

Identificação de Regiões Homogêneas na Análise Regional de Precipitações e de Vazões Máximas na Bacia do Rio São Francisco (Minas Gerais)

Bruno Rabelo Versiani e Rafael de Melo Franco Carneiro

Depto. de Engenharia Hidráulica e Recursos Hídricos – EE-UFMG – Av. do Contorno, 842/8º andar
30110-060 Belo Horizonte, MG - Fone (31) 238-1055 - versiani@ehr.ufmg.br

Recebido: 14/08/00 - revisão: 12/12/00 - aceito: 31/05/01

RESUMO

O método de regionalização TCEV é aplicado à bacia hidrográfica do rio São Francisco em Minas Gerais. Este é o caso de aplicação em uma bacia de grande superfície onde os dados de precipitação e de vazão máxima disponíveis não são distribuídos de maneira uniforme na área, mas são de boa qualidade. Há, desta forma, uma grande heterogeneidade quanto à formação hidrometeorológica disponível na região, tanto no espaço quanto no tempo. O modelo estatístico TCEV é baseado no produto de duas exponenciais, cada uma representando um processo de Poisson: o primeiro corresponde aos eventos mais frequentes e o outro aos eventos mais raros (os outliers). Na aplicação, duas regiões homogêneas são identificadas, dividindo a bacia hidrográfica em parte sul e parte norte, tanto para precipitações diárias máximas anuais quanto para vazões máximas anuais. Estas regiões foram confirmadas por análise de amostragem e por comparação com a metodologia de regionalização por momentos-L. No caso de precipitações máximas, a definição de regiões homogêneas reflete as características fisiográficas e climatológicas da bacia hidrográfica.

Palavras-chave: vazão máxima; região homogênea; precipitação.

INTRODUÇÃO

Um dos problemas clássicos que se apresentam em estudos relacionados aos recursos hídricos é a escassez de dados, motivada pela ausência de informações ao longo do tempo ou em dadas regiões da bacia hidrográfica.

A regionalização hidrológica consiste na utilização de um conjunto de ferramentas que buscam formas de transferência de informações de um local para outro da bacia hidrográfica, com o intuito de suprir esta carência de informações numa dada região. As diversas fases e procedimentos de um estudo de regionalização são resumidas por Tucci (1993). Uma das técnicas é a análise de frequência regional, que parte do estudo da frequência dos eventos hidrológicos, buscando um modelo estatístico que melhor se ajuste aos dados de diversos pontos da região estudada, para estimativa da variável hidrológica em locais sem informações. A análise de frequência regional “substitui o tempo pelo espaço”, ou seja: dados de diversos locais são usados para estimar frequências de eventos em um só local.

No presente trabalho desenvolve-se um estudo de análise de regionalização de precipitações e

de vazões máximas na bacia hidrográfica do rio São Francisco, no Estado de Minas Gerais. A bacia hidrográfica do rio São Francisco é uma área estratégica no tocante ao aproveitamento de recursos hídricos no Brasil. Na seção da barragem de Três Marias encontra-se um dos principais complexos hidroenergéticos da região. A operação do reservatório tem como objetivos a produção de energia elétrica, o controle de cheias a jusante e a navegação, e procura compatibilizar os usos múltiplos da água. Portanto estudos relacionados a eventos hidrológicos extremos, tais como precipitações e vazões máximas, são de grande interesse para a gestão de recursos hídricos na bacia.

Com o objetivo de análise do método de regionalização TCEV (*Two Component Extreme Value*) e dos efeitos da heterogeneidade espacial e temporal da informação disponível na distribuição regional das precipitações e das vazões extremas, o presente trabalho procura:

- i. descrever sucintamente as metodologias de regionalização pela TCEV e por momentos-L;
- ii. analisar o modelo estatístico TCEV ajustado a estações isoladas, destacando-se a presen-

- ca de observações *outliers* nas séries disponíveis;
- iii. analisar o problema de identificação de regiões homogêneas, e sua implicação nos métodos de regionalização e;
- iv. comparar os resultados da aplicação do método de regionalização TCEV com a metodologia de regionalização por momentos-L.

O MODELO TCEV

Para a descrição estatística de eventos extremos, o processo composto de Poisson é considerado por vários autores como um modelo válido para uma série de valores de K variáveis Z_j ($j = 1, \dots, K$) independentes e identicamente distribuídas (i.i.d.), que excedem um dado nível de base z_0 , dado que z_0 seja suficientemente grande. Este modelo de Poisson considera os valores máximos:

$$X = X_j = \max Z_j, \quad 1 \leq j \leq K \quad (1)$$

dado que a variável aleatória K siga a função de distribuição de Poisson. A função de distribuição acumulada (FDA) para o processo X, definido pela Equação (1) foi descrita por Todorovic e Zelenhasic (1970) como:

$$F_X(x) = P[X \leq x] = \exp\{-\Lambda[1 - F_Z(x)]\} \quad (2)$$

onde $\Lambda = E\{K\}$ é o parâmetro do processo de Poisson e $F(\cdot)$ é a FDA comum para todos os Z_j . É interessante notar que, supondo Z uma variável aleatória exponencial, temos:

$$F_Z(z) = 1 - \exp(-z/\theta), \quad z \geq 0 \quad (3)$$

com $\theta = E\{Z\}$, portanto a Equação (2) passa a ser:

$$F_X(x) = \exp[-\Lambda \exp(-x/\theta)] \quad (4)$$

A Equação (4) é a forma da FDA assintótica de valores extremos do tipo 1 (EV1) ou FDA de *Gumbel*:

$$F_X(x) = \exp\{-\exp[-(x - x_0)/\theta]\} \quad (5)$$

com

$$x_0 = \theta \ln \Lambda \quad (6)$$

onde θ é o parâmetro de escala (o *gradex*) e x_0 é o parâmetro de locação (a *moda*). O parâmetro *gradex* (θ) é um parâmetro básico importante, utilizado no *Método do Gradex*, descrito por Duband e Garros-Berthet (1994) para estimar vazões de cheia a partir de precipitações máximas.

Waylen e Woo (1982) estudaram uma bacia onde vazões máximas deveriam ser consideradas como provenientes de dois processos hidrológicos (vazões geradas por precipitações e vazões geradas por fusão de neve), sendo que as subpopulações correspondentes possuem características estatísticas diferentes. Neste caso, o modelo TCEV, proposto por Rossi et al. (1984), considera que existem duas séries independentes de variáveis $Z_{1,j}$, ($j = 1, \dots, K_1$) e $Z_{2,j}$ ($j = 1, \dots, K_2$) i.i.d., geradas por dois processos de Poisson, com os respectivos parâmetros $\Lambda_1 = E\{K_1\}$ e $\Lambda_2 = E\{K_2\}$, sendo $\Lambda_1 > \Lambda_2$. Definindo-se X_1 e X_2 como máximos anuais de $Z_{1,j}$ e $Z_{2,j}$, respectivamente, e se $Z_{1,j}$ e $Z_{2,j}$ são variáveis exponencialmente distribuídas, com parâmetros $\theta_1 = E\{Z_1\}$ e $\theta_2 = E\{Z_2\}$, teremos:

$$F_x(x) = \exp[-\Lambda_1 \exp(-x/\theta_1) - \Lambda_2 \exp(-x/\theta_2)], \quad x \geq 0 \quad (7)$$

A Equação (7) define o modelo TCEV (*Two Components Extreme Value*), com as seguintes condições:

$$x \geq 0, \quad \theta_2 \geq \theta_1 \geq 0, \quad \Lambda_1 \geq \Lambda_2 \geq 0.$$

Os parâmetros θ_1 e Λ_1 descrevem os eventos mais frequentes (a série básica), enquanto θ_2 e Λ_2 descrevem os eventos raros (a série dos *outliers*). Recentemente Francès (1998) mostrou que um modelo não tradicional como a TCEV pode também descrever o comportamento das vazões máximas em rios da costa mediterrânea da Europa, destacando valores extremos (*outliers*) obtidos por informações não-sistemáticas.

