

## Regionalização Hidrológica de Vazões Mínimas por meio dos Métodos OLS e WLS Aplicada à Bacia do Alto São Francisco

Sarah Figueiredo Luiz<sup>1</sup>, Wilson dos Santos Fernandes<sup>1</sup>, Dirceu Silveira Reis Júnior<sup>2</sup>

sarahflui@yaho.com.br; wilson@ehr.ufmg.br; dirceu.reis@gmail.com

Recebido: 04/03/13 - revisado: 25/04/13 - aceito: 05/08/13

---

### RESUMO

O conhecimento de vazões em locais não monitorados é de grande importância em hidrologia. Daí a utilização de técnicas que possibilitem a transferência de informações de locais monitorados para locais sem monitoramento. Além disso, o estudo das vazões mínimas também desempenha um importante papel na avaliação da disponibilidade hídrica, na elaboração de projetos hidroelétricos e de irrigação, além de serem utilizadas como critério de outorga. Portanto, neste trabalho foi empregado o modelo de regionalização de vazões mínimas ( $Q_{7,10}$ ) para a bacia do Alto São Francisco, sub-bacias 40 e 41. Os métodos de regressão utilizados foram OLS e WLS, sendo este último aplicado como proposto por Tasker e Stedinger. Para aplicação do método de regionalização, dividiu-se a área de estudo em 4 regiões supostamente homogêneas, as quais são referenciadas com os códigos BA1, BA2, BA3 e BA4. Realizou-se a análise de frequência local para todas as estações que possuíam mais de 10 anos de dados observados, utilizando a distribuição de probabilidade Weibull 2 parâmetros. A análise e a determinação do modelo final para cada uma das 4 regiões foi feita mediante a avaliação da variância média de predição (AVP). Os resultados obtidos mostraram que o método WLS teve melhor desempenho que o método OLS, em relação à variância do erro do modelo. A fim de que se pudessem avaliar os critérios relacionados aos resíduos do método OLS, aplicou-se testes de hipóteses de independência, normalidade e variância constante ao nível de significância de 5%. Essa análise permitiu evidenciar que para as regiões BA1, BA2 e BA4, a utilização do método OLS não é recomendada, pois pelo menos uma premissa referente aos resíduos foi violada.

**Palavras Chave:** Variância do erro do modelo. Variância média de predição. Homocedasticidade.

---

### INTRODUÇÃO

De acordo com WMO (1974), as vazões mínimas são definidas como sendo “a vazão que escoar em uma determinada seção de um rio durante um período prolongado de seca”.

Segundo Tucci (2002), a vazão mínima diária é pouco útil, por não ser representativa do período de estiagem. Sendo assim, usualmente utilizam-se vazões mínimas com durações médias entre 7 e 30 dias, associadas a um determinado período de retorno.

A principal justificativa da utilização de vazões mínimas anuais, tal como a  $Q_{7,10}$ , reside no fato destas serem as principais balizadoras no que se refere a licenças ambientais, avaliação de impacto ambiental de obras hidráulicas, além de ser um parâmetro de disponibilidade hídrica muito empregado

em análises de concessão de outorga. De fato, no estado de Minas Gerais, por exemplo, a Portaria administrativa IGAM nº 010/98 de 30 de dezembro de 1998 em seu artigo 8º regulamenta como vazão de referência o equivalente à  $Q_{7,10}$  (vazão mínima de 7 dias de duração e dez anos de tempo de retorno).

Assim, torna-se evidente a importância de estudos no sentido de estimar tais vazões de referência, sobretudo em locais não monitorados.

De acordo com CPRM (2001), “uma rede hidrometeorológica, mesmo que densa, dificilmente atenderá com seus dados às necessidades de informação dos diversos setores que utilizam os recursos hídricos. Sempre haverá a necessidade de se determinar a disponibilidade hídrica onde se originam as demandas, que muitas vezes são locais sem monitoramento ou com dados constituindo séries de curta duração ou com longos períodos apresentando falhas de observação. Para contornar esta situação, são utilizadas técnicas de regionalização que possibilitam a transferência de informações de uma região monitorada para outra que não possua coleta sistemática de dados”. Uma das formas de realizar a re-

---

<sup>1</sup> Universidade Federal de Minas Gerais

<sup>2</sup> Universidade de Brasília

gionalização é mediante o ajuste de um método de regressão.

O Método dos Mínimos Quadrados Ordinário - OLS (*Ordinary Least Square*), também conhecido como método dos mínimos quadrados convencional, tem diversas limitações. De acordo com Stedinger e Tasker (1985), este método assume que os erros associados às observações individuais são independentes, homocedásticos, ou seja, possuem a mesma variância e, são igualmente distribuídos.

Segundo Reis *et al.* (2005), a hipótese de homocedasticidade é frequentemente violada em problemas hidrológicos, uma vez que a precisão das estatísticas de interesse são diferentes em cada local, já que o tamanho das séries hidrológicas também o são. Além disso, dificilmente as estatísticas de interesse são independentes, já que os dados de vazão possuem, em geral, uma forte correlação espacial. Por outro lado, a correlação serial local é pequena, uma vez que as amostras são formadas por um único valor extremo anual.

Taylor *et al.* (2011) realizaram a análise regional de frequência de vazões para o Oeste da Austrália, onde aplicaram o método de regressão OLS para estimativa do quantil de vazões e para estimativa do parâmetros da distribuição de probabilidade Log-Pearson tipo 3, determinando aos quantis de vazões com períodos de retorno variando de 2, 5, 10, 20, 50 e 100. O estudo foi aplicado em 206 bacias hidrográficas. Tais autores relataram que o resultado de todos os modelos foram bons, com base em testes estatísticos e análise do  $R^2$  ajustado. Contudo, estes salientam que os resultados encontrados por meio da regressão GLS (Método dos Mínimos Quadrados Generalizados - *Generalized Least Square*), no estudo de Haddad *et al.* (2011), são provavelmente mais confiáveis.

