

Uso de Informação Regional na Estimativa de Quantis de Cheia GEV

Eduardo Sávio Passos Rodrigues Martins, Dirceu Silveira Reis Junior, Alan Michell B. Alexandre

FUNCEME – Fundação Cearense de Meteorologia e Recursos Hídricos
esm9@secrel.com.br; dirceu.reis@gmail.com; alan@funceme.br

Jery R. Stedinger

School of Civil and Environmental Engineering - Cornell University
jrs5@cornell.edu

Recebido: 20/05/07 – revisado: 23/10/07 – aceito: 25/06/08

RESUMO

Este artigo foca no uso de informações regionais na estimativa de quantis de cheia com a distribuição de Valores Extremos Generalizada (GEV). Estudos anteriores mostram que os estimadores de quantis, baseados no método da Máxima Verossimilhança (ML), são muito instáveis quando aplicados a amostras de tamanho pequeno e moderado, usualmente encontradas na prática, principalmente em países em desenvolvimento. A razão da instabilidade dos estimadores ML observada em amostras pequenas está na alta incerteza da estimativa do parâmetro de forma κ , que pode resultar em valores absurdos de quantis. Por causa disso, estimadores baseados em momentos lineares (MOM-L) são frequentemente empregados. Este artigo mostra que o uso de uma distribuição a priori informativa para κ , através do procedimento de Máxima Verossimilhança Generalizada (GML), resolve este problema. Resultados baseados em Simulação Monte Carlo mostram que o estimador GML é mais preciso que o estimador MOM-L quando $-0,4 \leq \kappa \leq 0$. O artigo apresenta ainda um procedimento, baseado em regionalização hidrológica, para derivar uma distribuição a priori de κ ainda mais informativa com intuito de aumentar a precisão dos estimadores de quantis de cheia.

Palavras-chave: Estimativa de Quantis de cheia; GEV.

INTRODUÇÃO

Muitos problemas em recursos hídricos requerem a identificação de uma relação de frequência. Estas relações são usualmente definidas pelo ajuste de um conjunto de dados a uma distribuição probabilística, que fornece uma representação compacta e suave da distribuição de frequência. Uma vez ajustada a distribuição, podem-se obter estimadores de quantis de interesse, associados a probabilidades de ocorrência, muitas vezes menores do que as frequências amostrais de excedência obtidas do conjunto de dados.

Entretanto, a estimativa destes quantis está sujeita a importantes fontes de incerteza, que incluem erros de medição, escolha do modelo de probabilidades, e ajuste dos parâmetros. O comprimento dos registros de uma estação restringe a precisão dos estimadores locais, o que aumenta a importância do uso de métodos eficientes para a estimativa de parâmetros e quantis, e a necessidade de utilização de

informações regionais e/ou históricas, quando disponíveis (Jin e Stedinger, 1989; Reis e Stedinger, 2005). Este estudo apresenta uma metodologia para incorporar informação regional à análise estatística utilizando como exemplo o estudo de cheias.

O modelo utilizado aqui para descrever a distribuição de cheias é a distribuição de Eventos Extremos Generalizada (GEV), introduzida por Jenkinson (1955). Esta distribuição tem sido muito empregada em hidrologia e áreas afins. Ela foi recomendada para análise de frequência de cheias na Grã-Bretanha (Natural Environment Research Council, 1975), precipitação nos EUA (Willeke et al., 1995), e ondas do mar (Haan e de Ronde, 1998). A distribuição GEV recebeu atenção especial na análise regional de frequência desde a introdução do procedimento *index-flood*, baseado nos momentos de pesos probabilísticos (PWM) de Wallis (1980), Greis e Wood (1981), e Hosking et al. (1985b). Vários estudos de frequência regional empregaram a distribuição GEV (Hosking et al., 1985b; Wallis e Wood, 1985; Lettenmaier et al., 1987; Hosking e Wallis, 1988; Chowdhury et al., 1991; Stedinger e Lu, 1995;

Madsen et al., 1997b; Hosking e Wallis, 1997). A origem física de máximos de variáveis ambientais sugere que suas distribuições podem ser de um dos tipos representados pela distribuição GEV (EV tipos I, II, e III). A distribuição GEV é muito atrativa porque sua inversa possui solução analítica, e seus parâmetros podem ser facilmente estimados pelos métodos dos momentos e momentos-L (MOM-L) (Hosking et al., 1985a; Hosking, 1990).

Hosking et al. (1985a) mostraram, para a distribuição GEV, que os estimadores PWM, ou seu equivalente MOM-L, são melhores, em termos de viés e variância, do que os estimadores de máxima verossimilhança (ML) para amostras de tamanho (n) entre 15 a 100. Madsen et al. (1997a) mostrou que o estimador da vazão com tempo de retorno de 100 anos, baseado no método dos momentos (MOM), tem menor raiz do erro quadrático médio (REQM), para $-0,25 < \kappa < 0,30$, do que aqueles baseados em MOM-L e ML para $10 \leq n \leq 50$ anos. Estimadores ML são preferíveis somente quando $\kappa > 0,3$ e $n \geq 50$. Uma preocupação é entender porque os estimadores ML se comportam tão pobremente em relação a outros estimadores, em especial para amostras pequenas.

O comportamento dos estimadores ML em amostras pequenas pode ser explicado pela alta incerteza na estimativa do parâmetro de forma κ , que acaba por resultar em estimativas absurdas de quantis. Isto explica o comportamento instável dos estimadores ML de quantis em amostras pequenas (Hosking et al., 1985a). Em contraposição aos estimadores ML, os estimadores MOM restringem $\hat{\kappa}$ a valores maiores que $-1/3$, enquanto que os estimadores MOM-L restringem $\hat{\kappa}$ a valores maiores que -1 (Coles e Dixon, 1998). Ambos estimadores MOM e MOM-L incorporam implicitamente alguma restrição nos valores de κ no procedimento de estimativa dos parâmetros da distribuição, enquanto que o estimador ML não o faz.

O foco deste estudo é na incerteza paramétrica causada pelos erros amostrais nos valores estimados do parâmetro de forma κ da distribuição GEV. Uma estrutura Bayesiana é empregada na análise ML para restringir os valores de κ a uma faixa de variação razoável do ponto de vista estatístico e físico para séries de vazões máximas anuais. Esta faixa de variação de κ é descrita por uma distribuição informativa a priori, que associa pesos a diferentes valores de κ dentro da faixa permitida. Este estudo demonstra que tal distribuição a priori, combinada com os dados do local de interesse, resulta em estimadores de quantis mais precisos do que os estimadores

MOM e MOM-L para amostras de tamanho pequeno e moderado. O novo estimador é aqui denominado de estimador Máxima Verossimilhança Generalizada (GML). O artigo também apresenta o uso dos estimadores GML com uma distribuição a priori de κ mais informativa, obtida através de regionalização hidrológica baseada em modelos de regressão. O método para obtenção da priori mais informativa para κ é aplicado às bacias dos rios Muskingum (USA, Ohio) e Tibagi (Paraná), e diversas bacias no Estado do Ceará.

