

## Regionalização de Vazões Médias de Longo Período Para o Estado do Ceará

*Alan Michell Barros Alexandre*

UFC – Departamento de Engenharia Hidráulica e Ambiental  
alan@funceme.br

*Eduardo Sávio P. R. Martins*

FUNCEME – Departamento de Recursos Hídricos  
martins@funceme.br, esm9@secrel.com.br

Recebido: 21/12/04 revisado: 26/07/05 aceito: 30/08/05

---

### RESUMO

*Estudos em recursos hídricos, e em particular para suprimento hídrico, necessitam de estimativas de vazão média de longo período, e quase sempre para bacias hidrográficas não monitoradas. Assim, objetivando determinar esta vazão média para uma bacia sem dados, torna-se necessário o uso de técnicas de regionalização. Este artigo introduz o uso dos mínimos quadrados generalizados (GLS) na regionalização da vazão média de longo período, o que permite não só levar em consideração o erro amostral e o erro do modelo, mas também a correlação cruzada entre os dados das diversas estações. Estes resultados são comparados com mínimos quadrados ponderados (WLS) e mínimos quadrados (OLS). Para geração dos modelos foram utilizados como variáveis explanatórias características climáticas e fisiográficas de 25 bacias do Estado do Ceará. Os diferentes modelos foram comparados utilizando como critérios a variância média de predição, a variância do erro do modelo e um coeficiente de determinação modificado. Devido à baixa correlação cruzada entre as vazões médias de anos concomitantes, os modelos GLS e WLS apresentam performances similares. O modelo adotado como regional é um WLS que utiliza como variáveis explanatórias a declividade, a precipitação média, e percentual da bacia situada no cristalino.*

**Palavras-chave:** Regionalização; vazões médias; suprimento hídrico.

---

### INTRODUÇÃO

A vazão média de longo período permite caracterizar a maior vazão possível de ser regularizada em uma bacia permitindo a avaliação dos limites superiores (abstraindo as perdas) da disponibilidade de água de um manancial. A vazão média de longo período é definida como a média das vazões médias anuais para toda a série de dados, sendo denominada como específica quando dividida pela área da bacia hidrográfica de interesse.

Para uma bacia com dados este cálculo é uma tarefa simples. No entanto, nem sempre há disponibilidade de dados de vazão ou séries suficientemente longas para um determinado local de interesse. Nestes casos o uso de informação regional torna-se importante para a obtenção de estatísticas mais precisas em locais com séries curtas ou sem dados.

Diversos estudos mostram a preferência de alguns pesquisadores ao uso do modelo de

regressão GLS (*Generalized Least Squares* – mínimos quadrados generalizados), o qual leva em consideração a heterogeneidade da variância amostral e a correlação cruzada dos resíduos do modelo de regressão, tornando o modelo mais robusto (Stedinger e Tasker, 1985; Stedinger e Tasker, 1986b; Tasker e Stedinger, 1989; Madsen e Rosbjerg, 1997; Muller e Fill, 1997; Pandey e Nguyen, 1999; Kroll e Stedinger, 1999; Madsen et al., 2002; Martins e Stedinger, 2002b; Markus et al, 2003; Reis Jr. et al, 2003; Reis Jr. et al, 2004; Reis Jr. et al, 2005). A dificuldade na estimativa da matriz de covariância ainda é um limitador no uso mais amplo deste tipo de regressão.

Para a proposta de regionalização, as características fisiográficas e climáticas médias são consideradas como uniformes dentro de uma bacia, podendo assim ser utilizadas como variáveis explanatórias para os modelos que determinam a vazão média de longo período. Vários estudos ilustram o uso destas características (Reimers, 1990;

Riggs, 1990; Tucci, 1991; Vogel et al., 1999; Euclides e Ferreira, 2001; Góes e Campana, 2000; Tucci, 2002; Porto, 2003; Pedrollo e Ferreira, 2003; Euclides e Ferreira, 2003; Pinto e Alves, 2003).

No Estado do Ceará existe uma pequena rede de estações fluvimétricas em operação com séries não muito longas e mal distribuídas espacialmente. Torna-se evidente a importância do uso adequado dos dados disponíveis, justificando o estudo no escopo deste trabalho.

Neste trabalho, é introduzida uma abordagem baseada no uso do modelo de regressão GLS para a regionalização das vazões médias de longo período. Os resultados obtidos com o GLS foram comparados com WLS (*Weighted Least Squares* – mínimos quadrados ponderados) e o OLS (*Ordinary Least Squares* – mínimos quadrados).

### Caracterização dos dados

Tabela 1 – Intervalo de variação das variáveis explanatórias

Variável	Mínimo	Máximo
Dec (m/km)	1,7	22,1
P (mm)	590	1152
CT (km)	9	264
CTD (km)	37	14309
DD (km <sup>-1</sup> )	0,48	0,99
CAD (mm)	40	97
CN (mm)	64	83
Kc	1,2	1,9
Dir	1	8
Cr	0	100
L (m <sup>2</sup> /km <sup>2</sup> )	0	6146

Para este estudo foram utilizadas 25 estações fluviométricas (ver Figura 2) que cobrem aproximadamente 50% do Estado do Ceará (-02°46', -41°26' e -37°13', -07°50'). As estações selecionadas têm uma série de no mínimo 10 anos de dados e não estão sob a influência de reservatórios artificiais. A média de registros nas estações é de 28 anos e as vazões médias específicas de longo período variam entre 21 e 348 mm. As características das estações estão apresentadas na Tabela 5. Os dados climáticos e fisiográficos das estações listados como variáveis explanatórias das regressões são: (Dec) declividade em m/km, (P) precipitação média em mm, (CT) comprimento do talvegue em km, (CTD) comprimento total de drenagem em km, (DD) densidade de drenagem

km<sup>-1</sup>, (CAD) capacidade de armazenamento do solo em mm (valor estimado considerando uma profundidade média de 1 m), (CN) número de curva do SCS médio, (Kc) índice de compactidade da bacia, (Dir) direção dominante do rio principal, seguindo uma numeração de 1 a 8 com o sentido sul-norte igual a 1 e modificando a numeração a cada 45° no sentido horário, e (Cr) percentual da bacia situada no cristalino e (L) espelhos de lagos naturais e artificiais na bacia em m<sup>2</sup>/km<sup>2</sup>. A Tabela abaixo mostra o intervalo de variação de cada uma das variáveis explanatórias utilizadas.

