variância comum**

Na equação (3), as variáveis  $X_i$  são repartidas em duas partes incorrelacionadas, matematicamente:

$$X_i = C_i + e_i \quad (4)$$

em que  $C_i = \lambda_{i1}f_1 + \lambda_{i2}f_2 + \dots + \lambda_{iq}f_q$  é a parte de cada variável comum às outras  $p-1$  variáveis, e  $e_i$  é a parte única dessa variável. Visto que as partes comuns e únicas são incorrelacionadas, e os fatores  $f_q$  têm variância unitária, pode-se dividir a variância de  $X_i$  em:

$$\text{var}(X_i) = \text{var}(C_i) + \text{var}(e_i) \quad (5)$$

em que,  $\text{var}(C_i)$  e  $\text{var}(e_i)$  representam as variâncias comum e única de  $X_i$ , respectivamente. A variância comum de uma variável é também chamada de comunalidade da variável. A comunalidade é a porção da variância total da variável que é explicada pelos fatores comuns. Denominando a comunalidade da  $i$ -

ésima variável por  $h_i^2$  e a variância única por  $\Psi_i$ , pode-se escrever:

$$\text{var}(X_i) = h_i^2 + \Psi_i \quad (6)$$

note que  $\text{var}(C_i) = h_i^2 = \sum_{j=1}^q \lambda_{i,j}^2$  é simplesmente a soma quadrática dos elementos da  $i$ -ésima coluna da matriz  $\Lambda$ . A variância única  $\Psi_i$  é chamada de unicidade da variável e reflete o grau de não-explicação da variância da variável. A contribuição total do fator  $f_j$  à variância total da serie de variáveis é dada pelo autovalor do fator  $f_j$ , dado por:

$$V_j = \sum_{i=1}^p \lambda_{i,j}^2 = \lambda_j' \lambda_j \quad (7)$$

em que,  $\lambda_j$  denota a carga da  $j$ -ésima coluna e  $\lambda_j'$  a sua transposta. A equação (7) nada mais é do que a soma quadrática das cargas dos fatores, para  $j = 1, 2, \dots, q$ . A contribuição de todos os fatores comuns à variância total de todas as variáveis, é dada pela comunalidade total:

$$V = \sum_{j=1}^q v_j \quad (8)$$

A variância das variáveis, explicada por um fator  $f_j$ , pode ser dada por uma percentagem da variância total, matematicamente dada por:

$$VT = \sum_{i=1}^p \sum_{j=1}^q \lambda_{i,j}^2 + \sum_{i=1}^p \Psi_i \quad (9)$$

O objetivo de estimativas das comunalidades é o de ponderar a importância de cada variável, de acordo com o seu grau de correlação, com as outras variáveis. As soluções com correlação igual à unidade não é de interesse em AFC. Dentre os vários métodos de estimativas das comunalidades, optar-se-á pelo método de correlação múltipla quadrática (CMQ). Segundo Walsh et al. (1982) esse método é o que melhor estima as comunalidades, pois apresenta soluções estáveis.

## Solução de fatores

O mais importante tópico no estudo da AFC é a extração de fatores. Por meio dessa técnica, a matriz de correlação pode ser reduzida à dimensão de seus fatores. Rummel (1970) dá ênfase às técnicas disponíveis, suas características, o critério para selecionar uma alternativa e as explicações dos conceitos envolvidos no processo. Ainda segundo Rummel (1970), dentre os métodos utilizados na extração dos fatores podem ser citados: Método da Análise Residual Mínima, da Fatoração Canônica de Rao, da Análise de Imagem, da Análise Alfa, da Máxima Verossimilhança e o do Fator Principal, este último será adotado aqui devido a sua simplicidade e rapidez computacional. O Método do Fator Principal (MFP) extrai os fatores, de tal modo que cada fator explica a máxima variância possível contida na série que está sendo fatorada. O MFP inicia gerando os coeficientes  $\lambda_{11}, \lambda_{21}, \dots, \lambda_{p1}$  para o fator  $f_1$  de tal modo que a contribuição de  $f_1$  para a comunalidade total  $V$  seja maximizada, e sujeita a  $p(p-1)/2$  correlações e  $p$  comunalidades específicas. Harman (1960) mostra que essa solução é equivalente a encontrar os autovalores e autovetores da matriz de correlação reduzida  $R^*$ . Se  $R^*$  for uma matriz simétrica e real, existe uma matriz  $E$ , ortogonal por colunas, tal que  $E'E=I$  e uma matriz diagonal  $D$ , tal que:  $EDE'=R^*$ . Nessa relação são válidas as permutações dos elementos da diagonal da matriz  $D$  com as correspondentes permutações das colunas da matriz  $E$  e das linhas da matriz  $E'$ . Portanto, os elementos da diagonal principal da  $D$  podem ser arrançados em ordem decrescentes de magnitudes. Os elementos da diagonal principal da  $D$  são denominados autovalores, ou raízes latentes, raízes características ou valores próprios da matriz  $R^*$ . As colunas da matriz  $E$  são chamadas de autovetores, ou vetores latentes, vetores característicos ou vetores próprios da matriz  $R^*$ . As provas dessas proposições podem ser encontradas em Murdoch (1972) e as aplicações em Stidd (1967).