A DISTRIBUIÇÃO GEV: MODELO E ESTIMATIVA DE PARÂMETROS

A distribuição GEV incorpora as distribuições Tipo I de Gumbel ($\kappa = 0$), Tipo II de Frechet ($\kappa < 0$) e Weibull ou Tipo III ($\kappa > 0$). A função de probabilidade acumulada é dada por:

$$F(x) = \exp\left\{-\left[1 - \kappa \frac{(x - \xi)}{\alpha}\right]^{1/\kappa}\right\} \quad \text{para } \kappa \neq 0 \quad (1)$$

$$= \exp\left\{-\exp\left[-\frac{(x - \xi)}{\alpha}\right]\right\} \quad \text{para } \kappa = 0$$

em que $\xi + \alpha/\kappa \leq x < \infty$ para $\kappa < 0$, $-\infty < x < +\infty$ para $\kappa = 0$, e $-\infty < x \leq \xi + \alpha/\kappa$ para $\kappa > 0$. Aqui ξ , α e κ são os parâmetros de posição, escala e forma, respectivamente. Os quantis da distribuição GEV são dados em termos dos parâmetros e da probabilidade acumulada p:

$$x_p = \xi + \frac{\alpha}{\kappa} \left\{1 - [-\ln(p)]^\kappa\right\} \quad \text{para } \kappa \neq 0 \quad (2)$$

$$= \xi - \alpha \ln[-\ln(p)] \quad \text{para } \kappa = 0$$

Os métodos considerados aqui para estimativa destes parâmetros são apresentados na Tabela 1: MOM-L (Landwehr et al., 1979; Hosking, 1990), MOM (Stedinger et al., 1993; Madsen et al., 1997a) e ML (Hosking, 1985; Hosking et al., 1985a; MacLeod 1989; Martins e Stedinger, 2000).

Tabela 1 - Estimativa dos parâmetros da distribuição GEV.

Método	Equações
Momentos-L	$\hat{\xi} = \hat{\lambda}_1 - \frac{\hat{\alpha}}{\hat{\kappa}} \{1 - \Gamma(1 + \hat{\kappa})\} \quad \hat{\alpha} = \frac{\hat{\lambda}_2 \hat{\kappa}}{(1 - 2^{-\hat{\kappa}}) \Gamma(1 + \hat{\kappa})} \quad \hat{\kappa} = 7.8590 c + 2.9554 c^2, \text{ com } c = 2/(3 + \hat{\tau}_3) - \log(2)/\log(3)$ $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3 \text{ e } \hat{\tau}_3 = \frac{\hat{\lambda}_3}{\hat{\lambda}_2} (L - \text{assimetria})$ <p>são obtidos usando um estimador não tendencioso para os três primeiros momentos de pesos probabilísticos (PWMs, β_r) definidos como:</p> $\beta_r = \sum_{i=1}^n \left[\frac{(i-1)(i-2)(i-3)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_{(i)} \right] \quad r=0,1,2,\dots$ <p>onde os $x_{(i)}$ são observações ordenadas de uma amostra de tamanho n $\{x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}\}$, e $\lambda_1 = \beta_0$, $\lambda_2 = 2\beta_1 - \beta_0$, $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$.</p>
Método dos Momentos	$\hat{\xi} = \hat{\mu} - \frac{\hat{\alpha}}{\hat{\kappa}} \{1 - \Gamma(1 + \hat{\kappa})\} \quad \hat{\alpha} = \frac{\hat{\sigma} \hat{\kappa} }{(\Gamma(1 + 2\hat{\kappa}) - [\Gamma(1 + \hat{\kappa})]^2)^{1/2}}$ $\hat{\gamma} = \text{Sign}(\hat{\kappa}) \frac{-\Gamma(1 + 3\hat{\kappa}) + 3\Gamma(1 + \hat{\kappa})\Gamma(1 + 2\hat{\kappa}) - 2[\Gamma(1 + \hat{\kappa})]^3}{(\Gamma(1 + 2\hat{\kappa}) - [\Gamma(1 + \hat{\kappa})]^2)^{3/2}}$ <p>onde $\text{Sign}(\hat{\kappa})$ é mais ou menos 1 dependendo do sinal de $\hat{\kappa}$, $\Gamma(\cdot)$ é a função gama, e $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\gamma}$ são a média, desvio padrão e assimetria amostrais, respectivamente. O momento de ordem r da distribuição GEV existe somente se $\hat{\kappa} > -\frac{1}{r}$. A equação da assimetria requer uma solução iterativa.</p>
Máxima Verossimilhança	<p>A função verossimilhança para uma amostra de n observações $\{x_1, x_2, \dots, x_n\}$ de uma distribuição GEV é</p> $\ln(L(\theta \underline{x})) = -n \cdot \ln(\alpha) + \sum_{i=1}^n \left[(1/\kappa - 1) \cdot \ln(y_i) - (y_i)^{1/\kappa} \right], \text{ onde } \theta = (\xi, \alpha, \kappa) \text{ e } y_i = [1 - (\kappa/\alpha) \cdot (x_i - \xi)].$ <p>Os estimadores de máxima verossimilhança de ξ, α, e κ podem ser identificados resolvendo o sistema de equações:</p> $\frac{1}{\alpha} \cdot \sum_{i=1}^n \left[1 - \kappa \cdot (y_i)^{1/\kappa} \right] / y_i = 0$ $-S/\alpha + \frac{1}{\alpha} \cdot \sum_{i=1}^n \left[1 - \kappa \cdot (y_i)^{1/\kappa} \right] \cdot (x_i - \xi) / \alpha = 0$ $-1/\kappa^2 \cdot \sum_{i=1}^n \ln(y_i) \cdot [1 - \kappa \cdot (y_i)^{1/\kappa}] + \left[1 - \kappa \cdot (y_i)^{1/\kappa} \right] / y_i \cdot \kappa \cdot (x_i - \xi) / \alpha = 0$

MÁXIMA VEROSSIMILHANÇA EM AMOSTRAS PEQUENAS

O método de máxima verossimilhança possui uma série de propriedades ótimas baseadas na Teoria de Grandes Amostras, entretanto, seu desempenho em amostras pequenas e moderadas não tem sido muito bem estudado (Hosking et al., 1985a). Resultados de Hosking et al. (1985a) indicam que quantis baseados em MOM-L são tendenciosos, mas preferíveis aos correspondentes estimadores ML, porque estes últimos são mais variáveis. O reexame do comportamento dos estimadores ML em amostras pequenas demonstra que valores absurdos do parâmetro de

forma κ da GEV podem ser obtidos. Para ilustrar o problema, considere a amostra gerada aleatoriamente e apresentada na Tabela 2.

Tabela 2 - Amostra (n = 15) da GEV($\xi=0, \alpha=1, \kappa = -0.20$)

-0,4154	-0,3999	-0,3973	-0,3397	-0,3094
-0,0997	0,1546	0,4539	1,0572	1,4462
1,4480	2,3660	2,4663	2,8629	4,6820

As estimativas ML dos parâmetros da GEV são: $\hat{\xi} = -0,26$, $\hat{\alpha} = 0,33$ e $\hat{\kappa} = -2,00$. O valor populacional do quantil 0,999 ($\xi=0, \alpha=1, \kappa = -0,20$) é 14,9, enquanto que o valor estimado ($\hat{\xi} = -0,26$, $\hat{\alpha} = 0,33$ e

$\hat{\kappa} = -2,00$) é da ordem de 1×10^5 . A Figura 1 compara a distribuição verdadeira e a estimada para esta pequena amostra.

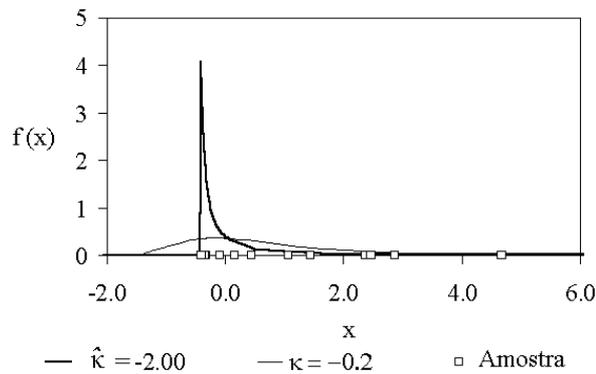


Figura 1 - Distribuição verdadeira ($\kappa = -0,20$) e “estimada” ($\hat{\kappa} = -2,00$) para uma pequena amostra de tamanho $n = 15$ (valores amostrais indicados por quadrados).