### METODOLOGIA

A regressão GLS (*Generalized Least Squares* – mínimos quadrados generalizados) foi primeiramente utilizada pelo USGS (*US Geological Survey*) e documentada em vários estudos (Stedinger e Tasker, 1985, 1986a, 1986b; Tasker e Stedinger, 1989). Essa análise leva em consideração diferentes pesos para as vazões observadas de acordo com o tamanho da amostra, a correlação cruzada com outras estações e a estrutura do erro do modelo. Em Stedinger e Tasker (1985) e (1986b) é demonstrado que, quando o tamanho dos registros varia amplamente e as vazões apresentam correlação cruzada, o uso da técnica GLS resulta em maior acuracidade na estimativa dos parâmetros, e em Stedinger e Tasker (1986b) que este apresenta uma tendenciosidade muito menor no estimador da variância dos resíduos quando comparados com WLS (*Weighted Least Squares* – mínimos quadrados ponderados) e OLS (*Ordinary Least Squares* – mínimos quadrados).

O estimador regional para a vazão média específica de longo período ( $\mu$ ) de um dado local  $i$ , denotado por  $\hat{\mu}_i$  (o valor “observado” da variável resposta), está sujeito a um erro amostral  $\varepsilon_i$ , de tal forma que

$$\hat{\mu}_i = \mu_i + \varepsilon_i \quad \text{para } i = 1, 2, \dots, M \text{ estações} \quad (1)$$

onde

$$E[\varepsilon_i] = 0 \text{ e } \Sigma_{ij} = \text{Cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma_{\varepsilon_i}^2 & i = j \\ \sigma_{\varepsilon_i} \sigma_{\varepsilon_j} \rho_{\varepsilon_i \varepsilon_j} & i \neq j \end{cases} \quad (2)$$

Aqui  $\sigma_{\varepsilon_i}^2$  é a variância do erro amostral de  $\hat{\mu}_i$  para o local  $i$ , e  $\rho(\varepsilon_i, \varepsilon_j)$  é o coeficiente de correlação do erro amostral devido à correlação

entre as vazões médias dos locais  $i$  e  $j$ , ou seja,  $\rho(\hat{\mu}_i, \hat{\mu}_j)$ .

O valor de  $\mu$  pode ser relacionado às características climáticas e fisiográficas de uma bacia (Vogel *et al.*, 1999) por

$$\mu_i = \beta_0 + \sum_{j=1}^p \beta_j A_{ij} + \delta_i \quad (3)$$

onde  $\beta_0$  é o valor constante do modelo de regressão;  $\beta_j$  são os coeficientes que multiplicam as variáveis explanatórias  $A_{ij}$  ( $j=1\dots p$ ), ou seja, as características fisiográficas e climáticas utilizadas no modelo de regressão;  $p$  é o número de variáveis explanatórias; e  $\delta_i$  o erro do modelo adotado. Assim, combinando (1) e (3) tem-se que

$$\hat{\mu}_i = \beta_0 + \sum_{j=1}^p \beta_j A_{ij} + \eta_i \quad (4)$$

com  $\eta_i = \varepsilon_i + \delta_i$ . O erro do modelo segue as seguintes características

$$E[\delta_i] = 0 \quad \text{e} \quad \text{Cov}(\delta_i, \delta_j) = \begin{cases} \sigma_\delta^2 & i = j \\ 0 & i \neq j \end{cases} \quad (5)$$

Combinando as características dos erros amostrais (2) com as características do erro do modelo (5) obtêm-se a matriz de covariância  $\Lambda$  com elementos

$$\Lambda_{ij} = \text{cov}(\eta_i, \eta_j) = \begin{cases} \sigma_{\varepsilon_i}^2 + \sigma_\delta^2 & i = j \\ \sigma_{\varepsilon_i} \sigma_{\varepsilon_j} \rho(\hat{\mu}_i, \hat{\mu}_j) & i \neq j \end{cases} \quad (6)$$

Em notação matricial o modelo pode ser escrito da seguinte forma

$$\begin{aligned} \hat{\mu} &= X\beta + \eta \\ \hat{\mu} &= (\hat{\mu}_1 \hat{\mu}_2 \dots \hat{\mu}_M)^T \\ \beta &= (\beta_0 \beta_1 \dots \beta_p)^T \\ \eta &= (\eta_1 \eta_2 \dots \eta_M)^T \end{aligned} \quad X = \begin{pmatrix} 1 & A_{11} & \dots & A_{1p} \\ \dots & \dots & \dots & \dots \\ 1 & A_{M1} & \dots & A_{Mp} \end{pmatrix} \quad (7)$$

O estimador de  $\beta$  e sua covariância são dados pela equação (8), proposta por Tasker e Stedinger (1986),

$$\begin{aligned} \hat{\beta} &= (X^T \Lambda^{-1} X)^{-1} X^T \Lambda^{-1} \hat{\mu} \\ \text{Cov}(\hat{\beta}) &= (X^T \Lambda^{-1} X)^{-1} \end{aligned} \quad (8)$$

onde a variância dos estimadores de  $\beta$ ,  $\text{Var}[\hat{\beta}]$ , é igual a diagonal principal da matriz de covariância destes estimadores,  $\text{Cov}(\hat{\beta})$ , definida na equação (8). Como existe uma co-dependência entre a matriz de covariância ( $\Lambda$ ) e o erro do modelo ( $\sigma_\delta^2$ , equação 12), as estimativas de  $\beta$ , sua covariância e a estimativa do erro do modelo são determinadas iterativamente utilizando-se tanto a equação (8) como a (12).