Tasker (1980) recomendou a utilização do Método dos Mínimos Quadrados Ponderados - WLS (*Weighted Least Square*) em análises regionais de vazões máximas. Os resultados encontrados mostraram que a variância do erro do modelo para o método WLS foi muito inferior quando comparada à regressão OLS, isto porque no caso do OLS não há nenhuma distinção entre o erro do modelo regional e o erro amostral, relacionado com o tamanho da amostra.

Tasker e Stedinger (1986) empregaram a regressão WLS para obter um estimador generalizado para o coeficiente de assimetria. Os pesos para o método WLS foram determinados separando-se a variância do erro em duas partes, variância do erro do modelo e variância do erro amostral. Tais autores relataram que resultados encontrados através do

modelo WLS foram melhores que os do modelo OLS, principalmente no que se refere à variância do erro do modelo.

O presente artigo tem como objetivo avaliar o desempenho dos métodos OLS e WLS em um estudo de regionalização de vazões mínimas de 7 dias de duração e 10 anos de período de retorno,  $Q_{7, 10}$ , para bacia do Alto São Francisco.

## DESCRIÇÃO DA ÁREA DE ESTUDO

A área de estudo compreende as sub-bacias 40 e 41 do Rio São Francisco.

As sub-bacias 40 e 41 (classificação ANEEL) estão localizadas na região denominada como Alto São Francisco, cuja área total é de 90.634,16 km<sup>2</sup>, abrangendo 170 municípios no Estado de Minas Gerais e contendo uma população total de aproximadamente 6.774.302 habitantes (CPRM, 2001). O alto São Francisco apresenta variabilidade climática relacionada à transição de um clima úmido e subúmido (Silva e Clarke, 2004).

De acordo com CPRM (2001), a sub-bacia 40 está localizada na região central do Estado de Minas Gerais, possui uma área de cerca de 51.000 km<sup>2</sup>, compreendida entre as nascentes do rio São Francisco e a Barragem de Três Marias, com latitudes 18° 15' e 20° 40' Sul e longitudes 43° 50' e 46° 10' Oeste.

A sub-bacia 40 abrange o alto São Francisco e seus afluentes até a barragem de Três Marias, além dos rios Paraopeba, Pará, Borrachudo, Indaiá e Ajudas. A Figura 1 apresenta a localização da sub-bacia 40 no estado de Minas Gerais.

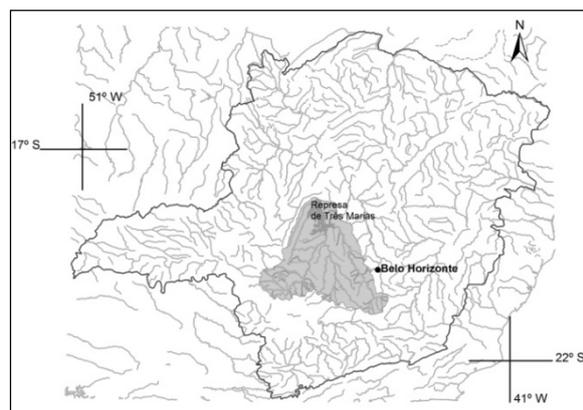
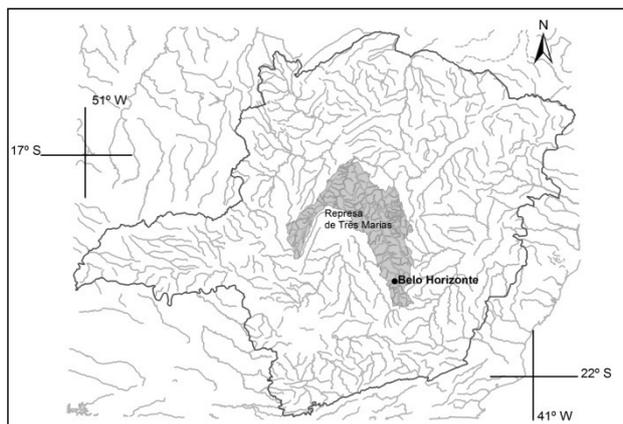


Figura 1 - Localização da sub-bacia 40 no Estado de Minas Gerais

A sub-bacia 41, de acordo com CPRM (2001), contém a parcela da bacia do rio São Francisco, com área aproximada de 39.000 km<sup>2</sup>, compreende o rio das Velhas e exclusive a barragem de Três Marias. Está localizada no centro do estado de Minas Gerais, entre as latitudes 17° 20' e 20° 30' Sul e longitudes 43° 30' e 46° 20' Oeste. O rio das Velhas e o rio Abaeté são os maiores tributários do rio São Francisco nesta bacia. A Figura 2 apresenta a localização da sub-bacia 41 no estado de Minas Gerais.



**Figura 2 - Localização da sub-bacia 41 no Estado de Minas Gerais**

## METODOLOGIA

Os métodos OLS e WLS foram aplicados às sub-bacias 40 e 41, do rio São Francisco, localizadas na região central de Minas Gerais. Estas bacias foram objeto de recentes estudos de regionalização de vazão e precipitação (CPRM, 2001; Euclides *et al.*, 2001, Lima, 2005 e IGAM, 2012).