Em particular, pela expressão (7), o valor máximo de  $V_1 = \lambda'_1 \lambda_1$  é o maior autovalor da  $R^*$ . em seguida, busca-se  $\lambda_2$  para maximizar  $V_2 = \lambda'_2 \lambda_2$ . As etapas subsequêntes seguem de forma similar. O processo finda com uma série de vetores. O "Scree" Teste será utilizado como critério para definir o número de fatores a ser extraídos. Aqui, esse número será igual ao de sub-regiões plu-

viometricamente homogêneas. Segundo Cattell (1978) esse teste baseia-se no exame do plote do número de autovetores versus os autovalores. Quando ocorre uma mudança brusca, da direita para a esquerda, na inclinação da curva, o número de fatores a ser extraído é igual ao valor numérico da abscissa daquele ponto.

## Rotação de fatores

Existem dois tipos de rotação de eixos: ortogonal e oblíqua. A ortogonal rotaciona rigidamente um predeterminado número de fatores, em um dado ângulo, mantendo a ortogonalidade dos mesmos. A oblíqua, mais flexível, permite rotacionar os fatores sem limitar-se à ortogonalidade. Essa solução rotineiramente produz fatores com maior clareza interpretativa do que a solução ortogonal. Burroughs; Miller (1961) apud Richman (1981), estudando Análise de Componentes Principais (ACP) concluíram que as componentes principais (CPs), após a rotação, representavam melhor os grupos de variáveis do que as componentes antes da rotação. Karl e Koscielny (1982) aplicaram e interpretaram os dois tipos de rotação. Quando o objetivo da rotação de fatores for o de realçar grupos de variáveis interrelacionadas, o que é o caso deste estudo, a rotação ortogonal não é adequada para essa tarefa. Portanto, será utilizada aqui a rotação oblíqua.

## Estimativa de adequabilidade amostral

A estimativa de adequabilidade da amostra na AFC é dada pelo índice Kaiser-Meyer-Olkin (KMO), segundo Friel (1997). O KMO compara a magnitude dos coeficientes de correlação observados ( $r_{i,j}$ ) com a magnitude dos coeficientes de correlação parciais ( $a_{i,j}$ ). Se a soma dos quadrados de ( $a_{i,j}$ ), entre todos os pares de variáveis, for pequena quando comparada com a soma dos quadrados de ( $r_{i,j}$ ), o KMO será próximo de um. Pequenos valores do KMO indicam que a metodologia de AFC não é adequada para as variáveis testadas, visto que as correlações entre pares de variáveis podem não ser explicadas pelos fatores comuns. Kaiser (1974) apud Richman (1986) classificou o KMO igual a 90% como maravilhoso, igual a 80% como meritório, igual a 70% como regular e igual a 60% como medíocre.

## RESULTADOS E DISCUSSÃO

A área escolhida para a aplicação da metodologia (parte do Estado de São Paulo) está compreendida entre 46 e 53° de longitude oeste, e entre 20 e 23° de latitude sul. A leste e ao norte a área é limitada pelo Estado de Minas Gerais, a oeste pelo Estado do Mato Grosso do Sul, e ao sul pelo Estado do Paraná. Essa área, com cerca de 80% da área do Estado, apresenta clima muito úmido sem estação seca. Os meses menos chuvosos, com normais pluviométricas inferiores a 100 mm, são os de inverno: junho a agosto, segundo o Departamento de Água e Energia Elétrica do Estado de São Paulo – DAEE (1970).