A função de distribuição TCEV é um modelo de quatro parâmetros ($\Lambda_1, \Lambda_2, \theta_1, \theta_2$) equivalente ao produto de duas FDA de *Gumbel*:

$$F_x(x) = \exp\{-\exp[-(x - x_{01})/\theta_1]\} \cdot \exp\{-\exp[-(x - x_{02})/\theta_2]\} \quad (8)$$

com

$$x_{0j} = \theta_j \ln \Lambda_j, \quad (j = 1, 2).$$

Uma descrição detalhada das propriedades estatísticas do modelo TCEV, com a definição dos momentos e das probabilidades inerentes aos *outliers*, foi apresentada por Beran et al. (1986) e Arnell e Gabriele (1988). Os parâmetros θ_* e Λ_* , usados no método de regionalização, são definidos por:

$$\theta_* = \frac{\theta_2}{\theta_1} \tag{9}$$

$$\Lambda_* = \frac{\Lambda_2}{\Lambda_1^{(1/\theta_*)}} \tag{10}$$

MÉTODO TCEV DE REGIONALIZAÇÃO

O modelo estatístico TCEV pode ser aplicado em estações isoladas, mas em numerosos casos foram encontradas dificuldades na identificação de seus quatro parâmetros, Arnell e Gabriele (1988) e Fiorentino et al. (1987a). O estudo de estações isoladas faz parte de uma primeira etapa de análise de regionalização hierárquica, aplicada em regiões da Itália e da Inglaterra por diversos autores como Cannarozzo et al. (1995) e Fiorentino et al. (1987b), que descrevem uma síntese dos trabalhos de pesquisa sobre o método de regionalização baseado na utilização da TCEV.

De acordo com os autores, o método é desenvolvido em dois níveis hierárquicos. O primeiro nível consiste em se procurar as regiões homogêneas, as quais são caracterizadas por coeficientes de assimetria pouco variáveis, de tal forma que podem ser considerados como constantes.

Considerando-se I séries de máximos numa região considerada homogênea, podemos definir a variável *standard*:

$$Y = Y_{ij} = \frac{X_{ij}}{\theta_{1i}} - \ln \Lambda_{1i}, \tag{11}$$

$$i = 1, \dots, I \quad ; \quad j = 1, \dots, n_i$$

com I igual ao número de estações (ou de séries); n_i o número de dados da série i e; θ_{1i} , Λ_{1i} os parâmetros da componente básica da série i.

A variável Y é uma variável aleatória i.i.d. segundo a função de distribuição TCEV dada por:

$$F_y(y) = P(Y \leq y) = \exp[-e^{-y} - \Lambda_* e^{-y/\theta_*}] \tag{12}$$

com os parâmetros θ_* e Λ_* dados pelas Equações (9) e (10).

Os autores mostram que, se a variável Y é distribuída segundo a Equação (12), o coeficiente de assimetria teórico é único em toda a região e que as relações entre os parâmetros θ_{2i} e θ_{1i} (entre a segunda e a primeira componentes) são constantes para todas as estações que pertencem a esta região.

O segundo nível de regionalização consiste na pesquisa de sub-regiões homogêneas, ou seja onde os coeficientes de variação são pouco variáveis, o que implica em que os parâmetros Λ_1 da TCEV sejam constantes nas respectivas sub-regiões. No interior de cada região homogênea existem sub-regiões homogêneas.

Aplicação do modelo TCEV ajustado

Depois de ajustado o modelo TCEV para regiões e sub-regiões, uma aplicação interessante é a estimativa de valores máximos para a variável X, com período de retorno T (x_T), quando dispõe-se de somente alguns anos de observação em uma estação pertencente a uma sub-região, dado que esta estação não foi usada no processo de calibração. Neste caso tem-se os parâmetros θ_* , Λ_* e Λ_1 da TCEV e quer-se estimar x_T . Tomando:

$$X' = \frac{X}{E\{X\}}, \tag{13}$$

Gabriele e Iiritano (1994) definiram a "Curva de Crescimento" dada por:

$$F(x') = \exp\{-\Lambda_1 \exp(-\eta x') - \Lambda_* [\Lambda_1^{(1/\theta_*)}] \exp(-\eta x' / \theta_*)\} \tag{14}$$

$$x' > 0 ; \Lambda_1 > 0 ; \Lambda_* > 0 ; \theta_* > 0$$

com:

$$\eta = \ln \Lambda_1 + 0,5772 - \sum_{j=1}^{\infty} \frac{(-1)^j \Lambda_*^j}{j!} \Gamma(j/\theta_*) \tag{15}$$

A FDA $F(x')$ depende somente dos parâmetros θ_* , Λ_* e Λ_1 obtidos no primeiro e segundo níveis de regionalização. Desta forma pode-se estimar o quantil x'_T invertendo-se a Equação (14) e calcular x_T como:

$$x_T = x_T' \cdot \bar{x} \quad (16)$$

onde \bar{x} é a estimativa local da média da série de máximos disponível.

METODOLOGIA DOS MOMENTOS-L PARA ANÁLISE DE FREQUÊNCIA REGIONAL

Introdução

Uma das técnicas usadas atualmente para análise de frequência regional foi desenvolvida por Hosking e Wallis (1997) e utiliza os momentos-L na definição de regiões homogêneas e da função de distribuição regional.

A maior parte dos processos de análise de frequência regional tende a ajustar os dados a uma função de distribuição cuja forma é especificada de acordo com um número finito de parâmetros indeterminados. Os momentos estatísticos amostrais convencionais, particularmente assimetria e curtose, são normalmente usados para verificar o ajuste de uma amostra de dados observados a uma função de distribuição dada. Entretanto estas estatísticas, em alguns casos, são insatisfatórias, pois especialmente em pequenas amostras, os valores calculados não são boas estimativas dos valores populacionais. Estes autores recomendam, portanto, uma abordagem alternativa, baseada nas estatísticas dos momentos-L.

Conceitos básicos sobre os momentos-L

Os momentos-L são um sistema alternativo de descrição de formas de funções de distribuição de probabilidades.

A estimativa dos momentos-L, a partir de uma amostra finita de tamanho n , inicia-se com a ordenação de seus elementos constituintes em ordem crescente, ou seja $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$. Um estimador não-enviesado, ou não tendencioso, dos momentos ponderados β_r pode ser escrito como:

$$b_r = \hat{\beta}_r = \frac{1}{n} \cdot \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (17)$$

Os momentos-L amostrais são definidos pelas seguintes expressões:

$$\ell_1 = b_0 \quad (18)$$

$$\ell_2 = 2b_1 - b_0 \quad (19)$$

$$\ell_3 = 6b_2 - 6b_1 + b_0 \quad (20)$$

$$\ell_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \quad (21)$$

Da mesma forma, os quocientes de momentos-L amostrais são dados por:

$$t_r = \frac{\ell_r}{\ell_2}; \quad r \leq 3 \quad (22)$$

enquanto o L-CV amostral calcula-se através de:

$$t = \frac{\ell_2}{\ell_1} \quad (23)$$

MÉTODO DE REGIONALIZAÇÃO BASEADO NOS MOMENTOS-L

No contexto da análise de frequência regional utilizando os momentos-L, Hosking e Wallis (1997) mostram que informações bastante úteis podem ser obtidas comparando-se os quocientes de momentos-L amostrais dos diferentes postos de observação. Valores incorretos de dados, *outliers*, tendências, e alterações na média sempre podem ser refletidos nos momentos-L de uma amostra e podem ser detectados pela *medida de discordância* a seguir definida.