**Estimativa da matriz de covariância**

A estimativa da matriz de covariância, excetuando-se a variância do erro do modelo, pode ser feita sabendo que

$$\sigma_{\varepsilon_i}^2 = \text{VAR}[\hat{\mu}_i] = \frac{\sigma_i^2}{n} \quad (9)$$

e que a covariância entre  $\hat{\mu}_i$  e  $\hat{\mu}_j$  é  $\text{Cov}(\hat{\mu}_i, \hat{\mu}_j) = \frac{\sigma_i \sigma_j}{n} \rho_{ij}$  (Stuart e Keith, 1994), onde  $n$  é o número de anos da série de vazões médias anuais;  $\sigma_i$  é desvio padrão da série de vazões médias anuais na estação  $i$ ; e  $\rho_{ij}$  é a correlação entre as vazões médias anuais das estações  $i$  e  $j$ .  $\text{VAR}[\hat{\mu}_i]$  é a estimativa local da variância do erro amostral de  $\hat{\mu}_i$ . Assim, tem-se

$$\begin{aligned} \rho(\hat{\mu}_i, \hat{\mu}_j) &= \frac{\text{Cov}(\hat{\mu}_i, \hat{\mu}_j)}{\sqrt{\text{VAR}[\hat{\mu}_i] \cdot \text{VAR}[\hat{\mu}_j]}} = \frac{\frac{\sigma_i \sigma_j}{n} \rho_{ij}}{\frac{\sigma_i \sigma_j}{n}} = \rho_{ij} \quad (10) \\ \therefore \rho(\hat{\mu}_i, \hat{\mu}_j) &= \rho_{ij} \end{aligned}$$

Para levar em consideração a possível diferença entre os tamanhos dos registros, pode-se usar a correção proposta por Martins e Stedinger (2002a).

$$\rho(\hat{\mu}_i, \hat{\mu}_j) = \frac{n_{ij}}{\sqrt{(n_i + n_{ij})(n_j + n_{ij})}} \rho_{ij} \quad (11)$$

onde  $n_{ij}$  é o período com dados em ambas estações  $i$  e  $j$ , e  $n_i$  ( $n_j$ ) o período com dados somente na estação  $i$  ( $j$ ).

A estimativa de  $\sigma_\delta^2$  pelo método dos momentos segue a solução dada por Stedinger e Tasker (1985), sujeita à condição  $\sigma_\delta^2 \geq 0$  é

$$\sigma_{\delta}^2 = \frac{(\mu - X\hat{\beta})^T \Lambda^{-1} (\mu - X\hat{\beta})}{M - p - 1} \quad (12)$$

onde  $\mu$  é um vetor com os valores amostrais de vazões médias específicas de longo período, e  $M$  o número de estações utilizadas na regionalização. Como mostrado anteriormente a  $\varphi$  é função de  $\sigma_{\delta}^2$  assim como  $\hat{\beta}$ . Como já mencionado na seção anterior, os valores de  $\sigma_{\delta}^2$  e  $\hat{\beta}$  são calculados simultaneamente utilizando as equações (8) e (12).

O valor de  $\sigma_{\delta}^2$  pode ser visto como uma medida de heterogeneidade dos resíduos do modelo. Quando o modelo regional selecionado é uma constante, o referido valor pode ser visto como uma medida da heterogeneidade das vazões médias de longo período (Madsen e Rosbjerg, 1997).

Quando na matriz de covariância ( $\varphi$ ), apresentada na equação (6), a parcela  $\rho(\hat{\mu}_i, \hat{\mu}_j)$  é nula para qualquer par de estações (i,j), o GLS reduz-se ao WLS. Neste caso, tem-se elementos apenas na diagonal principal, ou seja, quando  $i=j$  na equação 6. Se além disto, a variância amostral ( $\sigma_{\hat{\mu}_i}^2$ ) é nula para qualquer que seja  $i$ , a matriz de covariância reduz-se a  $\sigma_{\delta}^2 I$  (equação 6, quando  $i=j$ ), onde  $I$  é a matriz identidade, obtêm-se o OLS clássico.

Deve-se ressaltar que o uso de estimadores amostrais de  $\rho_{ij}$  pode resultar em uma matriz covariância ( $\hat{\Lambda}$ ) que nem sempre é positiva definida devido a incertezas e variações na extensão de dados concomitantes. Para evitar estes tipos de problemas Tasker e Stedinger (1989) propuseram relacionar  $\rho_{ij}$  com a distância entre as estações ( $d_{ij}$ ). Então, em vez de utilizar o estimador da correlação amostral  $\rho_{ij}$ , pode-se utilizar como estimador:

$$\hat{\rho}_i = \hat{\rho}(d_{ij}) \quad (13)$$

Se a relação correlação-distância não é clara, pode se assumir uma estrutura correlacional homogênea ( $\hat{\rho}=\bar{\rho}$ ), onde  $\bar{\rho}$  é o coeficiente de correlação regional médio (Martins e Stedinger, 2000). Logo, o estimador da matriz de covariância é obtido substituindo os seus componentes definidos pelas equações 9, 10, 11, 12 e 13 na equação 6.

### Análise da performance dos modelos

Na análise de performance dos modelos são utilizadas a variância média de predição, a variância do erro do modelo e uma versão modificada do coeficiente de determinação.

A variância média de predição (VMP) é utilizada como medida de quão bem os modelos OLS, WLS e GLS podem prever a variável resposta para uma nova bacia. Esta medida foi proposta por Tasker e Stedinger (1989) da seguinte forma

$$VMP = \sigma_{\delta}^2 + \frac{1}{M} \sum_{i=1}^M x_i (X^T \Lambda^{-1} X)^{-1} x_i^T \quad (14)$$

onde  $x_i$  é o vetor linha com o valor um na primeira coluna e nas restantes as características fisiográficas de uma nova bacia (Tasker e Stedinger, 1986).