As estações utilizadas no presente estudo são as de responsabilidade da Agência Nacional das Águas (ANA). Do total de 206 postos de observação, apenas 119 foram previamente selecionados, por possuírem pelo menos 10 anos de dados observados, um limite mínimo também empregado em vários estudos de regionalização da USGS (*United States Geological Survey*), como por exemplo, em Funkhouser *et al.* (2008), Capesius e Stephens (2009), Ahearn (2010), e Southard (2013). A definição do número mínimo de anos é sempre subjetiva, e deve-se ter em mente que quanto menor o número de anos, maior a incerteza na estimativa da estatística de interesse, nesse caso a  $Q_{7,10}$ . Porém, quanto menor o número de postos empregados na análise regional, maior a

incerteza na predição do modelo. Tasker (1989), por exemplo, num estudo de regionalização de vazões mínimas, empregou dados de 55 postos fluviométricos, sendo que em 6 deles a série histórica era igual ou menor há 10 anos. É importante notar que o método WLS, por considerar explicitamente os erros amostrais na estimativa da  $Q_{7,10}$ , reduz drasticamente a influência dessa escolha na análise, já que postos com série histórica curta terão menor influência na estimativa dos parâmetros do modelo regional, assim como na variância do erro do modelo. Outros estudos de regionalização empregaram postos fluviométricos com séries históricas menores do que 10 anos (por ex., RIES III, 1994; VOGEL *et al.*, 1999; WOTLING *et al.*, 2000). Após a análise de frequência local, restaram apenas 90 estações consistidas. Estas foram escolhidas devido ao fato de apresentarem melhor ajuste da curva empírica à distribuição Weibull 2P.

No levantamento das características físicas e climáticas, utilizou-se a mesma base de informações desenvolvida no estudo da CPRM (2001). Isto é devido ao fato deste estudo ter sido realizado com base em dados de vazões de períodos anteriores a 2000. Contudo, no presente trabalho, as séries de vazões utilizadas são mais extensas, o que possibilitou o melhor ajuste da curva empírica à distribuição teórica Weibull 2P.

Sendo assim, fez-se a coleta de informações utilizando o software Arcgis (9.3).

As características físicas e climáticas que foram avaliadas nas regressões são: Área de drenagem ( $A$ ), comprimento do talvegue principal ( $L$ ), declividade equivalente ( $I$ ), precipitação média anual ( $P$ ), coeficiente de escoamento superficial ( $C$ ) e densidade de drenagem ( $Dd$ ), embora se saiba que em estudos de vazões mínimas, seria mais apropriada a incorporação de grandezas hidrogeológicas. Entretanto, tais dados, quando existentes, não são facilmente disponibilizados.

### Ajuste da distribuição de frequência para determinação da $Q_{7,10}$

Na etapa de análise de frequência local, escolheu-se a distribuição de probabilidade Weibull dois parâmetros – Weibull (2P) para determinação do quantil  $Q_{7,10}$ . A escolha de tal distribuição é devida ao fato da mesma poder assumir forçosamente um limite inferior igual a zero, tornando assim uma forte candidata na modelagem de eventos hidrológicos mínimos (NAGHETTINI e PINTO, 2007).

De acordo com a Eletrobrás (1985), a prática tem demonstrado que vazões mínimas, com

quaisquer durações, se ajustam satisfatoriamente à distribuição de probabilidade Weibull.

Kaviski (1983) realizou um estudo de vazões de estiagens em pequenas bacias hidrográficas do Estado de Santa Catarina e aplicou as distribuições de probabilidades Log-Normal 3 parâmetros, Gumbel e Weibull. A distribuição que melhor se ajustou a seus dados foi a Weibull.

Mamun *et al.* (2010) em um estudo de regionalização de vazões mínimas para a Malásia, testaram os três tipos de distribuições Generalizadas de valores Extremos (GEV), a Log-Normal e a Log-Pearson tipo 3. A distribuição que melhor se ajustou foi Generalizada de Valores Extremos III (Weibull).

### Divisão da área de estudo em regiões homogêneas

A área de estudo foi dividida em 4 regiões supostamente homogêneas (Figura 3), não tendo sido realizada nenhuma análise estatística para corroborar que estas regiões realmente são homogêneas do ponto de vista hidrológico e estatístico. A divisão se justifica pela facilidade e clareza ao se apresentar os resultados, uma vez que os métodos OLS e WLS não necessitam dessa premissa, ao contrário que ocorre no método *index-flood*.

O critério utilizado para divisão dessas regiões foi através da análise das Unidades de Planejamento e Gestão de Recursos Hídricos (UPGRH), utilizadas pelo Instituto Mineiro de Gestão das Águas (IGAM). Sendo assim, as 4 regiões homogêneas são:

- Bacia do rio das Velhas - BA1;
- Bacia do rio Paraopeba - BA2;
- Bacia do rio Pará - BA3; e
- A união das bacias dos rios Ajudas, Bambuí, Indaiá, Borrachudo, Abaeté e a parcela do rio São Francisco na sub-bacia 40 - BA4.

Tabela 1 - Número de estações das regiões homogêneas

Regiões	Número de estações
<b>BA1</b>	29
<b>BA2</b>	20
<b>BA3</b>	18
<b>BA4</b>	23
Total = 90	

Na Figura 3 é apresentado o mapa da delimitação das regiões homogêneas na escala gráfica de 1:250.000. O número de estações para cada uma das regiões é apresentado na Tabela 1.