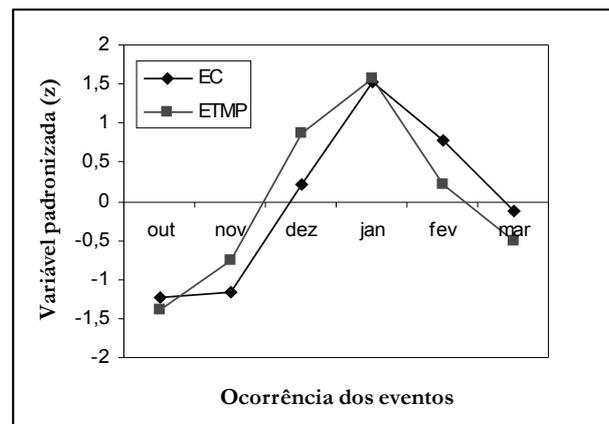
O primeiro passo na aplicação da metodologia foi verificar a validade de se representar a variabilidade das vazões máximas diárias anuais ou vazões de cheias, pelos totais mensais precipitados no período úmido: outubro a março. Para essa tarefa selecionou-se 64 postos fluviométricos e 148 postos pluviométricos, especialmente distribuídos na área de estudo, em um período base comum de 27 anos, de 1969 a 1995. Por estatística básica se obtiveram os valores da variabilidade percentual normalizada para os dois eventos, no período estudado. Para os ETMPs, os percentuais mensais foram calculados em relação aos totais anuais precipitados e para os ECs, contaram-se as freqüências de cheias e em seguida obtiveram-se os percentuais mensais em relação aos meses restantes dos anos. O passo seguinte foi padronizar as ocorrências desses eventos por  $z_i = (x_i - \bar{X})/S$ , em que  $\bar{X}$  e  $S$  são, a média e o desvio-padrão desses valores, respectivamente. A Tabela 1 lista esses valores.

**Tabela 1 Variabilidade percentual normalizada dos eventos de cheias (EC) e dos totais mensais precipitados (ETMP) no período úmido**

	out	nov	dez	jan	fev	mar
EC	-1,24	-1,17	0,22	1,53	0,79	-0,13
ETMP	-1,38	-0,75	0,87	1,56	0,22	-0,51

A Figura 3 resume as estatísticas da Tabela 1. Nela, os eventos das cheias (EC) e dos totais mensais precipitados (ETMP) apresentam proporções diferentes, mas mesmos sinais ao longo do período úmido. Dessa relação pode-se concluir que os totais mensais precipitados e as vazões máximas diárias

anuais (cheias) têm a mesma variabilidade espaço-temporal, no período úmido. Portanto, é possível se utilizar esses totais para representar a variabilidade das cheias e delimitar as sub-regiões hidrologicamente homogêneas.



**Figura 3 Percentuais normalizados de eventos de cheias (Q) e de totais mensais precipitados (TMP)**

Adicionalmente, verificou-se que cerca de 93% dos eventos de cheias ocorrem no período úmido de outubro a março; 71% no trimestre janeiro - março e 22% no trimestre outubro - dezembro.

**Tabela 2 Resultado da AFC após a rotação dos quatro primeiros fatores**

Fator	Autovalor	V.E (%)	V.T.E (%)
1°	81,87	55,3	55,3
2°	12,68	8,5	63,8
3°	5,99	4,1	67,9
4°	3,63	2,5	70,4

Utilizando as equações de (1) a (9) tendo como dados de entrada a matriz 162x148 (seis meses vezes 27 anos e 148 postos pluviométricos), e usando o “Scree” teste como critério para definir o número de fatores (ou sub-regiões hidrologicamente homogêneas), obteve-se a solução inicial de quatro fatores (ou quatro sub-regiões hidrologicamente homogêneas). Adicionalmente, foi necessário se aplicar, na solução inicial, o método de rotação oblíqua desses fatores, pois pelos valores numéricos das cargas ou “pesos” dos fatores da solução inicial, não seria possível distinguir qual carga pertencia a qual fator. Em outras palavras, não seria possível designar, com clareza, os postos pluviométricos que pertenciam, de fato, a cada uma das quatro sub-regiões hidrológica-



Figura 4 - Quatro sub-regiões hidrologicamente homogêneas, segundo o evento cheia, do Estado de São Paulo, exceto a região metropolitana da cidade de São Paulo e o Vale do rio Paraíba do Sul

mente homogêneas. A Tabela 2 resume o resultado da AFC, após a aplicação desse método de rotação.