No método dos mínimos quadrados clássicos (OLS) pode-se dividir a variabilidade total (SST) presente nos dados em duas componentes: a variabilidade não explicada pelo modelo (SSE) e aquela explicada pelo modelo (SSR), assim a proporção de variabilidade nos dados explicada pelo modelo é o coeficiente de determinação  $R^2$ , dado por

$$R^2 = 1 - \frac{SSE}{SST} \quad (15)$$

com  $SST = \sum (\hat{\mu}_i - \bar{\mu})^2$  e  $SSE = \sum (\hat{\mu}_i - \mu_i)^2$

Entretanto, para os modelos GLS e WLS essa fórmula não provê a quantidade desejada porque os termos SSE e SST levam em conta tanto o erro amostral como o erro do modelo. Logo, um estimador natural para estes modelos que descreve a real soma do erro quadrado do modelo é  $M\sigma_{\delta}^2$ , o que nos leva ao uso de um coeficiente de determinação como aquele proposto por Reis *et al.* (2004). Este descreve a fração da real variabilidade de  $\hat{\mu}$  explicada pelo modelo como

$$R_{GLS/WLS}^2 = 1 - \frac{M\sigma_{\delta}^2}{M\sigma_{\delta}^2 + \sum (\hat{\mu}_i - \hat{\mu}_0)^2} \quad (16)$$

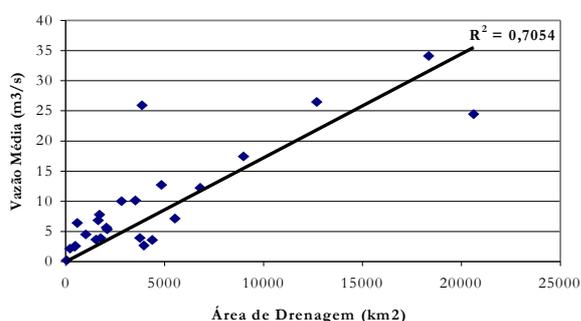
onde  $\hat{\mu}_0$  é a constante do modelo WLS ou GLS quando nenhuma variável explanatória é usada.

## RESULTADOS E DISCUSSÃO

A Figura 1 mostra a dependência linear da vazão média anual com as áreas das bacias. No entanto, não podemos considerar a vazão específica constante, pois esta consideração significaria

incorrer em erros significativos para nossa região de estudo (ver Tucci, 2002).

A fim de aplicar o método GLS, a estrutura de correlação entre as vazões das diversas bacias deve ser identificada. A estrutura correlacional da região foi analisada em maior detalhe plotando-se os coeficientes de correlação estimados e as distâncias entre as respectivas estações, gráfico este não apresentado no trabalho. Analisando este gráfico, verifica-se que há uma total dispersão dos pontos plotados, logo, conclui-se que a distância entre as estações não explica a variação da correlação cruzada entre vazões médias anuais. Logo, como sugerido por Martins e Stedinger (2000), uma estrutura de correlação homogênea foi assumida, ou seja, um coeficiente de correlação médio foi admitido.



**Figura 1 – Relação linear entre as vazões médias anuais e as áreas das bacias hidrográficas das estações fluviométricas.**

Foram testados 2048 modelos em cada método de regressão para a vazão média específica de longo período, correspondendo a todas as combinações possíveis das 11 características das bacias utilizadas como variáveis explanatórias. Destes foram selecionados previamente, baseado na precisão da estimativa dos parâmetros  $\beta$  dos modelos, 15 modelos para o GLS, 15 para o WLS e 18 para o OLS (Foram selecionados os modelos cujos parâmetros  $\beta$ 's eram significativos pelo menos ao nível de confiança de 90%). A partir desta seleção foram escolhidos os que apresentavam as menores variâncias médias de predição, as menores variâncias do erro do modelo e maiores coeficientes de determinação.

Os modelos GLS e WLS foram selecionados pelos critérios acima mencionados e utilizam combinações das variáveis explanatórias: declividade, precipitação média, comprimento do talvegue, comprimento total de drenagem, CN

médio e percentual da bacia situada no cristalino. Os valores dos parâmetros e seus desvios padrões (entre parênteses) podem ser observados nas Tabelas 2 e 3, respectivamente. Nas referidas tabelas todos os parâmetros são significativos ao nível de confiança de 99%, com exceção daqueles indicados por \* ou \*\*, os quais refere-se a parâmetros significativos a 95% e 90%, respectivamente.

Na regressão OLS foi utilizado o mesmo procedimento acima citado. De acordo com os critérios analisados destacam-se os modelos que utilizam como variáveis explanatórias: declividade, precipitação média, comprimento de talvegue, direção dominante do rio principal e percentual da bacia no cristalino. No entanto, a título de comparação a Tabela 4 apresenta os mesmos modelos utilizados no GLS e WLS.

Nas Tabelas 2, 3 e 4 são também apresentados: os coeficientes das variáveis explanatórias e os respectivos desvios padrão destes coeficientes, o coeficiente de determinação modificado ( $R^2_M$ ), a variância do erro do modelo e a variância média de predição para os modelos GLS, WLS e OLS, respectivamente. Nestes modelos as variáveis explanatórias foram centralizadas pelo valor médio da respectiva variável utilizando os valores de todas as bacias. Para os modelos GLS e WLS, a comparação do  $R^2$  com o  $R^2_M$  mostrou claramente que o uso do  $R^2$  subestima o percentual da variância explicada pelo modelo, devido esta última medida não levar em consideração o erro do modelo em estudo, como podemos observar nas equações (15) e (16).

Usando a variância média de predição e a variância do erro do modelo como critérios de performance para a seleção do modelo verifica-se que os modelos de número 3 apresentam os menores valores para ambas as medidas, sendo os valores para o GLS e WLS similares e o do OLS bem superiores aos primeiros.