ESTIMADORES DE MÁXIMA VEROSSIMILHANÇA GENERALIZADA

Uma distribuição a priori para

Suponha que o conhecimento que se tem do verdadeiro parâmetro κ da distribuição GEV, com função densidade de probabilidade $f(x|\xi, \alpha, \kappa)$, pode ser representado por uma distribuição a priori ($\pi(\kappa)$) no intervalo $[\kappa_L, \kappa_U]$. A experiência hidrológica indica que o intervalo $-0,30 \leq \kappa \leq 0$ é o mais provável para κ . Muitos estudos mostram que cheias parecem ter caudas de distribuição mais pesadas ($\kappa < 0$) do que a de uma distribuição Gumbel ($\kappa = 0$) (Farquharson *et al.*, 1992; Madsen e Rosbjerg, 1997). Adicionalmente, um exame das funções de densidade de probabilidade (fdp) da GEV revela que formas razoáveis são obtidas para κ entre $[-0,3; +0,3]$. A Figura 2 mostra fdp da distribuição GEV para diferentes valores de κ . Nesta figura, com $\kappa = -0,1$ e $-0,3$, ambas as fdp possuem uma assimetria positiva com uma longa cauda superior, como pode-se esperar para picos de cheia e fenômenos similares estritamente positivos como, por exemplo, precipitação máxima anual. Além disso, a distribuição GEV possui variância finita somente quando $\kappa > -0,5$, e assimetria finita quando $\kappa > -1/3$. Para $\kappa < 0$, a EV Tipo II tem uma cauda tipo Pareto, tal que $F(x) = 1 - [1 + |\kappa|/\alpha \cdot (\xi - x)]^{-1/|\kappa|}$ à medida que $x \rightarrow \infty$. Para $\kappa > 0$, a EV Tipo III tem uma cauda polinomial, tal que $F(x) = 1 - [(\kappa/\alpha) \cdot (\xi + \alpha/\kappa - x)]^{1/\kappa}$ à medida que $x \rightarrow \xi + \alpha/\kappa$. De

modo geral, uma distribuição a priori para o fator de forma κ , no caso de cheias máximas anuais, pareceria com a fdp apresentada na Figura 3.

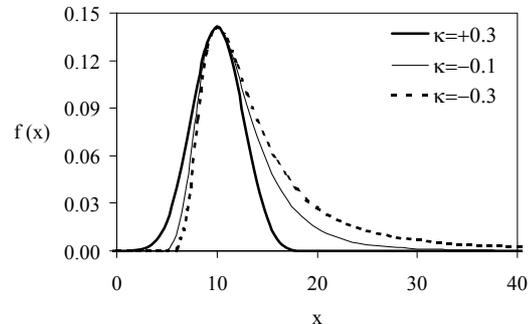


Figura 2 - FDPs da distribuição GEV para $\kappa = -0,3, -0,1$ e $+0,3$ (onde $\xi = 10$ e $\alpha = 2,6$).

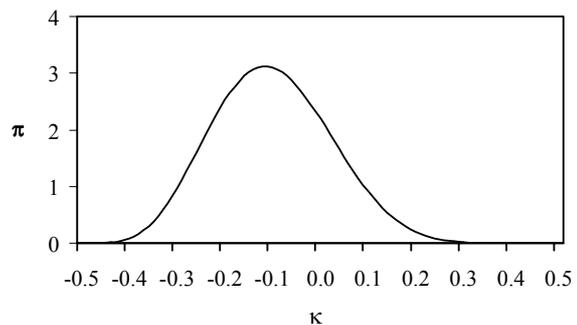


Figura 3 - *Priori* Geofísica para o parâmetro de forma da GEV. Distribuição Beta com média = $-0,10$ e var = $(0,122)^2$.

A distribuição a priori empregada na Figura 3, chamada aqui de *priori* Geofísica, é uma distribuição Beta, $\pi(\kappa) = (0,5 + \kappa)^{p-1} (0,5 - \kappa)^{q-1} / B(p, q)$, entre $[-0,5; +0,5]$, com $p=6$ e $q=9$, onde $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q)$. Ela possui $E[\kappa] = -0,10$ e $Var[\kappa] = (0,122)^2$. A escolha da distribuição Beta, e respectivos parâmetros, se justifica pela conformidade com as características descritas no início desta seção. Maiores detalhes sobre esta escolha podem ser encontrados em Martins e Stedinger (2000). Embora não seja evidente a partir da Figura 3, esta *priori* é relativamente suave em $-0,30 \leq \kappa \leq 0,15$, que é o intervalo de interesse (ver Figura 8). Informações de cheias provenientes de outras estações na região podem ser utilizadas para desenvolver uma *priori*

mais informativa para κ (Madsen e Rosbjerg, 1997), conforme apresentado mais adiante neste artigo, podendo trazer melhorias nos estimadores de quantis extremos (probabilidade de excedência $\leq 1\%$).

O Estimador Máxima Verossimilhança Generalizado

A função verossimilhança generalizada é obtida por $GL(\xi, \alpha, \kappa | x) = L(\xi, \alpha, \kappa | x) \pi(\kappa)$. Poder-se-ia ser mais rigoroso e incluir também uma *priori* conjunta para ξ e α , mas isto não foi feito. Um procedimento Bayesiano completo para análise de frequência de cheias é apresentado por Kuczera (1999). Tal procedimento seria apropriado se tivéssemos informações sobre os parâmetros de posição e escala da distribuição, o que não é o caso.

Os estimadores GML de ξ , α e κ são obtidos pela maximização da função log-verossimilhança generalizada, que corresponde à moda da distribuição a posteriori dos parâmetros (Berger, 1985, p. 133). Aqui, o máximo da função foi obtido através do método Newton-Raphson.

Com os estimadores GML, à medida que o tamanho da amostra aumenta, a informação contida na verossimilhança deve dominar a informação provida pela *priori* (Robert, 1994; p. 138), de modo que, assintoticamente, os estimadores ML e GML possuem as mesmas propriedades. Em geral, os estimadores GML terão as mesmas propriedades ótimas dos estimadores ML caso a função de verossimilhança e a distribuição a *priori* satisfaçam umas poucas condições de regularidade (Lehman e Casella, 1996). Com a *priori* geofísica, nosso estimador GML irá herdar as propriedades ótimas assintóticas do estimador ML para $-0,5 < \kappa < 0,5$.

Resultados Monte Carlo com informação local

Estudos com base em simulação Monte Carlo foram realizados para avaliar a eficiência de diferentes estimadores. Foram geradas amostras com tamanhos entre 15 e 100, com $\xi = 0$, $\alpha = 1$ e $-0,4 \leq \kappa \leq +0,4$. Para cada distribuição populacional e extensão da série, foram geradas 10.000 amostras que serviram de base para o cálculo de viés, variância e REQM dos estimadores dos quantis 0,001 - 0,999 com base em quatro métodos: ML, MOM-L, MOM e GML.

Problemas no algoritmo Newton-Raphson para estimativa ML são bem documentados na literatura (Hosking, 1985; Madsen *et al.*, 1997a). Nas simulações aqui realizadas estes problemas também foram identificados, a saber: a não convergência do

algoritmo Newton-Raphson para estimadores ML e a geração de quantis muito extremos em amostras pequenas. A obtenção de estimativa de quantis muito extremos estava associada a valores muito negativos de κ , resultando em estimativas instáveis da REQM para o estimador MV. O uso do método GML eliminou estes problemas.

As Figuras 4 e 5 apresentam, respectivamente, os valores de REQM dos estimadores de quantis 0,99 e 0,999 para diferentes valores do parâmetro de forma e amostras de tamanho 25. Para $\kappa < 0$, o estimador GML possui o melhor desempenho dentre os estimadores avaliados em termos de REQM. O REQM de todos os estimadores decresce à medida que o tamanho da amostra aumenta, mas o desempenho do GML ainda é o melhor.