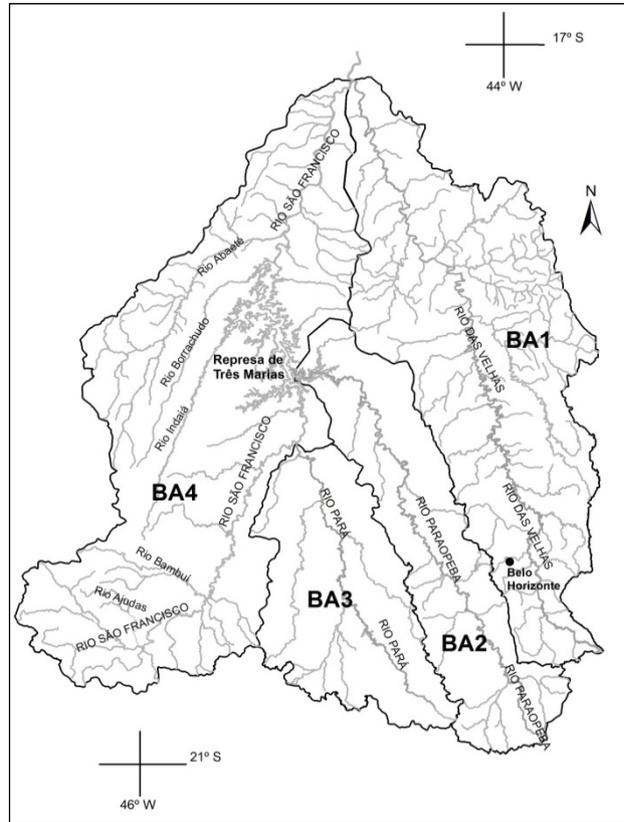


Figura 3 - Delimitação das regiões homogêneas

## MÉTODOS DE REGRESSÃO

### Método OLS

O método OLS assume que a variável hidrológica do vetor  $\hat{y}$ , pode ser descrita por uma função linear das características fisiográficas das bacias com uma componente de erro. Considere uma região com  $N$  postos de observação, sendo que cada posto possui  $P$  características fisiográficas (área de drenagem, declividade, precipitação média anual, etc.). O modelo de regressão na sua forma mais genérica para a variável  $\hat{y}$  é apresentado na Equação 1.

$$\hat{y} = X\beta + \varepsilon \quad (1)$$

onde  $\hat{y}$  é um vetor que contém as estimativas do valor verdadeiro de  $y$  em cada estação,  $X$  é uma matriz ( $N \times P+1$ ) contendo o valor um na primeira coluna e as  $P$  variáveis explicativas nas demais colunas,  $\beta$  é um vetor a ser estimado com os  $P+1$  parâmetros do modelo e  $\varepsilon$  é um vetor contendo os erros para cada um das  $N$  estações fluviométricas. Então, de acordo com o que é estabelecido na hipótese de homocedasticidade, a covariância entre os erros é igual a zero e a média também. Sendo assim, o estimador  $\hat{\beta}$  é dado por

$$\hat{\beta} = (X^T X)^{-1} X^T \hat{y} \quad (2)$$

De acordo com Stedinger e Tasker (1986b), o modelo básico do OLS assume que  $E(\hat{y}) = X\beta$ , e os resíduos são independentes e homocedásticos, de forma que

$$E[(\hat{y} - X\hat{\beta})(\hat{y} - X\hat{\beta})^T] = \delta^2 I \quad (3)$$

na qual,  $I$  é a matriz identidade ( $N \times N$ ) e  $\delta^2$  é a variância do erro do modelo.

A covariância dos erros amostrais para o estimador  $\hat{\beta}$  é dada por

$$\Sigma(\hat{\beta}_{OLS}) = \delta^2 (X^T X)^{-1} \quad (4)$$

O melhor estimador não enviesado de  $\delta^2$  é dado por

$$\delta^2_{OLS} = \frac{(\hat{y} - X\hat{\beta})^T (\hat{y} - X\hat{\beta})}{N - P - 1} \quad (5)$$

na qual  $N$  é o número de estações e  $P$  o número de variáveis explicativas introduzidas no modelo de regressão.

### Método WLS clássico

O método de regressão WLS, assim como o método OLS, é descrito pela Equação 1. No método WLS os resíduos são heterocedásticos. Assumindo-se que os erros ( $\varepsilon$ ) possuem média igual a zero, a matriz de covariância destes erros é dada por

$$E(\varepsilon \varepsilon^T) = \delta^2 W^{-1} \quad (6)$$

Onde  $W$  é uma matriz ( $N \times N$ ) com a diagonal principal  $w_i$  (pesos) e  $\delta^2$  é igual à variância do erro do modelo. Se  $W$  for igual à matriz identidade, os erros

são homocedásticos e o modelo se reduz ao Método OLS. No caso mais geral,  $W$  é definida de forma a representar a heterocedasticidade e a correlação entre os resíduos.

### Método WLS Tasker-Stedinger

A principal mudança introduzida no método WLS por Tasker e Stedinger (1986) é a separação da matriz de covariância dos erros do modelo ( $W$ ). O método assume que o erro total resulta de duas fontes: o erro do modelo  $\varepsilon_i$ , que é identicamente distribuído com média zero e variância constante, e os erros amostrais, advindos do fato do real valor de  $y_i$  não ser conhecido, sendo que somente as estimativas da variável de interesse são disponíveis. Nesse caso, a Equação (1) pode ser reescrita na Equação (7).

$$\hat{y} = X\beta + \omega + \varepsilon = X\beta + \eta \quad (7)$$

onde  $\omega$  é o erro dos estimadores amostrais. Assim, os erros do modelo de regressão  $\eta_i$  são a combinação do erro amostral da estimativa  $\hat{y}_i$  de  $y_i$  e do erro do modelo  $\varepsilon_i$ . O erro total  $\eta$  tem média zero e matriz de covariância dada pela Equação (8).

$$E(\eta \eta^T) = W = \delta^2 I + \Sigma \quad (8)$$

na qual  $\Sigma$  é a matriz diagonal das variâncias dos erros amostrais.

De acordo com Tasker e Stedinger (1986), o estimador dos parâmetros  $\beta$  para o método WLS é dado pela relação descrita na Equação 9.

$$\hat{\beta} = (X^T W^{-1} X)^{-1} X^T W^{-1} \hat{y} \quad (9)$$

A variância do erro do modelo é dada pela Equação 10.

$$\delta^2 = \frac{(y - X\hat{\beta})^T W^{-1} (y - X\hat{\beta})}{N - P - 1} \quad (10)$$

Stedinger e Tasker (1985), por meio de um estudo de simulação Monte Carlo, concluíram que o modelo WLS é mais eficiente do que o modelo OLS, em termos de erro médio quadrático, na estimativa de quantis regionais de variáveis hidrológicas. Quando a precisão das estimativas locais da variável de interesse são similares, ambos os métodos possuem desempenho semelhante.