Medida de discordância

Supondo que existam N postos de observação. Seja um vetor $u_i = [t^{(i)} \ t_3^{(i)} \ t_4^{(i)}]^T$ contendo os valores dos quocientes de momentos-L amostrais t , t_3 e t_4 de um posto i : o subscrito T denota a transposição de um vetor ou matriz. Seja:

$$\bar{u}_i = N^{-1} \sum_{i=1}^N u_i \quad (24)$$

a média do grupo (não ponderada). Define-se a matriz A de soma de quadrados e produtos cruzados como:

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (25)$$

A medida de discordância de um posto i é definida como:

$$D_i = \frac{1}{3} N(u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \quad (26)$$

Os autores recomendam que qualquer posto cujo valor de D_i for maior que 3 seja considerado como discordante, tendo este posto quocientes de momentos-L diferentes da média dos outros postos da região.

Identificação de regiões homogêneas

De todas as etapas de uma análise de frequência regional envolvendo vários postos de observação, a identificação de regiões homogêneas é normalmente a mais difícil e requer uma grande soma de julgamentos subjetivos. Vários autores, dentre os quais NERC (1975), Beable e McKerchar (1982), Matalas et al. (1975), Whiltshire (1986), e GREHYS (1996) tentaram desenvolver métodos no intuito de diminuir esta subjetividade. O objetivo é formar grupos de postos que aproximadamente satisfazem as condições de homogeneidade, ou seja, que as funções de distribuição destes postos sejam idênticas, a menos de um fator de escala específico de cada posto.

Para classificação das regiões homogêneas pelo método dos momentos-L, foi utilizada a técnica proposta por Hosking e Wallis (1997) conforme o procedimento seguinte:

- i. análise de *clusters* para precipitações máximas e vazões máximas;
- ii. cálculo dos momentos-L e da medida de discordância para todos os postos de cada *cluster*;
- iii. cálculo das medidas de heterogeneidade e de aderência para as regiões definidas pela análise de *clusters*.

Análise de "clusters"

A análise de *clusters* é um método de estatística multivariada para divisão de uma série de dados em grupos e tem sido usada com sucesso na determinação de regiões, na análise de frequência regional. Resumidamente, um vetor de dados é associado a cada estação e as estações são agregadas em grupos de acordo com a similaridade de cada vetor. Estes vetores podem ser formados por estatísticas das estações, características das estações ou alguma combinação entre as duas.

Medida de heterogeneidade

Nesta fase do procedimento, o objetivo é estimar o grau de heterogeneidade em um grupo de estações e verificar quais estações podem ser razoavelmente tratadas como pertencentes a uma região homogênea. Especificamente, a medida de heterogeneidade, H, compara a variação dos momentos-L amostrais com momentos-L simulados através de uma distribuição teórica. Os autores recomendam como função de distribuição teórica o modelo Kappa de quatro parâmetros para simulação dos dados, no cálculo de H. O modelo Kappa é dado pela Equação (27), onde os parâmetros são ξ (locação), α (escala), k e h:

$$F(x) = \left[1 - h \left\{ 1 - k(x - \xi) / \alpha \right\}^{1/k} \right]^{1/h} \quad (27)$$

A medida de heterogeneidade é calculada como:

$$H = \frac{(V - \mu_V)}{\sigma_V} \quad (28)$$

onde V é o desvio padrão ponderado amostral para L-CV, μ_V a média e σ_V o desvio padrão para os valores de V simulados.

A região será considerada heterogênea se H for suficientemente grande. Os autores sugerem que a região seja considerada "aceitavelmente homogênea" se $H < 1$, "possivelmente heterogênea" se $1 \leq H \leq 2$, e "definitivamente heterogênea" se $H > 2$.

Medida de aderência

A medida de aderência sugerida pelos autores permite a escolha, dentre um número de distribuições de frequência candidatas, aquela que melhor se ajusta aos dados. Trata-se de um procedimento que utiliza diretamente as estatísticas de médias de momentos-L regionais, e é denotada por Z. O valor de Z permite verificar se a distribuição candidata fornece um bom ajuste aos dados regionais. Este valor, calculado para cada distribuição candidata, é função da relação entre a L-curtose da distribuição ajustada, a média regional da L-curtose amostral e do desvio padrão da L-curtose simulada. Aceita-se a distribuição candidata como adequada se $|Z| \leq 1,64$. Segundo os autores, a estatística Z tem a forma de um teste de significância para a aderência e segue uma distribuição normal. Neste caso, o critério corresponde à aceitação da função de distribuição ao nível de confiança de 90%.

ESTUDO NA BACIA HIDROGRÁFICA DO ALTO SÃO FRANCISCO

Área de estudo

A bacia hidrográfica do Alto São Francisco, na seção da barragem de Três Marias, é mostrada na Figura 1. Trata-se da sub-bacia 40, que compreende as nascentes do rio São Francisco, Pará, Parapeba, Indaiá, Borrachudo e seus afluentes até a barragem de Três Marias, abrangendo cerca de 100 municípios. Os totais pluviométricos anuais variam, em geral, entre 1700 mm nas cabeceiras, parte sul, e 1150 mm na região próxima à barragem, ao norte.

Esta zona de estudo, com uma área de 50870 km², encontra-se na região sudeste do Brasil, com características meteorológicas do clima da região central do Estado de Minas Gerais: o trimestre mais chuvoso (novembro - dezembro - janeiro), que corresponde às estações do ano onde ocorrem os valores máximos, contribui com 55 a 60% do total anual precipitado e o mais seco (junho - julho - agosto) com menos de 5% da precipitação anual. As tipologias climáticas de Köppen encontradas na região são Cwb, Cwa e Aw. A rede de observações compreende 92 pluviômetros e 55 fluviógrafos, que apresentam dados de boa qualidade, mas não uniformemente distribuídos na região. Os dados utilizados neste estudo consistem de séries de precipita-

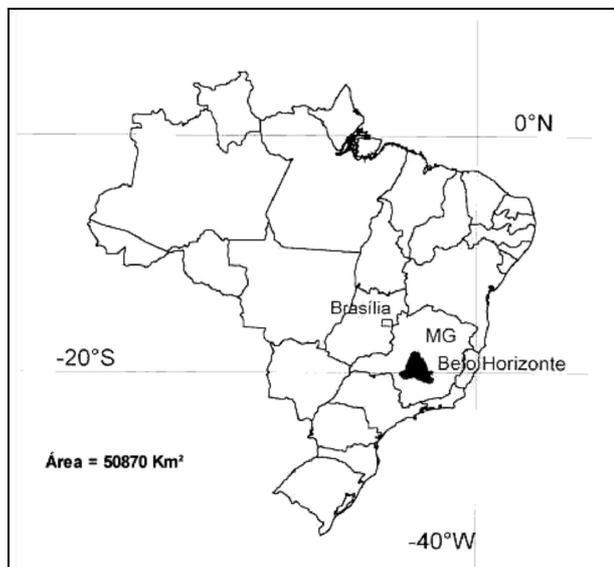


Figura 1. Localização da área de estudo.