No que concerne ao coeficiente de determinação modificado ( $R^2_M$ ) observa-se que os modelos (3) nas Tabelas 2, 3 e 4 apresentam altos coeficientes de determinação, com valores 92%, 92% e 74% para o GLS, WLS e OLS, respectivamente.

Contudo, para estes modelos, o teste de significância dos parâmetros para os métodos GLS e o WLS mostra que o valor do parâmetro do CN médio é significativo somente ao nível de 95%. Já o modelo de número 4 tem todos seus parâmetros significativos ao nível de 99% e conta com os valores para os critérios de performance analisados próximos do modelo 3. Além disso, o uso do

modelo 4 é mais simples pois contém três variáveis explana-

Tabela 2 – Modelos de regressão GLS selecionados para  $\mu$  em milímetros.

Modelos	Coeficientes							$R^2_M$ (%)	Variância do Erro do Modelo	Variância do Média Amostral	Variância Média de Predição
	Const'	D'	P'	CT'	CTD'	CN'	Cr'				
1	91,38 (7,09)	—	0,25 (0,05)	—	-0,004* (0,002)	—	0,96 (0,24)	80	(23,0) <sup>2</sup>	(13,0) <sup>2</sup>	(26,4) <sup>2</sup>
2	91,00 (6,68)	—	0,24 (0,04)	-0,30 (0,10)	—	—	0,90 (0,23)	84	(20,3) <sup>2</sup>	(12,4) <sup>2</sup>	(23,8) <sup>2</sup>
3	92,38 (6,09)	—	0,29 (0,04)	-0,41 (0,10)	—	-4,36* (1,81)	1,31 (0,25)	92	(15,0) <sup>2</sup>	(12,0) <sup>2</sup>	(19,2) <sup>2</sup>
4	89,80 (6,26)	7,96 (2,39)	0,19 (0,04)	—	—	—	0,74 (0,22)	90	(17,9) <sup>2</sup>	(13,2) <sup>2</sup>	(22,2) <sup>2</sup>

Tabela 3 – Modelos de regressão WLS selecionados para  $\mu$  em milímetros.

Modelos	Coeficientes							$R^2_M$ (%)	Variância do Erro do Modelo	Variância do Média Amostral	Variância Média de Predição
	Const'	D'	P'	CT'	CTD'	CN'	Cr'				
1	91,26 (6,88)	—	0,25 (0,05)	—	-0,004* (0,002)	—	0,96 (0,24)	80	(23,1) <sup>2</sup>	(12,9) <sup>2</sup>	(26,4) <sup>2</sup>
2	91,92 (6,47)	—	0,24 (0,04)	-0,31 (0,10)	—	—	0,90 (0,23)	84	(20,4) <sup>2</sup>	(12,3) <sup>2</sup>	(23,8) <sup>2</sup>
3	92,90 (5,84)	—	0,29 (0,04)	-0,41 (0,10)	—	-4,33* (1,81)	1,31 (0,25)	92	(15,0) <sup>2</sup>	(11,9) <sup>2</sup>	(19,1) <sup>2</sup>
4	90,81 (6,06)	8,02 (2,40)	0,19 (0,04)	—	—	—	0,74 (0,22)	90	(18,0) <sup>2</sup>	(13,2) <sup>2</sup>	(22,3) <sup>2</sup>

Tabela 4 – Modelos de regressão OLS selecionados para  $\mu$  em milímetros.

Modelos	Coeficientes							$R^2_M$ (%)	Variância do Erro do Modelo	Variância do Média Amostral	Variância Média de Predição
	Const	D	P	CT	CTD	CN	Cr				
1	103,51 (7,93)	—	0,30 (0,06)	—	-0,005* (0,002)	—	1,01 (0,29)	67	(39,6) <sup>2</sup>	(15,9) <sup>2</sup>	(42,7) <sup>2</sup>
2	103,51 (7,45)	—	0,28 (0,05)	-0,43 (0,14)	—	—	0,94 (0,27)	70	(37,2) <sup>2</sup>	(14,9) <sup>2</sup>	(40,1) <sup>2</sup>
3	103,51 (7,02)	—	0,33 (0,06)	-0,49 (0,13)	—	-4,88** (2,55)	1,45 (0,37)	74	(35,1) <sup>2</sup>	(15,7) <sup>2</sup>	(38,4) <sup>2</sup>
4	103,51 (6,99)	7,72 (2,03)	0,23 (0,05)	—	—	—	0,75 (0,27)	74	(35,0) <sup>2</sup>	(14,0) <sup>2</sup>	(37,7) <sup>2</sup>

Nas Tabelas 2, 3 e 4: D' – declividade, P' – precipitação média, CT' – comprimento do talvegue, CTD' – comprimento total de drenagem, CN' – CN médio e Cr' – percentual da bacia no cristalino. Todos estes centralizados por suas respectivas médias.

Desvio padrão de cada coeficiente encontra-se entre parênteses abaixo de seu valor identificado. Todos os parâmetros são significativos ao nível de confiança de 99%, exceto os assinalados (\*) 95% e (\*\*)90%.