No modelo WLS, a matriz  $\Sigma$  é elaborada levando em conta a distribuição de probabilidade definida na análise de frequência local (Weibull 2P) e o intervalo de confiança para os quantis. Os estimadores destes quantis podem ser obtidos através dos métodos de estimação de parâmetros, por exemplo, Métodos dos Momentos (MOM) e Método da Máxima Verossimilhança (MVS). No método WLS, assim como proposto por Tasker e Stedinger (1986), a matriz de covariância é dividida em duas partes (Equação 11) (REIS, 2005).

$$E(\eta(\tau)) = W = \begin{cases} \delta_\epsilon^2 + \delta_{\omega_{ij}}^2 & i = j \\ 0 & i \neq j \end{cases} \quad (11)$$

onde  $\delta_\epsilon^2$  é a variância do erro do modelo e  $\delta_{\omega_i}^2$  é a variância dos erros amostrais. De acordo com Tasker e Stedinger (1986), qualquer correlação entre  $\eta_i$  e  $\eta_j$  para  $i \neq j$  deve ser negligenciada.

Segundo Kite (1988), o estimador da variância dos erros amostrais para a distribuição Weibull é dada pela Equação (12). No entanto, para Weibull 2P, utiliza-se somente o termo  $T_1$  desta equação. Isso é devido o fator de frequência, depender somente do período de retorno, e do parâmetro de forma da distribuição Weibull, que por sua vez depende do coeficiente de variação. Entretanto, a Equação 12 foi apresentada em sua forma geral. No entanto, os termos  $T_2$  e  $T_3$  possuem valores pequenos e, têm pouca contribuição no resultado final, calculado por meio da Equação (12).

$$\delta_{\omega_i}^2 = \text{var}(\hat{y}_i) = \frac{\mu_2}{N} \{T_1 + T_2 + T_3\} \quad (12)$$

Onde  $T_1$ ,  $T_2$  e  $T_3$  são dados, respectivamente, pelas Equações 13, 14 e 15.

$$T_1 = 1 + K\gamma_1 + \frac{K^2}{4}[\gamma_2 - 1] \quad (13)$$

$$T_2 = \frac{\partial K}{\partial \gamma_1} \left[ 2\gamma_2 - 3\gamma_1^2 - 6 + K \left( \gamma_3 - \frac{6\gamma_1\gamma_2}{4} - \frac{10\gamma_1}{4} \right) \right] \quad (14)$$

$$T_3 = \left( \frac{\partial K}{\partial \gamma_1} \right)^2 \left[ \gamma_4 - 3\gamma_1\gamma_3 - 6\gamma_2 + \frac{9\gamma_1^2\gamma_2}{4} + \frac{35\gamma_1^2}{4} + 9 \right] \quad (15)$$

na qual  $\mu_2$  é a variância da amostra,  $K$  é o fator de frequência associado à distribuição de probabilidade e ao período de retorno,  $N$ ,  $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$  e  $\gamma_4$  correspondem, respectivamente, ao número de dados para cada estação, ao coeficiente de assimetri-

a, ao coeficiente de curtose, ao coeficiente de 5ª ordem e ao coeficiente de 6ª ordem populacionais.

### Construção do modelo final

Para cada uma das 4 regiões construiu-se um modelo regional cuja estimativa dos parâmetros foi feita através de um aplicativo em linguagem de programação Matlab. O modelo de regressão avaliado foi do tipo potencial. A utilização da equação do tipo potencial resulta numa dificuldade adicional, pois a metodologia para estimativa da matriz de variância é baseada nos quantis propriamente ditos, e não no logaritmo dos quantis. Essa transformação não linear certamente causa uma mudança nos elementos da matriz de covariância. Porém, não foi possível, até esse momento, avaliar as repercussões desse fato nos resultados finais do estudo de regressão.

Os procedimentos para a determinação dos parâmetros do modelo de regressão estão descritos abaixo:

- Primeiramente, criou-se uma matriz de modelos, a qual contém todas as combinações possíveis de modelos das características físicas e climáticas, juntamente com o intercepto. Sendo assim, fez-se o teste para 6 variáveis fisiográficas, correspondendo a 64 modelos;
- Encontrou-se a variância do erro do modelo (Equação 10) resolvendo-se, iterativamente, a Equação 8 para as 6 variáveis explicativas.
- Por fim, determinaram-se os parâmetros  $\beta$  do modelo de regressão.

A fim de avaliar a precisão dos modelos de regressão, utilizou-se a variância média de predição (*average variance of prediction - AVP*). De acordo com Tasker e Stedinger (1986), a AVP (Equação 16) é uma medida do desempenho médio do modelo regional em estimar os valores de  $Q_{7,10}$  em bacias similares àquelas utilizadas no estudo de regressão. Valores pequenos de AVP correspondem a melhores modelos de regressão (REIS *et al.*, 2005).

$$AVP = \delta_\epsilon^2 + \frac{1}{N} \sum_{i=1}^N x_i [X^T W^{-1} X]^{-1} x_i^T \quad (16)$$

Na avaliação dos resultados dos modelos OLS e WLS incorporaram-se outras análises estatísticas, tais como, o  $R^2$  (coeficiente de determinação).

Tabela 2 - Resultado dos modelos OLS e WLS

Regiões	Métodos de regressão	R <sup>2</sup>	R <sup>2</sup> ajustado	δ <sub>e</sub> <sup>2</sup>	AVP	Modelos
BA1	OLS	0,937	0,929	0,328	0,37	$Q_{7,10}=2,30 \times 10^4 A^{1,09} P^{8,38} Dd^{0,60}$
BA2	OLS	0,985	0,981	0,066	0,08	$Q_{7,10}=6,24 \times 10^4 A^{1,28} P^{0,55} C^{1,71} Dd^{0,43}$
BA3	OLS	0,925	0,902	0,186	0,24	$Q_{7,10}=5,74 \times 10^2 A^{0,85} L^{0,671} P^{5,57} C^{2,30}$
BA4	OLS	0,931	0,92	0,423	0,50	$Q_{7,10}=3,99 \times 10^5 A^{1,25} P^{9,70} C^{2,178}$
BA1	WLS	0,922	0,912	0,266	0,36	$Q_{7,10}=8,55 \times 10^4 A^{0,91} P^{7,92} Dd^{0,65}$
BA2	WLS	0,982	0,978	0,038	0,07	$Q_{7,10}=1,94 \times 10^2 A^{1,19} L^{-0,75} C^{2,99} Dd^{0,83}$
BA3	WLS	0,987	0,985	0,015	0,05	$Q_{7,10}=4,88 \times 10^3 A^{0,49} L^{0,64}$
BA4	WLS	0,986	0,984	0,055	0,12	$Q_{7,10}=6,96 \times 10^4 A^{0,85} P^{8,06} C^{1,60}$