Embora não seja aqui apresentado, (veja Martins e Stedinger, 2000), é interessante ressaltar que: (1) O estimador GML tem no mínimo um desempenho tão bom quanto os estimadores MOM-L e MOM, e tem a menor REQM para $p \geq 0,98$, e $p \leq 0,01$; (2) Para os quantis estudados (0,001-0,999), os estimadores GML são relativamente não tendenciosos. Os estimadores MOM, que possuem a segunda menor REQM para $p \geq 0,98$, exibem um grande viés negativo. Nestes casos, os estimadores GML possuem o menor viés e a menor RMSE; (3) À medida que o tamanho da amostra aumenta, os estimadores passam a ter desempenho semelhante.

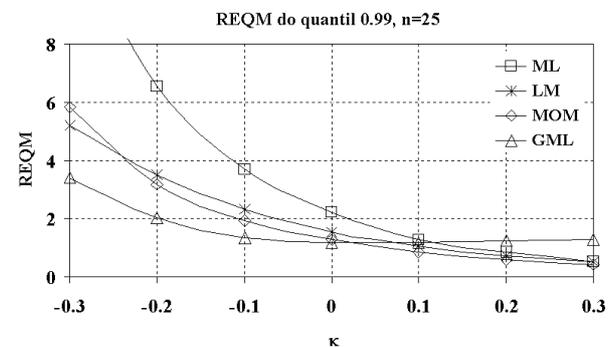


Figura 4 - REQM dos estimadores do quantil 0,99 para amostras com tamanho 25.

A Figura 6 ilustra, também com base em simulação Monte Carlo, como distribuições a *priori* mais informativas influenciam a precisão de estimadores de quantis de cheia para amostras com distribuição GEV. Verifica-se que os estimadores de quantis tornam-se mais precisos à medida que a variância

da *priori* de κ decresce. Deste modo, espera-se que o uso de *prioris* mais informativas, obtidas através de regionalização hidrológica com base em modelos de regressão, possam trazer benefícios adicionais à precisão destes estimadores. De forma a ilustrar esta idéia, a Figura 6 apresenta três linhas verticais que indicam o desvio padrão da *priori* de κ . As duas mais à esquerda representam a raiz quadrada da variância média de predição dos modelos de regressão regional de κ para as bacias dos rios Muskingum e Tibagi, enquanto que a linha à direita representa o desvio padrão da *priori* Geofísica.

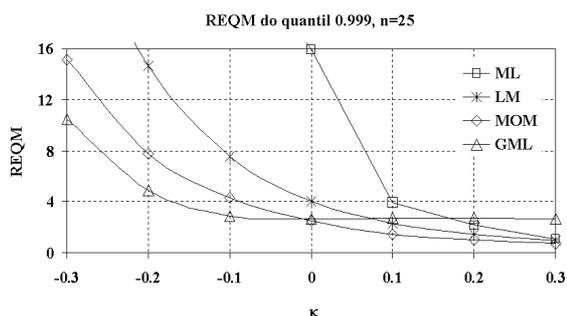


Figura 5 - REQM dos estimadores do quantil 0,999 para amostras com tamanho 25.

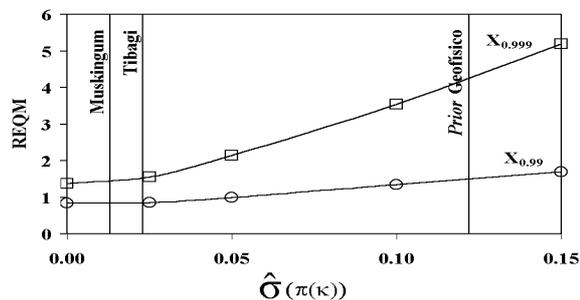


Figura 6 - REQM dos estimadores GML dos quantis 0,99 e 0,999 para desvio-padrão da *priori* variando entre 0 e 0,15 ($n=50$). As primeiras duas linhas verticais indicam os resultados correspondentes à aplicação do procedimento de regionalização de κ para Muskingum e Tibagi. A terceira linha corresponde à *priori* Geofísica. $\sigma(\pi(\kappa)) = 0$ indica que o valor de κ é sempre -0,10, portanto o REQM reflete somente a incerteza em $\hat{\xi}$ e $\hat{\alpha}$.

ESTIMADORES DE REGRESSÃO REGIONAL

Como foi dito anteriormente, um estudo de regionalização hidrológica pode ser utilizado para

gerar uma distribuição a *priori* para κ mais informativa do que a *priori* Geofísica.

Esta seção do artigo apresenta uma metodologia baseada em modelos de regressão para obter o valor esperado e a variância regional do parâmetro de forma κ , de forma a derivar uma distribuição a *priori* mais informativa baseada na distribuição de probabilidades Beta.

Introdução

Muitos estudos têm mostrado o valor da informação regional na análise de freqüência de cheias (NERC, 1975; Lettenmaier e Potter, 1985; Hosking *et al.*, 1985b; Lettenmaier *et al.*, 1987; Fill, 1994; Stedinger e Lu, 1995; Madsen e Rosbjerg, 1997; Fill e Stedinger, 1998). Um dos métodos empregados é o *Index-flood*, cuja premissa básica é que as cheias numa dada região, quando divididas por uma “cheia índice” local, são igualmente distribuídas (Dalrymple, 1950) (ou próximo disso). Usualmente utiliza-se a média da amostra local como “cheia índice” (Stedinger *et al.*, 1993). As vazões máximas anuais, transformadas pela média amostral, são chamadas de “cheias normalizadas”.

Admitir que as cheias normalizadas são igualmente distribuídas equivale a dizer que tanto o coeficiente de variação (CV), quanto a razão dos momentos de alta ordem, são iguais na região. Supõe-se, portanto, que as diferenças observadas nos dados são o resultado de erros amostrais. O *Index-flood* pode ser adequado para regiões homogêneas, onde as cheias possuem características similares. Entretanto, dados coletados em várias regiões do mundo mostram que o CV depende das características das bacias. Por exemplo, CVs normalmente decrescem com um acréscimo na área de drenagem (Thomas e Benson, 1970). Isto sugere que a estimativa de variáveis hidrológicas regionais, tal como CV, deva levar em conta as características fisiográficas das bacias. O mesmo pode ser feito para estimadores de razão entre momentos de mais alta ordem (ou dos parâmetros que os controlam).

Tasker e Stedinger (1986) utilizaram regressão WLS (Mínimos Quadrados Ponderados) para estimar o coeficiente de assimetria como função das características fisiográficas da bacia. Madsen e Rosbjerg (1997b) fizeram o mesmo para o parâmetro de forma κ da distribuição GEV usando a regressão GLS (Mínimos Quadrados Generalizados).

Hosking e Wallis (1988) examinaram o impacto da correlação espacial de vazões no método regional *Index-flood* e verificaram que o efeito é a

redução da precisão dos estimadores de quantis de cheias obtidos pela análise regional. A variabilidade regional dos parâmetros dos modelos, tal como o parâmetro de forma κ da distribuição GEV, pode ser descrita pelo uso do modelo de regressão GLS (Madsen e Rosbjerg, 1997b), que considera a heterogeneidade da variância amostral e a correlação cruzada dos resíduos do modelo de regressão. Isso pode melhorar o resultado do modelo para o parâmetro de forma κ , e conseqüentemente pode refinar a precisão dos estimadores de quantis de interesse. Para correlação espacial pequena e moderada, os modelos de regressão WLS e GLS são similares (Stedinger e Tasker, 1986ab).

MÍNIMOS QUADRADOS GENERALIZADOS

Estimativa da distribuição a priori

Dados de outras estações na região podem ser utilizados para estabelecer uma relação empírica entre o parâmetro de forma κ e as características físicas das bacias através de análise de regressão regional. Esta análise fornece, para o local de interesse, uma estimativa regional de κ , e respectiva incerteza, que podem ser utilizadas para definir uma distribuição a priori mais informativa de κ .