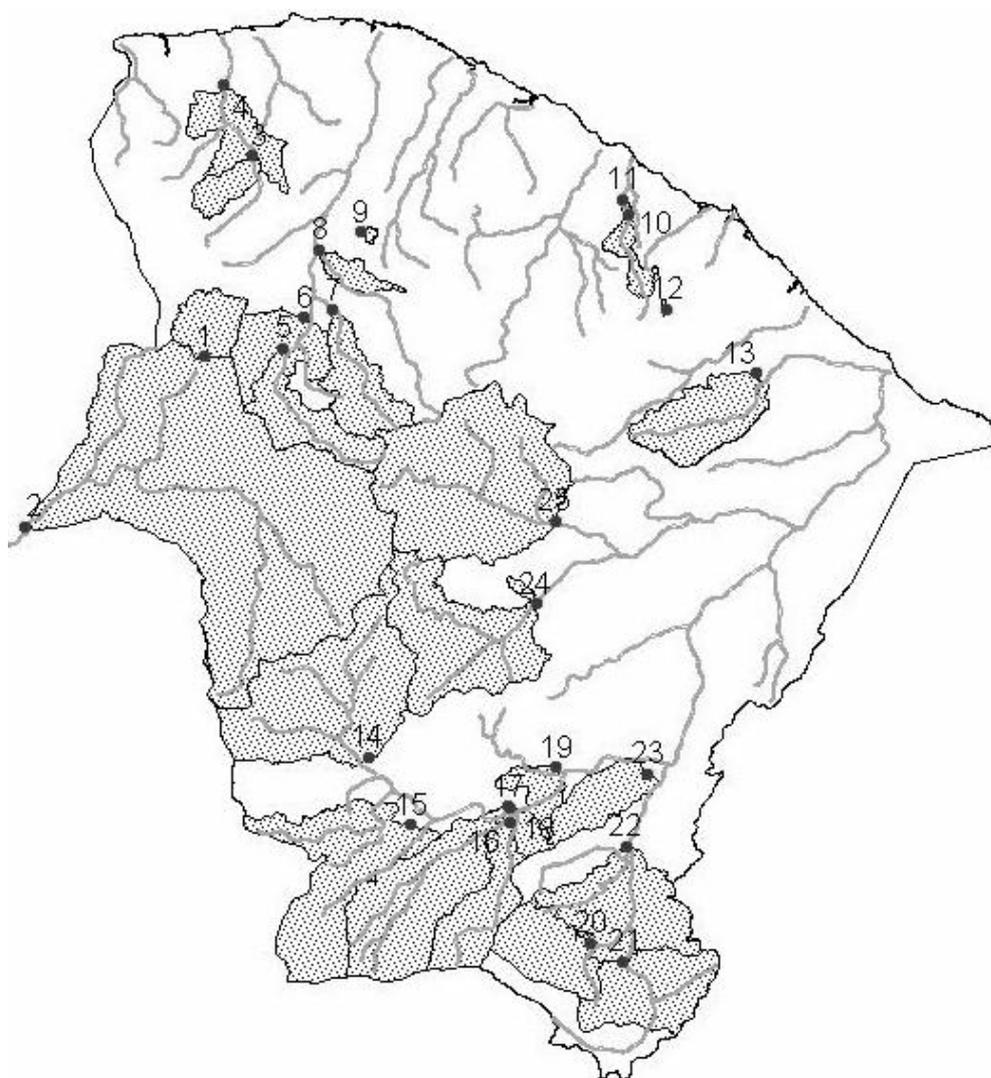


Figura 2 – Distribuição espacial das estações fluviométricas utilizadas e suas respectivas bacias de contribuição.

tórias enquanto o modelo 3 tem quatro variáveis explanatórias a serem determinadas.

Os modelos GLS e WLS apresentam resultados, bem semelhantes devido à baixa correlação cruzada entre os dados de vazão concomitantes,  $\hat{\rho}_{\{\hat{\mu}_i; \hat{\mu}_j\}} = 8 \times 10^{-4}$ . Logo, o estimador regional para a vazão específica média de longo período ( $\hat{\mu}$ ), em mm, para os dados analisados é o modelo WLS-4.

No modelo apresentado na equação (17) o valor esperado da vazão média de longo período, em mm, é expresso em termos das variáveis explanatórias e não das correspondentes normalizadas, assim este pode ser utilizado

aplicando diretamente os valores das variáveis da bacia para a qual deseja-se estimar o parâmetro.

$$E[\hat{\mu}] = -159,53 + 8,02D + 0,29P + 0,74Cr \quad (17)$$

onde a declividade  $D$  é utilizada em m/km, a precipitação média  $P$  é utilizada em mm, e o  $Cr$  é o percentual da bacia situada no cristalino.

Para a maior acuracidade dos resultados é importante utilizar as variáveis dentro da variação dos dados utilizados no modelo identificado, estando  $D$  entre 1,7 e 22,1 m/km,  $P$  entre 590 e 1152 mm, e  $Cr$  entre 0 e 100%.

A Figura 3 mostra as séries de vazões observadas e calculadas para cada uma das estações. Para a determinação dos valores

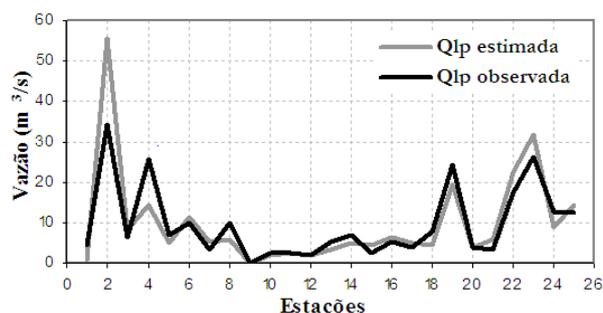
estimados das estações foi omitida uma estação após outra, seqüen