Contudo, devido ao fato do coeficiente R<sup>2</sup> aumentar à medida que se inclui as variáveis explicativas, decidiu-se também avaliar o R<sup>2</sup> ajustado, o qual tende diminuir a tendência de aumento do R<sup>2</sup>, pois leva em consideração os graus de liberdade. Também avaliou-se, a variância do erro do modelo.

Realizaram-se testes para verificar se os resíduos da regressão OLS eram independentes, normalmente distribuídos e com variância constante (homocedástico). Os testes realizados foram: análise gráfica dos resíduos, teste de aderência de Anderson-Darling, teste da independência dos resíduos de Durbin-Watson e testes da variância constante Breusch-Pagan e White.

**RESULTADOS E DISCUSSÃO**

Os resultados dos modelos para as 4 regiões estão apresentados na Tabela 2. A escolha destes modelos foi feita segundo o critério da AVP.

Os modelos WLS obtiveram os menores valores da variância do erro do modelo e da variância média de predição. Isto acontece porque o modelo OLS não faz qualquer distinção entre o erro do modelo propriamente dito e os erros amostrais na estimativa da variável de interesse. Em outras palavras, a presença de erros amostrais causa uma superestimativa da variância do erro do modelo no caso OLS. As diferenças relativas encontradas na variância do erro do modelo entre os modelos OLS e WLS são significativas, variando aproximadamente entre 20 e 90%, embora as diferenças entre AVP sejam

menores, entre 3 e 79%, mas mesmo assim não desprezíveis, especialmente nas regiões BA3 e BA4.

Ao analisar o R<sup>2</sup> e o R<sup>2</sup> ajustado, os modelos do OLS se sobressaíram para as regiões BA1 e BA2, e os modelos do WLS tiveram melhor R<sup>2</sup> nas regiões BA2, BA3 e BA4. É importante salientar que, o R<sup>2</sup> e o R<sup>2</sup> ajustado foram avaliados unicamente entre os modelos de um mesmo método. Pois, de acordo com Reis (2005, p.98), no WLS e no GLS, o coeficiente de determinação é diferente do método OLS. Eles medem coisas diferentes já que no modelo OLS não se faz distinção entre erro do modelo e erro amostral.

Com o objetivo de avaliar se o modelo OLS poderia ser descartado a priori, avaliaram-se os resíduos da regressão OLS por meio de testes de hipóteses sobre independência, normalidade e constância da variância (Figura 4), ao nível de significância de 5%. Os resultados referentes à não rejeição (NR-H<sub>0</sub>) ou à rejeição (R-H<sub>0</sub>) da hipótese nula (H<sub>0</sub>), estão discriminados na Tabela 3.

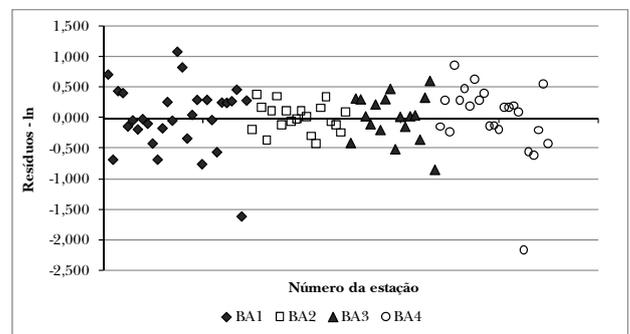


Figura 4 – Resíduos dos modelos OLS

É perceptível a presença de possíveis *outliers*, indicando a necessidade de avaliação dos pontos influentes dos modelos de regressão. Esta avaliação não foi realizada neste estudo, mas pode ser feita através de medidas de influência, tais como *leverage*, Cook's D, DFITS e DFBETAS (VEILLEUX, 2011).

**Tabela 3 - Testes dos resíduos para os modelos do OLS**

Regiões	Testes		
	Independência	Normalidade	Variância Constante
<b>BA1</b>	NR-H <sub>0</sub>	NR-H <sub>0</sub>	R-H <sub>0</sub>
<b>BA2</b>	NR-H <sub>0</sub>	NR-H <sub>0</sub>	R-H <sub>0</sub>
<b>BA3</b>	NR-H <sub>0</sub>	NR-H <sub>0</sub>	NR-H <sub>0</sub>
<b>BA4</b>	NR-H <sub>0</sub>	R-H <sub>0</sub>	NR-H <sub>0</sub>

Portanto, para as regiões BA1, BA2 e BA4 a utilização do método OLS não é recomendado, pois pelo menos uma das premissas referente aos resíduos foi violada. No entanto para a região BA3, a utilização do modelo de regressão OLS não pode ser descartada a *priori*. Contudo, sabe-se, com base em estudos de simulação Monte Carlo (STEDINGER e TASKER, 1985), que tanto os coeficientes da regressão WLS, quanto a estimativa da variância do erro do modelo, tendem a ser mais precisos que as do OLS, e por isso devem ser empregadas sempre que possível.

Na Tabela 4, estão apresentados o intervalo de validade, para cada uma das quatro regiões homogêneas.