A contribuição de cada um dos fatores à variância total explicada (VTE) é dada pelos autovalores desses fatores, e é expressa pela equação (7). Essa equação representa a soma quadrática das cargas ou “pesos” de cada um dos quatro primeiros fatores da Tabela 2. Por exemplo, o valor numérico do autovalor do primeiro fator, que é igual a 81, 87, é resultado da soma quadrática dos autovetores, relativos ao primeiro fator e assim sucessivamente. A variância explicada (VE), por cada um dos fatores, é dada pela razão entre o autovalor do fator e o traço da matriz de entrada. Por exemplo, a VE do primeiro fator é dado por  $81,87/148=55,3$  e assim sucessivamente. A VTE expressa a variância total explicada. A variância total explicada pelo método neste estudo é 70,4%.

A rotação de fatores é responsável pela composição linear das variáveis (postos pluviométricos) em torno dos fatores (ou seja, é responsável por designar, cada um dos postos pluviométricos, de acordo com sua carga ou “peso”, em cada um dos quatro primeiros fatores).

As correlações entre os fatores e as variáveis (postos pluviométricos) reforçam as designações ou alocações dos postos pluviométricos em cada um dos fatores. Essas correlações dão o grau de aproximação entre os postos pluviométricos e os fatores. Aqui cerca de 90% dessas correlações apresentaram valores acima de 70%. A Figura 4 mostra as quatro sub-regiões hidrologicamente homogêneas, obtidas pela aplicação da AFC. Desse modo, podem-se utilizar os métodos AFCR, que requeiram a homogeneidade, para estimar os quantis de cheias das quatro sub-regiões.

## CONCLUSÕES

Com base em técnica de Análise Multivariada e em uma nova abordagem para regionalizar cheias foi possível apresentar uma metodologia alternativa aos métodos tradicionais e identificar quatro sub-regiões hidrologicamente homogêneas, suscetíveis à ocorrência de eventos de cheia similares. Surgiu uma dificuldade inicial para se delimitar as sub-regiões, visto que os valores das variâncias intragrupos são baixos. Esses baixos valores são resultados da ótima distribuição espaço-temporal dos totais mensais precipitados do Estado de São Paulo. Essa dificuldade foi contornada, em parte, pela aplicação da Rotação Oblíqua à solução inicial. Mesmo após a aplicação dessa Rotação, a VTE explicou apenas 70,4% da variância total do sistema e deixou de explicar, portanto, 29,6 %.

Considerando que os baixos valores das variâncias intragrupos é um fator limitante para a aplicação da AFC, o autor sugere que se utilize a estatística multivariada denominada Análise de Agrupamento (“Cluster Analysis”) a fim de que se possa obter resultados (delimitações de sub-regiões hidrologicamente homogêneas) com melhores definições e maior facilidade.