Modelo de regressão GLS para o parâmetro de forma κ

Considera-se que o valor amostral de κ em um dado local i , $\hat{\kappa}_i$, pode ser descrito como uma função linear das características fisiográficas da bacia,

$$\hat{\kappa}_i = \beta_0 + \sum_{j=1}^p \beta_j A_{ij} + \eta_i \quad i = 1, 2, \dots, M \text{ estações} \quad (3)$$

onde A_{ij} ($j=1\dots p$) são as características físicas utilizadas no modelo de regressão e $\eta_i = \varepsilon_i + \delta$ a soma do erro do modelo δ com os erros amostrais ε_i . A matriz de covariância dos erros Λ é dada por:

$$\Lambda_{ij} = \text{cov}(\eta_i, \eta_j) = \begin{cases} \sigma_{\varepsilon i}^2 + \sigma_{\delta}^2 & i = j \\ \sigma_{\varepsilon i} \sigma_{\varepsilon j} \rho(\hat{\kappa}_i, \hat{\kappa}_j) & i \neq j \end{cases} \quad (4)$$

onde $\sigma_{\varepsilon i}^2$ é a variância do erro amostral do estimador local $\hat{\kappa}_i$, σ_{δ}^2 é a variância do erro do modelo e

$\rho_{\varepsilon ij}$ é o coeficiente de correlação do erro amostral devido à correlação espacial entre as vazões máximas anuais das estações i e j .

Em notação matricial, tem-se:

$$K = X\beta + \eta$$

$$\text{onde } K = (\hat{\kappa}_1 \hat{\kappa}_2, \dots, \hat{\kappa}_M)^T; \beta = (\beta_0 \beta_1 \dots \beta_p)^T \quad (5)$$

$$\eta = (\eta_1, \eta_2, \dots, \eta_M)^T; \quad X = \begin{bmatrix} 1 & A_{11} & \dots & A_{1p} \\ \circ & \circ & \dots & \circ \\ \dots & \dots & \dots & \dots \\ 1 & A_{M1} & \dots & A_{Mp} \end{bmatrix}$$

O estimador GLS de β é determinado por

$$(X^T \Lambda^{-1} X)\beta = X^T \Lambda^{-1} K \quad (6)$$

O estimador de σ_{δ}^2 baseado no método dos momentos é obtido pela solução conjunta de (6) e (7) dado que $\sigma_{\delta}^2 \geq 0$ (Stedinger e Tasker, 1985)

$$(K - X\hat{\beta})^T \Lambda^{-1} (K - X\hat{\beta}) = M - p - 1 \quad (7)$$

O valor de σ_{δ}^2 pode ser entendido como uma medida de heterogeneidade da região (Madsen e Rosbjerg, 1997a). Os modelos WLS e OLS (Mínimos Quadrados) podem ser visualizados como casos particulares do modelo GLS. Quando $\hat{\rho}(\hat{\kappa}_i, \hat{\kappa}_j) = 0$ para qualquer par de estações (i, j), GLS reduz-se a WLS. Quando a matriz de covariância apresenta zero nos elementos fora da diagonal e apenas σ_{δ}^2 na diagonal, o modelo GLS reduz-se ao modelo OLS. Para uma bacia em particular, com um vetor de características fisiográficas $x_0 = \{1, x_{01}, x_{02}, \dots, x_{0p}\}$, a variância de predição da variável de interesse κ_0 em relação a sua estimativa $x_0 \hat{\beta}$ é dada por:

$$E[(\kappa_0 - x_0 \hat{\beta})^2] = \sigma_{\delta}^2 + x_0 (X^T \Lambda^{-1} X)^{-1} x_0^T \quad (8)$$

A estatística proposta por Tasker e Stedinger (1986) para avaliar o desempenho do modelo numa dada região é a variância média de predição (VMP),

$$\text{VMP} = \sigma_{\delta}^2 + \frac{1}{M} \sum_{i=1}^M x_i (X^T \Lambda^{-1} X)^{-1} x_i^T \quad (9)$$

onde o segundo termo é a variância média amostral (VMA).

Estimando a matriz covariância amostral

A fim de aplicar o procedimento GLS é necessário estimar σ_{ei}^2 e $\hat{\rho}(\hat{\kappa}_i, \hat{\kappa}_j)$. Seguindo Tasker (1980) e Madsen e Rosbjerg (1997a), um estimador de σ_{ei}^2 que é aproximadamente independente de $\hat{\kappa}_i$ pode ser obtido por:

$$\sigma_{ei}^2 = \frac{N_A}{N_i} \text{var}[\hat{\kappa}_A | N_A] \tag{10}$$

onde: $\hat{\kappa}_A = \frac{1}{M} \sum_{i=1}^M \hat{\kappa}_i$ e $N_A = \frac{1}{M} \sum_{i=1}^M N_i$

Aqui $\text{var}[\hat{\kappa}_A | N_A]$ é a variância do erro amostral do estimador local, calculada usando a média regional de $\hat{\kappa}_i$ e o número médio de observações N_A .

Martins e Stedinger (2002) demonstraram que o coeficiente de correlação espacial de $\hat{\kappa}$ pode ser expresso em termos do coeficiente de correlação espacial de vazões, ρ_{ij} , através da seguinte expressão:

$$\hat{\rho}(\hat{\kappa}_i, \hat{\kappa}_j) = \rho_{ij}^\delta \tag{11}$$

onde δ depende do valor regional para κ , de acordo com a Tabela 3, proposta pelos mesmos autores.

Tabela 3 – Estimativa dos valores de δ para experimentos Monte Carlo com a distribuição GEV

κ	GEV
	Δ
-0,3	2,4
-0,2	2,6
-0,1	2,8
0,0	2,9
0,1	3,1

O uso de estimadores amostrais de ρ_{ij} pode resultar em uma matriz covariância (Λ) que nem sempre é positiva definida devido às incertezas e variações na extensão de dados concomitantes (Tasker e Stedinger, 1989). Tasker e Stedinger (1989) propuseram o uso de função suavizada entre ρ_{ij} e a distância entre as estações (d_{ij}). Portanto, em vez de

utilizar a estimativa amostral de ρ_{ij} , emprega-se o seguinte estimador:

$$\hat{\rho}(\hat{\kappa}_i, \hat{\kappa}_j) = [\hat{\rho}(d_{ij})]^\delta \tag{12}$$

Se a relação correlação-distância para as estações não é clara, pode-se assumir uma estrutura correlacional homogênea ($\hat{\rho} = \bar{\rho}$), onde $\bar{\rho}$ é o coeficiente de correlação regional médio,

$$\bar{\rho} = \left\{ \frac{1}{2} N_{ij} (N_{ij} - 1) \right\}^{-1} \sum_{1 \leq i \leq j \leq N} \bar{\rho}_{ij} \tag{13}$$

ESTIMADOR REGIONAL PARA T-ANOS

Com o valor estimado de β , pode-se calcular o valor esperado e a variância de predição de κ_i :

$$\begin{aligned} \hat{\mu}_{\kappa_i} &= x_i^T \hat{\beta} \\ \hat{\sigma}_{\kappa_i}^2 &= x_i^T \text{Var}[\hat{\beta}] x_i + \hat{\sigma}_\delta^2 \end{aligned} \tag{14}$$

onde x_i^T é a linha i em X e $\text{Var}[\hat{\beta}] = [X^T \Lambda^{-1} X]^{-1}$ é a matriz de covariância dos parâmetros do modelo regional. Estes dois momentos são utilizados para estimar os parâmetros da distribuição *a priori* para κ . Os parâmetros p e q da distribuição Beta são determinados resolvendo o seguinte sistema de equações:

$$\hat{\mu}_{\kappa_i} = \frac{p}{p+q} - 0,5 \quad \hat{\sigma}_{\kappa_i}^2 = \frac{pq}{(p+q+1)(p+q)^2} \tag{15}$$

As estimativas GML dos parâmetros da GEV para cada local de interesse são obtidas através da maximização da função verossimilhança generalizada, $GL(\xi, \alpha, \kappa | x) = L(\xi, \alpha, \kappa | x) \pi(\kappa)$, que mostra a relação entre a função verossimilhança generalizada e a verossimilhança. As informações regionais (sistemáticas ou históricas) são transmitidas através da *priori* $\pi(\kappa)$. Os quantis de interesse para estação i são então calculados pela equação (16), onde p é a probabilidade acumulada,

$$\hat{x}_p = \hat{\xi} + \frac{\hat{\alpha}}{\hat{\kappa}_R} \{1 - \exp[\hat{\kappa}_R \ln(-\ln p)]\} \tag{16}$$

APLICAÇÃO

Esta seção ilustra a aplicação do método de regionalização hidrológica, baseado no modelo GLS, para derivar distribuições *a priori* mais informativas do parâmetro de forma k da GEV.