**Tabela 4 - Intervalo de validade das equações regionais**

Região homogênea	Intervalo de validação para cada região
	A= área de drenagem (km <sup>2</sup> )
<b>BA1</b>	$85,0 \leq A \leq 26.553,0$
<b>BA2</b>	$10,2 \leq A \leq 8.734,0$
<b>BA3</b>	$144,0 \leq A \leq 11.350,0$
<b>BA4</b>	$259,0 \leq A \leq 14.180,0$

## CONCLUSÕES

Este artigo trata do uso dos modelos OLS e WLS na regionalização de variáveis hidrológicas, mais especificamente na regionalização de vazões

mínimas. Estudos anteriores, baseados em simulação Monte Carlo, mostraram que o uso do modelo WLS fornece estimativas mais precisas dos coeficientes do modelo regional, assim como da variância do erro do modelo. Esse resultado está ligado ao fato do modelo WLS, diferentemente do modelo OLS, levar em consideração a variância do erro amostral na estimativa das vazões mínimas, que está ligado ao tamanho da série histórica de cada posto fluviométrico empregado no estudo de regionalização. A dificuldade na utilização do modelo WLS está, muitas vezes, na dificuldade de se estimar a matriz de covariância dos erros amostrais.

Este artigo apresenta uma maneira de se estimar a matriz de covariância dos erros amostrais a ser utilizada num estudo de regressão de vazões mínimas para o caso específico da distribuição Weibull 2P, permitindo assim a utilização do modelo WLS. A metodologia proposta foi aplicada à bacia do Alto São Francisco.

Os resultados obtidos mostraram, como já era esperado, que o método WLS obteve valores menores da variância do erro do modelo, em relação ao OLS, assim como em relação à variância média de predição. Isto se deve ao fato do modelo OLS não fazer qualquer distinção entre erro do modelo regional propriamente dito e os erros amostrais associados às estimativas das vazões mínimas computadas em cada estação fluviométrica. Os erros amostrais, que são inevitáveis, acabam por causar uma superestimativa da variância do erro do modelo, o que leva também a distorções na estimativa dos coeficientes do modelo regional.

Teste de hipótese foram utilizados para avaliar se seria possível identificar, a *priori*, se o modelo OLS não seria recomendado para os dados em questão. Com base nesses testes, foi possível verificar que as regiões BA1, BA2 e BA4, o método OLS não é recomendado, pois foi evidenciada a violação de pelo menos uma das premissas referente aos resíduos. Entretanto, isso não significa dizer que o modelo OLS seria mais indicado que o modelo WLS na região BA3. Resultados do estudo de regionalização mostraram que há diferenças importantes, entre os modelos OLS e WLS, nos valores da variância do erro do modelo para a região BA3. As diferenças para essa região chegam a aproximadamente 90%, mostrando que os erros amostrais possuem um papel importante nessa região.

Os resultados obtidos foram condizentes com os encontrados por Tasker e Stedinger (1986), principalmente no que se refere à análise da variância do erro do modelo, mesmo que estes estudos tenham sido aplicados para variáveis e locais dife-

rentes. Por exemplo, em Tasker e Stedinger (1986), a variável regionalizada foi o coeficiente de assimetria da distribuição Log-Pearson tipo 3 e a área de estudo foi Illinois, nos EUA.

Por fim, vale notar que as vazões mínimas são fortemente correlacionadas, pois um período de seca é comum para todas as estações localizadas em uma mesma região. O método WLS, embora mais eficiente que o OLS, não leva em consideração a correlação espacial. Nessas situações, o método GLS (Método dos Mínimos Quadrados Generalizados - *Generalized Least Square*), sistematizado por Stedinger e Tasker (1985, 1986a e 1986b) é o mais indicado, porém o desafio de construir uma matriz de covariância dos erros é maior, pois é necessário entender as correlações da variável de interesse, que é função dos fluxos concorrentes.

## AGRADECIMENTOS

Os dois primeiros autores agradecem à FAPEMIG (APQ 04418-10) pelo apoio concedido a esta pesquisa, enquanto o terceiro autor agradece à FINEP pelo financiamento do projeto GERHINE (Chamada Pública MCT/FINEP/CT-Hidro - IGRH01/2007).

## REFERÊNCIAS

AHEARN, E. A. *Regional regression equations to estimate flow-duration statistics at ungaged stream sites in Connecticut*. U.S. Geological Survey, *Scientific Investigations Report 2010-5052*, in cooperation with the Connecticut Department of Environmental Protection, 2010.

CAPESIUS, J. P.; STEPHENS, V. C. *Regional regression equations for estimation of natural streamflow statistics in Colorado*. U. S. Geological Survey, *Scientific Investigations Report 2009-5136*, 2009.

CPRM. Companhia de Pesquisa de Recursos Minerais. Regionalização de vazões. Sub-bacias 40 e 41: Relatório Final. CPRM/ANEEL. Belo Horizonte, 2001.

EUCLYDES, H. P.; FERREIRA, P. A.; RUBERT, O. A. V.; SANTOS, R. M. Regionalização hidrológica na bacia do Alto São Francisco a Montante da Barra-

gem de Três Marias, Minas Gerais. *Revista Brasileira de Recursos Hídricos*, v. 6, n. 2, p. 81-105, 2001.

FUNKHOUSER, J. E.; ENG, K.; MOIX, M. W. Low-flow characteristics and regionalization of low-flow characteristics for selected streams in Arkansas. U. S. Geological Survey, *Scientific Investigations Report 2008-5065*, 2008.

HADDAD, K.; RAHMAN, A. AND KUCZERA, G. Comparison of Ordinary and Generalized Least Squares Regression Models in Regional Flood Frequency Analysis: A Case Study for New South Wales (technical paper). *Australian Journal of Water Resources*, v. 15, n.1, p. 59-70, 2011.

IGAM. Instituto Mineiro de Gestão das Águas. *Estudo de regionalização de vazão para o aprimoramento do processo de outorga no Estado de Minas Gerais / Grupo de Pesquisas em Recursos Hídricos da UFV*. Belo Horizonte, 2012.