## REFERÊNCIAS

- ABRAHAMS, A.D. Factor analysis of drainage basin properties: Evidence for stream abstraction accompanying the degradation of relief. **Water Resources Research**, v.8, n.3, p.624-633, 1972.
- ACREMAN, M. C. Predicting the mean annual flood from basin characteristics in Scotland. **Hydrological Sciences Journal**, v.30, n.1, p.37-49, 1985.
- ACREMAN, M.C.; SINCLAIR, C.D. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. **Journal of Hydrology**, v.84, p.365-380, 1986.
- CATTEL, R. B. **The Scientific use of Factor Analysis in Behavioral and Life Sciences**. New York, Plenum Press, 618p., 1978.
- CHIANG, SHIH-MIN; TSAY, TING-KUEI; J. NIX, STEPHAN. hydrologic regionalization of watersheds. I: Methodology development, **Journal Water Resources Planning and Management**, v.128, n. 1, p. 3-11, 2002a.
- CHIANG, SHIH-MIN; TSAY, TING-KUEI; J. NIX, STEPHAN. Hydrologic regionalization of watersheds. II: Applications, **Journal Water Resources Planning and Management**, v. 128, n. 1, p. 12-20, 2002b.
- CUNNANE, C. Methods and merits of regional flood frequency analysis. **Journal of Hydrology**, v.100, p.269-290, 1998.
- DAEE. Atlas pluviométrico do Estado de São Paulo. São Paulo, 84p, 1970.
- FILL, H.D. **Improving flood quantile estimates using regional information**. Cornell, 1994. 256p. Thesis (Doctor of Philosophy) -Faculty of the Graduate School of Cornell University.
- FRIEL, M. CHARLES. Notes on Factor Analysis. **Criminal Justice Center**, Sam Houston State University, 45p., 1997, Texas- EUA. www.shsu.edu.
- HARMAN, H. H. **Modern Factor Analysis**. Chicago, University of Chicago Press, 469p, 1960.
- KARL, T. R.; KOSCIELNY, A. J. Drought in the United States: 1895-1981. **Journal of Climatology**, v.2, p.313-329, 1982.
- MATALAS, N.C.; REIHER, B.J. Some comments on the use of factor analyses. **Water Resources Research**, v.3, n.1, p.213-223, 1967.
- MIMIKOU, M. ; KAEMAKI, S. Regionalization of flow duration characteristics. **Journal of Hydrology**, v.82, p.77-91, 1985.
- MOSLEY, M. P. Delimitation of New Zealand hydrologic regions. **Journal of Hydrology**, v.49, p.173-192, 1981.
- MURDOCH, D. C. **Álgebra Linear**. Rio de Janeiro, Livros Técnicos e Científicos Editora LTDA, 310p., 1972.
- NATHAN, R. J; McMAHON, T. A. Identification of homogeneous regions for the purposes of regionalization. **Journal of Hydrology**, v.121, p.217-238, 1990.
- RAO, A. R.; HSIEH, C. H. Estimation of variables at ungaged locations by empirical orthogonal functions. **Journal of hydrology**, v.123, p.51-67, 1991.
- RAO, A. RAMACHANDRA; SRINIVAS, V. V. Regionalization of watersheds by hybrid-cluster analysis, **Journal of Hydrology**, v. 318, n. 1-4, p. 37-56, 2006.
- RIBEIRO-CORRÉA, J.; CAVADIAS, G. S.; CLÉMENT, B.; ROUSSELLE, J. Identification of hydrological neighborhoods using canonical correlation analysis, **Journal of Hydrology**. v. 173, n. 1-4, p. 71-89, 1995.
- RICHMAN, M. B. Rotation of principal components. **Journal of Climatology**, v.6, p.293-335, 1986.
- RICHMAN, M. B. Obliquely rotated principal components: an improved meteorological map typing technique? **Journal Applied Meteorology**, v.20, p.1145-1159, 1981.
- RUMMEL, R. J. **Applied Factor Analysis**. Hawaii, Evanston: Northwestern University Press, 617p, 1970.
- STIDD, C. K. The use of eigenvectors for climatic estimates. **Journal Applied Meteorology**, v.6, p.255-271, 1967.

- THURSTONE, L. L. **Multiple Factor Analysis**. Chicago, The University of Chicago Press, 535p, 1965.
- WALSH, J. E. et al. Spatial coherence of monthly precipitation in the United States. **Monthly Weather Review**, v.110, p.272-285, 1982.
- WILTSHIRE, E. S. Regional flood frequency analysis II: Multivariate classification of drainage basins in Britain. **Hydrological Sciences Journal**, v.31, n.3, p.335-346, 1986c.
- WHITE, E.L. Factor analysis of drainage basin properties: classification of flood behavior in terms of basin geomorphology. **Water Resources Bulletin**, v.11, n.4, p.676-687, 1975.
- WONG, S. T. A dimensionally homogeneous and statistically optimal model for predicting mean annual flood. **Journal of Hydrology**, v.42, p.269-279, 1979.

### ***Use of Multivariate Analysis in Flood Regionalization in São Paulo State***

#### **ABSTRACT**

*In hydrology flood regionalization methods or practices are traditionally related to the physical characteristics of river basins. Sub-regions are commonly defined hydrologically, based on these physical features. The disadvantage of this method is that, in order to achieve the best results, it depends on the spatial distribution of stream gauging. This study aims at proposing a multivariate technique as an alternative to circumscribe sub-regions with a homogeneous rainfall pattern, based on total amount of monthly rainfall, and then regionalize the floods in those sub-regions. One of the advantages of the method is that it employs the variables with spatial and temporal perspectives, and not only spatial, as in traditional flood regionalization methods. Using this approach, the method does not depend on spatial distribution of stream gauging stations. An area of São Paulo state was chosen to apply this method. The result shows that the four circumscribed sub-regions accounted for 70.4 % of the variance of the total system. It is believed that the result could have been better if the inter-group variance were greater.*

*Key-words: multivariate analysis; flood regionalization.*