ções diárias máximas anuais e de vazões máximas anuais, fornecidos pela CPRM (Companhia de Pesquisa em Recursos Minerais, Serviço Geológico do Brasil), relativos ao período de 1939 a 1998. Na maior parte dos casos, as séries de precipitações máximas são curtas (57% com menos de 20 anos de observações) e somente algumas estações têm 35 a 48 anos. No caso das vazões máximas, 30% das séries têm menos de 20 anos de observações. Portanto, há uma grande heterogeneidade da informação disponível sobre a bacia hidrográfica, tanto no tempo quanto no espaço.

Aplicação do modelo TCEV em estações isoladas

Os ajustes das funções de distribuição de Gumbel e TCEV aos dados observados foram comparados, para todas as estações da região de estudo. A estimativa dos quatro parâmetros da TCEV foi feita pelo método da máxima verosimilhança, de acordo com Rossi et al. (1984), adotando-se o programa adaptado de Jolley (1987), que aplicou o modelo em regiões do Chile. Dois exemplos de resultados são mostrados na Tabela 1. Esta tabela mostra nas colunas 2 a 8 algumas características estatísticas como o número de iterações para se atingir a convergência (iter), o número de dados, a média, o desvio padrão, os valores máximo e mínimo e o coeficiente de assimetria (g) para precipitação máxima anual (em milímetros) na estação pluviométrica de Caeté (código 1943010) e vazão

Tabela 1. Características estatísticas e parâmetros da TCEV para estações isoladas.

(1) Estação	(2) iter	(3) n	(4) Média	(5) D.P.	(6) min	(7) max	(8) g	(9) Λ_1	(10) Λ_2	(11) θ_1	(12) θ_2	(13) θ^*	(14) Λ^*
1943010	87	37	87,4	28,1	47,2	210,2	2,220	66,194	0,169	17,726	67,829	3,826	0,057
40170000	74	23	112,1	66,2	51,8	340,0	2,238	28,617	0,721	21,933	93,606	4,268	0,328

máxima (em m^3/s) na estação de Marilândia (código 40170000). As colunas de 9 a 12 mostram os parâmetros Λ_1 , Λ_2 , θ_1 , θ_2 da TCEV. As colunas 13 e 14 mostram os parâmetros dados pelas Equações (9) e (10).

Um exemplo da comparação entre o ajuste da TCEV e da distribuição de Gumbel é mostrado nas Figuras 2 e 3, para as estações 1943010 e 4017000. Os dados são plotados no papel probabilístico de Gumbel, com a abscissa dada por $u = -\ln\{-\ln[F(x)]\}$ e ordenadas por precipitação máxima (Figura 2) e vazão máxima (Figura 3). Estas figuras mostram exemplos de estações em que o modelo TCEV é melhor ajustado em relação à distribuição de Gumbel, ressaltando a presença dos *outliers*, principalmente no caso da Figura 3, onde os 5 valores mais fortes de vazão são melhor ajustados. Nestes casos, para valores fortes de u , a curva TCEV é assintótica a uma reta de coeficiente angular θ_2 (o *gradex* dos *outliers*). Da Tabela 1 verifica-se que $\theta^* = \theta_2/\theta_1$ tem valores fortes, o que significa que θ_2 é maior que θ_1 (o *gradex* da série básica), como mostra a Equação (8). Dentre todas as estações analisadas 20% das estações pluviométricas e 25% das fluviométricas apresentam *outliers*. Apesar destas estações encontrarem-se espalhadas sobre toda a bacia, nota-se que existe uma maior concentração de estações com *outliers* na parte oriental, ao longo das sub-bacias dos rios Paraopeba e Pará, afluentes da margem direita do rio São Francisco.

Estimativa dos parâmetros regionais do modelo TCEV

O algoritmo de estimativa de parâmetros regionais do modelo TCEV adotado foi o proposto por Gabriele e Liritano (1994). Trata-se de uma estimativa conjunta de Λ^* , θ^* e Λ_{1s} , $s = 1, \dots, M$, sendo M o número de sub-regiões.

O seguinte procedimento foi adotado:

- i. análise de mapas de isolinhas de coeficientes de assimetria (g) para as regiões de interesse;

- ii. escolha das regiões homogêneas, onde g apresenta pequena variabilidade;
- iii. análise dos mapas de coeficiente de variação (c_v) e;
- iv. escolha das sub-regiões homogêneas, onde c_v apresenta pequena variabilidade.

A Tabela 2 mostra os resultados obtidos para estimativa dos parâmetros regionais da TCEV dos dois conjuntos de dados.

As Figuras 4a e 5a mostram, respectivamente para precipitações máximas e vazões máximas, os limites da bacia, a localização das estações pluviométricas e fluviométricas utilizadas no estudo, em coordenadas UTM.

Mostram também as isolinhas do coeficiente de assimetria, feitas pelo método de krigagem, e a delimitação das regiões e sub-regiões homogêneas.

Para precipitações máximas (Figura 4a) nota-se que a bacia hidrográfica é constituída por duas regiões homogêneas, uma ao sul e outra ao norte, separadas por uma linha curva que parte a leste aproximadamente nas coordenadas $X = 580000$, $Y = 7843000$, até o limite oeste, nas coordenadas $X = 395000$, $Y = 7887000$. Percebe-se também que a região sul possui duas sub-regiões (valores fortes e valores fracos de c_v).

Para vazões máximas (Figura 5a), nota-se que a bacia hidrográfica também é constituída por duas regiões homogêneas, uma ao sul e outra ao norte, separadas por uma linha curva que parte a leste aproximadamente nas coordenadas $X = 575000$, $Y = 7840000$, até o limite oeste, nas coordenadas $X = 380000$, $Y = 7845000$. A região sul também possui duas sub-regiões (valores fortes e valores fracos de c_v).

Tanto para precipitações quanto para vazões, obtiveram-se núcleos de valores fortes e de valores fracos para os c_v coincidentes com os obtidos para g .

Nota-se que existe uma coincidência aproximada na delimitação das regiões homogêneas para precipitações e para vazões (uma ao norte e