KAVISKI, E. *Vazões de estiagens em pequenas bacias hidrográficas do estado de Santa Catarina*. In: *Simpósio Brasileiro de Hidrologia e Recursos Hídricos*, 5, 1983. Florianópolis. *Anais ... Florianópolis*: ABRH, 1983, p.43-67.

KITE, G. W. *Frequency and Risk Analysis in Hydrology*. Fort Collins (CO): Water Resources Publications, 1988.

LIMA, A. A. *Metodologia integrada para determinação da enchente de projeto de estruturas hidráulicas por meio de séries sintéticas de precipitação e modelos chuva-vazão*. 2005. Dissertação (Mestrado - Programa de Pós-Graduação em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

MAMUN, A. A.; HASHIM, A.; DAOUD, J. I. Regionalisation of low flow frequency curves for the Peninsular Malaysia. *Journal of Hydrology*, 381, p. 174-180, 2010.

NAGHETTINI, M.; PINTO, E. J. A. *Hidrologia estatística*. Belo Horizonte: CPRM, 2007.552p.

REIS, D. S. Jr. Flood frequency analysis employing bayesian regional regression and imperfect historical information. 2005. Tese (Ph. D. thesis) - Faculty of the Graduate School of Cornell University, Ithaca, New York, 2005.

- REIS, D. S., Jr.; STEDINGER, J. R.; MARTINS, E. S. Bayesian GLS Regression with application to LP3 Regional Skew Estimation. *Water Resources Research*, v. 41, W10419, 2005.
- RIES III, K. G. Development and application of Generalized Least Squares Regression Models to estimate Low flow duration discharges in Massachusetts. *U. S. Geological Survey, Water-Resources Investigations Report 94-4155*, 1994.
- SILVA, B. M.; CLARKE, R. T. Análise estatística de chuvas intensas na bacia do rio São Francisco. *Revista Brasileira de Meteorologia*, v. 19, n.3, p. 265-272, 2004.
- SOUTHARD, R. E. Computed statistics at streamgages, and methods for estimating low-flow frequency statistics and development of regional regression equations for estimating low-flow frequency statistics at ungaged locations in Missouri. *U. S. Geological Survey, Scientific investigations Report 2013-5090*, 2013.
- STEDINGER, J. R.; TASKER, G. D. Correction to Regional hydrologic analysis 1. Ordinary, weighted, and generalized least squares compared. *Water Resources Research*, v. 22, n. 5, p. 844, 1986a.
- STEDINGER, J. R.; TASKER, G. D. Regional Hydrologic Analysis, 2 Model-Error Estimators, Estimation of Sigma and Log-Pearson Type 3 Distributions. *Water Resources Research*, v. 22, n. 10, p. 1487, 1499, 1986b.
- STEDINGER, J. R.; TASKER, G. D. Regional hydrologic analysis 1: Ordinary, weighted, and generalized least squares compared. *Water Resources Research*, v. 21, n.9, p. 1421-1432, 1985.
- TASKER, G. D. Hydrologic Regression With Weighted Least Squares. *Water Resources Research*, v. 16, n. 6, p.1107-1113, 1980.
- TASKER, G. D. *Regionalization of low flow characteristics using logistic and GLS regression*. In proceedings of symposium on new directions for surface water modeling, Baltimore, IAHS Publ. No. 181, p. 323-331, 1989.
- TASKER, G. D.; STEDINGER, J. R. Regional Skew with Weighted LS Regression. *Journal of Water Resources Planning and Management*, v. 112, n. 2, p. 225-237, 1986.
- TAYLOR, M.; HADDAD, K.; ZAMAN, M.; RAHMAN, A. *Regional flood modelling in Western Australia: Application of regression based methods using Ordinary Least Squares*. International Congress on Modelling and Simulation, 19th, 2011, Perth, Austrália. Austrália: MODSIM2011, p.43-67, 2011.
- TUCCI, C. E. M. *Regionalização de vazões*. Porto Alegre, ANEEL-UFRGS, 2002.
- VEILLEUX, A. G.; *Bayesian GLS regression, Leverage, and influence for regionalization of hydrologic statistics*. 2011. Dissertation (Doctor of Philosophy) - Faculty of the Graduate School of Cornell University, Ithaca, New York, 2011.
- VOGEL, R. M.; WILSON, I.; DALY, C. Regional regression models of annual streamflow for the United States. *Journal of irrigation and drainage engineering, paper No. 18974*, 1999.
- WMO - World Meteorological Organization. *International Glossary of Hydrology*, WMO, Geneva, 1974.
- WOTLING, G.; BOUVIER, Ch.; DANLOUX, J.; FRITSCH, J. M. Regionalization of extreme precipitation distribution using the principal components of the topographical environment. *Journal of Hydrology*, v. 233, p. 86-101, 2000.

***Hydrological Regionalization Of Minimum Flows Using The OLS And WLS Method Applied To The Alto São Francisco River Basin***

**ABSTRACT**

*The development of techniques that allow one to transfer hydrologic information from one site to another is crucial for the estimation of flow statistics at ungauged sites, which is still a great challenge in Hydrology. Besides, estimation of low-flow statistics plays a key role in water availability studies, design of hydroelectric power plants and irrigation projects, and on water licensing and allocation processes. This paper employs OLS and WLS regional models for the 7-day 10-year low-flow statistic ( $Q_{7,10}$ ) for the Alto São Francisco River Basin. The whole basin was divided into four supposedly homogeneous regions (BA1, BA2, BA3, and BA4). Local frequency analyses were carried out for all sites with at least 10 years of data employing the 2-parameter Weibull distribution. The selection of the best regional model was based upon the average variance of prediction (AVP). Results show that the WLS model*

*error variances and AVP were smaller than those provided by the OLS model. In order to evaluate the OLS assumption of homoscedasticity, hypothesis tests were applied to OLS residuals (independence, normality and constant variance). This analysis showed that the OLS model is not appropriate for regions BA1, BA2 e BA4 because at least one of the OLS assumptions is violated.*

**Keywords:** *Model error variance. Average variance of prediction. Homoscedasticity.*