O procedimento foi aplicado para as bacias dos rios Muskingum (EUA, Ohio) e Tibagi (Brasil, PR), com séries de dados entre 23 e 93 anos, e 15 e 65 anos, respectivamente, e para as bacias do Estado do Ceará, com séries entre 10 e 63 anos.

As seguintes variáveis explanatórias foram utilizadas nos estudos de regressão das bacias dos rios Muskingum e Tibagi: Área de drenagem (A), comprimento do talvegue (CT) e declividade (D). Para as bacias do Estado do Ceará foram ainda empregadas: precipitação média anual (P), comprimento total de drenagem (CTD), densidade de drenagem (DD), capacidade de armazenamento do solo (CAD), *curve number* médio (CN), índice de compactidade da bacia (Kc), direção dominante do rio principal (Dir), porcentagem da bacia situada no cristalino (Cr) e a razão (L) entre a área de lagos naturais ou artificiais e a área da bacia.

Correlação espacial de

Na bacia do Rio Muskingum, a distância entre as estações não explica a variação da correlação espacial das vazões máximas anuais. Portanto, adotou-se uma estrutura de correlação homogênea ($\hat{\rho} = \bar{\rho} = 0,26$). Deste modo, o modelo empregado para a correlação espacial de κ é

$$\hat{\rho}(\hat{\kappa}_i; \hat{\kappa}_j) = (cf_{ij}) \cdot \text{sign}(\bar{\rho}) \cdot \bar{\rho}^\delta \quad (17)$$

onde $cf_{ij} = n_{ij} / \sqrt{(n_{ij} + n_i)(n_{ij} + n_j)}$ (n_{ij} é o período comum de dados, n_i e n_j são as observações extras do período para as estações i e j , respectivamente). O valor médio de κ na região é de -0,15, de modo que $\delta = 2,7$ (Ver Tabela 3). O valor médio da correlação espacial de κ é $\hat{\rho}(\hat{\kappa}_i; \hat{\kappa}_j) = 0,02$, um valor modesto.

Situação semelhante ocorre nas bacias do Estado do Ceará, onde também adotou-se uma estrutura correlacional homogênea com coeficiente de correlação médio de 0,01. O valor médio de κ é igual a -0,153, portanto, $\delta = 2,7$. Neste caso, $\hat{\rho}(\hat{\kappa}_i; \hat{\kappa}_j) = 0,008$, um valor muito pequeno.

A Figura 7 mostra que a correlação espacial das vazões máximas anuais para o Rio Tibagi varia

linearmente com a distância entre estações, $\hat{\rho}_{ij} = 1 + \beta_1 d_{ij}$. O modelo empregado é dado por

$\hat{\rho}(\hat{\kappa}_i; \hat{\kappa}_j) = (cf_{ij}) \cdot \text{sign}(\hat{\rho}_{ij}) \cdot \hat{\rho}_{ij}^\delta$, onde $\hat{\rho}_{ij}$ é definida pela equação de regressão acima. O valor médio de κ na região é de -0,10, portanto, $\delta = 2,8$.

Resultados

O estudo de regionalização do parâmetro para a três regiões levou em consideração todas as combinações possíveis das variáveis explanatórias listadas anteriormente. Estas variáveis foram transformadas para o espaço logarítmico, e centralizadas por suas respectivas médias. A fim de ilustrar as diferenças entre as análises de regressão, OLS, WLS e GLS, as mesmas foram empregadas para cada uma das combinações de variáveis explanatórias.

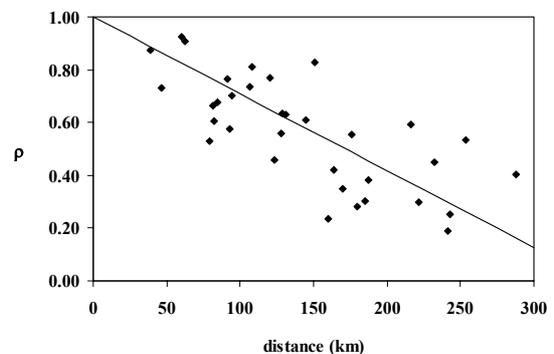


Figura 7 - Correlação espacial das vazões ($n_{ij} \geq 25$) da bacia do Rio Tibagi em função das distâncias entre estações.

O critério de seleção dos melhores modelos incluiu a variância média de predição (VMP) e a significância dos parâmetros do modelo. Apesar do número relativamente elevado de variáveis explanatórias utilizadas, pelo menos no caso do Estado do Ceará, o melhor modelo para as três regiões foi a média regional, independente da análise utilizada, OLS, WLS ou GLS.

Cabe esclarecer que a média regional obtida pelas análises WLS e GLS é diferente da média aritmética dos valores amostrais de κ . Isto porque essas análises levam em consideração os erros amostrais em cada estação e a correlação espacial dos estimadores de κ , no caso GLS.

Estes resultados podem causar a falsa impressão de que a média regional (empregando análise GLS) é sempre a melhor opção. Contudo, vale

mencionar dois estudos em andamento, um na bacia do Rio Illinois, e um outro nas bacias da Korea, onde, em ambos os casos, o melhor modelo regional de κ inclui características das bacias.

As Tabelas 4, 5 e 6 mostram as características dos melhores modelos regionais de κ para as bacias do Muskingum, Tibagi e do Ceará, respectivamente. O desvio-padrão da constante do modelo encontra-se entre parêntesis.

Os resultados de cada região mostram que não há muita diferença entre as estimativas da média regional de κ obtidas pelas análises OLS, WLS e GLS. As diferenças são mais importantes na estimativa do desvio-padrão da constante do modelo regional, na VEM, VMA e VMP. A estimativa da VMP é de suma importância nesta análise, pois ela será utilizada para representar a incerteza do estimador regional na obtenção da distribuição a priori mais informativa para κ . Deste modo, o emprego de uma metodologia que forneça uma estimativa precisa e realista da VEM, VMA e VMP do modelo regional é imprescindível para o correto emprego do procedimento GML.

Tabela 4 - Bacia do Muskingum: Regressão de 100. κ

Modelo	Coeficiente	VEM	VMA	VMP
	Constante			
OLS	-15,3 (1,13)	56,5	1,3	57,8
WLS	-14,9 (1,27)	0	2,0	2,0
GLS	-14,8 (1,68)	0	3,1	3,1

Tabela 5 - Bacia do Tibagi: Regressão de 100. κ

Modelo	Coeficiente	VEM	VMA	VMP
	Constante			
OLS	-10,5 (1,31)	29,3	1,7	31,0
WLS	-11,1 (2,25)	0,0	5,1	5,1
GLS	-10,9 (3,70)	0	13,7	13,7

Os resultados mostram que a análise OLS fornece valores muito maiores para a VEM, e consequentemente para a VMP, como observado antes por Stedinger e Tasker (1985). Isto se deve ao fato da análise OLS não fazer qualquer distinção entre a variância oriunda dos erros amostrais e a variância do erro do modelo propriamente dita. Portanto, a análise OLS não fornece uma representação adequada dos dados.