**Tabela 5 – Caracterização das 25 estações fluviométricas utilizadas neste estudo.**

ID	Nome	Rio	Município	Respons.	Operadora	Latit.	Long.	Nº Dados
01	Croatá	Macambira	Croatá	ANA	CPRM	-4:24:56	-40:54:42	18
02	Faz. boa esperança	Poti	Castelo do Piauí	ANA	CPRM	-5:13:23	-41:44:15	35
03	Moraujo	Coreaú	Moraújo	ANA	CPRM	-3:27:53	-40:41:50	19
04	Granja	Coreaú	Granja	ANA	CPRM	-3:07:15	-40:49:15	25
05	Faz. cajazeiras	Acaraú	Hidrolandia	ANA	CPRM	-4:22:43	-40:32:48	37
06	Açude araras	Acaraú	Santa Quitéria	DNOCS	DNOCS	-4:14:00	-40:27:00	13
07	Trapiá	Rch. dos Macacos	Santa Quitéria	ANA	CPRM	-4:11:42	-40:19:40	30
08	Groairas	Groairas	Groairas	ANA	CPRM	-3:54:32	-40:22:48	18
09	Faz. bela vista	Conceição	Forquilha	ANA	CPRM	-3:49:29	-40:11:15	17
10	Sítios Novos	São Gonçalo	S.Gonç. Amarante	ANA	Desativada	-3:44:59	-38:57:18	25
11	Umarituba novo	São Gonçalo	S.Gonç. Amarante	ANA	CPRM	-3:40:14	-38:58:57	13
12	Barra nova	Pacoti	Redenção	ANA	CPRM	-4:11:36	-38:46:47	18
13	Cristais	Pirangi	Cascavel	ANA	CPRM	-4:29:59	-38:21:40	32
14	Arneiroz	Jaguaribe	Arneiroz	ANA	CPRM	-6:19:26	-40:09:38	54
15	Malhada	Rch. Conceição	Saboeiro	ANA	CPRM	-6:38:44	-39:57:34	21
16	Sítio conceição	Cariús	Cariús	ANA	Desativada	-6:34:00	-39:30:00	13
17	Sítio Poço Dantas	Dos Bastidores	Cariús	ANA	CPRM	-6:33:30	-39:30:33	35
18	Cariús	Cariús	Cariús	ANA	CPRM	-6:38:50	-39:29:56	16
19	Iguatú	Jaguaribe	Iguatú	ANA	CPRM	-6:22:22	-39:17:33	62
20	Sítio lapinha	Salgado	Missão velha	ANA	CPRM	-7:12:43	-39:07:57	26
21	Podimirim	Rch. dos porcos	Milagres	ANA	CPRM	-7:17:59	-38:59:40	30
22	Lavras mangabeira	Salgado	Lavras mangabeira	ANA	Desativada	-6:45:00	-38:58:00	28
23	Icó	Salgado	Icó	ANA	CPRM	-6:24:23	-38:52:20	32
24	Senador Pompeu	Banabuiú	Senador Pompeu	ANA	CPRM	-5:35:42	-39:22:52	51
25	Quixeramobim	Quixeramobim	Quixeramobim	ANA	CPRM	-5:12:40	-39:17:33	43

Fonte: Agência Nacional de Águas - ANA

cialmente, sendo um modelo de regressão gerado a cada passo com as estações restantes para a estimativa da vazão média de longo período da estação omitida.



**Figura 3 – Vazões médias anuais de longo período observadas e estimadas pela equação (17).**

## CONCLUSÃO

No estudo de regionalização da vazão média específica de longo período para o Estado do Ceará, foram utilizadas como variáveis explanatórias sete características climáticas e fisiográficas para 25 estações fluviométricas.

Para a vazão média específica de longo período foi utilizado o método de regressão GLS (*Generalized Least Squares* – mínimos quadrados generalizados) e seus casos particulares, WLS (*Weighted Least Squares* – mínimos quadrados ponderados) e OLS (*Ordinary Least Squares* – mínimos quadrados). Foram analisadas todas as combinações possíveis para as onze características das bacias, sendo testados 2048 modelos para cada método de regressão. Entre estes foram selecionados os modelos cujos parâmetros fossem

significantes estatisticamente ao nível de 95%, além de apresentarem pequenos valores para as variâncias do erro dos modelos e variância média de predição.

Os modelos do tipo GLS e WLS obtiveram valores muito parecidos devido à baixa correlação cruzada entre vazões, tornando quase indiferente à utilização de qualquer um dos métodos. Por simplicidade foi adotado como estimador regional da vazão específica média de longo período o modelo WLS, o qual utiliza como variáveis explanatórias a declividade, a precipitação média e percentual da bacia situada no cristalino.

Assim, para essa região de estudo a aplicação da estrutura GLS não determinou melhorias na acuracidade do modelo regional. No entanto com o uso desta para outros locais pode-se obter modelos com significativas melhorias sobre os WLS.

## AGRADECIMENTOS

Os autores agradecem aos dois revisores da RBRH, cujas sugestões foram prontamente aceitas e com certeza serviram para melhorar a clareza dos resultados aqui apresentados. Os autores ainda agradecem as sugestões dos Drs. Robin Thomas Clarke e Dirceu Silveira Reis Junior para a melhoria do presente artigo.

## REFERÊNCIAS

- EUCLYDES, H. P. & FERREIRA, P. A., *Recursos hídricos e suporte tecnológico a projetos hidroagrícolas: Bacia do Alto São Francisco*. Viçosa: UFV; Belo Horizonte: RURALMINAS; Brasília: ANA. 2001.
- EUCLYDES, H. P., FERREIRA, P. A., FARIA FILHO, R. F. & DA SILVA, M. V. J., Regionalização hidrológica na bacia do rio Paranaíba, Minas Gerais. In: *Anais do XV Símposio Brasileiro de Recursos Hídricos*, Curitiba: Associação Brasileira de Recursos Hídricos - 2003. CD-ROM.
- GÓES, V. C. & CAMPANA, N. A., Regionalização de vazões no Distrito Federal. In: *Anais do XVI Símposio Brasileiro de Recursos Hídricos*, Aracajú: Associação Brasileira de Recursos Hídricos - 2000. CD-ROM.
- KROLL, C.N., & STEDINGER, J.R. Development of Regional Regression Relationships with Censored Data, *Water Resources Research* 35(3), 775-784, 1999.
- MADSEN, H. & ROSBJERG, D., Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modeling, *Water Resources Research*, 33(4), 771-782, 1997.
- MADSEN, H., MIKKELSEN, P. S., ROSBJERG, D. & HARREMOES, P., Regional Estimation of Rainfall Intensity-Duration-Frequency Curves Using Generalized Least Squares Regression of Partial Duration Series Statistics, *Water Resources Research* 38(11), 1239, doi:10.129/2001 WR001125, 2002.
- MARKUS, M., KNAPP, H. V. & TASKER, G. D., Entropy and generalized least square methods in assessment of regional value of streamgages, *Journal of Hydrology*, 283, 107-121, 2003.
- MARTINS, E.S. & STEDINGER, J. R., Generalized Maximum Likelihood GEV Quantile Estimators for Hydrologic Data, *Water Resources Research*, 28(11), 3001-3010, 2000.
- MARTINS, E.S., & STEDINGER, J. R., Cross-correlation among estimators of shape, *Water Resources Research*, 38(11), 1252-1258, 2002a.
- MARTINS, E.S., & STEDINGER, J. R., Efficient Regional Estimates of LP3 Skew using GLS Regression, *Proceedings of the ASCE Conference on Water Resources Planning and Management*, May 19-22, 2002b.
- MÜLLER, F. & FILL, H. D., Quantis de cheia GEV regionais – uma aplicação prática, *RBRH*, vol. 2, n. 2, pp. 53-64, 1997.
- PANDEY, G. R. & NGUYEN, V.-T.-V., A comparative study of regression based methods in regional flood frequency analysis, *Journal of Hydrology*, 225, 92-101, 1999.
- PEDROLLO, M. C., & FERREIRA, P. H., Regionalização de vazões na bacia da lagoa mirim. In: *Anais do XV Símposio Brasileiro de Recursos Hídricos*, Curitiba: Associação Brasileira de Recursos Hídricos - 2003. CD-ROM.
- PINTO, E. J. A. & ALVES, M. M. S., Regionalização de vazões da bacia do alto São Francisco. In: *Anais do XV Símposio Brasileiro de Recursos*