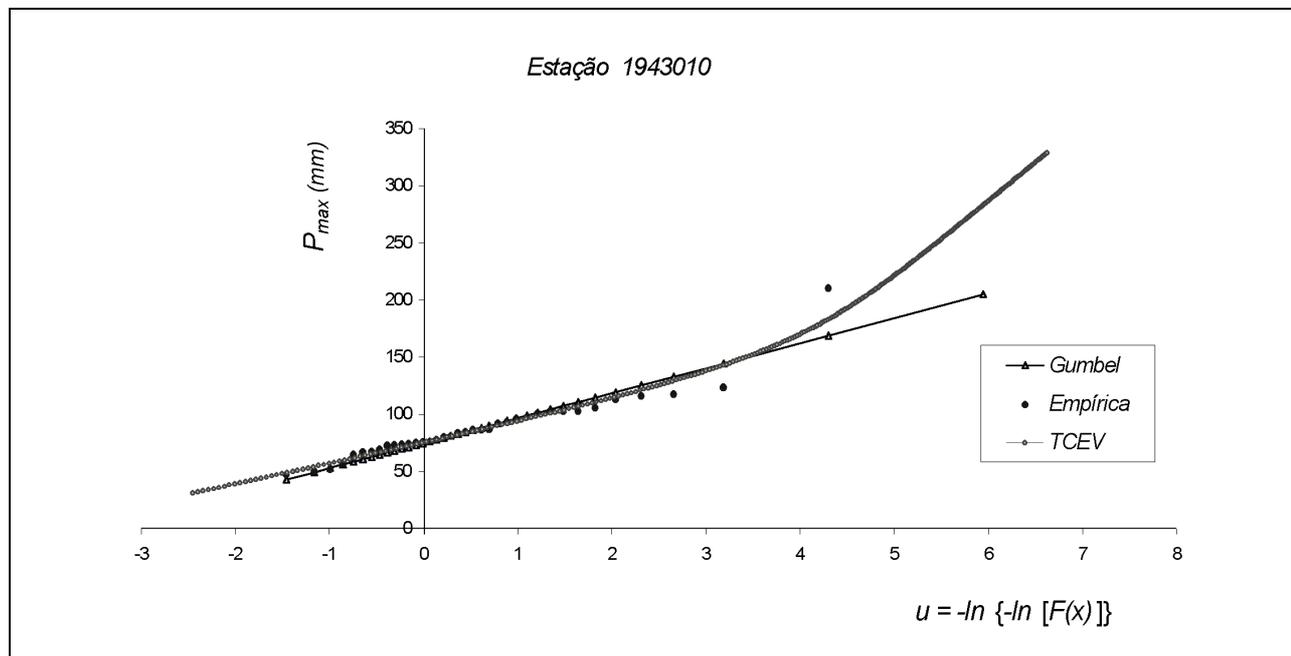


Figura 2. Comparação do ajuste entre o modelo TCEV e o modelo de Gumbel para a estação 1943010 (precipitação).

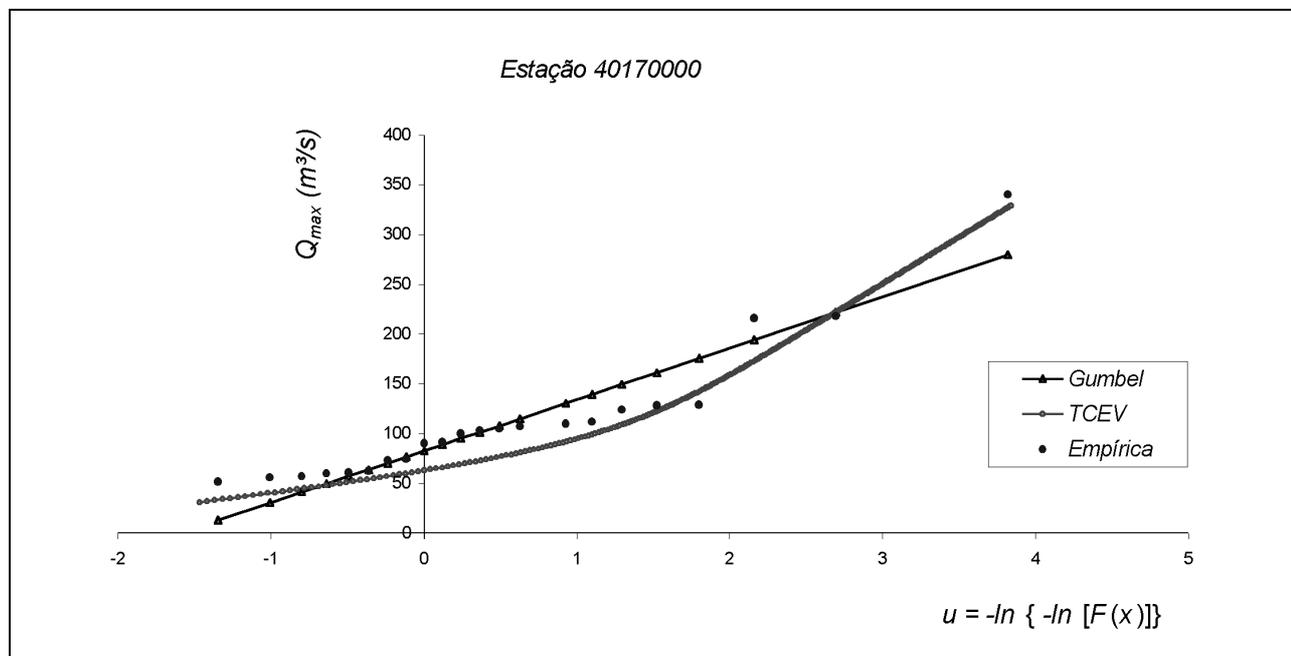


Figura 3. Comparação do ajuste entre o modelo TCEV e o modelo de Gumbel para a estação 40170000 (vazão).

Tabela 2. Parâmetros regionais da TCEV para precipitações e vazões máximas.

Dado	Região	Número de sub-regiões	Número de estações	Número de dados	θ^*	Λ^*	$\Lambda_{1,1}$	$\Lambda_{1,2}$
Precipitação	Norte	1	32	534	1,7570	0,6337	114,4791	-
	Sul	2	60	1384	1,3877	0,8710	53,7622	75,8463
Vazão	Norte	1	13	316	4,5531	0,0285	12,8831	-
	Sul	2	42	1110	2,5769	0,0500	10,3369	21,0542

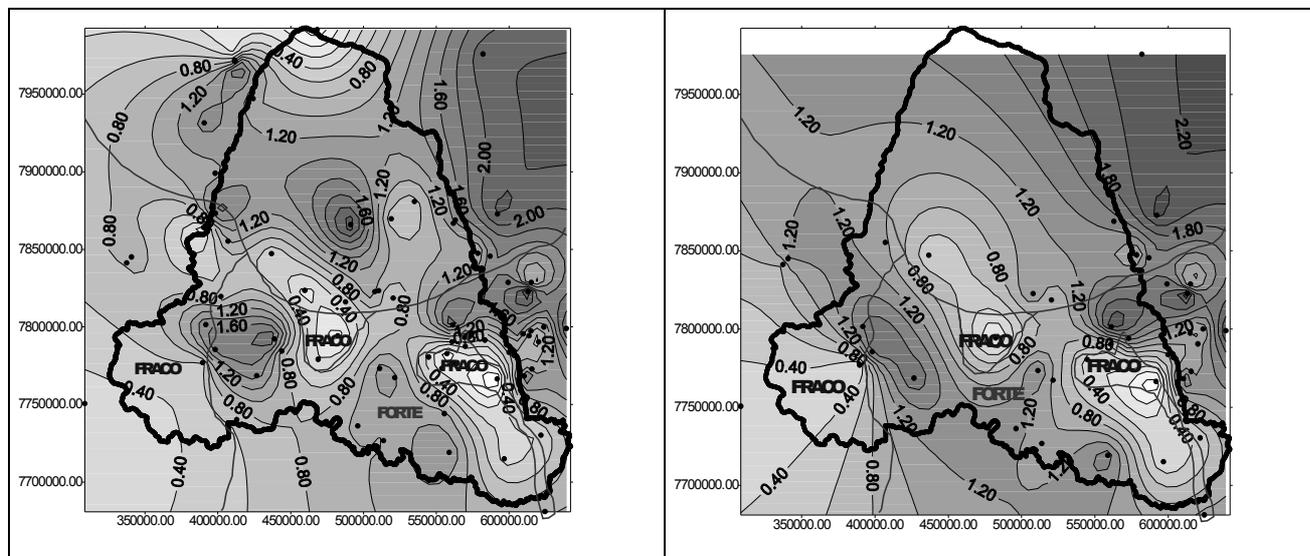


Figura 4. Isolinhas de coeficiente de assimetria (P15 à esquerda e P2028 à direita), precipitação máxima.

outra ao sul da bacia). Entretanto, para as sub-regiões (ao sul da bacia), não existe coincidência na sua delimitação.