Tabela 6 - Bacias do Estado do Ceará: Regressão de 100. κ

Modelos	Coeficientes	VEM	VMA	VMP
	Constante			
OLS	-15,4 (1,18)	34,8	1,2	36,0
WLS/GLS	-16,5 (1,96)	0	4,0	4,0

A única diferença entre as análises WLS e GLS é que a última leva em consideração a correlação espacial entre os estimadores de κ . No Estado do Ceará esta correlação é muito baixa, de modo que os resultados das análises WLS e GLS são idênticos. No caso do Muskingum, embora a correlação espacial de κ seja pequena, não é desprezível. A VMP obtida pela análise GLS é em torno de 50% maior do que aquela obtida pela análise WLS, muito embora as duas sejam pequenas. Já no caso do Rio Tibagi, a correlação espacial de κ é bem expressiva, justificando as diferenças significativas encontradas em termos de desvio-padrão da média regional e VMP. A VMP obtida pela análise GLS é quase três vezes maior do que aquela obtida pela análise WLS. Como a análise GLS representa melhor os dados hidrológicos, tendo em vista que leva em consideração não apenas a variância amostral dos estimadores locais, mas também a correlação espacial que por ventura exista na região, a sugestão é de que ela seja utilizada para a derivação da distribuição a priori para κ .

É interessante notar que as análises WLS e GLS obtiveram zero para a variância do erro do modelo nas três regiões estudadas. Reis et al. (2005) apresentam uma discussão sobre este fato. Existe a possibilidade de ocorrer uma superestimativa da precisão do estimador regional. Reis et al. (2005) desenvolveram uma abordagem Bayesiana para regressão hidrológica que evita este potencial problema.

A Tabela 7 mostra a média e a variância da distribuição a priori para κ , e os parâmetros p e q da distribuição Beta. As características da priori Geofísica também são apresentadas para servir de base de comparação com as *prioris* obtidas através de modelos de regressão. Observe que para cada uma destas três regiões, existe uma única distribuição a priori. Isto só acontece porque o melhor modelo de κ é a média regional. Se o modelo regional incluísse alguma característica da bacia, como por exemplo, sua declividade média, a distribuição a priori seria específica da estação de interesse, já que tanto o valor esperado de κ , quanto a variância de predi-

ção do modelo, seriam dependentes da característica da bacia em questão (ver eq. 14).

Tabela 7 - Média e variância da *priori* Regional para Muskingum, Tibagi e Estado do Ceará.

Parâmetro	<i>Priori</i>			
	Geofísica	Muskingum	Tibagi	Ceará
$\hat{\mu}_{\kappa_i}$	-0,100	-0,15	-0,11	-0,16
$\hat{\sigma}_{\kappa_i}^2$	$(0,122)^2$	$(0,013)^2$	$(0,023)^2$	$(0,020)^2$
p	6	259	68	186
q	9	476	105	370

A comparação entre as distribuições a *priori* pode ser feita visualmente através da Figura 8. Observa-se claramente que as *prioris* com base em modelos de regressão são mais informativas que a *priori* Geofísica, o que deve promover uma redução na incerteza dos estimadores de quantis de cheia, conforme ilustrado na Figura 6.

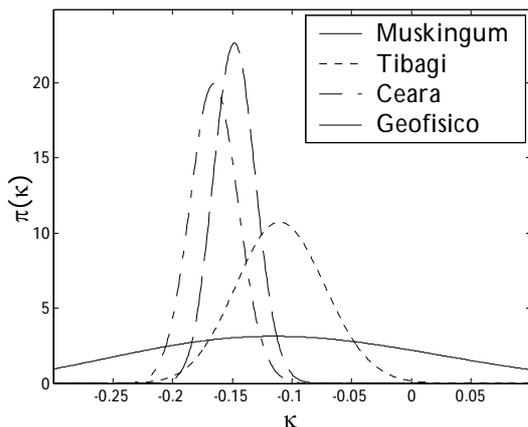


Figura 8 - Distribuições a *priori* obtidas através de modelos de regressão e a *priori* Geofísica.

CONCLUSÃO

Esse artigo compara, através de simulação Monte Carlo, o desempenho dos estimadores de quantis GML, ML, MOM e MOM-L para a distribuição GEV. Estimadores ML podem produzir estimativas absurdas do parâmetro de forma κ para amostras pequenas, o que explica o seu fraco desempenho para estimativa de quantis.

O uso do procedimento GML, baseado numa análise Bayesiana, emprega uma distribuição a

priori que restringe os valores de κ a uma faixa factível, tanto do ponto de vista físico das cheias quanto do ponto de vista estatístico, eliminando o problema que o estimador ML tem em amostras pequenas. O estimador GML de quantis utiliza uma distribuição a *priori* Beta para κ com $p=6$ e $q=9$ entre $[-0,5;0,5]$, de forma que $E[\kappa]=-0,10$ e $Var[\kappa]=0,015$. Esta distribuição foi denominada de *priori* Geofísica. O estimador GML de quantis, baseado na *priori* Geofísica, obteve um desempenho melhor que os estimadores MOM e ML para $\kappa \leq 0$, que é a região de interesse. Se κ é completamente positivo em uma dada região, uma *priori* mais apropriada pode ser adotada.

Os resultados deste estudo indicam que os estimadores de quantis GEV/GML devem ser empregados para análises locais no lugar dos estimadores MOM-L e MOM. Havendo a disponibilidade de dados fluviométricos de outros postos na região, é possível incorporar esta informação à análise de frequência de cheias através do uso de uma distribuição a *priori* mais informativa de κ , obtida através de modelos de regressão.

O artigo apresenta um procedimento, baseado em regionalização hidrológica, para obtenção de distribuições a *priori* mais informativas para o parâmetro de forma κ . Este procedimento é baseado em modelos de regressão, que definem uma relação empírica entre os valores de κ e as características climáticas e fisiográficas da bacia. O valor esperado e a variância de predição destes modelos de regressão regional são utilizados para derivar uma *priori* mais informativa para κ baseada na distribuição Beta, que pode ser então utilizada com o procedimento GML para obtenção de quantis de cheias mais precisos.

Modelos regionais de κ foram desenvolvidos para bacias dos rios Muskingum e Tibagi e do Estado do Ceará. Como forma de ilustrar a importância de se considerar a variância amostral dos estimadores de κ , assim como a correlação espacial entre eles, os modelos regionais foram obtidos com o emprego das análises OLS, WLS e GLS. As diferenças encontradas nos resultados destas análises foram discutidas.

Os resultados do estudo de regionalização hidrológica foram utilizados para obter distribuições a *priori* mais informativas para κ . Com base na variância de predição destes modelos regionais, pode-se perceber que as *prioris* obtidas neste estudo, baseadas em modelos de regressão, são mais informativas do que a *priori* Geofísica sugerida por Martins e Steidinger (2000), o que deve proporcionar um aumento na precisão dos estimadores GML de quantis de cheias.