- Hídricos, Curitiba: Associação Brasileira de Recursos Hídricos - 2003. CD-ROM.
- PORTO, M. M., 2003. Regionalização de vazões médias de longo período para o Estado do Ceará. 68p *Dissertação de Mestrado em Agronomia, área de concentração em irrigação e drenagem*, UFC, 2003.
- REIMERS, W., Estimating Hydrological parameters from basin characteristics. *Regionalization in Hydrology. (proceedings of the Ljubljana Symposium, April 1990)*. IAHS Publ. N° 191. 1990.
- REIS, JR., D. S., STEDINGER, J. R. & MARTINS, E. S., Bayesian GLS Regression with Application to LP3 Regional Skew Estimation. In: *ASCE World Water e Environmental Resources Congress 2003 e Related Symposia, Philadelphia, PA, 2003*.
- REIS, JR., D. S., STEDINGER, J. R. & MARTINS, E. S., Operational Bayesian GLS Regression for Regional Hydrologic Analyses. In: *ASCE World Water & Environmental Resources Congress 2004, Salt Lake City, PA, 2004*.
- REIS JR., D. S., STEDINGER, J. R AND MARTINS, E. S., Bayesian GLS Regression with application to LP3 Regional Skew Estimation, *Water Resources Research*, 2005 (in press).
- RIGGS, H. C., Estimating flow characteristics at ungauged sites. *Regionalization in Hydrology. (proceedings of the Ljubljana Symposium, April 1990)*. IAHS Publ. N° 191. 1990.
- STEDINGER, J.R. & TASKER, G. D., Regional hydrologic analysis. 1. Ordinary, weighted, and generalized least squares compared. *Water Resources Research* 21(9), 1421-1432, 1985.
- STEDINGER, J.R. & TASKER, G. D., Regional hidrologic analysis. 2. Model-error estimators, estimation os sigma and log-Pearson type 3 distribution. *Water Resources Research* 22(10), 1487-1499. 1986a.
- STEDINGER, J.R. & TASKER, G. D., Correction to Regional hydrologic analysis. 1. Ordinary, weighted, and generalized least squares compared. *Water Resources Research* 22(5), 844. 1986b.
- STUART, A. & KEITH, J., *Kendall's advanced theory of statistics, volume 1: Distribution Theory*, 6° edição, Washington DC, USA. 1994.
- TASKER, G. D. & STEDINGER, J.R., Regional skew with weighted LS regression. *Journal of Water Resources Planning and Management*. 112-2, 225-237. 1986.
- TASKER, G. D. & STEDINGER, J.R., An operational GLS model for hydrologic regression. *Journal of Hydrology*. 111, 361-375. 1989.
- TUCCI, C. E. M., (coord.) *Regionalização de vazões do Rio Grande do Sul*. Porto Alegre: Companhia Estadual de Energia Elétrica (CEEE)/Instituto de Pesquisas Hidráulicas/ UFEGS. 1°v. 1991.
- TUCCI, C. E. M., *Regionalização de vazões*, Ed. Universidade/UFRGS, 1° edição, Porto Alegre-RS. 2002.
- VOGEL, R. M.; WILSON, I. & DALY, CHRIS, Regional regression models of annual streamflow for the United States. *Journal of Irrigation and Drainage Engineering*, vol. 125, N°3. 1999.

### ***Regional Regression Models of Annual Streamflow Ceará State***

#### **ABSTRACT**

*Water resources studies, particularly for water supply, require estimates of annual streamflow, very often for ungauged basins. Thus, regionalization is needed to determine the water yield of a basin in the State of Ceará with little or no data. This paper introduces the use of Generalized Least Squares (GLS) for regionalizing annual streamflow, by which it is possible to take into account not only the sampling variance and the model error, but also the inter-site annual streamflow correlation. The GLS results were compared with the results obtained with Weighted Least Squares (WLS) and Ordinary Least Squares (OLS) regressions. Different models were generated by using different combinations of climatic and physiographic characteristics of 25 basins as explanatory variables. These models were analyzed using as criteria the standard deviations of all parameters, the model error variance, the average variance of prediction and a pseudo determination coefficient. Due to the low inter-site correlation observed between concurrent annual streamflow records, both GLS and WLS have a similar performance. The regional model adopted for the annual streamflow of Ceará State was a WLS with slope, average precipitation and percentage of the watershed located in the crystalline soil as explanatory variables.*

*Key-words: Regionalization; water supply; mean streamflows*