Influência do tamanho da amostra na variabilidade espacial dos coeficientes de variação e assimetria

A cartografia do coeficiente de assimetria (g, Figuras 4a, 4b, 5a, 5b) e do coeficiente de variação (c_v , Figuras 5c e 5d) permite caracterizar a informação sobre a variabilidade espacial das precipitações extremas na região de estudo.

Na presente análise, foi feita uma eliminação progressiva de dados das séries de precipitações máximas em cada estação, com a finalidade de verificar se houve alteração significativa na configuração das isolinhas de g e de c_v . Trata-se de um problema clássico de amostragem, para verificação de estabilidade da configuração das isolinhas.

Procedeu-se, portanto, à comparação entre os mapas de isolinhas obtidos com estações pluviométricas com no mínimo 15 anos de observações de precipitações máximas anuais, variando-se o tamanho das amostras até o número máximo disponível de anos de observação, conforme a Tabela 3.

Esta tabela mostra a eliminação progressiva de dados, com a finalidade de se verificar a influência da degradação da informação (diminuição do tamanho das séries) na configuração das isolinhas. Por exemplo: as amostras P2033 são formadas por

aquelas estações pluviométricas que têm, no mínimo, 20 anos de observações de precipitações máximas anuais, sendo que foram eliminadas as observações das séries (estações) disponíveis que superem 33 anos de observações. As amostras F2028 são formadas por aquelas estações pluviométricas que têm, no mínimo, 20 anos de observações de vazões máximas anuais, sendo que foram eliminadas as observações das séries (estações) disponíveis que superem 28 anos de observações.

Um exemplo de comparação entre as isolinhas de coeficiente de assimetria das amostras P15 e P2028 é mostrado nas Figuras 4a e 4b. Para as amostras F15 e F2028 esta comparação é mostrada nas Figuras 5a e 5b.

Da análise dos mapas de isolinhas obtidos pode-se concluir que existe uma variabilidade de amostragem espacial significativa. Entretanto, pode-se observar que apesar da diminuição do tamanho das séries, nota-se a presença de alguns núcleos de valores fortes e de valores fracos de g, o que nos permite confirmar a delimitação de regiões aproximadamente homogêneas com respeito a estes parâmetros estatísticos.

Validação do modelo TCEV de regionalização ajustado

Para validar o modelo de regionalização TCEV, escolheu-se uma estação que não integrou a

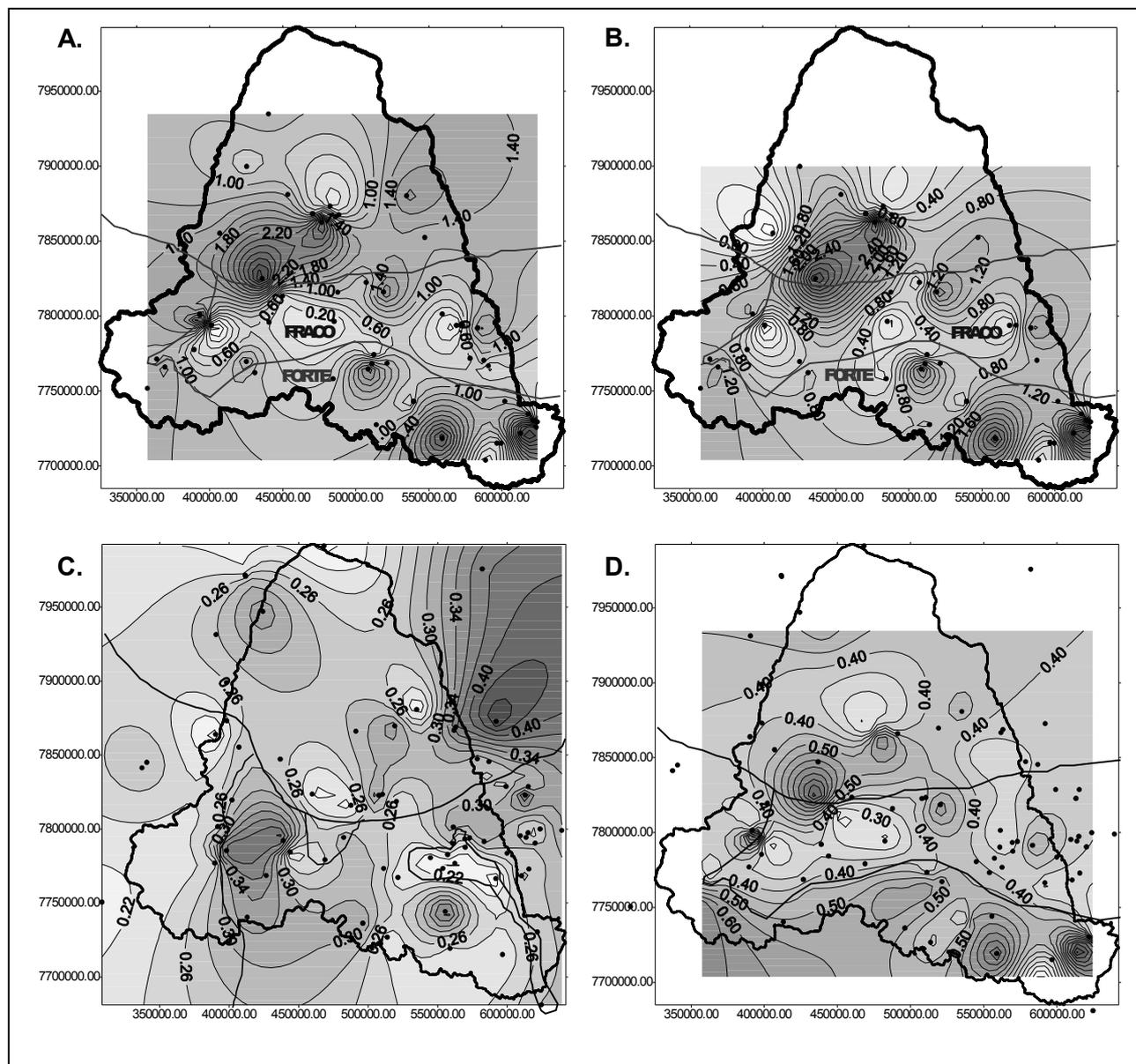


Figura 5. Isolinhas de coeficiente de: A) assimetria (F15), vazão máxima; B) assimetria (F2028), vazão máxima; C) variação (P15), precipitação máxima; e D) variação (F15), vazão máxima.

série usada na calibração da Curva de Crescimento (Equação 14). Esta estação tem 10 anos de observações de vazões máximas (estação Ponte Nova do Paraopeba - código 40800002), situada na região sul e na sub-região 2 ($\Lambda_1 = 21,0542$, $\Lambda^* = 0,0500$, $\theta^* = 2,5769$).

A Figura 6 mostra o papel probabilístico com a comparação entre o modelo de Gumbel, a Curva de Crescimento e os pontos observados. Pode-se ver que a Curva de Crescimento é bem ajustada aos pontos observados e valida o modelo de regionalização TCEV. O modelo de Gumbel é bem ajustado, mas com uma amostra de apenas 10

valores, ao passo que o modelo de regionalização foi ajustado com um número bem maior, pois utilizou a informação regionalizada das estações vizinhas.