REFERÊNCIAS

- BERGER, J.O. Statistical decision theory and Bayesian analysis (Springer series in statistics), 2nd edition, Springer-Verlag, New York, 1985.
- CHOWDHURY, J.U.; STEDINGER, J.R. e LU, L.-H. Goodness of fit tests for regional generalized extreme value flood distributions, *Water Resour. Res.*, 27(7), 1765-1776, 1991.
- COLES, S.G. e DIXON, M.J. Likelihood-based inference for extreme value models, submitted to *Extremes*, 1998.
- DALRYMPLE, T. Flood frequency analysis, *U.S. Geol. Surv. Water Supply Pap.*, 1543-A, 1960.
- JENKINSON, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quart. J.R. Met.Soc.*, 81, 158-171, 1955.
- JIN, M., e STEDINGER, J.R. Flood frequency analysis with regional and historical information, *Water Resour. Res.*, 25(5), 925-36, 1989.
- de HAAN, L. e de RONDE, J. Sea and wind: multivariate extremes at work, *Extremes*, 1(1), 7-45, 1998.
- FARQUHARSON, F.A.K.; MEIGH, J.R. e SUTCLIFFE, J.V. Regional flood frequency analysis in arid and semi-arid areas, *J. of Hydrol.*, 138, 487-501, 1992.
- FILL, H. D. Improving flood quantile estimates using regional information, Ph.D. Dissertation, Cornell University, Ithaca, New York, 1994.
- FILL, H.D. e STEDINGER, J. R. Using regional regression within index flood procedures and an empirical Bayesian estimator, *J. of Hydrol.*, 210, pp. 128-145, 1998.
- GREIS, N.P. e WOOD, E.F. Regional flood frequency estimation and network design, *Water Resour. Res.*, 17(4), 1167-1177, 1981. (Correction: *Water Resour. Res.*, 19(2), 589-590, 1983.)
- HOSKING, J.R.M. *Regional Frequency Analysis: An Approach Based on L-moments*, Cambridge University Press, 1997.
- HOSKING, J.R.M. L-Moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. of Royal Statistical Society, B*, 52(2), 105-124, 1990.
- HOSKING, J.R.M. Algorithm AS 215: Maximum-Likelihood Estimation of the parameter of the Generalized Extreme-Value distribution, *Appl. Statist.*, 34, 301-310, 1985.
- HOSKING, J.R.M, e WALLIS, J.R. The effect of intersite dependence on regional frequency analysis, *Water Resour. Res.*, 24(1), 588-600, 1988.
- HOSKING, J.R.M.; WALLIS, J.R. e WOOD, E.F. Estimation of the generalized extreme-value distribution by the method of probability weighted moments, *Technometrics*, 27(3), 251-261, 1985a.
- HOSKING, J.R.M.; WALLIS, J.R. e WOOD, E.F. An appraisal of the regional flood frequency procedure in the UK Flood Studies Report, *Hydrol. Sci. J.*, 30(1), 85-109, 1985b.
- KUCZERA, G. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian Inference, *Water Resour. Res.*, 35(5), 1551-1557, 1999.
- LANDWEHR, J.M.; MATALAS, N.C. e WALLIS, J.R. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resour. Res.*, 15(5), 1055-1064, 1979.
- LEHMANN, E.L. e CASELLA, G. *Theory of point estimation*, 2nd Edition, Chapman and Hall, New York, 1996.
- LETTENMAIER, D.P.; WALLIS, J.R. e WOOD, E.F. Effect of regional heterogeneity on flood frequency estimation, *Water Resour. Res.*, 23(2), 313-323, 1987.
- MACLEOD, A.J. AS R76-A remark on algorithm AS 215:Maximum-likelihood estimation of the parameters of the GEV distribution, *Appl. Statist.*, 38, 198-199, 1989.
- MADSEN, H. e ROSBJERG, D. Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modeling, *Water Resour. Res.*, 33(4), 771-782, 1997.
- MADSEN, H.; RASMUSSEN, P.F. e ROSBJERG, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling, *Water Resour. Res.*, 33(4), 747-758, 1997a.
- MADSEN, H.; PEARSON, C.P. e ROSBJERG, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling, *Water Resour. Res.*, 33(4), 759-770, 1997b.
- MARTINS, E.S. e STEDINGER, J.R. Cross-correlations among estimators of shape, submitted to *Water Resour. Res.*, 2002.
- Martins, E. S., J. R. STEDINGER, Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resour. Res.*, 36(3), 737-744, 10.1029/1999WR900330, 2000.
- NATURAL ENVIRONMENT RESEARCH COUNCIL. *Flood Studies Report*, vol. 1, London, 1975
- REIS, D.S., Jr & STEDINGER, J.R.. Bayesian MCMC Flood Frequency Analysis with Historical Information. *Journal of Hydrology*, vol. 313, no 1-2, pp. 97-116, 2005.
- REIS, D. S., JR., J. R. STEDINGER, E. S. MARTINS, Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation, *Water Resources Research*, 41, W10419, doi:10.1029/2004WR003445, 2005.

- ROBERT, C.P. *The Bayesian choice: a decision-theoretic motivation (Springer texts in statistics)*, Springer-Verlag, New York, 1994.
- STEDINGER, J.R.; e LU, L.-H. Appraisal of regional and index flood quantile estimators, *Stochastic Hydrol. Hydraul.*, 9(1), 49-75, 1995.
- STEDINGER, J.R. e TASKER, G. Correction to "Regional hydrologic analysis, 1, Ordinary, weighted and generalized least squares compared", *Water Res. Research*, 22(5), 844, 1986a.
- STEDINGER, J.R. e TASKER, G. Regional hydrologic analysis, 2: Model-error estimators, estimation of sigma and log-Pearson Type 3 distributions, *Water Res. Research*, 22(10), 1487-1499, 1986b.
- STEDINGER, J.R. e G.D. TASKER, Regional Hydrologic Analysis, 1. Ordinary, Weighted and Generalized Least Squares Compared, *Water Resources Research*, 21(9), 1421-1432, 1985. [with correction, *Water Resour. Res.* 22(5), 844, 1986].
- STEDINGER, J.R.; VOGEL, R.M. e FOUFOULA-GEORGIOU, E. Frequency analysis of extreme events, in *Handbook of Applied Hydrology*, chapt. 18, editor-in-chief D.A. Maidment, McGraw-Hill, New York, 1993.
- TASKER, G. D. Hydrologic regression and weighted least squares, *Water Resour. Res.*, 16 (6), 1107-1113, 1980.
- TASKER, G. D. e STEDINGER, J.R. An operational GLS model for hydrologic regression, *J. of Hydrol.*, 111, 361-357, 1989.
- TASKER, G. D. e STEDINGER, J.R. Estimating generalized skew with weighted least squares regression, *J. Water Resour. Plann. Manage.*, 112(2), 225-237, 1986.
- THOMAS, D. M. e BENSON, M. A. Generalization of stream-flow characteristics from drainage-basin characteristics, *U.S. Geol. Surv. Water Supply Pap.*, 1975, 55 pp., 1970.
- WALLIS, J.R.. Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures, in *Piense e Siccità*, edited by E. Guggino, G. Rossi, and E. Todini, pp. 3-36, Fondazione Politecnica del Mediter., Catania, Italy, 1980.
- WALLIS, J.R., e WOOD, E.F. Relative accuracy of log Pearson III procedures, *J. Hydraul. Eng.*, 111(7), 1043-1056, 1985.
- WILLEKE, G.E.; HOSKING, J.R.M.; WALLIS, J.R. e GUTTMAN, N.B. The National Drought Atlas (draft). *IWR Report 94-NDS-4*, U.S. Army Corps of Engineers, Fort Belvoir, Va, 1995.

Use of Regional Information to Estimate GEV Flood Quantiles

ABSTRACT

This paper focuses on the use of regional information in GEV flood quantiles estimates. Previous studies show that the Maximum Likelihood (ML) estimator is very unstable when applied to small and moderate samples, often found in practice, especially in developing countries. The reason for this instability is the high uncertainty in the shape parameter (κ) estimate, which may lead to absurd flood quantile estimates. Because of this, L-moment estimators are often preferred. This paper shows that the use of a prior informative distribution using a Generalized Maximum Likelihood (GML) procedure can solve the instability problems in MLE. Results based on Monte Carlo simulation show that the GML estimator is more precise than MOM-L estimators for $-0,4 \leq \kappa \leq 0$. This paper also introduces a procedure, based on regionalization of hydrologic variables, to derive even more informative prior distributions in order to improve the precision of flood quantile estimates.

Keywords: Flood quantile estimates; GEV