Identificação de regiões homogêneas utilizando a análise de momentos-L

A técnica descrita na metodologia dos momentos-L para análise de frequência regional foi aplicada na bacia hidrográfica do Alto São Francisco. Este enfoque permite a realização da análise de

Tabela 3. Tamanho das amostras e dados eliminados de precipitação e vazão.

Precipitação	Vazão	Mínimo número de dados em cada estação (n)	Dados eliminados em cada estação (n_e)	Número de estações (precipitação)	Número de estações (vazão)
P15	F15	$n \geq 15$	-	70	54
P20	F20	$n \geq 20$	-	40	44
P2033	F2033	$n \geq 20$	$n_e > 33$	40	44
P2028	F2028	$n \geq 20$	$n_e > 28$	40	44
P2023	F2023	$n \geq 20$	$n_e > 23$	40	44
P2020	F2020	$n \geq 20$	$n_e > 20$	40	44
P25	F25	$n \geq 25$	-	27	24
P2533	F2533	$n \geq 25$	$n_e > 33$	27	24
P2528	F2528	$n \geq 25$	$n_e > 28$	27	24
P30	F30	$n \geq 30$	-	20	15
P3033	F3033	$n \geq 30$	$n_e > 33$	20	15

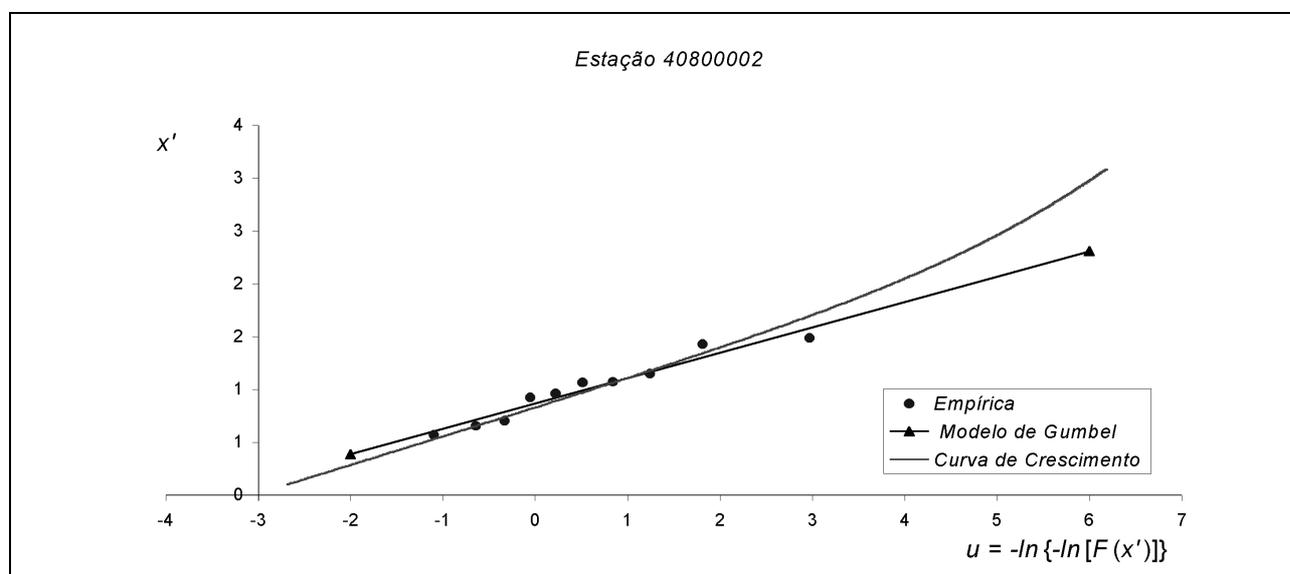


Figura 6. Comparação entre a curva de crescimento e o modelo de Gumbel.

clusters, estimativa dos momentos-L para todas as estações e cálculo das medidas de heterogeneidade e aderência, objetivando a identificação de regiões homogêneas. Os programas utilizados nestas etapas são parte do pacote LMOMENTS, gratuitamente disponibilizado eletronicamente pelos autores Hosking e Wallis (1997).

Da análise de *clusters* foram pré-definidos dois agrupamentos de estações, para os dois tipos de dados (precipitações e vazões máximas): uma região na parte norte da bacia e outra na parte sul.

Para as precipitações, a região norte reuniu 34 estações e a região sul, 55 estações (três estações foram eliminadas da região sul pelo critério de medida de discordância). Para as vazões, cada regi-

ão reuniu 26 estações (uma estação da região norte e duas estações da região sul foram eliminadas pelo critério de medida de discordância).

A Tabela 4 mostra os resultados obtidos do cálculo das medidas de heterogeneidade e aderência, através dos momentos-L e dos quocientes de momentos-L estimados na análise de frequência regional. Podemos observar que no caso das precipitações máximas existem duas regiões aceitavelmente homogêneas (regiões norte e sul). Entretanto a medida de aderência foi verificada somente na região sul ($|Z| \leq 1,64$), sendo que a função de distribuição melhor ajustada foi a *Generalizada Logística*. No caso das vazões máximas não foi possível identificar regiões homogêneas utilizando o método dos momentos-L ($H > 1,0$).

Tabela 4. Medidas de heterogeneidade e aderência para precipitações e vazões máximas.

Dados	Região	Número de estações	Medida de heterogeneidade (H)	Medida de aderência (Z)	Distribuição regional
Precipitação	Norte	34	-0,99	-4,60	Gen. Logística
	Sul	55	-1,87	-0,54	Gen. Logística
Vazão	Norte	26	3,23	-0,83	Gen. Logística
	Sul	26	3,75	0,42	G.E.V.

A Figura 7 mostra os limites da bacia com a localização das regiões homogêneas identificadas pelos dois métodos (TCEV e momentos-L). Pode-se observar que os limites das regiões definidas pelos dois métodos coincidem em quase sua totalidade, para o caso das precipitações máximas.

CONCLUSÕES E DISCUSSÃO

O modelo TCEV é bem ajustado para algumas estações individuais e destaca a presença dos *outliers*, em ambos os casos para precipitações e vazões máximas.

O método de regionalização baseado no modelo TCEV permitiu a identificação de duas regiões homogêneas que são aproximadamente iguais para as precipitações e vazões. Duas sub-regiões para precipitações e vazões máximas foram delimitadas em cada caso, ambas na respectiva região sul, porém elas não coincidem.

Para as precipitações máximas, as regiões homogêneas identificadas utilizando o método de regionalização baseado nos momentos-L são praticamente as mesmas que foram identificadas pelo método TCEV, satisfeitos os critérios de discordância e heterogeneidade. Pinto e Naghettini (1999) definiram regiões homogêneas para esta bacia hidrográfica baseados em informações fisiográficas e climatológicas. Eles concluíram que a configuração das isoietas de precipitações máximas reflete a topografia da bacia. Os dois núcleos de maior valor para precipitação máxima anual correspondem às maiores elevações na região sul. A região norte da bacia corresponde às menores declividades e elevações, e possui menores valores de precipitação máxima anual. De acordo com as informações fisiográficas e climatológicas eles definiram duas sub-regiões para a região sul, e a localização destas sub-regiões coincide com as definidas neste trabalho pelo método TCEV. Igualmente eles não definiram sub-regiões para a região norte da bacia.

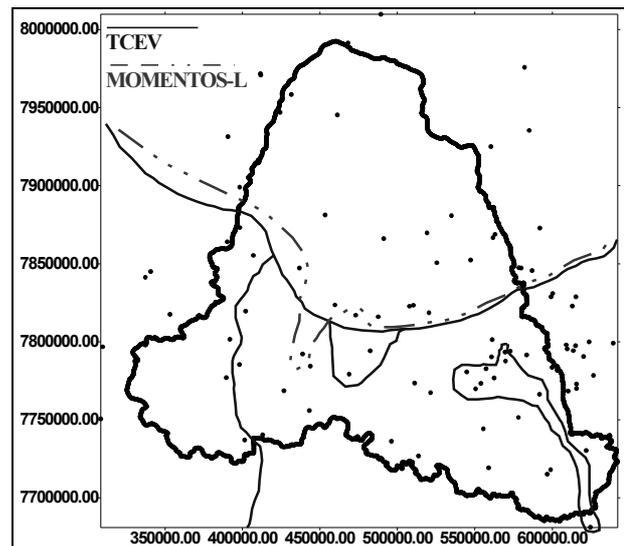


Figura 7. Comparação entre a definição de regiões homogêneas identificadas pelos métodos TCEV e momentos-L, para precipitações máximas.

No caso das vazões máximas não foi possível comparar os resultados da identificação de regiões homogêneas utilizando os dois métodos, pois no método dos momentos-L o critério de heterogeneidade não foi verificado. Entretanto, observa-se que a Curva de Crescimento Regional é bem ajustada na validação do modelo de regionalização TCEV, no exemplo mostrado. O método mostra-se como uma boa ferramenta no estudo de previsão de vazões máximas, inclusive em locais que não apresentam dados ou têm poucos dados de observação.

Como continuidade da pesquisa haverá um empenho no sentido de aprofundar o estudo de regionalização de vazões, nos seguintes tópicos:

- i. aprofundar a análise de *clusters* para o estudo de identificação de regiões aceitavelmente homogêneas, para efeito de comparação com o método TCEV;

- ii. desenvolver o terceiro nível de regionalização pela TCEV, ou seja, determinar um critério regional para estimativa da média (valor esperado de vazões máximas, Equação 13) em locais sem informação ou com pouca informação, para efeito de estimativa de quantis através da Curva de Crescimento (Equação 14).

AGRADECIMENTOS

Este trabalho foi possível graças ao apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Os autores são também gratos à CPRM - Serviço Geológico do Brasil pelo fornecimento dos dados utilizados.

REFERÊNCIAS

- ARNELL, N. W. e GABRIELE, S. (1988). The performance of the Two Component Extreme Value distribution in regional flood frequency analysis. *Water Resources Research*, 24(6), p. 879-887.
- BEABLE, M. E. e MCKERCHAR, A. I. (1982). Regional flood estimation in New Zealand. *Water and Soil Technical Publication 20*, Ministry of Works and Development, Wellington, N. Z.
- BERAN, M.; HOSKING, J. R. M. e ARNELL, N. (1986). Comment on "Two Component Extreme Value distribution for flood frequency analysis". *Water Resources Research*, 22(2), p. 263-266.
- CANNAROZZO, M., D'ASSARO, F. e FERRO, V. (1995). Regional rainfall and flood frequency analysis for Sicily using the Two Component Extreme Value distribution. *Hydrological Sciences Journal*, 40(1), p. 19-42.
- DUBAND, D. e GARROS-BERTHET, H. (1994). Design flood determination by the gradex method. *Bulletin du Comité Français des Grands Barrages, Barrages et Reservoirs*, n°2. Le Bourget du Lac, France.
- FIORENTINO, M.; GABRIELE, S.; ROSSI, F. e VERSACE, P. (1987a). Hierarchical approach for regional flood frequency analysis. In: Singh, V. P. (ed). *Regional Flood Frequency Analysis*, Reidel, Dordrecht, the Netherlands, p. 35-49.
- FIORENTINO, M.; GABRIELE, S.; ROSSI, F. e VERSACE, P. (1987b). Regional flood frequency analysis using the Two Component Extreme Value. A key reference abstract. *Excerpta*, vol. 2, p. 39-50.
- FRANCÈS, F. (1998). Using the TCEV distribution function with systematic and non-systematic data in a regional flood frequency analysis. *Stochastic Hydrology and Hydraulics*, vol. 12, p. 267-283.
- GABRIELE, S. e IIRITANO, G. (1994). Alcuni aspetti teorici ed applicativi nella regionalizzazione delle piogge con il modello TCEV. *Prediction and prevention of extreme hydrological events and their mitigation*, n°1089, GNDICI - CNR/IRPI (ed), Rende, Italy.
- GREHYS - GROUPE DE RECHERCHE EN HYDROLOGIE STATISTIQUE (1996). Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology*, 186, p. 63-84.
- HOSKING, J. R. M. e WALLIS, J. R. (1997). *Regional Frequency Analysis, an approach based on L-moments*. Cambridge University Press, U.K.
- JOLLEY, T. J. (1987). *The application of a regional Two Component Extreme Value distribution to Chilean flood data*. M.Sc. thesis, Department of Civil Engineering, Imperial College of Science and Technology, London, U.K.
- MATALAS, N. C.; SLACK, J. R. e WALLIS, J. R. (1975). Regional skew in search of a parent. *Water Resources Research*, 11, p. 815-826.
- NERC - NATURAL ENVIRONMENT RESEARCH COUNCIL (1975). *Flood Studies Report*, vol. 1. Natural Environment Research Council, London.
- PINTO, E. J. A. e NAGHETTINI, M. C. (1999). Definição de regiões homogêneas e a regionalização de precipitações máximas diárias da bacia do alto São Francisco. *Anais do XIII Simpósio Brasileiro de Recursos Hídricos*. Associação Brasileira de Recursos Hídricos, ABRH, Belo Horizonte, Brasil.
- ROSSI, F.; FIORENTINO, M. e VERSACE, P. (1984). Two Component Extreme Value distribution for flood frequency analysis. *Water Resources Research*, 20(7), p. 847-856.
- TODOROVIC, P. e ZELENHASIC, E. (1970). A stochastic model for flood analysis. *Water Resources Research*, 6(6), p. 1641-1648.
- TUCCI, C. E. M. (1993). *Hidrologia: Ciência e Aplicação*. Editora da Universidade. ABRH. EDUSP - Porto Alegre, p. 573-582.
- WAYLEN, P. e WOO, M. (1982). Prediction of annual floods generated by mixed processes. *Water Resources Research*, 18(4), p. 1283-1286.
- WILTSHIRE, S. E. (1986). Identification of homogeneous regions for flood frequency analysis. *Journal of Hydrology*, 84, p. 287-302.

Identification of Homogeneous Regions in the Regional Analysis of Precipitations and Maximum Floods in the São Francisco River Basin (Minas Gerais)

ABSTRACT

The TCEV regionalization model is applied to the São Francisco river basin in the State of Minas Gerais, Brazil. This is the case of a big watershed where the available data on maximum precipitation and floods are not uniformly distributed in the area but both present good quality. Therefore, there is heterogeneity of the available hydrometeorological information about the watershed, in both time and space. This statistical model is based on the product of two exponentials, both occurring according to a Poisson process: the first corresponds to the most frequently observed events and the other corresponds to the outliers. After applying this model, two homogeneous regions were identified, dividing the river basin into south and north, for both annual maximum precipitation and annual maximum floods. These regions were confirmed by sampling analysis and by comparison with the L-moments regionalization methodology. In the case of the maximum precipitation, the definition of the homogeneous regions reflects the available physiographic and climatological information on the watershed.

Key-words: maximum flood; homogeneous region; precipitation.